I. A Pressure Poisson Method for the Incompressible Navier-Stokes Equations

II. Long Time Behavior of the Klein-Gordon Equations

by

David Shirokoff

Submitted to the Department of Mathematics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2011

Author ....
Department of Mathematics
July 21, 2011

Certified by ....
Rodolfo Ruben Rosales
Professor of Applied Mathematics
Thesis Supervisor

Accepted by ....
Michel Goemans
Chairman, Applied Mathematics

I. A Pressure Poisson Method for the Incompressible Navier-Stokes Equations

II. Long Time Behavior of the Klein-Gordon Equations

by

David Shirokoff

Submitted to the Department of Mathematics
on July 21, 2011, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

In this thesis, we address two problems involving partial differential equations. In the first problem, we reformulate the incompressible Navier-Stokes equations into an equivalent pressure Poisson system. The new system allows for the recovery of the pressure in terms of the fluid velocity, and consequently is ideal for efficient but also accurate numerical computations of the Navier-Stokes equations. The system may be discretized in theory to any order in space and time, while preserving the accuracy of solutions up to the domain boundary. We also devise a second order method to solve the recast system in curved geometries immersed within a regular grid. In the second problem, we examine the long time behavior of the Klein-Gordon equation with various nonlinearities. In the first case, we show that for a positive (repulsive) strong nonlinearity, the system thermalizes into a state which exhibits characteristics of linear waves. Through the introduction of a renormalized wave basis, we show that the waves exhibit a renormalized dispersion relation and a Planck-like energy spectrum. In the second case, we discuss the case of attractive nonlinearities. In comparison, here the waves develop oscillons as long lived, spatially localized oscillating fields. With an emphasis on their cosmological implications, we investigate oscillons in an expanding universe, and study their profiles and stability. The presence of a saturation nonlinearity results in flat-topped oscillons, which are relatively stable to long wavelength perturbations.

Thesis Supervisor: Rodolfo Ruben Rosales
Title: Professor of Applied Mathematics

Acknowledgments

To all of those who helped me on my way. My mother Sheila for showing me the joy of learning, my father Peter for showing me the world of engineering, my siblings, Heather and Isaac for sharing their laughter and care free disposition. My grandmothers Margaret and Natalka.

I am also grateful to my advisor Ruben Rosales for teaching me his tricks, and for the many entertaining conversations we have shared. In addition, throughout my work, Mustafa Amin and Matthew Ueckermann have provided a continuous stream of helpful comments and ideas (I especially owe Matt, who as my roommate cheerfully ignores the fact that our internal clocks are six hours out of phase). I will also look back fondly on my conversations with Alan Guth, Tristan Gilet, Lyuba Chumakova, Benjamin Seibold, David McKay, Roger Mong, Evangelos Sfakianaki, Mark Hertzberg and J. C. Nave.

Perhaps most of all, my memories of MIT will be filled with the joyous times and friendships I have made: Michael Matejek, David Gosset, Eric Heubel, David Tax, Jennifer Park, Rahul Bhattacharyya, Patricia Engel, Camille Kemble, Michael Zabek. I count you among my equations as *friends for life.*

Contents

List of Figures

List of Tables

Chapter 1

A Tale of Two PDEs

1.1 Setting the Stage

In this thesis we addresses two problems involving physically motivated partial differential equations (PDE). The first problem is in numerical analysis and concerns the problem of how one computes solutions to the Navier-Stokes equations. In contrast, the second problem is physical in nature and deals with describing the long time behavior of solutions to the Klein-Gordon equation.

1.2 The Navier-Stokes Equations

We encounter fluids such as air and water everyday in our lives. Due to their practical importance, engineers and physicists often seek to model fluid behavior through analytic methods and physical principles. These principles result in the Navier-Stokes equations.

A fluids behavior is often modeled using a continuum theory by its velocity vector field. Through the conservation of mass and momentum, or equivalently Newton's law, one obtains a system of differential equations that describes the evolution of the fluid field. In the case of Navier-Stokes, one retains a fluid dissipative force, a pressure force and possibly body forces in the momentum balance. This force balance defines one equation for each velocity component. Moreover, in many cases, such as water, the behavior of the fluid is essentially incompressible, and consequently the conservation of mass reduces to a divergence free condition on the fluid velocity. The result is the incompressible Navier-Stokes equations:

$$\rho\mathbf{u}_t + \rho(\mathbf{u}\cdot\nabla)\mathbf{u} = \mu\Delta\mathbf{u} - \nabla p + \mathbf{f}, \tag{1.1}$$

$$\nabla\cdot\mathbf{u} = 0. \tag{1.2}$$

Here μ and ρ are constants representing the fluid viscosity and density, while $\mathbf{u}$ and p are the dependent variables representing the velocity field and pressure. The term $\mathbf{f}$ represents body forces such as gravity.

The equations (1.1–1.2) have recently received widespread interest in both the mathe-

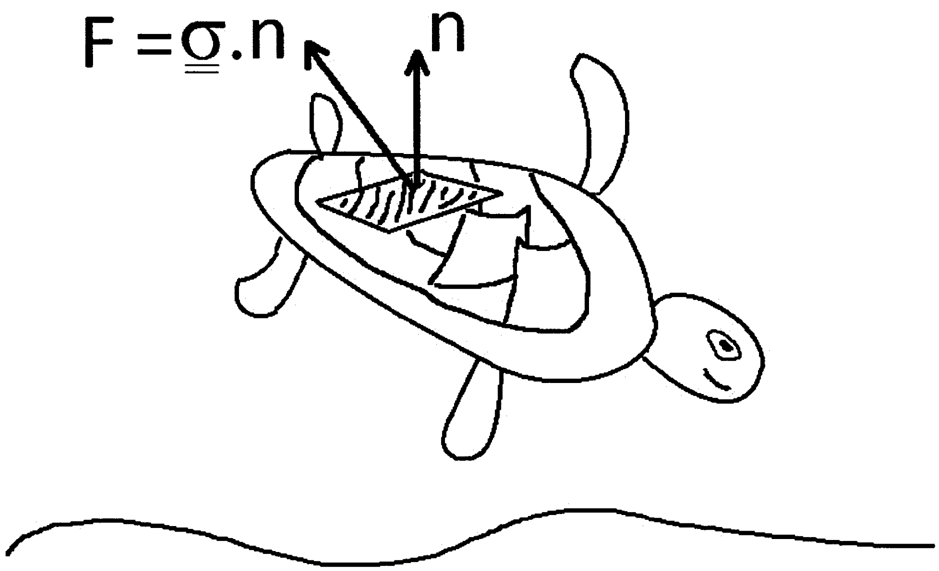

Figure 1-1: The computation of forces requires the evaluation of the stress tensor σ near domain boundaries. The stress tensor components are functions of the pressure and the velocity gradients.

matics community and general public due to their inclusion in one of the Clay Institute Millennium prize problems [26]. In particular, despite the practical importance of equations (1.1–1.2), the question of whether three dimensional solutions exist and stay smooth for all time remains an open problem[1]. In addition to the theoretical, analytic difficulties of Navier-Stokes, the equations also present many practical problems. The equations form a nonlinear, coupled set of PDEs which admit analytical solutions in a limited number of cases. As a result, one often resorts to solving equations (1.1–1.2) numerically on a computer. Due to their widespread use, however, the Navier-Stokes equations arise in many different scenarios with each problem presenting its own numerical difficulties. For instance, a few cases where the equations present numerical challenges include systems with small[2] or large μ, domains with moving or deformable boundaries, or domains with free boundaries. Consequently,

---

[1] We currently know that two dimensional solutions stay smooth for all time – see [26] for the millennium problem statement.

[2] Or more specifically small and large Reynolds numbers

there is no magic bullet for numerically solving Navier-Stokes.

A critical issue in the numerical solution of the incompressible Navier-Stokes equations is the question of how to implement the incompressibility constraint (1.2). As described theoretically in [98], and first exploited numerically by Chorin [20] and Temam [99], the pressure acts[3] to control the divergence constraint in the velocity field. As a result, efficient numerical schemes rely on the ability to quickly recover the pressure as a function of the velocity field $p = p[\mathbf{u}]$. In domains with symmetry, such as periodicity, or for fluid problems where the pressure is prescribed around the entire domain boundary, the accurate recovery of the pressure is possible with current schemes. In contrast, many physically interesting problems specify velocity boundary conditions. For instance, one often models the fluid interaction with a hard wall by specifying no flux (ie. no fluid can penetrate the wall) and no slip (ie. the fluid sticks to the wall at very close distances). For such domains with prescribed velocity boundary conditions, the accurate and efficient recovery of the pressure from the velocity field becomes more difficult. This also follows from the fact that the equations do not provide any boundary condition for the pressure. As a result, the recovery of the pressure has been an area of intense research, ever since the pioneering MAC scheme [50] of Harlow and Welch in 1965. Of course, one can avoid the problem by simultaneously discretizing the momentum and the divergence free equations, as in the difference scheme proposed by Krzywicki and Ladyzhenskaya [67]. For instance inside the domain, equations (1.1–1.2) can be written as:

$$\begin{pmatrix} \rho\partial_t - \mu\Delta & \nabla \\ \nabla\cdot & 0 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ p \end{pmatrix} = \begin{pmatrix} \mathbf{f} - \rho(\mathbf{u}\cdot\nabla)\mathbf{u} \\ 0 \end{pmatrix}. \tag{1.3}$$

Along with augmenting additional equations for the velocity boundary conditions, discretizing and inverting the left hand side of (1.3) avoids the need for pressure boundary conditions. This process does not however lead to efficient schemes since the velocity and pressure remain coupled together in a very large system of equations.

---

[3] Also as a Lagrange multiplier.

Generally the dilemma when numerically solving the Navier-Stokes equations has been that of a trade-off between efficiency and accuracy of the solution near the boundary. Many applications, however require both efficiency, and accuracy. For example, to calculate fluid solid interactions, both the pressure and gradients of the velocity are needed at the solid walls, as they appear in the components of the stress tensor. Furthermore, computation of the pressure and velocity must be achievable for *arbitrary* geometries, not just simple ones with exploitable symmetries. Unfortunately, these requirements are not something that current algorithms are generally well suited for. We now present a brief review of some popular current methods. However, we believe that algorithms based on a Pressure Poisson Equation (PPE) reformulation of the Navier-Stokes equations, reviewed towards the end of this section, offer a path out of the dilemma. The work presented in chapter 2 is such an approach. Namely the pressure Poisson formulation allows for the efficient recovery of the pressure from the velocity field.

1.2.1 Projection methods

Projection methods are very popular in practice because they are efficient. They achieve this efficiency by: (i) Interpreting the pressure as a projection of the flow velocity into the set of incompressible fields. That is, writing the equations in the form

$$\rho \mathbf{u}_t = \mathcal{P}\left(\mu \Delta \mathbf{u} - \rho(\mathbf{u}\cdot\nabla)\,\mathbf{u} + \mathbf{f}\right), \tag{1.4}$$

where $\mathcal{P}$ is the appropriate projection operator. (ii) Directly evolving the flow velocity. The question is then how to compute $\mathcal{P}$. Here the nature of a projection follows from the Hodge-Helmholtz decomposition theorem. Namely, given any vector field $\mathbf{w}$ (in L^2), then there are functions [98] b and vector fields $\mathbf{a}$ and $\mathbf{h}$ such that[4]:

$$\mathbf{w} = \nabla \times \mathbf{a} + \nabla b + \mathbf{h}. \tag{1.5}$$

---

[4] Here $\mathbf{h}$ is a harmonic vector field with $\nabla\cdot\mathbf{h} = \nabla\times\mathbf{h} = 0$. Such fields exist in domains with nontrivial homology groups.

To see how the pressure plays the role of a projection, one can apply the decomposition (1.5) on the field $\mathbf{w} = \mu\Delta\mathbf{u} - \rho(\mathbf{u}\cdot\nabla)\mathbf{u} + \mathbf{f}$. Noting that $\nabla\cdot\mathbf{u} = 0$, implies that the divergence of $\mathcal{P}\mathbf{w}$ should vanish. To preserve $\nabla\cdot\mathbf{u} = 0$, the projection $\mathcal{P}$ must therefore remove all gradient terms (at least those which are not Harmonic fields) in the field $\mathbf{w}$. This then suggests that for suitable b and $\mathbf{a}$, $p = b$ and $\mathbf{u}_t = \mu\Delta\mathbf{u} + \nabla\times\mathbf{a} + \mathbf{h}$. Hence computing the Hodge-Helmholtz decomposition of $\mathbf{w}$, which is a function of $\mathbf{u}$, *almost* provides a method to efficiently recover the pressure and evolve the Navier-Stokes equations. The problem is that, in bounded domains, the Hodge-Helmholtz decomposition (1.5) requires boundary conditions to make the individual terms (∇b) and $(\nabla\times\mathbf{a})$ unique[5]. Specifically, the terms (∇b) and $(\nabla\times\mathbf{a})$, and even whether they are mutually orthogonal in an L^2 norm, depend on their boundary values[6]. For a more detailed summary, Liu *et al.* [72] discuss the pressure boundary conditions in the context of a Hodge-Helmholtz decomposition, and include applications to fractional step methods.

In their original formulation by Chorin [20] and Temam [99], the projection method was formulated as a time splitting scheme in which: first an intermediate velocity is computed, ignoring incompressibility. Second, this velocity is projected onto the space of incompressible vector fields — by solving a Poisson equation for pressure. Unfortunately this process introduces numerical boundary layers into the solution, which can be improved (but not completely suppressed) for simple geometries — e.g. ones for which a staggered grid approach can be implemented [23]. Interestingly enough, as discussed in [47, 79, 97], the projection method is an approximate LU factorization of the differential operator matrix on the left hand side of equation (1.3). As a result of "splitting" the matrix into approximate operators L and U, one introduces an irreducible temporal error into the scheme. Namely, naive higher order discretizations of the time derivative in the projection method do not actually improve the order of the method. Consequently, new difficulties are introduced when devising high order (in time) projection and fractional steps methods.

---

[5] Note, b and $\mathbf{a}$ are never unique. For instance $\mathbf{a}$ can always vary by gauge field $\nabla\chi$.

[6] Note, there are two natural types of boundary conditions to take when constructing the Hodge-Helmholtz decomposition. These are somewhat analogous to Dirichlet and Neumann conditions.

The development of *second order projection methods* [9, 60, 63, 77, 103] provided greater control over the numerical boundary layers and accuracy in the pressure [15]. These are the most popular schemes used in practice. However, particularly for moderate or low Reynolds numbers, the effects of the numerical boundary layers can still be problematic [47]. Non-conforming boundaries add an extra layer of difficulty. The search for means to better control these numerical artifacts is an active area of research.

The numerical boundary layers in projection methods are reflected in the known convergence results (e.g. [83, 85, 93]). Convergence is stated only in terms of integral norms, with the main difficulties lying near the boundary. There point-wise convergence (and even less convergence of the flow velocity gradient) cannot be guaranteed — even if the solution is known to be smooth. Hence the accurate calculation of wall stresses with these methods is problematic. Guermond, Minev and Shen [47] provide further details on convergence results, as well as an extensive review of projection methods and the improved pressure-correction schemes.

1.2.2 Other methods

Two other methods for solving the Navier-Stokes equations are the *immersed boundary* [71, 75, 80, 101, 81, 97], and the *vortex-streamfunction* [10, 16, 76] methods. These also decouple the calculation of the velocity and of the pressure. The immersed boundary method does so by introducing Dirac forces to replace the domain walls, which makes obtaining high order implementation of the boundary conditions difficult. The vortex-streamfunction formulation decouples the equations, but at the expense of introducing integral boundary conditions for the vorticity, which negatively impacts the efficiency. An interesting variation of the vortex-streamfunction approach, using only local boundary conditions, is presented in reference [49].

Closely related to the immersed boundary methods are the *penalty* (alternatively: *fictitious domain* or *domain embedding*) methods — e.g. see [8, 40, 62]. These methods, effectively, replace solid walls in the fluid by a porous media with a small porosity $0 < \eta \ll 1$. In the limit $\eta \to 0$, this yields no slip and no flow-through at the solid walls. Two important advantages of this approach are that complicated domains are easy to implement, and that the total fluid-solid force can be computed using a volume integral, rather than an integral over the boundary of the solid. Unfortunately, the parameter η introduces $\sqrt{\eta}$ boundary layers which make convergence slow and high accuracy computations expensive, since η cannot be selected independently of the numerical grid size. *Spectral methods* [18, 42] are very efficient and accurate, but can have grid generation issues for complicated geometries.

1.2.3 Pressure Poisson equations

Finally, we mention the algorithms based on a *Pressure Poisson Equation (PPE) reformulation of the Navier-Stokes equations* [46, 57, 58, 64, 84, 86, 90]. The work presented in chapter 2 falls within this class of methods. In this approach the incompressibility constraint for the flow velocity is replaced by a Poisson[7] equation for the pressure. This then allows an extra boundary condition which must be selected so that incompressibility is maintained by the resulting system. Such a strategy was first proposed by Gresho and Shani [46], who pointed out that adding $\nabla \cdot \mathbf{u} = 0$ as a boundary condition yields a system of equations that is equivalent to the Navier-Stokes equations. Unfortunately, their particular PPE formulation incorporates no explicit boundary condition that can be used to recover the pressure from the velocity through the solution of a Poisson problem. In [51, 52] the issue was resolved at the discrete numerical level, where they demonstrate high order schemes. For instance [51] demonstrates a fourth-order in space and second-order in time implementation using overlapping grids. Subsequent work at the continuum level was then introduced by Henshaw et. al. [53] and Johnston and Liu [58]. Recently, work in PPE formulations have led to interesting analysis and improvements on projection methods [72]. In chapter 2, we present another PPE system, also equivalent to the Navier-Stokes equations, which allows an explicit recovery of the pressure given the flow velocity.

PPE reformulations of the Navier-Stokes equations, such as the one we present in chapter 2, or in reference [58], have important advantages over the standard form of the equations.

---

[7] The choice of the Poisson equation for the pressure is not unique, e.g. see [57].

First, the pressure is not implicitly coupled to the velocity through the momentum equation and incompressibility. Hence it can be directly (and efficiently) recovered from the velocity field by solving a Poisson equation. This allows one to march the velocity field in time using the momentum equation, with the pressure interpreted as some (complicated) function of the velocity. Second, no spurious boundary layers are generated for neither the velocity, nor the pressure. This follows since:

- There are no ambiguities as to which boundary conditions to use for the pressure — hence errors induced by not-quite-correct boundary conditions do not occur[8].
- Incompressibility is enforced at all times.

Hence pressure and velocities that are accurate everywhere can be obtained, in particular: near the boundaries. Finally, PPE formulations allow, at least in principle, for the systematic generation of higher order approximations.

It follows that PPE strategies offer the promise of a resolution for some of the difficulties with evolving the incompressible Navier-Stokes equations. They retain many of the advantages that have made projection methods popular, while not suffering from the presence of numerical boundary layers, or restrictions in temporal accuracy. On the negative side, the boundary conditions for PPE systems tend to be more complicated than the simple ones from other methods.

In chapter 2 we present a PPE formulation, using an alternative form of the velocity boundary conditions, that allows a complete splitting of the momentum and pressure equations. Namely, the pressure can be recovered from the flow velocity without boundary condition ambiguities. In addition, we resolve some of the numerical issues that arise when solving the resulting PPE formulation through the introduction of an extended Navier-Stokes system. This extension is done at the continuum (PDE) level so that the resolution remains independent of any numerical details. Moreover, the system is valid for an extended class

---

[8] Note that these errors should not be confused with the truncation errors that any discretization of the equations will produce. Truncation errors are controlled by the order of the approximation and, for smooth solutions, are uniformly small

of initial conditions, and for smooth solutions contain the Navier-Stokes equations as an attractor. Lastly, we describe a second order solver for the new system on an irregular domain embedded within a Cartesian staggered grid.

1.3 The Klein-Gordon Equation

The Klein-Gordon equation describes a wide variety of phenomena, including both classical wave systems, such as the displacement of a string attached to an elastic bed [105], or semi-classical and quantum systems based on scalar field theories [104]. Despite Klein-Gordon's simplicity, the fact that the equation is both dispersive, yet also hyperbolic, with the possibility of adding nonlinear potentials (such as a quartic φ^4 potential) has lead to a variety of interesting problems for the solutions.

As with many equations in physics, the Klein-Gordon equation often arises through the application of Newton's law, or the conservation of energy. For instance, in classical wave systems, such as a string on an elastic bed, application of Newton's law $\mathbf{f} = m\mathbf{a}$ to the string yields

$$\rho\partial_t^2 h - T\partial_x^2 h + kh + V'(h) = 0, \tag{1.6}$$

where $h(x,t)$ is the height of the string, ρ, T are the density and tension of the string, k is the linear elastic constant of the bed, and $V(h)$ describes a nonlinear restoring potential. Moreover, in quantum systems, identifying energy and momentum with the operators $\widehat{E} = \imath\hbar\partial_t$ and $\widehat{p} = -\imath\hbar\partial_x$, followed by the relation $E^2 = m^2c^4 + p^2c^2$ yields a scalar field equation of the form

$$\hbar^2\partial_t^2\varphi - \hbar^2\partial_x^2\varphi + m^2c^4\varphi + V'(\varphi) = 0, \tag{1.7}$$

where we have retained the additional $V'(\varphi)$ term to include nonlinear effects.

In both equations (1.6–1.7), the addition and relative sign of the nonlinearity $V(h)$ or $V(\varphi)$ can drastically change the nature of the wave solutions. Of particular importance is the resulting behavior over long periods of time. For example, depending on the type of physical system, the long time wave behavior can govern what we hear, see or observe. We will now discuss more generally the physical importance, and types of behavior that may result in systems with positive and negative nonlinearities.

1.3.1 Positive nonlinearities and thermalization

In this section we provide some background on Klein-Gordon systems with positive nonlinearities. For example, modern applications of the classical Klein-Gordon wave equation, with a positive $V(\varphi) = \varphi^4$ potential, include models of both early cosmology and ultrarelativistic heavy ion collisions [14]. In the context of a string on an elastic bed, the addition of a $V(h) = h^4$ potential acts to harden the response of the bed.

As we will show in chapter 3, a positive nonlinearity tends to alter the linear Klein-Gordon dispersion relation of $\omega^2 = m^2 + k^2$, while also redistributing the wave energy throughout different Fourier modes. For a simple analytic discussion, Whitham [105] outlines the dispersive effects for a weak nonlinearity in the Klein-Gordon equation. As discussed in chapter 3, if the waves are restricted to a fixed domain, the continuous sharing of energy between Fourier modes tends to approach a reasonably stable distribution. In these cases, the wave field is said to thermalize (a more precise definition of thermalization may be found in chapter 3). In applications such as the early universe [14], understanding thermalization for different out of equilibrium initial conditions may lead to a deeper understanding of a waves long time behavior. Specifically, recent studies have focused on the thermalization for the Klein-Gordon equation in both quantum [59, 17], and classical [14, 2, 1] field theories. These studies indicate that generic initial wave fields tend to thermalize into a state with large quantities of energy in a wide range of Fourier modes.

In previous studies [1, 2, 14, 89, 91, 66], the thermalization of the classical Klein-Gordon equation has been examined with an emphasis on the applications to quantum field theory or quantum systems. For example, Boyanovsky *et al.* [14] examine the approach to thermalization for out of equilibrium initial conditions. Meanwhile, Aarts *et al.* [2] consider both the thermalization of single fields as well as the statistical average of many initial fields

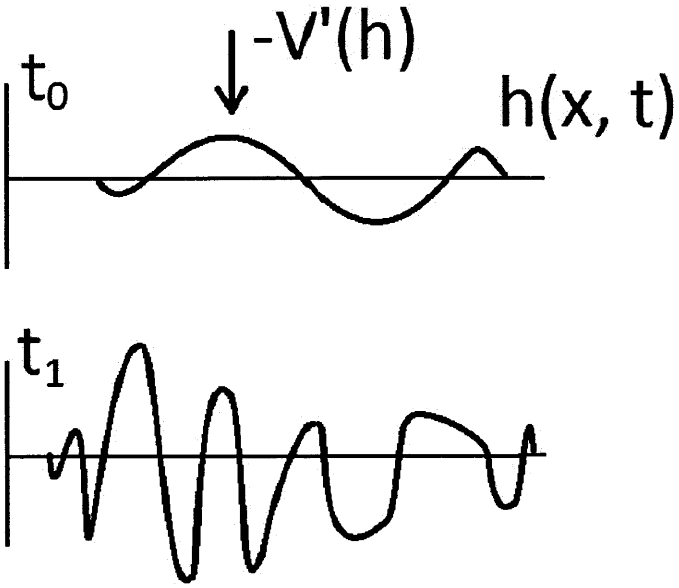

Figure 1-2: The diagram loosely illustrates how a string with some initial condition (at time t_0) starts to thermalize under a positive nonlinearity (time t_1).

(canonical ensemble average). Moreover, they also consider the interesting case of a strong nonlinearity as a nonperturbative system. One general trend in the previous work is the existence of a local thermodynamic equilibrium (LTE). Here an LTE refers to a wave solution which exhibits characteristics of a thermal equilibrium, such as a stable sharing of energy between Fourier modes. The term *local*, however, refers to the fact that such distributions are defined only locally in time, and may drift slowly over longer time scales.

One trend in the previous work on thermalization, is the use an effective Green's function (two point correlation function) [2, 14] or the application of functional integrals for computing thermodynamic quantities [66, 91]. In contrast, a recent approach for studying classical systems is through the use of renormalized waves introduced by Gershgorin *et al.* [34, 35]. Although the introduction of renormalized waves has been primarily restricted to the case of a finite lattice, a recent study [70], uses the approach for the infinite dimensional, Majda-McLaughlin-Tabak wave system. Specifically, they examine the resulting renormalized dispersion relation, and demonstrate how the new dispersion relation can effect the dynamics of the wave resonance structure.

In analogy with work by Gershgorin *et al.*, in chapter 3 we show that the thermal (LTE) state of the Klein-Gordon wave equation may be examined using renormalized waves. Specifically, we show that the resulting renormalized basis exhibits features of a weakly nonlinear system, even in the presence of strong nonlinearities. As a result, we obtain a simple form for the renormalized wave dispersion relation. The net result is a mass shift in the Klein-Gordon dispersion relation that is related to the waves mean field. Moreover, we find that the effective mass shift is different than that suggested by the simple (Hartree) nonperturbative approach. Lastly, we verify that the stable sharing of energy between Fourier modes achieves a Planck-like spectrum. Namely there is equipartition of energy in the low frequency modes, followed by an exponential decay in the high modes. In the classical case of a string on an elastic bed, the characterization of the LTE may be posed as understanding the long time-averaged sound of the string. For instance, what frequencies does the string make (dispersion relation), and how loud do they sound (energy distribution)?

1.3.2 Negative nonlinearities and oscillons

A number of physical phenomenon from water waves traveling in narrow canals [87], to phase transitions in the early Universe [32] exhibit the formation of localized, energy density configurations. The reason for their longevity are varied. Some configurations are stable due to conservation of charge, while some are stable due to a dynamical balance between the nonlinearities and dissipative forces.

Relativistic, scalar field theories (with nonlinear potentials) form simple yet interesting candidates for studying such phenomenon. Some well-studied examples include topological solitons in the $1+1$-dimensional Sine-Gordon model and nontopological solitons such as Q-balls [21]. The Sine-Gordon soliton is stationary in time whereas the Q-balls are oscillatory in nature. Both have conserved charges which make them stable (at least without coupling to gravity). In chapter 4 we discuss another interesting example of such localized configurations called oscillons (also called breathers). Like the Sine-Gordon soliton, they can exist in real scalar fields, and like the Q-balls they are oscillatory in nature. Unlike both of the above examples they do not have any known conserved charges (however, see [61] for an adiabatic invariant). In general they decay, however their lifetimes are significantly longer than any natural time scales present in the Lagrangian. Along with their longevity, another fascinating aspect of oscillons is that they emerge naturally from relatively arbitrary initial conditions. For instance, see [4] for a study of oscillons emerging from fluctuations in a homogeneous background field.

Oscillons first made their appearance in the literature in the 1970s [13]. They were subsequently rediscovered in the 1990s [37]. Oscillons are not exact solutions and (very slowly) radiate their energy away. The amplitude of the outgoing radiation (in the small amplitude expansion) has been calculated by a number of authors, see for example [29, 30, 92]. Characterization of their lifetimes and related properties using the "Gaussian" ansatz for the spatial profile was done in [39] (also see references therein). The importance of the dimensionality of space for these objects has been discussed in [38, 88].

Their possible applications in early Universe physics has not gone unnoticed. For example, they could be relevant for axion dynamics near the QCD phase transition [65]. The properties of oscillons in a $1+1$-dimensional expanding universe (in the small amplitude limit) have been discussed in [24, 44]. Their importance during bubble collisions and phase transitions have been discussed in [22]. In [55], interactions of oscillons with each other and with domain walls were studied in $2+1$ dimensions.

Not all scalar field theories support oscillons, however, we note that the requirement is satisfied by a large number of physically well-motivated examples. For example, the potential for the axion, as well as almost any potential near a vacuum expectation value related to symmetry breaking, support oscillons. Oscillons have also been found in the restricted standard model $SU(2) \times U(1)$, [25, 43, 44].

Since oscillons are oscillatory, localized field configurations, we find it convenient to visualize them as spatially localized wave envelopes, oscillating with a constant frequency. To get a heuristic understanding for the types of potentials that support oscillons, let us consider the Klein-Gordon equation (1.7) with an even nonlinear potential $V(-\varphi) = V(\varphi)$. For a small amplitude oscillon, we may seek an ansatz of the form $\varphi(t,x) \sim \Phi(x)\cos[\omega t]$, where $|\Phi| \ll 1$ for all x[9]. To have a localized field configuration, as we move far enough away from the center (whereby the nonlinearity in the potential is irrelevant), the oscillon must satisfy the linearized equation (we have set $\hbar = c = 1$):

$$-\omega^2\Phi - \partial_x^2\Phi + m^2\Phi \sim 0. \tag{1.8}$$

Again, because we are looking for smooth, localized configurations which vanish exponentially as $x \to \infty$, we must have $\omega^2 < m^2$. In addition, for the lowest energy oscillon configuration, we expect that the field decays monotonically to zero (ie. has no nodes), and is an even symmetric function about the origin. As a result, $\partial_x^2\Phi = 0$ for some value of x, and hence $\partial_x^2\Phi < 0$ for some range of x. Without loss of generality in the following argument, assume that $\partial_x^2\Phi < 0$ at $x = 0$, so that $(m^2 - \omega^2)\Phi - \partial_x^2\Phi|_{x=0} < 0$. Also, the linear profile approximately satisfies the equation

$$[(m^2 - \omega^2)\Phi - \partial_x^2\Phi]\cos(\omega t) \sim -V'[\cos(\omega t)\Phi]. \tag{1.9}$$

Multiplying (1.9) by $\cos(\omega t)$ and integrate over one period, at $x = 0$ we obtain

$$\int_0^{2\pi} V'[\cos(\theta)\Phi_0]\cos(\theta)\,\mathrm{d}\theta < 0, \tag{1.10}$$

$$\tag{1.11}$$

where $\Phi_0 = \Phi(0)$. For a symmetric potential, the product $V'(x)x$ is also an even function. The relation (1.10) then implies

$$\int_0^{1} V'[x\Phi_0]\frac{x}{\sqrt{1-x^2}}\,\mathrm{d}x < 0. \tag{1.12}$$

$$\tag{1.13}$$

---

[9] The ansatz may be justified by retaining the first term in a small amplitude expansion for φ.

By the mean value theorem for integrals, there exists an x^* with $0 < x^* < 1$ such that

$$V'[x^*\Phi_0]\frac{x^*}{\sqrt{1-(x^*)^2}} < 0, \tag{1.14}$$

Hence, for oscillons to exist in potentials with a quadratic minimum, we require the nonlinearity $V' < 0$ for some range of field values.

The above (heuristic) argument does not provide a reason for the longevity of oscillons. For oscillons, their shape, which determines their Fourier content, guarantees that the amplitude at the wave number of the outgoing radiation is exponentially suppressed (at least for the small amplitude oscillons). For details see [54] and the subsection in chapter 4 on radiation.

It is not too difficult to think of physically motivated potentials satisfying the above requirement. For example, the potential for the QCD axions $V(\varphi) = m^2f^2\left[1-\cos(\varphi/f)\right]$ where f is the Peccei-Quinn scale and m is the mass, or any symmetry breaking potential expanded about it's vacuum expectation value. Both potentials "open up" a little when we move away from the minimum.

In chapter 4, we examine oscillons in a scalar field theory for a class of nonlinear potentials. To study individual oscillon properties, we first consider a theory without expansion. We then derive the oscillon frequency, as well as the spatial profile, for both $1+1$ (analytically) and $3+1$ (numerically) dimensions. In particular, we show a nonmonotonic relationship between the height and the width of the oscillons, and discuss their stability to small perturbations (see [69] for a somewhat related analysis for Q-balls). For instance, in the model under consideration, large amplitude oscillons become very wide and develop flat tops. In contrast to the small amplitude oscillons, these flat topped oscillons may be physically important since they are more stable to long wavelength perturbations. Secondly, since oscillons could have important applications in cosmology, especially in the early Universe, we discuss the effect of expansion on their profiles and lifetimes. Specifically, when oscillons are placed in an expanding universe, we show the effect of expansion can tend to stretch the energy in the oscillon tails. As a result, there is additional outgoing radiation which can reduce the oscillon lifetime.

The work in chapter 4 appears as a first step towards understanding oscillons in an expanding universe. Subsequent work [5] shows that the effect of expansion does not tend to significantly alter the stability results we derive in chapter 4. Hence, aside from a reduced lifetime, qualitative oscillon properties, such as profile shapes and stability, remain intact in an expanding universe. There are, however, quantitative properties that do change. For example, the effect of expansion can tend to modify the growth of oscillons [4]. Lastly, expansion may also be of practical importance as a mechanism to first spread out generic initial conditions which then expedites the formation of oscillons. Note however, expansion may not be a necessary ingredient for the spontaneous formation of many oscillons. For instance, even in the absence of expansion, parametric instabilities about a homogeneous background field can potentially generate oscillons.

Chapter 2

A Pressure Poisson Approach for Navier-Stokes

2.1 Motivation

As outlined in the introduction, there is often a trade off between accuracy and efficiency when computing numerical solutions to the incompressible Navier-Stokes equations. For example, many common efficient schemes for the incompressible Navier-Stokes equations, such as projection or fractional step methods, have limited temporal accuracy as a result of matrix splitting errors, or introduce errors near the domain boundaries.

In this chapter we present a reformulation of the incompressible Navier-Stokes equations, and a corresponding numerical method to solve the resulting system [95]. The advantage of our approach is that we maintain both numerical accuracy and computational efficiency. Specifically, we recast the constant density Navier-Stokes equations, with velocity prescribed boundary conditions for the primary variables velocity and pressure. We do this in the usual way away from the boundaries, by replacing the incompressibility condition on the velocity with a Poisson equation for the pressure. The key difference from other common methods occurs at the boundaries, where we use boundary conditions that unequivocally allow the pressure to be recovered from knowledge of the velocity at any fixed time. This avoids the common difficulty of an, apparently, over-determined Poisson problem. Since in this alternative formulation the pressure can be accurately and efficiently recovered from the velocity, the recast equations are ideal for numerical marching methods. The new system can be discretized using a variety of methods, in principle to any desired order of accuracy.

This chapter begins with an introduction to the Navier-Stokes equations and the pressure Poisson approach. We first discuss our new formulation in section 2.3 and outline potential difficulties that may arise in a numerical implementation of the new system. In sections 2.4 and 2.5 we resolve these potential difficulties by introducing an extended Navier-Stokes system. The resulting extended system is then suitable for numerical implementation.

In section 2.7 we illustrate the new approach with a 2-D second order finite difference scheme on a Cartesian grid. Here we devise an algorithm to solve the equations on domains with curved (non-conforming) boundaries and in section 2.8 we present tests for the scheme in both a square domain, and a domain with a circular obstruction. Our tests verify that the algorithm achieves second order accuracy in the L^∞ norm, for both the velocity and the pressure.

Finally, we conclude the chapter with a section on potential future work. These including the starting framework for a Galerkin formulation of the new system, as well as an introduction to a second pressure Poisson reformulation of the Navier-Stokes equations.

2.2 The Pressure Poisson Equation.

In this section we introduce the well-known pressure Poisson equation (PPE), and use it to construct a system of equations (and boundary conditions) equivalent to the constant-density (hence incompressible) Navier-Stokes equations, with the velocity prescribed at the boundaries. Specifically, consider the incompressible Navier-Stokes equations in a connected domain $\Omega \in \mathbb{R}^D$, where $D = 2$ or $D = 3$, with a piece-wise smooth boundary $\partial\Omega$. Inside Ω, the flow velocity field $\mathbf{u}(\mathbf{x},\, t)$ satisfies the equations

$$\begin{aligned} \mathbf{u}_t + (\mathbf{u}\cdot\nabla)\,\mathbf{u} &= \mu\,\Delta\mathbf{u} - \nabla p + \mathbf{f}, && (2.1)\\ \nabla\cdot\mathbf{u} &= 0, && (2.2) \end{aligned}$$

where μ is the kinematic viscosity,[1] $p(\mathbf{x},\, t)$ is the pressure, $\mathbf{f}(\mathbf{x},\, t)$ are the body forces, ∇ is the gradient, and $\Delta = \nabla^2$ is the Laplacian. Equation (2.1) follows from the conservation of momentum, while (2.2) is the incompressibility condition (conservation of mass).

In addition, the following boundary conditions apply

$$\mathbf{u} = \mathbf{g}(\mathbf{x},\, t) \qquad \text{for } \mathbf{x} \in \partial\Omega, \tag{2.3}$$

where

$$\int_{\partial\Omega} \mathbf{n}\cdot\mathbf{g}\,\mathrm{d}A = 0, \tag{2.4}$$

---

[1] We work in non-dimensional variables, so that $\mu = 1/Re$ (where Re is the Reynolds number) and the fluid density is $\rho = 1$.

$\mathbf{n}$ is the outward unit normal on the boundary, and dA is the area (length in 2D) element on $\partial\Omega$. Equation (2.4) is the consistency condition for $\mathbf{g}$, since an incompressible fluid must have zero net flux through the boundary.

Finally, we assume that initial conditions are given

$$\begin{aligned} \mathbf{u}(\mathbf{x}, 0) &= \mathbf{u}_0 && \text{for } \mathbf{x} \in \Omega, && (2.5)\\ \nabla \cdot \mathbf{u}_0 &= 0 && \text{for } \mathbf{x} \in \Omega, && (2.6)\\ \mathbf{u}_0(\mathbf{x}) &= \mathbf{g}(\mathbf{x}, 0) && \text{for } \mathbf{x} \in \partial\Omega. && (2.7) \end{aligned}$$

Remark 2.1 *Of particular interest is the case of fixed impermeable walls, where no flux* $\mathbf{u} \cdot \mathbf{n} = 0$ *and no slip* $\mathbf{u} \times \mathbf{n} = 0$ *apply at* $\partial\Omega$. *This corresponds to* $\mathbf{g} = 0$ *in (2.3). Note that the no-slip condition is equivalent to* $\mathbf{u} \cdot \mathbf{t} = 0$ *for all unit tangent vectors* $\mathbf{t}$ *to the boundary.* ♠

Remark 2.2 *In this chapter we will assume that the domain* Ω *is fixed. Situations where the boundary of the domain,* $\partial\Omega$*, can move — either by externally prescribed factors, or from interactions with the fluid, are of great physical interest.* ♠

Next we introduce the Pressure Poisson Equation (PPE). To obtain the pressure equation, take the divergence of the momentum equation (2.1), and apply equation (2.2) to eliminate the viscous term and the term with a time derivative. This yields the following Poisson equation for the pressure

$$\Delta p = \nabla \cdot (\mathbf{f} - (\mathbf{u} \cdot \nabla)\,\mathbf{u}). \tag{2.8}$$

Two crucial questions are now (for simplicity, *assume solutions that are smooth all the way up to the boundary*)

2.2a *Can this equation be used to replace the incompressibility condition (2.2)?* Since — given (2.8) — the divergence of (2.1) yields the heat equation

$$\phi_t = \mu \Delta \phi, \tag{2.9}$$

for $\phi = \nabla \cdot \mathbf{u}$ and $\mathbf{x} \in \Omega$, it would seem that the answer to this question is yes — provided that the initial conditions are incompressible (*i.e.:* $\phi = 0$ for $t = 0$). However, this works only if we can guarantee that, at all times,

$$\phi = 0, \tag{2.10}$$

for $\mathbf{x} \in \partial\Omega$.

2.2b *Given the flow velocity* $\mathbf{u}$*, can (2.8) be used to obtain the pressure* p*?* Again, at first sight, the answer to this question appears to be yes. After all, (2.8) is a Poisson equation for p, which should determine it uniquely — given appropriate boundary conditions. The problem is: what boundary conditions? Evaluation of (2.1) at the boundary, with use of (2.3), shows that the flow velocity determines the whole gradient of the pressure at the boundary, which is too much for (2.8). Further, if only a portion of these boundary conditions are enforced when solving (2.8) — say, the normal component of (2.1) at the boundary, then how can one be sure that the whole of (2.1) applies at the boundary?

Remark 2.3 *From an algorithmic point of view, an affirmative answer to the questions above would very useful, for then one could think of the pressure as some (global) function of the flow velocity, in which case (2.1) becomes an evolution equation for* $\mathbf{u}$*, which could then be solved with a numerical "marching" method.* ♠

The issue in item 2.2a can be resolved easily, and we do so next. We postpone dealing with the issue in item 2.2b till the next section, § 2.3. Since the addition of an equation for the pressure allows the introduction of one extra boundary condition, we propose to replace the system in (2.1), (2.2), and (2.3) by the following Pressure Poisson Equation (PPE) formulation:

$$\begin{aligned} \mathbf{u}_t + (\mathbf{u} \cdot \nabla)\,\mathbf{u} &= \mu\,\Delta\mathbf{u} - \nabla p + \mathbf{f}, & (2.11) \\ \Delta p &= \nabla \cdot (\mathbf{f} - (\mathbf{u} \cdot \nabla)\,\mathbf{u}), & (2.12) \end{aligned}$$

for $\mathbf{x} \in \Omega$, with the boundary conditions

$$\mathbf{u} = \mathbf{g}(\mathbf{x}, t), \tag{2.13}$$

$$\nabla \cdot \mathbf{u} = 0, \tag{2.14}$$

for $\mathbf{x} \in \partial\Omega$ — where, of course, the restriction in (2.4) still applies. The extra boundary condition is precisely what is needed to ensure that the pressure enforces incompressibility throughout the flow (see item 2.2a)

Remark 2.4 *It can be seen that for smooth enough solutions* $(\mathbf{u}, p)$*, the pair of equations (2.11–2.12), accompanied by the boundary conditions (2.13–2.14), are equivalent to the incompressible Navier-Stokes equations in (2.1), (2.2), and (2.3). Of course, this result is not new. This reformulation of the Navier-Stokes equations was first presented by Gresho and Sani [46] — it can also be found in reference [84]. However, it should be pointed out that Harlow and Welch [50], in their pioneering work, had already noticed that the boundary condition* $\nabla \cdot \mathbf{u} = 0$ *was needed to guarantee, within the context of their MAC scheme, that* $\nabla \cdot \mathbf{u} = 0$ *everywhere.*

This formulation does not provide any boundary conditions for the pressure, which means that one ends up with a "global" constraint on the solutions to the Poisson equation (2.12). Hence it does not yield a satisfactory answer to the issue in item 2.2b, since recovering the pressure from the flow velocity is a hard problem with this approach.

Direct implementations of (2.11–2.14) have only been proposed for simple geometries — in particular: grid-conforming boundaries. In [64] a spectral algorithm for plane channel flows is presented. We have already mentioned [50], where they use a staggered grid on a rectangular domain, and the condition $\nabla \cdot \mathbf{u} = 0$ *is used (at the discrete level) to close the linear system for the pressure. In [57], by manipulating the discretization of the boundary conditions and of the momentum equation (2.11), they manage to obtain "local" approximate Neumann conditions for the pressure — both on rectangular, as well as circular, domains where* $\mathbf{u} = 0$ *on the boundary. Unfortunately, the approaches in [50, 57] seem to be very tied* *up to the details of particular discretizations, and require a conforming boundary.* ♠

2.3 Theoretical Reformulation

We still need to deal with the issue raised in item 2.2b. In particular, in order to implement the ideas in remark 2.3, we need to split the boundary conditions for (2.11–2.12), in such a way that: (i) there is a specific part of the boundary conditions that is used with (2.11) to advance the velocity field in time, given the pressure. (ii) The remainder of the boundary conditions is used with (2.12) to solve for the pressure — at each fixed time, given the flow velocity field $\mathbf{u}$. This is the objective of this section.

The conventional approach in projection or fractional step methods is to associate (2.13) with equation (2.11). This is reasonable for evolving the heat-like equation (2.11). Unfortunately, it has the drawback of only implicitly defining the boundary conditions for the Poisson equation (2.12). Specifically, the correct pressure boundary conditions are those that guarantee $\nabla \cdot \mathbf{u} = 0$ for $\mathbf{x} \in \partial\Omega$, and such boundary conditions cannot easily be known a priori as a function of $\mathbf{u}$. Hence one is left with a situation where the appropriate boundary conditions for the pressure are not known. This leads to errors in the pressure, and in the incompressibility condition, which are difficult to control. In particular, errors are often most pronounced near the boundary where no local error estimates can be produced, even for smooth solutions. By the latter we mean that the resulting schemes cannot be shown to be consistent, all the way up to the boundary, in the classical sense of finite differences introduced by Lax [68].

In our formulation we take a different approach, which has a similar spirit to the one used by Johnston and Liu [58] — see remark 2.5. Rather than associate all the D components of (2.13) with equation (2.11), we enforce the $D-1$ tangential components only, and complete the set of boundary conditions for (2.11) with (2.14). Hence, when evolving equation (2.11) we do not specify the normal velocity on the boundary, but — through the divergence condition (2.14) — specify the normal derivative of the normal velocity. Finally, the (as yet unused) boundary condition on the normal velocity, $\mathbf{n} \cdot (\mathbf{u} - \mathbf{g}) = 0$ for $\mathbf{x} \in \partial\Omega$, is employed to obtain an explicit boundary condition for equation (2.12). We do this by requiring that the pressure boundary condition be equivalent to $(\mathbf{n}\cdot(\mathbf{u}-\mathbf{g}))_t = 0$ for $\mathbf{x}\in\partial\Omega$ — which then guarantees that the normal component of (2.13) holds, as long as the initial conditions satisfy it. This objective is easily achieved: dotting equation (2.11) through with $\mathbf{n}$, and evaluating at the boundary yields the desired condition. The equations, with their appropriate boundary conditions, are thus:

$$\left.\begin{aligned} \mathbf{u}_t-\mu\Delta\mathbf{u} &= -\nabla p-(\mathbf{u}\cdot\nabla)\,\mathbf{u}+\mathbf{f} && \text{for } \mathbf{x}\in\ \Omega,\\ \mathbf{n}\times(\mathbf{u}-\mathbf{g}) &= 0 && \text{for } \mathbf{x}\in\partial\Omega,\\ \nabla\cdot\mathbf{u} &= 0 && \text{for } \mathbf{x}\in\partial\Omega, \end{aligned}\right\} \tag{2.15}$$

and

$$\left.\begin{aligned} \Delta p &= -\nabla\cdot((\mathbf{u}\cdot\nabla)\,\mathbf{u})+\nabla\cdot\mathbf{f} && \text{for } \mathbf{x}\in\ \Omega,\\ \mathbf{n}\cdot\nabla p &= \mathbf{n}\cdot(\mathbf{f}-\mathbf{g}_t+\mu\Delta\mathbf{u}-(\mathbf{u}\cdot\nabla)\,\mathbf{u}) && \text{for } \mathbf{x}\in\partial\Omega. \end{aligned}\right\} \tag{2.16}$$

Again, for smooth (up to the boundary) enough solutions $(\mathbf{u},\, p)$ of the equations: the incompressible Navier-Stokes equations (2.1–2.2), with boundary conditions as in (2.3), are equivalent to the system of equations and boundary conditions in (2.15–2.16).

For the sake of completeness, we display now the calculation showing that the boundary condition splitting in (2.15–2.16) recovers the normal velocity boundary condition $\mathbf{n}\cdot\mathbf{u}=\mathbf{n}\cdot\mathbf{g}$. To start, dot the first equation in (2.15) with the normal $\mathbf{n}$, and evaluate at the boundary. This yields

$$\mathbf{n}\cdot\mathbf{u}_t=\mathbf{n}\cdot(\mu\Delta\mathbf{u}-\nabla p-(\mathbf{u}\cdot\nabla)\,\mathbf{u}+\mathbf{f}) \quad \text{for } \mathbf{x}\in\partial\Omega. \tag{2.17}$$

Next, eliminate $\mathbf{n}\cdot\nabla p$ from this last equation — by using the boundary condition for the pressure in (2.16), to obtain

$$(\mathbf{n}\cdot(\mathbf{u}-\mathbf{g}))_t=0 \quad \text{for } \mathbf{x}\in\partial\Omega. \tag{2.18}$$

This is a trivial ODE for the normal component of the velocity at each point in the boundary. Thus, provided that $\mathbf{n}\cdot(\mathbf{u}-\mathbf{g})=0$ initially, it holds for all time.

An important final point to check is the solvability condition for the pressure problem. Given the flow velocity $\mathbf{u}$ at time t, equation (2.16) is a Poisson problem with Neumann boundary conditions for the pressure. This problem has a solution (unique up to an additive constant) if and only if the "flux equals source" criteria

$$\int_{\partial\Omega} \mathbf{n}\cdot(\mathbf{f}-\mathbf{g}_t+\mu\Delta\mathbf{u}-(\mathbf{u}\cdot\nabla)\,\mathbf{u})\;\mathrm{d}A=\int_{\Omega}\nabla\cdot(\mathbf{f}-(\mathbf{u}\cdot\nabla)\,\mathbf{u})\;\mathrm{d}V, \tag{2.19}$$

applies. This is satisfied because:

2.3a $\int_{\partial\Omega}\mathbf{n}\cdot(\mathbf{f}-(\mathbf{u}\cdot\nabla)\,\mathbf{u})\;\mathrm{d}A=\int_{\Omega}\nabla\cdot(\mathbf{f}-(\mathbf{u}\cdot\nabla)\,\mathbf{u})\;\mathrm{d}V$,

2.3b $\int_{\partial\Omega}\mathbf{n}\cdot\Delta\mathbf{u}\,\mathrm{d}A=\int_{\Omega}\Delta\;(\nabla\cdot\mathbf{u})\;\mathrm{d}V=0$,

2.3c $\int_{\partial\Omega}\mathbf{n}\cdot\mathbf{g}_t\,\mathrm{d}A=\frac{\partial}{\partial t}\int_{\partial\Omega}\mathbf{n}\cdot\mathbf{g}\,\mathrm{d}A=0$,

where we have used Gauss' theorem, incompressibility, and (2.4).

2.4 Modification for Stability

For the system of equations in (2.15–2.16), it is important to notice that

2.4a The tangential boundary condition on the flow velocity, $\mathbf{n}\times(\mathbf{u}-\mathbf{g})=0$ for $\mathbf{x}\in\partial\Omega$, is enforced explicitly.

2.4b The incompressibility condition, $\nabla\cdot\mathbf{u}=0$ for $\mathbf{x}\in\Omega$, is enforced "exponentially". By this we mean that any errors in satisfying the incompressibility condition are rapidly damped, because $\phi=\nabla\cdot\mathbf{u}$ satisfies (2.9–2.10). Thus this condition is enforced in a robust way, and we do not expect it to cause any trouble for "reasonable" numerical discretizations of the equations.

2.4c By contrast, the normal boundary condition on the flow velocity, $\mathbf{n}\cdot(\mathbf{u}-\mathbf{g})=0$ for $\mathbf{x}\in\partial\Omega$, is enforced in a rather weak fashion. By this we mean that errors in satisfying this condition are not damped at all by equation (2.18). Thus, this condition lacks the inherent stability provided by the heat equation. In practice, numerical errors add (effectively) noise to equation (2.18), resulting in a drift of the normal velocity component. This can have de-stabilizing effects on the behavior of a numerical scheme. Hence it is a problem that must be corrected.

In this section we alter the PPE equations (2.15–2.16) to address the problem pointed out in item 2.4c. We do this by adding an appropriate "stabilizing" term. The idea here is similar in nature to the feedback controller [41] proposed to control the boundary velocity for immersed boundary methods, as well as the divergence stabilizing term introduced by Henshaw [51, 53]. Our goal here is to develop a pair of differential equations — fully equivalent to (2.15–2.16) — which are suitable for numerical implementation. In order to resolve the issue in item 2.4c, we add a feedback term to the equations, by altering the pressure boundary condition. Specifically, we modify the equations from (2.15–2.16) to

$$\left.\begin{aligned} \mathbf{u}_t - \mu\,\Delta\mathbf{u} &= -\nabla p - (\mathbf{u}\cdot\nabla)\,\mathbf{u} + \mathbf{f} && \text{for } \mathbf{x}\in\ \Omega, \\ \mathbf{n}\times(\mathbf{u}-\mathbf{g}) &= 0 && \text{for } \mathbf{x}\in\partial\Omega, \\ \nabla\cdot\mathbf{u} &= 0 && \text{for } \mathbf{x}\in\partial\Omega, \end{aligned}\right\} \tag{2.20}$$

and

$$\left.\begin{aligned} \Delta p &= -\nabla\cdot((\mathbf{u}\cdot\nabla)\,\mathbf{u}) + \nabla\cdot\mathbf{f} && \text{for } \mathbf{x}\in\ \Omega, \\ \mathbf{n}\cdot\nabla p &= \mathbf{n}\cdot(\mathbf{f}-\mathbf{g}_t+\mu\,\Delta\mathbf{u}-(\mathbf{u}\cdot\nabla)\,\mathbf{u}) && \\ &+\ \lambda\,\mathbf{n}\cdot(\mathbf{u}-\mathbf{g}) && \text{for } \mathbf{x}\in\partial\Omega, \end{aligned}\right\} \tag{2.21}$$

where $\lambda > 0$ is a numerical parameter — see § 2.4.1. This system is still equivalent to the incompressible Navier-Stokes equations and boundary conditions in (2.1–2.3), for smooth (up to the boundary) enough solutions ($\mathbf{u}$, p). The heat equation (2.9–2.10) for $\phi = \nabla\cdot\mathbf{u}$ still applies, while the equation for the evolution of the normal velocity at the boundary changes from (2.18) to:

$$(\mathbf{n}\cdot(\mathbf{u}-\mathbf{g}))_t = -\lambda\,\mathbf{n}\cdot(\mathbf{u}-\mathbf{g}) \quad \text{for } \mathbf{x}\in\partial\Omega. \tag{2.22}$$

Thus, if $\mathbf{n} \cdot (\mathbf{u} - \mathbf{g}) = 0$ initially, it remains so for all times. In addition, this last equation shows that this new system resolves the issue pointed out in item 2.4c.

Finally, we check what happens to the solvability condition for the pressure problem, given the system change above. Clearly, all we need to do is to modify equation (2.19) by adding — to its left hand side, the term

$$\lambda \int_{\partial\Omega} \mathbf{n} \cdot (\mathbf{u} - \mathbf{g}) \, \mathrm{d}A = \lambda \int_{\Omega} \nabla \cdot \mathbf{u} \, \mathrm{d}V - \lambda \int_{\partial\Omega} \mathbf{n} \cdot \mathbf{g} \, \mathrm{d}A = 0, \tag{2.23}$$

where we have used incompressibility, and equation (2.4). It follows that solvability remains valid.

Remark 2.5 *The reformulation of the Navier-Stokes equations in (2.20–2.21) is similar to the one used by Johnston and Liu in [58]. In their paper the authors propose methods of solution to the Navier-Stokes equations based on a similar system (for simplicity, we set* $\mathbf{g} = 0$*, as done in [58]) where*

(a) For $\mathbf{x} \in \Omega$*, the same equations as in (2.20–2.21) apply.*

(b) For $\mathbf{x} \in \partial\Omega$*,* $\mathbf{u} = 0$ *is used for the momentum equation.*

(c) For $\mathbf{x} \in \partial\Omega$*,* $\mathbf{n} \cdot \nabla p = \mathbf{n} \cdot (-\mu \nabla \times \nabla \times \mathbf{u} + \mathbf{f})$ *is used for the Poisson equation.*

For this system the incompressibility condition $\phi = \nabla \cdot \mathbf{u} = 0$ *follows because these equations yield*

(d) $\phi_t = \mu \Delta \phi$ *for* $\mathbf{x} \in \Omega$*, with* $\mathbf{n} \cdot \nabla \phi = 0$ *for* $\mathbf{x} \in \partial\Omega$*.*

Thus, if ϕ *vanishes initially, it will vanish for all times. Note that the* $\mathbf{n} \cdot \nabla \phi = 0$ *boundary condition follows from using* $\mathbf{n} \cdot (\nabla \times \nabla \times \mathbf{u})$ *instead of* $\mathbf{n} \cdot (\Delta \mathbf{u})$ *in the pressure boundary condition.*

(e) The main advantage *of this reformulation over the one in (2.20–2.21) is that depending on the implementation, the boundary condition* $\nabla \cdot \mathbf{u} = 0$ *in (2.20) may couple the components of the flow velocity field — see § 2.7.3. Thus an implicit treatment of the*

viscous terms in (2.20–2.21) would be more expensive (and complicated) than for the system in items a–c above. However, it is not clear to us at this moment how much of a problem this is. The reason is that the coupling is "weak", by which we mean that: in the $N_G \times N_G$ *discretization matrix for the Laplacian — where* N_G *is the number of points in the numerical grid, the coupling induced by the boundary condition affects only* $O(N_G^{1/2})$ *entries in 2-D, and* $O(N_G^{1/3})$ *entries in 3-D — at least with the type of discretization that we use in § 2.7.3.*

(f) The velocity divergence is controlled through the damping of the heat equation. If the heat equation cannot suppress the divergence fast enought (for example in the case of large Reynolds numbers), the method can be modified using an idea by Henshaw [51, 53]. For example: modify the Poisson equation for the pressure to

$$\Delta p = -\nabla \cdot ((\mathbf{u} \cdot \nabla)\,\mathbf{u}) + \nabla \cdot \mathbf{f} + \lambda\, \nabla \cdot \mathbf{u}.$$

This then changes the system in item d to

$$\phi_t = \mu\, \Delta\, \phi - \lambda\, \phi \;\; for \;\; \mathbf{x} \in \Omega, \;\; with \;\; \mathbf{n} \cdot \nabla\, \phi = 0 \;\; for \;\; \mathbf{x} \in \partial\Omega.$$

Remark 2.6 *It would be nice to be able to use* $\lambda = \lambda(\mathbf{x})$*, so as to optimize the implementation of the condition* $\mathbf{n} \cdot \mathbf{u} = \mathbf{n} \cdot \mathbf{g}$ *for different points along* $\partial\Omega$*. However, this is not a trivial extension, since it destroys the solvability condition for the pressure. Thus, other (compensating) corrections are needed as well. We postpone the study of this issue for future work.* ♠

Remark 2.7 *As show earlier in (2.19) and (2.23), the validity of the solvability condition, for the problem in (2.21), relies on the velocity field satisfying the incompressibility constraint* $\nabla \cdot \mathbf{u} = 0$*. On the other hand, in the course of a numerical calculation, the discretization errors result in a small, but non-zero, divergence — thus solvability fails. However, the errors in solvability are small. Hence, a least squares solution of the discretized linear equation for* *the pressure provides an approximation within the order of the method — which is as good as can be expected. These considerations motivate the following theoretical question:* can the equations in (2.20–2.21) be modified, so they make sense even for $\nabla \cdot \mathbf{u} \neq 0$? *In section 2.5 we show that this is possible.* ♠

2.4.1 Selection of the parameter λ.

For numerical purposes, here we address the issue of how large λ should be, by using a simple model for the flow's normal velocity drift. Notice that no precision is needed for this calculation, just order of magnitude. In actual practice, one can monitor how well the normal velocity satisfies the boundary condition, and increase λ if needed. In principle one should be wary of using large values for λ, since this will yield stiff behavior in time. However, the calculation below shows that λ does not need to be very large, and does not depend on the grid size Δx.

It seems reasonable to assume that one can model how the numerical errors affect the ODE (2.22) for $\mathcal{E} = \mathbf{n} \cdot (\mathbf{u} - \mathbf{g})$, by perturbing the coefficients of the equation, and adding a forcing term to it. Hence we modify equation (2.22) as follows

$$\mathcal{E}_t = -\lambda\, c_p\, \mathcal{E} + \epsilon\, \gamma \quad \text{for } \mathbf{x} \in \partial\Omega, \tag{2.24}$$

where $\epsilon \ll 1$ characterizes the size of the errors (determined by the order of the numerical method), while $c_p = c_p(\mathbf{x},\, t) = 1 + O(\epsilon)$ and $\gamma = \gamma(\mathbf{x},\, t) = O(1)$ are functions encoding the numerical errors. What exactly they are depends on the details of the numerical discretization, but for this calculation we do not need to know these details. All we need is that[2]

$$0 < C_M \leq c_p \quad \text{and} \quad |\gamma| \leq \Gamma, \tag{2.25}$$

where C_M and Γ are some positive constants, with $C_M \approx 1$, and $\Gamma = O(1)$ — but not necessarily close to one.

---

[2] For smooth enough solutions, where the truncation errors are controlled by some derivative of the solution.

The solution to (2.24) is given by

$$\mathcal{E} = \mathcal{E}_0 \, e^{-\lambda I_1} + \epsilon \underbrace{\int_0^t \gamma(\mathbf{x},\, s)\, e^{-\lambda I_2}\, \mathrm{d}s}_{J}, \tag{2.26}$$

where $\mathcal{E}_0$ is the initial value, $I_1 = I_1(\mathbf{x},\, t) = \int_0^t c_p(\mathbf{x},\, s)\, \mathrm{d}s$, and $I_2 = I_1(\mathbf{x},\, t) - I_1(\mathbf{x},\, s)$. The crucial term is J, since the first term decreases in size, and starts at the initial value. However

$$|J| \leq \Gamma \int_0^t e^{-\lambda C_M (t-s)}\, \mathrm{d}s \leq \frac{\Gamma}{\lambda C_M}. \tag{2.27}$$

Within the framework of a numerical scheme, the normal boundary velocity may deviate from the prescribed velocity by some acceptable error δ. Thus we require $\epsilon\, J = O(\delta)$ or less, which — given (2.27) — will be satisfied if

$$\lambda \sim \frac{\epsilon\, \Gamma}{\delta\, C_M} \approx \frac{\epsilon\, \Gamma}{\delta}, \quad \text{or larger.} \tag{2.28}$$

For the second order numerical scheme in § 2.7, it is reasonable to expect that $\epsilon = (\Delta x)^2$, and to require that $\delta = (\Delta x)^2$. Then (2.28) reduces to $\lambda \geq \Gamma$. Of course, we do not know (a priori) what Γ is; this is something that we need to find by numerical experimentation — see the first paragraph in this § 2.4.1. For the numerical calculations reported in § 2.8, we found that values in the range $10 \leq \lambda \leq 100$ gave good results.

For this scheme, comparing the time step restriction imposed by feedback ($\Delta t_\lambda < O(1/\lambda)$) to the one imposed by diffusion $\Delta t_\mu < O((\Delta x)^2/\mu)$, results in a stiffness ratio that scales as

$$\frac{\Delta t_\mu}{\Delta t_\lambda} \sim \frac{\lambda}{\mu} \Delta x^2. \tag{2.29}$$

Hence for low to moderate Reynolds numbers, λ can be chosen quite large while maintaining a ratio well below unity. As a result, in these regimes, the feedback term does not introduce additional stiffness in the equations.

2.5 Further Modification for Solvability

In this section we address the question posed and motivated in remark 2.7. Namely: *Can the equations in (2.20–2.21) be modified in such a way that they make sense even for initial conditions that are not incompressible?* In fact, in such a way that if a solution starts with $\nabla \cdot \mathbf{u} \neq 0$ and $\mathbf{n} \cdot (\mathbf{u} - \mathbf{g}) \neq 0$, then (as $t \to \infty$) $\nabla \cdot \mathbf{u} \to 0$ and $\mathbf{n} \cdot (\mathbf{u} - \mathbf{g}) \to 0$ — so that the solution converges towards a solution of the Navier-Stokes equation.

As pointed out in remark 2.7, the problem with (2.20–2.21) is that the solvability condition for the Poisson equation (2.21) is not satisfied when $\nabla \cdot \mathbf{u} \neq 0$. Hence the equations become ill-posed, as they have no solution. The obvious answer to this dilemma is to interpret the solution to (2.21) in an appropriate least squares sense, which is equivalent to modifying the non-homogeneous terms in the equation by *projecting* them onto the space of right hand sides for which the Poisson equation has a solution. Symbolically, write (2.21) in the form

$$\left.\begin{aligned} \Delta p &= g \quad \text{for } \mathbf{x} \in \ \Omega, \\ \mathbf{n} \cdot \nabla p &= h \quad \text{for } \mathbf{x} \in \partial\Omega, \end{aligned}\right\} \tag{2.30}$$

where g and h are defined in (2.21). Then modify the equation to

$$\left.\begin{aligned} \Delta p &= g_p \quad \text{for } \mathbf{x} \in \ \Omega, \\ \mathbf{n} \cdot \nabla p &= h_p \quad \text{for } \mathbf{x} \in \partial\Omega, \end{aligned}\right\} \tag{2.31}$$

where $(g_p,\, h_p) = \mathbb{P}\,(g,\, h)$ for some projection operator $\mathbb{P}$ such that

$$\int_\Omega g_p \,\mathrm{d}V = \int_{\partial\Omega} h_p \,\mathrm{d}A. \tag{2.32}$$

The question, however, is: which projection?

The discretization failures in solvability arising during the course of a numerical calculation are small, since it should be $(g_e,\, h_e) - (g,\, h) = O\left((\Delta\, x)^q\right)$ — where q is the order of the method, and $(g_e,\, h_e)$ is the exact right hand side. Thus one can argue that, as long as $(g_p,\, h_p) - (g,\, h) = O((\Delta\, x)^q)$, the resulting numerical solution will be accurate to within the appropriate order — see remarks 2.7 and 2.11. In this section, however, the aim is to consider situations where there is no small parameter (i.e. Δx) guaranteeing that solvability is "almost" satisfied. In particular, we want to consider situations where $\nabla \cdot \mathbf{u} \neq 0$, and $|\nabla \cdot \mathbf{u}| \ll 1$ does not apply — leading to errors in solvability which are not small. It follows that here we must be careful with the choice of the projection.

Obviously, a very desirable property of the selected projection is that it should preserve the validity of equations (2.9–2.10) — so that the time evolution drives $\nabla \cdot \mathbf{u}$ to zero. Hence: *it must be that* $g_p = g$, with only h affected by the projection. Further: since the solvability condition involves h only via its mean value over $\partial\Omega$, the simplest projection that works is one that appropriately adjusts the mean of h, and nothing else. Thus we propose to modify the equations in (2.20–2.21) as follows: leave (2.20) as is, as well as the imposed boundary condition constraint (2.4), but replace (2.21) by

$$\left.\begin{aligned} \Delta p &= -\nabla \cdot ((\mathbf{u} \cdot \nabla)\, \mathbf{u}) + \nabla \cdot \mathbf{f} && \text{for } \mathbf{x} \in \Omega, \\ \mathbf{n} \cdot \nabla p &= \mathbf{n} \cdot (\mathbf{f} - \mathbf{g}_t + \mu \Delta \mathbf{u} - (\mathbf{u} \cdot \nabla)\, \mathbf{u}) && \\ &+ \lambda\, \mathbf{n} \cdot (\mathbf{u} - \mathbf{g}) - \mathcal{C} && \text{for } \mathbf{x} \in \partial\Omega, \end{aligned}\right\} \tag{2.33}$$

where

$$\mathcal{C} = \frac{1}{S} \int_{\partial\Omega} \mathbf{n} \cdot (\mu \Delta \mathbf{u} + \lambda \mathbf{u})\ \mathrm{d}A, \tag{2.34}$$

and $S = \int_{\partial\Omega} \mathrm{d}A$ is the surface area of the boundary. In terms of (2.30–2.31) this corresponds to the projection

$$g_p = g \quad \text{and} \quad h_p = h - \mathcal{C}, \tag{2.35}$$

where

$$\mathcal{C} = \frac{1}{S} \left(\int_{\partial\Omega} h\, \mathrm{d}A - \int_{\Omega} g\, \mathrm{d}V \right). \tag{2.36}$$

This is clearly a projection, since $\mathcal{C} = 0$ for $(g_p,\, h_p)$, so that $\mathbb{P}^2 = \mathbb{P}$. Further, since the solvability condition for (2.30) is precisely $\mathcal{C} = 0$ — see equation (2.32), the solvability condition for (2.33) is satisfied — even if $\nabla \cdot \mathbf{u} \neq 0$, though (of course) $\mathcal{C} = 0$ when $\nabla \cdot \mathbf{u} = 0$.

Finally, we remark (again) that the projection in (2.35) is not unique. In particular, numerical implementations of the Poisson equation with Neumann conditions often use least squares projections, which alter both the boundary condition h and the source term g. This makes sense if the solvability errors are small. However, in general it seems desirable to not alter the source term, and keep (2.9–2.10) valid. This still does not make (2.35) unique, but it makes it the simplest projection. Others would also alter (in some appropriate eigenfunction representation) the zero mean components of h.

The system of equations in (2.20) and (2.33) makes sense for arbitrary flows $\mathbf{u}$, which are neither restricted by the incompressibility condition $\phi = \nabla \cdot \mathbf{u} = 0$ in Ω, nor the normal velocity boundary condition $\mathcal{E} = \mathbf{n} \cdot (\mathbf{u} - \mathbf{g}) = 0$. Furthermore: *this system, at least in bounded domains Ω, includes the Navier-Stokes equations as a global attractor for the smooth solutions.* This is easy to see as follows:

First, because of equations (2.9–2.10), ϕ decays exponentially, at a rate controlled by the smallest eigenvalue of $L = -\Delta$ in Ω, with Dirichlet boundary conditions. In particular:

$$\mathcal{C} = \frac{1}{S} \int_{\partial\Omega} \mathbf{n} \cdot (\mu \Delta \mathbf{u} + \lambda \mathbf{u}) \, \mathrm{d}A = \frac{1}{S} \int_{\Omega} (\mu \Delta \phi + \lambda \phi) \, \mathrm{d}V,$$

vanishes exponentially.

Second, it is easy to see that, for the system in (2.20) and (2.33), equation (2.22) is modified to

$$\mathcal{E}_t = -\lambda \mathcal{E} + \mathcal{C} \quad \text{for } \mathbf{x} \in \partial\Omega, \tag{2.37}$$

where $\mathcal{E} = \mathbf{n} \cdot (\mathbf{u} - \mathbf{g})$. Hence $\mathcal{E}$ also vanishes exponentially.

Of course, if $\phi = 0$ and $\mathcal{E} = 0$ initially, then they remain so for all times, and the evolution provided by (2.20) and (2.33) is, exactly, the Navier-Stokes evolution.

In conclusion, the formulation in this section is not only an interesting theoretical fact. It also provides a robust framework within which numerical solvers for the incompressible Navier Stokes equations can be developed, without having to worry about the (potentially deleterious) effects that discretization (or initial condition) errors, can cause when they violate mass conservation — because either $\nabla \cdot \mathbf{u} = 0$ in Ω, or $\mathbf{n} \cdot (\mathbf{u} - \mathbf{g}) = 0$ in $\partial\Omega$, fail. In addition, the formulation eliminates the necessity of having to enforce the condition $\nabla \cdot \mathbf{u} = 0$ directly, which is a core difficulty for the solution of the Navier-Stokes equations.

2.6 Stability of Semi-implicit Schemes

In the case of moderate to low Reynolds number flows, the stiff viscosity term $\mu \Delta \mathbf{u}$ requires very small time steps. Hence, there is a large practical interest in treating $\mu \Delta \mathbf{u}$ implicitly while keeping the associated pressure explicit. In this section we write down and analyze semi-implicit schemes for the PPE splitting (2.15–2.16). Inspired by many of the ideas from [58], we then show the subsequent schemes are unconditionally stable.

To analyze the stability of the semi-implicit schemes, we require several properties of the Hodge-Helmholtz decomposition. Given $\mathbf{w} \in L^2$, then $\mathbf{w}$ has a unique orthogonal decomposition as

$$\mathbf{w} = \mathbf{a} + \nabla b, \tag{2.38}$$

where b is determined by $\Delta b = \nabla \cdot \mathbf{w}$ with boundary conditions $\mathbf{n} \cdot \nabla b = \mathbf{n} \cdot \mathbf{w}$. Hence $\mathbf{a}$ is divergence free with zero normal boundary component. Moreover the component $\mathbf{a}$ is orthogonal in the L^2 norm to every gradient field (i.e., $\langle \mathbf{a}, \nabla \phi \rangle = \int_\Omega \mathbf{a} \cdot \nabla \phi \, \mathrm{d}V = 0$ is the standard inner product on $L^2(\Omega)$). We may therefore write $\nabla b = \mathcal{Q}\mathbf{w}$ and $\mathbf{a} = \mathcal{P}\mathbf{w}$ where $\mathcal{P}$ and $\mathcal{Q}$ are complementary orthogonal projections (ie. $\mathcal{P} + \mathcal{Q} = \mathcal{I}$, $\mathcal{P}^2 = \mathcal{P}$ and $\mathcal{P} = \mathcal{P}^\dagger$). As a result, the following identities hold vector fields $\mathbf{w}$ and $\mathbf{v}$

$$\begin{aligned} \langle \mathbf{w}, \mathcal{P}\mathbf{v} \rangle &= \langle \mathcal{P}\mathbf{w}, \mathcal{P}\mathbf{v} \rangle & (2.39) \\ &= \langle \mathcal{P}\mathbf{w}, \mathbf{v} \rangle, & (2.40) \end{aligned}$$

and

$$\langle \mathbf{w}, \mathbf{w} \rangle = \langle \mathcal{P}\mathbf{w}, \mathcal{P}\mathbf{w} \rangle + \langle \mathcal{Q}\mathbf{w}, \mathcal{Q}\mathbf{w} \rangle. \tag{2.41}$$

To motivate the stability proofs for semi-implicit time discretizations, note that the Stokes equation can also be written in an equivalent projection form

$$\left.\begin{aligned} \mathbf{u}_t &= \mu\mathcal{P}\Delta\mathbf{u} && \text{for } \mathbf{x}\in\ \Omega, \\ \mathbf{n}\times\mathbf{u} &= 0 && \text{for } \mathbf{x}\in\partial\Omega, \end{aligned}\right\} \tag{2.42}$$

Dotting equation (2.42) through by[3] $-\Delta\mathbf{u} = \nabla\times\nabla\times\mathbf{u} - \nabla(\nabla\cdot\mathbf{u})$ and integrating yields

$$\frac{1}{2}\frac{\partial}{\partial t}\Big(\langle\nabla\times\mathbf{u},\nabla\times\mathbf{u}\rangle + \langle\nabla\cdot\mathbf{u},\nabla\cdot\mathbf{u}\rangle\Big) \;=\; -\mu\langle\Delta\mathbf{u},\mathcal{P}\Delta\mathbf{u}\rangle. \tag{2.43}$$

Here the first line follows directly from integration by parts, noting that the boundary integrals vanish. The last equation can be written in a more compact form:

$$\frac{1}{2}\frac{\partial}{\partial t}\Big(||\nabla\times\mathbf{u}||^2 + ||\nabla\cdot\mathbf{u}||^2\Big) \;=\; -\mu||\mathcal{P}\Delta\mathbf{u}||^2. \tag{2.44}$$

From this it follows that the curl and divergence are bounded in the L^2 sense. As a first example of a semi-implicit scheme, we analyze the backward Euler discretization for the Stokes equation:

$$\left.\begin{aligned} \mathbf{u}^{n+1} - \mathbf{u}^n - \Delta t\,\mu\,\Delta\mathbf{u}^{n+1} &= -\Delta t\,\nabla p^n && \text{for } \mathbf{x}\in\ \Omega, \\ \mathbf{n}\times\mathbf{u}^{n+1} &= 0 && \text{for } \mathbf{x}\in\partial\Omega, \\ \nabla\cdot\mathbf{u}^{n+1} &= 0 && \text{for } \mathbf{x}\in\partial\Omega, \end{aligned}\right\} \tag{2.45}$$

and

$$\left.\begin{aligned} \Delta p^n &= 0 && \text{for } \mathbf{x}\in\ \Omega, \\ \mathbf{n}\cdot\nabla p^n &= \mathbf{n}\cdot\mu\,\Delta\mathbf{u}^n && \text{for } \mathbf{x}\in\partial\Omega. \end{aligned}\right\} \tag{2.46}$$

We note that the scheme (2.45)–(2.46) preserves the divergence free condition on the velocity exactly. Assume $\nabla\cdot\mathbf{u}^n = 0$ for $\mathbf{x}\in\Omega$. Taking the divergence of (2.45), and setting

---

[3] To obtain the energy bound (2.44) by a straightforward calculation, we have kept the $\nabla\cdot\mathbf{u}$ terms throughout the computation even though they are in fact zero.

$\phi(\mathbf{x}) = \nabla \cdot \mathbf{u}^{n+1}$ yields

$$\left.\begin{array}{rcll} -\Delta t\, \mu\, \Delta\phi &=& -\phi & \text{for } \mathbf{x} \in \Omega, \\ \phi &=& 0 & \text{for } \mathbf{x} \in \partial\Omega, \end{array}\right\} \tag{2.47}$$

Since $-\Delta$ does not have a negative eigenvalue, $\phi = 0$ for all $\mathbf{x} \in \Omega$ is the only solution to (2.47). Consequently, the pressure equation for p^{n+1} automatically satisfies the Neumann consistency condition and is therefore well defined. As a direct consequence of preserving the divergence condition, the pressure as defined by (2.46) is equivalent to the projection:

$$\nabla p^n = \mathcal{Q}\Delta \mathbf{u}^n. \tag{2.48}$$

Remark 2.8 *Since the scheme (2.45)–(2.46) preserves the divergence constraint, there is never the need for a projection step (ie. one may associate $\mathbf{u}^{n+1} = \mathbf{u}^{*n+1}$, where $\mathbf{u}^{*n+1}$ is the conventional intermediate velocity field with non-zero divergence).* ♠

A stability proof for equations (2.45)–(2.46) closely follows the one for periodic channel flow in [58]. Specifically, dot both sides of equation (2.45) by $-\Delta(\mathbf{u}^{n+1} + \mathbf{u}^n)$, and integrate (this is the discrete analog of the steps required to derive equation (2.44)). The first two terms on the left hand side of equation (2.45) become

$$\langle \mathbf{u}^{n+1} - \mathbf{u}^n, -\Delta(\mathbf{u}^{n+1} + \mathbf{u}^n)\rangle = ||\nabla \times \mathbf{u}^{n+1}||^2 + ||\nabla \cdot \mathbf{u}^{n+1}||^2 \tag{2.49}$$

$$- \; ||\nabla \times \mathbf{u}^n||^2 - ||\nabla \cdot \mathbf{u}^n||^2. \tag{2.50}$$

In addition, we have

$$\langle \Delta\mathbf{u}^{n+1}, \Delta(\mathbf{u}^{n+1} + \mathbf{u}^n)\rangle = \frac{1}{2}||\Delta\mathbf{u}^{n+1} + \Delta\mathbf{u}^n||^2 + \frac{1}{2}||\Delta\mathbf{u}^{n+1}||^2 - \frac{1}{2}||\Delta\mathbf{u}^n||^2, \tag{2.51}$$

and

$$\langle \mathcal{Q}\Delta\mathbf{u}^n, \Delta(\mathbf{u}^{n+1} + \mathbf{u}^n)\rangle = \frac{1}{2}||\mathcal{Q}\Delta\mathbf{u}^{n+1} + \mathcal{Q}\Delta\mathbf{u}^n||^2 + \frac{1}{2}||\mathcal{Q}\Delta\mathbf{u}^n||^2 - \frac{1}{2}||\mathcal{Q}\Delta\mathbf{u}^{n+1}||^2. \tag{2.52}$$

For brevity, we also introduce the energy $\mathcal{E}$ where

$$\mathcal{E}[\mathbf{u}] = ||\nabla \times \mathbf{u}||^2 + ||\nabla \cdot \mathbf{u}||^2 + \mu\Delta t||\Delta \mathbf{u}||^2 + \mu\Delta t||\mathcal{Q}\Delta \mathbf{u}||^2. \tag{2.53}$$

Finally, combining everything, we have

$$\begin{aligned} \mathcal{E}[\mathbf{u}^{n+1}] - \mathcal{E}[\mathbf{u}^n] &= -\frac{\mu\Delta t}{2}||\Delta \mathbf{u}^{n+1} + \Delta \mathbf{u}^n||^2 + \frac{\mu\Delta t}{2}||\mathcal{Q}\Delta \mathbf{u}^{n+1} + \mathcal{Q}\Delta \mathbf{u}^n||^2 & (2.54) \\ &= -\frac{\mu\Delta t}{2}||\mathcal{P}\Delta \mathbf{u}^{n+1} + \mathcal{P}\Delta \mathbf{u}^n||^2, & (2.55) \end{aligned}$$

so that

$$\mathcal{E}[\mathbf{u}^{n+1}] \leq \mathcal{E}[\mathbf{u}^n]. \tag{2.56}$$

For simply connected domains, the energy bound (2.56) implies the scheme (2.45)–(2.46) is unconditionally stable. Unfortunately the bound (2.56) is not sufficient to prove stability in periodic domains, or domains with holes. The reason here is that $\mathcal{E}[\mathbf{h}] = 0$ for any vector field $\mathbf{h}$ that has $\nabla \cdot \mathbf{h} = \nabla \times \mathbf{h} = 0$ in Ω, and $\nabla \cdot \mathbf{h} = \mathbf{n} \times \mathbf{h} = 0$ in $\partial\Omega$. In simply connected domains, the only such vector fields are $\mathbf{h} = 0$. On the other hand, in periodic domains or domains with holes, there exist vector fields $\mathbf{h} \neq 0$, which satisfy $\nabla \cdot \mathbf{h} = 0$ and $\mathbf{n} \times \mathbf{h} = 0$ on $\partial\Omega$ and $\mathcal{E}[\mathbf{h}] = 0$. Here the energy bound (2.56) does not control the growth of these modes. We note however that such vector fields $\mathbf{h}$ have non zero flux $\mathbf{h} \cdot \mathbf{n} \neq 0$ at the boundary. Hence, we expect that the modified scheme (2.20)–(2.20) with addition of the term $\lambda \mathbf{n} \cdot \mathbf{u}$ in the boundary condition for the pressure will stabilize the growth of such modes.

2.6.1 Stability of a second order scheme

Following a procedure analogous to the one in [58], we discuss the stability of a second order Crank-Nicholson scheme where we treat the pressure with a second order Adams-Bashforth extrapolation:

$$\left.\begin{aligned} \mathbf{u}^{n+1} - \mathbf{u}^{n} &= \tfrac{\mu}{2}\Delta t\left(\Delta \mathbf{u}^{n+1} + \Delta \mathbf{u}^{n}\right) - \tfrac{3}{2}\Delta t \nabla p^{n} + \tfrac{1}{2}\Delta t \nabla p^{n-1} && \text{for } \mathbf{x} \in \Omega, \\ \mathbf{n} \times \mathbf{u}^{n+1} &= 0 && \text{for } \mathbf{x} \in \partial\Omega, \\ \nabla \cdot \mathbf{u}^{n+1} &= 0 && \text{for } \mathbf{x} \in \partial\Omega, \end{aligned}\right\} \quad (2.57)$$

where p^n is given by (2.46). Proceeding with a normal mode analysis[4], we set $\mathbf{u}^n = \alpha^n \tilde{\mathbf{u}}$ where $\tilde{\mathbf{u}}$ satisfies the boundary conditions in (2.57), and α is an eigenvalue of the time stepping operator. Upon substitution we obtain

$$(\alpha^2 - \alpha)\tilde{\mathbf{u}} = \frac{\mu}{2}\Delta t(\alpha^2 + \alpha)\Delta\tilde{\mathbf{u}} - \frac{3}{2}\Delta t \alpha \mathcal{Q}\Delta\tilde{\mathbf{u}} + \frac{1}{2}\Delta t \mathcal{Q}\Delta\tilde{\mathbf{u}}. \quad (2.58)$$

In analogy with the steps in the previous section, we dot equation (2.58) through by $-\Delta\tilde{\mathbf{u}}$ and integrating by parts to obtain

$$2(\alpha^2 - \alpha)\big(||\nabla \times \tilde{\mathbf{u}}||^2 + ||\nabla \cdot \tilde{\mathbf{u}}||^2\big) + \mu\Delta t(\alpha^2 + \alpha)||\Delta\tilde{\mathbf{u}}||^2 = \mu\Delta t(3\alpha - 1)||\mathcal{Q}\Delta\tilde{\mathbf{u}}||^2. \quad (2.59)$$

Equation (2.59) is now a quadratic of the form $a\alpha^2 - b\alpha + c = 0$ for the eigenvalues α where

$$a = 2||\nabla \times \tilde{\mathbf{u}}||^2 + 2||\nabla \cdot \tilde{\mathbf{u}}||^2 + \mu\Delta t||\Delta\tilde{\mathbf{u}}||^2, \quad (2.60)$$

$$b = 2||\nabla \times \tilde{\mathbf{u}}||^2 + 2||\nabla \cdot \tilde{\mathbf{u}}||^2 - \mu\Delta t||\Delta\tilde{\mathbf{u}}||^2 + 3\mu\Delta t||\mathcal{Q}\Delta\tilde{\mathbf{u}}||^2, \quad (2.61)$$

$$c = \mu\Delta t||\mathcal{Q}\Delta\tilde{\mathbf{u}}||^2. \quad (2.62)$$

Since the coefficients a, b, c satisfy the following inequalities $0 \leq c < a$ and $|b| < a + c$, it follows [96] that the eigenvalues α lie within the unit circle. The preceding argument indicates that the normal modes $\tilde{\mathbf{u}}$ remain bounded for the second order Crank-Nicholson scheme (2.57). Hence the scheme is stable for domains with simple geometries[5].

---

[4] We will not discussed here the questions of whether the resulting eigenvectors forms a complete basis.

[5] Note that in domains which are not simply connected, and have nontrivial homology groups, there is a family of harmonic modes which satisfy $\Delta\tilde{\mathbf{u}} = 0$. For these modes one can check that $a = b = c = 0$ and the argument fails to show that $|\alpha| < 1$

2.6.2 Domains with holes and the role of feedback

In domains with nontrivial topologies, controlling the velocity gradients in the scheme (2.45)–(2.46) is not enough to ensure stability. For example, in simple domains, the Laplacian on vector fields (with $\nabla \cdot \mathbf{u} = 0$ and $\mathbf{n} \times \mathbf{u} = 0$ boundary conditions) is self-adjoint and positive definite [98]:

$$\begin{aligned} ||\nabla \times \mathbf{u}||^2 + ||\nabla \cdot \mathbf{u}||^2 &= \langle -\Delta \mathbf{u}, \mathbf{u} \rangle && (2.63) \\ &\geq \alpha_{min} ||\mathbf{u}||^2, && (2.64) \end{aligned}$$

where $\alpha_{min} > 0$ is the minimum eigenvalue. Hence, in this case $||\nabla \times \mathbf{u}||^2 + ||\nabla \cdot \mathbf{u}||^2 \rightarrow 0$ implies $\frac{\partial}{\partial t}||\mathbf{u}|| \rightarrow 0$ as well. In domains with nontrivial homologies, for instance periodic domains or domains with holes, the Laplacian has nonzero eigenvalues. Hence, controlling $||\nabla \times \mathbf{u}||^2 + ||\nabla \cdot \mathbf{u}||^2$ does not necessarily control $||\mathbf{u}||$. In fact, in cases such as periodic flow, the velocity field may have a component corresponding to a zero eigenvalue of the Laplacian which grows in time. As a result, we now show that at the continuum PDE level, the addition of the λ feedback term indeed controls the growth of such modes. Written in a projection form, the extended Navier-Stokes system with the feedback term has the form

$$\left.\begin{aligned} \mathbf{u}_t &= \mu \mathcal{P} \Delta \mathbf{u} - \lambda \mathcal{Q} \mathbf{u} && \text{for } \mathbf{x} \in \Omega, \\ \mathbf{n} \times \mathbf{u} &= 0 && \text{for } \mathbf{x} \in \partial\Omega, \end{aligned}\right\} \qquad (2.65)$$

where the pressure now contains two projection components. Dotting equation (2.65) by $-\Delta \mathbf{u}$ and integrating yields

$$\frac{1}{2}\frac{\partial}{\partial t}\Big(||\nabla \times \mathbf{u}||^2 + ||\nabla \cdot \mathbf{u}||^2 \Big) = -\mu ||\mathcal{P} \Delta \mathbf{u}||^2 + \lambda \langle \Delta \mathbf{u}, \mathcal{Q} \mathbf{u} \rangle. \qquad (2.66)$$

In addition, dotting (2.65) by $\mathbf{u}$ and integrating yields

$$\frac{1}{2}\frac{\partial}{\partial t}||\mathbf{u}||^2 = \mu \langle \mathbf{u}, \Delta \mathbf{u} \rangle - \mu \langle \mathbf{u}, \mathcal{Q} \Delta \mathbf{u} \rangle - \lambda ||\mathcal{Q} \mathbf{u}||^2. \qquad (2.67)$$

Multiplying (2.66) by μ, (2.67) by λ and adding the two equations we obtain

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial t}\Big(\lambda||\mathbf{u}||^2 + \mu||\nabla\times\mathbf{u}||^2 + \mu||\nabla\cdot\mathbf{u}||^2\Big) &= -\mu^2||\mathcal{P}\Delta\mathbf{u}||^2 - \lambda\mu||\nabla\times\mathbf{u}||^2 && (2.68)\\
&- \lambda\mu||\nabla\cdot\mathbf{u}||^2 - \lambda^2||\mathcal{Q}\mathbf{u}||^2 && (2.69)\\
&\leq 0. && (2.70)
\end{aligned}$$

Hence the addition of the λ feedback term explicitly controls an appropriate (H^1) velocity norm.

2.7 Numerical Scheme

In this section we outline an efficient numerical scheme for solving the coupled differential equations (2.20–2.21) on a two-dimensional irregular domain. As should be clear from the prior sections, the main issue we aim to address is that of how to effectively implement the incompressibility condition and the boundary conditions for the pressure, avoiding the difficulties that projection and fractional step methods have. These are problems that are not related to the nonlinearities in the Navier-Stokes equations, and occur even for the linearized equations. Hence, in the spirit of focusing on the key issues only — also see remark 2.9, the calculations presented in § 2.8 are for the linear equations. Furthermore, while in the description of the scheme in this section we carry through the nonlinear terms, we do not describe their implementation. We note that there is extensive literature on this topic, and one may refer to the book by J. Ferziger and M. Peric [28] for more details.

To achieve an efficient scheme, we decouple the pressure and velocity fields, and explicitly treat each term in the time evolution of (2.20–2.21). Specifically, since both the right hand side and boundary conditions of equation (2.21) depend solely on $\mathbf{u}$, we may view the pressure as a computable functional of the velocity, $p = p[\mathbf{u}]$. The computation of $p[\mathbf{u}]$ requires the solution of a Poisson equation with Neumann boundary conditions. With this in mind, the momentum equation then has the form $\mathbf{u}_t = \mathbf{F}[\mathbf{u}]$, where $\mathbf{F}[\mathbf{u}]$ has a complicated, yet numerically computable form:

$$\mathbf{F}[\mathbf{u}] = \mu\,\Delta\mathbf{u} - \nabla p[\mathbf{u}] - (\mathbf{u}\cdot\nabla)\,\mathbf{u} + \mathbf{f} \quad \text{for } \mathbf{x}\in\Omega. \tag{2.71}$$

We now use an explicit forward Euler scheme to discretize the time evolution for (2.20–2.21), paired with an appropriate discretization in space described later in this section. This yields the scheme

$$\frac{1}{\Delta t}\left(\mathbf{u}^{n+1} - \mathbf{u}^{n}\right) = \mu\,\Delta\mathbf{u}^{n} - \nabla p^{n} - (\mathbf{u}^{n}\cdot\nabla)\,\mathbf{u}^{n} + \mathbf{f}^{n} \quad \text{for } \mathbf{x}\in\Omega, \tag{2.72}$$

with boundary conditions $\mathbf{n}\times\mathbf{u} = \mathbf{n}\times\mathbf{g}$ and $\nabla\cdot\mathbf{u} = 0$ for $\mathbf{x}\in\partial\Omega$, where the pressure is given by

$$\Delta p^{n} = -\nabla\cdot\left((\mathbf{u}^{n}\cdot\nabla)\,\mathbf{u}^{n}\right) + \nabla\cdot\mathbf{f}^{n} \quad \text{for } \mathbf{x}\in\Omega, \tag{2.73}$$

with the boundary condition

$$\mathbf{n}\cdot\nabla p^{n} = \mathbf{n}\cdot\left(\mathbf{f}^{n} - \mathbf{g}_t^{n} + \mu\,\Delta\mathbf{u}^{n} - (\mathbf{u}^{n}\cdot\nabla)\,\mathbf{u}^{n}\right) + \lambda\,\mathbf{n}\cdot(\mathbf{u}^{n} - \mathbf{g}^{n}),$$

for $\mathbf{x}\in\partial\Omega$. Here, starting with the initial data $\mathbf{u}^0$, a superscript n is used to denote a variable at time $t = n\,\Delta t$, where $0 < \Delta t \ll 1$ is the time step.

Remark 2.9 *Our purpose here is to illustrate the new approach with a simple scheme that does not obscure the ideas in the method with technical complications. Hence, the scheme here is first order in time (explicit) and second order in space, with the stability restriction* $\Delta t \propto (\Delta x)^2$. *However, unlike projection methods and other approaches commonly used to solve the Navier-Stokes equations, this new formulation does not seem to have any inherent order limitations. Unfortunately, the Navier-Stokes equations are stiff and nonlinear, which means that the fact that higher order extensions are possible does not mean that they are trivial.* ♠

2.7.1 Space grid and discretization

To discretize the equations in space, we use finite differences over a cartesian, square ($\Delta x = \Delta y$), staggered grid. The pressure values are stored at the nodes of the grid, while the horizontal and vertical components of the velocity are stored at the mid-points of the edges connecting the grid nodes (horizontal component on the horizontal edges and vertical component on the vertical edges).

When handling an arbitrary curved boundary, we cannot conform the boundary to the grid, but rather we immerse it within the regular mesh — see figure 2-1. Then, to numerically describe the domain boundary, we identify a set $\mathcal{C}_b$ of N_e points in $\partial\Omega$, say $\mathbf{x}_{bj} = (x_b, y_b)_j$ for $1 \le j \le N_e$ — see item 2.7c below. These N_e points are located at $O(\Delta x)$ distances apart, so that the resolution of the boundary is comparable with that of the numerical grid.

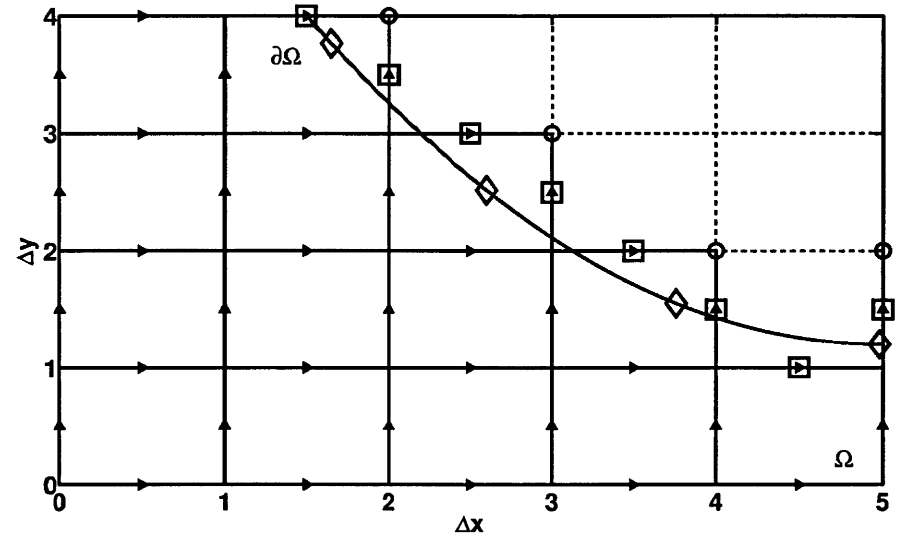

Figure 2-1: This plot shows the staggered grid, and the boundary. The numerical pressure values correspond to the graph nodes, while the velocities (arrows) correspond to the edge midpoints. Here the circles (∘) and squares (□) denote ghost pressure points, and boundary velocities, respectively. These are used to implement the boundary conditions in the Poisson and momentum equations, respectively. The diamonds (◇) denote the points $(x_b, y_b)_i$, used to represent the boundary $\partial\Omega$.

For the heat equation $T_t = \mu\,\Delta T$, the stability restriction for the standard scheme using a 5 point centered differences approximation for the Laplacian, and forward Euler in time, is

$$\Delta t \leq C \frac{(\Delta x)^2}{\mu}, \tag{2.74}$$

where $C = \frac{1}{2D}$ and $D = 1, 2, \ldots$ is the space dimension. Since the method described here uses exactly the same approach to advance the velocity flow field $\mathbf{u}$, we expect the same restriction (with, perhaps, a different constant C) to apply. For the 2D numerical calculations presented in § 2.8, we found that the algorithm was stable with $C \leq 0.2$, while $C \geq 0.3$ generally produced unstable behavior. Below we separately address the numerical implementation of equations (2.72) and (2.73).

2.7.2 Poisson equation

To solve the pressure Poisson equation (2.73) with Neumann boundary conditions, we use the ghost point idea with a method by Greenspan [45] to implement the boundary conditions. As discussed above, we embed the domain Ω within a cartesian square grid, and classify the computational points in the grid into inner and ghost points (see next paragraph). Then: (i) for the inner points we discretize the Laplace operator using the standard 5 point centered differences stencil. (ii) For the ghost points we obtain equations from an appropriate discretization of the Neumann boundary condition. In particular, if there are N_a inner points, and N_b ghost points, then the discretized pressure is represented by a vector $\widehat{p} \in \mathbb{R}^N$ where $N = N_a + N_b$.

The computational domain, $\mathcal{C}_p$, for the pressure is comprised by the *inner* points and the *ghost* points, defined as follows:

2.7a The *inner* points are the points in the cartesian grid located inside Ω.

2.7b The *ghost* points are the points in the cartesian grid that are either outside Ω or on $\partial\Omega$, and which are needed to complete the 5 point stencil for some inner point.

2.7c Construct the set $\mathcal{C}_b$ used to track the boundary $\partial\Omega$ as follows: For every pressure ghost point in the grid: $\mathbf{x}_j^{pg}$, $1 \leq j \leq N_b$, select the closest point in the boundary $\mathbf{x}_{bj} \in \partial\Omega$.

We will use these points to produce equations that approximate the Neumann boundary conditions (one equation per point).

Remark 2.10 *With the set $\mathcal{C}_b$ as above, $N_e = N_b$. This is not the only possibility. In the end, the discretized equations for the pressure have to be solved in the least squares sense (see remark 2.11). Thus, we could have elected to over-determine the implementation of the boundary condition by selecting $N_e > N_b$ points in $\partial\Omega$. One reason for doing this being to obtain a "smoother" implementation, avoiding the irregularities that occur as the boundary placement (relative to the cartesian grid) changes along $\partial\Omega$. In our calculations we found that this is not a must for the solution of the Poisson equation. On the other hand, for the momentum equation this strategy is needed to avoid numerical instabilities — see § 2.7.3.*

♠

It follows that, on the computational domain $\mathcal{C}_p$, the Poisson equation (including the boundary conditions) is discretized by N_a equations inside Ω, and N_b equations derived from the boundary conditions. Specifically

2.7d We use the standard 5 point centered differences stencil for the Laplacian, to obtain one equation for each of the N_a inner points. Similarly, we use a second order approximation for the forcing terms on the right hand side of the equation.

2.7e We construct one boundary condition equation for each pressure ghost point $\mathbf{x}_j^{pg}$, $1 \leq j \leq N_b$. This is done by using 6 nearby points from the computational domain to obtain a second order approximation to the normal derivative at the corresponding point $\mathbf{x}_{bj} \in \partial\Omega$ — see item 2.7c above. We use a similar approach to approximate the terms on the right hand side of the Neumann boundary conditions.

Hence, each boundary condition equation involves the ghost point, points inside the domain, and (possibly) other nearby ghost points.

Thus the Poisson equation can be written in a matrix form with the following structure

$$\begin{pmatrix} L \\ B \end{pmatrix} \widehat{p} = \begin{pmatrix} a \\ b \end{pmatrix} . \tag{2.75}$$

Here $B \in \mathbb{R}^{N_b \times N}$ and $L \in \mathbb{R}^{N_a \times N}$ are the discrete matrix representations for the derivatives $(\mathbf{n} \cdot \nabla)$ in the Neumann boundary condition,, and Laplacian (Δ), respectively. The terms $a \in \mathbb{R}^{N_a}$ and $b \in \mathbb{R}^{N_b}$ are the discrete representations of the source terms and applied boundary conditions for the Poisson equation:

$$\begin{aligned} a &\approx \nabla \cdot (\mathbf{f} - (\mathbf{u} \cdot \nabla)\mathbf{u}) , && (2.76) \\ b &\approx \mathbf{n} \cdot (\mathbf{f} - \mathbf{g}_t + \mu \Delta \mathbf{u} - (\mathbf{u} \cdot \nabla)\mathbf{u}) + \\ & \quad \lambda \mathbf{n} \cdot (\mathbf{u} - \mathbf{g}). && (2.77) \end{aligned}$$

Thus the pair of equations

$$L\widehat{p} = a \quad \text{and} \quad B\widehat{p} = b \tag{2.78}$$

are the discrete analog of equation (2.21).

Remark 2.11 *As pointed out earlier in remark 2.7, the discretization errors cause solvability for equation (2.75) to fail. These solvability errors are "small" (second order). Hence, as it is commonly the case with numerical solutions of the Poisson equation with Neumann boundary conditions, we solve equation (2.75) in the least squares sense. This provides an approximation to the pressure which is within the discretization error.* ♠

2.7.3 Momentum equation

The main numerical difficulty with implementing (2.72) is produced by the boundary conditions. Specifically: on a cartesian staggered grid the boundary conditions $\nabla \cdot \mathbf{u} = 0$ and $\mathbf{n} \times (\mathbf{u} - \mathbf{g}) = 0$ couple the "horizontal", u, and "vertical", v, components of the flow velocity — with the exception of the special case of a boundary aligned with the grid, where there is no coupling. Hence, in general, the implementation of the boundary conditions requires the solution (at every time step) of a linear system of equations that couples all the boundary velocities (these are defined below).

The *computational domain for the velocities,* $\mathcal{C}_{\mathbf{u}}$, is defined in terms of the edges in the cartesian grid with which the velocities are associated (see figure 2-1). To define $\mathcal{C}_{\mathbf{u}}$, it is convenient to first introduce the *extended set of pressure nodes*, $\mathcal{E}_p$, from which $\mathcal{C}_{\mathbf{u}}$ is easily constructed:

2.7f A pressure node in the cartesian grid belongs to $\mathcal{E}_p$ if and only if it either belongs to $\mathcal{C}_p$, or if it is connected (by an edge in the grid) to a ghost pressure node that lies on $\partial\Omega$.

2.7g A velocity is in $\mathcal{C}_{\mathbf{u}}$ if and only if: (a) Its corresponding edge connects two points in $\mathcal{E}_p$. (b) At least one of the two points is in Ω (or $\partial\Omega$).

Remark 2.12 *Notice that $\mathcal{E}_p$ is exactly what $\mathcal{C}_p$ becomes if $\partial\Omega$ is modified by an "infinitesimal" perturbation that turns all the ghost pressure points on $\partial\Omega$ into inner pressure points.*

The computational domain $\mathcal{C}_{\mathbf{u}}$ is, in turn, sub-divided into inner and boundary velocity edges

2.7h An edge in $\mathcal{C}_{\mathbf{u}}$ is an *inner* velocity if and only if $\mathcal{C}_{\mathbf{u}}$ includes the four other edges needed to compute the Laplacian (either Δu or Δv) at the edge mid-point, using the 5 point centered differences approximation.

2.7i An edge in $\mathcal{C}_{\mathbf{u}}$ is a *boundary* velocity if and only if it is not an inner velocity. Let M *be the number of boundary velocities.*

The solution of the momentum equation (2.20) is thus performed as follows:

2.7j At the start of each time step the right hand side in (2.72) is approximated (to second order, using centered differences) at the inner velocity locations, which can then be updated to their values at the next time.

2.7k Next, using the boundary conditions, the values of the boundary velocities at the next time step are constructed from the inner velocities. This "extension" process is explained below.

Let $\mathbf{y} \in \mathbb{R}^M$ be the vector of boundary velocities. Then $\mathbf{y}$ is determined by two sets of equations, corresponding to the discretization on $\partial\Omega$ of $\nabla \cdot \mathbf{u} = 0$ and $\mathbf{n} \times (\mathbf{u} - \mathbf{g}) = 0$. In our approach the divergence free criteria is enforced "point-wise", while the condition on the tangential velocity is imposed in a least squares sense.

Implementation of the divergence free $\nabla \cdot \mathbf{u} = 0$ *boundary condition.* At first sight, this boundary condition appears to be the hardest to implement, since it is a Neumann condition (essentially) prescribing the value of the normal derivative of the normal component of the velocity, in terms of the tangential derivative of the tangential velocity. However, because (for the exact solution) $\nabla \cdot \mathbf{u} = 0$ everywhere, an implementation of this condition which is second order consistent (in the classical sense of finite differences introduced by Lax [68]) is easy to obtain, as follows:

First: identify the M_d pressure nodes in $\mathcal{C}_p$, which have at least one boundary velocity as an adjacent edge. These M_d pressure nodes lie either on $\partial\Omega$, or inside Ω, and are all within a distance $O(\Delta x)$ of $\partial\Omega$ — definitely no further away than $\sqrt{2}\,\Delta x$. Second: for each of these M_d points, use centered differences to approximate the flow divergence at the point, and set $\nabla \cdot \mathbf{u} = 0$. This provides M_d equations that couple the (unknown) boundary velocities to the (known) inner velocities. In matrix form this can be written as

$$D\mathbf{y} = \mathbf{s} \tag{2.79}$$

where D is the portion of the discrete divergence operator acting on the unknown boundary velocities, while $\mathbf{s}$ is the associated flux derived from the known inner velocities. D is a rectangular, very sparse, matrix whose entries are 0 and ±1 — note that Δx can be eliminated from these equations.

Remark 2.13 *Notice that* $\nabla \cdot \mathbf{u} \equiv 0$ *for the exact solution. Hence, setting* $\nabla \cdot \mathbf{u} = 0$ *at* *the nodes near the boundary (as done above) involves no approximation. The second order nature of (2.79) is caused by the error in computing* $\nabla \cdot \mathbf{u}$ *— there is no "extrapolation" error.*

♠

Implementation of the tangential velocity $\mathbf{n} \times (\mathbf{u} - \mathbf{g}) = 0$ *boundary condition.* It is easy to see that $M_d \approx M/2$, since most of the M_d pressure nodes in the selected set will connect with two boundary boundary velocities. Thus (2.79) above provides approximately one half the number of equations needed to recover $\mathbf{y}$ from the inner velocities. It would thus seem natural to seek for $M - M_d$ additional equations using the other boundary condition. Namely, find $M - M_d$ points on $\partial\Omega$, and at each one of them write an approximation to $\mathbf{n} \times (\mathbf{u} - \mathbf{g}) = 0$ using nearby points in $\mathcal{C}_{\mathbf{u}}$. Unfortunately, this does not work. It is very hard to do the needed approximations in a fashion that is robust relative to the way Ω is embedded in the rectangular grid. Our attempts at this simple approach almost always lead to situations where somewhere along $\partial\Omega$ an instability was triggered.

To avoid the problem stated in the previous paragraph, we over-determine the implementation of the tangential velocity boundary condition, and solve the resulting system in the least squares sense. The boundary condition is replaced by the minimization of a (discrete version) of a functional of the form

$$\int_{\partial\Omega} |\mathbf{n} \times (\mathbf{u} - \mathbf{g})|^2 \, w \, \mathrm{d}A, \tag{2.80}$$

where w is some (strictly positive) weight function. This approach yields a robust, numerically stable, approximation — fairly insensitive to the particular details of how Ω is embedded within the cartesian grid.

In fact, we do not implement the minimization of a functional as in (2.80), but a simpler process which is (essentially) equivalent. To be specific, we start with the set $\mathcal{C}_b$ of points in $\partial\Omega$ — see item 2.7c. For each $\mathbf{x}_{bj} \in \mathcal{C}_b$, where $1 \le j \le N_e$, we identify a local horizontal, $\mathcal{P}_j^u$, and vertical, $\mathcal{P}_j^v$, velocity "patch". These patches — see figure 2-2 — have the following properties

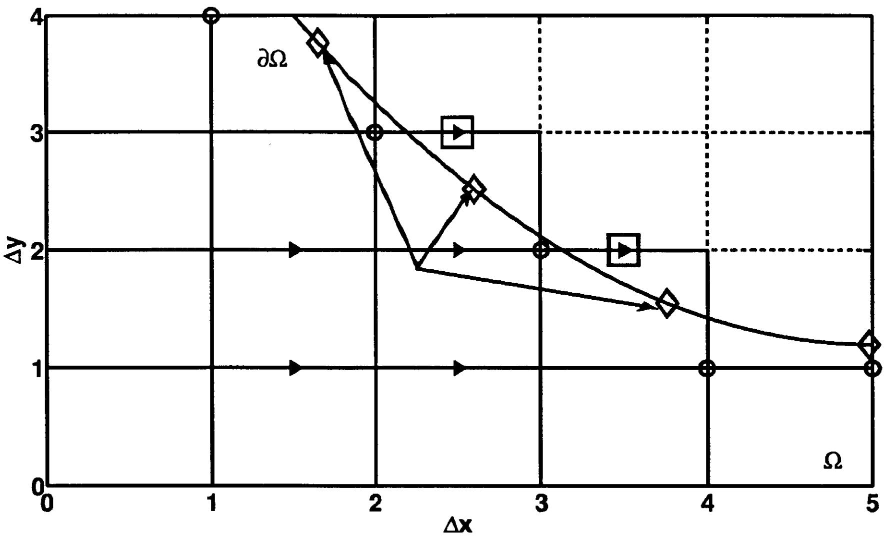

Figure 2-2: This plot illustrates the implementation of the momentum equation boundary conditions. The circles (o) indicate the points at which the velocity divergence is set to zero. The six arrows indicate the horizontal velocity components in the patch, $\mathcal{P}^{u}_{j\nu}$, used to extrapolate u to the three boundary points indicated by the diamonds (◇). The squares (□) denote the boundary velocities — which are part of the patch $\mathcal{P}^{u}_{j\nu}$.

2.7l Each patch contains both inner and boundary velocities.

2.7m Each patch contains 6 velocities, in an appropriate structure, so that these velocities can be used to extrapolate (with second order accuracy) values of the corresponding velocity (u or v) to nearby points along the boundary. For example: the 5 point stencil used to approximate the Laplacian, plus one of the four next closest points to the center point.

2.7n The union of all the patches contains all the boundary velocities.

These patches are used to (linearly) interpolate/extrapolate the velocities to nearby points along the boundary. Each patch is used to extrapolate the velocity to the three "closest" boundary points to the patch. In this fashion, for every $\mathbf{x}_{bj}$, $1 \leq j \leq N_e$, three different approximations to the tangential velocity at the boundary follow. These involve different (linear) combinations of velocities at the nearby edges in $\mathcal{C}_{\mathbf{u}}$, including both boundary and inner velocities. In this fashion, a set of $3\,N_e$ additional linear equations — beyond those in equation (2.79) — for the boundary velocities follow.

Remark 2.14 *The idea in the process above is to "link" the patches used to extrapolate the velocity to the boundary. This is so that no "gaps" occur in the implementation of the tangential boundary conditions, as the positioning of* $\partial\Omega$ *relative to the grid changes. We elected to use minimal coupling (each patch linked to its two neighbors). In principle one could increase the amount of over-determination of the tangential boundary condition. This would require using a set of points in* $\partial\Omega$ *that is "denser" than* $\mathcal{C}_b$*, but* in our calculations we found that this was not needed.

Finally, an interesting question is: why is a process similar to this one not needed for the implementation of the boundary conditions for the Poisson equation in § 2.7.2? The obvious answer is that the Poisson equation itself couples everything, so that no extra coupling needs to be added. ♠

Putting everything together, an overdetermined system of equations for the boundary velocity vector $\mathbf{y}$ follows, which can be written in the form

$$\begin{aligned} D\mathbf{y} &= \mathbf{s}, && (2.81) \\ E\mathbf{y} &= \mathbf{t}. && (2.82) \end{aligned}$$

Here $E \in \mathbb{R}^{3\,N_e \times M}$, and the second equation is to be solved in the least squares sense, subject to the constraint imposed by the first equation. One way to do this is as follows: write

$$\mathbf{y} = \mathbf{y}_p + P\mathbf{c} \tag{2.83}$$

where P is a matrix whose columns are a basis for the kernel of D — hence $D\,P = 0$, $\mathbf{y}_p$ is a particular solution of (2.81), and $\mathbf{c}$ is a constant vector parameterizing the space of (numerical) divergence free boundary velocities. Substituting the ansatz (2.83) into (2.82) yields

$$(E\,P)\,\mathbf{c} = \mathbf{t} - E\,\mathbf{y}_p, \tag{2.84}$$

which is a constraint-free least squares problem for the vector $\mathbf{c}$. Thus we can write

$$\mathbf{c} = (E\,P)^{+}\,(\mathbf{t} - E\mathbf{y}_p)\,, \tag{2.85}$$

which, together with (2.83), gives the solution $\mathbf{y}$.

Remark 2.15 *In (2.85), $(E\,P)^{+}$ is the pseudo-inverse, defined by the singular value decomposition of the matrix $F = E\,P$. In fact, since $F^T\,F$ should be invertible (see below), $F^{+} = (F^T\,F)^{-1}\,F^T$.*

One of the objectives of the above construction is to ensure that the boundary velocities are completely determined by the boundary conditions (and the inner velocities). If the problem in (2.82) for $\mathbf{y}$, with the constraint in (2.81), were to have more than one solution that minimizes the L^2 norm of $\mathbf{t} - E\,\mathbf{y}$, then this would be a sure sign that the boundary condition implementation is flawed, and there is missing information (i.e.: more equations are needed, see remark 2.14).

Of course, the requirement in the prior paragraph is equivalent to the statement that $F^T\,F$ is invertible. Finally, notice that in a domain with a fixed boundary, one may preprocess the matrices E, D and P. ♠

Remark 2.16 *It should be clear that the scheme developed in this section, is second order consistent[6] (all the way up to the boundary) in the classical sense of finite differences introduced by Lax [68].* ♠

2.7.4 Comparison with the projection method

The scheme proposed here — which involves the implementation of the equations in (2.72–2.73), appears to be quite similar to fractional step methods, such as Chorin's [20] original projection method. Specifically: advancing one time step requires both the evolution of a diffusion equation for the flow velocity, and the inversion of a Poisson equation for the pressure — which is the same as in projection methods. But there are important differences,

---

[6] For the system of PDE in (2.20–2.21).

mainly related to the implementation of the boundary conditions, and of the incompressibility condition. Below we highlight the similarities and differences between the pressure Poisson approach proposed here and, for simplicity, the projection method in its original formulation [20]. A (fairly recent) thorough review of projection methods, exploring the improvements to the approach since [20], as well as their drawbacks, can be found in [47].

First, the starting point for the method here is the discretization of a reformulation of the equations: (2.20–2.21), not the "original" Navier-Stokes equations (2.1–2.2). In this equivalent set: (i) there is a natural way to recover the pressure from the flow field at any given time. (ii) The time evolution automatically enforces incompressibility. As a consequence, there is (in-principle) no limitation to the order in time to which the reformulated equations can be numerically discretized. In contrast, the projection method is equivalent to an approximate $L\,U$ matrix factorization [47, 79, 97] of the discrete differential operators coupling $\mathbf{u}$ and p. This approximate factorization yields a splitting error in time, which is very hard to circumvent in order to achieve higher accuracy.

Second, the pressure Poisson formulation used here ensures both that, at every point in the time integration: (i) the normal and tangential velocity boundary conditions are accurately satisfied. (ii) The zero divergence condition is accurately satisfied. On the other hand, in the original projection method, the step advancing the flow velocity forward in time enforces the velocity boundary conditions, but cannot guarantee that incompressibility is maintained. In the other step, the pressure is used to recover incompressibility — by projecting the flow velocity field onto the space of divergence free fields. Unfortunately, while removing the divergence, this second step does not necessarily preserve the correct velocity boundary values [47].

Finally the modified equation (2.72) implemented here differs from the projection's method velocity forward step primarily by the boundary conditions imposed.[7] Specifically, the projection method imposes Dirichlet boundary conditions for each component of the velocity field when propagating the flow velocity. Dirichlet boundary conditions have the

---

[7] In the projection method the contribution from the pressure is ignored in the momentum equation step, but we will not focus on this here — the prior two paragraphs deal with the consequences of this difference.

obvious advantage of decoupling the velocity components. Numerically this means that an implicit treatment of the stiff viscosity operator $\partial_t - \mu\,\Delta$ is straightforward. By contrast, on a regular grid, the boundary conditions in (2.72) couple the boundary velocities. Although this coupling does not pose a serious difficulty for explicit schemes, it adds a (potentially serious) difficulty to the implementation of any scheme that treats the viscosity operator in the equations implicitly.

2.8 Implementation

The objective of this section is to study the convergence and accuracy of the scheme proposed in § 2.7. In particular: to verify the theoretical prediction that the scheme is second order in space, all the way up to the boundary. To this end, here we present the results from numerical computations done with two examples of flows driven by an external forcing $\mathbf{f}$ on a domain: (i) Flow on a square domain in § 2.8.1, and (ii) Flow on an irregular domain in § 2.8.2. In these examples the external forcing is selected so that an exact (analytic) solution to the equations can be produced [47].[8] In each case we compare the numerically integrated fields, $\mathbf{u}$ and p, with the known exact fields, and the errors are computed in an appropriate norm (see below). Note that, as pointed out at the beginning of § 2.7, here we only solve the linear Navier-Stokes equations. At some level, the nonlinear terms can be thought of as a non-zero divergence external forcing function. Consequently, to ensure a fair test, we choose $\mathbf{f}$ so that it is not divergence free.

In what follows the numerical errors are measured using the discrete $L^{\infty}_{\Delta x}$ grid norm defined by

$$\|g\|_{\infty} = \max\{|g_{ij}|\}. \tag{2.86}$$

Here the maximum is over all the indexes i and j — corresponding to the appropriate grid coordinates $(x_i,\, y_j)$ in the computational domain — for the field g in the staggered grid. Namely: nodes for the pressure p, mid-points of the horizontal edges for u, and mid-points

---

[8] That is, proceed backwards: first write an incompressible flow, and then compute the forcing, and boundary conditions, it corresponds to.

of the vertical edges for v. Note that neither ghost pressure points, nor boundary velocities, are included within the index set in (2.86) — we consider these as auxiliary numerical variables, introduced for the purpose of implementing the boundary conditions, but not part of the actual solution. Finally, for the exact (analytic) solution, grid values are obtained by evaluation at the points $(x_i,\, y_j)$. For example, to compute the error in the pressure, first define the discrete field $e_{ij} = \hat{p}_{ij} - p(x_i,\, y_j)$ — where $\hat{p}$ is the numerical pressure field and p is the exact continuous field, and then compute the norm above for e.

2.8.1 Flow on a square domain

Here we present the results of solving the linear Navier-Stokes equations on the unit square $0 \leq x,\, y \leq 1$, with no-slip and no flux boundary conditions (i.e.: $u = v = 0$ on $\partial\Omega$), and viscosity $\mu = 1$. We selected the forcing function $\mathbf{f} = \mathbf{u}_t + \nabla p - \mu \Delta \mathbf{u}$ to yield the following (incompressible) velocity and pressure fields:

$$
\begin{aligned}
u(x,\, y,\, t) &= \pi \cos(t) \sin(2\pi y) \sin^2(\pi x), && (2.87)\\
v(x,\, y,\, t) &= -\pi \cos(t) \sin(2\pi x) \sin^2(\pi y), && (2.88)\\
p(x,\, y,\, t) &= -\cos(t) \cos(\pi x) \sin(\pi y). && (2.89)
\end{aligned}
$$

Using as initial conditions those provided by (2.87–2.88) at $t = 0$, we solved the equations for various grid sizes Δx, with time step $\Delta t = 0.2\,(\Delta x)^2$ — for stability: see § 2.7.1, equation (2.74). Then we compared the numerical results with the exact fields in (2.87–2.89). For instance, figure 2-3 shows the error fields for the velocity and pressure at time $t = 4\pi$, for a (fairly coarse) 40×40 grid. Note that the error is fairly uniform in size across the whole domain. Unlike the common situation with projection methods [47], there are neither numerical boundary layers, nor numerical corner layers.

Since the algorithm is second order consistent in space, and first order in time, with $\Delta t \sim (\Delta x)^2$, we expect the errors to scale with $(\Delta x)^2$. This is precisely what we observe, with the errors measured in the L^∞ norm — see equation (2.86). We omit the convergence

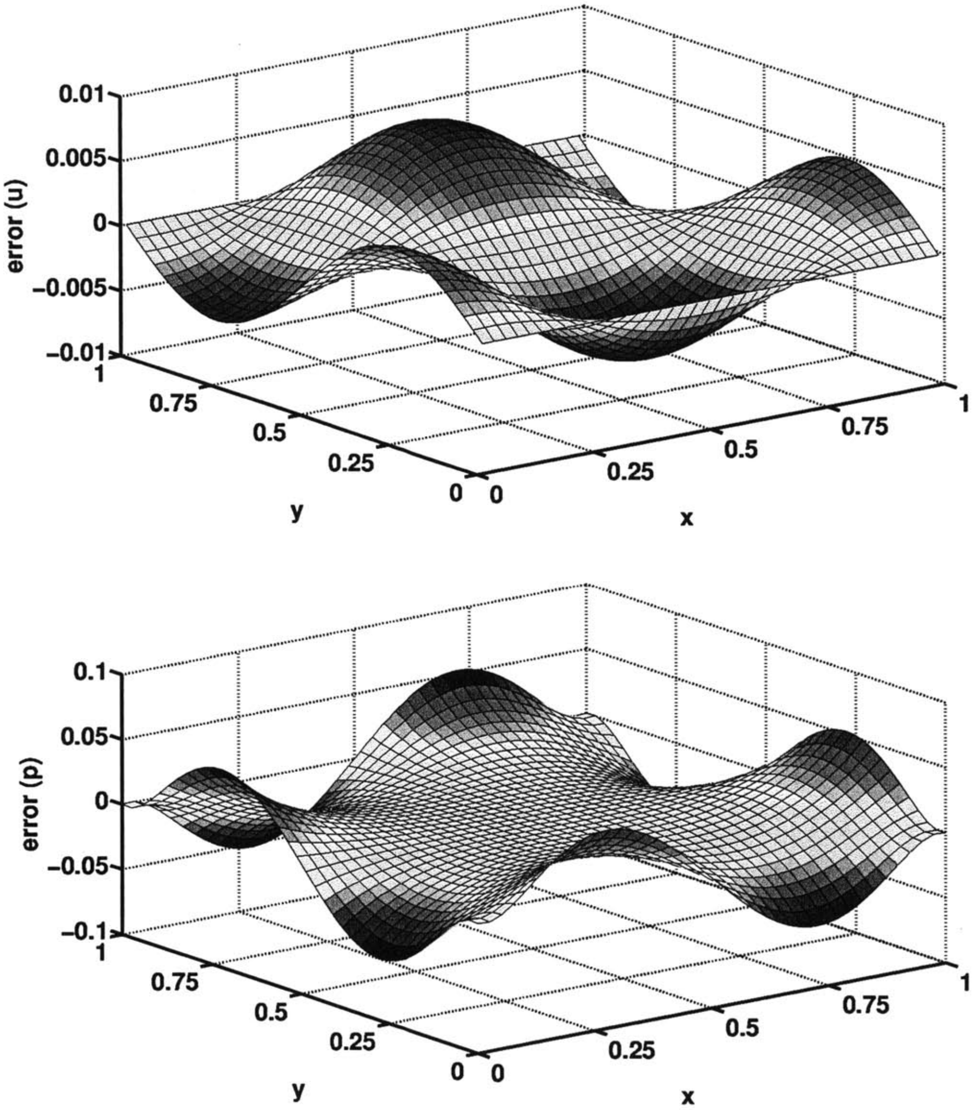

Figure 2-3: Error fields for the numerical solution corresponding to the exact formulas in (2.87–2.89). The plots are at time $t = 4\pi$, for a 40×40 grid, with $\Delta t = 0.2\,(\Delta x)^2$ and $\lambda = 100$. The horizontal velocity u (top) and the pressure p (bottom) error fields are shown. The error is uniform in size across the domain. There are neither numerical boundary layers, nor numerical corner layers.

plots for the current case, and show them for the case presented in § 2.8.2 only.

This example is not particularly taxing in terms of the algorithm implementation, since the boundaries are parallel to the grid lines. Instead, the aim here is to illustrate the role of the feedback parameter λ. For this purpose we present here the results of two tests. In the first test, we evolved the numerical solutions (with a fixed Δx) for different values of λ. To be precise, we used a 40×40 grid.[9] Figure 2-4 illustrates the results of this first test, with plots of the L^∞ error in the horizontal velocity u, as a function of time, for various values of λ. The errors in the pressure p, and in the vertical velocity v, exhibit the same qualitative behavior. In the absence of feedback — i.e. $\lambda = 0$, the errors grow steadily in time, as can be expected from equation (2.18) when influenced by the presence of numerical noise. Even though this effect does not constitute an instability (in the sense of exponential growth), there appears to be no bound to the growth. Thus, after a sufficiently long time, the errors can become substantial. The figure shows also that moderate values of λ — i.e. $\lambda \sim 10$ or bigger, are enough to control the errors.

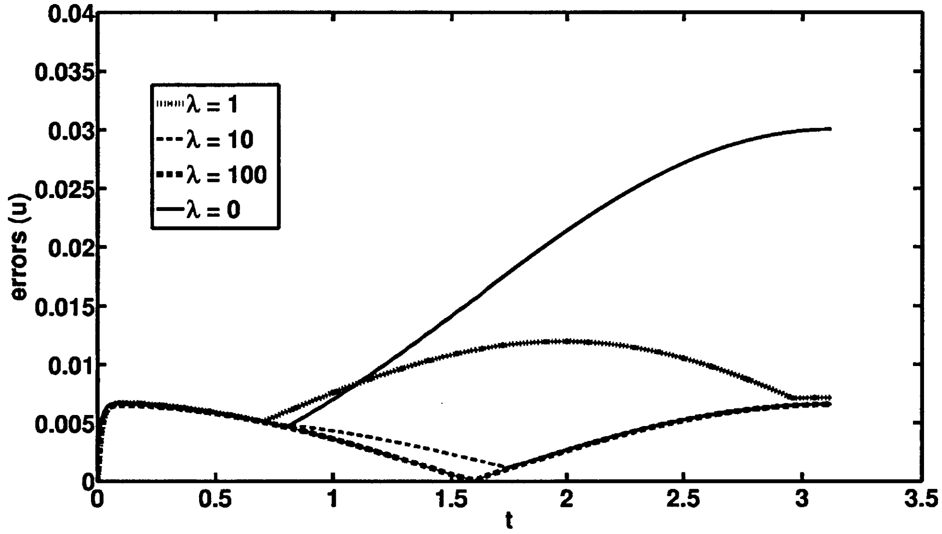

Figure 2-4: Time evolution of the L^∞ error in u, for the flow in a square domain, on a 40×40 mesh, with different values of the parameter λ. For $\lambda = 0$ the error grows in time. Values larger than $\lambda \sim 10$ control the error.

Physically, the errors that occur when λ is too small correspond to fluid leakage through the domain walls. However, these errors cause no net mass loss or gain, since $\nabla \cdot \mathbf{u} = 0$ applies — actually, $\nabla \cdot \mathbf{u} = O\left((\Delta x)^2\right)$, as we show later: see figure 2-14. Any positive flow across some part of the boundary must be compensated by a negative flow elsewhere.

---

[9] With viscosity $\mu = 1$ and a a time step $\Delta t = 0.2 (\Delta x)^2$.

In the second test, we introduce artificial random errors to the normal velocity along the boundary. The purpose of this second test is to study, under a controlled situation, the error sensitivity of the normal velocity boundary condition implementation — the purpose for which the parameter λ was introduced in § 2.4. Specifically, the main concern are the errors that are introduced by the implementation of the Neumann boundary condition for the pressure on non-conforming boundaries (as happens for the example in § 2.8.2). For simple geometries, such as a square, a finite differences discretization of the equation can be easily designed so that the discrete analog of the solvability condition applies. For curved geometries, however, this is generally not possible — therefore, a least squares solution of the equation is required, which introduces errors throughout the pressure field. Further, even though these errors are small (of the same order as the approximation scheme used), they are hard to quantify in detail — e.g. write a leading order approximation in terms of derivatives of the exact solution.

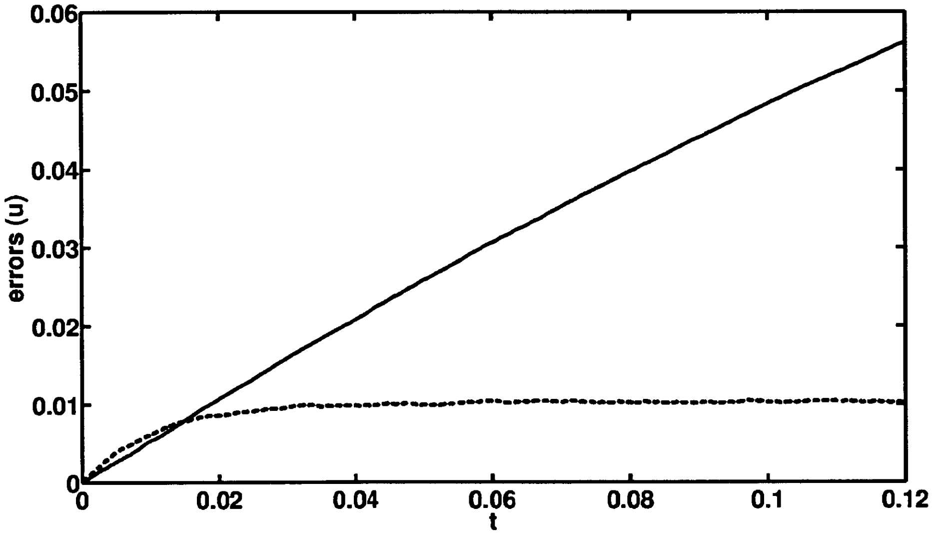

Figure 2-5: Time evolution of the L^∞ error in u, for the flow in a square domain, on a 40×40 mesh, with a random forcing on the boundary. Without feedback ($\lambda = 0$, solid line) the error grows steadily. With feedback ($\lambda = 100$, dashed line) the error growth saturates at a small value.

To perform the test, we introduce independent random errors at each point on the domain boundary, at the start of each Euler time step. To be precise, the boundary condition for equation (2.73) is taken as

$$\mathbf{n} \cdot \nabla p^n = \mathbf{n} \cdot (\mathbf{f}^n + \Delta \mathbf{u}^n + \lambda\, \mathbf{u}^n) + r, \tag{2.90}$$

where, at each boundary point r is sampled randomly from the interval $[0, 1]$. This represents an $O(1)$ random perturbation to the normal component of the desired velocity, $\mathbf{g} = 0$, at the boundary.[10] Hence, it is a very demanding test of how insensitive the boundary condition implementation is to errors. On the other hand, because of the highly "oscillatory" nature of the perturbation, the effect it has is to merely produce a thin boundary layer on the pressure — with the perturbations damped away from the boundary. Hence the perturbation only affects the solution at the boundary, and a large enough value of λ can keep the solution under control.

Figure 2-5 shows the error in the horizontal velocity u as a function of time, both for the feedback controlled solution with $\lambda = 100$, and the undamped solution with $\lambda = 0$. The errors saturate in the feedback solution, but they grow steadily for the undamped solution. Without damping, the drift errors contribute sizable effects to the solution after a short time period.

Finally, figure 2-6 illustrates the flow leakage produced by the random errors. This figure shows a cross section of the horizontal velocity u after 1000 time steps, for two values of the control parameter: $\lambda = 0$ and $\lambda = 100$. The first value shows a significant flow through the boundary, while the second does not. This plot also illustrates the point made earlier: there is no significant mass loss (gain) even when $\lambda = 0$, with positive flows compensated by negative flows elsewhere.

2.8.2 Flow on an irregular domain

In this section we implement the numerical scheme described in § 2.7, for an example with an externally forced velocity field on an irregular shaped domain. An exact solution in

---

[10] Recall that in this section we take $\mu = 1$, and neglect the nonlinear terms.

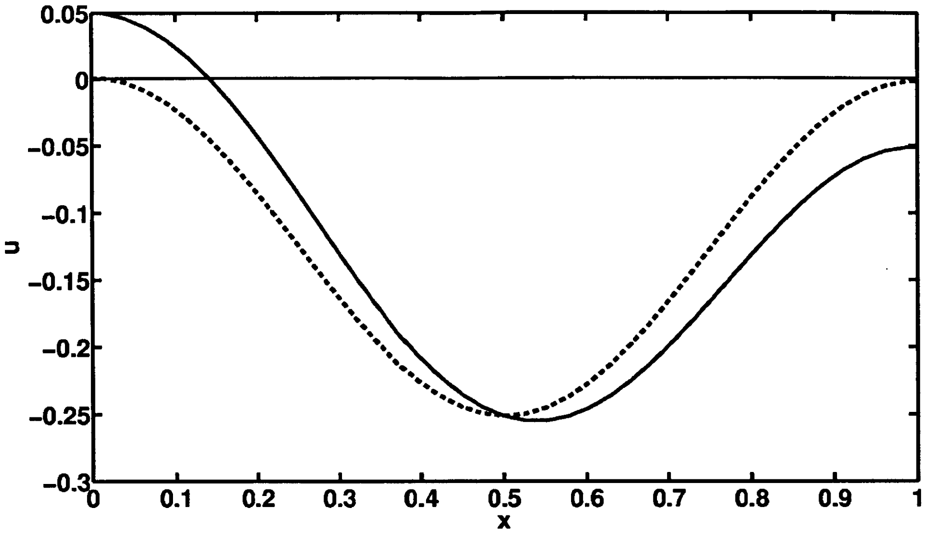

Figure 2-6: Velocity cross-section $u(x,\, y = 0.4872)$, at time $t = 0.1315$ (1000 time steps), for the flow in a square domain, on a 40×40 mesh, with a random forcing on the boundary. The dashed line ($\lambda = 100$) reproduces the correct field with zero flux at the boundaries. The solid line ($\lambda = 0$) shows a non-zero flow through the boundary.

the domain, needed to compute the errors, is constructed in a similar fashion to that in § 2.8.1. However, instead of the unit square, the selected domain Ω is the 2×2 square $0 < x,\, y < 2$, with the disk of radius $1/4$ centered at $(3/4,\, 1)$ removed. In addition, we use periodic boundary conditions (with period 2) in the y direction, and impose a nonzero flux ($\mathbf{u} = \mathbf{g} \neq 0$) on the rest of the boundary.

To construct an incompressible velocity field in the domain Ω, we use a stream function ψ, with $\mathbf{u} = (u,\, v) = (\psi_y,\, -\psi_x)$. Then we prescribe a pressure, and choose the forcing function by $\mathbf{f} = \mathbf{u}_t + \nabla p - \mu \Delta \mathbf{u}$, with viscosity $\mu = 1$. The function $\mathbf{g}$, used to prescribe the boundary condition for the velocity, is selected to match the values given by $(\psi_y,\, -\psi_x)$.

In this example we set the feedback parameter λ to $\lambda = 100$, considerably larger than the minimum needed to get good behavior in the calculations in § 2.8.1. This is to be expected from the results of the last test in § 2.8.1 — see figures 2-5 and 2-6. Because of the curved inner boundary in the current example, errors in the implementation of the Neumann boundary condition for the pressure occur, which trigger a drift in the normal boundary condition for the velocity — unless a large enough λ is used. Thus, for example, values where $\lambda = O(10)$ did not allow the feedback to quickly track fluctuations in the normal velocity component of the boundary conditions.

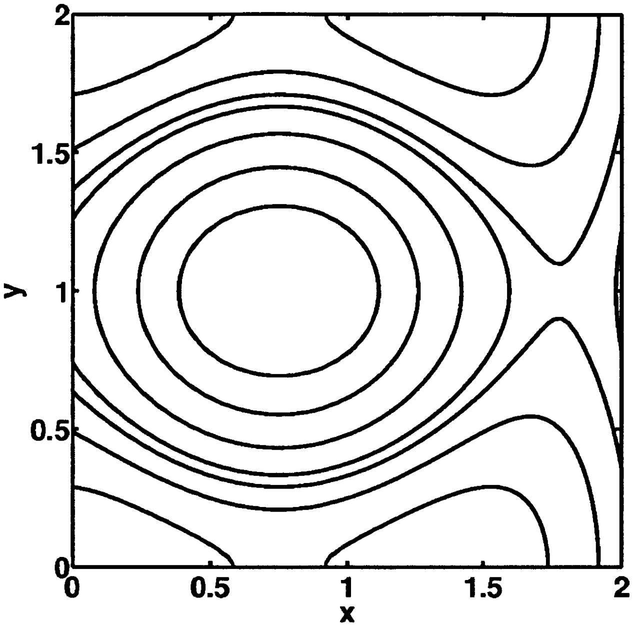

Figure 2-7: Contour lines for the stream function ψ, defined in equation (2.92), on the irregular domain of the second example. The domain is a 2×2 square, with a hole of radius $1/4$ centered at $(x,\, y) = (3/4,\, 1)$. The stream function is periodic, of period 2, in the y-direction.

The stream function $\psi = \psi(x,\, y,\, t)$ (see figure 2-7) is defined as follows: First, introduce the function $\psi_0 = \psi_0(x,\, y)$, defined on $\mathbb{R}^2$, by

$$\psi_0(x, y) = r^2 e^{-2r}, \tag{2.91}$$

where $r = \sqrt{(x - 3/4) + (y - 1)^2}$. Then construct a 2-periodic function in y by adding shifted copies ψ_0, as follows

$$\psi(x,\, y,\, t) = \cos(t) \sum_{k=-\infty}^{\infty} \psi_0(x,\, y + 2k). \tag{2.92}$$

Finally, the pressure is given by the formula

$$p(x,\, y,\, t) = \sin(t)\, \cos(\pi\, x)\, \sin(\pi\, y) \tag{2.93}$$

Figure 2-7 shows the contour lines for ψ on the domain Ω. Note that, although both ψ_0 and $\frac{\partial}{\partial r}\psi_0$ vanish at the edge of the circle centered at $(3/4,\, 1)$ — the inner boundary of Ω, the shifted copies of ψ_0 do not. Therefore, the resulting flow velocity $(\psi_y,\, -\psi_x)$ has some small components, both in the normal and tangent directions to the circle.

Second order convergence.

To illustrate the second order convergence of the scheme, below we present the results of evolving the velocity and pressure fields (for the problem described above), for varying grid sizes $\Delta\, x$, up to the fixed time $t = 0.0657$, using $\lambda = 100$ and $\Delta\, t = 0.2\, (\Delta\, x)^2$. Note that the selected final time appears small, but it is large enough to require a number of time steps in the range $O(10^3)$ to $O(10^4)$ — for the grid sizes we used. This is large enough to provide a reliable measure of the order of the scheme.

Figure 2-8 shows the convergence of velocity and pressure, while figure 2-9 shows the convergence of the partial derivatives of the horizontal velocity u. The figures indicate second order L^∞ convergence for the velocity u, the pressure p, and even the velocity gradient $(u_x,\, u_y)$ — see § 2.8.3.

Typical error behavior.

We illustrate the (typical) error behavior — both in time and in space, by showing the results of a calculation done at a fixed resolution. Specifically, we take an 80 × 80 grid, with $\Delta t = 0.2(\Delta x)^2$ and $\lambda = 100$, and solve the forced system of equations in this "flow on an irregular domain" example — recall that $\mu = 1$.

Figure 2-10 shows the time evolution of the L^∞ (spatial) errors in the pressure and in the horizontal velocity. It should be clear that, while the errors oscillate (over almost one decade in amplitude), they do not exhibit any measurable growth with time.

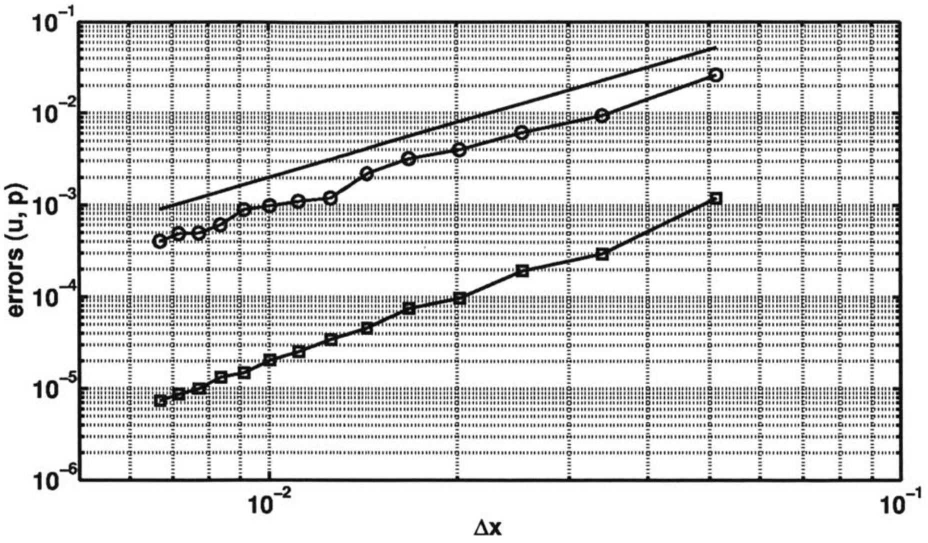

Figure 2-8: (Flow on an irregular domain example). Convergence plot for the pressure (circles: ∘) and the horizontal velocity (squares: □). The errors (in the L^∞ norm) are computed at the fixed time $t = 0.0657$, for different grid resolutions, with $\Delta t = 0.2\,(\Delta x)^2$. The slope of the solid straight line corresponds to the second order scaling error $\propto (\Delta x)^2$.

Figure 2-11 shows the horizontal velocity field at time $t = 4.25\,\pi$, while figure 2-12 shows the associated error field. Notice that, while the error is largest near the internal boundary — as expected from the difficulties that a curved, non-conforming, boundary causes — it is fairly well behaved, without abrupt transitions on the scale of the grid size. This explains why the error in the gradient of the velocity is also second order accurate — see figure 2-9 and remark 2-9, a feature that should be particularly useful for the calculation of fluid-solid forces (stresses) along domain boundaries.

Similarly, figure 2-13 shows plots of the pressure and of the associated pressure error field, again for the time $t = 4.25\,\pi$. Just as in the case of the velocity, the errors are dominated by what happens at the internal circular boundary. The error near the internal wall is a bit more jagged than the error for the velocity field. We did not check the order of convergence for the gradient of the pressure since one does not need the gradient of the pressure to compute forces.

Finally, figure 2-14 shows a plot of the error in the numerical divergence of the flow field, also for time $t = 4.25\,\pi$. The errors in the divergence are also second order. However:

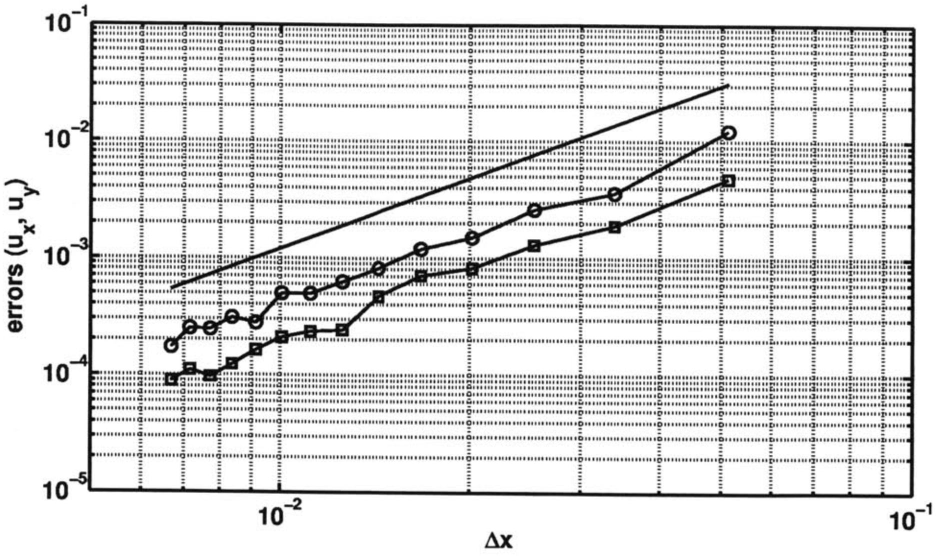

Figure 2-9: (Flow on an irregular domain example). Convergence plots for the partial derivatives of the horizontal velocity: u_y (circles: ○) and u_x (squares: □). The errors (in the L^∞ norm) are computed at the fixed time $t = 0.0657$, for different grid resolutions, with $\Delta t = 0.2\,(\Delta x)^2$. The slope of the solid straight line corresponds to the second order scaling error $\propto (\Delta x)^2$. The fact that the errors for the gradient of the velocity are also second order is important — see § 2.8.3.

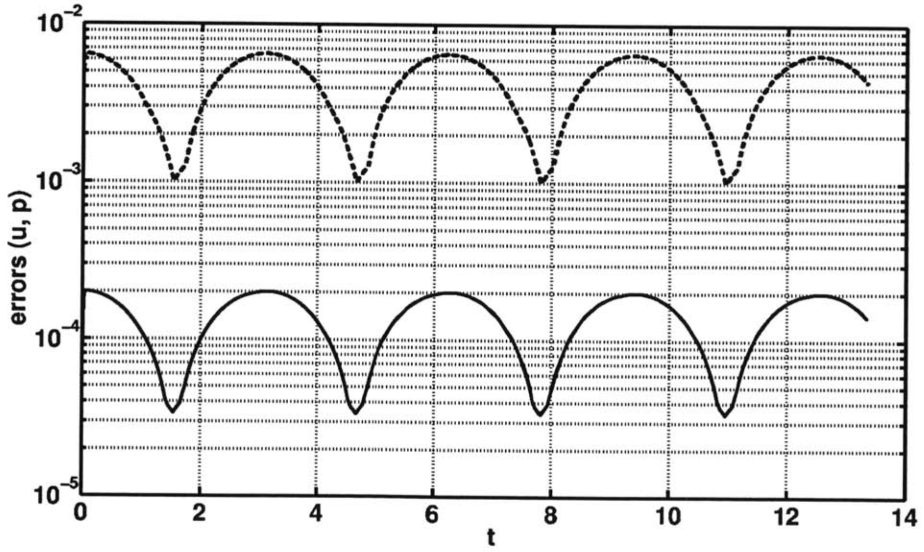

Figure 2-10: (Flow on an irregular domain example, on an 80×80 grid). Evolution in time of the L^∞ norms of the (spatial) errors for the pressure (dashed curve) and horizontal velocity (solid curve).

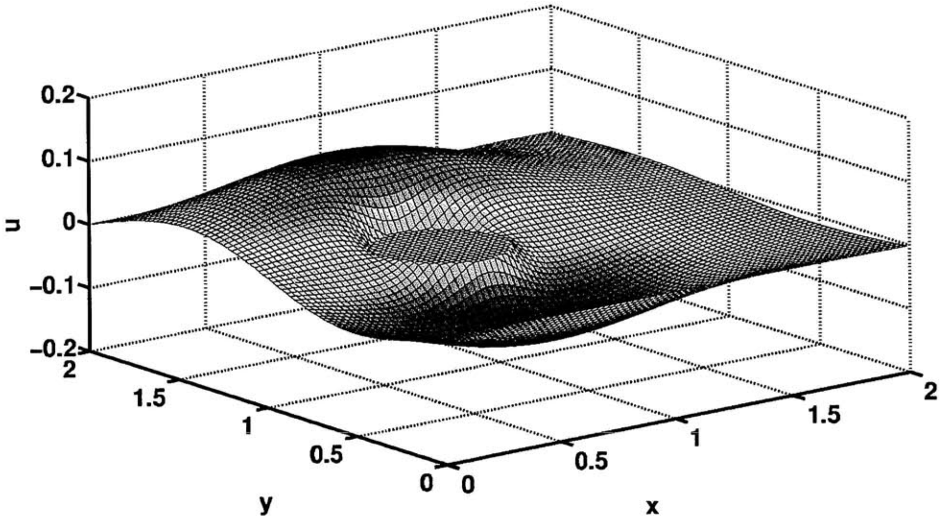

Figure 2-11: (Flow on an irregular domain example, on an 80×80 grid). Numerical horizontal velocity field u at $t = 4.25\,\pi$.

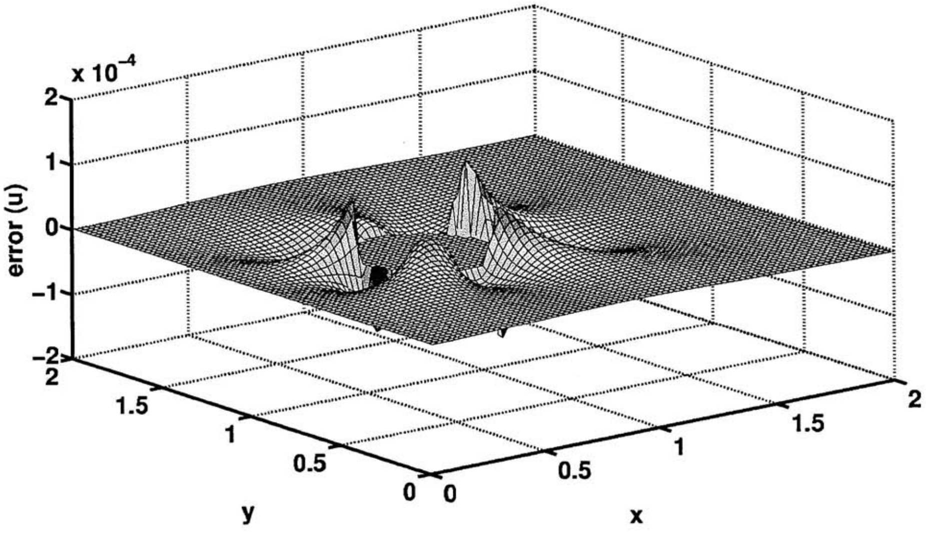

Figure 2-12: (Flow on an irregular domain example, on an 80×80 grid). Error field for the horizontal velocity u at $t = 4.25\,\pi$.

notice that they are much smaller than the errors in the flow field or in the pressure. This is indicative of the rather strong enforcement of incompressibility that equations (2.9) and (2.10) guarantee.

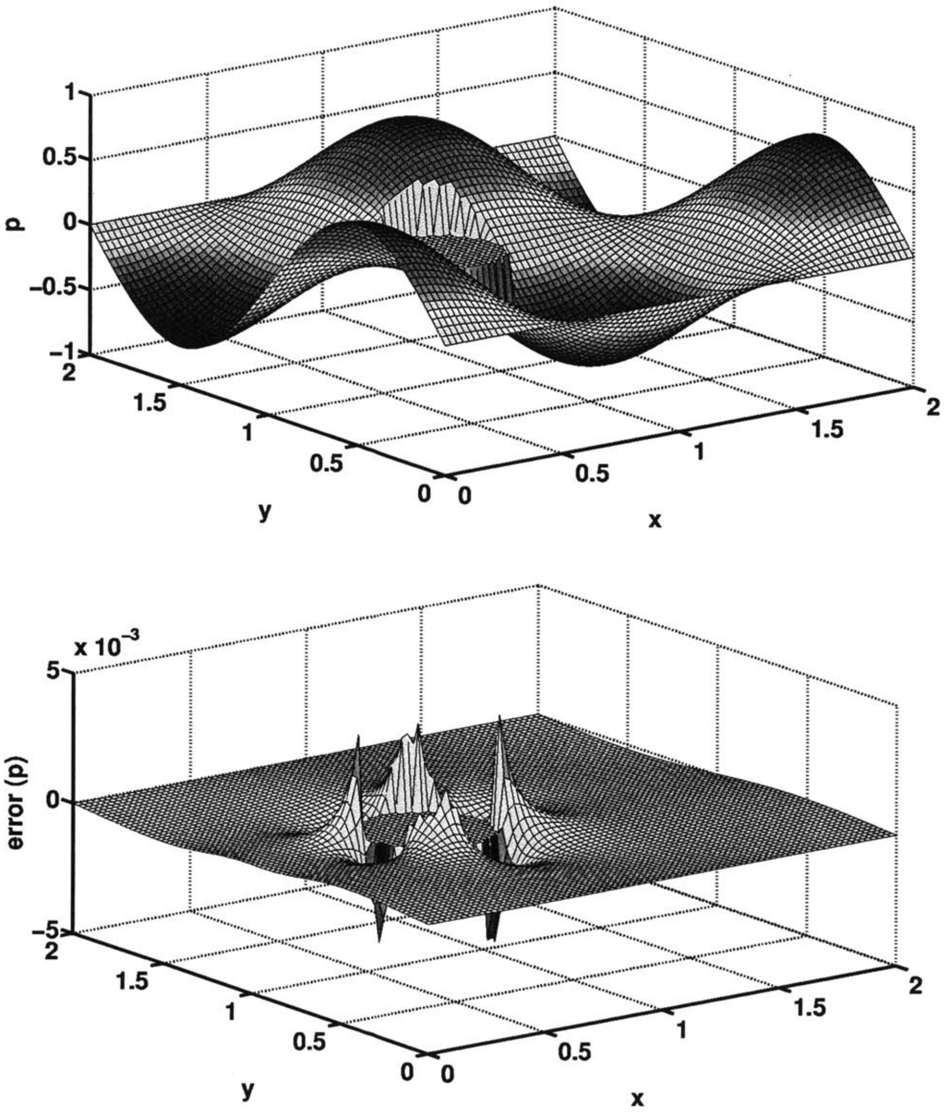

Figure 2-13: (Flow on an irregular domain example, on an 80 × 80 grid). Left: numerical pressure field p at $t = 4.25\,\pi$. Right: associated error field.

2.8.3 Convergence of the derivatives

A key reason to worry about uniform convergence, up to the boundary, of the numerical solution is that this is a necessary condition for the accurate modeling of solid-fluid interactions. The evaluation of fluid-solid stresses requires both the pressure, as well as the derivatives of the velocities, to be accurate at the boundaries. Hence, we also investigate the behavior of

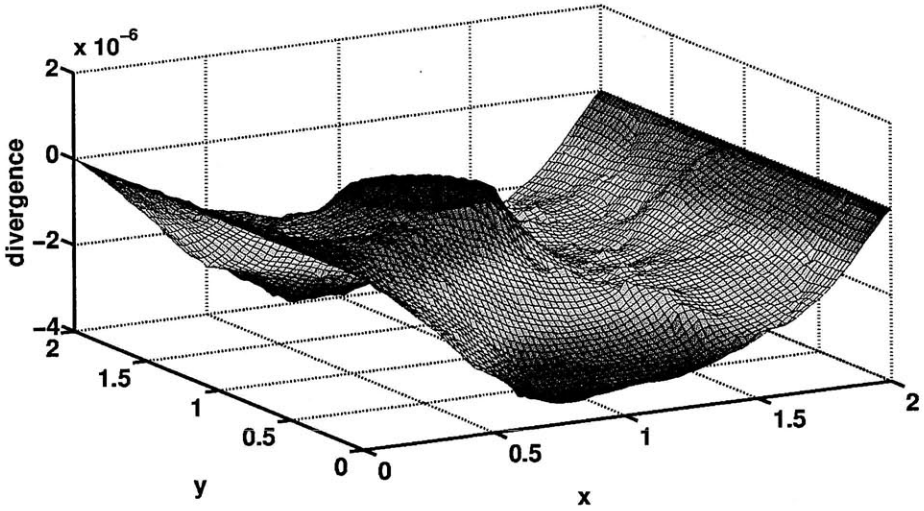

Figure 2-14: (Flow on an irregular domain example, on an 80 × 80 grid). Numerical error in the divergence of the velocity at $t = 4.25\,\pi$. The $O(10^{-6})$ amplitude is the result of the $O(\Delta x^2)$ order of the scheme.

the errors for not just the pressure p and the velocity fields $(u,\, v)$, but for the gradient of the flow velocity as well.

An *important point* to notice is that figure 2-9 shows that

$$\left.\begin{array}{l}\text{The derivatives of the velocities exhibit errors that appear to be}\\ \text{second order as well— in the } L^\infty \text{ norm.}\end{array}\right\} \tag{2.94}$$

At first sight this may seem surprising: since the velocities are second order accurate, one would expect their derivatives to be first order only. However, this is a worse case scenario, based on the assumption of random errors. On the other hand, the errors for finite differences approximations (on regular grids) are typically *not* random: If the solutions are smooth enough, the errors can be expanded in powers of Δx, with coefficients that involve derivatives of the solution. Hence, in this case, taking low order derivatives[11] does not degrade the order of convergence. This, of course, is an important advantage of finite differences over other approaches (e.g. projection) which do not have this property.

---

[11] At high order, round off errors dominate. There is an irreducible $O(\epsilon\,(\Delta x)^{-q})$ contribution from them, where ϵ is the round off error, and q is the order of the derivative.

The astute reader may have noticed that the argument in the prior paragraph avoids the issue of boundary condition implementation, which can ruin the smooth error expansions available for finite differences. This requires some extra considerations:

2.8a On conforming boundaries, it is usually possible to approximate the boundary conditions in such a way that smooth expansions (in powers of Δx) remain available for the truncation errors.

2.8b On non-conforming boundaries, on the other hand, it is very easy to ruin the smooth error expansions. The local stencils used to approximate the boundary conditions along the boundary will, typically, experience abrupt changes in response to the placement of the boundary relative to the regular grid.

This is where the global approach that we used in § 5.2 to implement the boundary conditions comes to the rescue: by linking each local stencil to its neighbors, the non-smoothness caused by abrupt stencil changes is smeared. Thus a better behaved error is obtained, which then explains why the result in (2.94) occurs. Notice that, as of now, we neither know if the smearing process is enough to make the errors C^1 at leading order — which is what is needed to make the errors in the velocity gradient second order, nor if the errors in the velocity gradient are actually second order. However, the numerical evidence seems to point in this direction.

A final point is that, the "simplest" way to get around the issue in item 2.8b above, is to implement the boundary conditions at a higher order than required. This has the disadvantage that it can lead to a messier than needed algorithm, but (generally) it should not significantly increase the computational cost — since it only involves the boundary points of the grid.

2.8.4 Driven cavity

In this subsection, we solve for the fluid velocity field in a driven cavity. We use the second order semi-implicit scheme (2.57) and the regular staggered grid described above. In addition, this test also demonstrates the application of a non-vanishing boundary condition where $\mathbf{n} \times \mathbf{u} \neq 0$. Here we include the nonlinear terms and treat them explicitly in the numerical scheme. To discretize them in space, we use a standard, second order upwind finite difference scheme. In this test, we take the velocity of the moving wall $U_{wall} = 1$ and consider different Reynolds numbers. Starting with rest initial conditions, we evolve the fields, setting $\lambda = 30$, until they reach a steady state flow. Figures (2-15) and (2-16) show the midpoint steady state velocities for $\mathbf{u}$ at Reynolds numbers $Re = 100$ and $Re = 400$ respectively. The plots are done with a 220×220 grid and also compare the computation to the standard benchmark of [36]. In addition figure (2-17) shows the streamfunction contours for a Reynolds number of $Re = 400$.

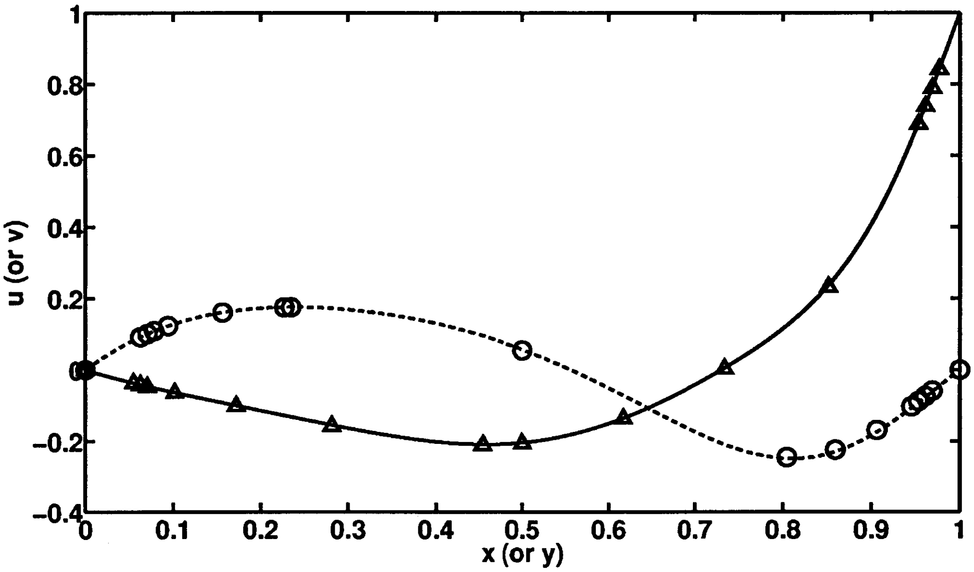

Figure 2-15: Midpoint steady state velocities (after $t = 20$) for $Re = 100$ and a 220×220 grid. For $\mathbf{u} = (u, v)$, the solid line shows $u(0.5, y)$, the dashed line shows $v(x, 0.5)$, while the circles and triangles correspond to data from [36].

2.9 The $\nabla \cdot \mathbf{u} = 0$ Boundary Condition

In the numerical implementation section of this chapter, the $\nabla \cdot \mathbf{u} = 0$ boundary condition is discretized using a regular Cartesian grid. In Cartesian coordinates, this boundary condition

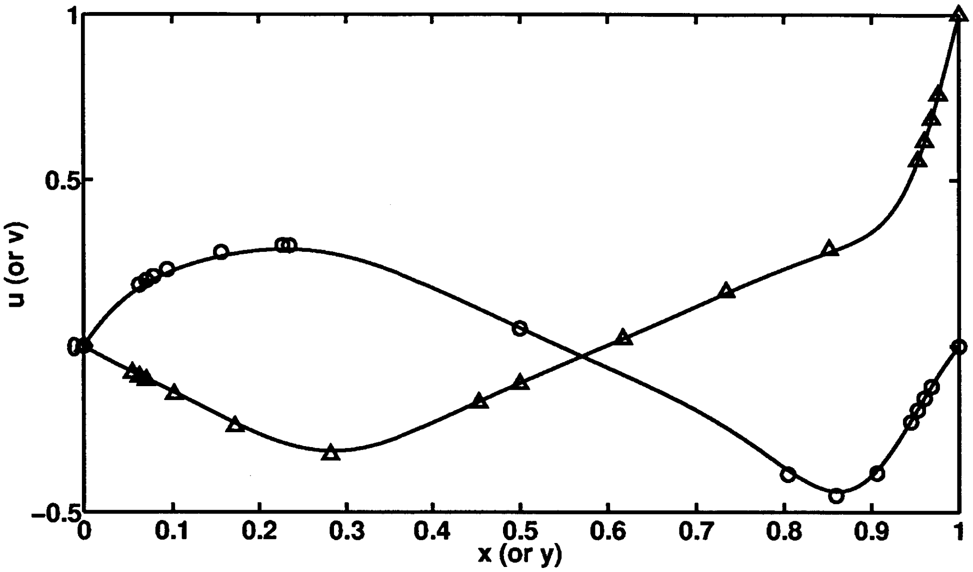

Figure 2-16: Midpoint steady state velocities (after $t = 30$) for $Re = 400$ and a 220×220 grid. For $\mathbf{u} = (u, v)$, the solid line shows $u(0.5, y)$, the dashed line shows $v(x, 0.5)$, while the circles and triangles correspond to data from [36].

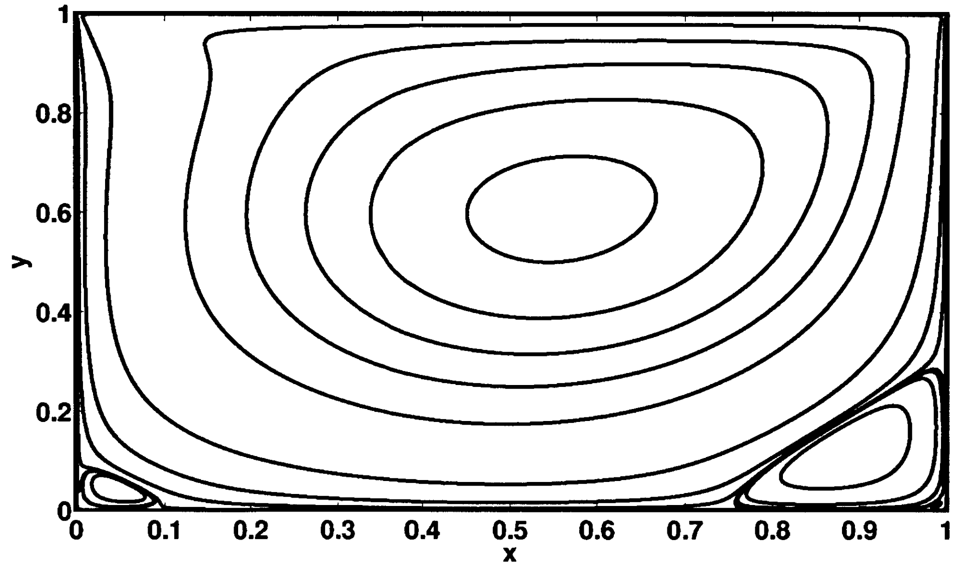

Figure 2-17: Plots shows steady state streamfunction contours for the driven cavity with $Re = 400$.

relates the horizontal and vertical velocities (for two dimensions) via the standard formulas

$$\mathbf{u} = u\,\hat{\imath} + v\,\hat{\jmath}, \tag{2.95}$$

$$\nabla \cdot \mathbf{u} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}, \tag{2.96}$$

where $\hat{\imath}$ and $\hat{\jmath}$ are the coordinate unit vectors.

For some applications, it may be more convenient to implement the numerical computation on a coordinate system where the boundaries are conforming. In general curvilinear coordinate systems, the divergence acquires a more complicated form than the one above. Hence, it is our purpose here to display the form that the $\nabla \cdot \mathbf{u} = 0$ boundary condition takes for a conforming boundary in a general orthogonal curvilinear coordinate system. We do the 2-D case only — the 3-D case is quite similar.

2.9.1 Curvilinear coordinates and a conforming boundary

In some region near the boundary $\partial\Omega$ of the domain of integration, assume that an orthogonal, curvilinear, set of coordinates (η, ξ) has been selected — such that the boundary is given by $\eta = c$, for some constant c. In terms of the coordinates (η, ξ), the vector field $\mathbf{u}$ can then be written in the form

$$\mathbf{u} = u_\eta\,\hat{\boldsymbol{\eta}} + u_\xi\,\hat{\boldsymbol{\xi}}, \tag{2.97}$$

where u_η and u_ξ are the velocity components in the coordinate directions given by the unit vectors $\hat{\boldsymbol{\eta}}$ and $\hat{\boldsymbol{\xi}}$, respectively. Let now the functions $x = x(\eta, \xi)$ and $y = y(\eta, \xi)$ describe the relationship of the curvilinear coordinates with a Cartesian system.

Then (e.g. see [56])

$$\hat{\boldsymbol{\eta}} = \frac{1}{s_\eta}\left(\frac{\partial x}{\partial \eta}\,\hat{\imath} + \frac{\partial y}{\partial \eta}\,\hat{\jmath}\right) \quad \text{and} \quad \hat{\boldsymbol{\xi}} = \frac{1}{s_\xi}\left(\frac{\partial x}{\partial \xi}\,\hat{\imath} + \frac{\partial y}{\partial \xi}\,\hat{\jmath}\right), \tag{2.98}$$

where s_η and s_ξ are local normalization factors

$$s_\eta^2 = \left(\frac{\partial x}{\partial \eta}\right)^2 + \left(\frac{\partial y}{\partial \eta}\right)^2 \quad \text{and} \quad s_\xi^2 = \left(\frac{\partial x}{\partial \xi}\right)^2 + \left(\frac{\partial y}{\partial \xi}\right)^2 . \tag{2.99}$$

Further, re-interpreting s_η and s_ξ as functions of (η, ξ), we have

$$\nabla \cdot \mathbf{u} = \frac{1}{s_\eta\, s_\xi} \left(\frac{\partial}{\partial \eta} (s_\xi\, u_\eta) + \frac{\partial}{\partial \xi} (s_\eta\, u_\xi) \right) . \tag{2.100}$$

The $\nabla \cdot \mathbf{u} = 0$ boundary condition is accompanied by the Dirichlet boundary condition $\mathbf{n} \times \mathbf{u} = \mathbf{n} \times \mathbf{g}$ for the vector momentum equation. For a conforming boundary, this is just $u_\xi = g_\xi$ along the boundary curve where $\eta = c$. Thus the $\nabla \cdot \mathbf{u} = 0$ boundary condition reduces to a *Robin* boundary condition on the normal velocity (provided that the boundary is piecewise smooth)

$$\frac{\partial}{\partial \eta} (s_\xi\, u_\eta) + \frac{\partial}{\partial \xi} (s_\eta\, g_\xi) = 0 \tag{2.101}$$

along $\eta = c$.

2.9.2 Example: polar coordinates

Consider the special case of a polar coordinate system $x = r \cos\theta$ and $y = r \sin\theta$, with the boundary at $r = 1$. Let the angular velocity at the boundary be given by $u_\theta(1,\, \theta) = g_\theta(\theta)$. Then since (in this case) $s_\theta = r$ and $s_r = 1$, the zero divergence boundary condition becomes

$$\frac{\partial}{\partial r} (r\, u_r) + \frac{\partial}{\partial \theta} g_\theta = 0 \quad \text{for } r = 1. \tag{2.102}$$

2.10 Steps Toward a Finite Element Implementation

As is often the case in engineering applications, finite element methods are a popular choice for the spatial discretization of a partial differential equation. One reason for their popularity is that they reduce some of the difficulties of computing solutions in arbitrary geometries. In this section we lay some preliminary ideas for a weak formulation of the recast Navier-Stokes system. With the weak formulation, one may then seek to develop appropriate finite element schemes.

In our recast Navier-Stokes system, the recovery of the pressure from a velocity field involves the solution of a Poisson equation, with Neumann boundary conditions. In this case, the weak formulation of a Poisson equation may be found in standard texts on finite elements methods [100]. Here, we are ultimately interested in the weak formulation for the recast Navier-Stokes momentum equation. As a first step, we examine the corresponding, vector Poisson equation (in a domain Ω with a smooth boundary):

$$\left.\begin{aligned} -\Delta \mathbf{u} &= \mathbf{f} && \text{for } \mathbf{x} \in \Omega, \\ \mathbf{n} \times (\mathbf{u} - \mathbf{g}) &= 0 && \text{for } \mathbf{x} \in \partial\Omega, \\ \nabla \cdot \mathbf{u} &= 0 && \text{for } \mathbf{x} \in \partial\Omega, \end{aligned}\right\} \tag{2.103}$$

Now introduce a standard L^2 inner product on functions and vector fields:

$$(u, v) = \int_\Omega uv \, \mathrm{d}V, \tag{2.104}$$

$$(\mathbf{u}, \mathbf{v}) = \int_\Omega \mathbf{u} \cdot \mathbf{v} \, \mathrm{d}V, \tag{2.105}$$

where $\mathbf{u} \cdot \mathbf{v}$ is the dot product for vectors. We now list several integral properties which will be useful in the variational formulation of the vector Poisson equation:

$$(\nabla \times \mathbf{u}, \mathbf{v}) = (\mathbf{u}, \nabla \times \mathbf{v}) + \int_{\partial\Omega} \mathbf{v} \cdot (\mathbf{n} \times \mathbf{u}) \, \mathrm{d}A \quad \text{for } 3D, \tag{2.106}$$

$$(\nabla \times \mathbf{u}, v) = (\mathbf{u}, \nabla^\perp v) + \int_{\partial\Omega} v(\mathbf{n} \times \mathbf{u}) \, \mathrm{d}A \quad \text{for } 2D, \tag{2.107}$$

$$(\nabla \cdot \mathbf{u}, v) = (\mathbf{u}, -\nabla v) + \int_{\partial\Omega} v(\mathbf{n} \cdot \mathbf{u}) \, \mathrm{d}A \quad \text{for all } D. \tag{2.108}$$

In 2D, the term $\nabla \times \mathbf{u} = \partial_1 u_2 - \partial_2 u_1$ is a scalar, while $\nabla^\perp v = \partial_2 v \hat{\mathbf{x}}_1 - \partial_1 v \hat{\mathbf{x}}_2$ is a standard representation in Cartesian coordinates. Lastly, we note that the Laplacian on a vector field in $3D$ may be written as $-\Delta = \nabla\times\nabla\times - \nabla(\nabla\cdot)$, while in $2D$ we have $-\Delta = \nabla^\perp(\nabla\times) - \nabla(\nabla\cdot)$

[12].

We now introduce the positive (semi)definite functional (defined on H^1):

$$\begin{aligned}
\mathcal{E}[\mathbf{u}] &= \frac{1}{2}\int_\Omega \sum_{k=1}^{D} |\nabla u_k|^2 \,\mathrm{d}V &(2.109)\\
&= \frac{1}{2}\int_\Omega |\nabla\cdot\mathbf{u}|^2 + |\nabla\times\mathbf{u}|^2 \,\mathrm{d}V &(2.110)\\
&= \frac{1}{2}(\nabla\cdot\mathbf{u},\nabla\cdot\mathbf{u}) + \frac{1}{2}(\nabla\times\mathbf{u},\nabla\times\mathbf{u}), &(2.111)
\end{aligned}$$

where u_k are the Cartesian coordinates of $\mathbf{u}$. With the boundary conditions appearing in the identities (2.106-2.108), let us also define:

$$\mathcal{W} = \{\mathbf{u}\in H^2(\Omega)\ :\ \mathbf{n}\times\mathbf{u}=0\ \text{for}\,\mathbf{x}\in\partial\Omega, \nabla\cdot\mathbf{u}=0\ \text{for}\,\mathbf{x}\in\partial\Omega\}. \qquad (2.112)$$

Remark 2.17 *For* $\mathbf{u}\in\mathcal{W}$*:*

$$\mathcal{E}[\mathbf{u}] = \frac{1}{2}(-\Delta\mathbf{u},\mathbf{u}). \qquad (2.113)$$

This follows since the boundary integrals in equations (2.106-2.108) vanish for $\mathbf{u}\in\mathcal{W}$*.*

Remark 2.18 *For* $\mathbf{u},\mathbf{v}\in\mathcal{W}$*:*

$$(-\Delta\mathbf{u},\mathbf{v}) = (\mathbf{u},-\Delta\mathbf{v}). \qquad (2.114)$$

Again, the identity follows from the vanishing boundary integrals. As a consequence, for $\mathbf{u},\mathbf{v}$ *restricted to these boundary conditions, the vector Laplacian* $(-\Delta)$ *is self-adjoint.*

Remark 2.19 *For the self-adjoint operator* $-\Delta$*, there exists:*

1. *A spectral decomposition,* $-\Delta\mathbf{u}_k = \lambda_k\mathbf{u}$ *with* $\lambda_k\geq 0$ *and* $\mathbf{u}_k\in\mathcal{W}$*. The dimension of the* $\lambda_k=0$ *eigenspace (Harmonic fields) depends on the topology of the domain.*

---

[12] These identities are very nicely combined using the theory of differential forms on $\mathcal{R}^D$

2. *A Green's function* $\mathbf{G}$ *such that* $-\Delta\mathbf{G}\mathbf{u} = (\mathbf{I} - \mathbf{P})\mathbf{u}$ *where* $\mathbf{P}$ *is the projection onto the nullspace of* $-\Delta$ *(ie. onto the space of harmonic fields) and* $\mathbf{I}$ *is the identity operator. For functions* $\mathbf{f} \in L^2$ *with no harmonic components (ie.* $\mathbf{Pf} = 0$*), the function* $\mathbf{u} = \mathbf{Gf} \in \mathcal{W}$ *solves* $-\Delta\mathbf{u} = \mathbf{f}$ *with the boundary conditions specified in* $\mathcal{W}$*.*

Remark 2.20 *(Hodge-Helmholtz decomposition) For any vector field* $\mathbf{v}$ *(in* L^2*), there are vectors and scalars* $\mathbf{a}, b, \mathbf{h}$ *such that:*

$$\mathbf{v} = \nabla \times \mathbf{a} + \nabla b + \mathbf{h}, \tag{2.115}$$

The components ∇b *and* $\nabla \times \mathbf{a}$ *are in general not unique, but depend on their boundary values (ie. one could take* $\mathbf{n} \cdot \nabla b = 0$ *or* $\mathbf{n} \times \nabla b = 0$ *and have different decompositions). One special decomposition, that may be useful is:*

$$\begin{aligned} \mathbf{a} &= \nabla \times \mathbf{Gv}, & (2.116)\\ b &= \nabla \cdot \mathbf{Gv}, & (2.117)\\ \mathbf{h} &= \mathbf{Pv}. & (2.118) \end{aligned}$$

Using this choice, the components $\nabla \times \mathbf{a}$*,* ∇b *and* $\mathbf{h}$ *are mutually orthogonal under the vector inner product* $(\cdot, \cdot)$*.*

2.10.1 A Galerkin approach

To approach the topic of a Galerkin method, we compute the variational derivative of $\mathcal{E}$. For fixed $\mathbf{u}$, the variational derivative of $\mathcal{E}$ is a linear functional in $\mathbf{v}$, defined as follows (assume that **u** and **v** are smooth enough but with no particular boundary conditions):

$$\begin{aligned}\frac{\delta\mathcal{E}}{\delta\mathbf{v}} &= \frac{d}{d\epsilon}\mathcal{E}[\mathbf{u}+\epsilon\mathbf{v}]|_{\epsilon=0} && (2.119)\\ &= (\nabla\cdot\mathbf{u},\nabla\cdot\mathbf{v})+(\nabla\times\mathbf{u},\nabla\times\mathbf{v}) && (2.120)\\ &= (-\Delta\mathbf{u},\mathbf{v})+\int_{\partial\Omega}(\nabla\cdot\mathbf{u})(\mathbf{n}\cdot\mathbf{v})+(\nabla\times\mathbf{u})\cdot(\mathbf{n}\times\mathbf{v})\,\mathrm{d}A. && (2.121)\\ & && (2.122)\end{aligned}$$

Note that the boundary integral contains two terms, which we will use when defining the appropriate space of functions to minimize in. We now seek a variational solution to (2.103). Minimize the energy functional

$$\mathcal{E}'[\mathbf{u}] = \frac{1}{2}(\nabla\cdot\mathbf{u},\nabla\cdot\mathbf{u})+\frac{1}{2}(\nabla\times\mathbf{u},\nabla\times\mathbf{u})-(\mathbf{f},\mathbf{u}), \qquad (2.123)$$

for functions $\mathbf{u}\in\mathcal{T}$ where

$$\mathcal{T} = \{\mathbf{u}\in H^1(\Omega)\ :\ \mathbf{n}\times\mathbf{u}=0\ \text{ for }\mathbf{x}\in\partial\Omega\}. \qquad (2.124)$$

The interesting point here is that $\mathcal{T}$ only contains a Dirichlet boundary condition on the velocity field.

Remark 2.21 *For vector fields* $\mathbf{u},\mathbf{v}\in\mathcal{T}$*, the energy functional* $\mathcal{E}'$ *has the variation*

$$\begin{aligned}\frac{\delta\mathcal{E}}{\delta\mathbf{v}} &= (-\Delta\mathbf{u}-\mathbf{f},\mathbf{v})+\int_{\partial\Omega}(\nabla\cdot\mathbf{u})(\mathbf{n}\cdot\mathbf{v})\,\mathrm{d}A. && (2.125)\\ & && (2.126)\end{aligned}$$

The minimum functional $\mathcal{E}'$ *occurs when the variation vanishes for all* $\mathbf{v}$. *Namely* $-\Delta\mathbf{u}=\mathbf{f}$ *in* Ω *and* $\nabla\cdot\mathbf{u}=0$ *in* $\partial\Omega$.

Therefore, to construct a Galerkin solution:

1. Imagine having a suitable mesh for a domain. Enumerate the mesh elements with *i*.

2. Choose a finite dimensional basis $\mathbf{u}_k^i$ for each element i. For elements on the boundary, choose $\mathbf{u}_k^i = 0$ on $\partial\Omega$.
3. Construct an approximate solution $\mathbf{u} = \sum_i \sum_k c_{k,i}\mathbf{u}_k^i \in \mathcal{T}$ by substituting into $\mathcal{E}'$ and minimizing with respect to each $c_{k,i}$.
4. This yields an approximate solution to problem (2.103).

2.11 General Comments on PPE Formulations

The PPE formulations proposed by Henshaw, and later Johnston and Liu, as well as the one we propose in section 2.3 share some common traits. Most notably, the formulations i) allow for the recovery of the pressure as a function of the velocity field, ii) are defined for extended Navier-Stokes systems, ie. systems which do not necessarily satisfy the required initial data constraint $\nabla\cdot\mathbf{u}_0 = 0$. In such cases where the initial data starts off the Navier-Stokes solution manifold, the equations act as attractors, pulling the solutions back to Navier-Stokes. In this section, we examine properties, i) and ii), that PPE formulations share.

We start by considering the following linearized equations with a no slip boundary condition:

$$\left.\begin{aligned} \mathbf{u}_t - \mu\Delta\mathbf{u} &= -\nabla p + \mathbf{f} && \text{for } \mathbf{x}\in\Omega, \\ \mathbf{u} &= 0 && \text{for } \mathbf{x}\in\partial\Omega, \end{aligned}\right\} \tag{2.127}$$

and

$$\left.\begin{aligned} \Delta p &= \nabla\cdot\mathbf{f} && \text{for } \mathbf{x}\in\Omega, \\ \mathbf{n}\cdot\nabla p &= \mathbf{n}\cdot(\mathbf{f}+\mu\Delta\mathbf{u}) && \text{for } \mathbf{x}\in\partial\Omega. \end{aligned}\right\} \tag{2.128}$$

We assume smooth initial data $\mathbf{u}_0$ satisfying

$$\mathbf{u}_0 = 0 \qquad \text{for } \mathbf{x}\in\partial\Omega, \tag{2.129}$$

$$\int_\Omega (\nabla\cdot\mathbf{u}_0)\,\mathrm{d}V = 0, \tag{2.130}$$

$$\int_{\partial\Omega} \mathbf{n}\cdot(\Delta\mathbf{u}_0)\,\mathrm{d}A = 0. \tag{2.131}$$

$$\tag{2.132}$$

Here the second line, via the divergence theorem, follows as the result of the velocity boundary condition. The third line is a solvability condition on the pressure (which is automatically satisfied for velocity fields with $\nabla \cdot \mathbf{u}_0 = 0$). Note however that these conditions are more general than those required for the original Navier-Stokes, namely $\nabla \cdot \mathbf{u}_0 = 0$ may not be satisfied at every point in space.

Even when the initial data $\mathbf{u}_0$ is divergence free, these equations are not equivalent to Navier-Stokes. In fact, they are ill-posed in the sense that solutions are not unique. Specifically, the equations form a linear system whose null space is equivalent to the space of harmonic functions[13]. We may construct the null space of equations (2.127–2.128) as follows. Suppose the functions $(\mathbf{u}, p)$ solve equations (2.127–2.128) for initial data $\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}_0(\mathbf{x})$. Then construct a new solution $(\mathbf{u} + \mathbf{v}, p + q)$ with functions $(\mathbf{v}, q)$ as follows. Take any boundary valued function $h(\mathbf{x}, t)$ for $\mathbf{x} \in \partial\Omega$ such that:

$$\int_{\partial\Omega} h(\mathbf{x}, t)\, \mathrm{d}A = 0, \tag{2.133}$$

and define q as

$$\begin{aligned} \Delta q &= 0 \quad \text{for } \mathbf{x} \in \Omega, & (2.134) \\ \mathbf{n} \cdot \nabla q &= h \quad \text{for } \mathbf{x} \in \partial\Omega. & (2.135) \end{aligned}$$

For instance q is any harmonic function with zero mean. Once we have q, we can now define $\mathbf{v}$ as the unique solution to the following heat equations:

$$\begin{aligned} \mathbf{v}_t - \mu \Delta \mathbf{v} &= -\nabla q \quad \text{for } \mathbf{x} \in \Omega, & (2.136) \\ \mathbf{v} &= 0 \quad \text{for } \mathbf{x} \in \partial\Omega. & (2.137) \end{aligned}$$

By construction, the pair $(\mathbf{v}, q)$ solve (2.127–2.128) with $\mathbf{f} = 0$. Hence, every zero mean harmonic function can be mapped to a unique null solution of (2.127–2.128). Note that the

---

[13] The pressure may always vary by a constant. We consider two pressure solutions equivalent if they differ by a constant.

previous construction contains the entire null space. For instance, given a solution $(\mathbf{v}, q)$, to (2.127–2.128) with $\mathbf{f} = 0$, then q is already a harmonic function.

The system above has a very large null space, and hence only contains the Navier-Stokes equations as solutions. In fact, these null space functions result in velocity fields with a nonzero divergence. In other words, the system (2.127–2.128) fails to control the divergence of the velocity field. The idea behind the various PPE formulations is to alter either the boundary conditions, or right hand side of equations (2.127–2.128) so that:

1. the velocity evolution becomes unique (ie. the kernel of the linear system (2.127–2.128) is trivial),
2. allow for the explicit recovery of the pressure from the velocity,
3. velocity fields starting on the $\nabla \cdot \mathbf{u} = 0$ surface remain constrained their for all time,
4. the $\nabla \cdot \mathbf{u} = 0$ surface is an attractor in the larger extended Navier-Stokes system.

In [58] Johnston and Liu propose several ways to modify the boundary conditions of equations (2.127–2.128) so that they satisfy items 1, 2 and 3 in the list above. Only one method satisfies item 4. The following is a brief description of these methods. For brevity, as in the previous sections of this chapter, let $\phi = \nabla \cdot \mathbf{u}$.

1. Alter the right hand side of the Poisson equation (2.128) to

$$\Delta p = \nabla \cdot \mathbf{f} + \mu \Delta (\nabla \cdot \mathbf{u}). \tag{2.138}$$

   The divergence field ϕ then satisfies $\phi_t = 0$. Hence, if $\phi = 0$ at $t = 0$, then $\phi = 0$ for all time. Navier-Stokes are not an attractor in this system. In fact, any initial condition with $\nabla \cdot \mathbf{u}_0 \neq 0$ will preserve the nonzero divergence for all time.

2. Alter the right hand side of the momentum equation (2.127) to

$$\mathbf{u}_t - \mu \nabla \times \nabla \times \mathbf{u} = -\nabla p + \mathbf{f}. \tag{2.139}$$

Using the identity $\nabla \times \nabla \times = \nabla(\nabla \cdot) - \Delta$, the divergence field ϕ then satisfies $\phi_t = 0$. Hence, if $\phi = 0$ at $t = 0$, then $\phi = 0$ for all time. In analogy with the previous method, this system does not contain Navier-Stokes as an attractor.

3. Alter the boundary conditions of the Poisson equation (2.128) to

$$\mathbf{n} \cdot \nabla p = \mathbf{n} \cdot (\mathbf{f} - \nabla \times \nabla \times \mathbf{u}) . \tag{2.140}$$

The divergence then satisfies

$$\begin{aligned} \phi_t &= \mu \Delta \phi \quad \text{for } \mathbf{x} \in \Omega, && (2.141) \\ \mathbf{n} \cdot \nabla \phi &= 0 \qquad \text{for } \mathbf{x} \in \partial\Omega. && (2.142) \end{aligned}$$

Hence, if $\phi = 0$ at $t = 0$, then $\phi = 0$ for all time. Moreover, one can even relax the constrain (2.131) in the initial conditions (ie. these equations are valid for a larger class of functions).

This system does contain Navier-Stokes as a global attractor. Specifically, all solutions with the extended Navier-Stokes initial conditions (2.129) converge to Navier-Stokes solutions. To show that initial conditions converge to ones with $\phi = 0 \; \forall \mathbf{x} \in \Omega$, first note that solutions to the heat equation (2.141) all converge to constant, steady state values as $t \to \infty$. Hence $\phi(\mathbf{x}) \to C$. To show that the constant $C = 0$ we observe that the heat equation preserves the average *temperature*:

$$\begin{aligned} \frac{d}{dt} \int_\Omega \phi &= \mu \int_\Omega \Delta \phi \, \mathrm{d}V && (2.143) \\ &= \mu \int_{\partial\Omega} (\mathbf{n} \cdot \nabla \phi) \, \mathrm{d}A && (2.144) \\ &= 0. && (2.145) \end{aligned}$$

Since $\int_\Omega \phi \, \mathrm{d}V = 0$ at time $t = 0$, we obtain $C = 0$ so that $\phi(\mathbf{x}) \to 0$ as $t \to \infty$.

Remark 2.22 *The* $\nabla \times \nabla \times \mathbf{u}$ *boundary condition in method 3 is not unique. For instance,* *equivalently, one can think of the pressure boundary condition as:*

$$\mathbf{n}\cdot\nabla p = \mathbf{n}\cdot(\mathbf{f}+\mu\Delta\mathbf{u}) + h, \tag{2.146}$$

where $h = -\mu\mathbf{n}\cdot\nabla\phi$. The correction h is only one possible choice. In fact, one may achieve the same results using $h = \lambda\mathbf{n}\cdot\nabla\phi$ for any λ. In this case, the solvability condition for the pressure is still preserved, however requires the constraint (2.131).

2.11.1 Generalizations

In the most general form (with a $\mathbf{u} = 0$ boundary condition for the momentum equation) the PPE equations have the form:

$$\left.\begin{array}{rcll} \mathbf{u}_t - \mu\Delta\mathbf{u} & = & -\nabla p + \mathbf{f} + \alpha(\mathbf{u}) & \text{for } \mathbf{x}\in\ \Omega, \\ \mathbf{u} & = & 0 & \text{for } \mathbf{x}\in\partial\Omega, \end{array}\right\} \tag{2.147}$$

and

$$\left.\begin{array}{rcll} \Delta p & = & \nabla\cdot\mathbf{f} + \beta(\mathbf{u}) & \text{for } \mathbf{x}\in\ \Omega, \\ \mathbf{n}\cdot\nabla p & = & \mathbf{n}\cdot(\mathbf{f}+\mu\Delta\mathbf{u}) + \gamma(\mathbf{u}) & \text{for } \mathbf{x}\in\partial\Omega. \end{array}\right\} \tag{2.148}$$

where $\alpha(\mathbf{u})$, $\beta(\mathbf{u})$ and $\gamma(\mathbf{u})$ are functions of the velocity field. The functions are not completely independent. For instance, through the solution of a Poisson equation, nonzero β and γ can be thought of as an equivalent α. These functions do satisfy the constraints:

1. For the system (2.147–2.148) to contain Navier-Stokes, we require $\alpha = \beta = \gamma = 0$ when $\phi = \nabla\cdot\mathbf{u} = 0$.

2. Poisson solvability criteria $\int_{\partial\Omega}\mathbf{n}\cdot\gamma\,\mathrm{d}A = \int_\Omega\beta\,\mathrm{d}V$.

These constraints are necessary, but not sufficient. For instance, the well-posedness of the system (so that solutions are unique) places additional restrictions on the functions.

2.12 Summary

Through the introduction of the pressure Poisson equation with consistent boundary conditions, we give an equivalent formulation of the incompressible Navier-Stokes equations. In this formulation the momentum equation takes the form of a vector heat equation with unconventional boundary conditions, while the pressure Poisson equation can be used to explicitly describe the pressure as a function of velocity at any fixed time. It follows that the reformulated system of equations is ideal for using efficient numerical marching methods, where there are no particular theoretical limitations to the order of accuracy or the method of implementation. In addition, we devise and implement a second order discretization (uniform up to the boundary) of the equations on irregular domains. We address the issue of numerical stability for the normal boundary velocity by adding an appropriate feedback term to the equations.

There are several issues and extensions that we hope to address. First, for irregular domains on a regular grid, the proposed momentum equation boundary conditions implicitly couple all the components of the velocity field. Therefore, the resulting vector heat operator, $\partial_t - \mu\,\Delta$, cannot be solved "component by component". This makes the implementation of schemes that do a naive implicit treatment of the stiff viscosity term computationally expensive, is there a better way? If so, then can a simple implicit scheme be extended to achieve high order in time? Second, our PPE formulation is for fixed domains, but extensions to deformable and or moving geometries seem possible. Third, can the ideas here be extended to flows with variable densities (e.g.: stratified flows)?

Chapter 3

Thermalization of the Klein-Gordon Equation

3.1 An Introduction to Thermalization

In the present chapter we discuss the thermalization of the Klein-Gordon (KG) equation due to a nonlinear (quadratic) interaction. Here by thermalization, we essentially seek to describe the evolution of generic, out of equilibrium, initial wave fields. Although the results and discussion in this chapter focus on characterizing such long time wave behavior, the problem is also physically motivated as follows. Consider some physical system described by the Klein-Gordon equation, such as a finite length string on a nonlinear[1] elastic bed. As an isolated, energy conserving system, what is the long time behavior of an initial disturbance? For instance, if the wave field is confined to a fixed domain in space, does the nonlinear interaction result in a long time behavior with general trends and features? In short, the answer is yes – the wave field thermalizes into a local thermodynamic equilibrium (LTE) [94].

In particular, we numerically show that in the presence of strong nonlinearities, the local thermodynamic equilibrium state exhibits a weakly nonlinear behavior in a renormalized wave basis. Starting with the Klein-Gordon equation, section 3.2 introduces the notion of a renormalized wave basis as a linear transformation of the wave field Fourier modes and the requirement of vanishing wave-wave correlations. In section 3.3, we discuss properties of the renormalized waves and show that they oscillate around a single frequency. The section also addresses several characteristics of the renormalized waves, most notably the nonlinear frequency dispersion relation and energy spectrum. Specifically, the renormalized waves exhibit a Klein-Gordon dispersion relation with a nonlinear shift in the mass (or $k = 0$ Fourier mode frequency), as well as a Planck-like energy spectrum. Namely, there is equipartition of energy in the low-frequency modes described by a Boltzmann distribution, followed by a linear exponential decay in the high-frequency modes.

---

[1] As described in section 3.2, the nonlinearity here is a positive contribution to the potential.

3.2 The Klein-Gordon Equation

In this section, we introduce the classical u^4 Klein-Gordon (KG) field[2]. In 1+1 dimensions, on the fixed domain $x \in [0, 2\pi]$, the KG equation admits a Hamiltonian structure, where the Hamiltonian is given by[3]

$$H = \int_0^{2\pi} \frac{1}{2}\left[(\partial_t u)^2 + (\partial_x u)^2 + u^2\right] + \frac{\lambda}{4} u^4 \, \mathrm{d}x, \tag{3.1}$$

Letting $p = \partial_t u$ so that $H = H[p, u]$ is a functional of the fields p and u, one obtains the KG equation via, $\partial_t p = -\frac{\delta H}{\delta u}, \partial_t u = \frac{\delta H}{\delta p}$:

$$\partial_t^2 u = \partial_x^2 u - u - \lambda u^3. \tag{3.2}$$

In addition to the conservation of total energy H, periodic boundary conditions $u(0,t) = u(2\pi, t)$ result in a second integral of motion, the total momentum

$$P = -\int_0^{2\pi} (\partial_x u)(\partial_t u) \, \mathrm{d}x, \tag{3.3}$$

In this chapter, we focus explicitly on the case where the potential $V(u) = \frac{1}{2}u^2 + \frac{\lambda}{4}u^4$ has a coupling constant $\lambda > 0$. This is in stark contrast to chapter 4 where we examine potentials with negative nonlinear contributions to the energy. In such cases where $\lambda < 0$, waves may spatially trap energy through the development of coherent structures [37]. As a result, the resulting thermalization discussed in this chapter is a direct consequence of the positive coupling.

In our work we study the LTE solutions for a wide variety of initial wave configurations. Hence, we evolve each solution from a fixed, out of equilibrium initial condition and do not consider ensemble averages. Instead, all appropriate thermal quantities are obtained by

---

[2] We consider real valued fields u.

[3] For brevity, $\partial_t u = \frac{\partial u}{\partial t}$, $\partial_x u = \frac{\partial u}{\partial x}$.

averaging over time. For instance

$$\langle A\rangle(t) = \frac{1}{\Delta T}\int_{t-\Delta T/2}^{t+\Delta T/2} A(t')\,\mathrm{d}t', \tag{3.4}$$

where ΔT is the averaging interval. The quantity $\langle A\rangle$, however, is not fixed for all times, but exhibits characteristics of an LTE and (for a fixed averaging interval) may drift in time.

Lastly, in the following subsections, we make repeated use of the waves square mean field $S(t)$, as it appears in the nonlinear Klein-Gordon dispersion relation:

$$S(t) = \frac{1}{2\pi}\int_0^{2\pi} u^2(x,t)\,\mathrm{d}x. \tag{3.5}$$

3.2.1 The linear case

Although we will ultimately focus on the study of equation (3.2) for strongly interacting waves, to motivate the introduction of renormalized waves, we first address (3.2) in the linear case, followed by the case of weakly interacting waves. In the absence of nonlinearity, $\lambda = 0$, (3.2) reduces to the linear KG equation. Hence, wave solutions decouple into linearly independent Fourier modes, with conservation of energy in each mode. Specifically, the Fourier series for $p(x,t)$ and $u(x,t)$ are

$$\begin{aligned} p(x,t) &= \frac{1}{\sqrt{2\pi}}\sum_k p_k(t)e^{-\imath kx}, \\ u(x,t) &= \frac{1}{\sqrt{2\pi}}\sum_k u_k(t)e^{-\imath kx} \end{aligned}$$

with summation k over all integer values. Taking $p(x,t)$ and $u(x,t)$ as real fields restricts $u_k = u^*_{-k}$ and $p_k = p^*_{-k}$ so that H becomes

$$H = \sum_k E_k, \tag{3.6}$$

$$E_k = \frac{1}{2}(|p_k|^2 + \omega_k^2|u_k|^2). \tag{3.7}$$

Here E_k is the linear energy in mode k, while $\omega_k^2 = 1 + k^2$ is the linear dispersion relation. In addition to constructing Fourier mode solutions, one may also make a secondary transformation to study the kinetics of interacting waves

$$a_k = \frac{p_k - \imath\omega_k u_k}{\sqrt{2\omega_k}}, \tag{3.8}$$

In this new basis, the equations of motion become

$$\imath\dot{a}_k = \frac{\partial}{\partial a_k^*} H(a_k, a_k^*), \tag{3.9}$$

where the Hamiltonian, $H(a_k, a_k^*)$ also acquires a convenient form:

$$H = \sum_k \omega_k |a_k|^2. \tag{3.10}$$

With the dynamic equations written in the variables a_k, the linear oscillator solutions take the form $a_k(t) = A_k e^{-\imath\omega_k t}$, where the amplitude and phase are contained in the complex variable A_k. As a result, each wave number oscillates with a negative frequency so that the Fourier transform of a linear wave is a Dirac δ function

$$\begin{aligned} \widehat{a}_k(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} a_k(t) e^{-\imath\omega t}\, \mathrm{d}t & (3.11) \\ &= A_k \delta(\omega_k + \omega). & (3.12) \end{aligned}$$

Secondly, the infinite time correlation of any two linear waves, including the case of $k = \pm l$, vanish:

$$\langle a_k a_l \rangle = 0. \tag{3.13}$$

Hence, linear solutions are uncorrelated waves, which oscillate at a single frequency.

3.2.2 Weakly nonlinear renormalized waves

In the presence of nonlinearity, $\lambda > 0$, the waves a_k no longer decouple into linear oscillators and therefore no longer conserve the energy E_k. For instance, the wave amplitudes $|a_k(t)|$ do not remain constant but fluctuate from the nonlinear interaction. As a result, we seek to characterize the nonlinear effect on both the amplitude and phase behavior of the waves a_k. Before proceeding to the case of strongly interacting waves, we first examine the case of a weak nonlinearity. By weak nonlinearity we mean the maximum amplitude of u is order 1, while the coupling strength remains small $\lambda \ll 1$. Moreover, with the onset of nonlinearity, especially in the presence of strong interactions, the waves a_k no longer exhibit the linear properties (3.12) and (3.13). We therefore follow the ideas of Gershgorin *et al.* [34] and introduce renormalized waves c_k in lieu of (3.8). Renormalized waves form a useful basis since they exhibit characteristics of linear waves. Namely, renormalized waves have vanishing correlators and approximately oscillate with a single frequency. In analogy with the linear waves (3.8), renormalized waves are defined through a linear combination of the Fourier transforms u_k and p_k, however, linear frequencies are altered to renormalized frequencies:

$$c_k = \frac{1}{\sqrt{2\tilde{\omega}_k}}(p_k - \imath\tilde{\omega}_k u_k). \tag{3.14}$$

The transformation to renormalized waves is canonical (up to a factor of $\imath$) provided the yet to be determined frequencies satisfy $\tilde{\omega}_k > 0$ and $\tilde{\omega}_k = \tilde{\omega}_{-k}$.

We now examine the evolution of renormalized waves in the presence of a weak nonlinearity $\lambda \ll 1$ through a perturbation expansion. Specifically, we show that the standard procedure of choosing renormalized frequencies to eliminate resonant terms yields the same result as choosing frequencies to enforce a vanishing correlator for waves at $\pm k$ wave numbers, i.e., $\langle c_k c_{-k}\rangle$. Written in terms of renormalized waves, the dynamic equations for c_k

are

$$
\begin{aligned}
\imath \dot{c}_k &= \frac{\partial H(c_k, c_k^*)}{\partial c_k^*}, && (3.15)\\
\imath \dot{c}_k &= \frac{\tilde{\omega}_k}{2}\left[(1+\frac{\omega_k^2}{\tilde{\omega}_k^2})c_k + (1-\frac{\omega_k^2}{\tilde{\omega}_k^2})c_{-k}^*\right] && (3.16)\\
&+ \lambda \sum_{l,m,n} T_{klmn}\Big[4c_l^* c_m^* c_n^* \delta_{k+l+m+n} \\
&- 12 c_l c_m^* c_n^* \delta_{l-k-m-n} - 4 c_l c_m c_n \delta_{k-l-m-n} \\
&+ 12 c_n c_m c_l^* \delta_{l+k-m-n}\Big], \\
T_{klmn} &= \frac{1}{32\pi(\tilde{\omega}_k \tilde{\omega}_l \tilde{\omega}_m \tilde{\omega}_n)^{\frac{1}{2}}},
\end{aligned}
$$

where T_{klmn} is the series coefficient, and $\delta_x = 1$ when $x = 0$ and $\delta_x = 0$ otherwise. The summation is over all integers l, m, n. As is commonly the case in kinetic theories, wave behavior is dominated by interacting resonant terms, for instance, those which force the order-one linear oscillators c_k on resonance. The nonlinear term in (3.15) admits only one such resonance, which comes from the term $c_n c_m c_l^* = |c_n|^2 c_k$ when $k = n, m = l$ or $k = m, n = l$. The presence of only one resonant term follows from the fact that the linear dispersion relation $\omega_k^2 = 1 + k^2$ is concave up [106]. For instance, the harmonics generated by the products $c_l^* c_m^* c_n^*$, $c_l c_m^* c_n^*$, and $c_l c_m c_n$ cannot oscillate with frequency $\tilde{\omega}_k$ provided $\tilde{\omega}_k$ is a concave function of k.

In addition to each resonant term in the series $c_n c_m c_l^*$, there is an equivalent term in $c_l c_m^* c_n^*$ containing the variable c_{-k}^* when $n = -k, l = m$ or $m = -k, l = n$. We now remove both the resonant terms and their equivalent c_{-k}^* terms from the nonlinear series, and absorb them into the coefficients of c_k and c_{-k}^*. We do this so that upon seeking a small amplitude ($\lambda \ll 1$) solution, we may simultaneous handle both the zeroth, $O(1)$, and first order, $O(\lambda)$, corrections to the renormalized frequencies. The dynamic equations then become:

$$\begin{aligned}
\imath\dot{c}_k &= g_k(t)c_k + h_k(t)c^*_{-k} + \lambda r_k(t), \\
g_k(t) &= \frac{\tilde{\omega}_k}{2}\Big(1+\frac{\omega_k^2}{\tilde{\omega}_k^2}\Big) + \lambda\mu_k(t), \\
h_k(t) &= \frac{\tilde{\omega}_k}{2}\Big(1-\frac{\omega_k^2}{\tilde{\omega}_k^2}\Big) - \lambda\mu_k(t), \\
\mu_k(t) &= 24\sum_m T_{mmkk}|c_m|^2 - 12T_{kkkk}|c_k|^2, \\
r_k(t) &= \sum_{l,m,n}{}' T_{klmn}\Big[4c_l^*c_m^*c_n^*\delta_{k+l+m+n} \\
&- 12c_lc_m^*c_n^*\delta_{l-k-m-n} - 4c_lc_mc_n\delta_{k-l-m-n} \\
&+ 12c_nc_mc_l^*\delta_{l+k-m-n}\Big].
\end{aligned} \tag{3.17}$$

Here the prime in r_k restricts the summation to exclude all resonant terms having either c_k or their equivalent terms containing c^*_{-k}. Meanwhile, the time dependent coefficients $g(t)$ and $h(t)$ are related by

$$g_k(t) + h_k(t) = \tilde{\omega}_k. \tag{3.18}$$

Thus far, equation (3.17) is exact.

We now extract approximate solutions to (3.17), in the limit $\lambda \ll 1$, for the initial value problem:

$$c_k(0) = C_k. \tag{3.19}$$

To remain consistent in the small amplitude approximation, we assume the initial field is small with finite energy. Hence, each $C_k < 1$ while $C_k \to 0$ as $k \to \infty$. To obtain a solution, we seek an asymptotic series for c_k in powers of λ with a single frequency leading term:

$$c_k(t) = C_ke^{-\imath\tilde{\omega}_kt} + \lambda c_k^{(1)}(t) + \lambda^2c_k^{(2)}(t) + \ldots. \tag{3.20}$$

In general, even at zeroth order in λ, an arbitrary choice of $\tilde{\omega}_k$ will couple solutions $c_{\pm k}$ together through the coefficients $g(t)$ and $h(t)$. As a result, the proposed ansatz (3.20) requires the following two consistency conditions on $\tilde{\omega}_k$:

$$g_k = \tilde{\omega}_k, \tag{3.21}$$

$$h_k = 0. \tag{3.22}$$

The first consistency condition (3.21) comes from the fact that we have chosen c_k to oscillate with frequency $\tilde{\omega}_k$. Meanwhile, the second condition (3.22) follows from the fact that our ansatz has no dependence on the initial value C_{-k}, or equivalently that c_k decouples from c_{-k}. Relation (3.18) however, implies the equivalence of the two conditions (3.21), (3.22). Therefore, choosing $\tilde{\omega}$ through (3.21) as the c_k oscillator single frequency automatically guarantees (3.22), thereby decoupling the fields c_k and c^*_{-k}. Using equation (3.21) to extract the expression for $\tilde{\omega}_k$ yields

$$\tilde{\omega}_k = \frac{\tilde{\omega}_k}{2}\left(1+\frac{\omega_k^2}{\tilde{\omega}_k^2}\right) + \lambda\mu_k^{(0)} + O(\lambda^2), \tag{3.23}$$

$$\mu_k^{(0)} = 24\sum_m T_{mmkk}|C_m|^2 - 12T_{kkkk}|C_k|^2. \tag{3.24}$$

To solve for $\tilde{\omega}_k$ to $O(\lambda^2)$, we assume the amplitude in a single mode is much smaller than the total averaged wave field. Therefore, neglecting the term $T_{kkkk}|C_k|^2$ results in

$$\tilde{\omega}_k^2 = \omega_k^2 + 3\lambda\langle S\rangle + O(\lambda^2) \tag{3.25}$$

$$= 1 + 3\lambda\langle S\rangle + k^2 + O(\lambda^2). \tag{3.26}$$

Here we have made use of the fact that μ_k^0 is proportional to the time-averaged value of the wave field

$$\begin{aligned} \langle S(t)\rangle &= \langle\frac{1}{2\pi}\int_0^{2\pi} u^2(x,t)\,\mathrm{d}x\rangle & (3.27)\\ &= \sum_m \frac{1}{4\pi\tilde{\omega}_m}(|C_{-m}|^2+|C_m|^2). & (3.28)\end{aligned}$$

Note that within the framework of the small amplitude ansatz, the approximation (3.25) only remains valid over time scales $O(\lambda^{-2})$. As a result, to remain consistent, the time average taken in (3.27) should be made over, at most, comparable time scales, $O(\lambda^{-2})$.

With the values of $\tilde{\omega}_k$ solved to $O(\lambda)$, we may substitute $c_k^{(0)} = C_k e^{-\imath\tilde{\omega}_k t}$ into the nonlinear term $r(t)$ and solve for the first-order contribution to $c_k(t)$. Consequently, $c_k^{(1)}(t)$ satisfies the equation

$$\begin{aligned} \imath\dot{c}_k^{(1)} &= \tilde{\omega}_k c_k^{(1)} + r_k^{(0)}(t), & (3.29)\\ c_k^{(1)}(0) &= 0. & (3.30)\end{aligned}$$

Here $r_k^{(0)}(t)$ is the nonlinear term $r(t)$ evaluated using $c_k^{(0)} = C_k e^{-\imath\tilde{\omega}_k t}$. Equation (3.29) therefore describes a first-order linear oscillator forced off resonance by $r_k^{(0)}(t)$. The solution for $c_k^{(1)}(t)$ is

$$\begin{aligned} c_k^{(1)}(t) &= -{\sum_{l,m,n}}' T_{klmn}\Big[\frac{4C_l^*C_m^*C_n^*e^{\imath(\tilde{\omega}_l+\tilde{\omega}_m+\tilde{\omega}_n)t}}{\tilde{\omega}_k+\tilde{\omega}_l+\tilde{\omega}_m+\tilde{\omega}_n}\delta_{k+l+m+n}\\ &- \frac{12C_lC_m^*C_n^*e^{\imath(-\tilde{\omega}_l+\tilde{\omega}_m+\tilde{\omega}_n)t}}{\tilde{\omega}_k-\tilde{\omega}_l+\tilde{\omega}_m+\tilde{\omega}_n}\delta_{l-k-m-n}\\ &- \frac{4C_lC_mC_ne^{-\imath(\tilde{\omega}_l+\tilde{\omega}_m+\tilde{\omega}_n)t}}{\tilde{\omega}_k-\tilde{\omega}_l-\tilde{\omega}_m-\tilde{\omega}_n}\delta_{k-l-m-n}\\ &+ \frac{12C_nC_mC_l^*e^{\imath(\tilde{\omega}_l-\tilde{\omega}_m-\tilde{\omega}_n)t}}{\tilde{\omega}_k+\tilde{\omega}_l-\tilde{\omega}_m-\tilde{\omega}_n}\delta_{l+k-m-n}\Big]\\ &+ De^{-\imath\tilde{\omega}_k t}. & (3.31)\end{aligned}$$

In the solution for $c_k^{(1)}$, the prime in the summation indicates that there are no terms oscillating with frequencies $\pm\tilde{\omega}_k$. This follows from the fact that we have removed the terms in $r(t)$ oscillating at frequencies $\pm\tilde{\omega}_k$. Meanwhile, the constant D, from the homogeneous term, is chosen to satisfy the initial value $c_k^{(1)}(0) = 0$. With the asymptotic solution (3.20) solved to $O(\lambda^2)$, we may calculate the time averaged correlator, $\langle c_k c_{-k} \rangle$, for waves $\pm k$:

$$\langle c_k c_{-k} \rangle = C_k C_{-k} \langle e^{-\imath 2\tilde{\omega}_k t} \rangle + \lambda C_k \langle e^{-\imath\tilde{\omega}_k t} c_{-k}^{(1)} \rangle \tag{3.32}$$

$$+ \lambda C_{-k} \langle e^{-\imath\tilde{\omega}_k t} c_k^{(1)} \rangle + O(\lambda^2). \tag{3.33}$$

Using formula (3.4) with an averaging time of $O(\lambda^2)$ will result in $\langle e^{-\imath 2\tilde{\omega}_k t} \rangle = O(\lambda^2)$. Meanwhile, the long time average $\langle e^{-\imath\tilde{\omega}_k t} c_{-k}^{(1)} \rangle$ picks out the component of $c_{-k}^{(1)}$ with frequency $\tilde{\omega}_k$. Since $c_{\pm k}^{(1)}(t)$ contains no term oscillating at frequency $\tilde{\omega}_k$, taking a time average over length $O(\lambda^{-2})$ will also tend this term to zero:

$$\langle c_k c_{-k} \rangle = O(\lambda^2). \tag{3.34}$$

Therefore, to first order in λ, choosing the renormalized frequencies $\tilde{\omega}_k$ to remove the resonant terms in (3.15) is equivalent to choosing them to enforce a vanishing correlator.

3.2.3 Strongly nonlinear waves

In studying LTE solutions, we are specifically interested in the case of strongly interacting waves, $\lambda \gg 1$. Even in the presence of a strong nonlinearity, one may introduce renormalized waves defined by (3.14). Extending the properties of linear waves, the new renormalized frequencies $\tilde{\omega}$ are then found by imposing that waves c_k and c_{-k} remain uncoupled and their correlations vanish [34, 35]:

$$\langle c_k c_{-k} \rangle = 0. \tag{3.35}$$

If one has a KG solution $u(x,t)$ for some long time interval, substituting the definition of c_k into (3.35) yields an expression for $\tilde{\omega}_k$:

$$\tilde{\omega}_k^2 = \frac{\langle |p_k|^2 \rangle}{\langle |u_k|^2 \rangle}. \tag{3.36}$$

Equation (3.36) provides a useful formula for numerically obtaining the renormalized frequencies. Despite the fact that the renormalized transformation defined by equations (3.14) and (3.36) is always mathematically possible, there is no guarantee that the waves c_k stand out as a useful coordinate system. The waves c_k do however form a useful basis provided the Fourier modes u_k exhibit narrow band single frequency oscillations. When u_k does oscillate with roughly one frequency, the formula (3.36) identifies $\tilde{\omega}_k$ with the frequency of oscillation. In the case of the KG equation with a strong u^4 interaction, we do in fact find that the waves c_k form narrow band oscillators with a central frequency $\tilde{\omega}_k$. We verify these wave properties in our numerical experiments, implying renormalized waves are a natural basis to study the LTE spectra. We demonstrate in our numerical experiments that in the limit $\lambda \gg 1$, the renormalized frequencies are very well approximated by the dispersion relation

$$\tilde{\omega}_k^2 \approx 1 + 2.59\lambda\langle S \rangle + k^2. \tag{3.37}$$

Here we refer to $2.59\lambda\langle S \rangle = \tilde{\omega}_k^2 - \omega_k^2$ as the mass shift.

In the case of large λ, the nonlinear time scale may be estimated by comparing the size of terms in the KG equation. Specifically, assuming the field $u(x,t)$ is order one, the time derivative will balance the nonlinear term, $u_{tt} \sim \lambda u^3$, provided that the characteristic nonlinear timescale $\tau_0 \sim \lambda^{-1/2}$. Note that this estimate only holds for large λ, or when λ dominates the linear dispersive term in the KG equation. In our numeric experiments, we find that LTE solutions exhibit $O(\lambda^{-1/2})$ as the natural fundamental period, while earlier work focusing on the analytic development of short time KG solutions [48] further verifies $\lambda^{-1/2}$ as the strong amplitude time scale. Consequently, we define LTE time averaged quantities

$\langle A \rangle$ over many periods of the time scale τ_0:

$$\langle A \rangle(t) = \lim_{\Delta T \gg \tau_0} \frac{1}{\Delta T} \int_{t-\Delta T/2}^{t+\Delta T/2} A(t')\,\mathrm{d}t'. \tag{3.38}$$

Our numeric experiments try to capture solution properties at large wave numbers, especially within the exponential decay of the spectrum. We therefore use a pseudospectral method to maintain high numeric accuracy in both space and time. The pseudospectral method consists of exact spectral propagation for linear terms, coupled with a Richardson extrapolation [82] algorithm for the propagation of the nonlinear terms. For smooth solutions, u, the spectral basis ensures high spatial accuracy, while the Richardson extrapolation guarantees high accuracy in time. Section 3.6 contains a more detailed discussion of the numerical algorithm.

3.3 The Local Thermodynamic Equilibrium

3.3.1 Equilibrium properties

In our numeric experiments, we study LTE solutions for a wide range of initial conditions. Specifically, we focus on initial conditions set over a low band of Fourier modes, and vary both the number and amplitudes of modes excited. In analogy with experiments on the Fermi-Pasta-Ulam (FPU) model [12, 33, 11], we find that there is a generic transient stage where energy in the initial modes is redistributed to higher ones. For large nonlinearities, $\lambda \gg 1$, the transient stage appears to occur on a timescale at least as fast as $\tau_0 \sim \lambda^{-1/2}$. Upon redistribution, the energy spectrum settles into an LTE, where a majority of the energy is shared within a finite band of lower modes. Here the LTE persists for time scales much longer than τ_0 and appears similar in nature to the intermediate metastable state realized by similar experiments [12, 33, 11] in the FPU lattice.

In a recent work concerning the FPU lattice, [11] the authors show that initial conditions, specifically those with Fourier components having random phases versus coherent phases, can alter the long time behavior of the solutions. Although most of our test cases under consideration have initial conditions with coherent phases, we also ran experiments with uniformly distributed random phases. We found that the effect of the phases did not effect the onset, or characteristics of the LTE.

Since the characteristics of a local thermodynamic equilibrium appear to be robust for the various initial conditions that we test, we first focus on describing in detail the generic long time behavior of one solution $u(x,t)$ to equation (3.2), and defer a description of the general trends for the following subsection.

Rescaling the field $u(x,t) \rightarrow \lambda^{-\frac{1}{2}}u(x,t)$ effectively sets the coupling constant to unity: $u_{tt} = u_{xx} - u - u^3$. Hence, without loss of generality, we take $\lambda = 1$ and control the coupling strength through the initial Fourier amplitudes. For the test case under consideration, we naturally initialize $u(x,0)$ via Fourier components [4] to

$$\dot{u}_k = 0, \tag{3.39}$$

$$u_k = \begin{cases} 2.08 & 0 < |k| \leq 50 \\ 0 & k = 0,\ |k| > 50 \end{cases}$$

Although each u_k is $O(1)$, the initial mean squared field amplitude is in fact strong: $S = 68.7$. The total energy and momentum are $H \sim 6.7 \times 10^5$ and $P = 0$.

We obtain the renormalized waves and study the KG thermalization in two steps. First, we evolve the field $u(x,t)$ for a long period of time, i.e., from $T = 0$ to T_1. We choose T_1 long enough so that the field amplitudes $|u_k|^2$ complete the initial transient stage and redistribute their energy into higher modes. More specifically, following [12, 33, 11] one may introduce $M(t)$ as the number of modes which share most (i.e., over 90%) of the energy. During the transient stage, the value of $M(t)$ rapidly, over a time scale at least as fast as $\lambda^{-1/2}$, increases to a plateau. We use the onset of a plateau region in $M(t)$, and hence the stable sharing of

---

[4] The somewhat odd numerical values in the initial conditions found in equation (3.39), or even table (3.1), arise as a consequence of normalization factors in the numerical fast Fourier transform. For instance, we used simple initial values in the numerical codes, however, converting these initial values with the definition (3.6), introduce extra numerical factors such as $\sqrt{2\pi}$, resulting in more obscure values (ie. $u_k = 2.08$).

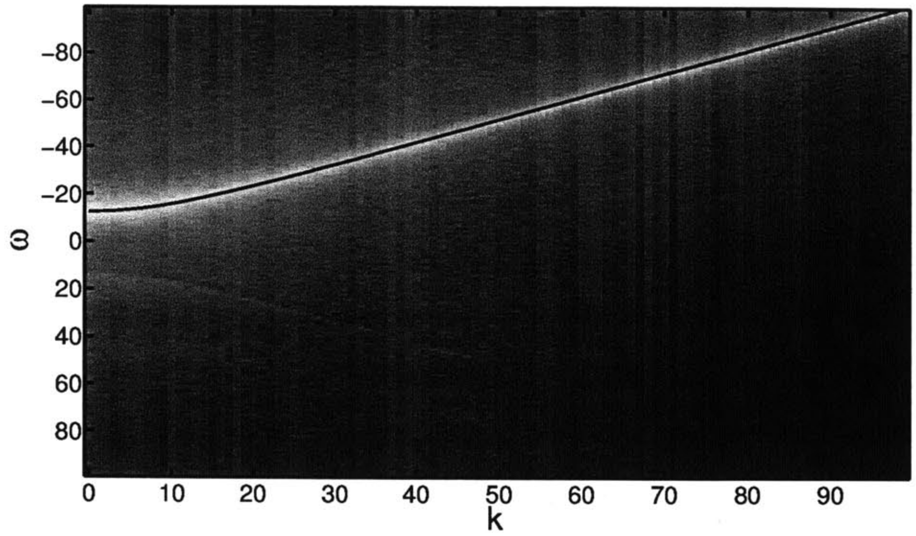

Figure 3-1: Spatiotemporal spectrum of renormalized waves taken over the interval (153, 230) for initial data (3.39). The shaded plot shows $|\widehat{c}_k(\omega)|^2$ while solid line corresponds to the dispersion relation defined by (3.36).

energy between a finite band of low wave number modes, as an indication for an LTE.

Once the field has thermalized, we then further evolve the modes u_k and p_k over some time interval (T_1, T_2) to obtain a description of the thermodynamic equilibrium state. Specifically, we recast the fields u_k and p_k into renormalized waves by obtaining the appropriate time averaged quantities, $\langle|u_k|^2\rangle$ and $\langle|p_k|^2\rangle$, followed by $\tilde{\omega}_k$ and c_k using equations (3.14) and (3.36). In agreement with previous studies, we find that waves achieve only a local thermodynamic equilibrium and both averages and frequencies $\tilde{\omega}_k$ do depend on the interval (T_1, T_2). Despite the fact that the wave equation only achieves a LTE, the LTE does exhibit generic properties.

To verify that renormalized waves oscillate with approximately one frequency, and that the frequency is in fact $\tilde{\omega}_k$, we calculate the spatiotemporal transform of c_k over interval times (T_1, T_2). Figure 3-1 shows the spatiotemporal transform $|\widehat{c}_k(\omega)|^2$ for an early time interval $T_1 = 153$, $T_2 = 230$. Here the choice of times T_1 and T_2 is somewhat arbitrary. Indeed we also obtain similar plots for other time intervals, such as $T_1 = 767$, $T_2 = 843$ or even $T_1 = 153$ and $T_2 = 843$. To calculate the spatiotemporal transform, we first compute $\tilde{\omega}_k$ using (3.36)

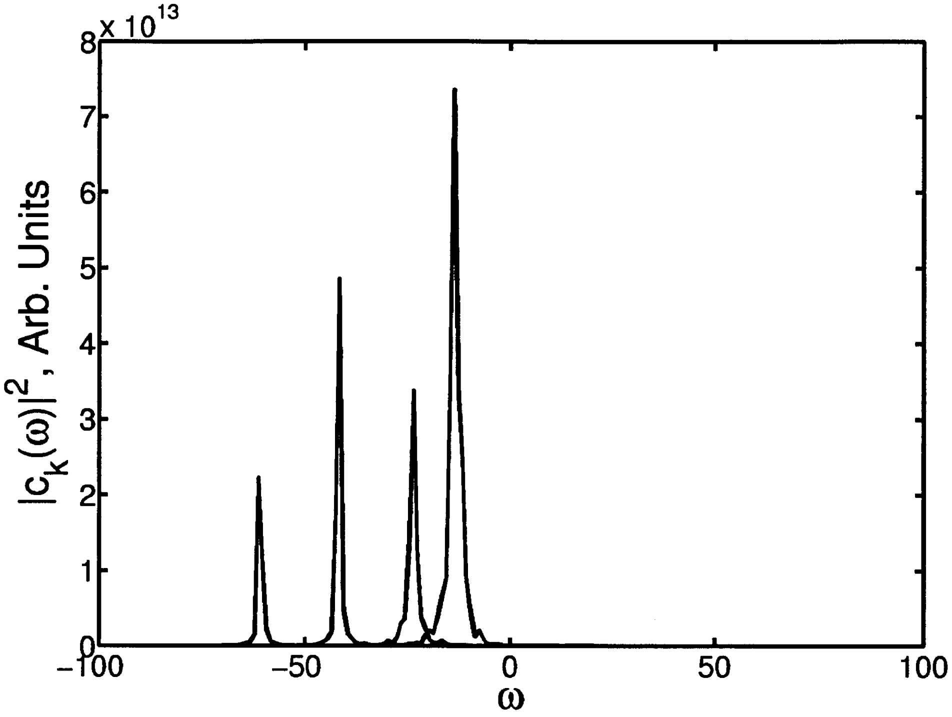

Figure 3-2: Power spectrum of renormalized waves $|\widehat{c}_k(\omega)|^2$ calculated over the interval $T_1 = (153, 230)$ for initial data (3.39). The waves stay localized around $-\tilde{\omega}_k$ while the oscillations around $\tilde{\omega}_k$ have been completely removed. From right to left, the peaks correspond to $k = 5, 20, 40, 60$.

averaged over (T_1, T_2) to obtain $c_k(t)$. For each k, we compute the power spectrum $|\widehat{c}_k(\omega)|^2$. We find that the energy in each $\widehat{c}_k(\omega)$ is sharply peaked, with a small amount of background noise. The background noise can be removed by treating each $c_k(t)$ as a stochastic signal, and averaging multiple power spectra. For instance, to remove the background noise, we divide the interval (T_1, T_2), which contains 25000 data points, into eight partitions. We calculate the power spectrum of each partition independently, and as shown in figure 3-2, average the results together. In addition to the spatiotemporal transform, the thick line in figure 3-1 is the dispersion relation $\tilde{\omega}_k$ obtained via (3.36). The dispersion relation $\tilde{\omega}_k$ corresponds very well with the localized magnitude of $|\widehat{c}_k(\omega)|^2$, indicating the waves c_k effectively oscillate with frequency $-\tilde{\omega}_k$. The faint line observed at $\tilde{\omega}_k$ is a small residual of the coupling found between c_k and c^*_{-k}. Since the plot is on a log shading scale, the coupling effect is in fact several orders of magnitude smaller than the amplitude found at $-\tilde{\omega}_k$. The removal of the coupling found between c_k and c^*_{-k} is further verified by the absence of a peak in the power spectrum at $\tilde{\omega}_k$ seen in figure 3-2.

In addition to numerically computing $\tilde{\omega}_k$, we find that, in the case of strong nonlinearities, the frequencies always satisfy a shifted Klein-Gordon dispersion relation of the form

$$\tilde{\omega}_k^2 \approx 1 + 2.59\lambda\langle S\rangle + k^2. \tag{3.40}$$

Again S is the mean squared averaged field computed over the same time interval (T_1, T_2) as $\tilde{\omega}_k$. For example, figure 3-7 shows the dispersion relation for several different initial conditions. Moreover, as a result of the LTE, both the value of S and therefore $\tilde{\omega}_k$ drift over long times. Figure 3-3 verifies the relation (3.40) on two separate time intervals, with $\langle S\rangle = 67.4$ on $(153, 230)$ and $\langle S\rangle = 60.24$ on $(767, 843)$. Figure 3-4 also shows a plot of $S(t)$ highlighting the two averaging intervals. The deviation in $\langle S\rangle$ between the two intervals corresponds to a drift of roughly 5% in the smallest frequency $\tilde{\omega}_0$.

Since the spatiotemporal transform verifies that the waves c_k form narrow band oscillators centered at frequency $\tilde{\omega}_k$, we may introduce the effective energy [35] in Fourier mode k in analogy with a linear oscillator

$$\begin{aligned} \langle E_k\rangle &= \frac{1}{2}(\langle|p_k|^2\rangle + \tilde{\omega}_k^2\langle|u_k|^2\rangle) && (3.41)\\ &= \langle|p_k|^2\rangle. && (3.42) \end{aligned}$$

Here the definition of the renormalized frequency over a given interval implies the waves on average satisfy the virial theorem, equally splitting the kinetic and potential energy. The quantity $\langle|p_k|^2\rangle$ therefore is a measure of the approximate energy in mode k, independent of renormalized frequency. Although the energy E_k describes the modal distribution of energy, one should note that $\sum_k\langle E_k\rangle$ is not a conserved quantity and that the values of $\langle E_k\rangle$ exist only in an LTE state, and weakly depend on the interval (T_1, T_2). The energy E_k can also

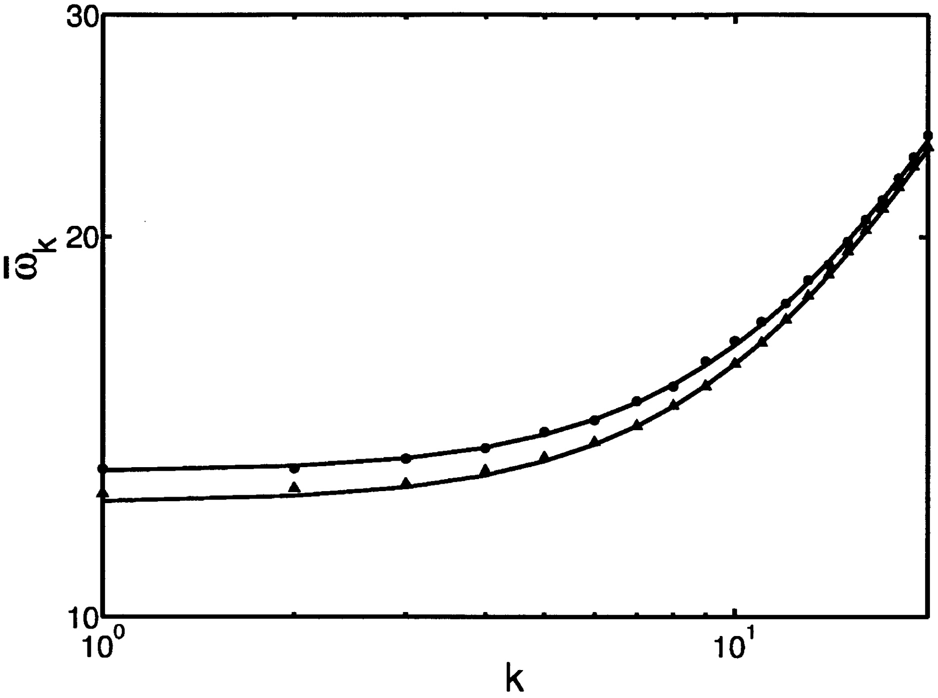

Figure 3-3: Frequency drift of renormalized waves. As the average S drifts in time, so do the renormalized frequencies. The top and bottom curves represent the spectrum of renormalized waves over the time intervals $(153, 230)$ and $(767, 843)$, respectively, for initial data (3.39). The dotted curves show the frequencies obtained by equation (3.36), while the solid lines represent fits obtained using (3.40). The quantity S is averaged over the same interval used to calculate the wave spectra.

be related to the renormalized wave amplitudes via

$$E_k = \frac{\tilde{\omega}_k}{2}\Big(\langle |c_k|^2\rangle + \langle |c_{-k}|^2\rangle\Big). \tag{3.43}$$

Consequently, in the special case when $u_k = u_{-k}$, such as the initial data (3.39), then $E_k = \tilde{\omega}_k \langle |c_k|^2\rangle$ can also be used as a measure of energy in mode k.

In general, over the thermalization stage $(0, T_1)$, solutions initialized to lower Fourier modes leak out into the higher modes. When viewed on a log scale, the energy spectrum evolves into a flat distribution in the low Fourier modes accompanied by a linear exponential decay in the high modes. For example, figure 3-5 shows the energy spectrum $\langle E_k\rangle$ averaged over the long time interval $(152, 843)$. As solutions propagate through time, the spectrum

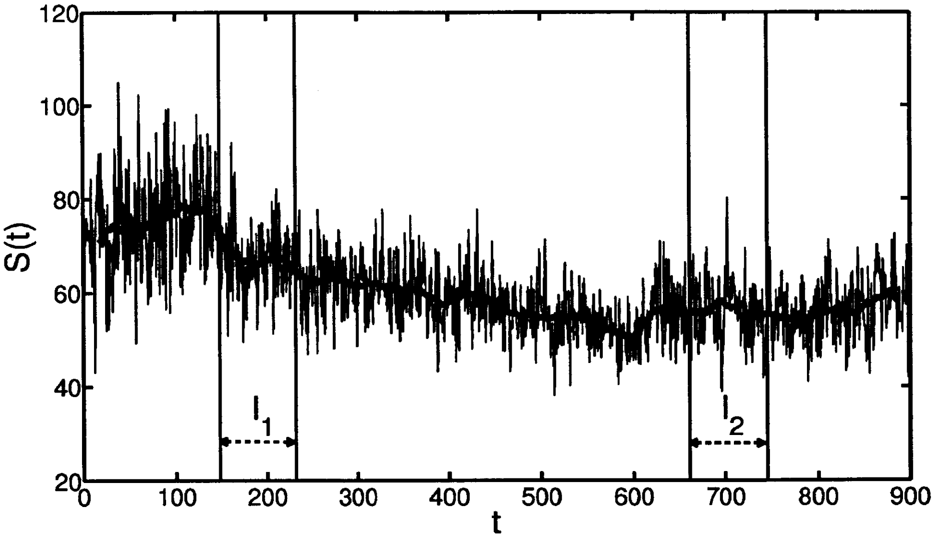

Figure 3-4: Plot of the mean field $S(t)$ for initial data (3.39). The dark curve represents a local time average $\langle S\rangle$ [over the interval $(t-14, t+14)$], and demonstrates a drift over time. The double arrows highlight the intervals $I_1 = (153, 230)$ and $I_2 = (767, 843)$.

retains a straight exponential decay, however the slope of decay may drift mildly (e.g., 10%) over time. We also calculate the sum $\sum_k \langle E_k\rangle \sim 6.8 \times 10^5$ which differs from the exact energy $H \sim 6.7 \times 10^5$ by $\approx 2.1\%$. Qualitatively, the spectrum appears very similar to the one predicted by the Planck blackbody distribution: $|c_k|^2 \propto (e^{\beta\tilde{\omega}_k} - 1)^{-1}$. The presence of a Planck-like spectrum arising from a classical system, such as a system of weakly coupled oscillators, is also discussed by Carati and Galgani [19]. For our data, the fit is only heuristic in that we can not simultaneous match the flat, long wave spectrum as well as the short wave exponential decay with a fixed temperature.

In addition to studying initial conditions (3.39) for $N = 2048$ modes, we also performed several convergence tests with varying grid sizes $N = 512, 1024$, and 4096 and different time steps Δt. We also found that numeric solutions do not depend on grid spacing provided the power spectrum remains exponentially suppressed over the integration times. For instance, $N = 4096$ with initial (3.39) yields identical solutions to $N = 2048$. Meanwhile, $N = 512$ develops an energy cascade into the ultraviolet spectrum, accompanied by the onset of equipartition of energy. Hence, for the initial conditions (3.39), taking $N = 2048$ yields a consistent solution to the PDE over the integration times $0 < T < 843$, while inconsistencies develop for smaller $N < 1024$.

3.3.2 Kinetics and fluctuations of the LTE

In addition to the power spectrum and renormalized dispersion relation, we also study the kinetic behavior of the LTE as well as fluctuations about equilibrium values. Since the frequency shift tracks the spatial average S, we plot S as a function of time to gain an understanding of the dispersion relation evolution. Figure 3-4 shows the value S over our integration time, along with a local time average value of S. We take the local time average S over intervals long enough for the waves c_k to exhibit numerous oscillations. Although the waves c_k appear well defined for any fixed time interval (T_1, T_2), there appears no well defined equilibrium, or more precisely if even the limit $\lim_{T\to\infty} \frac{1}{T}\int_0^T S\,\mathrm{d}t$ exists. As shown in figure 3-4, S generally appears to decay, however drifts in time.

To further characterize the nature of the LTE, we examine fluctuations by plotting probability distributions of the squared field strength S and energy spectrum for different time intervals. We find that for a sufficiently long time interval, S acquires a Gaussian probability distribution with varying mean and width. Figure 3-6 shows the probability distribution S, as well as a Gaussian fit, over the interval $(153, 230)$. Although not shown, the interval $(767, 843)$ also exhibits a Gaussian probability distribution with a different mean. Only at very short time intervals, does S not widen out to a Gaussian.

We also study the probability distributions (PDF) for wave amplitudes or particle numbers $|c_k|^2$. The distribution of these variables are related to the energy spectrum since one can identify $\tilde{\omega}_k |c_k|^2$ as the energy in mode k. Using the initial conditions (3.39) we plot the distribution of $|c_k|^2$ over the interval $(767, 843)$ for different wave numbers. We find that low wave numbers, $k < 200$, which achieve equipartition of energy, exhibit a classical Boltzmann distribution. For example figure 3-10 illustrates the distribution for mode $k = 2$ along with an exponential fit of the form $e^{-\beta|c_k|^2}$. Meanwhile, the exponential distributions found in wave modes near the natural spectral cutoff, $k \approx 200$, start to shift their peaks

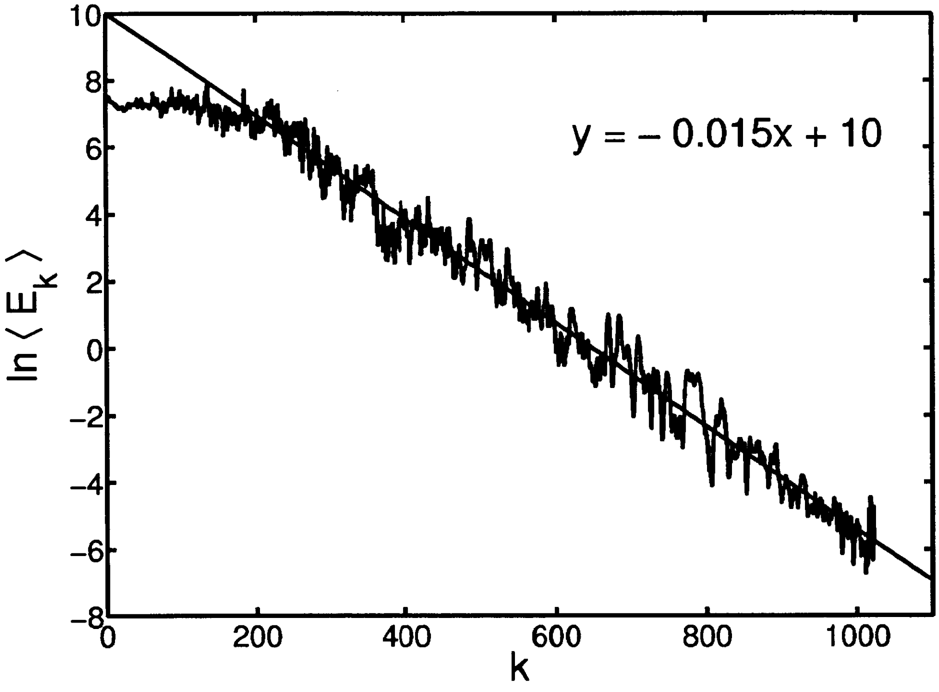

Figure 3-5: Energy spectrum of renormalized waves (on log scale), $E_k = \langle |p_k|^2 \rangle$. Spectrum is qualitatively similar to the distribution derived by Planck, equipartition of energy in the low modes, exponential decay in the high modes. Data is averaged over the time interval (153, 843) for initial data (3.39).

away from $|c_k|^2 = 0$, as seen in figure 3-11. Lastly, there does not appear to be a consistent distribution of energy at large wave numbers $k > 200$. Specifically, adjacent modes k and $k + 1$ may exhibit very different distributions. Despite the lack of a unifying distribution, many large wave modes do exhibit a multipeaked distribution. Figures 3-12 and 3-13 show the amplitude distribution for $k = 510$ and $k = 520$, respectively.

3.4 General Trends

In the following section we study numerical solutions for a wide range, roughly 40 trials, of initial data. In our experiments, we evolve initial data for a fixed time $T_1 = 767$, after which we calculate quantities characteristic of the LTE. We test three generic sets of initial conditions shown in Tables 3.1, 3.2, and 3.3. Specifically, the tables show data for the average

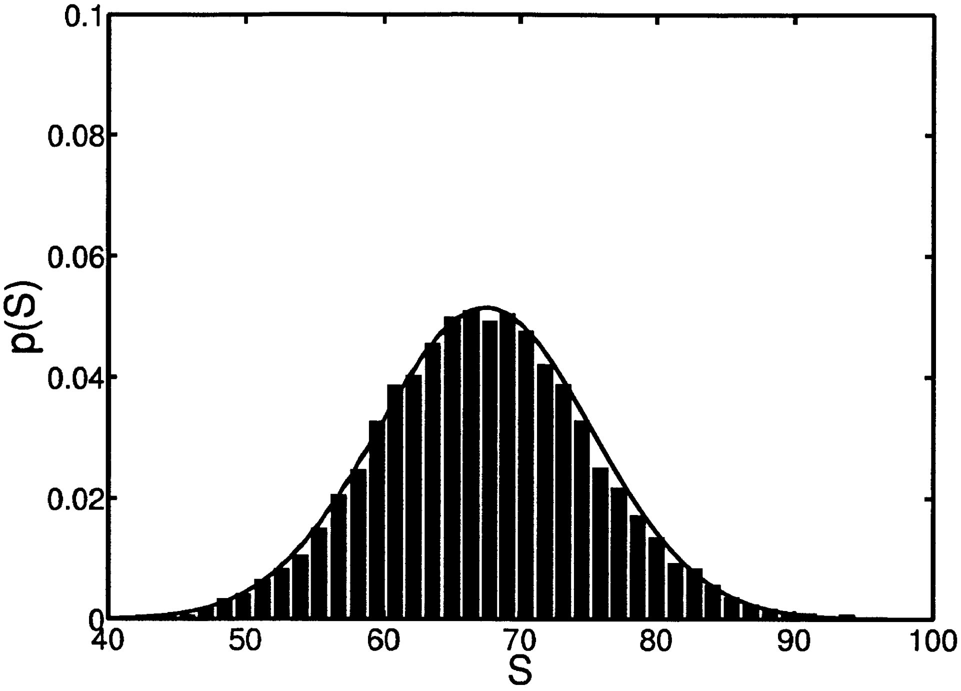

Figure 3-6: Histogram of mean field S over time interval (153, 230) for initial data (3.39). The mean field S is well fit by a Gaussian distribution with mean 67.4.

$\langle S\rangle$, the frequency shift $\tilde{\omega}_0^2$ obtained by a best fit to the renormalized dispersion relation, the variance of S, the exponential slope in the power spectrum and the classical analog of the particle number. Here the slope of the power spectrum refers to the slope obtained by a least squares linear fit to $\ln\langle E_k\rangle$ over the large wave numbers k. Physically, the slope corresponds to one measure of the inverse temperature $\beta = (k_b T)^{-1}$ from a Plank spectrum. In addition, the particle number is defined by

$$\langle N\rangle = \sum_k \langle |c_k|^2\rangle. \tag{3.44}$$

We include the values of $\langle N\rangle$ as they often play a role in quantum mechanics.

In our first test trials, we fix the initial shape $u(x, 0)$ by setting $u_k = C$ a constant over the first 50 modes ($0 < |k| \leq 50$). Altering the initial constant C varies the nonlinear wave strength. Phenomenologically, trials for varying coupling constants exhibit behavior

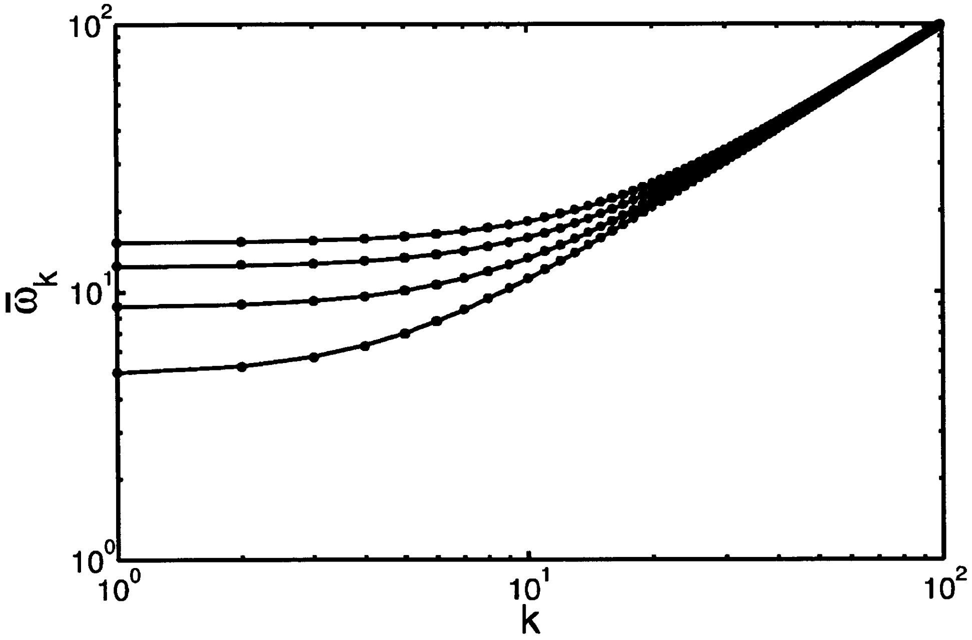

Figure 3-7: Renormalized dispersion relation for different coupling strengths over time interval (767, 843). Initial conditions are taken as a constant over the first 50 modes. The curves (bottom to top) have initial Fourier amplitudes $u_k = 0.69, 1.39, 2.08, 2.77$. The dotted line corresponds to the exact renormalized frequencies from (3.36), while the solid line is a fit with the numerically calculated dispersion relation $\tilde{\omega}_k^2 = 1 + 2.59\lambda\langle S\rangle + k^2$.

identical to the LTE described in the previous section. For instance, figure 3-7 shows the renormalized dispersion relation at different coupling strengths.

The second set of tests fix the coupling strength $\lambda = 1$, and varying the initial conditions so that energy H and momentum $P = 0$ remain constant. The primary goal is to determine whether drastically different LTE solutions arise from initial conditions with the same energy and momentum P, H. Table 3.2 shows LTE trends for initial data $u_k = 0.380, 0 < |k| \leq 150$, $u_k = 0.648, 0 < |k| \leq 100$, and $u_k = 2.08, 0 < |k| \leq 50$. The data show that the renormalized frequencies, mean field strength, and exponential slope vary dramatically over the initial data.

Lastly, the third set of initial data fixes $\lambda = 1$ and varies both the initial conditions $\dot{u}_k$ and u_k so that the total momentum $P \neq 0$. The goal in this case is to test whether

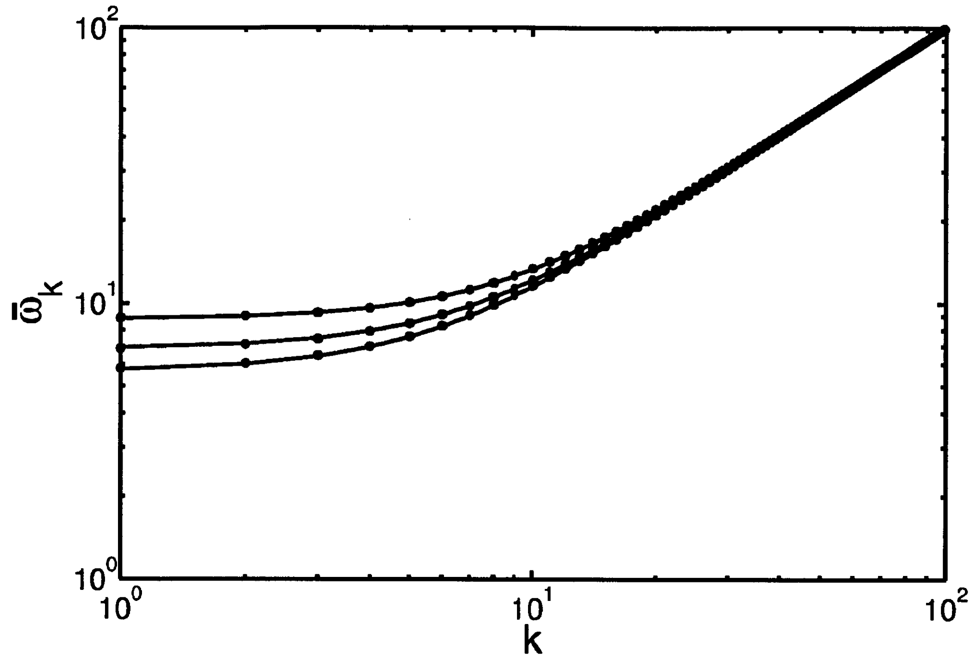

Figure 3-8: Renormalized dispersion relation for initial conditions with fixed energy $H \sim 1.8 \times 10^5$ and total momentum $P = 0$. Data is taken over the time interval $(767, 843)$. From bottom to top, the curves correspond to initial data: $u_k = 0.380, 0 < |k| \leq 150$, $u_k = 0.648, 0 < |k| \leq 100$ and $u_k = 1.38, 0 < |k| \leq 50$.

moving waves alter thermalization phenomenology. In general, the wave field thermalize into a renormalized wave basis with the characteristic dispersion relation $\tilde{\omega}_k$ and Planck-like spectrum. As shown by the data, moving waves exhibit trends identical to stationary ones.

3.5 A Lattice versus a Partial Differential Equation

In the present chapter we have focused on the long time behavior of the Klein-Gordon PDE. There is however, a long history of numerical science based on the long time behavior of discrete, finite dimensional, lattices. As a result, in this section we present some background on the lattice experiments, and contrast them with the numerical solution of a PDE.

In 1965, Fermi, Pasta and Ulam [27] (FPU) performed a numerical experiment to show that a weak nonlinearity would thermalize a discrete lattice. To their surprise, they discov-

Table 3.1: Different coupling constants. The first column shows the initial value for u_k over modes $0 < |k| \leq 50$. Averages are taken over the time interval $(767, 843)$.

Initial u_k	H	P	$\langle S \rangle$	$\tilde{\omega}_0^2$	$Var(S)$	-Slope	$\langle N \rangle$
0.69	26607	0	8.9	25.02	1.9	0.035	756
1.39	178502	0	29.3	77.7	4.6	0.023	3860
2.08	671909	0	60.3	154.0	9.9	0.015	10550
2.77	1867200	0	89.5	232.8	10.8	0.011	21175

Table 3.2: Trials with fixed energy $H \sim 1.8 \times 10^5$ and momentum $P = 0$. Initial conditions are taken as constants for u_k over the first $50, 100, 150$ modes. Averages are taken over the time interval $(767, 843)$.

Num. modes	Initial u_k	$\langle S \rangle$	$\tilde{\omega}_0^2$	$Var(S)$	-Slope	$\langle N \rangle$
50	1.39	29.3	77.7	4.6	0.023	3860
100	0.65	18.1	48.0	2.8	0.017	2450
150	0.38	12.1	33.3	2.2	0.012	1760

Table 3.3: The trials have data of the form $u_k = A$ with $p_k = \dot{u}_k = \imath B\sqrt{1+k^2}$, $\dot{u}_k^* = \dot{u}_{-k}$ for $0 < |k| \leq 50$. Trials 1 and 2 correspond to amplitudes $A = B = 0.69, 1.38$, respectively. Trial 3 corresponds to amplitudes $A = 1.38$, $B = 2A = 2.76$. Note, we choose $p_k \propto \omega_k$ to mimic the initial conditions one would require for traveling linear waves.

Trial	H	P	$\langle S \rangle$	$\tilde{\omega}_0^2$	$Var(S)$	-Slope	$\langle N \rangle$
1	47208	41180	12.2	33.5	1.7	0.021	1218
2	260900	164700	33.9	89.2	4.6	0.016	5047
3	508120	329420	51.1	136.5	7.6	0.015	8664

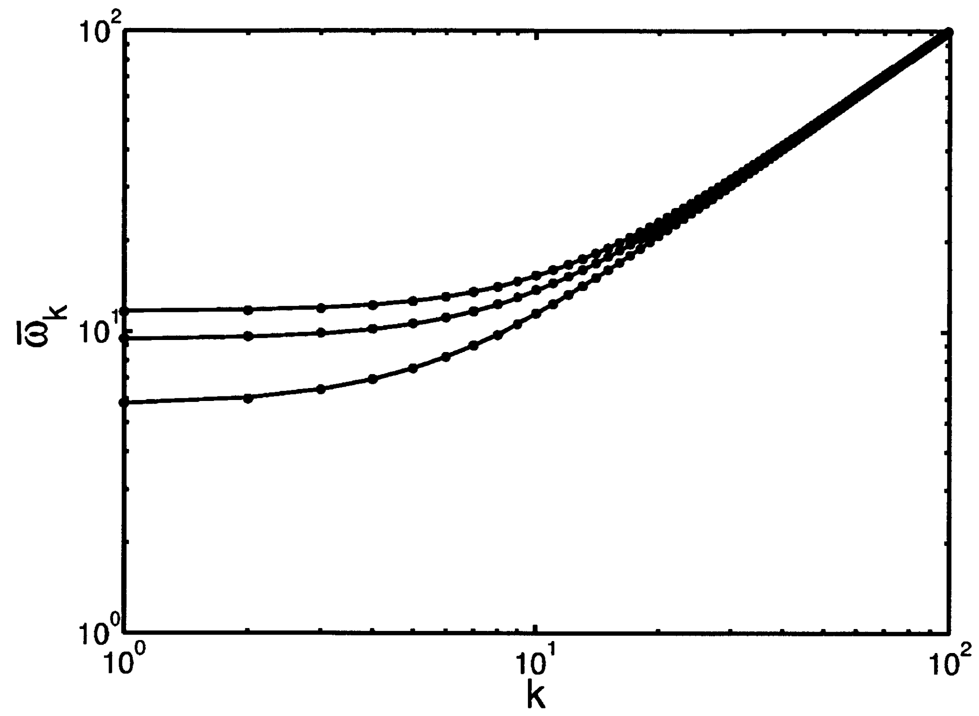

Figure 3-9: Renormalized dispersion relation for initial conditions in first 50 modes, with non-zero total momentum. Data is taken over the interval (767, 843). From bottom to top, the curves correspond to integrals of motion: $(H, P) \sim \{(4.7 \times 10^4, 4.1 \times 10^4), (2.6 \times 10^5, 1.6 \times 10^5), (5.1 \times 10^5, 3.3 \times 10^5)\}$.

ered that the lattice did not thermalize to the expected equipartition of energy, but in fact the energy returned periodically to its initial configuration. In the flurry of work following FPU, numeric results exposed that the quasiperiodic solutions were in fact dependent on initial data and nonlinear coupling strength. For sufficiently strong nonlinearities, the FPU lattice does achieve equipartition of energy [12].

In contrast to the finite dimensional lattice, wave equations contain an infinite number of modes. In the context of wave thermodynamics, the difficulties with infinitely many modes has been known for over 100 years, dating back to the Rayleigh-Jeans paradox and ultraviolet catastrophe. In particular, the assumption of equipartition of energy implies that the total wave energy diverges for any fixed thermodynamic temperature. Although this difficulty was resolved by Planck for the problem of radiation, the thermalization of a partial differential equation still requires one to work with an infinite number of modes.

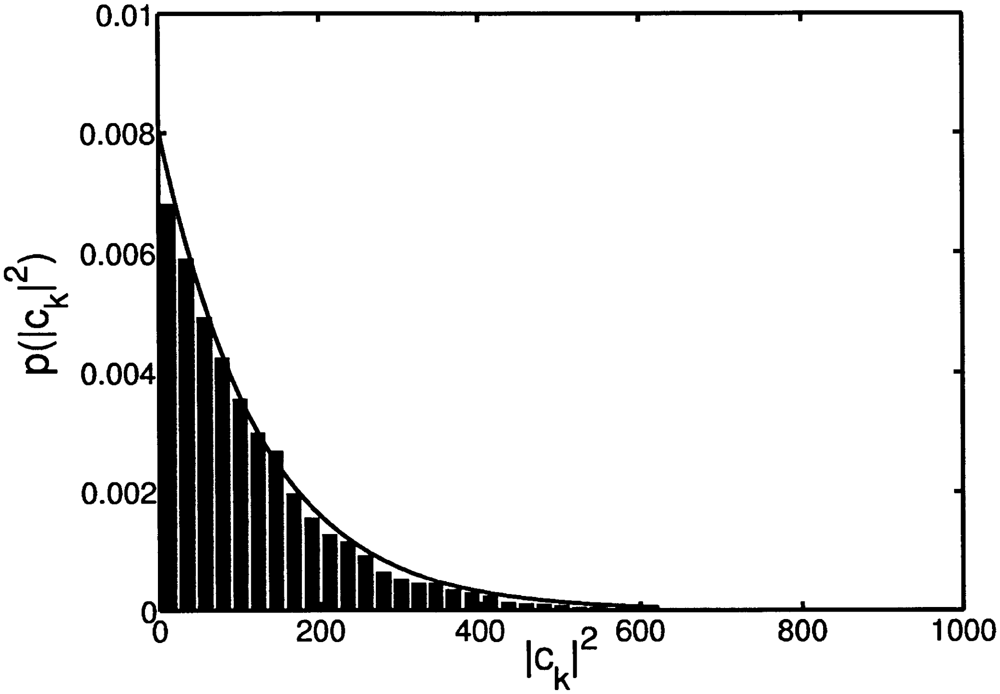

Figure 3-10: PDF for renormalized wave at small wave number, $(k = 2)$ $|c_2|^2$. Data are taken over the time interval $(767, 843)$ for initial data (3.39). The curve represents an exponential, Boltzmann distribution.

The partial differential equation (PDE) limit, which roughly speaking takes the number of discrete lattice points $N \to \infty$ while keeping energy, E and volume V constant, is also radically different from the thermodynamic limit which takes $N \to \infty$ at the same rate as the system size $V \to \infty$. One consequence of this difference is the absence of equipartition. For instance regularity results [73, 78] suggest the absence of a Rayleigh-Jeans divergence in classical field theories. Moreover, the finite truncation in any numeric experiment requires a careful interpretation of the results. For instance, numerically solving a PDE over very long computational times is mathematically identical to integrating a discrete lattice with many points. More specifically, by truncating the PDE to a finite lattice with N modes, nonlinear terms may artificially introduces aliasing effects that remain absent in the PDE. In the case of u^4 nonlinearity, upon discretization, Fourier modes (k, l, m, n) satisfying the relation $k + l = m + n + N$ strongly interact. These aliasing effects, which are absent

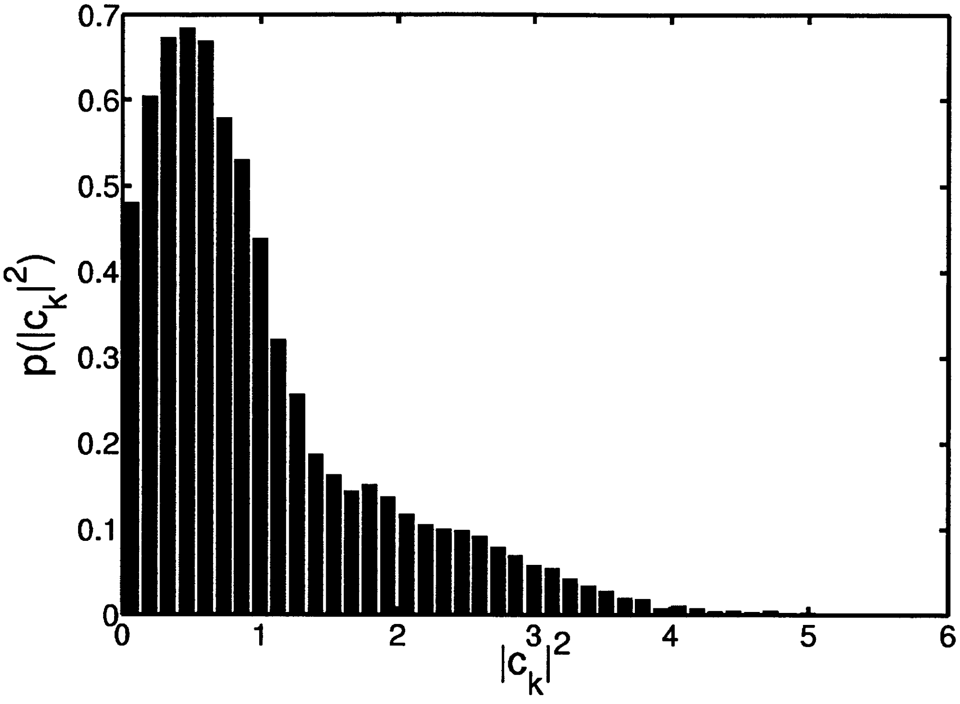

Figure 3-11: PDF for renormalized wave for mid ranged wave number ($k = 209$) $|c_{209}|^2$. Data are taken over the time interval (767, 843) for initial data (3.39). Mode $k = 209$ is on the edge of the spectrum between equipartition in the low modes and exponential decay in the high modes. The decrease in probability at low amplitude shifts the peak to the right.

in the PDE system, provide a mechanism for energy transport from low to high Fourier modes. As shown by De Luca and Lichtenberg [74], such interactions are responsible for the equipartition of energy in a lattice. As a result, to distinguish $u(x,t)$ as a solution to the PDE and not a discrete lattice approximation, one requires that the numerical solutions converge, in an appropriate norm, to the continuous field as the number of lattice points go to infinity.

3.6 Numerical Method

In the following section, we describe in detail our numerical method for evolving solutions to the Klein-Gordon equation. As mentioned in section 3.2, we use a pseudo-spectral method, which combines a spectral propagation for the linear terms with a Richardson extrapolation

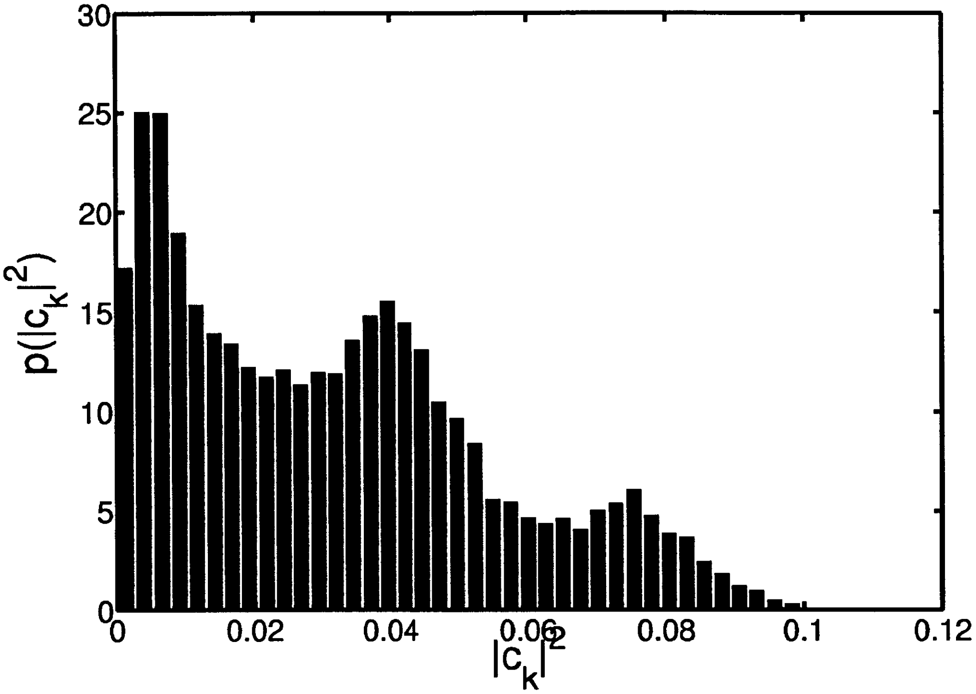

Figure 3-12: PDF for renormalized wave for large wave number ($k = 510$) $|c_{510}|^2$. Data are taken over the time interval $(767, 843)$ Higher wave numbers exhibit a variety of behavior. In the $k = 510$, wave number, the amplitude jumps around between three peaks.

for the nonlinear ones.

To march equation (3.2) through time, we convert the PDE into an equivalent first-order system:

$$\mathbf{U}_t - \imath\widehat{L}(\mathbf{U}) = \widehat{N}(\mathbf{U}) \tag{3.45}$$

where $\mathbf{U} = (u, w)^T$, $\widehat{L}$ and $\widehat{N}$ are linear and nonlinear operators respectively. We take coordinates for u and w along the characteristics of (3.2) so that $\widehat{L}$ and $\widehat{N}$ are:

$$\begin{aligned} \widehat{L} &= \begin{pmatrix} \imath\partial_x & -\imath \\ \imath & -\imath\partial_x \end{pmatrix}, \\ \widehat{N} &= \begin{pmatrix} 0 \\ -\lambda u^3 \end{pmatrix}. \end{aligned} \tag{3.46}$$

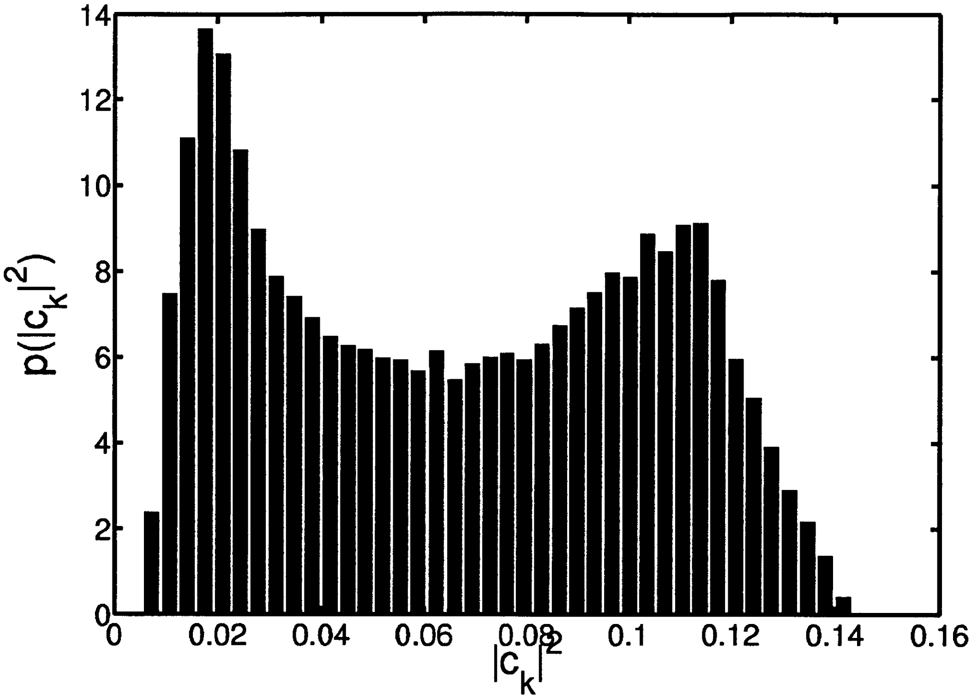

Figure 3-13: Distribution of renormalized wave for large wave number ($k = 520$) $|c_{520}|^2$. Data are taken over the time interval (767, 843).

Using matrix exponentials, we may integrate the linear term in (3.45) to obtain an expression for the propagation over a small time h:

$$\begin{aligned}
\left(e^{-i\hat{L}t}\mathbf{U}\right)_t &= e^{-i\hat{L}t}\widehat{N}(\mathbf{U}), \\
\mathbf{U}(t+h) &= e^{ih\hat{L}}\mathbf{U}(t) + e^{ih}\int_t^{t+h} e^{i(t-s)}\widehat{N}[\mathbf{U}(s)]\,\mathrm{d}s, \\
\mathbf{U}(t+h) &\approx e^{ih\hat{L}}\mathbf{U}(t) + h\widehat{N}[\mathbf{U}(t)]. \qquad (3.47)
\end{aligned}$$

When written in Fourier space, the differential operator $e^{ih\hat{L}}$ decouples into 2×2 matrices acting on each Fourier component of $\mathbf{U}$. We therefore use Fourier modes and a fast Fourier transform when evaluating the matrix exponential. Meanwhile, we adapt the Richardson extrapolation algorithm for propagating ODEs [82], to handle the nonlinear term in (3.47).

The Richardson extrapolation routine is a method for evaluating numerical limits. In our case, starting with $\mathbf{U}(t)$ at time t, we seek to obtain the solution $\mathbf{U}_h(t + \Delta t)$ at a fixed time Δt later, in the limit $h \to 0$. Once we have extrapolated the solution a time step Δt, we repeat the process M times to obtain a solution $\mathbf{U}(T)$ at time $T = M\Delta t$. To accomplish the extrapolation over a step Δt, we truncate our system to $N = 2048$ spectral modes and a spatial stencil $\Delta x = \frac{2\pi}{N}$. We also fix $\Delta t = 0.5\Delta x \ll 1$. The limit, $\lim_{h\to 0} \mathbf{U}(t + \Delta t)$ can now be thought of as the simultaneous limit of a finite number of variables. Following the Richardson extrapolation routine, we integrate solutions over a time Δt to obtain $\mathbf{U}_h(t+\Delta t)$ using successively smaller values of an intermediate step h. For instance, taking $h_0 = \Delta t$ in equation (3.47) yields a one step Euler approximation $\mathbf{U}_{h_0}(t + \Delta t)$. Taking $h_1 = \Delta t/2$, then requires 2 evaluations of (3.47) to obtain $\mathbf{U}_{h_1}(t + \Delta t)$. In general, the evaluation of $\mathbf{U}_h(t+\Delta t)$, requires one to divide the time step Δt into l pieces of length $h = \Delta t/l$, followed by integrating (3.47) l times. Provided the nonlinearity in (3.47) is an analytic function, and the solutions u are smooth, a polynomial extrapolation of the sequence $\mathbf{U}_{h_j}$, for successively smaller values of h_j, will converge with error $O(\Delta t^{2r+1})$. Here r is the number of sampled values h_j, $0 \leq j < r$. In our numerics, we have $r = 7$, which gives us a relative accuracy of $\Delta u \sim O(10^{-12})$.

Over the PDE integration times, we track numerical errors by estimating the absolute errors, and recording the two integrals of motion H and P. From the Richardson extrapolation routine we obtain an L^1 error estimate on the propagated solution for each time step Δt. Summing the L^1 errors at each step then provides an absolute bound on the L^1 error between the exact and numeric solutions. As a consistency check, we also calculate the integrals of motion, and find that they remain constant to within roughly one part in 10^{12} over each time step Δt. In the worst test cases, the integrals of motion deviate to one part in 10^6 after 3×10^5 time iterations. Meanwhile, the summed L_1 error remains bounded to roughly one order of magnitude less than the error in the integrals of motion. Lastly, to guarantee the u^4 nonlinearity does not introduce aliasing effects, we restrict our initial data so that Fourier amplitudes at wave numbers $k_{max} \approx 600$ remain exponentially small over the entire integration time.

3.7 Periodic Klein-Gordon Solutions

In the previous sections of this chapter, we evolve Klein-Gordon solutions for initial data with energy in a large number of Fourier modes, and show that they thermalize into an LTE. Although these LTE solutions appear to arise from generic initial data, the Klein-Gordon equation also admits traveling wave, exactly periodic solutions [105]. For instance, initial data starting on a periodic solution will not thermalize, but will remain periodic for all time. In this section, we investigate these periodic solutions and extract the large amplitude limit $\lambda \gg 1$. We show that solutions with spatial period $2\pi/k$ oscillate at a frequency $\omega_k^2 \approx 0.95 + 1.57\lambda\langle u^2\rangle_{2\pi} + k^2$, where $\langle u^2\rangle_{2\pi}$ is defined below.

Starting with the Klein-Gordon equation

$$\begin{aligned} \partial_t^2 u - \partial_x^2 u + V'(u) = 0, \qquad (3.48)\\ V(u) = \frac{1}{2}u^2 + \frac{\lambda}{4}u^4, \qquad (3.49)\end{aligned}$$

a periodic, traveling wave solution u, with velocity v, has the form

$$\begin{aligned} u(\theta) &= u(\theta + 2\pi), \qquad (3.50)\\ \theta &= x - vt. \qquad (3.51)\end{aligned}$$

Substitution then yields an integrable ODE[5]

$$\begin{aligned} (v^2-1)u'' + V'(u) &= 0, \qquad (3.52)\\ \frac{1}{2}(v^2-1)(u')^2 + V(u) &= V(u_m). \qquad (3.53)\end{aligned}$$

Here $V(u_m)$ is the constant of integration, where u_m is the maximum amplitude of the field u. Provided $v^2 > 1$, the equation (3.53) describes a nonlinear oscillator for the variable u. Therefore, for a fixed maximum amplitude u_m, only specific values of v will yield periodic, $u(\theta + 2\pi) = u(\theta)$, solutions. Namely, these values of v correspond to solutions u which

---

[5] Note that $V'(u) = \frac{dV}{du}$, while $u' = \frac{du(\theta)}{d\theta}$.

oscillate k times over the domain $0 < \theta < 2\pi$. Hence, v must satisfy

$$4k \int_0^{u_m} \frac{2^{-1/2}\,\mathrm{d}u}{\sqrt{V(u_m) - V(u)}} = \frac{2\pi}{\sqrt{v^2 - 1}}. \tag{3.54}$$

For brevity, introduce the effective mass

$$f(u_m)^{-1} = \frac{2}{\pi} \int_0^{u_m} \frac{2^{-1/2}\,\mathrm{d}u}{\sqrt{V(u_m) - V(u)}} \tag{3.55}$$

which yields

$$v^2 = 1 + \frac{f^2(u_m)}{k^2}. \tag{3.56}$$

The frequency in time ω then becomes

$$\begin{aligned} \omega_k &= kv, &(3.57)\\ \omega_k &= k\sqrt{1 + f^2(u_m)/k^2}, &(3.58)\\ \omega_k &= \sqrt{k^2 + f^2(u_m)}. &(3.59) \end{aligned}$$

Thus far, the nonlinear frequency relation (3.59) is general in the sense that one only requires the potential $V(u)$ to be monotonic and symmetric, $V(-u) = V(u)$, in u. In the original variables x and t, the kth traveling wave solution to (3.53), $u_k(\theta)$, is periodic in space $x \to x + \frac{2\pi}{k}$ and time $t \to t + \frac{2\pi}{\omega_k}$. For the special case of a u^4 nonlinear potential, the integral $f(u_m)$ and solution $u(\theta)$ may be written down explicitly in terms of elliptic functions [3]. As a result, we may extract the asymptotic behavior for large fields $\lambda u_m^2 \gg 1$:

$$f^2(u_m) \sim 0.7178\lambda u_m^2 + 1.0458 + O\Big(\frac{1}{\lambda u_m^2}\Big). \tag{3.60}$$

To compare the frequency shifts for the periodic, traveling wave solutions to those found in the renormalized waves, we recast the maximum field amplitude u_m in terms of the square averaged field:

$$
\begin{aligned}
\langle u^2\rangle_{2\pi} &= \frac{1}{2\pi}\int_0^{2\pi} u^2\,\mathrm{d}\theta && (3.61)\\
&= \frac{k}{2\pi}\int_0^{2\pi/k} u^2\,\mathrm{d}\theta && (3.62)\\
&= \frac{4k}{2\pi}\int_0^{u_m} u^2\frac{d\theta}{du}\,\mathrm{d}u && (3.63)\\
&= \frac{4k}{2\pi}\Big(\frac{v^2-1}{2}\Big)^{-1/2}\int_0^{u_m} u^2\,\mathrm{d}\mu && (3.64)\\
&= \frac{\int_0^{u_m} u^2\,\mathrm{d}\mu}{\int_0^{u_m}\,\mathrm{d}\mu}, && (3.65)
\end{aligned}
$$

where we have used the ODE (3.53) to replace $\frac{\mathrm{d}\theta}{\mathrm{d}u}$ with a function of u. Here $\mathrm{d}\mu$ is the measure defined as

$$\mathrm{d}\mu = \frac{\mathrm{d}u}{\sqrt{V(u_m) - V(u)}}. \qquad (3.66)$$

In the large amplitude limit, the averaged field scales with the peak field amplitude u_m

$$\langle u^2\rangle_{2\pi} \sim 0.456947u_m^2 + \frac{0.061347}{\lambda} + O\Big(\frac{1}{\lambda^2 u_m^2}\Big). \qquad (3.67)$$

Hence, in terms of the averaged squared field $\langle u^2\rangle_{2\pi}$, the frequency shift is

$$
\begin{aligned}
f^2(u_m) &\sim 1.5708\lambda\langle u^2\rangle_{2\pi} + 0.9494 + O\Big(\frac{1}{\lambda u_m^2}\Big), && (3.68)\\
\omega_k^2 &\approx 0.95 + 1.57\lambda\langle u^2\rangle_{2\pi} + k^2. && (3.69)
\end{aligned}
$$

As with the renormalized wave solutions, the periodic solutions $u_k(\theta)$ exhibit a mass frequency shift proportional to the averaged field $\lambda\langle u^2\rangle$. In addition, like the renormalized waves, the leading term in the large amplitude shift does not depend on the number of oscillations k, however, does admit a different leading coefficient of 1.57 as opposed to 2.59.

3.7.1 Summary

In this chapter, we numerically studied the long time behavior of the classical Klein-Gordon equation with a strong u^4 nonlinear interaction. By introducing a renormalized wave basis, we showed that the system exhibits characteristics, locally in time, similar to a weakly nonlinear system. Specifically, the renormalized waves remain uncorrelated and form narrow band oscillators centered around one frequency. In addition, the renormalized waves, in their LTE state, achieve a renormalized dispersion relation described by equation (3.40). Here (3.40) applies to the case of a strong nonlinear coupling strength, but appears qualitatively similar, with a different constant, to the one found in the weakly nonlinear analysis (3.25). In addition, the mean field $\langle S \rangle$, and subsequently the nonlinear dispersion relation may drift as much as 10% over long times, e.g., timescales $T \gg \tilde{\omega}_0^{-1}$. However, the LTE renormalized dispersion relation (3.40) still holds over any time interval (T_1, T_2).

We also found that fluctuations about the LTE are described by several characteristic distributions. For strong nonlinearities, the probability distribution of the mean squared wave field, S, appears as a Gaussian. Meanwhile, for low wave numbers, the amplitude of the renormalized waves $|c_k|^2$ exhibit a Boltzmann distribution. As a result, there is no longer equipartition of energy throughout the large wave numbers.

In the current study, certain results, such as the exact form of the nonlinear dispersion relation, depend explicitly on the u^4 potential. It would be interesting to determine whether similar relations hold for generic potentials.

Chapter 4

Oscillons with Flat Tops

4.1 Oscillons in Scalar Fields

In many dispersive wave systems, the presence of a focusing nonlinearity can result in soliton and breather solutions, or approximate solutions, which remain spatially localized for very long times. For example, a large number of nonlinear field theories support such localized solutions. Here, in the context of scalar field theory, the Klein-Gordon equation in the presence of a $\lambda < 0$ nonlinearity changes the wave behavior described in chapter 3, by supporting oscillon solutions. These oscillons, which develop somewhat spontaneously, exist as slowly radiating, oscillating, but spatially localized waves. As a result, they may drastically change the thermal behavior described in the previous chapter.

With an eye towards their cosmological implications, in this chapter we investigate oscillons in an expanding universe [6]. After introducing the Klein-Gordon model, in section 4.3 we provide an analytic solution for the resulting one-dimensional oscillons and discuss their generalization to three dimensions. In section 4.4, we discuss the oscillon stability to long wavelength perturbations, and in particular, we discuss a new, extended class of oscillons with flat tops. We show that these flat-topped oscillons, which acquire their shape from a saturating nonlinearity, are more robust against collapse instabilities in (3+1) dimensions than their usual counterparts. Meanwhile in section 4.5 we estimate the effects of expansion on oscillon shapes and lifetimes.

4.2 The Model

We begin with the action for a real scalar field in a $d+1$-dimensional, spatially flat, homogeneous, expanding universe ($\hbar = c = 1$):

$$S_{d+1} = \int (adx)^d dt \left[\frac{1}{2}\left(\partial_t \varphi\right)^2 - \frac{1}{2a^2}\left(\nabla\varphi\right)^2 - V(\varphi)\right], \tag{4.1}$$

where

$$V(\varphi) = \frac{1}{2}m^2\varphi^2 - \frac{\lambda}{4}\varphi^4 + \frac{g}{6}\varphi^6, \tag{4.2}$$

$a(t)$ is the dimensionless scale factor and $\lambda, g > 0$. As discussed in chapter one, what is crucial for the existence of oscillons is

$$V'(\varphi) - m^2\varphi < 0,$$

for some range of the field. The potential [Eq. (4.2)] is the simplest model which captures the effect we wish to explore, namely, a nonmonotonic relationship between the height and width and its implication for stability. However, apart from detailed expressions, our results are general and not restricted to the particular shape of the potential.

We find it convenient to work with dimensionless space-time variables as well as fields. Using $(t, x) \to m^{-1}(t, x)$ and $\varphi \to m\lambda^{-1/2}\varphi$ as well as $g \to (\lambda/m)^2 g$ the action becomes

$$S_{d+1} = m^{3-d}\lambda^{-1} \int (adx)^d dt \left[\frac{1}{2}\left(\partial_t\varphi\right)^2 - \frac{1}{2a^2}\left(\nabla\varphi\right)^2 - V(\varphi)\right], \tag{4.3}$$

with

$$V(\varphi) = \frac{1}{2}\varphi^2 - \frac{1}{4}\varphi^4 + \frac{g}{6}\varphi^6.$$

The classical equations of motion are given by

$$\partial_t^2\varphi - \frac{\nabla^2}{a^2}\varphi + H\partial_t\varphi + \varphi - \varphi^3 + g\varphi^5 = 0, \tag{4.4}$$

where g is the only free parameter in the potential and $H = \dot{a}/a$. We will be concentrate on the case where $g \gg 1$. This gives a controlled expansion in powers of $g^{-1/2}$, which allows us to derive an analytic form for the profile. Our approach is similar to the small amplitude expansion, with the important difference that it captures the entire range of amplitudes for which oscillons exist. Moreover, in our analysis we show that the flat-top oscillons are stable against small amplitude, long wavelength perturbations on time scales of order g.

4.3 Oscillon Profile and Frequency

In this section we derive the spatial profile of the oscillons in our model in a $1+1$- and $3+1$-dimensional Minkowski universe. We include the effects of expansion in Sec. 4.5. For simplicity, we begin with the $1+1$-dimensional case.

4.3.1 Profile and frequency in $1+1$ dimensions

The equation of motion is

$$\partial_t^2\varphi - \partial_x^2\varphi + \varphi - \varphi^3 + g\varphi^5 = 0. \tag{4.5}$$

To extract the oscillon profile, we introduce the following change of variables:

$$\begin{aligned} \varphi(t,x) &= \frac{1}{\sqrt{g}}\phi(\tau,y), \\ t &= \omega^{-1}\tau, \\ y &= x/\sqrt{g}, \end{aligned} \tag{4.6}$$

where

$$\omega^2 = 1 - g^{-1}\alpha^2.$$

Here, α^2 characterizes the change in frequency due to the nonliear potential. We define $\Phi_0 = \phi(0,0) = \sqrt{g}\varphi_0$ and choose $\partial_t\varphi(0,x) = \partial_\tau\phi(0,x) = 0$. Note that α and Φ_0 are not independent of each other. Their relationship will be determined from the requirement that the solution is periodic in time, smooth at the origin and vanishing at spatial infinity. With the change of variables (4.6), and collecting powers of g, the equations become

$$\partial_\tau^2\phi + \phi + g^{-1}\left[-\alpha^2\partial_\tau^2\phi - \partial_y^2\phi - \phi^3 + \phi^5\right] = \mathcal{O}[g^{-3/2}]. \tag{4.7}$$

Let us consider solutions of the form

$$\phi(\tau, y) = \phi_1(\tau, y) + g^{-1}\phi_3(\tau, y) + \dots \tag{4.8}$$

Again collecting powers of g, we get

$$\begin{aligned} &\partial_\tau^2\phi_1 + \phi_1 = 0, \\ &\partial_\tau^2\phi_3 + \phi_3 = \alpha^2\partial_\tau^2\phi_1 + \partial_y^2\phi_1 + \phi_1^3 - \phi_1^5. \end{aligned} \tag{4.9}$$

The first equation in (4.9) has a solution of the form

$$\phi_1(\tau, y) = \Phi(y)\cos\tau. \tag{4.10}$$

To determine the profile $\Phi(y)$, we look at the second equation in (4.9). Substituting $\phi_1(\tau, y)$ into this equation, we get

$$\begin{aligned} \partial_\tau^2\phi_3 + \phi_3 = &\left[-\alpha^2\Phi + \partial_y^2\Phi + \frac{3}{4}\Phi^3 - \frac{5}{8}\Phi^5\right]\cos\tau, \\ &+ \left[\frac{1}{4}\Phi^3 - \frac{5}{16}\Phi^5\right]\cos 3\tau - \frac{1}{16}\Phi^5\cos 5\tau. \end{aligned} \tag{4.11}$$

We are looking for solutions that are periodic in τ. The term $[\dots]\cos\tau$ will lead to a term linearly growing with τ. Hence, we must have

$$\partial_y^2\Phi - \alpha^2\Phi + \frac{3}{4}\Phi^3 - \frac{5}{8}\Phi^5 = 0. \tag{4.12}$$

This equation has a first integral, the "conserved energy"

$$E_y = \frac{1}{2}\left(\partial_y\Phi\right)^2 + \mathcal{U}(\Phi), \tag{4.13}$$

where $\mathcal{U}(\Phi) = -\frac{1}{2}\alpha^2\Phi^2 + \frac{3}{16}\Phi^4 - \frac{5}{48}\Phi^6$. If we demand spatially localized solutions, we require $E_y = 0$. Furthermore, requiring that the profile be smooth at the origin, we must have

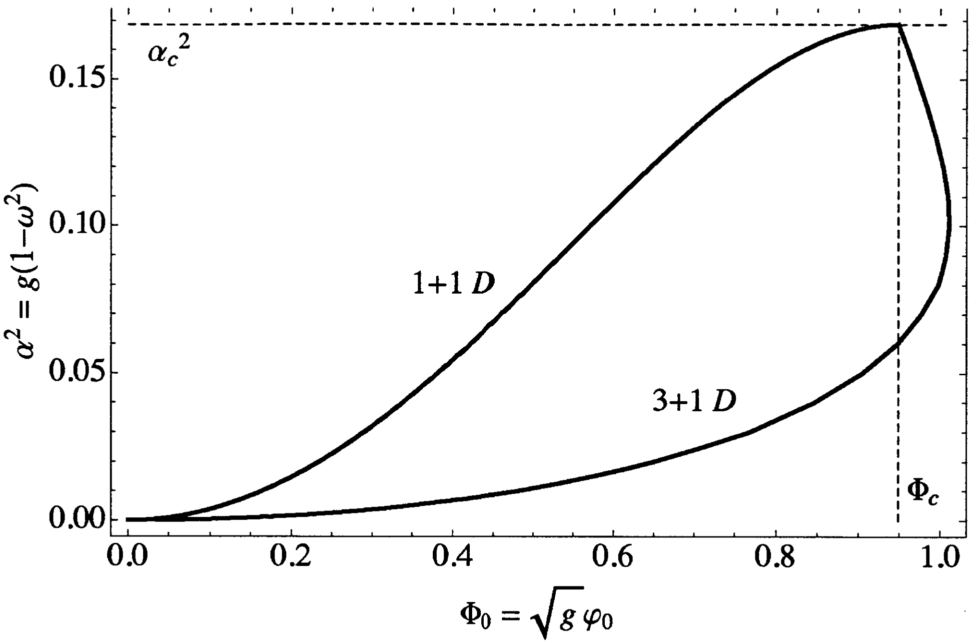

Figure 4-1: The above figure shows α^2 which characterizes the change in frequency of oscillation due to the nonlinearities in the potential. The critical $\alpha_c = \sqrt{27/160}$ can be obtained from the requirement that the nodeless solution is smooth and localized in space. Note that in 1+1 dimensions, α is monotonic in Φ_0. This is not the case in 3+1 dimensions. Frequency is measured in units of the m.

$\partial_y\Phi(0) = 0$. This immediately yields (also see Fig. 4-1)

$$\alpha^2 = \frac{3}{8}\Phi_0^2 - \frac{5}{24}\Phi_0^4. \tag{4.14}$$

Note that there is a critical value $\alpha_c = \sqrt{27/160}$ (or $\Phi_c = \sqrt{9/10}$) beyond which localized solutions do not exist. The profile equation becomes

$$(\partial_y\Phi)^2 = \alpha^2\Phi^2 - \frac{3}{8}\Phi^4 + \frac{5}{24}\Phi^6. \tag{4.15}$$

Integrating the above equation yields

$$\Phi(y) = \Phi_0\sqrt{\frac{1+u}{1+u\cosh[2\alpha y]}}, \tag{4.16}$$

where

$$
\begin{aligned}
u &= \sqrt{1-(\alpha/\alpha_c)^2}, \\
\Phi_0 &= \Phi_c\sqrt{1-u}.
\end{aligned}
\tag{4.17}
$$

We have introduced the variable $0 < u < 1$ which simplifies the appearance of the equations and controls the shape of the oscillons through α. We will come back to a more detailed analysis of this solution, but first we solve for the second order correction to this solution, ϕ_3:

$$\partial_\tau^2\phi_3 + \phi_3 = \left[\frac{1}{4}\Phi^3 - \frac{5}{16}\Phi^5\right]\cos 3\tau - \frac{1}{16}\Phi^5\cos 5\tau. \tag{4.18}$$

The solution with $\phi_3(0,y) = \partial_\tau\phi_3(0,y) = 0$ is

$$\phi_3(\tau,y) = \frac{1}{96}\left(3\Phi^3 - 4\Phi^5\right)\cos\tau + \frac{1}{128}\left(5\Phi^5 - 4\Phi^3\right)\cos 3\tau + \frac{1}{384}\Phi^5\cos 5\tau. \tag{4.19}$$

The full solution becomes

$$\phi(\tau,y) = \Phi\cos\tau + \frac{\Phi^3}{24g}\left[\frac{1}{4}\left(3-4\Phi^2\right)\cos\tau - \frac{3}{16}\left(4-5\Phi^2\right)\cos 3\tau + \frac{1}{16}\Phi^2\cos 5\tau\right]. \tag{4.20}$$

Note that the corrections to ϕ_1 are strongly suppressed for $g \gg 1$. Even for moderately large $g \sim 5$, the factor in the denominator is ~ 100, making ϕ_1 a rather good approximation. From now on we will mainly concern ourselves with ϕ_1.

Reverting back to the original variables (4.6), the solution for $\alpha < \alpha_c$ (equivalently, $\Phi_0 < \Phi_c$) is

$$\varphi(t,x) = \varphi_0\sqrt{\frac{1+u}{1+u\cosh[2\alpha x/\sqrt{g}]}}\cos(\omega t) + \mathcal{O}[g^{-3/2}], \tag{4.21}$$

where

$$
\begin{aligned}
u &= \sqrt{1-(\alpha/\alpha_c)^2}, \\
\varphi_0 &= \frac{\Phi_0}{\sqrt{g}} = \frac{\Phi_c}{\sqrt{g}}\sqrt{1-u}, \\
\omega^2 &= 1 - g^{-1}\alpha^2.
\end{aligned}
\tag{4.22}
$$

Here, φ_0 is the amplitude of the profile at the origin and scales as $1/\sqrt{g}$.

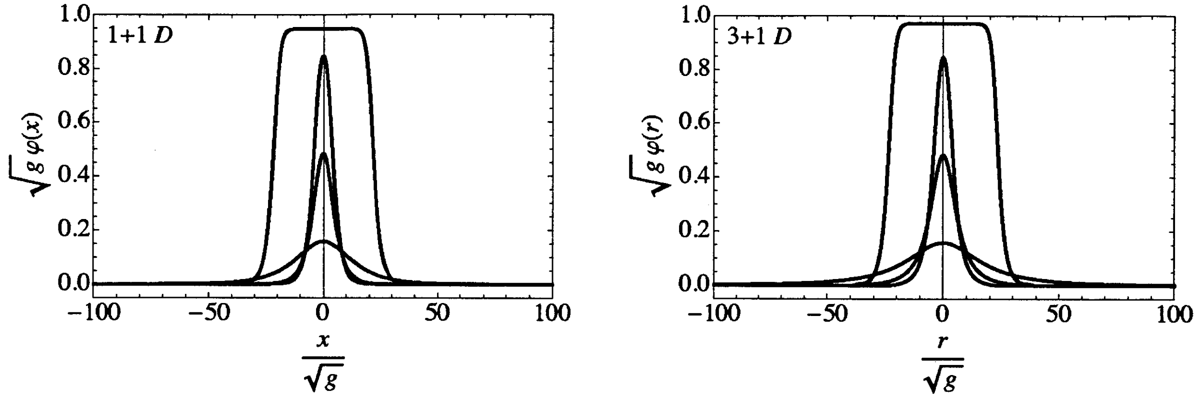

Figure 4-2: The above figure shows the spatial profiles of oscillons for different values of the amplitude at the center. For $\Phi_0 \ll \Phi_c = \sqrt{9/10}$ we get the usual sech-like profile, which is consistent with the small amplitude analysis. As Φ_0 approaches Φ_c, the oscillons become wider with surprisingly flat tops. Unlike the $1+1$-dimensional case, in $3+1$ dimensions, we approach the flat-top profiles from above. Distances are measured in units of the m^{-1}.

Let us now investigate the solution for the profile. Figure 4-2 (top left) shows this solution for different valued of α. Notice that as α approaches α_c (equivalently, $\Phi_0 \to \Phi_c$, $u \to 0$), the oscillon profile begin to deviate from the "sech" profile and has a flat top. Given this solution, one can derive the width of the oscillon as a function of its height. Defining the width to be the x value where the profile falls by $1/e$ of its maximum

$$x_e = \frac{1}{\varphi_0}\frac{2}{\sqrt{3}}(1+u)^{-1/2}\cosh^{-1}\left[\frac{e^2(1+u)-1}{u}\right]. \tag{4.23}$$

As $u \to 0$ we simply have $x_e \sim 1/\varphi_0$, which is consistent with the small amplitude analysis (see Fig. 4-3). Meanwhile, $u \to 1$ yields a spatially uniform solution.

We end this subsection by writing down an expression for the energy of these oscillons:

$$E_{osc} = \varphi_0 \frac{4}{\sqrt{3(1-u)}}\tanh^{-1}\left[\sqrt{\frac{1-u}{1+u}}\right] + \mathcal{O}[g^{-3/2}]. \tag{4.24}$$

Note that as $\alpha \to 0$ ($u \to 1$) we have $E_{osc} \sim 2\sqrt{2/3}\varphi_0$ whereas for $\alpha \to \alpha_c$ ($u \to 0$), $E_{osc} \to \infty$. In the next subsection we extend our results to $3+1$ dimensions.

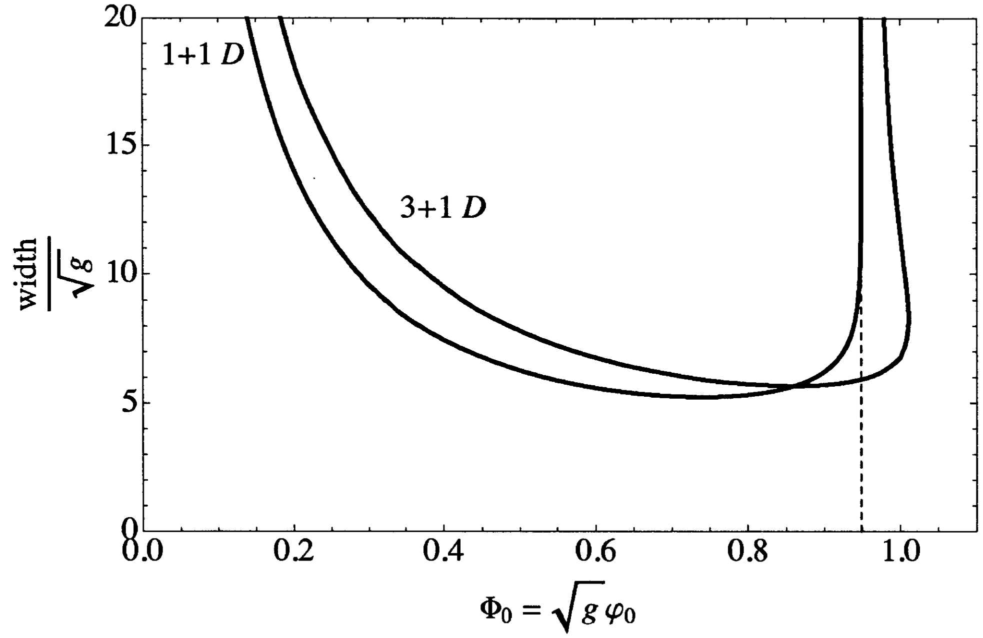

Figure 4-3: The above figure shows the nonmonotonic relationship between the width and height of oscillons in 1+1 and 3+1 dimensions. Note that as Φ_0 approaches $\Phi_c = \sqrt{9/10}$, the oscillons become wider with flat tops. Unlike the 1+1-dimensional case, in 3+1 dimensions, we obtain flat-top profiles when Φ_0 approaches Φ_c from above. Distances are measured in units of the m^{-1}.

4.3.2 Profile and frequency in $3+1$ dimensions

In this section we extend the results of the previous section to a 3+1-dimensional Minkowski space-time. Although we are unable to obtain an analytic form for the profile, the important qualitative (and some quantitative) aspects of the solutions can still be understood. In particular, we derive a critical amplitude and frequency for which the solution becomes spatially homogeneous and argue that the relationship between the height and width is nonmonotonic.

The equation of motion (assuming spherical symmetry) is given by

$$\partial_t^2\varphi - \partial_r^2\varphi - \frac{2}{r}\partial_r\varphi + \varphi - \varphi^3 + g\varphi^5 = 0. \tag{4.25}$$

We can follow the same procedure used in the previous subsection to arrive at the equation for the profile

$$\partial_\rho^2\Phi + \frac{2}{\rho}\partial_\rho\Phi - \alpha^2\Phi + \frac{3}{4}\Phi^3 - \frac{5}{8}\Phi^5 = 0, \tag{4.26}$$

where $\rho = r/\sqrt{g}$. This is where we first encounter the difficulty associated with three dimensions. We can no longer obtain a first integral due to the $2/\rho(\partial_\rho\Phi)$ term. However, we can still get a bound on α by requiring that the solutions are spatially localized (see [7] for an analysis of a similar profile equation in the context of Q-balls). It is convenient to define an energy E_ρ, which in the absence of the $(2/\rho)\partial_\rho\Phi$ term, is a constant of motion:

$$E_\rho = \frac{1}{2}\left(\partial_\rho\Phi\right)^2 + \mathcal{U}(\Phi), \tag{4.27}$$

where $\mathcal{U}(\Phi) = -\frac{1}{2}\alpha^2\Phi^2 + \frac{3}{16}\Phi^4 - \frac{5}{48}\Phi^6$. With this definition the equation of motion takes on an intuitive form

$$\frac{dE_\rho}{d\rho} = -\frac{2}{\rho}\left(\partial_\rho\Phi\right)^2. \tag{4.28}$$

This means that as we move away from $\rho = 0$, we move from a higher E_ρ trajectory to a lower one. With the requirement that the solution is "localized" (more specifically, $\Phi \propto \rho^{-1}e^{-\alpha\rho}$ as $\rho \to \infty$), we need $E_\rho \to 0$ as $\rho \to \infty$. Requiring that the solution is smooth at $\rho = 0$ requires $\partial_\rho\Phi = 0$ at $\rho = 0$. This implies that for a localized solution we must have $E_\rho \geq 0$. Equivalently, $U(\Phi_0) \geq 0$, which in turn implies that $\alpha \leq \alpha_c = \sqrt{27/160}$. For this critical value α_c, we get a special solution [with $\Phi_c(\rho) = \sqrt{9/10}$], which is homogeneous in space. For $0 < \alpha < \alpha_c$ we get nonzero spatial derivatives.

For each α in the range $0 < \alpha < \alpha_c$, only special, discrete values of $\Phi_0(n)$ will yield solutions that satisfy our requirement $\Phi \propto \rho^{-1}e^{-\alpha\rho}$ as $\rho \to \infty$. Here n enumerates the discrete number of nodes for the oscillon. From these, the $n = 0$ ones are the oscillon profiles we are looking for. The numerically obtained profiles are shown on the right in Fig. 4-2.

From figure 4-3, it is easy to see that the relationship between the heights and widths of the oscillons is nonmonotonic. We know that for $\alpha \ll \alpha_c$ (ie. $\Phi_0 \ll \Phi_c$), the usual small amplitude expansion yields solutions that have the property that their widths decrease with increasing amplitude. We also know that for $\alpha = \alpha_c$ ($\Phi_0 = \Phi_c$) the width will be infinite.

Thus, as in the $1+1$-dimensional scenario, we expect the width to be a nonmonotonic function of the central amplitude. This is indeed what is seen from the numerical solutions of the profile equation as shown in Fig. 4-3. Note that the width is a multivalued function of the amplitude beyond $\Phi_0 = \Phi_c$. Nevertheless, it still approaches the homogeneous solution via the flat-top profiles. The multivalued relationship between α and Φ_0 is shown in Fig. 4-1.

4.3.3 Radiation

The oscillon solution does not solve the equation of motion exactly. We have ignored terms of $\mathcal{O}[g^{-3/2}]$ as well as outgoing radiation. The problem of calculating the outgoing radiation in the small amplitude limit (not the flat tops) has been addressed in the literature (see [92, 30]). Our intention here is to point out that for flat tops, the radiation will still be small.

As shown in [54], the amplitude of the outgoing radiation can be estimated by the amplitude of the Fourier transform of the oscillon at the radiation wave number $k_r \sim \sqrt{8}m$ (also see [39]). For small amplitude oscillons, this is exponentially small $\sim e^{-1/\varphi_0}$. Let us estimate what changes are expected when we move to the flat-top oscillons. As we have seen, already our solutions have the form,

$$\varphi(t,x) = \frac{1}{\sqrt{g}}\phi\left(\tau, \frac{x}{\sqrt{g}}\right). \tag{4.29}$$

Where the function $\phi(\tau, y)$ is independent of g. The Fourier transform of φ can be determined from the Fourier transform of ϕ using

$$\varphi(t,k) = \phi(\tau, \sqrt{g}k). \tag{4.30}$$

Now $\phi(t,x)$ is determined entirely by α. Hence, for any given α, the Fourier transform of φ gets narrower as g is increased. Thus, by increasing g we can make the amplitude at the radiating wave number as small as we want. Note that even though the Fourier transform of a flat-top oscillon resembles a "sinc" function, rather than a sech, this is true only for wave

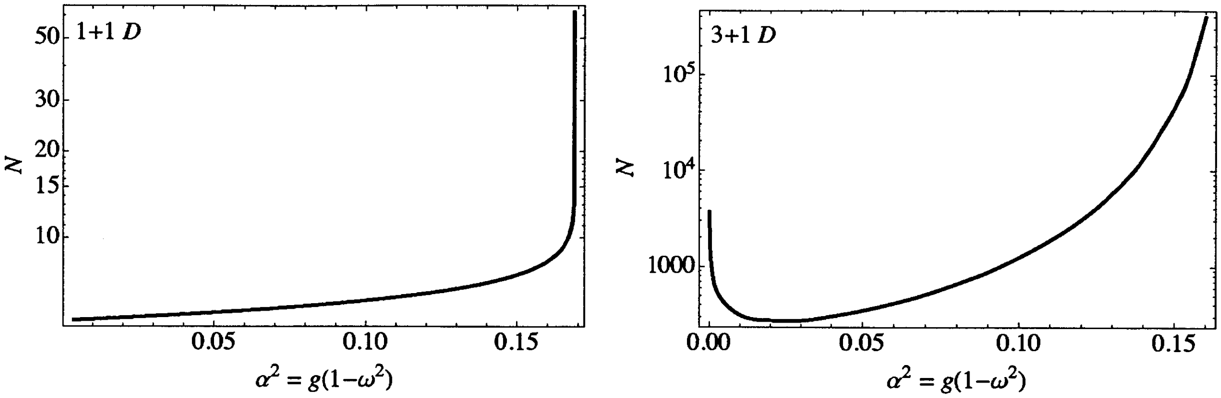

Figure 4-4: In the above figure we plot $N = \int \Phi^2 d\mathbf{x}$ for the oscillons in $1+1$ (left) and $3+1$(right) dimensions. The stability of oscillons is determined by the sign of $dN/d\alpha^2$ where $\alpha^2 = g(1-\omega^2)$. Note the important difference between the curves in the two cases. While all oscillons are stable to long wavelength perturbations in $1+1$ dimensions, this is not the case in $3+1$ dimensions. Only those with small frequency (or equivalently, towards the flat-top regime) are robust. Frequency is measured in units of m.

numbers near zero. Since the flat top oscillons are smooth solutions, their Fourier transforms still exhibit a rapid asymptotic decay. The argument in 3+1 dimensions will be similar.

4.4 Linear Stability Analysis

In this section we investigate whether oscillons are stable against small, localized perturbations. As discussed in the previous section, the periodic oscillon expansion, formulated in powers of $g^{-1/2}$, fails to solve the governing field equations and must expel radiation. In our stability analysis we ignore the effects of the exponentially suppressed radiation and focus on perturbing the oscillon profile. The main results of this section are as follows: (i) On the time scale of order g, $3+1$-dimensional oscillons with large amplitudes are robust (their small amplitude counterparts are not) against localized perturbations with spatial variations comparable to the width of the oscillon. (ii) For small wavelength perturbations (compared to the width of of the oscillon), instabilities *could* exist in discrete, extremely narrow bands in k space.

We now provide the details essential for reaching the above conclusions. As done previously, we discuss the 1 + 1-dimensional case first, and then extend the results to 3 + 1 dimensions. Starting with a fixed oscillon profile φ_{osc} [see Eq. (4.21)], we linearize about the oscillon by an arbitrary function χ. Provided the field χ remains smaller than φ_{osc}, the linearized dynamics will approximately describe the perturbation. Let

$$\varphi(t,x) = \varphi_{osc}(\tau,x) + \delta\chi(t,x), \tag{4.31}$$

where $\delta \ll g^{-1/2}$ is the amplitude of the perturbation and we keep $\chi \sim O(1)$. Note that for a linear analysis, χ must also vanish at infinity so that the perturbation $\delta\chi$ remains smaller than the original oscillon. Therefore, we restrict our analysis to spatially localized perturbations. The field χ then satisfies

$$\partial_t^2\chi - \partial_x^2\chi + \chi - 3\varphi_{osc}^2\chi + 5g\varphi_{osc}^4\chi = 0. \tag{4.32}$$

We now wish to determine if all initial conditions χ remain bounded, or whether there exists an unstable initial profile $\chi(0,x)$. The $\varphi_{osc}^2(t,x), \varphi_{osc}^4(t,x)$ terms act as periodic forcing functions[1]. This periodic forcing, somewhat analogous to pumping ones legs back and forth on a swing, may deposit energy into the field χ and consequently excite an instability.

A complete treatment of stability may require one to solve (4.32) for a complete basis of initial conditions. Because of the spatially dependent oscillon solution, a Fourier analysis is difficult. With this in mind, we split the set of initial conditions into two groups. The first with spatial variations comparable to the size of the oscillons and another which varies on much shorter length scales. In the second case we can approximate the oscillon as a spatially constant oscillating background. This allows us to carry out a standard Floquet analysis. Such an analysis reveals the most dangerous instability band at $k \sim \sqrt{3}$, with a width $\Delta k \lesssim g^{-1}\Phi_0^4$. For large g, this becomes extremely narrow. In addition, we expect the time scale of these instabilities to be $\sim g\Phi_0^{-4}$. Nevertheless, one should bear in mind that the slow spatial variation of the oscillon could still be important.

---

[1] Since ϕ_{osc} oscillates in time, the problem is essentially one of parametric resonance stability/instability. We are really diagonalizing the Floquet matrix - which in this case would really be an integral operator.

Now, let us look at the case where the perturbations vary on length scales comparable to the width of the oscillon in detail. Note that to leading order, the forcing potential: $\sim \varphi_{osc}^2$ is (i) $\mathcal{O}[g^{-1}]$, (ii) smoothly varying with a natural length $x_e \propto \sqrt{g}$, and (iii) oscillating with period 1 in the variable $\tau = \omega t$. The first observation implies we may use perturbation theory and seek an expansion for χ in inverse powers of g:

$$\chi = \chi_0 + g^{-1}\chi_1 + \ldots \tag{4.33}$$

In the following analysis we shall work out the linear instabilities to first order in g^{-1}. However, we must keep in mind that solutions stable to order g^{-1}, may in fact develop higher order instabilities over longer time scales. Since we are interested in perturbation with wavelengths comparable to x_e, we rescale the length $x = \sqrt{g}y$. To capture the instability, however, we introduce two times: the original oscillatory time $\tau = \omega t$ and a slow time $T = g^{-1}t^2$. The introduction of τ follows from our focus on perturbations which oscillate near the oscillon frequency. In addition, we require a slow time T to capture variations in the perturbation. Hence, the field $\chi = \chi(\tau, T, y)$ and the derivative ∂_t becomes

$$\begin{aligned} \partial_t &= \omega\partial_\tau + g^{-1}\partial_T, , && (4.34)\\ \partial_t^2 &= \partial_\tau^2 + g^{-1}[2\partial_T\partial_\tau - \alpha^2\partial_\tau^2] + \mathcal{O}[g^{-2}]. && (4.35)\end{aligned}$$

Upon substitution of Eqs. (4.33), (4.34) and (4.35) in Eq. (4.32) and collecting powers of g^{-1} we obtain

$$\begin{aligned} \partial_\tau^2\chi_0 + \chi_0 &= 0, && (4.36)\\ \partial_\tau^2\chi_1 + \chi_1 &= -[2\partial_T\partial_\tau - \alpha^2\partial_\tau^2 - \partial_y^2 - 3\cos^2\tau\Phi^2(y) + 5\cos^4\tau\Phi^4(y)]\chi_0. && (4.37)\end{aligned}$$

---

[2] One may want to know why $T = g^{-1}t$ provides the important slow time scale. A back of the envelop calculation is as follows - consider a homogeneous background oscillating at the oscillon frequency. A naive perturbation series for χ in powers of g^{-1} exhibits an oscillating term for χ_0, and then a term which grows linearly in t for χ_1. This means χ_1 becomes the same order as χ_0 when $tg^{-1} \sim 1$. Hence, there is a characteristic slow time $T = tg^{-1}$. If we let χ_0 be a function of both τ and T, and we choose χ_0 correctly, we can ensure that χ_1 remains small for long times. In a similar fashion, if we consider higher order terms in the oscillon expansion, there could be additional instabilities excited over time scales g^{-2}, g^{-3} etc.

From the zeroth order equation, the most general solution for $\chi_0(\tau, T, y)$ is

$$\chi_0(\tau, T, y) = u(T, y)\cos\tau + v(T, y)\sin(\tau) \tag{4.38}$$

Here, $u(T, y)$ and $v(T, y)$ are real functions that depend on the slow time T and space y. Eliminating the secular terms from the right hand-side of the χ_1 equation, we obtain

$$2\partial_T u = Lv, \tag{4.39}$$

$$2\partial_T v = -Mu, \tag{4.40}$$

where the L and M are both Hermitian operators. Explicitly,

$$L = -\partial_y^2 + \alpha^2 - \frac{3}{4}\Phi^2(y) + \frac{5}{8}\Phi^4(y), \tag{4.41}$$

$$M = -\partial_y^2 + \alpha^2 - \frac{9}{4}\Phi^2(y) + \frac{25}{8}\Phi^4(y). \tag{4.42}$$

Since Eqs. (4.39) and (4.40) are linear, we can separate variables via $u(T, y) = e^{\frac{1}{2}\Omega T}u(y)$, $v(T, y) = e^{\frac{1}{2}\Omega T}v(y)$:

$$\Omega u = Lv, \tag{4.43}$$

$$\Omega v = -Mu, \tag{4.44}$$

or equivalently

$$\Omega^2 u = -LMu, \tag{4.45}$$

$$\Omega^2 v = -MLv. \tag{4.46}$$

$$\tag{4.47}$$

Since both u and v are real fields and L and M are real operators, the eigenvalues[3] Ω^2 must

---

[3] If they exist.

also be real. Hence, all exponents Ω are either purely real or purely imaginary. Then, oscillon stability is guaranteed when $max(\Omega^2) < 0$, or equivalently when the largest real eigenvalue of $-ML$ is negative. Determining the largest real eigenvalue of $-ML$ can be done using the analysis performed by Vakhitov and Kolokolov [102]. Specifically, they exploit properties of the operator potentials found in L and M to show that $max(\Omega^2) < 0$ if and only if $dN/d\alpha^2 > 0$. Here, N is the integral over all space:

$$N = \int \Phi^2(y)dy \tag{4.48}$$

and $\alpha^2 = g(1-\omega^2)$. From Fig. 4-4, we can see that $dN/d\alpha^2 > 0$ for all allowed α in $1+1$ dimensions. Thus, $1+1$-dimensional oscillons are stable against small perturbations with long wavelengths. Note that to order g^{-1}, $N = 2g^{1/2}E_{osc}$ in $1+1$ dimensions (in $3+1$ dimensions $N = 2g^{-1/2}E_{osc}$.)

The argument of [102] holds for dimensions $D = 1, 2, 3$. The discussion above carries over to $3+1$ dimensions through the following identifications: $y \to \rho$ and $\partial_y^2 \to \partial_\rho^2 + (2/\rho)\partial_\rho$ and $N = 4\pi \int \Phi^2(\rho)\rho^2 d\rho$. The result in $3+1$ dimensions is in sharp contrast with that in $1+1$ dimensions (see Fig. 4-4). Unlike the $1+1$-dimensional result, not all oscillons are robust against long wavelength perturbations. Only oscillons with large α (equivalently small frequency or large amplitudes) are robust. This result makes the large amplitude, flat-topped oscillons in $3+1$ dimensions particularly interesting.

In the context of Q-balls, N is proportional to the conserved particle number and plays a role in the stability [69]. A similar interpretation might be possible here, since to leading order in $g^{-1/2}$, our solution is periodic in time. Finally, we note that the behavior of N in $1+1$ and $3+1$ dimensions can be understood heuristically. In $1+1$ dimensions, for small α, the amplitude of the profile at the origin $\sim \alpha$ whereas the width $\sim 1/\alpha$. Hence, $N \sim \alpha$. For $\alpha \to \alpha_c$ we have increasingly wide oscillons with amplitudes $\sim \Phi_c$. Hence, N diverges. Now for $3+1$ dimensions, the behavior at $\alpha \to \alpha_c$ is similar to the $1+1$-dimensional case. However, at small α, due to the different spatial volume factor, we get $N \sim 1/\alpha$, therefore implying a nonmonotonic behavior in N. Based on a numerical analysis in the case of dilatonic scalar fields, it was conjectured in [30], that the stability of oscillon like configurations is related to the slope of the E_{osc} vs amplitude curve. In the large g limit, $E_{osc} \propto N$ and the amplitude $\propto \alpha$. Hence, their conjecture is in agreement with our analytic result.

4.4.1 Spectral stability

Here we review the linear stability argument of Vakhitov and Kolokolov[102]. Specifically, the system

$$\begin{aligned} \Omega^2 u = & \ -LMu \\ \Omega^2 v = & \ -MLv \end{aligned} \tag{4.49}$$

where

$$L = -\partial_y^2 + \alpha - \frac{3}{4}\Phi^2(y) + \frac{5}{8}\Phi^4(y) \tag{4.50}$$

$$M = -\partial_y^2 + \alpha - \frac{9}{4}\Phi^2(y) + \frac{25}{8}\Phi^4(y) \tag{4.51}$$

admits no solutions $\Omega^2 > 0$ when $dN/d\alpha > 0$, for

$$N = \int \Phi(y)^2 dy. \tag{4.52}$$

Computationally, the result provides a dramatic simplification in determining stability of the perturbed system. For example, analysis of equation (4.49) requires one to solve an eigenvalue problem, while the result of [102] reduces the problem to calculating integrals. Although the proof requires several technical details involving the operators L and M, the general idea is to relate the signs of $max(\Omega^2)$ and $dN/d\alpha$ by use of a variational calculation.

The first step in the calculation requires converting equation (4.49) into a variational problem:

$$min(\Omega^2) = min \frac{\langle u|M|u\rangle}{\langle u|L^{-1}|u\rangle} \tag{4.53}$$

$$\langle u|\Phi\rangle = 0 \tag{4.54}$$

The product (4.53) follows from the observation that $L\Phi = 0$. Specifically, since $\Omega^2\langle\Phi|u\rangle = -\langle\Phi|LM|u\rangle = 0$, then either $\Omega^2 = 0$ or $\langle\Phi|u\rangle = 0$. Hence, to determine the sign of the smallest nonzero Ω^2, we can restrict ourselves to functions u such that $\langle u|\Phi\rangle = 0$. Now since L is Hermitian, and Φ is a radial function with zero nodes, it follows that Φ is also the ground mode of L. Therefore, L is positive definite and invertible on the space of functions orthogonal to Φ. As a result, the product $\langle u|L^{-1}|u\rangle$ is a well defined norm (squared) on functions u, and thus by the variational principle $min(\Omega^2)$ is equivalent to the right hand side of equation (4.53).

The second step of the calculation involves showing the sign of the variational product (4.53) is related to the sign of a simpler product $\langle\Phi|M^{-1}|\Phi\rangle$. This procedure requires two technical observations concerning the potential in M. Namely, for a large class of potentials, as shown in section 4.4.2, the radially symmetric modes have exactly one eigenvalue less than zero, while the first angular eigenvector has eigenvalue zero. To determine the sign of $min(\Omega^2)$, we use the fact that the product $\langle u|L^{-1}|u\rangle$ is positive definite, and therefore without loss of generality, we can replace the denominator in (4.53) with $\langle u|u\rangle$. The problem of finding

$$sign\ min(\Omega^2) = sign\ min\frac{\langle u|M|u\rangle}{\langle u|u\rangle} \tag{4.55}$$

$$\langle u|\Phi\rangle = 0 \tag{4.56}$$

then becomes equivalent to minimizing $\langle u|M|u\rangle$ subject to the constraints $\langle u|u\rangle = 1$ and $\langle u|\Phi\rangle = 0$. Introducing appropriate Lagrange multipliers λ and α we have:

$$Mu = \lambda u + \alpha\Phi \tag{4.57}$$

Taking the inner product $\langle u|$ to both sides of equation (4.57), shows that $min\langle u|M|u\rangle$ corresponds to the smallest Lagrange multiplier solution λ. To determine the sign of the smallest such λ, expand both u and Φ in an eigenbasis of M:

$$M\psi_n = \lambda_n \psi_n \tag{4.58}$$

$$u = \sum_n d_n \psi_n \tag{4.59}$$

$$\Phi = \sum_n c_n \psi_n \tag{4.60}$$

A direct substitution of u and Φ into (4.57) yields the relation $d_n = \alpha \frac{c_n}{\lambda_n - \lambda}$ between coefficients d_n and c_n. Lastly, by enforcing the constraint $\langle u|\Phi\rangle = 0$ we obtain an equation for the Lagrange multipliers:

$$\alpha \sum_n \frac{c_n^2}{\lambda_n - \lambda} = 0 \tag{4.61}$$

Equation (4.61) admits the solution $\alpha = 0$, as well as a countable number of λ, of which we are concerned with the sign of λ_{min}. We first show that $\alpha = 0$ corresponds to stable perturbations by exploiting properties of the operator M. Specifically, $\alpha = 0$ implies λ and u are eigenvalues and eigenvectors of M. There are now two cases to consider: u radially symmetric and u non-radially symmetric. In the former, M contains only one negative eigenvector corresponding to the ground mode. Hence, this ground eigenvector, having no nodes, would have a nonzero overlap with Φ, and therefore fail to satisfy the constraint imposed on u. It follows that any radially symmetric u would correspond to eigenvalues with positive λ. In the latter case, every non-symmetric eigenvector has a positive definite eigenvalue, again implying $\lambda \geq 0$. Therefore all $\alpha = 0$ solutions yield stable perturbations. For $\alpha \neq 0$, λ must solve:

$$g(\lambda) = \sum_n \frac{c_n^2}{\lambda_n - \lambda} \tag{4.62}$$

$$g(\lambda) = 0 \tag{4.63}$$

Note that every nonsymmetric eigenfunction is orthogonal to the space of radially symmetric functions. Hence, every non-symmetric eigenfunction has coefficient $c_n = 0$. In addition, $g(\lambda)$ varies continuously, and monotonically between the poles at eigenvalues λ_n. The minimum root λ_{min} therefore occurs somewhere between the smallest two (radially symmetric) eigenvalues $\lambda_1 < 0$ and $\lambda_2 > 0$. Since $g(\lambda)$ varies monotonically, we can relate the sign of $g(0)$ to the sign of λ_{min}. Specifically, $g(0) < 0$ implies $\lambda_{min} > 0$, while $g(0) > 0$ yields $\lambda_{min} < 0$. Consequently, we have:

$$\begin{aligned} sign(\langle u|M|u\rangle_{min}) &= -sign(g(0)) && (4.64)\\ &= -sign[\sum_n \frac{|c_n|^2}{\lambda_n}] && (4.65)\\ &= -sign[\langle\Phi|M^{-1}|\Phi\rangle] && (4.66)\end{aligned}$$

The last step of the proof replaces the inverse M^{-1}, appearing in the inner product $\langle\Phi|M^{-1}|\Phi\rangle$, with the derivative $\frac{d}{d\alpha}$. To obtain the required relation, differentiate the equation $L\Phi = 0$ with respect to α:

$$\begin{aligned} M\frac{d\Phi}{d\alpha} + \Phi &= 0 && (4.67)\\ \frac{d\Phi}{d\alpha} &= -M^{-1}\Phi && (4.68)\end{aligned}$$

Hence we have,

$$\begin{aligned} \langle\Phi|M^{-1}|\Phi\rangle &= -\langle\Phi|\frac{d}{d\alpha}|\Phi\rangle && (4.69)\\ &= -\frac{1}{2}\frac{d}{d\alpha}\langle\Phi|\Phi\rangle && (4.70)\end{aligned}$$

Equivalently, the oscillon is stable, with $\lambda_{min} > 0$, if and only if:

$$\frac{d}{d\alpha}\langle\Phi|\Phi\rangle > 0 \qquad (4.71)$$

4.4.2 Some properties of the operator M

Here we verify that the operator M admits a zero eigenvalue for the lowest non-symmetric angular mode, as well as exactly one negative eigenvalue for the radially symmetric modes. We work in 3D and begin by noting that the operators

$$\begin{aligned} L &= -\partial_\rho^2 - \frac{2}{\rho}\partial_\rho + \alpha - \frac{3}{4}\Phi^2(\rho) + \frac{5}{8}\Phi^4(\rho) && (4.72) \\ M &= -\partial_\rho^2 - \frac{2}{\rho}\partial_\rho + \alpha - \frac{9}{4}\Phi^2(\rho) + \frac{25}{8}\Phi^4(\rho) && (4.73) \end{aligned}$$

are approximately related by derivatives. Specifically, in the case of the non-symmetric angular mode, we can differentiate the equation $L\Phi(\rho)$ through by ρ:

$$\begin{aligned} \partial_\rho\left(-\partial_\rho^2\Phi - \frac{2}{\rho}\partial_\rho\Phi + \alpha\Phi - \frac{3}{4}\Phi^3 + \frac{5}{8}\Phi^5\right) &= 0 && (4.74) \\ \left(-\partial_\rho^2 - \frac{2}{\rho}\partial_\rho + \frac{2}{\rho^2} + \alpha - \frac{9}{4}\Phi^2 + \frac{25}{8}\Phi^4\right)\partial_\rho\Phi &= 0 && (4.75) \\ M(Y_m^{l=1}(\theta,\phi)\partial_\rho\Phi) &= 0 && (4.76) \end{aligned}$$

Here $Y_m^{l=1}(\theta,\phi)$ is the $l = 1$ spherical harmonic. Since $\partial_\rho\Phi$ contains no nodes, by construction the function $Y_m^{l=1}(\theta,\phi)\partial_\rho\Phi$ is the lowest non-symmetric angular eigenvector and has eigenvalue zero.

Next, to verify that M only contains one negative radial eigenvalue, we first note that the weak amplitude equation, $\alpha \to 0$:

$$\begin{aligned} -\partial_\rho^2\psi - \frac{2}{\rho}\partial_\rho\psi + \alpha\psi - \frac{3}{4}\Phi^2(\rho)\psi &= \epsilon\psi && (4.77) \\ & && (4.78) \end{aligned}$$

contains exactly one negative eigenvalue[4]. The goal now is to show that as alpha continuously increases, no eigenvalue ever crosses zero. We do so by studying solutions to the nonlinear

---

[4] For example, this has been checked numerically.

equation:

$$-\partial_\rho^2 f - \frac{2}{\rho}\partial_\rho f + \alpha f - \frac{3}{4}f^3 + \frac{5}{8}f^5 = 0 \qquad (4.79)$$

with initial data

$$f(0) = s \qquad (4.80)$$

$$\partial_\rho f(0) = 0 \qquad (4.81)$$

Here α is understood to be a fixed constant so that for each value of s, we obtain a solution to the differential equation. Moreover, there is a critical $s = \Phi_0$ which yields the zero node oscillon profile. For other values of s, even ones close to Φ_0, solutions $f(\rho) \to C$ as $\rho \to \infty$. In this case C is a fixed point of the differential equation and solves the algebraic equation:

$$\alpha - \frac{3}{4}C^2 + \frac{5}{8}C^4 = 0 \qquad (4.82)$$

We may now construct a function parameterized by α:

$$\psi_\alpha = \frac{d}{ds}f(\rho)|_{s=\Phi_0} \qquad (4.83)$$

The function ψ_α satisfies the boundary conditions $\partial_\rho \psi_\alpha(0) = 0$, $\psi_\alpha = 1$. In addition, ψ_α diverges as $y \to \infty$ since taking $s = \Phi_0 + \delta$ yields $|f_s(\rho) - f_{\Phi_0}(\rho)| \to C$. Lastly, by differentiating equation (4.79) with respect to s and evaluating at $s = \Phi_0$, we have:

$$M\psi_\alpha = 0 \qquad (4.84)$$

Hence, any function ψ which satisfies the boundary condition $\partial_\rho \psi(0) = 0$ and corresponds to a zero eigenvalue must diverge as $\rho \to \infty$. Therefore, M has no bounded radially symmetric, zero eigenfunctions. Equivalently M must have exactly one negative eigenvalue for all α.

4.5 Including Expansion

In this section we consider the effects of expansion on the lifetimes and shapes of oscillons. We closely follow the procedure provided in [24] for the small amplitude oscillons. Here, applying their procedure is somewhat subtle since in the limit $g \gg 1$, oscillons tend to be very wide ($\propto \sqrt{g}$), and the width grows without bound when $\alpha \to 0, \alpha_c$. Consequently, in these regimes it is easier to break up the oscillons due to Hubble horizon effects. Nevertheless we construct approximate solutions when the oscillon width is small compared to the Hubble horizon.

4.5.1 Including expansion in $1+1$ dimensions

As before, we begin with $1+1$ dimensions and generalize to $3+1$ dimensions. We will work in static deSitter co-ordinates where the metric is given by

$$ds^2 = -(1 - x^2H^2)dt^2 + (1 - x^2H^2)^{-1}dx^2. \tag{4.85}$$

Here, H is a constant Hubble parameter[5]. In these co-ordinates, the equation of motion becomes

$$(1 - x^2H^2)^{-1}\partial_t^2\varphi + 2xH^2\partial_x\varphi - (1 - x^2H^2)\partial_x^2\varphi = -V'(\varphi), \tag{4.86}$$

where $(xH) < 1$. We will assume that $H \ll 1$ and that $H = \bar{H}/g$ where $\bar{H}$ is a small number. The effects of expansion can be ignored when $x \ll H^{-1}$. For oscillons with widths satisfying $x_e(\alpha) \ll H^{-1}$, the solution to the above equation is well approximated by the Minkowski space solution. However, in the tail of the oscillon profile we cannot ignore the effects of expansion. Nevertheless, taking advantage of the exponential decay of the profile in the tails, we can linearize Eq. (4.86) and obtain a solution using the WKB approximation.

We carry out the change of space-time variables and redefinition of the field as was done in the nonexpanding case, Eq. (4.6). Again collecting powers of g, we get

---

[5] The assumption of H being constant is for simplicity. The analysis carries over to a time dependent H as long as the frequency of oscillation $\omega \gg H$.

$$\begin{aligned}&\partial_\tau^2\phi_1+\phi_1=0,\\&\partial_\tau^2\phi_3+\phi_3=\{-\alpha^2+y^2\bar{H}^2\}\partial_\tau^2\phi_1-\partial_y^2\phi_1-\phi_1^3+\phi_1^5.\end{aligned}\tag{4.87}$$

In the case of the Minkowski background, we chose an initial condition $\partial_t\varphi(0,x)=0$, which picked out one of the two linearly independent solutions of the first equation in (4.87). However, in the expanding universe we need to keep the general solution

$$\phi_1(\tau,y)=\frac{\Phi(y)}{2}e^{-i\tau}+c.c,\tag{4.88}$$

where $\Phi(y)$ can be complex and *c.c* stands for complex conjugate. The "profile" equation is given by

$$\left\{\alpha^2-(y\bar{H})^2\right\}\Phi-\partial_y^2\Phi-\frac{3}{4}|\Phi|^2\Phi+\frac{5}{8}|\Phi|^4\Phi=0\tag{4.89}$$

and includes the effect of expansion through the $(y\bar{H})^2$ term. We now analyze different regimes as seen in Fig. 4-5. For $(y\bar{H})^2\ll\alpha^2$, the equation admits solutions identical to the nonexpanding case [see Eq. (4.21)]. In the region $y_e(\alpha)\ll y\ll\alpha\bar{H}^{-1}$, where $y_e(\alpha)$ is the approximate width of the oscillon [Eq. (4.23)], the profile has the form

$$\Phi(y)\approx\Phi_0\sqrt{\frac{2(1+u)}{u}}\exp\left[-\alpha y\right]\qquad y_e(\alpha)\ll y\ll\alpha\bar{H}^{-1}.\tag{4.90}$$

Since this is an exponentially decaying solution, we can ignore the nonlinear terms in the potential when $y\gg y_e(\alpha)$:

$$\partial_y^2\Phi+\left\{(y\bar{H})^2-\alpha^2\right\}\Phi\approx 0\qquad y_e(\alpha)\ll y.\tag{4.91}$$

For $y>\alpha\bar{H}^{-1}$, the above equation has a WKB solution[6] in the form of an outgoing wave:

$$\Phi(y)\approx\Phi_0\sqrt{\frac{2\alpha(1+u)}{u\bar{H}y}}\exp\left[-\frac{\pi\alpha^2}{4\bar{H}}+\frac{i}{2}\bar{H}y^2\right].\tag{4.92}$$

---

[6] assuming the WKB condition $\bar{H}/\alpha^2\ll 1$ is satisfied

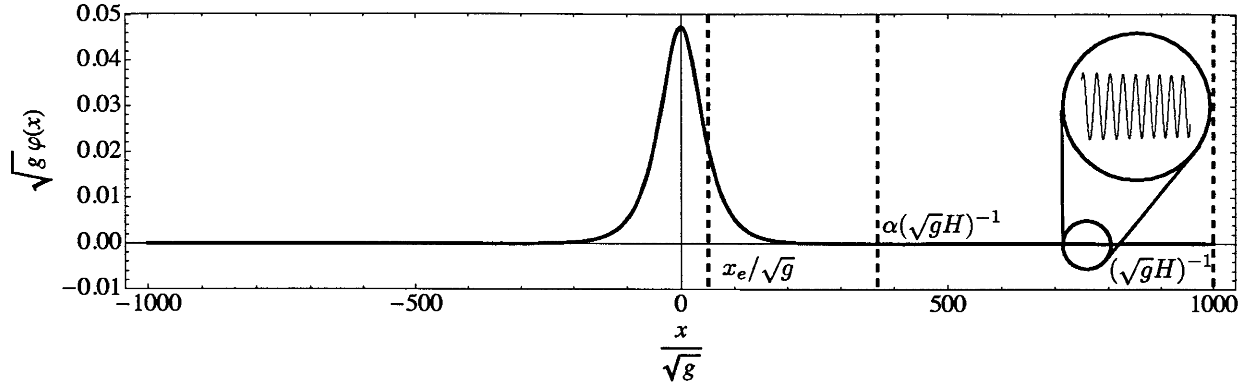

Figure 4-5: In an expanding background, if the width is small compared to H^{-1}, the flat space solution is adequate for distances much less than $\alpha(\sqrt{g}H)^{-1}$. For distances larger than this, but still smaller than H^{-1}, the oscillon feels the expansion, and loses energy in the form of outgoing waves (see inset in Fig. 4-5). Distance is measured in units of the m^{-1}.

The amplitude of the outgoing wave was chosen using the WKB connection formula to match the oscillon profile in Eq. (4.90). In terms of the original variables, we obtain

$$\begin{aligned} \varphi(t,x) &= \varphi_0\sqrt{\frac{1+u}{1+u\cosh\left[2\alpha x/\sqrt{g}\right]}}\cos(\omega t)+\mathcal{O}[g^{-3/2}] && |x|\ll\alpha(\sqrt{g}H)^{-1}, \\ \varphi(t,x) &= \varphi_0\sqrt{\frac{2(1+u)\alpha}{g^{1/2}H|x|u}}e^{-\frac{\pi\alpha^2}{4gH}}\cos\left[\omega t-\frac{1}{2}Hx^2\right] && \alpha(\sqrt{g}H)^{-1}\ll|x|<H^{-1}, \end{aligned} \tag{4.93}$$

where

$$\begin{aligned} u &= \sqrt{1-(\alpha/\alpha_c)^2}, \\ \varphi_0 &= \frac{\Phi_c}{\sqrt{g}}\sqrt{1-u}, \\ \omega^2 &= 1-g^{-1}\alpha^2. \end{aligned} \tag{4.94}$$

Our solution matches that of [24] in the limit $\alpha \ll \alpha_c$. However, as α gets larger the coefficient in front of the traveling wave captures the effects of the flat-top solutions. We will return to the above solution when we discuss the rate of energy loss by oscillons after considering the effects of expansion in $3+1$ dimensions.

We end the section by reminding ourselves of the assumptions required for this solution to be valid: (i) $g \gg 1$, (ii)$H < \mathcal{O}[g^{-1}]$, and (iii)

$$x_e(\alpha) \ll \alpha(\sqrt{g}H)^{-1}. \tag{4.95}$$

For any H, the solution is not valid when $\alpha \to 0$ or α_c. Also note that for a given $H \ll 1$, there always exists a g, which violates condition (iii) for all allowed α.

4.5.2 Including expansion in $3+1$ dimensions

Now, let us include the effects of expansion for the $3+1$-dimensional cases. The metric in the static deSitter co-ordinates (assuming spherical symmetry) is given by

$$ds^2 = -(1 - r^2H^2)dt^2 + (1 - r^2H^2)^{-1}dr^2 + r^2d\Omega^2. \tag{4.96}$$

Following a procedure similar to the one we laid out for the $1+1$-dimensional case, we get the profile equation:

$$\left\{\alpha^2 - (\rho\bar{H})^2\right\}\Phi - \partial_\rho^2\Phi - \frac{2}{\rho}\partial_\rho\Phi - \frac{3}{4}|\Phi|^2\Phi + \frac{5}{8}|\Phi|^4\Phi = 0, \tag{4.97}$$

where $\rho = r/\sqrt{g}$. The effect of expansion is included through the $(\rho\bar{H})^2$ term. For a given α, let the approximate width of the oscillon be $\rho_e(\alpha)$. In the region $\rho_e(\alpha) \ll \rho \ll \alpha\bar{H}^{-1}$, the profile has the form

$$\Phi(\rho) \approx f(\alpha)\frac{1}{\rho}\exp\left[-\alpha\rho\right] \qquad \rho_e(\Phi_0) \ll \rho \ll \alpha\bar{H}^{-1} \tag{4.98}$$

The lack of an analytic solution, prevents us from specifying $f(\alpha)$. Reverting back to the original variables, the solution in the spatially oscillatory regime is given by

$$\varphi(t,r) = (g^{-1/2}\alpha)^{1/2}\frac{f(\alpha)}{\sqrt{Hr^3}}e^{-\frac{\pi\alpha^2}{4gH}}\cos\left[\omega t - \frac{1}{2}Hr^2\right] \qquad \alpha(\sqrt{g}H)^{-1} \ll r < H^{-1}, \tag{4.99}$$

where $\omega^2 = 1 - g^{-1}\alpha^2$.

4.5.3 Energy loss due to expansion

In this subsection, we discuss the energy loss suffered by oscillons due to the expanding background. As before, we start with the $1+1$-dimensional scenario and then generalize to $3+1$ dimensions. The energy lost by an oscillon whose width is small compared to H^{-1} is given by

$$\frac{dE_{osc}}{dt} = -T^{x}_{t}|^{X}_{-X}, \tag{4.100}$$

where T^{μ}_{ν} is the energy momentum tensor of the scalar field. We have ignored the dependence of the metric on x. We take X to be in the region sufficiently far away from the center. More explicitly, we consider X such that

$$\alpha(\sqrt{g}H)^{-1} \ll |X| < H^{-1}. \tag{4.101}$$

Using Eq. (4.93) in Eq. (4.100), we get (to leading order in HX^2 and $g^{-1/2}$)

$$\frac{dE_{osc}}{dt} \approx -\varphi_0^3\left(\sqrt{\frac{3}{4}}\frac{(1+u)^{3/2}}{u}\right)\exp[-\pi\alpha^2/2gH] \tag{4.102}$$

In [24], a similar expression was provided in the limit where $u \to 1$. Oscillons are known to lose energy very slowly for a long time and then suddenly disintegrate. Our calculation cannot capture this disintegration. We therefore take the following as an upper bound on the lifetime of an oscillon (approximated by the time it take for it to lose most of its energy due to expansion effects)

$$\mathcal{T} \lesssim E_{osc}\left|\frac{dE_{osc}}{dt}\right|^{-1} = \frac{1}{\varphi_0^2}\left(\frac{8u\tanh^{-1}\left[\sqrt{\frac{1-u}{1+u}}\right]}{3\sqrt{1-u^2}(1+u)}\right)\exp[\pi\alpha^2/2gH] \tag{4.103}$$

where

$$\begin{aligned} u &= \sqrt{1-(\alpha/\alpha_c)^2}, \\ \varphi_0 &= \frac{\Phi_c}{\sqrt{g}}\sqrt{1-u}, \end{aligned} \tag{4.104}$$

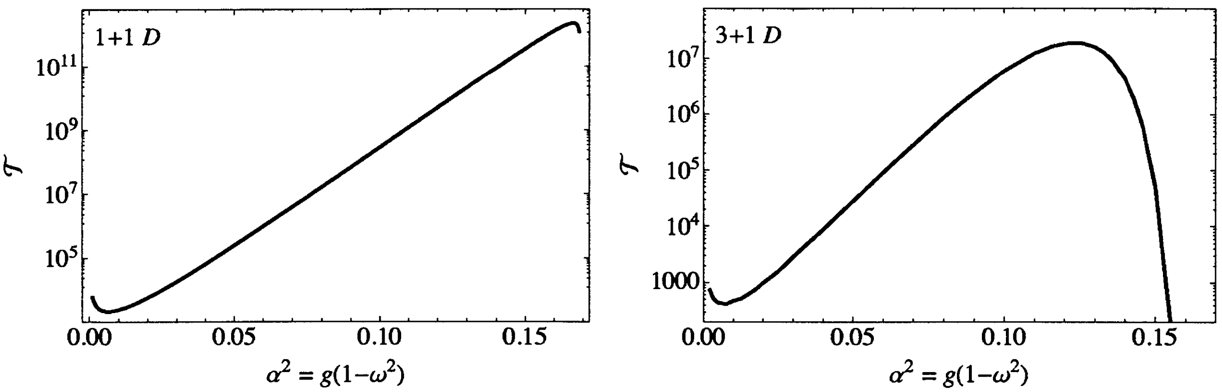

Figure 4-6: The above plot shows the time it takes for an oscillon to lose most of its energy due to the expanding background. The plot assumes an $H = 10^{-3}$ and $g = 10$. The lifetime is sensitive to the combination $\sim g\exp[\alpha^2/(gH)]$. Our result is only valid when α is not too close to 0 or $\alpha_c = \sqrt{27/160}$, since for these value our assumption $r_e \ll \alpha(\sqrt{g}H)^{-1}$ breaks down. Note that the lifetime should be interpreted qualitatively, since we are not allowing for instabilities that lead to the oscillon's abrupt disintegration. Time is measured in units of m^{-1}.

The terms in $(\ldots)$ are corrections for the deviation from the sech profile. We plot the lifetime as a function of the α^2 in Fig. 4-6. We should not trust this curve in the limit $\alpha \to 0, \alpha_c$, since our assumption of $x_e(\alpha) \ll (\sqrt{g}H)^{-1}$ is broken here. Note the scaling with g: $\mathcal{T} \sim g\exp[\alpha^2/(gH)]$.

A numerical investigation of oscillon lifetimes in a $1+1$-dimensional expanding background was carried out in [44]: however, no analytic calculation was provided for the lifetime. Our analytic results seem to be in good qualitative agreement with their paper. Note that $E_{osc} \sim g^{-1/2}\alpha^2$ for small α, which tells us that the lifetime $\mathcal{T} \sim \exp[E_{osc}/g^{3/2}H]$, is in agreement with the empirical formula inferred in their paper (also see Fig. 2 in [44]).

A similar calculation can be carried out in $3+1$ dimensions. The rate of energy loss is then given by (to leading order in $g^{-1/2}$)

$$\frac{dE_{osc}}{dt} \approx -\frac{4\pi}{\sqrt{g}} f^2(\alpha)\alpha\, e^{-\pi\alpha^2/(2gH)}. \tag{4.105}$$

Here, we have used Eq. (4.99). The lifetime is then given by

$$\mathcal{T} \lesssim \frac{N(\alpha)}{8\pi f^2(\alpha)\alpha} g e^{\pi\alpha^2/(2gH)}. \tag{4.106}$$

Note that $2E_{osc}(\alpha) = g^{1/2}N(\alpha)$ up to leading order in powers of g. We plot the approximate lifetime of an oscillon in a $3+1$-dimensional scenario in Fig. 4-6. The lifetime is maximized at $\alpha < \alpha_c$. In plotting the lifetime, the numerically obtained $N(\alpha)$ and $f(\alpha)$ had errors of about 1 percent. Again, we stress that one should think about the above curve somewhat qualitatively since higher order instabilities are ignored.

4.6 Summary

In this chapter, we have found that the spatial oscillon profiles can be very different from a Gaussian, an ansatz often made in the literature. In particular, we derived the nonmonotonic relationship between the height and the width of the oscillons, and discussed the importance of this feature for the stability of oscillons. We showed that the flat-topped oscillons are more stable in three dimensions to long wavelength perturbations as compared to their usual Gaussian counterparts.

A number of questions related to this chapter require further investigation. Our expressions for lifetime, especially in $3+1$ dimensions in the flat-top regimes should be checked with a detailed numerical investigation. The question of the possible small wavelength, narrow band instability needs to be resolved rigorously. Recently, [31] discussed oscillons in the presence of gravity (an oscillaton). It would also be nice to obtain oscillon number density estimates for a class of physically important potentials that have arisen in the context of string theory.

Bibliography

[1] G. Aarts, G. F. Bonini, and C. Wetterich. Exact and truncated dynamics in nonequilibrium field theory. *Phys. Rev. D*, 63(025012), 2000.

[2] G. Aarts, G. F. Bonini, and C. Wetterich. On thermalization in classical scalar field theory. *Nucl. Phys. B*, 587:403–418, 2000.

[3] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions.* Dover Publications, Inc., New York, 1972.

[4] M. A. Amin. *Inflaton fragmentation: Emergence of pseudo-stable inflaton lumps (oscillons) after inflation*, 2010. [arXiv:1006.3075v2].

[5] M. A. Amin, R. Easther, and H. Finkel. Inflaton fragmentation and oscillon formation in three dimensions. *J. Cosmol. Astropart. Phys.*, 12(001), 2010.

[6] M. A. Amin and D. Shirokoff. Flat-top oscillons in an expanding universe. *Phys. Rev. D*, 81(085045), 2010.

[7] D. L. T. Anderson. Stability of time-dependent particlelike solutions in nonlinear field theories. 2. *J. Math. Phys.*, 12:945–952, 1971.

[8] P. Angot, C.-H. Bruneau, and P. Fabrie. A penalization method to take into account obstacles in viscous flows. *Numerische Mathematik*, 81:497–520, 1999.

[9] J. B. Bell, P. Collela, and H. M. Glaz. A second-order projection method for the incompressible Navier-Stokes equations. *J. Comput. Phys.*, 85:257–283, 1989.

[10] M. Ben-Artzi, J.-P. Croisille, D. Fishelov, and S. Trachtenberg. A pure-compact scheme for the streamfunction formulation of the Navier-Stokes equations. *J. Comput. Phys.*, 205:640–644, 2005.

[11] G. Benettin, R. Livi, and A. Ponno. The Fermi-Pasta-Ulam problem: Scaling laws vs. initial conditions. *J. Stat. Phys.*, 135:873–893, 2009.

[12] G. P. Berman and F. M. Izrailev. The Fermi-Pasta-Ulam problem: 50 years of progress. *Chaos*, 15(015104), 2005.

[13] I. L. Bogolyubsky and V. G. Makhankov. Lifetime of pulsating solitons in some classical models. *Pisma Zh. Eksp. Teor. Fiz.*, 24(15), 1976.

[14] D. Boyanovsky, C. Destri, and H. J. de Vega. Approach to thermalization in the classical ϕ^4 theory in $1+1$ dimensions: Energy cascades and universal scaling. *Phys. Rev. D*, 69(045003), 2004.

[15] D. L. Brown, R. Cortez, and M. L. Minion. Accurate projection methods for the incompressible Navier–Stokes equations. *J. Comput. Phys.*, 168:464–499, 2001.

[16] D. Calhoun. A cartesian grid method for solving the two-dimensional stream function-vorticity equations in irregular regions. *J. Comput. Phys.*, 176:231–275, 2002.

[17] E. A. Calzetta and B. L. Hu. *Thermalization of an Interacting Quantum Field in the CTP-2PI Next-to-leading-order Large N Scheme*, 2002. [hep-ph/0205271].

[18] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang. *Spectral Methods: Evolution to Complex Geometries and Applications to Fluid Dynamics.* Springer-Verlag, Berlin Heidelberg, 2007.

[19] A. Carati and L. Galgani. Planck's formula and glassy behavior in classical nonequilibrium statistical mechanics. *Physica A*, 280:106–114, 2000.

[20] A. J. Chorin. Numerical solutions of the Navier-Stokes equations. *Math. Comput.*, 22:745–762, 1968.

[21] S. R. Coleman. Q balls. *Nucl. Phys. B*, 262, 263, 1985.

[22] I. Dymnikova, L. Koziel, M. Khlopov, and S. Rubin. Quasilumps from first-order phase transitions. *Grav. Cosmol.*, 6:311–318, 2000.

[23] W. E and J.-G. Liu. Guage method for viscous incompressible flows. *Commun. Math Sci.*, 1(317332), 2003.

[24] E. Farhi, N. Graham, A. H. Guth, N. Iqbal, R. R. Rosales, and N. Stamatopoulos. Emergence of oscillons in an expanding background. *Phys. Rev. D*, 77(085019), 2008.

[25] E. Farhi, N. Graham, V. Khemani, R. Markov, and R. Rosales. An oscillon in the SU(2) gauged higgs model. *Phys. Rev. D*, 72(101701), 2005.

[26] C. L. Fefferman. Existence and smoothness of the Navier-Stokes equation. Technical report, Clay Mathematics Institute, 2000. Millenium prize problem.

[27] E. Fermi, J. Pasta, and S. Ulam. Studies of nonlinear problems: I. *Los Alamos Scientific Laboratory Report No. LA-1940*, 1955. reprinted in Fermi E. Collected papers (University of Chicago Press, Chicago, 1965), Vol II, 978.

[28] J. Ferziger and M. Peric. *Computational Methods for Fluid Dynamics.* Springer-Verlag, New York, third edn. edition, 2002.

[29] G. Fodor, P. Forgacs, Z. Horvath, and A. Lukacs. Small amplitude quasi-breathers and oscillons. *Phys. Rev. D*, 78(025003), 2008.

[30] G. Fodor, P. Forgacs, Z. Horvath, and M. Mezei. Radiation of scalar oscillons in 2 and 3 dimensions. *Phys. Lett. B*, 674:319–324, 2009.

[31] G. Fodor, P. Forgacs, and M. Mezei. Mass loss and longevity of gravitationally bound oscillating scalar lumps (oscillatons) in d-dimensions. *Phys. Rev. D*, 81(064029), 2010.

[32] J. A. Frieman, G. B. Gelmini, M. Gleiser, and E. W. Kolb. Solitogenesis: Primordial origin of nontopological solitons. *Phys. Rev. Lett.*, 60(2101), 1988.

[33] G. Gallavotti, editor. *The Fermi-Pasta-Ulam problem: A status report*, volume 728 of *Lect. Notes Phys.* Springer, Berlin, 2008.

[34] B. Gershgorin, Y. V. Lvov, and D. Cai. Renormalized waves and discrete breathers in Beta-Fermi-Pasta-Ulam chains. *Phys. Rev. Lett.*, 95(264302), 2005.

[35] B. Gershgorin, Y. V. Lvov, and D. Cai. Interactions of renormalized waves in thermalized Fermi-Pasta-Ulam chains. *Phys. Rev. E*, 75(046603), 2007.

[36] U. Ghia, K. N. Ghia, and C. T. Shin. High-re solutions for incompressible flow using the Navier-Stokes equations and a multigrid method. *J. Comput. Phys.*, 48:387–411, 1982.

[37] M. Gleiser. Pseudostable bubbles. *Phys. Rev. D*, 49(2978), 1994.

[38] M. Gleiser. d-dimensional oscillating scalar field lumps and the dimensionality of space. *Phys. Lett. B*, 600:126–132, 2004.

[39] M. Gleiser and D. Sicilia. A general theory of oscillon dynamics. *Phys. Rev. D*, 80(125037), 2009.

[40] R. Glowinski, T.-W. Pan, A. J. Kearsley, and J. Periaux. Numerical simulation and optimal shape for viscous flow by a fictitious domain method. *Int. J. Numer. Meth. Fluids*, 20:695–711, 1995.

[41] D. Goldstein, R. Handler, and L. Serovich. Modeling a no-slip flow boundary with an external force field. *J. Comput. Phys.*, 105:354–366, 1993.

[42] D. Gottlieb and S. Orzag. *Numerical Analysis of Spectral Methods: Theory and Applications.* SIAM, Philadelphia, 1977.

[43] N. Graham. An electroweak oscillon. *Phys. Rev. Lett.*, 98(101801), 2007.

[44] N. Graham and N. Stamatopoulos. Unnatural oscillon lifetimes in an expanding background. *Phys. Lett. B*, 639:541–545, 2006.

[45] D. Greenspan. *Introduction to Numerical Analysis of Elliptic Value Problems.* Int. Edn., Harper New York, 1965.

[46] P. M. Gresho and R. L. Sani. On presssure boundary conditions for the incompressible Navier-Stokes equations. *Int. J. Numer. Methods Fluids*, 7:1111–1145, 1987.

[47] J. L. Guermond, P. Minev, and Jie Shen. An overview of projection methods for incompressible flows. *Comput. Methods in Appl. Mech. Eng.*, 195:6011–6045, 2006.

[48] C. Guha-Roy, B. Bagchi, and D. K. Sinha. Short-time evolution of nonlinear Klein-Gordon systems. *Int. J. Theor. Phys.*, 26(4), 1987.

[49] A. Hammouti. *Simulation Numérique Directe en différence finie de l'écoulement d'un fluide incompressible en présence d'une interface rigide.* PhD thesis, Ecole Nationale des Ponts et Chaussés, Paris, France, December 2009.

[50] F. H. Harlow and J. E. Welch. Numerical calculation of time-dependent viscous incompressible flow of fluid with free surface. *Phys. Fluids*, 8:2182–2189, 1965.

[51] William D. Henshaw. A fourth-order accurate method for the incompressible Navier-Stokes equations on overlapping grids. *J. Comput. Phys.*, 113(1):13–25, July 1994.

[52] William D. Henshaw, H.-O. Kreiss, and L. G. M. Reyna. A fourth-order accurate difference approximation for the incompressible Navier-Stokes equations. *Comput. Fluids*, 23(4):575–593, 1994.

[53] William D. Henshaw and N. Anders Petersson. A split-step scheme for the incompressible Navier-Stokes equations. In M.M. Hafez, editor, *Numerical Simulation of Incompressible Flows*, pages 108–125. World Scientific, 2003.

[54] M. Hertzberg. Quantum radiation of oscillons. *Phys.Rev.D*, 82(045022), 2010.

[55] M. Hindmarsh and P. Salmi. Oscillons and domain walls. *Phys. Rev. D*, 77(105025), 2008.

[56] T. A. S. Jackson. Orthogonal curvilinear coordinates. *The Mathematical Gazette*, 50(371):28–30, 1966.

[57] H. Johnston and J.-G. Liu. A finite difference method for incompressible flow based on local pressure boundary conditions. *J. Comput. Phys.*, 180:120–154, 2002.

[58] H. Johnston and J.-G. Liu. Accurate, stable and efficient Navier-Stokes solvers based on explicit treatment of the pressure term. *J. Comput. Phys.*, 199:221–259, 2004.

[59] S. Juchem, W. Cassing, and C. Greiner. Quantum dynamics and thermalization for out-of-equilibrium ϕ^4 theory. *Phys. Rev. D*, 69(025006), 2004.

[60] G. Karniadakis, M. Israeli, and S. A. Orszag. High-order splitting methods for the incompressible Navier-Stokes equations. *J. Comput. Phys.*, 97:414–443, 1991.

[61] S. Kasuya, M. Kawasaki, and F. Takahashi. I-balls. *Phys. Lett. B*, 559:99–106, 2003.

[62] K. Khadra, P. Angot, S. Parneix, and J.-P. Caltagirone. Fictitious domain approach for numerical modelling of Navier-Stokes equations. *Int. J. Numer. Meth. Fluids*, 34:651–684, 2000.

[63] J. Kim and P. Moin. Application of a fractional step method to incompressible Navier-Stokes equations. *J. Comput. Phys.*, 59:308–323, 1985.

[64] L. Kleiser and U. Schumann. Treatment of the incompressibility and boundary conditions in 3-d numerical spectral simulation of plane channel flows. In E. H. Hirschel, editor, *Notes on Numerical Numerical Fluid Mechanics*, pages 165–173. Vieweg, Braunschweig, 1980.

[65] E. W. Kolb and I. I. Tkachev. Nonlinear axion dynamics and formation of cosmological pseudosolitons. *Phys. Rev. D*, 49(5040), 1994.

[66] J. A. Krumhansl and J. R. Schrieffer. Dynamics and statistical mechanics of a one-dimensional model hamiltonian for structural phase transitions. *Phys. Rev. B*, 11(3535), 1975.

[67] A. Krzywicki and O. A. Ladyzhenskaya. A grid method for the Navier-Stokes equations. *Dokl. Akad. Nauk SSSR*, 167:309–311, 1966. English transl., Soviet Physics Dokl., 11:212–213, 1966.

[68] P. D. Lax and R. D. Richtmyer. Survey of the stability of linear finite difference equations. *Comm. Pure Appl. Math.*, 9:267–293, 1956.

[69] T. D. Lee and Y. Pang. Nontopological solitons. *Phys. Rept.*, 221(251), 1992.

[70] W. Lee, G. Kovačič, and D. Cai. Renormalized resonance quartets in dispersive wave turbulence. *Phys. Rev. Lett.*, 103(024502), 2009.

[71] M. Linnick and H. Fasel. A high-order immersed interface method for simulating unsteady incompressible flows on irregular domains. *J. Comput. Phys.*, 204:157–192, 2005.

[72] J.-G. Liu, J. Liu, and R. Pego. Stable and accurate pressure approximation for unsteady incompressible viscous flow. *J. Comput. Phys.*, 229:3428–3453, 2010.

[73] R. Livi, M. Pettini, S. Ruflo, and A. Vulpiani. The ultraviolet problem and analytical properties of classical field theory. *J. Phys. A: Math. Gen.*, 20:577–586, 1987.

[74] J. De Luca and A. Lichtenberg. Transitions and time scales to equipartition in oscillator chains: Low-frequency initial conditions. *Phys. Rev. E*, 66(026206), 2002.

[75] R. Mittal and G. Iaccarino. Immersed boundary methods. *Ann. Rev. Fluid Mech.*, 37:39–261, 2005.

[76] M. Napolitano, G. Pascazio, and L. Quartapelle. A review of vorticity conditions in the numerical solution of the $\zeta - \psi$ equations. *J. Comput. Fluids*, 28:139–185, 1999.

[77] S. A. Orszag, M. Israeli, and M. Deville. Boundary conditions for incompressible flows. *J. Sci. Comput.*, 1:75–111, 1986.

[78] A. Patrasciou. Are there any ergodic local relativistic field theories? *Phys. Lett.*, 104A(87), 1984.

[79] J. B. Perot. An analysis of the fractional step method. *J. Comput. Phys.*, 108:51–58, 1993.

[80] C. S. Peskin. Flow patterns around heart valves: a numerical method. *J. Comput. Phys.*, 10:252–271, 1972.

[81] C. S. Peskin. The immersed boundary method. *Acta Numerica*, 11:479–517, 2002.

[82] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, Cambridge, 1988.

[83] A. Prohl. Projection and quasi-compressibility methods for solving the incompressible Navier-Stokes equations. In *Advances in Numerical Mathematics, Wiley-Teubner series, advances in numerical mathematics.* B.G. Teubner, Stuttgart, 1997.

[84] L. Quartapelle. *Numerical Solution of the Incompressible Navier-Stokes Equations.* Birkhauser Verlag, 1993.

[85] R. Rannacher. On Chorin's projection method for the incompressible Navier-Stokes equations. *Lecture Notes in Mathematics*, 1530:167–183, 1991.

[86] D. Rempfer. On boundary conditions for the incompressible Navier-Stokes problems. *Applied Mechanics Reviews*, 59:107–125, 2006.

[87] J. S. Russell. Report on waves. *Report of the 14th Meeting of the British Association for the Advancement of Science, York*, pages 311–390, September 1844. (John Murray, London, 1844).

[88] P. M. Saffin and A. Tranberg. Oscillons and quasi-breathers in d+1 dimensions. *JHEP*, 0701(030), 2007.

[89] M. Sallé, J. Smit, and J. Vink. Staying thermal with Hartree ensemble approximations. *Nuclear Physics B*, 625:495–511, 2002.

[90] R. Sani, J. Shen, O. Pironneau, and P. Gresho. Pressure boundary condition for the time-dependent incompressible Navier-Stokes equations. *Int. J. Numer. Methods Fluids*, 50:673–682, 2006.

[91] D. J. Scalapino, M. Sears, and R. A. Farrell. Statistical mechanics of one-dimensional Ginzburg-Landau fields. *Phys. Rev. B*, 6(3409), 1972.

[92] H. Segur and M. D. Kruskal. Nonexistence of small amplitude breather solutions in ϕ^4 theory. *Phys. Rev. Lett.*, 58(747), 1987.

[93] J. Shen. On error estimates of the projection methods for the Navier-Stokes equations: first-order schemes. *SIAM J. Numer. Anal.*, 29:57–77, 1992.

[94] D. Shirokoff. Renormalized waves and thermalization of the Klein-Gordon equation. *Phys. Rev. E*, 85(046217), 2011.

[95] D. Shirokoff and R. R. Rosales. *An efficient method for the incompressible Navier-Stokes equations on irregular domains with no-slip boundary conditions, high order up to the boundary*, 2010. [arxiv:1011.3589].

[96] J. C. Strikwerda. *Finite Difference Schemes and Partial Differential Equations.* Wadsworth and Brooks Cole, 1989.

[97] K. Taira and T. Colonius. The immersed boundary method: A projection approach. *J. Comput. Phys.*, 225:2118–2137, 2007.

[98] M. Taylor. *Partial differential equations III, Nonlinear equations.* Springer New York, 1996.

[99] R. Temam. Sur l'approximation de la solution des equations de Navier-Stokes par la methode des pas fractionnaries, ii. *Arch. Ration. Mech. Anal.*, 33:377–385, 1969.

[100] V. Thomee. *Galerkin finite element methods for parabolic problems.* Springer-Verlag Berlin Heidelberg, second edition, 2006.

[101] C. Tu and C. S. Peskin. Stability and instability in the computation of flows with moving immersed boundaries: A comparison of three methods. *SIAM, Journal on Scientic and Statistical Computing*, 13:1361–1376, 1992.

[102] N.G. Vakhitov and A.A. Kolokolov. Stationary solutions of the wave equation in a medium with nonlinear saturation. *Radiophys. Quantum Electron.*, 16:783–789, 1973.

[103] J. van Kan. A second-order accurate pressure-correction scheme for viscous incompressible flow. *SIAM J. Sci. Stat. Comput.*, 7:870–891, 1986.

[104] S. Weinberg. *The Quantum Theory of Fields, Vol. 1.* Cambridge University Press, Cambridge, 1995.

[105] G. B. Whitham. *Linear and Nonlinear Waves.* John Wiley & Sons, New York, 1974.

[106] V. E. Zakharov, V. S. L'vov, and G. Falkovich. *Kolmogorov Spectra of Turbulence I.* Springer, Berlin, 1992.