Geometric integration of differential equations

Piggott, Matthew David

Geometric integration of differential equations

submitted by

Matthew David Piggott

for the degree of PhD

of the

University of Bath

2002

Signature of Author ....

Matthew David Piggott

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

This work focuses on the geometric integration of ordinary and partial differential equations. That is, the design, analysis and testing of numerical methods designed to capture qualitative geometric properties that may be present in a problem. Of primary importance here are the properties of scaling invariance and Hamiltonian structure.

In order to capture the property of scaling invariance in numerical methods, much use is made throughout of both spatial and temporal adaptivity. The adaptivity in the methods considered here is achieved through coordinate transformations between adapted physical variables and fixed computational variables.

Self-similar solutions are known to be of importance in scaling invariant problems. As well as being exact solutions for which it is often possible to find closed form expressions, they often also represent the singular or asymptotic behaviour of more general solutions. In this thesis it is proved that the scaling invariant numerical methods, constructed through the appropriate use of adaptivity, admit discrete self-similar numerical solutions. In the case of ordinary differential equations these are rigorously shown to uniformly approximate the true self-similar solutions for all time, and also to inherit their stability. As a result examples are given where the numerical methods capture the correct (asymptotic or singular) behaviour of the problem, for example the singular nature of gravitational collapse in the two-body problem is shown to be captured by the constructed methods. In the case of partial differential equations analogous results are demonstrated for the porous medium equation, where a maximum principle is also established and used to prove convergence of the method to self-similarity.

The field of numerical (in particular symplectic) methods for Hamiltonian problems has received much attention in the literature, these methods and problems are also considered at points throughout this thesis. In particular, due to the close relationship between symmetries (e.g. scaling invariance), Hamiltonian structures and conservation laws, an examination is made of the possibility of constructing methods which respect both the scaling and Hamiltonian nature for problems which possess both. Again this is achieved with standard methods, but following the correct type of coordinate transformation. Conditions are derived upon this transformation which guarantee that corresponding conservation laws are not destroyed by the discretization.

Finally much space is reserved for the consideration of the semi-geostrophic problem, which is an important equation set in meteorology. This problem has much complementary geometric structure, a lot of which has many similarities with the properties and methods considered here, including a natural coordinate transformation and a Hamiltonian structure. The final part of this thesis therefore looks at how geometric integration may be applied to this complicated and very interesting problem. The work on adaptivity in the thesis is linked to the semi-geostrophic coordinate transformation and a mesh for a particular model problem is constructed. The Hamiltonian problem is considered using semi-Lagrangian methods, and a useful reformulation of the noncanonical problem in terms of canonical Clebsch variables is given.

Acknowledgements

I would like to use this opportunity to thank my supervisor Chris Budd for his excellent guidance, advice and encouragement throughout the last three years, and for giving me the opportunity to work on such an interesting project.

For academic help and advice over the years I would also like to thank Mike Baines, Sergio Blanes, Mike Cullen, Philip Drazin, Ivan Graham, Arieh Iserles, Ben Leimkuhler, Bill Morton, Sebastian Reich, Bob Russell, Alastair Spence, and JF Williams.

I am grateful for the financial support of the EPSRC under the computational partial differential equations grant GR/M30975.

I would also like to thank my parents for their support, and for enabling me to attend University in the first place.

For company during many many hours of breaks and cups of coffee I would like to thank Bob, Sarah and Marc.

Finally I would like to say a big thank you to Zoka for her support and for putting up with me whilst I was trying to get the work for this thesis finished.

Contents

List of Figures

Chapter 1

Introduction

This thesis is concerned with the design, analysis and application of adaptive numerical methods constructed to integrate differential equations whilst preserving certain geometric properties of the underlying problems.

1.1 Overview

The modern study of natural phenomena described by both ordinary and partial differential equations usually requires a significant application of computational effort. The majority of methods and algorithms employed in this effort are generally based upon the standard technique of performing a stable discretization of the problem in such a way as to keep local truncation errors as small as possible. For many problems this is accompanied by the use of adaptive methods. Through many varied techniques these attempt to adjust the (spatial and temporal) meshes so as to constrain the local truncation errors not to exceed specified tolerances. When combined these algorithms and techniques lead to methods with the ability of being able to compute very accurate solutions to fairly general classes of differential equations. In general this is provided that the times for integration are not long and the solutions remain reasonably well behaved. Numerical analysis (and in particular, the numerical analysis of methods for differential equations) is the field of mathematics which is used to design and analyse these methods [100, 108, 137, 62].

The topic of this thesis is *geometric integration* — a relatively new field of numerical analysis. In general the methods mentioned above which are based primarily on the analysis of local truncation errors do not necessarily respect, or even take into account, the qualitative and global features of a problem or equation. This possible shortcoming of a method can lead to unreliable results, for example the computation of spurious solutions, or solutions which have the wrong qualitative properties and are therefore physically impossible and often completely useless. Many equations used to describe physical phenomena have geometric and qualitative features in common with the physics of the underlying problem being modelled. For example, Newton's second law ($m\ddot{x} = F$) is not just simply a statement relating an acceleration experienced by a body to an exerted force, it also tells us about all the physical laws relevant to the particular situation [121]. It can be argued in many situations that the qualitative structures present actually tell us more about the underlying problem than the local information given by the expression of the problem in terms of differentials. This observation motivates the study of numerical methods which, although possibly having larger local truncation errors and costs (however in some situations geometric integration methods actually turn out to be cheaper), attempt to systematically incorporate some of the qualitative information of the underlying problem into their structure. Geometric integration is the name given to the design and rigorous analysis of such methods. These are sometimes existing methods where geometric integration has led to new insight into their behaviour and performance (for example the Gauss-Legendre Runge-Kutta methods [108, 146], the Störmer-Verlet-leapfrog method [173, 155] popular in molecular dynamics, and the Newmark algorithm [185, 104] popular in structural mechanics), or entirely new methods where special techniques are employed to incorporate qualitative structure into the algorithm. Since the geometric properties of a problem are generally so fundamental and natural, the geometric methods may be very simple and fast, and the results guaranteed to be qualitatively correct, or even in some situations quantitatively more accurate than other methods. These methods often turn out to be more efficient than other schemes for certain problems and certain applications, they are also often easier to analyse, this is ultimately because we may exploit the qualitative theory of the underlying differential equations. Further details of geometric integration can be found in the recent reviews and discussions listed in the following references [32, 121, 84, 101, 36, 145, 146].

One of the aims of this thesis is to investigate how the geometric integration approach may be beneficial for solving both ordinary and partial differential equations with a scaling invariance property. In addition, one of the ongoing unsolved problems in geometric integration is if, how, and what the benefits are, of trying to preserve more than one qualitative property of a problem in a numerical method. For reasons that shall become apparent in due course, a natural combination of qualitative properties to consider is that of a Hamiltonian as well as a scaling (or more general symmetry) invariance structure. Techniques for incorporating both of these properties into a numerical method are also considered here. Finally, a large part of the current literature on geometric integration still focuses on model problems and not necessarily serious applications (although there are noticeable exceptions to this comment, for example the problem of stellar and molecular dynamics [166, 155, 109], an ODE application with more details later). The final aim of this thesis is therefore to investigate how some of the geometric integration techniques may be applied to a particular problem of interest to geophysicists and meteorologists, namely a PDE system called the semi-geostrophic equations which are a simplification of equations used to describe the large scale motion of the atmosphere and oceans. The ultimate hope being for a two-way exchange of ideas where, in addition to possibly designing methods which capture qualitative properties of the problem, problem specific procedures used by the meteorologists may be generalized to improve the techniques available in geometric integration and maybe also in numerical analysis in general.

1.2 Qualitative and geometric properties

In this section we shall take a brief look at some of the qualitative and geometric features of a system described by a differential equation which we have in mind when we talk about geometric integration.

There are many possible qualitative features which may be present in problems modelled by systems of ordinary or partial differential equations. An attempt is not made to give a complete listing here. However, below a partial listing is presented which covers a wide variety of possibilities, focusing attention on those properties that shall be encountered at various points throughout this thesis. In addition, these sometimes wildly different properties may be linked to one another in beautiful and deep mathematical ways, an attempt is made to mention at least some of these below.

1. *Geometrical structure.* The phase space in which a problem is defined may have deep mathematical properties which give enormous insight into the overall properties of its solutions. For example problems with fixed points or invariant sets, problems with a conservative or Hamiltonian structure (see Chapter 2), dissipative problems or problems possessing a Lyapunov function. For a discussion of these along with other geometric structures see [7, 164, 121, 120, 82, 138].

2. *Conservation laws.* Underlying many systems are conservation laws. These may include the conservation of total quantities such as mass, momentum and energy, or instead quantities which are conserved along particle trajectories and flows such as fluid density or potential vorticity, see Chapter 6. The calculation of solutions which do not respect the particular conservation laws present in a problem can lead to physically meaningless behaviour. For example the loss of energy in a system describing planetary motion will inevitably lead to the planet being modelled spiralling into the sun, which is clearly incorrect qualitatively, we shall see this in Chapter 2. Similarly it is widely accepted [53] that in many systems used to model the large scale behaviour of the oceans and atmosphere it is essential to conserve potential vorticity in order to retain the overall qualitative dynamics

of the solution. In addition, there are also more abstract quantities conserved by certain systems. For example phase space volume preservation in divergence free systems and the conservation of a symplectic structure in Hamiltonian systems. As is common with many such properties, there are deep relationships between them. For example the value of the Hamiltonian following the solution is conserved in autonomous Hamiltonian systems. Casimirs (a specific type of conservation law) and other functions may also be conserved along trajectories of certain Hamiltonian systems, for further details see Chapter 2 as well as [7, 128, 124, 125, 154, 181].

3. *Symmetries.* Many systems are invariant under the actions of symmetries such as Lie group, scaling and involution symmetries. Such symmetries may or may not be retained in the underlying solution of the system, but as is discussed in Chapter 2 there may exist important solutions which do not change when the symmetry group acts. The possible symmetries may include the following:

   - *Galilean symmetries.* These basically describe the invariance of a problem to a change of frame of reference. This includes for example space and time translations, rotations and boosts (moving frames of reference). They are important for example in travelling wave solutions to problems, in computer vision and in the motion of rigid bodies, see [114] for more details.

   - *Time reversal and involution symmetries.* The solar system is an example of a system which is invariant under a reversal of the time variable. That is given a solution describing an evolution of the system, the evolution obtained by playing back the motion in reverse is equally well a solution to the equations describing the system. More generally many physical systems are invariant under involution symmetries ρ satisfying the identity $\rho^2 = Id$, for a review and discussion of this type of structure see [107]. More details shall also be given in Chapter 4.

   - *Scaling symmetries.* Many physical problems have the property that they are invariant under rescalings in either time or space. This partly reflects the fact that the laws of physics should not depend upon the units in which they are measured or indeed should not have an *intrinsic* length scale [15]. An example of such a scaling law is Newton's law of gravitation which is invariant under a rescaling in time and space. This underlying property of a system shall be encountered at numerous points throughout this thesis and many more details and references may be found there.

   - *Lie group symmetries.* These are deeper symmetries (which generalize the scaling and Galilean symmetries mentioned above), often involving the invariance of a system to a (nonlinear) Lie group of transformations. See the start of Chapter 2 for many further details.

The presence of symmetries in a system may have a profound affect on the behaviour of solutions to that system. For example it may impose the constraints of conservation laws and particular asymptotic behaviour on solution structures, see later Chapters as well as [15, 4]. It may also govern and control interesting local solution behaviour, for example singularities, shocks and fronts, and boundary layers or interfaces. Again see later Chapters as well as [142, 177, 64, 21]. Finally, symmetries may also control and affect the types and multiplicity of solutions that bifurcate from steady states [78].

4. *Asymptotic behaviour.* Many problems have the property that they evolve in time so that asymptotically their dynamics in some sense simplifies. For example they may ultimately evolve so that the dynamics is restricted to a lower dimensional (possibly chaotic) attractor, or complex structures starting from arbitrary initial data may simplify into regular patterns [168, 164, 81]. Alternatively, the problem may have solutions which form singularities in finite time such as weather fronts (see Chapters 6 and 7), or combustion in which the solution itself becomes singular at a point (or along a line in two spatial dimensions etc.), see Chapter 3 and [142, 31]. All of these features can be incorporated into the design of a numerical scheme and should be reproduced by a good numerical method.

5. *Orderings in the solutions.* Many differential equations possess some form of maximum or comparison principle, this can lead to a preservation of the ordering between different solutions. For example, given two sets of initial data $u_0(x)$ and $v_0(x)$ for a partial differential equation, the solutions may respect the ordering that if $u_0(x) < v_0(x)$ for all x, then $u(x,t) < v(x,t)$ for all x and t. The linear heat equation $u_t = u_{xx}$ has this property as do many other parabolic problems, see [134]. The ordering of solutions and maximum principles can also have important and interesting combinations with the symmetries present in a problem. We use this to great effect in Chapter 5, see also [68, 61].

It is an important point to realize that many of these geometric properties may be closely linked to one another. For example, if the differential equation is derived from a variational principle linked to a Lagrangian function then, via Noether's theorem, each continuous symmetry of the Lagrangian leads directly to a conservation law for the underlying equation, see [128] and Chapter 2 for more precise details. This result may be generalized to apply to discretizations of problems with symmetries. If a numerical method is also based upon a (discrete) Lagrangian and this Lagrangian has symmetries then the numerical method automatically has a discrete conservation law associated with this symmetry, see [58], also see [116] for similar results in terms of momentum maps [114]. These and other points shall be expanded upon in later Chapters.

A combination of many properties from the above list are used in Chapter 5. Specifically, when symmetry is coupled with solution orderings this frequently leads to an understanding of the asymptotic behaviour of a problem. In particular, self-similar solutions (which are invariant under the action of a scaling group, see Chapter 2) can be used to bound the actual solution from above and below. The solution behaviour is then constrained to follow that of the self-similar solution. We shall consider this in more detail in Chapter 5 where we use a discrete form of this idea to prove convergence of our numerical method.

1.3 The motivation for preserving geometric features in algorithms

There are several reasons why it may be worthwhile to preserve qualitative structure in algorithms. Firstly, many of the properties of the previous section can be found in systems which occur naturally in applications. For example, large scale molecular or stellar dynamics can be described by Hamiltonian systems with many conservation laws. Mechanical systems evolve under rotational constraints, as do many of the problems of fluid mechanics. Partial differential equations possessing scaling symmetries and self-similarity arise in fluid and gas dynamics, combustion, nonlinear diffusion and mathematical biology. Partial differential equations with a Hamiltonian structure are important in the study of solitons, in the Korteweg-de Vries equation for example. As we shall see in Chapter 6, the semi-geostrophic equations of meteorology also have a Hamiltonian structure as well as whole families of conservation laws. They also possess a natural coordinate transformation which itself can be shown to have various properties, for example a Legendre transform structure, see Chapter 6 for many additional details and references.

In designing our numerical method to preserve certain geometrical properties we ultimately hope to see some kind of improvement in our computations. For a start we will ultimately end up with a discrete dynamical system which has many properties in common with the continuous one (see [164]), and thus can be thought of as being in some sense close to the underlying problem in that stability, orbits, long-time behaviour and other structural properties may be common to both systems. The technique of backward error analysis [80, 136, 85, 176] is often used to prove these results. Geometric structures often (for example, using backward error analysis and exploiting the geometric structure of the discretizations) make it easier to estimate errors, and in fact local and in particular global errors may well be smaller for no extra computational expense. Geometric integration methods designed to capture specific qualitative properties may also preserve or nearly preserve additional properties of the solution *for free.* For example symplectic methods for Hamiltonian problems have excellent energy conservation properties and can conserve angular momentum or other invariants (which may not even be known in advance), see Chapter 2 and [146].

The original philosophy behind the development of numerical methods for calculating solutions to differential equations centred on the study of minimizing errors over a fixed finite time T. However, in many situations we may be interested in numerically studying the dynamics of a system, possibly to gain insight into its long term or asymptotic behaviour. We are therefore often interested in the alternative philosophy of taking a method and applying it for an undefined number of time steps. Note that in a problem with widely varying time scales (such as molecular dynamics [2] where the time taken for atom interactions is vastly smaller than the time we are interested in integrating for, and also in PDEs and their discretizations [153]) the study of long term behaviour is unavoidable. For such studies we are not necessarily always interested in computing accurate individual solution trajectories (due to the possibly chaotic behaviour of solutions for example), but rather we desire statistically correct solution behaviour in our numerics. We therefore need to focus on the ability of a numerical method to preserve the structural properties of a system. A very thorough review of some of these issues and the way that numerical methods are used to study, and thought of in terms of, dynamical systems (and vice-versa) is given in [164]. An excellent example of an application of some of these points is in the long term study of the solar system, see for example [166] where specially designed methods are used to investigate whether or not the solar system exhibits chaotic behaviour. As the errors of many *conventional* methods accumulate at least quadratically or even exponentially with time, accurate qualitative investigations over long time using these methods is not possible, even if the methods are of high order and have very small local errors. Thus it is essential to use methods for which there is some control over the long term growth or accumulation of errors, even if the local error made by such methods may appear to be very large in comparison to others.

In conclusion, the motivation for considering geometric integration when designing numerical methods is that for difficult problems the methods so constructed can often *go where other methods cannot*, or at least have enormous difficulties. These methods have had many successes, for example in the accurate computation of singularities [30, 31], in molecular dynamics [109, 155, 148], and in the long term integration of the solar system [179, 166]. The list of possible application areas keeps on growing, in particular in this thesis an attempt is made in the direction of applying the geometric integration ideas to geophysical problems.

At this point it is worth noting that although the term geometric integration has only really been used during the last decade, people have been noting and considering the benefits of preserving geometric structures in algorithms much before this. For example, methods for Hamiltonian problems have been studied in great detail since the mid-Eighties (see [65, 140] for early references, see also [146] and the references therein), and even earlier [174]. Also, the Arakawa scheme [6] was directly designed to preserve conservation laws specifically because of the excellent stability properties this implied.

Although the majority of what has been said above expresses the positive view of geometric integration, it should also be noted that there is a negative side. For a start the numerical analyst needs to have a working knowledge of a wide variety of geometric theory before deciding what properties his or her system possesses and therefore which method and technique to employ. For the casual user much of this process may be automated using symbolic packages, but this is maybe not an ideal solution. In a similar vein some properties of a system maybe so *hidden* or subtle that they remain unknown even to experts, other properties may be impossible to know until after a solution has already been found. This shall be mentioned again in later Chapters, we shall also see a partial solution to this problem where methods designed to preserve one feature inherit others (which don't even need to be known or considered a priori) for free. Another point, already mentioned above, is the problem of having several qualitative properties present in a problem and deciding if it is possible or worth attempting to preserve them all in a method. Of course in many situations the properties fit together very nicely, but if it is not possible to preserve them all the problem of which properties are more important in the sense of leading to 'better' computations needs to be considered. Finally, perhaps the most obvious downside is that the extra effort and expense incurred by preserving qualitative properties in a method may not be outweighed by a sufficient improvement in performance over standard methods. Having said that, geometric integration is a subject still in its relative infancy and all of these issues need to be addressed in the future. Hopefully the positive points are sufficiently numerous and persuasive to warrant much further work in this subject area, and to show that the geometric integration based methods should at least be considered whenever one encounters and considers a new problem.

1.4 The main results and structure of the thesis

This thesis is organized as follows. In Chapter 2 some notation is introduced and a fairly complete discussion is made of symmetry and the theory underlying the invariance of a differential equation to a Lie group of transformations. The benefits of considering such a structure are mentioned and a special group invariant type of solution and its nice properties are introduced. This topic is applicable equally well to both ordinary and partial differential equations. Some simple examples are given, and in Appendix A a fairly detailed example of the standard but possibly complex process of deriving a group of invariant transformations is presented. Special attention is paid throughout to scaling transformations. This is followed by a review of the Hamiltonian structure which may be present in both ordinary and partial differential equations. Symplecticity and other properties of Hamiltonian problems are considered. Symplectic numerical methods constitute perhaps the most studied of the geometric integration techniques and some quick examples of their construction, properties and behaviour are given. The superior performance of these methods over similar, but non-symplectic, methods in simple experiments is noted and used as further motivation for the continued study of methods of this type.

In Chapter 3 the very natural way in which the scaling invariance property introduced in Chapter 2 and temporal adaptivity fit together is explained. It is demonstrated how adaptivity may be achieved through the use of a (Sundman) coordinate transformation of the time variable, and how this may be used to construct methods which mimic the scaling invariance structure of the problem. This is shown to lead to the methods possessing discrete self-similar solutions analogous to the continuous self-similar solutions of the problem. Some rigorous results are proved which lead to the fact that the discrete self-similar solution approximates the continuous self-similar solution with a relative error which is uniform in time, and that the stability properties of the self-similar solution are also inherited by the method. Some conclusions are made, in particular attractive results regarding the correct asymptotic behaviour and accurate computation of difficult solutions. The Chapter concludes with several numerical examples demonstrating in action the results proved in earlier sections. The examples include the Kepler problem where a discrete analogue of Kepler's third law is shown to follow from the preservation by the method of the scaling invariance property of the problem.

In Chapter 4 the application of geometric methods to problems with both a Hamiltonian structure and the property of being invariant to a symmetry group are considered. This very natural combination was also discussed in Chapter 2 where some useful standard results were reviewed. This motivates the search for a technique which enables the construction of methods which are both symplectic and invariant under a scaling transformation. The time transformation technique employed in Chapter 3 is shown in general to destroy the Hamiltonian structure of a problem. The Poincaré transformation is introduced as a generalization of the Sundman transformation method for performing temporal adaptivity. It is shown that this method allows the construction of methods which are both symplectic and scaling invariant. Further conditions on the coordinate transformation are given under which it is proved that conservation laws may also be preserved by the constructed methods. The Kepler problem is considered in detail and a comparison is made between the various methods constructed in the thesis to this point. This Chapter is presented as a partial investigation into the open problem of how different qualitative properties may be incorporated into algorithms.

In Chapter 5 partial differential equations invariant under scaling transformations are considered, motivated by, and as a natural extension to, the work of earlier Chapters. The approach used in Chapter 3 to perform the temporal adaptivity is generalized to the spatial dimension. The principle of equidistribution is introduced and used to define a spatial coordinate transformation, from which scaling invariance and adaptivity are again shown to fit naturally together. The properties and usefulness of self-similar solutions to many PDE problems is explained and the possible benefits of retaining them in numerical methods is discussed. A large proportion of the Chapter is taken up with a detailed examination of how these ideas apply to a fairly complex model problem (the porous medium equation). Scale invariant methods are constructed which have some remarkable properties. The scale invariance property is shown to yield a method which not only possesses a semi-discrete self-similar solution, but which also respects the underlying conservation laws and correct interface behaviour of the problem. In addition the method is shown to retain a comparison principle present in the PDE, this incredibly useful property is used to prove that the semi-discrete self-similar solution acts as an attractor for more general numerical solutions. The method therefore mimics exactly the behaviour of the continuous problem and some experiments are used to demonstrate this. A brief discussion of how these ideas may be generalized from finite difference methods to finite elements is also given.

In Chapter 6 the semi-geostrophic equations of meteorology are introduced and a detailed discussion of their structure is given. This system is used as an example to conclude the thesis with because it is a problem where possible geometric integration applications are not obvious at first sight, and also because efficient numerical methods for computing difficult solutions to this problem are of interest to people in industry (specifically in numerical weather prediction, oceanography and meteorology). It is a problem possessing many different geometrical properties, many of these are mentioned in this Chapter but special attention is paid to both a canonical and a noncanonical Hamiltonian structure underlying the equations (corresponding respectively to a Lagrangian and an Eulerian viewpoint of the problem), and a coordinate transformation which is often used to simplify the analysis of the problem. The coordinate transformation is shown to allow the problem to be rewritten in terms of a nonlinear elliptic PDE governing the transformation and an advection equation for a variable representing the potential vorticity of the system.

Chapter 7 considers in more detail the coordinate transformation introduced in Chapter 6. Some links between this so-called semi-geostrophic coordinate transformation and various techniques for performing spatial adaptivity, in particular higher dimensional versions of those used in Chapter 5, are demonstrated. The numerically challenging problem of integrating the equations through the formation of an idealized weather front (a discontinuity in the coordinate transformation) is also discussed. The accurate and efficient computation of solutions of this type which exhibit rapid local variations generally requires the use of some form of mesh adaptivity and the links with the semi-geostrophic coordinate transformation are used to motivate a particular strategy for performing this adaptivity. This is demonstrated through an example focusing on a parabolic umbilic model problem which is taken from the current literature, for completeness additional details behind this example are given in Appendix B. The exchange of ideas is completed with a study of a possible way in which the analytic work behind the semi-geostrophic coordinate transformation may be used to motivate a new type of mesh generation and adaptivity technique.

In Chapter 8 the derived advection equation for potential vorticity and the Hamiltonian structure underlying the semi-geostrophic problem are considered in more detail. A note here is made of the impossibility in general of spatially truncating a noncanonical Hamiltonian PDE system to obtain a Hamiltonian ODE system. The Hamiltonian structure behind the semi-geostrophic equations is shown to be very similar to that of the Euler equations and a discussion of how the numerical methods for the latter may be generalized and applied to the former is made. An interesting reformulation in terms of Clebsch variables is given, these inflate a noncanonical system so that it may be written as a canonical system, the motivation being that methods which respect an infinite-dimensional canonical Hamiltonian structure may be found and these are briefly mentioned. Due to their popularity with meteorologists and geophysicists, semi-Lagrangian methods are analysed and are shown, with the correct formulation, to possess some very nice geometric properties. These include firstly the preservation of the non-negativity of potential vorticity which is vital physically, as well computationally for the methods considered in Chapter 7, and secondly a possible way of performing the integration in a 'Hamiltonian way'.

Finally in Chapter 9 some conclusions and possible future research avenues are presented. Although the aims and goals of the thesis as stated above were fairly broad and wide ranging, it is argued that most have been addressed and many partial answers have been found. It is further claimed that several of the open problems of geometric integration have been considered and discussed, and as with many pieces of scientific work the list of unanswered questions is larger there than in this introduction.

1.5 A guide to original material

We shall now give a brief guide to make explicit which results in this work are original, distinguishing them from survey material. Firstly, Chapter 2 contains no original material, it is simply a review of some background theory which shall be useful throughout this Thesis. Appendix A presents an example of an application of this theory in a simple case and is well known material. Similarly Chapter 6 is purely a survey of current knowledge on the semi-geostrophic equations, focusing on those issues that are of importance for the final Chapters. Appendix B contains some extra background results as well as a summary of a specific model problem appearing in the literature.

Chapter 3 contains the first original material in this Thesis. The coordinate transformation employed to achieve adaptivity has been used before, although the motivation in terms of scaling invariance is new. All results then proved are original, including the scaling invariance of both Runge-Kutta and linear multistep methods and the corresponding admittance of discrete self-similar solutions which uniformly approximate the true self-similar solutions and also inherit their stability.

In Chapter 4, again the coordinate transformation employed here has been used before to obtain adaptive methods, although its motivation in terms of preserving *both* Hamiltonian and scaling invariance structures is original. The result establishing conditions under which conservation laws are preserved by the transformation is also original.

In Chapter 5 the starting discussion describing the extensions of the ideas of Chapter 3 to PDEs is a survey of material currently in the literature. The choice of adaptivity applied to the porous medium equation in the second half of the Chapter is new and (apart from a brief review of theory underlying this equation) all results established here are original, including the maximum principle and the convergence of the method to self-similarity.

In Chapter 7 the links between the Monge-Ampére equation and moving mesh theory established and the application to the parabolic umbilic example are all original results. The discussion of the deformation method simply reviews a current technique, although its possible applications from a geometric integration viewpoint are new. Finally the brief comment on a new adaptivity technique is simply a pointer to current and future research.

In Chapter 8 the sine bracket truncation comment is an application of a current technique to a new problem. Similarly the use of semi-Lagrangian methods is standard, although the discussion regarding the Hamiltonian structure in the problem is original material. Finally the Section on Clebsch variables reviews current knowledge, although again the application to the semi-geostrophic equations is original.

Chapter 2

The geometric theory of differential equations and an introduction to geometric integration

2.1 Overview of Chapter

In this Chapter we shall give a brief introduction to the geometric properties of problems that we shall be considering throughout this thesis. In Section 2.2 we shall mention quickly some of the fundamental ideas behind the Lie group invariance of a differential equation (of either ordinary or partial type) before moving on to consider the particular case of invariance under a scaling transformation, where we also introduce the concept of a self-similar solution. We follow this up in Section 2.3 with a discussion of Hamiltonian structures in differential equations. The reasonably well developed field of numerical methods for Hamiltonian ODEs is discussed, we use this opportunity to present some simple examples of methods designed under the geometric integration philosophy, namely methods which inherit the symplectic transformation property of the continuous flow. We perform a few very simple experiments to demonstrate some of the advantages of the geometric approach and use this as further motivation for this work. Finally, Hamiltonian PDEs are considered towards the end of this Chapter. Geometric methods for these problems are far less developed than they are for Hamiltonian ODEs and there are still many difficulties and unanswered questions in this part of the field, these are briefly mentioned. The Korteweg-de Vries equation and the Euler equations (which have many properties in common with the semi-geostrophic equations to be considered in later Chapters) are given as examples where some progress has been made in developing methods which preserve geometric structures.

2.2 Lie group invariance structure

2.2.1 General theory

Additional details on much of the following material may be found in the references [128, 99, 161, 60, 61]. We begin with the simplest case and suppose that we have a single dependent variable u, a function of an independent variable x. The relation between x and u may be governed, for example, by a differential equation of the form

$$\frac{du}{dx} = f(u),$$

where f is a given function. We firstly introduce the concept of a set of *point transformations* which, for our simple case, is a set of transformations of the variable x and u of the form

$$\tilde{x} = \tilde{x}(x, u; \lambda), \quad \tilde{u} = \tilde{u}(x, u; \lambda), \tag{2.1}$$

where λ is an arbitrary parameter. For each particular value of λ (2.1) gives us a mapping between the points (x, u) and $(\tilde{x}, \tilde{u})$, by varying λ we therefore have a set of such transformations. For example, a rotation anti-clockwise of the point (x, u) about $(0, 0)$ an angle λ is given by

$$\tilde{x} = x\cos\lambda - u\sin\lambda, \quad \tilde{u} = x\sin\lambda + u\cos\lambda. \tag{2.2}$$

If we further assume that each member of the set of transformations has an inverse also contained in the set (for (2.2) this is obviously the transformation where λ is replaced by $-\lambda$), the set contains an identity element (for (2.2) this is obviously the transformation given by taking $\lambda = 0$), and also that two transformations carried out in succession are equivalent to another single transformation in the set (for (2.2), given transformations characterized by λ_1 and λ_2, this single equivalent transformation is given by taking $\lambda = \lambda_1 + \lambda_2$). We call a set of transformations of this type a *one-parameter Lie group of point transformations*, (the Lie[1] that appears in this definition actually refers to some additional smoothness conditions required on the set of transformations, we omit the details here as we shall not directly need them, and refer the reader to the references stated above.)

Now suppose that we have a point (x_0, u_0) in the plane and consider what happens to this point as we apply (2.1) to it for λ varying continuously from zero. We obviously map out a path in the plane, for example with (2.2) we will ultimately map out a circle. We can repeat this procedure for different (x_0, u_0), each resulting path we call an *orbit* of the group. We see straight away the analogy with the flow induced by a differential equation for example. In particular each line may be completely characterized by the

[1] After the Norwegian Mathematician Marius Sophus Lie (1842–1899).

field of its tangent vectors. Expanding (2.1) in a Taylor series about an arbitrary point $(x, u; 0)$ we have

$$\tilde{x}(x,u;\lambda) = x + \lambda\xi(x,u) + \mathcal{O}(\lambda^2), \quad \tilde{u}(x,u;\lambda) = u + \lambda\eta(x,u) + \mathcal{O}(\lambda^2),$$

where the so-called *coordinate functions* of the group are given by

$$\xi(x,u) = \left.\frac{\partial\tilde{x}}{\partial\lambda}\right|_{\lambda=0}, \quad \eta(x,u) = \left.\frac{\partial\tilde{u}}{\partial\lambda}\right|_{\lambda=0}. \tag{2.3}$$

We can now write down an operator, which we call the *infinitesimal generator* of our group,

$$\mathbf{X} = \xi(x,u)\partial_x + \eta(x,u)\partial_u.$$

The term generator is used since this object contains sufficient information for us to recover transformation (2.1) exactly, simply through integrating

$$\frac{\partial\tilde{x}}{\partial\lambda} = \xi(\tilde{x},\tilde{u}), \quad \frac{\partial\tilde{u}}{\partial\lambda} = \eta(\tilde{x},\tilde{u}),$$

with the initial conditions $\tilde{x} = x$ and $\tilde{u} = u$ at $\lambda = 0$, i.e. by integrating the vector field of tangent vectors to give the group orbits, as mentioned above. For example, for (2.2) we have

$$\mathbf{X} = -u\partial_x + x\partial_u.$$

We shall also make extensive use of the scaling transformation and the translation of independent variables given by

$$\mathbf{X} = x\partial_x + \alpha u\partial_u \quad \text{and} \quad \mathbf{X} = \partial_x, \tag{2.4}$$

respectively. This can be extended to the case when u is a vector of dependent variables and we shall see examples of this when we consider systems of ODEs in Chapter 3, as well as the case of x being a vector of independent variables as we shall see when we consider partial differential equations in later Chapters. For the PDE case with one dependent variable u and two independent variables x and t, the scaling transformation above generalizes to

$$\mathbf{X} = \alpha_0 t\partial_t + \alpha_1 x\partial_x + \alpha_2 u\partial_u.$$

where x and t represent the spatial and temporal variables respectively. Now that we have given some background on Lie groups of point transformations we shall go on to show their relevance as *symmetries* of differential equations. We first note however that the Lie group method of analysing and finding solutions (see below) to differential equations is one of the only 'standard' methods which apply equally well to both linear and nonlinear problems. Also, since many other techniques (e.g. separation of variables, various transformations (e.g. hodograph), etc.) can be seen as special cases of the Lie group technique, it is often described as the most general and useful method for analysing differential equations.

Given an ordinary or partial differential equation problem $\mathbf{H} = 0$ where $\mathbf{H}$ is a function on the independent variables $\mathbf{x}$ and a differential operator on the dependent variables $\mathbf{u}$. Consider what happens to $\mathbf{H}(\mathbf{x}, \mathbf{u}) = 0$ when we substitute in the transformed variables (2.1), when we say that the differential equation is invariant we simply mean that we also have that $\mathbf{H}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}}) = 0$. Since this must be true for all λ we can differentiate with respect to λ and then set $\lambda = 0$ to give

$$0 = \left.\frac{\partial \mathbf{H}(\tilde{\mathbf{x}}, \tilde{\mathbf{u}})}{\partial \lambda}\right|_{\lambda=0} = \left.\left(\frac{\partial \mathbf{H}}{\partial \mathbf{x}}\frac{\partial \tilde{\mathbf{x}}}{\partial \lambda} + \frac{\partial \mathbf{H}}{\partial \mathbf{u}}\frac{\partial \tilde{\mathbf{u}}}{\partial \lambda} + \ldots\right)\right|_{\lambda=0}, \tag{2.5}$$

where the '...' in (2.5) denotes higher order terms arising from the differentiated terms present in $\mathbf{H}$. (Note that here differentiation with respect to a vector quantity denotes a gradient and scalar products have been dropped where it is obvious that they are needed.) But notice that this is equivalent to

$$\xi\frac{\partial \mathbf{H}}{\partial \mathbf{x}} + \eta\frac{\partial \mathbf{H}}{\partial \mathbf{u}} + \ldots = 0.$$

Therefore we may say that $\mathbf{H} = 0$ is invariant under the transformation described by the infinitesimal generator $\mathbf{X}$ if and only if, (see the references for a proof of the 'if' part of this statement),

$$\mathbf{X}^{(n)}\mathbf{H} = 0 \quad \text{whenever} \quad \mathbf{H} = 0. \tag{2.6}$$

Where $\mathbf{X}^{(n)}$ denotes the *prolongation* of $\mathbf{X}$, which simply means that $\mathbf{X}^{(n)}$ contains the extra transformation terms which correspond to the higher order terms in (2.5), and simply tells us how the various derivatives (up to the highest order, which we denote by n) in the problem transform. For example, under the scaling transformation $\tilde{x} = \lambda x$, $\tilde{u} = \lambda^\alpha u$, some $\lambda > 0$, we can quickly see by substitution and the chain rule for differentiation that the derivatives scale as

$$\frac{d\tilde{u}}{d\tilde{x}} = \lambda^{\alpha-1}\frac{du}{dx}, \quad \frac{d^2\tilde{u}}{d\tilde{x}^2} = \lambda^{\alpha-2}\frac{d^2u}{dx^2}, \quad \ldots,$$

and so for the first generator appearing in (2.4) we have

$$\mathbf{X}^{(2)} = x\partial_x + \alpha u\partial_u + (\alpha-1)u_x\partial_{u_x} + (\alpha-2)u_{xx}\partial_{u_{xx}}.$$

General formulae for the terms in $\mathbf{X}^{(n)}$ which correspond to the way in which derivatives transform under (2.1) can be written down, they can be calculated from the elements of the generator $\mathbf{X}$ directly using the chain rule. Since we shall primarily be interested in scalings in this thesis, where we have just seen that we immediately know how derivatives transform, we omit the formulae here and refer the reader to the references stated above.

2.2.2 Group invariant solutions

Given a differential equation there are systematic procedures for finding its symmetries, these generally reduce to solving an over-determined set of linear equations, see Appendix A for an example. Much of this process may be automated, and various computer algebra packages exist to perform the calculations, see [99] for a discussion of some of these packages. Having found the symmetries it may then be possible in many cases to use this additional information to go on to find exact solutions of the differential equation by first using the ansatz of a solution given by a function that is itself invariant under the symmetry group. These solutions are therefore termed *group invariant* and shall be considered further in the remainder of this Section on Lie group invariance.

Due to the richness in the behaviour of solutions to ordinary and particularly partial differential equations, which may have arbitrary initial conditions and complex boundary conditions, it is unlikely that the general solution of the differential equation will itself be invariant under the action of the symmetries that leave the equation invariant. However those special solutions which are we call *group invariant.* A most significant feature of group invariant solutions is that they need not be invariant under the full group of symmetries that leave the underlying equations invariant. In particular they may only be invariant under a particular sub-group.

For example consider the nonlinear wave equation

$$u_{tt} = u_{xx} + f(u). \tag{2.7}$$

For a general function f this equation is invariant under the two individual translation group actions ∂_t and ∂_x as well as obviously the combined action $\mathbf{X} = \partial_t + c\partial_x$, for any constant $c \in \mathbb{R}$. The latter leaves invariant travelling wave solutions, but only those moving at certain speeds are actually solutions to (2.7). A solution $u(x,t)$ of (2.7) which is itself invariant under the action of the group generated by $\mathbf{X}$ must take the form $u(x,t) = v(x-ct)$ and the function $v \equiv v(y)$ then satisfies (on substitution into (2.7)) the ordinary differential equation

$$c^2 v_{yy} = v_{yy} + f(v).$$

In this equation the wave speed c is an unknown, its value must be determined as part of the solution (see [81] for more details). That is the group invariant solutions in this case are invariant under a one-dimensional sub-group of the full two-dimensional group acting on the underlying differential equation. More generally, for a partial differential equation invariant under the action of several groups, determining the group under which the group invariant solution is actually invariant will form part of the solution process.

Group invariant solutions are of interest firstly because they provide exact solutions which may be used, for example, in the test of numerical methods. However their main usefulness lies in the fact that (as we shall see for solutions invariant under a scaling symmetry, and usually termed self-similar) they often describe the asymptotic limit of more general solutions which do not obey the group invariance. Many examples of this property may be found in [15, 184].

Group invariant, and in particular for this thesis self-similar, solutions play an incredibly important rôle in applied mathematics. Under certain circumstances they can be attractors [102, 172] for more general solutions, and hence give excellent approximations of asymptotic behaviour. They can also differentiate between different types of initial data which lead to qualitatively different forms of solution behaviour. More significantly, they often describe the *intermediate asymptotics* of a problem [15, 16]. That is, the behaviour of an evolutionary system at sufficiently long times so that the effects of initial data are not important, but before times in which the effects of boundary conditions dominate the solution (which shall see an example of this below). A self-similar solution also satisfies a *simpler* equation than the underlying differential equation. For example if we are considering a PDE with two independent variables the self-similar solution satisfies an ordinary differential equation. This has made them popular for computation — although they are normally singular, homoclinic or heteroclinic solutions of the ordinary differential equation and thus still remain a numerical challenge.

In a large number of interesting cases the precise group action under which a self-similar solution is invariant can not be found a priori and must be found as part of the solution procedure (as for the nonlinear wave equation in the case of translations, which are actually very closely related to scaling transformations through the exponential and logarithmic operations). Problems of this type are generally termed *self-similar of the second kind.* However in certain special cases the precise group action under which the self-similar solution is invariant may be determined from considerations such as dimensional analysis and conservation laws. This is of considerable advantage when computing the solution. For example if the speed of the travelling wave solution is known then the mesh can be required to move along with the solution. Such (less common) problems are termed *self-similar of the first kind.* We shall see an example of this kind of self-similarity when we consider the porous medium equation below.

Self-similar asymptotics of non self-similar general solutions

Consider the linear heat equation

$$u_t = u_{xx}, \tag{2.8}$$

which possesses the very well known exact solution

$$u(x,t) = \frac{1}{\sqrt{4\pi t}} e^{-x^2/4t}, \tag{2.9}$$

which just so happens to also be group invariant under a scaling transformation, i.e. is *self-similar.* See Section 2.2.6 for the full derivation of the self-similar solution to the nonlinear form of (2.8). In actual fact this is the exact solution for the problem posed with the idealized initial condition of an instantaneous release of 'heat' at the origin, i.e.

$$u(x,0) = A\,\delta(x),$$

where $\delta(\cdot)$ represents the Dirac delta function and $A \in \mathbb{R}$ is a constant. Now for the general problem with initial data given by $u(x,0) = u_0(x)$ the solution is given by the convolution

$$u(x,t) = \frac{1}{\sqrt{4\pi t}} \int_{-\infty}^{\infty} u_0(y) e^{-(x-y)^2/4t}\, dy. \tag{2.10}$$

Now that we have a special group invariant as well as a general solution we can make a non-rigorous comparison of the two and demonstrate that solution (2.9) does indeed here represent the asymptotic behaviour of (2.10). Expanding the integrand of (2.10) we have

$$u(x,t) = \frac{1}{\sqrt{4\pi t}} e^{-\xi^2} \left\{ \int_{-\infty}^{\infty} u_0(y)\, dy + \frac{\xi}{\sqrt{t}} \int_{-\infty}^{\infty} u_0(y) y\, dy + \dots \right\}, \tag{2.11}$$

where

$$\xi = \frac{x}{\sqrt{4t}}.$$

Each successive term of the expansion in (2.11) possesses t to the inverse power one half higher than the previous term. Therefore in the limit of large time the solution (2.10) corresponds to the idealized solution (2.9), with the constant A given by the initial *mass* of the solution,

$$A = \int_{-\infty}^{\infty} u_0(y)\, dy.$$

Hence for a wide range (for full rigour this statement would need to be clarified) of initial data the asymptotic behaviour of the solution is given by the special self-similar solution (2.9).

This very attractive property of self-similarity is not restricted to the simple example of linear diffusion. The references [184, 15] are replete with examples of this for many diverse physical problems.

2.2.3 Scaling invariance of ODEs

In this thesis we are particularly interested in problems which are invariant under the action of a scaling group. The invariance of physical equations under scaling groups is universal and expresses the deep physics that the equations representing physical processes should not depend upon the units in which they are measured [15].

Consider the ordinary differential equation system

$$\frac{d\mathbf{u}}{dt} = \mathbf{f}(\mathbf{u}), \quad \mathbf{u} = (u_1, u_2, \dots, u_N)^T, \quad \mathbf{f} = (f_1, f_2, \dots, f_N)^T. \tag{2.12}$$

We can immediately see that this system is invariant under the action of the time-translation symmetry

$$\mathbf{X} = \partial_t, \quad \text{i.e.} \quad t \rightarrow t + \lambda, \quad \forall\lambda. \tag{2.13}$$

A *linear* rescaling of the dependent and independent variables can be described by an $(N+1)$-tuple $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \dots, \alpha_N)$, such that if we introduce a rescaling parameter λ then the dependent and independent variables scale in the manner,

$$\mathbf{X} = \alpha_0 t \partial_t + \alpha_1 u_1 \partial_{u_1} + \dots + \alpha_N u_N \partial_{u_N}$$

i.e.

$$t \rightarrow \lambda^{\alpha_0} t, \quad u_i \rightarrow \lambda^{\alpha_i} u_i, \quad i = 1, \dots, N, \quad \forall\lambda > 0. \tag{2.14}$$

A typical example of this would be a change in units of measurement, for example if u is a velocity and t is time, then scaling $t \rightarrow \lambda t$ induces a scaling of $u \rightarrow u/\lambda$. Clearly, a physical problem should not depend upon the units of measurement and this leads to the concept of scale invariance for a system of equations. Note that in situations where we are considering the action of only one such scaling transformation we shall, without loss of generality, usually implicitly assume that $\alpha_0 = 1$. A system such as (2.12) is invariant under the action of the rescalings (2.14) provided that for each $i = 1, \dots, N$,

$$\lambda^{\alpha_i - \alpha_0} f_i(u_1, \dots, u_N) = f_i(\lambda^{\alpha_1} u_1, \dots, \lambda^{\alpha_N} u_N). \tag{2.15}$$

It is quite possible (and is often the case) that there may be many such $(N+1)$-tuples $\boldsymbol{\alpha}$ leaving (2.12) invariant. Indeed if $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$ are two such $(N+1)$-tuples then any linear combination corresponds to a further invariant transformation, and all such transformations commute (i.e. our groups of scaling transformations are Abelian). Hence the set of all admissible transformations is actually a vector space over $\mathbb{R}^{N+1}$.

We now give two examples of ODEs invariant under scaling transformations, both shall be considered further at later points.

Example 2.1 The blow-up equation.
Consider the ODE

$$\frac{du}{dt} = u^4, \tag{2.16}$$

this equation is invariant under the scaling transformation

$$t \to \lambda t, \quad u \to \lambda^{-1/3} u,$$

and so in the above notation we have $\boldsymbol{\alpha} = (1, -1/3)$.

Example 2.2 Gravitational collapse.
Consider the motion of a particle in a one-dimensional gravitational field given by

$$\frac{d^2 r}{dt^2} = -\frac{1}{r^2}, \quad \text{equivalently} \quad \frac{dv}{dt} = -\frac{1}{r^2}, \quad \frac{dr}{dt} = v, \tag{2.17}$$

which is invariant under the scaling transformation

$$t \to \lambda t, \quad r \to \lambda^{2/3} r, \quad \text{with} \quad v \to \lambda^{-1/3} v.$$

We therefore have $\boldsymbol{\alpha} = (1, 2/3, -1/3)$ (for the first-order formulation, which as should obviously be the case is identical to the once prolonged action applicable to the second-order formulation). See also Appendix A.

Both of these problems have solutions which develop singularities in finite time, and the scaling invariance of the problem plays a significant role in describing this.

2.2.4 Self-similar solutions of ODEs

As stated above, most solutions of an ordinary differential equation invariant under a scaling group are (due to the prescription of arbitrary initial (and in the case of PDEs boundary) conditions) not themselves invariant under the action of the group. An important class of solutions do however have this property and are called *self-similar.* These solutions are only admitted if the initial data satisfies certain algebraic constraints. Similarly to the asymptotic behaviour discussed above, self-similar solutions are especially important in the rôle that they play in describing singularity formation after the effects of initial conditions have decayed away. The ability of a numerical method to accurately represent self-similarity is therefore an important test both for its ability to compute singular behaviour and for it to represent the true long time asymptotics of the problem.

A solution to problem (2.12) is termed *self-similar* if it is itself invariant under the action of the transformation (2.14), that is if (assuming $\alpha_0 = 1$)

$$u_i(\lambda t) = \lambda^{\alpha_i} u_i(t), \quad i = 1, \ldots, N. \tag{2.18}$$

Differentiating with respect to λ and setting $\lambda = 1$ we obtain

$$t\frac{du_i}{dt} = \alpha_i u_i, \quad i = 1, \ldots, N,$$

which we call the *invariant curve condition* for the case of scaling transformations, from which we may find every curve $\mathbf{u} \equiv \mathbf{u}(t)$ in phase space invariant under (2.14). Solving in this case gives

$$u_i(t) = t^{\alpha_i} U_i, \quad i = 1, \ldots, N, \tag{2.19}$$

(up to translations and reflexions in t). We may now use this as a solution ansatz and determine the values of the constants U_i by substituting (2.19) into the original differential equation to give the *algebraic* system

$$\alpha_i U_i = f_i(\mathbf{U}), \quad i = 1, \ldots, N, \quad \mathbf{U} = (U_1, U_2, \ldots, U_N). \tag{2.20}$$

For example, for problem (2.16) we need to solve

$$-\frac{1}{3}U = U^4,$$

which has the solution $U = -3^{-1/3}$, (see example 3.3). An application of the translational symmetry (2.13), as well as the reflexional symmetry ($t \to -t$), gives us the family of self-similar solutions

$$u(t) = (3(C - t))^{-1/3}. \tag{2.21}$$

In fact this is the general solution for this problem, and in particular shows that the problem with initial condition $u(0) = u_0$ *blows up* at time $t = u_0^{-3}/3$.

Similarly for problem (2.17) we solve

$$\frac{2}{3}R = V, \quad -\frac{1}{3}V = -R^{-2}.$$

This yields

$$R = \left(\frac{9}{2}\right)^{1/3}, \quad V = \frac{2}{3}\left(\frac{9}{2}\right)^{1/3},$$

and we have the *expanding* self-similar solution

$$r = Rt^{2/3}, \quad v = -Vt^{-1/3},$$

as well as, following an application of translational and reflexional symmetries, the *collapsing* self-similar solution

$$r = R(T-t)^{2/3}, \quad v = -V(T-t)^{-1/3},$$

where T is an arbitrary finite (collapse) time. This solution is of interest to us as it forms a singularity in a finite time in which $r \to 0$ and $v \to -\infty$ as $t \to T$. We immediately observe that it is difficult to capture such behaviour if a fixed time step is used. Indeed, an explicit method will always give a bounded solution, and an implicit method may not have soluble algebraic equations.

2.2.5 Scaling invariance of PDEs

Now suppose that $\mathbf{u}(\mathbf{x}, t)$ satisfies a partial differential equation, say of the form

$$\mathbf{N}(\mathbf{u}, \mathbf{u}_\mathbf{x}, \mathbf{u}_{\mathbf{xx}}, \mathbf{u}_t, \mathbf{u}_{tt}, \mathbf{x}, t) = \mathbf{0}. \tag{2.22}$$

As above we define a *symmetry* of this equation to be any transformation of $\mathbf{u}$, $\mathbf{x}$ and t which leaves it unchanged (or invariant). Note that any PDE problem will generally be posed with initial and boundary conditions (IBCs), for the entire problem to possess a symmetry the IBCs also need to be invariant under the symmetry. However, in this work we shall be concerned with the construction of numerical methods which compute solutions to problems with arbitrary IBCs, these are then free to converge to (if this is their true behaviour) self-similar solutions of (2.22) not necessarily obeying the same IBCs as the problem being integrated. We shall therefore pay no further attention to IBCs when we think about symmetries of differential equation problems.

As in the ODE discussion above, here we are primarily interested in scaling transformations of the form, for $i = 1, \ldots, N_1$, $j = N_1 + 1, \ldots, N_1 + N_2$

$$t \to \lambda^{\alpha_0} t, \quad x_i \to \lambda^{\alpha_i} x_i, \quad u_j \to \lambda^{\alpha_j} u_j. \tag{2.23}$$

Here λ is considered to be an arbitrary positive quantity.

The book [15] gives many examples of systems of partial differential equations with such symmetries. These arise very naturally in many problems as they express the way that a differential equation changes when the units of measurement in which it is expressed also change.

It is an observed fact [15], as mentioned in previous Sections, that scaling (power-law) relationships arise naturally and have wide applications in science and in engineering. It is for this reason that many of the conservation laws of mathematical physics have polynomial nonlinearities as it is precisely these which are invariant under the action of such symmetries. Far from being an approximation of the actual behaviour of the equations, such scalings give evidence of deep properties of the phenomena they represent, which may have no intrinsic time or length scale, and which have solutions that *reproduce themselves in time and space* under rescaling. This is an example of a covariance principle in physics that *the underlying solutions of a partial differential equation representing a physical phenomenon should not have a form which depends upon the location of the observer or the units that the observer is using to measure the system.*

Indeed, far from being a special case, scaling symmetries of the form (2.23) are universal in physics and the applied sciences. They can be found, and have important applications, in for example, fluid mechanics [21, 89, 144, 143], turbulence [70, 14, 17], the theory of detonation and combustion [184, 142], heat diffusion and filtration [105, 172], gas dynamics [184, 177], mathematical biology [127], free boundary Stefan problems [64], boundary layer theory (including the well known Blasius example) [21, 18], general relativity [42]. The list and references can be continued to include virtually every application considered by applied mathematicians in the physical and applied sciences. Scaling invariance is also closely tied up with the theory of fractals [170], and with the general theory of dimensional analysis [151] (including the famous Buckingham-Pi theorem [27] for example) and renormalization group theory [72, 43].

Motivated by the definition (2.23) we introduce a vector $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_{N_1+N_2})$ to describe the scaling group. As was the case for ODEs, it is evident that for any such $\boldsymbol{\alpha}$ the vector $\mu\boldsymbol{\alpha}$ also describes the same scaling transformation. It is quite possible for the same system of partial differential equations to be invariant under several such scaling transformations. It is then easy to check that the scaling operations described by two separate vectors commute. Indeed, the set of vectors corresponding to scaling transformations which leave the partial differential equation invariant form a *commutative vector space.*

Example 2.3 The porous medium equation.
We consider here the example

$$u_t = (uu_x)_x, \tag{2.24}$$

usually termed the porous medium equation. This is an example of a nonlinear diffusion equation and is considered in far greater detail in Chapter 5. This problem admits four continuous transformation groups, the two groups of translations in time and space

$$\mathbf{X} = \partial_t, \quad \text{and} \quad \mathbf{X} = \partial_x,$$

and the two-dimensional vector space of scaling symmetry groups spanned by the operators

$$\mathbf{X}_1 = t\frac{\partial}{\partial t} + \frac{1}{2}x\frac{\partial}{\partial x} \quad \text{and} \quad \mathbf{X}_2 = t\frac{\partial}{\partial t} - u\frac{\partial}{\partial u}.$$

In particular (2.24) is invariant under the transformation generated by

$$\mathbf{X} = k_1\mathbf{X}_1 + k_2\mathbf{X}_2 = (k_1 + k_2)t\frac{\partial}{\partial t} + \frac{k_1}{2}x\frac{\partial}{\partial x} - k_2 u\frac{\partial}{\partial u}, \tag{2.25}$$

for arbitrary constants k_1 and k_2.

2.2.6 Self-similar solutions of PDEs

As was mentioned in the Section on self-similar solutions to ODE problems, the majority of solutions to PDEs are not themselves invariant under the same transformations as the PDE. Solutions which do have this property we again term self-similar and we now give a derivation of the self-similar solution to the porous medium equation (2.24).

Any solution u of (2.24) will become another solution under the transformation generated by (2.25). To ensure that this solution is itself invariant under the transformation we require that (where we assume without loss of generality that $k_1 + k_2 = 1$),

$$\lambda^\gamma u(x,t) = u(\lambda^\beta x, \lambda t), \tag{2.26}$$

where $\beta = k_1/2$, $\gamma = -k_2$ and we must have $2\beta - \gamma = 1$. Such a u is called a *self-similar solution* of (2.24). Differentiating (2.26) with respect to λ and setting $\lambda = 1$ yields the *invariant surface condition*

$$\gamma u = \beta x u_x + t u_t,$$

which has the characteristic equations

$$\frac{du}{\alpha u} = \frac{dx}{\beta x} = \frac{dt}{t},$$

from which we obtain two quantities invariant under the action of the transformation group, these are

$$\frac{u}{t^\gamma} \quad \text{and} \quad y := \frac{x}{t^\beta}$$

Setting one of these to be an arbitrary function of the other gives us the most general form of the self-similar solution, that is

$$u(x,t) = t^\gamma \tilde{u}(x/t^\beta). \tag{2.27}$$

Without additional conditions any such solution is possible, however, if we impose the condition that $u(x,t)$ decays sufficiently fast as $|x| \to \infty$ then a simple calculation shows that if the mass and centre of mass of the solution are given by

$$I_1 = \int u(x,t)\,dx \quad \text{and} \quad I_2 = \int xu(x,t)\,dx, \tag{2.28}$$

respectively, then both are constant for all t since

$$\frac{d}{dt}\int_{-\infty}^{\infty} u\,dx = \int_{-\infty}^{\infty} u_t\,dx = \int_{-\infty}^{\infty} (uu_x)_x\,dx = 0,$$

and

$$\frac{d}{dt}\int_{-\infty}^{\infty} xu\,dx = \int_{-\infty}^{\infty} x(uu_x)_x\,dx = -\int_{-\infty}^{\infty} uu_x\,dx = -\frac{1}{2}\int_{-\infty}^{\infty} (u^2)_x\,dx = 0.$$

In what follows assume that the initial first integral takes the value 1, and the initial centre of mass is at the origin. Since we are seeking a self-similar solution, which by definition is unchanged under the scaling transformation generated by (2.25), its mass is obviously also left unchanged under the transformation, i.e.

$$\int_{-\infty}^{\infty} u\,dx = t^{\gamma}\int_{-\infty}^{\infty} \tilde{u}(x/t^{\beta})\,dx = t^{\gamma+\beta}\int_{-\infty}^{\infty} \tilde{u}(y)\,dy = \text{const},$$

where, as above, $y = xt^{-\beta}$. Therefore $\gamma + \beta = 1$ and we deduce that $\beta = -\gamma = 1/3$. So although the equation itself is invariant under a two-dimensional space of scaling transformations, there only actually exists a self-similar solution corresponding to a one-dimensional subspace of these transformations. Substituting (2.27) into (2.24) gives the *principal ODE*

$$-\frac{1}{3}(\tilde{u} + y\tilde{u}') = (\tilde{u}\tilde{u}')'.$$

In passing from (2.24) to the above (this procedure is called *symmetry reduction*), the various powers of x and t that appeared in the derivatives of u either cancelled or could be combined into powers of y. This could only happen because we wrote our most general form for the self-similar solution in terms of the invariants derived above. One of the central rôles of group invariance is to help us to discover these correct variables, they are called dimensionless groupings or *similarity variables* [61]. Integrating, and using an assumption of reflexional symmetry about $x = 0$ to eliminate the constant of integration, shows that $\tilde{u}$ satisfies the first order ODE

$$\frac{y\tilde{u}}{3} + \left(\frac{\tilde{u}^2}{2}\right)' = 0.$$

Integrating again gives

$$\tilde{u}(y) = \left(a - \frac{y^2}{6}\right)_+,$$

where a can be determined by specifying the mass I_1 and noting that $\tilde{u}(y)$ has support on $[-\sqrt{6a}, \sqrt{6a}]$. For example, for $I_1 = 1$ a simple calculation gives $a = (3/32)^{1/3}$. Hence finally we have the self-similar solution

$$\hat{u}(x,t,a) = t^{-1/3}\left(a - \frac{1}{6}\left(\frac{x}{t^{1/3}}\right)^2\right)_+, \tag{2.29}$$

where we use the notation $(\cdot)_+ = \max\{\cdot, 0\}$. These solutions were discovered independently by Barenblatt and Pattle, see [15].

This self-similar solution can be shown (see [183]), as we did earlier for the linear diffusion equation, to describe the asymptotic behaviour for general non self-similar solutions to the porous medium equation. In Chapter 5 we shall derive a numerical method which, due to the fact that it does not destroy the scaling invariance property which leads to self-similarity, also exhibits this convergence to self-similarity in its discrete solutions.

Barenblatt [15, 16] explains in more generality that self-similar solutions do not only describe the evolution of a system under special idealized initial and boundary conditions. But that they also describe the *intermediate asymptotic behaviour* of a wider class of solutions in the ranges where they no longer depend upon the details of the initial and boundary conditions, but before the problem has reached a limiting case.

2.3 Hamiltonian ODEs and symplectic integration

The first serious and significant application of geometric ideas to numerical analysis was in the integration of Hamiltonian ordinary differential equations, (for general introductions and reviews of this material see [146, 164, 84, 35, 34, 36]). This is natural as Hamiltonian systems possess many strong structural properties and appear frequently in many varied applications, for example celestial and molecular dynamics, hydrodynamics, classical field theories (Maxwell's equations for example), classical and quantum mechanics as well as general relativity. We shall consider an application in geophysical fluid dynamics in Chapters 6–8. In general the analysis of Hamiltonian problems has almost always centred on underlying geometrical structures. In this Section we shall firstly present some reasonably well-known background information on Hamiltonian ODEs and their properties. This will be followed by some examples of integration methods designed to preserve some of these properties. Having introduced these methods we shall use this opportunity to demonstrate the superior behaviour of geometric integration methods in certain situations. We also mention the interesting links between these Hamiltonian properties, conservation laws and symmetries, and finish with a discussion of how to extend some of these results to infinite dimensions, i.e. to Hamiltonian partial differential equations.

2.3.1 Background theory on Hamiltonian ODEs

For classical introductions to this material see [7, 73]. Consider initially a mechanical system with (generalized) coordinates $\mathbf{q} \equiv \mathbf{q}(t) \in \mathbb{R}^d$ and Lagrangian L (a real-valued function of $\mathbf{q}$, $\dot{\mathbf{q}}$ and t), which for classical mechanics generally takes the form $L = T - V$, where $T \equiv T(\mathbf{q}, \dot{\mathbf{q}})$ represents the kinetic energy of the system and $V \equiv V(\mathbf{q})$ its potential energy.

Example 2.4 Kepler's problem
The Kepler (or two-body) problem shall be a heavily used model problem throughout this and the following two Chapters. It describes the motion of two bodies attracted by one another, for example heavenly bodies under the action of gravity. If we choose one of the bodies as the centre of our coordinate system, then $\mathbf{q} = (q_1, q_2)^T$ represents the position of the second body, and $\dot{\mathbf{q}} = (\dot{q}_1, \dot{q}_2)^T$ its velocity. Assuming normalized masses for the two bodies and an inverse square law attraction force, the kinetic and potential energies are given by, respectively

$$T(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2), \quad V(\mathbf{q}) = -\frac{1}{\sqrt{q_1^2 + q_2^2}}.$$

The Lagrangian for this system is therefore

$$L = \frac{1}{2}(\dot{q}_1^2 + \dot{q}_2^2) + \frac{1}{\sqrt{q_1^2 + q_2^2}}.$$

The dynamics of such a system can be studied in terms of the calculus of variations [77] by considering the action functional

$$\int_{t_0}^{t_1} L(\mathbf{q}, \dot{\mathbf{q}}, t)\, dt, \tag{2.30}$$

which is simply the integral of L along a curve $\mathbf{q}(t)$. We then compute variations of the action whilst holding the endpoints of the curve $\mathbf{q}(t)$ fixed. See details of this derivation in [116] together with a detailed discussion of the numerical methods derived by actually discretizing this action functional. Hamilton's principle of least action tells us that motions of mechanical systems coincide with extremals of (2.30). It can be shown [77] that this procedure leads to the following Euler-Lagrange equations describing the motion

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\mathbf{q}}}\right) - \frac{\partial L}{\partial \mathbf{q}} = 0. \tag{2.31}$$

Following through this procedure for Kepler's problem we easily find the equations of motion to be

$$\ddot{q}_i = -\frac{q_i}{(q_1^2 + q_2^2)^{3/2}}, \qquad i = 1, 2.$$

Hamilton recognized that the Euler-Lagrange equations (2.31) could be put into a form which allowed a more geometrical analysis. In particular he introduced the coordinates

$$\mathbf{p} := \frac{\partial L}{\partial \dot{\mathbf{q}}} \in \mathbb{R}^d,$$

which are the conjugate generalized momenta of the system. He further defined the Hamiltonian via a Legendre transformation (see Chapters 6 and 7 for more details on Legendre transforms) as

$$H(\mathbf{p}, \mathbf{q}) = \mathbf{p}^T \dot{\mathbf{q}} - L(\mathbf{q}, \dot{\mathbf{q}}) \tag{2.32}$$

and showed that (2.31) is equivalent to the following system of $2d$ first-order equations,

$$\dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}}, \quad \dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}. \tag{2.33}$$

This is called the *canonical form* for a Hamiltonian system. Note that for our mechanical problem $H \equiv T + V$, and thus the Hamiltonian actually represents the total energy present in the system.

Returning to Kepler's problem we have $\mathbf{p} = \dot{\mathbf{q}}$, the Hamiltonian

$$H(p_1, p_2, q_1, q_2) = \frac{1}{2}(p_1^2 + p_2^2) - \frac{1}{\sqrt{q_1^2 + q_2^2}}, \tag{2.34}$$

and the Hamiltonian formulation for the problem is given by

$$\frac{dq_i}{dt} = p_i, \quad \frac{dp_i}{dt} = -\frac{q_i}{(q_1^2 + q_2^2)^{3/2}}, \qquad i = 1, 2. \tag{2.35}$$

More generally, if a system of ordinary differential is defined in terms of $\mathbf{u} \in \mathbb{R}^{2d}$, where $\mathbf{u} = (\mathbf{p}, \mathbf{q})^T$ with $\mathbf{p}, \mathbf{q} \in \mathbb{R}^d$, such that

$$\dot{\mathbf{u}} = \mathbf{f}(\mathbf{u}), \tag{2.36}$$

then this system is canonically Hamiltonian if

$$\mathbf{f}(\mathbf{u}) = J^{-1} \nabla H \tag{2.37}$$

where $H = H(\mathbf{p}, \mathbf{q})$ is the Hamiltonian function, ∇ is the gradient operator

$$\nabla \equiv \left(\frac{\partial}{\partial p_1}, \ldots, \frac{\partial}{\partial p_d}, \frac{\partial}{\partial q_1}, \ldots, \frac{\partial}{\partial q_d} \right)^T,$$

and J is the skew-symmetric matrix,

$$J = \begin{pmatrix} 0 & I_d \\ -I_d & 0 \end{pmatrix} \tag{2.38}$$

where I_d represents the identity matrix of dimension d. In this case $\mathbf{f}$ is called a Hamiltonian vector field. It is easy to see that if $\mathbf{f}_1$ and $\mathbf{f}_2$ are Hamiltonian vector fields then so is the vector field $\mathbf{f}_1 + \mathbf{f}_2$. We shall exploit this simple result below to derive some simple geometric integration methods.

Possibly the simplest geometric property of systems which admit the formulation (2.33) is conservation of the Hamiltonian following the flow. This can easily be verified through the following, where we make use of the chain rule and Hamilton's equations (2.33),

$$\frac{dH}{dt} = \frac{\partial H}{\partial \mathbf{p}} \cdot \dot{\mathbf{p}} + \frac{\partial H}{\partial \mathbf{q}} \cdot \dot{\mathbf{q}} \equiv 0. \tag{2.39}$$

(Note that this property no longer holds for Hamiltonians which depend explicitly upon time.) In addition to this property a key feature of a Hamiltonian system is the *symplecticity* of its flow. The solution or flow of the system (2.36) induces a transformation $\psi(t)$ on the phase space $\mathbb{R}^{2d}$, with associated Jacobian ψ'. Such a map is said to be *symplectic* if,

$$\psi'^T J \psi' = J, \tag{2.40}$$

where J is defined as above. Symplectic maps have the highly useful property that they combine to give other symplectic maps,

Lemma 2.1. *If ψ and φ are symplectic maps then so is the composition $\psi \circ \varphi$.*

Proof. If ψ and φ are both symplectic maps then

$$(\psi \circ \varphi)'^T J (\psi \circ \varphi)' = (\psi'\varphi')^T J (\psi'\varphi') = \varphi'^T \psi'^T J \psi' \varphi' = \varphi'^T J \varphi' = J,$$

yielding the desired result. □

Symplecticity (or the preservation of a symplectic structure Ω) has the following important geometric interpretation. If M is any 2-dimensional manifold in $\mathbb{R}^{2d}$, we can define Ω to be the integral of the sum over the orientated areas, of its projections onto the (p_i, q_i) plane (so that if $d = 1$ this is simply the area of M). If ψ is a symplectic map, then $\Omega(M)$ is conserved throughout the evolution. In particular in one-dimension ($d = 1$) areas are conserved. The most succinct way to properly define (and definitely prove results) is to introduce some differential geometry notation and theory, in particular differential forms, wedge products, etc, [7, 114]. The preserved object just mentioned can then be defined in terms of a differential two-form. However this is not really necessary for this thesis and would probably simply confuse matters, we therefore omit some of the rigour in this Section. A key result of Poincaré relating symplectic flows to Hamiltonian systems is the following, again for proofs see any of the previous references.

Lemma 2.2. *The flow $\psi(t)$ induced by a Hamiltonian function H via the differential equation (2.36) with (2.37) is symplectic, and in actual fact (given the correct technical assumptions) symplecticity of a flow holds if and only if the flow is Hamiltonian.*

The symplecticity property is much stronger than simple preservation of $2d$-dimensional volume, for example Hamiltonian systems preserve volume (Liouville's theorem) but it is possible to find volume preserving systems which are not Hamiltonian. From the perspective of dynamics, symplecticity plays a central role. In particular it means that the behaviour of Hamiltonian systems is recurrent with solutions from any point in $\mathbb{R}^{2d}$ returning arbitrarily closely to their starting point. Furthermore (unlike dissipative systems) the dynamics of a Hamiltonian system can not evolve onto a low dimensional attractor. Hamiltonian dynamics may be described by the celebrated KAM (Kolmogorov-Arnold-Moser) theorem [7, 67] which describes how the solution space of integrable and near integrable Hamiltonian systems (whose solutions are generically periodic or are confined to tori) perturb under Hamiltonian perturbations to tori surrounded by regions of chaotic behaviour. Further details may be found in [7, 114, 128].

2.3.2 Abstract notation and theory

We now move a little into abstraction and consider arbitrary Hamiltonian problems which do not necessarily arise from mechanical systems or classical mechanics. Given a finite-dimensional smooth manifold M, a *Poisson bracket* on M is an operation that assigns a smooth real-valued function $\{F, G\}$ on M for each pair of smooth real-valued functions F, $G : M \to \mathbb{R}$, and such that the following hold:

Bilinearity:

$$\{\alpha F + \beta G, H\} = \alpha\{F, H\} + \beta\{G, H\}, \quad \{F, \alpha G + \beta H\} = \alpha\{F, G\} + \beta\{F, H\},$$

Skew-symmetry:

$$\{F, G\} = -\{G, F\},$$

Jacobi identity:

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

Leibniz' Rule:

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

for all real constants α and β, and smooth real-valued functions F, G and H on M.

Given local coordinates $\mathbf{x}=(x_1,\dots,x_m)^T$ on the m-dimensional manifold M, we can write a given Poisson bracket in the form

$$\{F,G\}=\nabla F\cdot J\nabla G, \tag{2.41}$$

where ∇F represents the gradient of F with respect to $\mathbf{x}$, and $J\equiv J(\mathbf{x})$ is an $m\times m$ dimensional matrix called the *structure matrix.* Given a bracket in the form (2.41) the conditions to be a Poisson bracket listed above can now be given in terms of J alone, they are:

Skew-symmetry:

$$J^{ij}(\mathbf{x})=-J^{ji}(\mathbf{x}),\quad i,j=1,\dots,m,$$

Jacobi identity:

$$\sum_{l=1}^{m}\left(J^{il}\frac{\partial}{\partial x_l}J^{jk}+J^{kl}\frac{\partial}{\partial x_l}J^{ij}+J^{jl}\frac{\partial}{\partial x_l}J^{ki}\right),\quad i,j,k=1,\dots,m.$$

Conditions analogous to bilinearity and the Leibniz rule are not required since they are automatically satisfied by bracket operations of the form (2.41).

Given a Hamiltonian function H on M we may now write Hamilton's equations as,

$$\frac{d\mathbf{x}}{dt}=J(\mathbf{x})\nabla H(\mathbf{x})=\{\mathbf{x},H\}.$$

Example 2.5 With $M=\mathbb{R}^{2n}$, coordinates $\mathbf{x}=(q_1,\dots,q_n,p_1,\dots,p_n)^T$ on M, and structure matrix

$$J=\begin{pmatrix}0 & I_n\\ -I_n & 0\end{pmatrix}, \tag{2.42}$$

the Poisson bracket as well as Hamilton's equations take canonical form, i.e.

$$\frac{dq_i}{dt}=\frac{\partial H}{\partial p_i},\quad \frac{dp_i}{dt}=-\frac{\partial H}{\partial p_i},\quad i=1,\dots,n. \tag{2.43}$$

For example the harmonic oscillator, the simple pendulum, and the Kepler two-body problem all have this form.

Example 2.6 Now suppose that the structure matrix has the special form of being linear in $\mathbf{x}$, i.e.

$$J^{ij} = \sum_{k=1}^{n} c_{ij}^{k} x_k, \tag{2.44}$$

(the c_{ij}^k being constants, in fact structure constants for some Lie algebra $\mathfrak{g}$). The equations for rigid body motion can be put into this form and in this case the associated Lie-algebra is $\mathfrak{so}(3)$. Systems which have this property are termed *Lie-Poisson*, see [114, 128]. The associated Lie-Poisson bracket gives a natural Poisson structure (a generalization of the symplectic structure) on spaces given by the dual of Lie algebras. Physically these systems often arise through the *reduction* [114, 124] of other systems. For example (in a PDE context, where details shall be given later and in Chapter 6) the reduction for an ideal fluid (based upon a particle relabelling symmetry [129] and the preservation of potential vorticity [141]) from a canonical Lagrangian formulation results in a noncanonical Lie-Poisson Eulerian formulation.

Given Poisson manifolds M and N, that is manifolds endowed with Poisson brackets $\{\cdot,\cdot\}_M$ and $\{\cdot,\cdot\}_N$ respectively, we call a smooth mapping $\varphi : M \to N$ a *Poisson map* if it preserves the bracket operations, i.e. if

$$\{F\circ\varphi, G\circ\varphi\}_M = \{F,G\}_N \circ \varphi, \quad \forall F,G : N \to \mathbb{R}.$$

In particular, the flow corresponding to Hamilton's equations, as defined above, determines a Poisson map from M to itself. This is one of the properties of Hamiltonian problems which will interest us when we come to look at numerical methods.

Although it shall not be necessary to go into the differences or details too much in this thesis, this definition of Poisson systems and maps is actually more general and contains as a special case symplectic maps and systems. The difference basically boils down to the invertibility or noninvertibility of the matrix J appearing in (2.41). For example for the rigid body problem [114] the system is three-dimensional, and the structure matrix J for this system is therefore of size 3×3 and noninvertible. Therefore although the system is Hamiltonian it has a Poisson rather than a symplectic structure underlying it. For some further details see [114, 128, 63, 36] for example.

2.3.3 Symplectic numerical methods

Suppose now that a numerical one-step method of constant step size Δt is applied to approximate the solution $\mathbf{u}(t)$ of (2.36) at time $t = n\Delta t$ by the vector $\mathbf{u}_n \in \mathbb{R}^{2d}$. The numerical scheme will induce a discrete flow mapping $\Psi_{\Delta t}$ on $\mathbb{R}^{2d}$ which will be an approximation to the continuous flow map $\psi(\Delta t)$. We define the map $\Psi_{\Delta t}$ (and therefore also the method) to be symplectic if it also satisfies the identity (2.40). A natural geometric property that we would require of $\Psi_{\Delta t}$ is that it should be symplectic whenever ψ is. We will now consider the construction of numerical methods with this property. As we shall see, such methods retain many of the other features (in particular the ergodic properties) of their continuous counterparts.

A simple example of a symplectic method is the implicit midpoint rule, which for $\dot{\mathbf{u}} = \mathbf{f}(\mathbf{u})$ takes the form

$$\mathbf{u}_{n+1} = \mathbf{u}_n + \Delta t\, \mathbf{f}\left(\frac{1}{2}(\mathbf{u}_n + \mathbf{u}_{n+1})\right).$$

If the time step Δt is not constant, such as in an adaptive method, then particular care has to be taken with this definition of numerical symplecticity, and much of the advantage in using a symplectic scheme is lost [146] unless the technique defining Δt at each time step also preserves the geometric structure in some way. We shall return to consider this issue further in Chapter 4.

Early work which specifically aimed to construct symplectic methods for ordinary differential equations is given in [174, 140, 65]. These early constructions were rather involved and much simpler derivations have followed, in particular see [146]. However, symplectic integrators themselves have been used for far longer than this. Some well established and very effective numerical methods have been successful precisely because they are symplectic even though this fact may not have been recognized when they were originally constructed. For example the Gauss-Legendre methods and the Störmer-Verlet (leapfrog) method mentioned in Chapter 1. Symplectic methods are, in general, constructed using one of four different methods. These are generating function methods, certain Runge-Kutta methods, splitting methods and variational methods. We shall demonstrate some simple examples of symplectic splitting methods presently.

Whilst such methods can be constructed to preserve symplecticity, and hence many of the qualitative features of a Hamiltonian problem such as invariant sets and orbit statistics, they do not (in particular) conserve the Hamiltonian itself (see [76]) unless an adaptive time step is used [103]. However they can remain exponentially (in Δt) close to H for exponentially long times, this result may be established using the useful technique of modified equation or backward error analysis [85, 84, 136, 36]. Importantly, symplectic methods can have far more favourable error growth properties. It is important also to observe that Hamiltonian systems arise naturally in partial differential equations for which the associated systems (say obtained through a semi-discretization) are typically stiff. A conventional stiff solver such as a BDF method may introduce artificial dissipation into higher order modes, producing quite false qualitative behaviour. To resolve the energy transfer into the higher modes and to retain the correct dynamics of these modes a symplectic solver is ideal.

Some simple examples of symplectic methods

We shall consider here three numerical schemes. The simplest possible is the first-order explicit Forward Euler method which when applied to problem (2.33) takes the form,

$$\begin{aligned} \mathbf{p}^{n+1} &= \mathbf{p}^n - \Delta t H_q(\mathbf{p}^n, \mathbf{q}^n), \\ \mathbf{q}^{n+1} &= \mathbf{q}^n + \Delta t H_p(\mathbf{p}^n, \mathbf{q}^n). \end{aligned}$$

However the natural partitioning present in (2.33) suggests the use of partitioned methods [86]. If we combine the Backward (implicit) Euler method for one equation, and the Forward (explicit) Euler method for the other, we get the following implicit first-order scheme — the Symplectic Euler method,

$$\begin{aligned} \mathbf{p}^{n+1} &= \mathbf{p}^n - \Delta t H_q(\mathbf{p}^{n+1}, \mathbf{q}^n), \\ \mathbf{q}^{n+1} &= \mathbf{q}^n + \Delta t H_p(\mathbf{p}^{n+1}, \mathbf{q}^n). \end{aligned}$$

For systems with separable Hamiltonians, that is those for which we may write $H = T+V$, where $T \equiv T(\mathbf{p})$ and $V \equiv V(\mathbf{q})$, (2.34) is an example of this case, it turns out that partitioned Runge-Kutta methods may be used to yield *explicit* symplectic methods, a result which is not true for discretizations of problems with general Hamiltonians. For example, the symplectic Euler method is explicit when applied to problems of this form.

For our final scheme, if we now consider the separable Hamiltonian case and apply the two-stage Lobatto IIIA-B Runge-Kutta pair (see [84]) we obtain the following second-order explicit symplectic method.

$$\begin{aligned} \mathbf{q}^{n+1/2} &= \mathbf{q}^{n-1/2} + \Delta t T_p(\mathbf{p}^n), \\ \mathbf{p}^{n+1} &= \mathbf{p}^n - \Delta t V_q(\mathbf{q}^{n+1/2}). \end{aligned}$$

This scheme is usually termed the Störmer-Verlet or leapfrog method. A form of this method appeared in the molecular dynamics literature [173] many years before anyone realized that its remarkable success in that field was due to the fact that it was actually a very efficient symplectic method.

This idea of decomposing the Hamiltonian (or equivalently the differential system arising from it) into more than one part turns out to be a good motivation for the class of *splitting* methods, of which symplectic Euler and Störmer-Verlet are two examples. In such methods the whole problem is *split* into simpler problems and each then solved separately. For example, for our Hamiltonian of the form $H = T(\mathbf{p}) + V(\mathbf{q})$ consider the Hamiltonian systems generated by $T(\mathbf{p})$ and $V(\mathbf{q})$ separately, i.e.

$$\begin{array}{l|l} \dot{\mathbf{p}} = 0 & \dot{\mathbf{p}} = -V_q(\mathbf{q}) \\ \dot{\mathbf{q}} = T_p(\mathbf{p}) & \dot{\mathbf{q}} = 0, \end{array}$$

which can be solved exactly to give

$$\begin{array}{l|l} \mathbf{p}(t) = \mathbf{p}_0 & \mathbf{p}(t) = \mathbf{p}_0 - V_q(\mathbf{q}_0)t \\ \mathbf{q}(t) = \mathbf{q}_0 + T_p(\mathbf{p}_0)t & \mathbf{q}(t) = \mathbf{q}_0. \end{array}$$

If we now denote the time-t flows of these split systems by φ_t^T and φ_t^V respectively, then it can be easily checked that the symplectic Euler method is given by the composition $\varphi_{\Delta t}^T \circ \varphi_{\Delta t}^V$, and the Störmer-Verlet method by $\varphi_{\Delta t/2}^V \circ \varphi_{\Delta t}^T \circ \varphi_{\Delta t/2}^V$. The latter is often called the *Strang splitting.* Since φ_t^T and φ_t^V are the exact flows for Hamiltonian problems they, and their compositions, are all symplectic mappings by Lemmas 2.1 and 2.2. This gives both a quick proof of the symplecticity of these two methods, as well as an introduction to the ideas behind splitting and composition methods.

Applications to Kepler's problem and stellar dynamics

As we introduced above, the Kepler (or two-body) problem may be written in the Hamiltonian form (2.33) with Hamiltonian given by (2.34). The dynamics of this system exactly preserve H which represents total energy, as well as the angular momentum given by (unfortunately L is standard notation for both angular momentum and the Lagrangian)

$$L = q_1 p_2 - q_2 p_1. \tag{2.45}$$

In addition, the problem has rotational, time-reversal and scaling symmetries, which we shall consider presently. For the initial data used here (see Chapter 4 for more details) the exact solution is periodic and lies on an ellipse of eccentricity $e = 0.5$ with the origin at one focus.

In figure 2-1 we consider both the growth in the trajectory error (computed using the Euclidean norm in $\mathbb{R}^4$) and the conservation (or lack of it) of the Hamiltonian for our methods. The forward Euler method acts to increase the energy of the system leading to a monotonic growth in the Hamiltonian, a plot of the computed solution confirms this as the trajectory spirals outwards and so does not accurately reproduce the periodic solutions to this problem. In contrast the Hamiltonian whilst not constant for the symplectic methods exhibits a bounded error. Also, for this problem the symplectic methods have *linear* trajectory error growth as opposed to the *quadratic* growth observed in the non-symplectic method. The various peaks in these graphs correspond to close approaches between the two bodies.

These results are summarized in the following table, given in [84]. Note that both the symplectic Euler and Störmer-Verlet methods preserve the quadratic invariant of angular momentum exactly.

Method	Global error	Error in H	Error in L
FE	$\mathcal{O}(t^2h)$	$\mathcal{O}(th)$	$\mathcal{O}(th)$
SE	$\mathcal{O}(th)$	$\mathcal{O}(h)$	0
SV	$\mathcal{O}(th^2)$	$\mathcal{O}(h^2)$	0

See [84, 146, 36] for similar experiments and discussions, as well as proofs and explanations of the apparent superiority of symplectic over non-symplectic methods.

Note that methods based upon these symplectic ideas have been developed in the astrophysics community for the computation of the more complex N-body problem. They have been used to compute the evolution of the solar system for many millions of years, the excellent long time qualitative properties of the methods being of vital importance in this case. For example in [166] the evolution of the nine planets was computed for 100 million years using a time step of 7.2 days and the solar system was found to be chaotic. Similar results were obtained in [179] where the evolution of the five outer planets was computed for 1.1 billion years with a time step of 1 year.

Note that for problems with planetary near collisions, and hence large forces and velocities, some form of adaptivity often needs to be employed. We discuss this in Chapters 3 and 4.

2.3.4 Symmetries and conservation laws

In classical mechanics as well as many other branches of applied mathematics it is usual to discuss symmetries in association with conservation laws. This correspondence is basically due to Noether's theorem which we shall now discuss.

Before doing this however we shall look at a result which shall prove useful in later Chapters but which does not require an application of (nor fits into the framework of) Noether's theorem. Kepler's problem as defined earlier is clearly invariant under time translations and spatial rotations, with generators given by

$$\mathbf{X}_1 = \frac{\partial}{\partial t} \quad \text{and} \quad \mathbf{X}_2 = q_1\frac{\partial}{\partial q_2} - q_2\frac{\partial}{\partial q_1},$$

respectively. It is also invariant under the scaling transformation given by the generator

$$\mathbf{X}_3 = t\frac{\partial}{\partial t} + \frac{2}{3}q_1\frac{\partial}{\partial q_1} + \frac{2}{3}q_2\frac{\partial}{\partial q_2} - \frac{1}{3}p_1\frac{\partial}{\partial p_1} - \frac{1}{3}p_2\frac{\partial}{\partial p_2},$$

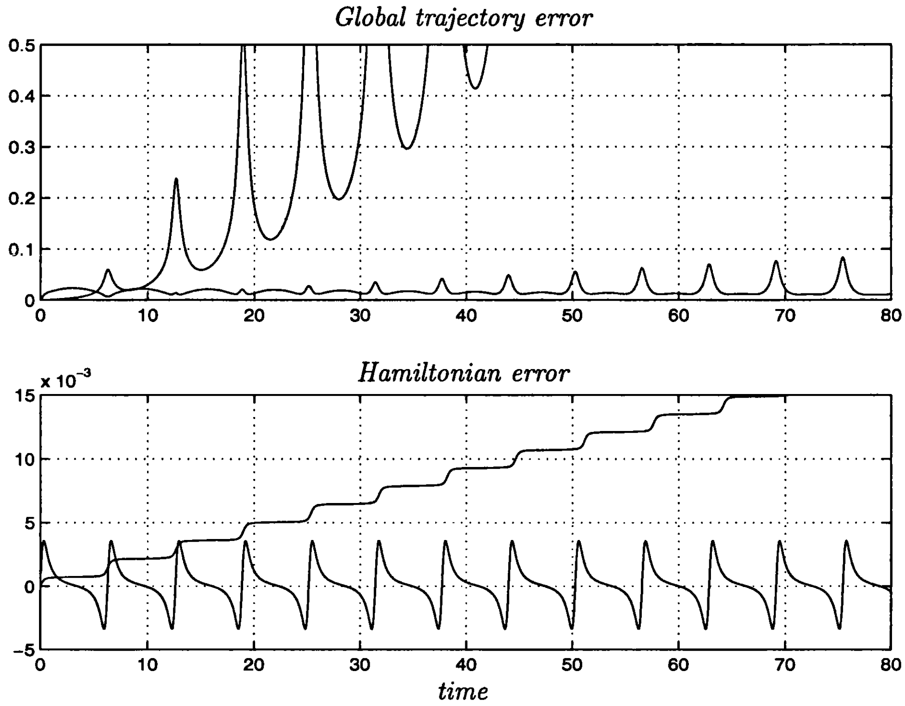


Figure 2-1: *Global trajectory error measured using the Euclidean norm in four-dimensional phase space, and error in the Hamiltonian for the Kepler problem with eccentricity $e = 0.5$. Methods shown are the forward Euler ($h = 0.0001$) lying in general above symplectic Euler ($h = 0.005$).*

where we have actually written down the generator corresponding to the prolongation of the transformation to the first jet space (this is, given how t and q transform, we can work out how p transforms by noting that $p = dq/dt$). The first two symmetries reflect the constancy and the rotationally invariant nature of the gravitational field between the bodies. The scaling invariance also has a physical meaning that is well known, writing the scaling generator in terms of polar coordinates we have

$$\mathbf{X}_3 = t\frac{\partial}{\partial t} + \frac{2}{3}r\frac{\partial}{\partial r}, \quad r = \sqrt{q_1^2 + q_2^2},$$

we know from earlier that this transformation will map one solution of the problem into another, i.e. given a solution with typical temporal and spatial length scales t and r, we immediately have another solution with corresponding length scales,

$$\tilde{t} = \lambda t, \quad \tilde{r} = \lambda^{2/3} r.$$

We can conclude the following which is generally known as Kepler's third law,

$$\frac{\tilde{t}^2}{\tilde{r}^3} = \frac{t^2}{r^3},$$

i.e. for solutions to Kepler's problem the square of the period is proportional to the cube of the distance from the origin. For a full and general discussion of the symmetries and conservation laws for Kepler's problem see [133].

Noether's theorem

We shall now briefly go through a basic discussion of Noether's theorem which makes concrete the relation between symmetries and conservation laws for systems which are derivable from a Lagrangian formulation. For additional details, including many generalization, to PDEs for example, see [77, 73, 128, 147].

We have actually already introduced an example of a correspondence between a symmetry and conservation law in this Chapter. Recall that we showed in (2.39) that if the Hamiltonian H (and equivalently the Lagrangian L) is explicitly independent of t, then H is a conserved quantity following the flow induced by the problem. However note that L (or H) being independent of t can be seen to be equivalent to the invariance of the Lagrangian L and the action functional under translations in time generated by ∂_t. Consider the action functional

$$\mathscr{L}[\mathrm{q}] = \int_{t_0}^{t_1} L(t, \mathrm{q}, \dot{\mathrm{q}})\, dt, \tag{2.46}$$

for arbitrary t_0 and t_1. A transformation of the form

$$\tilde{t} = \tilde{t}(t, \mathrm{q}; \lambda), \quad \tilde{\mathrm{q}} = \tilde{\mathrm{q}}(t, \mathrm{q}; \lambda), \tag{2.47}$$

shall be called a *variational symmetry* if is leaves (2.46) invariant, i.e. if

$$\mathscr{L}[\tilde{\mathrm{q}}] = \mathscr{L}[\mathrm{q}],$$

where q and $\tilde{\mathrm{q}}$ represent the curves,

$$\mathrm{q} \equiv \mathrm{q}(t), \quad t_0 \le t \le t_1, \quad \text{and} \quad \tilde{\mathrm{q}} \equiv \tilde{\mathrm{q}}(\tilde{t}), \quad \tilde{t}_0 \le \tilde{t} \le \tilde{t}_1.$$

For example if $L \equiv L(\mathrm{q}, \dot{\mathrm{q}})$ then under the transformation $\tilde{t} = t + \lambda$, $\tilde{\mathrm{q}} = \mathrm{q}$, that is

$$\tilde{\mathrm{q}}(\tilde{t}) = \mathrm{q}(t) = \mathrm{q}(\tilde{t} - \lambda),$$

we have

$$\int_{\tilde{t}_0}^{\tilde{t}_1} L\left(\tilde{\mathrm{q}}(\tilde{t}), \frac{d\tilde{\mathrm{q}}}{d\tilde{t}}(\tilde{t})\right) d\tilde{t} = \int_{t_0+\lambda}^{t_1+\lambda} L\left(\mathrm{q}(\tilde{t} - \lambda), \frac{d\mathrm{q}}{d\tilde{t}}(\tilde{t} - \lambda)\right) d\tilde{t} = \int_{t_0}^{t_1} L(\mathrm{q}(t), \dot{\mathrm{q}}(t))\, dt,$$

and so in this case the time-translation is a variational symmetry. Noether's theorem now basically goes on to say that if (2.47) is a variational symmetry for (2.46) then the quantity

$$\sum_{i=1}^{n} \eta_i L_{\dot{q}_i} + \xi \left(L - \sum_{i=1}^{n} \dot{q}_i L_{\dot{q}_i} \right), \tag{2.48}$$

is conserved along extrema of (2.46), i.e. as we saw earlier in this Chapter, along the solution to Hamilton's equations. We are using here the notation introduced in (2.3), i.e.

$$\xi(t,\mathbf{q}) = \left.\frac{\partial \tilde{t}}{\partial \lambda}\right|_{\lambda=0}, \quad \eta_i(t,\mathbf{q}) = \left.\frac{\partial \tilde{q}_i}{\partial \lambda}\right|_{\lambda=0}, \quad i = 1,\dots,n.$$

For example for the time-translational example above we have $\xi = 1$ and $\eta_i = 0$, and therefore Noether's theorem gives us the conserved quantity,

$$L - \sum_{i=1}^{n} \dot{q}_i L_{\dot{q}_i},$$

but notice that as claimed above this is precisely the negative Hamiltonian as defined in (2.32).

With a view to later Chapters we give one final example. The Lagrangian and action functional for the Kepler problem defined above are invariant under the transformation

$$\tilde{t} = t, \quad \tilde{q}_1 = q_1 \cos\lambda + q_2 \sin\lambda, \quad \tilde{q}_2 = -q_1 \sin\lambda + q_2 \cos\lambda.$$

Correspondingly we have $\xi = 0$, $\eta_1 = q_2$ and $\eta_2 = -q_1$. Noether's theorem now gives us the conserved quantity $\dot{q}_1 q_2 - \dot{q}_2 q_1$, which was defined as angular momentum in (2.45).

Similar results relating symmetries and conservation laws can be derived for the Hamiltonian formulation of problems. But now the relation is given in terms of group actions and *momentum maps* [114, 128, 8]. We do not go into detail here, except to say that this theory may be used, for example in [116], to demonstrate the preservation of conservation laws for certain discretization methods.

2.4 Hamiltonian PDEs

In moving from finite to infinite dimensions we essentially replace functions with functionals, gradients with variational derivatives and structure matrices with Hamiltonian operators. In particular, dependent variables are now functions $\mathbf{u}(\mathbf{x}, t)$, of space defined over some spatial domain, as well as time. Let M be the space of dependent and independent variables. Following [128] we shall use the notation $\mathscr{A}$ to denote the algebra of differential functions $P(\mathbf{x}, \mathbf{u}^{(n)}) \equiv P[\mathbf{u}]$ over M. Denote the quotient space under the image of total divergence by $\mathscr{F}$, that is the space of all functionals $\mathscr{P} = \int P\, dx$.

For functionals $\mathscr{G},\mathscr{H} \in \mathscr{F}$ define a Poisson bracket by

$$\{\mathscr{G},\mathscr{H}\} = \int \delta\mathscr{G} \cdot \mathscr{D}\,\delta\mathscr{H}\,dx, \tag{2.49}$$

where $\delta\mathscr{G} \in \mathscr{A}^q$ represents the variational derivative of $\mathscr{G}$ with respect to $\mathbf{u}$. The Hamiltonian operator $\mathscr{D} : \mathscr{A}^q \to \mathscr{A}^q$ is a linear operator (where q is the dimension of $\mathbf{u}$), which in analogy with the finite dimensional case must satisfy the following for (2.49) to define a Poisson bracket:

Skew-symmetry:

$$\{\mathscr{G},\mathscr{H}\} = -\{\mathscr{H},\mathscr{G}\}, \tag{2.50}$$

Jacobi identity:

$$\{\{\mathscr{G},\mathscr{H}\},\mathscr{K}\} + \{\{\mathscr{K},\mathscr{G}\},\mathscr{H}\} + \{\{\mathscr{H},\mathscr{K}\},\mathscr{G}\} = 0, \tag{2.51}$$

for all $\mathscr{F},\mathscr{G},\mathscr{K} \in \mathscr{F}$. Condition (2.50) holds if and only if $\mathscr{D}$ is skew-adjoint, the procedure for verifying (2.51) is complex and discussed at great length in [128]. A simple case which is straightforward to verify is when $\mathscr{D}$ is independent of $\mathbf{u}$ and its derivatives. Once we have defined our spaces and Poisson bracket, given a Hamiltonian functional $\mathscr{H}$, the infinite-dimensional form of Hamilton's equations can be written as the PDE

$$\frac{\partial \mathbf{u}}{\partial t} = \mathscr{D}\delta\mathscr{H}. \tag{2.52}$$

Example 2.7 Taking $\mathscr{D}$ to have the canonical form analogous to (2.42) (that is with the identity matrices replaced by identity operators) yields the canonical Poisson bracket, see (8.24).

Example 2.8 Taking $\mathscr{D}$ to be the operator $\partial/\partial x$ and Hamiltonian $\mathscr{H} = \int u^2\,dx/2$ yields the first order wave equation $u_t = u_x$.

2.4.1 The Korteweg-de Vries equation

An interesting example of a problem which may be written in an infinite dimensional Hamiltonian form is the Korteweg-de Vries (KdV) equation [59, 128],

$$u_t + uu_x + u_{xxx} = 0. \tag{2.53}$$

The Hamiltonian formulation for this problem is given by (2.52) with the Hamiltonian operator and functional given by

$$\mathscr{D} = \frac{\partial}{\partial x}, \qquad \mathscr{H} = \int \left(\frac{1}{2}u_x^2 - \frac{1}{6}u^3\right)\,dx. \tag{2.54}$$

Interestingly the KdV equation has another distinct Hamiltonian formulation, i.e. it can be written in the form (2.52) for a different $\mathscr{D}$ and $\mathscr{H}$. For more details of this *bi-Hamiltonian* system see [128], where it is shown that this property results in the recursive construction of an infinite hierarchy of symmetries and conservation laws.

Now for the Hamiltonian operator given in (2.54) skew-symmetry of the associated bracket (2.49) is simple to establish following an application of integration by parts. Since $\mathscr{D}$ in this case is independent of u or its derivatives this is actually sufficient to prove that the operator is Hamiltonian and the corresponding bracket Poisson [128]. However for the second formulation alluded to above, the Hamiltonian operator now depends on u and we must therefore explicitly verify the Jacobi identity. This can be an extremely complex and time-consuming undertaking, additional notation and theory is given in [128] which helps to simplify this procedure.

Given the large body of work on symplectic methods for Hamiltonian ODEs reviewed in the previous Sections, an attractive approach for numerically tackling Hamiltonian PDEs is first to semi-discretize in space to obtain a system of Hamiltonian ODEs which we can then symplectically integrate in time. There are many possible advantages in doing this, for example from the previous Sections it should be obvious that if we can preserve the Hamiltonian structure in passing from the infinite to finite spatial dimension and then exploit the advantages of symplectic methods we should achieve good long time results. There is also the possibility of obtaining numerical results which respect closely some of the conservation laws inherent in the problem. Lastly, and more practically, the use of (conservative) symplectic methods opens up the possibility of using explicit methods to efficiently integrate the typically stiff resulting systems of ODEs. For a review of methods for Hamiltonian problems and a discussion of there advantages see [119].

The dual formulation of the KdV equation can now be used to illustrate the following important point, in general (although there are a few exceptions, see the next Section on the Euler equations) semi-discretization in space fails to preserve any Hamiltonian structure present in a problem. However there are important exceptions, for example it is possible to do this when the Poisson bracket (2.49) takes the canonical form as in Example 2.7, and also when the Hamiltonian operator $\mathscr{D}$ is constant as in (2.54).

We have included this comment on discretizing Hamiltonian PDEs because in Chapters 6–8 we shall be considering a problem which has a non-canonical non-constant Hamiltonian formulation. Which, if possible, we would like to integrate whilst preserving as much as possible the properties of the system that can be associated with the Hamiltonian formulation. For brevity we shall not go through an explicit example of any discretization methods here, we simply refer to [119, 175] where spectral methods are employed to spatially truncate problems, and [93] where a Petrov-Galerkin (for background details see [185]) is used.

Linking in with earlier discussions, group invariant solutions can also be found for the KdV problem [128, 59]. For example the travelling wave solution invariant under the translation of independent variables (∂_x and ∂_t) leads to the classical one-soliton solutions. In addition consideration of the solutions associated with the Galilean boost group generated by the operator $t\partial_x + \partial_u$, leads to a reduction of the KdV equation to the famous class of ODE known as the *first Painlevé transcendent.* Similarly under the scaling transformation $x\partial_x + 3t\partial_t - 2u\partial_u$ we arrive at the so-called *second Painlevé transcendent.* Finally, it is known that the asymptotic behaviour of the KdV equation for large x values is given by a sequence of solitary waves moving in the same direction but apart from each other. It is shown in [16] that these asymptotics are actually self-similar in form.

2.4.2 The Euler equations

In Chapters 6, 7 and 8 we shall consider the semi-geostrophic equations — a fluids problem of interest to geophysicists. As motivation for this system we shall consider here the Euler equations describing inviscid incompressible fluid flow. This problem is of relevance here both because it has many properties in common with the semi-geostrophic equations, but also because it has been studied from a geometric integration viewpoint. This problem appears extensively in the mathematics literature as well as the meteorology literature where it is sometimes referred to as the baratropic vorticity equation.

The Euler equations of an inviscid, incompressible ideal fluid in a three-dimensional region Ω may be written

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u} = -\nabla p, \tag{2.55}$$

$$\nabla\cdot\mathbf{u} = 0, \tag{2.56}$$

where $\mathbf{x} = (x, y, z)^T$ are spatial coordinates, t is time, $\mathbf{u} = (u, v, w)^T$ is the velocity field, and p the pressure. A useful step is to rewrite these equations in terms of the vorticity $\omega = \nabla \times \mathbf{u}$. Taking the curl of (2.55) gives,

$$\frac{\partial \omega}{\partial t} = \omega\cdot\nabla\mathbf{u} - \mathbf{u}\cdot\nabla\omega. \tag{2.57}$$

In the two-dimensional case we have $\mathbf{u} = (u, v)^T$ a function of (x, y, t), then the vorticity is simply the scalar $w = v_x - u_y$ and (2.57) becomes

$$\frac{\partial \omega}{\partial t} = -u\omega_x - v\omega_y = \psi_y\omega_x - \psi_x\omega_y, \tag{2.58}$$

where the streamfunction Ψ satisfies $\Psi_x = v$, $\Psi_y = -u$. This may obviously be written as

$$\dot{\omega} = \frac{\partial(\omega, \Psi)}{\partial(x,y)}, \tag{2.59}$$

where the vorticity is related to the streamfunction through the relation $\omega = \nabla^2\Psi$.

This is another example of a problem which may be written in Hamiltonian form (see [128, 114]). The Hamiltonian functional being given by

$$\mathscr{H} = \frac{1}{2}\int d\mathbf{x}\,|\mathbf{u}|^2 = \frac{1}{2}\int d\mathbf{x}\,|\nabla\Psi|^2 = -\frac{1}{2}\int d\mathbf{x}\,\Psi\omega \tag{2.60}$$

with variational derivative

$$\frac{\delta\mathscr{H}}{\delta\omega} = -\Psi,$$

see [154] for details as well as a discussion on boundary conditions for this problem. The corresponding Poisson bracket is

$$\begin{aligned}\{\mathscr{F},\mathscr{G}\} &= -\int d\mathbf{x}\frac{\delta\mathscr{F}}{\delta\rho(\mathbf{x})}\left(\frac{\partial\left(\omega(\mathbf{x}), \frac{\delta\mathscr{G}}{\delta\rho(\mathbf{x})}\right)}{\partial(x,y)}\right)\\ &= \int d\mathbf{x}\,\omega(\mathbf{x})\left(\frac{\partial\left(\frac{\delta\mathscr{F}}{\delta\rho(\mathbf{x})}, \frac{\delta\mathscr{G}}{\delta\rho(\mathbf{x})}\right)}{\partial(x,y)}\right),\end{aligned} \tag{2.61}$$

and therefore, following the notation introduced earlier in this Section, we have

$$\mathscr{D}\bullet = -\frac{\partial(\omega, \bullet)}{\partial(x,y)}.$$

The Arakawa Jacobian

As a short aside we shall now mention one of the first examples of geometric integration for PDE problems. With the correct assumptions on boundary conditions the Euler equations can be shown to preserve the domain integrated vorticity, the enstrophy and the energy, given respectively by

$$\int \omega\, d\mathbf{x}, \quad \int \omega^2\, d\mathbf{x}, \quad \text{and} \quad \int |\nabla\Psi|^2\, d\mathbf{x}. \tag{2.62}$$

Although finite element discretizations can automatically obey discrete analogues of these conservation laws, the same is not in general true for finite difference spatial approximations. However, Arakawa [6] constructed finite difference analogues of the Jacobian operator in (2.59) which do preserve discrete analogues of (2.62). Numerical methods for solving the Euler equations may be subject to nonlinear instabilities, in particular aliasing errors where there is a spurious transfer of energy from large spatial scales to unresolvable smaller spatial scales. Since Arakawa's discretization preserves analogues of (2.62) it can be shown that this method actually prevents these nonlinear instabilities.

Sine bracket type truncation

For the KdV equation we discussed the problem with spatially discretizing a problem whilst retaining a Hamiltonian structure (as well as in some sense the entire underlying group structure), for general non-canonical Poisson brackets. Since the Euler equations have a very similar structure to the semi-geostrophic equations which shall be studied in detail in later Chapters we use this opportunity to demonstrate one of the exceptions where it is possible to obtain a Hamiltonian semi-discretization. For references to this material see [8, 118, 123, 182].

Many Hamiltonian systems turn out not to have a canonical formulation but often can naturally be written as Poisson systems, which generally arise as reductions from canonical formulations in more variables. The most common type are Lie-Poisson systems, these are distinguished by having a Poisson bracket which is linear in phase-space coordinates and reflects the symmetry of the problem, see Example 2.6. The approach discussed here hinges on the fact that the bracket (2.61) is of Lie-Poisson type, the (infinite-dimensional) Lie algebra suitable for use here is that of divergence free vector fields in our domain which are tangent to the boundary ($\mathfrak{X}_{\text{div}}(\Omega)$), this arises through the particle relabelling symmetry [129, 141] inherent in the problem. This is the Lie algebra associated with the Lie group of volume preserving diffeomorphisms of our region ($\text{Diff}_{\text{Vol}}(\Omega)$), see [8] for additional details on these objects.

We begin with our advection equation in the following form

$$\frac{\partial \omega}{\partial t} = \frac{\partial(\omega, \Psi)}{\partial(x, y)} \equiv \omega_x \Psi_y - \omega_y \Psi_x, \tag{2.63}$$

We assume (2π) periodic boundary conditions and therefore we evolve on the torus T^2. We now decompose our system into Fourier modes

$$\omega = \sum_{\mathbf{m}} \omega_{\mathbf{m}} e^{i(\mathbf{m},\mathbf{x})}$$

($(\mathbf{m}, \mathbf{x}) = \mathbf{m} \cdot \mathbf{x}$) and consider the resulting system of infinitely many ODEs describing their evolution in time. Decomposing Ψ in a similar manner and using the above relation between ω and Ψ we arrive at

$$\Psi_{\mathbf{m}} = \frac{\omega_{\mathbf{m}}}{|\mathbf{m}|^2}.$$

If we substitute into (2.63) we get after some cancellations

$$\dot{\omega}_{\mathbf{m}}(t) = \sum_{\mathbf{n}\neq \mathbf{0}} \frac{\mathbf{m}\times\mathbf{n}}{|\mathbf{n}|^2}\,\omega_{\mathbf{m}+\mathbf{n}}\,\omega_{-\mathbf{n}}$$

where $\mathbf{m}\times\mathbf{n} = m_1n_2 - m_2n_1$, and for real ω we have that $\omega_{-\mathbf{n}} = \omega^\star_{\mathbf{n}}$.

This turns out to be Lie-Poisson with the following

$$H(\omega) = \frac{1}{2}\sum_{\mathbf{n}\neq\mathbf{0}} \frac{\omega_{\mathbf{n}}\omega_{-\mathbf{n}}}{|\mathbf{n}|^2}, \quad \nabla H(\omega)_{\mathbf{k}} = \frac{\omega_{-\mathbf{k}}}{|\mathbf{k}|^2},$$

and Poisson structure (structure constants) defined by

$$J_{\mathbf{mn}}(\omega) = (\mathbf{m}\times\mathbf{n})\omega_{\mathbf{m}+\mathbf{n}} \equiv \sum_{\mathbf{k}} C^{\mathbf{k}}_{\mathbf{mn}}\omega_{\mathbf{k}}, \quad C^{\mathbf{k}}_{\mathbf{mn}} = (\mathbf{m}\times\mathbf{n})\delta_{\mathbf{m}+\mathbf{n}-\mathbf{k},\mathbf{0}},$$

where $\delta_{i,j}$ is the Kronecker delta. So that finally we may write our system in the form

$$\dot{\omega}_{\mathbf{m}} = \sum_{\mathbf{k},\mathbf{l},\mathbf{n}} a^{\mathbf{nl}}\, C^{\mathbf{k}}_{\mathbf{mn}}\,\omega_{\mathbf{k}}\,\omega_{\mathbf{l}},$$

where the metric (inverse inertia tensor) is given by

$$a^{\mathbf{nl}} = \frac{1}{|\mathbf{n}|^2}\delta_{\mathbf{n}+\mathbf{l},\mathbf{0}}.$$

The finite dimensional truncation of our bracket is now achieved by defining the new structure constants (N finite)

$$C^{\mathbf{k}}_{\mathbf{mn}} = \frac{N}{2\pi}\sin\left(\frac{2\pi}{N}(\mathbf{m}\times\mathbf{n})\right)\delta_{\mathbf{m}+\mathbf{n}-\mathbf{k},\mathbf{0}}$$

and so we have the finite-dimensional (Poisson) bracket

$$J_{\mathbf{mn}} = \frac{N}{2\pi}\sin\left(\frac{2\pi}{N}(\mathbf{m}\times\mathbf{n})\right)\omega_{\mathbf{m}+\mathbf{n}}\Big|_{\text{mod N}}.$$

Note that these structure constants are those for the algebra $\mathfrak{su}(N)^2$, and the consistency of this truncation relies on the fact that, in some sense, $SU(N) \to \text{Diff}_{\text{Vol}}(T^2)$ as $N \to \infty$.

[2] $SU(N)$ denotes the special unitary group in N dimensions. It is a subgroup of $GL(N,\mathbb{C})$ (the general linear group in N complex dimensions), the group of all complex $N\times N$ matrices with nonvanishing determinant. The unitary group $U(N)$ is the subgroup whose elements satisfy $U^{-1} = U^*$, its Lie algebra consists of all $N\times N$ anti-Hermitian matrices ($A^* = -A$). Its subgroup $SU(N)$ is the set of all matrices with unit determinant. Its Lie algebra $\mathfrak{su}(N)$ is the set of all anti-Hermitian matrices with zero trace, for some additional details see [149].

We then reduce indices modulo N to the periodic lattice $-M \leq m_1, m_2 \leq M$ where $N = 2M+1$. The Hamiltonian is truncated to a finite sum and we can now write down the Sine-Euler equations

$$\begin{aligned} \dot{\omega}_{\mathbf{m}} &= J_{\mathbf{mn}}(\omega)\nabla H_{\mathbf{n}}(\omega) \\ &= \sum_{\substack{n_1, n_2 = -M \\ \mathbf{n} \neq \mathbf{0}}}^{M} \frac{N}{2\pi|\mathbf{n}|^2} \sin\left(\frac{2\pi}{N}(\mathbf{m} \times \mathbf{n})\right) \omega_{\mathbf{m}+\mathbf{n}}\, \omega_{-\mathbf{n}}. \end{aligned} \tag{2.64}$$

In [118] a Poisson integrator for (2.64) is constructed which is explicit, fast, and preserves analogues of $N-1$ Casimirs (a special type of conserved quantity) to within round-off error. Baroclinic instability in a two-layer quasi-geostrophic type model was studies in [123] using these ideas, this shall be of relevance when we discuss the semi-geostrophic equations in Chapter 6–8.

2.5 Summary of Chapter

In this preliminary introductory Chapter some underlying geometric properties of differential equations have been discussed. The content has primarily focused on those properties which shall arise time and time again throughout this Thesis, for example symmetries and Hamiltonian structures. Some rigour and detail has been omitted in the interests of conciseness, however many references to the literature have been given.

Some examples of geometric integration methods designed to respect certain properties have been given and tested, mainly to provide motivation for the design of new methods. Minor changes to standard integration methods have been shown to result in vastly improved algorithms for certain Hamiltonian problems.

Although symmetries were discussed in a fair amount of detail in this Chapter, no symmetry respecting methods have been discussed. This shall be the topic of new work presented in the following three Chapters.

Following on from the comment made in Chapter 1 regarding the possibility of designing methods to preserve multiple geometric properties we have discussed the interesting links between symmetries, Lagrangian or Hamiltonian structures and conservation laws. This topic shall be revisited — primarily in Chapter 4.

Finally, with a view to Chapters 6, 7 and 8 the Euler equations and their properties were discussed. Since this system really does have some important similarities with the semi-geostrophic system, some of the current literature on geometric methods for this problem were reviewed.

Chapter 3

Scale invariant methods for ODEs

3.1 Overview of Chapter

In Section 2.2.3 we discussed ODEs invariant under a scaling transformation. We saw that the invariance property of a problem may lead to special solutions which we called self-similar. We noted that in many interesting situations (including especially those where standard numerical methods may experience difficulties) these self-similar solutions give important information about more general solutions to the problem.

In this Chapter we shall construct numerical methods which inherit the scale invariance property of a problem. What we exactly mean by this shall become apparent later on, but for now we simply note that for a numerical method using a fixed time step (and in Chapter 5, in addition a fixed spatial mesh) the method imposes an intrinsic length scale on the problem making it impossible for the method to admit scale invariant discrete solutions. We show below that this problem disappears when we consider a method which uses an adaptive time stepping strategy.

As was mentioned in Chapter 1 we shall think of an adaptive time stepping strategy in terms of a coordinate transformation between true or physical time t and an artificial or computational time τ. We discuss this in Section 3.2. We then go on to show that with the correct choice of time step our constructed method inherits the scaling invariance structure of an underlying problem, and in fact the operations of scaling and discretizing actually commute. We then go on to prove that the resulting method possesses a discrete self-similar solution which uniformly approximates the true self-similar solution as well as inheriting its stability, but with the freedom that these are not the only solutions admitted by the method. We therefore conclude that in many interesting situations we can be confident that our methods are accurately capturing the correct asymptotic behaviour of problems. The Chapter concludes with several examples of this occurring in practice.

Note that in this Chapter the problems we are considering do not necessarily also possess a Hamiltonian structure. However problems which do possess both a Hamiltonian and a scaling invariance structure shall be considered in Chapter 4.

Note that some of this work has appeared in the paper [33]. For brevity we shall use the following notation throughout this and later Chapters,

$$\lambda^{\boldsymbol{\alpha}}\mathbf{u} = (\lambda^{\alpha_1}u_1, \ldots, \lambda^{\alpha_N}u_N),$$

where $\mathbf{u} = (u_1, \ldots, u_N)^T$ and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$.

3.2 Time transformations and adaptivity

An adaptive choice of time step is a commonly used tool for reducing computational costs when numerically solving ODEs. The traditional form of adaptivity is to choose a step size so that some estimate of the local truncation error does not exceed a given tolerance over each discrete time interval. Whilst this approach is successfully used in many codes it departs from the spirit of geometric integration in that it is not attempting (necessarily) to respect any underlying qualitative structures present in a problem.

In this Chapter we shall look at an alternative method for performing adaptive time integrations which yields very nice results when applied to problems (seen in Section 2.2.3) which are invariant under a scaling transformation.

Suppose that the ODE problem we wish to solve is given by

$$\frac{d\mathbf{u}}{dt} = \mathbf{f}(\mathbf{u}), \tag{3.1}$$

Suppose further that the independent variable t is itself a function of a *fictive* computational variable τ such that

$$\frac{dt}{d\tau} = g(\mathbf{u}), \tag{3.2}$$

we shall call the transformation this induces between the variables t and τ a *Sundman transform*. Under this transformation equation (3.1) becomes

$$\frac{d\mathbf{u}}{d\tau} = g(\mathbf{u})\mathbf{f}(\mathbf{u}). \tag{3.3}$$

Ideally we would seek to choose the transformation such that the transformed equation (3.3) in the transformed variables (3.2) is in some way easier to solve either analytically or numerically. For example, singular solutions in the variable t should ideally become regular in the rescaled variable τ. Such a transformation may be made analytically and the resulting system (3.3) then solved numerically. We consider the advantages of this approach here. On the other hand, if we discretize (3.1) first and identify natural coordinates (3.2) *in the course of the calculation* then this process is at the heart of the adaptive approach. To see this we might suppose that in an adaptive procedure we find an approximate solution to problem (3.1) at a series of discrete times t_n. Typically we would want $t_{n+1} - t_n$ to be small if some measure of the solution (such as the estimated local truncation error) is large. If we take $g(\mathbf{u})^{-1}$ to be this measure and $\Delta\tau$ some prescribed constant, then a natural adaptive procedure for determining $t_{n+1} - t_n$ is to set

$$\Delta t_n \equiv t_{n+1} - t_n = \Delta\tau g(\mathbf{u}). \tag{3.4}$$

In the limit of small $\Delta\tau$ the equation (3.4) is simply a discretization of the transformation implied by (3.2). Note that the idea of starting from the time stepping strategy based upon (3.4) and using the fact that it is an approximation to (3.2) to study the dynamics of the numerical scheme is given in [79]. See [109] for a discussion and applications of the approach based upon (3.2).

We shall term the function g used in either (3.2) or (3.4) the temporal *monitor function* for the time transformation. This is since it is generally defined to give some monitor of the complexity of the problem being solved, and also to link the notation of this approach to that employed for performing the spatial adaptivity we shall see in Chapter 5.

In general the choice of the function g in either (3.2) or (3.4) is often somewhat arbitrary. However, for problems invariant under a scaling transformation we shall demonstrate a way of finding a suitable function $g(\mathbf{u})$. We show how the function g can be determined *a priori* for these differential equations by scaling arguments, making the condition that the transformed system (3.2), (3.3) should have the same scaling invariance as the original equation (3.1). In this case the resulting rescaled equations (3.2), (3.3) can be discretized in a way which inherits the original scaling invariance. We also demonstrate that for this class an *a posteriori* estimate of $g(\mathbf{u})$ in the formula (3.4) and related (and more sophisticated) adaptive formulae leads to essentially the same results, automatically identifying an appropriate fictive variable.

Both of the techniques of a priori and a posteriori scaling are effective in resolving singular structures especially when these structures have a self-similar form. We shall give evidence for this by looking at the solution of the Kepler problem under gravitational collapse.

In this Chapter we shall also consider both Runge-Kutta and multistep type discretizations of the transformed coupled system (3.2), (3.3), we shall show that the operations of discretization and scaling commute. We shall prove that the discretizations admit exact discrete self-similar solutions which *uniformly* approximate the true self-similar solutions over arbitrary long times and inherit the stability of these solutions.

3.3 Scaling invariance of the transformed problem

We firstly establish an important result relating the scaling properties of the original problem, the transformed problem, and the choice of the function g.

Lemma 3.1. *If the original ODE problem (3.1) is invariant under the scaling*

$$t \to \lambda t, \quad u_i \to \lambda^{\alpha_i} u_i, \quad i = 1, \dots, N, \quad \forall \lambda > 0, \tag{3.5}$$

then the transformed system (3.2), (3.3) will also be invariant (crucially without the need to scale the new variable τ*) if and only if (3.2) is invariant under (3.5).*

Proof. Notice that the scaling invariance of (3.1) is equivalent to the property that $\mathbf{f}(\lambda^{\boldsymbol{\alpha}}\mathbf{u}) = \lambda^{\boldsymbol{\alpha}-\mathbf{1}}\mathbf{f}(\mathbf{u})$, where $\mathbf{1} = (1, \dots, 1) \in \mathbb{R}^N$. Similarly, scaling invariance of (3.3) is equivalent to $(\mathbf{f}g)(\lambda^{\boldsymbol{\alpha}}\mathbf{u}) = \lambda^{\boldsymbol{\alpha}}(\mathbf{f}g)(\mathbf{u})$, recalling that we do not scale τ. Therefore given that (3.1) is invariant under (3.5), then (3.3) will also be invariant if and only if (3.2) is, i.e. equivalently $g(\lambda^{\boldsymbol{\alpha}}\mathbf{u}) = \lambda g(\mathbf{u})$. □

Substituting the rescaled variables into (3.2) we obtain

$$g(\lambda^{\boldsymbol{\alpha}}\mathbf{u}) = \lambda g(\mathbf{u}), \tag{3.6}$$

which on differentiating with respect to λ and setting $\lambda = 1$ gives

$$g = \sum_{i=1}^{N} \alpha_i u_i \frac{\partial g}{\partial u_i}. \tag{3.7}$$

We may now solve this to give a suitable function g, it shall transpire that effectively all solutions of (3.7) lead to essentially equivalent numerical schemes in terms of local error control and the admissibility of discrete self-similar solutions.

A very simple such choice for the function g is

$$g = u_j^{1/\alpha_j}, \tag{3.8}$$

where, for example, it may be wise to use j such that

$$|\alpha_j| = \max_i\{|\alpha_i| : i = 1, \dots, N\}.$$

For the blow-up problem (2.16) we have

$$g = -\frac{1}{3}ug_u \quad \Longrightarrow \quad g = u^{-3}.$$

We therefore transform (2.16) into the system

$$\frac{du}{d\tau} = u, \quad \frac{dt}{d\tau} = u^{-3}. \tag{3.9}$$

Similarly, for the Kepler problem (2.35), we have

$$g = \frac{2}{3}xg_x + \frac{2}{3}yg_y - \frac{1}{3}ug_u - \frac{1}{3}vg_v. \tag{3.10}$$

If we set $r = \sqrt{x^2 + y^2}$ and suppose that $g \equiv g(r)$ then (3.10) reduces to

$$g = \frac{2}{3}rg_r \quad \Longrightarrow \quad g = r^{3/2}.$$

We then have

$$\frac{d}{d\tau}\begin{pmatrix} x \\ y \end{pmatrix} = r^{3/2}\begin{pmatrix} u \\ v \end{pmatrix}, \quad \frac{d}{d\tau}\begin{pmatrix} u \\ v \end{pmatrix} = -r^{-3/2}\begin{pmatrix} x \\ y \end{pmatrix}, \quad \frac{dt}{d\tau} = r^{3/2}. \tag{3.11}$$

For convenience in further calculations we set

$$h_i(\mathbf{u}) = f_i(\mathbf{u})g(\mathbf{u}).$$

A nice feature of this rescaling is that it linearizes scalar equations. Suppose that u and f are both scalars. Now set $h(u) = g(u)f(u)$. We have that $h(\lambda^\alpha u) = \lambda^\alpha h(u)$. As this must be true for all scalars λ we deduce that

$$h(u) = \beta u,$$

for some appropriate β, and hence

$$\frac{du}{d\tau} = \beta u,$$

linearizing the differential equation, as we witnessed in (3.9).

Recall from Section 2.2.4 we defined what was meant by a self-similar solution to an ODE of the form (3.1) and also derived $u_i(t) = t^{\alpha_i}U_i$, $i = 1, \ldots, N$, for the form it takes, where the constants U_i are found after substitution into (3.1). In the rescaled system (3.2), (3.3) we now also have

$$\frac{dt}{d\tau} = g(\mathbf{u}).$$

Thus,

$$\frac{dt}{d\tau} = g(t^{\alpha_i}U_i) = tg(\mathbf{U}),$$

where the latter result follows from the scaling invariance of the function g. Without loss of generality we set $t = 1$ when $\tau = 0$. The self-similar solutions of (3.2), (3.3) therefore have the form

$$t = \exp(\mu\tau), \quad u_i = U_i \exp(\mu\alpha_i\tau), \quad i = 1, \ldots, N, \tag{3.12}$$

where we have that μ and $\mathbf{U}$ satisfy the algebraic system

$$\mu = g(\mathbf{U}), \quad \alpha_i\mu U_i = h_i(\mathbf{U}), \quad i = 1, \ldots, N. \tag{3.13}$$

Applying this procedure to problem (3.9) we have a self-similar solution of the form

$$u = U\exp(-\mu\tau/3), \quad t = \exp(\mu\tau),$$

which on substitution into (3.9) yields $\mu = -3$ and $U = (-3)^{-1/3}$. We therefore have the self-similar solution

$$u = (-3)^{-1/3}\exp(\tau), \quad t = \exp(-3\tau),$$

notice therefore that

$$u = (-3)^{-1/3}t^{-1/3},$$

compare this with (2.21).

For the rescaled Kepler problem (3.11) we have a self-similar solution of the form

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} X \\ Y \end{pmatrix}\exp(2\mu\tau/3), \quad \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} U \\ V \end{pmatrix}\exp(-\mu\tau/3), \quad t = \exp(\mu\tau).$$

Thus, X, Y, U, V, μ satisfy the algebraic system

$$\frac{2}{3}\mu X = (X^2+Y^2)^{3/4}U, \quad -\frac{1}{3}\mu U = -\frac{X}{(X^2+Y^2)^{3/4}},$$

with an almost identical equation for Y and V and

$$\mu = (X^2+Y^2)^{3/4}.$$

This has the solution

$$X = \left(\frac{9}{2}\right)^{1/3}, \quad U = \frac{2}{3}X, \quad \mu = \left(\frac{9}{2}\right)^{1/2}, \quad Y = V = 0. \tag{3.14}$$

Now, observing the further symmetries $t \to T - t, \quad \tau \to -\tau, \quad x \to x, \quad u \to -u$, we also have a self-similar solution of the form

$$x = \left(\frac{9}{2}\right)^{1/3} e^{-2\mu\tau/3}, \quad u = -\frac{2}{3}\left(\frac{9}{2}\right)^{1/3} e^{\mu\tau/3}, \quad t = T - e^{-\mu\tau}. \tag{3.15}$$

which describes a gravitational collapse at time T.

3.4 Discretizations of scale invariant ODEs

In this Section we consider what we achieve by discretizing the transformed system (3.2), (3.3) rather than the original problem (3.1). Although the transformed system is slightly larger than the original system (3.1) (if the original is N-dimensional the transformed system is $(N + 1)$-dimensional, which is of course only an enlargement relative to the size of N), solving the new system has distinct advantages over the original. Listed below are three advantages that shall be discussed throughout the remainder of this Chapter.

1. Multistep and Runge-Kutta discretizations of (3.2), (3.3) have relative local truncation errors which are independent of scale.

2. Any continuous self-similar solutions of the (original or transformed) problem are uniformly (in time) approximated by discrete self-similar solutions admitted by the numerical method. These discrete solutions also inherit the stability of the continuous ones.

3. Global properties of the continuous solution which are derived from the scaling invariance property (an example being Kepler's third law for planetary motion which we first met in Chapter 2) may be automatically inherited by the numerical method.

These properties all follow fundamentally from the fact that the two operations of linear scaling and discretization *commute* when applied to the problem (3.2), (3.3). This commutativity property is discussed now.

3.4.1 Commutativity of scaling and discretization

Recall that the transformed problem is to find $\mathbf{u}(\tau)$ and $t(\tau)$ such that

$$\frac{d\mathbf{u}}{d\tau} = \mathbf{f}(\mathbf{u})g(\mathbf{u}) \equiv \mathbf{h}(\mathbf{u}), \quad \frac{dt}{d\tau} = g(\mathbf{u}), \tag{3.16}$$

where, as a consequence of the choice of the function g

$$\mathbf{h}(\lambda^\alpha \mathbf{u}) = \lambda^\alpha \mathbf{h}(\mathbf{u}), \quad g(\lambda^\alpha \mathbf{u}) = \lambda g(\mathbf{u}). \tag{3.17}$$

Consider first a linear multistep [108, 100] discretization of (3.16) so that

$$\mathbf{u}_n \equiv (u_{1,n}, u_{2,n}, \dots, u_{N,n})^T \approx \mathbf{u}(n\Delta\tau), \quad t_n \approx t(n\Delta\tau),$$

are approximations to $\mathbf{u}$ and t at the n-th discrete time level with $\Delta\tau$ fixed. For appropriate $\{\beta_j\}$ and $\{\gamma_j\}$ a linear multistep method with l steps takes the form

$$\sum_{j=0}^{l} \beta_j \mathbf{u}_{n+j} = \Delta\tau \sum_{j=0}^{l} \gamma_j \mathbf{h}(\mathbf{u}_{n+j}), \quad \sum_{j=0}^{l} \beta_j t_{n+j} = \Delta\tau \sum_{j=0}^{l} \gamma_j g(\mathbf{u}_{n+j}). \tag{3.18}$$

We now establish the important result that the linear multistep method inherits exactly the same scaling invariance property as the original system.

Lemma 3.2. *With the correct choice of function g such that (3.17) holds, the linear multistep method (3.18) with a fixed step size $\Delta\tau$ inherits exactly the same scaling invariance as the original system (3.1). That is, if $(t_n, \mathbf{u}_n)$ is a solution of (3.18) for $n = 0, 1, \dots$, then so is the rescaled solution $(\lambda t_n, \lambda^\alpha \mathbf{u}_n)$.*

Proof. Substituting the sequence $(\lambda t_n, \lambda^\alpha \mathbf{u}_n)$ into the method (3.18), and exploiting the scaling structure (3.2) of the functions $\mathbf{h}$ and g we have

$$\sum_{j=0}^{l} \beta_j \lambda^\alpha \mathbf{u}_{n+j} = \Delta\tau \sum_{j=0}^{l} \gamma_j \lambda^\alpha \mathbf{h}(\mathbf{u}_{n+j}), \quad \sum_{j=0}^{l} \beta_j \lambda t_{n+j} = \Delta\tau \sum_{j=0}^{l} \gamma_j \lambda g(\mathbf{u}_{n+j}).$$

On cancelling the positive constants λ^α and λ we see immediately that this rescaled algebraic system is equivalent to the original (3.18), hence we have the desired result that if $(t_n, \mathbf{u}_n)$ is a solution of (3.18) for $n = 0, 1, \dots$, then so is the rescaled solution $(\lambda t_n, \lambda^\alpha \mathbf{u}_n)$. □

This is an important result and the remaining results in this Chapter depend crucially upon it. We are saying that the two operations of scaling and discretization when applied to a differential equation system *commute.* This result is simply not true when applied to a non-adaptive fixed time step method. It is implicit in the rescaling that the time step Δt can be rescaled along with the solution. If this is not done then rescaling in time is simply not possible.

In [122] the property of a method producing the same integrator in different coordinate systems in called *covariance.* It is proved that, in general, discrete approximations will only be covariant with respect to the group of affine transformations, i.e. to scalings and translations. This again shows the special rôle scalings can play in the analysis of both continuous problems and discrete numerical methods.

If instead we use an s-stage Runge-Kutta [108, 100] discretisation of the system as given by

$$\begin{aligned} \begin{pmatrix} \mathbf{u} \\ t \end{pmatrix}_{n+1} &= \begin{pmatrix} \mathbf{u} \\ t \end{pmatrix}_{n} + \Delta\tau \sum_i b_i \mathbf{k}_i, \\ \mathbf{k}_i &= \begin{pmatrix} \mathbf{h} \\ g \end{pmatrix} \left(\begin{pmatrix} \mathbf{u} \\ t \end{pmatrix}_{n} + \Delta\tau \sum_j a_{i,j} \mathbf{k}_j \right), \quad i = 1, \ldots, s, \end{aligned} \tag{3.19}$$

then we may establish the following analogous result to Lemma 3.2.

Lemma 3.3. *With the correct choice of function g such that (3.17) holds, the Runge-Kutta method (3.19) with a fixed step size $\Delta\tau$ inherits exactly the same scaling invariance as the original system (3.1). That is, if $(t_n, \mathbf{u}_n)$ is a solution of (3.18) for $n = 0, 1, \ldots,$ then so is the rescaled solution $(\lambda t_n, \lambda^{\alpha} \mathbf{u}_n)$.*

Proof. From the scaling properties of $\mathbf{h}$ and g it is straightforward to see that the non-linear system defining the $\mathbf{k}_i$ has the admissible solution $(\lambda^{\alpha_1} k_{i,1}, \ldots, \lambda^{\alpha_N} k_{i,N}, \lambda k_{i,N+1})$ given data $(\lambda t_n, \lambda^{\alpha} \mathbf{u}_n)$, whenever $(k_{i,1}, \ldots, k_{i,N}, k_{i,N+1})$ is a solution with data $(t_n, \mathbf{u}_n)$. Given this, it is then immediate upon substituting the sequence $(\lambda t_n, \lambda^{\alpha} \mathbf{u}_n)$ into the method (3.19), and exploiting the scaling structure (3.2) that the rescaled algebraic system is equivalent to (3.19). The desired result therefore follows. □

Certain global properties of a continuous solution derived from a scaling invariance property can be inherited exactly by a discrete numerical solution. This result follows immediately from the corresponding scaling invariance of the numerical method. For example, suppose that $(x(t), y(t))$ is a periodic solution of Kepler's problem (2.35) with period T, then Kepler's third law [73] states that the square of the period is proportional to the cube of the major axis of the orbit. This is equivalent to saying that given the above periodic solution, the orbit $(\lambda^2 x(t), \lambda^2 y(t))$ is also a solution with period $\lambda^3 T$, and this is precisely a statement of scaling invariance. Now suppose that (x_n, y_n) is a discrete periodic solution with period T_Δ obtained from a scaling invariant numerical method, then immediately from this invariance, $(\lambda^2 x_n, \lambda^2 y_n)$ is also a discrete periodic solution with period $\lambda^3 T_\Delta$, and hence Kepler's law also holds in the discrete case. This observation demonstrates some of the possibilities of the adaptive approach considered here. However note that typically we would not expect the existence of discrete periodic solutions for general methods, although the similar concept of closed invariant curves may well be present. See [164], as well as [163] for an analysis of such matters in the context of the adaptivity procedure employed in this Chapter.

3.4.2 Scaling of local truncation errors

In this Section we show that the relative local truncation error of linear multistep and Runge-Kutta discretization is independent of scale. As the errors in such discretizations involve either higher derivatives of **u** or **h** we must firstly determine the way in which these derivatives scale themselves.

Lemma 3.4. *The m-th derivatives $u_i^{(m)}$ of the components of* **u** *with respect to τ satisfy the equations*

$$u_i^{(m)} = h_i^{(m-1)}(\mathbf{u}), \quad m \in \mathbb{N}, \tag{3.20}$$

where under the rescaling (3.5) the functions $h_i^{(m)}$, as defined in (3.16), obey

$$h_i^{(m-1)}(\lambda^\alpha \mathbf{u}) = \lambda^{\alpha_i} h_i^{(m-1)}(\mathbf{u}), \quad m \in \mathbb{N}. \tag{3.21}$$

Here we have that $h_i^{(0)} \equiv h_i$ and we see that each of the subsequent functions $h_i^{(m)}$ scales in an identical manner to h_i.

Proof. The proof of this result is by mathematical induction on m. Clearly from the definition (3.16) and the result (3.17) the result is true for $m = 1$. Now on differentiation of (3.20) with respect to τ we have, following an application of the chain rule and using definition (3.16),

$$h_i^{(m)}(\mathbf{u}) = u_i^{(m+1)} = \frac{du_i^{(m)}}{d\tau} = \sum_k \frac{\partial h_i^{(m-1)}}{\partial u_k}\frac{du_k}{d\tau} = \sum_k \frac{\partial h_i^{(m-1)}}{\partial u_k} h_k.$$

Therefore we must have

$$h_i^{(m)}(\lambda^\alpha \mathbf{u}) = \sum_k \frac{\partial h_i^{(m-1)}}{\partial u_k}(\lambda^\alpha \mathbf{u}) h_k(\lambda^\alpha \mathbf{u}). \tag{3.22}$$

But differentiating the identity (3.21) with respect to u_k gives

$$\frac{\partial h_i^{(m-1)}}{\partial u_k}(\lambda^\alpha \mathbf{u}) = \lambda^{\alpha_i - \alpha_k} \frac{\partial h_i^{(m-1)}}{\partial u_k}(\mathbf{u}). \tag{3.23}$$

Hence, on substitution of this result as well as (3.17) into (3.22) we have

$$h_i^{(m)}(\lambda^{\alpha}\mathbf{u}) = \sum_k \lambda^{\alpha_i - \alpha_k} \frac{\partial h_i^{(m-1)}}{\partial u_k}(\mathbf{u})\lambda^{\alpha_k} h_k(\mathbf{u}) = \lambda^{\alpha_i} h_i^{(m)}(\mathbf{u}),$$

and the result follows. □

Thus the derivatives of u_i with respect to τ scale in exactly the same manner as u_i itself. In particular, the *relative derivatives*

$$v_i^{(m)} := \frac{u_i^{(m)}}{u_i}, \quad i = 1, \ldots, N, \quad m \in \mathbb{N},$$

are invariant under rescalings.

Now consider the linear multistep method (3.18) of order p. From the standard theory of such methods [86, 100, 108], it follows that the *local truncation error* $\mathbf{e}$ is given by

$$\mathbf{e} = C(\Delta\tau)^{p+1}\mathbf{u}^{(p+1)} + \mathcal{O}(\Delta\tau^{p+2}),$$

where C is some constant. To leading order the relative local error contributions are given by

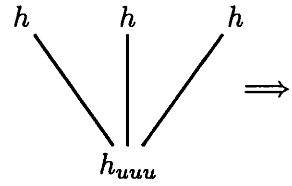
$$\frac{e_i}{u_i} = C(\Delta\tau)^{p+1}\frac{u_i^{(p+1)}}{u_i}.$$

From the above reasoning we see that these are invariant under rescaling. Thus if we define the relative local truncation error to be

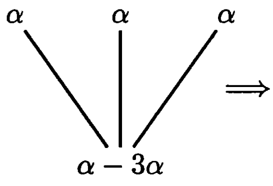
$$E = \max_i(e_i/u_i), \tag{3.24}$$

then for a method with constant step size $\Delta\tau$ this error is also invariant under rescaling.

Now for Runge-Kutta methods the local truncation error $\mathbf{e}$ is given as a linear sum of so-called *elementary differentials*, for details see [38, 86, 100, 108]. It follows that we will have a similar result regarding the scaling invariance of the relative local truncation error for Runge-Kutta methods if we can establish that the elementary differentials scale as $\mathbf{u}$ does, just as we proved for the $\mathbf{u}^{(m)}$ in Lemma (3.4). Now elementary differentials can be characterized using the graph-theoretical concept of *rooted trees*, we sacrifice rigour here in order to avoid the need to introduce additional notation and theory, we simply sketch some basic ideas for scalar systems and refer the reader to the previously quoted references for additional background detail. For the tree $\tilde{t}$ given below for example, we have replaced the vertices of the tree with a corresponding $h_{uu\ldots u}$ where the order of the derivative is given by the number of children (in the direction South–North) each vertex possesses.



$\Longrightarrow \; H = h_{uuu}hhh,$



$$\Longrightarrow \; \mathrm{sc}(H) = \alpha - 3\alpha \\ +\alpha + \alpha + \alpha \\ = \alpha$$

As we see the elementary differential for this tree is then simply given by the product of the terms present. Hence here we have $H[\tilde{t}] = h_{uuu}h^3$. To the right of the figure we have included the same tree, this time with the vertices replaced by the scaling powers of the corresponding terms, c.f. (3.23). The scaling of the elementary differential, $\mathrm{sc}(H)$, is then given by the sum of these individual scaling powers. As we see these powers cancel down so that the elementary differential here simply scales as h, or equivalently u. Since every rooted tree is made up of sub-trees of this type, with varying numbers of children, an induction argument establishes the following result, for which the extension to systems is straightforward.

Lemma 3.5. *For any rooted tree $\tilde{t}$, the corresponding elementary differential $H[\tilde{t}]$ scales under (3.5) as $H[\tilde{t}](\lambda^{\alpha}\mathbf{u}) = \lambda^{\alpha}H[\tilde{t}](\mathbf{u})$.*

Lemmas 3.4 and 3.5 and the earlier comments in this Section allow us to immediately establish the following result.

Theorem 3.1. *For both linear multistep and Runge-Kutta methods with local truncation errors $\mathbf{e}$, the relative local truncation errors as defined as $E = \max(e_i/u_i)$ are invariant under the rescaling (3.5).*

It is hard to overemphasize the importance of this result. For example if a singularity forms in the solution which is progressively described in terms of the action of the scaling group, then the resulting adaptive numerical method will continue to compute this solution with no overall loss of relative accuracy. We shall see this clearly demonstrated when we look at the problem of gravitational collapse. Observe, however, that an overall error defined by $||\mathbf{e}||$ is not invariant under rescaling, and indeed has no nice scaling properties.

3.4.3 Admittance of discrete self-similar solutions

A key feature of the linear multistep and Runge-Kutta discretizations of the transformed system, is that they are able to approximate the self-similar manifold for all time to a constant discretization error which does not grow with increasing τ even if the solution exhibits complex behaviour, for example if it forms a singularity.

To state the main result of this Chapter consider the scale invariant problem

$$\dot{u}_i = f_i(\mathbf{u}),$$

with $i = 1, \dots, N$. The self-similar solution for this problem as was already described in (2.19) is given by

$$u_i(t) = t^{\alpha_i} U_i. \tag{3.25}$$

The constants U_i then satisfy the *algebraic* system

$$\alpha_i U_i = f_i(\mathbf{U}), \quad \text{where} \quad \mathbf{U} = (U_1, U_2, \dots, U_N)^T. \tag{3.26}$$

From (3.12) we have that the general self-similar solution to (3.2), (3.3) is given by

$$u_i = U_i e^{\mu\alpha_i\tau}, \quad t = e^{\mu\tau}, \tag{3.27}$$

where, if without loss of generality we set $t = 1$ at $\tau = 0$,

$$\mu = g(\mathbf{u}).$$

Consider using the linear multistep method (3.18) to solve (3.2), (3.3). The following theorem is an immediate consequence of the scaling invariance of the multistep method.

Theorem 3.2. *For a consistent (p-th order) and zero stable linear multistep method with l steps of the form (3.18), with $u_{i,n}$ and t_n the discrete approximants to u_i and t at the n-th time level for problem (3.2), (3.3) with $(\mathbf{h}, g) \in C^p(\mathbb{R}^{N+1}, \mathbb{R}^{N+1})$. For suitable constants $\hat{U}_i$ and z satisfying an appropriate nonlinear equation, which for now is assumed to be soluble, there exists a discrete self-similar solution of the form*

$$t_n = z^n, \quad u_{i,n} = z^{\alpha_i n}\hat{U}_i, \quad i = 1, \dots, N, \tag{3.28}$$

which is valid for all $n \geq 0$. For this solution it is immediate that

$$u_{i,n} = t_n^{\alpha_i}\hat{U}_i. \tag{3.29}$$

Proof. Observe that from (3.17),

$$h_i(z^{\alpha(n+j)}\hat{\mathbf{U}}) = z^{(n+j)\alpha_i}h_i(\hat{\mathbf{U}}), \quad g(z^{\alpha(n+j)}\hat{\mathbf{U}}) = z^{(n+j)}g(\hat{\mathbf{U}}).$$

Thus the expression (3.28) satisfies the linear multistep method (3.18) provided that for $i = 1, \dots, N$ we have

$$\sum_j \beta_j z^{(n+j)\alpha_i} \hat{U}_i = \Delta\tau \sum_j \gamma_j z^{(n+j)\alpha_i} h_i(\hat{\mathbf{U}}),$$

and

$$\sum_j \beta_j z^{(n+j)} = \Delta\tau \sum_j \gamma_j z^{(n+j)} g(\hat{\mathbf{U}}).$$

Most significantly, this system is satisfied for all values of n provided that $\hat{\mathbf{U}}$ and z satisfy the following nonlinear algebraic system

$$\sum_j \beta_j z^{j\alpha_i} \hat{U}_i - \Delta\tau \sum_j \gamma_j z^{j\alpha_i} h_i(\hat{\mathbf{U}}) = 0, \tag{3.30}$$

and

$$\sum_j \beta_j z^j - \Delta\tau \sum_j \gamma_j z^j g(\hat{\mathbf{U}}) = 0. \tag{3.31}$$

□

We now proceed to show that the nonlinear algebraic equations (3.30), (3.31) yielding $\hat{U}_i$ and z, have a solution in a sense close to the underlying self-similar solution.

Theorem 3.3. *Suppose that the self-similar solution (3.27) to problem (3.2), (3.3), with $(\mathbf{h}, g) \in C^p(\mathbb{R}^{N+1}, \mathbb{R}^{N+1})$, exists and is locally unique. Suppose furthermore that the linear multistep method is zero stable and has a local truncation error which is of order $\mathcal{O}\left(\Delta\tau^{p+1}\right)$. Then for sufficiently small $\Delta\tau$ the system (3.30), (3.31) has a locally unique solution satisfying*

$$\hat{U}_i = U_i + \mathcal{O}\left(\Delta\tau^p\right), \tag{3.32}$$

$$z = e^{\mu\Delta\tau} + \mathcal{O}\left(\Delta\tau^{p+1}\right). \tag{3.33}$$

Proof. Consider the true self-similar solution given by (3.27). If we set $x = \exp(\mu\Delta\tau)$ we have immediately that, for $i = 1, \dots, N$

$$u_i(j\Delta\tau) = x^{j\alpha_i} U_i, \quad t(j\Delta\tau) = x^j.$$

Hence, substituting into the linear multistep scheme (3.18), exploiting the scaling properties of the scheme and looking at the local truncation error we have

$$\sum_j \beta_j x^{j\alpha_i} U_i - \Delta\tau \sum_j \gamma_j x^{j\alpha_i} h_i(\mathbf{U}) = \mathcal{O}\left(\Delta\tau^{p+1}\right), \tag{3.34}$$

$$\sum_j \beta_j x^j - \Delta\tau \sum_j \gamma_j x^j g(\mathbf{U}) = \mathcal{O}\left(\Delta\tau^{p+1}\right). \tag{3.35}$$

Thus $(\mathbf{U}, x)$ is an approximate solution of (3.30), (3.31) with a residual of $\mathcal{O}\left(\Delta\tau^{p+1}\right)$. Now, for (3.27) to be a true self-similar solution the values of U_i and μ must satisfy the nonlinear algebraic equations (c.f. (3.13))

$$\mu\alpha_i U_i - h_i(\mathbf{U}) = 0, \quad \mu - g(\mathbf{U}) = 0. \tag{3.36}$$

Thus the continuous self-similar solution is locally unique provided that the Jacobian matrix for this system given by

$$J = \left[\begin{array}{c|c} \alpha_i U_i & \mu\alpha_i\delta_{i,j} - \partial h_i/\partial U_j \\ \hline 1 & -\partial g/\partial U_j \end{array}\right] \tag{3.37}$$

(where $\delta_{i,j}$ is the Kronecker delta and its usage here should be obvious) has a bounded inverse. Now consider the Jacobian of the operator defined by the left hand side of the equations (3.34), (3.35). Define the *characteristic polynomials* of (3.18) by

$$\rho(\zeta) = \sum_j \beta_j \zeta^j, \quad \sigma(\zeta) = \sum_j \gamma_j \zeta^j,$$

for $\zeta \in \mathbb{C}$, and note that for consistency and zero-stability [108] of the linear multistep method we must have

$$\rho(1) = 0, \quad \sigma(1) = \rho'(1) \neq 0.$$

Furthermore, observe from the definition of x that for small $\Delta\tau$ we have

$$x = 1 + \mu\Delta\tau + \mathcal{O}\left(\Delta\tau^2\right),$$

so that

$$\rho(x^{\alpha_i}) = \rho'(1)\Delta\tau\alpha_i\mu + \mathcal{O}\left(\Delta\tau^2\right).$$

Combining these results, it follows immediately that the Jacobian of the nonlinear system (3.34), (3.35) is given by J_Δ, where

$$J_\Delta = \rho'(1) \left[\begin{array}{c|c} \alpha_i U_i + \mathcal{O}(\Delta\tau) & \Delta\tau(\mu\alpha_i\delta_{i,j} - \partial h_i/\partial U_j) + \mathcal{O}(\Delta\tau^2) \\ \hline 1 + \mathcal{O}(\Delta\tau) & -\Delta\tau\partial g/\partial U_j + \mathcal{O}(\Delta\tau^2) \end{array}\right].$$

Now suppose that we rescale the system (3.34), (3.35) by setting $\hat{U}_i = U_i + \delta_i/\Delta\tau$, $z = x + \delta z$. The Jacobian of the resulting rescaling acting on the vector $[\delta z, \delta_i]$ is given by

$$\hat{J}_\Delta = \rho'(1) \left[\begin{array}{c|c} \alpha_i U_i & \mu\alpha_i\delta_{i,j} - \partial h_i/\partial U_j \\ \hline 1 & -\partial g/\partial U_j \end{array} \right] + \mathcal{O}(\Delta\tau)$$

$$= \rho'(1)J + \mathcal{O}(\Delta\tau),$$

where J is the matrix given in (3.37). Now, as J has a bounded inverse, it follows that in the limit of small $\Delta\tau$ the matrix $\hat{J}_\Delta$ also has a bounded inverse, see [74]. Therefore since this Jacobian has a bounded inverse and the left hand side of (3.34), (3.35) is satisfied with zero residual at $(\delta z = 0, \delta_i = 0; \Delta\tau = 0)$ we can apply the implicit function theorem [3] to find a solution to (3.30), (3.31) for sufficiently small $\Delta\tau$. In addition

$$[\delta z, \delta_i] = \mathcal{O}(\Delta\tau^{p+1}).$$

Thus

$$z = x + \mathcal{O}\left(\Delta\tau^{p+1}\right), \quad \hat{U}_i = U_i + \mathcal{O}\left(\Delta\tau^{p}\right),$$

and the theorem follows. □

We now establish analogous results to Theorems 3.2 and 3.3 in the context of Runge-Kutta methods.

Theorem 3.4. *For a consistent s-stage Runge-Kutta method of the form (3.19), with $u_{i,n}$ and t_n the discrete approximants to u_i and t at the n-th time level for problem (3.2), (3.3) with $(\mathbf{h}, g) \in C^p(\mathbb{R}^{N+1}, \mathbb{R}^{N+1})$. For suitable constants $\hat{U}_i$ and z satisfying an appropriate nonlinear equation, which for now is assumed to be soluble, there exists a discrete self-similar solution of the form*

$$t_n = z^n, \quad u_{i,n} = z^{\alpha_i n}\hat{U}_i, \quad i = 1, \dots, N, \tag{3.38}$$

which is valid for all $n \geq 0$. For this solution it is immediate that

$$u_{i,n} = t_n^{\alpha_i}\hat{U}_i. \tag{3.39}$$

Proof. In a similar manner to the proof of Theorem 3.2, the expression (3.38) satisfies the Runge-Kutta method (3.19) provided that for $i = 1, \dots, N$ and for all values of n

we have that $\hat{\mathbf{U}}$ and z as well as $\hat{\mathbf{k}}_i$ satisfy the following nonlinear algebraic system

$$\begin{aligned} &\begin{pmatrix} \hat{\mathbf{U}} z^{\alpha} \\ z \end{pmatrix} - \begin{pmatrix} \hat{\mathbf{U}} \\ 1 \end{pmatrix} - \Delta\tau \sum_i b_i \hat{\mathbf{k}}_i = 0, \\ &\hat{\mathbf{k}}_i = \begin{pmatrix} \mathbf{h} \\ g \end{pmatrix} \left(\begin{pmatrix} \hat{\mathbf{U}} \\ 1 \end{pmatrix} + \Delta\tau \sum_j a_{i,j} \hat{\mathbf{k}}_j \right), \quad i = 1, \ldots, s. \end{aligned} \tag{3.40}$$

□

As before we now proceed to demonstrate that the nonlinear algebraic equations (3.40) yielding $\hat{U}_i$, z and $\hat{\mathbf{k}}_i$, have a solution in a sense close to the underlying self-similar solution.

Theorem 3.5. *Suppose that the self-similar solution (3.27) to problem (3.2), (3.3), with* $(\mathbf{h}, g) \in C^p(\mathbb{R}^{N+1}, \mathbb{R}^{N+1})$*, exists and is locally unique. Suppose furthermore that the Runge-Kutta method has a local truncation error which is of order* $\mathcal{O}\left(\Delta\tau^{p+1}\right)$*, (and that the problem being integrated results in a soluble nonlinear system for the* $\mathbf{k}_i$*). Then for sufficiently small* $\Delta\tau$ *the system (3.40) has a locally unique solution satisfying*

$$\hat{U}_i = U_i + \mathcal{O}\left(\Delta\tau^p\right), \tag{3.41}$$

$$z = e^{\mu\Delta\tau} + \mathcal{O}\left(\Delta\tau^{p+1}\right). \tag{3.42}$$

Proof. The proof of this result follows very closely that of Theorem 3.3. Consider the solution given by (3.27), for this to be a true self-similar solution the values of U_i and μ must satisfy the set of nonlinear algebraic equations (3.36), and for it to be locally unique the Jacobian matrix (3.37) must has a bounded inverse. Again set $x = \exp(\mu\Delta\tau)$, substituting the true self-similar solution into the Runge-Kutta method (3.19), exploiting the scaling properties of the scheme and looking at the local truncation error gives that $(\mathbf{U}, x)$, (and their corresponding $\mathbf{k}_i$), is an approximate solution of (3.40) with a residual of order $\mathcal{O}\left(\Delta\tau^{p+1}\right)$. Noting that in the limit of small $\Delta\tau$ the nonlinear system yielding the $\mathbf{k}_i$ in (3.19) is soluble with

$$\operatorname{diag}(x^{-\alpha_1 n}, x^{-\alpha_2 n}, \ldots, x^{-\alpha_N n}, x^{-n})\mathbf{k}_i = \begin{pmatrix} \mathbf{h} \\ g \end{pmatrix} \begin{pmatrix} \mathbf{U} \\ 1 \end{pmatrix} + \mathcal{O}(\Delta\tau),$$

and that consistency of the method implies that $\sum b_i = 1$, the Jacobian of the operator defined by the continuous self-similar solution substituted into the remainder of (3.19) is given by

$$J_\Delta = \left[\begin{array}{c|c} \alpha_i U_i + \mathcal{O}(\Delta\tau) & \Delta\tau(\mu\alpha_i\delta_{i,j} - \partial h_i/\partial U_j) + \mathcal{O}(\Delta\tau^2) \\ \hline 1 & -\Delta\tau\partial g/\partial U_j + \mathcal{O}(\Delta\tau^2) \end{array}\right].$$

The proof now follows virtually identically that of Theorem 3.3, and we have as required that a solution to (3.40) exists, for sufficiently small $\Delta\tau$, with

$$z = x + \mathcal{O}\left(\Delta\tau^{p+1}\right), \quad \hat{U}_i = U_i + \mathcal{O}\left(\Delta\tau^p\right).$$

□

Note that we can use the time-translational symmetry generally present in the problems we are considering to generalize the above results to the case where $t_n = a + bz^{\alpha_0 n}$. In general $b = 1$ but a time-reversal symmetry implies that we may also take $b = -1$, this is especially useful in problems with finite-time singularities, see the examples in Chapter 2 as well as those later in this Chapter.

Corollary 3.1. *With the assumptions of Theorem 3.2 or Theorem 3.4 as necessary, for all time it follows that for both linear multistep and Runge-Kutta methods*

$$\frac{u_{i,n}}{t_n^{\alpha_i}} = \frac{u_i(t_n)}{t_n^{\alpha_i}} + \mathcal{O}\left(\Delta\tau^p\right), \tag{3.43}$$

where the implied constant in the $\mathcal{O}(\cdot)$ term does not depend upon either the solution or on t.

Having established that we approximate the manifold geometry of the self-similar solutions with a *uniform* accuracy for all time we now turn our attention to looking at (i) the dynamics on this manifold, and (ii) the dynamics close to this manifold.

The dynamics on the self-similar manifold

We have from (3.33) that

$$z = e^{\mu\Delta\tau} + \mathcal{O}\left(\Delta\tau^{p+1}\right),$$

and from the definitions of the continuous (3.27) and discrete (3.28) self-similar solutions at the fictive time $\tau = n\Delta\tau$ we also have respectively

$$t = e^{n\mu\Delta\tau}, \quad t_n = z^n.$$

Thus at this fictive time

$$t_n = \left[e^{\mu\Delta\tau} + \mathcal{O}(\Delta\tau^{p+1})\right]^n = e^{n\mu\Delta\tau}\left[1 + \mathcal{O}(n\Delta\tau^{p+1})\right].$$

But $\log(t) = n\mu\Delta\tau$ and so

$$\frac{t_n}{t} = 1 + \mathcal{O}(\Delta\tau^p \log(t)). \tag{3.44}$$

Therefore the relative error in t_n (and hence in each of the terms $u_{i,n}$) grows very slowly for large t.

Of course, in a sense this is a fictional error which is introduced due to our use of a fictive time and the fact that we are making comparisons at the same fictive time. As an alternative (and more realistic) measure of the error we can compare $u_{i,n}$ with u_i at the same real time t_n. It is trivial to establish the following result from (3.25) and (3.29) or (3.39).

Corollary 3.2. *Let $u_i(t)$ be a self-similar solution of the ordinary differential equation, then there is a discrete self-similar solution $(u_{i,n}, t_n)$ of the discrete scheme (with the assumption of Theorem 3.2 or Theorem 3.4 as necessary) such that for all n*

$$\frac{u_i(t_n)}{u_{i,n}} - 1 = \frac{U_i}{\hat{U}_i} - 1 = \mathcal{O}(\Delta\tau^p).$$

It is useful to rewrite this result as

$$u_i(t_n) = u_{i,n}(1 + C_{i,n}\Delta\tau^p), \tag{3.45}$$

where $C_{i,n}$ is a constant bounded for all i and n, independently of n. Therefore we see that the discrete self-similar solution approximates the true self-similar solution with uniform accuracy *for all times*, with a relative error that *does not grow* with time.

Observe that this result does not depend upon the form of the self-similar solution. Thus, if this solution is developing a singularity (for example in the problem of gravitational collapse seen in Chapter 2 as well as later on in this Chapter) the discrete self-similar solution continues to approximate it with a uniform relative error.

However as we discussed in Chapter 2 a solution to a problem with arbitrary initial conditions is almost definitely not self-similar in form, however it may well converge with t to a self-similar form. We now consider this property in our numerical method.

The dynamics close to the self-similar manifold

For a general system with general initial conditions, the self-similar solutions, though invariants of the system, do not satisfy the initial conditions. However, the most interesting (and also frequently occurring) physical self-similar solutions are those which act as attractors (and therefore determine asymptotic behaviours) for more general solutions with arbitrary initial conditions. In order to accurately compute solutions for large times it is desirable that the numerical method should preserve this structure, we establish this result in the following.

Theorem 3.6. *For small $\Delta\tau$ the (neutral) stability of the true self-similar solution is inherited by the discrete self-similar solution of both consistent and zero stable linear multistep methods (3.18), and consistent Runge-Kutta (3.19) methods.*

Proof. The general form of the perturbation to a continuous self-similar solution such as (3.27) is

$$u_i(\tau) = e^{\mu\alpha_i\tau}\left[U_i + a_i\right], \quad t(\tau) = e^{\mu\tau}\left[1 + s\right]. \tag{3.46}$$

Where, using the scaling invariance of the functions, we have to leading order

$$\dot{a}_i + \mu\alpha_i a_i = \sum \frac{\partial h_i}{\partial u_j} a_j, \quad \dot{s} + \mu s = \sum \frac{\partial g}{\partial u_j} a_j. \tag{3.47}$$

In general this system will have solutions of the form

$$a_i = A_i e^{\kappa_k\tau}, \quad s = S e^{\kappa_k\tau}, \tag{3.48}$$

where

$$\kappa_k A_i + \mu\alpha_i A_i = \sum \frac{\partial h_i}{\partial u_j} A_j, \quad \kappa_k S + \mu S = \sum \frac{\partial g}{\partial u_j} A_j. \tag{3.49}$$

and we have eigenvalues κ_k, $k = 1, \ldots, N+1$. The solutions of this eigenvalue problem then determine the stability of the self-similar solution. We observe immediately that there are two solutions to this problem which can be obtained from symmetry arguments. The first follows from the observation that we may make an arbitrary perturbation to τ of the form $\tau \to \tau + \varepsilon$. This is equivalent to a rescaling of the original self-similar solution and corresponds to taking

$$\kappa_1 = 0, \quad A_i = \varepsilon\mu\alpha_i U_i, \quad S = \varepsilon\mu.$$

The second follows from the observation that the original equation (3.1) is invariant under the action of $t \to t + \varepsilon$ and this corresponds to taking

$$\kappa_2 = -\mu, \quad A_i = 0, \quad S = \varepsilon.$$

For stability of the continuous self-similar solution we require that $\mathrm{Re}(\kappa_k) < 0$ for all $k > 2$.

Now consider the discrete self-similar solution. A perturbation to this takes the form

$$u_{i,n} = z^{\alpha_i n}\left[\hat{U}_i + \hat{a}_{i,n}\right], \quad t_n = z^n\left[1 + \hat{s}_n\right]. \tag{3.50}$$

Arguing in a similar manner to before we may pose a solution of (3.50) of the form

$$\hat{a}_{i,n} = \hat{A}_i z^{\nu_k n}, \quad \hat{s}_n = \hat{S} z^{\nu_k n} \tag{3.51}$$

In the following Lemma we establish that (up to a rescaling) the eigenvalues ν_k are (noting the equality of the eigenvalues when $k = 1$) perturbations of the eigenvalues κ_k. Therefore we have the desired result that for small $\Delta\tau$ the stability of the true self-similar solution is inherited by the discrete self-similar solution. □

Lemma 3.6. *For both consistent and zero stable linear multistep methods (3.18), and consistent Runge-Kutta (3.19) methods. The eigenvalue equation satisfied by the terms ν_k is identical to that satisfied by the eigenvalues κ_k (i.e. (3.49)) up to a perturbation of $\mathcal{O}(\Delta\tau)$ and a rescaling.*

Proof. We note firstly that since the same symmetries are acting on the discrete and continuous systems we may deduce, in a similar manner to before, that there are two eigenmodes with corresponding eigenvalues,

$$\nu_1 = 0, \quad \text{and} \quad \nu_2 = -1.$$

Now in the case of linear multistep methods substituting (3.51) into the discretized equation (3.18) we have, after some manipulation, that for $i = 1, \ldots, N$,

$$\sum_j \beta_j z^{j\alpha_i}\left[\hat{U}_i + \hat{A}_i z^{\nu_k(n+j)}\right] - \Delta\tau \sum_j \gamma_j z^{j\alpha_i} h_i\left(\hat{\mathbf{U}} + \hat{\mathbf{A}} z^{nu_k(n+j)}\right) = 0,$$
$$\sum_j \beta_j z^j\left[1 + \hat{S} z^{\nu_k(n+j)}\right] - \Delta\tau \sum_j \gamma_j z^j g\left(\hat{\mathbf{U}} + \hat{\mathbf{A}} z^{nu_k(n+j)}\right) = 0.$$

Now upon expanding h_i and g in Taylor series in the above, and using the fact that $(\hat{\mathbf{U}}, z)$ characterizes a discrete self-similar solution to (3.18) we may cancel terms to give, for $i = 1, \ldots, N$,

$$\sum_j \beta_j \hat{A}_i z^{j(\alpha_i+\nu_k)} - \Delta\tau \sum_j \gamma_j z^{j(\alpha_i+\nu_k)} \sum_l \hat{A}_l \frac{\partial h_i}{\partial u_l}(\hat{\mathbf{U}}) = 0,$$
$$\sum_j \beta_j z^{j(1+\nu_k)} \hat{S} - \Delta\tau \sum_j \gamma_j z^{j(1+\nu_k)} \sum_l \hat{A}_l \frac{\partial g}{\partial u_l}(\hat{\mathbf{U}}) = 0.$$

Recall from Theorem 3.3 and its proof results and notation such as $z = 1 + \mu\Delta\tau +$

$\mathcal{O}\left(\Delta\tau^2\right)$ and

$$\rho\left(z^{(\alpha_i+\nu_k)}\right)=\rho'(1)\Delta\tau\mu(\alpha_i+\nu_k)+\mathcal{O}\left(\Delta\tau^2\right),\quad \sigma\left(z^{(\alpha_i+\nu_k)}\right)=\rho'(1)+\mathcal{O}\left(\Delta\tau^2\right).$$

We then further simplify to give

$$\mu(\alpha_i+\nu_k)\hat{A}_i=\sum_l \hat{A}_l\frac{\partial h_i}{\partial u_l}+\mathcal{O}\left(\Delta\tau\right),\quad \mu(1+\nu_k)\hat{S}=\sum_l \hat{A}_l\frac{\partial g}{\partial u_l}+\mathcal{O}\left(\Delta\tau\right),$$

and hence comparing with (3.49) we may conclude the desired result, with the rescaling being that the κ_k corresponds to $\mu\nu_k$.

Now the same result can be established for the Runge-Kutta method (3.19) in an analogous manner to the above, similar to the way in which the proof of Theorem 3.5 mirrored that of Theorem 3.3, we therefore omit the proof in this case. □

3.4.4 Classical adaptive methods for ODEs

We now consider further the relation between adaptivity and scaling. The previous Sections have shown that a discretization of an appropriately rescaled equation has many desirable properties. However we may also ask the converse question, given a differential equation with scaling invariance, can an adaptive method automatically identify a suitable time step to capture the scaling invariance property? We shall now demonstrate that this does indeed follow provided a suitable adaptive strategy is used.

Consider a linear multistep discretization of the original problem (3.1) of the form

$$\sum_j \beta_j \mathbf{u}_{n+j}=\Delta t_n\sum_j \gamma_j \mathbf{f}(\mathbf{u}_{n+j}),\tag{3.52}$$

where $\Delta t_n \equiv t_{n+1}-t_n$ is the time step chosen by the method and u_n is an approximation to the true solution at time t_n. A common practical strategy for determining Δt_n is to make some estimate of the local error and to then choose Δt_n so that this estimate is below a user-defined tolerance over each time interval.

As a first error estimate consider the relative local truncation error of the method. The local truncation error for the linear multistep method is given [108] to leading order by

$$\mathbf{e}=C(\Delta t_n)^{p+1}\mathbf{u}^{(p+1)},$$

where in this Section we consider all derivatives of $\mathbf{u}$ to be with respect to t. Following the definition (3.24) we now consider the relative local truncation error

$$E(\Delta t_n)=\max_i |e_i/u_i|.$$

Assuming for now that we can estimate this error estimate, we can choose Δt_n to bound it by a given tolerance. For now set this tolerance to be (the constant) $(\Delta\tau)^{p+1}$ for some fixed $\Delta\tau$. An obvious method for computing Δt_n is then to set

$$E(\Delta t_n) = \Delta\tau^{p+1}, \tag{3.53}$$

and to solve this for Δt_n. Using our estimate for E and for simplicity setting all constants to unity, we have

$$\Delta t_n = \Delta\tau \left(\min_i |u_i/u_i^{(p+1)}|\right)^{1/(p+1)} \equiv \Delta\tau g(\mathbf{u}). \tag{3.54}$$

But notice the very encouraging result that in the limit of small $\Delta\tau$ the relation (3.54) is precisely a leading order discretization of the ordinary differential equation reminiscent of earlier Sections

$$\frac{dt}{d\tau} = g(\mathbf{u}).$$

Now consider how the function g scales. A straightforward extension to Lemma 3.4 gives that under the scaling (3.5) we have that

$$u_i^{(p+1)}(\lambda t) \to \lambda^{\alpha_i-(p+1)} u_i^{(p+1)}(t), \quad i = 1, \dots, N.$$

Hence from the definition of g

$$g(\lambda^{\boldsymbol{\alpha}}\mathbf{u}) \to \left(\min_i \left(\frac{\lambda^{\alpha_i}}{\lambda^{\alpha_i-(p+1)}} |u_i/u_i^{(p+1)}|\right)\right)^{1/(p+1)} = \lambda g(\mathbf{u}),$$

and thus g, which was derived from the relative local truncation error estimate, scales in precisely the manner required by the function specified in Section 3.3. Therefore the adaptive method so constructed inherits the scaling invariance property of the continuous problem and the results obtained in the previous Section follow for this technique. However, note that the discretization implied by this adaptive approach is only first order accurate, thus whilst it will follow a self-similar solution it may (though not necessarily) do so at a reduced level of accuracy.

In practice of course the local truncation error is not available to us and it must be estimated. Moreover it is generally hard to solve (3.53) exactly for Δt_n, and instead Δt_n is often successively halved until $E(\Delta t_n)$ or some other error measure is bounded above by $\Delta\tau^{p+1}$. One such method is to use the Milne device [86], [100] in which two computations of an approximate solution are made using two different multistep methods, and the difference between them used as an estimate of the error E. In principle, provided the leading order behaviour of the error accurately reflects the true error, the estimate for E based upon the Milne device should have exactly the same scaling properties as the above estimate. Thus if E is estimated in this manner and (3.53) solved, the resulting method will be scale invariant.

3.5 Numerical examples

We shall now illustrate some of the results and issues raised in this Chapter by considering several simple examples.

Example 3.1[1]

Suppose that $u(t)$ satisfies the ordinary differential equation

$$\frac{du}{dt} = -4u^4, \quad u(0) = u_0.$$

This is invariant under the transformation

$$t \to \lambda t, \quad u \to \lambda^{-1/3} u.$$

Although an invariant choice of temporal monitor function g is not difficult to spot straight off, in general it may be found by solving the problem (3.7), which in this case reads

$$g = -\frac{1}{3} u g_u$$

the immediate solution being $g(u) = u^{-3}$, using this choice in (3.16) we then have

$$\frac{du}{d\tau} = -4u, \quad \frac{dt}{d\tau} = u^{-3} \tag{3.55}$$

with $u(0) = u_0$, $t(0) = 0$. It is easy to see that equation (3.55) is scale invariant, so that if $(u(\tau), t(\tau))$ is a solution then so is $(\lambda^{-1/3} u(\tau), \lambda t(\tau))$. It has the self-similar solution

$$u = A \exp(-4\tau), \quad t + C = \frac{A^{-3}}{12} \exp(12\tau), \quad A = 12^{-1/3}.$$

We now discretize (3.55) using (for ease of exposition) the forward Euler method. This gives

$$u_{n+1} - u_n = -4\Delta\tau u_n, \quad t_{n+1} - t_n = \Delta\tau u_n^{-3}, \tag{3.56}$$

with $u_0 = u_0$ and $t_0 = 0$. This scale invariant discretization admits a discrete self-similar solution of the form

$$u_n = V y^n, \quad t_n = a + b y^{-3n} \tag{3.57}$$

[1] This example is motivated by the porous medium equation studied in Chapter 5, it arises from the semi-discretization of the PDE if three spatial mesh points are used with the end nodes being fixed at zero ($W_0 = W_2 = \dot{W}_0 = \dot{W}_2 = 0$). The ODE considered in this example is then simply the equation describing the evolution of $W_1(t)$.

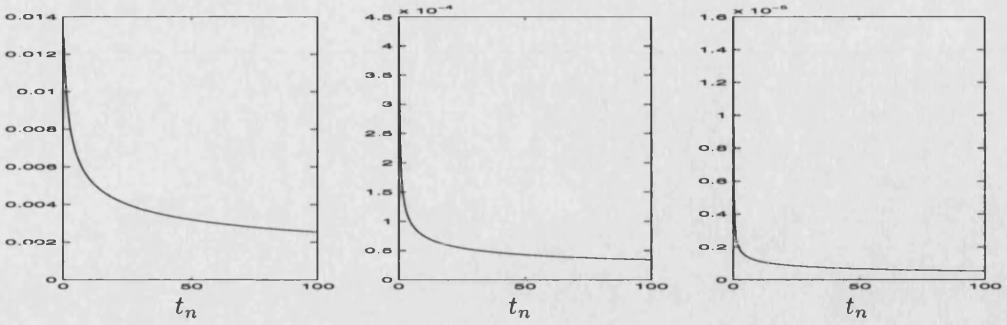


Figure 3-1: *Plots of the difference between the computed solution U_n to Example 3.1 and the calculated discrete self-similar solution (3.58). (Left) with (3.58) truncated after first term (note, this is equivalent to comparing with the exact solution (3.59)). (Centre) with (3.58) truncated after second term. (Right) with (3.58) truncated after third term.*

where V, $y = z^{-1/3}$, a and b are to be determined. Substituting into the first equation in (3.56) (and dividing by the constant factor of Vy^n) we have

$$y = 1 - 4\Delta\tau = \exp(-4\Delta\tau) + \mathcal{O}(\Delta\tau^2).$$

Similarly, from the second equation in (3.56) we have (on division by the constant factor of y^{-3n})

$$b(y^{-3} - 1) = \frac{\Delta\tau}{2} V^{-3},$$

therefore,

$$b = V^{-3}\left\{12 + 96\Delta\tau + 640\Delta\tau^2 + \mathcal{O}(\Delta\tau^3)\right\}^{-1}.$$

The initial conditions give that $V = u_0$ and $a + b = 0$. Thus, from (3.57) we have

$$t_n = -b + b\left(\frac{u_0}{u_n}\right)^3 = \left(u_n^{-3} - u_0^{-3}\right)\left\{12 + 96\Delta\tau + 640\Delta\tau^2 + \mathcal{O}(\Delta\tau^3)\right\}^{-1},$$

which we may rearrange to give

$$\begin{aligned} u_n &= u_0\left[1 + t_n u_0^3\left\{12 + 96\Delta\tau + 640\Delta\tau^2 + \mathcal{O}(\Delta\tau^3)\right\}\right]^{-1/3} \\ &= u_0\big[(1 + 12t_n u_0^3)^{-1/3} - 32t_n u_0^3(1 + 12t_n u_0^3)^{-4/3}\Delta\tau \\ &\quad -\frac{128}{3}t_n u_0^3(5 + 12t_n u_0^3)(1 + 12t_n u_0^3)^{-7/3}\Delta\tau^2 + \mathcal{O}(\Delta\tau^3)\big]. \end{aligned} \tag{3.58}$$

We now compare expression (3.58) with the exact solution which is given by

$$u(t) = u_0\left(1 + 12tu_0^3\right)^{-1/3}. \tag{3.59}$$

This is, in fact, a self-similar solution with respect to the translated time $s = t + u_0^{-3}/12$.

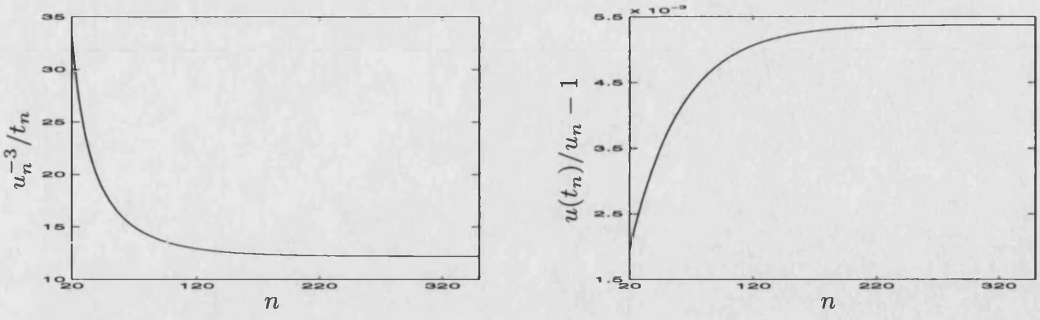


Figure 3-2: *For Example 3.1, (Left) Plot of u_n^{-3}/t_n for Example 3.1, note convergence to value close to* 12. *(Right) Plot of $u(t_n)/u_n - 1$.*

We have

$$\begin{aligned} u_n - u(t_n) = & - 32t_n u_0^4(1 + 12t_n u_0^3)^{-4/3}\Delta\tau \\ & - \frac{128}{3}t_n u_0^4(5 + 12t_n u_0^3)(1 + 12t_n u_0^3)^{-7/3}\Delta\tau^2 + \mathcal{O}(\Delta\tau^3). \end{aligned} \tag{3.60}$$

Note that all terms in this error expansion have the property that they possess the factor $t_n^{-1/3}$ and as such the error decrease with t_n.

In figure 3-1 results from a calculation of this problem using the forward Euler method with $\Delta\tau = 0.01$, and (non self-similar) initial conditions $u = 1$ at $t = 0$ are given. The difference between the computed solution and truncations of the discrete self-similar solution (3.58) are shown. From this we may conclude that the numerics are indeed approaching the discrete self-similar solution, which in turn, thanks to (3.60), is uniformly close and actually converging to the exact solution. Correspondingly in figure 3-2 some plots are given which demonstrate the results of Theorem 3.1 and Corollary 3.2.

Example 3.2

We now consider a problem where for demonstration purposes we shall invoke both the a priori and a posteriori methods of performing adaptivity. Consider the problem

$$\frac{du}{dt} = u^{-4}, \tag{3.61}$$

with $u = 1$ at $t = 1$. This problem is invariant under the transformation

$$t \to \lambda t, \quad u \to \lambda^{1/5}u.$$

The exact solution to our problem is given by

$$u(t) = (5t - 4)^{1/5},$$

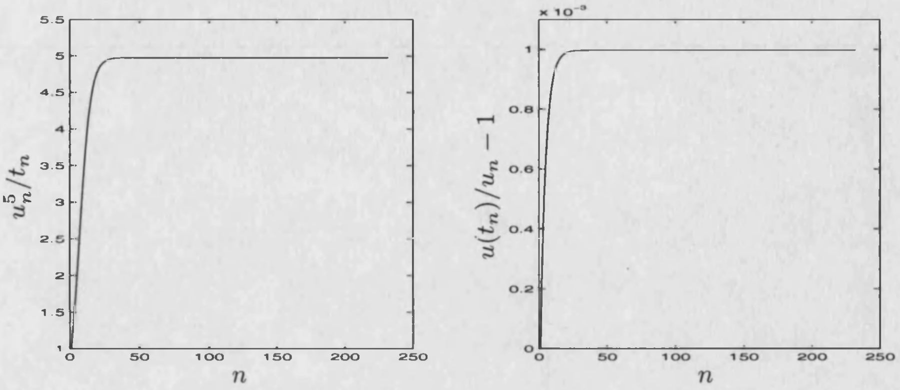


Figure 3-3: *(Left) Plot of u_n^5/t_n for Example 3.2, note convergence to value close to* 5. *(Right) Plot of $u(t_n)/u_n - 1$ for Example 3.2.*

for which

$$t^{-1/5}u \to 5^{1/5}. \tag{3.62}$$

A priori rescaling. The specific form of the scaling underlying this problem implies that a suitable choice of monitor function is $g = u^5$, leading to

$$\frac{du}{d\tau} = u, \quad \frac{dt}{d\tau} = u^5. \tag{3.63}$$

Consider now a trapezoidal discretization of (3.63),

$$u_{n+1} - u_n = \frac{\Delta\tau}{2}(u_n + u_{n+1}), \quad t_{n+1} - t_n = \frac{\Delta\tau}{2}\left(u_n^5 + u_{n+1}^5\right).$$

This system admits the discrete self-similar solution

$$u_n = Vy^n, \quad t_n = y^{5n} \tag{3.64}$$

Substituting into our discretization yields

$$y = \frac{1 + \Delta\tau/2}{1 - \Delta\tau/2} = \exp(\Delta\tau) + \mathcal{O}(\Delta\tau^3),$$

and

$$V^5 = \frac{2}{\Delta\tau}\frac{y^5 - 1}{y^5 + 1} = 5 + \mathcal{O}(\Delta\tau^2),$$

which are consistent with the theoretical results obtained in Section 3.4.3. Notice that the constant V here is approximately equal to the $5^{1/5}$ which appears in (3.62). We see this behaviour in figure 3-3, with the convergence of V^5 to a value very close to 5, further experimentation with different values for $\Delta\tau$ confirms that this figure is indeed in error by $\mathcal{O}(\Delta\tau^2)$.

A posteriori rescaling. We shall now attempt to determine the time step $\Delta t_n \equiv t_{n+1} - t_n$ in the course of the calculation. Consider again a trapezoidal discretization, this time using the (still to be computed) time step Δt_n,

$$u_{n+1} - u_n = \frac{\Delta t_n}{2}(u_n^{-4} + u_{n+1}^{-4}). \tag{3.65}$$

The local truncation error here is given by $e = C(\Delta t_n)^3 u^{(3)}$ for some constant C, [108]. But on differentiation (3.61) gives

$$u^3 = 36u^{-14},$$

and so the relative local truncation error here is given by (ignoring constants for the moment)

$$E = (\Delta t_n)^3 u^{-15}.$$

Evaluating this expression at u_n and following Section 3.4.4 to set this equal to the tolerance $(\Delta\tau)^3$, we have

$$\Delta t_n = u_n^5 \Delta\tau.$$

Using this time step in the discretization (3.65) gives the method

$$u_{n+1} - u_n = \frac{\Delta\tau}{2} u_n^5 (u_n^{-4} + u_{n+1}^{-4}), \quad t_{n+1} - t_n = u_n^5 \Delta\tau. \tag{3.66}$$

As for the a priori case we can now look for a discrete self-similar solution of the form (3.64) for this discrete scheme. Following similar details to above, but this time asking Maple to solve a nonlinear algebraic system, yields

$$y = 1 + \Delta\tau - 2\Delta\tau^2 + \mathcal{O}(\Delta\tau^3) = \exp(\Delta\tau) + \mathcal{O}(\Delta\tau^2),$$

and

$$V = 5^{1/5} + \mathcal{O}(\Delta\tau^2).$$

Therefore we can conclude that the use of a method with a correctly chosen time step yields similar results to the a priori technique of rescaling the problem. Notice here however that we have dropped an order of accuracy in the approximation y. This is due to the fact that in (3.66) the discretization of (3.2) yielding t_{n+1} is one order of accuracy lower than the approximation of (3.3). This reenforces the fact that the equation describing the coordinate transformation should be discretized to the same order of accuracy as the equation describing the dependent variable.

Example 3.3

We now look at a problem which has a solution which blows up in finite time. Consider

$$\frac{du}{dt} = u^4.$$

As in example 3.1 this is invariant under the transformation

$$t \longrightarrow \lambda t, \quad u \longrightarrow \lambda^{-1/3} u,$$

and so a suitable choice of monitor function is $g = u^{-3}$, leading to

$$\frac{du}{d\tau} = u, \quad \frac{dt}{d\tau} = u^{-3}. \tag{3.67}$$

Firstly consider a forward Euler discretization,

$$u_{n+1} - u_n = \Delta\tau u_n, \quad t_{n+1} - t_n = \Delta\tau u_n^{-3},$$

from which we have (assuming now that $t_0 = 0$)

$$u_n = u_0(1+\Delta\tau)^n, \quad t_n = \Delta\tau \sum_{i=0}^{n-1} u_i^{-3}. \tag{3.68}$$

Notice firstly that as $n \longrightarrow \infty$ we have $u_n \longrightarrow \infty$ as required, see Section 2.2.4 for the exact solution. Secondly if in (3.68) we substitute the first expression into the second we arrive at

$$t_n = \Delta\tau u_0^{-3} \frac{1-(1+\Delta\tau)^{-3n}}{1-(1+\Delta\tau)^{-3}},$$

which as $n \longrightarrow \infty$ gives us

$$t_n \longrightarrow \Delta\tau u_0^{-3} \left(\frac{1}{3}\Delta\tau^{-1} + \frac{2}{3} + \mathcal{O}(\Delta\tau) \right) = T + \mathcal{O}(\Delta\tau),$$

where $T = u_0^{-3}/3$ as we found in Section 2.2.4, hence our adaptive approach is accurately approximating the finite blow-up time to an accuracy consistent with the method. This result was established in [28].

Consider now a trapezoidal discretization

$$u_{n+1} - u_n = \frac{\Delta\tau}{2}(u_n + u_{n+1}), \quad t_{n+1} - t_n = \frac{\Delta\tau}{2}(u_n^{-3} + u_{n+1}^{-3}).$$

This system admits the discrete self-similar solution

$$u_n = Vy^n, \quad t_n = y^{-3n}$$

where, following a similar procedure to the previous examples, we find that

$$y = \frac{1+\Delta\tau/2}{1-\Delta\tau/2}, \quad \text{and} \quad V = -3^{-1/3} - 2.3^{-7/3}\Delta\tau^2 + \mathcal{O}(\Delta\tau^4).$$

So that

$$u_n = \left(-3^{-1/3} - 2.3^{-7/3}\Delta\tau^2 + \mathcal{O}(\Delta\tau^4)\right) t_n^{-1/3},$$

which uniformly approximates the true solution

$$u(t) = (-3t)^{-1/3}$$

to an accuracy of $\mathcal{O}(\Delta\tau^2)$.

For obvious reasons numerical methods in general will display difficulties when computing singular solutions to problems such as the present one. We have already seen above that our new adaptive methods cope well in this situation. The following discussion looks at how powerful such methods may be.

Due to the theory presented earlier in this Chapter we know that for this example the continuous problem and the numerical method both admit self-similar solutions of the form

$$u^{-3}(T-t)^{-1} = \text{Const} \approx 3. \tag{3.69}$$

For the continuous problem (and correspondingly the continuous self-similar solution) we know that the constant appearing in (3.69) must be 3. However for the numerical method (and correspondingly the discrete self-similar solution) the constant is actually $3 + \mathcal{O}(\Delta\tau^p)$. The problem is that prior to the calculation we do not know the discrete blow-up time T_Δ. If we were to approximate T_Δ by the true blow-up time T for example then rather than converging to a constant (3.69) would either tend to zero or infinity. Alternatively suppose for example that we compute until our discrete numerical solution u_n is greater than 10 say, if we take the corresponding t_n as a guess to T_Δ then as we are obviously underestimating the discrete blow-up time (3.69) will tend to infinity. However we can easily plot the left hand side of ((3.69)) using the guess

$$T = t_n|_{u_n \approx 10} + \varepsilon,$$

and by trial and error find the value of ε which best satisfies (3.69).

For example, in this current problem if we compute (using forward Euler with $\Delta\tau = 0.1$) until $u_n \geq 10$ we find the initial guess $T_\Delta = 1.33302955217525$. A simple trial and error plotting strategy, as described above, then yields $\varepsilon = 3.26003195126 \times 10^{-4}$ and hence the discrete blow-up time $T = 1.33335555537037$. But now we have the interesting result that computing again with our method, this time until $u_n \geq 10^{10}$ yields this exact result for all 14 decimal places. Note firstly that as expected the discrete blow-up time approximates the true blow-up time to order $\Delta\tau^p$ (further experiments confirm this fact). Secondly, even though we are only computing a very small distance into the blow-up regime, the numerics are already following the discrete self-similar solution accurately at this early stage in the evolution. Therefore with very little computational effort we can have confidence in predicting from the early numerical data when the discrete solution will blow-up.

Note that this technique of using a discrete solution ansatz to attempt to predict the longer (in n) behaviour of a numerical method really does depend upon the scale invariance of the method. As a test, applying this technique to the adaptive methods provided by MATLAB (e.g. ODE23) does not lead to such satisfying results. For example, it is not possible in general to choose a time T_Δ such that (3.69) is close to a constant. The 'best' effort however, on longer integration, is shown not to be a good approximation to the true discrete blow-up time for the numerical scheme.

Example 3.4 Gravitational collapse.
We now consider the gravitational collapse problem introduces in the previous Chapter.

$$\frac{dr}{dt} = v, \quad \frac{dv}{dt} = -r^{-2}. \tag{3.70}$$

This has the scaling invariance

$$t \to \lambda t, \quad r \to \lambda^{2/3} r, \quad v \to \lambda^{-1/3} v$$

for any arbitrary positive constant λ. Notice that this problem can also be written in the canonical Hamiltonian form (2.33), with Hamiltonian or energy

$$H(r,v) = \frac{v^2}{2} - \frac{1}{r}.$$

(The system is also invariant under reflexions and translations in time.) A singularity can occur in this system in finite time T. This is called *gravitational collapse* and occurs when a particle falls into the sun. An example of a solution with this property is the following self-similar solution found in Chapter 2,

$$r = R(T-t)^{2/3}, \quad v = -V(T-t)^{-1/3} \tag{3.71}$$

where T is an arbitrary finite (collapse) time and

$$R^3 = \frac{9}{2}, \quad V = \frac{2}{3}\left(\frac{9}{2}\right)^{1/3}. \tag{3.72}$$

This solution is of interest to us as it forms a singularity in a finite time in which $r \longrightarrow 0$ and $v \longrightarrow -\infty$. We immediately observe that it is difficult to capture such behaviour if a fixed time step is used.

If we set $g = r^{3/2}$ then we obtain the invariant system given by

$$\frac{dr}{d\tau} = r^{3/2}v, \quad \frac{dv}{d\tau} = -r^{-1/2}, \quad \frac{dt}{d\tau} = r^{3/2}. \tag{3.73}$$

For this system a collapsing self-similar solution can be given by

$$r = Re^{-2\mu\tau/3}, \quad v = -Ve^{\mu\tau/3}, \quad t = T - e^{-\mu\tau} \tag{3.74}$$

where

$$\mu = \left(\frac{9}{2}\right)^{1/2}.$$

Observe that as $\tau \longrightarrow \infty$ we have $t \longrightarrow T$, and that $t = T - 1$ at $\tau = 0$.

Now consider solving (3.73) by using the trapezoidal rule with step size $\Delta\tau$. In this case the scheme admits a discrete collapsing self-similar solution given by

$$r_n = \hat{R}z^{2n/3}, \quad v_n = -\hat{V}z^{-n/3}, \quad t_n = T_{\Delta\tau} - z^n.$$

Comparing with (3.74) we see that z is an analogue of $\exp(-\mu\Delta\tau)$ if $n\Delta\tau = \tau$. Here $|z| < 1$ so that $r_n \longrightarrow 0$, $|v_n| \longrightarrow \infty$ and $t_n \longrightarrow T_{\Delta\tau}$ as $n \longrightarrow \infty$. Here $T_{\Delta\tau}$ is a discrete collapse time which need not necessarily coincide with the true collapse time T. The constants $\hat{R}, \hat{V}$ and $z < 1$ then satisfy the algebraic equations

$$\hat{R}(z^{2/3} - 1) + \frac{\Delta\tau}{2}\hat{R}^{3/2}\hat{V}(z^{2/3} + 1) = 0, \tag{3.75}$$

$$\hat{V}(z^{-1/3} - 1) - \frac{\Delta\tau}{2}\hat{R}^{-1/2}(z^{-1/3} + 1) = 0, \tag{3.76}$$

$$-(z - 1) - \frac{\Delta\tau}{2}\hat{R}^{3/2}(z + 1) = 0, \tag{3.77}$$

Recall that following the predictions of Theorem 3.3 we have

$$\hat{R} = R(1 + \mathcal{O}(\Delta\tau^2)), \quad \hat{V} = V(1 + \mathcal{O}(\Delta\tau^2)), \quad z = e^{\mu\Delta\tau}(1 + \mathcal{O}(\Delta\tau^3))$$

where, giving their numerical values, we have

$$R = 1.650964, \quad V = 1.100642.$$

The resulting values for the trapezoidal discretization are as given below,

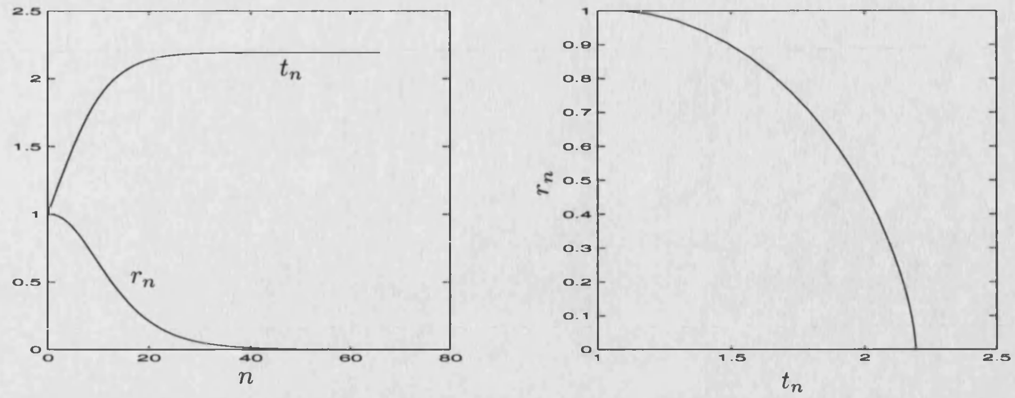


Figure 3-4: *(Left) Plot of r_n and t_n against iteration number n in gravitational collapse Example 3.4, here $\Delta\tau = 0.1$. (Right) Plot of r_n against t_n demonstrating the singular nature of collapse as $t_n \to T_{\Delta\tau}$.*

$\Delta\tau$	$\hat{R}$	$\hat{V}$	z
0.1	1.647989	1.100947	0.8086789
0.01	1.650934	1.100645	0.9790100
0.001	1.650963	1.100642	0.9978809

and $\exp(\mu\Delta\tau)$ takes the values 0.808858, 0.9790102, and 0.9978809 for the three choices of $\Delta\tau$ above. The results in this table are fully consistent with the given error estimates. Similarly we may also use a forward Euler discretization of the same system.

$$\begin{aligned} r_{n+1} - r_n &= r_n^{3/2} v_n \Delta\tau \\ v_{n+1} - v_n &= -r_n^{-1/2}\Delta\tau \\ t_{n+1} - t_n &= r_n^{3/2}\Delta\tau. \end{aligned} \tag{3.78}$$

This gives a very similar discrete self-similar collapse solution for which the corresponding values are given by,

$\Delta\tau$	$\hat{R}$	$\hat{V}$	z
0.1	1.556330	1.074403	0.8058432
0.01	1.641262	1.098045	0.9789735
0.001	1.649991	1.100383	0.9978806

As expected these values converge more slowly to the true values, exhibiting a first order rate of convergence.

Now consider a numerical implementation of the forward Euler method, for this problem. For initial values we take $r = 1$ and $v = 0$ at (without loss of generality) $t = 1$. This problem then has the exact solution given by the quadrature

$$t = 1 + \int_r^1 \frac{\sqrt{s}}{\sqrt{2(1-s)}} ds, \tag{3.79}$$

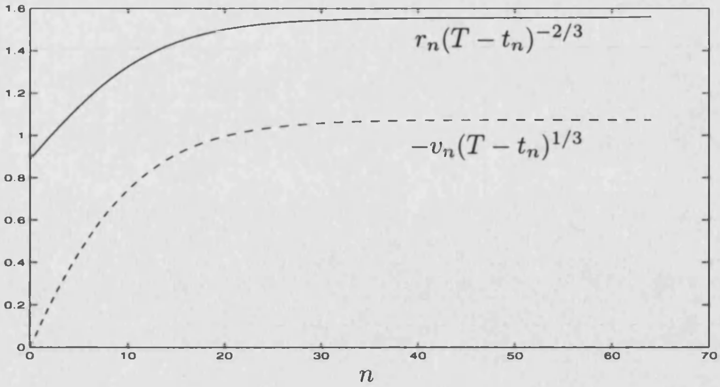


Figure 3-5: *Convergence of scaled solutions in gravitational collapse Example 3.4.*

with a gravitational collapse occurring when

$$T = 1 + \frac{\pi}{2\sqrt{2}} = 2.110720735.$$

The true solution for these values is not self-similar, but it does converges toward a self-similar solution as the collapse time is approached.

Using the rescaled method we firstly calculate the value of the discrete collapse time T_Δ as a function of $\Delta\tau$, where T_Δ is estimated as the first value of t_n at which $r_n < 10^{-6}$. These results are given below,

$\Delta\tau$	$T_{\Delta\tau}$	$T_{\Delta\tau} - T$
0.1	2.1925	0.0818
0.01	2.1188	0.0081
0.001	2.1115	0.0008

These results give convincing evidence that for this method

$$T_{\Delta\tau} \approx T + 0.8\Delta\tau$$

thus exhibiting first order convergence to the true collapse time, consistent with the rate of convergence of the forward Euler scheme.

Now consider the behaviour of the method close to collapse with $\Delta\tau = 0.1$. In Figure 3-4 we plot t_n and r_n both as functions of τ. Observe that t_n tends towards the constant value of T_Δ whilst r_n tends to zero. Also shown is a plot or r_n as a function of t_n in this case. Observe the singular nature of collapse of the solution. Now, using the collapse time given above we may rescale the solution by calculating $r_n(T_{\Delta\tau} - t_n)^{-2/3}$ and $v_n(T_{\Delta\tau} - t_n)^{1/3}$. These quantities are plotted in Figure 3-5 as functions of n. Observe that they converge as n increases to the respective constants $\hat{R} = 1.55633$ and $\hat{V} = 1.07440$ identified in the earlier analysis. Thus, as for the continuous problem, the numerical solution converges to the discrete self similar solution when rescaled with the correct discrete collapse time.

3.6 Summary of Chapter

In this Chapter adaptivity (achieved through the use of a temporal coordinate transformation) has been used in a novel way to allow standard ODE integration methods to respect scaling symmetries of certain continuous problems.

Much use has been made of the result that scalings represent a transformation group for which discretization and equation invariance commute. For example it was proved that correctly constructed numerical methods accurately admit discrete self-similar solutions for arbitrarily large (fictive) times with an error that does not grow. Significantly this result holds whatever the form of the solution, even if it develops a singularity in finite (real) time. The stability of said solutions was also shown to be reproduced in the numerics. In addition a posteriori rather than a priori techniques where shown to exhibit similar properties.

Finally, several model problems with a variety of solution types were considered and integrated using the newly constructed methods. The rigorous results made in this Chapter were demonstrated *in action* for each example.

In the following Chapter the question of what to do when confronted with a problem with both a scaling invariance and a Hamiltonian formulation is considered. In Chapter 5 the extension of this Chapter to PDEs is addressed, where use is made of certain methods to perform spatial adaptivity.

Chapter 4

A comparison of symplectic and scale invariant methods for Hamiltonian ODEs

4.1 Overview of Chapter

Noether's theorem was discussed in Chapter 2, we saw that the combination of Lagrangian or Hamiltonian structures in a problem with (special types of) symmetries yields very useful results. Specifically it gives us the existence of, as well as explicit expressions for, conservation laws of the system in question. The following question naturally arises: is it possible to preserve both a symplectic structure and symmetries in a method applied to problems which, such as Kepler's problem, possess both properties? In addition, if this is possible does the method inherit any corresponding conservation laws, in some sense *for free*. This is of course a single example of the important and fundamental point, which was raised in Chapter 1, of whether it is possible to preserve more than one geometric property in a numerical method. We may also ask the question, again given in more generality in Chapter 1, of whether it is more beneficial to preserve symplecticity or symmetries in a method if only one of these is possible. We attempt to investigate some of these properties in this Chapter.

The technique introduced in the previous Chapter for preserving scaling symmetries in a numerical method depended crucially on making a time transformation, or equivalently performing a special type of temporal adaptivity. As shall be discussed below, a problem arises in attempting to preserve both scaling symmetries and symplecticity in a numerical method. Specifically, numerical evidence suggests that the use of adaptive time stepping destroys the very desirable features of symplectic numerical methods. We therefore employ below an alternate time transformation strategy which avoids this deficit.

4.2 Systems possessing both symmetries and Hamiltonian structure

The Kepler two-body problem which was introduced in Chapter 2 and further analysed in Chapter 3 shall be used extensively throughout this Chapter. This tends to be the prototypical example used in geometric integration. This is because it is a relatively simple but very interesting problem which possesses many different but complimentary geometric properties. For example we witnessed its Hamiltonian formulation in Section 2.3.1. Its symmetries, conservation laws and the links between them was then given in Section 2.3.4.

As was discussed in Chapter 1, some of the unresolved questions in geometric integration revolve around how to handle problems with multiple geometric properties. In this Chapter we therefore make a comparison of geometric integration methods designed to capture different properties. Specifically the scaling invariant methods from Chapter 3, the symplectic methods briefly described in Chapter 2, as well as hybrid methods which shall be developed in this Chapter and which aim to preserve both properties. To again demonstrate the superiority of geometric integration methods in certain circumstances, standard methods shall also occasionally be used for comparison purposes.

Discrete versions of Noether's theorem have been established [58], and numerical methods preserving both symmetries and Hamiltonian or Lagrangian structures have been shown to automatically inherit discrete analogues of continuous conservation laws [103, 116]. It is therefore now apt to investigate the possible preservation of conservation laws in, if they exist, symplectic versions of the scaling invariant adaptive methods developed in Chapter 3.

As was mentioned above, Kepler's problem shall again be used as a model problem in this Chapter. The Hamiltonian for Kepler's problem was given in (2.34). Hamilton's corresponding equations describing the motion of the system are then

$$\dot{p}_i = -\frac{q_i}{(q_1^2+q_2^2)^{3/2}}, \quad \dot{q}_i = p_i, \quad i = 1, 2. \tag{4.1}$$

We may think of $\mathbf{q}$ representing the position and $\mathbf{p}$ the velocity of a heavenly body moving (in a two-dimensional plane) around the Sun positioned at the origin of our coordinate system. In the idealized state considered here the two objects have equal mass. Throughout this Section take the initial conditions for the problem to be

$$q_1(0) = 1 - e, \; q_2(0) = 0, \; p_1(0) = 0, \; p_2(0) = \sqrt{\frac{1+e}{1-e}}.$$

In which case the exact solution to the problem is given by a conic Section of type and precise shape controlled by the value of the parameter e — the *eccentricity*. We focus here on the case of the ellipse (periodic solutions of period 2π), i.e. we consider eccentricities between zero and one, $0 \leq e < 1$. Along with the Hamiltonian

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

which represents the total energy of the system, the angular momentum of the system given by

$$L(\mathbf{p},\mathbf{q}) = q_1 p_2 - q_2 p_1, \tag{4.2}$$

is also a conserved quantity.

For eccentricities $e \uparrow 1$ the exact solution is given by a very elongated ellipse with the planet experiencing a very close approach (near collision) with the Sun. During this close approach the planet experiences very large forces and accelerations. Whereas away from the close approach the behaviour of the orbit, or solution, is far more sedate. This is precisely the type of situation where the use of adaptive time stepping can be incredibly beneficial, with the use of larger time steps away from the close approach and smaller time steps and hence the concentration of computational effort during times of close approach. Since methods designed to preserve scaling invariances have just been developed and based fundamentally on the use of adaptivity it is natural to test the methods of Chapter 3 on this problem. However in Chapter 2 we showed how well symplectic methods cope with Kepler's problem, therefore it is natural to attempt to construct symplectic scale invariant methods. We now arrive at a problem however, since in [146] symplectic methods with standard adaptive time stepping strategies are tested, and disappointingly they are shown to behave in a non-symplectic way in that the advantages due to using a symplectic method (for example, the near conservation of the Hamiltonian, and the linear in time growth in trajectory errors) are lost. Thus although an individual step of an individual orbit may be symplectic, the overall map induced by the method on the whole of phase space may not be symplectic. Moreover, there is no obvious shadowing property, in that backward error analysis is hard to apply and the solution of the numerical method is not closely approximated by the solution of a nearby Hamiltonian system. As a consequence, the symplectic method with a (badly chosen) variable time step exhibits quadratic (which standard methods demonstrate) rather than the linear error growth for large times expected of symplectic methods when applied to Hamiltonian problems with periodic solutions (see [164]) or to integrable Hamiltonian problems (see [146]), such as the Kepler problem. For other discussions regarding this matter see [39, 40, 156, 162].

4.3 Time transformations

An obvious method for performing temporal adaptivity is through the Sundman transform we discussed in the previous Chapter for preserving scaling symmetries. However, as we shall now see there is a problem with this approach for Hamiltonian problems. Now Kepler's problem (4.1) is invariant under the scaling transformation

$$t \to \lambda t, \quad (q_1, q_2) \to \lambda^{2/3}(q_1, q_2), \quad (p_1, p_2) \to \lambda^{-1/3}(p_1, p_2),$$

for any arbitrary positive constant λ. Suppose that we perform the Sundman transform (3.2), (3.3) on this system with the scale invariant choice of

$$g = (q_1^2 + q_2^2)^{3/4}. \tag{4.3}$$

This yields the new scale invariant system,

$$\frac{dp_i}{d\tau} = -q_i(q_1^2 + q_2^2)^{-3/4}, \quad \frac{dq_i}{d\tau} = p_i(q_1^2 + q_2^2)^{3/4}, \quad i = 1, 2. \tag{4.4}$$

The choice of power 3/4 in the function g is also arrived at in [23] by equalizing the amount of fictive time required for both strong and weak collision events. This property has some similarity with Kepler's third law which our scaling invariant method automatically inherits (recall Section 3.4.1), and therefore the common choice of 3/4 should be unsurprising. Also, in [83] (where actually the Poincaré transformation to be given presently is used) the same function g is employed where, via experimentation, the optimum power is found to be somewhere between 0.5 and 1, dependent on the eccentricity of the orbit. The scale invariance thus gives us a choice which does not disagree with these recommendations.

The transformed system (4.4) has no Hamiltonian formulation, and in general the Sundman transform procedure fails to preserve any Hamiltonian structure present in a problem. But we do have here that since the function g has been chosen in a special way any numerical method applied to (4.4) will preserve the scaling invariance of the problem, the results of Chapter 3 will then follow. There would appear to be no reason to use a symplectic method on the transformed problem (4.4). The question arises as to whether there is any way to construct a method which manages to preserve both the Hamiltonian and scaling invariance properties for problems which possess them both.

We do note here however that if g is chosen to satisfy $g(q, -p) = -g(q, p)$ then we can construct methods that will preserve time reversal symmetries [107], see also Section 1.2 and [109, 88, 87, 162]. Note that Kepler's problem (4.4) also possesses this (discrete) symmetry. We shall not spend any more time discussing these methods in detail, we simply state that they exhibit many of the long time benefits of symplectic methods as discussed in Chapter 2.

Reich [136] and Hairer [83] combine the use of symplectic methods with adaptive time stepping through the use of the *Poincaré transformation* which we shall now introduce, see also [109]. Suppose that the original Hamiltonian H is time independent and therefore a conserved quantity of Hamilton's corresponding problem. Now, with $e := H(\mathbf{p}_0, \mathbf{q}_0)$, the constant 'energy' of the system, introduce a modified Hamiltonian $\hat{H}$ defined by

$$\hat{H}(\mathbf{p}, \mathbf{q}, t, e) = g(\mathbf{p}, \mathbf{q})\{H(\mathbf{p}, \mathbf{q}) - e\}. \tag{4.5}$$

The Hamiltonian system corresponding to $\hat{H}$ is then given by,

$$\begin{aligned}
\frac{d\mathbf{p}}{d\tau} &= -g\nabla_q H - \{H - e\}\nabla_q g, \\
\frac{d\mathbf{q}}{d\tau} &= g\nabla_p H + \{H - e\}\nabla_p g, \\
\frac{dt}{d\tau} &= g, \\
\frac{de}{d\tau} &= 0.
\end{aligned} \tag{4.6}$$

Here $(\mathbf{p}, t)^T$ and $(\mathbf{q}, e)^T$ are now conjugate variables in the extended phase space $\mathbb{R}^{2d} \times \mathbb{R}^2$. Along the *exact* solution of the problem $H(\mathbf{p}, \mathbf{q}) = e$, and thus the first two equations of (4.6) simply reduce in this case to a system transformed through the use of the Sundman transformation, as in (4.4) for example. We may thus think of the extra terms in (4.6) as perturbations of the Sundman transformed system which make the system Hamiltonian. We can obviously now apply a symplectic method with fixed time step $\Delta\tau$ to the transformed system (4.6) and the favourable properties (as in Section 2.3) of such methods should follow. However due to the third equation of (4.6) this can be thought of as being equivalent to an adaptive time stepping method in terms of t. So although our method yields numerical approximations on a non-equidistant temporal grid, it can be considered as a fixed step size, symplectic method applied to a different Hamiltonian system. This interpretation allows us to apply standard results and to draw conclusions such as good long time (in τ) error growth and near conservation of $\hat{H}$, depending crucially on g this implies near conservation of H. Notice in addition that the function g is performing exactly the rôle that it did in the Sundman transform method of performing time reparameterization. Therefore the scale invariant choices for g derived in (4.3) and Chapter 3 must also gives a Poincaré transformed system which is scale invariant without the need to scale τ. We have therefore developed a means for constructing methods which may be both symplectic and scale invariant.

4.4 Some methods, computations and comparisons

For simplicity we discretize (4.6) using the first-order symplectic Euler method (SEP). The Poincaré transformation does have the disadvantage that it destroys the separability property of the Hamiltonian and therefore the symplectic Euler method applied to this problem is now implicit. We use Newton iteration to solve the nonlinear equations here, however it is possible in this case to simply solve a quadratic equation [83].

For comparison we also form a *time-reversible*, second order, angular momentum conserving, explicit method by applying the second-order Lobatto IIIA-B [84] pair to the Sundman transformed system, this method is usually termed the adaptive Verlet method [94], see also [88, 109]. For the Kepler problem described earlier, with function g depending only on $\mathbf{q}$ this scheme (SVS) can be written as,

$$\begin{aligned}
\mathbf{q}_{n+1/2} &= \mathbf{q}_n + \frac{\Delta\tau}{2\rho_n}\mathbf{p}_n,\\
\rho_{n+1} &= \frac{2}{g(\mathbf{q}_{n+1/2})} - \rho_n,\\
\mathbf{p}_{n+1} &= \mathbf{p}_n - \frac{\Delta\tau}{2}\left\{\frac{1}{\rho_n} + \frac{1}{\rho_{n+1}}\right\}\frac{\mathbf{q}_{n+1/2}}{r^3_{n+1/2}},\\
\mathbf{q}_{n+1} &= \mathbf{q}_{n+1/2} + \frac{\Delta\tau}{2\rho_{n+1}}\mathbf{p}_{n+1},\\
t_{n+1} &= t_n + \frac{\Delta\tau}{2}\left\{\frac{1}{\rho_n} + \frac{1}{\rho_{n+1}}\right\},
\end{aligned}$$

where $r = \sqrt{q_1^2 + q_2^2}$, and an explanation for the reciprocal choice of time step update is given in [47].

In figure 4-1 we show the results of applying SE and SV to the untransformed Kepler problem (4.1) as well as SES (the symplectic Euler method applied to the Sundman transformed (non Hamiltonian) system), SVS and SEP. For this experiment we only integrate for 10 orbits of eccentricity 0.5 — a fairly simple problem. Straight away we see the correct order for the methods with both the fixed and variable step size formulations. The adaptive methods which preserve either the symplectic or time reversal properties can be seen to have better performance even for this problem where adaptivity is not vital for efficiency. However SES (the symplectic Euler method applied to the Sundman transformed system) which is neither symplectic nor time reversible demonstrates a definite reduction in performance compared both to SEP and in particular to SE itself.

We perform a similar experiment in figure 4-2, this time for a more difficult problem with a higher eccentricity of 0.9. Adaptive methods should begin to come into their own now. We see the desired result of the adaptive methods designed to preserve

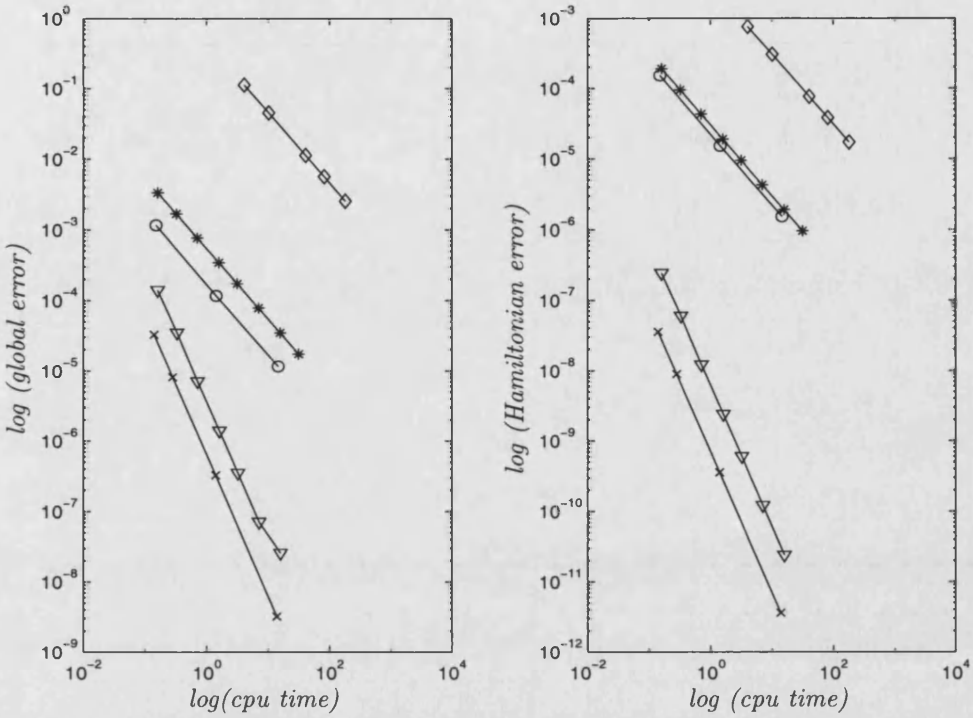


Figure 4-1: *Kepler's problem with eccentricity of 0.5 for 10 orbits. SE (⋆), SV (▽), SES (◇), SEP (○), SVS (×).*

symplecticity or reversibility performing well in comparison to their fixed step size counterparts. Again, SES is seen to perform poorly. Ruth's method [140, 146], which is symplectic and third order, applied to (4.1) is also included for comparison purposes.

In figure 4-3 we again perform computations for the problem with an eccentricity of 0.9, but now integrate over the much longer time scale of 1000 orbits. Again we see the improvements the use of adaptive time stepping affords, witness the close to two orders of magnitude improvement of SVS over SV. This example clearly demonstrates that for high accuracy the use of high-order methods appears to be beneficial. Note that SE and SES are omitted from this figure as both are totally uncompetitive. SE due to the fact that adaptive time stepping really is needed if low-order methods are to be efficiently used on this problem, and SES since the method is neither symplectic nor time-reversible and thus suffers in this long time simulation.

Finally in figure 4-4 we demonstrate the desirable linear error growth property of the methods applied to this problem which preserve symplecticity or time reversibility. We also include here the classical third-order Runge-Kutta method [106, 86], it can be seen to exhibit quadratic rather than linear error growth demonstrating the fact that for long time simulations a geometric integrator is to be preferred in many situations. Note that no effort has been made to choose fictive time steps in a consistent manner here, and therefore no conclusions regarding the relative accuracies of these methods should

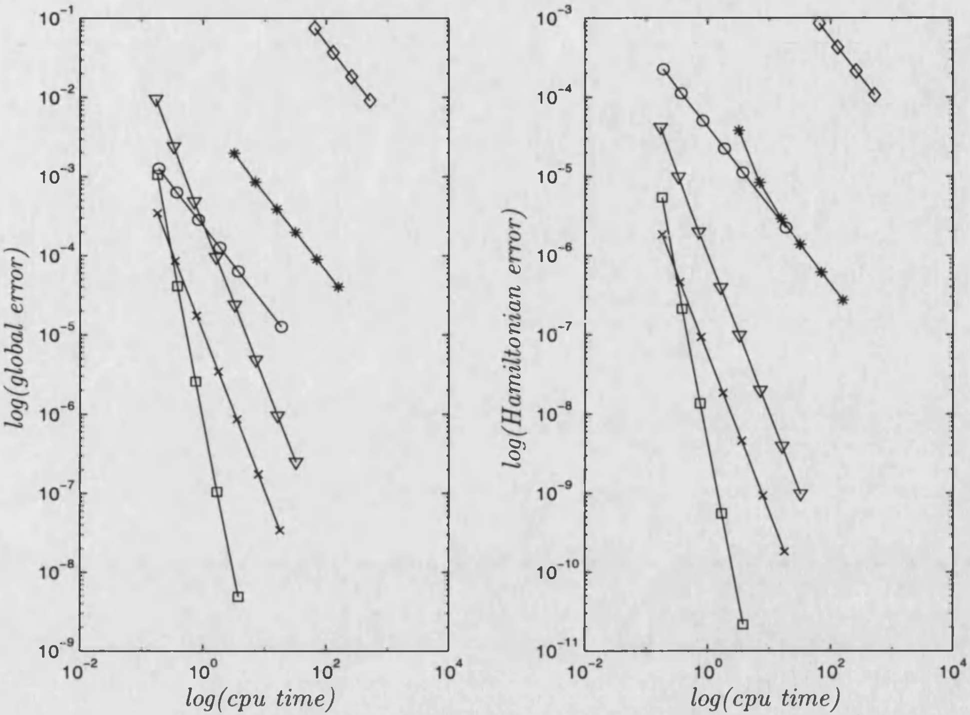


Figure 4-2: *Kepler's problem with eccentricity of 0.9 for 10 orbits. SE ($\star$), SV (∇), SES ($\diamond$), SEP ($\circ$), SVS ($\times$), Ruth ($\Box$).*

be inferred.

Similar experiments are carried out in [40]. They come to the similar conclusion that for Hamiltonian problems a code based on a high-order Gauss-Legendre with Poincaré transformation may out perform standard software.

4.5 Preservation of conservation laws under transformations of variables

Suppose that we have a system which possesses a conservation law L, so that

$$\frac{dL}{dt} = 0,$$

along solutions of the problem. For example the angular momentum (4.2) in Kepler's problem (4.1). Notice that if we now perform a transformation of the time variable through the use of the Sundman transform (3.2), (3.3) then the same conservation law holds for the new system in terms of τ, this is simply due to

$$\frac{dL}{d\tau} = \frac{dL}{dt}\frac{dt}{d\tau} = g\frac{dL}{dt} = 0.$$

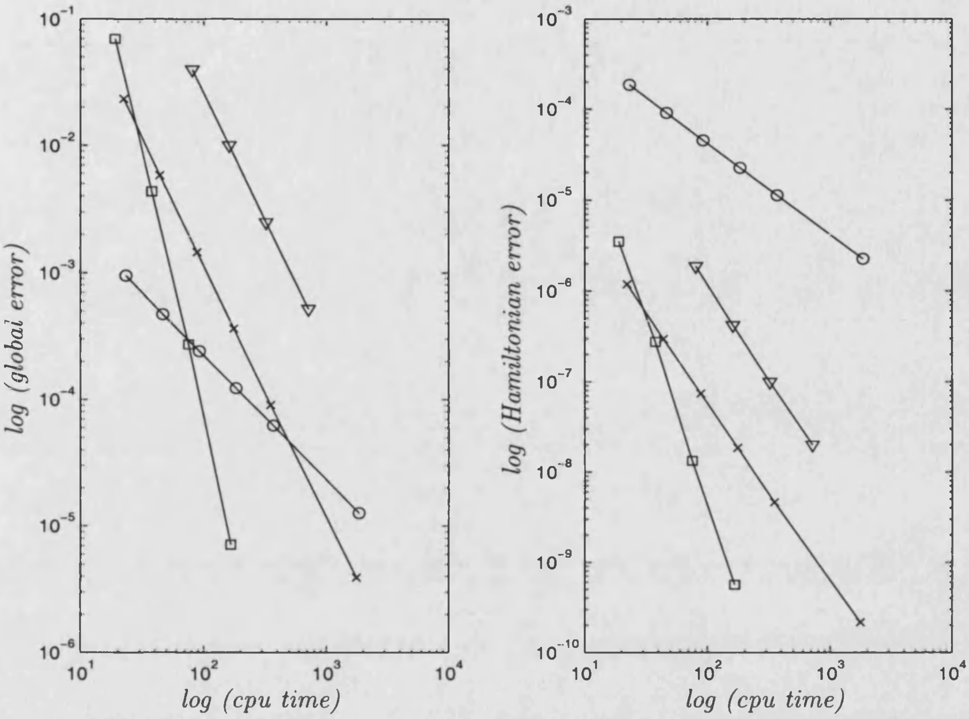


Figure 4-3: *Kepler's problem with eccentricity of 0.9 for 1000 orbits. SV (▽), SEP (○), SVS (×), Ruth (□).*

However, assuming the system is of Hamiltonian type and following a transformation induced by the Poincaré transform (4.5) the above argument no longer holds. We therefore ask the question of if and when conservation laws are inherited by the Poincaré transformed system. The intimate relationship which exists between Hamiltonian systems, symmetries and conservation laws (see Section 2.3.4) means that this is an important point to consider, as well as giving us an indication of how to proceed.

Lemma 4.1. *Suppose that a time independent Hamiltonian H with corresponding Hamiltonian (or Poisson) system (recall Section 2.3.2),*

$$\frac{d\mathbf{x}}{dt} = J(\mathbf{x})\nabla H(\mathbf{x}) = \{\mathbf{x}, H\},$$

has an invariant L (i.e. L Poisson commutes with H), then the Poincaré transformed system inherits the invariant L if the temporal adaptivity monitor function g also Poisson commutes with L.

Proof. First of all note that the conservation of L for the original system defined by the Hamiltonian H follows, using the chain rule, from

$$0 = \frac{dL}{dt} = \nabla L \cdot \dot{\mathbf{x}} = \nabla L \cdot J\nabla H = \{L, H\}, \tag{4.7}$$

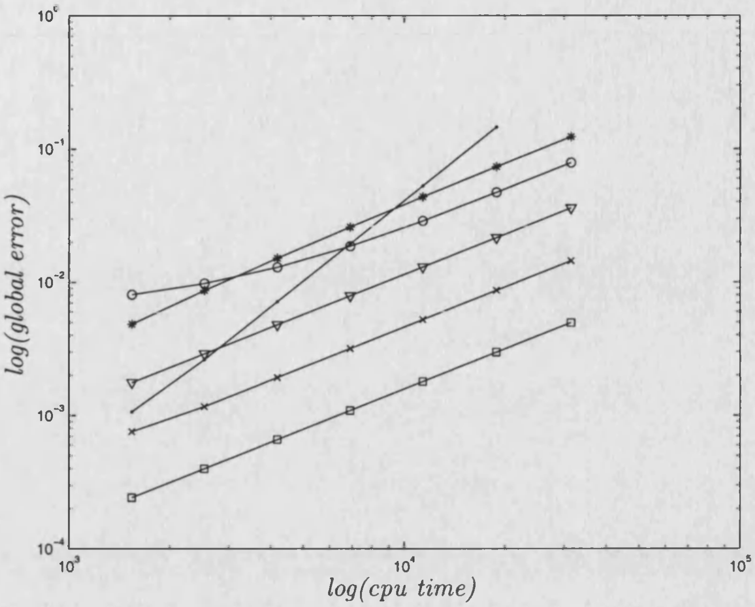


Figure 4-4: *Linear error growth of methods applied to Kepler's problem with eccentricity 0.5. SE ($\star$), SV ($\triangledown$), SEP ($\circ$), SVS ($\times$), Ruth ($\Box$). For comparison third-order classical Runge-Kutta ($\cdot$) is shown, clearly exhibiting quadratic error growth.*

that is from the Poisson commutativity of L and H. However, for the Poincaré transformed system (i.e. with Hamiltonian given by (4.5) in a similar manner we have,

$$\frac{dL}{d\tau} = \{L, \hat{H}\} = \{L, g(H-e)\}.$$

Now using the bilinearity property and Leibniz' rule for the Poisson bracket yields

$$\frac{dL}{d\tau} = g\left(\{L, H\} - \{L, e\}\right) + \{L, g\}(H-e).$$

However, since e is a constant, and $\{L, H\} = 0$ from (4.7), we are left simply with

$$\frac{dL}{d\tau} = \{L, g\}(H-e).$$

Therefore, we may conclude that L is an invariant of the new system if $\{L, g\} = 0$.

□

As an example where $\{L, g\} \neq 0$ leads to non-preservation of L consider the conservation of the angular momentum (4.2) in the Kepler problem (4.1), recall from Section 2.3.4 that via Noether's theorem this can be shown to be a consequence of the rotational invariance (i.e. O_2 symmetry) of the problem. Now consider using the temporal adaptivity monitor function given by

$$g = \left(q_1^2 + c\, q_2^2\right)^{3/4}, \quad c \in \mathbb{R}.$$

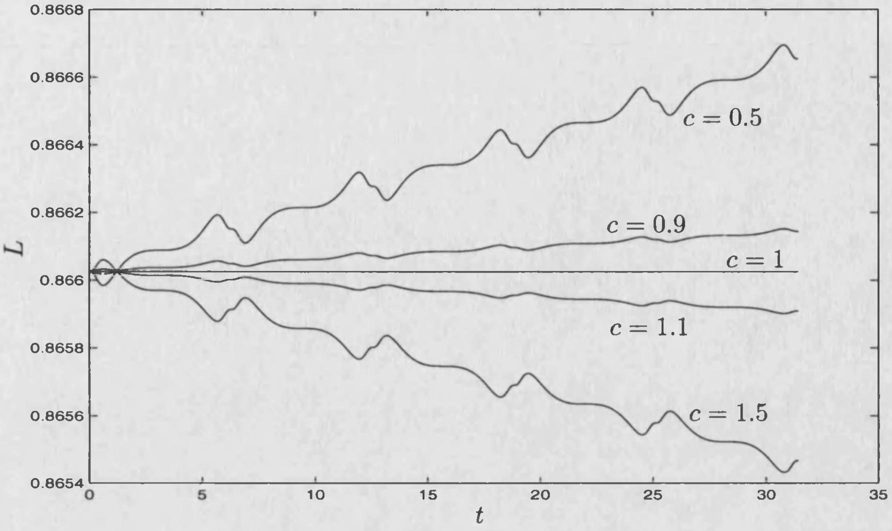


Figure 4-5: *Conservation of L for the rotationally symmetric choice of monitor function, and non-preservation for the non-rotationally symmetric choice.*

For $c = 1$ we are in the case investigated in the previous Section where, although it wasn't stated, L was automatically conserved by the geometric methods considered. However for $c \neq 1$ the transformed system is no longer rotationally invariant and correspondingly (also from Lemma 4.1) the new system is no longer guaranteed to possess an associated conservation law. This is due to the fact that the choice of g given by (4.5) does not Poisson commute with L for $c \neq 1$, i.e. the symmetry inherent in the original problem has been destroyed by the wrong choice of g. This result is verified in figure 4-5 where the Poincaŕe transformed system is integrated with the symplectic Euler method for different choices of c. As can be seen the discrete numerical solution does indeed conserve angular momentum for $c = 1$, whilst it does not for $c \neq 1$.

4.6 Summary of Chapter

In this Chapter we have demonstrated a means of extending the time transformation employed in Chapter 3 so that scale invariant symplectic methods may be constructed. The experiments carried out demonstrated that the symplectic (or time reversal) property was the important geometric feature to preserve in long time simulations of this problem with a periodic solution where singularities do not occur. However recall Example 3.4 which used exactly the problem used in this Chapter except the initial conditions where such that the planet fell into the Sun in finite time, here an adaptive time step is vital and the solution behaviour is governed by a scaling invariance rather than any symplectic property. We saw that a scaling invariant method performed extremely well, but a symplectic method (with fixed time step) would not be expected to work so well. This indicates that the techniques developed in this Chapter can yield methods which work well on problems which could either exhibit finite time singularities or rather evolve for long times. In addition the combination of both properties yields the possibility of preserving conservation laws which may otherwise be lost if only one of symplecticity or scaling (and indeed any other) invariance is preserved.

Chapter 5

Scale invariant methods for PDEs

5.1 Overview of Chapter

It is now natural to attempt to extend the methods and results of Chapter 3 to problems with more than one independent variable, i.e. to problems described by partial differential equations. Partial differential equation problems often involve a complex interaction between temporal and spatial structures. This can take many forms, but a common interaction concerns scalings, so that a change in the temporal scale of the solution is related to a change in the spatial scale. It is possible to capture this behaviour using an adaptive method based upon geometric ideas. We shall describe here a general method for adapting the spatial mesh and then show how geometric ideas can naturally be incorporated into it. In an analogous manner to the way in which we constructed an adaptive temporal mesh through a time reparameterization or transformation function, where the time t was described in terms of a differential equation in a fictive variable τ, we can think of a spatial mesh X as being a function of a fictive spatial variable ξ such that X satisfies a differential equation in ξ. Here we will assume that this function has a high degree of regularity (i.e. we progress beyond thinking of a mesh as a piecewise constant function of ξ.)

We shall primarily be interested in semi-discretizations in space (i.e. the method of lines [100]) and allow the mesh points to depend continuously on the time variable. We shall show that with the correct choice of adaptivity procedure (basically the correct choice of monitor function) the same scaling invariance property of the fully continuous PDE problem holds for the system of ODEs obtained following the semi-discretization. At this point the procedures and results obtained in Chapter 3 may then be employed to integrate the ODEs.

5.2 Scaling and adaptivity

5.2.1 Self-similarity and adaptive discretizations

We have shown in Chapter 2 that scaling invariance plays an important rôle in the theory and behaviour of the solutions to a partial differential equation. It is desirable that a numerical method to discretize such an equation should have a similar invariance principle. Ideally such a numerical method should possess *discrete self-similar solutions* which are scale invariant and which uniformly approximate the true self-similar solutions of the partial differential equation over all times. If these are global attractors (or at least have the same local stability as solutions of the underlying PDE) then we will have a numerical method which has the correct asymptotic properties, and for example, may also have excellent accuracy when approximating singular solutions.

Scaling invariance of a partial (or ordinary as seen in Chapter 3) differential equation and adaptivity of the spatial and temporal meshes fit very naturally together. This is because the use of a fixed mesh in a discretization automatically imposes an underlying spatial and temporal scale on the problem. This makes it impossible to consider scale invariant solutions. This difficulty disappears when we introduce adaptivity as now the spatial and temporal grids become part of the solution and can easily adjust to any appropriate length and time scale consistent with the underlying problem. When considering such an approach, it is natural to look at methods of *r-adaptivity* in which spatial mesh points are moved continuously throughout the solution procedure, rather than *h-adaptivity* in which new points are added or old points removed in a discontinuous manner, or the slightly different *p-adaptivity* in which the order of accuracy of the solution approximation is varied. The reason for doing this is that then the solution approximation, as well as the spatial and temporal mesh become one (large) dynamical system which has a wealth of structure, reflecting the underlying scalings of the original problem. This structure may then be analysed using dynamical systems techniques. A very general account of the interaction between adaptivity in space on a moving mesh for problems with a wide class of symmetries is given by the work of Dorodnitsyn [56, 57]. In this Chapter we shall look at the specific case of scaling symmetries and will call a numerical method which inherits these underlying symmetries *scale invariant.*

5.2.2 Coordinate transformations and semi-discrete self-similarity

The advantage of using an adaptive method which is invariant under the action of a scaling is that such methods should, if correctly designed, admit discrete self-similar solutions. If we conserve maximum principles or the stability of the underlying self-similar solution, then such numerical methods will have excellent asymptotic properties, as we shall see later on in this Chapter. However, the discrete self-similar solutions will not be the only solutions admitted by the numerical method and thus the effect of boundary conditions and arbitrary initial conditions may be taken into account. Thus, when applicable, the synthesis of adaptivity with symmetry invariance provides a flexible, general and powerful numerical tool.

To make things more precise, in this case consider a partial differential equation of the form

$$u_t = f(u, u_x, u_{xx}), \tag{5.1}$$

generalizations and extensions to higher dimensions etc. are straightforward. To discretize (5.1) consider the solution function $u(x,t)$ to be approximated on a spatial mesh $X_i(t)$ which is allowed to evolve with time. Let $U_i(t)$ denote the approximation to $u(x,t)$ at the mesh point $X_i(t)$, i.e.

$$U_i(t) \approx u(X_i(t), t).$$

We consider discretizations with a fixed number N of spatial mesh points $X_i(t)$, $i = 1, \ldots, N$. In a spatially adaptive numerical scheme the values of $X_i(t)$ are computed along with the solution $U_i(t)$. Suppose that, in the absence of boundary conditions, the differential equation (5.1) is invariant under the action of the scaling transformation

$$t \longrightarrow \lambda t, \quad x \longrightarrow \lambda^\beta x, \quad u \longrightarrow \lambda^\gamma u. \tag{5.2}$$

Now, consider the approximation $(U_i(t), X_i(t), t)$ to $u(x,t)$ obtained by our method, so that $(U_i(t), X_i(t), t)$ is the solution of a semi-discrete ODE system. We shall say that the semi-discretization is *scale invariant* if (again in the absence of boundary conditions) the set of points

$$(\lambda^\gamma U_i(t), \lambda^\beta X_i(t), \lambda t) \tag{5.3}$$

is also a solution of the semi-discrete ODE system defined by the numerical method.

Note that this definition can immediately be extended to that of a fully scale invariant discretization using the definitions and ideas of Chapter 3. In particular once an adaptive method of integrating the ODEs has been chosen, we require that whenever the set of points $(U_{i,n}, X_{i,n}, T_n)$ is a solution of the fully discrete system then so is $(\lambda^\gamma U_{i,n}, \lambda^\beta X_{i,n}, \lambda T_n)$. Where hopefully the use of notation should be fairly obvious, for example $U_{i,n} \approx u(X_{i,n}, T_n)$ where the T_n are a set of discrete time levels, etc.

Using the results of Chapter 3 as motivation, due to the scale invariance of the derived semi-discrete system we now consider *semi-discrete self-similar* solutions. Recall from

Chapter 2 that a self-similar solution of (5.1) invariant under (5.2) takes the form

$$u(x,t) = t^{\gamma}v(xt^{-\beta}),$$

where (in the one spatial dimension case) the function $v(y)$ satisfies an ordinary differential equation. In comparison, a semi-discrete self-similar solution possessing the same invariance must satisfy the condition

$$X_i(t) = t^{\beta}Y_i, \quad U_i(t) = t^{\gamma}V_i, \tag{5.4}$$

for constants Y_i and V_i. The existence of such a semi-discrete self-similar solution follows immediately from the scaling invariance condition (5.3). The vectors V_i and Y_i then satisfy algebraic equations obtained by substituting (5.4) into the semi-discretization and cancelling any t dependence. Now as we shall see for both finite difference (see also Theorem 5.1) and finite element discretizations of the porous medium equation later in this Chapter, it is easy to verify that the two operations of scaling and discretizing a PDE *commute*, with details identical to Lemmas 3.2 and 3.3 of Chapter 3. It follows that if the semi-discretization is consistent with the underlying PDE then the algebraic equations satisfied by Y_i and V_i are a consistent discretization of the ordinary differential equation satisfied by $v(y)$, and hence $V_i \approx v(Y_i)$. Observe that the error implicit in this approximation does not depend upon the value of n (i.e. on time). Therefore we may *uniformly* approximate the self-similar solution over arbitrarily long times, compare this with the results of Chapter 3, and see [29] and the example later on in this Chapter for a demonstration of this in practice. What we have said is that the property of being able to perform a symmetry reduction is preserved by the method, and by rescaling the method in terms of the discrete invariants implied by the scaling group the semi-discrete system may be reduced to a discretization of an ODE. The error induced during this process may also be rescaled, as described above, to be independent of the time variable. Therefore, as in Chapter 3 this results in a semi-discrete self-similar solution approximating the continuous self-similar solution uniformly in time.

The condition that (5.3) is a solution to the semi-discrete system for all $\lambda \geq 0$ gives a means of defining conditions for an adaptive mesh to respect the scaling invariance property. Observe that these are *global* conditions related to underlying scaling properties of the equation, rather than the usual *local* conditions of adaptivity in which (for example) we may choose to cluster mesh points in regions where an approximation to the local truncation error of the method is high. The reason for this choice of condition on the mesh is that it accurately reflects the underlying geometry of the problem.

As was witnessed in Chapter 3 the process of introducing adaptivity into a numerical scheme can be closely linked with rescaling. Suppose that τ and ξ are computational variables in the sense of [96]. We may consider an adaptive mesh $(X_{i,n}, T_n)$, on the underlying physical space (x, t) over which the PDE is defined, to be the image under a given mapping of a fixed mesh on the computational space (ξ, τ). This mapping or coordinate transformation may be given in terms of the maps $X \equiv X(\xi, \tau)$ and $T \equiv T(\tau)$. If the computational space is covered by a uniform mesh of spacing $(\Delta\xi, \Delta\tau)$ then we may use

$$T_n = T(n\Delta\tau), \quad X_{i,n} = X(i\Delta\xi, n\Delta\tau), \quad i,\, n \in \mathbb{Z}.$$

A similar procedure may be employed in higher spatial dimensions, see [97, 98] as well as Chapter 7. The differential equation (5.1) when expressed in terms of the computational variables then becomes, using a Lagrangian derivative,

$$u_\tau - u_x X_\tau = T_\tau f(u, u_x, u_{xx}), \tag{5.5}$$

where $u_x = u_\xi / X_\xi$ etc. This retains the same scaling invariance property as (5.1). An r-adaptive approach is then equivalent to discretizing the equation (5.5) on a uniform partition of the computational variables.

An essential part of this process is the determination of suitable functions $X(\xi, \tau)$ and $T(\tau)$. There is much arbitrariness about how this may be done, but we may be guided in our choice by the scaling invariance condition involving (5.3). In particular, if we have a set of conditions for the mesh which lead to the solutions $T(\tau)$, $X(\xi, \tau)$, and $U(\xi, \tau)$, then these conditions should also admit a rescaled solution of the form $\lambda T(\tau)$, $\lambda^\beta X(\xi, \tau)$, and $\lambda^\gamma U(\xi, \tau)$.

Theorem 5.1. *The standard finite difference operators given by*

$$\Delta_+ z_k = z_{k+1} - z_k \ , \quad \Delta_- z_k = z_k - z_{k-1}, \quad \Delta_0 z_k = z_{k+1/2} - z_{k-1/2},$$
$$\mathcal{E} z_k = z_{k+1}, \quad \Upsilon z_k = \frac{1}{2}(z_{k+1} + z_k),$$

for a real or complex sequence $\{z_k\}$, do not effect scaling, i.e. they all scale as z. As a consequence, assuming X and T have been chosen to be scaling invariant, (5.5) discretized using standard finite differences (constructed through successive applications of the difference operators above), is scaling invariant under (5.3).

Proof. On observing that the operators introduced in the statement of this theorem are all linear it is immediate that they have no effect on rescalings, e.g. suppose that $z_k \to \lambda^\alpha z_k$ for all k, then

$$\Delta_+ z_k = z_{k+1} - z_k \quad \longrightarrow \quad \lambda^\alpha z_{k+1} - \lambda^\alpha z_k = \lambda^\alpha \Delta_+ z_k.$$

Consequently any composition of these operators also does not effect scalings. Now suppose that we discretize (5.5) using these operators on a fixed and uniform grid in ξ and τ, of respective spacing $\Delta\xi$ and $\Delta\tau$. Then noting that by scaling invariance of the original PDE the function f must scale as $f \to \lambda^{\gamma-1} f$ under the appropriate scaling of its arguments, and also that the discretized T and X scale as, respectively, t and x. We have immediately that with fixed $\Delta\xi$ and $\Delta\tau$ the discretized version of every term in (5.5) scales as u, i.e. a factor λ^{γ} is present. As a consequence the discretized version of (5.5) is scale invariant, importantly with $\Delta\xi$ and $\Delta\tau$ held fixed. Therefore the resulting discrete set of equations is scale invariant. For example we may discretize (5.5) by

$$\frac{\Delta_+^t U_{i,n}}{\Delta\tau} - \frac{\Delta_0^x U_{i,n}/\Delta\xi}{\Delta_0^x X_{i,n}/\Delta\xi}\frac{\Delta_0^x X_{i,n}}{\Delta\tau} = f\left(U_{i,n}, \frac{\Delta_0^x U_{i,n}/\Delta\xi}{\Delta_0^x X_{i,n}/\Delta\xi}, \dots\right),$$

where the superscript x and t appearing here simply indicates whether the difference is being taken in the spatial or temporal direction. For fixed $\Delta\xi$ and $\Delta\tau$ the two terms on the left and the one on the right of this expression scale as u, as claimed and desired. □

Note that the related result for the case where only spatial derivatives are taken with this method and then the methods of Chapter 3 are employed (recall Lemmas 3.2 and 3.3) is straightforward to establish.

5.2.3 Mesh movement strategies

Lagrangian type methods

Strategies for calculating the mesh function X vary in the literature. One direct method is to use the Sundman time transformation method of Chapter 3 as motivation and to introduce a further function H such that

$$\frac{dX}{dt} = H(X, U). \tag{5.6}$$

This is a natural strategy for certain hyperbolic like equations since it could, with the correct choice of H, correspond to advecting the mesh along the flow of the solution. This strategy is adopted by Dorodnitsyn for a general class of groups [56, 57]. To give a scale invariant scheme we require that $H(X, U)$ satisfy,

$$H(\lambda^{\beta} X, \lambda^{\gamma} U) = \lambda^{\beta-1} H(X, U),$$

or equivalently, upon differentiating with respect to λ and setting $\lambda = 1$, the function H should satisfy the linear hyperbolic partial differential equation

$$\beta X H_X + \gamma U H_U = (\beta - 1) H.$$

A disadvantage of this approach it that it is rather local in form in that the individual mesh points themselves are moved. As was discussed in Chapter 1 we generally prefer methods which take into account global information, we therefore now introduce an alternate method for controlling the mesh movement where now the density of the mesh points is controlled.

Equidistribution based methods

The approach we prefer to use here, in one-dimension (for a similar method in higher dimensions see Chapter 7), is based on the framework which is developed in [95, 96]. Equidistribution can loosely be thought of as a process for changing the *density* of the mesh points in response to the solution (as opposed to Lagrangian type methods which tend to change the mesh points themselves). In contrast to parts of the discussion of Section 5.2.2, we shall now simply consider semi-discretizations on a spatial mesh, maintaining a continuous time dependence in the problem.

As we have mentioned many times before we think here of adaptivity in terms of a time varying coordinate transformation between a *computational* and a *physical* domain. As such we define the physical mesh, that is the mesh upon which the physical problem is posed, in terms of a particular realization of a mesh function $X(\xi, t)$ which maps a computational (fictive) coordinate $\xi \in [0,1]$ one-to-one and onto a physical coordinate x (which we assume without loss of generality to be in $[0,1]$) such that

$$X\left(\frac{j}{N}, t\right) = X_j(t), \qquad j = 0, \ldots, N.$$

Therefore our $N+1$ spatial (x) mesh points $X_j(t)$, which are permitted to vary with time, are simply the image of a uniform computational mesh under a time dependent mesh transformation function.

We now introduce a *monitor function* $M(x, u, u_x)$ which classically represents some measure of computational difficulty in the problem, and therefore characterizes regions of the domain where it would be desirable to employ higher numerical resolution. For example M may represent the local truncation error of the approximation, or some quantity such as the arclength of the solution, i.e.

$$M(x, u, u_x) = \sqrt{1 + u_x^2}. \tag{5.7}$$

Even more simply, and especially useful when we need to cluster points where the solution is large, for example in solutions exhibiting singular behaviour, we may take

$$M(x, u, u_x) = u^\sigma, \quad \sigma \in \mathbb{R}^+.$$

We can now invoke the principle of equidistribution by defining our mesh function X through the relation,

$$\int_0^X M\,dx = \xi \int_0^1 M\,dx = \xi f(t), \tag{5.8}$$

where f is a function of time only. Observe that this principle is closely related to the geometric idea of conserving the function M over mesh intervals, and through the correct choice of M we may exploit this feature to help design meshes which automatically retain invariants of the evolution [29, 35]. Differentiating (5.8) with respect to ξ we see that the mesh-density X_ξ satisfies the equation

$$X_\xi = \frac{1}{M}\int_0^1 M\,dx, \tag{5.9}$$

so that the mesh density is inversely proportional to M. Differentiated again gives the following partial differential equation for the mesh — referred to as MMPDE1 (moving mesh partial differential equation 1),

$$(MX_\xi)_\xi = 0. \tag{5.10}$$

This equation may then be solved for the mesh by discretizing the function X appropriately [95, 96].

In practice equation (5.10) can lead to instabilities [95, 96], even if a scale invariant monitor function is used [31]. Furthermore, it requires the use of a mesh which is initially equidistributed and this can be hard to achieve. To allow both for arbitrary initial meshes and to stabilize the system, a relaxed form of (5.10) is often used. One of the more popular versions is the so called MMPDE6 which takes the form

$$\varepsilon X_{t\xi\xi} = -\left(MX_\xi\right)_\xi, \tag{5.11}$$

where $\varepsilon > 0$ is a small relaxation parameter. This equation has been used with great success in many applications, see [95, 96]. See also [31] for a study of its applications to singular problems. Additional theoretical results including the impossibility of mesh points crossing and mesh smoothness issues are also established in [95, 96].

To solve the original problem, both the PDE for the mesh function X and the (coupled) PDE for $u(x,t)$ are discretized. Note as we proved in the case of adaptivity for scale invariant ODEs in Chapter 3, it is wise not to use a lower order discretization to solve for the mesh — something which is common in applications. This is due to the close coupling of the solution with space and time (especially for problems where scalings are important) which means that a reduction in accuracy of the solution of the mesh equations may convert to a reduction in accuracy of the solution of the underlying

PDE.

The new coupled system may or may not inherit the qualitative features of the original problem. The geometric integration viewpoint is to produce a mesh in which the mesh equation inherits (or equivalently does not destroy) as many qualitative features as possible. As an important example, we may seek a mesh so that the new system has the same scaling invariance as the original. As the mesh is governed by the monitor function M, this problem reduces to that of choosing M such that the coupled system is invariant with respect to the same transformation group as the original equation. By doing this we ensure that the resulting numerical method itself inherits the scaling symmetry structure of the underlying PDE which the choice of M preserves. It is possible to do this for a wide variety of problems with relatively simple choices of M leading to some elegant scaling invariant methods as we shall see.

It is simple to see that (5.10) is scale invariant provided M satisfies a relation of the form,

$$M(\lambda^\beta x, \lambda^\gamma u, \lambda^{\gamma-\beta} u_x) = \lambda^\delta M(x, u, u_x) \tag{5.12}$$

where the value of δ can be very general, meaning that (5.10) can be invariant for a wide variety of different scalings. We see significantly that the monitor function $M = u^\sigma$ satisfies this condition for any choice of β and γ, whereas the arc-length monitor function (5.7) only satisfies it in the very restrictive case of $\beta = \gamma$. Thus arc-length does not fit in well with the theory of invariant methods. Although it can be seen to be approximately invariant in regions where u_x is large, and so it may well still have its uses.

The scale invariance of MMPDE6 (5.11) can also be ensured by a suitable choice of M. This must now satisfy the more restrictive condition that,

$$M(\lambda^\beta x, \lambda^\gamma u, \lambda^{\gamma-\beta} u_x) = \lambda^{-1} M(x, u, u_x). \tag{5.13}$$

The distinction between the two conditions (5.12) and (5.13) is not important if a single scaling group acts on the system. However in problems, such as the linear heat equation, where several independent scaling groups may act, (5.10) is invariant under all such actions whereas (5.11) will (in general) only be invariant under the action of a single scaling group, in which case a decision as to what to aim to preserve must be taken.

5.3 Symmetry and the maximum principle

Before considering numerical discretizations of scale invariant problems, it is worthwhile also looking at the rôle played by the maximum principle as the combination of the maximum principle and scaling invariance can tell us a great deal about the asymptotic behaviour of partial differential equations. Suppose that we have a partial differential equation from which may be derived a semi-group operator φ_t such that if $u(x, 0)$ is some initial data then $u(x, t) = \varphi_t(u(x, 0))$. Such a partial differential equation has a strong maximum principle if the ordering of solutions is preserved under the action of the semi-group [134]. Thus if $u(x, 0) < v(x, 0)$ for all x then $\varphi_t(u) < \varphi_t(v)$ for all x and $t > 0$. Many parabolic partial differential equations (for example the nonlinear heat equation $u_t = u_{xx} + f(u)$) satisfy strong maximum principles, these are used extensively throughout the analysis of these problems. Such maximum principles are invaluable when studying the dynamics of the equation. For example, if $v(x, t)$ is a known solution which is bounded above and which satisfies the partial differential equation and if $u(x, 0) < v(x, 0)$, then we have that $u(x, t)$ is also bounded above. Such an exact solution could easily be a self-similar solution. Furthermore, if v_1 and v_2 are two self-similar solutions such that $v_1 \to v_2$ as $t \to \infty$ then if $v_1(x, 0) < u(x, 0) < v_2(x, 0)$ we deduce immediately that $u \to v_2$, i.e. the solution u converges to self-similarity (c.f. Section 2.2.2).

Techniques similar to this are described in [172] to prove the L_1 global attractivity of the self-similar solution of the porous medium equation, although there are considerable additional analytic difficulties due to the existence of a non-regular interface where differentiability of the solution is lost and the equation is only satisfied in a weak sense. It is anticipated that a numerical method which has both a strong maximum principle and discrete self-similar solutions will, similarly, give the correct global asymptotic behaviour of the underlying partial differential equation. This is precisely the type of result we seek to achieve when using a geometric integration approach, and we shall see this in the following Sections.

5.4 The porous medium equation

5.4.1 Background theory

Consider here the porous medium equation (PME) given by

$$u_t = \frac{1}{m}(u^m)_{xx} \equiv (u^{m-1}u_x)_x. \tag{5.14}$$

Recall that the case $m = 2$ was considered in Chapter 2. Note that much of the published literature on this equation poses it in the form

$$v_t = (v^m)_{xx}, \tag{5.15}$$

however it is easy to show that (5.14) and (5.15) are equivalent through the scaling,

$$u = m^{1/(m-1)}v$$

and all standard results follow.

The PME, one of the simplest examples of nonlinear degenerate diffusion, arises in the study of the diffusion of a perfect gas through a porous medium under the action of Darcy's law which relates velocity to pressure gradient [61], as well as many other applications, see [172]. Notice that in the case $m = 1$ (5.14) actually represents the linear heat equation considered in Chapter 2. If $m < 1$ then the diffusion 'coefficient' $u^{m-1} \uparrow \infty$ as $u \downarrow 0$ and this case is often termed *fast diffusion* [105]. However the case we consider here is when $m > 1$, often called *slow diffusion* since $u^{m-1} \downarrow 0$ as $u \downarrow 0$. This case has found many applications since it removes the infinite speed of propagation (a shortcoming for many modelling processes) of the linear heat equation.

From now on we shall exclusively be considering the problem with $m > 1$ and initial conditions given by

$$u_0(x) \geqslant 0, \quad u_0 \in L_1(\mathbb{R}). \tag{5.16}$$

An example of the evolution of a solution to (5.14) for $m = 2$ is given in the top left of figure 5-2. An introduction to the theory of this equation is given in [172].

The PME is parabolic at those points where $u > 0$ and degenerates when $u = 0$, we call it a degenerate parabolic equation. As a consequence of this we do not expect to have classical solutions when the initial data takes the value zero at some points, which is the case that shall be considered here. We therefore need to introduce the concept of *generalized (or weak) solutions* which include classical solutions where appropriate, i.e. when they exist. The solutions we shall be looking at are actually classical *a.e.* Much rigorous work has been carried out on this problem, see [9, 69, 102, 131, 171, 172] for example. Existence and uniqueness results have been established given that $u^0(x) \in L_1(\mathbb{R})$, as well as strong regularity results. Also, of special interest here, if $u^0(x)$ has compact support given by the interval $[x_L(0), x_R(0)]$, then the solution $u(x,t)$ also has compact support on the expanding interval $[x_L(t), x_R(t)]$, where for $t_1 < t_2$ we have

$$[x_L(t_1), x_R(t_1)] \subset [x_L(t_2), x_R(t_2)], \tag{5.17}$$

see figure 5-2 for example, as well as [9] for a discussion of this result. Therefore we have sharply defined interfaces $x_L(t)$ and $x_R(t)$, where it may be shown (again see [9]) that

$$\dot{x}_L = -\frac{1}{m-1}(u_L^{m-1})_x, \quad \dot{x}_R = -\frac{1}{m-1}(u_R^{m-1})_x. \tag{5.18}$$

Note also that since a maximum principle holds for this second-order parabolic problem [134], for non-negative initial data it is immediate that the solution remains non-negative for all time.

This equation admits four continuous transformation groups, the two groups of translations in time and space, and the two-dimensional vector space of scaling symmetry groups spanned by the operators

$$\mathbf{X}_1 = t\frac{\partial}{\partial t} + \frac{1}{2}x\frac{\partial}{\partial x} \quad \text{and} \quad \mathbf{X}_2 = t\frac{\partial}{\partial t} - \frac{1}{m-1}u\frac{\partial}{\partial u}.$$

We now follow through the procedure which was carried out for the case $m = 2$ in Chapter 2 and look for solutions invariant under the transformation $X = k_1X_1 + k_2X_2$. Omitting details we look for a solution of the form

$$u(x,t) = t^{\gamma}\tilde{u}(xt^{-\beta}), \tag{5.19}$$

where, again as in Chapter 2, the conservation of mass property of the solutions we are interested in fixes the constants to be

$$\beta = \frac{1}{m+1}, \quad \gamma = -\frac{1}{m+1}.$$

Therefore,

$$\mathbf{X} = t\frac{\partial}{\partial t} + \frac{1}{m+1}x\frac{\partial}{\partial x} - \frac{1}{m-1}u\frac{\partial}{\partial u}. \tag{5.20}$$

Substituting (5.19) into (5.14), reducing to an ODE and solving as in Chapter 2 we arrive at the self-similar solution

$$\hat{u}(x,t) = t^{\gamma}\left(C - \frac{m-1}{2(m+1)}y^2\right)_+^{1/(m-1)}, \quad y = xt^{-\beta}, \tag{5.21}$$

which we again call the Barenblatt-Pattle solution, see [15]. As was mentioned in Chapter 2, C is a free constant which characterizes the mass of the self-similar solution. The following result illustrates the importance of this special solution.

Theorem 5.2. *Let $u(x,t)$ be a solution of (5.14) and (5.16) with integral I and centre of mass x_0. Then if $\hat{u}(x,t)$ is the self-similar solution with the same integral and centre of mass,*

$$t^{1/(m+1)}|u - \hat{u}| \to 0, \quad \textit{as} \quad t \to \infty, \tag{5.22}$$

uniformly with respect to x.

Proof. See [102] for the original proof in one-dimension, however see [9, 69, 102, 131, 171, 172] for alternative methods of proof, discussions of improvements on this result and extensions to higher dimensions. □

Note since $u = \mathcal{O}\left(t^{-1/(m+1)}\right)$ as $t \to \infty$, the factor $t^{1/(m+1)}$ in the above is simply a normalizing factor.

5.4.2 Invariant spatial discretizations

To discretize the problem described by (5.14) we introduce an adaptive mesh $X(\xi, t)$ such that $X(0,t) = x_L$ and $X(1,t) = x_R$. To determine X we use a monitor function and a moving mesh partial differential equation. As the evolution of the porous medium equation is fairly gentle [29] it is possible to use the mesh equation MMPDE1 without fear of instability [95, 96] in the mesh. This then allows a wide possible choice of scale invariant monitor functions of the form $M(u) = u^\sigma$, for any $\sigma \in \mathbb{R}$. A convenient function to use is

$$M(u) = u$$

the choice of which is strongly motivated by the conservation law

$$\int_{x_L}^{x_R} u \, dx = C.$$

Here C is a constant (the mass of the solution) which we can take to equal 1 without loss of generality. Setting $M = u$ and $C = 1$ in the equidistribution principle yields,

$$\int_{x_L}^{X} u \, dx = \xi \tag{5.23}$$

so that differentiation with respect to ξ we have,

$$uX_\xi = 1, \tag{5.24}$$

as the equation for the mesh which we will discretize. Note that this is invariant under the group action $u \to \lambda^{-1/(m+1)}u$, $X \to \lambda^{1/(m+1)}X$. Now, differentiating (5.23) with respect to t gives

$$0 = X_t u + \int_{x_L}^{X} u_t \, dx = X_t u + \int_{x_L}^{X} (u^{m-1}u_x)_x \, dx = u\left[X_t + \frac{1}{m-1}(u^{m-1})_x\right].$$

Thus, for the continuous problem we also have that X satisfies the Lagrangian equation

$$X_t = -\frac{1}{m-1}(u^{m-1})_x, \tag{5.25}$$

which is also invariant under the scaling (5.20). Note here that we have the early encouragement that we have exactly recreated the true velocity of the interface (5.18). Therefore any mesh point placed on the interface will follow exactly the correct evolution (in terms of the numerical approximation to u rather than the true u of course). Recall our earlier discussion on Lagrangian mesh movement strategies. If we had decided to use this technique then from the earlier theory we could have written down this result straight away, either through the interface velocity result, or simply by considering the flux associated with the continuous problem.

Lemma 5.1. *The monitor function motivated to satisfy conservation of mass, i.e. $M = u$ (or any constant multiple of u), is the only choice which gives the correct interface velocity.*

Proof. Suppose that the monitor function is an arbitrary function of u

$$M = f(u).$$

Equidistribution implies that

$$f(u)x_t + \int^x f_t \, dx = 0.$$

Enforcing the correct interface velocity

$$x_t = -\frac{1}{m-1}\left(u^{m-1}\right)_x,$$

gives

$$\frac{f(u)}{m-1}\left(u^{m-1}\right)_x = \int^x f_t \, dx = \int^x f'(u)\, u_t \, dx = \int^x f'(u)\ \left(u^{m-1}u_x\right)_x \, dx.$$

Differentiating with respect to x yields,

$$\begin{aligned} &\frac{1}{m-1}\left\{\left(u^{m-1}\right)_x f_x(u) + \left(u^{m-1}\right)_{xx} f(u)\right\} = \left(u^{m-1}u_x\right)_x f'(u) \\ \Longrightarrow\ &\frac{1}{m-1}u_x\left(u^{m-1}\right)_x f_x(u) + \left(u^{m-2}u_x\right)_x f(u) = \left\{\left(u^{m-1}\right)_x u_x + u^{m-1}u_{xx}\right\} f'(u) \\ \Longrightarrow\ &f'(u)\left\{u_x\left(u^{m-1}\right)_x\left[\frac{1}{m-1} - 1\right] - u^{m-1}u_{xx}\right\} = \left(u^{m-2}u_x\right)_x f(u). \end{aligned}$$

Now, upon separating variables and integrating we have

$$\begin{aligned}\int \frac{df}{f} &= \int \frac{\left(u^{m-2}u_x\right)_x}{\frac{m-2}{m-1}u_x\left(u^{m-1}\right)_x + u^{m-1}u_{xx}}\,du \\ &= \int \frac{\left((m-2)u^{m-3}u_x^2 + u^{m-2}u_{xx}\right)}{(m-2)u^{m-2}u_x^2 + u^{m-1}u_{xx}}\,du \\ &= \int \frac{du}{u},\end{aligned}$$

from which we deduce the desired result that

$$f(u) = cu,$$

for some constant c. □

Now consider posing the problem in the computational domain and define $w(\xi, t) = u(x(\xi, t), t)$, then using (5.14), (5.24) and (5.25) we have,

$$w_t = u_t + u_x x_t = u_t - \frac{1}{m-1}u_x(u^{m-1})_x = (m-2)u^{m-2}u_x^2 + u^{m-1}u_{xx}.$$

Where $w_\xi = u_x x_\xi = u_x/u$ and so $u_x = uw_\xi = ww_\xi$. We may say further that,

$$u_{xx} = u_x w_\xi + u w_{\xi x} = w(w_\xi)^2 + w^2 w_{\xi\xi} = w(ww_\xi)_\xi.$$

Substituting into the above yields,

$$w_t = \frac{w^2}{m}(w^m)_{\xi\xi}. \tag{5.26}$$

Notice that this equation has no dependence on the variable x, i.e. the evolution of the solution has been decoupled from the movement of the mesh and if we wish we may consider the two problems separately. Now notice that (5.26) is invariant under the transformation $t \longrightarrow \lambda t$, $w \longrightarrow \lambda^{-1/(m+1)}w$ and so we can look for a self-similar solution in the computational domain by considering a solution of the form,

$$w(\xi, t) = (t + c)^{-1/(m+1)}\theta(\xi), \tag{5.27}$$

where c is an arbitrary constant, and the function θ satisfies the following ODE in ξ

$$\theta^2 \frac{d^2}{d\xi^2}(\theta^m) = -\frac{m}{m+1}\theta. \tag{5.28}$$

For simplicity we consider the case $m = 2$ here. In this case the left hand side of (5.28) may be expanded, and upon multiplying through by θ'/θ we have

$$\theta\frac{d\theta}{d\xi}\frac{d}{d\xi}\left(\theta\frac{d\theta}{d\xi}\right) = -\frac{1}{3}\frac{d\theta}{d\xi}.$$

Integrating this expression up once yields the relation

$$\theta\frac{d\theta}{d\xi} = \sqrt{C - \frac{2}{3}\theta},$$

where C is a constant of integration. Now on separating variables and integrating we have

$$\sqrt{C - \frac{2}{3}\theta}\left\{C + \frac{1}{3}\theta\right\} = D - \frac{1}{3}\xi,$$

where D is a second constant of integration. Now assume that at $\xi = 0$ and $\xi = 1$ we have $\theta = 0$, which yields the relations

$$C^{3/2} = D, \quad \text{and} \quad -C^{3/2} = D - \frac{1}{3},$$

where we have taken the negative root in the second case. This gives us the values

$$D = \frac{1}{6}, \quad \text{and} \quad C = \left(\frac{1}{6}\right)^{2/3},$$

and therefore (5.28) has a solution given by the implicit relation

$$\left(\left(\frac{1}{6}\right)^{2/3} - \frac{2}{3}\theta(\xi)\right)^{1/2}\left(\left(\frac{1}{6}\right)^{2/3} + \frac{1}{3}\theta(\xi)\right) = \frac{1}{3}\left(\frac{1}{2} - \xi\right). \tag{5.29}$$

We shall see now that an explicit solution, found after a change of variables, may be given in closed form. We consider the behaviour of solutions to the new continuous problem (5.26), again in the case of $m = 2$ for simplicity. In terms of the physical coordinate x we have the convergence result Theorem 5.2. We shall now extend this result to give us a convergence result for the analytic solution to the continuous problem in the computational domain.

Theorem 5.3. *Let $w(\xi, t)$ be a solution of (5.26), which we assume without loss of generality to have first integral one and centre of mass zero. Then if $\theta(\xi)$ is a solution to (5.28) we have*

$$t^{1/3}|w(\xi,t) - t^{-1/3}\theta(\xi)| \to 0, \quad \text{as} \quad t \to \infty, \tag{5.30}$$

uniformly with respect to ξ.

Proof. Recall the Barenblatt-Pattle self-similar solution to (5.14),

$$\hat{u}(x,t) = t^{-1/3}B(y), \quad B(y) = \left(\left(\frac{3}{32}\right)^{1/3} - \frac{1}{6}y^2 \right)_+, \quad y = xt^{-1/3}, \tag{5.31}$$

with first integral 1, and support on $[-L, L]$,

$$L = \sqrt{6\left(\frac{3}{32}\right)^{1/3}}.$$

From Theorem 5.2 we know that for a solution u of (5.14) we have convergence to the Barenblatt-Pattle self-similar solution of the same first integral and centre of mass, i.e. we have

$$t^{1/3}|u(x,t) - \hat{u}(x,t)| \to 0, \quad \text{as} \quad t \to \infty, \tag{5.32}$$

uniformly with respect to x. Recall that $u(x,t) \equiv w(\xi, t)$ and hence notice immediately how close (5.32) is to the result we are attempting to establish here. We are simply left to show that $\theta(\xi) := B(y)$ is the solution to (5.28).

Now the equidistribution principle invoked above gave us $ux_\xi = 1$, therefore on making the assumption (which we know to be valid asymptotically in time) that we have the Barenblatt-Pattle self-similar solution in physical space, (i.e. $u \equiv \hat{u}$) we can derive a transformation between the computational coordinate ξ and the physical similarity variable y. That is,

$$y_\xi \left\{ \left(\frac{3}{32}\right)^{1/3} - \frac{y^2}{6} \right\} = 1.$$

This can be integrated up to give

$$\left(\frac{3}{32}\right)^{1/3} y - \frac{y^3}{18} = \xi + E,$$

for some constant of integration E. The assumption of symmetric data ($y(1/2) = 0$) yields $E = -1/2$, again this can be made more general and rigorous by allowing a translation of the x coordinate and noting the asymptotic behaviour of arbitrary solutions u. Therefore we have

$$\xi = \frac{1}{2} + \left(\frac{3}{32}\right)^{1/3} y - \frac{y^3}{18}. \tag{5.33}$$

The inverse of this transformation between computational and physical coordinates is therefore given by the following root (the one symmetric about $\xi = 1/2$) of this cubic,

$$y(\xi) = -\frac{1}{4}\alpha^{1/3} - \frac{3}{2}\left(\frac{6}{\alpha}\right)^{1/3} - i\frac{\sqrt{3}}{2}\left(\frac{1}{2}\alpha^{1/3} - 3\left(\frac{6}{\alpha}\right)^{1/3}\right),$$

where

$$\alpha \equiv \alpha(\xi) = 36 - 72\xi + 72\sqrt{\xi(\xi-1)}.$$

Note that consideration of the cubic discriminant (see [1]) demonstrates that all roots of (5.33) are real, for $\xi \in [0,1]$. Finally, $\theta(\xi)$ is given by substituting this expression for y into the $B(y)$ which appeared in the Barenblatt-Pattle solution, that is

$$\theta(\xi) = \left(\frac{3}{32}\right)^{1/3} - \frac{1}{6}y^2. \tag{5.34}$$

A plot of θ and y is given in figure 5-1. We now verify that (5.34) indeed satisfies (5.28). With the notation

$$\beta \equiv \beta(\xi) = 6 - 32\xi + 32\xi^2, \quad \gamma \equiv \gamma(\xi) = 1 - 18\xi + 48\xi^2 - 32\xi^3,$$

and following some manipulation, from (5.34) we have

$$\left(\theta^2\right)'' = -\frac{36^3}{18\alpha^{10/3}\sqrt{\xi(-1+\xi)}}\left\{\left(\gamma + \beta\sqrt{\xi(-1+\xi)}\right)\left[\left(-6^{4/3} + \alpha^{2/3}\right) + i\sqrt{3}\left(6^{4/3} + \alpha^{2/3}\right)\right]\right\}.$$

Multiplying this result by θ given in (5.34), and again after much manipulation we find, in accordance with (5.28), the desired result that

$$\theta\left(\theta^2\right)'' = -\frac{2}{3}.$$

Note that due to the way in which ξ appears above, the manipulations required to derive this result are somewhat simplified by employing the substitution

$$\xi = \sin^2(\varphi), \quad \varphi \in \left[0, \frac{\pi}{2}\right],$$

and then making use of standard trigonometric identities. □

Similarly, the two equations describing the mesh movement, (5.24) and (5.25), are invariant under the transformation

$$t \longrightarrow \lambda t, \quad x \longrightarrow \lambda^{1/(m+1)}x, \quad u \longrightarrow \lambda^{-1/(m+1)}u,$$

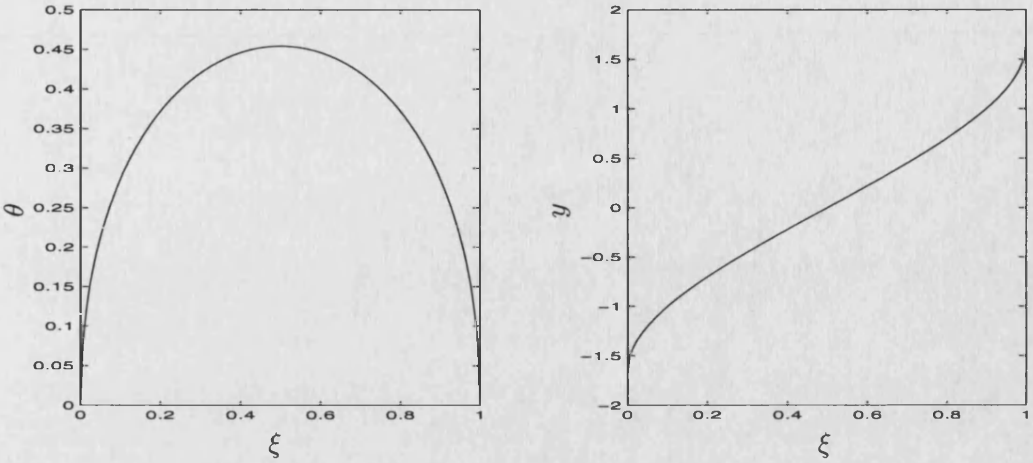


Figure 5-1: *Derived reduction functions θ and y plotted against ξ for the porous medium equation .*

treating X in the same manner as x here. Then the mesh has the self-similar form

$$\hat{X}(\xi,t) = (t+C)^{1/(m+1)}Y(\xi), \tag{5.35}$$

which substituting in to (5.24) and (5.25), and making use of the expression (5.27) for $u \equiv w$, gives the following system for $Y(\xi)$

$$\theta\frac{dY}{d\xi} = 1, \quad \frac{1}{m+1}Y = -\theta\frac{d\theta}{d\xi}, \tag{5.36}$$

which straight away can be seen to be consistent with (5.28).

Now, consider semi-discretizations of (5.24) and (5.26) so that we introduce discrete approximations $W_i(t)$ and $X_i(t)$ to the continuous functions $w(\xi,t)$ and $X(\xi,t)$ over the computational mesh

$$\xi = \frac{i}{N}, \quad i = 1,\ldots,(N-1),$$

with $W_0(t) = W_N(t) = 0$. A simple centred semi-discretization of (5.26) is given by

$$\frac{dW_i}{dt} = \frac{N^2}{m}W_i^2\left(W_{i+1}^m - 2W_i^m + W_{i-1}^m\right), \quad i = 1,\ldots,(N-1). \tag{5.37}$$

To define the mesh X_i we discretize (5.24) to give the algebraic system

$$(X_{i+1} - X_i)(W_{i+1} + W_i) = \frac{2}{N}, \quad i = 1,\ldots,(N-1). \tag{5.38}$$

We observe that this procedure has the geometric property of automatically conserving the discrete mass

$$\sum_{i=0}^{N-1}(X_{i+1} - X_i)\left(W_{i+1} + W_i\right). \tag{5.39}$$

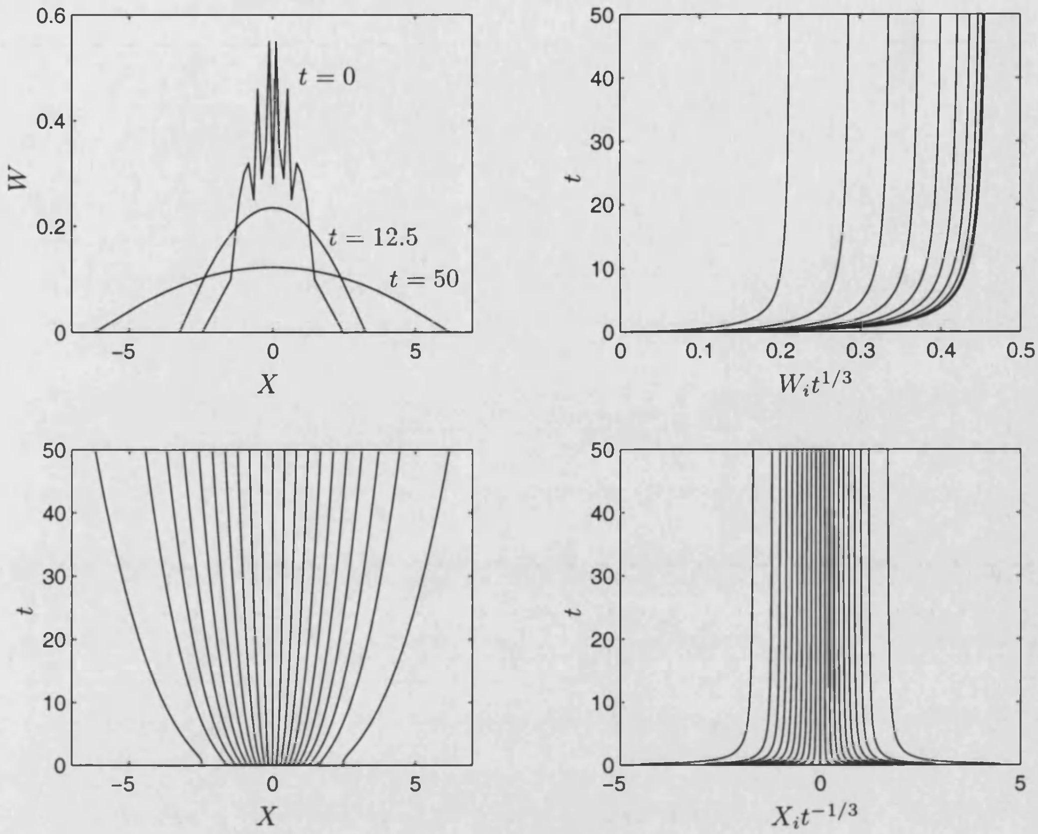


Figure 5-2: *The evolution and invariance of a solution and mesh for the porous medium equation, with $N = 19$.*

An additional equation is needed to close the set of equations for the unknowns X_i and we do this by insisting that (as in the true solution) the discrete centre of mass is conserved (without loss of generality at 0) so that

$$\sum_{i=0}^{N-1} (X_{i+1}^2 - X_i^2)(W_{i+1} + W_i) = 0. \tag{5.40}$$

Observe that the equation (5.37) for the solution and the equations (5.38), (5.40) for the mesh have decoupled in this system. This makes it much easier to analyse. In particular (5.37) has two key geometrical features. *Firstly*, it is invariant under the group action

$$t \to \lambda t, \quad W_i \to \lambda^{-1/3} W_i.$$

Thus it admits a semi-discrete self-similar solution of the form

$$\hat{W}_i(t) = t^{-1/(m+1)} \Theta_i. \tag{5.41}$$

Performing the symmetry reduction, i.e. substituting (5.41) into (5.37), gives that

$\Theta_0 = \Theta_N = 0$, and Θ_i for $i = 1, \ldots, (N-1)$ satisfies the algebraic equation

$$-\frac{1}{m+1}\Theta_i = \frac{N^2}{m}\Theta_i^2(\Theta_{i+1}^m - 2\Theta_i^m + \Theta_{i-1}^m). \tag{5.42}$$

Note that $\Theta_i \equiv 0$ is a solution to (5.42), however for use here we are interested in a solution with $\Theta_i > 0$ for $i = 1, \ldots, (N-1)$. We now proceed to demonstrate the existence of such a solution in, for simplicity, the case $m = 2$.

Theorem 5.4. *There exists a unique solution $\{\Theta_i\}$ to the algebraic equation (5.42) with $N \geq 2$, satisfying $\Theta_0 = \Theta_N = 0$, and $\Theta_i > 0$ for $i = 1, \ldots, (N-1)$.*

Proof. In the case N=2 we have the unique positive solution given by $\Theta_1 = (1/12)^{1/3}$ (c.f. example 3.1) and we are finished. In the case $N > 2$, with $\Theta_0 = 0$ write $\Theta_1 := \alpha$ then with $K := 2/(3N^2) > 0$ we have, for $2\alpha^3 \geq K$, the unique non-negative solution

$$\Theta_2 \equiv f_2(\alpha) = \left(2\alpha^2 - K\alpha^{-1}\right)^{1/2}. \tag{5.43}$$

Notice that by continuity of the operations of squaring, taking square roots and inverses, as well as the composition and addition of such functions, we may conclude that f_2 is a continuous function of $\alpha \in (0, \infty)$, for α sufficiently large. Now suppose that $\Theta_{i-1} \equiv f_{i-1}(\alpha)$ and $\Theta_i \equiv f_i(\alpha)$, then from (5.42) we have for α sufficiently large the unique non-negative solution

$$\Theta_{i+1} \equiv f_{i+1}(\alpha) = \left(2f_i(\alpha)^2 - f_{i-1}(\alpha)^2 - Kf_i(\alpha)^{-1}\right)^{1/2}. \tag{5.44}$$

Following the same argument as above we can conclude that f_{i+1} is continuous given that f_i and f_{i-1} are both also continuous. Since f_2 and the identity map are both continuous we may conclude by mathematical induction that f_i is a continuous function of $\alpha \equiv \Theta_1 \in (0, \infty)$, for $i = 1, \ldots, N$. Now from (5.43) and (5.44) we see that for any N, for α sufficiently large, $\alpha_R > 0$ say, $f_N(\alpha_R) > 0$. In this case define a new function $F(\alpha) := f_N(\alpha) > 0$, continuous and well-defined for all α sufficiently large. In addition we have already seen (in the case $N = 2$ above) that we can find an $\alpha_L := (K/2)^{1/3} > 0$ such that $f_2(\alpha_L) = 0$. Therefore in the case of $\alpha \geq \alpha_L$ sufficiently small, where i is the largest integer less than N such that $f_i(\alpha) \in \mathbb{R}^+$, further define

$$F(\alpha) = -N + i + 1 - \frac{f_i^{\max} - f_i(\alpha)}{f_i^{\max}} \in [-N+2, 0], \tag{5.45}$$

where $f_i^{\max}$ is defined to be the maximum value (dependent on N but not on α) of f_i such that f_{i+1} is not defined in $\mathbb{R}^+$ via (5.42), i.e. $f_i^{\max} = f_i(\alpha_i^{\max})$, where

$$\begin{aligned}\alpha_i^{\max} = \max\{\alpha : \alpha > 0,\ & f_1(\alpha), \ldots, f_i(\alpha) \in \mathbb{R}^+, \\ & \text{and}\ (2f_i(\alpha)^2 - f_{i-1}(\alpha)^2 - Kf_i(\alpha)^{-1}) < 0\}.\end{aligned}$$

To establish the continuity of F consider an $\alpha > \alpha_L$ and an $\varepsilon > 0$, for α sufficiently large and $\alpha' > 0$ satisfying $|\alpha - \alpha'| < \delta$, for $\delta := \varepsilon > 0$ say, we have that both $F(\alpha) = f_N(\alpha)$ and $F(\alpha') = f_N(\alpha')$, and hence continuity of F in this case follows from the continuity of f_N. Now for $\alpha > \alpha_L$ not so large, and for α' satisfying $|\alpha - \alpha'| < \delta > 0$ we have three further cases to consider. Firstly, if α and δ are such that $F(\alpha)$ and $F(\alpha')$ are defined by (5.45) with the same i, then continuity of F follows by continuity of f_i. Secondly, if $F(\alpha)$ and $F(\alpha')$ are defined by (5.45) with, respectively and without loss of generailty, i and $i+1$, then by continuity of f_i and f_{i+1}, as well as properties of the square root operation and the definition of $f_j^{\max}$, we may choose a $\delta \equiv \delta(\varepsilon) > 0$ such that $|\alpha - \alpha'| < \delta$ implies

$$f_i^{\max} - f_i(\alpha) < \frac{\varepsilon}{2} f_i^{\max}, \quad \text{and} \quad f_{i+1}^{\max} - f_{i+1}(\alpha') < \left(1 - \frac{\varepsilon}{2}\right) f_{i+1}^{\max},$$

are both true. We then have that $|F(\alpha') - F(\alpha)| < \varepsilon$ and continuity is proved. Finally, again without loss of generality suppose that α and δ are such that $F(\alpha') = f_N(\alpha')$ and $F(\alpha)$ is given by (5.45) with $i = N-1$, then by continuity of f_{N-1} and f_N, and the definition of $f_j^{\max}$, we may choose a $\delta > 0$ such that $|\alpha - \alpha'| < \delta$ implies

$$f_{N-1}^{\max} - f_{N-1}(\alpha) < \frac{\varepsilon}{2} f_{N-1}^{\max}, \quad \text{and} \quad f_N(\alpha') < \frac{\varepsilon}{2},$$

are both true. We then have that $|F(\alpha') - F(\alpha)| < \varepsilon$ and continuity is proved. We therefore have a continuous, well-defined function $F : [\alpha_L, \infty) \to [-N+2, \infty)$ for which there exists $\alpha_R > 0$ in addition to $\alpha_L > 0$, such that $F(\alpha_L) < 0 < F(\alpha_R)$. We may therefore conclude by continuity (the intermediate value theorem) that there exists an $\alpha \in (\alpha_L, \alpha_R)$ such that $F(\alpha) = 0$, i.e. $\Theta_N = f_N(\alpha) = 0$ and $\Theta_i > 0$, $i = 1, \ldots, (N-1)$ satisfies (5.42).

To establish uniqueness of non-trivial non-negative solutions consider two such solutions $\{\Theta_i\}$ and $\{\tilde{\Theta}_i\}$ to (5.42). Assume that $\Theta_i \neq \tilde{\Theta}_i$ for some i, then since given $\Theta_0 = 0$ and $\Theta_1 > 0$, (5.42) constructs (in $\mathbb{R}^+$ up to a point) a unique non-negative sequence of Θ_i, we must have that $\Theta_1 \neq \tilde{\Theta}_1$. Without loss of generality assume that $\tilde{\Theta}_1^2 - \Theta_1^2 =: \varepsilon_1 > 0$. This then yields from (5.42) (c.f. (5.43)) that

$$\varepsilon_2 := \tilde{\Theta}_2^2 - \Theta_2^2 = 2(\tilde{\Theta}_1^2 - \Theta_1^2) - K(\tilde{\Theta}_1^{-1} - \Theta_1^{-1}) > 2\varepsilon_1. \tag{5.46}$$

We can similarly further say that

$$\varepsilon_{i+1} := \tilde{\Theta}_{i+1}^2 - \Theta_{i+1}^2 = 2(\tilde{\Theta}_i^2 - \Theta_i^2) - (\tilde{\Theta}_{i-1}^2 - \Theta_{i-1}^2) - K(\tilde{\Theta}_i^{-1} - \Theta_i^{-1}) > 2\varepsilon_i - \varepsilon_{i-1}.$$

Hence we have that $\varepsilon_i > \varepsilon_{i-1} \Rightarrow \varepsilon_{i+1} > \varepsilon_i$. Therefore, in the light of (5.46) and the principle of mathematical induction, we have the contradiction that $\tilde{\Theta}_N \neq \Theta_N$, and

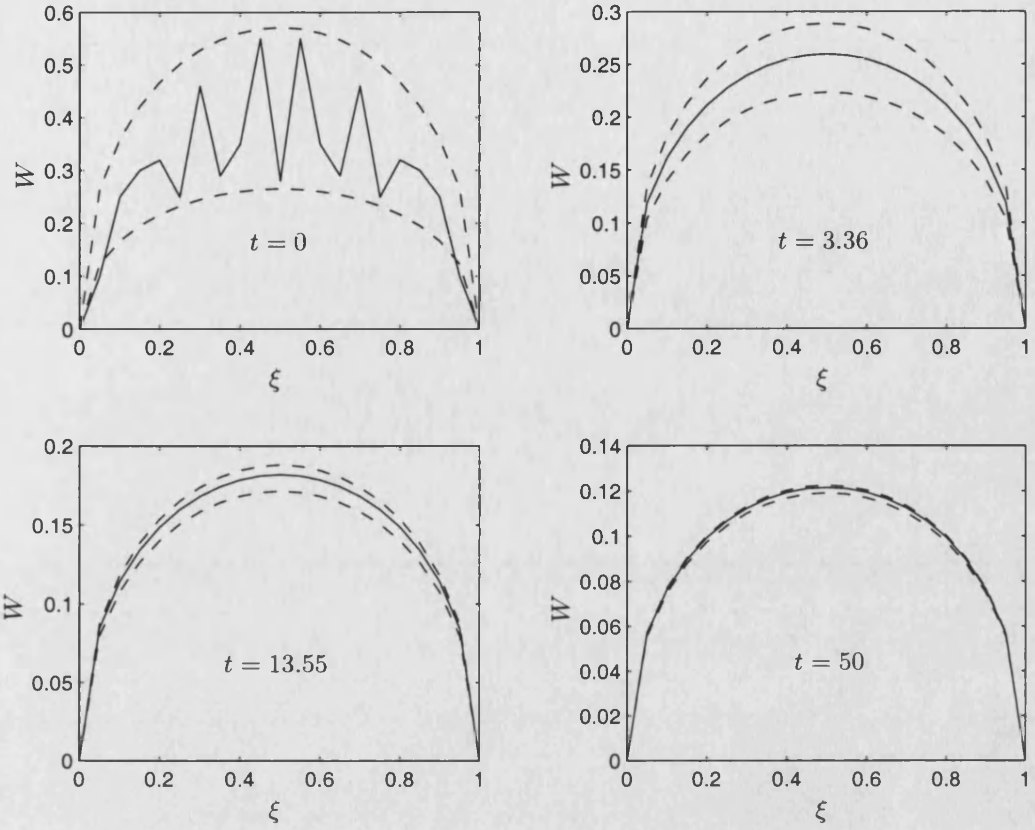


Figure 5-3: *The convergence of a solution to the porous medium equation to the semi-discrete self-similar solution using the 'pinch and squeeze argument', with $N = 19$.*

uniqueness is established. □

Now (5.42) is a consistent discretization of (5.28) with an error that is independent of t. Therefore the semi-discrete self-similar solution uniformly approximates the self-similar solution over arbitrarily long times, c.f. the analogous results of Chapter 3 regarding ODEs.

Following an application of the translational symmetry we have that

$$\hat{W}_i(t) = (t + C)^{-1/(m+1)}\Theta_i,$$

which for all constants C we notice that

$$t^{1/(m+1)}\hat{W}_i \to \Theta_i.$$

Secondly, the discretization satisfies the following comparison or maximum principle.

Theorem 5.5 (Comparison Principle). *Consider two (exact) solutions of (5.37) $\{W_{1,i}(t)\}$ and $\{W_{2,i}(t)\}$ where $W_{1,i}(0) \le W_{2,i}(0)$, $\forall i = 0, \ldots, N$, and $W_{1,0} = W_{1,N} =$*

$W_{2,0} = W_{2,N} = 0$. *Then we have*

$$W_{1,i}(t) \leq W_{2,i}(t), \quad \forall i = 0, \ldots, N, \quad \forall t > 0. \tag{5.47}$$

Proof. Firstly, assume that at some $t \geq 0$ there exists $j \in \{2, \ldots, N-1\}$ such that

$$W_{1,j}(t) = W_{2,j}(t), \quad W_{1,j-1}(t) < W_{2,j-1}(t), \quad W_{1,j+1}(t) < W_{2,j+1}(t),$$

then from (5.37) $\dot{W}_{1,j}(t) < \dot{W}_{2,j}(t)$, therefore for all sufficiently small time increments δt we have $W_{1,j}(t+\delta t) < W_{2,j}(t+\delta t)$.

As the second case assume that at some $t \geq 0$ there exists $j \in \{2, \ldots, N-1\}$ such that

$$W_{1,j}(t) = W_{2,j}(t), \quad W_{1,j-1}(t) = W_{2,j-1}(t), \quad W_{1,j+1}(t) < W_{2,j+1}(t),$$

then from (5.37) we again have that $\dot{W}_{1,j}(t) < \dot{W}_{2,j}(t)$, and also following a renumbering $j \to j+1$ in the above (and assuming that $W_{1,j-2}(t) < W_{2,j-2}(t)$, otherwise see the next case) that $\dot{W}_{1,j-1}(t) < \dot{W}_{2,j-1}(t)$. Therefore for all sufficiently small time increments δt we have $W_{1,k}(t+\delta t) < W_{2,k}(t+\delta t)$ for $k = j,\ j-1$.

As the third case assume that at some $t \geq 0$ there exists $j \in \{2, \ldots, N-1\}$ such that

$$W_{1,j}(t) = W_{2,j}(t), \quad W_{1,j-1}(t) = W_{2,j-1}(t), \quad W_{1,j+1}(t) = W_{2,j+1}(t),$$

then $\dot{W}_{1,j}(t) = \dot{W}_{2,j}(t)$. Taking the time derivative of (5.37) yields

$$\begin{aligned} \frac{d^2 W_i}{dt^2} = \frac{N^2}{m} &\left(2W_i\dot{W}_i\left(W_{i+1}^m - 2W_i^m + W_{i-1}^m\right)\right. \\ &\left.+mW_i^2\left(W_{i+1}^{m-1}\dot{W}_{i+1} - 2W_i^{m-1}\dot{W}_i + W_{i-1}^{m-1}\dot{W}_{i-1}\right)\right), \end{aligned} \tag{5.48}$$

for $i = 1, \ldots, (N-1)$. Assume now without loss of generality that j has been chosen such that $W_{1,j-2}(t) < W_{2,j-2}(t)$ (and $W_{1,j+2}(t) \leq W_{2,j+2}(t)$) then from the second case above we have $\dot{W}_{1,j-1}(t) < \dot{W}_{2,j-1}(t)$ (and $\dot{W}_{1,j+1}(t) \leq \dot{W}_{2,j+1}(t)$), and therefore from (5.48) $\ddot{W}_{1,j}(t) < \ddot{W}_{2,j}(t)$. Hence for all sufficiently small time increments δt we have $W_{1,k}(t+\delta t) < W_{2,k}(t+\delta t)$ for $k = j,\ j-1$, and again following as necessary either the second case above or a renumbering of this case we may extend this to hold for $k = j+1$, etc.

We now extend the last result for the case where at some $t \geq 0$ and for $l \in \mathbb{N}$, $2l+2 < N$, we have $W_{1,k} = W_{2,k}$ for $k = j-l, \ldots, j, \ldots, j+l$, as well as (without loss of generality)

$$W_{1,k} < W_{2,k} \text{ for } k = j-l-1, \text{ and } W_{1,k} \leq W_{2,k} \text{ for } k = j+l+1. \tag{5.49}$$

Notice that (as a straightforward induction argument will verify) the order s derivative of W_i involves the values of W at $2s+1$ consecutive points centred about point i. Hence in the case in question here we have that $W_{1,j}^{(l)} = W_{2,j}^{(l)}$. Also notice that

$$\frac{d^{s+1}}{dt^{s+1}} W_i = N^2 W_i^2 \left(W_{i+1}^{m-1} W_{i+1}^{(s)} - 2W_i^{m-1} W_i^{(s)} + W_{i-1}^{m-1} W_{i-1}^{(s)} \right) + \text{L.O.T.}, \qquad (5.50)$$

where the lower order terms here involve derivatives of order s and lower evaluated at point i and derivatives of order $s-1$ and lower evaluated at the two points either side, i.e. only involve values of W at $2s+1$ points centred on point i. Therefore since $W_{1,i}$ and $W_{2,i}$ coincide at $2l+1$ about point j, the lower order terms appearing in (5.50) are equal for both in the case $s = l$, $i = j$, and subtracting we therefore have

$$\begin{aligned} \frac{d^{l+1}}{dt^{l+1}} (W_{2,j} - W_{1,j}) = N^2 W_{1,j}^2 \Big(& W_{2,j+1}^{m-1} W_{2,j+1}^{(l)} - W_{1,j+1}^{m-1} W_{1,j+1}^{(l)} \\ & + W_{2,j-1}^{m-1} W_{2,j-1}^{(l)} - W_{1,j-1}^{m-1} W_{1,j-1}^{(l)} \Big). \end{aligned}$$

But now from (5.49) and the properties of the order l derivatives of W_i, we have that $W_{1,j-1}^{(l)}(t) < W_{2,j-1}^{(l)}(t)$ and $W_{1,j+1}^{(l)}(t) \le W_{2,j+1}^{(l)}(t)$, from which it follows that $W_{1,j}^{(l+1)}(t) < W_{2,j}^{(l+1)}(t)$, and as in the previous cases for all sufficiently small time increments δt we have $W_{1,k}(t+\delta t) < W_{2,k}(t+\delta t)$ for $k = j-l-1, \dots, j, \dots, j+l+1$.

The final case is where $W_{1,i}(t)$ and $W_{2,i}(t)$ coincide for all i at some $t \ge 0$, in which case they coincide for all further time. We conclude that $W_{1,j}(t) > W_{2,j}(t)$ for some j and some t is not possible, leaving the desired result. □

For the discretizations of ODEs in Chapter 3 we had a useful stability result, for this example our semi-discretization has an even stronger result. We now use this to prove the convergence of an arbitrary semi-discrete solution with arbitrary initial data to the semi-discrete self-similar solution (which from above we know converges to the fully continuous self-similar solution to the problem).

Theorem 5.6. *For a solution $\{W_i(t)\}$ of (5.37) with general initial conditions satisfying $W_0(0) = W_N(0) = 0$, and $W_i(0) > 0$ for $i = 1, \dots, (N-1)$. We have that $t^{1/(m+1)} W_i(t) \to \Theta_i$, $\forall i = 0, \dots, N$, as $t \to \infty$, where Θ_i is the non-negative solution to (5.42) which was shown to exist in Theorem 5.4.*

Proof. First note that $U_i(t)$ and $L_i(t)$, defined by

$$U_i(t) := (t + c_1)^{-1/(m+1)} \Theta_i, \quad L_i(t) := (t + c_2)^{-1/(m+1)} \Theta_i,$$

are both solutions to the semi-discretization (5.37) for fixed c_1, $c_2 \in \mathbb{R}$ where Θ_i satisfies (5.42). Therefore both are actually semi-discrete self-similar solutions. Now note that as c_1 and c_2 are both fixed, we have the convergence property

$$U_i(t) = t^{-1/(m+1)}(\Theta_i + o(1)), \quad L_i(t) = t^{-1/(m+1)}(\Theta_i + o(1)). \tag{5.51}$$

For given initial conditions $W_i(0)$, we can choose c_1 and c_2 such that

$$U_i(0) = c_1^{-1/(m+1)}\Theta_i \geq W_i(0) \geq c_2^{-1/(m+1)}\Theta_i = L_i(0).$$

Then from Theorem 5.5 we know that

$$L_i(t) \leq W_i(t) \leq U_i(t) \qquad \forall t \geq 0,\ \forall i = 0, \ldots, N. \tag{5.52}$$

So finally we may conclude the desired result from (5.51). □

Note that in the above proof we assumed that we solved the algebraic system (5.42) exactly. Suppose however that there was some small error here, which is obviously independent of time. But wherever the quantities Θ_i are used they are multiplied by $t^{-1/(m+1)}$, therefore any error in the numerical solution will converge to zero for large times.

An example of Theorem 5.6 in action is given in figure 5-3, where the evolution of a numerical solution to the PME with general initial conditions is shown, together with two semi-discrete self-similar solutions with the constants taking the values $c_1 = 0.5$ and $c_2 = 5$. As can be seen (ignoring the end regions of the domain where the plotting routine and lack of resolution spoil things slightly) the general solution is 'squeezed' demonstrating the convergence of the numerics to self-similarity. Correspondingly the mesh also converges to self-similarity, see figure 5-2. We also see the convergence to self-similarity of both the solutions and mesh in figure 5-4 where the correctly scaled numerical quantities can be seen to converge to the functions θ and y derived above and shown in figure 5-1.

Due to the fact that the solution and mesh equations decoupled for this problem we are able to solve the two systems separately. Here we have solved the system (5.37) using a third-order BDF method using a time step implied by the scaling of the problem, see Section 5.4.4, and recall Chapter 3 where it was seen that a posteriori and a priori time step selection essentially led to the same results. To obtain the corresponding mesh at a particular time we then solved the system given by (5.38) and (5.40) using a standard nonlinear equation solver. In general, for other problems, it will not be possible to decouple the two problems in such a way. In this case it may become necessary to solve the index-one differential-algebraic equation (DAE) given by (5.5) and (for example) (5.24). Solving this with the DAE solver DASSL [26] demonstrates that although the comparison principle, which only applied in the computational domain, is no longer

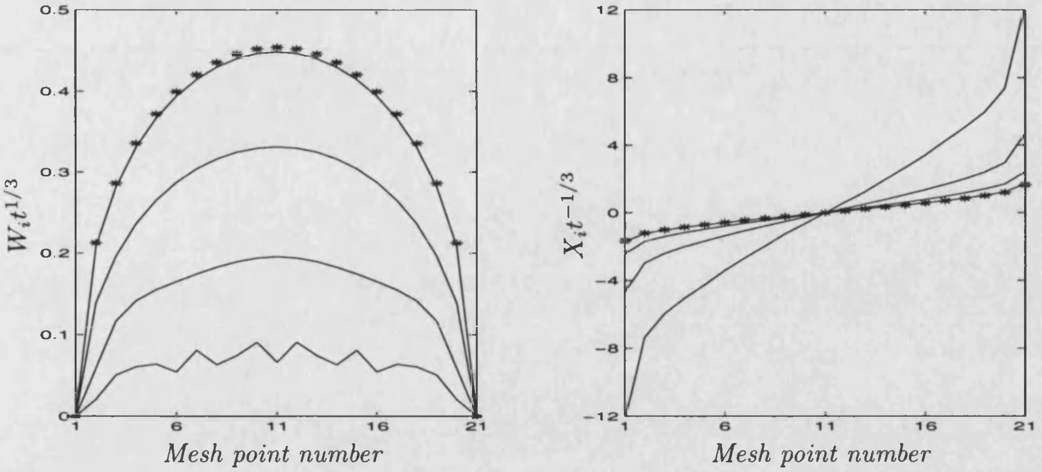


Figure 5-4: *Convergence of the solution (left) and mesh (right) to discrete self-similarity, stars indicate the θ and y derived analytically as a self-similar solution earlier. The four solid lines show $X_i t^{-1/3}$ and $W_i t^{-1/3}$ during the evolution at times 0.0085, 0.149, 1.184, and 50. The solution is computed from (5.26) using $N = 21$ and a third-order adaptive BDF method for the time integration.*

applicable, very similar numerical results are observed and the excellent long term behaviour of the method is preserved.

5.4.3 A finite element approach

We shall now take a look at whether a finite element approach may handle our scaling invariant problem in a manner which inherits the desirable behaviour the simple finite difference approach exhibited.

Consider the weak formulation of the transformed porous medium equation (5.26), in the case $m = 2$,

$$(w_t, v) - \left(\frac{1}{2}w^2(w^2)_{\xi\xi}, v\right) = 0,$$

for all v in some appropriately defined function space V, where here (f, g) represents the Euclidean inner product, i.e. the integral of fg over the spatial domain. Integrating by parts the second inner product in the above yields

$$\left(w_t + 2(ww_\xi)^2, v\right) + \left(w^3 w_\xi, v_\xi\right) = 0. \tag{5.53}$$

We immediately note that this expression is left invariant under the scaling generated by

$$\mathbf{X} = t\partial_t - \frac{1}{3}w\partial_w, \tag{5.54}$$

where, since the test function v depends only on the computational spatial variable ξ

it is itself unaffected by the scaling. If we now assume that we may write

$$w = \sum_{j=1}^{N} W_j(t)\varphi_j(\xi),$$

where $\{\varphi_j : j = 1, \ldots, N\}$ forms a basis for V_h, an N-dimensional subspace of V. Substituting into the weak form (5.53) where we also assume a standard Galerkin weighting, i.e. we take the test functions v to simply be the expansion functions φ_i, we obtain

$$\sum_j (\varphi_j, \varphi_i) \frac{dW_j}{dt} + 2\left(\left\{\sum_j W_j\varphi_j\right\}^2 \left\{\sum_j W_j \frac{\partial\varphi_j}{\partial\xi}\right\}^2, \varphi_i\right) \\ + \left(\left\{\sum_j W_j\varphi_j\right\}^3 \left\{\sum_j W_j \frac{\partial\varphi_j}{\partial\xi}\right\}, \frac{\partial\varphi_i}{\partial\xi}\right) = 0, \qquad (5.55)$$

for $i = 1, \ldots, N$. Notice here that under the scaling (5.54) each of the three terms above scales in the same manner, and hence the semi-discrete expression (5.55) is invariant under the transformation corresponding to (5.54).

Following our earlier investigation of the finite difference formulation we now consider a symmetry reduction of the finite element formulation. Due to the scaling invariance property of (5.55) we set

$$W_j = t^{-1/3}\Theta_j,$$

and substitute into (5.55). It is immediately seen that the variable t cancels throughout as desired. We are left with the steady problem,

$$-\frac{1}{3}\sum_j (\varphi_j, \varphi_i)\,\Theta_j + 2\left(\left\{\sum_j \Theta_j\varphi_j\right\}^2 \left\{\sum_j \Theta_j \frac{\partial\varphi_j}{\partial\xi}\right\}^2, \varphi_i\right) \\ + \left(\left\{\sum_j \Theta_j\varphi_j\right\}^3 \left\{\sum_j \Theta_j \frac{\partial\varphi_j}{\partial\xi}\right\}, \frac{\partial\varphi_i}{\partial\xi}\right) = 0, \qquad (5.56)$$

for $i = 1, \ldots, N$. But notice that if we now consider the weak formulation of the reduced ODE (5.28), in the case $m = 2$, and integrate by parts we obtain,

$$2\left(\theta^2(\theta')^2, v\right) + \left(\theta^3\theta', v'\right) = \frac{1}{3}(\theta, v),$$

which, as we hoped for, has (5.56) as a consistent Galerkin discretization. As in the earlier more lengthy discussion on a finite difference formulation, this demonstrates that semi-discrete self-similar solutions are admitted by our Galerkin formulation.

We end our consideration of finite element type techniques here and leave any further work in this direction as a topic for the future. However note that due to the equivalence of the Galerkin formulation to a least squares minimization problem, which may well exhibit oscillations about the continuous solution, there appear to be immediate pointers to the possibility that we will no longer be able to be establish a comparison principle as in the finite difference discretization.

5.4.4 Invariant temporal discretizations

Following the correct choice of monitor function M in the spatial adaptivity step and the transformation of our original PDE to the computational domain, we considered both a finite difference and finite element semi-discretization. These both yielded systems of ODEs, (5.37) and (5.55), which are themselves invariant under a scaling transformation corresponding to that of the original infinite dimensional problem. Of course now the techniques developed in Chapter 3 may be employed to integrate numerically forward in time whilst preserving the scaling property in the fully discrete solution. For the particular case of the porous medium equation here the semi-discrete self-similar solution admitted by the semi-discretization will, using the theory from Chapter 3, be preserved as an attractor by the temporal discretization. See Example 3.1, which exhibits this property in the simple case of $N = 2$.

5.5 Summary of Chapter

In this Chapter extensions of the work of Chapter 3 to problems with one spatial dimension are given. Again this is based upon the use of adaptivity and a review of some moving mesh methods in one dimension was given. Definitions of what it means for a method to be scale invariant were stated and a discussion was made of how the admittance of self-similar solutions and comparison principles may be combined.

The porous medium equation was used as an interesting model problem possessing many qualitative features. Adaptive methods were shown to be ideally suited to this problem where interfaces move with finite velocity. The scale invariant method developed here captures this behaviour precisely. In addition the scale invariance results in a method which preserves conservation laws as well as admitting semi-discrete self-similar solutions.

We note again here the important point that although these methods admit, and indeed converge to, (semi-) discrete self-similar solutions, they are entirely flexible enough not to exclude the computation of solutions starting from general initial conditions. See [143, 144] for an example where it is precisely the self-similar solutions that are solved for numerically, following the symmetry reduction.

Due to the special nature of the PME problem we were able to establish a discrete comparison principle which enabled us to prove the stability and non-negativity of the semi-discretization, as well as the convergence to self-similarity, and so rigorously extend the convergence proofs of Chapter 3. However the discrete comparison principle will not hold for other problems in general and the extensions of the results of Chapter 3 to the full and semi-discretizations of (the infinite dimensional) PDEs still need to be proved, with numerical results pointing to the fact that they do indeed hold.

Chapter 6

The semi-geostrophic equations

6.1 Overview of Chapter

The remainder of this thesis studies the application of geometric integration methods to problems arising in fluid dynamics. In particular we look at the challenging problem of large scale atmospheric motion described by the semi-geostrophic equations. This problem is of interest to meteorologists and geophysicists as well as people involved in numerical weather prediction.

The semi-geostrophic equations represent a problem which contains large amounts of geometric structure, and is also far from trivial to consider numerically. For example, linking nicely with earlier Chapters, they possess Hamiltonian structures, symmetries, singularities, conservation laws, as well as a natural coordinate transformation which itself has many geometric features. This Chapter describes the semi-geostrophic equations as well as their geometric properties. Numerical issues and geometric integration are considered in Chapters 7 and 8.

We shall use this problem as a case study on how ideas from geometric integration may be extended to higher-dimensional and fundamentally more complex problems.

6.2 The semi-geostrophic equations

Semi-geostrophic (SG) theory attempts to model atmospheric flows that vary on scales of typical synoptic (large scale, that is for the atmosphere at least one horizontal scale of the order of 1000km and a vertical scale of order 10km) patterns with Lagrangian (following the motion) time scales of at least several hours. We shall now spend a little time giving a sketch of the equations derivation and justification, as well as some of their properties.

6.2.1 Background physics and derivation from Navier-Stokes

The underlying equations for describing fluid flow are the Navier-Stokes equations, assuming inviscid flow gives us the Euler equations (as seen in Chapter 2), however for large scale atmospheric motions the rotation of the Earth must be taken into account and hence the Coriolis force must be included. The underlying equations are thus (in the horizontal)

$$\frac{D\mathbf{u}}{Dt} + f\mathbf{k} \times \mathbf{u} + \nabla_x \varphi = 0, \tag{6.1}$$

where $\mathbf{k} = (0,0,1)^T$, $\nabla_x = (\partial/\partial x, \partial/\partial y)^T$, $\mathbf{u} = (u,v)^T$ represents the horizontal velocity, g is the acceleration due to gravity, f is the Coriolis parameter (which depends upon latitude, but here is taken as a constant — this is the f-plane assumption), $\mathbf{x} = (x,y,z)^T$ is a local Cartesian coordinate system with y pointing northward and z (a function of pressure) vertically, finally φ represents the geopotential. We assume a stratified fluid and thus *hydrostatic balance* in the vertical, i.e.

$$\frac{\partial \varphi}{\partial z} = g\frac{\theta}{\theta_0}, \tag{6.2}$$

which says that gravitational and vertical pressure gradient terms are in balance and therefore we may neglect the effects of vertical inertia. The *Boussinesq approximation* is made which basically assumes that the density of the fluid in question is approximately constant, (the exact meaning and result of this assumption has some subtlety and ambiguity, see [159, 130, 62]). Here we simply note that this leads to the incompressibility constraint

$$\nabla_x \cdot \mathbf{u} = 0. \tag{6.3}$$

Finally, we assume conservation of potential temperature (i.e. *adiabatic flow*)

$$\frac{D\theta}{Dt} = 0. \tag{6.4}$$

The system given by (6.1)–(6.4) constitutes the *primitive equation set.* Note that the material (or Lagrangian) derivative is given by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u}.\nabla_x = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z}.$$

Boundary conditions for this system are taken to be no flow through upper and lower boundaries, i.e. $w = 0$ at $z = 0,\ H$. Along with suitable (e.g. periodic) conditions on the lateral boundaries.

The primitive equation set as defined here is very similar to the actual systems solved by weather forecasters and ocean modellers. Finding exact solutions is of course impossible in all but the most idealized situations and so numerical simulations are a necessity, both for predictive work as well as for gaining understanding and insight into the equations. However, the primitive equations describe a wide range of solution phenomena. For example they admit two distinct wave-like motions. Firstly the 'slow' moving synoptic patterns (Rossby waves) mentioned above which can be shown to result from an approximate *balance* (the so-called geostrophic balance, see below) between pressure gradient and Coriolis terms. As well as 'fast' inertia-gravity waves. (Fast moving sound waves have already been removed by making the hydrostatic assumption.) Of course these fast (high frequency) solutions imply great difficulties with stability bounds in numerical calculations (c.f. the CFL condition [137]) and may force the use of implicit methods or excessively small time steps. It is often argued [49, 130, 141] that these fast motions are of little importance in comparison to the slow motions in the study of large scale geophysical phenomena. Due to these facts geophysicists have put much effort into the derivation of further approximations (or simplifications) to the primitive equations which attempt to *filter* out the unwanted solution behaviours. Thereby resulting in systems which, although still highly complex and able to accurately model the large scale flow, are less susceptible to strict numerical limitations.

For our synoptic scales we can argue that the Rossby number R_0 is relatively small, for example with the typical values $f = 10^{-4}\text{s}^{-1}$, $L = 10^6\text{m}$ and $V = 10\text{ms}^{-1}$ we have

$$R_0 = \frac{V}{fL} \approx 0.1.$$

Expanding the momentum equations in terms of R_0 gives, to the lowest order,

$$u_g = -\frac{1}{f}\frac{\partial \varphi}{\partial y}, \quad v_g = \frac{1}{f}\frac{\partial \varphi}{\partial x}. \tag{6.5}$$

where u_g and v_g are known as the horizontal geostrophic velocity components. The relations (6.5) state that to a first approximation the horizontal pressure forces on a fluid element are exactly balanced by the Coriolis force.

We may now define the semi-geostrophic (SG) model which is a leading order approximation to the primitive equations. We do this by replacing the advected quantity in the horizontal momentum equation (6.1) by the geostrophic velocities, importantly however we leave the advecting velocity implied by the total derivative notation D/Dt as the original velocity $\mathbf{u}$. We leave the advecting velocity unchanged to increase the accuracy of the approximation. If we had also approximated the advecting velocity by the geostrophic velocity we would arrive at the quasi-geostrophic system which, although well studied [141, 130], does not possess the interesting physically realistic features of the SG system. Our derived approximation to the primitive equations is thus

$$\left.\begin{aligned}\frac{Du_g}{Dt} - fv + \frac{\partial\varphi}{\partial x} = \frac{Du_g}{Dt} - f(v - v_g) = 0,\\ \frac{Dv_g}{Dt} + fu + \frac{\partial\varphi}{\partial y} = \frac{Dv_g}{Dt} + f(u - u_g) = 0,\end{aligned}\right\} \tag{6.6}$$

$$\frac{D\theta}{Dt} = 0, \tag{6.7}$$

$$\nabla_x \cdot \mathbf{u} = 0, \tag{6.8}$$

$$\nabla_x \varphi = \left(fv_g, -fu_g, g\frac{\theta}{\theta_0}\right)^T. \tag{6.9}$$

For many further details see [90, 91, 51, 52]. The existence of global weak solutions to this problem is established in [19], however due to the nonlinearities present uniqueness is still an open question. The SG equation set is of interest for many reasons, one being that it allows the study of idealized atmospheric weather fronts. That is, just as for some of the ordinary differential equation examples considered in Chapter 3, it admits solutions which form singularities in finite time.

6.2.2 Coordinate transformation, Monge-Ampère / vorticity advection formulation

We now define a coordinate transformation from the physical $(x, y, z)^T$ coordinates to isentropic (on surfaces of constant potential temperature) *geostrophic momentum coordinates*

$$\mathbf{X} \equiv (X, Y, Z)^T = \left(x + \frac{v_g}{f}, y - \frac{u_g}{f}, \frac{g\theta}{f^2\theta_0}\right)^T. \tag{6.10}$$

In order to simplify the representation and to aid analysis this transformation is almost always invoked in studies of the SG problem, see for example [50, 51, 52]. In a similar manner to the previous Chapters we may think of $(X, Y, Z)^T$ as being (fictive) computational type coordinates as introduced in earlier Chapters, (6.10) then describes the correspondence between a computational and physical spatial mesh. In terms of these new coordinates (6.6) and (6.7) become (using $D\mathbf{x}/Dt = \mathbf{u}$)

$$\frac{D\mathbf{X}}{Dt} = \mathbf{u}_g \equiv (u_g, v_g, 0)^T, \tag{6.11}$$

and hence the motion in these new coordinates is exactly geostrophic in the horizontal and constrained to Z (isentropic) surfaces in the vertical. Very importantly the motion is also nondivergent in $\mathbf{X}$ space, since

$$\nabla_X \cdot \mathbf{u}_g = \frac{\partial u_g}{\partial X} + \frac{\partial v_g}{\partial Y} = f\left(\frac{\partial y}{\partial X} - \frac{\partial x}{\partial Y}\right). \tag{6.12}$$

The nondivergence now follows if we note that later on in (6.18) we shall show that it is possible to write $\mathbf{x} = \nabla_X R$, for some convex function $R(\mathbf{X})$. A numerical method based on (6.10) and (6.11) will perform in an adaptive way — a Lagrangian form of mesh adaptivity where the mesh is moved at exactly the speed of the underlying velocity field.

We shall now give a sketch to demonstrate that the Jacobian determinant

$$q = \frac{\partial(X,Y,Z)}{\partial(x,y,z)}, \tag{6.13}$$

representing the ratio of volume elements in dual space to those in physical space is conserved following the flow. Thinking in terms of mappings between physical and computational domains, this obviously relates the scaling of the spatial mesh to the computational mesh. We call q the SG potential vorticity (PV) due to the fact that it satisfies

$$\frac{Dq}{Dt} = 0,$$

and also because it can be thought of as a consistent form of the Ertel [141] potential vorticity of the primitive equations. Potential vorticity type quantities are of vital importance in the analysis of the evolution of geophysical fluid systems [130, 141, 49, 53].

Following [10], let subscript 0 denote function values at an arbitrary initial time. Split the expression (6.13) into three parts as follows,

$$q = \frac{\partial(X,Y,Z)}{\partial(X_0,Y_0,Z_0)}\frac{\partial(X_0,Y_0,Z_0)}{\partial(x_0,y_0,z_0)}\frac{\partial(x_0,y_0,z_0)}{\partial(x,y,z)} =: J_1\, J_2\, J_3. \tag{6.14}$$

Now, since the fluid motion here is incompressible in physical space $J_3 \equiv 1$. We know from (6.11) that the fluid motion in dual space is confined to surfaces of constant Z, therefore $Z \equiv Z_0$, and so

$$J_1 \equiv \frac{\partial(X,Y)}{\partial(X_0,Y_0)}. \tag{6.15}$$

Now, motion in dual space on constant Z surfaces is divergence free (from (6.12), therefore $J_1 \equiv 1$. Hence, we may conclude that $q \equiv J_2$, and so the quantity q is conserved following fluid trajectories.

It is possible to derive this same result using other geometric techniques. See [139] for example, where the proof is based upon a Hamiltonian formulation of the problem (see later). It should also be said that the conservation of potential vorticity in many fluid systems arises naturally in mathematical terms from the particle relabelling symmetry in the Lagrangian framework, again see later.

Having shown that the coordinate transformation (6.10) leads to an incredibly useful conserved scalar, we now illustrate some additional structure underlying the mapping. It is possible to write the coordinate transformation (6.10) as

$$\mathbf{X} = \nabla_x P, \quad \text{where} \quad P(\mathbf{x}) = \frac{\varphi}{f^2} + \frac{1}{2}(x^2 + y^2). \tag{6.16}$$

Hence q, the PV, is equivalently the determinant of the Hessian matrix of P with respect to the coordinates $\mathbf{x}$, i.e.

$$q = \det(\mathrm{Hess}_x(P)). \tag{6.17}$$

This is a nonlinear elliptic equation of Monge-Ampère type, see [13]. The Jacobian of the coordinate transformation (the Hessian of P) is a symmetric matrix, and hence when it is nonsingular its inverse is also symmetric, therefore on taking the curl and using standard vector calculus results we can conclude that $\mathbf{x}$ may be written as the gradient of some function $R(\mathbf{X})$,

$$\mathbf{x} = \nabla_{\mathbf{X}} R, \tag{6.18}$$

where $P(\mathbf{x})$ and $R(\mathbf{X})$ are a pair of functions dual to each other under the Legendre transform

$$P + R = \mathbf{x} \cdot \mathbf{X}.$$

It is now possible for us to write

$$u_g = f\left(\frac{\partial R}{\partial Y} - Y\right), \quad v_g = -f\left(\frac{\partial R}{\partial X} - X\right).$$

We may therefore introduce a streamfunction for the geostrophic velocities

$$\Psi = f^2\left(\frac{1}{2}(X^2 + Y^2) - R(\mathbf{X})\right), \quad (u_g, v_g) = \frac{1}{f}\left(-\frac{\partial \Psi}{\partial Y}, \frac{\partial \Psi}{\partial X}\right). \tag{6.19}$$

Defining ρ (often referred to as the *pseudo-density*) to be

$$\rho \equiv q^{-1} = \det(\mathrm{Hess}_X(R)), \tag{6.20}$$

it can be shown that

$$\frac{D_X \rho}{Dt} \equiv \left(\frac{\partial}{\partial t} + \mathbf{u}_g \cdot \nabla_X\right)\rho \equiv \frac{\partial \rho}{\partial t} - \frac{1}{f}\frac{\partial(\rho, \Psi)}{\partial(X, Y)} = 0. \tag{6.21}$$

It is now possible to compute the evolution of this system using the following procedure,

1. given an initial distribution of pseudo-density solve the nonlinear elliptic equation of Monge-Ampère type (6.20) for R,

2. using the streamfunction (6.19) compute a new velocity field (u_g, v_g),
3. advect the pseudo-density distribution using (6.21) and return to start.

The reason for computing on ρ and R rather than q and P shall be discussed in the following Chapter, along with a discussion of the correct boundary conditions to impose on the problem.

We thus have two distinct numerical problems to solve. The first being the computation of a solution to the Monge-Ampére equation (6.20). This is obviously linked to determining the coordinate transformation (6.10), since for a given R we have $\mathbf{x} = \nabla_{\mathbf{X}} R$, and hence this fits in well with our discussions of coordinate transformations and adaptivity from earlier Chapters. The second numerical challenge is that of solving the advection equation (6.21). We shall show that this also has nice geometric structure in Section 6.3, and then return to the adaptivity connection in Chapter 7.

6.3 Hamiltonian formulations of the SG problem

6.3.1 Canonical formulation

We now consider Hamiltonian formulations for this problem. We follow [139] where two distinct Hamiltonian formulations of the semi-geostrophic equations are given. The first, a canonical (infinite-dimensional extension of (2.33)) representation of the equations of motion (6.11), with Hamiltonian functional

$$\mathscr{H}[\mathbf{X}] = f \int d\mathbf{a} \left(\frac{1}{2}(X^2(\mathbf{a}) + Y^2(\mathbf{a})) - R(\mathbf{X}(\mathbf{a})) \right),$$

where $\mathbf{a}$ is a Lagrangian particle labelling coordinate. The standard canonical Poisson bracket is given by

$$\{\mathscr{F}, \mathscr{G}\}_c = \int d\mathbf{a} \left(\frac{\delta \mathscr{F}}{\delta X(\mathbf{a})} \frac{\delta \mathscr{G}}{\delta Y(\mathbf{a})} - \frac{\delta \mathscr{F}}{\delta Y(\mathbf{a})} \frac{\delta \mathscr{G}}{\delta X(\mathbf{a})} \right). \tag{6.22}$$

In this formulation it is possible to prove conservation of PV (equivalently pseudo-density) along trajectories by demonstrating that

$$\{q, \mathscr{H}\}_c = 0.$$

Although in Chapter 8 we shall be concerned with geometric discretizations of the noncanonical formulation for this problem (see the next Section), we note here for completeness that Hamiltonian truncations of this canonical formulation are discussed in [10, 11].

6.3.2 Noncanonical formulation

Again following [139] we may write our advection equation (6.21) in noncanonical Hamiltonian form, c.f. Section 2.4, especially the discussion of the Euler equations and the references given. Using our previous Hamiltonian, this time evaluated in phase space solely as a functional of ρ, that is

$$\mathscr{H}[\rho] = f\int d\mathbf{X}\,\rho(\mathbf{X})\left(\frac{1}{2}(X^2+Y^2)-R(\mathbf{X})\right),$$

where the extra ρ appears as a result of the change of variables in comparison with the previous $\mathscr{H}$. Our Hamiltonian functional here has the variational derivative (see [139] for the details)

$$\frac{\delta\mathscr{H}}{\delta\rho} = f\left(\frac{1}{2}(X^2+Y^2)-R\right) \equiv \frac{1}{f}\Psi.$$

where Ψ is the streamfunction defined in (6.19).

We are now in a position to write the equations of motion (6.21) in the Hamiltonian form

$$\frac{\partial\rho(\mathbf{X})}{\partial t} = \{\rho(\mathbf{X}),\mathscr{H}\}, \tag{6.23}$$

where the noncanonical Poisson bracket (see [139, 114, 128]) is given by

$$\{\mathscr{F},\mathscr{G}\} = \int d\mathbf{X}\,\frac{\delta\mathscr{F}}{\delta\rho(\mathbf{X})}\left(\frac{\partial\left(\rho(\mathbf{X}),\frac{\delta\mathscr{G}}{\delta\rho(\mathbf{X})}\right)}{\partial(X,Y)}\right). \tag{6.24}$$

See [139] for precise definitions of the domains of integration in both the Hamiltonian functionals and the brackets given above.

In a similar way to the Euler equations discussed in Section 2.4, it is possible to see that the above noncanonical bracket is associated with the following cosymplectic operator

$$\mathscr{D}\bullet = \frac{\partial(\rho,\bullet)}{\partial(X,Y)},$$

such that

$$\{\mathscr{F},\mathscr{G}\} = \int d\mathbf{X}\,\frac{\delta\mathscr{F}}{\delta\rho(\mathbf{X})}\,\mathscr{D}\,\frac{\delta\mathscr{G}}{\delta\rho(\mathbf{X})}.$$

Rather than beginning with the canonical formulation and essentially by inspection writing down this noncanonical formulation it is possible to begin with the former and perform a *reduction* [114, 124] to arrive at the latter involving fewer variables. It is the particle relabelling symmetry [141, 129] that allows this to be done as in the classical reduction from Lagrangian to Eulerian variables. Since we effectively build in the relabelling symmetry into the new bracket (6.24) of Lie-Poisson type (c.f. Example

2.6), conservation of potential vorticity (or pseudo-density) now becoming implicit in the definition of our new phase space.

The geometric integration problem of numerically integrating these equations whilst preserving the pseudo-density or potential vorticity along the flow turns out to be intimately related to preserving the Hamiltonian (or Poisson bracket) structure of the problem. See [139] for more details, and also [118, 123] for some applications to similar problems where explicit methods capturing the Hamiltonian structure and the Casimir invariants are derived.

To conclude this Section we note that in the noncanonical formulation the pseudo-density ρ becomes a *Casimir invariant* of the system. A functional $\mathscr{C}$ is a Casimir if and only if its Poisson bracket with *every other* functional vanishes, i.e. if

$$\{\mathscr{C},\mathscr{F}\} = 0 \quad \forall \mathscr{F}, \tag{6.25}$$

equivalently $\mathscr{D}\delta\mathscr{C} \equiv 0$. Hence, Casimirs arise from degeneracies in the Poisson bracket, conservation of ρ or q now becomes implicit in the definition of the noncanonical phase space. For our system under consideration we have the operator

$$\mathscr{D} = \rho_X \partial_Y - \rho_Y \partial_X,$$

and so notice that $\mathscr{D}P = 0$ whenever P is a smooth function of ρ. Therefore the following infinitely many functionals

$$\mathscr{C}[\rho] = \int C(\rho)\, dX\, dY, \tag{6.26}$$

are all Casimirs, where $C(\rho)$ is any smooth function of ρ. The quantities given by (6.26) are sometimes referred to as the area integrals and reflect the pointwise conservation of PV or pseudo-density for our system, see [128].

6.4 Summary of Chapter

In this Chapter we have given a derivation of the semi-geostrophic equation set. We have discussed the underlying physics and approximations imposed. Of vital importance to the problem, the geostrophic coordinate transformation was introduced. This is a problem specific change of variables which has been widely used in much analytic work on the system. The Legendre transform structure of the transformation was discussed and this shall be used in the following Chapter where links to the adaptivity and moving mesh ideas from previous Chapters is explored. Following the coordinate transformation an advection problem was obtained, we showed that this can be written in Hamiltonian form, linking in with Chapters 2 and 4. The advection equation part of the problem shall be considered further in Chapter 8.

Chapter 7

The SG coordinate transformation, Monge-Ampére equations and adaptivity

7.1 Overview of Chapter

In this Chapter we shall attempt to apply some of the coordinate transformation and adaptivity ideas we met in Chapter 5 to the semi-geostrophic equations. From the results of Chapters 3, 4 and 5 this has the possible immediate advantages of being able to accurately compute self-similar solutions, which may for example exhibit singular behaviour as in the process of frontogenesis. Although some discussion of the scalings present in certain special solutions to this problem does appear in the literature [135], a complete scaling analysis of the invariances of the equations appears to be lacking, and it is not the topic of this thesis to carry this out. We therefore rather use as motivation for this Chapter the fact that large amounts of the analytic theory developed for this problem hinges crucially on the coordinate transformation (6.10) which in some sense simplifies the equations. Amongst other features, this transformation leads to the Hamiltonian formulation of the transformed problem and ties this work in nicely with earlier parts of this Thesis. We consider further the Hamiltonian issues in Chapter 8, focusing in this Chapter on adaptivity and the coordinate transformation.

It shall be mentioned that at the tip (the cusp point — see Appendix B) of a front the geostrophic coordinate transformation has a singularity and the potential vorticity is infinite at this point. Due to jumps in temperature and momentum across the frontal region it seems sensible if possible to compute on a mesh which aligns and concentrates itself about the front. Therefore from the very outset some kind of adaptive numerical method which uses a monitor function based on the magnitude of potential vorticity seems sensible. We shall explore this further in this Chapter.

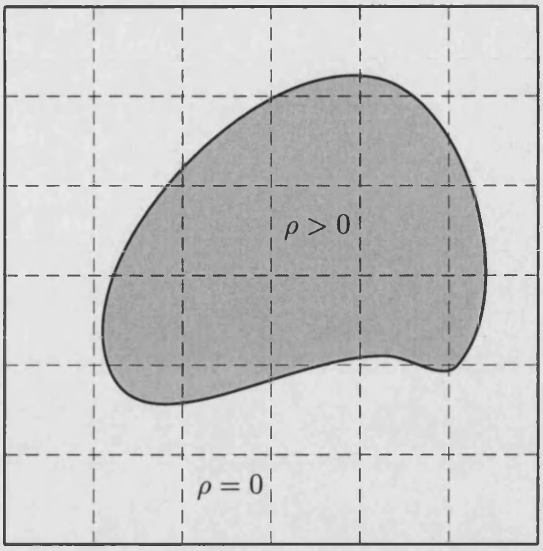


Figure 7-1: Support of ρ in dual space for the semi-geostrophic problem.

7.2 Invertibility relations and Monge-Ampére equations

7.2.1 Invertibility relations

For certain flows satisfying a dynamical balance (for example geostrophic balance) the spatial distribution of potential vorticity in principle determines all other dynamical fields like velocity, pressure and temperature. We call this an invertibility relation [92]. The relation in our context here takes the form of a Monge-Ampére equation in physical space linking the distribution of potential vorticity to a potential for the geostrophic coordinate transformation (6.10). An important point to recall from the previous Chapter is that the invertibility principle has an alternative formulation as the corresponding Monge-Ampére equation in dual space.

Consider the following equation of Monge-Ampére type in only two dimensions

$$\det(\mathrm{Hess}_X(R)) \equiv \frac{\partial^2 R}{\partial X^2}\frac{\partial^2 R}{\partial Y^2} - \left(\frac{\partial^2 R}{\partial X \partial Y}\right)^2 = \rho(X,Y). \tag{7.1}$$

At points of the (X, Y) domain where $\rho > 0$ (7.1) is of elliptic type, and at points where $\rho = 0$ it is parabolic (see [13] for additional background theory on Monge-Ampére type equations). Since regions where ρ is positive and zero are both of importance in semi-geostrophic theory we need to consider the equation in both regions and therefore consider solving a problem which changes type across discontinuities in the right hand side. This implies possible problems with solving the equation with standard methods for nonlinear elliptic problems for example.

7.2.2 Boundary conditions

The solution strategy posed in physical space of solving a Monge-Ampére equation, advecting q with the derived velocity field, and then repeating, gives rise to serious complications with respect to the boundary conditions to use with the Monge-Ampére equation. A simple illustration of this is as follows, it can be shown that in semi-geostrophic theory the boundaries of the physical and dual domains map to one another (apart from in frontal regions). Now the physical domain, i.e. the region of the Earth we are interested in computing on, is fixed and known throughout the integration procedure. However the dual domain is evolving and the position of its boundary must be computed as part of the solution procedure, much like a Stefan problem [64]. The gradient of the potential R (respectively P) gives us $\mathbf{x}$ as a function of $\mathbf{X}$ (respectively $\mathbf{X}$ as a function of $\mathbf{x}$), since we know the values of $\mathbf{x}$, but not $\mathbf{X}$, on the boundary, it appears wiser to attempt to compute R rather than P. This justification is rather loose, see some of the references given in the previous Chapter for further rigorous discussions.

Now, as was said above, away from frontal regions boundaries map to boundaries in physical and dual space. In the presence of a front parts of the boundary of the dual domain map into the interior of the physical domain, we shall see an example of this later in this Chapter following on from details given in Appendix B. Up until now we have been using the support of ρ as our definition of the dual domain, as in the shaded region in figure 7-1. However, recall that by definition ρ represents the ratio of volume elements between physical and dual space. Therefore values of zero for ρ implies that the corresponding regions of physical space must have zero volume, i.e. we may extend the dual domain to infinity with all those regions outside the support of ρ mapping to the boundary (or front) in physical space. We write this as the following nonstandard boundary condition for the Monge-Ampére problem,

$$\nabla_X R \in \partial\Omega, \quad |\mathbf{X}| \to \infty, \tag{7.2}$$

where Ω represents the physical domain. The nonstandard form of this boundary condition means that standard methods for solving nonlinear elliptic equations cannot be applied without some thought on how to correctly apply the boundary condition.

In the semi-geostrophic context some work has already been done on solving the Monge-Ampére equation (but with simpler boundary conditions), in [10, 24, 25] Gauss-Seidel iterations improved by over-relaxation and alternative sweeping orders were considered. In [71] a multigrid technique is demonstrated to converge 50-80 times faster than a simple relaxation method based on Gauss-Seidel on a single grid.

The problems with solving the Monge-Ampére equation in the semi-geostrophic context discussed above means that we shall consider its numerical solution no further here. We shall rather take a closer look at the coordinate transformation that the Monge-Ampére equation describes, and link it in with the earlier work on numerical adaptivity from previous Chapters. Before this, in the next Section we shall mention a method for constructing the coordinate transformation based heavily upon the Legendre transformation properties of the problem.

7.3 The geometric method

The *geometric method* gives a means of finding an approximate solution to the Monge-Ampére problem, (i.e. of finding the convex potential P) in the case that we assume P to be piecewise linear. It is based intimately on the Legendre transform structure of the coordinate transformation Some background of the underlying ideas are given in [51, 52] and the development of a numerical realization of the method is discussed in [44, 45].

In two dimensions and in a very simplified way the method involves the following. Given a set of (M_i, θ_i) points, for example as given by the grid in figure 7-2, and a guess to the piecewise linear $P(x, z)$. Project the potential P down onto the (x, z) plane to give a set of 'elements', so that the face of P above element i has gradient (M_i, θ_i). We can easily calculate the areas of the elements, and these should be equal to an imposed set of ρ values. However in general there shall be some error (or residual) and the faces of P are adjusted (keeping the gradients fixed). The iteration being continued until some tolerance is achieved. Of course there are many other issues here that need to be addressed, for example the initial guess, precise method of iteration etc. We refer to [44, 46] here for more details, we simply note that although some very attractive results are achieved in two spatial dimensions, it is noted in [44] that the method has some severe limitations in its ability to compute time dependent solutions to problems in three spatial dimensions. This is due to the poor resolution and inaccuracies associated with the piecewise linear representation of P and equivalently the piecewise constant representation of the M, θ and ρ fields. We can draw here an analogy with what we concluded in Chapters 3 and 5, that is that in many situations where adaptivity in the form of dynamically evolving coordinate transformations is used the accuracy with which we solve the grid equations impinges directly on the overall solution accuracy. For example in Chapter 3 we rigorously proved that in order not to lose accuracy in the approximation of a self-similar solution we need to solve the underlying ODE and the coordinate transformations to the same order.

See [48] for a comparison of this method with a carefully designed 'conventional' implicit finite difference method. Good agreement with the geometric method is shown for some model problems and the finite difference method is importantly demonstrated to be able to handle discontinuous solutions representing atmospheric fronts.

7.4 Links with adaptivity

7.4.1 The moving mesh technique

Introduction to moving mesh adaptivity in two or more dimensions

Before we start to consider any possible similarities or links between adaptivity and the geostrophic coordinate transformation we shall review the moving mesh technique of adaptivity in several spatial dimensions. This shall be a natural extension of the equidistribution method described in Chapter 5.

In three dimensions the moving mesh approach [97, 98] to constructing coordinate transformations ($\boldsymbol{\xi} = \boldsymbol{\xi}(\mathbf{x}, t)$ or equivalently $\mathbf{x} = \mathbf{x}(\boldsymbol{\xi}, t)$) is to define $\boldsymbol{\xi}$ to be the function of $\mathbf{x}$ which minimizes a functional involving various adaptation properties (for example orthogonality, mesh smoothness, as well as adapting to a given rule or solution). For example consider the adaptation functional

$$I[\boldsymbol{\xi}] = \frac{1}{2} \int \sum_{i=1}^{3} (\nabla \xi_i)^T G_i^{-1} \nabla \xi_i \, d\mathbf{x},$$

where ∇ represents gradient with respect to $\mathbf{x}$ and the G_i are monitor functions, three by three symmetric positive definite matrices (this procedure acts to concentrate mesh points in regions where G_i is 'large') which are exact analogues of the monitor function M considered in Chapter 5. The Euler-Lagrange equations for which are

$$-\frac{\delta I}{\delta \xi_i} = \nabla \cdot \left(G_i^{-1} \nabla \xi_i\right) = 0, \quad i = 1, 2, 3. \tag{7.3}$$

It is now possible to construct the coordinate transformation by directly solving the nonlinear elliptic equations given by (7.3). However, as in the discussion of the one-dimensional case in Chapter 5, it is often advantageous to introduce a time derivative and relax the mesh towards that given by the exact solution of (7.3). We do that here by considering the gradient flow of $I[\boldsymbol{\xi}]$,

$$\frac{\partial \xi_i}{\partial t} = -\frac{1}{\tau} \frac{\delta I}{\delta \xi}, \quad i = 1, 2, 3. \tag{7.4}$$

Since this implies, for $\delta I / \delta \xi \neq 0$, that

$$\frac{dI}{dt} = \frac{\delta I}{\delta \xi} \cdot \frac{\partial \xi}{\partial t} = -\frac{1}{\tau} \left\| \frac{\delta I}{\delta \xi} \right\|^2 < 0,$$

we converge to a stable stationary point of the functional.

In the simplest case the monitor functions G_i can be identical for $i = 1, 2, 3$, and simply a scalar multiplied by the three-dimensional identity matrix (this is generally referred to as Winslow's method). For example the analogue of arc-length could be used by taking

$$G = \sqrt{1 + |\nabla u|^2} I_3,$$

where u is the solution to the underlying PDE we are trying to solve on our moving adaptive mesh. The monitor function could well include additional terms to control mesh orthogonality etc in the form of a penalty function. In practice it proves simpler and more convenient to solve (7.4) after interchanging dependent and independent variables. This is straightforward and we now solve a coupled nonlinear PDE for $\mathbf{x} = \mathbf{x}(\boldsymbol{\xi}, t)$, we do not give its form explicitly here but refer to [97, 98].

Of course to properly define the transformation PDE we need to impose some boundary conditions on the problem. The simplest possible case is simply to take Dirichlet conditions with the boundary points held fixed. This is fine if the solution to the underlying problem is evolving so that the behaviour we wish to use higher resolution on is away from the boundary of the physical domain. However in some situations, most importantly here the semi-geostrophic problem, all the interesting and complex behaviour occurs on, or near, the boundary. In which case we would like to be able to move boundary points along the boundary. We do this by solving a lower dimensional moving mesh equation on the boundary, e.g. if our problem was in two spatial dimensions we would solve one of the one dimensional MMPDEs given in Chapter 5, with a monitor function given by the projection along the boundary of the higher dimensional monitor function. For additional details and practical issues regarding the coupling of the different moving mesh equations see [97, 98].

Links with the geostrophic coordinate transformation

We now take a closer look at the coordinate transformation from physical to geostrophic or dual coordinates, we also choose to sometimes use the term computational coordinates for X since these are the variables in which computing will be carried out. Recall from earlier we had

$$\mathbf{X} \equiv (X, Y, Z)^T = \left(x + \frac{v_g}{f}, y - \frac{u_g}{f}, \frac{g\theta}{f^2\theta_0}\right)^T,$$

and

$$q = \frac{\partial(X, Y, Z)}{\partial(x, y, z)} = \det(\mathrm{Hess}_x(P)), \tag{7.5}$$

we shall now show some links with the theory of moving mesh partial differential equations as introduced above.

It is possible to write the continuous form of the first one-dimensional moving mesh partial differential equation (MMPDE) in the form (c.f. (5.9))

$$\frac{\partial x}{\partial \xi} = \frac{1}{M}, \tag{7.6}$$

where $M(x,u)$ is our monitor function. Notice the similarity with (7.5) if we take $M = q$, and identify the computational coordinates X and ξ. Now (7.6) becomes (on inverting both sides)

$$\frac{\partial X}{\partial x} = q.$$

Which is exactly the one-dimensional form of (7.5). Therefore the one dimensional MMPDE theory based on equidistribution as introduced in Chapter 5 exactly recreates the PDE controlling the geostrophic coordinate transformation if the monitor function is taken to be the potential vorticity, i.e. if we move our mesh points into regions of higher potential vorticity. We shall now see what happens if we use this as motivation for the three dimensional case which is the situation of physical interest.

Recall the equations (7.3) which control the numerical coordinate transformation in three dimensions. Notice what happens when we take our monitor functions to be equal $G_i \equiv G$ and, using the one-dimensional case considered above as motivation,

$$G = \mathrm{Hess}_x(P).$$

For a start the determinant of our monitor function is simply the potential vorticity, i.e.

$$\det(G) = q,$$

and one possible solution to the Euler-Lagrange equations (7.3) is

$$\mathbf{X} \equiv \boldsymbol{\xi} = \nabla_x P, \tag{7.7}$$

since then (using the symmetry of the Hessian if necessary),

$$G^{-1}\nabla\xi_i = \mathbf{e}_i, \quad i = 1,2,3; \qquad \mathbf{e}_1 = (1,0,0)^T, \text{ etc.}$$

Therefore given P the moving mesh theory recreates exactly the geostrophic coordinate transformation (7.7). In practice of course we would not ordinarily have access to P throughout the course of the integration and so we need to consider an alternative choice of monitor function which we can readily calculate throughout the integration. Now the monitor function that achieved what we desired above would have the effect

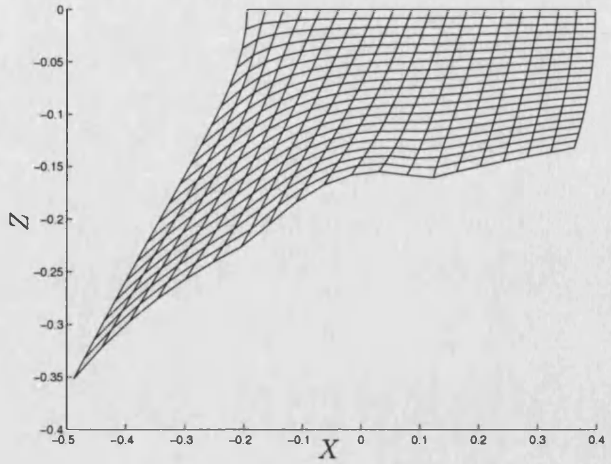


Figure 7-2: *The dual* (X, Z) *domain and an associated simple mesh of* 20×20 *points constructed using the Poisson mesh generation technique (7.8).*

of moving mesh points into regions where it was 'large', we can always approximate this by asking our mesh points to move into regions where the determinant of the above monitor function is large. In other words we can consider what happens if we actually take our monitor function to be $G = qI_3$. Hopefully this should give results not too far removed from the geostrophic coordinate transformation, and since we are not necessarily looking to recreate the coordinate transformation exactly, but rather to construct a mesh upon wish to discretize the problem, this may well be sufficient. We shall see an example of this monitor function correctly clustering mesh resolution around a front in the next Section.

We have thus shown a link between the usual analytical transformation found in the literature and our moving mesh adaptivity ideas discussed in previous Sections. A key point to take on board from these is that the equations governing the mesh transformation should be solved to high order, i.e. a smooth mesh should be used. This contrasts with the piecewise constant mesh transformation discussed in [51], as well as the geometric method discussed above

7.4.2 The parabolic umbilic example

We shall now use the so called parabolic umbilic as an example of a Legendre transform which may be used to give a solution to the geostrophic coordinate transformation which models an atmospheric front. The example is taken from [45, 44] and many further background details are given in Appendix B. Here we identify the dual domain in the SG notation with the computational domain in the adaptivity notation and feel free to use $\boldsymbol{\xi}$ and $\mathbf{X}$ synonymously.

The dual domain calculated in Appendix B is shown in figure 7-2. Overlaying the

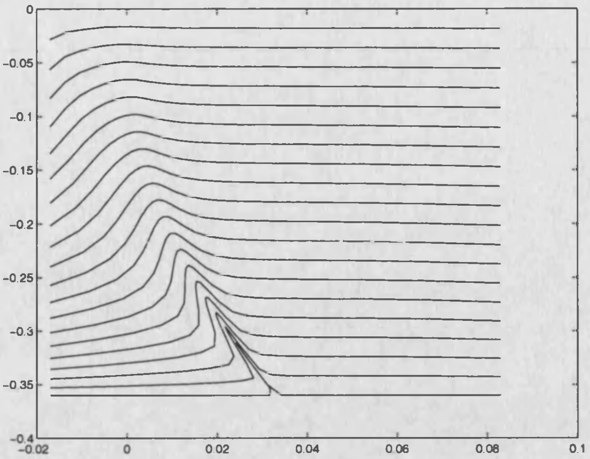


Figure 7-3: *The structure of the transformation in physical (x, z) space obtained via the parabolic umbilic example. Obtained by mapping the 'horizontal' grid lines shown in figure 7-2 to the physical domain under the exact transformation given by (B.3).*

domain is a simple grid used to obtain data values of ρ for use in the mesh generation. We obtain this mesh by simply solving Laplace's equations

$$\nabla^2 \xi = \nabla^2 \eta = 0, \tag{7.8}$$

on this dual domain. We employ a boundary condition for the ξ component as given by, $\xi = 0, 1$ on the left and right boundaries respectively, and ξ equal to the relative arclength along each of the top and bottom boundaries. Boundary conditions for the η equation are defined analogously. This method can be defined in the framework of Section 7.4.1 with the functional I given by

$$\iint \left((\nabla \xi)^2 + (\nabla \eta)^2 \right) \, dx \, dy,$$

i.e. with monitor functions $G_i = I$. The generalizations of this method where the right hand side of (7.8) is not identically zero are sometimes called Poisson or Thompson grid generators, see [167] for details. Note that the computational domain used for this adaptivity procedure is given by $[0, 1]^2$, however also note that the resulting adapted (in the adaptivity theory notation, physical) mesh as shown in figure 7-2 will itself serve as a computational grid for constructing the true physical mesh later, to avoid confusion we therefore refer to this as the dual domain/mesh. In the computations given here we use 20 mesh points in ξ and 20 in η.

In figure 7-3 we demonstrate the exact behaviour of the coordinate transformation by plotting the images of the 20 'horizontal' grid lines from figure 7-2 under the Legendre transform. Due to the form of this example (see Appendix B) we have an expression for the exact transformation. The front can be clearly seen. If we consider the 'bottom'

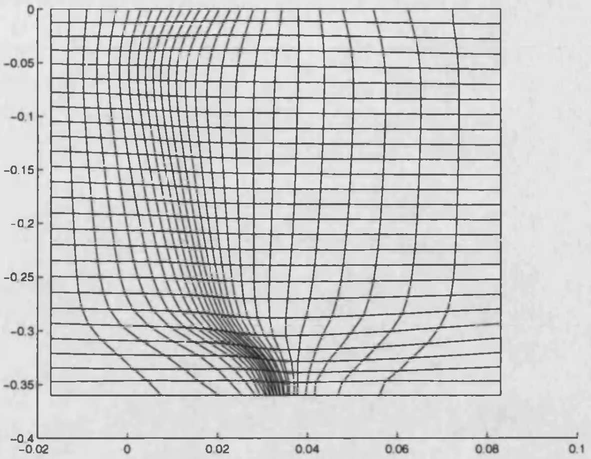


Figure 7-4: *The constructed mesh for the parabolic umbilic example with the (exact, c.f. (B.6)) potential vorticity as monitor function. Computed by integrating (7.4) to steady state using DASSL, see the discussion below.*

boundary from figure 7-2 and follow it along, noting that it maps to the 'bottom' line given in figure 7-3, notice that at a value of approximately $x = 0.35$ the line moves into the physical domain before coming back and (almost) meeting itself back on the boundary. Along this frontal line intruding into the physical domain each point must map from two points along the corresponding line in the dual domain. Since the variables X and Z can be thought of in terms of momentum and temperature respectively, this demonstrates the reason why the line in physical space can be thought of as representing a weather front. The discontinuity also demonstrating why an increase in numerical resolution is desirable around the frontal region.

Taking the monitor function to be simply the potential vorticity, which can easily be extracted from the exact coordinate transformation given by this example, we can now solve the mesh equations (7.4) or rather the corresponding problem given by interchanging dependent and independent variables as discussed above. Since all the interesting behaviour is occurring on the bottom boundary in this problem we also need to employ a one dimensional MMPDE as given in Chapter 5 along each boundary. For simplicity the two dimensional PDE was semi-discretized in space using centred differences and then solved using the ODE/DAE solver DASSL [26]. The integration was carried out for a sufficiently long time that the constructed mesh was qualitatively at a steady state. We argue that this statement is true both from inspecting results at earlier times, as well as the fact that a maximum allowed time step was imposed on the adaptive time stepping procedure DASSL employs, and that this very small (in comparison to the total length of the integration) time step was actually attained early on in the computation.

In figure 7-4 we see the resulting computed mesh or numerical coordinate transformation. As desired the mesh resolution is indeed concentrated about the frontal region.

7.4.3 The deformation method

An alternative method for performing mesh adaptivity, or equivalently for finding a coordinate transformation between meshes, is given by the deformation method [22, 112]. This method aims to construct the transformation between physical and computational space given a function representing the Jacobian determinant of the transformation, i.e. given the ratio between the volumes of elements, or in the semi-geostrophic context given the potential vorticity (or its inverse). So, immediately this deformation method appears to have some possible links with the semi-geostrophic theory and its coordinate transformation, which taken with the above work on relating the moving mesh technique to the semi-geostrophic theory could possibly lead to new discoveries on connections between these two alternative adaptivity methods. In addition the deformation method as a numerical technique is based upon a constructive proof by Moser [126, 54] on the existence of diffeomorphisms between volume elements on Riemannian manifolds. The geometric roots of this method in addition therefore imply the possibility of this method having some as yet unknown applications in geometric integration. No more shall be said on this method now, however due to the two interesting reasons just given these links shall be the subject of future work.

7.4.4 A new adaptivity technique

The work above on establishing links between the geostrophic coordinate transformation and the moving mesh technique for grid adaptivity provided a motivation for the use of potential vorticity as an error, or solution complexity, measure for use in the adaptivity procedure. We therefore used the well established, problem specific, geostrophic coordinate transformation to tell us something about a current method for performing mesh adaptivity. However, the links between semi-geostrophic theory and adaptivity have additional applications. For example the idea of imposing the value of the Jacobian determinant between physical and computational space can be seen as a higher dimensional analogue of the equidistribution principle from Chapter 5. Unfortunately this only yields one equation, and so in more than one dimension the problem of finding the coordinate transformation is under-determined. Semi-geostrophic theory and its links with an established adaptivity technique as given above now point to a solution to this problem, namely in considering one set of coordinate variables to be the gradient of some convex potential. This results in a Monge-Ampére equation controlling the adaptive coordinate transformation. We mentioned problems with boundary conditions for the Monge-Ampére equation above, but this was a problem intimately related to the formulation and structure of the semi-geostrophic system. This problem would no longer arise if the Monge-Ampére equation was used to provide an adapted mesh for a general problem where the physical and computational domains are assumed to be known. In addition the large body of theoretical results on elliptic equations, in particular those of Monge-Ampére type, could be used to provide rigorous (regularity for example) results on the resulting mesh. This is the topic of current ongoing work [178].

Note that a method with the same philosophy of controlling grid cell volumes via the Jacobian of the transformation is given in [5] and is based upon a generalization of the Poisson grid generator (7.8).

7.5 Summary of Chapter

In this Chapter, following a brief review of moving mesh methods in higher than one dimension, we have shown links between the coordinate transformation this technique yields and the geostrophic coordinate transformation of so much use in semi-geostrophic theory. Firstly, these links imply the possible use of potential vorticity as (at least a component of) a physically realistic and useful monitor function, rather than the more standard measures based on gradients or curvatures of the underlying solution to the problem. We have provided an example using a Legendre transform between physical and dual space which provides a model for an atmospheric weather front. Using potential vorticity as a monitor function was shown to yield a mesh which adapts well to the structure of the front. Since the numerical challenge with this system is in the vicinity of a front this appears to be a promising method for providing a mesh for solving the underlying equations (6.6)–(6.9). The monitor function, or equivalently the potential vorticity, may then be advected as a passive tracer for the system. In addition the mesh should adapt to and follow the front as it evolves in time, and the grid points should automatically redistribute themselves as the front eventually disappears. Finally, in a rather speculative note a possible way of unifying ideas with the deformation method and a new adaptivity technique were mentioned, with additional work in this direction left for the future.

Chapter 8

Potential vorticity advection

8.1 Overview of Chapter

In the previous chapter the Monge-Ampère equation, or equivalently the coordinate transformation, part of the semi-geostrophic system was discussed. It was also mentioned that potential vorticity is an incredibly important variable in this system since it can be used to determine all other dynamic fields. The accurate time integration (possibly for a large number of time steps) of the potential vorticity advection equation is therefore vital. In addition we are obviously interested in physically realistic solutions and should therefore attempt to preserve in our numerics as much of the underlying structure of the advection equation and its solution as possible. This problem therefore provides an ideal situation for the application of geometric integration ideas. We shall briefly consider some possibilities in this Chapter. The Sine-Euler truncation introduced in Chapter 2 shall be discussed. Due to its popularity with geophysicists semi-Lagrangian methods shall also be considered. Finally due to the problem with performing Hamiltonian truncations of noncanonical Hamiltonian PDEs a reformulation of the problem in terms of Clebsch variables shall be given.

8.2 Sine bracket type truncation

As was described in Section 2.4.2 the Sine-Euler truncation provides a means of discretizing the Euler equations (2.63) whilst capturing the Hamiltonian nature of the problem. Due to the similarities in structure between the Euler and semi-geostrophic equations it is natural to ask whether the Sine-Euler truncation has any possible application to the semi-geostrophic equations.

A problem arises now due to the more complex relation between the advected quantity (vorticity ω for the Euler equations and potential vorticity q (or its inverse ρ) for the semi-geostrophic equations) and the streamfunction. Recall for the Euler equations we had

$$\omega = \nabla^2 \Psi, \tag{8.1}$$

and for the semi-geostrophic equations

$$\rho = \det\left(\mathrm{Hess}_X\left(\frac{1}{2}(X^2+Y^2) - f^{-2}\Psi\right)\right). \tag{8.2}$$

Now we saw in Section 2.4.2 that in Fourier space, due to its linearity, relation (8.1) takes a simple form which can be used to arrive at an equation describing the Fourier modes of ω independently of Ψ. However in the case of relation (8.2) this is no longer possible, because the more complex nonlinear Monge-Ampére operator can not be explicitly inverted in Fourier space.

A possible solution (but almost certainly of no practical use) could be to hold Ψ constant (although in a numerical scheme it could be updated after each time step by solving the Monge-Ampére equation), following on from Section 2.4.2 we could then consider

$$\dot{\rho}_{\mathbf{m}} = \sum_{\mathbf{n}\neq\mathbf{0}} (\mathbf{m}\times\mathbf{n})\,\rho_{\mathbf{m}+\mathbf{n}}\Psi_{-\mathbf{n}}, \quad H(\rho) = \sum_{\mathbf{n}\neq\mathbf{0}} \rho_{\mathbf{n}}\Psi_{-\mathbf{n}}, \tag{8.3}$$

with the Poisson structure left unchanged. Although the possibility of retaining some structure of the problem is kept alive here, the constraint seems too harsh and unphysical. We therefore consider the Sine-Euler truncation no further and move on to other geometric possibilities for the advection problem.

8.3 Semi-Lagrangian methods

Semi-Lagrangian methods [160, 62] are very popular within the computational geophysical fluid dynamics community and we consider them here. In a sense they combine both the Eulerian and Lagrangian perspectives to fluid dynamics, and this could possibly open up some fresh ideas in terms of geometric integration. For this reason we consider semi-Lagrangian methods in general, and in particular when applied to problems of advection by an incompressible flow, i.e. by a nondivergent velocity field as in both the Euler and semi-geostrophic problems. To explain some of the basic ideas behind the method consider the advection of the quantity $\psi(x,t)$ in a one-dimensional flow field. The problem may be defined in Lagrangian form as

$$\frac{D\psi}{Dt} = 0, \tag{8.4}$$

or equivalently in Eulerian form as

$$\frac{\partial\psi}{\partial t} + u\frac{\partial\psi}{\partial x} = 0. \tag{8.5}$$

The equivalence of the two formulations follows from the definition of the material derivative operator, and also

$$\frac{dx}{dt} = u(x,t), \tag{8.6}$$

where we shall assume that $u(x,t)$ is a given velocity function. The basic idea behind semi-Lagrangian methods is very straight forward. Equation (8.4) tells us that ψ is constant along solutions of (8.6). We therefore simply integrate the differential equation (8.6) backwards in time a distance of one time step from a mesh point x_j to get the *departure* point x_j^d of the trajectory, we then set

$$\psi(x_j, t^{n+1}) = \psi(x_j^d, t^n). \tag{8.7}$$

However, at time level n we only know values of $\psi(x, t^n)$ at the mesh points $x = x_i$, $i = 0, \ldots, N$, we therefore need to interpolate to find a value of $\psi(x, t^n)$ at $x = x_j^d$ since x_j^d will generally not coincide with a mesh point.

Now suppose that we are in the case of M spatial dimensions and have the advection problem which when written in Lagrangian form is simply (8.4), or in Eulerian from we now write

$$\frac{\partial \psi}{\partial t} + \mathbf{u} \cdot \nabla \psi = 0, \tag{8.8}$$

where

$$\frac{d\mathbf{x}}{dt} = \mathbf{u}(\mathbf{x}, t). \tag{8.9}$$

The fundamental theorem of calculus tells us that

$$\psi(\mathbf{x}, t^{n+1}) = \psi(\mathbf{x}^d, t^n) + \int_C (dt, d\mathbf{x}) \cdot \left(\frac{\partial \psi}{\partial t}, \nabla \psi \right), \tag{8.10}$$

where C is an arbitrary contour in $M+1$-dimensional space-time connecting the points $(\mathbf{x}, t^{n+1})$ and $(\mathbf{x}^d, t^n)$. Given (8.8) we may rewrite (8.10) as

$$\psi(\mathbf{x}, t^{n+1}) = \psi(\mathbf{x}^d, t^n) + \int_C (d\mathbf{x} - \mathbf{u}dt) \cdot \nabla \psi. \tag{8.11}$$

Now notice what the freedom to choose $\mathbf{x}$, $\mathbf{x}^d$ and C allows us. For example, taking $\mathbf{x}^d = \mathbf{x} = \mathbf{x}_j$ a mesh point, and the contour C to be a straight line parallel to the t-axis (see figure 8-1a) gives us a purely Eulerian scheme.

Taking C to be a fluid particle trajectory (a solution of (8.9) beginning at a mesh point $\mathbf{x}^d = \mathbf{x}$, as in figure 8-1b) gives us a purely Lagrangian scheme, however we encounter the possibility of loss of resolutions in certain regions. Notice also that in this case the integral appearing in (8.11) vanishes.

The semi-Lagrangian method introduced above is given by taking C to be the trajectory

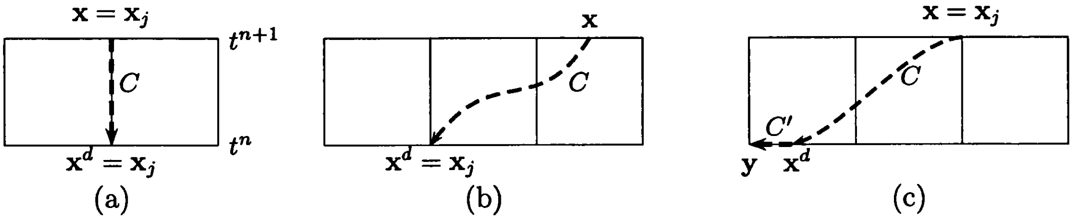


Figure 8-1: *Choice of integration contour C for the semi-Lagrangian method.*

of a fluid particle arriving at a grid point ($\mathbf{x} = \mathbf{x}_j$) at time t^{n+1}, as given in figure 8-1c. In this case the interpolation procedure is still required to deduce a value of $\psi(\mathbf{x}^d, t^n) = \psi(\mathbf{x}_j, t^{n+1})$ given discrete values of ψ at mesh points. There are obviously a wide range of interpolation techniques that may be employed here. We shall look at a way of reformulating this part of the overall method in the next Section.

For the advection equations in two spatial dimensions, if as in the advection parts of the Euler and semi-geostrophic problems we assume incompressible flow, i.e. there exists a streamfunction Ψ such that

$$\mathbf{u} = (u, v)^T = \left(-\frac{\partial\Psi}{\partial y}, \frac{\partial\Psi}{\partial x}\right)^T, \tag{8.12}$$

then the Lagrangian form of the problem way be written

$$\frac{d}{dt}\begin{pmatrix} x \\ y \\ \psi \end{pmatrix} = J\nabla\Psi, \quad J = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{8.13}$$

Therefore for the pure Lagrangian contour above, our numerical method should be Hamiltonian provided the trajectory calculations are performed with a symplectic algorithm. However, for the semi-Lagrangian method the interpolation procedure also needs to respect the Hamiltonian structure, we discuss this in the following Section.

An important point to note, especially in terms of geometric integration, is that for problems of the form (8.5) the integral of the transported field is conserved by the exact flow, but not by semi-Lagrangian methods in general. However it has been shown [20] that conservation can be recovered through the use of cubic spline interpolation. Also in [150] the conservation property is enforced through a finite volume approach to the interpolation step.

8.3.1 Interpolation via parameterized advection

Given that $\mathbf{y}$ is the mesh point closest to $\mathbf{x}^d$ (see fig 8-1c), i.e. we know the value of $\psi(\mathbf{y}, t^n)$. Suppose we augment the contour C chosen for the semi-Lagrangian method with another contour C' chosen to lie in the t-plane and connect the points $(\mathbf{x}^d, t^n)$ and $(\mathbf{y}, t^n)$ (in two or more spatial dimensions we could choose C' to be the union of contours parallel to the spatial coordinate axes). Extending (8.11) to this case, we can obviously write

$$\psi(\mathbf{x}, t^{n+1}) = \psi(\mathbf{y}, t^n) + \int_{C \cup C'} (d\mathbf{x} - \mathbf{u}dt) \cdot \nabla\psi,$$

which reduces to

$$\psi(\mathbf{x}, t^{n+1}) = \psi(\mathbf{x}^d, t^n) = \psi(\mathbf{y}, t^n) + \int_{C'} \nabla\psi \cdot d\mathbf{x}. \tag{8.14}$$

If we now take C' to be the contour

$$\mathbf{x}(\mathbf{y}, \tau) = \mathbf{y} - (\mathbf{y} - \mathbf{x}^d)\tau, \quad \tau \in [0, 1],$$

then, using the notation $\varphi(\mathbf{y}, \tau) \equiv \psi(\mathbf{x}(\mathbf{y}, \tau), t^n)$, (8.14) gives us that

$$\psi(\mathbf{x}, t^{n+1}) = \varphi(\mathbf{y}, \tau = 1) = \varphi(\mathbf{y}, \tau = 0) - \int_0^1 \nabla \cdot (\mathbf{U}\varphi)\, d\tau,$$

where we have introduced the notation $\mathbf{U} = \mathbf{y} - \mathbf{x}^d$. This is a formal solution to the constant velocity advection equation

$$\frac{\partial\varphi}{\partial\tau} + \nabla \cdot (\mathbf{U}\varphi) = 0, \tag{8.15}$$

over the τ interval $[0, 1]$ at grid point $\mathbf{y}$. Given $\varphi(\mathbf{y}, 0) = \psi(\mathbf{y}, t^n)$, a known value, we simply solve (8.15) up to $\tau = 1$ to give us

$$\varphi(\mathbf{y}, 1) = \psi(\mathbf{x}^d, t^n) = \psi(\mathbf{x}_j, t^{n+1}).$$

We have therefore expressed the interpolation problem as an equivalent advection problem for which many 'standard' algorithms exist [110]. Notice that (8.15) may be written as a Hamiltonian partial differential equation with Hamiltonian operator and functional given by

$$\mathscr{D} = -U\frac{\partial}{\partial x} - V\frac{\partial}{\partial y}, \quad \mathscr{H} = \frac{1}{2}\int \varphi^2\, d\mathbf{x}. \tag{8.16}$$

However we are now in a situation (since $\mathscr{D}$ does not now depend on ψ) where it is possible to find a discretization of (8.16) which preserves the Hamiltonian nature of the problem. For example we could again employ a spectral truncation. As stated in [175], for the infinite dimensional Hamiltonian system whose Hamiltonian operator is a constant differential operator the Fourier semi-discrete system is also Hamiltonian. The standard symplectic methods described in Chapter 2 could then be employed to step forward in time. However for an integration of the full problem (8.4) we are

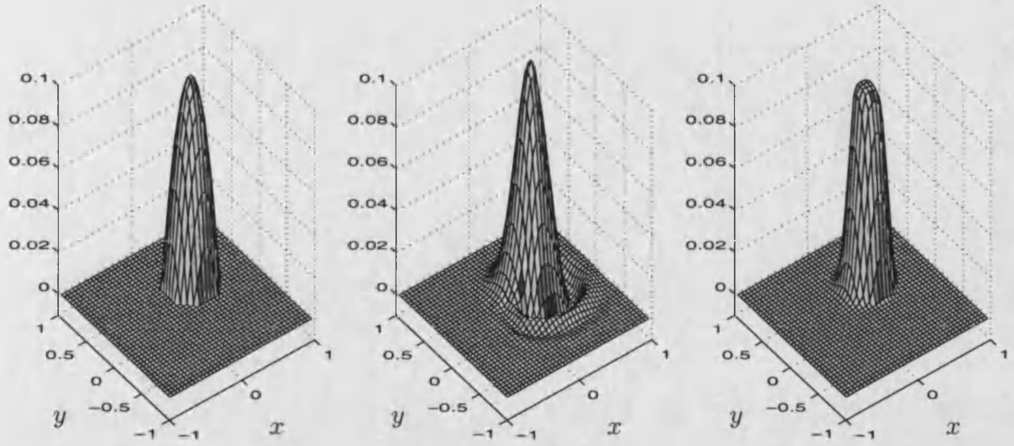


Figure 8-2: *Semi-Lagrangian method applied to the constant velocity field advection problem. (Left) Initial conditions. (Middle) Using cubic spline interpolation. (Right) Using superbee limited advection for the interpolation step.*

now in the situation of performing a distinct sequence of these smaller Hamiltonian problems. This is reminiscent of the problem we mentioned at the end of Section 8.2 where again a single continuous Hamiltonian problem could not be solved to construct a discrete solution of the original problem in time. There is however another attractive feature of the approach considered in this Section, again in [175] it is shown that if the Hamiltonian operator does not depend on ψ or its derivatives and if the Hamiltonian functional is quadratic then the semi-discrete Fourier spatial discretization preserves all conservation laws of the original Hamiltonian system.

8.3.2 Positivity preservation

There exist various methods for the integration of the advection problem (8.15) whilst preserving certain properties of the continuous problem, for example slope-limited, flux-limited and flux-corrected methods, which can possibly be shown to be total variation diminishing (TVD), essentially nonoscillatory (ENO), and also positive definite where the method never generates negative values from nonnegative initial data. Interpolation procedures based on these ideas (following on from the previous Section) are often termed *shape preserving* [157]. In the semi-geostrophic equations for example we are ultimately advecting an area ratio (ρ), physically of course this is not allowed to become negative, and in actual fact the Monge-Ampére equation associated with this problem will cease to be elliptic at points where this occurs. Now as was mentioned in Chapter 7 the distribution of ρ experiences a discontinuous jump from being positive and $\mathcal{O}(1)$ to zero across the boundary of its support. The problem of our chosen advection scheme giving negative values of ρ needs to be taken seriously since it is a standard problem in methods for hyperbolic problems [110] that for example a second order method such as Lax-Wendroff gives oscillatory solutions when confronted with discontinuous (or steep) data. This same problem will occur in all higher than first order (standard) interpolation procedures for use in the semi-Lagrangian algorithm. An example of the oscillations or wiggles causing negative values that can occur if standard interpolation (in actual fact cubic spline interpolation) is used is shown in figure 8-2. A qualitatively better picture is obtained if we use a two-dimensional flux-limited method to integrate (8.15), again see figure 8-2. For simplicity we use here the method supplied in conjunction with [111] and make use of the superbee limiter. Now even though this method is not strictly TVD or positive definite the superior qualitative performance can be seen in figure 8-2. The cubic spline method experiences an undershoot approximately ten times larger than that experienced by the method which uses the flux-limited scheme, both methods use the same ODE solution method with a time step of $2\pi/300$. This example was used for illustrative purposes only and in practise a fully positive definite scheme would be sought. See [169] for a related discussion of two-dimensional TVD schemes, and [165] for a discussion of the use of limiters to achieve genuinely positive schemes in multidimensions.

8.4 Clebsch variables

Given a noncanonical Lie-Poisson system it is possible to obtain an inflated canonical system of equations in terms of so called Clebsch variables. Note that if the Lie-Poisson system is arrived at by reducing a canonical system (e.g. in passing from Lagrangian to Eulerian variables in many fluid systems), the inflated system is not the same as the original formulation. If the inflated canonical system is solved for the Clebsch variables then a solution to the noncanonical system is immediately given. For further details see [106, 124, 125, 115, 75]. We shall discuss the general procedure, followed by the cases of the two-dimensional Euler equations and the semi-geostrophic equations. This has obvious applications to geometric integration since canonical systems provide no problem in general, whereas noncanonical systems provide huge problems, in finding truncations which respect the Hamiltonian nature of the system, as discussed in Chapter 2. See [37] for some additional merits of using Clebsch variables for numerical simulations in the context of an ideal compressible fluid.

We shall follow the descriptions given in [125, 124] very closely throughout this Section. Beginning with the finite dimensional case (recall Example 2.6), suppose we have a problem written in terms of a noncanonical Lie-Poisson bracket (using summation over repeated indices)

$$\{f,g\}_{LP} = w_k c^k_{ij} \frac{\partial f}{\partial w_i} \frac{\partial f}{\partial w_j}, \tag{8.17}$$

associated with the Lie algebra $\mathfrak{g}$ with structure constants c^k_{ij} (for example the problem of rigid body motion in three dimensions mentioned earlier, in which case $\mathfrak{g} = \mathfrak{so}(3)$). We would like to find new canonical variables for the problem which 'reduce' to the noncanonical variables. For the ideal fluid Clebsch found a set of variables that uniquely determine the usual physical variables, but the inverse of the transformation does not exist, this is where gauge conditions come in.

Suppose we write our noncanonical variables in terms of new canonical (which has yet to be shown) variables $\mathbf{p}$, $\mathbf{q}$

$$w_i = c_{ij}^k p_k q_j. \tag{8.18}$$

Given that the description of the problem in terms of $\mathbf{p}$, $\mathbf{q}$ is canonical with bracket

$$\{f,g\}_c = \frac{\partial f}{\partial q_i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q_i}, \tag{8.19}$$

where repeated indices are summed, we now show that (8.17) is obtained from this canonical bracket via the reduction defined through (8.18). Notice that

$$\frac{\partial f}{\partial p_i} = \frac{\partial f}{\partial w_j}\frac{\partial w_j}{\partial p_i} = \frac{\partial f}{\partial w_j}c_{jk}^i q_k, \quad \frac{\partial f}{\partial q_i} = \frac{\partial f}{\partial w_j}\frac{\partial w_j}{\partial q_i} = \frac{\partial f}{\partial w_j}c_{ji}^k p_k,$$

substituting into (8.19) gives

$$\begin{aligned}\{f,g\}_c &= \frac{\partial f}{\partial w_i}\frac{\partial j}{\partial w_j}(c_{jt}^k c_{ik}^r - c_{it}^k c_{jk}^r)p_r q_t \\ &= \frac{\partial f}{\partial w_i}\frac{\partial j}{\partial w_j}(c_{kt}^r c_{ij}^k)p_r q_t \\ &= w_k c_{ij}^k \frac{\partial f}{\partial w_i}\frac{\partial f}{\partial w_j} = \{f,g\}_{LP},\end{aligned}$$

where we have used the skew-symmetric and Jacobi identity properties[1] of structure constants. We have just shown that, if given a Lie-Poisson problem in terms of $\mathbf{w}$, we can inflate the system so that it has a canonical form in terms of the Clebsch variables $\mathbf{q}$ and $\mathbf{p}$. We can therefore numerically solve the canonical problem for $\mathbf{q}(t)$ and $\mathbf{p}(t)$ and then construct a solution to the noncanonical problem via the transformation (8.18).

We may follow a similar procedure in the infinite dimensional case. Suppose we have a noncanonical Lie-Poisson system in terms of the variable $\omega \in \mathfrak{g}^\star$, with the bracket taking the form

$$\{\mathscr{F},\mathscr{G}\}_{LP} = \left\langle \omega, \left[\frac{\delta\mathscr{F}}{\delta\omega}, \frac{\delta\mathscr{G}}{\delta\omega}\right]\right\rangle, \tag{8.20}$$

where $\langle\cdot,\cdot\rangle : \mathfrak{g}^\star \times \mathfrak{g} \rightarrow \mathbb{R}$ is a natural pairing between the Lie-algebra and its dual, and $[\cdot,\cdot] : \mathfrak{g} \times \mathfrak{g} \rightarrow \mathfrak{g}$ is the Lie-algebra bracket. See (6.24) for example, where the

[1] Skew-symmetry: $c_{ij}^k = -c_{ji}^k$. Jacobi identity: $c_{ij}^k c_{kl}^m + c_{li}^k c_{kj}^m + c_{jl}^k c_{ki}^m$.

pairing is simply the L_2 inner product and the Lie bracket is the canonical $[F,G] = F_XG_Y - F_YG_X$. Define the dual operator $[\cdot,\cdot]^\dagger : \mathfrak{g}^\star \times \mathfrak{g} \to \mathfrak{g}^\star$ as that which satisfies

$$\langle \omega, [F,G]\rangle = \left\langle [\omega, G]^\dagger, F\right\rangle . \tag{8.21}$$

The Clebsch transform analogous to (8.18) is then given by

$$\omega = [\Pi, Q]^\dagger .$$

Following the construction set out in the finite dimensional case we now use the variational chain rule to give

$$\frac{\delta \mathscr{F}}{\delta Q} = -\left[\Pi, \frac{\delta \mathscr{F}}{\delta \omega}\right]^\dagger, \quad \frac{\delta \mathscr{F}}{\delta \Pi} = \left[\frac{\delta \mathscr{F}}{\delta \omega}, Q\right].$$

Substituting into the canonical bracket (6.22), which may be written

$$\{\mathscr{F},\mathscr{G}\}_c = \left\langle \frac{\delta \mathscr{F}}{\delta Q}, \frac{\delta \mathscr{G}}{\delta \Pi}\right\rangle - \left\langle \frac{\delta \mathscr{G}}{\delta Q}, \frac{\delta \mathscr{F}}{\delta \Pi}\right\rangle, \tag{8.22}$$

we find, just as for the finite dimensional case (following the use of the Jacobi identity satisfied by $[\cdot,\cdot]$) that we arrive at the noncanonical Lie-Poisson bracket (8.20), see [125, 124] for the precise details.

8.4.1 Two-dimensional Euler equations

Following [124, 125], the Clebsch variables $Q(x,y,t)$ and $\Pi(x,y,t)$ are related to the scalar vorticity via

$$\omega(x,y,t) = [Q,\Pi], \tag{8.23}$$

where the bracket is given by $[f,g] = f_xg_y - f_yg_x$ and is therefore skew-adjoint. Substituting (8.23) into our Hamiltonian $\mathscr{H}[\omega]$ given in (2.60) we can calculate the equations of motion for Q and Π, they are

$$\frac{\partial Q}{\partial t} = \frac{\delta \mathscr{H}}{\delta \Pi}, \quad \frac{\partial \Pi}{\partial t} = -\frac{\delta \mathscr{H}}{\delta Q}.$$

The chain rule for functional differentiation [125] gives

$$\frac{\delta \mathscr{F}}{\delta Q} = \left[\Pi, \frac{\delta \mathscr{F}}{\delta \omega}\right], \quad \frac{\delta \mathscr{F}}{\delta \Pi} = \left[\frac{\delta \mathscr{F}}{\delta \omega}, Q\right],$$

and therefore the bracket for our system in terms of Clebsch variables is the *canonical*,

$$\{\mathscr{F},\mathscr{G}\}_c = \int \frac{\delta \mathscr{F}}{\delta Q}\frac{\delta \mathscr{G}}{\delta \Pi} - \frac{\delta \mathscr{F}}{\delta \Pi}\frac{\delta \mathscr{G}}{\delta Q}. \tag{8.24}$$

8.4.2 The semi-geostrophic equations

Following the previous Section on the Euler equations we write, using the notation of Chapter 6,

$$\rho(X,Y,t) = [Q,\Pi] = Q_X\Pi_Y - Q_Y\Pi_X, \tag{8.25}$$

notice that we immediately have the useful result that

$$Q := R_X = x, \quad \Pi := R_Y = y \tag{8.26}$$

satisfy relation (8.25). Recall from (6.21) that the evolution equation satisfied by ρ is (assuming here that $f = 1$),

$$\rho_t = -[\Psi, \rho]. \tag{8.27}$$

Substituting (8.25) into this yields

$$[x_t, y] + [x, y_t] = -[\Psi, [x,y]] = [x,[y,\Psi]] + [y,[\Psi,x]],$$

where we have made use of Jacobi's identity. Rearranging gives

$$[x, y_t + [\Psi,y]] + [x_t + [\Psi,x], y] = 0,$$

which is satisfied if

$$\begin{aligned}\frac{\partial x}{\partial t} &= -[\Psi,x] + \frac{\partial \gamma}{\partial y},\\ \frac{\partial y}{\partial t} &= -[\Psi,y] - \frac{\partial \gamma}{\partial x}.\end{aligned} \tag{8.28}$$

The function γ is an arbitrary function, representing a gauge invariance in this reformulation. Solutions of this system for any γ can be used to construct solutions of (6.21). Assume from here on that $\gamma \equiv 0$.

Note that (8.28) is exactly the canonical system derived in the previous Section, with the caveat that a minus sign appears, this is due to the fact that the semi-geostrophic Lie-Poisson bracket is minus the Euler one, and also (8.22), hence we need to take the negative of (8.24). In detail,

$$\mathscr{H} = \int \rho\Psi = \int \Psi(x_X y_Y - x_Y y_X) =: \int h,$$

with variational derivatives

$$\begin{aligned}\frac{\delta H}{\delta x} &= \frac{\partial h}{\partial x} - \frac{\partial}{\partial X}\frac{\partial h}{\partial x_X} - \frac{\partial}{\partial Y}\frac{\partial h}{\partial x_Y}\\ &= -\Psi_X y_Y + \Psi_Y y_X - \Psi y_{YX} + \Psi y_{XY} = -[\Psi, y],\end{aligned}$$

similarly

$$\frac{\delta H}{\delta y} = [\Psi, x].$$

Confirming that (8.28) is indeed a canonical system. We therefore now have

$$\begin{aligned}\frac{\partial x}{\partial t} &= -u_g x_X - v_g x_Y = -(y-Y)x_X - (X-x)x_Y,\\ \frac{\partial y}{\partial t} &= -u_g y_X - v_g y_Y = -(y-Y)y_X - (X-x)y_Y.\end{aligned}$$

So $\mathbf{x}$ is constant along a flow defined by $\dot{\mathbf{X}} = \mathbf{u}_g$. Now by thinking about the 'Lagrangian' viewpoint of SG theory given in Chapter 6 it can be seen this result, although slightly unconventional, is exactly what was to be expected.

We now mention a further useful result, see [180]. Suppose that the functions A_i, $i = 1, 2$, satisfy the equation

$$\frac{\partial A_i}{\partial t} = -[\Psi, A_i], \qquad (8.29)$$

then their Jacobian $[A_1, A_2]$ satisfies the same equation. This therefore simply restates that as desired we now have a solution to (8.27). Further, let the functions A_i, $i = 1, \ldots, n$, satisfy (8.29), then the functional

$$I = \int F(A_1, \ldots, A_n),$$

is an integral of motion of the system described by (8.29). Importantly note that

$$\int F(\rho) = \int F([Q, \Pi]),$$

is a possible example of this functional. Hence the family (6.26) of 'PV' conservation laws is preserved in this canonical reformulation in terms of Clebsch variables.

8.5 Summary of Chapter

In this Chapter we have looked at discretizations of the advection part of the SG problem. We specifically looked for methods which could preserve the Hamiltonian nature of the problem. To begin with we considered generalizations of the Sine-Euler truncation of the Euler equations which was discussed in Chapter 2. However problems due to the nonlinearity of the SG problem meant that we only considered this method very briefly.

We then looked at semi-Lagrangian methods. The original motivation for considering this family of methods was their popularity in the geophysical fluid dynamics community. The semi-Lagrangian method may be thought of as a combination of a particle trajectory calculation followed by an interpolation step. We showed that both of these components of the method can be interpreted as canonical Hamiltonian problems, and as such can each be solved by symplectic methods. However since the full method will then be given by an alternating sequence of these steps it is unclear at present exactly how much of the benefits of symplectic integrators will be inherited here. We then outlined a technique for preserving the nonnegativity of the advected ρ field, a property which is vital for constructing the coordinate transformations of Chapter 7.

Due to the problems with directly constructing Hamiltonian truncations of noncanonical Hamiltonian PDEs we concluded the Chapter with a discussion of a reformulation of the advection problem in terms of Clebsch variables. We showed that a new canonical system may be constructed from which the evolution of ρ may be recovered. This technique has the advantage that a Hamiltonian truncation of the new system is now easily achieved.

Chapter 9

Conclusions and further work

9.1 Conclusions

In this work, through the use of coordinate transformations or adaptivity, numerical methods invariant under scaling transformations have been constructed. For ODE problems these methods where shown to admit discrete self-similar solutions which uniformly approximate the true self-similar solutions for all time. These discrete solutions where also shown to inherit the stability of the continuous ones. This resulted in the conclusion, seen in practice in several examples, that general numerical solutions evolving from non self-similar initial data can converge to self-similarity. Of course this only occurs when this is the true behaviour of the problem being solved. Important for the applicability of these methods, this property is one which occurs often in problems arising from physical systems.

Again through the use of adaptivity, scaling invariant methods were also developed for PDE problems. Since maximum principles are of vital importance and often used in the analysis of PDEs, their uses in conjunction with self-similar solutions were discussed. In particular maximum principles were shown to be a means of extending the ODE results to prove the convergence of arbitrary semi-discrete numerical solutions to self-similarity. These results were all demonstrated for the porous medium equation, numerical experiments indeed showing the convergence of the method to self-similarity. It was however stated that the maximum principle will not hold in general for the transformed problem, and so a more general extension of the ODE theory is still needed. This is the subject of future work.

The interesting relationship between problems with symmetries and Hamiltonian structure was used as motivation for the development of numerical methods for ODEs which are both scaling invariant and symplectic. Problems with the use of adaptivity in symplectic methods were discussed, and an alternate coordinate transformation was utilized here. The resulting method was shown to work well on the Kepler problem. The coordinate transformation implied by scaling considerations was noted to have been derived by other means and shown to perform well in other sources. In addition symmetry considerations were shown to enable the construction of coordinate transformations which respect certain conservation laws of problems. This part of the thesis goes some way to exploring the important question of how different properties can be combined and preserved in geometric integrators.

Finally, the semi-geostrophic equations were considered. This problem was shown to possess a large amount of qualitative structure on which to base geometric methods. Linking in with other parts of this work, two particular properties were focused upon. The first a coordinate transformation which was shown to have many similarities with the adaptivity motivated coordinate transformations used to construct scale invariant methods for PDEs. The SG theory was used to inspire the use of potential vorticity as a monitor function to control adaptivity, as opposed to the more common use of arclength or curvature etc. This was shown to yield a mesh which adapted extremely well to the structure of an idealized atmospheric front. Secondly the Hamiltonian nature of the SG problem was considered and some possible means of constructing Hamiltonian truncations of the infinite-dimensional problem were discussed. Some interesting semi-Lagrangian techniques were developed following the correct interpretation of the interpolation step. Finally a reformulation of the problem in terms of Clebsch variables was given, this allowed the use of 'standard' Hamiltonian truncations.

9.2 Future work

As mentioned above a rigorous extension of the ODE theory of Chapter 3 to PDEs is required. Specifically the proof of the existence of discrete self-similar solutions uniformly approximating the true ones, and the convergence of general discrete solutions to self-similarity.

In [136] the backward error results for symplectic methods applied to Hamiltonian problems are shown to also apply to Lie group methods applied to problems invariant under a Lie group of transformations (e.g. scalings). It would therefore be interesting to see if this framework for analysing methods could be used to establish backward error results for scaling invariant methods.

Work is currently ongoing on recovering the separability property lost following the use of the Poincaré transformation in Chapter 4. This holds the promise of yielding very cheap methods through the use of adaptivity and efficient high order explicit splitting methods.

More additional work is needed on developing and assessing the possibilities of the 'new' approach to adaptivity in higher dimensions discussed in Chapter 7 and motivated by the SG coordinate transformation and its properties. Also note that a means of generalizing the equidistribution principle of Chapter 5 to allow for initially uniform meshes was outlined during the course of this work, it was shown to work well on the porous medium equation but needs further study.

It is known that in one-dimension equidistribution can be shown to be equivalent to a Legendre transform [12]. In addition in [46] the relation between mesh and Legendre duality is discussed, in particular Delaunay and Veronoi meshes are shown to be 'dual' to one another. It would be interesting to investigate further the relation between adaptivity interpreted in terms of a coordinate transformation and Legendre transforms between coordinates. A starting point could be the link shown in Chapter 7 between the SG coordinate transformation and higher dimensional moving mesh methods.

In [117] higher order analogues of the semi-geostrophic equations, also possessing geometric properties such as the useful coordinate transformation and Hamiltonian structure, are investigated. The existence of these models implies that the results of Chapters 7 and 8 may have direct applications to problems other than the semi-geostrophic equations, possibly ultimately to models used for operational forecasting of atmospheric and oceanic circulations. As these new higher order systems are developed some thought should therefore be given to the uses of geometric integration on them.

In [55, 66] adaptive methods of the type used in this thesis are employed in computations of meteorological flows. They demonstrate the possible usefulness and advantages of these adaptive methods when applied to geophysical problems. However further experimentation with moving mesh type (possibly geometric) adaptive methods on a range of serious model problems is required to truly assess their potential for future use. In particular after motivation from SG theory a study should be made of the use of potential vorticity instead of, and in conjunction with, quantities such as arclength in the design of appropriate monitor functions for use in geophysical applications.

Appendix A

Finding the symmetries of a differential equation

For completeness an example of the general procedure for obtaining the Lie point symmetries of a differential equation is demonstrated below, for additional details see [99, 128, 161]. Consider the gravitational collapse problem discussed in Chapters 2 and 3,

$$\frac{d^2}{dt^2}r + r^{-2} = 0. \tag{A.1}$$

We shall look for a transformation of the form

$$\mathbf{X}: \quad \tilde{t} = t + \lambda T(t,r), \quad \tilde{r} = r + \lambda R(t,r),$$

which leaves (A.1) invariant, i.e.

$$\mathbf{X}\left(\frac{d^2}{dt^2}r + r^{-2}\right) = 0 \quad \text{whenever (A.1) holds.}$$

Substituting the variables $\tilde{t}$, $\tilde{r}$ into (A.1) and using the chain rule leads to

$$\begin{aligned}\frac{d^2}{d\tilde{t}^2}\tilde{r} + \tilde{r}^{-2} = \frac{d^2}{dt^2}r + r^{-2} &+ \lambda\{R_{tt} + \dot{r}(2R_{tr} - T_{tt}) + \ddot{r}(R_r - 2T_t) \\ &+ \dot{r}^2(R_{rr} - 2T_{tr}) - 3\dot{r}\ddot{r}T_r - \dot{r}^3T_{rr} - 2Rr^{-3}\} + \mathcal{O}(\lambda^2),\end{aligned}$$

where $\dot{r} = dr/dt$. The $\mathcal{O}(\lambda)$ terms give the *linearized symmetry condition*

$$\begin{aligned}R_{tt} + \dot{r}(2R_{tr} - T_{tt}) - r^{-2}(R_r - 2T_t) &+ \dot{r}^2(R_{rr} - 2T_{tr}) \\ &- 3\dot{r}r^{-2}T_r - \dot{r}^3T_{rr} - 2Rr^{-3} = 0,\end{aligned}$$

where we have assumed (A.1). However R and T are independent of $\dot{r}$ and so we can equate powers of $\dot{r}$ to obtain the following determining equations for R and T,

$$\begin{aligned} R_{tt} - r^{-2}(R_r - 2T_t) - 2Rr^{-3} &= 0, \\ 2T_{tr} - T_{tt} + 3T_r r^{-2} &= 0, \\ R_{rr} - 2T_{tr} &= 0, \\ T_{rr} &= 0. \end{aligned}$$

The last two of which yield

$$T(r,t) = a(t)r + b(t), \quad R(r,t) = a'(t)r^2 + c(t)r + d(t),$$

where the functions of time a, b, c and d are constants of integration. Substituting these into the second determining relation and equating powers of r gives

$$a(t) = 0, \quad 2c'(t) = b''(t),$$

and so we have that

$$T(r,t) = b(t), \quad R(r,t) = \frac{1}{2}(b'(t) + A_1)r + d(t),$$

where A_1 is a constant. Now finally substituting these into the first determining equation and equating powers of r again yields

$$d(t) = 0, \quad b'''(t) = 0, \quad b'(t) = 3A_1,$$

therefore $b(t) = 3A_1 t + A_2$, for some constant A_2. We are left with

$$T = C_1 t + C_2, \quad R = \frac{2}{3}C_1 r,$$

for some constants C_1 and C_2. Hence every infinitesimal generator leaving (A.1) invariant is of the form

$$\mathbf{X} = C_1\mathbf{X}_1 + C_2\mathbf{X}_2,$$

where

$$\mathbf{X}_1 = \partial_t, \quad \mathbf{X}_2 = t\partial_t + \frac{2}{3}r\partial_r,$$

i.e. translations in time and a scaling transformation.

Appendix B

The parabolic umbilic and atmospheric fronts

B.1 Catastrophe theory

In general a mathematical model of a problem represents a state or configuration of the system of interest. As parameters in the model are smoothly and slowly varied, the structure of the solution manifold (for example the position of equilibrium points, etc.) will also change in a smooth manner (although of course bifurcations could occur). On the other hand *catastrophe theory* [113, 132, 152] is concerned with sudden and discontinuous changes in the solution to the system arising from small and smooth changes in one or more system parameter.

Consider for example the gradient system

$$\frac{du}{dt} = -\frac{\partial f(u;a)}{\partial u},$$

where a is a fixed parameter. The solution to this system converges to one of the equilibrium points which are given by the local minima of the potential function $f(u;a)$, assuming of course that the function f possesses any local minima.

Now suppose that the potential is given by

$$f(u;a) = u^3 + au. \tag{B.1}$$

Notice from figure B-1 that the qualitative behaviour of f changes as a is allowed to vary. As a passes from negative to positive the only local minima of f vanishes. Correspondingly the solution to this problem displays wildly different behaviour as a varies by only a small amount around zero.

This very simple example illustrates what is known as the *fold catastrophe*, and since

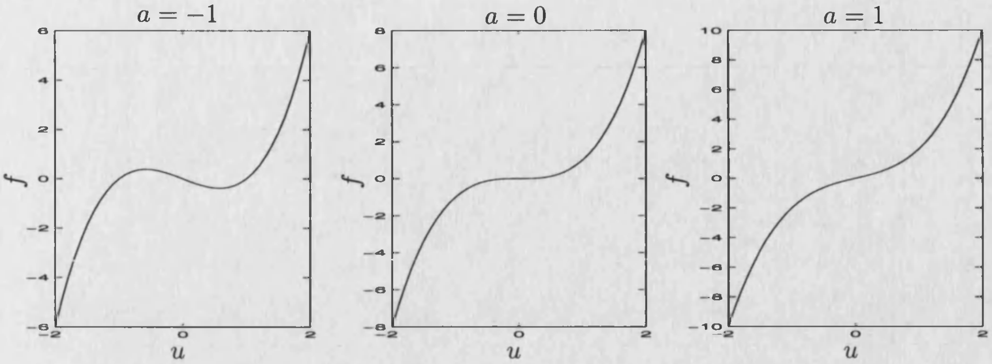


Figure B-1: *Behaviour of* $f(u;a) = u^3 + au$ *as a passes through zero.*

the problem has only one *state variable* u and only one *control variable* a, it is in fact the simplest possible example of a catastrophe. This is also used as a simple example of *bifurcation theory* [82] as two steady states (one unstable, the other stable) meet and annihilate each other as a passes through zero.

The function (B.1) is called the *universal unfolding* for the fold. This is due to the fact that it is a canonical form for this catastrophe, and all other polynomials of the same form as (B.1) with the same number of parameters must possess the same type of catastrophe. The existence of canonical forms describing the type of catastrophes that all other functions of the same family possess can only be carried on until a certain complexity is reached. There are seven of these so-called *elementary catastrophes.* From the simplest, the fold, after passing by the cusp, swallowtail, butterfly, hyperbolic umbilic, elliptic umbilic, the last and most complex on the list is the parabolic umbilic which is characterized by two state variables and four control variables. Although by this point in the list the high dimensional structures involved mean that this elementary catastrophe is actually far from elementary. These catastrophes have found a wide variety of applications in explaining and modelling many processes in a wide range of fields. We shall now look more closely at the parabolic umbilic which has been shown in [45, 44] to describe a 'solution' to part of the semi-geostrophic problem, and also to provide a model for an atmospheric front.

B.2 The parabolic umbilic

The presentation in [45, 44] shall be followed here to describe how the parabolic umbilic may be used to provide a Legendre transformation between the physical and geostrophic variables. This transformation contains a singularity which may be used as an idealized model of an atmospheric front. The data which this example provides shall be employed to construct an adapted mesh or numerical coordinate transformation in Chapter 7.

The universal unfolding of the parabolic umbilic is given by the single valued function

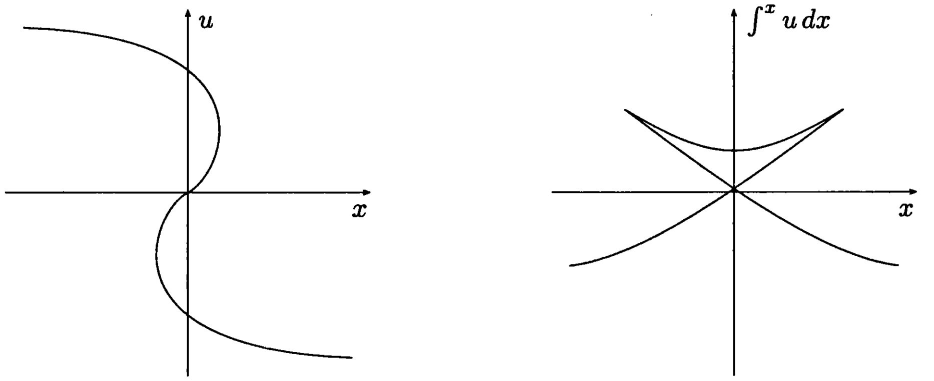


Figure B-2: *An idealized overturned solution and corresponding swallowtail structure.*

(defined over the geostrophic variable space, where M, θ are equivalent to X, Z upto a factor f)

$$R[M,\theta] = \frac{1}{4}M^4 + M\theta^2 + \frac{1}{40}M^2 + \theta^2. \tag{B.2}$$

The four control parameters have already been chosen, with the parameters acting as coefficients to the linear terms of the unfolding set to zero since these terms simply correspond to a change of coordinates under the Legendre transform.

Consider new variables x and z which span a so-called dual space (here the physical space),

$$x = \frac{\partial R}{\partial M} = M^3 + \theta^2 + \frac{1}{20}M, \quad z = \frac{\partial R}{\partial \theta} = 2M\theta + 2\theta. \tag{B.3}$$

The Legendre dual function to R is given, by definition, through

$$P[x,z] = xM + z\theta - R[M,\theta], \tag{B.4}$$

where (B.3) is inverted to give M and θ as function of x and z, which are then substituted into (B.4). This has the additional symmetrical relation of Legendre transforms

$$M = \frac{\partial P}{\partial x}, \quad \theta = \frac{\partial P}{\partial z}. \tag{B.5}$$

The functions P and R are called the Legendre transforms of one another. For our example (B.2) we have

$$P[x,z] = \frac{3}{4}M^4 + 2M\theta^2 + \frac{1}{40}M^2 + \theta^2,$$

It turns out that, due to the fact that we still need to invert (B.3), although R is single-valued P may actually be multi-valued. Note that

$$\begin{pmatrix} x \\ z \end{pmatrix} = \frac{\partial(x,z)}{\partial(M,\theta)} \begin{pmatrix} M \\ \theta \end{pmatrix} = \frac{\partial^2 R}{\partial(M,\theta)^2} \begin{pmatrix} M \\ \theta \end{pmatrix},$$

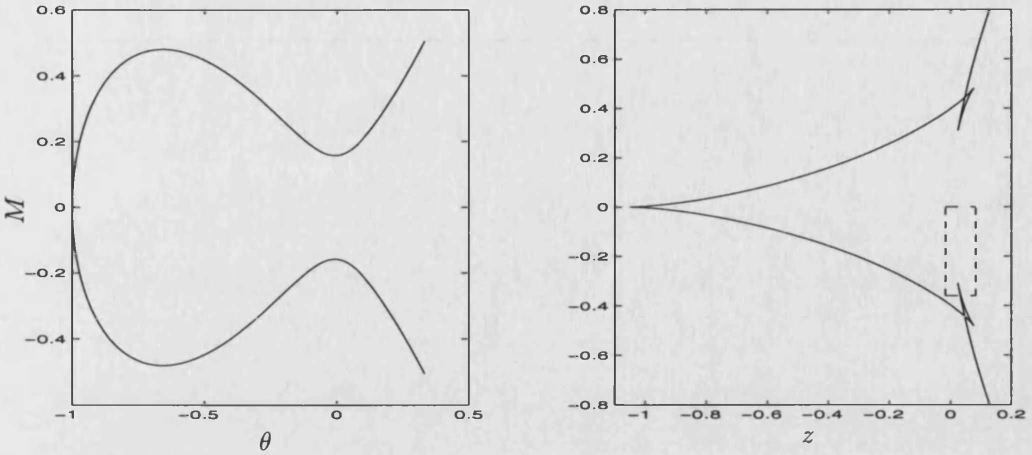


Figure B-3: *The locus of singularities (left) and its mapping under the Legendre transform into the physical space (right) for the parabolic umbilic.*

and so the inversion is not possible (the Legendre transform has a singularity) when the Hessian of R (equivalently the Jacobian of the coordinate transformation) has zero or infinite determinant. However due to the form of R and since M and θ are finite we only need concern ourselves with the zero case, i.e. when

$$0 = R_{MM}R_{\theta\theta} - R_{M\theta}^2 = 6M^3 + 6M^2 + \frac{1}{10}M + \frac{1}{10} - 4\theta^2. \tag{B.6}$$

Note that the right hand side of (B.6) gives us as expression for ρ in the notation of Chapter 6. The points in the (M, θ) plane which satisfy (B.6) are called the locus of singularities in the transformation and bound regions of different qualitative properties. This set of points may be mapped to the (x, z) space under the Legendre transform to give the bifurcation set of the parabolic umbilic which provides regions in which P has different multiplicities.

Chynoweth, Porter and Sewell [45] go through an extensive analysis of this parabolic umbilic example. In particular they consider the multi-valuedness of the function $P(x, z)$. With the atmospheric front example in mind they use a *physical stability argument* to select from the branches of P a convex single valued part, this is shown to contain an isolated gradient discontinuity representing an idealized atmospheric front (since there are jumps in the values of M and θ across it). The process of removing the unwanted branches of P can be thought of as a higher dimensional analogue of the process of removing the tail of a swallowtail which results in the insertion of a shock in an overturned solution, e.g. as can occur in Burgers' equation $u_t + uu_x = 0$. The left hand side of figure B-2 demonstrates an idealized example of an overturned (multivalued) solution, the right hand side is a plot of $\int^x u\,dx$ against x. It corresponds to a two-dimensional cross section of the three-dimensional bifurcation set of the swallowtail

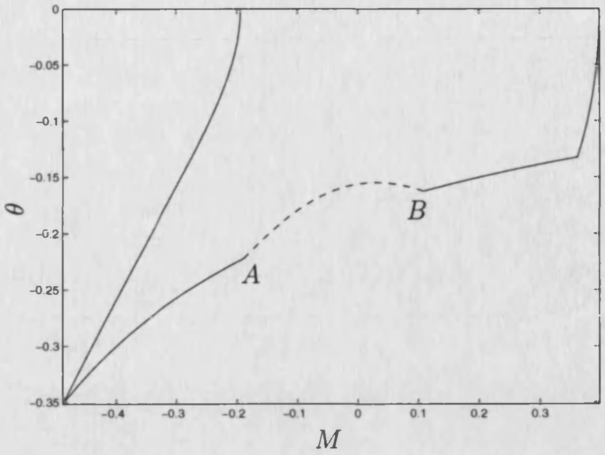


Figure B-4: *Domain of interest in* (M, θ) *space.*

catastrophe. The well known method of constructing physically meaningful solutions, through the insertion of a shock whose position is calculated using a conservation law argument [158, 177] to remove lobes of equal areas, can be shown to be equivalent to removing the tail from the swallowtail.

A plot of the locus of singularities (B.6) is given in the left of figure B-3, and in the right the mapping of this set under the Legendre transform into (x, z) space is given.

The right hand of figure B-3 gives domains in physical space where $P(x, z)$ has different multiplicities. In particular [45] it has multiplicity one outside of the transformed curve, e.g. for x less than some value. It has multiplicity three inside the curve, i.e. for x greater than the previously mentioned constant and z sufficiently small. Finally, it has multiplicity five inside the two swallowtail structures present.

Now, in order to satisfactorily model a state of the atmosphere (as defined through the semi-geostrophic equations) possessing a single weather front we choose only a small part of the physical domain portrayed above. The chosen region is given by the dashed box on the right of figure B-3. Within the box $P(x, z)$ and its gradient has multiplicity three or five, and therefore three or five (M, θ) points map from each (x, z) point in the region. By considering sets in (M, θ) space which map to the rectangle in (x, z) space, it is possible to disregard certain regions of (M, θ) and equivalently excise unwanted branches of $P(x, z)$. This procedure was indicated above where the analogy was drawn with the swallowtail and overturning solution. The resulting region in (M, θ) space is given in figure B-4. The four corners and 'sides' of the region in figure B-4 map to the four corners and sides of the physical domain as given by the dashed rectangle in figure B-3. However, the dashed portion of the bottom boundary in figure B-4 actually maps to the 'front' given by part of the swallowtail within the physical domain, the distinct points A and B both mapping to a single point on the bottom physical boundary. This gives us our jump in both the 'momentum' M and the 'temperature' θ across the front.

We use the domain so constructed and the ρ field over it as given by (B.6) as model data for constructing an adapted grid in Chapter 7.

Bibliography

[1] M. Abramowitz and I. Stegun, *Handbook of mathematical functions*, Dover Press, New York, (1964).

[2] M.P. Allen and D.J. Tildesley, *Computer simulations of liquids*, Clarendon Press, Oxford, (1987).

[3] A. Ambrosetti and G. Prodi, *A primer of nonlinear analysis*, CUP, (1993).

[4] S.C. Anco and G. Bluman, *Direct construction of conservation laws from field equations*, Phys. Rev. Lett. **78**, 2869–2873, (1997).

[5] D.A. Anderson, *Grid cell volume control with an adaptive grid generator*, Appl. Math. Comp. **35**, 209–217, (1990).

[6] A. Arakawa, *Computational design for long-term numerical integration of the equations of fluid motion: two-dimensional incompressible flow*, J. Comput. Phys. **1**, 119–143, (1966).

[7] V.I. Arnold, *Mathematical Methods of Classical Mechanics*, Springer-Verlag, New York, (1998).

[8] V.I. Arnold and B.A. Kheshin, *Topological methods in hydrodynamics*, Springer, (1998).

[9] D.G. Aronson, *The porous medium equation*, in 'Nonlinear diffusion problems', eds. A. Fasano and M. Primicerio, Springer lecture notes in mathematics 1224, 1–46, (1986).

[10] S. Baigent, *On the Integration of the Semi-Geostrophic Equations*, DPhil thesis, Oxford University, (1994).

[11] S. Baigent and J. Norbury, *Two discrete models for semi-geostrophic dynamics*, Physica D, **109**, 333–342, (1997).

[12] M.J. Baines and S.L. Wakelin, *Equidistribution and the Legendre transformation*, Univ. of Reading, Numer. Anal. Rep. 4/91, (1991).

[13] I.J. Bakelman, *Convex analysis and nonlinear geometric elliptic equations*, Springer, (1994).

[14] G.I. Barenblatt, *Selfsimilar turbulence propagation from an instantaneous plane source*, in 'Nonlinear dynamics and turbulence', eds. G.I. Barenblatt, G. Iooss and D.D. Joseph, Pitman, 48–60, (1980).

[15] G.I. Barenblatt, *Scaling, self-similarity and intermediate asymptotics*, CUP, (1996).

[16] G.I. Barenblatt and Ya.B. Zel'dovich, *Self-similar solutions as intermediate asymptotics*, Ann. Rev. Fluid Mech. **4**, 285–312, (1972).

[17] G.I. Barenblatt and A.J. Chorin, *New perspectives in turbulence: scaling laws, asymptotics, and intermittency*, SIAM Rev. **40**, 265–291, (1998).

[18] G.K. Batchelor, *An introduction to fluid dynamics*, CUP, (1967).

[19] J.-D. Benamou and Y. Brenier, *Weak existence for the semigeostrophic equations formulated as a coupled Monge-Ampére/transport problem*, SIAM J. Appl. Math. **58**, 1450–1461, (1998).

[20] R. Bermejo, *On the equivalence of semi-Lagrangian schemes and particle-in-cell methods*, Mon. Wea. Rev. **118**, 979–987, (1990).

[21] G. Birkhoff, *Hydrodynamics: a study in logic, fact and similitude*, Second edition, Princeton University Press, (1960).

[22] P. Bochev, G. Liao and G. de la Pena, *Analysis and computation of adaptive moving grids by deformation*, Numer. Meth. PDEs **12**, 489–506, (1996).

[23] S.D. Bond and B.J. Leimkuhler, *Time-transformations for reversible variable stepsize integration*, Numer. Algor. **19**, 55–71, (1998).

[24] J.G. Bonekamp, *Iterative solutions of the invertibility relation in semi-geostrophic dynamics*, Eindhoven University thesis, (1994).

[25] J.G. Bonekamp, *Iterative solutions of the 3D Monge-Ampere equation*, Met. Office forecasting research division report No. 126, (1994).

[26] K.E. Brenan, S.L. Campbell and L.R. Petzold, *Numerical solution of initial-value problems in differential-algebraic equations*, North-Holland, (1989).

[27] E. Buckingham, *On physically similar systems; illustrations of the use of dimensional equations*, Phys. Rev. **4**, 345–376, (1914).

[28] C.J. Budd and G.J. Collins, *Symmetry based numerical methods for partial differential equations*, in "Numerical Analysis 1997", eds. D. Griffiths, D. Higham and G. Watson, Pitman Research Notes in Mathematics Series **380**, 16–36, (1998).

[29] C.J. Budd, G.J. Collins, W-Z Huang and R.D. Russell, *Self-similar discrete solutions of the porous medium equation*, Phil. Trans. Roy. Soc. Lond. A **357**, 1047–1078, (1999).

[30] C.J. Budd, S. Chen and R.D. Russell, *New self-similar solutions of the nonlinear Schrödin ger equation with moving mesh computations*, J. Comp. Phys. **152**, 756–789, (1999).

[31] C.J. Budd, W-Z Huang and R.D. Russell, *Moving mesh methods for problems with blowup*, SIAM J. Sci. Comp. **17**, 305–327, (1996).

[32] C.J. Budd and A. Iserles (editors), *Geometric integration: numerical solution of differential equations on manifolds*, Phil. Trans. Roy. Soc. Lond. A **357**, 943–1133, (1999).

[33] C.J. Budd, B. Leimkuhler and M.D. Piggott, *Scaling invariance and adaptivity*, Appl. Numer. Math. **39**, 261–288, (2001).

[34] C.J. Budd and M.D. Piggott, *Geometric integration and its applications*, Proccedings of ECMWF workshop on *Developments in numerical methods for very high resolution global models*, 93–118, (2000).

[35] C.J. Budd and M.D. Piggott, *The geometric integration of scale invariant ordinary and partial differential equations*, J. Comp. Appl. Math. **128**, 399–422, (2001).

[36] C.J. Budd and M.D. Piggott, *Geometric integration and its applications*, 100pp, (2001), to appear in 'Foundations of Computational Mathematics', a volume of the 'Handbook of Numerical Analysis', ed. F. Cucker.

[37] O. Buneman, *Advantages of Hamiltonian formulations in computer simulations*, in Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems, AIP conference proceedings No. 88, eds. M. Tabor and Y. Treve, 137–143, (1982).

[38] J.C. Butcher, *The numerical analysis of ordinary differential equations*, John Wiley, (1987).

[39] M.P. Calvo and J.M. Sanz-Serna, *The development of variable-step symplecic integrators, with application to the two-body problem*, SIAM J. Sci. Comput. **14**, 936–952, (1993).

[40] M.P. Calvo, M.A. López-Marcos and J.M. Sanz-Serna, *Variable step implementation of geometric integrators*, Appl. Numer. Math. **28**, 1–16, (1998).

[41] M. Cannone and F. Planchon, *Self-similar solution for Navier-Stokes equations in* $\mathbb{R}^3$, Comm. PDEs, **21**, 179–193, (1996).

[42] B.J. Carr and A.A. Coley, *Self-similarity in general relativity*, Class. Quantum Grav. **16**, R31–R71, (1999).

[43] L.Y. Chen and N. Goldenfeld, *Numerical renormalization group calculations for similarity solutions and travelling waves*, Phys. Rev. E **51**, 5577–5581, (1995).

[44] S. Chynoweth, *The semi-geostrophic equations and the Legendre transform*, Ph.D thesis, University of Reading, UK, (1987).

[45] S. Chynoweth, D. Porter and M.J. Sewell, *The parabolic umbilic and atmospheric fronts*, Proc. R. Soc. Lond. A **419**, 337–362, (1988).

[46] S. Chynoweth and M.J. Sewell, *Mesh duality and Legendre duality*, Proc. R. Soc. Lond. A **428**, 351–377, (1990).

[47] S. Cirilli, E. Hairer and B. Leimkuhler, *Asymptotic error analysis of the adaptive Verlet method*, BIT, **39**, 25–33, (1999).

[48] M.J.P. Cullen, *Implicit finite difference methods for modelling discontinuous atmospheric flows*, J. Comp. Phys. 319–348, (1989).

[49] M.J.P. Cullen, *New mathematical developments in atmosphere and ocean dynamics, and their application to computer simulations*, Met. Office forecasting research scientific paper No. 48, (1997). To appear in 'Large-scale atmosphere-ocean dynamics', eds. I. Roulstone and J. Norbury, CUP, (2001).

[50] M.J.P. Cullen, J. Norbury, and R.J. Purser, *Generalised Lagrangian solutions for atmospheric and oceanic flows*, SIAM J. Appl. Math. **51**, 20–31, (1991).

[51] M.J.P. Cullen and R.J. Purser, *An extended Lagrangian theory of semi-geostrophic frontogenesis*, J. Atmos. Sci. **41**, 1477–1497, (1984).

[52] M.J.P. Cullen and R.J. Purser, *Properties of the Lagrangian semi-geostrophic equations*, J. Atmos. Sci. **46**, 2684-2697, (1987).

[53] M. Cullen, D. Salmond and P. Smolarkiewicz, *Key numerical issues for the development of the ECMWF model*, Proccedings of ECMWF workshop on *Developments in numerical methods for very high resolution global models*, 183–206, (2000).

[54] B. Dacorogna and J. Moser, *On a partial differential equation involving the Jacobian determinant*, Ann. Inst. Henri Poincaré, **7**, 1-26, 1990.

[55] G.S. Dietachmayer and K.K. Droegemeier, *Application of continuous dynamic grid adaptation techniques to meteorological modeling. Part I: basic formulation and accuracy*, Mon. Wea. Rev. **120**, 1675–1706, (1992); G.S. Dietachmayer, *Part II: efficiency*, Mon. Wea. Rev. **120**, 1707–1722, (1992).

[56] V.A. Dorodnitsyn, *Symmetry of Finite-Difference Equations*, in 'CRC Handbook of Lie Group Analysis of Differential Equations, Volume I : Symmetries, Exact Solutions and Conservation Laws', ed. N.Ibragimov, CRC Press, 365–403, (1993).

[57] V.A. Dorodnitsyn, *Finite-difference models exactly inheriting symmetry of original differential equations*, in 'Modern group analysis: Advanced analytical and computational methods in mathematical physics', Kluwer, 191–201, (1993).

[58] V.A. Dorodnitsyn, *Noether-type theorems for difference equations*, IHES/M/98/27, Bures-sur-Yvette(France), (1998).

[59] P.G. Drazin and R.S. Johnson, *Solitons: an introduction*, CUP, (1989).

[60] L. Dresner, *Similarity solutions of nonlinear partial differential equations*, Pitman Research Notes in Mathematics, Longman, **88**, (1983).

[61] L. Dresner, *Applications of Lie's theory of ordinary and partial differential equations*, IOP Publishing, (1999).

[62] D.R. Durran, *Numerical methods for wave equations in geophysical fluid dynamics*, TAM 32, Springer, (1999).

[63] K. Engø and S. Faltinsen, *Numerical integration of Lie-Poisson systems while preserving coadjoint orbits and energy*, SIAM J. Numer. Anal. **39**, 128–145, (2001).

[64] R. Fazio, *A similarity approach to the numerical solution of free boundary problems*, SIAM Rev. **40**, 616–635, (1998).

[65] Feng Kang, *On difference schemes and symplectic geometry*, Proc. 1984 Beijing symp. diff. geometry and diff. equations, Beijing, Science Press, 42–58, (1985).

[66] B.H. Fiedler and R.J. Trapp, *A Fast Dynamic Grid Adaption Scheme for Meteorological Flows*, Mon. Wea. Rev. **121**, 2879–2888, (1993).

[67] J. Ford, *The Fermi-Pasta-Ulam problem: paradox turns discovery*, Phys. Rep. **213**, 271–310, (1992).

[68] L.E. Fraenkel, *An introduction to maximum principles and symmetry in elliptic problems*, CUP, (2000).

[69] A. Friedman and S. Kamin, *The asymptotic behaviour of gas in an n-dimensional porous medium*, Trans. Amer. Math. Soc. **262**, 551–563, (1983).

[70] E. Frisch, *Turbulence: The legacy of A.N. Kolmogorov*, CUP, (1995).

[71] S.R. Fulton, *Multigrid solution of the semigeostrophic invertibility relation*, Mon. Wea. Rev. **117**, 2059–2066, (1989).

[72] N. Goldenfeld, O. Martin and Y. Oono, *Intermediate asymptotics and renormalization group theory*, J. Scient. Comput. **4**, 355–372, (1989).

[73] H. Goldstein, *Classical Mechanics*, Second edition, Addison-Wesley Publishing, (1980).

[74] G.H. Golub and C.F. van Loan, *Matrix Computations*, Second edition, Johns Hopkins University Press, (1989).

[75] V. Goncharov and V. Pavlov, *On the Hamiltonian approach: applications to geophysical flows*, Nonlinear Processes in Geophysics, **5**, 219–240, (1998).

[76] Z. Ge and J.E. Marsden, *Lie-Poisson Hamilton-Jacobi theory and Lie-Posson integrators*, Phys. Lett. A. **133**, 134–139, (1988).

[77] I.M. Gelfand and S.V. Fomin, *Calculus of variations*, Prentice-Hall, (1963). Reprinted by Dover (2000).

[78] M. Golubitsky, I. Stewart and D.G. Schaeffer, *Singularities and groups in bifurcation theory: volume II*, Springer-Verlag, (1988).

[79] D.F. Griffiths, *The dynamics of some linear multistep methods with step-size control*, in 'Numerical Analysis 1987', eds. D.F. Griffiths and G.A. Watson, Pitman, (1988).

[80] D.F. Griffiths and J.M. Sanz-Serna, *On the scope of the method of modified equations*, SIAM J. Sci. Stat. Comput. **7**, 994–1008, (1986).

[81] P. Grindrod, *Patterns and waves*, OUP, (1991).

[82] J. Guckenheimer and P. Holmes, *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, Springer-Verlag, (1983).

[83] E. Hairer, *Variable time step integration with symplectic methods*, Appl. Numer. Math. **25**, 219–227, (1997).

[84] E. Hairer, *Numerical geometric integration.* Unpublished lecture notes, (1999), available at http://www.unige.ch/math/folks/hairer/.

[85] E. Hairer and Ch. Lubich, *The life-span of backward error analysis for numerical integrators*, Numer. Math. **76**, 441–462, (1997).

[86] E. Hairer, S.P. Nørsett and G. Wanner, *Solving ordinary differential equations I.* 2nd edition, Springer series in computational mathematics **8**, Springer-Verlag, Berlin, (1993).

[87] E. Hairer and D. Stoffer, *Reversible long-term integration with variable stepsizes*, SIAM J. Sci. Comput. **18**, 257–269, (1997).

[88] T. Holder, B. Leimkuhler and S. Reich, *Explicit variable step-size and time-reversible integration*, Appl. Numer. Math. **39**, 367–377, (2001).

[89] D.D. Holm, *Symmetry breaking in fluid dynamics: Lie group reducible motions in real fluids*, PhD thesis, University of Michigan, (1976).

[90] B.J. Hoskins, *The geostrophic momentum approximation and the semi-geostrophic equations*, J. Atmos. Sci. **32**, 233–242, (1975).

[91] B.J. Hoskins and F.P. Bretherton, *Atmospheric Frontogenesis Models: Mathematical Formulation and Solution*, J. Atmos. Sci. **29**, 11–37, (1972).

[92] B.J. Hoskins, M.E. McIntyre and A.W. Robertson, *On the use and significance of isentropic potential vorticity maps*, Q. J. R. Meteorol. Soc. **111**, 877–946, (1985).

[93] M. Huang, *A Hamiltonian approximation to simulate solitary waves of the Korteweg-de Vries equation*, Math. Comp. **56**, 607–620, (1991).

[94] W. Huang and B. Leimkuhler, *The adaptive Verlet method*, SIAM J. Sci. Comput. **18**, 239–256, (1997).

[95] W. Huang, Y. Ren and R.D. Russell, *Moving mesh partial differential equations (MMPDES) based on the equidistribution principle*, SIAM J. Numer. Anal. **31**, 709–730, (1994).

[96] W. Huang, Y. Ren and R.D. Russell, *Moving mesh methods based on moving mesh partial differential equations*, J. Comp. Phys. **112**, 279–290, (1994).

[97] W. Huang and R.D. Russell, *A high dimensional moving mesh strategy*, Appl. Numer. Math. **26**, 998–1015, (1999).

[98] W. Huang and R.D. Russell, *Moving mesh strategy based on a gradient flow equation for two-dimensional problems*, SIAM J. Sci. Comput. **20**, 998–1015, (1999).

[99] P.E. Hydon, *Symmetry methods for differential equations*, CUP, (2000).

[100] A. Iserles, *A first course in the numerical analysis of differential equations*, CUP, (1996).

[101] A. Iserles, H.Z. Munthe-Kaas, S.P. Nørsett and A. Zanna, *Lie-group methods*, Acta Numerica 215–365, (2000).

[102] S. Kamenomostskaya, *The asymptotic behaviour of the solution of the filtration equation*, Israel J. Math. **14**, 279–290, (1973).

[103] C. Kane, J.E. Marsden and M. Ortiz, *Symplectic-energy-momentum preserving variational integrators*, J. Math. Phys. **40**, 3353–3371, (1999).

[104] C. Kane, J.E. Marsden, M. Ortiz and M. West, *Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems*, Int. J. Num. Meth. Eng. **49**, 1295–1325, (2000).

[105] J.R. King, *Self-similar behaviour for the equations of fast nonlinear diffusion*, Phil. Trans. R. Proc. Soc. A **343**, 337–375, (1993).

[106] H. Lamb, *Hydrodynamics*, Sixth edition, (1932), Dover, N.Y. Reprinted by CUP, (1971).

[107] J.S.W. Lamb and J.A.G. Roberts, *Time-reversal symmetry in dynamical systems: a survey*, Physica D **112**, 1–39, (1998).

[108] J.D. Lambert, *Numerical methods for ordinary differential equations: the initial value problem*, Wiley, (1991)

[109] B. Leimkuhler, *Reverible adaptive regularization: perturbed Kepler motion and classical atomic trajectories*, Phil. Trans. Roy. Soc. Lond. A **357**, 1101–1133, (1999).

[110] R.J. LeVeque, *Numerical methods for conservation laws*, Birkhauser-Verlag, (1990).

[111] R.J. LeVeque, *High-resolution algorithms for advection in incompresible flow*, SIAM J. Numer. Anal. **33**, 627–665, (1996).

[112] G. Liao, T-W. Pan and J. Su, *Numerical grid generator based on Moser's deformation method.* Numer. Meth. PDEs **10**, 21–31, (1994).

[113] Y.-C. Lu, *Singularity theory and an introduction to catastrophe theory*, Springer-Verlag, (1976).

[114] J.E. Marsden and T.S. Ratiu, *Introduction to mechanics and symmetry*, Second edition, Springer-Verlag, (1999).

[115] J.E. Marsden and A. Weinstein, *Coadjoint orbits, vortices, and Clebsch variables for incompressible fluids*, Physica D **7**, 305–332, (1983).

[116] J.E. Marsden and M. West, *Discrete mechanics and variational integrators*, Acta Numerica 2001, 357–514, (2001).

[117] M.E. McIntyre and I. Rolustone, *Are there higher-accuracy analogues of semigeostrophic theory?*, to appear in 'Large-scale atmosphere-ocean dynamics II: Geometric methods and models', eds. I. Roulstone and J. Norbury, CUP, (2001).

[118] R.I. McLachlan, *Explicit Lie-Poisson integration and the Euler equations*, Phys. Rev. Lett. **71**, 3043–3046, (1993).

[119] R.I. McLachlan, *Symplectic integration of Hamiltonian wave equations*, Numer. Math. **66**, 465–492, (1994).

[120] R.I. McLachlan, *The world of symplectic space*, New Scientist, 19 Mar 1994, (1994).

[121] R.I. McLachlan and G.R.W. Quispel, *Six lectures on the geometric integration of ODEs.* in 'Foundations of computational mathematics', eds. R. DeVore, A. Iserles and E. Süli, LMS Lecture note series 284, 155–210, (2001).

[122] R.I. McLachlan, G.R.W. Quispel and G.S. Turner, *Numerical integrators that preserve symmetries and reversing symmetries*, SIAM J. Numer. Anal. **35**, 586–599, (1998).

[123] R. McLachlan, I. Szunyogh and V. Zeitlin, *Hamiltonian finite-dimensional models of baroclinic instability*, Phys. Lett. A **229**, 299–305, (1997).

[124] P.J. Morrison, *Poisson brackets for fluids and plasmas*, in Mathematical Methods in Hydrodynamics and Integrability in Dynamical Systems, AIP conference proceedings No. 88, eds. M. Tabor and Y. Treve, 13–46, (1982).

[125] P.J. Morrison, *Hamiltonian description of the ideal fluid*, Rev. Mod. Phys. **70**, 467–521, (1998).

[126] J. Moser, *On the volume elements on a manifold*, Trans. A.M.S. **120**, 286–294, (1965).

[127] J.D. Murray, *Mathematical biology*, Springer-Verlag, (1989).

[128] P.J. Olver, *Applications of Lie groups to differential equations*, Springer, New York, (1986).

[129] N. Padhye and P.J. Morrison, *Fluid element relabeling symmetry*, Physics letters A, **219**, 287-292, (1996).

[130] J. Pedlosky, *Geophysical fluid dynamics*, Second edition, Springer-Verlag, (1979).

[131] L.A. Peletier, *The porous medium equation*, in 'Applications of nonlinear analysis', eds. H. Amann, N. Bazley and K. Kirchgrassner, Pitman, 229–241, (1981).

[132] T. Poston and I. Stewart, *Catastrophe theory and its applications*, Pitman, (1978).

[133] G.E. Prince and C.J. Eliezer, *On the Lie symmetries of the classical Kepler problem*, J. Phys. A: Math. Gen. **14**, 587–596, (1981).

[134] M.H. Protter and H.F. Weinberger, *Maximum principles in differential equations*, Springer-Verlag, New York, (1984).

[135] R.J. Purser and M.J.P. Cullen, *A duality principle in semi-geostrophic theory,* J. Atmos. Sci. **44**, 3449–3468, (1987).

[136] S. Reich, *Backward Error Analysis for Numerical Integrators*, SIAM J. Numer. Anal. **36**, 1549–1570, (1999).

[137] R.D. Richtmyer and K.W. Morton *Difference methods for initial value problems*, John Wiley and Sons, Inc., New York, (1967).

[138] J. Robinson, *Infinite-dimensional dynamical systems*, CUP, (2001).

[139] I. Roulstone and J. Norbury, *A Hamiltonian structure with contact geometry for the semi-geostrophic equations*, J. Fluid. Mech. **272**, 211–233, (1994).

[140] R.D. Ruth, *A canonical integration technique*, IEEE Trans. Nucl. Sci. **30**, 2669–2671, (1983).

[141] R. Salmon, *Lectures on geophysical fluid dynamics*, OUP, (1998).

[142] A. Samarskii, V. Galaktionov, S. Kurdyumov and A. Mikhailov, *Blow-up in quasilinear parabolic equations*, Walter de Gruyter, (1995).

[143] R. Samtaney, *Computational methods for self-similar solutions of the compressible Euler equations*, J. Comput. Phys. **132**, 327–345, (1997).

[144] R. Samtaney and D.I. Pullin, *On initial-value and self-similar solutions of the compressible Euler equations*, Phys. Fluids, *8*, 2650–2655, (1996).

[145] J.M. Sanz-Serna, *Geometric Integration* in 'The state of the art in numerical analysis', eds. I.S. Duff and G.A. Watson, Oxford: Clarendon, 121–143, (1997).

[146] J.M. Sanz-Serna and M.P. Calvo, *Numerical Hamiltonian problems*, Chapman and Hall, London, (1994).

[147] W. Sarlet and F. Cantrijn, *Generalizations of Noether's theorem in classical mechanics*, SIAM Rev. **23**, 467–494, (1981).

[148] Ch. Schlier and A. Seiter, *Symplectic Integration and Classical Trajectories: A Case Study*, J. Chem. Phys. A **102**, (102), 9399–9404, (1998).

[149] B.F. Schutz, *Geometrical methods of mathematical physics*, CUP, (1980).

[150] J.S. Scroggs and F.H.M. Semazzi, *A conservative semi-Lagrangian method for multidimensional fluid dynamics applications*, Num. Meth. PDEs **11**, 445–452, (1995).

[151] L.A. Segel, *Simplification and scaling*, SIAM Rev. **14**, 547–571, (1972).

[152] M.J. Sewell, *On Legendre transformations and umbilic catastrophes*, Math. Proc. Camb. Phil. Soc. **83**, 273–288, (1978).

[153] L.F. Shampine, *ODE solvers and the method of lines*, Num. Meth. for PDEs **10**, 739–755, (1994).

[154] T.G. Shepherd, *Symmetries, Conservation laws, and Hamiltonian structure in Geophysical Fluid Dynamics*, Advances in Geophysics, **32**, 287–338, (1990).

[155] R.D. Skeel, *Integration schemes for molecular dynamics and related applications*, in 'The graduate student's guide to numerical analysis 1998', eds. M. Ainsworth, J. Levesley and M. Marletta, Springer-Verlag, 119–176, (1999).

[156] R.D. Skeel and C.W. Gear, *Does variable step size ruin a symplectic integrator?*, Physica D **60**, (1992), 311-313.

[157] P.K. Smolarkiewicz and G.A. Grell, *A class of monotone interpolation schemes* J. Comp. Phys. **101**, 431–440, (1992).

[158] J. Smoller, *Shock waves and reaction-diffusion equations*, Springer-Verlag, (1983).

[159] E.A. Spiegel and G. Veronis, *On the Boussinesq approximation for a compressible fluid*, Astrophys. J. **131**, 442–447, (1960).

[160] A. Staniforth and J. Côté, *Semi-Lagrangian integration methods for atmospheric models: a review*, Mon. Wea. Rev. **119**, 2206–2223, (1991).

[161] H. Stephani, *Differential equations: Their solution using symmetries*, CUP, (1989).

[162] D.M. Stoffer, *Variable steps for reversible integration methods*, Computing **55**, 1–22, (1995).

[163] D. Stoffer and K. Nipp, *Invariant curves for variable step size integrators*, BIT **31**, 169–180, (1991).

[164] A.M. Stuart and A.R. Humphries, *Dynamical Systems and Numerical Analysis*, CUP, (1996).

[165] A. Suresh, *Positivity-preserving schemes in multidimensions*, SIAM J. Sci. Comput. **22**, 1184–1198, (2000).

[166] G.J. Sussman and J. Wisdom, *Chaotic evolution of the solar system*, Science, **257**, 56–62, (1992).

[167] J.F. Thompson, Z.U.A. Warsi and C.W. Mastin, *Numerical grid generation*, North-Holland, (1985).

[168] J.M.T. Thompson and H.B. Stewart, *Nonlinear dynamics and chaos*, Wiley, (1986).

[169] J.D. Towers, *TVD schemes for two-dimensional scalar conservation laws*, preprint, (2001).

[170] D.L. Turcotte, *Fractals in fluid mechanics*, Ann. Rev. Fluid Mech. **20**, 5–16, (1988).

[171] J.L. Vazquez, *Asymptotic behaviour and propogation properties of the one-dimensional flow of gas in a porous medium*, Trans. Amer. Math. Soc. **277**, 507–527, (1983).

[172] J.L. Vazquez, *An introduction to the mathematical theory of the porous medium equation*, in 'Shape optimisation and free boundaries' ed. M. Delfour, Kluwer Academic, 347–389, (1992).

[173] L. Verlet, *Computer "experiments" on classical fluids. I: Thermodynamic properties of Lennard-Jones molecules*, Phys. Rev. **159**, 98–103, (1967).

[174] R. de Vogelaere, *Methods of integration which preserve the contact transformation property of Hamiltonian equations*, Tech. report No. 4, Dept. Math., Univ. Notre Dame, (1956).

[175] D. Wang, *Semi-discrete Fourier spectral approximations of infinite dimensional Hamiltonian systems and conservation laws*, Computers Math. Applic. **21**, 63–75, (1991).

[176] R. Warming and B. Hyett, *The modified equation approach to the stability and accuracy of finite-difference methods*, J. Comp. Phys. **14**, 159–179, (1974).

[177] G.B. Whitham, *Linear and nonlinear waves*, Wiley, (1974).

[178] J.F. Williams, Private communication, (2001).

[179] J. Wisdom and M. Holman, *Symplectic maps for the N-body problem*, Astron. J. **102**, 1528–1538, (1991).

[180] V.E. Zakharov, *The algebra of integrals of motion of two-dimensional hydrodynamics in Clebsch variables*, Funct. Anal. Appl. **23**, 189–196, (1989).

[181] V.E. Zakharov and E.A. Kuznetsov, *Hamiltonian formalism for systems of hydrodynamic type*, Sov. Sci. Rev, Section C: Math. Phys. Rev. **4**, 167-220, (1984). See also: *Hamiltonian formalism for nonlinear waves*, Physics–Uspekhi, **40**, 1087–1116, (1997).

[182] V. Zeitlin, *Finite-mode analogs of 2D ideal hydrodynamics: coadjoint orbits and local canonical structure*, Physica D, **49**, 353–362, (1991).

[183] Ya.B. Zel'dovich and G.I. Barenblatt, *The asymptotic properties of self-modelling solutions of the nonstationary gas filtration equations*, Soviet Phys. Doklady, **3**, 44–47, (1958).

[184] Ya.B. Zel'dovich and Yu.P. Raizer, *Physics of shock waves and high-temperature hydrodynamic phenomena*, Volumes I & II, Academic Press, (1967).

[185] O.C. Zienkiewicz and R.L. Taylor, *The finite element method, Volume 2: Solid and fluid mechanics, dynamics and non-linearity*, Fourth edition, McGraw-Hill, (1991).