Nonlinearity 3 (1990) 231–259. Printed in the UK

Symplectic integration of Hamiltonian systems*

P J Channell† and C Scovel‡

† Accelerator Theory and Simulation Group, AT-6, and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

‡ Mathematical Analysis and Applications Group, T-7, and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received 22 August 1988, in final form 22 September 1989

Abstract. We survey past work and present new algorithms to numerically integrate the trajectories of Hamiltonian dynamical systems.These algorithms exactly preserve the symplectic 2-form, i.e. they preserve all the Poincaré invariants. The algorithms have been tested on a variety of examples and results are presented for the Fermi–Pasta–Ulam nonlinear string, the Henon–Heiles system, a four-vortex problem, and the geodesic flow on a manifold of constant negative curvature. In all cases the algorithms possess long-time stability and preserve global geometrical structures in phase space.

AMS classification scheme numbers: 65L05, 70H99

1. Introduction

Hamiltonian dynamical systems are studied using a variety of techniques including, increasingly, numerical integration. However, standard numerical integration schemes neglect important special features of the dynamics, in particular the fact that the time-δt map of phase space is symplectic [1] (e.g., the motion of the phase-space points from time 0 to time t preserves the Poincaré invariants). A symplectic transformation is one that satisfies [2]

$$f'^{\mathrm{T}}(q,p)Jf'(q,p) = J$$

where f' is the Jacobian matrix of derivatives of the time-δt map, $f(q,p)$, of phase space, q and p are vectors of canonically conjugate coordinates, and J is the matrix

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

with I the identity matrix. Note that this condition holds at every point of phase space and that a consequence of it is Liouville's theorem: phase-space volume is preserved. However, it is much stronger than Liouville's theorem and, as a result, places stringent conditions on the global geometry of the dynamics. Standard numerical integration schemes do not respect these restrictions even if energy conservation is built into them.

* This work supported by the US Department of Energy, Office of High Energy and Nuclear Physics and Office of Mathematics and Computation.

0951-7715/90/020231+29$03.50

As a result, numerical integration has usually been limited to investigations over a limited time; one suspects that these investigations are prone to miss global structures commonly known to be present because of the Hamiltonian nature of the dynamics. Thus, standard integration techniques are useful in the investigation of short-time quantitative phenomena, but may be severely limited in the investigation of long-time qualitative phenomena. Throughout this paper we use the word 'stability' to mean the preservation of global structures and qualitative phenomena, and not in the way commonly found in the numerical analysis literature.

It is possible to devise numerical integration algorithms that approximate the time-δt map of the exact dynamics to any desired order in the timestep and that are exactly symplectic. We refer to these as symplectic integration algorithms (SIAs). These algorithms give the exact evolution of a Hamiltonian system with a Hamiltonian function that is close to the Hamiltonian of interest to any desired order and preserve all the Poincaré invariants [3], namely the integrals

$$\int\int \sum_i \mathrm{d}q_i\,\mathrm{d}p_i \qquad \int\int\int\int \sum_{i\neq k} \mathrm{d}q_i\,\mathrm{d}p_i\,\mathrm{d}q_k\,\mathrm{d}p_k \qquad \ldots$$

over any appropriate even-dimensional surface are unchanged by a timestep.

Symplectic integration algorithms were first introduced by DeVogelaere [4] in 1956 in a series of unpublished reports. They were rediscovered in 1983 by Ruth [5] who devised an analogue of the Runge–Kutta algorithms for Hamiltonians of potential form and independently by one of the authors [6] who developed implicit schemes for general Hamiltonians similar to those of DeVogelaere. Subsequently and independently, Kang [7] investigated the same problem and developed algorithms based on Padé approximants. Neri [8] has shown how to extend the method of Ruth to arbitrary order. The applicability of canonical integrators to discretisations of Hamiltonian field equations [9] has been demonstrated. At this time, Ruth and Forest [10] are continuing the development of the Runge–Kutta class of explicit algorithms. In addition, Menyuk [11] has developed a symplectic differencing scheme for a particular class of Hamiltonians and noted the possibility of its extension to the general case.

A general technique, used by most authors, for deriving SIAs is to make use of the observation [3] that any transformation derived from a generating function is a symplectic map†. Thus, in order to produce a symplectic time-δt map accurate to some given order in δt, we compute the time-δt map of the Hamiltonian system to that order (with a typically non-symplectic result) and require that the map produced from a generating function agree with it to that order; we then use this generating function to derive the timestep algorithm. There are many ways to approximate the time-δt map of the Hamiltonian system, and many ways to match the generating function map to it, so that many different SIAs are possible. Algorithms investigated so far have included Runge–Kutta techniques, Adams techniques, Padé approximants and analytic formulations. The relative efficiency and/or simplicity of the methods vary from one problem to another, and there appears to be no 'best' approach in general. Nonetheless, they all share the stability properties of symplectic integration and are superior to standard ODE integrators in the investigation of global, long-time phase-space structures, though presently standard integrators are likely to be more useful for short-time quantitative phenomena because the packages have been more thoroughly debugged and tested.

† The work of Neri, however, is based on Lie algebraic techniques.

2. Hamiltonians of potential form

Hamiltonians of potential form, i.e.

$$H = \frac{p^2}{2} + V(q)$$

have been the systems most commonly studied, primarily because models that decompose into a kinetic energy term with constant mass and a potential term have been the easiest to understand geometrically (by drawing pictures of the potential function) and thus have been considered the most useful models. There is therefore a strong historical and practical motivation to study these systems despite the fact that they are not generic in the category of Hamiltonian systems. In this section we will develop a set of SIAs for these systems and examine two examples, the Fermi–Pasta–Ulam nonlinear string and the Hénon–Heiles model, to illustrate the power of SIAs. The most effective approach we have discovered to finding a generating function for Hamiltonians of potential form is with the Hamilton–Jacobi equation:

$$\frac{\partial S}{\partial t} + \frac{1}{2}\left(\frac{\partial S}{\partial q}\right)^2 = -V(q).$$

Recall that the negative of the action S is a generating function of the first kind [12] for the motion through the equations

$$p_0 = -\frac{\partial S}{\partial q_0} \tag{1}$$

and

$$p = \frac{\partial S}{\partial q}. \tag{2}$$

To develop a useful form for the solution, we introduce an artificial parameter ϵ, which we will eventually set equal to 1. We then let $V(q) \to \epsilon V(q)$ and expand the action S as

$$S = S_0 + \sum_{n=1}^{\infty} \epsilon^n S_n.$$

Using this expansion in the Hamilton–Jacobi equation and equating equal powers of ϵ we obtain

$$\frac{\partial S_1}{\partial t} + \frac{\partial S_0}{\partial q} \cdot \frac{\partial S_1}{\partial q} = -V(q) \tag{3}$$

and

$$\frac{\partial S_n}{\partial t} + \frac{\partial S_0}{\partial q} \cdot \frac{\partial S_n}{\partial q} = -\frac{1}{2} \sum_{\substack{l=1,m=1 \\ l+m=n}} \frac{\partial S_l}{\partial q} \cdot \frac{\partial S_m}{\partial q} \qquad \text{for } n > 2. \tag{4}$$

In general, a solution to the Hamilton–Jacobi equation generates a solution to Hamilton's equations, but the relation $\dot{q} = \partial S/\partial q$ implies that $p = \dot{q}$ is determined as a function of q, which makes it difficult to solve the initial value problem. However, if we choose the zero-order action as

$$S_0 = \frac{(q - q_0)^2}{2t} \tag{5}$$

then the singularity at $t = 0$ removes the dependence of p on q and gives us enough freedom to specify the momentum at $t = 0$ while forcing q to be q_0. The solution determined by the truncated generating function

$$S_N = \sum_{i=0}^{N} S_i$$

is an exact solution for the time-dependent Hamiltonian given by solving for q_0 in terms of $p = \partial S_N/\partial q$ and q and inserting the result in

$$H_N = -\frac{\partial S_N(q, q_0)}{\partial t}.$$

This Hamiltonian approximates the complete Hamiltonian to order N. The proof of this, assuming sufficient smoothness of the Hamiltonian, is a straightforward application of the Taylor series. For the exact system, the action for the flow backward in time can be obtained by letting $t \to -t$ or by interchanging q and q_0 and letting $S \to -S$. With the above choice of S_0 this property also holds for truncation at any order S_N, i.e.

$$S_i(q, q_0, t) = -S_i(q_0, q, -t). \tag{6}$$

As a consequence, the numerical solution is also time reversible.

To solve (3) for S_1, we expand $V(q)$ in a Taylor series about an arbitrary point, $\bar{q}$

$$V(q) = \sum_{n=0}^{\infty} \frac{(q - \bar{q})^n}{n!} \overline{\frac{\partial^n V}{\partial q^n}}$$

where the bar indicates that the function is to be evaluated at $\bar{q}$. Using the fact that the characteristics of the left-hand side of (3) are

$$q = q_0 + v_0 t$$

where q_0 and v_0 are the initial position and velocity, along with the binomial theorem, we find that S_1 is given by

$$S_1 = -\bar{V}t - t \sum_{n=1}^{\infty} \frac{1}{n!} \overline{\frac{\partial^n V}{\partial q^n}} \sum_{l=0}^{n} \frac{1}{n - l + 1} \binom{n}{l} (q_0 - \bar{q})^l (q - \bar{q})^{n-l}. \tag{7}$$

We can then compute $\partial S_1/\partial q$ and insert the result into the equation for S_2, which has the same characteristics as the equation for S_1. It can be integrated to give

$$S_2 = -\frac{t^3}{2} \sum_{m,n=1}^{\infty} \frac{1}{m!n!} \overline{\frac{\partial^m V}{\partial q^m}} \overline{\frac{\partial^n V}{\partial q^n}} \sum_{k=0}^{m} \sum_{l=0}^{n} \left(\frac{n-l}{n-l+1}\right) \left(\frac{m-k}{m-k+1}\right) \binom{n}{l} \binom{m}{k}$$
$$\times \frac{(q_0 - \bar{q})^l (q_0 - \bar{q})^k (q - q_0)^{m-k-1} (q - q_0)^{n-l-1}}{m + n - l - k + 1}. \tag{8}$$

We can now evaluate $\partial S_2/\partial q$ and compute the right-hand side of the equation for S_3, which can again be integrated to give

$$S_3 = -t^5 \sum_{m,n,s=1}^{\infty} \frac{1}{m!n!s!} \overline{\frac{\partial^s V}{\partial q^s}} \overline{\frac{\partial^m V}{\partial q^m}} \overline{\frac{\partial^n V}{\partial q^n}} \sum_{j=0}^{s}\sum_{k=0}^{m}\sum_{l=0}^{n} \binom{n}{l}\binom{m}{k}\binom{s}{j}\left(\frac{s-j}{s-j+1}\right)\left(\frac{n-l}{n-l+1}\right)$$
$$\times \left(\frac{m-k}{m-k+1}\right)\left(\frac{m-k-1}{m+n-l-k+1}\right)\left(\frac{1}{m+n+s-j-k-l+1}\right)$$
$$\times (q_0-\bar{q})^l (q_0-\bar{q})^j (q_0-\bar{q})^k (q-q_0)^{s-j-1}(q-q_0)^{n-l-1}(q-q_0)^{m-k-2}. \quad (9)$$

Observe that $S_i = t^{2i-1}$ times a function of q and q_0. In particular, the symmetry (6) implies that each S_i and therefore S_N is a symmetric function of q and q_0. We can proceed in this way to calculate S to any order desired. Note that the expansion of V still contains all the derivatives of V. This expansion should be truncated at an order consistent with the truncation in powers of ϵ. Once the derivatives of S have been computed to find the transformation, the point $\bar{q}$ can be set to any convenient value; in particular, one choice that insures that $q-\bar{q}$ is small (at least of order δt) is $\bar{q} = q_0$. An alternative approach to the solution of the Hamilton–Jacobi equation has been presented by Molzahn and Osborn [13].

Given S_N for some N, (1) and (2) are used to generate the time step. Note that (1) is usually an implicit equation for q but, given its solution, (2) is then explicit for p. Let us now consider two examples using algorithms derived from (7), (8) and (9).

2.1. The Fermi–Pasta–Ulam problem

To test this procedure we have applied it to the Fermi–Pasta–Ulam problem which is governed by the potential

$$V = \beta \sum_{i=1}^{n} q_i \frac{(q_{i+1}+q_{i-1}-2q_i)}{2}\left[1+\alpha(q_{i+1}-q_{i-1})\right]$$

where q_i are the components of q, β and α are constants (taken to be 1.0 and 0.5 in our examples), and where we set $n = 64$. The Hamiltonian system with this potential is one possible discretisation of the partial differential equation

$$\frac{\partial^2\phi}{\partial t^2} = \beta \frac{\partial^2\phi}{\partial x^2}\left(1+2\alpha\frac{\partial\phi}{\partial x}\right)$$

that is a nonlinear string with a derivative-dependent mass. Note, by the way, that a Hamiltonian field equation can be differenced so that the difference equations are no longer Hamiltonian; the above differencing is Hamiltonian, but in the general case care must be exercised to obtain equations in Hamiltonian form. This system has been extensively studied [14] and motivated much of the work that led to the discovery [15] of exactly soluble nonlinear partial differential equations; the system is still under active investigation. Our purpose in this section is not to add to the understanding of the Fermi–Pasta–Ulam problem itself, but to use it to illustrate that SIAs can be applied to large problems, to introduce the kinds of numerical checks we applied to all our examples, and to give a first indication of the kind of stability that one can expect of SIAs. For the Hamiltonian with this potential, we computed S_1 and S_2 as well as the

partial derivatives with respect to q and q_0; then $\bar{q}$ was set equal to q_0. Using these results, a straightforward iterative solution to (1) was programmed; that is, using the expression for S_0, (1) becomes

$$q = q_0 + p_0 t - \frac{\partial(S_1 + S_2)}{\partial q_0}. \tag{10}$$

On the first iteration, q was set equal to q_0 on the right-hand side of this equation, q was computed and then reinserted into the right-hand side, and this procedure was repeated until the change in q was less than round-off. This iteration seemed to converge well in all cases. Note that more sophisticated strategies for solving (10) exist; some have been tested with a considerable speed-up of the algorithm; see, for example, appendix A.

The resulting algorithm is second order in the timestep. Some call the order of an algorithm the order of the error after a single timestep. We, however, will call the order of an algorithm the order of the error to reach a fixed time; this is one order less than the single timestep order. We numerically tested the order of the algorithm with the results shown in table 1. As can be seen, the algorithm is confirmed to be of second order.

Table 1. Error scaling as the timestep is successively halved for the Fermi–Pasta–Ulam problem using an SIA. The constant ratio of successive errors confirms that the algorithm is second order.

Timestep (δt_n)	Error (ϵ_n)	$\epsilon_{n-1}/\epsilon_n$
0.05	5.779×10^{-5}	—
0.025	1.444×10^{-5}	4.00
0.0125	3.61×10^{-6}	4.00
0.00625	9.025×10^{-7}	4.00

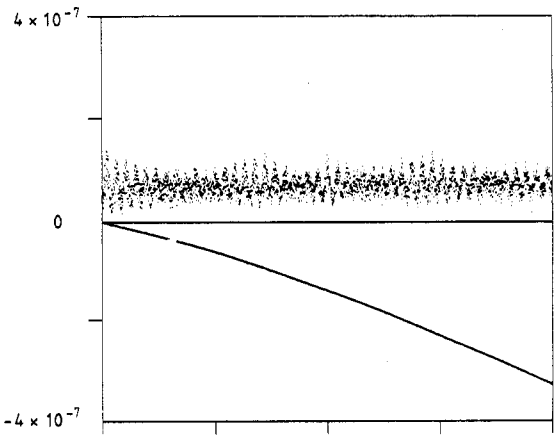

Figure 1. Comparison of relative energy error of a second-order SIA and a fourth-order RKI for the Fermi–Pasta–Ulam system of oscillators. The timestep for the SIA was 0.001 and 0.02 for the RKI and the final time was 7500.

To check the algorithm further, we compared energy conservation using the symplectic algorithm to the energy error using a standard fourth-order Runge–Kutta integrator (RKI), with fixed timestep. The results are shown in figure 1. Although the initial energy error using the RKI was about three orders of magnitude smaller than that of the symplectic integrator because of its higher order, the energy error using the RKI increased monotonically with time and eventually grew larger than the error using the symplectic integrator, which seems to have a quasiperiodic oscillation about a slightly displaced energy and shows no secular growth with time. Continuing the computation for longer times gives continued growth in the Runge–Kutta error until the error is so large that the calculation is nonsense, while the SIA error only shows continued oscillation with the same central value and amplitude. If a smaller timestep is used with the RKI, the time at which its error magnitude exceeds that of the SIA can be delayed, but the qualitative behaviour is the same and the finer timestep integration with the RKI must also be terminated eventually.

In all the examples we discuss in this paper, the order of the algorithm used was checked with tests identical to that used in constructing table 1, and the putative order was always confirmed. In addition, the behaviour of the energy error was always monitored and compared with an RKI with results similar to those of figure 1. Therefore, we will not repeat this figure and table for each of the examples. We will, however, show the energy error figure for the completely chaotic system to be considered in section 3.2 where the symplectic and RKI performed similarly.

2.2. *The Hénon–Heiles problem*

The classic paper of Hénon and Heiles [16] investigated the question of the existence of a third isolating integral in a six-dimensional system modeling stellar motion in a galaxy. Using the known second integral, their system reduces to that of a Hamiltonian given by

$$H(q,p) = \tfrac{1}{2}(q_1^2 + q_2^2 + p_1^2 + p_2^2) + q_1^2 q_2 - \tfrac{1}{3}q_2^3.$$

They found that the third integral seemed to exist, numerically, at low energy but seemed to disappear as the energy was increased. Gustavson [17] has shown that when the integral exists numerically, it is approximated accurately by a formal expansion similar to that of Birkhoff [18], though the expansion cannot be expected to converge in general. Gustavson's results do show, however, that for low energy, this Hamiltonian is nearly integrable. We will use this system to illustrate the applicability of the above algorithms, to contrast the behaviour of SIAs and standard RKIs, and to demonstrate the variety and importance of global geometrical structures in Hamiltonian systems.

Hénon and Heiles used a surface-of-section technique to visualise the behaviour of the trajectories; the coordinates (p_2, q_2) were plotted in the plane whenever the trajectory crossed the hyperplane $q_1 = 0$. Note that the fourth coordinate, p_1, is not independent, as it can be determined from the other three coordinates and the energy. We will, for the moment, employ this same technique in our comparison of integrators. Figure 2 shows the results of integrating a single initial condition for 1 200 000 timesteps of $\delta t = 1/6$. In figure 2(*a*), the SIA has produced a sharply defined level surface with every indication that the trajectory will remain on a well defined submanifold almost indefinitely. In figure 2(*b*), the RKI follows approximately the same trajectory but, as can be seen, the object is much less well defined and there is little indication that it might be a submanifold. Though typically one would not integrate for so long with

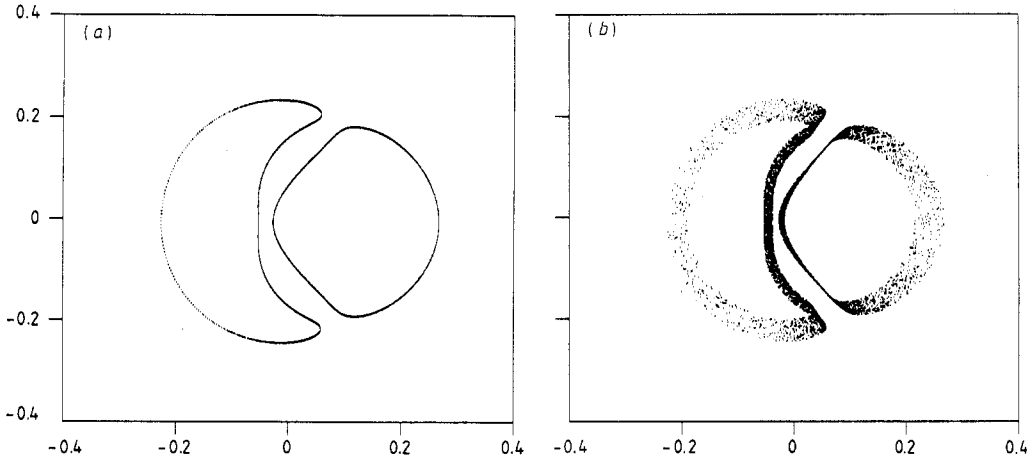

Figure 2. Comparison of third-order SIA (a) and a fourth-order RKI (b) for the Hénon–Heiles system. The initial condition was (0.12, 0.12, 0.12, 0.12) and energy 0.029 952. The timestep was 1/6 and 1200 000 timesteps were computed.

such a crude timestep using the RKI, integrations with smaller timesteps merely reduce the scale at which the destruction of global structures occurs. To investigate these smaller structures, however, requires longer time runs with more steps during which the deviation of the RKI from a global structure continues to accumulate. The SIA preserves these structures even with a large timestep.

In figure 3, we contrast the behaviour of the SIA and the RKI for a different kind of trajectory. The SIA, after 100 000 timesteps, traced out a resonant island structure (figure 3(a)). The RKI (figure 3(b)), after 100 000 timesteps, remained somewhat localised but gave no indication that the islands exist. In figure 3(c), the SIA, with a slightly different initial condition, produces a trajectory that is ergodic over a large region but that carefully avoids some embedded regions, presumably surrounded by intact closed curves. The RKI with the same initial condition produced a trajectory that is visually identical to figure 3(b) and that neither avoids the embedded stable regions nor ergodically fills the large region indicated in figure 3(c). Thus, the SIA seems to yield trajectories that lie on actual submanifolds when the motion is regular and seems to avoid stable regions when the motion is irregular—the RKI does neither of these. Figure 4 shows a period doubling transition that occurs because of a very small shift in the energy (5 parts in 10^5). Note that the SIA detected this transition even though the timestep was relatively crude (about 20 time steps per linear period). Finally, the results of running a variety of different initial conditions at the same energy are shown (figure 5) in a blow-up of the region $|q_2| \leq 0.1$ and $|p_2| \leq 0.1$. An assortment of behaviours is clearly distinguished by the SIA: trajectories that fill out a curve (torus in 3D), low-order resonances, high-order resonances, and a trajectory that fills out some region.

The ability of a SIA to clearly define global objects is best illustrated using three-dimensional graphics; we used an Evans–Sutherland PS-300 to view a projection (instead of a slice) of the 3D objects defined by (p_2, q_2, q_1) along a trajectory. In figure 6(a) a two-dimensional slice at $q_1 = 0$ of what appears to be a high-order resonance is shown. In figure 6(b), a 3D projection onto the $q_1 = 0$ plane reveals the underlying global structure that contains this slice. A different perspective (figure 6(c)),

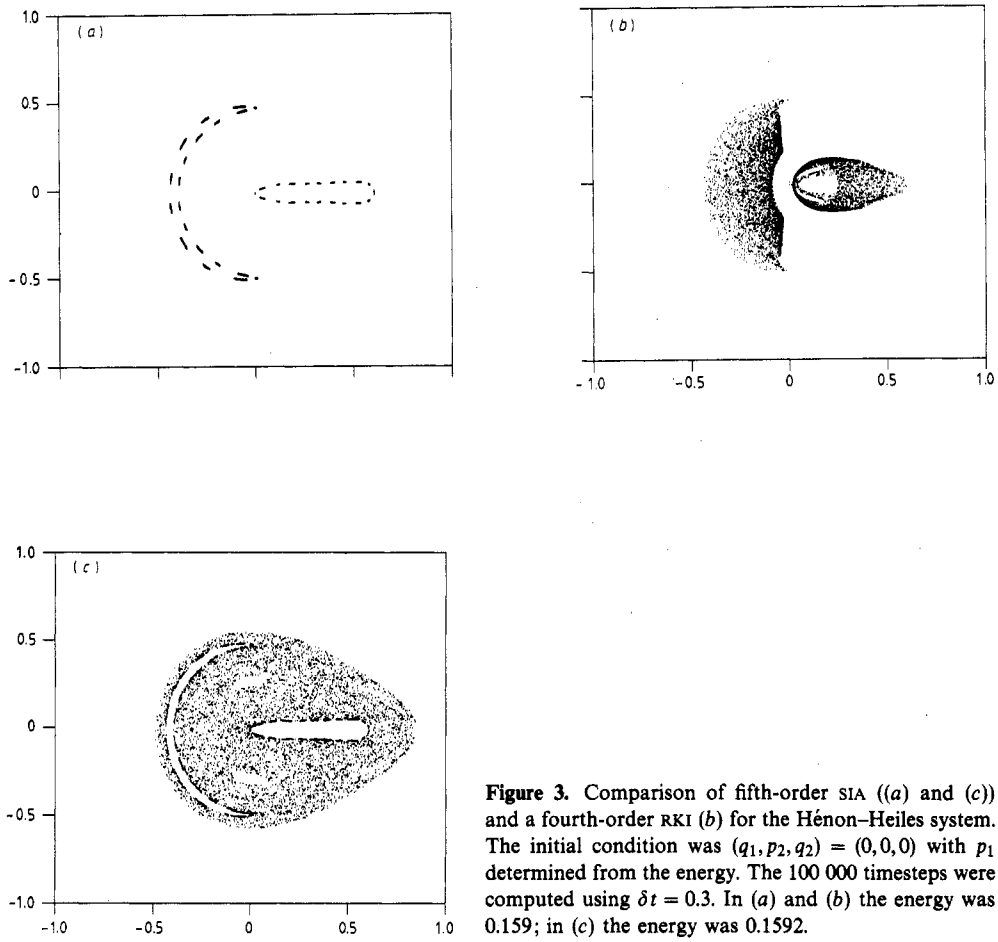

Figure 3. Comparison of fifth-order SIA ((*a*) and (*c*)) and a fourth-order RKI (*b*) for the Hénon–Heiles system. The initial condition was $(q_1, p_2, q_2) = (0,0,0)$ with p_1 determined from the energy. The 100 000 timesteps were computed using $\delta t = 0.3$. In (*a*) and (*b*) the energy was 0.159; in (*c*) the energy was 0.1592.

rotated 90° about the q_2 axis, of this same object reveals the discrete D^3 symmetry, i.e. invariance under 120° rotations and reflections due to the D^3 invariance of the potential. Notice also the precisely closed precessional behaviour in this perspective and the fact that the SIA achieves this by maintaining the D^3 symmetry of the exact system. In figure 6(*d*), a close-up from a different perspective of the central loop in figure 6(*b*) was made from runs using timesteps of $\delta t = 0.3$ (yellow) and $\delta t = 0.15$ (blue). This figure shows that halving the time step produces a structure that is essentially the same, differing only in certain local details. In addition, closer inspection using the interactive capability of the PS-300 reveals that the trajectory triply intertwines with itself forming an intricate local structure that would not result were it not for the global properties of the SIA.

A nice framework to discuss geometrical structures in Hamiltonian systems has been provided by Fomenko [19] in his treatment of the topology of energy surfaces of integrable systems. Though the Hénon–Heiles system is not integrable, it displays many vestiges of integrable behaviour, as was pointed out by Gustavson [17]. Figure 7 shows two structures. One is a twisted torus that is probably a remnant of a two-dimensional subset of what Fomenko calls a non-orientable saddle (A^3). The second structure appears to be a Klein bottle but a normal form analysis provided by Dragt

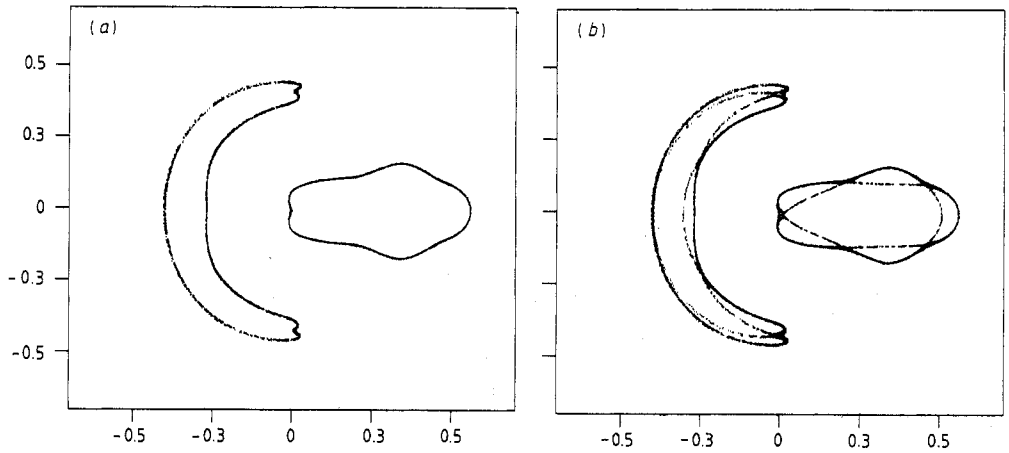

Figure 4. Results of a fifth-order SIA for the Hénon–Heiles system. The initial condition was $(q_1, p_2, q_2) = (0, 0, 0)$ with p_1 determined from the energy. The 200 000 timesteps were computed using $\delta t = 0.3$. In (a) the energy was 0.117 83, and in (b) the energy was 0.117 835.

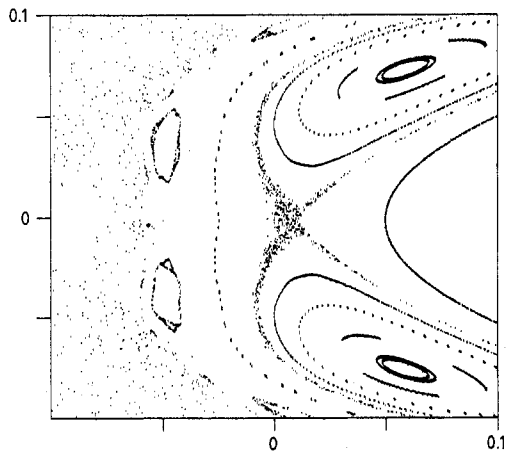

Figure 5. Magnification of results of ten runs with different initial conditions of 100 000 timesteps of $\delta t = 0.3$ at an energy of 0.117 85 using a fifth-order SIA for the Hénon–Heiles system.

and Neri reveals that it is just a singular projection of a standard 2-torus. Of course, not all structures are remnants of integrable objects, so not all will fall into Fomenko's classification; figure 6 might be one example of a non-integrable object. Nevertheless, many objects can be identified as belonging to one of Fomenko's categories.

Paralleling the sequence of figures $4(a, b)$, we demonstrate in figure 8 the ability of the SIA to distinguish the different global objects that can result from slightly different initial conditions. For 5000 timesteps of $\delta t = 0.3$, we integrated twenty trajectories with initial values evenly spaced from $(p_2, q_2) = (0, 0)$ to $(p_2, q_2) = (0, -0.1)$. We have chosen six representative trajectories, beginning in figure $8(a)$, with simple transitive flow on a 2-torus. In figure $8(b)$, it appears that the flow on the torus has coalesced onto a periodic orbit; but in fact, by linear interpolation of successive points, we see that

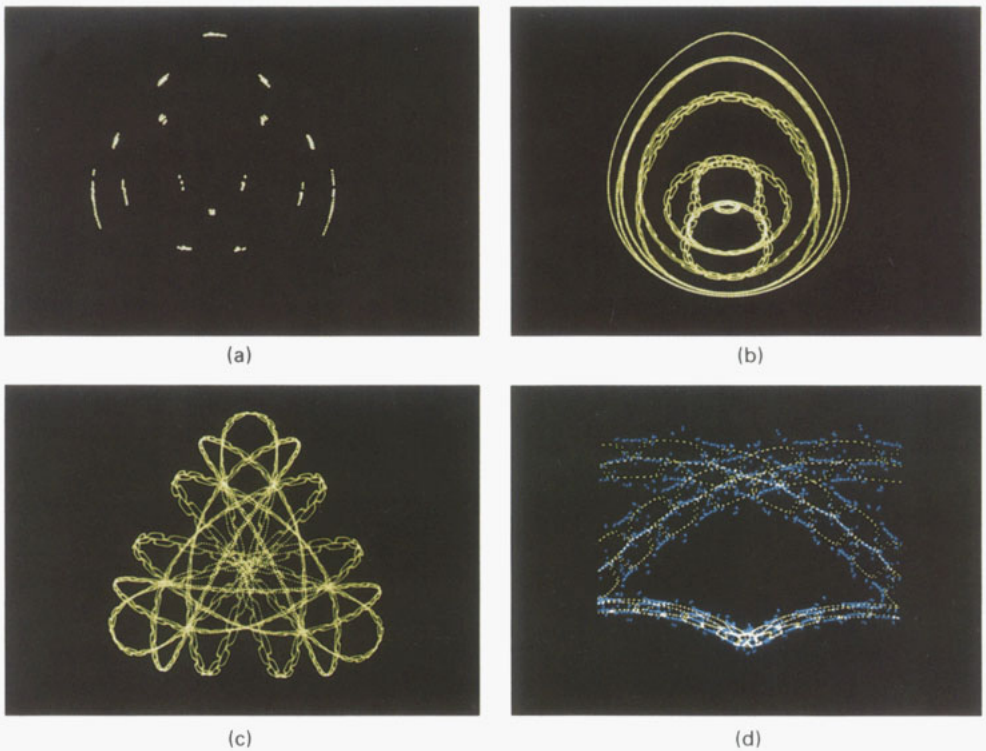

Figure 6. Four views of the results from a single initial condition, $(q_1, p_2, q_2) = (0, 0, -0.1)$, with energy 0.117 835: (*a*) two-dimensional slice at the $q_1 = 0$ plane; (*b*) projection of the three-dimensional phase space onto the $q_1 = 0$ plane; (*c*) projection of the three-dimensional phase space onto the $p_2 = 0$ plane; (*d*) Comparison of results using timesteps of $\delta t = 0.3$ (yellow) and $\delta t = 0.15$ composed twice (blue).

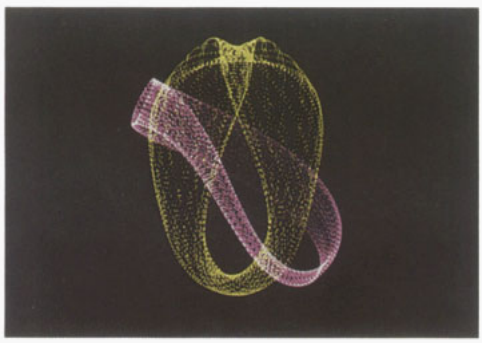

Figure 7. The 5000 timesteps using a fifth-order SIA at two different initial conditions for the Hénon–Heiles system with timestep $\delta t = 0.3$. One initial condition was $(q_1, p_2, q_2) = (0, 0.1, 0.15)$ with an energy of 0.1, while the other was at $(q_1, p_2, q_2) = (0, 0, 0.2)$ with an energy of 0.117 835.

the dynamics is again transitive. The size of the timestep is resonant with the periods on this torus, giving a numerical flow that is not ergodic on it [20]. In figure 8(*c*), we notice the beginning of the break-up of the 2-torus into the structure shown in figure 8(*d*) by forming a (genuine) resonance and the emergence near the centre of a

separate strand. In figure 8(*e*), the structure has almost collapsed onto a periodic orbit, but in figure 8(*f*), it has re-emerged as a broad structure similar to figure 8(*d*).

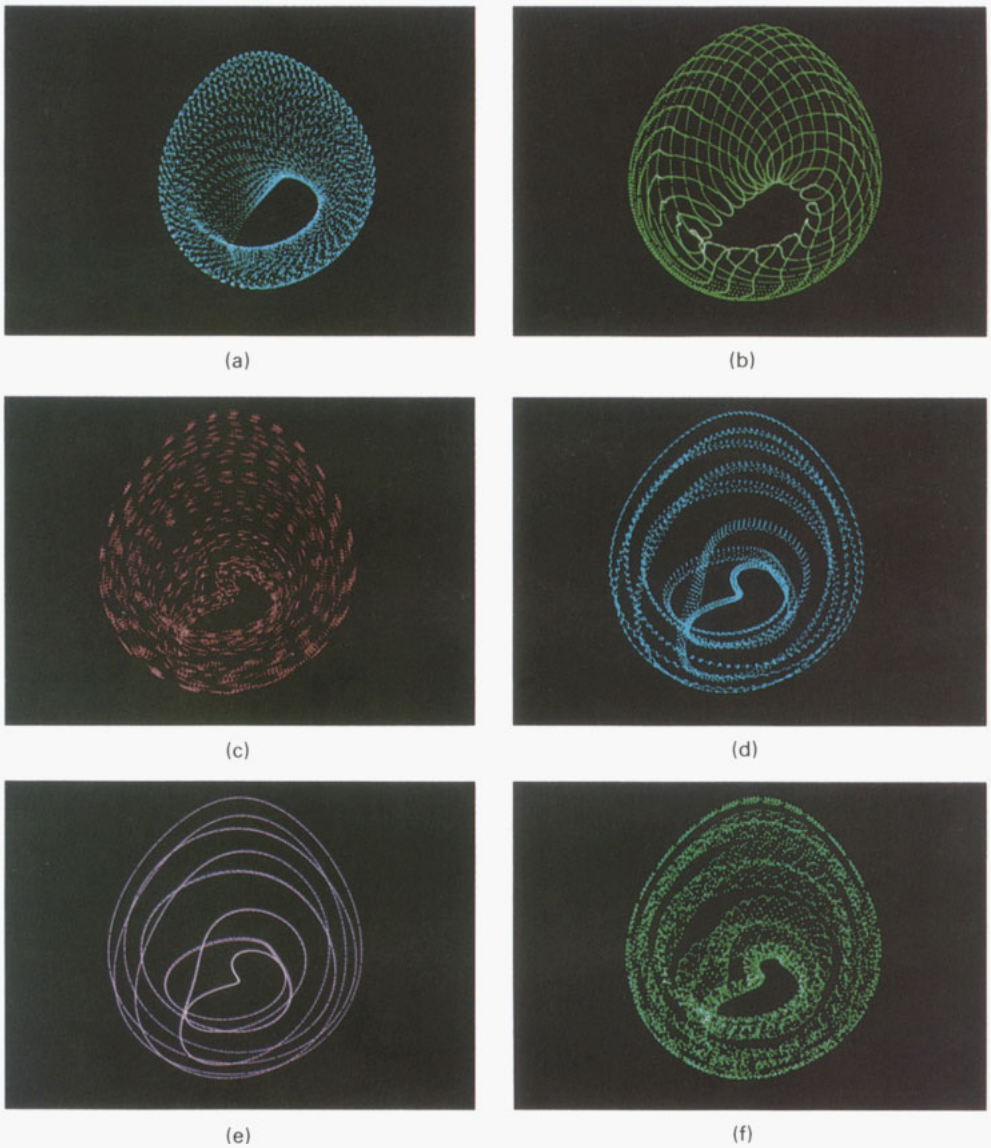

Figure 8. The 5000 timesteps using a fifth-order SIA at six different initial conditions for the Hénon–Heiles system with timestep $\delta t = 0.3$ and the same energy of 0.11. (*a*) $(q_1, p_2, q_2) = (0, 0, -0.010)$, (*b*) $(q_1, p_2, q_2) = (0, 0, -0.035)$, (*c*) $(q_1, p_2, q_2) = (0, 0, -0.055)$, (*d*) $(q_1, p_2, q_2) = (0, 0, -0.065)$, (*e*) $(q_1, p_2, q_2) = (0, 0, -0.070)$, (*f*) $(q_1, p_2, q_2) = (0, 0, -0.080)$.

The full elucidation of phase-space structures requires, by its global nature, long time integrations. It should not be construed that all those phase-space structures displayed by the SIA are those of the actual Hamiltonian system; some of the local structures, as we have seen, result from the finite timestep used in the integration. However, stable structures, such as the invariant tori of the KAM theorem or stable and unstable manifolds, will be faithfully represented for a small enough timestep. The

local details resulting from the finite timestep can change qualitatively as the timestep is varied but do not change quantitatively very much (for small timesteps) because of the accuracy of the algorithm.

3. General Hamiltonians

For Hamiltonians of general form, approaches starting from the Hamilton–Jacobi equation can be developed, but seem to yield algorithms that are difficult to implement and thus are not very useful. Instead, we adopt a different strategy and attempt to match the timestep given by a generating function directly with the timestep given by the equations of motion. It is convenient to use a generating function of the third kind, $K(p_0, q)$, which generates the transformation from the equations

$$p = -\frac{\partial K}{\partial q} \tag{11}$$

and

$$q_0 = -\frac{\partial K}{\partial p_0}. \tag{12}$$

We expand the generating function as

$$K = \sum_{m=0}^{\infty} \frac{\delta t^m}{m!} K_m(p_0, q)$$

where δt is the size of the timestep and where $K_0 = -p_0 \cdot q$ generates the identity. If we let

$$p = p_0 + \sum_{m=1}^{\infty} \frac{\delta t^m}{m!} p_m(p_0, q)$$

we see that

$$p_m = -\frac{\partial K_m(p_0, q)}{\partial q}.$$

Recall that the equations of motion for a Hamiltonian system are

$$\dot{p} = -\frac{\partial H}{\partial q} \tag{13}$$

and

$$\dot{q} = \frac{\partial H}{\partial p}. \tag{14}$$

The time derivative of p is given by

$$\dot{p} = \frac{\partial p}{\partial t} + \frac{\partial p}{\partial q} \cdot \dot{q}.$$

Using the expansion for p and the equation of motion for $\dot{q}$, we find that

$$\sum_{m=1}^{\infty} \frac{p_m \delta t^{m-1}}{(m-1)!} + \sum_{m=1}^{\infty} \frac{\delta t^m}{m!} \frac{\partial p_m}{\partial q} \cdot \dot{q} = -\frac{\partial H}{\partial q}\left(p_0 + \sum_{m=1}^{\infty} \frac{\delta t^m p_m}{m!}, q\right).$$

If we define

$$\Delta p = \sum_{m=1}^{\infty} \frac{\delta t^m}{m!} p_m(p_0, q)$$

and use the Taylor series to expand the arguments, we find that

$$\sum_{m=0}^{\infty} \frac{p_{m+1} \delta t^m}{m!} = \sum_{m=1}^{\infty} \frac{\delta t^m}{m!} \frac{\partial^2 K_m}{\partial q \partial q} \cdot \left[\sum_{l=0}^{\infty} \frac{(\Delta p \cdot \partial/\partial p)^l}{l!}\right] \frac{\partial H}{\partial p} - \sum_{l=0}^{\infty} \frac{(\Delta p \cdot \partial/\partial p)^l}{l!} \frac{\partial H}{\partial q}.$$

Using the binomial theorem to expand the powers of Δp and equating coefficients of equal powers of δt, we obtain

$$\begin{aligned}\frac{\partial K_{m+1}}{\partial q} = -m! \sum_{s=1}^{s} \frac{1}{s!} \frac{\partial^2 K_s}{\partial q \partial q} \cdot \sum_{l=0}^{m-s} \sum_{\substack{r_1,\ldots,r_m=0 \\ \sum r_i = l \\ \sum i r_i = m-s}}^{l} \frac{1}{r_1! \ldots r_m!} \\ \times \left(\frac{p_1}{1!} \cdot \frac{\partial}{\partial p_0}\right)^{r_1} \cdots \left(\frac{p_m}{m!} \cdot \frac{\partial}{\partial p_0}\right)^{r_m} \frac{\partial H}{\partial p_0} \\ + m! \sum_{s=0}^{m} \sum_{\substack{r_1,\ldots,r_m=0 \\ \sum r_i = s \\ \sum i r_i = m}}^{s} \frac{1}{r_1! \cdots r_m!} \left(\frac{p_1}{1!} \cdot \frac{\partial}{\partial p_0}\right)^{r_1} \cdots \left(\frac{p_m}{m!} \cdot \frac{\partial}{\partial p_0}\right)^{r_m} \frac{\partial H}{\partial q}. \end{aligned} \tag{15}$$

This equation is the general result. From this equation one can determine K to any desired order. Evaluating (15) for $m = 0, \ldots, 5$, it is easy to identify the right-hand sides as partial derivatives with respect to q of functions and to find that

$$K_1 = H \tag{16}$$

$$K_2 = -\frac{\partial K_1}{\partial q} \cdot \frac{\partial H}{\partial p_0} \tag{17}$$

$$K_3 = -\frac{\partial K_2}{\partial q} \cdot \frac{\partial H}{\partial p_0} + \sum_{i,j=1}^{n} \frac{\partial K_1}{\partial q_i} \frac{\partial K_1}{\partial q_j} \frac{\partial^2 H}{\partial p_{0i} \partial p_{0j}} \tag{18}$$

$$K_4 = -\frac{\partial K_3}{\partial q} \cdot \frac{\partial H}{\partial p_0} + 3 \sum_{i,j=1}^{n} \frac{\partial K_1}{\partial q_i} \frac{\partial K_2}{\partial q_j} \frac{\partial^2 H}{\partial p_{0i} \partial p_{0j}} - \sum_{i,j,k=1}^{n} \frac{\partial K_1}{\partial q_i} \frac{\partial K_1}{\partial q_j} \frac{\partial K_1}{\partial q_k} \frac{\partial^3 H}{\partial p_{0i} \partial p_{0j} \partial p_{0k}} \tag{19}$$

$$\begin{aligned} K_5 = -\frac{\partial K_4}{\partial q} \cdot \frac{\partial H}{\partial p_0} + 3 \sum_{i,j=1}^{n} \frac{\partial K_2}{\partial q_i} \frac{\partial K_2}{\partial q_j} \frac{\partial^2 H}{\partial p_{0i} \partial p_{0j}} - 6 \sum_{i,j,k=1}^{n} \frac{\partial K_1}{\partial q_i} \frac{\partial K_1}{\partial q_j} \frac{\partial K_2}{\partial q_k} \frac{\partial^3 H}{\partial p_{0i} \partial p_{0i} \partial p_{0i}} \\ + 4 \sum_{i,j=1}^{n} \frac{\partial K_1}{\partial q_i} \frac{\partial K_3}{\partial q_j} \frac{\partial^2 H}{\partial p_{0i} \partial p_{0j}} + \sum_{i,j,k,l=1}^{n} \frac{\partial K_1}{\partial q_i} \frac{\partial K_1}{\partial q_j} \frac{\partial K_1}{\partial q_k} \frac{\partial K_1}{\partial q_l} \frac{\partial^4 H}{\partial p_{0i} \partial p_{0j} \partial p_{0k} \partial p_{0l}} \end{aligned} \tag{20}$$

and

$$K_6 = -\frac{\partial K_5}{\partial q}\cdot\frac{\partial H}{\partial p_0} + 10\sum_{i,j=1}^{n}\frac{\partial K_2}{\partial q_i}\frac{\partial K_3}{\partial q_j}\frac{\partial^2 H}{\partial p_{0i}\partial p_{0j}}$$
$$-10\sum_{i,j,k=1}^{n}\frac{\partial K_1}{\partial q_i}\frac{\partial K_1}{\partial q_j}\frac{\partial K_3}{\partial q_k}\frac{\partial^3 H}{\partial p_{0i}\partial p_{0i}\partial p_{0i}} + 5\sum_{i,j=1}^{n}\frac{\partial K_1}{\partial q_i}\frac{\partial K_4}{\partial q_j}\frac{\partial^2 H}{\partial p_{0i}\partial p_{0j}}$$
$$+10\sum_{i,j,k,l=1}^{n}\frac{\partial K_1}{\partial q_i}\frac{\partial K_1}{\partial q_j}\frac{\partial K_1}{\partial q_k}\frac{\partial K_2}{\partial q_l}\frac{\partial^4 H}{\partial p_{0i}\partial p_{0j}\partial p_{0k}\partial p_{0l}}$$
$$-15\sum_{i,j,k=1}^{n}\frac{\partial K_1}{\partial q_i}\frac{\partial K_2}{\partial q_j}\frac{\partial K_2}{\partial q_k}\frac{\partial^3 H}{\partial p_{0i}\partial p_{0i}\partial p_{0i}}$$
$$-\sum_{i,j,k,l,m=1}^{n}\frac{\partial K_1}{\partial q_i}\frac{\partial K_1}{\partial q_j}\frac{\partial K_1}{\partial q_k}\frac{\partial K_1}{\partial q_l}\frac{\partial K_1}{\partial q_m}\frac{\partial^5 H}{\partial p_{0i}\partial p_{0j}\partial p_{0k}\partial p_{0l}\partial p_{0m}}. \tag{21}$$

Using these expressions, we can evaluate K to sixth order (or any lower order if we like). We then use (11) and (12) to find the timestep. Equation (12) is implicit for q, but when this is solved, (11) is explicit for p. Equation (12) can be rewritten as

$$q = q_0 - \sum_{m=1}^{\infty}\frac{\delta t^m}{m!}\frac{\partial K_m(p_0, q)}{\partial p_0}.$$

As in section 2, this equation can be solved by straightforward iteration, i.e. by substituting q_0 on the right-hand side, computing q, reinserting q on the right-hand side and repeating until q changes by less than round-off. Again, it is possible to use more sophisticated strategies to solve this equation, but we have not exhaustively investigated all the possibilities. Note also that the above treatment can be extended to include time-dependent Hamiltonians [6].

Let us now consider two examples of algorithms derived from (16)–(21).

3.1. The four-vortex problem

The evolution of the positions, represented by complex coordinates $z_j = x_j + \mathrm{i}y_j$, of n vortices in ideal two-dimensional hydrodynamics is governed by the Hamiltonian [21]

$$H = -\frac{1}{4\pi}\sum_{j\neq k}^{n}\Gamma_j\Gamma_k \ln|z_j - z_k|$$

where the bars denote absolute magnitude. The canonically conjugate variables are $q_j = x_j$ and $p_j = \Gamma_j y_j$. Because the Hamiltonian depends only on the relative positions of the vortices, it is invariant under the group of translations and rotations so that

$$\sum \Gamma_j z_j \equiv Q + \mathrm{i}P$$

and

$$\sum \Gamma_j |z_j|^2 \equiv L^2$$

are integrals of the motion. It can be shown that H, L^2, and $P^2 + Q^2$ are integrals in involution [21] so that the three-vortex problem is completely integrable. The four-vortex problem is not integrable, however, and makes a good test for the algorithms for general Hamiltonians. It also allows us to test whether integrals of the motion other than the Hamiltonian are preserved as well as the Hamiltonian is preserved.

The generating function was computed to second order, i.e. through K_2; the algorithm was generated using the method described in appendix B. After confirming the order of the algorithm, runs were made using $\delta t = 0.0025$. Table 2 shows a sampling of the maximum of the error of the invariants P, Q, and L^2 for a time $10\delta t$. The integrals P and Q are preserved to within round-off even though the global error in the energy was of order 10^{-4}. The term L^2 exhibits an error similar in magnitude to that of the Hamiltonian. Therefore, not only are additional integrals preserved as well as the Hamiltonian is, but in some cases they may be preserved much better, because of symmetries of the exact dynamics being retained by the truncated generating function [22].

Table 2. Errors for the exact system constants in the four-vortex problem using a second-order SIA.

H error	P error	Q error	L^2 error
$3.23 \cdot 10^{-5}$	8.88×10^{-16}	1.77×10^{-15}	1.87×10^{-5}

Similar behaviour was found using the algorithms of section 2 for a two-degree-of-freedom quartic oscillator with parameters chosen to make the total angular momentum a constant; the angular momentum, a non-Cartesian function, was preserved to within round-off. Application to the three-body problem preserves the nine integrals corresponding to centre-of-mass, centre-of-velocity, and three angular momenta to within round-off. Marsden [22] has shown that this is a general feature and that the approximate map on the full phase space is a lift of an approximate map on the reduced space, T^*Q/G, where T^*Q is the cotangent bundle of configuration space and G is the Lie group of symmetries of the Hamiltonian.

3.2. Geodesic flow on a doughnut with two holes

The examples we have considered so far have been systems that were either nearly integrable or had chaotic regions that were limited in extent. To check whether SIAs would work as well for completely chaotic systems, we investigated the geodesic flow on a two-dimensional manifold of constant negative curvature of genus 2 (a two-holed doughnut). The Hamiltonian for the geodesic flow on the covering space, the Lobachevski plane, is

$$H = \frac{(1 - q_1^2 - q_2^2)^2 (p_1^2 + p_2^2)}{4}$$

where we are using the Poincaré representation of the Lobachevski plane [23]. This geodesic flow is defined on the entire region $q_1^2 + q_2^2 < 1$. To obtain the geodesic flow on the compact manifold of genus 2, we could integrate entirely in the Lobachevski plane and when done, project the results down onto the compact two-manifold. This would be the analogue of integrating a flow on the standard torus by allowing the

angles to range to infinity and, when the calculation was done, projecting them onto the region 0–2π using the modulo function. In practice, this procedure runs into severe round-off problems (with any integrator) because of the singularity at $q_1^2 + q_2^2 = 1$, where, to preserve energy, the momenta must become very large. Instead, we integrated the trajectories in the Lobachevski plane; when the trajectory crossed the boundary of the fundamental polygon (an octagon in this case), we used the discrete isometry to translate it back into the fundamental polygon. This prevented trajectories from coming too close to $q_1^2 + q_2^2 = 1$ and avoided round-off problems. A good description of this construction of the two-holed doughnut and the necessary isometries can be found in [24]. Briefly, the boundary of the fundamental region is, near the positive q_1 axis, given by

$$1 + q_1^2 + q_2^2 - 2q_1 \frac{1 + R_0^2}{1 - R_0^2} = 0$$

where

$$R_0 = \frac{\sin \beta}{\cos \beta + \sqrt{\cos 2\beta}} \approx 0.216\,845\,34$$

with $\beta = \pi/8$. The boundary in the other octants can be obtained from this expression by rotating by multiples of $\pi/4$. When a trajectory crossed this boundary in the first octant, it was transformed back inside the fundamental polygon by the transformation

$$\bar{q}_1 = \frac{ab(q_1^2 + q_2^2) + q_1(a^2 + b^2) + ab}{b^2(q_1^2 + q_2^2) + 2abq_1 + a^2} \tag{22}$$

$$\bar{q}_2 = \frac{q_2(a^2 - b^2)}{b^2(q_1^2 + q_2^2) + 2abq_1 + a^2} \tag{23}$$

where $a = 1 + 1/R_0^2$ and $b = 1 - 1/R_0^2$. The transformation in the other octants can be obtained from this expression by rotating by multiples of $\pi/4$. Denoting the transformation of coordinates, including rotations, by f, the transformation of the momenta is given by

$$\bar{p} = (T^* f)^{-1} p \tag{24}$$

where we are using the notation of [1].

This dynamical system is known to be strongly ergodic; in fact, it is rigorously Bernoulli [25]. This is confirmed by figure 9, where, instead of plotting q_1 and q_2 at intervals of $\Delta t = 1.0$, we have scaled the variables radially so that equal visual areas correspond to equal metric areas. When the energy error was monitored for both a fourth-order SIA and a fourth-order Runge–Kutta integrator, we obtained the results shown in figure 10. As can be seen, there is no significant difference between the two sets of results. In fact, both integrators will duplicate exactly the topological structure of the time-δt map of the exact system for all time. This is a consequence of the Anosov theorem [26], which states that any sufficiently nearby map is topologically conjugate to a given Anosov map, though the conjugacy required to map the trajectories to those of the exact system need not be symplectic, nor even differentiable. Note, however, that because this system is known to have positive Kolmogorov entropy [27], the accuracy

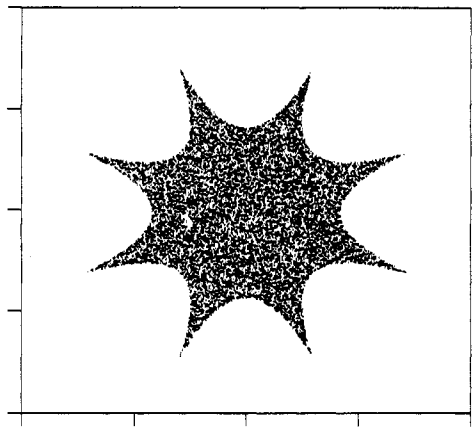

Figure 9. Plot of (q_1, q_2) scaled radially at intervals of $\Delta t = 1$ for the geodesic flow on a two-holed doughnut from a fourth-order SIA. The timestep was $\delta t = 0.001$ and the energy was 0.2. The 5000 points plotted have the uniform distribution one expects for this system.

of any integrator decreases exponentially with time, meaning that a certain number of bits of accuracy are lost per timestep. After a relatively short time, any integrator will have errors in the positions that are as large as the entire manifold. Thus, if one knows *a priori* that one's system is Anosov there is no structural reason to choose one integrator over another, and one can simply use the most convenient integrator. However, the number of cases that are rigorously known to be completely chaotic is small and if any remnant of non-chaotic behaviour is suspected, we will argue in section 5 that there are good reasons to use a SIA.

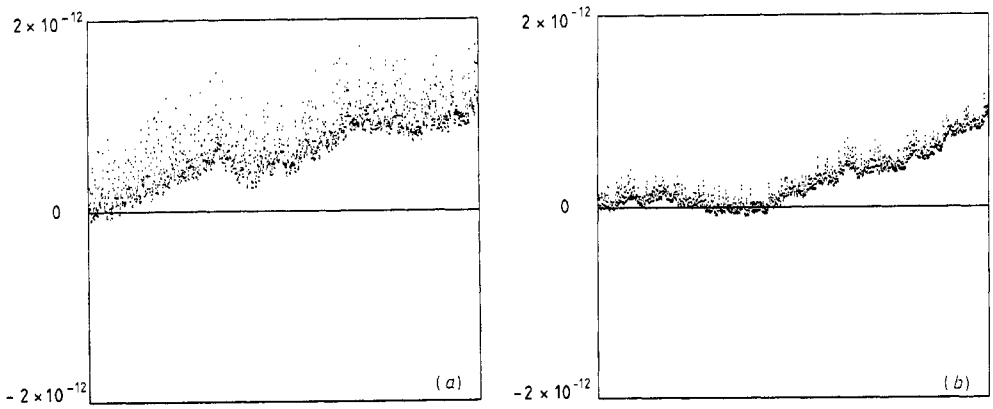

Figure 10. Comparison of energy errors using a fourth-order SIA (*a*) and a fourth-order RKI (*b*) as a function of time for the geodesic flow on a two-holed doughnut as in figure 9.

Let us note in passing an interesting aspect of this system. All Hamiltonian systems are locally integrable [28]. For this system the local integrals are obvious, namely, as long as the trajectory remains inside the fundamental polygon, both the energy and angular momentum are constant. When the trajectory crosses the boundary of the fundamental polygon, the energy is unchanged by the transformation (22)–(24), but

the angular momentum of almost all trajectories jumps a discrete amount. Thus, for this system, we know precisely the domain of definition of these local integrals.

4. Alternative techniques

The algorithms of section 3 can also be applied when the Hamiltonian is of potential form, that is, when it can be integrated using the algorithms of section 2. We have done this with the examples of section 2, generating a fifth-order integrator for the Fermi–Pasta–Ulam problem and a sixth-order integrator for the Hénon–Heiles problem. The stability properties were the same as those found using the algorithms of section 2. Therefore, it is a matter of choice as to which algorithm one should use; typically, the 'best' algorithm is highly problem dependent.

In our examples thus far, we have concentrated on problems in which the Hamiltonian is known analytically; as this is often the case, our concentration is not inappropriate. There are many cases, however, when either the Hamiltonian is not known analytically or the analytic evaluation is awkward or time consuming. Fortunately, for these cases, analogues of both the Runge–Kutta and Adams–Bashforth integrators exist.

Ruth [5] has developed the Runge–Kutta approach to symplectic integration for Hamiltonians of potential form in which n evaluations per timestep of the force generate an nth-order algorithm. We have not tested these algorithms, but we have seen numerical results of Laslett [29] demonstrating that these algorithms have the same remarkable stability properties that we have found with the analytic algorithms.

The Runge–Kutta type of symplectic algorithms are particularly useful when force evaluations are not time consuming. In many cases, however, force evaluations may be the most time-consuming part of the calculation; for these problems, the Runge–Kutta algorithms may not be very efficient. One answer to this problem when using standard integrators is to use algorithms of the Adams–Bashforth type, that is algorithms that store the results of n past force evaluations to give algorithms of nth order. Thus, these algorithms gain in speed if force evaluations are slow, because only one new force evaluation is needed per timestep, but they require more memory. It turns out that symplectic integration algorithms of the Adams–Bashforth type of any desired order can be produced based on any of the analytic algorithms we have discussed before. The basic idea is really very simple. One uses past force evaluations to evaluate the coefficients in a local expansion of the Hamiltonian. In other words, if $(\bar{q}, \bar{p})$ is a point near the present location in phase space, we assume that the Hamiltonian is of the form

$$H = \bar{H} + \sum_{i,j} c_{ij}(q - \bar{q})^i (p - \bar{p})^j$$

and we use past force evaluations to compute the c_{ij} to the desired order. Once the c_{ij} are known, the truncated Hamiltonian can be used in any of the analytic algorithms we have presented before. We have applied this procedure to a variety of one-degree-of-freedom problems to fairly high order and to the Fermi–Pasta–Ulam problem through second order. In all cases we obtained the same order and stability results we obtained using other symplectic algorithms. For the one-degree-of-freedom problems, these algorithms were also significantly faster than other symplectic integrators and did not

use much extra memory. However, as the number of degrees of freedom increased, the symplectic Adams–Bashforth integrators decreased rapidly in speed and increased dramatically in memory requirements. The reason for this behaviour is easy to see; if n is the number of degrees of freedom, the number of coefficients, c_{ij}, required to find a first-order timestep is $2n$, the number required to find a second-order timestep is $n(2n+1)$, and the number required to find the third-order timestep is $4n^3/3+2n^2+2n/3$. To obtain this information one must store one right-hand side evaluation for first order, $n+1$ right-hand side evaluations for second order, $(2n^2+1)/3+n$ force evaluations for third order, etc. Not only does the amount of memory required increase dramatically with order and number of degrees of freedom, but the size of the linear problem to be solved to determine the c_{ij} also increases rapidly, thus increasing the amount of computation required to make a time step. The reason that the memory and time requirements increase so quickly is that a symplectic integrator not only constrains the motion along the trajectory as in a standard Adams–Bashforth method, it also constrains all the trajectories in a plane transverse to the trajectory and requires information about the variation of the Hamiltonian in all directions to do this. In fact, it is this transverse constraint that, in some sense, probably accounts for the stability of symplectic algorithms, namely the trajectory has many neighboring trajectories that it cannot approach. The implication for the Adams–Bashforth type of symplectic integrators, however, is that they are likely to be of use only if the system has a small number of degrees of freedom or if only a low-order integrator is required.

In summary, a set of rules of thumb, with many exceptions, is that the analytic symplectic algorithms are better and should be used when they can be practically evaluated. The Runge–Kutta symplectic algorithms are better if the Hamiltonian is of potential form and if force evaluations are fast. The Adams–Bashforth symplectic algorithms may be useful either for a small number of degrees of freedom or for lower-order integration.

5. Stability results

5.1. Linear stability

When the Hamiltonian is of potential form with a quadratic potential, and thus the dynamics is linear, the technique of section 2 gives very nice results. Let $H(p,q) = p^2/2+\langle Lq,q\rangle$, where L can be chosen symmetric and where the angle brackets indicate scalar contraction. Then each term S_i of the generating function $S=\sum\epsilon^i S_i$ satisfies (3) or (4) and because the potential is quadratic, induction on i shows that S_i is quadratic in q and q_0. From the symmetry argument at the end of section 2, S_i is a symmetric function of q and q_0. The most general function of this form is

$$S_i = \tfrac{1}{2}\langle A_i q,q\rangle + \langle B_i q,q_0\rangle + \tfrac{1}{2}\langle A_i q_0,q_0\rangle$$

where $A_i = A_i(L)$ and $B_i = B_i(L)$ are symmetric. The truncated generating function is

$$S_N = \sum_{i=0}^{N}\epsilon^i S_i = \tfrac{1}{2}\langle A_N q,q\rangle + \langle B_N q,q_0\rangle + \tfrac{1}{2}\langle A_N q_0,q_0\rangle$$

where

$$A_N(L) = \sum_{i=0}^{N}\epsilon^i A_i(L)$$

and

$$B_N(L) = \sum_{i=0}^{N} \epsilon^i B_i(L).$$

We search for quadratic invariants of the map determined by

$$p = \frac{\partial S_N}{\partial q} = A_N q + B_N q_0 \tag{25}$$

and

$$p_0 = -\frac{\partial S_N}{\partial q_0} = -A_N q_0 - B_N q. \tag{26}$$

Let $\langle \mathcal{U} p, p\rangle + \langle \mathcal{V} p, q\rangle + \langle \mathcal{W} q, q\rangle$ be invariant under this map. Substituting (25) and (26) in the invariance condition and matching terms in qq, $q_0 q$, and $q_0 q_0$ gives

$$A\mathcal{U}A + \mathcal{V}A + \mathcal{W} = B\mathcal{U}B \tag{27}$$

$$A\mathcal{U}A - \mathcal{V}A + \mathcal{W} = B\mathcal{U}B \tag{28}$$

and

$$2B\mathcal{U}A + B\mathcal{V} = 2A\mathcal{U}B - \mathcal{V}B. \tag{29}$$

Equations (27) and (28) immediately imply that $\mathcal{V} = 0$ because A is close to a multiple of the identitity for small δt. The terms A and B commute because they are both functions of L; therefore, if we look for solutions of the form $\mathcal{U} = \mathcal{U}(L)$, then A, B, and $\mathcal{U}$ all commute and (29) is trivial. A solution is then $\mathcal{W} = (B^2 - A^2)\mathcal{U}$ where $\mathcal{U}$ is any function of L. Because every function of L can be constructed from the set of orthogonal projections P_i onto the ith eigenspace with corresponding eigenvalue λ_i of L, the set of general solutions is $\mathcal{U} = P_i$ and $\mathcal{W} = (B^2 - A^2)P_i$. Note that because

$$(B^2(L) - A^2(L))P_i = (B^2(\lambda_i) - A^2(\lambda_i))P_i$$

the i^{th} quadratic form

$$I_i = \langle P_i p_0, p_0\rangle + (B^2(\lambda_i) - A^2(\lambda_i))\langle P_i q_0, q_0\rangle$$

is exactly invariant. Similar results can be obtained for the generating function technique of section 3.

The existence of these invariants implies that the numerical phase space is foliated into n-dimensional tori in the same way as in the exact system. In other words, trajectories of the numerical system not only stay near an energy surface for all time, they also remain for all time on the integral surfaces determined by these invariants. Though invariant tori of linear systems are usually unstable to perturbations, the above result shows that they are stable under the perturbations generated by this class of symplectic integrators.

Not only are these surfaces invariant under the numerical flow, we will now show that they are near the invariants of the exact system. Observe that the functions $B(L)$

and $A(L)$ are functions of δt. This δt is held fixed in determining a timestep map. Inserting S_N into the Hamilton–Jacobi equation and evaluating at a fixed small time δt shows that

$$\tfrac{1}{2}\dot{A}_N + \tfrac{1}{2}A_N^2 + L = \mathcal{O}(\delta t^{2N}) \tag{30}$$

$$\dot{B}_N + A_N B_N = \mathcal{O}(\delta t^{2N}) \tag{31}$$

and

$$\tfrac{1}{2}\dot{A}_N + \tfrac{1}{2}B_N^2 = \mathcal{O}(\delta t^{2N}). \tag{32}$$

Subtracting (32) from (30) gives

$$B_N^2(L) - A_N^2(L) = 2L + \mathcal{O}(\delta t^{2N})$$

so that the invariants

$$I_i = \langle P_i p_0, p_0\rangle + (B_N^2(\lambda_i) - A_N^2(\lambda_i))\langle P_i q_0, q_0\rangle = \langle P_i p_0, p_0\rangle + 2\lambda_i\langle P_i q_0, q_0\rangle + \mathcal{O}(\delta t^{2N}).$$

The first two terms on the right-hand side are the invariants of the exact system. That is, the numerical invariants are at least asymptotic (in δt) to the invariants of the analytical system.

The ansatz that the full generating function be quadratic,

$$S = \tfrac{1}{2}\langle Aq, q\rangle + \langle Bq, q_0\rangle + \tfrac{1}{2}\langle Aq_0, q_0\rangle$$

gives an analytic solution to the full Hamilton–Jacobi equation in the form

$$A = \sqrt{2L}\cot\sqrt{2L}t$$

and

$$B = -\frac{\sqrt{2L}}{\sin\sqrt{2L}\,t}$$

provided L has a real square root. This is true when L has no eigenvalues on the negative real axis; the root can then be found by computing a contour integral in the complex plane, symmetric about the real axis using the functional calculus of self-adjoint operators [30]. Indeed, when there are no eigenvalues on the negative real axis, this indicates that the series, $\sum \delta t^i A_i$, $\sum \delta t^i B_i$, converge to A and B, respectively. If L does have eigenvalues on the negative real axis, these series may only be asymptotic. In any case, whether or not these invariants, I_i, are asymptotic or convergent to the invariants of the analytic system, they are still strictly invariant so that the numerical system fully reflects the structure of the exact system.

5.2. Nonlinear stability

The preservation of phase-space substructures in the linear case is one example of the global stability properties to be expected and which we have found from symplectic integrators in general. If a Hamiltonian is nonlinear but nearly integrable, the Kolmogorov–Arnold–Moser (KAM) theorem [27, 31] asserts that any sufficiently small symplectic perturbation (such as that determined by a symplectic numerical algorithm) of the Hamiltonian dynamics preserves a great deal of the global structure; namely, most of the incommensurate, non-degenerate, invariant tori that are remnants of the integrable system are preserved under the additional small perturbation that is due to the integrator for small enough timestep. The example of section 2.2 seems to confirm this. In addition, unstable periodic orbits [1] and lower dimensional tori [32] and their stable and unstable manifolds [33] are structurally stable and will be well represented by SIAs.

If the Hamiltonian system is an Anosov system [27], is such that almost all sufficiently nearby points diverge exponentially (an example was given in section 3B), then the Anosov theorem implies that the trajectories of the time-δt map of the Hamiltonian system are topologically identical to those of the symplectic integrator, or for that matter of any integrator of sufficient accuracy. For these systems, as we pointed out, all integrators quickly lose their accuracy, that is their tracking of the particular initial condition in the exact system, but they capture exactly the topological structure of the trajectories in phase space.

We see, then, that SIAs replicate the structure of large parts of the exact system if the system is linear, if it is nonlinear but nearly integrable, or if it is nonlinear and completely chaotic. In the general case, though no rigorous results are known, it is believed that the phase space of a Hamiltonian system decomposes into primary invariant tori (remnants of tori from a possibly nearby integrable system), secondary invariant tori (which arise from various resonances of the perturbation and their interactions), and 'chaotic' regions that are thought to consist of components that are almost Anosov systems. Not only do the linear, KAM, or Anosov stability theorems apply to many of these structures, but also many of them, KAM tori and stable and unstable manifolds, are Lagrangian submanifolds [34], which means that the symplectic 2-form (the collection of Poincaré invariants) vanishes on them. Because SIAs preserve exactly the symplectic 2-form, their structure should be well represented by symplectic integrators. Though naively one might think that SIAs and standard integrators should be equally good in the chaotic regions, SIAs are also superior there because they avoid the boundaries and the embedded islands of secondary tori, which are Lagrangian submanifolds that the integrator will not 'cross' (strictly speaking these submanifolds separate the phase space only for two degrees of freedom). Thus, using a symplectic integrator insures that the region within which the motion is chaotic is more accurately and sharply defined, maintaining more of the global structure caused by stable invariant objects.

6. Summary

A variety of symplectic integration algorithms of any desired order now exist for both the class of potential Hamiltonians and for general Hamiltonians. The algorithms developed by us and by others include analogues of Padé approximant, Runge–Kutta, Adams–Bashforth, and analytic series algorithms with their relative ease of use,

efficiency, and speed varying from one problem to another. More work is needed, principally to investigate techniques for optimising existing algorithms and to develop a standard package that is easy to use and does not require the extensive customisation that each problem necessitates. We have presented numerical evidence from a number of examples to show that SIAs are extremely stable. We have argued that SIAs replicate the structure of large parts of the exact system when it is linear, nonlinear but nearly integrable, or if it is nonlinear and completely chaotic. There is thus a great deal of evidence that symplectic integration algorithms are excellent tools for the investigation of the global geometry of phase space, reproducing most of the phase-space structures of interest.

Appendix A. Computational strategies

In some cases, the algorithms derived from equations (7)–(9) or (16)–(21) can be written down and programmed by hand. For many systems, however, involving either a high-order integrator or a Hamiltonian with a large number of degrees-of-freedom, we found it useful to do the computations and FORTRAN programming using the computer algebra program SMP (a product of Inference Corporation, Inc., of Los Angeles, CA). Similar results have been obtained using less memory with the computer algebra program Mathematica (a product of Wolfram Research Inc., of Urbana, IL). In one case, using SMP, a 4900 line FORTRAN program was generated and compiled in less than two hours real time. A listing of the program that implements (16)–(21) is shown as appendix B. Though this SMP program should work in principle for any analytically given Hamiltonian, in practice, it is often necessary to change many of the details of the calculation. Changes are necessary as to when expansions are done and how the FORTRAN programming is done, to stay within computer memory limits. Occasionally, it was necessary to change the details of the calculation or the approach because the computer time required was excessive; this was especially true when the Hamiltonian described a system with a large number of degrees of freedom. In fact, in most of the examples we have encountered, the type and order of algorithm we used was dictated by the constraints of computer memory and time. We have not found any general rule that indicates what approach will result in the best combination of algorithm generation ease, reasonable compile time, and speed of program execution.

Once the FORTRAN statements were generated using SMP, an editor was used to add subroutine header lines, RETURN and END statements, and to merge the subroutines with a driver program. The driver program initialised the variables, obtained run information, and computed the requested number of time steps.

In most cases a simple iteration was used to solve the implicit equation, either (1) or (12), in the manner described in section 2 (in rare cases these equations are not implicit). Simple iteration is not the only technique, of course, for solving the implicit equation. A Newton method could also be used to speed up the calculation, though we have not tried this. Instead, we have tested the idea of using a better first guess for the straightforward iteration technique. Instead of inserting the old coordinates into the non-identity piece of the equation, we compute the new coordinates using a standard (non-symplectic) Runge–Kutta integrator and use these to initialise the iteration. The results depended strongly on the particular problem and on the timestep. With a low-order algorithm or a small timestep, this improved initialisation resulted only in a small improvement in computation time, about 10–20%. In other cases, the speed-up

was significant, namely factors of 2–10 increase in computation speed were found. In any case, it is clear that most of the implementations of the algorithms we have used are far from optimised and that a great deal of work remains to be done in this area.

Appendix B. Symbolic code generation

In this appendix we give the listings of two SMP files that implement the algorithm derived from (16)–(21). The first instruction in the first file is to read the second file (presented at the end of this appendix) to find out the number of degrees-of-freedom and the specific Hamiltonian

```
<hamin
```

The expressions for the first two orders of the generating function are trivial and are implemented first

```
k0:-Sum[po[i]*qn[i],{i,1,n}];
k1:h;
```

We then evaluate the derivatives of `k1`.

```
i:1;
Loop[i<=n,dk1[i]:D[k1,qn[i]];Inc[i]];
```

We next evaluate all the momentum derivatives of the Hamiltonian that occur in the equations.

```
i:1;
Loop[i<=n,dhp[i]:D[h,po[i]];Inc[i]];
i:1;
Loop[i<=n,j:1;Loop[j<=n,dhpp[i,j]:D[dhp[i],po[j]];Inc[j]];Inc[i]];
i:1;
Loop[i<=n,j:1;Loop[j<=n,k:1;Loop[k<=n,dhppp[i,j,k]:D[dhpp[i,j],po[k]];\
Inc[k]];Inc[j]];Inc[i]];
i:1;
Loop[i<=n,j:1;Loop[j<=n,k:1;Loop[k<=n,l:1;Loop[l<=n,\
dhpppp[i,j,k,l]:D[dhppp[i,j,k],po[l]];\
Inc[l]];Inc[k]];Inc[j]];Inc[i]];
i:1;
Loop[i<=n,j:1;Loop[j<=n,k:1;Loop[k<=n,l:1;Loop[l<=n,\
m:1;Loop[m<=n,dhppppp[i,j,k,l,m]:D[dhpppp[i,j,k,l],po[m]];\
Inc[m]];Inc[l]];Inc[k]];Inc[j]];Inc[i]];
```

We now begin the evaluation of the higher-order generating functions, beginning with `k2`.

```
k2:Sum[-dk1[i]*dhp[i],{i,1,n}];
```

Given `k2`, we can now compute the terms necessary to find `k3`.

```
i:1;
Loop[i<=n,dk2[i]:D[k2,qn[i]];Inc[i]];
```

We now compute k3.

```
k3:Sum[-dk2[i]*dhp[i],{i,1,n}]+Sum[Sum[dk1[i]*dk1[j]*dhpp[i,j],\
{j,1,n}],{i,1,n}];
```

Given k3, we can now compute the terms necessary to find k4.

```
i:1;
Loop[i<=n,dk3[i]:D[k3,qn[i]];Inc[i]];
```

We now compute k4

```
k4:Sum[-dk3[i]*dhp[i],{i,1,n}]+Sum[Sum[3*dk2[i]*dk1[j]*dhpp[i,j],\
{j,1,n}],{i,1,n}]+Sum[Sum[Sum[-dk1[i]*dk1[j]*dk1[k]*dhppp[i,j,k],\
{k,1,n}],{j,1,n}],{i,1,n}];
```

Given k4, we can now compute the terms necessary to find k5

```
i:1
Loop[i<=n,dk4[i]:D[k4,qn[i]];Inc[i]];
```

We now compute k5

```
k5:Sum[-dk4[i]*dhp[i],{i,1,n}]+Sum[Sum[3*dk2[i]*dk2[j]*dhpp[i,j],\
{j,1,n}],{i,1,n}]+Sum[Sum[Sum[-6*dk1[i]*dk1[j]*dk2[k]*dhppp[i,j,k],\
{k,1,n}],{j,1,n}],{i,1,n}]+Sum[Sum[4*dk1[i]*dk3[j]*dhpp[i,j],\
{j,1,n}],{i,1,n}]+Sum[Sum[Sum[Sum[dk1[i]*dk1[j]*dk1[k]\
*dk1[l]*dhpppp[i,j,k,l],{l,1,n}],\
{k,1,n}],{j,1,n}],{i,1,n}];
```

Given k5, we can now compute the terms necessary to find k6

```
i:1;
Loop[i<=n,dk5[i]:D[k5,qn[i]];Inc[i]];
```

We now compute k6

```
k6:Sum[-dk5[i]*dhp[i],{i,1,n}]+Sum[Sum[Sum[Sum[Sum[-dk1[i]*dk1[j]*\
dk1[k]*dk1[l]*dk1[m]*dhppppp[i,j,k,l,m],{m,1,n}],{l,1,n}],{k,1,n}],\
{j,1,n}],{i,1,n}]+Sum[Sum[Sum[Sum[10*dk1[i]*dk1[j]*dk1[k]*dk2[l]*\
dhpppp[i,j,k,l],{l,1,n}],{k,1,n}],{j,1,n}],{i,1,n}]+Sum[Sum[\
5*dk1[i]*dk4[j]*dhpp[i,j],{j,1,n}],{i,1,n}]+Sum[Sum[10*dk2[i]*\
dk3[j]*dhpp[i,j],{j,1,n}],{i,1,n}]+Sum[Sum[Sum[-10*dk1[i]*\
dk1[j]*dk3[k]*dhppp[i,j,k],{k,1,n}],{j,1,n}],{i,1,n}]+\
Sum[Sum[Sum[-15*dk1[i]*dk2[j]*dk2[k]*dhppp[i,j,k],{k,1,n}],\
{j,1,n}],{i,1,n}];
```

Given all these, we now compute the part of k we will have to iterate, i.e., the non-identity piece of k; we call it kp. We also compute the total k, kt, by adding back the identity. Note that del is the size of the time step

```
kp:del*k1+del^2*k2/2+del^3*k3/6+del^4*k4/24\
  +del^5*k5/120+del^6*k6/720;
kt:k0+kp;
```

We can now compute the new momentum and the iterative piece of the next q

```
i:1;
Loop[i<=n,pn[i]:Ex[-D[kt,qn[i]]];Inc[i]];
i:1;
Loop[i<=n,qt[i]:Ex[D[kp,po[i]]];Inc[i]];
```

We now break these expressions into pieces small enough to program in FORTRAN without exceeding memory limits; whether this is necessary or not depends on the particular problem. We first compute the lengths of the various expressions and compute how many blocks of 10 terms are in each, and how many terms are in the final, incomplete block

```
i:1;
Loop[i<=n,plen[i]:Len[pn[i]];rp[i]:Floor[plen[i]/10];rem[i]:\
Mod[plen[i],10];Inc[i]];
i:1;
Loop[i<=n,qlen[i]:Len[qt[i]];rq[i]:Floor[qlen[i]/10];req[i]:\
Mod[qlen[i],10];Inc[i]];
```

Given the relevant lengths, we now define the appropriate sums of the pieces of the expressions

```
i:1;
Loop[i<=n,j:1;Loop[j<=rp[i],psn[i,j]:Sum[pn[i,k],{k,10*j-9,10*j}]\
;Inc[j]];\
If[rem[i]~~=0,psn[i,rp[i]+1]:Sum[pn[i,k],\
{k,10*rp[i]+1,10*rp[i]+rem[i]}],,];\
Inc[i]];
i:1;
Loop[i<=n,j:1;Loop[j<=rq[i],qst[i,j]:Sum[qt[i,k],{k,10*j-9,10*j}];\
Inc[j]];\
If[req[i]~~=0,qst[i,rq[i]+1]:Sum[qt[i,k],\
{k,10*rq[i]+1,10*rq[i]+req[i]}],,];\
Inc[i]];
```

We now have expressions small enough to program within memory restrictions

```
Prog[qst,hsix,{{qn,po},,,,,,1,},2];
Prog[psn,hsix,{{qn,po},,,,,,1,},2];
```

The file `hsix` now needs to be worked on a little bit with an editor, principally to define the resummations of the pieces of `qt` and`pn`. We also have to add subroutine headers and `return` and `end` statements. The subroutines are then incorporated in a driver program.

We also include, as an example, the listing of `hamin`, an SMP file that defines the Hénon–Heiles Hamiltonian for use in a symplectic integrator. We first specify that there are two degrees of freedom

```
n:2;
```

We now specify the Hamiltonian. Because the Hamiltonian will be used in a generating function, we label the momentum `po` for 'old' and the coordinates `qn` for 'new'

```
h: (po[1]^2+po[2]^2+qn[1]^2+qn[2]^2)/2+qn[1]^2*qn[2]-qn[2]^3/3;
```

These two files have been run in SMP and produced subroutines for an integrator that was confirmed to be sixth order with the expected long time stability.

Acknowledgments

The authors wish to express their appreciation to A Weinstein and F Molzahn for pointing out corrections to a preliminary draft of this paper, to C Menyuk for pointing out [11], to J Marsden for discussions on the role of symmetry, and to G Mayer-Kress for assistance with the graphics.

References

[1] Abraham R and Marsden J E 1978 *Foundations of Mechanics* 2nd edn (Reading, MA: Benjamin-Cummings)
[2] Arnold V I 1974 *Mathematical Methods of Classical Mechanics* (Moscow: Nauka) (1978 Engl. transl. (New York: Springer) p 225)
[3] Landau L D and Lifshitz E M 1960 *Mechanics* (New York: Pergamon)
[4] De Vogelaere 1956 Methods of integration which preserve the contact transformation property of the Hamiltonian equations *Report 4* Department of Mathematics, University of Notre Dame
[5] Ruth R 1983 A canonical integration technique *IEEE Trans. Nucl. Sci.* **30** 2669
[6] Channell P J 1983 Symplectic integration algorithms *Internal report AT-6:ATN-83-9* Los Alamos National Laboratory
[7] Kang F 1986 Difference schemes for Hamiltonian formalism and symplectic geometry *J. Comput. Math.* **4** 279
[8] Neri F 1987 Lie algebras and canonical integration *Technical Report* University of Maryland Department of Physics
[9] Channell P J 1983 Initial value and eigenvalue problems for field equations using symplectic integration algorithms *Internal report AT-6:ATN-83-18* Los Alamos National Laboratory
[10] Ruth R and Forest E in preparation
[11] Menyuk C R 1984 Some properties of the discrete Hamiltonian method *Physica* **11D** 109
[12] Goldstein H 1950 *Classical Mechanics* (Reading, MA: Addison-Wesley) p 240
[13] Molzahn F H and Osborn T A 1958 Tree graphs and the solution to the Hamilton–Jacobi equation *J. Math. Phys.* **27** 88
[14] Fermi E, Pasta J R and Ulam S 1955 Studies of nonlinear problems, part I *Report LA-1940* Los Alamos National Laboratory
[15] Gardner C S, Greene J M, Kruskal M D and Miura R M 1968 Method for solving the Korteveg–deVries equation *Phys. Rev. Lett.* **19** 1095
[16] Hénon M and Heiles C 1964 The applicability of the third integral of motion: some numerical experiments *Astron. J.* **69** 73
[17] Gustavson F G 1966 On constructing formal integrals of a Hamiltonian system near an equilibrium point *Astron. J.* **71** 670
[18] Birkhoff G D 1950 Dynamical systems *Colloq. Publ. IX* 2nd edn (Providence, RI: American Mathematical Society)
[19] Fomenko A T 1987 The topology of surfaces of constant energy in integrable Hamiltonian systems and obstructions to integrability *Math. USSR Izv.* **20** No 3
[20] Cornfeld I P, Fomin S V and Sinai Ya G *Ergodic Theory* (New York: Springer)
[21] Aref H and Pomphrey N 1982 Integrable and chaotic motions of four vortices I. The case of identical vortices *Proc. R. Soc.* A **380** 359
[22] Marsden J E 1988 Lie Poisson Hamilton Jacobi theory and Lie Poisson integrators *Phys. Lett.* A submitted
[23] Dubrovin B A, Fomenko A T and Novikov S P 1979 *Modern Geometry—Methods and Applications* vol I (Moscow: Nauka) (1984 Engl. transl. (New York: Springer) p 93)

[24] Dubrovin B A, Fomenko A T and Novikov S P 1979 *Modern Geometry—Methods and Applications* vol II (Moscow: Nauka) (1985 Engl. transl. (New York: Springer) p 166)
[25] Weiss B and Ornstein D 1973 Geodesic flows are Bernoulli *Israel J. Math.* **14** 184
[26] Anosov D V 1962 Roughness of geodesic flows on compact Riemannian manifolds of negative curvature *Sov. Math. Dokl.* **3** 1068
[27] Arnold V I and Avez A 1968 *Ergodic Problems of Classical Mechanics* (New York: Benjamin)
[28] Duff G F D 1956 *Partial Differential Equations* (Toronto: University of Toronto Press) p 66
[29] Laslett L J private communication
[30] Kato T 1976 *Perturbation Theory for Linear Operators* 2nd edn (New York: Springer)
[31] Moser J 1973 Stable and random motions in dynamical systems *Ann. Math. Stud.* No 77
[32] Graff S M 1974 On the conservation of hyperbolic invariant tori for Hamiltonian systems *J. Diff. Eq.* **15** 1
[33] Hirsch M W, Pugh C and Shub M 1977 Invariant manifolds *Lecture Notes in Mathematics* **583** (Berlin: Springer)
[34] Weinstein A 1971 Symplectic manifolds and their Lagrangian submanifolds *Adv. Math.* **6** 1040