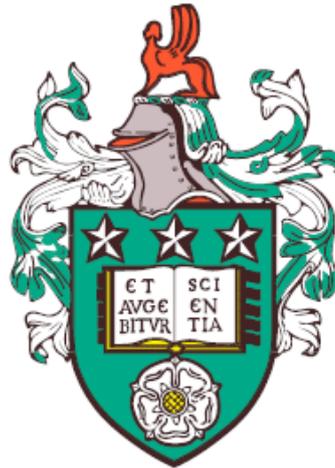


Numerical Methods as Applied to
Linear and Nonlinear Wave
Equations with Solutions having
Different Levels of Smoothness



Laya Yaqub Haweil

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

Department of Applied Mathematics

University of Leeds

May 2023

The candidate confirms that the work submitted is her own and that appropriate credit has been given where reference has been made to the work of others.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

©2023 The University of Leeds and Laya Yaqub Haweil

The right of Laya Yaqub Haweil to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988.

*For my family
with love and appreciation*

Acknowledgements

First and foremost, I would like to give All Glory, Praise and Thanks to my God and the Rock of my life, the Lord **Jesus Christ**, whom I love everlastingly; without Him, I could never have accomplished this challenging PhD journey. As the Bible says “But the Lord stood with me and strengthened me” (2 Timothy 4:17).

I would like to express my sincere and heartfelt gratitude to my supervisors **Dr. Jitse Niesen** and **Dr. Evy Kersalé**, who have been with me since the start of my PhD, for their constant guidance, patience and assistance through their knowledge and expertise. I would also like to extend very special thanks to my supervisor **Professor Rainer Hollerbach**, who joined the supervision team lately. His extensive knowledge, encouragement, and support have been a bright light on my dark path, leading me at times when I had lost all hope academically– thank you for everything. I am also grateful to my examiners **Professor Andrew Hazel** and **Professor Steven Tobias** for examining my thesis and the useful comments that came from my viva.

I would like to thank the postgraduates and lecturers in the School of Mathematics for the stimulating conversations, advice and support that kept me going, especially during the pandemic and working from home.

Being alone in the UK for many years with my family in Iraq, I am appreciative of all the kind people who offered me assistance in many general aspects of life. I am also incredibly grateful for the healthcare (physical and mental) I had from Leeds Student Medical Practice, Leeds Hospitals, Leeds University Services, etc.

Furthermore, I am forever indebted to my family for their unconditional love, as well as for their support, encouragement and inspiration throughout my academic journey, from primary school to this level. I would like to express my most sincere thanks to my lovely mother **Khezimy** and my oldest brother **Yalda** for always going out of their way for us all. I love you all endlessly. At this time, I think of, and give thanks to, my late father **Yaqub**, who constantly pushed me to get a higher education.

This project was carried out with the program of Human Capacity Development Programme (HCDP), to whom I thank. I am also grateful for the funding received from the School of Mathematics/Leeds University for one academic year 2019-2020.

Abstract

A variety of numerical methods are applied to solving the wave equations $u_{tt} = u_{xx}$ and $u_{tt} = u_{xx} - u^3$ on the bounded interval $x = [-1, 1]$, with Dirichlet boundary conditions $u = 0$ at the endpoints, and initial conditions $u = f(x)$ and $u_t = 0$. For the linear equation the d'Alembert solution on an infinite interval can be adapted to yield an analytical solution also on the bounded interval. This d'Alembert solution shows that discontinuities can develop in the second (or higher even) derivatives, even if the initial condition $f(x)$ itself is infinitely differentiable. The exact solution of our nonlinear wave equation is unknown, but the discontinuities presented in the linear equation are the same for this nonlinear equation as they share the same characteristics. On the other hand, this nonlinear equation also has an analytic result, namely energy conservation. In short, the exact solution of the linear/nonlinear problems have these different properties:

- 1: The solution is infinitely smooth at some special times, $t = 0, 2, 4, 6, 8, \text{etc.}$, only.
- 2: The energy is conserved, thus, the overall solution amplitude remains a constant.

A major focus in this thesis will be on how well following various numerical methods cope with three different initial conditions carefully chosen to have different levels of smoothness in their later evolution.

The numerical methods used are the spectral methods (SM) and the finite element methods (FEM). Many time stepping methods are studied: Euler's, modified Euler, implicit Euler and trapezoidal methods, as well as the exponential methods ETD1 and ETD2RK. The results of the SM show that the ETD methods for the linear problem recover the smoothness of the solution at the specified times in an amazing way far more accurately than the trapezoidal method (which does it down to $\mathcal{O}(h^2)$ only). Moreover, the ETD methods preserve the energy exactly. On the other hand, the EDT2RK method for the nonlinear problem recovers the smoothness of the solution at these times more faithfully than the trapezoidal method (similarly down to $\mathcal{O}(h^2)$ only). However, the energy decreases for EDT2RK, so the overall solution amplitude decays to zero, but on an extremely long time scale. Meanwhile, the trapezoidal rule almost preserves energy, thus, the overall solution amplitude remains a constant. In contrast to the SM, nonsmoothness does not have major impacts for the FEM.

Contents

1	Introduction and background	1
1.1	Motivation of the thesis	2
1.2	The classification of PDEs	5
1.3	The method of characteristics	6
1.4	Mathematical formulation	8
1.5	Numerical methods for discretizing partial differential equations . .	10
1.5.1	Finite difference methods (FDM)	10
1.5.2	Finite element methods (FEM)	11
1.5.3	Spectral method (SM)	12
1.5.4	Comparison of the previous three numerical methods	13
1.6	Published mathematical literature	14
1.7	Time Stepping	19
1.7.1	Euler's method	20
1.7.2	The modified Euler method	20
1.7.3	The implicit Euler method	21
1.7.4	The trapezoidal method	22
1.7.5	Theta methods	22
1.7.6	The implicit midpoint and Runge-Kutta methods	23

1.7.7	Exponential time differencing methods (ETD)	23
1.8	Stability of PDEs	24
1.9	D'Alembert solution of the wave equation	25
1.10	Separation of variables	39
1.11	Energy conservation	42
1.12	Detailed Motivation	43
1.13	Outline of the thesis	47
2	General concepts	49
2.1	Truncation errors	50
2.1.1	Global truncation error	50
2.2	Absolute stability region	51
2.3	Stiff differential equations	53
2.4	Hamiltonian ODEs	53
2.5	Symplectic maps	55
3	Linear PDE with Dirichlet boundary conditions	57
3.1	Spectral approximation	59
3.1.1	Chebyshev polynomials	60
3.1.2	Discrete Chebyshev transform	62
3.1.3	Spectral convergence	65
3.1.4	Differentiation	66
3.2	Solving the wave equation	69
3.3	The pseudospectral method	70
3.3.1	Spatial discretization	70

3.3.2	Time stepping	72
3.4	Numerical results	74
3.4.1	Convergence of Euler's method	76
3.4.2	Convergence of the modified Euler method	77
3.4.3	Convergence of the implicit Euler method	79
3.4.4	Convergence of the trapezoidal method	79
3.4.5	Stability of Euler's method	91
3.4.6	Stability of the modified Euler method	92
3.4.7	Stability of the implicit Euler and the trapezoidal methods	93
3.5	The collocation method	96
3.5.1	Spatial discretization	96
3.5.2	Time stepping	97
3.6	Numerical results	98
3.6.1	Convergence of Euler's and the trapezoidal methods	98
3.6.2	Stability of Euler's, the implicit Euler and the trapezoidal methods	98
3.7	The tau method	99
3.7.1	Spatial discretization	99
3.7.2	Time stepping	100
3.8	Numerical results	103
3.8.1	Convergence of Euler's and the trapezoidal methods	103
3.8.2	Stability of Euler's, the implicit Euler and the trapezoidal methods	104
3.9	The Galerkin method	105
3.9.1	Spatial discretization	106

3.9.2	Time stepping	109
3.10	Numerical results	110
3.10.1	Convergence of Euler's and the trapezoidal methods	110
3.10.2	Stability of Euler's, the implicit Euler and the trapezoidal methods	110
3.11	Summary and discussion	111
4	Nonlinear wave equation	115
4.1	Energy preservation for the exact solution	116
4.2	The pseudospectral method	119
4.2.1	Spatial discretization	119
4.2.2	Time stepping	121
4.2.3	Solving the nonlinear equation using Newton's method	122
4.3	Numerical results	123
4.3.1	Convergence of Euler's method	123
4.3.2	Convergence of the modified Euler method	125
4.3.3	Convergence of the implicit Euler method	126
4.3.4	Convergence of the trapezoidal method	126
4.3.5	Stability for Euler's and the modified Euler methods	134
4.3.6	Stability for the implicit Euler and the trapezoidal methods	135
4.4	Hamiltonian structure	141
4.4.1	Symplecticity of the numerical schemes	142
4.4.2	Symplecticity of Euler's and the implicit Euler methods	142
4.4.3	Symplecticity of the trapezoidal method	143
4.5	Summary and discussion	151

5	Exponential integrators	155
5.1	Derivation of exponential integrators	156
5.1.1	Exponential time differencing method (ETD1)	156
5.1.2	Exponential time differencing method with Runge-Kutta time stepping	157
5.2	Solving the linear wave equation	157
5.2.1	Convergence of the ETD1 method	158
5.2.2	Symplecticity and stability of the ETD1 method	167
5.3	Solving the nonlinear wave equation	167
5.3.1	Convergence of the ETD1 and ETD2RK methods	168
5.3.2	Stability of the ETD1 and ETD2RK methods	176
5.3.3	Symplecticity of the ETD1 and ETD2RK methods	177
5.4	Summary and discussion	179
6	Finite element methods	183
6.1	Finite element methods	184
6.1.1	Linear basis functions	184
6.1.2	Spatial discretization and weak formulation	185
6.1.3	Time stepping	190
6.1.4	Solving the nonlinear equation using Newton's method	191
6.2	Numerical results	193
6.2.1	Convergence of solution	193
6.2.2	Stability for Euler's and the trapezoidal methods	196
6.2.3	Symplecticity of the FEM for Euler and trapezoidal methods	196
6.3	Summary and discussion	198

7	General conclusion	201
7.1	Summary and contributions	203
7.2	Future work	211

List of Figures

- 1.1 The characteristics propagating from the interval $[-1, 1]$ for the finite string problem (left panel). Extending the characteristics throughout the domain strip $[-1, 1]$ (right panel). 3
- 1.2 The characteristics propagating from the interval $[-1, 1]$ for the finite string problem (1.47). The slope of the lines is ± 1 as $c = 1$ 27
- 1.3 The function $F(x)$, (1.51), in infinite domain, but here we take a part of the domain. It shows how the original function $f(x)$ is extended to the whole domain. The figures show that the function $f(x)$ is in the interval $[-1, 1]$, the red line, then it is extended oddly through the boundary point -1 to the interval $[-3, -1]$, the black line, the right panel show this clearly, similarly $f(x)$ is extended oddly from the boundary point 1 to the interval $[1, 3]$, the green line. Here we consider $f(x) = e^{-\sin(\pi(x+1))} - 1$ 28
- 1.4 The d'Alembert solution (1.46), $u(x, t)$, the black line. The time used here is $t = 0.2$. This solution is half of the initial condition $F(x)$, (1.51), which is the blue dashed line, shifted to the left, $F(x + t)$ the cyan dashed line, and half shifted to the right, $F(x - t)$ the red dashed line, and then adding them up. The left panel is the solution on the infinite domain. The right panel is the solution on one period $[-3, 1]$ 29
- 1.5 Extending the characteristics throughout the domain strip $[-1, 1]$ 30

-
- 1.6 The contour of the second derivative of the exact d'Alembert solution (1.46), $\frac{\partial^2 u}{\partial x^2}(x, t)$, panel (c), and its components (odd component, $\frac{\partial^2 u}{\partial x^2}(x, t)_{odd}$, panel (a), and even component, $\frac{\partial^2 u}{\partial x^2}(x, t)_{even}$, panel (b)) for the linear wave equation (1.47). The initial condition used is the C^1 IC (1.55). 36
- 1.7 The function $H(x)$ in infinite domain, but here we take a part of the domain. It shows that the original function $h(x)$ is extended to the whole domain as even-type extension. The figure shows that the function $h(x)$ is in the interval $[-1, 1]$, the red line, then is extended from the boundary point -1 to the interval $[-3, -1]$, the black dashed line, similarly $h(x)$ is extended from the boundary point 1 to the interval $[1, 3]$, the green dashed line. 38
- 3.1 The error $\|u(x) - I_N u(x)\|_{L^\infty(-1,1)}$ against the order of the truncation N , showing the exponential convergence of the function (3.32) (left panel), using *semilogy* scale, and algebraic convergence of (3.33) (right panel), using *loglog* scale. 66
- 3.2 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the wave equation (3.52) solved using the pseudospectral method with Euler's method for the C^∞ IC (3.81). The step size is $h = 10^{-5}$ and the error is evaluated at $t = 0.9$. Here u is the exact solution and $u_{num}(x, t)$ is the numerical solution. 77
- 3.3 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the wave equation (3.52) using the pseudospectral method with the modified Euler method for the C^∞ IC (3.81). The step size is $h = 10^{-5}$ and the error is evaluated at $t = 0.9$ 78
- 3.4 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the inverse step size $1/h$ for the wave equation (3.52) solved using the pseudospectral method for the C^∞ initial condition (3.81). The left panel with Euler's method and the right panel with the modified Euler method. The Chebyshev series is truncated at $N = 20$ and the error is evaluated at $t = 0.9$ 78

- 3.5 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the wave equation (3.52) solved using the pseudospectral method with the trapezoidal method for the C^∞ IC (3.81), the C^1 IC (3.82) and the C^3 IC (3.83) (a, b and c respectively). The error is evaluated at $h = 10^{-3}$ and $t = 0.9$ using $nx = 100N$. The blue straight line indicates order $\frac{1}{N^3}$ for the IC (3.82) and order $\frac{1}{N^5}$ for the IC (3.83). 80
- 3.6 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) for the C^∞ IC (3.81) solved using the pseudospectral method with trapezoidal method at $t = 0.9$. The coefficients are evaluated for $h = 10^{-3}$ and $N = 100$. The blue dots are the even coefficients component and the red dots are the odd coefficients component. Similar plot is obtained for $t = 1$ and $t = 4$. 83
- 3.7 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) for the C^3 IC (3.83) solved using the pseudospectral method with the trapezoidal method at $t = 0.9, 1, 4$ (a, b, c respectively). The coefficients are evaluated for $h = 10^{-3}$ and $N = 2000$. The black straight line indicates order $\frac{1}{k^5}$, more clearly for $t = 0.9$ and $t = 1$ 83
- 3.8 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with the trapezoidal method at $t = 0.04, 0.9, 1, 2, 4, 5.3$ (a, b, c, d,e ,f panels respectively). The step size is $h = 10^{-3}$ with $N = 2000$ for the C^1 IC (3.82). The blue dots are the even coefficients component and the red dots are the odd coefficients component. The black straight line indicates order $\frac{1}{k^3}$ 84
- 3.9 The even Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with the trapezoidal method for three step sizes $h = 10^{-2}, 10^{-3}, 10^{-4}$ (magenta, blue, cyan plots respectively). This is evaluated at $t = 0.16, 1, 2$ (a, b, c panels respectively) using the C^1 IC (3.82) with $N = 2000$ 85

- 3.10 The Chebyshev coefficients $|\hat{u}_k|$ of the exact d'Alembert solution against the mode k for the wave equation (3.52) for the C^∞ IC (3.81) at $t = 0.9$. The coefficients are evaluated for $h = 10^{-3}$ and $N = 100$. Similar plot is obtained for $t = 1$ and $t = 4$ 88
- 3.11 The Chebyshev coefficients $|\hat{u}_k|$ of the exact d'Alembert solution against the mode k for the wave equation (3.52) for the C^3 IC (3.83) at $t = 0.9, 1, 4$ (a, b, c respectively). The coefficients are evaluated for $h = 10^{-3}$ and $N = 2000$. The black straight line indicates order $\frac{1}{k^3}$. 89
- 3.12 The Chebyshev coefficients $|\hat{u}_k|$ of the exact d'Alembert solution against the mode k for the wave equation (3.52) for the C^1 IC (3.82) at $t = 0.9, 1, 4$ (a, b, c respectively). The coefficients are evaluated for $h = 10^{-3}$ and $N = 2000$. The black straight line indicates order $\frac{1}{k^3}$. 89
- 3.13 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the time t for the wave equation (3.52) solved using the pseudospectral method for the C^∞ initial condition (3.81) with Euler's method, the left panel, and with the modified Euler method, the right panel. The Chebyshev series is truncated at $N = 10$ and the step size is $h = 0.01$ 92
- 3.14 The exact solution $u(x, t)$, blue line, and the numerical solution $u_{num}(x, t)$, red line, at $x = \frac{1}{2}$, for the wave equation (3.52) solved for $N = 10$. Using the pseudospectral method with the trapezoidal method (a, c panels), $\theta = 0.5$, and the implicit Euler method (b, d panels), $\theta = 1$, for the C^∞ initial condition (3.81). Using two time steps, $h = 10^{-1}$ (a, b panels) and $h = 10^{-3}$ (c, d panels). 94
- 3.15 The numerical solution, $u_{num}(x, t)$, at $x = \frac{1}{2}$, for the wave equation (3.52). Using the pseudospectral method with the trapezoidal method (left panel), $\theta = 0.5$, and the implicit Euler method (right panel), $\theta = 1$, for the C^1 initial condition (3.82), for $N = 10$ and $h = 10^{-3}$ 95
- 3.16 The contour of the numerical solution $u_{num}(x, t)$ for the wave equation (3.52), using the pseudospectral method with the trapezoidal method for the initial conditions, the C^∞ IC (3.81)(left panel) and the C^1 IC (3.82)(right panel), at $N = 100$ and $h = 10^{-2}$. 95

- 3.17 The numerical solution $u_{num}(x, t)$ for the wave equation (3.52), using the pseudospectral method with the trapezoidal method for the initial conditions, the C^∞ IC (3.81)(left panel) and the C^1 IC (3.82)(right panel), at $N = 100$, $h = 10^{-2}$ and for $t = 0 - 3$ 96
- 3.18 The exact solution $u(x, t)$, red line, and the numerical solution $u_{num}(x, t)$, blue line, at $x = \frac{1}{2}$, for the wave equation (3.52) solved using the tau method, with the C^∞ initial condition (3.81), $N = 10$ and the step size is $h = 10^{-3}$. Panel (a) shows the explicit Euler method, panel (b) shows the implicit Euler method, and panel (c) shows the trapezoidal method. 105
- 4.1 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the nonlinear wave equation (4.1), for the initial condition (3.81), $f(x) = \sin(\pi x)$, using Euler's method. The estimated error is evaluated at $t = 0.9$ with time step $h = 10^{-3}$ and $nx = 100N$. Here u_{num} is the numerical solution and u is the reference solution. 125
- 4.2 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the inverse step size $1/h$ for the nonlinear wave equation (4.1), for the initial condition (3.81) using Euler's method (left panel) and the modified Euler method (right panel). The estimated error is evaluated $t = 0.9$ using $N = 30$. The blue straight line indicates order one for Euler's method and order two for the modified Euler method. 126
- 4.3 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the nonlinear wave equation (4.1) using the pseudospectral method with the trapezoidal method for the C^∞ IC (3.81), C^1 IC (3.82) and C^3 IC (3.83) (a, b and c respectively). The error is evaluated at $t = 0.9$ using $h = 10^{-3}$ and $nx = 100N$. The blue straight line indicates order $\frac{1}{N^3}$ for the IC (3.82) and order $\frac{1}{N^5}$ for the IC (3.83). 127

- 4.4 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with trapezoidal method for the C^∞ IC (3.81) at $t = 0.9$. The coefficients are evaluated for $h = 10^{-3}$ and $N = 100$. Similar plot is obtained for $t = 1$ and $t = 4$. The blue dots are the even coefficients component and the red dots are the odd coefficients component. 130
- 4.5 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with trapezoidal method for the C^3 IC (3.83) at $t = 0.9, 1, 4$ (a, b, c respectively). The coefficients are evaluated for $h = 10^{-3}$ and $N = 2000$. The black straight line indicates order $\frac{1}{k^5}$ 130
- 4.6 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with the trapezoidal method at $t = 0.04, 0.9, 1, 2, 4, 5.3$ (a, b, c, d, e, f panels respectively). The step size is $h = 10^{-3}$ with $N = 2000$ for the C^1 IC (3.82). The blue dots are the even coefficients component and the red dots are the odd coefficients component. The black straight line indicates order $\frac{1}{k^3}$ 131
- 4.7 The numerical solutions $u_{num}(x, t)$ for the nonlinear wave equation (4.1) solved using Euler's method, at $x = \frac{1}{2}$, for two time steps, $h = 10^{-2}$ (red line), $h = 10^{-3}$ (dashed blue line). Applied for the C^∞ IC (3.81) and for $N = 10$ 135
- 4.8 The numerical solutions $u_{num}(x, t)$ for the nonlinear wave equation (4.1) using the implicit Euler method, at $x = \frac{1}{2}$, for two time steps, $h = 10^{-1}$ (red line), $h = 10^{-2}$ (dashed blue line) (a,b panels) and $h = 10^{-2}$ (red line), $h = 10^{-3}$ (dashed blue line) (c,d panels). Applied for the initial condition, (3.81), $f(x) = \sin(\pi x)$, (left panels) and for the initial condition (3.82), $f(x) = e^{\sin(\pi x)} - 1$, (right panels) and for $N = 10$ 137

- 4.9 The numerical solutions $u_{num}(x, t)$ for the nonlinear wave equation (4.1) using the trapezoidal method, at $x = \frac{1}{2}$, for two time steps, $h = 10^{-1}$ (red line), $h = 10^{-2}$ (dashed blue line) (a,b,c panels) and $h = 10^{-2}$ (red line), $h = 10^{-3}$ (dashed blue line) (d,e,f panels). Applied for the initial condition (3.81) (a,d panels), for the initial condition (3.82) (b,e panels) and for the initial condition (3.83) (c,f panels) and for $N = 10$ 138
- 4.10 The contour of the numerical solution $u_{num}(x, t)$ for the nonlinear wave equation (4.1), using the pseudospectral method with the trapezoidal method, at $N = 100$, $h = 10^{-2}$ and for $t = 100 - 108$ (left panels) and $t = 1492 - 1500$ (right panels). Using the initial condition (3.81)(panels a and b), the initial condition (3.82)(panels c and d) and the initial condition (3.83)(panels e and f). 139
- 4.11 The numerical solution $u_{num}(x, t)$ for the nonlinear wave equation (4.1), using the pseudospectral method with the trapezoidal method, at $N = 100$, $h = 10^{-2}$ and for $t = 100 - 104$ (left panels) and $t = 1496 - 1500$ (right panels). Using the initial condition (3.81)(panels a and b), the initial condition (3.82)(panels c and d) and the initial condition (3.83)(panels e and f). 140
- 4.12 The energy of the nonlinear wave equation (4.1) for the C^∞ IC (3.81), $f(x) = \sin \pi x$, using Euler's method (left panel), for $N = 10$ with time step $h = 10^{-2}$, and the implicit Euler method (right panel), for $N = 100$ with time step $h = 10^{-3}$ 143
- 4.13 The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1), solved using the pseudospectral method with the trapezoidal method for the C^∞ IC (3.81), with $N = 100$, $nx = 200$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a, b, and c respectively) for both short and long periods of time. Here E is the numerical energy and E_{exact} is the exact energy. . . . 147

- 4.14 The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1), solved using the pseudospectral method with the trapezoidal method for the C^1 IC (3.82), with $N = 1000$, $nx = 100N$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a, b, and c respectively) for both short and long periods of time. 148
- 4.15 The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1), solved using the pseudospectral method with the trapezoidal method for the C^3 IC (3.83), with $N = 100$, $nx = 100N$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a, b, and c respectively) for both short and long periods of time. 149
- 4.16 The energy error, $E - E_{exact}$, of the nonlinear wave equation (4.1) solved using the trapezoidal method with $nx = 200$. The C^1 IC (3.82) using $N = 1000$, $h = 10^{-3}$ and the C^3 IC (3.83), using $N = 100$, $h = 10^{-4}$ (panels a and b respectively). 150
- 4.17 The energy error, $E - E_{exact}$, of the nonlinear wave equation (4.1) solved using the trapezoidal method with $h = 10^{-4}$, $nx = 100N$ and $N = 100, 200, 400, 800$ (a, b, c and d respectively) using the C^1 IC (3.82). 151
- 5.1 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the linear wave equation (3.52), for the C^∞ IC (3.81) solved using the pseudospectral method with the ETD1 method. The exact error is evaluated at $h = 10^{-1}, 10^{-2}, 10^{-3}$ and $h = 0.9$ for $t = 0.9$ (blue, red, black and green respectively). Here u is the exact solution and u_{num} is the numerical solution. 159
- 5.2 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with ETD1 method at $t = 0.9, 1, 4$ (a, b, c panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^3 IC (3.83). The blue dots are the even coefficients component and the red dots are the odd coefficients component. The black straight line indicates order $\frac{1}{k^3}$ 163

- 5.3 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with ETD1 method at $t = 0.04, 0.9, 1, 2, 4, 6.3$ (a, b, c, d, e, f panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^1 IC (3.82). The blue dots are the even coefficients component and the red dots are the odd coefficients component. The black straight line indicates order $\frac{1}{k^3}$ 164
- 5.4 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with ETD1 method at times near $t = 2$, i.e., $t = 1.9, 1.99, 1.999$ (a, b, c panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^1 IC (3.82). 165
- 5.5 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the nonlinear wave equation (4.1), for the C^∞ IC (3.81), C^1 IC (3.82) and C^3 IC (3.83) (a, b and c respectively), solved using the pseudospectral method with the ETD2RK method. The estimated error is evaluated at time step $h = 10^{-3}$ and $t = 0.9$. The blue straight line shows order $\frac{1}{N^3}$ for the C^1 IC (3.82) and order $\frac{1}{N^5}$ for the C^3 IC (3.83). Here u_{num} is the numerical solution and for u we take the reference solution in Subsection 4.3.1. 169
- 5.6 The errors against the inverse step size $1/h$ for the nonlinear wave equation (4.1) solved using the pseudospectral method C^∞ IC (3.81). The left panel with the ETD1 method and the right panel with the ETD2RK method. The estimated error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, is evaluated at $t = 0.9$ and $N = 100$. The blue straight line shows order one for the ETD1 method and order two for the ETD2RK method. 169
- 5.7 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD2RK method at $t = 0.9, 1, 4$ (a, b, c panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^3 IC (3.83). The black straight line indicates order $\frac{1}{k^5}$ 172

- 5.8 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD2RK method at $t = 0.04, 0.9, 1, 2, 4, 6.3$ (a, b, c, d, e, f panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^1 IC (3.82). The black straight line indicates order $\frac{1}{k^3}$ 173
- 5.9 The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD2RK method at times near $t = 2$, i.e., $t = 1.9, 1.99, 1.999$ (a, b, c panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^1 IC (3.82). 174
- 5.10 The numerical solutions $u_{num}(x, t)$ for the nonlinear wave equation (4.1), at $x = \frac{1}{2}$, for $N = 10$ using two time steps, $h = 10^{-2}$ (red line), $h = 10^{-3}$ (dashed blue line). The solutions are computed using ETD1 method (left panel) and ETD2RK (right panel), applied for the C^1 IC (3.82). 177
- 5.11 The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD1 for $N = 100$, $nx = 100N$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a,b, and c respectively) and initial condition (3.82). Here E is the numerical energy and E_{exact} is the exact energy. 178
- 5.12 The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD2RK, for $N = 1000$, $nx = 100N$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a,b, and c respectively) and initial condition (3.82). 179

- 6.1 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the number of elements M for the nonlinear wave equation (4.1) solved using the FEM with the trapezoidal method for the C^∞ IC (3.81), the C^1 IC (3.82) and the C^3 IC (3.83) (a, b and c respectively). The error is evaluated at $h = 10^{-3}$ and $t = 0.9$. The red straight line indicates order two, $\frac{1}{M^2}$, for all the three ICs. Here u_{num} is the numerical solution and u is the reference solution in Subsection 4.3.1. 195
- 6.2 The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the inverse step size $1/h$ for the nonlinear wave equation (4.1) solved using the FEM method for the C^∞ initial condition (3.81). The left panel with Euler's method for $M = 100$ and the right panel with the trapezoidal method for $M = 1000$. The error is evaluated at time $t = 0.9$. The red straight line indicates order one for Euler's method and order two for the trapezoidal method. 195
- 6.3 The energy error, $E - E_{exact}$, of the FEM with Euler's rule applied to the nonlinear wave equation (4.1) for $M = 10$, $h = 10^{-2}$ and the C^1 IC (3.82). Here E is the numerical energy and E_{exact} is the exact energy. We get a similar plot for the energy difference $E - E_0$, where E_0 is the numerical energy at $t = 0$ 197
- 6.4 The energy difference, $E - E_0$, for the nonlinear wave equation (4.1) using the FEM with the trapezoidal rule, for $M = 100$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a,b, and c respectively) and the C^1 IC (3.82). 197
- 7.1 The characteristics of the problem in the domain strip $[-1, 1]$ 213

List of Tables

1.1	The classification of PDEs	5
-----	--------------------------------------	---

Chapter 1

Introduction and background

Many problems in applied science and engineering are modelled mathematically by time-dependent partial differential equations (PDEs). A partial differential equation is an equation stating a relationship between a function of two or more independent variables and the partial derivatives of this function with respect to these independent variables. For further information about PDEs see (Gockenbach, 2002; Hoffman & Frankel, 2001; Stavroulakis & Tersian, 2004; Strauss, 1992). The dependent variable u is used as a generic dependent variable throughout this thesis. In recent years, seeking exact solutions of the PDEs has been an important topic in the study of nonlinear physical phenomena. The reason for that is, it appears that these PDEs are mathematical models of complex phenomena arising in physics, chemistry, biology and engineering (Yomba, 2005). However, in most cases it is extremely difficult, if not impossible, to obtain exact solutions of these problems, therefore, numerical methods are used to provide accurate approximations. The discretization of the equation in space first, then in time, is a commonly used technique in the design of numerical solutions for PDEs (Boyd, 2001; Minchev, 2004). There exist many efficient and accurate methods for spatial discretization; in this thesis we use finite element and spectral methods.

This introductory chapter is divided into two main parts. The first is about the mathematical formulation of PDEs and numerical methods for discretizing PDEs in space and time, also a review of published mathematical literature. The second part is about the analytical solution of the linear wave problem and the level of nonsmoothness of solutions obtained using three infinitely smooth initial conditions,

ICs, on a finite interval, followed by the analytical concepts of the nonlinear wave problem.

1.1 Motivation of the thesis

The core topic of this thesis is essentially about how does information propagate along the characteristics for the wave equations $u_{tt} = u_{xx}$ and $u_{tt} = u_{xx} - u^3$, however, not on infinite domain but rather on a bounded interval $[-1, 1]$. In particular the effect of nonsmoothness of different levels emanating from the boundaries is studied.

The characteristics of the linear wave equation are defined in the literature and given in the left panel of Figure 1.1. In addition, the exact d'Alembert solution for the linear wave equation is well-known for the infinite domain. However, the exact solution for the nonlinear wave equation is unknown, but this equation shares the same characteristics as for the linear wave equation. Nevertheless, people have tried to find the solution of the nonlinear problem on the infinite domain, specifically for periodic boundary conditions (BCs) in x , but not on the finite bounded domain.

In the present thesis we consider a harder case; we take the same problems but put some boundaries on the infinite domain at $x = -1$ and $x = 1$, thus we get the bounded domain $[-1, 1]$, see the left panel of Figure 1.1. Using Dirichlet boundary conditions $u = 0$ at the endpoints, and initial conditions $u = f(x)$ and $u_t = 0$. Three ICs are considered, which are infinitely smooth on the bounded interval $[-1, 1]$.

The characteristics play a significant role in the finite domain. They are the same as for the infinite domain but they hit the boundaries and reflect off them, see the right panel of Figure 1.1. The reflection depends on the boundary conditions used; either in odd-type way or even-type for Dirichlet and Neumann BCs, respectively. In our problems what matters with us are the characteristics that originate in the corners, i.e., $x = -1$ and $x = 1$, not the characteristics that originate from all other interior points of the interval.

For the linear equation the d'Alembert solution on the infinite interval can be adapted to yield an analytical solution also on the bounded interval. This has consequences for the higher derivatives; this d'Alembert solution shows that

discontinuities can develop in the second (or higher even) derivatives for our Dirichlet BCs, even if the initial condition $f(x)$ itself is infinitely differentiable. This is due to the fact that this solution will have two interfaces at the boundaries $x = \pm 1$ as it will have two adjacent intervals at the boundaries. Thus, it turns out that the characteristics generate discontinuities in the higher solution derivatives that propagate along the characteristics which originally happen at $x = -1$ and $x = 1$ only. More precisely, these discontinuities will move along these characteristics; these discontinuities, which originally occurred only at the boundaries $x = \pm 1$, now will move into the interior at later times t . That is, the problem is triggered by the initial condition at the boundaries, then it propagates throughout the interior along the characteristics. Incredibly, we discover that there are some special times, $t = 0, 2, 4, 6, 8, \text{etc.}$, where the discontinuities go back to the end points, i.e., the discontinuities disappear again, thus the solution becomes infinitely smooth at these special times only. This is because the characteristics for $x = -1$ and $x = 1$ will both hit the boundaries at exactly the same times, the special times above, as shown in the right panel. Therefore, the discontinuities for both characteristics disappear at exactly the same times. That is, we start out with infinitely smooth ICs and at all the intermediate times there are the discontinuities inside our interval then at these special times the discontinuities have temporarily disappeared again because they hit the boundaries, so temporarily at these special times the solution will become infinitely smooth again.

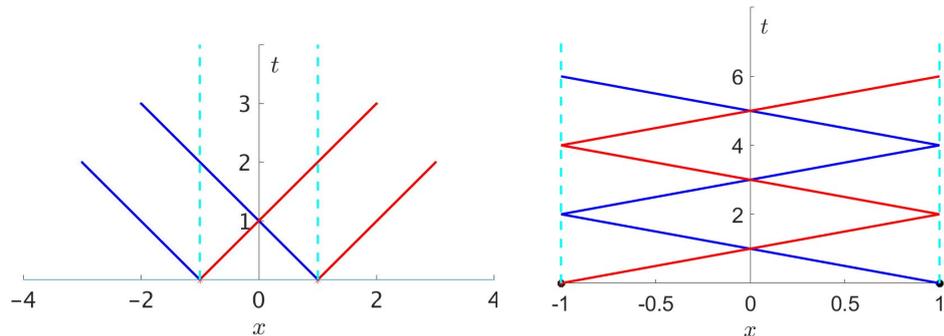


Figure 1.1: The characteristics propagating from the interval $[-1, 1]$ for the finite string problem (left panel). Extending the characteristics throughout the domain strip $[-1, 1]$ (right panel).

In our case the discontinuities of the solution originally happen at $x = -1$ and $x = 1$ only because our ICs are infinitely smooth on the interval $[-1, 1]$. While for the other weird cases when the IC is not smooth that will lead to have the discontinuities to happen at all the points of the interval not only at $x = -1$ and $x = 1$. In this case the characteristics that propagate from inside the interval $(-1, 1)$ will not have special times that the solution is smooth. The reason for this is that for any $x \in (-1, 1)$ the right characteristic $x + t$ will hit one boundary at a different time, say at $t = a$, than the left characteristic $x - t$ when will hit the other boundary, say at $t = b$, where a and b are constants. That means the discontinuities are still present and have not vanished, say at $t = a$, because the other characteristic has not yet reached the other boundary but is still in the interior at this time. Therefore, for such ICs the special times do not exist for the solution to become infinitely smooth, in contrast to our ICs case.

The exact d'Alembert solution for the linear wave equation on the finite bounded domain is already known in the literature, however, to the best of our knowledge, we are the first to illustrate this discontinuity problem that occurs for the finite d'Alembert solution.

Although the exact solution for the nonlinear wave equation is unknown, but the discontinuity aspects presented above for the linear wave equation are the same for this nonlinear wave equation, because the characteristics are the same as the right panel of Figure 1.1. In other words, the unknown exact solution will be nonsmooth in general but only at these specific times it will become temporarily infinitely smooth again as well.

Given that smoothness is a very important aspect in the entire field of numerical analysis, our interest is to see how well numerical methods cope with nonsmoothness that occur in the analytical solution; will they be so good as to recognise the special times, i.e., to recover a smooth solution; will they notice that at all or not.

All of these aspects are thoroughly investigated throughout the present thesis, beginning with this chapter.

1.2 The classification of PDEs

The classification of PDEs is most easily explained for a single second order PDE. The classification of higher order PDEs, large systems of PDEs, and PDEs having more than two independent variables is considerably more complicated.

One can start by classifying three types of equations under investigation. A partial differential equation of the form

$$Au_{xx} + Bu_{xy} + Cu_{yy} = f(x, y, u, u_x, u_y), \quad (1.1)$$

where A , B and C are constants or functions of x and y , is called quasilinear. There are three types of quasilinear equations:

Table 1.1: The classification of PDEs

$B^2 - 4AC$	Classification
Negative	Elliptic
Zero	Parabolic
Positive	Hyperbolic

As an example of hyperbolic equations, we consider the one-dimensional model for a vibrating string. The displacement $u(x, t)$ is governed by the wave equation,

$$\rho \frac{\partial^2 u}{\partial t^2}(x, t) = k \frac{\partial^2 u}{\partial x^2}(x, t) \quad \text{for } a < x < b \text{ and } t > 0, \quad (1.2)$$

where a , b , ρ and k are constants (see [Hoffman & Frankel, 2001](#); [Mathews *et al.*, 2004](#), for more details). Equation (1.2) can be solved subject to the initial conditions,

$$u(x, 0) = \varphi(x), \quad x \in [a, b], \quad (1.3)$$

and

$$\frac{\partial u}{\partial t}(x, 0) = \psi(x), \quad x \in [a, b], \quad (1.4)$$

where φ and ψ represent the initial displacement and velocity, respectively. Different boundary conditions can be considered, such as periodicity conditions or the following BCs,

$$u(a, t) = P_a(t) \quad \text{and} \quad u(b, t) = P_b(t), \quad t \in [0, \infty), \quad (1.5)$$

or

$$\frac{\partial u}{\partial x}(a, t) = P_a(t) \quad \text{and} \quad \frac{\partial u}{\partial x}(b, t) = P_b(t), \quad t \in [0, \infty), \quad (1.6)$$

where P_a and P_b are given functions. The boundary conditions in (1.5) are called Dirichlet BCs and in (1.6) are Neumann BCs. One can, moreover, mix Dirichlet and Neumann BCs; this is called mixed (Robin's) boundary conditions. In this study only Dirichlet BCs are employed. More details about initial and boundary conditions can be found in (Benmansour, 1993; Gockenbach, 2002).

1.3 The method of characteristics

The method of characteristics provides solutions to the PDEs. We begin by considering the following simple wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \quad (1.7)$$

where u represents the displacement of a vibrating string from its equilibrium position and c represents the speed of wave propagation. Equation (1.7) can be solved subject to two initial conditions, specifying the initial position $u(x, 0)$ and the initial velocity $u_t(x, 0)$,

$$u(x, 0) = F(x), \quad (1.8)$$

$$\frac{\partial u}{\partial t}(x, 0) = 0, \quad (1.9)$$

see (1.2)–(1.4).

The method of characteristics solves the PDE (1.7) by factoring it to two separate operators, as follows

$$\left(\frac{\partial}{\partial t} - c \frac{\partial}{\partial x} \right) \left(\frac{\partial}{\partial t} + c \frac{\partial}{\partial x} \right) u = 0. \quad (1.10)$$

First, we assume that

$$u_+ := \left(\frac{\partial u}{\partial t} - c \frac{\partial u}{\partial x} \right) = 0. \quad (1.11)$$

If we consider the solution of (1.11) along the curve $(x, t) = (X(t), t)$, then by chain rule we have

$$\frac{du}{dt} \left(X(t), t \right) = \partial_t u + \frac{dX}{dt} \partial_x u = \partial_x u \left(\frac{dX}{dt} + c \right), \quad (1.12)$$

because of our assumption.

Thus, if we choose the curve $(X(t), t)$ such that

$$\frac{dX}{dt} = -c, \quad (1.13)$$

then $\frac{du}{dt}(X(t), t) = 0$; in other words, the solution is constant along the curves.

Similarly, if we assume that

$$u_- := \left(\frac{\partial u}{\partial t} + c \frac{\partial u}{\partial x} \right) = 0, \quad (1.14)$$

then

$$\frac{dX}{dt} = +c. \quad (1.15)$$

So, equations (1.13) and (1.15) can be rewritten as

$$\frac{dX}{dt} = \pm c, \quad (1.16)$$

this is called **the equation of characteristics**.

If the speed is constant, $c = \text{constant}$, then these characteristics are nothing but the straight lines $x = X = \pm ct + b$, where b is a constant. Therefore, the solutions of the homogeneous linear equations in (1.11) and (1.14) are exactly the functions u that are constant on the integral curves $(X(t), t)$. This means that we have

$$u_{\pm}(x, t) = u_{\pm}(ct + b, t) = u_{\pm}(b, 0) = F_{\pm}(b) = F_{\pm}(x \mp ct). \quad (1.17)$$

Thus, the solution is

$$u_{\pm}(x, t) = F_{\pm}(x \mp ct). \quad (1.18)$$

In the general case, the more general operator (1.1) can be factored in the same way as (1.10) with purely real characteristics and the condition $B^2 - 4AC$ in the classification table 1.1 to be positive is all that will be required, i.e., this is true for all the hyperbolic problems. Noteworthy, the characteristics do not get affected by the term $f(x, y, u, u_x, u_y)$ in (1.1).

The material discussed above in this section is standard and can be found in many textbooks, such as (Rauch, 2012; Zauderer, 1983).

1.4 Mathematical formulation

In this section we are considering in more detail some different wave equations and various generalizations.

The most widespread and best known model for waves in linear non-dispersive media is the so-called linear wave equation. It has the form

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \quad (1.19)$$

see (1.2) and (1.7), where u represents the displacement of a vibrating string from its equilibrium position and c represents the speed of wave propagation. In this study, we are considering this equation but for $c = 1$ and the interval $[-1, 1]$ as follows,

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \quad \text{for } -1 < x < 1 \text{ and } t > 0, \quad (1.20)$$

see (1.47)–(1.50).

The most natural generalization of the wave equation is the **nonlinear Klein–Gordon equation**, which its general form is

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} - f(u), \quad (1.21)$$

also called the general **Klein–Gordon equation**. Equation (1.21) describes the vibration of a string that lies on an elastic foundation with nonlinear elastic forces. The elastic force density is described by the function $f(u)$ (Landa, 1996; Whitham, 1999, 2011). For $f(u) = au$, where a is a real constant, equation (1.21) is called the **linear Klein–Gordon equation** (Landa, 1996; Whitham, 1999, 2011).

The nonlinear Klein–Gordon equation appears with many types of nonlinearities. The Klein–Gordon equation plays a significant role in many scientific applications such as solid state physics, nonlinear optics and quantum field theory, see (Dehghan & Shokri, 2009; Wazwaz, 2008).

The Klein–Gordon equation was actually first found by Schrödinger, before he made the discovery of the equation that now bears his name. He rejected the Klein–Gordon equation because he could not make it fit the data (the equation does not take in account the spin of the electron). In 1926 several physicists, among them Klein, Fock, Schrödinger, and de Broglie, announced this equation as a candidate for a relativistic generalization of the usual Schrödinger equation.

In most of the early versions the Klein–Gordon equation was connected with the general theory of relativity. This origin background is taken from (Hepson *et al.*, 2016; Kragh, 1984; Rani & Kumar, 2016).

The choice of $f(u)$ in (1.21) can generate various important equations of mathematical physics. In particular, for the φ^4 Klein–Gordon equation, which is also called φ^4 equation,

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - mu + \varepsilon u^3, \quad (1.22)$$

see (Golmankhaneh & Baleanu, 2011; Merdan, 2014; Tamsir & Srivastava, 2016; Tariq & Akram, 2017; Ye & Zhang, 2010). This equation arises in quantum field theory with m denoting mass, $m > 0$, and ε is a coupling constant, $\varepsilon \in R$ (McOwen, 1996). And another example is the sine–Gordon equation,

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - \sin u, \quad (1.23)$$

found by Zabusky and Kruskal in 1965 (Bobenko & Kuksin, 1995; Grundland & Infeld, 1992; Landa, 1996; Scott, 2003; Scott *et al.*, 1999).

In this thesis the nonlinear Klein–Gordon equation (1.22) of the form

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3 \quad \text{for } -1 < x < 1 \quad \text{and } t > 0, \quad (1.24)$$

is considered. This nonlinear wave equation requires the same initial and boundary conditions as the wave equation (1.2).

In this thesis we are studying homogeneous linear wave equation (1.20) and nonlinear wave equation (1.24) as they are time-dependent problems which can be solved using a variety of spatial discretization schemes combined with different time stepping schemes. There are many kinds of the nonlinear wave equations, but we are choosing this particular one for the following reasons:

- It is an easy nonlinear wave equation which could serve as a useful diagnostic equation to explore the accuracy of numerical solvers.
- Numerical stability of the wave equation is very common; this will be used as a basis to study the stability of a more complicated problem. Stability theory for the nonlinear problems is more difficult generally.

- To test and investigate our numerical experiments in the light of nonsmoothness to any nonlinear equation; we have chosen this one as more people have looked at but from different perspectives.

1.5 Numerical methods for discretizing partial differential equations

The partial differential equations (PDEs) that arise in applications can only seldom be solved in closed form. Even when they can be, the solutions are often impractical to work with and to imagine. On the other hand, numerical techniques can be applied successfully to almost all well-posed PDEs (Fornberg, 1998).

Most of the large literature on the numerical solution of PDEs falls into three categories: finite difference methods, finite element methods, and spectral methods (Su, 1998; Trefethen, 2000). We will describe finite difference methods briefly in this chapter but they will not be discussed further in the rest of this thesis. However, finite element methods and spectral methods are described in more detail here as they are the main topics in this thesis. These three big technologies came into their own in nearly consecutive decades:

1950s: finite difference methods

1960s: finite element methods

1970s: spectral methods

Naturally, the root of each technology can be traced further back (Trefethen, 2000). This section uses the references (Boyd, 2001; Fornberg, 1998; Su, 1998; Trefethen, 2000; Zhao, 1993).

1.5.1 Finite difference methods (FDM)

Finite difference methods are a classic numerical method of solution of differential equations using finite difference equations to approximate derivatives. The FDM approximate the unknown function $u(x)$ by a sequence of overlapping polynomials which interpolate $u(x)$ at a set of grid points. The approximation of $u(x)$ is done by using the local polynomial interpolant. The derivatives of $u(x)$ are then

approximated by differentiating this local polynomial. The general manner for solving PDEs by finite difference is by obtaining the solution at grid points (Smith, 1985).

1.5.2 Finite element methods (FEM)

In the finite element methods (FEM) the solution domain is discretized by dividing it into non-overlapping elements of arbitrary shape and size. Within each element a certain number of nodes are defined at which the unknown nodal values are to be determined. These nodal values are used to approximate the exact solution by a finite linear combination of local basis functions (Boyd, 2001). In more detail, considering a solution domain Ω , one can split it into a mesh containing N_e elements, denoted by Ω^e , such that the union of the elements equal the original domain, that is,

$$\Omega = \cup_{e=1}^{N_e} \Omega^e, \quad \text{where } \Omega^e \cap \Omega^{e'} = \emptyset \text{ for all } e \neq e',$$

(see Karniadakis & Sherwin, 2005). In the finite element method, we seek an approximation to the function u over each finite element. The polynomial approximation of the solution within a typical finite element Ω^e is of the form

$$u^e(x) = \sum_{k=0}^N u_k^e \phi_k^e(x),$$

where N is an integer number and u_k^e are the values of the solution at the nodes of the finite element and ϕ_k^e are the basis functions (Reddy, 1993).

Finite element methods are a collection of theory-rich techniques that can produce approximate solutions to equations common to engineering and mathematical physics (Baker & Pepper, 1991). Earliest examples of the FEM usage date from the early forties of the 20th century; certain key features of the FEM can be found in the works of (Hrennikoff, 1941) and (Courant, 1943). The term “finite element” was first used by (Clough, 1960). The first paper was published by (Turner *et al.*, 1956), followed by (Clough, 1960) and (Argyris, 1963), among others.

There is a vast literature on the FEM, therefore, in addition to the books mentioned above, we list here some more: (Becker *et al.*, 1981), (Whiteman, 1996) and (Babuska *et al.*, 2010), which all have several editions.

1.5.3 Spectral method (SM)

Spectral methods are a more recent technique to discretize PDEs. In contrast to traditional methods such as finite difference and finite element methods, spectral methods can achieve higher degree of accuracy only limited by computer accuracy for smooth problems, which require less modes. Spectral methods are **global in nature**: that is, a computation at any point will depend on information on the whole domain of computation, not only on information at neighbouring points (Boyd, 2001).

Chebyshev polynomials are one of the most interesting polynomials in approximation theory; these polynomials are used in this thesis. A common way to introduce them starts by approximating the function wanted to differentiate as a sum of very smooth basis functions,

$$u(x) = \sum_{k=0}^N \hat{u}_k \phi_k(x), \quad (1.25)$$

where the $\phi_k(x)$ are for example Chebyshev polynomials or trigonometric functions (Fornberg, 1998).

Spectral methods are used in a diverse range of mathematical, engineering and physical applications (Boyd, 2001; Canuto *et al.*, 1988). For a historical point of view, since 1970s spectral methods have been used more and more for the solution of PDEs. Obviously, their origin goes back years before. Some ideas are as old as expansion, interpolation, and more specifically algorithmic developments came with Lanczos in early 1938 (Lanczos, 1938, 1956) and with Clenshaw, Elliott, Fox, and others in the 1960s (Fox & Parker, 1968). Then, a transformation of the field was initiated in 1970s by work by Orszag and others on problems in fluid dynamics and meteorology, and spectral methods became well known. Then in 1977 Orszag and Gottlieb published in SIAM the book titled *Numerical analysis of spectral methods: Theory and applications*, in which they gave a systematic description of the theory and applications of the spectral methods (Fornberg, 1998), (Zhao, 1993). Furthermore, two landmarks of the early modern spectral methods literature were by Hussaini, Gottlieb and Voigt (Hussaini *et al.*, 1984) and by Canuto, Hussaini, Quarteroni, and Zang (Canuto *et al.*, 1988). Other books have been published since then by (Mercier, 1989), (Boyd, 2001) (first edition in 1989), (Funaro, 1992),

(Fornberg, 1998) and (Boyd, 2001) (revised second edition in 2001).

The origin of Chebyshev points and Chebyshev polynomials goes back more than a century: they were discovered by the Russian mathematician Pafnuty L'vovich Chebyshev (1821-1899) (Fox & Parker, 1968; Trefethen, 2013). There have been many books and papers discussing the Chebyshev polynomials and their applications, (Snyder, 1966), (Fox & Parker, 1968), (Rivlin, 1974) and (Trefethen, 2013) are but a few of them.

The Chebyshev series are best used for non-periodic functions as basis function on a finite interval. The Fourier spectral approximation is the traditional and original spectral method, and it is important in the solution of periodic problems, but it is not applicable for non-periodic functions (Zhao, 1993).

There are several spectral methods that can be employed to solve PDEs. The most representative methods are: the pseudospectral, the collocation, the tau and the Galerkin methods, (Canuto *et al.*, 1988), which are considered in this thesis.

1.5.4 Comparison of the previous three numerical methods

Moving to **comparison of the previous three methods**, typical finite difference methods use a local stencil to compute the derivative at a given point. In finite difference and finite element methods, a low order local function is used to approximate the unknown function; in other words, these methods are **local methods**. However, spectral methods are **global methods**. They use globally smooth functions (usually a set of orthogonal polynomials) to approximate the unknown function.

Another important point to compare is the speed of convergence of the methods. The convergence of spectral methods is exponential for smooth problems, while the convergence of finite difference and finite element methods is only algebraic. Because of exponential convergence, spectral methods can achieve much higher accuracy than finite difference and finite element methods with the same number of grid points.

One advantage of finite difference methods is that they are easy to use, easier than finite element and spectral methods. And an advantage of using finite difference and finite element methods is that both can use parallel computers in their calculations because they are both local methods. In contrast, spectral methods can not use

this type of computer so easily because they are global methods. However, spectral methods strength lies in the exponential convergence of the solutions for smooth problems, therefore, they are more efficient than the other two methods for those problems.

For further understanding, see (Boyd, 2001; Fornberg, 1998; Su, 1998; Trefethen, 2000; Zhao, 1993).

1.6 Published mathematical literature

The following is a review of published mathematical literature divided into two parts: The first is the literature of our nonlinear cubic wave equation (1.24),

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3,$$

and the other is the literature of the nonsmoothness aspect for some problems generally.

Let us consider the nonlinear wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} + mu + f(u) = 0.$$

When $m = 0$ the equation is called “**completely resonant**”, this name was given for the first time by (Pöschel, 1996, p. 274). In this thesis we are considering $\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} + u^3 = 0$; this completely resonant equation arises in different physical problems, where $u(t, x)$ is a real-valued function (Lidskii & Shulman, 1989).

The reason that our equation is called “**completely resonant**” is that the linearised equation at $u = 0$ is the linear wave equation,

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} = 0,$$

whose solutions are all periodic in time with period 2π when solved subject to Dirichlet $u(0, t) = u(\pi, t) = 0$ or space-periodic $u(x, t) = u(x + 2\pi, t)$ boundary conditions, (Berti & Bolle, 2008; Gentile *et al.*, 2005; Paleari *et al.*, 2001; Procesi, 2005).

For the completely resonant PDEs, existence of small-amplitude periodic or quasi-periodic solutions have attracted much attention for a long time, and there is a

massive literature for different such PDEs. For example, since 1989 existence of periodic solutions for the equation $f = u^3$ has been studied, (see [Bambusi & Paleari, 2001](#); [Berti & Bolle, 2004](#); [Bourgain, 1999](#); [Craig & Wayne, 1993](#); [Gentile *et al.*, 2005](#); [Giuliani *et al.*, 2020, 2021](#); [Lidskii & Shulman, 1989](#); [Paleari *et al.*, 2001](#)) and references therein. In particular, in the original work of ([Craig & Wayne, 1993](#)), periodic solutions for one-dimensional nonlinear wave equations have been extensively studied by using the Lyapunov-Schmidt method and Newton iterations. And quasi-periodic solutions with two frequencies have been proved in ([Baldi, 2005, 2007](#); [Procesi, 2005](#)) for the specific nonlinearities $f = u^3 + O(u^5)$.

The first existence results for small-amplitude periodic solutions of $\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} + u^3 = 0$ were obtained in ([Lidskii & Shulman, 1989](#)) with periodic (with respect to space) boundary conditions. They have found a family of small (in amplitude) periodic (with respect to t and x) solutions.

[Bourgain \(1999\)](#) has proved the existence of periodic solutions (for positive measure sets of frequencies) for this equation in the case of periodic boundary conditions in x , in the form of traveling waves $u(t, x) = \alpha p_0(\omega t + x)$, where ω is the frequency, $\omega^2 = 1 + \alpha^2$ and p_0 is a nontrivial 2π -periodic solution of the ordinary differential equation $p_0'' = -p_0^3$ (see [Berti & Bolle, 2008](#), for more information).

More recently, ([Yuan, 2006](#)) has studied the existence of quasi-periodic solutions for our completely resonant nonlinear cubic wave equation, subject to periodic boundary conditions. The main conclusion arrived in this paper is that there are plenty of quasi-periodic solutions of this equation in the neighbourhood of the solution $u_0(t)$ of $u_0'' + u_0^3 = 0$, which is uniform in space and periodic in time.

Moreover, the obtained solution $u(t, x)$ is even and 2π -periodic in x and quasi-periodic in time t .

[Bambusi & Paleari \(2001\)](#) construct some families of small-amplitude periodic solutions theoretically close to a completely resonant equilibrium point of a semilinear partial differential equation. For this purpose, using averaging methods, they construct a suitable map from the configuration space to itself. The paper proves that to each nondegenerate zero of such a map there corresponds a family of small-amplitude periodic solutions of the system. Lyapunov-Schmidt decomposition is used as a basis for the proof. This leads to establish a relation between Lyapunov-Schmidt decomposition and averaging theory. As an application to the above, they consider the nonlinear string equation $u_{tt} - u_{xx} \pm u^3 = 0$ (and of its perturbations)

with Dirichlet boundary conditions, which is Hamiltonian, see Section 2.4 for more information about Hamiltonian. For such an equation they compute all the zeroes of the considered map and prove that they are nondegenerate. They obtain countable many families of periodic solutions and find their fundamental periods. They prove that the fundamental periods of solutions belonging to the n^{th} family converge to $2\pi/n$ when the amplitude tends to zero.

In January 1983 during the Sixth I. G. Petrovskii memorial meeting of the Moscow Mathematical Society Professor V.E. Zakharov proposed the following problem. Numerical experiments demonstrated that the initial-boundary-value problem of the semilinear wave equation,

$$u_{tt} - u_{xx} + u^3 = 0$$

subject to periodic boundary conditions $u(t, 0) = u(t, 2\pi)$, $u_t(t, 0) = u_t(t, 2\pi)$ and initial conditions $u(0, x) = u_0(x)$, $u_t(0, x) = v_0(x)$, where $u_0(x)$ and $v_0(x)$ satisfy the same periodic conditions, possesses the "returning" property. The "returning" property means that the solutions ultimately return to a neighbourhood of the initial state $u_0(x)$, $v_0(x)$ after undergoing a possibly chaotic evolution. The paper (Friedlander, 1985) explains this phenomenon. Friedlander is taking into consideration the classical Poincaré theorem that every flow which preserves a finite measure has the returning property modulo a set of measure zero. So, an appropriate abstract space is considered in the paper. In this space a finite measure is constructed. This measure is invariant under the flow generated by the Hamiltonian system which corresponds to the original equation. This enabled Friedlander to verify the above "returning" property.

In our thesis we are actually coming across the phenomena of returning property, when we are solving analytically the linear wave equation using the d'Alembert solution, see Section 1.9; the analytical solution is having the returning property to the initial conditions. This happens at $t = 0, 4, 8, \text{etc.}$, because when extending the initial data $f(x)$ to the whole line to be an infinite interval the new function $F(x)$ used to find the exact solution is a periodic function of period four. So, that leads to get the solution to be exactly the initial condition function at those times. We can see that in the numerical solutions for the linear and nonlinear wave equations as well, although it does not go back exactly, but still one can see a bit about it,

we see periodicity of period 4 for the linear equation and quasi-periodicity, with similarity 4 units later, for the nonlinear wave equation.

As noticed in all the papers above that they are solving our completely resonant nonlinear wave equation, (1.24), theoretically considering mainly infinite domains. However, the solutions given are a very special class of solutions that can be constructed satisfying some given assumptions. They do not formulate a completely general solution. In addition, no theoretical solution or numerical method/solution were presented on finite domains in the light of such ICs that generate nonsmoothness solution in later derivatives. Therefore, this thesis considers the numerical methods to solve the equation. And as stated in (Friedlander, 1985) and (Bambusi & Paleari, 2001) that this nonlinear cubic wave equation is Hamiltonian, so in the light of this fact this thesis presents the Hamiltonian system of the equation and then studies the symplecticity, see Section 2.5 for more information about the symplecticity, of the numerical schemes employed when studying the energy conservation.

The papers are proving the existence of small-amplitude periodic or quasi-periodic solutions of the problem. Our numerical results in the present work show that the analytical solutions of the linear wave equation are also indeed periodic, of period 4, and also the numerical solutions are periodic for all the ICs implemented. Also, our result show that the numerical solutions obtained for the nonlinear wave equation, $u_{tt} - u_{xx} + u^3 = 0$, are behaving more like periodic or quasi-periodic according to the ICs and BCs implemented, and also the time used. However, we can not be certain about the behaviour of the solutions of the nonlinear problem for a very large time.

All the above papers, when giving the solutions according to their assumptions, are considering periodic boundary conditions in x because their methods and assumptions about the nature of the solutions fundamentally do not work with other BCs. While in this work we are going beyond the simple problem they do with periodic in x , we are considering a harder problem, Dirichlet boundary conditions, where we are including how the wave actually reflects off those boundaries. In addition, all papers do not ever consider the possibility that we could have nonsmooth solutions as it will be difficult for them to implement their methods and assumptions in the nonsmoothness case or even do not work. However,

the nonsmoothness issue is something very important to be considered because this equation certainly can have nonsmooth solutions or can even itself generate nonsmooth solutions even if the IC is infinitely differentiable on the bounded intervals, as we will see in this work.

As we said all the above literature does not mention about the nonsmooth data; instead, they are automatically considering smooth data without even stating it. While the next papers acknowledge about nonsmoothness of the data by taking problems related but not the nonlinear problems.

Application of the Fourier method to linear systems of hyperbolic equations with periodic boundary conditions having nonsmooth initial data is considered by (Majda *et al.*, 1978), both theoretically and computationally. A detailed examination of the effect of filtering for these problems is made. Two main results are proven: the first shows that appropriate smoothing techniques, a proper filtering, applied to the equation gives stability; and the second states that this smoothing combined with a certain smoothing of the initial data leads to infinite order accuracy.

A hyperbolic time-dependent partial differential equation that is periodic discontinuous function is considered in (Vandeven, 1991), and a technique to accelerate the rate of convergence of its Fourier series is given. A class of filters are presented and analyzed that allows the recovering of the solution with an exponential accuracy.

Dutt (1999) considers hyperbolic partial differential equations with boundary conditions having nonsmooth data. The paper recovers pointwise values with spectral accuracy, provided that the actual solution is piecewise smooth. This is done after some changes regarding the time domain and then solving the problem using a filtered version of the data by a specific spectral scheme. A local smoothing of the computed solution is performed.

Further literature exists in spectral methods for nonsmooth problems, for example see (Abarbanel *et al.*, 1985; Cai *et al.*, 1989; Dutt & Singh, 1993; Eckhoff, 1993, 1994).

The above literature is trying to resolve the nonsmooth data by adding additional factors, such as filters, to get the outcome they needed. The nonsmooth data are either in the initial condition or the solution itself. Mostly they deal with linear problems generally, and some are dealing with linear hyperbolic partial differential

equations. However, in this thesis we are doing something related but very different. We are considering a linear wave equation, (1.20), and a nonlinear wave equation, (1.24), and implementing some smooth ICs, infinitely differentiable. The analytic solutions produced by those ICs are smooth and also their first derivatives as well. Nevertheless, the analytic solutions at the second and fourth derivatives are nonsmooth depending on the IC employed. In this thesis we are not attempting to resolve the nonsmoothness in the second and fourth derivatives of the solutions; rather, we are examining the consequence of this nonsmoothness by studying and comparing the results produced by such different levels of smoothness using spectral methods and finite element methods.

Based on all the above literature research, this thesis aims to fill in the gaps on the nonsmoothness issue, i.e., it considers problems, linear and nonlinear, that have discontinuities in certain higher derivatives, and to control precisely the level of the nonsmoothness. In addition, this thesis seeks to fill in the gaps on the numerical solution for our nonlinear equation on a bounded interval in consideration of nonsmoothness.

1.7 Time Stepping

Numerical initial value problems (IVPs) play a critical part in the design and analysis of partial differential equations (PDEs) (Iserles, 1996), which are the equations that we are solving in this study. Therefore, in this section some ideas are given for solving initial value problems numerically. After discretizing the PDEs in space, see Section 1.5, the produced ODEs can be solved using the time stepping methods.

Let us consider the initial value problem,

$$y' = f(y, t), \quad a \leq t \leq b, \quad y(a) = \alpha. \quad (1.26)$$

There are a number of different time stepping methods for solving the initial value problem (1.26), but in this section we are focusing on methods that will be used for solving the stiff problems which will appear in solving PDEs in later chapters of the thesis. Stiff problems are described in Section 2.3.

In the process of finding the formulas of those methods the given interval $[a, b]$ is divided into N equally spaced intervals. The mesh points,

$$t_i = a + ih, \quad \text{for each } i = 0, 1, 2, \dots, N,$$

are defined, where N is a chosen positive integer number and $h = (b - a)/N = t_{i+1} - t_i$, and i is the step index; h is called the step size. The approximate solution is given as $u_i \simeq y(t_i)$, for each $i = 1, 2, \dots, N$.

The material discussed in this section is standard and can be found in many textbooks (e.g., [Atkinson, 1989](#); [Burden & Faires, 2001, 2011](#)).

1.7.1 Euler's method

Euler's method is the most basic approximation technique for solving initial value problems. Although it is seldom used in practice, the simplicity of its derivation can be used to illustrate the techniques involved in the construction of some of the more advanced techniques.

The Euler's method formula is,

$$\begin{aligned} u_0 &= \alpha, \\ u_{i+1} &= u_i + hf(u_i, t_i), \quad \text{for each } i = 0, 1, \dots, N - 1. \end{aligned} \tag{1.27}$$

This method is called **explicit** because it gives u_{i+1} explicitly in terms of previously determined values. Euler's method is a model for explicit methods. This method has a global truncation error of order $\mathcal{O}(h)$.

1.7.2 The modified Euler method

In addition to Euler's method, we can also define the modified Euler method. It has the formula,

$$u_{i+1} = u_i + \frac{h}{2} \left[f(u_i, t_i) + f(u_i + hf(u_i, t_i), t_{i+1}) \right], \quad \text{for each } i = 0, 1, \dots, N - 1, \tag{1.28}$$

(see [Burden & Faires, 2001, 2011](#)).

This method is also explicit. It is more accurate than Euler's method because it

is of order $\mathcal{O}(h^2)$. Both methods, Euler's and the modified Euler, have stability problems as we will see later in subsequent chapters.

1.7.3 The implicit Euler method

The implicit Euler method is also of order $\mathcal{O}(h)$ same as Euler's method; its formula is,

$$u_{i+1} = u_i + hf(u_{i+1}, t_{i+1}), \quad \text{for each } i = 0, 1, \dots, N - 1. \quad (1.29)$$

This method requires the solution of an equation to determine the new value u_{i+1} such methods are known as **implicit methods**.

The implicit Euler method is used to introduce implicit methods and ideas associated with them; it is a model for implicit methods as Euler's is a model for explicit methods. In addition, this method is one of the algorithms for the solution of the stiff ODEs, where the discussion of stiff methods is given in the next Chapter. We need to solve the following equation,

$$F(u) = u - u_i - hf(u, t_{i+1}) = 0, \quad (1.30)$$

to obtain u_{i+1} , see (1.29). Generally, this is done iteratively, for instance by the secant method, but often Newton's method is used. Hence, we give details on how to use the implicit rule with Newton iteration to solve (1.30).

The **Newton-Raphson formula for the multi-variable problem** $F(p) = 0$ is,

$$p^{(k)} = p^{(k-1)} - J(p^{(k-1)})^{-1}F(p^{(k-1)}), \quad \text{for } k \geq 1, \quad (1.31)$$

where $J(p)$ is the Jacobian matrix of function $F(p)$,

$$J(p) = \frac{\partial F}{\partial p}, \quad (1.32)$$

and here $p = (u_0, u_1, \dots, u_{N+1})$.

To approximate the solution of (1.30), select $u_{i+1}^{(0)}$ usually as u_i , and generate $u_{i+1}^{(k)}$ by applying Newton's method,

$$u_{i+1}^{(k)} = u_{i+1}^{(k-1)} - \frac{F(u_{i+1}^{(k-1)})}{F'(u_{i+1}^{(k-1)})}, \quad (1.33)$$

$$= u_{i+1}^{(k-1)} - \frac{u_{i+1}^{(k-1)} - u_i - hf(u_{i+1}^{(k-1)}, t_{i+1})}{1 - hf'(u_{i+1}^{(k-1)}, t_{i+1})}, \quad \text{for } k \geq 1, 0 \leq i < N, \quad (1.34)$$

until $|u_{i+1}^{(k)} - u_{i+1}^{(k-1)}|$ is sufficiently small (see [Burden & Faires, 2011](#)). Here k is the iteration index and i is the step index. For vector equations, the term $\frac{F}{F'}$ means multiplying with matrix inverse.

Normally only three or four iterations per step are required, because of the quadratic convergence of Newton's method.

1.7.4 The trapezoidal method

This method is another implicit method. It is also complicated as the implicit Euler method, however, it is of order $\mathcal{O}(h^2)$, same as modified Euler method.

The formula of the trapezoidal method is,

$$u_{i+1} = u_i + \frac{h}{2} \left[f(u_{i+1}, t_{i+1}) + f(u_i, t_i) \right], \quad \text{for each } i = 0, 1, \dots, N-1. \quad (1.35)$$

In order to find the solution to

$$F(u) = u - u_i - \frac{h}{2} \left[f(u, t_{i+1}) + f(u_i, t_i) \right] = 0, \quad (1.36)$$

we use **Trapezoidal rule with Newton iteration**. This means that we find u_{i+1} in (1.35) by applying Newton's method to (1.36). Thus select $u_{i+1}^{(0)}$ and generate $u_{i+1}^{(k)}$ by

$$u_{i+1}^{(k)} = u_{i+1}^{(k-1)} - \frac{F(u_{i+1}^{(k-1)})}{F'(u_{i+1}^{(k-1)})} \quad (1.37)$$

$$= u_{i+1}^{(k-1)} - \frac{u_{i+1}^{(k-1)} - u_i - \frac{h}{2} [f(u_i, t_i) + f(u_{i+1}^{(k-1)}, t_{i+1})]}{1 - \frac{h}{2} f'(u_{i+1}^{(k-1)}, t_{i+1})}, \quad \text{for } k \geq 1, 0 \leq i < N, \quad (1.38)$$

until $|u_{i+1}^{(k)} - u_{i+1}^{(k-1)}|$ is sufficiently small (see [Burden & Faires, 2011](#)).

1.7.5 Theta methods

The previous methods, Euler's, the implicit Euler and the trapezoidal, are included as special cases of the following formula

$$u_{i+1} = u_i + h \left[\theta f(u_{i+1}, t_{i+1}) + (1 - \theta) f(u_i, t_i) \right], \quad \text{for each } i = 0, 1, \dots \quad (1.39)$$

Euler's method corresponds to the choice $\theta = 0$, the implicit Euler to $\theta = 1$ and the trapezoidal rule to $\theta = \frac{1}{2}$. In general, formula (1.39) for $\theta \in [0, 1]$ is known as the theta method (see [Hundsdorfer & Savcenca, 2009](#)).

1.7.6 The implicit midpoint and Runge-Kutta methods

There are a number of other methods for solving IVPs, and here some examples are given, namely the implicit midpoint method and Runge-Kutta methods. The implicit midpoint is used briefly in Chapter 4 and Runge-Kutta method of order two is used in exponential integrators in the next subsection, which are implemented in Chapter 5.

The midpoint method is another implicit method; its formula is,

$$u_{i+1} = u_i + hf \left[u_i + \frac{h}{2} f(u_i, t_i), t_i + \frac{h}{2} \right], \quad \text{for each } i = 0, 1, \dots, N-1. \quad (1.40)$$

The general formula for Runge-Kutta methods is,

$$u_{i+1} = u_i + \sum_{j=1}^s b_j K_j, \quad i = 0, 1, \dots, N-1,$$

where

$$K_j = hf \left[t_i + c_j h, u_i + \sum_{n=1}^{s-1} a_{jn} K_n \right], \quad j = 1, \dots, s,$$

where the constant coefficients $\{c_j, a_{jn}, b_j\}$, for $1 \leq j \leq s$, and $1 \leq n < s$, where s is an integer number, specify the method and $u_i \simeq y(t_i)$. Another method is Runge-Kutta-Fehlberg method, which uses Runge-Kutta methods of two different orders in order to determine the truncation error. This is used to choose h in order to avoid spending too much time with excessively small step sizes. The general formula for Runge-Kutta methods and Runge-Kutta-Fehlberg method are not used in this thesis but included for completeness.

1.7.7 Exponential time differencing methods (ETD)

In addition to the previous time stepping methods, some other ones are given here; the methods are called exponential time differencing methods (ETD) or exponential

integrators methods, in fact they have many names, which use the exponential and often functions which are closely related to the exponential function inside the numerical method. The exponential time differencing (**ETD1**) and the exponential time differencing with Runge-Kutta time stepping (**ETD2RK**) are used in this study.

When discretizing the spatial part of the PDE a system of ODEs is obtained, which can be written in vector form,

$$\vec{u}' = L\vec{u} + \vec{\Gamma}(\vec{u}, t), \quad (1.41)$$

where L and $\vec{\Gamma}$ are linear and nonlinear operators, respectively.

The (**ETD1**) method has the formula,

$$\vec{u}_{n+1} = e^{Lh}\vec{u}_n + L^{-1}(e^{Lh} - I)\vec{\Gamma}_n, \quad (1.42)$$

where we denote the numerical approximation to $\vec{u}(t_n)$ by \vec{u}_n and $\vec{\Gamma}(\vec{u}(t_n), t_n)$ as $\vec{\Gamma}_n$. This method is of order $\mathcal{O}(h)$.

The **ETD2RK** formula is,

$$\vec{u}_{n+1} = \vec{a}_n + (\vec{\Gamma}(\vec{a}_n, t_n + h) - \vec{\Gamma}_n)(e^{Lh} - I - hL)L^{-2}/h, \quad (1.43)$$

where \vec{a}_n is given by

$$\vec{a}_n = e^{Lh}\vec{u}_n + L^{-1}(e^{Lh} - I)\vec{\Gamma}_n, \quad (1.44)$$

the method is of order $\mathcal{O}(h^2)$.

Generally, the exponential methods solve the linear equations exactly, therefore, they could serve as useful diagnostic methods to explore the accuracy of numerical solvers for the nonlinear wave equation later.

The references (Cox & Matthews, 2002; Kassam & Trefethen, 2005; Krogstad, 2005; Minchev & Wright, 2005; Minchev, 2004) are used to write this subsection.

1.8 Stability of PDEs

The term stability is used to designate that any numerical errors introduced at some stage of the calculations do not blow up in the subsequent steps of the method. In other words, stability means that errors at any stage of the computation are not

amplified but instead are bounded.

Classical tools to assess stability in the numerical solution of partial differential equations a priori include Fourier transformation and the corresponding famous Von Neumann condition for the stability of finite difference methods. Further tools of recognized merit for assessing stability in the solution of ordinary differential equations (ODEs) comprise so-called stability regions in the complex plane. It is important to know to what extent stability regions can be relied upon in assessing stability in the numerical solution of differential equations (Spijker, 1996).

Stability of spectral methods for time dependent PDEs requires that the eigenvalues of the spatial discretization operator, scaled by $h = \Delta t$, lie in the stability region of the time stepping formula; this is only for linear stability. Hence, for a given spectral discretization, we must pick h small enough that this stability region encloses the eigenvalues of the spectral discretization operator, scaled by h . Because of large eigenvalues, especially in the Chebyshev case, time step limits for explicit methods may be very severe, making it advantageous to use implicit or semi-implicit methods (Trefethen, 2000).

Nonlinear stability results can be proved either by the energy method or by a generalized variational principle for finite element and spectral methods (Canuto *et al.*, 1988). We talk about Hamiltonian differential equations in Chapter 2 and use it to assess the stability of numerical methods in Chapter 4, and this is closely related to energy method.

1.9 D'Alembert solution of the wave equation

The method of d'Alembert provides a solution to the one-dimensional linear wave equation on the domain $(-\infty, \infty)$. We consider the following wave equation of an infinite domain with two initial conditions, specifying the initial position $u(x, 0)$ and the initial velocity $u_t(x, 0)$,

$$\begin{aligned} \frac{\partial^2 u}{\partial t^2} &= c^2 \frac{\partial^2 u}{\partial x^2}, \quad \text{for } -\infty < x < \infty, t > 0, \\ u(x, 0) &= F(x), \quad \text{for } -\infty < x < \infty, \\ \frac{\partial u}{\partial t}(x, 0) &= G(x), \quad \text{for } -\infty < x < \infty. \end{aligned}$$

The d'Alembert general solution of this equation is

$$u(x, t) = \frac{1}{2} [F(x + ct) + F(x - ct)] + \frac{1}{2c} \int_{x-ct}^{x+ct} G(s) ds. \quad (1.45)$$

When the initial velocity is zero, $G(x) = 0$, as we are considering in this thesis, the d'Alembert solution becomes

$$u(x, t) = \frac{1}{2} [F(x + ct) + F(x - ct)], \quad (1.46)$$

which has been obtained from (1.18) using the same procedure.

The material discussed above in this section for the infinite domain is standard and can be found in many textbooks, such as (Gockenbach, 2002; Strauss, 1992, 2007; Vasy, 2015).

This solution generally is used for infinite domains but it can be modified to also provide an exact analytic solution on a finite interval in space, as what we are doing in this thesis, where we have the finite interval $[-1, 1]$. The linear wave equation we have is,

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \quad (1.47)$$

on $(-1, 1)$ and $t \geq 0$, see (1.20), to be solved subject to Dirichlet boundary conditions

$$u(-1, t) = 0, \quad \text{and} \quad u(1, t) = 0, \quad (1.48)$$

and initial conditions

$$u(x, 0) = f(x), \quad (1.49)$$

and

$$\frac{\partial u}{\partial t}(x, 0) = 0. \quad (1.50)$$

The d'Alembert solution is not well-defined for this problem because $F(x + ct)$ in (1.46) is not defined for $x + ct > 1$ for $c, t > 0$, similarly $F(x - ct)$ is not defined for $x - ct < -1$. This can be seen by looking at the characteristics in the xt -plane. In Figure 1.2 there are characteristics propagating from the points -1 and 1 . However, for this problem we can only obtain information for values of x and t such that $-1 \leq x - ct \leq 1$ and $-1 \leq x + ct \leq 1$. In other words, the main ABC triangle, which is the domain of dependence, is the only region in which the solution can be found based solely on the initial conditions. The point of intersection of point $P1$ has a domain of dependence entirely in the region $[-1, 1]$ for $t > 0$, however, the

domain of dependence of point $P2$ reaches outside this region; both characteristics $x + ct$ and $x - ct$ go beyond the interval $[-1, 1]$ at $t = 0$, where in our problem (1.47) we have $c = 1$. But, we need $F(x)$ for $x > 1$ and $F(x)$ for $x < -1$ to form a solution. Therefore, we must extend the initial data to the whole line. That means, we must change it to an infinite domain first. See Section 1.1 for more information.

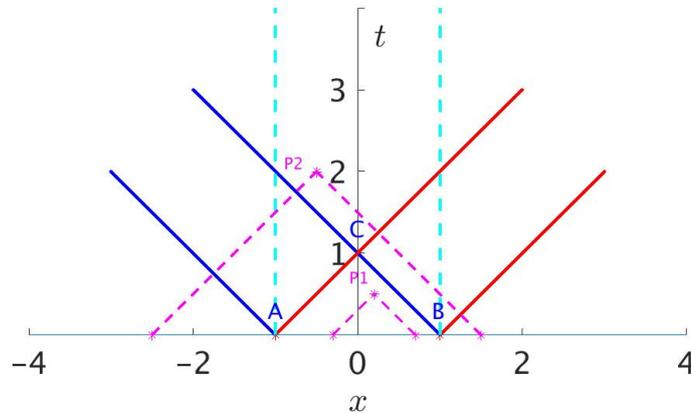


Figure 1.2: The characteristics propagating from the interval $[-1, 1]$ for the finite string problem (1.47). The slope of the lines is ± 1 as $c = 1$.

Our initial condition function $f(x)$ is in the interval $[-1, 1]$. We can extend it into the region $x < -1$ and to the region $x > 1$. For the homogeneous Dirichlet data the appropriate choice is sort of the odd extension, see (Daileda, R. C., 2014; Grigoryan, V., 2011). Given the function $f(x)$ with domain $[-1, 1]$ we first extend it to an odd-like function on $[-3, -1]$ by reflecting its graph through the boundary point -1 and similarly at 1 . The transform of the function $f(x)$ into an infinite interval can be done in such a way that the new function, modified from $f(x)$, is periodic. The extension formula of extending the function $f(x)$, which is in $[-1, 1]$, to the left side $[-3, -1]$ from the boundary point -1 and to the right side $[1, 3]$ from the end point 1 , is

$$F(x) = \begin{cases} -f(-2-x), & \text{for } -3 \leq x < -1, \\ f(x), & \text{for } -1 \leq x < 1, \\ -f(2-x), & \text{for } 1 \leq x < 3. \end{cases} \quad (1.51)$$

The formula is clear in Figure 1.3, where the red plot is the original function $f(x)$, then its extension to the left is the black plot and to the right is the green plot. Now after getting the function $F(x)$ in $[-3, 1]$ (see the right panel), or in $[-1, 3]$, one can then extend this to all infinite domain by repeatedly “cutting and pasting” its graph. This leads to a periodic function on the infinite domain of period 4; the function on the interval $[-3, 1]$ or $[-1, 3]$ is one period. This is called the 4-periodic odd-type extension of $f(x)$. This formula will help to know the behaviour of $F(x)$ when studying the continuity of the solution especially at the boundary points -1 and 1 .

To illustrate the above one can take for example $f(x) = e^{-\sin(\pi(x+1))} - 1$, we will obtain the following figures:

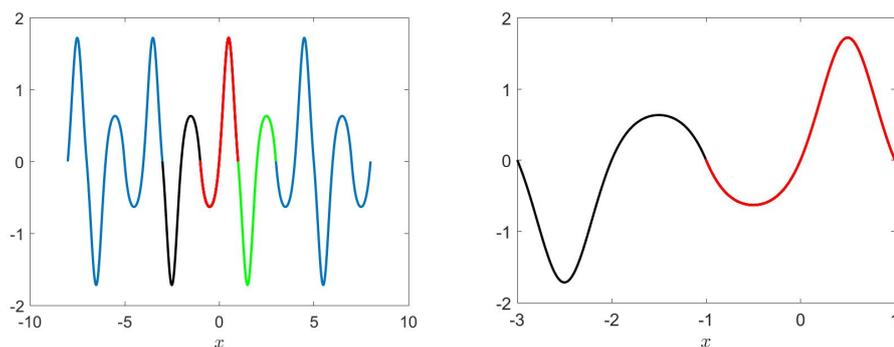


Figure 1.3: The function $F(x)$, (1.51), in infinite domain, but here we take a part of the domain. It shows how the original function $f(x)$ is extended to the whole domain. The figures show that the function $f(x)$ is in the interval $[-1, 1]$, the red line, then it is extended oddly through the boundary point -1 to the interval $[-3, -1]$, the black line, the right panel show this clearly, similarly $f(x)$ is extended oddly from the boundary point 1 to the interval $[1, 3]$, the green line. Here we consider $f(x) = e^{-\sin(\pi(x+1))} - 1$.

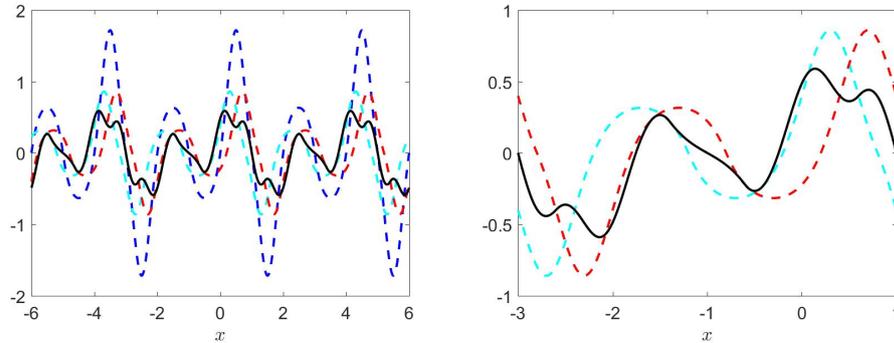


Figure 1.4: The d'Alembert solution (1.46), $u(x, t)$, the black line. The time used here is $t = 0.2$. This solution is half of the initial condition $F(x)$, (1.51), which is the blue dashed line, shifted to the left, $F(x + t)$ the cyan dashed line, and half shifted to the right, $F(x - t)$ the red dashed line, and then adding them up. The left panel is the solution on the infinite domain. The right panel is the solution on one period $[-3, 1]$.

This extended periodic function $F(x)$, defined for all x , is used as an initial condition for the wave equation defined for all x , then the resulting solution will very conveniently satisfy the desired boundary conditions that $u = 0$ at $x = -1$ and 1 , and for all times t (also at all new boundary points made through the extension process). This is clear in Figure 1.4. This new IC will help us to get the d'Alembert solution of the linear wave equation in infinite domain from the formula (1.46). The reason that the d'Alembert solution on the infinite interval will satisfy $u = 0$ at $x = -1$ and 1 at all times t is because that this solution ends up being just half of the initial condition, $F(x + t)$, moving to the left, and half, $F(x - t)$, moving to the right and then adding them up; the components, $F(x + t)$ is the cyan dashed line and $F(x - t)$ is the red dashed line in Figure 1.4, summation will satisfy the BCs at the boundaries at any time t although the components themselves do not satisfy the BCs at the boundaries, see the right panel.

Once this has been established, one then can effectively restrict the solution back to the finite interval $[-1, 1]$. Therefore, after implementing the d'Alembert solution formula, (1.46), in the whole domain, the rest of the graph outside the interval $[-1, 1]$ must be cut off. Consequently, Figure 1.5 shows how the characteristics are extended throughout the domain strip $[-1, 1]$. In other words, we end up to find

the solution of our linear wave equation, (1.47)-(1.50), in the finite domain $[-1, 1]$ only, as it should be.

Wherefore, the exact d'Alembert solution of the wave equation at any time of multiple 4, i.e., at $t = 0, 4, 8, \text{etc.}$, is exactly the function $f(x)$ in the initial conditions because the function $F(x)$ is a periodic function with period four and by inserting this into (1.46) we will end up to have the solution to be exactly $f(x)$. Figure 1.5 mirrors this property as it is apparent that the initial data go back to their initial place after each period (of period 4). In addition, the characteristics also touch the boundaries at other times of multiple 2, i.e., at $t = 2, 6, 10, \text{etc.}$, and the exact d'Alembert solution at these times is the function $-f(-x)$. See Section 1.1 for further information.

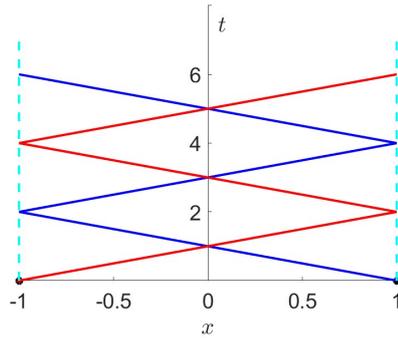


Figure 1.5: Extending the characteristics throughout the domain strip $[-1, 1]$.

Nevertheless, if $f(x)$ were extended in some wrong ways, then inserting the resulting $F(x)$ into the d'Alembert formula would yield a solution that does not satisfy the boundary conditions. Therefore, we must extend it this way so that the resulting extension function is periodic and satisfy the BCs.

The d'Alembert solution for the bounded interval can yield solutions with discontinuities in second or fourth or any higher even derivative, even if the initial condition is infinitely differentiable on the bounded interval, as it happened in this study. In fact, the infinite smoothness of the ICs could be a tricky thing that one cannot anticipate and pay close attention that this could eventually lead to nonsmooth solutions later. Thus, one must be careful. To test these assumptions we will go back to (1.51) and take its general derivatives. Directly from the separate rules on the different intervals, we evaluate the derivatives on each interval

separately as follows,

$$F^{(n)}(x) = \begin{cases} (-1)^{n+1} f^{(n)}(-2-x), & \text{for } -3 \leq x < -1, \\ f^{(n)}(x), & \text{for } -1 \leq x < 1, \\ (-1)^{n+1} f^{(n)}(2-x), & \text{for } 1 \leq x < 3, \end{cases} \quad (1.52)$$

for any integer value of n .

In order to test their continuity one needs to see whether the two different rules across the interface are matching, i.e., at the boundaries $x = \pm 1$. In other words, given $F(x)$ on two adjacent intervals we need to see whether these two rules agree when the first rule approaches the interface point from below, and the second rule approaches it from above.

Coming now to test those derivatives at the interface points $x = \pm 1$. The function $F(x)$ is continuous, as long as the original function $f(x)$ is satisfying the boundary conditions, i.e. $F(x) = 0$ at $x = \pm 1$ because $f(x) = 0$ at those boundaries. The first derivative $F'(x)$ will also always be continuous because it agrees at the interface points, as one can see that $F'(x) = f'(x)$ at $x = \pm 1$.

Going now to the more important one $F''(x)$. In order to study this we need to work directly on the second derivatives on each interval separately and try to match them across the interfaces $x = \pm 1$. It will all depend on the original $f(x)$ because the first rule of $F(x)$ to approach the interface point at $x = -1$ is,

$$F''(x) = \begin{cases} -f''(-1), & \text{as } x \uparrow -1, \\ f''(-1), & \text{as } x \downarrow -1, \end{cases} \quad (1.53)$$

and the second rule to approach at $x = 1$ will have similar behaviour. So unless the original $f(x)$ satisfies $f''(x) = 0$ at $x = \pm 1$, F'' will not be continuous.

If we assume $F''(x)$ is continuous, then by continuing we find that $F^{(3)}(x)$ is again always continuous, same as $F'(x)$, but $F^{(4)}(x)$ is only continuous if again it is 0 at the endpoints. One can say that all the odd derivatives of the generated function $F(x)$ will be continuous, while the even derivatives will only be continuous when $F^{(n)}(x) = 0$ at $x = \pm 1$, for any even number n .

This solution has this interesting property that, in general after the extension there will be discontinuity in the second or fourth or higher even derivatives although on the original interval there were no discontinuities. Accordingly, from Figure 1.5 it is clear that these discontinuities will move along these characteristics; these

discontinuities in $F(x)$, which originally occurred only at the boundaries $x = \pm 1$ (the black dots on the x -axis in the figure), now will move into the interior at later times t . That is, the problem is triggered by the initial condition at the boundaries, then it propagates throughout the interior along the characteristics. In other words, this discontinuity that is originally in $F(x)$ will be also in the half of $F(x+t)$ and $F(x-t)$ in (1.46), which will be shifted into the inside of the interval at later times t . As a result, it is at this point when we anticipate that the computations will be difficult and that extra conditions will be required to obtain satisfactory results. To the best of our knowledge, we are the first to illustrate this discontinuity problem for the finite d'Alembert solution as the literature has not addressed it. Based on all this general theory it is obvious that some particular ICs that would be interesting to consider, such as

$$f(x) = \sin(\pi x), \quad (1.54)$$

$$f(x) = e^{\sin(\pi x)} - 1, \quad (1.55)$$

and

$$f(x) = \frac{1}{11} e^x (11 - 8x - 10x^2 + 8x^3 - x^4). \quad (1.56)$$

These initial conditions have been carefully chosen to yield solutions with different levels of smoothness. They are selected so that the initial condition itself is infinitely differentiable on the entire interval $[-1, 1]$, but the analytic solution at later times will have discontinuities in the second derivatives for the second IC (1.55) and discontinuities in the fourth derivative for the third IC (1.56).

These three initial conditions are having different properties. The first initial condition (1.54) yields an infinite smooth solution, it is infinitely differentiable in space for $[-1, 1]$ for all derivatives. In other words, the solution is continuous at all its derivatives, i.e. $F^{(n)}(x) = 0$ at the interfaces $x = \pm 1$ for all even n . This IC is giving such a smooth solution because $F^{(n)}(x)$ is continuous as the original $f(x)$ satisfies $f^{(n)}(x) = 0$ at $x = \pm 1$.

However, the second initial condition (1.55) introduces nonsmoothness at the second derivative because the second derivative is non-zero at the endpoints, i.e. $F''(x) \neq 0$ at the interfaces $x = \pm 1$ as $f''(x) \neq 0$ at $x = \pm 1$. This means $F(x)$ does not have a continuous second derivative at the boundaries, $x = \pm 1$, so the worst possible discontinuity. While the third IC, (1.56), yields a twice continuously differentiable solution, as $F''(x) = 0$ at $x = \pm 1$, but the fourth derivative is not continuous

because $F^{(4)}(x) \neq 0$ is not zero at the boundaries as $f^{(4)}(x) \neq 0$. This IC has been chosen to be zero at the endpoints, and also have the second derivative zero at the endpoints, but the fourth derivative to be non-zero at the endpoints, so smoother solutions than for second IC (1.55), but still not infinitely differentiable as for first IC (1.54). Having discontinuity at the boundary points this leads to have the numerical experiments of the third IC (which is C^2 but not C^4) fall between the other two ICs and this is clear throughout the thesis.

Another aspect we will study relates to splitting the solution in even and odd parts. As is generally known, every function $f(x)$ can be uniquely decomposed as the sum of an even and an odd function, which are called respectively the even part and the odd part of the function; if one defines

$$f_e(x) = \frac{f(x) + f(-x)}{2}$$

and

$$f_o(x) = \frac{f(x) - f(-x)}{2}$$

where f_e is the even part, f_o the odd part, and then

$$f(x) = f_e(x) + f_o(x).$$

Therefore, the function $f(x)$ of our initial conditions can be decomposed in a such manner. The first IC (1.54) is an odd function. However, the other two ICs, the second IC (1.55) and the third IC (1.56), are mixed; they have both odd and even components.

The solution of the homogeneous linear wave equation preserve the symmetry properties of the initial condition $f(x)$. If $f(x)$ is odd (resp. even) then the exact solution remains odd (resp. even). For our linear wave equation, the odd and even components of the initial condition and the exact solution can be evolved independently. As mentioned before, the first IC (1.54) yields an infinitely smooth solution because $F^{(n)}(x)$ is continuous as $f^{(n)}(\pm 1) = 0$. Therefore, the whole solution evolves in such a way that the odd function is always smooth. However, the second IC (1.55) introduces nonsmoothness at the second derivative because $F'''(\pm 1) \neq 0$ as $f''(\pm 1) \neq 0$; the values are $f''(\pm 1) = \pi^2$, the important point is that the two values are the same. Consequently, this leads to $f_o''(\pm 1) = 0$ and

$f_e''(\pm 1) = \pi^2$. In particular, that means all the discontinuities generated from f'' are associated with the even part of f , and none of it with the odd part of f ; the discontinuities generated from f'' are completely even. The discontinuities is not affecting the odd part that means this part will behave exactly the same as the first IC (1.54), which is totally an odd function. If we evaluate the higher even derivatives we can see that $f^{(4)}(\pm 1) = -3\pi^4$ and $f^{(6)}(\pm 1) = -3\pi^6$; again the two values are the same and probably if one works out all the higher even derivatives the values will be again identically the same at the interfaces $x = \pm 1$.

While for the third IC (1.56) the second derivative is zero at the endpoints, i.e. $F''(\pm 1) = 0$ as $f''(\pm 1) = 0$, but the fourth derivative is non-zero at the endpoints, i.e. $F^{(4)}(\pm 1) \neq 0$ as $f^{(4)}(\pm 1) \neq 0$; the values are $f^{(4)}(+1) = 136e$ and $f^{(4)}(-1) = -56e^{-1}$; the important point is that the two values are totally different magnitudes and not opposite. This leads to have $f_o^{(4)}(\pm 1) \neq 0$ and $f_e^{(4)}(\pm 1) \neq 0$. That means, in contrast to the second IC, the discontinuities generated from $f^{(4)}$ are associated with both components, the even part and the odd part of f . If we evaluate the higher even derivatives we can see that $f^{(6)}(+1) = 312e$, $f^{(6)}(-1) = 120e^{-1}$ and $f^{(8)}(+1) = 48e$, $f^{(8)}(-1) = 432e^{-1}$. We believe if one evaluates the rest of the higher even derivatives all will behave similarly; having two different values at the interfaces.

Remarkably, the odd/even phenomena can be presented in the exact d'Alembert solution itself; that is the exact d'Alembert solution itself has odd solutions at some specific times. For instance, let us consider the solution at $t = 1$, which we can obtained using (1.51) and after few steps we get

$$F(x+1) = \begin{cases} f(x+1), & \text{for } -1 \leq x < 0, \\ -f(1-x), & \text{for } 0 \leq x < 1. \end{cases} \quad (1.57)$$

and

$$F(x-1) = \begin{cases} -f(-1-x), & \text{for } -1 \leq x < 0, \\ f(x+1), & \text{for } 0 \leq x < 1. \end{cases} \quad (1.58)$$

and from (1.46) the d'Alembert solution will be

$$u(x,1) = \begin{cases} \frac{1}{2}f(x+1) - \frac{1}{2}f(-1-x), & \text{for } -1 \leq x < 0, \\ -\frac{1}{2}f(1-x) + \frac{1}{2}f(x-1), & \text{for } 0 \leq x < 1. \end{cases} \quad (1.59)$$

One can see that this solution is an odd function, and the same holds for all the d'Alembert solutions at $t = 3, 5, 7, \text{etc.}$ The reason for getting an odd function we suspect to be because the d'Alembert solution reflects off the boundaries in an odd way since we structured the solution as an odd-type extension, see (1.51).

In addition, there is a crucial fact to know about the exact d'Alembert solution which is that at any time which is a multiple of 4, i.e., at $t = 0, 4, 8, \text{etc.}$, the exact d'Alembert solutions are infinitely differentiable, infinitely smooth, for all the above initial conditions employed because, as mentioned above, the exact d'Alembert solution is of period 4 and has the returning property to the initial conditions; the solutions are just the functions $f(x)$ of the initial conditions, which are infinitely differentiable in $[-1, 1]$. Furthermore, at times that are a multiple of 2 but not of 4, i.e., at $t = 2, 6, 10, \text{etc.}$, the exact d'Alembert solutions are also infinitely differentiable, infinitely smooth, since the characteristics touch the boundaries at these specific times, as given above, meaning that the discontinuities go away again then. Indeed, using (1.46) and (1.51) one can see that the exact d'Alembert solution is the function $-f(-x)$.

To illustrate the above a space-time plot of the second derivative of the exact d'Alembert solution (1.46) and its odd/even components for one initial condition, the C^1 IC (1.55), is given in Figure 1.6. This figure demonstrates that indeed the discontinuities move along characteristics into the interior of domain, which originally occurred only at the boundaries $x = \pm 1$ now they move into the interior at later times t . One can see that there is a clash in the figure along characteristics, where the shape and colours of the figure change, while at the special even times this does not occur but rather the solution is smooth. To be more pedantic, these discontinuities is only in the even component, Figure 1.6b, not the odd, Figure 1.6a, which approve our explanation above. Consequently, they are in the solution figure as well, Figure 1.6c.

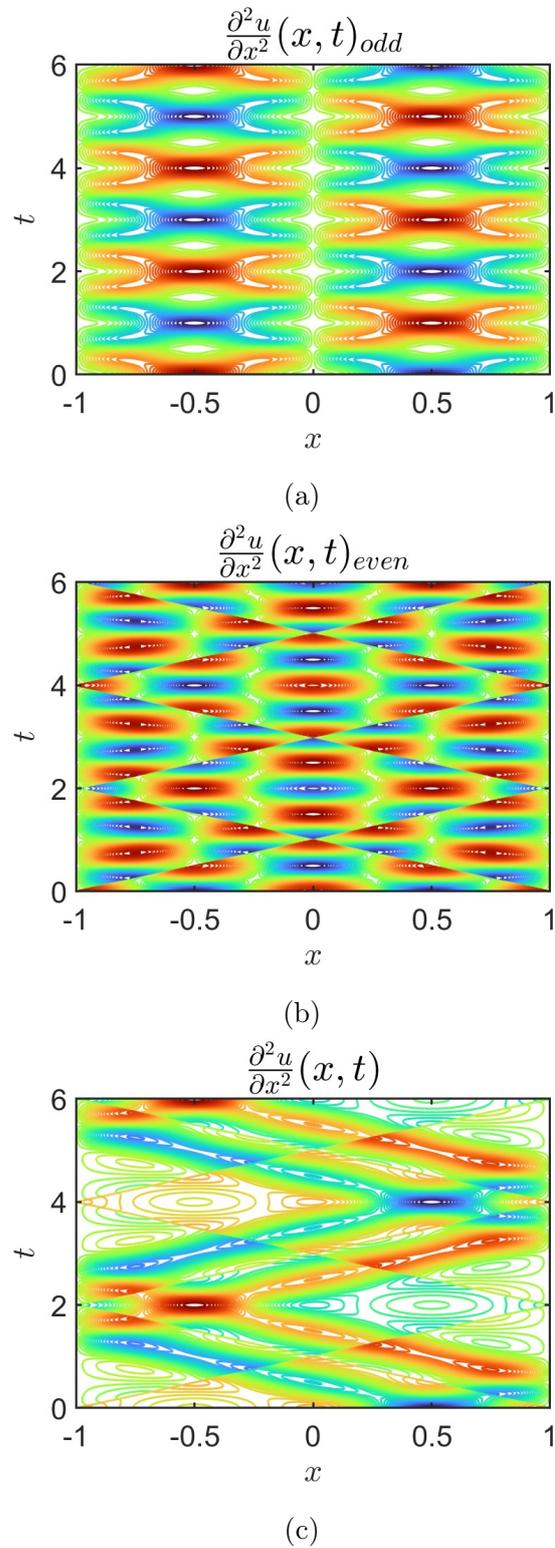


Figure 1.6: The contour of the second derivative of the exact d'Alembert solution (1.46), $\frac{\partial^2 u}{\partial x^2}(x, t)$, panel (c), and its components (odd component, $\frac{\partial^2 u}{\partial x^2}(x, t)_{odd}$, panel (a), and even component, $\frac{\partial^2 u}{\partial x^2}(x, t)_{even}$, panel (b)) for the linear wave equation (1.47). The initial condition used is the C^1 IC (1.55).

In later subsequent chapters, we will examine the impacts of all this on the numerical solutions and consequential results in detail; that is the impacts of the odd and even phenomena that is presented either in the ICs (consequently on their generated exact solutions) or the odd phenomena presented in the exact d'Alembert solutions themselves at $t = 1, 3, 5, \text{etc.}$, and the infinitely smooth of the exact d'Alembert solutions at $t = 0, 2, 4, 6, 8, \text{etc.}$

In this thesis we are only considering Dirichlet BCs, but one can implement other BCs as well such as Neumann BCs and mixed conditions, see Section 1.2. However, the behaviour of each is different than the others.

For Neumann BCs, $\frac{\partial u}{\partial x}(\pm 1, t) = 0$, the appropriate choice for the extension is sort of the even extension. Let us take a simple example of a wave equation with Neumann BCs,

$$\frac{\partial u}{\partial x}(-1, t) = 0, \quad \text{and} \quad \frac{\partial u}{\partial x}(1, t) = 0, \quad (1.60)$$

and initial conditions

$$u(x, 0) = h(x), \quad (1.61)$$

and

$$\frac{\partial u}{\partial t}(x, 0) = 0. \quad (1.62)$$

The extension of $h(x)$ must be even-like as one can see in Figure 1.7. The extension function formula will be different from $F(x)$, in (1.51), let us call the function here $H(x)$. Figure 1.7 shows that the original function $h(x)$ is extended to the whole domain. It shows that the function $h(x)$ is extended from the boundary point -1 to the interval $[-3, -1]$ and it is extended from the boundary point 1 to the interval $[1, 3]$. The difference with Dirichlet BCs is that in the case of Neumann BCs, the graph of H is symmetric in the lines $x = -1$ and $x = 1$. Now after getting the function $H(x)$ in $[-3, 1]$, or in $[-1, 3]$, one can then extend this to all infinite domain by repeatedly “cutting and pasting” its graph. This leads to a periodic function on the infinite domain of period 4; the function on the interval $[-3, 1]$ or $[-1, 3]$ is one period. Thus, this new function, $H(x)$, same as $F(x)$, is periodic of period 4.

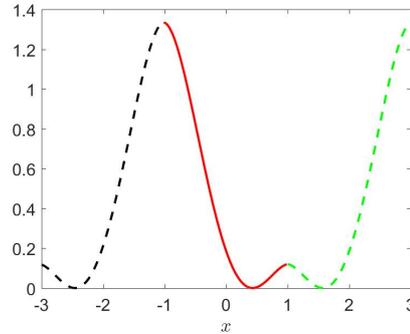


Figure 1.7: The function $H(x)$ in infinite domain, but here we take a part of the domain. It shows that the original function $h(x)$ is extended to the whole domain as even-type extension. The figure shows that the function $h(x)$ is in the interval $[-1, 1]$, the red line, then is extended from the boundary point -1 to the interval $[-3, -1]$, the black dashed line, similarly $h(x)$ is extended from the boundary point 1 to the interval $[1, 3]$, the green dashed line.

In this case all the even derivatives of the generated function $H(x)$ at the interface points, at $x = \pm 1$, will be continuous. However, the odd derivatives $H^{(n)}(x)$, for any odd number $n > 1$, is only continuous if again it is 0 at the endpoints, so unless the original $h(x)$ satisfies $h^{(n)}(x) = 0$ at $x = \pm 1$, $H^{(n)}(x)$ will not be continuous.

We anticipate that when using Neumann BCs the d'Alembert solution at $t = 1, 3, 5, \text{etc.}$, would probably be an even function, in contrast when using Dirichlet BCs as the solution is an odd function at these specific times. We imagine that would be again because the solution is shaped as an even-type extension. However, we are not certain as we have not investigated on this matter for Neumann BCs.

For mixed BCs, which are the combination of the previous two BCs, so at each end point will satisfy the BCs employed at that point, regarding the even or odd properties above. The extension function for the mixed BCs, when having Dirichlet BCs at one end point and Neumann BCs at the other end point, will turn out be of period 8, not 4 as for the previous two BCs.

For all these BCs the function $F(x)$ will always be continuous, as long as $f(x)$ satisfies the original boundary conditions, and $F'(x)$ will also always be continuous as well. One can say that our BCs case is the more complicated one than the ones mentioned above because we proved analytically that it has more severe level of

nonsmoothness as it will start having the discontinuity starting from the second derivative at both end points $x = \pm 1$, meanwhile, the other worst case is the Neumann BCs when it will start having the discontinuity from the third derivative.

Unfortunately, the above finite-interval d'Alembert solution is mainly for the linear wave equations and this solution is not applicable for our nonlinear wave equation, which is also on a finite domain. However, the discontinuity aspects presented in the linear wave equation are probably the same for this nonlinear wave equation, because the characteristics are the same as in Figure 1.5, i.e., both of the equations share the same characteristics equation

$$\frac{dX}{dt} = \pm c = \pm 1, \quad (1.63)$$

see (1.16), because they have the same speed $c = 1$. Noteworthy, the characteristics do not get affected by the nonlinear term that comes out from $f(x, y, u, u_x, u_y)$ in (1.1), see Section 1.3.

1.10 Separation of variables

We can also apply the method of separation of variables to find the exact solution of the linear wave equation (1.47) for the ICs (1.54) and (1.56).

The problem (1.47)–(1.50) is a linear homogeneous initial boundary value problem. Recall that at the first stage of the method, we compute nontrivial separated solutions of the PDE (1.47), i.e. solutions of the form

$$u(x, t) = F(x)G(t), \quad (1.64)$$

that also satisfy the boundary conditions (1.48). Here, as usual, F, G are functions of the variables x and t respectively. At this stage, we do not take into account the initial conditions (1.49) and (1.50).

Our boundary conditions are Dirichlet BCs at $x = \pm 1$ that we transform into BCs at $x = 0, \pi$ in order to get the standard Fourier sine transform formula. This can be done by considering the following assumptions.

If we define $y = \frac{1}{2}\pi(x + 1)$, then this implies $x = \frac{2}{\pi}y - 1$. Thus, the wave equation:

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2},$$

becomes

$$\frac{\partial^2 u}{\partial t^2} = \frac{\pi^2}{4} \frac{\partial^2 u}{\partial y^2}. \quad (1.65)$$

Let

$$w(t, y) = u(t, x),$$

so the new wave equation can be written as

$$\frac{\partial^2 w}{\partial t^2} = \frac{\pi^2}{4} \frac{\partial^2 w}{\partial y^2}, \quad (1.66)$$

see (1.65), with the new boundary conditions

$$w(0, t) = 0, \quad \text{and} \quad w(\pi, t) = 0, \quad (1.67)$$

and the initial conditions

$$w(y, 0) = f\left(\frac{2y}{\pi} - 1\right), \quad (1.68)$$

and

$$\frac{\partial w}{\partial t}(y, 0) = 0. \quad (1.69)$$

Now the separable solution is

$$w(y, t) = F(y)G(t). \quad (1.70)$$

Differentiating the separated solution (1.70) twice in y and twice in t , and then substituting these derivatives into the wave equation (1.66), we infer

$$F(y)G''(t) = \frac{\pi^2}{4} F''(y)G(t). \quad (1.71)$$

By separating the variables, we see that

$$\frac{F(y)}{F''(y)} = \frac{\pi^2}{4} \frac{G(t)}{G''(t)}, \quad (1.72)$$

assume that,

$$\frac{F(y)}{F''(y)} = \frac{\pi^2}{4} \frac{G(t)}{G''(t)} = -k^2, \quad (1.73)$$

where k is a constant. Equation (1.73) implies

$$F''(y) = -k^2 F(y), \quad (1.74)$$

$$G''(t) = -\frac{k^2 \pi^2}{4} G(t). \quad (1.75)$$

By implementing the BCs and ICs this leads to have the solution to be

$$w(y, t) = \sum_{k=1}^{\infty} A_k \sin(ky) \cos\left(\frac{\pi kt}{2}\right), \quad \text{for } k = 1, 2, 3, \dots \quad (1.76)$$

We get

$$w(y, 0) = f\left(\frac{2y}{\pi} - 1\right) = \sum_{k=1}^{\infty} A_k \sin(ky), \quad (1.77)$$

which is Fourier sine series, where

$$A_k = \frac{2}{\pi} \int_0^{\pi} f\left(\frac{2y}{\pi} - 1\right) \sin(ky) dy, \quad \text{for } k = 1, 2, 3, \dots \quad (1.78)$$

The material discussed above in this section is standard and can be found in many textbooks, such as (Pinchover & Rubinstein, 2005).

For the particular initial condition (1.54), we have $u(x, 0) = f(x) = \sin(\pi x)$, the exact solution can be found explicitly as follows

$$\sin(\pi x) = \sin\left(\pi\left(\frac{2y}{\pi} - 1\right)\right) \quad (1.79)$$

$$= \sin(2y - \pi) \quad (1.80)$$

$$= -\sin(2y). \quad (1.81)$$

Substituting $f(x)$ into (1.76) implies

$$w(y, 0) = \sum_{k=1}^{\infty} A_k \sin(ky) = -\sin(2y), \quad (1.82)$$

it follows that

$$A_2 = -1, A_k = 0, \quad \forall k \neq 2.$$

So, the solution (1.76) becomes

$$w(y, t) = -\sin(2y) \cos(\pi t), \quad (1.83)$$

and from (1.81) and (1.83) the final solution can be written as

$$u(x, t) = \sin(\pi x) \cos(\pi t). \quad (1.84)$$

Hence, this exact solution is infinitely smooth as we have shown also in Section 1.9.

For the polynomial initial condition (1.56), the calculation is more difficult and by using Maple we get that

$$A_k = \frac{2176k\pi^2}{11(k^2\pi^2 + 4)^5} \left((-1)^k e \left(k^4\pi^4 + \frac{184}{17}\pi^2k^2 - \frac{304}{17} \right) + \frac{7e^{-1}}{17} \left(\pi^4k^4 + \frac{200}{7}k^2\pi^2 + 208 \right) \right), \quad (1.85)$$

as $k \rightarrow \infty$, this can be simplified to become

$$A_k = Ck^{-5} + \mathcal{O}(k^{-7}), \quad (1.86)$$

where C is a constant. Thus, the exact solution is not infinitely smooth but only C^3 as we have discussed also in Section 1.9.

1.11 Energy conservation

Moving now to the nonlinear wave equations, the full exact solution $u(x, t)$ of the nonlinear wave equations is not known generally, therefore, numerical methods usually are used to find their approximate solutions. The nonlinear equations are often far more challenging than linear equations even in finding their numerical solutions, in many ways. For example, solving the linear wave equations numerically using something like backward Euler is usually no harder than forward Euler, but for nonlinear equations it immediately presents the challenge that the equations to be solved for the new values at the next time step are nonlinear and thus will require some special solver all on their own (such as the Newton-Raphson that we are using in this thesis).

The particular nonlinear equation that we study is the u^3 extension of the previous linear wave equation, $\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3$, see (1.24). The full exact solution $u(x, t)$ of this nonlinear wave equation is not known. The previous analytic solution no longer applies but the discontinuity aspects are probably the same, however, this nonlinear wave equation also has an analytic result, namely the energy conservation, which serves as a useful diagnostic quantity to explore the accuracy of numerical solvers.

The theoretical energy conservation proof for the general nonlinear wave equation is shown next:

Let us consider the semi-linear wave equation of the form

$$u_{tt} = u_{xx} - f(u), \quad (1.87)$$

where $f(u)$ and $u(x, t)$ are smooth functions. (Leimkuhler & Reich, 2004) proved that this wave equation is Hamiltonian, conservative, under specific prescriptions of boundary and initial data. They imposed periodic boundary conditions with period L . The general energy functional is given as

$$E(u) = \int_0^L \left[\frac{1}{2}u_t^2 + \frac{1}{2}u_x^2 + F(u) \right] dx, \quad (1.88)$$

where $F(u) = \int_0^u f(s)ds$.

In our equation the interval is $[-1, 1]$ and $f(u) = u^3$, see equation (1.24); this implies that $F(u) = \frac{1}{4}u^4$. Hence, (1.88) becomes

$$E(u) = \int_{-1}^1 \left[\frac{1}{2}u_t^2 + \frac{1}{2}u_x^2 + \frac{1}{4}u^4 \right] dx. \quad (1.89)$$

After few steps this leads to

$$\frac{d}{dt}E(u) = [u_x u_t]_{-1}^1 = 0. \quad (1.90)$$

This shows that the energy is conserved for this nonlinear wave equation; this is given in more detail in Subsection 4.1. The energy is preserved, in other words, it is constant while increasing the time.

1.12 Detailed Motivation

Section 1.9 shows that we can obtain the analytical solution for our linear wave equation, moreover, we can take the analysis even further as a logical consequence of these equations, (1.46) and (1.51). We can show, at least for the linear wave equation, that the equation itself will develop discontinuities in certain higher derivatives, in other words, we have this systematic way of constructing solutions that have built into them different levels of smoothness even if the ICs are infinitely differentiable on the bounded interval. In fact, we can systematically control the level of which even derivative do we want that discontinuity to show up, i.e., we can control precisely how smooth or not smooth we want our solutions to be, either at the second derivative or fourth derivative or higher. To the best of

our knowledge the literature does not consider this discontinuity issue that we demonstrate for the finite d'Alembert solution, for example see (Daileda, R. C., 2014; Grigoryan, V., 2011) who present the finite-interval d'Alembert solution but not the implications it has for discontinuities in higher derivatives. Section 1.11 shows that our nonlinear wave equation also has an analytic result, namely the energy conservation, which serves as a useful diagnostic quantity to explore the accuracy of numerical solvers. Therefore, the analytical results we have in Sections 1.9 and 1.11 for our two wave equations are ideal test cases to systematically explore a variety of numerical methods, to see how well they can cope with something as challenging as discontinuities in higher derivatives, and how well they conserve a global quantity like energy.

Consequently, we combine the property of constructing solutions, with different levels of smoothness, with the numerical analysis aspects. Combine it with the numerical analysis concepts that smoothness of solutions is an incredibly important thing in spectral methods and a bit less important in finite element methods. Accordingly, to take some of these solutions and load them into some of these numerical methods and see what does that really do to the behaviour? How well or not well do the different numerical methods then still perform and to systematically compare it. In view of the point above that smoothness of the solutions is incredibly important in studying any numerical method, it is therefore of interest to conduct a systematic comparison study of a variety of methods, all applied to particular initial conditions to this linear wave equation that have been carefully chosen to yield solutions with different levels of smoothness. And to study its impacts on solving any nonlinear wave equations, in particular on our nonlinear equation studied in this thesis. We strongly believe that this will have big impacts on the spectral methods as this method demands the solution to be smooth. However, for the finite elements methods it may not have such an important role as the FEM have low order of convergence; the convergence requirements are only just few levels of smoothness such as C or C^1 or C^2, \dots, C^r , for a limited integer number r not infinite (see Babuska *et al.*, 2010; Huebner *et al.*, 2001; Rao, 2005). So, we will study how badly will the convergence be degraded especially for the last two ICs, and whether the algebraic convergence order will differ. Accordingly, study the impact of this on stability and energy. Most literature addresses different cases than us, they consider either very smooth problems directly, infinitely smooth, or they consider

some simple nonsmooth problems, mainly for linear problems, either the initial data or the solution itself are nonsmooth, not its higher derivatives, see Section 1.6. This draws attention to the need for someone to take a step back and study all this in a systematic comparison.

The numerical methods used here are for how to discretise in space and for how to advance in time, which are almost completely separate and pretty much any spatial discretisation method can be combined with pretty much any timestepping method. This leads to have two kinds of errors, the discretisation errors in space and time, and both contribute to the numerical errors.

The discontinuity that occurs for x in the linear wave equation will also occur for t because $u_{tt} = u_{xx}$. For example, for the second IC, when u_{xx} is discontinuous then u_{tt} will also be discontinuous, similarly for the third IC, when u_{xxxx} is discontinuous then by using $u_{tt} = u_{xx}$ twice, it will end up with u_{tttt} to be necessarily having exactly the same discontinuity too. So, we assume that for all the three spectral methods used, the pseudospectral, the tau and the Galerkin methods, the lack of smoothness could affect both the spatial discretization accuracy and the time stepping accuracy. Bearing in mind that the general expectation for some time stepping methods is to be of the first order, such as the Euler's forward and backward methods, and ETD method, and others to be of the second order, such as the modified Euler method, trapezoidal method and ETD2RK method, regarding the time stepping error.

All the standard theory for the spectral methods assumes infinite level of smoothness of the solutions, i.e., the function to be infinitely differentiable and continuous with all its derivatives on an interval. The theoretical expectation for convergence is to be exponential for the smooth functions. Admittedly, it is known that the spectral accuracy will drop off from exponential convergence for the smooth functions, to be only algebraic convergence for the nonsmooth functions, see (Boyd, 2001; Canuto *et al.*, 1988, 2006; Hesthaven *et al.*, 2007; Trefethen, 2000). Therefore, we believe that the accuracy related to spatial error to be changed very significantly. This is because all of these accuracy results depend crucially on various assumptions about the level of smoothness of the expected solutions. If solutions are less smooth than the theory expects, the accuracy will be less just because of the extremely close connection between "degree of smoothness" and "efficiency of convergence". The first IC (1.54), which is infinitely differentiable and continuous at any derivative

in $[-1, 1]$, is expected to be exponentially convergent for all the time stepping methods applied. While the other two ICs, (1.55) and (1.56), they should drop off only algebraically, respectively, because at any time there is a discontinuity in the second and fourth derivatives, respectively, in the interior of the interval somewhere.

As a result of the changes expected above regarding the convergence we assume that the stability of our wave equations will also be affected by the discontinuity of the expected solutions, mainly for the trapezoidal method and ETD2RK as the other methods are not stable.

We believe that the discontinuities presented in the exact solutions of the linear wave equation using our ICs are very likely to be presented in the exact solutions of our nonlinear wave equation as well. Although we do not know the exact solutions of our nonlinear wave equation but we believe that the nonsmoothness presented in the linear wave equation will still have a significant role in the exact solution of the nonlinear wave equation too. Accordingly, the smoothness level will be a crucial part in our numerical study too for solving the nonlinear wave equation, which are using the pseudospectral method only. Nevertheless, all the numerical properties for the nonlinear wave equation are far more challenging than of the linear equations. The theoretical studies in Section 1.11 shows that our nonlinear wave equation has an analytic result, namely the energy conservation, the energy is conserved. Therefore, we will study whether our numerical schemes are also having energy preserved or not; this will help to study the stability of those methods. The nonsmoothness of the solutions produced by the last two ICs, (1.55) and (1.56), we expect to be so crucial in studying the energy conservation, mostly regarding those two ICs. The expectation is that those last two ICs will require larger number of modes than the first IC requires in order to get the energy conserved because as we said we expect the convergence for those two ICs to drop off algebraically and because for the IC (1.55), where u_{xx} and u_{tt} are discontinuous, that means u_x and u_t would have graphs with corners. That in turn means that evaluating integrals of $(u_x)^2$ and $(u_t)^2$ in (1.89) via the trapezoid rule is inappropriate, and would require a large number of points to obtain accurate results. So one can say that, our nonlinear equation can have accuracy, stability and energy conservation properties that are significantly more complicated and challenging than the linear wave equation. For example, something like exponential integrator methods for the linear wave equation should give exact solution, or let us say the exact energy value

regardless of the nonsmoothness, but for the nonlinear problem the nonsmoothness will have a major role in studying the energy for the numerical methods. It is thus important to also consider nonlinear equations in the comparison study that we are doing in this thesis. The second method used here to discretise in space is finite elements methods, FEM. Generally, FEM have low order of convergence so we expect it will not suffer much from the lack of smoothness, we anticipate a minor impact, and we will see how minor it is. We think the error here is very different regarding the convergence with respect to spatial discretization than for the spectral methods. The theoretical expectation convergence is to be algebraic for all methods for all the ICs used. In contrast to the spectral methods, we expect that the convergence will remain algebraic for all the ICs because smoothness of the solution does not have a big role here. Therefore, the smoothness of the ICs has no big part in studying the stability and energy conservation as well. In comparing the different methods against each other, for spectral methods the three initial conditions will yield very different rates of convergence, whereas for finite elements the three initial conditions will be much more similar.

All these special properties of the linear and nonlinear wave equations make these equations ideal test cases for a variety of numerical methods, to systematically see how well they can cope with solutions that have these discontinuities in their various derivatives.

1.13 Outline of the thesis

Our starting point is in **Chapter 2**; this chapter is a continuation of the previous chapter in terms of giving the general basic concepts. The stability region and truncation error are explained, also some more concepts, such as Hamiltonian and symplecticity, which are used to assess the stability of the nonlinear PDE in the later chapters.

In **Chapter 3** spectral approximations based on Chebyshev series are used to solve the linear wave equation (1.20) with Dirichlet BCs using our three infinitely differentiable ICs on the interval $[-1, 1]$. The discretization of the equation in space first, then in time, procedure is followed here. In physical space (the pseudospectral method) and spectral space (the collocation, the tau and the Galerkin methods)

are implemented, then different time stepping methods are used. The convergence and the stability of each method is studied. **It is at this point when the effects of the levels of nonsmooth solutions of the problem generated by ICs start to be obvious in terms of convergence and stability.**

Chapter 4 is considering the nonlinear wave equation (1.24) following the same procedure as in the previous chapter. This chapter outlines the numerical solution using the pseudospectral method for this PDE to study the accuracy, such as, the convergence and stability. In addition, we investigate the energy conservation for this problem, so the Hamiltonian properties of our PDE and the symplectic properties of each time stepping method is studied; this helps to study the stability of each time stepping method. **Despite the existing literature in terms of the nonsmoothness aspect that is mostly for the linear problems, and for the solution itself not for its derivatives; the present approach is distinguished from prior work though by investigating the nonsmoothness of the derivatives of the expected solutions for the nonlinear problem, starting from this chapter.**

Additionally, some other time stepping methods are employed in **Chapter 5** to study the accuracy for the linear equation (1.20) and nonlinear wave equation (1.24) in the previous chapters, in the light of the nonsmooth level of solutions. The methods considered are exponential integrators methods, mainly ETD and ETD2RK methods.

Next, in **Chapter 6** other spatial discretization methods are employed to explore all the above numerical investigation for the nonlinear problem; the methods considered are finite element methods, as these methods do not require high level of smoothness of the problem as for the spectral methods.

Finally, general conclusions and suggestions for possible future work are given in **Chapter 7**. A systematic comparison between all the methods explored in the previous chapters is made. This comparison is based on the nonsmoothness level of the solutions when two spatial discretization methods are used, spectral and finite elements methods, each of which has different smoothness level requirements.

Chapter 2

General concepts

Numerical methods offer approximate solutions, with specific errors, to mathematical problems. One needs to determine accurate approximations with minimal effort and small error values. Therefore, we need a means for comparing the efficiency of various approximation methods.

In the previous chapter the crucial role that initial value problems (IVPs) play in the design and analysis of partial differential equations (PDEs) is mentioned. Time stepping methods are given there to solve IVPs. The general goal of every time stepping method is to obtain approximations to the solution of IVPs.

We are studying the stability of some PDEs in the next chapters. Accordingly, it is important to know the stability of each numerical method used to solve IVPs and the error of its approximate solution. Hence, the stability region and truncation error are explained in this chapter. Also, some more concepts are explained, such as Hamiltonian and symplecticity, which are used to assess the stability of the nonlinear PDE in the later chapters.

The material discussed in this chapter is standard and can be found in many textbooks, such as, ([Atkinson, 1989](#); [Burden & Faires, 2001, 2011](#)).

2.1 Truncation errors

In this section, we shall consider the truncation error of the time stepping methods, mentioned in previous chapter, used to solve first order IVPs (initial value problems) of the form

$$y' = f(y, t), \quad \text{for } a \leq t \leq b \quad \text{with initial condition } y(a) = \alpha. \quad (2.1)$$

In order to find the truncation errors the given interval $[a, b]$ is divided into N intervals of equal length and the mesh points,

$$t_i = a + ih, \quad \text{for each } i = 0, 1, 2, \dots, N,$$

are defined, where N is a chosen positive integer number, $h = (b - a)/N = t_{i+1} - t_i$ and i is the step index.

The material described in this section is standard and can be found in many textbooks, such as, (Burden & Faires, 2001), (Burden & Faires, 2011) and (Süli & Mayers, 2003).

2.1.1 Global truncation error

The global truncation error measures the accuracy of the method over the entire range of the approximations

$$e_{i+1}(h) = y(t_{i+1}) - u_{i+1}, \quad \text{for } i = 0, 1, \dots, N - 1,$$

where $y(t_{i+1})$ is the exact solution and u_{i+1} is the approximate solution of the initial value problem at t_{i+1} .

For example, the global error of the explicit and the implicit Euler method is $\mathcal{O}(h)$, so we call them first order, and of the modified Euler and the trapezoidal method is $\mathcal{O}(h^2)$, meaning that they are second order. We will be using the global truncation error in solving our PDEs using the time stepping methods described before.

A numerical method is **convergent** if the global truncation error goes to zero as $h \rightarrow 0$, i.e.

$$\lim_{h \rightarrow 0} \max_n |e_n| = 0,$$

in other words, the numerical solution converges to the exact solution. All methods, Euler's, the implicit Euler, the modified Euler and the trapezoidal method, are convergent.

2.2 Absolute stability region

It is very important to know the stability region for time stepping methods when we solve PDEs in the next chapters to know when the method is stable in order to avoid the unstable solution which will be outside the stability region. We say that a method is **unstable** if errors grow rapidly.

Indeed, we will see later that we have stability problems when solving PDEs; this manifests itself by rapidly growing errors. Knowing the stability region helps to avoid this instability.

Suppose we have the linear equation,

$$y' = \lambda y, \quad t \geq 0, \quad y(0) = y_0, \quad (2.2)$$

where $\lambda \in \mathbb{C}$. The exact solution of this equation is $y(t) = y_0 e^{\lambda t}$, hence $\lim_{t \rightarrow \infty} y(t) = 0$ if and only if $\operatorname{Re} \lambda < 0$. If a given numerical method is applied with a constant step size $h > 0$ to (2.2), then the linear stability domain R of this method is the set of all numbers $\lambda h \in \mathbb{C}$ such that,

$$\lim_{n \rightarrow \infty} u_n = 0. \quad (2.3)$$

The expressions for the results of the numerical solution using Euler's method (1.27) are,

$$u_n^E = (1 + \lambda h)^n u_0, \quad \text{after } n \text{ iterations.}$$

When $\lambda < 0$ and $h > 0$, we have the condition,

$$|1 + \lambda h| < 1,$$

which comes from (2.3). This condition is satisfied if and only if $0 < h|\lambda| < 2$. This gives that the sequence $|u_n^E|$ decreases monotonically, if $h < \frac{2}{|\lambda|}$. For this reason, the absolute region of the explicit Euler's method is,

$$R_{Euler}^E = \{Z = \lambda h \in \mathbb{C} : |1 + Z| < 1\}, \quad (2.4)$$

For the modified Euler method (1.28) we have,

$$u_n = \left(1 + \lambda h + \frac{1}{2}\lambda^2 h^2\right)^n u_0, \quad \text{after } n \text{ iterations,}$$

and the stability region for it is,

$$R_{Euler}^M = \{Z = \lambda h \in \mathbb{C} : |1 + Z + \frac{1}{2}Z^2| < 1\}. \quad (2.5)$$

If Z is real, then the method is stable if $Z \in (-2, 0)$ and unstable otherwise.

The expressions for the results of the numerical solution using the implicit Euler method (1.29) are,

$$u_n^I = (1 - \lambda h)^{-n} u_0, \quad \text{after } n \text{ iterations.}$$

When $\lambda < 0$ and $h > 0$, we have

$$\left| \frac{1}{1 - \lambda h} \right| < 1,$$

therefore, the sequence $|u_n^I|$ decreases monotonically with increasing n .

The stability region of the implicit Euler method is,

$$R_{Euler}^I = \{Z = \lambda h \in \mathbb{C} : |1 - Z| > 1\}. \quad (2.6)$$

If Z is real, then it is unstable if $Z \in [0, 2]$ and stable otherwise.

In addition to the previous methods, the expression of the trapezoidal method (1.35) has to be described, which is,

$$u_n = \left(\frac{1 + \frac{\lambda h}{2}}{1 - \frac{\lambda h}{2}} \right)^n u_0, \quad \text{after } n \text{ iterations,}$$

and the absolute region is,

$$R_{Trapezoidal} = \{Z = \frac{\lambda h}{2} \in \mathbb{C} : \operatorname{Re}(Z) = \frac{h}{2}\operatorname{Re}(\lambda) < 0\}, \quad (2.7)$$

(see in Iserles, 1996; Süli & Mayers, 2003).

2.3 Stiff differential equations

We explain stiff equations now because we will use it later when we solve PDEs with spatial discretization methods, which always leads to stiff equations.

For some initial value problems the exact solution includes terms that decay to zero much more rapidly than other. For instance, consider the system of differential equations,

$$\vec{y}' = D\vec{y}, \quad \vec{y}(0) = \vec{y}_0, \quad \text{where } D = \begin{pmatrix} -100 & 1 \\ 0 & -\frac{1}{10} \end{pmatrix}. \quad (2.8)$$

The exact solution of (2.8) is

$$\vec{y}(t) = e^{-100t}\vec{x}_1 + e^{-t/10}\vec{x}_2, \quad t \geq 0, \quad (2.9)$$

where \vec{x}_1 and \vec{x}_2 are two vectors dependent on y_0 but not on t .

This system is an example of a stiff problem, because the function $g(t) = e^{-100t}$ decays very fast, while the decay of $e^{-t/10}$ is more sedate. In order to solve it numerically, the step size h should be chosen carefully such that it is in the stability region. For example, solving the system by explicit Euler's method, the value of h depends on $|\lambda|$. For this example $\lambda = -100$ and $-\frac{1}{10}$, the smallest value of h should be taken (here $h < \frac{2}{100}$ and $h < 20$) and this is not very useful because h has to be very small. The technique commonly used for stiff systems are implicit methods, such as, the implicit Euler method or the trapezoidal method, to avoid such small h . The methods are used with Newton iteration (see subsections 1.7.3 and 1.7.4). The stability region for these methods does not restrict the choice of h , as it does for Euler's method, because λh is negative so it is always in the stability region; (2.6) and (2.7). Instead, the choice of h is governed by the required accuracy and how long the method takes (Burden & Faires, 2001, 2011; Iserles, 1996; Süli & Mayers, 2003).

2.4 Hamiltonian ODEs

This section gives an introduction to Hamiltonian problems for ODEs. We discuss this here because we will be studying some Hamiltonian PDEs in this thesis.

Hamiltonian ODEs form the basis of the study of Hamiltonian PDEs, which will be taken up in Section 4.4. We will subsequently use this to analyze the stability of spectral methods for the nonlinear wave equation. There are many ways to study the stability of numerical methods for the nonlinear wave equation, one method is to use the Hamiltonian structure.

We start by characterizing the class of problems we shall be concerned with and by presenting some notation. Given a phase space $\mathbb{R}^d \times \mathbb{R}^d$ of the points,

$$(p, q) := (p_1, p_2, \dots, p_d, q_1, q_2, \dots, q_d),$$

and an arbitrary (smooth) Hamiltonian function $H : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, then the **Hamiltonian system** of differential equations with this Hamiltonian H is given by,

$$\frac{dp}{dt} = -\frac{\partial H(p, q)}{\partial q}, \quad (2.10)$$

$$\frac{dq}{dt} = +\frac{\partial H(p, q)}{\partial p}. \quad (2.11)$$

When discussing Hamiltonian systems, it is often convenient to use the notation,

$$z := (p, q)^T,$$

with $p, q \in \mathbb{R}^d$, $z \in \mathbb{R}^{2d}$ and to introduce the $2d \times 2d$ structure matrix J ,

$$J := \begin{pmatrix} 0 & +I_d \\ -I_d & 0 \end{pmatrix}, \quad (2.12)$$

(I_d and 0 respectively represent the identity and zero $d \times d$ matrices).

Then the Hamiltonian system (2.10)-(2.11) can be rewritten in compact form,

$$\frac{dz}{dt} = J \frac{\partial H(z)}{\partial z}. \quad (2.13)$$

If $z(t)$ is a solution of (2.13), then $H(z(t))$ is constant. In other words, H is a conserved quantity. In applications, this usually corresponds to **conservation of energy** (this writing uses the ideas from the textbooks [Bridges & Reich, 2006](#); [Hairer *et al.*, 2002](#); [Leimkuhler & Reich, 2004](#); [Sanz-Serna & Calvo, 1994](#)).

2.5 Symplectic maps

We need to introduce the concept of flow map. So, let us take general systems of differential equations (2.1), which is,

$$y' = f(y, t), \quad (2.14)$$

where y represents a point in the phase space. The flow map of the given system (2.14) is the mapping $\Phi_f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ which to any point y_0 in the phase space associates the value $y(t)$ of the solution with initial value $y(0) = y_0$:

$$y(t) = \Phi_f(t, t_0)y_0. \quad (2.15)$$

A smooth map Ψ on the phase space \mathbb{R}^{2d} is called a **symplectic map** if its Jacobian $\frac{\partial \Psi}{\partial Z}(Z)$ satisfies,

$$\left[\frac{\partial \Psi}{\partial Z}(Z) \right]^T J^{-1} \frac{\partial \Psi}{\partial Z}(Z) = J^{-1} \quad (2.16)$$

for all Z in the domain of definition of Ψ , where J is given in (2.12).

For a linear transformation $\Psi(Z) = AZ$, the condition of symplecticity reduces to $A^T J^{-1} A = J^{-1}$ and A is referred to as a **symplectic matrix**.

The flow map of a Hamiltonian system is always symplectic. Symplecticity is a characterization of Hamiltonian systems in terms of their solutions, rather than in terms of the actual form of the differential equations.

A one-step numerical method generates a smooth mapping $\Psi_f(t_{n+1}, t_n)$ that effects the step from one time level to the next,

$$u_{i+1} = \Psi_f(t_{i+1}, t_i)u_i, \quad (2.17)$$

where as before, u_i denotes the numerical approximation at time t_i to the value $y(t_i)$ of a solution of (2.14). For example, the explicit Euler method is given by (1.27), so the map Ψ_f for the explicit Euler method is given by

$$\Psi_f(t_{i+1}, t_i)u_i = u_i + hf(u_i, t_i).$$

A method is said to be symplectic if its map Ψ_f is symplectic whenever the method is applied to a Hamiltonian ODE. Euler's, the implicit Euler and the modified Euler methods are not symplectic methods. However, the implicit midpoint method is a

symplectic method.

The symplecticity requirement might be too strong if we are interested in a correct long-time behaviour of a numerical method. (Stoffer, 1988) suggests considering methods that are not necessarily symplectic but **conjugate** to a symplectic method. Two numerical methods $\tilde{\Psi}_h(y)$ and $\Psi_h(y)$ are mutually **conjugate**, if there exists a transformation $\chi_h(y)$, such that,

$$\tilde{\Psi}_h = \chi_h^{-1} \circ \Psi_h \circ \chi_h, \quad (2.18)$$

assuming that $\chi_h(y) = y + \mathcal{O}(h)$ uniformly for y varying in a compact set (Hairer *et al.*, 2002). The most prominent pair of conjugate methods are the trapezoidal rule, (1.35), and the implicit midpoint method, (1.40).

The implicit midpoint method is symplectic, hence one can say that the trapezoidal method is conjugate symplectic. Both, symplectic methods and conjugate symplectic methods have good energy conservation. This leads to say that, the trapezoidal method has good energy conservation.

The information in this section is taken from (Hairer *et al.*, 2002; Leimkuhler & Reich, 2004; Sanz-Serna & Calvo, 1994).

We will come back to these Hamiltonian and symplecticity concepts in the later Chapters, where we study a nonlinear wave equation. It will be used in those Chapters to assess the stability of numerical methods, especially the trapezoidal method, by studying energy conservation, which is closely related to the energy method in Section 1.8.

Chapter 3

Linear PDE with Dirichlet boundary conditions

All ideas mentioned in the previous chapter are used in this chapter and the next chapters to find numerical solutions to some partial differential equations (PDEs). We will use spectral methods, which are global methods where the computation at any given point uses not only information at neighbouring points, but information in the whole domain.

At the heart of spectral methods is the fact that any nice enough function $u(x)$ can be expanded in the form,

$$u(x) = \sum_{k=0}^{\infty} \tilde{u}_k \phi_k(x), \quad (3.1)$$

where \tilde{u}_k are the expansion coefficients and $\phi_k(x)$ are the global basis functions, in this thesis trial functions $\phi_k(x)$ are Chebyshev polynomials. If $u(x)$ is periodic then it is better to use Fourier series, where the ϕ_k are trigonometric functions (sine and cosine) (Hesthaven *et al.*, 2007). However, if $u(x)$ is not periodic, then it makes sense to use a non-periodic basis, such as a polynomial basis. Spectral methods may be seen as a development of the class of discretization schemes for differential equations known generally as the method of weighted residuals (**MWR**) (see Canuto *et al.*, 1988, 2006). The key elements of **MWR** are the trial functions, also called the expansion or approximation functions, and the test functions, also known as weight functions.

The trial functions are used as the basis functions for a truncated series expansion of

the solution. The test functions are used to guarantee that the differential equation is satisfied as accurately as possible by the truncated series expansion.

The main techniques used to determine the expansion coefficients \tilde{u}_k are the pseudospectral, collocation, tau and Galerkin methods. In all cases, we consider the residual $R(x)$ when an expansion is substituted into the governing equation. We want to keep the residual as small as possible across the domain while satisfying the boundary conditions. The residual will not vanish everywhere. Test functions, therefore, are selected so that the residual function $R(x)$ is minimized, e.g., the weighted average of the residual over the domain of interest is set to zero.

The choice of the trial functions is the feature which distinguishes spectral methods from finite-element and finite-difference methods. The functions for spectral methods are infinitely differentiable global functions. The choice of test functions distinguishes between the most commonly used spectral schemes, the pseudospectral, collocation, tau and Galerkin methods.

The pseudospectral and collocation methods: The test functions are represented by delta functions at special points, called collocation points. These approaches require the differential equation to be satisfied exactly at the collocation points, make the residual zero at the collocation points. However, the pseudospectral method is in physical space while the collocation method is in transform space.

The tau method: The test functions are the same as trial functions. However, none of the test functions need satisfy the boundary conditions. The boundary conditions are enforced by an additional set of equations. This method requires that the residual be orthogonal to the basis functions.

The Galerkin method: The test functions are the same as trial functions. This method requires to combine the original basis functions into a new set in which all the functions satisfy the boundary conditions. Then the method requires that the residual be orthogonal to these new basis functions.

This chapter is about solving PDE with Dirichlet boundary condition using spectral approximations, mainly using Chebyshev series. In this chapter one linear equation, namely the wave equation, is solved in physical space (the pseudospectral method) and spectral space (the collocation, tau and Galerkin methods).

Most of material discussed in this chapter is standard and can be found in many text books, such as, ([Canuto *et al.*, 1988, 2006](#); [Fornberg, 1998](#); [Fox & Parker, 1968](#); [Hesthaven *et al.*, 2007](#)).

3.1 Spectral approximation

In order to study the spectral approximation, the following theorem must be introduced.

The projection theorem: Let H be a Hilbert space and let M be a finite-dimensional subspace of H . For every $v \in H$ there is a unique $p \in M$ such that

$$\|v - p\| = \min_{m \in M} \|v - m\|$$

if and only if M is a closed subspace of H . We also call p the projection of v onto M and write

$$p = \text{proj}_M v.$$

If $\{m_1, m_2, \dots, m_n\}$ is a basis for M , then

$$\text{proj}_M v = \sum_{i=1}^n x_i m_i,$$

where the vector x satisfies the matrix-vector equation

$$Gx = b, \text{ with } G_{ij} = \langle m_j, m_i \rangle \text{ and } b_i = \langle m_i, v \rangle.$$

If the basis is orthogonal, then the best approximation to v from M is

$$\text{proj}_M v = \sum_{i=1}^n \frac{\langle m_i, v \rangle}{\langle m_i, m_i \rangle} m_i,$$

see (Gockenbach, 2002).

Let a family of orthogonal functions $\{\phi_k\}$ form a basis of the infinite-dimensional Hilbert space $L_w^2(a, b)$ of real functions of one variable, $u(x)$, defined on the finite interval (a, b) , with the inner product

$$\langle f, g \rangle = \int_a^b f(x)g(x)w(x) dx, \quad (3.2)$$

where w is the weight function (the standard book (Canuto *et al.*, 1988) has defined this term).

By the projection theorem the function $u(x)$ can be represented by its orthogonal projection on the basis $\{\phi_k(x)\}$ as

$$u(x) = \sum_{k=0}^{\infty} \tilde{u}_k \phi_k(x)$$

with the coefficients,

$$\tilde{u}_k = \frac{\langle \phi_k, u \rangle}{\langle \phi_k, \phi_k \rangle}. \quad (3.3)$$

The material discussed in this section is standard and can be found in many text books, such as (Boyd, 2001; Canuto *et al.*, 1988, 2006; Hesthaven *et al.*, 2007; Trefethen, 2000).

3.1.1 Chebyshev polynomials

The Chebyshev polynomials of first kind of degree k are defined as,

$$T_k(x) = \cos(k \cos^{-1}(x)) = \cos(k\theta), \quad k = 0, 1, \dots, \quad (3.4)$$

where $x = \cos(\theta)$, $-1 \leq x \leq 1$ and $0 \leq \theta \leq \pi$.

Alternatively, the Chebyshev polynomials may be defined by the following recurrence relation. Set $T_0(x) = 1$ and $T_1(x) = x$, then the other polynomials are defined by

$$T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x). \quad (3.5)$$

This follows from (3.4) by using the trigonometric relation

$$\cos((k+1)\theta) + \cos((k-1)\theta) = 2 \cos(\theta) \cos(k\theta).$$

The Chebyshev polynomials (3.4) are the eigenfunctions of the singular Sturm-Liouville problem

$$\left(\sqrt{1-x^2} T_k'(x) \right)' + \frac{k^2}{\sqrt{1-x^2}} T_k(x) = 0.$$

They form an orthogonal basis of $L_w^2(-1, 1)$, where the weight function of Chebyshev polynomial is $w(x) = \frac{1}{\sqrt{1-x^2}}$. Specifically,

$$\langle T_l, T_k \rangle = \int_{-1}^1 \frac{1}{\sqrt{1-x^2}} T_l(x) T_k(x) dx = \frac{\pi}{2} (1 + \delta_{0k}) \delta_{lk}, \quad (3.6)$$

where

$$\delta_{lk} = \begin{cases} 1 & \text{if } l = k, \\ 0 & \text{otherwise.} \end{cases} \quad (3.7)$$

The previous equation (3.6) can be rewritten as

$$\langle T_k, T_p \rangle = \begin{cases} 0 & \text{if } k \neq p, \\ \pi & \text{if } k = p, k = 0, \\ \frac{\pi}{2} & \text{if } k = p, k \neq 0. \end{cases} \quad (3.8)$$

We define γ_k by

$$\gamma_k = \langle T_k, T_k \rangle = \frac{\pi}{2}(1 + \delta_{0k}). \quad (3.9)$$

For factorization purposes specifically in finding the differentiation matrices, the following properties are useful ,

$$|T_k(x)| \leq 1, \quad -1 \leq x \leq 1,$$

$$T_k(\pm 1) = (\pm 1)^k,$$

$$|T'_k(x)| \leq k^2, \quad -1 \leq x \leq 1.$$

By taking only at the end points of the interval, 1 and -1 , the following formula is defined,

$$T'_k(\pm 1) = (\pm 1)^k k^2,$$

where $k \geq 0$ for all above formulas. We can also define $T_k(x)$ in terms of differentiation as follows,

$$2T_k(x) = \frac{T'_{k+1}(x)}{k+1} - \frac{T'_{k-1}(x)}{k-1}, \quad k > 1. \quad (3.10)$$

The expression for the integral is given from (3.6) by,

$$\int_{-1}^1 T_k^2(x) \frac{dx}{\sqrt{1-x^2}} = c_k \frac{\pi}{2},$$

where

$$c_k = \begin{cases} 2 & \text{if } k = 0, \\ 1 & \text{if } k > 0. \end{cases} \quad (3.11)$$

The Chebyshev expansion of a function $u \in L_w^2(-1, 1)$ is

$$u(x) = \sum_{k=0}^{\infty} \tilde{u}_k T_k(x), \quad \text{where } \tilde{u}_k = \frac{2}{c_k \pi} \int_{-1}^1 u(x) T_k(x) w(x) dx. \quad (3.12)$$

where the weight function is $w(x) = \frac{1}{\sqrt{1-x^2}}$.

This is found using the projection theorem and is the same as (3.3). In the case of Chebyshev polynomials we have $\phi_k(x) = T_k(x)$, see (3.4).

The truncated Chebyshev expansion of the function u , is defined on the interval $[-1, 1]$, in terms of orthogonal polynomials $T_k(x)$. Using the projection theorem in Section 3.1, we get,

$$P_N u(x) = \sum_{k=0}^N \tilde{u}_k T_k(x), \quad (3.13)$$

where

$$\tilde{u}_k = \frac{\langle T_k, u \rangle}{\langle T_k, T_k \rangle}, \quad (3.14)$$

with

$$\langle f, g \rangle = \int_{-1}^1 f(x)g(x)w(x) dx. \quad (3.15)$$

3.1.2 Discrete Chebyshev transform

The continuous Chebyshev series method requires the evaluation of the coefficients (3.12), which are

$$\tilde{u}_k = \frac{2}{c_k \pi} \int_{-1}^1 u(x) T_k(x) w(x) dx. \quad (3.16)$$

In general, these integrals cannot be computed analytically. To overcome this issue a quadrature formula is used in the discrete Chebyshev transform.

Concretely, the coefficients (3.16) can be approximated using Gauss-Lobatto integration is the most common way to evaluate numerically integrals of polynomials and hence applies naturally to the integral (3.16). The method is explained next in detail.

Gauss quadrature and discrete polynomial transform:

We shall consider here from a general point of view the problem of expanding of a function in terms of a system of orthogonal polynomials. We denote by \mathbb{P}_N the space of all polynomials of degree $\leq N$. Assume that $\{p_k\}_{k=0,1,\dots}$ is a system of algebraic polynomials (with the degree of p_k equal to k) that are mutually orthogonal over the interval $(-1, 1)$ with respect to a weight function $w(x)$:

$$\int_{-1}^1 p_k(x)p_m(x)w(x) dx = 0 \quad \text{for } k \neq m. \quad (3.17)$$

We discuss here the close relation between orthogonal polynomials and Gauss-type integration formulas on the interval $[-1, 1]$. The material of this subsection includes interpolation formulas and discrete transforms. First, we review Gaussian integration formulas. The result about Gauss integration can be found in most textbooks on numerical analysis. For completeness we include the proofs concerning Gauss-Lobatto formulas, see also (Canuto *et al.*, 1988) and (Canuto *et al.*, 2006).

Gauss integration: Let $x_0 < x_1, \dots < x_N$ be the $N + 1$ roots of the $(N + 1)$ -th orthogonal polynomial $p_{N+1}(x)$ and the weights w_j are solutions of the linear system,

$$\int_{-1}^1 x^k w(x) dx = \sum_{j=0}^N (x_j)^k w_j, \quad k = 0, \dots, N. \quad (3.18)$$

For polynomials $p(x)$ of degree $< 2N+1$, Gauss integration gives the exact result,

$$\int_{-1}^1 p(x)w(x) dx = \sum_{j=0}^N p(x_j)w_j, \quad \text{for all } p \in \mathbb{P}_{2N+1}. \quad (3.19)$$

The points $x = -1$ and $x = 1$ are not roots of $p_{N+1}(x)$; consequently, the values of $p(x)$ at the boundaries are not included in equation (3.19). This version of Gauss integration is quite well known. However, the roots are all in the interior of $(-1, 1)$. The requirement of imposing boundary conditions at both end points creates the need for generalized Gauss integration formulas which include these points.

To obtain the Gauss-Lobatto formula let us consider the polynomial,

$$q(x) = p_{N+1}(x) + \lambda p_N(x) + \mu p_{N-1}(x) \quad (3.20)$$

with λ and μ chosen such that $q(-1) = q(1) = 0$.

Gauss-Lobatto integration: Let $-1 = x_0, x_1, \dots, x_N = 1$ be the $N + 1$ roots of the polynomial (3.20), and let w_0, w_1, \dots, w_N be the solution of the linear system,

$$\int_{-1}^1 x^k w(x) dx = \sum_{j=0}^N (x_j)^k w_j, \quad k = 0, \dots, N. \quad (3.21)$$

Then the result is exact for polynomials of degree $< 2N - 1$, i.e.,

$$\int_{-1}^1 p(x)w(x) dx = \sum_{j=0}^N p(x_j)w_j, \quad \text{for all } p \in \mathbb{P}_{2N-1}. \quad (3.22)$$

Concretely, explicit formulas for the quadrature points and weights in the Chebyshev case, where the weight function is $w(x) = \frac{1}{\sqrt{1-x^2}}$, are:

1. Chebyshev-Gauss,

$$x_j = \cos \frac{(2j+1)\pi}{2N+2}, \quad \text{with } w_j = \frac{\pi}{N+1}, \quad j = 0, \dots, N. \quad (3.23)$$

2. Chebyshev-Gauss-Lobatto,

$$x_j = \cos \frac{j\pi}{N}, \quad j = 0, \dots, N \quad (3.24)$$

with

$$w_j = \begin{cases} \frac{\pi}{2N}, & \text{if } j = 0, N, \\ \frac{\pi}{N}, & \text{if } 1 \leq j \leq N-1. \end{cases} \quad (3.25)$$

The most commonly used points to solve PDEs are those for the Chebyshev-Gauss-Lobatto case, which include the boundary points; we are also using them in this thesis.

The coefficients of the continuous polynomial expansion (3.13), \hat{u}_k , defined by (3.14) and (3.16) can be approximated by the discrete polynomial coefficients computed using Chebyshev-Lobatto quadrature as

$$\hat{u}_k = \frac{2}{Nc_k} \sum_{j=0}^N \frac{1}{c_j} u(x_j) T_k(x_j), \quad c_j = 1 + \delta_{j0} + \delta_{jN}, \quad k = 0, 1, \dots, \quad (3.26)$$

which is **the discrete Chebyshev transform**.

Since, in general, we cannot compute the coefficients \tilde{u}_k (integral form) but only the coefficients \hat{u}_k (quadrature formula), the actual continuous numerical representation of the function $u(x)$ is

$$I_N u(x) = \sum_{k=0}^N \hat{u}_k T_k(x), \quad (3.27)$$

not the truncated orthogonal projection $P_N u(x)$ in (3.13).

This interpolating polynomial satisfies the following

$$I_N u(x_j) = u(x_j), \quad 0 \leq j \leq N. \quad (3.28)$$

The formula,

$$u(x_j) = \sum_{k=0}^N \hat{u}_k T_k(x_j), \quad 0 \leq j \leq N, \quad (3.29)$$

is called **the inverse discrete Chebyshev transform**.

3.1.3 Spectral convergence

If u is continuously differentiable in $[-1, 1]$, then for $k \neq 0$,

$$\hat{u}_k = \mathcal{O}(k^{-1}). \quad (3.30)$$

Repeating this line of argument we have the following result for smooth functions.

Theorem: If u is m -times continuously differentiable in $[-1, 1]$ ($m \geq 1$) then

$$\hat{u}_k = \mathcal{O}(k^{-m}), \quad k = \pm 1, \pm 2, \dots, \quad (3.31)$$

It follows that the smoother the function u , the faster the convergence of its truncated spectral expansion. We conclude that the k -th Chebyshev coefficients of a function which is infinitely differentiable with all its derivatives on $[-1, 1]$, decay faster than any algebraic order of k . See (Boyd, 2001; Canuto *et al.*, 1988, 2006; Hesthaven *et al.*, 2007; Trefethen, 2000).

Let us consider the following examples:

1. The smooth function

$$u_\infty(x) = e^{\cos(8x^3+1)}, \quad (3.32)$$

is infinitely differentiable in $[-1, 1]$. Its coefficients decay exponentially, hence the error $\|u(x) - I_N u(x)\|_{L^\infty(-1,1)}$ converges exponentially to zero as is shown in the left hand side of Figure 3.1. The error decreases to 10^{-14} .

In order to compute the error as plotted in the figure, a sufficient number of linearly spaced points, namely 1000, are taken on the interval $[-1, 1]$. The Matlab command **linspace** generates these points. Both solutions, the exact and the numerical, are then evaluated on these 1000 points. The final step is to apply the ∞ -**norm** on the difference of those solutions, where the definition of ∞ -**norm** of vectors $x = (x_1, \dots, x_n)$ is

$$\|x\|_\infty := \max_i |x_i|$$

see (Dym, 2013).

2. The function

$$u_1(x) = |x|, \quad (3.33)$$

is non-smooth because its first derivative is not continuous, the series converges slowly at the point where u'_1 is discontinuous. Therefore, the error $\|u(x) - I_N u(x)\|_{L^\infty(0,2\pi)}$ also converges only algebraically to zero (see the right hand side of Figure 3.1).

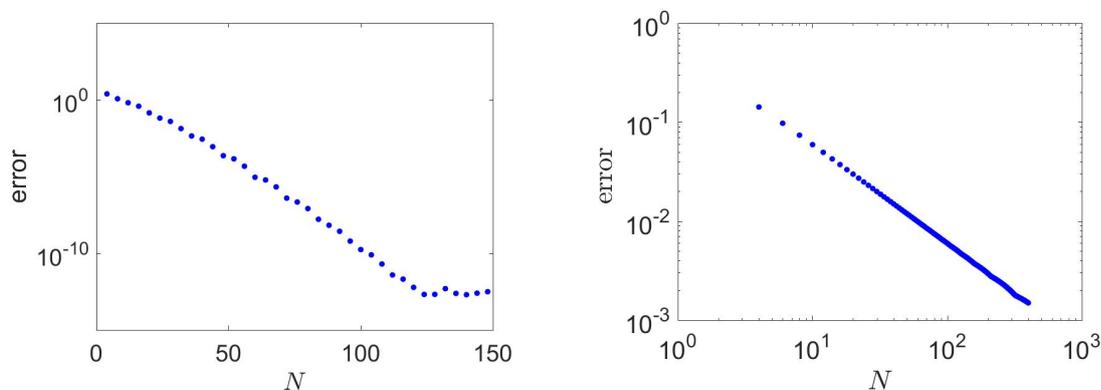


Figure 3.1: The error $\|u(x) - I_N u(x)\|_{L^\infty(-1,1)}$ against the order of the truncation N , showing the exponential convergence of the function (3.32) (left panel), using *semilogy* scale, and algebraic convergence of (3.33) (right panel), using *loglog* scale.

3.1.4 Differentiation

The material discussed in this subsection is standard and can be found in many text books on spectral methods, such as, in (Canuto *et al.*, 1988), (Canuto *et al.*, 2006) and (Hesthaven *et al.*, 2007).

The derivative of a function u in (3.12) can be expanded in Chebyshev polynomials as in (3.12) as

$$u' = \sum_{k=0}^{\infty} \hat{u}_k^{(1)} T_k. \quad (3.34)$$

It turns out that the coefficients $\hat{u}_k^{(1)}$ are given by

$$\hat{u}_k^{(1)} = \frac{2}{c_k} \sum_{\substack{p=k+1 \\ p+k \text{ odd}}}^{\infty} p \hat{u}_p, \quad k \geq 0. \quad (3.35)$$

The key to prove this formula is the relation (3.10). Remember $T_0 = 1$ and $T_1 = x$, therefore $T'_0 = 0$ and $T'_1 = 1$. So,

$$u'(x) = \sum_{k=0}^{\infty} \hat{u}_k^{(1)} \left(\frac{T'_{k+1}(x)}{2(k+1)} - \frac{T'_{k-1}(x)}{2(k-1)} \right). \quad (3.36)$$

On the other hand,

$$u'(x) = \sum_{k=0}^{\infty} \hat{u}_k T'_k(x), \quad (3.37)$$

and since the T_k are linearly independent, it follows that

$$2k \hat{u}_k = c_{k-1} \hat{u}_{k-1}^{(1)} - \hat{u}_{k+1}^{(1)}, \quad k \geq 1; \quad (3.38)$$

whence (3.35) follows.

Since $\hat{u}_k^{(1)} = 0$ for $k \geq N$, the non-zero coefficients can be computed in decreasing order by the recurrence relation,

$$c_k \hat{u}_k^{(1)} = \hat{u}_{k+2}^{(1)} + 2(k+1) \hat{u}_{k+1}^{(1)}, \quad 0 \leq k \leq N-1. \quad (3.39)$$

Similarly, the q -th derivative of a function u can be expanded in Chebyshev polynomials as

$$u^{(q)} \equiv \sum_{k=0}^{\infty} \hat{u}_k^{(q)} T_k = \sum_{k=0}^{\infty} \hat{u}_k^{(q-1)} T'_k, \quad (3.40)$$

where the coefficients are given by generalization of (3.39):

$$c_k \hat{u}_k^{(q)} = \hat{u}_{k+2}^{(q)} + 2(k+1) \hat{u}_{k+1}^{(q-1)}, \quad q \geq 1, \quad k \geq 0. \quad (3.41)$$

Hence, the second derivative is

$$u'' = \sum_{k=0}^{\infty} \hat{u}_k^{(2)} T_k, \quad (3.42)$$

where the coefficients are

$$\hat{u}_k^{(2)} = \frac{1}{c_k} \sum_{\substack{p=k+2 \\ p+k \text{ even}}}^{\infty} p(p^2 - k^2) \hat{u}_p, \quad k \geq 0. \quad (3.43)$$

The **Chebyshev projection derivative** is just $(P_N u)'$. The **Chebyshev interpolation derivative** of a function u using either the Chebyshev-Gauss quadrature nodes (3.23) or the Chebyshev-Gauss-Lobatto nodes (3.25) is defined as the derivative of the discrete Chebyshev series representing u at the same nodes,

$$Du = (I_N u)'. \quad (3.44)$$

Chebyshev truncation and interpolation do not commute with differentiation. In order to find the values $(Du)(x_j)$, for $j = 0, 1, \dots, N$, from the values $u(x_l)$, for $l = 0, 1, \dots, N$, one can use formula (3.26) to get the discrete Chebyshev coefficients of u , then use (3.35) to differentiate in transform space and finally compute $(Du)(x_j)$ through (3.29). However, this procedure is not efficient for N of practical interest in the absence of a fast transform method for the Chebyshev expansion. Therefore, it is preferable to get the Chebyshev interpolation derivative at the grid points through matrix multiplication, namely,

$$Du(x_j) = \sum_{l=0}^N D_{jl} u(x_l), \quad j = 0, \dots, N. \quad (3.45)$$

The matrix D is called the first-order Chebyshev differentiation matrix and its entries and they can be found in texts on spectral methods, see for example, in (Canuto *et al.*, 1988), (Canuto *et al.*, 2006) and (Trefethen, 2000). The matrix entries are readily determined to be

$$D_{jl} = \begin{cases} \frac{1 + \delta_{j0} + \delta_{jN} (-1)^{j+l}}{1 + \delta_{l0} + \delta_{jN} x_j - x_l}, & \text{if } j \neq l, \\ \frac{-x_l}{2(1 - x_l^2)}, & \text{if } 1 \leq j = l \leq N - 1, \\ \frac{1 + 2N^2}{6}, & \text{if } j = l = 0, \\ -\frac{1 + 2N^2}{6}, & \text{if } j = l = N. \end{cases} \quad (3.46)$$

More details are given in (Canuto *et al.*, 1988, 2006; Hurst, 2014). In addition to this matrix they give the second derivative matrix $D^{(2)}$, where

$$D^{(2)}u = (I_N u)'' \quad (3.47)$$

The second Chebyshev interpolation derivative can also be represented in matrix form as,

$$D^{(2)}u(x_j) = \sum_{l=0}^N D_{jl}^{(2)} u(x_l), \quad j = 0, \dots, N. \quad (3.48)$$

The second derivative matrix entries are determined to be

$$D_{jl}^{(2)} = \begin{cases} \frac{(-1)^{j+l}}{c_l} \frac{x_j^2 + x_j x_l - 2}{(1 - x_j^2)(x_j - x_l)^2}, & \text{if } 1 \leq j \leq N-1, \ 0 \leq l \leq N, \ l \neq j, \\ -\frac{(N^2 - 1)(1 - x_j^2) + 3}{3(1 - x_j^2)^2}, & \text{if } 1 \leq j = l \leq N-1, \\ \frac{2(-1)^l(2N^2 + 1)(1 - x_l) - 6}{3c_l(1 - x_l)^2}, & \text{if } 1 \leq l \leq N, \ j = 0, \\ \frac{2(-1)^{(l+N)}(2N^2 + 1)(1 + x_l) - 6}{3c_l(1 + x_l)^2}, & \text{if } 0 \leq l \leq N-1, \ j = N, \\ \frac{N^4 - 1}{15}, & \text{if } j = l = 0, \\ \frac{N^4 - 1}{15}, & \text{if } j = l = N. \end{cases} \quad (3.49)$$

Furthermore, this second derivative matrix can be found from the first derivative matrix by the following formula,

$$D_{jl}^{(m)} = [(D)^m]_{jl}, \quad (3.50)$$

see (Hesthaven *et al.*, 2007) and (Hurst, 2014).

In general,

$$D^{(q)}u = (I_N u)^{(q)}, \quad (3.51)$$

where $D^{(3)}$ and $D^{(4)}$ are given in (Hurst, 2014).

3.2 Solving the wave equation

The linear wave equation,

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}, \quad (3.52)$$

on $(-1, 1)$ and $t \geq 0$, see (1.47)-(1.50), has to be solved subject to Dirichlet boundary conditions,

$$u(-1, t) = P_{-1}(t), \quad t \in [0, \infty), \quad (3.53)$$

$$\mu u(1, t) + (1 - \mu)u_x(1, t) = P_1(t), \quad t \in [0, \infty), \quad (3.54)$$

where $\mu \in \{0, 1\}$ with $\mu = 1$ for Dirichlet boundary condition. In (3.53) and (3.54), P_{-1} and P_1 are given functions. In this thesis Dirichlet boundary conditions are taken as homogeneous,

$$u(-1, t) = 0, \quad \text{and } u(1, t) = 0, \quad (3.55)$$

and initial conditions

$$u(x, 0) = f(x), \quad (3.56)$$

and

$$\frac{\partial u}{\partial t}(x, 0) = 0. \quad (3.57)$$

All four methods mentioned at the start of this chapter (pseudospectral, collocation, tau and Galerkin) are used to solve the wave equation (3.52). For time stepping, we use the explicit Euler method and the trapezoidal method in all four spectral methods. The modified Euler method and the implicit Euler method are also used in the pseudospectral and tau methods. However, the calculations for the implicit Euler method are not given in as much detail because they are similar to the calculations for the trapezoidal method.

3.3 The pseudospectral method

The pseudospectral method requires the residual to vanish at a set of grid points called the collocation points (the collocation points are usually the same as the quadrature points).

3.3.1 Spatial discretization

The chosen trial functions for any number N are,

$$T_k(x) = \cos(k \cos^{-1}(x)), \quad k = 0, 1, \dots, N, \quad (3.58)$$

as it is shown in (3.4).

From (3.27) the semi-discretization solution has the representation

$$u_{sd}(x, t) = \sum_{k=0}^N \hat{u}_k(t) T_k(x), \quad (3.59)$$

where from (3.26)

$$\hat{u}_k(t) = \frac{2}{Nc_k} \sum_{j=0}^N \frac{1}{c_j} u_{sd}(x_j, t) T_l(x_j), \quad c_j = 1 + \delta_{j0} + \delta_{jN}, \quad k = 0, \dots, N. \quad (3.60)$$

The pseudospectral method requires the residual to vanish at the collocation points

$$R(x_j) = 0, \quad j = 1, \dots, N - 1. \quad (3.61)$$

For the wave equation, this is

$$\left. \frac{\partial^2 u_{sd}}{\partial t^2} - \frac{\partial^2 u_{sd}}{\partial x^2} \right|_{x=x_j} = 0, \quad j = 1, \dots, N - 1. \quad (3.62)$$

The boundary conditions

$$u_{sd}(-1, t) = 0, \quad u_{sd}(1, t) = 0 \quad (3.63)$$

and the initial conditions

$$u_{sd}(x_k, 0) = f(x_k), \quad k = 0, \dots, N \quad (3.64)$$

and

$$\frac{\partial u_{sd}}{\partial t}(x_k, 0) = 0, \quad k = 0, \dots, N \quad (3.65)$$

accompany (3.62).

Since the approximate solution is required to satisfy the differential equation exactly at a set of discrete points, equation (3.62) are based on the strong formulation of the differential equation. One can formally obtain the same equation (3.62) by taking the test functions

$$\psi_j(x) = \delta(x - x_j), \quad j = 1, \dots, N - 1,$$

and enforcing the conditions,

$$\int_{-1}^1 \left[\frac{\partial^2 u_{sd}}{\partial t^2} - \frac{\partial^2 u_{sd}}{\partial x^2} \right] \psi_j(x) dx = 0, \quad j = 1, \dots, N - 1.$$

The exact derivative of (3.59) is

$$\frac{\partial^2 u_{sd}(x, t)}{\partial x^2} = \sum_{k=0}^N \hat{u}_k^{(2)}(t) T_k(x). \quad (3.66)$$

The coefficients $\hat{u}_k^{(2)}$ obviously depend linearly on the nodal values $u_{sd}(x_l, t)$ hence they satisfy (3.48), where the matrix $D^{(2)}$ is given in (3.49). Then (3.48) and (3.59) lead (3.66) to become

$$\frac{\partial^2 u_{sd}(x_j, t)}{\partial x^2} = \sum_{k=0}^N \hat{u}_k^{(2)}(t) T_k(x_j) = \sum_{l=0}^N D_{jl}^{(2)} u_{sd}(x_l, t), \quad j = 1, \dots, N-1. \quad (3.67)$$

From (3.28) we have,

$$u_{sd}(x_j, t) = u_j(t) \quad j = 1, \dots, N-1, \quad (3.68)$$

where $u_j(t) \approx u(x_j, t)$ and at $t = 0$ they are equal, $u_j(0) = u(x_j, 0) = f(x_j)$, then equation (3.67) can be rewritten as

$$\frac{\partial^2 u_j(t)}{\partial x^2} = \sum_{k=0}^N \hat{u}_k^{(2)}(t) T_k(x_j) = \sum_{l=0}^N D_{jl}^{(2)} u_l(t), \quad j = 1, \dots, N-1. \quad (3.69)$$

By (3.63), we have $u_0(0) = u_N(0) = 0$.

Substituting (3.69) into (3.62), we get

$$\frac{\partial^2 u_j(t)}{\partial t^2} = \sum_{l=0}^N D_{jl}^{(2)} u_l(t), \quad j = 1, \dots, N-1, \quad (3.70)$$

where $D^{(2)}$ is given in (3.49).

The second-order ODE (3.70) can be written in first-order form as

$$\frac{\partial y_j(t)}{\partial t} = w_j(t), \quad (3.71)$$

$$\frac{\partial w_j(t)}{\partial t} = \sum_{l=0}^N D_{jl}^{(2)} y_l(t), \quad j = 1, \dots, N-1. \quad (3.72)$$

3.3.2 Time stepping

If Euler's method is used to solve (3.71) and (3.72), then

$$y_j(t+h) = y_j(t) + hw_j(t), \quad (3.73)$$

$$w_j(t+h) = w_j(t) + h \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t) \quad j = 1, \dots, N-1, \quad (3.74)$$

where from the boundary conditions (3.55) we have $y_0(0) = y_N(0) = 0$ and $w_0(0) = w_N(0) = 0$, and from the initial conditions (3.56) and (3.57) we have

$$y_j(0) = f(x_j) \quad \text{and} \quad w_j(0) = 0, \quad j = 1, \dots, N-1.$$

The quadrature points used are Gauss-Lobatto [see (3.24)]; we rewrite them here again,

$$x_j = \cos \frac{j\pi}{N}, \quad j = 0, \dots, N. \quad (3.75)$$

If the modified Euler method is implemented to solve (3.71) and (3.72) then

$$y_j(t+h) = y_j(t) + hw_j(t) + \frac{h^2}{2} \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t), \quad (3.76)$$

$$w_j(t+h) = w_j(t) + h \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t) + \frac{h^2}{2} \sum_{l=1}^{N-1} D_{jl}^{(2)} w_l(t) \quad j = 1, \dots, N-1, \quad (3.77)$$

with the same initial and boundary conditions.

In addition to the previous methods, we apply the trapezoidal method to solve (3.71) and (3.72),

$$y_j(t+h) = y_j(t) + \frac{h}{2} \{w_j(t) + w_j(t+h)\}, \quad (3.78)$$

$$w_j(t+h) = w_j(t) + \frac{h}{2} \left\{ \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t) + \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t+h) \right\}, \quad j = 1, \dots, N-1. \quad (3.79)$$

Finally, in order to find the numerical solution, a sufficient number of equally spaced points, denoted by nx , are taken on the interval $(-1, 1)$. The Matlab command `linspace(-1, 1, nx)` generates these points, where $nx > N$. Then substituting the u_j we have got from these time stepping methods, into (3.60) leads to the numerical solution,

$$u_{num}(x, t) = \sum_{k=0}^N \hat{u}_k(t) T_k(x), \quad (3.80)$$

where x are the equally spaced points. They are different than the Chebyshev points x_j in (3.75), and they must be more than the Chebyshev points.

3.4 Numerical results

The initial condition

$$f(x) = \sin(\pi x), \quad (3.81)$$

is used, which is infinitely differentiable, smooth, or C^∞ .

In addition to this first initial condition other initial conditions are taken, such as,

$$f(x) = e^{\sin(\pi x)} - 1, \quad (3.82)$$

which is not continuous at the second derivative, nonsmooth at the second derivative, or C^1 . And

$$f(x) = \frac{1}{11}e^x(11 - 8x - 10x^2 + 8x^3 - x^4), \quad (3.83)$$

which is not continuous at the fourth derivative, nonsmooth at the fourth derivative, or C^3 , see Chapter 1, IC (1.54), IC (1.55) and IC (1.56).

A point worth clarifying is that this is a slight abuse of terminology because it is the solution that is smooth or not smooth, not the initial condition, but we label them this way to make it clearer for the reader. The three ICs are considered in order to investigate more in the behaviour of the pseudospectral scheme when solving the wave equation (3.52). They are used to study the numerical solution, the convergence and the stability of this linear wave equation for the time stepping methods used; this is discussed next. In addition to the linear wave equation, these initial conditions are used in next chapters when solving a nonlinear wave equation. We intentionally choose these ICs because they are having different properties regarding the smoothness. The first initial condition (3.81) yields a smooth solution, it is symmetric and infinitely continuously differentiable in space for $[-1, 1]$ for even derivatives. However, the second initial condition (3.82) introduces non-smoothness at the second derivative because it does not satisfy the PDE in the corners of the domain where the second derivative is non-zero at the endpoints, which means that it does not have continuous second derivative at the boundaries, $x = \pm 1$, so the worst possible discontinuity among the ICs used here. While the third IC, (3.83), is continuously differentiable for the second derivative in space for $[-1, 1]$, but is not continuously differentiable for the fourth derivative because it is not zero at the boundaries, so smoother solutions than for second IC (3.82), but still not infinitely

differentiable as for first IC (3.81).

We need to find the exact solution of the linear wave equation (3.52) for all the three ICs mentioned, so d'Alembert methods is used for this purpose, see Section 1.9 for more information.

In our calculations we are using $t = 0.9$, so from Section 1.9 using (1.46) and (1.51), the d'Alembert solution at this t , after few steps, will be

$$F(x + 0.9) = \begin{cases} f(x + 0.9), & \text{for } -1 \leq x < 0.1, \\ -f(1.1 - x), & \text{for } 0.1 \leq x < 1. \end{cases} \quad (3.84)$$

and

$$F(x - 0.9) = \begin{cases} -f(-1.1 - x), & \text{for } -1 \leq x < -0.1, \\ f(x + 0.9), & \text{for } -0.1 \leq x < 1. \end{cases} \quad (3.85)$$

so, the d'Alembert solution by (1.46) is

$$u(x, t) = \begin{cases} \frac{1}{2}f(x + 0.9) - \frac{1}{2}f(-1.1 - x), & \text{for } -1 \leq x < -0.1, \\ \frac{1}{2}f(x + 0.9) + \frac{1}{2}f(x - 0.9), & \text{for } -0.1 \leq x < 0.1, \\ -\frac{1}{2}f(1.1 - x) + \frac{1}{2}f(x - 0.9), & \text{for } 0.1 \leq x < 1. \end{cases} \quad (3.86)$$

Applying the ICs into this formula, (3.86), will give us the exact solution at this specific time, $t = 0.9$.

For the particular initial condition (3.81), we have $u(x, 0) = \sin(\pi x)$, the exact solution can also be found using the separation of variables, which is

$$u(x, t) = \sin(\pi x) \cos(\pi t), \quad (3.87)$$

as explained in Section 1.10, see equation (1.84). At $t = 0.9$ the solution is same as the one we get from (3.86).

Moreover, as explained in Section 1.9 by using (1.46) and (1.51), the exact d'Alembert solutions at $t = 0, 4, 8, \text{etc.}$, have the returning property to the initial conditions; they are exactly the functions $f(x)$ in the initial conditions, because the function $F(x)$, (1.51), is a periodic function with period four, so the solutions are infinitely smooth at these specified times for all three ICs. Likewise, at the other times of multiple 2, i.e., at $t = 2, 6, 10, \text{etc.}$, the exact d'Alembert solutions are infinitely differentiable again, infinitely smooth, since the characteristics touch

the boundaries at these specific times, thus the exact d'Alembert solution is the function $-f(-x)$, meaning that the discontinuities go away again then. Figure 1.5 mirrors these properties. We used this to verify our implementation. In Subsection 3.4.4 we investigate the conduct of the numerical solutions at these specific times to determine whether or not they possess the same property as well.

In this chapter we are studying the convergence and stability of the time stepping methods employing these ICs.

3.4.1 Convergence of Euler's method

Figures 3.2 and 3.4 show the convergence error in space and time, respectively, of the spectral methods when solving the wave equation (3.52); the error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, is plotted against the truncation index N in Figure 3.2 using the pseudospectral method with Euler's method for the C^∞ IC (3.81). Additionally, the error is plotted against the inverse step size $1/h$ in Figure 3.4 with the Euler's method (left panel) and with the modified Euler method (right panel) for the same IC. It can be seen in Figures 3.2 and the left panel of 3.4 that Euler's method is convergent since the convergence error is decaying to zero correspondingly. Figure 3.2 shows that the IC (3.81) converges exponentially. Figure 3.4, left panel, shows that the order of Euler's method is $\mathcal{O}(h)$.

It can be seen in Figure 3.2 that the error is consistent with the $\mathcal{O}(h)$ error expected for Euler's method; the spatial error decreases until it gets smaller than the error that comes from the time stepping error of Euler's method, which is $\mathcal{O}(h)$, then it remains there with no further decaying. The step size used here is $h = 10^{-5}$, thus, the error reaches to about 10^{-5} .

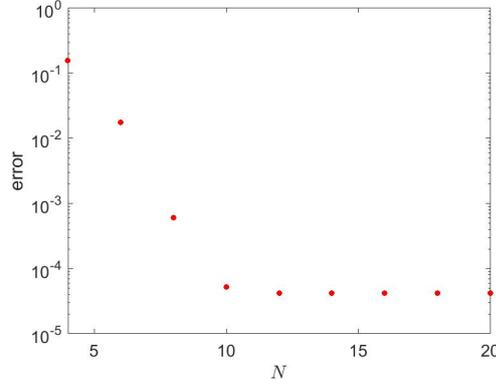


Figure 3.2: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the wave equation (3.52) solved using the pseudospectral method with Euler's method for the C^∞ IC (3.81). The step size is $h = 10^{-5}$ and the error is evaluated at $t = 0.9$. Here u is the exact solution and $u_{num}(x, t)$ is the numerical solution.

3.4.2 Convergence of the modified Euler method

The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, is plotted against the truncation index N in Figure 3.3 using the pseudospectral method with the modified Euler method and against the inverse step size $1/h$ in Figure 3.4 for the C^∞ IC (3.81). Similarly, the combination of the modified Euler method with pseudospectral method converges exponentially in N for solving the wave equation (3.52) for this IC. One can see in the right panel of Figure 3.4 that the modified Euler method for this IC is of order $\mathcal{O}(h^2)$.

It can be seen in Figure 3.3 that the error is consistent with the $\mathcal{O}(h^2)$ error expected for the modified Euler method; the spatial error decreases until it gets smaller than the error that comes from the time stepping error of the modified Euler method, which is $\mathcal{O}(h^2)$. The step size used here is $h = 10^{-5}$, thus, the error reaches to about 10^{-10} .

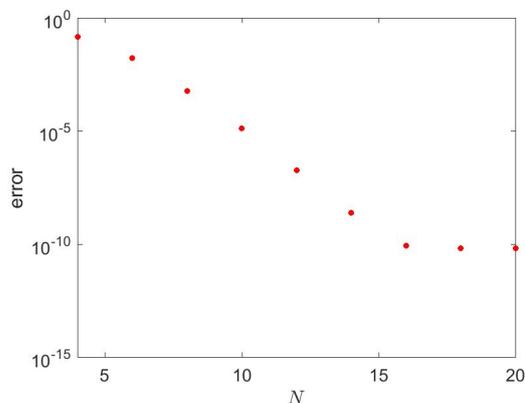


Figure 3.3: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the wave equation (3.52) using the pseudospectral method with the modified Euler method for the C^∞ IC (3.81). The step size is $h = 10^{-5}$ and the error is evaluated at $t = 0.9$.

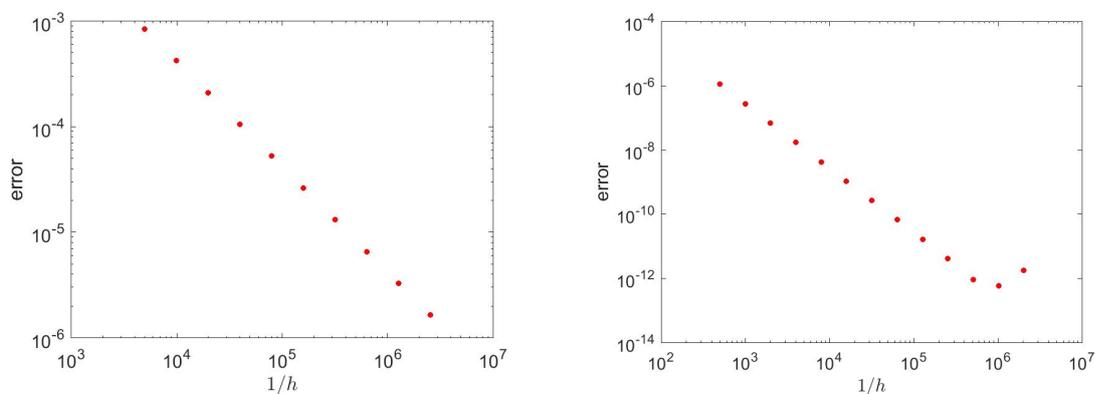


Figure 3.4: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the inverse step size $1/h$ for the wave equation (3.52) solved using the pseudospectral method for the C^∞ initial condition (3.81). The left panel with Euler's method and the right panel with the modified Euler method. The Chebyshev series is truncated at $N = 20$ and the error is evaluated at $t = 0.9$.

3.4.3 Convergence of the implicit Euler method

Again same as Euler's method, the combination of the implicit Euler method with pseudospectral method converges exponentially in N for solving the wave equation (3.52) using the pseudospectral method for the C^∞ IC (3.81) and this method is $\mathcal{O}(h)$ as well.

3.4.4 Convergence of the trapezoidal method

The convergence for the combination of the trapezium method with pseudospectral method in N is studied for all the three ICs, the C^∞ IC (3.81), the C^1 IC (3.82) and the C^3 IC (3.83) here as it shows in Figure 3.5, because this method is more important and the main method that we are studying in Chapter 4 when studying the energy conservation using all these three ICs.

Figure 3.5 shows that the error for C^∞ IC (3.81) converges exponentially, panel (a), while for the other ICs, the C^1 (3.82), panel (b), and the C^3 (3.83), panel (c), converges algebraically with order seems to be $\frac{1}{N^3}$ and $\frac{1}{N^5}$, respectively.

Similar to the modified Euler method, the trapezoidal method converges in h with order two for all the three ICs.

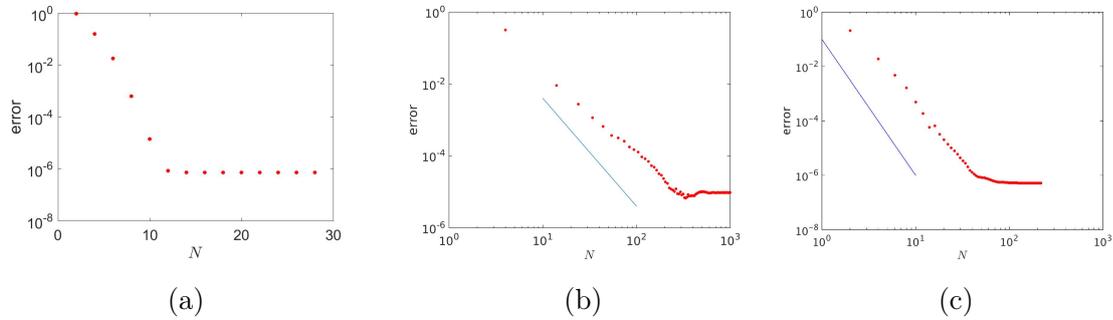


Figure 3.5: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the wave equation (3.52) solved using the pseudospectral method with the trapezoidal method for the C^∞ IC (3.81), the C^1 IC (3.82) and the C^3 IC (3.83) (a, b and c respectively). The error is evaluated at $h = 10^{-3}$ and $t = 0.9$ using $nx = 100N$. The blue straight line indicates order $\frac{1}{N^3}$ for the IC (3.82) and order $\frac{1}{N^5}$ for the IC (3.83).

One thing that we note in Figure 3.5b is that the points do not lie on a very straight line. Therefore, to gain a better understanding of how the method behaves, we plot the Chebyshev coefficients of the numerical solution \hat{u}_k against the mode k , for all three ICs, the C^∞ IC (3.81), the C^3 IC (3.83) and the C^1 IC (3.82) in Figures 3.6, 3.7 and 3.8, respectively. We colour the even coefficients blue and the odd coefficients red.

We see that for the first IC (3.81) the coefficients decay exponentially, see Figure 3.6. More precisely, for the first IC (3.81) the odd spectral coefficients decay exponentially, the red dots plots, whereas the even spectral coefficients are just round-off, the blue dots plots, at any t but in this figure we have considered only at $t = 0.9$. We can understand the reasons of this from Section 1.9, which are explained next in detail. As mentioned in Section 1.9 that every function $f(x)$ of the initial conditions can uniquely decomposed as the sum of an even and an odd function. And from the Chebyshev polynomials (3.4) one can see that the polynomials have the even and odd parts; it is even when k is even; and odd when k is odd. Consequently, the solution in (4.23) can be written as the even and odd components. Certainly, for the linear wave equation, the odd and even components of the initial condition, the Chebyshev polynomials and the solution can be evolved independently. Thus, clearly, nonsmoothness only affects the non-zero component

of the solution. As explained in Section 1.9 that the IC (3.81) is an odd function, there is no even component of the function at all. The solution of the homogeneous linear wave equation preserves the symmetry properties of the initial condition f . If f is odd (resp. even) then the solution remains odd (resp. even). Accordingly, the whole solution evolves in such a way that the odd function is always the case, so sure enough, the even spectral coefficients are always basically just round-off.

In contrast, for the third IC (3.83) the spectral coefficients decay changes from decaying exponentially to only algebraically with order $\frac{1}{k^5}$, see Figure 3.7. As one can see from the figures, especially 3.7a, that the odd/even spectral coefficients both drop off only algebraically. More details are given later in this subsection. That is because the function of this IC itself has both odd and even components, as explained in Section 1.9. Thus, we have the odd/even spectral coefficients plots; they are not just round-off for both even and odd k values. The reason for this is that for the third IC (3.83) both of the function components are non-zero; as $f_o^{(4)}(\pm 1) \neq 0$ and $f_e^{(4)}(\pm 1) \neq 0$. That means, the discontinuities generated from $f^{(4)}$ are associated with both components, the even part and the odd part of f , thence, nonsmoothness affects both components of the solution as they are non-zero. Accordingly, the spectral coefficients in this case are much the same for even versus odd, whence, we have both odd/even spectral coefficients drop off algebraically. Furthermore, at $t = 0$ the solutions are just the IC function so both of the plots drop off exponentially for all ICs accordingly.

The more interesting case is the second IC (3.82), as we have two different plots in the figures; one plot decays exponentially and the other one only algebraically with order $\frac{1}{k^3}$, see Figure 3.8. More precisely, the odd spectral coefficients decay exponentially, the red dots plots, while the even spectral coefficients decay algebraically with order $\frac{1}{k^3}$, the blue dots plots, at any $t > 0$ but we have considered only six times $t = 0.04, 0.9, 1, 2, 4$ and $t = 5.3$ in Figure 3.8 and $t = 0.16$ in Figure 3.9a. We are having two coefficients components because, similar to the third IC, the function of second IC (3.82) itself has both odd and even components as well, as explained in Section 1.9. In contrast to the third IC (3.83), for the second IC (3.82) the discontinuities generated from f'' are associated with the even component of f , and none of it with the odd component of f because $f_o''(\pm 1) = 0$ and $f_e''(\pm 1) \neq 0$ as $f_e''(\pm 1) = \pi^2$, see Section 1.9 for more understanding. So, as mentioned above, nonsmoothness affects any non-zero component of the solution.

Consequently, it is not surprising then that the odd spectral coefficients of this IC behave exactly as if that discontinuity were not there at all, therefore, they continue to drop off exponentially, and it is only the even spectral coefficients that see the discontinuities generated from f'' at all, and hence drop off only algebraically; the spectral coefficients are determined entirely by the fact that the discontinuities generated from f'' are completely even. The discontinuities is not affecting the odd part that means this part behaves exactly as the first IC (3.81), which is totally an odd function, hence, it decays exponentially. More details are given later in this subsection.

Figure 3.8a, shows that the even spectral coefficients component has not changed completely to decay algebraically; there is a fast drop, transition, in the plots from algebraically to round-off, for instance for $t = 0.04$ the drop off is for k between 900 and 1000, but as time increases this is pushed forward. Therefore, the range of k giving algebraic convergence grows in time. Surprisingly, Figure 3.9 shows how the drop in range decreases as step size decreases, $h = 10^{-2}, 10^{-3}, 10^{-4}$. In addition, it shows how the range of k giving algebraic convergence grows as step size decreases at any fixed time, here is at $t = 0.16, 1, 2$. More information is given next. One can observe that the drop in solely affects the even coefficients component of the plots, not the odd coefficients component. The spectra figures of the numerical solution in this subsection show that the drop in occurs only when using the trapezoid method with the second IC (3.82) but it does not happen when using the third IC (3.83). The reason for this is perhaps that because as known the solution generated by the C^1 IC (3.82) is the more nonsmooth than the solution generated by the C^3 IC (3.83). We have tested this at various times for the third IC (3.83), including the above times used for the second IC (3.82) and for all three step sizes $h = 10^{-2}, 10^{-3}, 10^{-4}$, there is no drop in the plots.

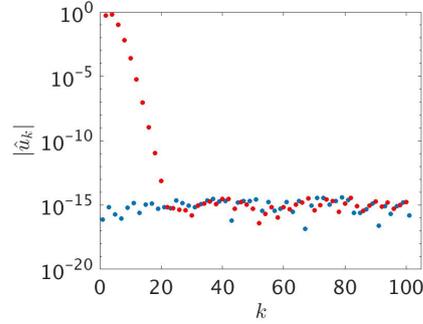


Figure 3.6: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) for the C^∞ IC (3.81) solved using the pseudospectral method with trapezoidal method at $t = 0.9$. The coefficients are evaluated for $h = 10^{-3}$ and $N = 100$. The blue dots are the even coefficients component and the red dots are the odd coefficients component. Similar plot is obtained for $t = 1$ and $t = 4$.

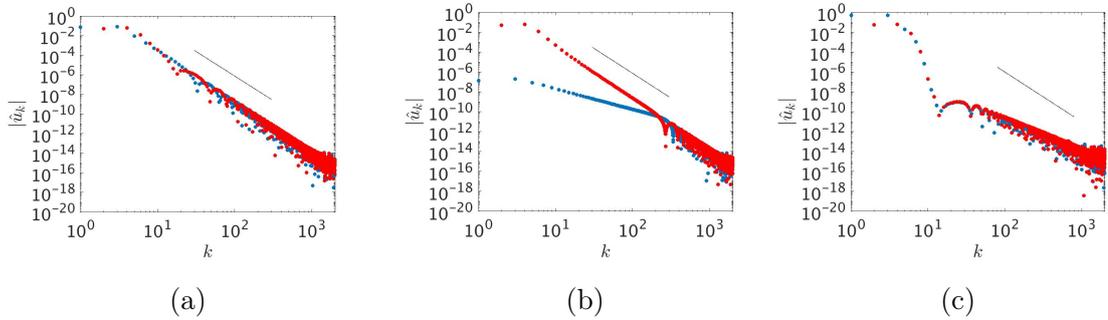


Figure 3.7: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) for the C^3 IC (3.83) solved using the pseudospectral method with the trapezoidal method at $t = 0.9, 1, 4$ (a, b, c respectively). The coefficients are evaluated for $h = 10^{-3}$ and $N = 2000$. The black straight line indicates order $\frac{1}{k^5}$, more clearly for $t = 0.9$ and $t = 1$.

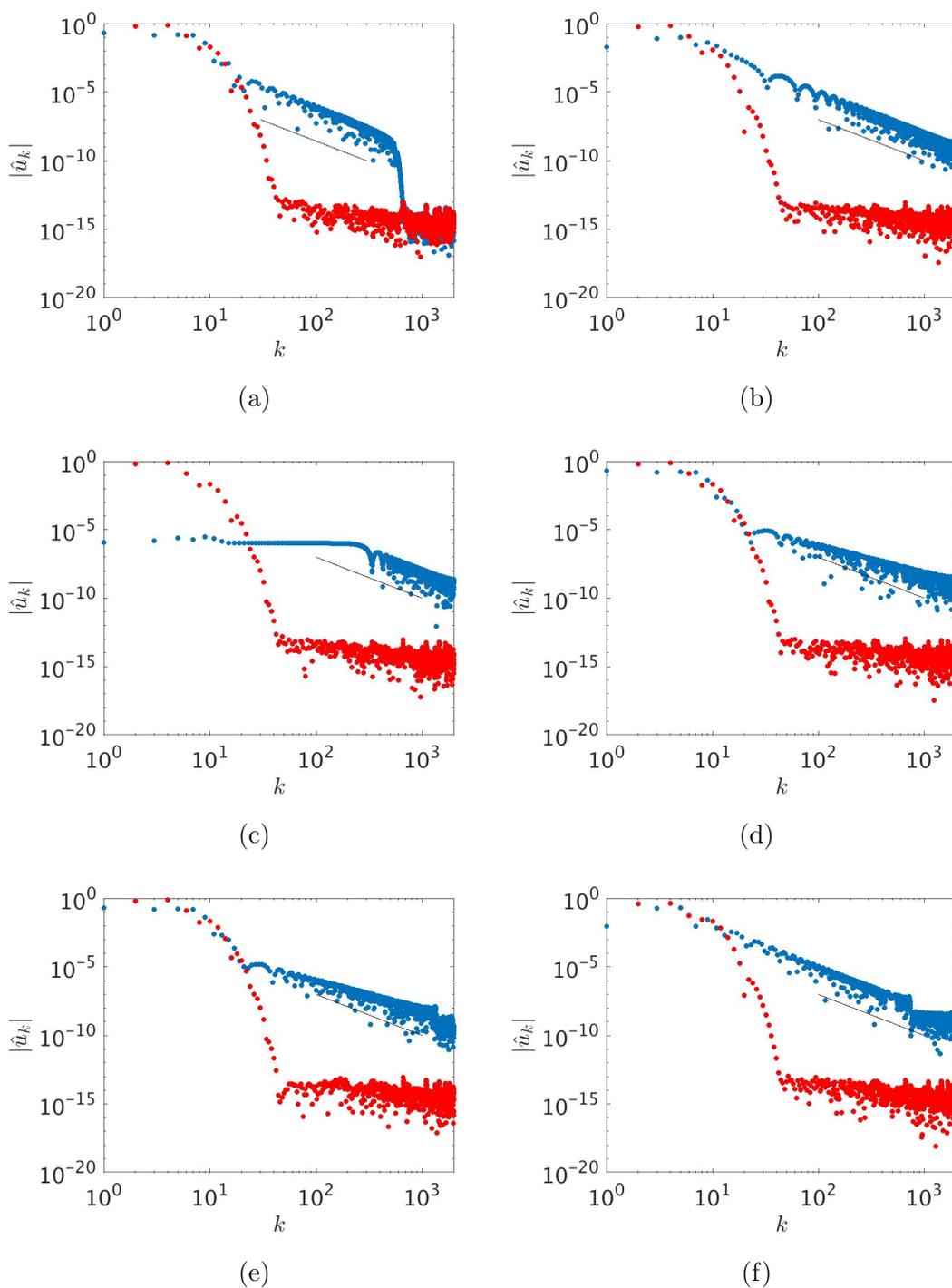


Figure 3.8: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with the trapezoidal method at $t = 0.04, 0.9, 1, 2, 4, 5.3$ (a, b, c, d, e, f panels respectively). The step size is $h = 10^{-3}$ with $N = 2000$ for the C^1 IC (3.82). The blue dots are the even coefficients component and the red dots are the odd coefficients component. The black straight line indicates order $\frac{1}{k^3}$.

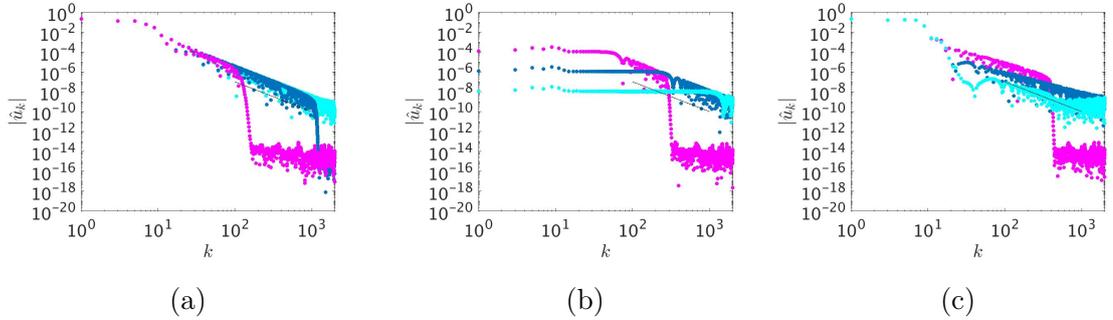


Figure 3.9: The even Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with the trapezoidal method for three step sizes $h = 10^{-2}, 10^{-3}, 10^{-4}$ (magenta, blue, cyan plots respectively). This is evaluated at $t = 0.16, 1, 2$ (a, b, c panels respectively) using the C^1 IC (3.82) with $N = 2000$.

For the purpose of better comprehension of all of the above spectra plots we plot the Chebyshev coefficients \hat{u}_k of the exact d'Alembert solution against the mode k for all three ICs, the C^∞ IC (3.81), the C^3 IC (3.83) and the C^1 IC (3.82) in Figures 3.10, 3.11 and 3.12, respectively. The figures explain behaviour of the convergence with N in Figure 3.5 more clearly as it shows that at $t = 0.9$ actually the spectra of the exact solution of the C^∞ IC (3.81) converges exponentially, while of the other ICs, the C^1 (3.82) and the C^3 (3.83), converge algebraically. The Chebyshev coefficients of the C^1 (3.82) is $\frac{1}{k^3}$, see Figure 3.12. And the Chebyshev coefficients of the C^3 IC (3.83) is $\frac{1}{k^5}$, see Figure 3.11, we found in Section 1.10 that the Fourier coefficients of the exact solution go as $\frac{1}{k^5}$ too, see (1.86).

The goal for now is to gain a clearer understanding of all the figures behaviour; by considering the figures of numerical solution spectra, Figures 3.6, 3.7, 3.8 and 3.9, and compare them with their corresponding figures of the exact solution spectra, Figures 3.10, 3.11 and 3.12.

The figures of the numerical solution spectra and the exact solution spectra of solving the wave equation (3.52) using the C^∞ IC (3.81), Figures 3.6 and 3.10, respectively, possess the same conduct, they drop off exponentially at all times.

However, it is obvious in all the spectra figures of solving the wave equation (3.52) using the C^3 IC (3.83) and the C^1 IC (3.82) that we have some specific times when

the figures are similar, the times are at $t = 0, 2, 4, 6, 8, \text{etc.}$, $t = 1, 3, 5, \text{etc.}$, and all the times in between, as we explained here in detail: We have observed from the figures that we get similar spectral coefficients plots of the numerical solution, the odd coefficients component and the even coefficients component, as their corresponding figures in Figures 3.7 and 3.8, respectively. That is, the figures at $t = 2, 6, 8, 10, \dots \text{etc.}$ are like their corresponding figures at $t = 4$; which means that at all even positive t the plots are similar. Likewise, the figures at $t = 3, 5, 7, \text{etc.}$, are similar to their corresponding figures at $t = 1$; which means that at all odd positive t the plots are similar. In addition, at all other times in between the plots are like their corresponding figures at $t = 0.9$. Furthermore, the same holds for the figures of the spectrum of the exact solution, Figures 3.11 and 3.12. Consequently, we get similar even coefficients plots conduct of the numerical solution in Figure 3.9 at $t = 4, 6, 8, 10, \text{etc.}$, as the plot at $t = 2$ and at $t = 3, 5, 7, \text{etc.}$, as the plot at $t = 1$. Moreover, at all other times the plots behaviour is equivalent as the plot at $t = 0.16$. In general, the spectra figures at each time group, mentioned above, are almost different than the figures at other times group; the figures have some special conducts and shapes as we explain next in detail.

Firstly, we will investigate the above explanation for the figure of the more interesting IC, the C^1 (3.82), of the exact solution spectra Figure 3.12 and of the numerical solution spectra Figures 3.8 and 3.9.

Three times are reviewed for the Chebyshev coefficients of the exact solution in Figure 3.12, $t = 0.9, 1$ and $t=4$. The figure at $t = 0.9$, Figure 3.12a, is not special because its exact d'Alembert solution, in equation (3.86), is not smooth at this time, thus, the even spectra component decay algebraically and the odd component decay exponentially. This is the case for all such times. However, the figure at $t = 1$, Figure 3.12b, is special and different from Figure 3.12a. The even component of the spectra are just round-off but the odd component remain decay exponentially like the odd component in Figure 3.12a. The same figure is obtained at all other odd positive times, $t = 3, 5, 7, \text{etc.}$ In addition, the figure at $t = 4$, Figure 3.12c, is special again and different from Figures 3.12a and 3.12b. In Figure 3.12c both spectra components (odd/even) change to decay exponentially not algebraically. The same figure is obtained at all other even positive times, $t = 2, 8, 10, \text{etc.}$ In order to understand this we go back to Section 1.9. In this section we have proved that the exact d'Alembert solution is very special. We have shown that the exact

d'Alembert solution at $t = 1$, in equation (1.59), is an odd function. Therefore, the even component of the exact d'Alembert solution is zero, thus, the Chebyshev component of the exact d'Alembert solution are just round-off in Figure 3.12b. This holds at all other odd positive times, $t = 3, 5, 7, \text{etc.}$ Also, we have proved that the exact d'Alembert solution is periodic of period 4 and the exact solution becomes the initial condition function $f(x)$, thus, it is infinitely smooth at $t = 4$, and at all times of multiple 4, i.e., at $t = 0, 4, 8, \text{etc.}$, Consequently, we see that the Chebyshev component of the exact d'Alembert solution in Figure 3.12c go back to decay exponentially again. Likewise, at the other times of multiple 2, i.e., at $t = 2, 6, 10, \text{etc.}$, the exact d'Alembert solutions is infinitely differentiable as well, infinitely smooth, so above holds here too. See Sections 1.9 and 3.4 for a better understanding.

It can be seen that six times are considered for the Chebyshev coefficients of the numerical solution in Figure 3.8, $t = 0.04, 0.9, 1, 2, 4$ and $t = 5.3$. Obviously, the figures are not identical but they are different in each time group explained above. For instance, the Chebyshev coefficients of the numerical solution spectra at $t = 2$ and $t = 4$, Figures 3.8d and 3.8e, respectively, are similar. That is the Chebyshev coefficients of the numerical solution at both times decay exponentially up to the time stepping error $\mathcal{O}(h^2)$ but after that they decay algebraically. More specifically, the even Chebyshev coefficients decay exponentially down to 10^{-6} for $h = 10^{-3}$ up to the time stepping error $\mathcal{O}(h^2)$ but after that they decay algebraically, which is consistent with the $\mathcal{O}(h^2)$ error expected for the trapezoid method. See the corresponding figure of the exact solution spectra, Figure 3.12c, described above. From Figure 3.9c one can see that the range giving exponential convergence grows as the time step decreases. The reason for this perhaps is the accumulated error that was occurring with the trapezoidal method at all the times before when the solutions were not smooth. We have examined and seen that, the same holds at any other times of multiple 2, i.e., at $t = 6, 8, 10, \text{etc.}$ However, the even Chebyshev coefficients of the numerical solution at $t = 1$, Figure 3.8c, are zero down to 10^{-6} for $h = 10^{-3}$, up to the time stepping error $\mathcal{O}(h^2)$, but after that they decay algebraically, which is consistent with the $\mathcal{O}(h^2)$ error expected for the trapezoid method. See the corresponding figure of the exact solution spectra, Figure 3.12b, described above. From Figure 3.9b one can see that the place where getting zero spectra goes down as the time step decreases. The above reason is valid here too.

Again we have examined and seen that, the same holds at any other odd positive times, i.e., at $t = 3, 5, 7, \text{etc.}$ The figures at all other times, Figures 3.8a, 3.8b and 3.8f, behave as the exact solution figure at $t = 0.9$, Figure 3.12a, the even spectra component decay algebraically and the odd component decay exponentially. This is the case for all such times.

Accordingly, the similar behaviours explained above are applicable for the third IC (3.83) in Figures 3.11 and 3.7, apart from the drop in as there is no fast drop in the plots of the numerical results for this IC as explained above. The conduct in Figure 3.9 happens for the C^3 IC (3.83) either at $t = 1$ and $t = 4$ in Figures 3.7b and 3.7c. See the corresponding figure of the exact solution spectra, Figures 3.11b and 3.11c, respectively. Figure 3.7c show that the numerical solution spectra decay exponentially down to 10^{-10} for $h = 10^{-3}$. One can see from Figures 3.7c and 3.8e, that the range of the exponential convergence is bigger for the C^3 IC (3.83) than for the IC (3.82) for the very same $h = 10^{-3}$ used. This holds at all times of multiple 2, i.e., at $t = 0, 2, 4, 6, 8, \text{etc.}$, for both ICs above respectively. Thus, one concludes that the trapezoidal method deals better with the C^3 IC (3.83) than the C^1 IC (3.82), probably mainly to the nonsmoothness problem presented more severely by the C^1 IC (3.82).

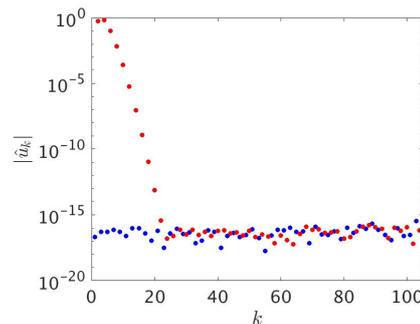


Figure 3.10: The Chebyshev coefficients $|\hat{u}_k|$ of the exact d'Alembert solution against the mode k for the wave equation (3.52) for the C^∞ IC (3.81) at $t = 0.9$. The coefficients are evaluated for $h = 10^{-3}$ and $N = 100$. Similar plot is obtained for $t = 1$ and $t = 4$.

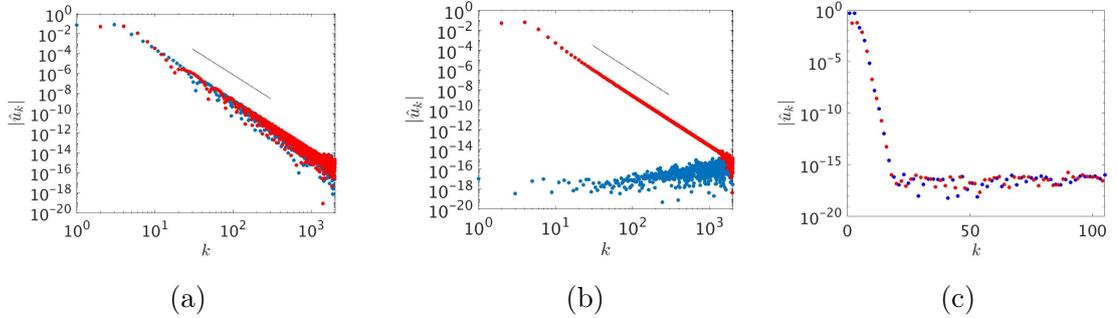


Figure 3.11: The Chebyshev coefficients $|\hat{u}_k|$ of the exact d'Alembert solution against the mode k for the wave equation (3.52) for the C^3 IC (3.83) at $t = 0.9, 1, 4$ (a, b, c respectively). The coefficients are evaluated for $h = 10^{-3}$ and $N = 2000$. The black straight line indicates order $\frac{1}{k^5}$.

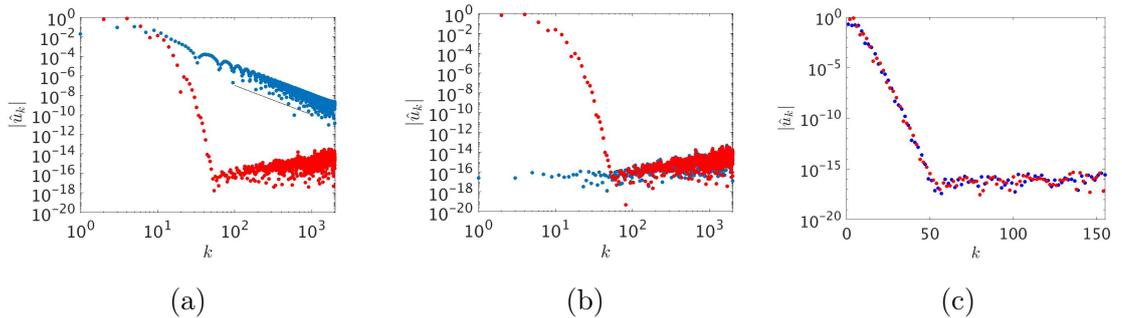


Figure 3.12: The Chebyshev coefficients $|\hat{u}_k|$ of the exact d'Alembert solution against the mode k for the wave equation (3.52) for the C^1 IC (3.82) at $t = 0.9, 1, 4$ (a, b, c respectively). The coefficients are evaluated for $h = 10^{-3}$ and $N = 2000$. The black straight line indicates order $\frac{1}{k^3}$.

In Sections 1.9 and 3.4 we have shown that the exact d'Alembert solutions for the linear wave equation (3.52) at any time which is of multiple 2, i.e., at $t = 0, 2, 4, 6, 8, \text{etc.}$, are infinitely smooth for all three ICs. Thus, it is of interest to examine the conduct of the numerical solution at these specified times to determine whether they behave the same as the exact solutions or not, especially for the two nonsmooth ICs, the C^1 (3.82) and the C^3 (3.83), as the first IC (3.81) is already infinitely smooth. Consequently, here we look at the numerical solution with respect to the spectrum since one cannot rely much on the plots of the numerical solution

with respect to x even if they indicate indistinguishable difference. Therefore, we plot the Chebyshev coefficients at $t = 4$ using the three ICs (3.81), (3.82) and (3.83) in Figures 3.6, 3.7 and 3.8, respectively. Figures shows that the numerical solution for the C^∞ (3.81) is as expected; it decays exponentially. On the contrary, the numerical solutions for the very nonsmooth IC, the C^1 (3.82), Figures 3.8d and 3.8e, are not very special at these specified times than at other times, in terms of still getting the algebraical convergence in the plots, as the Chebyshev coefficients of the numerical solution decay algebraically not exponentially after reaching the time stepping error $\mathcal{O}(h^2)$. More specifically, the Chebyshev coefficients of the numerical solution, of the trapezoidal method combined with pseudospectral method, decay exponentially down to 10^{-6} for $h = 10^{-3}$ (up to the time stepping error $\mathcal{O}(h^2)$) but after that they decay algebraically. The reason for this perhaps is the accumulated error that was occurring at all the times before when the solutions were not smooth. Nevertheless, our numerical results of using the C^3 (3.83) at $t = 4$ in Figure 3.7c show that the numerical solution spectra decay exponentially down to 10^{-10} for $h = 10^{-3}$ at this time, which means more faithfully then using the C^1 (3.82) in Figures 3.8d and 3.8e. This holds at all times of multiple 2, i.e., at $t = 0, 2, 4, 6, 8, etc.$, for both ICs above respectively.

As a result, one can observe and conclude that the numerical solution spectra using the trapezoidal method combined with the pseudospectral method attempts to mirror the behaviour of the exact solution spectra (recover smoothness of the solution) at the specified times described above, however, it is not able to reflect (recover) it faithfully but only partially to a limited extent.

Noteworthy, the figures of the spectra for the trapezoidal method combined with pseudospectral method behave similarly when considering other modes, N ; for instance, using $N = 1000$ or $N = 4000$ give us the same plots as plots of this section for both the C^1 IC (3.82) and the C^3 IC (3.83), where the plots are for $N = 2000$, but only the plot in the x -axis gets shorter or longer up to the N used; taking into consideration that N to be considered according to the values of N in the spatial convergence as shown in Figure 3.5. For instance, for the C^3 IC (3.83) N must be bigger than $N = 100$; any bigger N than that has no impact on the plots. In contrast, choosing different h changes the plots as shown and explained above.

3.4.5 Stability of Euler's method

To find the stability region of Euler's method when applied to the wave equation (3.52), the two equations (3.73) and (3.74) can be written as

$$\begin{aligned}y_{n+1} &= y_n + hw_n, \\w_{n+1} &= w_n + hD^{(2)}y_n,\end{aligned}$$

i.e.

$$\begin{pmatrix} y \\ w \end{pmatrix}_{n+1} = \begin{pmatrix} I & hI \\ hD^{(2)} & I \end{pmatrix} \begin{pmatrix} y \\ w \end{pmatrix}_n, n = 0, 1, \dots$$

Let

$$A = \begin{pmatrix} I & hI \\ hD^{(2)} & I \end{pmatrix}.$$

The eigenvalue equation of A is

$$A \begin{pmatrix} y \\ w \end{pmatrix} = \lambda \begin{pmatrix} y \\ w \end{pmatrix},$$

we can write this as

$$y + hw = \lambda y, \tag{3.88}$$

and

$$hD^{(2)}y + w = \lambda w. \tag{3.89}$$

Suppose that $\tilde{\lambda}$ is an eigenvalue of $D^{(2)}$: $D^{(2)}y = \tilde{\lambda}y$, where y is the eigenvector.

Then $y + hw = \lambda y$ implies

$$w = \frac{\lambda - 1}{h}y. \tag{3.90}$$

Substituting $D^{(2)}y = \tilde{\lambda}y$ in (3.89), we get $h\tilde{\lambda}y + w = \lambda w$ and by substituting (3.90) in this equation we get

$$h\tilde{\lambda}y + \frac{\lambda - 1}{h}y = \lambda \frac{\lambda - 1}{h}y,$$

which leads to

$$h^2\tilde{\lambda} + \lambda - 1 = \lambda(\lambda - 1).$$

Thus, we find that the eigenvalues for Euler's method are

$$\lambda = 1 \pm \sqrt{h^2\tilde{\lambda}} = 1 \pm h\sqrt{\tilde{\lambda}}. \tag{3.91}$$

For the method to be stable, the modulus of (3.91) needs to be less than one, i.e., $|\lambda| < 1$. However, Weideman & Trefethen (1988) said that the eigenvalues $\tilde{\lambda}$ of $D^{(2)}$ are all negative, hence

$$|\lambda| = \left| 1 \pm h\sqrt{\tilde{\lambda}} \right| = \left| 1 \pm ih\sqrt{|\tilde{\lambda}|} \right| = \sqrt{1 + h^2\tilde{\lambda}} > 1. \quad (3.92)$$

We conclude that $|\lambda| > 1$. In other words, Euler's method for the wave equation (3.52) is always unstable. This is illustrated experimentally in Figure 3.13. It can be seen in the left panel that the error grows rapidly as time increases.

3.4.6 Stability of the modified Euler method

The modified Euler method for the wave equation (3.52) is also unstable. This is illustrated in Figure 3.13, the right panel, which shows that, the error grows as time increases.

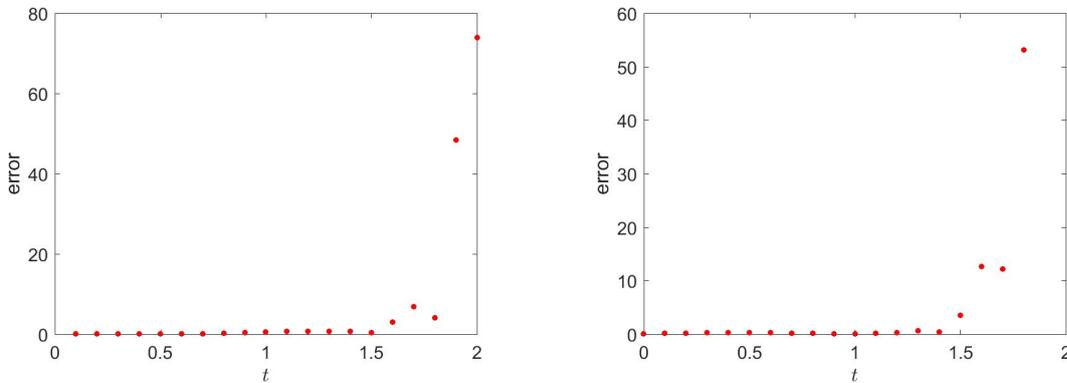


Figure 3.13: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the time t for the wave equation (3.52) solved using the pseudospectral method for the C^∞ initial condition (3.81) with Euler's method, the left panel, and with the modified Euler method, the right panel. The Chebyshev series is truncated at $N = 10$ and the step size is $h = 0.01$.

3.4.7 Stability of the implicit Euler and the trapezoidal methods

The implicit Euler and the trapezoidal methods are considered to solve the wave equation using the C^∞ IC (3.81) and the C^1 IC (3.82). In order to study the stability two main figures are obtained, Figures 3.14 and 3.15, where the exact solution, u , and the numerical solution, u_{num} , at $x = \frac{1}{2}$ are plotted. From the right panels of the figures one can notice that the numerical solutions using the implicit Euler method decrease and approach zero as time $\rightarrow \infty$. Moving to the trapezoidal method, clearly from the left panels of the figures the numerical solutions are bounded and vibrate between -2 and 2 . We conclude that both methods, the implicit Euler and trapezoidal, are stable, but it is important to know which one is better. The solutions of the implicit Euler method decay but the solutions of the trapezium method oscillate with constant amplitude. Since the exact solution does the same, the trapezium method reproduces the exact solution more faithfully. This is more obvious in Figure 3.14.

In order to understand these figures, further investigation is done next. In addition, some more graphs for the numerical solution, u_{num} , are given, Figures 3.16 and 3.17. These figures give a clearer idea about the behaviour of numerical solution for both initial conditions. In Figure 3.16 the contour plots look smooth for both initial conditions, but the plot for the first initial condition (3.81) is more regular and symmetrical. The contour plots using the IC (3.81) is regular circles but for the IC (3.82) the graph is not having an organized shape. In Figure 3.17 the numerical solution using the IC (3.81) gives very regular and symmetric plot, specifically, it is the same from -1 to 0 and from 0 to 1 except for a sign, this solution is symmetric about $x = 0$. Whereas, this is not the case when the other initial condition is used.

Figures, 3.16 and 3.17, show that the numerical solutions are periodic of period 4, similar to the exact d'Alembert solutions in Section 1.9, for the ICs implemented. Here the first IC (3.81) is a special case and the period is indistinguishable at also period 2, see the left panels of the figures.

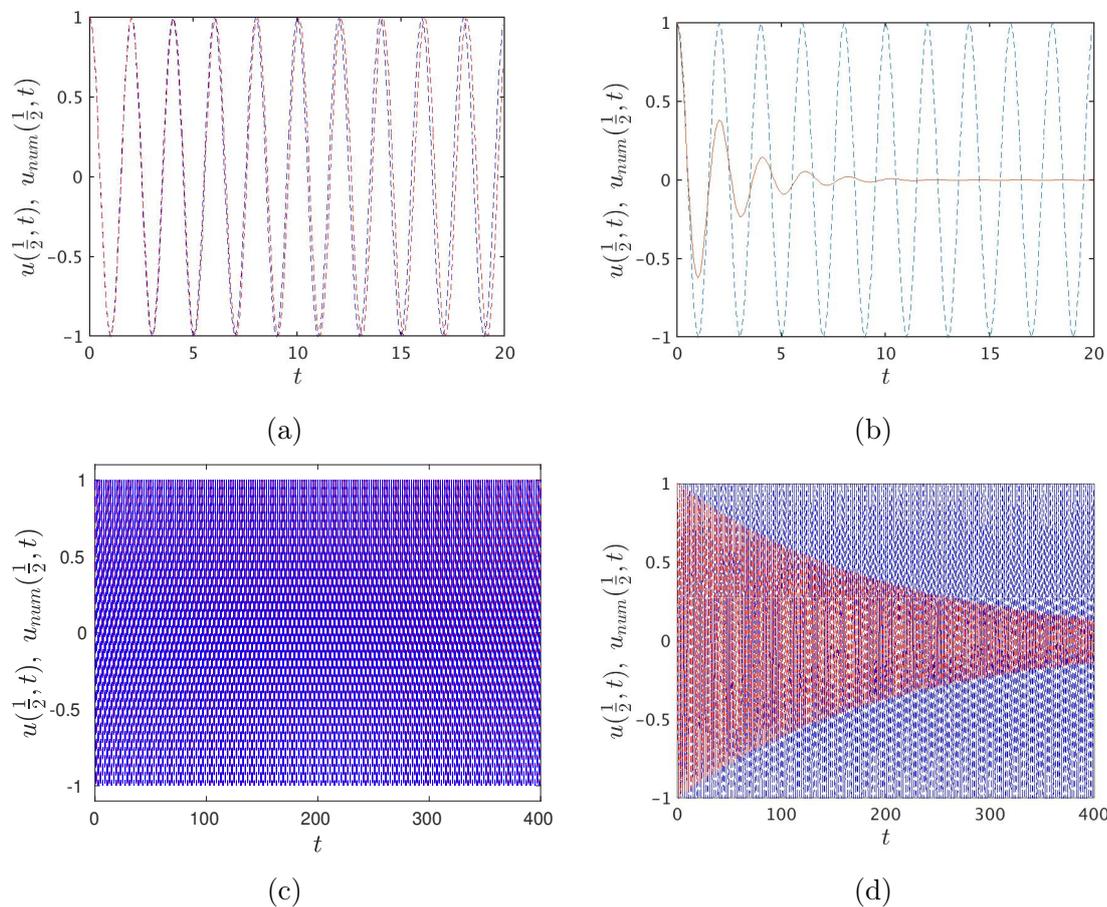


Figure 3.14: The exact solution $u(x, t)$, blue line, and the numerical solution $u_{num}(x, t)$, red line, at $x = \frac{1}{2}$, for the wave equation (3.52) solved for $N = 10$. Using the pseudospectral method with the trapezoidal method (a, c panels), $\theta = 0.5$, and the implicit Euler method (b, d panels), $\theta = 1$, for the C^∞ initial condition (3.81). Using two time steps, $h = 10^{-1}$ (a, b panels) and $h = 10^{-3}$ (c, d panels).

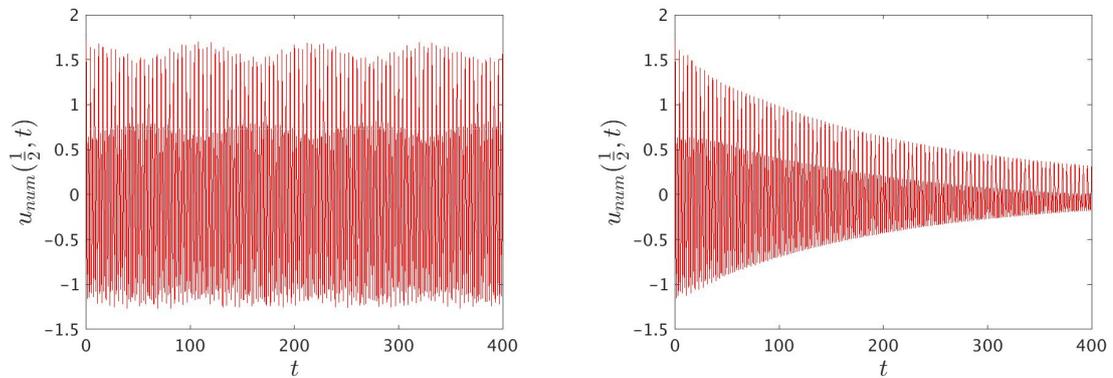


Figure 3.15: The numerical solution, $u_{num}(x, t)$, at $x = \frac{1}{2}$, for the wave equation (3.52). Using the pseudospectral method with the trapezoidal method (left panel), $\theta = 0.5$, and the implicit Euler method (right panel), $\theta = 1$, for the C^1 initial condition (3.82), for $N = 10$ and $h = 10^{-3}$.

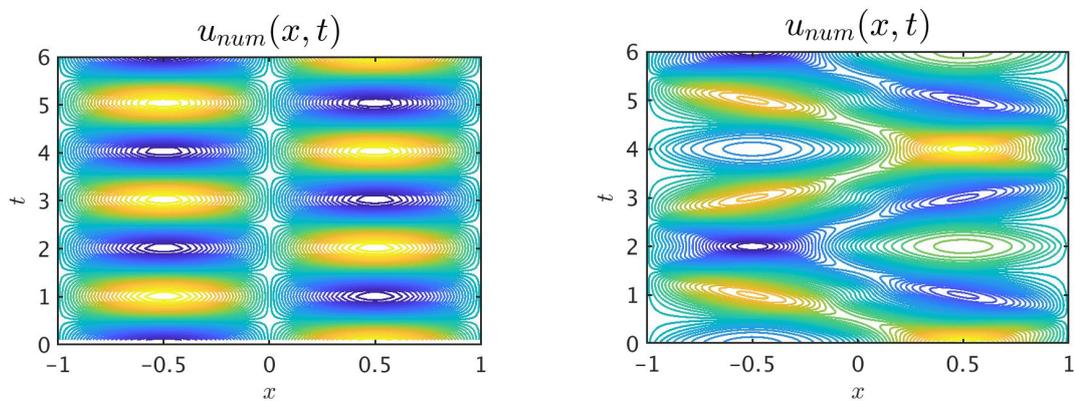


Figure 3.16: The contour of the numerical solution $u_{num}(x, t)$ for the wave equation (3.52), using the pseudospectral method with the trapezoidal method for the initial conditions, the C^∞ IC (3.81)(left panel) and the C^1 IC (3.82)(right panel), at $N = 100$ and $h = 10^{-2}$.

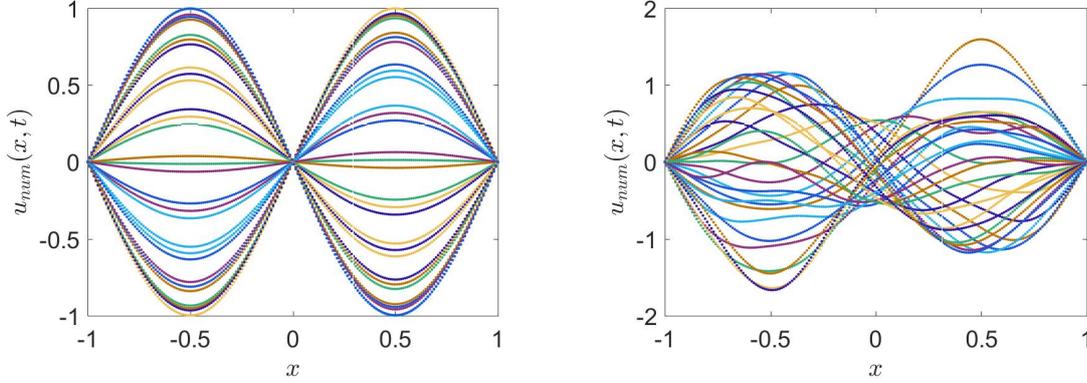


Figure 3.17: The numerical solution $u_{num}(x, t)$ for the wave equation (3.52), using the pseudospectral method with the trapezoidal method for the initial conditions, the C^∞ IC (3.81)(left panel) and the C^1 IC (3.82)(right panel), at $N = 100$, $h = 10^{-2}$ and for $t = 0 - 3$.

3.5 The collocation method

The collocation method is mainly in transform space and the pseudospectral method is in physical space. But in most cases the collocation method is equivalent to the pseudospectral method as we have found in our case. The collocation methods cancel the residual on a set of grid points called collocation points (the collocation points are usually the same as the quadrature points),

$$R(x_j) = 0, \quad j = 1, \dots, N - 1, \quad (3.93)$$

i.e.,

$$\left. \frac{\partial^2 u_{sd}}{\partial t^2} - \frac{\partial^2 u_{sd}}{\partial x^2} \right|_{x=x_j} = 0, \quad j = 1, \dots, N - 1. \quad (3.94)$$

3.5.1 Spatial discretization

The coefficient $\hat{u}_k^{(2)}$ obviously depend linearly on the nodal values $u_{sd}(x_l, t)$, hence there exists a matrix $D^{(2)}$ such that

$$\frac{\partial^2 u_j(t)}{\partial x^2} = \sum_{k=0}^N \hat{u}_k^{(2)}(t) T_k(x_j) = \sum_{k=0}^N \sum_{l=0}^N D_{kl}^{(2)} \hat{u}_l(t) T_k(x_j), \quad j = 1, \dots, N - 1. \quad (3.95)$$

Substituting (3.95) into (3.94) we get

$$\frac{\partial^2 u_j(t)}{\partial t^2} = \sum_{k=0}^N \sum_{l=0}^N D_{kl}^{(2)} \hat{u}_l(t) T_k(x_j), \quad j = 1, \dots, N-1. \quad (3.96)$$

and by (3.63), we have $u_0(0) = u_N(0) = 0$.

The ODE (3.96) can be written as

$$\frac{\partial y_j(t)}{\partial t} = w_j(t), \quad (3.97)$$

$$\frac{\partial w_j(t)}{\partial t} = \sum_{k=0}^N \sum_{l=0}^N D_{kl}^{(2)} \hat{y}_l(t) T_k(x_j). \quad (3.98)$$

3.5.2 Time stepping

Euler's method is used to solve (3.97) and (3.98) and this leads to

$$y_j(t+h) = y_j(t) + hw_j(t), \quad (3.99)$$

$$w_j(t+h) = w_j(t) + \sum_{k=0}^N \sum_{l=0}^N D_{kl}^{(2)} \hat{y}_l(t) T_k(x_j) \quad j = 1, \dots, N-1, \quad (3.100)$$

where from the boundary conditions (3.55) we have $y_0(0) = y_N(0) = 0$ and $w_0(0) = w_N(0) = 0$, and from the initial conditions (3.56) and (3.57) we have

$$y_j(0) = f(x_j) \quad \text{and} \quad w_j(0) = 0, \quad j = 1, \dots, N-1.$$

The value of $\hat{y}(0)$ can be found from (3.60) and the initial conditions as

$$\hat{u}_l(0) = \frac{2}{Nc_l} \sum_{j=0}^N \frac{1}{c_j} u_j(0) T_l(x_j), \quad c_j = 1 + \delta_{j0} + \delta_{jN}, \quad l = 0, \dots, N, \quad (3.101)$$

so

$$\hat{y}_l(0) = \frac{2}{Nc_l} \sum_{j=0}^N \frac{1}{c_j} y_j(0) T_l(x_j), \quad c_j = 1 + \delta_{j0} + \delta_{jN}, \quad l = 0, \dots, N. \quad (3.102)$$

Also, the trapezoidal method is implemented to solve (3.97) and (3.98).

3.6 Numerical results

3.6.1 Convergence of Euler's and the trapezoidal methods

The results of figures, show that the collocation method for solving wave equation (3.52) is almost the same as the pseudospectral method, in terms they show that for the C^∞ initial condition (3.81), both Euler's and the trapezoidal methods combined with the collocation method converge exponentially in N . Also, for this IC Euler's method is $\mathcal{O}(h)$ and the trapezoidal method is $\mathcal{O}(h^2)$. However, there are some differences in the figures, for example the convergence in space results, the figures for Euler's method show that the first three dots in Figure 3.2 are identically the same but dots after that are between 10^{-6} and 10^{-7} not between 10^{-4} and 10^{-5} as for the pseudospectral method that are up to the $\mathcal{O}(h)$ error. More precisely, we get four points that decrease not only three; that starting from $N = 12$ not $N = 10$ the dots remain in the same place, between 10^{-6} and 10^{-7} . The same holds for the figures of the trapezoidal method, see Figure 3.5a, the differences are at the points down to the $\mathcal{O}(h^2)$ error only; that the dots are below the $\mathcal{O}(h^2)$ error and we have additional dot that decreases.

Overall, only the C^∞ IC (3.81) is considered when using the collocation method; the other two ICs are not examined.

3.6.2 Stability of Euler's, the implicit Euler and the trapezoidal methods

Regarding the stability properties, from some figures, not shown, it seems that we obtain the same results to the pseudospectral method. This when the method is used in solving the wave equation (3.52) with the time stepping methods, Euler's, the implicit Euler and the trapezoidal methods for the C^∞ IC (3.81). That is, Euler's method is not stable, while the other two methods are stable; the numerical solution using the implicit Euler method decreases and approaches zero as time $\rightarrow \infty$. And for the trapezoidal method, the numerical solution is bounded and vibrate between -1 and 1 .

3.7 The tau method

The basis functions that were used are the trial functions, i.e., $T_k(x)$ for $k = 0, \dots, N$, and they do not satisfy the boundary conditions. The boundary conditions are enforced by an additional set of equations. This method requires that the residual be orthogonal to the basis functions. Therefore, we have to solve simultaneously for the $N - 1$ equations

$$\langle T_q, R \rangle = 0, \quad q = 0, \dots, N - 2, \quad (3.103)$$

and for the two boundary conditions $u(-1, t) = 0$ and $u(1, t) = 0$, where

$$R = \frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2}.$$

3.7.1 Spatial discretization

We have

$$u_{sd}(x, t) = \sum_{k=0}^N \hat{u}_k(t) T_k(x), \quad (3.104)$$

where

$$T_k(x) = \cos(k \cos^{-1}(x)), \quad k = 0, 1, \dots, N.$$

The exact derivatives of (3.104) are

$$\frac{\partial^2 u_{sd}(x, t)}{\partial t^2} = \sum_{k=0}^N \frac{d^2}{dt^2} \hat{u}_k(t) T_k(x), \quad (3.105)$$

and

$$\frac{\partial^2 u_{sd}(x, t)}{\partial x^2} = \sum_{k=0}^N \sum_{j=0}^N D_{kj}^{(2)} \hat{u}_j(t) T_k(x). \quad (3.106)$$

Now by substituting (3.105) and (3.106) in (3.103) we get

$$\left\langle T_q, \sum_{k=0}^N \frac{d^2}{dt^2} \hat{u}_k(t) T_k - \sum_{k=0}^N \sum_{j=0}^N D_{kj}^{(2)} \hat{u}_j(t) T_k \right\rangle = 0, \quad q = 0, \dots, N - 2.$$

Expanding the inner product yields

$$\sum_{k=0}^N \frac{d^2}{dt^2} \hat{u}_k(t) \langle T_q, T_k \rangle - \sum_{k=0}^N \sum_{j=0}^N D_{kj}^{(2)} \hat{u}_j(t) \langle T_q, T_k \rangle = 0. \quad (3.107)$$

We have $\langle T_q, T_k \rangle = 0$ if $q \neq k$, see (3.8), so this leads (3.107) to become

$$\frac{d^2}{dt^2} \hat{u}_q(t) \langle T_q, T_q \rangle - \sum_{j=0}^N D_{qj}^{(2)} \hat{u}_j(t) \langle T_q, T_q \rangle = 0, \quad (3.108)$$

and using (3.9) this becomes

$$\frac{d^2}{dt^2} \hat{u}_q(t) \gamma_q = \sum_{j=0}^N D_{qj}^{(2)} \hat{u}_j(t) \gamma_q,$$

from which it follows that

$$\frac{d^2}{dt^2} \hat{u}_q(t) = \sum_{j=0}^N D_{qj}^{(2)} \hat{u}_j(t). \quad (3.109)$$

The ODEs (3.109) can be written as

$$\frac{d}{dt} \hat{y}_q(t) = \hat{w}_q(t), \quad (3.110)$$

$$\frac{d}{dt} \hat{w}_q(t) = \sum_{j=0}^N D_{qj}^{(2)} \hat{y}_j(t). \quad (3.111)$$

3.7.2 Time stepping

By applying Euler's method to (3.110) and (3.111) we get

$$\hat{y}_q(t+h) = \hat{y}_q(t) + h \hat{w}_q(t), \quad (3.112)$$

$$\hat{w}_q(t+h) = \hat{w}_q(t) + h \sum_{j=0}^N D_{qj}^{(2)} \hat{y}_j(t). \quad (3.113)$$

Substituting the boundary conditions into (3.104) yields

$$\sum_{k=0}^N \hat{u}_k(t) = 0, \quad \text{and} \quad \sum_{k=0}^N \hat{u}_k(t) (-1)^k = 0. \quad (3.114)$$

Because the basis functions do not satisfy the boundary conditions, the two last rows (highest order terms) of the system have to be replaced by the relation (3.114), so this leads (3.112) and (3.113) to be written as

$$A \begin{pmatrix} \vec{\hat{y}} \\ \vec{\hat{w}} \end{pmatrix}^{n+1} = B \begin{pmatrix} \vec{\hat{y}} \\ \vec{\hat{w}} \end{pmatrix}^n, n = 0, 1, \dots,$$

where the matrices A and B are of the forms

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & & & & & & & & & & & \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 1 & 1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & -1 & 1 & \dots & (-1)^{N-2} & (-1)^{N-1} & (-1)^N & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & & & & & & & & & \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 & 0 & h & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 & 0 & 0 & h & 0 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & & & & & & & & \\ 0 & 0 & 0 & \dots & 1 & 0 & 0 & 0 & 0 & 0 & \dots & h & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ hD_{0,0}^{(2)} & hD_{0,1}^{(2)} & hD_{0,2}^{(2)} & \dots & hD_{0,N-2}^{(2)} & hD_{0,N-1}^{(2)} & hD_{0,N}^{(2)} & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ hD_{1,0}^{(2)} & hD_{1,1}^{(2)} & hD_{1,2}^{(2)} & \dots & hD_{1,N-2}^{(2)} & hD_{1,N-1}^{(2)} & hD_{1,N}^{(2)} & 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ \vdots & & & & & & & & & & & & & \\ hD_{N,0}^{(2)} & hD_{N,1}^{(2)} & hD_{N,2}^{(2)} & \dots & hD_{N,N-2}^{(2)} & hD_{N,N-1}^{(2)} & hD_{N,N}^{(2)} & 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}.$$

Both matrices are of size $(2N + 2) \times (2N + 2)$.

The value of $\vec{\hat{y}}(0)$ and $\vec{\hat{w}}(0)$ can be found from (3.60) and the initial conditions as

$$\hat{u}_l(0) = \frac{2}{Nc_l} \sum_{j=0}^N \frac{1}{c_j} u_j(0) T_l(x_j), \quad c_j = 1 + \delta_{j0} + \delta_{jN}, l = 0, \dots, N, \quad (3.115)$$

so

$$\hat{y}_l(0) = \frac{2}{Nc_l} \sum_{j=0}^N \frac{1}{c_j} y_j(0) T_l(x_j), \quad c_j = 1 + \delta_{j0} + \delta_{jN}, l = 0, \dots, N, \quad (3.116)$$

and

$$\vec{w}(0) = 0. \quad (3.117)$$

In addition to Euler's method, the trapezoidal rule is also applied to solve (3.110) and (3.111)

$$\hat{y}_q(t+h) = \hat{y}_q(t) + \frac{h}{2} \{ \hat{w}_q(t) + \hat{w}_q(t+h) \}, \quad (3.118)$$

$$\hat{w}_q(t+h) = \hat{w}_q(t) + \frac{h}{2} \left\{ \sum_{j=0}^N D_{qj}^{(2)} \hat{y}_j(t) + \sum_{j=0}^N D_{qj}^{(2)} \hat{y}_j(t+h) \right\}. \quad (3.119)$$

By applying the relation (3.114), we get

$$A \begin{pmatrix} \vec{\hat{y}} \\ \vec{\hat{w}} \end{pmatrix}^{n+1} = B \begin{pmatrix} \vec{\hat{y}} \\ \vec{\hat{w}} \end{pmatrix}^n, \quad n = 0, 1, \dots,$$

where A is of the form

$$A = \begin{pmatrix} A_{11} & 0 \\ A_{21} & 0 \\ 0 & A_{32} \end{pmatrix}_{(2N+2) \times (2N+2)}.$$

With the blocks A_{11} , A_{21} and A_{32}

$$A_{11} = \begin{pmatrix} 1 - \frac{h^2}{4} D_{0,0}^{(2)} & -\frac{h^2}{4} D_{0,1}^{(2)} & \cdots & -\frac{h^2}{4} D_{0,N-2}^{(2)} & -\frac{h^2}{4} D_{0,N-1}^{(2)} & -\frac{h^2}{4} D_{0,N}^{(2)} \\ -\frac{h^2}{4} D_{1,0}^{(2)} & 1 - \frac{h^2}{4} D_{1,1}^{(2)} & \cdots & -\frac{h^2}{4} D_{1,N-2}^{(2)} & -\frac{h^2}{4} D_{1,N-1}^{(2)} & -\frac{h^2}{4} D_{1,N}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{h^2}{4} D_{N-2,0}^{(2)} & -\frac{h^2}{4} D_{N-2,1}^{(2)} & \cdots & 1 - \frac{h^2}{4} D_{N-2,N-2}^{(2)} & -\frac{h^2}{4} D_{N-2,N-1}^{(2)} & -\frac{h^2}{4} D_{N-2,N}^{(2)} \end{pmatrix}$$

of size $(N-1) \times (N+1)$, where the third column is removed to get good layout,

$$A_{21} = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 & 1 & 1 \\ 1 & -1 & 1 & \cdots & (-1)^{N-1} & (-1)^N & (-1)^{N+1} \end{pmatrix}_{(2) \times (N+1)}$$

and

$$A_{32} = \begin{pmatrix} 1 - \frac{h^2}{4} D_{0,0}^{(2)} & -\frac{h^2}{4} D_{0,1}^{(2)} & -\frac{h^2}{4} D_{0,2}^{(2)} & \cdots & -\frac{h^2}{4} D_{0,N-2}^{(2)} & -\frac{h^2}{4} D_{0,N-1}^{(2)} & -\frac{h^2}{4} D_{0,N}^{(2)} \\ -\frac{h^2}{4} D_{1,0}^{(2)} & 1 - \frac{h^2}{4} D_{1,1}^{(2)} & -\frac{h^2}{4} D_{1,2}^{(2)} & \cdots & -\frac{h^2}{4} D_{1,N-2}^{(2)} & -\frac{h^2}{4} D_{1,N-1}^{(2)} & -\frac{h^2}{4} D_{1,N}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -\frac{h^2}{4} D_{N,0}^{(2)} & -\frac{h^2}{4} D_{N,1}^{(2)} & -\frac{h^2}{4} D_{N,2}^{(2)} & \cdots & -\frac{h^2}{4} D_{N,N-2}^{(2)} & -\frac{h^2}{4} D_{N,N-1}^{(2)} & 1 - \frac{h^2}{4} D_{N,N}^{(2)} \end{pmatrix}$$

of size $(N + 1) \times (N + 1)$.

And the matrix B is of the form

$$B = \begin{pmatrix} B_{11} & B_{13} \\ 0 & 0 \\ B_{31} & B_{33} \end{pmatrix}_{(2N+2) \times (2N+2)}.$$

With the blocks B_{11} , B_{13} , B_{31} and B_{33}

$$B_{11} = \begin{pmatrix} 1 + \frac{h^2}{4} D_{0,0}^{(2)} & \frac{h^2}{4} D_{0,1}^{(2)} & \frac{h^2}{4} D_{0,2}^{(2)} & \cdots & \frac{h^2}{4} D_{0,N-2}^{(2)} & \frac{h^2}{4} D_{0,N-1}^{(2)} & \frac{h^2}{4} D_{0,N}^{(2)} \\ \frac{h^2}{4} D_{1,0}^{(2)} & 1 + \frac{h^2}{4} D_{1,1}^{(2)} & \frac{h^2}{4} D_{1,2}^{(2)} & \cdots & \frac{h^2}{4} D_{1,N-2}^{(2)} & \frac{h^2}{4} D_{1,N-1}^{(2)} & \frac{h^2}{4} D_{1,N}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{h^2}{4} D_{N-2,0}^{(2)} & \frac{h^2}{4} D_{N-2,1}^{(2)} & \frac{h^2}{4} D_{N-2,2}^{(2)} & \cdots & 1 + \frac{h^2}{4} D_{N-2,N-2}^{(2)} & \frac{h^2}{4} D_{N-2,N-1}^{(2)} & \frac{h^2}{4} D_{N-2,N}^{(2)} \end{pmatrix}$$

of size $(N - 1) \times (N + 1)$,

$$B_{13} = \begin{pmatrix} h & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & h & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & h & 0 & 0 \end{pmatrix}_{(N-1) \times (N+1)}$$

$$B_{31} = \begin{pmatrix} hD_{0,0}^{(2)} & hD_{0,1}^{(2)} & hD_{0,2}^{(2)} & \cdots & hD_{0,N-2}^{(2)} & hD_{0,N-1}^{(2)} & hD_{0,N}^{(2)} \\ hD_{1,0}^{(2)} & hD_{1,1}^{(2)} & hD_{1,2}^{(2)} & \cdots & hD_{1,N-2}^{(2)} & hD_{1,N-1}^{(2)} & hD_{1,N}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ hD_{N,0}^{(2)} & hD_{N,1}^{(2)} & hD_{N,2}^{(2)} & \cdots & hD_{N,N-2}^{(2)} & hD_{N,N-1}^{(2)} & hD_{N,N}^{(2)} \end{pmatrix}$$

of size $(N + 1) \times (N + 1)$,

$$B_{33} = \begin{pmatrix} 1 + \frac{h^2}{4} D_{0,0}^{(2)} & \frac{h^2}{4} D_{0,1}^{(2)} & \frac{h^2}{4} D_{0,2}^{(2)} & \cdots & \frac{h^2}{4} D_{0,N-2}^{(2)} & \frac{h^2}{4} D_{0,N-1}^{(2)} & \frac{h^2}{4} D_{0,N}^{(2)} \\ \frac{h^2}{4} D_{1,0}^{(2)} & 1 + \frac{h^2}{4} D_{1,1}^{(2)} & \frac{h^2}{4} D_{1,2}^{(2)} & \cdots & \frac{h^2}{4} D_{1,N-2}^{(2)} & \frac{h^2}{4} D_{1,N-1}^{(2)} & \frac{h^2}{4} D_{1,N}^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{h^2}{4} D_{N,0}^{(2)} & \frac{h^2}{4} D_{N,1}^{(2)} & \frac{h^2}{4} D_{N,2}^{(2)} & \cdots & 1 + \frac{h^2}{4} D_{N,N-2}^{(2)} & \frac{h^2}{4} D_{N,N-1}^{(2)} & \frac{h^2}{4} D_{N,N}^{(2)} \end{pmatrix}$$

of size $(N + 1) \times (N + 1)$.

3.8 Numerical results

3.8.1 Convergence of Euler's and the trapezoidal methods

The results of figures, show that the tau method for solving wave equation (3.52) is almost the same as the pseudospectral method, in terms they show that for the C^∞

initial condition (3.81), both Euler's and the trapezoidal methods combined with the collocation method converge exponentially in N . Also, for this IC Euler's method is $\mathcal{O}(h)$ and the trapezoidal method is $\mathcal{O}(h^2)$. However, there are some differences in the figures, for example when using Euler's method, the figures of the convergence in space show that the first three dots in Figure 3.2 are not identically the same but are slightly higher; they start at $4 \cdot 10^{-1}$ not at $2 \cdot 10^{-1}$, as for the pseudospectral method in Figure 3.2, and so on about the other two points, however, the dots after that are up to the $\mathcal{O}(h)$ error; between 10^{-4} and 10^{-5} , identically the same as in Figure 3.2. The same holds for the figures of the trapezoidal method, see Figure 3.5c, the differences are at the points before the $\mathcal{O}(h^2)$ error only; that the dots are slightly higher.

Overall, only the C^∞ IC (3.81) is considered when using the tau method; the other two ICs are not addressed.

3.8.2 Stability of Euler's, the implicit Euler and the trapezoidal methods

Figure 3.18 shows that the numerical solutions of the explicit Euler method, the implicit Euler method, and the trapezoidal method are all increasing for large time. Therefore, one can confidently say all three methods are not stable for solving the wave equation (3.52) using the tau method.

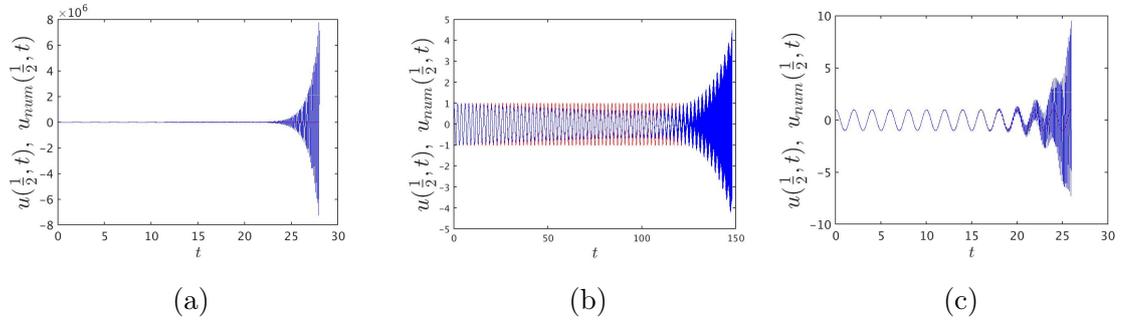


Figure 3.18: The exact solution $u(x,t)$, red line, and the numerical solution $u_{num}(x,t)$, blue line, at $x = \frac{1}{2}$, for the wave equation (3.52) solved using the tau method, with the C^∞ initial condition (3.81), $N = 10$ and the step size is $h = 10^{-3}$. Panel (a) shows the explicit Euler method, panel (b) shows the implicit Euler method, and panel (c) shows the trapezoidal method.

3.9 The Galerkin method

The Galerkin method assumes that the boundary conditions of the problem are satisfied by all basis functions $\phi_k(x)$. We choose the test functions to be same as the basis functions, i.e. $T_k(x)$ for $k = 0, \dots, N$, where $T_k(\pm 1) = (\pm 1)^k$.

In order to solve the wave equation (3.52) using the Galerkin method, we need to find a new basis, $\{\phi_k(x) : k = 0, 1, \dots, N-2\}$, that satisfies the boundary conditions $\phi_k(\pm 1) = 0$. We can define this basis in terms of the Chebyshev polynomials as

$$\phi_{2l}(x) = T_{2(l+1)}(x) - T_0(x) \quad \text{and} \quad \phi_{2l+1}(x) = T_{2l+3}(x) - T_1(x). \quad (3.120)$$

As before, we impose the conditions $\langle \phi_k(x), R \rangle = 0$, meaning that the residual is orthogonal to the trial space.

3.9.1 Spatial discretization

The basis $\{\phi_k(x) : k = 0, 1, \dots, N-2\}$ can be written with respect to the basis $\{T_k(x) : k = 0, 1, \dots, N\}$ as follows

$$\phi_l(x) = \sum_{k=0}^N \Phi_{kl} T_k(x), \quad \text{for all } l = 0, \dots, N-2, \quad (3.121)$$

$$\phi(x) = \Phi^T T(x).$$

We illustrate this by a simple example; for $N = 4$ we have,

$$\Phi = \begin{pmatrix} -1 & 0 & -1 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The semi-discretization solution can be expanded in the basis of $\{T_k\}$ or $\{\phi_k\}$:

$$u_{sd}(x, t) = \sum_{k=0}^N \hat{u}_k(t) T_k(x) = \sum_{l=0}^{N-2} \hat{v}_l(t) \phi_l(x). \quad (3.122)$$

Substituting equation (3.121) in (3.122) gives

$$u_{sd}(x, t) = \sum_{k=0}^N \hat{u}_k(t) T_k(x) = \sum_{l=0}^{N-2} \hat{v}_l(t) \sum_{k=0}^N \Phi_{kl} T_k(x) = \sum_{k=0}^N \left(\sum_{l=0}^{N-2} \hat{v}_l(t) \Phi_{kl} \right) T_k(x), \quad (3.123)$$

so

$$\hat{u}_k(t) = \sum_{l=0}^{N-2} \hat{v}_l(t) \Phi_{kl}. \quad (3.124)$$

The Galerkin approach for solving the wave equation (3.52) leads to the equations $\langle \phi_q, R \rangle = 0$, i.e., $\langle \phi_q, u_{tt} - u_{xx} \rangle = 0$, where $q = 0, \dots, N-2$. Using (3.122) we get

$$\left\langle \phi_q, \left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} \right) \sum_{l=0}^{N-2} \hat{v}_l(t) \phi_l \right\rangle = 0,$$

which implies

$$\sum_{l=0}^{N-2} \frac{d^2}{dt^2} \hat{v}_l(t) \langle \phi_q, \phi_l \rangle - \sum_{l=0}^{N-2} \hat{v}_l(t) \left\langle \phi_q, \frac{d^2}{dx^2} \phi_l \right\rangle = 0.$$

By substituting equation (3.121) into this equation we get

$$\sum_{l=0}^{N-2} \left(\frac{d^2}{dt^2} \hat{v}_l(t) \left\langle \sum_{k=0}^N \Phi_{kq} T_k, \sum_{p=0}^N \Phi_{pl} T_p \right\rangle - \hat{v}_l(t) \left\langle \sum_{k=0}^N \Phi_{kq} T_k, \frac{d^2}{dx^2} \sum_{p=0}^N \Phi_{pl} T_p \right\rangle \right) = 0.$$

Simplifying this equation leads to

$$\sum_{l=0}^{N-2} \left(\frac{d^2}{dt^2} \hat{v}_l(t) \sum_{k=0}^N \sum_{p=0}^N \Phi_{kq} \Phi_{pl} \langle T_k, T_p \rangle - \hat{v}_l(t) \sum_{k=0}^N \sum_{p=0}^N \Phi_{kq} \Phi_{pl} \left\langle T_k, \frac{d^2}{dx^2} T_p \right\rangle \right) = 0. \quad (3.125)$$

Using (3.59) we have,

$$\frac{d}{dx} u_{sd}(x, t) = \frac{d}{dx} \sum_{k=0}^N \hat{u}_k(t) T_k(x) \equiv \sum_{k=0}^N \hat{u}_k^{(1)}(t) T_k(x), \quad (3.126)$$

where from (3.35) and (3.43) one can use the equivalent matrix formulation

$$\hat{u}_k^{(1)}(t) = \frac{2}{c_k} \sum_{\substack{m=k+1 \\ m+k \text{ odd}}}^N m \hat{u}_m(t), \quad \hat{u}_k^{(2)}(t) = \frac{1}{c_k} \sum_{\substack{m=k+1 \\ m+k \text{ even}}}^N m(m^2 - k^2) \hat{u}_m(t), \quad k = 0, \dots, N, \quad (3.127)$$

where

$$c_k = \begin{cases} 2 & \text{if } k = 0, N, \\ 1 & \text{if } 1 \leq k \leq N - 1. \end{cases} \quad (3.128)$$

Equation (3.127) can be rewritten as

$$\hat{u}_k^{(1)}(t) = \sum_{l=0}^N D_{kl}^{(1)} \hat{u}_l(t), \quad \hat{u}_k^{(2)}(t) = \sum_{l=0}^N D_{kl}^{(2)} \hat{u}_l(t), \quad (3.129)$$

for example, for $N = 4$ we have,

$$D^{(1)} = \begin{pmatrix} 0 & 1 & 0 & 3 & 0 \\ 0 & 0 & 4 & 0 & 8 \\ 0 & 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad D^{(2)} = \begin{pmatrix} 0 & 0 & 4 & 0 & 32 \\ 0 & 0 & 0 & 24 & 0 \\ 0 & 0 & 0 & 0 & 48 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Substituting (3.129) in (3.126) gives

$$\sum_{k=0}^N \hat{u}_k(t) \frac{d}{dx} T_k(x) = \sum_{k=0}^N \sum_{l=0}^N D_{kl}^{(1)} \hat{u}_l(t) T_k(x) = \sum_{l=0}^N \hat{u}_l(t) \sum_{k=0}^N D_{kl}^{(1)} T_k(x)$$

$$\equiv \sum_{k=0}^N \hat{u}_k(t) \sum_{l=0}^N D_{lk}^{(1)} T_l(x).$$

This leads to

$$\frac{d}{dx} T_k(x) = \sum_{l=0}^N D_{lk}^{(1)} T_l(x),$$

and

$$\frac{d^2}{dx^2} T_k(x) = \sum_{l=0}^N D_{lk}^{(2)} T_l(x).$$

From this equation we get

$$\left\langle T_k, \frac{d^2}{dx^2} T_p \right\rangle = \left\langle T_k, \sum_{r=0}^N D_{rp}^{(2)} T_r \right\rangle = \sum_{r=0}^N D_{rp}^{(2)} \langle T_k, T_r \rangle.$$

This equation with equations (3.8) and (3.9) lead the equation (3.125) to become

$$\sum_{l=0}^{N-2} \left(\frac{d^2}{dt^2} \hat{v}_l(t) \sum_{k=0}^N \Phi_{kq} \Phi_{kl} \gamma_k - \hat{v}_l(t) \sum_{k=0}^N \sum_{p=0}^N \Phi_{kq} \Phi_{pl} D_{kp}^{(2)} \gamma_k \right) = 0, \text{ for all } q = 0, \dots, N-2. \quad (3.130)$$

Define the matrices A , B and Θ by:

$$A_{ql} = \sum_{k=0}^N \Phi_{kq} \Phi_{kl} \gamma_k, \quad B_{ql} = \sum_{k=0}^N \sum_{p=0}^N \Phi_{kq} \Phi_{pl} D_{kp}^{(2)} \gamma_k,$$

and

$$\Theta_{kq} = (1 + \delta_{k0}) \Phi_{kq}.$$

For $N = 4$, we have

$$\Theta = \begin{pmatrix} -2 & 0 & -2 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

We have $A = \Theta^T \Phi$ and $B = \Theta^T D^{(2)} \Phi$.

Hence equation (3.130) becomes

$$\sum_{l=0}^{N-2} \frac{d^2}{dt^2} \hat{v}_l(t) A_{ql} = \sum_{l=0}^{N-2} B_{ql} \hat{v}_l(t),$$

which we can write as

$$A \frac{d^2}{dt^2} \vec{v}(t) = B \vec{v}(t),$$

or even more simply as

$$\frac{d^2}{dt^2} \vec{v}(t) = C \vec{v}(t), \quad (3.131)$$

where $C = A^{-1}B$.

The ODEs (3.131) can be written as

$$\frac{d}{dt} \vec{y}(t) = \vec{w}(t), \quad (3.132)$$

$$\frac{d}{dt} \vec{w}(t) = C \vec{y}(t). \quad (3.133)$$

3.9.2 Time stepping

By applying Euler's method to (3.132) and (3.133) we get

$$\vec{y}(t+h) = \vec{y}(t) + h\vec{w}(t), \quad (3.134)$$

$$\vec{w}(t+h) = \vec{w}(t) + hC\vec{y}(t). \quad (3.135)$$

To find $\vec{v}(0)$, we use (3.124) as follows

$$\hat{u}_k(t) = \sum_{l=0}^{N-2} \Phi_{kl} \hat{v}_l(t), \quad k = 0, \dots, N.$$

This can be written as

$$\vec{\hat{u}}(0) = \Phi \vec{\hat{v}}(0).$$

To simplify this equation we multiply both sides of the equation from the left with the transpose of the matrix Φ :

$$\Phi^T \vec{\hat{u}}(0) = \Phi^T \Phi \vec{\hat{v}}(0).$$

Then, multiply with the inverse of $\Phi^T \Phi$

$$(\Phi^T \Phi)^{-1} \Phi^T \vec{\hat{u}}(0) = \vec{\hat{v}}(0), \quad (3.136)$$

and thus

$$\vec{y}(0) = (\Phi^T \Phi)^{-1} \Phi^T \vec{\hat{u}}(0).$$

From the second initial condition we have,

$$\vec{w}(0) = 0.$$

The value of $\vec{u}(0)$ can be found using (3.115).

The second method used is the trapezoidal rule, as follows

$$\vec{y}(t+h) = \vec{y}(t) + \frac{h}{2}\{\vec{w}(t) + \vec{w}(t+h)\}, \quad (3.137)$$

$$\vec{w}(t+h) = \vec{w}(t) + \frac{h}{2}\{C\vec{y}(t) + C\vec{y}(t+h)\}. \quad (3.138)$$

3.10 Numerical results

3.10.1 Convergence of Euler's and the trapezoidal methods

Similar figures to the pseudospectral method for solving the wave equation (3.52), of the convergence in space and time, were obtained here as well, which show that for the C^∞ initial condition (3.81), both Euler's and the trapezoidal methods combined with the Galerkin method converge exponentially in N . Also, for the IC (3.81) Euler's method is $\mathcal{O}(h)$ and the trapezoidal method is $\mathcal{O}(h^2)$. The figures look almost identical.

Overall, only the C^∞ IC (3.81) is considered when using the Galerkin method; the other two ICs are not studied.

3.10.2 Stability of Euler's, the implicit Euler and the trapezoidal methods

From some figures, not shown, it seems that the stability properties for the Galerkin method are the same results to the pseudospectral method. This when the method is used in solving the wave equation (3.52) with the time stepping methods, Euler's, the implicit Euler and the trapezoidal methods for the C^∞ IC (3.81). That is, Euler's method is not stable, while the other two methods are stable; the numerical solution using the implicit Euler method decreases and approaches zero as time $\rightarrow \infty$. And for the trapezoidal method, the numerical solution is bounded and vibrates between -1 and 1 . The figures look almost identical.

3.11 Summary and discussion

In this chapter a linear partial differential equation, namely the wave equation (3.52), is solved using the spectral methods, with Chebyshev polynomials, in physical space (the pseudospectral method) and spectral space (the collocation, tau and Galerkin methods).

Solving the linear wave equation goes through two main steps: discretizing this PDE to ODEs, then solving those ODEs using time stepping methods. Therefore, there are two main types of errors that occur and affect on our calculations, which are the error of the spectral method and the error of the time stepping methods (Euler's, the modified Euler, the implicit Euler and the trapezoidal method).

This chapter studies the convergence, in space and time, and stability. Overall, one initial condition, the C^∞ IC (3.81), is implemented for solving this linear equation throughout this chapter for all methods above. Only some minor differences are obtained in the figures when using the spectral methods in physical space (the pseudospectral method) than when using them in spectral space (the collocation, tau and Galerkin methods). The error converges exponentially with respect to N for all the time stepping methods addressed. However, for the pseudospectral method combined with the trapezoidal method two more ICs, the C^1 IC (3.82) and the C^3 IC (3.83), are employed. These ICs are carefully chosen to generate solutions with different levels of smoothness. This leads to spectral convergence with respect to N for the C^∞ IC, (3.81), but for the other ICs, (3.82) and (3.83), the solution converges algebraically, with order seems to be $\frac{1}{N^3}$, for (3.82), and order $\frac{1}{N^5}$, for (3.83). Furthermore, the convergence is thoroughly investigated by using the spectrum of the numerical solution of the combination the trapezium method with pseudospectral method. The Chebyshev coefficients of the numerical solution \hat{u}_k are plotted against the mode k , the even coefficients component and the odd coefficients component, for all three ICs. Similarly, the Chebyshev coefficients of the exact d'Alembert solution. However, we have not studied the spectra of the numerical solution using other time stepping methods, when using the pseudospectral method, also, we have not investigated this for other spectral methods (the collocation, tau and Galerkin methods).

From Section 1.9 we know that every function can be uniquely decomposed as the sum of an even and an odd function. In a Chebyshev series, the terms with the

even coefficients sum to be the the even component and the terms with the odd coefficients sum to be the the odd component. As explained in Section 1.9 that the C^∞ IC (1.54) is purely an odd function. However, the other two ICs, the IC (3.83) and the IC (3.82), have both the odd and even components. The C^∞ (3.81) yields an infinitely smooth solution, that is also odd. We see that, the odd coefficients of the numerical solution decay exponentially, but the even spectral coefficients are basically just round-off. The C^3 IC (3.83) introduces nonsmoothness at the fourth derivative. The nonsmoothness affects both the odd and even component of the solution as they are nonsmooth. Thus, the spectral coefficients of the generated numerical solution do not decay exponentially but only algebraically with order $\frac{1}{k^5}$ at general t . The C^1 IC (3.82) introduces nonsmoothness at the second derivative. However, in this case only the even component is not smooth, while the odd component is smooth. We see that, the odd spectral coefficients of the generated numerical solution decay exponentially, while the even spectral coefficients decay algebraically with order $\frac{1}{k^3}$ at general t .

Section 1.9 shows that the exact d'Alembert solution of the linear problem is infinitely smooth when t is an even integer, therefore, the exact solution spectra decay exponentially for both components. While, the exact d'Alembert solution is an odd function when t is an odd integer, thus, the even coefficients are zero. The Chebyshev coefficients of the numerical solution, of the trapezoidal method combined with pseudospectral method, attempt to mirror the behaviour of the exact solution (i.e., recover smoothness of the solution or the difference of the odd/even of the solution). However, it is not able to reflect it faithfully but only partially to a limited extent, especially for the C^1 IC (3.82) up to the time stepping error $\mathcal{O}(h^2)$ only. For this IC, (3.82), the even numerical Chebyshev coefficients at even integer times decay exponentially as they should, but only up to $\mathcal{O}(h^2)$, after that they decay algebraically. The even coefficients at odd integer times should vanish but again they only do this up to $\mathcal{O}(h^2)$, after that they decay algebraically.

Furthermore, there is a fast drop, transition for large k , in the plots of the even coefficients of the numerical solution of the C^1 IC (3.82) only but not of the C^3 IC (3.83). One concludes that the trapezoidal method deals better with the C^3 IC (3.83) than the C^1 IC (3.82) in simulating the behaviour of the numerical solution spectra, perhaps because the solution generated by the C^1 IC (3.82) is the less smooth than the solution generated by the C^3 IC (3.83).

Moving now to the convergence with respect to h , the global error of the Euler's method is $\mathcal{O}(h)$ and the modified Euler method is $\mathcal{O}(h^2)$, similarly the order of the implicit Euler method is $\mathcal{O}(h)$ and of the trapezoidal method is $\mathcal{O}(h^2)$. Nevertheless, our results of the pseudospectral method combined with the trapezoidal method for the C^1 IC (3.82) and the C^3 IC (3.83) are different from our expectation in Chapter 1 as we assumed that the discontinuity will impact on the time stepping error as well as the spatial error.

Turning now to discuss in more detail the results of the wave equation (3.52), in terms of stability. Euler's method is not stable when solving the wave equation (3.52), for all spatial discretization schemes (the pseudospectral, tau and Galerkin methods). In addition to Euler's method, the modified Euler method is also unstable for solving this wave equation using the pseudospectral method; the error is growing rapidly as time is increasing. The modified Euler method is not used in this Chapter for the tau and Galerkin methods.

The stability of the implicit Euler and the trapezoidal methods when solving the wave equation (3.52) using the pseudospectral method for two ICs, the C^∞ IC (3.81) and the C^1 IC (3.82), is studied in detail. It seems that, both of these methods are stable, in other words, the error does not grow without bounds as time increases but instead is bounded. Nevertheless, the trapezoidal method is better than the implicit Euler method because the numerical solution, with the C^∞ IC (3.81), for the implicit Euler method is not behaving like the exact solution: the numerical solution decays to zero, while the solution of the trapezoidal method is behaving like the exact solution as the overall amplitude remains a constant; it does not decay to zero. However, we found that the tau method is unstable when solving the wave equation for all the three time stepping methods applied (Euler's, the implicit Euler and the trapezoidal methods). The error grows rapidly for all the time stepping methods implemented. Finally coming to stability of the collocation and Galerkin method, the results discussed for stability of the pseudospectral method when solving this wave equation can be applied for these two methods too because they have almost similar stability properties results. Noteworthy, we have not investigated on the energy for any method in this chapter.

Many spectral methods are used to solve the linear wave equation throughout this chapter. We concluded that the pseudospectral method is more practical for our problems. This leads us to choose the pseudospectral method as the most practical

method in the next chapter, when we will numerically solve a nonlinear partial differential equation whose solution we do not know, namely the nonlinear wave equation (1.24).

Chapter 4

Nonlinear wave equation

We continue in this chapter the study of spectral methods, especially the pseudospectral method using Chebyshev polynomials. However, here we will solve a nonlinear wave equation subject to non-periodic boundary conditions and this equation will be studied in the following chapters as well with different methods. The nonlinear wave equation we are solving is (1.24), which is as follows

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3, \quad (4.1)$$

for $x \in (-1, 1)$ and $t \geq 0$. This equation is solved subject to Dirichlet boundary conditions

$$u(-1, t) = 0, \quad \text{and} \quad u(1, t) = 0, \quad (4.2)$$

and initial conditions

$$u(x, 0) = f(x), \quad (4.3)$$

and

$$\frac{\partial u}{\partial t}(x, 0) = 0. \quad (4.4)$$

The equation and its background are discussed in Section 1.4. Three different initial conditions presented in Chapter 1, which are (1.54), (1.55) and (1.56), are implemented in solving this equation.

This chapter starts by taking the spatial discretization of this PDE, Subsection 4.2.1, and then using time stepping methods, used in previous chapter, to find the solution of the resulting ODEs, in Subsection 4.2.2. Furthermore, in Section 4.3

the convergence and stability of each time stepping method is studied. In addition, more investigation on numerical schemes is done by studying the symplecticity of them in Section 4.4, which the Hamiltonian structure of this PDE analytically is given in Subsection 4.1.

4.1 Energy preservation for the exact solution

In this section we study the Hamiltonian formulation of the PDE (4.1). The introduction of Hamiltonian for ODEs is given in Section 2.4.

In order to study the Hamiltonian of this equation, we start by considering a general semi-linear wave equation of the form

$$u_{tt} = u_{xx} - f(u), \quad (4.5)$$

where $f(u)$ and $u(x, t)$ are smooth functions. (Leimkuhler & Reich, 2004) proved that this wave equation is Hamiltonian and conservative under specific prescriptions of boundary and initial data. For example, they imposed periodic boundary conditions with period L . The general energy functional is given as

$$E(u) = \int_0^L \left[\frac{1}{2}u_t^2 + \frac{1}{2}u_x^2 + F(u) \right] dx, \quad (4.6)$$

where $F(u) = \int_0^u f(s)ds$.

We can now implement this proof to our equation (4.1) in which the interval is $[-1, 1]$ and $f(u) = u^3$, this implies that $F(u) = \frac{1}{4}u^4$. Hence, (4.6) becomes

$$E(u) = \int_{-1}^1 \left[\frac{1}{2}u_t^2 + \frac{1}{2}u_x^2 + \frac{1}{4}u^4 \right] dx. \quad (4.7)$$

Taking the derivative of (4.7) and using integrating by parts yield

$$\frac{d}{dt}E(u) = \int_{-1}^1 \frac{d}{dt} \left[\frac{1}{2}u_t^2 + \frac{1}{2}u_x^2 + \frac{1}{4}u^4 \right] dx = \int_{-1}^1 [u_t u_{tt} - u_{xx} u_t + u^3 u_t] dx + [u_x u_t]_{-1}^1.$$

Now substituting the PDE (4.1) yields

$$\frac{d}{dt}E(u) = [u_x u_t]_{-1}^1. \quad (4.8)$$

The Dirichlet boundary conditions, (4.2), are

$$u(\pm 1, t) = 0,$$

so,

$$\frac{du}{dt}(\pm 1, t) = 0,$$

and thus (4.8) implies

$$\frac{d}{dt}E(u) = [u_x u_t]_{-1}^1 = 0. \quad (4.9)$$

This shows that E is conserved. Furthermore, the second order equation (4.1) can be equivalently written as a first order system of PDEs,

$$u_t = v, \quad (4.10)$$

$$v_t = u_{xx} - u^3. \quad (4.11)$$

The nonlinear equation (4.1) can be viewed as a Hamiltonian PDE with $Z = (u, v)^T$ and Hamiltonian functional,

$$H[u, v] = \int_{-1}^1 \left[\frac{1}{2}v^2 + \frac{1}{2}u_x^2 + \frac{1}{4}u^4 \right] dx. \quad (4.12)$$

The Hamiltonian H is the same as E , except that H is function of u and v while E is function of u only.

In order to prove that the PDE is indeed Hamiltonian, we need to show that

$$\frac{\delta H}{\delta v} = u_t, \quad (4.13)$$

$$\frac{\delta H}{\delta u} = -v_t, \quad (4.14)$$

see Section 2.4. We need to replace the partial derivatives in Section 2.4 by functional derivatives, which are defined by

$$H[u, v + \varepsilon \tilde{v}] = H[u, v] + \varepsilon \left\langle \tilde{v}, \frac{\delta H}{\delta v} \right\rangle + \mathcal{O}(\varepsilon^2), \quad (4.15)$$

and

$$H[u + \varepsilon \tilde{u}, v] = H[u, v] + \varepsilon \left\langle \tilde{u}, \frac{\delta H}{\delta u} \right\rangle + \mathcal{O}(\varepsilon^2). \quad (4.16)$$

The inner product definition is

$$\left\langle \tilde{u}, \frac{\delta H}{\delta u} \right\rangle = \int_{-1}^1 \frac{\delta H}{\delta u} \tilde{u} dx. \quad (4.17)$$

Using (4.12),

$$\begin{aligned} H[u + \varepsilon\tilde{u}, v] &= \int_{-1}^1 \left[\frac{1}{2}v^2 + \frac{1}{2} \frac{\partial}{\partial x} (u + \varepsilon\tilde{u})^2 + \frac{(u + \varepsilon\tilde{u})^4}{4} \right] dx \\ &= \int_{-1}^1 \left[\frac{1}{2}v^2 + \frac{1}{2}u_x^2 + \varepsilon u_x \tilde{u}_x + \frac{1}{2}\varepsilon^2 \tilde{u}_x^2 + \frac{(u + \varepsilon\tilde{u})^4}{4} \right] dx, \end{aligned}$$

this leads to,

$$\begin{aligned} H[u + \varepsilon\tilde{u}, v] &= \int_{-1}^1 \left[\frac{1}{2}v^2 + \frac{1}{2}u_x^2 + \varepsilon u_x \tilde{u}_x + \frac{1}{2}\varepsilon^2 \tilde{u}_x^2 + \frac{u^4}{4} + \varepsilon\tilde{u}u^3 + \frac{6}{4}\varepsilon^2 \tilde{u}^2 u^2 + \varepsilon^3 \tilde{u}^3 u + \frac{\varepsilon^4 \tilde{u}^4}{4} \right] dx \\ &= \int_{-1}^1 \left[\frac{1}{2}v^2 + \frac{1}{2}u_x^2 + \frac{u^4}{4} + \varepsilon u_x \tilde{u}_x + \varepsilon\tilde{u}u^3 + \varepsilon^2 \left(\frac{1}{2}\tilde{u}_x^2 + \frac{6}{4}\tilde{u}^2 u^2 + \varepsilon\tilde{u}^3 u + \frac{\varepsilon^2 \tilde{u}^4}{4} \right) \right] dx \\ &= \int_{-1}^1 \left[\frac{1}{2}v^2 + \frac{1}{2}u_x^2 + \frac{u^4}{4} \right] dx + \int_{-1}^1 \varepsilon u_x \tilde{u}_x dx + \int_{-1}^1 \varepsilon\tilde{u}u^3 dx + \mathcal{O}(\varepsilon^2) \\ &= H[u, v] + \int_{-1}^1 \varepsilon u_x \tilde{u}_x dx + \int_{-1}^1 \varepsilon u^3 \tilde{u} dx + \mathcal{O}(\varepsilon^2). \end{aligned} \quad (4.18)$$

Integration by part must be used on the term $\int_{-1}^1 \varepsilon u_x \tilde{u}_x dx$ in order to find $\frac{\delta H}{\delta u}$ because the equation has \tilde{u}_x instead of \tilde{u} . Hence, (4.18) becomes,

$$H[u + \varepsilon\tilde{u}, v] = H[u, v] + \varepsilon \left([u_x \tilde{u}]_{-1}^1 - \int_{-1}^1 u_{xx} \tilde{u} dx \right) + \varepsilon \int_{-1}^1 u^3 \tilde{u} dx + \mathcal{O}(\varepsilon^2). \quad (4.19)$$

Using the initial conditions (4.2) we get $u_x(-1, t) = u_x(1, t) = 0$, this is what we need, and by using the inner product definition (4.17), Eq. (4.19) becomes

$$\begin{aligned} H[u + \varepsilon\tilde{u}, v] &= H[u, v] - \varepsilon \int_{-1}^1 u_{xx} \tilde{u} dx + \varepsilon \int_{-1}^1 u^3 \tilde{u} dx + \mathcal{O}(\varepsilon^2) \\ &= H[u, v] + \int_{-1}^1 (-u_{xx} + u^3) \varepsilon \tilde{u} dx + \mathcal{O}(\varepsilon^2) \\ &= H[u, v] + \varepsilon \langle \tilde{u}, -u_{xx} + u^3 \rangle + \mathcal{O}(\varepsilon^2). \end{aligned}$$

So according to the functional derivative (4.16), we get

$$\frac{\delta H}{\delta u} = -u_{xx} + u^3 = -v_t. \quad (4.20)$$

Now to find $\frac{\delta H}{\delta v}$, (4.12) is used again,

$$\begin{aligned} H[u, v + \varepsilon\tilde{v}] &= \int_{-1}^1 \left[\frac{1}{2}(v + \varepsilon\tilde{v})^2 + \frac{1}{2}u_x^2 + \frac{1}{4}u^4 \right] dx \\ &= \int_{-1}^1 \left[\frac{1}{2}v^2 + \frac{1}{2}u_x^2 + \frac{1}{4}u^4 \right] dx + \int_{-1}^1 \varepsilon v \tilde{v} dx + \frac{1}{2} \int_{-1}^1 \varepsilon^2 \tilde{v}^2 dx \\ &= H[u, v] + \varepsilon \langle \tilde{v}, v \rangle + \mathcal{O}(\varepsilon^2), \end{aligned}$$

comparing the last equation with (4.16), we get

$$\frac{\delta H}{\delta v} = v. \quad (4.21)$$

Combining (4.20) and (4.21) with the first order equations (4.10) and (4.11), we see that (4.13) and (4.14) are satisfied. Therefore, one can say that our PDE (4.1) is Hamiltonian. The consequence is that

$$\frac{\partial H}{\partial t} = 0,$$

meaning that this wave equation (4.1) is Hamiltonian, in other words, the energy is conserved. This was already shown in (4.9).

This section is using the ideas presented in the following references and applying them on our equation (4.1) (Bridges & Reich, 2006; Craig, 2008; Hairer *et al.*, 2002; Leimkuhler & Reich, 2004; Sanz-Serna & Calvo, 1994)

4.2 The pseudospectral method

As mentioned in the chapter before, the pseudospectral method cancels the residual on a set of grid points called the collocation points (the collocation points are usually the same as the quadrature points). The basis functions are taken to be the same as the trial functions, i.e., $T_k(x)$ for $k = 0, \dots, N$.

There are many time stepping methods which can be used to find the solution of the nonlinear wave equation (4.1) using the pseudospectral method. In total, we use four methods, which are Euler's method, the modified Euler, the implicit Euler and trapezoidal method. However, only for Euler's method, the modified Euler method and the trapezoidal method the calculations are given in more detail because the calculations for the implicit Euler method are similar as for the trapezoidal method.

4.2.1 Spatial discretization

The chosen trial functions are the Chebyshev polynomial of first kind, which are given in (3.4),

$$T_k(x) = \cos(k \cos^{-1}(x)) = \cos(k\theta), \quad k = 0, 1, \dots, N, \quad (4.22)$$

where $x = \cos(\theta)$, $-1 \leq x \leq 1$ and $0 \leq \theta \leq \pi$.

From (3.27) the semi-discretization solution has the representation,

$$u_{sd}(x, t) = \sum_{k=0}^N \hat{u}_k(t) T_k(x), \quad (4.23)$$

where from (3.26)

$$\hat{u}_k(t) = \frac{2}{Nc_k} \sum_{j=0}^N \frac{1}{c_j} u_{sd}(x_j, t) T_k(x_j), \quad c_j = 1 + \delta_{j0} + \delta_{jN}, k = 0, \dots, N. \quad (4.24)$$

The pseudospectral method solves the PDE by requiring that the residual on a set of grid points is zero:

$$\frac{\partial^2 u_{sd}}{\partial t^2} - \frac{\partial^2 u_{sd}}{\partial x^2} + u_{sd}^3 \Big|_{x=x_j} = 0, \quad j = 1, \dots, N-1. \quad (4.25)$$

The boundary conditions

$$u_{sd}(-1, t) = 0, \quad u_{sd}(1, t) = 0 \quad (4.26)$$

and the initial conditions

$$u_{sd}(x_k, 0) = f(x_k), \quad k = 0, \dots, N \quad (4.27)$$

and

$$\frac{\partial u_{sd}}{\partial t}(x_k, 0) = 0, \quad k = 0, \dots, N \quad (4.28)$$

accompany (4.25).

The coefficients $\hat{u}_k^{(2)}$, (3.60),

$$\hat{u}_k(t) = \frac{2}{Nc_k} \sum_{j=0}^N \frac{1}{c_j} u_{sd}(x_j, t) T_l(x_j) \quad c_j = 1 + \delta_{j0} + \delta_{jN}, k = 0, \dots, N$$

depend linearly on the nodal values $u_{sd}(x_l, t)$, hence they satisfy (3.48), where the matrix $D^{(2)}$ is given in (3.49). This leads,

$$\frac{\partial^2 u_j(t)}{\partial x^2} = \sum_{k=0}^N \hat{u}_k^{(2)}(t) T_k(x_j) = \sum_{l=0}^N D_{jl}^{(2)} u_l(t), \quad j = 1, \dots, N-1, \quad (4.29)$$

see (3.67)-(3.69), where $u_j(t) \approx u(x_j, t)$ and at $t = 0$ they are equal, $u_j(0) = u(x_j, 0) = f(x_j)$. From (4.26) we have $u_0(0) = u_N(0) = 0$. Substituting (4.29) into (4.25), we get

$$\frac{\partial^2 u_j(t)}{\partial t^2} = \sum_{l=0}^N D_{jl}^{(2)} u_l(t) - u_j^3(t), \quad j = 1, \dots, N-1. \quad (4.30)$$

The ODEs (4.30) can be written as

$$\frac{\partial y_j(t)}{\partial t} = w_j(t), \quad (4.31)$$

$$\frac{\partial w_j(t)}{\partial t} = \sum_{l=0}^N D_{jl}^{(2)} y_l(t) - y_j^3(t), \quad j = 1, \dots, N-1. \quad (4.32)$$

4.2.2 Time stepping

Euler's method is used to solve (4.31) and (4.32) and this implies

$$y_j(t+h) = y_j(t) + hw_j(t), \quad (4.33)$$

$$w_j(t+h) = w_j(t) + h \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t) - hy_j^3(t), \quad j = 1, \dots, N-1, \quad (4.34)$$

where from the boundary conditions, (4.2), we have $y_0(0) = y_N(0) = 0$ and $w_0(0) = w_N(0) = 0$, and from the initial conditions, (4.3) and (4.4), we have,

$$y_j(0) = f(x_j) \quad \text{and} \quad w_j(0) = 0, \quad j = 1, \dots, N-1,$$

where x_j are the Chebyshev points in (3.75).

The modified Euler method is also used to solve (4.31) and (4.32) and this yields

$$y_j(t+h) = y_j(t) + hw_j(t) + \frac{h^2}{2} \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t) - \frac{h^2}{2} y_j^3(t), \quad (4.35)$$

$$\begin{aligned} w_j(t+h) = & w_j(t) + h \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t) + \frac{h^2}{2} \sum_{l=1}^{N-1} D_{jl}^{(2)} w_l(t) \\ & - \frac{h}{2} y_j^3(t) - \frac{h}{2} (y_j(t) + hw_j(t))^3, \quad j = 1, \dots, N-1. \end{aligned} \quad (4.36)$$

In addition to the previous methods, the trapezoidal method is used to solve (4.31) and (4.32) and calculations are given in detail next in Subsection 4.2.3. Applying the trapezoidal method on (4.31) and (4.32) gives,

$$y_j(t+h) = y_j(t) + \frac{h}{2}\{w_j(t) + w_j(t+h)\}, \quad (4.37)$$

$$w_j(t+h) = w_j(t) + \frac{h}{2}\left\{\sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t) + \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l(t+h), -y_j^3(t+h) - y_j^3(t)\right\}, \quad j = 1, \dots, N-1. \quad (4.38)$$

One can rewrite (4.37) and (4.38) as,

$$y_j^{n+1} = y_j^n + \frac{h}{2}\{w_j^{n+1} + w_j^n\} \quad (4.39)$$

$$w_j^{n+1} = w_j^n + \frac{h}{2}\left\{\sum_{l=1}^{N-1} D_{jl}^{(2)} y_l^{n+1} + \sum_{l=1}^{N-1} D_{jl}^{(2)} y_l^n - (y_j^{n+1})^3 - (y_j^n)^3\right\}. \quad (4.40)$$

4.2.3 Solving the nonlinear equation using Newton's method

The trapezoidal rule leads to a system of nonlinear equations, (4.39) and (4.40), which needs to be solved for y_j^{n+1} and w_j^{n+1} . To achieve this, Newton's method, which is given in Section 1.7.3, is used. It can be written as

$$g_{j+1} = g_j - J^{-1}(g_j)F_n(g_j), \quad (4.41)$$

where j is the iteration index of the Newton's method, g_j is taken to be the solution of the previous step and g_0 can be found from the initial condition, i.e., $g_0 = \{y_j^0, w_j^0\}$, for $j = 1, \dots, N-1$. The function F in Newton's method is

$$F(g_{n+1}) = \begin{pmatrix} H_j(g_{n+1}) \\ G_j(g_{n+1}) \end{pmatrix}_{2(N-1) \times 1}, \quad j = 1, \dots, N-1,$$

where from (4.39) and (4.40),

$$H_j(g_{n+1}) = y_j^{n+1} - y_j^n - \frac{h}{2}w_j^{n+1} - \frac{h}{2}w_j^n, \quad (4.42)$$

and

$$G_j(g_{n+1}) = w_j^{n+1} - w_j^n - \frac{h}{2}\sum_{l=1}^{N-1} D_{jl}^{(2)} y_l^{n+1} - \frac{h}{2}\sum_{l=1}^{N-1} D_{jl}^{(2)} y_l^n + \frac{h}{2}(y_j^{n+1})^3 + \frac{h}{2}(y_j^n)^3, \quad \text{for } j = 1, \dots, N-1, \quad (4.43)$$

and from the boundary conditions we have $y_0^{n+1} = 0$ and $w_0^{n+1} = 0$.

The Jacobian matrix J is given in Section 1.7.3, which is

$$J(g_{n+1}) = \begin{pmatrix} \frac{\partial H}{\partial y^{n+1}} & \frac{\partial H}{\partial w^{n+1}} \\ \frac{\partial G}{\partial y^{n+1}} & \frac{\partial G}{\partial w^{n+1}} \end{pmatrix}_{2(N-1) \times 2(N-1)},$$

which implies,

$$J(g_{n+1}) = \begin{pmatrix} I & -\frac{h}{2}I \\ A & I \end{pmatrix}_{2(N-1) \times 2(N-1)},$$

where $A = \frac{3h}{2}\tilde{A} - \frac{h}{2}D^{(2)}$ with \tilde{A} being the diagonal matrix with $(y_j^{n+1})^2$ for $j = 1, \dots, N-1$ on the diagonal.

4.3 Numerical results

The convergence and stability of each method are discussed in detail, using figures, in the following. The particular initial conditions used in this chapter are the three ICs used in Subsection 3.4. The ICs are,

$$f(x) = \sin \pi x,$$

$$f(x) = e^{\sin(\pi x)} - 1,$$

and

$$f(x) = \frac{1}{11}e^x(11 - 8x - 10x^2 + 8x^3 - x^4),$$

see (3.81), (3.82) and (3.83), which are given first in Section 1.9 see (1.54), (1.55) and (1.56).

4.3.1 Convergence of Euler's method

Figures in this chapter are using two solutions, the reference solutions (explained below) and the numerical solutions, of the nonlinear wave equation (4.1) to find

the estimated error. Those Figures are the infinite norm of the estimated error. Noteworthy, a sufficient number of equally spaced points, denoted by nx , are taken on the interval $(-1, 1)$ in order to complete finding the numerical solution. The Matlab command `linspace(-1, 1, nx)` generates these points, where $nx > N$. Then by substituting these u_j , we obtained from the time stepping methods, into (4.24) and using (4.23) we obtain the numerical solution,

$$u_{num}(x, t) = \sum_{k=0}^N \hat{u}_k(t) T_k(x), \quad (4.44)$$

where x are the equally spaced points mentioned above. They are different from the Chebyshev points x_j in (3.75), they must be more than the Chebyshev points.

The reference solutions are found using the trapezoidal method for solving the nonlinear wave equation (4.1) in this chapter for specific N 's and h 's, according to each of the three ICs used. For example, they are found for the ICs (3.81) and (3.83) at modes $N = 100$, $t = 0.9$, $nx = 100N$ but at $h = 10^{-8}$, and for the second initial condition, (3.82), at modes $N = 1000$, $t = 0.9$, $nx = 100N$ but at $h = 10^{-6}$.

An important point to mention that there is a relation between h and N in finding the convergence and stability, as mentioned before. Hence, the process of choosing h and N considers all these points for all the time stepping methods employed, especially because the reference solutions are using some certain values of N and h so this is leading us to be more careful in choosing the values of N and h in our numerical experiments. For instance, for large h the methods may not appear to behave as expected for the first and second order methods. However, the large h can not be depended on to know the exact order. Therefore, the order is studied for smaller h 's.

Figure 4.1 and the left panel of Figure 4.2 show that Euler's method converges when solving the nonlinear wave equation (4.1) using the pseudospectral method for the initial condition (3.81). Figure 4.1 shows that the method converges exponentially in space for this IC. However, for the other two ICs, the method does not converge exponentially but only algebraically, similar as for the trapezoidal rule which will be discussed in the forthcoming Subsection 4.3.3.

Figure 4.2 (left panel) shows that Euler's method for the initial condition (3.81) is first order. The same holds for the other two ICs.

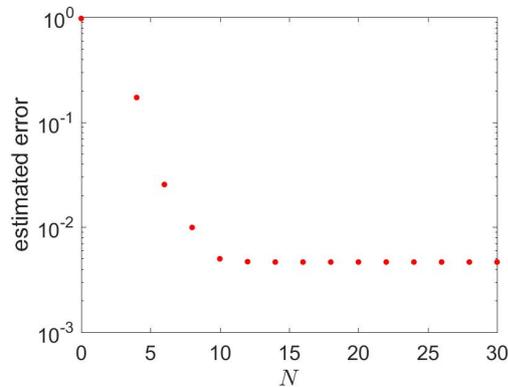


Figure 4.1: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the nonlinear wave equation (4.1), for the initial condition (3.81), $f(x) = \sin(\pi x)$, using Euler's method. The estimated error is evaluated at $t = 0.9$ with time step $h = 10^{-3}$ and $nx = 100N$. Here u_{num} is the numerical solution and u is the reference solution.

4.3.2 Convergence of the modified Euler method

In addition to Euler's method, numerical experiments are conducted for the modified Euler method applied to the nonlinear wave equation (4.1) discretized with the pseudospectral method. They show that the method also converges with respect to N for the three ICs, with the same behaviour as for Euler's and trapezoidal methods.

And it can be noticed in the right panel of Figure 4.2 that this method for the IC (3.81) is second order in h , $\mathcal{O}(h^2)$. The same holds for the other two ICs.

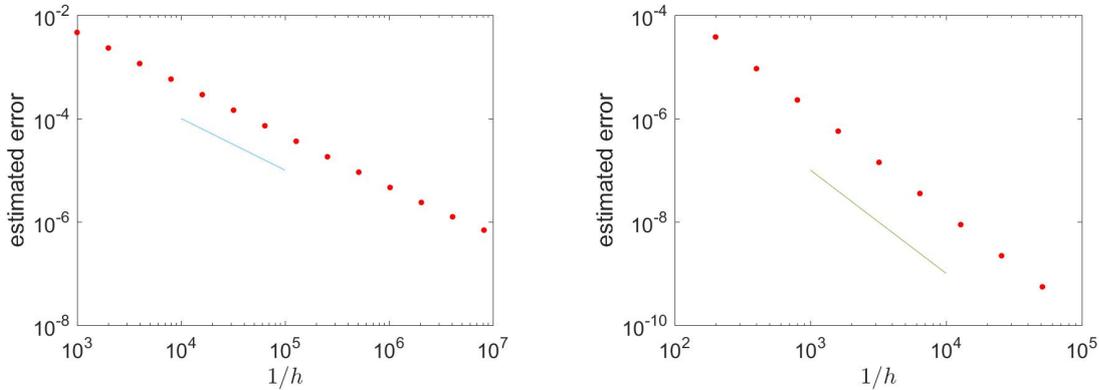


Figure 4.2: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the inverse step size $1/h$ for the nonlinear wave equation (4.1), for the initial condition (3.81) using Euler's method (left panel) and the modified Euler method (right panel). The estimated error is evaluated $t = 0.9$ using $N = 30$. The blue straight line indicates order one for Euler's method and order two for the modified Euler method.

4.3.3 Convergence of the implicit Euler method

The implicit Euler method converges like the explicit Euler method when solving this nonlinear wave equation using the pseudospectral method: we have exponential convergence in N for the first IC (3.81) but algebraic convergence for the other two ICs, and first order convergence in h . The convergence plots for the implicit Euler method look qualitatively the same as for the explicit Euler method; they have the same order of convergence in space and time but with different pre-factor.

4.3.4 Convergence of the trapezoidal method

The trapezoidal method converges with order two in the step size h for the three ICs, (3.81), (3.82) and (3.83), similar to the modified Euler. More interesting is the convergence with respect to N , and this is studied in detail here. This will also prepare us to study the conservation of energy in subsequent sections.

The d'Alembert solution (3.86) for the linear wave equation (3.52) no longer applies here for the nonlinear wave equation (4.1); the analytic solution is not known for

this equation. However, the discontinuities present in the linear wave equation are very likely to be present here as well as both of the equations share the same characteristics, and we believe there is just nothing about the nonlinearity that would make the discontinuities go away. The experiments study the convergence in N and the results are as in the linear wave regime. Figure 4.3 shows that the method with IC (3.81), see panel (a), converges exponentially, while for the other two ICs, (3.82) in panel (b), and (3.83) in panel (c), the method converges algebraically with order seems to be $\frac{1}{N^3}$ and $\frac{1}{N^5}$, respectively. This shows that the method is convergent and confirms our assumption that the discontinuity is present here as well.

If we compare this figure for the nonlinear wave equation with the related figure for the linear wave equation in Chapter 3 (Figure 3.5), we see that the figures look very similar. This gives more credibility to our results for the nonlinear wave equation here.

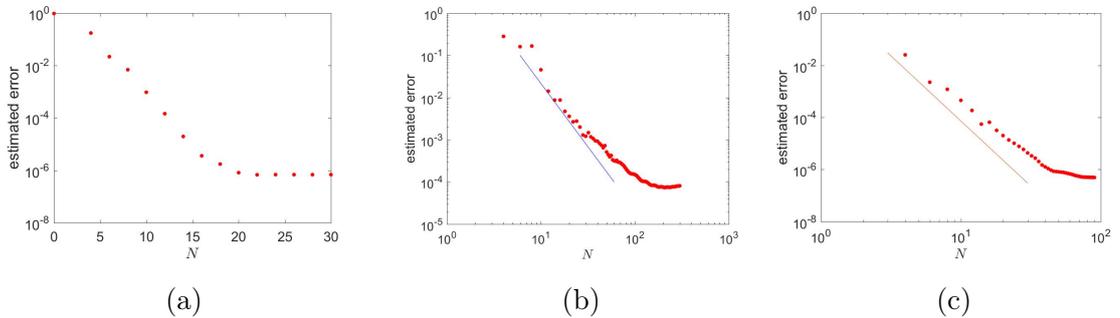


Figure 4.3: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the nonlinear wave equation (4.1) using the pseudospectral method with the trapezoidal method for the C^∞ IC (3.81), C^1 IC (3.82) and C^3 IC (3.83) (a, b and c respectively). The error is evaluated at $t = 0.9$ using $h = 10^{-3}$ and $nx = 100N$. The blue straight line indicates order $\frac{1}{N^3}$ for the IC (3.82) and order $\frac{1}{N^5}$ for the IC (3.83).

Similar as Subsection 3.4.4 we note in Figure 4.3 that the points do not lie on a straight line. However, the plot looks similar to the corresponding plot for the linear wave equation (Figure 3.5). Thus, to gain a better understanding of how the method behaves, we plot the Chebyshev coefficients of the numerical solution \hat{u}_k against the mode k , for all three ICs, the C^∞ IC (3.81), the C^3 IC (3.83) and the

C^1 IC (3.82) in Figures 4.4 and 4.5 and 4.6, respectively. Here we address the first and third ICs first since they are more straightforward as they behave very similar as for the linear equation, we will then discuss the second IC.

We see that for the first IC (3.81) the coefficients decay exponentially, see Figure 4.4, similar to the corresponding plots for the linear wave equation in Subsection 3.4.4 (Figure 3.6), because the solution of the homogeneous nonlinear wave equation preserves the symmetry properties of the initial condition f . If f is odd (resp. even) then the solution remains odd (resp. even), similar to the linear wave equation, as we will explain below. For the third IC (3.83), in Figure 4.5, both the odd and the even coefficients components drop off algebraically with order $\frac{1}{k^5}$, similar to the linear wave equation Figure 3.7, that is more clearly in Figure 4.5a at $t = 0.9$. See Subsection 3.4.4 for a better understanding. More explanation is given later in this subsection. Surprisingly, for the second IC (3.82) both of the spectral coefficients components, the even and odd, decay algebraically at any $t > 0$ with order $\frac{1}{k^3}$, see Figures 4.6, in contrast to the linear wave equation, see Figure 3.8, more details are given later in this subsection. That is for the second IC (3.82) one set (the even set) decays algebraically just like the linear equation, see Subsection 3.4.4 when nonsmoothness affects only the non-zero component (the even component) of the solution for the linear equation so it decays algebraically, whereas the other set (the odd set) changes from decaying exponentially for the linear equation to algebraically for the nonlinear equation.

We now try to explain these observations. From the Chebyshev polynomials (4.22), and before in (3.4), one can see that the polynomials have even and odd parts; it is even when k is even; and odd when k is odd, as mentioned before. In contrast to the linear equation, for the nonlinear wave equation the odd and even components of the initial condition, the Chebyshev polynomials and the solution cannot be evolved independently. Thus, the solution in (4.23) can be written as the sum of its even and odd components,

$$u_{sd} = u_{sd}^o + u_{sd}^e, \quad (4.45)$$

where u_{sd}^o is the odd part and u_{sd}^e is the even part. Therefore, the nonlinear term in (4.25) becomes

$$\begin{aligned} u_{sd}^3 &= (u_{sd}^o + u_{sd}^e)^3 \\ &= (u_{sd}^o)^3 + 3(u_{sd}^o)^2(u_{sd}^e) + 3(u_{sd}^o)(u_{sd}^e)^2 + (u_{sd}^e)^3. \end{aligned}$$

This means the odd component of the cubic term is $(u_{sd}^o)^3 + 3(u_{sd}^o)(u_{sd}^e)^2$ and the even component is $3(u_{sd}^o)^2(u_{sd}^e) + (u_{sd}^e)^3$, which means that each component has the odd and the even parts of the solution; they are mixed. Therefore, the nonlinear PDE mixes the odd and the even parts over time. For the second IC (3.82), only the even part of the solution is nonsmooth initially, but due to the mixing the nonsmoothness leaks to the odd part as shows in the numerical results in Figure 4.6. While for the third IC (3.83) the mixing up is not clear as both the odd and the even component of the IC introduces nonsmoothness, similar to the linear equation, thus both the odd and the even coefficients decay algebraically, behaving no different than for the linear equation in Subsection 3.4.4.

Figure 4.6 shows that as t increases, both the odd and even plots approach one another, especially in Figures 4.6d, 4.6c, 4.6d, 4.6e and 4.6f. Whereas, Figure 4.6a is more like Figure 3.8a for the linear equation, which are at $t = 0.04$; both figures have the odd components dropping off exponentially and the even component increases to decay algebraically, and the even component of both figures are very similar as there is a fast drop in the plot for k between 900 and 1000. That means, for a very small time, $t = 0.04$, the mixing up that we get for the nonlinear equation is still not affecting much and things are the same as for the linear equation, mainly for the odd component, but as time increases the impacts become very obvious.

We get similar behaviour of fast drop in the plots for the nonlinear wave equation for the C^1 IC (3.82), Figure 4.6, as for the linear wave equation, Figure 3.8. In addition, Figure 3.9 is applicable here too as we get the same figure at $t = 0.16$, Figure 3.9a, and $t = 2$, Figure 3.9c, here as well. In short, the fast drop happens to both equations plots for the same times, N and the range of k , see Subsection 3.4.4 for more understanding. It is stated in Subsection 3.4.4 that Figure 3.9 shows how the drop in range decreases as step size decreases, $h = 10^{-2}, 10^{-3}, 10^{-4}$. In addition, it shows how the range of k giving algebraic convergence grows as step size decreases at any fixed time, and the same holds here too. However, the figure of the nonlinear equation at $t = 1$, Figure 4.6c, does not possess the same conduct as for the linear equation in Figure 3.9b, the reason for this is given later. Noteworthy, only the even coefficients component of the plots of the nonlinear problem experience the drop in; the odd coefficients component does not, similar to the linear problem. The spectra figures of the numerical solution in this subsection show that the drop in occurs only when using the trapezoid method with the second IC (3.82) but it

does not happen when using the third IC (3.83), similar to the linear problem, we believe that is due to the same reason given there regarding the nonsmoothness. We have also seen that changing N gives the same plots for all ICs correspondingly, similar as for the linear problem, see Subsection 3.4.4 for more information.

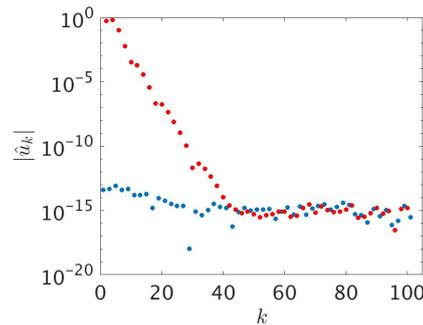


Figure 4.4: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with trapezoidal method for the C^∞ IC (3.81) at $t = 0.9$. The coefficients are evaluated for $h = 10^{-3}$ and $N = 100$. Similar plot is obtained for $t = 1$ and $t = 4$. The blue dots are the even coefficients component and the red dots are the odd coefficients component.

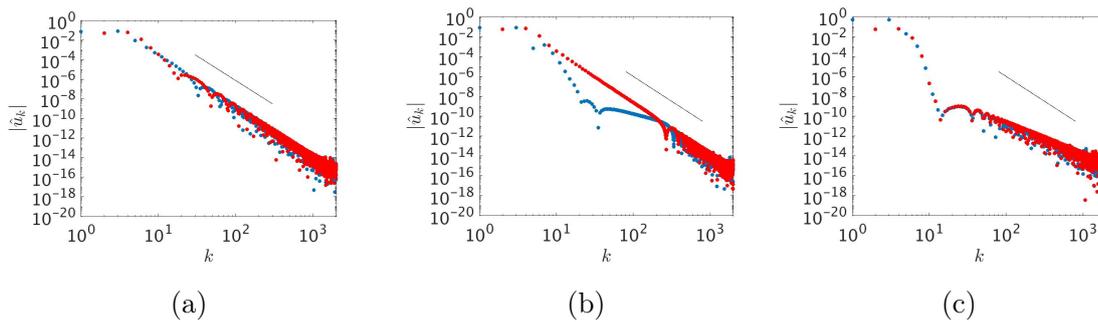


Figure 4.5: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with trapezoidal method for the C^3 IC (3.83) at $t = 0.9, 1, 4$ (a, b, c respectively). The coefficients are evaluated for $h = 10^{-3}$ and $N = 2000$. The black straight line indicates order $\frac{1}{k^5}$.

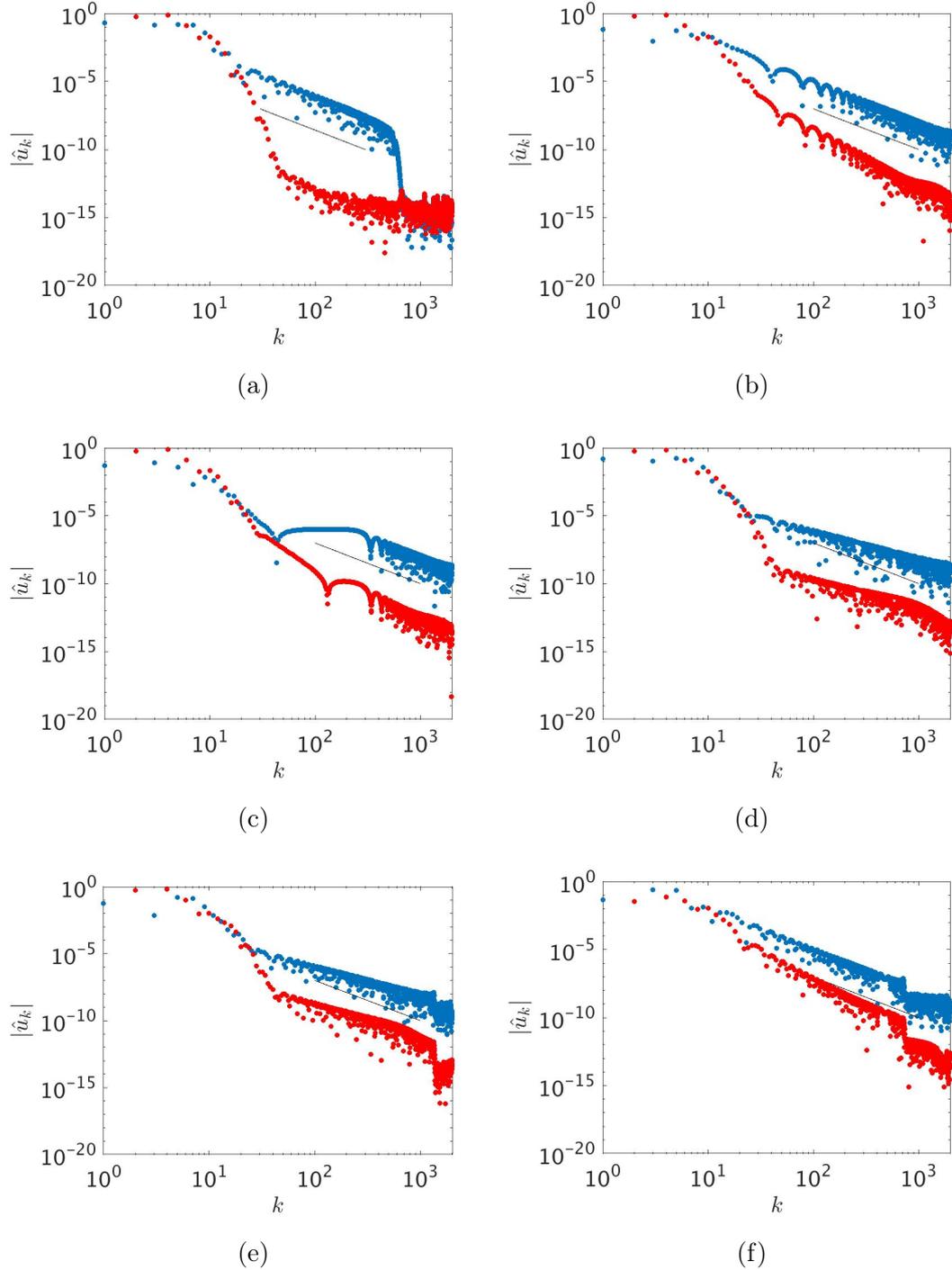


Figure 4.6: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with the trapezoidal method at $t = 0.04, 0.9, 1, 2, 4, 5.3$ (a, b, c, d, e, f panels respectively). The step size is $h = 10^{-3}$ with $N = 2000$ for the C^1 IC (3.82). The blue dots are the even coefficients component and the red dots are the odd coefficients component. The black straight line indicates order $\frac{1}{k^3}$.

Noticeably, as described in Subsection 3.4.4, we get similar spectral coefficients plots when solving the nonlinear wave equation (4.1) of the numerical solution using the trapezoidal method combined with pseudospectral method, for both ICs, the C^1 IC (3.82) and the C^3 IC (3.83), respectively to their corresponding figures above. For example, the plots at $t = 2, 6, 8, 10, \dots$ etc., are as the plots at $t = 2$ and $t = 4$; at all even positive t are similar, and at $t = 3, 5, 7, \dots$ etc., are like at $t = 1$; at all odd positive t are similar. In all other times, bigger than $t = 0.04$, the plots are like at $t = 0.9$. This is clarified next in detail.

Firstly, we will explain and investigate the above for the figure of the C^1 (3.82) of the numerical solution spectra Figure 4.6. Similar to previous chapter, six times are considered for the Chebyshev coefficients of the numerical solution in Figure 4.6, $t = 0.04, 0.9, 1, 2, 4$ and $t = 5.3$. The figures conduct is similar to Figure 3.8. That is figures at $t = 6, 8, 10, \dots$ etc., are similar to each other, i.e., they are like Figures 4.6d and 4.6e. And at $t = 3, 5, 7, \dots$ etc., are like Figure 4.6c, and the figures at rest of times are similar to Figures 4.6b and 4.6f. Nevertheless, for the nonlinear problem only figures at $t = 0, 2, 4, 6, \dots$ etc., have the special behaviour as in Figure 3.8 for such times. For instance, the Chebyshev coefficients of the numerical solution spectra at $t = 2$ and $t = 4$ in Figures 4.6d and 4.6e, decay exponentially up to the $\mathcal{O}(h^2)$ error expected for the trapezoid method but after that they decay algebraically, which is consistent with the $\mathcal{O}(h^2)$ error expected for the trapezoid method. More specifically, the even Chebyshev coefficients at $t = 2$ and $t = 4$ in Figures 4.6d and 4.6e, respectively, decay exponentially down to 10^{-6} for $h = 10^{-3}$, up to the time stepping error $\mathcal{O}(h^2)$, similar to the linear problem in Figures 3.8d and 3.8e, and the odd Chebyshev coefficients decay exponentially down to 10^{-10} , but after that they both decay algebraically. The conduct in Figure 3.9 happens for the C^1 (3.82) too for Figures 4.6d and 4.6e; the range giving exponential convergence grows as the time step decreases. However, figures at $t = 1, 3, 5, \dots$ etc., do not have the same special conduct as in Figures 3.8 and 3.9 for such times, which we do not expect to have so as we do not know the exact solution behaviour at these times, while the exact solution for the linear problem is an odd function at these times. That is, the even Chebyshev coefficients of the numerical solution at $t = 1, 3, 5, \dots$ etc., are not zero down to 10^{-6} for $h = 10^{-3}$ down to $\mathcal{O}(h^2)$ and then decaying algebraically; they are not zero up to the $\mathcal{O}(h^2)$ error expected for the trapezoid method. Nevertheless, all figures are similar at these specific times as explained above.

Consequently, the similar behaviours explained above are applicable for the third IC (3.83) in Figure 4.5, apart from the drop in as there is no fast drop in the plots of the numerical results for this IC similar as the linear problem. Figure 4.5 is like the corresponding figure of linear problem, Figure 3.7, apart from the figures at $t = 1$. Figure 4.5c, at $t = 4$, which is like Figure 3.7c, shows that the numerical solution spectra decay exponentially more faithfully than using the C^1 (3.82) at this time; they decay exponentially down to 10^{-10} for $h = 10^{-3}$, which means the range giving exponential convergence is bigger than using the C^1 (3.82) in Figures 4.6d and 4.6e, see the explanation of Figure 3.7c in Subsection 3.4.4. The conduct in Figure 3.9 happens for the C^3 IC (3.83) as well for Figure 4.5c. While again the figure at $t = 1$, Figure 4.5a, is not special, similar as the C^1 (3.82) case given above.

According to our numerical results above of the nonlinear wave equation (4.1), one concludes that the trapezium rule deals more accurately with the C^3 IC (3.83) than the C^1 IC (3.82), similar to the linear problem. We assume that is again perhaps due to the nonsmoothness problem presented more severely by the C^1 IC (3.82). That is clear in Figures 4.5 and 4.6, of the C^3 IC (3.83) and the C^1 IC (3.82), respectively, see Subsection 3.4.4 for a better understanding.

In Sections 1.9 and 3.4 we have shown that the exact d'Alembert solutions for the linear wave equation (3.52) at any time which is of multiple 4, i.e., at $t = 0, 4, 8, \text{etc.}$, are infinitely smooth for all three ICs since they have the returning property; they are the initial conditions functions $f(x)$ at these times. Also, at the other times of multiple 2, i.e., at $t = 2, 6, 10, \text{etc.}$, the exact d'Alembert solutions are infinitely smooth for all three ICs because the solutions are $-f(-x)$. On the contrary, in Subsection 3.4.4 we see that the numerical solutions for the very nonsmooth IC, the C^1 (3.82), when solving the linear wave equation (3.52) are not very special at these specified times than at other times as the Chebyshev coefficients of the numerical solution decay algebraically not exponentially; that is perhaps because of the accumulated error. To be more precise, the even Chebyshev coefficients of the numerical solution, of the trapezoidal method combined with pseudospectral method, decay exponentially up to the time stepping error $\mathcal{O}(h^2)$ but after that they decay algebraically. Our numerical results in Figure 4.6 show that the same holds for the numerical solutions for the nonlinear wave equation (4.1); as one can see that Figures 4.6d and 4.6e, which are at $t = 2$ and $t = 4$, respectively. Figures have similar behaviour to their corresponding plots for the linear wave equation

in Subsection 3.4.4 (Figures 3.8d and 3.8e). Nevertheless, our numerical results of using the C^3 (3.83) at $t = 4$ in Figure 4.5c show that the numerical solution spectra decay exponentially more faithfully than using the C^1 (3.82) at this time; the range giving exponential convergence is bigger, similar to the Figure 3.7c. This holds at all times of multiple 2, i.e., at $t = 0, 2, 4, 6, 8, \text{etc.}$, for both of the ICs above respectively.

The d'Alembert solution of the linear wave equation is smooth for the first IC, (3.81), but it is only C^1 for the second IC, (3.82), and it is C^3 for the third IC, (3.83), however, the solutions are infinitely smooth at $t = 0, 2, 4, 6, 8, \text{etc.}$, for all three ICs, see Sections 1.9 and 3.4. We do not know the exact solution for the nonlinear wave equation, but the decay of the Chebyshev coefficients of the numerical solution, of the trapezoidal method combined with pseudospectral method, in Figures 4.5 and 4.6 strongly suggest that the same holds here. Furthermore, the nonlinear wave equation share the same characteristics with the linear equation; they both have the same characteristics Figure 1.5. Therefore, based on the characteristics fact, we expect the exact solutions for the nonlinear wave equation to be smooth at $t = 0, 2, 4, 6, 8, \text{etc.}$, as well for all three ICs. Nevertheless, that is not very clear in the numerical results here because as one can notice that the numerical solution, of the trapezoidal method combined with pseudospectral method, show the exponential convergence only to a limited extent in the plots but not fully, especially for the very nonsmooth IC, the C^1 (3.82), at the specified times above, similar as the linear wave equation, see Subsection 3.4.4. In Chapter 5 some other time stepping methods, namely the exponential integrators, ETD1 and ETD2RK, will be used to investigate the behaviour of the numerical solution of the nonlinear wave equation, and we will then examine our assumption again.

Finally, we can say that: it is clear the plots of the spectra (Figures 4.5 and 4.6) show the algebraic decay more clearly than the convergence plots in Figure 4.3, thus, we understand Figure 4.3 more explicitly now.

4.3.5 Stability for Euler's and the modified Euler methods

Euler's method for solving the nonlinear wave equation (4.1) is always unstable as is shown in Figure 4.7, which is for the IC (3.81). One can see in this Figure that the numerical solution is increasing rapidly and this will cause the error to grow

rapidly. Similar behaviour is occurring for the other two ICs as well. In addition to Euler's method, the modified Euler method is unstable for solving (4.1). The results here are similar as for Euler's method.

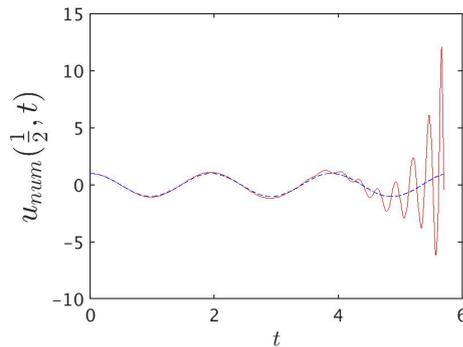


Figure 4.7: The numerical solutions $u_{num}(x, t)$ for the nonlinear wave equation (4.1) solved using Euler's method, at $x = \frac{1}{2}$, for two time steps, $h = 10^{-2}$ (red line), $h = 10^{-3}$ (dashed blue line). Applied for the C^∞ IC (3.81) and for $N = 10$.

4.3.6 Stability for the implicit Euler and the trapezoidal methods

The implicit Euler and the trapezoidal methods are used to solve the nonlinear wave equation (4.1) using the three initial conditions with different smoothness properties. In order to study the stability two main figures are obtained, where the numerical solution at $x = \frac{1}{2}$ is plotted, Figures 4.8 and 4.9 for the implicit Euler and the trapezoidal methods, respectively. From Figure 4.8, one can notice that the numerical solutions using the implicit Euler method are decreasing and approaching zero as time is growing. Moving to the trapezoidal method, all the three initial conditions are considered in order to investigate its stability more fully. Clearly from Figure 4.9 all the numerical solutions using the trapezoidal method are bounded and vibrate between -2 and 2 . Hence, both methods are stable.

If we compare the figures for the nonlinear wave equation, Figures 4.8 and 4.9, with the related figures for the linear wave equation in Chapter 3, Figures 3.14 and 3.15, respectively, the figures look very similar. The results for the nonlinear equation

are behaving the same as for the linear equation: the numerical solution computed by the implicit Euler method decays to zero and the solution computed by the trapezoidal method is bounded.

In order to understand those Figures, 4.8 and 4.9, more investigations are done. Some graphs for the numerical solution, u_{num} , are given, see Figures 4.10 and 4.11. Those figures are giving a clearer idea about the behave of numerical solutions at all the three initial conditions and it is more obvious in the contour Figure, 4.10. Apparently the ICs (3.81) and (3.83) are giving a very regular numerical solution, see Figures 4.10 and 4.11, the panels (a) and (b) for the IC (3.81) and the panels (e) and (f) for the IC (3.83).

Again by comparing the figures here, Figures 4.10 and 4.11, with the related figures for the linear wave equation in Chapter 3, Figures 3.16 and 3.17, respectively, the figures seem to be similar again. It is obvious that the numerical solution and contour of the linear and nonlinear wave equations are almost very close to each other.

Figures 4.10 and 4.11 show that the numerical solutions are quasi-periodic for all the ICs implemented, with similarity 4 units later, as literature in Section 1.6 proved for our nonlinear wave equation on infinite intervals but here we have a finite domain. Here again the first IC is a special case and the period is indistinguishable at also period 2, see Figures 4.10a, 4.10b, 4.11a and 4.11b, the same as for the linear wave equation in Subsection 3.4.7 (see the left panels of Figures 3.16 and 3.17).

Generally, the comparison made throughout this chapter, and in the upcoming chapters, between figures of the linear and nonlinear wave equations is giving more credibility to our results for the nonlinear wave equation.

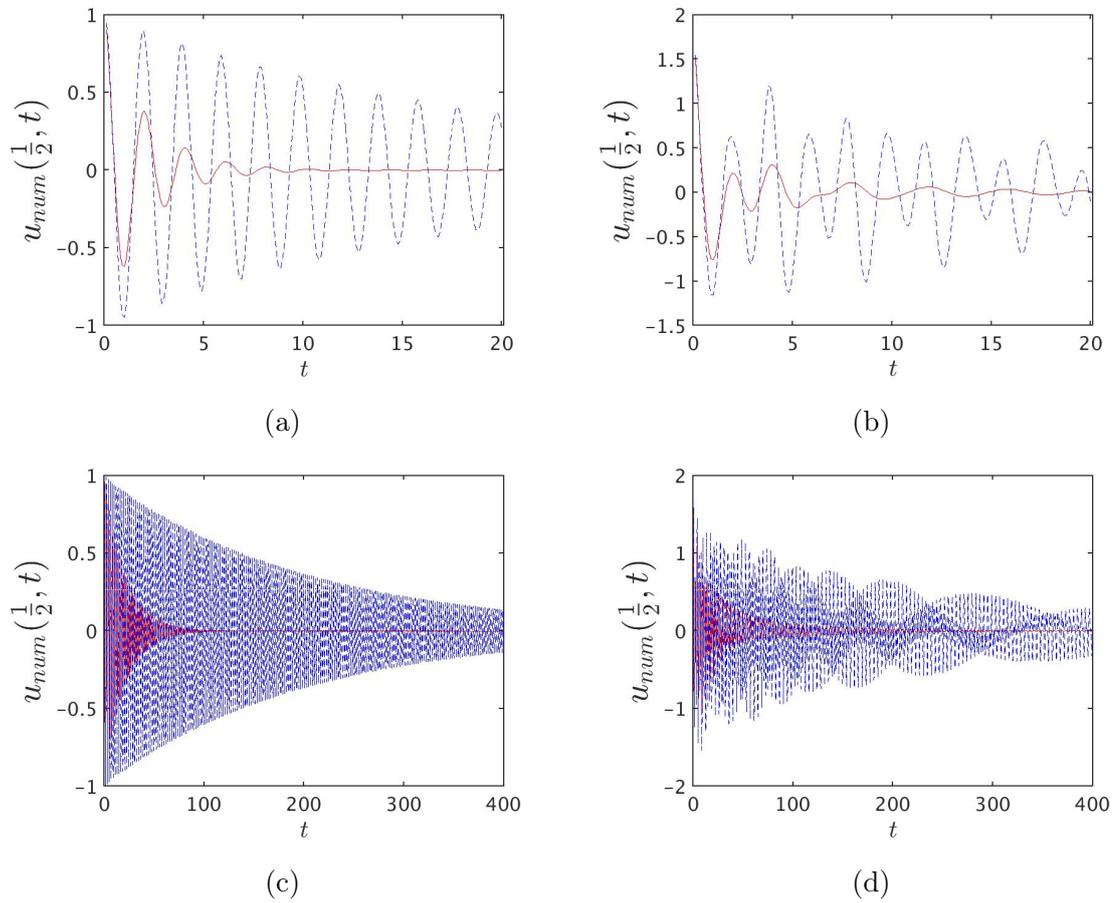


Figure 4.8: The numerical solutions $u_{num}(x, t)$ for the nonlinear wave equation (4.1) using the implicit Euler method, at $x = \frac{1}{2}$, for two time steps, $h = 10^{-1}$ (red line), $h = 10^{-2}$ (dashed blue line) (a,b panels) and $h = 10^{-2}$ (red line), $h = 10^{-3}$ (dashed blue line) (c,d panels). Applied for the initial condition, (3.81), $f(x) = \sin(\pi x)$, (left panels) and for the initial condition (3.82), $f(x) = e^{\sin(\pi x)} - 1$, (right panels) and for $N = 10$.

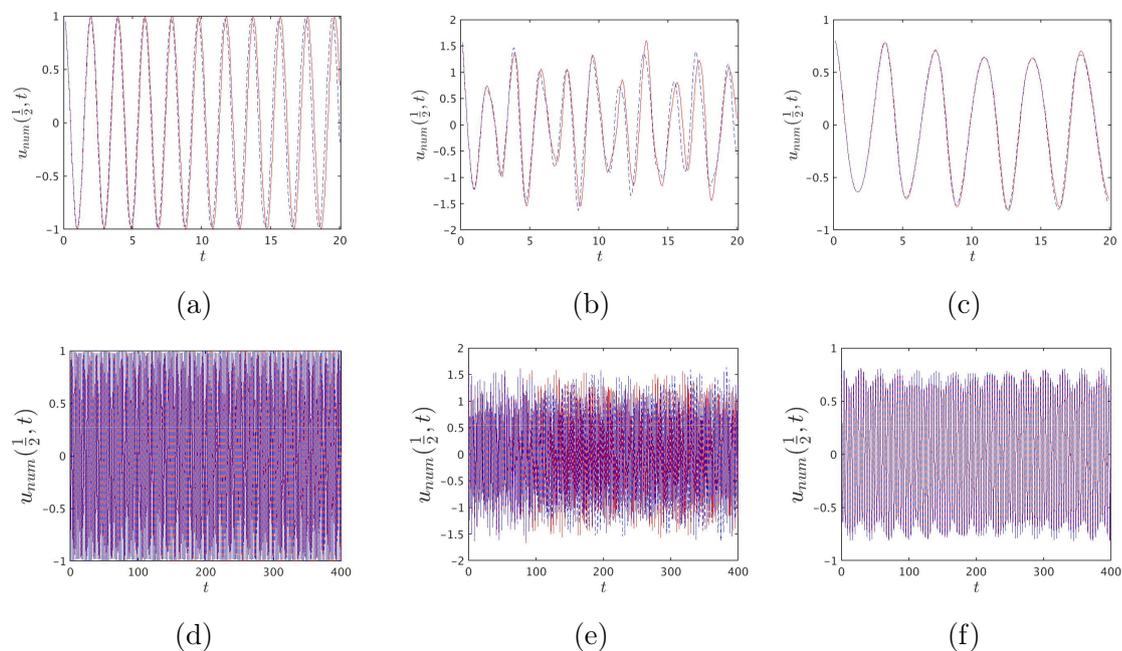


Figure 4.9: The numerical solutions $u_{num}(x, t)$ for the nonlinear wave equation (4.1) using the trapezoidal method, at $x = \frac{1}{2}$, for two time steps, $h = 10^{-1}$ (red line), $h = 10^{-2}$ (dashed blue line) (a,b,c panels) and $h = 10^{-2}$ (red line), $h = 10^{-3}$ (dashed blue line) (d,e,f panels). Applied for the initial condition (3.81) (a,d panels), for the initial condition (3.82) (b,e panels) and for the initial condition (3.83) (c,f panels) and for $N = 10$.

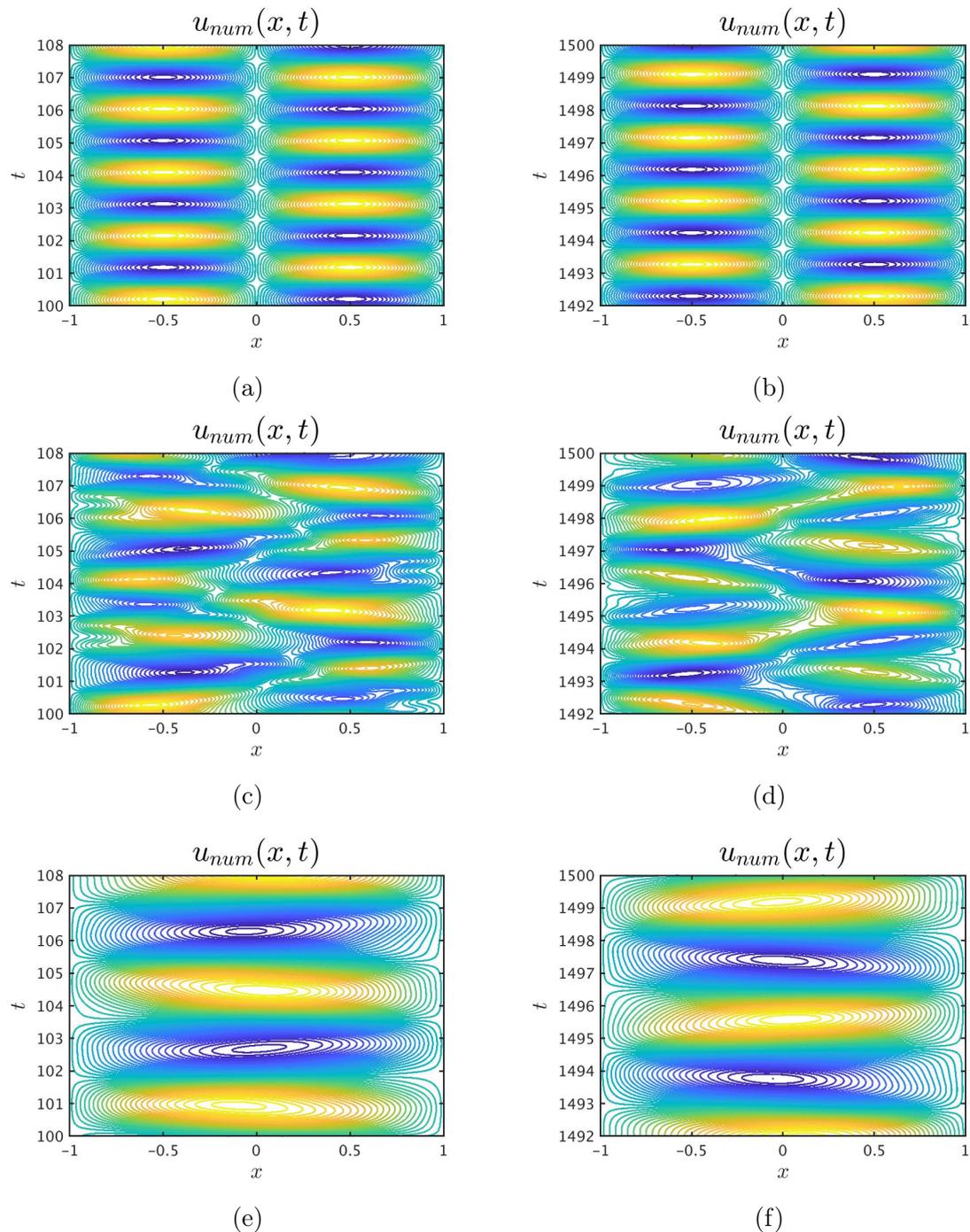


Figure 4.10: The contour of the numerical solution $u_{num}(x, t)$ for the nonlinear wave equation (4.1), using the pseudospectral method with the trapezoidal method, at $N = 100$, $h = 10^{-2}$ and for $t = 100 - 108$ (left panels) and $t = 1492 - 1500$ (right panels). Using the initial condition (3.81) (panels a and b), the initial condition (3.82) (panels c and d) and the initial condition (3.83) (panels e and f).

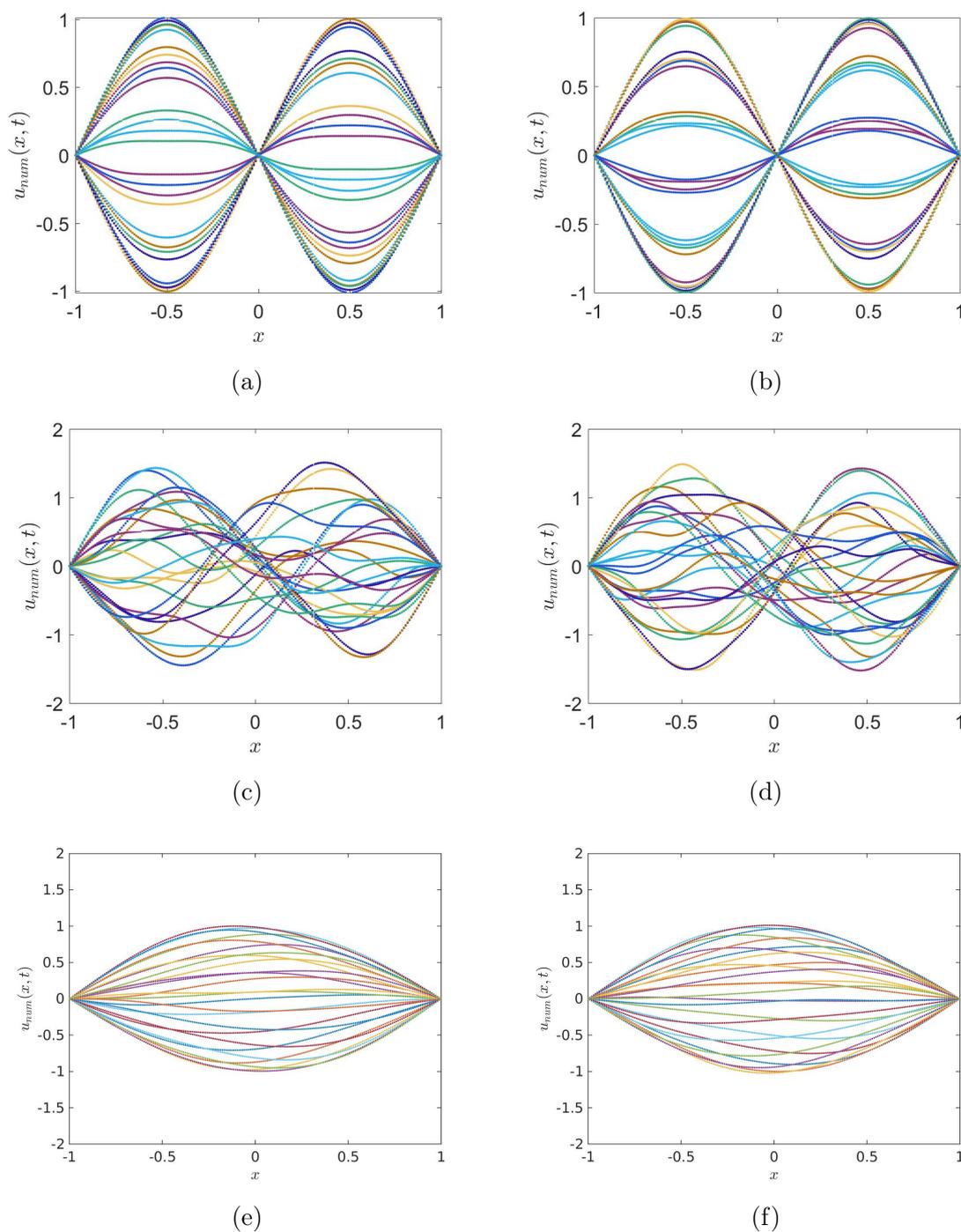


Figure 4.11: The numerical solution $u_{num}(x, t)$ for the nonlinear wave equation (4.1), using the pseudospectral method with the trapezoidal method, at $N = 100$, $h = 10^{-2}$ and for $t = 100 - 104$ (left panels) and $t = 1496 - 1500$ (right panels). Using the initial condition (3.81) (panels a and b), the initial condition (3.82) (panels c and d) and the initial condition (3.83) (panels e and f).

4.4 Hamiltonian structure

The preservation of the Hamiltonian structure of the nonlinear wave equation by the explicit Euler method, the implicit Euler method and the trapezoidal method, i.e., symplecticity of each numerical scheme, is studied in this section.

One can see from Figures 4.8 and 4.9 in Section 4.3.6 that the error for the implicit Euler and the trapezoidal methods is bounded and that both of the methods are stable. The solution for the implicit Euler method approaches zero as time increases, and the solution for the trapezoidal method neither decays nor grows in amplitude. The implicit Euler method is more stable than the explicit Euler method and the trapezoidal method for the linear problem due to its stability region, see Section 2.2, and our examples are evidence for that. In this section it is found that the energy is not conserved by either Euler's or the implicit Euler method. On the other hand, the trapezoidal method does more or less preserve the energy so it simulates the structure of the nonlinear wave equation more faithfully than the Euler's and implicit Euler methods.

In order to know whether the trapezoidal method is stable generally or not for the nonlinear wave equation (4.1), i.e., the energy is preserved, the Hamiltonian structure of this PDE is studied in Section 4.1, and the symplecticity of the trapezoidal method is examined in depth here, the symplecticity of the other two schemes is also briefly studied. The theoretical studies in Section 4.1 show that our nonlinear wave equation is Hamiltonian, in other words, the energy is conserved. We need to study whether our numerical schemes also preserve the energy or not, if so that leads to a proof that the method is stable.

Knowing whether the numerical scheme is symplectic will help to know its stability. The explanation of Hamiltonian equations and symplectic methods is given in Section 2.5. A symplectic (or conjugate symplectic) scheme applied to a Hamiltonian ODE almost preserves the energy, as explained in Section 2.5. It follows that the solution cannot grow without bound, so the method is stable. It is not clear how much this generalizes to PDEs, which can be seen as infinitely dimensional ODEs. Nevertheless, after semi-discretization the PDE transforms into a system of ODEs. Although our nonlinear wave equation (4.1) is Hamiltonian, this does not mean that the system of nonlinear ODEs that results from semi-discretization is Hamiltonian. We can, however, say that the ODEs approach the

PDE in some sense as $N \rightarrow \infty$, so they must be approximately Hamiltonian for large enough N , see (Li *et al.*, 2017). As discussed in Section 2.5, the trapezoidal method is conjugate symplectic and thus its energy error is bounded over long time intervals if applied to Hamiltonian ODEs. While our semi-discretized ODEs are not necessarily Hamiltonian, they are almost Hamiltonian, so we expect the energy error to be bounded.

To lend some credence to the reasoning, we study the extent to which the energy is conserved in the following sections.

4.4.1 Symplecticity of the numerical schemes

In order to assess the nonlinear stability of numerical schemes for the nonlinear wave equation, we need to see if the numerical solution has the same behaviour as the exact solution. In the following subsections, the symplecticity for the explicit Euler, the implicit Euler and the trapezoidal methods, is studied.

4.4.2 Symplecticity of Euler's and the implicit Euler methods

We know from Section 2.5 that Euler's and the implicit Euler methods are not symplectic and this is confirmed in this section using our results.

First one needs to know the exact energy for all initial conditions, (3.81), (3.82) and (3.83) in order to study the symplecticity of the numerical schemes.

The exact energy for the first initial condition (3.81), $f(x) = \sin \pi x$, is found by evaluating the integral in (4.7) analytically, where $u(x, 0) = \sin \pi x$. The result of this integration gives that the exact energy is $\pi^2/2 + 6/32$ which is approximately 5.12230220, we denote the exact energy by E_{exact} . For the second initial condition (3.82), $f(x) = e^{\sin(\pi x)} - 1$, one can approximate the exact energy numerically using Maple by substituting this IC into (4.7) getting its approximate value to be 8.54547821. Similarly, for the third IC (3.83), $f(x) = \frac{1}{11}e^x(11 - 8x - 10x^2 + 8x^3 - x^4)$, the exact energy value is 1.50533917.

Figure 4.12, the left panel, shows that the energy for Euler's method is not constant and it is increasing rapidly. Hence, the method is not symplectic. If we look at the energy of the implicit Euler method, Figure 4.12, the right panel, we see that the

energy approaches zero as time goes to ∞ . However, these results are not accurate when compared with the exact energy of the solution of our PDE (4.1), which is a constant number, 5.12230220. The same holds for the other two ICs. Therefore, one can say that the implicit Euler method is also not symplectic for our PDE.

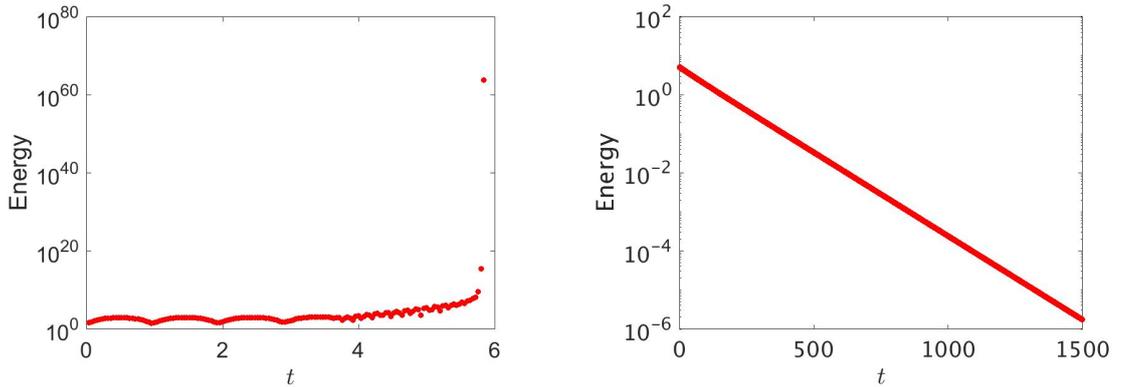


Figure 4.12: The energy of the nonlinear wave equation (4.1) for the C^∞ IC (3.81), $f(x) = \sin \pi x$, using Euler's method (left panel), for $N = 10$ with time step $h = 10^{-2}$, and the implicit Euler method (right panel), for $N = 100$ with time step $h = 10^{-3}$.

4.4.3 Symplecticity of the trapezoidal method

The trapezoidal rule is not symplectic but the implicit midpoint method is symplectic. And because the trapezoidal rule is conjugate to the implicit midpoint rule, hence one can say that the trapezoidal method is conjugate symplectic, see Section 2.5. More information one can find in (Hairer *et al.*, 2002; Leimkuhler & Reich, 2004; Sanz-Serna & Calvo, 1994).

We compute the energy of the numerical solution as computed by the trapezoidal method, starting from three initial conditions, $f(x) = \sin(\pi x)$ (3.81), $f(x) = e^{\sin(\pi x)} - 1$ (3.82) and $f(x) = \frac{1}{11}e^x(11 - 8x - 10x^2 + 8x^3 - x^4)$ (3.83). As mentioned before in Chapters 1 and 3, we choose these ICs because they yield solutions with different levels of smoothness for the linear wave equation (3.52). They are selected so that the initial condition itself is infinitely differentiable on the entire interval, but the generated analytic solution for the equation (3.52) at later times will have

discontinuities in the second derivative for the second IC (3.82) and discontinuities in the fourth derivative for the third IC (3.81). The first initial condition (3.81) yields a smooth solution, it is symmetric and infinitely continuously differentiable in space for $[-1, 1]$ for all derivatives. However, the second initial condition (3.82) introduces non-smooth solution at the second derivative, While the third IC, (3.83), generates continuously differentiable solution for the second derivative, but is not for the fourth derivative, so smoother solutions than for the second IC (3.82), but still not infinitely differentiable as for first IC (3.81). In Subsection 4.3.4 we showed that the discontinuities present in the linear wave equation are also present for the nonlinear wave equation. The following numerical experiments and results study the discontinuity from a different perspective, namely from the perspective of energy conservation.

Figures 4.13-4.15 show the energy error of the numerical solution computed by the trapezoidal rule for these three initial conditions. The energy error $|E - E_{exact}|$ is small for large N used here, as one can see in those figures, where E is the numerical energy and E_{exact} is the exact energy given in Subsection 4.4.2. The reason is that the numerical energy is very close to the exact energy value. But for small N , $N = 5$, the error will be bigger. It is obvious in these figures that the energy error is bounded for all the three ICs. This means that the trapezoidal method is stable for solving the Hamiltonian PDE, (4.1), for these three initial conditions. To be careful, we should only conclude that the method is stable for the initial conditions and step sizes considered here.

There are three main sources of numerical error in our numerical computation, which are; the step size h , the number of modes N and the number of points that are used for the trapezoid rule integration, which are equally spaced quadrature points denoted by nx (see Section 4.1 and Subsection 4.3.1).

For all of these three initial conditions the energy error decreases to zero as the step size h decreases to zero. More precisely, the energy error behaves like $\mathcal{O}(h^2)$ as expected for second order methods: the energy error decreases by a factor of 100 when step size h decreases by a factor of 10. For example, with the first initial condition the energy error is approximately 10^{-5} for $h = 10^{-2}$, for $h = 10^{-3}$ it is around 10^{-7} and for $h = 10^{-4}$ it is 10^{-9} , see Figure 4.13. The other two ICs behaves similarly, see Figures 4.14 and 4.15, respectively.

The number of modes N has a big effect on the results of these calculations, and the

effect varies depending on the smoothness of the solution. For the first IC, (3.81), it suffices to take N around 20; as explained before in Section 4.3.4 the numerical solution for the first IC (3.81) is exponentially convergent with N , see Figures 4.3-4.6, so the number of modes needed is small. However, the numerical solution for other two ICs are algebraically convergent with N , with order approximately $\frac{1}{N^3}$ for (3.82) and order $\frac{1}{N^5}$ for (3.83), see the same figures. Therefore, the number of modes must be large for the last two ICs especially for the second IC, which is giving the most nonsmooth solution. For the third IC, (3.83), a good choice is $N = 100$, see Figure 4.15; for consistency, we use the same value of the first IC. In contrast, for the second IC (3.82) it must be around $N = 1000$, see Figure 4.14. Even with this large number of modes the energy began to be somehow irregular for 10^{-4} , see Figure 4.14c. Therefore, choosing smaller h than the ones used here will force us to choose even bigger N . Figure 4.17, shows how the choice of N affects the results for the second IC (3.82). In this figure it is clear that choosing $N = 100, 200, 400, 800$ is giving us wrong results especially for the smallest N here, $N = 100$, if compared with our accurate results in Figure 4.14 (panels c and f) where the energy error is approximately 10^{-8} for $h = 10^{-4}$ and the figure is more regular. The smaller N is the more wrong outcome one gets. For example, $N = 100$ is giving totally not good result, even though $N = 100$ is a good number of modes for the third IC and surely for the first IC as well. One can notice in this figure that by increasing N the results are improving, but still they are not accurate though. This is showing that N must be chosen to be larger than the N used in this Figure, 4.17, for the time step used here $h = 10^{-4}$, i.e., larger than $N = 800$, because Figure 4.14c is showing that the energy began to be somehow irregular even for $N = 1000$.

We identified another source of numerical error when computing the numerical energy, namely the number of points nx that are used for the trapezoid rule integration. These are different from the Chebyshev points, see Section 4.1 and Subsection 4.3.1. This source of error does not affect that much the first IC, which has a very smooth solution, but it does affect the other two ICs especially the second IC (3.82), the very non-smooth solution with a discontinuous second derivative. Therefore, for the first IC (3.81) even choosing $nx = 200$, for $N = 100$, we get that the energy conserved for h 's used here even for the small ones, see Figure 4.13. Whilst, for the IC (3.82), u_{xx} and u_{tt} are discontinuous, so u_x and u_t are not continuously differentiable. That in turn means that evaluating the integrals

of $(u_x)^2$ and $(u_t)^2$ in (4.6) via the trapezoid rule requires a large number of points to obtain accurate results. Consequently, for the second IC (3.82) we take the number of quadrature points, the equal spaced points nx , to be very large, to be something like $nx = 100N$, see Figure 4.14. In addition, for the third IC (3.83) we take $nx = 100N$ and this helps for small values of h , e.g. $h = 10^{-4}$, Figure 4.15. However, for larger h a fairly small number of quadrature points, say $nx = 200$ for $N = 100$, works well. Figure 4.16 shows the effect of reducing nx . The right panel shows that the energy error for the third IC with $h = 10^{-4}$ and $N = 100$ seems to be of the order of 10^{-8} if the energy is computed with $nx = 200$ quadrature points. However, this is not because of any error when solving the PDE but because of the error committed when computing the energy. Indeed, we know from Figure 4.15 that if we use $nx = 10000$, the energy error is of the order of 10^{-9} and the figure is more regular. However, the left panel shows that the energy error for the IC (3.82) with $h = 10^{-3}$ and $N = 1000$ is stuck at 10^{-4} regardless decreasing h to be 10^{-3} , the same thing occurs for other small values of the quadrature points, $200 < nx < 100N$, see Figure 4.14 for more clarification. The most important point in this figure regarding the IC (3.83) is that we have this inaccurate result for the very small value of $h = 10^{-4}$, but for $h = 10^{-3}$ things were working correctly in contradiction with what we have for the IC (3.82). The energy error for the second initial condition does not seem to go down as $h \rightarrow 0$: the energy error is of order 10^{-4} , regardless of whether h equals 10^{-2} , 10^{-3} or 10^{-4} . Noteworthy, increasing nx further, more than $100N$, does not have a noticeable effect.

One concludes that the energy is almost preserved when solving the nonlinear wave equation (4.1) using the trapezoidal method, at least for the three initial conditions considered here. This agrees well with the theory of Hamiltonian equations and symplectic methods, as explained in Section 2.5. It follows that the trapezoidal method is stable for the nonlinear wave equation. To be pedantic, we only showed this for the three initial conditions considered here, but it is not possible to say that the trapezoidal method is stable generally.

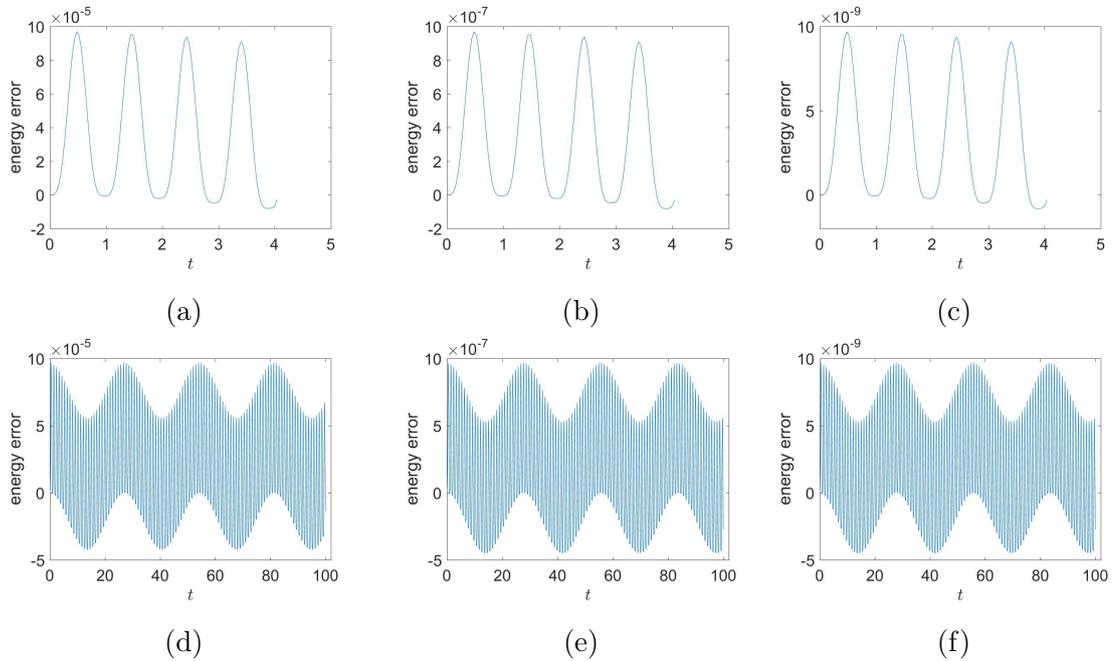


Figure 4.13: The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1), solved using the pseudospectral method with the trapezoidal method for the C^∞ IC (3.81), with $N = 100$, $nx = 200$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a, b, and c respectively) for both short and long periods of time. Here E is the numerical energy and E_{exact} is the exact energy.

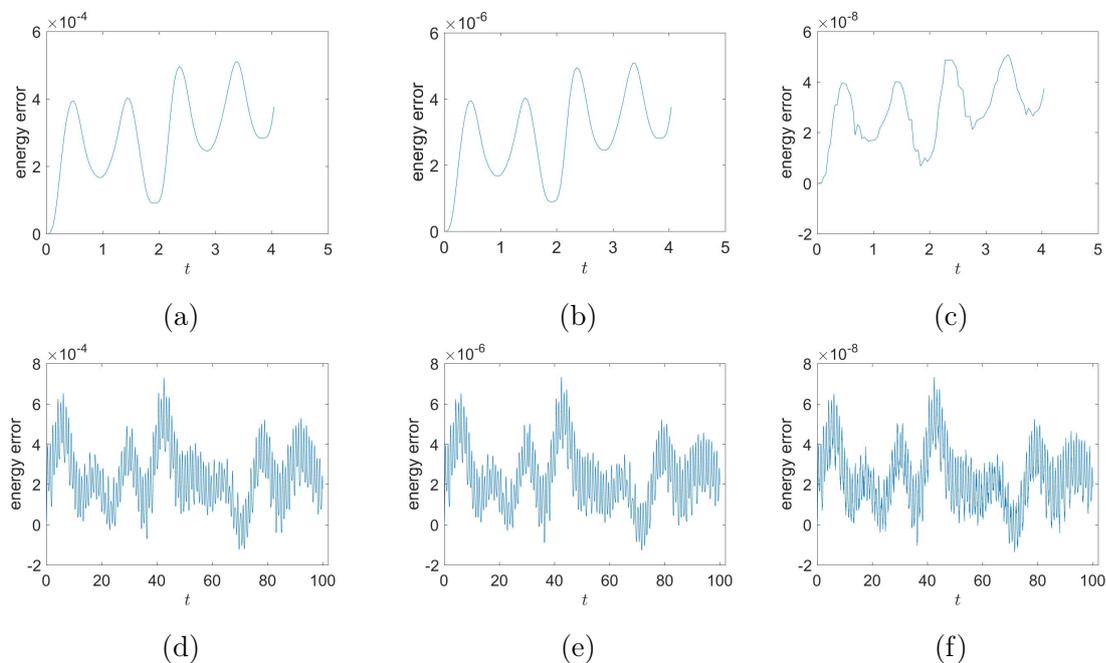


Figure 4.14: The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1), solved using the pseudospectral method with the trapezoidal method for the C^1 IC (3.82), with $N = 1000$, $nx = 100N$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a, b, and c respectively) for both short and long periods of time.

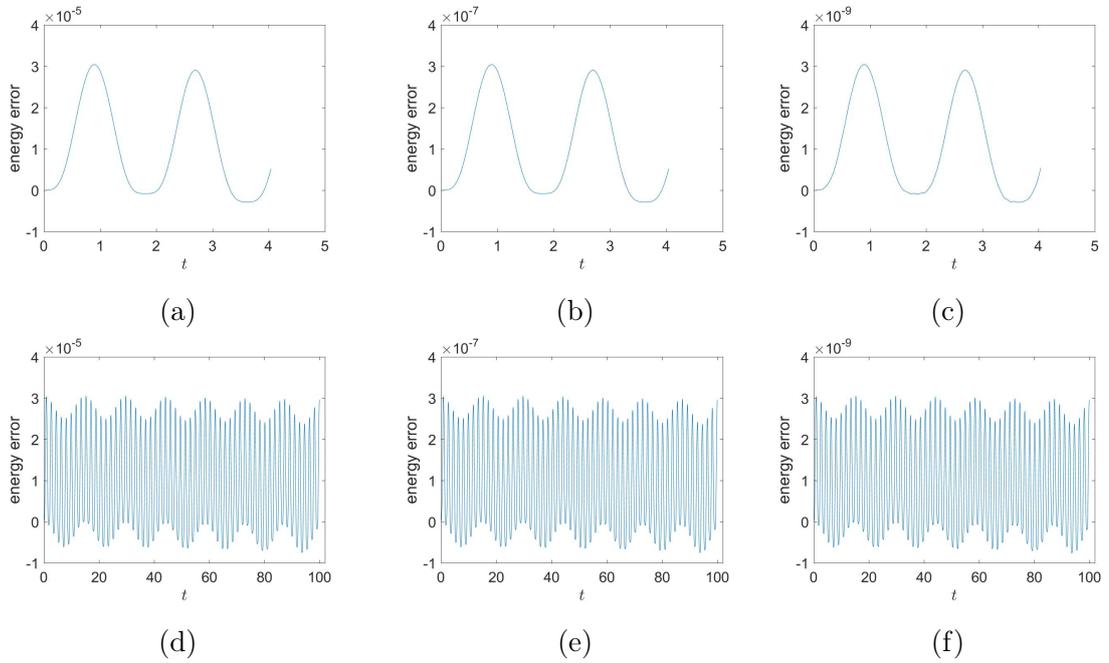


Figure 4.15: The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1), solved using the pseudospectral method with the trapezoidal method for the C^3 IC (3.83), with $N = 100$, $nx = 100N$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a, b, and c respectively) for both short and long periods of time.

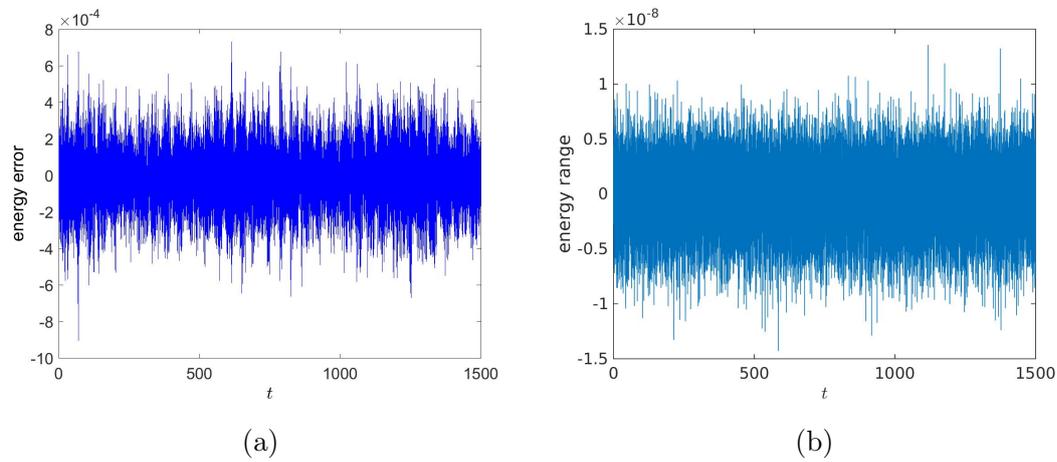


Figure 4.16: The energy error, $E - E_{exact}$, of the nonlinear wave equation (4.1) solved using the trapezoidal method with $nx = 200$. The C^1 IC (3.82) using $N = 1000$, $h = 10^{-3}$ and the C^3 IC (3.83), using $N = 100$, $h = 10^{-4}$ (panels a and b respectively).

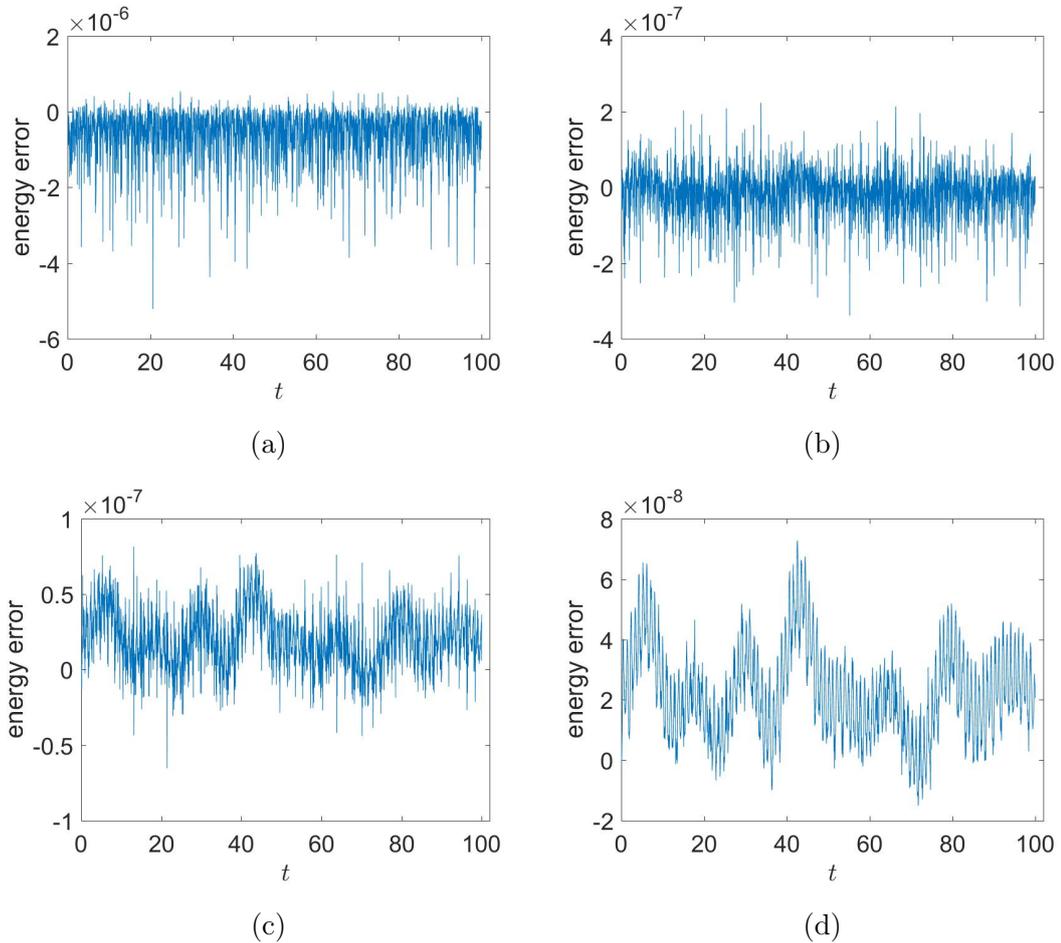


Figure 4.17: The energy error, $E - E_{exact}$, of the nonlinear wave equation (4.1) solved using the trapezoidal method with $h = 10^{-4}$, $nx = 100N$ and $N = 100, 200, 400, 800$ (a, b, c and d respectively) using the C^1 IC (3.82).

4.5 Summary and discussion

In this chapter the nonlinear wave equation $\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3$, (4.1), with homogeneous Dirichlet boundary conditions is solved in physical space using the pseudospectral method, with Chebyshev polynomials. Several time stepping methods are implemented: Euler's, the modified Euler, the implicit Euler and the trapezoidal methods.

This chapter, similar to previous chapter, studies the convergence, in space and time, and stability. In addition, the energies of the solutions (exact/numerical) are thoroughly investigated. Again three initial conditions are considered, the C^∞ IC (3.81), the C^1 IC (3.82), and the C^3 IC (3.83). These ICs are carefully chosen to generate solutions with different levels of smoothness. This leads to spectral convergence with respect to N for the C^∞ IC, (3.81), but for the other ICs, (3.82) and (3.83), the solution converges algebraically, with order approximately $\frac{1}{N^3}$, for (3.82), and order $\frac{1}{N^5}$, for (3.83). This holds for all time stepping methods employed.

Similar to Chapter 3, the convergence using the spectrum of the numerical solution of the combination of the trapezium method with pseudospectral method for the nonlinear wave equation (4.1) is thoroughly investigated. We see that for the first IC (3.81) plots of the coefficients decay exponentially similar to the linear wave equation in Chapter 3. In addition, the third IC (3.83) plots of the coefficients decay algebraically like the linear wave equation at general t . Interestingly, for the second IC (3.82) both of the spectral coefficients components, the even and odd, decaying algebraically, in contrast to the linear wave equation when only the even component decays algebraically, while the odd component decays exponentially at general t . We have seen that is because the nonlinearity in the equation mixes things up so that the discontinuity is no longer associated only with one part (the non-zero part) of the solution, but gets mixed to both of the odd and even parts (the zero and non-zero parts). Thus, each component of the solution has the odd and the even parts; they are mixed. That is clearer when using the second IC (3.82) as it has one non-zero component. Similar to previous chapter, the drop in happens only for the C^1 IC (3.82). Meaning that the trapezoidal method deals with the C^3 IC (3.83) better than the C^1 IC (3.82) for the nonlinear problem as well.

The special times here are only at $t = 0, 2, 4, 6, 8, \text{etc.}$, and at these times both of the Chebyshev coefficients components of the numerical solution decay exponentially, however, to a limited extent only but after that they both decay algebraically. The exact solution of the linear wave equation is known and is infinitely smooth at $t = 0, 2, 4, 6, 8, \text{etc.}$, for all three ICs. Meanwhile, the exact solution of the nonlinear wave equation is unknown, however, the nonlinear wave equation share the same characteristics with the linear equation, therefore, we anticipate that the exact solution for the nonlinear wave equation to be smooth at these specified times as well for all three ICs. Nevertheless, that is not very clear from our numerical results

as they show the exponential convergence only to a limited extent but not fully, especially for the very nonsmooth IC, the C^1 (3.82), at the special times above.

In addition, our results show that the global error of the Euler's and the implicit Euler methods is $\mathcal{O}(h)$, while for the modified Euler method and the trapezoidal method it is $\mathcal{O}(h^2)$. However, this is not our expectation in Chapter 1 as we assumed that the discontinuity will impact the time error as well as the spatial error, the reason for this contradiction probably because time stepping methods used here are not very high orders.

We now move to the stability of time stepping methods for solving the nonlinear wave equation (4.1). Similar as for the linear wave equation studied in previous chapter, both the Euler method and the modified Euler method are not stable at all. The stability of the other two methods, the implicit Euler and the trapezoidal, is assessed in more detail. Fortunately, both of these methods are stable for the initial conditions studied here, in other words, the error does not grow as time increases but instead it is bounded. For the implicit Euler method, the overall solution amplitude decays to zero as $t \rightarrow \infty$, so the error will approach a constant. For the trapezoidal method the overall solution amplitude remains a constant; it does not decay to zero, so the error will be bounded.

The stability of the trapezoidal method seems to be rather subtle, so it is not clear how strong a conclusion we can draw from numerical experiments using only three initial conditions. Therefore, in Section 4.4 the behaviour of the trapezoidal method is investigated from another point of view. Specifically, we look at the energy of the numerical solutions generated by time stepping methods, Euler's, the implicit Euler and trapezoid. As it turns out exact numerical energy conservation is too much to ask for, but the trapezoidal method does almost conserve the energy. Semi-discretization results in a system of ODEs which is not necessarily Hamiltonian, but they are expected to be almost Hamiltonian as they are resulted of the Hamiltonian PDE (4.1). The trapezoidal rule is conjugate symplectic, so it almost conserves energy when applied to Hamiltonian ODEs. Our results show that when applied to the semi-discretized nonlinear wave equation, the trapezoidal rule also almost conserves energy. It follows that the numerical solution is bounded; the overall amplitude solution remains a constant, and thus that the trapezoidal method is stable. Nevertheless, this is shown only for the three initial conditions considered here, however, it is not possible to say definitely that the trapezoidal

method is stable generally.

In order to determine which of the two time stepping methods, the implicit Euler and trapezoidal, is better the following three facts are considered. First regarding the solutions, the exact and numerical solutions. We know that the nonlinear wave equation (4.1) is a Hamiltonian PDE, so the energy of the exact solution is constant. Thus, in our case we have non-zero constant energy values for each IC used. That means, the exact solution for (4.1) is bounded and does not go to zero, and this is what we are getting from the numerical solution of the trapezoidal method. While the numerical solution of the implicit Euler method decays to zero as time increases. In other words, the numerical solution of the trapezoidal method behaves like the exact solution of the equation as both are bounded and not decaying to zero. The other fact is that the trapezoidal method is energy conserved for our equation, in contrast to the implicit Euler method, which does not. Another consideration is that the trapezoidal method is second order, while the implicit Euler method is first order. Accordingly, after taking the above points into account, one can say that the trapezoidal method is better than the implicit Euler for solving the nonlinear wave equation (4.1).

Chapter 5

Exponential integrators

In the present chapter, exponential integrators for time integration of two wave equations, the linear wave equation (3.52) and the nonlinear wave equation (4.1), are studied. These integrators, as their name suggests, use the exponential and often functions which are closely related to the exponential function inside the numerical method. Several kinds of methods belong to this class; among them, exponential time differencing (ETD) methods have received constant attention since the early 1960s (Certaine, 1960). The general formula for ETD-schemes of arbitrary order were developed by Cox & Matthews (2002), but their origin was years earlier: for example the first paper in this direction, as far as we know, was by (Certaine, 1960). Exponential integrators are a type of numerical method for solving ordinary differential equations, particularly initial value problems. The main feature of exponential integrators is that they solve exactly the linear part of the problem (Cox & Matthews, 2002; de la Hoz & Vadillo, 2008; Du & Zhu, 2004; Garrappa & Popolizio, 2011).

In earlier chapters, we utilised Chebyshev methods to discretize our two PDEs in space, and then solved the resulting ODEs in time using standard time stepping methods. In this chapter we use exponential integrators as time stepping methods. Specifically, we use the first order exponential time differencing (ETD1) and the second order exponential time differencing with Runge-Kutta time stepping (ETD2RK) to solve the linear wave equation (3.52) and the nonlinear wave equation (4.1) with homogeneous Dirichlet boundary conditions. Noteworthy for the linear wave equation, the resulting ODEs are linear so they will be solved exactly. The

performance of the numerical methods for the PDEs are studied regarding, the spatial and time stepping convergence, stability and energy conservation. As with the chapters before the three ICs, (3.81), (3.82) and (3.83), which generate different levels of smoothness, are implemented in our numerical experiments.

5.1 Derivation of exponential integrators

The numerical method treated in this chapter is the so-called "exponential time differencing" (ETD) schemes for solving two PDEs.

Once we discretize the spatial part of the PDE we get a system of ODEs, which can be written in vector form,

$$\vec{u}' = L\vec{u} + \vec{\Gamma}(\vec{u}, t), \quad (5.1)$$

where L and $\vec{\Gamma}$ are linear and nonlinear operators, respectively, where typically the operator L represents the stiff part of the equation and can be represented by a matrix. The ETD schemes solve ODEs of this form.

The material discussed in this section is standard, it is given in many articles (e.g., Cox & Matthews, 2002; de la Hoz & Vellido, 2008; Du & Zhu, 2004; Garrappa & Popolizio, 2011; Kassam & Trefethen, 2005; Krogstad, 2005; Minchev & Wright, 2005; Minchev, 2004).

5.1.1 Exponential time differencing method (ETD1)

To derive the ETD methods, we begin by multiplying (5.1) by the integrating factor e^{-Lt} and integrate from $t = t_n$ to $t = t_{n+1}$ to produce the exact result,

$$\vec{u}(t_{n+1}) = e^{Lh}\vec{u}(t_n) + e^{Lh} \int_0^h e^{-L\tau} \vec{\Gamma}(\vec{u}(t_n + \tau), t_n + \tau) d\tau, \quad (5.2)$$

where $h = t_{n+1} - t_n$. Then for the ETD1 the simplest approximation to the final term in (5.2) is that $\vec{\Gamma}$ is constant between $t = t_n$ to $t = t_{n+1}$, so the term becomes

$$e^{Lh} \int_0^h e^{-L\tau} \vec{\Gamma}(\vec{u}(t_n + \tau), t_n + \tau) d\tau = e^{Lh} \int_0^h e^{-L\tau} \vec{\Gamma}(\vec{u}(t_n), t_n) + \mathcal{O}(h^2) d\tau = \vec{M}_1 \vec{\Gamma}_n + \mathcal{O}(h^2), \quad (5.3)$$

where

$$\vec{M}_1 = L^{-1}(e^{Lh} - I), \quad (5.4)$$

and where I is the identity matrix. We denote the numerical approximation to $\vec{u}(t_n)$ by \vec{u}_n and $\vec{\Gamma}(\vec{u}(t_n), t_n)$ as $\vec{\Gamma}_n$.

This turns (5.2) into the scheme **ETD1**, given by

$$\vec{u}_{n+1} = e^{Lh}\vec{u}_n + L^{-1}(e^{Lh} - I)\vec{\Gamma}_n. \quad (5.5)$$

5.1.2 Exponential time differencing method with Runge-Kutta time stepping

In this section we obtain the ETD method of RK type of order 2. This method is analogous to the modified Euler method.

Instead of assuming that $\vec{\Gamma}$ is constant over the interval $t_n \leq t \leq t_{n+1}$ as we did for ETD1, we use the higher-order approximation that

$$\vec{\Gamma} = \vec{\Gamma}(\vec{u}_n, t_n) + (t - t_n)(\vec{\Gamma}(\vec{a}_n, t_n + h) - \vec{\Gamma}(\vec{u}_n, t_n))/h + \mathcal{O}(h^2), \quad (5.6)$$

where \vec{a}_n is given by

$$\vec{a}_n = e^{Lh}\vec{u}_n + L^{-1}(e^{Lh} - I)\vec{\Gamma}_n, \quad (5.7)$$

see (5.5).

We substitute this into (5.2) to yield the numerical scheme **ETD2RK**, given by

$$\vec{u}_{n+1} = \vec{a}_n + (\vec{\Gamma}(\vec{a}_n, t_n + h) - \vec{\Gamma}_n)(e^{Lh} - I - hL)L^{-2}/h. \quad (5.8)$$

5.2 Solving the linear wave equation

The linear wave equation (3.52):

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2},$$

is discretized to ODEs (3.71) and (3.72). This system of ODEs can be written as

$$\vec{u}' = L\vec{u}, \quad (5.9)$$

see (5.1), where $\vec{\Gamma} = 0$ here and we have

$$\vec{u} = \begin{bmatrix} \vec{y} \\ \vec{w} \end{bmatrix}, L = \begin{bmatrix} 0 & I \\ D^2 & 0 \end{bmatrix}. \quad (5.10)$$

In this section we study the convergence of N and h , the spectra of the numerical solution, the stability and energy conservation of the exponential time difference methods, mainly of the ETD1 method, when used to solve the above linear problem. We are implementing the same three initial conditions used in previous chapters, (3.81), (3.82) and (3.83), in our numerical experiments. The ICs are,

$$f(x) = \sin \pi x,$$

equation (3.81), and the second initial condition,

$$f(x) = e^{\sin(\pi x)} - 1,$$

equation (3.82). The third initial condition is,

$$f(x) = \frac{1}{11} e^x (11 - 8x - 10x^2 + 8x^3 - x^4),$$

equation (3.83). See Subsection 3.4.

5.2.1 Convergence of the ETD1 method

Exponential methods generally solve linear equations exactly. In other words, when the non-linear term $\vec{\Gamma}$ in (5.1) is zero the linear term can be solved exactly. This is clear when using ETD1 method, see (5.2), because the only term that remains in the right side of (5.5) is $e^{Lh}\vec{u}_n$, which is exact. The same holds for ETD2RK and any other exponential method and they no longer will be needed as there is no nonlinear term, hence, they are just ETD1 method. Therefore, they solve the generated ODEs of the linear wave equation (3.52), $\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$, which is discretized to ODEs (3.71) and (3.72), exactly. However, there is an error of the numerical solution and exact solution for the linear wave equation. This error is the spatial error, depending on N , not the time stepping error, depending on h , because the ETD methods solve only the ODEs exactly not the PDE. The spatial error is behaving similar as for

the nonlinear equation Figure 4.3, in other words, it converges exponentially for the IC (3.81) but algebraically for the ICs (3.82) and (3.83). Figure 5.1, in which the error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, is plotted against the truncation index N using the ETD1 method for the linear wave equation, (3.52), for the C^∞ IC (3.81), shows that there is no time stepping error; increasing the time steps has no impact. The same holds for the other two ICs. Admittedly, the t needs to be a multiple of h , hence, in this figure we use $h = 0.9$ instead of $h = 1$ as we use $t = 0.9$ in the other plots throughout the thesis.

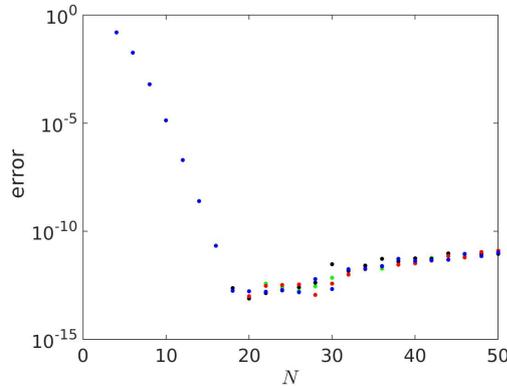


Figure 5.1: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the linear wave equation (3.52), for the C^∞ IC (3.81) solved using the pseudospectral method with the ETD1 method. The exact error is evaluated at $h = 10^{-1}, 10^{-2}, 10^{-3}$ and $h = 0.9$ for $t = 0.9$ (blue, red, black and green respectively). Here u is the exact solution and u_{num} is the numerical solution.

Following the same pattern of the previous two chapters, the Chebyshev coefficients for the ETD1 and ETD2RK methods combined with pseudospectral method have been investigated in this chapter as well for the linear wave equation (3.52), in this subsection, and for the nonlinear wave equation (4.1), in Subsection 5.3.1, for all three ICs employed.

The Chebyshev coefficients for the combination of the ETD1 and ETD2RK methods with pseudospectral method in this chapter possess the similar odd/even conduct as in Subsection 3.4.4 and 4.3.4, which are for the combination of the trapezium method with pseudospectral method when solving the linear and nonlinear equations, respectively. That is the odd/even Chebyshev coefficients

correspond to the odd/even components phenomena of the ICs; that is explained in Section 1.9, and to their generated exact d'Alembert solution and numerical solutions, where for the linear problem is given in Subsection 3.4.4 and for the nonlinear problem in Subsection 4.3.4. We will study this in detail throughout this chapter, for the linear problem in this subsection and for the nonlinear problem in 5.3.1.

As described and shown in Subsection 3.4.4 and 4.3.4, of solving the linear and nonlinear wave equations, respectively, that the spectra figures have some specific times when the figures are similar to each other, the times are at $t = 0, 2, 4, 6, 8, \text{etc.}$, $t = 1, 3, 5, \text{etc.}$, and all the times in between, for the C^3 IC (3.83) and the C^1 IC (3.82). This is the case in this chapter as well when using exponential integrators for both equations and ICs.

The linear wave equation is considered in this subsection, the Chebyshev coefficients for the numerical solution for the ETD1 method combined with pseudospectral method using the C^∞ IC (3.81) and the C^3 IC (3.83) are examined first. The Chebyshev coefficients of the IC (3.81) decay exponentially similar to the previous chapters; particularly the plot is like the Chebyshev coefficients plot of the numerical solution of the linear problem in Figure 3.6 and the Chebyshev coefficients of the exact solution in 3.10.

Moreover, the Chebyshev coefficients of the numerical solution using the C^3 IC (3.83) is studied in Figure 5.2. When comparing the three Figures for the linear problem of the third IC (3.83); the numerical solutions spectra plots of using the trapezoid and ETD1 methods combined with pseudospectral method, Figures 3.7 and 5.2, respectively, and the plot of the exact d'Alembert solution spectra, Figure 3.11, a clearer idea about the conduct of the spectra plots can be obtained. The plot at $t = 1$ of the numerical solutions spectra using the ETD1 technique, Figure 5.2b, can be seen to resemble the corresponding plot of the exact solution spectra, Figure 3.11b, better than the numerical solutions spectra using the trapezoidal method, Figure 3.7b, for the C^3 IC (3.83). Figure 5.2b show that the numerical solutions even spectra is zero down to 10^{-10} , then it decreases to round-off error reaching the exact solution even spectra plot in Figure 3.11b at around $N = 2000$. While the plot using the trapezoidal, Figure 3.7b, is zero down to only 10^{-6} . However, the numerical solutions spectra plots in Figures 3.7 and 5.2 at the other two times, at $t = 0.9$ and $t = 4$, are almost similar. In particular, at $t = 0.9$, Figures 3.7a

and 5.2a are almost identical and are similar to the exact solution spectra plot at this time, Figure 3.11a; they all drop off algebraically. One can see that also at $t = 4$ are almost similar; Figure 5.2c drop off exponentially down to 10^{-13} , with no decaying algebraically later, almost similar to the exact spectra plot conduct in Figure 3.11c (in terms of not decaying algebraically later) and the plot of the trapezoidal method, Figure 3.7c, drop off exponentially but down to 10^{-10} , then it decays algebraically.

Furthermore, similar to the previous chapters, the second IC is studied in more detail as shown in Figure 5.3. As one can see that for the C^1 IC (3.82) the odd Chebyshev coefficients decay exponentially and the even Chebyshev coefficients decay algebraically, except at some special times which we explain later, see the corresponding plots for the linear wave equation in Subsection 3.4.4 (Figure 3.8). It is obvious that the spectra of the ETD1 method in Figure 5.3 behaves very similarly as the spectra of the exact d'Alembert solution in Figure 3.12; it mirrors the spectrum of the exact solution more faithfully than the trapezoidal method in Figure 3.8 for the C^1 IC (3.82). Noticeably, we have some specific times, $t = 1, 3, 5, \text{etc.}$, and $t = 2, 4, 6, 8, \text{etc.}$, when the plots are special for the exact and numerical solutions spectra of the linear wave equations; the plots have some special shapes, as explained next. But at other general times the plots decay in the same manner as Figure 3.8. From Figure 5.3, of the C^1 IC (3.82), one can see that at all times before $t = 2$ (except at $t = 1$ as explained next) the plots of the even Chebyshev coefficients of the numerical solution using the ETD1 method decay algebraically. Suddenly, at $t = 2$ they change to decay exponentially not algebraically. In addition, this is the case at $t = 4$ as well. One can see this in Figures 5.3d and 5.3e, at $t = 2$ and $t = 4$, respectively. The same holds when comparing Figures 5.3d and 5.3e with the exact solution spectrum in Figure 3.12c. On the contrary, in Subsection 4.3.4 we see that the Chebyshev coefficients of the numerical solution when using the trapezium rule for the C^1 (3.82) are not very special at these specified times than at other times as the Chebyshev coefficients of the numerical solution decay exponentially only up to the time stepping error $\mathcal{O}(h^2)$, Figures 3.8d and 3.8e, perhaps due to the accumulated error. However, the figure at $t = 1$, Figure 5.3c, is special and different from Figures 5.3d and 5.3e. The even component of the Chebyshev coefficients are zero down to 10^{-10} , and they do not decay algebraically later, but the odd component remain decaying exponentially.

While the plot using the trapezoidal, Figure 3.8c, is zero down to only 10^{-6} then it decreases algebraically. That means, Figure 5.3c resembles the corresponding plot of the exact solution spectra, Figure 3.12b, more faithfully than the numerical solutions spectra using the trapezoidal method, Figure 3.8c. The same figure is obtained at all other odd positive times, $t = 3, 5, 7, \text{etc.}$ The above holds for the plots of the C^3 IC (3.83), Figure 5.2, as well, that figures are similar at each time group shown above. Surprisingly, the even Chebyshev coefficients of the numerical solution using the ETD1 method drop off algebraically again after these special times, for example, we have considered at $t = 6.3$ in Figure 5.3f. See Section 1.9 and 3.4.4 for a better understanding the conduct of the plots at these special times, at $t = 1, 2, 3, 4, 5, 6, \text{etc.}$, which mainly is because the exact d'Alembert solutions are very special at these times.

Obviously, the drop does not occur in the plots of the numerical solution spectra when using exponential integrators for all three ICs, on the contrary to the trapezoidal method which causes to have drop in the plots of the numerical solution spectra for the C^1 IC (3.82) in Figures 3.8 and 3.9, see Subsection 5.3.1 for more information and understanding.

When solving the linear wave equation (3.52) the same figures of the spectra of the numerical solution are obtained for $h = 10^{-3}$ and $h = 10^{-4}$ as for $h = 10^{-2}$, which is considered in the figures, for all three ICs. The reason for that is, as explained above, there is no time stepping error when using exponential integrators for the linear problem.

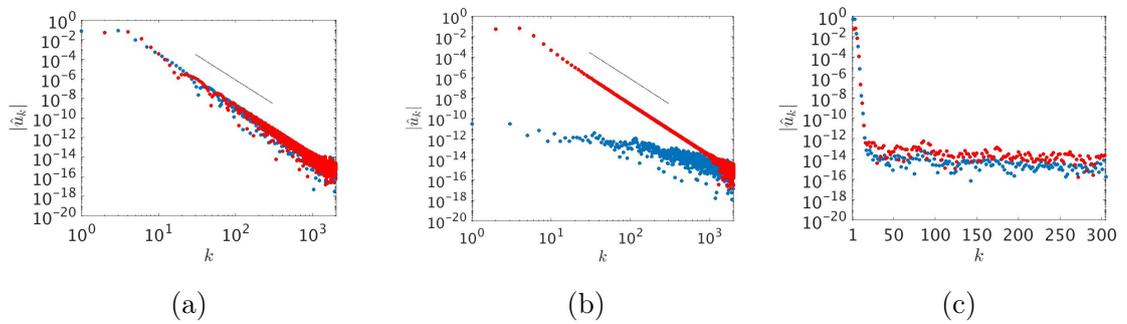


Figure 5.2: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with ETD1 method at $t = 0.9, 1, 4$ (a, b, c panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^3 IC (3.83). The blue dots are the even coefficients component and the red dots are the odd coefficients component. The black straight line indicates order $\frac{1}{k^5}$.

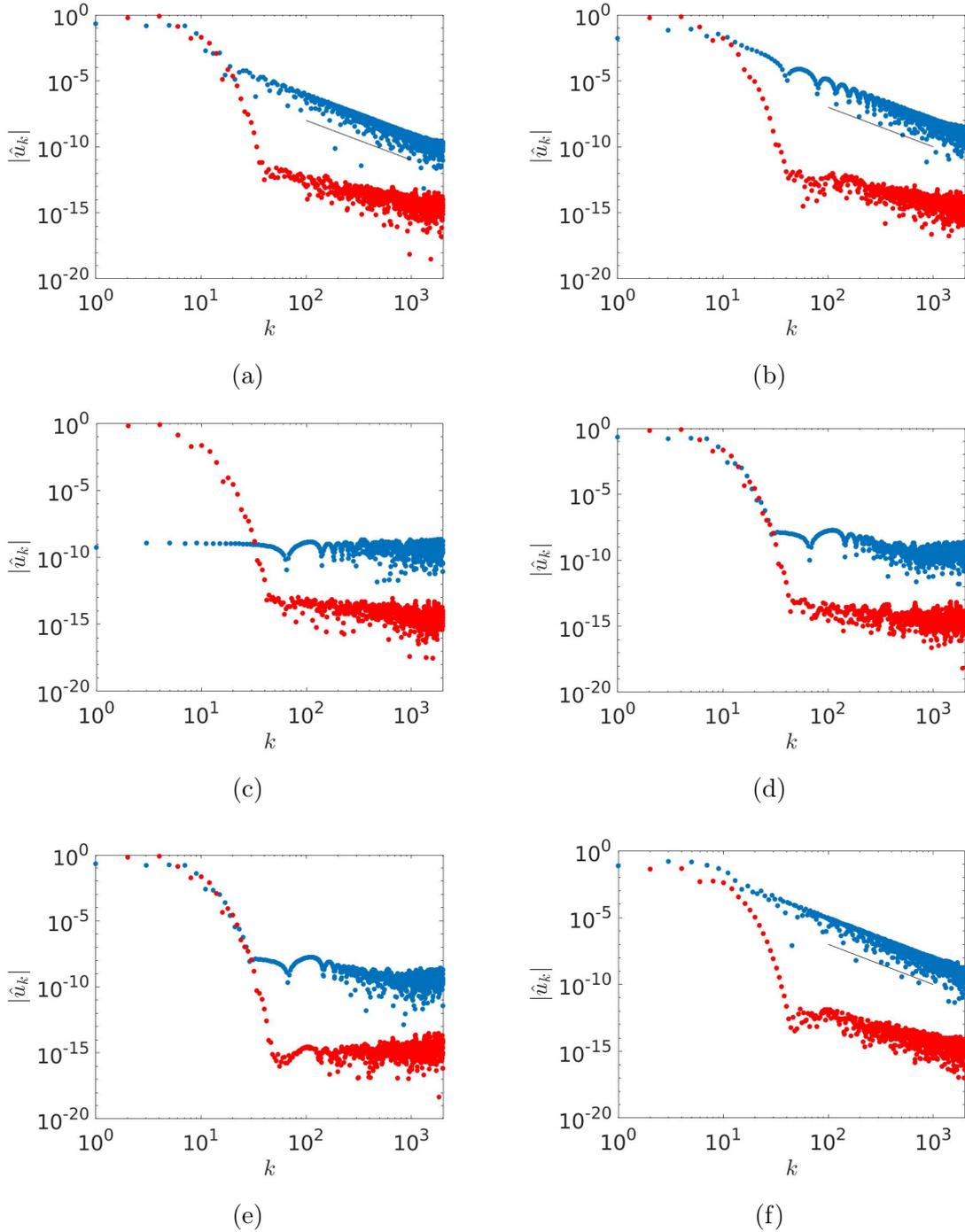


Figure 5.3: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with ETD1 method at $t = 0.04, 0.9, 1, 2, 4, 6.3$ (a, b, c, d, e, f panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^1 IC (3.82). The blue dots are the even coefficients component and the red dots are the odd coefficients component. The black straight line indicates order $\frac{1}{k^3}$.

Further analysis is taken into consideration in Figure 5.4 to show how convergence of numerical methods behaves close to recovery of smoothness, which is primarily at the even times. The figure represents the Chebyshev coefficients of the numerical solution at times near $t = 2$, i.e., $t = 1.9, 1.99, 1.999$ (a, b, c panels respectively). It indicates that the recovery of smoothness happens gradually near these special even times and mainly this occurs for the even component as the odd component is already smooth.

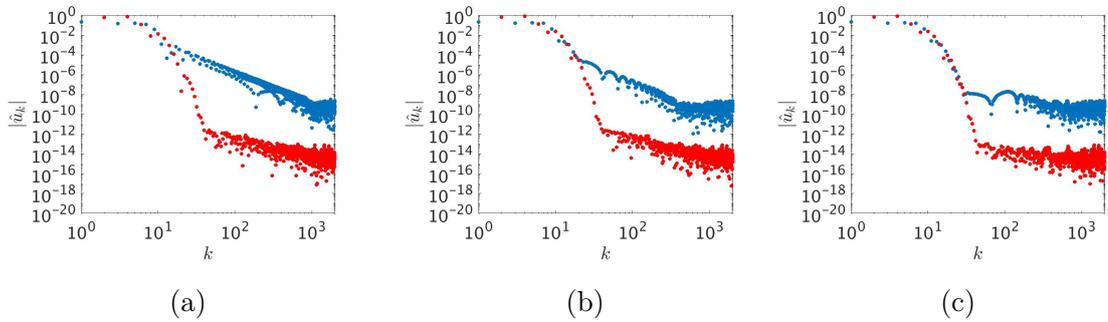


Figure 5.4: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the wave equation (3.52) solved using the pseudospectral method with ETD1 method at times near $t = 2$, i.e., $t = 1.9, 1.99, 1.999$ (a, b, c panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^1 IC (3.82).

In Sections 1.9 and 3.4 we have shown that the exact d'Alembert solutions for the linear wave equation (3.52) at any time which is of multiple 4, i.e., at $t = 0, 4, 8, \text{etc.}$, are infinitely smooth for all three ICs. Thus, we examine the conduct of the numerical solution at these specified times in Subsection 3.4.4 to determine whether they behave the same as the exact solutions or not, especially for the two nonsmooth ICs, the C^3 (3.83) and the C^1 (3.82). Accordingly, we examined the numerical solution with respect to the spectrum using the trapezoidal method combined with pseudospectral method.

Consequently, we do the same in this subsection using exponential integrators. We plot the Chebyshev coefficients of the numerical solution at $t = 4$ using the C^3 (3.83) and the C^1 (3.82) in Figures 5.2 and 5.3, respectively. Amazingly, figures show that indeed the conduct of the numerical solution goes back to be special again at these specific times as well, same as the exact solution behaviour. That is clear in Figures 5.2c and 5.3e, at $t = 4$, the numerical solution spectra of both components change

to decay exponentially not algebraically at these times. In like manner, at the other times of multiple 2, i.e., at $t = 2, 6, 10, \text{etc.}$, the numerical solution spectra of both components change to decay exponentially as well. One can see that clearly in Figure 5.3d, as the numerical solution spectra at $t = 2, 6, 10, \text{etc.}$, possess the same conduct as at $t = 0, 4, 8, \text{etc.}$ Moreover, the exponential integrators combined with pseudospectral method figure at $t = 1$, Figure 5.3c, is special and different from Figures 5.3d and 5.3e. The even coefficients component of the Chebyshev coefficients in Figure 5.3c are zero down to 10^{-10} , and they do not decay algebraically later, and the odd component remains decaying exponentially. In contrast, in Subsection 3.4.4, we see that the Chebyshev coefficients of the numerical solutions, of the trapezoidal method combined with pseudospectral method, especially for the very nonsmooth IC, the C^1 (3.82), are not very special at these specified times, at $t = 0, 1, 2, 3, 4, 5, 6, \text{etc.}$, than at other times. That is, we get accurate results up to the time stepping error $\mathcal{O}(h^2)$ but we still get the algebraic convergence in the plots after reaching $\mathcal{O}(h^2)$, see Subsection 3.4.4 for more information.

In short, it is obvious that the Chebyshev coefficients of the ETD1 method for solving the linear problem in Figure 5.2 and Figure 5.3, for the C^3 IC (3.83) and the C^1 IC (3.82), respectively, are almost similar to the corresponding plots of the exact solution spectrum in Figure 3.11 and 3.12, respectively, at the special times $t = 0, 1, 2, 3, 4, 5, 6, \text{etc.}$ Knowing that the exact d'Alembert solution of the linear wave equation is smooth at $t = 0, 2, 4, 6, 8, \text{etc.}$, and it is an odd function at $t = 1, 3, 5, \text{etc.}$, and seeing the numerical results that both components of the Chebyshev coefficients of the numerical solution drop off exponentially at $t = 0, 2, 4, 6, 8, \text{etc.}$, and the even component of the Chebyshev coefficients at $t = 1, 3, 5, \text{etc.}$, are zero down to 10^{-10} , and they do not decay algebraically later, and the odd coefficients component decays exponentially. Immediately after these times the even component of the Chebyshev coefficients drop of algebraically again, for instance, see Figure 5.3f at $t = 6.3$. That strongly says that the numerical solution recover smoothness of the solution and the difference of the odd/even of the solution at the specified times described above very accurately and faithfully.

5.2.2 Symplecticity and stability of the ETD1 method

Consequently from the subsection 5.2.1, since the ETD methods give the exact solution for the linear problem, this leads to have the energy to be exactly as the energy of the semi-discretized system, ODEs (3.71) and (3.72), constant up to an error. The energy is not affected by the time stepping error but it is still affected by the spatial error; N must be chosen accordingly. More precisely, the energy is preserved up to the round-off error 10^{-12} for any $N \geq 20$ for the C^∞ IC (3.81), and for the other two ICs, C^1 IC (3.82) and C^3 IC (3.83), it is preserved up to an error 10^{-7} for $N = 1000$ and up to an error 10^{-9} for $N = 100$, respectively. The energy will eventually be preserved for the last two ICs up to the round-off error of 10^{-12} by increasing N . However, the energy will almost be preserved for all three ICs when N decreases, using smaller values mentioned above. The reason for this is because ETD is exact for linear equations, so the energy for ETD is the same as the energy for the semi-discretized solution. The spectral discretization introduces a spatial error. If N is the above values, $N = 20$ for the C^∞ IC (3.81), $N = 1000$ for the C^1 IC (3.82) and $N = 100$ for the C^3 IC (3.83), then this spatial error means that the energy in the semi-discretized solution oscillates and the amplitude of these oscillations decrease with N , but after these values of N the amplitude is so small that one only sees the round-off error. The above holds for any implemented value of h , such as for $h = 10^{-2}$, $h = 10^{-3}$, or $h = 10^{-4}$, as there is no time stepping error. Accordingly, the ETD methods are stable because the numerical solution will be bounded not explode as the energy is conserved.

5.3 Solving the nonlinear wave equation

The nonlinear wave equation (4.1):

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3,$$

is discretized to ODEs (4.31) and (4.32). This system of ODEs can be written as

$$\vec{u}' = L\vec{u} + \vec{\Gamma}, \quad (5.11)$$

see (5.1), where here we have

$$\vec{u} = \begin{bmatrix} \vec{y} \\ \vec{w} \end{bmatrix}, L = \begin{bmatrix} 0 & I \\ D^2 & 0 \end{bmatrix}, \vec{\Gamma} = \begin{bmatrix} 0 \\ -\vec{y}^3 \end{bmatrix}. \quad (5.12)$$

In this section we study the convergence of N and h , the spectra of the numerical solution, the stability and energy conservation of the exponential time difference methods, the ETD1 and ETD2RK methods, when used to solve the above nonlinear problem. We are implementing the same three initial conditions used in previous chapters, (3.81), (3.82) and (3.83), in our numerical experiments. In order to find the estimated error we use the reference solutions obtained in Chapter 4 using the trapezoidal method for solving the nonlinear wave equation (4.1), see Subsection 4.3.1 for more information.

5.3.1 Convergence of the ETD1 and ETD2RK methods

The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, is plotted against the truncation index N in Figure 5.5 using the ETD2RK method for the nonlinear wave equation (4.1), for the C^∞ IC (3.81), C^1 IC (3.82) and C^3 IC (3.83). Additionally, the error against the inverse step size $1/h$ for the ETD1 method (left panel) and the ETD2RK (right panel) is plotted in Figure 5.6 for the C^∞ IC (3.81). It is easily observed in Figures 5.5 and 5.6 that both methods, ETD1 and ETD2RK, are convergent. Figure 5.5 shows that ETD2RK converges exponentially for the IC (3.81); the other two ICs, (3.82) and (3.83), are studied as well and the results, again, show that both ICs converge algebraically with order seems to be $\frac{1}{N^3}$ and $\frac{1}{N^5}$, respectively. Comparing this figure with the related figure, 4.3, in Chapter 4, one can see that this method behaves similarly to the trapezoidal rule. Similar results are obtained for ETD1.

Regarding the time step error, h , the left panel of 5.6 gives that the ETD1 method is of order one. Meanwhile, the ETD2RK method is of order two as shown in the right panel. The other two ICs also have the same orders for both of the methods.

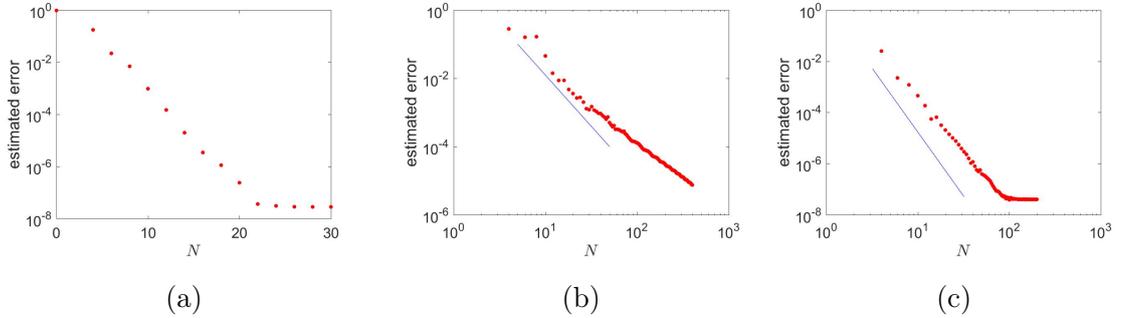


Figure 5.5: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the truncation index N for the nonlinear wave equation (4.1), for the C^∞ IC (3.81), C^1 IC (3.82) and C^3 IC (3.83) (a, b and c respectively), solved using the pseudospectral method with the ETD2RK method. The estimated error is evaluated at time step $h = 10^{-3}$ and $t = 0.9$. The blue straight line shows order $\frac{1}{N^3}$ for the C^1 IC (3.82) and order $\frac{1}{N^5}$ for the C^3 IC (3.83). Here u_{num} is the numerical solution and for u we take the reference solution in Subsection 4.3.1.

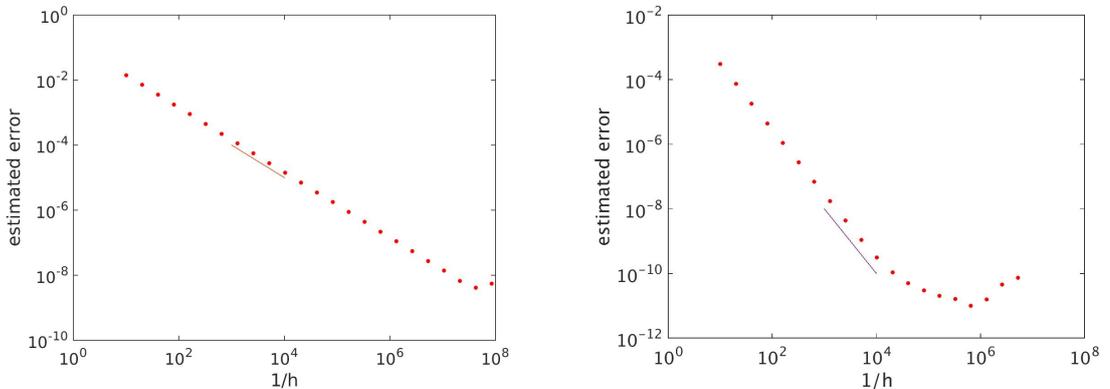


Figure 5.6: The errors against the inverse step size $1/h$ for the nonlinear wave equation (4.1) solved using the pseudospectral method C^∞ IC (3.81). The left panel with the ETD1 method and the right panel with the ETD2RK method. The estimated error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, is evaluated at $t = 0.9$ and $N = 100$. The blue straight line shows order one for the ETD1 method and order two for the ETD2RK method.

As mentioned in Subsection 5.2.1, the same structure of the previous two chapters is

followed in this chapter as well in terms of investigating the Chebyshev coefficients of the ETD1 and ETD2RK methods combined with pseudospectral method for the nonlinear wave equation (4.1) with all three ICs implemented.

The Chebyshev coefficients for the numerical solution using the ETD1 and ETD2RK methods when solving the nonlinear wave equation using the C^∞ IC (3.81), the C^1 IC (3.82) and the C^3 IC (3.83) are examined. The results show that the ETD1 method has the same Chebyshev coefficients results conduct as for the ETD2RK method for the times used in the figures, in terms of the exponential decay versus the algebraic decay behaviours. The Chebyshev coefficients using ETD2RK for the three ICs are studied thoroughly next.

The Chebyshev coefficients of the C^∞ IC (3.81) decay exponentially similar to the previous chapters and Subsection 5.2.1; particularly the plot is similar to the Chebyshev coefficients plot of the numerical solution of the nonlinear problem in Figure 4.4.

In addition, the Chebyshev coefficients of the numerical solution using the C^3 IC (3.83) is studied in Figure 5.7. The three Figures of the third IC (3.83), the numerical solutions spectra plots for the nonlinear wave equation (4.1) of using the trapezoid and ETD2RK methods combined with pseudospectral method, Figures 4.5 and 5.7, respectively, and the plot of the exact d'Alembert solution spectra for the linear wave equation, Figure 3.11, are compared in order to get a more transparent view about the conduct of the spectra plots. We use the exact solution plots of the linear problem just to help us understand our results of the nonlinear problem better. It can be seen that the ETD2RK technique plots are close to the trapezoidal method plots. In particular, at $t = 0.9$ are almost identical and are similar to the exact plot at this time; they all drop off algebraically, like Subsection 5.2.1. One can see that also at $t = 4$ are almost similar; Figure 5.7c drop off exponentially down to 10^{-12} , almost similar behaviour as of the exact spectra plot in Figure 3.11c, also like Subsection 5.2.1, and Figure 4.5c drop off exponentially but down to 10^{-10} then it decays algebraically. While both methods, the trapezoid and the ETD2RK, have different even spectra component plots of the numerical solutions at $t = 1$, Figures 4.5b and 5.7b, respectively, than the exact solution spectra, Figure 3.11b as stated in Subsection 4.3.4 that the nonlinear problem does not have a special conduct at $t = 1, 3, 5, etc.$

Furthermore, C^1 IC (3.82) is investigated in more detail as shown in Figure 5.8.

It is obvious the figure shows that the odd Chebyshev coefficients and the even Chebyshev coefficients of the numerical solution decay algebraically again, except at some times which we explain later; see the corresponding plots for the nonlinear wave equation using the trapezoidal method, Figure 4.6, in Subsection 4.3.4. It is noticeable that we have some specific times in this chapter, $t = 2, 4, 6, 8, \text{etc.}$, when the plots are special for the numerical solution spectra of the linear and nonlinear wave equations. The Chebyshev coefficients of the ETD2RK method in Figure 5.8 are almost similar to the corresponding plots for the linear wave equation using the ETD1 method in Figure 5.3 at these specified time. In other words, both Chebyshev coefficients components of the numerical solution at $t = 2, 4, 6, 8, \text{etc.}$, decay exponentially not algebraically for the C^1 IC (3.82) but at all other time the plots decay algebraically. That is clear in Figures 5.8d and 5.8e at $t = 2$ and $t = 4$, respectively. Surprisingly, after these special times the plots drop off algebraically again, see Figure 5.8f at $t = 6.3$. The same holds when compare Figures 5.8d and 5.8e with the exact solution spectrum of the linear problem in Figure 3.12c, that all plots at these times decay exponentially with no algebraic decay later. In contrast, in Subsection 4.3.4 the Chebyshev coefficients of the numerical solution when using the trapezium rule for the C^1 (3.82) are not very special at these specified times than at other times as the Chebyshev coefficients of the numerical solution decay exponentially only up to the time stepping error $\mathcal{O}(h^2)$, however, they decay algebraically again after that, see figures 4.6d and 4.6e, probably because of the accumulated error.

One of the general differences between the plots of the spectra of exponential integrator combined with pseudospectral method for the linear problem in Figure 5.3 and for the nonlinear problem in Figure 5.8 is that it seems the nonlinearity in equation leads to have some scatter in the plots of odd spectra components for the nonlinear equation plots, especially in the plots at $t = 2, 4, 6, 8, \text{etc.}$

Noteworthy, throughout this chapter, for the linear and the nonlinear problems, when employing exponential integrators the numerical solution spectra plots do not drop for all three ICs, unlike when using the trapezoidal technique in the previous chapters, which results in a drop in the spectra plots for the C^1 IC (3.82) for the linear problem, see Figures 3.8 and 3.9, and for the nonlinear problem, see Figure 4.6. This is more clear in the spectra plots at $t = 0.04$; in comparing Figures of the exponential integrators, Figure 5.3a (of the linear problem) and 5.8a (of the

nonlinear problem) with the corresponding plots of the trapezoidal method, Figure 3.8a (of the linear problem) and 4.6a (of the nonlinear problem). The step size used throughout this chapter for the plots of the numerical solution spectra is $h = 10^{-2}$, while in the previous chapters, Subsections 3.4.4 and 4.3.4, the step size is smaller which is $h = 10^{-3}$, see Subsection 3.4.4 and Figure 3.9 for more understanding about the impacts of the step size.

From our numerical results of the numerical solutions spectra in Subsections 3.4.4, 4.3.4, 5.2.1 and in this Subsection 5.3.1, we can see and understand that the fast drop in the plots of the C^1 IC (3.82) in Figures 3.8, 3.9 and 4.6 is caused by the time stepping method, i.e., it is caused mainly by the trapezoidal method. In addition, we have tested and found that the ETD1 method for the linear problem is having the same results conduct for up to $t = 1500$. Also, the ETD2RK method for the nonlinear problem is having the same results conduct for up to $t = 1500$, in terms of the exponential decay versus the algebraic decay behaviours.

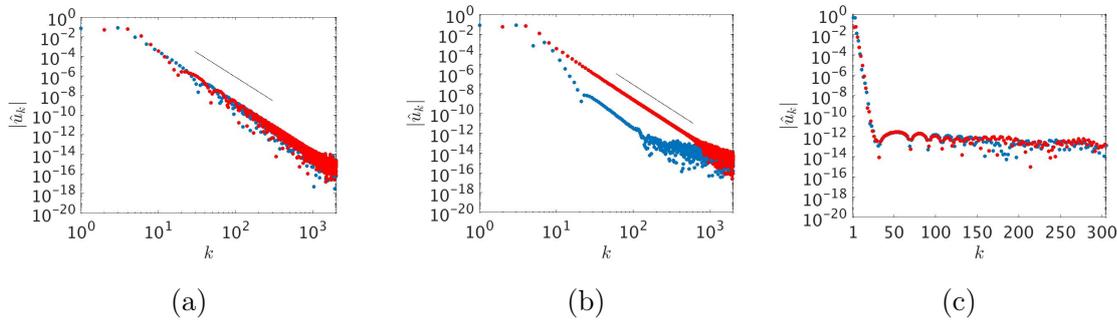


Figure 5.7: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD2RK method at $t = 0.9, 1, 4$ (a, b, c panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^3 IC (3.83). The black straight line indicates order $\frac{1}{k^5}$.

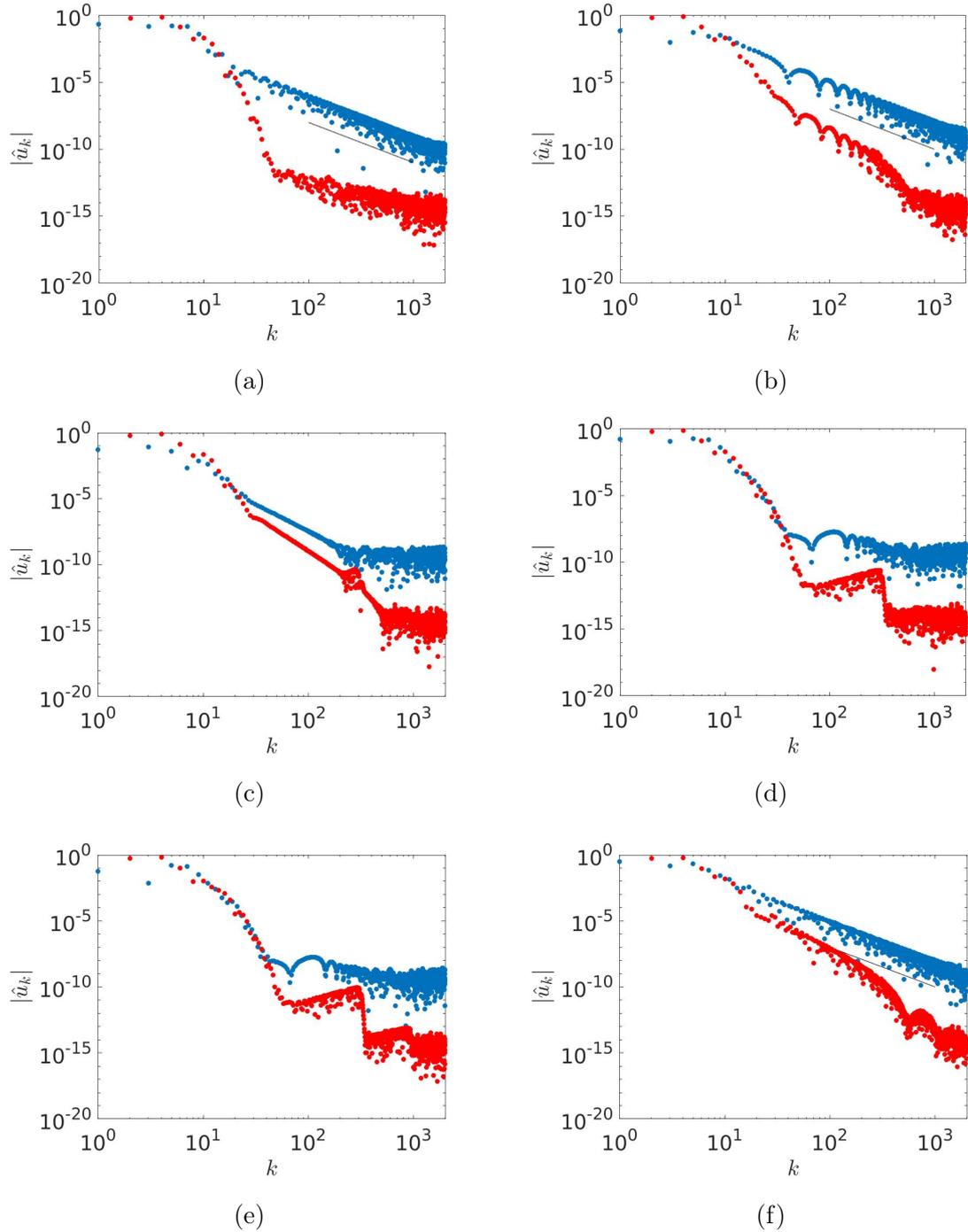


Figure 5.8: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD2RK method at $t = 0.04, 0.9, 1, 2, 4, 6.3$ (a, b, c, d, e, f panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^1 IC (3.82). The black straight line indicates order $\frac{1}{k^3}$.

In order to illustrate how convergence of numerical methods behaves near to recovery of smoothness, which is mainly at the even times, further investigation is considered in Figure 5.9. Similar to the linear problem in Figure 5.4, Figure 5.9 shows that the recovery of smoothness happens gradually near these special even times, however, for the nonlinear problem this occurs for the both components not only the even component as they are both nonsmooth.

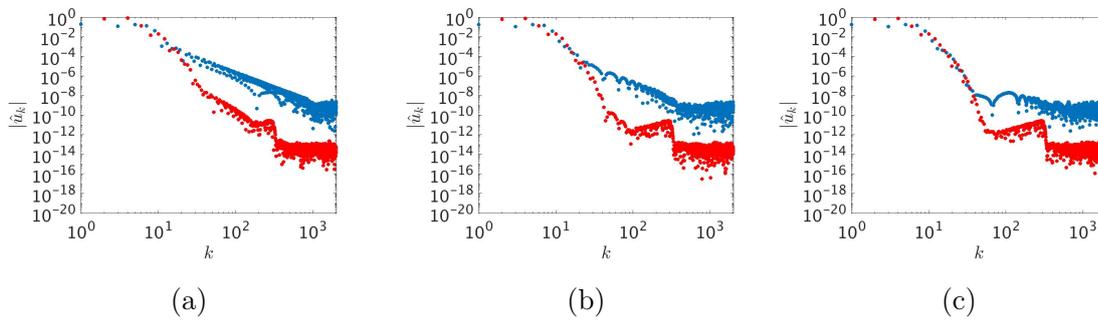


Figure 5.9: The Chebyshev coefficients $|\hat{u}_k|$ of the numerical solution against the mode k for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD2RK method at times near $t = 2$, i.e., $t = 1.9, 1.99, 1.999$ (a, b, c panels respectively). The step size is $h = 10^{-2}$ with $N = 2000$ for the C^1 IC (3.82).

In Subsection 4.3.4 we said that we do not know the exact solution for the nonlinear wave equation (4.1), however, the decay of the Chebyshev coefficients of the numerical solution in figures throughout the chapters, in Subsections 4.3.4 and 5.3.1, strongly suggest that the exact solution is smooth for the C^∞ IC (3.81), but it is only C^1 for the second IC, (3.82), and it is C^3 for the third IC, (3.83). However, the nonlinear wave equation share the same characteristics with the linear equation; they both have the same characteristics Figure 1.5. Therefore, based on the characteristics fact, we suspected that the exact solutions for the nonlinear wave equation to be smooth at $t = 0, 4, 8, \text{etc.}$, for all three ICs as well.

Astonishingly, our numerical results for solving the nonlinear wave equation using exponential integrators combined with pseudospectral method in this chapter, in this subsection, Figures 5.7 and 5.8 of the C^3 IC (3.83) and the C^1 IC (3.82) respectively, confirm our assumption in Subsection 4.3.4 that: indeed the exact solution of the nonlinear equation goes back to be special again at these specific

times. That is clear in Figures 5.7 and 5.8 and comparing the figures with the corresponding figures of the numerical solution spectra of the linear wave equation (3.52), Figures 5.2 and 5.3 of the C^3 IC (3.83) and the C^1 IC (3.82) respectively, when the numerical solution spectra of both components decay exponentially again not algebraically at these times, since we know the exact d'Alembert solution of the linear wave equation is indeed smooth at these specified times and the numerical spectra results mirror the exact solution behaviour of the linear equation thus the components drop off exponentially. This confirms our results and assumption. In other words, the exact solution of the nonlinear wave equation (4.1) becomes infinitely smooth again at $t = 0, 4, 8, \text{etc.}$, since the numerical solution spectra of both components decay exponentially again not algebraically at these times, similar to the exact d'Alembert solution of the linear wave equation (3.52). In the same way, at the other times of multiple 2, i.e., at $t = 2, 6, 10, \text{etc.}$, the exact d'Alembert solution of the nonlinear wave equation (4.1) is infinitely differentiable as well, infinitely smooth, since the characteristics touch the boundaries at these specific times, meaning that the discontinuities go away again then, see 1.9 and 3.4 for a better understanding. One can see that clearly in Figure 5.8d, as the numerical solution spectra at $t = 2, 6, 10, \text{etc.}$, possess the same conduct as at $t = 0, 4, 8, \text{etc.}$ On the contrary, in Subsection 4.3.4, we see that the numerical solution, of the trapezoidal method combined with pseudospectral method, show the exponential convergence only to a limited extent in the plots but not fully, especially for the very nonsmooth IC, the C^1 (3.82) at the specified times above. Thus, one can see that the trapezoidal method combined with pseudospectral method attempts to mirror the behaviour of the exact solution, however, it is able to reflect it only partially to a limited extent not faithfully as exponential integrators combined with pseudospectral method.

According to our numerical results of the nonlinear wave equation (4.1) in Subsection 4.3.4 and this Subsection 5.3.1, one concludes that at the specified times described above the spectrum behaviour of the numerical solution using exponential integrators mirror the exact solution coefficients behaviour (recover smoothness of the solution) more accurately and faithfully than using the trapezoidal method, for both ICs (3.82) and (3.83). In the same manner, from Subsections 3.4.4 and 5.2.1 we see that this is the case for the numerical solution spectrum conduct of the linear problem as well. Therefore, in the plots of the spectra for exponential

integrators intentionally we have considered a time step that is bigger than the time step in the previous two chapters of the trapezoidal method to demonstrate how well exponential integrators perform in comparison to the trapezoidal technique at the special times; we have chosen $h = 10^{-2}$ here, meanwhile, in the other chapters it is $h = 10^{-3}$.

5.3.2 Stability of the ETD1 and ETD2RK methods

The numerical solutions for the nonlinear wave equation (4.1), at $x = \frac{1}{2}$, against time are plotted in Figure 5.10, using ETD1 method (left panel) and ETD2RK (right panel) applied for the C^1 IC (3.82). The figure shows that the ETD1 method is not stable for $h = 10^{-2}$ used in the figure, because the numerical solution blows up. However, the ETD2RK is stable, because the solution is bounded for the time used in the figure; the time is very big $t = 6000$ but still the overall amplitude of solution remains a constant; it does not decay to zero. And also, as we will see next, because the energy error decreases means that the numerical solution will decay to zero, but in extremely long time. Regarding the ETD1 method, if we compare the left panel of the figure with the related figure for Euler's method in Chapter 4, Figure 4.7, we see that Euler's method starts to be unstable at $t < 6$. Meanwhile ETD1 becomes unstable after a longer time, around $t \sim 200$, additionally, rather than exploding rapidly like Euler's approach, this method gradually rises, for the times used here. The reason is that, for the Euler's method, the numerical solution is linearly unstable (the eigenvalues of D^2 are outside the linear stability domain). For ETD1, linear stability is marginal (the linear wave equation is stable with ETD1 and only the nonlinear term makes it unstable). With respect to the ETD2RK method, comparing the right panel of Figure 5.10 with Figure 4.9 of the trapezoidal rule in Chapter 4, we see that they have similar solution amplitude, although time used there is much smaller $t = 400$ and here is $t = 6000$. It is worth mentioning that Figure 5.10 is only for the IC (3.82): the other two ICs behave similarly.

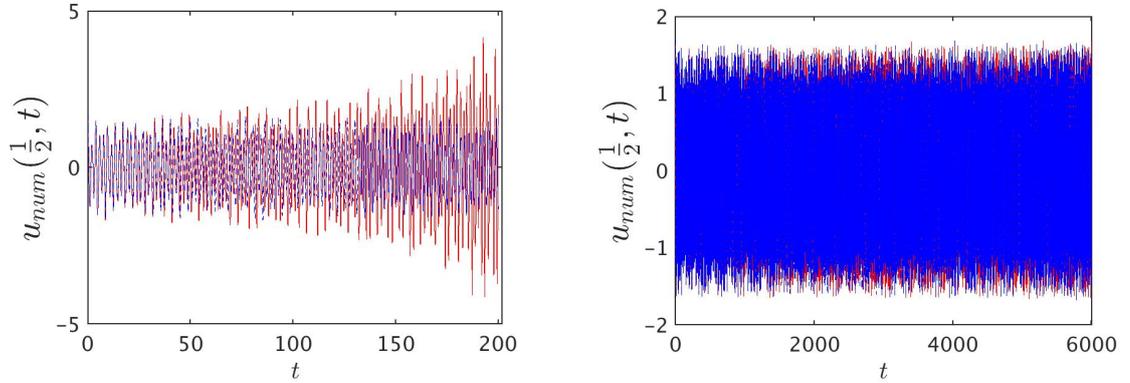


Figure 5.10: The numerical solutions $u_{num}(x, t)$ for the nonlinear wave equation (4.1), at $x = \frac{1}{2}$, for $N = 10$ using two time steps, $h = 10^{-2}$ (red line), $h = 10^{-3}$ (dashed blue line). The solutions are computed using ETD1 method (left panel) and ETD2RK (right panel), applied for the C^1 IC (3.82).

5.3.3 Symplecticity of the ETD1 and ETD2RK methods

Theoretically, ETD1 and ETD2RK are not symplectic because when $L = 0$ then (5.5) and (5.8) will be the same as Euler's and the modified Euler methods, respectively, and those methods are not symplectic theoretically, see Section 2.5. Our numerical results in this section confirm this.

The energy error, $E - E_{exact}$, against time is plotted in Figures 5.11 and 5.12 for the nonlinear wave equation (4.1) using ETD1 and ETD2RK, respectively, for the C^1 IC (3.82). One can easily observe that both methods, ETD1 and ETD2RK, have some energy drift, for ETD1 the energy is increasing and for ETD2RK it is decreasing, but slowly. Comparable behaviour is achieved for the other two ICs, (3.82) and (3.83), for ETD1 using the same indices, N and nx , used in Figure 5.11 and for ETD2RK using the same indices required in previous chapter for the trapezoidal method.

It is clear that for all regimes the error scales as expected for the first and second order methods, i.e., it is increasing or decreasing of order one or two for ETD1 and ETD2RK respectively, for all the values of h 's used in the figures, $h = 10^{-2}$, $h = 10^{-3}$ and $h = 10^{-4}$. Overall, the energy error is changing gradually not quickly.

One can compare the energy error for these methods with Euler's and the trapezoidal methods in Chapter 4; the energy error for Euler's method explodes in a very short time, see Figure 4.12, this is because Euler's method is linearly unstable. Consequently, the ETD1 method has a better energy than Euler method. The trapezoidal approach, however, almost preserves energy, see Figure 4.13 -4.15. On the other hand, the ETD2RK approach drifts it rather than preserving it, see Figure 5.12. However, the drift is very slow, and one would have to integrate for a very long time before the energy had decreased significantly from its initial value. Furthermore, once it had decreased significantly, the system would move closer to the linear problem where the ETD methods become exact. Measured on extremely long time scales the energy decrease would then be slower than exponential decay. For relatively short times as in Figure 5.12 this effect is not visible though. One can see that the energy accuracy for the trapezoidal and ETD2RK methods is similar and the computational time is close. For example, for the C^1 IC (3.82) used here, the energy error is approximately 10^{-4} for $h = 10^{-2}$, for $h = 10^{-3}$ it is around 10^{-6} and for $h = 10^{-4}$ it is 10^{-8} , see Figures 4.14 and 5.12.

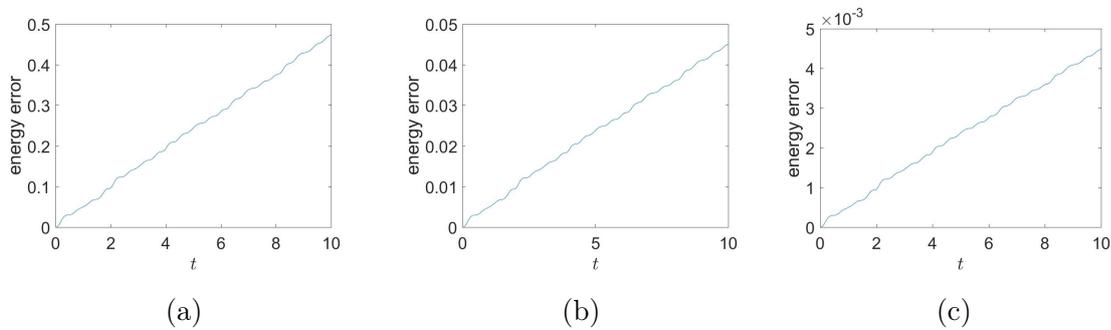


Figure 5.11: The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD1 for $N = 100$, $nx = 100N$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a,b, and c respectively) and initial condition (3.82). Here E is the numerical energy and E_{exact} is the exact energy.

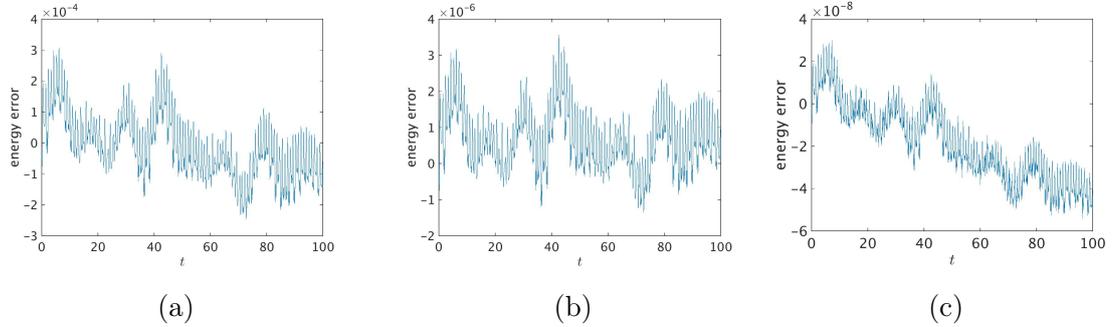


Figure 5.12: The energy error, $E - E_{exact}$, for the nonlinear wave equation (4.1) solved using the pseudospectral method with ETD2RK, for $N = 1000$, $nx = 100N$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a,b, and c respectively) and initial condition (3.82).

5.4 Summary and discussion

Exponential integrators are a type of numerical method for solving ordinary differential equations, particularly initial value problems. They are based on the exact integration of the linear part of the initial value problem.

In this chapter some other time stepping methods, namely the exponential integrators, ETD1 and ETD2RK, are used for solving the linear wave equation (3.52), $\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}$, and the nonlinear wave equation (4.1), $\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3$, with homogeneous Dirichlet boundary conditions.

Similar to earlier chapters, this chapter studies the convergence in space and time, stability and energy conservation of the numerical methods. In addition, the spectra of the solutions, of the exact solution, where available, and of the numerical solution, are thoroughly investigated. The spatial convergence for ETD1 and ETD2RK is similar as the previous time stepping methods in earlier chapters. The solution converges exponentially for the first IC, (3.81), but for the other ICs, (3.82) and (3.83), it converges only algebraically, with order seems to be $\frac{1}{N^3}$, for (3.82), and order $\frac{1}{N^5}$, for (3.83), for both of the PDEs.

Similar to Chapter 3 and 4, the convergence using the spectrum of the numerical solution of the combination of exponential integrators with pseudospectral method for the linear and nonlinear wave equations is thoroughly investigated. In comparison to these chapters, our numerical outcomes here are more precise and

accurate. This is made more obvious by the numerical solution spectra for the linear problem, which demonstrate a strong likeness to the exact solution spectra, particularly at some specified times.

In the same manner as the prior chapters, the special times for the linear problem are $t = 0, 1, 2, 3, 4, 5, 6, \text{etc.}$, and for the nonlinear problem are at $t = 0, 2, 4, 6, \text{etc.}$ The exact d'Alembert solution of the linear problem is infinitely smooth at $t = 0, 2, 4, 6, \text{etc.}$, **novel result** from Chapter 1. The exact solution of the nonlinear wave equation is unknown, however, in Chapter 4 we assumed the exact solution to be infinitely smooth too at $t = 0, 2, 4, 6, 8, \text{etc.}$, for all three ICs. Astonishingly, our results show that exponential integrators recover the smoothness of the numerical solution at $t = 0, 2, 4, 6, \text{etc.}$, for both equations, the linear and nonlinear wave equations. In other words, both components of the Chebyshev coefficients of the numerical solution drop off exponentially at these specified times. Wherefore, our assumption in Subsection 4.3.4 is confirmed and verified. In addition, for the linear problem, the numerical solution spectra at $t = 1, 3, 5, \text{etc.}$, the exact solution is an odd function, **novel result** from Chapter 1, thus, the even Chebyshev coefficients of the numerical solution for the the C^1 (3.82) are zero down to 10^{-10} , and they do not decay algebraically later, and also for the C^3 IC (3.83). Immediately after these special times the even component of the Chebyshev coefficients drop of algebraically again as at the other ordinary times. On the contrary, in Subsection 3.4.4, we see that the Chebyshev coefficients of the numerical solutions, of the trapezoidal method combined with pseudospectral method, especially for the very nonsmooth IC, the C^1 (3.82), are not very special at these specified times. That is, we get accurate results up to the time stepping error $\mathcal{O}(h^2)$ but we still get the algebraical convergence in the plots after reaching $\mathcal{O}(h^2)$, probably because of the accumulated error. As opposed to earlier chapters, none of the ICs experience the fast drop in here, indicating that the trapezoidal method from the earlier chapter is to be blamed.

According to our numerical results of the spectra for the linear and nonlinear wave equations in this chapter one observes that at the specified times described above exponential integrators recover the smoothness of the solutions and the difference of the even and odd coefficients of the solution (when exists) very precisely and accurately. Thus, one concludes that the coefficients behaviour of the numerical solution using exponential integrators, at the specified times described above, mirror

the exact solution coefficients behaviour more accurately and faithfully than using the trapezoidal method for both equations, for the C^1 (3.82) and the C^3 IC (3.83).

As for the time stepping error, the ETD method is of order one, like the explicit and implicit Euler methods, and ETD2RK is of order two, like the trapezoidal and modified Euler methods. However, there is no time stepping error for the linear wave equation, because the ETD method solve the resulting linear ODEs exactly. For the same reason, the exponential methods for the linear wave equation have the energy exactly as the energy of the semi-discretized system, ODEs (3.71) and (3.72), constant up to an error. More precisely, the energy is preserved up to the round-off error 10^{-12} for any $N \geq 20$ for the C^∞ IC (3.81), and for the other two ICs, C^1 IC (3.82) and C^3 IC (3.83), it is preserved up to an error 10^{-7} for $N = 1000$ and up to an error 10^{-9} for $N = 100$, respectively. The energy will eventually be preserved for the last two ICs up to the round-off error of 10^{-12} by increasing N . Thus, the ETD methods are stable because the numerical solution will be bounded not explode as the energy is conserved.

We now discuss energy conservation and the stability of the methods for the nonlinear wave equation. The ETD1 do not preserve energy but show a drift in energy, but it increases gradually. Thus, the ETD1 method is not stable but the instability is less severe as for Euler's method as its effects occur after a long time compared to Euler's method, furthermore, for the times employed here, this method steadily rises rather than rapidly expanding like Euler's method.

A crucial fact to know is that the exact solution of the nonlinear problem has these different properties, which are unrelated:

- The property that the energy is conserved, **known result** from chapter 4. So, the overall amplitude solution remains a constant; it does not decay to zero, which has nothing to do with the smoothness of the solution at special times.
- The property related to the smoothness, that the solution is infinitely smooth at special times, **novel result** from this chapter, which has nothing to do with the overall amplitude of the solution.

Our results in this chapter show that the ETD2RK method does not preserve energy but show a drift in energy, it decreases. However, the drift is very slow

because of the property that: the smaller the amplitude is the closer gets to the linear problem, thus, in fact it will drop off extremely slow, even slower than exponential decay. Therefore, ETD2RK is stable for the nonlinear wave equation. The trapezoidal approach in Chapter 4, however, almost preserves energy. One can see that the energy accuracy for the trapezoidal and ETD2RK methods is similar and the computational time is close. That implies: The decrease in energy for ETD2RK means the overall amplitude of the numerical solution will decay to zero, but obviously on an extremely long time scale. Meanwhile, the overall solution amplitude of the trapezoidal method remains a constant; it does not decay to zero. That means, the trapezoidal method simulates the overall solution amplitude structure of the nonlinear wave equation more faithfully than ETD2RK method. Whilst, according to our numerical results of the spectra we have shown that the EDT2RK method recovers the smoothness of the solution at the special times for the nonlinear wave equations far more accurately and faithfully than the trapezoidal method.

Chapter 6

Finite element methods

The finite element methods (FEM) are numerical methods that provide a general solution technique for partial differential equations with boundary conditions for a very wide range of applicability (Van der Vegt & Bokhove, 2009). As briefly introduced in Chapter 1, the FEM are a general and systematic computational procedure for approximately solving problems in physics and engineering. This chapter introduces the **weak statement** as the mathematical foundation for development of the FEM.

The concept of the FEM is to divide the domain into finitely many smaller areas called elements and approximate the solution of the differential equation on these elements using a suitable set of basis functions. The result is a system of equations which can either be solved directly or by another numerical method (Boyd, 2001). The nonlinear wave equation (4.1) that is considered in the previous chapters is considered in this chapter as well, which is

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3,$$

for the interval $(-1, 1)$ and $t \geq 0$. This equation is solved subject to Dirichlet boundary conditions (4.2),

$$u(-1, t) = 0, \quad \text{and} \quad u(1, t) = 0,$$

and initial conditions (4.3) and (4.4),

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

and

$$\frac{\partial u}{\partial t}(x, 0) = 0.$$

6.1 Finite element methods

In the finite element methods (FEM) the solution domain is discretized by dividing it into non-overlapping elements $e^{(i)}$ of arbitrary shape and size. Within each element a certain number of nodes are defined at which the unknown nodal values are to be determined. Unlike the spectral methods, which employ global basis functions, the FEM use basis functions that are locally defined polynomials within each element $e^{(i)}$ and zero outside a neighbourhood of $e^{(i)}$. The nodal values are used to define a finite linear combination of local basis functions which approximates the exact solution.

6.1.1 Linear basis functions

Here, we describe the simplest variant of the FEM, which uses linear basis functions. The domain, which is the interval $[-1, 1]$ in our problem, is divided into M subintervals separated by nodes, where M is an integer number. The nodes x_i with $i = 0, \dots, M$ are chosen such that $-1 = x_0 < x_1 < \dots < x_{M-1} < x_M = 1$ and the elements are defined as intervals

$$e^{(i)} := [x_i, x_{i+1}], \quad \text{for } i = 0, \dots, M - 1,$$

and the size of elements $e^{(i)}$ is given by Δ_{x_i} or more specifically

$$\Delta_{x_i} := x_{i+1} - x_i.$$

The linear basis function are considered in this study, which can be written as

$$\phi_i(x) = \begin{cases} \frac{x-x_{i-1}}{x_i-x_{i-1}}, & \text{if } x_{i-1} \leq x \leq x_i, \\ \frac{x_{i+1}-x}{x_{i+1}-x_i}, & \text{if } x_i \leq x \leq x_{i+1}, \\ 0, & \text{otherwise.} \end{cases} \quad \text{for } j = 1, \dots, M - 1, \quad (6.1)$$

The two remaining basis functions, with $i = 0$ or $i = M$, lie near the boundary of the interval $[-1, 1]$ and are defined such that they are non-zero on the elements $e^{(0)}$

and $e^{(M-1)}$ respectively and zero everywhere else. These functions can be written as

$$\phi_0(x) = \begin{cases} \frac{x_1-x}{x_1-x_0}, & \text{if } x_0 \leq x \leq x_1, \\ 0, & \text{otherwise,} \end{cases} \quad (6.2)$$

and

$$\phi_M(x) = \begin{cases} \frac{x-x_{M-1}}{x_M-x_{M-1}} & \text{if } x_{M-1} \leq x \leq x_M, \\ 0, & \text{otherwise.} \end{cases} \quad (6.3)$$

Using the just defined basis functions one can define the semi-discretization finite element solution as follows,

$$u_{sd}(x, t) = \sum_{i=0}^M u_i(t) \phi_i(x), \quad (6.4)$$

where $\phi_i(x)$ are the basis functions and $u_i(t) \approx u(x_i, t)$. At $t = 0$ the approximation is an equality $u_i(0) = u(x_i, 0) = f(x_i)$.

(See Babuska *et al.*, 2010; Huebner *et al.*, 2001; Van der Vegt & Bokhove, 2009).

6.1.2 Spatial discretization and weak formulation

Let $w(x)$ be a differentiable function that satisfies the BCs, $w(-1) = w(1) = 0$. We will call $w(x)$ the test function. By multiplying the nonlinear wave equation (4.1) with this function and integrating it over the domain the following expression will be obtained

$$0 = \int_{-1}^1 \left(\frac{\partial^2 u_{sd}(x, t)}{\partial t^2} - \frac{\partial^2 u_{sd}(x, t)}{\partial x^2} + u_{sd}^3(x, t) \right) w(x) dx. \quad (6.5)$$

This expression can be rewritten as,

$$0 = \int_{-1}^1 \frac{\partial^2 u_{sd}(x, t)}{\partial t^2} w(x) dx - \int_{-1}^1 \frac{\partial^2 u_{sd}(x, t)}{\partial x^2} w(x) dx + \int_{-1}^1 u_{sd}^3(x, t) w(x) dx, \quad (6.6)$$

using integration by part we get

$$0 = \int_{-1}^1 \frac{\partial^2 u_{sd}(x, t)}{\partial t^2} w(x) dx + \int_{-1}^1 \frac{\partial u_{sd}(x, t)}{\partial x} \frac{\partial w}{\partial x} dx + \int_{-1}^1 u_{sd}^3(x, t) w(x) dx. \quad (6.7)$$

The plan is now to substitute the solution (6.4) and find an equation for the nodal values $u_i(t)$.

In order to work out the final term in (6.7), we note that from (6.4) we have

$$u_{sd}^3(x, t) = \sum_{i=0}^M (u_i(t)\phi_i(x))^3 + 3 \sum_{\substack{l,i=0 \\ |l-i|=1}}^M (u_l(t)\phi_l(x))^2(u_i(t)\phi_i(x)). \quad (6.8)$$

Substituting (6.8) into (6.7) one gets

$$\begin{aligned} 0 = & \int_{-1}^1 \frac{\partial^2 u_{sd}(x, t)}{\partial t^2} w(x) dx + \int_{-1}^1 \frac{\partial u_{sd}(x, t)}{\partial x} \frac{\partial w}{\partial x} dx \\ & + \int_{-1}^1 \left[\sum_{i=0}^M (u_i(t)\phi_i(x))^3 + 3 \sum_{\substack{l,i=0 \\ |l-i|=1}}^M (u_l(t)\phi_l(x))^2(u_i(t)\phi_i(x)) \right] w(x) dx. \end{aligned} \quad (6.9)$$

For our finite element methods the test functions $w(x)$ are chosen to be the same as the basis functions $\phi_1(x), \dots, \phi_{M-1}(x)$; we will consider the two remaining basis functions, $\phi_0(x)$ and $\phi_M(x)$, later. Now by replacing w by ϕ_j throughout one gets

$$\begin{aligned} 0 = & \int_{-1}^1 \left(\sum_{i=0}^M \frac{\partial^2 u_i(t)}{\partial t^2} \phi_i(x)\phi_j(x) \right) dx + \int_{-1}^1 \left(\sum_{i=0}^M u_i(t) \frac{\partial \phi_i(x)}{\partial x} \frac{\partial \phi_j(x)}{\partial x} \right) dx \\ & + \int_{-1}^1 \left(\sum_{i=0}^M (u_i(t)\phi_i(x))^3 \phi_j(x) \right) dx + 3 \int_{-1}^1 \left(\sum_{\substack{l,i=0 \\ |l-i|=1}}^M (u_l(t)\phi_l(x))^2(u_i(t)\phi_i(x))\phi_j(x) \right) dx, \end{aligned} \quad (6.10)$$

for $j = 1, \dots, M - 1$.

This can be written in the form

$$0 = \sum_{i=0}^M \frac{\partial^2 u_i(t)}{\partial t^2} T_{i,j} + \sum_{i=0}^M u_i(t) S_{i,j} + \sum_{i=0}^M u_i(t)^3 V_{i,j} + 3 \sum_{\substack{l,i=0 \\ |l-i|=1}}^M u_l(t)^2 u_i(t) B_{l,i,j}. \quad (6.11)$$

For the two basis functions left out one can use the following equations

$$\frac{\partial^2 u_0(t)}{\partial t^2} = 0 \quad \text{and} \quad \frac{\partial^2 u_M(t)}{\partial t^2} = 0, \quad (6.12)$$

in order to make sure that the boundary conditions hold, as the boundary conditions lead to $u_0(t) = 0$ and $u_M(t) = 0$.

Equation (6.11) will result in $M - 1$ linear ordinary differential equations for $u_i(t)$. The values of $T_{i,j}$, $S_{i,j}$, $V_{i,j}$ and $B_{l,i,j}$ can be calculated. As the basis functions only have a maximum of two elements on which they are non-zero, the values can be easily calculated by integrating over one or two elements on which both the basis functions are non-zero.

The term $\sum_{i=0}^M \frac{\partial^2 u_i(t)}{\partial t^2} T_{i,j}$ in (6.11) will be as follows:

If $i = j$ then we will have

$$T_{i,i} = \int_{-1}^1 \phi_i(x)^2 dx = \int_{x_{i-1}}^{x_i} \phi_i(x)^2 dx + \int_{x_i}^{x_{i+1}} \phi_i(x)^2 dx = \frac{1}{3}(x_{i+1} - x_{i-1}),$$

and when $i \neq j$, and i is one apart from j then

$$T_{i,i+1} = \int_{-1}^1 \phi_i(x)\phi_{i+1}(x) dx = \int_{x_i}^{x_{i+1}} \phi_i(x)\phi_{i+1}(x) dx = \frac{1}{6}(x_{i+1} - x_i),$$

and if i and j are more than one apart then $T_{i,j} = 0$.

So, T is the $(M - 1) \times (M - 1)$ tridiagonal matrix given by:

$$T = \text{tridiag} \left(\frac{1}{6} \left([x_j]_{j=2}^{M-1} - [x_j]_{j=1}^{M-2} \right), \frac{1}{3} \left([x_j]_{j=2}^M - [x_j]_{j=0}^{M-2} \right), \frac{1}{6} \left([x_j]_{j=3}^M - [x_j]_{j=2}^{M-1} \right) \right),$$

where $\text{tridiag}(a, b, c)$ represents a tridiagonal matrix with the vectors a, b and c on the subdiagonal, main diagonal, and superdiagonal, respectively.

By defining the vector $\vec{u}(t) := (u_1(t), \dots, u_{M-1}(t))$ one can write the $M - 1$ differential equations in (6.11) as one differential equation for the vector $\vec{u}(t)$. Accordingly, we can write the term

$$\sum_{i=0}^M \frac{\partial^2 u_i(t)}{\partial t^2} T_{i,j} = \sum_{i=1}^{M-1} \frac{\partial^2 u_i(t)}{\partial t^2} T_{i,j} = T \frac{\partial^2 \vec{u}(t)}{\partial t^2}, \quad (6.13)$$

where $u_0(t) = 0$ and $u_M(t) = 0$.

The term $\sum_{i=0}^M u_i(t) S_{i,j}$ in (6.11) will be as follows:

If $i = j$ then

$$S_{i,i} = \int_{-1}^1 \left(\frac{\partial \phi_i(x)}{\partial x} \right)^2 dx = \int_{x_{i-1}}^{x_i} \left(\frac{\partial \phi_i(x)}{\partial x} \right)^2 dx + \int_{x_i}^{x_{i+1}} \left(\frac{\partial \phi_i(x)}{\partial x} \right)^2 dx = \frac{1}{x_i - x_{i-1}} + \frac{1}{x_{i+1} - x_i},$$

and when $i \neq j$ and i is one apart from j then

$$S_{i,i+1} = \int_{-1}^1 \left(\frac{\partial \phi_i(x)}{\partial x} \right) \left(\frac{\partial \phi_{i+1}(x)}{\partial x} \right) dx = \int_{x_i}^{x_{i+1}} \left(\frac{\partial \phi_i(x)}{\partial x} \right) \left(\frac{\partial \phi_{i+1}(x)}{\partial x} \right) dx = \frac{-1}{x_{i+1} - x_i},$$

and if i and j are more than one apart then $S_{i,j} = 0$.

So, S is the $(M-1) \times (M-1)$ tridiagonal matrix given by:

$$S = \text{tridiag} \left(\frac{-1}{[x_j]_{j=2}^{M-1} - [x_j]_{j=1}^{M-2}}, \left(\frac{1}{[x_j]_{j=1}^{M-1} - [x_j]_{j=0}^{M-2}} + \frac{1}{[x_j]_{j=2}^M - [x_j]_{j=1}^{M-1}} \right), \frac{-1}{[x_j]_{j=2}^{M-1} - [x_j]_{j=1}^{M-2}} \right).$$

Thus, we can write the term as

$$\sum_{i=0}^M u_i(t) S_{i,j} = S \vec{u}(t). \quad (6.14)$$

The term $\sum_{i=0}^M u_i(t)^3 V_{i,j}$ in (6.11) will be as follows:

If $i = j$ then we will have

$$V_{i,i} = \int_{-1}^1 \phi_i(x)^4 dx = \int_{x_{i-1}}^{x_i} \phi_i(x)^4 dx + \int_{x_i}^{x_{i+1}} \phi_i(x)^4 dx = \frac{1}{5}(x_{i+1} - x_{i-1}),$$

and when $i \neq j$ and i is one apart from j then

$$V_{i,i+1} = \int_{-1}^1 \phi_i(x)^3 \phi_{i+1}(x) dx = \int_{x_i}^{x_{i+1}} \phi_i(x)^3 \phi_{i+1}(x) dx = \frac{1}{20}(x_{i+1} - x_i),$$

and if i and j are more than one apart then $V_{i,j} = 0$.

So, V is the $(M-1) \times (M-1)$ tridiagonal matrix given by:

$$V = \text{tridiag} \left(\frac{1}{20} ([x_j]_{j=2}^{M-1} - [x_j]_{j=1}^{M-2}), \frac{1}{5} ([x_j]_{j=2}^M - [x_j]_{j=0}^{M-2}), \frac{1}{20} ([x_j]_{j=3}^M - [x_j]_{j=2}^{M-1}) \right).$$

Thus, we can write the term as

$$\sum_{i=0}^M u_i(t)^3 V_{i,j} = V \vec{u}^3(t), \quad (6.15)$$

where $\vec{u}^3(t)$ is the vector with each entry cubic: $\vec{u}^3(t) = (u_1^3(t), \dots, u_{M-1}^3(t))$.

The term $3 \sum_{\substack{l,i=0 \\ |l-i|=1}}^M u_l(t)^2 u_i(t) B_{l,i,j}$ in (6.11) will be as follows:

If l, i, j are all different, then $B_{l,i,j} = 0$.

If two indices are equal, say $j = l$, and i and j are one apart then

$$B_{j,j+1,j} = \int_{-1}^1 \phi_j(x)^3 \phi_{j+1}(x) dx = \int_{x_j}^{x_{j+1}} \phi_j(x)^3 \phi_{j+1}(x) dx = \frac{1}{20}(x_{j+1} - x_j).$$

If $j = i$ and l and j are one apart then

$$B_{i+1,i,i} = \int_{-1}^1 \phi_i(x)^2 \phi_{i+1}(x)^2 dx = \int_{x_i}^{x_{i+1}} \phi_i(x)^2 \phi_{i+1}(x)^2 dx = \frac{1}{30}(x_{i+1} - x_i).$$

The three indices l, i, j will never be all equal because we have $l \neq i$ as $|l - i| = 1$. In all other cases $B_{l,i,j} = 0$; if i and j are more than one apart, and if l and j are more than one apart.

We now define the matrix B of size $(M - 1) \times 2(M - 2)$ by

$$B_{ij} = \begin{cases} \frac{1}{20}(x_2 - x_1), & \text{if } i = j = 1, \\ \frac{1}{30}(x_2 - x_1), & \text{if } i = 1, j = 2, \\ \frac{1}{30}(x_i - x_{i-1}), & \text{if } 2 \leq i \leq M - 2, j = 2(i - 1) - 1, \\ \frac{1}{20}(x_i - x_{i-1}), & \text{if } 2 \leq i \leq M - 2, j = 2(i - 1), \\ \frac{1}{20}(x_{i+1} - x_i), & \text{if } 2 \leq i \leq M - 2, j = 2(i - 1) + 1, \\ \frac{1}{30}(x_{i+1} - x_i), & \text{if } 2 \leq i \leq M - 2, j = 2(i - 1) + 2, \\ \frac{1}{30}(x_i - x_{i-1}), & \text{if } i = M - 1, j = 2(M - 2) - 1, \\ \frac{1}{20}(x_i - x_{i-1}), & \text{if } i = M - 1, j = 2(M - 2), \end{cases}$$

the rest of entries are zeros.

Thus, we can write the term as

$$3 \sum_{\substack{l,i=0 \\ |l-i|=1}}^M u_l(t)^2 u_i(t) B_{l,i,j} = 3B\vec{R}(\vec{u}(t)), \quad (6.16)$$

where $R(\vec{u}(t))$ is

$$R(\vec{u}(t)) = \begin{pmatrix} u_1^2(t)u_2(t) \\ u_2^2(t)u_1(t) \\ u_2^2(t)u_3(t) \\ \vdots \\ u_{M-2}^2(t)u_{M-3}(t) \\ u_{M-2}^2(t)u_{M-1}(t) \\ u_{M-1}^2(t)u_{M-2}(t) \end{pmatrix}_{2(M-2) \times 1}, \quad (6.17)$$

the entries of the vector are of the form $u_l^2(t)u_i(t)$ where $|l - i| = 1$ for $l, i = 1, \dots, M - 1$.

Therefore, using (6.13), (6.14), (6.15) and (6.16) one can write (6.11) in the form

$$T \frac{\partial^2 \vec{u}(t)}{\partial t^2} + S\vec{u}(t) + V\vec{u}^3(t) + 3B\vec{R}(\vec{u}(t)) = 0. \quad (6.18)$$

Basically, the formulas get much easier when using equal sized elements, hence we will consider this case. For equal size points we have $\Delta_x = \frac{2}{M}$. Therefore, the nodes can be defined as

$$x_i := -1 + i\Delta_x, \quad \text{for } i = 0, \dots, M. \quad (6.19)$$

For such case we have

$$T_{i,i} = \frac{2}{3}\Delta_x, \quad T_{i,i+1} = \frac{1}{6}\Delta_x, \quad S_{i,i} = \frac{2}{\Delta_x}, \quad S_{i,i+1} = \frac{-1}{\Delta_x},$$

and

$$V_{i,i} = \frac{2}{5}\Delta_x, \quad V_{i,i+1} = \frac{1}{20}\Delta_x.$$

In order to solve (6.18) one can multiply it by T^{-1} :

$$\frac{\partial^2 \vec{u}(t)}{\partial t^2} = -T^{-1}S\vec{u}(t) - T^{-1}V\vec{u}^3(t) - 3T^{-1}BR(\vec{u}(t)). \quad (6.20)$$

Let $S^* = -T^{-1}S$, $V^* = -T^{-1}V$ and $B^* = -T^{-1}B$, then (6.20) becomes

$$\frac{\partial^2 \vec{u}(t)}{\partial t^2} = S^*\vec{u}(t) + V^*\vec{u}^3(t) + 3B^*R(\vec{u}(t)). \quad (6.21)$$

The ODE (6.21) can be written in first-order form as

$$\frac{\partial \vec{y}(t)}{\partial t} = \vec{w}(t), \quad (6.22)$$

$$\frac{\partial \vec{w}(t)}{\partial t} = S^*\vec{y}(t) + V^*\vec{y}^3(t) + 3B^*R(\vec{y}(t)). \quad (6.23)$$

6.1.3 Time stepping

Euler's method is used to solve (6.22) and (6.23), this gives

$$\vec{y}^{n+1} = \vec{y}^n + h\vec{w}^n, \quad (6.24)$$

$$\vec{w}^{n+1} = \vec{w}^n + h \left\{ S^*\vec{y}^n + V^*(\vec{y}^n)^3 + 3B^*R(\vec{y}^n) \right\}. \quad (6.25)$$

In addition to Euler's method, we also use the trapezoidal rule to solve (6.22) and (6.23):

$$\vec{y}^{n+1} = \vec{y}^n + \frac{h}{2} \left\{ \vec{w}^{n+1} + \vec{w}^n \right\}, \quad (6.26)$$

$$\vec{w}^{n+1} = \vec{w}^n - \frac{h}{2} \left\{ S^*\vec{y}^{n+1} + V^*(\vec{y}^{n+1})^3 + 3B^*R(\vec{y}^{n+1}) \right. \\ \left. + S^*\vec{y}^n + V^*(\vec{y}^n)^3 + 3B^*R(\vec{y}^n) \right\}. \quad (6.27)$$

6.1.4 Solving the nonlinear equation using Newton's method

The trapezoidal rule leads to a system of nonlinear equations, (6.26) and (6.27), which needs to be solved for y_j^{n+1} and w_j^{n+1} . To achieve this, Newton's method, which is given in Section 1.7.3, is used. It can be written as

$$g_{j+1} = g_j - J^{-1}(g_j)F_n(g_j), \quad (6.28)$$

where j is the iteration index of the Newton's method, g_j is taken to be the solution of the previous step and g_0 can be found from the initial condition, i.e., $g_0 = \{y_j^0, w_j^0\}$, for $j = 1, \dots, M - 1$. The function F in Newton's method is

$$F(g_{n+1}) = \begin{pmatrix} H_j(g_{n+1}) \\ G_j(g_{n+1}) \end{pmatrix}_{2(N-1) \times 1}, \quad j = 1, \dots, M - 1,$$

where from (6.26) and (6.27)

$$H_j(g_{n+1}) = y_j^{n+1} - y_j^n - \frac{h}{2} \{w_j^{n+1} + w_j^n\}, \quad (6.29)$$

and

$$\begin{aligned} G_j(g_{n+1}) = & w_j^{n+1} - w_j^n - \frac{h}{2} \sum_{l=1}^{M-1} S_{jl}^* y_l^{n+1} - \frac{h}{2} \sum_{l=1}^{M-1} V_{jl}^* (y_l^{n+1})^3 - 3 \frac{h}{2} \sum_{p=1}^{2(M-2)} B_{jp}^* R_p(y^{n+1}) \\ & - \frac{h}{2} \sum_{l=1}^{M-1} S_{jl}^* y_l^n - \frac{h}{2} \sum_{l=1}^{M-1} V_{jl}^* (y_l^n)^3 - 3 \frac{h}{2} \sum_{p=1}^{2(M-2)} B_{jp}^* R_p(y^n), \quad \text{for } j = 1, \dots, M - 1, \end{aligned} \quad (6.30)$$

and from the boundary conditions we have $y_0^{n+1} = 0$ and $w_0^{n+1} = 0$. This implies,

$$F(g_{n+1}) = \begin{pmatrix} y_1^{n+1} - y_1^n - \frac{h}{2}w_1^{n+1} - \frac{h}{2}w_1^n \\ y_2^{n+1} - y_2^n - \frac{h}{2}w_2^{n+1} - \frac{h}{2}w_2^n \\ \vdots \\ y_{M-1}^{n+1} - y_{M-1}^n - \frac{h}{2}w_{M-1}^{n+1} - \frac{h}{2}w_{M-1}^n \\ w_1^{n+1} - w_1^n - \frac{h}{2} \sum_{l=1}^{M-1} S_{1l}^* y_l^{n+1} - \frac{h}{2} \sum_{l=1}^{M-1} V_{1l}^* (y_l^{n+1})^3 - 3 \frac{h}{2} \sum_{p=1}^{2(M-2)} B_{1p}^* R_p(y^{n+1}) \\ - \frac{h}{2} \sum_{l=1}^{M-1} S_{1l}^* y_l^n - \frac{h}{2} \sum_{l=1}^{M-1} V_{1l}^* (y_l^n)^3 - 3 \frac{h}{2} \sum_{p=1}^{2(M-2)} B_{1p}^* R_p(y^n) \\ w_2^{n+1} - w_2^n - \frac{h}{2} \sum_{l=1}^{M-1} S_{2l}^* y_l^{n+1} - \frac{h}{2} \sum_{l=1}^{M-1} V_{2l}^* (y_l^{n+1})^3 - 3 \frac{h}{2} \sum_{p=1}^{2(M-2)} B_{2p}^* R_p(y^{n+1}) \\ - \frac{h}{2} \sum_{l=1}^{M-1} S_{2l}^* y_l^n - \frac{h}{2} \sum_{l=1}^{M-1} V_{2l}^* (y_l^n)^3 - 3 \frac{h}{2} \sum_{p=1}^{2(M-2)} B_{2p}^* R_p(y^n) \\ \vdots \\ w_{M-1}^{n+1} - w_{M-1}^n - \frac{h}{2} \sum_{l=1}^{M-1} S_{M-1l}^* y_l^{n+1} - \frac{h}{2} \sum_{l=1}^{M-1} V_{M-1l}^* (y_l^{n+1})^3 - 3 \frac{h}{2} \sum_{p=1}^{2(M-2)} B_{M-1p}^* R_p(y^{n+1}) \\ - \frac{h}{2} \sum_{l=1}^{M-1} S_{M-1l}^* y_l^n - \frac{h}{2} \sum_{l=1}^{M-1} V_{M-1l}^* (y_l^n)^3 - 3 \frac{h}{2} \sum_{p=1}^{2(M-2)} B_{M-1p}^* R_p(y^n) \end{pmatrix}$$

of size $2(M-1) \times 1$.

The Jacobian matrix J of the function F is

$$J(g_{n+1}) = \begin{pmatrix} \frac{\partial H}{\partial y^{n+1}} & \frac{\partial H}{\partial w^{n+1}} \\ \frac{\partial G}{\partial y^{n+1}} & \frac{\partial G}{\partial w^{n+1}} \end{pmatrix}_{2(M-1) \times 2(M-1)},$$

which implies,

$$J(g_{n+1}) = \begin{pmatrix} I & -\frac{h}{2}I \\ A & I \end{pmatrix}_{2(M-1) \times 2(M-1)},$$

where $A = -\frac{h}{2}S^* - \frac{3h}{2}V^*(y^{n+1})^2 - \frac{3h}{2}\tilde{A}$ with \tilde{A} a matrix of size $(M-1) \times (M-1)$ defined as follows:

Let

$$A_{ji}^* = \begin{cases} Q_{ji} = 2y_i^{n+1}y_{i+1}^{n+1}B_{j1}^* + (y_{i+1}^{n+1})^2B_{j2}^*, & \text{if } j, i = 1 \text{ or } i = j - 1, \\ E_{ji} = (y_{i-1}^{n+1})^2B_{jC}^* + 2y_{i-1}^{n+1}y_i^{n+1}B_{jC+1}^* + 2y_i^{n+1}y_{i+1}^{n+1}B_{jC+2}^* + (y_{i+1}^{n+1})^2B_{jC+3}^*, & \text{if } 2 \leq j \leq M-2 \text{ and } i = j, \\ W_{ji} = (y_{i-1}^{n+1})^2B_{jC}^* + 2y_{i-1}^{n+1}y_i^{n+1}B_{jC+1}^*, & \text{if } j, i = M-1 \text{ or } i = j + 1, \end{cases}$$

where $C = 2(i - 1) - 1$. Then

$$\tilde{A} = \begin{pmatrix} Q_{11} & W_{12} & 0 & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ Q_{21} & E_{22} & W_{23} & 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & Q_{32} & E_{33} & W_{34} & 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & Q_{43} & E_{44} & W_{45} & 0 & \dots & 0 & 0 & 0 & 0 \\ \vdots & & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & Q_{M-2M-3} & E_{M-2M-2} & W_{M-2M-1} \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & 0 & 0 & Q_{M-1M-2} & W_{M-1M-1} \end{pmatrix}.$$

6.2 Numerical results

In this section, we study the convergence and stability of the Euler and trapezoidal methods combined with the FEM when finding the solution of the nonlinear wave equation (4.1). As in the previous chapters, we estimate the error by comparing the numerical solution with the reference solution from Chapter 4 (see Subsection 4.3.1) computing with high accuracy. The error is evaluated as in the previous chapters, see Subsections 3.3.2 and 4.3.1. The particular initial conditions used in this chapter are the three ICs used in the previous chapters, see Subsection 3.4. The ICs are,

$$f(x) = \sin \pi x,$$

which is (3.81), and the initial condition,

$$f(x) = e^{\sin(\pi x)} - 1,$$

see (3.82). The third initial condition is,

$$f(x) = \frac{1}{11} e^x (11 - 8x - 10x^2 + 8x^3 - x^4),$$

see (3.83).

6.2.1 Convergence of solution

Figures 6.1 and 6.2 show the convergence error in space and time, respectively, of the FEM when solving the nonlinear wave equation (4.1) solved using the FEM with Euler's and the trapezoidal schemes. It can be seen in the figures that the

FEM are convergent as the convergence error is decaying to zero correspondingly as we will explain next.

As is generally known, the FEM converges algebraically, with respect to space. In order to get convergence the solution and its derivatives must be continuous up to one order less than the highest order derivative appearing in the PDE, (see Babuska *et al.*, 2010; Huebner *et al.*, 2001). The highest order derivative in our equation is 2 which means that the solution must be C^1 , which is indeed the case. However, this only guarantees convergence; in order for the convergence to be order 2, error $= \mathcal{O}(M^2)$, (Babuska *et al.*, 2010, p. 143) says that the solution must be C^2 . As explained in Section 1.9 the solution generated by the first IC (3.81) is C^∞ , the solution for the second IC (3.82) is C^1 , and the solution for the third IC (3.83) is C^3 . Thus, only for the first and third ICs is the solution at least C^2 , but for the second IC it is not. Nevertheless, Figure 6.1 shows that the FEM converges for all the three ICs, (3.81), (3.82) and (3.83), with order two. The discontinuity produced by the second IC (3.82) has only a minor impact; the graph is not as straight a line as it is for the other two ICs, but the convergence order is unchanged, it is second order as well. The reason for this is perhaps that while the solution for the second IC is not C^2 , it is piecewise C^2 with at most two discontinuities in the second derivative. Also, for Euler's method we are getting similar results for all the ICs implemented.

If we compare Figure 6.1 with the corresponding figures using spectral methods for the linear wave equation in Chapter 3, Figure 3.5, and for the nonlinear wave equation in Chapter 4, Figure 4.3, and Figure 5.5 in Chapter 5, we can see that the FEM behaves very differently. The accuracy of the spectral methods drops off from exponential convergence for smooth solutions, when using the first IC, (3.81), to only algebraic convergence for nonsmooth solutions, specifically when using (3.82) and (3.83) ICs. Meanwhile, the FEM maintains algebraic convergence of order two with respect to M for all these three ICs regardless of the discontinuity. So on conclusion, the nonsmoothness of the solutions generated from the second IC, (3.82) and the third IC, (3.83), has no impact on the order of convergence for the FEM, in contrast with the spectral methods. The only effect is that in the case of the second IC the plot is not a very nice straight line as the plots of other two ICs.

Regarding the time step error, h , Figure 6.2 for the IC (3.81) gives that Euler's

method is of order one, the left panel. Meanwhile, the trapezoidal method is of order two, the right panel. The other two ICs also have same orders for both methods.

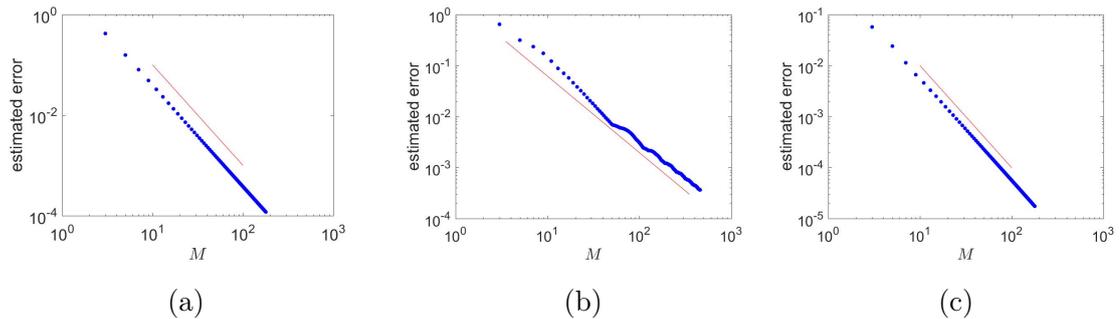


Figure 6.1: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the number of elements M for the nonlinear wave equation (4.1) solved using the FEM with the trapezoidal method for the C^∞ IC (3.81), the C^1 IC (3.82) and the C^3 IC (3.83) (a, b and c respectively). The error is evaluated at $h = 10^{-3}$ and $t = 0.9$. The red straight line indicates order two, $\frac{1}{M^2}$, for all the three ICs. Here u_{num} is the numerical solution and u is the reference solution in Subsection 4.3.1.

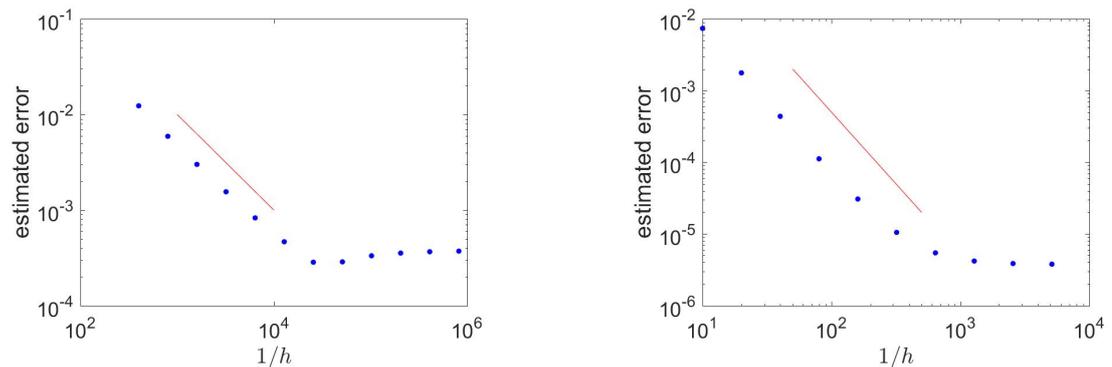


Figure 6.2: The error, $\|u - u_{num}\|_{L^\infty(-1,1)}$, against the inverse step size $1/h$ for the nonlinear wave equation (4.1) solved using the FEM method for the C^∞ initial condition (3.81). The left panel with Euler's method for $M = 100$ and the right panel with the trapezoidal method for $M = 1000$. The error is evaluated at time $t = 0.9$. The red straight line indicates order one for Euler's method and order two for the trapezoidal method.

6.2.2 Stability for Euler's and the trapezoidal methods

When studying the stability of Euler's and the trapezoidal methods combined with the FEM to solve the nonlinear wave equation (4.1) using our three initial conditions with different properties regarding the smoothness, similar figures as Figures 4.7 and 4.9 are obtained. The numerical solution is increasing rapidly for Euler's method, so Euler's method is also unstable when combined with the FEM. On the other hand, the numerical solutions of the trapezoidal method are bounded, i.e., the estimated error does not grow as a function of time for the three ICs employed. This means that the trapezoidal method is stable.

6.2.3 Symplecticity of the FEM for Euler and trapezoidal methods

Figure 6.3 shows the energy error of the FEM when solving the nonlinear wave equation using Euler's scheme for the C^1 IC (3.82). The energy error is not constant but increases rapidly in time. This behaviour occurs for all the three ICs employed and is the same as for the spectral methods, see Figure 4.12 in Chapter 4. Hence, the energy for Euler's method is not preserved.

Figure 6.4 shows the energy error for the trapezoid scheme for the C^1 IC (3.82). This figure shows that the energy for the trapezoidal method is almost preserved: it is almost constant and does not decay to zero as the time increases, and it is very close to the exact energy computed in Subsection 4.4.1. The energy error scales as expected for second order methods, i.e., it decreases by a factor of 100 when the step size h decreases by a factor of 10. The figure shows that the energy error is approximately 10^{-4} for $h = 10^{-2}$, for $h = 10^{-3}$ it is around 10^{-6} and for $h = 10^{-4}$ it is 10^{-8} . Comparable behaviour is achieved for the other two ICs, (3.81) and (3.83). All the three ICs require the same number of elements M and quadrature points nx . We do not have to increase M , making it very big as we did for the spectral methods with the ICs that produce the nonsmoothness, see Figures 4.14-4.15 in Chapter 4 and Figure 5.12 in Chapter 5. Remember that for the spectral methods one had to increase the number of modes to $N = 1000$ for the C^1 IC (3.82) and to

$N = 100$ for the C^3 IC (3.83) in order to get the trapezoidal method to have the energy almost conserved. In addition, the number of quadrature points, nx , when computing the numerical energy, does not need to be very large as for the spectral methods chapters when we had to take $nx = 100N$ in order to show clearly that the method is almost symplectic. The reason is that, as it is stated in the previous Subsection 6.2.1, the FEM is algebraically convergent with order two for all those three ICs used regardless of the nonsmoothness of the solutions produced for the ICs (3.82) and (3.83).

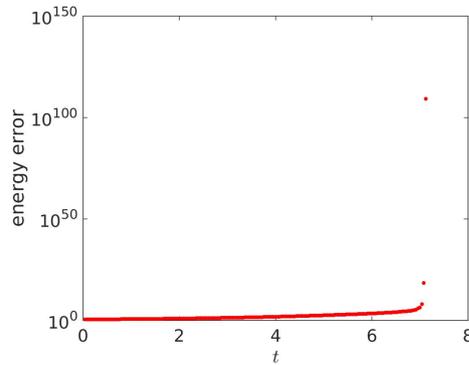


Figure 6.3: The energy error, $E - E_{exact}$, of the FEM with Euler's rule applied to the nonlinear wave equation (4.1) for $M = 10$, $h = 10^{-2}$ and the C^1 IC (3.82). Here E is the numerical energy and E_{exact} is the exact energy. We get a similar plot for the energy difference $E - E_0$, where E_0 is the numerical energy at $t = 0$.

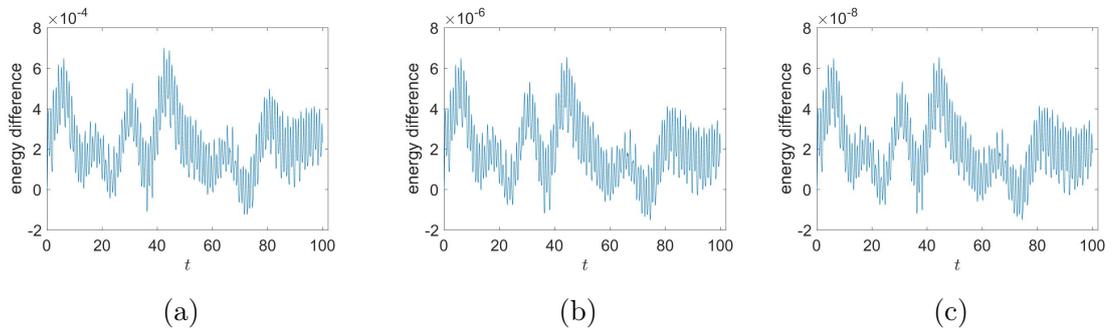


Figure 6.4: The energy difference, $E - E_0$, for the nonlinear wave equation (4.1) using the FEM with the trapezoidal rule, for $M = 100$, $h = 10^{-2}, 10^{-3}, 10^{-4}$ (a, b, and c respectively) and the C^1 IC (3.82).

6.3 Summary and discussion

In this chapter the nonlinear wave equation $\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2} - u^3$, (4.1), with homogeneous Dirichlet boundary conditions is solved using finite element methods (FEM) with linear basis functions. Two time stepping methods are implemented in finding the solution, Euler's and the trapezoidal methods. The three initial conditions are considered (3.81), (3.82) and (3.83), which generate solutions with different levels of smoothness. Similar to earlier chapters, many numerical concepts are studied in this chapter as well, such as the convergence in space M and time t , stability and energy conservation.

Theory says that the FEM converges algebraically if the solution and its derivatives are continuous up to one order less than the highest order derivative in the PDE. The order of convergence is two if the solution is C^2 . In contrast, the spectral methods require an infinitely smooth solution to get exponential convergence. Our results show that the error with respect to M converges algebraically with order two for all three ICs. Unlike in the spectral methods chapters, the discontinuity, yielded by the second IC (3.82) in the second derivative and by the third IC (3.83) in the fourth derivative, has no impact on the convergence as the order of convergence remains two for all the ICs employed. The discontinuity generated by the second IC (3.82) does have a very minor impact is that the graph is not as a straight line as it is for the other two ICs. In contrast, for the spectral methods the error with respect to N converges exponentially for the first IC (3.81) but for the other ICs, (3.82) and (3.83), the convergence drops off to be algebraic with order seems to be three for (3.82) and order five for (3.83), for all the time stepping methods employed. In addition, our results show that the nonsmoothness does not impact on the convergence in time: the error of Euler's method is $\mathcal{O}(h)$ and the trapezoidal method is $\mathcal{O}(h^2)$.

We now discuss stability and energy preservation. Similar to Chapters 3 and 4, Euler's method is not stable at all and the trapezoidal rule is stable for the initial conditions studied here. The error for the trapezoidal rule does not grow as time increases but instead it is bounded. Unsurprisingly, Euler's methods does not preserve energy, similar to the previous chapters. Whereas, the trapezoidal rule almost conserves energy; and it requires the same number of elements, M ,

and quadrature points, nx , for all the ICs implemented. In other words, the nonsmoothness we face has no impact on the energy either. In contrast, the spectral methods require a very large number of modes, N , and points, nx , when the discontinuity is present to get the energy to be almost preserved.

Chapter 7

General conclusion

In examining any numerical procedure, the smoothness of the solutions is very crucial. It plays a significant part, for instance, in the spectral methods (SM), which are exponentially convergent with regard to space for infinitely smooth problems. However, in other approaches with low order of convergence, such as the finite element methods (FEM), smoothness does not play such a significant role.

Prior to this study, much work in literature is performed for solving completely infinitely smooth problems along the lines of the linear or nonlinear wave equations. However, the literature is much scarcer on solving the nonsmooth problems, especially for the nonlinear problems. There are some papers that take into account a few simple nonsmooth problems; mostly they deal with linear problems generally and some with linear hyperbolic partial differential equations, either nonsmoothness is in the initial data or the solution itself, but not in its higher derivatives. In these papers, the authors tried to resolve nonsmoothness by introducing some filters in order to get the outcome wanted. Nevertheless, this thesis is doing something related but very different, it considers problems that nonsmoothness is in the later solution's derivatives for linear and nonlinear problems. More precisely, this approach has been devoted to conduct a systematic comparison study of a variety of numerical methods that are employed to solve the linear wave equation (3.52), $u_{tt} = u_{xx}$, and the nonlinear wave equation (4.1), $u_{tt} = u_{xx} - u^3$, on the bounded interval $x = [-1, 1]$ with Dirichlet boundary conditions $u = 0$ at the endpoints, all applied to particular initial conditions that have been carefully chosen to yield solutions with different levels of smoothness. Significantly, the derivative level at

which we want that discontinuity to appear is something that can systematically be controlled in the present work, i.e., we can control precisely how smooth or not smooth we want our solutions to be. The success of a systematic comparison depends on the availability and quality of literature, therefore, this thesis is studying all above in a systematic comparison so as to be a novel contribution added to the literature.

This thesis investigates the solutions, numerical and exact (where available), of these two equations in Chapters 1, 3, 4, 5 and 6. The linear wave equation (3.52) is investigated in Chapters 1, 3 and 5, and the nonlinear wave equation (4.1) in Chapters 4, 5 and 6. The numerical methods used are the spectral methods (SM), based on Chebyshev polynomials, in Chapters 3, 4 and 5, and the finite element methods (FEM), in Chapter 6. There are several spectral approximations that can be employed, such as the pseudospectral, collocation, tau and Galerkin methods, which are all used mainly in Chapter 3, while Chapters 4 and 5 implement the pseudospectral method only. Two main time stepping methods are used throughout this thesis in Chapters 3, 4 and 6, which are Euler's and trapezoidal. In addition, two more time stepping methods are considered in Chapters 3 and 4, which are the modified Euler and the implicit Euler. Moreover, some other non-standard time stepping techniques are considered in Chapter 5 which are exponential integrators, like exponential time differencing (ETD) methods (ETD1 and ETD2RK).

In the light of nonsmoothness phenomena, the main objective in this thesis is to investigate the numerical solution, convergence in space and time, stability and energy conservation of the numerical methods for the spectral methods and finite element methods. In addition, for the spectral methods, the convergence in space is thoroughly investigated from another point of view. Specifically, we look at the spectrum of the solutions. Throughout this thesis the numerical solutions are compared with their known analytical solutions (where available) and with the reference solutions otherwise. Section 7.1 provides the summary and contributions of the work presented in this thesis followed by section 7.2 which describes directions for future work.

7.1 Summary and contributions

In Chapter 1 we have structured the exact d'Alembert solution for the linear wave equation on a finite interval in space, $[-1, 1]$, which is adapted from the d'Alembert solution on an infinite interval. This solution is formed to be a periodic function with period four when considering Dirichlet BCs. This is done by extending the initial condition function $f(x)$ to all infinite domain using sort of odd extension; this leads to a periodic function on the infinite domain of period 4, then restrict the solution back to our finite interval. Three initial conditions (ICs), the C^∞ IC (3.81), the C^1 IC (3.82), and the C^3 IC (3.83) are addressed. These initial conditions have been carefully chosen to yield solutions with different levels of smoothness. They are selected so that the initial condition itself is infinitely differentiable on the entire interval, but the analytic solution at later times will have discontinuities in the derivatives. The first IC (3.81) yields an infinitely smooth solution. Nevertheless, the analytic solution will have discontinuities in the second derivatives for the second IC (3.82) and discontinuities in the fourth derivative for the third IC (3.83). This leads to have the numerical experiments of the third IC (which is C^2 but not C^4) fall between the other two ICs and this is clear throughout the thesis. In other words, we can control precisely how smooth or not smooth we want our solutions to be, **novel result** from Chapter 1. It is shown in Section 1.9 that the exact d'Alembert solution at any time which is a multiple of 4, i.e., at $t = 0, 4, 8, etc.$, are infinitely differentiable, infinitely smooth, for all the above initial conditions employed because the exact d'Alembert solution is of period 4 and have the returning property to the initial conditions; the solutions are just the functions $f(x)$ of the initial conditions, which are infinitely differentiable in $[-1, 1]$, **novel result** from Chapter 1. Likewise, at the other times of multiple 2, i.e., at $t = 2, 6, 10, etc.$, the exact d'Alembert solutions are infinitely differentiable again, infinitely smooth, since the characteristics touch the boundaries at these specific times, thus the exact d'Alembert solution is the function $-f(-x)$, **novel result** from Chapter 1. In addition, it is shown that the exact d'Alembert solution at $t = 1, 3, 5, etc.$, is an odd function, **novel result** from Chapter 1.

Firstly, we discuss the results of the SM:

All four spectral approximations mentioned, the pseudospectral, collocation, tau

and Galerkin methods, are used in Chapter 3. The results show that the pseudospectral method is more simple and accurate than the tau method, as the tau method is unstable. Moreover, the outcomes show that the pseudospectral method and the collocation method have approximately the same results. The Galerkin method also gives similar results to the pseudospectral method, but finding the new basis for solving the nonlinear wave equations is not that easy. Overall, only some minor differences are obtained in the results when using these methods, apart from the instability of the tau method. However, only some of the aspects are investigated, for the collocation, tau and Galerkin methods; the convergence using the spectra of the solution was not studied and only the C^∞ IC (3.81) is considered. Therefore, we have chosen to concentrate on the pseudospectral method to be the main and more practical method in our study in Chapters 3, 4 and 5.

The convergence with N is studied in Chapters 3, 4 and 5. The results show that the numerical solution converges exponentially for the first IC, (3.81), but for the other ICs, (3.82) and (3.83), it converges only algebraically, with order seems to be $\frac{1}{N^3}$, for (3.82), and order $\frac{1}{N^5}$, for (3.83), for both of the PDEs, for all time stepping methods employed. In addition, the convergence in space is thoroughly investigated from another point of view. Specifically, we look at the spectrum of the numerical solution of the combination the trapezium method with pseudospectral method for solving the linear wave equation (3.52) and the nonlinear wave equation (4.1), in Chapters 3, and 4, respectively. The spectrum of the numerical solution of the combination exponential integrators with pseudospectral method for solving the linear wave equation (3.52) and the nonlinear wave equation (4.1), in Chapter 5. The Chebyshev coefficients of the numerical solution \hat{u}_k are plotted against the mode k , the even coefficients component and the odd coefficients component, for all three ICs. From Section 1.9 we know that every function $f(x)$ of the initial conditions can be uniquely decomposed as the sum of even and odd functions. In a Chebyshev series, the terms with the even coefficients sum to be the even component and the terms with the odd coefficients sum to be the the odd component.

One can see that the odd/even coefficients of the solution conduct is different for all three ICs implemented, and, according to the times used. As known from Section 1.9 that the C^∞ IC (1.54) is an odd function, however, the other two ICs, the IC (3.83) and the IC (3.82), have both the odd and even components. The C^∞ (3.81) yields an infinitely smooth solution, that is also odd. We see that, the odd coefficients

of the numerical solution decay exponentially, but the even spectral coefficients are basically just round-off. While for the IC C^3 (3.83) introduces nonsmoothness at the fourth derivative since $F^{(4)}(\pm 1) \neq 0$ as $f^{(4)}(\pm 1) \neq 0$; this leads to have $f_o^{(4)}(\pm 1) \neq 0$ and $f_e^{(4)}(\pm 1) \neq 0$. We have seen that nonsmoothness affects any non-zero component of the solution. Therefore, both spectral coefficients components of the generated numerical solution change from decaying exponentially to only algebraically with order $\frac{1}{k^5}$, for both equations, at general t . Nevertheless, the C^1 IC (3.82) introduces nonsmoothness at the second derivative because $F''(\pm 1) \neq 0$ as $f''(\pm 1) = \pi^2$; this leads to have $f_o''(\pm 1) = 0$ and $f_e''(\pm 1) = \pi^2$. Therefore, the results of the linear problem, in Chapters 3 and 5, show that the odd spectral coefficients of the generated numerical solution decay exponentially, while the even spectral coefficients decay algebraically with order $\frac{1}{k^3}$ at general t . Interestingly, the results of the nonlinear wave equation (4.1), in Chapters 4 and 5, show that, for the second IC (3.82) both of the spectral coefficients components, the even and odd, decaying algebraically at general t . We have seen that is because the nonlinearity in the equation mixes the odd and the even parts over time. For the second IC (3.82), only the even part of the solution is nonsmooth initially, but due to the mixing the nonsmoothness leaks to the odd part. While for the third IC (3.83) the mixing up is not clear as both the odd and the even component of the IC introduces nonsmoothness, thus both the odd and the even coefficients decay algebraically, similar to the linear equation. To the best of our knowledge these differences of the odd/even components are **novel results**, in consideration of nonsmoothness.

We have observed from Chapters 3, 4 and 5, that there are some specific times, at $t = 0, 1, 2, 3, 4, 5, 6, etc.$, when the coefficients of the solutions have special shapes. Specifically, the three time groups are: $t = 0, 2, 4, 6, 8, etc.$, $t = 1, 3, 5, etc.$, and all times in between. The spectra figures are similar to each other at each time group, but, are different from the figures at the other time groups, for the C^3 IC (3.83) and the C^1 IC (3.82). The conduct of the plots is related to the conduct of the exact d'Alembert solution at each time group above. It is given in Section 1.9 that the exact d'Alembert solution of the linear problem is infinitely smooth at $t = 0, 2, 4, 6, 8, etc.$, therefore, the exact solution spectra decay exponentially for both components. While the exact d'Alembert solution at $t = 1, 3, 5, etc.$, is an odd function, thus, the even coefficients are just round-off.

Consequently, the Chebyshev coefficients of the numerical solution attempt to

mirror the behaviour of the exact solution spectra (i.e., recover smoothness of the solution or recover the difference of the odd/even components). However, the Chebyshev coefficients of the numerical solution, of the trapezoidal method combined with pseudospectral method, for solving the linear and nonlinear problems, in Chapters 3 and 4, respectively, are not able to reflect it faithfully but only partially, especially for the C^1 IC (3.82) up to the time stepping error $\mathcal{O}(h^2)$ only, but after that they decay algebraically, probably because of the accumulated error.

On the other hand, the special times when solving the nonlinear problem are only at $t = 0, 2, 4, 6, 8, \text{etc.}$ The exact solution of the nonlinear wave equation is unknown, however, the nonlinear wave equation share the same characteristics with the linear equation, therefore, in Chapters 1 and 4 we anticipated that the exact solution for the nonlinear wave equation to be smooth at these specified times as well for all three ICs. Astonishingly, our results in Chapter 5 show that exponential integrators recover the smoothness of the numerical solution at $t = 0, 2, 4, 6, \text{etc.}$, for both equations, the linear and nonlinear wave equations, **novel results** from Subsections 5.2.1 and 5.3.1, respectively. In other words, both components of the Chebyshev coefficients of the numerical solution drop off exponentially at these specified times. Wherefore, our assumption in Chapters 1 and 4 is confirmed and verified; that indeed the exact solution for the nonlinear wave equation is infinitely smooth at these specified times, **novel result** from Subsection 5.3.1. In addition, exponential integrators results for the linear problem in Subsection 5.2.1, show that the numerical solution spectra at $t = 1, 3, 5, \text{etc.}$, when the exact solution is an odd function, thus, the even Chebyshev coefficients of the numerical solution for the C^1 IC (3.82) are zero down to 10^{-10} , and they do not decay algebraically later, and also for the C^3 IC (3.83), **novel result** from Subsection 5.2.1. Immediately after these special times the even component of the Chebyshev coefficients drop of algebraically again as at the other ordinary times.

From our numerical results of the numerical solutions spectra in Subsections 3.4.4, 4.3.4, 5.2.1 and 5.3.1; it is clear that the trapezoidal method causes to have a fast drop, transition, in the plots of the even coefficients of the numerical solution of the C^1 IC (3.82) only, but not of the C^3 IC (3.83). One concludes that the trapezoidal method copes better with the C^3 IC (3.83) than the C^1 IC (3.82) in simulating the behaviour of the numerical solution spectra; perhaps because the solution generated

by the C^1 IC (3.82) is the less smooth than the solution generated by the C^3 IC (3.83). On the other hand, when employing exponential integrators the numerical solution spectra plots do not experience the fast drop for all three ICs. Meaning that, the drop in is caused mainly by the trapezoidal method.

In comparison with these chapters, 3, 4 and 5, the numerical solution spectra outcomes of using exponential integrators, in Chapter 5, outperform the outcomes of using the trapezoidal method; they are more precise and accurate. This is made more obvious by the numerical solution spectra for the linear problem, which demonstrate a strong likeness to the exact solution spectra, particularly at the special times.

As for the time stepping error our results in Chapters 3, 4 and 5, show that the global error of Euler's, the implicit Euler and the ETD method is $\mathcal{O}(h)$, while for the modified Euler, the trapezoidal and the ETD2RK methods it is $\mathcal{O}(h^2)$, for all three ICs. However, this is not our expectation in Chapter 1 as we assumed that the discontinuity will impact the time error as well as the spatial error, probably because time stepping methods used here are not very high orders. Noteworthy, there is no time stepping error when using exponential integrators for the linear wave equation, because exponential integrators solve the resulting linear ODEs exactly.

We now move to the stability of time stepping methods for solving the linear and nonlinear wave equations for the spectral methods. Euler's and the modified Euler methods are not stable. The stability of the other two methods, the implicit Euler and the trapezoidal, is assessed in more details, and results show that both methods are stable. In addition, the ETD methods are stable for the linear problem, however, unstable for the nonlinear problem, but the instability is less severe than for Euler's method. On the other hand, the ETD2RK method is stable for the nonlinear wave equation. More details are coming next about some methods.

In Chapter 4 another aspect is thoroughly investigated, namely the energy when solving the nonlinear wave equation (4.1) of the exact solution and the numerical solution (of the trapezoidal method with pseudospectral method) as well the stability of the numerical solution, in addition to the convergence described above. It is shown that the exact solution of the nonlinear problems has the energy conserved, **known result** from this chapter. The energy for the explicit Euler, implicit Euler and trapezoidal methods are studied, both the explicit and implicit Euler methods do not preserve energy. The trapezoidal rule is conjugate symplectic,

so it almost conserves energy when applied to Hamiltonian ODEs. Our numerical results show that when applied to the semi-discretized nonlinear wave equation, the trapezoidal rule also almost conserves energy. It follows that the trapezoidal method is stable. However, this was demonstrated only for the three initial conditions considered here.

We identified three main sources of numerical errors in our numerical computation for the energy when using the spectral methods for our ICs, which are; the step size h , the number of modes N and the number of points that are used for the trapezoid rule integration, which are equally spaced quadrature points denoted by nx . For all of these three initial conditions the energy error decreases to zero as the step size h decreases to zero. More precisely, the energy error behaves like $\mathcal{O}(h^2)$ as expected for second order methods: the energy error decreases by a factor of 100 when step size h decreases by a factor of 10. The number of modes N has a big effect on the results of these calculations, and the effect varies depending on the smoothness of the solution. For the first IC, (3.81), it suffices to take N around 20 as this IC (3.81) is exponentially convergent with N . However, the other two ICs are algebraically convergent with N , with order about $\frac{1}{N^3}$ for (3.82) and order $\frac{1}{N^5}$ for (3.83). Therefore, the number of modes must be large for the last two ICs especially for the second IC, which is giving the most nonsmooth solution. For the third IC, (3.83), a good choice is to be $N = 100$; for consistency, we use the same value of the first IC. In contrast, for the second IC (3.82) it must be around $N = 1000$. The other source of numerical error namely the number of points nx is also important. This source of error does not affect that much the first IC, but it does affect the other two ICs especially the IC (3.82). Therefore, for the IC (3.81) even choosing $nx = 200$, for $N = 100$, we get that the energy conserved for h 's used. While for the IC (3.82) we take the number of quadrature points to be very large, to be something like $nx = 100N$. In addition, for the third IC (3.83) we take $nx = 100N$ and this helps for small values of h , e.g. $h = 10^{-4}$.

In Chapter 5 the energy is again thoroughly investigated when solving the linear and nonlinear wave equations of the numerical solution, of exponential integrators with pseudospectral method, also the stability of the numerical solution, in addition to the convergence described above. Here as well it is required to take a very large number of modes, N , and points, nx , the same as for the trapezoidal method in Chapter 4. For the linear problem, the ETD method solves the resulting linear

ODEs of the linear wave equation exactly, thus, exponential integrators for the linear wave equation have the energy conserved; it is exactly as the energy of the semi-discretized system, constant up to an error. Thus, the ETD method is stable. While for the nonlinear wave equation the results are different. The ETD1 method does not preserve energy but shows a drift in energy, it increases gradually. Thus, the ETD1 method is not stable. In addition, the results of the ETD2RK method show that the method does not preserve energy but show a drift in energy, it decreases, so the overall solution amplitude decays to zero, but on an extremely long time scale. Therefore, ETD2RK is stable for the nonlinear wave equation.

From chapters 4 and 5 we have seen the impacts of nonsmoothness on the spectral methods in simulating the convergence, stability and energy, when using both time steps, the trapezoidal method and exponential integrators. It causes to require a very large number of modes, N , and points, nx , when the discontinuity is present to get the energy to be almost preserved, as explained above. The most important thing to remember is that using different values than mentioned above, for N and nx , will result in an inaccurate outcome. To the best of our knowledge these specific details of choosing N and nx to get the energy almost preserved are **novel results**, in consideration of nonsmoothness.

The following is a brief summary of the above results and contributions:

A crucial fact to know is that the exact solution of the linear/nonlinear problems has these different properties, which are unrelated:

- The property related to the smoothness, that **the solution is infinitely smooth at the special times**, $t = 0, 2, 4, 6, 8, \text{etc.}$, **novel result** from Chapters 1 and 5 (for the linear and nonlinear problems, respectively), which has nothing to do with the overall amplitude of the solution.
- The property that **the energy is conserved**. So, **the overall solution amplitude remains a constant**; it does not decay to zero, **known result** from Chapter 4, which has nothing to do with the smoothness of the solution at special times.

According to our numerical results above for the linear and nonlinear wave equations of the solutions spectra and the energy, we conclude the following:

The outcomes of the linear problem show that: The **ETD** method **recovers the smoothness of the solution at the special times** in an amazing way, **novel result** from Chapter 5, far more accurately and faithfully than the trapezoidal method (which does it down to $\mathcal{O}(h^2)$ only). On the other hand, the **ETD** method **simulates the energy exactly**; it is precisely as the energy of the semi-discretized system, constant up to an error.

The outcomes of the nonlinear problem show that: The **EDT2RK** method **recovers the smoothness of the solution at the special times, novel result** from Chapter 5, far more accurately and faithfully than the trapezoidal method (similarly down to $\mathcal{O}(h^2)$ only). Whilst, the decrease in energy for EDT2RK means the overall amplitude of the numerical solution will decay to zero, but obviously on an extremely long time scale. Meanwhile, the **trapezoidal** approach **almost preserves energy**, thus, **the overall solution amplitude remains a constant**; it does not decay to zero. That means, the trapezoidal method simulates the overall solution amplitude structure of nonlinear wave equation more faithfully than EDT2RK method.

Secondly, we discuss the results of the FEM:

In contrast to the above when nonsmoothness has a major impacts for the spectral methods (SM), as we have seen it causes to have three long Chapters, 3, 4 and 5, the impacts are very minor for the finite element methods (FEM), it causes to have only one Chapter 6, as we will see next. The convergence of the FEM is studied as well when solving the nonlinear problem. Our results show that the error with respect to M converges algebraically with order two for all three ICs. Unlike in the spectral methods chapters, the discontinuity, yielded by the second IC (3.82) in the second derivative and by the third IC (3.83) in the fourth derivative, has no impact on the convergence as the order of convergence remains two for all the ICs employed. The discontinuity generated by the second IC (3.82) does have a very minor impact in that the graph is not as a straight line as it is for the other two ICs, for the data used. Therefore, we have not investigated on the convergence using the spectra of the solution as nonsmoothness is not a major deal for the FEM as it is for the SM. Nevertheless similar to Chapters 3 and 4, Euler's method is not stable at all and the trapezoidal rule is stable for the initial conditions studied here. Unsurprisingly, Euler's methods does not preserve energy, similar to the previous

chapters. Whereas, the trapezoidal rule almost conserves energy; and it requires the same small number of elements, M , and quadrature points, nx , for all the ICs implemented. In other words, the nonsmoothness we face has no impact on the energy either. In contrast, the spectral methods require a very large number of modes, N , and points, nx , when the discontinuity is present to get the energy to be almost preserved.

7.2 Future work

The comparative study presented in this thesis concerning the spectral methods and finite element methods to solve some PDEs with solutions having different levels of smoothness can be developed. Some possible future work may consist of:

- Make the present comparison wider by adding the current exponential integrators to the FEM Chapter 6 and do the energy for the linear equation (3.52) of the SM in Chapter 3 and the FEM in Chapter 6.
- One can study this nonlinear wave equation (4.1) using different BCs. That is, imposing Neumann BCs, $\frac{\partial}{\partial x}u(\pm 1, t) = 0$, instead of Dirichlet BCs, $u(\pm 1, t) = 0$, and then studying the discontinuities in the third and fifth derivatives by imposing some new ICs, such as a C^2 IC and C^4 IC. As explained in Chapter 1 the appropriate choice for the extension is even-type extension of $f(x)$. Thus, the solution should converge only algebraically, with order approximately $\frac{1}{N^4}$ and $\frac{1}{N^6}$ for the C^2 IC and C^4 IC, respectively. There are many interesting aspects to investigate. One aspect to see how would the trapezoidal method function in terms of the spectra study when imposing the C^2 IC and the C^4 IC, would we get the fast drop in the plots of the C^2 IC? And what would be the range giving exponential convergence for both these ICs? Another aspect is in investigating the energy, does the SM still require a very large number of modes, N , and quadrature points, nx , when the discontinuity is present to get the energy to be almost preserved? For example, for the C^2 IC we should have $100 < N < 1000$, so what would be exactly and what would be nx required?

- One can also investigate more about the nonsmoothness issue by solving a system of the nonlinear wave equations, such as taking two equations and then coupled them via their nonlinear terms, suppose

$$u_{tt} = u_{xx} - uv^2, \quad (7.1)$$

and

$$v_{tt} = v_{xx} - vu^2. \quad (7.2)$$

Then to set up the initial conditions so that u on its own would be like the C^∞ IC (3.81) but v on its own would be like the C^1 IC (3.82). This coupled system of equations conserve the energy as shown next. It turns out that the energy is given by

$$E(u, v) = \frac{1}{2} \int_{-1}^1 [u_t^2 + u_x^2 + v_t^2 + v_x^2 + u^2v^2] dx. \quad (7.3)$$

Indeed, following the same procedure in Section 4.1 by taking the derivative and then integrating by parts, after few steps, we get

$$\frac{d}{dt}E(u, v) = \int_{-1}^1 [(u_{tt}u_t - u_{xx}u_t + uv^2u_t) + (v_{tt}v_t - v_{xx}v_t + vu^2v_t)] dx + [u_xu_t]_{-1}^1 + [v_xv_t]_{-1}^1.$$

Now substituting our nonlinear coupling equations (7.1) and (7.2) and using our Dirichlet BCs yields

$$\frac{d}{dt}E(u, v) = 0, \quad (7.4)$$

meaning that the energy is preserved. Thus, the question is, would the nonlinear coupling then mix up the discontinuities so that u would end up looking more like v after all?

- To modify the wave equation so that the wave speed is not constant, but is instead a function of x . So, the equation would just be $u_{tt} = c(x)^2u_{xx}$, where $c(x)$ is the prescribed wave speed. Suppose $c(x)$ varies in some smooth fashion across the domain $x = [-1, 1]$, so it is faster at one end and slower at the other. The characteristics of this hyperbolic equation would then no longer be straight lines, but would be curved as shown in Figure 7.1, see the corresponding figure in Chapter 1 (Figure 1.5). Figure 7.1 corresponds to the particular choice $c(x) = e^{-x}$. The characteristics are then given by $\frac{dt}{dx} = \pm \frac{1}{c(x)}$, or $t = \pm \int \frac{1}{c(x)} dx = \pm e^x + constant$ for this choice of $c(x)$.

The \pm indicates waves moving to the right or the left. In either direction, the time to cross the interval is given by $T = \int_{-1}^1 \frac{1}{c(x)} dx = e - e^{-1} \approx 2.35$. In other words, multiples of this T are our new special times where the characteristics touch the boundaries, where the dashed black lines are in the figure, which are different than the special times in this approach. That means, again, at all these new special times the characteristics coming from both boundaries at $t = 0$ would again exactly reach the other boundary. So, the solutions should still be smooth at these new special times. Thus, the question would be, are exponential integrators so good that they could still recover smoothness in this case with the variable wave speed?

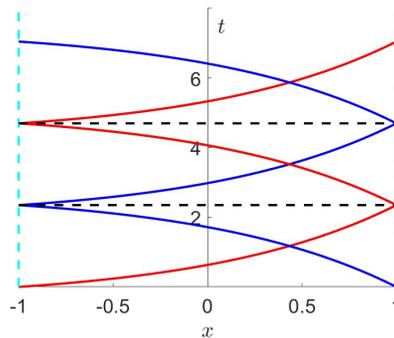


Figure 7.1: The characteristics of the problem in the domain strip $[-1, 1]$.

References

- ABARBANEL, S., GOTTLIEB, D. & TADMOR, E. (1985). *Spectral Methods for Discontinuous Problems*. ICASE report, Institute for Computer Applications in Science and Engineering, National Aeronautics and Space Administration, Langley Research Center. [18](#)
- ARGYRIS, J.H. (1963). Recent advances in matrix methods of structural analysis. *Pergamon Press, New York*. [11](#)
- ATKINSON, K. (1989). *An introduction to numerical analysis*. Second edition, John Wiley and Sons, Inc. [20](#), [49](#)
- BABUSKA, I., WHITEMAN, J. & STROUBOULIS, T. (2010). *Finite elements: an introduction to the method and error estimation*. Oxford University Press. [11](#), [44](#), [185](#), [194](#)
- BAKER, A.J. & PEPPER, D.W. (1991). *Finite Elements 1-2-3*. United Kingdom, McGraw-Hill. [11](#)
- BALDI, P. (2005). Quasi-periodic solutions of the equation $v_{tt} - v_{xx} + v^3 = f(v)$. *arXiv preprint math/0506089*. [15](#)
- BALDI, P. (2007). Bifurcation of free and forced vibrations for nonlinear wave and Kirchhoff equations via Nash-Moser theory. *PhD thesis, SISSA - International School for Advanced Studies*. [15](#)
- BAMBUSI, D. & PALEARI, S. (2001). Families of periodic solutions of resonant PDEs. *Journal of Nonlinear Science*, **11**, 69–87. [15](#), [17](#)
- BECKER, E.B., CAREY, G.F. & ODEN, J.T. (1981). *Finite elements, an introduction Vol 1*. Englewood Cliffs, NJ : Prentice-Hall. [11](#)

- BENMANSOUR, N. (1993). Boundary element solution to stratified shallow water wave Equations. *PhD thesis, University of Portsmouth*. 6
- BERTI, M. & BOLLE, P. (2004). Multiplicity of periodic solutions of nonlinear wave equations. *Nonlinear Analysis: Theory, Methods & Applications*, **56**, 1011–1046. 15
- BERTI, M. & BOLLE, P. (2008). Cantor families of periodic solutions for wave equations via a variational principle. *Advances in Mathematics*, **217**, 1671–1727. 14, 15
- BOBENKO, A.I. & KUKSIN, S.B. (1995). The nonlinear Klein–Gordon equation on an interval as a perturbed sine-Gordon equation. *Commentarii Mathematici Helvetici*, **70**, 63–112. 9
- BOURGAIN, J. (1999). Periodic solutions of nonlinear wave equations. In: Harmonic analysis and partial differential equations. *Chicago Lectures in Math., Univ. Chicago Press, Chicago, IL*. 15
- BOYD, J.P. (2001). *Chebyshev and Fourier spectral methods*. Dover Publications Inc, Mineola, NY., 2nd edn. 1, 10, 11, 12, 13, 14, 45, 60, 65, 183
- BRIDGES, T. & REICH, S. (2006). Numerical methods for Hamiltonian PDEs. *Journal of Physics A: Mathematical and General*, **39**. 54, 119
- BURDEN, R.L. & FAIRES, J.D. (2001). *Numerical analysis*. Thomson Learning, 7th edn. 20, 49, 50, 53
- BURDEN, R.L. & FAIRES, J.D. (2011). *Numerical analysis*. Cengage Learning, 9th edn. 20, 22, 49, 50, 53
- CAI, W., GOTTLIEB, D. & SHU, C.W. (1989). Essentially nonoscillatory spectral fourier methods for shock wave calculations. *Mathematics of Computation*, **52**, 389–410. 18
- CANUTO, C.M., HUSSAINI, M.Y., QUARTERONI, A. & ZANG, T.A. (1988). *Spectral methods in fluid dynamics*. Springer. 12, 13, 25, 45, 57, 58, 59, 60, 63, 65, 66, 68

- CANUTO, C.M., HUSSAINI, M.Y., QUARTERONI, A. & ZANG, T.A. (2006). *Spectral methods fundamentals in single domains*. Springer. 45, 57, 58, 60, 63, 65, 66, 68
- CERTAINE, J. (1960). The solution of ordinary differential equations with large time constants. *Mathematical methods for digital computers*, **1**, 128–132. 155
- CLOUGH, R.W. (1960). The finite element method in plane stress analysis. In *Proceedings of 2nd ASCE Conference on Electronic Computation*, Pittsburg, Pennsylvania, 345–378, Pittsburg, Pennsylvania. 11
- COURANT, R. (1943). Variational methods for the solution of problems of equilibrium and vibrations. *Bulletin of the American Mathematical Society*, **49**, 1–43. 11
- COX, S.M. & MATTHEWS, P.C. (2002). Exponential time differencing for stiff systems. *Journal of Computational Physics*, **176**, 430–455. 24, 155, 156
- CRAIG, W. (2008). Transformation theory of Hamiltonian PDE and the problem of water waves. *Craig, W. (eds). Hamiltonian Dynamical Systems and Applications. NATO Science for Peace and Security Series. Springer, Dordrecht*, 67–83. 119
- CRAIG, W. & WAYNE, C.E. (1993). Newton’s method and periodic solutions of nonlinear wave equations. *Communications on Pure and Applied Mathematics*, **46**, 1409–1498. 15
- DAILEDI, R. C. (2014). Solving the one-dimensional wave equation, part 2. URL: https://http://ramanujan.math.trinity.edu/rdailedi/teach/s14/m3357/lectures/lecture_1_28_slides.pdf. 27, 44
- DE LA HOZ, F. & VADILLO, F. (2008). An exponential time differencing method for the nonlinear Schrödinger equation. *Computer Physics Communications*, **179**, 449–456. 155, 156
- DEGHAN, M. & SHOKRI, A. (2009). Numerical solution of the nonlinear Klein–Gordon equation using radial basis functions. *Journal of Computational and Applied Mathematics*, **230**, 400–410. 8

- DU, Q. & ZHU, W.X. (2004). Stability analysis and application of the exponential time differencing schemes. *Journal of Computational Mathematics*, 200–209. [155](#), [156](#)
- DUTT, P. (1999). Spectral methods for initial boundary value problems with nonsmooth data. *Numerische Mathematik*, **81**, 323–344. [18](#)
- DUTT, P. & SINGH, A. (1993). Spectral methods for periodic initial value problems with nonsmooth data. *mathematics of computation*, **61**, 645–658. [18](#)
- DYM, H. (2013). *Linear algebra in action*. American Mathematical Society. [65](#)
- ECKHOFF, K.S. (1993). Accurate and efficient reconstruction of discontinuous functions from truncated series expansions. *Mathematics of computation*, **61**, 745–763. [18](#)
- ECKHOFF, K.S. (1994). On discontinuous solutions of hyperbolic equations. *Computer methods in applied mechanics and engineering*, **116**, 103–112. [18](#)
- FORNBERG, B. (1998). *A practical guide to pseudospectral methods*. Cambridge University Press. [10](#), [12](#), [13](#), [14](#), [58](#)
- FOX, L. & PARKER, I.B. (1968). *Chebyshev polynomials in numerical analysis*. Oxford University Press. [12](#), [13](#), [58](#)
- FRIEDLANDER, L. (1985). An invariant measure for the equation $u_{tt} - u_{xx} + u^3 = 0$. *Communications in mathematical physics*, **98**, 1–16. [16](#), [17](#)
- FUNARO, D. (1992). *Polynomial approximation of differential equations*. Springer. [12](#)
- GARRAPPA, R. & POPOLIZIO, M. (2011). Generalized exponential time differencing methods for fractional order problems. *Computers & Mathematics with Applications*, **62**, 876–890. [155](#), [156](#)
- GENTILE, G., MASTROPIETRO, V. & PROCESI, M. (2005). Periodic solutions for completely resonant nonlinear wave equations with dirichlet boundary conditions. *Communications in mathematical physics*, **256**, 437–490. [14](#), [15](#)

- GIULIANI, F., GUARDIA, M., MARTÍN, P. & PASQUALI, S. (2020). Chaotic-like transfers of energy in hamiltonian PDEs. *arXiv: Analysis of PDEs*. 15
- GIULIANI, F., GUARDIA, M., MARTIN, P. & PASQUALI, S. (2021). Chaotic-like transfers of energy in hamiltonian PDEs. *Communications in Mathematical Physics*, **384**, 1227–1290. 15
- GOCKENBACH, M.S. (2002). *Partial differential equations: analytical and numerical methods*. *Sima*. **1**, 6, 26, 59
- GOLMANKHANEH, A.K. & BALEANU, D. (2011). On nonlinear fractional Klein–Gordon equation. *Signal Processing*, **91**, 446–451. 9
- GRIGORYAN, V. (2011). Waves on the finite interval. URL: <https://http://web.math.ucsb.edu/~grigoryan/124A/lecs/lec14.pdf>. 27, 44
- GRUNDLAND, A. & INFELD, E. (1992). A family of nonlinear Klein–Gordon equations and their solutions. *Journal of Mathematical Physics*, **33**, 2498–2503. 9
- HAIRER, E., LUBICH, C. & WANNER, G. (2002). *Geometric numerical integration: structure-preserving algorithms for ordinary differential equations*. Springer-Verlag Berlin Heidelberg. 54, 56, 119, 143
- HEPSON, O.E., KORKMAZ, A. & DAG, I. (2016). On the numerical solution of the Klein–Gordon equation by exponential cubic B-spline collocation method. *Communications Faculty of Sciences University of Ankara Series A1 Mathematics and Statistics*, **68**, 412–421. 9
- HESTHAVEN, J.S., GOTTLIEB, S. & GOTTLIEB, D. (2007). *Spectral methods for time-dependent problems*. Cambridge University Press. 45, 57, 58, 60, 65, 66, 69
- HOFFMAN, J.D. & FRANKEL, S. (2001). *Numerical methods for engineers and scientists*. CRC Press. 1, 5
- HRENNIKOFF, A. (1941). Solution of problems of elasticity by the framework method. *Journal of Applied Mechanics*, **8**, 169–175. 11

- HUEBNER, K.H., DEWHIRST, D.L., SMITH, D.E. & BYROM, T.G. (2001). *The finite element method for engineers*. Fourth edition, John Wiley & Sons. 44, 185, 194
- HUNSDORFER, W. & SAVCENCO, V. (2009). Analysis of a multirate theta-method for stiff ODEs. *Applied Numerical Mathematics*, **59**, 693–706. 23
- HURST, G. (2014). Modelling and analysis of ophthalmic fluid dynamics. *PhD thesis, Leeds University*. 68, 69
- HUSSAINI, M.Y., STREETT, C.L. & ZANG, T.A. (1984). Spectral methods for partial differential equations. *II.S. Army Research Office*, 883. 12
- ISERLES, A. (1996). *A first course in the numerical analysis of differential equations*. Cambridge University Press. 19, 52, 53
- KARNIADAKIS, G. & SHERWIN, S. (2005). *Spectral/hp element methods for computational fluid dynamics*. Second edition, Oxford University Press. 11
- KASSAM, A. & TREFETHEN, L.N. (2005). Fourth-order time-stepping for stiff pdes. *SIAM Journal on Scientific Computing*, **26**, 1214–1233. 24, 156
- KRAGH, H. (1984). Equation with the many fathers. The Klein–Gordon equation in 1926. *American Journal of Physics*, **52**, 1024–1033. 9
- KROGSTAD, S. (2005). Generalized integrating factor methods for stiff PDEs. *Journal of Computational Physics*, **203**, 72–88. 24, 156
- LANCZOS, C. (1938). Trigonometric interpolation of empirical and analytical functions. *Studies in Applied Mathematics*, **17**, 123–199. 12
- LANCZOS, C. (1956). *Applied Analysis*. Prentice-Hall. 12
- LANDA, P.S. (1996). *Nonlinear Oscillations and Waves in Dynamical Systems*, vol. 360. Springer Science and Business Media. 8, 9
- LEIMKUHNER, B. & REICH, S. (2004). *Simulating Hamiltonian Dynamics*. Cambridge University Press. 43, 54, 56, 116, 119, 143

- LI, Y., WU, B. & LEOK, M. (2017). Spectral variational integrators for semi-discrete Hamiltonian wave equations. *Journal of Computational and Applied Mathematics*, **325**, 56–73. [142](#)
- LIDSKII, B.V. & SHULMAN, E. (1989). Periodic solutions of the equation $u_{tt} - u_{xx} + u^3 = 0$. *Functional Analysis and Its Applications*, **22**, 332–333. [14](#), [15](#)
- MAJDA, A., MCDONOUGH, J. & OSHER, S. (1978). The Fourier method for nonsmooth initial data. *Mathematics of Computation*, **32**, 1041–1081. [18](#)
- MATHEWS, J.H., FINK, K.D. *et al.* (2004). *Numerical methods using MATLAB*. Prentice Hall, 4th edn. [5](#)
- MCOWEN, R.C. (1996). *Partial differential equations: methods and applications*. Second edition, Prentice Hall. [9](#)
- MERCIER, B. (1989). An Introduction to the Numerical Analysis of Spectral Methods, Lecture Notes in Physics. *Springer Berlin Heidelberg*, **318**. [12](#)
- MERDAN, M. (2014). On the solutions of nonlinear fractional Klein—Gordon equation with modified Riemann-Liouville derivative. *Applied Mathematics and Computation*, **242**, 877–888. [9](#)
- MINCHEV, B. & WRIGHT, W. (2005). A review of exponential integrators for first order semi-linear problems. [24](#), [156](#)
- MINCHEV, B.V. (2004). Exponential integrators for semilinear problems. *PhD thesis, University of Bergen*. [1](#), [24](#), [156](#)
- PALEARI, S., BAMBUSI, D. & CACCIATORI, S. (2001). Normal form and exponential stability for some nonlinear string equations. *Zeitschrift für angewandte Mathematik und Physik ZAMP*, **52**, 1033–1052. [14](#), [15](#)
- PINCHOVER, Y. & RUBINSTEIN, J. (2005). *An introduction to partial differential equations*, vol. 10. Cambridge university press. [41](#)
- PÖSCHEL, J. (1996). Quasi-periodic solutions for a nonlinear wave equation. *Commentarii Mathematici Helvetici*, **71**, 269–296. [14](#)

- PROCESI, M. (2005). Quasi-periodic solutions for completely resonant non-linear wave equations in 1d and 2d. *Discrete and Continuous Dynamical Systems*, **13**, 541–552. [14](#), [15](#)
- RANI, R. & KUMAR, Y. (2016). An analysis on exact solution of Klein—Gordon equation: a relativistic version of the Schrödinger equation. *Journal of advances in science and technology*, **10**. [9](#)
- RAO, S.S. (2005). *The Finite Element Method in Engineering*. Fourth edition, Elsevier Science. [44](#)
- RAUCH, J. (2012). *Hyperbolic partial differential equations and geometric optics*, vol. 133. American Mathematical Soc. [7](#)
- REDDY, J. (1993). *An introduction to finite element methods*. Second edidtion, McGraw-Hill, Inc. [11](#)
- RIVLIN, T.J. (1974). *The Chebyshev polynomials*. John Wiley and Sons, New york. [13](#)
- SANZ-SERNA, J.M. & CALVO, M.P. (1994). *Numerical Hamiltonian problems*. Chapman and Hall. [54](#), [56](#), [119](#), [143](#)
- SCOTT, A. (2003). *Nonlinear science: emergence and dynamics of coherent structures*. Oxford University Press. [9](#)
- SCOTT, A., SRENSEN, M.P. & CHRISTIANSEN, P.L. (1999). *Nonlinear science: emergence and dynamics of coherent structures*. Oxford University Press. [9](#)
- SMITH, G.D. (1985). *Numerical solution of partial differential equations: finite difference methods*. Third edition, Oxford University Press. [11](#)
- SNYDER, M. (1966). *Chebyshev methods in numerical approximations*. Pentice Hall. [13](#)
- SPIJKER, M. (1996). Numerical Stability, Stability Estimates and Resolvent Conditions in the Numerical Solution of Initial Value Problems. *Lecture Notes, Department of Mathematics and Computer Science, Leiden University, The Netherlands*. [25](#)

- STAVROULAKIS, I.P. & TERSIAN, S.A. (2004). *Partial differential equations: an introduction with mathematica and maple*. Second edition, World Scientific Publishing Company. [1](#)
- STOFFER, D. (1988). On reversible and canonical integration methods. *Technical SAM-Report No. 88-05, ETH Zürich*. [56](#)
- STRAUSS, W.A. (1992). *Partial differential equations an introduction*. Second edition, John Wiley and Sons New York, NY, USA. [1](#), [26](#)
- STRAUSS, W.A. (2007). *Partial differential equations: An introduction*. Second edition, John Wiley & Sons. [26](#)
- SU, Y. (1998). Collocation spectral methods in the solution of Poisson equation. *MSc dissertation, British Columbia University*. [10](#), [14](#)
- SÜLI, E. & MAYERS, D. (2003). *An introduction to numerical analysis*. Cambridge University Press. [50](#), [52](#), [53](#)
- TAMSIR, M. & SRIVASTAVA, V. (2016). Analytical study of time-fractional order Klein–Gordon equation. *Alexandria Engineering Journal*, **55**, 561–567. [9](#)
- TARIQ, H. & AKRAM, G. (2017). New approach for exact solutions of time fractional Cahn–Allen equation and time fractional Phi-4 equation. *Physica A: Statistical Mechanics and its Applications*, **473**, 352–362. [9](#)
- TREFETHEN, L. (2013). *Approximation theory and approximation practice*. Siam. [13](#)
- TREFETHEN, L.N. (2000). *Spectral methods in MATLAB*. Siam. [10](#), [14](#), [25](#), [45](#), [60](#), [65](#), [68](#)
- TURNER, M.J., CLOUGH, R.W., MARTIN, H.C. & TOPP, L. (1956). Stiffness and deflection analysis of complex structures. *journal of the Aeronautical Sciences*, **23**, 805–823. [11](#)
- VAN DER VEGT, J. & BOKHOVE, O. (2009). Finite element methods for partial differential equations. *Department of Mathematics, UCLA, Los Angeles (non published article)*. [183](#), [185](#)

- VANDEVEN, H. (1991). Family of spectral filters for discontinuous problems. *Journal of Scientific Computing*, **6**, 159–192. [18](#)
- VASY, A. (2015). *Partial Differential Equations: An Accessible Route Through Theory and Applications*, vol. 169. American Mathematical Society. [26](#)
- WAZWAZ, A. (2008). New travelling wave solutions to the Boussinesq and the Klein—Gordon equations. *Communications in Nonlinear Science and Numerical Simulation*, **13**, 889–901. [8](#)
- WEIDEMAN, J. & TREFETHEN, L.N. (1988). The Eigenvalues of Second-Order Spectral Differentiation Matrices. *Siam Journal on Numerical Analysis*, **25**, 1279–1298. [92](#)
- WHITEMAN, J.R. (1996). *The mathematics of finite elements and Applications*. John Wiley and Sons. [11](#)
- WHITHAM, G.B. (1999). *Linear and nonlinear waves*. John Wiley and Sons. [8](#)
- WHITHAM, G.B. (2011). *Linear and nonlinear waves*, vol. 42. John Wiley and Sons. [8](#)
- YE, C. & ZHANG, W. (2010). New explicit solutions for the Klein—Gordon equation with cubic nonlinearity. *Applied Mathematics and Computation*, **217**, 716–724. [9](#)
- YOMBA, E. (2005). The general projective Riccati equations method and exact solutions for a class of nonlinear partial differential equations. *Chinese Journal of Physics*, **43**, 991–1003. [1](#)
- YUAN, X. (2006). Quasi-periodic solutions of completely resonant nonlinear wave equations. *Journal of Differential Equations*, **230**, 213–274. [15](#)
- ZAUDERER, E. (1983). *Partial differential equations of applied mathematics*. John Wiley and Sons. [7](#)
- ZHAO, S. (1993). Chebyshev spectral methods for potential field computation. *PhD thesis, British Columbia University*. [10](#), [12](#), [13](#), [14](#)