Analysis of Numerical Approximation Algorithms for Nonlinear Differential Equations Using a Discrete Multiple Scales Technique

Eben Maré

Submitted in accordance with the requirements for the degree of

Philosophiae Doctor

in the

Faculty of Natural and Agricultural Sciences
Department of Mathematics and Applied Mathematics
at the
University of the Free State
Bloemfontein 9300
South Africa

November 2002

Supervisor: Prof. S. W. Schoombie

Acknowledgements

I am grateful to:

- Prof. S. W. Schoombie, my supervisor, for his guidance, encouragement, patience and friendship.
- Prof. C. J. Wright, who introduced me to the field of numerical analysis and provided vital career guidance at crucial times.
- Previous and current colleagues and friends at Nedbank, FirstDerivatives, Gensec and NIB for their encouragement and support.
- The UNISA library.
- My family and friends for their support, especially my wife Amanda and my children Eben, Jean-Jacques and Desiré who shared in the delights (and miseries) of this entire project.

SOLI DEO GLORIA

Preface

Theory in Numerical Analysis is never,
or should never be,
an end in itself.

P. Henrici

Motivation for the research project

Perturbation techniques for the solution of differential equations form an essential ingredient of the tools of mathematics as applied to physics, engineering, finance and other areas of applied mathematics. A natural extension would be to seek perturbation-type solutions for discrete approximations of differential equations.

Objectives

The main objective of the research project was to develop a perturbation technique for the solution of discrete equations. To achieve this we gave attention to the following:

1. Application of the technique to discrete approximations of relevant equations.

2. Comparisons of the developed theory with observed computed results.

3. Investigate deviations between perturbation solutions and computed solutions and explain reasons.

Thesis Overview

This thesis is divided up as follows:

Chapter 1 contains a short exposition of singular perturbation techniques, highlighting in particular the method of multiple scales. These techniques are applied to the solution of the Van der Pol, Korteweg-de Vries (KdV) and Regularized Long Wave equations.

Chapter 2 contains the basic framework for the numerical methods we wish to examine, specifically finite difference methods. A number of different ways to derive finite difference methods are detailed and we show the connection between central finite difference methods and the pseudospectral method.

In Chapter 3 we discuss a discrete multiple scales methodology as derived by Schoombie [111]. We generalize the method for application to general finite difference approximations.

In Chapter 4 we apply the generalized discrete multiple scales analysis to the solution of the discrete KdV equation. We shall show the consistency of the method with the continuous analysis as the discretization parameters tend to zero. We show that the discrete multiple scales technique is a powerful tool for the examination of modulational properties of equations such as the KdV equation. We show that in the case of certain modes of the carrier wave, the multiple scales analysis breaks down, indicating that in these cases the numerical solution deviates in behavior from that of the KdV equation. Several numerical experiments are performed to examine the spurious behavior for different orders of approximation. We supplement the work of Chapter 4 with a Benjamin-Feir instability analysis in Chapter 5.

In Chapter 6 we show the application of the discrete multiple scales analysis to the solution of a specific discretization of the Van der Pol equation. We also discuss related work.

A short discussion on the work performed in this study, as well as a list of possible extensions and ideas for future research, is given in Chapter 7.

Contents

List of Tables

List of Figures

List of symbols

$a \approx b$	:	a is approximately equal to b
$a := b$	:	a is defined as b
$a \equiv b$	:	a is equivalent to b
$a \ll b$	:	a is much less than b
i	:	Complex unit
$\frac{d^k f(x)}{dx^k}$	:	k-th order derivative of $f(x)$
$f'(x)$	:	Derivative of $f(x)$
∂_x^p	:	Partial derivative of order p with respect to x
F, F^{-1}	:	Fourier transform and inverse
h	:	Grid spacing
τ	:	Step size of time integration procedure
x_j	:	Grid points or nodes
$u_j(t)$	:	Semi discrete approximation to $u(x_j, t)$
u_j^n	:	$u(x_j, n\tau)$
$\delta_{m,n}$	:	Kronecker delta
$\|\|\cdot\|\|_\infty$	:	L_∞-norm
$\|\|\cdot\|\|_{L_2}$	:	L_2-norm
N	:	Number of degrees of freedom in space discretizations
$\Re u$	:	Real part of u
$\Im u$	:	Imaginary part of u
$\prod_j$	:	Product over j terms
$\sum_j$	:	Summation over j terms

List of Abbreviations

AB	:	Adams-Bashfort
AM	:	Adams-Moulton
CC	:	Complex conjugate
CFL	:	Courant-Friedrichs-Lewy (stability condition)
DE	:	Differential equation
FD	:	Finite difference(s)
FFT	:	Fast Fourier transform
IVP	:	Initial value problem
KdV	:	Korteweg-de Vries (equation)
LF	:	Leapfrog
MOL	:	Method of Lines
MKdV	:	Modified Korteweg-de Vries (equation)
NLS	:	Nonlinear Schrödinger (equation)
ODE	:	Ordinary differential equation
PDE	:	Partial differential equation
RK	:	Runge-Kutta
RLW	:	Regularized Long Wave (equation)
ZK	:	Zabusky-Kruskal (approximation)

Chapter 1

Perturbation Techniques

A mathematician, like a painter or a poet, is a master of pattern.

G. H. Hardy

In this chapter we present some essential ideas used in the solution of singular perturbation problems. It is not intended to be a complete survey of the literature on the subject; even for the examples given and discussed here no attempt has been made to make the references complete. It is not a survey of results or techniques - for this the reader is referred to Nayfeh [93, 94], Ames [4] and Mason [83] amongst others.

Our main aim is to discuss the fundamental heuristic ideas which underlie certain analytical techniques with the aim of expanding the techniques towards the analysis of finite difference approximations to ordinary and partial differential equations.

We shall introduce the well-known method of multiple scales and show its use for the solution of the Korteweg-de Vries (KdV) and Van der Pol equations. We shall also demonstrate the use of an alternative multiple scales technique as applied to the KdV and Regularized Long Wave (RLW) equations. In particular, for the KdV and RLW equations the analysis shows that the envelopes of modulated waves are governed by the nonlinear Schrödinger (NLS) equation. The alternative multiple scales technique presents an ideal framework from which we shall devise a discrete multiple scales analysis methodology.

1.1 Multiple Scales - An Introduction

Consider the *Cauchy* problem for the parabolic equation

$$\partial_t u + \epsilon u = \partial_x^2 u, \tag{1.1}$$

where ∂_p denotes partial differentiation with respect to p. The equation is to be solved on the domain $-\infty < x < \infty$, $t > 0$ subject to initial data

$$u(x,t) = f(x), \ -\infty < x < \infty. \tag{1.2}$$

Equation (1.1) is used as a model equation for heat conduction in a rod in which there is a heat loss due to radiation on the surface. The radiative effect gives rise to the ϵu term in (1.1). (See [146], for example.)

Changing to a new dependent variable by means of the transformation

$$u(x,t) = e^{-\epsilon t} v(x,t), \tag{1.3}$$

yields the classical *heat equation* [15],

$$\partial_t v = \partial_x^2 v. \tag{1.4}$$

Thus the solution to (1.1) is given by the *Cauchy* problem for (1.4).

We shall now apply a simple perturbation method [9] to find an approximate solution to (1.1). For that purpose we use the expansion

$$u(x,t) = \sum_{n=0}^{\infty} u_n(x,t)\epsilon^n. \tag{1.5}$$

Substituting (1.5) into (1.1) we have

$$\begin{aligned} &\partial_t u_0 + \epsilon \partial_t u_1 + \dots \\ &+\epsilon[u_0 + \epsilon u_1 + \dots] \\ &= \partial_x^2 u_0 + \epsilon \partial_x^2 u_1 + \dots \end{aligned} \tag{1.6}$$

Equating the coefficients of like power in ϵ in (1.6) we have

$$Lu_0 = 0, \tag{1.7}$$

for the coefficient of ϵ^0 and

$$Lu_1 = -u_0, \tag{1.8}$$

for the coefficient of ϵ^1. The linear operator, L, is given by

$$L = \partial_t - \partial_x^2. \tag{1.9}$$

The initial conditions are given by

$$\begin{aligned} u_0(x,0) &= f(x), \\ u_1(x,0) &= 0. \end{aligned} \tag{1.10}$$

Solving (1.7) and (1.8) we find

$$u(x,t) = v(x,t) - \epsilon t v(x,t) + O(\epsilon^2), \tag{1.11}$$

where $v(x,t)$ is the solution to the *heat equation* (1.4).

On comparison with the exact solution (1.3) we conclude that (1.11) serves as a good approximate solution for $\epsilon t \ll 1$. However, if $\epsilon t = O(1)$ we find that the first perturbation term $\epsilon u_1 = -\epsilon t v(x,t)$ is of the same order of magnitude as the leading term u_0. Consequently, no matter how small ϵ there exists a finite time at which all terms in the perturbation series are of the same order in ϵ and cannot be neglected on the basis that they constitute small corrections for small ϵ. The result is that the perturbation solution (1.11) is not uniformly valid for all times. Terms of the form $\epsilon t\, v(x,t)$ are referred to as *secular terms.*

If we complete the full perturbation solution we obtain

$$Lu_n = -u_{n-1}, \; n \geq 2, \tag{1.12}$$

with the initial conditions

$$u_n(x,0) = 0. \tag{1.13}$$

Making use of Duhamel's principle [15, 146] it is readily shown that

$$u_n(x,t) = [(-t)^n/n!]\, v(x,t), \tag{1.14}$$

which, upon substitution into (1.5), leads to the following well-known convergent infinite series expansion for the exponential function

$$\begin{aligned} u(x,t) &= \sum_{n=0}^{\infty} \frac{(-\epsilon t)^n}{n!} v(x,t) \\ &= e^{-\epsilon t} v(x,t). \end{aligned} \tag{1.15}$$

Therefore, the result obtained from the simple perturbation method with an infinite number of terms is identical to the result (1.3) obtained by means of the continuous

change of variables approach. However, in practice only a few terms in the expansion (1.15) are retained which will lead to an invalid approximation to the initial problem at some stage as illustrated above.

There are a number of methods in the literature for remedying the problems caused by secular terms – the textbooks by Nayfeh [93] and [94], for example, cover the area with a significant number of examples and studies. Also of interest is the more recent work by [38, 77, 92]. Our focus in this work will be concentrated on multiple scales type methods, as described below briefly for equation (1.1).

As shown by the solution (1.3) to (1.1) the problem involves a slow time scale ϵt and a comparatively fast time scale t. Therefore, it seems reasonable to look for a solution to (1.1) of the form $u(x, T_0, T_1)$ where $T_0 = t$ and $T_1 = \epsilon t$.

By making use of the chain rule for derivatives we can write

$$\partial_t = \partial_{T_0} + \epsilon \partial_{T_1}, \tag{1.16}$$

which leads to

$$\partial_{T_0} u + \epsilon(\partial_{T_1} u + u) = \partial_x^2 u, \tag{1.17}$$

upon substitution into (1.1). Employing a perturbation series expansion

$$u = \sum_{n=0}^{\infty} \epsilon^n u_n, \tag{1.18}$$

yields the recursive system

$$\partial_{T_0} u_0 - \partial_x^2 u_0 = 0, \tag{1.19}$$

as well as

$$\partial_{T_0} u_n - \partial_x^2 u_n = -[\partial_{T_1} u_{n-1} + u_{n-1}]. \tag{1.20}$$

From (1.19) with the initial conditions given by (1.10) we find

$$u_0(x, T_0, T_1) = c(T_1) v(x, T_0), \tag{1.21}$$

with $v(x, T_0)$ defined as before and $c(T_1)$ an arbitrary function of T_1 that satisfies the condition $c(0) = 1$ (since $t = 0$ implies that $T_1 = \epsilon t$ vanishes). Substitution of (1.21) into (1.20) with $n = 1$ yields

$$\partial_{T_0} u_1 - \partial_x^2 u_1 = -[\partial_{T_1} u_0 + u_0] = -[\frac{dc}{dT_1} + c] v(x, T_0). \tag{1.22}$$

Because $c(T_1)$ and its first derivative are constants as far as the operator on the left of (1.22) is concerned, we obtain the solution as

$$u_1(x, T_0, T_1) = -t[\frac{dc}{dT_1} + c] v(x, T_0) + d(T_1), \tag{1.23}$$

where $d(T_1)$ is an arbitrary function that vanishes at $T_1 = 0$. Inspection of (1.23) reveals that we obtain a solution that grows with t. To remove such secular terms we require that

$$\frac{dc}{dT_1} + c = 0. \tag{1.24}$$

Making use of the initial condition $c(0) = 1$ gives

$$c(T_1) = e^{-\epsilon t}, \tag{1.25}$$

as the solution of (1.24). It is important to note that we also set $d(T_1) = 0$. Otherwise further secular terms would arise at the next level of approximation. Consequently, $u_1 = 0$ and therefore $u_n = 0$ for all $n > 1$. The perturbation solution (1.18) terminates after the first term and yields the exact solution as given by (1.3).

In Nayfeh [93], the comment is made that it is the rule rather than the exception that expansions of the form (1.5) are not uniformly valid and the approximations break down in regions called regions of non-uniformity [74], as shown earlier. As shown above the method of multiple scales provides a mechanism to render the solution uniformly valid. Nayfeh [93] provides a list of various applications in physics, engineering and applied mathematics.

In the following sections we shall show the application of the method of multiple scales to some ordinary and partial differential equations occuring in mathematical physics.

1.2 Van der Pol Equation

1.2.1 Introduction

We consider the Van der Pol differential equation in the form

$$\frac{d^2x}{dt^2} - \beta\epsilon(1 - x^2)\frac{dx}{dt} + \omega^2 x = 0, \tag{1.26}$$

with

$$(0 < \epsilon \ll 1),\ \beta > 0, \tag{1.27}$$

and ω constant.

This equation was first described by Van der Pol [135] in 1922 in the context of electronic circuits containing vacuum tubes. The Van der Pol equation is used to

describe damped oscillations, the damping dependent on the factor $(1 - x^2)$. For $|x| < 1$ the damping is negative implying an expanding solution of (1.26) while for $|x| > 1$ the damping is positive indicating a contracting solution. Since trajectories near the origin expand, trajectories from the outer regions contract, and the only equilibrium point is at $(0, 0)$, there must be a limit cycle encircling the origin [98]. If that were not the case the trajectory, constrained to lie on the plane, would intersect itself (a result of the Poincaré-Bendixson theorem only possible for a plane [101]).

The nonlinear Van der Pol differential equation serves as an important model equation for one-dimensional dynamic systems having a single, stable limit-cycle and is frequently used as a model equation in the study of differential equations.

More recent research on the Van der Pol equation focus on generalized equations. As an example, Moremedi and Mason [90] consider the generalized case where $(1 - x^2)$ is replaced by $(1 - x^{2n})$ with n any positive integer. Another case of interest is that where a forcing term of the form $Kcos(\alpha t)$, with K and α constant, is introduced to the right hand side of (1.26); see [32] and the references cited therein.

In the next section we discuss a multiple scales analysis of (1.26) in the manner of Nayfeh [93].

1.2.2 Multiple Scales Analysis

Following [83, 93] we assume the following multiple scales solution for (1.26)

$$\begin{aligned} x &= x(T_0, T_1, \epsilon) \\ &= x_0(T_0, T_1) + \epsilon x_1(T_0, T_1) + O(\epsilon^2), \end{aligned} \tag{1.28}$$

where we have introduced the independent variables T_0 and T_1 defined as

$$\begin{aligned} T_0 &= t, \\ T_1 &= \epsilon t. \end{aligned} \tag{1.29}$$

By making use of the chain rule for differentiation we can once more write d/dt in terms of ∂_{T_0} and ∂_{T_1}. Therefore

$$\frac{d}{dt} = \partial_{T_0} + \epsilon \partial_{T_1}, \tag{1.30}$$

and

$$\frac{d^2}{dt^2} = \partial_{T_0}^2 + 2\epsilon\partial_{T_0}\partial_{T_1} + O(\epsilon^2). \tag{1.31}$$

Substitution of (1.28), (1.30) and (1.31) into the differential equation (1.26) we obtain

$$\begin{aligned}&(\partial_{T_0}^2 + 2\epsilon\partial_{T_0}\partial_{T_1})(x_0 + \epsilon x_1)\\ &-\beta\epsilon(1 - (x_0 + \epsilon x_1)^2)(\partial_{T_0} + \epsilon\partial_{T_1})(x_0 + \epsilon x_1)\\ &+\omega^2(x_0 + \epsilon x_1) = 0.\end{aligned} \tag{1.32}$$

By equating coefficients of like power in ϵ we obtain a system of two partial differential equations in the two variables T_0 and T_1 for x_0 and x_1, namely

$$Lx_0 = 0, \tag{1.33}$$

and

$$Lx_1 = \beta(1 - x_0^2)\partial_{T_0}x_0 - 2\partial_{T_0}\partial_{T_1}x_0, \tag{1.34}$$

where L is defined by

$$L = \partial_{T_0}^2 + \omega^2. \tag{1.35}$$

The general solution of (1.33) is given by

$$x_0(T_0, T_1) = A(T_1)e^{i\omega T_0} + CC. \tag{1.36}$$

Substitution of (1.36) into (1.34) yields

$$\begin{aligned}Lx_1 &= -i\omega e^{i\omega T_0}(2\partial_{T_1}A - \beta A + \beta A|A|^2)\\ &\quad -i\omega e^{3i\omega T_0} + CC.\end{aligned} \tag{1.37}$$

To eliminate secular terms in x_1 we require the coefficients of $e^{i\omega T_0}$ in (1.37) to vanish. Thus

$$\partial_{T_1}A = \frac{\beta}{2}(A - A|A|^2). \tag{1.38}$$

To solve (1.38) we follow [10, 93]. Let

$$A = \frac{1}{2}a(T_1, T_2)e^{i\phi(T_1,T_2)}. \tag{1.39}$$

After substitution of (1.39) into (1.38) and separation of real and imaginary parts in (1.38), we obtain the logistic-type equation

$$\partial_{T_1}a = \frac{\beta}{2}(1 - \frac{a^2}{4})a, \tag{1.40}$$

and

$$\partial_{T_1}\phi = 0. \tag{1.41}$$

Making use of the initial condition

$$a(0) = a_0, \tag{1.42}$$

we obtain

$$a^2 = \frac{4}{[1 + (4/a_0^2 - 1)e^{-\beta T_1}]}. \tag{1.43}$$

Consequently, in terms of (1.36), we have

$$x_0 = \frac{2\cos(\omega t + \phi)}{\sqrt{(1 + (4/a_0^2 - 1)e^{-\beta\epsilon t})}}. \tag{1.44}$$

Therefore in terms of (1.28) we have

$$\begin{aligned} x(t,\epsilon) &= x(T_0, T_1, \epsilon) \\ &= x_0(T_0, T_1) + O(\epsilon) \\ &= \frac{2\cos(\omega t + \phi)}{\sqrt{(1 + (4/a_0^2 - 1)e^{-\beta\epsilon t})}} + O(\epsilon). \end{aligned} \tag{1.45}$$

We can proceed to find multiple scales solutions to the Van der Pol equation valid to higher orders of ϵ by using the methodology illustrated above as in Nayfeh [93]. However, for the purpose of our numerical investigation the presentation up to $O(\epsilon)$ given in (1.45) will be adequate.

1.3 Korteweg-de Vries Equation

The KdV equation is a nonlinear PDE arising in a number of different physical systems, e.g., water waves and elastic rods [68, 91]. It describes the long time evolution of small but finite amplitude dispersive waves. From detailed studies of properties of the equation and its solutions, the concept of solitons was introduced and the method for exact solution of the initial-value problem using inverse scattering theory was developed [36, 91]. In section 1.3.2 we show multiple scales analyses of the KdV equation.

1.3.1 Introduction

Consider the KdV equation in the form

$$\partial_t u + \eta \partial_x u + \zeta u \partial_x u + \gamma \partial_x^3 u = 0, \tag{1.46}$$

where ∂_p denotes partial differentiation with respect to p and η, ζ and γ are constants, with $\gamma \neq 0$.

Furthermore, for the purposes of our study, consider the following initial data

$$u(x,0) = \epsilon f(x), \quad f(x) = O(1), \tag{1.47}$$

where ϵ is a small, real, positive number. In addition, the following periodicity conditions are imposed on solutions of the KdV equation

$$u(x \pm L, t) = u(x,t), \quad f(x \pm L) = f(x), \quad t > 0, \; x \in \Re. \tag{1.48}$$

The KdV equation provides a useful model for describing the long-time evolution of wave phenomena in which the steepening effect of the nonlinear term $u\partial_x u$ is counterbalanced by dispersion [36]. For the KdV equation, the effect of dispersion is to prevent the formation of discontinuities [72]. The KdV equation was first derived by Korteweg and de Vries [71] to describe the propagation of unidirectional shallow water waves. For this specific case Vliegenthart [136] formulates the KdV equation in the form

$$\partial_t u + \sqrt{gh_0}[1 + \frac{3}{2}u/h_0]\partial_x u + \frac{1}{6}\sqrt{gh_0}h_0^2\partial_x^3 u = 0, \tag{1.49}$$

where $u(x,t)$ denotes the local wave-height above the undisturbed depth h_0, x the coordinate along the horizontal bottom, t the time and g the gravitational acceleration.

The KdV equation has been used to account adequately for observable phenomena such as the interaction of solitary waves and dissipationless, undular shocks. A *soliton* is defined as a localized or solitary entity that propagates at a uniform speed and preserves its structure (or shape) and speed in an interaction with another such solitary entity [61, 143]. In fact, Zabusky and Kruskal [144] discovered the concept of solitons while studying the results of a numerical computation (describing an anharmonic lattice) on the KdV equation [36]. The Zabusky-Kruskal (ZK) discretization of (1.46) will be described in greater detail later in this work.

The one-soliton solution of (1.46), for example, is given by [81]

$$u(x,t) = \frac{12\gamma\beta^2}{\zeta} sech^2(\beta(x - \eta t - \gamma\beta^2 t - x_0)), \tag{1.50}$$

where β and x_0 are free parameters and x_0 determines the initial position of the soliton. The one-soliton solution is a hump with height $12\gamma\beta^2/\zeta$ and width proportional to $1/\beta$, traveling to the right with velocity $\eta + \gamma\beta^2$. Consequently, the larger (in height) a soliton is, the thinner it is and the faster it travels to the right with respect to smaller solitons. The size of the coefficient of the convection term η merely adds velocity to all solitons.

The more complex two-soliton solution is given by [36, 81]:

$$u(x,t) = \frac{12\gamma}{\zeta}[\frac{A}{B}], \tag{1.51}$$

where

$$A = (\beta_1^2 f_1 + \beta_2^2 f_2 + 2(\beta_1 - \beta_2)^2 f_1 f_2$$
$$+(\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2})^2(\beta_1^2 f_1 f_2^2 + \beta_2^2 f_2 f_1^2),$$

and

$$B = (1 + f_1 + f_2 + (\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2})^2 f_1 f_2)^2.$$

Furthermore

$$f_1 = \exp(\beta_1(x - \eta t - x_1) - \gamma\beta_1^3 t),$$
$$f_2 = \exp(\beta_2(x - \eta t - x_2) - \gamma\beta_2^3 t),$$

and the β_n and x_n determine height and position of the n-th soliton.

The KdV equation has also been used as a model for ion acoustic waves in plasma; magnetohydrodynamic waves in plasma; the anharmonic lattice; longitudinal dispersive waves in elastic rods; pressure waves in liquid-gas mixtures; rotating flow down a tube and thermally exited phonon packets in low temperature nonlinear crystals. A number of these applications are described in [68] and [91].

It is also of interest to note that we can rewrite the KdV (1.46) in the conservation form

$$\partial_t T + \partial_x X = 0, \tag{1.52}$$

where $T = u$ and $X = \eta u + \zeta u^2/2 + \gamma\partial_x^2 u$. Assuming u is periodic in x or that u and its derivatives vanish sufficiently fast towards $\pm\infty$, integrating the conservation law yields

$$\partial_t \int T dx = 0, \tag{1.53}$$

where the limits of integration are $\pm\infty$ or two ends of a period in x. A second conservation law can be derived by multiplying (1.46) by u, with the result that $T = u^2/2$ and $X = \eta u^2/2 + \zeta u^3/3 + \gamma(u\partial_x^2 u - (\partial_x u)^2/2))$. The KdV is known to have an infinite number of polynomial conservation laws of the form (1.52) [14, 91].

Considering the large number of applications of the KdV equation as well as its historical importance in the study of solutions to nonlinear equations of evolution, it is impossible to do justice to the existing literature in our references. We do not, for example, make any reference to its historical significance in the development of the Inverse Scattering Transform theory [36, 61]. Therefore, only a small number of references relevant to this study are provided, these being [11, 14, 16, 31, 36, 43, 64, 68, 69, 70, 71, 72, 82, 91, 110, 127, 136, 139, 140, 143, 144] and [145].

1.3.2 Continuous Multiple Scales Analysis

For the purpose of a conventional multiple scales approach it is assumed that the solution $u(x,t)$ to (1.46) can be expanded in the following form [36, 110]

$$u(x,t) = \sum_{n=1}^{\infty} \epsilon^n u^{(n)}(X_0, X_1, T_0, T_1, T_2), \tag{1.54}$$

where

$$X_k = \epsilon^k x, \quad k = 0, 1, \tag{1.55}$$

and

$$T_k = \epsilon^k t, \quad k = 0, 1, 2. \tag{1.56}$$

Equations (1.55) and (1.56) are the spatial and time scales respectively. For $k = 0$ we have the fast scales in space and time and for higher values of k we have progressively longer space and slower time scales.

To proceed with the analysis the chain rule for derivatives is used with the result that the spatial derivative is transformed to

$$\partial_x = \sum_{n=0}^{1} \epsilon^n \partial_{X_n}. \tag{1.57}$$

The time derivative is written as

$$\partial_t = \sum_{n=0}^{2} \epsilon^n \partial_{T_n}. \tag{1.58}$$

The multiple scales analysis would ensue by substituting equations (1.54) through (1.58) into (1.46) and equating different orders of ϵ to zero. Thereby a hierarchy of perturbation equations is generated, the first three members of which are given by

$$Lu^{(1)} = 0, \tag{1.59}$$

and

$$Lu^{(2)} = -\partial_{T_1} u^{(1)} - \eta \partial_{X_1} u^{(1)} - \zeta u^{(1)} \partial_{X_0} u^{(1)} - 3\gamma \partial^3_{X_0} u^{(1)}, \tag{1.60}$$

and

$$\begin{aligned} Lu^{(3)} = & - \partial_{T_1} u^{(2)} - \partial_{T_2} u^{(1)} - \eta \partial_{X_1} u^{(2)} \\ & - \zeta [\partial_{X_0} (u^{(1)} u^{(2)}) + 1/2 \partial_{X_1} (u^{(1)})^2] \\ & - 3\gamma (\partial^2_{X_0} \partial_{X_1} u^{(2)} + \partial_{X_0} \partial^2_{X_1} u^{(1)}), \end{aligned} \tag{1.61}$$

where L is the linear operator

$$L = \partial_{T_0} + \eta \partial_{X_0} + \gamma \partial^3_{X_0}. \tag{1.62}$$

The solution to (1.59) is considered in the following form

$$u^{(1)} = A(X_1, T_1, T_2) e^{i\theta} + \text{CC}, \tag{1.63}$$

where the phase variable, θ, is given by

$$\theta = kX_0 - \omega T_0, \tag{1.64}$$

with the carrier wave number k related to ω, the carrier wave frequency, by the linear dispersion relation [14, 140]

$$\omega = \eta k - \gamma k^3. \tag{1.65}$$

To satisfy the periodicity conditions given in (1.48), we have to restrict the wave number k to the following values

$$k = k_m = 2\pi m / L, \quad m = 1, 2, \ldots. \tag{1.66}$$

We restrict k to nonzero values only as the trivial case $k = 0$ requires special treatment and is of little interest. Substituting (1.63) into (1.60) we find that we have to remove secular terms in order to obtain a bounded solution $u^{(2)}$ by imposing the condition

$$\partial_{T_1} A + C_g \partial_{X_1} A = 0, \tag{1.67}$$

where we define

$$C_g := \frac{d\omega}{dk} = \eta - 3\gamma k^2, \tag{1.68}$$

to be the group velocity associated with the operator L.

We find that (1.60) has a solution of the form

$$u^{(2)} = [\zeta/(6\gamma k^2)][A^2 e^{2i\theta} + A^{*2} e^{-2i\theta}] + B(X_1, T_1, T_2), \tag{1.69}$$

where B is a function yet to be determined.

Substituting (1.69) and (1.63) into (1.61) we have to impose the following conditions in order to remove secular terms

$$\partial_{T_1} B + \eta \partial_{X_1} B + \zeta \partial_{X_1} |A|^2 = 0, \tag{1.70}$$

and

$$\partial_{T_2} A + i[\zeta^2/(6\gamma k)]A|A|^2 + i\zeta k B A + 3i\gamma k \partial_{X_1}^2 A = 0. \tag{1.71}$$

Equations (1.67), (1.70) and (1.71) yield the required modulation equations, describing the behavior of the envelope A.

If we assume, on a physically reasonable basis, that B satisfies (1.67), i.e.,

$$\partial_{T_1} B + C_g \partial_{X_1} B = 0, \tag{1.72}$$

it follows upon substitution that

$$\partial_{X_1}[(\eta - C_g)B + \zeta|A|^2] = 0. \tag{1.73}$$

Therefore, (1.70) has a solution

$$B = -\zeta/(3\gamma k^2)|A|^2. \tag{1.74}$$

Consequently, (1.71) can be rewritten as

$$\partial_{T_2} A + 3i\gamma k \partial_{X_1}^2 A - [i\zeta^2/(6\gamma k)]A|A|^2 = 0, \tag{1.75}$$

the nonlinear (or cubic) Schrödinger (NLS) equation in the variables X_1 and T_2. The name *nonlinear Schrödinger* has been coined precisely because its structure is that of the Schrödinger equation of quantum mechanics with $|A|^2$ as a potential, although for most of the situations in which it occurs it has no relationship with the real quantum Schrödinger equation other than in name. The NLS equation serves as a model equation in its own right. This ubiquitous nonlinear wave problem of mathematical physics finds applications in such diverse fields as water waves, plasma physics and nonlinear optics [5, 36, 125, 138]. As the example illustrated above shows, it plays a significant role in the theory of the propagation of the envelopes of wave trains in many stable dispersive physical systems in which no dissipation occurs.

It is well known, for instance Zakharov and Kusnetsov [145] and Dodd *et al.* [36] that the nonlinear modulation properties of certain low amplitude periodic solutions of the KdV equation are described by a form of the nonlinear (or cubic) Schrödinger equation as given in (1.75).

It is important to note that (1.67) describes a linear modulation property of (1.46), whereas (1.75) gives information about the modulation effects of the nonlinear terms in (1.46).

In the numerical work that follows in a subsequent chapter we shall show that by using an exact discrete analogue of the continuous multiple scales techniques, a discrete version of the NLS is derived from the numerical scheme for the KdV, which should tell us something about the modulation of the numerical solution of the KdV equation, and therefore also point out any deviations from the behavior of the corresponding solution of the KdV (i.e. spurious behavior). In fact, it will be shown that this discrete version of the NLS is consistent with the continuous NLS obtained by the continuous multiple scales analysis, and could be viewed as a valid numerical scheme for the NLS as well.

1.3.3 Alternate Multiple Scales Expansion

The primary focus of this work will be the analysis of discretised equations by means of perturbation techniques. To that end, the following multiple scales method used by Tracy *et al.* [127] as well as Zakharov and Kusnetsov [145] is particularly suited for adaptation to the discrete case as illustrated by Schoombie [110, 111] and Maré and Schoombie [80, 112].

Consider the expansion

$$u(x,t) = \sum_{r=-\infty}^{\infty} u_r(X_1, T_1, T_2, \epsilon) e^{ir\theta}, \tag{1.76}$$

with θ given by (1.64) and k given by (1.66) as defined above. Then following Tracy *et al.* [127] as well as Zakharov and Kusnetsov [145] we use

$$u_r = \epsilon^{\delta_r} v_r(X_1, T_1, T_2, \epsilon), \tag{1.77}$$

with

$$\delta_0 = 2,\ \delta_r = |r|, \tag{1.78}$$

and

$$v_0 = V_0(X_1, T_1, T_2), \tag{1.79}$$

$$v_1 = V_1(X_1, T_1, T_2). \tag{1.80}$$

When $r > 1$ we have

$$v_r = V_r(X_1, T_1, T_2) + \sum_{s=r}^{\infty} \epsilon^{s+1-r} W_{rs}(X_1, T_1, T_2). \tag{1.81}$$

To enable a multiple scales analysis of (1.46), we use the expansions

$$\partial_t = -ir\omega + \epsilon\partial_{T_1} + \epsilon^2\partial_{T_2}, \tag{1.82}$$

and

$$\partial_x = irk + \epsilon\partial_{X_1}, \tag{1.83}$$

instead of the more general (1.57) and (1.58).

We now substitute (1.76) into the KdV equation (1.46) while making use of the expansions of the derivatives in (1.82) and (1.83). Putting the coefficient of each $e^{ir\theta}$ equal to zero, we find that (1.46) is equivalent to the following infinite set of equations for u_r

$$\begin{aligned} &-ir\omega u_r + \epsilon\partial_{T_1} u_r + \epsilon^2\partial_{T_2} u_r + ir\eta k u_r + \eta\epsilon\partial_{X_1} u_r \\ &+\gamma[irk + \epsilon\partial_{X_1}]^3 u_r + \zeta \sum_{s=-\infty}^{\infty} [iksu_s + \partial_{X_1}\epsilon u_s] u_{r-s} = 0. \end{aligned} \tag{1.84}$$

To proceed with the multiple scales analysis, we now wish to have (1.46), and therefore (1.84) satisfied for each r up to terms $O(\epsilon^3)$.

We commence by putting $r = 0$ in (1.84). By equating the $O(\epsilon^3)$ term (the lowest order term in ϵ) to zero we obtain

$$\partial_{T_1} V_0 + \eta\partial_{X_1} V_0 + \zeta\partial_{X_1}|V_1^2| = 0. \tag{1.85}$$

Next, put $r = 1$ in (1.84). Equating the $O(\epsilon)$ terms to zero reproduces the linear dispersion relation (1.65). Similarly the $O(\epsilon^2)$ and $O(\epsilon^3)$ terms yield the following equations respectively

$$\partial_{T_1} V_1 + C_g\partial_{X_1} V_1 = 0, \tag{1.86}$$

and

$$\partial_{T_2} V_1 + 3i\gamma k\partial_{X_1}^2 V_1 + ik\zeta[V_1 V_0 + V_1^* V_2] = 0, \tag{1.87}$$

where

$$C_g := \frac{d\omega}{dk} = \eta - 3\gamma k^2, \tag{1.88}$$

is the linear group velocity (as obtained in (1.68)).

When we use $r = 2$ we obtain from the $O(\epsilon^2)$ terms

$$V_2 = \zeta V_1^2/(6\gamma k^2), \tag{1.89}$$

by making use of the linear dispersion relation (1.65).

By making the physically meaningful assumption that V_0 also satisfies (1.86), that is

$$\partial_{T_1} V_0 + C_g \partial_{X_1} V_0 = 0, \tag{1.90}$$

with C_g as defined above, we obtain from (1.85) that

$$V_0 = -(\zeta/3\gamma k^2)|V_1|^2. \tag{1.91}$$

We rewrite (1.87) by making use of (1.91) and (1.89). The result

$$\partial_{T_2} V_1 - i[\zeta^2/(6\gamma k)]V_1(|V_1|^2) + 3i\gamma k\, \partial_{X_1}^2 V_1 = 0, \tag{1.92}$$

is the nonlinear Schrödinger equation in the variables T_2 and X_1 as obtained in the previous section, equation (1.75).

In terms of the expansion (1.76) as a solution of (1.46) it follows that

$$u(x,t) = \epsilon(V_1^* e^{-i\theta} + V_1 e^{i\theta}) - \epsilon^2(\zeta/6\gamma k^2)|V_1|^2 + O(\epsilon^3), \tag{1.93}$$

which can in turn be approximated by

$$u(x,t) \approx \epsilon(V_1^* e^{-i\theta} + V_1 e^{i\theta}), \tag{1.94}$$

since ϵ is small. Thus V_1 can be considered to be a small, variable amplitude of a monochromatic wave.

On the time scale T_1, (1.86) tells us that the modulation envelope V_1 moves at linear group velocity without changing its shape. On the time scale T_2, however, the envelope does change its shape, according to (1.87). Thus (1.86) describes the linear, and (1.92) describes the nonlinear modulation properties of the KdV equation.

If we identify V_1 in the above analysis with A in the previous analysis and correspondingly V_0 with B we observe that (1.85) and (1.86) is the same as (1.70) and (1.67) respectively. Furthermore, combining (1.87) and (1.89) we obtain (1.71). Thus exactly the same results are obtained as described in the previous section.

In the next section we shall illustrate the use of the alternative method of scales as illustrated above for the KdV equation for the Regularized Long Wave equation.

1.4 Regularized Long Wave (RLW) Equation

The RLW equation describes wave motion to the same order of approximation as the KdV equation (1.46) and could equally well model all of the applications of the KdV equation on the same formal basis of justification for both equations. Indeed [17], when the initial datum of both equations is restricted to conform to that arising in many physical applications, it can be shown that essentially the same solutions are obtained over a non-trivial time scale.

1.4.1 Introduction

We consider the Regularized Long Wave (RLW) equation in the form

$$\partial_t u + \eta \partial_x u + \zeta u \partial_x u - \gamma \partial_x^2 \partial_t u = 0, \tag{1.95}$$

where η, ζ and γ are constants with $\gamma \neq 0$. Similar to the boundary and initial conditions for the KdV equation we assume

$$u(x,0) = \epsilon f(x), \quad f(x) = O(1), \tag{1.96}$$

where ϵ is a small positive constant and

$$u(x \pm L, t) = u(x,t), \quad f(x \pm L) = f(x), \quad t > 0, \; x \in \Re. \tag{1.97}$$

The RLW equation was first put forward by Peregrine [100] to describe the temporal development of an undular bore.

Benjamin *et al.* [11] contend that "the RLW equation is in important respects the preferable model, obviating certain problematical aspects of the KdV equation and generally having more expedient mathematical properties".

The conditions for existence, stability and uniqueness of solutions of the IVP (1.95) to (1.97) was shown by Benjamin *et al.* [11]. Of some interest is the fact that the RLW has only three conservation laws as opposed to an infinite set of conservation laws for the KdV equation [17, 54, 116].

1.4.2 Continuous Multiple Scales Analysis

The perturbation approach we shall perform on the RLW equation (1.95) is the same as the alternate multiple scales analysis described for the KdV equation (1.46) in equations (1.76) through (1.94).

We start with the now familiar expansion [111, 127, 145],

$$u(x,t) = \sum_{r=-\infty}^{\infty} u_r(X_1, T_1, T_2, \epsilon) e^{ir\theta}, \tag{1.98}$$

where as in (1.76), (1.77) through (1.81)

$$u_r = \epsilon^{\delta_r} v_r(X_1, T_1, T_2, \epsilon), \tag{1.99}$$

with

$$\delta_0 = 2, \ \delta_r = |r|, \tag{1.100}$$

and

$$v_0 = V_0(X_1, T_1, T_2), \tag{1.101}$$

$$v_1 = V_1(X_1, T_1, T_2). \tag{1.102}$$

When $r > 1$ we have

$$v_r = V_r(X_1, T_1, T_2) + \sum_{s=r}^{\infty} \epsilon^{s+1-r} W_{rs}(X_1, T_1, T_2). \tag{1.103}$$

Furthermore θ is given by

$$\theta = kX_0 - \omega T_0, \tag{1.104}$$

where

$$\omega = \frac{\eta k}{\gamma k^2 + 1}, \tag{1.105}$$

subject to

$$\gamma k^2 + 1 \neq 0. \tag{1.106}$$

As before we also use

$$X_k = \epsilon^k x, \ \ k = 0, 1, \tag{1.107}$$

and

$$T_k = \epsilon^k t, \ \ k = 0, 1, 2. \tag{1.108}$$

We use the chain rule for derivatives equivalent equations given by (1.82) and (1.83). We substitute equations (1.98) through (1.108) into (1.95) and equate coefficients of $e^{ir\theta}$ to zero. This procedure yields the following system of equations:

$$\begin{aligned} &-ir\omega u_r + \epsilon \partial_{T_1} u_r + \epsilon^2 \partial_{T_2} u_r + ir\eta k u_r + \eta \epsilon \partial_{X_1} u_r \\ &-\gamma [irk + \epsilon \partial_{X_1}]^2 [-ir\omega + \epsilon \partial_{T_1} + \epsilon^2 \partial_{T_2}] u_r \\ &+\zeta \textstyle\sum_{s=-\infty}^{\infty} [iksu_s + \partial_{X_1} \epsilon u_s] u_{r-s} = 0. \end{aligned} \tag{1.109}$$

We now wish to have (1.95), and therefore (1.109) satisfied for each r up to terms $O(\epsilon^3)$.

We start with $r = 0$ in (1.109). The lowest order term in ϵ is $O(\epsilon^3)$ which yields the following equation upon equating the coefficient of $O(\epsilon^3)$ to zero

$$\partial_{T_1} V_0 + \eta \partial_{X_1} V_0 + \zeta \partial_{X_1} |V_1^2| = 0. \tag{1.110}$$

Next, we use $r = 1$. Equating the $O(\epsilon^2)$ term to zero we obtain

$$\partial_{T_1} V_1 + C_g \partial_{X_1} V_1 = 0, \tag{1.111}$$

where

$$C_g := \frac{d\omega}{dk} = \frac{\eta}{\gamma k^2 + 1} - \frac{2\eta\gamma k^2}{(\gamma k^2 + 1)^2}, \tag{1.112}$$

is the linear group velocity associated with the linearized RLW equation. The equation associated with setting the coefficient of the $O(\epsilon^3)$ term to zero is ($r = 1$)

$$(1 + \gamma k^2)\partial_{T_2} V_1 - 2i\gamma k \partial^2_{X_1 T_1} V_1 + \gamma i\omega \partial^2_{X_1} V_1 + ik\zeta[V_1 V_0 + V_1^* V_2] = 0. \tag{1.113}$$

Note that we can rewrite

$$-2i\gamma k \partial^2_{X_1 T_1} V_1 + \gamma i\omega \partial^2_{X_1} V_1 = i[\frac{3\eta\gamma k}{(\gamma k^2 + 1)} - \frac{4\eta\gamma^2 k^3}{(\gamma k^2 + 1)^2}]\partial^2_{X_1} V_1, \tag{1.114}$$

by making use of (1.105), (1.111) and (1.112).

By making use of $r = 2$ we find upon equating the $O(\epsilon^2)$ term to zero and making use of the linear dispersion relation (1.105) that

$$V_2 = \zeta \frac{(\gamma k^2 + 1)}{6\eta\gamma k^2} V_1^2. \tag{1.115}$$

To determine an expression for V_0 we make the reasonable assumption that it satisfies (1.111). Combining with (1.112) leads to the following expression

$$V_0 = -\zeta |V_1|^2 / (\eta - \frac{d\omega}{dk}). \tag{1.116}$$

Using (1.110) we find after algebraic manipulation that we can express V_0 in terms of V_1 as follows:

$$V_0 = -\zeta \frac{(\gamma k^2 + 1)^2}{\eta\gamma k^2(\gamma k^2 + 3)} |V_1|^2, \tag{1.117}$$

subject to

$$\gamma k^2 + 3 \neq 0, \tag{1.118}$$

and $k \neq 0$.

We are now in a position to combine equations (1.114), (1.115) and (1.117) in (1.113). The resulting equation

$$\partial_{T_2} V_1 - i\zeta^2 \frac{5\gamma k^2 + 3}{6\eta\gamma k(\gamma k^2 + 3)} V_1 |V_1|^2 - i/2 \frac{d^2\omega}{dk^2} \partial_{X_1}^2 V_1 = 0, \tag{1.119}$$

makes use of the fact that

$$\frac{d^2\omega}{dk^2} = [\frac{8\eta\gamma^2 k^3}{(\gamma k^2 + 1)^2} - \frac{6\eta\gamma k}{(\gamma k^2 + 1)}]. \tag{1.120}$$

Equation (1.119) is the nonlinear Schrödinger equation in the variables T_2 and X_1 as obtained in the previous section for the KdV equation (1.75).

This result is interesting for a number of reasons. Firstly, it shows that envelopes of modulated waves for the RLW equation are governed by the NLS equation, similar to the KdV equation. This confirms a result by Dodd *et al.* [36] namely, that for a class of partial differential equations

$$L(\partial_t, \partial_x)\phi = \sum_i (M^i\phi)(N^i\phi) + \sum_i (P^i\phi)(Q^i\phi)(R^i\phi),$$

where L, M, N, P, Q, and R are scalar differential operators in ∂_x and ∂_t, envelopes of modulated waves are governed by the NLS equation

$$2il_\omega \partial_{T_2} A + \partial_k^2 \omega \partial_X^2 A + \Gamma A|A|^2 = 0,$$

where $l(-i\omega, ik)$ describes the dispersion relation of L. Secondly, the result was obtained by making use of the alternative multiple scales analysis.

Chapter 2

Numerical Methods

"Can you do addition?" the White Queen asked.
"What's one and one and one and one and one and one and one and one and one and one and one?"
"I don't know," said Alice. "I lost count."

Through the Looking Glass. Lewis Carroll

Although this may seem a paradox, all exact science is dominated by the idea of approximation.

– Bertrand Russell (1872 - 1970)

While the multiple scales solutions to the equations in the previous chapter are elegant, detailed and accurate solutions are available only by making use of numerical methods. Our primary aim in this thesis is to investigate numerical solutions of the KdV equation (1.46) and the Van der Pol equation (1.26), specifically confining our attention to finite difference approximations.

In the next chapter we shall show that it is possible to use multiple scales techniques analogous to those described in Chapter 1 to analyze numerical approximations of the abovementioned differential equations. We shall particularly concentrate on a discrete multiple scales methodology applied by Schoombie [111] to the analysis of the Zabusky-Kruskal (ZK) approximation of the KdV equation. By comparing the results of the perturbation solutions to those of the corresponding analyses of the differential equations, spurious behavior in the numerical schemes can often be identified immediately as will be shown in subsequent chapters.

In this chapter we shall discuss finite difference methods. We shall devote the first two sections to preliminaries and notation, followed by a general derivation of finite difference formulae using the method of undetermined coefficients. This is followed by sections where we present an efficient methodology to compute finite difference coefficients on general grids and an analysis illustrating the accuracy of various orders of finite difference approximations focusing primarily on equi-spaced or regular grids. We also introduce the pseudospectral method and highlight the connection between central difference approximations of increasing orders of accuracy and the pseudospectral method.

2.1 Zabusky-Kruskal KdV Approximation

An example of a finite difference approximation to the KdV equation (1.46) is the ZK [144] explicit LF central finite difference scheme

$$\begin{aligned} u_j^{n+1} = u_j^{n-1} &- \frac{\tau}{h}(\eta + \frac{\zeta}{3}(u_{j+1}^n + u_j^n + u_{j-1}^n))(u_{j+1}^n - u_{j-1}^n) \\ &- \frac{\gamma\tau}{h^3}(u_{j+2}^n - 2u_{j+1}^n + 2u_{j-1}^n - u_{j-2}^n). \end{aligned} \tag{2.1}$$

We use the symbol u_j^n to denote the solution of a difference scheme at $x = hj,\ t = \tau n$ where j and n are given integers, i.e.,

$$u_j^n = u(hj, \tau n). \tag{2.2}$$

The parameters h and τ are discretization parameters to be defined below.

In the next section we shall provide operator notation for the analysis of finite difference approximations such as (2.1).

2.2 Preliminaries and Notation

To enable ourselves to analyze discretizations of the continuous differential equations we need to introduce some notation. Traditionally, but mostly for the sake of convenience, finite difference methods are considered on equally spaced grids. We therefore consider an equally spaced grid defined around the point x_0

$$x_0, x_0 \pm h, \ldots, x_0 \pm kh, \ldots \tag{2.3}$$

where h, a positive real number, is the grid spacing. In general, we are interested in grids defined on the plane (x, t). Therefore, by introducing a time step τ, we consider a grid defined by

$$(x_j, t_n) = (hj, \tau n), \tag{2.4}$$

for arbitrary integers j and n. We shall usually be interested in the solution of an initial value problem over some time period $t \in [0, T]$. The temporal discretization, τ, is typically given by T/N for some positive integer N, or determined by the specific time integration method.

On this grid we shall use the notation $u_j^n = u(hj, \tau n)$ as introduced in (2.2).

We define the following spatial shift operator ([18, 86, 111, 130], for example) for a general function $f = f(x, t)$:

$$E_x f(x, t) := f(x + h, t), \tag{2.5}$$

and similarly a temporal shift operator, namely

$$E_t f(x, t) := f(x, t + \tau). \tag{2.6}$$

Therefore using the notation defined in (2.2) above, we can write

$$E_x u_j^n \equiv u_{j+1}^n, \tag{2.7}$$

and

$$E_t u_j^n \equiv u_j^{n+1}, \tag{2.8}$$

for a function $u(x, t)$.

Making use of the shift operators, we next define divided difference operators on the grid defined above (2.3) as in the references [18, 57, 78, 86, 87, 103, 111]:

$$\Delta_x := (E_x - 1)/h, \tag{2.9}$$

$$\nabla_x := (1 - E_x^{-1})/h, \tag{2.10}$$

from which follows

$$\begin{aligned} \delta_x &:= (E_x - E_x^{-1})/2h \\ &\equiv (\Delta_x + \nabla_x)/2. \end{aligned} \tag{2.11}$$

Similarly

$$\Delta_t := (E_t - 1)/\tau, \tag{2.12}$$

$$\nabla_t := (1 - E_t^{-1})/\tau, \tag{2.13}$$

$$\delta_t \equiv (\Delta_t + \nabla_t)/2. \tag{2.14}$$

The operators could be defined for more arbitrary grids than (2.3), however, for the purpose of most of our work we shall consider equi-spaced grids. Since all of the above operators depend on h or τ a more complete notation would be to use $\Delta_x(h)$, $\nabla_x(h)$ and $\delta_x(h)$. Using the operator δ_x defined in equation (2.11), as an example, the symbol $\delta_x(2h)$ would be defined by

$$\begin{aligned}\delta_x(2h)\,u_j &= \frac{1}{2}(\Delta_x(2h) + \nabla_x(2h))\,u_j \\ &= \frac{1}{4h}(E_x^2 - E_x^{-2})\,u_j \\ &= \frac{1}{4h}(u_{j+2} - u_{j-2}).\end{aligned}$$

In subsequent formulas we shall only show the more complete notation when needed. These operators will be especially useful in later sections and generally to denote difference schemes for the numerical solution of ODEs and PDEs.

2.2.1 Order of Approximation

The operators introduced above are well-known in the literature. Consider the problem of evaluating du/dx at a grid point $x = x_0$ when u is defined only at the equally spaced grid points in (2.3). By making use of a Taylor series expansion [23, 62, 76] we have

$$u(x_0 + h) = u(x_0) + hu'(x_0) + \frac{h^2}{2!}u''(x_0) + \frac{h^3}{3!}u'''(\xi_1),$$

and

$$u(x_0 - h) = u(x_0) - hu'(x_0) + \frac{h^2}{2!}u''(x_0) - \frac{h^3}{3!}u'''(\xi_2),$$

where $\xi_1 \in (x_0, x_0 + h)$ and $\xi_2 \in (x_0 - h, x_0)$.

Applying the operator δ_x defined in (2.11) to $u(x)$ and making use of a Taylor series expansion of $u(x)$ above we have

$$\begin{aligned}\delta_x u(x_0) &= \frac{1}{2h}[u(x_0 + h) + u(x_0 - h)] \\ &= u'(x_0) + \frac{h^2}{3!}u'''(\xi), \qquad (2.15)\end{aligned}$$

where $\xi \in (x_0 - h, x_0 + h)$. The approximation shown here would be exact for second degree polynomials. This leads us to define the order of approximation of a finite difference expression as follows:

Definition 1 *[49, 99] $f(h) = O(g(h))$ as $h \to 0$ if there exists constants $C > 0$ and $h_0 > 0$ such that $|f(h)| \leq Cg(h)$ for all $h \leq h_0$. Provided $g(h) \neq 0$ this means that $|f(h)|/g(h)$ is bounded from above for all sufficiently small h.*

From the above definition the operator δ_x defined in (2.11) results in an $O(h^2)$ approximation to the first derivative of a function as shown in (2.15).

2.2.2 Zabusky-Kruskal Approximation Revisited

To illustrate the use of the operators defined above we rewrite the ZK approximation (2.1) to the KdV equation in operator form, thus [111]

$$\delta_t u_j^n + (\eta + \zeta(1 + h(\Delta_x - \nabla_x)/3)u_j^n)\delta_x u_j^n + \gamma \delta_x \Delta_x \nabla_x u_j^n = 0. \tag{2.16}$$

To complete the discrete initial value analogue to the continuous problem (1.46) to (1.48) we have to impose periodic boundary conditions [111]

$$u_{j\pm N}^n = u_j^n, \tag{2.17}$$

as well as prescribe initial data of the form

$$u_j^0 = \epsilon f_j, \tag{2.18}$$

with

$$f_j = O(1), \quad f_{j\pm N}^n = f_j^n, \tag{2.19}$$

where ϵ is a small, real, positive parameter as before. This now constitutes an example of the type of discrete initial boundary value problem that we wish to study. The initial conditions stated above could be more general; however, numerical studies in a later chapter will frequently use the above type of initial conditions.

It is important to note that the ZK approximation is consistent with the continuous KdV equation (1.46) with a truncation error of order $(O(h^2) + O(\tau^2))$.

We shall in particular consider the modulation properties of solutions of the discrete approximation techniques. For the purpose of our analysis we shall frequently use

the Method of Lines (MOL) to circumvent temporal discretization. As an example, using the MOL, the ZK discretization becomes

$$\frac{du_j}{dt} + (\eta + \zeta(1 + h(\Delta_x - \nabla_x)/3)u_j)\delta_x u_j + \gamma\delta_x\Delta_x\nabla_x u_j = 0.$$

We wish to study more general finite difference approximations to the KdV equation than that given by the ZK discretization above. In the next two sections we shall show how to derive difference expressions to prescribed order of accuracy for derivatives of sufficiently smooth functions. Using these expressions would lead to higher order, more general, difference approximations of the KdV equation.

2.3 General Derivation of Finite Differences

The ZK approximation (2.16) to the KdV equation (1.46) was obtained by replacing continuous partial derivatives with second order accurate difference operators. In this section we shall consider the derivation of difference operators with higher orders of accuracy using the method of undetermined coefficients.

We wish to show that the derivative d^k/dx^k for a sufficiently smooth function $u(x)$, of arbitrary order k, can be replaced by a difference expression such that the error induced by this replacement will be of any prescribed order, p, i.e., $O(h^p)$.

Following Godunov and Ryabenkii [57] we write an equation of the form,

$$\frac{d^k u(x)}{dx^k} = h^{-k} \sum_{s=-s_1}^{s_2} a_s E^s u(x) + O(h^p), \tag{2.20}$$

based on the grid defined in (2.3) as well as the shift operator E defined in equation (2.5). The limits of summation can be chosen arbitrarily provided that the order of the difference equation satisfies the inequality, $s_1 + s_2 \geq k + p - 1$ where $s_1, s_2 \geq 0$. By making use of a Taylor series expansion we have

$$\begin{aligned} E^s u(x) = u(x) + sh\frac{du(x)}{dx} + \frac{(sh)^2}{2!}\frac{d^2u(x)}{dx^2} + \dots \\ \dots + \frac{(sh)^{k+p-1}}{(k+p-1)!}\frac{d^{k+p-1}u(x)}{dx^{k+p-1}} + \frac{(sh)^{k+p}}{(k+p)!}\frac{d^{k+p}u(\xi)}{dx^{k+p}}, \end{aligned} \tag{2.21}$$

where $\xi \in (x, x + sh)$.

Substituting the above result into equation (2.20), in place of $E^s u(x)$, and collecting like terms we obtain

$$\frac{d^k u(x)}{dx^k} = h^{-k}[u(x)\sum a_s + h\frac{du(x)}{dx}\sum s a_s + \dots$$
$$\dots + \frac{h^{k+p-1}}{(k+p-1)!}\frac{d^{k+p-1}u(x)}{dx^{k+p-1}}\sum s^{k+p-1}a_s]$$
$$+\frac{h^p}{(k+p)!}\sum s^{k+p}a_s\frac{d^{k+p}u(\xi)}{dx^{k+p}}, \tag{2.22}$$

where we have suppressed the subscripts of the summation for ease of notation.

By equating coefficients of like powers h^s, where $s = -k, \dots, p-1$ on the left- and right-hand sides of equation (2.22) we derive the following system of equations for the a_s:

$$\begin{aligned}
&\sum a_s = 0,\\
&\sum s a_s = 0,\\
&\vdots\\
&\sum s^{k-1}a_s = 0,\\
&\sum s^k a_s = k!,\\
&\sum s^{k+1}a_s = 0,\\
&\vdots\\
&\sum s^{k+p-1}a_s = 0.
\end{aligned} \tag{2.23}$$

If $s_1 + s_2 = k + p - 1$, the $k + p$ equations in (2.23) form a linear system of the same number of unknowns a_s. The determinant of this system is the well-known Vandermonde determinant [55], and is different from zero [57, 132]. Therefore, there exist a unique set of coefficients a_s satisfying (2.23). Should $s_1 + s_2 \geq k + p$ then many such systems of coefficients would exist.

As an example of the application of this analysis we consider second-order difference expressions of the form

$$\frac{1}{h}[a_{-1}u(x-h) + a_0 u(x) + a_1 u(x+h)], \tag{2.24}$$

as approximations to du/dx. Clearly there are infinitely many approximations to first order in h ($p = 1$), however, only one solution is of second order accuracy. Solving

the system of equations (2.23) in this case we find that

$$\begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{-1} \\ a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Therefore, $a_{-1} = -1/2$, $a_0 = 0$, and $a_1 = 1/2$ yielding

$$\begin{aligned} \frac{du}{dx} &= \frac{u(x+h) - u(x-h)}{2h} + O(h^2) \\ &= \delta_x u + O(h^2), \end{aligned}$$

as obtained in (2.15).

By making use of the methodology described above we are able to construct difference schemes, by replacing the derivatives in a given differential equation with difference expressions as given by (2.20), with any prescribed order of approximation.

Finite difference formulae for equi-spaced grids are readily available in tables and can be obtained from symbolic manipulation of difference operators. In the next section we show an algorithm to calculate general finite difference approximations to higher order derivatives on arbitrary grids.

2.4 Polynomial Interpolation and FD Stencils

In this section we shall show an alternative methodology to determine weights in finite difference formulas. The methodology is elegant and especially suited for efficient computer implementation. The methodology can be implemented on arbitrary spaced grids. Although we shall confine ourselves for most of this study to equi-spaced grids we shall make an exception here in order to show the mathematical elegance of the derivation.

The approach is to construct Lagrange interpolatory polynomials of a specified order and subsequently evaluate the derivative of these polynomials at the grid points to obtain the coefficients of the finite-difference stencil.

We consider a set of values $u_i = u(x_i)$ at the locations x_i, $i = 0, 1, \ldots, N$ which are arbitrary, yet distinct points in $\Re$. It is well-known from the literature [19, 23] that there exists a unique polynomial P of degree at most N with the property that

$$P(x_i) = u(x_i), \quad \text{for each } i = 0, 1, \ldots, N. \tag{2.25}$$

In general [19, 23], one may fit any $N+1$ points by a polynomial of $N-$th degree via the Lagrange interpolation formula, namely

$$P_N(x) = \sum_{i=0}^{N} u(x_i) L_{i,N}(x), \tag{2.26}$$

where the $L_{i,N}(x)$ are defined by

$$L_{i,N}(x) = \prod_{\substack{j=0 \\ j \neq i}}^{N} \frac{x - x_j}{x_i - x_j}. \tag{2.27}$$

The N factors of $(x - x_j)$ ensure that $L_{i,N}(x)$ vanishes at all the interpolation points except x_i. The denominator forces $L_i(x)$ to equal 1 at the interpolation point $x = x_i$; at that point every factor in the product is $(x_i - x_j)/(x_i - x_j) = 1$. Therefore

$$L_{i,N}(x_j) = \delta_{i,j}, \tag{2.28}$$

where $\delta_{i,j}$ is the familiar Kronecker $\delta-$function.

Theorem 1 *Let $u(x)$ have at least $N+1$ derivatives on the interval of interest and let $P_N(x)$ be its Lagrangian interpolant of degree N. Then*

$$u(x) - P_N(x) = \frac{1}{(N+1)!} \frac{d^{N+1} f(\xi)}{dx^{N+1}} \prod_{i=0}^{N} (x - x_i), \tag{2.29}$$

for some ξ on the interval spanned by x and the interpolation points.

Proof: The proof of this theorem is well-known and contained in the text of [23], for example. □

From the theorem it is clear that interpolation for polynomials of degree N would be exact. It is both interesting and useful to note that we can rewrite the Lagrange polynomial. We denote

$$\omega_N(x) = \prod_{j=0}^{N} (x - x_j). \tag{2.30}$$

Evaluating

$$\omega'_N(x_i) = \prod_{\substack{j=0 \\ j \neq i}}^{N} (x_i - x_j),$$

we have, as in [45],

$$L_{i,N}(x) = \frac{\omega_N(x)}{(x - x_i)\omega'_N(x_i)}, \tag{2.31}$$

the analogue of (2.27). From (2.30) we find the following recursion relation

$$\omega_N(x) = (x - x_N)\omega_{N-1}(x), \tag{2.32}$$

from which follows that

$$\omega'_N(x) = (x - x_N)\omega'_{N-1}(x) + \omega_{N-1}(x). \tag{2.33}$$

Therefore, using (2.31) and (2.33) we have for $i < N$

$$L_{i,N}(x) = \frac{x - x_N}{x_i - x_N} L_{i,N-1}(x), \tag{2.34}$$

and (for $N > 1$) with $i = N$

$$\begin{aligned} L_{N,N}(x) &= \frac{\omega_{N-1}(x)}{\omega_{N-1}(x_N)} \\ &= (x - x_{N-1}) \frac{\omega_{N-2}(x_{N-1})}{\omega_{N-1}(x_N)} L_{N-1,N-1}(x). \end{aligned} \tag{2.35}$$

We shall now use the recursion relations (2.34) and (2.35) to generate finite difference weights. For the sake of simplicity we seek to approximate derivatives of $u(x)$ at $x = 0$. Considering Theorem 1 we shall approximate $u(x)$ by $P_N(x)$ and consider derivatives of $P_N(x)$ as approximations to derivatives of $u(x)$, i.e.,

$$\begin{aligned} \frac{d^k u(x)}{dx^k}|_{x=0} &\approx \frac{d^k p_j(x)}{dx^k}|_{x=0} \\ &= \sum_{i=0}^{j} \frac{d^k L_{i,j}(x)}{dx^k}|_{x=0}\, u_i \\ &\equiv \sum_{i=0}^{j} \delta_{i,j}^k u_i, \end{aligned} \tag{2.36}$$

where we define

$$\frac{d^k L_{i,j}(x)}{dx^k}|_{x=0} \equiv \delta_{i,j}^k. \tag{2.37}$$

By making use of Taylor's formula

$$L_{i,j}(x) = \sum_{k=0}^{j} \frac{d^k L_{i,j}(x)}{dx^k}|_{x=0}\, \frac{x^k}{k!}. \tag{2.38}$$

Consequently, by using the definition for $\delta_{i,j}^k$ above it follows that

$$L_{i,j}(x) = \sum_{k=0}^{j} \delta_{i,j}^k \frac{x^k}{k!}. \tag{2.39}$$

By substituting the expansion (2.39) into the recursion relations (2.34) and (2.35) respectively, and by equating powers of x, we obtain the following recursion relations between the weights [45, 47]:

$$i < N : \ \delta_{i,N}^k = \frac{1}{x_N - x_i}(x_N \delta_{i,N-1}^k - k\delta_{i,N-1}^{k-1}), \tag{2.40}$$

and

$$i = n : \ \delta_{N,N}^k = \frac{\omega_{N-2}(x_{N-1})}{\omega_{N-1}(x_N)}(k\delta_{N-1,N-1}^{k-1} - x_{N-1}\delta_{N-1,N-1}^k). \tag{2.41}$$

We could implement the recursion relations (2.40) and (2.41) in an algorithm. From this single algorithm one could derive coefficients for centered, one-sided and more general approximations to all kinds of derivatives.

The above method is particularly useful when dealing with adaptive methods where the grid is adjusted as well as the order of the finite difference stencil. It is important to realize that the above construction depends only on the grid points x_i. We have illustrated the method as it provides a methodology to derive high order finite difference approximations. In the next section we show finite difference coefficients for higher orders than two for equi-spaced grids.

2.5 Regular Grids

The special case of regular or equi-spaced grids is of specific practical importance as most practioners consider finite difference approximations to partial differential equations on such grids.

In Table 2.1 we consider finite difference weights for centered approximations to the first derivative of a function. The weights in the table correspond to the choice $h = 1$ for the grid spacing in (2.3). The weights are derived from the algorithm above but could equally well be derived from equation (2.23). We show approximations with orders of accuracy ranging from 2 to 10.

Order	-5	-4	-3	-2	-1	0	1	2	3	4	5
2					$\frac{-1}{2}$	0	$\frac{1}{2}$				
4				$\frac{1}{12}$	$\frac{-2}{3}$	0	$\frac{2}{3}$	$\frac{-1}{12}$			
6			$\frac{-1}{60}$	$\frac{3}{20}$	$\frac{-3}{4}$	0	$\frac{3}{4}$	$\frac{-3}{20}$	$\frac{1}{60}$		
8		$\frac{1}{280}$	$\frac{-4}{105}$	$\frac{1}{5}$	$\frac{-4}{5}$	0	$\frac{4}{5}$	$\frac{-1}{5}$	$\frac{4}{105}$	$\frac{-1}{280}$	
10	$\frac{-1}{1260}$	$\frac{5}{504}$	$\frac{-5}{84}$	$\frac{5}{21}$	$\frac{-5}{6}$	0	$\frac{5}{6}$	$\frac{-5}{21}$	$\frac{5}{84}$	$\frac{-5}{504}$	$\frac{1}{1260}$

Table 2.1: Finite difference weights for centered approximations to first order derivatives.

Similarly, in Table 2.2 we show finite difference weights for centered approximations to the second derivative of a function.

Order	-5	-4	-3	-2	-1	0	1	2	3	4	5
2					1	-2	1				
4				$\frac{-1}{12}$	$\frac{4}{3}$	$\frac{-5}{2}$	$\frac{-4}{3}$	$\frac{-1}{12}$			
6			$\frac{1}{90}$	$\frac{-3}{20}$	$\frac{3}{2}$	$\frac{-49}{18}$	$\frac{3}{2}$	$\frac{-3}{20}$	$\frac{1}{90}$		
8		$\frac{-1}{560}$	$\frac{8}{315}$	$\frac{-1}{5}$	$\frac{8}{5}$	$\frac{-205}{72}$	$\frac{8}{5}$	$\frac{-1}{5}$	$\frac{8}{315}$	$\frac{-1}{560}$	

Table 2.2: Finite difference weights for centered approximations to second order derivatives.

In Table 2.3 we show weights for the third derivative of a function. Note that we use the anti-symmetry of the weights for odd-order derivatives (as demonstrated for the first order derivative weights in Table 2.1) and consequently the weights are shown for positive references only.

We focus specifically on regular grids for the purpose of this study. Let D_{2p} denote the discrete first order derivative spatial difference operator obtained by interpolating $f(x_{j-p}), \ldots, f(x_{j+p})$ on the stencil

$$x_j + \alpha h, \quad \alpha = 0, \pm 1, \pm 2, \cdots \pm p,$$

by a polynomial of degree $2p$ and then differentiating it once at x_j.

$$\frac{df}{dx}|_{x=x_j} \approx D_{2p} f(x_j) = \frac{1}{h} \sum_{\alpha=-p}^{p} \delta_{p,\alpha} f(x_j + \alpha h). \tag{2.42}$$

Order	0	1	2	3	4	5	6
2	0	-1	$\frac{1}{2}$				
4	0	$\frac{-13}{8}$	1	$\frac{-1}{8}$			
6	0	$\frac{-61}{30}$	$\frac{169}{120}$	$\frac{-3}{10}$	$\frac{7}{240}$		
8	0	$\frac{-1669}{720}$	$\frac{4369}{2520}$	$\frac{-541}{1120}$	$\frac{1261}{15120}$	$\frac{-41}{6048}$	
10	0	$\frac{-1769}{700}$	$\frac{4469}{2240}$	$\frac{-4969}{7560}$	$\frac{643}{4200}$	$\frac{-19}{840}$	$\frac{479}{302400}$

Table 2.3: Finite difference weights for centered approximations to third order derivatives.

Fornberg [44] derived the following explicit formula for arbitrary order of accuracy $2p$ for the calculation of the $\delta_{p,\alpha}$ for this special case:

$$\delta_{p,\alpha} = \frac{(\alpha!)^2(-1)^{\alpha+1}}{\alpha(p+\alpha)!(p-\alpha)!}, \ \alpha \neq 0, \tag{2.43}$$

and $\delta_{p,0} = 0$. The $\delta_{p,\alpha}$ could alternatively be read from Table 2.1.

In general we denote D_{2p}^m as the discrete m-th order spatial difference operator obtained from interpolation of $f(x_{j-p}), \ldots, f(x_{j+p})$ by a polynomial of degree $2p$ and then differentiating m times and use the notation:

$$\frac{d^m f}{dx^m}|_{x=x_j} \approx D_{2p}^m f(x_j) = \frac{1}{h^m} \sum_{\alpha=-p}^{p} \delta_{p,\alpha}^m f(x_j + \alpha h), \tag{2.44}$$

where, as in Fornberg's paper [45],

$$\frac{1}{h^m}\delta_{p,\alpha}^m = \left[\frac{d^m}{dx^m} F_{p,\alpha}(x)\right]_{x=x_j}, \tag{2.45}$$

with

$$F_{p,\alpha}(x) = \frac{\omega_p(x)}{\omega_p'(x_j + \alpha h)(x - x_j - \alpha h)}, \tag{2.46}$$

and

$$\omega_p(x) = \prod_{\beta=-p}^{p} (x - x_j - \beta h). \tag{2.47}$$

Moreover,

$$p(x) = \sum_{\alpha=-p}^{p} F_{p,\alpha}(x) f(x_j + \alpha h), \tag{2.48}$$

is the Lagrange interpolation polynomial of degree $2p$ on our finite difference stencil.

We note from Table 2.1 (and formula (2.43)) the anti-symmetry of the finite difference coefficients, i.e.,

$$\delta^m_{p,\alpha} = -\delta^m_{p,-\alpha}, \tag{2.49}$$

whenever the order of the spatial derivative is odd. Using this property and the shift operator E defined in (2.5), we write the following expression for the finite difference operator D^m_{2p} (m odd):

$$D^m_{2p} = \frac{1}{h^m} \sum_{\alpha=1}^{p} \delta^m_{2p,\alpha}(E^\alpha - E^{-\alpha}). \tag{2.50}$$

Later on we shall also have need of the following identities which we formulate in the Theorem:

Theorem 2 *Schoombie and Maré [112]: Let m be an odd integer, with $1 \leq m < 2p$, and let $\delta^m_{2p,\alpha}$ be defined as in equation (2.45). Then*

$$\sum_{\alpha=1}^{p} \alpha^k \delta^m_{2p,\alpha} = 0 \text{ for } 0 < k < m, \tag{2.51}$$

with k an integer, and

$$\sum_{\alpha=1}^{p} \alpha^m \delta^m_{2p,\alpha} = \frac{m!}{2}. \tag{2.52}$$

Proof: The interpolation approximation in (2.48) is exact for $f(x)$ a polynomial of degree $\leq 2p$. Therefore it follows that

$$x^k = \sum_{\alpha=-p}^{p} F_{p,\alpha}(x)(x_j + \alpha h)^k, \tag{2.53}$$

with $0 < k \leq m$ an odd integer.

By differentiating the above expression m times with respect to x, and putting

$$x = x_j = jh,$$

we have that

$$\begin{aligned} \sum_{\alpha=-p}^{p} \delta^m_{p,\alpha}(j+\alpha)^k &= \sum_{\alpha=1}^{p} \delta^m_{2p,\alpha}[(j+\alpha)^k - (j-\alpha)^k] \\ &= \begin{cases} m! & \text{if } k = m \\ 0 & \text{if } k < m. \end{cases} \end{aligned} \tag{2.54}$$

Thus, for $m = 1$, the identity in (2.52) follows immediately. For $m > 1$, first let $k = 1$ in equation (2.54). Then the identity (2.51) follows for $k = 1$. For $k = 3 < m$, equation (2.54) becomes

$$6j^2 \sum_{\alpha=1}^{p} \alpha \delta_{2p,\alpha}^{m} + 2 \sum_{\alpha=-p}^{p} \alpha^3 \delta_{2p,\alpha}^{m} = 0.$$

The first term vanishes by virtue of (2.51) for $k = 1$, and what remains proves equation (2.51) for $k = 3$. Proceeding in this fashion, equation (2.51) is proved for any odd k such that $0 < k < m$.

The identity (2.52) is proved similarly, putting $k = m$ in (2.54) and removing superfluous terms by means of equation (2.51). □

This theorem could also be proved as a consequence of equations (2.22) and (2.23).

2.6 Pseudospectral Methods

Given $N > 0$, we consider the set of points

$$x_j = \frac{2\pi j}{N}. \tag{2.55}$$

The discrete Fourier coefficients of a complex-valued function u in $[0, 2\pi]$ with respect to these points are given by [25, 30, 46, 97]

$$\hat{u}_k = F u_k = \frac{1}{N} \sum_{j=0}^{N-1} u(x_j) e^{-ikx_j}, \tag{2.56}$$

with $-N/2 \leq k \leq N/2 - 1$. Due to the orthogonality relation [1]

$$\frac{1}{N} \sum_{j=0}^{N-1} e^{ipx_j} = \begin{cases} 1 & \text{if } p = Nm,\ m = 0, \pm 1, \pm 2, \ldots \\ 0 & \text{otherwise,} \end{cases}$$

we have the inversion formula

$$u(x_j) = F^{-1} \hat{u}_k = \sum_{k=-N/2}^{N/2-1} \hat{u}_k e^{ikx_j}, \tag{2.57}$$

with $j = 0, \ldots, N - 1$.

The transformations in (2.56) and (2.57) can be performed efficiently by means of the fast Fourier transform (FFT) algorithm [20, 46].

Once we have calculated the discrete coefficients $\hat{u}_k$, the approximate derivative u'_j can be computed through

$$u'_j = \sum_{k=-N/2}^{N/2-1} a_k e^{ikx_j}, \tag{2.58}$$

where

$$a_k = \frac{ik}{N} \sum_{j=0}^{N-1} u(x_j) e^{-ikx_j}, \tag{2.59}$$

a process requiring two FFTs.

We quote the following result from Canuto *et al.* [25]. If u is a 2π-periodic analytic function in the strip $|\Im(z)| < \eta_0$ then

$$||\frac{du}{dx} - u'||_{L_2} \leq C(\eta) N e^{-N\eta/2},$$

for all η such that $0 < \eta < \eta_0$. The error therefore decays essentially exponentially in N [118]. It is also of interest to note that in the case of an analytic function, the asymptotic rate of decay of the Fourier coefficients of the function $u(x)$ is $u_k = O(exp(-\eta|k|))$ as $|k| \to \infty$, see e.g., Dieudonné [34] and also [30].

2.7 Limiting Case

In this section we shall consider the connection between finite difference and pseudospectral methods. In particular we consider the limiting case of the finite difference method. (For ease of notation we consider an odd number of points, $N = 2p + 1$.)

By considering equation (2.43) we find:

$$\lim_{p\to\infty} \delta_{p,\alpha} = \frac{(-1)^{\alpha+1}}{\alpha}, \tag{2.60}$$

which would yield an infinite stencil with coefficients

$$\{\ldots, \frac{1}{4}, \frac{-1}{3}, \frac{1}{2}, -1, 0, 1, \frac{-1}{2}, \frac{1}{3}, \frac{-1}{4}, \ldots\}/h. \tag{2.61}$$

Following Fornberg [44], on a periodic grid the stencil above would collapse to a stencil of width $2p+1$ with the coefficients

$$\{\gamma_{p,-p}, \ldots, \gamma_{p,-1}, \gamma_{p,0}, \gamma_{p,1}, \ldots, \gamma_{p,p}\}/h, \tag{2.62}$$

where

$$\begin{aligned} \gamma_{p,\alpha} &= (-1)^{\alpha+1} \sum_{k=-\infty}^{\infty} \frac{(-1)^k}{Nk+\alpha}, \quad 0 < \alpha \leq p, \\ &= (-1)^{\alpha+1} \frac{\pi}{N \sin(\frac{\pi\alpha}{N})}. \end{aligned} \tag{2.63}$$

We evaluate the sum in (2.63) by making use of contour integrals [33, 44]. Note furthermore that $\gamma_{p,0} = 0$ and $\gamma_{p,\alpha} = -\gamma_{p,\alpha}$.

We shall now consider the pseudospectral method as discussed in Section 2.6. Consider a function $f(kh)$ defined at the grid points kh, $k = -p, \ldots, p$. We use $N = 2p+1$, hence the interpolating trigonometric polynomial, as defined before in equation (2.57), is

$$f(x) = \sum_{\omega=-p}^{p} \hat{f}(\omega) e^{i\omega x}, \tag{2.64}$$

where

$$\hat{f}(\omega) = \frac{1}{N} \sum_{k=-p}^{p} f(kh) e^{-ikh\omega}. \tag{2.65}$$

We evaluate the derivative of (2.64) at $x = \alpha h$, i.e.,

$$\begin{aligned} \frac{df}{dx}|_{x=\alpha h} &= i \sum_{\omega=-p}^{p} \omega \hat{f}(\omega) e^{i\omega\alpha h} \\ &= \frac{i}{N} \sum_{k=-p}^{p} f(kh) \sum_{\omega=-p}^{p} \omega e^{i\omega h(\alpha-k)}. \end{aligned} \tag{2.66}$$

We wish to find a closed-form expression for (2.66). Note that

$$\sum_{\omega=-p}^{p} e^{-i\omega x} = \frac{\sin(Nx/2)}{\sin(x/2)}, \quad x \neq 0. \tag{2.67}$$

Therefore

$$\frac{d}{dx} \sum_{\omega=-p}^{p} e^{-i\omega x}|_{x=\alpha h=\frac{2\pi\alpha}{N}} = i \frac{(-1)^\alpha N}{2 \sin(\frac{\pi\alpha}{N})}. \tag{2.68}$$

Hence

$$\begin{aligned}\delta_{p,\alpha} &= \frac{2\pi i}{N^2}\sum_{\omega=-p}^{p}\omega\hat{f}(\omega)e^{i\omega\alpha h} \\ &= \frac{(-1)^{\alpha+1}\pi}{N\sin(\frac{\pi\alpha}{N})}. \qquad (2.69)\end{aligned}$$

Note that (2.69) is identical to equation (2.63).

We have therefore shown that Fourier differencing can be viewed as a special centered finite differencing based on an ever increasing number of periodic stencils [44, 123, 130], i.e.,

$$F = \lim_{p\to\infty} D_{2p}. \qquad (2.70)$$

2.8 Finite Difference Fourier Mode Analysis

Suppose we wish to approximate d/dx. For a specific mode $e^{i\omega x}$ we obtain the exact answer

$$\frac{d}{dx}e^{i\omega x} = i\omega e^{i\omega x}. \qquad (2.71)$$

We wish to show the effect of finite difference discretizations applied to $e^{i\omega x}$.

A second order finite difference approximation would yield

$$\begin{aligned}D_2 e^{i\omega x} &= \frac{e^{i\omega(x+h)} - e^{i\omega(x-h)}}{2h} \\ &= i\frac{\sin(\omega h)}{h}e^{i\omega x} \qquad (2.72) \\ &\equiv if(2,h,\omega)e^{i\omega x}. \qquad (2.73)\end{aligned}$$

In general for a $2p$-th order approximation we would write

$$D_{2p}e^{i\omega x} = if(2p,h,\omega)e^{i\omega x}, \qquad (2.74)$$

subject to the difference coefficients derived in Section 2.5. In Table 2.4 we show $f(2p,h,\omega)$ for a range of values.

In Table 2.5 we show $f(2p,h,\omega)$ for a range of wavenumbers ($N = 128$) and different orders of approximation. For the second order approximation only a fraction of the

Order	$f(2p,h,\omega)$
2	$\sin(\omega x)/h$
4	$\frac{1}{3}(4-\cos(\omega x))\sin(\omega x)/h$
6	$(\frac{8}{5}-\frac{3}{5}\cos(\omega x)-\frac{2}{15}\sin^2(\omega x))\sin(\omega x)/h$
8	$(\frac{64}{35}-\frac{29}{35}\cos(\omega x)-\frac{32}{105}\sin^2(\omega x)+\frac{2}{35}\sin^2(\omega x)\cos(\omega x))\sin(\omega x)/h$
$\vdots$	$\vdots$
∞	ω

Table 2.4: Multiplicative factors $f(2p,h,\omega)$ arising from the $2p$-th order FD approximation to $ie^{i\omega x}$.

	Wavenumber					
Order	2	4	8	16	32	48
2	1.9967	3.9743	7.7959	14.4051	20.3718	14.4051
4	1.9999	3.9998	7.9937	15.8114	27.1624	22.6021
6	2.0000	3.9999	7.9997	15.9762	29.8787	28.1993
8	2.0000	4.0000	7.9999	15.9969	31.0428	32.2944
$\vdots$						
∞	2.0000	4.0000	8.0000	16.0000	32.0000	48.0000

Table 2.5: Values of $f(2p,h,\omega)$ for various wavenumbers.

modes are treated correctly. For the higher order aproximations shown, a significant increase in the accurate representation of modes is observed, although the very high mode numbers are still not accurate, i.e., the convergence for higher orders to the ideal line (the $i\omega$-factor in (2.71)) is slow - reminiscent of the convergence of a Taylor series where derivatives of successively higher orders match at the origin.

As an example, we note the approximation of the sixth order scheme, which provides very accurate approximations for mode numbers up to 16. It is also interesting to note that approximation orders higher than six do not increase the accuracy of representation of the lower mode numbers significantly.

Chapter 3

Discrete Perturbation Techniques

It is a mathematical fact that the casting of this pebble from my hand alters the centre of gravity of the universe.

– Thomas Carlyle (1795 - 1881)

We described the method of multiple scales for continuous partial differential equations in great detail in Chapter 1. In Chapter 2 we provided a framework from which finite difference approximations to partial differential equations could be derived.

Our aim in this work is to consider solutions of these finite difference approximations. To this end we shall develop an analysis technique to extend the continuous multiple scales methodology to discrete difference equations. We shall devote this chapter to show this development as well as cover some other approaches in the literature.

In the first section we shall provide notation needed for our analysis. This notation will build on the notation provided in Chapter 2. In the second section we shall show a special difference identity which serves the same purpose as the chain rule for derivatives of sufficiently regular functions.

3.1 Partial Difference Operators - Notation

Our aim in this section is to provide the framework to extend the continuous multiple scales methodology to discrete difference equations. In analogy with the continuous scales coordinates, for example (1.30), (1.55) and (1.56), we shall use the following

discrete multiple scales coordinates in space, notably [110, 111, 112]

$$X_p = \epsilon^p hj, \quad p = 0, 1, \ldots, \tag{3.1}$$

and for the temporal scales coordinates

$$T_p = \epsilon^p \tau n, \quad p = 0, 1, \ldots. \tag{3.2}$$

We define the partial shift operators E_{X_p} and E_{T_p} as follows:

$$E_{X_p} f(\ldots, X_p, \ldots) := f(\ldots, X_p + \epsilon^p h, \ldots), \tag{3.3}$$

in analogy with the definition of the spatial shift operator (2.5) and similarly

$$E_{T_p} f(\ldots, T_p, \ldots) := f(\ldots, T_p + \epsilon^p \tau, \ldots), \tag{3.4}$$

in analogy with the temporal shift operator (2.6).

It follows that for a discrete function $u(X_0, X_1, X_2)$, for example, we would have

$$E_x u = E_{X_0} E_{X_1} E_{X_2} u. \tag{3.5}$$

In the previous chapter we used the spatial and temporal shift operators to define divided difference operators, equations (2.9) to (2.14). Using the partial shift operators (3.3) and (3.4) above we also define partial divided difference operators:

$$\Delta_{X_p} := (E_{X_p} - 1)/(\epsilon^p h), \tag{3.6}$$

$$\nabla_{X_p} := (1 - E_{X_p}^{-1})/(\epsilon^p h), \tag{3.7}$$

in X_p, and for T_p

$$\Delta_{T_p} := (E_{T_p} - 1)/(\epsilon^p \tau), \tag{3.8}$$

$$\nabla_{T_p} := (1 - E_{T_p}^{-1})/(\epsilon^p \tau). \tag{3.9}$$

The analogues of the central difference operators δ_x and δ_t, defined in equations (2.11) and (2.14) respectively, follow similarly:

$$\delta_{X_p} \equiv (\Delta_{X_p} + \nabla_{X_p})/2, \tag{3.10}$$

and

$$\delta_{T_p} \equiv (\Delta_{T_p} + \nabla_{T_p})/2. \tag{3.11}$$

We shall use the operators defined above in our discrete multiple scales analysis.

3.2 Product Rule of Differences

It is also of interest to note the discrete analogue of the product rule of differentiation for the respective operators, which we state in the following theorem.

Theorem 3 *Consider* $V = V(X_1)$ *and* $W = W(X_1)$. *Then*

$$\Delta_{X_1}(VW) = (E_{X_1}W)\Delta_{X_1}V + V\Delta_{X_1}W, \tag{3.12}$$

and

$$\nabla_{X_1}(VW) = W\nabla_{X_1}V + (E_{X_1}^{-1}V)\nabla_{X_1}W. \tag{3.13}$$

Proof: Consider equation (3.12). From the definition of Δ_{X_1} (3.6) we have

$$\begin{aligned} \epsilon h\Delta_{X_1}(VW) &= E_{X_1}(VW) - VW \\ &= E_{X_1}VE_{X_1}W - VE_{X_1}W + VE_{X_1}W - VW \\ &= E_{X_1}W\epsilon h\Delta_{X_1}V + V\epsilon h\Delta_{X_1}W. \end{aligned}$$

The proof of (3.13) follows similarly. □

Theorem 4 *Consider* $V = V(X_1)$ *and* $W = W(X_1)$. *Then*

$$\Delta_{X_1}(VW) = V\Delta_{X_1}W + W\Delta_{X_1}V + \epsilon h\Delta_{X_1}V\Delta_{X_1}W, \tag{3.14}$$

and

$$\nabla_{X_1}(VW) = V\nabla_{X_1}W + W\nabla_{X_1}V - \epsilon h\nabla_{X_1}V\nabla_{X_1}W. \tag{3.15}$$

Proof: Using Theorem 3 we write

$$\begin{aligned} &\Delta_{X_1}(VW) - V\Delta_{X_1}W - W\Delta_{X_1}V \\ =\ & E_{X_1}V\Delta_{X_1}W + W\Delta_{X_1}V - V\Delta_{X_1}W - W\Delta_{X_1}V \\ =\ & \epsilon h\Delta_{X_1}V\Delta_{X_1}W. \end{aligned}$$

The proof of (3.15) follows similarly. □

Using the results of Theorem 4 we can find a corresponding product rule for δ_{X_1}.

Theorem 5 *Consider $V = V(X_1)$ and $W = W(X_1)$. Then*

$$\delta_{X_1}(VW) = V\delta_{X_1}W + W\delta_{X_1}V + \epsilon h(\Delta_{X_1}V\delta_{X_1}W - \nabla_{X_1}W\delta_{X_1}V). \tag{3.16}$$

Proof: From Theorem 4 we have

$$\begin{aligned} 2\delta_{X_1} &= (V\Delta_{X_1}W + W\Delta_{X_1}V + \epsilon h\Delta_{X_1}V\Delta_{X_1}W) \\ &\quad +(V\nabla_{X_1}W + W\nabla_{X_1}V - \epsilon h\nabla_{X_1}V\nabla_{X_1}W) \\ &= 2V\delta_{X_1}W + 2W\delta_{X_1}V + \epsilon h(\Delta_{X_1}V\Delta_{X_1}W - \nabla_{X_1}V\nabla_{X_1}W). \end{aligned}$$

We could have ended the result here, but the required result follows by noting, from the definition of δ_x, that $\Delta_{X_1}W = 2\delta_{X_1}W - \nabla_{X_1}W$. □

We could also prove Theorem 5 by noting the following theorem:

Theorem 6 *[111] Consider $V = V(X_1)$ and $W = W(X_1)$. Then*

$$\delta_{X_1}(VW) = (E_{X_1}V)\delta_{X_1}W + (E_{X_1}^{-1}W)\delta_{X_1}V. \tag{3.17}$$

Proof: Using the definition of δ_{X_1} in equation (3.10) we expand $\delta_{X_1}(VW)$

$$\begin{aligned} 2\epsilon h\delta_{X_1}(VW) &= E_{X_1}VE_{X_1}W - E_{X_1}^{-1}VE_{X_1}^{-1}W \\ &= E_{X_1}V(E_{X_1}W - E_{X_1}^{-1}W) + E_{X_1}^{-1}W(E_{X_1}V - E_{X_1}^{-1}V) \\ &= 2\epsilon h[(E_{X_1}V)\delta_{X_1}W + (E_{X_1}^{-1}W)\delta_{X_1}V], \end{aligned}$$

which again follows from repeated application of the definition. □

Using the result of Theorem 6 we could prove Theorem 5. We consider the sum

$$\delta_{X_1}(VW) - V\delta_{X_1}W - W\delta_{X_1}V,$$

which, using (3.12), can be written as

$$\begin{aligned} &[(E_{X_1} - 1)V]\delta_{X_1}W + [(E_{X_1}^{-1} - 1)W]\delta_{X_1}V \\ &\qquad = \epsilon h(\Delta_{X_1}V\delta_{X_1}W - \nabla_{X_1}W\delta_{X_1}V), \end{aligned}$$

by making use of the definitions (3.6) and (3.7) for Δ_{X_1} and ∇_{X_1} respectively.

3.3 Discrete Chain Rule Expansion

By making use of the definitions in equations (3.1) through (3.11) we are now in a position to state the following theorem proved by Schoombie [111], which forms the keystone to the discrete multiple scales analysis:

Theorem 7 *Schoombie [111]:*

Let for any function

$$f = f(X_0, X_1, \ldots, X_n, T_0, T_1, \ldots, T_m), \tag{3.18}$$

where

$$X_p = \epsilon^p hj, \quad p = 0, 1, \ldots, n,$$

and

$$T_p = \epsilon^p \tau n, \quad p = 0, 1, \ldots, m,$$

and furthermore E_{X_p} and E_{T_p} defined as in (3.3) and (3.4). Then the divided difference operators defined by (2.9) through (2.13) satisfy the relations:

$$\Delta_x = \Delta_{X_0} + \sum_{p=1}^{n} \epsilon^p \Delta_{X_p} E_{X_0} E_{X_1} \ldots E_{X_{p-1}}, \tag{3.19}$$

$$\nabla_x = \nabla_{X_0} + \sum_{p=1}^{n} \epsilon^p \nabla_{X_p} E_{X_0}^{-1} E_{X_1}^{-1} \ldots E_{X_{p-1}}^{-1}. \tag{3.20}$$

Also

$$\Delta_t = \Delta_{T_0} + \sum_{p=1}^{m} \epsilon^p \Delta_{T_p} E_{T_0} E_{T_1} \ldots E_{T_{p-1}}, \tag{3.21}$$

$$\nabla_t = \nabla_{T_0} + \sum_{p=1}^{m} \epsilon^p \nabla_{T_p} E_{T_0}^{-1} E_{T_1}^{-1} \ldots E_{T_{p-1}}^{-1}. \tag{3.22}$$

Proof: We follow the elegant proof of Schoombie [111]. We shall only prove (3.20); the rest of the relations are proved similarly. We note, as with the example given in equation (3.5), that

$$E_x^{-1} = E_{X_0}^{-1} E_{X_1}^{-1} \ldots E_{X_n}^{-1}. \tag{3.23}$$

Then

$$\begin{aligned}
&\nabla_{X_0} + \sum_{p=1}^{n} \epsilon^p \nabla_{X_p} E_{X_0}^{-1} E_{X_1}^{-1} \dots E_{X_{p-1}}^{-1} \\
&= \frac{1}{h}[\sum_{p=1}^{n} (E_{X_0}^{-1} E_{X_1}^{-1} \dots E_{X_{p-1}}^{-1} - E_{X_0}^{-1} E_{X_1}^{-1} \dots E_{X_p}^{-1}) + 1 - E_{X_0}^{-1}] \\
&= \frac{1}{h}[1 - E_{X_0}^{-1} E_{X_1}^{-1} \dots E_{X_n}^{-1}] \\
&= \frac{1}{h}(1 - E_x^{-1}) = \nabla_x.
\end{aligned}$$

□

Theorem 7 provides us with the discrete counterparts of the continuous chain rules as used in (1.57) and (1.58), for example. In the context of our analysis the case $n = 1$ will be of special significance:

Theorem 8 *Subject to the conditions of Theorem 7*

$$\Delta_x = \Delta_{X_0} + \epsilon \Delta_{X_1} E_{X_0}, \tag{3.24}$$

$$\nabla_x = \nabla_{X_0} + \epsilon \nabla_{X_1} E_{X_0}^{-1}. \tag{3.25}$$

Proof: The result is obtained by expanding the results of Theorem 7 for $n = 1$. □

Schoombie [111] showed how (3.24) and (3.25) may be used to perform a multiple scales analysis on the ZK discretization (2.16) of the KdV equation. In the next section we shall provide a generalization of these equations that will be used to analyze higher order finite difference schemes.

3.4 Generalization of Discrete Chain Rule

One should always generalize.

– Carl Jacobi

We shall now proceed to the following theorem which serves as a generalization of Theorem 7:

Theorem 9 *Schoombie and Maré [112]: The difference approximation to any derivative of odd order m of a suitable function can be expanded into the following discrete scales summation*

$$D_{2p}^{m} = D_{2p,0}^{m} + \sum_{r=1}^{p} \epsilon^{r} d_{r}^{m}, \tag{3.26}$$

where

$$D_{2p,0}^{m} = \frac{1}{h^{m}} \sum_{\alpha=1}^{p} \delta_{2p,\alpha}^{m} (E_{X_0}^{\alpha} - E_{X_0}^{-\alpha}), \tag{3.27}$$

and where the d_r^m are given by

$$d_{r}^{m} = h^{r-m} \sum_{\alpha=1}^{p} \begin{pmatrix} \alpha \\ r \end{pmatrix} \delta_{2p,\alpha}^{m} \{\Delta_{X_1}^{r} E_0^{\alpha} - \nabla_{X_1}^{r} E_0^{-\alpha}\}, \tag{3.28}$$

with

$$\begin{pmatrix} n \\ p \end{pmatrix} = \frac{n!}{p!(n-p)!}, \quad \begin{pmatrix} n \\ 0 \end{pmatrix} = 1. \tag{3.29}$$

Proof: Consider the generalized central difference approximation given by (2.50)

$$D_{2p}^{m} = \frac{1}{h^{m}} \sum_{\alpha=1}^{p} \delta_{2p,\alpha}^{m} [E_{x}^{\alpha} - E_{x}^{-\alpha}]. \tag{3.30}$$

By making use of (2.9) and (2.10) we can rewrite (3.30) in the following form

$$D_{2p}^{m} = \frac{1}{h^{m}} \sum_{\alpha=1}^{p} \delta_{2p,\alpha}^{m} [(1 + h\Delta_{x})^{\alpha} - (1 - h\nabla_{x})^{\alpha}]. \tag{3.31}$$

Applying the Binomial theorem [13] to the terms involving powers of α in (3.31) we see that

$$D_{2p}^{m} = \frac{1}{h^{m}} \sum_{\alpha=1}^{p} \delta_{2p,\alpha}^{m} \sum_{j=1}^{\alpha} \begin{pmatrix} \alpha \\ j \end{pmatrix} h^{j} [\Delta_{x}^{j} - (-1)^{j} \nabla_{x}^{j}]. \tag{3.32}$$

Using Theorem 7 with $n = 1$ as shown in (3.24) and (3.25) and using the Binomial theorem we have

$$\begin{aligned} \Delta_{x}^{j} &= (\Delta_{X_0} + \epsilon \Delta_{X_1} E_{X_0})^{j} \\ &= \sum_{r=0}^{j} \begin{pmatrix} j \\ r \end{pmatrix} \epsilon^{r} \Delta_{X_0}^{j-r} \Delta_{X_1}^{r} E_{X_0}^{r}, \end{aligned} \tag{3.33}$$

and similarly

$$\begin{aligned}\nabla_x^j &= (\nabla_{X_0} + \epsilon \nabla_{X_1} E_{X_0}^{-1})^j \\ &= \sum_{r=0}^{j} \binom{j}{r} \epsilon^r \nabla_{X_0}^{j-r} \nabla_{X_1}^r E_{X_0}^{-r}. \end{aligned} \tag{3.34}$$

By combining (3.33) and (3.34) with (3.32) we find that

$$\begin{aligned} &h^m D_{2p}^m = \textstyle\sum_{\alpha=1}^{p} \delta_{2p,\alpha}^m \sum_{j=1}^{\alpha} \binom{\alpha}{j} h^j \times \\ &(\textstyle\sum_{r=0}^{j} \binom{j}{r} \epsilon^r [\Delta_{X_0}^{j-r} \Delta_{X_1}^r E_{X_0}^r - (-1)^j \nabla_{X_0}^{j-r} \nabla_{X_1}^r E_{X_0}^{-r}]). \end{aligned} \tag{3.35}$$

Equation (3.35) can now be rearranged in the form of (3.26), with

$$\begin{aligned} &h^m d_r^m = \textstyle\sum_{\alpha=1}^{p} \delta_{2p,\alpha}^m \sum_{j=r}^{\alpha} \binom{\alpha}{j} h^j \times \\ &(\binom{j}{r} [\Delta_{X_0}^{j-r} \Delta_{X_1}^r E_{X_0}^r - (-1)^j \nabla_{X_0}^{j-r} \nabla_{X_1}^r E_{X_0}^{-r}]). \end{aligned} \tag{3.36}$$

We can simplify equation (3.36) by noting that, since

$$\binom{\alpha}{j}\binom{j}{r} = \binom{\alpha}{r}\binom{\alpha-r}{j-r}, \tag{3.37}$$

we obtain

$$\begin{aligned} &\sum_{j=r}^{\alpha} \binom{\alpha}{j} h^j \binom{j}{r} \Delta_{X_0}^{j-r} E_{X_0}^r \\ &= \binom{\alpha}{r} \sum_{j=r}^{\alpha} \binom{\alpha-r}{j-r} h^j \Delta_{X_0}^{j-r} E_{X_0}^r \\ &= \binom{\alpha}{r} \sum_{j=0}^{\alpha-r} \binom{\alpha-r}{j} h^{j+r} \Delta_{X_0}^j E_{X_0}^r \\ &= \binom{\alpha}{r} h^r (1 + h\Delta_0)^{\alpha-r} E_{X_0}^r \\ &= \binom{\alpha}{r} h^r E_{X_0}^{\alpha}, \end{aligned} \tag{3.38}$$

and likewise

$$\sum_{j=r}^{\alpha} \binom{\alpha}{j} h^j \binom{j}{r} (-1)^j \nabla_{X_0}^{j-r} E_{X_0}^{-r} = \binom{\alpha}{r} h^r (-1)^r E_{X_0}^{-\alpha}. \tag{3.39}$$

Hence equation (3.28) follows. □

3.5 Discrete Multiple Scales

By making use of the results obtained in the previous sections we are now in a position to proceed with the discrete multiple scales analysis technique in direct analogy with the continuous scales analysis as detailed in Section 1.3.3.

We start with an expansion in the same type of form as in the continuous case (1.76). In the discrete case, however, we have to take account of the effect of aliasing, i.e., only a finite number of modes can be resolved on a grid of discrete points.

Therefore, we obtain a discrete scales analysis by solving for an approximate solution $u_j(t)$ in the following form

$$u_j(t) = \sum_{r=-[l/2]}^{[l/2]} c_r u_r(X_1, t, \epsilon) e^{ir\theta}, \tag{3.40}$$

where θ is defined as

$$\theta = khj - \Omega t, \tag{3.41}$$

with k now restricted to the following finite set of values due to aliasing, namely

$$k = 2\pi m/L, \; m = -N/2 + 1, \ldots, N/2. \tag{3.42}$$

Following the details described in Schoombie [111], l is obtained by using the integers s and l with least absolute magnitude such that

$$\frac{m}{N} = \frac{s}{l}, \tag{3.43}$$

and

$$[l/2] = \begin{cases} l/2 & \text{if } l \text{ even} \\ (l-1)/2 & \text{if } l \text{ odd.} \end{cases} \tag{3.44}$$

We also use

$$c_r = \begin{cases} 1 & \text{if } |r| < l/2 \\ 1/2 & \text{if } |r| = l/2. \end{cases} \tag{3.45}$$

Then (in direct analogy with the continuous case) we use

$$u_r = \epsilon^{\delta_r} v_r(X_1, T_1, T_2, \epsilon), \tag{3.46}$$

with

$$\delta_0 = 2, \delta_r = |r|, |r| = 1, ..., [l/2], \tag{3.47}$$

and

$$v_0 = V_0(X_1, T_1, T_2), \tag{3.48}$$

$$v_1 = V_1(X_1, T_1, T_2). \tag{3.49}$$

When $r = 2, \ldots, [l/2]$ we have

$$v_r = V_r(X_1, T_1, T_2, \epsilon) + \sum_{s=r}^{\infty} \epsilon^{s+1-r} W_{rs}(X_1, T_1, T_2). \tag{3.50}$$

3.5.1 Second Order Analysis

We now wish to substitute the expansion (3.40) into the finite difference scheme to be analyzed, equation (2.16), for example. In this process we shall need to evaluate difference operators applied to (3.40). We apply the results of Theorems 7 and 9 to (3.40).

Consider firstly the central difference operator δ_x defined in (2.11) as a second order accurate approximation to d/dx. Using Theorem 8 we can write up to terms $O(\epsilon^2)$:

Theorem 10 *Consider $u_r e^{ir\theta}$ as defined in equations (3.40) through (3.44). Then*

$$\begin{aligned} &\delta_x(u_r e^{ir\theta}) \\ &= [i \sin(rkh)/h + \epsilon \cos(rkh)\delta_{X_1} \\ &+ i\epsilon^2 h \sin(rkh)\Delta_{X_1}\nabla_{X_1}/2]u_r e^{ir\theta}. \end{aligned} \tag{3.51}$$

Proof: By using Theorem 8 we can expand δ_x as

$$\delta_x = \delta_{X_0} + \frac{\epsilon}{2}(\Delta_{X_1} E_{X_0} + \nabla_{X_1} E_{X_0}^{-1}),$$

by making use of (3.24) and (3.25) respectively.

Considering $\delta_{X_0}(u_r e^{ir\theta})$ we have

$$\begin{aligned} \delta_{X_0}(u_r e^{ir\theta}) &= \frac{(e^{irkh} - e^{-irkh})}{2h} u_r e^{ir\theta} \\ &= i\frac{\sin(rkh)}{h} u_r e^{ir\theta}. \end{aligned}$$

Furthermore, for the terms of $O(\epsilon)$:

$$\begin{aligned}
&(\Delta_{X_1} E_{X_0} + \nabla_{X_1} E_{X_0}^{-1})(u_r e^{ir\theta}) \\
&= (\Delta_{X_1} + \nabla_{X_1}) \cos(rkh) u_r e^{ir\theta} + (\Delta_{X_1} - \nabla_{X_1}) i \sin(rkh) u_r e^{ir\theta} \\
&= [2\delta_{X_1} \cos(rkh) + i\epsilon h \Delta_{X_1} \nabla_{X_1} \sin(rkh)] u_r e^{ir\theta},
\end{aligned}$$

The last step follows by noting that

$$\Delta_{X_1} - \nabla_{X_1} = \epsilon h \Delta_{X_1} \nabla_{X_1},$$

from which the theorem is proved. □

If we consider equation (3.51) we find that in the limit $h \to 0$ we obtain the chain rule derivative expansion (1.83) utilized for the continuous multiple scales analysis, namely

$$\partial_x = ir k + \epsilon \partial_{X_1}.$$

If we proceed similarly and use the operators $\delta_x \Delta_x \nabla_x$ defined in (2.9) through (2.11) as an approximation to d^3/dx^3 as shown in (3.40), for example, we can again write up to terms $O(\epsilon^2)$:

Theorem 11 *Consider $u_r e^{ir\theta}$ as defined in equations (3.40) through (3.44). Then*

$$\begin{aligned}
&\delta_x \Delta_x \nabla_x (u_r e^{ir\theta}) \\
&= [(2i/h^3) \sin(rkh)(\cos(rkh) - 1) \\
&+ (2\epsilon/h^2)(\cos(rkh) - 1)(2\cos(rkh)\delta_{X_1} \\
&+ (i\epsilon^2/h) \sin(rkh)(4\cos(rkh) - 1)\Delta_{X_1} \nabla_{X_1}] u_r e^{ir\theta}.
\end{aligned} \tag{3.52}$$

The proof of the theorem follows along similar lines as the proof of Theorem 10 and will not be shown here.

As before, in the limit $h \to 0$, we obtain from (3.52) that

$$\partial_x^3 = -ir^3 k^3 - 3\epsilon r^2 k^2 \partial_{X_1} + 3i\epsilon^2 r k \partial_{X_1}^2,$$

which is to $O(\epsilon^2)$ the same as expanding $[irk + \epsilon \partial_{x_1}]^3$, the continuous multiple scales third order derivative expansion.

Therefore, as a result of Theorem 7 applied to the expression (3.40), we can find discrete chain rule expansions for the derivative approximation operators (2.9) to (2.11) and combinations thereof, as shown above for specific examples. We shall generalize this result in the next section to facilitate analysis of generalized central finite difference schemes.

3.5.2 General Analysis

In direct analogy with Section 3.5.1 we wish to substitute the expansion (3.40) into the finite difference scheme to be analyzed. In this process the following generalized result is of fundamental importance.

Theorem 12 *We can write up to terms* $O(\epsilon^2)$

$$D_{2p}^m(u_r e^{ir\theta}) = Q_{r,p}^m(u_r e^{ir\theta}), \tag{3.53}$$

where

$$\begin{aligned} Q_{r,p}^m/2 &= ih^{-m}\sum_{\alpha=1}^{p}\delta_{2p,\alpha}^m \sin(rkh\alpha) \\ &+ \epsilon h^{1-m}\sum_{\alpha=1}^{p}\alpha\delta_{2p,\alpha}^m \cos(rkh\alpha)\delta_{X_1} \\ &+ \epsilon^2 ih^{2-m}\sum_{\alpha=1}^{p}\alpha^2\delta_{2p,\alpha}^m \sin(rkh\alpha)\Delta_{X_1}\nabla_{X_1}. \end{aligned} \tag{3.54}$$

Proof: The proof follows by applying (3.26) to (3.28) from Theorem 9 to $(u_r e^{ir\theta})$. The proof relies heavily on the ideas used in Theorem 10 and will not be repeated here. □

The results (3.51) and (3.52) are special examples of (3.54).

The discrete multiple scales technique ensues by substitution of the expansion (3.40) as well as the derived discrete chain rule expansions into the discrete numerical approximation. The technique will be demonstrated in detail in Chapter 4 where generalized central finite difference approximations to the KdV equation are analyzed.

3.6 Discrete Perturbation Techniques – Literature

The analysis methodology presented above is convenient for discrete numerical approximation algorithms. The analysis was derived by Schoombie [111] and applied to the ZK discretization of the KdV. We note, however, following Schoombie [110], that we would be able to obtain similar results with an alternative expansion to (3.40), namely

$$u_j(t) = \sum_{s=1}^{\infty} \epsilon^s u_s(X_0, X_1, t), \tag{3.55}$$

with X_0 and X_1 defined by (3.1). In Chapter 6 we shall apply (3.55) to the perturbation analysis of a discrete van der Pol equation. It is important to note the role of Theorem 7 in this analysis, effectively yielding a discrete chain rule for derivatives.

Newell [96] used a more heuristic perturbation analysis to develop a theory to account for the effect of finite amplitude perturbations on the stability of partial difference equations. The technique does not, however, give a clear picture of the relationships between partial differences with respect to the various time and space scales. Stuart [120] discussed Newell's approach for a nonlinear stability analysis of dissipative schemes. Similarly, Cloot and Herbst [29] discussed stability of difference schemes and used a multiple scales analysis to demonstrate instabilities caused by a resonance effect introduced by discretization of the inviscid Burgers equation.

Huston [66] provided an extension of the now classical Krylov-Bogoliubov method (see [83, 93]) to nonlinear difference equations. We quote from [66]: "Unfortunately the extension of the Krylov-Bogoliubov method to difference equations is not straightforward. Indeed, some modifications must be made in the method. These include changes in the Fourier series expansions, the averaging processes, and the solutions of the resulting first order equations." Mickens [84] demonstrated that the approximate solutions arising from the averaged discrete equations are not uniformly valid approximations.

Mickens [84, 85] demonstrated the use of a so-called multi-discrete variable procedure. The approach ensues by defining two discrete-time variables k and $s = \epsilon k$. We assume a solution of the form

$$\begin{aligned} y_k &\equiv y(k, s, \epsilon) \\ &= y_0(k, s) + \epsilon y_1(k, s) + O(\epsilon^2), \end{aligned} \tag{3.56}$$

where y_k is assumed to have at least a first partial derivative with respect to s. On the basis of these assumptions, we have

$$\begin{aligned} y_{k+1} &= y(k+1, s+\epsilon, \epsilon) \\ &= y_0(k+1, s+\epsilon) + \epsilon y_1(k+1, s+\epsilon) + O(\epsilon^2), \end{aligned} \tag{3.57}$$

and

$$y(k+1, s+\epsilon) = y_0(k+1, s) + \epsilon \partial_s y_0(k+1, s) + O(\epsilon^2), \tag{3.58}$$

as well as

$$y_1(k+1, s+\epsilon) = y_1(k+1, s) + O(\epsilon). \tag{3.59}$$

Furthermore

$$y_{k+1} = y_0(k+1, s) + \epsilon[y_1(k+1, s) + \partial_s y_0(k+1, s)] + O(\epsilon^2), \tag{3.60}$$

and

$$y_{k-1} = y_0(k-1, s) + \epsilon[y_1(k-1, s) - \partial_s y_0(k-1, s)] + O(\epsilon^2). \tag{3.61}$$

The technique would ensue by substituting (3.56), (3.60) and (3.61) into the difference equation to be considered. This forms a reasonable approach with practical merit.

Chapter 4

KdV Discrete

Truth ... and if mine eyes
Can bear its blaze, and trace its symmetries,
Measure its distance, and its advent wait,
I am no prophet – I but calculate.

– Charles Mackay (1814 - 1889)

In this chapter we shall consider numerical approximation techniques to the solution of the KdV equation. We shall specifically consider a generalized central difference approximation technique to the KdV and we shall show, using the discrete multiple scales analysis technique discussed in Chapter 3, conditions for the existence of spurious numerical solutions.

4.1 Introduction

When analyzing the behavior of a numerical approximation technique to a given partial differential equation, numerical analysts usually ask the following type of questions. How accurate is the approximation [119], and is there consistency with the original problem? How stable is the approximation for a given choice of initial conditions? How efficient is the approximation technique? Fundamentally, does the approximation technique actually converge to a unique solution as the discretization parameters become arbitrarily small [87]?

In recent years researchers and practitioners have also paid more attention as regards

consistency of the approximation with the original problem formulation. As an example many physical models require conservation of energy and momentum - these would typically be *enforced* in the numerical approximation [114, 134]. The KdV equation (1.46), for example, is known to possess an infinite number of conservation laws [91] as referred to in Chapter 1.

Numerical analysts are used to the fact that the solution of difference equations usually differ quantitatively from the solution of the partial differential equation which it approximates. These solutions would usually be acceptable if the numerical scheme can be shown to be convergent, in which case the solution can be obtained to desired accuracy by a suitably small choice of discretization parameter(s). However, when qualitative differences between the solution of the difference scheme and that of the partial differential equation exist, the solution may be unacceptable (even if conservation laws have been satisfied). For our purpose differences in qualitative behavior will be referred to as *spurious behavior.*

In this chapter we describe spurious numerical solutions that arise specifically as a consequence of finite difference approximations to the KdV equation. One specific example is the well-known Zabusky-Kruskal (ZK) [144] discretization illustrated in operator form in equation (2.16). Using the discrete multiple scales technique illustrated earlier, Schoombie [110] analyzed the ZK discretization to show the occurrence of spurious solutions. We shall consider more general central difference type approximations to the KdV. We shall apply the generalized multiple scales analysis explained previously as an analysis tool to identify spurious modes.

In Section 4.2 we consider the finite difference treatment of the nonlinear term in the KdV equation. This is followed by the generalized finite difference approximation to the KdV equation with associated numerical experiments compared against analytical results. In Section 4.5 we illustrate the discrete multiple scales methodology for the generalized central finite difference approximations. This analysis is followed by a theorem which proves consistency with the continuous case. In the last sections we provide numerical experiments illustrating the spurious behavior of solutions of the KdV equation as predicted by the analysis.

4.2 Nonlinear Term

We consider the inviscid Burgers equation

$$\partial_t u + u\partial_x u = 0, \tag{4.1}$$

subject to the initial condition

$$u(x,0) = f(x). \tag{4.2}$$

Equation (4.1) has for many years played a major role in the study of discretizations of nonlinear conservation laws. It has been used amongst others as a model of the type of nonlinearity occuring in the Navier-Stokes equations of viscous incompressible flows.

Equation (4.1) does not necessarily have continuous solutions for all time. It has an analytical solution which develops a mathematical shock for certain initial conditions. Following [104] and [146], for example, its analytical solution is given by

$$u(x,t) = f(Z),\ Z = x - u(x,t)t. \tag{4.3}$$

From this solution we obtain by differentiation

$$\partial_x u = \frac{f'(Z)}{1 + tf'(Z)}. \tag{4.4}$$

Therefore, for an initial condition such that $1 + tf'(Z) \to 0$ for some $t \to t_B$, we would have the shock

$$\partial_x u(t_B) \to -\infty. \tag{4.5}$$

Such PDE solutions which are not differentiable are termed *weak solutions* as opposed to *genuine solutions*. The shock joins two regions of genuine solutions.

As an example, using the cosine initial condition

$$u(x,0) = \cos(\pi x), \tag{4.6}$$

we obtain from (4.3) the implicit solution

$$u(x,t) = \cos(\pi(x - ut)). \tag{4.7}$$

As is well-known [68] this solution tends to steepen its shape and becomes multivalued at $x = 1/2$ when $t_B = 1/\pi$, the so-called breakdown time. Historically speaking, Zabusky and Kruskal [144] used this initial condition in their numerical study of the KdV equation, constituting the first example of an initial value problem for the KdV equation solved numerically.

It is also known that solution of (4.1) can be continued beyond the time of the breakdown as solutions in the integral sense of the conservation law

$$\partial_x u + \partial_x(\frac{1}{2}u^2) = 0. \tag{4.8}$$

These solutions in the integral sense contain discontinuities, the mathematical representation of shock waves; they are uniquely determined by their initial data provided that the discontinuities are constrained to satisfy an entropy condition.

It is known that solutions of dispersive approximations to equation (4.1) behave quite differently. As an example of a dispersive approximation consider the following:

$$\partial_t u_k + u_k \delta_x u_k = 0. \tag{4.9}$$

To see the dispersive nature of the approximation we use Taylor's expansion, namely

$$\delta_x u_k = \partial_x u + \frac{h^2}{6}\partial_x^3 u + \dots$$

which, upon substitution in the approximation (4.9) above yields

$$\partial_t u + u\partial_x u + \frac{h^2}{6} u \partial_x^3 u \approx 0,$$

an equation very much like the KdV equation (1.46)

$$\partial_t u + u\partial_x u + \epsilon^2 \partial_x^3 u = 0, \tag{4.10}$$

as long as u does not change sign.

Specifically it is known [56] that solutions $u(x,t;\epsilon)$ of the KdV equation stated in (4.10) behave as follows as $\epsilon \to 0$. As long as the solution of equation (4.1) subject to initial value (4.2) has a smooth solution, $u(x,t;\epsilon)$ tends uniformly to that solution. However, when t exceeds the time when the solution of (4.1) breaks down, $u(x,t;\epsilon)$ behaves in an oscillatory manner over some x-interval; as ϵ tends to zero, the amplitude of the oscillations remain bounded but does not tend to zero and the wave length is of order ϵ. Goodman *et al.* [56] showed the validity of the analogy between equations (4.10) and (4.9), namely that solutions of (4.9) behave analogously in their dependence on the discretization parameter h as solutions of (4.10) do in their dependence on ϵ.

We consider equation (4.1) as a model nonlinear equation, specifically owing to the special numerical treatment that it should be allowed. In the rest of this section we shall restrict ourselves to the problem of the nonlinear instability of some finite difference discretizations to (4.1).

The description *nonlinear instability* refers to situations where linear stability theory involving perturbation around a spatially fixed point may be inconclusive [58]. Typically, numerical solutions may remain bounded over long time intervals before showing a sudden catastrophic growth over a relatively short time interval [129].

Our primary objective is to seek an adequate numerical treatment of (4.1), in particular when we consider finite difference approximations. The numerical treatment of (4.1) is by no means new in the literature. In this regard we note important work by [6, 22, 28, 42, 46, 56, 58, 79, 87, 96, 106, 129, 133] and [137]. We make specific use of the seminal research by Fornberg [42] and Briggs *et al.* [22].

Following Fornberg [42] we rewrite (4.1) as

$$\partial_t u + \frac{\theta}{2}\partial_x u^2 + (1-\theta)u\partial_x u = 0, \tag{4.11}$$

where θ is a parameter of arbitrary value to provide for general discretizations of the nonlinear term. The choice $\Theta = 1$, for example, would give rise to the product approximation method as described by Christie *et al.* [27].

A second order central finite difference spatial approximation of (4.11) with a Leapfrog discretization in time would be [6, 42]:

$$\delta_t u_j^n + \frac{\theta}{2}\delta_x[u_j^n]^2 + (1-\theta)u_j^n\delta_x u_j^n = 0. \tag{4.12}$$

To draw the analogy with the ZK [144] finite difference approximation of the KdV equation as shown in equation (2.1) we rewrite (4.12) without the operator formalism thus:

$$u_j^{n+1} = u_j^{n-1} - \tau/(2h)[\theta u_{j+1}^n + 2(1-\theta)u_j^n + \theta u_{j-1}^n](u_{j+1}^n - u_{j-1}^n). \tag{4.13}$$

For comparative purposes when $\theta = 2/3$ we can write equation (4.13) as

$$u_j^{n+1} = u_j^{n-1} - \tau/(3h)[u_{j+1}^n + u_j^n + u_{j-1}^n](u_{j+1}^n - u_{j-1}^n), \tag{4.14}$$

similar to the treatment of the nonlinear term used by Zabusky and Kruskal [144] as shown in equation (2.1).

Using the results of Briggs *et al.* [22] equation (4.13) has a one mode solution of the form

$$u_j^n = a(n)e^{i(2\pi/3)j} + a^*(n)e^{-i(2\pi/3)j}, \tag{4.15}$$

where $a(n)$ is given by the following difference relation

$$a(n+1) = a(n-1) + i\frac{\sqrt{3}\tau}{2h}(2-3\theta)a^*(n)^2. \tag{4.16}$$

Following [22] we note the result of Fornberg [42] namely, that in the continuous limit ia behaves in time like $1/(t_0 - t)$ for $\theta \neq 2/3$. We thus have a solution for u_j^n in (4.12) which is nonlinearly unstable whenever $\theta \neq 2/3$.

Fornberg [42] showed that the condition $\theta = 2/3$ is a necessary condition for stability. Stability here means that the L_2 norm of the difference approximation does not increase faster in time than a fixed exponential function even if the mesh is refined. Fornberg, [46] notes that with $\theta = 2/3$ for all centered FD schemes in space and any spatial period,

$$\frac{d}{dt}\sum_i u_i^2 = 0; \tag{4.17}$$

that is, spatially discrete solutions cannot grow in the L^2 norm. We can also show [22], that

$$\frac{d}{dt}\sum_i u_i = 0. \tag{4.18}$$

The above equations correspond to conservation of energy and momentum respectively. To provide for the temporal discretization one would use a variety of ODE solvers. Shampine [114], for example, discussed techniques for the temporal discretization of systems with conservation laws.

It is important to note that the condition $\theta = 2/3$ is only a necessary condition for stability but not sufficient. As mentioned above, Fornberg [42] showed that $\theta = 2/3$ is a necessary condition for stability of Leapfrog discretizations and necessary and sufficient for stability of the Crank-Nicolson discretization. It is well known in the literature ([6, 7, 28, 42, 106, 122, 133, 134], for example) that Leapfrog discretizations of PDEs may have unbounded solutions for any choice of mesh size (even for choices satisfying the conditions for linear stability).

The result above shows that one needs to exercise caution in the discretization of nonlinear terms by finite difference methods. The type of instability encountered here is only a function of the spatial discretization of (4.1) and cannot be removed by a suitable choice of temporal discretization [137]. Using the same reference we note, interestingly, that should we choose a Galerkin method to provide for the spatial discretization of (4.1), it would not be necessary to apply the $\theta = 2/3$ rule [87].

By making use of the above discussion we shall use the following central finite difference spatial approximation to (4.1) with a method of lines (MOL) for the temporal discretization, namely

$$\partial_t u_j + [\Theta D_{2p}^1[(u_j)^2/2] + (1-\Theta)u_j D_{2p}^1 u_j] = 0, \tag{4.19}$$

where D_p is given by (2.50). For the purposes of preventing numerical nonlinear instability we shall generally make use of the $\theta = 2/3$ rule but we consider the general case for the purpose of analysis.

There are several methods available in the literature for time stepping of finite difference methods. The MOL utilized in (4.19) above generally advances a system that has been discretized in space by application of a "package" ODE solver. The study of ODE solvers is extensive - we mention only the following references [3, 19, 23, 46, 60, 75, 87, 102, 122] and [141].

The benefit of the method is that many ODE solution techniques have been thoroughly analyzed and convenient program packages developed. In its simplest form the MOL could consist of a fixed time step and fixed-order implementation of a Leapfrog of Runge-Kutta method that the user would incorporate in the code. The MOL is particularly useful for the purpose of analysis as we specifically wish to examine spatial difference techniques.

4.3 Numerical Approximations of the KdV

In this section we consider numerical solution techniques for the KdV equation. We shall provide a brief survey of methods in the literature but due to the exhaustive amount of literature available, as in the continuous case, we can only cite specific relevant papers. Omissions are made purely on the basis that this is not a survey article. We discuss some methods specifically for their historical importance, numerical intricacies or practical use and efficiency.

The KdV approximation most frequently referred to is surely the Zabusky and Kruskal finite difference approximation illustrated in (2.1), and in operator form in (2.16). In this approximation Zabusky and Kruskal [144] used a second order central difference approximation in space and a Leapfrog discretization in time. Similar non-dissipative (or conservative [52, 103]) type methods were followed by Greig and Morris [53], Sanz-Serna and Christie [107] and Sanz-Serna [105]. Other less-known approaches include those of Iskandar [67], El-Zoheiry *et al.* [39] and Feng and Mitsui [41].

As an example, the Greig-Morris Hopscotch algorithm [2, 53, 126] proceeds as follows:

$$\Delta_t u_j^n + \eta \delta_x u_j^n + \zeta \delta_x (u_j^n)^2 + \gamma \delta_x \Delta_x \nabla_x u_j^n = 0, \tag{4.20}$$

where $j+n$ is even and

$$\Delta_t u_j^n + \eta \delta_x u_j^{n+1} + \zeta \delta_x (u_j^{n+1})^2 + \gamma \delta_x \Delta_x \nabla_x u_j^{n+1} = 0, \tag{4.21}$$

where $j+n$ is odd. To implement the scheme we would employ equation (4.20) at those grid points for which $j+n$ is even and (4.21) otherwise. The scheme has a

truncation error of order $(O(\tau^2) + O(h^2))$. In section 4.4.3 we shall briefly compare the ZK and Hopscotch methods.

Furihata [51] applied a procedure to design difference schemes that inherit energy conservation properties to the KdV. Zhen *et al.* [147] also proposed a scheme which possesses the first four conserved quantities of the KdV. Huang [65] proposed a Hamiltonian approximation to the KdV to obtain conservation results (see also, van Beckum and van Groesen [134]). De Frutos and Sanz-Serna [35] considered numerical integration techniques that preserve invariants with specific attention to the KdV.

Vliegenthart [136] examined the ZK discretization and suggested some dissipative finite difference schemes. One such approximation is

$$\begin{aligned}\Delta_t u_j^n &= \delta_x u_j^n + \frac{\zeta}{2}\delta_x (u_j^n)^2 - \frac{\gamma}{2}\Delta_x \nabla_x \delta_x u_j^n \\ &+ \frac{\tau}{2}\Delta_x \nabla u_j^n + \frac{\tau\zeta}{2}[\Delta_x \nabla_x (u_j^n)^2] \\ &+ \gamma\tau(1+\zeta u_j^n)(\Delta_x^2 \nabla_x^2) u_j^n + \frac{\gamma^2}{2}\Delta_x^3 \nabla_x^3 u_j^n,\end{aligned}$$

a second order accurate scheme which in our belief never gained popularity due to its lack of simplicity as opposed to the ZK approximation.

Other dissipative methods include a variety of Petrov-Galerkin methods, e.g., Schoombie [109].

Pen-Yu and Sanz-Serna [73] provided results concerning stability and convergence of a family of methods which includes as particular cases some of the schemes mentioned here.

In a series of articles Taha and Ablowitz [124, 125, 126] provided comparisons of some different finite difference approximations, notably the ZK discretization [144], the Hopscotch discretization [53] and some schemes based on the Inverse Scattering Transform. They also compared results to the Pseudospectral method considered by Fornberg and Whitham [43] which reads as follows:

$$\delta_t u_j^n + \zeta i u_j F^{-1}\{\nu F u_j\} - \gamma i F^{-1}\{\nu^3 F u_j\} = 0, \tag{4.22}$$

where Fu and F^{-1} denotes the discrete Fourier expansion and its inverse as defined in equations (2.56) and (2.57) respectively.

In particular Fornberg and Whitham considered a modification to the last term in (4.22); the approximation then reads:

$$\delta_t u_j^n + \zeta i u_j F^{-1}\{\nu F u_j\} - i F^{-1}\{\sin(\gamma\nu^3\tau)/\tau F u_j\} = 0. \tag{4.23}$$

The difference between (4.22) and (4.23) is that the approximation to the linear KdV equation is not subject to differencing errors in equation (4.23) with a subsequent relaxation of linear time-step stability requirements as well. Fundamentally, consistent approximations to DEs are accurate for low wavenumbers but, in general, lose accuracy for increasing wavenumbers, specifically applicable to equation (4.22). Chan and Kerkhoven [26] considered a similar approach by modifying basis functions to solve the linear dispersive part of the KdV equation. Nouri and Sloan [97] compared a variety of pseudospectral methods for the KdV equation. Cox and Mortell [31] applied the Fornberg-Whitham pseudospectral method to a forced KdV equation with damping.

More recently Fornberg and Driscoll [48] presented an easily implementable time stepping strategy for spatially discretized spectral numerical solutions. They combine AB and AM methods for the time stepping of the nonlinear and stiff linear parts respectively, for the KdV and NLS equations, adding the novel feature that different methods are used in different wavenumber ranges with a result that combines high temporal accuracy with good stability properties. The method compares very favorably with the linearly exact method described above.

It is also interesting to note the application of a class of methods involving adaptive spatial grid refinements to the solution of the KdV/NLS equations as described by Fraga and Morris [50]. These methods are specifically designed to cater for the evolutionary nature of the KdV equation, for example, where moving waves are frequently exhibited.

4.4 Central Difference Approximations

In this work we shall specifically consider a generalized central difference approximation for the spatial discretization of the KdV equation. Using (2.50) we write a generalized central finite difference spatial approximation for the approximation of the initial value problem of the KdV equation (1.46) to (1.48):

$$\partial_t u_j + \eta D^1_{2p} u_j + \zeta[\Theta D^1_{2p}[(u_j)^2/2] + (1-\Theta) u_j D^1_{2p} u_j] + \gamma D^3_{2q} u_j = 0. \qquad (4.24)$$

In direct analogy with equation (4.19), the approximation does not take account of temporal discretization, i.e., we use a MOL approach for the time variable. For the purposes of numerical experiments we shall frequently use the Leapfrog discretization for temporal integration.

An illustration of approximation (4.24) is given by the exemplary case:

$$\begin{aligned} \delta^1_{2,1} &= \tfrac{1}{2} \\ \delta^3_{2,1} &= -1 \\ \delta^3_{2,2} &= \tfrac{1}{2} \\ \Theta &= \tfrac{2}{3}. \end{aligned} \tag{4.25}$$

These values are contained in the first rows of Tables 2.1 and 2.3 respectively. Substituting the parameters for the exemplary case into the generalized finite difference approximation (4.24) we obtain the spatial approximation for the KdV equation as given by the ZK discretization (2.1).

4.4.1 Stability

In this section we shall consider linear stability constraints for the central difference approximation (4.24) to the KdV equation as described above, with the temporal variable approximated by a Leapfrog discretization. We perform the standard Von Neumann analysis [103, 119].

Consider first the linearized version of the ZK scheme (2.16)

$$\delta_t u^n_j + \eta \delta_x u^n_j + \gamma \delta_x \Delta_x \nabla_x u^n_j = 0, \tag{4.26}$$

where we note the use of the Leapfrog discretization for the temporal variable.

We put [103, 117, 122]

$$u^n_j = r^n \exp(ikhj), \tag{4.27}$$

and substitute into (4.26). We find that

$$r^2 - 1 + 2\tau i a(kh) r = 0, \tag{4.28}$$

where

$$a(kh) = (\eta \sin(kh)/h + (2\gamma/h^3)\sin(kh)(cos(kh) - 1)). \tag{4.29}$$

Solving for r we find that the amplification factor is given by

$$r = -i\tau a(kh) \pm \sqrt{1 - \tau^2 a(kh)^2}. \tag{4.30}$$

Therefore the requirement $|r| \leq 1$ for linear stability leads to

$$|\tau a(kh)| \leq 1. \tag{4.31}$$

When the scheme is applied to the nonlinear KdV equation we require (4.31) to hold for any η such that $u_{min} \leq \eta \leq u_{max}$, where u_{min} and u_{max} are the smallest values of $u(x,t)$ respectively. We assume $u_{min} = 0$. Then the condition [105]

$$\max_{0 \leq kh \leq \pi} \{(\frac{2\tau\gamma}{h^3}) \sin(kh)(1 - \cos(kh))\} \leq 1, \tag{4.32}$$

is necessary if (4.31) is to hold for all relevant values of η. Since the maximum of $\sin(kh)(1 - \cos(kh))$ is $3\sqrt{3}/4$ we can rewrite the condition as

$$\tau \leq \frac{2h^3}{3\sqrt{3}\gamma} \approx 0.3849 \frac{h^3}{\gamma}. \tag{4.33}$$

The condition (4.33) is also sufficient if (4.31) is to hold provided that [105]

$$\tau \leq h/\eta. \tag{4.34}$$

Because the pseudospectral method is the limiting method for finite difference approximations of increasing orders of accuracy it is important to consider stability bounds for the pseudospectral method as discussed in (4.22). We can perform a similar analysis for the pseudospectral method as for the finite difference method illustrated above - the result being that [26, 43],

$$\tau \leq \frac{h^3}{\pi^3\gamma} \approx 0.0323 \frac{h^3}{\gamma}, \tag{4.35}$$

showing that the linear stability restriction for the pseudospectral method (or for higher order finite difference approximations) is more restrictive than for low order approximations. However, the increased accuracy compensates for this restriction as is well-known [25, 46].

In Table 4.1 we present linear stability restrictions for central difference approximations of different orders (including the second order and pseudospectral approximation restrictions (4.33) and (4.35) respectively). The restrictions were also derived using the analysis shown above applied to the generalized central difference approximation (4.24), using the finite difference weights provided in Tables 2.1 and 2.3, with a Leapfrog temporal discretization.

Since we are considering higher order central difference approximations, the limit of which is the pseudospectral method [44], we shall for the purposes of our calculations frequently use the stability restriction given by (4.35).

	Stability Restriction
Order	$\tau/h^3 \leq$
2	$0.3849\ /\gamma$
4	$0.2171\ /\gamma$
6	$0.1621\ /\gamma$
8	$0.1347\ /\gamma$
10	$0.1181\ /\gamma$
$\vdots$	$\vdots$
∞	$0.0323\ /\gamma$

Table 4.1: Linear stability restrictions on τ/h^3 for central difference approximations to the KdV equation with increasing orders of accuracy. Temporal discretization by LF.

In Table 4.2 we present maximum time steps allowable under the linear stability bounds (4.33) and (4.35) for given values of h. We use $\gamma = 0.000484$ for illustrative purposes - a value that we shall frequently use for further calculations corresponding to the analysis performed by Zabusky and Kruskal [144], and others [8, 53, 105, 107].

h	Eq. (4.33)	Eq. (4.35)
0.10	7.95(-1)	6.66(-2)
0.02	6.36(-3)	5.33(-4)
0.01	7.95(-4)	6.66(-5)
0.005	9.94(-5)	8.33(-6)

Table 4.2: Maximum time step for h given. (Notation: $6.36(-3) = 6.36 \times 10^{-3}$.)

4.4.2 Analytical Reference Solution

We consider the following parameters for the continuous KdV equation:

$$\begin{aligned} \zeta &= 1 \\ \gamma &= 0.000484 = 0.022^2 \\ \eta &= 0, \end{aligned} \tag{4.36}$$

as used by [7, 8, 53, 105, 107, 144].

We shall consider the analytical 1-soliton solution of the KdV (refer to Section 1.3). For the parameters (4.36) we use the same 1-soliton solution as [26, 53, 97, 105, 107, 144]. We therefore consider the initial condition

$$u(x,0) = 3c\,sech^2(bx+d), \tag{4.37}$$

which has the theoretical solution

$$u(x,t) = 3c\,sech^2(bx - bct + d), \tag{4.38}$$

where $c = 0.3$,

$$b = \sqrt{\frac{c}{4\gamma}},$$

and $d = -b$, representing a single soliton with amplitude 0.9 and speed 0.3.

We shall use equation (4.38) as an analytical benchmark against which to compare numerical results.

4.4.3 Numerical Tests

As stated previously, we shall use the LF method to advance solutions of the type (4.24) in time. The LF approximation to (4.24) reads:

$$\delta_t u_j^n + \eta D_{2p}^1 u_j^n + \zeta[\Theta D_{2p}^1[(u_j^n)^2/2] + (1-\Theta)u_j^n D_{2p}^1 u_j^n] + \gamma D_{2q}^3 u_j^n = 0. \tag{4.39}$$

To use the LF discretization we need to supplement the initial condition $u_j^0 = f(jh)$ with a starting procedure to compute u_j^1. This is usually performed via a standard Euler starter [105, 126], i.e.,

$$(u_j^1 - u_j^0)/\tau + \eta D_{2p}^1 u_j^0 + \zeta[\Theta D_{2p}^1[(u_j^0)^2/2] + (1-\Theta)u_j^0 D_{2p}^1 u_j^0] + \gamma D_{2q}^3 u_j^0 = 0. \tag{4.40}$$

In Table 4.3 we compare the ZK and Hopscotch approximations for two sets of discretization parameters subject to (4.36). From the results displayed we conclude that the performances of the methods were comparable for $0 \leq t \leq 1$.

Time	ZK h=0.05 $\tau = 0.025$	Hopscotch h=0.05 $\tau = 0.025$	ZK h=0.01 $\tau = 0.0005$	Hopscotch h=0.01 $\tau = 0.0005$
0.25	1.94(-2)	3.27(-2)	2.05(-3)	1.11(-3)
0.50	6.35(-2)	6.74(-2)	4.22(-3)	2.14(-3)
0.75	1.22(-1)	9.93(-2)	6.36(-3)	3.54(-3)
1.00	1.61(-1)	1.42(-1)	8.13(-3)	4.91(-3)

Table 4.3: l_∞ error of one-soliton solutions of the KdV equation with ZK and Hopscotch approximations.

We shall consider three sets of discretization parameters for the solution of (4.24). The first set is given by

$$\begin{aligned} N &= 50 \\ h &= 0.04 \\ \tau &= 0.0005. \end{aligned} \tag{4.41}$$

The second set is given by

$$\begin{aligned} N &= 100 \\ h &= 0.02 \\ \tau &= 0.0005, \end{aligned} \tag{4.42}$$

while the third set is given by

$$\begin{aligned} N &= 200 \\ h &= 0.01 \\ \tau &= 0.00005. \end{aligned} \tag{4.43}$$

We shall use $\theta = 2/3$ throughout our experiments.

In Tables 4.4, 4.5 and 4.6 we show the l_∞ error of solutions of (4.24) for different orders of discretization measured against the analytical 1-soliton solution (4.38) for the KdV equation for $0 \leq t \leq 1$.

	Order				
Time	2	4	6	8	10
0.25	5.21(-2)	1.04(-2)	6.69(-3)	5.07(-3)	3.93(-3)
0.50	8.83(-2)	1.77(-2)	7.06(-3)	3.75(-3)	2.21(-3)
0.75	1.17(-1)	2.29(-2)	6.62(-3)	5.97(-3)	2.45(-3)
1.00	1.43(-1)	2.66(-2)	7.38(-3)	5.87(-3)	3.18(-3)

Table 4.4: l_∞ error of one-soliton solutions of the KdV equation subject to parameters in (4.36) with various order approximations. Discretization parameters (4.41).

	Order				
Time	2	4	6	8	10
0.25	1.25(-2)	1.15(-3)	3.12(-4)	1.02(-4)	3.11(-5)
0.50	1.94(-2)	1.31(-3)	2.48(-4)	8.31(-5)	3.14(-5)
0.75	3.01(-2)	1.19(-3)	3.03(-4)	7.21(-5)	3.80(-5)
1.00	3.62(-2)	1.74(-3)	4.24(-4)	1.27(-4)	5.05(-5)

Table 4.5: l_∞ error of one-soliton solutions of the KdV equation subject to parameters in (4.36) with various order approximations. Discretization parameters (4.42).

	Order				
Time	2	4	6	8	10
0.25	3.18(-3)	8.44(-5)	6.44(-6)	6.75(-7)	1.22(-7)
0.50	5.46(-3)	8.08(-5)	5.96(-6)	5.78(-7)	8.13(-8)
0.75	7.19(-3)	7.93(-5)	6.24(-6)	6.03(-7)	1.14(-7)
1.00	9.52(-3)	1.21(-4)	8.83(-6)	8.81(-7)	1.31(-7)

Table 4.6: l_∞ error of one-soliton solutions of the KdV equation subject to parameters in (4.36) with various order approximations. Discretization parameters (4.43).

In Table 4.7 we show the speed of the numerical soliton solution. As mentioned in Section 4.4.2 the analytical soliton speed is 0.3, a value which is relatively accurately attained by most of the higher order methods. Herman and Knickerbocker [64] report a numerically induced phase shift in the 1-soliton solution of the KdV equation for the ZK approximation. The results shown here seem to indicate that the phase shift

disappears for higher order approximations.

	Soliton Speed		
Order	Eq. (4.41)	Eq. (4.42)	Eq. (4.43)
2	0.2905570	0.2997599	0.3002505
4	0.3029317	0.3004325	0.3000374
6	0.3017023	0.3001414	0.3000029
8	0.3015519	0.3000425	0.3000003
10	0.3006850	0.3000168	0.3000000

Table 4.7: Numerical solution soliton speed subject to (4.36) and the discretization parameters (4.41), (4.42) and (4.43).

It is interesting to consider the stability of the KdV equation. We consider the parameters given by (4.36). We examine a small perturbation $\epsilon(x,t)$ on the solution $u(x,t)$ which leads to the perturbation equation [88]

$$\partial_t \epsilon + u\partial_x \epsilon + \epsilon \partial_x u + \partial_x(\frac{1}{2}\epsilon^2) + \gamma \partial_x^3 \epsilon = 0. \tag{4.44}$$

Multiplying equation (4.44) by ϵ and integrating over the intended interval and applying periodic boundary conditions we obtain

$$\frac{d}{dt}||\epsilon(t)||_{L_2}^2 + \int \epsilon^2(x,t)\partial_x u(x,t)dx = 0.$$

Note that when $\partial_x u < 0$ we would have $\frac{d}{dt}||\epsilon(t)||_{L_2}^2 > 0$, hence $\epsilon(t)$ would be an increasing function. It would therefore appear that a negative gradient in the initial condition, for example, would trigger off an instability in the solution of the KdV equation.

In Figure 4.1 we show the solution of the KdV equation at different times, subject to the initial condition (4.6), namely $u(x,0) = \cos(\pi x)$, for which $\partial_x u < 0$. The solution was obtained using the numerical discretization parameters in (4.42), with a tenth order finite difference approximation. The results are qualitatively similar to that reported in the original Zabusky and Kruskal paper [144]. Note the train of solitons which have developed at time $t = 1.15$ from the shock-type profile at $t = 0.325$.

In the next section we shall apply the discrete multiple scales analysis developed in Chapter 3 to the generalized central difference approximation (4.24) to the KdV equation.

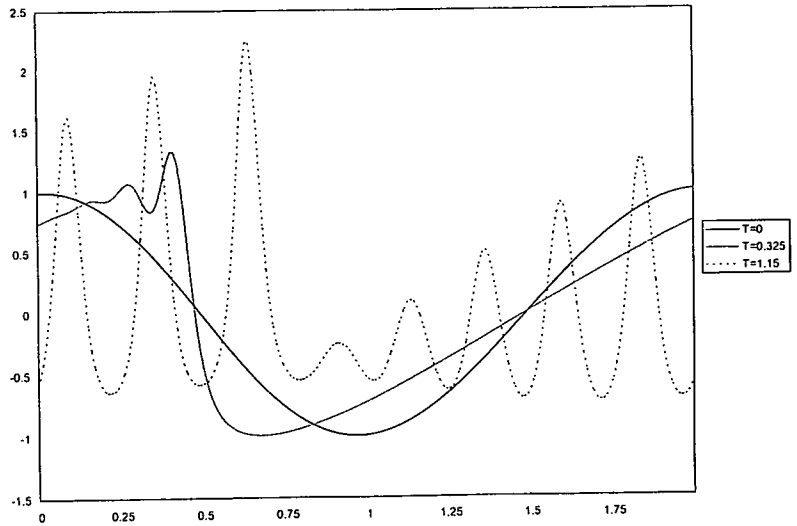

Figure 4.1: Solution of the KdV subject to initial condition (4.6).

4.5 Discrete Multiple Scales Analysis

We shall now proceed with the discrete multiple scales analysis of the generalized finite difference approximation (4.24) to the KdV equation. The analysis follows along similar lines as in the continuous case. We obtain a discrete scales analysis by solving for an approximate solution $u_j(t)$ in the following form [80, 111, 112]

$$u_j(t) = \sum_{r=-\lfloor l/2 \rfloor}^{\lfloor l/2 \rfloor} c_r u_r(X_1, t, \epsilon) e^{ir\theta}, \tag{4.45}$$

subject to the definitions given in (3.40) to (3.50). Substitution of (4.45) into (4.24) and making use of the discrete spatial derivative expansion (2.50) applied to $e^{ir\theta}$ (given in equations (3.53) and (3.54)) as well as the continuous time derivative (1.82) we obtain

$$\begin{aligned} &\epsilon^{\delta_r} \partial_t v_r + \epsilon^{\delta_r} \eta Q^1_{r,p} v_r + \gamma \epsilon^{\delta_r} Q^3_{r,q} v_r \\ &+\zeta \sum_w [(1-\Theta)(Q^1_{r,p} v_r) v_{r-w} + \Theta/2 Q^1_{r,p}(v_r v_{r-w})] \epsilon^{\delta_{r-w}+\delta_w} = 0, \end{aligned} \tag{4.46}$$

after setting coefficients of $c_r e^{ir\theta}$ to zero. In analogy with the continuous case we now study equation (4.46) up to terms $O(\epsilon^3)$. This leads us to consider $r = 0, 1$, and 2.

If we put $r = 0$ in (4.46) the lowest order term in ϵ is found to be $O(\epsilon^3)$, yielding the following equation when equated to zero

$$\partial_{T_1} V_0 + \eta \delta_{X_1} V_0 + \zeta[2(1-\Theta) \sum_{\alpha=1}^{p} \alpha \delta^1_{2p,\alpha} \cos(k\alpha h) + \Theta] \delta_{X_1} |V_1|^2 = 0. \tag{4.47}$$

It is straightforward to show that equation (4.47) becomes the continuous multiple scales analogue (1.85) as $h \to 0$. In deriving (4.47) we make use of Theorem 2 with $m = 1$ and $m = 3$ respectively, i.e.,

$$\sum_{\alpha=1}^{p} \alpha \delta^1_{2p,\alpha} = \frac{1}{2},$$

and

$$\sum_{\alpha=1}^{p} \alpha \delta^3_{2p,\alpha} = 0.$$

If we substitute the Zabusky-Kruskal specific parameters as shown in (4.25) into (4.47) we obtain

$$\partial_{T_1} V_0 + \eta \delta_{X_1} V_0 + (\zeta/3)[\cos(kh) + 2] \delta_{X_1} |V_1|^2 = 0,$$

directly comparable with the result obtained by Schoombie [110, 111].

By putting $r = 1$ in (4.46) and equating coefficients of the $O(\epsilon)$ terms to zero we obtain

$$\frac{\Omega}{2} = \frac{\eta}{h} \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(kh\alpha) + \frac{\gamma}{h^3} \sum_{\alpha=1}^{p} \delta^3_{2p,\alpha} \sin(kh\alpha), \tag{4.48}$$

the discrete linear dispersion relation. (See Sei and Symes [113] for an interesting application of dispersion analysis to numerical wave propagation.)

For the Zabusky-Kruskal specific parameters (4.25) applied to (4.48) we obtain

$$\Omega = \frac{\eta}{h} \sin(kh) + \frac{2\gamma}{h^3} \sin(kh)(\cos(kh) - 1).$$

Subsequently, for $r = 1$, by equating the $O(\epsilon^2)$ terms to zero, the resulting equation is found to be

$$\partial_{T_1} V_1 + V_g \delta_{X_1} V_1 = 0, \tag{4.49}$$

where

$$V_g = \frac{d\Omega}{dk}$$
$$= 2\eta \sum_{\alpha=1}^{p} \alpha \delta^1_{2p,\alpha} \cos(kh\alpha) + \frac{2\gamma}{h^2} \sum_{\alpha=1}^{p} \alpha \delta^3_{2p,\alpha} \cos(kh\alpha), \tag{4.50}$$

is the discrete linear group velocity. Trefethen [128, 129], demonstrated the use of group velocity in the analysis of finite difference methods.

Finally, for $r = 1$ terms, the equation obtained by putting $O(\epsilon^3)$ terms equal to zero is found to be

$$\partial_{T_2} V_1 + \eta i h \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \alpha^2 \sin(kh\alpha) \Delta_{X_1} \nabla_{X_1} V_1$$
$$+ \frac{\gamma i}{h} \sum_{\alpha=1}^{p} \delta^3_{2p,\alpha} \alpha^2 \sin(kh\alpha) \Delta_{X_1} \nabla_{X_1} V_1$$
$$+ \frac{2\zeta i}{h} \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(kh\alpha) V_1 V_0$$
$$+ \frac{2\zeta i}{h} [(1-\Theta) \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} (\sin(2kh\alpha) - \sin(kh\alpha))$$
$$+ \Theta \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(kh\alpha)] V_1^* V_2 = 0. \tag{4.51}$$

Using $r = 2$ we obtain terms $O(\epsilon^2)$

$$-\Omega V_2 i + \frac{\eta i}{h} \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(2\alpha kh) V_2$$
$$+ \frac{\gamma i}{h^3} \sum_{\alpha=1}^{p} \delta^3_{2p,\alpha} \sin(2\alpha kh) V_2$$
$$+ \frac{\zeta i}{h} [\Theta/2 \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(2\alpha kh) V_1^2$$
$$+ (1-\Theta) \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(\alpha kh) V_1^2] = 0, \tag{4.52}$$

which can be rewritten, using Ω derived in (4.48), as

$$V_2 = \zeta \Lambda V_1^2, \tag{4.53}$$

where

$$\Lambda = -\frac{i}{h}\frac{[\Theta/2\sum_{\alpha=1}^{p}\delta_{2p,\alpha}^{1}\sin(2\alpha kh) + (1-\Theta)\sum_{\alpha=1}^{p}\delta_{2p,\alpha}^{1}\sin(\alpha kh)]}{g(h,k,t)}, \tag{4.54}$$

and

$$\begin{aligned} g(h,k,t) &= \frac{\eta}{h}\sum_{\alpha=1}^{p}\delta_{2p,\alpha}^{1}(\sin(2\alpha kh) - 2\sin(\alpha kh)) \\ &+ \frac{\gamma}{h^3}\sum_{\alpha=1}^{p}\delta_{2p,\alpha}^{3}(\sin(2\alpha kh) - 2\sin(\alpha kh)), \end{aligned} \tag{4.55}$$

provided that

$$g(h,k,t) \neq 0. \tag{4.56}$$

Should condition (4.56) be violated we would not be able to remove all the $O(\epsilon^2)$ terms in (4.46) with the result that the expansion (4.45) would only be valid to terms $O(\epsilon)$. Condition (4.56) does not have a continuous counterpart, therefore its violation should correspond to spurious behavior of the numerical scheme (4.24). We shall return to this in Section 4.7.

Note that V_0 is not uniquely defined by (4.47). As in the continuous analysis we make the physically reasonable assumption that V_0 must also satisfy (4.49), i.e.,

$$\partial_{T_1}V_0 + V_g\delta_{X_1}V_0 = 0. \tag{4.57}$$

Hence, after combining equations (4.47) and (4.57) we find that

$$\partial_{X_1}[V_0 + \frac{\zeta[2(1-\Theta)\sum_{\alpha=1}^{p}\alpha\delta_{2p,\alpha}^{1}\cos(k\alpha h) + \Theta]|V_1|^2}{(\eta - V_g)}] = 0, \tag{4.58}$$

provided that

$$\eta - V_g \neq 0. \tag{4.59}$$

Solution of (4.58) leads to the following relation for V_0

$$V_0 = -\frac{\zeta[2(1-\Theta)\sum_{\alpha=1}^{p}\alpha\delta_{2p,\alpha}^{1}\cos(k\alpha h) + \Theta]}{(\eta - V_g)}|V_1|^2. \tag{4.60}$$

If we apply the special case (4.25) to equation (4.60) we find

$$V_0 = \frac{\zeta[\cos(kh) + 2]}{3(V_g - \eta)}|V_1|^2, \tag{4.61}$$

the same equation found by [111], also subject to (4.59).

As with equation (4.56) no continuous analogue exists for violation of condition (4.59). Its violation could similarly lead to *spurious* behavior of the solution of (4.24). In Section 4.7 we show that violation of condition (4.59) does indeed lead to spurious numerical solutions.

It is interesting to investigate the meaning of violation of the condition (4.59). Should $\eta = V_g$ we would have both V_1 and V_0 satisfy the same difference equation, namely

$$\partial_{T_1} V + \eta \delta_{X_1} V = 0, \tag{4.62}$$

with no specific implied relationship between V_1 and V_0. The only conclusion being that

$$\delta_{X_1} |V_1|^2 = 0. \tag{4.63}$$

Furthermore, V_1 and V_0 will both represent a motion along the discrete grid with a velocity η.

Proceeding as in the continuous case, we can construct the discrete equivalent of the NLS equation, i.e., we replace in (4.51) the solutions for V_0 obtained in (4.61) and V_2 obtained in (4.53). The resulting equation is

$$\begin{aligned}
&\partial_{T_2} V_1 + i \sum_{\alpha=1}^{p} (\eta h \delta^1_{2p,\alpha} + \frac{\gamma}{h} \delta^3_{2p,\alpha}) \alpha^2 \sin(kh\alpha) \Delta_{X_1} \nabla_{X_1} V_1 \\
&- \frac{2\zeta^2 i \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(kh\alpha) [2(1-\Theta) \sum_{\alpha=1}^{p} \alpha \delta^1_{2p,\alpha} \cos(k\alpha h) + \Theta]}{h(\eta - V_g)} V_1 |V_1|^2 \\
&+ \frac{2\zeta^2 i}{h} [(1-\Theta) \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} (\sin(2kh\alpha) - \sin(kh\alpha)) \\
&+ \Theta \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(kh\alpha)] \Lambda V_1 |V_1|^2 = 0.
\end{aligned} \tag{4.64}$$

In the next section we shall show that equation (4.64) becomes the NLS equation in the limit as the discretization parameter tends to zero.

In Section 4.7 we shall conduct numerical experiments that will serve to illustrate spurious behavior of the solutions to the finite difference approximations when conditions (4.56) ($g \neq 0$) and (4.59) ($\eta \neq V_g$) are violated.

4.6 Limiting Case

Here we shall show that the discrete multiple scales analysis yields the same results as the continuous case in the limit as the discretization parameter tends to zero. To that end we wish to obtain the limit case of equation (4.64) as $h,\ kh \to 0$.

Theorem 13 *In the limit as $h \to 0$, equation (4.64) becomes the nonlinear Schrödinger equation*

$$\partial_{T_2} V_1 + 3i\gamma k \partial^2_{X_1} V_1 - i\zeta^2/(6\gamma k) V_1 |V_1|^2 = 0. \tag{4.65}$$

Proof: We consider the limit as $h \to 0$ of different parts of equation (4.64).

Firstly, the $\Delta_{X_1}\nabla_{X_1}$ term:

$$\begin{aligned}
&\lim_{h\to 0} i \sum_{\alpha=1}^{p} (\eta h \delta^1_{2p,\alpha} + \frac{\gamma}{h}\delta^3_{2p,\alpha})\alpha^2 \sin(kh\alpha)\Delta_{X_1}\nabla_{X_1} V_1 \\
&= \gamma i k \sum_{\alpha=1}^{p} \alpha^3 \delta^3_{2p,\alpha}\, \partial^2_{X_1} V_1 \\
&= 3i\gamma k\, \partial^2_{X_1} V_1.
\end{aligned} \tag{4.66}$$

The last step follows by making use of Theorem 2 with $m = 3$, namely

$$\sum_{\alpha=1}^{p} \alpha^3 \delta^3_{2p,\alpha} = 3.$$

Secondly, we consider the various $V_1|V_1|^2$ terms. From the discrete linear group velocity V_g, detailed in equation (4.50), follows

$$\lim_{h\to 0} \frac{d\Omega}{dk} = \eta - 3\gamma k^2,$$

by making use of Theorem 2 repeatedly. Therefore, in the limiting case

$$\begin{aligned}
\lim_{h\to 0} \; & - \frac{2\zeta^2 i \sum_{\alpha=1}^{p} \delta^1_{2p,\alpha} \sin(kh\alpha)[2(1-\Theta)\sum_{\alpha=1}^{p} \alpha\delta^1_{2p,\alpha}\cos(k\alpha h) + \Theta]}{h(\eta - V_g)} V_1|V_1|^2 \\
&= \frac{-\zeta^2}{3\gamma k}.
\end{aligned} \tag{4.67}$$

Note the interesting fact that the limit is independent of Θ.

Furthermore, by making use of the identities

$$\lim_{x\to 0} \frac{\sin(2kx) - 2\sin(kx)}{x^3} = -k^3,$$

and

$$\lim_{x\to 0} \frac{2\sin(kx) - \sin(2kx)}{x} = 0,$$

in equation (4.55), we observe that

$$\lim_{h\to 0} \Lambda = \frac{1}{6\gamma k^2}.$$

Therefore, in the limit $h \to 0$,

$$\begin{aligned} &\frac{2\varsigma^2 i}{h}[(1-\Theta)\sum_{\alpha=1}^{p} \delta^1_{2p,\alpha}(\sin(2kh\alpha) - \sin(kh\alpha)) \\ &+\Theta\sum_{\alpha=1}^{p} \delta^1_{2p,\alpha}\sin(kh\alpha)]\Lambda V_1|V_1|^2 \\ &\to \frac{\varsigma^2}{6\gamma k}. \end{aligned} \tag{4.68}$$

By virtue of equation (4.66), and the sum of (4.67) and (4.68) we are led to conclude that in the limit $h \to 0$ equation (4.64) becomes (4.65), the continuous nonlinear Schrödinger equation as previously obtained in the continuous multiple scales analysis, equation (1.92). □

As an interesting by-product of our multiple scales analysis, we can view equation (4.64) as a spatial finite difference approximation for the NLS equation as we have demonstrated consistency with the continuous PDE in the analysis above.

4.7 Numerical Experiments

In this section we shall report the results of several numerical experiments. We are especially interested in the the cases where the conditions (4.56) ($g \neq 0$) and (4.59) ($\eta \neq V_g$) are violated, since in these cases the multiple scales analysis breaks down.

From equation (4.45) and the analysis in Section 4.5 it follows that, since ϵ is a small number, we are essentially looking at solutions of the discretized KdV (4.24) of the form

$$u_j(t) = \epsilon V_1 e^{i\theta} + \epsilon V_1^* e^{-i\theta}, \tag{4.69}$$

subject to the definitions in (3.40) to (3.50) and where V_1 must satisfy equations (4.49) and (4.64).

In the case of carrier wave numbers for which one or both of the conditions (4.56) or (4.59) are violated, the higher order terms in ϵ may become comparable to, or even dominate, the first order terms to such an extent that (4.69) could no longer be considered to be true.

For our numerical experiments we consider the following values of the parameters in (4.24):

$$\begin{aligned} N &= 128 \\ L &= 2 \\ \tau &= 0.0005 \\ \gamma &= 0.000484 \\ \zeta &= 1. \end{aligned} \tag{4.70}$$

Again, we shall use $\theta = 2/3$ throughout our experiments.

We use the initial condition

$$u_j^0 = \epsilon \cos(h\pi j)\cos(m\pi hj), \tag{4.71}$$

where ϵ is a small number, typically $\epsilon = 0.01$. Since we can write (4.71) in the form

$$u_j^0 = (\epsilon/2)\cos(\pi(m+1)hj) + (\epsilon/2)\cos(\pi(m-1)hj), \tag{4.72}$$

there are initially only two modes with wave numbers $(m-1)\pi$ and $(m+1)\pi$ present. Therefore, when we consider the time evolution of Fourier modes of the solution of (4.24), subject to the initial condition (4.71), nonlinear effects where the solution deviates significantly from (4.69) would show up clearly.

We use the LF method as described in equation (4.39) with a Euler starter (4.40) to advance (4.24) in time.

4.7.1 Violation of condition $g \neq 0$ (4.56)

Using the parameters in (4.70) we solve for violation of equation (4.56) for various values of η subject to different mode numbers m and different orders of approximation. We show the results of the calculation in Table 4.8.

	Order of Approximation				
m	2	4	6	8	10
20	−2.8883	−8.9082	−19.3247	−36.0562	−62.2975
25	0.3934	−1.4114	−4.0405	−7.3908	−11.5101
30	3.1115	3.2137	3.0383	2.7372	2.3489
35	4.9578	5.7895	6.3034	6.6640	6.9310
40	5.8383	6.8273	7.4161	7.7987	8.0603
45	5.8748	6.8749	7.4701	7.8553	8.1180
50	5.3563	6.5029	7.2779	7.8381	8.2578
55	4.6532	6.1484	7.2142	8.0026	8.6074
60	4.1144	5.9788	7.2510	8.1866	8.9124

Table 4.8: Values of η which violate (4.56) for various orders of approximation and different mode numbers (m) and subject to the parameters in (4.70).

We first consider spurious solutions of fourth order approximations. To illustrate the behavior of the solution of (4.24) when $g = 0$ we consider a specific value of η, as obtained from Table 4.8, specifically

$$\eta = 5.7895, \tag{4.73}$$

which causes the 35th carrier mode i.e., $k = 35\pi$, for a fourth order approximation, to violate (4.56).

Figures 4.2 through 4.6 show the time evolution of the Fourier modes of the fourth order solution plotted as a three-dimensional graph. For modes $m = 33$ and $m = 37$ only two modes are distinctly present in the solution - any nonlinear effects here are so small that they do not show up in the graphs.

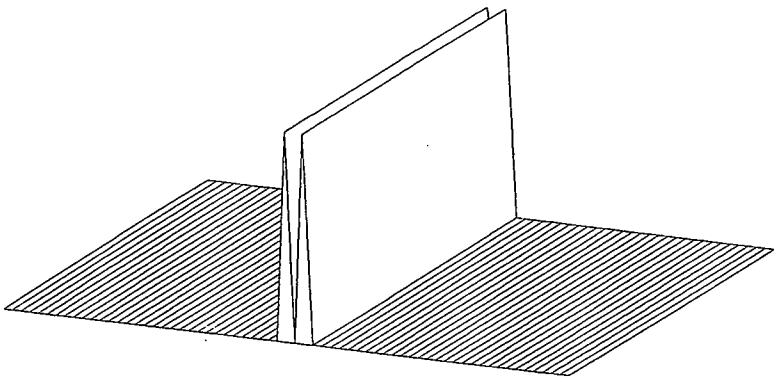

Figure 4.2: Time evolution of Fourier modes of the solution of the fourth order (4.24). Initial data is (4.71), parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 5.7895$, and $m = 33$.

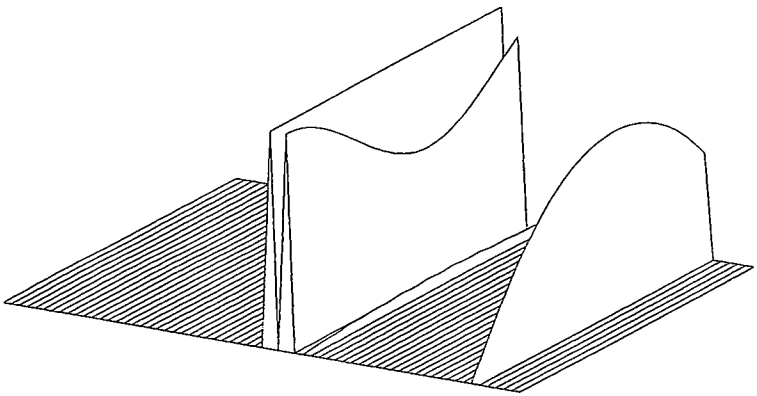

Figure 4.3: Time evolution of Fourier modes of the solution of the fourth order (4.24). Initial data is (4.71), parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 5.7895$, and $m = 34$.

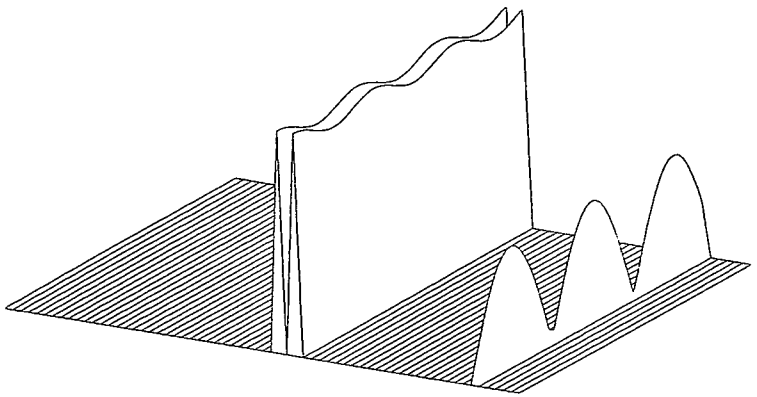

Figure 4.4: Time evolution of Fourier modes of the solution of the fourth order (4.24). Initial data is (4.71), parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 5.7895$, and $m = 35$.

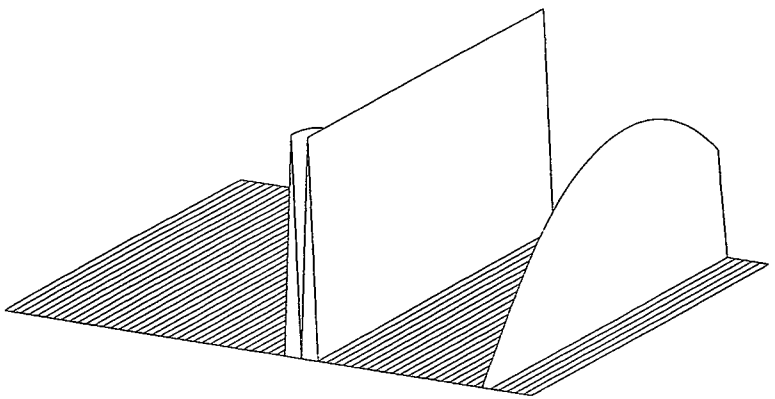

Figure 4.5: Time evolution of Fourier modes of the solution of the fourth order (4.24). Initial data is (4.71), parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 5.7895$, and $m = 36$.

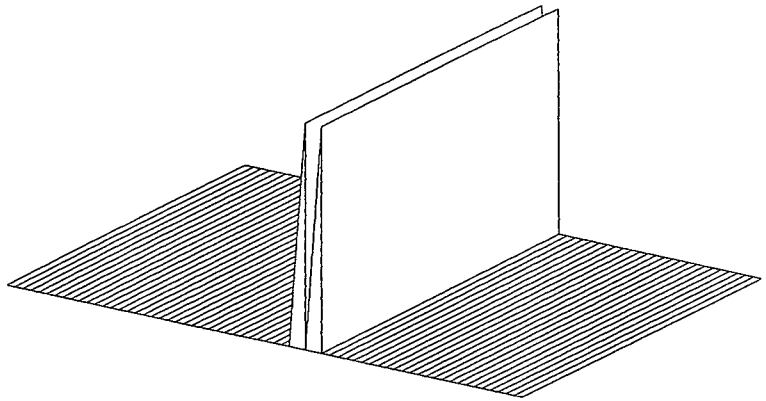

Figure 4.6: Time evolution of Fourier modes of the solution of the fourth order (4.24). Initial data is (4.71), parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 5.7895$, and $m = 37$.

For $m = 34, 35$, and 36, however, significant nonlinear activity is exhibited. As can be seen from Figures 4.3 through 4.5 the 58th mode is excited in all the cases. It is interesting and important to note that the 2θ term in (4.45) contains a carrier wave which on a finer grid would have had a wave number of 70π, which aliases to a mode with wave number 58π on our grid.

The spurious solutions discussed for the fourth-order approximation are also present for higher order approximations. In Figures 4.7 through 4.9 we show solutions for orders of approximation six, eight and ten respectively, with values of η which cause violation of (4.56) for $m = 35$.

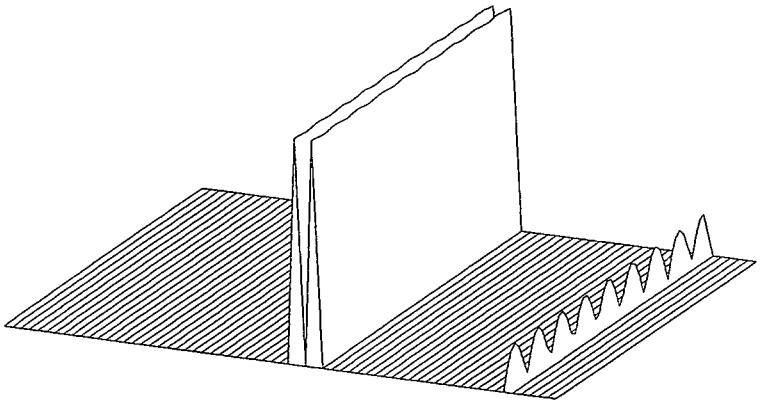

Figure 4.7: Sixth order solution subject to parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 6.3034$, with $m = 35$.

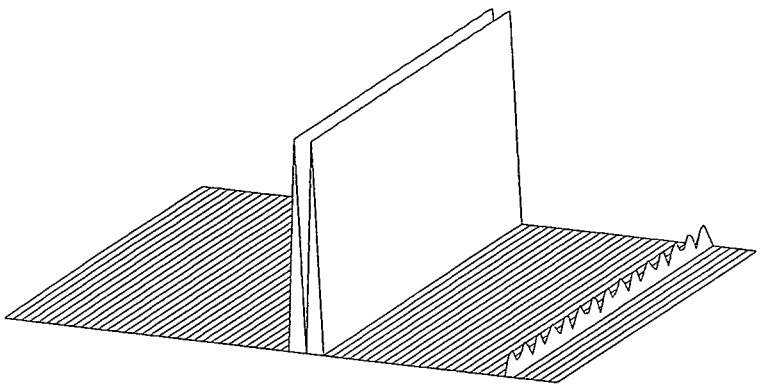

Figure 4.8: Eighth order solution subject to parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 6.6640$, with $m = 35$.

It is interesting to note from Figures 4.7 through 4.9 that the absolute magnitude of the spurious solutions tend to decrease for increasing orders of approximation.

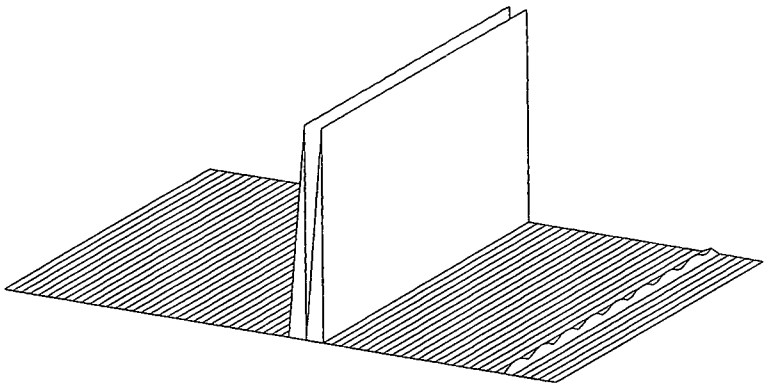

Figure 4.9: Tenth order solution subject to parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 6.9309$, with $m = 35$.

In Table 4.9 we show the l_2 and l_∞ values of the amplitudes of the spurious modes of (4.24) for various values of m and different orders of approximation, given the values of η which would violate (4.56) for the specific mode, all calculations subject to parameters (4.70).

m	30		40		50		60	
Order	l_2	l_∞	l_2	l_∞	l_2	l_∞	l_2	l_∞
2	2.8(−2)	5.2(−3)	8.8(−4)	1.8(−4)	1.6(−3)	3.1(−4)	2.0(−3)	4.0(−4)
4	1.1(−2)	2.2(−3)	2.2(−3)	4.3(−4)	5.2(−4)	1.0(−4)	3.3(−4)	6.5(−5)
6	8.3(−3)	1.6(−3)	5.2(−3)	9.9(−4)	7.4(−5)	1.5(−5)	1.9(−4)	3.8(−5)
8	7.9(−3)	1.6(−3)	2.0(−3)	4.0(−4)	1.2(−4)	2.3(−5)	2.4(−4)	4.8(−5)
10	2.2(−3)	4.3(−4)	2.5(−4)	5.0(−5)	2.4(−4)	4.9(−5)	3.3(−4)	6.5(−5)

Table 4.9: Amplitudes of the spurious mode for various values of m and various orders of approximation. (Notation: $2.8(-2) = 2.8 \times 10^{-2}$.)

As can be seen from Table 4.9, the magnitude of spurious solutions tend to decrease for higher order approximations relative to lower order approximations.

We also performed numerical experiments using the pseudospectral method to examine whether spurious effects are detected. No spurious effects were found using the

present form of analysis.

4.7.2 Violation of condition $\eta \neq V_g$ (4.59)

Using the parameters in (4.70) we solve for violation of equation (4.59) for various values of η subject to different mode numbers m and different orders of approximation. The results are shown in Table 4.10 below.

	Order of Approximation				
m	2	4	6	8	10
30	−4.7422	−14.7081	−35.0227	−76.3982	−162.9125
35	−2.8014	−8.3028	−17.6485	−33.0786	−58.1964
40	−0.9303	−3.5427	−7.6685	−13.6316	−21.9750
45	0.7589	0.0803	−1.3437	−3.3956	−6.0873
50	2.1649	2.7803	2.8264	2.5358	2.0030
55	3.2036	4.6410	5.4872	6.0217	6.3656
60	3.8126	5.6894	6.9223	7.8165	8.5014

Table 4.10: Values of η which violate (4.59) for various orders of approximation and different mode numbers (m) and subject to the parameters in (4.70).

As before we made use of the initial condition (4.71) with $\epsilon = 0.01$ and ran the scheme (4.24) for various orders of approximation. In Figure 4.10 we show the time evolution of Fourier modes of the second order solution plotted as a three-dimensional graph, subject to $\eta = 1.6423$ which violates (4.59) for $m = 48$. In this case a solution with a spurious low wave number mode is obtained.

Similarly, in Figure 4.11 we ran a fourth order solution with $\eta = 1.8038$, again violating (4.59) for $m = 48$. Figure 4.12 shows the Fourier evolution of modes of the sixth order scheme with $\eta = 1.3643$ violating (4.59) for $m = 48$.

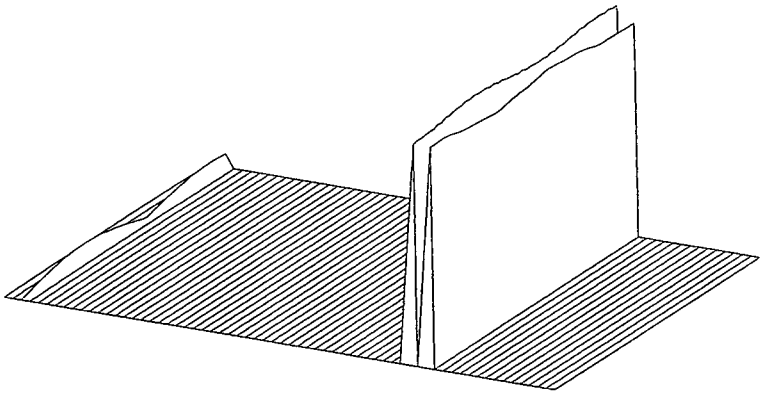

Figure 4.10: Second order solution subject to parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 1.6423$, with $m = 48$.

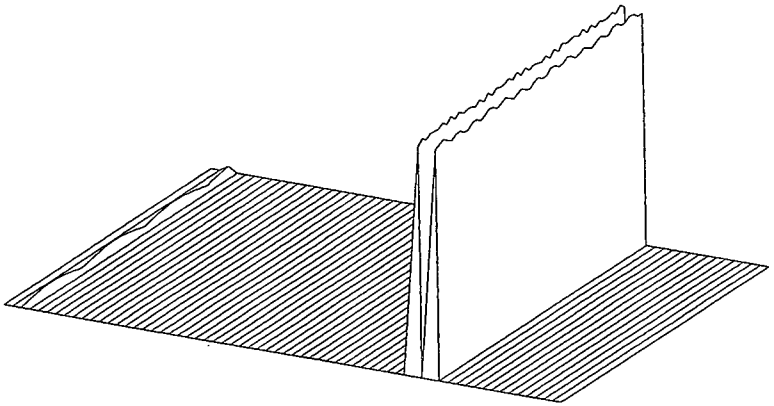

Figure 4.11: Fourth order solution subject to parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 1.8038$, with $m = 48$.

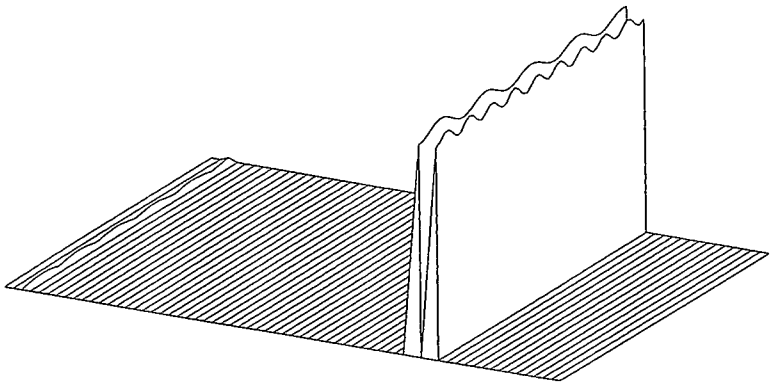

Figure 4.12: Sixth order solution subject to parameter values given by (4.70), $\epsilon = 0.01$, $\eta = 1.3643$, with $m = 48$.

4.8 Spurious Solution Discussion

The discrete multiple scales analysis described in Section 4.5 is a powerful tool for the examination of of the modulation properties of equations such as the KdV equation. We showed how the technique could identify modes of the carrier wave of the envelope of a small modulated harmonic wave, for which the solution of the numerical approximation deviates sharply from that of the continuous KdV equation.

In Sections 4.7.1 and 4.7.2 we have demonstrated spurious solutions to the KdV equation that arise from central difference approximations of various orders with a Leapfrog discretization in time. The spurious solutions were predicted from the discrete multiple scales analysis performed in Section 4.5. It is important to note that the spurious behaviour occurs as a function of the spatial discretization. The spurious behaviour tends to dissipate for higher order approximations but still forms a substantial part of the approximation.

Spurious solutions arising from discretizations to PDEs is a subject of intense interest. We cite some interesting references from the literature, namely [7, 21, 22, 37, 58, 81, 82, 89, 108]. Schreiber and Keller [108], for example, remark that for the solution of the steady Navier-Stokes equations, "there is at present no good theory to determine when a solution of the approximating problem is spurious and when it is legitimate".

Parasitic or nonphysical waves occuring in the numerical solution of certain wave equations can be due to several causes. In some cases such spurious waves are caused by the use of high order difference methods and they can often be filtered out by use of suitable difference operators applied to the initial data. In other cases the parasitic waves are observed in the numerical solution when discontinuities in the coefficients of the differential equation or the initial data are present. In these cases such waves correspond to high wave numbers and are the effect of numerical dispersion, as discussed by Trefethen [128], who considers the numerical solution of non-dispersive, linear wave equations - hence the only dispersion introduced would be that introduced by discretization.

Maritz and Schoombie [82] showed that with rectangular pulse intial data, discretized versions of the KdV and MKdV equations have high wave number components in their solutions which are not present in the analytical solutions of these equations. In particular they reported the occurence of small amplitude, saw toothed wave packets when using the ZK approximation to the KdV equation with rectangular pulse initial data. Sloan [115] also considered the ZK approximation to the KdV equation showing that the presence of the dispersive term causes modulational instabilities. In the precursor to the present work, Schoombie [111] illustrated spurious solutions which occur for specific choices of parameters in the solution of the KdV equation by the ZK method.

An important question that needs to be asked is how to suppress or prevent the spurious solution behaviour. We performed all our numerical experiments using the pseudospectral method with the same parameters that caused spurious solutions for the central difference approximations - no spurious solutions were found. Indeed, from the experiments we concluded that the effect of the spurious solutions becomes smaller for higher order approximations, cf. Table 4.9. As the pseudospectral method is the limiting method of central difference approximations of increasing orders we would postulate that the specific spurious solutions examined here would not be present in pseudospectral solutions. Adaptive methods [40, 50] or methods involving grid-changes [21] would also prevent perpetuation of the spurious solutions.

Chapter 5

Benjamin-Feir Analysis

I wish these calculations had been executed by steam.

Charles Babbage (1792 - 1871)

Do not worry about your difficulties in mathematics, I assure you that mine are greater.

Albert Einstein (1879 - 1955)

We have endeavored to illustrate the application of perturbation methods for the analysis of numerical approximation techniques, in particular finite difference methods. We provided a discussion on perturbation techniques as applied to continuous problems. We showed specifically that for the KdV and RLW equations the envelopes of modulated waves are governed by the nonlinear Schrödinger (NLS) equation.

The specific multiple scales technique that we used was shown to be extendable to a discrete multiple scales analysis. We showed the application of the discrete multiple scales analysis technique to the analysis of generalized central finite difference approximations to the KdV equation. Of specific importance the discrete multiple scales analysis was shown to be able to identify spurious behavior of the numerical approximations. The spurious behavior was shown in several numerical case studies. We also obtained a discrete version of the NLS equation which described the modulation properties of solutions of the generalized central finite difference approximations to the KdV equation in Section 4.5. In Section 4.6 we showed that in the limit as

$h \to 0$ the discrete NLS equation tends to the continuous equation, hence the discrete version can be considered to be a valid approximation to the continuous NLS equation also.

In this chapter we consider a so-called Benjamin-Feir [12, 138] analysis of the continuous and discrete NLS equations. The purpose of the analysis would be to define regions of instability (subject to parameters or discretization) of the NLS equation.

In the first section we show the methodology applied to the NLS in simple form followed by an application to the NLS obtained in Section 1.3.2. In the third section we consider the analysis for the discrete NLS from Section 4.5.

5.1 Benjamin-Feir Instability Analysis - NLS

In this section we consider an instability analysis of the NLS equation. For this purpose we use the canonical form of the NLS, to wit

$$i\partial_t u + \partial_x^2 u + q|u|^2 u = 0, \quad i^2 = -1, \tag{5.1}$$

assuming periodic boundary conditions

$$u(x,t) = u(x \pm L, t), \tag{5.2}$$

throughout.

Considering the special case $q = 0$, the linear dispersion equation

$$i\partial_t u + \partial_x^2 u = 0, \tag{5.3}$$

has solutions of the form

$$u(x,t) = a\exp[i(kx - \omega t)], \tag{5.4}$$

where a is a complex constant and ω satisfies the dispersion relationship

$$\omega = k^2.$$

The NLS equation (5.1) also has solutions of the form (5.4) where the dispersion relationship is given by

$$\omega = k^2 - q|a|^2. \tag{5.5}$$

Our purpose is to investigate the stability of the solution (5.4). We therefore perturb this solution to

$$\begin{aligned} u(x,t) &= a\exp[i(kx-\omega t)](1+\epsilon(x,t)) \\ &\equiv A_0(1+\epsilon), \end{aligned} \tag{5.6}$$

where $||\epsilon|| \ll 1$.

Noting that

$$q|u|^2u = q|A_0|^2A_0(1+2\epsilon+\epsilon^*),$$

and using the fact that A_0 is a solution of (5.1) we obtain

$$i\partial_t(A_0\epsilon) + \partial_x^2(A_0\epsilon) + q|A_0|^2A_0(2\epsilon+\epsilon^*) = 0. \tag{5.7}$$

Developing the derivatives in the equation and noting that $\partial_x A_0 = ikA_0$, we have

$$\partial_t\epsilon = i\partial_x^2\epsilon - 2k\partial_x\epsilon + iq|a|^2(\epsilon+\epsilon^*). \tag{5.8}$$

Following [138] we assume ϵ to be periodic in the interval $[-L/2, L/2]$ and express it as the Fourier series

$$\epsilon(x,t) = \sum_{n=-\infty}^{\infty} \hat{\epsilon}_n(t)\exp(i\mu_n x), \tag{5.9}$$

with wave numbers

$$\mu_n = \frac{2\pi n}{L}, \tag{5.10}$$

leading to the following system of ODEs upon substitution into (5.8)

$$\begin{aligned} \frac{d}{dt}\begin{pmatrix}\hat{\epsilon}_n \\ \hat{\epsilon}^*_{-n}\end{pmatrix} &= i\begin{pmatrix} q|a|^2 - 2k\mu_n - \mu_n^2 & q|a|^2 \\ -q|a|^2 & -q|a|^2 - 2k\mu_n + \mu_n^2\end{pmatrix}\begin{pmatrix}\hat{\epsilon}_n \\ \hat{\epsilon}^*_{-n}\end{pmatrix} \\ &\equiv iG_n\begin{pmatrix}\hat{\epsilon}_n \\ \hat{\epsilon}^*_{-n}\end{pmatrix}, \end{aligned} \tag{5.11}$$

where $n = -\infty, ..., \infty$ but $n \neq 0$.

The *characteristic polynomial* [55] of the matrix G_n is given by

$$p(\lambda) = \lambda^2 + 4k\mu_n\lambda + (2q\mu_n^2|a|^2 + 4k^2\mu_n^2 - \mu_n^4), \tag{5.12}$$

the roots of which are the eigenvalues

$$\lambda_n = -2k\mu_n \pm \mu_n\sqrt{\mu_n^2 - 2q|a|^2}. \tag{5.13}$$

One of the eigenvalues will be negative imaginary and accordingly the $|\hat{\epsilon}_n|$ would grow exponentially should

$$0 < \mu_n^2 < 2q|a|^2, \tag{5.14}$$

subject to $q > 0$. This is the so-called Benjamin-Feir [12, 121, 138] region of instability.

We note, as in [63], that the analysis would only be valid for a small time period while the perturbation ϵ is small. The long time behavior of the solution would be determined by various conservation laws such as

$$\frac{d}{dt}\int_{-\infty}^{\infty}|u(x,t)|^2dx = 0,$$

and

$$\frac{d}{dt}\int_{-\infty}^{\infty}(|\partial_x u(x,t)|^2 - \frac{q}{2}|u(x,t)|^4)dx = 0.$$

5.2 KdV Modulational Stability Analysis

In Chapter 1 we showed that the NLS equation, written in the form (1.75)

$$\partial_{T_2}A + 3i\gamma k\partial_{X_1}^2 A - [i\zeta^2/(6\gamma k)]A|A|^2 = 0,$$

describes nonlinear modulation properties of the KdV equation (1.46). We now wish to apply the results of Section 5.1 to this version of the NLS. We rewrite the NLS as

$$i\partial_{T_2}A - 3\gamma k\partial_{X_1}^2 A + [\zeta^2/(6\gamma k)]A|A|^2 = 0. \tag{5.15}$$

To avoid confusion of notation we shall here consider solutions of the form

$$u(x,t) = a\exp[i(lx - \omega t)].$$

Applying the methodology we find the growth matrices G_n as previously defined in equation (5.11) to be given by

$$G_n = \begin{pmatrix} \zeta^2|a|^2/(6\gamma k) + 6\gamma kl\mu_n + 3\gamma k\mu_n^2 & \zeta^2|a|^2/(6\gamma k) \\ -\zeta^2|a|^2/(6\gamma k) & -\zeta^2|a|^2/(6\gamma k) + 6\gamma kl\mu_n - 3\gamma k\mu_n^2 \end{pmatrix}, \tag{5.16}$$

the eigenvalues of which are given by

$$\lambda_n = 6\gamma k l \mu_n \pm \mu_n \sqrt{9\mu_n^2\gamma^2k^2 + \zeta^2|a|^2}, \tag{5.17}$$

and are real. Therefore, no Benjamin-Feir region of instability exists for the NLS equation in (5.15).

In the next section we shall perform the analysis on the discrete version of (5.15) as obtained in Section 4.5.

5.3 Benjamin-Feir Analysis - Discrete

Our purpose in this section in to extend the Benjamin-Feir analysis conducted in Section 5.1 for the continuous NLS equation to a discrete version of the NLS. We consider the discrete NLS equation (4.64) obtained from the discrete multiple scales analysis of the ZK approximation to the KdV equation, namely

$$\partial_{T_2}V_1 + i(h^2/2)[\frac{\eta}{h}\sin(kh) + \frac{2\gamma}{h^3}(2\sin(2kh) - \sin(kh))]\Delta_{X_1}\nabla_{X_1}V_1$$

$$+i\frac{\zeta^2\sin(kh)}{3h}[\frac{\cos(kh)+2}{V_g-\eta} + \Lambda(1+2\cos(kh))]V_1|V_1|^2 = 0, \tag{5.18}$$

where V_g and Λ follows from equations (4.50) and (4.54) respectively, subject to the ZK-specific parameters given by (4.25).

We write equation (5.18) as follows:

$$\partial_{T_2}V_1 + iA\Delta_{X_1}\nabla_{X_1}V_1 + iBV_1|V_1|^2 = 0, \tag{5.19}$$

where we define

$$A := (h^2/2)[\frac{\eta}{h}\sin(kh) + \frac{2\gamma}{h^3}(2\sin(2kh) - \sin(kh))], \tag{5.20}$$

and

$$B := \frac{\zeta^2\sin(kh)}{3h}[\frac{\cos(kh)+2}{V_g-\eta} + \Lambda(1+2\cos(kh))], \tag{5.21}$$

from equation (5.18).

Following from equation (5.4) we consider solutions of (5.19) of the form

$$V_1 = a\exp[i(KX_1 - \omega t)], \tag{5.22}$$

where the discrete dispersion relationship is given by

$$\omega = -\frac{2A}{\epsilon^2 h^2}[\cos(K\epsilon h) - 1] - B|a|^2. \tag{5.23}$$

In the derivation of (5.23) we use the following result:

$$\begin{aligned}\Delta_{X_1}\nabla_{X_1}V_1 &= \frac{2}{\epsilon^2 h^2}[\cos(K\epsilon h) - 1]\, a\exp[i(KX_1 - \omega t)] \\ &= \frac{2}{\epsilon^2 h^2}[\cos(K\epsilon h) - 1]V_1. \end{aligned}\tag{5.24}$$

Note in the limit as $\epsilon h \to 0$ we regain the analogous continuous nonlinear dispersion relationship (5.5) from (5.23). The discrete dispersion relationship is $O(h^2)$ accurate.

As before, we consider the stability of the solution (5.22). Our approach shall be to perturb the solution by $\Sigma(X_1, t)$, subject to $|\Sigma|^2 \ll 1$, i.e., we write

$$V_1 = \bar{V}_1(1 + \Sigma(X_1, t)), \tag{5.25}$$

where $\bar{V}_1$ is

$$\bar{V}_1 = a\exp[i(KX_1 - \omega t)]. \tag{5.26}$$

Noting that

$$\Delta_{X_1}\nabla_{X_1}V_1 = \Delta_{X_1}\nabla_{X_1}\bar{V}_1 + \Delta_{X_1}\nabla_{X_1}(\bar{V}_1\Sigma), \tag{5.27}$$

and

$$|V_1|^2 V_1 \approx |\bar{V}_1|^2 V_1(1 + 2\Sigma + \Sigma^*), \tag{5.28}$$

we find upon substitution in equation (5.19) that

$$\partial_{T_2}(\bar{V}_1\Sigma) + iA\Delta_{X_1}\nabla_{X_1}(\bar{V}_1\Sigma) + iB|\bar{V}_1|^2 V_1(2\Sigma + \Sigma^*) = 0. \tag{5.29}$$

We proceed by expanding

$$\begin{aligned}&\Delta_{X_1}\nabla_{X_1}(\bar{V}_1\Sigma) \\ =\ &\frac{1}{\epsilon^2 h^2}[E_{X_1} - 2 + E_{X_1}^{-1}](\bar{V}_1\Sigma) \\ =\ &\frac{1}{\epsilon^2 h^2}[\cos(K\epsilon h)\{E_{X_1}\Sigma + E_{X_1}^{-1}\Sigma\} + i\sin(K\epsilon h)\{E_{X_1}\Sigma - E_{X_1}^{-1}\Sigma\} - 2\Sigma]\bar{V}_1,\end{aligned}$$

from which follows that

$$\begin{aligned}&\Delta_{X_1}\nabla_{X_1}(\bar{V}_1\Sigma)\\ &= \frac{1}{\epsilon^2h^2}[2(\cos(K\epsilon h)-1)\Sigma+\epsilon^2h^2\cos(K\epsilon h)\Delta_{X_1}\nabla_{X_1}\Sigma+2i\epsilon h\sin(K\epsilon h)\delta_{X_1}\Sigma]\bar{V}_1,\end{aligned} \tag{5.30}$$

upon making use of the definitions of δ_{X_1}, Δ_{X_1} and ∇_{X_1} as defined in equations (3.10), (3.6) and (3.7) respectively, i.e.,

$$E_{X_1}+E_{X_1}^{-1}=2+\epsilon^2h^2\Delta_{X_1}\nabla_{X_1},$$

and

$$E_{X_1}-E_{X_1}^{-1}=2\epsilon h\delta_{X_1}.$$

Noting also equation(5.24) we can rewrite (5.29) in the form

$$\partial_{T_2}\Sigma+i\frac{A}{\epsilon h}[\epsilon h\cos(K\epsilon h)\Delta_{X_1}\nabla_{X_1}\Sigma+2i\sin(K\epsilon h)\delta_{X_1}\Sigma]+i\zeta^2B|a|^2(\Sigma+\Sigma^*)=0, \tag{5.31}$$

Expansion of Σ in the discrete Fourier series

$$\Sigma(x,t)=\sum_{n=-N/2}^{N/2}\hat{\Sigma}_n(t)\exp(i\mu_nX_1), \tag{5.32}$$

with

$$\mu_n=\frac{2\pi n}{L},$$

leads to the following system of ODEs upon substitution into (5.8)

$$\begin{aligned}\frac{d}{dt}\begin{pmatrix}\hat{\Sigma}_n\\ \hat{\Sigma}^*_{-n}\end{pmatrix} &= iG_n\begin{pmatrix}\hat{\Sigma}_n\\ \hat{\Sigma}^*_{-n}\end{pmatrix}\\ &\equiv i\begin{pmatrix}G_{11}&G_{12}\\ G_{21}&G_{22}\end{pmatrix}\begin{pmatrix}\hat{\Sigma}_n\\ \hat{\Sigma}^*_{-n}\end{pmatrix},\end{aligned} \tag{5.33}$$

where

$$\begin{aligned}G_{11} &= \frac{-2A}{\epsilon^2h^2}[\cos(K\epsilon h)(\cos(\mu_n\epsilon h)-1)-\sin(K\epsilon h)\sin(\mu_n\epsilon h)]-B|a|^2,\\ G_{12} &= -B|a|^2,\\ G_{21} &= \frac{2A}{\epsilon^2h^2}[\cos(K\epsilon h)(\cos(\mu_n\epsilon h)-1)+\sin(K\epsilon h)\sin(\mu_n\epsilon h)]+B|a|^2,\\ G_{22} &= B|a|^2.\end{aligned}$$

The characteristic polynomial of G_n is given by

$$(\lambda - \frac{2A}{\epsilon^2 h^2}\sin(K\epsilon h)\sin(\mu_n \epsilon h))^2$$

$$-(\frac{2A}{\epsilon^2 h^2}(\cos(K\epsilon h)(\cos(\mu_n \epsilon h) - 1) + B|a|^2)^2 + (B|a|^2)^2 = 0, \qquad (5.34)$$

the roots of which produce the eigenvalues of G_n.

Based on the characteristic polynomial, the eigenvalues of G_n would have negative imaginary parts, and hence allow exponential growth in the side-band amplitudes $|\hat{\Sigma}_n|$, whenever

$$(\frac{2A}{\epsilon^2 h^2}(\cos(K\epsilon h)(\cos(\mu_n \epsilon h) - 1) + B|a|^2)^2 - (B|a|^2)^2 < 0. \qquad (5.35)$$

Considering the special (but important) case $K = 0$, condition (5.35) becomes

$$\frac{8A}{\epsilon^2 h^2}\sin^2(\mu_n \epsilon h/2)\{\frac{2A}{\epsilon^2 h^2}\sin^2(\mu_n \epsilon h/2) - B|a|^2\} < 0. \qquad (5.36)$$

Therefore, if $A > 0$ we would have instability when

$$\frac{2A}{\epsilon^2 h^2}\sin^2(\mu_n \epsilon h/2) < B|a|^2, \qquad (5.37)$$

implying that for $B < 0$ we would have no unstable modes, but for $B > 0$ all modes would be unstable whenever

$$\epsilon^2 h^2 B|a|^2 > 2A. \qquad (5.38)$$

Similarly, if $A < 0$ we would have instability when

$$\frac{2A}{\epsilon^2 h^2}\sin^2(\mu_n \epsilon h/2) > B|a|^2, \qquad (5.39)$$

implying that for $B > 0$ we would have no unstable modes, but for $B < 0$ all modes would be unstable whenever condition (5.38) holds.

For consistency with the results of Section 5.2 we consider the limit case $h \to 0$. Then

$$\lim_{h \to 0} A = 3\gamma k,$$

and

$$\lim_{h \to 0} B = \frac{-\zeta^2}{6\gamma k},$$

implying that in the limit $AB = -\zeta^2/2 < 0$; hence the continuous equation would be stable, the result obtained in Section 5.2.

Chapter 6

Van der Pol Discrete

I have a theory that whenever you want to get in trouble with a method, look for the Van der Pol equation.

P. E. Zadunaisky

In this chapter we shall analyze some numerical approximations to the Van der Pol equation (1.26). In [90], for example, IMSL routines DIVPRK and DIVPAG are used to solve for limit cycles of the generalized Van der Pol equation. The IMSL routines DIVPRK [102], efficient for non-stiff systems, uses RK formulae of order five and six whereas DIVPAG utilizes the implicit Adams method up to order twelve as well as Geer's stiff method. A similar approach is followed in [60] where the results of extrapolation based codes (with stepsize control) and Runge-Kutta-Nyström codes are compared. Details of these methods are described in [75] and [141], for example.

The specific method examined in this chapter follows the work described in Cai *et al.* [24] where a Leapfrog scheme is applied to the solution of the Van der Pol equation. The specific method is interesting because it leads to spurious solution behavior. We apply a basic discrete multiple scales method to this equation as originally performed by Schoombie [110] for the KdV equation. We shall show details of the numerical solution as well as an interpretation of the inherent faults in the method as identified by the multiple scales solution of the numerical technique.

6.1 Leapfrog Scheme

We rewrite the Van der Pol equation (1.26) for the sake of convenience

$$\frac{d^2x}{dt^2} - \beta\epsilon(1-x^2)\frac{dx}{dt} + \omega^2 x = 0, \tag{6.1}$$

subject to the conditions $\beta > 0$ and ω a constant. We assume that ϵ satisfies

$$(0 < \epsilon \ll 1). \tag{6.2}$$

Also, for ease of reference, we repeat the perturbation solution to $O(\epsilon)$ as given by equation (1.45):

$$x(t,\epsilon) = \frac{2\cos(\omega t + \phi)}{\sqrt{(1 + (4/a_0^2 - 1)e^{-\beta\epsilon t})}} + O(\epsilon). \tag{6.3}$$

For the purpose of derivation of the numerical method we rewrite the Van der Pol equation in the form of a system of simultaneous first order ODEs, namely [24, 60]

$$\frac{dx}{dt} = y, \tag{6.4}$$

$$\frac{dy}{dt} = \beta\epsilon(1-x^2)y - \omega^2 x. \tag{6.5}$$

Following the approach by Cai *et al.* [24] we employ the central difference approximation (2.14) as an approximation to the differential operators in (6.4) and (6.5). We use the notation from Chapter 2 namely, $x^n = x(n\tau)$ and $y^n = y(n\tau)$ where τ is a fixed time step. We therefore obtain the following system of difference equations:

$$\delta_t x^n = y^n, \tag{6.6}$$

$$\delta_t y^n = \beta\epsilon(1-(x^n)^2)y^n - \omega^2 x^n. \tag{6.7}$$

By substituting (6.6) into (6.7) and using the basic definitions (2.6) and (2.14) we obtain

$$\begin{aligned}&(E^2 - 2 + E^{-2})x^n/(4\tau^2)\\ &-\beta\epsilon(1-(x^n)^2)(E - E^{-1})x^n/(2\tau) + \omega^2 x^n = 0.\end{aligned} \tag{6.8}$$

To write (6.8) in operator form, we note that

$$\begin{aligned}&(E^2 - 2 + E^{-2})/(4\tau^2)\\ &= [(1+\tau\Delta_t)^2 - 2 + (1+\tau\nabla_t)^2]/(4\tau^2)\\ &= [\Delta_t\nabla_t + \tau^2/4\Delta_t^2\nabla_t^2].\end{aligned} \tag{6.9}$$

Therefore, by virtue of (6.9) we rewrite (6.8) as

$$[\Delta_t \nabla_t + \tau^2/4\Delta_t^2\nabla_t^2]x^n - \beta\epsilon(1-(x^n)^2)\delta_t x^n + \omega^2 x^n = 0. \tag{6.10}$$

In Figure 6.1 we show x^n obtained from equations (6.6) and (6.7) for $\epsilon = 0.025$ and $\tau = 0.2$. Initial values are given by $x^0 = 0$ and $y^0 = 0.5$. The x^1 and y^1 are calculated by making use of an Euler starter with time increment τ.

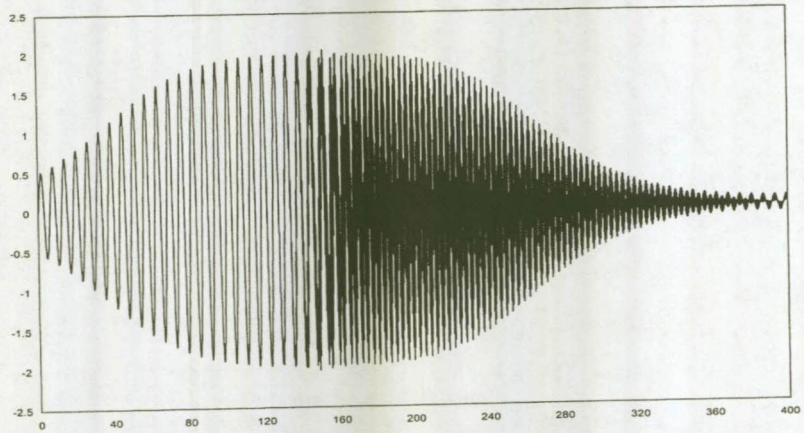

Figure 6.1: Solution of the Van der Pol equation using the Leapfrog scheme (6.6) and (6.7) with $\epsilon = 0.025, \tau = 0.2$. Initial conditions are $x^0 = 0$ and $y^0 = 0.5$.

From Figure 6.1 it is quite clear that the numerical solution deviates significantly from the perturbation solution (6.3). Of particular significance is the failure to obtain the asymptotic solution behavior discussed in Section 1.2, namely

$$\lim_{t\to\infty} x(t) = 2\cos(t+\phi) + O(\epsilon).$$

The numerical solution becomes meaningless for times $t > 100$. This is a good example of spurious behavior of numerical solutions to ODEs where the spurious nature renders the solution meaningless.

Also of interest, the solution pattern starts repeating itself as time increases. This behavior is illustrated in Figure 6.2.

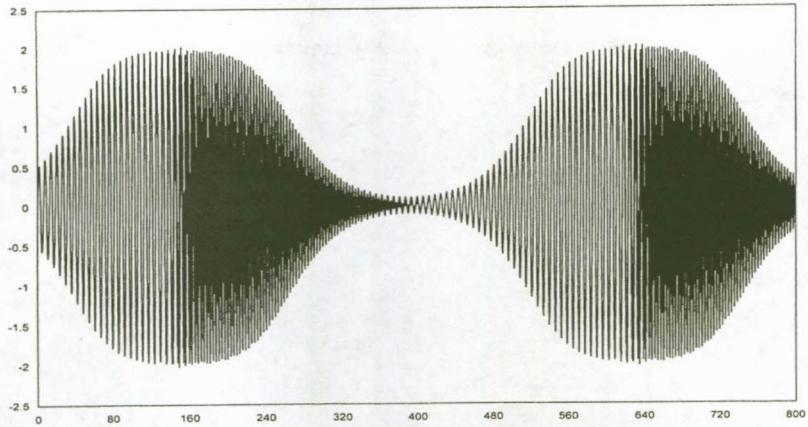

Figure 6.2: Solution of the Van der Pol equation using the Leapfrog scheme (6.6) and (6.7) with $\epsilon = 0.025, \tau = 0.2$ over a longer time period. Initial conditions are $x^0 = 0$ and $y^0 = 0.5$.

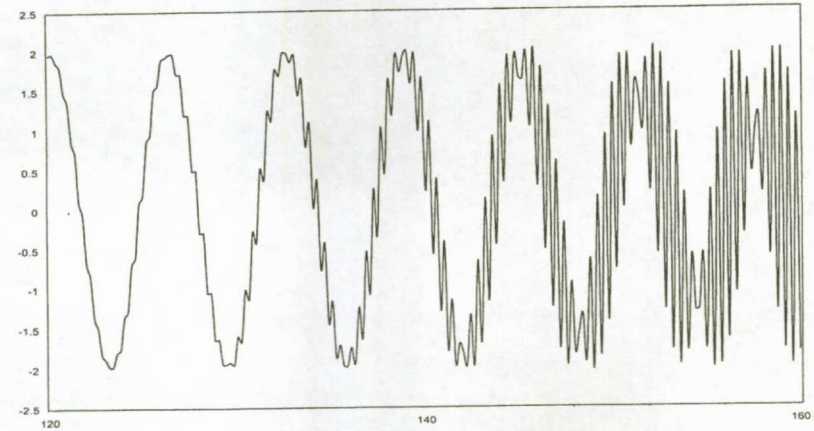

Figure 6.3: Solution of the Van der Pol equation using the Leapfrog scheme (6.6) and (6.7) with $\epsilon = 0.025, \tau = 0.2$ – solution magnified between times $t = 120$ and $t = 160$. Initial conditions are $x^0 = 0$ and $y^0 = 0.5$.

In Figure 6.3 we show x^n for $t \in [120, 160]$. The figure shows the onset of the spurious solution behavior.

In the following two sections we shall provide a discrete multiple scales solution of (6.10) and for comparison the results of a linear analysis as performed by Cai *et al.* [24].

6.2 Discrete Scales Solution

To perform a multiple scales analysis of (6.10) we shall work in direct analogy with the continuous multiple scales analysis performed for the Van der Pol equation. We shall use two discrete time scales, namely

$$T_p = \epsilon^p n\tau, \quad p = 0, 1, \tag{6.11}$$

and consider the expansion [110]

$$u^n = u_0(T_0, T_1) + \epsilon u_1(T_0, T_1) + O(\epsilon^2). \tag{6.12}$$

By making use of Theorem 7 we note the discrete multiple scales expansion of the differential approximations Δ_t and ∇_t, (as shown previously in equations (3.24) and (3.25))

$$\Delta_t = \Delta_{T_0} + \epsilon \Delta_{T_1} E_{T_0}, \tag{6.13}$$

$$\nabla_t = \nabla_{T_0} + \epsilon \nabla_{T_1} E_{T_0}^{-1}. \tag{6.14}$$

Substitution of (6.13) and (6.14) into (6.9) has the following result:

$$\begin{aligned} &\Delta_t \nabla_t + \tau^2/4 \Delta_t^2 \nabla_t^2 \\ &= \Delta_{T_0} \nabla_{T_0} + \tau^2/4 \Delta_{T_0}^2 \nabla_{T_0}^2 \\ &+\epsilon(\Delta_{T_0}\Delta_{T_1} + \nabla_{T_0}\nabla_{T_1} + \tau^2/2\Delta_{T_0}^2 \nabla_{T_0}\Delta_{T_1} + \tau^2/2\nabla_{T_0}^2 \Delta_{T_0}\nabla_{T_1}). \end{aligned} \tag{6.15}$$

We introduce the following notation in (6.15)

$$\begin{aligned} L_{T_0T_0} &= \Delta_{T_0}\nabla_{T_0} + \tau^2/4\Delta_{T_0}^2\nabla_{T_0}^2 \\ L_{T_0T_1} &= \Delta_{T_0}\Delta_{T_1} + \nabla_{T_0}\nabla_{T_1} + \tau^2/2\Delta_{T_0}^2\nabla_{T_0}\Delta_{T_1} + \tau^2/2\nabla_{T_0}^2\Delta_{T_0}\nabla_{T_1}. \end{aligned} \tag{6.16}$$

By making use of the definition of δ_t in (2.14) and substitution of (6.13) and (6.14) therein, we also obtain that

$$\delta_t = \delta_{T_0} + \epsilon/2(\Delta_{T_1} E_{T_0} + \nabla_{T_1} E_{T_0}^{-1}). \tag{6.17}$$

Now, when substituting (6.12), (6.15) as well as (6.17) into (6.10) and collecting terms with equal powers of ϵ we generate a set of equations (similar to the continuous equations (1.33) to (1.35)) given by

$$L_{T_0T_0}u_0 + \omega^2 u_0 = 0, \tag{6.18}$$

relevant to $O(1)$ and for $O(\epsilon)$

$$L_{T_0T_0}u_1 + \omega^2 u_1 + L_{T_0T_1}u_0 - \beta(1-u_0^2)\delta_{T_0}u_0 = 0, \tag{6.19}$$

with $L_{T_0T_0}$ and $L_{T_0T_1}$ defined in (6.16).

Equation (6.18) has a solution of the form

$$u_0(T_0, T_1) = A(T_1)e^{i\Omega\tau n} + CC. \tag{6.20}$$

We observe that

$$\begin{aligned} &L_{T_0T_0}A(T_1)e^{i\Omega\tau n} \\ &= (\Delta_{T_0}\nabla_{T_0} + \tau^2/4\Delta_{T_0}^2\nabla_{T_0}^2)A(T_1)e^{i\Omega n\tau} \\ &= (\cos(2\Omega\tau)-1)/(2\tau^2)A(T_1)e^{i\Omega n\tau}, \end{aligned} \tag{6.21}$$

and therefore the relationship between Ω and ω is obtained by substitution of (6.20) into (6.18) and making use of (6.21):

$$\cos(2\Omega\tau) = 1 - 2\tau^2\omega^2. \tag{6.22}$$

By virtue of the familiar Maclaurin power series expansion of $\cos(\cdot)$ [1] we have that

$$\Omega^2 = \omega^2 + O(\tau^2). \tag{6.23}$$

To proceed we substitute (6.20) in (6.19). We note that

$$\begin{aligned} &(\Delta_{T_0}\Delta_{T_1} + \nabla_{T_0}\nabla_{T_1})A(T_1)e^{i\Omega\tau n} \\ &= (\epsilon\Delta_{T_1}\nabla_{T_1}A(T_1)(\cos(\Omega\tau)-1) + 2i\delta_{T_1}A(T_1)\sin(\Omega\tau)/\tau)e^{i\Omega\tau n}. \end{aligned} \tag{6.24}$$

Furthermore

$$\begin{aligned} &\tau^3\Delta_{T_0}^2\nabla_{T_0}\Delta_{T_1}A(T_1)e^{i\Omega n\tau} \\ &= [2(\cos(\Omega\tau)-1)^2 + 2i\sin(\Omega\tau)(\cos(\Omega\tau)-1)]\Delta_{T_1}A(T_1)e^{i\Omega n\tau}, \end{aligned} \tag{6.25}$$

and similarly

$$\begin{aligned} &\tau^3\nabla_{T_0}^2\Delta_{T_0}\nabla_{T_1}A(T_1)e^{i\Omega n\tau} \\ &= [-2(\cos(\Omega\tau)-1)^2 + 2i\sin(\Omega\tau)(\cos(\Omega\tau)-1)]\nabla_{T_1}A(T_1)e^{i\Omega n\tau}. \end{aligned} \tag{6.26}$$

Combining (6.25) and (6.26) results in

$$\begin{aligned}&\tau^3[\Delta_{T_0}^2\nabla_{T_0}\Delta_{T_1}+\nabla_{T_0}^2\Delta_{T_0}\nabla_{T_1}]A(T_1)e^{i\Omega n\tau}\\&=[4i\sin(\Omega\tau)(\cos(\Omega\tau)-1)]\delta_{T_1}A(T_1)e^{i\Omega n\tau}+O(\epsilon).\end{aligned}\tag{6.27}$$

We also find that

$$\begin{aligned}&\beta(1-u_0^2)\delta_{T_0}u_0\\&=\beta i\sin(\Omega\tau)/\tau(A-|A|^2A)e^{i\Omega n\tau}-A^3e^{3i\Omega n\tau}+CC.\end{aligned}\tag{6.28}$$

Therefore, by combining (6.24), (6.27) and (6.28) in (6.19) we have that

$$\begin{aligned}&L_{T_0T_0}u_1+\omega^2u_1\\&=i\sin(2\Omega\tau)/\tau\delta_{T_1}Ae^{i\Omega n\tau}-\beta i\sin(\Omega\tau)/\tau(A-|A|^2A)e^{i\Omega n\tau}\\&-A^3e^{3i\Omega n\tau}+CC.\end{aligned}\tag{6.29}$$

From (6.29) we find that in order to obtain bounded solutions for u_1 we need to remove secular terms by imposing the following requirement:

$$\sin(2\Omega\tau)\delta_{T_1}A-\beta\sin(\Omega\tau)(A-|A|^2A)=0.\tag{6.30}$$

Based on the condition that $\sin(\Omega\tau)\neq 0$ (which we exclude for all practical purposes) we obtain the discrete analogue of (1.38) from (6.30), i.e.,

$$\cos(\Omega\tau)\delta_{T_1}A-\beta/2(A-|A|^2A)=0.\tag{6.31}$$

As in the continuous case we assume a solution of (6.31) in the form

$$A=\frac{1}{2}a(T_1,T_2)e^{i\phi(T_1,T_2)}.\tag{6.32}$$

from which we obtain

$$\cos(\Omega\tau)\delta_{T_1}a=\beta/2(1-a^2/4)a,\tag{6.33}$$

and

$$\delta_{T_1}\phi=O(\epsilon^2),\tag{6.34}$$

after substitution into (6.31) and separation of real and imaginary parts. We rewrite (6.33) as

$$a^{n+1}-a^{n-1}=\alpha\tau a^n(1-(a^n)^2/4),\tag{6.35}$$

where we define α as

$$\alpha=\frac{\beta\epsilon}{\cos(\Omega\tau)}.\tag{6.36}$$

Note that we shall in practice use an Euler-type starter

$$a^1 = a^0 + \alpha\tau a^0(1 - (a^0)^2), \tag{6.37}$$

given the initial value a^0.

We shall first examine the fixed points of a continuous version of (6.33), similar to equation (1.40), namely

$$\frac{da}{dt} = \alpha a(1 - a^2/4) = \alpha S(a). \tag{6.38}$$

The fixed-points of this equation are roots of $a(1 - a^2/2)$; it has three fixed-points $\bar{a} = 0$, $\bar{a} = 2$ and $\bar{a} = -2$. To study the stability of these fixed-points we perturb the fixed-point with a disturbance ξ to obtain the perturbed equation

$$\frac{d\xi}{dt} = \alpha S(\bar{a} + \xi). \tag{6.39}$$

Following [142] we expand $S(\bar{a} + \xi)$ into a Taylor series around $\bar{a}$, so that

$$\frac{d\xi}{dt} = \alpha[S(\bar{a}) + \xi\frac{dS}{da}|_{\bar{a}} + \ldots]. \tag{6.40}$$

Stability is normally obtained by examining a small neighborhood of the fixed-point provided for a given value of α we have that

$$\alpha\frac{dS}{da} \neq 0. \tag{6.41}$$

Using this condition and neglecting successive powers of ξ we obtain the linear perturbed equation

$$\frac{d\xi}{dt}/\xi = \alpha\frac{dS}{da}|_{\bar{a}}. \tag{6.42}$$

The fixed point $\bar{a}$ is asymptotically stable if the right side of (6.42) is negative and unstable if it is positive. Higher order perturbations are needed if the right side of (6.42) equals 0.

Perturbing the Logistic-type equation (6.38) around its fixed points we find that $\bar{a} = 0$ is unstable whereas $\bar{a} = 2$ and $\bar{a} = -2$ respectively are stable. Note that the latter two would correspond to a stable limit cycle.

Considering the actual discrete problem on the other hand, we note that (6.33) has three constant solutions or fixed-points. They are

$$a^n = \bar{a}^0 = 0, \tag{6.43}$$

$$a^n = \bar{a}^1 = 2, \tag{6.44}$$

$$a^n = \bar{a}^2 = -2. \tag{6.45}$$

To investigate the stability of $a^n = \bar{a}^0$ we set [85]

$$a^n = \bar{a}^0 + \gamma^n, \quad |\gamma^n| \ll 1, \tag{6.46}$$

substitute this into (6.33) and neglect all but the linear terms. Doing this gives

$$\gamma^{n+1} - \gamma^{n-1} = \alpha\tau\gamma^n. \tag{6.47}$$

The solution to the above second-order difference equation is

$$\gamma^n = A(r_+)^n + B(r_-)^n, \tag{6.48}$$

where A and B are arbitrary constants, and

$$r_\pm = \frac{\alpha\tau \pm \sqrt{\alpha^2\tau^2 + 4}}{2}. \tag{6.49}$$

From (6.49) it can be concluded that the first term on the right of (6.48) is exponentially increasing, while the second term oscillates with an exponentially decreasing amplitude.

Similarly, we represent small perturbations to the fixed-point $\bar{a}^1$ by

$$a^n = \bar{a}^1 + \eta^n = 2 + \eta^n, \tag{6.50}$$

which results in a linear perturbation equation

$$\eta^{n+1} - \eta^{n-1} = -2\alpha\tau\eta^n, \tag{6.51}$$

whose solution is

$$\eta^n = C(s_+)^n + D(s_-)^n, \tag{6.52}$$

where C and D are small arbitrary constants and

$$s_\pm = -\alpha\tau \pm \sqrt{\alpha^2\tau^2 + 1}. \tag{6.53}$$

Therefore, the first term on the right-side of (6.49) exponentially decreases and the second term oscillates with exponentially increasing amplitude.

The same result holds for the fixed-point $\bar{a}^2$.

Putting these results together it follows that the central difference scheme (6.35) has exactly the same three fixed-points as the Logistic-type equation (6.38). However, while $a(t) = 0$ is (linearly) unstable and $a(t) = \pm 2$ is (linearly) stable for the differential equation all the fixed-points are linearly unstable for the difference scheme.

We also give a short account of the phase plane solution of (6.35). Following a technique of Sanz-Serna [106] we put:

$$a^n = \begin{cases} V^{n/2} & \text{if } n \text{ even} \\ W^{(n-1)/2} & \text{if } n \text{ odd.} \end{cases} \tag{6.54}$$

Therefore, we can express the scheme (6.35) as the augmented system [88, 106, 131]

$$V^{n+1} - V^n = \alpha\tau W^n(1 - (W^n)^2/4), \tag{6.55}$$

$$W^{n+1} - W^n = \alpha\tau V^n(1 - (V^n)^2/4). \tag{6.56}$$

The dynamics of the above system can be understood by taking the continuous limit and forming the following system:

$$\partial_t v = \alpha w(1 - w^2/4), \tag{6.57}$$

$$\partial_t w = \alpha v(1 - v^2/4). \tag{6.58}$$

The phase solution of this system is obtained by integration and is given by

$$8(v^2 - w^2) + (w^4 - v^4) = K, \tag{6.59}$$

where K is a constant of the integration.

In fact, the above system is a Hamiltonian system [59], which has only saddles and centers as equilibrium points. The equilibrium points (0,0), (-2,-2), (-2,2), (2,-2) and (2,2) correspond to saddles while (-2,0), (0,-2), (0,2) and (2,0) would be classified as centers. The full computer generated phase diagram is shown in Figure 6.4.

We can explain the mechanics of the Leapfrog solution of the nonlinear Logistic-type equation (6.35) as follows: The solution would follow one of the orbits in the phase plane. In the ideal situation the solution should follow the line $v = w$. However, as a result of errors in the relation between the initial condition and the next value provided by the Euler-type starter, or due to numerical round-off, it does not follow this orbit, but follows one close to it. Consequently, for initial conditions $u^0 \in (0, 2)$ it would follow one of the closed loops, approximating the line $v = w$ well for some time and then displaying large discrepancies between odd and even levels for some time as it returns to its initial state. For initial conditions $u^0 > 2$ it would follow an infinite orbit – the solution would follow the line $v = w$ closely but would deviate away from it as it approaches the fixed point $a^n = 2$ and would eventually move off to infinity, i.e., blow-up.

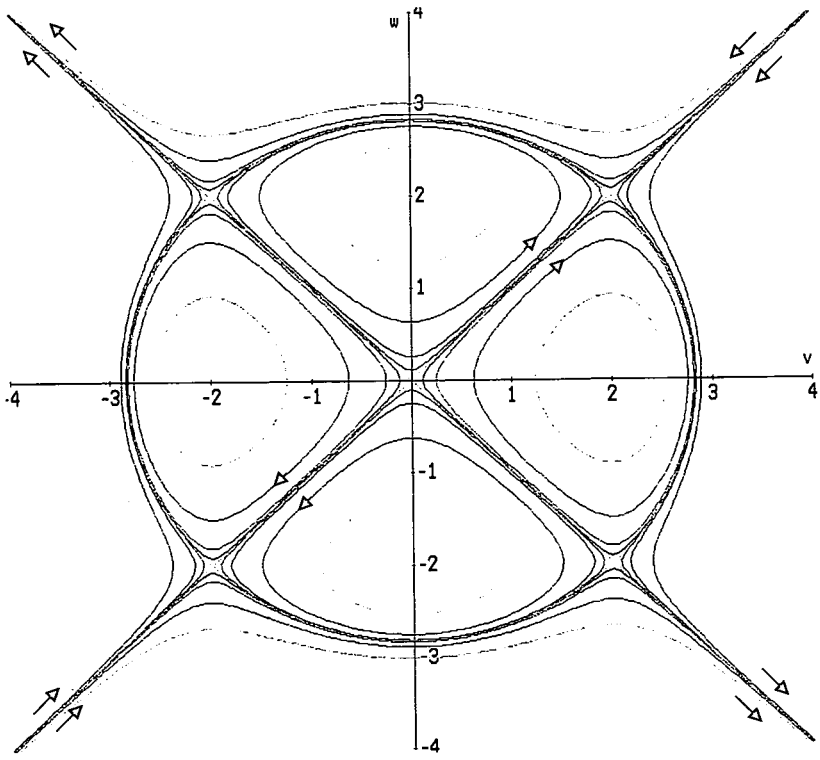

Figure 6.4: Phase diagram for a^n in equation (6.35).

6.3 Linear Analysis

In the manner of Cai *et al.* [24] a linear version of (6.10) is obtained by replacing the nonlinear term $(1-(x^n)^2)$ with $(1-X^2)$ where X is a constant. Therefore

$$[\Delta_t \nabla_t + \tau^2/4\Delta_t^2\nabla_t^2]x^n - \beta\epsilon(1-X^2)\delta_t x^n + \omega^2 x^n = 0, \tag{6.60}$$

which has a solution of the form $e^{in\Gamma}$, where

$$\sin^2(\Gamma) + \beta i\epsilon\tau(1-X^2)\sin(\Gamma) - \omega^2\tau^2 = 0, \tag{6.61}$$

or alternatively

$$\sin(\Gamma) = -i(1-X^2)\epsilon\tau\beta/2 \pm \sqrt{\omega^2 - (\beta/2)\epsilon^2(1-X^2)^2\tau}. \tag{6.62}$$

The general solution of (6.60) is therefore given as

$$x^n = e^{(\beta/2)(1-X^2)\epsilon n\tau}[A\cos(\phi n\tau) + B\sin(\phi n\tau)]$$
$$+(-1)^n e^{-(\beta/2)(1-X^2)\epsilon n\tau}[C\cos(\phi n\tau) + D\sin(\phi n\tau)], \tag{6.63}$$

where

$$\phi \equiv \sqrt{\omega^2 - (\beta/2)\epsilon^2(1-X^2)^2}, \tag{6.64}$$

and A, B, C, and D are constants.

The second term of equation (6.63) represents a computational mode which would become unstable when

$$|X| > 1. \tag{6.65}$$

Chapter 7

Epilogue

Euclid taught me that without assumptions there is no proof.
Therefore, in any argument, examine the assumptions.

Eric Temple Bell (1883 - 1960)

When asked how soon he expected to reach certain mathematical conclusions,
Gauss replied that he had them long ago,
all he was worrying about was how to reach them!

René J. Dubos

The main objective of the research project was to develop a perturbation technique for the solution of discrete equations. Furthermore we wished to apply the technique to discrete approximations of relevant equations, compare the theory with observed computed results and investigate deviations between the perturbation solutions and computed results with suitable explanations.

We discussed perturbation techniques in Chapter 1, specifically concentrating on an alternative multiple scales expansion [111, 127, 145]. Applying the technique we showed that for the KdV and RLW equations the envelopes of modulated waves are governed by the NLS equation, confirming a result of [36]. Our aim was to show that the alternative multiple scales methodology could be used for the analysis of more general PDEs than the KdV equation, the RLW being an example.

The alternative multiple scales analysis was chosen specifically for its ease of adap-

tation to the discrete case. In fact, Schoombie [111] demonstrated the use of the extension of the continuous methodology to perform a discrete multiple scales analysis of the ZK approximation to the KdV. We extended this analysis to cater for more general discrete approximation techniques.

To obtain his analysis, Schoombie [110, 111] derived a discrete analogue of the chain rule for differentiation. We showed how this result would be generalized for central difference schemes of arbitrary finite orders of approximation in Chapter 3 following from a detailed exposition on how these methods could be derived in Chapter 2.

In Chapter 4 we applied the discrete multiple scales analysis to the numerical solution of the KdV when discretized using generalized central finite differences. We demonstrated the consistency of the method with the continuous perturbation analysis as the discretization parameters tend to zero.

The discrete multiple scales technique is a powerful tool for the examination of modulational properties of the KdV equation. In the case of certain modes of the carrier wave, the discrete multiple scales analysis breaks down, indicating that the numerical solution deviates in behavior from that of the KdV equation. Several numerical experiments were performed to examine the spurious behavior for different orders of approximation. The spurious behavior, predicted theoretically, was shown to be present experimentally, independent of temporal discretization.

We showed that for the higher order approximations the spurious behavior tends to have a smaller effect on numerical solutions. For the pseudospectral method no spurious solutions were observed – we have not yet extended the analysis technique to cater for the pseudospectral method although we emphasized the result that Fourier differencing can be viewed as special centered finite differencing based on an ever increasing number of periodic stencils in Chapter 2. This represents an area which should be researched further both theoretically and experimentally.

An open question remains how to rid computations of the spurious behavior. We mentioned adaptive grids [40, 50] as a possibility, however, the spurious nature would remain, albeit locally. An exciting area of future research would be automatic recognition of conditions that would generate spurious solutions for a given PDE using a symbolic manipulation language.

We also detailed some comparisons of central difference solutions to the KdV equation. The results show a clear benefit of higher order central difference relative to lower order methods. The benefit of the central difference methodology would also extend to more general regions over which we would solve PDEs.

Although we placed specific emphasis on the KdV equation in this thesis we believe that the analyses are sufficiently general to be extended to other nonlinear wave equations. However, it is important to note that it is characteristic of nonlinear problems that each displays its own important features. Suitable numerical methods would generally be designed around those special features sometimes resulting in *ad hoc* methods. Accordingly it was not the aim of this thesis to identify best numerical methods for all nonlinear wave problems - we believe such a method probably does not exist.

The discrete scales method of analysis illustrated in this thesis can be applied to discretizations of other dispersive wave equations such as the MKdV equation or could be useful when dealing with models which are discrete to start with. This is an area where further research would prove fruitful.

Utilizing the discrete multiple scales analysis we also obtained a discrete version of the NLS equation which describes the modulation properties of solutions of the generalized central finite difference approximations to the KdV equation. We also showed that in the limit as $h \to 0$ the discrete NLS tends to the continuous equation, hence the discrete version can be considered to be a valid approximation to the continuous NLS equation as well. We supplemented this work of Chapter 4 with a Benjamin-Feir instability analysis for the ZK approximation to the KdV equation in Chapter 5. We note, as in [63], that the analysis would only be valid for a small time period while the perturbation is small. The long time behavior of the solution would ultimately be determined by various conservation laws. This analysis demands further numerical experimentation and could also be extended to higher order difference schemes.

We applied the discrete scales methodology to a particular discretization of the Van der Pol ODE. We demonstrated the mechanism by which numerical instability is created. This is also an area in which substantial future research work could be performed.

To conclude we quote from Richard Feynman's 1966 Nobel Lecture: "We have a habit in writing articles published in scientific journals to make the work as finished as possible, to cover up all the tracks, to not worry about the blind alleys or describe how you had the wrong idea first, and so on. So there isn't any place to publish, in a dignified manner, what you actually did in order to get to do the work." We wrote this thesis with this quote in mind, trying to show to the reader the process involved in obtaining the stated results.

Bibliography

[1] M. Abramowitz and I. A. Stegun (1965), *Handbook of Mathematical Functions*, New York: Dover.

[2] K. Abe and O. Inoue (1980), Fourier expansion solution of the Korteweg-de Vries equation, *J. Comp. Phys.* **34**, 202 – 210.

[3] P. Albrecht (1996), The Runge-Kutta theory in a nutshell, *SIAM J. Numer. Anal.* **33**, 1712 – 1735.

[4] W. F. Ames (1972), *Nonlinear Partial Differential Equations in Engineering*, Vol II, Academic Press.

[5] M. J. Ablowitz, B. M. Herbst and J. A. C. Weideman (1991), Dynamics of semi-discretizations of the defocusing nonlinear Schrödinger equation, *IMA J. Num. Anal.* **11**, 539 – 552.

[6] A. Aoyagi (1995), Nonlinear Leapfrog instability for Fornberg's pattern, *J. Comp. Phys.* **120**, 316 – 322.

[7] A. Aoyagi and K. Abe (1989), Parametric excitation of computational modes inherent to Leapfrog scheme applied to the Korteweg-de Vries equation, *J. Comp. Phys.* **83**, 447 – 462.

[8] A. Aoyagi and K. Abe (1991), Runge-Kutta smoother for suppression of computational - mode instability of Leapfrog scheme, *J. Comp. Phys.* **93**, 287 – 296.

[9] R. Bellman (1964), *Perturbation Techniques in Mathematics, Physics, and Engineering*, Holt, Rinehart and Winston, Inc.

[10] C. M. Bender and S. A. Orszag (1978), *Advanced Mathematical Methods for Scientists and Engineers*, New York: McGraw-Hill.

[11] T. B. Benjamin, J. L. Bona and J. J. Mahony (1971), Model equations for long waves in nonlinear dispersive media, *Phil. Trans. Roy. Soc. London* **A272**, 47 – 78.

[12] T. B. Benjamin and J. Feir (1966), Disintegration of wave trains on deep water, *J. Fluid Mech.* **27**, 417 – 430.

[13] F. Berman and K. D. Fryer (1972), *Introduction to Combinatorics*, Academic Press.

[14] P. L. Bhatnagar (1979), *Nonlinear Waves in One-dimensional Dispersive Systems*, Clarendon Press, Oxford.

[15] D. Bleecker and G. Csordas (1992), *Basic Partial Differential Equations*, Van Nostrand Reinhold.

[16] J. L. Bona (1978), Model equations for waves in nonlinear dispersive media, *Proceedings of the International Congress of Mathematicians*, Helsinki.

[17] J. L. Bona, W. G. Pritchard and L. R. Scott (1985), Numerical schemes for a model for nonlinear dispersive waves, *J. Comp. Phys.* **60**, 167 – 186.

[18] G. Boole (1960), *Finite Differences*, Fourth ed., Chelsea Publishing Company.

[19] J. P. Boyd (1989), *Chebyshev & Fourier Spectral Methods*, Springer Verlag.

[20] W. L. Briggs and V. E. Henson (1995), *The DFT: An Owner's Manual for the Discrete Fourier Transform*, SIAM.

[21] J. U. Brackbill and G. Lapenta (1994), A method to suppress the finite-grid instability in plasma simulations, *J. Comp. Phys.* **114**, 77 – 84.

[22] W. L. Briggs, A. C. Newell and T. Sarie (1983), Focusing: A mechanism for instability of nonlinear finite difference equations, *J. Comp. Phys.* **51**, 83 – 106.

[23] R. L. Burden and J. D. Faires (1985), *Numerical Analysis*, Third edition, Prindle Weber and Schmidt, Boston.

[24] D. Cai, A. Aoyagi and K. Abe (1993), Parametric excitation of computational mode of the Leapfrog scheme applied to the van der Pol equation, *J. Comp. Phys.* **107**, 146 – 151.

[25] C. Canuto, M. Y. Hussaini, A. Quarteroni and T. A. Zang (1988), *Spectral Methods in Fluid Dynamics*, Springer Verlag.

[26] T. F. Chan and T. Kerkhoven (1985), Fourier methods with extended stability intervals for the Korteweg-de Vries equation, *SIAM J. Numer. Anal.* **22**, 441 – 454.

[27] I. Christie, D. F. Griffiths, A. R. Mitchell and J. M. Sanz-Serna (1981), Product approximation for nonlinear problems in the finite element method, *IMA J. Num. Anal.* **1**, 253 – 266.

[28] S. R. Choudhury (1993), Some results on the stability and dynamics of finite difference approximations to nonlinear partial differential equations, *Num. Meth. PDE* **9**, 117 – 133.

[29] A. Cloot and B. M. Herbst (1988), Grid resonances, focusing and Benjamin-Feir instabilities in Leapfrog time discretizations, *J. Comp. Phys.* **75**, 31 – 53.

[30] A. Cloot and J. A. C. Weideman (1992), An adaptive algorithm for spectral computations on unbounded domains, *J. Comp. Phys.* **102**, 398 – 406.

[31] E. A. Cox and M. P. Mortell (1986), The evolution of resonant water-wave oscillations, *J. Fluid Mech.* **162**, 99 – 116.

[32] M. B. Dadfar and J. F. Geer (1990), Resonances and power series solutions of the forced van der Pol oscillator, *SIAM J. Appl. Math.* **50**, 1496 – 1506.

[33] J. Dieudonné (1960), *Foundations of Modern Analysis*, Academic Press, New York.

[34] J. Dieudonné (1971), *Infinitesimal Calculus*, Houghton-Mifflin, Boston.

[35] J. de Frutos and J. M. Sanz-Serna (1997), Accuracy and conservation properties in numerical integration: the case of the Korteweg-de Vries equation, *Numer. Math.* **75**, 421 – 445.

[36] R.K. Dodd, J.C. Eilbeck, J.D. Gibbon and H.C. Morris (1982), *Solitons and Nonlinear Wave Equations*, Academic Press, London.

[37] J. W. Eastwood and W. Arter (1987), Spurious behaviour of numerically computed fluid flow, *IMA J. Num. Anal.* **7**, 205 – 222.

[38] D. A. Edwards (2000), An alternative example of the method of multiple scales, *SIAM Rev.* **42**, 317 – 332.

[39] H. El-Zoheiry, L. Iskandar and B. El-Naggar (1994), The Quintic spline for solving the Korteweg-de Vries equation, *Math. Comp. Sim.* **37**, 539 – 549.

[40] I. Fatkullin and J. S. Hesthaven (2001), Adaptive high-order finite-difference method for nonlinear wave problems, *J. Sci. Comp.* **16**, 47 – 67.

[41] B. Feng and T. Mitsui (1998), A finite difference method for the Korteweg-de Vries and the Kadomtsev-Petviashvili equations, *J. Comput. Appl. Math.* **90**, 95 – 116.

[42] B. Fornberg (1973), On the instability of Leapfrog and Crank-Nicolson approximations of a nonlinear partial differential equation, *Math. Comp.* **121**, 45 – 57.

[43] B. Fornberg and G. B. Whitham (1978), A numerical and theoretical study of certain nonlinear wave phenomena, *Philos. Trans. Roy. Soc. London* **A289**, 373 – 404.

[44] B. Fornberg (1987), The pseudospectral method: Comparisons with finite differences for the elastic wave equation. Geophys. **52**, 483 – 501.

[45] B. Fornberg (1988), Generation of finite difference formulas on arbitrarily spaced grids, *Math. Comp.* **31**, 699 – 706.

[46] B. Fornberg (1996), *A Practical Guide to Pseudospectral Methods*, Cambridge University Press.

[47] B. Fornberg (1998), Calculation of weights in finite difference formulas, *SIAM Rev.* **40**, 685 – 691.

[48] B. Fornberg and T. Driscoll (1999), A fast spectral algorithm for nonlinear wave equations with linear dispersion, *J. Comp. Phys.* **155**, 456 – 467.

[49] G. E. Forsythe and W. R. Wasow (1967), *Finite Difference Methods for Partial Differential Equations*, John Wiley and Sons.

[50] E. S. Fraga and J. Ll. Morris (1996), A piecewise uniform adaptive grid algorithm for nonlinear dispersive wave equations, *Numerical analysis: A R Mitchell 75th birthday volume*, ed. D. F. Griffiths and G. A. Watson, World Scientific Publishing, Singapore.

[51] D. Furihate (1999), Finite difference schemes for $\partial_t u = \partial_x^\alpha \frac{\delta G}{\delta u}$ that inherit energy conservation or dissipation property, *J. Comp. Phys.* **156**, 181 – 205.

[52] I. Gladwell and R. Wait (1979), *A Survey of Numerical Methods for Partial Differential Equations*, Clarendon Press, Oxford.

[53] I. S. Greig and J. L. Morris (1976), A Hopscotch method for the Korteweg-de Vries equation, *J. Comp. Phys.* **20**, 64 – 80.

[54] K. Goda and Y. Fukui (1980), Numerical studies of the Regularized Long Wave equation, *J. Phys. Soc. Japan* **48**, 623 – 630.

[55] G. H. Golub and C. F. van Loan (1990), *Matrix Computations*, Second Edition, John Hopkins University Press.

[56] J. Goodman and P. D. Lax (1988), On dispersive difference schemes. I, *Comm. Pure Appl. Math.* **XLI**, 591 – 613.

[57] S. K. Godunov and V. S. Ryabenkii (1987), *Difference Schemes*, North Holland.

[58] D. F. Griffiths and A. R. Mitchell (1990), Spurious behaviour and nonlinear instability in discretised partial differential equations, IMA Conference Proceedings Series, Oxford University Press.

[59] D. W. Jordan and P. Smith (1999), *Nonlinear Ordinary Differential Equations*, Third Edition, Oxford University Press.

[60] E. Hairer, S.P. Nørsett and G. Wanner (1988), *Solving Ordinary Differential Equations I*, Springer Verlag.

[61] M. A. Helal (2002), Soliton solution of some nonlinear partial differential equations and its applications in fluid mechanics, *Chaos, Solitons and Fractals* **13**, 1917 – 1929.

[62] P. Henrici (1964), *Elements of Numerical Analysis*, John Wiley & Sons.

[63] B. M. Herbst, A. R. Mitchell and J. A. C. Weideman (1985), On the stability of the nonlinear Schrödinger equation, *J. Comp. Phys.*, **60**, 263 – 281.

[64] R. L. Herman and C. J. Knickerbocker (1993), Numerically induced phase shift in the KdV soliton, *J. Comp. Phys.* **104**, 50 – 55.

[65] M. Huang (1991), A Hamiltonian approximation to simulate solitary waves of the Korteweg-de Vries equation, *Math. Comp.* **194**, 607 – 620.

[66] R. L. Huston (1970), Krylov-Bogoljubov method for difference equations, *SIAM J. Appl. Math.* **19**, 334 – 339.

[67] L. Iskandar (1989), New numerical solution of the Korteweg-de Vries equation, *Appl. Num. Math.* **5**, 215 – 221.

[68] A. Jeffrey and T. Kakutani (1972), Weak nonlinear dispersive waves: A discussion centered around the Korteweg-de Vries equation, *SIAM Rev.* **14**, 582 – 643.

[69] T. Kawahara (1975), Derivative-expansion method for nonlinear waves on a liquid layer of slowly varying depth, *J. Phys. Soc. Japan* **38**, 1200 – 1206.

[70] T. S. Kivshar and B. A. Malomed (1989), Dynamics of solitons in nearly integrable systems, *Rev. Mod. Phys.* **61**, 763 – 915.

[71] D. J. Korteweg and G. de Vries (1895), On the change of form of long waves advancing in a rectangular channel, and on a new type of long stationary waves, *Philos. Mag.* **39**, 422 – 443.

[72] M. D. Kruskal (1974), The Korteweg-de Vries equation and related evolution equations, *Lectures in Applied Mathematics* **Vol. 15**, American Mathematical Society.

[73] K. Pen-Yu and J. M. Sanz-Serna (1981), Convergence of methods for the numerical solution of the Korteweg-de Vries equation, *IMA J. Num. Anal.* **1**, 215 – 221.

[74] P. A. Lagerstrom and R. G. Casten (1972), Basic concepts underlying singular perturbation techniques, *SIAM Rev.*, **14**, 63 – 120.

[75] J. D. Lambert (1973), *Computational Methods in Ordinary Differential Equations*, Wiley.

[76] C. Lanczos (1961), *Linear Differential Operators*, D. Van Nostrand Company Limited.

[77] C. L. Lee and C. T. Lee (1997), A higher order method of multiple scales (1997), *J. Sound Vibration* **202**, 284 – 287.

[78] H. Levy and F. Lessman (1992), *Finite Difference Equations*, Dover.

[79] J. C. López-Marcos and J. M. Sanz-Serna (1988), Stability and convergence in numerical analysis III: Linear investigation of nonlinear stability, *IMA J. Num. Anal.* **8**, 71 – 84.

[80] E. Maré and S. W. Schoombie (1996), A review of multiple scales techniques used for perturbation analysis of numerical approximation methods, Differential Equations and Chaos, Ed. N. H. Ibragimov et. al., New Age International Publishers.

[81] M. F. Maritz (1990), *A Study of the Spurious Solutions of Some Nonlinear Partial Difference Equations of Evolution*, Ph.D. Thesis, University of the Orange Free State.

[82] M. F. Maritz and S. W. Schoombie (1987), Parasitic waves and solitons in the numerical solution of the Korteweg-de Vries and modified Korteweg-de Vries equation, *J. Comp. Phys.* **73**, 244 – 266.

[83] D. P. Mason (1992), *Singular Perturbation Methods for Initial Value Problems*, Lecture Notes, Centre for Nonlinear Studies and Department of Computational and Applied Mathematics, University of the Witwatersrand.

[84] R. E. Mickens (1987), Periodic solutions of second-order nonlinear difference equations containing a small parameter – IV. Multi-discrete time method, *J. Franklin Inst.* **324**, 263 – 271.

[85] R. E. Mickens (1994), *Nonstandard Difference Models of Differential Equations*, World Scientific Publishers.

[86] L. M. Milne-Thomson (1951), *The Calculus of Finite Differences*, Macmillan and Co., London.

[87] A. R. Mitchell and D. F. Griffiths (1980), *The Finite Difference Method in Partial Difference Equations*, Wiley, New York.

[88] A. R. Mitchell and D. F. Griffiths (1980), Beyond the linearized stability limit in nonlinear problems, *Numerical Analysis*, Pitman Research Notes in Mathematics **140**, *eds. D. F. Griffiths and G. A. Watson*, Longman Scientific, New York.

[89] D. W. Moore (1983), Resonances introduced by discretization, *IMA J. Appl. Math.* **31**, 1 – 11.

[90] G. M. Moremedi and D. P. Mason (1994), Analysis of the limit cycle of a generalized van der Pol equation by a time transformation method, *Q. Math.* **17**, 349 – 380.

[91] R. M. Muira (1976), The Korteweg-de Vries equation: A survey of results, *SIAM Rev.* **18**, 412 – 459.

[92] J. Murdock and L. Wang (1996), Validity of the multiple scale method for very long intervals, *Z. Angew. Math. Phys.* **47**, 760 – 789.

[93] A. H. Nayfeh (1973), *Perturbation Methods*, Wiley, New York.

[94] A. H. Nayfeh (1981), *Introduction to Perturbation Techniques*, John Wiley & Sons, New York.

[95] A. H. Nayfeh (1996), On the discretization of weakly nonlinear spatially continuous systems, Differential Equations and Chaos, Ed. N. H. Ibragimov et. al., New Age International Publishers.

[96] A. C. Newell (1977), Finite amplitude instabilities of partial difference equations, *SIAM J. Appl. Math.* **33**, 133 – 160.

[97] F. Z. Nouri and D. M. Sloan (1989), A comparison of Fourier pseudospectral methods for the solution of the Korteweg-de Vries equation, *J. Comp. Phys.* **83**, 324 – 344.

[98] T. S. Parker and L. O. Chua (1989), *Practical Numerical Algorithms for Chaotic Systems*, Springer Verlag.

[99] C. E. Pearson (1974), *Handbook of Applied Mathematics*, Van Nostrand Reinhold.

[100] D. H. Peregrine (1966), Calculations on the development of an undular bore, *J. Fluid Mech.* **25**, 321 – 330.

[101] R. Redheffer and D. Port (1991), *Differential Equations*, Jones and Bartlett Publishers, Boston.

[102] J. R. Rice (1983), *Numerical Methods, Software, and Analysis: IMSL Reference Edition*, McGraw-Hill Book Company.

[103] R. D. Richtmyer and K. W. Morton (1967), *Difference Methods for Initial Value Problems*, Wiley, New York.

[104] P. J. Roache (1985), *Computational Fluid Dynamics*, Hermosa Publishers.

[105] J. M. Sanz-Serna (1982), An explicit finite difference scheme with exact conservation properties, *J. Comp. Phys.* **47**, 199 – 210.

[106] J. M. Sanz-Serna (1985), Studies in numerical nonlinear instability I. Why do Leapfrog schemes go unstable?, *SIAM J. Sci. Stat. Comput.* **6**, 923 – 938.

[107] J. M. Sanz-Serna and I. Christie (1981), Petrov-Galerkin methods for nonlinear dispersive waves, *J. Comp. Phys.* **39**, 94 – 102.

[108] R. Schreiber and H. B. Keller (1983), Spurious solutions in driven cavity calculations, *J. Comp. Phys.* **49**, 165 – 172.

[109] S. W. Schoombie (1982), Spline Petrov-Galerkin methods for the numerical solution of the Korteweg-de Vries equation, *IMA J. Num. Anal.* **2**, 95 – 109.

[110] S. W. Schoombie (1987), A discrete multiple scales analysis of a discrete version of the Korteweg-de Vries equation, University of Dundee, Report NA/109.

[111] S. W. Schoombie (1992), A discrete multiple scales analysis of a discrete version of the Korteweg-de Vries equation, *J. Comp. Phys.* **101** (1992), 55 – 70.

[112] S. W. Schoombie and E. Maré (1996), A generalized discrete multiple scales analysis technique, *Numerical Analysis: A R Mitchell 75th Birthday Volume*, ed. D. F. Griffiths and G. A. Watson, World Scientific Publishing, Singapore.

[113] S. Sei and W. Symes (1995), Dispersion analysis of numerical wave propagation and its computational consequences, *J. Scientific Computing* **10**, 1 – 27.

[114] L. F. Shampine (1986), Conservation laws and the numerical solution of ODEs, *Comp. Maths. Appls.* **12B**, 1287 - 1296.

[115] D. M. Sloan (1988), On modulational instabilities in discretisations of the Korteweg-de Vries equation, *J. Comp. Phys.* **79**, 167 - 183.

[116] D. M. Sloan (1992), Fourier pseudospectral solution of the regularised long wave equation, *J. Comp. Appl. Math.* **36**, 159 - 179.

[117] G. D. Smith (1978), *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, Second ed., Clarendon Press.

[118] A. Solomonoff and E. Turkel (1989), Global properties of pseudospectral methods, *J. Comp. Phys.* **81**, 239 - 276.

[119] G. Strang (1968), On the construction and comparison of difference schemes, *SIAM J. Numer. Anal.* **5**, 506 - 517.

[120] A. Stuart (1989), Nonlinear instability in dissipative finite difference schemes, *SIAM Rev.* **31**, 191 - 220.

[121] J. T. Stuart and R. C. DiPrima (1978), The Eckhaus and Benjamin-Feir resonance mechanisms, *Proc. R. Soc. London* **A362**, 27 - 41.

[122] J. C. Strikwerda (1989), *Finite Difference Schemes and Partial Differential Equations*, Wadsworth & Brooks/Cole, Pacific Grove.

[123] E. Tadmor (1987), Stability analysis of finite difference, pseudospectral and Fourier - Galerkin approximations for time - dependent problems, *SIAM Rev.* **29**, 525 - 555.

[124] T. R. Taha and M. J. Ablowitz (1984), Analytical and numerical aspects of certain nonlinear evolution equations. I. Analytical, *J. Comp. Phys.* **55**, 192 - 202.

[125] T. R. Taha and M. J. Ablowitz (1984), Analytical and numerical aspects of certain nonlinear evolution equations. II. Numerical, Nonlinear Schrödinger Equation, *J. Comp. Phys.* **55**, 203 - 230.

[126] T. R. Taha and M. J. Ablowitz (1984), Analytical and numerical aspects of certain nonlinear evolution equations. III. Numerical, Korteweg-de Vries equation, *J. Comp. Phys.* **55**, 231 - 253.

[127] E. R. Tracy, J. W. Larson, A. R. Osborne and L. Bergamasco (1988), On the nonlinear Schrödinger limit of the Korteweg-de Vries equation, *Physica D* **32**, 83 - 106.

[128] L. N. Trefethen (1982), Group velocity in finite difference schemes, *SIAM Rev.* **24**, 113 – 136.

[129] L. N. Trefethen (1986), Dispersion, dissipation and stability, *Numerical Analysis*, Pitman Research Notes in Mathematics, **140**, Eds. D. F. Griffiths and G. A. Watson, Longman Scientific and Technical, Longman Scientific, New York.

[130] L. N. Trefethen (1996), *Finite Difference and Spectral Methods for Ordinary and Partial Differential Equations*, Unpublished text.

[131] S. Ushiki (1982), Central difference scheme and chaos, *Physica D* **4**, 407 – 424.

[132] J. V. Uspensky (1948), *Theory of Equations*, McGraw-Hill Book Company.

[133] F. Vadillo and J. M. Sanz-Serna (1986), Studies in numerical nonlinear instability II. A new look at $U_t + UU_x = 0$, *J. Comp. Phys.* **66**, 225 – 238.

[134] F. P. H. van Beckum and E. van Groesen (1987), Discretizations conserving energy and other constants of the motion, *Proc. ICIAM*, Eds. A. H. van der Burgh and R. M. M. Mattheij, CWI Tract.

[135] B. Van der Pol (1922), On a type of oscillation hysteresis in a simple triode generator, *Philos. Mag.* **43**, 177 – 193.

[136] A. C. Vliegenthart (1971), On finite difference methods for the Korteweg-de Vries equation, *J. Eng. Math.* **5**, 137 – 155.

[137] R. Wait (1981), Evolutionary partial differential equations *Numerical Solution of Nonlinear Problems ed. C. T. H. Baker and C. Phillips*, Oxford Science Publications, Clarendon Press.

[138] J. A. C. Weideman (1986), *Computation of Instability and Recurrence Phenomena in the Nonlinear Schrödinger Equation*, Ph.D. Thesis, University of the Orange Free State.

[139] G. B. Whitham (1974), *Linear and Nonlinear Waves*, Wiley, New York.

[140] G. B. Whitham (1984), Comments on periodic waves and solitons, *IMA J. Appl. Math.* **32**, 353 – 366.

[141] W. L. Wood (1990), *Practical Time Stepping Schemes*, Oxford Applied Mathematics and Computing Science Series, Clarendon Press, Oxford.

[142] H. C. Yee, P. K. Sweby and D. F. Griffiths (1991), Dynamical approach study of spurious steady-state numerical solutions of nonlinear differential equations. 1. The dynamics of time discretization and its implications for algorithm development in computational fluid dynamics, *J. Comp. Phys.* **97**, 249 – 310.

[143] N. J. Zabusky (1981), Computational synergetics and mathematical innovation, *J. Comp. Phys.* **43**, 195 – 249.

[144] N. J. Zabusky and M. D. Kruskal (1965), Interaction of solitons in a collisionless plasma and the recurrence of initial states, *Phys. Rev. Lett.* **15**, 240 – 243.

[145] V. E. Zakharov and E. A. Kusnetsov (1986), Multi-scale expansions in the theory of systems integrable by the Inverse Scattering Transform, *Physica D* **18**, 455 – 463.

[146] E. Zauderer (1983), *Partial Differential Equations of Applied Mathematics*, Wiley Interscience.

[147] H. Zhen and S. Long-jun (1994), A better difference scheme with four near-conserved quantities for the KdV equation, *J. Comp. Math.* **12**, 224 – 234.

Summary

Keywords: Perturbation techniques, Multiple scales, Korteweg-de Vries equation, Regularized Long Wave equation, Van der Pol equation, Finite difference approximation, Numerical solution, Discrete multiple scales.

Perturbation techniques for the solution of differential equations form an essential ingredient of the tools of mathematics as applied to physics, engineering, finance and other areas of applied mathematics. A natural extension would be to seek perturbation-type solutions for discrete approximations of differential equations.

The main objective of the research project is to develop a perturbation technique for the solution of discrete equations.

We discuss the well-known method of multiple scales and show its use for the solution of the Korteweg-de Vries (KdV), Regularized Long Wave (RLW) and Van der Pol equations. In particular, for the KdV and RLW equations the analysis shows that the envelopes of modulated waves are governed by the nonlinear Schrödinger equation. We present a variation of the multiple scales technique which presents an ideal framework from which we devise a discrete multiple scales analysis methodology.

We discuss a discrete multiple scales methodology derived by Schoombie [111], as applied to the Zabusky-Kruskal approximation of the KdV equation. This discrete multiple scales analysis methodology is generalized and applied to the solution of a generalized finite difference approximation of the KdV equation. We show the consistency of the method with the continuous analysis as the discretization parameters tend to zero.

The discrete multiple scales technique is a powerful tool for the examination of modulational properties of the KdV equation. In the case of certain modes of the carrier wave, the discrete multiple scales analysis breaks down, indicating that the numerical solution deviates in behavior from that of the KdV equation. Several numerical experiments are performed to examine the spurious behavior for different orders of

approximation. The spurious behavior, predicted theoretically, is shown to be present experimentally, independent of temporal discretization.

We also detail some comparisons of central difference solutions of different orders of approximation to the KdV equation. The results show a clear benefit of higher order central differences relative to lower order methods. The benefit of the central difference methodology would also extend to more general regions over which we would solve partial differential equations.

We also show that the method of multiple scales can provide an adequate explanation for spurious behavior in a difference scheme for the Van der Pol equation.

Opsomming

Perturbasie tegnieke vorm 'n integrale deel van die gereedskap van wiskundige tegnieke om differensiaal vergelykings op te los in fisika, ingenieurswese, finansiële en verwante areas in toegepaste wiskunde. Dit is gevolglik 'n natuurlike uitbreiding om perturbasie oplossings te soek vir die numeriese benaderings van differensiaal vergelykings.

Die hoofdoel van dié navorsingsprojek is om perturbasie tegnieke te vind vir die oplossing van diskrete vergelykings.

Ons bespreek die bekende veelvuldige skale tegniek en toon die gebruik daarvan aan vir die oplossing van die Korteweg-de Vries (KdV), RLW en Van der Pol vergelykings. Vir die KdV en RLW vergelykings is 'n gevolg van die analise dat die omhulsel van gemoduleerde golwe beheer word deur die nie-lineêre Schrödinger vergelyking. Ons bespreek 'n spesifieke veelvuldige skale tegniek wat 'n ideale raamwerk bied om 'n diskrete tegniek te skep.

Ons ondersoek 'n diskrete veelvuldige skale tegniek soos deur Schoombie [111] ontwikkel en toegepas op die Zabusky-Kruskal benadering van die KdV vergelyking. Die tegniek word veralgemeen en toegepas op 'n algemene sentraal verskil benadering van die KdV vergelyking. Ons toon aan dat die diskrete metode konsistent is met die kontinue geval as die diskretiserings parameters na nul neig.

Die diskrete skale tegniek is 'n geskikte tegniek vir die ondersoek van modulasie eienskappe van die KdV vergelyking. Vir spesifieke modes van die draer golf word oplossings van die diskrete skale tegniek onwenslik wat aandui dat die numeriese oplossing wat ons ondersoek verskil van die oplossing van die KdV vergelyking. Verskeie numeriese eksperimente word uitgevoer om die vals oplossings te ondersoek. Die vals oplossings, soos teoreties voorspel, word eksperimenteel aangetoon, onafhanklik van die diskretisasie tegniek in tyd.

Ons benadruk ook oplossings van sentraal verskil benaderings met verskillende ordes

van akkuraatheid vir die KdV vergelyking. Die resultate toon 'n duidelike voordeel aan van hoër orde metodes teenoor laer orde metodes. Die voordeel van die sentraal verskil vergelykings is dat ons dit op 'n veralgemeende gebied kan gebruik vir die oplossing van parsiële differensiaal vergelykings.

Ons beskou ook 'n eindige verskil benadering van die Van der Pol vergelyking en toon aan dat die diskrete skale tegniek 'n bevredigende verduideliking bied vir vals oplossings veroorsaak deur die spesifieke metode.