Geometric integrators for ODEs

2006 J. Phys. A: Math. Gen. 39 5251

(http://iopscience.iop.org/0305-4470/39/19/S01)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **39** (2006) 5251–5285 doi:10.1088/0305-4470/39/19/S01

Geometric integrators for ODEs

Robert I McLachlan[1] **and G Reinout W Quispel**[2]

[1] Massey University, Palmerston North, New Zealand
[2] La Trobe University, Bundoora, VIC 3083, Australia

Received 11 October 2005, in final form 22 December 2005
Published 24 April 2006
Online at stacks.iop.org/JPhysA/39/5251

Abstract

Geometric integration is the numerical integration of a differential equation, while preserving one or more of its 'geometric' properties exactly, i.e. to within round-off error. Many of these geometric properties are of crucial importance in physical applications: preservation of energy, momentum, angular momentum, phase-space volume, symmetries, time-reversal symmetry, symplectic structure and dissipation are examples. In this paper we present a survey of geometric numerical integration methods for ordinary differential equations. Our aim has been to make the review of use for both the novice and the more experienced practitioner interested in the new developments and directions of the past decade. To this end, the reader who is interested in reading up on detailed technicalities will be provided with numerous signposts to the relevant literature.

PACS number: 45.10.−b

1. Introduction

1.1. Geometric integration

The topic of this paper is structure-preserving numerical integration methods for ordinary differential equations (ODEs)[3]. We will concentrate mainly on Hamiltonian equations and on methods that preserve their symplectic structure, first integrals, symmetries, or phase-space volume. We also consider Hamiltonian systems with a dissipative perturbation. More details on these topics and extensive references can be found in several recent books [47, 72, 109, 115] and survey articles [18, 19, 36, 48, 61, 74, 76, 89, 90]. Physical applications are ubiquitous, varying from particle accelerators [38] to systems with multiple time scales [63] and many areas in between (see, e.g., applications in molecular dynamics [72], quantum mechanics [120], celestial mechanics [70, 123, 124] and complex systems [1]).

[3] Geometric integration for *partial* differential equations is addressed by Bridges and Reich [16].

In recent decades the theme of structure preservation has emerged in nearly every branch of numerical analysis, most notably in numerical differential equations and numerical linear algebra. The motivations for preserving structure include

(i) it may yield methods that are faster, simpler, more stable and/or more accurate, as in methods for structured eigenvalue problems and for some types of ODEs;
(ii) it may yield methods that are more robust and yield qualitatively better results than standard methods, even if the standard (quantitative) numerical errors are not smaller;
(iii) it may suggest new types of calculations previously thought to be impossible, as in the long-time integration of Hamiltonian systems;
(iv) it may be essential to obtain a useful (e.g. convergent) method, as in the discrete differential complexes used in electromagnetism [2, 56]; and
(v) the development of general-purpose methods may be nearing completion, as in Runge–Kutta methods for ODEs.

Most geometric properties are not preserved by traditional numerical methods. The requirement of structure preservation necessarily restricts the choice of method and may impose a cost. In each case the costs and benefits have to be weighed. In some cases (as in some types of structured matrices and some classes of differential equations) it can be impossible to preserve the given structure [41], while in others it is unknown whether or not such a method exists ([89], section 6).

These possibilities are illustrated by the two most popular geometric integrators. The *leapfrog method* [48] for the simple mechanical system $\dot{q} = p, \dot{p} = -\nabla V(q)$ with the potential energy $V(q)$ is given by

$$\begin{aligned} q_{k+1/2} &= q_k + \tfrac{1}{2}\tau p_k, \\ p_{k+1} &= p_k - \tau \nabla V(q_{k+1/2}), \\ q_{k+1} &= q_{k+1/2} + \tfrac{1}{2}\tau p_{k+1}. \end{aligned} \tag{1}$$

Here (and below) τ is the time step. The leapfrog method (1) is explicit and second order[4], despite using only one evaluation of the force $-\nabla V$ per time step (compared to two for a second-order Runge–Kutta method). It does not require the storage of any intermediate values. It is symplectic, preserving the canonical symplectic form $\sum_i \mathrm{d}q_i \wedge \mathrm{d}p_i$, and time reversible. The total energy is not conserved, but nor does the energy error grow with time, even for time steps of the same order of magnitude as the frequencies of the system. Invariant sets such as periodic, quasiperiodic and chaotic orbits are well preserved in phase space. Leapfrog preserves linear and angular momentum (where applicable) up to round-off error. On the other hand, it does become unstable if the time step is too large (larger than T/π, where T is the shortest natural period of the system), and it does not preserve quadratic first integrals such as $\|q\|^2$ where applicable.

The *midpoint rule* for the first-order system $\dot{x} = f(x), x \in \mathbb{R}^n$, is given by

$$x_{k+1} = x_k + \tau f\left(\frac{x_k + x_{k+1}}{2}\right). \tag{2}$$

It is implicit, and hence more expensive (uses more CPU time) than the leapfrog method. But it too is symplectic, this time for *any* Hamiltonian system with any constant symplectic or Poisson structure. It preserves any linear symmetries of the system, and is time reversible with respect to any linear reversing symmetry of phase space. It preserves not just linear and angular momentum but any quadratic first integrals of the system. It is linearly stable for all

[4] The local error is $\mathcal{O}(\tau^3)$ per time step, and the global error after integrating for a fixed time is $\mathcal{O}(\tau^2)$.

time steps. Many more accurate or more widely applicable variants of these two methods have been developed.

On the other hand, no method can preserve energy and symplecticity in general [41]. No method is known that preserves volume in phase space and all linear symmetries [89].

The traditional problem in numerical ODEs is to compute the solution to the initial value problem $\dot{x} = f(x)$, $x(0) = x_0$, at a fixed time $t = k\tau$, to within a given global error $\|x_k - x(t)\|$, as efficiently as possible. The class of method, its order and local error, and choice of time steps are all tailored to this end [20, 21, 49, 57]. In contrast, a typical application of a geometric integrator is to fix a (sometimes moderately large) time step and compute very long orbits, with perhaps many different initial conditions. Although the global error of each orbit is large, the phase portrait that emerges can still be close, in some sense, to the phase portrait of the differential equation. Once small global errors are abandoned; however, the convergence of each feature (e.g. energy, other first integrals, periodic and quasiperiodic orbits, phase-space averages) has to be examined separately.

While the leapfrog and midpoint methods (equations (1) and (2), respectively) are sufficient for many applications, there have been many developments since their geometric properties were discovered. Research in geometric integration has focused on three main areas: (i) developing new types of integrators, and integrators preserving new structures; (ii) improving the efficiency of the integrators, by finding methods of higher order, smaller local errors, or allowing larger time steps, and by finding methods tuned for special classes of systems and (iii) studying the behaviour of the integrators, especially their long-time dynamics and the extent to which the phase portrait of the system is preserved.

1.2. An illustrative example

We illustrate these themes by applying them to a venerable example, the Hénon–Heiles system. This is a Hamiltonian system with two degrees of freedom and Hamiltonian

$$\begin{aligned} H &= T(p) + V_1(q) + V_2(q), & T(p) &= \tfrac{1}{2}\bigl(p_1^2 + p_2^2\bigr), \\ V_1(q) &= \tfrac{1}{2}\bigl(q_1^2 + q_2^2\bigr), & V_2(q) &= q_1^2 q_2 - \tfrac{1}{3} q_2^3. \end{aligned} \tag{3}$$

The standard geometric integrator for this system is the leapfrog method, equation (1). Numerous articles have illustrated the preservation of the phase portrait by the leapfrog method, the convergence of its invariant sets (e.g. quasiperiodic and chaotic) as $\tau \to 0$, and the relative efficiency of various higher-order methods depending on the time step.

Here we take a different approach and illustrate geometric integrators as they are often used in practice. That is, rather than investigate what happens as $\tau \to 0$, we study long runs with a relatively large time step. We will also take advantage of the special structure of (3) and use a new, highly tuned fourth-order method.

We split H, not into kinetic and potential energies, but into linear ($T(p) + V_1(q)$) and nonlinear ($V_2(q)$) parts. The flow[5] of each is easily calculated, giving an explicit, second-order generalized leapfrog algorithm, equation (7) below ('LF2' in the figures). We compare this to a fourth-order (composition, i.e. leapfrog-like) method with six force evaluations per time step due to Blanes and Moan [14] that has been tuned for this type of splitting (see equation (56)) ('LF4'), and to two non-geometric, standard algorithms: the 'classic' fourth-order Runge–Kutta method with fixed step size ('RK4'), and a modern, adaptive fifth-order Runge–Kutta method, MATLAB's `ode45` ('RK45'). The step sizes and tolerances were adjusted so that

[5] The mapping in phase space defined by the exact solution of an ODE.

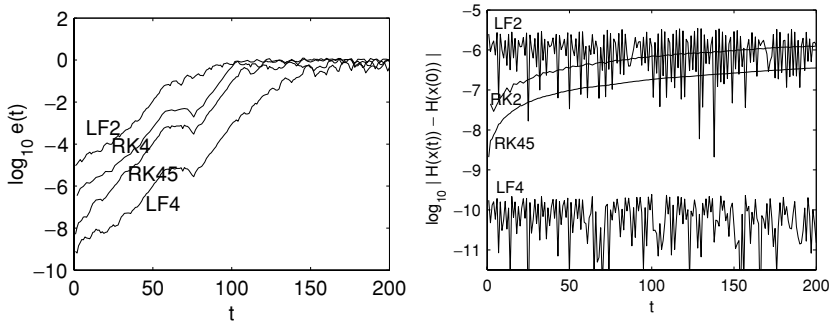

Figure 1. Advantage of newer, highly tuned geometric integrators. Global error (left) and energy error (right) for the Hénon–Heiles system with initial condition $(q_1, q_2, p_1, p_2) = (0.25, 0.5, 0, 0)$, which leads to a chaotic orbit that covers most of the energy surface. RK4: classical fourth-order Runge–Kutta method. RK45: MATLAB's `ode45`, a modern, adaptive, fifth-order Runge–Kutta method. LF2: second-order leapfrog, a symplectic integrator with one force evaluation per time step. LF4: a fourth-order symplectic integrator with six force evaluations per time step.

each method used the same number of force evaluations[6]. Our aim will be to show that, for this application, both geometric methods (LF2 and LF4) are better than both non-geometric ones (RK4 and RK45), and that, between the geometric integrators, LF4 is better than LF2. While the latter observation would be obvious if τ were very small, it is non-trivial that this also remains the case for moderately large values of τ.

Figure 1 shows the results for a *chaotic* orbit integrated to time $t = 200$. The global errors grow *exponentially* in time for all integrators, as expected. The time step for LF4 is a moderate $\tau = 0.1$. The energy errors grow *linearly* for RK4 and RK45 but are *bounded* for the symplectic integrators LF2 and LF4. Despite the relatively large time step, the fourth-order method LF4 easily beats the standard leapfrog LF2. (In fact, this advantage still persists even at $\tau = 1$.) Although we report energy and global errors, we specifically do *not* want to imply that these are directly related to the preservation of the phase portrait. Although it is necessary in a symplectic integration to check that τ is small enough that the energy errors are bounded, this is not sufficient to ensure that the errors in any other quantity are small.

Figure 2 shows the results for a *quasiperiodic* orbit integrated to time $t = 2000$. The time step for LF4 is now $\tau = 1$, which is of the same order as the period 2π of the harmonic oscillators in $T(p) + V_1(q)$. RK4 and RK45 now show a *quadratic* growth of global errors, while LF2 and LF4 show a *linear* growth.

Figure 3 shows the results for a *small-amplitude quasiperiodic* orbit integrated to time $t = 2000$ with $\tau = 1$. Now the global errors of RK4 and RK45 are growing *linearly*[7]. Although the small amplitude helps RK4 and RK45, it helps LF2 and LF4 even more, because they capture the linear dynamics exactly and hence have global errors that vanish as $q \to 0$. (See also section 4.2.)

The advantage of the modern method LF4 over leapfrog LF2 shown in figures 1, 2 and 3 would disappear if we had used instead the much cruder fourth-order method $\varphi(z\tau)\varphi((1-2z)\tau)\varphi(z\tau)$, where $z = (2 - 2^{1/3})^{-1}$ and φ is the standard leapfrog (equation (1)). This method, the first fourth-order symplectic integrator to be discovered [31, 116, 125] has errors 100 to 1000 times larger than LF4. LF4 and its higher-order variants [13]) is also much

[6] The number of force or vector field evaluations is often used as a machine-independent measure of the amount of work an algorithm needs to do.

[7] The second period of the orbit is very large. On the time scale shown, only the amplitudes of the harmonic oscillators drift; but this does not change their frequencies.

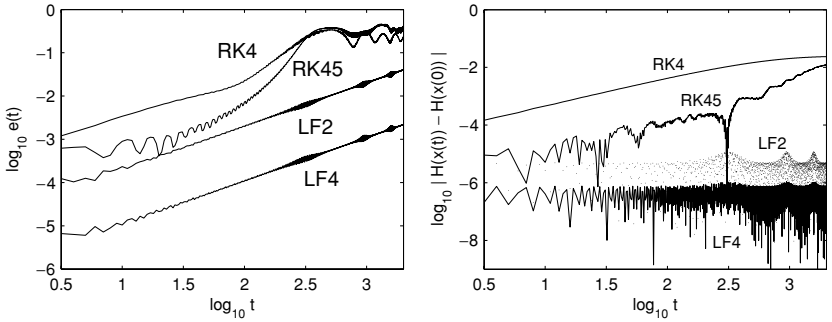

Figure 2. Advantage on quasiperiodic orbits. Global error (left) and energy error (right) for the Hénon–Heiles system with initial condition $(q_1, q_2, p_1, p_2) = (0.1, 0.2, 0, 0)$, which leads to a quasiperiodic orbit.

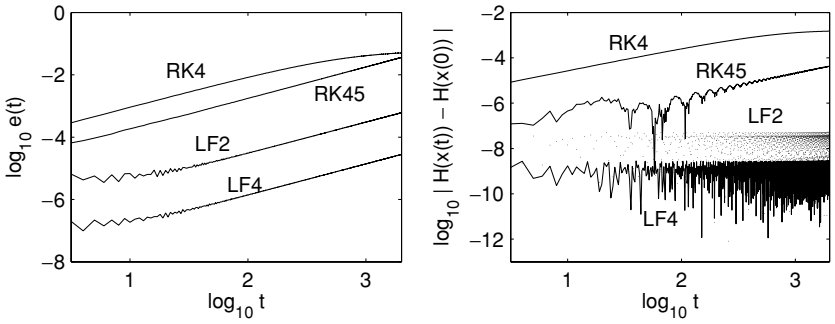

Figure 3. Advantage on nearly-integrable systems. Global error (left) and energy error (right) for the Hénon–Heiles system with initial condition $(q_1, q_2, p_1, p_2) = (0.025, 0.05, 0, 0)$, which leads to a small-amplitude quasiperiodic orbit.

better than the earliest methods tuned for simple mechanical systems (see section 4.1), such as those in McLachlan and Atela [83].

Figure 4 shows the dynamics of this small-amplitude orbit out to $t = 20000$. The periods of the orbit are about 2π and 28 084. Remarkably, the integrator LF4 captures this motion with a relative frequency error of about 1.4×10^{-4} at a time step of $\tau = 1$. (In the figure, it is indistinguishable from the exact orbit.) The orbit for RK4 decays to 0, while the orbit for the adaptive method RK45 is more complex, but still bears little relation to the actual orbit; moreover, the dynamics of RK45 depends sensitively on the chosen tolerance.

Returning to the simulation of chaotic orbits, for an orbit $x(t)$ let $v(t)$ be a tangent vector (i.e. v obeys $\dot{v} = f'(x)v$) and let $\lambda(t)$ be the rate of expansion in the direction $v(t)$, i.e. $\frac{\mathrm{d}}{\mathrm{d}t}\|v\| = \lambda\|v\|$ or $\lambda = (v^{\mathrm{T}} f'(x)v)/(v^{\mathrm{T}}v)$. For generic $v(0)$, the mean $\bar{\lambda}$ of $\lambda(t)$ is the largest Lyapunov exponent of the orbit.

Figure 5 shows the probability density function $p(\lambda)$ of λ as calculated by LF2 for the chaotic orbit of figure 1 calculated for time $t = 5 \times 10^5$. For symplectic integrators like LF2, the finite-time estimates of the Lyapunov exponents come in exact opposite $(+/-)$ pairs; in systems with widely varying exponents, this can lead to greater reliability in estimating them [91]. The exact value of the Lyapunov exponent $\bar{\lambda}$ is 0.0812; an integration with $\tau = 0.5$ (close to the largest stable value of the time step) gives $\bar{\lambda} = 0.0724$; an integration with $\tau = 0.125$ gives 0.0819. The sampling errors are about 5×10^{-4} in $\bar{\lambda}$ and less than 0.04 in $p(\lambda)$. This illustrates the convergence of statistical properties of chaotic orbits. However, note that the

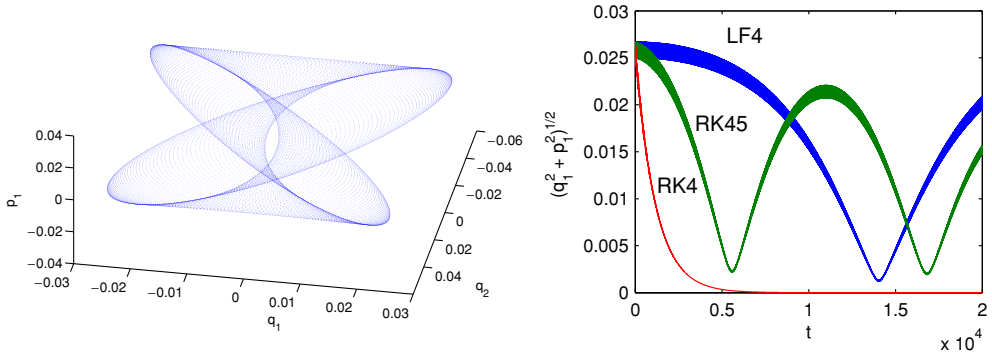

Figure 4. A longer run with the initial conditions of figure 3. Left: a projection of the quasiperiodic orbit in phase space. Right: the slow exchange of energy between the two oscillators is only captured by the leapfrog scheme. The method LF4 yields a numerical solution that is indistinguishable from the exact orbit on this scale.

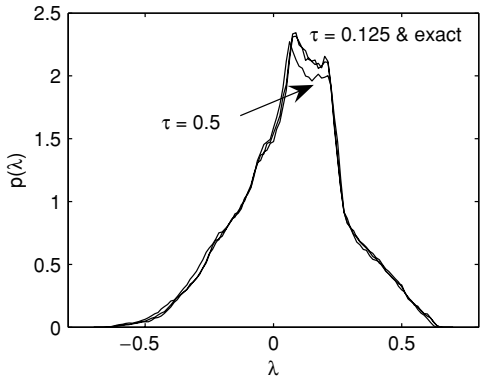

Figure 5. The probability distribution $p(\lambda)$ of the rate of expansion λ of the chaotic orbit of the Hénon–Heiles system shown in figure 2, calculated by leapfrog (LF2) for time $t = 5 \times 10^5$. The results for $\tau = 0.125$ agree with the exact value to within the sampling error.

error in these observables (up to 10% at $\tau = 0.5$) is much larger than the *energy* error, which is only 0.6% at $\tau = 0.5$. Lyapunov exponents are almost impossible to calculate reliably with non-geometric integrators because the results are contaminated by sampling errors if t is too small and excessive dissipation if t is too large. We therefore do not present any results for $\bar{\lambda}$ calculated by non-geometric methods.

1.3. Outline of the paper

In section 2 we introduce the main types of methods used in geometric integration of ODEs: splitting and composition, Runge–Kutta and its variants, projection, variational and multistep methods. In section 3 we survey the main types of equations to which the methods are applied: Hamiltonian and Poisson systems, volume-preserving systems, and systems with symmetries, reversing symmetries, first integrals and foliations. Here the properties considered are all intrinsic (i.e., coordinate independent), geometric properties that control and characterize the dynamics of the system. Some dissipative, non-Hamiltonian types of systems are also covered. In section 4 we specialize to consider properties that, although not intrinsic, nevertheless are numerically significant because they allow vastly more efficient methods: examples are

separable or polynomial Hamiltonian systems, nearly integrable systems, and systems with multiple time scales[8]. In section 5 we cover the behaviour of geometric integrators, especially their stability, the convergence of invariant sets in phase space, and the rate of error growth of the solution or of the first integrals of the system. Section 6 surveys other computational problems: solving boundary-value and optimization problems, non-trajectory methods for analysing dynamical systems and the computation of Poincaré maps.

2. The main geometric integrators for ODEs

In this section we briefly define the main classes of numerical integrators that have been found to have useful geometric properties.

2.1. Splitting and composition methods

With the phase space M, differential equation $\dot{x} = f(x)$, $x \in M$, and f a vector field on M, splitting methods involve three steps: (i) choosing a set of vector fields f_i such that $f = \sum_{i=1}^{n} f_i$; (ii) integrating either exactly or approximately each f_i; and (iii) combining these solutions to yield an integrator for f. For example, writing the solution of the ODE $\dot{x} = f(x), x(0) = x_0$ as $x(t) = \mathrm{e}^{tf}(x_0)$, we might use the first-order-accurate composition method

$$\mathrm{e}^{\tau f_1}\,\mathrm{e}^{\tau f_2}\cdots\mathrm{e}^{\tau f_n} = \mathrm{e}^{\tau f} + \mathcal{O}(\tau^2). \tag{4}$$

The pieces f_i should be simpler than the original vector field f; most commonly, they can be integrated exactly in closed form, as in equation (4). If so, and if the f_i lie in the same linear space of vector fields as f (e.g. of Hamiltonian vector fields, or of vector fields preserving a symmetry, reversing symmetry, first integral etc.), then the composition method (4) is explicit and preserves the appropriate geometric property automatically. For a detailed review of splitting methods, see [89].

The advantages of splitting methods are that they are usually simple to implement, explicit, and can preserve a wide variety of structures. Their disadvantages are that, except in particular special cases, the higher-order methods are relatively expensive, and that the splitting may break some particular property, such as a symmetry, that one may want to preserve.

Details specific to particular geometric properties are deferred to section 3.1; here, we discuss general aspects of composition methods. The composition (4) is only first order. The order can be increased by including more exponentials in a time step. For a splitting into two parts, $f = f_1 + f_2$, we have the general nonsymmetric composition

$$\mathrm{e}^{a_m\tau f_1}\,\mathrm{e}^{b_m\tau f_2}\cdots\mathrm{e}^{a_1\tau f_1}\,\mathrm{e}^{b_1\tau f_2}\,\mathrm{e}^{a_0\tau f_1}. \tag{5}$$

By convention, we only count the evaluations of the flow of f_2, and refer to equation (5) as an m-stage method. The number of stages and the coefficients a_i and b_i are to be chosen to ensure that the method has some order p, i.e. $\varphi = \mathrm{e}^{\tau(f_1+f_2)} + \mathcal{O}(\tau^{p+1})$. The method $\varphi(\tau)$ is *time symmetric* if

$$\varphi(\tau)\varphi(-\tau) = 1 \tag{6}$$

for all τ. It is easy to find time-symmetric methods, for if $\varphi(\tau)$ is any method of order p, then $\varphi\left(\frac{1}{2}\tau\right)\varphi^{-1}\left(-\frac{1}{2}\tau\right)$ is time symmetric and of order at least $p+2$ (if p is even) or at least $p+1$ (if p is odd). In general, if φ is an explicit method, then φ^{-1} is implicit. However, if

[8] In the latter two examples, an intrinsic property emerges in an asymptotic limit.

φ is a composition of (explicitly given) flows, then φ^{-1} is also explicit. Applied to the basic composition (4), this leads to the explicit 'generalized leapfrog' method of order 2,

$$\mathrm{e}^{\frac{1}{2}\tau f_1}\cdots\mathrm{e}^{\frac{1}{2}\tau f_n}\,\mathrm{e}^{\frac{1}{2}\tau f_n}\cdots\mathrm{e}^{\frac{1}{2}\tau f_1}, \tag{7}$$

which is widely used in many applications. For many purposes it is the most sophisticated method needed. From the flow property $\mathrm{e}^{\tau f}\,\mathrm{e}^{\sigma f} = \mathrm{e}^{(\tau+\sigma)f}$, the two central stages coalesce, and the last stage coalesces with the first stage of the next time step, so we evaluate $2n-2$ flows per time step, or $2-2/n$ as much as the first-order method equation (4). Therefore there is an advantage in searching for splittings for which the number n of parts is small (say 2 or 3).

The simplest way to increase the order is to iteratively apply the following construction. If $\varphi(\tau)$ is a time-symmetric method of order $2k>0$, then the method

$$\varphi(\alpha\tau)^n\varphi(\beta\tau)\varphi(\alpha\tau)^n, \qquad \alpha = (2n-(2n)^{1/(2k+1)})^{-1}, \qquad \beta = 1-2n\alpha \tag{8}$$

is time symmetric and has order $2k+2$. This yields methods of order 4 containing three applications of equation (7) when $n=1$, of order 6 containing nine applications of equation (7) when $n=2$, and so on [31, 116, 125]. The truncation errors of these methods tend to be rather large, however, although the order 4 methods with $n\geqslant 2$ can be competitive and have some interesting properties [82]. Interestingly, the $n=k+1$ methods were also recommended by their discoverers [31].

A good, general purpose fourth-order method is given by (8) with $n=2$ and $k=1$.

To get methods with fewer stages, one can use the composition

$$\varphi(a_1\tau)\cdots\varphi(a_m\tau)\cdots\varphi(a_1\tau), \tag{9}$$

where $\varphi(\tau)$ is any time-symmetric method. See [114] for the best known high-order methods of this type. Methods of orders 4, 6, 8 and 10 require at least 3, 7, 15 and 31 stages (i.e., applications of φ), but in practice the methods that give the smallest error for a given amount of work have even more stages.

It is also possible to consider composition methods formed from an arbitrary (usually first order and non-time-symmetric) method $\varphi(\tau)$ and its inverse,

$$\prod_{i=1}^{m}\varphi^{-1}(-c_i\tau)\varphi(d_i\tau), \tag{10}$$

which is time symmetric if $c_i = d_{m+1-i}$ for all i. In fact, methods of class (10) are in 1–1 correspondence with methods of class (5), so a high-order method derived for two flows, such as the typical case of the flows of the kinetic and potential parts of the Hamiltonian, can be directly applied to any more general situation [80]. (A method (10) generates a method (5) by choosing $\varphi(\tau)=\mathrm{e}^{\tau f_2}\,\mathrm{e}^{\tau f_1}$.)

A good fourth-order method of this class, due to Blanes and Moan [14], is given by equation (5) together with $m=6$ and

$$\begin{aligned}
a_0 &= a_6 = 0.079\,203\,696\,431\,1957,\\
a_1 &= a_5 = 0.353\,172\,906\,049\,774,\\
a_2 &= a_4 = -0.042\,065\,080\,357\,7195,\\
a_3 &= 1-2(a_0+a_1+a_2),\\
b_1 &= b_6 = 0.209\,515\,106\,613\,362,\\
b_2 &= b_5 = -0.143\,851\,773\,179\,818,\\
b_3 &= b_4 = \tfrac{1}{2}-(b_1+b_2),
\end{aligned} \tag{11}$$

or, equivalently, by equation (10) together with

$$\begin{aligned} b_1 &= c_6 = 0.079\,203\,696\,431\,1957, \\ b_2 &= c_5 = 0.222\,861\,495\,867\,6077, \\ b_3 &= c_4 = 0.324\,648\,188\,689\,7062, \\ b_4 &= c_3 = 0.109\,688\,477\,876\,7498, \\ b_5 &= c_2 = -0.366\,713\,269\,047\,4257, \\ b_6 &= c_1 = 0.130\,311\,410\,182\,1663. \end{aligned} \tag{12}$$

The coefficients of τ, called variously a_i, b_i, c_i and d_i above, cannot all be positive if the order of the method is greater than 2 [7, 111]. This can prevent the application of this kind of method to dissipative systems. If one allows *complex* coefficients, the number of stages can be greatly reduced and the real parts of the coefficients can be positive, even at high order; this requires complexifying phase space but it has yielded promising results [27]. It could be a natural choice for complex Hamiltonian systems.

Various extensions have been considered in order to find more efficient methods. We consider the two that apply most generally here. The first is the use of a 'corrector' (also known as processing or effective order) [6, 9, 124]. Suppose the method φ can be factored as

$$\varphi = \chi \psi \chi^{-1}. \tag{13}$$

Then to evaluate n time steps, we have $\varphi^n = \chi \psi^n \chi^{-1}$, so only the cost of ψ is relevant. The maps φ and ψ are conjugate by the map χ, which can be regarded as a change of coordinates. Many dynamical properties of interest (to a theoretical physicist, *all* properties of interest) are invariant under changes of coordinates; in this case we can even omit χ entirely and simply use the method ψ. For example, calculations of Lyapunov exponents, phase space averages, partition functions, existence and periods of periodic orbits, etc, fall into this class. Initial conditions, of course, are not invariant under changes of coordinates, so applying χ *is* important if one is interested in a particular initial condition, such as one determined experimentally.

The simplest example of a corrector is the following:

$$\mathrm{e}^{\tau f_1}\,\mathrm{e}^{\tau f_2} = \mathrm{e}^{\frac{1}{2}\tau f_1}\left(\mathrm{e}^{\frac{1}{2}\tau f_1}\,\mathrm{e}^{\tau f_2}\,\mathrm{e}^{\frac{1}{2}\tau f_1}\right)\mathrm{e}^{-\frac{1}{2}\tau f_1}, \tag{14}$$

showing that the first-order method of equation (4) is conjugate to a second-order time-symmetric method, when f is split into $n = 2$ pieces.

The error can be substantially reduced by the use of a corrector, with particularly marked improvement for nearly integrable systems (section 4.2) and for Hamiltonian systems of kinetic plus potential type (section 4.1). These improvements will typically reduce the error by 2 or 3 orders of magnitude.

The second extension uses the flows not just of the f_i, where $f = \sum f_i$, but also the flows of their commutators $[f_i, f_j]$, $[f_i, [f_j, f_k]]$, etc. The simplest such method, for $\dot{x} = f_1 + f_2$, is given by

$$\mathrm{e}^{C}\,\mathrm{e}^{\frac{1}{2}\tau f_1}\,\mathrm{e}^{\tau f_2 - \frac{1}{24}\tau^3[f_2,[f_2,f_1]]}\,\mathrm{e}^{\frac{1}{2}\tau f_1}\,\mathrm{e}^{-C} = \mathrm{e}^{\tau(f_1+f_2)} + \mathcal{O}(\tau^5), \tag{15}$$

where the corrector C is itself a composition approximating $\frac{1}{24}\tau^2[f_1, f_2]$ to within $\mathcal{O}(\tau^4)$. For simple mechanical systems with $H = \frac{1}{2}\|p\|^2 + V(q)$, this is the same as applying the leapfrog method (equation (1)) with the modified potential energy $V - \frac{\tau^2}{24}\|\nabla V\|^2$, effectively giving a fourth-order method with a single force evaluation [117].

In this paper, we have only specified the coefficients of four particularly useful fourth-order methods, namely equations (8), (11), (15) and (56) below. For other methods, and especially for higher-order methods, we refer the reader to the original sources [6, 9–14, 24, 25, 29, 61, 75, 80–82, 101, 114].

2.2. *Runge–Kutta-like methods*

Runge–Kutta (RK) methods are defined for systems with linear phase space $\mathbb{R}^n$ [21]. For the system

$$\dot{x} = f(x), \qquad x(0) = x_0, \qquad x \in \mathbb{R}^n, \tag{16}$$

the s-stage RK method with parameters a_{ij}, b_i $(i, j = 1, \ldots, s)$ is given by

$$X_i = x_k + \tau \sum_{j=1}^{s} a_{ij} f(X_j), \qquad x_{k+1} = x_k + \tau \sum_{j=1}^{s} b_j f(X_j). \tag{17}$$

RK methods are 'linear', that is, the map from vector field f to the map $x_k \mapsto x_{k+1}$ commutes with linear changes of variable $x \mapsto Ax$ [95]. (Alternatively, the method is independent of the basis of $\mathbb{R}^n$). This implies, for example, that all RK methods preserve all linear symmetries of the system. They are explicit if $a_{ij} = 0$ for $j \geqslant i$, although they cannot then be symplectic. The most useful Runge–Kutta methods in geometric integration are the Gaussian methods ([49], section II.7). They are implicit, A-stable (stable for all τ for all linear systems with bounded solutions), have the maximum possible order for an s-stage RK method (namely $2s$), preserve all quadratic first integrals of f, and are symplectic for canonical Hamiltonian systems. The 1-stage Gaussian method, usually known as the (implicit) midpoint rule, reduces to equation (2), and for linear systems, coincides with the well-known Crank–Nicolson method.

RK methods can be expanded in a Taylor series of the form

$$x_{k+1} = a_0 x_k + a_1 \tau f + a_2 \tau^2 f'(f) + \tau^3 (a_3 f''(f, f) + a_4 f'(f'(f))) + \cdots, \tag{18}$$

where each term on the right-hand side is evaluated at $x = x_k$ and the derivative $f^{(m)}$ is a multilinear mapping from m vector fields to single vector fields. The terms in this series are called elementary differentials of f and the series equation (18) is called a B-series. (For a consistent method, $a_0 = a_1 = 1$.) There are other integration methods that also have a B-series. We call any such method a B-series method. A very general form is to generalize the right-hand sides of equation (17) to be a sum over elementary differentials, each term of which is evaluated at each of the X_i, i.e.

$$\begin{aligned} X_i &= x_k + \tau \sum_{j=1}^{s} a_{ij} f(X_j) + \tau^2 \sum_{j,l=1}^{s} a_{ijl} f'(X_j)(f(X_l)) + \cdots, \\ x_{k+1} &= x_k + \tau \sum_{j=1}^{s} b_j f(X_j) + \tau^2 \sum_{j,l=1}^{s} b_{jl} f'(X_j)(f(X_l)) + \cdots. \end{aligned} \tag{19}$$

Three special cases are (i) exponential integrators, in which an analytic function of the Jacobian $f'(X_j)$ is incorporated [99]; (ii) elementary differential (EDRK) methods, in which all terms of each elementary differential are evaluated at a single X_i [101], and (iii) multiderivative (MDRK) methods, in which, in addition, only those combinations of elementary differentials that appear in the derivatives of x (i.e. $\dot{x} = f, \ddot{x} = f'(f), \dddot{x} = f'(f'(f)) + f''(f, f), \ldots$) appear [49]. EDRK methods can be symplectic but MDRK methods cannot (except when one

restricts to RK methods). Exponential integrators can be exact on all linear or affine systems and have been applied to stiff and highly oscillatory systems.

If we fix a basis in $\mathbb{R}^n$, writing $x = (x^1, \ldots, x^n)$, we can define the *partitioned* Runge–Kutta (PRK) methods, in which a different set of coefficients are used for each component of x:

$$X_i^l = x_k^l + \tau \sum_{j=1}^{s} a_{ij}^l f^l(X_j), \qquad x_{k+1}^l = x_k^l + \tau \sum_{j=1}^{s} b_j^l f^l(X_j). \tag{20}$$

They are not linear methods and do not have a B-series (although they can be expanded in a so-called P-series). Some PRK methods are splitting methods, but apart from these, the most useful PRK methods in geometric integration are the Lobatto IIIA-IIIB methods, of order $2s - 2$ ([49], p 550). These are symplectic for natural Hamiltonian systems and have two sets of coefficients, one for the q and one for the p variables. They also have a natural extension that is symplectic for holonomically constrained Hamiltonian systems [62].

2.3. Projection methods

Some properties can be preserved by simply taking a step of an arbitrary method and then enforcing the property. Integrals and weak integrals (invariant manifolds) can be preserved by projecting onto the desired manifold at the end of a step or steps, typically using orthogonal projection with respect to a suitable metric (see section 3.9). For example, energy-preserving integrators have been constructed using discrete gradient methods, a form of projection [94, 98]. Although it is still used, projection is something of a last resort, as it typically destroys other properties of the method (such as symplecticity) and may not give good long-time behaviour. Reversibility (section 3.6) is an exception, for R-reversibility can be imposed on the map φ by using $\varphi R \varphi^{-1} R^{-1}$, where $R : M \to M$ is an arbitrary diffeomorphism of phase space. Since this is a composition, it can preserve group properties of φ such as symplecticity [95]. Symmetries are a partial exception; the composition $\varphi S \varphi S$ is not S-equivariant, but it is closer to equivariant than φ is, when $S^2 = 1$ [60].

2.4. Variational methods

Many ODEs and PDEs of mathematical physics are derived from variational principles with natural discrete analogues. For an ODE with Lagrangian $L(q, \dot{q})$, one can construct an approximate discrete Lagrangian $L_d(q_0, q_1)$ such as $L_d(q_0, q_1) = L(q_0, (q_1 - q_0)/\tau)$ and form an integrator by requiring that the discrete action $\sum_{i=0}^{N} L_d(q_i, q_{i+1})$ be stationary with respect to all variations of the q_i, $i = 1, \ldots, N-1$, with fixed end points q_0 and q_N [15, 78]. For regular Lagrangians, the integrator can be seen to be symplectic in a very natural way, and in fact the standard symplectic integrators such as leapfrog and the symplectic Runge–Kutta methods can be derived in this way. The advantage of the discrete Lagrangian approach is that it acts as a guide in new situations. The Newmark method, well known in computational engineering, is variational [78], a fact that allowed the determination of the (nonstandard) symplectic form it preserves; singular Lagrangians can be treated; it suggests a natural treatment of holonomic (position) and nonholonomic (velocity) constraints and of nonsmooth problems involving collisions [103]; and powerful 'asynchronous' variational integrators can be constructed, which use different, even incommensurate time steps on different parts of the system [73]. In these situations variational integrators appear to be natural, and to work extremely well in practice, even if the reason for their good performance (e.g., by preserving some geometric feature) is not yet known [74].

2.5. Linear multistep methods

These are not at first sight good candidates for geometric integrators. Defined on a linear space $M = \mathbb{R}^n$ for $\dot{x} = f(x)$, a linear s-step method [21] is

$$\sum_{j=0}^{s} \alpha_j x_{k+j} = \tau \sum_{j=0}^{s} \beta_j f(x_{k+j}). \tag{21}$$

Such methods define a map on the product space M^s, and can sometimes preserve a structure (e.g. a symplectic form) in this space. However, this will not usually give good long-time behaviour of the sequence of points $x_k \in M$.

Instead, one considers the so-called underlying one-step method $\varphi : M \to M$, which is defined so that the sequence of points $x_k := \varphi^k(x_0)$ satisfies the multistep formula. (It always exists, at least as a formal power series in τ.) Often the dynamics of φ dominates the long-term behaviour of the multistep method. Recently it has been proved [28, 46] that the underlying one-step methods for a class of time-symmetric multistep methods for second-order problems $\ddot{x} = f(x)$ are conjugate to symplectic, which explains their near conservation of energy over long times and their practical application in solar system dynamics.

3. Preservation of intrinsic properties of differential equations

3.1. Types of differential equations

In this section we survey first the types of differential equations that are considered in this paper, then their intrinsic properties, and finally the preservation of these properties by geometric integrators. First, Hamiltonian systems. In the most common case, a Hamiltonian system has the following form [77]:

$$\begin{pmatrix} \dot{q} \\ \dot{p} \end{pmatrix} = \mathcal{J} \nabla H(q, p), \tag{22}$$

where q and $p \in \mathbb{R}^m$, where the symplectic structure $\mathcal{J}$ is given by

$$\mathcal{J} := \begin{pmatrix} 0 & Id \\ -Id & 0 \end{pmatrix}, \tag{23}$$

and where the Hamiltonian function H is the sum of kinetic and potential energy:

$$H(q, p) = \frac{1}{2}\|p\|_2^2 + V(q), \qquad \|p\|_2^2 = \sum_i p_i^2. \tag{24}$$

More general is the case

$$\dot{x} = \mathcal{J} \nabla H(x), \tag{25}$$

where $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, and $\mathcal{J}$ is now allowed to be any constant skew-symmetric matrix, and H is an arbitrary (differentiable) function.

A simple example is provided by a classical particle of mass m moving in a potential $V(q_1, q_2, q_3)$ and subject to a constant magnetic field $b{:=}(b_1, b_2, b_3)$. This can be modelled by equation (25), with $x{:=}(q_1, q_2, q_3, p_1, p_2, p_3)$, $H = \frac{1}{2m}\left(p_1^2 + p_2^2 + p_3^2\right) + V(q_1, q_2, q_3)$ and

$$\mathcal{J} = \begin{pmatrix} 0 & Id_3 \\ -Id_3 & \hat{b} \end{pmatrix}, \qquad \text{where} \quad \hat{b} := \begin{pmatrix} 0 & -b_3 & b_2 \\ b_3 & 0 & -b_1 \\ -b_2 & b_1 & 0 \end{pmatrix}. \tag{26}$$

In this example, as in equation (23), the matrix $\mathcal{J}$ has full rank. In general, however, the determinant of $\mathcal{J}$ in equation (25) may be zero. In this case, the system (25) is called a Poisson system.

More general again is the case where the Poisson structure $\mathcal{J}$ depends on x:

$$\dot{x} = \mathcal{J}(x)\nabla H(x). \tag{27}$$

Here, in addition to being skew, $\mathcal{J}$ must satisfy the so-called Jacobi identity:

$$\sum_{\ell} \left\{ \mathcal{J}_{i\ell} \frac{\partial}{\partial x_\ell} \mathcal{J}_{jk} + \mathcal{J}_{j\ell} \frac{\partial}{\partial x_\ell} \mathcal{J}_{ki} + \mathcal{J}_{k\ell} \frac{\partial}{\partial x_\ell} \mathcal{J}_{ij} \right\} = 0, \quad i, j, k = 1, \ldots, n. \tag{28}$$

Finally, the most general case we will consider is the nonautonomous version of (27):

$$\dot{x} = \mathcal{J}(x)\nabla H(x, t). \tag{29}$$

3.2. Symplectic structure

Considering these systems of increasing generality (22), (25), (27), (29), we note that in all cases they preserve the symplectic two-form[9]

$$\Omega := \sum_{i,j} \mathcal{J}_{ij}^{-1}(x)\, \mathrm{d}x_i \wedge \mathrm{d}x_j. \tag{30}$$

This is equivalent to saying that these equations can be derived from an underlying variational problem using Hamilton's principle. This symplectic/variational structure has many important physical and mathematical consequences, and it is therefore usually important to preserve if possible.

3.3. Conservation of phase-space volume

One consequence of the symplectic structure of equations (22) and (25) is that these systems preserve Euclidean volume, while equations (27) and (29) preserve a non-Euclidean volume, i.e., are measure preserving. For example, if the matrix $\mathcal{J}$ in (27)–(28) is nonsingular, the system preserves the measure[10]

$$\frac{\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n}{\sqrt{\det \mathcal{J}(x)}}. \tag{31}$$

3.4. Conservation of energy

The next property we consider is conservation of the Hamiltonian/energy. This property is possessed by those equations above that are autonomous (i.e. by (22), (25) and (27), but in general not by (29)), as is not hard to show

$$\begin{aligned} \frac{\mathrm{d}H}{\mathrm{d}t} &= \sum_i \frac{\partial H(x)}{\partial x_i} \frac{\mathrm{d}x_i}{\mathrm{d}t} \\ &= \sum_{i,j} \frac{\partial H(x)}{\partial x_i} \mathcal{J}_{ij}(x) \frac{\partial H(x)}{\partial x_j} \\ &= 0, \end{aligned} \tag{32}$$

where we have used the fact that $\mathcal{J}$ is skew, i.e. $\mathcal{J}_{ij} = -\mathcal{J}_{ji}$.

[9] If $\mathcal{J}$ is singular, the form Ω on a leaf is determined by $\Omega(\mathcal{J}\nabla F, \mathcal{J}\nabla G) = \{F, G\}$ [77].

[10] Note that for the non-singular skew $\mathcal{J}$, $\sqrt{\det \mathcal{J}(x)}$ can be expressed as a Pfaffian.

3.5. Other first integrals

In addition to the above properties, shared by all (respectively all autonomous) Hamiltonian systems, it is not unusual for Hamiltonian systems to possess additional symmetries, foliations, or first integrals. We start by discussing first integrals, also known as constants of the motion.

By definition, $I(x)$ is a first integral of a system with Hamiltonian H, if $\frac{\mathrm{d}I(x)}{\mathrm{d}t} = 0$ provided $x(t)$ is a solution of Hamilton's equations, i.e.

$$\begin{aligned}\frac{\mathrm{d}I}{\mathrm{d}t} &= \sum_i \frac{\partial I(x)}{\partial x_i}\frac{\mathrm{d}x_i}{\mathrm{d}t} \\ &= \sum_{i,j} \frac{\partial I(x)}{\partial x_i}\mathcal{J}_{ij}(x)\frac{\partial H(x)}{\partial x_j} \\ &= 0. \end{aligned} \tag{33}$$

Well-known physical examples of first integrals are momentum and angular momentum. More generally, the so-called momentum map associated with certain symmetries (via Noether's theorem) is a first integral. In addition there exist many other examples of first integrals, including Casimirs (i.e. I that satisfy $\sum_i \frac{\partial I(x)}{\partial x_i}\mathcal{J}_{ij}(x) = 0, \forall j$). Systems with a sufficient number of integrals to allow a solution of the system in closed form (in principle) are called (completely) integrable; some examples are the Kepler problem and the Toda chain. (For a discussion of foliations see [88].)

3.6. Symmetries

Symmetries occur in many physical systems. We can distinguish linear/nonlinear, continuous/discrete, and ordinary/reversing symmetries [42, 68, 104].

We start with a simple example of a linear, discrete, reversing symmetry. Note that the Hamiltonian (24) is invariant under $p \mapsto -p$. It is easy to see that the corresponding equations (22) are invariant under $\binom{p}{t} \mapsto \binom{-p}{-t}$. This is the most common form of time-reversal symmetry, and it has important physical and mathematical consequences. Common (continuous) symmetries are invariance under translation ($q \mapsto q + a$) and rotation ($q \mapsto Rq$). An example of a system with a nonlinear symmetry is generated by the Hamiltonian $H = \frac{1}{2}\left(p_1^2 + p_2^2\right)^2 + q_1^2 + 4q_2^2 + aq_1^{-2}$. This system has an additional integral $p_2p_1^2 + 8q_1q_2p_1 + 2\left(aq_1^{-2} - q_1^2\right)p_2$ that (via Noether's theorem) corresponds to a nonlinear symmetry [55].

A map $S : M \to M$ of phase space is a symmetry of f if it leaves f invariant (i.e. $TS \cdot f \circ S^{-1} = f$, where TS is the tangent map (Jacobian derivative) of S), while a map $R : M \to M$ of phase space is a reversing symmetry of f if it maps f to $-f$ (i.e. $TR \cdot f \circ R^{-1} = -f$). In general, a system may have both symmetries and reversing symmetries, and the group combining both of these is called the reversing symmetry group of the system [68].

3.7. Preserving symplectic structure

The symplectic structure for natural Hamiltonian systems described by equations (22)–(24) can be preserved by splitting the Hamiltonian function into kinetic and potential energy, i.e., applying equation (5) with $f_1 = \mathcal{J}\nabla\frac{1}{2}\|p\|_2^2$, $f_2 = \mathcal{J}\nabla V(q)$.

Each can then be integrated exactly, yielding

$$\begin{aligned} \mathrm{e}^{\tau f_1} &: q_{k+1/2} = q_k + \tau p_k, & p_{k+1/2} &= p_k, \\ \mathrm{e}^{\tau f_2} &: q_{k+1} = q_{k+1/2}, & p_{k+1} &= p_{k+1/2} - \tau\nabla V(q_{k+1/2}). \end{aligned} \tag{34}$$

Now we can use the composition (5) to obtain any desired order of accuracy (cf section 2) [125].

In general, we must split H into a sum of simpler Hamiltonians. Many simple Hamiltonians have been proposed and used [89]: quadratic Hamiltonians with linear dynamics; kinetic energies $p^{\mathrm{T}}M(q)^{-1}p$, of which many integrable cases are known, and if nonintegrable, can be diagonalized and further split into integrable 2D systems; two-body Hamiltonians such as central force problems and point vortices; monomials; the noninteracting (e.g. checkerboard) parts of lattice (e.g. quantum or semidiscrete PDE) systems with only near-neighbour interactions. A very general approach [37, 93] is to find vectors $a_1, \ldots, a_n$ such that $a_i^{\mathrm{T}}\mathcal{J}a_j = 0$ for all $i, j = 1, \ldots, n$. Then $H\left(a_1^{\mathrm{T}}x, \ldots, a_n^{\mathrm{T}}x\right)$ is integrable for any function H, and its flow is given simply by Euler's method $x_{k+1} = x_k + \tau\mathcal{J}\nabla H(x_k)$. Choosing suitable sets of vectors a_i then gives explicit splitting methods for any polynomial, an approach pioneered for use in accelerator physics [38].

However, for arbitrary Hamiltonians the only known symplectic methods are the symplectic Runge–Kutta methods and their generalizations, the symplectic partitioned Runge–Kutta and symplectic B-series and P-series methods. Symplectic Runge–Kutta methods are methods of the form (17) that satisfy the symplecticity condition [47]

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \qquad i, j = 1, \ldots, s, \tag{35}$$

of which the most useful are the Gaussian methods (see section 2.2). These will outperform all other symplectic integrators in the limit of very high accuracy.

Symplectic B- and P-series methods can be constructed using generating functions. Each type of generating function has associated with a Hamilton–Jacobi equation, whose exact solution generates the exact flow. Solving the Hamilton–Jacobi equation approximately (e.g., by Taylor series) yields a symplectic integrator. For example, the Poincaré generating function $x_{k+1} = x_k + \tau\mathcal{J}\nabla S((x_k + x_{k+1})/2)$ generates the midpoint rule on taking $S = H$, and generates a time-symmetric, fourth-order, linear, symplectic, EDRK, B-series method on taking $S = H - \frac{1}{24}\tau^2(\nabla H)^{\mathrm{T}}\mathcal{J}H''\mathcal{J}\nabla H$ [65].

Symplectic RK and B-series methods also preserve any constant symplectic structure for autonomous or nonautonomous Hamiltonians $H(x, t)$. However, there are no known methods that can preserve a general Poisson structure $\mathcal{J}(x)$. Splitting is the most practical method when it applies, the most famous example being the free rigid body, a Poisson, momentum-preserving integrator being just a sequence of planar rotations [79, 107, 63]. Methods that apply to arbitrary Hamiltonians for particular Poisson structures rely on finding either canonical coordinates or symplectic coordinates (a Poisson map from a symplectic vector space to the Poisson manifold). This can be done when $\mathcal{J}(x)$ is linear in x, the so-called Lie–Poisson case. Such $\mathcal{J}$ always arise from symplectic reduction of a canonical system on T^*G where G is a Lie group; a symplectic integrator for this system can reduce to a Poisson integrator [97].

3.8. Preserving phase-space volume

Because the preservation of phase-space volume plays an important role in many applications (e.g. in incompressible fluid flow), we here present volume-preserving methods for arbitrary divergence-free vector fields (this includes, but is more general than, Hamiltonian systems of the form (25)) [90].

We recall that, by definition, the ODE

$$\dot{x} = f(x) \tag{36}$$

is divergence-free if

$$\sum_{i=1}^{n} \frac{\partial f_i(x)}{\partial x_i} = 0. \tag{37}$$

Using this, we have

$$\begin{aligned} f_n(x) &= f_n(\bar{x}) + \int_{\bar{x}}^{x_n} \frac{\partial f_n(x)}{\partial x_n}\, \mathrm{d}x_n \\ &= f_n(\bar{x}) - \int_{\bar{x}}^{x_n} \left(\sum_{i=1}^{n-1} \frac{\partial f_i(x)}{\partial x_i} \right) \mathrm{d}x_n, \end{aligned} \tag{38}$$

where $\bar{x}$ is an arbitrary point which can be chosen conveniently (e.g. if possible such that $f_n(\bar{x}) = 0$).

Substituting (38) in (36), we get

$$\begin{aligned} \dot{x}_1 &= f_1(x), \\ &\vdots \\ \dot{x}_{n-1} &= f_{n-1}(x), \\ \dot{x}_n &= f_n(\bar{x}) - \sum_{i=1}^{n-1} \int_{\bar{x}}^{x_n} \frac{\partial f_i(x)}{\partial x_i}\, \mathrm{d}x_n. \end{aligned} \tag{39}$$

We now split this as the sum of $n-1$ vector fields

$$\begin{aligned} \dot{x}_i &= 0 \qquad i \neq j, n, \\ \dot{x}_j &= f_j(x), \\ \dot{x}_n &= f_n(\bar{x})\delta_{j,n-1} - \int_{\bar{x}}^{x_n} \frac{\partial f_j(x)}{\partial x_j}\, \mathrm{d}x_n, \end{aligned} \tag{40}$$

for $j = 1, \ldots, n-1$. Here δ is the Kronecker delta. Note that

(i) Each of these $n-1$ vector fields is divergence-free.
(ii) We have split one big problem into $n-1$ small problems. But each small problem has a simpler structure than (36); each corresponds to a two-dimensional Hamiltonian system

$$\dot{x}_j = \frac{\partial H_j}{\partial x_n}, \qquad \dot{x}_n = -\frac{\partial H_j}{\partial x_j}, \tag{41}$$

with Hamiltonian $H_j = -f_n(\bar{x})\delta_{j,n-1}x_j + \int_{\bar{x}}^{x_n} f_j(x)\, \mathrm{d}x_n$, treating x_i for $i \neq j, n$ as fixed parameters. Each of these 2D problems can either be solved exactly (if possible), or approximated with any symplectic integrator ψ_j. Even though ψ_j is not symplectic in the whole space $\mathbb{R}^n$, it is volume preserving. A volume-preserving integrator for f is then given by $\psi = \psi_1 \psi_2 \cdots \psi_{n-1}$.

3.9. Preserving energy and other first integrals

Linear and quadratic integrals are all automatically preserved by symplectic Runge–Kutta methods. However, no other integrals can be preserved by such methods, so if the integral one wishes to preserve is, for example, cubic, one has to make the choice of whether to preserve the symplectic structure or to preserve the integral. For the latter purpose there are two main methods: discrete-gradient methods [94, 98] and projection methods. Each of these

has its advantages and disadvantages. Here we will focus on projection methods, as they are somewhat simpler [47].

Define the integral manifold

$$M := \{x \mid g(x) = 0\}, \tag{42}$$

where

$$g(x) := I(x) - I(x_0) \tag{43}$$

and $I : \mathbb{R}^n \to \mathbb{R}^s$ defines s first integrals of the system. The standard projection method is now defined as follows.

(i) Compute $\bar{x}_{k+1} = \varphi_\tau(x_k)$ by any integration method φ_τ.
(ii) Project $\bar{x}_{k+1}$ onto M, to obtain $x_{k+1} \in M$.

The projection step is usually done by minimizing the Lagrange function $\mathcal{L}(x_{k+1}, \lambda) = \frac{1}{2}\|x_{k+1} - \bar{x}_{k+1}\|^2 - g(x_{k+1})^{\mathrm{T}}\lambda$, where $\lambda := (\lambda_1, \ldots, \lambda_s)$ represent Lagrange multipliers. Stationarity of the Lagrange function is then given by $\frac{\partial \mathcal{L}}{\partial x_{k+1}} = 0$, $\frac{\partial \mathcal{L}}{\partial \lambda} = 0$, i.e.

$$x_{k+1} = \bar{x}_{k+1} + g'(x_{k+1})^{\mathrm{T}}\lambda, \qquad g(x_{k+1}) = 0. \tag{44}$$

To speed up the numerical process it is convenient to replace (44) by

$$x_{k+1} = \bar{x}_{k+1} + g'(\bar{x}_{k+1})^{\mathrm{T}}\lambda, \tag{45}$$

$$g(x_{k+1}) = 0. \tag{46}$$

Since (45) is now explicit we can substitute it in (46) to obtain implicit equations for the Lagrange multipliers λ_i.

Reversible integral-preserving methods can be obtained by imposing reversibility by composition (equation (53) below), or by the symmetric projection method [47]. A recent novel method [22] projects to the desired integral surface yet retains automatic preservation of affine invariants.

When a system has a large number of integrals which it is desirable to preserve it is better to see if they arise from some structural feature of the equation. If solutions are confined to lie in the orbit of a group action, each orbit is a *homogeneous space* and it is possible to make big progress using *Lie group integrators* [61].

A famous example is the Toda lattice, which can be written in the form $\dot{L} = [A(L), L]$, where L is a symmetric, tridiagonal matrix and $A(L)$ is the antisymmetric, tridiagonal matrix with the same upper diagonal as L. Solutions are confined to the isospectral manifolds $\{UL(0)U^{-1} : U \in \mathrm{O}(n)\}$ and the eigenvalues of L are first integrals. Here the orthogonal matrices act on the symmetric matrices by conjugation.

Let $g \cdot x$ be the action of $g \in G$ on $x \in M$, where G is a Lie group and M is the phase space of the system. Write

$$\gamma(v, x) := \frac{\mathrm{d}}{\mathrm{d}t} \exp(tv) \cdot x|_{t=0} \tag{47}$$

so that all vector fields tangent to the orbits can be written as

$$\dot{x} = \gamma(a(x), x) \tag{48}$$

for some function $a : M \to \mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra of G. Classes of methods that can preserve the group orbits include Runge–Kutta–Munthe-Kaas (RKMK) integrators, Magnus methods (especially for nonautonomous linear systems), and splitting.

Instead of solving (48) directly, in RKMK methods its solution is represented as $x(t) = \exp(y(t)) \cdot x(0)$ where now $y(t) \in \mathfrak{g}$. Since $\mathfrak{g}$ is a linear space, any approximate

solution for y will yield an x lying in the correct orbit. Now $y(t)$ obeys the so-called 'dexpinv' equation

$$\dot{y} = \mathrm{dexp}_y^{-1} a(\exp(y) \cdot x(0)), \quad y(0) = 0, \tag{49}$$

which can be solved by any integrator. Here dexp is the derivative of the exponential map $\exp : \mathfrak{g} \to G$ and its inverse is given by $\mathrm{dexp}_y^{-1} a = a - \frac{1}{2}[y, a] + \frac{1}{6}[y, [y, a]] + \cdots$ where the coefficients are the Bernoulli numbers; typically this series has to be truncated to the desired order. If an explicit third-order Runge–Kutta method is chosen as the base method, then the Lie group integrator,

$$\begin{aligned} k_1 &= a(x_k), \\ k_2 &= a\big(\exp\big(\tfrac{1}{2}\tau k_1\big) \cdot x_k\big), \\ k_3 &= a(\exp(-\tau k_1 + 2\tau k_2) \cdot x_k), \\ l &= \tau\big(\tfrac{1}{6}k_1 + \tfrac{2}{3}k_2 + \tfrac{1}{6}k_3\big), \\ x_{k+1} &= \exp\big(l + \tfrac{1}{6}[l, k_1]\big) \cdot x_k, \end{aligned} \tag{50}$$

is third order and leaves x_{k+1} on the same group orbit as x_k. Many higher-order methods using as few as possible function evaluations, exponentials and commutators have been derived [24, 25, 105]. Attention also has to be given to evaluating the exponentials exactly, or at least approximating them within G; this can be computationally expensive.

Magnus methods can be used for linear problems $\dot{x} = \gamma(a(t), x)$, such as the time-dependent Schrödinger equation. Again they allow one to use evaluations, commutators and exponentials of a to approximate x_{k+1} in the correct group orbit [12, 54, 59].

Splitting methods are more ad hoc. One approach is to choose a basis for $\mathfrak{g}$, say $v_1, \ldots, v_n$, and write $a(x) = \sum a_i(x) v_i$. The vector field $\gamma(a_i(x), x)$ is tangent to the one-dimensional group orbit $\exp(t v_j)(x)$; hence if these are integrable, the ODE $\dot{x} = \gamma(a_i(x), x)$ is integrable by quadratures. Splitting is preferred when it is desired to preserve some other property as well as the integrals, such as phase-space volume. RKMK and Magnus methods, on the other hand, are 'linear', and hence preserve orbits of subgroups of G (as well as orbits of G) automatically.

3.10. Preserving symmetries and reversing symmetries [95]

Linear and affine symmetries are automatically preserved by all Runge–Kutta methods. Linear and affine reversing symmetries are preserved by all so-called time-symmetric Runge–Kutta methods, i.e. methods of the form (17) that satisfy the time-symmetry condition

$$a_{s+1-i,s+1-j} + a_{ij} = b_j, \qquad \forall i, j. \tag{51}$$

It follows that time-symmetric symplectic Runge–Kutta methods, apart from being symplectic and preserving any linear and quadratic first integrals, also automatically preserve any linear symmetries and reversing symmetries.

Starting from any numerical integration method $x_{k+1} = \varphi_\tau(x_k)$, a single nonlinear reversing symmetry

$$x \mapsto R(x), \qquad t \mapsto -t, \tag{52}$$

can be preserved by

$$x_{k+1} = R\varphi_{\frac{\tau}{2}}^{-1} R^{-1} \varphi_{\frac{\tau}{2}}(x_k), \tag{53}$$

provided φ_τ commutes with R^2.

Preserving nonlinear symmetries is in general more complicated, and the reader is referred to [35, 56].

3.11. Dissipative systems

We will use the term 'dissipative' to refer to systems in which the time derivative of a differential k-form is nonpositive. When $k = 0$ the system has a Lyapunov function $V(x)$ which obeys $\dot{V} := \frac{\mathrm{d}}{\mathrm{d}t} V(x(t)) \leqslant 0$. When $k = 2$ the system has a symplectic form ω which obeys $\dot{\omega} := \frac{\mathrm{d}}{\mathrm{d}t}((\mathrm{e}^{tf})^*)\omega) \leqslant 0$. When $k = n$ the system has a volume form μ which obeys $\dot{\mu} \leqslant 0$, i.e. the system is volume contracting. (Here the forms are evaluated on k arbitrary vectors.)

An important special case is when the k-form contracts at a constant rate, i.e. $\dot{V} = -cV$ (in which case the foliation defined by $\{x : V(x) = \text{const.}\}$ is preserved), $\dot{\omega} = -c\omega$ (in which case the system is said to be conformal Hamiltonian [84]; in fact symplectic forms can only contract in this way), or $\dot{\mu} = -c\mu$ (in which case the system is said to be conformal volume preserving). In these cases the maps contracting the given structure actually form a group; for general contraction they only form a semigroup [91].

An example is provided by simple mechanical systems with Rayleigh dissipation. If the Hamiltonian is $H = \frac{1}{2} p^{\mathrm{T}} M(q)^{-1} p + V(q)$, the equations of motion are

$$\dot{q} = \nabla_p H, \qquad \dot{p} = -\nabla_q H - R(q)M(q)^{-1} p, \tag{54}$$

where $R(q)$ is positive definite. Now the energy H is a Lyapunov function, for $\dot{H} = -\dot{q} R(q) \dot{q} \leqslant 0$, and the Euclidean volume $\mu = \prod_i \mathrm{d}q_i \wedge \mathrm{d}p_i$ contracts, for $\dot{\mu} = -\mathrm{tr}(R(q)M(q)^{-1})\mu \leqslant 0$. The symplectic form $\sum \mathrm{d}q_i \wedge \mathrm{d}p_i$ contracts conformally if and only if $R(q) = cM(q)$ for some constant c, as in the important and simplest case $R(q) = cI$, $M(q) = I$.

Methods are available that preserve these properties. Doing so is particularly important when one wants to know the effect of very slow dissipation over long times. However, it is in precisely these cases that we should add the requirement that the differential form is in fact preserved when $\dot{V}$ (respectively $\dot{\omega}, \dot{\mu}$) $= 0$. This will ensure both that the dissipation proceeds at the correct rate and that it vanishes if the solution tends towards a dissipation-free submanifold as $t \to \infty$.

To ensure that a Lyapunov function (or functions) decreases, one can use a form of projection combined with Runge–Kutta [44]. First, evaluate the stages of the RK method (17) applied to the original system $\dot{x} = f(x)$. Then, apply the *same* RK method to $\dot{V} = \nabla V(x)^{\mathrm{T}} f(x)$. If the weights b_i of the RK method are all positive, then V will not increase. Finally, project the original RK solution for x_{k+1} to the manifold $\{x : V(x) = V_{k+1}\}$.

Conformal Hamiltonian systems can be integrated [84] by splitting off the dissipative part, which (without loss of generality) always has the form $\dot{q} = 0$, $\dot{p} = -cp$ and can be integrated explicitly. The non-dissipative part can be treated with any symplectic integrator.

Volume-contracting systems can also be integrated by splitting them into contracting and divergence-free parts. For example, the vector field g defined by

$$\dot{x}_1 = f_1 + \sum_{i=2}^{n} \int_0^{x_1} \frac{\partial f_i}{\partial x_i} \, \mathrm{d}x_1, \qquad \dot{x}_i = 0, \quad i = 2, \ldots, n, \tag{55}$$

has divergence equal to $\nabla \cdot f \leqslant 0$. But for a scalar system, contractivity simply requires that solutions approach each other. Runge–Kutta methods that obey this are called B-stable [115], which is equivalent to $b_i \geqslant 0$ and $b_i a_{ij} + b_j a_{ji} - b_i b_j \geqslant 0$ for all i, j. In particular, Gaussian Runge–Kutta methods are B-stable. The remainder $f - g$ is divergence-free and can be treated with any volume-preserving integrator. Volume contractivity is only a semigroup property, however, and so the order cannot be increased beyond 2 by composition (because this would require negative time steps [7] that increase the volume). Higher-order volume-contracting integrators require special techniques [45].

This leaves the cases when the system has several of these properties. These are unsolved in general, although there is some evidence that applying an integrator to (54) that would be symplectic in the absence of dissipation does provide good long-time behaviour and good tracking of the energy decay. Such methods can be constructed by splitting or by including dissipative forces in variational integrators by the Lagrange–d'Alembert principle [74].

4. Special cases

The ODEs considered in section 3 fall into natural classes defined by intrinsic, geometric properties. Their flows and integrators inherit these properties, which directly influence the nature of the dynamics of the system. In this section, by contrast, we consider properties that may not be intrinsic or influence the dynamics, but can nevertheless be usefully taken into account in the integrator design. Splitting itself is an example: while writing $f = \sum f_i$ with f_i integrable is (in a sense) intrinsic, it does not in itself influence the dynamics. We first look at special cases of splitting.

4.1. Simple mechanical systems

These are canonical Hamiltonian systems where the Hamiltonian has the form kinetic plus potential energy, i.e. $H = \frac{1}{2}\|p\|^2 + V(q)$, the norm $\|p\|$ being induced from a metric on the configuration space; in coordinates, $\|p\|^2 = p^{\mathrm{T}} M(q)^{-1} p$. Let $f_2 = \mathcal{J}\nabla V$ and $f_1 = \mathcal{J}\nabla\frac{1}{2}\|p\|^2$ be the Hamiltonian vector fields of the potential and kinetic energies, so that f_2 is always integrable and f_1 is integrable when the geodesics of the metric are integrable. In the simplest case, $M(q) = I$ and the symmetrized composition (7) gives the standard leapfrog method, equation (1). The Lie algebra generated by f_1 and f_2 is far from being generic, however, and its special properties can be exploited in the design of the integrator.

First, note that $[f_2, [f_2, f_1]] = \mathcal{J}\nabla\|\nabla V(q)\|^2$ (the Hamiltonian vector field of the function $\|\nabla V(q)\|^2$) and hence $[f_2, [f_2, f_1]]$ commutes with any other function of the coordinates only; for example, $[f_2, [f_2, [f_2, f_1]]] \equiv 0$. This means that $[f_2, [f_2, [f_2, f_1]]]$ and numerous terms in the error expansion of a composition method are identically zero and that their order conditions (their coefficients expressed as polynomials in the parameters of the method, which must normally be set to zero) can be dropped. This allows methods, usually known as 'Runge–Kutta–Nyström' methods, with smaller errors and often with fewer force evaluations as well. A full analysis of the Lie algebra generated by f_1 and f_2 for simple mechanical systems is provided in [96].

A good fourth-order method of this class, due to Blanes and Moan [14], who also give good sixth-order methods, is given by equation (5) together with $m = 7$ and

$$\begin{aligned}
a_0 &= a_7 = 0,\\
a_1 &= a_6 = 0.245\,298\,957\,184\,271,\\
a_2 &= a_5 = 0.604\,872\,665\,711\,080,\\
a_3 &= a_4 = \tfrac{1}{2} - (a_2 + a_3),\\
b_1 &= b_7 = 0.082\,984\,406\,417\,4052,\\
b_2 &= b_6 = 0.396\,309\,801\,498\,368,\\
b_3 &= b_5 = -0.039\,056\,304\,922\,3486,\\
b_4 &= 1 - 2(b_1 + b_2 + b_3).
\end{aligned} \tag{56}$$

It supersedes previous methods of this class [83]. This is the method 'LF4' used in figures 1–4. Note that, after the first step, the initial force is available from the previous time step, so only six force evaluations per time step are needed.

A further extension uses derivatives of the force. Because $f_3 := [f_2, [f_2, f_1]] = \mathcal{J}\nabla(\|\nabla V(q)\|^2)$ is itself integrable, its flow can be included in the integrator. The flow of $f_2 + \alpha f_3$ is given by

$$q(t) = q(0), \qquad p(t) = p(0) - t(f_2 + \alpha(N'(f_2, f_2) + 2f_2'(Nf_2)))(q(0))), \tag{57}$$

which only involves evaluating one derivative of the force evaluated in one direction Nf_2. (Here $N = M^{-1}$.) This can be very cheap for some problems, e.g. n-body systems with two-body interactions, for which it costs about the same as one force evaluation, or if the latter is dominated by expensive square root calculations, much less. For example, let $W : \mathbb{R}^3 \to \mathbb{R}$ be a potential such that $V = \frac{1}{2}\sum_{\substack{i,j \\ j\neq i}} W(q_i - q_j)$, $q_i \in \mathbb{R}^3$. Then

$$\sum_j \frac{\partial f_i}{\partial q_j} v_j = \sum_j W''(q_i - q_j)(v_i - v_j), \tag{58}$$

where W'' is the Hessian of W.

The simplest method of this type uses a corrector to achieve order 4 with only a single force evaluation, equation (15) [117]. It can be extended [10, 11, 14, 29] by including more stages (this decreases the local error); going to higher order (up to 8th order methods have been found); considering near-integrable systems, described below; and including higher derivatives of the force.

This class of methods can also be used when $f_1 = \mathcal{J}\nabla\left(\frac{1}{2}\|p\|^2 + V_1(q)\right)$, $f_2 = \mathcal{J}\nabla V_2(q)$ (i.e., $V = V_1 + V_2$) if f_1 is integrable, this being of particular advantage when f_2 is simpler (e.g. smaller or slower) than f_1. The most famous example is the solar system, for which f_1 contains the sun–planet and f_2 the planet–planet interactions. The class can also be used on non-Hamiltonian systems of the form

$$\dot{x} = f(x) + L(x)y, \qquad \dot{y} = g(x), \tag{59}$$

where $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$ and $L(x) \in \mathbb{R}^{m\times n}$. For example, high-order systems $\frac{\mathrm{d}^k z}{\mathrm{d}t^k} := z^{(k)} = f(z, \dot{z}, \ldots, z^{(k-2)})$ have this form when written as the (divergence-free) first-order system

$$\dot{x}_i = x_{i+1}, \qquad i = 0, \ldots, k-2, \qquad \dot{x}_{k-1} = f(x_0, \ldots, x_{k-2}), \tag{60}$$

where $x_i = z^{(i)}$.

4.2. Near-integrable systems

If we are splitting $f = f_1 + f_2$ (f_1 integrable) and $\|f_2\| \ll \|f_1\|$, then f is near-integrable and this can be exploited in the construction of integrators. Specifically, introduce a small parameter ε and (by rescaling f_2) let $f = f_1 + \varepsilon f_2$. The error of a composition method is automatically $\mathcal{O}(\varepsilon)$ and vanishes with ε. (See figures 2, 3.) Splitting is superb for such nearly-integrable systems.

But, one can do even better. The error can be expanded as a double Taylor series in τ and ε, and if $\varepsilon \ll \tau$ then one can preferentially eliminate error terms with small powers of ε. The number of these terms is only polynomial in the order n, instead of exponential. Thus, we can beat the large cost of high-order methods in this case. For example, there is only one error term of each order $\mathcal{O}(\varepsilon\tau^n)$ and $\lfloor\frac{1}{2}(n-1)\rfloor$ of order $\mathcal{O}(\varepsilon^2\tau^n)$. One can easily build, for example, a 2-stage method of order $\mathcal{O}(\varepsilon^2\tau^4 + \varepsilon^3\tau^3)$, a 3-stage method of order $\mathcal{O}(\varepsilon^2\tau^6 + \varepsilon^3\tau^4)$, and so on [10, 69, 81].

This idea combines particularly well with the use of correctors. For *any* composition, even standard leapfrog, for all n there is a corrector that eliminates the $\mathcal{O}(\varepsilon\tau^p)$ error terms for all $1 < p < n$. Thus, any splitting method is 'really' $\mathcal{O}(\varepsilon^2)$ accurate on near-integrable problems. This construction was used by Wisdom [124] to improve the accuracy of a long leapfrog integration of the solar system years after it had actually been performed! A combination of correctors and preferential elimination of error terms was also used by Laskar [70] in a solar system simulation which contributed to the 2004 geophysical time scale.

4.3. PDEs and linear parts

Consider the wave equation $\ddot{q} = q_{xx} + f(q)$ with periodic boundary conditions in space. A pseudo-spectral spatial discretization leads to a set of ODEs of the form $\ddot{q} = Lq + f(q)$. Although the linear part $\dot{q} = p$, $\dot{p} = Lq$ could be split as in the standard leapfrog, it is also possible to split the system into linear and nonlinear parts, solving the linear part exactly. This is the approach traditionally used for the nonlinear Schrödinger equation, for example [122]. This has the advantage of removing any stability restriction due to the splitting of the oscillators, and if the nonlinearity is small, we have the advantages discussed above for near-integrable systems. Furthermore, the highly accurate Runge–Kutta–Nyström corrector, and multiderivative methods can also be used, and the time step can be increased using special methods for highly oscillatory systems (section 4.7).

4.4. Nonautonomous systems

The usual way to treat these is to split the corresponding autonomous systems in an extended phase space. In the Hamiltonian case with $H = \frac{1}{2}\|p\|^2 + V(q,t)$, for example, the extended phase space is (q, ξ, p, p_ξ) and the extended Hamiltonian is $(\frac{1}{2}\|p\|^2 + p_\xi) + V(q,\xi)$. Applying leapfrog to the indicated splitting gives

$$\begin{aligned} q_{k+1/2} &= q_k + \tfrac{1}{2}\tau p_k, \\ p_{k+1} &= p_k - \tau \nabla V\left(q_{k+1/2}, t_k + \tfrac{1}{2}\tau\right), \\ q_{k+1} &= q_k + \tfrac{1}{2}\tau p_{k+1}. \end{aligned} \tag{61}$$

However, Blanes and Moan [13] have proposed an interesting alternative based on the Magnus expansion. To integrate the ODE $\dot{x} = f(x,t)$ on $[t_0, t_0 + \tau]$, first calculate the autonomous vector fields

$$f_0 := \int_{t_0}^{t_0+\tau} f(x,t)\,\mathrm{d}t, \qquad f_1 := \frac{1}{\tau}\int_{t_0}^{t_0+\tau}\left(t - \frac{1}{2}\tau\right) t f(x,t)\,\mathrm{d}t. \tag{62}$$

Then a second-order approximation of the flow of f is given by e^{f_0}, and a fourth-order approximation is given by

$$\mathrm{e}^{\frac{1}{2}f_0 - 2f_1}\,\mathrm{e}^{\frac{1}{2}f_0 + 2f_1}. \tag{63}$$

Each of these vector fields is then split and an integrator constructed by composition. This can be cost effective because more information about the t-dependence of f is used [8].

4.5. Systems with constraints

An ODE with a constraint forms a *differential-algebraic equation* (DAE), e.g. $\dot{x} = f(x,\lambda)$, $g(x) = 0$. A great deal is known about how to integrate DAEs in general ([51] chapters VI, VII, [4]). Here we are interested in the case when the DAE has extra structure which must also be preserved.

Consider a Hamiltonian system with the configuration manifold Q, phase space T^*Q and Hamiltonian $H : T^*Q \to \mathbb{R}$. If the motion is subject to a holonomic (position) constraint that $q \in M \subset Q$, then the constrained system is again Hamiltonian, this time with the phase space T^*M. The structure to be preserved is that the integrator should be a symplectic map on T^*M, expressed; however, in coordinates on the original phase space T^*Q, because often we have no good global description of M. Examples are bond length and bond angle constraints in molecular dynamics.

Suppose that we do have a second-order, time-symmetric, symplectic integrator $\varphi : T^*Q \to T^*Q$ for the unconstrained motion, and that the constraints are specified by $M = \{q : g(q) = 0\}$ where $g : Q \to \mathbb{R}^k$. Then the constrained equations are

$$\dot{q} = \nabla_p H, \qquad \dot{p} = -\nabla_q H - G(q)^{\mathrm{T}}\lambda, \qquad 0 = g(q), \tag{64}$$

where $G_{ij} = \partial g_i/\partial q^j$ and $\lambda \in \mathbb{R}^k$. If the initial data (q_k, p_k) satisfy the constraints, i.e. $g(q_k) = 0$ and $G(q_k)p_k = 0$, then a second-order, time-symmetric, symplectic integrator ψ is given by first calculating the new position by solving for the Lagrange multipliers λ in

$$(q_{k+1}, \bar{p}_{k+1}) = \varphi(q_k, p_k - G(q_k)^{\mathrm{T}}\lambda), \qquad g(q_{k+1}) = 0, \tag{65}$$

and then projecting the momentum to the constraint manifold by solving for the Lagrange multipliers μ in

$$p_{k+1} = \bar{p}_{k+1} - G(q_{k+1})^{\mathrm{T}}\mu, \qquad G(q_{k+1})p_{k+1} = 0, \tag{66}$$

although the second step (66) can be omitted without affecting the essential behaviour of the method.

When $H = \frac{1}{2}\|p\|^2 + V(q)$ and φ is given by the standard leapfrog method (1), this reduces to an integrator known as RATTLE. Although nonlinear equations have to be solved for the Lagrange multipliers λ, typically by iteration, the force ∇V only needs to be evaluated once.

Since the method (65), (66) is time symmetric, its order can be increased as desired by composition.

Another approach is to extend partitioned Runge–Kutta methods to constrained systems. The Lobatto IIIA–IIIB method [47, 62] has order $2s-2$ with s stages and is time symmetric and symplectic for systems with holonomic constraints. It is fully implicit, so a simple method like RATTLE would be preferred where possible, but if high order is needed, or if the Hamiltonian is not separable, then partitioned Runge–Kutta methods can be competitive. (Many complex mechanical systems have a nontrivial mass matrix $M(q)$ and hence are not separable.)

The dynamics of nonholonomically constrained Hamiltonian systems is far less well understood. It is not, in general, symplectic. The Lagrange–d'Alembert principle which generates the nonholonomic equations of motion can be discretized to yield useful integrators which preserve the key geometric features of the equations, namely the constraints and reversibility [30, 86]. For example, if the Lagrangian $\frac{1}{2}\|\dot{q}\|_2^2 - V(q)$ is subject to the nonholonomic constraint $A(q)\dot{q} = 0$, then one choice of discrete Lagrangian yields the nonholonomic analogue of RATTLE, namely

$$\begin{aligned} q_{k+1/2} &= q_k + \tfrac{1}{2}\tau v_k, \\ v_{k+1} &= v_k + \tau(-\nabla V(q_{k+1/2}) + A(q_{k+1/2})^{\mathrm{T}}\lambda_{k+1}), \\ q_{k+1} &= q_{k+1/2} + \tfrac{1}{2}\tau v_{k+1}, \\ A(q_{k+1})v_{k+1} &= 0, \end{aligned} \tag{67}$$

where $v = \dot{q}$ is the velocity. Like RATTLE, it is second order, time symmetric, reversible, and has only one force evaluation. The reversibility appears to ensure favourable long-time energy behaviour [85].

4.6. Variable time steps

In general purpose ODE software it is standard to adapt the time step during an integration so as to control the local error [49]. This becomes progressively more efficient than using a fixed time step as the diversity of regimes the orbit passes through becomes more extreme. In geometric integration, however, simply choosing the time step as some function of x (related to the force or its gradient, or to an estimate of the local error) will in general destroy any geometric properties of the integrator and hence destroy its good long-time behaviour.

Instead the idea is to adapt the system so that it *can* be integrated well using a constant time step. The adaptation must preserve the geometric properties of the system. Typically, however, the adaptation does make the system and any suitable geometric integrators more complicated.

The *Poincaré transformation* of Hamiltonian H on the energy level E is $\widetilde{H} := g(x)(H - E)$. Orbits $\tilde{x}(\tilde{t})$ of $\widetilde{H}$ with energy 0 coincide with orbits $x(t)$ of H and with energy E, but are followed at a different speed: $\tilde{x}(\tilde{t}) = x(t)$ where $\mathrm{d}t/\mathrm{d}\tilde{t} = g(x)$. The monitor function $g(x)$ is chosen to be small in regions of phase space that are difficult to integrate and the system with Hamiltonian $\widetilde{H}$ is integrated with a constant time step $\tilde{\tau}$, effectively reducing the time step in the difficult regions. However, even if H is separable then $\widetilde{H}$ is not, so an implicit symplectic integrator must be used for $\widetilde{H}$. To reduce the cost, one can choose $g(x)$ to be a smooth function that equals to 1 away from the difficult regions, and choose an integrator such as two-stage Lobatto IIIA–IIIB Runge–Kutta that reduces to the leapfrog method on separable Hamiltonians. Then the calculation proceeds using standard leapfrog until the orbit enters the difficult region, when it automatically switches to the (more expensive) implicit method.

A related method has been proposed for the N-body gravitational problem [26]. The system has Hamiltonian $H = H_1 + H_2$ where H_2 is singular at collisions. The leapfrog method is applied to the splitting $H = (H_1 + (1 - K)H_2) + K H_2$ where the monitor function K is equal to 0 near collisions, equal to 1 far away from collisions, and varies smoothly in between. Far from collisions, this reduces to standard, explicit leapfrog for $H_1 + H_2$. When $K \neq 1$, the first term $H_1 + (1 - K)H_2$ can no longer be integrated explicitly; when this happens, its solution is evaluated numerically up to roundoff error using a standard integrator. (One could equally well use a symplectic integrator with a smaller time step.) Provided the near collisions are rare enough, the extra cost is negligible.

There is more scope for transforming reversible systems by $\frac{\mathrm{d}t}{\mathrm{d}\tilde{t}} = g(x)$. When the original system is Hamiltonian and separable, there are reversible integrators for the transformed system [72, 50] that are either explicit or semiexplicit (require solving just a scalar nonlinear equation). Now there is essentially no cost to including adaptivity, but symplecticity has been sacrificed, raising questions about the long-time reliability of the integration. (For example, energy errors may be no longer bounded.) This method is in widespread use in molecular dynamics [72].

Asynchronous variational integrators [73] have been proposed in which each element (for example, each pair of particles in an N-body system) has its own time step, which can be either prescribed or determined variationally. The method looks very promising but has not yet been fully developed for ODE applications.

4.7. Highly oscillatory problems

The leapfrog method for the oscillator $\ddot{q} = \omega^2 q$ is stable only for $\tau < 2/\omega$. Step size restrictions like this can be a major performance bottleneck for large systems and many

methods have been proposed to overcome them. Three broad classes of systems have been tackled:

(i) the phase-space coordinates can be partitioned into fast and slow variables (in molecular dynamics, the fast variables might be the positions and momenta within each molecule, and the slow variables the positions and momenta between them);
(ii) the forces can be separated into fast and slow forces;
(iii) a special case of (ii) in which the motion due to the fast forces can be solved exactly.

In each case the new methods allow the time step to be increased significantly, by a factor of 5 or more, while still obtaining reliable results.

The 'reversible averaging' method of Leimkuhler and Reich [71] applies to problems of type (i). The fast subsystem is still integrated with a small time step, the slow variables being interpolated as needed, while the slow subsystem is integrated with a large time step, but with the force due to the fast variables being averaged along their orbit. This relies on the fast system being much faster to solve, as in molecular dynamics where it decouples into a sum of small systems. It is reversible and not symplectic, but appears to have excellent stability properties for large time steps on linear and nonlinear problems.

The 'mollified impulse method' of García-Archilla, Sanz-Serna and Skeel [40] applies to problems of type (ii). Suppose the system is described by $H = H_f(p,q) + V_s(q)$, where $\|H_f''\| \gg \|V_s''\|$. The method is the composition

$$\mathrm{e}^{N\tau/2\mathcal{J}\nabla\bar{V}_s}(\varphi_\tau(H_f))^N\,\mathrm{e}^{N\tau/2\mathcal{J}\nabla\bar{V}_s}, \tag{68}$$

where $\varphi_\tau(H_f)$ is a time-symmetric, symplectic integrator for the fast system H_f, such as leapfrog. The expensive, slow forces ∇V_s are only evaluated for every N time steps. The mollified potential $\bar{V}_s$ is given by $\bar{V}_s(q) = V_s(A(q))$, where the mollifier $A(q)$ has the effect of averaging over the fast motion. A typical choice is $A(q) = (2N\tau)^{-1}\int_{-N\tau}^{N\tau} q(t)\,\mathrm{d}t$ where $q(t)$ satisfies the reduced fast problem $\dot{q} = \nabla_p H_f$, $\dot{p} = -\nabla_q H_f$, $q(0) = q$, $p(0) = 0$, which itself has to be integrated for $2N\tau$ time steps using the method φ_τ. This time-symmetric, symplectic method yields dramatic improvements in accuracy and stability over standard leapfrog, although some instabilities remain when the step size is in resonance with the fast frequencies.

These can best be analysed by passing to the simplest case (iii), for which a key test problem is the system [47]

$$\ddot{q} + \Omega^2 q = -\nabla V_s(q), \tag{69}$$

where Ω is symmetric and positive definite, so that the fast frequencies are the eigenvalues of Ω. In this case the mollified impulse method reduces to

$$q_{k+1} - 2\cos(\tau\Omega)q_k + q_{k-1} = -\tau^2\psi(\tau\Omega)\nabla V_s(\phi(\tau\Omega)q_k) \tag{70}$$

with $\psi(\xi) = \sin(\xi)/\xi$ and $\phi(\xi) = 1$, a method first introduced by Deuflhard [33]. In fact, the method (70) is time symmetric, second order and exact when $\nabla V_s(q) =$ constant for any even functions ψ, ϕ with $\psi(0) = \phi(0) = 1$. It is symplectic if $\psi(\xi) = (\sin(\xi)/\xi)\phi(\xi)$. It suffers from resonance instabilities when $\tau\lambda$ is a multiple of π (λ an eigenvalue of Ω), but these can be greatly reduced by a careful choice of ψ and ϕ. Dramatic improvements in accuracy and increases in the step size can be obtained compared with leapfrog. Unfortunately it appears that the requirements of symplecticity, lack of resonance instabilities, good long-time energy conservation, good long-time conservation of adiabatic invariants and accurate slow-time-scale dynamics cannot all be achieved simultaneously [47].

Exponential integrators can be applied to highly oscillatory problems [43]. These are methods for $\dot{x} = f(x)$ that incorporate entire functions of the Jacobian $f'(x)$ or an

approximation of the Jacobian. Most practically, one writes $\dot{x} = Lx + N(x)$ and approximates $f'(x)$ by the constant matrix L. The main strand of research seeks methods that are explicit, exact when $N(x) = 0$, and show good stability and high order on stiff problems (those with $\|L\| \gg 1$), especially including discretizations of parabolic and hyperbolic PDEs. Enormous gains can be made compared to the common approach of simply treating the linear terms with Crank–Nicolson. See [99] for a survey of exponential integrators.

Lie group integrators based on the Magnus expansion can also be adapted to solve highly oscillatory nonautonomous linear ODEs [58].

5. Dynamical behaviour of geometric integrators

The main goal of geometric integration is to preserve phase portraits as well as possible. Therefore we first consider how well various invariant sets in phase space are preserved, either exactly or with a perturbation, before looking at the error growth in quantities that are not preserved.

5.1. Preservation of invariant sets

A useful reference for this subsection is Stuart and Humphries [115].

5.1.1. Preservation of equilibrium points. All B-series methods and all linear multistep methods retain all equilibrium points of $\dot{x} = f(x)$ exactly. Some B-series methods, however, may generate additional, so-called spurious, fixed points [115]. The one-stage theta method

$$x_{k+1} = x_k + \tau f(\theta x_{k+1} + (1-\theta)x_k) \tag{71}$$

has no spurious fixed points for any τ (for $\theta = \frac{1}{2}$ it is the symplectic midpoint rule), and neither does the (symplectic) fourth-order Gaussian Runge–Kutta method. It has been conjectured that there are no higher-order Runge–Kutta methods without spurious fixed points.

Composition can, in general, introduce spurious fixed points. However, if we split $f = \sum f_i$ such that $\sum f_i(y_i) = 0 \Longleftrightarrow f_i(y_i) = 0$ when $y_i - y_j = \mathcal{O}(\tau)$, then the first-order composition (5) has no spurious fixed points, and nor does the leapfrog method (1). The existence of higher-order composition methods with this property is unknown.

5.1.2. Preservation of quasiperiodic orbits. One of the most important sources of nonlinear stability is given by the existence of invariant tori in phase space, guaranteed by the Kolmogorov–Arnold–Moser (KAM) theorems. Such theorems have been proved for Hamiltonian systems, reversible systems and volume-preserving systems. This is one of the most important motivations for using integrators that preserve these geometric properties (usually a necessary condition for preservation of KAM tori).

Proofs that geometric integrators actually do preserve KAM tori seem to mainly have been given for Hamiltonian systems [47, 100, 110] and reversible systems [47]. The step size must not be resonant with the frequencies of the torus. Of course, the torus does undergo an $\mathcal{O}(\tau^p)$ shift in its position.

Preservation of periodic orbits for Hamiltonian, reversible and volume-preserving systems seems to have received much less attention.

5.1.3. Preservation of integrals. This can of course be enforced if desired (see section 3.5). But how do integrals fare if their preservation is not explicitly enforced?

Energy cannot, in general, be preserved by a symplectic integrator [41]. However, error analysis shows that energy errors can be bounded for long times (section 5.2). Some other integrals can be preserved exactly: splitting methods preserve any first integral of f that is also a first integral of each f_i, where $f = \sum f_i$. For simple mechanical systems with $H = \frac{1}{2}\|p\|^2 + V(q)$, this includes any integral associated with a spatial (q) symmetry. Thus the leapfrog method for separable systems preserves both linear and angular momentum. Quadratic integrals are preserved by any symplectic Runge–Kutta method, as mentioned before. Variational integrators can preserve integrals provided a discrete Lagrangian that preserves the associated symmetry can be found.

5.1.4. Preservation of spectral properties and bifurcations. Of crucial importance for the dynamics near an equilibrium point are its linear eigenvalues. These eigenvalues are generally grouped as follows:

$$\begin{aligned} &\lambda_i \in \text{a stable set}, && \text{if} \quad \operatorname{Re}\lambda_i < 0, \\ &\lambda_i \in \text{a centre set}, && \text{if} \quad \operatorname{Re}\lambda_i = 0, \\ &\lambda_i \in \text{an unstable set}, && \text{if} \quad \operatorname{Re}\lambda_i > 0. \end{aligned} \tag{72}$$

A necessary condition for an equilibrium point to be stable is that its unstable set be empty. To each of the above three sets is associated a linear subspace, and these subspaces are tangent to their nonlinear generalizations, the so-called stable, centre and unstable manifolds.

A second reason that eigenvalues are important is that bifurcations typically occur when the real part of one of the eigenvalues equals zero.

For both of these reasons, it is important to preserve the spectral properties of equilibrium points. This means that, corresponding to (72), we would like an integrator to have a fixed point with eigenvalues

$$\begin{aligned} &\mu_i \in \text{a stable set}, && \text{if} \quad |\mu_i| < 1, \\ &\mu_i \in \text{a centre set}, && \text{if} \quad |\mu_i| = 1, \\ &\mu_i \in \text{an unstable set}, && \text{if} \quad |\mu_i| > 1. \end{aligned} \tag{73}$$

Time-symmetric Runge–Kutta methods that are also A-stable (such as Gaussian Runge–Kutta) have the property that if $\lambda_i \in$ a stable, centre, respectively unstable set, then $\mu_i \in$ a stable, centre, respectively unstable set, for all τ.

Nongeometric methods will typically not have this property for any τ.

Geometric methods will have this property for sufficiently small τ provided that the spectrum of the equilibrium point is generic to the class of systems. The linearization of the flow will be perturbed by the integrator to a nearby matrix lying in the same class. Hence its spectrum will retain any forced or generic properties of this class: for Hamiltonian systems, the eigenvalues will come in $\lambda, \bar{\lambda}, \lambda^{-1}, \bar{\lambda}^{-1}$ quadruplets, and, if they are elliptic and generic, will remain so. If the eigenvalues are nongeneric (e.g. double), or if the time step is too large, then the eigenvalues may change type.

Geometric methods, therefore, will undergo generic bifurcations at parameter values close to those at which the ODE bifurcates. (Nongeneric bifurcations are usually due to the presence of some other structure, such as a symmetry, which should therefore be preserved by the integrator to restore genericity.) Time-symmetric, A-stable Runge–Kutta methods do even better, for they perturb each eigenvalue by the same function, by $\lambda_i \mapsto \mu_i := R(\tau\lambda_i)$ where R is analytic at 0. Bifurcations therefore occur at precisely the correct parameter values, and even nongeneric bifurcations with double eigenvalues are captured correctly. Only eigenvalue resonances such as $\lambda_i/\lambda_j \in \mathbb{Q}$ are not captured correctly.

5.1.5. Preservation of chaotic properties ([72], pp 124–126). By their very nature, chaotic orbits are hard (if not impossible) to compute numerically. This is due to the exponential divergence of nearby orbits in chaotic regions. In addition, the long-time behaviour of some orbits can be extraordinarily subtle. This means that in general much care is needed and the convergence as $\tau \to 0$ of any solution features should be carefully checked.

KAM tori can often act as a (partial) boundary between 'ordered' and 'chaotic' regions in a phase space, and this is another reason why it is good to preserve such tori, e.g. by using symplectic integrators.

On the other hand, chaotic orbits in hyperbolic systems possess a shadowing property [66]. This property is structurally stable, and hence preserved by symplectic integrators. Reich [108] has used this property to prove that numerically computed time averages are close to the real averages. Geometric properties of the system (especially integrals, symplectic structure and volume preservation) constrain the Lyapunov exponents of a chaotic orbit; hence geometric integrators that obey these constraints can give more accurate estimates of the exponents. (See figure 5.)

5.2. Error growth

Suppose the ODE $\dot{x} = f(x)$ with solution $x(t)$ is integrated by $x_{k+1} = \varphi_\tau(x_k), x_0 = x(0)$. The traditional objects of study in numerical analysis are the local error $x(\tau) - x_1$ after one step and the global error $x_k - x(t)$, typically with $\tau \to 0$ and the total integration time $t = k\tau$ held fixed. For a method of order p, the local error will be $\mathcal{O}(\tau^{p+1})$ and the global error $\mathcal{O}((e^{Lt} - 1)\tau^p)$, where L is the Lipshitz constant of f in the domain of integration ($L = \max\|f'(x)\|$). However, in geometric integration, the step size is often held fixed (and is even quite large) while the number of steps $k \to \infty$. The global error bound is useless in this situation, and indeed, the global errors themselves typically would be large, especially on chaotic orbits.

If the global errors are small, then the error in any observed feature of the solution is also small. If the global errors are large, then the error in an observed feature can be small or large; the convergence of each feature has to be studied in its own right.

There are two situations, however, in which the global error of a geometric integrator can in fact be dramatically smaller than that of a standard integrator. The first case is when the system has some special structure that can be exploited to decrease the error, as in nearly-integrable systems (section 4.2) and systems for which the method (15) is efficient. The second case is when the solution is quasiperiodic. In a standard integrator applied to an integrable Hamiltonian system, for example, the action errors for almost all initial conditions are $\mathcal{O}(t)$, leading to frequency errors also being $\mathcal{O}(t)$; these compound to give rise to $\mathcal{O}(t^2)$ global errors. In a symplectic integrator the action and frequency errors are bounded on exponentially long-time intervals; the orbit moves on a torus with slightly wrong frequencies, leading to $\mathcal{O}(t)$ global errors. (See figure 2.) In a system with a mixture of regular and chaotic motion, one could still expect some favourable error behaviour near the regular orbits.

If the global errors are $\mathcal{O}(\tau^p t)$, then it can be tempting to obtain small global errors for extremely long runs. For example, this has been attempted for the solar system. In this case, the error will in practise be dominated by round-off error which grows at best like $\mathcal{O}(t^{3/2}/\tau)$. The Extreme case must be taken to minimize the round-off error, including the use of compensated summation [53] and extended double precision.

The integrator φ_τ and the flow $e^{\tau f}$ are both maps of phase space in the same class: symplectic, reversible, volume preserving, etc. They are both $\mathcal{O}(\tau)$-close to the identity and $\mathcal{O}(\tau^p)$ close to each other. Any property of maps that is preserved by arbitrarily

small perturbations within the appropriate class will be preserved by the integrator. This includes many of the key features of the class of dynamics, such as preservation of KAM tori, genericity of bifurcations, Poincaré recurrence, and so on. However, sometimes one can say even more, using the theory of *modified equations* or *backward error analysis* [108]. This states that the integrator $\varphi(\tau)$ is, up to a small error, close to the time-τ flow of an *autonomous* modified vector field $\tilde{f}$ satisfying $f - \tilde{f} = \mathcal{O}(\tau^p)$. If $\varphi(\tau)$ is analytic in τ then the error $\varphi(\tau) - \mathrm{e}^{\tau\tilde{f}}$ is exponentially small in τ. If $\varphi(\tau)$ is symplectic (respectively reversible, respectively volume preserving) then $\tilde{f}$ is Hamiltonian (respectively reversible, respectively divergence-free). Backward error analysis can be applied both to reduce the study of maps to that of differential equations, which can be simpler, and to any case in which the dynamics of maps and flows is different.

The champion example is energy preservation. Hamiltonian systems preserve energy but symplectic maps, generically, do not. However, the modified vector field $\tilde{f}$ is Hamiltonian and hence $\tilde{f} = \mathcal{J}\nabla\widetilde{H}$; $\widetilde{H}$ is called the modified Hamiltonian of φ. Taking care of the small error terms gives the result that there exists a time step τ_0 such that

$$\widetilde{H}(x_k) = \widetilde{H}(x_0) + \mathcal{O}(\mathrm{e}^{-\tau_0/2\tau}), \tag{74}$$

implying

$$H(x_k) = \widetilde{H}(x_0) + \mathcal{O}(\tau^p), \tag{75}$$

for all $|\tau| < \tau_0$ and for all $|t| < \mathrm{e}^{\tau_0/2\tau}$. The energy error oscillates within fixed bounds for exponentially long times. Decreasing the time step until the energy errors are bounded is a crucial part of any symplectic integration, although it does not, of course, ensure that any other features are yet correct.

The behaviour of other first integrals that are not exactly preserved by the integrator can be difficult to determine. Numerically, both linear $\mathcal{O}(t)$ increases in error and bounded $\mathcal{O}(1)$ errors are observed in different cases. Bounded errors are presumably due to the presence of a modified first integral, but these have only been constructed in special cases, for example, when the Hamiltonian and integral are both separable [126].

When reversible systems are integrated by reversible integrators, a variety of behaviour in first integrals can be observed. For nonsymmetric orbits, reversibility is irrelevant, and $\mathcal{O}(t)$ error growth is to be expected. For symmetric orbits the errors are bounded for quasiperiodic orbits and $\mathcal{O}(\sqrt{t})$ for ergodic orbits [85]. While this random-walk-like drift can still be an obstacle to, e.g., very long molecular dynamics simulations, it is still much better than the $\mathcal{O}(t)$ drift of a standard integrator.

Systems with multiple time scales can possess adiabatic invariants J that obey $|J(t) - J(0)| = \mathcal{O}(\varepsilon)$ for $|t| < \mathrm{e}^{c/(2\varepsilon)}$, where $\varepsilon \ll 1$ is the ratio of slow-to-fast frequencies. If the time step is small enough for backward error analysis to apply, i.e. if the fast frequencies are well resolved, then one can show that the integrator obeys $|J(x_k) - J(x_0)| = \mathcal{O}(\tau^p) + \mathcal{O}(\varepsilon)$ also for exponentially long-time intervals [108]. Even if the time step is larger than the period of the fast motion, and one is applying a special method for such highly oscillatory systems as in section 4.7, and time step resonances are avoided, then adiabatic invariants are observed not to drift, although the magnitude of their oscillations may not be correct.

5.3. Stability

It might be expected that splitting methods, being nominally explicit, have only modest stability. This is not necessarily the case.

(i) Because they are 1-step methods, they are automatically 0-stable (stable for $\dot{x} = 0$), unlike multistep methods.

(ii) If the phase space is compact, all integrators are unconditionally stable. Examples are integrators preserving the integral $\|x\|^2$, unitary and orthogonal integrators used in quantum mechanics, and symplectic integrators for Poisson systems with compact symplectic leaves, such as the rigid body.

(iii) In general, most geometric integrators are only conditionally stable. One chooses a simple test problem of the relevant type with bounded solutions and computes the stability limit of various methods. For Hamiltonian systems, the test problem is usually the harmonic oscillator with $T(p) + V(q)$ splitting. Leapfrog is stable for $\tau < 2$ and higher-order m-stage integrators are stable for $\tau < \tau^*$ where usually $\tau^* \approx \pi$. Special compositions can be found that are more stable and allow the use of a larger time step. Typically the stability limit τ^* decreases with increasing order, while τ^*/m increases slightly with m, where m is the number of stages.

(iv) Some nonlinear stability can follow merely from the group property, such as existence of KAM tori in Hamiltonian, volume-preserving, or reversible systems [17, 110], preservation of invariant hypersurfaces which partition phase space, or the nonlinear stability of fixed points of Hamiltonian systems [112].

(v) The modified Hamiltonian of symplectic integrators can confer nonlinear stability. In particular, for some splitting methods, the critical points of H do not move or change their value under the perturbation due to the integrator [87].

(vi) Implicit geometric integrators such as symplectic RK and PRK methods can be unconditionally stable for linear systems. However, this does not always allow one to increase the time step over that of an explicit method, because the implicit equations become more expensive to solve and may not even have a solution, and because nonlinear stability for large time steps is not guaranteed [113].

(vii) Integrators for highly oscillatory systems (section 4.7) are expressly designed to allow much greater time steps and better stability [71].

6. Other problems in numerical ODEs

In this survey we have mostly considered the solution of initial value problems for ODEs. Other types of problems include the following.

6.1. Boundary-value problems

The geometry of boundary-value problems for the various classes of systems (Hamiltonian, symmetric etc.) has received extremely little attention in the literature. The same can be said about the development and testing of geometric integrators for these problems. The close relationship between boundary- and initial-value problems does suggest, however, that structure preservation should be useful in much the same situations in the two cases. There are two key differences. First, boundary-value problems are usually solved on fixed and moderate time intervals (partly because they become increasingly ill-conditioned as the time interval increases), so the utility of having a good long-time dynamics in the integrator is decreased. Second, the balance between explicit and implicit integrators shifts markedly. For general boundary-value problems, one can choose between shooting (typically with an explicit integrator) or a direct solution (typically with an implicit integrator). In the latter case, Gaussian Runge–Kutta is *already* the preferred method for boundary-value problems for first-order ODEs [3]. Since it preserves so much structure as well, this appears to make it the method of choice for geometric boundary-value problems, despite being implicit.

6.2. *Optimization and control problems*

In optimization and control problems, and when continuing the solution of a boundary-value problem with parameters, one has to repeatedly solve (sometimes approximately) a sequence of nearby systems. This further favours implicit methods, because good initial guesses, factored Jacobian matrices, etc, are known. With an implicit Runge–Kutta method one could expect to apply just one or two Newton-like updates per step. There can also be an advantage to treating the discretization and the optimization in parallel, as in the 'discrete mechanics and optimal control' method [64], in which both the dynamics and optimization are specified, and discretized, variationally.

6.3. *Non-trajectory methods*

In the *Hybrid Monte Carlo* method [31] for the potential $V(q)$, a sequence of orbit segments are computed for the Hamiltonian (e.g.) $H = \frac{1}{2}\|p\|^2 + V(q)$, each initialized with a random value of p. As fidelity to the dynamics of H is not required for good results, there is scope for adapting the integrator. For example, in [92], explicit integrators that preserve volume, reversibility and energy are constructed for some systems, leading to correct statistics with no rejected steps.

Subdivision Techniques [32] discretize the phase space itself to a set of cells, with the dynamics replaced by a Markov chain on the cells. The transition probabilities of the Markov chain are computed by a Monte Carlo method from a set of orbit segments with random initial conditions. Analysis of the Markov chain yields approximates of global quantities like invariant measures and almost invariant sets. Although the methods can be successfully applied to low-dimensional Hamiltonian systems [34], as far as we know the question of what geometric structure the Markov chain can inherit from the ODE has not yet been examined.

6.4. *Poincaré maps*

The computation of Poincaré maps is often done without preserving any structure; if it is determined that the section is crossed between x_k and x_{k+1}, say, then the value $\hat{x}$ of the map can be approximated by interpolation (often inverse Hermite interpolation is used, because the derivatives of $x(t)$ are known from the differential equation), while the orbit calculation proceeds from x_{k+1}, so that the overall orbit calculation is given by a geometric integrator. Unfortunately, this Poincaré map does not necessarily inherit the geometric structure from the ODE, which can be important if the map is to be used, for example, in a study of bifurcation of periodic or quasiperiodic orbits. Other popular methods (such as solving $\varphi_\tau(x_k) \in \Sigma$ for τ, or, when the section Σ is defined by $g(x) = 0$, letting $g(x)$ be a new dependent variable so that Σ can be reached in one step [118]) are not geometric either. However, some geometric approaches have been explored. First, If the section has small enough dimension, the Poincaré map can be (nonsymplectically) approximated on a grid on the section; then a (symplectic) Poincaré map can be constructed by interpolating a generating function for the map on the grid [39]. The map will be extremely fast to evaluate, but will suffer from an interpolation error. Second, in accelerator physics, the Poincaré map is known to have an elliptic fixed point, and it can be approximated by its Taylor (or Taylor–Fourier) series. A symplectic map is then constructed that has the same Taylor series up to some order [38, 121]. Third, a more conventional approach is to seek a function $C(x)$ such that $g(x) := \dot{C} > 0$ in the domain of interest. Under the transformed Hamiltonian $g(x)^{-1}(H - E)$ on energy level E, $\dot{C} = 1$. Hence, the section defined by $=$ constant can be reached with a constant step size.

Acknowlegments

The authors are grateful to W-J Beyn and Shang Zai-jiu for useful correspondence, to David McLaren, Stephen Marsland, Dion O'Neale, Brett Ryland, Will Wright, Philip Zhang and the referees for their comments, and to Maaike Wienk for editorial assistance. This work was supported by the Australian Research Council, the Marsden Fund of the Royal Society of New Zealand, and the New Zealand Institute of Mathematics and its Applications.

References

[1] Anteneodo C and Tsallis C 1998 Breakdown of exponential sensitivity to initial conditions: role of the range of interactions *Phys. Rev. Lett.* **102** 5313–6
[2] Arnold D N 2002 Differential complexes and numerical stability *Proceedings of the International Congress of Mathematicians (Beijing, 2002)* vol 1 (Beijing: Higher Education Press) pp 137–57
[3] Ascher U M, Mattheij R M M and Russell R D 1995 *Numerical Solution of Boundary Value Problems for Ordinary Differential Equations* (Philadelphia: SIAM)
[4] Ascher U M and Petzold L R 1998 *Computer Methods for Ordinary Differential Equations and Differential-Algebraic Equations* (Philadelphia: SIAM)
[5] Barth E, Leimkuhler B and Reich S 1999 A time-reversible variable-stepsize integrator for constrained dynamics *SIAM J. Sci. Comput.* **21** 1027–44
[6] Blanes S 2001 High order numerical integrators for differential equations using composition and processing of low order methods *Appl. Numer. Math.* **37** 289–306
[7] Blanes S and Casas F On the existence of positive coefficients for operator splitting schemes of order higher than two *Preprint* GIPS 2003-04 http://www.focm.net/gips
[8] Blanes S and Casas F 2006 Explicit Magnus expansions for nonlinear equations *J. Phys. A: Math. Gen.* **39** 5405–23
[9] Blanes S, Casas F and Ros J 2000 Symplectic integrators with processing: a general study *SIAM J. Sci. Comput.* **21** 711–27
[10] Blanes S, Casas F and Ros J 2000 Processing symplectic methods for near-integrable Hamiltonian systems *Celest. Mech. Dyn. Astron.* **77** 17–35
[11] Blanes S, Casas F and Ros J 2001 High-order Runge–Kutta–Nyström geometric methods with processing *Appl. Numer. Math.* **39** 245–59
[12] Blanes S, Casas F and Ros J 2000 Improved high-order integrators based on the Magnus expansion *BIT* **40** 434–50
[13] Blanes S and Moan P C 2001 Splitting methods for non-autonomous Hamiltonian equations *J. Comput. Phys.* **170** 205–30
[14] Blanes S and Moan P C 2002 Practical symplectic partitioned Runge–Kutta and Runge–Kutta–Nyström methods *J. Comput. Appl. Math.* **142** 313–30
[15] Bobenko A I and Suris Yu B 1999 Discrete time Lagrangian mechanics on Lie groups, with an application to the Lagrange top *Commun. Math. Phys.* **204** 147–88
[16] Bridges T and Reich S 2006 Numerical method for Hamiltonian PDEs *J. Phys. A: Math. Gen.* **39** 5287–320
[17] Broer H W, Huitema G B and Sevryuk M B 1996 Quasi-periodic motions in families of dynamical systems *(Lecture Notes in Mathematics* vol 1645*)* (Berlin: Springer)
[18] Budd C J and Iserles A 1999 Geometric integration: numerical solution of differential equations on manifolds *Phil. Trans. R. Soc.* A **357** 945–56
[19] Budd C J and Piggott M D 2003 Geometric integration and its applications *Handbook of Numerical Analysis* vol 11 (Amsterdam: North-Holland) pp 35–139
[20] Butcher J C 1987 *The Numerical Analysis of Ordinary Differential Equations* (Chichester: Wiley)
[21] Butcher J C 2003 *Numerical Methods for Ordinary Differential Equations* (Chichester: Wiley)
[22] Calvo M, Hernández-Abreu D, Montijano J I and Randez L Explicit Runge–Kutta methods for the preservation of invariants *SIAM J. Sci. Comput.* in press
[23] Calvo M P, Iserles A and Zanna A 1997 Numerical solution of isospectral flows *Math. Comput.* **66** 1461–86
[24] Casas F and Owren B 2003 Cost-efficient Lie group integrators in the RKMK class *BIT* **43** 723–42
[25] Celledoni E, Marthinsen A and Owren B 2003 Commutator-free Lie group methods *Future Gener. Comput. Syst.* **19** 341–52
[26] Chambers J E 1999 A hybrid symplectic integrator that permits close encounters between massive bodies *Mon. Not. R. Astron. Soc.* **304** 793–9

[27] Chambers J E 2003 Symplectic integrators with complex time steps *Astron. J.* **126** 1119–26
[28] Chartier P, Faou E and Murua A An algebraic approach to invariant preserving integrators: the case of quadratic and Hamiltonian invariants *Numer. Math.* (submitted)
[29] Chin S A and Kidwell D W 2000 Higher-order force gradient symplectic algorithms *Phys. Rev.* E **62** 8746–52 (Part B)
[30] Cortes J and Martinez S 2001 Non-holonomic integrators *Nonlinearity* **14** 1365–92
[31] Creutz M and Gocksch A 1989 Higher-order hybrid Monte Carlo algorithms *Phys. Rev. Lett.* **63** 9–12
[32] Dellnitz M and Junge O 2002 Set-oriented numerical methods for dynamical systems *Handbook of Dynamical Systems* vol 2 (Amsterdam: North-Holland) pp 221–64
[33] Deuflhard P 1979 A study of extrapolation methods based on multistep schemes without parasitic solutions *Z. Angew. Math. Phys.* **30** 177–89
[34] Deuflhard P, Dellnitz M, Junge O and Schütte Ch 1998 Computation of essential molecular dynamics by subdivision techniques. I. Basic concept *Computational Molecular Dynamics: Challenges, Methods, Ideas (Lecture Notes in Computational Science and Engineering* vol 4*)* ed P Deuflhard, J Hermans, B Leimkuhler, A E Mark, S Reich and R D Skeel (Berlin: Springer) pp 98–115
[35] Dorodnitsyn V A 1994 Finite difference models entirely inheriting continuous symmetry of original differential equations *Int. J. Mod. Phys.* C **5** 723–34
[36] Feng K and Wang D-L 1994 Dynamical systems and geometric construction of algorithms *Contemporary Mathematics* vol 163 (Providence: AMS) pp 1–32
[37] Feng K and Wang D-L 1998 Variations on a theme by Euler *J. Comput. Math.* **16** 97–106
[38] Forest E 2006 Geometric integration for particle accelerators *J. Phys. A: Math. Gen.* **39** 5321–77
[39] Froeschle C and Lega E 1997 Global symplectic polynomial approximation of area-preserving maps *J. Diff. Eqns. Appl.* **3** 169–84
[40] Garcıa-Archilla B, Sanz-Serna J-M and Skeel R D 1999 Long-time-step methods for oscillatory differential equations *SIAM J. Sci. Comput.* **20** 930–63
[41] Ge Z and Marsden J 1998 Lie–Poisson Hamilton–Jacobi theory and Lie–Poisson integrators *Phys. Lett.* A **133** 135–9
[42] Golubitsky M and Stewart I 2002 The symmetry perspective. From equilibrium to chaos in phase space and physical space *Progress in Mathematics vol 200* (Basel: Birkhäuser Verlag)
[43] Grimm V and Hochbruck M 2006 Non-smooth error bounds for a family of exponential integrators applied to highly oscillatory second-order differential equations *J. Phys. A: Math. Gen.* **39** 5495–507
[44] Grimm V and Quispel G R W 2005 Geometric integration methods that preserve Lyapunov functions *BIT* **45** 709–23
[45] Grimm V and Quispel G R W Geometric integration methods that unconditionally contract volume *Appl. Numer. Math.* submitted
[46] Hairer E and Lubich C 2004 Symmetric multistep methods over long times *Numer. Math.* **97** 699–723
[47] Hairer E, Lubich C and Wanner G 2002 *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations* (Berlin: Springer)
[48] Hairer E, Lubich C and Wanner G 2003 Geometric integration illustrated by the Störmer–Verlet method *Acta Numer.* 399–450
[49] Hairer E, Nørsett S P and Wanner G 1993 *Solving Ordinary Differential Equations. I. Nonstiff Problems* 2nd edn (Berlin: Springer)
[50] Hairer E and Söderlind G 2005 Explicit, time reversible, adaptive step size control *SIAM J. Sci. Comput.* **26** 1838–51
[51] Hairer E and Wanner G 1996 *Solving Ordinary Differential Equations. II. Stiff and Differential-Algebraic Problems* 2nd edn (Berlin: Springer)
[52] Hardy D J, Okunbor D I and Skeel R D 1999 Symplectic variable step size integration for N-body problems *Appl. Numer. Math.* **29** 19–30
[53] Higham N J 1993 The accuracy of floating point summation *SIAM J. Sci. Comput.* **14** 783–99
[54] Hochbruck M and Lubich C 2003 On Magnus integrators for time-dependent Schrödinger equations *SIAM J. Numer. Anal.* **41** 945–63
[55] Holt C R 1982 Construction of new integrable Hamiltonians in two degrees of freedom *J. Math. Phys.* **23** 1037–46
[56] Hydon P and Mansfield E L 2004 A variational complex for difference equations *Found. Comput. Math.* **4** 187–217
[57] Iserles A 1996 *A first course in the numerical analysis of differential equations* (Cambridge: Cambridge University Press)
[58] Iserles A 2004 On the method of neumann series for highly oscillatory equations *BIT* **44** 473–88

[59] Iserles A and Nørsett S P 1999 On the solution of linear differential equations in Lie groups *Phil. Trans. R. Soc.* A **357** 983–1019
[60] Iserles A, McLachlan R I and Zanna A 1999 Approximately preserving symmetries in the numerical integration of ordinary differential equations *Eur. J. Appl. Math.* **10** 419–45
[61] Iserles A, Munthe-Kaas H Z, Nørsett S P and Zanna A 2000 Lie group methods *Acta Numer.* **9** 215–365
[62] Jay L 1996 Symplectic partitioned Runge-Kutta methods for constrained Hamiltonian systems *SIAM J. Numer. Anal.* **33** 368–87
[63] Jia Z and Leimkuhler B 2006 Geometric integrators for multiple time-scale simulation *J. Phys. A: Math. Gen.* **39** 5379–403
[64] Junge O, Marsden J E and Blöbaum S Discrete mechanics and optimal control *Preprint*
[65] Kang F, Qin M-Z, Wang D-L and Wu H-M 1989 Construction of canonical difference schemes for Hamiltonian formalism via generating functions *J. Comput. Math.* **7** 71–96
[66] Katok A and Hasselblatt B 1995 *Introduction to the modern theory of dynamical systems* (Cambridge: Cambridge University Press)
[67] Kværnø A and Leimkuhler B 2000 A time-reversible, regularized, switching integrator for the N-body problem *SIAM J. Sci. Comput.* **22** 1016–35
[68] Lamb J S W and Roberts J A G 1998 Time-reversal symmetry in dynamical systems: a survey *Phys. D* **112** 1–39
[69] Laskar J and Robutel P 2001 High order symplectic integrators for perturbed Hamiltonian systems *Celest. Mech.* **80** 39–62
[70] Laskar J, Robutel P, Joutel F, Gastineau Correia A C M and Levrard B 2004 A long-term numerical solution for the insolation quantities of the earth *Astron. Astrophys.* **428** 261–85
[71] Leimkuhler B and Reich S 2001 A reversible averaging integrator for multiple time-scale dynamics *J. Comput. Phys.* **171** 95–114
[72] Leimkuhler B and Reich S 2004 *Simulating Hamiltonian Dynamics* (Cambridge: Cambridge University Press)
[73] Lew A, Marsden J E, Ortiz M and West M 2003 Asynchronous variational integrators *Arch. Rat. Mech. Anal.* **167** 85–146
[74] Lew A, Marsden J E, Ortiz M and West M 2004 Variational time integrators *Int. J. Numer. Methods Eng.* **60** 153–212
[75] López-Marcos M A, Sanz-Serna J M and Skeel R D 1997 Explicit symplectic integrators using Hessian-vector products *SIAM J. Sci. Comput.* **18** 223–38
[76] Marsden J E, Patrick G W and Shadwick W F (ed) 1996 *Integration Algorithms and Classical Mechanics* (Providence: AMS)
[77] Marsden J E and Ratiu T S 1999 *Introduction to Mechanics and Symmetry* 2nd edn (New York: Springer-Verlag)
[78] Marsden J E and West M 2001 Discrete mechanics and variational integrators *Acta Numer.* **10** 357–514
[79] McLachlan R I 1993 Explicit Lie–Poisson integration and the Euler equations *Phys. Rev. Lett.* **71** 3043–6
[80] McLachlan R I 1995 On the numerical integration of ordinary differential equations by symmetric composition methods *SIAM J. Sci. Comput.* **16** 151–68
[81] McLachlan R I 1995 Composition methods in the presence of small parameters *BIT* **35** 258–68
[82] McLachlan R I 2002 Families of high-order composition methods *Numer. Alg.* **31** 233–46
[83] McLachlan R I and Atela P 1992 The accuracy of symplectic integrators *Nonlinearity* **5** 541–62
[84] McLachlan R I and Perlmutter M 2001 Conformal Hamiltonian systems *J. Geom. Phys.* **39** 276–300
[85] McLachlan R I and Perlmutter M 2004 Energy drift in reversible time integration *J. Phys. A: Math. Gen.* **37** L593–8
[86] McLachlan R I and Perlmutter M Integrators for nonholonomic mechanical systems *J. Nonlinear Sci.* at press
[87] McLachlan R I, Perlmutter M and Quispel G R W 2004 On the nonlinear stability of symplectic integrators *BIT* **44** 99–117
[88] McLachlan R I, Perlmutter M and Quispel G R W 2003 Lie group foliate integrators *Future Gener. Comput. Syst.* **19** 1207–19
[89] McLachlan R I and Quispel G R W 2002 Splitting methods *Acta Numer.* **11** 341–434
[90] McLachlan R I and Quispel G R W 2001 Six lectures on geometric integration *Foundations of Computational Mathematics* ed R DeVore, A Iserles and E Süli (Cambridge: Cambridge University Press) pp 155–210
[91] McLachlan R I and Quispel G R W 2001 What kinds of dynamics are there? Lie pseudogroups, dynamical systems, and geometric integration *Nonlinearity* **14** 1689–706
[92] McLachlan R I and Quispel G R W 2003 Geometric integration of conservative polynomial ODEs *Appl. Numer. Math.* **45** 411–8

[93] McLachlan R I and Quispel G R W 2004 Explicit geometric integration of polynomial vector fields *BIT* **44** 515–38
[94] McLachlan R I, Quispel G R W and Robidoux N 1999 Geometric integration using discrete gradients *Phil. Trans. R. Soc.* A **357** 1021–46
[95] McLachlan R I, Quispel G R W and Turner G S 1998 Numerical integrators that preserve symmetries and reversing symmetries *SIAM J. Numer. Anal.* **35** 586–99
[96] McLachlan R I and Ryland B N 2003 The algebraic entropy of classical mechanics *J. Math. Phys.* **44** 3071–87
[97] McLachlan R I and Scovel C 1995 Equivariant constrained symplectic integration *J. Nonlinear Sci.* **5** 233–56
[98] McLaren D I and Quispel G R W 2004 Integral-preserving integrators *J. Phys. A: Math. Gen.* **37** L489–95
[99] Minchev B V and Wright W M A review of exponential integrators for first order semi-linear problems *Preprint* http://www.math.ntnu.no/preprint/numerics/2005
[100] Moan P C 2004 On the KAM and nekhoroshev theorems for symplectic integrators and implications for error growth *Nonlinearity* **17** 67–83
[101] Moan P C, Murua A, Quispel G R W, Sofroniou M and Spaletta G Symplectic elementary differential Runge–Kutta methods *Preprint*
[102] Munthe-Kaas H and Zanna A 1997 Numerical integration of differential equations on homogeneous manifolds *Foundations of Computational Mathematics (Rio de Janeiro, 1997)* ed F Cucker and M Shub (Berlin: Springer) 305–15
[103] Pandolfi A, Kane C, Marsden J E and Ortiz M 2002 Time-discretized variational formulation of nonsmooth frictional contact *Int. J. Numer. Methods Eng.* **53** 1801–29
[104] Olver P J 1993 *Applications of Lie groups to differential equations* 2nd edn (New York: Springer-Verlag)
[105] Owren B 2006 Order conditions for commutator-free Lie group methods *J. Phys. A: Math. Gen.* **39** 5585–99
[106] Quispel G R W 1995 Volume-preserving integrators *Phys. Lett.* A **206** 226–30
[107] Reich S 1993 Numerical integration of the generalized Euler equations *Technical Report* 93–20 Department of Computer Science, University of British Columbia
[108] Reich S 1993 Backward error analysis for numerical integrators *SIAM J. Numer. Anal.* **36** 1549–70
[109] Sanz-Serna J-M and Calvo M P 1994 *Numerical Hamiltonian Problems* (London: Chapman and Hall)
[110] Shang Z J 2000 Resonant and diophantine step sizes in computing invariant tori of Hamiltonian systems *Nonlinearity* **13** 299–308
[111] Sheng Q 1989 Solving linear partial differential equations by exponential splitting *IMA J. Numer. Anal.* **9** 199–212
[112] Skeel R D and Srinivas K 2000 Nonlinear stability analysis of area-preserving integrators *SIAM J. Numer. Anal.* **38** 129–48
[113] Simo J C and Tarnow N 1994 A new energy and momentum conserving algorithm for the non-linear dynamics of shells *Int. J. Numer. Methods Eng.* **37** 2527–49
[114] Sofroniou M and Spaletta G 2005 Derivation of symmetric composition constants for symmetric integrators *Optimization Methods and Software* **20** 597–613
[115] Stuart A M and Humphries A R 1996 *Dynamical systems and numerical analysis* (Cambridge: Cambridge University Press)
[116] Suzuki M 1990 Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations *Phys. Lett.* A **146** 319–23
[117] Takahashi M and Imada M 1984 Monte Carlo calculations of quantum systems. II higher order correction *J. Phys. Soc. Japan* **53** 3765–9
[118] Tucker W 2002 Computing accurate Poincaré maps *Phys.* D **171** 127–37
[119] Verlet L 1967 Computer 'experiments' on classical fluids. I thermodynamical properties of Lennard–Jones molecules *Phys. Rev.* **159** 98–103
[120] Walton A R and Manolopoulos D E 1996 A new semiclassical initial value method for Franck–Condon spectra *Mol. Phys.* **87** 961–78
[121] Warnock R L and Ellison J A 1999 From symplectic integrator to Poincaré map: spline expansion of a map generator in Cartesian coordinates *Appl. Numer. Math.* **29** 89–98
[122] Weideman J A C and Herbst B M 1986 Split-step methods for the solution of the nonlinear Schrödinger equation *SIAM J. Numer. Anal.* **23** 485–507
[123] Wisdom J and Holman M 1991 Symplectic maps for the N-body problem *Astron. J.* **102** 1528–38
[124] Wisdom J, Holman M and Touma J 1996 Symplectic correctors *Integration Algorithms and Classical Mechanics* ed J E Marsden, G W Patrick and W F Shadwick (Providence: AMS) pp 217–44
[125] Yoshida H 1990 Construction of higher order symplectic integrators *Phys. Lett.* A **150** 262–8
[126] Yoshida H 2001 Non-existence of the modified first integral by symplectic integration methods *Phys. Lett.* A **282** 276–83