A spectral Lagrange-Galerkin method for convection-dominated diffusion problems

A. F. Ware

October 1991

Contents

Chapter 1

Introduction

The modelling of pure convection or convection-dominated processes is a central problem in fields such as meteorology, oil reservoir simulation, modelling of aerodynamic or geophysical flows, magnetohydrodynamics and many others. For many standard numerical methods there is a trade-off between excessive numerical diffusion on the one hand, and unphysical oscillations on the scale of the mesh on the other. However, at least as far back as the work of Courant, Isaacson and Rees [17] and Ansorge [3], it was realised that exploitation of the characteristics could give rise to effective numerical schemes for the solution of hyperbolic problems. When other terms, such as diffusion, are present in small amounts, the equation may be regarded as 'nearly hyperbolic', and various options for the incorporation of the extra terms into a scheme designed for the purely hyperbolic case are possible, giving rise to methods that are capable of avoiding the undesirable trade-off mentioned above.

A closely related idea is the use of Lagrangian coordinates, which may be used to improve the treatment of convective terms in fluid dynamical applications. As opposed to Eulerian methods, where an observer associated with a grid point watches the world evolve around him from a *fixed* vantage point, in a scheme based on Lagrangian coordinates a similar observer finds himself travelling with a fluid particle. From this perspective the time evolution of the flow may well be smoother, and thus easier to model numerically.

Such concepts have formed the basis of a wide variety of numerical schemes since their initial application, and are the central theme of this Thesis. We shall begin by giving a brief review of a selection of the range of methods that have been devised for convection-dominated problems, giving particular emphasis to those based on the characteristics/Lagrangian ideas discussed above.

1.1 Particle and free-Lagrange methods

The common feature of the methods to be described in this section is the presence of a set of points or particles that are transported under the action of some convective field. Consider a time-dependent convection-diffusion equation

$$\begin{aligned} u_t + \nabla \cdot f(u) &= \nabla \cdot (b \nabla u) \quad \text{in } R^n \times (0,T], && (1.1\text{a}) \\ u(\cdot, 0) &= u_0, && (1.1\text{b}) \end{aligned}$$

where $\nabla \cdot f(u) = a \cdot \nabla u$; a, b are functions of x and t, and a may depend on u. Let $x(t)$ satisfy the equation

$$\frac{dx(t)}{dt} = a(x(t), t). \tag{1.2}$$

Then, writing $u = u(x(t), t)$, (1.1a) may be rewritten in terms of coordinates satisying (1.2), giving

$$\frac{du}{dt} = \nabla \cdot (b \nabla u), \tag{1.3}$$

which is in the form of a heat equation.

1.1.1 Particle methods

Particle methods (see [63]) model the solution of (1.1) by calculating the evolution of a set of particles whose coordinates satisfy (1.2). In the absence of diffusion ($b = 0$) (1.1) thus reduces to the problem of solving the set of ordinary differential equations (1.2). In the fluid dynamics arena, particle methods were first applied to the solution of Euler's equations in vorticity-streamfunction form. Based on the principle that, for non-viscous flows, the vorticity in the field is transported by the flow, the particles are taken to be vortex 'blobs', and particle methods in this context are known as vortex methods.

The earliest examples of vortex methods involved the use of point vortices, so that the solution was represented by a linear combination of Dirac distributions. However, when two vortices approach one another, the velocities that they induce upon one another may become arbitrarily large. Thus the Dirac distributions are *mollified* [14] to give 'blob' vortices. Calculations based on these have been very successful and are surveyed in [45].

The convergence of vortex methods was first proved in [32], [33], and these results were generalised and extended to three dimensions by Beale and Majda [6], where they demonstrated that with a judicious choice of the mollifier used for the Dirac distributions, infinite order convergence analogous to that of spectral methods may be obtained. In [2], Anderson and Greengard give convergence results for a fully discrete scheme relying on stability and consistency results obtained in [6]. A common feature of the above convergence results is that if δ is a measure of the support of the mollified Dirac distributions and h the initial spacing between the particles, then the convergence rate depends on the way in which $\delta/h \to 0$. Typically, a relationship of the form $\delta = h^{1+\epsilon}$ is enforced, with $\frac{1}{2} \leq \epsilon < 1$.

The extension of the vortex method to viscous flows introduces issues which have still to be fully resolved. Two main approaches have been advocated. Chorin [14] used a random walk method to simulate the diffusion process. This approach has been successfully applied in a variety of situations, but suffers from low accuracy. An alternative is a deterministic vortex method: one possibility is to introducing a splitting of the equations as studied by Beale and Majda [7], involving a redistribution of the vorticity at each time level. More recently various generalisations that involve no splitting have been proposed [51], [25]. One problem here is that, since particle methods are automatically adaptive, the particles are likely to congregate in areas of high spatial activity, and to be missing from others, and the accurate treatment of diffusion will be difficult in this case. Such considerations lead to the proposal [15] of a particle-grid superposition scheme, whereby a grid of fixed particles is maintained along with the moving grid to help with the diffusion.

1.1.2 Free-Lagrange methods

Most of the algorithms mentioned in the previous section are truly Lagrangian, in the sense that the particles are allowed to move freely throughout the computation according to the flow field. Another type of method that fits this description is the free-Lagrange method [27]. Here there is a grid which develops according to the flow field, thus incorporating the convective terms, and the remaining terms may be dealt with in a finite-difference, finite-element or finite-volume manner. For each point a record is kept of which are the neighbouring grid points, which must be updated regularly to avoid problems such as mesh tangling. A finite element/volume formulation is developed in [64], [65], based on a locally-Delaunay triangulation with an efficient algorithm for the update. Such methods seek to avoid the problems associated with points absenting themselves from areas where they might be required for the approximation of diffusion etc. by allowing for the addition and removal of grid points as and when required. An alternative strategy, a moving point method as described by Farmer [23], is to interpolate on to a fixed grid at each time step. Such an approach combines advantages of both the Eulerian and Lagrangian frameworks, but the use of interpolation requires care to be taken to avoid numerical diffusion.

1.1.3 Moving-grid methods

We mention here a further class of methods, generically termed moving-grid methods. These methods are adaptive, in that the grid points are required to move around according to some criterion, but they are more general than those described above, since this criterion need not be the satisfaction of (1.2). A commonly applied criterion is the principle of *equidistribution*, enforced by the use of an arc-length monitor function. A comparison of two finite difference moving grid methods and a moving finite element method is carried out in [28], where their comparison of the methods takes into account their robustness and also the time-smoothness of the resulting system of ordinary differential equations.

In the moving finite element method [53, 54], [4], [90], the movement of the nodes is controlled by a minimisation procedure, in line with the finite element approach. The method has encountered problems with points drifting close together, and with parallelism, causing the mass matrix to become singular, and has needed problem-dependent tuning to avoid these difficulties. Controlling the mesh in a systematic way remains an area of active research.

1.2 The semi-Lagrangian method

Purely Lagrangian methods of the type discussed in the previous section are susceptible to the fact that an initially regularly-spaced set of points will generally evolve into a highly irregular set. In meteorological circles *semi-Lagrangian* methods have achieved great popularity. Here a different set of points is used at each time step. These are chosen so that they arrive exactly at points of a regular mesh at the top time level; in general then the feet of the trajectories of these points will not be points on the regular mesh, and a characteristic feature of the semi-Lagrangian methods used in meteorological circles has been the use of interpolation from values at nearby mesh points. We shall return to this point below.

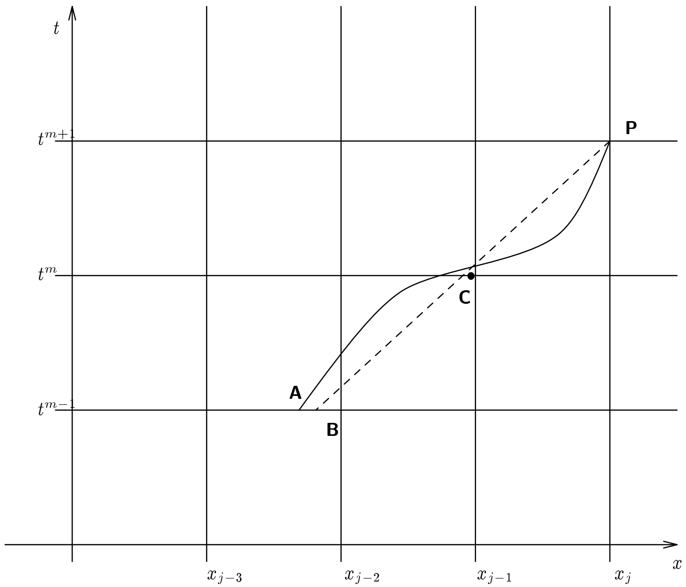

Figure 1.1: The trajectory tracing of the semi-Lagrangian method.

The recent review by Staniforth and Côté [73] gives an excellent overview of work in this area. A major problem faced by numerical methods in meteorology is that the maximum timestep is governed by stability rather than accuracy considerations. The semi-Lagrangian treatment introduced by Robert [69] lead to a reported factor of six increase in the size of the allowable timestep. The size of the timestep is here dictated primarily by accuracy, since the scheme is unconditionally stable. Morevover, the phase error is greatly reduced as compared to an Eulerian method.

1.2.1 Trajectory tracing

The tracing of the trajectories is found to be a decisive factor in the accuracy of the schemes. Consider equation (1.1) in the absence of diffusion ($b = 0$) and with $\nabla \cdot a = 0$. The three time-level scheme first utilised by Robert is given by

$$\frac{u^{m+1}(x) - u^{m-1}(x - 2\alpha^m(x))}{2\Delta t} = 0, \tag{1.4a}$$

$$\alpha^m(x) = \Delta t a^m(x - \alpha^m(x)). \tag{1.4b}$$

Equation (1.4b) represents the use of the midpoint rule for the integration of (1.2), and is solved by a Picard iteration. Although (1.4a) places no stability-related timestep restriction, the convergence of this iteration does introduce a restriction on Δt, but this remains an order of magnitude larger than the stability restriction to which an Eulerian timestepping scheme is subject.

Figure 1.1 illustrates the solution of (1.2) by (1.4b). The solid line represents the exact trajectory passing through the point (x_j, t^{m+1}), landing at **A**, and the dashed line is the approximation given by (1.4b), which passes through **C** (coordinates $(x_j - \alpha^m(x_j), t^m)$) and lands at **B** (coordinates $(x_j - 2\alpha^m(x_j), t^{m-1})$).

The three level scheme (1.4) is almost a set of two decoupled two level integrations, apart from the fact that in applications the velocity field a depends on u either explicitly or implicitly through other equations to which (1.1) is coupled. Temperton and Staniforth [88], by extrapolating a to t^m from t^{m-1} and t^{m-3}, managed to effectively decouple these processes and arrived at a two time level scheme, reducing the amount of work involved.

1.2.2 Interpolation

A second crucial factor in the success of these methods is the interpolation that is used at the feet of the trajectories. Too low an interpolation introduces excessive damping—too high a formula is expensive. For most applications, cubic interpolation is found to be an effective compromise. For some meteorological applications, the avoidance of overshoots and undershoots is of prime importance; with this in mind Williamson and Rasch [91] make use of shape-preserving and monotone interpolation. Another alternative is the non-interpolating semi-Lagrangian method of Ritchie [66]. Here the velocity field a is split up into two parts. The first part a^* is such that each trajectory through a grid point at the top time level lands exactly on the nearest grid point to the base of the original trajectory through that grid point, thus requiring no interpolation. The residual velocity field $a^+ = a - a^*$ is dealt with in a standard Eulerian fashion. The stability advantages of the semi-Lagrangian method are maintained, since this residual velocity always satisfies a unit CFL condition, and the damping associated with the interpolation is avoided. However, the dispersion properties of Eulerian methods have returned, albeit with smaller coefficients associated with them (since they are associated with the residual velocity). These ideas have been generalised and extended by Smolarkiewicz and Rasch [72], who develop a formalism which may be used to convert any advection algorithm into a semi-Lagrangian format.

Ritchie [66] also argues that the non-interpolating scheme opens the way for the incorporation of a *spectral* space-discretisation into the semi-Lagrangian method. Spectral methods have long been used in meteorological models, but their advantages would have been lost with the low order grid point interpolation of the semi-Lagrangian method. Ritchie [67, 68] has incorporated a spectral discretisation and compared the interpolating and non-interpolating versions; he finds the interpolation does indeed lead to a deterioration of performance at short scales.

1.2.3 Other interpolations

Similar ideas have appeared in the engineering literature, with applications in such areas as subsurface hydrology. In the Holly-Preissman scheme [38], two functions are evolved in time, representing the solution and its first derivative at the grid points. These are found at the new time level by tracing back along characteristics through grid points and using a cubic Hermite polynomial interpolation of the solution and its first derivatives at the grid points adjacent to the feet of the characteristics to find the function values there. Diffusion and other terms are incorporated using a splitting technique. Recent extensions include the inclusion of second derivatives into the above scheme, resulting in a fifth-order accurate method [92], and allowing the characteristics to reach back more than one time level, with a view to reducing the numerical dispersion associated with the interpolation [93].

1.3 Finite element methods

It is well known that finite element approximations enjoy optimal approximation properties in appropriate integral norms, most straightforwardly when applied to self-adjoint elliptic problems. Similar results for non self-adjoint problems can be obtained by the Petrov-Galerkin approach, whereby the test functions are selected with a view to symmetrising the bilinear form associated with the method [5].

Semidiscrete Galerkin finite element approximations of convection-diffusion problems also possess many attractive approximation and conservation properties. However, when time discretisations are introduced, these properties deteriorate noticeably, and unless severe restrictions are placed on the scheme, unphysical oscillations occur on the scale of the mesh. Attempts to avoid such phenomena include the adapting of the Petrov-Galerkin approach, choosing the test functions to satisfy a unit CFL condition [56], and the Taylor-Galerkin approach of Donea [19], which enjoyed similar accuracy properties yet was easier to extend to more complicated problems. Another approach, which has been extremely successful, is the streamline-diffusion method, first introduced by Hughes and Brookes [39]. The method can be viewed as a Petrov-Galerkin method with a test function modified by adding a multiple of a linearised form of the hyperbolic operator applied to the test function. This gives added stability without sacrificing accuracy, and in the scalar one-dimensional case can be viewed as adding artificial diffusion along the streamlines. References to recent work can be found in [42].

1.3.1 The characteristic Galerkin method

Here we consider a method primarily designed for the treatment of hyperbolic conservation laws, although it is intimately related to more generally applicable schemes to be discussed below. The characteristic Galerkin method, as the name suggests, combines the method of characteristics with a Galerkin approach, offering distinct improvements over the use of interpolation. For an extensive account of the method we refer to [13]; here we give a brief description of some of the main features.

Consider (1.1), with $b = 0$. Since u is constant along the characteristics, they are straight lines. Thus we may write

$$u^{m+1}(y) = u^m(x) \quad \text{where} \quad y = x + a(u^m(x))\Delta t. \tag{1.5}$$

A finite element representation of u^m in terms of the basis functions ϕ_i may be written

$$U^m = \sum_i U_i^m \phi_i, \tag{1.6}$$

and (1.5) is then discretised by

$$\langle U^{m+1}, \phi_i \rangle = \int U^m(x)\phi_i(y)dy, \tag{1.7}$$

where here $\langle \cdot, \cdot \rangle$ represents the L^2 inner product over R^n.

To obtain the characteristic Galerkin method, (1.7) is then rewritten in an equivalent form,

$$\langle U^{m+1} - U^m, \phi_i \rangle + \Delta t \langle \nabla \cdot f(U^m), \Phi_i \rangle = 0, \tag{1.8a}$$

where

$$\Phi_i(x) = \frac{1}{|a(U^m(x))|\Delta t} \int_x^{x+a(U^m(x))\Delta t} \phi_i(z)dz. \tag{1.8b}$$

The scheme (1.8) is unconditionally stable, conservative, and highly accurate (splines of order s give accuracy of order $2s-1$). Moreover, the accuracy can be enhanced through recovery procedures, and (1.8) is still valid when shocks form. However, there is a difficulty in the application of the method to systems of conservation laws, which has inhibited the use of the method.

1.3.2 Lagrange-Galerkin methods

Let us return to (1.1) and (1.2) and rewrite (1.2) in the more explicit form

$$\begin{aligned} \frac{d}{ds}X(x,t;s) &= a(X(x,t;s),s) && \text{(1.9a)} \\ X(x,t;t) &= x. && \text{(1.9b)} \end{aligned}$$

For a function w on R^n we define $[E(t;s)w](x) = w(X(x,t;s))$. Then the basic Lagrange-Galerkin method for approximating (1.1) takes the form

$$\langle U^{m+1}, \phi_i \rangle + \Delta t \langle b^{m+1} \nabla U^{m+1}, \nabla \phi_i \rangle = \langle E(t^{m+1}; t^m) U^m, \phi_i \rangle. \tag{1.10}$$

When a depends on u, the discretised form of (1.9) must usually be solved by some iterative means. Often a forward Euler discretisation of (1.9) will suffice, although the midpoint rule of the semi-Lagrangian method is also an option. The method is, of course, similar to the semi-Lagrangian methods already described, although it incorporates the accuracy and flexibility of the finite element method, and, at least in its pure form, it avoids the disadvantages of the use of interpolation at the feet of the characteristics.

Such methods have been developed in various settings ([21], [8], [9], [36], [49], [60], [75] and [76]), with each group of authors using a different name. The *modified method of characteristics* [21] was developed with a view to applications in petroleum engineering. Generalisations to higher order in time versions were studied in [22]; we shall return to this subject in Chapter 4. The problem of boundary conditions is one that has received scanty treatment in the literature—this issue was addressed by Russell and his co-workers in [71], where they use space-time elements. The *transport-diffusion algorithm* was applied to the Navier-Stokes equations and analysed in [60] (obtaining slightly sub-optimal convergence rates), and the *Lagrange-Galerkin method* for the same equations was analysed in [76] (obtaining optimal convergence rates and demonstrating non-linear stability). The engineering literature has preferred the name *Eulerian-Lagrangian* methods for these and related ideas (e.g. [58]), and this nomenclature was adopted by Russell in [71].

The convergence and stability properties of the method (1.10) are impressive (the scheme is unconditionally stable). However, much of the above mentioned analysis was carried out under the assumption that the inner products involved in evaluating the right hand side are carried out exactly. In practice some form of quadrature is necessary. Unfortunately this causes the unconditional stability to be lost. Instability was first observed by Priestley [61] and the analysis in [57] and [77] showed that for most commonly used quadrature schemes, the method is only conditionally stable, and that for some the stability condition cannot be met by reducing the size of the timestep. The presence of diffusion does have a stabilising influence on the method, however, and Jack [40], [41] has investigated the impact of these theoretical results in practice.

In [57], an alternative technique was presented, using ideas developed for particle-in-cell methods. This is termed area weighting and involves transporting the support of each finite element basis function without rotation according to the path taken by the centroid of the element, before the integrals in (1.10) are carried out. This has the effect of restoring the stability properties of the exactly-integrated scheme without significantly deteriorating the accuracy [57]. Its effectiveness, however, is limited to cartesian meshes, and this led Jack [41] to investigate another alternative, called subdivision. However, in practice, he concluded that this was less effective than the careful use of quadrature. It is worth commenting at this point that the area-weighted Lagrange-Galerkin method is very similar to a particle-in-cell finite element method which Bermejo [10] has shown, on rectangular cartesian meshes, to be equivalent to the semi-Lagrangian method with cubic spline interpolation.

The deterioration of the stability properties of the Lagrange-Galerkin method when quadrature is introduced can be attributed to the lack of smoothness in the basis functions. For Eulerian schemes this is not a problem, but here the basis functions are shifted, and the discontinuity becomes apparent to the integration formula. Such considerations were a major factor in the original proposal of the spectral Lagrange-Galerkin method [78], which will be investigated in detail below.

1.3.3 The weak Lagrange-Galerkin method

We bring this brief survey to a close by describing an alternative finite element method for (1.1), which may be formulated by shifting the test functions instead of the trial functions. Let us integrate (1.1) in space and time against time-dependent test functions. We have

$$\int_{t^m}^{t^{m+1}} \langle \partial_t u + \nabla \cdot f - \nabla \cdot (b\nabla u), \psi_i \rangle dt = 0,$$

which, after some integration by parts, gives

$$\langle u^{m+1}, \psi_i^{m+1} \rangle - \langle u^m, \psi_i^m \rangle + \int_{t^m}^{t^{m+1}} \langle b\nabla u, \nabla \psi_i \rangle dt = \int_{t^m}^{t^{m+1}} [\langle u, \partial_t \psi_i \rangle + \langle f, \nabla \psi_i \rangle] dt. \qquad (1.11)$$

Let us set $c = f/u$. Then if we replace a by c in (1.9) and set $\psi_i(x,t) = \phi_i(X(x, t^{m+1}; t))$, the right hand side of (1.11) vanishes, and on approximating the integral on the left hand side in some way we arrive at the weak Lagrange-Galerkin method of Benqué et. al. [8]. This formulation has the advantage of being conservative in the case $b = 0$, and for this reason is favoured by Priestley [62] in his implementation of the spectral Lagrange-Galerkin method on a meteorological test problem. It is also the formulation of an independent development of spectral Lagrange-Galerkin methods by Ho *et. al.* [37]. They use a quadrature formula on the term $\langle u^m, \psi_i^m \rangle$ which uses points which have been shifted according to (1.9). To evaluate u^m at those points they use an efficient 'regridding' technique which is based on solving a pure advection subproblem for which they use an Adams-Bashforth time integrator. All this is combined with a spectral element method to enable the solution of problems on non-rectangular domains.

1.4 Spectral methods

Having discussed various implementations of Lagrangian approaches for convection-diffusion problems, we now outline by way of introduction some aspects of spectral methods, and mention some work that is of relevance to us, before setting out the programme for the remainder of this Thesis.

Spectral methods have developed into a widely used technique for the solution of differential equations. Their increase in popularity is due in part to the fact that the approximations underlying spectral methods have very high accuracy and that, with the availability of the Fast Fourier Transform (FFT), they may be implemented efficiently. More recently, much work has been put into overcoming the limitation of standard spectral methods to simple domains, and into extending their remarkable effectiveness to problems with non-smooth solutions. Spectral element, and spectral domain-decomposition techniques, are areas of particularly active development in this respect. The standard reference on spectral methods is [12], which contains a unified theory of their mathematical analysis, together with a fairly comprehensive description of applications of spectral methods in fluid dynamics.

1.4.1 Spectral approximation

The expansion of a function in terms of an infinite sequence of orthogonal functions underlies many numerical methods of approximation, including spectral methods. The usefulness of these methods depends upon their accuracy and upon the efficiency of their implementation. With regard to the accuracy of spectral methods, the most familiar results concern the approximation of periodic functions by Fourier series. Here the kth coefficient of the expansion decays faster than any inverse power of k when the function is infinitely smooth. (For holomorphic functions, this decay is actually exponential [80]). A particular characteristic of this decay is that it is not observed until the expansion contains enough terms to effectively resolve the salient features of the function. However, once this point is reached the series will provide a very good approximation with the inclusion of a very few additional terms. This behaviour is referred to as the *spectral accuracy* of the Fourier method. For smooth, non-periodic functions, such spectral accuracy may be attained by expanding in terms of Jacobi polynomials.

1.4.2 The discrete Fourier transform

This expansion in terms of an orthogonal system introduces a linear transformation between u and the sequence of its expansion coefficients $\{\hat{u}(k)\}$. If the system is complete in a suitable Hilbert space, this transform can be inverted. Hence functions can be described both through their values in physical space and through their coefficients in transform space. These coefficients can rarely be calculated exactly, since they are integrals depending on all the values of u in physical space. If a high precision quadrature formula is used, approximate values for a finite number of expansion coefficients may be calculated. Such a process describes a *discrete transform* between the values of u at the nodes of the quadrature formula and the set of discrete coefficients. The quadrature points are chosen so that an approximation defined by such discrete coefficients has the same accuracy properties as the exactly integrated expansion, and it may be used instead in practical computations. If the quadrature formula is chosen appropriately, the approximating function is actually the

interpolant of u at the quadrature nodes, so that this is the suitable transform to use in a collocation method for, say, the evaluation of a derivative at the collocation/quadrature points. For the most common systems (Fourier and Chebyshev polynomials), the discrete transform may be calculated efficiently by using the FFT in $\frac{5}{2}N \log N$ operations, where N is the number of polynomials in the expansion, compared with $2N^2$ operations without the use of the FFT.

1.4.3 Spectral methods for time-dependent problems

Most standard spectral methods for time-dependent problems fall within the framework of the method of lines (m.o.l.) approach, where the partial differential equation is discretised first in space, giving rise to a system of ordinary differential equations which may be solved by the method of choice. Analysis of such schemes thus centres first on the semidiscrete equations formed by the spatial discretisation. The review by Tadmor [81] is interesting for our present purpose since he deals with the periodic, hyperbolic case exclusively. As expected, the stability of the Fourier-Galerkin method exactly mirrors the well-posedness of the original equation. However, once the discrete Fourier transform, and along with it aliasing, is introduced, stability is less straightforward, at least in theory. In fact, the theoretical stability of the method remains an open problem, although in practical computations the introduction of aliasing has not been observed to cause any problems. (For a discussion of this point see also [12]).

There are at present two ways around this difficulty [44]. One path one might take is the skew-symmetrisation of the spatial differentiation operator. Pasciak [59] analyses a problem already in this form. The second way to obtain stability is the introduction of filtering. This is discussed in detail by Kreiss and Oliger in [44], where they propose a delicate filtering procedure that enforces a minimal decay rate in the higher region of the spectrum, and leaves unchanged polynomials which are sufficiently smooth, in the sense that they satisfy an inequality in which some high order Sobolev norm is bounded by the L^2-norm. As in Tadmor [81], stability is more straightforward to demonstrate with the (less accurate) application of the crudest of filters—merely cutting off a fixed proportion of the highest modes. This can be seen as corresponding to the use of quadrature with the number of quadrature points proportional to and strictly greater than the number of modes in the solution. For the spectral Lagrange-Galerkin method, the situation is analogous, although we of necessity consider the fully discrete case, since the time discretisation is carried out *first* (see Chapter 4 for a full discussion of this point). An exactly-integrated Fourier-Galerkin spatial approximation gives rise to a scheme whose stability properties mirror the well-posedness of the original equation, whereas the use of quadrature introduces aliasing errors which must be controlled in some way. This is achieved by ensuring that the number of quadrature points exceeds that of the number of coefficients in the solution. The resulting scheme can also be viewed as a filtered *collocation* method. What is in some ways the central result of this Thesis is that the filtering that is used by Tadmor [81] in his proof is sufficient to guarantee unconditional stability for the spectral Lagrange-Galerkin method. A final point to make in this comparison is that in practical computations the need for such filtering has not been observed.

For time-dependent problems, the ordinary differential equations resulting from spectral discretisations in space are often solved by standard methods, such as a linear multistep, or a Runge-Kutta method. In contrast with the spectral Lagrange-Galerkin method, such calculations are usually constrained by a CFL-type condition, which for Jacobi polynomials takes the form $\Delta t \leq CN^{-2}$ (In the Fourier case this is replaced by the less restrictive $\Delta t \leq CN^{-1}$). This is sometimes heuristically viewed as being a relation between the size of the timestep and the minimum gridsize. However, Gottlieb and Tadmor [30] show that a more rigorous explanation is that the N^2-term corresponds to the growth of the eigenvalues of the Sturm-Liouville problem associated with the Jacobi polynomials.

With this procedure, a high order discretisation in space is coupled with a compara tively low order discretisation in time. For problems in which the solution changes over a time scale which is comparable to the spatial scale, as in many hyperbolic problems and advection-dominated diffusion problems, the time errors will dominate the spatial errors, necessitating the use of very small timesteps. Gottlieb and Turkel [31] are led by this observation to introduce a coupling of the space and time discretisations. They obtain schemes which are unconditionally stable, but still are subject to accuracy restrictions on the size of the timestep.

Tal-Ezer [83] has proposed a spectral method *in time* for hyperbolic equations (and for parabolic equations [84]), also with a view to eliminating the discrepancy between the temporal and spatial discretisations. His method applies to linear problems with known bounds on the spectrum of the spatial operator L and is based on an orthogonal polynomial expansion of the evolution operator generated by L. His method is relatively efficient; in Chapter 3 we shall compare its performance with that of the spectral method of characteristics for a linear hyperbolic problem.

1.5 Summary of contents

We shall begin, in the next chapter, by considering a simple advection equation—a scalar hyperbolic problem. Our purpose here will be to investigate the properties of the numerical treatment of the evolution operator associated with evolving the solution of the hyperbolic equation through one time step. This is the spectral-Lagrange treatment of the convective terms, and it is shown to be unconditionally stable, both in the L^2-norm and in higher order Sobolev norms.

The central result here is that these results continue to hold even with the introduction of quadrature, which amounts to turning the Fourier-Galerkin method into a Fourier pseudospectral method with truncation of higher modes.

The method as it is described in Chapter 2, although it has very desirable stability and convergence properties, is very inefficient to implement, especially in higher dimensions. In Chapter 3 we describe an additional approximation which, if included in the method, serves to maintain the spectral accuracy of the original scheme and yet may be implemented efficiently. After providing stability and convergence results for this we illustrate these results and those of the previous chapter by means of a series of numerical experiments.

Chapter 4 describes the application of the method to convection-diffusion problems. The equation is rewritten in Lagrangian coordinates and is discretised in this form firstly in time and then in space, and backward difference methods up to order 6 are analysed within a general framework provided by Le Roux [46]. Again the theoretical results are illustrated by numerical experiments. We then begin to turn our attention to non-linear problems, and we consider first a discretisation of Burgers equation. Here we break with the pattern of previous investigations and content ourselves with numerical results. They provide a lead in to the final chapter, where we consider the Navier-Stokes equations. A full analysis for the spectral Lagrange-Galerkin method with a backward-Euler timestep is given, followed by some results obtained by a numerical study of the evolution of a perturbed double shear layer. Scope for future investigation is discussed, as well as conclusions drawn from the current work.

1.6 Notations

Here we introduce much of the notation that will be used below, and describe some of the main function spaces in which we will be operating.

Z denotes the set of integers, N is the set of positive integers and N_0 the set of nonnegative integers. R denotes the set of real numbers and C is the set of complex numbers. The integer part of a nonnegative real number σ is denoted by $[\sigma]$, and we further define $[\sigma]^* = \min\{n \in N_0 : n \geq \sigma\}$. For $x = (x^1, \ldots, x^n)$, $y = (y^1, \ldots, y^n) \in R^n$, we define

$$x \cdot y := x^i y^i,$$

where we have made use of the summation convention, and

$$|x|_1 := \sum_{i=1}^{n} |x^i|,$$

$$|x|_\infty := \max_{1 \leq j \leq n} |x^j|.$$

$|x|$ will denote the Euclidean norm of the vector x.

For $s \in N$, let $\Lambda_0^s = \{0,1\}^s$, and $\Lambda^s = \{0,1\}^s \backslash \{(0,\ldots,0)\}$. Then the binomial theorem generalises to

$$\prod_{r=1}^{s} (a_r + b_r) = \sum_{\lambda \in \Lambda_0^s} \prod_{r=1}^{s} a_r^{\lambda_r} b_r^{1-\lambda_r}. \tag{1.12}$$

Suppose that $K \in N_0$ and $\nu = (\nu_1, \ldots, \nu_n)$, $0 \leq \nu_j \leq K$, then

$$\sum_{\nu} := \sum_{\nu_1=0}^{K} \cdots \sum_{\nu_n=0}^{K}.$$

Let $\Omega = [0, 2\pi)^n$. For any $N \in N_0$, we denote by S_N the space of trigonometric polynomials on Ω with degree $\leq N$, i.e.

$$S_N = \left\{ v = \sum_{|k|_\infty \leq N} \hat{v}_k \Phi_k \right\},$$

where

$$\Phi_k(x) := e^{ik \cdot x}.$$

Let $L^2_\#(\Omega)$ be the space of square-integrable functions on Ω, periodically extended to the whole of R^n. Then Π_N will denote the orthogonal projector in $L^2_\#(\Omega)$ onto S_N, and, for u, $v \in L^2_\#(\Omega)$, (u, v) will denote the $L^2_\#(\Omega)$ inner product

$$(2\pi)^{-n} \int_\Omega u(x)\overline{v(x)} dx;$$

we denote the norm on $L^2_\#(\Omega)$ by $\|\cdot\|$. We note that the set $\{\Phi_k\}_{k \in Z^n}$ forms a Hilbert basis for $L^2_\#(\Omega)$, so that any $v \in L^2_\#(\Omega)$ may be written

$$v = \sum_{k \in Z^n} \hat{v}(k) \Phi_k, \tag{1.13}$$

where

$$\hat{v}(k) = (v, \Phi_k), \quad k \in Z^n,$$

are the Fourier coefficients of v.

For 2π-periodic functions f, g we define the convolution by

$$(f * g)(x) = (2\pi)^{-n} \int_\Omega f(x-y) g(y) dy. \tag{1.14}$$

This enjoys all the usual properties of convolution in R^n.

For $k \in N_0 \cup \{\infty\}$, $C^k_\#(\overline{\Omega})$ is the space of functions from Ω to R which are the restrictions to Ω of functions from R^n to R which are k-times continuously differentiable and 2π-periodic in each coordinate direction. (For $k = 0$ we write $C_\#(\overline{\Omega})$). We note that for $v \in C^k_\#(\overline{\Omega})$ the Fourier coefficients $\hat{v}(p)$ of v decay like $|p|^{-k}$.

The norm on $C_\#(\overline{\Omega})$ will be denoted by $|\cdot|_\infty$, and for $m \geq 1$, the norm of $u = (u^1, \ldots, u^m) \in C_\#(\overline{\Omega})^m$ will be given by

$$|u|_\infty = \left(\sum_{i=1}^{m} |u^i|_\infty^2 \right)^{\frac{1}{2}}.$$

For $i = 1, \ldots, n$, let $D_i = \partial / \partial x^i$; the gradient operator $\nabla : C^{k+1}_\#(\overline{\Omega})^m \mapsto (C^k_\#(\overline{\Omega})^m)^n$ may then be written $\nabla = (D_1, \ldots, D_n)$. If $\alpha = (\alpha_1, \ldots, \alpha_k)$, with each $\alpha_i \in \{1, \ldots, n\}$, we write $n(\alpha) = k$ and $D^\alpha = D_{\alpha_1} \ldots D_{\alpha_k}$. (We shall make use of this *non-standard* multiindex notation throughout this Thesis.) Then the norm of $u \in C^k_\#(\overline{\Omega})^m$ is given by

$$|u|_{k,\infty} = \left(\sum_{j=0}^{k} \binom{k}{j} \sum_{n(\alpha)=j} |D^\alpha u|_\infty^2 \right)^{\frac{1}{2}},$$

which may also be written

$$|u|_{k,\infty} = \left(\sum_{j=0}^{k} \binom{k}{j} |\nabla^j u|_\infty^2 \right)^{\frac{1}{2}}.$$

We introduce the dual $\mathcal{D}'_\#(\Omega)$ of $C^\infty_\#(\overline{\Omega})$ (consisting of conjugate-linear functionals), with duality pairing $\langle \cdot, \cdot \rangle$; this is the space of periodic distributions. For further details to the description below, we refer to [52]. For $f \in \mathcal{D}'_\#(\Omega)$ we can define the Fourier coefficients of f by

$$\hat{f}(p) = \langle f, \Phi_p \rangle.$$

f may be expanded in a Fourier series as in (1.13), but with convergence in $\mathcal{D}'_\#(\Omega)$ instead of in $L^2_\#(\Omega)$. An important result in this respect is that $f \in \mathcal{D}'_\#(\Omega)$ if and only if $\exists \sigma > 0 : (1 + |p|^2)^{-\sigma} \hat{f}(p) \to \infty$.

Periodic distributional derivatives are defined by

$$\langle D^{(\alpha)} f, \phi \rangle = (-1)^\alpha \langle f, D^{(\alpha)} \phi \rangle, \quad \forall \phi \in C^\infty_\#(\overline{\Omega}). \tag{1.15}$$

We note that derivatives in the above sense do not in general coincide with those in the sense of $\mathcal{D}'(\Omega)$;when a function is regular but not periodic, the non-periodicity on the boundary of Ω will induce the same behaviour in the (periodic distributional) derivative of that function as would a corresponding discontinuity in the interior of Ω. For $\sigma \in N_0$, we define the Sobolev space $H^\sigma_\#(\Omega)$ by

$$H^\sigma_\#(\Omega) = \{u : D^\alpha u \in L^2_\#(\Omega), 0 \leq \alpha \leq \sigma\},$$

(where the derivatives are in the sense of periodic distributions) with norm

$$\|u\|_\sigma = \left(\sum_{k=0}^{\sigma} \binom{\sigma}{k} \sum_{n(\alpha)=k} \|D^\alpha u\|^2 \right)^{\frac{1}{2}}. \tag{1.16}$$

We also define the semi-norm $|\cdot|_\sigma$ on $H^\sigma_\#(\Omega)$ by

$$|u|_\sigma = \left(\sum_{n(\alpha)=\sigma} \|D^\alpha u\|^2 \right)^{\frac{1}{2}}.$$

Defining the norm on $H^\sigma_\#(\Omega)^m$ equivalently to that on $C^k_\#(\overline{\Omega})^m$, we may rewrite

$$|u|_\sigma = \|\nabla^\sigma u\|,$$

and $\|u\|_\sigma$ may be rewritten

$$\|u\|_\sigma = \left(\sum_p (1+|p|^2)^\sigma |\hat{u}(p)|^2\right)^{\frac{1}{2}}. \tag{1.17}$$

This allows us to extend the definition of $H^\sigma_\#(\Omega)$ to non-integer σ:

$$H^\sigma_\#(\Omega) = \{u \in \mathcal{D}'_\#(\Omega) : \|u\|_\sigma < \infty\}. \tag{1.18}$$

For $\sigma < 0$, $H^\sigma_\#(\Omega)$ is defined to be the dual space of $H^{-\sigma}_\#(\Omega)$, and may also be characterised by (1.18) and (1.17).

Let $T > 0$ and $0 = t_0 < t_1 < \ldots < t_M = T$. For a Banach space X,

$$l^\infty(X) := \left\{ v : \{t_0, \ldots, t_M\} \to X | \ \|v\|_{l^\infty(X)} = \max_{0 \le m \le M} \|v(t_m)\|_X < \infty \right\}.$$

Further, $C^k(X)$ denotes the space of k-times continuously differentiable mappings from $[0,T]$ into X; for $1 \le p \le \infty$, $L^p(X)$ will often be used as shorthand for $L^p(0,T;X)$, and $W^{k,p}(X)$ for $W^{k,p}(0,T;X)$, for $k \ge 0$. For any positive integer m, when there is no risk of confusion with the norm on $C^k_\#(\overline{\Omega})^m$, the norm on $L^p(C^k_\#(\overline{\Omega}))^m$ will be denoted $|\cdot|_{k,p}$. For $s,t \in [0,T]$ the norm on $L^p(s,t;C^k_\#(\overline{\Omega}))^m$ will be denoted $|\cdot|_{k,p;[s,t]}$. For any Banach spaces Y and Z, $L(Y;Z)$ denotes the Banach space of continuous linear operators from Y to Z equipped with the induced operator norm. The letter C will denote a generic positive constant, and C$_0$, C$_1$, etc. will stand for specific positive constants. In most places, the nature of the dependence of these constants will be made explicit.

Chapter 2

A model hyperbolic problem

As indicated in the previous chapter, the Lagrange-Galerkin method is designed for problems in which convection is the dominant term in the governing equation. It falls into a particular class of schemes, each of which is based on a particular strategy for dealing with the convective term. In this chapter we focus attention on the strategy that is characteristic of the spectral Lagrange-Galerkin method, and so we consider the application of the scheme to a linear scalar hyperbolic problem.

We begin by posing the problem in a weak form. We construct solutions by means of the method of characteristics, and discuss their regularity properties. This involves results concerning the well-posedness of the problem in various norms, and leads naturally to the formation of a numerical method which has analogous stability properties, and which forms the basis of the spectral Lagrange-Galerkin method. In subsequent sections we consider the effects of further approximations that must be made in order to create a scheme that may be implemented efficiently. These are: the use of quadrature in the evaluation of the integrals that form an essential part of the method, and the introduction of a local interpolation (in order to improve efficiency). In both cases we show that the original stability and accuracy properties can be maintained.

2.1 Discussion of the model problem; the evolution operator E

For $u_0 \in L^2_\#(\Omega)$ and $a \in L^1(C^1_\#(\overline{\Omega})^n)$, consider the following initial boundary value problem: find u in $C(L^2_\#(\Omega)) \cap W^{1,1}(H^{-1}_\#(\Omega))$ such that

$$\begin{aligned} \frac{\partial u}{\partial t} + a \cdot \nabla u &= 0 \quad \text{in } \Omega \times (0,T], \qquad (2.1\text{a}) \\ u(\cdot, 0) &= u_0 \quad \text{in } \Omega. \qquad (2.1\text{b}) \end{aligned}$$

We shall establish the existence and uniqueness of solutions to (2.1), and discuss their regularity, using the method of characteristics as our main tool. This approach yields results which are central to much of the analysis in this Thesis. We want to construct curves along which solutions of (2.1) remain constant. With this end in view we proceed as follows: for $x \in R^n$ and $t_0 \in [0,T]$, we define the mapping $X(x,t_0;\cdot) : [0,T] \to R^n$ as the solution of the initial value problem

$$\begin{aligned} \frac{d}{dt} X(x,t_0;t) &= a(X(x,t_0;t),t), \quad a.e. \quad t \in [0,T] \backslash \{t_0\}, \qquad (2.2\text{a}) \\ X(x,t_0;t_0) &= x, \qquad (2.2\text{b}) \end{aligned}$$

where we have periodically extended a to the whole of R^n. As we shall show below, the solutions of (2.2) are continuous curves in space-time passing through the points (x,t_0). Given sufficient smoothness of u and a, we can apply the chain rule and obtain

$$\begin{aligned} \frac{d}{dt} u(X(x,t_0;t),t) &= u_t(X(x,t_0;t),t) + \frac{dX}{dt}(x,t_0;t) \cdot \nabla u(X(x,t_0;t),t) \\ &= [u_t + a \cdot \nabla u](X(x,t_0;t),t). \end{aligned}$$

Thus smooth solutions of (2.1) will satisfy, for each $x \in R^n$,

$$u(X(x,t_0;t),t) = \text{ constant},$$

and then (2.2) defines the curves we seek.

For each $x \in R^n$ and $t_0 \in [0,T]$, (2.2) is mathematically equivalent to the following problem:

$$\begin{aligned} \frac{d}{dt}Y(x,t_0;t) &= a(x+Y(x,t_0;t),t), \quad a.e. \quad t \in [0,T]\backslash\{t_0\}, && (2.3\text{a}) \\ Y(x,t_0;t_0) &= 0, && (2.3\text{b}) \end{aligned}$$

where $Y(x,t_0;t) = X(x,t_0;t) - x$, and we note that solutions to (2.3) will be 2π-periodic in their spatial variables.

Theorem 2.1 *Let $a \in L^1(C^1_\#(\overline{\Omega})^n)$, $x \in R^n$ and $t_0 \in [0,T]$. Then (2.3) has a unique solution $Y(x,t_0;\cdot) \in W^{1,1}(0,T)$. Moreover, if $a \in L^1(C^r_\#(\overline{\Omega})^n)$ for integer $r \geq 1$, then $Y(x,t_0;t)$ and $\partial Y/\partial t_0(x,t_0;t)$ are continuous in each of their variables, and for each $t_0 \in [0,T]$, the map $\phi : (x,t) \mapsto Y(x,t_0;t)$ satisfies $\phi \in W^{1,1}(C^r_\#(\overline{\Omega})^n)$ and the map $\psi : (x,t) \mapsto \partial Y/\partial t_0(x,t_0;t)$ satisfies $\psi \in W^{1,1}(C^{r-1}_\#(\overline{\Omega})^n))$. Setting*

$$X(x,t_0;t) = x + Y(x,t_0;t), \quad x \in R^n, \ \ t_0,t \in [0,T], \tag{2.4}$$

then provides the unique solution to (2.2), with corresponding smoothness properties.

Remark. This is only a slight generalisation on standard results for such systems of ordinary differential equations (as found in [35], for example). We give a sketch of the proof.

Proof Integrating (2.3) formally, we obtain

$$Y(x,t_0;t) = \int_{t_0}^t a(x+Y(x,t_0;\tau),\tau)d\tau. \tag{2.5}$$

Defining, iteratively,

$$\begin{aligned} Y_0(x,t_0;t) &= 0 \\ Y_{n+1}(x,t_0;t) &= \int_{t_0}^t a(x+Y_n(x,t_0;\tau),\tau)d\tau, \end{aligned}$$

we obtain a sequence converging uniformly for t in $B_{t_0} = \{t \in [0,T] : |\nabla a|_{0,1;[t_0,t]} < 1\}$. Since each $Y_n(x,t_0;t)$ is continuous for $t \in B_{t_0}$, this holds true for the limit $Y(x,t_0;t)$. It also follows readily that $Y(x,t_0;\cdot)$ solves (2.3) on B_{t_0} and is unique there. To see how to extend this construction of the solution to the whole of $[0,T]$, we take some $t_1 \in B_{t_0}$ $(t_1 \geq t_0)$ such that $|\nabla a|_{0,1;[t_0,t_1]} \geq \frac{1}{2}$. We know that such a point exists, except perhaps if $T \in B_{t_0}$, because of the continuity of the map $t \mapsto |\nabla a|_{0,1;[t_0,t]}$. Now we let B_{t_1} be defined analogously to B_{t_0} and define for $t \in B_{t_1}$ the iteration

$$\begin{aligned} Y'_0(x,t_0;t) &= Y(x,t_0;t_1) \\ Y'_{n+1}(x,t_0;t) &= Y(x,t_0;t_1) + \int_{t_1}^t a(x+Y'_n(x,t_0;\tau),\tau)d\tau. \end{aligned}$$

The limit $Y'(x,t_0;t)$ solves (2.5) on B_{t_1}, by uniqueness, and agrees with $Y(x,t_0;t)$ on $B_{t_0} \cap B_{t_1}$. Continuing in this way we can extend the solution on to a sequence of intervals increasing to the right. It is impossible that these intervals should remain within $[0,T]$ indefinitely, since the way we constructed the points $t_0, t_1, \ldots$ would then imply that $|\nabla a|_{0,1;[0,T]}$ is unbounded, contradicting $a \in L^1(C^1_\#(\overline{\Omega})^n)$. Thus we can extend the solution to $[t_0,T]$, and by changing directions in the above argument, to the whole of $[0,T]$. We obtain a unique continuous map $t \in [0,t] \mapsto Y(x,t_0;t)$; since $a \in L^1(C^1_\#(\overline{\Omega})^n)$ this map is differentiable almost everywhere and satisfies (2.3). Moreover, $Y(x,t_0;\cdot) \in W^{1,1}(0,T)$ as required, and continuity in both x and t_0 is also assured.

The first step towards obtaining the further regularity results is to differentiate formally (2.3). Making use of the summation convention, we obtain

$$\begin{aligned} \frac{\partial}{\partial t}D_iY_j(x,t_0;t) &= (\delta_{ik} + D_iY_k(x,t_0;t))D_ka_j(x+Y(x,t_0;t),t), && (2.6\text{a}) \\ D_iY_j(x,t_0;t_0) &= 0, && (2.6\text{b}) \end{aligned}$$

and

$$\begin{aligned} \frac{\partial}{\partial t}\frac{\partial Y_i}{\partial t_0}(x,t_0;t) &= \frac{\partial Y_j}{\partial t_0}(x,t_0;t)D_ja_i(x+Y(x,t_0;t),t), && (2.7\text{a}) \\ \frac{\partial Y_i}{\partial t_0}(x,t_0;t_0) &= -a_i(x,t_0), && (2.7\text{b}) \end{aligned}$$

with (2.6a) and (2.7a) holding a.e. $t \in [0,T]\backslash\{t_0\}$.

Analogous arguments to those applied to (2.3) may be applied to (2.6) and to (2.7), and the conclusions of the theorem for $r = 1$ follow. For higher values of r an inductive argument is used.

Corollary 2.1 *For $x \in R^n$ and $t_0, t, \tau \in [0,T]$,*

$$X(x,t_0;t) = X(X(x,t_0;\tau),\tau;t). \tag{2.8}$$

Proof Since both the left and the right hand sides of (2.8) (regarded as functions of t) are solutions to (2.2), they satisfy

$$\begin{aligned} \frac{d}{dt}\chi(t) &= a(\chi(t),t), \quad a.e. \quad t \in [0,T], \\ \chi(\tau) &= X(x,t_0;\tau), \end{aligned}$$

the result follows since such solutions are unique.

We now define the evolution operator E.

Definition 2.1 *Let w be a function from a subset A of R^n to R. For $t_0, t \in [0,T]$ and $x, X(x,t_0;t) \in A$ define*

$$E(t_0;t)w(x) = w(X(x,t_0;t)). \tag{2.9}$$

Motivated by the discussion following (2.2), our solution to (2.1) will be defined by

$$u(x,t) = u_0(X(x,t;0)) = E(t;0)u_0(x). \tag{2.10}$$

In order to ascertain whether (2.10) will define a solution to (2.1) under the weaker assumptions we have made on u_0, and to be able to discuss its regularity properties, we have to answer some questions about E.

The first question is whether $E(t_0;t) : L^2_\#(\Omega) \to L^2_\#(\Omega)$. This is answered in the following lemma.

Lemma 2.1 *Let $t_0, t \in [0,T]$; then $E(t_0;t)$ is a quasi-isometry of $L^2_\#(\Omega)$ onto itself, with*

$$\|E(t_0;t)\|_{L(L^2_\#(\Omega))} \leq e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t_0,t]}}. \tag{2.11}$$

Proof Define $\phi : R^n \to R^n$ by $\phi(x) = X(x,t_0;t)$, $x \in R^n$. By virtue of Liouville's formula applied to (2.6) (see [35]), the Jacobian $J(x,t_0;t)$ of the mapping $x \to X(x,t_0;t)$ is given by

$$J(x,t_0;t) = e^{\int_{t_0}^t (\nabla\cdot a)(X(x,t_0;\tau),\tau)d\tau}.$$

Thus

$$\int_\Omega |v(\phi(x))|^2 dx = \int_{\phi(\Omega)} |v(y)|^2 J(y,t;t_0)dy,$$

where, (making use of (2.8)), we have replaced x by $y = \phi(x) = X(x,t_0;t)$ (so that $x = X(y,t;t_0)$). We want to replace the integral over $\phi(\Omega)$ by one over Ω. For $j \in Z^n$ define the set Ω_j by

$$\Omega_j := [0,2\pi)^n + 2\pi j.$$

Then, since $\phi(x) - x$ is 2π-periodic,

$$\phi(\Omega_k) \cap \Omega_l = (\phi(\Omega_{k-l}) \cap \Omega_0) + 2\pi l \quad \forall k,l \in Z^n.$$

The sets Ω_j (and $\phi(\Omega_j)$) are mutually disjoint and cover R^n, and so it follows that, if f is a locally integrable 2π-periodic function on R^n,

$$\begin{aligned} \int_{\phi(\Omega_0)} f(x)dx &= \sum_{l\in Z^n} \int_{\phi(\Omega_0)\cap\Omega_l} f(x)dx \\ &= \sum_{l\in Z^n} \int_{(\phi(\Omega_{-l})\cap\Omega_0)+2\pi l} f(x)dx \\ &= \sum_{l\in Z^n} \int_{\phi(\Omega_{-l})\cap\Omega_0} f(x)dx \\ &= \int_{\Omega_0} f(x)dx. \end{aligned}$$

Thus

$$\int_{\phi(\Omega)} |v(y)|^2 J(y,t;t_0)dy = \int_\Omega |v(y)|^2 J(y,t;t_0)dy,$$

and the result follows.

The next question concerns the action of E on the spaces $H^\sigma_\#(\Omega)$, $\sigma \geq 0$, and is addressed by the following lemma.

Lemma 2.2 *Let $\sigma \geq 0$, and suppose that $a \in L^1(C_\#^{[\sigma]^*}(\overline{\Omega})^n)$. Let $t_0, t \in [0,T]$; then $E(t_0;t)$ is a quasi-isometry of $H^\sigma_\#(\Omega)$ onto itself, with*

$$\|E(t_0;t)\|_{L(H^\sigma_\#(\Omega))} \leq e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t_0,t]}+C_1|a|_{[\sigma]^*,1;[t_0,t]}}, \tag{2.12}$$

where $C_1 = C_1(n,\sigma,|a|_{[\sigma]^*,1;[t_0,t]})$, and if $\sigma = 0$, $C_1 = 0$.

In addition, for integer σ, $E(t_0;t)$ is also a quasi-isometry of $C^\sigma_\#(\overline{\Omega})$ to itself, with

$$\|E(t_0;t)\|_{L(C^\sigma_\#(\overline{\Omega}))} \leq e^{C_1|a|_{[\sigma]^*,1;[t_0,t]}}. \tag{2.13}$$

Finally, an estimate for the size of C_1 may be provided. If we define, inductively,

$$c_0 = 1; \qquad c_k = e^{\mu_k \sum_{j=0}^{k-1} c_j |a|_{j+1,1;[t_0,t]}}, \tag{2.14}$$

where $\mu_k = (\sum_{j=1}^k \binom{k}{j})^{\frac{1}{2}}$, then C_1 is such that

$$e^{C_1|a|_{[\sigma]^*,1;[t_0,t]}} \leq c_{[\sigma]^*}.$$

Proof We prove the theorem first for σ integer. For non-integer σ, the result may be deduced by interpolation. We use induction on σ, and so we suppose that the statement of the theorem holds when σ is replaced by $0,\ldots,\sigma-1$, and deduce that it holds for σ.

Since $C^\infty_\#(\overline{\Omega})$ is dense in $H^\sigma_\#(\Omega)$, we first consider the action of $E(t_0;t)$ on $u \in C^\infty_\#(\overline{\Omega})$. We denote $E(t_0;t)$ by $E(t)$. We shall also denote $X(\cdot,t_0;t)$ by $X(t)$.

For $v = \sum_p \hat{v}(p)\phi_p \in H^\sigma_\#(\Omega)$,

$$\begin{aligned} \|v\|_\sigma &= \left(\sum_p (1+|p|^2)^\sigma |\hat{v}(p)|^2\right)^{\frac{1}{2}} \\ &= \left(\sum_{s=0}^\sigma \binom{\sigma}{s} \sum_p |p|^{2s}|\hat{v}(p)|^2\right)^{\frac{1}{2}} \\ &= \left(\sum_{s=0}^\sigma \binom{\sigma}{s} |v|_s^2\right)^{\frac{1}{2}}. \end{aligned}$$

Now, if $\alpha = (\alpha_1,\ldots,\alpha_s) \in \{1,\ldots,n\}^s$, and we define $D^\alpha = D_{\alpha_1}\ldots D_{\alpha_s}$, we can write

$$|v|_s = \left(\sum_{n(\alpha)=s} \|D^\alpha v\|^2\right)^{\frac{1}{2}}.$$

Applying D^α to $E(t)u$, we obtain

$$\begin{aligned} D^\alpha E(t)u &= D_{\alpha_1}\ldots D_{\alpha_s}E(t)u \\ &= D_{\alpha_1}\ldots D_{\alpha_{s-1}}E(t)D_{\alpha_s}u + D_{\alpha_1}\ldots D_{\alpha_{s-1}}[D_{\alpha_s},E(t)]u \\ &= D^{\alpha'}E(t)D_{\alpha_s}u + D^{\alpha'}[D_{\alpha_s},E(t)]u, \end{aligned} \tag{2.15}$$

where $D^{\alpha'} = D_{\alpha_1}\ldots D_{\alpha_{s-1}}$. Now, for $j = 1,\ldots,n$,

$$\begin{aligned} [D_j,E(t)]u &= D_jX_k(t)E(t)D_ku - E(t)D_ju \\ &= (D_jX_k(t)-\delta_{jk})E(t)D_ku \\ &= \xi_{jk}(t)E(t)D_ku, \end{aligned} \tag{2.16}$$

with $\xi_{jk}(t) = D_jX_k(t) - \delta_{jk}$, $j,k = 1,\ldots,n$. (Here and elsewhere in this proof we use the summation convention.)

The application of $D^{\alpha'}$ to the right hand side of (2.16) will necessitate the use of Leibnitz' rule. We define, for $\lambda \in \Lambda_0^{s-1}$, $D^{\alpha',\lambda} = D^{\lambda_1}_{\alpha_1}\ldots D^{\lambda_{s-1}}_{\alpha_{s-1}}$. Then, combining (2.15) and (2.16), we have

$$\begin{aligned} D^\alpha E(t)u &= D^{\alpha'}E(t)D_{\alpha_s}u + D^{\alpha'}\xi_{\alpha_s k}(t)E(t)D_ku \\ &= D^{\alpha'}E(t)D_{\alpha_s}u + \sum_{\lambda\in\Lambda_0^{s-1}} \left(D^{\alpha',\lambda}\xi_{\alpha_s k}(t)\right)\left(D^{\alpha',1-\lambda}E(t)D_ku\right). \end{aligned} \tag{2.17}$$

By density, (2.17) holds, in the sense of distributions, for $u \in H^\sigma_\#(\Omega)$, and henceforth we only assume that u comes from $H^\sigma_\#(\Omega)$.

We multiply both sides of (2.17) by $D^\alpha E(t)u$, integrate over Ω, sum over all α such that $n(\alpha) = s$, and obtain, by the Cauchy-Schwarz inequality,

$$\begin{aligned} |E(t)u|_s &\leq |E(t)\nabla u|_{s-1} + \sum_{\lambda\in\Lambda_0^{s-1}} \left(\sum_{j,k=1}^n |\xi_{jk}(t)|^2_{|\lambda|_1,\infty}\right)^{\frac{1}{2}} |E(t)\nabla u|_{s-1-|\lambda|_1} \\ &= |E(t)\nabla u|_{s-1} + \sum_{r=0}^{s-1}\binom{s-1}{r}\xi_r(t)|E(t)\nabla u|_{s-1-r}, \end{aligned} \tag{2.18}$$

where $\xi_r(t) = \left(\sum_{j,k=1}^n |\xi_{jk}(t)|^2_{r,\infty}\right)^{\frac{1}{2}}$. Again using the Cauchy-Schwarz inequality, and writing $\mathcal{E}_s(t) = \left(\sum_{r=0}^s \binom{s}{r}\xi_r^2(t)\right)^{\frac{1}{2}}$, we find

$$|E(t)u|_s \leq |E(t)\nabla u|_{s-1} + \mathcal{E}_{s-1}(t)\,\|E(t)\nabla u\|_{s-1}. \tag{2.19}$$

But then, repeating the above calculation, and writing $|v|_s = \|\nabla^s v\|$,

$$|E(t)u|_s \leq \|E(t)\nabla^s u\| + \sum_{r=0}^{s-1}\mathcal{E}_r(t)\,\left\|E(t)\nabla^{s-r}u\right\|_r. \tag{2.20}$$

If we denote $e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t_0,t]}}$ by c, we can bound $\|E(t)\nabla^{s-r}u\|_r$ (for $r \leq s-1$, $s \leq \sigma$) by $cc_r\,\|u\|_s$, and this in turn by $cc_r\,\|u\|_\sigma$. Thus, multiplying (2.20) by $\binom{\sigma}{s}|E(t)u|_s$ and summing, we obtain

$$\begin{aligned} \|E(t)u\|_\sigma &\leq c\,\|u\|_\sigma + \|u\|_\sigma \left(\sum_{s=1}^\sigma \left(\sum_{r=0}^{s-1}\mathcal{E}_r(t)cc_r\right)^2 \binom{\sigma}{s}\right)^{\frac{1}{2}} \\ &\leq c\,\|u\|_\sigma \left[1 + \mu_\sigma \sum_{r=0}^{\sigma-1}\mathcal{E}_r(t)c_r\right]. \end{aligned} \tag{2.21}$$

It remains to bound $\mathcal{E}_r(t)$, $0 \leq r \leq \sigma-1$. Integrating (2.6) from t_0 to t, we find

$$\xi_{ij}(t) = \int_{t_0}^t [\xi_{ik}(\tau)E(\tau)D_ka_j(\tau) + E(\tau)D_ia_j(\tau)]d\tau. \tag{2.22}$$

Differentiating (2.22),

$$\begin{aligned} D^\alpha\xi_{ij}(t) &= \sum_{\lambda\in\Lambda_0^s}\int_{t_0}^t \left(D^{\alpha,\lambda}\xi_{ik}(\tau)\right)\left(D^{\alpha,1-\lambda}E(\tau)D_ka_j(\tau)\right)d\tau \\ &\quad + \int_{t_0}^t D^\alpha E(\tau)D_ia_j(\tau)d\tau. \end{aligned} \tag{2.23}$$

Again, we may multiply both sides of (2.23) by $D^\alpha\xi_{ij}(t)$, take the maximum norm in space, and sum over $|\alpha|_1 = s$, $i,j = 1,\ldots,n$, to obtain

$$\xi_s(t) \leq \sum_{\lambda\in\Lambda_0^s}\int_{t_0}^t \xi_{|\lambda|_1}(\tau)|E(\tau)\nabla a(\tau)|_{s-|\lambda|_1,\infty}d\tau$$

$$
\begin{aligned}
&\quad + \int_{t_0}^{t} |E(\tau)\nabla a(\tau)|_{s,\infty} d\tau \\
&= \sum_{r=0}^{s} \binom{s}{r} \int_{t_0}^{t} \xi_r(\tau)|E(\tau)\nabla a(\tau)|_{s-r,\infty} d\tau \\
&\quad + \int_{t_0}^{t} |E(\tau)\nabla a(\tau)|_{s,\infty} d\tau \\
&\leq \int_{t_0}^{t} \mathcal{E}_s(\tau)A_s(\tau)d\tau + \int_{t_0}^{t} |E(\tau)\nabla a(\tau)|_{s,\infty} d\tau, \qquad (2.24)
\end{aligned}
$$

where $A_s(t) = \left(\sum_{r=0}^{s} \binom{s}{r} |E(t)\nabla a(t)|_{r,\infty}^2\right)^{\frac{1}{2}}$. Now (2.24) holds for $0 \leq s \leq \sigma - 1$. For these values of s we deduce

$$
\begin{aligned}
\mathcal{E}_s(t) &\leq \int_{t_0}^{t} \left(\sum_{r=0}^{s} \binom{s}{r} \mathcal{E}_r(\tau)^2 A_r(\tau)^2 \right)^{\frac{1}{2}} d\tau + \int_{t_0}^{t} A_s(\tau)d\tau \\
&\leq \mu_\sigma \int_{t_0}^{t} \mathcal{E}_s(\tau)A_s(\tau) + \int_{t_0}^{t} A_s(\tau)d\tau. \qquad (2.25)
\end{aligned}
$$

Gronwall's Lemma now gives

$$
\begin{aligned}
\mathcal{E}_s(t) &\leq \int_{t_0}^{t} A_s(\tau)e^{\mu_\sigma \int_{t_0}^{\tau} A_s(\tau')d\tau'} d\tau \\
&= \frac{1}{\mu_\sigma}\left[e^{\mu_\sigma \int_{t_0}^{t} A_s(\tau)d\tau} - 1 \right]. \qquad (2.26)
\end{aligned}
$$

But, by the inductive hypothesis,

$$
\int_{t_0}^{t} A_s(\tau)d\tau \leq c_s|a|_{s+1,1;[t_0,t]}. \qquad (2.27)
$$

Substituting (2.27) into (2.21), we obtain

$$
\begin{aligned}
\|E(t)u\|_\sigma &\leq c\,\|u\|_\sigma \left[1 + \sum_{r=0}^{\sigma-1} c_r \left[e^{\mu_\sigma c_r |a|_{r+1,1;[t_0,t]}} - 1 \right] \right] \\
&= c\,\|u\|_\sigma\, e^{\mu_\sigma \sum_{r=0}^{\sigma-1} c_r|a|_{r+1,1;[t_0,t]}} \\
&= cc_\sigma \|u\|_\sigma .
\end{aligned}
$$

The establishment of the bound on $\|E(t_0;t)\|_{L(C^\sigma_\#(\overline{\Omega}))}$ follows an analogous argument to that used in the first part of this proof.

We are now in a position to give the following existence and uniqueness theorem for (2.1).

Theorem 2.2 *Let* $u_0 \in L^2_\#(\Omega)$ *and* $a \in L^1(C^1_\#(\overline{\Omega})^n)$. *Then there is a unique solution* u *of (2.1) satisfying* $u \in C(L^2_\#(\Omega)) \cap W^{1,1}(H^{-1}_\#(\Omega))$, *and given by (2.10). Moreover, if* $u_0 \in H^\sigma_\#(\Omega)$, $\sigma \geq 0$, *and* $a \in L^1(C^{[\sigma]^*}_\#(\overline{\Omega})^n)$, *then* $u \in C(H^\sigma_\#(\Omega)) \cap W^{1,1}(H^{\sigma-1}_\#(\Omega))$, *and satisfies*

$$
\|u(\cdot,t)\|_\sigma \leq e^{C_2(0,t)} \|u_0\|_\sigma , \quad t \in [0,T], \qquad (2.28)
$$

where $C_2 = C_2(0,t) = \frac{1}{2}|\nabla \cdot a|_{0,1;[0,t]} + C_1|a|_{[\sigma]^*,1;[0,t]}$, *and* C_1 *is as in Lemma 2.2.*

Proof Let u be defined by (2.10). Then by Lemma 2.1, $u \in C(L^2_\#(\Omega))$. By density, we may take a sequence of functions $u_0^{(j)}$ in $C^\infty_\#(\overline{\Omega})$ converging to u_0 in $L^2_\#(\Omega)$. If, analogously to the definition of u in terms of u_0 by (2.10), we define $u^{(j)}$ by

$$
u^{(j)}(x,t) = u_0^{(j)}(X(x,t;0)) = E(t;0)u_0^{(j)}(x), \qquad (2.29)
$$

then we find that $u^{(j)} \to u$ in $C(L^2_\#(\Omega))$ by Lemma 2.1. By Theorem 2.2, $\frac{\partial X}{\partial t}(x,t;0)$ is continuous in x and t, and so we may differentiate (2.29) to obtain

$$
u_t^{(j)}(x,t) = \frac{\partial X_k}{\partial t}(x,t;0)E(t;0)D_k u_0^{(j)}(x). \qquad (2.30)
$$

Now, rewriting (2.6) and (2.7) in terms of X instead of Y and comparing them, we find that

$$
\frac{\partial X_k}{\partial t}(x,t;0) = -a_i(x,t)D_iX_k(x;t;0),
$$

and so we obtain

$$
\begin{aligned}
u_t^{(j)}(x,t) &= -a_i(x,t)D_iX_k(x;t;0)E(t;0)D_k u_0^{(j)}(x) \\
&= -a_i(x,t)D_iE(t;0)u_0^{(j)}(x). \qquad (2.31)
\end{aligned}
$$

Letting $j \to \infty$, we see that, in $L^1(H^{-1}_\#(\Omega))$,

$$
u_t = -a \cdot \nabla u, \qquad (2.32)
$$

and so $u \in W^{1,1}(H^{-1}_\#(\Omega))$, and is a solution of (2.1).

To show uniqueness we prove that any solution of (2.1) must satisfy (2.10). Let ρ_δ be a periodic Friedrichs' mollifier, constructed as follows. Let ρ be a real-valued function in $C^\infty(\Omega')$ (where $\Omega' = (-\pi,\pi)^n$) satisfying

$$
\rho \geq 0; \quad supp\rho \subset \{x : |x|_\infty \leq 1\}; \quad \int_{\Omega'} \rho dx = 1.
$$

For example,

$$
\rho(x) = \begin{cases} Ce^{-\frac{1}{(1-|x|^2)}} & |x| < 1; \\ 0 & otherwise. \end{cases}
$$

Now define ρ_δ by

$$
\rho_\delta(x) = \left(\frac{1}{\delta}\right)^n \rho\left(\frac{x}{\delta}\right),
$$

and periodically extend ρ_δ to the whole of R^n. Then $\rho_\delta \in C^\infty_\#(\overline{\Omega})$ and satisfies

$$
\rho_\delta \geq 0; \quad supp\rho_\delta \cap \Omega \subset \{x : \max_i |x^i|_{mod 2\pi} \leq \delta\}; \quad \int_\Omega \rho_\delta dx = 1.
$$

Suppose that $v \in L^2_\#(\Omega)$, and write $v_\delta = \rho_\delta * v$. Then $v_\delta \in C^\infty_\#(\overline{\Omega})$, and $v_\delta \to v$ in $L^2_\#(\Omega)$ as $\delta \to 0^+$. For

$$
\begin{aligned}
|v_\delta(x) - v(x)|^2 &= \left(\int_\Omega \rho_\delta(y)(v(x-y) - v(x))dy \right)^2 \\
&\leq \int_{supp\rho_\delta} |v(x-y) - v(x)|^2 \rho_\delta(y)dy,
\end{aligned}
$$

so that

$$
\begin{aligned}
\|v_\delta - v\|^2 &\leq \int_\Omega \int_{supp\rho_\delta} |v(x-y) - v(x)|^2 \rho_\delta(y)dydx \\
&\leq \sup_{y \in supp\rho_\delta} \int_\Omega |v(x-y) - v(x)|^2 dx,
\end{aligned}
$$

and this tends to zero as $\delta \to 0^+$ by Lebesgue's Lemma (see [55]).

Now let u be a solution of (2.1), and let $u_\delta = \rho_\delta * u$. Then

$$
\begin{aligned}
E(0;t)u_\delta(x,t) - u_{\delta,0}(x) &= \int_0^t \frac{d}{ds}E(0;s)u_\delta(x,s)ds \\
&= \int_0^t E(0;s)[\frac{\partial}{\partial t}u_\delta + a \cdot \nabla u_\delta](x,s)ds. \qquad (2.33)
\end{aligned}
$$

Now, the left hand side of (2.33) converges in $C(L^2_\#(\Omega))$ to

$$
E(0;t)u(x,t) - u_0(x).
$$

For the right hand side we have

$$
\frac{\partial}{\partial t}u_\delta + a \cdot \nabla u_\delta = \phi_\delta * (u_t + a \cdot \nabla u) + [a \cdot \nabla, \phi_\delta *]u.
$$

Since u satisfies (2.1) a.e. $t \in [0,T]$, the first term on the right hand side is zero. The second term converges to zero in $L^2_\#(\Omega)$ by Lemma 2.3 below, and so u satisfies (2.10).

To complete the proof, we notice that, because of Lemma 2.2, if $u_0 \in H^\sigma_\#(\Omega)$ then $u \in H^\sigma_\#(\Omega)$. The remaining assertions of the theorem now follow from (2.1) and from Lemma 2.2.

We quote the following lemma, which was used in the proof of Theorem 2.2. For the proof in the non-periodic case we refer to [55].

Lemma 2.3 (Friedrichs) *Let $b \in C^1_\#(\overline{\Omega})$, $u \in L^2_\#(\Omega)$. For a commutator C_δ which is defined as*

$$\begin{aligned} C_\delta u &= [a \cdot \nabla, \phi_\delta *]u \\ &= a \cdot \phi_\delta * \nabla u - \phi_\delta * a \cdot \nabla u, \end{aligned}$$

we have:
1. *$\|C_\delta u\| \le C\,\|u\|$, where C depends only on ϕ_δ;*
2. *$C_\delta u \to 0$ in $L^2_\#(\Omega)$ as $\delta \to 0^+$.*

We have completed our discussion of the model problem, and are now ready to describe the spectral Lagrange-Galerkin method as it applies the problem we have been considering, and to discuss how well the approximations it produces conform to the exact solutions of the problem.

2.2 Construction of the numerical method

As will become a familiar procedure in later chapters, we construct our solution by first discretising *in time* (although we note that in this instance our time discretisation involves no approximation). For $m = 0, \ldots, M$, let $t^m := m\Delta t$, where $\Delta t = T/M$. Then, because of (2.10), the solution u of (2.1) satisfies, for each $\phi \in L^2_\#(\Omega)$

$$\begin{aligned} (u(\cdot,t^m),\phi) &= (E(t^m;t^{m-1})u(\cdot,t^{m-1}),\phi), \quad m = 1,\ldots,M && (2.34a)\\ (u(\cdot,0),\phi) &= (u_0,\phi). && (2.34b) \end{aligned}$$

This leads us to introduce the following numerical method for approximating (2.1): find $U = \{U^m\}_{m=1}^M \subset S_N$ such that

$$\begin{aligned} (U^m,\phi) &= (E(t^m;t^{m-1})U^{m-1},\phi), \quad \forall \phi \in S_N,\ m = 1,\ldots,M && (2.35a)\\ (U^0,\phi) &= (u_0,\phi), \quad \forall \phi \in S_N. && (2.35b) \end{aligned}$$

By taking $\phi = \Phi_p$, $|p|_\infty \le N$, then we obtain expressions for the coefficients $\widehat{U^m}(p)$ of U^m, in terms of an integral involving $E(t^m;t^{m-1})U^{m-1}$. Unfortunately, calculating these integrals exactly entails at least as much work as finding the exact solution of the original differential equation. Thus some form of numerical quadrature must be used. We will return to this subject in the next section, after having discussed the stability and convergence properties of the exactly-integrated scheme (2.35).

2.2.1 Stability

The only approximation involved in (2.35) is the introduction of the projections carried out in obtaining the initial datum and at each time step. Since the projection operator Π_N is non-expansive in $H^\sigma_\#(\Omega)$, $\sigma \ge 0$, the stability properties of (2.35) exactly mirror those of (2.1). This result is expressed in the following theorem, which is an immediate consequence of Lemma 2.1.

Theorem 2.3 *Let $u_0 \in L^2_\#(\Omega)$ and $a \in L^1(C^1_\#(\overline{\Omega})^n)$. Then the method (2.35) is unconditionally stable in $l^\infty(L^2_\#(\Omega))$, and*

$$\|U\|_{l^\infty(L^2_\#(\Omega))} \le e^{\frac{1}{2}|\nabla\cdot a|_{0,1}} \left\|U^0\right\|. \tag{2.36}$$

Moreover, if $a \in L^1(C^{[\sigma]^*}_\#(\overline{\Omega})^n)$, for some $\sigma \ge 0$, then unconditional stability in $H^\sigma_\#(\Omega)$ applies, and we have

$$\|U\|_{l^\infty(H^\sigma_\#(\Omega))} \le e^{C_2} \left\|U^0\right\|_{H^\sigma_\#(\Omega)}, \tag{2.37}$$

where C_2 is as in Theorem 2.2.

Remark. If $\|\nabla\cdot a\|_{L^1(0,\infty;C_\#(\overline{\Omega}))} < \infty$, then the method is unconditionally asymptotically stable in $L^2_\#(\Omega)$.

2.2.2 Convergence

Here we will be mainly concerned to establish the convergence (as $\Delta t \to 0$ and $N \to \infty$) of U to u in the $l^\infty(L^2_\#(\Omega))$-norm, although, as above, extensions of the results to higher order norms will be indicated. At each time step, as the projection is carried out, an error is incurred, and these errors accumulate as the calculation proceeds. Our concern, then, will be to estimate the size of these errors, and how they are affected by the size of the timestep. The principal result of this subsection is the following.

Theorem 2.4 *Let $u_0 \in H^\sigma_\#(\Omega)$, $a \in L^1(C^{[\sigma]^*}_\#(\overline{\Omega})^n)$, $\sigma \ge 0$. Then*

$$\|u - U\|_{l^\infty(L^2_\#(\Omega))} \le e^{C_2} N^{-\sigma} \left[1 + \min(\frac{T}{\Delta t}, n^{\frac{1}{2}} N |a|_{0,1})\right] \|u_0\|_\sigma\,. \tag{2.38}$$

Remark. We note that the term $\min(\frac{T}{\Delta t}, n^{\frac{1}{2}}N|a|_{0,1})$ which appears in (2.38) is bounded by $\frac{T}{\Delta t}\min(1, n^{\frac{1}{2}}N\Delta t|a|_{0,\infty})$, thus indicating the role played by the Courant number in this estimate.

Proof For $m = 0, \ldots, M$, denote $u(\cdot,t^m)$ by u^m, and define $\eta^m = (I - \Pi_N)u^m$ and $\xi^m = \Pi_N u^m - U^m$. Then, for each $m = 1, \ldots, M$,

$$\xi^m = \Pi_N E(t^m;t^{m-1})(\xi^{m-1} + \eta^{m-1}), \tag{2.39}$$

so that

$$\|\xi^m\| \le e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{m-1},t^m]}}(\left\|\xi^{m-1}\right\| + \left\|\eta^{m-1}\right\|). \tag{2.40}$$

Alternatively, we may expand $\Pi_N E(t^m;t^{m-1})\eta^{m-1}$ to give

$$\begin{aligned} \Pi_N E(t^m;t^{m-1})\eta^{m-1} &= \Pi_N(E(t^m;t^{m-1}) - I)\eta^{m-1} \\ &= \Pi_N \int_{t^m}^{t^{m-1}} \frac{d}{dt}E(t^m;t)\eta^{m-1}dt \\ &= \Pi_N \int_{t^m}^{t^{m-1}} E(t^m;t)(a\cdot\nabla\eta^{m-1})dt, \end{aligned}$$

so that

$$\begin{aligned} \left\|\Pi_N E(t^m;t^{m-1})\eta^{m-1}\right\| &\le \int_{t^{m-1}}^{t^m} \left\|E(t^m;t)(a\cdot\nabla\eta^{m-1})\right\| dt \\ &\le e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{m-1},t^m]}} \int_{t^{m-1}}^{t^m} \left\|a\cdot\nabla\eta^{m-1}\right\| dt; \end{aligned}$$

thus we obtain

$$\|\xi^m\| \le e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{m-1},t^m]}}(\left\|\xi^{m-1}\right\| + |a|_{0,1;[t^{m-1},t^m]}\left\|\nabla\eta^{m-1}\right\|). \tag{2.41}$$

Since $\left\|\xi^0\right\| = 0$, it follows by induction (and by combining (2.40) and (2.41)) that, for $m = 1, \ldots, M$,

$$\begin{aligned} \|\xi^m\| &\le \sum_{k=0}^{m-1} e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{k-1},t^k]}} \min(\left\|\eta^k\right\|, |a|_{0,1;[t^{k-1},t^k]}\left\|\nabla\eta^k\right\|) \\ &\le N^{-\sigma} e^{\frac{1}{2}|\nabla\cdot a|_{0,1}} \sum_{k=0}^{m-1} \min(1, n^{\frac{1}{2}}N|a|_{0,1;[t^{k-1},t^k]}) \left\|u^k\right\|_\sigma, && (2.42) \end{aligned}$$

where we have made use of the estimate

$$\|(I - \Pi_N)v\|_s \le N^{s-\sigma}\|v\|_\sigma, \quad \sigma \ge s,\ v \in H^\sigma_\#(\Omega) \tag{2.43}$$

(see, for example, [52]). The theorem now follows by adding $\|\eta^m\|$ to both sides of (2.42), again applying (2.43), and applying (2.2) to bound each $\left\|u^k\right\|_\sigma$ by $\|u_0\|_\sigma$.

An alternative analysis of the error terms may be carried out by writing, for $m = 0, \ldots, M$, $\zeta^m = \xi^m + \eta^m$. Then we find that

$$\zeta^m = E(t^m;t^{m-1})\zeta^{m-1} + (I - \Pi_N)E(t^m;t^{m-1})U^{m-1}. \tag{2.44}$$

The mechanism for the generation and transmission of the error terms as the calculation proceeds is perhaps clearer from this perspective than in (2.39). We see from (2.44) that, in the mth time step, the error as it stands at t^{m-1} is evolved according to the exact equation, but an additional error is incurred in projecting $E(t^m;t^{m-1})U^{m-1}$. The analysis may be continued in a similar way to the proof of Theorem 2.4 by making use of Lemma 2.2.

The accumulation of errors committed at each time step (corresponding to the term $\frac{T}{\Delta t}$) indicates that the total error is minimised when only one time step is used. This is of course true in this instance. However, when other terms (such as diffusion) are included, or when nonlinear problems are considered, other error terms will appear, which can only be reduced in size when the timestep Δt is reduced.

Here, as Δt is reduced, the analysis predicts that the error will increase, up to a maximum value which will be bounded from above by

$$e^{C_2}N^{-\sigma}\left[1+N|a|_{0,1}\right]\|u_0\|_\sigma .$$

As regards the behaviour as $N\to\infty$, we see from (2.38) that, if $\Delta t\to 0$ simultaneously, the convergence rate is $N^{1-\sigma}$, whereas for each fixed Δt, the error will decrease as $N^{-\sigma}$. This suboptimal convergence in the former case is consistent with standard results in the semidiscrete case for the Fourier pseudospectral method for advection problems [52], [59]. Pasciak [59] comments that since in the constant coefficient case an optimal estimate holds, this result is possibly not sharp; in practice he observes the faster convergence rate. Our numerical experiments (Chapter 3) also point to this conclusion, and we shall discuss this point further in that chapter.

2.3 The introduction of numerical quadrature

As already mentioned, the integrals appearing on the right hand side of (2.35) can be evaluated exactly only in exeptional cases (such as when a is constant); in general it is necessary to employ some form of numerical quadrature. Here, a compound trapezium rule is used with nodes $x_\nu=\nu\pi/L$, $\nu=(\nu_1,\ldots,\nu_n)$, $0\le\nu_j\le 2L-1$, and then the inner product on $L^2_\#(\Omega)$ is approximated by the *discrete* inner product defined, for u, $v\in C_\#(\overline{\Omega})$, by

$$(u,v)_L:=(2L)^{-n}\sum_\nu u(x_\nu)\overline{v(x_\nu)}.$$

If u, $v\in S_N$ and $N<L$, then

$$(u,v)_L=(u,v).$$

By analogy with the continuous case, we define the discrete Fourier coefficients of $v\in C_\#(\overline{\Omega})$ by

$$\hat{v}_p=(v,\Phi_p)_L,\ |p|_\infty\le L;$$

then we have, according to the Poisson summation formula, the aliasing relation

$$\hat{v}_p=\sum_{k\in Z^n}\hat{v}(p+2kL),\ |p|_\infty\le L. \tag{2.45}$$

We demonstrate that the results of the previous section hold when all inner products are replaced by their discrete counterparts, as long as L is large enough in comparison to N. Thus the numerical method (2.35) is replaced by: find $U=\{U^m\}_{m=1}^M\subset S_N$ such that

$$\begin{aligned}(U^m,\phi)_L &= (E(t^m;t^{m-1})U^{m-1},\phi)_L,\quad \forall\phi\in S_N,\ m=1,\ldots,M && (2.46a)\\ (U^0,\phi)_L &= (u_0,\phi)_L,\quad \forall\phi\in S_N. && (2.46b)\end{aligned}$$

By taking $\phi=\Phi_p$, $|p|_\infty\le N$, then we obtain expressions for the coefficients $\widehat{U^m_p}$ of U^m, in terms of a summation involving values of $E(t^m;t^{m-1})U^{m-1}$ at the quadrature points. Once $\{X(x_\nu;t^m;t^{m-1})\}$ is known, these values may themselves be calculated using the coefficients of U^{m-1}.

The use of the discrete inner product above is equivalent to the application of an interpolation operator to $E(t^m;t^{m-1})U^{m-1}$, with the interpolation points being the same as the quadrature points. (When $N=L$, the resulting scheme is actually a *collocation* method.) Let P_L denote the interpolation operator from $C_\#(\overline{\Omega})$ into S_L defined by

$$P_Lv=\sum_{|p|_\infty\le L}\frac{1}{c(p)}\hat{v}_p\Phi_p,$$

where $c(p)=\prod_{i=1}^n\kappa(p_i)$, with

$$\kappa(k)=\left\{\begin{array}{ll}1 & \text{if } |k|\neq L\\ 2 & \text{if } |k|=L\end{array}\right. .$$

It is readily checked that P_L interpolates at the quadrature points x_ν (this is ensured by the term $\frac{1}{c(p)}$), and that, for $v\in C_\#(\overline{\Omega})$,

$$(P_Lv,\Phi_p)_L=\hat{v}_p,$$

so that P_L is a projection in the sense that $P_L^2=P_L$.

We may now rewrite (2.46) in a more compact operator notation as

$$\begin{aligned}U^m &= \Pi_NP_LE(t^m;t^{m-1})U^{m-1},\quad m=1,\ldots,M && (2.47a)\\ U^0 &= \Pi_NP_Lu_0. && (2.47b)\end{aligned}$$

2.3.1 Accuracy

Later in this section our concern will be to discuss the effect of the introduction of quadrature on the stability and convergence properties established in the previous section for the exactly-integrated scheme (2.35). Here we discuss the extra error incurred, which may be seen to be a result of the aliasing phenomenon (2.45). The effect of aliasing has been the subject of much discussion among researchers and users of spectral methods. Several ingeneous approaches towards controlling (or eliminating) its effect have been employed. Now the consensus appears to be that its effect on the accuracy of the spectral approximation is not significant, and that its presence will not be detrimental to the success of any particular computation. Theoretical justification for this confidence is the following approximation result, which demonstrates that the interpolation error decays at the same rate as the projection error when the degree of the approximating polynomial is increased.

Lemma 2.4 *Let $u\in H^\sigma_\#(0,2\pi)$, $\sigma>\frac{1}{2}$, and let $0\le\rho\le\sigma$. Then*

$$\|(P_L-I)u\|_\rho\le C_\sigma L^{\rho-\sigma}\|u\|_\sigma , \tag{2.48}$$

where $C_\sigma=(2\zeta(2\sigma))^{\frac{1}{2}}$, and $\zeta(\cdot)$ is Riemann's zeta function.

The proof of this standard result may be found in [52]. The extension of the result to general dimensions is straightforward. We quote it below, and run through its proof, since the method of proof, and some of the notation introduced in the course of the proof, will be used again in later results.

Lemma 2.5 *Let $u\in H^\sigma_\#(\Omega)$, $\sigma>\frac{n}{2}$, and let $0\le\rho\le\sigma$. Then*

$$\|(P_L-I)u\|_\rho\le C_{n,\sigma}L^{\rho-\sigma}\|u\|_\sigma , \tag{2.49}$$

where $C_{n,\sigma}=[(1+C_{\sigma/n})^n-1]$.

Proof We must first introduce some notation. For $x\in R^n$, write

$$x=\sum_{i=1}^n x^ie_{(i)},$$

where $e_{(i)}$ is a unit vector in the i^{th} coordinate direction. Further, write

$$x^-_{(i)}=x-x^ie_{(i)}.$$

We can now define $P_L^{(i)}$, the interpolation operator in the ith coordinate direction, by

$$P_L^{(i)}u(x)=\sum_{|p_i|\le L}\frac{1}{\kappa(p_i)}\left(\frac{1}{2L}\sum_{\nu=0}^{2L-1}u(x^-_{(i)}+y_\nu e_{(i)})\overline{\phi_{p_i}(y_\nu)}\right)\phi_{p_i}(x^i),$$

where $y_\nu=\nu\pi/2L$, $\nu=0,\ldots,2L-1$. Then the following operator identity holds:

$$P_L=\prod_{i=1}^n P_L^{(i)}.$$

We should like to expand P_L-I as a sum of products of terms such as $P_L^{(i)}-I$, since we can apply Lemma 2.4 to such terms. Making use of (1.12), the above operator identity may be rewritten

$$\begin{aligned}P_L &= \prod_{i=1}^n P_L^{(i)}=\prod_{i=1}^n((P_L^{(i)}-I)+I)\\ &= \sum_{\lambda\in\Lambda_0^n}\prod_{i=1}^n(P_L^{(i)}-I)^{\lambda_i},\end{aligned}$$

on the understanding that, if F is an operator,

$$F^{\lambda_i} = \begin{cases} F & \text{if } \lambda_i = 1 \\ I & \text{if } \lambda_i = 0. \end{cases}$$

Thus

$$P_L - I = \sum_{\lambda \in \Lambda^n} \prod_{i=1}^{n} (P_L^{(i)} - I)^{\lambda_i}, \tag{2.50}$$

and we have, by (2.48),

$$\begin{aligned} \|(P_L - I)u\|_\rho &= \left\| \sum_{\lambda \in \Lambda^n} \prod_{i=1}^{n} (P_L^{(i)} - I)^{\lambda_i} u \right\|_\rho \\ &\le \sum_{\lambda \in \Lambda^n} \left\| \prod_{i=1}^{n} (P_L^{(i)} - I)^{\lambda_i} u \right\|_\rho \\ &\le \sum_{\lambda \in \Lambda^n} C_{\sigma/|\lambda|_1}^{|\lambda|_1} L^{\rho-\sigma} \|u\|_\sigma \\ &\le \sum_{\lambda \in \Lambda^n} C_{\sigma/n}^{|\lambda|_1} L^{\rho-\sigma} \|u\|_\sigma \end{aligned}$$

and (2.49) then follows by applying (1.12) in reverse.

2.3.2 Stability analysis

The above results indicate that the introduction of quadrature (and therefore aliasing) is not detrimental to the accuracy of a spectral approximation—at least not to the asymptotic rate of convergence. Its effect on the stability (and therefore convergence) of the method as applied to a time-dependent p.d.e. is perhaps more delicate. In the exactly integrated scheme, stability was an automatic consequence of the well-posedness of the exact problem and the fact that the operator Π_N is non-expansive. Unfortunately, as a result of (2.45), far from being non-expansive, P_L is not even bounded in $L^2_\#(\Omega)$. (This is indicated by the restriction $\sigma > n/2$ in Lemma 2.5.) Thus in this case stability will not be so straightforward to establish.

Our results will be based on regarding the term $\Pi_N P_L E(t^m; t^{m-1})U^{m-1}$ as a perturbation of $\Pi_N E(t^m; t^{m-1})U^{m-1}$, and using the aliasing relation (2.45) to bound the difference. We begin with the following lemma, which makes use of Lemma 2.5[1].

Lemma 2.6 *Let $a \in L^1(C_\#^{[\sigma]^*}(\overline{\Omega})^n)$, $\sigma > \frac{n}{2}$, and let $0 \le \rho \le \sigma$. Let $\Delta t_0 > 0$. Then there is a constant $C_3 = C_3(n, \sigma, \Delta t_0, |a|_{[\sigma]^*,1})$ such that for $t_0, t \in [0, T]$, with $|t_0 - t| \le \Delta t_0$, and for $V \in \Pi_N P_L S_N$ with $L \ge N$,*

$$\|(P_L - I)E(t_0; t)V\|_\rho \le e^{C_2(t_0,t)} C_3 |a|_{[\sigma]^*,1;[t_0,t]} L^{\rho-\sigma} \|\nabla V\|_\sigma , \tag{2.51}$$

where $C_2(t_0, t)$ is as given in Theorem 2.2.

Proof Since $P_L \Pi_N P_L = \Pi_N P_L$ when $N \le L$, $(P_L - I)V = 0$, so that

$$(P_L - I)E(t_0; t)V = (P_L - I)(E(t_0; t) - I)V.$$

Now

$$\begin{aligned} (E(t_0; t) - I)V &= \int_{t_0}^{t} \frac{d}{d\tau}(E(t_0; \tau)V) d\tau \\ &= \int_{t_0}^{t} E(t_0; \tau)(a \cdot \nabla V) d\tau \quad \text{(by the chain rule).} \end{aligned}$$

Application of Lemma 2.5 and Lemma 2.2 now gives

$$\begin{aligned} &\|(P_L - I)E(t_0; t)V\|_\rho \\ &\le C_{n,\sigma} L^{\rho-\sigma} \left\| \int_{t_0}^{t} E(t_0; \tau)(a \cdot \nabla V) d\tau \right\|_\sigma \\ &\le C_{n,\sigma} L^{\rho-\sigma} e^{C_2(t_0,t)} \|a \cdot \nabla V\|_{L^1(t_0,t; H^\sigma_\#(\Omega))} , \end{aligned}$$

[1] We note that (2.45) plays a key role in the proof of Lemma 2.4, and therefore also in that of Lemma 2.5, of which the first stability result is an almost direct application.

and (2.51) follows since, by function space interpolation,

$$\|a \cdot \nabla V\|_{L^1(t_0,t; H^\sigma_\#(\Omega))} \le C |a|_{[\sigma]^*,1;[t_0,t]} \|\nabla V\|_\sigma .$$

Lemma 2.6 works because of the smoothness we have imposed on a, and therefore on $E(t_0; t)$. P_L *is* bounded on S_N, and this smoothness ensures that the application of $E(t_0; t)$ does not take V too far away from S_N, so that the effect of P_L on $E(t_0; t)V$ is controllable. In fact we are able to bound the difference between $P_L E(t_0; t)V$ and $E(t_0; t)V$ by $O(|t_0 - t|)$, albeit at the cost of one derivative. Stability follows from a simple corollary of Lemma 2.6 and Lemma 2.2, which we now state.

Corollary 1 *Suppose that the conditions of Lemma 2.6 hold, and that $L > N$. Then there is a constant $C_4 = C_4(n, \sigma, \Delta t_0, |a|_{[\sigma]^*,1})$ such that, for $|t_0 - t| \le \Delta t_0$, and for $V \in S_N$,*

$$\|\Pi_N P_L E(t_0; t)V\|_\rho \le e^{C_2} \left[1 + C_4 \left(\frac{N^{1+\sigma-\rho}}{L^{\sigma-\rho}} \right) |a|_{[\sigma]^*,1;[t_0,t]} \right] \|V\|_\rho , \tag{2.52}$$

where $C_2(t_0, t)$ is as given in Theorem 2.2.

To ensure the stability of (2.47) via (2.52), the loss of one derivative must be countered by making $L > N$ in such a way that as N and L are increased,

$$(N^{1+\sigma-\rho} / L^{\sigma-\rho}) \le C. \tag{2.53}$$

[We note that the more smoothness we have on a, the closer N and L can remain.] When (2.53) holds, and taking $(t_0, t) = (t^m, t^{m-1})$ at successive time steps, the stability of (2.47) is guaranteed independently of Δt, so that the unconditional stability of the exactly-integrated scheme (2.35) is carried over to the method (2.47).

The relationship (2.53) is one between the number of quadrature points and the smallest scale being represented by the discrete solution. A similar relationship, between the number of particles and the smallest scale represented by the solution, is often required in the context of vortex methods (e.g. [16]).

Stability was proved in [79] assuming the slightly more restrictive relationship

$$L \ge (1 + \epsilon N^{n^2/\sigma})N. \tag{2.54}$$

Tadmor [81] gives a relationship of the form

$$L \ge (1 + \epsilon)N \tag{2.55}$$

as a sufficient condition for the stability of a standard *semidiscrete* Fourier-Galerkin approximation. In [78], (2.55) was used, together with the additional assumption $N\Delta t \le C$, to establish the stability of the method (2.47). This latter assumption could be viewed as a CFL-type restriction on the method, which is something we are trying to avoid by using the Lagrange-Galerkin approach. The approach we shall now employ is similar to that in [79], but with some refinement, and we are able to prove stability subject to a relationship of the form (2.55), with no timestep limitation.

Lemma 2.7 (c.f. Lemma 2.6) *Let $a \in L^1(C_\#^{[\sigma]^*}(\overline{\Omega})^n)$, with $\sigma > \frac{n+1}{2}$ when $n > 1$, and with $\sigma \ge 1$ when $n = 1$, and let $0 \le \rho \le \sigma$. Let $\Delta t_0 > 0$, and let L and N be positive integers related by (2.55), for some $\epsilon > 0$. Then there is a constant $C_5 = C_5(n, \sigma, \Delta t_0, \epsilon, |a|_{[\sigma]^*,1})$ such that for $t_0, t \in [0, T]$, with $|t_0 - t| \le \Delta t_0$, and for $V \in S_N$,*

$$\|\Pi_N (P_L - I)E(t_0; t)V\|_\rho \le e^{C_2(t_0,t)} C_5 |a|_{[\sigma]^*,1;[t_0,t]} N^{\rho-\sigma} \|V\|_\sigma , \tag{2.56}$$

where $C_2(t_0, t)$ is as given in Theorem 2.2.

Proof In order to communicate effectively the main ideas in the proof, we demonstrate the result first in one dimension, for the case $\sigma = 1$, $\rho = 0$. The proof involves temporarily expanding the left hand side in terms of the Fourier coefficients of V, and exploiting the nature of the Fourier basis functions Φ_p.

From the aliasing relation (2.45) we can write

$$\Pi_N (P_L - I)E(t_0; t)V = \sum_{|q| \le N} \frac{1}{c(q)} \sum_{k \ne 0} (E(t_0; t)V, \Phi_{q+2kL}) \Phi_q .$$

The right hand side can be expanded further to give

$$\Pi_N (P_L - I)E(t_0; t)V = \sum_{|q| \le N} \frac{1}{c(q)} \sum_{|p| \le N} \hat{V}(p) \sum_{k \ne 0} (E(t_0; t)\Phi_p, \Phi_{q+2kL}) \Phi_q .$$

Now

$$E(t_0;t)\Phi_p(x) = \Phi_p(X(x,t_0;t)) = \Phi_p(X(x,t_0;t)-x)\Phi_p(x).$$

Thus if we drop the explicit dependence on t_0 and t and define

$$\mathcal{E}\Phi_p(x) = \Phi_p(X(x,t_0;t)-x),$$

then $E(t_0;t)\Phi_p = (\mathcal{E}\Phi_p)\Phi_p$, and, taking the bounds $|p|,|q| \leq N$ and $|q| \leq L$L to be understood,

$$\Pi_N(P_L - I)E(t_0;t)V = \sum_p \sum_q \frac{1}{c(q)} \hat{V}(p) \sum_{k\neq 0} (\mathcal{E}\Phi_p, \Phi_{q-p+2kL})\Phi_q. \tag{2.57}$$

Now, we may integrate by parts the term $(\mathcal{E}\Phi_p, \Phi_{q-p+2kL})$, giving

$$(\mathcal{E}\Phi_p, \Phi_{q-p+2kL}) = \frac{(D(\mathcal{E}\Phi_p), \Phi_{q-p+2kL})}{i(q-p+2kL)}. \tag{2.58}$$

The term appearing in the integrand on the right hand side of (2.58) may be expanded in the following way:

$$D(\mathcal{E}\Phi_p(x)) = \left(\frac{d}{dx}(X(x,t_0;t)-x)\right)(ip)\mathcal{E}\Phi_p(x).$$

In order to ease presentation, we write $\frac{d}{dx}(X(x,t_0;t)-x) = B(x)$; then in notation drawn from the proof of Lemma 2.2 we have that $\|B\|_{C^\infty_\#(\overline{\Omega})} = \mathcal{E}_0(t)$, and (2.26) provides the bound

$$\mathcal{E}_0(t) \leq e^{|a|_{1,1;[t_0,t]}} - 1. \tag{2.59}$$

The denominator of (2.58) may be rewritten in a binomial expansion

$$\frac{1}{i(q-p+2kL)} = \sum_{r=0}^{\infty} \frac{(ip)^r}{[i(q+2kL)]^{r+1}},$$

so that (2.58) becomes

$$\begin{aligned}
(\mathcal{E}\Phi_p, \Phi_{q-p+2kL}) &= \frac{(D(\mathcal{E}\Phi_p), \Phi_{q-p+2kL})}{i(q-p+2kL)} \\
&= \sum_{r=0}^{\infty} \frac{(ip)^{r+1}}{[i(q+2kL)]^{r+1}}(B\mathcal{E}\Phi_p, \Phi_{q-p+2kL}) \\
&= \sum_{r=0}^{\infty} \frac{(ip)^{r+1}}{[i(q+2kL)]^{r+1}}(BE\Phi_p, \Phi_{q+2kL}).
\end{aligned}$$

Thus, substituting this back into (2.57) we have

$$\Pi_N(P_L - I)E(t_0;t)V = \sum_{r=0}^{\infty} \sum_q \sum_{k\neq 0} \frac{1}{c(q)} \frac{(BED^{r+1}V, \Phi_{q+2kL})}{[i(q+2kL)]^{r+1}} \Phi_q. \tag{2.60}$$

We shall write the right hand side of (2.60) as $\sum_{r=0}^{\infty} A_r$, and proceed to bound the terms A_r in turn. We have

$$\begin{aligned}
\|A_r\|^2 &\leq \sum_q \left(\frac{1}{c(q)} \sum_{k\neq 0} \frac{|(BED^{r+1}V, \Phi_{q+2kL})|}{|q+2kL|^{r+1}}\right)^2 \\
&\leq \sum_q \left(\frac{1}{c(q)} \sum_{k\neq 0} |(BED^{r+1}V, \Phi_{q+2kL})|^2\right) \left(\frac{1}{c(q)} \sum_{k\neq 0} |q+2kL|^{-2(r+1)}\right) \\
&\leq \|B\|^2_{C_\#(\overline{\Omega})} \left\|ED^{r+1}V\right\|^2 \max_q \left(\sum_{k\neq 0} |q+2kL|^{-2(r+1)}\right).
\end{aligned} \tag{2.61}$$

With L and N related by (2.55),

$$\begin{aligned}
\max_q \sum_{k\neq 0} |q+2kL|^{-2(r+1)} &\leq 2\sum_{k=1}^{\infty} |2kL - L|^{-2(r+1)} \\
&\leq 2[(1+\epsilon)N]^{-2(r+1)} \sum_{k=1}^{\infty} (2k-1)^{-2(r+1)} \\
&\leq 2[(1+\epsilon)N]^{-2(r+1)}\zeta(2),
\end{aligned} \tag{2.62}$$

where ζ is Riemann's zeta function. Moreover,

$$\begin{aligned}
\left\|ED^{r+1}V\right\|^2 &\leq e^{2C_2(t_0,t)} \sum_q |q|^{2(r+1)} |\hat{V}(q)|^2 \\
&\leq e^{2C_2(t_0,t)} N^{2r} \|DV\|^2.
\end{aligned} \tag{2.63}$$

Together, (2.59), (2.61), (2.62) and (2.63) give that

$$\|A_r\| \leq (2\zeta(2))^{\frac{1}{2}}(e^{|a|_{1,1;[t_0,t]}} - 1)e^{C_2(t_0,t)}(1+\epsilon)^{-(r+1)}N^{-1}\|V\|_1,$$

so that

$$\|\Pi_N(P_L - I)E(t_0;t)V\| \leq \frac{(2\zeta(2))^{\frac{1}{2}}}{\epsilon}(e^{|a|_{1,1;[t_0,t]}} - 1)e^{C_2(t_0,t)}N^{-1}\|V\|_1. \tag{2.64}$$

In this way the proof for the one-dimensional case with $\sigma = 1$ and $\rho = 0$ is completed. The case $\rho > 0$ (in general dimensions) may be reduced to the case $\rho = 0$ by noting that the left hand side of (2.56) is a norm of a function in S_N, and for $v \in S_N$ we have the inverse inequality

$$\|v\|_\rho \leq (1 + nN^2)^{\rho/2}\|v\|.$$

The extension to higher dimensions is somewhat of a notational headache, although there is otherwise no particular technical difficulty. The basic idea is the following. We expand $P_L - I$ by the formula (2.50), and note that, since $\lambda = 0$ is excluded, each term in the summation involves the application of $P_L^{(i)} - I$ for at least one i. The result is then obtained by applying the one-dimensional version of Lemma 2.7 in that direction, and by applying Lemma 2.5 in the other directions for which $\lambda_i = 1$.

By rearranging (2.50) we can write, for $n > 1$,

$$\Pi_N(P_L - I) = \Pi_N^{(1)}(P_L^{(1)} - I) + \sum_{j=2}^{n} \left[\sum_{\lambda \in \Lambda_0^{j-1}} \prod_{i=1}^{j-1} \Pi_N^{(i)}(P_L^{(i)} - I)^{\lambda_i}\right] \Pi_N^{(j)}(P_L^{(j)} - I),$$

where $\Pi_N^{(i)}$ is projection with respect to the ith coordinate direction only, and use is made of the fact that $\Pi_N^{(i)}$ commutes with $P_L^{(j)}$ for $i \neq j$. We set $W_j = \Pi_N^{(j)}(P_L^{(j)} - I)E(t_0;t)V$, $j = 1,\ldots,n$. Then Lemma 2.5 implies that, so long as $\sigma - 1 > (n-1)/2$,

$$\|\Pi_N(P_L - I)E(t_0;t)V\| \leq \|W_1\| + \sum_{j=2}^{n} C_{j-1,\sigma}L^{-(\sigma-1)}\|W_j\|_{\sigma-1}. \tag{2.65}$$

Thus we seek to bound terms of the form $\|W_j\|_\rho$, for $\rho = 0$ and $\rho = \sigma - 1$. We shall investigate the case σ integer—the corresponding result for non-integer σ being obtained thence by interpolation.

We may expand W_j to give

$$\begin{aligned}
W_j &= \Pi_N^{(j)}(P_L^{(j)} - I)E(t_0;t)V \\
&= \sum_{|q_j|\leq N} \frac{1}{\kappa(q_j)} \sum_{|p|_\infty \leq N} \hat{V}(p) \sum_{k_j \neq 0} (E\Phi_p, \phi_{q_j+2k_jL})_j \phi_{q_j},
\end{aligned} \tag{2.66}$$

where $(\cdot,\cdot)_j$ denotes an integral carried out in the jth coordinate direction only, and where ϕ_{q_j} is a function of x^j only. The term involving the inner product may be rewritten

$$(E\Phi_p, \phi_{q_j+2k_jL})_j = \left(\left(\prod_{i\neq j} E\phi_{p_i}\right)\mathcal{E}\phi_{p_j}, \phi_{q_j-p_j+2k_jL}\right)_j$$

$$
\begin{aligned}
&= \frac{\left(D_j\left(\prod_{i\neq j} E\phi_{p_i}\right)\mathcal{E}\phi_{p_j}, \phi_{q_j-p_j+2k_jL}\right)_j}{i(q_j-p_j+2k_jL)} \\
&= \sum_{l=1}^{n} ip_l \frac{(\xi_{jl}E\Phi_p, \phi_{q_j+2k_jL})_j}{i(q_j-p_j+2k_jL)},
\end{aligned}
$$

where ξ_{jl} denotes $D_jX_l - \delta_{jl}$ as in the proof of Lemma 2.2. Now let α be a multiindex of order $\leq \rho$, and let β be another multiindex of order $\sigma - 1$, which is such that $\beta_i = \alpha_i$ when $i \neq j$. After some integration by parts in the inner product, and expanding $1/(q_j - p_j + 2k_jL)$ as in the one-dimensional case, we obtain

$$
\begin{aligned}
D^\alpha W_j &= \sum_{r=0}^{\infty}\sum_{q_j}\sum_{l=1}^{n}\sum_{k_j\neq 0} \frac{1}{\kappa(q_j)}(D^\beta(\xi_{jl}ED_j^rD_lV), \phi_{q_j+2k_jL})_j \times \\
&\quad \times \frac{(iq_j)^{\alpha_j}}{[i(q_j+2k_jL)]^{\beta_j+r+1}}\phi_{q_j}. \qquad (2.67)
\end{aligned}
$$

The remaining steps are now very similar to the one-dimensional case. We write the right hand side as a sum of terms A_r, and bound each one as follows.

$$
\begin{aligned}
\|A_r\|^2 &\leq \sum_{q_j}\left(\sum_{l=1}^{n}\sum_{k_j\neq 0}\frac{1}{\kappa(q_j)}|(D^\beta(\xi_{jl}ED_j^rD_lV), \phi_{q_j+2k_jL})_j| \times \right. \\
&\quad \left. \times \left|\frac{q_j}{(q_j+2k_jL)}\right|^{\alpha_j} |q_j+2k_jL|^{\rho-\sigma-r}\right)^2 \\
&\leq \sum_{l=1}^{n}\left\|D^\beta(\xi_{jl}ED_j^rD_lV)\right\|^2 \max_{q_j}\sum_{k_j\neq 0}|q_j+2k_jL|^{2(\rho-\sigma-r)} \\
&\leq 2\zeta(2)L^{2(\rho-\sigma-r)}\left\|D^\beta(\xi_{jl}ED_j^rD_lV)\right\|^2 \\
&\leq 2\zeta(2)L^{2(\rho-\sigma-r)}\mathcal{E}_{\sigma-1}^2(t)e^{2C_2(t_0,t)}N^{2r}\|V\|_\sigma^2, \qquad (2.68)
\end{aligned}
$$

where $\mathcal{E}_{\sigma-1}(t)$ is as in the proof of Lemma 2.2, and may be bounded in terms of $|a|_{[\sigma]^*,1;[t_0,t]}$ by use of (2.26) and (2.27). The proof may now be completed by summing over r and over all multiindices α of order $\leq \rho$.

We now end our discussion of stability with the following theorem.

Theorem 2.5 (Stability) *Suppose that the conditions of Lemma 2.7 hold, and that $\Delta t \leq \Delta t_0$. Then the method (2.47) is unconditionally stable, and for any $0 \leq \rho \leq \sigma$,*

$$
\|U\|_{l^\infty(H^\rho_\#(\Omega))} \leq e^{C_2(0,T)+C_5|a|_{[\sigma]^*,1}}\left\|U^0\right\|_\rho \qquad (2.69)
$$

Proof The proof is by induction. The inductive step is provided by

$$
\|U^m\|_\rho \leq \left\|\Pi_N E(t^{m-1};t^m)U^{m-1}\right\|_\rho + \left\|\Pi_N(P_L-I)E(t^{m-1};t^m)U^{m-1}\right\|_\rho.
$$

Use of Lemma 2.2 and Lemma 2.7 now gives

$$
\|U^m\|_\rho \leq e^{C_2(t^{m-1},t^m)+C_5|a|_{[\sigma]^*,1;[t^{m-1},t^m]}}\left\|U^{m-1}\right\|_\rho,
$$

as is required.

2.3.3 Convergence

Our goal here is to give an analogous convergence result to Theorem 2.4, showing that the method (2.47) enjoys similar spectral accuracy to (2.35), and that this is irrespective of the size of the timestep under the condition on L and N given by (2.55). Our result is the following.

Theorem 2.6 (Convergence) *Suppose that the conditions of Lemma 2.7 hold, and that $u_0 \in H^\sigma_\#(\Omega)$. Suppose also that $\Delta t \leq \Delta t_0$. Then*

$$
\begin{aligned}
&\|u-U\|_{l^\infty(L^2_\#(\Omega))} \leq \\
&\quad e^{C_2+C_5|a|_{[\sigma]^*,1}}N^{-\sigma}\|u_0\|_\sigma\left[1+C_{n,\sigma}+\min(\frac{T}{\Delta t}, n^{\frac{1}{2}}N|a|_{0,1})+C_5|a|_{[\sigma]^*,1}\right], \qquad (2.70)
\end{aligned}
$$

where C_2 and C_5 are as in Lemma 2.7.

Proof For $m = 0, \ldots, M$, we denote $u(\cdot, t^m)$ by u^m, and we define $\eta^m = (I - \Pi_N)u^m$ and $\xi^m = \Pi_N u^m - U^m$. Then, for each $m = 1, \ldots, M$,

$$
\begin{aligned}
\xi^m &= \Pi_N P_L E(t^m;t^{m-1})\xi^{m-1} + \Pi_N E(t^m;t^{m-1})\eta^{m-1} \\
&\quad -\Pi_N(P_L-I)E(t^m;t^{m-1})\Pi_N u^{m-1}. \qquad (2.71)
\end{aligned}
$$

As in Theorem 2.5,

$$
\begin{aligned}
&\left\|\Pi_N P_L E(t^m;t^{m-1})\xi^{m-1}\right\| \\
&\quad \leq e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{m-1},t^m]}+C_5|a|_{[\sigma]^*,1;[t^{m-1},t^m]}}\left\|\xi^{m-1}\right\|, \qquad (2.72)
\end{aligned}
$$

and as in Theorem 2.4,

$$
\begin{aligned}
&\left\|\Pi_N E(t^m;t^{m-1})\eta^{m-1}\right\| \\
&\quad \leq N^{-\sigma}e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{m-1},t^m]}}\min(1, n^{\frac{1}{2}}N|a|_{0,1;[t^{m-1},t^m]})\left\|u^{m-1}\right\|_\sigma. \qquad (2.73)
\end{aligned}
$$

It remains to bound the third term on the right hand side of (2.71). Lemma 2.7 provides

$$
\begin{aligned}
&\left\|\Pi_N(P_L-I)E(t^m;t^{m-1})\Pi_N u^{m-1}\right\| \\
&\quad \leq e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{m-1},t^m]}}C_5|a|_{[\sigma]^*,1;[t^{m-1},t^m]}N^{-\sigma}\left\|u^{m-1}\right\|_\sigma. \qquad (2.74)
\end{aligned}
$$

The inequalities (2.72), (2.73) and (2.74) may now be combined with (2.71) to give

$$
\begin{aligned}
\|\xi^m\| &\leq e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{m-1},t^m]}+C_5|a|_{[\sigma]^*,1;[t^{m-1},t^m]}}\left\|\xi^{m-1}\right\| \\
&\quad +N^{-\sigma}\left\|u^{m-1}\right\|_\sigma e^{\frac{1}{2}|\nabla\cdot a|_{0,1;[t^{m-1},t^m]}}\left[C_5|a|_{[\sigma]^*,1;[t^{m-1},t^m]}\right. \\
&\quad \left. +\min(1, n^{\frac{1}{2}}N|a|_{0,1;[t^{m-1},t^m]})\right]
\end{aligned}
$$

and this may be used to show inductively that

$$
\begin{aligned}
\|\xi^m\| &\leq e^{C_2+C_5|a|_{[\sigma]^*,1}}\left[\left\|\xi^0\right\|\right. \\
&\quad \left. +N^{-\sigma}\|u_0\|_\sigma\left(\min(\frac{T}{\Delta t}, n^{\frac{1}{2}}N|a|_{0,1})+C_5|a|_{[\sigma]^*,1}\right)\right].
\end{aligned}
$$

The result now follows by adding $\|\eta^m\|$ to both sides, applying (2.43), applying (2.2) to bound $\|u_m\|_\sigma$ by $\|u_0\|_\sigma$, and by including an estimate for $\left\|\xi^0\right\|$ provided by Lemma 2.5.

The theoretical estimates we have given in this chapter indicate that the spectral Lagrange-Galerkin method is unconditionally stable and retains the accuracy of the underlying spectral approximation, even when quadrature is introduced. In the next chapter we concern ourselves with the implementation of the method, first describing an additional approximation that must be made in order for the method to be implemented efficiently, and then illustrating the theory with various numerical experiments.

Chapter 3

An efficient implementation

As is common when discussing spectral methods, we are concerned with the efficiency of the scheme, because the global nature of the basis functions can mean that the evaluation of derivatives or of inner products of functions is very expensive. Such operations usually involve the use of a discrete Fourier transform. Using a Fast Fourier Transform (FFT), this may be performed efficiently (by which is usually meant in $O(N^n \log N)$ operations, where n is the dimension of the problem). Without the use of the FFT, the operations count is still only $O(N^{n+1})$, because of the tensor-product nature of the basis functions.

The timestepping procedure central to the spectral Lagrange-Galerkin method (e.g. (2.47)) involves, at each time step, the evaluation of $E(t^m;t^{m-1})U^{m-1}(x_\nu) = U^{m-1}(X_\nu)$ at the $(2L)^n$ quadrature points x_ν, where $U^{m-1} \in S_N$ and $X_\nu = X(x_\nu, t^m; t^{m-1})$. Unfortunately, the points X_ν are not in general evenly spaced, and so the FFT cannot be used in these evaluations. Also, when the evolution operator $E(t^m;t^{m-1})$ is applied to the basis functions, they lose their tensor-product properties, so that in general, the operation count for the evaluation of $E(t^m;t^{m-1})U^{m-1}(x_\nu)$ is in fact $O(N^{2n})$. This is a severe disadvantage, especially in more than one dimension.

3.1 Local interpolation

In this section we describe and analyse an interpolation process which is designed to enable the efficient and accurate evaluation of a trigonometric polynomial at a randomly distributed set of points. The method involves first setting down an equally-spaced grid of points and, in regions surrounding those points, replacing the trigonometric polynomial by a local Chebyshev polynomial interpolant, with the degree of the interpolating polynomial identical in each region. The error incurred by the use of this local interpolation is shown to decay faster than factorially with the degree of the interpolating polynomial. Thus, when the process is applied to U^{m-1} at each time step, it gives rise to a method which retains the stability and convergence properties described in the previous section. We also show that, especially for problems where a high degree of resolution is required, this modification leads to substantial savings in time with effectively no loss of accuracy.

Let $V \in S_N$. Then V may be expanded in the form

$$V = \sum_{p \in Z^n} \hat{V}(p)\Phi_p,$$

where $\hat{V}(p) = 0$ if $|p|_\infty > N$. Let $\Gamma_{2L-1} = \{0, \ldots, 2L-1\}^n$, and let $\mathcal{G} = \{x_\nu\}_{\nu \in \Gamma_{2L-1}}$ be an equally-spaced grid of points indexed by Γ_{2L-1}, where $x_\nu = (x^1_{\nu_1}, \ldots, x^n_{\nu_n})$, and $x^i_{\nu_i} = \frac{\nu_i \pi}{2L}$, $i = 1, \ldots, n$. Let $\mathcal{X} = \{X_\nu\}_{\nu \in \Gamma_{2L-1}}$ be some other set of points (also indexed by Γ_{2L-1}) which we may take to be a subset of Ω by the periodicity of V. A typical distribution of these two sets of points in two dimensions is given in Figure 3.1.

Let $R^n_{2\pi}$ be the vector space of equivalence classes of points in R^n, where two such points are regarded as equivalent if they are separated by an integer multiple of 2π in each coordinate direction. In the sequel we shall speak freely of points in R^n (and often Ω), but we shall regard them as representatives of their particular equivalence classes. We can define a metric d on $R^n_{2\pi}$ by

$$d(x,y) = \max_i |x^i - y^i|_{mod 2\pi}. \tag{3.1}$$

Then, for $x \in \Omega$ we denote by $\lfloor x \rceil$ the nearest grid point to x with respect to d. Let

$$b = \max_{\nu \in \Gamma_{2L-1}} d(\lfloor X_\nu \rceil, X_\nu). \tag{3.2}$$

We note that (3.2) implies that $b \leq \frac{\pi}{2L}$.

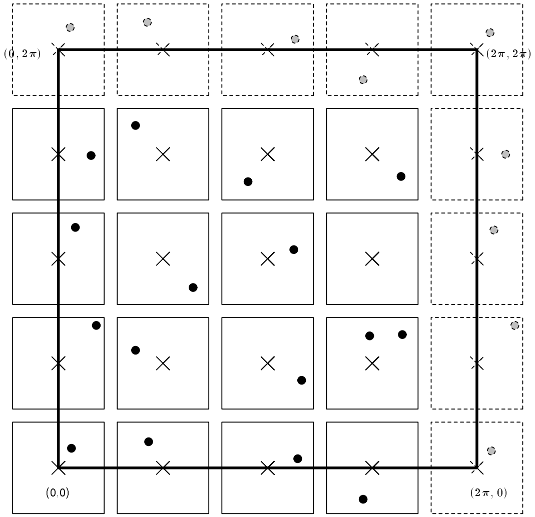

Figure 3.1: A typical distribution of points X_ν (marked •) in Ω, together with the grid points x_ν (marked by ×) and the sets $\mathcal{B}_\nu$ (denoted by) surrounding the grid points.

We define, for $\nu \in \Gamma_{2L-1}$,

$$\mathcal{B}_\nu = \{x \in R^n | d(x, x_\nu) \leq b\},$$

and

$$\mathcal{B} = \cup_{\nu \in \Gamma_{2L-1}} \mathcal{B}_\nu,$$

so that $\mathcal{X} \subset \mathcal{B}$. Thus, for our interpolation procedure to yield an accurate approximation to $\{V(X_\nu)\}_{\nu \in \Gamma_{2L-1}}$, we are mainly concerned that it should be accurate in $\mathcal{B}$. Our strategy is to replace V in each component $\mathcal{B}_\nu$ by a finite sum of (tensor product) Chebyshev polynomials. This is achieved by laying down an auxiliary grid of interpolation points on $\mathcal{B}$, identical in each component $\mathcal{B}_\nu$. Once the coefficients of the polynomials in each component are known, the value of V at any point in $\mathcal{X}$ may be calculated from the polynomial in the component of $\mathcal{B}$ in which it lies. In practice we exploit the fact that the sets $\mathcal{B}_\nu$ are centered around the points of $\mathcal{G}$ (which are equally-spaced) to calculate the coefficients of the polynomials in each of the sets $\mathcal{B}_\nu$ simultaneously. This computation takes advantage of the FFT and may be performed in $O(L^n \log L)$ operations.

In order to meaningfully speak of a polynomial defined on $\mathcal{B}_\nu$ we shall have to identify $\mathcal{B}_\nu$ with a particular subset of R^n which shall be the canonical representation of $\mathcal{B}_\nu$. As indicated in Figure 3.1, this shall be the set of points in R^n whose distance in the ordinary l^∞-norm from the grid point $x_\nu \in \Omega$ is not greater than b.

Let K be a positive integer (K will be the degree of the local interpolating polynomials). Let $y_j = \cos\frac{j\pi}{K}$, $j = 0, \ldots, K$. We define, for $p = -N, \ldots, N$ and $0 < k \leq K$,

$$\hat{a}_{p,k} = \frac{2}{K} \sum_{j=0}^{K}{}'' \phi_p(by_j) T_k(y_j),$$

where $T_k(y) = \cos(k \cos^{-1} y)$ ($y \in [-1,1]$) is the k^{th} Chebyshev polynomial on $[-1,1]$, and the double prime denotes the fact that the first and last terms in the summation are halved; for $k = 0$, $\hat{a}_{p,0}$ is defined by

$$\hat{a}_{p,0} = \frac{1}{K} \sum_{j=0}^{K}{}'' \phi_p(by_j),$$

and for $k = K$, $\hat{a}_{p,K}$ is defined by

$$\hat{a}_{p,K} = \frac{1}{K} \sum_{j=0}^{K}{}'' (-1)^j \phi_p(by_j).$$

With these definitions, $\sum_{k=0}^{K} \hat{a}_{p,k} T_k(\frac{y}{b})$ interpolates $\phi_p(y)$ at the points $\{by_j\}_{j=0}^{K}$ in $[-b, b]$. For $x \in \Omega$ define

$$\tilde{\phi}_p(x^i) = \begin{cases} \sum_{k=0}^{K} \hat{a}_{p,k} T_k(\frac{x^i - \lfloor x \rfloor^i}{b}) \phi_p(\lfloor x \rfloor^i) & x \in \mathcal{B} \\ \phi_p(x^i) & \text{otherwise;} \end{cases}$$

$\tilde{\phi}_p(x^i)$ agrees with $\phi_p(x^i)$ when $x^i = x^i_\nu + by_j$, for any $\nu \in \Gamma_{2L-1}$ and any $j = 0, \ldots, K$. Then, for $p \in Z^n$ we put $\tilde{\Phi}_p = \prod_{i=1}^{n} \tilde{\phi}_{p_i}$ and we find that $\sum_{p \in Z^n} \hat{V}(p) \tilde{\Phi}_p$ interpolates V on the set $\cup_{\nu \in \Gamma_{2L-1}} \cup_{\mu \in \Gamma_K} \{x_\nu + (by_{\mu_1}, \ldots, by_{\mu_n})\}$ (the auxiliary grid). We define the local interpolation operator $\mathcal{P}_{K,L}$ (acting on functions in S_N) by

$$\mathcal{P}_{K,L} V = \sum_{p \in Z^n} \hat{V}(p) \tilde{\Phi}_p \tag{3.3}$$

3.2 The algorithm

We want to evaluate $\mathcal{P}_{K,L} V$ at the points X_ν of $\mathcal{X}$. Expanding $\mathcal{P}_{K,L} V(X_\nu)$ we have

$$\begin{aligned} & \mathcal{P}_{K,L} V(X_\nu) \\ &= \sum_{p \in Z^n} \hat{V}(p) \sum_{k \in \Gamma_K} \prod_{i=1}^{n} \left(\hat{a}_{p_i,k_i} T_{k_i} \left(\frac{X^i_\nu - \lfloor X_\nu \rfloor^i}{b} \right) \right) \Phi_p(\lfloor X_\nu \rfloor) \\ &= \sum_{k \in \Gamma_K} \left[\prod_{i=1}^{n} T_{k_i} \left(\frac{X^i_\nu - \lfloor X_\nu \rfloor^i}{b} \right) \right] \left[\sum_{p \in Z^n} \hat{V}(p) \prod_{i=1}^{n} \hat{a}_{p_i,k_i} \Phi_p(\lfloor X_\nu \rfloor) \right]. \end{aligned} \tag{3.4}$$

(3.4) indicates how the calculation may be performed inexpensively. For each $k \in \Gamma_K$ we have a term of the form

$$\sum_p \hat{V}(p) f_k(X_\nu - \lfloor X_\nu \rfloor) g_k(p) \Phi_p(\lfloor X_\nu \rfloor),$$

and the algorithm is as follows:

- multiply the $\hat{V}(p)$ by the $g_k(p)$,
- evaluate $W(x_\nu) = \sum_p \hat{V}(p) g(p) \Phi_p(x_\nu)$ by the use of a FFT,
- multiply $W(\lfloor X_\nu \rfloor)$ by $f_k(X_\nu - \lfloor X_\nu \rfloor)$.

The total number of operations for one loop is $O(L^n \log L)$ (assuming that $L \geq N$), so that the operation count for evaluating $\{\mathcal{P}_{K,L} V(X_\nu)\}_{\nu \in \Gamma_{2L-1}}$ is now $O(K^n L^n \log L)$. We shall show, both by means of theoretical estimates and numerical experiments, that this represents a significant saving over the cost of a full evaluation ($O(L^{2n})$).

We note here that the local interpolation described above is not the only means of obtaining an approximation to $\{V(X_\nu)\}_{\nu \in \Gamma_{2L-1}}$ that may be implemented by the use of this algorithm. The relevant property of the approximation of $\Phi_p(X_\nu - \lfloor X_\nu \rfloor)$ by $\sum_k f_k(X_\nu - \lfloor X_\nu \rfloor) g_k(p)$ is that the dependences on p and on ν have been *decoupled*. Any approximation with this property will provide an efficient procedure, but one based on the use of Chebyshev polynomials will in general be amongst the most accurate [34].

The incorporation of this local interpolation into the timestepping procedure of the previous chapter results in the following numerical method. Find $U = \{U^m\}_{m=1}^{M} \subset S_N$ such that

$$U^m = \Pi_N P_L E(t^m; t^{m-1}) \mathcal{P}_{K,L} U^{m-1}, \quad m = 1, \ldots, M \tag{3.5a}$$

$$U^0 = \Pi_N P_L u_0. \tag{3.5b}$$

Here, the points X_ν in $\mathcal{X}$ used to define $\mathcal{P}_{K,L}$ are given by $X_\nu = X(x_\nu, t^m; t^{m-1})$, with $\nu \in \Gamma_{2L-1}$. Calculation of U^m thus involves evaluating $\mathcal{P}_{K,L} U^{m-1}$ on the set $\{X_\nu\}_{\nu \in \Gamma_{2L-1}}$. These X_ν are *not* randomly distributed and in fact, for small Courant numbers ($|a|_{0,\infty} L \Delta t$), they will be grouped around their respective grid points. Also, when a is smooth, there will be a corresponding smooth variation between neighbouring points in $\mathcal{X}$. In the next section we shall give an error analysis for the local interpolation as it used in (3.5) which exploits these observations.

3.3 Error analysis

The analysis rests on the standard error estimate for Lagrange interpolation of a smooth function. Most of the subsequent complication arises when (in more than one dimension) (3.4) is expanded further, resulting in terms where the local interpolation has been carried out in some coordinate directions but not in others. We shall describe the one-dimensional case first.

For $-N \leq p \leq N$, let $\theta_p = \tilde{\phi}_p - \phi_p$. Let $\nu \in \Gamma_{2L-1} = \{0, \ldots, 2L-1\}$, and consider $\theta_p(x)$ for $x \in \mathcal{B}_\nu$. We can write such an x as $x_\nu + by$ for some $y \in [-1, 1]$. Then

$$\theta_p(x) = \phi_p(x_\nu) (\sum_{k=0}^{K} \hat{a}_{p,k} T_k(y) - \phi_p(by)).$$

When x is not in $\mathcal{B}$, $\theta_p(x) = 0$, so that, from the standard error estimate for Lagrange interpolation,

$$|\theta_p|_\infty = \max_{y \in [-1,1]} |\sum_{k=0}^{K} \hat{a}_{p,k} T_k(y) - \phi_p(by)| \leq \frac{|pb|^{K+1}}{2^K (K+1)!}. \tag{3.6}$$

Now for $V \in S_N$ we can write

$$(\mathcal{P}_{K,L} - I) V = \sum_{p \in Z^n} \hat{V}(p) \theta_p,$$

so that (3.6) implies that

$$\begin{aligned} \|(\mathcal{P}_{K,L} - I) V\|_\infty &\leq \frac{b^{K+1}}{2^K (K+1)!} \sum_{p \in Z^n} |\hat{V}(p)| |p|^{K+1} \\ &\leq |V|_{K+1} \frac{b^{K+1} \surd(2N)}{2^K (K+1)!}. \end{aligned} \tag{3.7}$$

So far this concerns the local interpolation, without reference to the set of points at which V is to be evaluated. For any particular set of points $\{X_\nu\}$, we shall be able to say more about b. For instance, we note that here b satisfies

$$b \leq \min\{\frac{\pi}{2L}, |a|_{0,1;[t^{m-1},t^m]}\} \tag{3.8}$$

$$\leq \min\{\frac{\pi}{2L}, |a|_{0,\infty} \Delta t\}, \tag{3.9}$$

so that we have

$$\begin{aligned} & \|P_L E(t^m; t^{m-1}) (\mathcal{P}_{K,L} - I) V\| \\ &= \left(\frac{1}{2L} \sum_{\nu=0}^{2L-1} |(\mathcal{P}_{K,L} - I) V(X_\nu)|^2 \right)^{\frac{1}{2}} \\ &\leq \max_{\nu \in \Gamma_{2L-1}} |(\mathcal{P}_{K,L} - I) V(X_\nu)| \\ &\leq C_6 |a|_{0,1;[t^{m-1},t^m]} |V|_{K+1} \left(\frac{\pi}{4L} \right)^K \frac{\surd(2N)}{(K+1)!} \end{aligned} \tag{3.10}$$

In the one-dimensional case, the constant C_6 appearing in (3.10) may be taken to be unity. In general dimensions, however, it depends on C_5 , $|a|_{[\sigma]^*,1}$, n and Δt_0, where C_5 is as in Lemma 2.7 and Δt_0 is an upper bound on Δt. In order to obtain a stability result analogous to Theorem 2.5 we let $0 \leq \rho \leq \min(K+1, \sigma)$ and we have

$$\begin{aligned} \left\| \Pi_N P_L E(t^m; t^{m-1}) (\mathcal{P}_{K,L} - I) V \right\|_\rho &\leq (1 + nN^2)^{\rho/2} \left\| \Pi_N P_L E(t^m; t^{m-1}) (\mathcal{P}_{K,L} - I) V \right\| \\ &\leq C_7 |a|_{0,1;[t^{m-1},t^m]} \left\| V \right\|_\rho \left(\frac{N\pi}{4L} \right)^K \frac{N^{\frac{3}{2}}}{(K+1)!}, \end{aligned} \tag{3.11}$$

where C_7 is just a multiple of C_6.

We seek to bound the right hand sides of (3.11) independently of K or N, and so we require that K and N are related by, for example,

$$N^{\frac{3}{2}} \leq \left(\frac{4L}{\pi N} \right)^K (K+1)!. \tag{3.12}$$

In order to illustrate the relationship between N and K that (3.12) implies, we consider sample values of K and N satisfying (3.12), shown in Table 3.1. Since the saving in time resulting from the use of the procedures described in this section is of the order of N/K, we see that this can be quite substantial. The numerical experiments in the next section will show how substantial these savings are in practice. It is perhaps worth remarking that a relationship of the form (3.12) implies that $K = O(\log N)$ so that the asymptotic operations count is $O(L^n (\log L)^{n+1})$.

K	N
1	1
2	4
3	13
4	46
5	179
6	772
7	3630
8	18451
9	100608
10	584571

Table 3.1: Sample values of K and N.

Making use of (3.11) we can ensure the stability of (3.5) (under the conditions of Theorem 2.5, together with (3.12)) as follows:

$$\begin{aligned}\|U^m\|_\rho &= \left\|\Pi_N P_L E(t^m;t^{m-1})\mathcal{P}_{K,L}U^{m-1}\right\|_\rho\\ &\leq \left\|\Pi_N P_L E(t^m;t^{m-1})U^{m-1}\right\|_\rho + \left\|\Pi_N P_L E(t^m;t^{m-1})(\mathcal{P}_{K,L}-I)U^{m-1}\right\|_\rho\\ &\leq e^{C_2(t^{m-1},t^m)+C_5|a|_{[\sigma]^*,1;[t^{m-1},t^m]}}\left\|U^{m-1}\right\|_\rho\\ &\quad +C_7|a|_{0,1;[t^{m-1},t^m]}\left\|U^{m-1}\right\|_\rho .\end{aligned}$$

As in the proof of Theorem 2.5 we deduce that

$$\|U^m\|_\rho \leq e^{(C_2+C_5|a|_{[\sigma]^*,1}+C_7|a|_{0,1})}\left\|U^0\right\|_\rho, \quad 0\leq m\leq M. \tag{3.13}$$

Convergence also follows, analogously to Theorem 2.6. We find that (2.71) is replaced by

$$\begin{aligned}\xi^m \;=\;& \Pi_N P_L E(t^m;t^{m-1})\mathcal{P}_{K,L}\xi^{m-1} + \Pi_N E(t^m;t^{m-1})\eta^{m-1}\\ & -\Pi_N(P_L-I)E(t^m;t^{m-1})\Pi_N u^{m-1}\\ & -\Pi_N P_L E(t^m;t^{m-1})(\mathcal{P}_{K,L}-I)\Pi_N u^{m-1}.\end{aligned}\tag{3.14}$$

Making use of the above stability result, the first three terms are dealt with in the same way as the corresponding terms in (2.71). By (3.10) and (3.12) we may bound the final term by

$$C_6|a|_{0,1;[t^{m-1},t^m]}N^{-(K+1)}|u^{m-1}|_{K+1}.$$

As in the proof of Theorem 2.6, this, when included in with the other terms, yields the following convergence result.

$$\begin{aligned}&\|u-U\|_{l^\infty(L^2_\#(\Omega))} \leq\\ &\quad e^{C_2+C_5|a|_{[\sigma]^*,1}+C_7|a|_{0,1}}N^{-\sigma'}\|u_0\|_\sigma \times\\ &\quad \times\left[1+C_{n,\sigma}+\min(\frac{T}{\Delta t},n^{\frac{1}{2}}N|a|_{0,1})+C_5|a|_{[\sigma]^*,1}+C_6|a|_{0,1}\right],\end{aligned}\tag{3.15}$$

where $\sigma' = \min\{1+K,\sigma\}$.

We have now proved, for the one-dimensional case, the following stability and convergence result for the method (3.5).

Theorem 3.1 *Suppose that the conditions of Theorem 2.5 and Theorem 2.6 hold, together with the condition (3.12). Then (3.13) and (3.15) hold.*

Proof All that remains to be proved is that (3.10) and (3.11) hold in general dimensions. We define, for $p\in Z^n$, $x\in R^n$,

$$\Theta_p(x) = \prod_{i:p_i\neq 0}\theta_{p_i}(x^i).$$

Let $\lambda\in\Lambda^n$. We note that $Z^n = Z^n_\lambda \oplus Z^n_{1-\lambda}$ where Z^n_λ is as defined by

$$Z^n_\lambda = \{k\in Z^n | k_i = 0 \Leftrightarrow \lambda_i = 0, i=1,\ldots,n\}; \tag{3.16}$$

then, for $V\in S_N$, $(\mathcal{P}_{K,L}-I)V$ may be expanded (c.w. (2.50))

$$(\mathcal{P}_{K,L}-I)V = \sum_{\lambda\in\Lambda^n}\sum_{q\in Z^n_\lambda}\left(\sum_{p\in Z^n_{1-\lambda}}\hat{V}(p+q)\Phi_p\right)\Theta_q. \tag{3.17}$$

We further define

$$V_{\lambda,q} = \sum_{p\in Z^n_{1-\lambda}}\hat{V}(p+q)\Phi_p,$$

and note that $V_{\lambda,q}\in S_N$. Then

$$\begin{aligned}&\left\|P_L E(t^m;t^{m-1})(\mathcal{P}_{K,L}-I)V\right\|\\ &\quad = \left\|P_L E(t^m;t^{m-1})\sum_{\lambda\in\Lambda^n}\sum_{q\in Z^n_\lambda}V_{\lambda,q}\Theta_q\right\|\\ &\quad \leq \sum_{\lambda\in\Lambda^n}\sum_{q\in Z^n_\lambda}\left\|P_L E(t^m;t^{m-1})V_{\lambda,q}\Theta_q\right\|\\ &\quad \leq \sum_{\lambda\in\Lambda^n}\sum_{q\in Z^n_\lambda}\|\Theta_q\|_{C_\#(\overline{\Omega})^n}\left\|P_L E(t^m;t^{m-1})V_{\lambda,q}\right\|.\end{aligned}\tag{3.18}$$

Under the conditions of the theorem,

$$\left\|P_L E(t^m;t^{m-1})V_{\lambda,q}\right\| \leq (1+C_5|a|_{[\sigma]^*,1;[t^{m-1},t^m]})\left\|V_{\lambda,q}\right\|. \tag{3.19}$$

Moreover

$$\|V_{\lambda,q}\|^2 = \sum_{p\in Z^n_{1-\lambda}}|\hat{V}(p+q)|^2, \tag{3.20}$$

and, for $q\in Z^n_\lambda$,

$$\|\Theta_q\|_{C_\#(\overline{\Omega})^n} \leq \prod_{i=1}^n\left(\frac{|q_i b|^{K+1}}{2^K(K+1)!}\right)^{\lambda_i}. \tag{3.21}$$

By the Cauchy-Schwarz inequality, (3.20) and (3.21) imply that

$$\begin{aligned}&\sum_{q\in Z^n_\lambda}\|\Theta_q\|_{C_\#(\overline{\Omega})^n}\|V_{\lambda,q}\|\\ &\quad\leq \left(\frac{b^{K+1}\surd(2N)}{2^K(K+1)!}\right)^{|\lambda|_1}\left(\sum_{p\in Z^n_{1-\lambda}}\sum_{q\in Z^n_\lambda}\prod_{i=1}^n|q_i|^{2(K+1)\lambda_i}|\hat{V}(p+q)|^2\right)^{\frac{1}{2}}\\ &\quad\leq \left(\frac{b^{K+1}\surd(2N)}{2^K(K+1)!}\right)^{|\lambda|_1}N^{(K+1)(|\lambda|_1-1)}|V|_{K+1}\\ &\quad= \left(\frac{(Nb)^{K+1}\surd(2N)}{2^K(K+1)!}\right)^{|\lambda|_1}\frac{|V|_{K+1}}{N^{K+1}}.\end{aligned}$$

Then, from (3.18) and (3.19),

$$\begin{aligned}&\left\|\Pi_N P_L E(t^m;t^{m-1})(\mathcal{P}_{K,L}-I)V\right\|\\ &\quad\leq \frac{|V|_{K+1}}{N^{K+1}}(1+C_5|a|_{[\sigma]^*,1})\left[\left(1+\frac{(Nb)^{K+1}\surd(2N)}{2^K(K+1)!}\right)^n-1\right]\end{aligned}$$

and (3.10) follows, with (3.11) a straightforward consequence.

3.4 Numerical experiments

In this section we perform various numerical experiments in order to illustrate the theoretical predictions of the previous two chapters. We begin by considering the performance of the method on two one-dimensional advection equations.

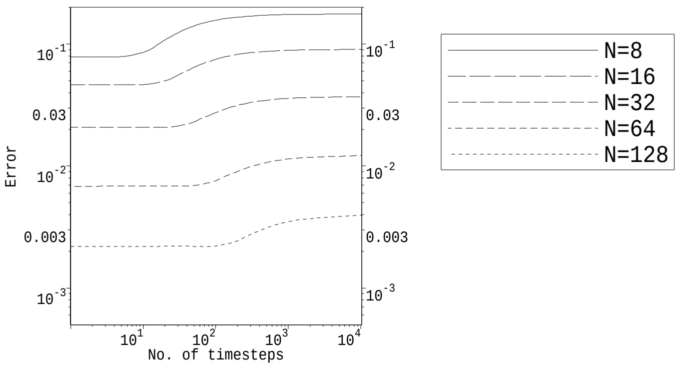

Figure 3.2: Results from Experiment 1.

3.4.1 Experiment 1

In this section our model problem is the following.

$$u_t - (\sin x)u_x = 0, \quad x \in [0, 2\pi), \quad t \in (0, 1.571],$$

with initial condition

$$u(x, 0) = f(x),$$

for which the exact solution is

$$u(x,t) = f\left(2\tan^{-1}(e^t \tan\frac{x}{2})\right).$$

The advection is 'pushing' the solution out towards the boundaries of the domain, and so in general the solution tends to develop sharp layers at $x = 0$ and $x = 2\pi$.

We consider the case

$$f(x) = \begin{cases} \sin x & 0 \le x \le \pi \\ 0 & \pi \le x \le 2\pi. \end{cases}$$

This enables us to illustrate the behaviour of the method (2.35) as Δt is reduced when the initial data and the velocity field a have only finite smoothness. Figure 3.2 shows the L^2 error for various values of N and Δt when (2.35) was used to integrate the equation up to time $t = 3.14$. (In fact we had to use the scheme (2.47), but we took $L = 2048$ to effectively eliminate the quadrature error).

As predicted by Theorem 2.4, for each fixed N the error increases up to a maximum value as more time steps are taken. In Figure 3.2 this increase does not show itself until $N\Delta t \approx 1.5$. Until that point it is overshadowed by the contribution to the error of the projection of the initial data. Although this agrees with Theorem 2.4 in a qualitative sense, the predicted reduction in the order of convergence for the time-converged solution is *not* evident in this experiment. A more careful analysis might be able to reveal the balance between the various contributions to the error more clearly.

3.4.2 Experiment 2

Here we consider the scalar hyperbolic equation

$$u_t - (2 + \cos x)^{-1} u_x = 0, \quad x \in [0, 2\pi), \quad t \in (0, 50.27],$$

subject to periodic boundary conditions and the initial condition

$$u(x, 0) = \sin(2x + \sin x).$$

The exact solution is

$$u(x, t) = \sin(2x + \sin x + t).$$

Tal-Ezer has also studied this problem [83]. He uses a spectral method in time, based on a Chebyshev polynomial expansion of the evolution operator. His theory predicts that in order to effectively resolve the behavior of the solution in time, the degree of the polynomial needs to be proportional to both the time over which he is integrating, and to N, where S_N is the space in which the problem is being approximated. In order to demonstrate his theoretical predictions, he integrates the above problem up to time 50.27. To achieve machine accuracy ($\approx 10^{-13}$) he requires 840 polynomials, with N=16.

K	$N = 3$, $L = 4$	$N = 7$, $L = 8$	$N = 15$, $L = 16$	$N = 31$, $L = 32$
4	1.880×10^{-1}	4.457×10^{-5}	8.560×10^{-7}	4.597×10^{-8}
6	1.882×10^{-1}	3.209×10^{-5}	8.620×10^{-10}	8.672×10^{-12}
8	1.882×10^{-1}	3.209×10^{-5}	4.277×10^{-13}	1.629×10^{-15}
10	1.882×10^{-1}	3.209×10^{-5}	1.100×10^{-15}	3.953×10^{-16}

Table 3.2: $l^\infty(L^2)$ errors for Experiment 2.

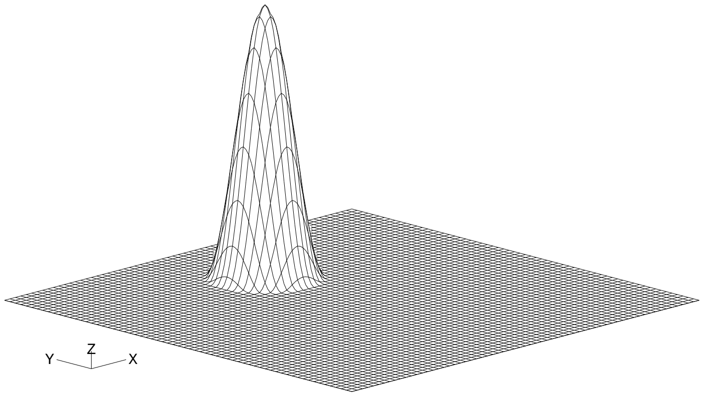

Figure 3.3: Initial data for the rotating cone problem ($p = 1$).

For comparison we also integrated the above problem up to time 50.27 ($\approx 16\pi$). Because of the fact that the exact solution is 2π-periodic in space *and* in time, integration to this final time in just one time step using the spectral method of characteristics (with the additional interpolation) is almost a trivial problem. We therefore used three time steps. The feet of the characteristics were found to machine accuracy by repeated bisection. The results obtained for various values of K, N and L are shown in Table 3.2. We note that for $N = 15$ and $L = 16$, the $l^\infty(L^2)$ error in the computed solution was down to $\approx 10^{-15}$ with the use of the additional interpolation $\mathcal{P}_{K,L}$ with K=10, so that the amount of work required was a fraction of that for the time-integration scheme of Tal-Ezer. The results also illustrate the high accuracy of the spectral approximation for such smooth problems.

3.4.3 Experiment 3

This test problem, known as the rotating cone problem, is the one in which Priestley (see [76]) first observed the instability to which the finite-element version of the method is prey when quadrature is introduced. For our purposes it also serves to illustrate the theoretically predicted convergence rates for initial datum that is not infinitely smooth, and to illustrate the savings arising from use of the approximate evolution operator.

The governing equation is

$$u_t + a(x) \cdot \nabla u = 0, \quad x \in \Omega', \quad t \in (0, 20],$$

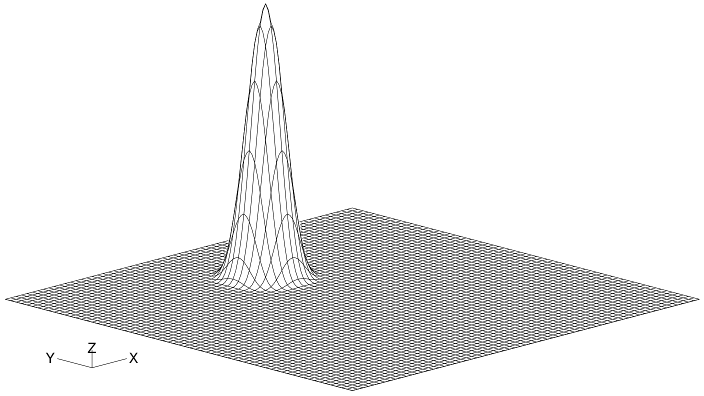

Figure 3.4: Initial data for the rotating cone problem $(p = 2)$.

method	N/L	U_{max}	U_{min}	$l^\infty(L^2)$ error
exactly-integrated	$7/8$	0.9917	-4.112×10^{-2}	6.840×10^{-3}
quadrature-based		0.9847	-2.802×10^{-2}	6.827×10^{-3}
approx. $(K = 5)$		0.9853	-2.829×10^{-2}	6.842×10^{-3}
exactly-integrated	$15/16$	1.0001	-6.544×10^{-3}	9.598×10^{-4}
quadrature-based		0.9991	-4.138×10^{-3}	7.468×10^{-4}
approx. $(K = 5)$		0.9992	-5.145×10^{-3}	9.081×10^{-4}

Table 3.3: Rotating Cone Problem $(p = 1)$

subject to periodic boundary conditions and the initial condition

$$u(x, 0) = \begin{cases} \cos^{2p}(2|x - x_0|), & |x - x_0| \le \pi/4, \\ \\ 0, & \text{otherwise}, \end{cases}$$

where $p \in N$, $a(x) = 2\pi(x_2, -x_1)$, $x_0 = (0, \pi/2)$ and $\Omega' = (-\pi, \pi)^2$. The initial datum is shown, for $p = 1$ and $p = 2$, in Figs. 3.3 and 3.4. Again the characteristics may be calculated explicitly. Moreover, because of the linearity of the velocity field it is possible to perform the inner products involved in the scheme (2.35) exactly. Thus we can observe the effect of quadrature. The solutions obtained were visibly indistinguishable from the initial data, so we illustrate the performance of the schemes by tabulating the errors. In Table 3.3, the results show the $l^\infty(L^2)$ error after five revolutions $(t = 5)$ with 250 time steps, for the exactly integrated scheme (2.35), the quadrature-based scheme (2.46), and for the scheme (3.5) involving the additional interpolation. The value of K given here is the minimum required to reproduce the results obtained by the scheme (2.46).

The parameter p in the description of the problem enables us to control the smoothness of the initial datum. We find that $u_0 \in H_{\#}^{2p+1/2-\epsilon}(\Omega')$ for any $\epsilon > 0$. Thus we can check the theoretically predicted orders of convergence for the method. Results for $p = 1$ and $p = 2$ are given in Table 3.4 for the method (2.46), together with the orders of convergence represented by the results. 25 time steps were performed to reach a final time of $t = 0.5$. The results for $N/L = 63/64$ were obtained with the approximate method (3.5) with $K = 5$.

These calculations were performed on a Convex C1/XP, which is a vector machine. On this machine the scheme (3.5) was approximately seven times as fast per time step as the scheme (2.46) for $N = 63$, $L = 64$ and $K = 5$. On the VAX 11/785, a scalar machine, this difference is increased from a factor of seven to a factor of fifty. The reason for this disparity is that the FFT routine used in the approximate trajectories scheme does not take full advantage of the vectorizing capabilities of the Convex machine.

p	N/L	U_{max}	U_{min}	$l^\infty(L^2)$ error	order
1	$7/8$	0.9996	-2.1×10^{-2}	5.112×10^{-3}	
	$15/16$	1.0008	-4.9×10^{-3}	8.172×10^{-4}	2.65
	$31/32$	1.0002	-1.4×10^{-3}	1.448×10^{-4}	2.50
	$63/64$	1.00005	-4.4×10^{-4}	2.660×10^{-5}	2.44
2	$7/8$	0.8676	-6.0×10^{-2}	1.432×10^{-2}	
	$15/16$	0.9998	-7.6×10^{-4}	1.385×10^{-4}	6.69
	$31/32$	$.99999$	-3.7×10^{-5}	5.185×10^{-6}	4.74
	$63/64$	$.999999$	-2.5×10^{-6}	2.301×10^{-7}	4.49

Table 3.4: Rotating Cone Problem

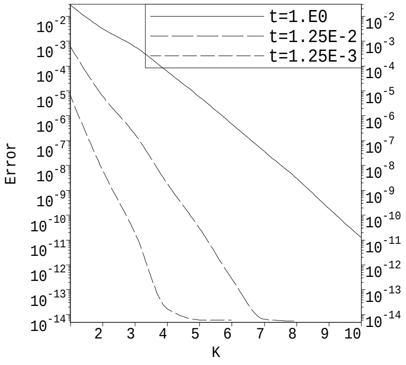

Figure 3.5: Results from Experiment 4.

K	c.p.u. time	speedup
1	2.0s	630
2	3.4s	370
3	5.2s	242
4	7.5s	168
5	10.3s	122
6	13.4s	94
7	17.1s	73
8	21.1s	59
9	25.7s	49
10	30.0s	42

Table 3.5: Example c.p.u. times for Experiment 4.

3.4.4 Experiment 4

Our final experiment in this chapter is a more thorough investigation of the behaviour of the additional interpolation than was afforded by the rotating cone problem. We shall consider the problem of the evaluation of a two-dimensional trigonometric polynomial at set of points randomly distributed in $(0, 2\pi)^2$. The polynomial U will be a function from S_N, with the value of N being 32, and we take 64^2 points. We choose U to be the projection of the vorticity field obtained at the final time level from the Navier-Stokes calculation performed at the end of the final chapter. We shall actually perform two experiments here. In the first, we shall first lay down an equally-spaced grid of 64^2 points on $(0, 2\pi)^2$, and then, to obtain the points at which we are to evaluate U, we allow each grid point to move to a random position within a ball of radius t (in the l^∞-norm) of its original position. Then we allow the degree (K) of the local interpolation to increase, and measure the maximum error. The results are illustrated, for various values of t, in Figure 3.5. Note that the spacing between the points of the original equally-spaced grid is $\approx 9.8 \times 10^{-2}$, so that setting $t = 1$ means that the final set of points is randomly distributed throughout the domain. In Table 3.5 we illustrate the c.p.u. times corresponding to various values of K, and compare these to the c.p.u. time (1260s) taken by a full evaluation of U at the random points. The savings are substantial, and are even more dramatic at higher spatial resolutions (i.e. higher values of N and L). The exponential accuracy of the interpolation is evident in these results. Moreover, in the context of the spectral Lagrange-Galerkin method, Fig. 3.5 indicates that the smaller the value of Δt, the less work necessary for each time step.

Our second experiment examines a method (implemented in the Navier-Stokes calculations of Chapter 5) for avoiding unnecessary work in the evaluation of U. We notice that in the expansion (3.4) the Chebyshev coefficients of $\phi_{p_i}(by)$ appear in a product. For high values of k_i, the coefficient $\hat{a}_{p_i,k_i}$ will be small. When two or more coefficients each corresponding to high values of the respective component of k are multiplied together, the result will be negligibly small. Yet the calculation of the term corresponding to this outlying value of k takes just as much work as the calculation of the term corresponding to $k = (0, \ldots, 0)$ for example, which will be of the same order of magnitude as the entire function. In order to avoid wasting time on such calculations, we set a tolerance and employ an upper bound of the size of the contribution arising from a particular value of k, and exclude those values of k from our calculation for which this measure falls below the set tolerance.

The simplest bound on the size of the contribution arising from the k-term on the right hand side of (3.4) is given by

$$\max_{\nu} \left| \left[\prod_{i=1}^{n} T_{k_i} \left(\frac{X_\nu^i - \lfloor X_\nu \rfloor^i}{b} \right) \right] \left[\sum_{p \in Z^n} \hat{V}(p) \prod_{i=1}^{n} \hat{a}_{p_i,k_i} \Phi_p(\lfloor X_\nu \rfloor) \right] \right|$$

$$\leq \sum_{p \in Z^n} \left| \hat{V}(p) \prod_{i=1}^{n} \hat{a}_{p_i,k_i} \right| \tag{3.22}$$

$$\leq \|V\| \left(\sum_{p \in Z^n} \left| \prod_{i=1}^{n} \hat{a}_{p_i,k_i} \right|^2 \right)^{\frac{1}{2}}$$

$$= \|V\| \prod_{i=1}^{n} \left(\sum_{p_i \in Z} |\hat{a}_{p_i,k_i}|^2 \right)^{\frac{1}{2}} . \tag{3.23}$$

The expression (3.23) is easy to calculate, and provides a first check for which values of $k \in \Gamma_K$ to ignore. A tighter bound is given by (3.22). This expression is much more expensive to calculate, and yet need only be checked for those values of k that managed to sneak through the first test. This then is the procedure we have used in practice: having assigned a tolerance at the beginning of the run, we first discard those values of k for which (3.23) exceeds that tolerance, and then check the remaining values according to (3.22).

In Figures 3.6 and 3.7 the shaded areas represent those values of k which were actually used in each particular calculation. In each table the tolerance parameter is shown, together with the resulting l^∞ error. The results indicate that the tolerance parameter is a fairly reliable guide to the size of the resulting error. Table 3.5 shows how the c.p.u. time decreases with K; correspondingly, the shaded areas in the diagrams in Figs. 3.6 and 3.7 directly relate to the c.p.u. time taken by the respective calculations.

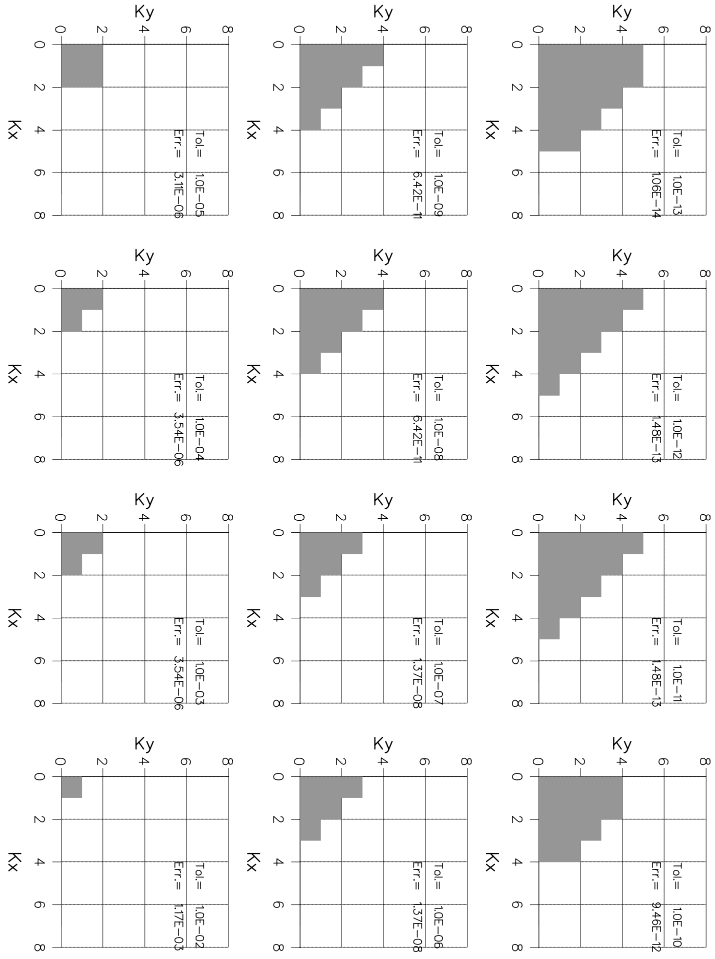

Figure 3.6: Results from Experiment 4, $\Delta t = 1/80$.

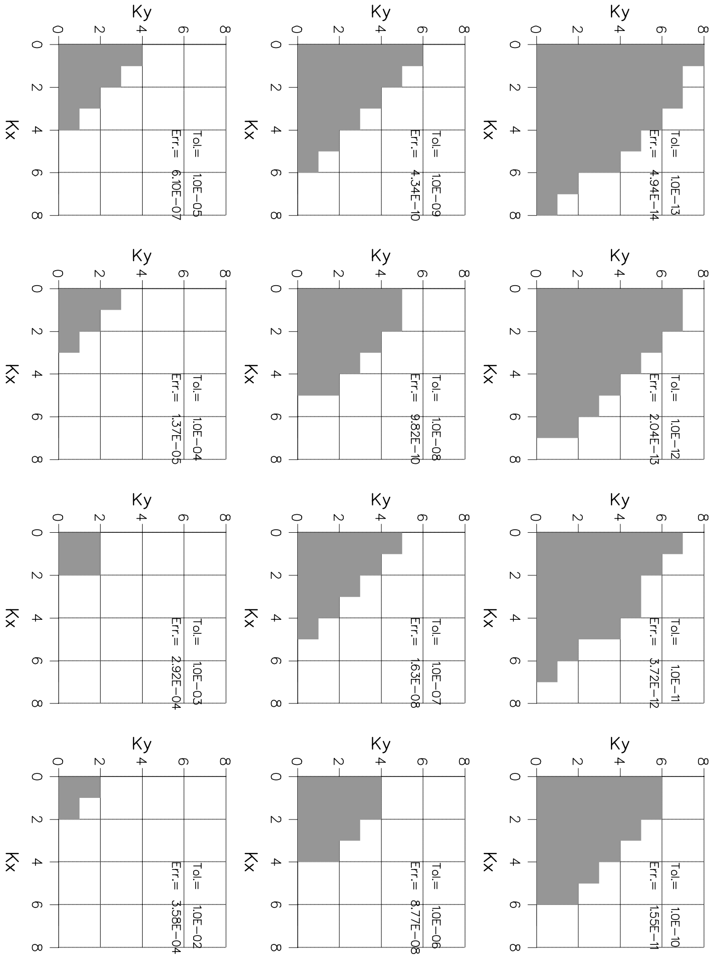

Figure 3.7: Results from Experiment 4, $\Delta t = 1/800$.

Chapter 4

Advection–diffusion equations

4.1 Introduction

In this chapter we consider the model advection–diffusion problem

$$\frac{\partial u}{\partial t} + a \cdot \nabla u - \nu \nabla^2 u = 0 \quad \text{in } \Omega \times (0, T], \tag{4.1a}$$

subject to periodic boundary conditions and the initial condition

$$u(x, 0) = u_0(x) \quad \text{in } \Omega, \tag{4.1b}$$

where $\Omega = (0, 2\pi)^n$, T a fixed final time and ν a non-negative diffusion coefficient. The initial data u_0 and the velocity field a are assumed to be periodic in the spatial variables, and we assume that $\nabla \cdot a = 0$.

The combination of the advection and diffusion phenomena modelled in this problem is characteristic of a wide variety of physical processes. In isolation the two phenomena merit quite distinct approaches towards the numerical approximation of equations in which they occur. Our interest is in cases where the advection is the dominant term, and so our numerical approximation of this problem will be based on the treatment of the advective term presented in the preceding chapters. We shall find that our discrete equations will reduce to those of Chapter 2 in the limit as $\nu \to 0$.

We shall begin, in the remainder of this section, by describing the mathematical setting of the problem upon which the analysis of this chapter is founded, and giving a preliminary discussion of a backward-Euler discretisation of this problem, based on the approach introduced in Chapter 2. In the second section a Lagrangian form of the equation is introduced, and shown to be equivalent to the original form. The third section introduces and analyses time-discrete schemes derived from the Lagrangian form, and in the final section the fully discrete schemes are discussed.

4.1.1 Mathematical setting of the problem

Suppose that u is a classical solution of (4.1). It is a straightforward consequence of the divergence theorem that if $\nabla \cdot a = 0$ then $\int_\Omega u$ is constant, so that we can assume that u has zero mean. With this assumption, we define two function spaces that will be central to our discussion of the problem. Let

$$H := H^0_\#(\Omega)/R$$

with norm denoted by $\|\cdot\|$, and

$$V := H^1_\#(\Omega)/R,$$

with the norm on V given by

$$\|u\|_V = \|\nabla u\|.$$

We define the bilinear form $\mathcal{A}(\cdot,\cdot)$ on $V \times V$ by

$$\mathcal{A}(u, v) = \nu(\nabla u, \nabla v), \quad u, v \in V.$$

By the Riesz Representation Theorem we may associate with $\mathcal{A}$ a continuous linear operator A from V to V'. For $u \in V$, Au is that element of V' which satisfies

$$\langle Au, v\rangle = \mathcal{A}(u, v) \quad \forall v \in V.$$

A is an isomorphism, and may be associated with $-\nu\nabla^2$. We define

$$D(A) = \{v \in H | Av \in H\},$$

which, with the graph norm, is a Hilbert space, isomorphic to $H^2_\#(\Omega) \cap H$.

A weak form of (4.1) is

Problem 1 *Let $a \in L^1(C^1_\#(\overline{\Omega})^n) \cap L^\infty(C_\#(\overline{\Omega})^n)$ with $\nabla \cdot a = 0$. Given $u_0 \in H$, find $u : [0, T] \to V$ satisfying (in the sense of distributions, with values in V')*

$$\begin{aligned} \frac{du}{dt} + (a \cdot \nabla)u + Au &= 0, && (4.2a) \\ u(0) &= u_0. && (4.2b) \end{aligned}$$

The following theorem expresses a standard existence and uniqueness result for Problem 1. (See for example [87], Theorem 3.1.)

Theorem 4.1 *There exists a unique solution u of Problem 1 such that*

$$\begin{aligned} u &\in L^2(V) \cap C(H), && (4.3a) \\ u' &\in L^2(V'). && (4.3b) \end{aligned}$$

Regularity results concerning the solution of Problem 1 may also be found in [87].

4.1.2 The material derivative

In line with the discretisation of the hyperbolic problem in the previous chapter, our approach is to rewrite the advection terms in (4.1a) as a directional derivative along the particle paths defined by the flow field a. This is seen to be equivalent to rewriting the equation in a Lagrangian form with the Lagrangian variables being given by the particle paths. In the hyperbolic case, this gives the solution explicitly. Here the process results in a new differential equation for which appropriate higher-order discretisations present themselves fairly readily.

We recall for completeness the definition of the Lagrangian variables. For $x \in \Omega$, $t \in [0, T]$, $X(x, t; \cdot)$ is defined to be the solution of the initial value problem

$$\begin{aligned} \frac{dX}{ds}(x, t; s) &= a(X(x, t; s), s), \quad s \in [0, T]\backslash\{t\}, && (4.4a) \\ X(x, t; t) &= x; && (4.4b) \end{aligned}$$

various properties of solutions to (4.4) are given in Theorem 2.1. The operator $E(t; s)$ is defined by its action on a function w on R^n by

$$E(t; s)w(\cdot) = w(X(\cdot, t; s)).$$

We recall that when $\nabla \cdot a = 0$, $E(t; s)$ is an isometry on H. We also have, for $t, s, \tau \in [0, T]$, by Corollary 2.1,

$$E(t; s) = E(t; \tau)E(\tau; s).$$

The material derivative $D_t w$ of a function w on $R^n \times [0, T]$ is defined, where it exists, by

$$D_t w(\cdot, t) = \frac{d}{dt} E(s; t) w(\cdot, t)\Big|_{s=t},$$

and, for smooth w, we have, by the chain rule,

$$D_t w = \frac{\partial w}{\partial t} + a \cdot \nabla w. \tag{4.5}$$

Thus for smooth solutions u, (4.1a) may be replaced by

$$D_t u + Au = 0. \tag{4.6}$$

We note that D_t is a directional derivative, the direction at (x, t) being that of the tangent to the curve $(X(x, t; s), s)$.

4.1.3 First-order timestepping—an example

A straightforward backward-Euler discretisation of (4.6) between times t^{m-1} and t^m results in

$$u^m + \Delta t A u^m = E(t^m; t^{m-1}) u^{m-1}. \tag{4.7}$$

When $u^{m-1} \in H$, $E(t^m; t^{m-1})u^{m-1} \in H$, and so, by the Lax-Milgram Theorem, (4.7) has a unique solution $u^m \in H$. We shall sketch the derivation of an error estimate for the above method.

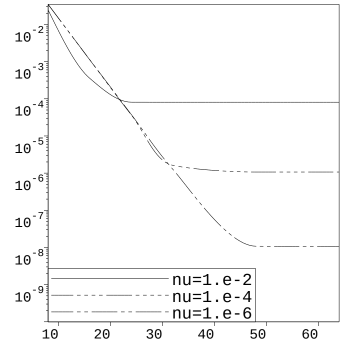

Figure 4.1: Results for the advection-diffusion test problem

Suppose that u is a smooth solution of Problem 1. Then

$$\begin{aligned} u(t^m) + \Delta t A u(t^m) - E(t^m; t^{m-1}) u(t^{m-1}) &= \Delta t A u(t^m) + \int_{t^{m-1}}^{t^m} \frac{d}{ds} E(t^m; s) u(s) ds \\ &= \int_{t^{m-1}}^{t^m} A u(t^m) - E(t^m; s) A u(s) ds \\ &= \int_{t^{m-1}}^{t^m} \int_s^{t^m} \frac{d}{d\tau} E(t^m; \tau) A u(\tau) d\tau ds \\ &= \int_{t^{m-1}}^{t^m} (s - t^{m-1}) E(t^m; s) D_t A u(s) ds, \end{aligned}$$

and so

$$\left\| u(t^m) + \Delta t A u(t^m) - E(t^m; t^{m-1}) u(t^{m-1}) \right\| \leq \Delta t \left\| D_t A u \right\|_{L^1(t^{m-1}, t^m; H)}.$$

It follows that

$$\max_{k=0,\ldots,m} \left\| u(t^k) - u^k \right\| \leq \nu \Delta t \left\| D_t \nabla^2 u \right\|_{L^1(H)}. \tag{4.8}$$

It will be shown below that (4.8) still holds if u is the solution to Problem 1 furnished by Theorem 4.1. We see that the time-truncation properties of the scheme are good: when $\nu \to 0$ the discretisation is exact, recovering the situation already observed in the hyperbolic case. This is in contrast to the errors committed by standard Eulerian timestepping schemes, where the truncation error is of the form $\left\| \frac{d^2}{dt^2} u \right\|_{L^1(H)}$, and will not vanish as $\nu \to 0$.

Example 1. In order to illustrate the meaning of (4.8), let us consider the application of (4.7), when coupled with the spectral discretisation in space described in Chapter 2, to the test problem

$$\begin{aligned} \frac{\partial u}{\partial t} - (\sin x) \frac{\partial u}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} &= 0 \quad \text{in } (0, 2\pi) \times (0, T] \qquad (4.9a) \\ u(x, 0) &= \sin x \quad \text{in } (0, 2\pi). \qquad (4.9b) \end{aligned}$$

The exact solution to this problem develops sharp layers at $x = 0$ and $x = 2\pi$, which eventually decay. Although the flow field here is not divergence-free, an error estimate similar to (4.8) may still be obtained, since $E(t; s)$ maps $H^* \to H^*$ (and then also $V^* \to V^*$ via Lemma 2.2), where H^* (respectively V^*) is the restriction of H (respectively V) to the subspace consisting of odd functions.

A standard Fourier Galerkin method with a fourth order Runge-Kutta time integrator at high resolution was employed to obtain an 'exact' solution for the purpose of comparison at time $t = 1.57$. In Figure 4.1 the error measured in the L^2-norm is plotted on a log-scale against L (with $N = L - 1$), for various values of the diffusion ν. 320 time steps of the scheme (4.7) were taken to reach the final time. The exponential accuracy of the spectral approximation is evident, although eventually it is obscured by the error due to the time discretisation. This final error is proportional to ν, as predicted by the theory, but only for very small values of ν will it be comparable to the spatial discretisation error.

When approximating problems with smooth solutions, for which the spectral method is evidently capable of producing extremely accurate results, we would like to obtain a more balanced global error. Thus we shall introduce discretisation schemes that are higher-order in time. These can be obtained in a most natural way from the Lagrangian form of the equations to be derived below.

4.2 A Lagrangian form of the equation

Writing formally $A(t; s) = E(t; s) A E(s; t)$, and $v(t; s) = E(t; s) u(\cdot, s)$, (4.6) takes the form

$$\left(\frac{d}{ds} + A(t; s) \right) v(t; s) \Big|_{s=t} = 0. \tag{4.10}$$

Our initial aim in this section is to make the definition of the operator $A(t; s)$ precise and to describe some of its properties. We shall then be in a position to give a weak form of (4.10) (removing the restriction $s = t$) and to apply standard existence and uniqueness theorems, on which the analysis of the numerical schemes will be based. Our first task is derive some preliminary results concerning the operator $E(t; s)$.

4.2.1 Further properties of $E(t; s)$

The first property we require concerns the commutator $[E(t; s), \nabla]$ of $E(t; s)$ with the gradient operator ∇ and is summed up in the following lemma

Lemma 4.1 *Let $a \in L^1(C^1_{\#}(\overline{\Omega})^n)$. Then for each $t, s \in [0, T]$ the commutator $[E(t; s), \nabla]$ belongs to $L(V, H^n)$, and we have*

$$\| [E(t; s), \nabla] \|_{L(V, H^n)} \leq e^{|a|_{1,1;[t,s]}} - 1. \tag{4.11}$$

Proof The first statement of the lemma is established by Lemma 2.2, once we have shown that $\int_\Omega u = 0 \Rightarrow \int_\Omega E(t; s) u = 0$, which is a consequence of $\nabla \cdot a = 0$. To establish (4.11) we use the notation of Lemma 2.2 to write

$$[D_j, E(t)] u = \xi_{jk}(t) E(t) D_k u,$$

where we have taken $s = 0$ without loss of generality. It follows that

$$\| [\nabla, E(t)] u \| \leq \xi_0(t) \| \nabla u \|.$$

Now, by (2.26),

$$\xi_0(t) \leq e^{\int_0^t A_0(\tau) d\tau} - 1 \leq e^{|a|_{1,1;[0,t]}} - 1,$$

and the result follows.

Next, we present a result which, like Lemma 4.1, is almost a corollary of Lemma 2.2, and which will be useful in the analysis of time-discrete schemes in the next section.

Lemma 4.2 *Let $a \in L^1(C^2_{\#}(\overline{\Omega})^n)$. Then for each $t, s \in [0, T]$ the commutator $[E(t; s), A]$ is in $L(D(A), H)$, and*

$$\| [E(t; s), A] \|_{L(D(A), H)} \leq e^{2|a|_{1,1;[t,s]} + |a|_{2,1;[t,s]}} - 1. \tag{4.12}$$

Proof Again following the notation of Lemma 2.2 (and setting $s = 0$), we have

$$\begin{aligned} D_i D_i E(t) u &= D_i (E(t) D_i u + \xi_{ij}(t) E(t) D_j u) \\ &= E(t) D_i D_i u + 2 \xi_{ij}(t) E(t) D_j D_i u \\ &\quad + (D_i \xi_{ij}(t)) E(t) D_j u + \xi_{ij}(t) \xi_{ik}(t) E(t) D_k D_j u. \end{aligned}$$

It follows that

$$\begin{aligned} \| [E(t), \Delta] u \| &\leq 2 \xi_0(t) \| \Delta u \| + \xi_1(t) \| \nabla u \| + \xi_0^2(t) \| \Delta u \| \\ &\leq (\xi_1(t) + (\xi_0(t) + 1)^2 - 1) \| \Delta u \|. \end{aligned}$$

Now, (2.23) translates to

$$\begin{aligned} D_i \xi_{ij}(t) &= \int_0^t [(D_i \xi_{ik}(s)) E(s) D_k a_j(s) + 2 \xi_{ik}(s) E(s) D_i D_k a_j(s) \\ &\quad + \xi_{ik}(s) \xi_{il}(s) E(s) D_k D_l a_j(s) + E(s) D_i D_i a_j(s)] ds, \end{aligned}$$

so that

$$\begin{aligned}\xi_1(t) &\leq \int_0^t [\xi_1(s)|\nabla a(s)|_\infty + 2\xi_0(s)|\Delta a(s)|_\infty \\ &\quad +\xi_0^2(s)|\Delta a(s)|_\infty + |\Delta a(s)|_\infty]ds \\ &\leq \int_0^t (\xi_1(s) + e^{2|a|_{1,1;[0,s]}})|\Delta a(s)|_\infty ds,\end{aligned}$$

using the fact that $\xi_0(t) \leq e^{|a|_{1,1;[0,s]}} - 1$. Then, by Gronwall's lemma,

$$\begin{aligned}\xi_1(t) &\leq e^{2|a|_{1,1;[0,t]}} \int_0^t |\Delta a(s)|_\infty e^{\int_s^t |\Delta a(\tau)|_\infty d\tau} ds \\ &= e^{2|a|_{1,1;[0,t]}}(e^{|a|_{2,1;[0,t]}} - 1),\end{aligned}$$

and the result follows.

Let $u \in H$ and $v \in V$. Lemma 2.1 implies that

$$(E(t;s)u, v) = (u, E(s;t)v),$$

where $E(s;t)v \in V$. This motivates the following extension of $E(t;s)$ to V', the dual space of V.

Definition 4.1 *Let $v' \in V'$; then $E(t;s)v'$ is that element of V' such that*

$$\langle E(t;s)v', v\rangle = \langle v', E(s;t)v\rangle \quad \forall v \in V,$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between V and V'.

Then, for each $s, t \in [0, T]$, $E(t;s)$ is an isometry on V' and bounds similar to (2.12) hold.

4.2.2 The operator $A(t;s)$

Our first step towards a precise definition of the operator $A(t;s)$ is now open to us. For $s, t \in [0, T]$ we define the bilinear form $\mathcal{A}(s;t|\cdot,\cdot)$ on $V \times V$ by

$$\mathcal{A}(s;t|u,v) = \nu(\nabla E(s;t)u, \nabla E(s;t)v), \quad \forall u, v \in V. \tag{4.13}$$

The continuity of $\mathcal{A}(s;t|\cdot,\cdot)$ is assured by Lemma 4.1. By the Riesz Representation Theorem we may associate with $\mathcal{A}(s;t|\cdot,\cdot)$ a continuous linear operator $A(t;s)$ from V to V' as follows: for every $u \in V$, $A(t;s)u$ is defined to be that element of V' which satisfies

$$\langle A(t;s)u, v\rangle = \mathcal{A}(s;t|u,v) \quad \forall v \in V.$$

We note that for all $s, t \in [0, T]$,

$$A(t;s) = E(t;s)AE(s;t) \quad \text{in} \ \ L(V, V').$$

The following lemma encapsulates further properties of the bilinear form $\mathcal{A}(s;t|\cdot,\cdot)$ (and so of $A(t;s)$).

Lemma 4.3 *Let $a \in L^1(C^1_\#(\overline{\Omega})^n) \cap L^\infty(C_\#(\overline{\Omega})^n)$. Then*

For every $u, v \in V$, $t \in [0,T]$, $s \to \phi(s) = \mathcal{A}(s;t|u,v)$ is measurable; *(4.14a)*

$$\mathcal{A}(s;t|u,v)| \leq \nu e^{2|\nabla a|_{0,1;[t,s]}} \|u\|_V \|v\|_V, \quad \forall u, v \in V, \ \forall s, t \in [0,T]; \tag{4.14b}$$

$$\mathcal{A}(s;t|u,u) \geq \nu e^{-2|\nabla a|_{0,1;[t,s]}} \|u\|_V^2, \quad \forall u \in V, \ \ \forall s, t \in [0,T]; \tag{4.14c}$$

For every $u, v \in V$,
$\phi(s)$ is absolutely continuous with derivative $s \to \phi'(s) = \mathcal{A}'(s;t|u,v)$; *(4.14d)*

For a.e. $s, t \in [0, T]$, $\mathcal{A}'(s;t|\cdot,\cdot)$ is a bilinear continous form on V with

$$|\mathcal{A}'(s;t|u,v)| \leq 2|\nabla a|_{0,\infty} \nu e^{2|\nabla a|_{0,1;[t,s]}} \|u\|_V \|v\|_V, \quad \forall u, v \in V. \tag{4.14e}$$

Proof The properties (4.14b), (4.14c) follow immediately from Lemma 2.2. Let $u, v \in V$ and $t \in [0, T]$. Consider $\phi(s^*) - \phi(s)$ for $s, s^* \in (0, T)$. We have, denoting $E(s;t)$ by E and $E(s^*;s)$ by E^*,

$$\begin{aligned}\phi(s^*) - \phi(s) &= \nu(\nabla E^*Eu, \nabla E^*Ev) - \nu(\nabla Eu, \nabla Ev) \\ &= \nu(\nabla E^*Eu, \nabla E^*Ev) - \nu(E^*\nabla Eu, E^*\nabla Ev) \\ &= \nu([\nabla, E^*]Eu, \nabla E^*Ev) + \nu(E^*\nabla Eu, [\nabla, E^*]Ev).\end{aligned}$$

Thus, by Lemma 4.1,

$$|\phi(s^*) - \phi(s)| \leq \nu(e^{2|\nabla a|_{0,1;[s^*,s]}} - 1)e^{2|\nabla a|_{0,1;[t,s]}} \|u\|_V \|v\|_V, \tag{4.15}$$

so that ϕ is Lipschitz, which implies its absolute continuity. Hence ϕ is differentiable a.e. on $[0, T]$. (4.14a) and (4.14d) follow. Suppose $s \in (0, T)$ is a point at which ϕ is differentiable. Then (4.14e) is obtained from (4.15) by dividing through by $|s^* - s|$ and taking the limit as $s^* \to s$.

Since the bilinear form $\mathcal{A}(s;t|\cdot,\cdot)$ is coercive, the Lax-Milgram theorem allows us to deduce that $A(t;s)$ is an isomorphism from V onto V'. We can also define the domain of $A(t;s)$ in H by

$$D(A(t;s)) = \{u \in H \mid A(t;s)u \in H\}.$$

The form $\mathcal{A}(s;t|\cdot,\cdot)$ is symmetric so that the operator $A(t;s)$ is self-adjoint in H. Its inverse $A^{-1}(t;s)$ is also self-adjoint, in H.

We are now ready to state a Lagrangian form of Problem 1 as an abstract Cauchy problem in V. More precisely, we state:

Problem 2 *Given $t^* \in [0, T]$, $u_{t^*,0} \in H$, find $u_{t^*} : [0, T] \to V$ satisfying (in the sense of distributions, with values in V')*

$$\frac{d}{dt}u_{t^*}(t) + A(t^*;t)u_{t^*}(t) = 0 \quad \textit{on} \ (0, T], \tag{4.16a}$$

$$u_{t^*}(0) = u_{t^*,0}. \tag{4.16b}$$

The following results for Problem 2 are standard (see Theorems 3.4–3.6 of [87] for a general statement).

Theorem 4.2 *Suppose that (4.14a–4.14c) hold. Then there exists a unique solution u_{t^*} of Problem 2 such that*

$$u_{t^*} \in L^2(V) \cap C(H), \tag{4.17}$$

$$u'_{t^*} \in L^2(V'). \tag{4.18}$$

If, moreover, (4.14d) and(4.14e) hold, and we assume that

$$u_{t^*,0} \in D(A(t^*;0)), \tag{4.19}$$

then

$u_{t^*}(t) \in D(A(t^*;t)), \forall t \in [0, T]$ and
the map $t \to A(t^*;t)u_{t^*}(t)$ is continuous from $[0, T]$ into H, *(4.20)*

and

$$u'_{t^*} \in L^2(V) \cap C(H), \tag{4.21}$$

$$u''_{t^*} \in L^2(V'). \tag{4.22}$$

Suppose now that $a \in L^1(C^2_\#(\overline{\Omega})^n)$, so that (see Lemma 4.2) $D(A(t^*;t)) = D(A) = H^2_\#(\Omega) \cap H$. Replacing assumptions (4.19) by

$$u_{t^*,0} \in V, \tag{4.23}$$

we have

$$u_{t^*} \in L^2(D(A)) \cap C(V). \tag{4.24}$$

q	α_0	α_1	α_2	α_3	α_4	α_5	α_6	β_0
1	1	-1						1
2	1	-$\frac{4}{3}$	$\frac{1}{3}$					$\frac{2}{3}$
3	1	-$\frac{18}{11}$	$\frac{9}{11}$	-$\frac{2}{11}$				$\frac{6}{11}$
4	1	-$\frac{48}{25}$	$\frac{36}{25}$	-$\frac{16}{25}$	$\frac{3}{25}$			$\frac{12}{25}$
5	1	-$\frac{300}{137}$	$\frac{300}{137}$	-$\frac{200}{137}$	$\frac{75}{137}$	-$\frac{12}{137}$		$\frac{60}{137}$
6	1	-$\frac{360}{147}$	$\frac{450}{147}$	-$\frac{400}{147}$	$\frac{225}{147}$	-$\frac{72}{147}$	$\frac{10}{147}$	$\frac{60}{147}$

Table 4.1: Coefficients for the backward differentiation formulae.

Remark. It is perhaps worth noting that of the results in Theorem 4.2 only (4.24) depends on the symmetry of $\mathcal{A}(s;t|\cdot,\cdot)$.

It would seem at first sight that Problem 2 is actually a one-parameter family of problems $\{\mathsf{P2}(t^*)\}_{t^*\in[0,T]}$. However, it is readily seen that, if $s^*, t^* \in [0,T]$ and u_{t^*} is the solution of P2(t^*) with initial datum $u_{t^*,0}$, then $E(s^*;t^*)u_{t^*}$ satisfies P2(s^*) with initial datum $E(s^*;t^*)u_{t^*,0}$. Thus, since $E(s^*;t^*)$ is an isomorphism on V, the problems P2(t^*) and P2(s^*) are equivalent.

The central question of whether Problem 2 is equivalent to Problem 1 is more delicate. The answer is contained in the following lemma.

Lemma 4.4 *Problem 1 and Problem 2 are equivalent, in the sense that the unique solution to Problem 1 (with initial datum u_0) provided by Theorem 4.1 may be identified with the unique solution to Problem 2 (with initial datum $E(t^*;0)u_0$) provided by Theorem 4.2.*

Proof Let u be the solution to Problem 1. We shall show that u also satisfies Problem 2 and so it may be identified with the unique solution provided by Theorem 4.2.

Our first step is to show that

$$\frac{d}{dt}E(t^*;t)u(t) = E(t^*;t)\left(\frac{d}{dt} + a\cdot\nabla\right)u(t) \tag{4.25}$$

in $\mathcal{D}'(0,T;V')$. Since $u \in H^1(V')\cap L^2(V)$, there is (by density) a sequence $\{u_j\} \subset C^\infty((0,T);C^\infty_\#(\overline{\Omega}))$ which converges to u in the norm $\|\cdot\|_{H^1(V')\cap L^2(V)} = \|\cdot\|_{H^1(V')} + \|\cdot\|_{L^2(V)}$. For each j, (4.25) holds in the classical sense with u replaced by u_j. Let $\phi \in \mathcal{D}(0,T)$ and $v \in V$. Integrating (4.25) against v over Ω and against ϕ over $[0,T]$ we obtain

$$-\int_0^T \langle E(t^*;t)u_j(t), v\rangle\,\phi'(t)dt = \int_0^T \left\langle E(t^*;t)\left(\frac{d}{dt} + a\cdot\nabla\right)u_j(t), v\right\rangle\phi(t)dt, \tag{4.26}$$

where $\langle\cdot,\cdot\rangle$ is the duality pairing between V' and V. Since, for each $t\in[0,T]$, $E(t^*;t)$ is an isomorphism on V', uniformly bounded in t, and since $\frac{d}{dt}u_j \to \frac{d}{dt}u$ in $L^2(V')$ and $a\cdot\nabla u_j \to a\cdot\nabla u$ in $L^2(H)$, we may take the limit through (4.26) as $j\to\infty$ and replace u with u_j, so that (4.25) holds in the sense of distributions (with values in V').

In this sense, we may write

$$\left(\frac{d}{dt} + A(t^*;t)\right)u_{t^*}(t) = E(t^*;t)\left(\frac{d}{dt} + a\cdot\nabla + A\right)u(t),$$

where $u_{t^*} = E(t^*;\cdot)u$, and we deduce that u_{t^*} satisfies (4.16a).

The equivalence of Problems 1 and 2 means that we can derive discrete approximations to Problem 1 from schemes designed to approximate Problem 2. A natural way to construct such approximations is to consider the time-discrete case *first*, introducing discretisations in the space variables as a *second* step. This is the pattern our analysis will follow.

4.3 Time discretisation

Several authors (see, for example, [18, 46, 47, 48, 89]) have considered the problem of the time discretisation of abstract Cauchy problems such as Problem 1 and Problem 2. The results of Le Roux in [46] concerning higher-order time discretisations are especially relevant here, and we shall draw heavily on her work in this section.

Suppose that we discretise Problem 2 in time by a q-step linear multistep method. We obtain

$$\sum_{i=0}^{q}(\alpha_i + \Delta t\beta_i A(t^*;t^{m-i}))u_{t^*}^{m-i} = 0, \tag{4.27}$$

where m ranges from q to M, and $\Delta t = T/M$. We shall be interested primarily in the *backward difference formulae* (see for example Gear [29]), which are strongly A_0-stable for $q \le 6$. These have $\beta_i = 0$ for $i > 0$. Full details of the coefficients of these formulae are given in Table 4.1. Under stronger assumptions on the velocity field a than are necessary for the backward difference formulae, we shall also consider extensions to the case $\beta_i \ge 0$ for $i > 0$. We shall in either case assume that the method used is strongly A_0-stable. Le Roux's results may be applied to these schemes to give stability and optimal-order convergence to the solution of Problem 2. (4.27) represents a time-discretisation of Problem 1 in Lagrangian coordinates which originate at time t^*. We shall derive from (4.27) an approximation to Problem 1 involving a *different* set of Lagrangian coordinates at each time level (with the coordinates used to obtain u^m from (4.27) originating at time t^m) and present stability and convergence results for this. We shall only sketch the proofs, since they are at all points analogous to the proofs to be found in [46].

Setting $u^{m,i} = E(t^{m+i};t^*)u_{t^*}^m = E(t^{m+i};t^m)u^m$, $i = 0,\ldots,q$, we find, when $\beta_i = 0$ for $i > 1$, that the method represented by (4.27) is equivalent to

$$u^{m,0} = u^m \quad = \quad -(\alpha_0 + \Delta t\beta_0 A)^{-1}\sum_{i=1}^{q}E(t^m;t^{m-1})(\alpha_i + \Delta t\beta_i A)u^{m-i,i-1} \tag{4.28a}$$

$$u^{m-i,i} \quad = \quad E(t^m;t^{m-1})u^{m-i,i-1}, \quad i = 1,\ldots,q-1, \tag{4.28b}$$

which is the form used in practice.

Fig. 4.2 illustrates (4.28) for a second-order backward difference time step with $q = 2$.

$$u^{m,0} \quad = \quad (1 + \frac{2}{3}\Delta tA)^{-1}(\frac{4}{3}u^{m-1,1} - \frac{1}{3}u^{m-2,2}) \tag{4.29a}$$

$$u^{m-i,i} \quad = \quad E(t^m;t^{m-1})u^{m-i,i-1}, \quad i = 1,2. \tag{4.29b}$$

The meaning of the labelling is as follows. For example, the positioning of the label $u^{m-1,1}$ corresponds to the point X_j, the foot of the trajectory through (x_j,t^m). We emphasise that the $m-1$ in the superscript indicates that $u^{m-1,1}$ represents the solution at time t^{m-1}, and the 1 corresponds to the fact that $x_j \to X_j$ represents evolution through one time step ($E(t^m;t^{m-1})$). One can think of the procedure given by (4.29) as being

$$(u^{m-1,0}, u^{m-2,1}) \quad \underset{\text{(4.29b)}}{\rightarrow} \quad (u^{m-1,1}, u^{m-2,2}) \quad \underset{\text{(4.29a)}}{\rightarrow} \quad (u^{m,0}, u^{m-1,1}).$$

For the purposes of our analysis we shall define the $q\times q$ matrices of operators

$$\Lambda^* = \begin{pmatrix} -(\alpha_0+\beta_0\Delta tA)^{-1} & & & & \\ & I & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot & \\ & & & & & I \end{pmatrix},$$

$$\Lambda^{**} = \begin{pmatrix} (\alpha_1+\beta_1\Delta tA) & \cdot & \cdot & \cdot & \cdot & (\alpha_q+\beta_q\Delta tA) \\ I & & & & 0 & 0 \\ & \cdot & & & \cdot & \cdot \\ & & \cdot & & \cdot & \cdot \\ & & & \cdot & 0 & \cdot \\ & & & & I & 0 \end{pmatrix},$$

$$E_m = \begin{pmatrix} E(t^m;t^{m-1}) & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & E(t^m;t^{m-1}) \end{pmatrix},$$

$$\Lambda_\infty = \begin{pmatrix} \frac{\beta_1}{\beta_0}I & 0 & \cdot & \cdot & 0 & 0 \\ I & & & & \cdot & \cdot \\ & \cdot & & & \cdot & \cdot \\ & & \cdot & & \cdot & \cdot \\ & & & \cdot & 0 & \cdot \\ & & & & I & 0 \end{pmatrix},$$

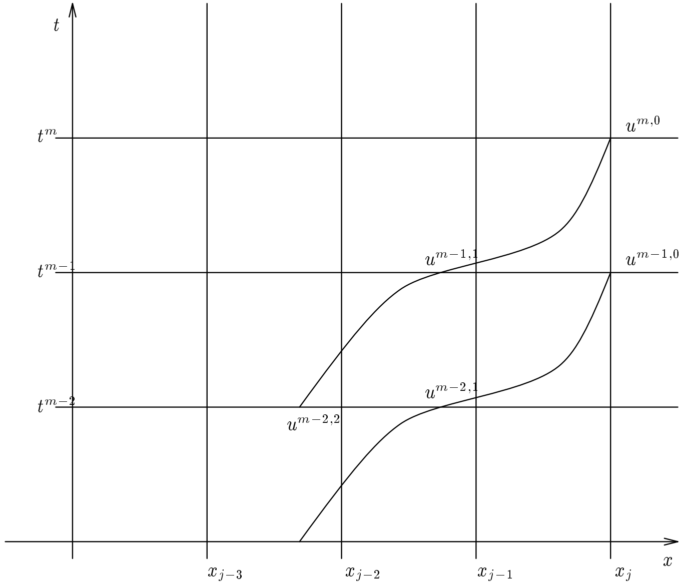

Figure 4.2: Illustration of a second order time step

and we shall denote

$$\Lambda_m = \Lambda^* E_m \Lambda^{**} \quad \text{and} \quad \Lambda = \Lambda^* \Lambda^{**}.$$

Writing $u^m = (u^{m,0}, \ldots, u^{m-q+1,q-1})^T$, the method (4.28) may then be written

$$u^m = \Lambda_m u^{m-1}. \tag{4.30}$$

4.3.1 Stability

Our first concern will be to demonstrate that the scheme is stable. By (4.30) this is equivalent to showing that there is a constant C such that

$$\|\Lambda_k \ldots \Lambda_q\|_{L(H^q)} \leq C,$$

uniformly in k. The positive constants C and μ will, in the following discussion, be generic constants depending only on the particular multistep formula used (i.e. on $\alpha_0, \ldots, \alpha_q$ and $\beta_0, \ldots, \beta_q$), and sometimes on Δt_0, an upper bound on the size of the timestep. μ will also be proportional to $\min\{\nu, \frac{1}{\Delta t_0}\}$. Any other dependence will be explicitly stated. We have the following operator identities.

$$\begin{aligned} \Lambda_k \ldots \Lambda_q &= E_k \ldots E_q \Lambda^{k-q+1} \\ &\quad + \sum_{p=q+1}^{k+1} E_k \ldots E_p (\Lambda^{k+1-p} \Lambda_{p-1} - E_{p-1} \Lambda^{k+2-p}) \Lambda_{p-2} \ldots \Lambda_q, \end{aligned} \tag{4.31}$$

and

$$\Lambda^k \Lambda_p - E_p \Lambda^{k+1} = [\Lambda^{k+1}, E_p] + \Lambda^k (\Lambda_p - \Lambda E_p). \tag{4.32}$$

As in [46], Proposition 4, we have, for $k \geq 1$,

$$\left\|\Lambda^k\right\|_{L(H^q)} \leq C e^{-\mu k \Delta t}. \tag{4.33}$$

It remains to bound the two terms appearing on the right hand side of (4.32). For the first, the following operator identity holds.

$$[\Lambda^{k+1}, E_p] = \sum_{j=0}^{k} \Lambda^j [\Lambda, E_p] \Lambda^{k-j}$$

$$\begin{aligned} &= \sum_{j=1}^{k-1} (\Lambda^j - \Lambda^j_\infty)[\Lambda, E_p](\Lambda^{k-j} - \Lambda^{k-j}_\infty) + \sum_{j=0}^{k-1} \Lambda^j_\infty [\Lambda, E_p](\Lambda^{k-j} - \Lambda^{k-j}_\infty) \\ &\quad + \sum_{j=1}^{k} (\Lambda^j - \Lambda^j_\infty)[\Lambda, E_p]\Lambda^{k-j}_\infty + \sum_{j=0}^{k} \Lambda^j_\infty [\Lambda, E_p] \Lambda^{k-j}_\infty. \end{aligned} \tag{4.34}$$

From [46], p.133, we have

$$\left\|A^{1/2}(\Lambda^j - \Lambda^j_\infty)\right\|_{L(H^q)} = \left\|(\Lambda^j - \Lambda^j_\infty)A^{1/2}\right\|_{L(H^q)} \leq \frac{Ce^{-\mu j \Delta t}}{(j\Delta t)^{1/2}}, \tag{4.35}$$

and from [46], Lemma 3, we have

$$\left\|\Lambda^j_\infty\right\|_{L(H^q)} \leq Ce^{-\mu j}. \tag{4.36}$$

We thus seek suitable bounds on $\|A^{-\epsilon_1}[\Lambda, E_p]A^{-\epsilon_2}\|_{L(H^q)}$ for $\{\epsilon_1, \epsilon_2\} = \{\frac{1}{2}, \frac{1}{2}\}$, $\{\frac{1}{2}, 0\}$, $\{0, \frac{1}{2}\}$, and $\{0, 0\}$. We have

$$\begin{aligned} &\left\|A^{-\epsilon_1}[\Lambda, E_p]A^{-\epsilon_2}\right\|_{L(H^q)} \\ &\quad \leq \sum_{i=1}^{q} \left\|A^{-\epsilon_1}[(\alpha_0 + \beta_0 \Delta t A)^{-1}(\alpha_i + \beta_i \Delta t A), E(t^p; t^{p-1})]A^{-\epsilon_2}\right\|. \end{aligned} \tag{4.37}$$

Now, denoting $E(t^p; t^{p-1})$ by E,

$$A^{-\epsilon_1}[(\alpha_0 + \beta_0 \Delta t A)^{-1}(\alpha_i + \beta_i \Delta t A), E]A^{-\epsilon_2} \tag{4.38}$$

$$= \Delta t A^{-\epsilon_1}(\alpha_0 + \beta_0 \Delta t A)^{-1}(\beta_i \alpha_0 - \beta_0 \alpha_i)[A, E]A^{-\epsilon_2}(\alpha_0 + \beta_0 \Delta t A)^{-1}, \tag{4.39}$$

and we have (see [46], p.132)

$$\left\|A^{-\epsilon_1}(\alpha_0 + \beta_0 \Delta t A)^{-1}\right\|_{L(H,V)} \leq C\Delta t^{\epsilon_1 - \frac{1}{2}}, \tag{4.40}$$

and

$$\left\|(\alpha_0 + \beta_0 \Delta t A)^{-1} A^{-\epsilon_2}\right\|_{L(V',H)} \leq C\Delta t^{\epsilon_2 - \frac{1}{2}}. \tag{4.41}$$

A bound on $\|[A, E]\|_{L(V,V')}$ may be obtained as follows. Let $u, v \in V$: then

$$\begin{aligned} \langle [A, E]u, Ev \rangle &= \nu(\nabla E u, \nabla E v) - \nu(\nabla u, \nabla v) \\ &= \nu([\nabla, E]u, \nabla E v) + \nu(E \nabla u, [\nabla, E]v), \end{aligned}$$

so that

$$\left\|[A, E(t^p; t^{p-1})]\right\|_{L(V,V')} \leq \nu(e^{2|\nabla a|_{0,1;[t^{p-1},t^p]}} - 1). \tag{4.42}$$

Combining (4.37–4.42) we obtain

$$\left\|A^{-\epsilon_1}[\Lambda, E_p]A^{-\epsilon_2}\right\|_{L(H^q)} \leq C(e^{2|\nabla a|_{0,1;[t^{p-1},t^p]}} - 1)\Delta t^{\epsilon_1 + \epsilon_2}. \tag{4.43}$$

(4.43), together with (4.33)–(4.36), may be used as in [46] pp.131–134 to give

$$\left\|[\Lambda^{k+1}, E_p]\right\|_{L(H^q)} \leq C(e^{2|\nabla a|_{0,1;[t^{p-1},t^p]}} - 1)e^{-\mu(k+1)\Delta t}. \tag{4.44}$$

The second term appearing on the right hand side of (4.32), $\Lambda^k(\Lambda_p - \Lambda E_p)$, vanishes when the multistep scheme being used is a backward difference formula ($\beta_i = 0, i > 0$). When for example $\beta_1 \neq 0$, we must assume that $a \in L^1(C^2_\#(\overline{\Omega})^n)$ so that Lemma 4.2 applies. Then we have

$$\left\|\Lambda^k(\Lambda_p - \Lambda E_p)\right\|_{L(H^q)} \leq C|a|_{2,1;[t^{p-1},t^p]} e^{-\mu(k+1)\Delta t}. \tag{4.45}$$

We can now deduce from (4.31), (4.32), (4.44) and (4.45) that

$$\begin{aligned} \|\Lambda_k \ldots \Lambda_q\|_{L(H^q)} &\leq Ce^{-\mu(k-q+1)\Delta t} \\ &\quad + C \sum_{p=q+1}^{k+1} |a|_{1,1;[t^{p-1},t^p]} e^{-\mu(k+2-p)\Delta t} \|\Lambda_{p-2} \ldots \Lambda_q\|_{L(H^q)}, \end{aligned} \tag{4.46}$$

with $|a|_{1,1;[t^{p-1},t^p]}$ replaced by $|a|_{2,1;[t^{p-1},t^p]}$ if $\beta_1 \neq 0$. If we write $b_q = 1$, $a_k = |a|_{1,1;[t^k,t^{k+1}]}$ and $b_k = e^{\mu(k-q)\Delta t} \|\Lambda_k \dots \Lambda_q\|_{L(H^q)}$, for $k \geq q+1$, (4.46) becomes

$$b_{k+1} \leq C(1 + \sum_{p=q}^{k} a_p b_p).$$

The discrete Gronwall lemma now allows us to deduce that

$$b_{k+1} \leq Ce^{C\sum_{p=q}^{k} a_p},$$

which when translated back into the original terms gives us the following result.

Theorem 4.3 (Stability) *Suppose that the multistep method (4.27) is strongly A_0-stable, and is such that $\beta_i = 0$ for $i > 0$, or, if $\beta_i geq$ for $i \geq 1$, that $a \in L^1(C^2_\#(\overline{\Omega})^n)$. Then there is a positive constant C depending only on $\alpha_0, \dots, \alpha_q$, $\beta_0, \dots, \beta_q$ and Δt_0, and a constant μ depending on Δt_0 and ν, such that, for $m = q, \dots, M$,*

$$\|\Lambda_m \dots \Lambda_q\|_{L(H^q)} \leq Ce^{C|a|_{1,1} - \mu(m-q+1)\Delta t}, \tag{4.47}$$

with $|a|_{1,1}$ replaced by $|a|_{2,1}$ when $\beta_{i(\geq 1)} \neq 0$.

4.3.2 Convergence

Let u be the solution to Problem 1 with initial datum u_0. Then, by Theorem 4.1, $u \in C(H)$ and so we can sensibly refer to $u(t)$, for $t \in [0,T]$. Suppose that $u^{0,q-1}, \dots, u^{q-1,0}$ are given as starting values for the method (4.28), and that they correspond to approximations of the terms $E(t^{q-1};0)u_0, E(t^{q-1};t^1)u(t^1), \dots, u(t^{q-1})$. Then we have the following convergence result.

Theorem 4.4 (Convergence) *We suppose that the conditions of Theorem 4.3 are satisfied, and that $D_t^{p+1}u \in L^1(H)$, where the multistep scheme (4.27) has order of accuracy p. Then, for $m = q-1, \dots, M$,*

$$\begin{aligned} \|u(t^m) - u^{m,0}\| \quad &\leq \quad Ce^{C|a|_{1,1}} \left(e^{-\mu(m-q+1)\Delta t} \sum_{i=0}^{q-1} \|E(t^{q-1};t^i)u(t^i) - u^{i,q-1-i}\| \right. \\ &\qquad \left. + \Delta t^p \int_0^T e^{-\mu(T-t)} \left\| D_t^{p+1} u(t) \right\| dt \right). \end{aligned} \tag{4.48}$$

Proof We begin by defining the *truncation error* T^m at time t^m by

$$T^m = \frac{1}{\Delta t} \sum_{i=0}^{q} E(t^m;t^{m-i})(\alpha_i + \Delta t \beta_i A) u(t^{m-i}). \tag{4.49}$$

Now, let $u_{t^*}(t) = E(t^*;t)u(t)$, $t, t^* \in [0,T]$. Lemma 4.4 established that u_{t^*} is then a solution of Problem 2, and it is readily seen that, writing $T^m_{t^*} = E(t^*;t^m)T^m$,

$$\begin{aligned} T^m_{t^*} \quad &= \quad -\frac{1}{\Delta t} \sum_{i=0}^{q} (\alpha_i + \Delta t \beta_i A(t^*;t^{m-i})) u_{t^*}(t^{m-i}) \\ &= \quad -\frac{1}{\Delta t} \sum_{i=0}^{q} \left(\alpha_i u_{t^*}(t^{m-i}) - \Delta t \beta_i u'_{t^*}(t^{m-i}) \right), \end{aligned}$$

and standard Taylor-series expansion then yields, changing back to our original terms,

$$\|T^m\| \leq C\Delta t^{p-1} \int_{t^{m-q}}^{t^m} \left\| D_t^{p+1} u \right\| dt. \tag{4.50}$$

Now let us define, for $m = q-1, \dots, M$,

$$\zeta^m = \left(u_{t^m}(t^m) - u^{m,0}, \dots, u_{t^m}(t^{m-q+1}) - u^{m-q+1,q-1} \right)^T.$$

Then we have

$$\begin{aligned} \zeta^m \quad &= \quad \Lambda_m \zeta^{m-1} + \Delta t \Lambda^* \left(T^m, 0, \dots, 0 \right)^T \\ &= \quad \Lambda_m \dots \Lambda_q \zeta^{q-1} + \Delta t \sum_{k=1}^{m-q} \Lambda_m \dots \Lambda_{m-k+1} \Lambda^* \left(T^{m-k}, 0, \dots, 0 \right)^T \\ &\qquad + \Delta t \Lambda^* \left(T^m, 0, \dots, 0 \right)^T, \end{aligned}$$

so that, making use of Theorem 4.3,

$$\begin{aligned} \|\zeta^m\|_{H^q} \quad &\leq \quad \|\Lambda_m \dots \Lambda_q\|_{L(H^q)} \left\| \zeta^{q-1} \right\|_{H^q} \\ &\qquad + \Delta t \sum_{k=1}^{m-q} \|\Lambda_m \dots \Lambda_{m-k+1}\|_{L(H^q)} \left\| T^{m-k} \right\| + \Delta t \|T^m\| \\ &\leq \quad Ce^{C|a|_{1,1} - \mu(m-q+1)\Delta t} \left\| \zeta^0 \right\|_{H^q} \\ &\qquad + Ce^{C|a|_{1,1}} \Delta t^p \sum_{k=1}^{m-q} e^{-\mu k \Delta t} \left\| D_t^{p+1} u \right\|_{L^1(t^{m-q-k}, t^{m-k}; H)} \end{aligned}$$

and (4.48) follows.

These results concern the performance of the Lagrange-Galerkin method's approach to timestepping, which may be implemented in combination with a variety of spatial discretisation methods, such as finite difference methods, finite element methods, or spectral methods. In the next section we shall examine the effect of the incorporation of the spectral projection and interpolation operators introduced in Chapter 2.

4.4 Space discretisation

In this section we turn our attention to fully-discrete practicable methods, and we restrict ourselves to those arising from backward difference time discretisations (4.27) with $\beta_i = 0, i > 0$. Let K, L and N be positive integers, with L and N related by (2.53). Corresponding to (4.28) we have, for $m = q, \dots, M$,

$$U^{m,0} \quad = \quad -(\alpha_0 + \Delta t \beta_0 A)^{-1} \sum_{i=1}^{q} \alpha_i \Pi_N P_L E(t^m;t^{m-1}) \mathcal{P}_{K,L} U^{m-i,i-1} \tag{4.51a}$$

$$U^{m-i,i} \quad = \quad \Pi_N P_L E(t^m;t^{m-1}) \mathcal{P}_{K,L} U^{m-i,i-1}, \quad i = 1, \dots, q-1, \tag{4.51b}$$

and we assume that $U^{0,q-1}, \dots, U^{q-1,0}$ (each in S_N) are either given or have been calculated by some other means to approximate $u_{t^{q-1}}(0), \dots, u_{t^{q-1}}(t^{q-1})$, where u is the solution to Problem 1, and, for $s, t \in [0,T]$, $u_s(t) = E(s;t)u(t)$.

Defining $\Pi_N P_L$ and $\mathcal{P}_{K,L}$ on H^q in the obvious way, we further define, for $m = q, \dots, M$,

$$\widetilde{\Lambda}_m = \Lambda^* \Pi_N P_L E_m \mathcal{P}_{K,L} \Lambda^{**},$$

and, for $m = q-1, \dots, M$,

$$U^m = \left(U^{m,0}, \dots, U^{m-q+1,q-1} \right)^T,$$

so that (4.51) may be rewritten

$$U^m = \widetilde{\Lambda}_m U^{m-1}. \tag{4.52}$$

4.4.1 Stability

Our concern is to obtain a bound on $\widetilde{\Lambda}_m \dots \widetilde{\Lambda}_q$ in the $L(S_N^q)$-norm. We have, for $k = q, \dots, m$,

$$\begin{aligned} \widetilde{\Lambda}_k \dots \widetilde{\Lambda}_q &= \Pi_N \Lambda_k \dots \Pi_N \Lambda_q \\ &\quad + \sum_{j=q}^{k} \Pi_N \Lambda_k \dots \Pi_N \Lambda_{j+1} (\widetilde{\Lambda}_j - \Pi_N \Lambda_j) \widetilde{\Lambda}_{j-1} \dots \widetilde{\Lambda}_q. \end{aligned} \tag{4.53}$$

The stability of the first term follows exactly as in Theorem 4.3 since we have, for $j = q, \dots, k$,

$$\Pi_N \Lambda_k \dots \Pi_N \Lambda_j = \Pi_N E_k \dots \Pi_N E_j \Lambda^{k+1-j} \tag{4.54}$$

$$+ \sum_{p=j}^{k} \Pi_N E_k \dots \Pi_N E_{p+1} \Lambda^{k-p} \Pi_N (\Lambda_p - E_p \Lambda) \Pi_N \Lambda_{p-1} \dots \Pi_N \Lambda_j, \tag{4.55}$$

and it is readily seen that, with the positive constants C and μ exactly as in Theorem 4.3,

$$\|\Pi_N\Lambda_k\ldots\Pi_N\Lambda_j\|_{L(S_N^q)} \leq Ce^{C|a|_{1,1}-\mu(k+1-j)\Delta t}. \tag{4.56}$$

Our attention thus centres on the term

$$(\widetilde{\Lambda}_j - \Pi_N\Lambda_j) = \Lambda^*(\Pi_N P_L E_j \mathcal{P}_{K,L} - \Pi_N E_j)\Lambda^{**}$$

appearing in (4.53). The central term may be expanded

$$\Pi_N P_L E_j \mathcal{P}_{K,L} - \Pi_N E_j = \Pi_N P_L E_j(\mathcal{P}_{K,L} - I) + \Pi_N(P_L - I)E_j,$$

so that, under the conditions of Theorem 3.1, there is a positive constant C depending on C_5 and on C_6 such that

$$\|\Pi_N P_L E_j \mathcal{P}_{K,L} - \Pi_N E_j\|_{L(S_N^q)} \leq C|a|_{[\sigma]^*,1;[t^{j-1},t^j]}, \tag{4.57}$$

so that, from (4.53) and (4.56),

$$\begin{aligned}\left\|\widetilde{\Lambda}_k\ldots\widetilde{\Lambda}_q\right\|_{L(S_N^q)} \leq\; & Ce^{C|a|_{1,1}-\mu(k-q+1)\Delta t} \\ & +C\sum_{j=q}^{k}|a|_{[\sigma]^*,1;[t^{j-1},t^j]}e^{C|a|_{1,1}-\mu(k-j)\Delta t}\left\|\widetilde{\Lambda}_{j-1}\ldots\widetilde{\Lambda}_q\right\|_{L(S_N^q)}.\end{aligned}$$

In the same way that Theorem 4.3 was deduced from (4.46), we obtain the following stability result for the fully discrete scheme (4.52).

Theorem 4.5 *Suppose that the multistep method (4.27) is strongly A_0-stable, and is such that $\beta_i = 0$ for $i > 0$. Suppose also that the conditions of Theorem 3.1 hold. Then there is a positive constant C, depending only on $\alpha_0, \ldots, \alpha_q$, β_0, Δt_0, C_5 and C_6, and another constant μ proportional to $\min\{\nu, \frac{1}{\Delta t_0}\}$, such that, for $m = q, \ldots, M$,*

$$\left\|\widetilde{\Lambda}_m\ldots\widetilde{\Lambda}_q\right\|_{L(S_N^q)} \leq Ce^{C|a|_{[\sigma]^*,1}-\mu(m-q+1)\Delta t}. \tag{4.58}$$

4.4.2 Convergence

The strategy employed in the proof of Theorem 4.5 (i.e. that of regarding $\Pi_N P_L E_j \mathcal{P}_{K,L}$ as a perturbation of $\Pi_N E_j$) is exactly the same as that employed in the proofs of the stability results in Chapter 2. In the same way, our demonstration of the convergence of the fully-discrete scheme here will, in its essential ideas, resemble that of the convergence results in that chapter.

For $m = q-1, \ldots, M$ and $i = 0, \ldots, q-1$ we define

$$\xi^{m,i} = \Pi_N u_{t^{m+i}}(t^m) - U^{m,i} \quad \text{and} \quad \eta^{m,i} = (I - \Pi_N)u_{t^{m+i}}(t^m).$$

If we denote $E(t^m; t^{m-1})$ by E, and write

$$\begin{aligned}\widetilde{T}^{m,i} \;=\;& \Pi_N E\eta^{m-i,i-1} - \Pi_N(P_L - I)E\Pi_N u_{t^{m-1}}(t^{m-i}) \\ & -\Pi_N P_L E(\mathcal{P}_{K,L} - I)\Pi_N u_{t^{m-1}}(t^{m-i}), \quad i = 1, \ldots, q.\end{aligned} \tag{4.59}$$

we find, from (4.49) and (4.51), that

$$\xi^{m,0} = -(\alpha_0 + \Delta t\beta_0 A)^{-1}\left(\sum_{i=1}^{q}\alpha_i(\Pi_N P_L E\mathcal{P}_{K,L}\xi^{m-1,i-1} + \widetilde{T}^{m,i}) + \Delta t\Pi_N T^m\right) \tag{4.60a}$$

$$\xi^{m-i,i} = \Pi_N P_L E\mathcal{P}_{K,L}\xi^{m-i,i-1} + \widetilde{T}^{m,i}, \quad i = 1, \ldots, q-1, \tag{4.60b}$$

(compare (4.59) and (4.60) with (3.14)). Defining $\widetilde{T}^m = (\sum_{i=1}^q \widetilde{T}^{m,i}, \widetilde{T}^{m,1}, \ldots, \widetilde{T}^{m,q-1})^T$ and $\xi^m = (\xi^{m,0}, \ldots, \xi^{m-q+1,q-1})^T$ we can rewrite (4.60) as

$$\xi^m = \widetilde{\Lambda}_m\xi^{m-1} + \Delta t\Lambda^*(T^m, 0, \ldots, 0)^T + \Lambda\widetilde{T}^m. \tag{4.61}$$

Then, by expanding the right hand side of (4.61),

$$\begin{aligned}\|\xi^m\|_{H^q} \leq\; & \left\|\widetilde{\Lambda}_m\ldots\widetilde{\Lambda}_q\right\|_{L(S_N^q)}\|\xi^{q-1}\|_{H^q} \\ & +\sum_{j=q}^{m}\left\|\widetilde{\Lambda}_m\ldots\widetilde{\Lambda}_{j+1}\right\|_{L(S_N^q)}(\Delta t\left\|T^j\right\| + \left\|\Lambda\widetilde{T}^j\right\|_{H^q}).\end{aligned} \tag{4.62}$$

We must now obtain a bound on $\left\|\widetilde{T}^m\right\|_{H^q}$. To begin with, we bound the terms on the right hand side of (4.59). Combining the results of (2.73), (2.74) and (3.11), we obtain

$$\left\|\widetilde{T}^{m,i}\right\| \leq CN^{-\sigma'}\|u_0\|_{\sigma'}[\min(1, n^{\frac{1}{2}}N|a|_{0,1;[t^{m-1},t^m]}) + |a|_{[\sigma]^*,1;[t^{m-1},t^m]}], \tag{4.63}$$

where $\sigma' = \min\{K+1, \sigma\}$.

Continuing as in previous convergence proofs, we obtain the following result.

Theorem 4.6 *Suppose that the multistep method (4.27) is strongly A_0-stable, and is such that $\beta_i = 0$ for $i > 0$, and that $D_t^{p+1}u \in L^1(H)$, where the multistep scheme (4.27) has order of accuracy p. Suppose also that (3.10) and the conditions of Lemma 2.6 hold. Then, for $m = q-1, \ldots, M$,*

$$\begin{aligned}\left\|u(t^m) - u^{m,0}\right\| \leq\; & Ce^{C|a|_{[\sigma]^*,1}}\Bigg(e^{-\mu(m-q+1)\Delta t}\sum_{i=0}^{q-1}\left\|E(t^{q-1}; t^i)u(t^i) - u^{i,q-1-i}\right\| \\ & +N^{-\sigma'}[\min\left\{\frac{1}{\mu\Delta t}(1 - e^{-\mu T}), n^{\frac{1}{2}}N|a|_{0,1}\right\} + |a|_{[\sigma]^*,1}] \\ & + \Delta t^p\int_0^T e^{-\mu(T-t)}\left\|D_t^{p+1}u(t)\right\|dt\Bigg).\end{aligned} \tag{4.64}$$

4.4.3 The initialisation procedure

In all of the above analysis it has been assumed that suitable starting values have been obtained for the quantities $E(t^{q-1}; t^i)u(t^i)$, $i = 0, \ldots, q-1$. In this section we describe one procedure for obtaining approximations to these values with appropriate accuracy, which is used in each of the calculations presented in the remainder of this Thesis.

The initial data for our numerical procedure consists of the projection of the initial data for the original problem, $\Pi_N P_L u_0$. We set $\Delta t_0 = 2^{-s}\Delta t$, where s is an integer which is such that $\Delta t_0 \leq \Delta t^{q-1}$ (e.g. $s = [(q-2)|\log\Delta t|/\log 2] + 1$), and perform $q-1$ first-order time steps with timestep Δt_0. This enables us to find approximate values for $E((q-1)\Delta t_0; i\Delta t_0)u(i\Delta t_0)$, $i = 0, \ldots, q-1$; this is enough information to be able to perform a qth-order time step with a timestep of Δt_0. Enough of these are then performed to provide enough information to carry out a qth-order time step but with double the size of the timestep, and this process is repeated until a timestep of size Δt may be used.

The algorithm is illustrated in Fig. 4.3, for a fourth-order time step with $s = 3$ (so that $\Delta t_0 = \frac{\Delta t}{8}$). Time is advancing to the right, and each successive vertical level represents a step in the calculation. On any particular level, the circles represent times at which a value for the approximate solution is known at that step; solid circles indicate those values which will be used in subsequent steps. For example, at the 6th stage, the solution is calculated at times $3\frac{\Delta t}{8}, \ldots, 6\frac{\Delta t}{8}$. Values at times 0 and $\frac{\Delta t}{4}$ are carried over from previous calculations. Of these values, only those at times 0, $\frac{\Delta t}{4}$, $\frac{\Delta t}{2}$ and $3\frac{\Delta t}{4}$ are kept, and are used for the next stage of the calculation.

4.4.4 An example revisited

In Figure 4.4 the method (4.51) is used, for $q = 1, \ldots, 4$, on Problem 1 above. The final time is (as before) 1.57, and the error is plotted against the number of time steps taken; the value of ν is 10^{-2}, and the values of N and L are 63 and 64 respectively, so that spatial resolution has been obtained. For the higher values of q, the imbalance between spatial and temporal error has to some extent been redressed. Figure 4.5 plots results from the same set of experiments as Figure 4.4, but this time the error is plotted against the c.p.u. time used (on a Sun SPARC Station). The numbers of time steps used to obtain these results were as shown in the following table.

q	range
1	10–1280
2	10–640
3	10–400
4	10–240

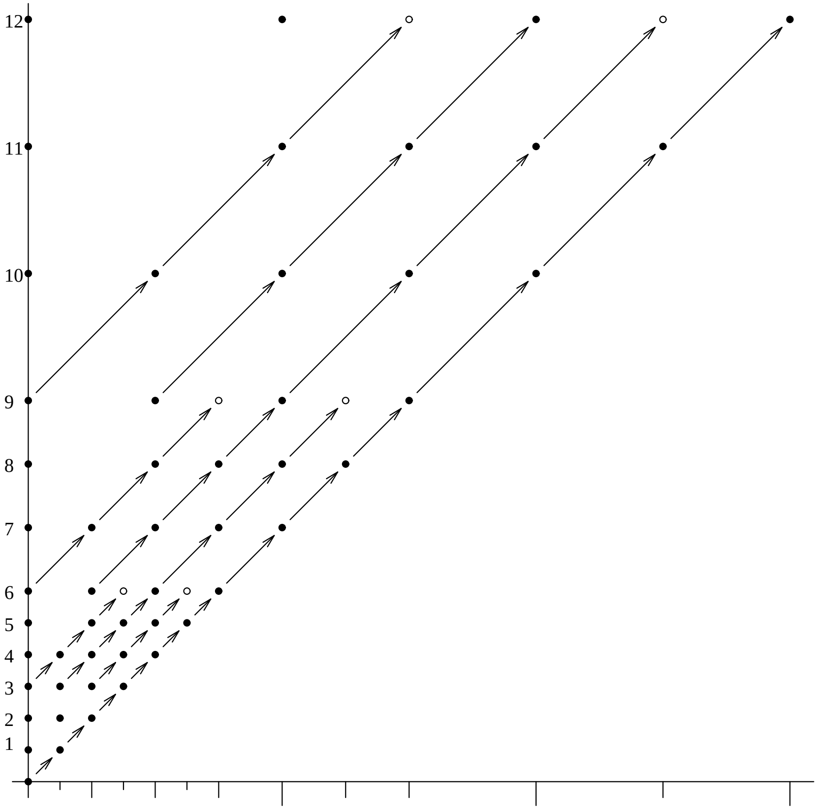

Figure 4.3: Initialisation procedure for fourth order time step.

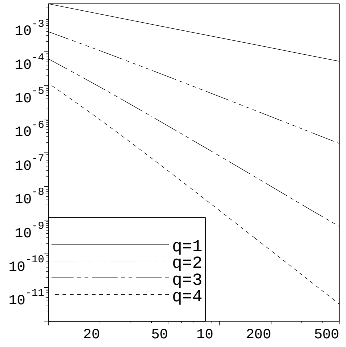

Figure 4.4: L^2 error against number of time steps for the advection-diffusion test problem.

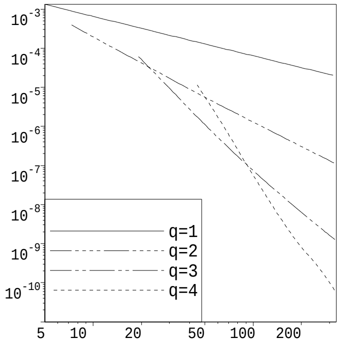

Figure 4.5: L^2 error against c.p.u. time (secs.) for the advection-diffusion test problem.

q	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
1	1					
2	$\frac{3}{2}$	$-\frac{1}{2}$				
3	$\frac{23}{12}$	$-\frac{16}{12}$	$\frac{5}{12}$			
4	$\frac{55}{24}$	$-\frac{59}{24}$	$\frac{37}{24}$	$-\frac{9}{24}$		
5	$\frac{1901}{9982}$	$-\frac{2774}{9982}$	$\frac{2616}{9982}$	$-\frac{1274}{9982}$	$\frac{251}{9982}$	
6	$\frac{4277}{1440}$	$-\frac{7923}{1440}$	$\frac{9982}{1440}$	$-\frac{7298}{1440}$	$\frac{2877}{1440}$	$-\frac{475}{1440}$

Table 4.2: Coefficients for the Adams-Bashforth formulae.

The results are only illustrative, since the measure of c.p.u. time used can vary slightly from run to run. They indicate that for maximum efficiency the order of the method should be selected with a view to the accuracy required: for high accuracy the 4th order time step is to be preferred, whereas for low to moderate accuracy the 2nd or 3rd order schemes are quite adequate.

4.5 A nonlinear example

Thus far our discussion has been restricted to linear problems. However, our eventual aim is the application of these ideas to the solution of (for example) the Navier-Stokes equations, which are, of course, non-linear. A suitable model problem for the development of techniques for the solution of the Navier-Stokes equations is the viscous Burgers equation

$$\begin{aligned} u_t + uu_x &= \nu u_{xx} \quad \text{in } (0,2\pi)\times(0,T], && \text{(4.65a)} \\ u(t=0) &= u_0 \quad \text{in } (0,2\pi). && \text{(4.65b)} \end{aligned}$$

This generalises (4.1) in that it is nonlinear, and in that the velocity field is not divergence-free. The analysis of this chapter will not apply straightforwardly to this problem. For the time being we content ourselves with experiments, although the theory does promise to be interesting and will be pursued elsewhere.

The main additional problem to be overcome here by the numerical method is the solution of the trajectory equations

$$\begin{aligned} \frac{d}{dt}X(x,t^*;t) &= u(X(x,t^*;t),t) \quad \text{in } (0,2\pi)\times(0,T], && \text{(4.66a)} \\ X(x,t^*;t^*) &= x \quad \text{in } (0,2\pi). && \text{(4.66b)} \end{aligned}$$

In order to be able to implement (4.51), we must know how to apply the operator $E(t^m;t^{m-1})$, and to do this we must know the solution to the trajectory equation (4.66), which itself depends on the solution to (4.65). The approach taken is to discretise (4.66) by an Adams-Bashforth scheme of order q, and to solve this in tandem with the discretisation of (4.65) given by (4.51). Since we seek a different set of coordinates at each time level the initial conditions are set at t^m, and so the fully-discrete equations are

$$x_\nu - X(x_\nu, t^m; t^{m-1}) = \Delta t \sum_{i=1}^{q} \gamma_i U^{m-i,i-1}(X(x_\nu, t^m; t^{m-1})), \quad 0 \le \nu \le 2L-1, \tag{4.67}$$

(where γ_i are given in Table 4.2) which must be solved by some iterative procedure, which we shall discuss below. Thus, to obtain $\{U^{m-i,i}\}_{0\le i\le q-1}$ from $\{U^{m-1-i,i}\}_{0\le i\le q-1}$ we first solve (4.67), then use the values of $X(x_\nu, t^m; t^{m-1})$ to define the operator $\Pi_N P_L E(t^m; t^{m-1})$, and finally we can solve (4.51) just as in the linear case.

4.5.1 A modified Newton method

In this section we describe the iterative method that is used to solve (4.67), which we shall write as the set of equations

$$y(x_\nu) + f(x_\nu + y(x_\nu)) = 0, \quad 0 \le \nu \le 2L-1. \tag{4.68}$$

Here $y(x_\nu) = X(x_\nu, t^m; t^{m-1}) - x_\nu$ and $f = \Delta t \sum_{i=1}^{q} \gamma_i U^{m-i,i-1}$. Now, in our case, the function f can be evaluated rapidly (by means of a FFT) at the points $\{x_\nu\}$, but the further the arguments $\{x_\nu + y(x_\nu)\}$ deviate from the set $\{x_\nu\}$, the more expensive this evaluation becomes. (This is graphically illustrated by the results from Experiment 4 in Chapter 3). With an iterative scheme for solving (4.68) we obtain sequences of points $\{y^m(x_\nu)\}$ tending towards the points $\{y(x_\nu)\}$ that solve (4.68). Each iteration, however, involves an evaluation of f whose cost is of the same order as the evaluation of f at the points $\{y(x_\nu)\}$. This is precisely because, for each ν, $y^m(x_\nu) \to y(x_\nu)$: yet it *ought* to be possible to use this convergence to reduce the cost of the later iterations, since the change between subsequent iterates is reducing to zero. This hope is what underlies the development of the modified iteration described below.

We shall first describe our method in the context of a fixed-point iteration, and then show how the approach can be applied to Newton's method, which is what we use in practice. Consider the functional equation

$$y(x) + f(x + y(x)) = 0. \tag{4.69}$$

f will be small, so y will be small and we take our first iteration to be

$$y_0(x) = -f(x).$$

This will give

$$y_0(x) + f(x + y_0(x)) \approx 0;$$

we can then use y_0 to define a new function *smaller* than f (if y_0 is a good guess),

$$f_0(x) = y_0(x) + f(x + y_0(x)).$$

Let us seek now a solution ζ to the equation

$$\zeta(x) + f_0(x + \zeta(x)) = 0. \tag{4.70}$$

If we can find such a solution then we can construct a solution to (4.69) by

$$y(x) = \zeta(x) + y_0(x + \zeta(x)). \tag{4.71}$$

However, from our point of view, (4.70) has an advantage over (4.69) since, usually, ζ will be much smaller than y, and so $f_0(x + \zeta(x))$ will be much cheaper to evaluate than $f(x + y(x))$. We have carried out, in going from (4.69) to (4.70) and (4.71), one loop in our modified iteration procedure. The full method is outlined below.

Equation to be solved

$$y(x) + f(x + y(x)) = 0. \tag{4.72a}$$

Starting step

$$\zeta_0(x) = -f(x), \tag{4.72b}$$

$$y_0(x) = \zeta_0(x), \tag{4.72c}$$

$$f_0(x) = y_0(x) + f(x + y_0(x)). \tag{4.72d}$$

Recursive step

$$\zeta_{n+1}(x) = -f_n(x), \tag{4.72e}$$

$$y_{n+1}(x) = \zeta_{n+1}(x) + y_n(x + \zeta_{n+1}(x)), \tag{4.72f}$$

$$f_{n+1}(x) = \zeta_{n+1}(x) + f_n(x + \zeta_{n+1}(x)), \tag{4.72g}$$

and it is shown by induction that

$$f_{n+1}(x) = y_{n+1}(x) + f(x + y_{n+1}(x)). \tag{4.72h}$$

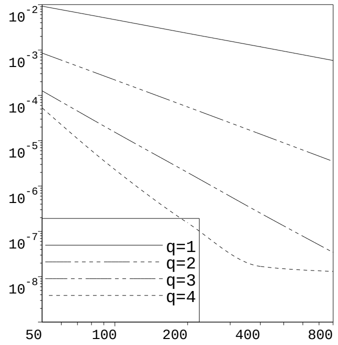

Figure 4.6: L^2 error against number of time steps for Burgers equation

When Newton's method is to be used, all that changes is that (4.72b) and (4.72e) are replaced by

$$\begin{aligned} \zeta_0(x) &= -f(x)/(1 + f'(x)), \\ \zeta_{n+1}(x) &= -f_n(x)/(1 + f_n'(x)) \end{aligned}$$

respectively.

The analysis of this procedure is awkward and is incomplete at the present time. The method is designed so that $y_n(x) \to y(x)$ for each x as $n \to \infty$, and our experience with the method suggests that this occurs (rapidly), at least when f is small.

We have described the procedure as applied to a continuous functional equation. Of course in (4.68) we are interested in the solution at only a finite number of points x_ν. We thus start by finding $\{\zeta_0(x_\nu)\}$ and $\{y_0(x_\nu)\}$. These are used to define

$$f_0 = P_L(y_0 + E_{y_0}\mathcal{P}_{K,L} f), \tag{4.73}$$

where for functions g and h, $E_g h(x) = h(x + g(x))$, and we proceed inductively by using $\{\zeta_{n+1}(x_\nu)\}$ to define

$$f_{n+1} = P_L(\zeta_{n+1} + E_{\zeta_{n+1}}\mathcal{P}_{K,L} f_n). \tag{4.74}$$

This process describes $2L$ semi-independent iterations (one for each value of ν)—they are coupled together only in as much as all of the $2L$ points are used in the definitions of the new functions f_n. (4.68) is thus solved to spectral accuracy. We note that the value of K in the right hand side of (4.74) will be very much smaller than that in the right hand side of (4.73), an observation which forms the justification for the introduction of this procedure.

4.5.2 Numerical results

The method (4.51), coupled with (4.67) solved by the Newton form of (4.72), was used to solve (4.65) in several test runs, with the same initial data as Example 1.

In Fig. 4.6 the behaviour of the error as the timestep is decreased is illustrated. The values of N and L are 63 and 64 respectively, the final time is $t = 1.57$, and the value of ν is 10^{-2}. Thus the calculation is analogous to that illustrated in Fig. 4.4. Here we also see that the high order timestepping schemes are producing the required convergence. Moreover, for the fourth order time step we see that we have indeed achieved a balance between spatial and temporal error. The spatial error here is higher than in Fig. 4.4 because the gradients in the solution are higher in this case.

Chapter 5

The Navier-Stokes equations

5.1 Introduction

As mentioned in Chapter 1, the *finite element* Lagrange-Galerkin approximation of the Navier-Stokes equations has been analysed by Pironneau [60] and subsequently by Süli in [76], where optimal-order error estimates were obtained in both the L^2 and H^1 norms and the non-linear stability of the scheme was demonstrated subject to certain constraints on the size of the timestep in terms of the mesh spacing. Our initial aim in this chapter is to describe the spectral Lagrange-Galerkin approximation of the Navier-Stokes equations, and to derive error estimates that correspond to those in [76]. The stability and convergence results obtained are subject to weaker constraints on the timestep than for the finite element version of the method. The effect of quadrature in the nonlinear case will be the subject of future investigation.

We begin by introducing some of the notation that will be employed in the remainder of the chapter. Section 2 describes the incompressible Navier-Stokes equations and recalls some well known existence and uniqueness results. The concept of Lagrangian coordinates is also recalled and the Lagrange-Galerkin time discretisation is described. Finally the fully discrete method is presented, based on a spectral spatial discretisation. Section 3 is devoted to the convergence analysis of the spectral Lagrange-Galerkin method, beginning with some technical lemmata followed by statements and proofs of the error estimates. Convergence is shown directly; no particular concept of stability is explicitly referred to. However, the stability of the method is still important, and so is dealt with in the next section. There we briefly discuss the concept of non-linear stablity employed, and state and prove the relevant stability result. The penultimate section describes the application of the method to the two-dimensional Navier-Stokes equations. The particular setting is a study of a perturbed double shear layer. We end this chapter by setting out the conclusions of the Thesis and indicating areas of future development for the work.

5.1.1 Notation

We begin with some definitions specific to this chapter. We define $\dot{L}^2(\Omega)$ (resp. $\dot{H}^s_{\#}(\Omega)$, $s \geq 0$) as the set of all functions u in $L^2(\Omega)$ (resp. $H^s_{\#}(\Omega)$) such that $\int_\Omega u dx = 0$.

Let $\nu = (\nu_1, \ldots, \nu_n)$ denote the unit outward normal to $\partial\Omega$ and let Γ_i (resp. Γ_{i+n}) be the face of Ω with $\nu_i = 1$ (resp. $\nu_i = -1$). Two spaces central to our discussion will be

$$H := \{u \in \dot{L}^2(\Omega)^n | \nabla \cdot u = 0, u_{i|\Gamma_i} = -u_{i|\Gamma_{i+n}}, i = 1, \ldots, n\},$$

where the latter condition is that $u \cdot \nu$, with ν the outward normal, is periodic (the trace of $u \cdot \nu$ on $\partial\Omega$ exists when $u \in L^2(\Omega)^n$ and $\nabla \cdot u \in L^2(\Omega)$, [86]); and

$$V := \{u \in \dot{H}^1_{\#}(\Omega)^n | \nabla \cdot u = 0\}.$$

For $s \geq 1$ we set $V_s = H^s_{\#}(\Omega)^n \cap V$. We define a norm on V_s equivalent to that induced by $H^s_{\#}(\Omega)^n$ by

$$\|u\|_{V_s} = \left(\sum_p |p|^{2s} |\hat{u}(p)|^2 \right)^{1/2}.$$

Finally we denote by $C^{0,1} = C^{0,1}(\overline{\Omega})$ the space of Lipshitz-continuous functions on the closure of Ω.

5.2 The equations and their approximation

5.2.1 Statement of the problem

The Navier-Stokes equations for a viscous incompressible fluid in two or three space dimensions take the form

$$u_t + (u \cdot \nabla)u - \nu\Delta u + \nabla p = f \quad \text{on } \Omega \times (0,T), \tag{5.1a}$$

$$\nabla \cdot u = 0 \quad \text{on } \Omega \times (0,T), \tag{5.1b}$$

$$u(x,0) = u_0(x) \quad \text{on } \Omega, \tag{5.1c}$$

where $u(x,t)$ is the velocity of the fluid, $p(x,t)$ the kinematic pressure, ν the kinematic viscosity, $f(x,t)$ the density of body force per unit mass, and u_0 the initial velocity. The problem is completed in this case by periodic boundary conditions

$$u(x + 2\pi e_i, t) = u(x,t) \quad \forall x \in R^n, \ \forall t \in [0,T], \tag{5.1d}$$

where $e_1, \ldots, e_n$ is the canonical basis for R^n. We assume that the mean flow is zero.

The following functional-analytic formulation of the problem is due to J. Leray. Details may be found in Temam's works [85, 86].

For f and u_0 given,

$$f \in L^2(H), \tag{5.2a}$$

$$u_0 \in V, \tag{5.2b}$$

find the *strong solution* u satisfying

$$u \in L^2(H^2_{\#}(\Omega)^n) \cap L^\infty(V), \tag{5.2c}$$

$$\frac{d}{dt}(u,v) + ((u \cdot \nabla)u, v) + \nu(\nabla u, \nabla v) = (f,v) \quad \forall v \in V, \tag{5.2d}$$

and

$$u(0) = u_0. \tag{5.2e}$$

The following theorem encapsulates some standard existence and uniqueness results for this problem. Again we refer to [85, 86] for further details.

Theorem 5.1 *For $n = 2$ there exists a unique solution to problem (5.2) satisfying*

$$u \in L^2(H^2_{\#}(\Omega)^n) \cap C(V) \cap H^1(H). \tag{5.3}$$

For $n = 3$, given $f \in L^\infty(H)$ instead of (5.2a), the same result holds for T small enough.

Under the assumptions of Theorem 5.1, $D_t u = \partial u/\partial t + (u \cdot \nabla)u$, the material derivative of u, belongs to $L^2(L^2(\Omega)^n)$, so that (5.2d) may be rewritten as

$$(D_t u, v) + \nu(\nabla u, \nabla v) = (f,v) \quad \forall v \in V. \tag{5.4}$$

The crucial aspect of the Lagrange-Galerkin approach is the discretisation of the material derivative along particle trajectories. To this end we seek to cast the equations into a Lagrangian form.

5.2.2 Lagrangian form

We assume here that the solution u of (5.2) satisfies, in addition to (5.3),

$$u \in C(C^{0,1}(\overline{\Omega})^n). \tag{5.5}$$

Using von Neuman's measurable selection theorem, it is possible to give a description of a Lagrangian representation of the flow under weaker assumptions on u (see for example [87]). Since, however, we shall need u to satisfy (5.5) for most of the results in this chapter, we are happy to assume it here.

For $x \in R^n$, $t \in [0,T]$, let $X_u(x,t;\cdot)$ denote the trajectory of the particle of fluid whose motion is governed by the velocity field u and which is at position x at time t. Then $X_u(x,t;\cdot)$ is the solution of the initial value problem

$$\frac{d}{ds}X_u(x,t;s) = u(X_u(x,t;s),s), \quad s \in [0,T]\backslash\{t\}, \tag{5.6a}$$

$$X_u(x,t;t) = x. \tag{5.6b}$$

The map $x \to X_u(x,t;s)$ is, for each $s,t \in [0,T]$, an isometric homeomorphism from R^n onto itself, and by virtue of Theorem 2.1 it is differentiable almost everywhere. Moreover, since u is divergence-free, the map has the volume-preserving property, i.e. its Jacobian is equal to 1 almost everywhere.

Associated with $X_u(x,t;s)$, for each $t,s \in [0,T]$, we can define the linear operator $E_u(t;s)$ on $\dot{H}^0_{\#}(\Omega)^n$ (which coincides exactly with $\dot{L}^2(\Omega)^n$) as follows. Each $v \in \dot{H}^0_{\#}(\Omega)^n$ may be periodically extended to the whole of R^n. We then define $(E_u(t;s)v)(x) := v(X_u(x,t;s))$, $x \in R^n$. The restriction of $E_u(t;s)v$ to Ω is in $\dot{H}^0_{\#}(\Omega)^n$. Moreover $E_u(t;s)$ is invertible (with inverse $E_u(s;t)$) and, because of the volume preserving property,

$$\|E_u(t;s)v\| = \|v\|.$$

Under the assumptions of Theorem 5.1, together with the additional assumption (5.5), we may replace the term $D_t u$ in (5.4) by $\left(\frac{d}{ds}u(X_u(\cdot,t;s),s)\right)\Big|_{s=t}$. The first step in obtaining the discrete equations described in the next subsection is to replace this time derivative by a first-order backward difference formula. Higher order backward difference formulae may also be used; their properties in the non-linear case will be the subject of future investigations.

5.2.3 Spectral Lagrange-Galerkin approximation

Having described how we will discretise (5.4) in time, we form our fully discrete equations by employing a spectral Galerkin method for our spatial discretisation. Thus we seek $\mathcal{U} = (U^0,\ldots,U^M)^T \in (\Pi_N V)^{M+1}$ satisfying

$$\begin{aligned} \frac{1}{\Delta t}(U^k,v) + \nu(\nabla U^k,\nabla v) &= \frac{1}{\Delta t}(E_{U^{k-1}}(t^k;t^{k-1})U^{k-1},v) + (f^k,v) \\ &\quad \forall v \in V_N = \Pi_N V \subset V, \quad k = 1,\ldots,M, \end{aligned} \tag{5.7}$$

with $U^0 = \Pi_N u_0$, where, for $x \in \Omega$ and $t \in [0,T]$, we define $X_{U^{k-1}}(x,t;s)$ by

$$\frac{d}{ds}X_{U^{k-1}}(x,t;s) = U^{k-1}(X_{U^{k-1}}(x,t;s)), \quad s \in [0,T]\backslash\{t\}, \tag{5.8a}$$

$$X_{U^{k-1}}(x,t;t) = x, \tag{5.8b}$$

and then $E_{U^{k-1}}(t;s)$ is defined by

$$\left(E_{U^{k-1}}(t;s)w\right)(x) = w(X_{U^{k-1}}(x,t;s)), \quad x \in \Omega.$$

We note that, similarly to $E_u(t;s)$, $E_{U^{k-1}}(t;s)$ is volume-preserving and invertible (with inverse $E_{U^{k-1}}(s;t)$). These properties will be used repeatedly in the subsequent analysis.

Given the solution U^{k-1} at time t^{k-1}, $X_{U^{k-1}}(x,t^k;t^{k-1})$ is found by solving (5.8) to whatever accuracy is required. The right hand side of (5.7) is then well defined, giving a diagonal system to solve for U^k. In practice, the solution of (5.8) is carried out by using a forward Euler discretisation to reduce it to a functional equation, which is solved at equally spaced values of x by Newton iteration. These values of x are the quadrature points (for the trapezium rule) which are then used in the evaluation of the integrals implicit in the right hand side of (5.7). An additional approximation is employed, taking the form of a piecewise Chebyshev interpolant of U^{k-1}, in order to speed up the evaluation of $E_{U^{k-1}}(t^k;t^{k-1})U^{k-1}$. The effect of these approximations is analysed in the linear case in [79] and shown to be minimal. Their effect in the nonlinear case will be the subject of future investigation.

5.3 Error analysis

In what follows, $C_1, C_2, \ldots$ will stand for specific constants whose values will often be given explicitly. They will be different from those constants appearing in previous chapters. For simplicity of presentation, we shall sometimes write $u(\cdot,t)$ as $u(t)$, and we define without ambiguity, for $k = 1,\ldots,M$,

$$\begin{aligned} u^k &:= u(t^k), \\ E_u &:= E_u(t^k;t^{k-1}), \\ E_{U^{k-1}} &:= E_{U^{k-1}}(t^k;t^{k-1}), \\ X_u &:= X_u(x,t^k;t^{k-1}), \\ X_{U^{k-1}} &:= X_{U^{k-1}}(x,t^k;t^{k-1}). \end{aligned}$$

We set $\xi^k = \Pi_N u^k - U^k$, $\eta^k = (I - \Pi_N)u^k$ and $\zeta^k = \xi^k + \eta^k$, and derive, from (5.4) and (5.7),

$$\begin{aligned} &\frac{1}{\Delta t}(\xi^k,v) + \nu(\nabla\xi^k,\nabla v) \\ &= \frac{1}{\Delta t}(E_{U^{k-1}}\xi^{k-1},v) + \frac{1}{\Delta t}((E_u - E_{U^{k-1}})u^{k-1},v) \\ &+ \frac{1}{\Delta t}(E_{U^{k-1}}\eta^{k-1},v) + (\frac{1}{\Delta t}(u^k - E_u u^{k-1}) - D_t u^k,v) \quad \forall v \in V_N. \end{aligned} \tag{5.9}$$

Our first error estimate will be obtained from (5.9), taking $v = \xi^k$. First, however, we present some preliminary results.

Lemma 5.1 *Let* $v \in \dot{S}_N = \Pi_N \dot{L}^2(\Omega)^n$, $n = 1,2,3$. *Then*

$$\|v\|_\infty \le D_n(N)\,\|v\|_1,$$

where

$$\begin{aligned} D_1(N) &\le \pi^{1/2}, \\ D_2(N) &\le [\pi(2 + 4\ln(1+N^2))]^{1/2}, \\ D_3(N) &\le [\pi(9 + 12\ln(1+N^2) + 32\sqrt{3N})]^{1/2}. \end{aligned}$$

Proof . We note that

$$\begin{aligned} \|v\|_\infty &\le \sum_{0 \ne |p|_\infty \le N} |\hat{v}(p)| \\ &\le \left[\sum_{0 \ne |p|_\infty \le N} (1 + |p|^2)^{-1}\right]^{1/2} \|v\|_1 \\ &=: D_n(N)\,\|v\|_1, \end{aligned}$$

and the result follows readily.

Lemma 5.2 *For* $k = 1,\ldots,M$, *let* $c_k = \left|u^k\right|_{C^{0,1}}$*: then*

$$\|X_u - X_{U^{k-1}}\| \le \frac{1}{c_{k-1}}(e^{c_{k-1}\Delta t} - 1)(\left\|\zeta^{k-1}\right\| + \left\|\frac{du}{dt}\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)}). \tag{5.10}$$

Proof Integrating (5.6) and the corresponding equation for $X_{U^{k-1}}$ between $t^k - t$ and t^k we obtain

$$\begin{aligned} &(X_u - X_{U^{k-1}})(\cdot,t^k,t^k - t) \\ &= \int_{t^k}^{t^k - t} E_u(t^k;s)u(s) - E_{U^{k-1}}(t^k;s)U^{k-1}ds \\ &= -\int_0^t \{E_{U^{k-1}}(t^k;t^k - s)(u^{k-1} - U^{k-1}) \\ &+ (E_u - E_{U^{k-1}})(t^k;t^k - s)u^{k-1} \\ &+ E_u(t^k;t^k - s)(u(t^k - s) - u^{k-1})\}ds, \end{aligned}$$

so that

$$\begin{aligned} &\left\|(X_u - X_{U^{k-1}})(\cdot,t^k,t^k - t)\right\| \\ &\le \int_0^t \{\left\|E_{U^{k-1}}(t^k;t^k - s)(u^{k-1} - U^{k-1})\right\| \\ &+ \left\|(E_u - E_{U^{k-1}})(t^k;t^k - s)u^{k-1}\right\| \\ &+ \left\|E_u(t^k;t^k - s)(u(t^k - s) - u^{k-1})\right\|\}ds \\ &\le t(\left\|\zeta^{k-1}\right\| + \left\|\frac{du}{dt}\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)}) \\ &+ c_{k-1}\int_0^t \left\|(X_u - X_{U^{k-1}})(\cdot,t^k,t^k - s)\right\| ds. \end{aligned}$$

A straightforward application of Gronwall's lemma yields the required result.

Lemma 5.3 *For* $k = 1,\ldots,M$, $v \in V$, *and* $t,s \in [0,T]$,

$$\|(E_{U^{k-1}}(t;s) - I)v\| \le |t - s|\left\|U^{k-1}\right\|_\infty \|\nabla v\|. \tag{5.11}$$

Proof We have

$$\begin{aligned}(E_{U^{k-1}}(t;s) - I)v &= \int_t^s \frac{d}{d\tau}\left(E_{U^{k-1}}(t;\tau)v\right) d\tau \\ &= \int_t^s E_{U^{k-1}}(t;\tau)\left(U^{k-1}\cdot\nabla v\right) d\tau.\end{aligned}$$

The result then follows by taking the norm of both sides.

The final lemma deals with the time truncation error in the backward Euler approximation of the material derivative along particle trajectories.

Lemma 5.4 *For $k = 1, \dots, M$,*

$$\left\|(D_t u)^k - \frac{u^k - E_u u^{k-1}}{\Delta t}\right\| \leq \left\|D_t^2 u\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)}. \tag{5.12}$$

Proof

$$\begin{aligned}&\frac{u^k - E_u u^{k-1}}{\Delta t} - (D_t u)^k \\ &= \frac{1}{\Delta t}\int_{t^{k-1}}^{t^k} \frac{d}{ds}E_u(t^k;s)u(s) - \frac{d}{dt}E_u(t^k;t)u(t)\Big|_{t=t^k} ds \\ &= -\frac{1}{\Delta t}\int_{t^{k-1}}^{t^k}\int_s^{t^k} \frac{d^2}{dt^2}E_u(t^k;t)u(t)dt\, ds \\ &= -\frac{1}{\Delta t}\int_{t^{k-1}}^{t^k}(t - t^{k-1})\frac{d^2}{dt^2}E_u(t^k;t)u(t)dt,\end{aligned}$$

and the result follows.

5.3.1 L^2 error estimates

We assume that

$$f \in C(H) \quad \text{and} \quad u_0 \in C^{0,1} \cap V_s, \quad s > n/2, \tag{5.13}$$

and that the corresponding solution u of (5.2) satisfies

$$\begin{aligned}u &\in C(C^{0,1} \cap V_s), \quad s > n/2, & (5.14a)\\ du/dt &\in L^2(H) \quad \text{and} & (5.14b)\\ D_t^2 u &\in L^2(H). & (5.14c)\end{aligned}$$

Then we have the following

Theorem 5.2 *Suppose that u satisfies (5.2), $\mathcal{U}$ satisfies (5.7), and that (5.13) and (5.14) hold. Then there exist two positive constants Δt_0 and N_0 such that, for all $\Delta t \leq \Delta t_0$ and all $N \geq N_0$,*

$$\|u - U\|_{l^\infty(H)} \leq C_3(C_4 N^{-s} + C_5 \Delta t) \tag{5.15}$$

where

$$C_3 = \exp\{2\,|u|^2_{l^2(C^{0,1})} + |u|_{l^1(C^{0,1})})\},$$

$$C_4 = \|u\|_{L^\infty(V_s)} + \left(\frac{4C_1^2\,\|u\|^2_{L^\infty(V_s)}}{\nu} + 1\right)^{1/2}\|u\|_{l^2(V_s)},$$

$$C_5 = \left(\left\|\frac{du}{dt}\right\|_{L^2(H)} + \left(\frac{2}{\nu}\right)^{1/2}\left\|D_t^2 u\right\|_{L^2(H)}\right),$$

and C_1 is a positive constant which depends only on s.

The remainder of this subsection will be taken up with the proof of this theorem.

Proof We choose $v = \xi^k$ in (5.9) to obtain

$$\begin{aligned}&\frac{1}{\Delta t}\left\|\xi^k\right\|^2 + \nu\left\|\nabla\xi^k\right\|^2 \\ \leq\;& \frac{1}{\Delta t}\left\|\xi^k\right\|\left\|\xi^{k-1}\right\| + \frac{1}{\Delta t}\left\|(E_u - E_{U^{k-1}})u^{k-1}\right\|\left\|\xi^k\right\| \\ +\;& \frac{1}{\Delta t}|(E_{U^{k-1}}\eta^{k-1},\xi^k)| + \left\|\frac{u^k - E_u u^{k-1}}{\Delta t} - D_t u^k\right\|\left\|\xi^k\right\| \\ =\;& A_1 + \ldots + A_4.\end{aligned} \tag{5.16}$$

We shall bound the terms A_2 to A_4 in turn. We have, for A_2,

$$\left\|(E_u - E_{U^{k-1}})u^{k-1}\right\| \leq c_{k-1}\left\|X_u - X_{U^{k-1}}\right\|,$$

and so, from Lemma 5.2,

$$\begin{aligned}A_2 \;\leq\;& \frac{1}{\Delta t}(e^{c_{k-1}\Delta t} - 1)\left(\left\|\zeta^{k-1}\right\| + \left\|\frac{du}{dt}\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)}\right)\left\|\xi^k\right\| \\ \leq\;& \left\|\xi^k\right\|\left\{\frac{1}{\Delta t}(e^{c_{k-1}\Delta t} - 1)\;\left\|\xi^{k-1}\right\|\right. \\ &\left.+c_{k-1}e^{c_{k-1}\Delta t}\left(\left\|\eta^{k-1}\right\| + \left\|\frac{du}{dt}\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)}\right)\right\},\end{aligned}$$

where we have written ζ^{k-1} as $\xi^{k-1} + \eta^{k-1}$.

In order to bound A_3 we make use of the fact that E_u and $E_{U^{k-1}}$ are invertible and volume preserving. Thus we have, applying Lemma 5.1 and Lemma 5.3,

$$\begin{aligned}A_3 \;\leq\;& \frac{1}{\Delta t}|(\eta^{k-1},(E^{-1}_{U^{k-1}} - I)\xi^k)| \\ \leq\;& \left\|\eta^{k-1}\right\|\left\|U^{k-1}\right\|_\infty\left\|\nabla\xi^k\right\| \\ \leq\;& \left\|\eta^{k-1}\right\|\left\|\nabla\xi^k\right\|\left(\left\|\Pi_N u^{k-1}\right\|_\infty + D_n(N)\left\|\nabla\xi^{k-1}\right\|\right).\end{aligned}$$

Now, by Sobolev's imbedding theorem and the contractivity of Π_N in V_r for any real r, it follows that there is a constant C_1, independent of N but dependent on $s > n/2$, such that $\|\Pi_N u\|_{l^\infty(L^\infty)} \leq C_1\|u\|_{L^\infty(V_s)} =: C_2$. Including the bound on A_4 which follows from Lemma 5.4 we find, repeatedly making use of the inequality $ab \leq \frac{\epsilon}{2}a^2 + \frac{1}{2\epsilon}b^2$ (valid for all real a, b and for $\epsilon > 0$) and gathering like terms, that

$$\begin{aligned}&\frac{1}{\Delta t}\left\|\xi^k\right\|^2 + \nu\left\|\nabla\xi^k\right\|^2 \\ \leq\;& \frac{e^{c_{k-1}\Delta t}}{\Delta t}\left\|\xi^k\right\|\left\|\xi^{k-1}\right\| + \left\|\nabla\xi^k\right\|\left\|\eta^{k-1}\right\|\left(C_2 + D_n(N)\left\|\nabla\xi^{k-1}\right\|\right) \\ +\;& \left\|\xi^k\right\|\left\{c_{k-1}e^{c_{k-1}\Delta t}\left(\left\|\frac{du}{dt}\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)} + \left\|\eta^{k-1}\right\|\right)\right. \\ +\;& \left.\left\|D_t^2 u\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)}\right\} \\ \leq\;& \frac{1}{2\Delta t}\left(\left\|\xi^k\right\|^2 + e^{2c_{k-1}\Delta t}\left\|\xi^{k-1}\right\|^2\right) + c^2_{k-1}\left\|\xi^k\right\|^2 + \frac{\nu}{2}\left\|\nabla\xi^k\right\|^2 \\ +\;& \frac{e^{2c_{k-1}\Delta t}}{2}\left(\left\|\frac{du}{dt}\right\|^2_{L^1(t^{k-1},t^k;L^2(\Omega)^n)} + \left\|\eta^{k-1}\right\|^2\right) \\ +\;& \frac{2(\left\|\eta^{k-1}\right\|D_n(N))^2}{\nu}\left\|\nabla\xi^{k-1}\right\|^2 \\ +\;& \frac{1}{\nu}\left\|D_t^2 u\right\|^2_{L^1(t^{k-1},t^k;L^2(\Omega)^n)} + \frac{2C_2^2}{\nu}\left\|\eta^{k-1}\right\|^2,\end{aligned}$$

where we have also made use of the fact that $\|v\| \leq \|\nabla v\|$ for $v \in V$.

Because of the smoothness of u, there is an N_0 such that for $N \geq N_0$,

$$2\left(\|\eta\|_{C(H)}D_n(N)\right)^2 \leq \frac{\nu^2}{4}.$$

We assume that we are dealing with such an N, so that multiplying through by $2\Delta t$ we obtain

$$\begin{aligned}(1 - 2\Delta t c_{k-1}^2) \left\|\xi^k\right\|^2 + \nu\Delta t \left\|\nabla\xi^k\right\|^2 &\leq e^{2c_{k-1}\Delta t} \left\|\xi^{k-1}\right\|^2 \\ &+\Delta t \left(e^{2c_{k-1}\Delta t} + \frac{4C_2^2}{\nu}\right) \left\|\eta^{k-1}\right\|^2 + \frac{\nu\Delta t}{2} \left\|\nabla\xi^{k-1}\right\|^2 \\ &+\Delta t^2 \left(e^{2c_{k-1}\Delta t} \left\|\frac{du}{dt}\right\|^2_{L^2(t^{k-1},t^k;L^2(\Omega)^n)} + \frac{2}{\nu} \left\|D_t^2 u\right\|^2_{L^2(t^{k-1},t^k;L^2(\Omega)^n)}\right).\end{aligned} \tag{5.17}$$

We proceed under the assumption that $\Delta t \leq \Delta t_0$ where

$$\Delta t_0 = \min\left(T, \frac{1}{4\left|u\right|^2_{l^\infty(C^{0,1})}}\right). \tag{5.18}$$

For such Δt, we can bound $(1 - 2\Delta t c_{k-1}^2)$ from below by $e^{-4\Delta t c_{k-1}^2}$. Then we deduce from (5.17) that

$$\begin{aligned}&\left\|\xi^k\right\|^2 + e^{4\Delta t c_{k-1}^2}\nu\Delta t \left\|\nabla\xi^k\right\|^2 \\ \leq\ & e^{4\Delta t c_{k-1}^2 + 2\Delta t c_{k-1}} \left\{\left\|\xi^{k-1}\right\|^2 + \frac{\nu\Delta t}{2} e^{4\Delta t c_{k-2}^2} \left\|\nabla\xi^{k-1}\right\|^2\right. \\ +\ & \Delta t \left(1 + \frac{4C_2^2}{\nu}\right) \left\|\eta^{k-1}\right\|^2 \\ +\ & \left.\Delta t^2 \left\|\frac{du}{dt}\right\|^2_{L^2(t^{k-1},t^k;L^2(\Omega)^n)} + \frac{2\Delta t^2}{\nu} \left\|D_t^2 u\right\|^2_{L^2(t^{k-1},t^k;L^2(\Omega)^n)}\right\}.\end{aligned} \tag{5.19}$$

This inequality is valid for all $k \geq 1$ if we set $c_{-1} = 0$. It is slightly less sharp than (5.17), since we have multiplied some of the terms on the right hand side of (5.17) by terms greater than unity. Our reason for doing this is to enable us to make use of the following lemma, whose proof (by induction) is straightforward.

Lemma 5.5 *Let $\{a_k\}$, $\{b_k\}$, $\{d_k\}$ and $\{\lambda_k\}$ be sequences of non-negative real numbers satisfying, for $k = 1, \ldots, M$,*

$$a_k + b_k \leq e^{\lambda_{k-1}} \left(a_{k-1} + \frac{1}{2} b_{k-1} + d_{k-1}\right). \tag{5.20}$$

Then the following inequality holds:

$$a_M + \frac{1}{2}\sum_{k=0}^{M} b_k \leq \left(\prod_{k=0}^{M-1} e^{\lambda_k}\right) \left(a_0 + \frac{b_0}{2} + \sum_{k=0}^{M-1} d_k\right). \tag{5.21}$$

If we denote $2\Delta t \left(\sum_{k=0}^{M-1} 2c_k^2 + c_k\right)$ by α, then from (5.19)–(5.21) we deduce that

$$\begin{aligned}&\left\|\xi^M\right\|^2 + \frac{\nu\Delta t}{2} \sum_{k=1}^{M} e^{4\Delta t c_{k-1}^2} \left\|\nabla\xi^k\right\|^2 \\ &\leq e^\alpha \sum_{k=1}^{M} \Delta t \left(1 + \frac{4C_2^2}{\nu}\right) \left\|\eta^{k-1}\right\|^2 \\ &+\Delta t^2 e^\alpha \left(\left\|\frac{du}{dt}\right\|^2_{L^2(L^2(\Omega)^n)} + \frac{2}{\nu} \left\|D_t^2 u\right\|^2_{L^2(L^2(\Omega)^n)}\right).\end{aligned} \tag{5.22}$$

Since the same bound holds for $\left\|\xi^k\right\|^2$, for any $k \leq M$, the conclusions of the theorem follow.

Remark. The result is such that, given sufficient smoothness and decay of u, it is valid for a semi-infinite time interval.
Remark. (5.22) also provides us with a bound on the l^2 norm of $\nabla\xi$, which we encapsulate in the following lemma, and will make use of in the next subsection.

Lemma 5.6 *Under the conditions of Theorem 5.2, and with C_4 and C_5 as defined there, we have*

$$\left\|\nabla\xi^k\right\|_{l^2(H)} \leq \left(\frac{2}{\nu}\right)^{1/2} C_3 \left(C_4 N^{-s} + C_5\Delta t\right). \tag{5.23}$$

5.3.2 Error estimates in V

Theorem 5.3 *Suppose that the conditions of Theorem 5.2 hold. Suppose also that N and Δt are related by*

$$\Delta t = \begin{cases} O((\ln N)^{-1/2}), & n = 2 \\ O(N^{-1/2}), & n = 3, \end{cases} \tag{5.24}$$

and that Δt_0, N_0, C_1, C_3, C_4 and C_5 are as in Theorem 5.2. Then, for $\Delta t \leq \Delta t_0$ and $N \geq N_0$,

$$\|u - U\|_{l^\infty(V)} \leq C_6(C_7 N^{1-s} + C_8\Delta t) \tag{5.25}$$

where

$$C_6 = \exp\left\{\frac{1}{\nu}\left(C_1^2 \|u\|^2_{l^2(V_s)} + C_9^2\right)\right\},$$

$$\begin{aligned}C_7 \quad &= \quad \|u\|_{l^\infty(V_s)} \left(1 + \left(\frac{5}{\nu}\right)^{1/2} \left(C_1 \|u\|_{l^2(V_s)} + C_9\right)\right) \\ &+ \left(\frac{5}{\nu}\right)^{1/2} |u|_{l^2(C^{0,1})} C_3 C_4 e^{\Delta t_0 |u|_{l^\infty(C^{0,1})}},\end{aligned}$$

$$\begin{aligned}C_8 \quad &= \quad \left(\frac{5}{\nu}\right)^{1/2} e^{\Delta t_0 |u|_{l^\infty(C^{0,1})}} \left(|u|_{l^\infty(C^{0,1})} \left\|\frac{du}{dt}\right\|_{L^2(L^2(\Omega)^n)} + C_3 C_5 |u|_{l^2(C^{0,1})}\right) \\ &+ \left(\frac{5}{\nu}\right)^{1/2} \left\|D_t^2 u\right\|_{L^2(L^2(\Omega)^n)}\end{aligned}$$

and C_9 is an upper bound on $D_n(N) \left\|\nabla\xi^k\right\|_{l^2(H)}$.

Remark. The constant C_9 is guaranteed to exist and to be independent of N and Δt by Lemma 5.6 and (5.24).

Proof We begin by considering (5.9), adding $-\frac{1}{\Delta t}(\xi^{k-1}, v)$ to both sides, and taking $v = \frac{\xi^k - \xi^{k-1}}{\Delta t}$, giving

$$\begin{aligned}&\left\|\frac{\xi^k - \xi^{k-1}}{\Delta t}\right\|^2 + \frac{\nu}{2\Delta t}\left\{\left\|\nabla\xi^k\right\|^2 - \left\|\nabla\xi^{k-1}\right\|^2\right\} \\ &\leq \left\|\frac{\xi^k - \xi^{k-1}}{\Delta t}\right\| \left\{\frac{1}{\Delta t}\left\|(E_{U^{k-1}} - I)\xi^{k-1}\right\| + \frac{1}{\Delta t}\left\|(E_u - E_{U^{k-1}})u^{k-1}\right\|\right. \\ &\left.+\frac{1}{\Delta t}\left\|(E_{U^{k-1}} - I)\eta^{k-1}\right\| + \left\|\frac{u^k - E_u u^{k-1}}{\Delta t} - D_t u^k\right\|\right\}.\end{aligned} \tag{5.26}$$

We deal with each of the terms on the right hand side of (5.26) in turn to obtain

$$\begin{aligned}&\frac{1}{\Delta t}\left\|(E_{U^{k-1}} - I)\xi^{k-1}\right\| \\ &\leq \quad \left\|U^{k-1}\right\|_\infty \left\|\nabla\xi^{k-1}\right\| \\ &\leq \quad \left(C_1 \left\|u^{k-1}\right\|_{V_s} + D_n(N)\left\|\nabla\xi^{k-1}\right\|\right)\left\|\nabla\xi^{k-1}\right\|,\end{aligned} \tag{5.27a}$$

$$\begin{aligned}&\frac{1}{\Delta t}\left\|(E_u - E_{U^{k-1}})u^{k-1}\right\| \\ &\leq \quad c_{k-1}e^{c_{k-1}\Delta t}\left(\left\|\zeta^{k-1}\right\| + \left\|\frac{du}{dt}\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)}\right),\end{aligned} \tag{5.27b}$$

$$\begin{aligned}&\frac{1}{\Delta t}\left\|(E_{U^{k-1}} - I)\eta^{k-1}\right\| \\ &\leq \quad \left\|U^{k-1}\right\|_\infty \left\|\nabla\eta^{k-1}\right\| \\ &\leq \quad \left(C_1 \left\|u^{k-1}\right\|_{V_s} + D_n(N)\left\|\nabla\xi^{k-1}\right\|\right)\left\|\nabla\eta^{k-1}\right\|,\end{aligned} \tag{5.27c}$$

$$\begin{aligned}&\left\|\frac{u^k - E_u u^{k-1}}{\Delta t} - D_t u^k\right\| \\ &\leq \quad \left\|D_t^2 u\right\|_{L^1(t^{k-1},t^k;L^2(\Omega)^n)}.\end{aligned} \tag{5.27d}$$

Substituting (5.27a)–(5.27d) into (5.26) yields

$$\begin{aligned}
&\frac{\nu}{2\Delta t}\left\{\left\|\nabla\xi^k\right\|^2 - \left\|\nabla\xi^{k-1}\right\|^2\right\} \\
&\le \left(\left\|\nabla\xi^{k-1}\right\|^2 + \frac{5}{2}\left\|\nabla\eta^{k-1}\right\|^2\right)\left(C_1^2\left\|u^{k-1}\right\|_{V_s}^2 + D_n(N)^2\left\|\nabla\xi^{k-1}\right\|^2\right) \\
&+ \frac{5}{2}c_{k-1}^2 e^{2c_{k-1}\Delta t}\left(\left\|\zeta^{k-1}\right\|^2 + \Delta t\left\|\frac{du}{dt}\right\|_{L^2(t^{k-1},t^k;L^2(\Omega)^n)}^2\right) \\
&+ \frac{5}{2}\Delta t\left\|D_t^2 u\right\|_{L^2(t^{k-1},t^k;L^2(\Omega)^n)}^2 . \qquad (5.28)
\end{aligned}$$

Let us write

$$\lambda_{k-1} = \frac{2\Delta t}{\nu}\left(C_1^2\left\|u^{k-1}\right\|_{V_s}^2 + D_n(N)^2\left\|\nabla\xi^{k-1}\right\|^2\right).$$

Then, multiplying (5.28) by $2\Delta t/\nu$, we obtain

$$\begin{aligned}
\left\|\nabla\xi^k\right\|^2 &\le e^{\lambda_{k-1}}\left(\left\|\nabla\xi^{k-1}\right\|^2 + \frac{5\lambda_{k-1}}{2}\left\|\nabla\eta^{k-1}\right\|^2\right) \\
&+ \frac{5\Delta t}{\nu}e^{\lambda_{k-1}}c_{k-1}^2 e^{2c_{k-1}\Delta t}\left(\left\|\zeta^{k-1}\right\|^2 + \Delta t\left\|\frac{du}{dt}\right\|_{L^2(t^{k-1},t^k;L^2(\Omega)^n)}^2\right) \\
&+ \frac{5\Delta t^2}{\nu}e^{\lambda_{k-1}}\left\|D_t^2 u\right\|_{L^2(t^{k-1},t^k;L^2(\Omega)^n)}^2 .
\end{aligned}$$

A simple application of Lemma 5.5 completes the proof of Theorem 5.3.

5.4 Stability

Although in the previous section we have demonstrated the convergence of the spectral Lagrange-Galerkin method for the Navier-Stokes equations, we have not yet addressed the pertinent question of the stability properties of the scheme. These properties become important, for example, in the presence of rounding errors. Here we carry out a stability analysis within the framework introduced by López-Marcos and Sanz-Serna [50], and we begin with a brief discussion of this framework.

5.4.1 Definition of stability

Let u be a solution of the equation $\Phi(u) = 0$, where Φ is a mapping from a Banach space $\mathcal{X}$ into a Banach space $\mathcal{Y}$. (The Navier-Stokes equations as described earlier may be set in this form). Let $\mathcal{H}$ be a set of positive numbers (or vectors in R^2 with positive entries) with zero infimum. For each $h \in \mathcal{H}$, let $\mathcal{U}_h$ be a numerical approximation to u, obtained by solving

$$\Phi_h(\mathcal{U}_h) = 0, \qquad (5.29)$$

where Φ_h is a fixed mapping with domain $\mathcal{D}_h \in \mathcal{X}_h$ and taking values in $\mathcal{Y}_h$, with both $\mathcal{X}_h$ and $\mathcal{Y}_h$ being finite-dimensional linear spaces. We choose a norm $\|\cdot\|_{\mathcal{X}_h}$ in $\mathcal{X}_h$, a norm $\|\cdot\|_{\mathcal{Y}_h}$ in $\mathcal{Y}_h$ and a 'target element' u_h in the interior of $\mathcal{D}_h$, which should be some discrete representation of the analytical solution u in $\mathcal{X}_h$. Then we have the following definition of stability for (5.29) (c.f. [50]).

Definition 5.1 *For $h \in \mathcal{H}$, let $R_h \in (0,\infty]$. The discretisation (5.29) is said to be stable, restricted to the thresholds R_h, if there exist positive constants h_0 and S such that for $h \in \mathcal{H}$, $|h| \le h_0$, the open ball $B(u_h, R_h)$ is contained in $\mathcal{D}_h$ and for any $\mathcal{V}_h, \mathcal{W}_h$ in that ball,*

$$\|\mathcal{V}_h - \mathcal{W}_h\|_{\mathcal{X}_h} \le S\,\|\Phi_h(\mathcal{V}_h) - \Phi_h(\mathcal{W}_h)\|_{\mathcal{Y}_h} . \qquad (5.30)$$

The motivation for this definition was in part to introduce a concept of stability that could be used in a general "stability + consistency ⇒ convergence" type of Theorem. Knowing this, it would have been natural to have sought to prove the convergence of the method by this well-trodden route. However, although such a strategy would have indeed produced a convergence result, the assumptions of the smoothness of the exact solution u would have had to be more severe. The difference between this and the problem-specific proof given above is the introduction of the 'target element' in the above general definition of stability.

In order to make use of this definition we have to cast the spectral Lagrange-Galerkin method into an appropriate form. This is the aim of the next subsection.

5.4.2 Reformulation of the problem

In the previous section the discrete equations were parameterised by small positive numbers h. In our case, however, there are two parameters, Δt and N, associated with the discrete equations. We thus denote by h the pair $(\Delta t, 1/N) \in (0,\infty)^2$. In fact we invoke the limitations on Δt and N involved in Theorem 5.2, so that $h \in (0,\Delta t_0] \times (0, 1/N_0]$.

The spaces $\mathcal{X}_h$ and $\mathcal{Y}_h$ will be distinguished only by their norms. They will both consist of vectors

$$\mathcal{U} = (U^0,\ldots,U^M)^T \in \left(\dot{S}_N\right)^{M+1}.$$

For $\mathcal{U} \in \mathcal{X}_h$ we define

$$\|\mathcal{U}\|_{\mathcal{X}_h} := \|\mathcal{U}\|_{l^\infty(H)} + \nu^{1/2}\|\mathcal{U}\|_{l^2(V)},$$

and for $\mathcal{U} \in \mathcal{Y}_h$ we define

$$\|\mathcal{U}\|_{\mathcal{Y}_h} := \left\|U^0\right\| + \left(\Delta t\sum_{k=1}^{M}\left\|U^k\right\|^2\right)^{1/2}.$$

We suppose that u satisfies (5.2), (5.13) and (5.14), for some $s > (n+2)/2$. Then we set our target element $u_h \in \mathcal{X}_h$ to be $\Pi_N u$. We are now in a position to define the operator $\Phi_h : \mathcal{X}_h \to \mathcal{Y}_h$. Let $\mathcal{V} = (V^0,\ldots,V^M)^T \in \mathcal{X}_h$. For each $k = 1,\ldots,M$, we have $V^{k-1}, V^k \in \dot{S}_N$, and we construct $F^k \in \dot{S}_N$ as follows. For $(x,t) \in \Omega \times [0,T]$ we define $X_{V^{k-1}}(x,t;\cdot)$ to be the solution of the initial value problem

$$\begin{aligned}
\frac{d}{ds}X_{V^{k-1}}(x,t;s) &= V^{k-1}(X_{V^{k-1}}(x,t;s)), \quad s \in [0,T]\backslash\{t\}, \\
X_{V^{k-1}}(x,t;t) &= x.
\end{aligned}$$

We then define, for $s,t \in [0,T]$, $E_{V^{k-1}}(t;s) : H \to H$ analogously to $E_{U^{k-1}}(t;s)$ in the discussion of convergence, and similarly denote $E_{V^{k-1}}(t^k;t^{k-1})$ by $E_{V^{k-1}}$. Furthermore we define $F^k \in \dot{S}_N$ to be the solution of

$$(F^k, v) = \frac{1}{\Delta t}((V^k - E_{V^{k-1}}V^{k-1}), v) + \nu(\nabla V^k, \nabla v) - (f^k, v) \quad \forall v \in V_N,$$

and F^0 to be the solution of

$$(F^0, v) = (V^0 - \Pi_N u_0, v) + \frac{1}{2}\nu\Delta t\,(\nabla(V^0 - \Pi_N u_0), \nabla v) \quad \forall v \in V_N.$$

In this way we construct $\mathcal{F} = (F^0,\ldots,F^M)^T$ from $\mathcal{V}$, and we define $\Phi_h : \mathcal{X}_h \to \mathcal{Y}_h$ by

$$\Phi_h(\mathcal{V}) = \mathcal{F}.$$

The spectral Lagrange-Galerkin method for the Navier-Stokes equations may then be succinctly described as follows: find $\mathcal{U} \in \mathcal{X}_h$ such that

$$\Phi_h(\mathcal{U}) = 0. \qquad (5.31)$$

5.4.3 Stability analysis

Theorem 5.4 *There is a pair of positive numbers $(\Delta t_0, 1/N_0)$ such that for $N \ge N_0$ and $\Delta t \le \Delta t_0$ the spectral Lagrange-Galerkin method (5.31) is unconditionally non-linearly stable with stability threshold $R_h = (ND_n(N))^{-1}$.*

Proof Choose $\mathcal{V}, \mathcal{W} \in \mathcal{X}_h \cap B(u_h, R_h)$, and write $\mathcal{F} = \Phi_h(\mathcal{V})$ and $\mathcal{G} = \Phi_h(\mathcal{W})$. For $k = 0,\ldots,M$, we write $\zeta^k = V^k - W^k$. Thus, for $k = 1,\ldots,M$ and for all $v \in V_N$, we have

$$\frac{1}{\Delta t}\left(\zeta^k - (E_{V^{k-1}}V^{k-1} - E_{W^{k-1}}W^{k-1}), v\right) + \nu(\nabla\zeta^k, \nabla v) = (F^k - G^k, v),$$

so that, taking $v = \zeta^k$, we find

$$\begin{aligned}
\frac{1}{\Delta t}\left\|\zeta^k\right\|^2 + \nu\left\|\nabla\zeta^k\right\|^2 &\le \left\|\zeta^k\right\|\left\{\frac{1}{\Delta t}\left\|E_{V^{k-1}}V^{k-1} - E_{W^{k-1}}W^{k-1}\right\|\right. \\
&+ \left.\left\|F^k - G^k\right\|\right\}. \qquad (5.32)
\end{aligned}$$

The key part of the proof is the bound on $\left\|E_{V^{k-1}}V^{k-1} - E_{W^{k-1}}W^{k-1}\right\|$. We write

$$\left\|E_{V^{k-1}}V^{k-1} - E_{W^{k-1}}W^{k-1}\right\| \le \left\|\zeta^{k-1}\right\| + \left\|(E_{V^{k-1}} - E_{W^{k-1}})W^{k-1}\right\|,$$

where we have bounded $\left\|E_{V^{k-1}}\zeta^{k-1}\right\|$ by $\left\|\zeta^{k-1}\right\|$. The second term may be bounded by $\|X_{V^{k-1}} - X_{W^{k-1}}\| \left|W^{k-1}\right|_{C^{0,1}}$, where $|\cdot|_{C^{0,1}}$ is the Lipschitz seminorm, and it is relatively straightforward to show, by use of Gronwall's lemma, that

$$\|X_{V^{k-1}} - X_{W^{k-1}}\| \leq \Delta t \left\|\zeta^{k-1}\right\| e^{\Delta t \left|W^{k-1}\right|_{C^{0,1}}}.$$

Under the assumptions that we have on Δt and $\mathcal{W}$, and making use of Sobolev's embedding theorem, we have

$$\begin{aligned} \Delta t \left|W^{k-1}\right|_{C^{0,1}} &\leq \Delta t D_n(N) \left\|W^{k-1} - \Pi_N u^{k-1}\right\|_{V_2} + \Delta t \left|\Pi_N u^{k-1}\right|_{C^{0,1}} \\ &\leq \Delta t^{1/2} N D_n(N) R_h + \Delta t C_{10} \|u\|_{l^\infty(V_s)}, \end{aligned} \tag{5.33}$$

and so according to the assumptions of the theorem $e^{\Delta t \left|W^{k-1}\right|_{C^{0,1}}}$ may be bounded by a constant C_{11}. We deduce from (5.32), making use of the relation $\left\|\zeta^k\right\| \leq \left\|\nabla\zeta^k\right\|$, that

$$\begin{aligned} \frac{1}{\Delta t}\left\|\zeta^k\right\|^2 &+ \nu\left\|\nabla\zeta^k\right\|^2 \leq \frac{1}{2\Delta t}\left(\left\|\zeta^k\right\|^2 + \left\|\zeta^{k-1}\right\|^2\right) + \frac{\nu}{2}\left\|\nabla\zeta^k\right\|^2 \\ &+ \frac{1}{\nu}\left(C_{11}^2 \left|W^{k-1}\right|_{C^{0,1}}^2 \left\|\zeta^{k-1}\right\|^2 + \left\|F^k - G^k\right\|^2\right). \end{aligned} \tag{5.34}$$

Multiplying through by $2\Delta t$ we obtain

$$\begin{aligned} \left\|\zeta^k\right\|^2 + \nu\Delta t\left\|\nabla\zeta^k\right\|^2 &\leq \left\|\zeta^{k-1}\right\|^2 \left(1 + \frac{C_{11}^2\Delta t}{\nu}\left|W^{k-1}\right|_{C^{0,1}}^2\right) \\ &+ \frac{2\Delta t}{\nu}\left\|F^k - G^k\right\|^2. \end{aligned} \tag{5.35}$$

Bounding $1 + \frac{C_{11}^2\Delta t}{\nu}\left|W^{k-1}\right|_{C^{0,1}}^2$ by $\exp\{\frac{C_{11}^2\Delta t}{\nu}\left|W^{k-1}\right|_{C^{0,1}}^2\}$ and applying Lemma 5.6 gives that

$$\|\mathcal{V} - \mathcal{W}\|_{\mathcal{X}_h} \leq 2\left(\frac{1+\nu}{\nu}\right)^{1/2} e^{\frac{C_{11}^2}{2\nu}|\mathcal{W}|_{l^2(C^{0,1})}^2} \|\Phi_h(\mathcal{V}) - \Phi_h(\mathcal{W})\|_{\mathcal{Y}_h}. \tag{5.36}$$

We bound $\left|W^{k-1}\right|_{l^2(C^{0,1})}$ by employing an argument identical to (5.33), and the stability result follows.

5.5 Numerical experiments

In order to be able to complete the analysis of this chapter, we have had to restrict ourselves to considering the exactly-integrated version of the method with a backward-Euler time step. In practice quadrature must be introduced, the additional local interpolation may be used to improve efficiency, and higher-order timestepping procedures may be incorporated. All of these elements are included in the calculations presented below, and we shall now describe the algorithm in detail, giving the formulae which serve to define the method, and discussing some of the details of the computational procedure.

5.5.1 The algorithm

Let $\mathcal{R}$ be the L^2-projection from $\dot{S}_N$ to the subset V_N consisting of functions with zero divergence and zero mean. $\mathcal{R}$ may be identified with the operator $(I - \nabla\Delta^{-1}\nabla\cdot)$, and so may be applied straightforwardly. The canonical basis for V_N is $\{\mathcal{R}\Phi_p\}_{0\neq|p|_\infty\leq N}$. Functions are represented by the coefficients of their expansions in these basis functions. The approximation is obtained by first discretising in time the Lagrangian form of (5.1) and then projecting onto divergence-free finite-dimensional subspaces. The full set of formulae for the qth order method is:

$$\begin{aligned} (I - \nu\Delta t\beta_0\Delta)U^{m,0} &= \mathcal{R}\Pi_N P_L E(t^m; t^{m-1})\mathcal{P}_{K,L}\left[\sum_{i=1}^q \alpha_i U^{m-i,i-1}\right. \\ &\quad \left. - \Delta t\beta_1(\nabla P^{m-1} - F^{m-1} - \nu\Delta U^{m-1,0})\right] \\ &\quad + \Delta t\beta_0\mathcal{R}\Pi_N P_L F^m, \end{aligned} \tag{5.37a}$$

$$U^{m-i,i} = \Pi_N P_L E(t^m; t^{m-1})\mathcal{P}_{K,L}U^{m-i,i-1}, \quad i = 1, \ldots, q-1, \tag{5.37b}$$

$$\begin{aligned} \Delta t\beta_0\nabla P^m &= (I - \mathcal{R})\Pi_N P_L E(t^m; t^{m-1})\mathcal{P}_{K,L}\left[\sum_{i=1}^q \alpha_i U^{m-i,i-1}\right. \\ &\quad \left. - \Delta t\beta_1(\nabla P^{m-1} - F^{m-1} - \nu\Delta U^{m-1,0})\right] \\ &\quad + \Delta t\beta_0(I - \mathcal{R})\Pi_N P_L F^m, \end{aligned} \tag{5.37c}$$

No. of time steps	L^2-error	Order of convergence
10	2.65×10^{-2}	
15	7.04×10^{-3}	3.26
25	1.15×10^{-3}	3.55
50	8.48×10^{-5}	3.76
100	5.69×10^{-6}	3.90

Table 5.1: Results for the first Navier-Stokes experiment

where

$$F^m = \Pi_N P_L f(t^m), \tag{5.37d}$$

and where $P_L E(t^m; t^{m-1})$ may be implemented by solving

$$x_\nu - X_\nu = \Delta t\sum_{i=1}^q \gamma_i\mathcal{P}_{K,L}U^{m-i,i-1}(X_\nu), \quad \nu \in \Gamma_{2L-1}. \tag{5.37e}$$

The coefficients α_i, β_i and γ_i are as in (4.27) and (4.67). The calculation is commenced by means of the initialisation procedure described in Chapter 4. Once this is complete the solution is advanced in time by means of the qth-order algorithm with constant timestep. Each particular time step is performed as follows.

The first task is the solution of (5.37e). The Fourier coefficients of the sum $\sum_{i=1}^q \gamma_i U^{m-i,i-1}$ are formed and used to form the right hand side of (5.37e), which is then solved by means of the modified Newton iteration described in the context of Burgers equation in Chapter 4. The calculation is performed to a predetermined tolerance (TOL1). In practice, no more than three iterations have ever been needed to achieve machine accuracy in this calculation.

Once this is known, the coefficients of the term

$$\Pi_N P_L E(t^m; t^{m-1})\mathcal{P}_{K,L}\left[\sum_{i=1}^q \alpha_i U^{m-i,i-1} - \Delta t\beta_1(\nabla P^{m-1} - F^{m-1} - \nu\Delta U^{m-1,0})\right]$$

may be calculated and stored. It is then a simple matter to calculate $U^{m,0}$ by including the forcing term, applying $\mathcal{R}$, and inverting the left hand side of (5.37a), and to update the other terms $U^{m-i,i}$ and calculate the pressure according to (5.37b) and (5.37c) respectively.

The operator $\mathcal{P}_{K,L}$ is used repeatedly in this process. The time-saving devices described in Chapter 4 are used, thus introducing a second tolerance parameter (TOL2) which is again chosen at the beginning of the calculation.

5.5.2 The experiments

The experiments we have conducted thus far have been primarily for the purpose of verification of the code. More extensive experimentation will be carried out and presented elsewhere.

The first results show that the time-accuracy of the method is as required. It is easy to design a problem with a predetermined exact solution by choosing the forcing function appropriately. In this experiment the exact solution is

$$u(x,y,t) = \begin{pmatrix} \cos y e^{-t+\sin x+\sin y} \\ -\cos x e^{-t+\sin x+\sin y} \end{pmatrix}, \tag{5.38a}$$

$$p(x,y,t) = e^{-t+\sin x+\sin y}, \tag{5.38b}$$

and this is generated by the forcing function

$$f(x,y,t) = \begin{pmatrix} \alpha(\cos x(1+\alpha\sin y) - \cos y - \nu\cos y(\cos^2 x - \sin x - \sin^2 y - 3\sin y)) \\ \alpha(\cos x + \cos y(1+\alpha\sin x) + \nu\cos x(\cos^2 y - \sin y - \sin^2 x - 3\sin x)) \end{pmatrix}, \tag{5.39}$$

where $\alpha = e^{-t+\sin x+\sin y}$.

The calculations were carried out using a fourth-order backward difference scheme with the following parameters:

$$\begin{aligned} N &= 31 \\ L &= 32 \\ K &= 5 \\ T &= 1 \\ \nu &= 10^{-4} \\ TOL1 &= 10^{-8} \\ TOL2 &= 10^{-8}, \end{aligned}$$

and the results for various values of Δt are given in Table 5.1.

The second experiment is somewhat more realistic. We consider the evolution of a perturbed double shear layer, generated by an initial vorticity distribution

$$\omega(x,y) = sech^2(10(y-\alpha)) - sech^2(10(y-\beta)) + 10^{-3}\cos(2x) + 10^{-4}\cos(4x), \tag{5.40}$$

with $\alpha = (1-1/8)\pi$ and $\beta = (1+1/8)\pi$. This problem is considered by Kreiss *et. al.* [43], where they use a spectral discretisation of the equations in vorticity-velocity form combined with a fourth-order Runge-Kutta method for the discretisation in time. Our calculation was carried out with the following parameters:

$$\begin{aligned} N &= 63 \\ L &= 64 \\ K &= 5 \\ T &= 150 \\ \Delta t &= 5 \times 10^{-2} \\ \nu &= 10^{-4} \\ TOL1 &= 10^{-8} \\ TOL2 &= 10^{-8}. \end{aligned}$$

The results, shown in the form of contour plots of the vorticity in Figs. 5.1–5.4, are visibly indistinguishable from those obtained in [43]. They show a wake flow unstable to the perturbations imposed upon it. The $\cos(4x)$ perturbation governs the initial formation of a von Kármán vortex street, but eventually the $\cos(2x)$ perturbation grows and these vortices pair off into a final stable state.

5.6 Conclusions

We have in this Thesis presented and analysed a new numerical method for the solution of convection-dominated diffusion problems. The scheme is based on a Lagrangian-Eulerian approach that rewrites the equations in Lagrangian coordinates that take care of the convective terms, and avoids the distortion that might cause difficulties in dealing with the diffusion and other terms by using a different set of Lagrangian coordinates at each time step.

This approach has been widely and successfully implemented in combination with a variety of spatial discretisation methods. The use of spectral methods is relatively new. In our case it was motivated by the stability problems faced by the finite-element version of the method when quadrature was introduced. Our analysis has shown that the spectral version remains unconditionally stable, even with the introduction of quadrature.

The combination of spectral methods with Lagrangian approaches suffers from the drawback of unreasonable expense due to the distortion of the coordinates. We have introduced a new procedure by which this can be circumvented with negligible loss in accuracy. The algorithm is such that it could be readily implemented on a shared memory parallel machine, and would be able to make full use of a modestly parallel computer, with 10–20 processors. On such a machine the method would be highly competitive. This new procedure is based on the use of a local Chebyshev interpolation. Its application is more general than just the spectral Lagrange-Galerkin method. It will be of use whenever it is required to evaluate a trigonometric (or a Chebyshev) polynomial at an unstructured set of points, as might be the case if one were to seek to track a set of particles released into a velocity field which was represented in terms of such a polynomial.

In order to capitalise upon the high accuracy of spectral methods for space discretisation, a highly accurate time discretisation is required. We have described and analysed multistep methods (specifically backward difference methods) for the time discretisation of the equations in Lagrangian form. These give rise to fully discrete schemes with well balanced error terms without the need to use excessively small timesteps. Their implementation in non-linear problems involves tracking the 'particle trajectories' with corresponding accuracy. It was found that the most convenient way to do this was by means of an Adams-Bashforth formula used to extrapolate along the trajectory itself. This then gives rise to set of implicit problems which must be solved by some iterative scheme. In the search for greater efficiency we have introduced a modified Newton iteration, which leads to savings in time of at least 50 percent because of the nature of the local interpolation used in evaluating the functions involved. We are not aware of any other applications for this iterative procedure.

The culmination of the development of the method in this Thesis is its application to the Navier-Stokes equations. The analysis, applied to a slightly idealised form of the method, has shown it to be convergent (with optimal order of convergence) and non-linearly stable. Moreover, with suitable decay of the exact solution, these results hold for infinite time. Alongside the theory, the method has been made into a computer code, which has been successfully implemented on a simulation of two-dimensional wake instability.

5.6.1 Future developments

The most obvious extension of this work is the consideration of non-periodic problems. The major difficulty here will be the incorporation of boundary conditions, a long-standing thorn in the side for some Lagrangian methods. It is hoped that the detailed exploration of the properties of the spectral Lagrange-Galerkin method in the absence of boundaries contained herein will have shed light on efforts being made to apply the method to such problems.

As already mentioned, the algorithm is highly amenable to implementation on shared-memory parallel machines, and in such a setting would provide a highly competitive method for the solution of the Navier-Stokes equations. There is a long history of the use of Fourier methods in the modelling of two-dimensional turbulence. The spectral Lagrange-Galerkin method is a fresh approach and could make a significant contribution to the area.

A final area of interest is in the analysis of the method as applied to non-linear problems with non-trivial dynamics. There is an increasing awareness of the need to ensure that a numerical scheme mirrors the dynamics of the original system accurately, and the approach of timestepping along particle trajectories as opposed to timestepping along straight lines may significantly alter the dynamics of a numerical method. We feel that there is room for further investigation in this area; the stability results of this chapter may be a first step in this direction.

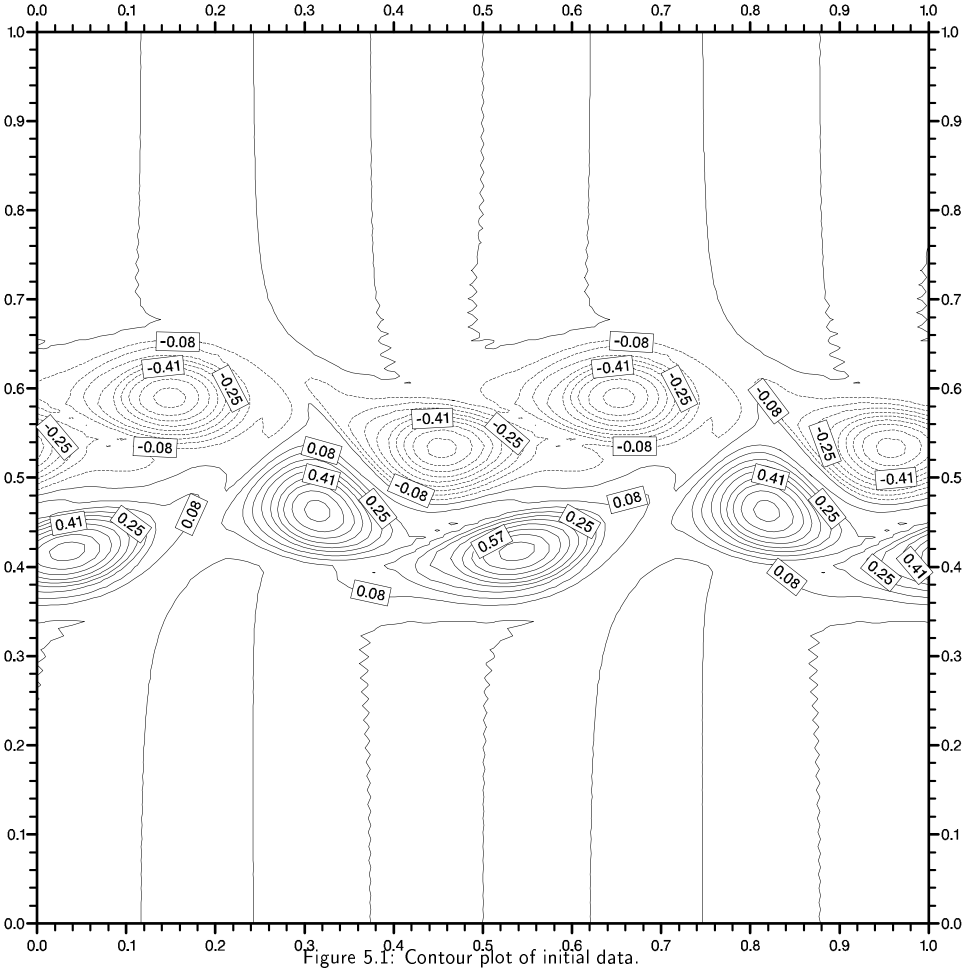

Figure 5.1: Contour plot of initial data.

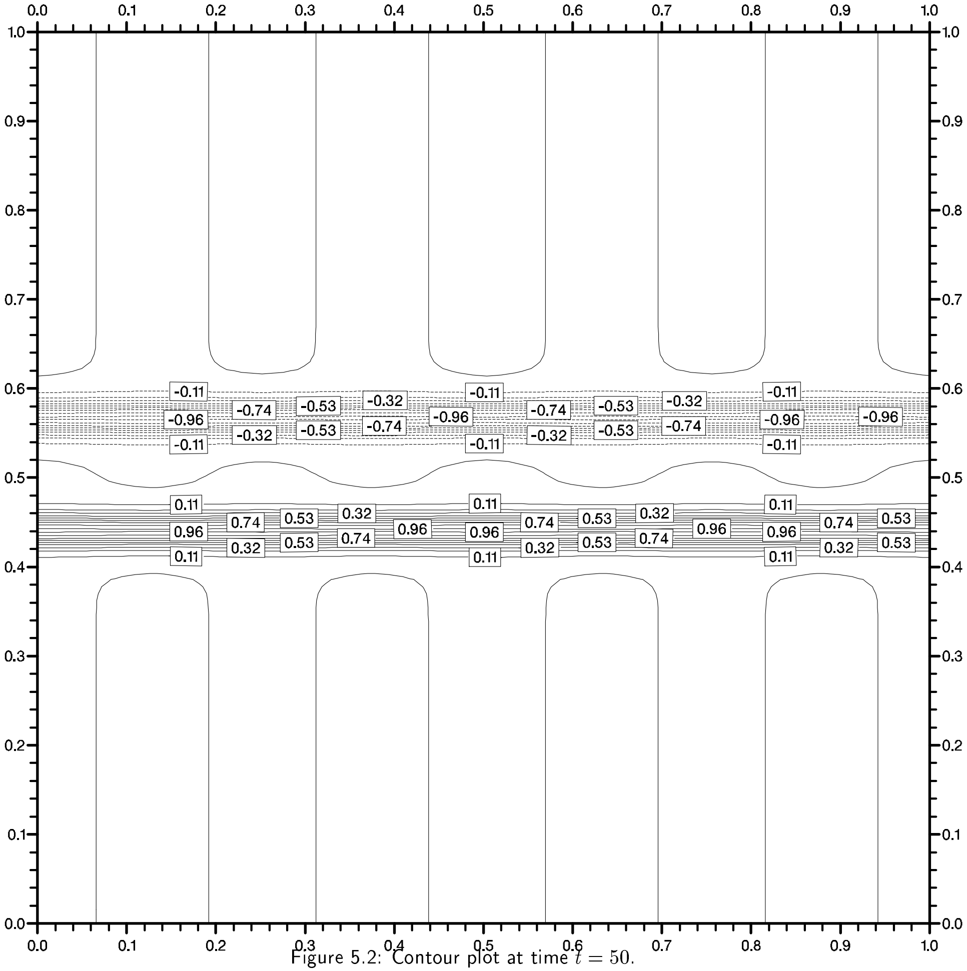

Figure 5.2: Contour plot at time $t = 50$.

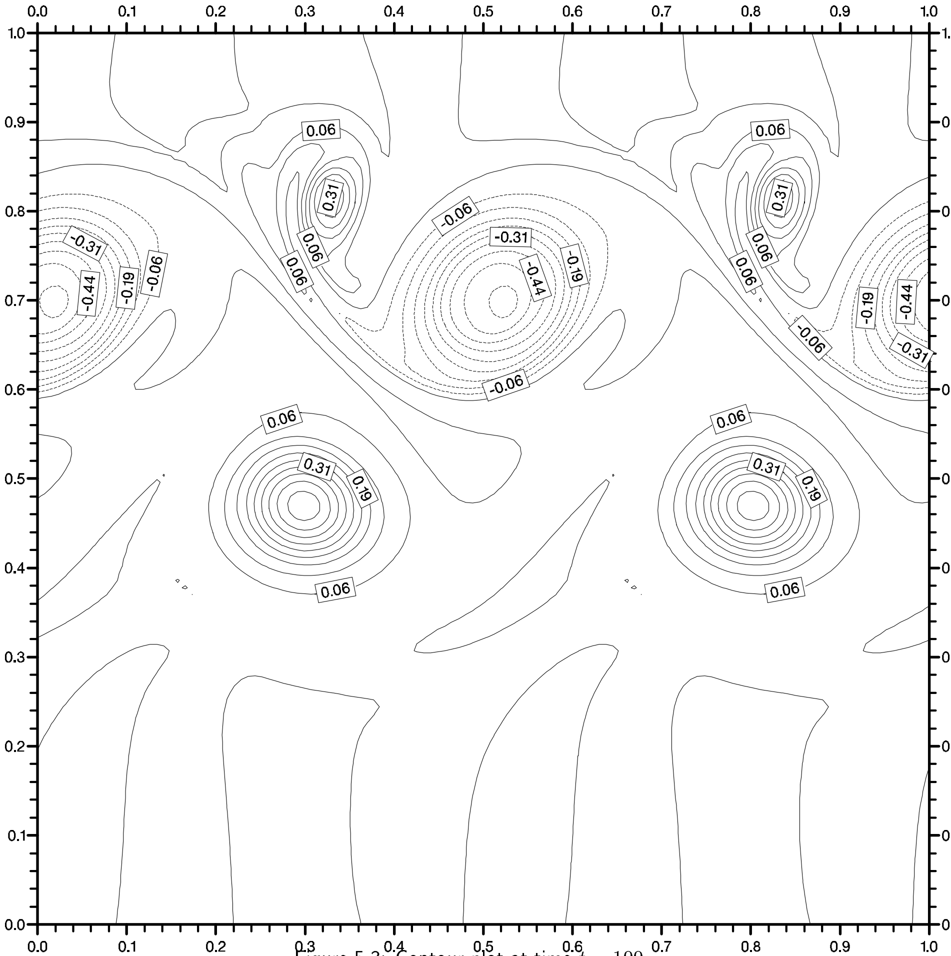

Figure 5.3: Contour plot at time $t = 100$.

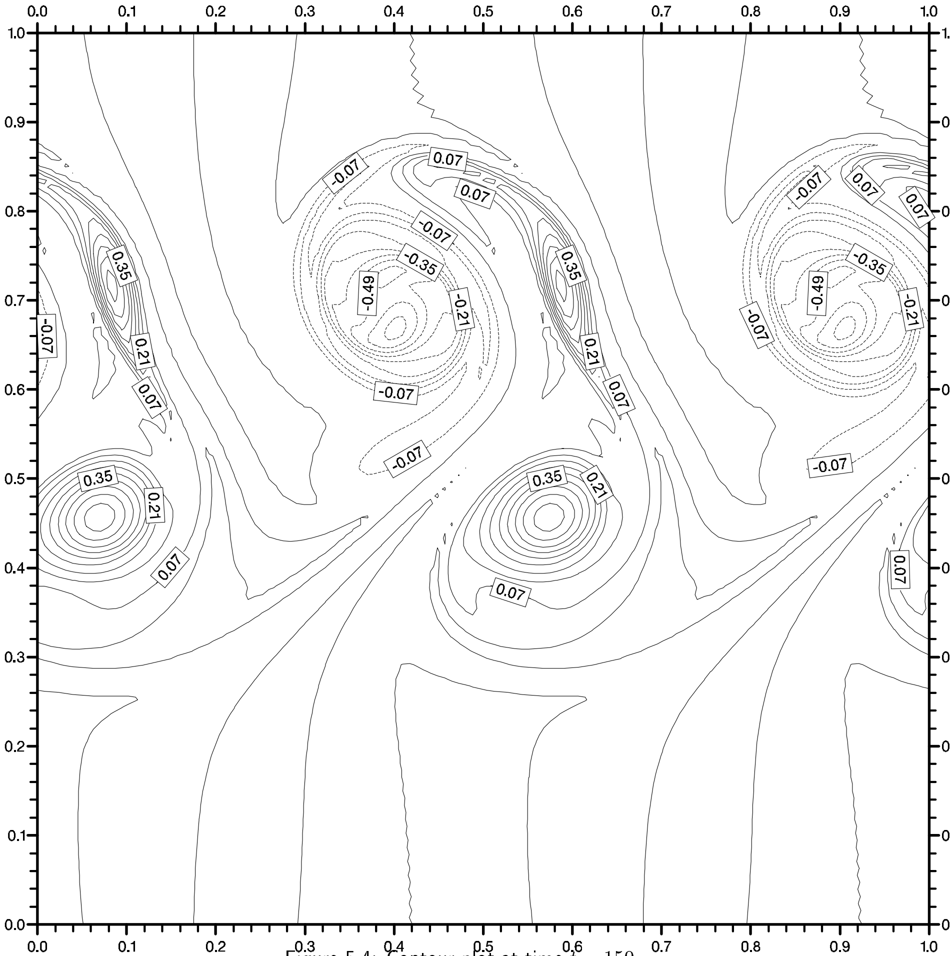

Figure 5.4: Contour plot at time $t = 150$.

Bibliography

[1] Adams, R.S. (1975). Sobolev Spaces. Academic Press, New York.

[2] Anderson, C. and Greengard, C. (1985). On vortex methods. *SIAM J. Numer. Anal.*, 22:413–440.

[3] Ansorge, R. (1963). Die Adams-Verfahren als Characteristikenverfahren Ordnung zur Lösung von hyperbolischen Systemen halblinearer Differentialgleichungen. *Numer. Math.*, 5: 443–460.

[4] Baines, M.J. (1991). Analysis of the moving finite element procedure. *SIAM J. Numer. Anal.*, To appear.

[5] Barrett, J. W. and Morton, K. W. (1984). Approximate symmetrisation and Petrov-Galerkin methods for diffusion-convection problems. *Comp. Meth. Appl. Mech. Engrg.*, 45:97–122.

[6] Beale, J. T. and Majda, A. (1982). Vortex methods. I: Convergence in three dimensions. *Math. Comp.*, 39:1–27.

[7] Beale, J. T. and Majda, A. (1981). Rates of convergence for viscous splitting of the Navier-Stokes equations. *Math. Comp.*, 37:243–259.

[8] Benqué, J. P., Labadie, G., and Ronat, J. (1982). A new finite element method for Navier-Stokes equations coupled with a temperature equation. In *Proc. 4th Int. Symp. on Finite Element Methods in Flow Problems*, Kawai, T., Ed., pp 295–301. North Holland.

[9] Bercovier, M. and Pironneau, O. (1982). Characteristics and the finite element method. In *Proc. 4th Int. Symp. on Finite Element Methods in Flow Problems*, Kawai, T., Ed., pp 67–73. North Holland.

[10] Bermejo, R. (1990). On the equivalence of semi-Lagrangian schemes and particle-in-cell finite element methods. *Mon. Wea. Rev.*, 118:979–987.

[11] Cai, W., Gottlieb, D. and Shu, C-W. (1989). Essentially nonoscillatory spectral Fourier methods for shock wave calculations. *Math. Comp.*, 52:389–410.

[12] Canuto, C., Hussaini, M. Y., Quarteroni, A. and Zang, T. A. (1988). Spectral Methods in Fluid Dynamics. Springer-Verlag.

[13] Childs, P.N. and Morton, K.W. (1990). Characteristic Galerkin methods for scalar conservation laws in one dimension. *SIAM J. Numer. Anal.*, 27:553–594.

[14] Chorin, A.J. (1973). A numerical study of slightly viscous flow. *J. Fluid Mech.* 57:785–796.

[15] Cottet, G-H. (1989). A particle-grid superposition method for the Navier-Stokes equations. *J. Comp. Phys.*, 89:301–318.

[16] Cottet, G-H. (1991). Large-time behaviour of deterministic particle approximations to the Navier-Stokes equations. *Math. Comp.*, Vol. 56, No. 193, pp. 45–59.

[17] Courant, R., Isaacson and E, Rees, M. (1952). On the solution of non-linear hyperbolic differential equations by finite differences. *Comm. Pure Appl. Math.*, 5: 243–264.

[18] Crouzeix, M. (1980). Une méthode multipas implicite-explicite pour l'approximation des équations d'évolution paraboliques. *Numer. Math.*, 35: 257–276.

[19] Donea, J. (1984). A Taylor-Galerkin method for convective-transport problems. *Int. J. Numer. Meth. Engrg.*, 20:101–119.

[20] Douglas, Jr., J. (1983). Finite difference methods for two phase incompressible flow in porous media. *SIAM J. Numer. Anal.*, 20:681–696.

[21] Douglas, Jr., J. and Russell, T. F. (1982). Numerical methods for convection-dominated diffusion problems based on combining the method of characteristics with finite element or finite difference procedures. *SIAM J. Numer. Anal.*, 19:871–885.

[22] Ewing, E. R. and Russell, T. F. (1981). Multistep Galerkin methods along characteristics for convection-diffusion problems. In *Advances in Computer Methods for Partial Differential Equations*, Vichnevetsky, R. and Stepleman, R. S., Eds., pp 28–36. IMACS, Rutgers University.

[23] Farmer, C. L. (1985). A moving point method for arbitrary Peclet number multi-dimensional convection-diffusion problems. *IMA J. Numer. Anal.*, 5:465–480.

[24] Federer, H. (1969). Geometric Measure Theory. Berlin Heidelberg New York: Springer.

[25] Fishelov, D. (1990). A new vortex scheme for viscous flows. *J. Comp. Phys.*, 86:211–224.

[26] Foias, C., Guillopé, C. and Temam, R. (1985). Lagrangian representation of a flow. *J. Diff. Eq.* 57:440–449.

[27] Fritts, M.J., Crowley, W.P. and Trease, H.E. eds., (1985). Free-Lagrange methods for compressible hydrodynamics in two space dimensions. *Lecture Notes in Physics 238, The Free-Lagrange Method*, Springer-Verlag.

[28] Furzeland, R.M., Verwer, J.G. and Zegeling, P.A. (1990). A numerical study of three moving-grid methods for one-dimensional partial differential equations which are based on the method of lines. *J. Comp. Phys.*, 89:349–388.

[29] Gear, C.W. (1971). Numerical Initial-Value Problems in Ordinary Differential Equations. Prentice-Hall.

[30] Gottlieb, D. and Tadmor, E. (1991). The CFL condition for spectral approximations to hyperbolic initial-boundary value problems. *Math. Comp.*, 56:565–588.

[31] Gottlieb, D. and Turkel, E. (1980). On time discretisation for spectral methods. *Stud. Appl. Math.*, 63:67–86.

[32] Hald, O. and Del Prete, V.M., (1978). Convergence of vortex methods for Euler's equations. *Math. Comp.* 32:791–809.

[33] Hald, O. (1979). Convergence of vortex methods for Euler's equations. II *SIAM J. Numer. Anal.*, 16:726–755.

[34] Handscomb, D.C. (1989). Private communication.

[35] Hartman, P., (1964). Ordinary Differential Equations. New York London Sydney: Wiley.

[36] Hasbani, Y., Livne, E., and Bercovier, M. (1983). Finite elements and characteristics applied to advection-diffusion equations. *Comput. & Fluids*, 11:71–83.

[37] Ho, L-W., Maday, Y., Patera, A.T. and Rønquist, E.M. (1989). A high-order Lagrangian-decoupling method for the incompressible Navier-Stokes equations. *ICASE* Report No. 89-57.

[38] Holly Jr., F.M. and Preissmann, A. (1977). Accurate evaluation of transport in two dimensions. *J. Hydraul. Div. ASCE*, 98:1259–1277.

[39] Hughes, T. J. R. and Brooks, A. (1979). A multi-dimensional upwind scheme with no crosswind diffusion, In *Finite element methods for convection-dominated flows*, Hughes, T. J. R., Ed., AMD Vol. 34, ASME, New York., (1979), pp 19–35.

[40] Jack, R. O. (1987). Convergence properties of the Lagrange-Galerkin method with and without exact integration. OUCL Report No. 87/10, Oxford.

[41] Jack, R. O. (1988). Stability of the Lagrange-Galerkin method: the performance of quadrature in theory and practice. OUCL Report No. 88/15, Oxford.

[42] Johnson, C., Szepessy, A. and Hansbo, P. (1990). On the convergence of shock-capturing streamline-diffusion finite element methods for hyperbolic conservation laws. *Math. Comp.*, Vol. 54, No. 189, pp. 107–129.

[43] Kreiss, H.-O., Henshaw, W.D. and Reyna, L.G. (1988). On the smallest scale for the incompressible Navier-Stokes equations in two dimensions. *ICASE* Report No. 88-8.

[44] Kreiss, H.-O. and Oliger, J. (1979). Stability of the Fourier method. *SIAM J. Numer. Anal.*, 16:421–433.

[45] Leonard, A. (1980). Vortex methods for flow simulations. *J. of Comp. Phys.*, 37:289–335.

[46] Le Roux, M-N. (1979). Semi-discretisation en temps pour les équations d'évolution paraboliques lorsque l'opérateur dépend du temps. *R.A.I.R.O. Numerical Analysis*, Vol. 13, No. 2, pp. 119–137.

[47] Le Roux, M-N. (1979). Semidiscretisation in time for parabolic problems. *Math. Comp.*, Vol. 33, No. 147, pp. 919–931.

[48] Le Roux, M-N. (1980). Méthodes multipas pour des équations paraboliques non linéaires. *Numer. Math.*, 35:143–162.

[49] Lesaint, P. (1977). Numerical solution of the equation of continuity. In *Topics in Numerical Analysis III*, J.J.H. Miller, Ed., Academic Press, London, New York, San Francsico, pp. 199-222.

[50] López-Marcos, J. and Sanz-Serna, J. (1987). Stability and convergence in numerical analysis III: Linear investigation of nonlinear stability. *IMA J. Numer. Anal.* 8: 71–84.

[51] Mas-Gallic, S. (1990). Deterministic particle methods: diffusion and boundary conditions. Internal report 90029, Analyse Numerique, Université Pierre et Marie Curie.

[52] Mercier, B. (1989). An Introduction to the Numerical Analysis of Spectral Methods. Lecture Notes in Physics 318. Springer-Verlag.

[53] Miller, K. and Miller, R.N. (1981). Moving finite elements part I. *SIAM J. Numer. Anal.*, 18:1019–1032.

[54] Miller, K. (1981). Moving finite elements part II. *SIAM J. Numer. Anal.*, 18:1033–1057.

[55] Mizohata, S. (1973). The Theory of Partial Differential Equations. Cambridge University Press.

[56] Morton, K. W. and Parrott, A.K. (1980). Generalised Galerkin methods for hyperbolic problems. *J. Comput. Phys.*, 36:247–270.

[57] Morton, K. W., Priestley, A., and Süli, E. (1988). Stability of the Lagrange-Galerkin method with non-exact integration. *RAIRO M^2AN*, vol. 22, no. 4, pp. 123–151.

[58] Neuman, S.P. (1981). A Eulerian-Lagrangian numerical scheme for the dispersion-convection equation using conjugate space-time grids. *J. Comp. Phys.*, 41:270–294.

[59] Pasciak, J.E. (1980). Spectral and pseudo-spectral methods for advection equations. *Math. Comp.*, Vol. 35, No. 152, pp. 1081–1092.

[60] Pironneau, O. (1982). On the transport-diffusion algorithm and its applications to the Navier-Stokes equations. *Numer. Math.*, 38:309–332.

[61] Priestley, A. (1986). Lagrange and characteristic Galerkin methods for evolutionary problems. D. Phil. Thesis, University of Oxford.

[62] Priestley, A. (1989). The spectral Lagrange-Galerkin method for the atmospheric transportation of pollutants. *Numerical Analysis Report* 2/89, University of Reading.

[63] Raviart, P.-A. (1985). An analysis of particle methods. Lecture Notes in Mathematics 1127, Springer.

[64] Rees, M.D. (1988). Moving point, particle and free-Lagrange methods. D.Phil Thesis, Oxford University.

[65] Rees, M.D. and Morton, K. W. (1991). Moving point, particle and free-Lagrange methods for convection-diffusion equations. *SIAM J. Sci. Statist. Comput.*, 12:547–572.

[66] Ritchie, H. (1986). Eliminiating the interpolation associated with the semi-Lagrangian method. *Mon. Wea. Rev.*, 114:135–146.

[67] Ritchie, H. (1987). Semi-Lagrangian advection on a Gaussian grid. *Mon. Wea. Rev.*, 115:608–619.

[68] Ritchie, H. (1988). Application of the semi-Lagrangian method to a spectral model of the shallow water equations. *Mon. Wea. Rev.*, 116:1587–1598.

[69] Robert, A. (1981). A stable numerical integration scheme for the primitive meteorological equations. *Atmos. Ocean*, 19:35–46.

[70] Russell, T. F. (1985). Timestepping along characteristics with incomplete iteration for a Galerkin approximation of miscible displacement in porous media. *SIAM J. Numer. Anal.*, 22:970–1013.

[71] Russell, T. F. (1989). Eulerian-Lagrangian localized adjoint methods for advection-dominated problems. In *Numerical Analysis 1989*, D.F. Griffiths and G.A. Watson Eds., Pitman Research Notes in Mathematics, Vol.228. Longman/Wiley. pp. 206–228.

[72] Smolarkiewicz, P.K. and Rasch, P.J. (1991). Monotone advection on a sphere: an Eulerian versus semi-Lagrangian approach. *J. Atmos. Sci.*, 48:793–810.

[73] Staniforth, A. and Côté, J. (1990). Semi-Lagrangian integration schemes for atmospheric models—a review. Submitted to Monthly Weather Review.

[74] Stuart, A. (1989). Private communication.

[75] Süli, E. (1985). Lagrange-Galerkin mixed finite element approximation of the Navier-Stokes equations. In *Numerical Methods for Fluid Dynamics*, Morton, K. W. and Baines, M. J., Eds. pp 439–448. Oxford University Press.

[76] Süli, E. (1988). Convergence and nonlinear stability of the Lagrange-Galerkin method for the Navier-Stokes equations. *Numer. Math.*, 53:459–483.

[77] Süli, E. (1988). Stability and convergence of the Lagrange-Galerkin method with non-exact integration. In *The Mathematics of Finite Elements and Applications VI*, Whiteman, J. R., editor, pages 435–442. Academic Press.

[78] Süli, E. and Ware, A. (1988). A spectral method of characteristics for first-order hyperbolic equations. OUCL Report No. 88/6, Oxford.

[79] Süli, E. and Ware, A. (1991). A spectral method of characteristics for first-order hyperbolic equations. *SIAM J. Numer. Anal.* Vol. 28, No. 2, pp. 423–445.

[80] Tadmor, E. (1986). The exponential accuracy of Fourier and Chebyshev differencing methods. *SIAM J. Numer. Anal.*, Vol. 23, pp. 1–10.

[81] Tadmor, E. (1987). Stability analysis of finite difference, pseudospectral and Fourier-Galerkin approximations for time-dependent problems. *SIAM Review*, Vol. 29, pp. 525–555.

[82] Tadmor, E. (1989). Convergence of spectral methods for nonlinear conservation laws. *SIAM J. Numer. Anal.*, Vol. 26, pp. 30–44.

[83] Tal-Ezer, H. (1986). Spectral methods in time for hyperbolic equations. *SIAM J. Numer. Anal.*, Vol. 23, pp. 11–26.

[84] Tal-Ezer, H. (1989). Spectral methods in time for parabolic equations. *SIAM J. Numer. Anal.*, Vol. 26, pp. 1–11.

[85] Temam, R. (1983). Navier-Stokes Equations and Nonlinear Functional Analysis. CBMS-NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia.

[86] Temam, R. (1984). The Navier-Stokes Equations, Theory and Numerical Analysis. 3rd revised edition, North-Holland.

[87] Temam, R. (1989). Infinite-Dimensional Dynamical Systems in Mechanics and Physics. Applied Mathematical Sciences 68, Springer-Verlag.

[88] Temperton, C. and Staniforth, A. (1987). An efficient two time-level semi-Lagrangian implicit integration scheme. *Q. J. Roy. Met. Soc.*, 113:1025–1039.

[89] Tomarelli, F. (1984). Regularity theorems and optimal error estimates for linear parabolic Cauchy problems. *Numer. Math.*, 45:23–50.

[90] Tourigny, Y. and Süli, E. (1991). The finite-element method with nodes moving along the characteristics for convection-diffusion equations. *Numer. Math.*, 59:399–412.

[91] Williamson, D. and Rasch, P. (1989). Two-dimensional semi-Lagrangian transport with shape-preserving interpolation. *Mon. Wea. Rev.*, 117:102–129.

[92] Yang, G., Belleudy, P. and Temperville, A. (1991). A higher-orderEulerian scheme for coupled advection-diffusion transport. *Int. J. Num. Meth. Fl.*, 12:43–58.

[93] Yang, J-C. and Hsu, E-L. (1991). On the use of the reach-back characteristics method for calcuation of dispersion. *Int. J. Num. Meth. Fl.*, 12:225–235.