Preservation of Phase Space Structure in Symplectic Integration

A thesis presented in partial
fulfilment of the requirements
for the degree of
Doctor of Philosophy
in Mathematics

at Massey University,
Palmerston North, New Zealand.

Dion Robert James O'Neale
2009

Abstract

This thesis concerns the study of geometric numerical integrators and how they preserve phase space structures of Hamiltonian ordinary differential equations.

We examine the invariant sets of differential equations and investigate which numerical integrators preserve these sets, and under what conditions. We prove that when periodic orbits of Hamiltonian differential equations are discretized by a symplectic integrator they are preserved in the numerical solution when the integrator step size is not resonant with the frequency of the periodic orbit.

The preservation of periodic orbits is the result of a more general theorem which proves preservation of lower dimensional invariant tori from dimension zero (fixed points) up to full dimension (the same as the number of degrees of freedom for the differential equation). The proof involves first embedding the numerical trajectory in a non-autonomous flow and then applying a KAM type theorem for flows to achieve the result. This avoids having to prove a KAM type theorem directly for the symplectic map which is generally difficult to do.

We also numerically investigate the break up of periodic orbits when the integrator's step size is resonant with the frequency of the orbit.

We study the performance of trigonometric integrators applied to highly oscillatory Hamiltonian differential equations with constant frequency. We show that such integrators may not be as practical as was first thought since they suffer from higher order resonances and can perform poorly at preserving various properties of the differential equation. We show that, despite not being intended for such systems, the midpoint rule performs no worse than many of the trigonometric integrators, and indeed, better than some.

Lastly, we present a numerical study of a Hamiltonian system consisting of two magnetic moments in an applied magnetic field. We investigate the effect of both the choice of integrator and the choice of coordinate system on the numerical solutions of the system. We show that by a good choice of integrator (in this case the generalised leapfrog method) one can preserve phase space structures of the system without having to resort to a change of coordinates that introduce a coordinate singularity.

Acknowledgements

Thanks first to my supervisor Robert McLachlan for guidance along the course of this thesis. Thanks also to other members of the mathematics department at I.F.S., including my co-supervisor Matt Perlmutter and fellow doctoral candidates Brett Ryland and Philip Zhang, for useful discussions, mathematical and otherwise.

During this thesis, financial support was received from the New Zealand Institute of Mathematics and its Applications, the New Zealand Mathematics Society, The Royal Society of New Zealand, Education New Zealand, and the IFS Postgraduate Research Fund. This was much appreciated and very helpful.

Much of this funding was used for travel to visit Zai-jiu Shang at the Chinese Academy of Sciences in Beijing, and to attend the program on highly oscillatory problems at the Isaac Newton Institute, in Cambridge, hosted by Arieh Iserles. The hospitality of both these people was greatly appreciated, as were the discussions.

Discussions with Ernst Hairer, Reinout Quispel and David Cohen during the INI program also contributed to work in this thesis.

Finally, thanks to family, friends and colleagues, for help and support over the duration of this thesis.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

> Ours, according to Leibniz, is the best of all possible worlds, and the laws of nature can therefore be described in terms of extremal principles. Thus, arising from corresponding variational problems, the differential equations of mechanics have invariance properties relative to certain groups of coordinate transformations.
>
> — C. L. Siegel and J. K. Moser, 1971

Ordinary differential equations are used to describe and model physical systems. The solution of such equations gives insight into how the systems evolve. Although ours may be the best of all possible worlds, it is, in general, extremely difficult, if not impossible, to find analytic solutions of differential equations. Consequently, numerical methods which are able to provide approximate solutions are a crucial tool. Encapsulated in the differential equations which describe it are *all* the properties of a dynamical system — information about invariant quantities, stability and much more. The best numerical methods therefore give solutions which are not just points or sets of points close to an analytic solution, but which capture as many properties as possible of the differential equation. This idea is the basis of geometric numerical integration. In section 1.2 a brief introduction is given to traditional methods of numerical integration for differential equations. This is followed in section 1.3 by a description of Hamiltonian systems of differential equations — the main class of differential equation which we will focus on in this thesis. A discussion of some well known geometric integrators and some of the ideas behind them is presented in section 1.4. The chapter closes with an outline of the rest of this thesis and a list of original contributions.

1.1 Framework

Assume we are given an initial value problem consisting of an autonomous differential equation and an initial condition at some point t_0,

$$\frac{\mathrm{d}y}{\mathrm{d}t} = \dot{y} = f(y(t)), \quad y(t_0) = y_0. \tag{1.1}$$

If the differential equation is non-autonomous, (i.e., it has explicit time-dependence) it can be easily represented in autonomous form by adding the trivial equation

$$\dot{t} = 1 \tag{1.2}$$

to the system. A reasonable requirement of any numerical method is that it can exactly solve equations of the form (1.2). Apart from increasing the dimension of the system by one, a consequence of representing a non-autonomous system in autonomous form is that a system which was originally decoupled becomes coupled upon introduction of the differential equation (1.2) for t. This can be important, for example, for splitting methods (cf. section 1.2.5) where one may wish to exploit the decoupled form of the original equations.

Generally, in (1.1), the solution $y(t)$ of the differential equation is a vector valued function:

$$y : \mathbb{R} \to \mathbb{R}^n, \quad t \mapsto y(t).$$

As is the vector field:

$$f : \mathbb{R}^n \to \mathbb{R}^n, \quad y \mapsto f(y),$$

where n is the dimension of the system. Given such an initial value problem, one may want to know the state of the system $y(t)$ at some time in the future, $t > t_0$. It is important to consider whether such a solution does, in fact, exist and whether it is unique. Both questions can be answered by checking whether the function f satisfies a Lipschitz condition.

Definition 1.1. A function $f : \mathbb{R}^n \to \mathbb{R}^n$ satisfies a Lipschitz condition if there exists a Lipschitz constant $L < \infty$ and a norm $\|\cdot\|$ such that for any $u, v \in \mathbb{R}^n$, $\|f(u) - f(v)\| \leq L\|u - v\|$.

The following theorem then ensures existence and uniqueness of solutions. The proof of the theorem can be found in the book by Hairer, Nørsett and Wanner [56, section I.7], or in any introductory textbook on differential equations.

Theorem 1.1. *Given an initial value problem (1.1), which has a continuous vector field f and which satisfies a Lipschitz condition, there exists $\varepsilon > 0$ and a unique solution to the problem on $|t - t_0| < \varepsilon$.*

If $y(t)$ is the solution to some differential equation $\dot{y} = f(y)$, one can (at least locally, i.e. for $|t - t_0|$ less than some bound) define the exact flow $\varphi_{t,f}$ of the equation as,

$$\varphi_{t,f}(y_0) = y(t), \quad y_0 = y(t_0).$$

For any point y_0 in the phase space the flow associates a value of the solution $y(t)$ with the initial value y_0. Fixing the time step h turns the flow into a map from the phase space (say $\mathbb{R}^n$) to itself; $\varphi_{h,f} : \mathbb{R}^n \to \mathbb{R}^n$. Since numerical integrators also define a map (though not necessarily from the phase space to itself; cf. linear multistep methods, mentioned in section 1.2.1), the exact flow map is a natural object with which to work.

1.2 Background & history of numerical integration

One of the first, and simplest, numerical integrators is due to Euler [38]. Given the solution at some point y_{n-1}, the explicit Euler method follows the tangent at that point, $f(y_{n-1})$, a distance $h = t_n - t_{n-1}$ to estimate the solution at a new point y_n. The method can be written

$$y_n = y_{n-1} + hf(y_{n-1}).$$

The difference between the exact solution $y(t_{n-1} + h) = \varphi_h(y_{n-1})$ and the point y_n calculated by the Euler method after one step is known as the *local truncation error*. The Taylor expansion of the exact solution, given by

$$y(t_{n-1} + h) = y_{n-1} + hf(y_{n-1}) + \frac{h^2}{2}\frac{\mathrm{d}^2 y}{\mathrm{d}t^2}(t_{n-1}) + \ldots,$$

agrees with the Euler method up to the first power of h, hence the method is said to be first order accurate.

The traditional approach to numerical integration has been to try to increase the accuracy of methods by minimising the local truncation error and increasing the order of the method, hopefully without increasing the computational cost disproportionately. Two main approaches have been used towards this end. The first approach increases accuracy by including more information from past steps in the integrator. Such methods are known as multistep (or, sometimes, multivalue) methods. The second approach increases accuracy by using approximations at internal points within each integration step — multistage methods.

1.2.1 Linear multistep methods

The 1883 paper by Adams and Bashforth [9] is the earliest example of the development of multistep methods. Linear multistep methods take the general form

$$y_n = \alpha_1 y_{n-1} + \alpha_2 y_{n-2} + \cdots + \alpha_k y_{n-k} + h\left(\beta_0 f(y_n) + \beta_1 f(y_{n-1}) + \cdots + \beta_k f(y_{n-k})\right).$$

Today the class of linear multistep methods with $\alpha_1 = 1$, $\alpha_2 = \cdots = \alpha_k = 0$ are known as Adams-Bashforth methods. Implicit linear multistep methods, i.e. with $\beta_0 \neq 0$ are known as Adams-Moulton methods due to Moulton's 1926 paper [102], though such methods were also considered in [9]. Other notable early contributions to linear multistep methods came in 1925 from Nyström [108] and in 1926 from Milne [94]. It was also Milne who suggested in 1949 [95] that linear multistep methods should be implemented as predictor-corrector pairs using an explicit method to calculate a predicted value for y_n and an implicit method to correct this value. These methods are often described as $\mathrm{P(EC)}^N\mathrm{E}$ or $\mathrm{P(EC)}^N$ methods; the index N refers to iterating the evaluation and correction of y_n, N times or, perhaps, until convergence. From a geometric viewpoint, an immediate disadvantage of linear multistep methods is that they do not preserve the phase space of the differential equation, swapping it for a product space of k copies of itself where k is the number of different times involved in one step of the method. Although this drawback does not preclude linear multistep methods from being used as geometric integrators, *per se*, we will not pay much attention to them in this thesis.

1.2.2 Runge-Kutta methods

The credit for the first multistage numerical integrators goes to Runge for his 1895 paper *Über die numerische Auflösung von Differentialgleichunen* [119]. Further early methods were developed by Heun [63] in 1900. In 1901 Kutta [78] completely characterised all methods of order four and proposed some order five methods.

For real numbers b_i and a_{ij}, an s stage Runge-Kutta method takes the following form:

$$\begin{aligned} Y_i &= \sum_{j=1}^{s} a_{ij} h f(Y_j) + y_{n-1}, \quad i = 1, 2, \ldots, s, \\ y_n &= \sum_{i=1}^{s} b_i h f(Y_i) + y_{n-1}. \end{aligned} \tag{1.3}$$

The Y_i are known as the *stage values* and are approximations to the solution at points in the interval $[t_{n-1}, t_{n-1} + h]$. Such methods are usually represented by a tableau of coefficients

$$\begin{array}{c|c} A \\ \hline b^T \end{array}, \tag{1.4}$$

where b is a vector of the coefficients b_i and A is a matrix with entries a_{ij}. The complete definition of a Runge-Kutta method involves an additional set of coefficients c_i which specify the position in time at which the stage values $Y_i := Y_i(t_{n-1} + c_i h)$ are evaluated. Since we are interested in autonomous differential equations, (and since non-autonomous differential equations can be put into autonomous form as described in section 1.1), we omit the c_i. The tableau (1.4), which usually also includes a column vector c with entries c_i, is known as a Butcher tableau after the 1967 article by Butcher [18]. If the coefficient

matrix A has entries on, or above, the main diagonal, then the method will be implicit, requiring the solution of a nonlinear system of equations at each step. Papers by Gill [43] and Merson [93], in 1951 and 1957 respectively, pointed the way to a full analysis of order conditions for Runge-Kutta integrators by considering the Taylor series expansions of solutions and expressing the exact solution in elementary differentials. However, it was not until the work of Butcher [18] in 1967, associating elementary differentials with rooted trees, that a complete theory of order was available for multistage methods.

Rooted trees and elementary differentials

For the differential equation (1.1) the Taylor coefficients of the exact solution are obtained by repeated differentiation with respect to t. Using the chain-rule, the second order term in the expansion is

$$\ddot{y} = \frac{\mathrm{d}}{\mathrm{d}t} f(y) = f'(y)\dot{y} = f'f, \tag{1.5}$$

where we have dropped the arguments in the last term, and in what follows. In component-by-component form (1.1) is

$$\dot{y}^i = f^i,$$

and hence (1.5) takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\dot{y}^i = \frac{\mathrm{d}}{\mathrm{d}t} f^i = f^i_j f^j,$$

where we have used the Einstein summation notation. It is at the third order term in the Taylor expansion that the distinction between elementary and total differentials becomes apparent. Since y is present in both factors of $f'f$ we get two terms for the third derivative of y:

$$\frac{\mathrm{d}^3}{\mathrm{d}t^3} y = f''(f, f) + f'(f'f),$$

or, in tensor notation,

$$\frac{\mathrm{d}^3}{\mathrm{d}t^3} y^i = f^i_{jk} f^j f^k + f^i_j f^j_k f^k.$$

At each order in the Taylor expansion, the derivative which defines the Taylor coefficient is given by a weighted sum of *elementary differentials*. The complexity of these increases with subsequent derivatives.

Each elementary differential can be represented by a *rooted tree*; a graph with a specified vertex — the root — and with no cycles. Following the convention of the texts [20, 56], (and the convention of trees), the root of a rooted tree is placed at the bottom of a graph and all vertices grow upwards from the root. The number of vertices of a rooted tree is its *order*. For a tree τ the order of the tree is denoted by $r(\tau)$. We will denote by $\mathcal{T}$ the set of all rooted trees, including the empty tree $\emptyset$. Each rooted tree $\tau \in \mathcal{T}$ is generated recursively by the following definition.

Definition 1.2 (Rooted trees)**.** The set of (unordered) rooted trees $\tau \in \mathcal{T}$ is recursively defined by

$$\tau = \begin{cases} \bullet & \text{if } r(\tau) = 1 \\ [\tau_1, \tau_2, \ldots, \tau_k] & \text{if } r(\tau) > 1, \tau_i \in \mathcal{T}, i = 1, 2, \ldots, k. \end{cases}$$

The notation $[\tau_1, \ldots, \tau_k]$ refers to a rooted tree constructed by grafting the roots of $\tau_1, \ldots, \tau_k$ onto a new root. The trees $\tau_1, \ldots, \tau_k$ need not be distinct and their order is not important — that is, any two branches of τ which share a vertex can be interchanged.

Definition 1.3 (Equivalence class of rooted trees)**.** Two rooted trees of the same order are said to be equivalent if the corresponding elementary differentials are the same up to an exchange of summation indices. An equivalence class of rooted trees consists of all equivalent rooted trees of a particular order.

For each rooted tree $\tau \in \mathcal{T}$ the associated elementary differential $\mathcal{F}_f(\tau)$ of the vector field $f(y)$ is given by definition 1.4.

Definition 1.4 (Elementary differentials [20])**.** Given a tree τ and a function $f : \mathbb{R}^n \to \mathbb{R}^n$, analytic in a neighbourhood of y, the *elementary differential* $\mathcal{F}_f(\tau)(y)$ is defined by

$$\mathcal{F}_f(\bullet)(y) = f(y), \tag{1.6}$$

$$\mathcal{F}_f\left([\tau_1, \ldots, \tau_k]\right)(y) = f^{(k)}(y)\left(\mathcal{F}_f(\tau_1)(y), \mathcal{F}_f(\tau_2)(y), \ldots, \mathcal{F}_f(\tau_k)(y)\right). \tag{1.7}$$

In tensor notation the differential (1.7) is written as

$$\mathcal{F}_f^i\left([\tau_1, \ldots, \tau_k]\right)(y) = f^i_{j_1,\ldots,j_k} \mathcal{F}_f^{j_1}(\tau_1)(y) \mathcal{F}_f^{j_2}(\tau_2)(y) \cdots \mathcal{F}_f^{j_k}(\tau_k)(y).$$

For a Runge-Kutta method to attain order p, the Taylor expansions of the numerical solution y_n and the exact solution $y(t_n)$ must agree for terms up to $\mathcal{O}\left(h^p\right)$. The set of polynomial equations which the coefficients of the Runge-Kutta method must satisfy to achieve order p are known as *order conditions* and can be expressed in terms of functions of rooted trees [20].

Example 1.1. We will find the order conditions necessary for a Runge-Kutta method with three stages to have order three. We begin by tabulating the trees up to order three, and their corresponding elementary differentials, in table 1.1. Included in table 1.1 are values of the functions $\gamma(\tau)$ and $\alpha(\tau)$ along with polynomials $\Phi_i(\tau)$. The meaning of these will become clear during the example.

The functions $\alpha(\tau)$ give the number of trees in the equivalence class of τ; that is, the number of monotonic relabellings of τ. If we denote by $\mathcal{T}_r$ the set of all rooted trees of order r then the rth derivative of the exact solution can be written as

$$y^{(r)}(t_0) = \sum_{\tau \in \mathcal{T}_r} \alpha(\tau) \mathcal{F}(\tau)(y_0). \tag{1.8}$$

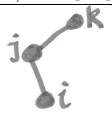
τ	$\emptyset$	$\tau_1 = \bullet$	$\tau_2 = [\tau_1]$	$\tau_{3,1} = [\tau_1, \tau_1]$	$\tau_{3,2} = [\tau_2]$
graph					
$\mathcal{F}(\tau)$	y^i	f^i	$f^i_j f^j$	$f^i_{jk} f^j f^k$	$f^i_j f^j_k f^k$
$\gamma(\tau)$	1	1	2	3	6
$\alpha(\tau)$	1	1	1	1	1
$\Phi_i(\tau)$		1	$\sum_j a_{ij}$	$\sum_{jk} a_{ij} a_{ik}$	$\sum_{jk} a_{ij} a_{jk}$

Table 1.1: Rooted trees and their associated elementary differentials up to order three. The tree $\tau_{i,j}$ is the j-th tree of order i. The choice of which tree, of order i, corresponds to a particular value of j is arbitrary. Here, e.g., $f^i_{jk} = \frac{\partial^2 f^i}{\partial x_j \partial x_k}$ and the summation convention is used.

The function $\gamma(\tau)$ can be calculated by labeling each vertex of τ with the order of the subtree that would result if all edges below that vertex were removed. That is, the "leaves" of the tree are labeled 1 and all other vertices are labeled with the 1 plus the sum of the labels attached to the outward growing neighbours of the vertex (the adjacent vertices which are further from the root.) The product of all such labels gives the value of $\gamma(\tau)$. We illustrate the calculation of γ in figure 1.1.

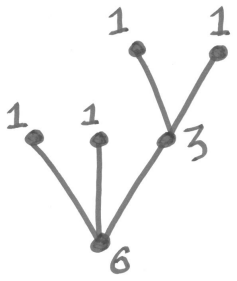


Figure 1.1: Calculation of the function $\gamma(\tau)$ for a tree of order six: $\gamma(\tau) = 1 \times 1 \times 3 \times 1 \times 1 \times 6 = 18$.

The calculation of derivatives of y_1 and Y_i necessary for the Taylor expansion of the numerical solution, although theoretically clear, lead to long and complicated formulas, even for order three. We are saved from these calculations by a theorem (see [20]) which states that the numerical solution y_1 given by the Runge-Kutta method (1.3) satisfies

$$y_1^{(r)}\Big|_{h=0} = \sum_{\tau \in \mathcal{T}_r} \alpha(\tau)\gamma(\tau) \sum_i b_i \Phi_i(\tau) \mathcal{F}(\tau)(y_0). \tag{1.9}$$

Comparing the expressions (1.8) and (1.9) leads to the conclusion that a Runge-Kutta

method (1.3) is of order p if and only if

$$\sum_{i=1}^{s} b_i \Phi_i(\tau) = \frac{1}{\gamma(\tau)} \tag{1.10}$$

for all trees of order $\leq p$. (While the "if" part of this claim is clear, the "only if" part requires some work to show.) For our example of methods of order three, the order equations (1.10) read

$$\begin{aligned} 1 &= \sum_{i}^{3} b_i, \\ \frac{1}{2} &= \sum_{ij}^{3} b_i a_{ij}, \\ \frac{1}{3} &= \sum_{ijk}^{3} b_i a_{ij} a_{ik}, \\ \frac{1}{6} &= \sum_{ijk}^{3} b_i a_{ij} a_{jk}. \end{aligned}$$

Butcher's paper [18] also included simplifying assumptions which reduced the number of conditions needed to obtain a particular order. This work not only made it possible to develop a large number of Runge-Kutta methods, optimal within their class[1], but formed the foundation for further analysis of multistage methods — including their stability and many of their geometric properties.

1.2.3 Partitioned Runge-Kutta methods

Partitioned Runge-Kutta methods involve partitioning a system of differential equations into two, or more, components, $\dot{y} = f(y, x)$, $\dot{x} = g(y, x)$ and discretizing each component with a different Runge-Kutta method;

$$\begin{aligned} Y_i &= \sum_{j=1}^{s} a_{ij} h f(Y_j, X_j) + y_{n-1}, \quad y_n = \sum_{i=1}^{s} b_i h f(Y_i, X_i) + y_{n-1}, \\ X_i &= \sum_{j=1}^{s} \hat{a}_{ij} h g(Y_j, X_j) + x_{n-1}, \quad x_n = \sum_{i=1}^{s} \hat{b}_i h g(Y_i, X_i) + x_{n-1}. \end{aligned}$$

[1] I.e. to find, for example, the diagonally implicit, fourth order method with the smallest local truncation method.

The two Runge-Kutta methods are specified by the coefficients a_{ij}, b_i for the first method and $\hat{a}_{ij}$, $\hat{b}_i$ for the second. (Again we have omitted the c_i and $\hat{c}_i$ coefficients, assuming that the differential equation is in autonomous form.) A necessary, but not sufficient, condition for a partitioned method to be order p, is that the component methods must also be order p. Order conditions for partitioned systems can be calculated using bi-coloured trees, a variation of the trees from Runge-Kutta order conditions [56, Chapter II.15], [55, Chapter III.2].

The first partitioned-Runge-Kutta methods are due to Hofer [67] and Griepentrog [45] who suggested partitioning a system into stiff and non-stiff components, applying an implicit method to the stiff components and an explicit method to the non-stiff components. More recently, partitioned methods have been used for solving Hamiltonian systems where the position and momentum variables of the system allow a natural partitioning and where well chosen pairs of methods result in a partitioned Runge-Kutta method with desirable properties such as symplecticity [55]. Even more recently, partitioned Runge-Kutta methods have been used as *multisymplectic* integrators for discretization of partial differential equations. See, for example, [68, 125] and [124, Chapter 3].

1.2.4 General linear methods

Linear multistep methods and Runge-Kutta methods both fit into the broader framework of general linear methods introduced by Butcher [17]. Such methods allow for information both from past steps and from within the current integration step to be used in determining the numerical solution y_n. An r step, s stage, general linear method is defined by four coefficient matrices, A, U, B and V with entries $[a_{ij}]_{s,s}$, $[u_{ij}]_{s,r}$, $[b_{ij}]_{r,s}$ and $[v_{ij}]_{r,r}$ respectively. The r quantities passed from step-to-step are denoted $y_1^{[n-1]}, y_2^{[n-1]}, \ldots, y_r^{[n-1]}$ at the start of the nth step. The corresponding values at the end of that step are denoted $y_1^{[n]}, y_2^{[n]}, \ldots, y_r^{[n]}$. During each step the s stage values $Y_1, Y_2, \ldots, Y_s$, and the corresponding stage derivatives $F_1(Y_1, \ldots, Y_s), F_2, \ldots, F_s$ are calculated. If supervectors containing either r or s subvectors are defined as follows:

$$y^{[n-1]} = \begin{bmatrix} y_1^{[n-1]} \\ y_2^{[n-1]} \\ \vdots \\ y_r^{[n-1]} \end{bmatrix}, \quad y^{[n]} = \begin{bmatrix} y_1^{[n]} \\ y_2^{[n]} \\ \vdots \\ y_r^{[n]} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_s \end{bmatrix}, \quad F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_s \end{bmatrix},$$

then the general linear method may be neatly described, for a system of dimension N, using Kronecker product notation, as

$$\begin{aligned} Y &= h(A \otimes I_N)F + (U \otimes I_N)y^{[n-1]}, \\ y^{[n]} &= h(B \otimes I_N)F + (V \otimes I_n)y^{[n-1]}, \end{aligned}$$

where I_N is the N-by-N identity matrix.

An advantage of such methods is the extra flexibility which results from having a larger number of coefficients to use as parameters. Hand-in-hand with this is the disadvantage that the methods are correspondingly more difficult to analyse. For examples of general linear methods and discussion of their derivation and practical implementation see [21, 22, 19, 20].

1.2.5 Splitting methods

Sometimes, a vector field contains terms that, while they are simple enough to integrate individually, are difficult to integrate together. For such systems it is often helpful to split the vector field into its simpler parts

$$\dot{y} = f(y) = \sum_i f_i(y), \tag{1.11}$$

and then integrate each of the parts individually before recombining the component solutions. Doing so results in a *splitting method*. These methods involve three equally important steps:

- choosing the components $f_i(y)$ of (1.11) into which to split the vector field;
- integrating the components, either exactly (as is usually the case), or approximately in some suitable fashion;
- recombining the component solutions so as to approximate the solution of the original vector field.

The final step amounts to choosing a set of coefficients a_{ij} so that the resulting method

$$\Phi_{h,f}(y_0) = \prod_{i,j} \psi^{[j]}_{(a_{ij}h),f_j}(y_0) \tag{1.12}$$

approximates the exact flow $\varphi_h(y_0)$, where the $\psi^{[j]}$ are the result of integrating, either exactly, or approximately, the component vector fields f_j.

The new vector fields may be simpler in two possible ways:

- the structure of the component vector fields may be simpler than the original, e.g., they may contain extra symmetries or admit an exact solution;
- the component vector fields may be easier to treat numerically, for example, fast Fourier transforms may be used in their solution.

Because they developed simultaneously in diverse areas, it is difficult to give a clear description of the origins of splitting methods. The notable early publications include

Trotter who, in 1959 [142], showed that, for A and B linear self-adjoint operators on a Banach space and bounded above,

$$\lim_{n\to\infty} (\exp(-hA/n)\exp(-hB/n))^n = \exp(-h(A+B)),$$

giving a first order accurate integrator for $n = 1$, and Strang [135] in 1968 who suggested the second order difference scheme

$$S = L^A_{h/2} L^B_h L^A_{h/2},$$

for vector fields $f = A + B$ and where L is an approximation to the solution of A or B which is second order accurate.

The Baker-Campbell-Hausdorff formula is a useful tool in determining the coefficients in (1.12). It gives an expansion in powers of h for $C(h)$ such that $\exp(hA)\exp(hB) = \exp(C(h))$ where the coefficients in the expansion of $C(h)$ are commutators and repeated commutators of A and B. Yoshida [147] and Suzuki [138, 139] have both presented systematic procedures for determining coefficients such that the resulting integrator has a particular order. The procedures involve solving systems of nonlinear equations which generally have several solutions and are only tractable numerically. McLachlan and Quispel [90] list good[2] values for the coefficients along with details on a variety of different types of splitting methods.

Because splitting methods integrate parts of the vector field separately they allow one to tailor different integration techniques to different pieces. For example, the cubic, nonlinear Schrödinger equation in one space dimension is $iu_t + u_{xx} + c|u|^2 u = 0, \quad c \in \mathbb{R}$. Splitting the vector field into the linear and nonlinear parts allows the linear part with the spatial derivatives to be calculated efficiently and with spectral accuracy in Fourier space while the nonlinear part is calculated in physical space [61, 141, 146]. Changing between physical and Fourier space is cheap due to the fast Fourier transform, and since the splitting methods do not need to store information about the vector field at intermediate steps (like Runge-Kutta methods) or from previous steps (like linear multistep methods) memory requirements are modest — an advantage for partial differential equations in several dimensions. The fact that splitting methods are explicit methods also makes them attractive.

1.3 Hamiltonian systems

Hamiltonian systems are perhaps the most common physical systems to appear in the applied mathematics and physics literature. They describe conservative (i.e. dissipation free) systems from the very large (e.g. celestial mechanics) to the very small (e.g. molecular

[2] I.e. values such that the resulting method has a small leading error coefficient for numerical methods of a particular order or for a particular number of stages

dynamics and quantum mechanics) as well as those in between (e.g. rigid-body dynamics). The differential equations which arise from them are the main focus of this thesis. A d degree of freedom autonomous[3] Hamiltonian system consists of $2d$ first order differential equations

$$X_H := \dot{y} = f(y) = J^{-1}\nabla H(y), \quad J = \begin{bmatrix} 0 & -I_d \\ I_d & 0 \end{bmatrix}, \tag{1.13}$$

where I_d is the $d \times d$ identity matrix.

Definition 1.5 (Poisson bracket)**.** The *canonical Poisson bracket* of two smooth, scalar functions $F(y)$ and $G(y)$ is the function

$$\{F, G\}(y) = \nabla F(y)^\top J^{-1} \nabla G(y), \tag{1.14}$$

where J is the same as in (1.13).

Using the Poisson bracket, the system (1.13) can be written

$$\dot{y}_i = \{y_i, H\}, \quad i = 1, \ldots, d. \tag{1.15}$$

More generally, the derivative of any function $F(y)$ along the flow of the Hamiltonian system is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} F(y(t)) = \{F, H\}(y(t)).$$

If the skew-symmetric matrix J^{-1} depends on y (i.e. $J^{-1} = J^{-1}(y)$) and satisfies the *Jacobi identity*,

$$\{F, \{G, H\}\} + \{H, \{F, G\}\} + \{G, \{H, F\}\} = 0,$$

and *Leibniz's rule*,

$$\{F \cdot G, H\} = F \cdot \{G, H\} + G \cdot \{F, H\},$$

then the system (1.15) is known as a *Poisson system*. On account of this, the skew-symmetric matrix J^{-1} is sometimes called the *Poisson matrix* even when it is constant.

From the skew-symmetric structure of J the system naturally splits into two sets of d equations with *canonically conjugate* variables, $y = (q, p)$, — that is, they satisfy the relations

$$\{q_i, q_j\} = \{p_i, p_j\} = 0, \quad \text{and} \quad \{q_i, p_j\} = \delta_{ij}.$$

This gives

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \text{and} \quad \dot{p}_i = -\frac{\partial H}{\partial q_i} \tag{1.16}$$

as the equations of motion in the conjugate variables.

[3] A non-autonomous Hamiltonian system, $H(y, t)$ is said to have $(d + \frac{1}{2})$ degrees of freedom. We will only deal with autonomous Hamiltonian systems, or non-autonomous systems which have been put into autonomous form by introducing a new variable conjugate to the time variable.

The d coordinates q represent the position — or configuration — variables of the system. The associated p variables represent the conjugate momenta. The Hamiltonian $H(q,p)$ generally represents the total energy of the system. For autonomous systems, where the Hamiltonian does not depend explicitly on time, it is an invariant quantity as we see by differentiating the Hamiltonian with respect to time and substituting (1.16):

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_{i=1}^{d} \left(\frac{\partial H}{\partial q_i} \frac{\mathrm{d}q_i}{\mathrm{d}t} + \frac{\partial H}{\partial p_i} \frac{\mathrm{d}p_i}{\mathrm{d}t} \right) = 0.$$

William Rowan Hamilton, for whom Hamiltonian dynamics are named, presented Hamiltonian dynamics as a reformulation of Lagrangian dynamics [58]. However, Hamiltonian dynamics are much more than just a reformulation; they provide a framework in which perturbation theory and integrability can be expressed (cf. Chapter 3), and give rise to *symplectic geometry*. The systems also exhibit remarkable stability properties for structures such as periodic orbits and invariant tori; this is manifested as KAM (Kolmogorov-Arnold-Moser) theory (cf. section 1.3.3 and chapters 3 and 4). Since (autonomous) Hamiltonian systems do not dissipate energy, their solution trajectories do not simply decay to a rest-state, and their long-time, or even infinite-time, dynamics are of interest. This presents a challenge to numerical integrators.

1.3.1 Symplectic transformations

The Hamiltonian form (1.13) is not preserved under arbitrary coordinate transformations; it is preserved by *symplectic* or *canonical* transformations.

Definition 1.6. A linear transformation $A : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ is a *symplectic transformation* if it satisfies

$$A^\top J A = J, \tag{1.17}$$

where J is the matrix given in (1.13). For a nonlinear transformation $\mathcal{A} : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ the transformation is symplectic if it satisfies (1.17) with A replaced by $D\mathcal{A}$, the Jacobian of the transformation.

Alternatively, one can think of symplectic transformations from the viewpoint of differential geometry: a symplectic transformation is a diffeomorphism $f : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ with coordinates (q,p) which preserves the symplectic two-form $\omega = \frac{1}{2} dy \wedge J dy = dp \wedge dq$. This anti-symmetric, bilinear form acts on pairs of tangent vectors and gives the sum of the oriented areas they define, that is, the areas of the parallelograms formed by projecting the tangent vectors onto the planes spanned by each pair of (canonical) coordinates (q_i, p_i), $i = 1, \ldots, d$. More explicitly,

$$\omega(u,v) = u^\top J v = \sum_{i=1}^{d} \left(u_{p_i} v_{q_i} - u_{q_i} v_{p_i} \right). \tag{1.18}$$

In terms of the symplectic form (1.18), definition 1.6 can be re-presented:

Definition 1.7. A differentiable map $\Phi : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$ is a *symplectic transformation* if the Jacobian of the map $D\Phi(q,p)$ satisfies

$$\omega\left(D\Phi(q,p)u, D\Phi(q,p)v\right) = \omega(u,v).$$

or, equivalently, $\Phi^*\omega = \omega$, where Φ^* is the pull-back of Φ.

An equivalent condition for symplecticity of Φ in terms of the canonical Poisson bracket is that $\{F,G\} \circ \Phi = \{F \circ \Phi, G \circ \Phi\}$ for all smooth functions F and G. Also, note that $\{F,G\} = \omega(X_F, X_G)$.

The following theorem is due to Poincaré [112]:

Theorem 1.2. *If the Hamiltonian $H(y)$ is a twice differentiable function on some neighbourhood of $\mathbb{R}^{2d}$, then the time-t flow map $\varphi_{t,f}(y_0)$ of the Hamiltonian vector field $f = X_H$ is a symplectic map for all t and wherever it is defined .*

Proof. We will show that the symplectic form $\omega(u,v) = u^\top Jv$ is constant along the solutions of $\dot{y} = J^{-1}\nabla H(y)$. The tangent vector $u(t) = \frac{\partial \varphi_t}{\partial y_0}(y_0)u(0)$ satisfies $\dot{u} = J^{-1}\nabla^2 H(y(t))u$, while $v(t)$ satisfies an equivalent equation. The time derivative of $\omega(u,v)$ is therefore given by

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\omega(u,t) &= \dot{u}^\top Jv + u^\top J\dot{v}, \\ &= u^\top \nabla^2 H J^{-\top} Jv + u^\top J J^{-1}\nabla^2 H v = 0,\end{aligned}$$

since $J^\top = -J$ and $J^{-\top}J = -I$. (We have used the notation $J^{-\top} := (J^{-1})^\top$.) □

A consequence is that if we take a set of initial conditions $S_0 \subset \mathbb{R}^{2d}$, for a Hamiltonian vector field, we find that the area, for $d = 1$, (the sum of the oriented areas for $d \geq 2$) of the set $S(t)S_0 := \{y(t) : y(0) \in S_0\}$ is preserved for all t.

We attempt to illustrate this in figure 1.2.

Figure 1.2: Area preservation of a (two-dimensional) symplectic flow. The area of the parallelogram defined by the tangent vectors u and v remains constant as the vectors are carried along the solution curve by the flow φ.

The flow of a Hamiltonian vector field also preserves volume. For every bounded, open set $A \subset \mathbb{R}^{2d}$ and for every t for which $\varphi_{t,f}(y)$ exists

$$\mathrm{vol}(\varphi_{t,f}(A)) = \mathrm{vol}(A),$$

where $f(y) = J^{-1}\nabla H(y)$ and where $\mathrm{vol}(A) = \int_A \mathrm{d}y$. A more general, necessary and sufficient condition for a differential equation to be volume preserving is that it be divergence free, $\nabla \cdot f = 0$, as the following lemma shows.

Lemma 1.3. *The flow of a differential equation $\dot{y} = f(y)$ in $\mathbb{R}^n$ is volume preserving if and only if $\nabla \cdot f = 0$.*

Proof. The Jacobian of the flow map $\varphi_{t,f}(y_0)$ defines the matrix function $Y(t) = D\varphi_{t,f}(y_0)$. The volume of the shape spanned by the columns of Y is given by $\det(Y)$, and $Y(t)$ is a solution of the equation

$$\dot{Y} = A(t)Y(t), \qquad Y(0) = I,$$

where $A(t) := Df(y(t))$ denotes the Jacobian of $f(y(t))$ evaluated at $y(t) = \varphi_{t,f}(y_0)$. To show that the volume defined by Y is constant we use the Abel-Liouville-Jacobi-Ostrogradskii identity to find the time derivative of $\det Y$:

$$\frac{\mathrm{d}}{\mathrm{d}t} \det Y(t) = \mathrm{trace}A(t) \cdot \det Y(t).$$

Since $A(t) = Df(y(t))$ we have $\mathrm{trace}A(t) = \sum_i \frac{\partial f_i}{\partial y_i}(y(t)) = \nabla \cdot f(y(t))$ and therefore $\det Y(t) = \det Y(0) = 1$ if and only if $\nabla \cdot f(y) = 0$. □

The divergence free condition is necessarily true for autonomous Hamiltonian systems;

$$\nabla \cdot f = \nabla \cdot J^{-1}\nabla H = \sum_i \frac{\partial^2 H}{\partial p_i \partial q_i} - \frac{\partial^2 H}{\partial q_i \partial p_i} = 0.$$

A symplectic coordinate transformation $(q, p) \to (\hat{q}, \hat{p})$, is usually obtained via a *generating function* $S(q, \hat{p})$ which implicitly generates a transformation through

$$\hat{q} = \frac{\partial S}{\partial q}(q, \hat{p}), \quad p = \frac{\partial S}{\partial \hat{p}}(q, \hat{p}).$$

The matrix $\frac{\partial^2 S}{\partial p \partial \hat{p}}$ must be non-singular in order that the second equation of the transformation may be inverted to find $\hat{p}(q, p)$. Such transformations are often employed so as to simplify the equations of motion in the new variables. If a symplectic transformation can be found such that the transformed Hamiltonian depends only on the new momentum variables then the system is said to be *integrable* and its dynamics *regular*. In general, such transformations do not exist, and as a consequence chaotic motion is a generic feature of Hamiltonian systems.

Figure 1.3: Nested tori of an integrable Hamiltonian system. A unique torus, carrying a conditionally periodic flow, exists for each value of the action variables.

1.3.2 Integrability

Roughly speaking, a set of differential equations is integrable if it can be solved explicitly for arbitrary initial conditions. Inverting the explicit solution returns the initial conditions, hence the initial conditions are invariant functions — or constants of motion — along the orbits of the system. When a Hamiltonian system can be transformed to a canonical system which depends only on its momenta it is said to be *Liouville integrable*. Arnold showed [4] that when the energy surfaces are compact and the momenta are everywhere independent — we will make these points more concrete in section 3.1 — it is always possible to choose the momentum variables of the new system such that their conjugate position variables are periodic and in the interval $[0, 2\pi]$. Such coordinates are called *angle-action* coordinates and denoted (θ, I). The equations of motion for the transformed Hamiltonian $H(I)$ are $\dot{I} = 0$, $\dot{\theta} = \omega(I) = \frac{\partial H}{\partial I}$. For a fixed value of I the trajectories of the system occur on nested d-dimensional tori and wind around the torus with frequencies $\omega \in \mathbb{R}^d$ (cf. figure 1.3). When the frequency vector is *non-resonant* the trajectories are dense on the torus; otherwise they form a degenerate lower-dimensional torus. We will describe integrable systems and invariant tori in more detail in chapters 3 and 4.

The concept of *first integrals* is important to Hamiltonian systems, particularly for integrability — the Arnold-Liouville theorem (theorem 3.2) presented in section 3.1 requires, amongst other things, that a d-degree of freedom Hamiltonian system possess d first integrals if the system is to be integrable.

Definition 1.8. A non-constant function $I(y)$ is said to be a *first integral* of $\dot{y} = f(y)$ if $I'(y)f(y) = 0$ for all y. That is, $I(y) = I(y_0) = \text{Const.}$ along every solution of $f(y)$.

An autonomous Hamiltonian system with one degree of freedom is always locally integrable since the Hamiltonian function itself is a first integral. Hamiltonian systems with

two or more degrees of freedom are typically non-integrable though there are well known exceptions such as the Toda lattice [55].

1.3.3 Chaotic motion and KAM theory

The Kepler problem — describing the gravitational interaction of two spherical bodies (see, for example [55, Chapter I]) — is an example of a famous, and well studied, integrable Hamiltonian system. However, when other effects, such as additional bodies, even of small mass, are included there appear to be no general, explicit solutions. Poincaré [112] addressed the question of the stability of the solar system and realised that the series used to study perturbations of the Kepler problem were divergent. That is, perturbations of integrable systems need not be integrable. The problem arises from the appearance of small denominators of the form $k \cdot \omega(I)$ in the terms of the series. These terms occur whenever ω is resonant, (i.e. when $\omega_i/\omega_j \in \mathbb{Q}$), or even close to resonant. The fact that the rational numbers are dense would seem to suggest that for arbitrarily small perturbations of integrable Hamiltonian systems none of the invariant tori of the system persist and the motion is entirely, or at least mostly, chaotic. The famous 1955 computational experiment of Fermi, Pasta and Ulam [39] failed to find this supposedly ubiquitous chaotic behaviour. This was explained by *KAM theory*, developed by Kolmogorov [76] and refined by Arnold [2] and Moser [100], which proved that when the perturbation is small enough, and provided that:

- the system satisfies a non-degeneracy condition, and
- is sufficiently differentiable, and
- the frequencies of the invariant torus are sufficiently far from resonance,

then the invariant torus survives the perturbation. The tori which are resonant or close to resonant are generally destroyed and either form degenerate lower dimensional tori — that is, secondary tori of lower dimension trapped in a resonance — or are replaced by regions of chaotic motion. The size of these chaotic regions grows with the perturbation size.

In two degrees of freedom the surviving two-tori form barriers on the three-dimensional energy surfaces. As a consequence, all trajectories inside such tori are bounded and the chaotic motion is restricted to bands between the tori [103]. When the system has three of more degrees of freedom the invariant tori no longer partition the energy surface and thus, there is no obstruction to (chaotic) trajectories connecting distant points in the phase space. The existence of such orbits is known as *Arnold diffusion*. It remains an open problem to prove that Arnold diffusion occurs in any appreciable sense. That is, to show that for small perturbations of an integrable Hamiltonian, a non-zero volume of initial conditions behaves chaotically in the sense that pairs of initial conditions, arbitrarily close, lead to trajectories which diverge from each other and therefore have positive

Lyapunov exponents. Any such diffusion must occur slowly, however, due to a theorem of Nekhoroshev [105, 104] which states that all trajectories of perturbed integrable systems stay close to the unperturbed tori for times which are exponentially long in the inverse perturbation size. The theorem requires a form of smoothness exceeding the existence of infinitely many derivatives of the Hamiltonian. For finitely differentiable Hamiltonians the trajectories remain close for only polynomial times. A local version of the theorem is presented in [55, section X.4.2].

1.4 Geometric numerical integration

In 1952 Curtiss and Hirchfelder [29] suggested that some classes of numerical integrators are better suited to certain types of differential equations. They realised that so called *stiff*[4] differential equations are best solved by implicit numerical methods even though such methods are more computationally expensive. This was the first example of tailoring a numerical method to a particular type of problem, in contrast to the traditional approach where practitioners would solve differential equations using their favourite numerical method. Although Curtiss and Hirchfelder, and those who followed them, recognised that the different intrinsic properties of differential equations made different integrators "good" integrators in different cases, the focus was still on the global error of solutions after finite integration times. In contrast, geometric numerical integration methods try to preserve geometric properties of the flow of a differential equation. The flow of a dynamical system is a map or diffeomorphism of the phase space to itself. The set of diffeomorphisms forms a group with composition of the flow maps as the group operation.

Since numerical integrators can't preserve specific orbits of the flow exactly — especially in the case of long integration periods and non-dissipative systems — geometric integration instead focuses on preserving the special features of the system which control its dynamics. Subgroups of the group of diffeomorphisms can correspond to different types of dynamical features or different classes of dynamical systems. If a numerical integrator is able to preserve these subgroups it may produce a phase portrait which has the same qualitative features as that of the original dynamical system, at least for those features which are structurally stable (i.e. they are shared by all maps close to the flow map).

If the goal of a numerical integrator is to follow a single trajectory as accurately as possible on some fixed time-interval then it is not necessary to distinguish between "classical" (i.e. non-geometric) integrators and geometric integrators[5]. One need only choose a method which is stable for an acceptable step size, and with the smallest local truncation error for the number of function evaluations per step that one is willing to make.

[4]Stiff equations may be most simply defined as being differential equations for which explicit methods don't work. Their linearization typically has some eigenvalues with large negative real part.

[5]Though it can be argued that it is still worth doing so — for example given the choice of two methods with similar local errors and computational cost, one of which preserves invariant sets and one which may not, it would seem to be advantageous to use the method which does.

If, however, one wishes to study a *dynamical system* (or a region of a dynamical system), or to perform calculations over long integration periods, (where long might mean, for the longest integration period one is willing to wait for), then it is important that geometric properties of the dynamical system are replicated by the numerical system. Examples could include:

- Determining whether the solar-system is chaotic, where one is not interested in the exact position of the planets after some millions of years, but rather in whether any Lyapunov exponents are converging to positive values as $t \to \infty$.
- Calculating the preferred structures of a molecule, one wants to know the configuration where the molecule spends most of its time and the probability of the molecule switching to other nearby configurations. However, one does not wish to know the exact positions of every atom at every instant in time.
- In order to simulate a particle accelerator and determine the best position of the magnets which control the beam of accelerated particles it is necessary to accurately predict the region of space that the particles will be confined to. The trajectories and positions of individual particles need not be known.

These examples all have in common that the global error associated with their calculation can become very large, that is, $\mathcal{O}(1)$, due to the length of the integration period and the fact that the systems are non-dissipative and do not decay to some rest state. They also have in common conserved quantities and underlying structure (all are Hamiltonian systems) which can be exploited in order to achieve good numerical results.

We will focus on symplectic integrators (and their application to Hamiltonian systems) though of course geometric numerical integration is far wider than such numerical methods and dynamical systems. By symplectic integrator we mean any numerical integrator whose Jacobian satisfies (1.17).

Numerical methods of symplectic integration were introduced by de Vogelaere [32] in 1956 in a report; however, this remained unpublished and it was not until the 1983 article by Ruth [122] that the idea of geometric integration — designing integrators which had the same physical properties as the differential equations they were to be used with — received attention. Ruth, a physicist working on problems related to particle accelerators, took the point of view that since differential equations derived from a Hamiltonian system had a symplectic structure (their flow generates a symplectic map), numerical integrators could never correctly solve such equations, no matter their accuracy, unless they too gave a symplectic map. Other early work on symplectic integration also came from within the physics community, notably in reports from Channell [23] in 1983 and Neri [106] in 1988. The initial approaches to constructing symplectic integrators commonly used the observation [79] that any coordinate transformation derived from a generating function gives a symplectic map. Publications by Kang [74] in 1986 and Sanz-Serna [126] and

Lasagni [80] in 1988, amongst others, helped establish the idea, within the numerical analysis community, that a geometric approach to integration could yield better qualitative results. By the early 1990s publications such as the book *Numerical Hamiltonian Problems* [127] by Sanz-Serna and Calvo and the survey article [24] by Channell and Scovel had established geometric integration as a distinct field. Further details on the development of geometric integration can be found in the survey articles by McLachlan and Quispel [89, 91], and Budd and Piggot [16] and in the book [55] by Hairer, Lubich and Wanner.

Probably the best known, and most widely used, geometric integrator today is the Störmer-Verlet or leapfrog method which, for a separable Hamiltonian $H(q,p) = T(p) + V(q)$ with $T(p) = \frac{1}{2}p^\top p$, reads:

$$q_{n+\frac{1}{2}} = q_n + \frac{h}{2}p_n, \quad p_{n+1} = p_n - h\nabla V(q_{n+\frac{1}{2}}), \quad q_{n+1} = q_{n+\frac{1}{2}} + \frac{h}{2}p_{n+1}. \tag{1.19}$$

The method was publicised, at least to the molecular dynamics community, by Verlet [144, 145] who used it to conduct computer "experiments" on a system of argon atoms. It is also often used in celestial mechanics (it is sometimes called Encke's method after J. F. Encke who, around 1860 used it for extensive calculations for the perturbation terms of planetary orbits [54]) and in the solution of partial differential equations for wave propagation (because it is easy to code, cheap to run and generally produces acceptable results). In fact the leapfrog method is perhaps the archetype of geometric numerical integrators — Hairer, Lubich and Wanner use it exclusively in [54], a 50 page expository article on geometric numerical integration. The method fits into the framework of both partitioned Runge-Kutta methods and splitting methods. The latter can be seen by splitting the Hamiltonian $H = T + V$ into $\varphi_{t,T}$ and $\varphi_{t,V}$, the flows of the kinetic- and potential-energy parts respectively. The composition

$$\varphi_{h/2,T} \circ \varphi_{h,V} \circ \varphi_{h/2,T}$$

then yields the leapfrog method (1.19).

However, the method is also one of the earliest examples of a (geometric) numerical integrator, appearing in 1687 in Book I of Newton's *Principia Mathematica* [107] (and reproduced in figure 1.4) where it was shown to conserve angular momentum and used by Newton to prove Kepler's second law of planetary motion — that planets sweep out equal areas during equal intervals of time. In his "Leçon d'Adieu", Gerhard Wanner calls this proof by Newton *"the first proof of modern science"*[6].

Newton's exposition of the algorithm also shows that it is reversible, that is, after running the method forwards in time one can reverse the direction of time and run the method backwards returning to the same points in phase space as during the forward

[6]The slides which accompanied Wanner's presentation can be found at `www.unige.ch/~wanner/Lecondadieu.pdf`

Dividatur tempus in partes æquales, & prima temporis parte deſcribat corpus vi inſita rectam *AB*. Idem ſecunda temporis parte, ſi nil impediret, recta pergeret ad *c*, (per Leg. I) deſcribens lineam *Bc* æqualem ipſi *AB*, adeo ut radiis *AS*, *BS*, *cS* ad centrum actis, confectæ forent æquales areæ *ASB*, *BSc*. Verum ubi corpus venit ad B, agat vis centripeta impulſu unico ſed magno, faciatq; corpus a recta *Bc* deflectere & pergere in recta *BC*. Ipſi *BS* parallela agatur *cC* occurrens *BC* in *C*, & completa ſecunda temporis parte, corpus (per Legum Corol. 1) reperietur in *C*, in eodem plano cum triangulo *ASB*. Junge

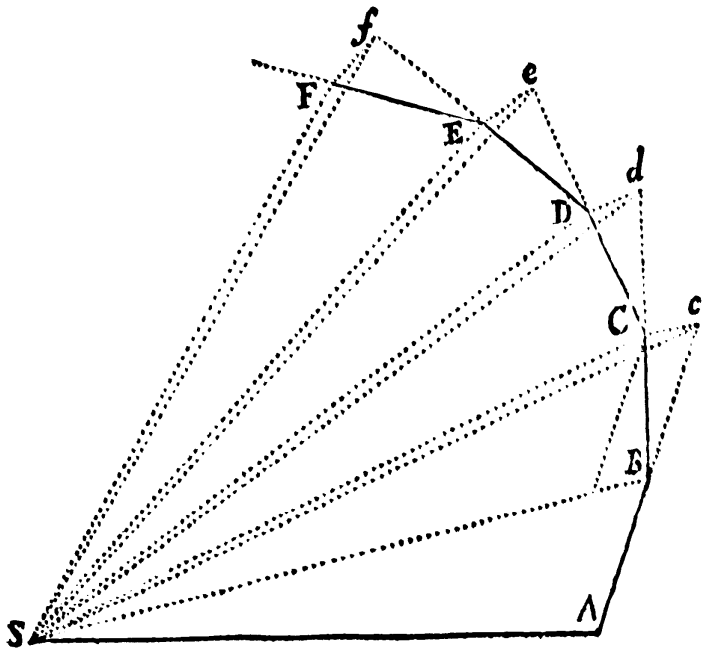


Figure 1.4: The leapfrog method in Book I of Newton's *Principia Mathematica*. The algorithm conserves area and angular momentum. By letting the step sizes shrink infinitesimally, Newton proves that under gravitational attraction, Kepler's second law holds: planets sweep out equal areas of their orbits during equal times.

iterations. More recently, the leapfrog method and its generalisations have been of interest as examples of symplectic integrators.

1.4.1 Symplectic integrators

If one takes phase space preservation to be a necessary property of geometric integration, a consequence is that linear multistep methods of numerical integration cannot be geometric integrators since they use values of the vector field and/or solution at several different time levels. (A similar argument applies to general linear methods.) However, one can, in the case of a strictly stable linear multistep method, talk about the associated equivalent or underlying one-step method [75]. Tang [140] proved in 1993 a conjecture of Kang that no linear multistep method can be symplectic, in the sense that the underlying one-step method can not be symplectic for $\dot{y} = f(y)$. It is believed, though not proven, that the same result holds for general linear methods [55]. Linear multistep methods can, however, be time reversible or symmetric making them "good" methods for reversible Hamiltonian systems. Such methods have received attention as integrators for long-time simulations in celestial mechanics [53, 50, 115]. Chartier, Faou and Murua [25] have shown that for Newton equations[7] if a linear multistep method is applied to the second order equation $\ddot{y} = f(y)$ then, if the method is symmetric its underlying one-step method is formally conjugate to a method which *is* symplectic. An example is the leapfrog method which is known to be symplectic, and can be expressed in the form of a linear multistep method:

$$y_{n+1} - 2y_n + y_{n-1} = h^2 f(y_n).$$

Here we will briefly mention two of the most popular types of symplectic integrators: Runge-Kutta/partitioned Runge-Kutta methods and splitting methods. The first class of methods benefits from the fact that (partitioned) Runge-Kutta methods are already widely known, well understood and generally trusted as classical numerical integrators, hence little thought is required to switch from a non-symplectic to a symplectic Runge-Kutta method.[8] Splitting methods, though more novel from a numerical analysis point of view, benefit from being explicit (hence fast), and very flexible. However, they require some analysis of the differential equation in question before they can be applied.

Symplectic (partitioned) Runge-Kutta methods

Runge-Kutta methods were one of the earliest classes of numerical integrators to be investigated with respect to symplecticity. For a Runge-Kutta method of the form (1.3),

[7] That is, equations with a Hamiltonian of the form $H(q,p) = \frac{1}{2}p^\top p + V(q)$.

[8] The issue of symplectic Runge-Kutta methods all being implicit — and, hence, relatively expensive in comparison to many (explicit) popular "classical" Runge-Kutta methods — is, however, a point *against* the idea that one can simply "upgrade" from an arbitrary non-symplectic Runge-Kutta method to a symplectic one.

independent results from Sanz-Serna [126] and Lasagni [80] show that the method is symplectic if and only if the condition

$$b_i a_{ij} + b_j a_{ji} - b_i b_j = 0, \quad i, j = 1, \ldots, s, \tag{1.20}$$

holds for the coefficients of the method. The left-hand side of this condition defines the matrix which must be non-negative definite if the Runge-Kutta method is to be algebraically stable [57]. Since the condition for algebraic stability of Runge-Kutta methods is essentially ensuring that the integrator damps the solutions of dissipative systems, the requirement that all entries in the matrix are zero for a symplectic integrator avoids the numerical solutions being either artificially damped or driven. Sanz-Serna and Calvo give a detailed study of symplectic Runge-Kutta and partitioned Runge-Kutta methods in [127].

The classes of Runge-Kutta methods which satisfy (1.20) include the Gauß methods of order $2s$, where s is the number of internal stages used by the method. Such methods are popular, partly on account of their high order but also because they can be implemented as a "black-box" where only the differential equation $\dot{y} = f(y)$ needs to be known. A disadvantage of the Gauß, and all other symplectic Runge-Kutta methods, is that they are implicit and a system of nonlinear equations must be solved at each step. If one wants to take advantage of the exact preservation of the symplectic form, the system of equations must be solved to machine precision.

Partitioned Runge-Kutta methods are symplectic if their coefficients satisfy

$$b_i \hat{a}_{ij} + \hat{b}_j a_{ji} - b_i \hat{b}_j = 0, \quad b_i = \hat{b}_i, \quad i, j = 1, \ldots, s. \tag{1.21}$$

The Lobatto IIIA-IIIB pair is such a class of methods.

It is worth mentioning at this point that much of the good behaviour of symplectic integrators relies on them being implemented with constant step size. Gladman, Duncan and Candy [44] observed that when such integrators are implemented with adaptive step sizes they lose many of their desirable qualities and behave like classical integrators, displaying, for example, linear energy error growth. Symplectic integrators form a group under composition, hence, a sequence of steps by a symplectic integrator, using various step sizes, yields a symplectic map:

$$\Phi_{h^*} = \Phi_{h_k} \circ \Phi_{h_{k-1}} \circ \cdots \circ \Phi_{h_2} \circ \Phi_{h_1}.$$

However, as we explain in section 1.4.2, the map generated by a symplectic integrator is exponentially close to the flow of a *modified Hamiltonian* which is constructed as a series expansion in the step size h. Changing step size therefore changes the modified Hamiltonian being approximated; such a process need not give a good approximation of the flow of the original Hamiltonian system. Additionally, many of the theorems concerning the good structure preservation properties of symplectic integrators rely on repeated iteration

of a single symplectic map for their proof (cf. KAM theory, chapter 3).

Splitting methods

Splitting methods (cf. section 1.2.5) offer an alternative approach for developing geometric integrators. They are more "hands on" than Runge-Kutta methods since they require that one must decide how the vector field is to be split. Sometimes this choice will be clear, such as when the vector field can be split into parts which can be integrated exactly, thereby preserving all the geometric properties of the individual components. In such cases it just remains to ensure that the composition of the individual solutions, to form the final map, retains the geometric structures. In the case of symplectic integrators, symplectic maps form a group under composition, and hence splitting methods preserve the symplecticity of the total flow whenever symplecticity is preserved for the individual flows. The same holds for any other vector fields whose flows form a group. Issues arise when it is not obvious which particular group of diffeomorphisims the flow of the vector field belongs to. The classification, and discussion of groups of diffeomorphisims, with respect to splitting methods and geometric integration, is treated in [90].

1.4.2 Backward error analysis

The technique of *backward error analysis* allows one to make comparisons between the numerical trajectory of an integrator and the exact solution of a *modified equation*. Often, these comparisons agree, up to terms which are exponential in some small parameter, (the integrator step size), and, as such, are useful for studying the behaviour of a numerical method over very long, but still finite, integration periods. As such, we will generally avoid taking the backward error approach to our analysis since we are interested in *exact* preservation of structures for unbounded integration periods. Nevertheless, we will need to understand backward error analysis in order to appreciate why such an approach is not sufficient for studying structure preservation. We will also make use of ideas from backward error analysis, such as the idea of modified Hamiltonians and modified vector fields, in chapter 4 when we prove that symplectic integrators preserve periodic orbits.

Consider an ordinary differential equation $\dot{y} = f(y)$ treated with a numerical method $\Phi_{h,f}(y)$, which produces a numerical trajectory $y_0, y_1, y_2, \ldots$ approximating $y(t)$. A forward error analysis consists of studying the local $\|y_1 - \varphi_{h,f}(y_0)\|$ and global $\|y_n - \varphi_{nh,f}(y_0)\|$ errors in the solution space. In contrast, the idea of backward error analysis is to find a *modified differential equation* $\dot{\tilde{y}} = \tilde{f}_h(\tilde{y})$ of the form

$$\dot{\tilde{y}} = f(\tilde{y}) + h\tilde{f}_2(\tilde{y}) + h^2\tilde{f}_3(\tilde{y}) + \ldots, \tag{1.22}$$

such that $y_n = \tilde{y}(nh)$. The numerical trajectory is then interpreted as the exact solution of a modified differential equation. One can then study the behaviour of the modified

equation in order to understand the numerical solution.

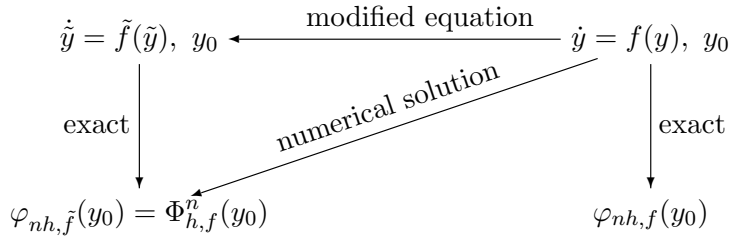


Figure 1.5: Schematic of backward error analysis using modified differential equations.

In general, the series (1.22) diverges and must be suitably truncated. The series can be computed, formally, by setting $\tilde{y}(t) = y$ for a fixed t and expanding the solution of (1.22) in a Taylor series

$$\begin{aligned}\tilde{y}(t+h) = y &+ h\left(f(y) + h\tilde{f}_2(y) + h^2\tilde{f}_3(y) + \ldots\right)\\ &+ \frac{h^2}{2}\left(f'(y) + h\tilde{f}_2'(y) + \ldots\right)\left(f(y) + h\tilde{f}_2(y) + \ldots\right)\\ &+ \frac{h^3}{3!}\left(f''(f,f)(y) + f'f'f(y) + \ldots\right) + \ldots,\end{aligned} \tag{1.23}$$

where the terms at order h^3 and higher involve elementary differentials of f and $\tilde{f}_j$, (cf. section 1.2.2). Writing the numerical method $\Phi_{h,f}(y)$ as a power series in h, and assuming that the method is consistent,[9] and hence, the coefficient of the order h term is $f(y)$, gives:

$$\Phi_{h,f}(y) = y + hf(y) + h^2c_2(y) + h^3c_3(y) + \ldots \tag{1.24}$$

The coefficient functions $c_2(y)$, $c_3,(y)$, *et cetera* are known from the construction of the numerical method and generally consist of combinations of $f(y)$ and its derivatives. We can now calculate the coefficients of (1.22) by requiring $\tilde{y}(nh) = y_n$, hence, $\tilde{y}(t+h) = \Phi_{h,f}(y)$. Comparing the coefficients of h in equations (1.23) and (1.24) yields recurrence relations for $\tilde{f}_j(y)$:

$$\begin{aligned}\tilde{f}_2 &= c_2(y) - \frac{1}{2}f'f(y),\\ \tilde{f}_3 &= c_3(y) - \frac{1}{3!}\left(f'f'f(y) + f''(f,f)(y)\right) - \frac{1}{2}\left(f'\tilde{f}_2(y) + \tilde{f}_2'f(y)\right).\end{aligned}$$

The coefficient functions $c_j(y)$ are necessarily such that the perturbation terms in the modified equation are $\mathcal{O}\left(h^p\right)$ where p is the order of the numerical integrator.

One of the most important consequences of modified equations is that for Hamiltonian

[9] Consistency is a prerequisite for any viable numerical integrator. It requires that the method is able to solve the equation $\dot{y} = 0$ exactly.

systems $\dot{y} = J^{-1}\nabla H(y)$, with an infinitely differentiable Hamiltonian $H(y)$, the modified equation of a symplectic method is also Hamiltonian. That is, when the continuous-time flow of a Hamiltonian system preserves the Hamiltonian function, the time-h map of a symplectic integrator preserves (up to the order at which the modified equations are truncated) a nearby Hamiltonian function. We illustrate this idea in figure 1.6.

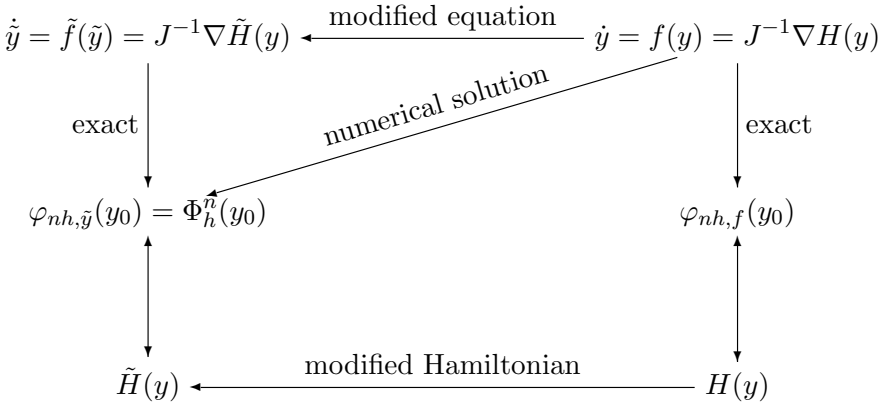


Figure 1.6: Schematic showing the idea of preservation of a modified Hamiltonian when a symplectic numerical integrator is used with a Hamiltonian differential equation.

The following theorem gives the existence of a *local* modified Hamiltonian. Whenever the symplectic method has a generating function (cf. section 1.3.1) then one can also prove existence of a *global* modified Hamiltonian defined on the same domain as the generating function.

Theorem 1.4 (Existence of a local modified Hamiltonian)**.** [55, theorem 3.2, section IX.3.2] *If a symplectic method $\Phi_h(y)$ is applied to a Hamiltonian system with a smooth Hamiltonian $H : \mathbb{R}^{2d} \to \mathbb{R}$, then the modified equation (1.22) is also Hamiltonian. More precisely, there exist smooth functions $\tilde{H}_j : \mathbb{R}^{2d} \to \mathbb{R}$ for $j = 2, 3, \ldots$, such that $\tilde{f}_j(y) = J^{-1}\nabla\tilde{H}_j(y)$.*

Similarly, the modified equation for a Poisson integrator applied to a Poisson system is also locally a Poisson system.

Of course, since it is necessary to truncate the modified equation (1.22) before it diverges, the numerical trajectory produced by $\Phi_{h,f}$ is not *exactly* the flow of a Hamiltonian system even though the modified equation is. This *almost* preservation of a modified Hamiltonian is responsible for the good long-time energy conservation of symplectic integrators applied to Hamiltonian systems. The following theorem is due to Benettin and Giorgilli [11]

Theorem 1.5. *Consider a Hamiltonian system with analytic $H : U \to \mathbb{R}$, and $U \subset \mathbb{R}^{2d}$, and apply a symplectic numerical method $\Phi_h(y)$ with step size h. If the numerical solution*

stays in a compact set $K \subset U$, then there exists a constant h_0, (which depends on the size of the domain of analyticity of H and on the bounds of the derivatives of the numerical solution), and an integer $N_0 := \max\{N \in \mathbb{Z} : hN < h_0\}$, such that

$$\tilde{H}(y_n) = \tilde{H}(y_0) + \mathcal{O}\left(e^{-h_0/2h}\right),$$
$$H(y_n) = H(y_0) + \mathcal{O}\left(h^p\right),$$

over exponentially long time intervals $nh \leq e^{h_0/2h}$.

The significance of the fact that symplectic algorithms fail to conserve the total energy of Hamiltonian systems can be well illustrated by comparing the numerical solution of the Hamiltonian pendulum

$$H(q,p) = \frac{p^2}{2} - \cos(q) \tag{1.25}$$

with its modified equation. For the (symplectic, second-order) leapfrog integrator (cf. (1.19)) the modified Hamiltonian, up to $\mathcal{O}\left(h^2\right)$, is

$$\tilde{H}(q,p) = H(q,p) + h^2\left(\frac{-1}{24}p^2\cos(q) + \frac{1}{12}\sin^2(q)\right). \tag{1.26}$$

In figure 1.7 we plot the numerical trajectories of the leapfrog integrator applied to the system generated by (1.25), and the level sets of the modified Hamiltonian (1.26) for various step sizes h and for initial conditions $(q_0, p_0) = (-1.5, 0)$, $(q_0, p_0) = (-2.5, 0)$, $(q_0, p_0) = (-\pi, 1.5)$, and $(q_0, p_0) = (-\pi, 2.5)$. The integration length is 200000 steps in all cases. The figures show that for $h \leq 0.6$, larger than would typically be used, there is excellent agreement between the numerical trajectories and that the modified Hamiltonian; that is, $\tilde{H}$ is well conserved. For larger step sizes the numerical trajectories become chaotic, highlighting the effect that the non-existence of an exact, (that is convergent), modified Hamiltonian has on the dynamics. Yoshida [148] discusses the relationship between this chaotic behaviour and the divergence of the un-truncated series for the modified Hamiltonian.

The figure also shows that the numerical trajectory closest to the fixed point at $(q,p) = (0,0)$ manages to retain the correct dynamics, even when the step size becomes very large. For $h = 1.2$ the periodic orbit is replaced by a chaotic trajectory with the initial condition close to a seperatrix. The chaotic trajectory, however, remains within a bounded region of the phase space. This aspect of the numerical dynamics can be explained by KAM type theorems, (cf. chapters 3 and 4), and comes from the fact that the leapfrog integrator is symplectic.

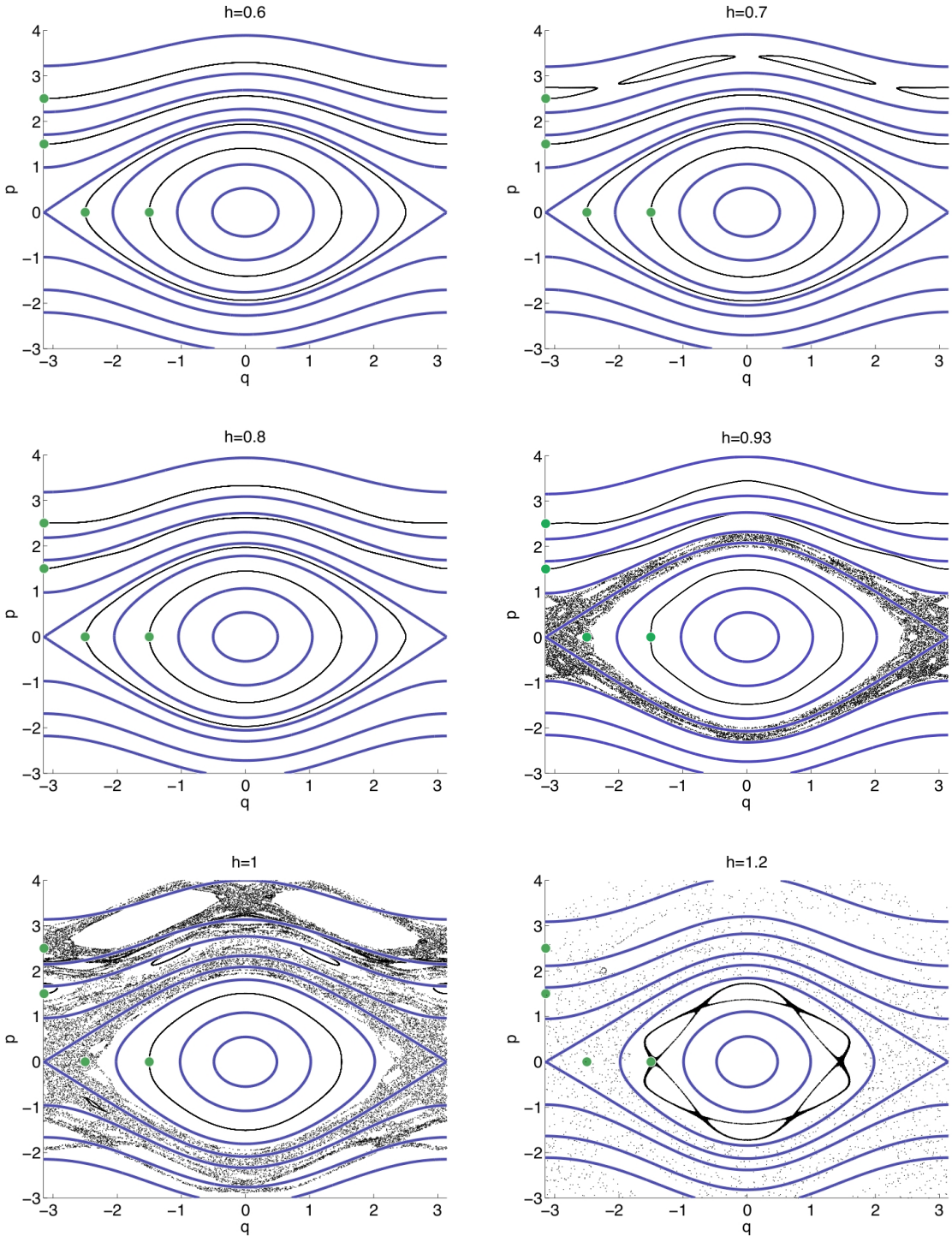


Figure 1.7: Numerical solutions of the pendulum by the leapfrog integrator for a range of large step sizes and for 200000 steps. The four initial conditions used are indicated. The level sets of the corresponding modified Hamiltonian truncated after the $\mathcal{O}\left(h^2\right)$ term, (cf. equation (1.26)), are the thick, solid lines.

1.5 Thesis outline

The aim of this thesis is to give insight into the preservation of phase space structure for Hamiltonian systems when the equations of motion are solved numerically. We present theorems and conjectures on when numerical integrators preserve such structures, and on what happens when they do not. We also make extensive use of a variety of applications and numerical experiments to illustrate why a geometric, or structure preserving, approach to numerical integration is important and how it is possible.

The original work in this thesis is based on three separate projects and, therefore, chapters 4, 5 and 6, where this is presented, can each stand alone to some extent. The common thread throughout the chapters is the importance of a "geometric" view of numerical integration: the idea that numerical integration is a tool for studying the dynamics of differential equations rather than just computing numerical trajectories to a certain accuracy.

The chapters of this thesis are organised as follows: Chapter 2 presents background on the preservation of invariant sets by numerical integrators. We discuss what is known about preservation of fixed points and their spectra, stable and unstable manifolds, periodic orbits, invariant tori, and chaotic invariant sets, as well as statistical properties of differential equations. Where possible, we consider the treatment of each of these sets by linear multistep, Runge-Kutta and splitting methods. As part of this, we present and prove a therom concerning the fixed points of splitting methods. We are not aware of this theorem existing previously in the literature.

Chapter 3 gives a detailed background to perturbation theory and Kolmogorov's iteration, which can be used to prove that most invariant tori of integrable Hamiltonian systems survive small Hamiltonian perturbations. We give a detailed proof, based on that in [55], that Kolmogorov's iteration converges.

This lays the foundation for chapter 4 where we prove that most periodic orbits of Hamiltonian differential equations persist under a symplectic discretization, though with a periodic perturbation. The proof uses existing results [114, 98] to embed the trajectory of a symplectic integrator in the non-autonomous flow of a modified Hamiltonian. It then uses a theorem from [73] to prove that the periodic orbit of the original flow is replaced by an invariant 2-torus in the perturbed flow. The motion in the additional dimension is very small, and comes about from the quasi-periodic perturbation due to the numerical integrator. The chapter also includes a conjecture on the breakup of periodic orbits of Hamiltonian systems with $d \geq 2$ degrees of freedom when the integrator step size is resonant with the frequency of the periodic orbit. Supporting numerical evidence is given and an application to Hamiltonian partial differential equations is discussed.

Chapter 5 concerns numerical integrators for highly oscillatory Hamiltonian systems with constant frequency. It studies so-called trigonometric integrators and uses a variety of numerical tests to show that such integrators are perhaps not as practical as previous

publications suggest. It is shown that the trigonometric integrators suffer from higher order resonances, in addition to the low order resonances already known, and that, in many cases, they perform poorly at preserving various properties of differential equations which may be important to the study of the dynamical system.

Chapter 6 presents a numerical study of two coupled magnetic moments in an applied magnetic field. It is motivated by the article [118] where the authors used numerical investigation of the system to better understand quantum chaos. However the approach of [118] did not appear to take a geometric point of view. We investigate the effect of choice of integrator and choice of coordinate system on the numerical solutions. The system is interesting to investigate in its own right with non-trivial dynamics.

The thesis concludes, in chapter 7, with a summary, and a discussion of open problems related to the preceding chapters.

1.6 List of original contributions

Chapter 2 (partly original)

- Theorems on preservation of fixed points and prevention of spurious fixed points for first order splitting methods.

Chapter 4 (entirely original)

- KAM-type theorem for preservation of periodic orbits of Hamiltonian systems by symplectic integrators with a non-resonant step size. The theorem also holds for other lower dimensional invariant tori, though the necessary assumptions can be more restrictive in this case.
- Numerical investigation of break-up of periodic orbits of Hamiltonian systems with two or more degrees of freedom when the frequency of the periodic orbit is resonant with the step size of the (symplectic) integrator. Based on the results of the investigation, a conjecture on resonant symplectic discretization of Hamiltonian periodic orbits is presented along with a new application for steady-state solutions of Hamiltonian partial differential equations.

Chapter 5 (entirely original)

- Numerical investigations showing evidence of previously unreported third order resonances in trigonometric integrators for highly oscillatory problems. These resonances can lead to unstable numerical solutions.
- Numerical investigation of previously unstudied structural properties and long-time dynamics of trigonometric integrators. The behaviour of the trigonometric methods is compared with that of a non-trigonometric integrator not specifically designed for highly oscillatory problems. It is found that in some respects the trigonometric methods perform no better than the midpoint rule.

Chapter 6 (entirely original)

- Investigation of a coupled two-spin system in an applied magnetic field including:
  - investigation of the effects of symplectic versus non-symplectic integration, and
  - investigation of the effect of the choice of coordinate system on accuracy and structure preservation of numerical solutions.

Chapter 2

Structure Preservation: Invariant Sets

> Numerical experiments are crucial to the development of insight into the behaviour of dynamical systems.... However, a central question must be faced when utilizing such numerical results: In what sense do the numerical experiments with their inherent computer roundoff reflect the true dynamics of the actual system?
>
> — S. M. Hammel, J. A. Yorke and C. Grebogi, 1987

This chapter reviews some of the issues when using numerical integrators as tools for the study of dynamical systems. As we mentioned in chapter 1, numerical analysis has typically been concerned with the study of methods as the step size tends to zero while the integration length remains finite. In contrast, for dynamical systems one wants to study the behaviour of systems as $t \to \infty$. In this case, the accuracy with which numerical objects are computed may be less important than the fact that numerically computed objects are qualitatively correct.

Invariant sets are an important aspect of dynamical systems; they are often of lower dimension (than the full system), and they may carry simple(r) dynamics.

Definition 2.1 (Invariant set)**.** For a dynamical system described by $f : \mathbb{R}^n \to \mathbb{R}^n$, the set $A \subset \mathbb{R}^n$ is an *invariant set* if $\varphi_{t,f}(A) \subset A$ for all t. If $\varphi_{t,f}(A) \subset A$ for all $t > 0$ then A is a *forward (or positively) invariant* set. It is *backward (or negatively) invariant* if $\varphi_{t,f}(A) \subset A$ for all $t < 0$.

When a property of a dynamical system, such as an invariant set, persists under perturbations of the same class as the dynamical system (for example, Hamiltonian perturbations of a Hamiltonian system), the property is said to be structurally stable within that class. It would seem desirable that numerical integrators should reproduce any structurally stable invariant sets of a dynamical system if they are to help us understand its

dynamics. The accuracy of the integrator does not necessarily help with this goal. Take the example of a chaotic dynamical system. Sensitive dependence on initial conditions, and trajectories which typically diverge exponentially quickly, are serious factors limiting the efficacy of numerical integrators as the error in each step is subsequently amplified by the divergent trajectories. Even if the integration can be done exactly at each step and one is simply iterating a diffeomorphism, the round-off error in each evaluation leads to a new trajectory. If the distance between pairs of orbits grows geometrically, then an error of machine epsilon[1] may grow to $\mathcal{O}(1)$ after only a few dozen iterations.

For any particular invariant set, it is not trivial to say what it means for such a set to be preserved by a numerical integrator, or for a spurious invariant set of the same type to be introduced. A useful definition is to say that an invariant set has been preserved if, each point in the invariant set of a numerical method lies within some — hopefully small — bounded region of a corresponding point in an invariant set of the original dynamical system, and that the size of the region about each point shrinks towards zero with decreasing step size.

In this chapter we will discuss the computation of various invariant sets: fixed points (and their spectra), stable/unstable manifolds, periodic orbits, quasi-periodic orbits and invariant tori and chaotic invariant sets. We also briefly discuss computation of statistical properties. We will try to say what is known (at least what is known by the author) about each property with respect to three classes of integrators: Runge-Kutta, linear multistep and composition/splitting methods. Many of the examples and theorems in sections 2.1 – 2.4 are taken from the article [136] and the book [137].

2.1 Fixed points & their spectra

Definition 2.2 (fixed point)**.** A point y^* is a *fixed point (or stationary point or equilibrium point)* of a differential equation $\dot{y} = f(y)$ (respectively, a map $\Phi(y_n) = y_{n+1}$) if $f(y^*) = 0$, (resp. $\Phi(y^*) = y^*$).

Preserving existing fixed points of dynamical systems is relatively easy for numerical integrators. For example, all Runge-Kutta and linear multistep methods preserve all fixed points exactly. Theorem 2.1 and its proof shows why this is true for Runge-Kutta methods.

Theorem 2.1 (Runge-Kutta methods preserve fixed points)**.** *A Runge-Kutta method applied to a differential equation $\dot{y} = f(y)$ retains all fixed points of the differential equation. that is,*

$$f(y^*) = 0 \Rightarrow y^* = \Phi_{h,f}(y^*),$$

Proof. Suppose $f(y^*) = 0$ and $y_n = y^*$, then the Y_i of the Runge-Kutta method (1.3) satisfy $Y_i = y^*$ for $i = 1, \ldots, s$ and hence $y_{n+1} = y_n = y^*$. □

[1] That is, the smallest number ε_m such that the computer calculation $(1 + \varepsilon_m) - 1 \neq 0$.

For linear multistep methods the proof is not so straightforward as the methods are defined on a product space rather than on the phase space of the differential equation.

It is important that preserved fixed points are of the same type for both the original dynamical system and its discrete approximation. Stability, in particular, is important to preserve since it influences nearby points in the phase space.

Definition 2.3 ((asymptotically) stable fixed point)**.** A fixed point y^* of a map $\Phi(y_n) = y_{n+1}$ (resp. a differential equation $\dot{y} = f(y)$) is said to be *stable* if, for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $y_0 = y(0) \in B(y^*, \delta)$ implies $y_n \in B(y^*, \varepsilon), \forall n \in \mathbb{N}$, (resp. $y(t) \in B(y^*, \varepsilon), \forall t \geq 0$). The fixed point is *asymptotically stable* if it is stable and $\|y_n - y^*\| \to 0$ as $n \to \infty$, (resp. $\|y(t) - y^*\| \to 0$ as $t \to \infty$) for $\|y_0 - y^*\|$ sufficiently small.

Hyperbolicity (or non-hyperbolicity) is another important property of fixed points. It essentially says that the neighbourhood of a fixed point is partitioned into attracting and repelling directions.

Definition 2.4 (hyperbolic fixed point of a map)**.** A fixed point y^* of a map Φ is said to be *hyperbolic* if none of the eigenvalues of the Jacobian $D\Phi|_{y^*}$ lie on the unit circle.

Definition 2.5 (hyperbolic fixed point of a differential equation)**.** A fixed point y^* of a differential equation $\dot{y} = f(y)$ is said to be *hyperbolic* if none of the eigenvalues of the Jacobian $Df|_{y^*}$ lie on the imaginary axis.

While definition 2.3 describes the behaviour of a stable fixed point, it does not give a way to test whether a fixed point is stable or not. The following two theorems say that the stability of a fixed point can be determined by simply looking at the stability of the fixed point in the linearized system. Proofs of these, and of other several other theorems presented in this chapter can be found in the book by Stuart and Humphries [137].

Theorem 2.2 (linear stability of maps implies asymptotic stability)**.** *A fixed point y^* of a twice differentiable, map Φ is asymptotically stable if all the eigenvalues of the Jacobian $D\Phi|_{y^*}$ lie strictly inside the unit circle. If any of the eigenvalues lies outside the unit circle the fixed point is unstable.*

Theorem 2.3 (linear stability of differential equations implies asymptotic stability)**.** *A fixed point y^* of a twice differentiable, differential equation $\dot{y} = f(y)$ is asymptotically stable if all the eigenvalues of the Jacobian $Df|_{y^*}$ lie strictly in the left half-plane. If any of the eigenvalues lies in the right half-plane, the fixed point is unstable.*

In order that we can precisely state a general theorem on when fixed points are preserved by a numerical integrator it is necessary to first introduce the following assumptions

Assumption 2.6. Consider a locally Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^n$ whose flow $\varphi_{t,f} : \mathbb{R}^n \to \mathbb{R}^n$ generates a (semi-)group for all $t > 0$. We will assume that given a bounded set

$B \subset \mathbb{R}^n$ there exists a constant $K = K(B) > 0$ and a step size $h_c = h_c(B) > 0$ such that for all $y, \hat{y} \in B$ and $h \in [0, h_c)$ the map $\Phi_{h,f}(y_n)$ given by an integrator (or its underlying one-step method) is defined and,

1. $\|D\Phi_{h,f}(y)\| \leq 1 + Kh$,
2. $\|D\Phi_{h,f}(y) - D\Phi_{h,f}(\hat{y})\| \leq Kh\|y - \hat{y}\|$,
3. $\|\Phi_{h,f}(y) - \varphi_{h,f}(y)\| \leq Kh^{r+1}$,
4. $\|D\Phi_{h,f}(y) - D\varphi_{h,f}(y)\| \leq Kh^{r+1}$.

These assumptions are not particularly strong; they hold for all Runge-Kutta methods. For the underlying one-step methods of linear multistep methods they hold for $r = 1$ and for points within some compact set [137]. The restriction to trajectories within a compact set for linear multistep methods is not of consequence since the theorems which make use of this assumption also apply to trajectories within a compact set. The first three points of the assumption simply say that the map corresponding to the numerical integrator is close to the identity, Lipschitz and of order r. It is now possible to state the following two theorems which, together, say that hyperbolic fixed points of dynamical systems are generally preserved by numerical integrators, remaining close to those of the original system and retaining their (in)stability.

Theorem 2.4. *If y^* is a hyperbolic fixed point of the differential equation $\dot{y} = f(y)$ with flow $\varphi_{t,f}$, then there exists $\beta > 0$ and $t_c > 0$ such that*

$$\left\| (I - D\varphi_{t,f}(y^*))^{-1} \right\| \leq \frac{\beta}{t}, \forall t \in (0, t_c).$$

Theorem 2.5 (fixed points are preserved)**.** *Let y^* be a hyperbolic fixed point of $\dot{y} = f(y)$ with $f \in C^3$, then, under assumption 2.6, there exist constants $h_c, K > 0$ such that $\Phi_{h,f}(y)$ has a fixed point $\hat{y}$ which is unique in $B(y^*, 2K\beta h^r)$, the ball about y^* with radius $2K\beta h^r$, for all $h \in (0, h_c)$ and where β is the same as in theorem 2.4. Furthermore, $\hat{y}$ is stable/unstable if y^* is stable/unstable.*

One can give a more general theorem than 2.5, covering collections of fixed points of a differential equation. Its statement necessitates the following definitions:

Definition 2.7 (sets of fixed points)**.** For convenience we define the following sets:

- $\mathcal{U} := \{y \in \mathbb{R}^n : f(y) = 0\}$,
- $\mathcal{U}(\varepsilon) := \{y \in \mathbb{R}^n : \|f(y)\| \leq \varepsilon\}$,
- $\mathcal{U}_h := \{y \in \mathbb{R}^n : y = \Phi(y)\}$,
- $\mathcal{U}_h(\varepsilon) := \{y \in \mathbb{R}^n : \|y - \Phi(y)\| \leq h\varepsilon\}$.

Theorem 2.6. *Under the assumptions 2.6 and given a bounded set B there exist constants $K = K(B) > 0$ and $h_c = h_c(B) > 0$ such that for $h \in (0, h_c)$ the (semi-)group generated by the map $\Phi_{h,f}$ satisfies*

$$\mathcal{U}_h \cap B \subseteq \mathcal{U}(hK)$$

and, for any $\varepsilon > 0$

$$\mathcal{U}_h(\varepsilon) \cap B \subseteq \mathcal{U}(\varepsilon + hK).$$

That is, fixed points (and their neighbouring points) of the map $\Phi_{h,f}$ are close to fixed points of the flow $\varphi_{t,f}$ and converge towards them linearly in h.

Numerical methods may also introduce spurious fixed points which are not part of the original system. These spurious fixed points will have their own stable and unstable manifolds and consequently the dynamics of the numerical and continuous systems will differ in their vicinity. Runge-Kutta methods are known to introduce spurious fixed points in some cases, though two popular symplectic methods, the midpoint rule and the two-stage (fourth order) Gaussian Runge-Kutta method, do not. Nor does the (less popular, non-symplectic) explict Euler method $y_{n+1} = y_n + hf(y_n)$ since for any choice of step size $y_{n+1} = y_n \iff f(y_n) = 0$. The following example, however, shows how spurious fixed points *can* be created by Runge-Kutta methods

Example 2.1. The differential equation

$$\dot{y} = f(y) = \frac{-\lambda y}{1 + y^2}$$

has a single fixed point; $y = 0$. Applying the Runge-Kutta method

$$\begin{aligned} Y_1 &= y_n + \frac{1}{2} h f(y_n), \\ y_{n+1} &= y_n + h f(Y_1), \end{aligned}$$

yields

$$y_{n+1} = y_n - \frac{h\lambda \left(y_n - \frac{1}{2} \frac{h\lambda y_n}{1+y_n^2} \right)}{1 + \left(y_n - \frac{1}{2} \frac{h\lambda y_n}{1+y_n^2} \right)^2}.$$

Fixed points of the numerical trajectory occur when $h\lambda \left(y_n - \frac{1}{2} \frac{h\lambda y_n}{1+y_n^2} \right) = 0$, that is, for $y_n = 0$ and for $y_n^2 - h\lambda/2 + 1 = 0$. The first solution matches that of the differential equation while the quadratic gives rise to two spurious fixed points at $y_n = \pm\sqrt{h\lambda/2 - 1}$ whenever $h > 2/\lambda$.

By requiring a stronger assumption — assumption 2.8 below — which holds for Runge-Kutta methods [137], one can achieve the following theorem which states that in any bounded region, the fixed points of a differential equation and the numerical method which approximates it are the same:

Assumption 2.8. We assume that, in addition to assumption 2.6, the following holds:

$$\|\Phi_{h,f}(y) - \varphi_{h,f}(y)\| \leq K\|f(y)\|h^{r+1}.$$

Theorem 2.7. *Under the assumption 2.8 and given a bounded region B, there exist constants $K = K(B) > 0$ and $h_c = h_c(B) > 0$ such that for all $h \in (0, h_c)$, the fixed points of the map $\Phi_{h,f}$ satisfy*

$$\mathcal{U}_h \cap B = \mathcal{U} \cap B,$$

and, for $\varepsilon > 0$

$$\mathcal{U}_h(\varepsilon) \cap B = \mathcal{U}\left(\frac{\varepsilon}{1 - hK}\right).$$

A consequence of theorem 2.7 is the following:

Theorem 2.8 (spurious fixed points)**.** *Let $v : (0, h^*) \to \mathbb{R}^n$ be a continuous function satisfying $v(h) \in \mathcal{U}_h$ and $v(h) \notin \mathcal{U}, \forall h \in (0, h^*)$. That is, v defines a continuous branch of spurious fixed points. Then, under assumption 2.8:*

$$\|v(h)\| \to \infty \text{ as } h \to 0.$$

That is, the spurious fixed points move towards infinity as the step size approaches zero.

Linear multistep methods are always free from spurious fixed points, though they can introduce period-two or saw-tooth solutions; pairs of points which the method switches between at each step. (Runge-Kutta methods can also introduce these spurious period-two solutions, as can composition methods).

Composition or splitting methods can also introduce spurious fixed points — though the symplectic Euler and leapfrog methods are free of them. The general question of which splitting methods do not introduce spurious fixed points seems to be unstudied.

The article [91] gives conditions which must be satisfied in order that a first order splitting method, or the leapfrog method, be free of spurious fixed points. The conditions in the article are slightly ambiguous and though they seem to ensure that existing fixed points of a differential equation are preserved exactly by a splitting method, it is not clear that they prevent spurious fixed points being introduced — no proof is given. We therefore state our own theorems on the preservation of fixed points by splitting methods.

We are primarily concerned with the first order splitting method

$$\Phi_{h,f}(y_n) = e^{hf_k} e^{hf_{k-1}} \cdots e^{hf_1} y_n = e^{hf} y_n + \mathcal{O}\left(h^2\right) \tag{2.1}$$

for a splitting $f = \sum_i f_i$, where e^{hf_i} is the exact time-h flow of the component vector field f_i. If the components of the vector field are to be solved approximately, then the approximation must also preserve fixed points (for theorem 2.9) and introduce no spurious

fixed points (for theorem 2.10). The following notation is useful in the proofs

$$\begin{aligned} y_n^{[1]} &:= e^{hf_1} y_n, \\ y_n^{[2]} &:= e^{hf_2} y_n^{[1]} = e^{hf_2} e^{hf_1} y_n, \\ &\textit{et cetera.} \end{aligned}$$

Theorem 2.9. *Consider a differential equation $\dot{y} = f(y)$ with a fixed point y^*. If the splitting $f = \sum_i f_i$ satisfies*

$$f_i(y) = 0 \iff \sum_i f_i(y) = 0, \quad i = 1, \ldots, k, \tag{2.2}$$

then y^* is a fixed point of the splitting method (2.1).

Proof. Since $f_1(y^*) = 0$, by condition (2.2), we have $y_n^{[1]} = y^*$. Similarly, $f_2(y_n^{[1]}) = f_2(y^*) = 0$ so $y_n^{[2]} = y^*$. Continuing in this fashion we get

$$\Phi_{h,f}(y^*) = e^{hf_k} y^{[k-1]} = \ldots = e^{hf_k} e^{hf_{k-1}} \cdots e^{hf_1} y^* = y^*.$$

□

Theorem 2.9 says that fixed points of the differential equation are preserved exactly by first order splitting methods satisfying (2.2). In fact, since the condition (2.2) implies that the value of the vector field at $y_n^{[i]}$ is always zero if $y_n = y^*$, theorem 2.9 holds for all splitting methods of the form

$$\Phi_{h,f}(y_n) = e^{a_{m,k} hf_k} e^{a_{m,k-1} hf_{k-1}} \cdots e^{a_{m,1} hf_1} e^{a_{m-1,k} hf_k} e^{a_{m-1,k-1} hf_{k-1}} \cdots e^{a_{1,1} hf_1} y_n,$$

of *any* order since the introduction of the coefficients $a_{i,j}$ has no effect when the vector field is zero. Theorem 2.9 is not particularly strong since it says nothing about the presence of spurious fixed points introduced by the splitting method. The following theorem is stronger, but its assumptions on the splitting are more restrictive.

Theorem 2.10. *Consider a differential equation $\dot{y} = f(y)$ with a fixed point y^*, and a first order splitting method (2.1). Then, if the splitting $f = \sum_i f_i$ satisfies the condition (2.2), and if the components of the vector field splitting are orthogonal to each other in a neighbourhood of y^*, the method (2.1) preserves the fixed point y^* exactly and does not introduce spurious fixed points in the neighbourhood of y^* where the f_i are orthogonal.*

Theorem 2.9 gives the exact preservation of y^*. It remains for us to prove that spurious fixed points can not be introduced by the integrator. We show that this is the case for a three part splitting, but the proof generalizes automatically to splittings into any number of parts.

Proof. Consider a differential equation $\dot{y} = f(y)$ where y is partitioned as $y^\top = (y_1^\top, y_2^\top, y_3^\top)$, and the vector field is split accordingly. The requirement that the splitting of f be orthogonal means that each component of the splitting affects only one of y_1, y_2, y_3 (though it may depend on all of them), and that each of $\dot{y}_1, \dot{y}_2, \dot{y}_3$ is associated with a single component, f_i, of the splitting. Hence,

$$\dot{y} = \begin{bmatrix} \dot{y}_1 \\ \dot{y}_2 \\ \dot{y}_3 \end{bmatrix} = f = \begin{bmatrix} f_1(y_1, y_2, y_3) \\ f_2(y_1, y_2, y_3) \\ f_3(y_1, y_2, y_3) \end{bmatrix} = \begin{bmatrix} f_1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ f_2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ f_3 \end{bmatrix}.$$

The splitting method (2.1) is then

$$\Phi_{h,f}(y) = e^{hf_3} e^{hf_2} e^{hf_1} y.$$

(We drop the subscript indicating the step number of the integrator in order to avoid confusion with the partitioning of y into y_1, y_2, y_3.) Proceeding as for the proof of theorem 2.9; the condition (2.2) implies that $\dot{y}_1 = f_1(y) = 0 \iff f_1(y) = f_1(y_1^*, y_2^*, y_3^*)$, hence $y^{[1]} = e^{hf_1} y = y^* \iff y = y^*$. By the orthogonality of f_i, $\dot{y}_2 = f_2(y_1^{[1]}, y_2^{[1]}, y_3^{[1]}) = f_2(y_1^{[1]}, y_2, y_3)$ and so $\dot{y}_2 = 0 \iff y_1^{[1]} = y_1^*$ and $y_2 = y_2^*$, $y_3 = y_3^*$. The first of these equations holds from the previous step and hence $y^{[2]} = e^{hf_2} y^{[1]} = y^* \iff y = y^*$. Similarly for $\dot{y}_3$, orthogonality gives $\dot{y}_3 = f_3(y_1^{[2]}, y_2^{[2]}, y_3^{[2]}) = f_3(y_1^{[2]}, y_2^{[2]}, y_3)$ so condition (2.2) implies $\dot{y}_3 = 0 \iff y_1^{[2]} = y_1^*$, $y_2^{[2]} = y^*$ and $y_3 = y_3^*$. Again, the first two conditions hold from the preceding step and hence $y^{[3]} = e^{hf_3} y^{[2]} = y^* \iff y = y^*$. Therefore $\Phi_{h,f}$ can have no spurious fixed points in the region where the orthogonality condition holds. □

A similar proof to that of theorem 2.10 can be used to show that the fixed points of the leapfrog integrator are exactly those of the differential equation it is applied to since the leapfrog splitting is orthogonal and satisfies condition (2.2) for the differential equation $\dot{q} = f(p)$, $\dot{p} = g(q)$. (I.e. for the equations of motion of the separable Hamiltonian $H(q,p) = T(p) + V(q)$, with $f = \nabla T$ and $g = -\nabla V$.)

Theorem 2.11. *The leapfrog integrator*

$$\begin{aligned} q_{n+\frac{1}{2}} &= q_n + \frac{h}{2} f(p_n), \\ p_{n+1} &= p_n + h g(q_{n+\frac{1}{2}}), \\ q_{n+1} &= q_{n+\frac{1}{2}} + \frac{h}{2} f(p_{n+1}), \end{aligned}$$

has no spurious fixed points and preserves the fixed points of the differential equation exactly.

Proof. In order to prove that the leapfrog method preserves fixed points of the differential

equation exactly, assume that (q^*, p^*) is a fixed point of $\dot{q} = f(p)$, $\dot{p} = g(q)$, then

$$\begin{aligned} q_{n+\frac{1}{2}} =& q^*, \\ p_{n+1} =& p^* + hg(q_{n+\frac{1}{2}}) = p^*, \\ q_{n+1} =& q_{n+\frac{1}{2}} + \frac{h}{2} f(p_{n+1}) = q^* + \frac{h}{2} f(p^*) = q^*. \end{aligned}$$

To prove that the leapfrog method has no spurious fixed points we assume that the method has a fixed point at (q_n, p_n) and show that it is also a fixed point of the differential equation. Since p_n is a fixed point of leapfrog $p_{n+1} = p_n$ and

$$\begin{aligned} q_{n+1} &= q_{n+\frac{1}{2}} + \frac{h}{2} f(p_{n+1}) \\ &= q_{n+\frac{1}{2}} + \frac{h}{2} f(p_n) \\ &= q_n + \frac{h}{2} f(p_n) + \frac{h}{2} f(p_n) \\ &= q_n + hf(p_n). \end{aligned}$$

Since q_n was assumed to be a fixed point of the integrator $q_n = q_{n+1}$ and $f(p_n) = 0$. Similarly,

$$\begin{aligned} p_{n+1} &= p_n + hg(q_{n+\frac{1}{2}}) \\ &= p_n + hg(q_n + \frac{h}{2} f(p_n)) \\ &= p_n + hg(q_n). \end{aligned}$$

Since $p_n = p_{n+1}$ we must have $g(q_n) = 0$. Hence, q_n and p_n are fixed points of the differential equation and the leapfrog method has no spurious fixed points. □

It is not clear whether a similar result holds for generalizations of the leapfrog method, or for splitting methods in general. While theorem 2.10 and its proof are not difficult, we are not aware of it already existing in the literature.

Spectral Properties

We have seen already that the eigenvalues of a system linearized at a fixed point are important for determining the stability of the fixed point (cf. theorems 2.2 and 2.3). The linear eigenvalues also give information about bifurcations and the dimension of stable/centre/unstable manifolds in the neighbourhood of the fixed point.

The eigenvalues of the linearized system can be categorised as:

$$\text{the stable set: } \lambda^s = \{\lambda_i \ : \ \mathrm{Re}(\lambda_i) < 0\},$$

$$\text{the centre set: } \lambda^c = \{\lambda_i \ : \ \mathrm{Re}(\lambda_i) = 0\},$$

$$\text{the unstable set: } \lambda^u = \{\lambda_i \ : \ \mathrm{Re}(\lambda_i) > 0\}.$$

For a fixed point to be stable it must have an empty unstable set. Corresponding to each set there is a linear subspace which is tangent to the nonlinear system at the fixed point. These are the stable, centre and unstable manifolds and dominate the dynamics of the system in the vicinity of the fixed point — we discus these further in section 2.2.

Bifurcations of the full system typically occur when the eigenvalues enter or leave the centre set.

The map given by discretising the differential equation with an integrator has related sets of eigenvalues at the fixed point:

$$\text{the stable set: } \mu^s = \{\mu_i \ : \ |\mu_i| < 1\},$$

$$\text{the centre set: } \mu^c = \{\mu_i \ : \ |\mu_i| = 1\},$$

$$\text{the unstable set: } \mu^u = \{\mu_i \ : \ |\mu_i| > 1\}.$$

In order to capture the dynamics of the system near the fixed point we need the eigenvalues of the integrator to be in the same sets as the corresponding eigenvalues of the continuous system. That is, the stable/centre/unstable sets of the differential equation and of the map given by the numerical integrator should have the same dimension.

Time-symmetric, A-stable[2] Runge-Kutta methods have this property regardless of step size, and could therefore be viewed as geometric integrators with regard to this rather weak criterion. Such methods perturb each eigenvalue by the same function:

$$\lambda_i \mapsto \mu_i = R(h\lambda_i),$$

where R is the stability function of the Runge-Kutta method. Bifurcations are therefore preserved, with only eigenvalue resonances such as $\lambda_i/\lambda_j \in \mathbb{Q}$ not captured correctly. Hamiltonian systems and symplectic integrators are a special case. The eigenvalues of symplectic maps come in $\lambda, \bar{\lambda}, \lambda^{-1}, \bar{\lambda}^{-1}$ quadruplets. Preserving this property, as symplectic integrators must, is sufficient to preserve the stable/centre/unstable sets for generic fixed points with no double eigenvalues.

More generally, for small enough step sizes, geometric integrators perturb the coefficient matrix of the linearized flow to a matrix of the same class, preserving nearby dynamics of the system. This means that bifurcations will occur at parameter values close

[2] A numerical method is A-stable if its stability region includes the entire left-hand half of the complex plane [30]. For Runge–Kutta methods with coefficients as defined by 1.3 this means that the stability function $R(z) = 1 + zb^T(I - zA)^{-1}\vec{1} = \frac{\det(I - zA + z\vec{1}b^T)}{\det(I - zA)}$, where $z \in \mathbb{C}$ must satisfy $|R(z)| \leq 1$ for $\mathrm{Re}(z) < 0$. ($\vec{1}$ is a vector whose entries are all 1.)

to those of the original system. Exceptions are situations such as where the eigenvalues are degenerate in which case the integrator may introduce a spurious bifurcation. By the same reasoning, non-geometric methods, typically do not preserve spectral properties, regardless of step size, since they may alter the class of the matrix of the linearized flow, for example a system with symplectic flow may be perturbed to a dissipative system. Integrators which are designed specifically to preserve linearization (and which are self-adjoint and/or symplectic) are presented by McLachlan, Quispel and Tse in [92].

2.2 Stable & unstable manifolds

Definition 2.9 (local/global unstable manifold)**.** The *unstable manifold* of a fixed point y^* of $\dot{y} = f(y)$ is the set

$$W^u(y^*) := \{y \in \mathbb{R}^n : y(t) \to y^* \text{ as } t \to -\infty\}.$$

The *local unstable manifold* of y^* is the set

$$W^u_\varepsilon(y^*) := \{y \in W^u(y^*) : \|y(t) - y^*\| \leq \varepsilon, \forall t \leq 0\}.$$

Definition 2.10 (local/global stable manifold)**.** The *stable manifold* of a fixed point y^* of $\dot{y} = f(y)$ is the set

$$W^s(y^*) := \{y \in \mathbb{R}^n : y(t) \to y^* \text{ as } t \to \infty\}.$$

The *local stable manifold* of y^* is the set

$$W^s_\varepsilon(y^*) := \{y \in W^s(y^*) : \|y(t) - y^*\| \leq \varepsilon, \forall t \geq 0\}.$$

The stable and unstable manifolds of a differential equation are tangent to those of the linearized system at a fixed point, and are well approximated by them in a neighbourhood around it. Hence, the first step towards preserving the stable and unstable manifolds is to preserve the spectral properties of the system. Away from the fixed point, the numerically computed manifolds are somewhat at the mercy of the local error of the numerical integrator as the following example shows.

Example 2.2. We consider the two dimensional system

$$\dot{q} = -q + p^2, \quad \dot{p} = p, \tag{2.3}$$

which has a fixed point at $(q, p) = (0, 0)$. Defining $z = q - p^2/3$ gives $\dot{z} = \dot{q} - \frac{2}{3}p\dot{p} =$

$-q + p^2/3 = -z$. Hence, if $z(0) = 0$ then $z(t) = 0$ for all time, and the curve

$$q = \frac{p^2}{3}$$

is invariant under the flow of the original differential equation. Solving the original equations gives

$$q(t) = \frac{A^2}{3}e^{2t}, \quad p(t) = Ae^t,$$

and shows that $(q, p) \to (0, 0)$ as $t \to -\infty$. That is, the invariant curve is the unstable manifold of the fixed point.

If the explicit Euler method is applied to (2.3) we get

$$\begin{bmatrix} q_{n+1} \\ p_{n+1} \end{bmatrix} = \begin{bmatrix} q_n - hq_n + hp_n^2 \\ p_n = p_n + hp_n \end{bmatrix} = \begin{bmatrix} (1-h)q_n + hp_n^2 \\ (1+h)p_n \end{bmatrix} \tag{2.4}$$

Does the system (2.4) have the same, or at least a similar, unstable manifold as the original system (2.3)? That is, can we find a coefficient $a \in \mathbb{R}$ such that

$$q_n = ap_n^2 \iff q_{n+1} = ap_{n+1}^2?$$

Inserting $q_n = ap_n^2$ in (2.4) gives the following necessary condition

$$q_{n+1} = (1-h)ap_n^2 + hp_n^2 = ap_{n+1}^2 = a(1+h)^2 p_n.$$

This is satisfied when $(1-h)a + h = a(1 + 2h + h^2)$, that is, when $a = 1/(3+h)$ and

$$q_n = \frac{p_n^2}{3+h}.$$

As $h \to 0$ the numerical invariant set tends towards that of the differential equation but only at a speed of $\mathcal{O}(h)$. The numerical trajectories in this invariant set are given by

$$q_n = \frac{1}{3+h}(1+h)^{2n}p_0^2, \quad p_n = (1+h)^n p_0,$$

hence, as $n \to -\infty$ $(q_n, p_n) \to (0, 0)$ and the invariant set retains the stability of the original system.

The following theorem says that under certain conditions, any local unstable manifold about the fixed point of a differential equation is close to a local unstable manifold about a fixed point of the numerical approximation and that the manifolds converge with the same order as the numerical integrator. The converse also holds; all local unstable manifolds of a numerical approximation converge to local unstable manifolds of the original differential equation. The theorem makes use of the following definition:

Definition 2.11 (distance between sets)**.** The distance between two sets, A and B is

$$\operatorname{dist}(B, A) := \sup_{u \in B} \operatorname{dist}(u, A),$$

and where

$$\operatorname{dist}(u, A) := \inf_{v \in A} \|u - v\|.$$

Theorem 2.12 (preservation of local unstable manifolds)**.** *Consider an approximation of the differential equation $\dot{y} = f(y)$ by a map $\Phi_{h,f}(y_n) = y_{n+1}$ generated by a numerical integrator (or its underlying one-step method) and satisfying assumption 2.6. Let $W^u_\varepsilon(y^*)$ denote the local unstable manifold of a hyperbolic equilibrium point y^* of f and let $W^u_{\varepsilon,h}(\hat{y})$ denote the local unstable manifold of the fixed point $\hat{y}$ of the map $\Phi_{h,f}$. (Theorem 2.5 ensures that such a fixed point exists.) Then, there exist constants $C, h_c, \varepsilon_c > 0$ such that for any $\varepsilon \in (0, \varepsilon_c)$ and for each $y \in W^u_\varepsilon(y^*)$ there exists $\varepsilon' > 0$ and $y_n \in W^u_{\varepsilon',h}(\hat{y})$ such that*

$$\|y - y_n\| \leq Ch^r, \quad \forall h \in (0, h_c),$$

and hence

$$\operatorname{dist}(W^u_\varepsilon(y^*), W^u_{\varepsilon',h}(\hat{y})) \leq Ch^r, \quad \forall h \in (0, h_c).$$

Furthermore, for any $y_n \in W^u_{\varepsilon,h}(\hat{y})$ there is a $y \in W^u_{\varepsilon'}(y^*)$ such that

$$\|y - y_n\| \leq Ch^r, \quad \forall h \in (0, h_c),$$

and hence

$$\operatorname{dist}(W^u_{\varepsilon,h}(\hat{y}), W^u_{\varepsilon'}(y^*)) \leq Ch^r, \quad \forall h \in (0, h_c).$$

An equivalent theorem also applies for local stable manifolds [137, chapter 6]. Since Runge-Kutta methods and first order linear multistep methods satisfy the necessary assumptions of the theorem, these methods preserve local stable and unstable manifolds.

2.3 Periodic orbits

Definition 2.12 (periodic orbit of a differential equation)**.** A *periodic orbit* of a differential equation $\dot{y} = f(y)$ is a continuous, closed curve $\gamma(t)$ satisfying $\dot{\gamma} = f(\gamma(t))$ and $\gamma(t) = \gamma(t+T)$ for all $t \in \mathbb{R}$ and some $T > 0$. The minimal value of T is the period of the periodic orbit.

Unlike for the case of fixed points, periodic orbits of differential equations and their numerical approximations can be qualitatively different. If a numerical integrator follows a periodic orbit with a step size h such that $T/h \in \mathbb{R} \setminus \mathbb{Q}$ then the numerical trajectory describes an invariant closed curve like that of the differential equation. However, since the trajectory never returns to its initial conditions it does not satisfy the obvious criterion

for a trajectory to be periodic. In fact such orbits mimic the behaviour of *quasi-periodic orbits* which are discussed in the following section. On the other hand, if the period T is a rational multiple of h then the numerical trajectory is a finite set of discrete points. It is this invariant set of discrete points which we identify as a periodic orbit of the map:

Definition 2.13 (periodic orbit of a map)**.** A point y^* is a *period-s point* of the map $\Phi(y_n) = y_{n+1}$ if $\Phi^s(y^*) = y^*$ and $\Phi^r(y^*) \neq y^*$ for $0 < r < s$, $r, s, \in \mathbb{N}$. If y^* is a period-s point then the invariant set $\{y^*, \Phi(y^*), \Phi^2(y^*), \ldots, \Phi^{s-1}(y^*)\}$ is a *period-s orbit* of Φ.

Changing the initial condition along the original periodic orbit may give rise to a continuous family of periodic orbits for the discrete system. Despite this qualitative discrepancy between periodic orbits of flows and maps, properties of the continuous case have an analogous version for the discrete case, and *vice versa*.

Definition 2.14 ((asymptotically) stable periodic orbit of a differential equation)**.** A periodic orbit $\gamma(t)$ of the differential equation $\dot{y} = f(y)$ is stable if, for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ and $\tau = \tau(\varepsilon)$ such that $\|y(0) - \gamma(0)\| \leq \delta$ implies $\|y(t) - \gamma(t+\tau)\| \leq \varepsilon, \forall t \geq 0$. If a stable periodic orbit satisfies $\|y(t) - \gamma(t+\tau)\| \to 0$ as $t \to \infty$ then $\gamma(t)$ is asymptotically stable.

Definition 2.15 (stable periodic orbit of a map)**.** A period-s orbit of the map Φ is (asymptotically) stable if each period-s point of the orbit is (asymptotically) stable under the map Φ^s.

Definition 2.16 (Floquet, or characteristic multipliers)**.** Consider a periodic orbit $\gamma(t)$, with period T, of a vector field with flow $\varphi_t(y)$. For $y \in \gamma$ the eigenvalues of the linearization $D\varphi_T(y)$ (i.e. the Jacobian of the time-T flow map) are called the *Floquet multipliers* or *characteristic multipliers* of γ. They are independent of the choice of $y \in \gamma(t)$. The matrix $L(T) = D\varphi_T(y)$ is known as the *monodromy matrix*.

Similar to the case for fixed points, the definition of stability for periodic orbits of differential equations is descriptive rather than providing a way that the stability of an orbit can be tested and determined. The following theorem provides such a test.

Theorem 2.13 (stability via characteristic multipliers)**.** *If* $\mu_n = 1$ *and* $|\mu_i| < 1, i = 1, \ldots, n-1$ *then the periodic orbit* $\gamma(t)$ *is asymptotically stable.*

Periodic orbits always have at least one characteristic multiplier which is 1 with an eigenvector $f(y)$. If 1 is a simple eigenvalue then the periodic orbit perturbs smoothly with f. Such orbits are called *regular* or *hyperbolic*.

Definition 2.17 (hyperbolic periodic orbit)**.** Let $\{\mu_i\}_{i=1}^n$ denote the characteristic multipliers of $Df(\gamma(t))|_T$ where $\gamma(t)$ is a periodic orbit of $\dot{y} = f(y)$ and the μ_i are ordered such that $|\mu_i| \leq |\mu_{i+1}|, i = 1, \ldots, n-1$, then the periodic orbit is said to be hyperbolic if $\mu_j = 1$ for a single fixed $1 \leq j \leq n$ and $|\mu_i| \neq 1$ for $i \neq j$.

The definition 2.15 implies that it is necessary to test the stability of each periodic point in order to determine whether a periodic orbit of a map is stable. The following theorem tells us that this is not the case.

Theorem 2.14 (for maps, one stable periodic point implies another)**.** *Let y^* be a point on a period-s orbit of a differentiable map Φ, then y^* is a hyperbolic, stable (resp. unstable) fixed point of Φ^s if, and only if, the period-s orbit is hyperbolic and stable (resp. unstable).*

Stable, hyperbolic periodic orbits are preserved by Runge-Kutta methods, or by first order linear multistep methods since these methods satisfy assumption 2.6, a prerequisite for theorem 2.15 below.

Definition 2.18 (Hausdorff distance)**.** The *Hausdorff distance* between two sets A and B is

$$d_H(A, B) := \max\left(\operatorname{dist}(A, B), \operatorname{dist}(B, A)\right),$$

where $(\operatorname{dist}(A, B), \operatorname{dist}(B, A))$ is given by definition 2.11.

Theorem 2.15 (stable hyperbolic periodic orbits are preserved)**.** *Assume that $\dot{y} = f(y)$ has a hyperbolic, stable periodic orbit $\gamma(t)$ comprising the set of points $\mathcal{P}$. Then, under assumption 2.6, the map $\Phi_{h,f}(y_n) = y_{n+1}$ has a closed invariant curve comprising the set of points $\mathcal{P}_h$ and there exists a constant $c > 0$ such that*

$$d_H(\mathcal{P}, \mathcal{P}_h) \leq ch^r.$$

The invariant circle $\mathcal{P}_h$ may, but need not, be comprised of periodic points.

The theorem says nothing about whether the frequency of the periodic orbit of the numerical method matches that of the differential equation. However, the definition of asymptotic stability for periodic orbits requires that trajectories near the periodic orbit circle the orbit with a frequency that approaches that of the periodic orbit. Numerical approximations to these trajectories behave similarly; therefore one can expect the frequency of the numerical periodic orbit to approach that of the original system with decreasing step size. One problem related to this is that if the period of the orbit is not an integer multiple of the step size, the numerical solution may circle the orbit multiple times before returning to its initial condition. This problem arises because periodic orbits are degenerate in the sense that they are unchanged by time translation: $\gamma(t) = \gamma(t+c) \quad \forall c \in \mathbb{R}$. One can remove this degeneracy by taking a Poincaré cross-section. This involves restricting the initial conditions to lie in a section Σ and defining a return map $\sigma : \Sigma \to \Sigma$ which maps points y^* in Σ to the next point on the trajectory $\varphi_{t,f}(y^*)$ (or $\Phi_{nh,f}(y^*)$) lying in Σ. Applying such an approach to the numerical orbit allows one to estimate the period of the original solution. The map $\Phi_{nh,f}$ given by a numerical method generally does not preserve the section Σ exactly. This issue is discussed further in section 4.6.

The following example illustrates the preservation of a periodic orbit as the limit set of numerical trajectories. It also raises the spectre of spurious periodic orbits.

Example 2.3. Consider the complex-valued differential equation

$$\dot{z} = (\alpha i + 1 - |z|^2)z. \tag{2.5}$$

This equation has a periodic solution

$$z(t) = e^{\alpha i t}$$

which corresponds to the set $\mathcal{P} := \{z \in \mathbb{C} : |z| = 1\}$. For $|z| > 1$, solutions collapse onto $\mathcal{P}$, and for $|z| < 1$ they expand towards $\mathcal{P}$.

The explicit Euler method applied to (2.5) yields

$$z_{n+1} = z_n + h(\alpha i + a - |z_n|^2)z_n. \tag{2.6}$$

We look for an invariant circle of points of (2.6) satisfying $|z_{n+1}|^2 = |z_n|^2$. These are fixed points of the map $|z_n|^2 \mapsto |z_{n+1}|^2$, that is, solutions of

$$|z_{n+1}|^2 = \left|z_n + h(\alpha i + 1 - |z_n|^2)z_n\right|^2 = |z_n|^2.$$

This holds when

$$\begin{aligned} 1 &= \left|1 + h\alpha i + h - h|z_n|^2\right|^2 \\ &= \left(1 + h - h|z_n|^2\right)^2 + h^2\alpha^2, \end{aligned}$$

hence $|z_n|^2 = R_\pm(\alpha)^2$, where

$$R_\pm(\alpha)^2 := 1 + \frac{1 \pm \sqrt{1 - \alpha^2 h^2}}{h}. \tag{2.7}$$

Thus, there is a circle of fixed points for $|z_n| = R_-(\alpha)$ and the mapping (2.6) has an invariant set

$$\mathcal{P}_h := \{z \in \mathbb{C} : |z| = R_-(\alpha)\}.$$

A series expansion shows $R_-(\alpha) = 1 + \mathcal{O}(h)$ and hence $\mathcal{P} \to \mathcal{P}_h$ as $h \to 0$. The system also has a spurious limit set given by the invariant circle $|z_n| = R_+(\alpha)$ which grows like $\mathcal{O}(1/h)$ as $h \to 0$.

Analysis such as that in the example above assumes that the periodic solutions are isolated structures. This holds for hyperbolic systems, but for Hamiltonian systems the periodic orbits occur in one-parameter families (cf. chapter 4). The following illustrates

why isolated orbits are necessary for the type of analysis used above, and motivates the work in chapter 4 of this thesis where KAM theory is used to show that under certain assumptions periodic orbits of Hamiltonian systems are in fact preserved.

Example 2.4. Consider the differential equation

$$\dot{q} = -p, \quad \dot{p} = q,$$

with periodic solution given by

$$q = A\cos(t), \quad p = A\sin(t), \quad A \in \mathbb{R}.$$

Since the parameter A is arbitrary, the phase portrait is full of invariant circles.

Now consider the map given by the explicit Euler method:

$$q_{n+1} = q_n - hp_n, \quad p_{n+1} = p_n + hq_n,$$

whose trajectories satisfy

$$q_{n+1}^2 + p_{n+1}^2 = (1+h^2)^{n+1}(q_0^2 + p_0^2).$$

As $n \to \infty$, $q_{n+1}^2 + p_{n+1}^2 \to \infty$ for all choices of h and all initial conditions other than the fixed point $q_0 = 0$, $p_0 = 0$. Hence, there can be no closed curves approximating a periodic orbit.

2.4 Quasi-periodic orbits & invariant tori

Quasi-periodic orbits and invariant tori are fundamental objects of symplectic flows (see section 1.3.2 for a definition and introduction, and chapters 3 and 4 for more details) but they can also arise in dissipative systems. In the case of flows for which a Poincaré section is defined globally, the winding number or rotation number of the return map is a unique invariant quantity of the torus. If the winding number is rational then the flow on the torus is a periodic orbit, if it is irrational then all trajectories on the torus are dense and quasi-periodic. We will mostly leave details of invariant tori and quasi-periodic orbits to chapters 3 and 4.

Similar to the case of periodic orbits, numerical integrators can both approximate existing quasi-periodic orbits and introduce spurious quasi-periodic orbits, as the following example shows. Quasi-periodic orbits can be destroyed by numerical integrators when the integrator step size is resonant with one of the frequencies of the orbit.

Example 2.5. Consider the complex, two dimensional system of differential equations:

$$\dot{z} = (i + 1 - |w|^2)z, \quad \dot{w} = (\sqrt{2}i + 1 - |z|^2)w, \quad z, w, \in \mathbb{C}. \tag{2.8}$$

This admits the solution

$$z = e^{it}, \quad w = e^{\sqrt{2}it},$$

which in turn defines the manifold

$$\mathcal{Q} := \{z, w \in \mathbb{C} : |z| = |w| = 1\}.$$

As previously, we discretize the system with the explicit Euler method

$$\begin{aligned} z_{n+1} &= z_n + h(i + 1 - |w_n|^2)z_n, \\ w_{n+1} &= w_n + h(\sqrt{2}i + 1 - |z_n|^2)w_n, \end{aligned}$$

and look for an invariant set satisfying $|z_{n+1}| = |z_n|$ and $|w_{n+1}| = |w_n|$ for all $n > 0$.

Similarly to example 2.3 we require $|z_{n+1}|^2 = \left|z_n\left(1 + h(i + 1 - |w_n|^2)\right)\right|^2 = |z_n|^2$ which gives

$$1 = h^2 + \left(1 + h - h|w_n|^2\right)^2.$$

The analogous calculation for $|w_{n+1}|^2 = |w_n|^2$ gives

$$1 = 2h^2 + \left(1 + h - h|z_n|^2\right)^2.$$

Hence, the invariant set satisfies

$$|w_n|^2 = 1 + \frac{1 \pm \sqrt{1 - h^2}}{h}, \qquad |z_n|^2 = 1 + \frac{1 \pm \sqrt{1 - 2h^2}}{h},$$

or

$$|w_n| = R_-(1), \qquad |z_n| = R_-(\sqrt{2}),$$

where $R_-(\alpha)$ is the same as that defined by (2.7). The same analysis as in section 2.3 shows that the invariant set $\mathcal{Q}_h := \{z, w \in \mathbb{C} : |z| = R_-(\sqrt{2}), |w| = R_-(1)\}$ converges to the set $\mathcal{Q}$ like $\mathcal{O}(h)$ as $h \to 0$ and that the solution $R_+(\alpha)$ gives rise to a spurious invariant set. Again, the analysis relies on the invariant set being isolated, a condition which, in general, does not hold for Hamiltonian systems. The analysis also does not say anything about the preservation of the frequencies of the quasi-periodic orbits — we can only resort to the accuracy of the numerical integrator to ensure any sort of fidelity in this respect.

2.5 Chaotic invariant sets

The term chaos is frequently applied to any dynamics whose behaviour is more irregular than quasi-periodic motion. While chaos is one of the most interesting properties of dynamical systems; it is also one of the more difficult to compute numerically. In chaotic

regions, arbitrarily close orbits diverge exponentially quickly making the solutions unpredictable over long times [49]. Since invariant tori can act as partial boundaries between chaotic and non-chaotic regions in phase space their preservation is important for correctly preserving chaotic dynamics, (cf. section 1.3.3).

A necessary and sufficient condition for the existence of chaos for diffeomorphisms is that there are transverse intersections of stable and unstable manifolds of a periodic orbit. This is the Smale-Birkhoff homoclinic theorem [49, chapter 5]. The theorem also holds for the return map of flows.

If it is possible to calculate the stable and unstable manifolds of a system then the Smale-Birkhoff theorem can be used to verify the existence of chaos. The Melnikov method [49, chapter 4] gives criteria whereby transverse intersections can occur when a system with non-transverse homoclinic orbits of periodic orbits is perturbed. However, the numerical verification of these intersections can be difficult since the angles between the stable and unstable manifolds at the intersection may be small "beyond all orders" [49].

Conversely, discrete dynamical systems can display chaos when the underlying continuous system has none. A particularly good example of such spurious chaos is described in the article by Ablowitz, Herbst and Schober [1] where chaos was observed in numerical discretizations of the nonlinear Schrödinger equation and the sine-Gordon equation even though both systems are entirely integrable (and hence their motion must be regular, not chaotic).

Another illustration of spurious chaos introduced by numerical discretization is shown in figure 2.1 where a magnified view of the phase portrait of the pendulum is shown in the vicinity of the hyperbolic fixed point at $(-\pi, 0)$. The figure shows the numerical trajectories of two different initial conditions both calculated by a symplectic integrator — the leapfrog method. One trajectory begins close to the hyperbolic fixed point and is chaotic. It follows the unstable manifold away from the hyperbolic fixed point, returning on the stable manifold resulting in a dimension-zero transverse intersection of the stable and unstable manifolds giving rise to chaos in the numerical trajectory even though the pendulum system is integrable. Another trajectory begins further away from the fixed point and displays the correct, regular, dynamics.

Such behaviour can be expected for Runge-Kutta, linear multistep and composition methods, even for geometric integrators. Of course, geometric integrators will be at an advantage over non-geometric integrators in such cases since any structures they preserve, such as invariant tori, will constrain (or slow) the spread of spurious chaotic trajectories.

In contrast to the spurious chaos which can be created by numerical integrators applied to integrable systems, the chaotic dynamics of the Lorenz attractor can be captured even by the explicit Euler method. The Lorenz system is given by three coupled, nonlinear

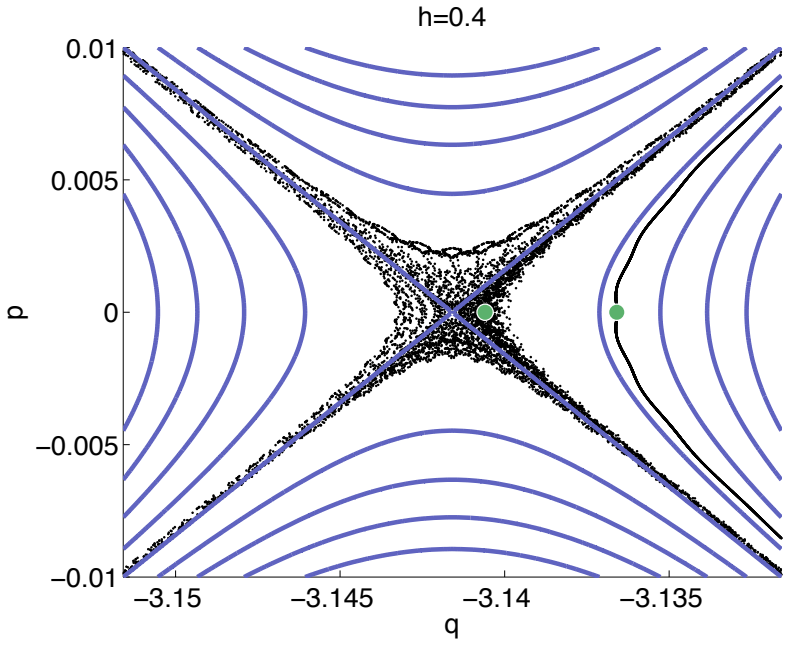


Figure 2.1: Spurious chaos near the hyperbolic fixed point of the pendulum. Initial conditions are indicated, by large green dots, for two numerical trajectories. The trajectory which begins close to the hyperbolic fixed point is chaotic, while the trajectory with initial condition slightly further away from the fixed point displays the correct, regular behaviour. The solid blue lines show the exact phase portrait of the system.

equations:

$$\begin{aligned}\dot{x} &= \sigma(y - x),\\ \dot{y} &= \rho x - y - xz,\\ \dot{z} &= xy - \beta z.\end{aligned}$$

The system is chaotic for the parameters $(\sigma, \rho, \beta) = (10, 28, \frac{8}{3})$ and the phase portrait is dominated by a *strange attractor* [137, section 2.5]. Trajectories within the attractor spend time in two distinct regions of the phase space. Figure 2.2 shows a single numerical trajectory of the Lorenz system calculated with the explicit Euler method and shown for various projections.

Perhaps more challenging to numerical integrators than the previous two examples are systems which exhibit both chaotic and regular dynamics within a given region. The Hénon-Heiles system is one such system. It has a separable Hamiltonian:

$$H(q,p) = T(p) + V(q), \quad q, p \in \mathbb{R}^2, \quad q^\top = (q_1, q_2), \quad p^\top = (p_1, p_2),$$

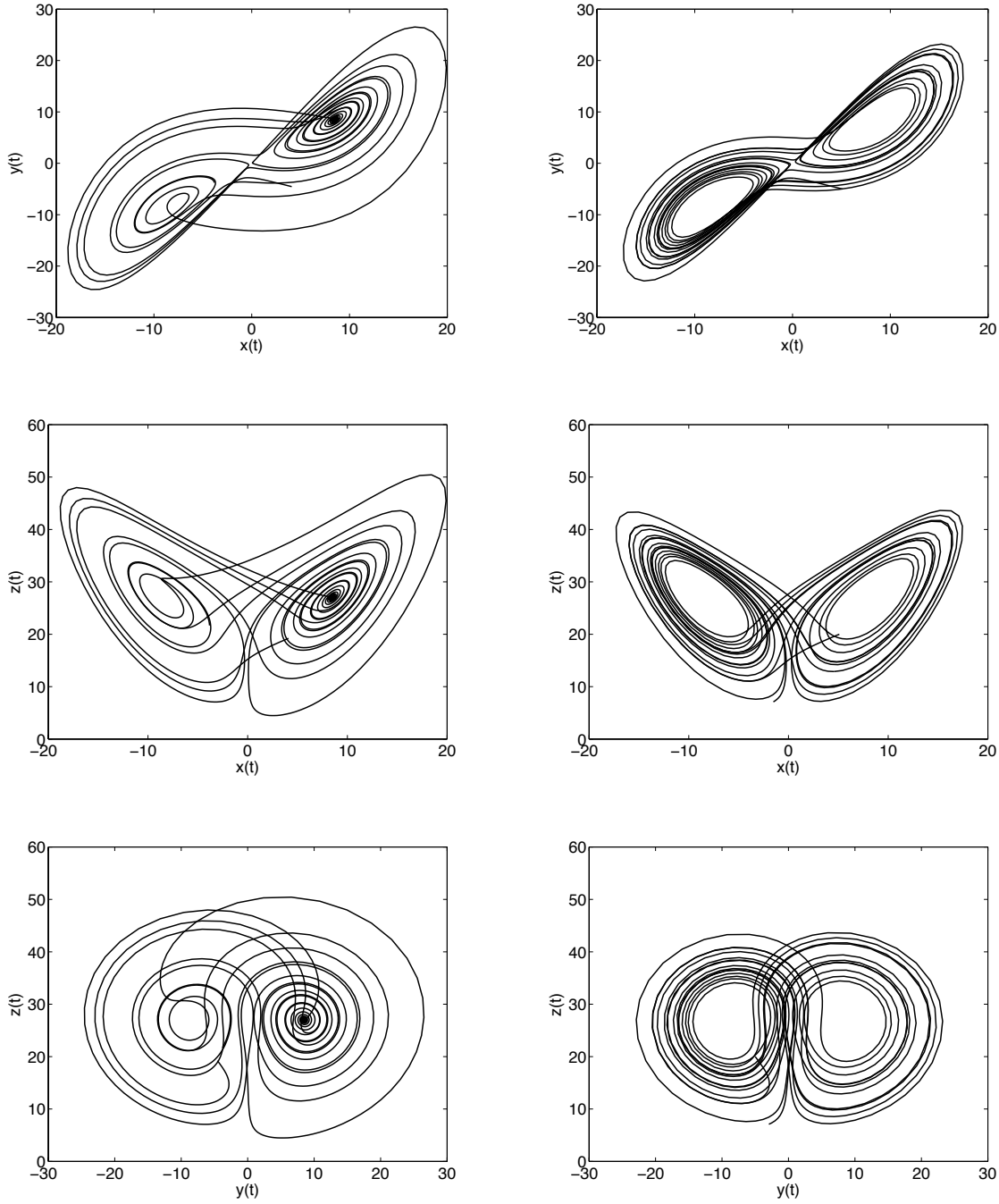


Figure 2.2: Chaos in the Lorenz system for a numerical integrator calculated with the explicit Euler method (*left*) for a step size of $h = 0.01$. While the individual points on the trajectory at any specific time can (and do) differ greatly from the corresponding point of the exact trajectory, the overall behaviour of the attractor for the explicit Euler method is close to that of the reference solution (*right*). The reference solution was computed with `ode45` [128] — an explicit, fifth order Runge-Kutta method with error estimation and variable step size control.

with

$$T(p) = \frac{1}{2}\|p\|^2 \quad \text{and} \quad V(q) = \frac{1}{2}\|q\|^2 + q_1 q_2^2 - \frac{1}{3}q_1^3$$

which makes it amenable to integration by a splitting method such as leapfrog. A Poincaré section for numerical trajectories of the system calculated with leapfrog, using $h = 0.1$, is shown in figure 2.3. Initial conditions are all chosen from the energy surface $H = 0.1$.

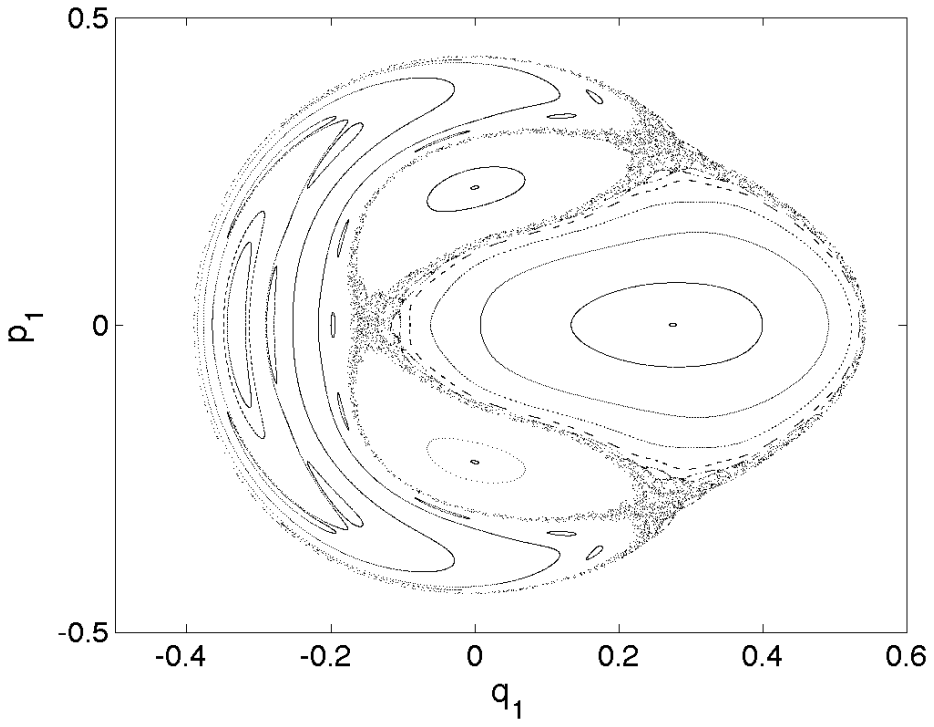


Figure 2.3: Poincaré section for numerical trajectories of the Hénon-Heiles system calculated with leapfrog (for a step size of $h = 0.1$). The section shows both regular and chaotic motion. The closed curves are sections of invariant tori. All initial conditions were taken from the energy surface with $H = 0.1$.

For energies below $H = \frac{1}{6}$ all trajectories of the system starting within a neighbourhood of the origin are bounded. Within this bounded region there is both regular motion, such as periodic orbits and invariant tori, and chaotic trajectories. For such systems geometric integrators such as symplectic methods come to the fore. In order to capture the dynamics of such a system an integrator needs to reproduce the features responsible for the regular motion (e.g. preserve invariant tori and periodic orbits) and to retain the statistical properties such as time averages of the chaotic regions. Non-geometric methods will not, in general, preserve the structures associated with the regular dynamics, or the statistical properties.

Chapter 3

KAM Theory

KAM tori are very sticky...

— A. D. Perry and S. Wiggins, 1994

This chapter covers the classical version of KAM theory for Hamiltonian systems. It begins with a description of conditionally periodic flow and the Arnold-Liouville theorem for completely integrable Hamiltonian systems. Lindstedt-Poincaré series are used to explain classical perturbation theory before a description of Kolmogorov's iteration is given. Finally Kolmogorov's theorem on the persistence of invariant tori is presented and proved. Results are mostly taken from the well known text [55] by Hairer, Lubich and Wanner. Publications by Arnold [4] and de la Llave [31] are also used.

3.1 Integrable Hamiltonian systems

Integrable Hamiltonian systems are interesting since their equations of motion can be solved analytically, however, it is not always obvious from inspection which systems are integrable and which are not. If a system is known to be integrable then the behaviour of nearby systems can often be studied by treating the new system as a perturbation of the integrable system and writing it as such using a power series in some small parameter.

If a Hamiltonian system

$$\dot{q} = \frac{\partial H}{\partial p}(q,p), \quad \dot{p} = -\frac{\partial H}{\partial q}(q,p), \tag{3.1}$$

with d degrees of freedom, can be transformed to the form

$$H(q,p) = \widehat{H}(I), \tag{3.2}$$

by a symplectic coordinate transformation $(q,p) \mapsto (\theta, I)$ then the flow of the transformed system is given by

$$\dot{I} = 0, \quad \dot{\theta} = \omega(I). \tag{3.3}$$

The flow of this system is simply uniform motion on a Cartesian product of straight lines and circles and is easily integrated to give

$$I = I(t_0), \quad \theta = \theta(t_0) + \omega(t_0) \cdot t.$$

The goal is therefore to construct a symplectic transformation $(q, p) \mapsto (\theta, I)$ so that the uniform motion can simply be read off from the Hamiltonian in the new coordinates. The transformation between neighbourhoods of (q_0, p_0) and (θ_0, I_0) can be constructed, (one hopes), via a *generating function* $S(q, I)$, according to

$$\theta = \frac{\partial S}{\partial I}(q, I), \tag{3.4}$$

$$p = \frac{\partial S}{\partial q}(q, I), \tag{3.5}$$

assuming that (3.4) and (3.5) hold for (q_0, p_0) and (θ_0, I_0) and that $S(q, I)$ has an invertible Hessian matrix at (q_0, I_0).

Equation (3.5) togther with (3.2) gives the Hamilton-Jacobi partial differential equation for S

$$H\left(q, \frac{\partial S}{\partial q}(q, I)\right) = \widehat{H}(I). \tag{3.6}$$

If $S(q, I)$ is a solution of (3.6) for some function $\widehat{H}(I)$ then (3.5) defines (implicity) d functions $I_i = F_i(q, p)$, $i = 1, \ldots, d$ which are all first integrals of the system (3.1) and are in involution: their Poisson brackets vanish pairwise $\{F_i, F_j\} = 0$, $i, j = 1, \ldots, d$. The converse result, due to Bour [13] and Liouville [83] — that a system with d first integrals in involution can be transformed to the form (3.3) — is known as Liouville's Lemma.

Lemma 3.1 (Liouville's Lemma)**.** *Let $F_1, \ldots, F_d$ be smooth, real-valued functions defined in a neighbourhood of $(q_0, p_0) \in \mathbb{R}^d \times \mathbb{R}^d$ and suppose that the Poisson brackets of all of these functions vanish pairwise, and that their gradients are linearly independent at (q_0, p_0). Then, there exist d additional smooth functions, $G_1, \ldots, G_d$ defined on a neighbourhood of (q_0, p_0), such that*

$$(F_1, \ldots, F_d, G_1, \ldots, G_d) : (q, p) \mapsto (\theta, I)$$

is a symplectic transformation which transforms $H(q, p)$ locally to $\widehat{H}(I)$.

This gives a sort of *local* integrability result. However, since it holds only in a neighbourhood of the initial conditions it is not much use for studying the dynamics of a system, hence one needs a stronger version of integrability.

Definition 3.1 (Completely Integrable Hamiltonian)**.** Let M be an open subset of $\mathbb{R}^d \times \mathbb{R}^d$. The Hamiltonian system $H : M \to \mathbb{R}$ is said to be *completely integrable* if there exist d smooth functions $F_1 = H, F_2, \ldots, F_d : M \to \mathbb{R}$ satisfying the following three properties:

1. $F_1, \ldots, F_d$ are in involution. (I.e. their Poisson brackets vanish pairwise.)

2. The gradients of $F_1, \ldots, F_d$ are linearly independent at every point in M.

3. The trajectories of the system given by $(\dot{q}^\top, \dot{p}^\top)^\top = J^{-1}\nabla H(q,p)$ exist for all times and remain in M.

Definition 3.2. ($d-$dimensional torus) We denote the $d-$dimensional torus

$$\mathbb{T}^d := \mathbb{R}^d/2\pi\mathbb{Z}^d = \{(\theta_1 \bmod 2\pi, \ldots, \theta_d \bmod 2\pi) : \theta_i \in \mathbb{R}\}.$$

The Arnold-Liouville theorem states that for completely integrable systems there is a *global* symplectic change of coordinates, (θ, I), the system has d first integrals which are independent of θ and the phase space of the system is filled with invariant tori parameterized by I. The (θ, I) coordinates are known as *angle-action* coordinates; $\theta_1 \bmod 2\pi, \ldots, \theta_d \bmod 2\pi$ being the angles and $I_1, \ldots, I_d$ being the actions.

Theorem 3.2 (Arnold-Liouville theorem, [55])**.** *Let $F_1, \ldots, F_d : M \to \mathbb{R}$ be the first integrals of a completely integrable system, as per definition 3.1. Suppose that the level sets*

$$M_x := \{(q,p) \in M : F_i(q,p) = x_i, \quad i = 1, \ldots, d \text{ for } x \in \mathbb{R}^d\}$$

are compact and connected for all x in a neighbourhood of $x_0 \in \mathbb{R}^d$. Then, there are neighbourhoods, B, of x_0 and, D, of $0 \in \mathbb{R}^d$ such that the following holds:

1. *For every $x \in B$, the level set M_x is a $d-$dimensional torus (see definition 3.2) that is invariant under the flow of the Hamiltonian F_i for $i = 1, \ldots, d$.*

2. *There exists a bijective symplectic transformation*

$$\psi : D \times \mathbb{T}^d \to \bigcup_{x \in B} M_x \subset \mathbb{R}^d \times \mathbb{R}^d : (\theta, I) \mapsto (q,p),$$

such that $(F_i \circ \psi)(\theta, I)$ depends only on I. That is, such that

$$F_i(q,p) = f_i(I) \text{ for } (q,p) = \psi(\theta, I), \quad i = 1, \ldots, d \text{ with } f_i : D \to \mathbb{R}.$$

If the Hamiltonian H is real-analytic, then the transformation ψ is also real-analytic.

The flow of the Hamiltonian system in angle-action coordinates is given by

$$\dot{\theta}_i = \omega_i(I), \quad \dot{I}_i = 0, \quad i = 1, \ldots, d,$$

where $\omega(I) = \frac{\partial \widehat{H}}{\partial I}(I)$, and $\widehat{H}(I) = H(q,p),$ for $(q,p) = \psi(\theta, I)$.

Definition 3.3 (conditionally periodic, periodic, quasi-periodic)**.** The flow of a system

$$\dot{\theta} = \omega, \quad \omega \in \mathbb{R}^d,$$

on the torus $\mathbb{T}^d$ is said to be *conditionally periodic*. The flow is *periodic* if

$$\frac{\omega_i}{\omega_j} = \frac{k_i}{k_j}, \quad k_i, k_j \in \mathbb{Z}$$

for all pairs of frequencies. Otherwise the flow is said to be *quasi-periodic*. This includes the case where the flow is *non-resonant* and all the frequencies are rationally independent.

For any initial conditions, the non-resonant flow is dense and equally distributed on the torus $\mathbb{T}^d$. It is also possible for the flow on $\mathbb{T}^d$ to have $2 \geq r < d$ resonant frequencies. In such cases the flow is dense on a (degenerate) lower dimensional torus of dimension $d - r$, but not on the original invariant torus of dimension d.

3.2 Perturbation theory & Lindstedt-Poincaré series

KAM theory has its origins in classical perturbation theory. Hence, it is useful to describe the basic scheme of perturbation theory and how it manages to overcome the small divisors problem in order to better understand Kolmogorov's iteration.

One motivation of perturbation theory is the case where one wants to study the structure (i.e. the global dynamics) of a completely integrable Hamiltonian system after a (small, real-analytic) Hamiltonian perturbation is added. For example, a perturbation of the completely integrable Hamiltonian $H_0(I)$ given by

$$H(\theta, I) = H_0(I) + \varepsilon H_1(\theta, I), \tag{3.7}$$

defined on $D \times \mathbb{T}^d$, where D is an open subset of $\mathbb{R}^d$, H_0 and H_1 are real-analytic with H_1 bounded by a constant on a complex neighbourhood of $D \times \mathbb{T}^d$, and ε is a small parameter.

The idea underlying perturbation theory is to study the perturbed Hamiltonian system by trying to find a generating function which yields a symplectic, near-identity change of coordinates such that the perturbed Hamiltonian has the desired form — say, is independent of θ up to some order in the perturbation parameter ε, — when written in the new coordinates. This is done by writing the generating function as a power series in ε and then comparing the Fourier coefficients of the perturbed Hamiltonian in the new and the old variables at each power of ε.

More precisely, we want to find a generating function

$$S(\theta, \hat{I}) = \hat{I}^\top \theta + \sum_{n=1}^{N-1} \varepsilon^n S_n(\theta, \hat{I}), \tag{3.8}$$

where the S_n are $2\pi-$periodic in θ and where the symplectic transformation defined by

$$I = \frac{\partial S}{\partial \theta}(\theta, \hat{I}) = \hat{I} + \sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}(\theta, \hat{I}), \tag{3.9}$$

$$\hat{\theta} = \frac{\partial S}{\partial \hat{I}}(\theta, \hat{I}), \tag{3.10}$$

is such that the expansion of the Hamiltonian in the new variables $H(\theta, I) = \widehat{H}(\hat{\theta}, \hat{I}) = H_0(\hat{I}) + \varepsilon H_1(\hat{\theta}, \hat{I}) + \dots$ is independent of $\hat{\theta}$ for all terms up to ε^N.

Equations (3.9) and (3.7) give

$$H(\theta, I) = H_0\left(\hat{I} + \sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}(\theta, \hat{I})\right) + \varepsilon H_1\left(\theta, \hat{I} + \sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}(\theta, \hat{I})\right),$$

which we expand as a Taylor series in I (dropping the arguments $(\theta, \hat{I})$ of S_n)

$$\begin{aligned} H(\theta, I) = H_0(\hat{I}) &+ \frac{\partial H_0}{\partial I}\left(\sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}\right) + \frac{1}{2}\frac{\partial^2 H_0}{\partial I^2}\left(\sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}, \sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}\right) + \dots \\ &+ \varepsilon H_1(\hat{I}) + \varepsilon \frac{\partial H_1}{\partial I}\left(\sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}\right) + \varepsilon \frac{1}{2}\frac{\partial^2 H_1}{\partial I^2}\left(\sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}, \sum_{n=1}^{N-1} \varepsilon^n \frac{\partial S_n}{\partial \theta}\right) + \dots \end{aligned}$$

and collect in powers of ε.

$$\begin{aligned} H(\theta, I) = H_0(\hat{I}) &+ \varepsilon \left\{ \omega(\hat{I})^\top \frac{\partial S_1}{\partial \theta} + H_1 \right\} \\ &+ \varepsilon^2 \left\{ \omega(\hat{I})^\top \frac{\partial S_2}{\partial \theta} + \frac{1}{2}\frac{\partial^2 H_0}{\partial I^2}\left(\frac{\partial S_1}{\partial \theta}, \frac{\partial S_1}{\partial \theta}\right) + \frac{\partial H_1}{\partial I}\left(\frac{\partial S_1}{\partial \theta}\right) \right\} \\ &+ \varepsilon^3 \left\{ \omega(\hat{I})^\top \frac{\partial S_3}{\partial \theta} + \frac{1}{2}\frac{\partial^2 H_0}{\partial I^2}\left(\frac{\partial S_1}{\partial \theta}, \frac{\partial S_2}{\partial \theta}\right) + \frac{1}{3!}\frac{\partial^3 H_0}{\partial I^3}\left(\frac{\partial S_1}{\partial \theta}, \frac{\partial S_1}{\partial \theta}, \frac{\partial S_1}{\partial \theta}\right) \right. \\ &\qquad \left. + \frac{\partial H_1}{\partial I}\left(\frac{\partial S_2}{\partial \theta}\right) + \frac{1}{2}\frac{\partial^2 H_1}{\partial I^2}\left(\frac{\partial S_1}{\partial \theta}, \frac{\partial S_1}{\partial \theta}\right) \right\} + \dots, \end{aligned}$$

where the terms

$$\frac{1}{i!}\frac{\partial^i H_{k_0}}{\partial I^i}\left(\frac{\partial S_{k_1}}{\partial \theta}, \dots, \frac{\partial S_{k_i}}{\partial \theta}\right), \tag{3.11}$$

with $k_0 + k_1 + \dots + k_i = j$, $k_0 = 0, 1$, $k_1, \dots, k_i \geq 1$, are multi-linear functions of i arguments, corresponding to the product of the i-th derivative of H_{k_0} with the derivatives $\frac{\partial S_{k_1}}{\partial \theta}$, *et cetera*. Comparing this expansion with the expansion of $\widehat{H}$ one sees that at each order of ε it is necessary to solve a partial differential equation of the form

$$\omega(\hat{I})^\top \frac{\partial S_j}{\partial \theta}(\theta, \hat{I}) + K_j(\theta, \hat{I}) = \widehat{H}(\theta, \hat{I}), \tag{3.12}$$

where K_j is the sum of terms (3.11) of order ε^j. Since the S_j are required to be 2π-periodic in θ the function $\widehat{H}$ must (for the first N terms) equal $\overline{H}$, the angular averages of H over the angles:

$$\overline{H}_1(\hat{I}) = \frac{1}{(2\pi)^d}\int_{\mathbb{T}^d} H_1(\theta, \hat{I})\mathrm{d}\theta. \tag{3.13}$$

Now, expanding $\widehat{H}_1(\theta, \hat{I})$ and $S_j(\theta, \hat{I})$ as Fourier series, equation (3.12) becomes, for $j = 1$,

$$\omega(\hat{I})^\top \hat{I} \sum_{k\in\mathbb{Z}^d} ik \cdot s_{1,k}(\hat{I}) e^{ik\cdot\theta} + \sum_{k\in\mathbb{Z}^d} h_k(\hat{I}) e^{ik\cdot\theta} = \overline{H}_1(\hat{I}). \tag{3.14}$$

Higher order terms are similar. For $k = 0$ the coefficient $s_{1,0}(\hat{I})$ is arbitrary; for $k \neq 0$ one can compare Fourier coefficients to get a formal solution to (3.14)

$$s_{1,k}(\hat{I}) = \frac{-h_k(\hat{I})}{ik^\top\omega(\hat{I})}. \tag{3.15}$$

Clearly there is a problem whenever the frequencies $\omega(\hat{I})$ are rationally dependent and $k^\top\omega(\hat{I}) = k_1\omega_1 + \ldots + k_d\omega_d = 0$. If the perturbation H_1 is such that only a finite number of h_k are non-zero then (3.15) excludes only finitely many resonant frequencies and small neighbourhoods around them where the denominator of (3.15) may become arbitrarily small. In the general case, one can only use the fact that the Fourier coefficients $h_k(\hat{I})$ of an analytic function decay exponentially with $|k| = \sum_i |k_i|$ and require that the small denominators satisfy a Diophantine condition so that they do not vanish faster than the Fourier coefficients in the numerator.

Definition 3.4 (Diophantine number)**.** A number $\omega_i \in \mathbb{R}$ is *Diophantine* (of type (K, ν)), for $K > 0$, $\nu \geq 1$, if

$$\left|\omega_i - \frac{r}{s}\right| > \frac{K}{|s|^{1+\nu}}$$

for all $r/s \in \mathbb{Q}$. A number which is not Diophantine is called *Liouville*.

The *strong non-resonance condition* — or *Siegel's Diophantine condition* —

$$|k^\top\omega| \geq \frac{\gamma}{|k|^\nu}, \quad k \in \mathbb{Z}^d, \quad \nu, \gamma > 0, \tag{3.16}$$

ensures that the Fourier coefficients (3.15) are such that the series expansion defining S converges. For $\nu > d - 1$ the set of frequencies ω in a fixed ball, which do *not* satisfy the strong non-resonance condition, have Lebesgue measure bounded by $\gamma \cdot \text{Const}$.

If, for some fixed $\hat{I} = \hat{I}_*$, all the series for S_j, and their partial derivatives, converge then the system has an invariant torus $\{\hat{I} = \hat{I}_*, \theta \in \mathbb{T}^d\}$. If the series converge for all choices of $\hat{I}_*$ then the transformation has returned the Hamiltonian to the form of a completely integrable system and the phase space of the perturbed system is foliated into invariant tori. Clearly, this can not happen in general since it would exclude the possibility

of chaotic motion for Hamiltonian systems. The proof that in general the series diverges was given by Poincaré in 1893 [112].

3.3 Kolmogorov's iteration

In 1954, Kolmogorov [76] described a convergent transformation for perturbation theory which gave quadratic reduction of the perturbation size. The resulting method gave a proof that invariant tori carrying quasi-periodic flow with Diophantine frequencies persist under small perturbations of the Hamiltonian. Here we describe Kolmogorov's transformation.

Assume that $H_0(I)$ is a completely integrable Hamiltonian (see definition 3.1). The phase space of the system is therefore filled with invariant tori parameterized by their action coordinate, I. In what follows we fix $I = I_*$ so as to pick one specific torus (without loss of generality we can pick $I_* = 0$). This follows the approach of Kolmogorov and gives a method which proves preservation of individual tori in comparison with Arnold's [2] or Moser's [100] methods which prove preservation of Cantor sets of tori. The torus $\{I_* = 0, \theta \in \mathbb{T}^d\}$ is invariant under the flow of all Hamiltonians of the form

$$H(\theta, I) = c + \omega^\top I + \frac{1}{2} I^\top M(\theta, I) I, \tag{3.17}$$

for $M(\theta, I)$ a real, symmetric matrix in $\mathbb{R}^{d\times d}$, analytic in all its arguments. That is, all those Hamiltonians for which the linear terms in the Taylor expansion with respect to I, at zero, are θ-independent.

For a real analytic perturbation of (3.17)

$$H_\varepsilon = H(\theta, I) + \varepsilon G(\theta, I), \tag{3.18}$$

with ε small, it is possible to find a near-identity, symplectic coordinate transformation $(\theta, I) \mapsto (\hat{\theta}, \hat{I})$, such that, in the new coordinates the perturbed Hamiltonian is returned to the form of (3.17) with the same frequencies ω and, hence, the same quasi-periodic flow on the invariant torus $\{\hat{I} = 0, \theta \in \mathbb{T}^d\}$. This transformation is constructed via an iterative procedure — Kolmogorov's iteration — which holds under the condition that the frequencies of the perturbed system satisfy the strong non-resonance condition (3.16) and that the angular average

$$\overline{M}_0 = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} M(\theta, I = 0) \mathrm{d}\theta$$

is an invertible matrix. We will use the notation $(\theta^{(1)}, I^{(1)})$ for the coordinates after one step of the transformation. The transformation is constructed as the time-ε flow of a generating function T of the form

$$T(\theta^{(1)}, I^{(1)}) = \xi^\top \theta^{(1)} + T_0(\theta^{(1)}) + \sum_{i=1}^{d} I_i^{(1)} T_i(\theta^{(1)}), \tag{3.19}$$

with $\xi \in \mathbb{R}^d$ a constant vector and $T_0, T_1, \ldots, T_d$ all $2\pi-$periodic functions. The terms of order $\|I^{(1)}\|^2$ and higher are not used in this construction so we omit them from (3.19) and whenever they occur in the construction which follows. The time-ε flow of (3.19) gives the following relationship between the old and new coordinates:

$$I = I^{(1)} + \varepsilon \frac{\partial T}{\partial \theta^{(1)}}(\theta^{(1)}, I^{(1)}) + \mathcal{O}\left(\varepsilon^2\right), \quad \theta = \theta^{(1)} - \varepsilon \frac{\partial T}{\partial I^{(1)}}(\theta^{(1)}, I^{(1)}) + \mathcal{O}\left(\varepsilon^2\right). \tag{3.20}$$

The construction is similar to that of perturbation theory, we set (3.20) into (3.18) and get

$$\begin{aligned} H_\varepsilon(\theta, I) = c + \omega^\top & \left[I^{(1)} + \varepsilon \frac{\partial T}{\partial \theta^{(1)}} \right] \\ & + \frac{1}{2} \left[I^{(1)} + \varepsilon \frac{\partial T}{\partial \theta^{(1)}} \right]^\top M\left(\theta^{(1)}, I^{(1)}\right) \left[I^{(1)} + \varepsilon \frac{\partial T}{\partial \theta^{(1)}} \right] \\ & + \varepsilon G\left(\theta^{(1)}, I^{(1)}\right), \end{aligned} \tag{3.21}$$

where all the partial derivatives of T are evaluated at $\left(\theta^{(1)}, I^{(1)}\right)$ and where we have written $M\left(\theta^{(1)} - \varepsilon \frac{\partial T}{\partial I^{(1)}}, I^{(1)} + \varepsilon \frac{\partial T}{\partial \theta^{(1)}}\right)$ and $G\left(\theta^{(1)} - \varepsilon \frac{\partial T}{\partial I^{(1)}}, I^{(1)} + \varepsilon \frac{\partial T}{\partial \theta^{(1)}}\right)$ as perturbations of $M\left(\theta^{(1)}, I^{(1)}\right)$ and $G\left(\theta^{(1)}, I^{(1)}\right)$.

Substituting the series expansion

$$G\left(\theta^{(1)}, I^{(1)}\right) = G_0\left(\theta^{(1)}\right) + \sum_{j=1}^{d} I_j^{(1)} G_j\left(\theta^{(1)}\right) + {I^{(1)}}^\top Q\left(\theta^{(1)}, I^{(1)}\right) I^{(1)} \tag{3.22}$$

and the ansatz (3.19) into (3.21), and expanding $M\left(\theta^{(1)}, I^{(1)}\right)$ about $I^{(1)} = 0$ we get

$$\begin{aligned} H_\varepsilon(\theta, I) = c + \omega^\top I^{(1)} + \frac{1}{2} {I^{(1)}}^\top & M(\theta^{(1)}, I^{(1)}) I^{(1)} \\ & + \varepsilon \left\{ \omega^\top \xi + \omega^\top \frac{\partial T_0}{\partial \theta^{(1)}} + \omega^\top \sum_{j=1}^{d} I_j^{(1)} \frac{\partial T_j}{\partial \theta^{(1)}} \right. \\ & \qquad + {I^{(1)}}^\top M(\theta^{(1)}, 0)\xi + {I^{(1)}}^\top M(\theta^{(1)}, 0) \frac{\partial T_0}{\partial \theta^{(1)}} \\ & \qquad \left. + G_0(\theta^{(1)}) + \sum_{j=1}^{d} I_j^{(1)} G_j\left(\theta^{(1)}\right) \right\} \\ & + \mathcal{O}\left(\varepsilon^2\right) + \mathcal{O}\left(\varepsilon \|I^{(1)}\|^2\right). \end{aligned} \tag{3.23}$$

The requirement that this expansion is independent of $\theta^{(1)}$ up to linear terms in $I^{(1)}$ is

satisfied when the term in braces is Const. $+ \mathcal{O}\left(\varepsilon\|I^{(1)}\|^2\right)$, that is when

$$\begin{aligned}\omega^\top\xi + \omega^\top\frac{\partial T_0}{\partial\theta^{(1)}}\left(\theta^{(1)}\right) + G_0\left(\theta^{(1)}\right) + \sum_{j=1}^{d} I_j^{(1)} G_j\left(\theta^{(1)}\right)\\ + \sum_{j=1}^{d} I_j^{(1)}\left(\omega^\top\frac{\partial T_j}{\partial\theta^{(1)}}\left(\theta^{(1)}\right) + u_j\left(\theta^{(1)}\right) + v_j\left(\theta^{(1)}\right)\right) = \text{Const.},\end{aligned} \tag{3.24}$$

where we have made the substitutions

$$u(\theta^{(1)}) = M(\theta^{(1)}, 0)\xi, \tag{3.25}$$

$$v(\theta^{(1)}) = M(\theta^{(1)}, 0)\frac{\partial T_0}{\partial\theta^{(1)}}(\theta^{(1)}). \tag{3.26}$$

Since the generating function $T(\theta^{(1)}, I^{(1)})$ was assumed to be 2π-periodic in $\theta^{(1)}$, the condition (that the term in braces from (3.23) be Const. $+ \mathcal{O}\left(\|I^{(1)}\|^2\right)$) is satisfied when

$$\omega^\top\frac{\partial T_0}{\partial\theta^{(1)}}(\theta^{(1)}) + G_0(\theta^{(1)}) = \overline{G}_0, \tag{3.27}$$

$$\omega^\top\frac{\partial T_j}{\partial\theta^{(1)}}(\theta^{(1)}) + u_j(\theta^{(1)}) + v_j(\theta^{(1)}) + G_j(\theta^{(1)}) = \bar{u}_j + \bar{v}_j + \overline{G}_j, \tag{3.28}$$

$$\bar{u}_j + \bar{v}_j + \overline{G}_j = 0, \tag{3.29}$$

where the bars denote angular averages, as defined by equation (3.13). One now follows the procedure of section 3.2, noting that equations (3.27)–(3.29) have the same form as equation (3.12), and replacing the terms in them with their Fourier series. By matching Fourier coefficients, one obtains a formal solution for the Fourier coefficients of $T(\theta^{(1)}, I^{(1)})$.

Solving equation (3.27) determines $T_0\left(\theta^{(1)}\right)$ which can then be used with (3.26) to calculate $v = (v_1, \ldots, v_d)^\top$, which, in turn, can be used with equation (3.29) to determine $\bar{u} = (\bar{u}_1, \ldots, \bar{u}_d)^\top$ since $\overline{G}_j$, $j = 1, \ldots, d$ can be calculated directly. Once $\bar{u}$ is known, we can calculate the angular average of both sides of equation (3.25) which determines ξ since $\overline{M}(\theta^{(1)}, 0)$ was assumed to be invertible. Finally, equation (3.28) can be used to calculate $T_j(\theta^{(1)})$, $j = 1, \ldots, d$ and we have constructed $T(\theta^{(1)}, I^{(1)})$.

The transformation generated by the time-ε flow of $T(\theta^{(1)}, I^{(1)})$,

$$\theta = \theta^{(1)} - \varepsilon\frac{\partial T}{\partial I^{(1)}}(\theta^{(1)}, I^{(1)}), \quad I = I^{(1)} + \varepsilon\frac{\partial T}{\partial\theta^{(1)}}(\theta^{(1)}, I^{(1)})$$

gives

$$\begin{aligned}H_\varepsilon(\theta^{(1)}, I^{(1)}) =& H(\theta, I) + \varepsilon G(\theta, I)\\ =& c^{(1)} + \omega^\top I^{(1)} + \frac{1}{2} {I^{(1)}}^\top M^{(1)}(\theta^{(1)}, I^{(1)}) I^{(1)} + \varepsilon^2 G^{(1)}(\theta^{(1)}, I^{(1)}).\end{aligned} \tag{3.30}$$

The terms $c^{(1)}$ and $M^{(1)}\left(\theta^{(1)}, I^{(1)}\right)$ are both $\mathcal{O}(\varepsilon)$ close to the original terms, while the

frequency vector ω is unchanged from the original, unperturbed system. Iterating this procedure turns out to be convergent and gives a symplectic change of coordinates such that the perturbed system (3.18) is returned to the form of the unperturbed system (3.17) with an invariant torus carrying quasi-periodic flow with frequencies ω.

3.4 Kolmogorov's theorem

Here we present, and prove, Kolmogorov's theorem which shows that the iteration described in section 3.3 converges and that it returns the perturbed Hamiltonian (3.18) to the form of the unperturbed system (3.17).

Theorem 3.3 (Kolmogorov's theorem)**.** *Consider a real-analytic Hamiltonian $H(\theta, I)$, which is defined for I in a neighbourhood of $0 \in \mathbb{R}^d$ and $\theta \in \mathbb{T}^d$, and for which the linearization at $I = I^* = 0$ is independent of the angles:*

$$H(\theta, I) = c + \omega^\top I + \frac{1}{2} I^\top M(\theta, I) I.$$

Suppose that $\omega \in \mathbb{R}^d$ satisfies the Diophantine equation (3.16), (i.e. $|k^\top \omega| \geq \frac{\gamma}{|k|^\nu}$ for $k \in \mathbb{Z}^d \setminus \{0\}$), and that the angular average $\overline{M}_0 = \frac{1}{(2\pi)^d} \int_{\mathbb{T}^d} M(\theta, 0) \mathrm{d}\theta$ is an invertible $d \times d$ matrix satisfying

$$\|\overline{M}_0 x\| \geq \mu \|x\| \text{ for } x \in \mathbb{R}^d, \tag{3.31}$$

with ν, γ and μ positive constants.

Let $H_\varepsilon(\theta, I) = H(\theta, I) + \varepsilon G(\theta, I)$ be a real-analytic perturbation of $H(\theta, I)$. Then, there exists $\varepsilon_0 > 0$ such that for every $|\varepsilon| \leq \varepsilon_0$ there is an analytic, symplectic transformation $\psi_\varepsilon : (\hat{\theta}, \hat{I}) \mapsto (\theta, I)$, which is $\mathcal{O}(\varepsilon)$ close to the identity, analytic with respect to the perturbation parameter ε and which puts the perturbed Hamiltonian back into the form

$$H_\varepsilon(\theta, I) = c_\varepsilon + \omega^\top \hat{I} + \frac{1}{2} \hat{I}^\top M_\varepsilon(\hat{\theta}, \hat{I}) \hat{I} \tag{3.32}$$

for $(\theta, I) = \psi_\varepsilon(\hat{\theta}, \hat{I})$. The threshold ε_0 depends on d and on the constants ν, γ and μ, and on the bounds of H and G in a complex neighbourhood of $\{0\} \times \mathbb{T}^d$.

The perturbed system has an invariant torus $\left\{\hat{I} = 0, \hat{\theta} \in \mathbb{T}^d\right\}$ carrying a quasi-periodic flow with the same frequencies ω as the unperturbed system.

A practical difficulty with Kolmogorov's theorem is that the threshold ε_0 can become extremely small. The following proof, which follows [55], requires that $\varepsilon_0 \leq \delta_0^{5\alpha}$, $\alpha = \nu + d + 1$, where δ_0 is proportionate to $1/\nu$. This requirement is overly pessimistic — for example, Rüssmann [120, 121] gives optimal values for α which are smaller.

We introduce the notation

$$\|G\|_\rho = \sup\{|G(\theta, I)| : \ \|I\| < \rho, \ \|\mathrm{Im}\theta\| < \rho\},$$

where G is a bounded, analytic function on $W_\rho := B_\rho(0) \times U_\rho$ with $B_\rho(0)$ the complex ball about zero of radius ρ and $U_\rho := \{\theta \in \mathbb{T}^d + i\mathbb{R}^d : \ \|\mathrm{Im}\theta\| < \rho\}$, the complex extension of $\mathbb{T}^d$ with width ρ. The same notation $\|\cdot\|_\rho$ is used for vector- and matrix-valued functions. In these cases, the underlying norm is the supremum norm $\|v\| = \sup |v_i|$, or its induced matrix-norm.

Kolmogorov's theorem requires the following lemma which takes place under the same assumptions as the theorem 3.3 and in the setting of section 3.3.

Lemma 3.4. *Suppose that H and G are real-analytic functions and are bounded on W_ρ. Then, there exists $\delta_0 > 0$ such that the following bounds hold for Kolmogorov's iteration whenever $0 < \delta < \delta_0$:*

$$\begin{aligned} \text{if } \|\varepsilon G\|_\rho \leq \delta^{5\alpha}, \text{ then } \left\|\varepsilon^2 G^{(1)}\right\|_{\rho-\delta} &\leq \left(\frac{1}{2}\delta\right)^{5\alpha}, \\ \text{and } \|\varepsilon \nabla T\|_{\rho-\delta} &\leq \delta^{3\alpha}, \\ \text{and } \left\|M^{(1)} - M\right\|_{\rho-\delta} &\leq \delta^{2\alpha}, \end{aligned}$$

where $\alpha = \nu + d + 1$ and the threshold δ_0 depends only on d, ν, γ, μ, and $\|H\|_\rho$.

The proof of lemma 3.4 follows the terms arising in the construction of Kolmogorov's iteration, finding estimates which bound the size of the terms. It makes use of the following additional lemma which we state without proof. (A proof can be found in [55, X.4.1])

Lemma 3.5. *Suppose $\omega \in \mathbb{R}^d$ satisfies the Diophantine condition (3.16) $|k^\top \omega| \geq \gamma/|j|^\nu$ for $k \in \mathbb{Z}^d \setminus \{0\}$. Let G be a bounded, real-analytic function on U_ρ, the complex extension of the torus $\mathbb{T}^d$ with width ρ, and let $\overline{G}$ denote the angular average of G over $\mathbb{T}^d$ (cf. equation (3.13)), then,*

$$\omega^\top \frac{\partial F}{\partial \theta} + G = \overline{G} \tag{3.33}$$

has a unique real-analytic solution F on U_ρ with $\overline{F} = 0$. For every δ satisfying $0 < \delta < \min(\rho, 1)$ F is bounded on $U_{\rho-\delta}$ by

$$\|F\|_{\rho-\delta} \leq \kappa_0 \|G\|_\rho \delta^{-\alpha+1}, \qquad \left\|\frac{\partial F}{\partial \theta}\right\|_{\rho-\delta} \leq \kappa_1 \|G\|_\rho \delta^{-\alpha}, \tag{3.34}$$

where $\alpha = \nu + d + 1$, $\kappa_0 = \gamma^{-1} 8^d 2^\nu \nu!$, and $\kappa_1 = \gamma^{-1} 8^d 2^{\nu+1} (\nu+1)!$.

The proof of lemma 3.4 and of theorem 3.3 follows that presented in [55], with only minor differences.

Proof: lemma 3.4. We first introduce the notation

$$\|\cdot\|_j := \|\cdot\|_{\rho - j\delta/4} \text{ for } j = 0, 1, 2, 3, 4.$$

The transformation $\left(\theta^{(1)}, I^{(1)}\right) \mapsto (\theta, I)$ was constructed as the time$-\varepsilon$ flow of $T(\theta, I)$. That is, $(\theta, I) = y(\varepsilon)$ where $y(t)$ is the solution of $\dot{y} = J^{-1}\nabla T(y)$, $y(0) = \left(\theta^{(1)}, I^{(1)}\right)$. Suppose for the moment that

$$\|\varepsilon \nabla T\|_3 \leq \frac{1}{4}\delta. \tag{3.35}$$

Let $\left(\theta^{(1)}, I^{(1)}\right) \in W_{\rho-\delta}$, then for $0 \leq t \leq \varepsilon$, $y(t) \in W_{\rho-3\delta/4}$ and, by (3.35), $\|y(t) - y(0)\| \leq \frac{1}{4}\delta$. In particular, $\| \left(\theta^{(1)}, I^{(1)}\right) - (\theta, I)\| \leq \frac{1}{4}\delta$. Now, define

$$\begin{aligned}\varepsilon^2 R(\theta^{(1)}, I^{(1)}) :=& \left(I - I^{(1)} - \varepsilon \frac{\partial T}{\partial \theta^{(1)}}(\theta^{(1)}, I^{(1)}), \theta - \theta^{(1)} + \varepsilon \frac{\partial T}{\partial I^{(1)}}(\theta^{(1)}, I^{(1)}) \right) \\ =& y(\varepsilon) - \left[y(0) + \varepsilon J^{-1} \nabla T\left(y(0)\right) \right].\end{aligned} \tag{3.36}$$

Since the term from (3.36) in square brackets is the beginning of the Taylor expansion for $y(\varepsilon)$ we have (via the remainder theorem)

$$\left\| R(\theta^{(1)}, I^{(1)}) \right\| \leq \frac{1}{2} \max_{0 \leq t \leq \varepsilon} \|\ddot{y}(t)\|. \tag{3.37}$$

The right-hand side of (3.37) is, in turn, bounded by

$$\frac{1}{2} \left\| \nabla^2 T \nabla T \right\|_3,$$

and hence

$$\|R\|_4 \leq \frac{1}{2} \left\| \nabla^2 T \right\|_3 \|\nabla T\|_3. \tag{3.38}$$

We now follow the construction of T in section 3.3. From the Taylor expansion of $H_\varepsilon(\theta, I)$ we have the following terms at $\mathcal{O}\left(\varepsilon \left\| {I^{(1)}}^2 \right\|\right)$:

$$\varepsilon \left\{ {I^{(1)}}^\top \left[Q(\theta^{(1)}, I^{(1)}) + \left(\frac{\partial M}{\partial I} \frac{\partial T}{\partial \theta^{(1)}} - \frac{\partial M}{\partial \theta} \frac{\partial T}{\partial I^{(1)}} \right) \right] I^{(1)} + {I^{(1)}}^\top P(\theta^{(1)}, I^{(1)}) I^{(1)} \right\},$$

where $Q(\theta^{(1)}, I^{(1)})$ comes from the $\mathcal{O}\left(\left\|I^{(1)}\right\|^2\right)$ term in (3.22), the expansion of $G\left(\theta^{(1)}, I^{(1)}\right)$, and where

$${I^{(1)}}^\top P(\theta^{(1)}, I^{(1)}) I^{(1)} = {I^{(1)}}^\top \left(M(\theta^{(1)}, I^{(1)}) - M(0, I^{(1)}) \right) \frac{\partial T}{\partial \theta^{(1)}}$$

is symmetric and includes the $\mathcal{O}\left(\left\|I^{(1)}\right\|\right)$ terms from the expansion of $M\left(\theta^{(1)}, I^{(1)}\right) \frac{\partial T}{\partial \theta^{(1)}}$. Hence the new matrix $M^{(1)}\left(\theta^{(1)}, I^{(1)}\right)$ in (3.30) has the form

$$M^{(1)}(\theta^{(1)}, I^{(1)}) = M(\theta^{(1)}, I^{(1)}) + \varepsilon L(\theta^{(1)}, I^{(1)}),$$

with

$$L(\theta^{(1)}, I^{(1)}) = \sum_{j=1}^{d} \left(\frac{\partial M}{\partial I_j} \frac{\partial T}{\partial \theta_j^{(1)}} - \frac{\partial M}{\partial \theta_j} \frac{\partial T}{\partial I_j^{(1)}} \right) (\theta^{(1)}, I^{(1)}) + P(\theta^{(1)}, I^{(1)}) + Q(\theta^{(1)}, I^{(1)}).$$

It follows that

$$\left\| M^{(1)} - M \right\|_4 = \|\varepsilon L\|_4 \leq 2\varepsilon \left(\|\nabla M\|_4 \|\nabla T\|_4 + \left\|\nabla^2 G\right\|_4 \right). \tag{3.39}$$

Similarly, we look at the $\mathcal{O}\left(\varepsilon^2\right)$ terms from the Taylor expansion of H_ε:

$$R^\top \nabla H + (J\nabla T)^\top \nabla^2 H J \nabla T + (\nabla G)^\top J \nabla T,$$

which gives the estimate

$$\left\| G^{(1)} \right\|_4 \leq \|\nabla H\|_3 \|R\|_4 + \|\nabla T\|_4^2 \left\|\nabla^2 H\right\|_3 + \|\nabla G\|_3 \|\nabla T\|_4 . \tag{3.40}$$

We now trace the construction of T in section 3.3. We first apply lemma 3.5 to equation (3.27) and get

$$\|T_0\|_1 \leq \kappa_0 (\delta/4)^{-\alpha+1} \|G_0\|_0 , \tag{3.41}$$

$$\left\| \frac{\partial T_0}{\partial \theta^{(1)}} \right\|_1 \leq \kappa_1 (\delta/4)^{-\alpha} \|G_0\|_0 . \tag{3.42}$$

By construction, u and v satisfy the estimates

$$\|v\|_1 \leq \|M\|_1 \left\| \frac{\partial T_0}{\partial \theta^{(1)}} \right\|_1 , \tag{3.43}$$

and

$$\|u\|_1 \leq \|M\|_1 \mu^{-1} \left(\|v\|_1 + \sum_{j=1}^{d} \|G_j\|_1 \right). \tag{3.44}$$

Applying lemma 3.5 to (3.28) gives

$$\|T_j\|_2 \leq \kappa_0 (\delta/4)^{-\alpha+1} \left(\|v\|_1 + \|u\|_1 + \|G_j\|_1 \right). \tag{3.45}$$

Direct substitution of (3.42) in (3.43) gives $\|v\|_1 \leq C\delta^{-\alpha} \|G_0\|_0$ which in turn gives $\|u\|_1 \leq C\delta^{-\alpha} \|G\|_0$ via (3.44) and then, via (3.45)

$$\|T_j\|_2 \leq C\delta^{-2\alpha+1} \left(\|G\|_0 + \|G_0\|_0 + \|G_j\|_1 \right). \tag{3.46}$$

Using the estimates (3.41) and (3.46) gives $\|T\|_0 \leq C\delta^{-2\alpha+1} \|G\|_0$ and applying Cauchy's estimates, $\|\nabla F\|_{\rho-\delta} \leq C \|F\|_\rho \delta^{-1}$, and $\left\|\nabla^2 F\right\|_{\rho-\delta} \leq C \|F\|_\rho \delta^{-2}$, gives the following

bounds:

$$\|\nabla T\|_3 \leq C\delta^{-2\alpha} \|G\|_0 , \qquad \left\|\nabla^2 T\right\|_3 \leq C\delta^{-2\alpha-1} \|G\|_0 . \tag{3.47}$$

The constant C depends on ν, d, γ, μ, and on $\|H\|_\rho$.

The bound

$$\|\varepsilon \nabla T\|_{\rho-\delta} \leq C\delta^{-2\alpha} \|\varepsilon G\|_\rho \tag{3.48}$$

follows directly from (3.47). Using equations (3.47) and (3.38) with equations (3.40) and (3.39), and using Cauchy's estimates again to bound the derivatives of H and G gives

$$\left\|\varepsilon^2 G^{(1)}\right\|_{\rho-\delta} \leq C\delta^{-4\alpha-4} \|\varepsilon G\|_\rho^2 \leq C\delta^{-5\alpha} \|\varepsilon G\|_\rho^2 , \tag{3.49}$$

and

$$\left\|M^{(1)} - M\right\|_{\rho-\delta} \leq C\delta^{-2\alpha-1} \|\varepsilon G\|_\rho \leq C\delta^{-3\alpha} \|\varepsilon G\|_\rho . \tag{3.50}$$

The working above assumes that condition (3.35) holds. From (3.47) we can see that this can be ensured by requiring

$$\|\varepsilon G\|_\rho \leq \delta^{5\alpha}$$

for $\delta \leq \delta_0$ with δ_0 sufficiently small. Substitution of this into the bounds (3.48) – (3.50) gives

$$\begin{aligned} \|\varepsilon \nabla T\|_{\rho-\delta} &\leq C\delta^{3\alpha}, \\ \left\|\varepsilon^2 G^{(1)}\right\|_{\rho-\delta} &\leq C\delta^{5\alpha}, \\ \left\|M^{(1)} - M\right\|_{\rho-\delta} &\leq C\delta^{2\alpha}. \end{aligned}$$

□

We are now in a position to prove that Kolmogorov's iterations converges.

Proof: Kolmogorov's theorem 3.3. Applying Kolmogorov's iteration to (3.18) repeatedly generates the sequences

$$\begin{aligned} G &= G^{(0)}, G^{(1)}, G^{(2)}, \ldots, \\ M &= M^{(0)}, M^{(1)}, M^{(2)}, \ldots, \\ T &= T^{(0)}, T^{(1)}, T^{(2)}, \ldots . \end{aligned}$$

Applying lemma 3.4 to $\varepsilon^2 G^{(1)}$, $M^{(1)} - M^{(0)}$, and $\varepsilon \nabla T^{(0)}$ gives

$$\begin{aligned} \left\| \varepsilon^2 G^{(1)} \right\|_{\rho-\delta} &\leq (\delta/2)^{5\alpha}, \\ \left\| \varepsilon \nabla T^{(0)} \right\|_{\rho-\delta} &\leq \delta^{3\alpha}, \\ \left\| M^{(1)} - M^{(0)} \right\|_{\rho-\delta} &\leq \delta^{2\alpha}. \end{aligned} \tag{3.51}$$

After a second step of Kolmogorov's iteration, one wishes to bound $\varepsilon^4 G^{(2)}$, $\varepsilon^2 \nabla T^{(1)}$, etc. We introduce the notation $\rho = \rho^{(0)}$, $\delta = \delta^{(0)}$, $\rho^{(k)} = \rho^{(k-1)} - \delta^{(k-1)} = \rho - \sum_{i=0}^{k-1} 2^{-i}\delta$, and $\delta^{(k)} = \delta^{(k-1)}/2 = 2^{-k}\delta$ for $k = 1, 2, \ldots$, and rewrite equation (3.51) above as

$$\left\| \varepsilon^2 G^{(1)} \right\|_{\rho^{(1)}} \leq {\delta^{(1)}}^{5\alpha}.$$

We can therefore continue to apply lemma 3.4. This gives the bounds

$$\begin{aligned} \left\| \varepsilon^4 G^{(2)} \right\|_{\rho^{(1)}-\delta^{(1)}} &\leq \left(\delta^{(1)}/2\right)^{5\alpha}, \\ \left\| \varepsilon^2 \nabla T^{(1)} \right\|_{\rho^{(1)}-\delta^{(1)}} &\leq {\delta^{(1)}}^{3\alpha}, \\ \left\| M^{(2)} - M^{(1)} \right\|_{\rho^{(1)}-\delta^{(1)}} &\leq {\delta^{(1)}}^{2\alpha}. \end{aligned}$$

Or, more generally,

$$\left\| \varepsilon^{2^k} G^{(k)} \right\|_{\rho^{(k)}} \leq \left(2^{-k}\delta\right)^{5\alpha}, \tag{3.52}$$

$$\left\| \varepsilon^{2^{(k-1)}} \nabla T^{(k-1)} \right\|_{\rho^{(k)}} \leq \left(2^{-(k-1)}\delta\right)^{3\alpha}, \tag{3.53}$$

$$\left\| M^{(k)} - M^{(k-1)} \right\|_{\rho^{(k)}} \leq \left(2^{-(k-1)}\delta\right)^{2\alpha}. \tag{3.54}$$

Equation (3.54) implies that the inverse of $M^{(k)}$ remains bounded for all k, and so, we are justified in using Kolmogorov's iteration and lemma 3.4 iteratively. At each step, the time-ε^{2^k} flow of $T^{(k)}$ defines a symplectic transformation $\varphi_\varepsilon^{(k)}$. By the bound (3.53) this transformation satisfies $\left\| \varphi_\varepsilon^{(k)} - \mathrm{Id} \right\|_{\rho/2} \leq \left(2^{-k}\delta\right)^{3\alpha}$, since $\rho^{(k)} > \rho/2$ for all k. This implies that the sequence

$$\psi_\varepsilon^{(k)} := \varphi_\varepsilon^{(0)} \circ \varphi_\varepsilon^{(1)} \circ \cdots \circ \varphi_\varepsilon^{(k)}(\theta^{(k)}, I^{(k)})$$

converges uniformly on $W_{\rho/2} \times (-\varepsilon_0, \varepsilon_0)$ to the limit $\psi_\varepsilon\left(\hat{\theta}, \hat{I}\right)$ which is analytic in ε, $\hat{\theta}$, and $\hat{I}$. It therefore follows that ψ_ε is $\mathcal{O}(\varepsilon)$ close to the identity on $W_{\rho/2}$ since it is analytic in ε and since $\psi_0 = \mathrm{Id}$.

The transformation $\psi_\varepsilon^{(k)}$ was constructed such that

$$H \circ \psi_\varepsilon^{(k-1)}\left(\theta^{(k)}, I^{(k)}\right) = c^{(k)} + \omega^\top I^{(k)} + \frac{1}{2} {I^{(k)}}^\top M^{(k)} I^{(k)} + \varepsilon^{2^k} G^{(k)}\left(\theta^{(k)}, I^{(k)}\right),$$

and so, by (3.52) the transformed Hamiltonian $H \circ \psi_\varepsilon \left(\hat{\theta}, \hat{I}\right)$ has the desired form, (3.32).
□

Chapter 4

Symplectic Integrators and Periodic Orbits

> The goal of the theory of numerical algorithms should be to design proper algorithms to preserve *most* of the dynamics of *most* systems.
>
> — Z.-J. Shang, 1999

When a symplectic integrator using step size h is applied to a d-degree of freedom Hamiltonian vector-field $f = X_H$ it gives a symplectic map $\Phi_{h,f} : \mathbb{R}^{2d} \to \mathbb{R}^{2d}$. We are interested in what happens to periodic orbits of a dynamical system when the system is discretized by a symplectic integrator. Does the numerical solution given by $\Phi_{h,f}$ preserve the periodic orbits of the original vector field? Since periodic orbits can be thought of as invariant tori with a single degree of freedom, KAM theory is a useful tool for addressing such questions.

We begin, in section 4.1, with some background on periodic orbits of Hamiltonian systems. In section 4.2 we describe the theorem by Shang [130, 129] that a symplectic integrator applied to a Hamiltonian vector field preserves most of the invariant tori with full dimension. We want to show that a similar result holds for lower dimensional tori and in particular, for periodic orbits. Section 4.3 sets the framework for the problem. In section 4.4 we make use of a theorem that it is possible to embed a symplectic map in a quasi-periodic non-autonomous flow. We combine this with a KAM result of Jorba and Villaneuva [73] to show that periodic orbits persist in the new flow; section 4.5. The preserved periodic orbits have a slightly perturbed period and gain an additional degree of freedom, or rather, as Jorba and Simó so elegantly put it in [72] "the perturbed tori are 'quasi-periodically dancing' to the 'rhythm' of the perturbation." Section 4.6 deals with the resonant case; when the integrator step size divides the period of the orbit. The work presented in this chapter has been accepted for publication in the article [88].

4.1 Background to periodic orbits of Hamiltonian systems

For periodic orbits, Floquet multipliers play a role similar to that of the eigenvalues of the linearization of a fixed point. One Floquet multiplier is always equal to 1 as it corresponds to the direction tangential to the periodic orbit. The other multipliers are in one-to-one correspondence with the eigenvalues of the linearization of the Poincaré mapping[1]. In the case of Hamiltonian systems, the Floquet multipliers occur in pairs (λ, λ^{-1}) of the same multiplicity. The "second" eigenvalue equal to 1 corresponds to one of the *normal* directions of the periodic orbit. The eigenvector associated with this multiplier spans the direction conjugate to that of the first multiplier 1. Consequently, and in contrast to the case of dissipative systems which may have isolated periodic orbits (cf. §2.3), periodic orbits of Hamiltonian systems form 1-parameter families. Rather than use the coordinate corresponding to the "second" Floquet multiplier of 1 (i.e. the coordinate conjugate to that along the periodic orbit), the period, or the value of the Hamiltonian is often used to parameterize the family of orbits. An illustration of this is given in figure 4.1 for the one degree of freedom Hamiltonian pendulum

$$H(q,p) = \frac{1}{2}p^2 - \cos(q),$$

where all (non-equilibrium and non-homoclinic) solutions are periodic, and in figure 4.2 for the two degree of freedom Hénon-Heiles system with Hamiltonian

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

which has both chaotic and regular motion, (cf. §4.6). (Actually, figures 4.1 and 4.2 show the continuation of the sets of n-periodic points which replace the periodic orbits of the continuous system when it is discretized with a (symplectic) integrator using a step size resonant with the period of the orbit. For large n, these discrete sets are an accurate approximation of the periodic orbits. We discuss exactly this point in section 4.6.)

The 1-parameter families can undergo a variety of bifurcations (for example, consider the pendulum when $H = 1$ or the Hénon-Heiles system when $H = \frac{1}{6}$.) We abstain entirely from discussing these and instead mention the monograph [60] by Hanßmann which gives comprehensive coverage of bifurcations in Hamiltonian systems.

Theorem 4.1 (Floquet's theorem)**.** *For the system*

$$\dot{y} = A(t)y, \tag{4.1}$$

let $A(t)$ be a continuous $n \times n$ matrix, defined for $-\infty < t < \infty$, which is periodic with

[1]The mapping σ from a section Σ, transverse to the flow, to itself such that $\sigma(y_0) = y_1 \implies \varphi_{t_1}(y_0) = y_1$ for some $t_1 \in \mathbb{R}$ and with $\varphi_t(y_0) \notin \Sigma$ for $0 < t < t_1$. Further details on Poincaré maps are in §4.6.

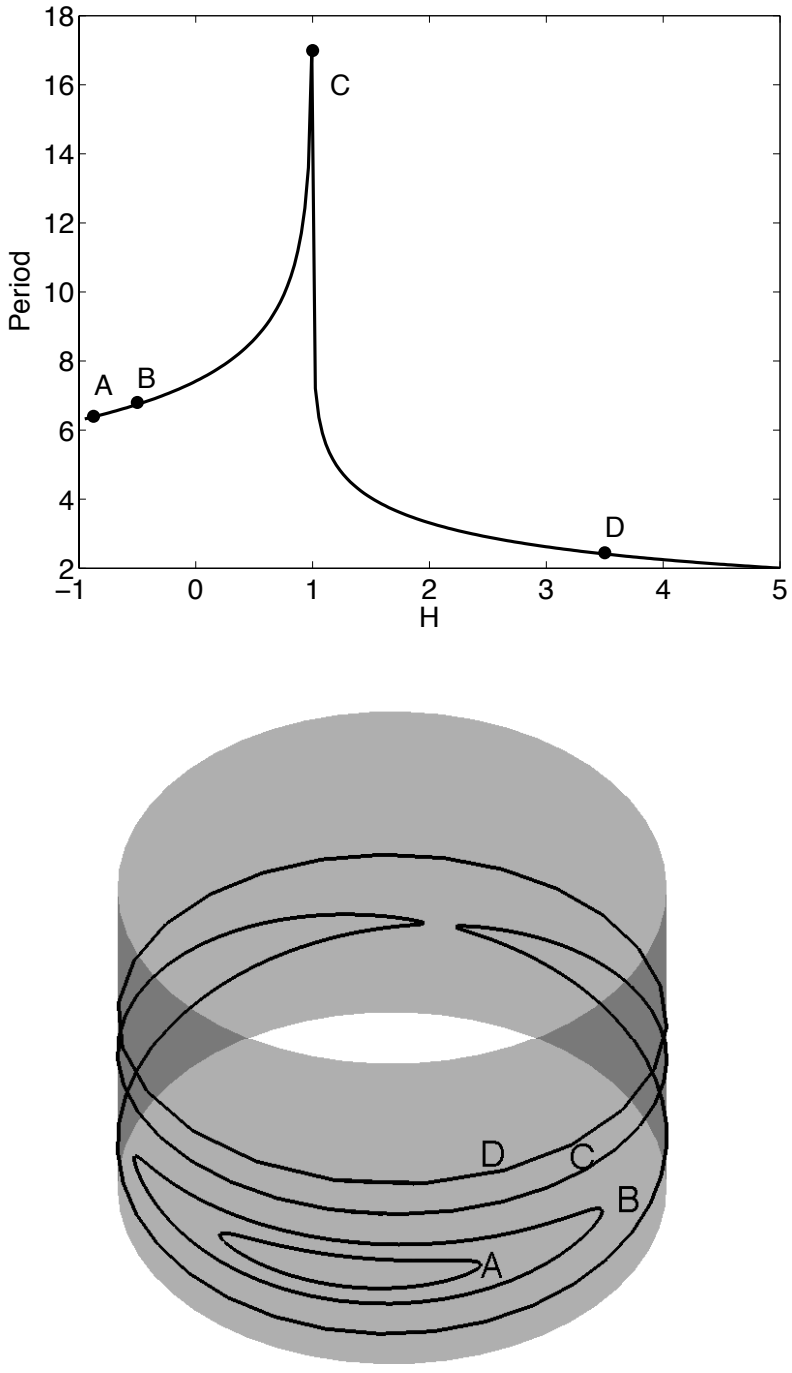


Figure 4.1: Plot of period against the value of the Hamiltonian for the 1-parameter family of periodic orbits of the pendulum (*above*), and a selection of four periodic orbits, plotted in the natural phase space of the pendulum — a cylinder (*lower*). The points in the 1-parameter family corresponding to the four orbits are indicated. Note that the homoclinic orbits connected to the hyperbolic fixed point mean that $T \to \infty$ as $H \to 1$.

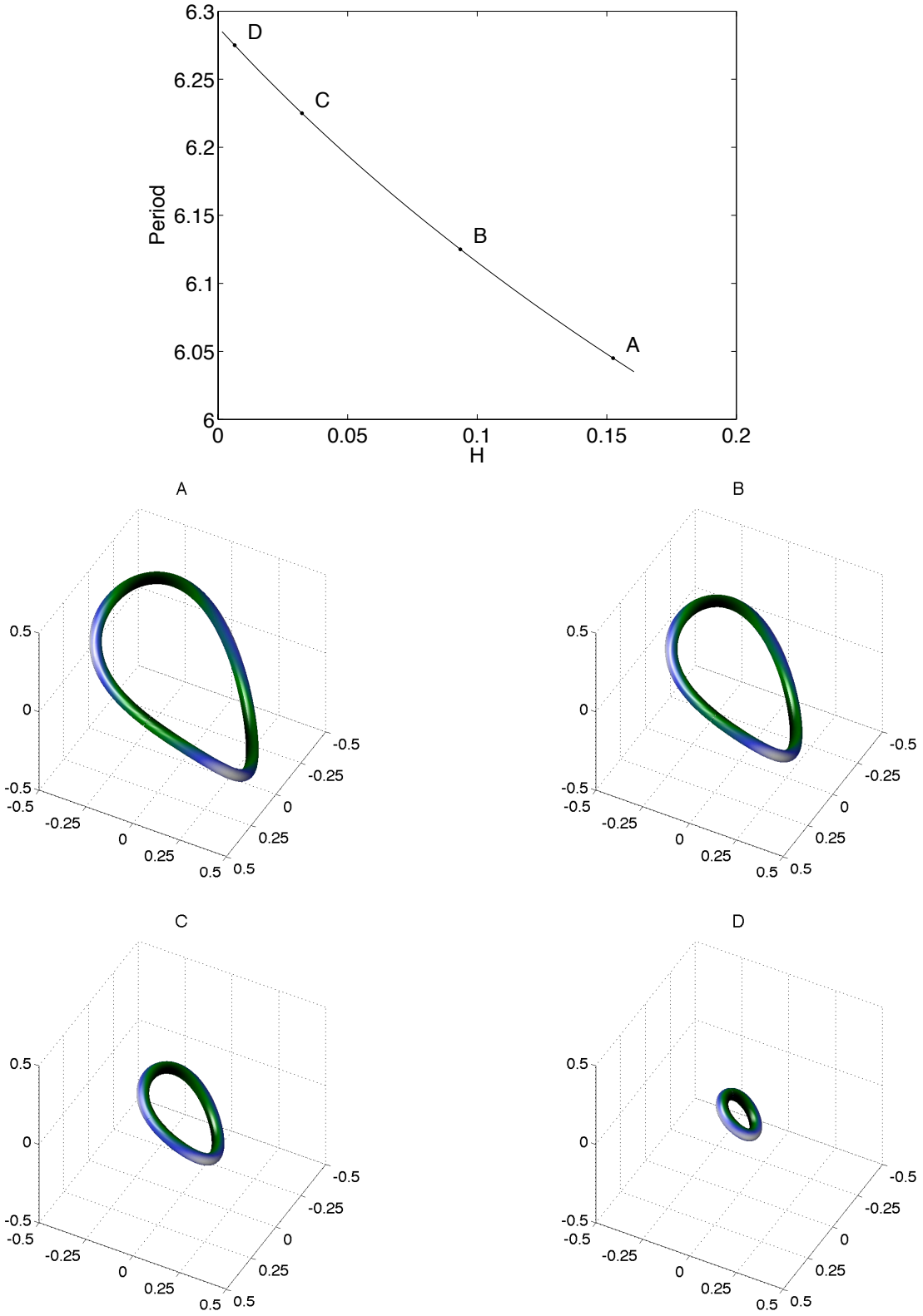


Figure 4.2: Plot of period against the value of the Hamiltonian for the 1-parameter family of periodic orbits of the Hénon-Heiles system (*above*), and a selection of four "fattened out" periodic orbits corresponding to the four points indicated in the 1-parameter family (*lower*). The periodic orbits are drawn in the (q_1, q_2, p_1) space.

period T,

$$A(t+T) = A(t).$$

Then any fundamental matrix $Y(t)$ of (4.1) can be written in the form

$$Y(t) = Z(t)e^{Rt}, \qquad \text{where } Z(t+T) = Z(t)$$

and where R is a constant matrix. Further, there exists a T-periodic change of coordinates $y = B(t)x$ such that in the new coordinates (4.1) becomes

$$\dot{x} = Rx.$$

More details on Floquet theory can be found in [62, 49, 5].

Much of the analysis of periodic orbits, and indeed of invariant tori (both lower and maximal dimensional), makes use of *reducibility* of the flow on these objects. For periodic orbits, Floquet's theorem gives symplectic coordinates (θ, I, z), where $\theta \in \mathbb{T} := \mathbb{R}/\mathbb{Z}$ is the coordinate along the periodic orbit $\gamma = \{(\theta, I, z) : (I, z) = (I_0, 0)\}$. The coordinate I is conjugate to θ, while z consists of all the directions "normal" to the periodic orbit [60]. The equations of motion in these variables read

$$\begin{aligned}
\dot{\theta} &= f(\theta, I, z),\\
\dot{I} &= g(\theta, I, z),\\
\dot{z} &= h(\theta, I, z),
\end{aligned}$$

and for $I = I_0$ the right-hand sides are of the form

$$\begin{aligned}
f(\theta, I_0, z) &= \omega(I_0) + \mathcal{O}\left(z^2\right),\\
g(\theta, I_0, z) &= \mathcal{O}\left(z^3\right),\\
h(\theta, I_0, z) &= \Omega(I_0)\cdot z + \mathcal{O}\left(z^2\right).
\end{aligned}$$

The linear part $\Omega(I_0)\cdot z$ is independent of the angle θ and the eigenvalues of Ω are the Floquet exponents[2] of the periodic orbit. For Floquet multipliers on the unit circle the periodic orbit is elliptic, and sits at the center of a family of nested invariant tori (cf. figure 1.3). The flow on these tori winds about the periodic orbit — we attempt to illustrate this in figure 4.3.

[2]Exponentiation of the *Floquet exponents* yields the *Floquet multipliers*

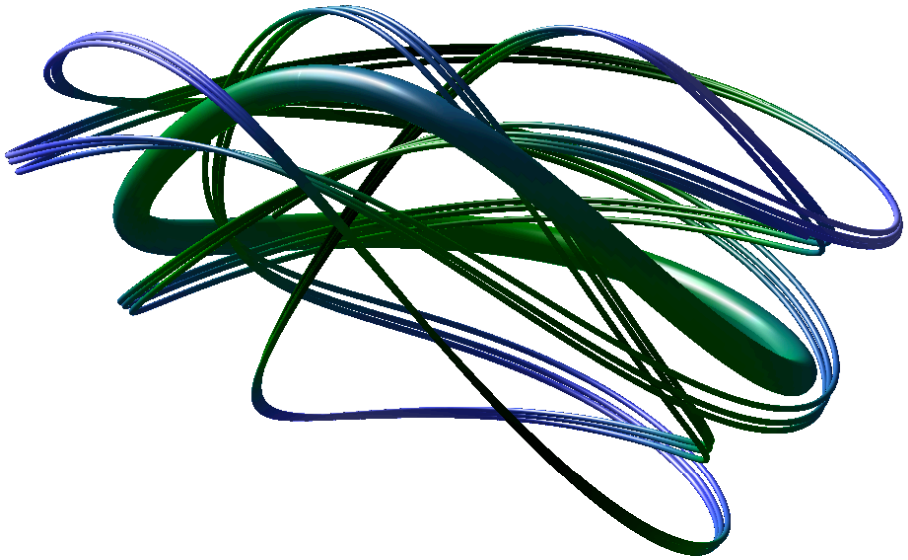


Figure 4.3: Trajectory on an invariant torus (*thin tube*) winding about an elliptic periodic orbit (*thick tube*) of the Hénon-Heiles system.

4.2 Symplectic integrators preserve invariant tori: Shang's theorem

Knowing that particular geometric structures persist in numerical solutions is important if we are to correctly understand the dynamical systems which the solutions represent. KAM-type theorems are a powerful tool in the analysis of dynamical systems because of the information they give about structural stability. Initially, KAM theory focused on the case of $H(\theta, I) = H_0(I)$; that is, perturbations of fully integrable systems (cf. §3.3 and 3.4). Since then, KAM theory has been extended in many directions, to the point where it is now difficult to succinctly state all the results covered by the theory.

Roughly speaking, modern KAM theory says that for C^k open sets (where k may be large depending on d) of d degree of freedom dynamical systems possessing some geometric property (e.g. Hamiltonian, volume-preserving, reversible, *et cetera*) there exist sets of positive measure covered by invariant tori. The tori need not have dimension d (e.g. lower dimensional tori) and d need not be finite (e.g. infinite dimensional KAM theory, KAM theory for PDEs). See [31], and the references therein, for a comprehensive survey and exposition of KAM theory and for a list of the current directions of KAM theory.

When a KAM-type theorem holds for a continuous dynamical system, one would like

an analogous result to also hold for a discrete version of the system given as the numerical solution of the differential equation. For example, when a symplectic integrator is applied to a Hamiltonian system containing a full dimensional (i.e. d-dimensional) invariant torus, and when the step size of the integrator is not resonant with any of the frequencies of the invariant torus, it is reasonable to hope that the torus persists in the numerical solution. The proof that such tori do indeed persist in the numerical solution is possible through a KAM theorem for symplectic maps, due to Shang [130, 129], who proved that the original KAM theory for full dimensional invariant tori of symplectic flows also holds for symplectic maps. That is, a symplectic integrator applied to a Hamiltonian system, with an invariant torus whose frequencies are strongly non-resonant, preserves the torus, with unchanged frequencies, for a Cantor set's worth of strongly non-resonant step sizes. The tori are only slightly deformed by the integrator and the density of the Cantor set of step sizes tends to one as $h \to 0$. To illustrate this, (and the converse point, that we must expect more and more invariant tori to be destroyed with increasing step size), we show, in figure 4.4, the (numerical) Poincaré map for the Hénon-Heiles system integrated with the leapfrog method.

The invariant tori of the full system are closed curves of the Poincaré map. As the step size h increases, and with it the size of the perturbation to the initial invariant tori, fewer and fewer closed curves are seen. The "smearing" of the curves for large step size is a separate effect which can be attributed to the need to interpolate between points in the numerical trajectories in order to estimate the position of points on the section[3] and to the fact that symplectic integrators keep the energy of trajectories close to the original value, rather than conserving it exactly. The preserved tori are therefore not exactly iso-energetic surfaces which "fattens" the closed curves in the numerical Poincaré sections for large h.

Shang's theorem states:

Theorem 4.2. *[130] Given an analytic, non-degenerate, and integrable Hamiltonian system of d-degrees of freedom, and given N Diophantine[4] frequency vectors ω^j, $j = 1, 2, \ldots, N$ in the domain of frequencies of the system, there exists a Cantor set $\mathcal{I} \subset \mathbb{R}$, depending on the N frequency vectors, such that for any symplectic algorithm applied to the system, there exists a positive number δ_0 such that if the step size h of the algorithm falls in the set $(-\delta_0, \delta_0) \cap \mathcal{I}$, then the algorithm has N invariant tori with frequency vectors $h\omega^j$, $j = 1, 2, \ldots, N$ when applied to the integrable system. These invariant tori approximate the corresponding tori of the system, in the sense of Hausdorff, with the order equal to the order of accuracy of the algorithm. The Cantor set $\mathcal{I}$ has density one at the origin.*

Our goal in this chapter is to develop a similar result for preservation of periodic orbits (and other lower dimensional tori). To do so, two possible approaches are available.

[3] See section 4.6 for details in the construction of numerical Poincaré sections.

[4] I.e. each $\omega^j \in \mathbb{R}^d$ satisfies $|k^\top \omega| \geq \frac{\gamma}{|k|^\nu}$, $0 \neq k = (k_1, \ldots, k_d) \in \mathbb{Z}^d$ for some $\gamma > 0$ and $\nu > 0$.

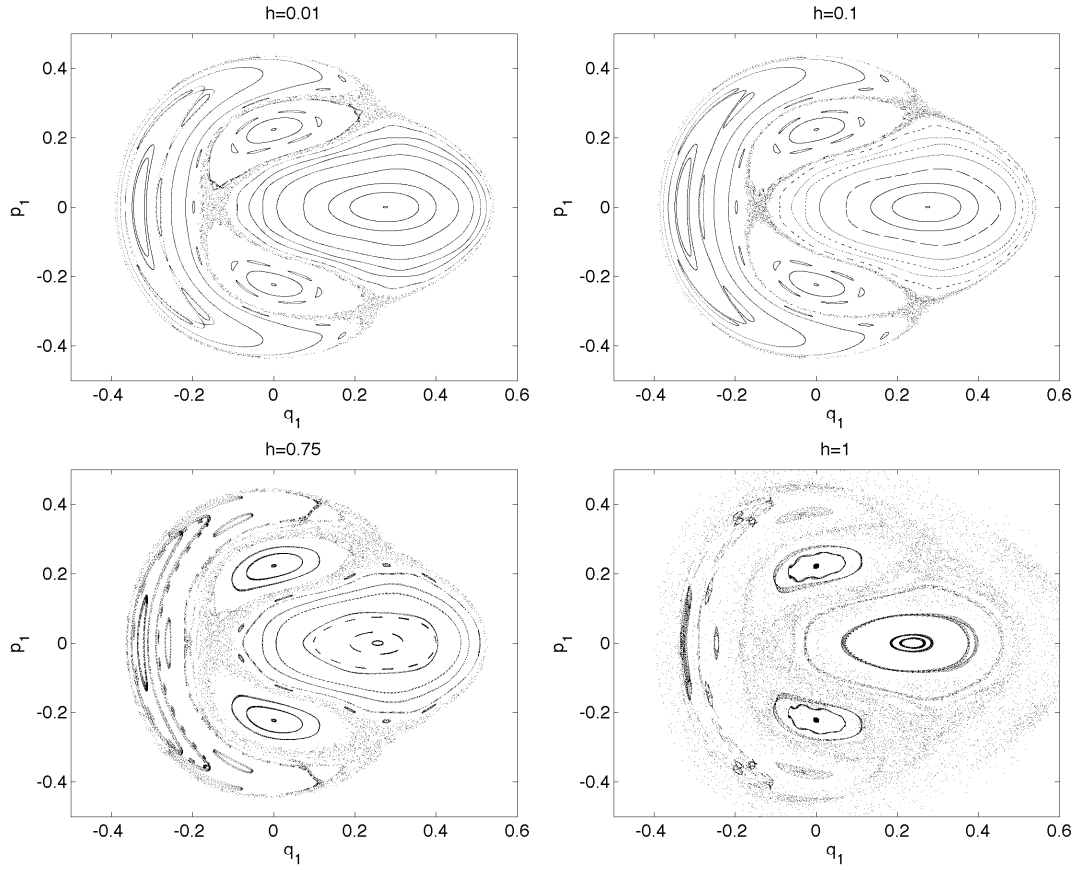


Figure 4.4: Poincaré sections of the Hénon-Heiles system showing the destruction of invariant tori with increasing perturbation size. The system was integrated using the (symplectic) leapfrog method for different step sizes. As the step size h, and hence the perturbation size, increases, fewer and fewer invariant tori (seen as closed curves in the section) remain.

The first is to follow the example of Shang who re-proved the original KAM theorem in the setting of analytic, symplectic maps. This is not a trivial undertaking. The second approach is to use interpolation to embed the map produced by a symplectic integrator into a symplectic flow which is sufficiently smooth (or even analytic). One can then try to use an existing KAM result for preservation of periodic orbits of perturbed symplectic flows to prove that such orbits are also preserved by the symplectic map. This second approach has the advantage that it is often easier to think of a problem in terms of maps while it is simpler to give a proof in terms of flows. Many KAM style results for lower dimensional invariant tori already exist and by using interpolation one can avoid redoing lengthy proofs for maps. It is this second approach which we take here.

For one degree of freedom systems (i.e. systems of dimension two) a periodic orbit is also a full dimensional invariant torus and both the resonant and non-resonant case are well understood. The non-resonant case is covered by Shang's theorem 4.2, while the resonant case is discussed at the beginning of section 4.6. We are therefore interested in

Hamiltonian systems with several degrees of freedom and which contain a periodic orbit. More generally, we can work with tori of any dimension from zero (i.e. fixed points) through to full dimension. The following section gives a full description of the conditions we require the Hamiltonian system to satisfy.

4.3 Framework & the P1 & P2 conditions

Here, and in sections 4.4 and 4.5, we use a slightly different notation from the rest of this thesis in order that the notation here is consistent with the article [73] which we refer to at several points. We use $\mathcal{H}$ to denote the original, unperturbed Hamiltonian function while H appears as a term in the perturbation, and subsequent modification, of $\mathcal{H}$.

The periodic orbits (or lower dimensional invariant tori) we work with are assumed to be non-degenerate. That is, they are not contained within a resonant invariant torus of higher dimension. For periodic orbits, this requirement is ensured by the assumption that the Floquet multipliers — the eigenvalues of the monodromy matrix — of the original system are distinct, and, therefore, by the canonical structure, are non-zero (cf. §2.3).

The exact conditions necessary to ensure non-degeneracy of lower dimensional tori, in general, are given explicitly in theorem 4.6.

We will assume that the Hamiltonian system $\mathcal{H}$ we are working with has the following properties.

1. The initial Hamiltonian $\mathcal{H}$ has d degrees of freedom and is autonomous, and analytic with respect to all its variables. It contains an invariant torus with linear, quasi-periodic flow with rationally independent frequencies $\hat{\omega}^{(0)} \in \mathbb{R}^r$, $0 \leq r \leq d$ ($r = 1$ corresponds to a periodic orbit).

2. The invariant torus is reducible; that is, the time-dependent linear equations which describe the linearization of the flow on the torus (e.g. $\dot{x} = A(\phi + \omega t)x$) can be transformed into linear, constant coefficient equations $\dot{y} = \hat{A}y$. It is known that reducibility holds automatically for periodic orbits due to Floquet theory (cf. §4.1). For invariant tori with more degrees of freedom there are various positive results concerning when a system is reducible (see, for example, [69, 26, 36, 71, 72, 70, 101, 123]); however, the question of reducibility remains open in general.

3. The initial periodic orbit or invariant torus is isotropic; that is the symplectic form evaluates to zero everywhere on it. Any one-dimensional manifold of a symplectic vector space is isotropic so the property holds automatically for periodic orbits.

There is a canonical change of coordinates such that the initial Hamiltonian can be written as a function of the coordinates $\hat{\theta}, \hat{I}, x, y$ with $\hat{\theta}, \hat{I} \in \mathbb{C}^r$, $x, y \in \mathbb{C}^m$, $d = r + m$, and $z^\top = (x^\top, y^\top)$. Here $\hat{\theta}$ and x are the position variables, with $\hat{\theta}$ being the angle coordinates on the periodic orbit or torus. $\hat{I}$ and y are their respective conjugate momenta. (We

use hats to denote variables pertaining to the initial torus. The x and y coordinates are the "normal" directions to the torus.) Since the original Hamiltonian was assumed to be analytic, it has a Taylor expansion (about $z = 0$, $\hat{I} = 0$). For periodic orbits, Floquet theory ensures that the expansion has constant coefficients for the $\hat{I}$ and $z^\top z$ terms and that the only linear term is $\hat{\omega}^{(0)}\hat{I}$ where $\hat{\omega}^{(0)} \in \mathbb{R}$ is the frequency of the periodic orbit. (For invariant tori of dimension two, or greater, the frequency is replaced by a frequency vector $\hat{\omega}^{(0)} \in \mathbb{R}^r$ and the linear term is $\hat{\omega}^{(0)\top}\hat{I}$.) More generally, the assumption of linear, reducible flow on the invariant torus ensures that the initial Hamiltonian can be put into the semi-normal form

$$\mathcal{H}(\hat{\theta}, x, \hat{I}, y) = \hat{\omega}^{(0)\top}\hat{I} + \frac{1}{2}z^\top \mathcal{B} z + \mathcal{H}_*(\hat{\theta}, x, \hat{I}, y). \tag{4.2}$$

This is sometimes referred to as Floquet form. The terms in the Taylor expansion of $\mathcal{H}_*$ begin at second order in $\hat{I}$ and z. The assumption that the flow on the torus can be reduced to the case of constant coefficients means that $\mathcal{H}_*$ has no quadratic terms in the z variables — all such terms are included in $\frac{1}{2}z^\top \mathcal{B} z$. In these variables $\mathcal{B}$ is a symmetric $2m \times 2m$ complex matrix. $\mathcal{H}_*$ is analytic with respect to all its arguments and is periodic in $\hat{\theta}$.

We also assume that:

4. The analyticity of $\mathcal{H}_*$ holds in a neighbourhood of $z = 0$, $\hat{I} = 0$ (the periodic orbit/torus is assumed to be centered about this point — if it is not, then a change of variables can be used to reduce to this case) and in a complex strip about the variable $\hat{\theta}$, that is for $|\mathrm{Im}(\hat{\theta}_j)| \leq \rho, \quad j = 1, 2, \ldots, r, \quad \rho \in \mathbb{R}$. Also, the matrix $J_m\mathcal{B}$ is diagonal with distinct eigenvalues

$$\lambda^\top = (\lambda_1, \ldots, \lambda_m, -\lambda_1, \ldots, -\lambda_m),$$

where J_m is the canonical symplectic form on $\mathbb{C}^{2m}$ (cf. equation (1.13)).

We will also require that the periodic orbit/torus satisfies a strong non-resonance condition and that the normal "frequencies" λ_j, $j = 1, \ldots, 2m$ satisfy a non-degeneracy condition. We delay giving the details of these conditions until section 4.5 where they arise naturally.

The method is as follows: we begin with the Hamiltonian in the form of equation (4.2). Before considering any perturbation it is helpful to put the initial Hamiltonian into a (semi-)normal form. One does this by expanding $\mathcal{H}_*$, the higher order part of (4.2), as a power series in $\hat{I}$ and z about $\hat{I} = 0$, $z = 0$. We get

$$\mathcal{H}_* = \sum_{p \geq 2} \mathcal{H}_p^{(0)},$$

where the degree p of a monomial $z^l \hat{I}^j$ is defined as $p = |l|_1 + 2|j|_1$ and where $\mathcal{H}_p^{(0)}$ are

homogeneous polynomials of degree p;

$$\mathcal{H}_p^{(0)} = \sum_{\substack{l\in\mathbb{N}^{2m},\ j\in\mathbb{N}^r,\\ |l|_1+2|j|_1=p}} h_{l,j}^{(0)}(\hat{\theta}) z^l \hat{I}^j.$$

The periodic coefficients $h_{l,j}^{(0)}(\hat{\theta})$ are defined by their Fourier series,

$$h_{l,j}^{(0)}(\hat{\theta}) = \sum_{k\in\mathbb{Z}^r} h_{l,j,k}^{(0)} \exp(ik^\top \hat{\theta}). \tag{4.3}$$

It is then possible to use three steps of an iterative KAM-like procedure to rewrite the initial Hamiltonian (4.2). Individual monomials in the expansion of $\mathcal{H}_*$ can be eliminated with a convergent change of variables using a generating function. Each step involves a generating function of the form

$$S^{(n)}(\hat{\theta}, x, \hat{I}, y) = \sum_{\substack{l\in\mathbb{N}^{2m},\ j\in\mathbb{N}^r,\\ |l|_1+2|j|_1=n}} s_{l,j}^{(n)}(\hat{\theta}) z^l \hat{I}^j, \quad n = 3, 4, 5, \tag{4.4}$$

where the periodic coefficients $s_{l,j}^{(n)}(\hat{\theta})$ are defined by their Fourier coefficients allowing us to give an expansion for $S^{(n)}$ based on $\mathcal{H}_n^{(n-3)}$;

$$s_{l,j,k}^{(n)} = \frac{h_{l,j,k}^{(n-3)}}{ik^\top \hat{\omega}^{(0)} + l^\top \lambda}. \tag{4.5}$$

The procedure is similar to that used at each step of the usual KAM method (see section 3.3) with a point of difference being that the small divisors which appear in the construction of the generating function take the form $ik^\top \hat{\omega}^{(0)} + l^\top \lambda$, with $k \in \mathbb{Z}^r \setminus \{0\}$, $l \in \mathbb{N}^{2m}$. To ensure convergence, the Diophantine condition

$$\left| ik^\top \hat{\omega}^{(0)} + l^\top \lambda \right| \geq \frac{\mu_0}{|k|_1^\gamma}, \quad |l|_1 \leq 2 \tag{4.6}$$

is assumed to hold for $k \in \mathbb{Z}^{r+s} \setminus \{0\}$, $l \in \mathbb{N}^{2m}$, $|l|_1 \leq 2$. This differs from the usual non-resonance condition of KAM theory in that it includes the effect of the normal frequencies λ. At each step $S^{(n)}$ is constructed so that the term H_* satisfies the two following conditions for monomials of degree$= 3, 4, 5$:

P1 The coefficients of the monomials $(z, \hat{I})$ (degree 3) and $(z, \hat{I}, \hat{I})$ (degree 5) are zero.

P2 The coefficients of the monomials $(z, z, \hat{I})$ and $(\hat{I}, \hat{I})$ (both of degree 4) do not depend on $\hat{\theta}$ and the coefficients of $(z, z, \hat{I})$ vanish, except for the trivial resonant terms.

The Diophantine condition (4.6) must hold in order that the procedure converges and that

the above two conditions can be satisfied. The resulting Hamiltonian has the form

$$\mathcal{H} = \hat{\omega}^{(0)\top}\hat{I} + \frac{1}{2}z^\top \mathcal{B} z + \frac{1}{2}\hat{I}^\top \mathcal{C}\hat{I} + H_*(\hat{\theta}, x, \hat{I}, y) \tag{4.7}$$

where $\mathcal{C}$ is a constant matrix with $\det(\mathcal{C}) \neq 0$ and where $H_*(\hat{\theta}, x, \hat{I}, y)$ satisfies the conditions **P1** and **P2**.

These manipulations of the Hamiltonian to reach the form (4.7) are purely formal in the sense that they are necessary to the proof of theorem 4.6 and to precisely state the assumptions of that theorem, but the symplectic integrator, whose behaviour we are ultimately interested in, is applied directly to the original Hamiltonian without first manipulating it into the form (4.2) or (4.7).

In the following section we consider the effect of applying a symplectic integrator to the Hamiltonian system: we try to find a modified Hamiltonian which is an $\mathcal{O}(h^p)$ perturbation of the original, and whose time-h flow is given by the trajectory of the symplectic integrator.

4.4 Embedding a map in the flow of a modified vector field

Associated with the original Hamiltonian[5] $\mathcal{H}$ is the real analytic vector field $X_\mathcal{H} = f(u)$, $u^\top = (\hat{\theta}^\top, x^\top, \hat{I}^\top, y^\top)$. When a symplectic integrator with step size h is applied to the vector field f it induces a symplectic map $\Phi_{h,f}$. In order to study whether periodic orbits of $f(u)$ persist in the numerical solution given by the integrator we want to embed $\Phi_{h,f}$ in a modified vector field close to $f(u)$ and ask whether the modified vector field still contains a periodic orbit. We assume that the symplectic integrator is given by a one-step method, $\Phi_{h,f}$, analytic in both h and u. Iterating the numerical method gives a numerical trajectory $\{u_n\}$ by generating the sequence of vectors u_n:

$$u_{n+1} = \Phi_{h,f}(u_n), \quad n = 0, 1, \ldots \quad u_0 = u(0).$$

The problem of embedding the map $\Phi_{h,f}$ in a flow means finding an analytic, modified vector field $\tilde{f}$ which exactly interpolates u_n.

It is well known that it is possible to find an autonomous vector field whose flow is close to the numerical trajectory — we already proved in section 1.4.2 (theorem 1.4) that a local modified Hamiltonian exists for a symplectic integrator applied to a Hamiltonian vector field. The following theorem, presented in [98] states that there is always a local modified vector field which comes exponentially close to interpolating the numerical solution.

Theorem 4.3. *[98] Let f be an analytic vector field in some open domain $\mathcal{D} \subset \mathbb{C}^d$ around the trajectory, and with a corresponding bound $\|f\|_\mathcal{D}$ and let $\varphi_{h,f}$ be the time-h flow of*

[5] We use the notation $\mathcal{H}$ for the unperturbed Hamiltonian both before and after any change of coordinates and/or transformation to the semi-normal/Floquet form (4.2) or (4.7).

f. Let $u_1 = \Phi_{h,f}(u_0) = u_0 + hf(u_0) + \mathcal{O}(h^2)$ be an approximation produced by a one-step method. Then there exists an autonomous modified vector field $\tilde{f}$, bounded on the smaller domain $\tilde{\mathcal{D}} \subset \mathcal{D}$ such that

$$\|u_1 - \varphi_{h,\tilde{f}}(u_0)\| = \mathcal{O}(h \exp(-h_0/h\|f\|_{\mathcal{D}}))$$

for a sufficiently small step size h and where the positive constant h_0 depends on the difference between $\mathcal{D}$ and $\tilde{\mathcal{D}}$ and on the method.

Proofs of theorem 4.3 are given by Hairer and Lubich [51] for any B-series methods (e.g. Runge-Kutta methods), by Benettin and Giorgilli [11] for any symplectic map, and, in a general setting, by Reich[6] [117]. Iterating the bound in theorem 4.3 for each step of a numerical integrator, one sees that the numerical trajectory stays exponentially close to $\varphi_{nh,\tilde{f}}$, the flow of the modified vector field, for some finite time. Unfortunately this result is too weak for our intended use since a trajectory which is close to an invariant curve may diverge from it and may do so after only a short time in the case of exponentially divergent systems. For an arbitrary numerical trajectory, an autonomous flow interpolating the trajectory need not exist, and, in general, rarely does [110]. The following proposition from [98] is an example of the failure of maps, or diffeomorphisms, to embed into flows. Further discussion and examples of this point can be found in the book by Banyaga [8, 1.3.6].

Proposition 4.4. *[98] There exist vector fields f and one-step methods $\Phi_{h,f}$ for which no time-independent vector field $\tilde{f}$ exists with time-h flow $\varphi_{h,\tilde{f}}$ equal to $\Phi_{h,f}$.*

Proof. Consider the following special case of a diffeomorphism defined on the circle $\mathbb{T}$. First, note that arbitrarily close to the identity there is a diffeomorphism $\Phi : \mathbb{T} \to \mathbb{T}$ with unstable period$-s$ points. That is, $\Phi^s(u_0) = \Phi^{s-1} \circ \Phi(u_0) = u_0$ and $|D\Phi^s(u_0)| > 1$. Therefore, the Jacobian $D\Phi^{ns}(u_0) = (D\Phi^s(u_0))^n$ can be made arbitrarily large for sufficiently large n.

On the other hand, if the map $\Phi = \varphi_{h,\tilde{f}}$ is without fixed points, then so too is the vector field $\tilde{f}$ which generates $\varphi_{h,\tilde{f}}$. Then $t_0 = \int_{\mathbb{T}} 1/|\tilde{f}(s)|\mathrm{d}s$ is finite and $\varphi_{t_0,\tilde{f}} = \mathrm{Id}$. For $n \in \mathbb{N}$, the decomposition $n = t_0\lfloor n/t_0 \rfloor + r_n$, $0 \le r_n \le t_0$ implies $\Phi^n = \varphi_{n\tilde{f}} = \varphi_{r_n,\tilde{f}}$, hence, $\varphi_{n,\tilde{f}}$ is uniformly bounded in C^1 and so $D\Phi^n$ is continuous and is also bounded. This is in contrast to the case for maps. □

The proof above is an example of the sort of topological mismatch that occurs between flows and diffeomorphisms. Most diffeomorphisms have periodic points which are isolated from other periodic points with the same period, while a periodic point of a flow is isolated only if it is a stationary point.

[6]The results concerning preservation of structural properties in [117] must be treated with caution as counter-examples are known for some cases [59].

If, however, we allow the modified vector field to be time dependent then the inclusion of an (analytic) symplectic map in an (analytic) Hamiltonian flow *is* possible. The first publications to this effect are by Douady [34, 35] (however, these are not easily accessible, the first being a Ph.D. thesis, and both being in French).

Kuksin and Pöschel [77] prove that inclusion in an analytic Hamiltonian flow with periodic time dependence is possible for analytic symplectic maps which are perturbations of integrable systems. They also suggest that such a result can be used to show preservation of lower dimensional tori by symplectic maps — by combining the non-autonomous inclusion result with a result from Pöschel [113] on preservation of elliptic lower dimensional tori of Hamiltonian flows. Compared with that in [113], the theorem by Jorba and Villaneuva in [73], which we make use of, has the advantage that it holds for lower dimensional tori with any combination of elliptic and hyperbolic eigenvalues for the normal directions, and is hence more general.

Pronin and Treschev [114] use a time-averaging procedure to construct an analytic, non-autonomous, periodic flow which exactly interpolates analytic maps isotopic to the identity[7] If the original map is symplectic, then the flow is Hamiltonian, (similarly for volume preserving or reversible maps). The theorem also holds for maps without such geometric properties.

If, in the symplectic case, the original map is close to integrable then so too is the Hamiltonian flow associated with it and the orders of the closeness are the same for both. More generally, if the map Φ is $\mathcal{O}(\varepsilon)$ close to a map $\hat{\Phi}$, where $\hat{\Phi}$ is already included in the flow of a periodic analytic vector field $\hat{X}$, then Φ can be included in a vector field X which is $\mathcal{O}(\varepsilon)$ close to $\hat{X}$. In the case of a symplectic integrator, of order p, applied to a Hamiltonian vector field X_H, which contains a periodic orbit or lower dimensional torus, we can take the time-h flow of the original vector field to be $\hat{X}$, that is $\varphi_{h,X_H} = \hat{\Phi}$. The map Φ_{h,X_H} generated by the symplectic integrator is $\mathcal{O}(h^p)$ close to the exact flow map $\hat{\Phi}$, and, hence, Φ_{h,X_H} can be included in a the flow of a Hamiltonian vector field $X = X_{\tilde{H}}$ which is $\mathcal{O}(h^p)$ close to the original flow which contained the periodic orbit or invariant torus.

In [96, 98] Moan presents the following theorem which gives estimates on the size of the non-autonomous component of the modified vector field $\tilde{f} = X_{\tilde{H}}$. The proof of this theorem remains unpublished; though it is available as a preprint [97].

Theorem 4.5. *[96, 97, 98] Let $\Phi_{h,f}$ be a one-step method and assume that $f(u)$ is analytic for $u \in \mathcal{D} \subset \mathbb{C}^d$. Then there exists a modified vector field*

$$\tilde{f}(u,t,h) = f(u) + \varepsilon r_1(u) + \varepsilon r_2(u,t;h) \tag{4.8}$$

[7] Two smooth maps $\Phi_i : M \to M'$, $i = 0, 1$ of manifolds M, M' are called *isotopic* if there exists a family of maps $\hat{\Phi}_s : M \to M'$ of the same smoothness class and continuous in the parameter $s \in [0,1]$, such that $\hat{\Phi}_0 = \Phi_0$ and $\hat{\Phi}_1 = \Phi_1$.

analytic in $\tilde{\mathcal{D}} \subset \mathcal{D}$, analytic and h-periodic in t and with a flow that exactly interpolates the numerical trajectory $\{u_n\}$ for all time. Additionally, if the step size is sufficiently small then the time-dependent term is exponentially small in h.

More precisely, for

$$h\|f\|_{\delta_1+\delta_2} < \frac{2\pi\delta_2}{e}$$

the size of the non-autonomous term is bounded by

$$\|\varepsilon r_2\|_{\delta_1} \leq C \cdot \exp\left(\frac{-2\pi\delta_2}{eh\|f\|_{\delta_1+\delta_2}}\right),$$

where

$$\|f\|_\delta = \sup_{z\in\mathcal{D}_\delta(x)} |f(z)|_\infty,$$

and

$$\mathcal{D}_\delta(x) = \{z \in \mathbb{C}^d : |z_i - x_i| \leq \delta,\ i = 1, \ldots, d\}, \quad x \in \mathbb{R}^d.$$

One can now see the $\mathcal{O}(h\exp(-h_0/h\|f\|_{\mathcal{D}}))$ term in theorem 4.3 as being a consequence of the non-autonomous term εr_2 in theorem 4.5.

By theorem 4.5, and the results of Pronin and Treschev *et al.* we know that there exists an analytic vector field $\tilde{f}(u,t;h) = f(u) + \varepsilon r_1(u) + \varepsilon r_2(u,t;h)$ which is also analytic and h-periodic in t. This modified vector field exactly interpolates the numerical trajectory $\{u_n\}$ and is symplectic. Since the non-autonomous perturbation is periodic in the time-like variable, we can, by Floquet's theorem write the associated Hamiltonian as

$$\mathcal{H}_{\text{pert}} = \tilde{\omega}^{(0)}\tilde{I} + \varepsilon\tilde{\mathcal{H}}(\theta, x, I, y, \varepsilon),$$

where $\varepsilon(r_1+r_2) = J^{-1}\nabla\mathcal{H}_{\text{pert}}$, and where we have extended the phase space of the Hamiltonian system to include the periodic non-autonomous component as a new time/angle variable $\tilde{\theta}$, with $\theta^\top = (\hat{\theta}^\top, \tilde{\theta})$, and with $\tilde{I}$ conjugate to $\tilde{\theta}$, and $I^\top = (\hat{I}^\top, \tilde{I})$. The frequency of the new angle variable is $\tilde{\omega}^{(0)} = \frac{2\pi}{h}$, $\omega^{(0)\top} = (\hat{\omega}^{(0)\top}, \tilde{\omega}^{(0)})$.

The perturbed Hamiltonian associated with the modified vector field $\tilde{f} = f + \varepsilon(r_1 + r_2) = X_H = X_{\mathcal{H}+\mathcal{H}_{\text{pert}}}$ can be written as

$$H(\theta, x, I, y, \varepsilon) = \underbrace{\hat{\omega}^{(0)\top}\hat{I} + \tilde{\omega}^{(0)}\tilde{I}}_{=\omega^{(0)\top}I} + \frac{1}{2}z^\top\mathcal{B}z + \frac{1}{2}\hat{I}^\top\mathcal{C}\hat{I} + H_*(\hat{\theta}, x, \hat{I}, y) + \varepsilon\tilde{\mathcal{H}}(\theta, x, I, y, \varepsilon). \quad (4.9)$$

The dependence of $\tilde{\mathcal{H}}$, and hence H, on ε is due to the dependence of the perturbation size $\|\varepsilon(r_1 + r_2)\|$ of the modified vector field on the step size h. However, for the result which follows, we don't make any use of ε as a parameter.

4.5 Main result — periodic orbits are preserved

Having found a perturbed Hamiltonian whose flow interpolates the numerical trajectory we are in a position to apply a KAM-type method to H in order to prove that it still contains the periodic orbit/lower dimensional torus of the initial Hamiltonian $\mathcal{H}$. First, however, we must specify a condition which will be required to hold for the modified modified Hamiltonian from (4.9). In order to do so, it is necessary to first give further details of the procedure in [73] where a KAM theorem for lower dimensional tori is given. We begin by recalling the perturbed Hamiltonian (4.9):

$$H(\theta, x, I, y, \varepsilon) = \omega^{(0)\top} I + \frac{1}{2} z^\top \mathcal{B} z + \frac{1}{2} \hat{I}^\top \mathcal{C} \hat{I} + H_*(\hat{\theta}, x, \hat{I}, y) + \varepsilon \tilde{\mathcal{H}}(\theta, x, I, y, \varepsilon)$$

and expanding the perturbation $\varepsilon\mathcal{H}$ in a power series about $\hat{I} = 0$, $z = 0$. Doing so allows us to group together the terms of the initial Hamiltonian and the perturbation, giving the following expression for the Hamiltonian (without explicitly writing the ε dependence)

$$H(\theta, x, I, y) = \tilde{\omega}^{(0)\top} \tilde{I} + H^*(\theta, x, \hat{I}, y), \tag{4.10}$$

where

$$H^* = a(\theta) + b(\theta)^\top z + c(\theta)^\top \hat{I} + \frac{1}{2} z^\top B(\theta) z + \hat{I}^\top E(\theta) z + \frac{1}{2} \hat{I}^\top C(\theta) \hat{I} + \Omega(\theta, x, \hat{I}, y), \tag{4.11}$$

and where Ω includes all the higher order terms in the expansion. The terms $b(\theta)$ and $E(\theta)$ are $\mathcal{O}(\varepsilon)$, while $a(\theta)$, $c(\theta)$, $B(\theta)$ and $C(\theta)$ are $\mathcal{O}(\varepsilon)$ close to $\bar{a}$, $\hat{\omega}^{(0)}$, $\mathcal{B}$, and $\mathcal{C}$ respectively. ($\bar{a}$ is the angular average of the periodic function $a(\theta)$, cf. equation (3.13).)

The idea is to then use a generating function to give a canonical change of coordinates and to kill one power of ε with a procedure similar to that of Kolmogorov [76], and described in section 3.3. Although the terms in (4.11) don't initially depend on θ they do during the iteration. The smallness of ε and the Diophantine condition (4.6) satisfied by the initial torus means the first step of the procedure can be taken with no small divisor problems and the initial generating function S can be specified. The resulting Hamiltonian is

$$H^{(1)} = H \circ X_S = \tilde{\omega}^{(0)\top} \tilde{I} + H^{(1)*}(\theta, x, \hat{I}, y), \tag{4.12}$$

with

$$\begin{aligned} H^{(1)*}(\theta, x, \hat{I}, y) = {} & a^{(1)}(\theta) + b^{(1)}(\theta)^\top z + c^{(1)}(\theta)^\top \hat{I} \\ & + \frac{1}{2} z^\top B^{(1)}(\theta) z + \hat{I}^\top E^{(1)}(\theta) + \frac{1}{2} \hat{I}^\top C^{(1)}(\theta) \hat{I} + \Omega^{(1)}(\theta, x, \hat{I}, y). \end{aligned}$$

If we rewrite the Hamiltonian (4.12) in the original form (4.9) we have

$$H^{(1)} = \omega^{(0)\top} I + \frac{1}{2} z^\top \mathcal{B}^{(0)}(\varepsilon) z + \frac{1}{2} \hat{I}^\top \mathcal{C}^{(0)}(\theta, \varepsilon) \hat{I} + H_*^{(0)}(\hat{\theta}, x, \hat{I}, y) + \varepsilon^2 \tilde{\mathcal{H}}(\theta, x, I, y, \varepsilon), \quad (4.13)$$

where $\mathcal{B}^{(0)}$, $\mathcal{C}^{(0)}$ and $H_*^{(0)}$ are all order ε close to the quantities $\mathcal{B}$, $\mathcal{C}$ and H^* from before the iteration, and where the properties **P1** and **P2** may no longer hold for $H_*^{(0)}$. The dependence of $\mathcal{B}^{(0)}$ and $\mathcal{C}^{(0)}$ on ε is due to the perturbation size affecting the choice of the generating function and, hence, the coefficient matrices of the new system. However, ε has been fixed throughout this procedure. Although we have not needed to make use of it, the iteration preserves the C^2 dependence of the Hamiltonian on the perturbation size ε. That is, the C^2 dependence of the Hamiltonian perturbation on ε has not played a role so far. (If the dependence was initially C^2, however, this property is preserved by the step above.)

We would now like to repeat the step above in order to further reduce the size of the perturbation, however, since the normal frequencies change during each step, we can no longer be sure that the Diophantine property which allows convergence, by preventing the small divisor problem, will hold for the eigenvalues of $J_m \mathcal{B}^{(0)}$. In order to control these eigenvalues at each step we need to introduce a new parameter. One possible parameter is the perturbation size ε. However we can't use it since, although we have sufficient continuity with respect to ε, it depends on the step size h which must remain fixed throughout the iteration. Instead, we introduce the frequencies of the invariant torus, or periodic orbit, as a parameter $\hat{\omega}$, or, more precisely, the difference between the perturbed frequencies and the initial frequencies $\hat{\omega} - \hat{\omega}^{(0)}$. We also introduce the change of variables $\hat{I} \mapsto \hat{I} + \mathcal{C}^{-1}(\hat{\omega} - \hat{\omega}^{(0)})$ and the parameter vector $\varphi^\top = (\hat{\omega}^\top, \varepsilon)$, $\varphi^{\top(0)} = (\hat{\omega}^{(0)\top}, 0)$. With the change of variables (4.13) becomes

$$\begin{aligned} H^{(1)}(\theta, x, I, y, \varphi) \;=\; & \tilde{\omega}^{(0)\top} \tilde{I} + \hat{\omega}^{(0)\top}(\hat{I} + \mathcal{C}^{-1}(\hat{\omega} - \hat{\omega}^{(0)})) + \frac{1}{2} z^\top \mathcal{B}^{(0)} z \\ & + \frac{1}{2}(\hat{I} + \mathcal{C}^{-1}(\hat{\omega} - \hat{\omega}^{(0)}))^\top \mathcal{C}^{(0)}(\hat{I} + \mathcal{C}^{-1}(\hat{\omega} - \hat{\omega}^{(0)})) \\ & + H_*^{(0)}(\theta, x, \hat{I} + \mathcal{C}^{-1}(\hat{\omega} - \hat{\omega}^{(0)}), y, \varepsilon) \\ & + \varepsilon^2 \mathcal{H}^{(0)}(\theta, x, \hat{I} + \mathcal{C}^{-1}(\hat{\omega} - \hat{\omega}^{(0)}), y, \varepsilon). \end{aligned}$$

Now, if we expand and use the fact that $H_*^{(0)}$ is $\mathcal{O}(\varepsilon)$ close to H_* which is in semi-normal form we get

$$H^{(1)} = c^{(1)}(\varphi) + \omega^\top I + \frac{1}{2} z^\top \mathcal{B}^{(1)}(\varphi) z + \frac{1}{2} \hat{I}^\top \mathcal{C}^{(1)}(\theta, \varphi) \hat{I} + H_*^{(1)}(\theta, x, \hat{I}, y, \varphi) + \tilde{H}^{(1)} \quad (4.14)$$

where $\tilde{H}^{(1)}$ contains all the terms that are of order $(\varphi - \varphi^{(0)})^2$ and higher.

By construction, the matrix $J_m \mathcal{B}^{(1)}$ is diagonal. Using the C^2 differentiability with

respect to φ, the eigenvalues of $J_m\mathcal{B}$ can be written as

$$\lambda_j^{(1)}(\varphi) = \lambda_j + iu_j\varepsilon + iv_j^\top(\hat{\omega} - \hat{\omega}^{(0)}) + \tilde{\lambda}_j^{(1)}(\varphi), \tag{4.15}$$

for $j = 1, \ldots, 2m$ with $u_j \in \mathbb{C}$ and $v_j \in \mathbb{C}^r$ and where the Lipschitz constant of $\tilde{\lambda}_j^{(1)}$ on the set $\mathcal{E}^{(1)} := \{\varphi \in \mathbb{R}^{r+1} : |\varphi - \varphi^{(0)}| \leq \nu,\ 0 \leq \nu \leq 1\}$ is of $\mathcal{O}(\nu)$. The $iu_j\varepsilon$ term can be fixed to give an ε independent result.

The remaining condition necessary for our main result can now be given explicitly.

NDC For any j such that $\operatorname{Re}\lambda_j = 0$ we have $u_j \neq 0$ and $\operatorname{Re} v_j \notin \mathbb{Z}^r$. Moreover, these same conditions hold for $u_{j,l} := u_j - u_l$ and $v_{j,l} := v_j - v_l$ for any $j \neq l$ such that $\operatorname{Re}(\lambda_j - \lambda_l) = 0$.

At last, we are in a position to state our main result.

Theorem 4.6. *Consider a Hamiltonian of the form (4.7) which contains a periodic orbit (or respectively, an r-dimensional invariant torus) about $z = 0$, $\hat{I} = 0$ and satisfying the following assumptions*

(i) H_* is analytic with respect to $(\hat{\theta}, x, \hat{I}, y)$ about $z = 0$, $\hat{I} = 0$ and satisfies the conditions **P1** *and* **P2** *of section 4.3.*

(ii) $\mathcal{B}$ is a constant symmetric matrix such that $J_m\mathcal{B}$ is diagonal with distinct eigenvalues $\lambda^\top = (\lambda_1, \ldots, \lambda_m, -\lambda_1, \ldots, -\lambda_m)$.

(iii) $\mathcal{C}$ is a constant symmetric matrix with non-zero determinant.

(iv) For $\mu_0 > 0$ and $\gamma > 1$ ($\gamma > r$ for the r-dimensional torus case) the following Diophantine condition holds

$$\left| ik^\top\omega^{(0)} + l^\top\lambda \right| \geq \frac{\mu_0}{|k|_1^\gamma}, \quad k \in \mathbb{Z}^{r+s} \setminus \{0\}, \quad l \in \mathbb{N}^{2m}, \quad |l|_1 \leq 2.$$

Then, under the ε-independent version of the non-degeneracy condition **NDC**, *which is given below, the following assertion holds.*

Given a fixed ε satisfying $0 \leq \varepsilon \leq R_0^{\frac{\gamma}{\gamma+1}}$ for $R_0 \in \mathbb{R}$ small enough, there exists a Cantor set $\mathcal{W}_*(\varepsilon, R_0) \subset \{\hat{\omega} \in \mathbb{R} : |\hat{\omega} - \hat{\omega}^{(0)}| \leq R_0\} =: \mathcal{V}(R_0)$ ($\hat{\omega} \in \mathbb{R}^r$ for the r-dim. torus) such that for every $\hat{\omega} \in \mathcal{W}_*(\varepsilon, R_0)$ the perturbed Hamiltonian H (as given by 4.9) corresponding to this fixed value of the perturbation size ε has a reducible 2-dimensional ($(r+1)$-dim.) invariant torus with a vector of frequencies $\omega^\top = (\hat{\omega}^\top, \tilde{\omega}^{(0)})$ on the torus. Moreover, if R_0 is small enough (depending on σ) then for $0 < \sigma < 1$, $\operatorname{mes}(\mathcal{V}(R_0) \setminus \mathcal{W}_*(\varepsilon, R_0)) \leq \exp\left(-R_0^{\frac{-\sigma}{\gamma+1}}\right)$ where $\operatorname{mes}(A)$ denotes the Lebesgue measure of the set A.

In contrast to Shang's result which gives preservation of a (full dimensional) torus with fixed frequencies for a Cantor set's worth of fixed step sizes, our theorem gives preservation

of a Cantor set's worth of frequencies (close to the initial frequencies) for a Cantor set's worth of fixed step sizes, namely, those step sizes that are strongly non-resonant in the sense of condition *(iv)* of theorem 4.6

Theorem 4.6 above is, in fact, an application of a more general result due to Jorba and Villaneuva [73]:

Theorem 4.7. *[73] Consider a d-degree of freedom Hamiltonian of the form (4.9), containing an r-dimensional invariant torus and where the perturbation $\tilde{\mathcal{H}}$ is quasi-periodic in $s \geq 1$ time-like coordinates $\tilde{\theta} \in \mathbb{C}^s$. Assume that $\tilde{\mathcal{H}}$ is analytic with respect to $(\theta, x, \hat{I}, y)$, $\theta^\top = (\hat{\theta}^\top, \tilde{\theta}^\top)$ about $z = 0$, $\hat{I} = 0$ with 2π periodic dependence on θ for any $\varepsilon \in \mathcal{I}_0 := [0, \varepsilon_0]$, in a domain that is independent of ε. The dependence of $\tilde{\mathcal{H}}$ on ε is assumed to be C^2 and the derivatives of $\tilde{\mathcal{H}}$ with respect to ε are also analytic in $(\theta, x, \hat{I}, y)$ on the same domain. Then, if assumptions (i) to (iv) of theorem 4.6 are satisfied, along with the full ε-dependent version of* **NDC** *(given above), the following two assertions hold.*

(a) There exists a Cantor set $\mathcal{I}_* \subset \mathcal{I}_0$, such that for every $\varepsilon \in \mathcal{I}_*$ the Hamiltonian H has a reducible $(r+s)$-dimensional invariant torus with a vector of basic frequencies $\omega^{(0)}$.

Moreover, for every $0 < \sigma < 1$, and for $\bar{\varepsilon}$ small enough (depending on σ), $\mathrm{mes}([0, \bar{\varepsilon}] \setminus \bar{\mathcal{I}}_*) \leq \exp(-(1/\bar{\varepsilon})^{\frac{\sigma}{\gamma}})$, where $\mathrm{mes}(A)$ denotes the Lebesgue measure of the set A and where, for every $\bar{\varepsilon}, \bar{\mathcal{I}}_* := \bar{\mathcal{I}}_*(\bar{\varepsilon}) = [0, \bar{\varepsilon}] \cap \mathcal{I}_*$.

(b) Given R_0 small enough and a fixed $0 \leq \varepsilon \leq R_0^{\frac{\gamma}{\gamma+1}}$, there exists a Cantor set $\mathcal{W}_*(\varepsilon, R_0) \subset \{\hat{\omega} \in \mathbb{R}^r : |\hat{\omega} - \hat{\omega}^{(0)}| \leq R_0\} =: \mathcal{V}(R_0)$, such that for every $\hat{\omega} \in \mathcal{W}_*(\varepsilon, R_0)$ the Hamiltonian H corresponding to this fixed value of ε has a reducible $(r+s)$-dimensional invariant torus with vector of basic frequencies ω, $\omega^\top = (\hat{\omega}^\top, \tilde{\omega}^{(0)\top})$.

Moreover, if R_0 is small enough (depending on σ), then for every $0 < \sigma < 1$, $\mathrm{mes}\left(\mathcal{V}(R_0) \setminus \mathcal{W}_*(\varepsilon, R_0)\right) \leq \exp\left(-R_0^{\frac{-\sigma}{\gamma+1}}\right)$.

Proof of theorem 4.6 therefore requires that for a Hamiltonian (4.7) there exists a modified Hamiltonian (4.9) with a corresponding modified vector field as described by theorem 4.5 such that the Hamiltonian (4.9), and in particular the perturbation $\varepsilon\tilde{\mathcal{H}}$, satisfies all the necessary assumptions of theorem 4.7. The proof proceeds via theorem 4.5: the trajectories of the symplectic integrator are interpolated by the flow of a modified Hamiltonian close to the exact Hamiltonian.

Proof. (of theorem 4.6) By theorem 4.5 we have that the modified vector field $\tilde{f}(u, t; h) = f(u) + \varepsilon r_1(u) + \varepsilon r_2(u, t; h)$, and hence $\tilde{\mathcal{H}}$, is analytic with respect to all its arguments and is periodic in the new time-like variable. The C^2 dependence of $\tilde{\mathcal{H}}$ on ε required in theorem 4.7 is not essential to us since we only require that the second result *(b)* of theorem 4.7 holds—this involves fixing the perturbation size ε, not varying it as a parameter to control the normal frequencies.

Properties *(ii)* and *(iii)* of theorem 4.6 (and the analogous properties of theorem 4.7) only apply to the initial Hamiltonian $\mathcal{H}$ (i.e equations (4.2) and (4.7)), and hence are unaffected by the perturbation $\varepsilon(r_1 + r_2)$ to the vector field. Similarly, once the initial frequencies $\hat{\omega}^{(0)}$ are given the Diophantine condition *(iv)* is further affected only by the choice of a (strongly non-resonant) step size h, not by the particular form of the perturbation $R = r_1 + r_2$ (though the choice of h clearly affects the size of R). We can therefore impose condition *(iv)* as an assumption on the frequencies of the initial torus, its normal frequencies and on the step size of the numerical method. These frequencies are preserved in the initial component f of the modified vector field $\tilde{f}$ given by (4.8) and so, the perturbed Hamiltonian $H = \mathcal{H} + \varepsilon\tilde{\mathcal{H}}$ satisfies those assumptions of theorem 4.7 necessary for the result *(b)* of that theorem. Hence, theorem 4.6 holds. □

Part *(a)* of theorem 4.7 seems to suggest that it should be possible to use the perturbation size ε of the modified vector field as a parameter and to achieve a result similar to that of KAM theory for full dimensional tori; namely that periodic orbits or invariant tori are preserved *with their frequencies unchanged*. However, our proof does not allow for this. The parameter ε arises directly from the step size of the symplectic integrator, it also directly affects the frequency of the periodic non-autonomous perturbation. Hence, if one uses the perturbation size, (and thus, the step size), as a parameter, one can no longer ensure that the frequency of the non-autonomous perturbation is not resonant with the existing frequency of the periodic orbit (or the vector of existing frequencies of the lower dimensional torus).

In order to use the perturbation size as a parameter, one would need to show that for a particular step size, the resulting perturbation size was in the Cantor set of values of ε for which the periodic orbit/invariant torus persists *and* that the frequency corresponding to that step size continued to satisfy the strong non-resonance conditions. That is, there must be a non-empty intersection of the Cantor set of step sizes allowed by the perturbation size, and the Cantor set of step sizes allowed by the non-resonance requirement. It does not seem likely that such an intersection would have positive measure.

4.6 Resonant periodic orbits

In this section we investigate the case of a symplectic integrator, applied to a Hamiltonian system containing a periodic orbit, when the step size of the integrator is exactly resonant with the (perturbed) period of the orbit of the modified system. From the results of the previous section we can not expect the orbit to persist in general.

We must be careful to specify exactly which periodic orbit we are referring to. For Hamiltonian systems, periodic orbits occur in one-parameter families. In section 4.4 we saw that applying a symplectic integrator, with step size h, to the vector field of a Hamiltonian system gives a symplectic map which may be included in the flow of a non-autonomous

modified Hamiltonian with period h. That is, the flow of the modified Hamiltonian exactly interpolates the numerical trajectory obtained by applying the symplectic integrator to the original system. (If the time-like coordinate of the periodic perturbation is paired with a conjugate momentum coordinate, the non-autonomous modified Hamiltonian may be written as an autonomous modified Hamiltonian with an additional (periodic) degree of freedom.)

The perturbation to the Hamiltonian consists of two components; an autonomous perturbation of size ε, (where ε is $\mathcal{O}(h^p)$ for an integrator of order p), and a non-autonomous perturbation which is exponentially small in ε. that is

$$\tilde{H}(y,t) = H(y) + \varepsilon \tilde{H}_1(y) + Ce^{-1/\varepsilon}\tilde{H}_2(y,t).$$

The autonomous Hamiltonian $H_1 = H + \varepsilon\tilde{H}_1$ can be thought of as a sort of intermediate modified Hamiltonian. It too has a one-parameter family of periodic orbits. We denote the period of the system H_1 by T_1. Since H_1 is an $\mathcal{O}(\varepsilon)$ perturbation of H, we expect $T - T_1 = \mathcal{O}(\varepsilon)$ too.

Theorem 4.6 of section 4.5 then says that if the original Hamiltonian system H has a periodic orbit with period T, and if the step size of the symplectic integrator is strongly non-resonant with the period of this orbit (along with some additional requirements which are given in the theorem), then the modified Hamiltonian system $\tilde{H}$ has an invariant two-torus and that one frequency on the torus is $2\pi/h$ (the amplitude of motion is small in the variable corresponding to this frequency) while the other frequency of the torus, $1/\tilde{T}$, is close to the frequency of the original periodic orbit, $1/T$.

(Since the numerical trajectory given by the symplectic integrator applied to the original Hamiltonian is the *exact* time-h flow of the modified Hamiltonian, and since the modified Hamiltonian is periodic in the new variable, one can think of the numerical trajectory as being the period-h return map (iso-energetic in the value of the modified Hamiltonian) of the modified flow. As such, the (non-resonant) two-tori of the modified flow are $\tilde{T}$-periodic orbits of the return map, and hence of the integrator.)

Since one frequency of the torus is fixed by the choice of step size, the two-torus is part of a one-parameter Cantor set parameterised by $\tilde{T}$. The gaps in this set occur when theorem 4.6 fails, that is whenever $\tilde{T}$ is (close to) resonant with h. It is this situation which we investigate in this section. That is, we ask whether the discrete system obtained by applying a symplectic integrator to a Hamiltonian system, with a periodic orbit of period T, has a set (or sets) of $\tilde{T}/h$-periodic points where $\tilde{T}$ is $\mathcal{O}(\varepsilon)$ close to T (and $\mathcal{O}\left(e^{-1/\varepsilon}\right)$ close to T_1) and where the periodic points of the discrete system, interpolated by the modified Hamiltonian $\tilde{H}$ are close to the periodic orbit of the original system. Figure 4.5 illustrates this point.

The resonant, one degree of freedom case is well understood; in fact, it is none other than the Poincaré–Birkhoff fixed point theorem [12]. Consider an annulus $A = \{(\theta, I) : 0 \leq$

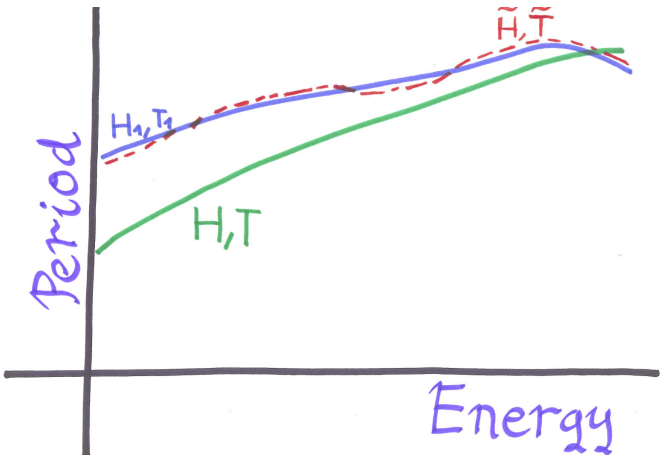


Figure 4.5: Possible one-parameter families of periodic orbits for a Hamiltonian H and an autonomous modified Hamiltonian H_1 which is an $\mathcal{O}(\varepsilon)$ perturbation of H. The (non-autonomous) Hamiltonian $\tilde{H}$ is generated by including the trajectory of a symplectic integrator in the flow of the original Hamiltonian vector field $X_{\tilde{H}}$ and my be thought of as an $\mathcal{O}\left(e^{-1/\varepsilon}\right)$ perturbation of H_1. The set of periodic orbits of $\tilde{H}$ exist for a Cantor set of periods $\tilde{T}$.

$\theta \leq 2\pi, a \leq I \leq b\}$ and an area-preserving twist map $\Gamma : A \to A$, $\Gamma : (\theta, I) \mapsto (\theta + \alpha(I), I)$. Let Γ_ε be an area-preserving perturbation of Γ; that is,

$$\Gamma_\varepsilon : (\theta, I) \mapsto (\theta + \alpha(I) + f(\theta, I, \varepsilon), I + g(\theta, I, \varepsilon))$$

such that for all ε

$$\int_\gamma I \mathrm{d}\theta = \int_{\Gamma_\varepsilon \gamma} I \mathrm{d}\theta$$

for γ any closed curve in A. Then, given any rational number m/n, satisfying $\alpha(a)/2\pi \leq n/m \leq \alpha(b)/2\pi$ there exist $2n$ fixed points of Γ_ε^n satisfying $\Gamma_\varepsilon^n : (\theta, I) \mapsto (\theta + 2\pi m, I)$ for ε sufficiently small. The fixed points of Γ_ε^n (i.e. the period-n points of Γ_ε) are alternately hyperbolic and elliptic. The eigenvalues of Γ_ε^n must satisfy $\lambda\lambda' = 1$ since the map is area preserving. The return map of the original periodic orbit described by Γ has a pair of degenerate eigenvalues $(1, 1)$. Under the perturbation, these split to give eigenvalues for the elliptic and hyperbolic fixed points which satisfy $\lambda' = \lambda^{-1}$. The pairs of eigenvalues are real for the hyperbolic case (i.e. $0 \leq \lambda \leq 1 \leq \lambda'$) and complex for the elliptic case (i.e. $\lambda' = \lambda^{-1}$, $|\lambda'| = |\lambda| = 1$). See figure 4.7 for an illustration.

We conjecture that the periodic orbits of Hamiltonian systems with more degrees of freedom behave in an analogous way when treated with a symplectic map which is resonant with the period of the orbit. More specifically, the return map of a periodic orbit within a d-degree of freedom Hamiltonian system has eigenvalues with a single degenerate pair $\lambda_0 = \lambda_0' = 1$ and $d - 1$ non-degenerate pairs satisfying $\lambda_i \lambda_i' = 1$, for $i = 1, \ldots, d - 1$. Applying

a symplectic integrator whose step size exactly divides the period of a closed orbit, we expect the periodic orbit to be destroyed, leaving $n = \tilde{T}/h$ elliptic and n hyperbolic periodic points, with the eigenvalues corresponding to the degenerate pair of the original orbit splitting as either a real or a complex pair satisfying $\lambda_0' = \lambda_0^{-1} \neq 1$. The remaining $2(d-1)$ eigenvalues are expected to remain of the same type (elliptic or hyperbolic) as they were for the original periodic orbit but, with a small perturbation due to the integrator.

In the remainder of this section, we investigate this numerically. We take, for our symplectic integrator, the leapfrog or Störmer-Verlet method, which, for a separable Hamiltonian, reads:

$$\begin{aligned} q_{n+1/2} &= q_n + \frac{h}{2} p_n, \\ p_{n+1} &= p_n - h \nabla V(q_{n+1/2}), \\ q_{n+1} &= q_{n+1/2} + \frac{h}{2} p_{n+1}. \end{aligned}$$

For a model system, we use the two degree of freedom Hénon–Heiles system given by the Hamiltonian,

$$H(q,p) = T(p) + V(q), \quad T(p) = \frac{1}{2}\|p\|^2 \quad \text{and} \quad V(q) = \frac{1}{2}\|q\|^2 + q_1 q_2^2 - \frac{1}{3} q_1^3,$$

where $q^\top = (q_1, q_2)$, and $p^\top = (p_1, p_2)$. The system is non-integrable[8] and when $H < \frac{1}{6}$ trajectories beginning in a region near the origin are bounded. For higher energies there are unbounded orbits.

We will show below how a Poincaré map is necessary as an intermediate step towards locating periodic orbits, although we stress that it plays no part in the final results. We use an iso-energetic Poincaré section in order to study the system. This involves first fixing the energy of the system (we use initial conditions satisfying $H = 0.1$ throughout) to reduce the system from four to three dimensions. Then, taking a transverse section of the flow to reduce the system by a further dimension leaving a two-dimensional map from the plane to itself. If the flow generating the Poincaré map is symplectic, then so too is the map. However, it is generally not possible to give the Poincaré map analytically. Rather, it must be calculated by (numerically) following the trajectories of points on the section until they return to it. Since, for a fixed step size, numerical trajectories will not necessarily land back on the section after an integer number of steps, it is necessary to use some sort of interpolation to define the numerical Poincaré map. We now sketch the construction of the Poincaré map.

1. Begin at a point (q_1, p_1) on the two-dimensional section defined by $q_2 = 0$ and

[8] Some similar systems, also referred to as Hénon–Heiles systems, are integrable. For example, if the coefficient of the term $q_1 q_2^2$ is negative then the system is integrable.

$H(q_1, q_2, p_1, p_2) = 0.1$.

2. Integrate forward using leapfrog until a pair of consecutive points, $(q, p) = (q_1, q_2, p_1, p_2)$ and $(q', p') = (q'_1, q'_2, p'_1, p'_2)$, are found which satisfy $q_1 < 0$, $q'_1 > 0$ and $p_2, p'_2 > 0$.

3. Use cubic Hermite interpolation and the values (q, p), $f(q, p)$, (q', p') and $f(q', p')$ to estimate the point (q^*, p^*) and the time t^* at which the trajectory crossed the section.

Although the leapfrog integrator is symplectic, the numerical Poincaré map from the algorithm above is not, due to the cubic Hermite interpolation used to find the crossing. However, by continuing from a point on the trajectory, rather than from the new point on the Poincaré section, the "distance from symplecticity" remains uniform when the numerical map is iterated.

Fixed points of the reduced system correspond to periodic orbits of the full four-dimensional system, while closed curves in the Poincaré section indicate invariant tori. The phase portrait for the reduced system is shown in figure 4.6. Elliptic and hyperbolic periodic orbits can be seen near $(q_1, p_1) = (0.27, 0)$ and $(-0.15, 0)$ respectively. The periods of the orbits are roughly 5.76 for the elliptic orbit and 6.47 for the hyperbolic. Invariant tori can also be clearly seen, particularly about the elliptic periodic orbit.

As we mentioned in section 1.3.3, the two-dimensional invariant tori form barriers in the three-dimensional surfaces of constant energy. Hence, for the continuous Hénon-Heiles system the chaotic trajectories are trapped between areas of regular motion due to these tori. For the discrete system due to the leapfrog integrator, the chaotic trajectories can leak across the invariant tori since symplectic integrators do not conserve energy exactly, but rather ensure that trajectories remain close to their initial energy for long times. Despite this, the very slow energy drift observed for symplectic integrators and the observation by Perry and Wiggins [111] that trajectories close to invariant tori remain close for long times, means that we can still hope to achieve an accurate representation of the true dynamics of the Poincaré section.

In order to study the periodic orbits of the discretized Hénon-Heiles system we use the following procedure:

Step 1 Use the numerical Poincaré map with a step size $\hat{h}$ to estimate approximate initial conditions (on the section) for the elliptic and hyperbolic periodic orbits, and to estimate the periods of those orbits.

The numerical Poincaré map should use a step size close to that which will be used for investigating the orbit of the full system in the next step. We used $\hat{h} = 1$ initially. We will discuss later the reasons for using a large step size.

Step 2 Beginning with the initial conditions estimated from the Poincaré map, resonant periodic orbits of the full (i.e. four-dimensional) Hénon-Heiles were calculated, for a

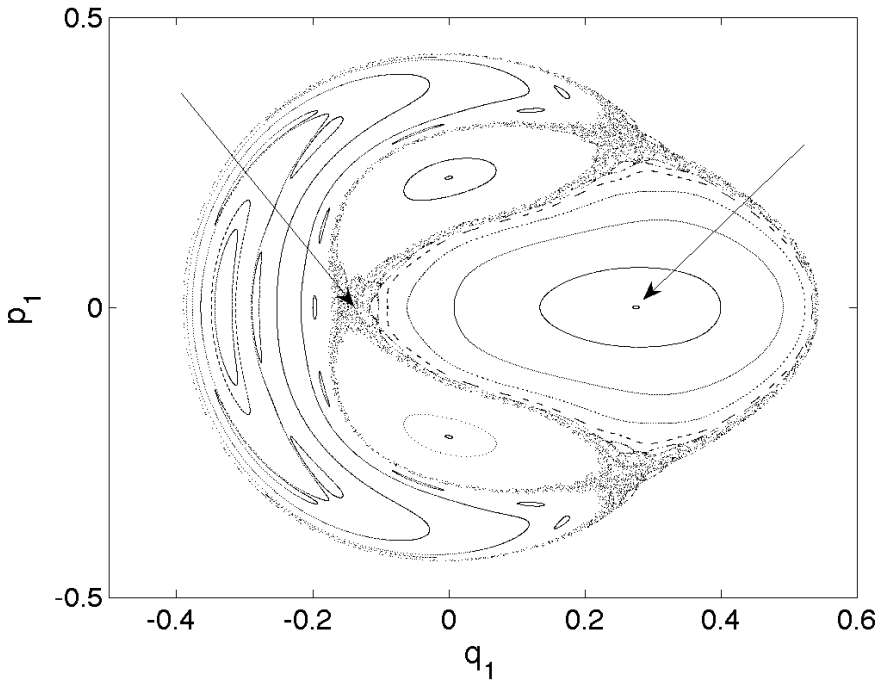


Figure 4.6: The Poincaré section of the Hénon–Heiles system calculated with the leapfrog method for $H = 0.1$, with step size $h = 0.1$. The section shows two fixed points (indicated by arrows)—the approximate locations of two periodic orbits of the full system—a hyperbolic point near $(q_1, p_1) = (-0.15, 0)$ and an elliptic point near $(q_1, p_1) = (0.27, 0)$.

particular step size, by minimizing the function

$$g(x) = |\Phi_h^n(x) - x|\,, \tag{4.16}$$

where Φ_h^n means taking n steps of size h using the leapfrog integrator.

The step size h was chosen in order to give a resonance of a particular order for a specified period. For example, the numerical Poincaré map (computed with $\hat{h} = 1$) indicated that the elliptic periodic orbit had $\tilde{T} \simeq 5.67$. We then fixed the step size at $h = 5.67/6$ and used Newton's method to solve $g(x) = \left|\Phi^6_{5.67/6}(x) - x\right|$ in order to find a period-6 point of the discrete system.

Finding numerical solutions of (4.16) involves overcoming a number of numerical difficulties. Firstly, the starting point of Newton's method must be close to the final solution, especially in the case of the hyperbolic periodic orbit for which nearby trajectories are chaotic and rapidly leave the neighbourhood of the periodic orbit. This necessitated the use of a nonlinear least-squares solver (we used the `lsqnonlin` solver from Matlab) to refine the initial estimates from step 1 before a suitable starting point for the Newton iterations was obtained. The reasons for not using the least-squares minimization to solve

(4.16), without involving Newton iterations, are two-fold: the `lsqnonlin` code did not converge with sufficient accuracy, and we also wished to calculate the eigenvalues of the return map $\Phi_h^n(x)$ which is best done by using information from the Jacobian $D\Phi_h(x)$ calculated during the Newton iterations. An alternative, however, would be to estimate $D\Phi_h^n$ using finite differences after the least-squares minimization.

The second numerical difficulty is that, during the Newton iterations, the condition number of the Jacobian $Dg(x)$ may become large, resulting in a loss of accuracy when solving the linear system of equations $Dg(x_n)\Delta_n = g(x_n)$. We counteract this by using one step of *iterative refinement*. Iterative refinement is an established technique for improving the accuracy of a computed solution $\hat{u}$ of a linear system $Au = b$ [64, Chapter 12]. The procedure has three steps:

i) Compute the residual: $r = b - A\hat{u}$.

ii) Solve $Ad = r$.

iii) Update the solution: $u^* = \hat{u} + d$

(If one wishes to use further iterations, then the procedure is repeated with $\hat{u}$ replaced by u^*.)

We implemented the linear solver using QR factorization of $Dg(x_n)$. This has the advantage that the factorization could be reused for solving the linear system in the iterative refinement step. The pseudo-code for the resulting Newton iterations is as follows:

Algorithm 4.1 (Newton iteration with iterative refinement)**.**

Set the initial quantities Δ_0^*, x_0, and tol.

While $\Delta_n^* > tol$:

Evaluate $g(x_n)$ and $Dg(x_n)$.

Factorize $Dg(x_n) = QR$ such that Q is orthonormal and R is upper triangular.

Solve $R\Delta_n = Q^\top g(x_n)$ for the increment Δ_n.

Compute the residual: $r = Dg(x_n)\Delta_n - g(x_n)$.

Solve $Rd = Q^\top r$ for the refinement d.

Update the increment: $\Delta_n^* = \Delta_n + d$.

Take a Newton step and update the solution: $x_{n+1} = x_n - \Delta_n^*$.

End.

Period-6 and period-12 points of the system were calculated for both the elliptic and the hyperbolic periodic orbits. We do not require the periods of the 6-step and 12-step orbits to be the same since changing the integrator step size, in order to change from six

to twelve steps per period, will generally cause the period $\tilde{T}$ of the modified system to change. Instead, the period is determined by using the numerical Poincaré map with an appropriate step size to determine an approximate value for the period in each case. The period is then fixed at that value and a periodic orbit with exactly that period is found by the minimization.

The positions of the points corresponding to initial conditions for the four pairs of resonant periodic orbits are listed in table 4.1. Results here are rounded to five decimal places (i.e. a value of zero indicates zero to machine precision while a value of 0.00000 indicates a value with magnitude smaller than 0.000005.)

	n	h	q_1	q_2	p_1	p_2	eigenvalues
Ellip.	6	5.67/6	0.24947	0	0	0.40239	-0.26057±0.96545i 0.99278±0.11993i
Ellip.	6	5.76/6	0.21945	-0.15126	0.09332	0.31512	-0.18830±0.98211i 1.12557 0.88844
Hyp.	6	6.47/6	-0.16707	0.24590	0.18003	0.45607	7.47324 0.13381 1.34600 0.74294
Hyp.	6	6.47/6	-0.14100	0.24422	-0.13168	0.22807	0.59441±0.80416i 1.38670 0.72114
Ellip.	12	6.03/12	0.26409	-0.00000	0.00000	0.37047	1.00000±0.00004i 0.04451±0.99901i
Ellip.	12	6.03/12	0.25775	0.088842	-0.04954	0.35192	1.00003 0.99997 0.04451±0.99901i
Hyp.	12	6.77/12	-0.17440	-0.00000	-0.00000	0.45137	3.63872 0.27482 1.00050 0.99950
Hyp.	12	6.77/12	-0.16637	0.12413	0.06132	0.44055	3.63872 0.27482 1.00000±0.00050i

Table 4.1: Periodic points corresponding to one elliptic and one hyperbolic perioic orbit of the Hénon–Heiles system after discretization with $n = 6$ and with $n = 12$ steps of size h per period. Eigenvalues are given for the corresponding return map, Φ_h^n, $n = 6, 12$. Figures are rounded to 5 d.p. but in all cases the pairs of eigenvalues satisfied the property $\lambda\lambda' = 1$ before rounding. A value of 0 indicates zero to machine precision.

Step 3 For each of the periodic orbits found in step 2, the eigenvalues of the linearization of the return map $\Phi_h^n(x)$, $n = 6, 12$ were calculated. In principle this is easy since the Newton iterations already calculate the Jacobian $D\Phi_h^n(x)$ of the return map.

This is done by composing the Jacobians from each step along the periodic orbit: $D\Phi_h^n(x) = D\Phi_h \circ D\Phi_h \circ \cdots \circ D\Phi_h(x)$.

The eigenvalues of $D\Phi_h^n$ allow us to classify the periodic points. The original system has one (degenerate) pair of double one eigenvalues — corresponding to the flow around the periodic orbit — and one pair, corresponding to the flow transverse to the periodic orbit, which is either elliptic (i.e. $\lambda' = \bar{\lambda}$, $\lambda \neq 1$) or hyperbolic (i.e. $\lambda' = 1/\lambda$, $\lambda \in \mathbb{R}$, $\lambda \neq 1$). In both cases $\lambda\lambda' = 1$ due to the symplecticity. Symplectic integrators preserve this property up to machine precision. Under the resonant symplectic discretization, the degenerate pair of eigenvalues may bifurcate into either an elliptic or a hyperbolic pair. The original elliptic or hyperbolic pair remains of the same class, though the eigenvalues will be slightly perturbed. The perturbation of the degenerate pair is also slight, though it changes the class of the eigenvalues.

The smallness of the perturbation to the eigenvalues makes it necessary to choose a large step size in order to see the effect of the discretization. However, as one increases the step size, the condition number of $D\Phi_h$ also increases resulting in a loss of accuracy due to fewer meaningful significant figures in the calculation of the eigenvalues. It also makes the convergence of Newton's method more difficult during the minimization procedure of step 2. Taking smaller step sizes reduces the condition number of $D\Phi_h$ but also reduces the size of the eigenvalue perturbation and requires more steps along the orbit (and hence in the calculation of $D\Phi_h^n(x) = D\Phi_h \circ \cdots \circ D\Phi_h(x)$) which risks the perturbation to the eigenvalues being swamped by the accumulated numerical errors. For the orbits considered here, six or twelve steps per period gave reliable results. Increasing or decreasing the number of steps per period much beyond these numbers would require more sophisticated numerical techniques or increased computational precision (e.g. quadruple precision arithmetic) or both. The calculated eigenvalues are included in table 4.1.

The results in table 4.1 show that in each case a periodic orbit of the original system gives rise to a pair of sets of n-periodic points. As expected, the $(1,1)$ pair of degenerate eigenvalues corresponding to the original periodic orbit splits into an elliptic and a hyperbolic pair; one pair associated with each of the sets of n-periodic points. In all cases, prior to rounding, the eigenvalues satisfy the property $\lambda\lambda' = 1$ to machine precision since the leapfrog method is symplectic. It is worth noting, also, the rapid convergence of the eigenvalues of the return maps towards the eigenvalues of the original periodic orbit as the step size used in the map decreases. For the six step return map, the eigenvalues corresponding to the splitting of the degenerate $(1,1)$ pair differ from one in the first or second significant figure. For the 12 step map, the effect is reduced to the fifth or sixth significant figure. That is, halving the step size reduces the perturbation size by a factor of about 1000.

Step 4 Use a numerical search in the vicinity of the periodic points and along the line segments connecting them, to see if any other set of periodic points with the same

period can be found.

In order to support our claim that the two sets of periodic points found in steps two and three are the unique sets with that period we used numerical searches in the neighbourhood of the periodic orbit of the continuous system. This involved making random perturbations, either to the periodic points, or to a point on the line segment connecting two periodic points. Step 2 was then repeated for the perturbed initial conditions. In all cases, the minimization procedure either returned to one of the existing periodic points, or failed to converge. This suggests that the sets of periodic points which replace the periodic orbits of the continuous system are indeed unique.

Based on the behaviour observed in these numerical experiments, and on generalization from the one degree of freedom case, we make the following conjecture:

Conjecture 4.8 (Break up of resonant periodic orbits)**.** *Consider a Hamiltonian system with $d \geq 2$ degrees of freedom and a periodic orbit $\gamma(t)$ of period T discretized by a symplectic integrator of order p, using step size h. When the period $\tilde{T}$ of the perturbed system generated by the integrator is such that $\tilde{T}/h = n \in \mathbb{N}$, then $\gamma(t)$ is replaced by two sets of n-periodic points; $\mathcal{P}_e$ and $\mathcal{P}_h$. For points x^* in $\mathcal{P}_e$, or $\mathcal{P}_h$, the eigenvalues of $D\left(\Phi_h^n(x^*)\right)$ are of the same type as those of $\varphi_T(x^*)$ with the exception of the $(1,1)$ degenerate pair which becomes elliptic (i.e. $\lambda \neq 1$, $\lambda' = \bar{\lambda}$) for $x^* \in \mathcal{P}_e$, or hyperbolic (i.e. $\lambda \neq 1$, $\lambda' = 1/\lambda$) for $x^* \in \mathcal{P}_h$. The points in $\mathcal{P}_e$ and $\mathcal{P}_h$ are $\mathcal{O}\left(h^p\right)$ away from $\gamma(t)$, evenly spaced and occur alternately about $\gamma(t)$, as illustrated in figure 4.7*

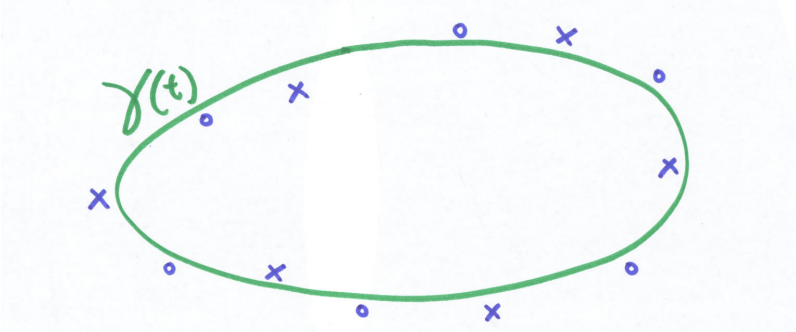


Figure 4.7: Elliptic (*dots*), and hyperbolic (*crosses*), periodic points result from the discretization of a periodic orbit $\gamma(t)$ by a symplectic integrator whose step size is resonant with the period of γ.

A proof of the behaviour conjectured here, on the basis of our numerical study, remains to be undertaken.

The implications for such a conjecture go beyond integration of Hamiltonian ordinary differential equations. For example, it is relevant to the discretization of Hamiltonian partial differential equations as we will now show. The concept of Hamiltonian ordinary

differential equations can be extended to so called *multi-Hamiltonian* partial differential equations where the dependent variable z is a function of multiple independent variables, typically, one space-like variable x and one time-like variable t. The equations of motion for Hamiltonian ordinary differential equations, $\frac{\mathrm{d}y}{\mathrm{d}t} = J^{-1}\nabla H(y)$, extend naturally to multi-Hamiltonian partial differential equations:

$$K\frac{\partial z}{\partial t} + L\frac{\partial z}{\partial x} = \nabla_z S(z), \tag{4.17}$$

where $z \in \mathbb{R}^n$, K and L are any skew-symmetric matrices in $\mathbb{R}^{n\times n}$, and $S : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, typically autonomous (i.e. $S(z)$ does not have any explicit dependence on t or x) [15].

As an example of such an equation, consider the wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = -\nabla_u V(u).$$

Setting $z = \begin{bmatrix} u \\ v \\ w \end{bmatrix}$, $K = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $L = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$, and $S(z) = \frac{1}{2}(w^2 - v^2) - V(u)$, equation (4.17) becomes

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} u_t \\ v_t \\ w_t \end{bmatrix} + \begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} u_x \\ v_x \\ w_x \end{bmatrix} = \begin{bmatrix} -\nabla_u V(u) \\ -v \\ w \end{bmatrix},$$

where we have adopted the notation $\frac{\partial u}{\partial t} = u_t$, *et cetera*. The following system of partial differential equations results:

$$v_t - w_x = -\nabla_u V(u), \tag{4.18}$$

$$-u_t = -v, \tag{4.19}$$

$$u_x = w. \tag{4.20}$$

Differentiating (4.20) and (4.19) with respect to x and t respectively, and substituting the results into (4.18) yields the wave equation in its familiar form:

$$u_{tt} - u_{xx} = -\nabla_u V(u). \tag{4.21}$$

A large class of partial differential equations can be formulated in this fashion including: the nonlinear wave equation, the Boussinesq equation, the Korteweg-de Vries equation, the Klein-Gordon equation, the nonlinear Schrödinger equation, the Padé-II equation, and the Sine-Gordon equation, amongst others [125, 99, 14, 15]. Even so, there is currently no systematic method for constructing the multi-Hamiltonian form of a Hamiltonian partial

differential equation — in fact, it is not even known if such a procedure is possible [124].

Our reason for introducing the multi-Hamiltonian form is to make clear a link between such systems and resonant periodic orbits of Hamiltonian ordinary differential equations. Given a (Hamiltonian) partial differential equation in the form (4.17), with periodic boundary conditions one may search for steady-state solutions of the equation. In this case the time derivative term of (4.17) vanishes and the equation becomes

$$Lz_x = \nabla_z S(z); \tag{4.22}$$

a (Hamiltonian) ordinary differential equation for the space-like variables. The periodic boundary conditions mean that solutions of (4.22) must satisfy $z(0) = z(l)$, where $l \in \mathbb{R}$ is the length of the space domain.

Steady state solutions of (4.17) are periodic solutions of (4.22) with period l (or l/k, $k \in \mathbb{N}$). The problem of finding such solutions is a boundary value problem — in contrast to the initial value problems we consider throughout this thesis. As such, the geometry is quite different from the initial value problems; exact periodicity of the numerical solution is imposed by construction. Solutions satisfying (4.22), but not the boundary conditions, and which are arbitrarily close to being periodic, (or, are periodic with period arbitrarily close, but not equal, to l/k), are *not* solutions of the boundary value problem which steady state solutions of the Hamiltonian partial differential equation (4.17) must satisfy.

Now, assume (4.17) is discretized by a *multisymplectic* integration scheme (i.e. by symplectic integrators in the time, and in the space variables). Preserving steady states of (4.17) is similar to preserving fixed points of an ordinary differential equation, so, if the integrator used for the spatial discretization preserves fixed points (cf. §2.1) then steady states of the multisymplectic discretization of (4.17) give periodic orbits of the discrete version of (4.22), with period $l/k\Delta x$, where Δx is the grid spacing of the spatial discretization, which is assumed to be constant. (If the integrator used for the spatial discretization permits spurious fixed points for ordinary differential equations, then, in the situation above, spurious steady states are possible. This situation should therefore be avoided.)

Conjecture 4.8 then applies, implying that the continuous family of steady state solutions of (4.17), with continuous translation symmetry in x, is replaced, in the discrete system, by two steady states, each approximating one of the continuous steady state solutions. These discrete solutions have a symmetry of translation in space by $\Delta x/2$. In fact, for scalar partial differential equations like the wave equation (4.21) or the nonlinear Schrödinger equation, conjecture 4.8 is known to be true — it is just the Poincaré-Birkhoff fixed point theorem. The resulting situation is sketched in figure 4.8 where the solid green line represents a continuous (exact) solution of the boundary value problem. It is approximated at the grid points by a discrete solution; the blue circles. The dashed green line is the $\Delta x/2$ forward translation of the solid line. It is approximated at the grid points by

the discrete solution represented by the blue crosses.

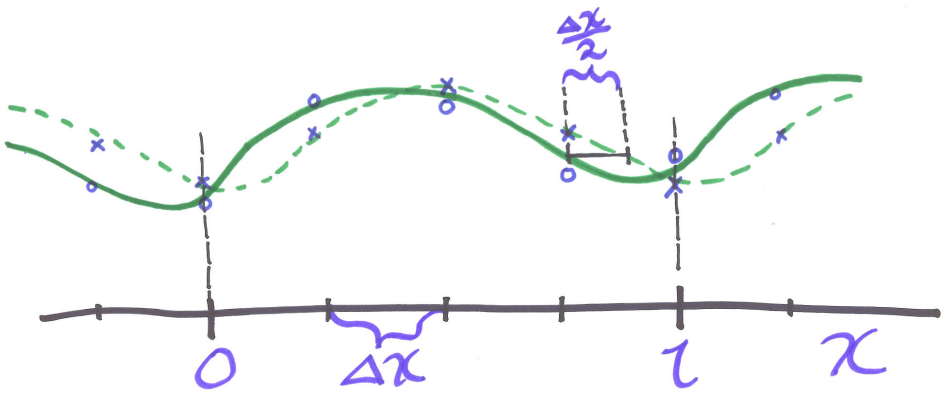


Figure 4.8: Steady state solutions of a Hamiltonian partial differential equation with periodic boundary conditions, discretized with a multisymplectic integrator (with grid spacing Δx in space). The steady states of the partial differential equation are solutions of a periodic boundary value problem. The continuous solution has a continuous translation symmetry, in the discrete solution this is replaced by a $\Delta x/2$ translation symmetry. The solid green line represents a continuous (exact) solution of the boundary value problem. It is approximated at the grid points by a discrete solution; the blue circles. The dashed green line is the $\Delta x/2$ forward translation of the solid line. It is approximated at the grid points by the discrete solution represented by the blue crosses.

By way of contrast, a non-multisymplectic discretization of (4.17) — for example a discretization with a wide finite-differences stencil, as is commonly used — can result in complete destruction of numerical steady state solutions since such a discretization need not have periodic points, and since only *exactly* periodic solutions (of the discrete version of (4.22)) are steady state solutions of the partial differential equation. This can be understood by noting that a necessary property of the twist map in the Poincaré-Birkhoff fixed point theorem (presented at the beginning of this section) is that the map be area preserving.

This shows that:

(i) symplectic integration in space does lead to qualitatively better representation of solutions, and

(ii) symplectic integration in space is valuable even though the integration length (i.e. $[0, l]$) does not tend to infinity, and is even rather short.

As an aside, we mention that for certain Hamiltonian partial differential equations, second order in space, discretized with a symmetric linear multistep method, the result by Chartier, Faou and Murua [25], mentioned in section 1.4.1, may imply that steady state solutions are preserved in such discretizations since the underlying one-step method associ-

ated with the linear multistep integrator is conjugate to symplectic. Such a discretization can be represented as a (wide) finite-differences stencil.

This section has only investigated the effect of resonances between the frequency of a periodic orbit and the step size of a numerical integrator. However, the strong non-resonance condition in theorem 4.6 is more general than this: it also includes the effects of the "normal" frequencies — the Floquet multipliers (or their generalizations). Investigation of the behaviour of a symplectic integrator for resonances involving this sort of condition appears to be an open (and difficult) question.

4.7 Conclusion

When a symplectic integrator is applied to a Hamiltonian differential equation, whose dynamics include a periodic orbit, the consequences can be split into two cases: when the integrator step size is resonant with the period of the orbit, and when it is strongly non-resonant.

In the strongly non-resonant case we showed that it is possible to apply a KAM-type theorem to prove that the periodic orbit persists but with a slightly perturbed frequency and with a quasi-periodic perturbation. The perturbation to the original frequency is exponentially small in the integrator step size. Put another way, the periodic orbit of the original system is replaced by an invariant 2-torus in a system with one additional degree of freedom. One frequency of the invariant torus is close to that of the of the original periodic orbit, while the other frequency is $2\pi/h$. The amplitude of the motion in the new periodic variable is exponentially small in h. This exponentially small perturbation in the new coordinate is a direct consequence of the exponentially small non-autonomous term in the embedding theorem of Moan. The position — that is, the value of the action variable — of the periodic orbit undergoes a perturbation of order h^p where p is the order of the symplectic integrator. This perturbation is due to the need to modify terms in the perturbed Hamiltonian during Kolmogorov's iteration so as to make the terms independent of the angles.

In the case where the integrator step size is resonant with the period of the orbit, the periodic orbit is destroyed. This can also be seen from the KAM-type theorem: the divisors in the series expansion of the generating function, necessary to return the perturbed system to a form with a periodic orbit, vanish and the series diverges. Hence, no such generating function exists. We conjecture that the breakup occurs in a fashion analogous to that for one degree of freedom systems as described by the Poincaré-Birkhoff fixed point theorem. The periodic orbit of the continuous system is replaced by a set of $2n$ periodic points when the resonance is of order n. When viewed as fixed points of the return map (i.e. Φ_h^n) half the points are elliptic and half are hyperbolic. The elliptic and hyperbolic points occur alternately and are $\mathcal{O}(h^p)$ close to the original periodic orbit. Numerical experiments with the two degree of freedom Hénon-Heiles system supported this conjecture.

The case when the integrator step size is close to resonant is somewhat of a grey area. The conjecture of the resonant case does not apply but the KAM type theorem may not hold due to the lack of a strong non-resonance condition preventing a proof of convergence of Kolmogorov's iteration. Numerical evidence would seem to suggest that the strong non-resonance bounds in KAM theory are rather pessimistic. Some sort of numerical quantification of the step size for which periodic orbits are destroyed would be interesting. For small step sizes, one may fall back on backwards error analysis which says only that something close to a periodic orbit is preserved over long, but finite, times.

Chapter 5

Highly Oscillatory Problems

> Our results are not encouraging.... one must require that [the product of the step size and the highest frequency] be small enough. Otherwise the method might become unstable and/or it might lead to a wrong approximation of the slowly varying solution components.
>
> — U. M. Ascher and S. Reich, 1997

In this chapter we look at integrators for second order differential equations with oscillatory solutions. It is desirable, at least on grounds of computational efficiency, if the numerical integrators used with this type of problem only require a new function evaluation after a time step of one, or even several, periods of the fastest oscillations of the system. In this case backwards error analysis is of no help; the product of the step size and the highest frequency of the problem cannot be assumed small and the "exponentially small" remainder terms are $\mathcal{O}(1)$. This leads to an "all-or-nothing" situation with structure preservation — only the properties captured exactly by the integrator can be expected to persist.

One highly popular class of integrators for oscillatory differential equations with constant high frequencies are the so-called *trigonometric integrators*. It is widely known that some of the trigonometric integrators suffer from low order resonances for particular step sizes. We show here that, in general, trigonometric integrators also suffer from higher order resonances which can lead to loss of nonlinear stability. We also show that in some cases trigonometric integrators yield misleading information: they preserve invariant or adiabatic quantities, but at the wrong values. The Fermi-Pasta-Ulam problem is a well known example of a nonlinear Hamiltonian system with highly oscillatory solutions. We use it here to illustrate and study the behaviour of trigonometric integrators. Implicit methods, such as the midpoint rule are less popular integration methods for highly oscillatory problems — in part due to the cautionary tone of articles by Ascher and Reich [6, 7]. However, we show that, despite the pitfalls of the midpoint rule, its performance may not be so much worse than that of the trigonometric methods, depending on what

one requires of the solution. We use statistical properties such as time averages, as well as more traditional criteria such as conservation of total and oscillatory energy, to evaluate the performance of the trigonometric methods and compare their performance with that of the midpoint rule.

In section 5.1 we give some background on trigonometric integrators, including their origins and their derivation. We present the celebrated Fermi-Pasta-Ulam (FPU) problem — a standard test problem for integrators for highly oscillatory problems — in section 5.2. In section 5.3 we illustrate the problem of higher order resonance for trigonometric integrators using conservation of an invariant quantity (the total energy) and an adiabatic invariant (oscillatory energy) as measures of performance. In section 5.6 we extend our comparison to other measures of performance and look at how the integrators treat some of the statistical properties of the FPU system. We summarise our findings in section 5.7.

Much of the work in this chapter has been published in the article [109] and the preprint [87].

5.1 Background to trigonometric integrators

Trigonometric integrators have enjoyed recent popularity due to their efficiency at solving constant frequency, highly oscillatory, differential equations [41, 46, 52, 55, 66]. For a particular class of second order differential equations with (highly) oscillatory solutions these methods require a new complete function evaluation only after a time step of one, or even many, periods of the fastest oscillations of the system.

Trigonometric methods have their origins in molecular dynamics and other n-body problems. The Newtonian equations of motion for a system of particles in a potential $V(q)$ are a given by the second order differential equation

$$\ddot{q} = -\nabla V(q).$$

In molecular dynamics, the standard integrator for such systems is the leapfrog method. The method is symmetric, symplectic and second order accurate but for linear stability (i.e. in order that the growth of errors remains bounded for linearized systems) it is necessary that the step size of the method be restricted to

$$h\omega < 2,$$

where ω is the highest frequency of the system — that is, $\omega^2 = \max\left(\lambda\left(\nabla^2 V(q)\right)\right)$, $\lambda(A) :=$ spectrum of A — along the numerical solution. Often, the potential is the sum of components which act on different time scales:

$$V(q) = W(q) + U(q), \text{ where } \|\nabla^2 W(q)\| \gg \|\nabla^2 U(q)\|,$$

and where $\nabla^2 W$ is positive semi-definite. Generally, the solutions of such systems are highly oscillatory on the slow time scale $1/\|\nabla^2 U(q)\|^{\frac{1}{2}}$. Often, the fast forces $-\nabla W(q)$ are cheaper to evaluate than the slow forces $-\nabla U(q)$ and one wants to exploit this by designing numerical methods where the number of slow force evaluations is not affected too severely by the fast forces.

5.1.1 Earlier trigonometric integrators: Gautschi, Deuflhard and the impulse method

Gautschi's method

Gautschi [42] proposed a class of methods for differential equations where the high oscillations arise from a quadratic potential $W(q) = \frac{1}{2}\omega^2 q^\top q$, $\omega \gg 1$. The differential equations take the form

$$\ddot{q} = -\omega^2 q + g(q). \tag{5.1}$$

(Such systems can be thought of as a collection of perturbed harmonic oscillators.) Gautschi's methods are of linear multistep type but with the coefficients of the methods α and β (cf. the description of linear multistep methods in §1.2.1) replaced by trigonometric functions of $h\omega$. The methods are constructed so as to be exact if the solution is a trigonometric polynomial in ωt. The simplest of Gautschi's methods, and the one used as a starting point for modern trigonometric integrators, reads

$$q_{n+1} - 2q_n + q_{n-1} = h^2 \left(\frac{\sin(h\omega/2)}{h\omega/2} \right)^2 \ddot{q}_n. \tag{5.2}$$

(In what follows, we will write $\mathrm{sinc}(x)$ for $\sin(x)/x$.) With use of the half-angle formula it is easy to show that (5.2) is equivalent to

$$q_{n+1} - 2\cos(h\omega) q_n + q_{n-1} = h^2 \mathrm{sinc}^2(h\omega/2) g_n, \quad \text{where} \quad g_n = g(q_n).$$

The method is exact for (5.1) in the case where $g = \text{Const.}$, as is the velocity approximation

$$\dot{q}_{n+1} - \dot{q}_{n-1} = 2h\mathrm{sinc}(h\omega)\ddot{q}_n.$$

Deuflhard's method

Written as a system of first order equations, the variation-of-constants formula for (5.1) reads

$$\begin{aligned} \begin{pmatrix} q(t) \\ \dot{q}(t) \end{pmatrix} &= \begin{pmatrix} \cos(t\omega) & \omega^{-1}\sin(t\omega) \\ -\omega\sin(t\omega) & \cos(t\omega) \end{pmatrix} \begin{pmatrix} q_0 \\ \dot{q}_0 \end{pmatrix} \\ &\quad + \int_0^t \begin{pmatrix} \omega^{-1}\sin((t-s)\Omega) \\ \cos((t-s)\omega) \end{pmatrix} g(q(s))\mathrm{d}s. \end{aligned} \tag{5.3}$$

Deuflhard [33] replaced the integral in (5.3) with its approximation by the trapezoidal rule to obtain the following method which is both explicit and symmetric:

$$\begin{pmatrix} q_{n+1} \\ \dot{q}_{n+1} \end{pmatrix} = \begin{pmatrix} \cos(h\omega) & \omega^{-1}\sin(h\omega) \\ -\omega\sin(h\omega) & \cos(h\omega) \end{pmatrix} \begin{pmatrix} q_n \\ \dot{q}_n \end{pmatrix} + \frac{h^2}{2} \begin{pmatrix} \operatorname{sinc}(h\omega) g_n \\ g_{n+1} + \cos(h\omega) g_n \end{pmatrix}. \quad (5.4)$$

Both Gautchi's and Dueflhard's method reduce to leapfrog for $\omega = 0$, and both extend to systems of multiple constant frequencies:

$$\ddot{q} = -Aq + g(q),$$

where $A = A^\top$, and A is positive semi-definite, by formally replacing ω with $\Omega = A^{\frac{1}{2}}$ in (5.4). The resulting methods require computation of products of matrix functions of $h\Omega$ with vectors. This can be done by diagonalizing A, which is practical for small systems or when A is the result of a pseudo-spectral discretization (e.g. from a nonlinear wave equation [146]), or by Krylov subspace methods [65] when the dimension of the system is large.

The impulse method

For systems with a Hamiltonian of the form $H(q,p) = T(p) + V(q)$, leapfrog can be viewed as a splitting method

$$\Phi_{h,X_H} = \varphi_{h/2,X_V} \circ \varphi_{h,X_T} \circ \varphi_{h/2,X_V}.$$

In cases where the potential can be split into fast and slow components $V = W + U$ it may, instead, be desirable to split the Hamiltonian as $H = (T+W) + U$ and approximate the flow φ_{h,X_H} by

$$\Phi_{h,X_H} = \varphi_{h/2,X_U} \circ \varphi_{h,X_{(T+W)}} \circ \varphi_{h/2,X_U}.$$

Such a splitting yields the *impulse method* [41] and was first proposed, in the context of molecular dynamics, by Grubmüller, Heller, Windemuth & Schulten [47] who named it the "Verlet-I scheme" and by Tuckerman, Berne & Martyna [143] who called the method "r-Respa". Often the flow $\varphi_{h,X_{(T+W)}}$ must be computed approximately by a numerical integrator using a smaller time step. If the inner integrator is symmetric and symplectic, then so too is the overall method. For quadratic potentials the impulse method is equivalent to Deuflhard's method (5.4).

5.1.2 Recent trigonometric integrators: The mollified impulse method and methods of Hairer, Lubich, Hochbruck and Grimm

The mollified impulse method

The impulse method is known to have instabilities when $h\omega$ is close to an integer multiple of π [41]. Such instabilities are known as *resonances*. The cause of the resonances is

that the slow forces $-\nabla U(q)$ are only evaluated at isolated points along the solution, at the end of each time step. Roughly speaking, the resonances mean the slow forces are always sampled at the same point during the oscillation leading to the integrator "seeing" the wrong effect for the slow forces. Garcia-Archilla, Sanz-Serna and Skeel [41] proposed evaluating the slow forces at some average value $\bar{q} = a(q_n)$, by replacing the potential $U(q)$ with a modified, or rather, a *mollified*, potential $\bar{U}(q) = U(a(q))$.

The *mollified impulse method* is the impulse method with the slow force $-\nabla U(q)$ replaced by the mollified force

$$-\nabla \bar{U}(q) = -a'(q)\nabla U(a(q)).$$

The method is still symplectic and symmetric since it is just the impulse method applied to a modified potential. There are numerous possibilities to choose for the average $a(q_n)$. The first (potential) restriction is that, if one wishes the algorithm to remain symmetric and symplectic the average must remain independent of the momenta. Hence, one cannot take averages of the solution during the oscillation step (i.e. the calculation of $\varphi_{h,X_{(T+W)}}$) of the algorithm.

Instead, one solves the initial value problem

$$\ddot{x} = -\nabla W(x(t)), \quad \text{with} \quad x(0) = q, \quad \dot{x}(0) = 0,$$

and the variational problem

$$\ddot{X} = -\nabla^2 W(x(t))X, \quad \text{with} \quad X(0) = I, \quad \dot{X}(0) = 0,$$

using the same integrator and step size for both. One can then compute the time average over an interval of length ch, $c > 0$.

$$a(q) = \frac{1}{ch}\int_0^{ch} x(t)\mathrm{d}t, \qquad a'(q) = \frac{1}{ch}\int_0^{ch} X(t)\mathrm{d}t.$$

Garcia-Archilla, *et al.* found that the best results were obtained with $c = 1$.

In the case of a quadratic potential for the fast forces $W(q) = \frac{1}{2}q^\top Aq$, with A positive semi-definite, it is possible to compute the average analytically:

$$\ddot{x}(t) = -Ax(t) \qquad \text{and} \qquad \ddot{X}(t) = -AX(t)$$

yield

$$x(t) = \cos(\Omega t)q \qquad \text{and} \qquad X(t) = \cos(\Omega t), \qquad \text{where } \Omega = A^{\frac{1}{2}}.$$

The averages can then be calculated:

$$a(q) = \frac{1}{ch}\int_0^{ch} \cos(\Omega t) q \mathrm{d}t = \mathrm{sinc}(ch\Omega)q, \qquad \text{and} \qquad a'(q) = \mathrm{sinc}(ch\Omega).$$

That is, $a(q) = \phi(h\Omega)q$ where $\phi(x) = \mathrm{sinc}(x)$ is called a *filter function*. With this choice of ϕ, the mollified impulse method reads:

$$\begin{pmatrix} q_{n+1} \\ \dot{q}_{n+1} \end{pmatrix} = \begin{pmatrix} \cos(h\omega) & \omega^{-1}\sin(h\omega) \\ -\omega\sin(h\omega) & \cos(h\omega) \end{pmatrix} \begin{pmatrix} q_n \\ \dot{q}_n \end{pmatrix} + \frac{h}{2} \begin{pmatrix} h\mathrm{sinc}(h\omega)\phi(h\Omega)g_n \\ \phi(h\Omega)g_{n+1} + \cos(h\omega)\phi(h\Omega)g_n \end{pmatrix}, \tag{5.5}$$

where $g_n = g(\phi(h\Omega)q_n) = -\nabla U(\phi(h\Omega)q_n)$.

More generally, for a pair of filter functions ϕ and ψ, one may write

$$\begin{pmatrix} q_{n+1} \\ \dot{q}_{n+1} \end{pmatrix} = \begin{pmatrix} \cos(h\omega) & \omega^{-1}\sin(h\omega) \\ -\omega\sin(h\omega) & \cos(h\omega) \end{pmatrix} \begin{pmatrix} q_n \\ \dot{q}_n \end{pmatrix} + \frac{h}{2} \begin{pmatrix} h\Psi g_n \\ \Psi_1 g_{n+1} + \Psi_0 g_n \end{pmatrix}, \tag{5.6}$$

where $\Phi = \phi(h\Omega)$, $\Psi = \psi(h\Omega)$ and where the ϕ and ψ are even, real-valued filter functions satisfying $\phi(0) = \psi(0) = 1$. Additionally, $g_n = g(\Phi q_n)$ and $\Psi_0 = \psi_0(h\Omega)$, $\Psi_1 = \psi_1(h\Omega)$ with even functions ψ_0, ψ_1 satisfying $\psi_0(0) = \psi_1(0) = 1$.

If the method is to be symmetric, then the ψ, ψ_0 and ψ_1 functions must satisfy

$$\psi(x) = \mathrm{sinc}(x)\psi_1(x) \qquad \text{and} \qquad \psi_0(x) = \cos(x)\psi_1(x), \tag{5.7}$$

as can be seen by exchanging $n \leftrightarrow n+1$ and $h \leftrightarrow -h$ in the method (5.6):

$$\begin{aligned} \begin{pmatrix} q_n \\ \dot{q}_n \end{pmatrix} = &\begin{pmatrix} \cos(-h\omega) & \omega^{-1}\sin(-h\omega) \\ -\omega\sin(-h\omega) & \cos(-h\omega) \end{pmatrix} \begin{pmatrix} q_{n+1} \\ \dot{q}_{n+1} \end{pmatrix} \\ &+ \begin{pmatrix} \frac{h^2}{2}\psi(-h\Omega)g(\phi(-h\Omega)q_{n+1}) \\ \frac{-h}{2}\psi_1(-h\Omega)g(\phi(-h\Omega)q_n) + \frac{-h}{2}\psi_0(-h\Omega)g(\phi(-h\Omega)q_{n+1}) \end{pmatrix}. \end{aligned}$$

Rearranging and making use of the properties of even and odd functions gives

$$\begin{aligned} \begin{pmatrix} q_{n+1} \\ \dot{q}_{n+1} \end{pmatrix} = &\begin{pmatrix} \cos(h\omega) & \omega^{-1}\sin(h\omega) \\ -\omega\sin(h\omega) & \cos(h\omega) \end{pmatrix} \begin{pmatrix} q_n \\ \dot{q}_n \end{pmatrix} \\ &+ \frac{h}{2} \begin{pmatrix} \cos(h\omega) & \omega^{-1}\sin(h\omega) \\ -\omega\sin(h\omega) & \cos(h\omega) \end{pmatrix} \begin{pmatrix} -h\Psi g_{n+1} \\ \Psi_1 g_n + \Psi_0 g_{n+1} \end{pmatrix}. \end{aligned} \tag{5.8}$$

Comparing (5.8) with the original integrator (5.6) we see that it is only the terms on the right with the $\frac{h}{2}$ prefactor which differ. Setting these equal, and comparing coefficients of

the g_n and g_{n+1} terms, we get the following two pairs of equations, each in three unknowns:

$$-h\Psi \cos h\Omega = \Psi_0 \Omega^{-1} \sin(h\Omega),$$
$$h\Psi = \Psi_1 \Omega^{-1} \sin(h\Omega).$$

and

$$h\Psi\Omega \sin h\Omega + \Psi_0 \cos(h\Omega) = \Psi_1,$$
$$h\Psi_1 \cos(h\Omega) = \Psi_0.$$

Solving either of these pairs of equations gives the symmetry requirements (5.7). It can also be shown that the method (5.6) is symplectic if and only if the filter functions satisfy

$$\Psi = \mathrm{sinc}(h\Omega)\Phi. \tag{5.9}$$

The methods of Hairer, Lubich, Hochbruck and Grimm

Recently, Hochbruck and Lubich [66] considered the question of "good" choices of filter functions. They required that the method be exact for all second order, linear problems with constant inhomogeneity (i.e. $\ddot{q} = -Aq + \text{Const.}$, for A positive semi-definite) and proposed several new filter functions. The choice

$$\psi(x) = \mathrm{sinc}^2(x/2), \qquad \phi(x) = \mathrm{sinc}(x)\left(1 + \frac{1}{3}\sin^2(x/2)\right) \tag{5.10}$$

was considered to give good results. The method is second order accurate, independent of frequency. Other choices of filter functions were proposed by Hairer and Lubich [52] and by Grimm and Hochbruck [46]. We focus on six[1] well known methods; the filter functions which define them are listed in table 5.1

	$\psi(x)$	$\phi(x)$	reference
(A)	$\mathrm{sinc}^2(\frac{1}{2}x)$	1	[42]
(B)	$\mathrm{sinc}(x)$	1	[33]
(C)	$\mathrm{sinc}^2(x)$	$\mathrm{sinc}(x)$	[41]
(D)	$\mathrm{sinc}^2(\frac{1}{2}x)$	$\mathrm{sinc}(x)(1 + \frac{1}{3}\sin^2(\frac{1}{2}x))$	[66]
(E)	$\mathrm{sinc}^2(x)$	1	[52]
(G)	$\mathrm{sinc}^3(x)$	$\mathrm{sinc}(x)$	[46]

Table 5.1: Filter functions for various trigonometric integrators

For comparison, and in spite of the caution against it by Ascher and Reich [7], we also consider the midpoint rule. The leapfrog/Verlet method is also used for comparison

[1]The methods are denoted (A) – (E) and (G), with (F) omitted to avoid confusion with the method (F) of [55, XIII.2.2]

where a reference solution is required. However in such cases it is used to fully resolve the high frequency components of the solution which does not agree at all with the spirit of integrators for highly oscillatory problems. Both leapfrog and the midpoint rule are symmetric, symplectic and second order accurate — at least when step size bounds such as that for linear stability are obeyed.

One can see that of the trigonometric methods in table 5.1 only (B) and (C) are symplectic. However, one can show that under the change of coordinates $y_n \mapsto \chi(h\Omega)y_n$ and with the filter functions ϕ, ψ, ψ_0 and ψ_1 replaced by $\chi\phi$, $\chi^{-1}\psi$, $\chi^{-1}\psi_0$ and $\chi^{-1}\psi_1$, then for $h\omega$ satisfying $\operatorname{sinc}(h\omega)\phi(h\omega)/\psi(h\omega) > 0$, it is possible to find $\chi(h\omega)$ such that the transformed method is symplectic [55, XIII.11-(3)]. That is, the methods preserve a frequency-dependent, modified symplectic form $dp \wedge dq + g(h\omega)dP \wedge dQ$ — a linear combination of two components of the original symplectic form — where (Q, P) are the "fast" variables corresponding to the ωI block of Ω. The consequences of preserving such a structure are not clear, though it would seem that the frequency dependent aspect is undesirable.

For short (i.e. $\mathcal{O}(1)$) time periods and small (i.e $h \to 0$) step size all the trigonometric methods are second order accurate in position; see [55, XIII.2.3].

5.1.3 Other methods

Multi-force methods

All the trigonometric methods mentioned thus far require only a single force evaluation per time step and are entirely specified by the choice of just two filter functions, ϕ and ψ. This is restrictive in the sense that additional restraints placed on the filter functions by requirements such as symplecticity, energy conservation, and oscillatory energy exchange are not always mutually compatible. To circumvent this difficulty, Hairer and Lubich [52] proposed generalizing the right-hand side of (5.6) to a linear combination of force evaluations, with different filter functions allowed for each. Written as a multistep method, the resulting integrators take the form

$$q_{n+1} - 2\cos(h\Omega)q_n + q_{n-1} = h^2 \sum_{j=1}^{k} \Psi_j g(\Phi_j q_n),$$

with $\Psi_j = \psi_j(h\Omega)$, $\Phi_j = \phi_j(h\Omega)$ and where ψ_j and ϕ_j are even functions satisfying

$$\sum_{j=1}^{k} \psi_j(0) = 1, \qquad \phi_j(0) = 1 \quad \text{for} \quad j = 1, \ldots, k.$$

Despite their possible advantages, multi-force trigonometric methods have not received much attention; nor are they considered in this thesis.

Hamilton-Jacobi methods

The Hamilton-Jacobi, or homogenization, methods of Le Bris and Legoll [81, 82] take an entirely different approach towards integrators for highly oscillatory Hamiltonian systems. The approach is similar to that of generating function methods. One considers a Newtonian system in Hamilton-Jacobi form. The generating function S defines, implicitly, a change of coordinates via a symplectic map.

In the methods of Le Bris and Legoll, an expansion in *two* small parameters, the step size h and the inverse frequency $1/\omega$, is used to define a generating function which, in turn, defines a map close to the flow of the highly oscillatory Hamiltonian system. By using a two-scale expansion, rather than expanding only for small time steps, it is possible to find methods with good behaviour, uniform in ω, at least under bounds on h and ω which are not excessively restrictive.

The methods themselves, however, are rather cumbersome and, potentially, computationally expensive, requiring the calculation of higher derivatives of the forces and the solution of systems of linear equations at each step. Despite these difficulties, the "geometric" approach offered by the use of generating functions, which automatically give symplectic methods, seems to offer some promise as a useful class of methods.

5.2 The Fermi-Pasta-Ulam problem

We are interested in second order differential equations of the form

$$\ddot{q} = -Aq + g(q), \quad \text{where} \quad A^{\frac{1}{2}} = \Omega = \begin{pmatrix} 0 & 0 \\ 0 & \omega I \end{pmatrix}, \quad \omega >> 1, \tag{5.11}$$

with blocks of arbitrary dimension and where $g(q) = -\nabla U(q)$. Partitioning q as (q_1, q_2) according to the blocks of Ω we see that the system (5.11) is Hamiltonian; the total energy is given by

$$H(q, \dot{q}) = \frac{1}{2}(|\dot{q_1}|^2 + |\dot{q_2}|^2) + \frac{1}{2}\omega^2|q_2|^2 + U(q). \tag{5.12}$$

The system also has an adiabatic invariant, or almost conserved quantity; the oscillatory energy:

$$I(q, \dot{q}) = \frac{1}{2}|\dot{q_2}|^2 + \frac{1}{2}\omega^2|q_2|^2. \tag{5.13}$$

Preservation of these quantities are two of the many possible criteria which one could use for judging the performance of a numerical integrator.

The main test problem we use in this article is the Fermi-Pasta-Ulam or FPU problem with three stiff (fast/harmonic) and four soft (nonlinear) springs as it is given in [55, I.5.1]

The Hamiltonian for the FPU system can be written as

$$\begin{aligned} H(y) &= \frac{1}{2}\sum_{i=1}^{3}(y_{0,i}^2 + y_{1,i}^2) + \frac{1}{2}\omega^2\sum_{i=1}^{3}x_{1,i}^2 \\ &\quad + \frac{1}{4}\left((x_{0,1} - x_{1,1})^4 + \sum_{i=1}^{2}(x_{0,i+1} - x_{1,i+1} - x_{0,1} - x_{1,i})^4 + (x_{0,3} + x_{1,3})^4\right), \end{aligned}$$

where:

- $x_{0,i}$ is the scaled displacement of ith stiff spring.
- $x_{1,i}$ is the scaled expansion (or compression) of the ith stiff spring.
- $y_{0,i}$ and $y_{1,i}$ are the conjugate momenta of the above.

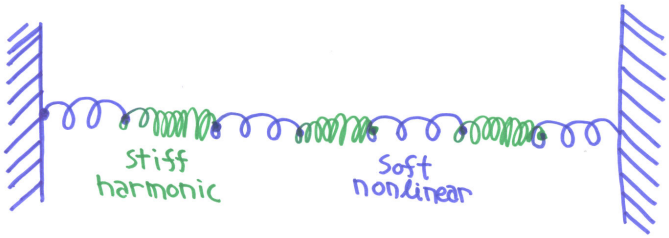


Figure 5.1: The Fermi-Pasta-Ulam problem: a chain of alternating soft, nonlinear, and stiff, harmonic, springs.

As in [55], we use initial conditions: $x_{0,1}(0) = 1$, $x_{1,1}(0) = \omega^{-1}$, $y_{0,1}(0) = 1$, $y_{1,1}(0) = 1$ and all other entries are zero. I.e.

$$z(0) = [1, 0, 0, \omega^{-1}, 0, 0, 1, 0, 0, 1, 0, 0]^\top,$$

where $z = [x_{0,i}^\top, x_{1,i}^\top, y_{0,i}^\top, y_{1,i}^\top], i = 1, 2, 3$. The Hamiltonian of the system is clearly a conserved quantity. The oscillatory energy of the ith stiff spring is given by

$$I_j = \frac{1}{2}\left(y_{1,j}^2 + \omega^2 x_{1,j}^2\right); \tag{5.14}$$

the total oscillatory energy $I = \sum_j I_j$ being an adiabatic invariant.

The FPU system is chaotic, and therefore, global errors are not sensible quantities to study as indicators of the performance of the numerical methods since arbitrarily nearby trajectories can diverge exponentially quickly. Instead, we use measures such as the change in the Hamiltonian and preservation of other physical properties to quantify the perfor-

mance of the methods. In the following section we use changes in the total and the oscillatory energy to show the effect of resonances on the numerical integrators.

5.2.1 Time scales in oscillatory problems

Before reporting on numerical results from the FPU problem we give a brief description, based on that in [28] and [55], of the behaviour at different time scales for Hamiltonian problems of the form (5.11). A useful tool in the analysis of such systems, and in the design of trigonometric integrators, is the modulated Fourier expansion:

$$q(t) = \sum_k z^k(t) e^{ik\omega}.$$

The coefficients $z^k(t)$ change on a time scale $\mathcal{O}(1)$ and multiply the exponentials which oscillate with frequency ω. Comparison of the coefficients from the expansion of exact and numerical solutions can give information about how the filter functions affect properties such as slow energy exchange.

Time scale $1/\omega$

On this time scale, the system (5.11) looks like a collection of slightly perturbed harmonic oscillators with frequency ω and amplitude $\mathcal{O}(1/\omega)$ in the fast variables. This behaviour can be reproduced by almost any integrator since the short time allows one to simply reduce the step size until the solution is sufficiently accurate.

Time scale ω^0

This is the time scale of motion for the slow variables q_1 under the influence of the potential $U(q)$. At this scale it is useful to have numerical methods whose error is small in the step size h and uniform in $h\omega$ — the product of the step size with the highest frequency. For trigonometric integrators, such a uniform error bound depends on the choice of filter functions. In general, numerical methods will not have uniform error in $h\omega$, even if they are geometric.

Time scale ω

At this time scale energy exchange between the different highly oscilatory components of the system can be observed. For example, exchange of oscillatory energy (5.14) between the stiff springs of the FPU problem. To leading order in $1/\omega$, the change in I_j is described by a differential equation that determines the coefficient of $e^{it\omega}$ in the modulated Fourier expansion of the solution. It turns out that if the corresponding coefficient of the modulated Fourier expansion for the solution of a trigonometric method is to agree with

that of the original system, then the filter functions must satisfy

$$\psi(x)\phi(x) = \operatorname{sinc}(x),$$

independent of frequency and step size. We will return to this point in section 5.5.

Time scale ω^N with $N \geq 2$

At this time scale, both the fast and slow variables appear highly oscilatory. However, one should observe the total energy of the system remaining constant and the total oscillatory energy remaining approximatly conserved — it varies only by $\mathcal{O}(1/\omega)$ [10, 27]. This is also the relevant time scale for the calculation of statistical properties. For chaotic systems, trajectories of this length can be expected to sample enough of the (chaotic) phase space in order to give accurate approximations to time averages of the system.

5.3 Nonlinear stability & resonances

It has already been mentioned, in this chapter, that trigonometric integrators can suffer from instabilities or resonances when $h\omega$, the product of the step size and the highest frequency, is an integer multiple of π. We will show here that rational multiples of π, in particular $h\omega = 2\pi/3$, can also lead to unstable resonances. These resonances are a type of nonlinear instability. For most systems, a linearized analysis is sufficient to study the stability of a dynamical system. However, Hamiltonian, or area-preserving systems, are neutrally stable and when integrated by a numerical method which does not introduce artificial damping, a linearized analysis is usually indeterminate. In such cases it is necessary to consider nonlinear effects. Instabilities arise when the integrator step size is one-third, one-quarter, or — rarely — some other fraction, of the period of the normal mode of the *discrete* dynamics [131]. Since trigonometric integrators are designed to get the normal modes exactly correct, (i.e. they reproduce the high frequencies ω exactly), it is trivial to see that the higher order resonances will occur at $h\omega = 2\pi/3$, $h\omega = 2\pi/4$, *et cetera*.

For Runge-Kutta methods — such as the midpoint rule, which we consider here — the period of the discrete dynamics can be determined by using the stability function of the methods:

$$R(z) = \frac{P(z)}{Q(z)} = \frac{\det(I - zA + z\vec{1}b^\top)}{\det(I - zA)}$$

with $z \in \mathbb{C}$ and where A and b are the coefficients from the Butcher tableau (cf. §1.2.2); $\vec{1}$ is a vector whose entries are all ones. Runge-Kutta methods map the eigenvalues λ of the continuous system to $\mu := R(h\lambda)$ for the discrete system. Hence, for the high frequency

oscillations of a planar Hamiltonian system

$$H(q,p) = \frac{1}{2}p^2 + \frac{1}{2}\omega q^2 + g(q),$$

discretized by a Runge-Kutta method, resonances of order n occur when the step size satisfies

$$R(ih\omega) = \exp(i2\pi/n).$$

For the midpoint rule $R(z) = \frac{1+z/2}{1-z/2}$ and resonances of order three and four occur for $h\omega = 2\sqrt{3}$ and $h\omega = 2$ respectively. Whether or not these resonances lead to instability (i.e. unbounded error growth) typically depends on higher order terms in the potential. The midpoint rule has no resonances of order one or two since as $h \to \infty$, $R(ih\omega) \to \exp(i\pi)$.

5.3.1 Resonances for a planar problem

Skeel and Srinivas, [131] showed that for a planar Hamiltonian

$$H(q,p) = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2 + \frac{1}{3}Bq^3 + \frac{1}{4}Cq^4 + \mathcal{O}(q^5) \tag{5.15}$$

the midpoint rule becomes unstable for order three resonances whenever the potential has a cubic term ($B \neq 0$) while the order four resonances are unstable when $(\omega^2 C - B^2)(\omega^2 C - 2B^2) < 0$.

This is related to a more general result, presented in [4, Appendix 7]; order three resonances of Hamiltonian systems are generically unstable whenever they occur. Order four resonances may be stable or unstable depending on nonlinear terms in the Hamiltonian equations of motion. The generic instability of order three resonances can be somewhat understood from figure 5.2 which shows the level sets of a one degree of freedom Hamiltonian which depends on a small parameter ε. For $\varepsilon \neq 0$, the fixed point at the origin of the system is surrounded by closed curves for a stable neighbourhood of radius $\mathcal{O}(\varepsilon)$. Points within this stable neighbourhood remain no further than $\mathcal{O}(\varepsilon)$ from the fixed point, while points outside the neighbourhood move away. At $\varepsilon = 0$ arbitrarily small perturbations of the fixed point give points which leave the neighbourhood of the origin and the fixed point becomes unstable.

An interesting case which does *not* show the effects of unstable order three resonances is the Hamiltonian pendulum

$$H(q,p) = \frac{1}{2}p^2 - \cos(q),$$

integrated with the midpoint rule. Since the Taylor expansion of $\cos(q)$ has only terms of even order the system does not support unstable order three resonances. The midpoint rule cannot have resonances of order one or two, so the first resonance which can occur is an order four resonance which occurs at $h\omega = h = 2$.

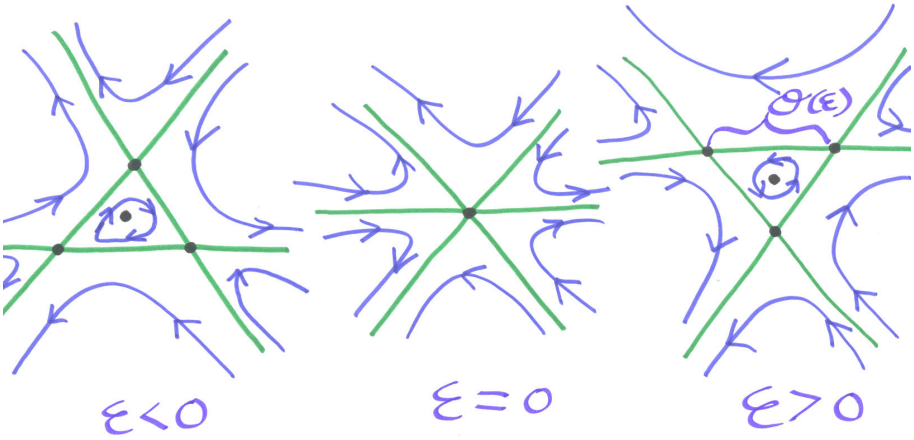


Figure 5.2: Illustration of an order three resonance which depends on the parameter ε. At $\varepsilon = 0$ the fixed point at the origin of the system becomes unstable — arbitrarily small perturbations lead to points which leave the neighbourhood of the origin.

Figure 5.3 shows the phase portrait of the pendulum for $h < 2$, $h = 2$ and $h > 2$. For $h > 2$ and $h < 2$, the phase portrait near the origin resembles that of the exact solution with invariant circles about the elliptic fixed point, even though the step size was large, particularly for $h > 2$. For $h = 2$, the invariant circles are distorted to diamonds. Picking one initial condition and plotting the maximum energy error as a function of step size shows a spike in the energy error around $h = 2$. The pendulum and the midpoint rule are a special case however. For more general systems one must be cautious of order three, and lower, resonances.

Here we investigate the effect of resonances on the solutions of the midpoint rule, and on the trigonometric integrators of table (5.1). We take the system given by (5.15), truncated after the q^4 term, and with $B = C = -1$. Fixing the step size at $h = 0.02$ we vary ω such that $0 < h\omega/\pi \leq 4.5$. Figure 5.4 shows the maximum deviation in the Hamiltonian for integration over the interval $[0, 1000]$.

The widths of the resonant bands for the symplectic trigonometric method (B) appear to render it unusable. Method (A) fares slightly better due to the absence of the wide resonance spikes at even multiples of π. Methods (C) and (G) seem to be free of the resonances at πn, $n \in \mathbb{N}$. More importantly, *all* the trigonometric methods show evidence of order three resonances despite the fact that such resonances are not reported in the literature.

The results for the midpoint rule show order three and four resonances as predicted. The method is not affected by resonances of order lower than three because its eigenvalues are limited to $\exp(i\theta)$, $\theta \in (0, \pi)$.

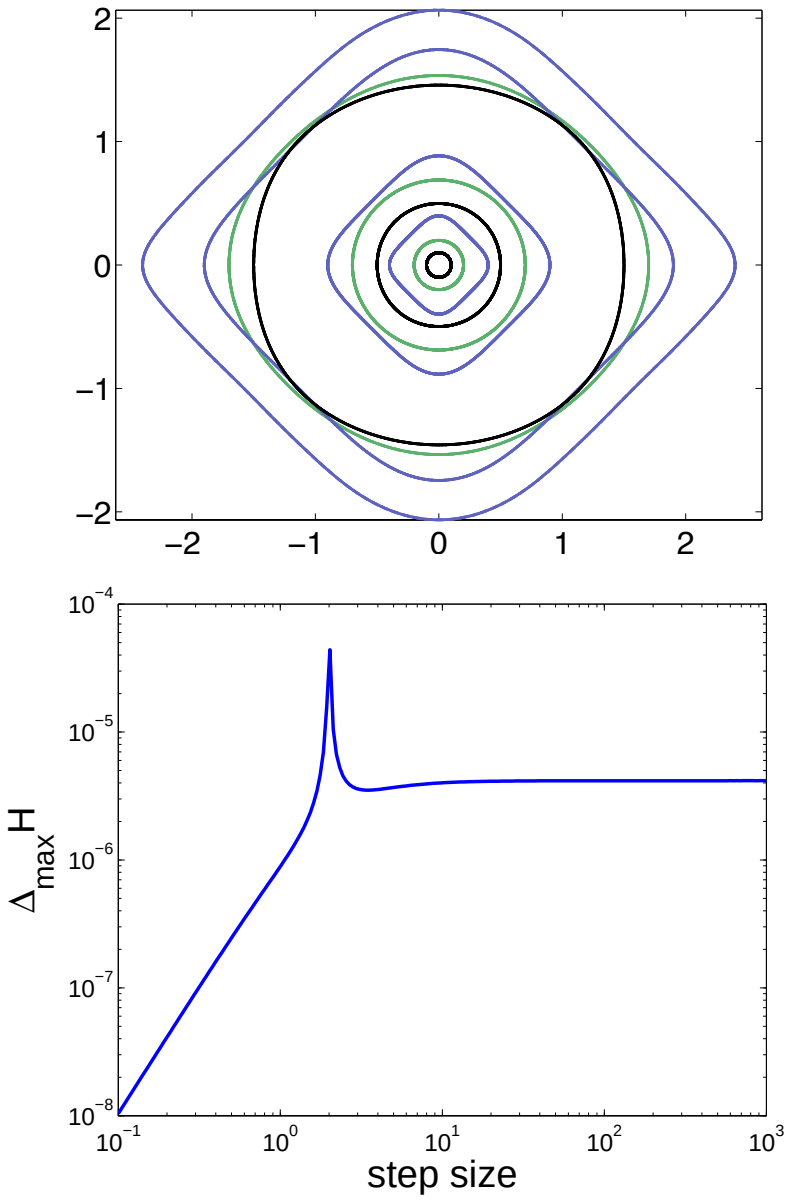


Figure 5.3: Phase portrait and energy error for an order four resonance of the pendulum integrated with the midpoint rule. The phase portrait (*top*) shows that for $h = 1$ (*green curves*) the phase portrait near the origin resembles the invariant circles of the exact solution. For $h = 2$ (*blue curves*) the order four resonance distorts the invariant circles to diamonds. The trajectories near the origin return to something close to those of the exact solution for $h = 3$ (*black curves*). A plot of maximum energy deviation as a function of step size (*bottom*), for one trajectory, clearly shows a spike in the energy error when the step size is close to the resonant value.

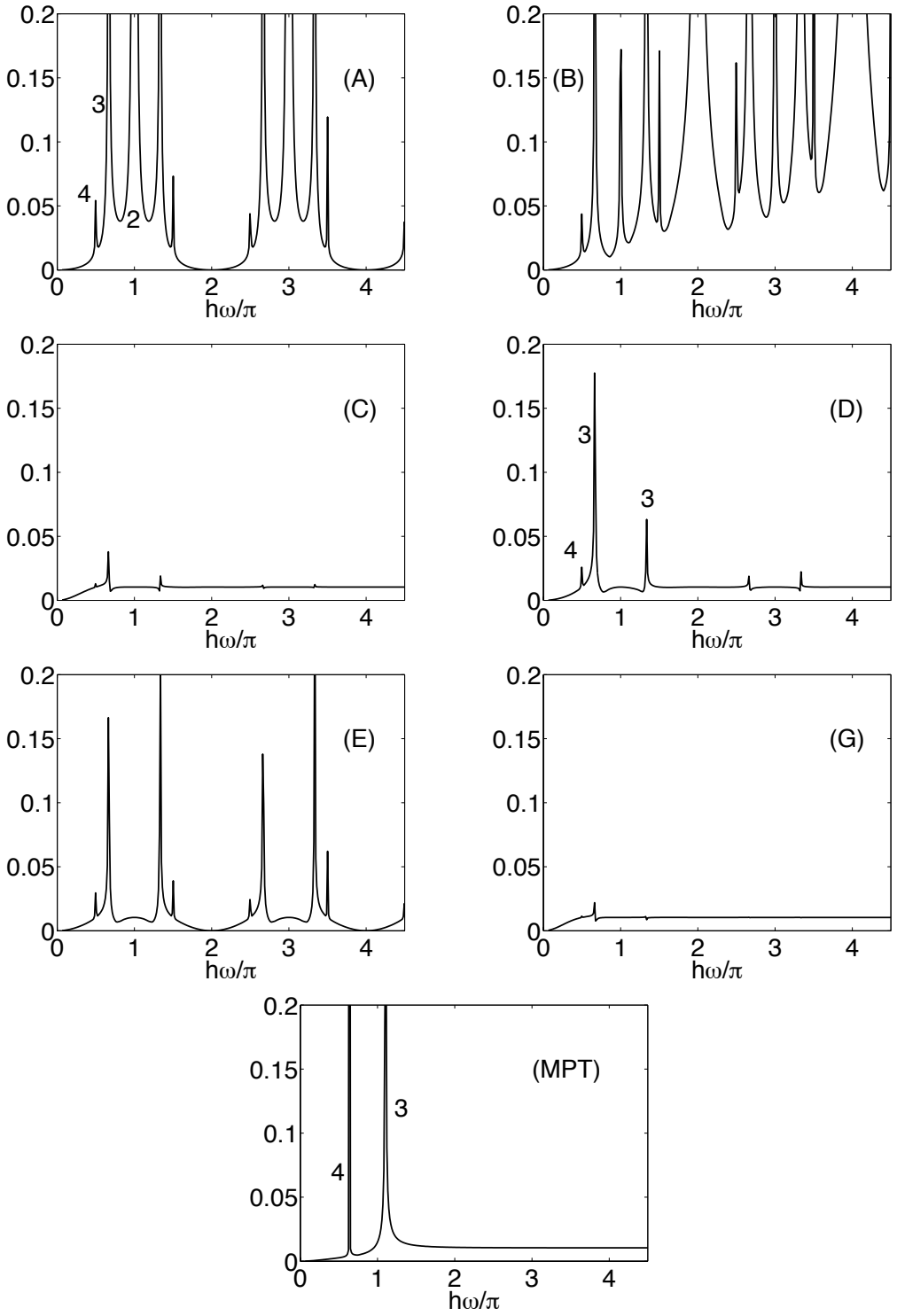


Figure 5.4: Maximum error in total energy on the interval $[0, 1000]$ as a function of $h\omega/\pi$ for the Hamiltonian $H(q,p) = \frac{1}{2}p^2 + \frac{1}{2}\omega^2 q^2 - \frac{1}{3}q^3 - \frac{1}{4}q^4$, (step size $h = 0.02$). For three of the methods — (A), (D) and (MPT) — we label some of the resonance spikes. For the trigonometric methods, the order m resonance spikes occur at $2\pi n/m$, $n \in \mathbb{N}$, when they are present. For the midpoint rule, the positions of the resonance spikes must be calculated as described at the beginning of section 5.3.

5.3.2 Consequences of resonance: Energy & oscillatory energy in a nonlinear system

We now turn to the Fermi-Pasta-Ulam problem with six degrees of freedom, as introduced in section 5.2. This has become a standard test system for integrators for highly oscillatory problems.

Energy deviation: Fixed step size

The FPU problem is Hamiltonian — its energy is an invariant quantity. Since the system is chaotic it doesn't make sense to look at the global error (in position) of an individual orbit, except for very short integration lengths. Therefore, to illustrate the effect of resonances we look at the errors in the energy of the numerical solutions for fixed step size and integration length. In figure 5.5 we plot the maximum deviation of the energy over the interval $[0, 1000]$ as a function of $h\omega/\pi$ for fixed $h = 0.02$.

With the exception of the inclusion of the midpoint rule, the results in figure 5.5 are identical to those in [55] where the use of trigonometric integrators is advocated.

In figure 5.5 the trigonometric methods no longer appear to show the order three and four resonances that they displayed for the planar system. This is due to relatively slow growth of these higher order resonances combined with the relatively short integration length, rather than the resonances no longer being present for higher dimensional systems. For longer integration times these resonances become apparent. We illustrate this in figure 5.6 by plotting the value of the Hamiltonian for three choices of step size using method (A) with an integration time roughly ten times longer than those used in figure 5.5. We take one step size $h = 0.04192 \simeq \pi/(3\omega)$ on the order three resonance and two step sizes immediately either side of the resonant value; $h = 0.044$ and $h = 0.04$.

We observe that the maximum energy error in figure 5.6 is no longer bounded for the resonant step size. This means that one must take care to avoid resonances due to rational as well as integer multiples of π for trigonometric integrators. It is also worth noting that for the midpoint rule, the error growth due to higher order resonances is extremely slow — orders of magnitude less than for the trigonometric integrators affected by resonances. While methods (E), (G), and to some extent (C) do not suffer from resonance spikes, they do show errors in the total energy, even over this relatively short integration period. What is perhaps worse, this energy error seems to vary, in a somewhat periodic fashion, as a function of $h\omega$.

Oscillatory energy deviation: fixed step size

The total oscillatory energy $I = \sum_j I_j$ of the FPU system is not a conserved quantity but, rather, an adiabatic invariant: a nearly conserved quantity which oscillates about its mean value with some standard deviation. The standard deviation of I is therefore a better characterisation of the oscillatory energy than the maximum deviation.

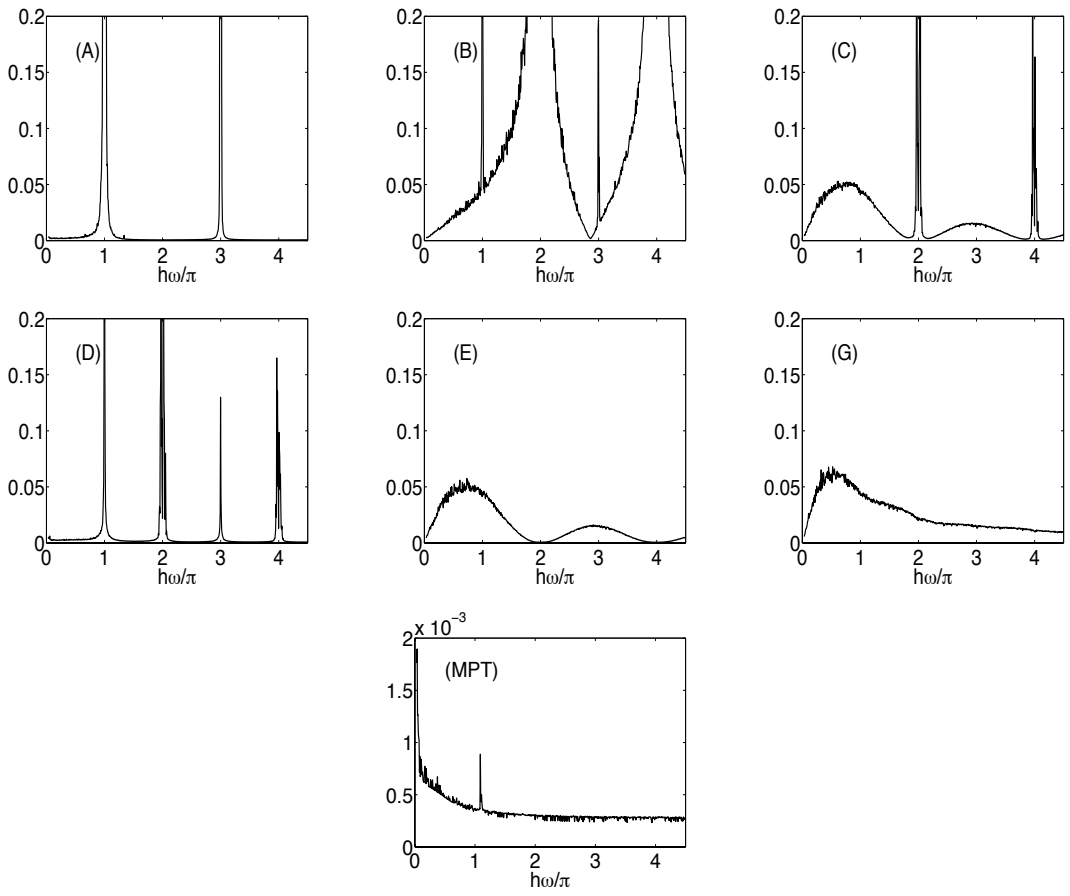


Figure 5.5: Maximum error of total energy on the interval $[0, 1000]$, (step size $h = 0.02$) for various integrators applied to the FPU problem. Note the different vertical axis for (MPT).

The exact solution of the FPU problem has $I(t) = \text{Const.} + \mathcal{O}(\omega^{-1})$. That is, the standard deviation of the total oscillatory energy as a function of ω should look like $C\omega^{-1}$. We used a very long ($T = 1 \times 10^6$) integration period and a small ($h = 0.002$) time step which resolved all the fast oscillations (for $\omega = 50$) to calculate, σI, the standard deviation of the oscillatory energy. This allowed us to determine the value of the coefficient ($C = 0.75$) and give a reference solution for the behaviour of $\sigma I(\omega)$.

Figure 5.7 shows $\sigma I \omega$ minus the reference solution, that is we plot $\sigma I \omega - 0.75$ against $h\omega/\pi$, for fixed step size $h = 0.02$. We can see that for the trigonometric method (G), the cost of preventing resonances is to also prevent the correct behaviour of the oscillatory energy. The only method which manages to approximate the correct behaviour, (a horizontal line at zero), is the midpoint rule. The trigonometric methods (A) and (D) show the correct behaviour away from resonant values of ω.

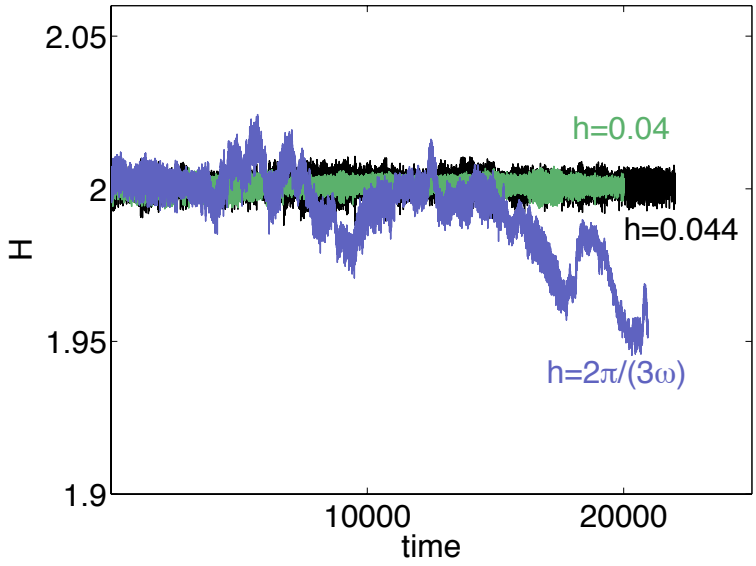


Figure 5.6: Total energy of solutions of the FPU problem, calculated with method (A) for step sizes slightly below ($h = 0.04$), on ($h = 2\pi/(3\omega) \simeq 0.0419,\ \omega = 50$), and slightly greater than ($h = 0.044$) the resonant value.

5.4 Energy conservation for fixed $h\omega$

Results for symplectic integrators concerning approximate energy conservation and preservation of a modified Hamiltonian hold in the limit of small step size. As the step size increases, one sees the difference between the modified and the original Hamiltonian grow; the energy of the numerical solution oscillates with greater amplitude as step size increases and backward error analysis — the usual tool for showing near-conservation of energy — no longer holds.

The condition on the filter functions ϕ and ψ for a trigonometric method to conserve total energy for linear systems $f(z) = Az$, up to $\mathcal{O}(h)$, independent of $h\omega$, (i.e. equation (5.16)), is incompatible with the condition (5.9) for the trigonometric methods to be symplectic. Since symplectic methods are not exactly energy conserving either, a fair question to ask is how well a particular numerical method is able to approximately conserve the total energy and how much extra work is needed in order to get better conservation of energy. That is, what is the order of the numerical scheme with respect to energy conservation.

In figure 5.8 we fix $h\omega = 0.5,\ 1.5$ and 5 and let h (and ω) vary. We plot the maximum deviation in the total energy on the interval $[0, 1000]$ against the step size h and use this to show the orders of the schemes with respect to energy conservation. The trigonometric methods (E) and (G) satisfy a condition

$$\psi(x) = \mathrm{sinc}^2(x)\phi(x) \tag{5.16}$$

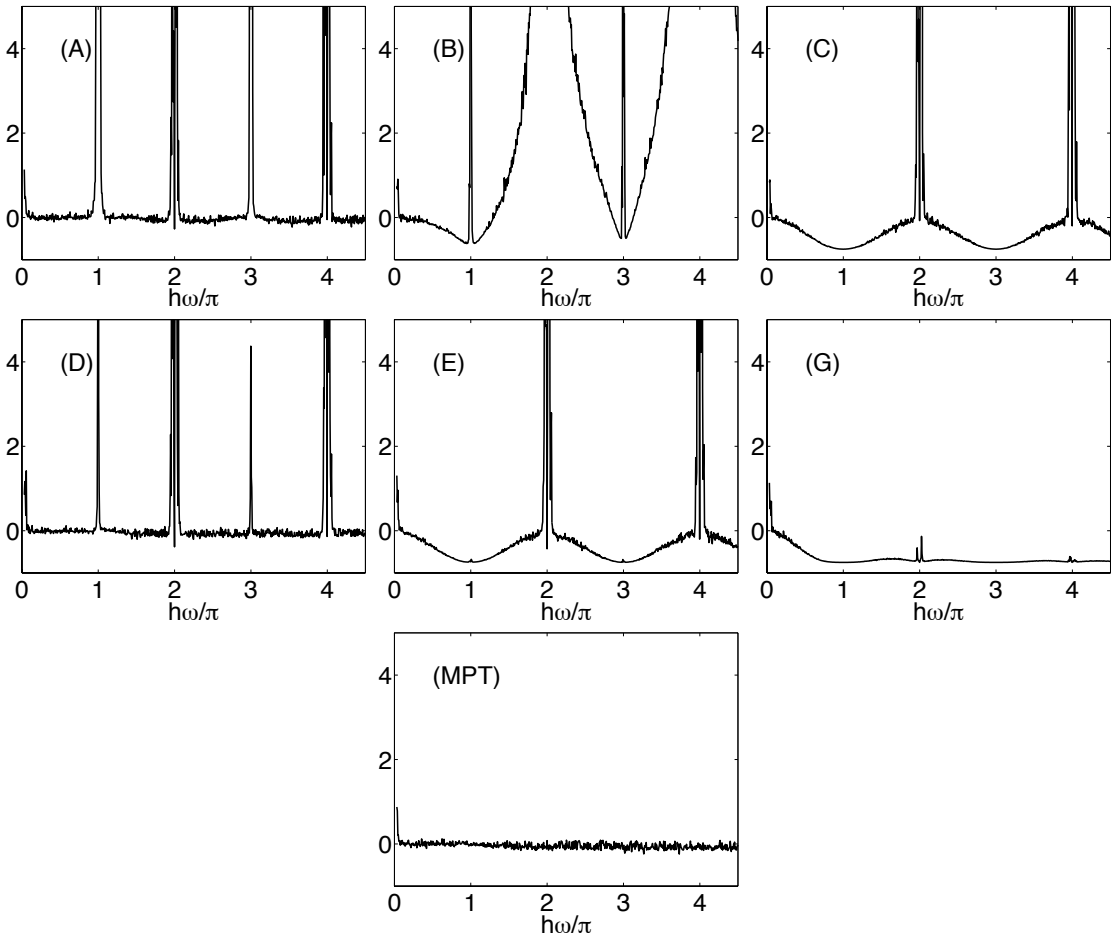


Figure 5.7: Standard deviation of oscillatory energy of the FPU problem, on the interval $[0, 1000]$, scaled by ω and shifted by a reference solution, (i.e. $\sigma I(\omega)\omega - 0.75$) such that the correct solution is a flat line at zero. (Step size $h = 0.02$).

necessary for the methods to conserve total energy up to $\mathcal{O}(h)$ independent of $h\omega$ [46, 52]. Methods (A) and (D) satisfy a different condition,

$$\psi(x) = \operatorname{sinc}^2(x/2), \tag{5.17}$$

which ensures that the energy conservation is $\mathcal{O}\left(h^2\right)$ accurate. These are also the methods which (away from resonances) correctly captured the oscillatory energy behaviour in section 5.3.2. It turns out that, even for $h\omega$ large, the midpoint rule is also second order accurate with respect to conserving the total energy. Methods (B), (C) and, of course, the midpoint rule are symplectic.

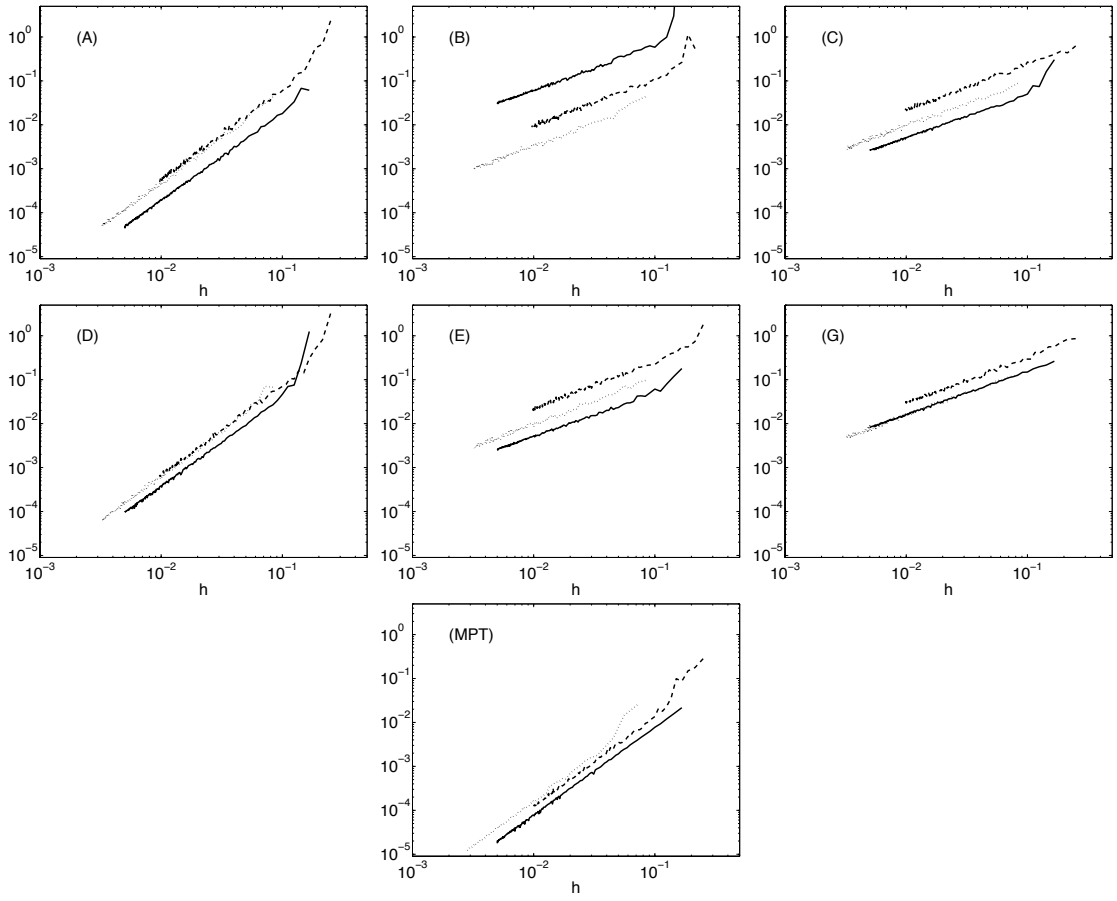


Figure 5.8: Maximum deviation in the total energy of solutions of the FPU system on the interval $[0, 1000]$ as a function of h for fixed $h\omega = 0.5$ (*dotted*), $h\omega = 1.5$ (*dashed*) and $h\omega = 5$ (*solid*).

5.5 Slow exchange of oscillatory energy

The FPU problem exhibits a slow exchange of energy between the stiff springs, that is, the distribution of I between I_1, I_2 and I_3 changes with time. These effects take place on a time scale of $\mathcal{O}(\omega)$. A good numerical method should capture the rate of the slow exchange and should give correct statistics for the mean distribution of energy between the stiff springs. For the trigonometric integrators to correctly approximate the slow exchange it is necessary that their filter functions satisfy [55, XIII.4.2]

$$\psi(h\omega)\phi(h\omega) = \operatorname{sinc}(h\omega). \tag{5.18}$$

In figure 5.9 we show the oscillatory energy in the stiff springs and the total oscillatory energy on the interval $[0, 200]$ for the trigonometric integrators and the midpoint rule. We use $\omega = 50$ and a step size of $h = 0.03$. For comparison we include the same results computed using the leapfrog integrator using a smaller step size (h=0.001) so that all the

fast oscillations are resolved.

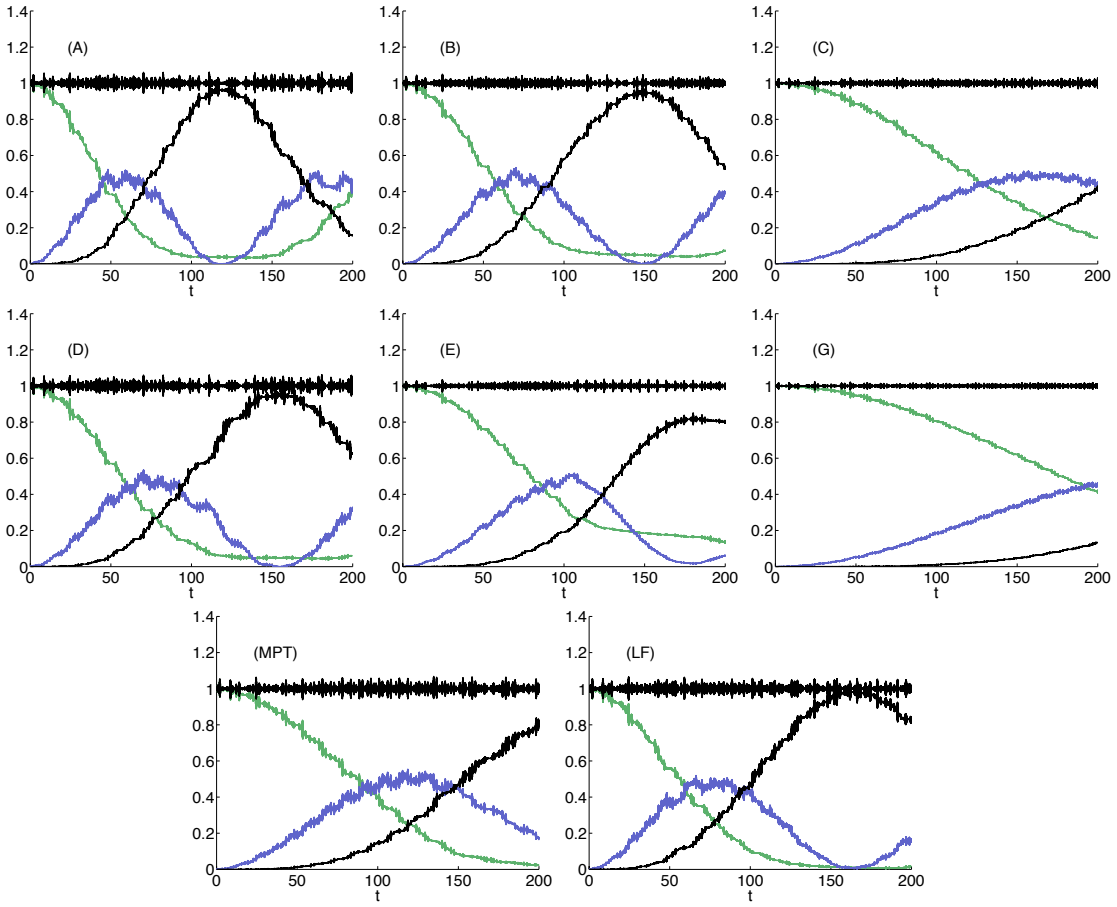


Figure 5.9: Oscillatory energy exchange between stiff springs of the FPU system on the interval $[0, 200]$ for fixed step size $h = 0.03$ and with $\omega = 50$. The solution for the leapfrog method, (LF), was computed with step size $h = 0.001$, resolving all oscillations (i.e. $h\omega = 0.05$); this can be regarded as the reference solution with which the other solutions should be compared. The slowly varying green, blue and black lines indicate the oscillatory energy in the I_1, I_2 and I_3 components respectively. The total oscillatory energy is also shown.

Of the trigonometric methods, only method (B) satisfies the slow exchange condition (5.18). However, for the values used here, $h\omega = 1.5$, method (D) almost satisfies (5.18) with $\psi(1.5)\phi(1.5) \simeq 0.95\text{sinc}(1.5)$ which accounts for its good behaviour in this case. Method (A) gets the energy exchange slightly too fast while method (E) and the midpoint rule get the exchange slightly too slow, methods (C) and (G) get the exchange slower still.

Although it doesn't get the rate of exchange correct for the oscillatory energy, the midpoint rule is the only method for which the energy in I_1 decreases all the way to zero before increasing again — the behaviour seen in the reference solution. The trigonometric methods transfer only a portion of the oscillatory energy between springs for this value of

$h\omega$. For example, method (B) has a turning point for the oscillatory energy in spring one (i.e. I_1) in the correct position, but at a higher value of I_1 than the reference solution. It seems likely that the failure to exchange the correct *amounts* of oscillatory energy between the modes would lead to incorrect long-time statistics of the trigonometric methods since they may not sample the same regions of phase space as the exact solutions do.

The condition for correctly approximating the slow exchange depends on $h\omega$. That is, for fixed ω the methods will give different rates of slow exchange for different step sizes — a purely numerical property since the rate of exchange should depend only on the parameters of the FPU problem itself. In figure 5.10 we plot I_1, the energy in the first stiff spring, on the interval $[0, 200]$, and with $\omega = 50$. We use three different step sizes so that $h\omega$ takes the values 0.5, 1.5 and 5.

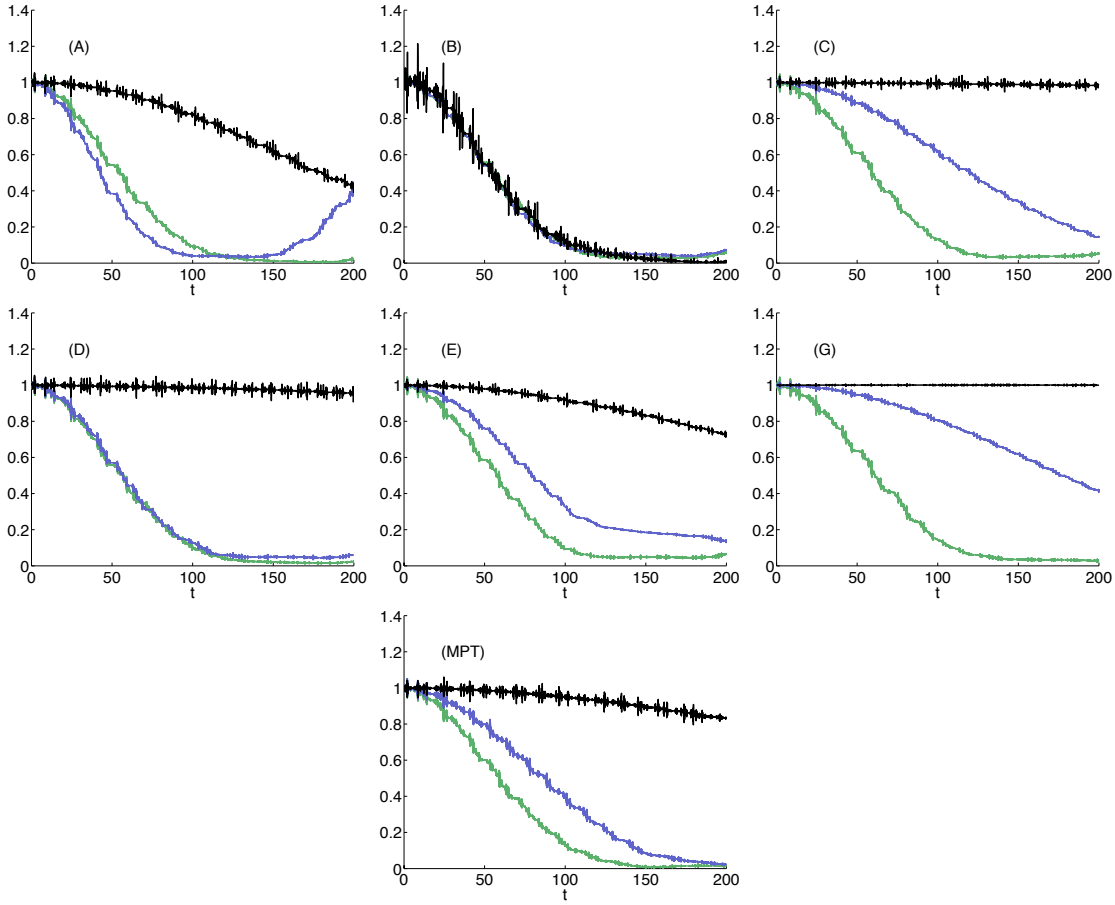


Figure 5.10: Oscillatory energy in the first stiff spring in the FPU system, on the interval $[0, 200]$ and with $\omega = 50$ for step sizes $h = 0.01$ (*green/lightest*, $h\omega = 0.5$), $h = 0.03$ (*blue*, $h\omega = 1.5$) and $h = 0.1$ (*black/darkest*, $h\omega = 5$).

For large values of $h\omega$, all the methods — with the exception of method (B) — perform badly. Some methods, however, perform worse than others: methods (G) and (C) do particularly badly. It might be sensible to ask questions about how quickly the behaviour

for the slow exchange converges to the correct behaviour as $h\omega$ decreases. Since the point of the trigonometric integrators is to be able to take step sizes larger than those allowed by traditional restrictions on $h\omega$ (most molecular dynamics calculations use a rule-of-thumb of $h\omega = 0.1$) one could ask "What is the slow exchange behaviour for $h\omega$ small but above the traditional limits?".

We also note that the amount of oscillatory energy transferred between springs is a function of $h\omega$. For example, method (B) has the correct positions of turning points for the oscillatory energy transfer independent of $h\omega$ but transfers the correct *amount* of oscillatory energy only as $h\omega \to 0$.

5.6 Statistical properties

We saw in section 5.5 that the slow exchange of energy between the stiff springs is a difficult property to capture for numerical methods, though it can be enforced by particular choices of filter functions. Here we look at some other quantities which are relevant for long integration lengths. We ask whether the numerical schemes get the correct mean values for oscillatory energy and we look at the distribution of the oscillatory energy between the stiff springs. We used a long integration interval of $[0, 1 \times 10^6]$ and frequency $\omega = 50$ for the fast oscillation. The step size was fixed at $h = 0.02$ for the trigonometric methods and the midpoint rule. As a reference solution, we used the leapfrog method with $h = 0.002$, a step size small enough to resolve the fast oscillations. We saved every 100th data point for the trigonometric methods and the midpoint rule and every 1000th for the leapfrog method. This was done purely in order to reduce the stored data stored to a manageable amount. Due to the very long integration period, and the fact that the solutions have bounded total and oscillatory energy, it is not expected that the discarded data would significantly alter the final values.

We present, in table 5.2, the average values of the oscillatory energy in each of the stiff springs, the standard deviation in the total oscillatory energy and the relative maximum deviation in the Hamiltonian for each of the methods. It is worth noting that for these calculations $h\omega = 1$ is only moderately large.

Another way to look at the data in table 5.2 is to subtract the leapfrog reference values from each row and then look at the absolute relative differences $\left|\frac{\bar{I}_j - \bar{I}_j^{ref}}{\bar{I}_j^{ref}}\right|$ We present this data in table 5.3 along with the relative absolute errors for the standard deviations of the total oscillatory energy and the mean of the relative absolute differences for the $\bar{I}_j$. In order to give a better overview of the data in table 5.3, we show the same data plotted as bar graphs in figures 5.11 and 5.12.

The methods which give the best results for the long-time statistics for the mean difference in the oscillatory energy with respect to the reference solution are the trigonometric methods (A) and (D) and the midpoint rule — the same methods which did well at captur-

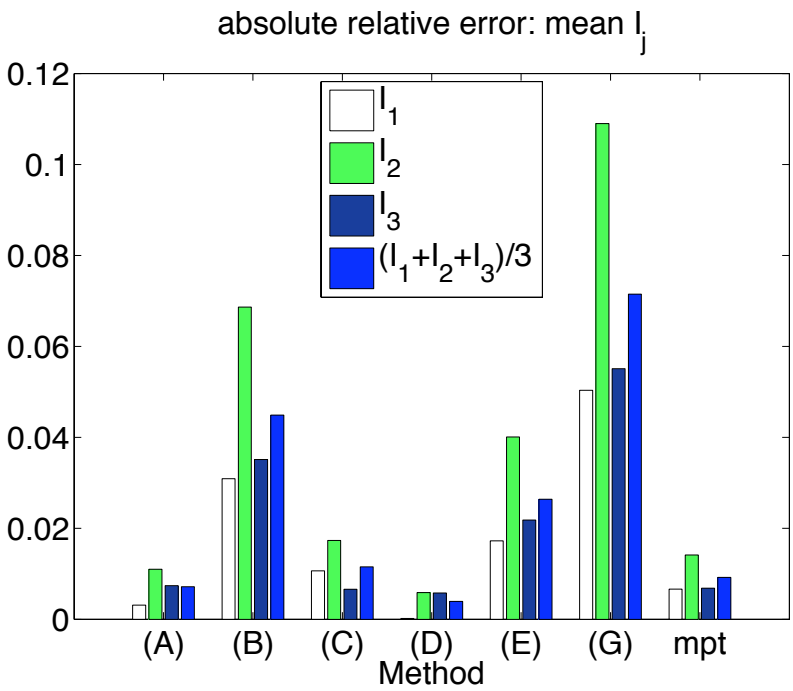


Figure 5.11: Bar graph of data from table 5.3: relative absolute differences in mean oscillatory energy of the FPU problem.

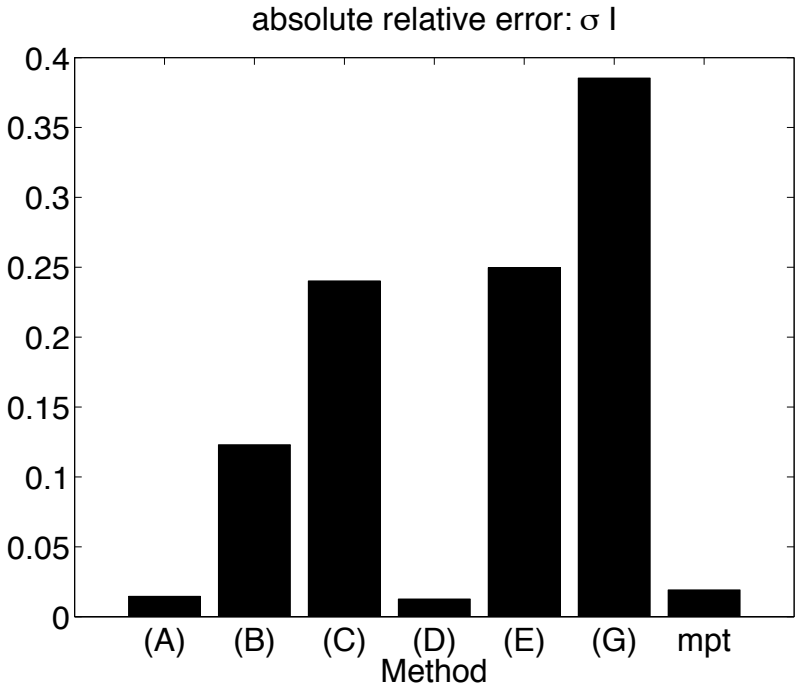


Figure 5.12: Bar graph of data from table 5.3: relative absolute differences in standard deviation of oscillatory energy of the FPU problem.

	$\bar{I}_1$	$\bar{I}_2$	$\bar{I}_3$	σI	$\Delta_{max} H/H(y(0))$
(A)	3.367e-01	3.279e-01	3.360e-01	1.475e-02	1.290e-03
(B)	3.273e-01	3.466e-01	3.266e-01	1.312e-02	7.800e-03
(C)	3.341e-01	3.300e-01	3.363e-01	1.137e-02	2.221e-02
(D)	3.378e-01	3.262e-01	3.366e-01	1.477e-02	1.616e-03
(E)	3.319e-01	3.373e-01	3.311e-01	1.123e-02	2.188e-02
(G)	3.207e-01	3.597e-01	3.199e-01	9.198e-03	3.280e-02
midpoint	3.400e-01	3.197e-01	3.408e-01	1.468e-02	3.789e-04
leapfrog	3.377e-01	3.243e-01	3.385e-01	1.496e-02	2.514e-03

Table 5.2: Mean values of I_1, I_2 and I_3, standard deviation in the total oscillatory energy and relative maximum deviation of the total energy for the trigonometric methods, the midpoint rule and the leapfrog method calculated on the interval $[0, 1 \times 10^6]$ with $\omega = 50$. The calculations with the trigonometric methods and the midpoint rule used a step size of $h = 0.02$ and saved every 100th point for the statistics. The calculation with the leapfrog method used $h = 0.002$ which resolves all the fast oscillations to give a reference solution. For the leapfrog method, every 1000th point was saved.

	$\left\lvert\frac{\bar{I}_1-\bar{I}_1{}^{ref}}{\bar{I}_1{}^{ref}}\right\rvert$	$\left\lvert\frac{\bar{I}_2-\bar{I}_2{}^{ref}}{\bar{I}_2{}^{ref}}\right\rvert$	$\left\lvert\frac{\bar{I}_3-\bar{I}_3{}^{ref}}{\bar{I}_3{}^{ref}}\right\rvert$	$\left\lvert\frac{\sigma I-\sigma I^{ref}}{\sigma I^{ref}}\right\rvert$	$\frac{1}{3}\sum\left\lvert\frac{\bar{I}_j-\bar{I}_j{}^{ref}}{\bar{I}_j{}^{ref}}\right\rvert$
(A)	3.11e-03	1.10e-02	7.40e-03	1.46e-02	7.17e-03
(B)	3.09e-02	6.86e-02	3.51e-02	1.23e-01	4.49e-02
(C)	1.07e-02	1.73e-02	6.61e-03	2.40e-01	1.15e-02
(D)	1.76e-04	5.88e-03	5.78e-03	1.27e-02	3.95e-03
(E)	1.72e-02	4.01e-02	2.18e-02	2.50e-01	2.64e-02
(G)	5.04e-02	1.09e-01	5.51e-02	3.85e-01	7.15e-02
midpoint	6.65e-03	1.41e-02	6.84e-03	1.92e-02	9.21e-03

Table 5.3: Relative absolute differences between the I_j and σI of table 5.2 and the reference solutions computed using leapfrog, along with the absolute mean of the relative differences in I_j.

ing the total oscillatory energy (see section 5.3.2). These methods show relative absolute errors of about one percent for the mean oscillatory energy. For methods (B) and (G) these errors are about five percent, though, for some components of the oscillatory energy, they are significantly larger. The relative absolute errors for the standard deviation of the oscillatory energy show even greater variation. The midpoint rule and methods (A) and (D) have errors of a couple of percent; the error for method (G) is close to 40 percent of the reference value.

In addition to getting the average values I_j, $j = 1, 2, 3$ correct, a numerical method should give the correct distribution of I_j values, that is, it should visit the appropriate parts of the I_j phase space for the correct amounts of time. We try to visualize this for the various methods by plotting the probability distribution functions for I_1 and I_2 — see figures 5.13 and 5.14.

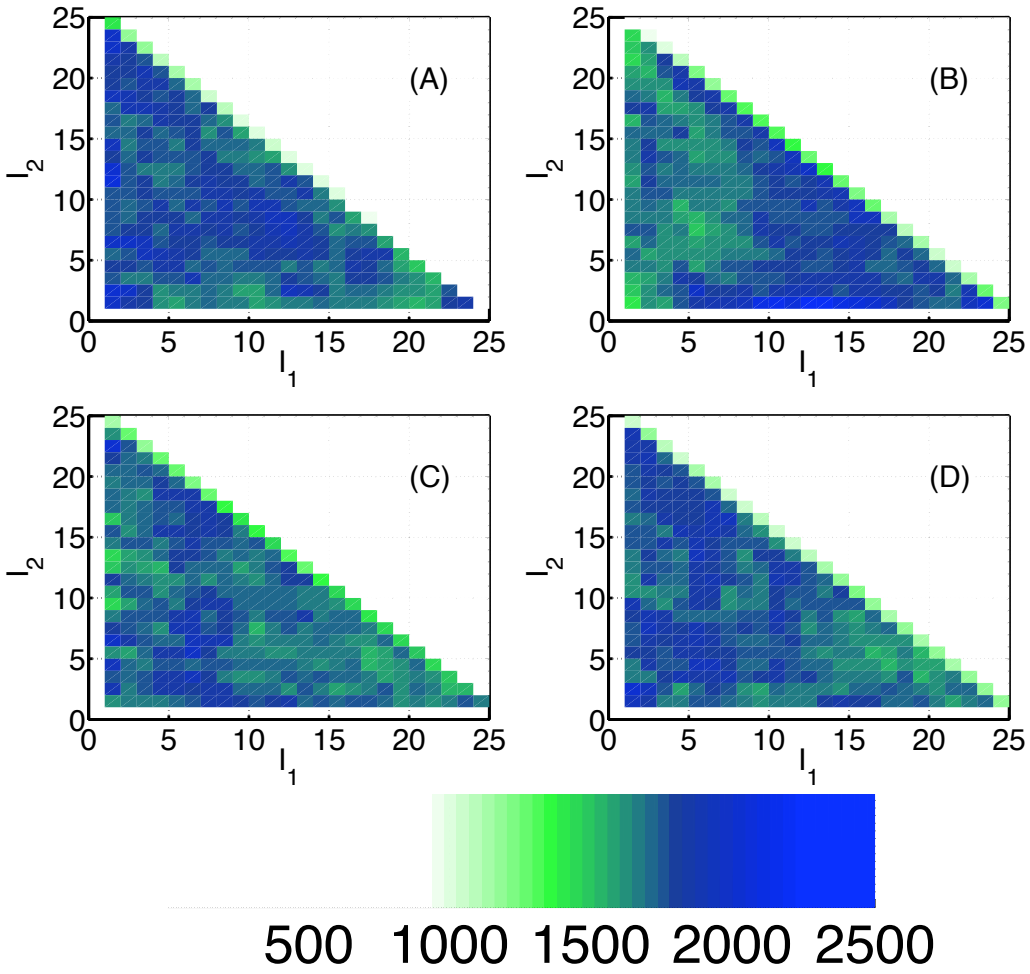


Figure 5.13: Flattened 3D histograms of the amount of time spent by solutions of the FPU system in regions of the $I_1 - I_2$ phase space. Solutions were calculated on the interval $[0, 10^6]$ with $\omega = 50$ and $h = 0.02$ (every 100th point saved), except for the leapfrog solution which used $h = 0.002$, to resolve all oscillations and saved every 1000th point. Plots for the remaining methods are shown in figure 5.14.

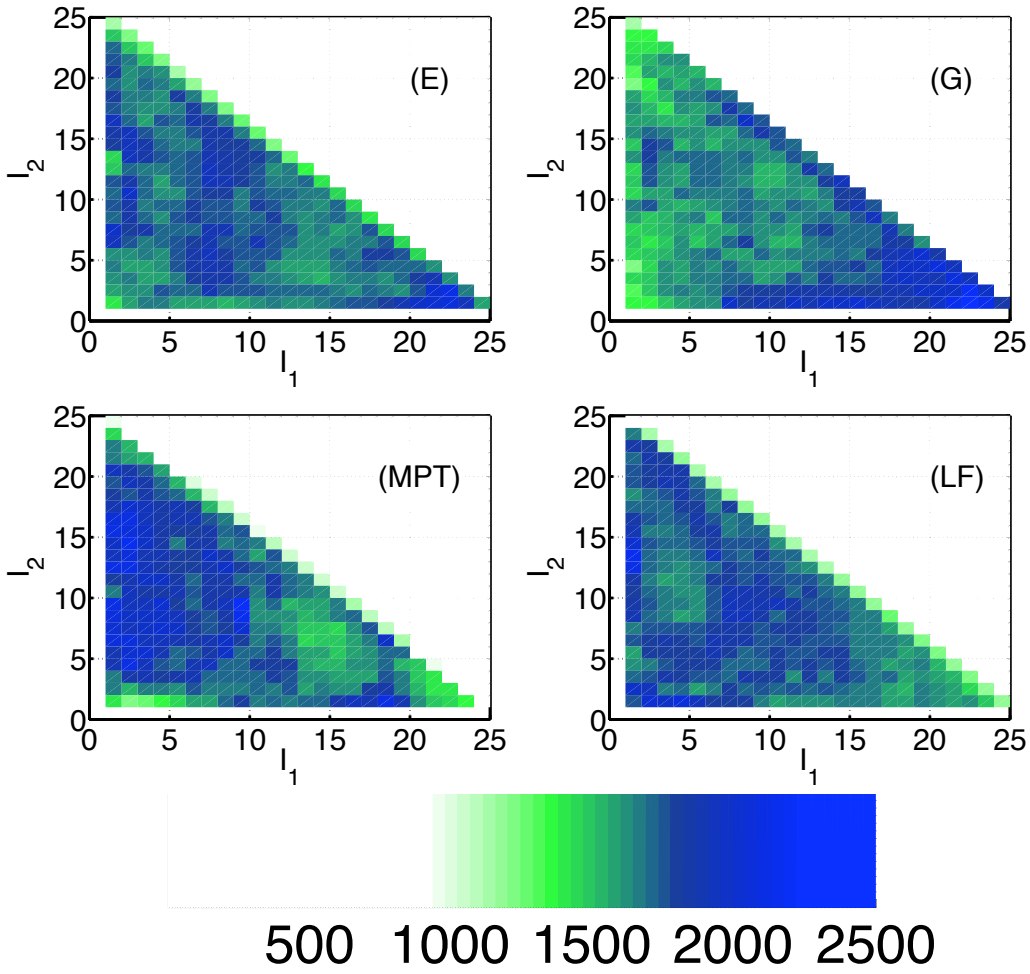


Figure 5.14: More flattened 3D histograms of the amount of time spent by solutions of the FPU system in regions of the $I_1 - I_2$ phase space. Solutions were calculated on the interval $[0, 10^6]$ with $\omega = 50$ and $h = 0.02$ (every 100th point saved), except for the leapfrog solution which used $h = 0.002$, to resolve all oscillations and saved every 1000th point. Plots for the remaining methods are shown in figure 5.13.

Method	**(A)**	**(B)**	**(C)**	**(D)**	**(E)**	**(G)**	**midpoint**
$\frac{\|P-P^{ref}\|}{\|P^{ref}\|}$	0.0547	0.1009	0.0685	0.0576	0.0904	0.1559	0.1233

Table 5.4: Relative normed differences for I_1, I_2 oscillatory energy probability distributions. (The distributions are plotted in figures 5.13 and 5.14).

Finding the best way to investigate the distribution is somewhat challenging: long integration periods are necessary in order to be sure that the trajectories have time to sufficiently sample the phase space of the system. This produces large amounts of data, hence, only every 1000th data point was saved for the leapfrog method, and every 100th data point from the other methods which used a larger step size. The sampling error due to this procedure is a possible obstacle to good characterization of phase space structure. In order to make probability distributions in figures 5.13 and 5.14 useful, a certain amount of coarse-graining is necessary so that the structure of the distribution can be seen. We divided the oscillatory energy I_1 and I_2 into 25 sections, or bins, in order that the three dimensional histogram of the probability density function showed sufficiently significant structures. Fewer bins tended to "smear out" features, while more bins made it difficult to distinguish structures from the inherent noise.

The reference solution shows quite a lot of structure, with certain regions of the $I_1 - I_2$ plane preferred over other regions. None of the methods considered here manage to reproduce all of this structure correctly, though, some methods perform worse than others. Methods (B) and (G) display almost the opposite behaviour to that of the reference solution, spending more time with no energy in the I_2 and I_3 components, while the reference solution rarely has large amounts of energy in I_1. Of all the methods, method (D) seems to best reproduce the distribution in the reference solution. This is supported by using the matrix of values with the I_1, I_2 frequency counts and calculating $\frac{\|\text{distribution}-\text{reference distribution}\|}{\|\text{reference distribution}\|}$ for each of the methods. Doing so gave the results in table 5.4 where one sees that, for this measure, the match between method (G) and the reference solution is about three times worse than for the best methods; (A) and (D).

In figure 5.15 we plot the probability distribution for I_1 by itself. These plots have the advantage that, by concentrating on a single component of the system, it is easier to identify similarities and differences between the behaviour of the integrators. Here again, differences between the distributions can be seen. The figure shows methods (B) and (G) spending more time than the reference solution with low values of I_1. The midpoint rule appears to favour moderate values of I_1 at the cost of time spent with I_1 close to zero.

5.7 Conclusion

We list here some of the main points from the preceding sections of this chapter:

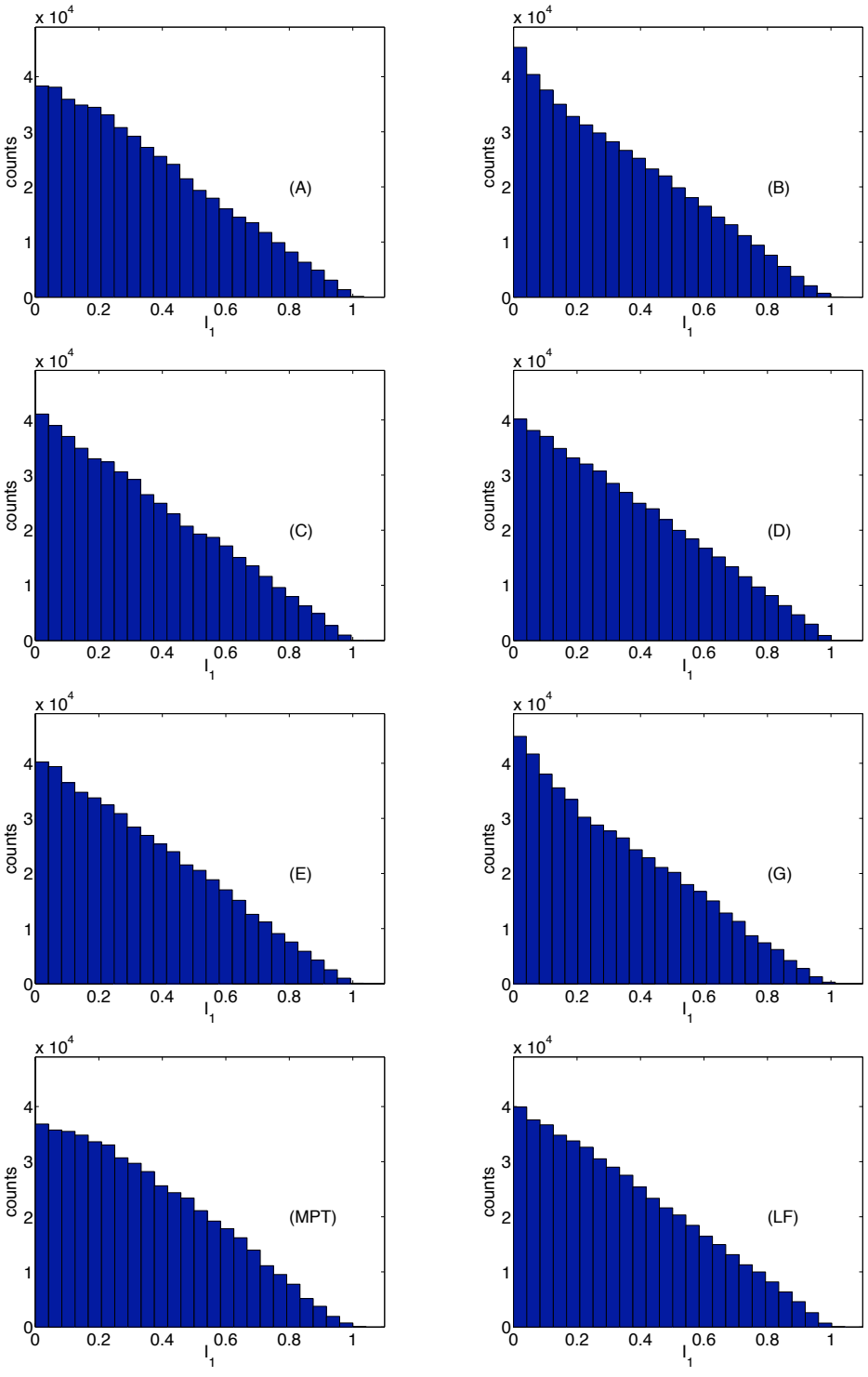


Figure 5.15: Histograms of the amount of time solutions of the FPU system spent with a particular fraction of the total oscillatory energy (*horizontal axis*) in I_1. Solutions were calculated on the interval $[0, 10^6]$ with $\omega = 50$ and $h = 0.02$ (every 100th point saved), except for the leapfrog solution which used $h = 0.002$, to resolve all oscillations and saved every 1000th point.

1. Order three resonances are generally unstable and cannot be avoided, As a consequence, trigonometric integrators are unstable for $3h\omega = 2n\pi$, $n \in \mathbb{Z}$.
2. It is not enough to simply suppress resonances or to bound the variation in conserved/adiabatic quantities as this can destroy other properties of the dynamical system. Some of the trigonometric integrators show conserved quantities being preserved but at entirely the wrong value.
3. None of the trigonometric methods manage to capture all properties and some perform worse than the midpoint rule. Although the midpoint rule is implicit, while the trigonometric integrators are explicit, the cost of evaluating matrix exponentials or similar, for the trigonometric integrators, in contrast to various techniques which reduce the cost of solving the system of implicit equations for the midpoint rule, mean that the midpoint rule can not be automatically discarded on grounds of computational cost.

   Previous studies of trigonometric integrators have not considered structural properties and, consequently, have advocated the use of integrators such as method (G) which we have shown gives incorrect, and misleading, results for the preservation of various quantities such as oscillatory energy. The likely consequences of this are that such methods are not able to correctly reproduce the behaviour of the dynamical systems they may be employed to study.

	(A)	(B)	(C)	(D)	(E)	(G)	(MPT)
$H(h\omega)$	✓	×	✓	✓	✓	✓	✓✓
$H(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$	$\mathcal{O}(h)$	$\mathcal{O}(h)$	$\mathcal{O}(h^2)$
I	✓	×	×	✓	×	×	✓
dI_j/dt	×	✓	×	×	×	×	×
I_j stats	×	×	×	×	×	×	×

Table 5.5: A quick comparison of the performance of various methods. $H(h\omega)$ refers to the effect of resonances on the energy error (see figures 5.4 and 5.5), while $H(h)$ refers to the energy error order behaviour of the methods for fixed $h\omega$.

Table 5.5 gives a comparison of the trigonometric methods and the midpoint rule for a selection of criteria. It is far from exhaustive. Other sensible measures of performance for highly oscillatory systems are other long-time averages, such as Lyapunov exponents. More investigation into the role (if any) of modified symplecticity of trigonometric integrators is also warranted, as is the investigation of multi-force trigonometric methods whose filter functions should be able to satisfy a larger number of the requirements for preserving various properties such as slow energy exchange. It is also important that future investigations of integrators for highly oscillatory problems make use of a variety of test problems; or demonstrate that the properties of the test problems used are generic. Because the FPU problem is so commonly used for comparison and evaluation of numerical methods,

there is the possibility that the development of new methods will be driven by the FPU problem leading to a narrowly applicable class of integrators.

Chapter 6

An Application: A Coupled Two-spin System

Data aequatione quotcunque fluentes quantitae involvente fluxions invenire et vice versa — It is useful to solve differential equations.

— I. Newton, reported in [3]

In this chapter we draw together some of the points made earlier in this thesis and illustrate them with a simple, yet non-trivial (and hopefully novel), application. We study a *coupled spin system in a magnetic field* — the classical analogue of two interacting magnetic moments, or spins, in an applied magnetic field. The motivation for studying this system is the article by Robb and Reichl [118] where the authors draw a link between chaos in the quantum Hamiltonian system of two coupled quantum spins, and the classical Hamiltonian system obtained as the limit of the quantum system; with the quantum spins replaced by angular momentum vectors. (We only look at the treatment of the classical Hamiltonian system and make no claims regarding the dynamics of the quantum system.)

The criterion used by Robb and Reichl for stating that a system is chaotic is that the iso-energetic Poincaré section of the system displays disorder. Such an approach allows one to be misled by the numerical dynamics and hence Guckenheimer [48] cautions against such an approach, suggesting instead that one should analyse the dynamical system; making use of the Smale-Birkhoff homoclinic theorem [132], and looking for transverse intersections of stable and unstable manifolds of periodic orbits. We take an approach somewhere between these two: numerical results can be sufficient to indicate chaotic dynamics but one must take care that the numerical trajectories accurately represent the dynamics of the original system. While chaotic dynamics do indeed lead to disorder in an iso-energetic Poincaré section, so too do poorly simulated dynamics, such as when the energy of trajectories is allowed to drift, or when invariant structures which would indicate regular dynamics (such

as KAM tori) are destroyed by a numerical integrator. Hence, a geometric approach to such calculations is needed.

It is also important that numerical integration take account of properties specific to the system being studied. The coupled two-spin system has a phase space given by the Cartesian product of two $\mathbb{S}^2$ spheres. When the system is in canonical form (that is, when it can be written in the form (1.13)) this property can be clearly seen and is preserved by any one-step method. However, in canonical form, the equations of motion have a coordinate singularity which numerical integrators may struggle to deal with. If the system is written as a more general *Poisson system* (cf. §1.3) the coordinate singularity is avoided but the equations of motion have coordinates in $\mathbb{R}^6$. Care must therefore be taken to restrict the numerical solutions to the correct manifold.

We discuss this point in more detail in section 6.1 where we present the coupled two-spin system and discuss its geometric structure. In section 6.2 we construct an appropriate (symplectic) integrator for the two-spin system — based on splitting. Numerical results are presented in section 6.3, where Poincaré sections calculated with the geometric integrator of section 6.2 are contrasted with those from a "black-box" integrator, and in section 6.4 where accuracy and energy conservation are investigated.

This chapter is based on work published in the article [86] and the report [85].

6.1 Description of the system

We begin with the Hamiltonian obtained as the classical limit of a system of two interacting magnetic moments/quantum spins in an applied magnetic field [118]:

$$H\left(S_1, S_2\right) = -\mu(S_{1x}S_{2x} + S_{1y}S_{2y}) + \lambda(S_{1x} + S_{2x}), \tag{6.1}$$

where $S_1 = (S_{1x}, S_{1y}, S_{1z})^\top$ and $S_2 = (S_{2x}, S_{2y}, S_{2z})^\top$ are the angular momentum vectors in $\mathbb{R}^3$. The system consists of an xy interaction between the two moments, or spins, given by the first term, and an interaction with a magnetic field applied in the x direction. The strength of the coupling between the spins is given by μ and the strength of the applied magnetic field is given by λ. Such systems may play a role in the investigation of quantum chaos: one asks whether a quantum system displays chaos for the same parameter values as its classical limit [118, 116, 37].

In Cartesian coordinates the equations of motion for the system (6.1) are given by

$$\frac{\mathrm{d}S_l}{\mathrm{d}t} = -S_l \times \frac{\partial H}{\partial S_l} = \begin{bmatrix} S_{lx} \\ S_{ly} \\ S_{lz} \end{bmatrix} \times \begin{bmatrix} \mu S_{mx} - \lambda \\ \mu S_{my} \\ 0 \end{bmatrix}, \tag{6.2}$$

where $\times$ is the usual cross product and where $l, m = 1, 2$, $l \neq m$. This is a *Poisson* system

(cf. §1.3) with the generalised Poisson bracket

$$\{F, G\}(S) = \nabla F(S)^\top J(S) \nabla G(S),$$

where $S^\top = [S_1^\top, S_2^\top]$, and where $J(S)$ depends linearly on S, namely,

$$J(S) = \begin{bmatrix} 0 & -S_{1z} & S_{1y} & 0 & 0 & 0 \\ S_{1z} & 0 & -S_{1x} & 0 & 0 & 0 \\ -S_{1y} & S_{1x} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -S_{2z} & S_{2y} \\ 0 & 0 & 0 & S_{2z} & 0 & -S_{2x} \\ 0 & 0 & 0 & -S_{2y} & S_{2x} & 0 \end{bmatrix}.$$

The equations of motion (6.2) can be expressed compactly as $\frac{\mathrm{d}S}{\mathrm{d}t} = J(S)\nabla H(S)$. The functions $\|S_1\|$ and $\|S_2\|$ are a special type of first integral — a *Casimir*: for an arbitrary scalar function G, they satisfy

$$\{\|S_l\|, G\} = 0, \qquad l = 1, 2,$$

as is easily verified by computing $(\nabla\|S_l\|)^\top J(S) = 0$. Since $\|S_1\|$ and $\|S_2\|$ are the lengths of the spin vectors, the fact that they are constant simplifies the dynamics of the system.

The system can be put into canonical form by first making the coordinate transformation

$$\begin{aligned} S_{lx} &\mapsto r \sin\theta_l \cos\phi_l, \\ S_{ly} &\mapsto r \sin\theta_l \sin\phi_l, \\ S_{lz} &\mapsto r \cos\theta_l, \end{aligned}$$

and then setting $q_l = \phi_l$, $p_l = r\cos\theta_l$. (Since the spin lengths are constant we promptly, and without loss of generality, fix $r = 1$.) In the canonical coordinates the Hamiltonian reads

$$\begin{aligned} H(q,p) = &-\mu\left(\sqrt{1-p_1^2}\sqrt{1-p_2^2}\cos(q_1-q_2)\right) \\ &+\lambda\left(\sqrt{1-p_1^2}\cos q_1 + \sqrt{1-p_2^2}\cos q_2\right). \end{aligned} \tag{6.3}$$

The corresponding equations of motion then have the familiar form $\frac{\mathrm{d}}{\mathrm{d}t}(q^\top, p^\top)^\top = J^{-1}\nabla H(q,p)$, where J is the usual structure matrix (e.g as defined in equation (1.13)). When written

in terms of ϕ_l and θ_l the equations of motion become

$$\dot{\phi}_l = \mu \cos(\phi_l - \phi_m) \cot\theta_l \sin\theta_m - \lambda \cot\theta_l \cos\phi_l \tag{6.4}$$

$$\dot{\theta}_l = -\mu(-1)^l \sin\theta_m \sin(\phi_l - \phi_m) - \lambda \sin\phi_l, \qquad l = 1, 2. \tag{6.5}$$

The equations of motion (6.4) and (6.5) have the advantage that solutions of the equations are automatically constrained to the correct manifold — $\mathbb{S}^2 \times \mathbb{S}^2$ — independent of the numerical integrator. The price of this, is that (6.4) the equations for $\dot{\phi}_l$ are singular at $\theta_l = 0, \pi$. This can cause problems for numerical integrators since the differential equation can become arbitrarily large when one of the spins passes close to the north or south pole of the sphere.

The Poisson system of equations (6.2) in Cartesian coordinates is amenable to solution via a geometric integrator consisting of the composition of three planar rotations [40, 133, 134]. This is similar to the well studied case of the free rigid-body [84, 90, 55]. Possible advantages of such a treatment are:

- The system of equations given by (6.2) is free from the coordinate singularity which affects the system in spherical coordinates; (6.4) and (6.5). The singularity may cause a numerical integrator to make large errors (or may require a large step size reduction to avoid them).

- The integrator preserves the symplectic structure of the system and almost preserves the value of a nearby Hamiltonian for exponentially long times. An important consequence in this case is that the energy error in the computed solution oscillates about the correct value rather than drifting away as it would for the case of a classical integrator. This is important if one wishes to calculate iso-energetic Poincaré sections.

- The geometric integrator can be implemented reasonably cheaply. A first order method requires the composition of three rotations (i.e. multiplication by a pair of 3×3 rotation matrices), a second order method needs five rotations (multiplications). The second order method is symmetric, so, if output is not needed at the end of each step, the rotations at the end of one step can be combined with those at the beginning of the next and the number of matrix multiplications can be reduced to four. The trigonometric functions in the rotation matrices can be replaced by an approximation to further reduce the cost of the method.

- The rotation matrices used for the geometric integrator all have a determinant of one. This conserves the magnitude of the spins $\|S_l\|$, so that solutions remain on the correct manifold. This is generally not the case for a classical integrator applied to equation (6.2).

6.2 The generalised leapfrog integrator

If the Hamiltonian (6.1) is split as $H = H_x + H_y + H_\lambda = (-\mu S_{1x} S_{2x}) + (-\mu S_{1y} S_{2y}) + \lambda(S_{1x} + S_{2x})$ then the component vector fields in $X_H = X_{H_x} + X_{H_y} + X_{H_\lambda}$ can be integrated exactly. The solutions to the individual components of the vector field are given by rotations about the x and y axes for the interaction between the two spins and a rotation about the x axis due to the interaction between the spins and the applied magnetic field. In matrix form, these rotations are given by

$$R_{lx}(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha t & \sin\alpha t \\ 0 & -\sin\alpha t & \cos\alpha t \end{bmatrix}, \qquad R_{ly}(t) = \begin{bmatrix} \cos\beta t & 0 & -\sin\beta t \\ 0 & 1 & 0 \\ \sin\beta t & 0 & \cos\beta t \end{bmatrix},$$

$$R_{l\lambda}(t) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma t & \sin\gamma t \\ 0 & -\sin\gamma t & \cos\gamma t \end{bmatrix}, \quad l = 1, 2,$$

where $\alpha = \mu S_{mx}(t_0)$, $\beta = \mu S_{my}(t_0)$, $m = 1, 2$, $m \neq l$, and where $\gamma = -\lambda$.

Replacing t with a time step h and composing the rotations gives a symplectic integrator for the system (6.2). Using the notation

$$R_\bullet(h) = \begin{bmatrix} R_{1\bullet}(h) & 0 \\ 0 & R_{2\bullet}(h) \end{bmatrix},$$

an integrator of order p is given by choosing $a_i, b_i, c_i, \quad i = 1, \ldots, k$ such that $S(t_0 + h) = \prod_{i=1}^{k} R_x(a_i h) R_y(b_i h) R_\lambda(c_i h) S(t_0) + \mathcal{O}\left(h^{p+1}\right)$. We use the second order method given by $k = 6$ and $a_1 = b_2 = c_3 = c_4 = b_5 = a_6 = \frac{1}{2}$ with all other a, b, c set to zero. This is the well known "generalised leapfrog" method. The central stages of the method coalesce and can be combined. Since the composition is symmetric, when output is not required at every step, the outer stages of the composition can also be combined so that the method requires only four stages per time step.

More details on the derivation of splitting methods for geometric integration, in particular higher order splittings with a minimal number of stages, can be found in [147]. The choice of optimal coefficients for such splittings is discussed in [90].

6.3 Poincaré sections of the two-spin system

We study the iso-energetic Poincaré section given by $\theta_2 = \pi/2$, $\dot{\theta_2} > 0$, and $E = -0.1$ (see §4.6 for details on the procedure of constructing a numerical Poincaré section). The change of variables $\theta_l \mapsto \pi - \theta_l$ leaves the energy and the sign of $\dot{\theta_2}$ unchanged — it is a symmetry of the Poincaré section. When either λ or μ is zero, the two-spin system is integrable. In [118], Robb and Reichl solved the system of equations given by (6.4) and (6.5) for

$\lambda = 0.02, 0.2, 0.5, 2.5$ and with $\mu = 1$. Their goal was to show that for small or large values of λ the system is close to the $\lambda = 0$ and $\mu = 0$ integrable limits and, for intermediate values of λ, the system is far from integrable and displays chaos. Our goal, on-the-other-hand, is to show that such calculations are best done in a geometric fashion, otherwise one encounters computationally expensive, or even misleading, numerical results.

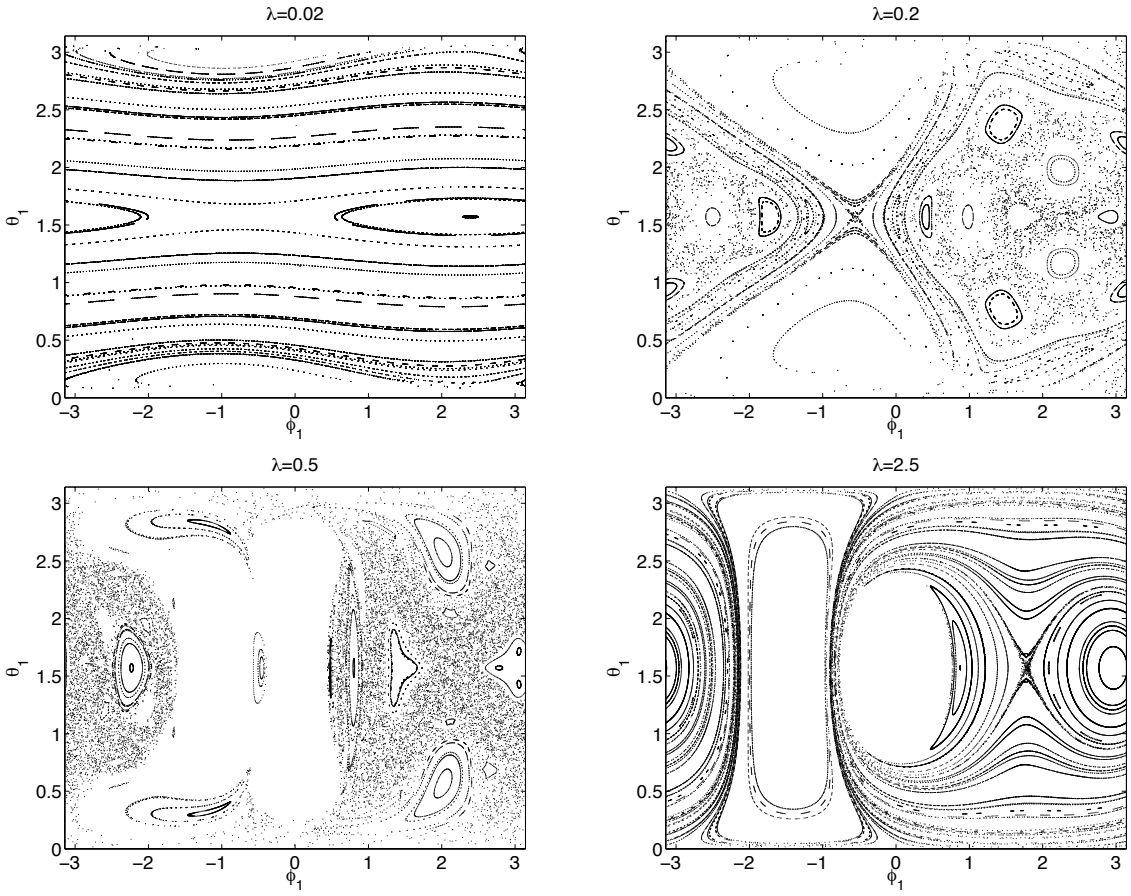


Figure 6.1: Poincaré sections of the coupled two-spin system for $\lambda = 0.02,\ 0.2,\ 0.5$ and 2.5 with energy $E = -0.1$ generated with the generalised leapfrog method with $\tau = 0.1$.

We calculate the same Poincaré sections as in [118], first using the same system of equations (6.4) and (6.5) as Robb and Reichl, and integrating the equations with a classical "black-box" integrator. (In this case `ode45` from the Matlab package *odesuite*, with `RelTol`=10^{-4}, an order (4,5) Dormand-Prince pair with step size control. See [128] for details). We then recalculate the same Poincaré sections using the Poisson system of equations (6.2) and the generalised leapfrog integrator of section 6.2. The Poincaré sections generated by the generalised leapfrog method are shown in figure 6.1, those from the "black-box" integrator are shown in figure 6.2.

The "natural" manifold for the Poincaré sections is not $\mathbb{R}^2$, but rather, the $\mathbb{S}^2$ sphere — plotting the sections on this surface prevents the regions about $\theta_1 = 0$ and $\theta_1 = \pi$

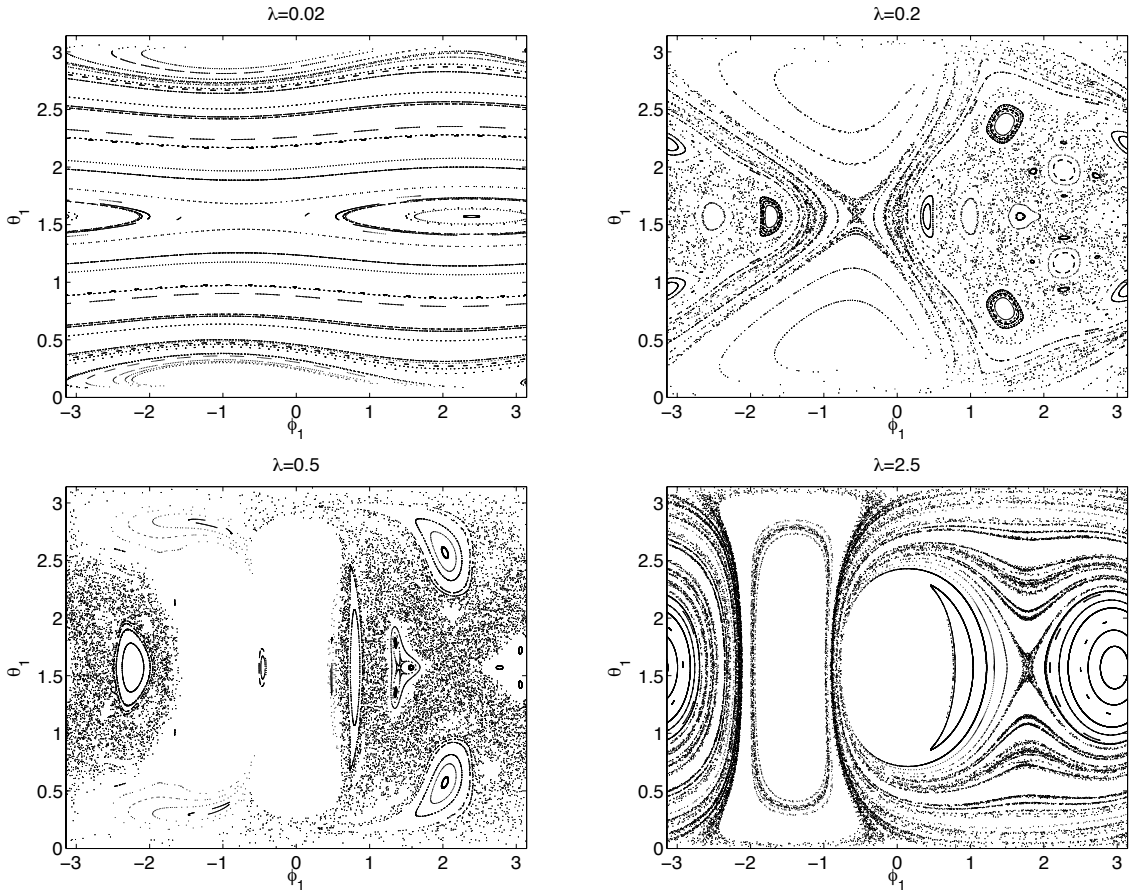


Figure 6.2: Poincaré sections of the coupled two-spin system with applied magnetic field for $\lambda = 0.02,\ 0.2,\ 0.5$ and 2.5 with energy $E = -0.1$ with a classical "black-box" integrator; `ode45` with `RelTol`=10^{-4}.

from being distorted. Figure 6.3 shows the resulting view of the Poincaré sections for the generalised leapfrog method.

Both sets of Poincaré sections manage to demonstrate that for $\lambda = 0.02$ the two-spin system is close to integrable, while for $\lambda = 0.5$ it is mostly chaotic, though the computational cost for the "black-box" method is an order of magnitude greater. This increased cost is most likely due to the integrator struggling to maintain accuracy for trajectories which pass close to the coordinate singularities at $\theta_l = 0, \pi$. (We look closer at the effect of the coordinate singularities on accuracy in §6.4.1.) Even when the Poincaré sections show large regions of chaos (e.g. for $\lambda = 0.2, 0.5$) the sections still show significant regions of regular dynamics. KAM tori can be clearly seen — particularly in the sections of the generalised leapfrog method, which being symplectic preserves most invariant tori. The classical integrator, which is not symplectic, does less well — though through brute force (via error estimation and step size control) — it manges to reproduce these structures to some extent for an integration period equivalent to about 9000 time steps of the leapfrog

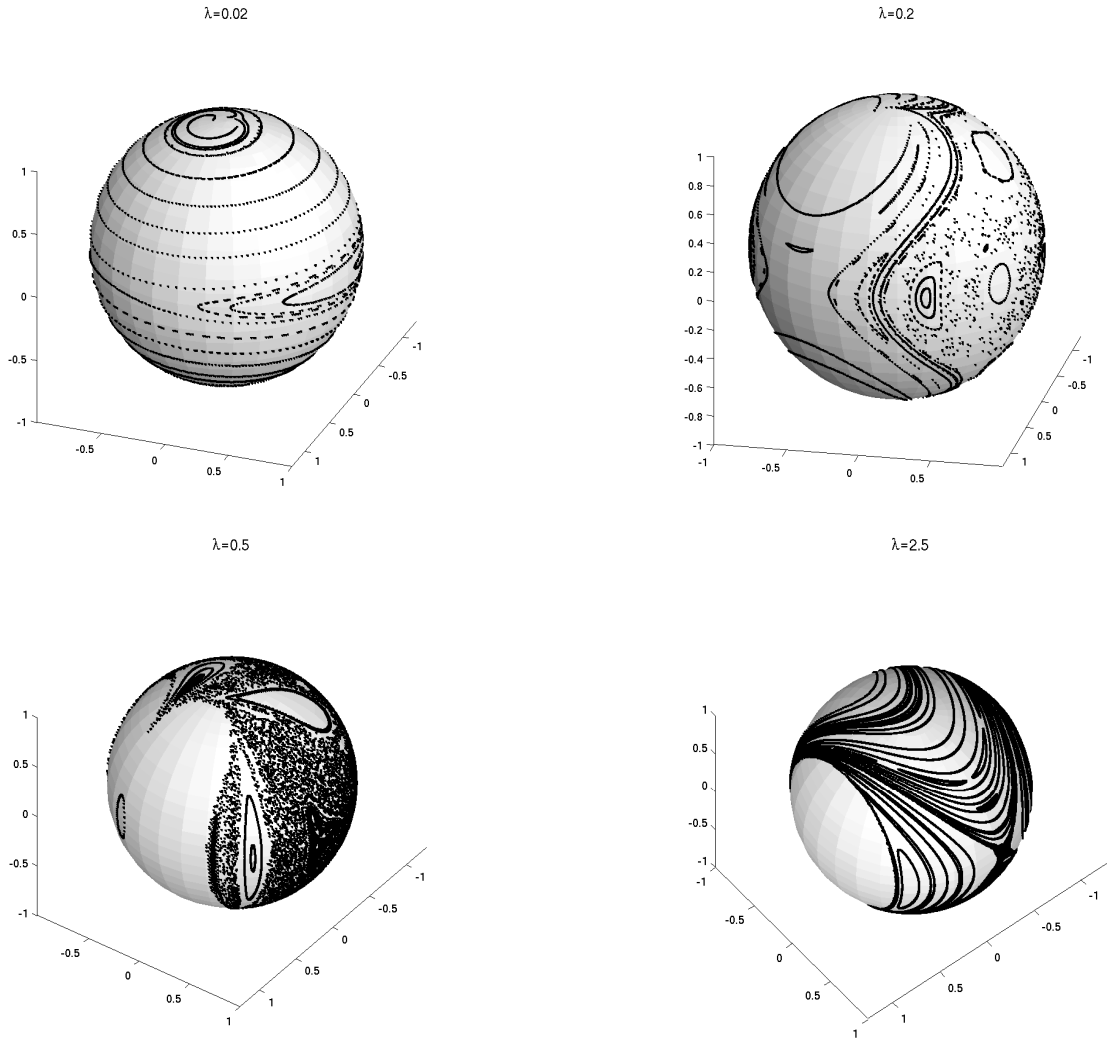


Figure 6.3: Poincaré sections of the coupled two-spin system with an applied magnetic field for $\lambda = 0.02,\ 0.2,\ 0.5$ and 2.5 with energy $E = -0.1$ generated with the generalised leapfrog method when plotted in spherical coordinates.

method.

Both integrators, to some extent, show "smearing" of the invariant curves in the Poincaré sections, though this effect is far more pronounced for the classical integrator (where one might be excused for confusing the distorted invariant curves of the $\lambda = 2.5$ section with chaotic behaviour). For the classical integrator, the energy of the trajectories drifts away from its initial value, rather than remaining close to it as the generalised leapfrog integrator does — again due to the fact that leapfrog is symplectic. We quantify this effect in section 6.4.1.

When calculating the Poincaré sections, one quickly sees that not all the points on the sphere correspond to initial conditions for the Poincaré map. That is, there are regions

of the Poincaré section which do not intersect the constant energy surface (in this case $E = -0.1$). In order to find the boundaries of such regions we set the Hamiltonian (6.1) equal to a fixed energy E. Since $S_{2z} = 0$ for the Poincaré section the requirement that $\|S_2\| = 1$ becomes $S_{2x}^2 + S_{2y}^2 = 1$, or rather, that S_2 lie on the unit circle. The intersection of the Poincaré section with the constant energy surface is therefore given by

$$-\mu(S_{1x}\cos\phi_2 + S_{1y}\sin\phi_2) + \lambda(S_{1x} + \cos\phi_2) = E.$$

This has at least one solution when $(S_{1x} - \lambda/\mu)^2 + S_{1y}^2 \geq (E - \lambda S_{1x})^2/\mu^2$. The projection of the excluded region onto the equatorial (xy) plane by $S_z = 0$ is given by the intersection of the unit disk with the interior of the closed curve

$$S_{1y}^2 = \frac{(E - \lambda S_{1x})^2 - (\mu S_{1x} - \lambda)^2}{\mu^2}, \qquad S_{1x} \in [-1, 1]. \tag{6.6}$$

Then, re-projecting onto the surface of the unit sphere, the boundary of the excluded region is given by $S_{1z}^2 = 1 - S_{1x}^2 - S_{1y}^2$.

In figure 6.4 we indicate the excluded regions of the Poincaré sections, plotted on the sphere, and viewed from the north pole. When λ is small the excluded regions include the north and south poles of the spheres. For large values of λ the excluded region also becomes large and its centre shifts towards the equatorial plane.

The equations of motion for the system allow for trajectories of S_2 which are tangential to, but do not cross, the S_{2z} equator. Such trajectories are not included in the Poincaré sections as they do not satisfy the condition $\dot{S}_{2z} < 0$. By solving the system of equations $\dot{S}_{2z} = 0$, $\|S_1\| = 1$, $\|S_2\| = 1$, $S_{2z} = 0$ and the constant energy equation $E = -\mu(S_{1x}S_{2x} + S_{1y}S_{2y}) + \lambda(S_{1x} + S_{2x})$ we can obtain the equation for the curve in the Poincaré section which corresponds to the tangential crossings. This is exactly the curve described by equation (6.6) and which gives the boundary of the excluded region of the Poincaré section. That is, the boundaries of the excluded regions in figure 6.4 are the curves where the flow is tangential to the Poincaré sections. The Poincaré map can become singular at the boundary of the excluded region, and this (rather than the coordinate singularity to which Robb and Reichl attribute this effect), explains the apparently singular behaviour of some of the invariant curves in figure 6.4.

6.4 Accuracy, local error & energy preservation

6.4.1 Effect of the coordinate singularity on local error

As mentioned in section 6.1, the coordinate singularity in the equations of motion (6.4) and (6.5) in spherical coordinates can cause large local errors in the solutions of numerical integrators when the solution passes close to the singularity. Quantifying the effect of the singularity on the error committed by the numerical integrator is complicated by the fact

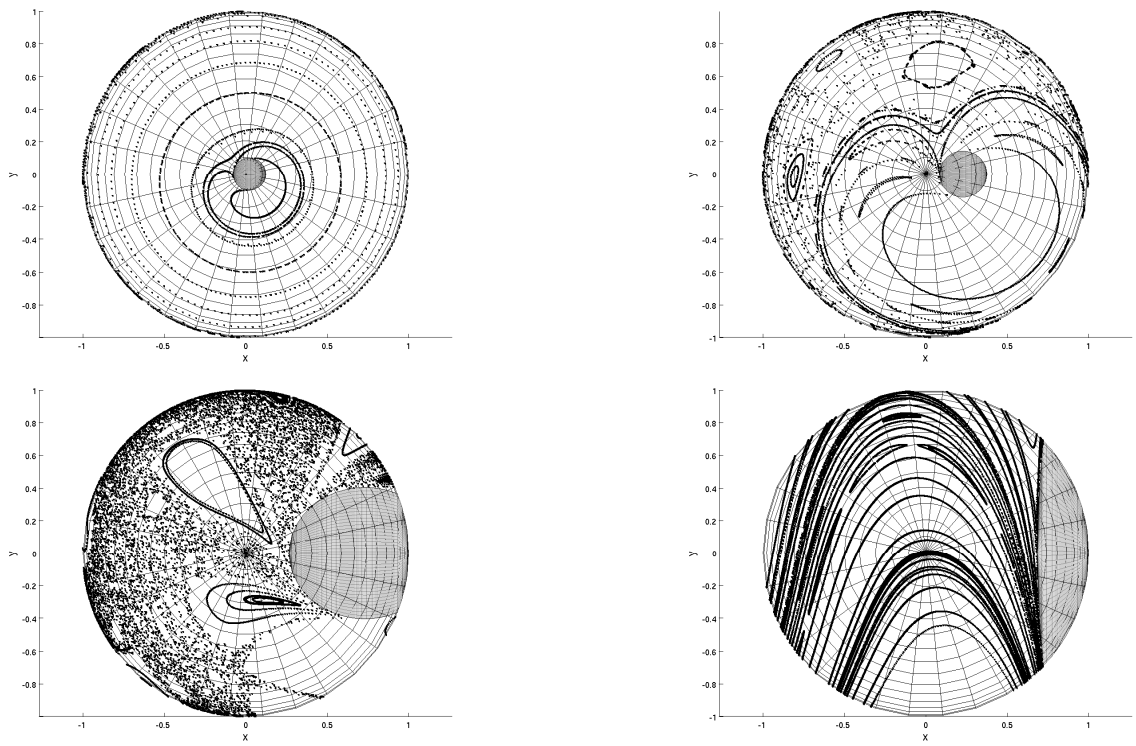


Figure 6.4: Poincaré sections of the coupled two-spin system with applied magnetic field for $E = -0.1$ and $\lambda = 0.02,\ 0.2,\ 0.5$ and 2.5 as viewed from the north pole of the sphere. The excluded regions are shaded.

that for a fixed value of, say, θ_1 we have (in spherical coordinates) four other parameters ($\phi_1,\ \phi_2,\ \theta_2$ and λ) which can affect the error in a single step.

Our approach is to hold ϕ_2 and θ_2 fixed at $\pi/4$ and $\pi/2$ respectively and to calculate the local error, with respect to a numerically computed reference solution, in the Euclidean norm for a range of values of ϕ_1, θ_1 and λ. We then take the worst case over the range of ϕ_1 values. That is, we calculate

$$\mathrm{Err}(\theta_1, \lambda) = \max_{\phi_1 \in \mathcal{U}} \left\| \Phi_h(\phi_1, \phi_2, \theta_1, \theta_2) - \Phi_h^{ref}(\phi_1, \phi_2, \theta_1, \theta_2) \right\|, \qquad (\phi_2, \theta_2) = (\pi/4, \pi/2),$$

where $\mathcal{U}$ is a grid of ϕ_1 values in the interval $[0, \pi]$. This allows the local error to be plotted as a function of θ_1 for various values of λ as shown in figure 6.5. We compare the results from Heun's method — the explicit, second order Runge-Kutta method with Butcher tableau $\begin{array}{c|cc} \frac{1}{2} & \frac{1}{2} & \\ \hline & 1 & \end{array}$ — applied to the system of differential equations (6.4), (6.5), with those from the generalised leapfrog method (also second order and explicit) applied to the singularity free system (6.2). Both integrators used a step size of $h = 0.1$.

By fixing θ_2 at $\pi/2$, well away from the value which would lead to a singularity in the

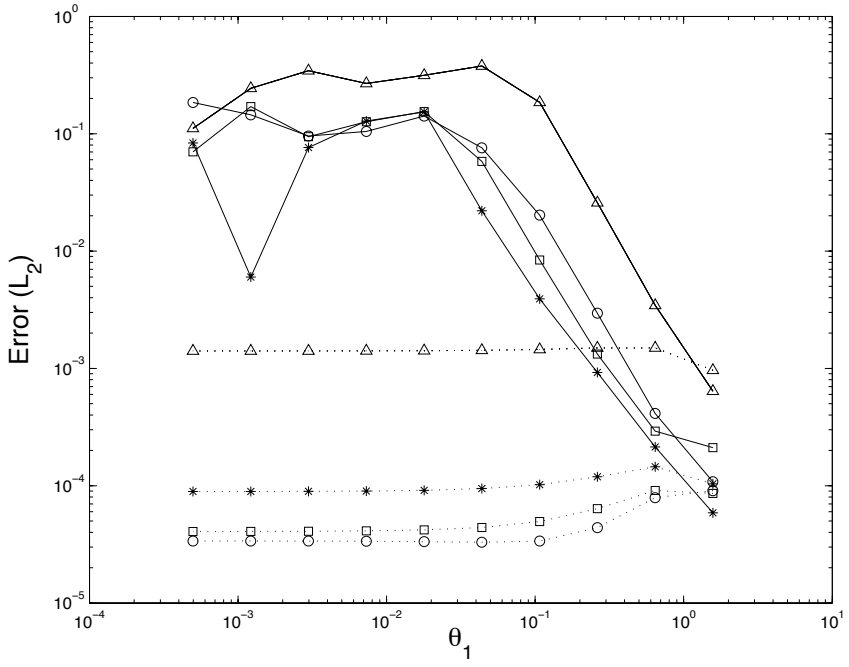


Figure 6.5: Local error as a function of θ_1 for Heun's method applied to (6.4) & (6.5) (*solid lines*) and for the generalised leapfrog method applied to (6.2) (*dotted lines*). Initial conditions were $\phi_2 = \pi/4$, $\theta_2 = \pi/2$ with ϕ_1 chosen so as to give an approximate "worst-case" for the local error from each method. Four values of λ were used: 0.02, 0.2, 0.5 and 2.5 (*circles, boxes, stars and triangles respectively*). As θ_1 approaches zero the local error for the system in spherical coordinates grows like $\mathcal{O}\left(\theta_1^{-2}\right)$.

equations of motion for $\dot{\phi}_2$ we attempt to isolate the effect of the singularity to the $\dot{\phi}_1$. Numerical investigation indicated that the effect of θ_1 and θ_2 both approaching zero does not appear to be additive, that is the error when both θ_1 and θ_2 are near zero is of the same size as when either θ_1 or θ_2 are near zero.

For values of θ_1 (and θ_2) away from the singularity the local errors for the two integrators, and two coordinate systems, are of the same order of magnitude. As θ_1 approaches zero the local error for Heun's method with the system (6.4) and (6.5) grows like $\mathcal{O}\left(\theta_1^{-2}\right)$. The same rate of error growth as $\theta_1 \to 0$ was seen for other initial conditions, hence, when the singular equations of motion are used, if the step size is not reduced (near the singularity) the local error is likely to shift the numerical solution away from the true solution.

6.4.2 Energy preservation

The Hamiltonian, or total energy, is an invariant of the two-spin system. Here we briefly look at the effect of both, the choice of integrator, and the choice of coordinate system, on energy conservation.

In figure 6.6 we plot the total energy and the value of θ_1 for the numerical solutions from Heun's method applied to (6.4), (6.5). The same calculation for the generalised leapfrog method, applied to the differential equation (6.2) is shown for comparison. Both methods used a step size of $h = 0.1$. Jumps in the energy of the solution from Heun's method can be seen when θ_1 passes close to zero.

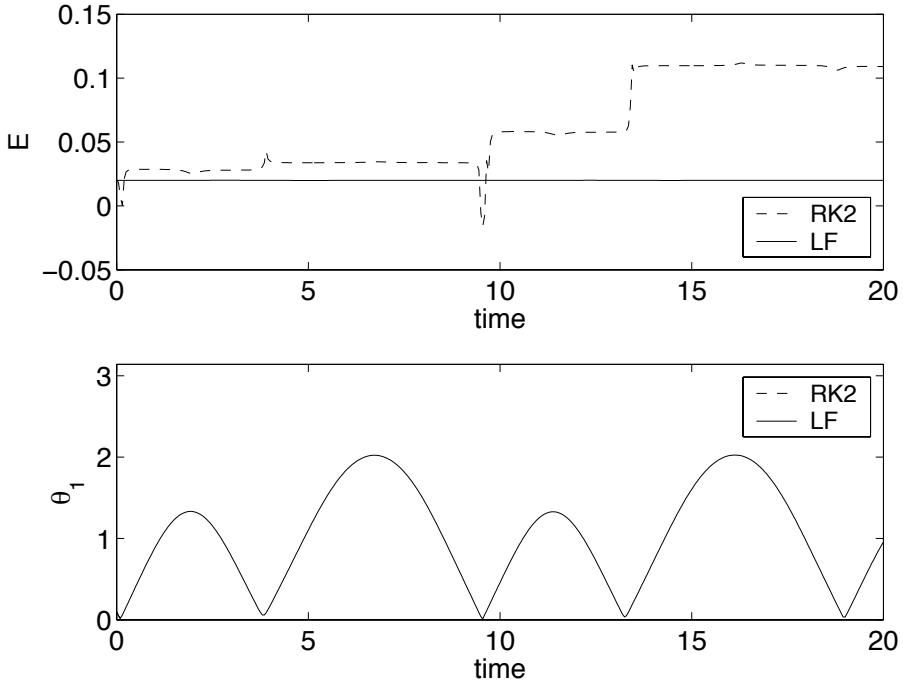


Figure 6.6: Energy E and θ_1 for solutions of the two-spin system with equations of motion (6.4) & (6.5) integrated with Heun's method (RK2 — *dashed lines*). The same calculation for the differential equation (6.2) and computed with the generalised leapfrog method (LF — *solid lines*) is given for comparison. The jumps in the energy of the RK2 solution can be seen to correspond to points where θ_1 (modulo π) approaches zero. For this example $\lambda = 0.2$, $h = 0.1$. The initial conditions were $(\phi_1, \phi_2, \theta_1, \theta_2) = (0, \pi/2, 0.1, \pi/2)$.

The same effect can be seen in a more striking fashion when the integration period is large; figure 6.7. In fact, the singularity due to the choice of coordinate system has a far larger effect on the energy error than the choice of integrator. When Heun's method is applied to the equations of motion (6.2), which are free of coordinate singularities, one observes energy drift, which though far worse than for the (symplectic) generalised leapfrog method, is never-the-less, an order of magnitude less than for the differential equations (6.4) and (6.5) — as figure 6.8 shows.

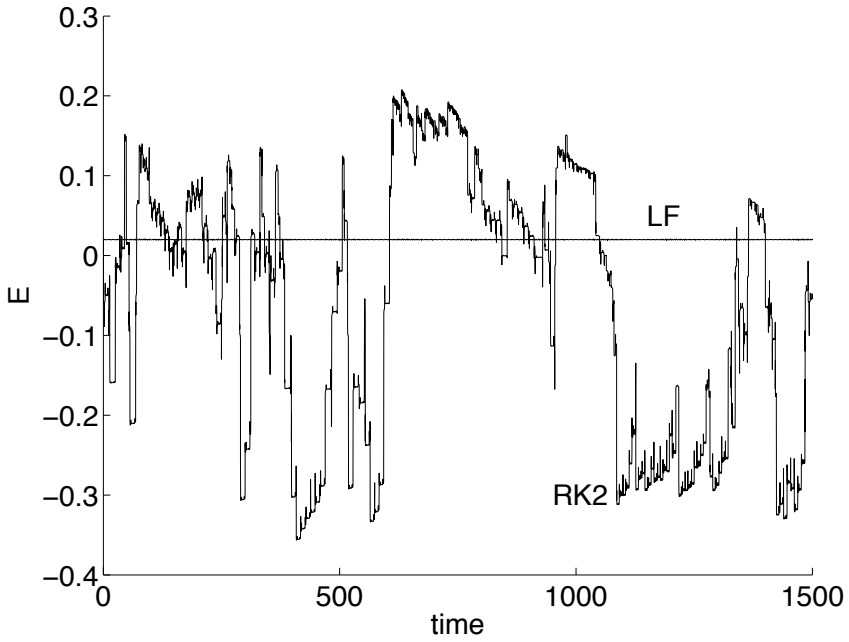


Figure 6.7: While the energy remains nearly constant for the system (6.2) in Cartesian coordinates when integrated with the generalised leapfrog method (LF), the energy of the system (6.4) & (6.5) in spherical coordinates, and integrated with Heun's method (RK2) fluctuates dramatically. The initial conditions used here were $(\phi_1, \phi_2, \theta_1, \theta_2) = (0, \pi/2, 0.1, \pi/2)$ with $\lambda = 0.2$ and a step size $h = 0.1$.

6.5 Conclusion

The choice of both coordinate system and integration method are significant for the solution of the equations of motion for the coupled two-spin system (or for spin systems more generally). Although the spherical coordinates used in [118] automatically restrict the solution to the correct manifold when using a classical integrator, the price of this is the singularities introduced to the equations of motion.

The geometric integrator consisting of a composition of planar rotations avoids the problem of singularities by working in Cartesian coordinates but does not destroy the property that the solutions remain restricted to the correct manifold ($\mathbb{S}^2 \times \mathbb{S}^2$), as a classical integrator might in the same coordinate system. In order to mitigate growth of the local error near the singularities (when using a classical integrator) it was necessary to adjust the step size of the integrator. This made calculation of trajectories (for the Poincaré sections) with the generalised leapfrog integrator significantly faster even when compared with a classical integrator of twice the order.

The Poincaré sections produced from the two different methods and coordinate systems are comparable in that both are able to show most of the same features, though those from the classical integrator and using the spherical coordinates are more expensive to

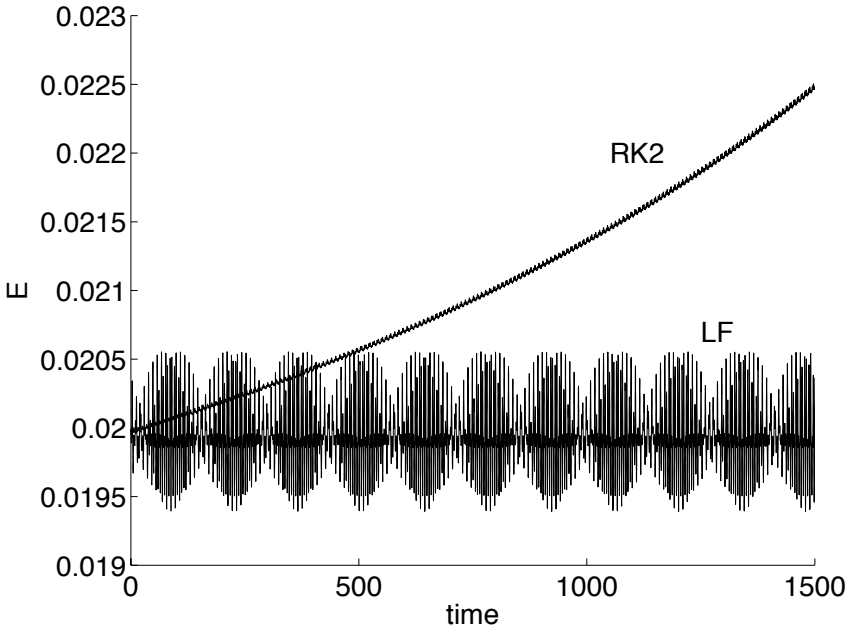


Figure 6.8: Solving the two-spin system (6.2) in Cartesian coordinates with Heun's method (RK2) causes a gradual drift in the value of the energy while the energy of the solution from the leapfrog integrator (LF) oscillates about the initial value. Again, $(\phi_1, \phi_2, \theta_1, \theta_2) = (0, \pi/2, 0.1, \pi/2)$ with $\lambda = 0.2$ and $h = 0.1$.

produce and are more susceptible to "smearing" of the invariant curves if care is not taken to control the local error near the singularities. Both methods display chaotic orbits for the same parameters and in the same regions of phase space. As expected, the generalised leapfrog method excels at preserving invariant structures, such as KAM tori and periodic orbits, on account of being symplectic. The persistence of empty regions in the Poincaré sections are due to the particular choice of the constant energy surface which does not satisfy the Hamiltonian for all points in the phase space. On the boundary of the excluded regions the trajectories of the system are tangential to the Poincaré section leading to the apparently singular behaviour of some of the orbits near this region.

The combination of the spherical coordinates and the classical integrator causes jumps in the energy when the solution passes near to a singularity. These are an order of magnitude larger than the energy itself and give greater energy errors than those due to the familiar phenomenon of energy drift which occurs with classical integrators.

Chapter 7

Closing Remarks and Open Questions

The aim of this thesis was to give insight into the preservation of phase space structure of Hamiltonian differential equations, in particular by symplectic numerical integration algorithms. This insight arises in two ways: theorems and conjectures on the conditions under which integrators preserve phase space structures, and applications illustrating the consequences of different treatments of these structures by different types of numerical integrators.

Chapter 1 gave a background to numerical integration of differential equations including linear multistep, (partitioned) Runge-Kutta, and splitting methods. Hamiltonian systems were introduced and their features and dynamics, such as energy- area- and volume-preservation and symplecticity, were discussed. Geometric numerical integration was introduced. Examples of current (in particular, symplectic) integrators were given and the difference between a classical and a geometric approach to numerical integration was presented. As part of this, backward error analysis was explained.

Chapter 2 discussed what is currently known about the preservation of invariant sets of differential equations by numerical integrators, geometric or otherwise. This included fixed points and their linearizations, stable/unstable manifolds, periodic orbits, invariant tori and chaotic sets as well as statistical properties. It also presented a theorem concerning preservation of fixed points by splitting methods, which we are unaware of existing previously in the literature.

In preparation for dealing with the preservation of periodic orbits, chapter 3 gave a background to KAM theory including a description of Kolmogorov's iteration and proof of its convergence for perturbations of integrable Hamiltonian systems.

Chapter 4 discussed the preservation of periodic orbits of Hamiltonian differential equations when discretized with a symplectic integrator. (Detailed summaries of chapters 4, 5 and 6 occur at the ends of the respective chapters.) The important differences between periodic orbits of Hamiltonian systems and of other differential equations were explained:

periodic orbits of Hamiltonian systems form 1-parameter families while periodic orbits of dissipative systems are generally isolated.

A result of Moan was used to embed the numerical trajectory of a symplectic integrator in a non-autonomous flow. By applying a KAM theorem of Jorba and Villaneuva it was then shown that when a periodic orbit of a Hamiltonian system is discretized by a symplectic integrator with a step size strongly non-resonant with the period of the orbit, the periodic orbit is replaced by an invariant 2-torus with one frequency close to that of the original frequency of the periodic orbit and one given by $2\pi/h$. The magnitude of the motion in the new periodic coordinate is exponentially small.

This result differs from the existing result of Shang, concerning preservation of *full dimensional* invariant tori by symplectic integrators, in that it perturbs the frequency of the periodic orbit. It is an interesting (and difficult) question as to whether an approach, such as Shang's, which gives a direct KAM theorem for symplectic maps, without any embedding step, can be obtained for lower dimensional tori such as periodic orbits.

The result for the preservation of periodic orbits holds for any lower dimensional tori which satisfy the conditions of the theorem. This includes fixed points (tori of dimension zero) and full dimensional tori. An interesting idea is to use the results of the theorem for lower dimensional tori to study cases such as fixed points, or full dimensional tori where results are already known from different theorems. Any differences or similarities may give a better understanding of the dynamics of Hamiltonian systems.

Chapter 4 also conjectured that for systems of two or more degrees of freedom, periodic orbits discretized by a symplectic integrator with a resonant step size break up into sets of elliptic and hyperbolic periodic points in the same fashion as periodic orbits of one degree of freedom systems break up under the Poincaré-Birkhoff fixed point theorem. A detailed numerical investigation of the discretization of periodic orbits in the two degree of freedom Hénon-Heiles system was used to support this conjecture. In light of this, the effect of this break up on steady-state solutions of Hamiltonian partial differential equations with periodic boundary solutions should be studied. We explained in section 4.6 of chapter 4 that in such a situation the conjecture applies, since the (symplectic) discretization of such a system necessarily leads to a periodic orbit discretized by an integrator with a resonant step size. The implication is that only symplectic discretizations of Hamiltonian partial differential equations with periodic boundary conditions reproduce the steady-state solutions of the system. Additionally, a similar, though more complicated, situation may occur for travelling waves with constant wave speed if the wave speed is a rational multiple of the spatial grid spacing.

In chapter 4, the numerical investigations of resonance only considered the effect of resonances between the frequency of a periodic orbit and the step size of a numerical integrator. However, the strong non-resonance condition in theorem 4.6 is more general than this: it also includes the effects of the "normal" frequencies — the Floquet multipliers (or their generalizations). Investigation of the behaviour of a symplectic integrator for

resonances involving this sort of condition appears to be an open (and difficult) question. A more complete description of the phase portrait in the neighbourhood of a periodic orbit also remains as an open problem.

The KAM theorem of chapter 4 holds the perturbation size fixed and uses the frequency of the periodic orbit as a parameter. The reason being that the fixed perturbation size corresponds to a fixed non-autonomous perturbation of the original differential equation due to the embedding theorem, and this perturbation size in turn, corresponds to a step size of the symplectic integrator. Hence, changing the step size of the integrator affects both the perturbation size and the perturbation frequency. It seems unlikely, therefore that it would be possible to use the perturbation size as parameter to ensure convergence in a KAM-type theorem. However, investigation of this possibility may yield a better understanding of the dynamics, if not a positive result for preservation of periodic orbits. One possibility in this regard is that the non-autonomous perturbation need not be unique. There may be a family of perturbations of different sizes but with the same step size and, hence, frequency.

Chapter 5 dealt with integrators designed for Hamiltonian systems of differential equations with constant, high frequency, oscillations arising from a quadratic term. It presented trigonometric integrators — a class of numerical integrator which can solve such problems with a step size equal to, or greater than, one period of the fast oscillation. Such integrators are computationally efficient but little is known about their geometric properties including their treatment of phase space structure. The methods can be symplectic but often are not. When they are, it does not seem to help them greatly in preserving the dynamics of differential equations.

Trigonometric integrators were studied with respect to a range of criteria: energy and oscillatory energy preservation, slow exchange of oscillatory energy and statistical properties concerning the amount of time their solutions spent in different regions of the phase space of a chaotic differential equation — the Fermi-Pasta-Ulam system. It was shown that the trigonometric integrators suffer from higher order resonances, in addition to the lower order resonances which were already known to affect them. It was also shown that those trigonometric integrators which suppress low order resonances do so at the cost of forcing invariant or adiabatic quantities to be conserved at the wrong values. Such results are concerning since there are no associated warning signs, such as instability or rapid energy growth, to indicate that something is awry. The results of the trigonometric integrators were compared to those of the midpoint rule using the same step size. It was shown that none of the trigonometric methods managed to perform well on all the tests, and some performed worse than the midpoint rule.

Open questions from this chapter include what sort of properties should one require such integrators to preserve, in addition to, (or instead of), the properties used as criteria in evaluating the trigonometric integrators. Also, is it possible to construct a numerical integrator which manages to take step sizes larger than one period yet still manages to

preserve structural properties of the differential equations? Two promising avenues for this last point are multi-force trigonometric integrators and Hamilton-Jacobi integrators. The first use a linear combination of filter functions and therefore allow greater flexibility in the conditions necessary for the method to preserve various properties. The Hamilton-Jacobi methods show promise since they are derived via a generating function and are automatically symplectic; results such as KAM theory can therefore be applied. Such methods, however, are rather complicated and, potentially, computationally expensive.

Chapter 6 considered the example of a coupled two-spin system in a magnetic field and dealt with the most fundamental phase space structure — the phase space itself. Depending on the system of coordinates in which it is written, the differential equations may automatically preserve the correct phase space with *any* numerical integrators, but at the cost of introducing a coordinate singularity into the equations of motion.

A geometric integrator (the generalised leapfrog method) was constructed such that the phase space was preserved in a coordinate system without singularities. Results from a classical integrator were compared with those of the generalised leapfrog method for producing Poincaré sections. The effect of the coordinate singularity (and the choice of integrator) on the accuracy and the energy preservation were illustrated.

Bibliography

[1] M. J. Ablowitz, B. M. Herbst, and C. M. Schober. Discretizations, integrable systems and computation. *J. Phys. A*, 43:10671–10693, 2001.

[2] V. I. Arnold. Small denominators and problems of stability of motion in classical and celestial mechanics, cited in [55]. *Russian Math. Surveys*, 18:85–191, 1963.

[3] V. I. Arnold. *Geometrical Methods in the Theory of Ordinary Differential Equations*. Springer-Verlag, New York, 2nd edition, 1988.

[4] V. I. Arnold. *Mathematical Methods of Classical Mechanics*, volume 60 of *Graduate Texts in Mathematics*. Springer-Verlag, Berlin, 2nd edition, 1989.

[5] D. K. Arrowsmith and C. M. Place. *An Introduction to Dynamical Systems*. Cambridge University Press, Cambridge, 1990.

[6] U. M. Ascher and S. Reich. On some difficulties in integrating highly oscillatory Hamiltonian systems. Preprint SC 97-50, Konrad-Zuse-Zentrum für Infomationstechnik, Berlin, 1997.

[7] U. M. Ascher and S. Reich. The midpoint scheme and variants for Hamiltonian systems: advantages and pitfalls. *SIAM J. Sci. Comput.*, 21(3):1045–1065, 1999.

[8] A. Banyaga. *The Structure of Classical Diffeomorphism Groups*, volume 400 of *Mathematics and Its Applications*. Kluwer Academic Publishers, Boston, 1997.

[9] F. Bashforth and J. C. Adams. *An attempt to test the theories of capillary action by comparing the theoritical and measured forms of drops of fluid. With an explanation of the method of integration employed in constructing the tables which give the theoretical forms of such drops*. Cambridge University Press, Cambridge, 1883.

[10] G. Benettin, L. Galgani, and A. Giorgilli. Realization of holonomic constraints and freezing of high frequency degrees of freedom in the light of classical perturbation theory. part I. *Comm. Math. Phys.*, 113:87–103, 1987.

[11] G. Benettin and A. Giorgilli. On the Hamiltonian interpolation of near-to-the-identity symplectic mappings with application to symplectic integration algorithms. *J. Stat. Phys.*, 74(5/6):1117–1143, 1994.

[12] G. D. Birkhoff. Proof of Poincaré's geometric theorem. *Transactions of the American Mathematical Society*, 14(1):14–22, 1913.

[13] E. Bour. L'intégration des équations différentielles de la mécanique analytique. *J. Math. Pures et Appliquées, cited in [55]*, 20:185–200, 1855.

[14] T. J. Bridges. Multi-symplectic structures and wave propagation. *Math. Proc. Cambridge Philos. Soc.*, 121:147–190, 1997.

[15] T. J. Bridges and S. Reich. Multi-symplectic integrators: numerical schemes for Hamiltonian PDEs that conserve symplecticity. *Phys. Lett. A*, 284(4–5):184–193, 2001.

[16] C. J. Budd and M. D. Piggott. Geometric integration and its applications. In *Handbook of Numerical Analysis*, pages 35–139. North-Holland, 2003.

[17] J. C. Butcher. On the convergence of numerical solutions to ordinary differential equations. *Math. Comp.*, 20:1–10, 1966.

[18] J. C. Butcher. A multistep generalization of Runge-Kutta methods with four or five stages. *J. Assoc. Comput. Mach.*, 14:84–99, 1967.

[19] J. C. Butcher. General linear methods. *Acta Numer.*, 15:157–256, 2006.

[20] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. Wiley, England, 2nd edition, 2008.

[21] J. C. Butcher and N. Moir. Experiments with a new fifth order method. *Numer. Algorithms*, 33:137–151, 2003.

[22] J. C. Butcher and W. M. Wright. The construction of practical general linear methods. *BIT Numerical Mathematics*, 43:695–721, 2003.

[23] P. J. Channell. Initial value and eigenvalue problems for field equations using symplectic integration algorithms. Internal report AT-6:ATN-83-18, Los Alamos National Laboratory, cited in [24], 1983.

[24] P. J. Channell and C. Scovel. Symplectic integration of Hamiltonian systems. *Nonlinearity*, 3:231–259, 1990.

[25] P. Chartier, E. Faou, and A. Murua. An algebraic approach to invariant preserving integrators: The case of quadratic and Hamiltonian invariants. *Num. Math.*, 103(4):575–590, 2006.

[26] L. Chierchia. Absolutely continuous spectra of quasi-periodic Schrödinger operators. *J. Math. Phys.*, 28:2891–2898, 1987.

[27] D. Cohen, E. Hairer, and Ch. Lubich. Modulated Fourier expansions of highly oscillatory differential equations. *Found. Comput. Math.*, 3:327–345, 2003.

[28] D. Cohen, T. Jahnke, K. Lorenz, and Ch. Lubich. Numerical integrators for highly oscillatory Hamiltonian systems: a review. In A. Mielke, editor, *Analysis, Modeling and Simulation of Multiscale Problems*. Springer, Berlin, 2006.

[29] C. F. Curtiss and J. O. Hirschfelder. Integration of stiff equations. *Proc. Natl Acad. Sci. USA*, 38:235–243, 1952.

[30] G. Dahlquist. A special stability property for linear multistep methods. *BIT*, 3:27–43, 1963.

[31] R. de la Llave. A tutorial on KAM theory. In A. Katok, R. de la Llave, Y. Pesin, and H. Weiss, editors, *Smooth ergodic theory and its applications. Proceedings of the AMS Summer Research Institute held at the University of Washington, Seattle, WA, July 26–August 13, 1999.*, volume 69 of *Proceedings of Symposia in Pure Mathematics*, pages 175–292. American Mathematics Society, 2001.

[32] R. de Vogelaere. Methods of integration which preserve the contact transformation property of the Hamiltonian equations. Technical Report 4, Dept. Math., University of Notre Dame, cited in [24], Notre Dame, Indiana, 1956.

[33] P. Deuflhard. A study of extrapolation methods based on multistep schemes without parasitic solutions. *Z. Angew. Math. Phys.*, 30:177–189, 1979.

[34] R. Douady. *Applications du théorème des tores invariantes*. PhD thesis, Université Paris VII, 1982.

[35] R. Douady. Une démonstration directe de l'équivalence des théorèms de tores invariants pour difféomorphismes et champs de vecteurs. *C. R. Acad. Sci. Paris*, pages 201–204, 1982.

[36] L. H. Eliasson. Floquet solutions for the 1-dimensional quasi-periodic Schrödinger equation. *Comm. Math. Phys.*, 146(3):447–482, 1992.

[37] J. Emerson and L. E. Ballentine. Characteristics of quantum-classical correspondence for two interacting spins. *Phys. Rev. A*, 63(5), 2001.

[38] L. Euler. *Institutiones Calculi Integralis, Volumen Primum.* Opera Omnia, Vol. XI, cited in [56], 1768.

[39] E. Fermi, J. Pasta, and S. Ulam. Studies of nonlinear problems I. In A. C. Newell, editor, *Nonlinear Wave Motion*, volume 15 of *Lectures in Applied Mathematics*, pages 143–156. AMS, 1974. First published as Los Alamos Report No. LA-1940 (1955).

[40] J. Frank, W. Huang, and B. Leimkuhler. Geometric integrators for classical spin systems. *J. Comput. Phys.*, 133:160–172, 1997.

[41] B. García-Archilla, J. M. Sanz-Serna, and R. D. Skeel. Long-time-step methods for oscillatory differential equations. *SIAM J. Sci. Comput.*, 20(3):930–963, 1998.

[42] W. Gautschi. Numerical integration of ordinary differential equations based on trigonometric polynomials. *Num. Math.*, 3(1):381–397, 1961.

[43] S. Gill. A process for the step-by-step integration of differential equations in an automatic digital computing machine. *Math. Proc. Cambridge Philos. Soc.*, 47:96–108, 1951.

[44] B. Gladman, M. Duncan, and J. Candy. Symplectic integrators for long-time integrations in celestial mechanics. *Celestial Mech. Dynam. Astronom.*, 52(3):221–240, 1991.

[45] E. Griepentrog. Gemischte Runge-Kutta-Verfahren für steife Systeme. In *Seminarbericht Sekt. Math.*, volume 11, pages 19–29. Humbolt-Univ. Berlin, 1978.

[46] V. Grimm and M. Hochbruck. Error analysis of exponential integrators for oscillatory second-order differential equations. *J. Phys. A*, 39:5495–5507, 2006.

[47] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.*, 6(1–3):121–142, 1991.

[48] J. Guckenheimer. Numerical analysis of dynamical systems. In B. Fiedler, editor, *Handbook of Dynamical Systems*, volume 2, pages 345–390. North-Holland, Amsterdam, 2002.

[49] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, volume 42 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, 1983.

[50] E. Hairer and M. Hairer. GniCodes — MATLAB programes for geometric numerical integration. In J. F. Blowey, A. W. Craig, and T. Shardlow, editors, *Frontiers in Numerical Analysis: Durham 2002*. Springer, Berlin, 2003.

[51] E. Hairer and Ch. Lubich. The life-span of backward error analysis for numerical integrators. *Num. Math.*, 76:441–462, 1997.

[52] E. Hairer and Ch. Lubich. Long-time energy conservation of numerical methods for oscillatory differential equations. *SIAM J. Numer. Anal.*, 38(2):414–441, 2000.

[53] E. Hairer and Ch. Lubich. Symmetric multi-step methods over long times. *Num. Math.*, 97(4):699–723, 2004.

[54] E. Hairer, Ch. Lubich, and G. Wanner. Geometric numerical integration illustrated by the Störmer/Verlet method. *Acta Numer.*, 12:399–450, 2003.

[55] E. Hairer, Ch. Lubich, and G. Wanner. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2nd edition, 2006.

[56] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2nd edition, 1993.

[57] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: stiff and differential-algebraic problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 2nd edition, 1996.

[58] W. R. Hamilton. On the application to dynamics of a general mathematical method previously applied to optics. In *British Association Report 1834*, pages 513–518. 1835.

[59] A. C. Hansen. A theoretical framework for backward error analysis on manifolds. preprint available at `http://www.acm.caltech.edu/~hansen/FOCM_paper.pdf`, 2008.

[60] H. Hanßmann. *Local and Semi-Local Bifurcations in Hamiltonian Dynamical Systems*. Number 1893 in Lecture Notes in Mathematics. Springer, Berlin, 2007.

[61] R. H. Hardin and F. D. Tappert. Application of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations. *SIAM Review*, 15:423, 1973.

[62] P. Hartman. *Ordinary Differential Equations*. Number 38 in Classics in Applied Mathematics. SIAM, Philadelphia, 2nd edition, 2002.

[63] K. Heun. Neue Methoden zur approximativen Integration der Differentialgleichungen einer unabhängigen Veränderlichen. *Z. Math. Phys.*, 45:23–38, 1900.

[64] N. J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM, Philadelphia, 2nd edition, 2002.

[65] M. Hochbruck and Ch. Lubich. On Krylov subspace approximations to the matrix exponential operator. *SIAM J. Numer. Anal.*, 34(5):1912–1925, 1997.

[66] M. Hochbruck and Ch. Lubich. A Gautschi-type method for oscillatory second order differential equations. *Num. Math.*, 83(3):403–426, 1999.

[67] E. Hofer. A partially implicit method for large stiff systems of ODEs with only few equations introducing small time-constants. *SIAM J. Numer. Anal.*, 13(5):645–663, 1976.

[68] J. L. Hong, Y. Liu, and G. Sun. The multi-symplecticity of partitioned Runge-Kutta methods for Hamiltonian PDEs. *Math. Comp.*, 75:167–181, 2005.

[69] R. A. Johnson and G. R. Sell. Smoothness of spectral subbundles and reducibility of quasi-periodic linear differential equations. *J. Differential Equations*, 41:262–288, 1981.

[70] À. Jorba. A methodology for the numerical computation of normal forms, centre manifolds, and first integrals of Hamiltonian systems. *Exp. Math*, 8(2):155–195, 1999.

[71] À. Jorba and C. Simó. On the reducibility of linear differential equations with quasi-periodic coefficients. *J. Differential Equations*, 98(1):111–124, 1992.

[72] À. Jorba and C. Simó. On quasi-periodic perturbations of elliptic equilibrium points. *SIAM J. Math. Anal.*, 27(6):1704–1737, 1996.

[73] À. Jorba and J. Villanueva. On the persistence of lower dimensional invariant tori under quasi-periodic perturbations. *J. Nonlinear Sci.*, 7(5):427–473, 1997.

[74] F. Kang. Difference schemes for Hamiltonian formalism and symplectic geometry. *J. Comput. Math.*, 4, 1986.

[75] U. Kirchgraber. Multi-step methods are essentially one-step methods. *Num. Math.*, 48(1):85–90, 1986.

[76] A. N. Kolmogorov. On conservation of conditionally periodic motions under small perturbations of the Hamiltonian. *Dokl. Akad. Nauk. SSSR, (cited in [55])*, 98:527–530, 1954.

[77] S. Kuksin and J. Pöschel. On the inclusion of analytic symplectic maps in analytic Hamiltonian flows and its applications. In S. Kuksin, V. Lazutkin, and J. Pöschel, editors, *Seminar on Dynamical Systems (St Petersberg 1991)*, pages 96–116. Birkhäuser, Basel, 1994.

[78] W. Kutta. Beitrag zur näherungsweisen Integration totaler Differentialgleichungen. *Z. Math. Phys.*, 46(435–453), 1901.

[79] L. D. Landau and E. M. Lifschitz. *Mechanics*, volume 1. Butterworth-Heinemann, 3rd edition, 1976.

[80] F. M. Lasagni. Canonical Runge-Kutta methods. *Z. Angew. Math. Phys.*, 39(6):952–953, 1988.

[81] C. Le Bris and F. Legoll. Dérivation de schémas numériques symplectiques pour des systèms Hamiltoniens hautment oscillants. *C. R. Acad. Sci. Paris, Ser I*, 344:277–282, 2007.

[82] C. Le Bris and F. Legoll. Integrators for highly oscillatory Hamiltonian systems: an homogenization approach. Technical Report 6252, INRIA, Rocquencourt, France, July 2007.

[83] J. Liouville. Note à l'occasion du mémoire précédent (de M. E. Bour). *J. Math. Pures et Appliquées, cited in [55]*, 20:201–202, 1855.

[84] J. E. Marsden and T. S. Ratiu. *Introduction to Mechanics and Symmetry.* Springer-Verlag, New York, 2nd edition, 1999.

[85] R. I. McLachlan and D. R. J. O'Neale. Geometric integration for a two-spin system. Reports in Mathematics 14, Massey University, Institute of Fundamental Sciences, Dec. 2005.

[86] R. I. McLachlan and D. R. J. O'Neale. Geometric integration for a two-spin system. *J. Phys. A*, 39:L447–L452, 2006.

[87] R. I. McLachlan and D. R. J. O'Neale. Comparison of integrators for the Fermi-Pasta-Ulam problem. Technical Report NI07052-HOP, Isaac Newton Institute for Mathematical Sciences, 2007.

[88] R. I. McLachlan and D. R. J. O'Neale. Preservation and destruction of periodic orbits by symplectic integrators. *Numer. Algorithms*, accepted, 2009.

[89] R. I. McLachlan and G. R. W. Quispel. Six lectures on geometric integration. In R. DeVore, A. Iserles, and E. Süli, editors, *Found. Comput. Math.*, pages 155–210. Cambridge University Press, Cambridge, 2001.

[90] R. I. McLachlan and G. R. W. Quispel. Splitting methods. *Acta Numer.*, 11:341–434, 2002.

[91] R. I. McLachlan and G. R. W. Quispel. Geometric integrators for ODEs. *J. Phys. A*, 39(19):5251–5285, 2006.

[92] R. I. McLachlan, G. R. W. Quispel, and P. S. P. Tse. Linearization-preserving self-adjoint and symplectic integrators. *BIT Numerical Mathematics*, 49:177–197, 2009.

[93] R. H. Merson. An operational method for the study of integration processes. *Proc. Symp. Data Processing*, pages 1–25, 1957.

[94] W. E. Milne. Numerical integration of ordinary differential equations. *Amer. Math. Monthly*, 33:455–460, 1926.

[95] W. E. Milne. A note on the numerical integration of ordinary differential equations. *J. Research Nat. Bur. Standards*, 43:537–542, 1949.

[96] P. C. Moan. On the KAM and Nekhoroschev theorems for symplectic integrators and implications for error growth. *Nonlinearity*, 17:67–83, 2004.

[97] P. C. Moan. On rigorous modified equations for discretizations of ODEs, preprint. available at `http://www.focm.net/gi/gips/2005/3.html`, 2005.

[98] P. C. Moan. On modified equations for discretizations of ODEs. *J. Phys. A*, 39:5545–5561, 2006.

[99] B. E. Moore and S. Reich. Multi-symplectic integration methods for Hamiltonian PDEs. *Future Generation Comput. Syst.*, 19:395–402, 2003.

[100] J. Moser. On invariant curves of area preserving mappings of an annulus, cited in [55]. *Nachr. Akad. Wiss. Gottingen, II Math.-Phys. Kl.*, pages 1–20, 1962.

[101] J. Moser and J. Pöschel. An extension of a result by Dinaburg and Sinai on quasiperiodic potentials. *Comment. Math. Helv.*, 59(1):39–85, 1984.

[102] F. R. Moulton. New methods in exterior ballistics. Technical report, University of Chicago, 1926.

[103] A. I. Neishtadt. Estimates in the Kolmogorov theorem on conservation of conditionally periodic motions. *J. Appl. Math. Mech.*, 45:766–772, 1982.

[104] N. N. Nekhoroschev. An exponential estimate of the time of stability of nearly integrable Hamiltonian systems II. In O. A. Olenik, editor, *Topics in Modern Mathematics, Petrovskii Seminar No. 5, cited in [60]*, pages 1–58. Consultants Bureau, 1985.

[105] N. N. Nekhoroshev. An exponential estimate of the time of stability of nearly-integrable Hamiltonian systems, cited in [60]. *Russian Math. Surveys*, 32:1–65, 1977.

[106] F. Neri. Lie algebras and canonical integration. Technical report, Dept. Phys., University of Maryland, 1988.

[107] I. Newton. *Philosophiae Naturalis Principia Mathematica, Londini anno MDCLXXXII.* Facsimile retrieved from `http://eebo.chadwyck.com`, London, 1687.

[108] E. J. Nyström. Über die numerische Integration von Differentialgleichungen. *Acta Soc. Sci. Fennicae*, 50:1–54, 1925.

[109] D. R. J. O'Neale and R. I. McLachlan. Reconsidering trigonometric integrators. *ANZIAM J.*, 50:320–332, 2009.

[110] J. Palis. Vector fields generate few diffeomorphisms. *Bull. Amer. Math. Soc.*, 80(3):503–505, 1974.

[111] A. D. Perry and S. Wiggins. KAM tori are very sticky: rigorous lower bounds on the time to move away from an invariant Lagrangian torus with linear flow. *Physica D*, 71(1–2):102–121, 1994.

[112] H. Poincaré. *Les Méthodes Nouvelles de la Mécanique Céleste*. NASA English translation of the Dover publication, 1957.

[113] J. Pöschel. On elliptic lower dimensional tori in Hamiltonian systems. *Math. Z.*, 202(559–608), 1989.

[114] A. V. Pronin and D. V. Treschev. On the inclusion of analytic symplectic maps into analytic flows. *Regular & Chaotic Dynamics*, 2(2):14–24, 1997.

[115] G. D. Quinlan and S. Tremaine. Symmetric multistep methods for the numerical integration. of planetary orbits. *Astron. J.*, 100(5):1694–1700, 1990.

[116] S. Ree and L. E. Reichl. Classical and quantum chaos in a circular billard with a straight cut. *Phys. Rev. E*, 60(2):1607–1615, 1999.

[117] S. Reich. Backward error analysis for numerical integrators. *SIAM J. Numer. Anal.*, 36(5):1549–1570, 1999.

[118] D. T. Robb and L. E. Reichl. Chaos in a two-spin system with applied magnetic field. *Phys. Rev. E*, 57(2):2458–2459, 1998.

[119] C. Runge. Über die numerische Auflösung von Differentialgleichungen. *Math. Ann.*, 46:167–178, 1895.

[120] H. Rüssmann. *Dynamical Systems, Theory and Applications: Battelle Seattle 1974 Rencontres*, volume 38 of *Lecture Notes in Physics*, chapter On optimal estimates for the solutions of linear partial differential equations of first order with constant coefficients on the torus, pages 598–624. Springer, New York, 1975.

[121] H. Rüssmann. On optimal estimates for the solutions of linear difference equations on the circle. *Celestial Mechanics*, 14:33–37, 1976.

[122] R. D. Ruth. A canonical integration technique. *IEEE Tranactions on Nuclear Science*, 30(4):2669–2671, 1983.

[123] M. Rychlik. Renormalization of cocycles and linear ODEs with almost periodic coefficients. *Invent. Math.*, 110(1):173–206, 1992.

[124] B. N. Ryland. *Multisymplectic Integration.* PhD thesis, Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand, 2007.

[125] B. N. Ryland and R. I. McLachlan. On multisymplecticity of partitioned Runge-Kutta methods. *SIAM J. Sci. Comp.*, 30(3):1318–1340, 2008.

[126] J. M. Sanz-Serna. Runge-Kutta schemes for Hamiltonian systems. *BIT*, 28(4):877–883, 1988.

[127] J. M. Sanz-Serna and M. P. Calvo. *Numerical Hamiltonian Problems*, volume 7 of *Applied Mathematics and Computation.* Chapman and Hall, London, 1994.

[128] L. F. Shampine and M. W. Reichelt. The MATLAB ODE suite. *SIAM J. Sci. Comput.*, 18(1):1–22, 1997.

[129] Z.-J. Shang. A note on the KAM theorem of symplectic mappings. *J. Dynam. Differential Equations*, 12(2):357–383, 2000.

[130] Z.-J. Shang. Resonant and Diophantine step sizes in computing invariant tori of Hamiltonian systems. *Nonlinearity*, 13:299–308, 2000.

[131] R. D. Skeel and K. Srinivas. Nonlinear stability analysis of area-preserving integrators. *SIAM J. Numer. Anal.*, 38(1):129–148, 2000.

[132] S. Smale. Differentiable dynamical systems. *Bull. Amer. Math. Soc.*, 73:747–817, 1967.

[133] R. Steinigeweg and H.-J. Schmidt. Symplectic integrators for classical spin systems. *Comput. Phys. Comm.*, 174(11):853–861, 2006.

[134] R. Steinigeweg and H.-J. Schmidt. Heisenberg-integrable spin systems. *Math. Phys. Anal. Geom.*, 12(1):19–45, 2009.

[135] G. Strang. On the construction and comparison of difference schemes. *SIAM J. Num. Anal.*, 5(3):506–517, 1968.

[136] A. M. Stuart. Numerical analysis of dynamical systems. *Acta Numer.*, pages 467–572, 1994.

[137] A. M. Stuart and A. R. Humphries. *Dynamical Systems and Numerical Analysis*, volume 2 of *Cambridge Monographs on Applied and Computational Mathematics.* Cambridge University Press, Cambridge, 1996.

[138] M. Suzuki. Fractal decomposition of exponential operators with applications to many-body theories and Monte Carlo simulations. *Phys. Lett. A*, 146(6):319–321, 1990.

[139] M. Suzuki. General theory of higher-order decomposition of exponential operators and symplectic integrators. *Phys. Lett. A*, 165(5,6):387–395, 1992.

[140] Y.-F. Tang. The symplecticity of multi-step methods. *Computers Math. Applic.*, 25(3):83–90, 1993.

[141] F. D. Tappert. Numerical solutions of the Korteweg-de Vries equation and its generalizations by the split-step Fourier method. In A. C. Newell, editor, *Nonlinear Wave Motion*, volume 15 of *Lectures in Applied Mathematics*, pages 215–216. AMS, 1974.

[142] H. F. Trotter. On the products of semi-groups of operators. *Proc. Am. Math. Soc.*, 10(4):545–551, 1959.

[143] M. Tuckerman, B. J. Berne, and G. J. Martyna. Reversible multiple timescale molecular dynamics. *J. Chem. Phys.*, 97(3):1990–2001, 1992.

[144] L. Verlet. Computer "experiments" on classical fluids, I: Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, 1967.

[145] L. Verlet. Computer "experiments" on classical fluids, II. Equilibrium correlation functions. *Phys. Rev.*, 165(1):201–214, 1968.

[146] J. A. C. Weideman and B. M. Herbst. Split-step methods for the solution of the nonlinear Schrödinger equation. *SIAM J. Numer. Anal.*, 23(3):485–507, 1986.

[147] H. Yoshida. Construction of higher order symplectic integrators. *Phys. Lett. A*, 150:262–268, 1990.

[148] H. Yoshida. Recent progress in the theory and application of symplectic integrators. *Celestial Mech. Dynam. Astron.*, 56:27–43, 1993.