NUMERICAL METHODS AND THEIR ANALYSIS FOR SOME NONLINEAR DISPERSIVE EQUATIONS

DONG XUANCHUN

NATIONAL UNIVERSITY OF SINGAPORE

2012

NUMERICAL METHODS AND THEIR ANALYSIS FOR SOME NONLINEAR DISPERSIVE EQUATIONS

DONG XUANCHUN

(B.Sc., Jilin University)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

NATIONAL UNIVERSITY OF SINGAPORE

2012

Acknowledgements

First and foremost, I owe my deepest gratitude to my supervisor Prof. Bao Weizhu, whose encouragement, patient guidance, generous support, invaluable help and constructive suggestion enabled me to conduct such an interesting research project.

I would like to express my appreciation to my other collaborators for their contribution to the work: Prof. Jack Xin and Mr. Zhang Yong. Special thanks go to Zhang Yong for reading the draft.

My heartfelt thanks go to all the former researchers, colleagues and fellow graduates in our group, for fruitful interactions and suggestions on my research. I sincerely thank my friends, for all the encouragement, emotional support, comradeship and entertainment they offered.

I would also like to thank NUS for awarding me the Research Scholarship which financially supported me during my Ph.D candidature. Many thanks go to IPAM at UCLA, and INIMS at Cambridge, for their financial assistance during my visits.

Last but not least, I am forever indebted to my beloved family, for their encouragement, steadfast support and endless love when it was most needed.

Dong Xuanchun

May 2012

Contents

Summary

The nonlinear dispersive equations, including a large body of classes, are wildely used models for a great number of problems in the fields of physics, chemistry and biology, and have gained a surge of attention from mathematicians ever since they were derived. In addition to mathematical analysis, the numerics of these equations is also a beautiful world and the studies on it have never stopped.

The aim of this thesis is to propose and analyze various numerical methods for some representative classes of nonlinear dispersive equations, which mainly arise in the problems of quantum mechanics and nonlinear optics. Extensive numerical results are also reported, which are geared towards demonstrating the efficiency and accuracy of the methods, as well as illustrating the numerical analysis and applications. Although the subjects considered here is merely a small sample of nonlinear dispersive equations, it is believed that the methods and results achieved for these equations can be applied or extended to more general cases.

The first part of this thesis is concerned with the Schrödinger–Poisson (SP) type equations, which can be derived as the single-particle approximations in taking the mean-field limit of Coulomb many-body quantum systems, in both nonrelativity and relativity theories. First, various numerical methods are proposed and compared for computing the ground states and dynamics of a nonrelativistic SP type equation, with motivation for the systems of electrons (fermions), in all space dimensions. In particular, when the equation is of spherical symmetry, the preferred methods suggested by extensive comparisons in general settings are significantly simplified. Later, as a benefit of the observations drawn in the nonrelativistic problem, efficient and accurate numerical methods are proposed for computing the ground states and dynamics of a SP type equation when relativity is taken into account.

The second part is to understand and compare various numerical methods for solving the nonlinear Klein–Gordon (KG) equation. The nonlinear KG equation might be viewed as the most simplest form of wave equations; however, here it is considered in a nonrelativistic scaling involving a small parameter $\varepsilon > 0$, in which scaling the solutions are highly oscillatory in time. Frequently used second-order finite difference time domain (FDTD) methods are first analyzed, concluding with rigorous and optimal error estimates with respect to the small ε. Then a new numerical integration, namely a Gautschi-type exponential wave integrator in time advances, is proposed and analyzed. Rigorous and optimal error estimates show that the Gautschi-type integrator offers compelling advantages over those FDTD methods regarding the meshing strategy requirement for resolving the oscillation structure.

The last part is to investigate the sine–Gordon (SG) equation and perturbed nonlinear Schrödinger (perturbed NLS) equation for modeling the light bullets in two space dimensions. Here, the primary focus is in the time regime beyond the collapse time of critical (cubic focusing) NLS equation. To this purpose, efficient and accurate numerical methods are proposed with rigorous error estimates. Comprehensive comparisons among the light bullets solutions of the SG, perturbed NLS and critical NLS equations are carried out. The results validate people's anticipation that cubic NLS fails to match SG well before and beyond the collapse time, whereas the perturbed NLS still agrees with SG beyond the critical collapse. Consequently, propagation of light bullets over long time is traced by solving the perturbed NLS equation.

List of Tables

List of Figures

List of Symbols and Abbreviations

1D, 2D and 3D	One, two and three-dimensional space
BEFC/TSFC	backward Euler/Time-splitting fast convolution
BEFP/TSFP	backward Euler/Time-splitting Fourier pseudospectral
BESP/TSSP	Backward Euler/Time-splitting sine pseudospectral
CNGF	continuous normalized gradient flow
c.c.	complex conjugate of previous term
FDTD	finite difference time domain
FFT	fast Fourier transform
FMM	fast multipole method
FST	fast sine transform
f^*	conjugate of a complex function f
$f * g$	convolution of function f with function g
Gautschi-FD/-SP	Gautschi-type exponential wave integrator finite difference/sine pseudospectral
GFDN	gradient flow with discrete normalization
h	mesh size
$\mathcal{I}$	interpolation operator
i	imaginary unit

KG	Klein–Gordon
LBs	light bullets
NLS	nonlinear Schrödinger
$\mathcal{P}$	projection operator
$p \lesssim q$	$\lvert p\rvert \le Cq$ for some generic constant C
RSP	relativistic Schrödinger–Poisson
SG	sine–Gordon
SN	Schrödinger–Newton
SP	Schrödinger–Poisson
SPS	Schrödinger–Poisson–Slater
t	time variable
$W_{p,q}$	standard Sobolev space
$\mathbf{x} = (x, y, z)^T$	Cartesian coordinate
τ	time step
∇	gradient
$\Delta = \nabla \cdot \nabla$	Laplacian

Chapter 1

Introduction

The term *dispersion*, occurring in a partial differential equation, generally refers to a frequency-dependent phenomenon in its wave propagation [33, 38, 103, 122, 142, 143]. It accounts for the fact that different frequencies in this equation tend to propagate at different phase velocities; and thus, a wave packet of mixed wavelengths tends to spread out in space over time. Dispersive equations are in contrast to transport equations, in which various frequencies travel at the same velocity, or dissipative equations such as the heat equation, in which frequencies do not propagate but instead simply attenuate to vanish.

1.1 Motivations of the study

The applications of dispersive equations are found in many branches of physical sciences from fluid dynamics, quantum machines, plasma physics to nonlinear optics and so forth, and in chemistry and biology as well [103, 122]. For instance, the Korteweg-de Vries equation and its various modifications serve as the modeling equations in several physical problems, such as the Fermi–Pasta–Ulam problem and the evolution of one-dimensional (1D) long waves in many settings [122, 124]. The Schrödinger equation is the fundamental governing equation in quantum machines and quantum field theory [33, 38, 46, 128, 142], which is used to describe, for example, many-body theory and condensed matter physics like the Bose–Einstein condensate. It is also a classical field equation with extensive applications to optics [6, 119] and water waves [33, 38, 142]. Also, certain problems in chemistry and biology obey the Schrödinger–Poisson type equations [27, 76]. The nonlinear wave equations such as the Klein–Gordon equation and sine–Gordon equation arise in the fields from acoustics, electromagnetics, fluid dynamics, to relativity in physics [3,35,36,122,143].

Over the past few decades, an extensive body of studies have contributed to the mathematical theories of various classes of dispersive equations; and the analytical results, like local and global well-posedness theory, existence and uniqueness of stationary states and so forth, are rich and vast in the literature (see, e.g., some recent monographs on this topic [103, 122, 143]). In parallel with the analytical studies, a surge of efforts have been devoted to the numerics of these equations, which is a topic of great interests from the point of view of concrete real-world applications to physics and other sciences. Although the numerical approximation of solutions of differential equations is a traditional topic in numerical analysis, has a long history of development and has never stopped, it remains as the beating heart in this field that to propose more sophisticated numerical methods for dispersive equations.

For some nonlinear dispersive equations, the computation concern involves several challenges. For example, long-time simulations call for much efficient and stable temporal solvers since the round-off error in discretizing dispersive equations will accumulate dramatically for the discretization with poor stability. And, applications to real-world problems in two or three space dimensions (2D, 3D) give rise to a demand placed on the spatial discretizing formulations with high resolution capacity and low computational and memory cost. Also, in some singular limit regimes (like semi-classical limit, nonrelativistic limit, subsonic limit, and so forth), the oscillatory nature inherent in the solutions would build up severe numerical burdens. In the scenario that oscillation occurs, even for those stable discretizations the oscillations may very well pollute the solutions unless the oscillatory profiles are fully resolved numerically, i.e., using many grid points per wavelength.

These potentials in applications and challenges in numerical solutions propel this study. In this work, the focus is put on some specific classes of nonlinear dispersive equations, which will be discussed in a nutshell in the forthcoming section.

1.2 The subjects

This thesis focuses primarily on five equations: the Schrödinger–Poisson–Slater equation, the nonlinear relativistic Hartree equation, the nonlinear Klein–Gordon equation, the sine–Gordon equation, and the perturbed nonlinear Schrödinger (perturbed NLS) equation. The former two equations can be viewed as the single-particle approximations, in the mean-field theory, of the multi-body quantum systems with Coulomb interaction in nonrelativity and relativity theories, respectively, from the point of view of mathematical physics. In fact, the relativistic Hartree equation is also called the relativistic Schrödinger–Poisson equation, which is a degenerate case of Schrödinger–Poisson–Slater and valid only for bosons. The nonlinear Klein–Gordon equation is considered in a nonrelativistic limit scaling, which explicitly leaves the inverse of the speed of light as a small parameter. The last two equations are investigated with motivation of their applications to nonlinear optics for modeling 2D localized optical pulses, i.e., the so-called 2D *light bullets*. These five equations are of course only a very small sample of the nonlinear dispersive equations, but they are reasonably representative in that the numerics of them showcase many of the techniques applicable or generalizable for more general equations.

I. The Schrödinger–Poisson–Slater equation

The Schrödinger–Poisson–Slater (SPS) equation, also named as the Schrödinger–Poisson–Xα equation, serves as a local single-particle approximation of the time-dependent Hartree-Fock system as the mean-field equations of N-particle quantum systems [23, 32, 111]. It reads, in scaled form,

$$i\partial_t\psi(\mathbf{x},t) = \left[-\frac{1}{2}\Delta + V_{\text{ext}}(\mathbf{x}) + C_P V_P \; - \alpha|\psi|^{\frac{2}{d}}\right]\psi, \quad t>0, \tag{1.1}$$

$$\Delta V_P(\mathbf{x},t) = -|\psi|^2, \quad \mathbf{x}\in\mathbb{R}^d \ (d=1,\ 2,\ 3), \quad t\geq 0, \tag{1.2}$$

with the following initial condition for dynamics

$$\psi(\mathbf{x},0) = \psi_0(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^d. \tag{1.3}$$

Here, the complex-valued function $\psi(\mathbf{x},t)$ (t is time, $\mathbf{x}$ is the Cartesian coordinates) with $\lim_{|\mathbf{x}|\to\infty}|\psi(\mathbf{x},t)| = 0$ stands for the single-particle wave function, $V_{\text{ext}}(\mathbf{x})$ is a given external potential, for example a confining potential, $V_P(\mathbf{x},t)$ denotes the Hartree potential with the same asymptotic far-field behavior as the fundamental solution of Poisson equation in $\mathbb{R}^d$, and C_P and α are interaction constants. The sign of Poisson constant C_P depends on the type of interaction considered: $C_P > 0$ in the repulsive case and $C_P < 0$ in the attractive case. Physically, the Slater constant $\alpha > 0$ for electrons. Note that if the Slater term is not considered, i.e. $\alpha = 0$, then the SPS equation (1.1)–(1.3) coincides with the Schrödinger–Poisson (SP) equation. Also, the attractive SP equation, i.e. (1.1)–(1.3) with $C_P < 0$ and $\alpha = 0$, is usually called as the Schrödinger–Newton (SN) equation which describes the particle moving in its own gravitational potential. Note that the rigorous derivation of SP equation, as a mean-field approximation, is only valid for bosons in that it disregards the "Pauli exclusion principle" for fermions. Derivation of the SPS equation (1.1)–(1.3) as an effective approximation of a Coulomb system of N electrons will be discussed in Chapter 2.

The SPS equation (1.1)–(1.2) is equivalent to a nonlinear Schrödinger (NLS) equation:

$$i\partial_t\psi(\mathbf{x},t) = \left[-\frac{1}{2}\Delta + V_{\text{ext}}(\mathbf{x}) + C_P V_P\left(|\psi|^2\right) - \alpha|\psi|^{\frac{2}{d}}\right]\psi. \tag{1.4}$$

Here, the Hartree potential V_P is rewritten as a function of $|\psi|^2$,

$$V_P\left(|\psi|^2\right) = G_d(\mathbf{x}) * |\psi|^2, \tag{1.5}$$

where $G_d(\mathbf{x})$ denotes the Green's function of the Laplacian on $\mathbb{R}^d$ $(d = 1,\ 2,\ 3)$:

$$G_d(\mathbf{x}) = \begin{cases} -\frac{1}{2}|\mathbf{x}|\,, & d = 1, \\ -\frac{1}{2\pi}\ln(|\mathbf{x}|)\,, & d = 2, \\ \frac{1}{4\pi}|\mathbf{x}|^{-1}\,, & d = 3. \end{cases} \tag{1.6}$$

In addition, the initial condition is usually normalized under the *normalization condition* by a proper non-dimensionalization

$$\|\psi_0\|^2 := \int_{\mathbb{R}^d} |\psi_0(\mathbf{x})|^2 \,\mathrm{d}\mathbf{x} = 1. \tag{1.7}$$

Part of this study will deal with the computation for the dynamics of the SPS equation and its ground states, i.e., one particular class of stationary states which minimize the total energy functional of the equation in its energy space under the normalization constraint (1.7).

II. The nonlinear relativistic Hartree equation for boson stars

The nonlinear relativistic Hartree equation in 3D, i.e. the relativistic Schrödinger–Poisson equation, is given as [55, 96, 97]

$$i\partial_t\psi(\mathbf{x},t) = \sqrt{-\Delta + m^2}\ \psi + V_{\mathrm{ext}}(\mathbf{x})\psi + \lambda\left(|\mathbf{x}|^{-1} * |\psi|^2\right)\psi,\ \ \mathbf{x} \in \mathbb{R}^3,\ \ t > 0, \tag{1.8}$$

with the following initial condition for dynamics

$$\psi(\mathbf{x}, 0) = \psi_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^3. \tag{1.9}$$

Here, t is time, $\mathbf{x} = (x, y, z)^T$ is the Cartesian coordinates, $\psi = \psi(\mathbf{x}, t)$ is a complex-valued dimensionless single-particle wave function, a real-valued function $V_{\mathrm{ext}}(\mathbf{x})$ stands for an external potential, $m \geq 0$ denotes the scaled particle mass ($m = 1$ in most cases) with $m = 0$ corresponding to massless particles, and $\lambda \in \mathbb{R}$ is a dimensionless constant describing the interaction strength. The sign of λ depends on the type of interaction: positive for the repulsive interaction and negative for the attractive interaction. The pseudodifferential operator $\sqrt{-\Delta + m^2}$ for the kinetic energy is defined via multiplication in the Fourier space with the symbol $\sqrt{|\xi|^2 + m^2}$

for $\xi \in \mathbb{R}^3$, which is frequently used in relativistic quantum mechanical models as a convenient replacement of the full (matrix-valued) Dirac operator [9,55,96,97]. The symbol $*$ stands for the convolution in $\mathbb{R}^3$.

The above nonlinear relativistic Hartree equation (1.8) was rigorously derived recently in [55] for a quantum mechanical system of N bosons with relativistic dispersion interacting through a gravitational attractive or repulsive Coulomb potential, which is often referred to as a boson star. Also, the initial condition is usually normalized under the *normalization condition* by a proper non-dimensionalization

$$\|\psi_0\|^2 := \int_{\mathbb{R}^3} |\psi_0(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x} = 1. \tag{1.10}$$

Again, the concern here is the computation for its dynamics and ground states.

III. The nonlinear Klein–Gordon equation in the nonrelativistic limit regime

The Klein–Gordon equation, which is also known as the relativistic version of the Schrödinger equation, describes the motion of a spinless particle with mass $m > 0$ (see, e.g. [46, 128], for its derivation). Denoting by c the speed of light and $\hbar$ the Planck constant, the nonlinear Klein–Gordon (KG) equation reads

$$\frac{\hbar^2}{mc^2}\partial_{tt}u - \frac{\hbar^2}{m}\Delta u + mc^2 u + g(u) = 0, \quad \mathbf{x} \in \mathbb{R}^d \ (d = 1,\ 2,\ 3), \quad t > 0, \tag{1.11}$$

where, $u = u(\mathbf{x}, t)$ is a real-valued field and $g(u)$ is a real-valued function, independent of c and m, describing the nonlinear interaction and satisfying $g(0) = 0$.

By introducing the dimensionless variables in (1.11): $t \to \frac{\hbar}{m\varepsilon^2 c^2}t$ and $\mathbf{x} \to \frac{\hbar}{m\varepsilon c}\mathbf{x}$ with a dimensionless parameter $\varepsilon > 0$ which is inversely proportional to the speed of light c, the following dimensionless KG equation is obtained,

$$\varepsilon^2\partial_{tt}u - \Delta u + \frac{1}{\varepsilon^2}u + f(u) = 0, \quad \mathbf{x} \in \mathbb{R}^d, \quad t > 0, \tag{1.12}$$

with initial conditions given as

$$u(\mathbf{x}, 0) = \phi(\mathbf{x}), \quad \partial_t u(\mathbf{x}, 0) = \frac{1}{\varepsilon^2}\gamma(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d. \tag{1.13}$$

Here, ϕ and γ are given real-valued functions and $f(u)$ is a dimensionless real-valued function independent of ε and satisfying $f(0) = 0$.

The KG equation (1.12) in the $O(1)$-speed of light regime, i.e., for fixed $\varepsilon > 0$, has been extensively studied in the literature. This study will mainly work in the regime that $0 < \varepsilon \ll 1$ (i.e. if the speed of light goes to infinity), under which limit the issues become substantially complicated in that in this regime the solutions are highly oscillating in time. In fact, the solutions are propagating waves with wavelength of $O(\varepsilon^2)$ and $O(1)$ in time and space, respectively.

IV. Sine–Gordon and perturbed NLS equations for light bullets

The *light bullets* (LBs), i.e., spatially localized particle-matter optical pulses, have been observed in the numerical simulations of the full Maxwell system with instantaneous Kerr ($\chi^{(3)}$ or cubic) nonlinearity in 2D [70]. Recently, by examining a distinguished asymptotic limit of the two level dissipationless Maxwell–Bloch system in the transverse electric regime, Xin [149] found that the well-known (2+1) sine–Gordon (SG) equation

$$\partial_{tt}u(\mathbf{x},t) - c^2\Delta u + \sin(u) = 0, \quad t > 0, \tag{1.14}$$

with initial conditions

$$u(\mathbf{x},0) = u^{(0)}(\mathbf{x}), \quad \partial_t u(\mathbf{x},0) = u^{(1)}(\mathbf{x}), \quad \mathbf{x} = (x,y) \in \mathbb{R}^2, \tag{1.15}$$

where $u(\mathbf{x},t)$ is a real-valued function and c is a given constant, has its own LBs solutions.

On the other hand, a new and complete perturbed NLS equation was also derived in [149] by Xin via removing all resonance terms (*complete NLS approximation*) in carrying out the envelope expansion of the SG-LBs solutions. Upon a proper rescaling, the perturbed NLS equation derived in [149] reads

$$i\partial_T A(\mathbf{X},T) - \frac{\varepsilon^2}{4\omega^2}\partial_{TT}A = -\Delta A - \frac{\varepsilon ck}{\omega}\partial_{XT}A + f_\varepsilon\big(|A|^2\big)A, \quad T > 0, \tag{1.16}$$

with initial conditions,

$$A(\mathbf{X},0) = A^{(0)}(\mathbf{X}), \quad \partial_T A(\mathbf{X},0) = A^{(1)}(\mathbf{X}), \quad \mathbf{X} \in \mathbb{R}^2, \tag{1.17}$$

where, $A(\mathbf{X},T)$ $(\mathbf{X}=(X,Y)\in\mathbb{R}^2)$ is a complex-valued function, and

$$\rho=|A|^2, \quad f_\varepsilon(\rho)=\sum_{l=0}^{\infty}\frac{(-1)^{l+1}\varepsilon^{2l}\rho^{l+1}}{(l+1)!(l+2)!}. \tag{1.18}$$

In this study, numerical comparisons will be carried out among the LBs solutions of the SG equation (1.14), the perturbed NLS equation (1.16) and its finite terms approximation in nonlinearity, and the critical (cubic focusing) NLS equation ($\varepsilon=0$ in (1.16)).

1.3 Overview of the thesis

Each subsequent chapter is devoted to one of the mentioned subjects. For each problem, various classes of numerical methods will be proposed and compared, and some of them will be rigourously analyzed in the concepts of stability and convergence.

The first part of Chapter 2 is devoted to the computation of ground states and dynamics of the Schrodinger–Poisson–Slater (SPS) equation (1.1)–(1.2) (or equivalently (1.4)–(1.5)) with general external potential and initial condition. To this end, efficient numerical methods, namely backward Euler and time-splitting pseudospectral methods are proposed for the NLS equation (1.4) with the nonlocal Hartree potential (1.5) approximated by various approaches. These approaches include fast convolution algorithms, which are accelerated by using FFT in 1D and fast multipole method (FMM) in 2D and 3D, and sine/Fourier pseudospectral methods. Numerical comparisons among all these approaches show that the methods based on sine pseudospectral formulation are the best candidates. Applications of the backward Euler and time-splitting sine pseudospectral methods to study the ground states and dynamics in different setups are also reported. The second part of Chapter 2 is concerned with the case that the external potential and initial condition are spherically symmetric. For the SPS equation with spherical symmetry, via applying a proper change of variables into the reduced quasi-1D model, the methods proposed for the general 3D case are simplified, such that both the memory and computational load are significantly reduced.

Chapter 3, to some extents, can be regarded as one application of the observations drawn in Chapter 2; in this chapter, efficient and accurate numerical methods are presented for computing the ground states and dynamics of 3D nonlinear relativistic Hartree equation (1.8) for boson stars. Those preferred numerical methods discussed in Chapter 2 are extended to the relativistic Hartree equation, i.e. relativistic Schrödinger–Poisson equation ($\alpha = 0$ in (1.4)). Also, when the external potential and initial data for dynamics are spherically symmetric, the original 3D problem collapses to a quasi-1D problem, for which the 3D spectral-type methods are extended and simplified successfully with a proper change of variables. Extensive numerical results are also reported to demonstrate the spectral accuracy of the methods and to show very intriguing and complicated phenomena in the mean-field dynamics of boson stars.

Chapter 4 analyzes rigourously error estimates and compares numerically temporal/spatial resolution of various numerical methods for solving the Klein–Gordon equation (1.12) in the nonrelativistic limit regime ($0 < \varepsilon \ll 1$). We begin with four frequently used finite difference time domain (FDTD) methods and obtain their rigorous error estimates for $0 < \varepsilon \ll 1$. The results show that, besides of the second-order accuracy, in order to compute "correct" solutions when $0 < \varepsilon \ll 1$, the four FDTD methods follow the same meshing strategy requirement: $\tau = O(\varepsilon^3)$ (τ is time step). Then new numerical methods are proposed by using either sine pseudospectral or finite difference approximation for spatial derivatives combined with the Gautschi-type exponential wave integrator for temporal derivatives. The new methods are unconditionally stable and their meshing strategy requirement is loosen to $\tau = O(1)$ and $\tau = O(\varepsilon^2)$ for linear and nonlinear problems, respectively, which is also rigorously proved.

In Chapter 5, the sine–Gordon (SG) equation (1.14) and the perturbed NLS equation (1.16) are studied numerically for modeling the 2D *light bullets* (LBs).

We begin with the derivation of the perturbed NLS equation (1.16) for the SG-LBs envelopes, which is globally well-posed and has all the relevant higher order terms to regularize the collapse of the standard critical (cubic focusing) NLS equation ($\varepsilon = 0$ in (1.16)), followed by the discussion that the perturbed NLS equation (1.16) is approximated by truncating the saturating nonlinearity into finite higher order terms undergoing focusing-defocusing cycles. Efficient methods for solving the SG and perturbed NLS equations are proposed with rigorous error estimates. Numerical comparison results validate that the LBs solutions of the perturbed NLS equation and its finite-term truncations are in qualitative and quantitative agreement with the ones of the SG equation even beyond the critical collapse time of the cubic focusing NLS equation. In contrast, the critical NLS-LBs is in qualitative agreement with the SG-LBs merely before the collapse time. As a benefit of such observations, LBs propagations are studied via solving the perturbed NLS equation truncated by reasonably many nonlinear terms, which is a much cheaper task than solving the SG equation directly.

Finally, the main results obtained for these subjects are summarized in Chapter 6. Also, some interesting topics for further work are addressed in Chapter 6.

Chapter 2

Methods for the Schrödinger–Poisson–Slater equation

In this chapter, various classes of efficient numerical methods are proposed and compared for computing the ground states and dynamics of the Schrödinger–Poisson–Slater (SPS) equation (1.1)–(1.2) (or equivalently (1.4)–(1.5)). The first part of this chapter (Sections 2.2 and 2.3) is concerned with the case of general external potential V_{ext} and initial condition in (1.3), and different methods are discussed there. The second part (Section 2.4) is devoted to a special 3D case where the SPS equation is of spherical symmetry.

2.1 The SPS equation: derivation and contemporary studies

One of the fundamental problems of many body quantum mechanics is seeking for the approximation of exact N-body problems by simpler models, in particular single-body equations. The following will sketch the formal derivation of the SPS equation (1.1)–(1.3) as an effective time-dependent single-particle approximation of a quantum system of N electrons interacting via Coulomb potential, with a local exchange correction term to the so-called mean-field approximation.

The linear Schrödinger equation for the wave function $\Psi = \Psi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N, t)$ of N electrons interacting via the Coulomb potential reads

$$i\partial_t \Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N, t) = -\frac{1}{2}\sum_{j=1}^{N} \Delta_{\mathbf{x}_j} \Psi + \sum_{j=1}^{N} \sum_{k=j+1}^{N} \frac{1}{|\mathbf{x}_j - \mathbf{x}_k|} \Psi, \quad t > 0, \tag{2.1}$$

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N, 0) = \Psi_0(\mathbf{x}_1, \ldots, \mathbf{x}_N), \quad \mathbf{x}_j \in \mathbb{R}^3, \quad j = 1, \ldots, N. \tag{2.2}$$

Here, the Planck constant, the mass and other physical constants are kept fixed and scaled to 1. To obtain the mean-field approximation from (2.1), the Hartree ansatz for the N-particle wave function Ψ, i.e.,

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N, t) = \Pi_{j=1}^{N} \psi(\mathbf{x}_j, t), \quad j = 1, \ldots, N, \tag{2.3}$$

yields the Schrödinger–Poisson (SP) equation ($\alpha = 0$ in (1.1)), for which rigorous derivations were given recently in [10] for the stationary case, and respectively, in [12] for the time-dependent case. However, the Hartree ansatz (2.3) writes the N-particle wave function as a simple product of single-particle wave functions; hence, in the SP model the "Pauli exclusion principle" for fermions is disregarded (the SP model is thus only valid for bosons), and the exchange effects of electrons are missing.

In contrast, the Hartree–Fock (HF) ansatz takes the N-particle wave function as a Slater determinant:

$$\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N, t) = \frac{1}{\sqrt{N!}} \det\left(\psi_j(\mathbf{x}_k, t)\right)_{j,k=1,\ldots,N}, \tag{2.4}$$

which vanishes for two particles occupying the same position, and thus realizes the antisymmetrization of the N-particle wave function so that the Pauli principle is respected. In the context of minimizing the total energy of an N-body system (therefore the variable t is not taken into account), with the HF ansatz (2.4), the original N-body problem reduces to a system of N coupled stationary one-electron Schrödinger equations. The stationary HF equations for the set of N orthonormal single-particle wavefunctions ψ_j are

$$-\frac{1}{2}\Delta\psi_j + V_{\text{ext}}\psi_j + V_P\psi_j + (V_{\text{exc}}\psi)_j = E_j\psi_j, \quad j = 1, \ldots, N, \tag{2.5}$$

where E_j is the j-th eigenvalue, V_{ext} refers to some given external potential, V_P is the Hartree potential with the local density ρ:

$$V_P(\mathbf{x}) := \int_{\mathbb{R}^3} \frac{\rho(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}\mathrm{d}\mathbf{y}, \quad \rho(\mathbf{x}) := \sum_{j=1}^{N} |\psi_j(\mathbf{x})|^2, \quad \mathbf{x} \in \mathbb{R}^3, \tag{2.6}$$

and $(V_{\text{exc}}\psi)_j$ stands for the exchange term, defined by

$$(V_{\text{exc}}\psi)_j(\mathbf{x}) := -\sum_{k=1}^{N} \left[\int_{\mathbb{R}^3} \frac{\psi_j(\mathbf{y})\psi_k^*(\mathbf{y})}{|\mathbf{x}-\mathbf{y}|}\mathrm{d}\mathbf{y}\right] \psi_k(\mathbf{x}), \quad j = 1, \ldots, N. \tag{2.7}$$

This HF model has been used to analyze vast phenomena in quantum chemistry and solid state physics. For the rigorous analysis of the stationary HF system, one can refer to [104] and references therein. For the time-dependent case, the HF equations formulated for the density matrix were rigorously derived by means of "mean-field limits" in [13] for the bounded interactions and, respectively, in [14] for the Coulomb case.

The HF equations (2.5) are too complex for numerical simulations since the nonlocal exchange term (2.7) is quite costly to calculate. Slater in [135] gave one simple approximation to the exchange term (2.7), which is $(V_{\text{exc}}\psi)_j(\mathbf{x}) = C\rho^\alpha\psi_j$ with $\alpha = 1/3$ and some constant C. This local expression was actually first introduced implicitly by Dirac while considering the exchange energy as a correction in the Thomas–Fermi model [49]. Such kind of ρ^α approximation is usually named as Xα-approach, in which α is taken as a parameter and differs as various limits. Such local approximation to the nonlocal HF exchange potential provides excellent results in the study of stationary states [51, 100, 101]. The rigorous derivations of this Xα approximation in the stationary case were given in [30,31] and the argument in time-dependent case is still an active research topic.

Therefore, so far only the SP equation has been rigorously derived as the time-dependent single-particle approximation. Hence, it is imperative to find appropriate corrections to the mean-field potential in the SP model so as to take into account the exchange effects. To this end, taking the more or less rigorously derived expression of the stationary case and hence adding the local Xα-approximated exchange term (with $\alpha = 1/d$ for the problem in $d = 1, 2, 3$ space dimensions according to the derivation in [13]) with t as an additional variable, to the effective potential in the SP model, the SPS model (1.1)–(1.3) was proposed in [111].

There are at least two important invariants of (1.1)–(1.2) or equivalently (1.4): the *mass of particles*

$$N(\psi(\cdot, t)) := \|\psi(\mathbf{x}, t)\|^2 = \int_{\mathbb{R}^d} |\psi(\mathbf{x}, t)|^2 \, \mathrm{d}\mathbf{x} = 1, \quad t \geq 0, \tag{2.8}$$

and the *total energy*

$$\begin{aligned} E(\psi(\cdot, t)) &:= \int_{\mathbb{R}^d} \left[\frac{1}{2} |\nabla\psi|^2 + \left(V_{\text{ext}}(\mathbf{x}) + \frac{C_P}{2} V_P(|\psi|^2) \right) |\psi|^2 - \frac{\alpha d}{d+1} |\psi|^{\frac{2}{d}+2} \right] \mathrm{d}\mathbf{x} \\ &\equiv E(\psi_0), \quad t \geq 0. \end{aligned} \tag{2.9}$$

The NLS equations have drawn a surge of attention from mathematicians, and for an overview of this subject one can refer to [33, 38, 142]. Also, there is a series of analytical results on the SPS equation in the literature. For (1.1)–(1.2) (or (1.4)), by the standard results in [38] the global existence of a unique solution in its energy space H^1 can be established for 3D [30]. The existence theory in 1D was given in [138] and the analysis in 2D was recently announced in [109]. Another interesting problem is the existence and uniqueness of the ground states, i.e. the solutions which minimize the total energy functional (2.9) under the normalization constraint (1.7). For the most simple-looking equation in the form of (1.1)–(1.3), i.e. the SN equation without external potential, the existence of a unique spherically symmetric ground state in 3D was proven by Lieb in [99], and in any dimension $d \leq 6$ was given in [43]. There is no global minimum of the energy functional for the repulsive SP equation without external potential since the infimum of its energy is always zero. When the Slater term in (1.4) is considered and in the absence of any external potential, the existence analysis of ground states in 3D was given in [129], and in particular the existence of a unique spherically symmetric ground state is proven in [32] for the attractive case. To our knowledge, so far the existence analysis of higher bound states remains open.

Along the numerical front, self-consistent solutions of the SPS equation are important in the simulations of a quantum system. For example, time-independent SP equation was solved in [27, 39] for the eigenstates of the quantum system, and time-dependent spherically symmetric SP equation was considered in [54] and time-dependent SN equation was treated in [78] with three kinds of symmetry: spherical, axial and translational symmetry. Most of the pervious work apply Crank–Nicholson time integration and finite difference for space discretization. Also, note that in general the ground states of the SPS equation will lose the symmetric profile due to the external potential and therefore one cannot obtain a reduced quasi-1D model from (1.1)–(1.3) as for the SN system, by studying which the SN equation was extensively investigated in [78]. On the other hand, the computation of stationary states and dynamics of the NLS equation (1.4) without Hartree potential, has been extensively studied. Among the numerical methods proposed in the literature, discretizations based on a gradient flow with discrete normalization (GFDN) [17,18,62] show more efficient in finding the ground and excited states of NLS modeling the Bose–Einstein condensates (BEC). For dynamics, a time-splitting pseudospectral discretization [20, 21, 26] shows its accuracy and efficiency in practice. Such results suggest that we can extend these successful tools to the computation of ground states and dynamics of the SPS equation. For example, similar methods were extended in [16] to treat a Gross–Pitaevskii–Poisson type system which is used to model dipolar BEC, and a time-splitting approach was used in [23] for computing the dynamics of the SPS equation with periodic boundary conditions in all space dimensions. However, there still remains an issue that how to approximate the Hartree potential (1.5) properly, which definitely affects the overall accuracy and efficiency.

2.2 Numerical studies for ground states

In this section, the GFDN of the SPS equation is given, and different numerical methods are presented and compared for computing the ground states.

2.2.1 Ground states and normalized gradient flow

To find the stationary states of (1.1)–(1.2), we take the ansatz

$$\psi(\mathbf{x},t) = e^{-i\mu t}\phi(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d, \quad t \geq 0, \tag{2.10}$$

where $\mu \in \mathbb{R}$ is the chemical potential and $\phi := \phi(\mathbf{x})$ is a time-independent real-valued function with $\lim_{|\mathbf{x}|\to\infty} |\phi(\mathbf{x})| = 0$. Inserting (2.10) into (1.1)–(1.3) leads to the time-independent Schrödinger equation (or a nonlinear eigenvalue problem)

$$\mu\phi(\mathbf{x}) = \left[-\frac{1}{2}\Delta + V_{\text{ext}}(\mathbf{x}) + C_P V_P(|\phi|^2) - \alpha|\phi|^{\frac{2}{d}}\phi\right]\phi, \quad \mathbf{x} \in \mathbb{R}^d, \tag{2.11}$$

under the constraint

$$\|\phi\|^2 := \int_{\mathbb{R}^d} |\phi(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x} = 1, \tag{2.12}$$

where, $V_P\left(|\phi|^2\right)$ satisfies (1.5). Mathematically the ground state is defined as the minimizer of the following nonconvex minimization problem:
Find $\phi_g \in S$ and $\mu_g \in \mathbb{R}$ such that

$$E_g := E(\phi_g) = \min_{\phi\in S} \; E(\phi), \quad \mu_g := \mu(\phi_g), \tag{2.13}$$

where the constraint set S is defined as $S := \{\phi \mid \|\phi\|^2 = 1, \, E(\phi) < \infty\}$ and the chemical potential (or eigenvalue of (2.11)) is defined as

$$\begin{aligned}\mu(\phi) &:= \int_{\mathbb{R}^d} \left[\frac{1}{2}|\nabla\phi|^2 + V_{\text{ext}}(\mathbf{x})|\phi|^2 + C_P V_P(|\phi|^2)|\phi|^2 - \alpha|\phi|^{\frac{2}{d}+2}\right] \mathrm{d}\mathbf{x} \\ &\equiv E(\phi) + \int_{\mathbb{R}^d} \left[\frac{C_P}{2} V_P(|\phi|^2)|\phi|^2 - \frac{\alpha}{d+1}|\phi|^{\frac{2}{d}+2}\right] \mathrm{d}\mathbf{x}.\end{aligned} \tag{2.14}$$

In above the energy functional $E(\phi)$ is defined according to (2.9). In fact, only the positive solution of (2.13) is of interests since for any $\phi(\mathbf{x}) \in S$ we always have

$E(\phi) \geq E(|\phi|)$. Also, the nonlinear eigenvalue problem (2.11) under the constraint (2.12) can be viewed as the Euler–Lagrangian equation of the nonconvex minimization problem (2.13). Any eigenfunction of (2.11) under the constraint (2.12) corresponds to the critical point of energy functional $E(\phi)$ over the unit sphere S. The eigenfunctions whose energy are larger than E_g are usually called as excited states in physics literature.

In order to solve the minimization problem (2.13) numerically, the gradient flow with discrete normalization (GFDN) is constructed via the similar procedure as in [5,18,41] for computing the stationary states of an NLS modeling BEC. Choose a time step $\tau = \Delta t > 0$ and set $t_n = n\tau$ for $n = 0, 1, \ldots$. Applying the steepest decent method to the energy functional $E(\phi)$ in (2.9) without the constraint (2.12), and then projecting the solution back to the unit sphere S at the end of each time interval $[t_n, t_{n+1}]$ to enforce the constraint (2.12), one can obtain the following gradient flow for $\phi(\mathbf{x}, t)$ with discrete normalization:

$$\partial_t \phi(\mathbf{x}, t) = -\frac{1}{2}\frac{\delta E(\phi)}{\delta \phi} = \left[\frac{1}{2}\Delta - V_{\text{ext}}(\mathbf{x}) - C_P V_P(|\phi|^2) + \alpha|\phi|^{\frac{2}{d}}\right]\phi, \tag{2.15}$$

$$\phi(\mathbf{x}, t_{n+1}) := \phi(\mathbf{x}, t_{n+1}^+) = \frac{\phi(\mathbf{x}, t_{n+1}^-)}{\|\phi(\mathbf{x}, t_{n+1}^-)\|}, \tag{2.16}$$

$$\lim_{|\mathbf{x}|\to\infty} |\phi(\mathbf{x}, t)| = 0, \quad \phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), \quad \text{with} \quad \|\phi_0\| = 1, \tag{2.17}$$

for $\mathbf{x} \in \mathbb{R}^d$, $t_n \leq t < t_{n+1}$ and $n \geq 0$, where $\phi(\mathbf{x}, t_n^{\pm}) := \lim_{t\to t_n^{\pm}} \phi(\mathbf{x}, t)$. In fact, the gradient flow (2.15) can also be obtained from the NLS equation (1.4) by setting time t to $\tilde{t} = it$, which refers to the imaginary time method in physics literature [45, 95, 126].

Letting $\tau \to 0$ in the GFDN (2.15)–(2.17), one can obtain the following continuous normalized gradient flow (CNGF) [18]:

$$\partial_t \phi(\mathbf{x}, t) = \left[\frac{1}{2}\nabla^2 - V_{\text{ext}}(\mathbf{x}) - C_P V_P(|\phi|^2) + \alpha|\phi|^{\frac{2}{d}} + \frac{\mu(\phi)}{\|\phi\|^2}\right]\phi, \tag{2.18}$$

$$\lim_{|\mathbf{x}|\to\infty} |\phi(\mathbf{x}, t)| = 0, \quad \phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), \quad \text{with} \quad \|\phi_0\| = 1, \tag{2.19}$$

for $\mathbf{x} \in \mathbb{R}^d$ and $t \geq 0$, where $\mu(\phi)$ is defined by (2.14). It can be justified by simple calculation that the CNGF (2.18)–(2.19) is normalization conserved and energy diminishing, i.e.,

$$\|\phi(\mathbf{x},t)\|^2 \equiv \|\phi_0\|^2 = 1, \quad \frac{\mathrm{d}}{\mathrm{d}t}E(\phi(\mathbf{x},t)) = -2\|\partial_t\phi(\mathbf{x},t)\|^2 \le 0, \quad t \ge 0,$$

which also implies that $E(\phi(\mathbf{x},t_2)) \le E(\phi(\mathbf{x},t_1))$ for $0 \le t_1 \le t_2 < \infty$.

The positive ground state $\phi_g(\mathbf{x})$ and its corresponding chemical potential μ_g can be obtained from the stationary solution of GFDN (2.15)–(2.17) or CNGF (2.18)–(2.19) with a positive initial condition $\phi_0(\mathbf{x}) \ge 0$.

2.2.2 Backward Euler spectral discretization

To compute the ground states, the starting model is the GFDN (2.15)–(2.17) constructed before. In practice, the whole space problem (2.15)–(2.17) is usually truncated into a bounded computation domain Ω with homogeneous Dirichlet or periodic boundary conditions. We choose Ω as an interval $[a,b]$ in 1D, a rectangle $[a,b]\times[c,d]$ in 2D, a box $[a,b]\times[c,d]\times[e,f]$ in 3D. For simplicity of notations, the discretization in 1D shall be introduced. Generalization to higher dimensions is straightforward due to tensor product grids. When $d=1$, for $x\in[a,b]$, $t_n \le t < t_{n+1}$ and $n \ge 0$,

$$\partial_t\phi(x,t) = \frac{1}{2}\partial_{xx}\phi - V_{\text{ext}}(x)\phi - C_P V_P(|\phi|^2)\phi + \alpha|\phi|^{\frac{2}{d}}\phi, \tag{2.20}$$

$$\phi(x,t_{n+1}) := \phi(x,t_{n+1}^+) = \frac{\phi(x,t_{n+1}^-)}{\|\phi(x,t_{n+1}^-)\|_{L^2(a,b)}}, \tag{2.21}$$

$$\phi(x,0) = \phi_0(x), \quad \text{with} \quad \|\phi_0\|^2_{L^2(a,b)} := \int_a^b |\phi_0(x)|^2 \mathrm{d}x = 1, \tag{2.22}$$

with homogeneous Dirichlet boundary conditions:

$$\phi(a,t) = \phi(b,t) = 0, \quad t > 0, \tag{2.23}$$

or periodic boundary conditions:

$$\phi(a,t) = \phi(b,t), \quad \phi_x(a,t) = \phi_x(b,t), \quad t > 0. \tag{2.24}$$

Choose the spatial mesh size $h = \Delta x > 0$ with $h = \frac{(b-a)}{M}$ for M being an even positive integer, and let the grid points be $x_j = a + jh, j = 0, 1, \ldots, M$. Define two function spaces

$$Y_M^{\mathcal{S}} = \text{span}\left\{\sin\left(\mu_l(x-a)\right),\ l = 1, \ldots, M-1,\ x \in [a,b]\right\},$$
$$Y_M^{\mathcal{F}} = \text{span}\left\{\exp\left(i\lambda_l\,(x-a)\right),\ l = -M/2, \ldots, M/2-1,\ x \in [a,b]\right\},$$

with

$$\mu_l = \frac{\pi l}{b-a} \quad (l = 1, \ldots, M-1), \qquad \lambda_l = \frac{2\pi l}{b-a} \quad (l = -M/2, \ldots, M/2-1).$$

Let $\mathcal{P}_M^{\mathcal{S}} : Y_0 := \{U(x) \in C(a,b) \,|\, U(a) = U(b) = 0\} \to Y_M^{\mathcal{S}}$ and $\mathcal{P}_M^{\mathcal{F}} : Y_{\rm p} := \{U(x) \in C(a,b) \,|\, U(a) = U(b), U'(a) = U'(b)\} \to Y_M^{\mathcal{F}}$ be the standard projection operators [71, 80, 133], i.e.,

$$\left(\mathcal{P}_M^{\mathcal{S}} U\right)(x) = \sum_{l=1}^{M-1} \widehat{(U)}_l^{\mathcal{S}} \sin\left(\mu_l(x-a)\right), \quad x \in [a,b], \quad \forall\, U(x) \in Y_0,$$
$$\left(\mathcal{P}_M^{\mathcal{F}} U\right)(x) = \sum_{l=-M/2}^{M/2-1} \widehat{(U)}_l^{\mathcal{F}} \exp\left(i\lambda_l(x-a)\right), \quad x \in [a,b], \quad \forall\, U(x) \in Y_{\rm p},$$

with

$$\widehat{(U)}_l^{\mathcal{S}} = \frac{2}{b-a}\int_a^b U(x)\sin\left(\mu_l(x-a)\right) \mathrm{d}x, \quad l = 1, \ldots, M-1, \tag{2.25}$$
$$\widehat{(U)}_l^{\mathcal{F}} = \frac{1}{b-a}\int_a^b U(x)\exp\left(-i\lambda_l(x-a)\right) \mathrm{d}x, \quad l = -M/2, \ldots, M/2-1. \tag{2.26}$$

Then for (2.20)–(2.22) with homogeneous Dirichlet boundary conditions (2.23), a backward Euler sine spectral discretization reads:

Find $\phi^{n+1}(x) \in Y_M^{\mathcal{S}}$ (i.e. $\phi^+(x) \in Y_M^{\mathcal{S}}$) such that

$$\frac{\phi^+(x) - \phi^n(x)}{\tau} = \frac{1}{2}\partial_{xx}\phi^+(x) - \mathcal{P}_M^{\mathcal{S}}\left\{\left[V_{\rm ext}(x) + C_P V_P^n(x) - \alpha|\phi^n(x)|^{\frac{2}{d}}\right]\phi^+(x)\right\} \tag{2.27}$$

$$\phi^{n+1}(x) = \frac{\phi^+(x)}{\|\phi^+(x)\|_{L^2(a,b)}}, \quad \phi^0(x) = \mathcal{P}_M^{\mathcal{S}}\left(\phi_0(x)\right), \quad x \in [a,b], \quad n \geq 0. \tag{2.28}$$

Here, $V_P^n(x)$ is a numerical approximation of the Hartree potential (1.5) at time t_n with $\psi(x,t_n)$ being taken as $\phi^n(x)$, for which the numerical methods will be discussed in the coming subsection.

The above discretization can be solved in phase space but it is not suitable in practice due to the difficulty in computing the integrals in (2.25). In fact, we apply an efficient implementation by choosing $\phi^0(x)$ as the interpolation of $\phi_0(x)$ on the grid points $\{x_j,\ j=0,\ldots,M\}$ and approximating the integrals in (2.25) by a numerical quadrature rule on the grid points [57,133]. Let ϕ_j^n be the approximation of $\phi(x_j,t_n)$ and ϕ^n be a vector with components ϕ_j^n; $(V_P)_j^n$ be the approximation of the Hartree potential $V_P(x_j,t_n)$ from ϕ^n and V_P^n be a vector with components $(V_P)_j^n$. Choosing $\phi_j^0=\phi_0(x_j)$, then for $n=0,1,\ldots,$ a backward Euler sine pseudospectral discretization for (2.20)–(2.22) with homogeneous Dirichlet boundary conditions (2.23) reads,

$$\frac{\phi_j^+-\phi_j^n}{\tau}=\frac{1}{2}\,D_{xx}^s\phi^+\big|_j-\left[V_{\rm ext}(x_j)+C_P(V_P)_j^n-\alpha|\phi_j^n|^{\frac{2}{d}}\right]\phi_j^+, \tag{2.29}$$

$$\phi_0^+=\phi_M^+=0,\quad \phi_j^{n+1}=\frac{\phi_j^+}{\|\phi^+\|_h},\quad j=1,\ldots,M-1. \tag{2.30}$$

Here, D_{xx}^s is the sine pseudospectral approximation of ∂_{xx}, defined as

$$D_{xx}^sU|_j=-\sum_{l=1}^{M-1}(\mu_l)^2\widetilde{(U)}_l^s\,\sin\left(\mu_l(x_j-a)\right),\quad j=1,2,\ldots,M-1,$$

with $\widetilde{(U)}_l^s$ the discrete sine transform coefficients of the vector $U=(U_0,U_1,\ldots,U_M)^T$ satisfying $U_0=U_M=0$,

$$\widetilde{(U)}_l^s=\frac{2}{M}\sum_{j=1}^{M-1}U_j\sin\left(\mu_l(x_j-a)\right),\quad l=1,\ldots,M-1.$$

The discrete l^2-norm is defined in standard way, $\|U\|_h^2=h\sum_{j=0}^{M-1}|U_j|^2$.

The nonlinear system (2.29)–(2.30) can be iteratively solved in phase space efficiently with the help of fast sine transform (FST). The procedure is similar to that in [17] and the details are omitted here for brevity.

For the problem (2.20)–(2.22) with periodic boundary conditions (2.24), with a similar procedure to above a backward Euler Fourier spectral discretization can be proposed, i.e., replacing $Y_M^{\mathcal{S}}$ and $\mathcal{P}_M^{\mathcal{S}}$ in (2.27)–(2.28) by $Y_M^{\mathcal{F}}$ and $\mathcal{P}_M^{\mathcal{F}}$ respectively. Similarly, a practical implementation, a backward Euler Fourier pseudospectral discretization, will be used in computation which is similar to (2.29)–(2.30) but defined on a proper index set with replacing $D_{xx}^{\mathcal{S}}$ by the Fourier pseudospectral approximation of ∂_{xx}, defined as

$$D_{xx}^{\mathcal{F}}U|_j = -\sum_{l=-M/2}^{M/2-1} (\lambda_l)^2 \widetilde{(U)}_l^{\mathcal{F}} \exp\left(i\lambda_l(x_j - a)\right), \quad j = 0, 1, \ldots, M-1,$$

with $\widetilde{(U)}_l^{\mathcal{F}}$ the discrete Fourier transform coefficients of the vector $U = (U_0, U_1, \ldots, U_M)^T$ satisfying $U_0 = U_M$,

$$\widetilde{(U)}_l^{\mathcal{F}} = \frac{1}{M}\sum_{j=0}^{M-1} U_j \exp\left(-i\lambda_l(x_j - a)\right), \quad l = -M/2, \ldots, M/2-1.$$

The backward Euler Fourier pseudospectral discretization can also be iteratively solved in phase space efficiently with the help of FFT.

2.2.3 Various methods for the Hartree potential

In this subsection, different ways to obtain the approximations $(V_P)_j^n$ from the vector ϕ^n are proposed. The methods proposed here include fast convolution, sine pseudospectral and Fourier pseudospectral approaches.

Fast convolution method is the approach to approximate the convolution (1.5) on grid points with fast algorithms. Since the convolution kernel changes with the dimension of space, the algorithms also vary in different dimensions.

In 1D, first consider the problem (2.20)–(2.22) with homogeneous Dirichlet boundary conditions (2.23). For $n \geq 0$, with $\rho^n := (|\phi_0^n|^2, |\phi_1^n|^2, \ldots, |\phi_M^n|^2)^T$ the Hartree potential approximation $(V_P)_j^n$ in (2.29) is obtained by

$$(V_P)_j^n = -\frac{1}{2}\sum_{l=1}^{M-1} \widetilde{(\rho^n)}_l^{\mathcal{S}} \int_a^b |x_j - y| \sin\left(\mu_l(y-a)\right) \mathrm{d}y, \; j = 1, \ldots, M-1. \quad (2.31)$$

The integrals in above can be evaluated exactly since

$$\int_a^b |x-y| \sin\left(\mu_l(y-a)\right) \mathrm{d}y = \frac{1}{\mu_l}\left[(1+(-1)^l)x - (a+(-1)^l b)\right]$$
$$-\frac{2}{(\mu_l)^2}\sin\left(\mu_l(x-a)\right), \quad x \in [a,b], \quad l = 1, \ldots, M-1. \tag{2.32}$$

Thus,

$$(V_P)_j^n = \sum_{l=1}^{M-1} \widetilde{(\rho^n)}_l^s \frac{a+(-1)^l b}{2\mu_l} - x_j \cdot \sum_{l=1}^{M-1} \widetilde{(\rho^n)}_l^s \frac{1+(-1)^l}{2\mu_l}$$
$$+\sum_{l=1}^{M-1} \frac{\widetilde{(\rho^n)}_l^s}{(\mu_l)^2} \sin\left(\mu_l(x_j-a)\right) := S_1 - x_j \cdot S_2 + S_3, \; j = 1, \ldots, M-1. \tag{2.33}$$

Since the summation terms S_1 and S_2 are uniform for any $j = 1, \ldots, M-1$ and S_3 can be evaluated efficiently with the help of FST, the overall computation cost reduces from $O(M^2)$ for direct convolution to $O(M\ln(M))$. Hereafter the fast algorithm (2.33) is referred as 1D fast convolution method in homogeneous Dirichlet boundary conditions case. Combining this 1D fast convolution method with (2.29)–(2.30) leads to a backward Euler sine pseudospectral+fast convolution (BSFC) discretization to compute the ground states in 1D. On the other hand, for the 1D problem (2.20)–(2.22) with periodic boundary conditions (2.24), a similar fast convolution algorithm can also be achieved with the help of FFT and noting that

$$\int_a^b |x-y| \exp\left(i\lambda_l(y-a)\right) \mathrm{d}y$$
$$= \begin{cases} \dfrac{2}{(\lambda_l)^2}\left[1-\exp\left(i\lambda_l(x-a)\right)\right] + \dfrac{(a+b-2x)}{i\lambda_l}, & l \neq 0, \\[2ex] x^2-(a+b)x+\dfrac{a^2+b^2}{2}, & l = 0, \end{cases}$$

which combines with the backward Euler Fourier pseudospectral discretization (BFFC) to compute the ground states in 1D with periodic boundary conditions.

In higher dimensions, i.e. $d = 2$ and 3, the above fast algorithms is difficult to be generalized since there is no analytical formula to evaluate the convolution of $G_d(\mathbf{x})$ with sine or Fourier base functions. In what follows, the 2D and 3D convolution are accelerated by fast multipole method (FMM), for which the computation cost is $O(N)$ with N being the number of target points (grid points). Backward Euler sine/Fourier pseudospectral discretization combined with such fast convolution approximation (BSFC/BFFC) is used to compute the 2D or 3D ground states, depending on the boundary conditions made on the wave function.

For simplicity of notations, the domain Ω is assumed to be a square and a cube in 2D and 3D respectively, i.e. $\Omega^2 := [a,b] \times [a,b]$ and $\Omega^3 := [a,b] \times [a,b] \times [a,b]$, and grid points in y-axis and z-axis to be $y_k = a + kh$ and $z_l = a + lh$ for $k, l = 0, 1, \ldots, M$. Given $\phi_{jk}^n \approx \phi(x_j, y_k, t_n)$ and $\phi_{jkl}^n \approx \phi(x_j, y_k, z_l, t_n)$, the density function $\rho(\mathbf{x}, t_n) := |\phi(\mathbf{x}, t_n)|^2$ is first interpolated by a piecewise bilinear and trilinear function $\rho_h^n(\mathbf{x})$ in 2D and 3D respectively. Then, $(V_P)_{jk}^n \approx V_P(x_j, y_k, t_n)$, and $(V_P)_{jkl}^n \approx V_P(x_j, y_k, z_l, t_n)$ are obtained by evaluating

$$-\frac{1}{2\pi} \int_{\Omega^2} \ln\left(|\mathbf{x} - \mathbf{y}|\right) \rho_h^n(\mathbf{y}) d\mathbf{y}, \quad \text{and} \quad \frac{1}{4\pi} \int_{\Omega^3} \frac{1}{|\mathbf{x} - \mathbf{y}|} \rho_h^n(\mathbf{y}) d\mathbf{y}, \tag{2.34}$$

at target points.

In order to calculate the above convolution efficiently, FMM is applied by following [56] for 2D and [40,67,151] for 3D. The procedure is sketched here in a nutshell. First, an oct-tree hierarchy is imposed on Ω^3 by dividing the cube into eight sub cubes recursively. Similarly, a quad-tree is superimposed on Ω^2 in 2D. One can refer to [56,66,67] for detailed tree structures and their adaptivity. In FMM, the far field interactions are calculated by means of multipole expansions (via upward pass) and it converts the multipole expansions into local expansions (via downward pass) relying on three kind of translation operators acting on multipole and local expansions in the tree hierarchy: multipole-to-multipole ($\mathcal{T}_{MM}$), multipole-to-local ($\mathcal{T}_{ML}$), and local-to-local ($\mathcal{T}_{LL}$) translations. Last, direct interactions (influence from neighbors of a leaf node and itself) are computed according to (2.34). The algorithms are omitted here for brevity and one can refer to [40, 56, 67] for the technical details. The most time-consuming translation operator $\mathcal{T}_{ML}$ is accelerated by plane wave method as described in [40, 85] for 2D and 3D. To calculate the integrals in the multipole and local expansions efficiently, recurrence formulas for the spherical harmonics are helpful (refer to [150]). For regular integral in (2.34), Gaussian quadrature is applied. In the implementation, both multipole and local expansions are truncated to $p = 18$ terms which allows a 6-digits precision.

Sine pseudospectral approximation is the approach to solve the Poisson equation (1.2) (or its modified equation) on the bounded domain Ω with homogeneous Dirichlet boundary conditions by using sine pseudospectral method. In 1D, consider the problem (2.20)–(2.22) with homogeneous Dirichlet boundary conditions (2.23), and at each time t_n, given ϕ^n, consider the following problem

$$\partial_{xx} V_P(x_j, t_n) = -|\phi_j^n|^2, \quad j = 1, \dots, M-1, \quad n \geq 0 \tag{2.35}$$

$$V_P(x_0, t_n) = (V_P)_0^n, \quad V_P(x_M, t_n) = (V_P)_M^n, \tag{2.36}$$

where, $(V_P)_0^n$ and $(V_P)_M^n$ are two approximated boundary conditions which, for example, can be obtained from (2.33) by letting $j = 0$ and M respectively. Then a sine pseudospectral discretization to a modified problem of (2.35)–(2.36) reads

$$D_{xx}^{\mathcal{S}} \underline{V_P}(x, t_n)\big|_j = -|\phi_j^n|^2, \quad j = 1, \dots, M-1, \quad n \geq 0 \tag{2.37}$$

$$\underline{V_P}(x_0, t_n) = \underline{V_P}(x_M, t_n) = 0, \tag{2.38}$$

where

$$\underline{V_P}(x, t) = V_P(x, t) - \frac{(V_P)_M^n - (V_P)_0^n}{b - a}(x - a) - (V_P)_0^n. \tag{2.39}$$

Solving (2.37)–(2.38) in phase space, for $j = 1, \dots, M-1$,

$$(V_P)_j^n = \sum_{l=1}^{M-1} \frac{\widetilde{(\rho^n)}_l^{\mathcal{S}}}{(\mu_l)^2} \sin\left(\mu_l(x_j - a)\right) + \frac{(V_P)_M^n - (V_P)_0^n}{b - a}(x_j - a) + (V_P)_0^n. \tag{2.40}$$

Note that if the external potential $V_{\text{ext}}(x)$ is symmetric, without loss of generality $V_{\text{ext}}(x)$ is an even function, then the solution of (2.15)–(2.17) $\phi(x, t)$ should also be even. Therefore, it is reasonable to choose $a = -b$ and the approximated boundary conditions $(V_P)_0^n = (V_P)_M^n$ in (2.37)–(2.38) due to (1.5). Then, the approximation in (2.40) is just a constant translation of the result by applying the sine pseudospectral discretization to (2.35) with $(V_P)_0^n = (V_P)_M^n = 0$. In view of any constant translation of external potential will leave the ground states unchanged, one can simply choose $(V_P)_0^n = (V_P)_M^n = 0$ when $V_{\text{ext}}(x)$ is an even function, i.e. the Hartree potential is approximated by

$$(V_P)_j^n = \sum_{l=1}^{M-1} \frac{\widetilde{(\rho^n)}_l^{\mathcal{S}}}{(\mu_l)^2} \sin\left(\mu_l(x_j - a)\right), \quad j = 1, \ldots, M-1. \tag{2.41}$$

In 3D, the far-field condition of $V_P(\mathbf{x}, t)$ being $\lim_{|\mathbf{x}|\to\infty} |V_P(\mathbf{x}, t)| = 0$ can be drawn from (1.5), and therefore the sine pseudospectral discretization in 3D is a straightforward generalization of (2.41) by tensor product grids without any modification provided that the bounded domain Ω is chosen large enough.

Hereafter in this chapter, (2.40) or (2.41) for an even external potential, and the generalization of (2.41) in 3D are referred as sine pseudospectral approximation of (1.5). Combining this method with (2.29)–(2.30), one can obtain a backward Euler sine pseudospectral (BESP) discretization to compute the ground states in $d = 1$, 3 space dimensions.

Remark 2.1. *In 2D, to obtain appropriate approximated boundary conditions with high-order of accuracy is a costly job itself. Meanwhile, no homogenization tool like (2.37)–(2.39) is available in general for 2D problems. Thus, the homogeneous boundary conditions cannot be satisfied, and the sine pseudospectral approach is not applicable in 2D. The work to propose a spectral-type approach in 2D with artificial boundary conditions is still on-going.*

Fourier pseudospectral approximation is the approach to solve the Poisson equation (1.2) (or its modified equation) on the bounded domain Ω with periodic boundary conditions by using Fourier pseudospectral method. In 1D, consider the problem (2.20)–(2.22) with periodic boundary conditions (2.24). At each time t_n, for (2.35)–(2.36) and introducing

$$\underline{V_P}(x, t) = V_P(x, t) - \frac{(V_P)_M^n - (V_P)_0^n}{b - a}(x - a), \tag{2.42}$$

one obtains a modified problem

$$\partial_{xx}\underline{V_P}(x_j,t_n) = -|\phi_j^n|^2, \quad j=0,\ldots,M-1, \quad n\geq 0 \tag{2.43}$$

$$\underline{V_P}(x_0,t_n) = \underline{V_P}(x_M,t_n), \tag{2.44}$$

which determines a unique $\underline{V_P}$ up to a constant translation. A Fourier pseudospectral discretization applying to the modified problem (2.43)–(2.44) reads

$$D_{xx}^{\mathcal{F}}\underline{V_P}(x,t_n)\big|_j = -|\phi_j^n|^2 + \frac{1}{b-a}, \quad j=0,\ldots,M-1, \quad n\geq 0 \tag{2.45}$$

$$\underline{V_P}(x_0,t_n) = \underline{V_P}(x_M,t_n). \tag{2.46}$$

Adding the last term in (2.45) is due to the consistency requirement in 0-mode after taking Fourier transform on both sides of (2.43) and the normalization condition of ϕ^n. Then $(V_P^n)_j$ for $j=0,1,\ldots,M-1$ is obtained by

$$(V_P)_j^n = \sum_{l=-M/2}^{M/2-1} \widetilde{(\underline{V_P})}_l^{\mathcal{F}} \exp\left(i\lambda_l(x_j-a)\right) + \frac{(V_P)_M^n - (V_P)_0^n}{b-a}(x_j-a), \tag{2.47}$$

where $\widetilde{(\underline{V_P})}_l^{\mathcal{F}} = \dfrac{\widetilde{(\rho^n)}_l^{\mathcal{F}}}{(\lambda_l)^2}$ for $l\neq 0$, and usually one can choose $\widetilde{(\underline{V_P})}_0^{\mathcal{F}} = 0$. In fact, $\widetilde{(\underline{V_P})}_0^{\mathcal{F}}$ can be chosen as any value since any constant translation of potential leaves the ground states unchanged.

One remark here is that although the above approximation is expected to have a spectral order of accuracy, the error from adding $(b-a)^{-1}$ in (2.45) to ensure the consistency in 0-mode will dominate (truncation error). It implies that the approximation will converge when $b-a$ becomes larger, as shown in the above method derivation and the numerical results reported in the next subsection. Therefore, in practice if periodic boundary conditions are made and the Fourier approach is applied, a large computation domain is necessary. However, it is noted that the Fourier approach for solving (2.45)–(2.46) is spectrally accurate as expected and numerically shown in the next subsection, hence with only a few grid points it can already achieve the conserved approximation with respect to the computation domain. On the other hand, it can be implemented very efficiently thanks to FFT.

Thus, to obtain a good approximation one can implement it on a large computation domain but with relatively few grid points, and the computation cost would be much less than other discretization methods, like finite difference or finite element approaches.

Similar to the sine pseudospectral discretization, in 1D if the external potential $V_{\text{ext}}(x)$ is an even function, then $(V_P^n)_j$ can be simply evaluated by

$$(V_P)_j^n = \sum_{l=-M/2}^{M/2-1} \widetilde{\left(\underline{V_P}\right)}_l^{\mathcal{F}} \exp\left(i\lambda_l(x_j - a)\right), \tag{2.48}$$

provided $a = -b$. Again, the Fourier pseudospectral discretization in 3D is a straightforward generalization of (2.48) with tensor product grids, while such discretization is not suitable to 2D case for similar reasons pointed out in Remark 2.1. With such approximation it leads to a backward Euler Fourier pseudospectral (BEFP) discretization for computing the ground states in $d = 1$, 3 space dimensions.

2.2.4 Numerical results

Numerical comparisons among all the discussed numerical methods for computing the ground states and application results of the BESP method to investigate the ground states of the SPS equation in 3D under various setups are shown here.

Comparison of different methods for ground states

In order to reflect the effects of different Hartree potential approximations on the computed ground states, we only present the results for the simplest form of (1.1)–(1.3), i.e., the SN equation.

Example 2.1. Ground states of 1D SN equation without external potential, i.e., $d = 1$, $V_{\text{ext}} = 0$, $C_P < 0$ and $\alpha = 0$ in (1.1), are examined with $C_P = -3$. In computation, the initial guess is chosen as $\phi_0 = \pi^{-1/4}\mathrm{e}^{-x^2/2}$, $x \in \mathbb{R}$, and the time step is $\tau = 0.005$. Let ϕ_g be the "exact" ground state obtained from BSFC with a very fine mesh size $h = 1/128$ on $\Omega = [-128, 128]$. $\phi_{g,h}$ denotes the approximated

Table 2.1: Ground state error analysis in Example 2.1. (1) $\|\phi_g - \phi_{g,h}\|_\infty$ versus mesh size h on $\Omega = [-16, 16]$ for BSFC, BESP and BEFP (upper part); (2) $\|\phi_g - \phi_{g,h}\|_\infty$ versus bounded domain $\Omega = [-a, a]$ with $h = 1/16$ for BEFP (last row).

mesh size	$h = 1$	$h = 1/2$	$h = 1/4$	$h = 1/8$	$h = 1/16$
BSFC	7.644E-03	4.076E-06	1.400E-12	<E-12	<E-12
BESP	7.644E-03	4.076E-06	1.400E-12	<E-12	<E-12
BEFP	5.725E-03	1.074E-02	1.074E-02	1.074E-02	1.074E-02
domain	$a = 8$	$a = 16$	$a = 32$	$a = 64$	$a = 128$
BEFP	2.297E-02	1.078E-02	5.235E-03	2.581E-03	1.281E-03

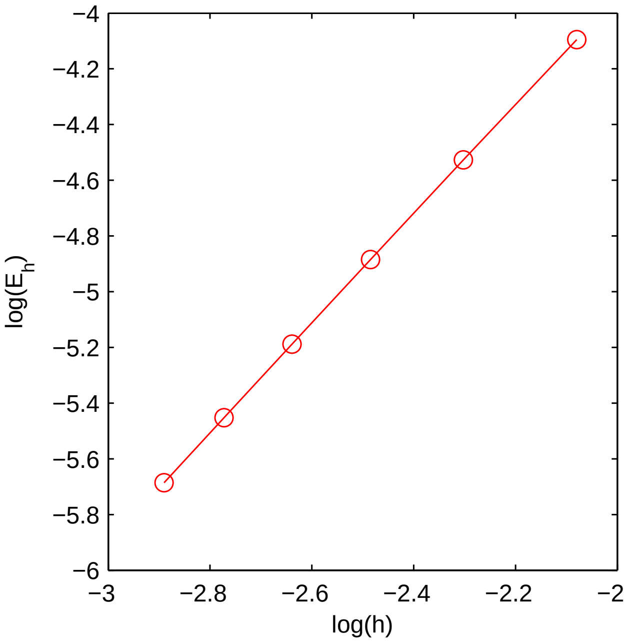


Figure 2.1: Ground state error analysis in Example 2.2. Plot of $\log(\|\phi_g - \phi_{g,h}\|_\infty)$ versus $\log(h)$ for 3D BSFC method on a cube $[-4, 4]^3$ with uniform grids in each axis.

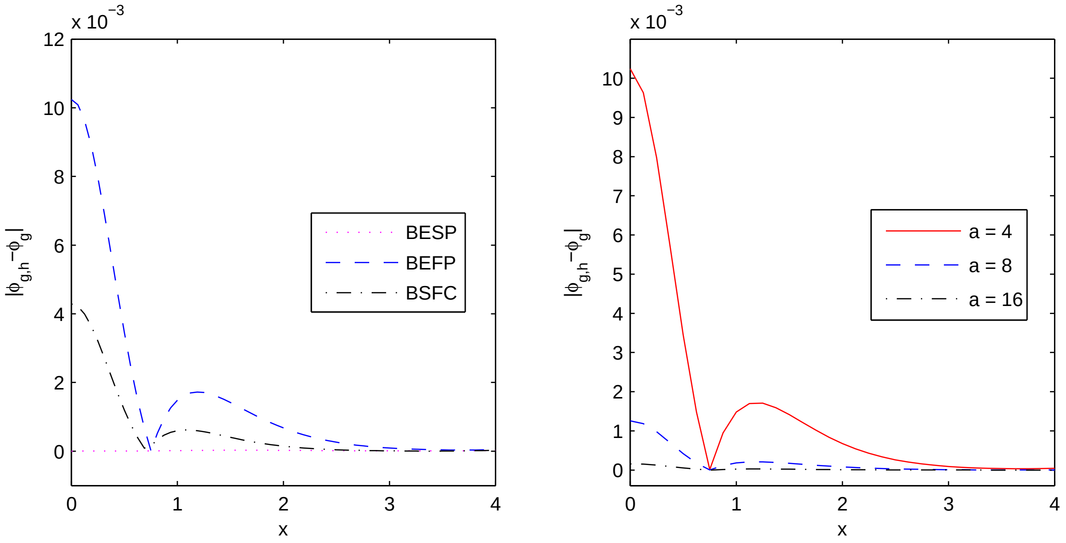


Figure 2.2: Ground state error analysis in Example 2.2. Left: slice plots of $|\phi_g - \phi_{g,h}|$ along x-axis for 3D BSFC, BESP and BEFP in a cube $[-4,4]^3$ with uniform mesh size $h = 1/16$ in each axis; right: slice plots of $|\phi_g - \phi_{g,h}|$ along x-axis for BEFP in different cubes $[-a,a]^3$ with uniform mesh size $h = 1/8$ in each axis.

ground states obtained from different methods with mesh size h. Tab. 2.1 shows the errors $\|\phi_g - \phi_{g,h}\|_\infty$ of the methods BSFC, BESP and BEFP on $\Omega = [-16,16]$ for various mesh sizes h and $\|\phi_g - \phi_{g,h}\|_\infty$ of BEFP on different domains Ω with $h = 1/16$.

Example 2.2. Ground states of 3D SN equation without external potential, i.e.,$d = 3$, $V_{\text{ext}} = 0$, $C_P < 0$ and $\alpha = 0$ in (1.1), are examined with $C_P = -75$. In computation, the initial guess is chosen as $\phi_0 = (6\pi)^{-3/4}\, \mathrm{e}^{-(x^2+y^2+z^2)/12}$, $(x,y,z) \in \mathbb{R}^3$, and time step is $\tau = 0.01$. Since the ground state of this SN equation is radially symmetric, a benchmark is achieved by using a Backward Euler finite difference method to the reduced quasi-1D model of GFDN (2.15)–(2.17) with Dirichlet boundary conditions of ϕ and Robin boundary conditions of V_P. The "exact" solution $\phi_g(r)$ is computed in a ball $0 \le r \le 8$ with a very fine mesh size $\Delta r = 1/1024$. Fig. 2.1 shows the convergence rate of BSFC method in 3D, which applies FMM to accelerate the direct convolution (1.5). Fig. 2.2 depicts the slice plots of error $|\phi_g - \phi_{g,h}|$

along x-axis for 3D BSFC, BESP and BEFP methods in a cube $[-4, 4]^3$ with uniform mesh size $h = 1/16$ in each axis, and for BEFP in different cubes with uniform mesh size $h = 1/8$ in each axis.

From Tab. 2.1, Fig. 2.1, 2.2 and additional results not shown here for brevity, the following observations are made:

(i). BESP and 1D BSFC methods both have spectral order of accuracy (cf. Tab. 2.1), and 2D and 3D BSFC methods have second-order of accuracy in spatial discretization (cf. Fig. 2.1).

(ii). For BEFP method, the error from the truncated computation domain dominates and it has a low order of accuracy instead of spectral order of accuracy expected for spectral-type method. This is observed in the error $\|\phi_g - \phi_{g,h}\|_\infty$ versus h of BEFP for a fixed domain (cf. the 3rd row in Tab. 2.1), which remains to be a uniform bound when h goes finer. In addition, as indicated by the method formulation, the approximated ground states will converge as the truncated domain is chosen larger (cf. the last row in Tab. 2.1 and (b) in Fig. 2.2).

(iii). In 3D, BEFP method is a better choice than BSFC method which applies FMM to accelerate the convolution. Comparing (a) and (b) in Fig. 2.2 (cf. "$-\cdot-\cdot-$" in (a) versus "$----$" in (b)), it shows that with the same number of grid points, BEFP method gives better approximations than BSFC method. In addition, the implementation of BEFP is much more efficient due to FFT.

(iv). In view of both efficiency and accuracy, BESP method is the best choice for computing the ground states of the SPS equation in 3D.

Applications of BESP method

Example 2.3. Ground states of 3D SPS are investigated in different setups:

Table 2.2: Results in Example 2.3. Different quantities in the ground states of the SPS equation for Poisson coefficient $C_P = 1$ with different exchange coefficients α under $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + 4z^2)$.

α	E_g^{kin}	E_g^{pot}	E_g^{int}	E_g^{exc}	E_g	μ_g	δ_x	δ_z	$\rho_g(0)$
0.1	0.999	1.001	0.031	-0.031	2.000	2.021	0.501	0.250	0.503
0.5	1.031	0.970	0.032	-0.157	1.876	1.855	0.481	0.245	0.519
1	1.074	0.932	0.032	-0.321	1.717	1.642	0.455	0.238	0.540
5	1.619	0.635	0.038	-2.013	0.279	-0.355	0.272	0.182	0.786
10	3.154	0.348	0.057	-5.677	-2.118	-3.953	0.128	0.110	1.357

Table 2.3: Results in Example 2.3. Different quantities in the ground states of the SPS equation without exchange term for different Poisson coefficients C_P under $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + z^2)$.

C_P	E_g^{kin}	E_g^{pot}	E_g^{int}	E_g^{exc}	E_g	μ_g	$\rho_g(0)$
-50	1.516	0.377	-1.845	0.000	0.048	-1.797	0.780
-10	0.839	0.671	-0.293	0.000	1.217	0.923	0.471
-5	0.792	0.710	-0.142	0.000	1.361	1.219	0.446
0	0.750	0.750	0.000	0.000	1.500	1.500	0.424
5	0.713	0.790	0.111	0.000	1.613	1.724	0.404
10	0.679	0.829	0.258	0.000	1.766	2.023	0.385
50	0.502	1.137	1.054	0.000	2.694	3.748	0.280

CASE I: fixed Poisson potential coefficient, e.g. $C_P = 1$, with different exchange coefficients α. Here, the equation under a trapping potential $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + 4z^2)$ is considered. Tab. 2.2 lists different quantities in the ground states for this case, which shows that for a fixed Poisson constant C_P, when α increases, energy E_g,

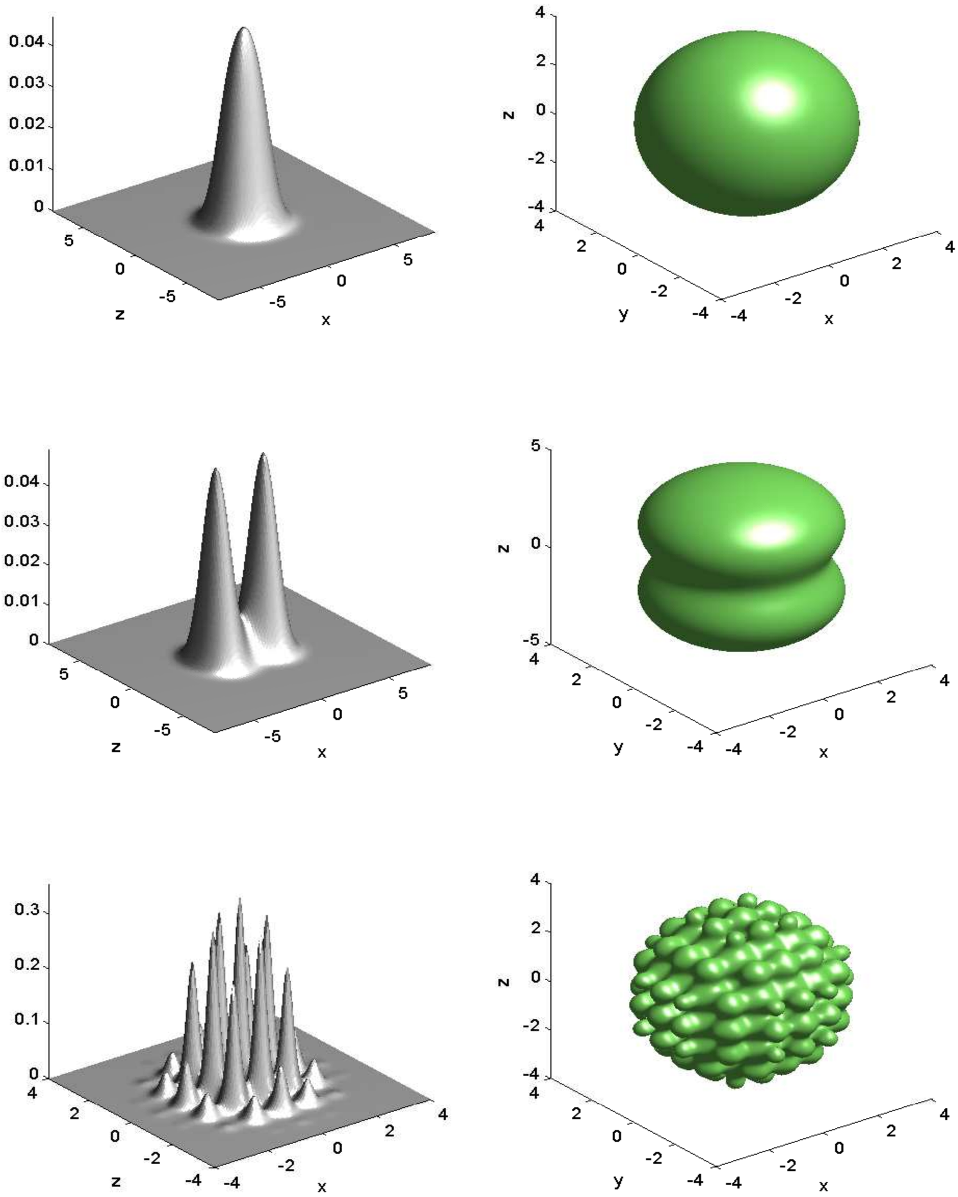


Figure 2.3: Results in Example 2.3. Surface plots of ground states $|\phi_g(x,0,z)|^2$ (left column) and isosurface plots of $|\phi_g(x,y,z)| = 0.01$ (right column) of the SPS equation (1.1) with $C_P = 100$ and $\alpha = 1$ under harmonic potential (top row), double-well potential (middle row) and optical lattice potential (bottom row).

chemical energy μ_g, potential energy E_g^{pot}, exchange energy E_g^{exc}, mean widths δ_x and δ_z decrease, whereas kinetic energy E_g^{kin}, interaction energy E_g^{int} and central density $\rho_g(0)$ increase. Here, the studied kinetic, potential, interaction and exchange energies account for the four consecutive terms appearing in the continuous energy functional (2.9), which can be evaluated by using the Parsaval's identity. Also, the mean width is defined by

$$\delta_x = \int_{\mathbb{R}^3} x^2 |\phi_g(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x},$$

and similar for δ_z, which can be computed numerically.

CASE II: different Poisson potential coefficients C_P without exchange term under a trapping potential $V_{\text{ext}} = \frac{1}{2}\left(x^2 + y^2 + 4z^2\right)$. Different quantities in the ground states for this case are listed in Tab. 2.3, which shows that without exchange effect, i.e. $\alpha = 0$, when the Poisson constant C_P increases from negative (attractive) to positive (repulsive), energy E_g, chemical energy μ_g, potential energy E_g^{pot} and interaction energy E_g^{int} increase, and the kinetic energy E_g^{kin}, whereas central density $\rho_g(0)$ decrease.

CASE III: ground states under various external potentials. Fig. 2.3 depicts the surface plots of ground states $|\phi_g(x, 0, z)|^2$ and isosurface plots of $|\phi_g| = 0.01$ of the SPS equation (1.1) with $C_P = 100$, $\alpha = 1$ and under: (1) harmonic potential $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + z^2)$; (ii) double-well potential $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + z^2) + 4e^{-\frac{1}{2}z^2}$; (iii) optical lattice potential $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + z^2) + 20\left[\sin(\pi x)^2 + \sin(\pi y)^2 + \sin(\pi z)^2\right].$

2.3 Numerical studies for dynamics

In this section, an efficient and accurate time-splitting sine or Fourier pseudospectral discretization is presented, coupled with the various approaches proposed in Section 2.2.3 for approximating the Hartree potential, to compute the dynamics of the SPS equation (1.1)–(1.3) or (1.4) with (1.3).

2.3.1 Efficient methods

Again, in practice the whole space problem is truncated into a bounded computation domain Ω with either homogeneous Dirichlet or periodic boundary conditions. For simplicity of notations, the discretization in 1D with homogeneous Dirichlet boundary conditions shall be introduced. Generalizations to higher space dimensions or periodic boundary conditions proceed in the same manner as in the last section. In 1D, from $t = t_n$ to $t = t_{n+1}$, the problem (1.4) on $\Omega = [a, b]$ with homogeneous Dirichlet boundary conditions splits into two steps, i.e., the so-called time-splitting technique which is widely and successfully used for evolution equations [20, 21, 23, 26, 29, 86, 105, 108, 139, 144]. One solves first the free Schrödinger equation

$$i\partial_t \psi(x,t) = -\frac{1}{2}\partial_{xx}\psi(x,t), \quad \psi(a,t) = \psi(b,t) = 0, \quad t_n \le t \le t_{n+1}, \tag{2.49}$$

for the time step with length τ, followed by solving

$$i\partial_t \psi(x,t) = \left[V_{\text{ext}}(x) + C_P V_P \left(|\psi|^2\right) - \alpha|\psi|^2\right]\psi, \quad t_n \le t \le t_{n+1}, \tag{2.50}$$

for the same time step. Similar to [26] the problem (2.49) is discretized in space by sine pseudospectral method and integrated in phase space exactly. For $t_n \le t \le t_{n+1}$, (2.50) leaves $|\psi|$ (so as V_P) unchanged in t and thus it collapses to

$$i\partial_t \psi(x,t) = \left[V_{\text{ext}}(x) + C_P V_P \left(|\psi(x,t_n)|^2\right) - \alpha|\psi(x,t_n)|^2\right]\psi, \tag{2.51}$$

The linear ODE (2.51) is integrated in time exactly with the Hartree potential V_P being approximated by methods proposed in Section 2.2.3. Let ψ_j^n be the approximation of $\psi(x_j, t_n)$ and ψ^n be the approximation vector with components ψ_j^n; $(V_P)_j^n$ be the approximation of the Hartree potential $V_P(x_j, t_n)$ from ψ^n and V_P^n be a vector with components $(V_P)_j^n$; and choose $\psi_j^0 = \psi_0(x_j)$ for $j = 0, \ldots, M$. For $n = 0, 1, \ldots,$ a detailed second-order time-splitting sine pseudospectral discretization via Strang formula [20,21,23,26,29,86,105,108,139,144], applied in computing, is as follows

$$\psi_j^{(1)} = \sum_{l=1}^{M-1} \exp\left(-i\tau\mu_l^2/4\right) \widetilde{(\psi^n)}_l^s \sin\left(\mu_l(x_j-a)\right), \tag{2.52}$$

$$\psi_j^{(2)} = \exp\left[-i\tau\left(V_{\rm ext}(x_j) + C_P(V_P^{(1)})_j - \alpha|\psi_j^{(1)}|^2\right)\right]\psi_j^{(1)}, \tag{2.53}$$

$$\psi_j^{n+1} = \sum_{l=1}^{M-1} \exp\left(-i\tau\mu_l^2/4\right) \widetilde{(\psi^{(2)})}_l^s \sin\left(\mu_l(x_j-a)\right), \tag{2.54}$$

for $j = 1, \dots, M-1$. Here, $\widetilde{(\psi^n)}_l^s$ is the discrete sine transform coefficients of ψ^n. Evaluating $(V_P^{(1)})_j$ via sine pseudospectral method (2.40) leads to a time-splitting sine pseudospectral (TSSP) discretization to compute the dynamics. Similarly, a time-splitting sine pseudospectral+fast convolution (TSFC) method is obtained by evaluating $(V_P^{(1)})_j$ via the fast convolution approach (2.33). These methods are explicit, unconditionally stable and time reversible. Moreover, for the L^2 stability, one has

Lemma 2.1. *TSFC and TSSP are normalization conserved, i.e.,*

$$\|\psi^n\|_h^2 := h\sum_{j=0}^{M-1} \left|\psi_j^n\right|^2 \equiv h\sum_{j=0}^{M-1} \left|\psi_j^0\right|^2 = \left\|\psi^0\right\|_h^2, \qquad n \geq 0. \tag{2.55}$$

Proof. The argument process in analogous lines as in [20,21] and the details are omitted here for brevity. □

In periodic boundary conditions case, a similar time-splitting Fourier pseudospectral discretization can be proposed. It combines with Fourier pseudospectral method (2.47), i.e. TSFP, or with fast convolution approach based on Fourier bases, i.e. TFFC, to compute the dynamics. The details are omitted here for brevity. Also, they are explicit, time reversible, time traversable and unconditionally stable.

Note that in the special case that the external potential is even, then $(V_P^{(1)})_j$ can be again simply obtained by (2.41) or (2.48). This is because from t_n to t_{n+1}, if a constant c is added to potential $(V_P^{(1)})_j$, then ψ_j^{n+1} obtained from time-splitting sine or Fourier pseudospectral approaches get multiplied by a phase factor $\exp(-i\tau C_P \cdot c)$, which leaves $|\psi_j^{n+1}|$ unchanged and so as for any discrete quadratic observables, for

Table 2.4: Density error analysis in Example 2.4. (1) $\|\rho-\rho_h\|_\infty$ at $t=1.0$ versus mesh size h on $\Omega=[-16,16]$ for TSFC, TSSP and TSFP (upper part); (2) $\|\rho-\rho_h\|_\infty$ at $t=1.0$ versus bounded domain $\Omega=[-a,a]$ with $h=1/32$ for BEFP (last row).

mesh size	$h=1$	$h=1/2$	$h=1/4$	$h=1/8$	$h=1/16$
TSFC	5.017E-02	1.531E-02	1.120E-05	1.412E-12	<E-12
TSSP	5.017E-02	1.531E-02	1.120E-05	1.396E-12	<E-12
TSFP	5.412E-02	3.968E-02	2.345E-02	2.345E-02	2.345E-02
domain	$a=8$	$a=16$	$a=32$	$a=64$	$a=128$
TSFP	6.207E-02	2.345E-02	1.107E-02	5.395E-03	2.654E-03

example the particle density $\rho_j^n=|\psi_j^n|^2$, and such observables are of real interests in applications.

2.3.2 Numerical results

Comparisons among different methods and application results of the TSSP method to study the dynamics of the SPS equation under various setups are reported here.

Comparison of different methods for dynamics

Again, comparisons are carried out for the SN equation in 1D and 3D.

Example 2.4. Dynamics of 1D SN equation without external potential, i.e., $d=1$, $V_{\text{ext}}=0$, $C_P<0$ and $\alpha=0$ in (1.1), is studied with $C_P=-20$. The initial value is taken as $\psi_0(\mathbf{x})=\pi^{-1/4}\mathrm{e}^{-x^2/2}$, $x\in\mathbb{R}$. Here, special focus is put on the spatial resolution capacity of different methods, and hence a very small time step $\tau=0.0001$ is chosen such that the error from time discretization is negligible. The "exact" solution of wave function ψ and density $\rho=|\psi|^2$ are computed from TSFC on $\Omega=[-64,64]$ with a very fine mesh size $h=1/32$. $\rho_h=|\psi_h|^2$ denotes

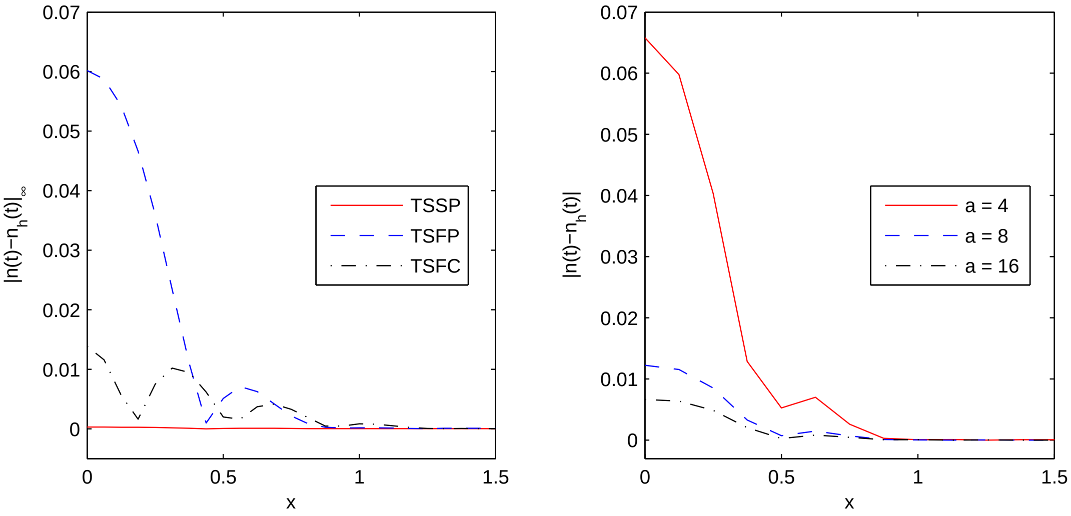
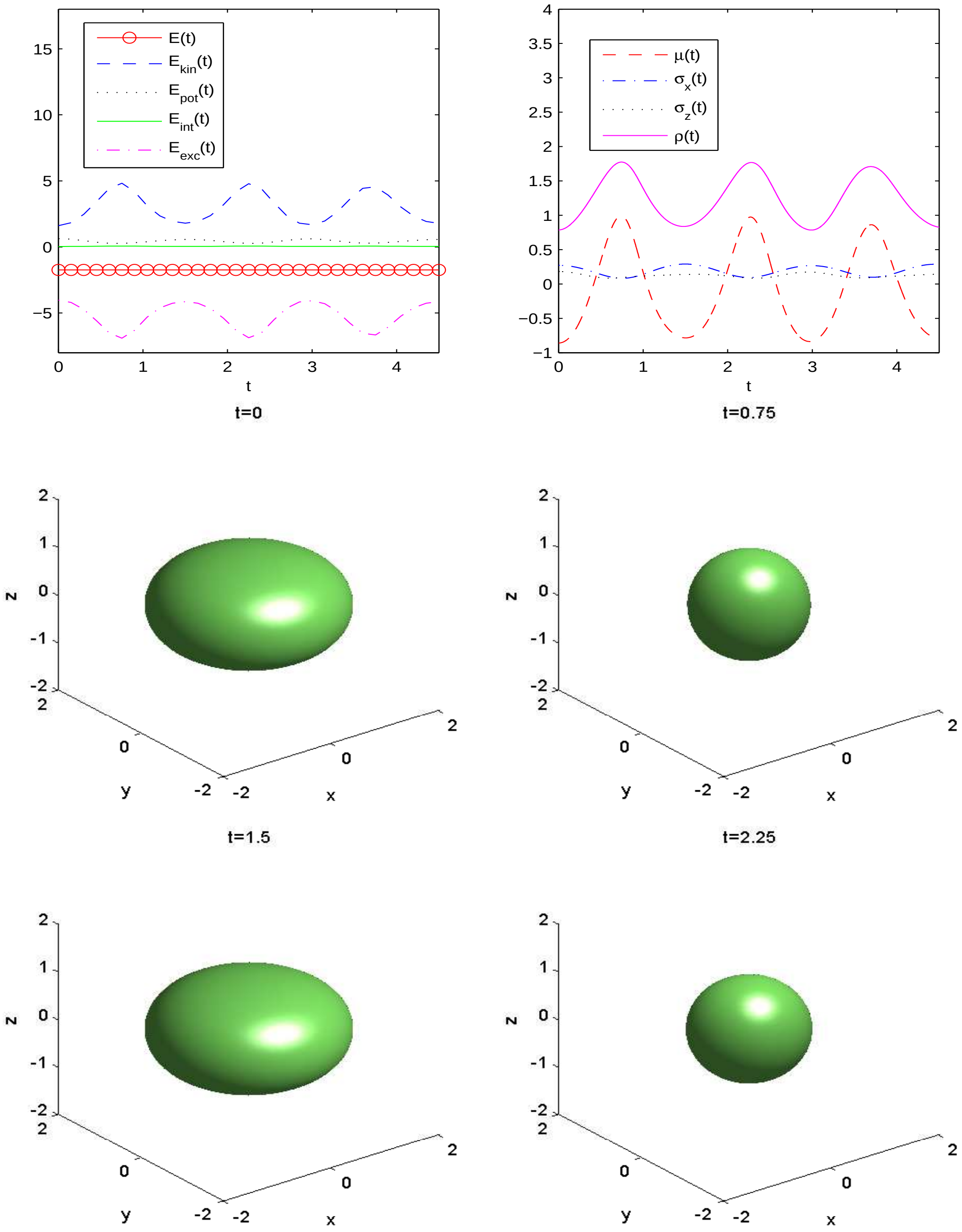


Figure 2.4: Density error analysis in Example 2.5. Left: slice plots of $|\rho - \rho_h|$ along x-axis for 3D TSFC, TSSP and TSFP in a cube $[-4,4]^3$ with uniform mesh size $h = 1/16$ in each axis; right: slice plots of $|\rho - \rho_h|$ along x-axis for TSFP in different cubes $[-a,a]^3$ with uniform mesh size $h = 1/8$ in each axis.

the approximated density with mesh size h. Tab. 2.4 shows the density errors $\|\rho(t) - \rho_h(t)\|_\infty$ at $t = 1.0$ of the methods TSFC, TSSP and TSFP with different mesh sizes h on $\Omega = [-16,16]$, and the similar error of TSFP on different domains Ω with $h = 1/32$.

Example 2.5. Dynamics of 3D SN equation without external potential, i.e., $d = 3$, $V_{\text{ext}} = 0$, $C_P < 0$ and $\alpha = 0$ in (1.1) is studied with $C_P = -200$. A radially symmetric initial value is chosen as $\psi_0 = (\pi/2)^{-3/4}\mathrm{e}^{-(x^2+y^2+z^2)}$. A benchmark is obtained by applying a Crank–Nicolson finite difference method to the reduced 1D model due to the radial symmetry property. The "exact" solution $\psi(r,t)$ is computed in a ball $0 \le r \le 8$ with a very fine mesh size $\Delta r = 1/1024$ and a very fine time step $\tau = 0.00001$. TSFC, TSSP and TSFP methods are compared with the same time step $\tau = 0.001$. Slice plots of $|\rho - \rho_h|$ along x-axis of these methods in a cube $[-4,4]^3$ with uniform mesh size $h = 1/16$ in each axis, and for TSFP in different cubes with uniform mesh size $h = 1/8$ in each axis are shown in Fig. 2.4.



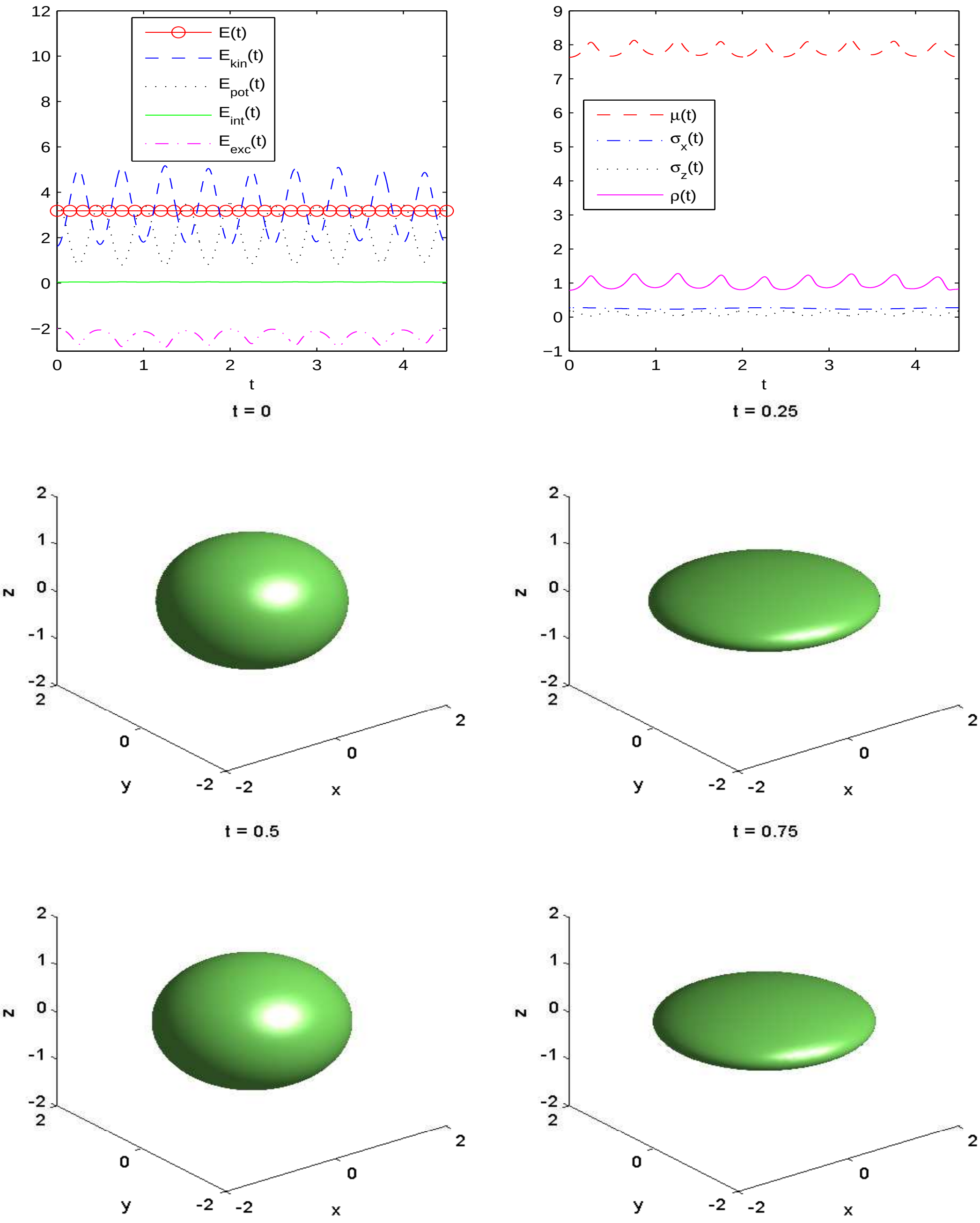
t=0

t=0.75

t=1.5

t=2.25

Figure 2.5: Results in Example 2.6. Time evolution of various quantities and iso-surface plots of density $\rho(\mathbf{x}, t) := |\psi(\mathbf{x}, t)|^2 = 0.01$ at different time points for 3D SPS with slater coefficient changing from $\alpha = 5$ to $\alpha = 10$ at $t = 0$.



t = 0

t = 0.25

t = 0.5

t = 0.75

Figure 2.6: Results in Example 2.6. Time evolution of various quantities and isosurface plots of density $\rho(\mathbf{x},t) := |\psi(\mathbf{x},t)|^2 = 0.01$ at different time points for 3D SPS under external potential changing from $V_{\text{ext}} = \frac{1}{2}(x^2+y^2+4z^2)$ to $V_{\text{ext}} = \frac{1}{2}(x^2+y^2+36z^2)$ at $t = 0$.

From Tab. 2.4, Fig. 2.4 and additional results not shown here for brevity, similar conclusions to those made after Example 1.1 and 1.2 can be drawn about the convergence in spatial discretization for TSFC, TSSP and TSFP methods. Also, TSSP method is the best choice for computing the dynamics of the SPS equation in 3D.

Applications of TSSP method

Example 2.6. Dynamics of 3D SPS equation is investigated in different setups. In the run, the initial data $\psi_0(\mathbf{x})$ is chosen as the ground state computed numerically for $C_P = 1, \alpha = 5, V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + 4z^2)$. First, the slater coefficient is instantly changed from $\alpha = 5$ to $\alpha = 10$ while all the other parameters are kept unchanged. Fig. 2.5 depicts the time evolution of total energy $E(t)$, kinetic energy $E^{\text{kin}}(t)$, potential energy $E^{\text{pot}}(t)$, interaction energy $E^{\text{int}}(t)$, exchange energy $E^{\text{exc}}(t)$, chemical potential $\mu(t)$, condensate width $\sigma_x(t)$, $\sigma_z(t)$, central density $\rho_0(t) := |\psi(0,0,0,t)|^2$ and the isosurface plots of density $\rho(\mathbf{x}, t) := |\psi(\mathbf{x}, t)|^2 = 0.01$ at different time points. Next, Fig. 2.6 shows the similar quantities for the case of instantly changing the trapping potential from $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + 4z^2)$ to $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + 36z^2)$ and keeping all the other parameters unchanged.

In Fig. 2.5 and 2.6, a periodic profile of kinetic energy, potential energy, interaction energy, exchange energy, chemical potential, condensate width and density is observed. In addition, the total energy is numerically conserved very well by the TSSP method.

2.4 Simplified spectral-type methods for spherically symmetric case

In the last two sections, various approaches were proposed and compared for computing the ground states and dynamics of the SPS equation (1.1)–(1.2) for general external potential and initial condition, with a conclusion that the methods based on sine pseudospectral discretization in space are the best candidates. This section is concerned with the case that the external potential and initial condition are spherically symmetric. For the SPS equation with spherical symmetry, via applying a proper change of variables into the reduced quasi-1D model the methods BESP and TSSP for the general 3D case are simplified. The simplified methods are still spectrally accurate in space, but reduce the memory cost from $O(J^3)$ to $O(J)$ and the computational cost per time step from $O(J^3 \ln(J^3))$ to $O(J \ln(J))$, where J is the number of mesh nodes.

2.4.1 A quasi-1D model in spherically symmetric case

Throughout this section, both the external potential V_{ext} and initial condition ψ_0 in (1.1) are assumed to be spherically symmetric, i.e. $V_{\text{ext}}(\mathbf{x}) = V_{\text{ext}}(r)$ and $\psi_0(\mathbf{x}) = \psi_0(r)$ with $r = |\mathbf{x}|$. In this case, the solution ψ of (1.1)-(1.3) and the ground states ϕ_g are also spherically symmetric, i.e.,

$$\psi(\mathbf{x}, t) = \psi(r, t), \quad t \geq 0, \quad \phi_g(\mathbf{x}) = \phi(r), \quad \mathbf{x} \in \mathbb{R}^3.$$

Thus, the SPS equation (1.1)–(1.3) collapses to the following quasi-1D problem

$$i\partial_t \psi(r,t) = -\frac{1}{2r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial \psi}{\partial r}\right) + V_{\text{ext}}(r)\psi + C_P V_P \psi - \alpha \left|\psi\right|^{\frac{2}{3}} \psi, \ t > 0, \tag{2.56}$$

$$-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial V_P(r,t)}{\partial r}\right) = \left|\psi\right|^2, \quad 0 < r < \infty, \quad t \geq 0, \tag{2.57}$$

$$\psi(r, 0) = \psi_0(r), \quad 0 \leq r < \infty, \tag{2.58}$$

with boundary conditions

$$\partial_r \psi(0,t) = \partial_r V_P(0,t) = 0, \ \lim_{r\to\infty} \psi(r,t) = 0, \ \lim_{r\to\infty} rV_P(r,t) = \frac{1}{4\pi}, \ t \geq 0, \tag{2.59}$$

due to the decay conditions of ψ and V_P, and the Green function of the Laplacian in $\mathbb{R}^3$ [92].

Introducing

$$\mathcal{U}(r,t) = 2\sqrt{\pi} r \psi(r,t), \quad \mathcal{V}(r,t) = 4\pi r V_P(r,t), \quad 0 \leq r < \infty, \quad t \geq 0, \tag{2.60}$$

a simple computation shows

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi}{\partial r}\right)=\frac{1}{2\sqrt{\pi}r}\partial_{rr}\mathcal{U},\quad \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V_P}{\partial r}\right)=\frac{1}{4\pi r}\partial_{rr}\mathcal{V}. \tag{2.61}$$

Plugging the above into (2.56)–(2.59), one can obtain

$$i\partial_t\mathcal{U}(r,t)=-\frac{1}{2}\partial_{rr}\mathcal{U}+V_{\text{ext}}(r)\mathcal{U}+\frac{C_P}{4\pi r}\mathcal{V}\mathcal{U}-\alpha\big(2\sqrt{\pi}r\big)^{-\frac{2}{3}}|\mathcal{U}|^{\frac{2}{3}}\mathcal{U},\ t>0, \tag{2.62}$$

$$-\partial_{rr}\mathcal{V}(r,t)=\frac{1}{r}|\mathcal{U}|^2,\quad 0<r<\infty,\quad t\geq 0, \tag{2.63}$$

$$\mathcal{U}(r,0)=\mathcal{U}_0(r)=2\sqrt{\pi}r\psi_0(r),\quad 0\leq r<\infty, \tag{2.64}$$

$$\mathcal{U}(0,t)=\mathcal{V}(0,t)=0,\quad \lim_{r\to\infty}\mathcal{U}(r,t)=0,\quad \lim_{r\to\infty}\mathcal{V}(r,t)=1,\quad t\geq 0. \tag{2.65}$$

Also, the above problem conserves the *mass*

$$\mathcal{N}(\mathcal{U}(\cdot,t)):=\|\mathcal{U}(\cdot,t)\|^2=\int_0^\infty|\mathcal{U}(r,t)|^2\,\mathrm{d}r=N(\psi(\cdot,t))=1,\quad t\geq 0,$$

and the *energy*

$$\begin{aligned}&\mathcal{E}(\mathcal{U}(\cdot,t))\\ &:=\int_0^\infty\left[\frac{1}{2}|\partial_r\mathcal{U}|^2+\left(V_{\text{ext}}(r)+\frac{C_P}{8\pi r}\mathcal{V}(r,t)\right)|\mathcal{U}|^2-\frac{3\alpha}{4}\left(2\sqrt{\pi}r\right)^{-\frac{2}{3}}|\mathcal{U}|^{\frac{8}{3}}\right]\mathrm{d}r\\ &=E(\psi(\cdot,t)),\quad t\geq 0.\end{aligned}$$

In what follows the problem (2.62)–(2.65) will be taken as the starting model to propose efficient numerical methods. After one obtains the solution $\mathcal{U}$ of (2.62)–(2.65), the solution ψ of (2.56)–(2.59) is obtained as

$$\psi(r,t)=\frac{1}{2\sqrt{\pi}}\begin{cases}\mathcal{U}(r,t)/r, & r>0,\\ \partial_r\mathcal{U}(r,t)=\lim_{s\to 0^+}\mathcal{U}(s,t)/s, & r=0,\end{cases}\quad t\geq 0. \tag{2.66}$$

Meanwhile, the minimization problem (2.13) to define the ground state collapses to

Find $\varphi_g\in\mathcal{S}=\{\varphi\,|\,\mathcal{E}(\varphi)<\infty,\,\mathcal{N}(\varphi)=1,\,\varphi(0)=0\}$ such that

$$\mathcal{E}_g:=\mathcal{E}(\varphi_g)=\min_{\varphi\in\mathcal{S}}\mathcal{E}(\varphi). \tag{2.67}$$

Again, after one obtains the minimizer of (2.67), the ground state ϕ_g of (2.56)–(2.58) is obtained as

$$\phi_g(r)=\frac{1}{2\sqrt{\pi}}\begin{cases}\varphi_g(r)/r, & r>0,\\ \partial_r\varphi_g(r)=\lim_{s\to 0^+}\varphi_g(s)/s, & r=0.\end{cases} \tag{2.68}$$

2.4.2 Efficient numerical methods

Backward Euler sine pseudospectral method for ground states

Choose a time step $\tau > 0$ and set $t_n = n\tau$ for $n = 0, 1, \dots$. Similar to Section 2.2, for the minimization problem (2.67), the following gradient flow with discrete normalization (GFDN) can be constructed as

$$\partial_t \varphi(r,t) = \frac{1}{2}\partial_{rr}\varphi - V_{\text{ext}}(r)\varphi - \frac{C_P}{4\pi r}\mathcal{V}(r,t)\varphi + \alpha\big(2\sqrt{\pi}r\big)^{-\frac{2}{3}}|\varphi|^{\frac{2}{3}}\varphi, \tag{2.69}$$

$$-\partial_{rr}\mathcal{V}(r,t) = \frac{1}{r}|\varphi|^2, \quad 0 < r < \infty, \quad t_n \le t < t_{n+1}, \tag{2.70}$$

$$\varphi(r, t_{n+1}) := \varphi(r, t_{n+1}^+) = \frac{\varphi(r, t_{n+1}^-)}{\|\varphi(r, t_{n+1}^-)\|}, \quad n \ge 0, \tag{2.71}$$

$$\varphi(r,0) = \varphi_0(r), \quad 0 \le r < \infty, \quad \text{with} \quad \mathcal{N}(\varphi_0) = 1, \tag{2.72}$$

$$\varphi(0,t) = \mathcal{V}(0,t) = 0, \quad \lim_{r\to\infty}\varphi(r,t) = 0, \quad \lim_{r\to\infty}\mathcal{V}(r,t) = 1, \quad t \ge 0, \tag{2.73}$$

where $\varphi(r, t_n^\pm) := \lim_{t\to t_n^\pm}\varphi(r,t)$ for $0 \le r < \infty$. In practical computation, the above problem is truncated into an interval $[0, R]$ with $R > 0$ sufficiently large, together with Dirichlet boundary conditions

$$\varphi(0,t) = \varphi(R,t) = \mathcal{V}(0,t) = 0, \quad \mathcal{V}(R,t) = 1, \quad t \ge 0.$$

Introducing a linear translation (homogenization)

$$\underline{\mathcal{V}}(r,t) = \mathcal{V}(r,t) - r/R \ \text{ for } \ 0 \le r \le R, \tag{2.74}$$

one can have,

$$-\partial_{rr}\underline{\mathcal{V}}(r,t) = -\partial_{rr}\mathcal{V}(r,t) = \frac{1}{r}|\varphi|^2, \quad 0 < r < R, \tag{2.75}$$

$$\underline{\mathcal{V}}(0,t) = \underline{\mathcal{V}}(R,t) = 0, \quad t \ge 0. \tag{2.76}$$

Then the problem is discretized in space by sine pseudospectral method and in time by a backward Euler integration similar to that used in Section 2.2. Choose a mesh size $h_r = \Delta r = R/J$ with some even integer $J > 0$, and denote the grid points as $r_j = jh_r$ for $j = 0, 1, \dots, J$. Let $\varphi_j^n \approx \varphi(r_j, t_n)$ and $\underline{\mathcal{V}}_j^n \approx \underline{\mathcal{V}}(r_j, t_n)$, and denote $\rho_j^n = \left|\varphi_j^n\right|^2/r_j$. Choosing $\varphi_j^0 = \varphi_0(r_j)$, a backward Euler sine pseudospectral discretization (BESP) reads: for $n = 0, 1, \ldots,$

$$\begin{aligned}\frac{\varphi_j^+ - \varphi_j^n}{\tau} &= -\left[V_{\text{ext}}(r_j) + \frac{C_P}{4\pi r_j}\underline{\mathcal{V}}_j^n + \frac{C_P}{4\pi R} - \alpha\big(2\sqrt{\pi}r_j\big)^{-\frac{2}{3}}\left|\varphi_j^n\right|^{\frac{2}{3}}\right]\varphi_j^+ \\ &\quad + \frac{1}{2}\left(D_{rr}^s\varphi^+\right)\big|_j, \quad j = 1, 2, \ldots, J-1, \end{aligned} \tag{2.77}$$

$$-\left(D_{rr}^s\underline{\mathcal{V}}^n\right)\big|_j = \rho_j^n, \quad j = 1, 2, \ldots, J-1, \quad \varphi_0^+ = \varphi_J^+ = \underline{\mathcal{V}}_0^n = \underline{\mathcal{V}}_J^n = 0, \tag{2.78}$$

$$\varphi_j^{n+1} = \frac{\varphi_j^+}{\|\varphi^+\|_h}, \quad j = 0, 1, \ldots, J, \quad \text{with} \quad \left\|\varphi^+\right\|_h^2 := h_r\sum_{j=1}^{J-1}\left|\varphi_j^+\right|^2, \tag{2.79}$$

where D_{rr}^s is the sine pseudospectral approximation of ∂_{rr}, defined via

$$-\left(D_{rr}^s\varphi^n\right)\big|_j = \sum_{k=1}^{J-1}\mu_k^2\widetilde{(\varphi^n)}_k\sin\left(\frac{jk\pi}{J}\right), \quad j = 1, 2, \ldots, J-1, \tag{2.80}$$

with $\widetilde{(\varphi^n)}_k$ the discrete sine transform coefficients

$$\widetilde{(\varphi^n)}_k = \frac{2}{J}\sum_{j=1}^{J-1}\varphi_j^n\sin\left(\frac{jk\pi}{J}\right), \quad \mu_k = \frac{k\pi}{R}, \quad k = 1, 2, \ldots, J-1. \tag{2.81}$$

Similar to Section 2.2, the linear system (2.77)–(2.79) can be iteratively solved efficiently in phase space with the help of discrete sine transform. After one gets the stationary solution $(\varphi_g)_j$ of the above problem, the ground state $(\phi_g)_j \approx \phi_g(r_j)$ of (2.56)–(2.58) is achieved via

$$(\phi_g)_j = \frac{1}{2\sqrt{\pi}}\begin{cases}(\varphi_g)_j/r_j, & j = 1, 2, \ldots, J, \\ \sum_{k=1}^{J-1}\mu_k\widetilde{(\varphi_g)}_k, & j = 0.\end{cases} \tag{2.82}$$

Note that the above numerical method is spectrally accurate and it works only when V_{ext} is spherically symmetric. Compared with the pseudospectral method proposed in Section 2.2 for general 3D problem, the memory cost is reduced from $O(J^3)$ to $O(J)$ and computational cost per time step is reduced from $O(J^3\ln(J^3))$ to $O(J\ln(J))$.

Time-splitting sine pseudospectral method for dynamics

Again, the problem is truncated into an interval $[0, R]$, with introducing the linear translation (2.74) for $\mathcal{V}$ into (2.62)–(2.65) such that both $\mathcal{U}$ and $\underline{\mathcal{V}}$ satisfy homogeneous Dirichlet boundary conditions. Similar to Section 2.3, for computing the dynamics, the time-splitting technique is applied to decouple the nonlinearity and then sine pseudospectral method is used to discretize the spatial derivatives. Denote $\mathcal{U}_j^n \approx \mathcal{U}(r_j, t_n)$ and $\underline{\mathcal{V}}_j^n \approx \underline{\mathcal{V}}(r_j, t_n)$ and choose $\mathcal{U}_j^0 = \mathcal{U}_0(r_j)$, a second-order time-splitting sine pseudospectral (TSSP) discretization reads

$$\mathcal{U}_j^{(1)} = \sum_{k=1}^{J-1} \exp\left(-i\tau\mu_k^2/4\right) \widetilde{(\mathcal{U}^n)}_k \sin\left(\frac{jk\pi}{J}\right), \tag{2.83}$$

$$\mathcal{U}_j^{(2)} = \exp\left[-i\tau\left(V_{\text{ext}}(r_j) + \frac{C_P}{4\pi r_j}\left(\underline{\mathcal{V}}_j^{(1)} + \frac{r_j}{R}\right) - \frac{\alpha|\mathcal{U}_j^{(1)}|^{2/3}}{(2\sqrt{\pi}r_j)^{2/3}}\right)\right] \times \mathcal{U}_j^{(1)}, \tag{2.84}$$

$$\mathcal{U}_j^{n+1} = \sum_{k=1}^{J-1} \exp\left(-i\tau\mu_k^2/4\right) \widetilde{(\mathcal{U}^{(2)})}_k \sin\left(\frac{jk\pi}{J}\right), \tag{2.85}$$

for $n \geq 0$ and $j = 1, 2, \ldots, J-1$. Here, $\underline{\mathcal{V}}_j^{(1)}$ is obtained from solving the Poisson equation via sine pseudospectral method (similar to Section 2.2), i.e.,

$$\underline{\mathcal{V}}_j^{(1)} = \sum_{k=1}^{J-1} \mu_k^{-2} \widetilde{(\rho^{(1)})}_k \sin\left(\frac{jk\pi}{J}\right), \quad \rho_j^{(1)} = \frac{1}{r_j}\left|\mathcal{U}_j^{(1)}\right|^2, \quad j = 1, 2. \ldots, J-1. \tag{2.86}$$

Again, after one gets the solution $\mathcal{U}_j^n$ from (2.83)–(2.86), the solution $\psi_j^n \approx \psi(r_j, t_n)$ of (2.56)–(2.58) is achieved via

$$\psi_j^n = \frac{1}{2\sqrt{\pi}} \begin{cases} \mathcal{U}_j^n / r_j, & j = 1, 2, \ldots, J, \\ \sum_{k=1}^{J-1} \mu_k \widetilde{(\mathcal{U}^n)}_k, & j = 0. \end{cases} \tag{2.87}$$

The above method is explicit, spectrally accurate in space and second-order accurate in time and it works only when both V_{ext} and ψ_0 are spherically symmetric. Again, compared with the method proposed in Section 2.3 for general 3D problem, the memory cost is reduced from $O(J^3)$ to $O(J)$ and computational cost per time step is reduced from $O(J^3 \ln(J^3))$ to $O(J \ln(J))$. In addition, similar to Section 2.3, one has,

Lemma 2.2. *The TSSP method (2.83)–(2.86) is normalization conservation, i.e.,*

$$\|\mathcal{U}^n\|_h^2 := h_r \sum_{j=1}^{J-1} \left|\mathcal{U}_j^n\right|^2 \equiv h_r \sum_{j=1}^{J-1} \left|\mathcal{U}_j^0\right|^2 = \left\|\mathcal{U}^0\right\|_h^2, \quad n \geq 0,$$

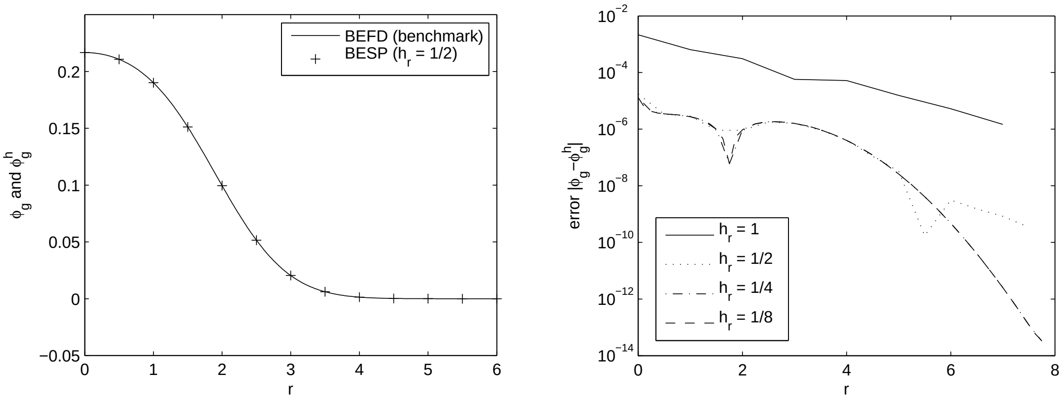


Figure 2.7: Results for spherical symmetric SPS. Accuracy analysis for BESP method: (1) ϕ_g obtained from BEFD method with $h_r = 1/64$ as benchmark and ϕ_g^h obtained from BESP method with $h_r = 1/2$ (left figure); (2) error $\left|\phi_g - \phi_g^h\right|$ with different h_r (right figure).

so it is unconditionally stable in L^2-norm.

2.4.3 Numerical results

Numerical results are reported here to demonstrate the accuracy and efficiency of the proposed simplified methods, with choosing $V_{\text{ext}} = \frac{1}{2}r^2$, $C_P = 100$ and $\alpha = 1$ in (2.56) as the example. For computing the ground states, the "exact" solution ϕ_g (benchmark) is achieved by applying a backward Euler finite-difference (BEFD) discretization to GFDN of the quasi-1D model (2.56)–(2.58) with Dirichlet boundary conditions of ϕ and Robin boundary conditions of V_P (similar to Example 2.2). ϕ_g is computed in a ball $0 \leq r \leq 8$ with a very fine mesh size $h_r = 1/64$. Let ϕ_g^h be the approximations obtained from BESP method (2.77)–(2.79), Fig. 2.7 plots ϕ_g and ϕ_g^h with $h_r = 1/2$, and the error $\left|\phi_g - \phi_g^h\right|$ with different h_r. The results show that the BESP method (2.77)–(2.79) gives the approximation of ground states with spectral order of accuracy in space; and therefore, it is more efficient in implementation than the standard finite-difference discretization for spherically symmetric case and the spectral-type method proposed in Section 2.2 for general 3D case. Similar accuracy

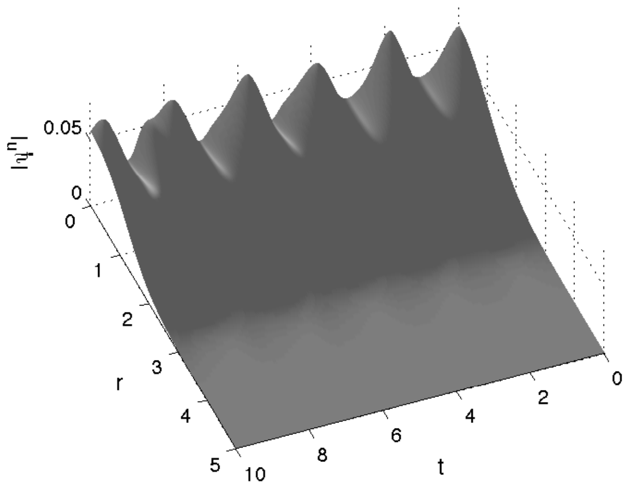


Figure 2.8: Results for spherical symmetric SPS. Dynamics computed by TSSP method: evolution of $|\psi^n|$ up to time $t_n = 10$.

and efficiency conclusions can be drawn for TSSP method (2.83)–(2.85). Fig. 2.8 plots the evolution of $|\psi^n|$ for $0 \leq t_n \leq 10$ when $\psi_0 = (2\pi)^{3/4} \exp\left(-r^2/4\right)$. Here, the computation is carried out in a ball $0 \leq r \leq 16$, with $h_r = 1/16$ and $\tau = 0.01$.

Chapter 3

Methods for the nonlinear relativistic Hartree equation

In the chapter, the computation for ground states and dynamics of the nonlinear relativistic Hartree equation (1.8) is considered. The methods proposed here can be viewed as an application of the results and observations obtained in Chapter 2 in that the relativistic Hartree equation also refers to the relativistic Schrödinger–Poisson equation.

3.1 Relativistic Hartree equation for boson stars

The nonlinear relativistic Hartree equation (1.8) was rigorously derived recently in [55] for a boson star, which refers to a quantum mechanical system of N bosons with relativistic dispersion interacting through a gravitational attractive or repulsive Coulomb potential. In fact, by starting from the N-body relativistic Schrödinger equation (replacing $-\Delta/2$ in the Schrödinger equation (2.1) to $\sqrt{-\Delta+m^2}$) and choosing the initial wave function to describe a condensate where N bosons are all in the same one-particle state, in the mean-field limit $N \to \infty$, one can prove that the time evolution of the one-particle density is governed by the nonlinear relativistic Hartree equation (under a proper non-dimensionalization) [55, 58, 59]. Also, one can refer to [9,96,97] and references therein (with a slightly different dimensionless scaling in some cases) for overviews of other physical backgrounds of (1.8).

It is easy to show that the equation (1.8) admits at least two important conserved quantities [9,55,58,59], i.e. the *mass* of the system

$$N(\psi(\cdot,t)) := \|\psi(\cdot,t)\|^2 = \int_{\mathbb{R}^3} |\psi(\mathbf{x},t)|^2 \, \mathrm{d}\mathbf{x} \equiv \int_{\mathbb{R}^3} |\psi_0(\mathbf{x})|^2 \, \mathrm{d}\mathbf{x} = 1, \ t \geq 0, \quad (3.1)$$

and the *energy*

$$\begin{aligned} E(\psi(\cdot,t)) \ &:= \int_{\mathbb{R}^3} \left[\psi^* \left(-\Delta + m^2\right)^{1/2} \psi + \left(V_{\mathrm{ext}}(\mathbf{x}) + \frac{\lambda}{2|\mathbf{x}|} * |\psi|^2 \right) |\psi|^2 \right] \mathrm{d}\mathbf{x} \\ &\equiv E(\psi_0), \quad t \geq 0. \end{aligned} \quad (3.2)$$

The well-posedness of the initial-value problem (1.8)–(1.9) was extensively studied in [9, 42, 59, 96] and references therein. Their results can be summarized as: (i) there exists a universal constant λ_{cr} (also referred to as the "Chandrasekhar limit mass" in physics [102] and with a lower bound $\lambda_{\mathrm{cr}} > 4/\pi$) such that, when $\lambda > -\lambda_{\mathrm{cr}}$, the solution is globally well-posed in the energy space $H^{1/2}(\mathbb{R}^3)$ provided that $V \in L^3(\mathbb{R}^3) \cap L^\infty(\mathbb{R}^3)$; (ii) when $\lambda \leq -\lambda_{\mathrm{cr}}$, the solution is locally well-posed; and (iii) when $\lambda < -\lambda_{\mathrm{cr}}$, the solution will blow up in finite time, which indicates the "gravitational collapse" of boson stars when the effective "mass" exceeds the critical value λ_{cr} [59]. Another problem of interests is the existence and uniqueness of the ground state for (1.8), similar to (2.13), which is defined as the minimizer of the following nonconvex minimization problem:

Find $\phi_g \in S = \left\{\phi \mid \phi \in H^{1/2}(\mathbb{R}^3), \|\phi\|^2 = 1\right\}$ such that

$$E_g := E(\phi_g) = \min_{\phi \in S} \ E(\phi). \quad (3.3)$$

If $V_{\mathrm{ext}}(\mathbf{x}) \equiv 0$, it was shown that the ground state exists iff $-\lambda_{\mathrm{cr}} < \lambda < 0$ [58,102] and any ground state is smooth, decays exponentially when $|\mathbf{x}| \to \infty$, and is identical to its spherically symmetric rearrangement up to phase and translation. Moreover, it was proven recently in [97] that, when $\lambda < 0$ and $|\lambda| \ll 1$, the spherical-symmetric ground state is unique up to phase and translation, and the author remarked there that whether such uniqueness result can be extended to the whole range of existence $-\lambda_{\rm cr} < \lambda < 0$ remains open. Thus, such critical value $\lambda_{\rm cr}$ plays an important role in investigating the ground states and dynamics of (1.8). One remark here is that based on numerical results $\lambda_{\rm cr} \approx 2.69 > 8/\pi$ (cf. Fig. 3.2).

For the Schrödinger–Poisson (or –Newton) equations, i.e. the pseudodifferential operator $\sqrt{-\Delta + m^2}$ in (1.8) is replaced by $-\Delta$ [27, 78], as discussed in Chapter 2, different numerical methods were presented in the literature based on finite difference discretization; see, e.g., [53, 54, 62, 78]. However, these numerical methods have some difficulties in discretizing the 3D relativistic Hartree equation efficiently and accurately due to the appearance of the pseudodifferential operator. The main aim of this chapter is to design efficient and accurate numerical methods for computing the ground states of (1.8) and the dynamics of the initial-value problem (1.8)–(1.9). For this purpose, let $\beta = 4\pi\lambda$ and

$$V_P(\mathbf{x},t) = \frac{1}{4\pi|\mathbf{x}|} * |\psi|^2 = \frac{1}{4\pi}\int_{\mathbb{R}^3} \frac{1}{|\mathbf{x}-\mathbf{x}'|}|\psi(\mathbf{x}',t)|^2 \, d\mathbf{x}', \quad \mathbf{x}\in\mathbb{R}^3, \quad t \geq 0,$$

then (1.8) is re-written as the relativistic Schrödinger–Poisson (RSP) equation

$$i\partial_t\psi(\mathbf{x},t) = \sqrt{-\Delta + m^2}\,\psi + V_{\rm ext}(\mathbf{x})\psi + \beta V_P\,\psi, \quad \mathbf{x}\in\mathbb{R}^3, \quad t>0, \tag{3.4}$$

$$-\Delta V_P(\mathbf{x},t) = |\psi|^2, \quad \mathbf{x}\in\mathbb{R}^3, \quad \lim_{|\mathbf{x}|\to\infty} V_P(\mathbf{x},t) = 0, \quad t \geq 0. \tag{3.5}$$

With this formulation, the energy functional (3.2) is re-written as

$$\begin{aligned} E(\psi(\cdot,t)) &= \int_{\mathbb{R}^3}\left[\psi^*\left(-\Delta+m^2\right)^{1/2}\psi + \left(V_{\rm ext}(\mathbf{x}) + \frac{\beta}{2}V_P\right)|\psi|^2\right] d\mathbf{x} \\ &= \int_{\mathbb{R}^3}\left[\left|\left(-\Delta+m^2\right)^{1/4}\psi\right|^2 + \left(V_{\rm ext}(\mathbf{x}) + \frac{\beta}{2}(-\Delta)^{-1}|\psi|^2\right)|\psi|^2\right] d\mathbf{x} \\ &= \int_{\mathbb{R}^3}\left[\left|\left(-\Delta+m^2\right)^{1/4}\psi\right|^2 + V_{\rm ext}(\mathbf{x})\,|\psi|^2 + \frac{\beta}{2}\,|\nabla V_P|^2\right] d\mathbf{x}, \quad t\geq 0. \end{aligned} \tag{3.6}$$

In order to design numerical methods for computing the ground states, similar to Chapter 2, a gradient flow with discrete normalization (GFDN) is first constructed. In the spirit of observations drawn in Chapter 2, the problem is then truncated into a box with homogeneous Dirichlet boundary conditions and a backward Euler sine pseudospectral method is applied to discretize it. For computing the dynamics, again the problem is truncated into a box with homogeneous Dirichlet boundary conditions and a time-splitting sine pseudospectral method is applied to discretize it. In particular, when the potential and initial data for dynamics are spherically symmetric, the problem collapses to a quasi-1D problem. Similar to Section 2.4, simplified numerical methods are designed by using a proper change of variables in the quasi-1D problem.

3.2 Numerical method for ground states

In this section, an efficient and accurate numerical method will be proposed for computing the ground states, i.e. solving the minimization problem (3.3). Similar to Chapter 2, it is readily to verify that its Euler–Lagrange equation is

$$\mu\phi(\mathbf{x}) = \sqrt{-\Delta + m^2}\,\phi(\mathbf{x}) + V_{\text{ext}}(\mathbf{x})\phi(\mathbf{x}) + \beta V_P(\mathbf{x})\phi(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^3, \tag{3.7}$$

$$-\Delta V_P(\mathbf{x}) = |\phi(\mathbf{x})|^2, \quad \mathbf{x}\in\mathbb{R}^3, \quad \lim_{|\mathbf{x}|\to\infty} V_P(\mathbf{x}) = 0, \tag{3.8}$$

under the constraint

$$\|\phi\|^2 := \int_{\mathbb{R}^3} |\phi(\mathbf{x})|^2 d\mathbf{x} = 1, \tag{3.9}$$

where the eigenvalue μ is usually called as the chemical potential in physics literature, which can be obtained by

$$\begin{aligned}\mu(\phi) &= \int_{\mathbb{R}^3}\left[\left|\left(-\Delta+m^2\right)^{1/4}\phi\right|^2 + \left(V_{\text{ext}}(\mathbf{x}) + \beta V_P\right)|\phi|^2\right] d\mathbf{x}\\ &= E(\phi) + \frac{\beta}{2}\int_{\mathbb{R}^3} V_P|\phi|^2 d\mathbf{x}.\end{aligned} \tag{3.10}$$

In fact, the above nonlinear eigenvalue problem can also be obtained by taking the ansatz

$$\psi(\mathbf{x},t) = e^{-i\mu t}\phi(\mathbf{x}), \quad \mathbf{x}\in\mathbb{R}^3, \quad t\geq 0, \tag{3.11}$$

in (3.4)–(3.5). Thus it is also called as the time-independent relativistic Schrödinger–Poisson equation.

3.2.1 Gradient flow with discrete normalization

In order to solve the nonconvex minimization problem (3.3) efficiently, the gradient flow with discrete normalization (GFDN) is constructed following the procedure in [5, 18, 41] and Chapter 2. Choose a time step $\tau = \Delta t > 0$ and set $t_n = n\tau$ for $n = 0, 1, \dots$. Applying the steepest decent method to the energy functional $E(\phi)$ in (3.2) without the constraint (3.9), and then projecting the solution back to the unit sphere S at the end of each time interval $[t_n, t_{n+1}]$ to enforce the constraint (3.9), one comes to the following gradient flow with discrete normalization in 3D (GFDN-3D) for $\phi(\mathbf{x}, t)$:

$$\partial_t\phi(\mathbf{x},t) = -\frac{1}{2}\frac{\delta E(\phi)}{\delta\phi} = -\sqrt{-\Delta + m^2}\,\phi - V_{\text{ext}}(\mathbf{x})\phi - \beta V_P\phi, \quad t_n \le t < t_{n+1}, \tag{3.12}$$

$$-\Delta V_P(\mathbf{x},t) = |\phi|^2, \quad \mathbf{x} \in \mathbb{R}^3, \quad \lim_{|\mathbf{x}|\to\infty} V_P(\mathbf{x},t) = 0, \quad t \ge 0, \tag{3.13}$$

$$\phi(\mathbf{x}, t_{n+1}) := \phi(\mathbf{x}, t_{n+1}^+) = \frac{\phi(\mathbf{x}, t_{n+1}^-)}{\|\phi(\mathbf{x}, t_{n+1}^-)\|}, \quad \mathbf{x} \in \mathbb{R}^3, \quad n \ge 0, \tag{3.14}$$

$$\phi(\mathbf{x}, 0) = \phi_0(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^3, \quad \text{with} \quad \|\phi_0\|^2 = \int_{\mathbb{R}^3} |\phi_0(\mathbf{x})|^2\,\mathrm{d}\mathbf{x} = 1, \tag{3.15}$$

where $\phi(\mathbf{x}, t_n^\pm) := \lim_{t\to t_n^\pm} \phi(\mathbf{x}, t)$. Again, the gradient flow (3.12) can also be obtained from (3.4) by setting time t to $\tilde{t} = it$, thus the above construction is also referred to as the imaginary time method in physics literature [45, 95, 126].

Letting $\tau \to 0$ in the GFDN-3D (3.12)–(3.15), similar to Chapter 2, one can obtain the following continuous normalized gradient flow (CNGF):

$$\partial_t\phi(\mathbf{x},t) = -\sqrt{-\Delta + m^2}\,\phi - V_{\text{ext}}(\mathbf{x})\phi - \beta V_P\phi + \frac{\mu(\phi)}{\|\phi\|^2}\phi, \quad t > 0, \tag{3.16}$$

$$-\Delta V_P(\mathbf{x},t) = |\phi|^2, \quad \mathbf{x} \in \mathbb{R}^3, \quad \lim_{|\mathbf{x}|\to\infty} V_P(\mathbf{x},t) = 0, \quad t \ge 0. \tag{3.17}$$

It is easy to justify that the above CNGF is normalization conserved and energy diminishing, i.e.,

$$\|\phi(\mathbf{x},t)\|^2 \equiv \|\phi_0\|^2 = 1, \quad \frac{\mathrm{d}}{\mathrm{d}t}E(\phi(\mathbf{x},t)) = -2\|\partial_t\phi(\mathbf{x},t)\|^2 \le 0, \quad t \ge 0.$$

Thus similar to the discussions made in Chapter 2, the positive ground state $\phi_g(\mathbf{x})$ can be obtained as the steady state solution of the GFDN-3D (3.12)–(3.15) or CNGF (3.16)–(3.17) with a positive initial data $\phi_0(\mathbf{x}) \geq 0$ for $\mathbf{x} \in \mathbb{R}^3$.

3.2.2 Backward Euler sine pseudospectral discretization

Similar to Chapter 2, in practical computation the whole space problem (3.12)–(3.15) is usually truncated into a bounded computation domain $\Omega = [a,b] \times [c,d] \times [e,f]$ for $|a|$, b, $|c|$, d, $|e|$ and f sufficiently large with homogeneous Dirichlet boundary conditions on $\partial\Omega$, i.e.,

$$\partial_t \phi(\mathbf{x},t) = -\sqrt{-\Delta + m^2}\,\phi - V_{\text{ext}}(\mathbf{x})\phi - \beta V_P \phi, \quad \mathbf{x} \in \Omega,\ t_n \leq t < t_{n+1}, \tag{3.18}$$

$$-\Delta V_P(\mathbf{x},t) = |\phi|^2, \quad \phi(\mathbf{x},t)|_{\partial\Omega} = V_P(\mathbf{x},t)|_{\partial\Omega} = 0, \quad t \geq 0, \tag{3.19}$$

$$\phi(\mathbf{x},t_{n+1}) := \phi(\mathbf{x},t_{n+1}^+) = \frac{\phi(\mathbf{x},t_{n+1}^-)}{\|\phi(\mathbf{x},t_{n+1}^-)\|}, \quad n \geq 0, \tag{3.20}$$

$$\phi(\mathbf{x},0) = \phi_0(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad \text{with} \quad \|\phi_0\|^2 = \int_\Omega |\phi_0(\mathbf{x})|^2\, \mathrm{d}\mathbf{x} = 1. \tag{3.21}$$

Let J, K, L be even positive integers and define the index sets,

$$\mathcal{T}_{JKL} = \{(j,k,l) : j = 1,2,\ldots,J-1,\ k = 1,2,\ldots,K-1,\ l = 1,2,\ldots,L-1\},$$
$$\mathcal{T}_{JKL}^0 = \{(j,k,l) : j = 0,1,\ldots,J,\ k = 0,1,\ldots,K,\ l = 0,1,\ldots,L\}.$$

Choose mesh sizes $h_x = (b-a)/J$, $h_y = (d-c)/K$ and $h_z = (f-e)/L$, let $h = \max\{h_x, h_y, h_z\}$, and define the grids

$$x_j = a + jh_x, \quad y_k = c + kh_y, \quad z_l = e + lh_z, \quad (j,k,l) \in \mathcal{T}_{JKL}^0.$$

Denote

$$Y_{JKL} = \text{span}\left\{\Phi_{pqs}(\mathbf{x}),\ \mathbf{x} \in \Omega,\ (p,q,s) \in \mathcal{T}_{JKL}\right\},$$

with

$$\Phi_{pqs}(\mathbf{x}) = \sin\left(\mu_p^x(x-a)\right)\sin\left(\mu_q^y(y-c)\right)\sin\left(\mu_s^z(z-e)\right), \quad \mathbf{x} \in \Omega,\ (p,q,s) \in \mathcal{T}_{JKL},$$

$$\mu_p^x = \frac{\pi p}{b-a}, \quad \mu_q^y = \frac{\pi q}{d-c}, \quad \mu_s^z = \frac{\pi s}{f-e}, \quad (p,q,s) \in \mathcal{T}_{JKL},$$

and $\mathcal{P}_{JKL} : Y = \{U(\mathbf{x}) \in C(\Omega) : U(\mathbf{x})|_{\partial\Omega} = 0\} \to Y_{JKL}$ the standard projection operator [71, 80, 133], i.e.,

$$\left(\mathcal{P}_{JKL}U\right)(\mathbf{x}) = \sum_{(p,q,s)\in\mathcal{T}_{JKL}} \widehat{U}_{pqs}\Phi_{pqs}(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad \forall U \in Y,$$

with $\widehat{U}_{pqs}$ the sine transform coefficients

$$\widehat{U}_{pqs} = \frac{8}{(b-a)(d-c)(f-e)} \int_\Omega U(\mathbf{x})\Phi_{pqs}(\mathbf{x})\mathrm{d}\mathbf{x}, \quad (p,q,s) \in \mathcal{T}_{JKL}. \tag{3.22}$$

Choosing $\phi^0(\mathbf{x}) = \left(\mathcal{P}_{JKL}\phi_0\right)(\mathbf{x})$, a backward Euler sine spectral discretization for (3.12)–(3.13) reads

Find $\phi^{n+1}(\mathbf{x}) \in Y_{JKL}$ (i.e. $\phi^+(\mathbf{x}) \in Y_{JKL}$) and $V_P^n(\mathbf{x}) \in Y_{JKL}$, such that,

$$\frac{\phi^+(\mathbf{x}) - \phi^n(\mathbf{x})}{\tau} = -\sqrt{-\Delta + m^2}\,\phi^+(\mathbf{x}) - \mathcal{P}_{JKL}\left\{\left(V_{\text{ext}}(\mathbf{x}) + \beta V_P^n(\mathbf{x})\right)\phi^+(\mathbf{x})\right\}, \tag{3.23}$$

$$-\Delta V_P^n(\mathbf{x}) = \left(\mathcal{P}_{JKL}\,|\phi^n|^2\right)(\mathbf{x}), \quad \phi^{n+1}(\mathbf{x}) = \frac{\phi^+(\mathbf{x})}{\|\phi^+(\mathbf{x})\|}, \quad \mathbf{x} \in \Omega, \quad n \geq 0. \tag{3.24}$$

The above discretization can be solved in phase space, but it is not suitable in practical computation due to the difficulty in evaluating the integrals in (3.22). Instead, an efficient implementation can be carried out by choosing $\phi^0(\mathbf{x})$ as the interpolation of $\phi_0(\mathbf{x})$ on the grids $\{(x_j, y_k, z_l),\ (j,k,l) \in \mathcal{T}^0_{JKL}\}$ and approximating the integrals in (3.22) by a quadrature rule on the grids [57, 133]. Let ϕ^n_{jkl} and $(V_P)^n_{jkl}$ be the approximations of $\phi(x_j, y_k, z_l, t_n)$ and $V_P(x_j, y_k, z_l, t_n)$, respectively, and denote $\rho^n_{jkl} = |\phi^n_{jkl}|^2$ and $V_{jkl} = V_{\text{ext}}(x_j, y_k, z_l)$ for $(j,k,l) \in \mathcal{T}^0_{JKL}$. Choosing $\phi^0_{jkl} = \phi_0(x_j, y_k, z_l)$ for $(j,k,l) \in \mathcal{T}^0_{JKL}$, for $n = 0, 1, \ldots,$ a backward Euler sine pseudospectral discretization in 3D (BESP-3D) for (3.12)–(3.13) reads

$$\frac{\phi^+_{jkl}-\phi^n_{jkl}}{\tau} = -\left(\sqrt{-\Delta^s+m^2}\,\phi^+\right)\Big|_{jkl} - \left(V_{jkl}+\beta (V_P)^n_{jkl}\right)\phi^+_{jkl}, \tag{3.25}$$

$$-\left(\Delta^s (V_P)^n\right)\big|_{jkl} = \rho^n_{jkl}, \quad \phi^{n+1}_{jkl} = \frac{\phi^+_{jkl}}{\|\phi^+\|_h}, \quad (j,k,l)\in\mathcal{T}_{JKL}, \tag{3.26}$$

$$\phi^{n+1}_{0kl}=\phi^{n+1}_{Jkl}=\phi^{n+1}_{j0l}=\phi^{n+1}_{jKl}=\phi^{n+1}_{jk0}=\phi^{n+1}_{jkL}=0, \quad (j,k,l)\in\mathcal{T}^0_{JKL}, \tag{3.27}$$

$$\begin{aligned}(V_P)^{n+1}_{0kl} &= (V_P)^{n+1}_{Jkl} = (V_P)^{n+1}_{j0l} = (V_P)^{n+1}_{jKl}\\ &= (V_P)^{n+1}_{jk0} = (V_P)^{n+1}_{jkL} = 0, \quad (j,k,l)\in\mathcal{T}^0_{JKL},\end{aligned} \tag{3.28}$$

where Δ^s is the sine pseudospectral approximation [57, 133] of the Laplacian Δ, defined as

$$-\left(\Delta^s\phi^n\right)\big|_{jkl} = \sum_{(p,q,s)\in\mathcal{T}_{JKL}} \Xi_{pqs}\,\widetilde{(\phi^n)}_{pqs}\Phi_{pqs}(x_j,y_k,z_l), \quad (j,k,l)\in\mathcal{T}_{JKL},$$

and the approximation to the operator $\sqrt{-\Delta+m^2}$ is defined as, for $(j,k,l)\in\mathcal{T}_{JKL}$,

$$\left(\sqrt{-\Delta^s+m^2}\,\phi^n\right)\Big|_{jkl} = \sum_{(p,q,s)\in\mathcal{T}_{JKL}} \sqrt{\Xi_{pqs}+m^2}\,\widetilde{(\phi^n)}_{pqs}\Phi_{pqs}(x_j,y_k,z_l),$$

with

$$\Xi_{pqs} = (\mu^x_p)^2+(\mu^y_q)^2+(\mu^z_s)^2, \quad (p,q,s)\in\mathcal{T}_{JKL}, \tag{3.29}$$

$\widetilde{(\phi^n)}_{pqs}$ ($(p,q,s)\in\mathcal{T}_{JKL}$) the discrete sine transform coefficients defined as

$$\widetilde{(\phi^n)}_{pqs} = \frac{8}{JKL}\sum_{(j,k,l)\in\mathcal{T}_{JKL}} \phi^n_{jkl}\Phi_{pqs}(x_j,y_k,z_l), \quad (p,q,s)\in\mathcal{T}_{JKL}, \tag{3.30}$$

and the discrete l^2-norm $\|\cdot\|_h$ defined as

$$\|\phi^+\|^2_h = h_x h_y h_z \sum_{(j,k,l)\in\mathcal{T}_{JKL}} \left|\phi^+_{jkl}\right|^2.$$

Similar to Chapter 2, the linear system (3.25)–(3.28) can be iteratively solved efficiently in phase space with the help of discrete sine transform and the details are omitted here for brevity. In fact, the above numerical method is spectrally accurate, works for general potential $V_{\rm ext}(\mathbf{x})$ and its memory cost is $O(JKL)$.

3.3 Numerical method for dynamics

In this section, an efficient and accurate numerical method is presented for computing the dynamics of the RSP equation (3.4)–(3.5) with the initial condition (1.9). Again, the whole space problem is truncated into a bounded computation domain $\Omega = [a,b] \times [c,d] \times [e,f]$ with homogeneous Dirichlet boundary conditions on $\partial\Omega$, i.e.,

$$i\partial_t \psi(\mathbf{x},t) = \sqrt{-\Delta + m^2}\,\psi + V_{\rm ext}(\mathbf{x})\psi + \beta V_P\,\psi, \quad \mathbf{x} \in \Omega, \quad t > 0, \tag{3.31}$$

$$-\Delta V_P(\mathbf{x},t) = |\psi|^2, \quad \mathbf{x} \in \Omega, \quad \psi(\mathbf{x},t)|_{\partial\Omega} = V_P(\mathbf{x},t)|_{\partial\Omega} = 0, \quad t \geq 0, \tag{3.32}$$

$$\psi(\mathbf{x},0) = \psi_0(\mathbf{x}), \quad \mathbf{x} \in \Omega. \tag{3.33}$$

In order to discretize the above system, the time-splitting technique is applied to decouple the nonlinearity. From time $t = t_n$ to $t = t_{n+1}$, one first solves

$$i\partial_t \psi(\mathbf{x},t) = \sqrt{-\Delta + m^2}\,\psi, \quad \mathbf{x} \in \Omega, \quad \psi(\mathbf{x},t)|_{\partial\Omega} = 0, \quad t_n \leq t \leq t_{n+1}, \tag{3.34}$$

for the time step of length τ, followed by solving

$$i\partial_t \psi(\mathbf{x},t) = [V_{\rm ext}(\mathbf{x}) + \beta V_P(\mathbf{x},t)]\,\psi(\mathbf{x},t), \quad \mathbf{x} \in \Omega, \quad t_n \leq t \leq t_{n+1}, \tag{3.35}$$

$$-\Delta V_P(\mathbf{x},t) = |\psi(\mathbf{x},t)|^2, \quad \mathbf{x} \in \Omega, \quad \psi(\mathbf{x},t)|_{\partial\Omega} = V_P(\mathbf{x},t)|_{\partial\Omega} = 0, \tag{3.36}$$

for the same time step. Similar to (3.23), equation (3.34) will be discretized in space by sine spectral method [71, 80, 133], and then in phase space integrated *exactly* in time. For $t_n \leq t \leq t_{n+1}$, (3.35)–(3.36) leaves $|\psi|$ (and V_P) invariant in time t, i.e.

$$|\psi(\mathbf{x},t)| \equiv |\psi(\mathbf{x},t_n)|, \quad V_P(\mathbf{x},t) \equiv V_P(\mathbf{x},t_n), \quad t_n \leq t \leq t_{n+1}, \quad \mathbf{x} \in \Omega.$$

Plugging them into (3.35) and (3.36),

$$i\partial_t \psi(\mathbf{x},t) = [V_{\rm ext}(\mathbf{x}) + \beta V_P(\mathbf{x},t_n)]\,\psi(\mathbf{x},t), \quad \mathbf{x} \in \Omega, \quad t_n \leq t \leq t_{n+1}, \tag{3.37}$$

$$-\Delta V_P(\mathbf{x},t_n) = |\psi(\mathbf{x},t_n)|^2, \quad \mathbf{x} \in \Omega, \quad \psi(\mathbf{x},t_n)|_{\partial\Omega} = V_P(\mathbf{x},t_n)|_{\partial\Omega} = 0. \tag{3.38}$$

Again, (3.38) will be discretized in space by sine spectral method and the linear ODE (3.37) will be integrated in time *exactly*.

Similar to Chapter 2 and previous section, in practical computation, the above sine spectral method will be replaced by sine pseudospectral method [57, 133]. Let ψ_{jkl}^n and $(V_P)_{jkl}^n$ be the approximations of $\psi(x_j, y_k, z_l, t_n)$ and $V_P(x_j, y_k, z_l, t_n)$, respectively, and choose $\psi_{jkl}^0 = \psi_0(x_j, y_k, z_l)$ for $(j, k, l) \in \mathcal{T}_{JKL}^0$. Here, a detailed second-order time-splitting sine pseudospectral discretization in 3D (TSSP-3D) for (3.31)–(3.33) is given as [20, 21, 23, 139]

$$\begin{aligned}
\psi_{jkl}^{(1)} &= \sum_{(p,q,s)\in\mathcal{T}_{JKL}} \exp\left(-\frac{i\tau}{2}\sqrt{\Xi_{pqs}+m^2}\right)\widetilde{(\psi^n)}_{pqs}\Phi_{pqs}(x_j, y_k, z_l), \\
\psi_{jkl}^{(2)} &= \exp\left[-i\tau\left(V_{jkl}+\beta(V_P)_{jkl}^{(1)}\right)\right]\psi_{jkl}^{(1)}, \quad (j,k,l)\in\mathcal{T}_{JKL}, \\
\psi_{jkl}^{n+1} &= \sum_{(p,q,s)\in\mathcal{T}_{JKL}} \exp\left(-\frac{i\tau}{2}\sqrt{\Xi_{pqs}+m^2}\right)\widetilde{(\psi^{(2)})}_{pqs}\Phi_{pqs}(x_j, y_k, z_l), \quad n \geq 0,
\end{aligned} \tag{3.39}$$

where Ξ_{pqs} is defined in (3.29), $\widetilde{(\psi^n)}_{pqs}$ and $\widetilde{(\psi^{(2)})}_{pqs}$ are the discrete sine transform coefficients of ψ^n and $\psi^{(2)}$, respectively, which are defined similar to (3.30), and

$$(V_P)_{jkl}^{(1)} = \sum_{(p,q,s)\in\mathcal{T}_{JKL}} \frac{1}{\Xi_{pqs}}\widetilde{(|\psi^{(1)}|^2)}_{pqs}\Phi_{pqs}(x_j, y_k, z_l), \quad (j,k,l)\in\mathcal{T}_{JKL}.$$

The above method is explicit, spectrally accurate in space and second-order accurate in time. Its memory cost is $O(JKL)$ and computation cost per time step is $O(JKL\ln(JKL))$. It works for general potential $V_{\rm ext}(\mathbf{x})$ and initial data $\psi_0(\mathbf{x})$. In addition, following the analogous proof in [20, 21], one can have

Lemma 3.1. *The TSSP-3D method (3.39) is normalization conservation, i.e.,*

$$\|\psi^n\|_h^2 := h_x h_y h_z \sum_{(j,k,l)\in\mathcal{T}_{JKL}} |\psi_{jkl}^n|^2 \equiv h_x h_y h_z \sum_{(j,k,l)\in\mathcal{T}_{JKL}} |\psi_{jkl}^0|^2 = \left\|\psi^0\right\|_h^2, \quad n \geq 0.$$

Hence the method is unconditionally stable in L^2.

By using the Parsaval's equality, the total energy and chemical potential can be approximated via the composite trapezoid quadrature, i.e.,

$$\begin{aligned}
E(\psi(\mathbf{x}, t_n)) &\approx E_h(\psi^n) = E_h^{\rm kin}(\psi^n) + E_h^{\rm exp}(\psi^n) + E_h^{\rm inp}(\psi^n), \\
\mu(\psi(\mathbf{x}, t_n)) &\approx \mu_h(\psi^n) = E_h^{\rm kin}(\psi^n) + E_h^{\rm exp}(\psi^n) + 2E_h^{\rm inp}(\psi^n), \quad n \geq 0,
\end{aligned}$$

where the kinetic energy, external potential energy and internal potential energy are defined as

$$\begin{aligned}
E_h^{\rm kin}(\psi^n) &= h_x h_y h_z \sum_{(j,k,l)\in\mathcal{T}_{JKL}} (\psi^n_{jkl})^* \left(-\Delta^s + m^2\right)^{1/2} \psi^n_{jkl}, \\
&= \frac{(b-a)(d-c)(f-e)}{8} \sum_{(p,q,s)\in\mathcal{T}_{JKL}} \sqrt{\Xi_{pqs}+m^2} \left|\widetilde{(\psi^n)}_{pqs}\right|^2, \\
E_h^{\rm exp}(\psi^n) &= h_x h_y h_z \sum_{(j,k,l)\in\mathcal{T}_{JKL}} V_{jkl} \left|\psi^n_{jkl}\right|^2, \quad n \ge 0, \\
\\
E_h^{\rm inp}(\psi^n) &= \frac{\beta h_x h_y h_z}{2} \sum_{(j,k,l)\in\mathcal{T}_{JKL}} (V_P)^n_{jkl} \left|\psi^n_{jkl}\right|^2 \\
&= \frac{(b-a)(d-c)(f-e)\beta}{16} \sum_{(p,q,s)\in\mathcal{T}_{JKL}} \frac{1}{\Xi_{pqs}} \left|\widetilde{(|\psi^n|^2)}_{pqs}\right|^2.
\end{aligned}$$

3.4 Simplified methods for spherical symmetry

In this section, the potential $V_{\rm ext}$ and initial data ψ_0 are assumed to be spherically symmetric, i.e. $V_{\rm ext}(\mathbf{x}) = V_{\rm ext}(r)$ and $\psi_0(\mathbf{x}) = \psi_0(r)$ with $r = |\mathbf{x}|$ for $\mathbf{x} \in \mathbb{R}^3$. Similar to Chapter 2, by using a proper change of variables BESP-3D and TSSP-3D methods in previous sections are simplified such that the memory cost (with $J = K = L$) is reduced from $O(J^3)$ to $O(J)$ and computation cost per step is reduced from $O(J^3 \ln(J^3))$ to $O(J \ln(J))$.

3.4.1 Quasi-1D problems

Under the spherically symmetric assumption, the solution ψ of (3.4)–(3.5) with the initial condition (1.9) and the ground state ϕ_g are also spherically symmetric, i.e.,

$$\psi(\mathbf{x}, t) = \psi(r, t), \quad \phi_g(\mathbf{x}) = \phi_g(r), \quad \mathbf{x} \in \mathbb{R}^3, \quad t \ge 0.$$

Thus, the RSP equation (3.4)–(3.5) collapses to

$$i\partial_t\psi(r,t) = \left[-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+m^2\right]^{1/2}\psi + V_{\text{ext}}(r)\psi + \beta V_P\,\psi, \quad t>0, \tag{3.40}$$

$$-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V_P(r,t)}{\partial r}\right) = |\psi|^2, \quad 0<r<\infty, \quad t\geq 0, \tag{3.41}$$

$$\partial_r\psi(0,t) = \partial_r V_P(0,t) = \lim_{r\to\infty}\psi(r,t) = 0, \quad \lim_{r\to\infty} rV_P(r,t) = \frac{1}{4\pi},\ t\geq 0, \tag{3.42}$$

with initial condition

$$\psi(r,0) = \psi_0(r), \quad 0\leq r<\infty. \tag{3.43}$$

Also, the normalization (3.1) collapses to

$$N(\psi(\cdot,t)) = 4\pi\int_0^\infty |\psi(r,t)|^2\, r^2\,\mathrm{d}r \equiv 4\pi\int_0^\infty |\psi_0(r)|^2\, r^2\,\mathrm{d}r = 1,\ t\geq 0, \tag{3.44}$$

and the energy (3.6) collapses to

$$\begin{aligned} &E(\psi(\cdot,t)) \\ &= \ 4\pi\int_0^\infty \left[\psi^*\left(-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+m^2\right)^{1/2}\psi + \left(V_{\text{ext}}(r)+\frac{\beta}{2}V_P\right)|\psi|^2\right] r^2\,\mathrm{d}r \\ &\equiv \ E(\psi_0), \quad t\geq 0. \end{aligned}$$

Similar to Chapter 2, introducing

$$\mathcal{U}(r,t) = 2\sqrt{\pi}r\,\psi(r,t), \quad \mathcal{V}(r,t) = 4\pi r\,V_P(r,t), \quad 0\leq r<\infty, \quad t\geq 0, \tag{3.45}$$

a detailed computation leeds to

$$\begin{aligned} &\left[-\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+m^2\right]^{1/2}\psi = \frac{1}{2\sqrt{\pi}r}\left(-\partial_{rr}+m^2\right)^{1/2}\mathcal{U}, \\ &\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial V_P}{\partial r}\right) = \frac{1}{4\pi r}\partial_{rr}\mathcal{V}, \quad 0<r<\infty, \quad t>0. \end{aligned}$$

Plugging the above equations and (3.45) into (3.40)–(3.42),

$$i\partial_t\mathcal{U} = \left(-\partial_{rr}+m^2\right)^{1/2}\mathcal{U} + V_{\text{ext}}(r)\mathcal{U} + \frac{\beta}{4\pi r}\mathcal{V}\mathcal{U}, \quad 0<r<\infty, \quad t>0, \tag{3.46}$$

$$-\partial_{rr}\mathcal{V} = \frac{1}{r}|\mathcal{U}|^2, \quad 0<r<\infty, \quad \lim_{r\to\infty}\mathcal{V}(r,t) = 1, \quad t\geq 0, \tag{3.47}$$

$$\mathcal{U}(0,t) = \mathcal{V}(0,t) = \lim_{r\to\infty}\mathcal{U}(r,t) = 0, \quad t\geq 0, \tag{3.48}$$

with initial condition

$$\mathcal{U}(r,0) = \mathcal{U}_0(r) = 2\sqrt{\pi}r\,\psi_0(r), \quad 0 \le r < \infty. \tag{3.49}$$

Again, it is easy to show that the above problem conserves the *mass*

$$\mathcal{N}(\mathcal{U}(\cdot,t)) := \|\mathcal{U}(\cdot,t)\|^2 = \int_0^\infty |\mathcal{U}(r,t)|^2\,\mathrm{d}r \equiv \int_0^\infty |\mathcal{U}_0(r)|^2\,\mathrm{d}r, \quad t \ge 0, \tag{3.50}$$

and the *energy*

$$\begin{aligned}\mathcal{E}(\mathcal{U}(\cdot,t)) &:= \int_0^\infty \left[\mathcal{U}^*\left(-\partial_{rr}+m^2\right)^{1/2}\mathcal{U} + \left(V_{\text{ext}}(r) + \frac{\beta}{8\pi r}\mathcal{V}\right)|\mathcal{U}|^2\right]\mathrm{d}r \\ &\equiv \mathcal{E}(\mathcal{U}_0), \quad t \ge 0.\end{aligned} \tag{3.51}$$

Plugging (3.45) into (3.50) and (3.51),

$$N(\psi(\cdot,t)) = \mathcal{N}(\mathcal{U}(\cdot,t)) \equiv 1, \quad E(\psi(\cdot,t)) = \mathcal{E}(\mathcal{U}(\cdot,t)), \quad t \ge 0.$$

Similar to Chapter 2, after one gets the solution $\mathcal{U}$ of (3.46)–(3.49), the solution ψ of (3.40)–(3.43) can be obtained as

$$\psi(r,t) = \frac{1}{2\sqrt{\pi}}\begin{cases}\mathcal{U}(r,t)/r, & r > 0,\\ \partial_r\mathcal{U}(0,t) = \lim_{s\to 0^+}\mathcal{U}(s,t)/s, & r = 0,\end{cases} \quad t \ge 0.$$

Meanwhile, the minimization problem (3.3) for ground state collapses to

Find $\varphi_g \in S = \left\{\varphi \mid \varphi \in H^{1/2}([0,\infty)), \varphi(0) = 0, \|\varphi\|^2 = \int_0^\infty |\varphi|^2\mathrm{d}r = 1\right\}$ such that

$$\mathcal{E}_g := \mathcal{E}(\varphi_g) = \min_{\varphi\in S}\ \mathcal{E}(\varphi). \tag{3.52}$$

Again, after one gets the ground state φ_g of (3.52), the solution ϕ_g of (3.3) can be obtained as

$$\phi_g(r) = \frac{1}{2\sqrt{\pi}}\begin{cases}\varphi_g(r)/r, & r > 0,\\ \partial_r\varphi_g(0) = \lim_{s\to 0^+}\varphi_g(s)/s, & r = 0.\end{cases}$$

3.4.2 Sine pseudospectral methods

Similar to Chapter 2 and Section 3.2, for computing the minimizer of (3.52), the following gradient flow with discrete normalization in 1D (GFDN-1D) is constructed,

$$\partial_t\varphi = -\left(-\partial_{rr} + m^2\right)^{1/2}\varphi - V_{\text{ext}}(r)\varphi - \frac{\beta}{4\pi r}\mathcal{V}\,\varphi, \quad 0 < r < \infty, \quad t_n \le t < t_{n+1}, \tag{3.53}$$

$$-\,\partial_{rr}\mathcal{V} = \frac{1}{r}|\varphi|^2, \quad 0 < r < \infty, \quad \lim_{r\to\infty}\mathcal{V}(r,t) = 1, \quad t \ge 0, \tag{3.54}$$

$$\varphi(r,t_{n+1}) := \varphi(r,t_{n+1}^+) = \frac{\varphi(r,t_{n+1}^-)}{\|\varphi(r,t_{n+1}^-)\|}, \quad n \ge 0, \tag{3.55}$$

$$\varphi(0,t) = \mathcal{V}(0,t) = \lim_{r\to\infty}\varphi(r,t) = 0, \quad t \ge 0, \tag{3.56}$$

$$\varphi(r,0) = \varphi_0(r), \quad 0 \le r < \infty, \quad \text{with} \quad \|\varphi_0\|^2 = \int_0^\infty |\varphi_0(r)|^2\,dr = 1, \tag{3.57}$$

where $\varphi(r,t_n^\pm) := \lim_{t\to t_n^\pm}\varphi(r,t)$ for $0 \le r < \infty$.

Again, in practical computation, the above GFDN-1D is truncated into an interval $[0,R]$ with $R > 0$ sufficiently large with homogeneous Dirichlet boundary conditions

$$\varphi(0,t) = \varphi(R,t) = \mathcal{V}(0,t) = 0, \quad t \ge 0.$$

Also, in order to implement sine pseudospectral discretization in space, same as (2.74), one can introduce a linear translation (homogenization)

$$\underline{\mathcal{V}}(r,t) = \mathcal{V}(r,t) - r/R \;\text{ for }\; 0 \le r \le R, \tag{3.58}$$

and have,

$$-\,\partial_{rr}\underline{\mathcal{V}}(r,t) = -\partial_{rr}\mathcal{V}(r,t) = \frac{1}{r}\left|\varphi\right|^2, \quad 0 < r < R, \tag{3.59}$$

$$\underline{\mathcal{V}}(0,t) = \underline{\mathcal{V}}(R,t) = 0, \quad t \ge 0. \tag{3.60}$$

Then it is discretized in space by sine pseudospectral method and in time by back Euler method. Let $J > 0$ be an even integer, choose mesh size $h_r = R/J$, and denote grid points as $r_j = jh_r$ for $j = 0, 1, \ldots, J$. Let φ_j^n and $\underline{\mathcal{V}}_j^n$ be the approximations of $\varphi(r_j,t_n)$ and $\underline{\mathcal{V}}(r_j,t_n)$, respectively, denote $V_j = V_{\text{ext}}(r_j)$ for $j = 0,1,\ldots,J$ and $\rho_j^n = \left|\varphi_j^n\right|^2/r_j$ for $j = 1,2,\ldots,J-1$. Choosing $\varphi_j^0 = \varphi_0(r_j)$ for $j = 0,1,\ldots,J$, for $n = 0,1,\ldots$, a backward Euler sine pseudospectral discretization in 1D (BESP-1D) reads

$$\frac{\varphi_j^+ - \varphi_j^n}{\tau} = -\left(\sqrt{-\partial_{rr}^s + m^2}\,\varphi^+\right)\Big|_j - \left(V_j + \frac{\beta}{4\pi r_j}\underline{\mathcal{V}}_j^n + \frac{\beta}{4\pi R}\right)\varphi_j^+, \tag{3.61}$$

$$-\left(\partial_{rr}^s \underline{\mathcal{V}}^n\right)\big|_j = \rho_j^n, \quad j = 1,2\ldots,J-1, \quad \varphi_0^+ = \varphi_J^+ = \underline{\mathcal{V}}_0 = \underline{\mathcal{V}}_J = 0, \tag{3.62}$$

$$\varphi_j^{n+1} = \frac{\varphi_j^+}{\|\varphi^+\|_h}, \quad j = 0,1,\ldots,J, \quad \text{with} \quad \left\|\varphi^+\right\|_h^2 := h_r\sum_{j=1}^{J-1}\left|\varphi_j^+\right|^2, \tag{3.63}$$

where ∂_{rr}^s is the sine pseudospectral approximation of ∂_{rr}, defined as

$$-\left(\partial_{rr}^s\varphi^n\right)\big|_j = \sum_{k=1}^{J-1}(\mu_k^r)^2\,\widetilde{(\varphi^n)}_k\,\sin\left(\frac{jk\pi}{J}\right), \quad j = 0,1,\ldots,J,$$

and the approximation to the operator $\sqrt{-\partial_{rr} + m^2}$ is defined as

$$\left(\sqrt{-\partial_{rr}^s + m^2}\,\varphi^n\right)\Big|_j = \sum_{k=1}^{J-1}\sqrt{(\mu_k^r)^2 + m^2}\,\widetilde{(\varphi^n)}_k\,\sin\left(\frac{jk\pi}{J}\right), \quad j = 0,1,\ldots,J,$$

with

$$\mu_k^r = \frac{k\pi}{R}, \quad k = 1,2,\ldots,J-1,$$

and $\widetilde{(\varphi^n)}_k$ $(k = 1,2,\ldots,J-1)$ the discrete sine transform coefficients defined as

$$\widetilde{(\varphi^n)}_k = \frac{2}{J}\sum_{j=1}^{J-1}\varphi_j^n\,\sin\left(\frac{jk\pi}{J}\right), \quad k = 1,2,\ldots,J-1. \tag{3.64}$$

Again, the linear system (3.61)–(3.63) can be iteratively solved efficiently in phase space with the help of discrete sine transform [17]. The above numerical method is spectrally accurate and it works only when $V_{\text{ext}}(\mathbf{x})$ is spherically symmetric, and its memory cost is only $O(J)$.

Similar to before, for computing the dynamics of (3.46)–(3.49), the time-splitting technique is first applied to decouple the nonlinearity and then sine pseudospectral method is used to discretize the spatial derivative. Let $\mathcal{U}_j^n$ and $\underline{\mathcal{V}}_j^n$ be the approximations of $\mathcal{U}(r_j,t_n)$ and $\underline{\mathcal{V}}(r_j,t_n)$, respectively, and choose $\mathcal{U}_j^0 = \mathcal{U}_0(r_j)$ for

$j = 0, 1, \ldots, J$. Then a second-order time-splitting sine pseudospectral discretization in 1D (TSSP-1D) [20, 21, 139] for (3.46)–(3.49) reads

$$\begin{aligned}
\mathcal{U}_j^{(1)} &= \sum_{k=1}^{J-1} \exp\left\{-\frac{i\tau}{2}\sqrt{(\mu_k^r)^2 + m^2}\right\} \widetilde{(\mathcal{U}^n)}_k \, \sin\left(\frac{jk\pi}{J}\right), \\
\mathcal{U}_j^{(2)} &= \exp\left\{-i\tau\left(V_j + \frac{\beta}{4\pi r_j}\mathcal{V}^{(1)} + \frac{\beta}{4\pi R}\right)\right\}\mathcal{U}_j^{(1)}, \quad j = 1, \ldots, J-1, \qquad (3.65) \\
\mathcal{U}_j^{n+1} &= \sum_{k=1}^{j-1} \exp\left\{-\frac{i\tau}{2}\sqrt{(\mu_k^r)^2 + m^2}\right\} \widetilde{(\mathcal{U}^{(2)})}_k \, \sin\left(\frac{jk\pi}{J}\right), \quad n \geq 0,
\end{aligned}$$

where $\widetilde{(\mathcal{U}^n)}_k$ and $\widetilde{(\mathcal{U}^{(2)})}_k$ are the discrete sine transform coefficients of $\mathcal{U}^n$ and $\mathcal{U}^{(2)}$, respectively, which are defined similar to (3.64),

$$\mathcal{V}_j^{(1)} = \sum_{k=1}^{J-1} \frac{1}{(\mu_k^r)^2} \widetilde{(\rho^n)}_k \, \sin\left(\frac{jk\pi}{J}\right), \quad j = 1, 2, \ldots, J-1,$$

with $\rho_j^n = |\mathcal{U}^{(1)}|^2/r_j$ for $j = 1, 2, \ldots, J-1$.

Again, the above method is explicit, spectrally accurate in space and second-order accurate in time, its memory cost is $O(J)$ and computational cost per time step is $O(J \ln(J))$. It works only when the potential $V_{\text{ext}}(\mathbf{x})$ and initial data $\psi_0(\mathbf{x})$ are spherically symmetric. In addition, following the analogue proof in [20, 21], one can have,

Lemma 3.2. *The TSSP-1D method (3.65) is normalization conservation, i.e.,*

$$\|\mathcal{U}^n\|_h^2 := h_r \sum_{j=1}^{J-1} \left|\mathcal{U}_j^n\right|^2 \equiv h_r \sum_{j=1}^{J-1} \left|\mathcal{U}_j^0\right|^2 = \left\|\mathcal{U}^0\right\|_h^2, \quad n \geq 0.$$

After one gets the solution $\mathcal{U}_j^n$ from (3.65), the solution ψ_j^n of (3.40)–(3.43) can be obtained as

$$\psi_j^n = \frac{1}{2\sqrt{\pi}} \begin{cases} \mathcal{U}_j^n / r_j, & j = 1, 2, \ldots, J, \\ \sum_{k=1}^{J-1} \mu_k^r \, \widetilde{(\mathcal{U}^n)}_k, & j = 0, \end{cases} \qquad n \geq 0.$$

And after one gets the ground state $(\varphi_g)_j$ $(j = 0, 1, \ldots, J)$ from (3.61)–(3.63), the solution $(\phi_g)_j$ $(j = 0, 1, \ldots, J)$ of (3.3) can be obtained as

$$(\phi_g)_j = \frac{1}{2\sqrt{\pi}} \begin{cases} (\varphi_g)_j / r_j, & j = 1, 2, \ldots, J, \\ \sum_{k=1}^{J-1} \mu_k^r \, \widetilde{(\varphi_g)}_k, & j = 0, \end{cases}$$

where $\widetilde{(\mathcal{U}^n)}_k$ and $\widetilde{(\varphi_g)}_k$ are the discrete sine transform coefficients of $\mathcal{U}^n$ and φ_g, respectively.

By using the Parsaval equality, the energy and chemical potential can also be approximated via the composite trapezoid quadrature, i.e.

$$\begin{aligned} E(\psi(\mathbf{x},t_n)) &\approx \underline{E}_h(\mathcal{U}^n) = \underline{E}_h^{\text{kin}}(\mathcal{U}^n) + \underline{E}_h^{\text{exp}}(\mathcal{U}^n) + \underline{E}_h^{\text{inp}}(\mathcal{U}^n), \\ \mu(\psi(\mathbf{x},t_n)) &\approx \underline{\mu}_h(\mathcal{U}^n) = \underline{E}_h^{\text{kin}}(\mathcal{U}^n) + \underline{E}_h^{\text{exp}}(\mathcal{U}^n) + 2\underline{E}_h^{\text{inp}}(\mathcal{U}^n), \quad n \geq 0, \end{aligned}$$

where the kinetic energy, external potential energy and internal potential energy are defined as

$$\begin{aligned} \underline{E}_h^{\text{kin}}(\mathcal{U}^n) &= h_r \sum_{j=1}^{J-1} (\mathcal{U}_j^n)^* \left(-\partial_{rr}^s + m^2\right)^{1/2} \mathcal{U}_j^n = \frac{R}{2} \sum_{k=1}^{J-1} \sqrt{(\mu_k^r)^2 + m^2} \left|\widetilde{(\mathcal{U}^n)}_k\right|^2, \\ \underline{E}_h^{\text{exp}}(\mathcal{U}^n) &= h_r \sum_{j=1}^{J-1} V_j \left|\psi_{jkl}^n\right|^2, \quad n \geq 0, \\ \underline{E}_h^{\text{inp}}(\mathcal{U}^n) &= \frac{\beta h_r}{8\pi} \sum_{j=1}^{J-1} \mathcal{V}_j^n \left(\frac{1}{r_j} \left|\mathcal{U}_j^n\right|^2\right) = \frac{\beta R}{16\pi} \sum_{k=1}^{J-1} \frac{1}{(\mu_k^r)^2} \left|\widetilde{(\rho^n)}_k\right|^2 + \frac{\beta}{8\pi R}, \end{aligned}$$

with $\rho_j^n = |\mathcal{U}_j^n|^2/r_j$ for $j = 1, 2, \ldots, J-1$.

3.4.3 Finite difference discretization

For comparison, a backward Euler finite difference (BEFD-1D) discretization can be applied to (3.53)–(3.57), after it is truncated on the interval $[0, R]$ with the linear translation (3.58), as

$$\frac{\varphi^+ - \varphi^n}{\tau} = -(A + m^2 I_{J-1})^{1/2}\, \varphi^+ - \left(F^n + \frac{\beta}{4\pi R} I_{J-1}\right) \varphi^+, \quad n \geq 0, \tag{3.66}$$

$$A\underline{\mathcal{V}}^n = \rho^n, \quad \varphi^{n+1} = \frac{\varphi^+}{\|\varphi^+\|_h}, \quad n \geq 0, \quad \varphi^0 = \varphi_0, \tag{3.67}$$

where I_{J-1} is the $(J-1) \times (J-1)$ identity matrix, $\varphi^+ = \left(\varphi_1^+, \varphi_2^+, \ldots, \varphi_{J-1}^+\right)^T$, $\varphi^n = \left(\varphi_1^n, \varphi_2^n, \ldots, \varphi_{J-1}^n\right)^T$, $\underline{\mathcal{V}}^n = \left(\underline{\mathcal{V}}_1^n, \underline{\mathcal{V}}_2^n, \ldots, \underline{\mathcal{V}}_{J-1}^n\right)^T$, $\varphi_0 = (\varphi_0(r_1), \varphi_0(r_2), \ldots, \varphi_0(r_{J-1}))^T$, $\rho^n = \left(|\varphi_1^n|^2/r_1, |\varphi_2^n|^2/r_2, \ldots, \left|\varphi_{J-1}^n\right|^2/r_{J-1}\right)^T$, $F^n = \text{diag}\{V_1 + \beta\underline{\mathcal{V}}_1^n/4\pi r_1, \ldots, V_{J-1} +$

$\beta\underline{\mathcal{V}}^n_{J-1}/4\pi r_{J-1}\}$, and A is a $(J-1)\times(J-1)$ tri-diagonal matrix defined as

$$A = \frac{1}{h_r^2}\begin{pmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & -1 & 2 \end{pmatrix}.$$

In computation, we need to factorize A as $A = Q\Lambda Q^T$ with Λ a diagonal matrix and Q an orthogonal matrix satisfying $Q^T = Q^{-1}$, then $(A + m^2 I_{J-1})^{1/2} = Q(\Lambda + m^2 I_{J-1})^{1/2}Q^T$.

Similarly, one can apply a time-splitting finite difference (TSFD-1D) discretization for (3.46)–(3.49) for dynamics after it is truncated on the interval $[0, R]$. The details are omitted here for brevity.

Remark 3.1. *If the Poisson equation (3.5) in the RSP equation is replaced by the Yukawa equation*

$$-\Delta V_P + \beta V_P = |\psi|^2, \quad \mathbf{x}\in\mathbb{R}^3, \quad \lim_{|\mathbf{x}|\to\infty} V_P(\mathbf{x},t) = 0, \quad t\geq 0,$$

with $\beta > 0$ a constant, the numerical methods BESP-3D and BESP-1D for computing the ground states and TSSP-3D and TSSP-1D for computing the dynamics can be extended straightforward.

Remark 3.2. *For the semirelativistic Hartree system considered in [9], i.e.*

$$i\varepsilon\partial_t\psi_j^\varepsilon(\mathbf{x},t) = \left[\sqrt{-\varepsilon^2\Delta + 1} + V_{\text{ext}}^\varepsilon(\mathbf{x}) + V_P^\varepsilon\right]\psi_j^\varepsilon, \quad \mathbf{x}\in\mathbb{R}^3,\ t>0,\ |j|\leq M, \tag{3.68}$$

$$-\Delta V_P^\varepsilon = \rho^\varepsilon := \sum_{j=-M}^{M}|\psi_j|^2, \quad \mathbf{x}\in\mathbb{R}^3, \quad \lim_{|\mathbf{x}|\to\infty} V_P^\varepsilon(\mathbf{x},t) = 0, \quad t\geq 0; \tag{3.69}$$

with $\varepsilon > 0$ a scaled Planck constant and $M \geq 0$ a non-negative integer, the numerical methods TSSP-3D and TSSP-1D for computing the dynamics can be extended straightforward.

Table 3.1: Spatial discretization error analysis of BESP-3D, BESP-1D and BEFD-1D for computing ground states of relativistic Hartree.

	$h = 2$	$h = 4/3$	$h = 1$	$h = 2/3$	$h = 1/2$
BESP-3D	1.3254E-2	9.3079E-5	1.2608E-6	1.4965E-9	<E-9
BESP-1D	3.2523E-2	3.4154E-4	8.9687E-6	5.7715E-9	<E-9
	$h = 1/2$	$h = 1/4$	$h = 1/8$	$h = 1/16$	$h = 1/32$
BEFD-1D	1.0394E-2	2.4597E-3	6.0795E-4	1.5157E-4	3.7867E-5

3.5 Numerical results

In this section, we first test the accuracy of methods BESP-3D, BESP-1D and BEFD-1D for computing the ground states, and TSSP-3D, TSSP-1D and TSFD-1D for computing the dynamics of the RSP system. Then we apply them to simulate the ground states and dynamics in different parameter regimes and external potentials, as well as with finite time blow-up. For simplification, we always choose $h_x = h_y = h_z := h$ in 3D in the computation.

3.5.1 Accuracy test

First, we test the spatial discretization errors of BESP-3D, BESP-1D and BEFD-1D methods for computing the ground states. In order to do so, we take $\beta = -16$, $m = 1$, $V_{\text{ext}}(\mathbf{x}) \equiv 0$ for $\mathbf{x} \in \mathbb{R}^3$ in (3.4). In computation, we choose $\tau = 0.01$, initial data $\phi_0(\mathbf{x}) = (\pi/2)^{-3/4} \mathrm{e}^{-(x^2+y^2+z^2)}$ in (3.15), $\Omega = [-16, 16]^3$ with $J = K = L$ (or $h_x = h_y = h_z = h$) for the 3D case; and respectively, $\varphi_0(r) = 2\sqrt{\pi} r (\pi/2)^{-3/4} \mathrm{e}^{-r^2}$ in (3.57), $R = 16$ for the 1D case. The ground state ϕ_g is reached when $\|\phi^n - \phi^{n+1}\|_\infty < 10^{-9}$. The "exact" ground state ϕ_g^e is obtained under a very fine mesh. Let ϕ_g^h be the numerical ground state under the mesh size h. Tab. 3.1 lists the errors $\left\|\phi_g^h - \phi_g^e\right\|_\infty$ by using BESP-3D, BESP-1D and BEFD-1D with different mesh sizes h.

Table 3.2: Spatial discretization error analysis of TSSP-3D, TSSP-1D and TSFD-1D for computing dynamics of relativistic Hartree.

	$h=1$	$h=2/3$	$h=1/2$	$h=1/3$	$h=1/4$
TSSP-3D	2.7987E-2	6.6190E-3	4.0541E-6	6.7901E-7	7.6630E-9
TSSP-1D	8.9639E-3	5.9967E-4	6.5654E-5	1.0935E-7	6.8056E-10
	$h=1/4$	$h=1/8$	$h=1/16$	$h=1/32$	$h=1/64$
TSFD-1D	1.1365E-2	3.3655E-3	8.7813E-4	2.2189E-4	5.5622E-5

Table 3.3: Temporal discretization error analysis of TSSP-3D, TSSP-1D and TSFD-1D for computing dynamics of relativistic Hartree.

	$\tau=0.2$	$\tau=0.1$	$\tau=0.05$	$\tau=0.025$
TSSP-3D	2.3918E-4	5.9753E-5	1.4892E-5	3.7201E-6
TSSP-1D	1.7504E-4	4.3414E-5	1.0832E-5	2.5067E-6
TSFD-1D	1.8826E-4	4.6948E-5	1.1975E-5	3.2543E-6

Then we test the spatial and temporal discretization errors of TSSP-3D, TSSP-1D and TSFD-1D methods for computing the dynamics. Again, we take $\beta = -16$, $m = 1$, $V_{\text{ext}}(\mathbf{x}) \equiv 0$ for $\mathbf{x} \in \mathbb{R}^3$ in (3.4), and the initial data $\psi_0(\mathbf{x}) = (\pi/2)^{-3/4}\mathrm{e}^{-(x^2+y^2+z^2)}$ in (1.9) and $\mathcal{U}_0(r) = 2\sqrt{\pi}r(\pi/2)^{-3/4}\mathrm{e}^{-r^2}$ in (3.49). In computation, we take $\Omega = [-6,6]^3$ with $J = K = L$ (or $h_x = h_y = h_z = h$) for the 3D case; and respectively, $R = 6$ for the 1D case. The "exact" solution ψ^e is obtained under a very fine mesh and small time step. Let $\psi^{h,\tau}$ be the numerical solution under the mesh size h and time step τ. Tab. 3.2 gives the errors $\left\|\psi^{h,\tau} - \psi^e\right\|_\infty$ at time $t = 1$ under $\tau = 10^{-5}$ by using TSSP-3D, TSSP-1D and TSFD-1D with different mesh sizes h, which demonstrates spatial discretization errors; and Tab.

Table 3.4: Various quantities in the ground states when $\beta = -10$ and $V_{\text{ext}}(\mathbf{x}) \equiv 0$ with different m for case (i) in Example 3.1.

m	E_g	E_g^{kin}	E_g^{inp}	μ_g	δ_r
1	0.9769	1.0380	-0.0611	0.9157	9.9553
2	1.9413	2.0761	-0.1347	1.8066	2.4889
3	2.9265	3.1141	-0.1876	2.7389	1.1062
4	3.9075	4.1521	-0.2446	3.6630	0.6222
5	4.8886	5.1902	-0.3016	4.5870	0.3982
6	5.8663	6.2282	-0.3619	5.5044	0.2765

3.3 shows similar results under $h = 1/8$ for TSSP-3D, TSSP-1D, and respectively, $h = 1/512$ for TSFD-1D with different time steps τ, which demonstrates temporal discretization errors.

From Tabs. 3.1, 3.2 and 3.3, we can draw the following conclusions: (i) both BESP-3D and BESP-1D are spectrally accurate and BEFD-1D is second-order accurate in spatial discretization for computing the ground states; (ii) both TSSP-3D and TSSP-1D are spectrally accurate and TSFD-1D is second-order accurate in spatial discretization for computing the dynamics, and all these three methods are second-order accurate in temporal discretization. Based on these observations, for computing ground states of the RSP equation, if the potential V_{ext} is spherically symmetric BESP-1D is suggested, otherwise, BESP-3D should be used; and for computing the dynamics, if the potential V and initial data ψ_0 are both spherically symmetric TSSP-1D is suggested, otherwise, TSSP-3D should be used.

3.5.2 Ground states of the RSP equation

To quantify the ground state $\phi_g(\mathbf{x})$, we will examine its total energy $E_g := E(\phi_g)$, chemical potential $\mu_g := \mu(\phi_g)$, kinetic energy $E_g^{\text{kin}} := E^{\text{kin}}(\phi_g)$, external potential

Table 3.5: Various quantities in the ground states when $m = 1$ and $V_{\text{ext}}(\mathbf{x}) \equiv 0$ with different $\beta < 0$ for case (ii) in Example 3.1.

β	E_g	E_g^{kin}	E_g^{inp}	μ_g	δ_r
-16	0.9434	1.1153	-0.1718	0.7716	3.1277
-14	0.9588	1.0825	-0.1237	0.8351	4.4562
-12	0.9679	1.0573	-0.0894	0.8785	6.5188
-10	0.9769	1.0380	-0.0611	0.9157	9.9553
-8	0.9842	1.0235	-0.0393	0.9449	16.3002
-6	0.9925	1.0128	-0.0204	0.9721	30.0289

Table 3.6: Various quantities in the ground states when $m = 1$ and $V_{\text{ext}}(\mathbf{x}) = V_{\text{ext}}(r) = \frac{1}{2}r^2$ with different $\beta > 0$ for case (iii) in Example 3.1.

β	E_g	E_g^{kin}	E_g^{inp}	E_g^{exp}	μ_g	δ_r
16	2.7164	1.5673	0.4408	0.7083	3.1572	0.4722
32	3.1292	1.4899	0.7764	0.8629	3.9055	0.5752
64	3.8349	1.4016	1.2947	1.1387	5.1296	0.7591
128	4.9784	1.3176	2.0488	1.6120	7.0271	1.0747
256	6.7429	1.2479	3.0960	2.3989	9.8390	1.5993
512	9.3476	1.1934	4.4745	3.6798	13.8221	2.4532

energy $E_g^{\text{exp}} := E^{\text{exp}}(\phi_g)$ and internal potential energy $E_g^{\text{inp}} := E^{\text{inp}}(\phi_g)$ as well as its mean width square δ_r defined as

$$\delta_r = \frac{1}{3}\int_{\mathbb{R}^3} |\mathbf{x}|^2 |\phi_g(\mathbf{x})|^2 \, d\mathbf{x} = \frac{1}{3}\int_{\mathbb{R}^3} (x^2 + y^2 + z^2)|\phi_g(\mathbf{x})|^2 \, d\mathbf{x},$$

which can be computed numerically in 3D as

$$\delta_r \approx \frac{h_x h_y h_z}{3} \sum_{(j,k,l)\in\mathcal{T}_{JKL}} \left(x_j^2 + y_k^2 + z_l^2\right) |(\phi_g)_{jkl}|^2 ,$$

and respectively, if ϕ_g is spherically symmetric in 1D as

$$\delta_r \approx \frac{4\pi h_r}{3} \sum_{j=1}^{J-1} r_j^4 \, |(\phi_g)_j|^2.$$

Example 3.1. Ground states of the RSP equation with spherically symmetric potential for different parameters m and β. We consider three cases: (i) $\beta = -10$ and $V_{\text{ext}}(\mathbf{x}) \equiv 0$ with different m; (ii) $m = 1$ and $V_{\text{ext}}(\mathbf{x}) \equiv 0$ with different $\beta < 0$; and (iii) $m = 1$ and a harmonic trapping potential $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + z^2) = \frac{1}{2}r^2$ with different $\beta \geq 0$. The problem is always computed on a sufficiently large bounded domain $\Omega = [0, R]$ by using BESP-1D with 257 grid points and time step $\tau = 0.01$. The initial data φ_0 is taken as $\varphi_0 = \frac{2\sqrt{\pi}r}{(\pi/2)^{3/4}}\mathrm{e}^{-r^2}$.

Tabs. 3.4, 3.5 and 3.6 show various quantities in the ground states in cases (i), (ii) and (iii), respectively, including total energy, kinetic energy, internal and external potential energy, chemical potential μ_g and mean width square δ_r. Fig. 3.1 depicts the plots of the ground state solution $\phi_g(r)$ in cases (i), (ii) and (iii) as well as the energy evolution while solving the gradient flow in case (i). In addition, from the results in cases (i) and (ii), we can numerically predict the “Chandrasekhar limit mass”, λ_{cr}. For each fixed $m > 0$, we can numerically fit a curve of δ_r versus $\beta < 0$, and then λ_{cr} is numerically obtained by finding the zero point of the fitting function. Fig. 3.2 shows the fitting curves of δ_r versus $\beta < 0$ when $m = 2$, 3 and 4; and the ground states $\phi_g(r)$ when $m = 4$ for $\beta = -32, -32.5, -33, -33.5$. From these numerical results, it is numerically found that $\beta_{\text{cr}} = -4\pi\lambda_{\text{cr}} \approx -33.8$, i.e. $\lambda_{\text{cr}} \approx 2.69$, which is independent of m.

Based on Tabs. 3.4, 3.5 and 3.6, and Figs. 3.1 and 3.2, one can conclude, for a large system with attractive self-interaction (i.e. $\lambda < 0$ in (1.8) or $\beta < 0$ in (3.4)) and without external potential, that: (i) as the particle mass m increases but for a fixed β in (3.4), the total energy, kinetic energy in ground states and the chemical potential increase, but the internal potential energy (negative) decreases. Also, as m increases, the attractive interaction becomes stronger. (ii) for fixed m, as $|\beta|$ increases in (3.4), the total energy, internal potential energy (negative) in the ground

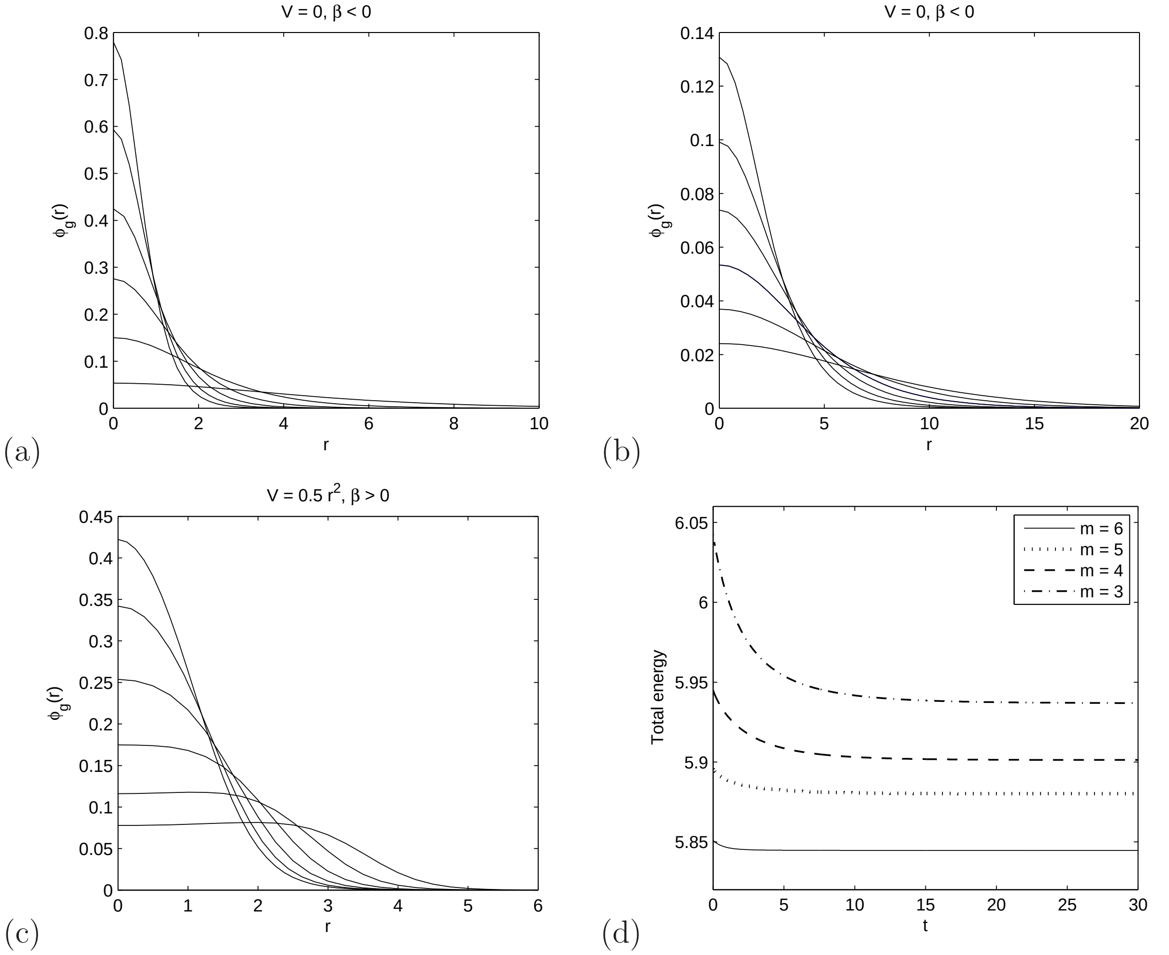


Figure 3.1: Ground states $\phi_g(r)$ in Example 3.1: (a) for case (i) with $m = 1, 2, \dots, 6$ (as peak increasing); (b) for case (ii) with $\beta = -6, -8, \dots, -16$ (as peak increasing); (c) for case (iii) with $\beta = 2^4, 2^5, \dots, 2^9$ (as peak decreasing); and (d) time evolution of energy in case (i).

states and chemical potential decrease, but the kinetic energy increases. Again, as $|\beta|$ increases, the attractive interaction becomes stronger, which also indicates that when the total mass exceeds certain critical value, the "gravitational collapse" of boson stars would occur. On the other hand, in a large system with repulsive self-interaction (i.e. $\lambda > 0$ in (1.8) or $\beta > 0$ in (3.4)) with a harmonic potential, for the fixed particle mass m as the total number of particle increases (i.e. β increases in (3.4)), the total energy, both internal and external potential energy in the ground

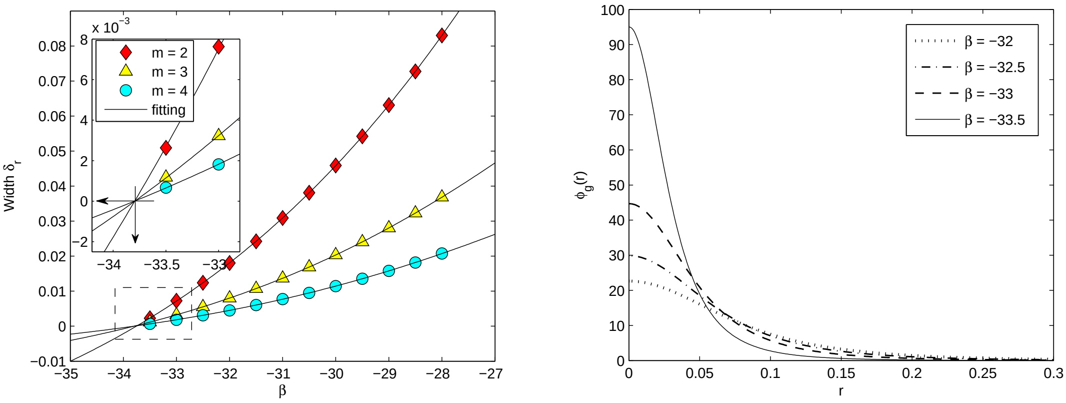


Figure 3.2: Numerical study of the "Chandrasekhar limit mass", i.e., $\lambda_{\text{cr}} = -\beta_{\text{cr}}/4\pi \approx 33.8/4\pi \approx 2.69$ in Example 3.1: fitting curves of δ_r versus $\beta < 0$ for $m = 2$, 3 and 4 (left column); and ground states $\phi_g(r)$ when $m = 4$ for $\beta = -32$, -32.5, -33, -33.5 (right column).

states, and the chemical potential increase, while the kinetic energy decreases. Also, in this case the repulsive interaction becomes stronger as β increases.

Example 3.2. Ground states of the RSP equation with different non-spherically symmetric potentials in (3.4). We consider three cases: (i) $\beta = -10$ and $m = 1$ with a harmonic potential $V_{\text{ext}}(x, y, z) = \frac{1}{32}(16x^2 + y^2 + z^2)$; (ii) $\beta = -10$ and $m = 1$ with a double-well potential $V_{\text{ext}}(x, y, z) = \frac{1}{32}((4 - x^2)^2 + y^2 + z^2)$; and (iii) $\beta = 64$ and $m = 1$ with an optical lattice potential $V_{\text{ext}}(x, y, z) = \frac{1}{2}(x^2 + y^2 + z^2) + 10\left(\sin^2(\pi x) + \sin^2(\pi y) + \sin^2(\pi z)\right)$.

The problem is computed on a bounded domain $\Omega = [-8, 8]^3$ by using BESP-3D with mesh size $h_r = 1/8$ and time step $\Delta t = 0.01$. The initial data is taken as $\phi_0(x, y, z) = (\pi/2)^{-3/4} e^{-(x^2+y^2+z^2)}$. Fig. 3.3 shows the surface plots of $\phi_g(x, y, 0)$ and isosurface plots of $|\phi_g| = 0.1$ for the above three cases. The results show that the BESP-3D method can compute the ground states very efficiently and accurately.

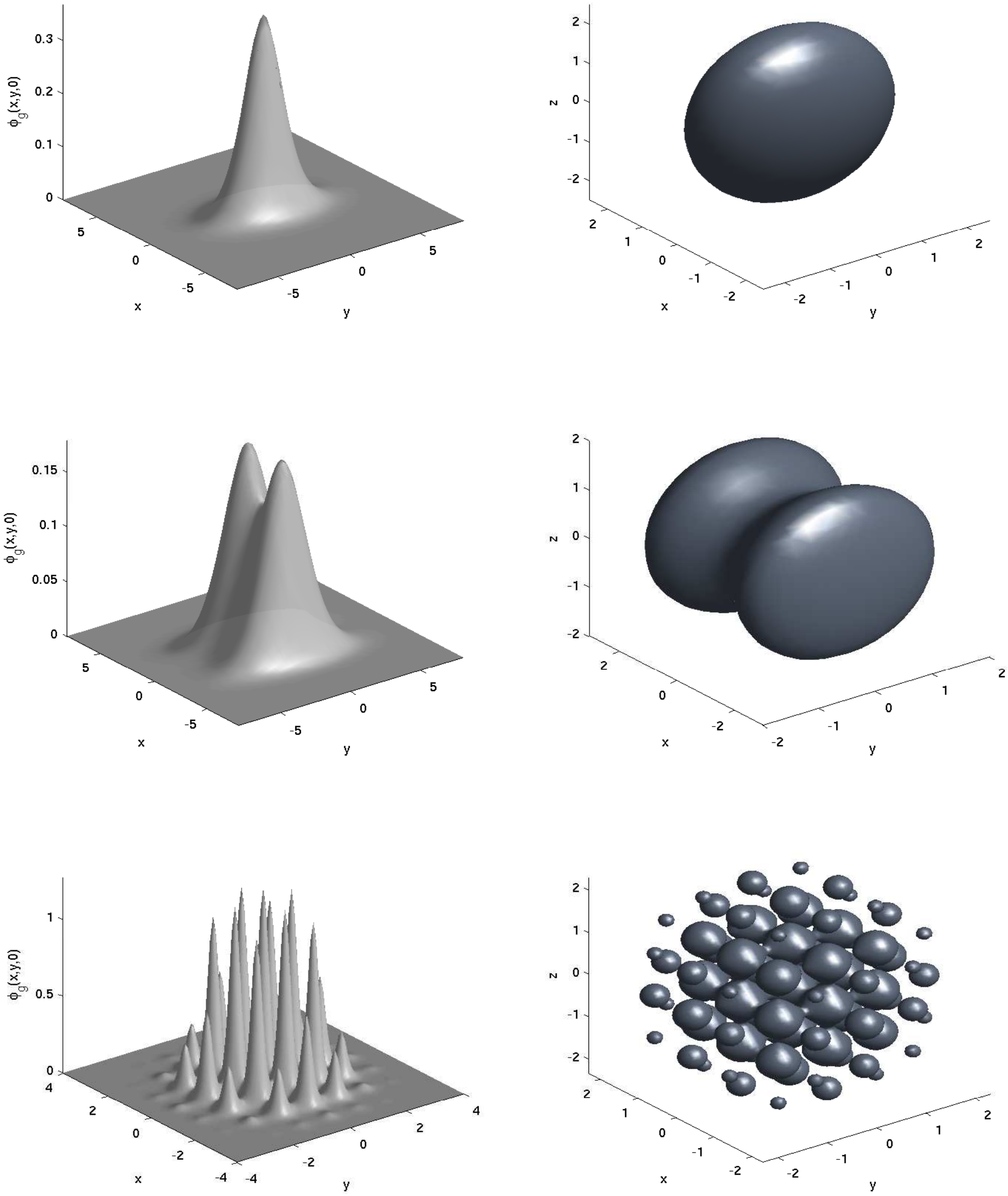


Figure 3.3: Ground state solution ϕ_g in Example 3.2 for case (i) (top row), case (ii) (middle row) and case (iii) (bottom row): surface plots of $\phi_g(x, y, 0)$ (left column); and isosurface plots of $|\phi_g| = 0.1$ (right column).

3.5.3 Dynamics of the RSP equation

Example 3.3. Dynamics of ground states under perturbation, i.e. we take initial condition as the ground state computed numerically by using the BESP-3D method. First, we study the evolution of the ground state under the potential $V_{\rm ext} = \frac{1}{2}(x^2 + y^2 + z^2)$ for $\beta = -1$ and $m = 1$, when the potential suddenly changes to $V_{\rm ext} = \frac{1}{2}(4x^2 + y^2 + z^2)$. We choose $\Omega = [-4, 4]^3$ with mesh size $h = 1/8$ and time step $\tau = 0.001$. Second, we look at the evolution of the ground state under a double-well potential $V_{\rm ext}(x, y, z) = \frac{1}{32}\left((4 - x^2)^2 + y^2 + z^2\right)$ for $\beta = -10$ and $m = 1$, when the potential suddenly changes to $V_{\rm ext} = \frac{1}{2}(x^2 + y^2 + z^2)$. In this case, we choose $\Omega = [-8, 8]^3$ with mesh size $h = 1/4$ and time step $\Delta t = 0.001$. Figs. 3.4 and 3.5 show the evolution of total energy, kinetic energy and external/internal potential energy, the evolutions of $\psi(x, 0, 0, t)$, and isosurface plots of $|\psi| = 0.1$ at different time points for these two cases. In these two cases, the existence of global-in-time solution is observed. Also, the method conserves the total energy very well.

Next, we study the dynamics of the center of mass. Let ϕ_g be the ground state under the potential $V_{\rm ext}(x, y, z) = \frac{1}{2}(x^2 + y^2 + z^2)$ with $\beta = -1$ and $m = 1$, which is obtained numerically by the BESP-3D method on $[-4, 4]^3$ with mesh size $h = 1/8$. The initial condition is taken as

$$\psi_0(x, y, z) = \phi_g(x, y, z)\mathrm{e}^{i\,(0.8x+0.5y+0.3z)},$$

and we apply the TSSP-3D method with mesh size $h = 1/4$ and time step $\tau = 0.001$. The center of mass, $(x_{\rm com}, y_{\rm com}, z_{\rm com})$, is evaluated by

$$x_{\rm com} = h_x h_y h_z \sum_{(i,j,k)\in\mathcal{T}_{JKL}} x_j |\psi^n_{jkl}|^2,$$

and similar for $y_{\rm com}$ and $z_{\rm com}$.

Fig. 3.6 shows the evolution of each component of the center of mass, various energy as well as the isosurface plots of $|\psi| = 0.1$ at different time points. An obvious damping phenomena in the center of mass is observed, and the damping frequencies in each component of the center of mass are identical even though the damping amplitudes differ.

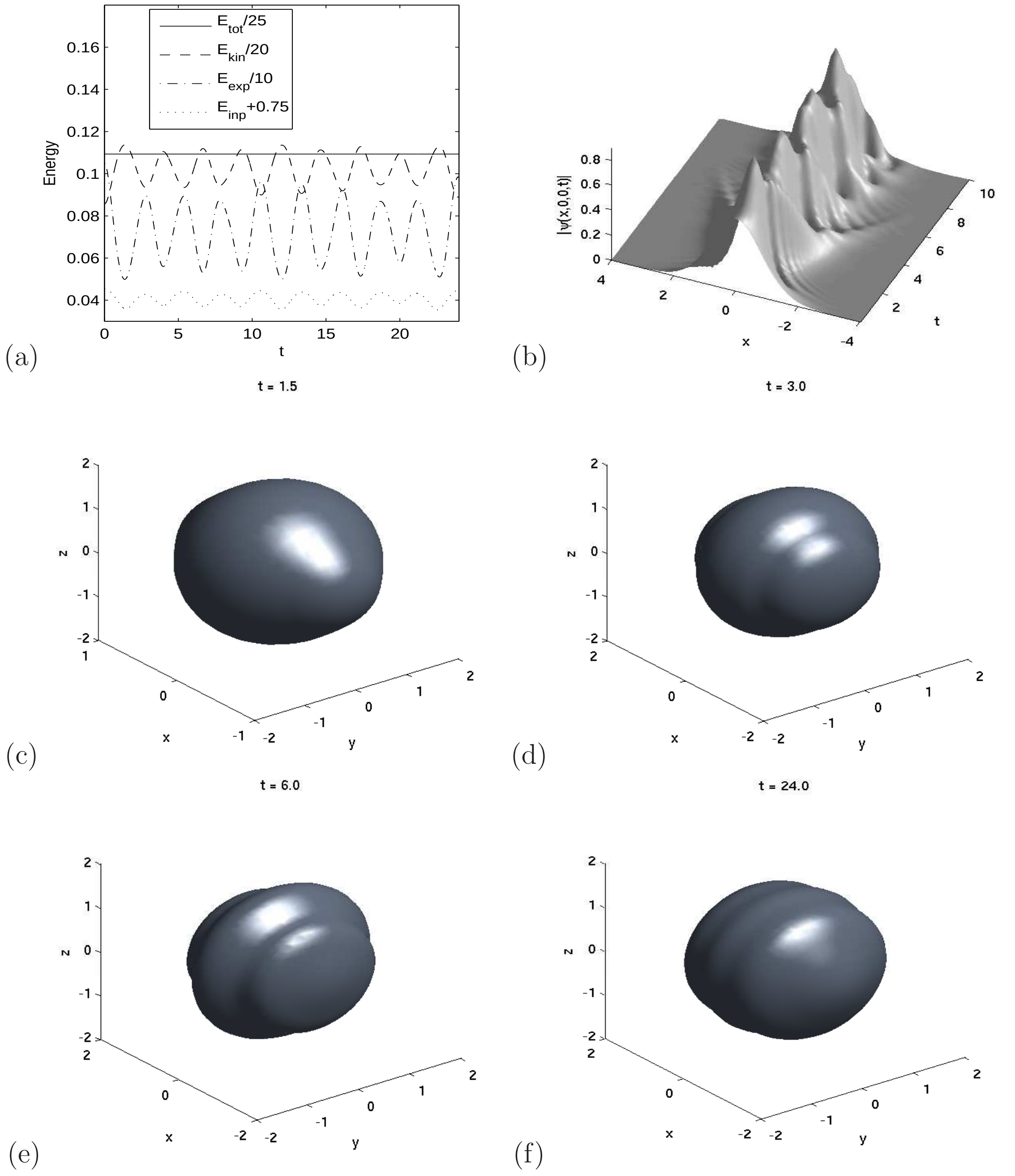


Figure 3.4: Dynamics of the ground state when potential changes instantly from $V_{\text{ext}} = \frac{1}{2}(x^2 + y^2 + z^2)$ to $V_{\text{ext}} = \frac{1}{2}(4x^2 + y^2 + z^2)$, for $\beta = -1$ and $m = 1$ in Example 3.3: (a) evolution of various energies; (b) evolution of $|\psi(x, 0, 0, t)|$; (c)-(f) isosurface plots of $|\psi| = 0.1$ at different times.

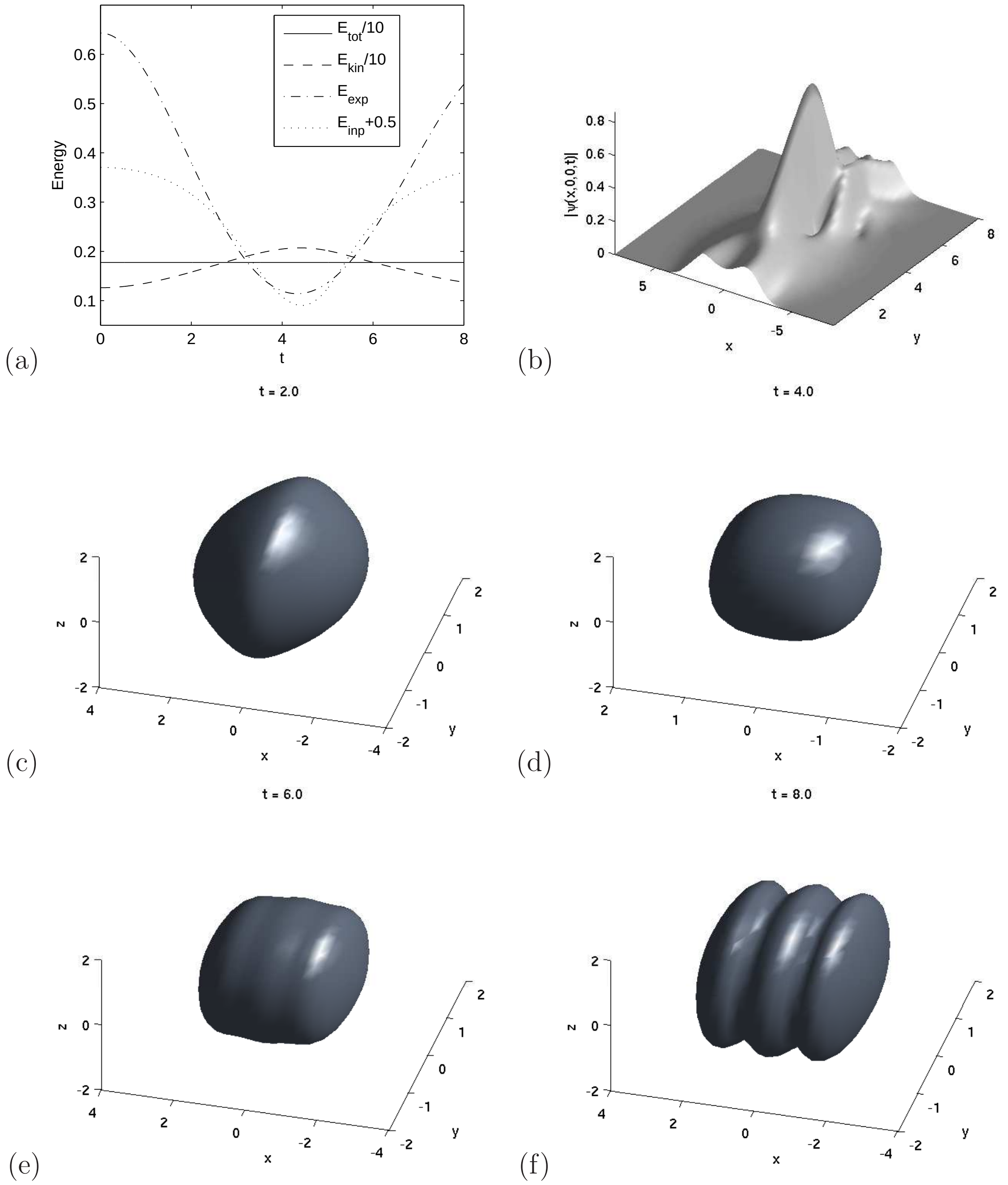


Figure 3.5: Dynamics of the ground state when potential changes instantly from $V_{\text{ext}} = \frac{1}{32}\left((4-x^2)^2 + y^2 + z^2\right)$ to $V_{\text{ext}} = \frac{1}{32}\left(4x^2 + y^2 + z^2\right)$, for $\beta = -10$ and $m = 1$ in Example 3.3: (a) evolution of various energies; (b) evolution of $|\psi(x,0,0,t)|$; (c)-(f) isosurface plots of $|\psi| = 0.1$ at different times.

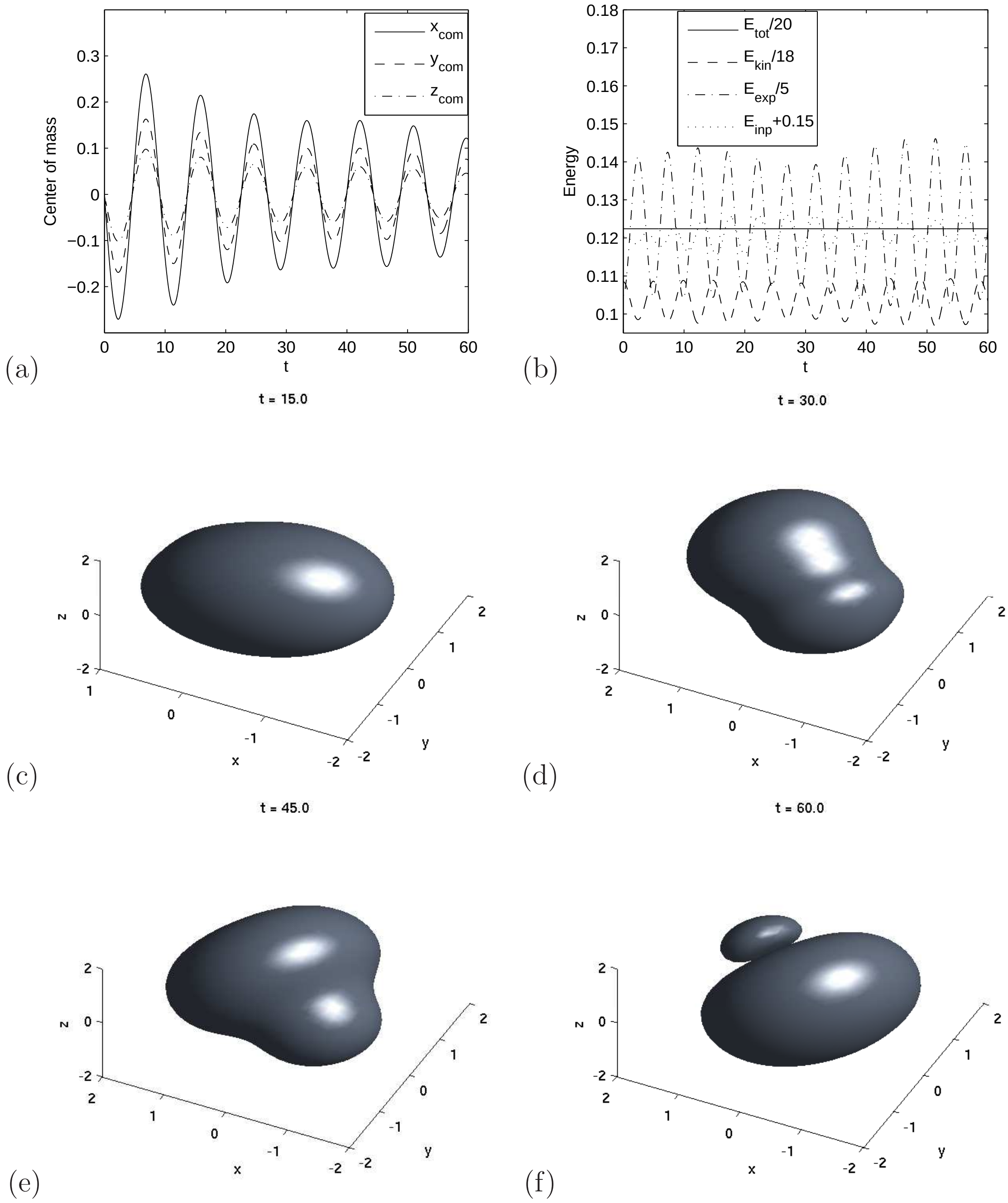


Figure 3.6: Dynamics of the ground state enforced an instant movement in Example 3.3: (a) evolution of the center of mass $(x_{\rm com}, y_{\rm com}, z_{\rm com})$; (b) evolution of various energies; (c)-(f) isosurface plots of $|\psi| = 0.1$ at different times. Here, $V_{\rm ext} = \frac{1}{2}\,(x^2 + y^2 + z^2)$, $m = 1$ and $\beta = -1$.

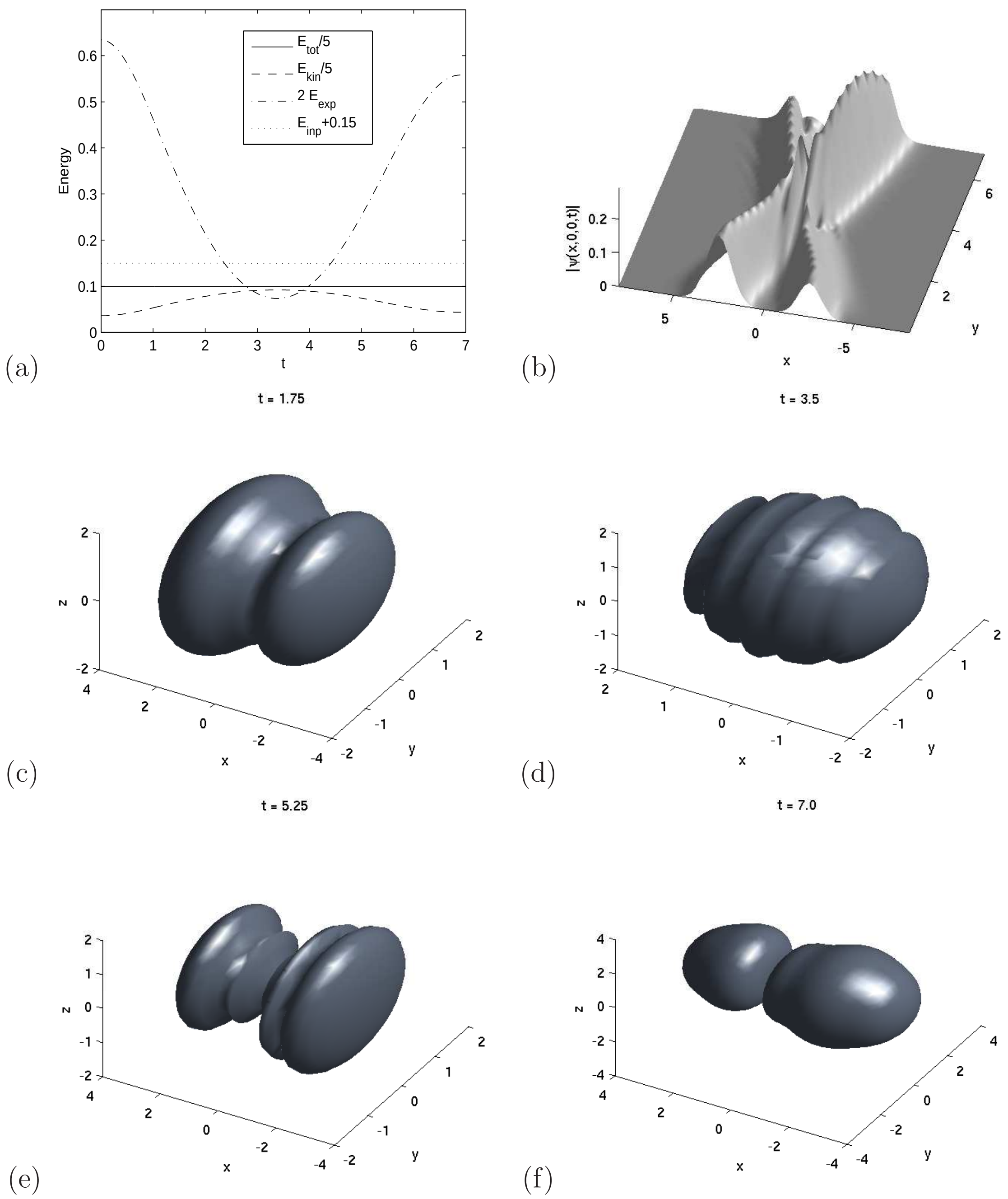


Figure 3.7: Results in Example 3.4. Dynamics of two Gaussian beams with opposite moving directions: (a) evolution of various energies; (b) evolution of $|\psi(x, 0, 0, t)|$; (c)-(f) isosurface plots of $|\psi| = 0.05$ at different times.

Example 3.4. Wave-collision in the RSP equation, i.e. we take the initial condition as

$$\psi_0 = \frac{1}{3(\pi/2)^{3/4}}\,\mathrm{e}^{-(y^2+z^2)}\left(\mathrm{e}^{i\,0.8x-(x+2.5)^2} + 2\mathrm{e}^{-i\,0.5x-(x-2.5)^2}\right),$$

which is two Gaussian beams in x-axis with opposite moving directions, $V_{\mathrm{ext}} = \frac{1}{2}(x^2+y^2+z^2)$, $\beta = -1$ and $m = 1$. We apply the TSSP-3D method by choosing $\Omega = [-8,8]^3$ with mesh size $h = 1/4$ and time step $\tau = 0.001$. Fig. 3.7 plots the evolution of various energies, the evolution of $|\psi(x,0,0,t)|$ and isosurface plots of $|\psi| = 0.05$ at different time points. It shows that after a collision of two Gaussian beams, which may have different amplitudes and opposite moving directions with various velocities, there is no significant new wave structure generated.

Example 3.5. Finite time blow-up in the RSP equation, i.e. we investigate the change of the "gravitational collapse" time with respect to the particle mass as well as the total number of particles in boson stars without external potentials. The initial condition is taken as

$$\psi_0(r) = \frac{1}{(\pi/50)^{3/4}}e^{-25r^2},$$

and the TSSP-1D method is applied with $\Omega = [0,1]$, $h_r = 1/256$ and $\tau = 0.0001$. The blow-up time is detected by looking at the evolution of the kinetic energy. First we fix the particle mass as $m = 1$ and change β from -50 to -200, and then choose $m = 1$, 40, 60 and 80, when $\beta = -50$. Fig. 3.8 shows the evolution of kinetic energy in these two settings, and depicts the evolution of $|\psi(r,t)|$ when $(\beta, m) = (-200, 1)$ and $(\beta, m) = (-50, 80)$. The results indicate a monotonic relation between the "gravitational collapse" time and both the particle mass and total particle number. More precisely, when either the total particle number increases or the particle mass decreases, the boson stars would collapse earlier.

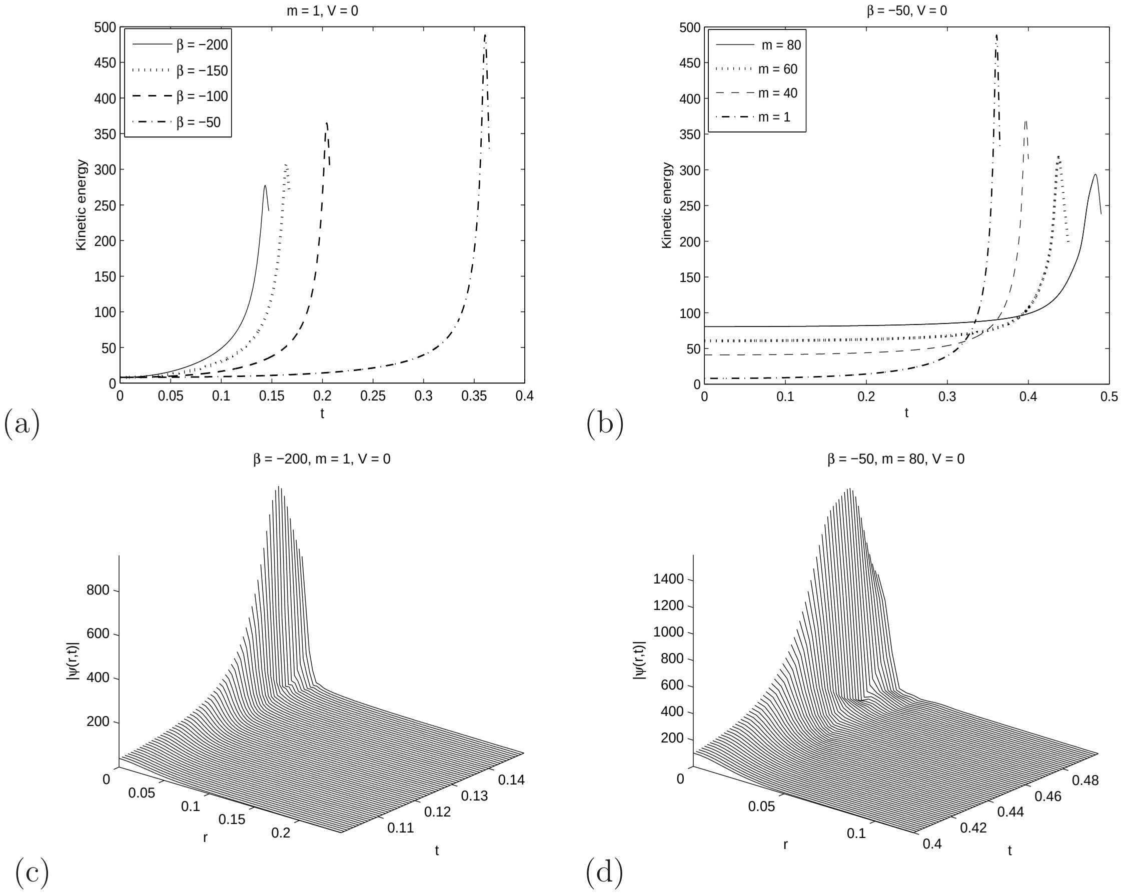


Figure 3.8: Time evolution of kinetic energy in the blow-up cases when $V_{\rm ext} = 0$ in Example 3.5: (a) for $\beta < 0$ and $m = 1$, and (b) for $\beta = -50$ and different m; and evolution of $|\psi(r,t)|$ close to the blow-up when $V_{\rm ext}(r) = 0$: (c) for $\beta = -200$ and $m = 1$, and (d) for $\beta = -50$ and $m = 80$.

Chapter 4

Methods and analysis for the Klein–Gordon equation

This chapter investigates the performance of various numerical methods for solving the Klein–Gordon equation (1.12)–(1.13) in the nonrelativistic limit regime, i.e. $0 < \varepsilon \ll 1$. The methods studied here include frequently-used finite difference time domain (FDTD) discretizations and the Gautschi-type exponential wave integrator combined with spectral or finite difference discretization in space. For all the methods considered here, rigorous error estimates are carried out with particular attention on how their optimal error bounds depend explicitly on the small parameter ε.

4.1 Introduction

As introduced in Section 1.2, the dimensionless relativistic Klein–Gordon (KG) equation in d-dimensions ($d = 1,\ 2,\ 3$) [106, 107, 110] is considered here,

$$\varepsilon^2 \partial_{tt} u - \Delta u + \frac{1}{\varepsilon^2} u + f(u) = 0, \quad \mathbf{x} \in \mathbb{R}^d, \quad t > 0, \tag{4.1}$$

with initial conditions given as

$$u(\mathbf{x}, 0) = \phi(\mathbf{x}), \quad \partial_t u(\mathbf{x}, 0) = \frac{1}{\varepsilon^2} \gamma(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^d. \tag{4.2}$$

Here $u = u(\mathbf{x}, t)$ is a real-valued field, $\varepsilon > 0$ is a dimensionless parameter which is inversely proportional to the speed of light [106, 107, 110], ϕ and γ are given real-valued functions, $f(u)$ is a dimensionless real-valued function independent of ε and satisfies $f(0) = 0$. In practice, the typical nonlinearity is the pure power case, i.e. $f(u) = \lambda u^{p+1}$ with $p \geq 0$ and $\lambda \in \mathbb{R}$ [35, 36, 64, 65, 68, 69, 106, 107, 110, 115, 123, 131, 134, 140]. In fact, the above KG equation is also known as the relativistic version of the Schrödinger equation under proper non-dimensionalization (cf. Section 1.2 and [106,107,110]) and it is used to describe the motion of a spinless particle [46,128]. The KG equation (4.1)–(4.2) is time symmetric or time reversible. In addition, if $u(\cdot, t) \in H^1(\mathbb{R}^d)$ and $\partial_t u(\cdot, t) \in L^2(\mathbb{R}^d)$, it also conserves the *energy* [106, 107, 110], i.e.,

$$\begin{aligned} E(t) &:= \int_{\mathbb{R}^d} \left[\varepsilon^2 \left(\partial_t u(\mathbf{x}, t)\right)^2 + |\nabla u(\mathbf{x}, t)|^2 + \frac{1}{\varepsilon^2} u^2(\mathbf{x}, t) + F\left(u(\mathbf{x}, t)\right) \right] d\mathbf{x} \\ &\equiv \int_{\mathbb{R}^d} \left[\frac{1}{\varepsilon^2} \gamma^2(\mathbf{x}) + |\nabla \phi(\mathbf{x})|^2 + \frac{1}{\varepsilon^2} \phi^2(\mathbf{x}) + F\left(\phi(\mathbf{x})\right) \right] d\mathbf{x} := E(0), \quad t \geq 0, \end{aligned} \tag{4.3}$$

where

$$F(u) = 2 \int_0^u f(s)\, ds, \quad u \in \mathbb{R}. \tag{4.4}$$

For fixed $\varepsilon > 0$ ($O(1)$-speed of light regime), e.g. $\varepsilon = 1$, the KG equation (4.1)–(4.2) has gained a surge of attention in both analytical and numerical aspects. Along the analytical front, the Cauchy problem was investigated, e.g. in [11, 36, 64, 68, 87, 94, 131, 134]. In particular, for the defocusing case (i.e. $F(u) \geq 0$ for $u \in \mathbb{R}$) the global existence of solutions was established in [36], and for the focusing case (i.e. $F(u) \leq 0$ for $u \in \mathbb{R}$) possible finite time blow-up was shown in [11]. For more results in this regime, one can refer to [4,35,115,118,123,132,140] and references therein. In the numerical aspect, various numerical schemes were proposed and studied in the literature. For instance, standard finite difference time domain (FDTD) methods such as energy conservative, semi-implicit and explicit finite difference discretizations were proposed and analyzed in [3, 52, 98, 121, 141]. Other approaches, like finite element or spectral discretization, were also studied in [37, 47, 146]. Comparisons of different methods in this regime were carried out in [88, 121].

However, in the nonrelativistic limit regime, i.e. if $0 < \varepsilon \ll 1$ or the speed of light goes to infinity, the analysis and efficient computation of the KG equation (4.1)–(4.2) are mathematically rather complicated issues. The difficulty in analysis is mainly due to that the energy $E(t)$ in (4.3) becomes unbounded when $\varepsilon \to 0$. Recently, Machihara et al. [107] studied such limit in the energy space, and Masmoudi et al. [110] analyzed such limit in a strong topology of the energy space. For more recent progresses made on this topic, one can refer to [116, 117, 147]. Their results show that the solution propagates waves with wavelength $O(\varepsilon^2)$ and $O(1)$ in time and space, respectively, when $0 < \varepsilon \ll 1$. On the other hand, this highly oscillatory nature in time provides severe numerical burdens, making the computation in the nonrelativistic limit regime extremely challenging. Tracing to the literature, so far there are few results on the numerics of the KG equation in this regime.

The aim of this chapter is to study the efficiency of frequently used FDTD methods applied in the nonrelativistic limit regime, to propose new numerical schemes and to compare their resolution capacities in this regime. In the following sections, we begin with the detailed analysis on the stability and convergence of four standard implicit/semi-implicit/explicit energy conservative or non-conservative FDTD methods. Here, particular attention is paid on how the error bounds depend explicitly on the small parameter ε in addition to the mesh size h and time step τ. Based on the estimates, in order to obtain "correct" numerical approximations when $0 < \varepsilon \ll 1$, the meshing strategy requirement (ε-scalability) for those frequently used FDTD methods is

$$\tau = O(\varepsilon^3), \quad h = O(1), \tag{4.5}$$

which suggests that the standard FDTD methods are computationally expensive for the KG equation (4.1)–(4.2) as $0 < \varepsilon \ll 1$. To relax the ε-scalability, we then propose new numerical methods, whose ε-scalability is optimal for both time and space in view of the inherent oscillatory nature. The key ideas of the new schemes are: (i) to apply either sine pseudospectral or centered finite difference discretization for spatial derivatives; and (ii) to discretize the highly oscillatory second-order ordinary differential equations (ODEs) in phase space by using the Gautschi-type exponential wave integrator [63, 77] which is well demonstrated in the literature that it has favorable properties compared with standard time integrators for oscillatory second-order differential equations [72,73,83,84]. For the linear KG equation, the Gautschi-type time integrator does not introduce any time discretization error. Rigorous error estimates show that the ε-scalability of the new methods is improved to

$$\tau = O(1), \quad h = O(1), \tag{4.6}$$

for the linear KG equation, and respectively, to

$$\tau = O(\varepsilon^2), \quad h = O(1), \tag{4.7}$$

for the nonliear KG equation. Thus, the Gautschi-type methods offer compelling advantages over commonly used FDTD methods in temporal resolution when $0 < \varepsilon \ll 1$.

4.2 FDTD methods and their analysis

In this section, commonly used FDTD methods are applied to the KG equation (4.1)–(4.2) [52, 88, 98, 121, 141], and their stability and convergence in the nonrelativistic limit regime are rigorously analyzed. For simplicity of notations, the numerical methods and their analysis shall be only presented in 1D. Generalization to higher dimensions is straightforward and results remain valid without modifications. Similar to most works in the literature for the analysis and computation of the KG equation (cf. [3, 37, 47, 52, 88, 98, 121, 141, 146] and references therein), in practical computation, the whole space problem is truncated into an interval $\Omega = (a, b)$ with homogeneous Dirichlet boundary conditions. In 1D, the KG equation (4.1)–(4.2) with homogeneous Dirichlet boundary conditions collapses to

$$\varepsilon^2\partial_{tt}u(x,t) - \partial_{xx}u + \frac{1}{\varepsilon^2}u + f(u) = 0, \quad x \in \Omega = (a,b), \quad t > 0, \tag{4.8}$$

$$u(a,t) = u(b,t) = 0, \quad t \geq 0, \tag{4.9}$$

$$u(x,0) = \phi(x), \quad \partial_t u(x,0) = \frac{1}{\varepsilon^2}\gamma(x), \quad x \in \bar{\Omega} = [a,b], \tag{4.10}$$

with $\phi(a) = \phi(b) = 0$ and $\gamma(a) = \gamma(b) = 0$.

4.2.1 FDTD methods

Choose mesh size $h := \Delta x = (b-a)/M$ with M being an even positive integer, time step $\tau := \Delta t > 0$ and denote grid points and time steps as

$$x_j := a + jh, \quad j = 0,1,\ldots,M, \quad t_n := n\tau, \quad n = 0,1,2,\ldots.$$

Let u_j^n be the approximation of $u(x_j,t_n)$ $(j = 0,1,\ldots,M, n = 0,1,\ldots)$ and introduce the finite difference discretization operators as

$$\delta_t^+ u_j^n = \frac{u_j^{n+1} - u_j^n}{\tau}, \quad \delta_t^- u_j^n = \frac{u_j^n - u_j^{n-1}}{\tau}, \quad \delta_t^2 u_j^n = \frac{u_j^{n+1} - 2u_j^n + u_j^{n-1}}{\tau^2},$$
$$\delta_x^+ u_j^n = \frac{u_{j+1}^n - u_j^n}{h}, \quad \delta_x^- u_j^n = \frac{u_j^n - u_{j-1}^n}{h}, \quad \delta_x^2 u_j^n = \frac{u_{j+1}^n - 2u_j^n + u_{j-1}^n}{h^2}.$$

It is easy to check that $\delta_t^2 = \delta_t^+\delta_t^- = \delta_t^-\delta_t^+$ and $\delta_x^2 = \delta_x^+\delta_x^- = \delta_x^-\delta_x^+$. Here, four frequently used FDTD methods [52,88,98,121,141] are considered to discretize the problem (4.8)–(4.10): for $j = 1,2,\ldots,M-1, \quad n = 1,2,\ldots,$

I. *Implicit energy conservative finite difference (Impt-EC-FD) method*

$$\varepsilon^2\delta_t^2 u_j^n - \frac{1}{2}\delta_x^2\left(u_j^{n+1} + u_j^{n-1}\right) + \frac{1}{2\varepsilon^2}\left(u_j^{n+1} + u_j^{n-1}\right) + G\left(u_j^{n+1}, u_j^{n-1}\right) = 0; \tag{4.11}$$

II. *Semi-implicit energy conservative finite difference (SImpt-EC-FD) method*

$$\varepsilon^2\delta_t^2 u_j^n - \delta_x^2 u_j^n + \frac{1}{2\varepsilon^2}\left(u_j^{n+1} + u_j^{n-1}\right) + G\left(u_j^{n+1}, u_j^{n-1}\right) = 0; \tag{4.12}$$

III. *Semi-implicit finite difference (SImpt-FD) method*

$$\varepsilon^2\delta_t^2 u_j^n - \frac{1}{2}\delta_x^2\left(u_j^{n+1} + u_j^{n-1}\right) + \frac{1}{2\varepsilon^2}\left(u_j^{n+1} + u_j^{n-1}\right) + f\left(u_j^n\right) = 0; \tag{4.13}$$

IV. *Explicit finite difference (Expt-FD) method*

$$\varepsilon^2\delta_t^2 u_j^n - \delta_x^2 u_j^n + \frac{1}{\varepsilon^2}u_j^n + f\left(u_j^n\right) = 0. \tag{4.14}$$

Here,

$$G(v,w) = \int_0^1 f\left(\theta v + (1-\theta)w\right)\,\mathrm{d}\theta = \frac{F\left(v\right) - F\left(w\right)}{2(v-w)}, \quad \forall\, v, w \in \mathbb{R}, \tag{4.15}$$

with $F(u)$ defined in (4.4). The initial and boundary conditions are discretized as

$$u_0^n = u_M^n = 0, \quad n \geq 0, \quad u_j^0 = \phi(x_j), \quad j = 0, 1, \ldots, M, \tag{4.16}$$

$$u_j^1 = \phi(x_j) + \frac{\tau}{\varepsilon^2}\gamma(x_j) + \frac{\tau^2}{2\varepsilon^2}\left[\delta_x^2\phi(x_j) - \frac{1}{\varepsilon^2}\phi(x_j) - f\left(\phi(x_j)\right)\right]. \tag{4.17}$$

Clearly, the above four FDTD methods are time symmetric or time reversible, i.e. they are unchanged if we interchange $n+1 \leftrightarrow n-1$ and $\tau \leftrightarrow -\tau$. Expt-FD is an explicit method, whereas Impt-EC-FD, SImpt-EC-FD and SImpt-FD are implicit methods. At each time step, SImpt-FD needs to solve a linear system, SImpt-EC-FD needs to solve a nonlinear decoupled system, and Impt-EC-FD needs to solve a fully nonlinear coupled system.

Denoting $X_M = \{v = (v_0, v_1, \ldots, v_M) \mid v_0 = v_M = 0\} \subset \mathbb{R}^{M+1}$, and letting $\left\{v_j^n,\ j = 0, 1, \ldots, M,\ n = 0, 1, \ldots\right\}$ be any grid function satisfying $v_0^n = v_M^n = 0$ $(n = 0, 1, \ldots)$, thus one has $v^n = (v_0^n, v_1^n, \ldots, v_M^n) \in X_M$ and can define its standard discrete l^2 norm, semi-H^1 norm, semi-H^2 norm and l^∞ norm as

$$\|v^n\|_{l^2}^2 = h\sum_{j=0}^{M-1}\left|v_j^n\right|^2, \quad \left\|\delta_x^+ v^n\right\|_{l^2}^2 = h\sum_{j=0}^{M-1}\left|\delta_x^+ v_j^n\right|^2, \tag{4.18}$$

$$\left\|\delta_x^2 v^n\right\|_{l^2}^2 = h\sum_{j=1}^{M-1}\left|\delta_x^2 v_j^n\right|^2, \quad \|v^n\|_{l^\infty} = \max_{0\leq j\leq M}\left|v_j^n\right|, \quad n \geq 0. \tag{4.19}$$

The results given in the following lemma are frequently used in the numerical analysis for finite difference schemes defined for $u^n \in X_M$.

Lemma 4.1. *For any $v^n \in X_M$ $(n \geq 0)$, the following equalities hold*

$$-h\sum_{j=0}^{M-1} v_j^n \delta_x^2 v_j^n = h\sum_{j=0}^{M-1} \left|\delta_x^+ v_j^n\right|^2 = \left\|\delta_x^+ v^n\right\|_{l^2}^2, \tag{4.20}$$

$$h\sum_{j=0}^{M-1} v_j^n v_j^{n+1} = \frac{1}{2}\left\|v^n\right\|_{l^2}^2 + \frac{1}{2}\left\|v^{n+1}\right\|_{l^2}^2 - \frac{\tau^2}{2}\left\|\delta_t^+ v^n\right\|_{l^2}^2, \tag{4.21}$$

$$\begin{aligned} h\sum_{j=0}^{M-1} \left(\delta_x^+ v_j^{n+1}\right)\left(\delta_x^+ v_j^n\right) = &\frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(v_{j+1}^{n+1} - v_j^n\right)^2 + \left(v_{j+1}^n - v_j^{n+1}\right)^2\right] \\ &-\frac{\tau^2}{h^2}\left\|\delta_t^+ v^n\right\|_{l^2}^2 \qquad n = 0, 1, \ldots. \end{aligned} \tag{4.22}$$

Proof. The equality (4.20) comes from the standard summation by parts formula (see, e.g. [98]) and (4.21) comes from

$$\begin{aligned} v_j^n v_j^{n+1} &= \frac{1}{2}\left[\left(v_j^{n+1}\right)^2 + \left(v_j^n\right)^2 - \left(v_j^{n+1} - v_j^n\right)^2\right] \\ &= \frac{1}{2}\left[\left(v_j^{n+1}\right)^2 + \left(v_j^n\right)^2 - \tau^2\left(\delta_t^+ v_j^n\right)^2\right]. \end{aligned}$$

From (4.21) and a straightforward computation, one gets

$$\begin{aligned} &h\sum_{j=0}^{M-1}\left(\delta_x^+ v_j^{n+1}\right)\left(\delta_x^+ v_j^n\right) = \frac{1}{2}\left\|\delta_x^+ v^{n+1}\right\|_{l^2}^2 + \frac{1}{2}\left\|\delta_x^+ v^n\right\|_{l^2}^2 - \frac{\tau^2}{2}\left\|\delta_t^+\delta_x^+ v^n\right\|_{l^2}^2 \\ &= \frac{h}{2}\sum_{j=0}^{M-1}\left[\left(\delta_x^+ v_j^{n+1}\right)^2 + \left(\delta_x^+ v_j^n\right)^2\right] - \frac{\tau^2}{2h}\sum_{j=0}^{M-1}\left(\delta_t^+ v_{j+1}^n - \delta_t^+ v_j^n\right)^2 \\ &= \frac{\tau^2}{h}\sum_{j=0}^{M-1}\left(\delta_t^+ v_{j+1}^n\right)\left(\delta_t^+ v_j^n\right) + \frac{h}{2}\sum_{j=0}^{M-1}\left[\left(\delta_x^+ v_j^{n+1}\right)^2 + \left(\delta_x^+ v_j^n\right)^2\right] - \frac{\tau^2}{h^2}\left\|\delta_t^+ v^n\right\|_{l^2}^2 \\ &= \frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(v_{j+1}^{n+1} - v_j^{n+1}\right)^2 + \left(v_{j+1}^n - v_j^n\right)^2 + 2\left(v_{j+1}^{n+1} - v_{j+1}^n\right)\left(v_j^{n+1} - v_j^n\right)\right] \\ &\quad -\frac{\tau^2}{h^2}\left\|\delta_t^+ v^n\right\|_{l^2}^2 \\ &= \frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(v_{j+1}^{n+1} - v_j^n\right)^2 + \left(v_{j+1}^n - v_j^{n+1}\right)^2\right] - \frac{\tau^2}{h^2}\left\|\delta_t^+ v^n\right\|_{l^2}^2, \end{aligned}$$

which immediately implies (4.22). □

For the first two methods Impt-EC-FD and SImpt-EC-FD, one can easily show that they conserve the energy in the discretized level, i.e.

Lemma 4.2. *The method Impt-EC-FD (4.11) conserves the discrete energy as*

$$\begin{aligned} E^n = {}& \varepsilon^2 \left\|\delta_t^+ u^n\right\|_{l^2}^2 + \frac{1}{2}\left(\left\|\delta_x^+ u^n\right\|_{l^2}^2 + \left\|\delta_x^+ u^{n+1}\right\|_{l^2}^2\right) + \frac{1}{2\varepsilon^2}\left(\left\|u^n\right\|_{l^2}^2 + \left\|u^{n+1}\right\|_{l^2}^2\right) \\ & + \frac{h}{2}\sum_{j=0}^{M-1}\left[F\left(u_j^n\right) + F\left(u_j^{n+1}\right)\right] \equiv E^0, \quad n = 0, 1, 2, \dots . \end{aligned} \tag{4.23}$$

Similarly, the method SImpt-EC-FD (4.12) conserves the discrete energy as

$$\begin{aligned} \widetilde{E}^n = {}& \varepsilon^2 \left\|\delta_t^+ u^n\right\|_{l^2}^2 + h\sum_{j=0}^{M-1}\delta_x^+ u_j^n \cdot \delta_x^+ u_j^{n+1} + \frac{1}{2\varepsilon^2}\left(\left\|u^n\right\|_{l^2}^2 + \left\|u^{n+1}\right\|_{l^2}^2\right) \\ & + \frac{h}{2}\sum_{j=0}^{M-1}\left[F\left(u_j^n\right) + F\left(u_j^{n+1}\right)\right] \equiv \widetilde{E}^0, \quad n = 0, 1, 2, \dots . \end{aligned} \tag{4.24}$$

Proof. The proof proceeds in the analogous lines as in [98, 141] for the standard KG equation, i.e. $\varepsilon = 1$ in (4.8)–(4.10), and the details are omitted here for brevity. □

4.2.2 Stability analysis

By using the standard von Neumann analysis [136], the following stability results for the FDTD methods can be obtained,

Theorem 4.1. *Suppose $f(u)$ is linear, i.e. $f(u) = \alpha u$ with α a constant satisfying $\alpha > -\varepsilon^{-2}$, then,*

(i) The method Impt-EC-FD (4.11) is unconditionally stable for any $\tau > 0$, $h > 0$ and $\varepsilon > 0$.

(ii) When $4\varepsilon^2 - h^2(1 + \varepsilon^2\alpha) \leq 0$, the method SImpt-EC-FD (4.12) is unconditionally stable for any $\tau > 0$ and $h > 0$; and when $4\varepsilon^2 - h^2(1 + \varepsilon^2\alpha) > 0$, it is conditionally stable under the stability condition

$$\tau \leq \frac{2h\varepsilon^2}{\sqrt{4\varepsilon^2 - h^2(1 + \varepsilon^2\alpha)}}. \tag{4.25}$$

(iii) When $-\varepsilon^{-2} < \alpha \leq \varepsilon^{-2}$, the method SImpt-FD (4.13) is unconditionally stable for any $\tau > 0$ and $h > 0$; and when $\alpha > \varepsilon^{-2}$, it is conditionally stable under the stability condition

$$\tau \leq \frac{2\varepsilon^2}{\sqrt{\varepsilon^2\alpha - 1}}. \tag{4.26}$$

(iv) The method Expt-FD (4.13) is conditionally stable under the stability condition

$$\tau \le \frac{2h\varepsilon^2}{\sqrt{4\varepsilon^2 + h^2(1+\alpha\varepsilon^2)}}. \tag{4.27}$$

Proof. Noticing $f(u) = \alpha u$, plugging

$$u_j^{n-1} = \mathrm{e}^{2ijl\pi/M}, \quad u_j^n = \xi_l \mathrm{e}^{2ijl\pi/M}, \quad u_j^{n+1} = \xi_l^2 \mathrm{e}^{2ijl\pi/M},$$

into (4.11)–(4.14), with ξ_l the amplification factor of the l-th mode in phase space, one can obtain the characteristic equation with the following structure

$$\xi_l^2 - 2\theta_l\,\xi_l + 1 = 0, \tag{4.28}$$

where $\theta_l \in \mathbb{R}$ is determined by the corresponding method and may vary for different methods. Solving the above equation, one has $\xi_l = \theta_l \pm \sqrt{\theta_l^2 - 1}$. The stability of numerical schemes amounts to

$$|\xi_l| \le 1 \quad \Longleftrightarrow \quad |\theta_l| \le 1. \tag{4.29}$$

(i) For the method Impt-EC-FD (4.11), noticing $\alpha > -\varepsilon^{-2}$, one has

$$0 \le \theta_l = \frac{2\varepsilon^4}{2\varepsilon^4 + \tau^2\left(\varepsilon^2\lambda_l^2 + \varepsilon^2\alpha + 1\right)} \le 1, \tag{4.30}$$

with

$$\lambda_l = \frac{2}{h}\sin\left(\frac{l\pi}{M}\right). \tag{4.31}$$

This implies that the method Impt-EC-FD (4.11) is unconditionally stable for any $\tau > 0$, $h > 0$ and $\varepsilon > 0$.

(ii) For the method SImpt-FD (4.13), one has

$$\theta_l = \frac{2\varepsilon^4 - \tau^2\varepsilon^2\lambda_l^2}{2\varepsilon^4 + \tau^2\left(\varepsilon^2\alpha + 1\right)}. \tag{4.32}$$

From (4.31),

$$0 \le \lambda_l^2 \le \frac{4}{h^2}. \tag{4.33}$$

Thus, when $4\varepsilon^2 - h^2(1+\varepsilon^2\alpha) \le 0$, or $4\varepsilon^2 - h^2(1+\varepsilon^2\alpha) > 0$ with the condition (4.25),

$$\left(\varepsilon^2\lambda_l^2 - \varepsilon^2\alpha - 1\right)\tau^2 \le \left(\frac{4\varepsilon^2}{h^2} - \varepsilon^2\alpha - 1\right)\tau^2 \le 4\varepsilon^4 \Longrightarrow |\theta_l| \le 1.$$

(iii) For the method SImpt-EC-FD (4.12), one has

$$\theta_l = \frac{2\varepsilon^4 - \tau^2\varepsilon^2\alpha}{2\varepsilon^4 + \tau^2\left(\varepsilon^2\lambda_l^2 + 1\right)}. \tag{4.34}$$

Noticing (4.33), when $-\varepsilon^{-2} < \alpha \le \varepsilon^{-2}$, or $\alpha > \varepsilon^{-2}$ with the condition (4.26),

$$\tau^2\left(\varepsilon^2\alpha - 1 - \varepsilon^2\lambda_l^2\right) \le \tau^2(\varepsilon^2\alpha - 1) \le 4\varepsilon^4 \Longrightarrow |\theta_l| \le 1.$$

(iv) For the method Expt-FD (4.14), one has

$$\theta_l = \frac{2\varepsilon^4 - \tau^2\left(\varepsilon^2\lambda_l^2 + \varepsilon^2\alpha + 1\right)}{2\varepsilon^4}. \tag{4.35}$$

Combining (4.33) and (4.27), one gets

$$\tau^2\left(\varepsilon^2\lambda_l^2 + 1 + \varepsilon^2\alpha\right) \le \tau^2\left(\frac{4\varepsilon^2}{h^2} + 1 + \varepsilon^2\alpha\right) \le 4\varepsilon^4 \Longrightarrow |\theta_l| \le 1.$$

The proof is completed. □ □

4.2.3 Main results on error estimates

Motivated by the analytical results in [107,110] for the KG equation, the following assumptions on the exact solution u of (4.8)–(4.10) are made

$$\text{(A)} \qquad u \in C^4([0,T];W^{1,\infty}) \cap C^3([0,T];W^{2,\infty}) \cap C^2([0,T];W^{3,\infty}) \cap C([0,T];W^{5,\infty}\cap H_0^1),$$

$$\left\|\frac{\partial^{r+s}}{\partial t^r \partial x^s}u(x,t)\right\|_{L^\infty(\Omega_T)} \lesssim \frac{1}{\varepsilon^{2r}}, \quad 0 \le r \le 4 \ \&\ 0 \le r+s \le 5,$$

where, $\Omega_T = \Omega \times [0,T]$ and $0 < T < T^*$ with T^* the maximum existence time of the solution. Define the grid "error" function $e^n \in X_M$ $(n \ge 0)$ as

$$e_j^n = u(x_j,t_n) - u_j^n, \quad j = 0,1,\ldots,M, \quad n = 0,1,2,\ldots, \tag{4.36}$$

with u_j^n the approximations obtained from FDTD methods.

For the method Impt-EC-FD (4.11), one can establish the following error estimate (see detailed proof in the forthcoming subsection):

Theorem 4.2. *Assume $\tau \lesssim \varepsilon^3$ and under assumptions (A) and $f \in C^3(\mathbb{R})$, there exist constants $\tau_0 > 0$ and $h_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \leq 1$, when $0 < \tau \leq \tau_0$ and $0 < h \leq h_0$, the following error estimate for the method Impt-EC-FD (4.11) with (4.16) and (4.17) holds,*

$$\|e^n\|_{l^2} + \|\delta_x^+ e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \leq n \leq \frac{T}{\tau}. \tag{4.37}$$

For Expt-FD method, one can have the following error estimate (see detailed proof in the forthcoming subsection):

Theorem 4.3. *Assume $\tau \lesssim \varepsilon^3$ and under assumptions (A) and $f \in C^2(\mathbb{R})$, there exist constants $\tau_0 > 0$ and $h_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \leq 1$, when $0 < \tau \leq \tau_0$ and $0 < h \leq h_0$ satisfying $\tau \leq \varepsilon h/2$, the following error estimate for the method Expt-FD (4.14) with (4.16) and (4.17) holds,*

$$\|e^n\|_{l^2} + \|\delta_x^+ e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \leq n \leq \frac{T}{\tau}. \tag{4.38}$$

Similarly, for the methods SImpt-EC-FD (4.12) and SImpt-FD (4.13), one can have,

Theorem 4.4. *Assume $\tau \lesssim \varepsilon^3$ and under assumptions (A) and $f \in C^3(\mathbb{R})$, there exist constants $\tau_0 > 0$ and $h_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \leq 1$, when $0 < \tau \leq \tau_0$ and $0 < h \leq h_0$ satisfying $\tau \leq \varepsilon h/\sqrt{2}$, the following error estimate for the method SImpt-EC-FD (4.12) with (4.16) and (4.17) holds,*

$$\|e^n\|_{l^2} + \|\delta_x^+ e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \leq n \leq \frac{T}{\tau}. \tag{4.39}$$

Theorem 4.5. *Assume $\tau \lesssim \varepsilon^3$ and under assumptions (A) and $f \in C^2(\mathbb{R})$, there exist constants $\tau_0 > 0$ and $h_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \le 1$, when $0 < \tau \le \tau_0$ and $0 < h \le h_0$, the following error estimate for the method SImpt-FD (4.13) with (4.16) and (4.17) holds,*

$$\|e^n\|_{l^2} + \|\delta_x^+ e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \le n \le \frac{T}{\tau}. \tag{4.40}$$

Based on Theorems 4.2, 4.3, 4.4 and 4.5, the four FDTD methods studied here exhibit the same temporal/spatial resolution capacity in the nonrelativistic limit regime. In fact, given an accuracy bound $\delta > 0$, the ε-scalability of four FDTD methods is

$$\tau = O\left(\varepsilon^3\sqrt{\delta}\right) = O(\varepsilon^3), \quad h = O\left(\sqrt{\delta}\right) = O(1), \quad 0 < \varepsilon \ll 1. \tag{4.41}$$

Remark 4.1. *The same kind of error bounds in 2D and 3D can be achieved by replacing the assumption $\tau \lesssim \varepsilon^3$ in Theorems 4.2, 4.3, 4.4 and 4.5 by $\tau \lesssim \varepsilon^3\sqrt{C_d(h)}$, with the use of the following discrete Sobolev inequality (inverse inequality) [15, 145],*

$$\|u^n\|_{l^\infty} \lesssim \frac{1}{C_d(h)}\left[\|\delta_x^+ u^n\|_{l^2} + \|u^n\|_{l^2}\right], \quad C_d(h) = \begin{cases} 1, & d = 1, \\ 1/|\ln h|, & d = 2, \\ h^{1/2}, & d = 3. \end{cases} \tag{4.42}$$

4.2.4 Proof of Theorem 4.2

Lemma 4.3. *Denote the local truncation error $\xi^n \in X_M$ for Impt-EC-FD (4.11) as*

$$\begin{aligned} \xi_j^0 &:= \delta_t^+ u(x_j, 0) - \frac{1}{\varepsilon^2}\gamma(x_j) - \frac{\tau}{2\varepsilon^2}\left[\delta_x^2\phi(x_j) - \frac{1}{\varepsilon^2}\phi(x_j) - f(\phi(x_j))\right], \\ \xi_j^n &:= \varepsilon^2\delta_t^2\left(u(x_j, t_n)\right) - \frac{1}{2}\left[\delta_x^2\left(u(x_j, t_{n+1})\right) + \delta_x^2\left(u(x_j, t_{n-1})\right)\right] \\ &\quad + \frac{1}{2\varepsilon^2}\left[u\left(x_j, t_{n+1}\right) + u\left(x_j, t_{n-1}\right)\right] + G\left(u\left(x_j, t_{n+1}\right), u\left(x_j, t_{n-1}\right)\right), \; n \ge 1, \end{aligned} \tag{4.43}$$

for $j = 1, 2, \ldots, M-1$, and $\xi_0^n = \xi_M^n = 0$ $(n = 0, 1, \ldots)$. Under the assumptions (A) and $f \in C^3(\mathbb{R})$,

$$\|\xi^n\|_{l^2} + \left\|\delta_x^+\xi^n\right\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \le n \le \frac{T}{\tau}, \quad \left\|\delta_x^2\xi^0\right\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}. \tag{4.44}$$

Proof. Taking Taylor's expansion in the local truncation error (4.43), noticing (4.15), (4.8)–(4.10), using the assumptions (A) and $f \in C^3(\mathbb{R})$, with the help of the triangle inequality and Cauchy-Schwartz inequality,

$$\left|\xi_j^0\right| \le \frac{\tau^2}{6}\left\|\partial_{ttt}u\right\|_{L^\infty(\Omega_T)} + \frac{h\tau}{6\varepsilon^2}\left\|\phi'''\right\|_{L^\infty(\Omega)} \lesssim \frac{\tau^2}{\varepsilon^6} + \frac{h\tau}{\varepsilon^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \tag{4.45}$$

for all $j = 0, 1, \ldots, M-1$ and therefore the first assertion in (4.44) is proved for $n = 0$. Also, for $j = 1, 2, \ldots, M-2$,

$$\left|\delta_x^+\xi_j^0\right| \le \frac{\tau^2}{6}\left\|\partial_{tttx}u\right\|_{L^\infty(\Omega_T)} + \frac{h\tau}{6\varepsilon^2}\left\|\phi''''\right\|_{L^\infty(\Omega)} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}. \tag{4.46}$$

For $j = 0$ and $M-1$, from the homogeneous boundary conditions one can deduce that $\partial_t^l u(x,t)\big|_{\partial\Omega} = 0$, $l \ge 0$, and the equation (4.8)–(4.10) itself indicates that $\partial_{xx}u(x,t)|_{\partial\Omega} = 0$, $\partial_{ttxx}u(x,t)|_{\partial\Omega} = 0$ and $\partial_{xxxx}u(x,t)|_{\partial\Omega} = 0$. Then one can get the same estimate as (4.46) for $j = 0$ and $M-1$. Similarly, for $j = 1, 2, \ldots, M-2$,

$$\left|\delta_x^2\xi_j^0\right| \le \frac{\tau^2}{6}\left\|\partial_{tttxx}u\right\|_{L^\infty(\Omega_T)} + \frac{h\tau}{6\varepsilon^2}\left\|\phi'''''\right\|_{L^\infty(\Omega)} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \tag{4.47}$$

$$\begin{aligned}
\left|\xi_j^n\right| \le{}& \frac{\varepsilon^2\tau^2}{12}\left\|\partial_{tttt}u\right\|_{L^\infty(\Omega_T)} + \frac{\tau^2}{2}\left\|\partial_{ttxx}u\right\|_{L^\infty(\Omega_T)} + \frac{h^2}{12}\left\|\partial_{xxxx}u\right\|_{L^\infty(\Omega_T)} \\
&+ \tau^2\left[\left\|f'\right\|_{L^\infty(\mathbb{R})}\left\|\partial_{tt}u\right\|_{L^\infty(\Omega_T)} + \left\|f''\right\|_{L^\infty(\mathbb{R})}\left\|\partial_t u\right\|^2_{L^\infty(\Omega_T)} + \frac{1}{2\varepsilon^2}\left\|\partial_{tt}u\right\|_{L^\infty(\Omega_T)}\right] \\
\lesssim{}& h^2 + \frac{\tau^2}{\varepsilon^6} + \frac{\tau^2}{\varepsilon^4} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 1 \le n \le \frac{T}{\tau},
\end{aligned} \tag{4.48}$$

$$\begin{aligned}
\left|\delta_x^+\xi_j^n\right| \le{}& \frac{\varepsilon^2\tau^2}{12}\left\|\partial_{ttttx}u\right\|_{L^\infty(\Omega_T)} + \frac{\tau^2}{2}\left\|\partial_{ttxxx}u\right\|_{L^\infty(\Omega_T)} + \frac{h^2}{12}\left\|\partial_{xxxxx}u\right\|_{L^\infty(\Omega_T)} \\
&+ \tau^2\Big[\left\|f''\right\|_{L^\infty(\mathbb{R})}\left\|\partial_{tt}u\right\|_{L^\infty(\Omega_T)}\left\|\partial_x u\right\|_{L^\infty(\Omega_T)} + \left\|f'\right\|_{L^\infty(\mathbb{R})}\left\|\partial_{ttx}u\right\|_{L^\infty(\Omega_T)} \\
&+ \left\|f'''\right\|_{L^\infty(\mathbb{R})}\left\|\partial_t u\right\|^2_{L^\infty(\Omega_T)}\left\|\partial_x u\right\|_{L^\infty(\Omega_T)} + \frac{1}{2\varepsilon^2}\left\|\partial_{ttx}u\right\|_{L^\infty(\Omega_T)} \\
&+ \left\|f''\right\|_{L^\infty(\mathbb{R})}\left\|\partial_t u\right\|_{L^\infty(\Omega_T)}\left\|\partial_{tx}u\right\|_{L^\infty(\Omega_T)}\Big] \\
\lesssim{}& h^2 + \frac{\tau^2}{\varepsilon^6} + \frac{\tau^2}{\varepsilon^4} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 1 \le n \le \frac{T}{\tau}.
\end{aligned} \tag{4.49}$$

These (the boundary cases are similar to above) immediately imply the estimates in (4.44). □

Lemma 4.4. *There exist $h_0 > 0$ and $\tau_0 > 0$ sufficiently small, under the assumption $f \in C^3(\mathbb{R})$ and when $0 < \tau \leq \tau_0$ and $0 < h \leq h_0$, there exists a unique solution u_j^n ($j = 0, 1, \ldots, M$, $n \geq 0$) of the problem (4.11) with (4.16) and (4.17).*

Proof. The argument follows the analogous lines as in [52,141] for the standard KG equation, i.e. $\varepsilon = 1$ in (4.8)–(4.10), and the details are omitted here for brevity. □

It is expected that the main difficulty in the rest arguments is the l^∞ control of the finite difference solutions. Traditional approaches to overcome such difficulty or to achieve the desired control rely on stronger assumptions on the nonlinear term, i.e. much stronger than merely to assume it is continuous, as well as the conservative property of the scheme. Here, instead of using those traditional approaches, the nonlinear term f is truncated to a global Lipschitz function with compact support by using a cut-off technique (cf. [15]). This is guaranteed provided that the continuous solutions are bounded and the finite difference solutions are closed to the continuous solutions. Noting the regularity assumption (A),

$$K_0 = \|u(x,t)\|_{L^\infty(\Omega_T)},$$

is well-defined. Choose a smooth function $\chi(s) \in C_0^\infty(\mathbb{R})$ such that

$$\chi(s) = \begin{cases} 1, & 0 \leq |s| \leq 1, \\ \in [0,1], & 1 \leq |s| \leq 2, \\ 0, & |s| \geq 2. \end{cases} \tag{4.50}$$

Denote $B = K_0 + 1$ and for $s \in \mathbb{R}$,

$$f_B(s) = f(s)\chi(s/B), \tag{4.51}$$

then it is readily to verify that f_B is global Lipschitz. Denote

$$F_B(v) = 2\int_0^v f_B(s)\mathrm{d}s,$$
$$G_B(v,w) = \int_0^1 f_B\left(\theta v + (1-\theta)w\right)\,\mathrm{d}\theta = \frac{F_B(v) - F_B(w)}{2(v-w)}, \quad \forall\, v, w \in \mathbb{R},$$

then one can have, $f_B(u(x,t)) = f(u(x,t))$ and $G_B(u(x,t), u(x,t')) = G(u(x,t), u(x,t'))$, for exact solution $u(x,t)$ of (4.8)–(4.10). Now, it is tempting to refer $u^n \in X_M$ $(n \geq 0)$ as solutions of the following scheme,

$$\varepsilon^2 \delta_t^2 u_j^n - \frac{1}{2}\delta_x^2 \left(u_j^{n+1} + u_j^{n-1}\right) + \frac{1}{2\varepsilon^2}\left(u_j^{n+1} + u_j^{n-1}\right) + G_B\left(u_j^{n+1}, u_j^{n-1}\right) = 0, \quad (4.52)$$

with initial and boundary conditions defined as (4.16) and (4.17). In fact, later one can show the scheme (4.52) and the original scheme (4.11) will coincide, provided that τ and h are small enough.

Lemma 4.5. *For $n \geq 1$, denote $\eta^n \in X_M$ with*

$$\eta_j^n = G_B\left(u(x_j, t_{n+1}), u(x_j, t_{n-1})\right) - G_B\left(u_j^{n+1}, u_j^{n-1}\right), \quad (4.53)$$

under the assumptions (A) and $f \in C^2(\mathbb{R})$, one has

$$\|\eta^n\|_{l^2}^2 \lesssim \left\|e^{n-1}\right\|_{l^2}^2 + \left\|e^{n+1}\right\|_{l^2}^2, \quad (4.54)$$

$$\left\|\delta_x^+ \eta^n\right\|_{l^2}^2 \lesssim \left\|e^{n-1}\right\|_{l^2}^2 + \left\|\delta_x^+ e^{n-1}\right\|_{l^2}^2 + \left\|e^{n+1}\right\|_{l^2}^2 + \left\|\delta_x^+ e^{n+1}\right\|_{l^2}^2. \quad (4.55)$$

Proof. From (4.53), noticing (4.15) and the assumption $f \in C^2(\mathbb{R})$ (which implies $f_B \in C_0^2(\mathbb{R})$),

$$\begin{aligned} \left|\eta_j^n\right| &= \left|\int_0^1 \left[f_B\left(\theta u(x_j, t_{n+1}) + (1-\theta)u(x_j, t_{n-1})\right) - f_B\left(\theta u_j^{n+1} + (1-\theta)u_j^{n-1}\right)\right] \mathrm{d}\theta\right| \\ &\leq \|f_B'\|_{L^\infty(\mathbb{R})} \int_0^1 \left[\theta\left|u(x_j, t_{n+1}) - u_j^{n+1}\right| + (1-\theta)\left|u(x_j, t_{n-1}) - u_j^{n-1}\right|\right] \mathrm{d}\theta \\ &\lesssim \left|e_j^{n-1}\right| + \left|e_j^{n+1}\right|, \quad j = 0, 1, \ldots, M-1, \quad n \geq 1. \end{aligned}$$

Using Hölder inequality, one gets (4.54) immediately. Similarly, for $j = 0, 1, \ldots, M-1$ and $n \geq 1$, one can obtain

$$|\delta_x^+ \eta_j^n| \lesssim |e_j^{n-1}| + |\delta_x^+ e_j^{n-1}| + |e_{j+1}^{n-1}| + |e_j^{n+1}| + |\delta_x^+ e_j^{n+1}| + |e_{j+1}^{n+1}|.$$

This, together with the Hölder inequality, implies (4.55) immediately. □

Lemma 4.6. *Under the assumptions (A) and $f \in C^3(\mathbb{R})$, the error bounds given in (4.37) hold for u^n obtained from (4.52) with (4.16) and (4.17).*

Proof. Subtracting (4.52) and (4.17) from (4.43), noticing (4.16) and (4.36), one can obtain that the error e_j^n satisfies

$$\begin{aligned}&\varepsilon^2\delta_t^2 e_j^n - \frac{1}{2}\left(\delta_x^2 e_j^{n+1} + \delta_x^2 e_j^{n-1}\right) + \frac{1}{2\varepsilon^2}\left(e_j^{n+1} + e_j^{n-1}\right)\\ &\qquad = \xi_j^n - \eta_j^n, \quad j = 1, 2, \ldots, M-1, \end{aligned} \tag{4.56}$$

$$e_j^0 = 0, \quad e_j^1 = \tau\xi_j^0, \quad e_0^n = e_M^n = 0, \quad n = 0, 1, \ldots, \quad j = 0, 1, \ldots, M. \tag{4.57}$$

Here, the local truncation error ξ_j^n is the same as that defined in Lemma 4.3 by noting that for exact solution $u(x_j, t_n)$, $f_B(u(x_j, t_n)) = f(u(x_j, t_n))$. Define the "energy" for the error vector $e^n \in X_M$ $(n = 0, 1, \ldots)$ as

$$\mathcal{E}^n = \varepsilon^2\left\|\delta_t^+ e^n\right\|_{l^2}^2 + \frac{1}{2}\left(\left\|\delta_x^+ e^n\right\|_{l^2}^2 + \left\|\delta_x^+ e^{n+1}\right\|_{l^2}^2\right) + \frac{1}{2\varepsilon^2}\left(\left\|e^n\right\|_{l^2}^2 + \left\|e^{n+1}\right\|_{l^2}^2\right). \tag{4.58}$$

Multiplying both sides of (4.56) by $h\left(e_j^{n+1} - e_j^{n-1}\right)$, then summing up for $j = 0, 1, \ldots, M-1$, noticing (4.20) and (4.58), one can get

$$\mathcal{E}^n - \mathcal{E}^{n-1} = h\sum_{j=0}^{M-1}\left(\xi_j^n - \eta_j^n\right)\left(e_j^{n+1} - e_j^{n-1}\right), \quad n \geq 1. \tag{4.59}$$

From (4.59), using Young's inequality, noticing Lemma 4.5,

$$\begin{aligned}\mathcal{E}^n - \mathcal{E}^{n-1} &\leq h\sum_{j=0}^{M-1}\left(\left|\xi_j^n\right| + \left|\eta_j^n\right|\right)\left|e_j^{n+1} - e_j^{n-1}\right|\\ &= \tau h\sum_{j=0}^{M-1}\left(\left|\xi_j^n\right| + \left|\eta_j^n\right|\right)\left|\delta_t^+ e_j^n + \delta_t^+ e_j^{n-1}\right|\\ &\leq \tau\left[\frac{1}{\varepsilon^2}\left(\|\xi^n\|_{l^2}^2 + \|\eta^n\|_{l^2}^2\right) + \varepsilon^2\left(\left\|\delta_t^+ e^n\right\|_{l^2}^2 + \left\|\delta_t^+ e^{n-1}\right\|_{l^2}^2\right)\right]\\ &\lesssim \tau\left(\mathcal{E}^n + \mathcal{E}^{n-1}\right) + \frac{\tau}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad n \geq 1. \end{aligned} \tag{4.60}$$

Thus, there exists a constant $\tau_0 > 0$ sufficiently small and independent of ε and h, such that when $0 < \tau \leq \tau_0$,

$$\mathcal{E}^n - \mathcal{E}^{n-1} \lesssim \tau\mathcal{E}^{n-1} + \frac{\tau}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad n \geq 1. \tag{4.61}$$

Summing the above inequality up for n,

$$\mathcal{E}^n - \mathcal{E}^0 \lesssim \tau\sum_{m=0}^{n-1}\mathcal{E}^m + \frac{T}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 1 \leq n \leq \frac{T}{\tau} - 1. \tag{4.62}$$

Using the discrete Gronwall's inequality [98, 120],

$$\mathcal{E}^n \lesssim \mathcal{E}^0 + \frac{T}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 1 \le n \le \frac{T}{\tau} - 1. \tag{4.63}$$

Combining (4.56)–(4.57), (4.58) for $n = 0$ and (4.44), one has

$$\begin{aligned} \mathcal{E}^0 &= \varepsilon^2\|\xi^0\|_{l^2}^2 + \frac{\tau^2}{2}\|\delta_x^+\xi^0\|_{l^2}^2 + \frac{\tau^2}{\varepsilon^2}\|\xi^0\|_{l^2}^2 \\ &\lesssim \left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2\left(\varepsilon^2 + \frac{\tau^2}{2} + \frac{\tau^2}{\varepsilon^2}\right) \lesssim \left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2\left(1 + \frac{\tau^2}{\varepsilon^2}\right). \end{aligned} \tag{4.64}$$

Plugging (4.64) into (4.63) leads to

$$\mathcal{E}^n \lesssim \frac{1}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 0 \le n \le \frac{T}{\tau} - 1. \tag{4.65}$$

Although from the above estimate one can achieve the semi-H^1 error estimate as well, it is not optimal. In order to get the optimal semi-H^1 error estimate, in addition, one can define another "energy" for the error vector $e^n \in X_M$ $(n = 0, 1, \ldots)$ as

$$\hat{\mathcal{E}}^n = \varepsilon^2\left\|\delta_x^+\delta_t^+e^n\right\|_{l^2}^2 + \frac{1}{2}\left(\left\|\delta_x^2e^n\right\|_{l^2}^2 + \left\|\delta_x^2e^{n+1}\right\|_{l^2}^2\right) + \frac{1}{2\varepsilon^2}\left(\left\|\delta_x^+e^n\right\|_{l^2}^2 + \left\|\delta_x^+e^{n+1}\right\|_{l^2}^2\right). \tag{4.66}$$

Multiplying both sides of (4.56) by $h\left(\delta_x^2e_j^{n+1} - \delta_x^2e_j^{n-1}\right)$, similar to the above procedure,

$$\hat{\mathcal{E}}^n \lesssim \frac{1}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 0 \le n \le \frac{T}{\tau} - 1. \tag{4.67}$$

Combining (4.65), (4.67), (4.58) and (4.66), noticing that $\|e^n\|_{l^2}^2 + \|e^{n+1}\|_{l^2}^2 \le 2\varepsilon^2\mathcal{E}^n$ and $\|\delta_x^+e^n\|_{l^2}^2 + \|\delta_x^+e^{n+1}\|_{l^2}^2 \le 2\varepsilon^2\hat{\mathcal{E}}^n$ when $0 < \varepsilon \le 1$, one can immediately obtain that error estimates in (4.37) hold for u^n solving the difference equation (4.52). □

With the results achieved in Lemma 4.6 at hands, now Theorem 4.2 can be proved:

Proof of Theorem 4.2 From Lemma 4.6, the H^1 estimate (4.37) holds for u^n obtained from the truncated scheme (4.52). By applying the inverse inequality, one has

$$\|e^n\|_{l^\infty} \lesssim \|e^n\|_{l^2} + \|\delta_x^+e^n\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6},$$

and thus, under the assumption $\tau \lesssim \varepsilon^3$,

$$\|e^n\|_{l^\infty} \le 1,$$

if τ and h are sufficiently small. Noting the properties of the cut-off function (4.50) and truncated nonlinear term (4.51), one immediately realize that the solutions u^n of (4.52) collapse to the solutions of the original finite difference scheme (4.11), under the assumptions put in Theorem 4.2. Therefore, the proof of Theorem 4.2 is accomplished. □

4.2.5 Proofs of Theorems 4.3, 4.4 and 4.5

Proof of Theorem 4.3 For simplicity of presenting, f is first treated as the truncated nonlinearity (4.51), i.e. f is assumed to be global Lipschitz and satisfy

$$\|f'(v)\|_{L^\infty(\mathbb{R})} + \|f''(v)\|_{L^\infty(\mathbb{R})} \lesssim 1. \tag{4.68}$$

Define

$$\begin{aligned}
\tilde{\xi}_j^n : &= \varepsilon^2\delta_t^2\left(u(x_j,t_n)\right) - \delta_x^2\left(u(x_j,t_n)\right) + \frac{1}{\varepsilon^2}u(x_j,t_n) + f\left(u(x_j,t_n)\right), & (4.69)\\
\tilde{\eta}_j^n : &= f\left(u(x_j,t_n)\right) - f\left(u_j^n\right), \quad j = 0,1,\ldots,M, \quad n \ge 1, & (4.70)
\end{aligned}$$

and $\tilde{\xi}_j^0$ is defined the same as in Lemma 4.3. Similar to Lemmas 4.3 and 4.5, one can prove

$$\begin{aligned}
&\left\|\tilde{\xi}^n\right\|_{l^2} + \left\|\delta_x^+\tilde{\xi}^n\right\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, \quad 0 \le n \le \frac{T}{\tau}, \quad \left\|\delta_x^2\tilde{\xi}^0\right\|_{l^2} \lesssim h^2 + \frac{\tau^2}{\varepsilon^6}, & (4.71)\\
&\|\tilde{\eta}^n\|_{l^2}^2 \lesssim \|e^n\|_{l^2}^2, \quad \left\|\delta_x^+\tilde{\eta}^n\right\|_{l^2}^2 \lesssim \|e^n\|_{l^2}^2 + \left\|\delta_x^+e^n\right\|_{l^2}^2, \quad n \ge 1. & (4.72)
\end{aligned}$$

Subtracting (4.14) from (4.69), noticing (4.16), (4.17) and (4.70),

$$\varepsilon^2\delta_t^2e_j^n - \delta_x^2e_j^n + \frac{1}{\varepsilon^2}e_j^n = \tilde{\xi}_j^n - \tilde{\eta}_j^n, \quad j = 1,2,\ldots,M-1, \tag{4.73}$$

$$e_j^0 = 0, \quad e_j^1 = \tau\tilde{\xi}_j^0, \quad e_0^n = e_M^n = 0, \quad j = 0,1,\ldots,M, \quad n = 0,1,\ldots. \tag{4.74}$$

Define the “energy” for the error vector e^n $(n = 0, 1, \ldots)$ as

$$\begin{aligned} \mathcal{S}^n \quad := \quad & \left(\varepsilon^2 - \frac{\tau^2}{2\varepsilon^2} - \frac{\tau^2}{h^2}\right) \|\delta_t^+ e^n\|_{l^2}^2 + \frac{1}{2\varepsilon^2}\left(\|e^{n+1}\|_{l^2}^2 + \|e^n\|_{l^2}^2\right) \\ & + \frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(e_{j+1}^{n+1} - e_j^n\right)^2 + \left(e_{j+1}^n - e_j^{n+1}\right)^2\right], \quad n \geq 0. \end{aligned} \tag{4.75}$$

Similar to the proof of Theorem 4.2, with the help of (4.21) and (4.22), noticing (4.73), (4.14), (4.16), (4.17), (4.75) and restriction on time step τ, in view of the estimates (4.71), one can obtain

$$\mathcal{S}^n \lesssim \mathcal{S}^0 + \frac{1}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 0 \leq n \leq \frac{T}{\tau} - 1. \tag{4.76}$$

Plugging (4.74) into (4.75) with $n = 0$,

$$\mathcal{S}^0 \lesssim \left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2 \left(1 + \frac{\tau^2}{\varepsilon^2}\right). \tag{4.77}$$

Similarly, define another “energy” as

$$\begin{aligned} \hat{\mathcal{S}}^n \quad := \quad & \left(\varepsilon^2 - \frac{\tau^2}{2\varepsilon^2} - \frac{\tau^2}{h^2}\right) \|\delta_x^+ \delta_t^+ e^n\|_{l^2}^2 + \frac{1}{2\varepsilon^2}\left(\|\delta_x^+ e^{n+1}\|_{l^2}^2 + \|\delta_x^+ e^n\|_{l^2}^2\right) \\ & + \frac{1}{2h}\sum_{j=0}^{M-1}\left[\left(\delta_x^+ e_{j+1}^{n+1} - \delta_x^+ e_j^n\right)^2 + \left(\delta_x^+ e_{j+1}^n - \delta_x^+ e_j^{n+1}\right)^2\right], \quad n \geq 0, \end{aligned} \tag{4.78}$$

one can obtain

$$\mathcal{S}^n \lesssim \frac{1}{\varepsilon^2}\left(h^2 + \frac{\tau^2}{\varepsilon^6}\right)^2, \quad 0 \leq n \leq \frac{T}{\tau} - 1. \tag{4.79}$$

Thus (4.38) is a combination of (4.76), (4.77), (4.79), (4.75) and (4.78) by noticing $\|e^n\|_{l^2}^2 + \|e^{n+1}\|_{l^2}^2 \leq 2\varepsilon^2\mathcal{S}^n$, $\|\delta_x^+ e^n\|_{l^2}^2 + \|\delta_x^+ e^{n+1}\|_{l^2}^2 \leq 2\varepsilon^2\hat{\mathcal{S}}^n$ and $0 < \varepsilon \leq 1$.

Noting that the above argument is based on a stronger assumption on the nonlinear term f, i.e. (4.68). To obtain the estimates (4.38) with a weaker assumption $f \in C^2(\mathbb{R})$ as put in Theorem 4.3, one can apply the cut-off technique used in the proof of Theorem 4.2, i.e. by requiring $\tau \lesssim \varepsilon^3$ to control the l^∞-error from H^1-error. The details are omitted here for brevity. □

Proof of Theorem 4.4 and 4.5 Follow the analogous proofs to Theorems 4.2 and 4.3, with the help of Lemma 4.1; we omit the details here for brevity. □

4.3 Exponential wave integrator and its analysis

In this section, new numerical methods, which have better temporal resolution capacity than that of the FDTD methods in the nonrelativistic limit regime, are proposed with rigorous stability and convergence analysis. Again, for simplicity of notations, the schemes and their analysis are only presented for 1D problem with homogeneous Dirichlet boundary conditions, i.e. (4.8)–(4.10). Generalization to higher dimensions is straightforward and the error estimates remain valid without modifications.

4.3.1 Numerical methods

First, the Gautschi-type exponential wave integrator sine spectral method is proposed, which is based on the application of sine spectral approach to spatial discretization followed by a Gautschi-type exponential wave integrator [63, 72, 73, 77, 83] to time discretization. Let

$$Y_M = \text{span}\{\sin\left(\mu_l(x-a)\right),\ x \in [a,b],\ l = 1, 2, \dots, M-1\},$$

with

$$\mu_l = \frac{\pi l}{b-a}.$$

For any function $v(x)$ on $[a, b]$ satisfying $v(a) = v(b) = 0$, and vector $v \in X_M$, define $\mathcal{P}_M :\ L^2(a,b) \to Y_M$ as the standard projection operator, $\mathcal{I}_M :\ C_0(a,b) \to Y_M$ and $\mathcal{I}_M :\ X_M \to Y_M$ as the trigonometric interpolation operators [133], i.e.

$$(\mathcal{P}_M v)(x) = \sum_{l=1}^{M-1} \widehat{v}_l \sin\left(\mu_l(x-a)\right), \quad (\mathcal{I}_M v)(x) = \sum_{l=1}^{M-1} \widetilde{v}_l \sin\left(\mu_l(x-a)\right), \quad a \le x \le b,$$

with $(l = 1, 2, \dots, M-1)$,

$$\widehat{v}_l = \frac{2}{b-a}\int_a^b v(x)\sin\left(\mu_l(x-a)\right)\mathrm{d}x, \quad \widetilde{v}_l = \frac{2}{M}\sum_{j=1}^{M-1} v_j \sin\left(\frac{jl\pi}{M}\right),$$

where v_j is interpreted as $v(x_j)$ for a function $v(x)$. In addition, the same notation as vector case (4.18) is adopted to define the discrete l^2-norm for a function $v(x) \in C_0(a,b)$, i.e. $\|v\|_{l^2}^2 = h\sum_{j=0}^{M-1} |v(x_j)|^2$.

The sine spectral method for (4.8)–(4.10) is as follows:

Find $u_M(x,t) \in Y_M$, i.e.

$$u_M(x,t) = \sum_{l=1}^{M-1} \widehat{u}_l(t) \sin\left(\mu_l(x-a)\right), \quad a \le x \le b, \quad t \ge 0, \tag{4.80}$$

such that

$$\varepsilon^2 \partial_{tt} u_M(x,t) - \partial_{xx} u_M + \frac{1}{\varepsilon^2} u_M + \mathcal{P}_M f(u_M) = 0, \quad a \le x \le b, \quad t \ge 0. \tag{4.81}$$

Plugging (4.80) into (4.81), noticing the orthogonality of the sine bases functions,

$$\varepsilon^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} \widehat{u}_l(t) + \frac{1+\varepsilon^2\mu_l^2}{\varepsilon^2} \widehat{u}_l(t) + \widehat{f(u_M)}_l(t) = 0, \quad l = 1,2,\ldots,M-1, \quad t \ge 0. \tag{4.82}$$

For each fixed l $(l = 1,2,\ldots,M-1)$, when t is near $t = t_n$ $(n \ge 0)$, the above ODEs are re-written as

$$\frac{\mathrm{d}^2}{\mathrm{d}w^2} \widehat{u}_l(t_n+w) + \left(\beta_l^n\right)^2 \widehat{u}_l(t_n+w) + \frac{1}{\varepsilon^2} \widehat{g}_l^n(w) = 0, \quad w \in \mathbb{R}, \tag{4.83}$$

where

$$\beta_l^n = \frac{1}{\varepsilon^2}\sqrt{1+\varepsilon^2\left(\mu_l^2+\alpha^n\right)}, \quad \widehat{g}_l^n(w) = \widehat{f(u_M)}_l(t_n+w) - \alpha^n \widehat{u}_l(t_n+w). \tag{4.84}$$

Here, a linear stabilization term with stabilizing constant α^n satisfying $1+\varepsilon^2\alpha^n > 0$ is introduced, such that the scheme is unconditionally stable (see below for its choice). Using the variation-of-constants formula as in the Gautschi-type exponential wave integrator for oscillatory second-order differential equations [72, 73, 77, 83], the general solution of the above second-order ODEs can be written as

$$\widehat{u}_l(t_n+w) = c_l^n \cos\left(w\beta_l^n\right) + d_l^n \frac{\sin\left(w\beta_l^n\right)}{\beta_l^n} - \int_0^w \widehat{g}_l^n(s) \frac{\sin\left(\beta_l^n(w-s)\right)}{\varepsilon^2\beta_l^n} \mathrm{d}s, \tag{4.85}$$

where c_l^n and d_l^n are two constants to be determined.

Now the key problem is how to choose two proper transmission conditions for the second-order ODEs (4.83) between different time intervals so that one can uniquely determine the two constants in (4.85). When $n = 0$, considering the solution (4.85) for $w \in [0, \tau]$, from the initial conditions these two conditions can be chosen naturally as

$$\widehat{u}_l(0) = \widehat{\phi}_l, \quad \frac{\mathrm{d}}{\mathrm{d}t}\widehat{u}_l(0) = \frac{1}{\varepsilon^2}\widehat{\gamma}_l. \tag{4.86}$$

Plugging (4.86) into (4.85) with $n = 0$ to determine the two constants c_l^0 and d_l^0 and then letting $w = \tau$ leads to

$$\widehat{u}_l(\tau) = \widehat{\phi}_l \cos\left(\tau\beta_l^0\right) + \widehat{\gamma}_l \frac{\sin\left(\tau\beta_l^0\right)}{\varepsilon^2\beta_l^0} - \int_0^\tau \widehat{g}_l^0(s)\frac{\sin\left(\beta_l^0(\tau - s)\right)}{\varepsilon^2\beta_l^0}\mathrm{d}s. \tag{4.87}$$

For $n > 0$, one can consider the solution in (4.85) for $w \in [-\tau, \tau]$ and require the solution to be continuous at $t = t_n$ and $t = t_{n-1} = t_n - \tau$. Plugging $w = 0$ and $w = -\tau$ into (4.85) to determine the two constants c_l^n and d_l^n and then letting $w = \tau$, noticing (4.84),

$$\widehat{u}_l(t_{n+1}) = -\widehat{u}_l(t_{n-1}) + 2\cos(\tau\beta_l^n)\widehat{u}_l(t_n) - \int_0^\tau \left[\widehat{g}_l^n(-s) + \widehat{g}_l^n(s)\right]\frac{\sin\left(\beta_l^n(\tau - s)\right)}{\varepsilon^2\beta_l^n}\mathrm{d}s. \tag{4.88}$$

In order to design an explicit scheme, the integrals in (4.87) and (4.88) are approximated by the following quadratures

$$\int_0^\tau \widehat{g}_l^0(s)\sin\left(\beta_l^0(\tau - s)\right)\mathrm{d}s \approx \widehat{g}_l^0(0)\int_0^\tau \sin\left(\beta_l^0(\tau - s)\right)\mathrm{d}s = \frac{\widehat{g}_l^0(0)}{\beta_l^0}\left[1 - \cos(\tau\beta_l^0)\right],$$
$$\int_0^\tau \left[\widehat{g}_l^n(-s) + \widehat{g}_l^n(s)\right]\sin\left(\beta_l^n(\tau - s)\right)\mathrm{d}s \approx \frac{2\widehat{g}_l^n(0)}{\beta_l^n}\left[1 - \cos(\tau\beta_l^n)\right].$$

Denote $\widehat{(u_M^n)}_l$ and $u_M^n(x)$ be the approximations of $\widehat{u}_l(t_n)$ and $u_M(x, t_n)$, respectively. Choosing $u_M^0(x) = (\mathcal{P}_M\phi)(x)$ and noticing (4.84), then a Gautschi-type exponential wave integrator sine spectral discretization for the KG equation (4.8)–(4.10) is

$$u_M^{n+1}(x) = \sum_{l=1}^{M-1}\widehat{(u_M^{n+1})}_l \sin\left(\mu_l(x - a)\right), \quad a \le x \le b, \quad n = 0, 1, \dots, \tag{4.89}$$

where

$$\widehat{(u_M^1)}_l = p_l^0\, \widehat{\phi}_l + q_l^0\, \widehat{\gamma}_l + r_l^0\, \widehat{(f(\phi))}_l, \quad l = 1, 2, \ldots, M-1, \tag{4.90}$$

$$\widehat{(u_M^{n+1})}_l = -\widehat{(u_M^{n-1})}_l + p_l^n\, \widehat{(u_M^n)}_l + r_l^n\, \widehat{(f(u_M^n))}_l, \quad n \geq 1, \tag{4.91}$$

with

$$p_l^0 = \cos(\tau\beta_l^0) + \frac{\alpha^0(1-\cos(\tau\beta_l^0))}{(\varepsilon\beta_l^0)^2}, \quad q_l^0 = \frac{\sin(\tau\beta_l^0)}{\varepsilon^2\beta_l^0}, \quad r_l^0 = \frac{\cos(\tau\beta_l^0)-1}{(\varepsilon\beta_l^0)^2}, \tag{4.92}$$

$$p_l^n = 2\left[\cos(\tau\beta_l^n) + \frac{\alpha^n(1-\cos(\tau\beta_l^n))}{(\varepsilon\beta_l^n)^2}\right], \quad r_l^n = \frac{2\,(\cos(\tau\beta_l^n)-1)}{(\varepsilon\beta_l^n)^2}, \quad n \geq 1. \tag{4.93}$$

As demonstrated in the literature [63, 72, 73, 77, 83], the above Gautschi-type exponential wave integrator gives exact solution to the linear second-order ODEs (4.83) and has favorable properties compared to standard time integrators for oscillatory second-order ODEs. The next two subsections will demonstrate that the above discretization gives exact solution in time to the linear KG equation (4.8)–(4.10), i.e. $f(u) = \alpha u$, under the choice of $\alpha^n = \alpha$ $(n \geq 0)$ in (4.83), and respectively, performs much better resolution in time than that of the FDTD methods for the nonlinear KG equation. One remark here is that similar techniques in time discretization have been used in discretizing wave-type equations in Zakharov system [24], Maxwell–Dirac equations [22] and Klein–Gordon–Schrödinger equations [25].

The above procedure is not suitable in practice due to the difficulty of computing the integrals in (4.90) and (4.91). An efficient implementation is achieved by choosing $u_M^0(x)$ as the interpolation of $\phi(x)$ on the grids $\{x_j,\ j = 0, 1, \ldots, M\}$, i.e. $u_M^0(x) = (I_M\phi)(x)$, and approximating the integrals in (4.90) and (4.91) by a quadrature rule on the grids. Let u_j^n be the approximation of $u(x_j, t_n)$ and denote $u_j^0 = \phi(x_j)$ $(j = 0, 1, \ldots, M)$. For $n = 0, 1, \ldots,$ a Gautschi-type exponential wave integrator sine pseudospectral (Gautschi-SP) discretization for the KG equation

(4.8)–(4.10) is

$$u_j^{n+1} = \sum_{l=1}^{M-1} \widetilde{(u^{n+1})}_l \sin\left(\frac{jl\pi}{M}\right), \quad j = 1, 2, \ldots, M, \tag{4.94}$$

where,

$$\begin{aligned} &\widetilde{(u^1)}_l = p_l^0\, \widetilde{\phi}_l + q_l^0\, \widetilde{\gamma}_l + r_l^0\, \widetilde{(f(\phi))}_l, \quad l = 1, 2, \ldots, M-1, \\ &\widetilde{(u^{n+1})}_l = -\widetilde{(u^{n-1})}_l + p_l^n\, \widetilde{(u^n)}_l + r_l^n\, \widetilde{(f(u^n))}_l, \quad n \geq 1, \end{aligned}$$

with p_l^n, q_l^0 and r_l^n are given in (4.92) and (4.93). Based on the results in Theorem 4.6 (see below), in practice, α^n is suggested to be chosen as: if $f(v) = \alpha v$ is a linear function with α a constant, choose $\alpha^n = \max\{-1/\varepsilon^2,\ \alpha\}$ for $n \geq 0$, and respectively, if $f(v)$ is a nonlinear function, choose $\alpha^{-1} = 0$ and

$$\alpha^n = \max\left\{\alpha^{n-1},\ \max_{u_j^n \neq 0,\ 1 \leq j \leq M-1} f(u_j^n)/u_j^n\right\}, \quad n \geq 0. \tag{4.95}$$

This Gautschi-SP discretization is explicit, time symmetric and easy to extend to 2D and 3D. The memory cost is $O(M)$ and computation cost per time step is $O(M \ln M)$ thanks fast sine transform (FST).

Remark 4.2. *Another way to approximate the integrals in (4.87) and (4.88) is to use the trapezoidal rule:*

$$\begin{aligned} &\int_0^\tau \widehat{g}_l^0(s) \sin\left(\beta_l^0(\tau - s)\right) ds \approx \frac{\tau}{2} \widehat{g}_l^0(0) \sin\left(\tau\beta_l^0\right), \\ &\int_0^\tau \left[\widehat{g}_l^n(-s) + \widehat{g}_l^n(s)\right] \sin\left(\beta_l^n(\tau - s)\right) ds \approx \tau\, \widehat{g}_l^n(0) \sin\left(\tau\beta_l^n\right). \end{aligned}$$

The rest of computations can be carried out in a similar manner.

For comparison, the Gautschi-type exponential wave integrator finite difference (Gautschi-FD) method is also introduced here, which is based on applying centered finite difference to spatial discretization followed by a Gautschi-type integrator to time discretization. The aim is to show that the temporal resolution capacity of the Gautschi-type integrator for wave-type equation is independent of the spatial discretization which it follows [74]. Let $u_j(t)$ be the approximation of $u(x_j, t)$ $(j =$

$0, 1, \ldots, M$). Applying a centered finite difference to the spatial derivative in (4.8)–(4.10),

$$\varepsilon^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} u_j(t) - \delta_x^2 u_j(t) + \frac{1}{\varepsilon^2} u_j(t) + f(u_j(t)) = 0, \quad j = 1, 2, \ldots, M-1, \tag{4.96}$$

with $u_0(t) = u_M(t) = 0$.

Let

$$U(t) = (u_1(t), u_2(t), \ldots, u_{M-1}(t))^T,$$

$$F(U(t)) = (f(u_1(t)), f(u_2(t)), \ldots, f(u_{M-1}(t)))^T,$$

then the above ODEs can be re-written as

$$\varepsilon^2 U''(t) + A\,U(t) + F(U(t)) = 0, \quad t \geq 0, \tag{4.97}$$

where A is a $(M-1) \times (M-1)$ matrix independent of t. Since A is symmetric, it is normal, i.e. there exists an orthogonal matrix P and a diagonal matrix Λ such that

$$A = P^{-1}\,\Lambda\,P.$$

Let $V(t) = P\,U(t)$ and multiply P to both sides of (4.97), one can get

$$\varepsilon^2 V''(t) + \Lambda V(t) + P\,F\,(U(t)) = 0, \quad t \geq 0. \tag{4.98}$$

The above second-order ODEs are similar to (4.83) and the Gautschi-type exponential wave integrator can be applied to discretize it, which immediately gives a discretization of (4.96). The resulting scheme is quite similar to (4.94), with μ_l in (4.84) replaced by $2\sin(l\pi/2M)/h$.

4.3.2 Stability and convergence analysis in linear case

In this subsection, $f(u)$ is assumed to be a linear function, i.e. $f(u) = \alpha u$ with α being a constant satisfying $\alpha > -\varepsilon^{-2}$. In this case, the solution of (4.8)–(4.10) is

$$u(x,t) = \sum_{l=1}^{\infty} \left[\widehat{\phi}_l \cos(t\beta_l) + \widehat{\gamma}_l \frac{\sin(t\beta_l)}{\varepsilon^2 \beta_l} \right] \sin\left(\mu_l(x-a)\right), \quad a \leq x \leq b, \ t \geq 0, \tag{4.99}$$

where

$$\beta_l = \frac{1}{\varepsilon^2}\sqrt{1+\varepsilon^2\left(\mu_l^2+\alpha\right)}, \quad l=1,2,\dots . \tag{4.100}$$

Again, by using the standard von Neumann analysis [136], one can have the following stability results for Gautschi-SP and Gautschi-FD:

Theorem 4.6. *If α^n in (4.84) is chosen such that $\alpha^n \geq \alpha$ for $n \geq 0$, then both Gautschi-SP and Gautschi-FD are unconditionally stable for any $\tau > 0$, $h > 0$ and $\varepsilon > 0$.*

Proof. Similar to the proof of Theorem 4.1, noticing (4.93) and (4.94), one has the same characteristic equation (4.28) for Gautschi-SP with

$$\begin{aligned}\theta_l = \cos(\tau\beta_l^n) &+ \frac{(\alpha^n-\alpha)(1-\cos(\tau\beta_l^n))}{(\varepsilon\beta_l^n)^2}\\ = \quad &\cos^2\left(\frac{\tau\beta_l^n}{2}\right) + \left[\frac{2(\alpha^n-\alpha)}{\varepsilon^{-2}+\mu_l^2+\alpha^n} - 1\right]\sin^2\left(\frac{\tau\beta_l^n}{2}\right).\end{aligned}$$

Since $\alpha^n \geq \alpha > -\varepsilon^{-2}$ $(n \geq 0)$,

$$0 \leq \frac{2(\alpha^n-\alpha)}{\varepsilon^{-2}+\mu_l^2+\alpha^n} \leq 2 \quad \Longrightarrow \quad |\theta_l| \leq 1, \tag{4.101}$$

which immediately leads to the unconditional stability of the Gautschi-SP. For Gautschi-FD, one only need to replace μ_l in (4.101) by $2\sin(l\pi/2M)/h$ and the stability claim follows immediately. □

Let $u_I(x,t)$ be the solution of the following problem

$$\varepsilon^2\partial_{tt}u_I(x,t) - \partial_{xx}u_I + \frac{1}{\varepsilon^2}u_I + \alpha u_I = 0, \quad a < x < b, \quad t > 0, \tag{4.102}$$

$$u_I(a,t) = u_I(b,t) = 0, \quad t \geq 0, \tag{4.103}$$

$$u_I(x,0) = \left(\mathcal{I}_M\phi\right)(x), \quad \partial_t u_I(x,0) = \frac{1}{\varepsilon^2}\left(\mathcal{I}_M\gamma\right)(x), \quad a \leq x \leq b. \tag{4.104}$$

It is easy to see that the solution of the above problem is

$$u_I(x,t) = \sum_{l=1}^{M-1}\left[\cos(t\beta_l)\widetilde{\phi}_l + \widetilde{\gamma}_l\frac{\sin(t\beta_l)}{\varepsilon^2\beta_l}\right]\sin\left(\mu_l(x-a)\right), \quad a \leq x \leq b,\ t \geq 0. \tag{4.105}$$

Denote

$$e_j^n = u(x_j, t_n) - u_j^n, \quad j = 0, 1, \ldots, M, \quad n \geq 0,$$
$$e^n(x) := u(x, t_n) - (\mathcal{I}_M u^n)(x), \quad a \leq x \leq b, \quad n \geq 0.$$

For Gautschi-SP, one can have the following error estimates:

Theorem 4.7. *Let u_j^n be the solution of Gautschi-SP (4.94) with $\alpha^n = \alpha$ in (4.84) for $n \geq 0$, then,*

$$u_j^n = u_I(x_j, t_n), \quad j = 0, 1, \ldots, M, \quad n \geq 0. \tag{4.106}$$

In addition, if ϕ, $\gamma \in H_s^m := \{v \in H^m(a,b) \mid \partial_x^{2l} v(a) = \partial_x^{2l} v(b) = 0, 0 \leq 2l \leq m\}$ with $m \geq 2$, when $\alpha \geq 0$ for any $\varepsilon > 0$ or when $\alpha < 0$ for $0 < \varepsilon \leq \varepsilon_0 := \frac{1}{\sqrt{2|\alpha|}}$, the following error estimates hold,

$$\|e^n(x)\|_{L^2} \lesssim h^m, \quad \|\partial_x e^n(x)\|_{L^2} \lesssim h^{m-1}, \quad n \geq 0. \tag{4.107}$$

Thus if initial conditions ϕ and γ are smooth, for the linear KG equation, the Gautschi-SP converges exponentially fast in space with no error in time discretization.

Proof. From (4.102)–(4.104), one has $u_I(x_j, 0) = (\mathcal{I}_M \phi)(x_j) = \phi(x_j) = u_j^0$ for $j = 0, 1, \ldots, M$. Thus (4.106) is valid for $n = 0$. From (4.84) and (4.100), when $\alpha^n = \alpha$ for $n \geq 0$,

$$\beta_l^n = \beta_l, \quad n \geq 0, \quad l = 1, 2, \ldots, M-1. \tag{4.108}$$

Plugging (4.92) into (4.94) with $n = 0$, noticing (4.108) and (4.105),

$$\begin{aligned} u_j^1 &= \sum_{l=1}^{M-1} \left[p_l^0 \widetilde{\phi}_l + q_l^0 \widetilde{\gamma}_l + r_l^0 \alpha \widetilde{\phi}_l \right] \sin\left(\frac{jl\pi}{M}\right) \\ &= \sum_{l=1}^{M-1} \left[\widetilde{\phi}_l \cos(t_1 \beta_l) + \widetilde{\gamma}_l \frac{\sin(t_1 \beta_l)}{\varepsilon^2 \beta_l} \right] \sin\left(\frac{jl\pi}{M}\right) \\ &= u_I(x_j, t_1), \quad j = 0, 1, \ldots, M. \end{aligned} \tag{4.109}$$

Thus (4.106) is valid for $n = 1$. Assume (4.106) is valid for $n = 0, 1, \ldots, m$. When $n = m + 1$, from (4.94) with $n = m$, noticing (4.93) and (4.108),

$$\begin{aligned}\widetilde{(u^{m+1})}_l &= -\widetilde{(u^{m-1})}_l + p_l^m\,\widetilde{(u^m)}_l + r_l^m\,\alpha\widetilde{(u^m)}_l = -\widetilde{(u^{m-1})}_l + 2\cos(\tau\beta_l)\,\widetilde{(u^m)}_l \\ &= \ -\left[\widetilde{\phi}_l\cos(t_{m-1}\beta_l) + \widetilde{\gamma}_l\frac{\sin(t_{m-1}\beta_l)}{\varepsilon^2\beta_l}\right] + 2\cos(\tau\beta_l)\left[\widetilde{\phi}_l\cos(t_m\beta_l) + \widetilde{\gamma}_l\frac{\sin(t_m\beta_l)}{\varepsilon^2\beta_l}\right] \\ &= \ \widetilde{\phi}_l\cos(t_{m+1}\beta_l) + \widetilde{\gamma}_l\frac{\sin(t_{m+1}\beta_l)}{\varepsilon^2\beta_l}, \quad l = 1, 2, \ldots, M-1.\end{aligned}$$

Plugging the above equality into (4.94) with $n = m$ and noticing (4.105) with $t = t_{m+1}$, one can obtain (4.106) for $n = m + 1$, thus the claim (4.106) is verified by mathematical induction. From (4.106), noticing (4.99) and (4.105),

$$\begin{aligned}&\|e^n(x)\|_{L^2}^2 \lesssim \|\phi - \mathcal{I}_M\phi\|_{L^2}^2 + \|\gamma - \mathcal{I}_M\gamma\|_{L^2}^2 \lesssim h^{2m}, \\ &\|\partial_x e^n(x)\|_{L^2}^2 \lesssim \|\partial_x(\phi - \mathcal{I}_M\phi)\|_{L^2}^2 + \|\partial_x(\gamma - \mathcal{I}_M\gamma)\|_{L^2}^2 \lesssim h^{2(m-1)},\end{aligned}$$

which complete the proof of (4.107). □

Also, the following error estimates hold for Gautschi-FD in linear case,

Theorem 4.8. *Let u_j^n be the solution of Gautschi-FD with $\alpha^n = \alpha$ for $n \geq 0$. If $\phi,\ \gamma \in W^{4,\infty} \cap H_0^1$, when $\alpha \geq 0$ for any $\varepsilon > 0$ or when $\alpha < 0$ for $0 < \varepsilon \leq \varepsilon_0 := \frac{1}{\sqrt{2|\alpha|}}$,*

$$\|e^n\|_{l^2} \lesssim h^2, \quad 0 \leq n \leq \frac{T}{\tau}. \tag{4.110}$$

Proof. Let $u_j(t)$ be the solution of (4.96) with initial condition

$$u_j(0) = \phi(x_j), \quad \frac{\mathrm{d}}{\mathrm{d}t}u_j(0) = \frac{1}{\varepsilon^2}\gamma(x_j), \quad j = 0, 1, \ldots, M.$$

Similar to the proof of Theorem 4.7, for $j = 0, 1, \ldots, M$,

$$u_j^n = u_j(t_n) = \sum_{l=1}^{M-1}\left[\widetilde{\phi}_l\cos(t\beta_l^h) + \widetilde{\gamma}_l\frac{\sin(t\beta_l^h)}{\varepsilon^2\beta_l^h}\right]\sin\left(\frac{jl\pi}{M}\right),$$

where

$$\beta_l^h = \frac{1}{\varepsilon^2}\sqrt{1 + \varepsilon^2\left(\frac{4\sin^2(l\pi/2M)}{h^2} + \alpha\right)} \geq \frac{1}{\sqrt{2}\varepsilon^2}, \quad l = 1, 2, \ldots. \tag{4.111}$$

Let

$$\begin{aligned}
&e_j(t) = u(x_j,t) - u_j(t), \quad j = 0,1,\ldots,M,\\
&\xi_j(t) = \varepsilon^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} u(x_j,t) - \delta_x^2 u(x_j,t) + \left(\frac{1}{\varepsilon^2} + \alpha\right) u(x_j,t)\\
&\qquad = \frac{h^2}{12}\partial_{xxxx} u(\tilde{x}_j(t),t), \quad j = 1,2,\ldots,M-1,\\
&\xi_0(t) = \xi_M(t) = 0,
\end{aligned} \tag{4.112}$$

where $\tilde{x}_j(t)$ is located between x_{j-1} and x_{j+1}. Subtracting (4.96) from (4.112),

$$\varepsilon^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} e_j(t) - \delta_x^2 e_j(t) + \left(\frac{1}{\varepsilon^2} + \alpha\right) e_j(t) = \xi_j(t), \ j = 1,2,\ldots,M-1, \ t > 0, \tag{4.113}$$

$$e_0(t) = e_M(t) = 0, \quad e_j(0) = 0, \quad \frac{\mathrm{d}}{\mathrm{d}t} e_j(0) = 0. \tag{4.114}$$

Taking discrete sine transform of (4.113),

$$\begin{aligned}
&\varepsilon^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} \widetilde{e}_l(t) + \left(\varepsilon\beta_l^h\right)^2 \widetilde{e}_l(t) = \widetilde{\xi}_l(t),\\
&\widetilde{e}_l(0) = 0, \quad \frac{\mathrm{d}}{\mathrm{d}t}\widetilde{e}_l(0) = 0, \quad l = 1,2,\ldots,M-1.
\end{aligned}$$

Solving the above ODEs,

$$\widetilde{e}_l(t) = \frac{1}{\beta_l^h \varepsilon^2} \int_0^t \sin(\beta_l^h (t-s)) \widetilde{\xi}_l(s) \mathrm{d}s, \quad l = 1,2,\ldots,M-1. \tag{4.115}$$

Plugging (4.112) into (4.115), noticing ϕ, $\gamma \in W^{4,\infty} \cap H_0^1$ and (4.99), using the Hölder's inequality and Parseval's identity, one can obtain

$$\begin{aligned}
\sum_{l=1}^{M-1} |\widetilde{e}_l(t)|^2 &\le 2 \sum_{l=1}^{M-1} \left[\int_0^t \left|\widetilde{\xi}_l(s)\right| \mathrm{d}s\right]^2 \le 2t \int_0^t \sum_{l=1}^{M-1} \left|\widetilde{\xi}_l(s)\right|^2 \mathrm{d}s\\
&\le \frac{4t}{M} \int_0^t \sum_{j=0}^{M-1} |\xi_j(t)|^2 \mathrm{d}s \le \frac{4T}{M} \int_0^T \sum_{j=0}^{M-1} |\xi_j(s)|^2 \mathrm{d}s \lesssim h^4, \quad 0 \le t \le T.
\end{aligned}$$

Noticing $e_j^n = e_j(t_n)$ ($j = 0,1,\ldots,M$, $0 \le n \le T/\tau$) and using the Parseval's equality, the estimate (4.110) follows immediately. □

Based on Theorems 4.7 and 4.8, both Gautschi-SP and Gautschi-FD introduce no error in time discretization for the linear KG equation, and exhibit the same temporal resolution in the nonrelativistic limit regime. In fact, for a given accuracy $\delta > 0$, for the linear KG equation the ε-scalability of the two methods is

$$\tau = O(1), \quad h \le O\left(\sqrt{\delta}\right) = O(1), \quad 0 < \varepsilon \ll 1, \tag{4.116}$$

i.e. both mesh size h and time step τ can be chosen independently of the small parameter ε.

4.3.3 Convergence analysis in the nonlinear case

In order to obtain an error estimate for Gautschi-SP method (4.89) with (4.95), let $0 < T < T^*$ with T^* the maximum existence time of the solution, motivated by the results in [107, 110], assume that there exists an integer $m_0 \ge 2$ such that

(B)
$$u \in C^2\left([0,T]; H^1\right) \cap C^1\left([0,T]; W^{1,4}\right) \cap C\left([0,T]; W^{1,\infty} \cap H^{m_0} \cap H_0^1\right),$$
$$\|\partial_t u(x,t)\|_{L^\infty([0,T];W^{1,4})} \lesssim \frac{1}{\varepsilon^2}, \quad \|\partial_{tt} u(x,t)\|_{L^\infty([0,T];H^1)} \lesssim \frac{1}{\varepsilon^4},$$
$$\|u(x,t)\|_{L^\infty([0,T];W^{1,\infty}\cap H_0^{m_0})} \lesssim 1.$$

Under the above assumption (B) and assume $f \in C^3(\mathbb{R})$, the following are well-defined,

$$M_1 := \max_{0\le t\le T} \|u(x,t)\|_{W^{1,\infty}} \lesssim 1, \quad M_2 := \max_{|v|\le 1+M_1} \sum_{l=1}^{3} |f^{(l)}(v)| \lesssim 1, \tag{4.117}$$

$$M_3 := \max\left\{0, \sup_{0\ne v,\, |v|\le 1+M_1} f(v)/v\right\} \le M_2 \lesssim 1. \tag{4.118}$$

Assuming

$$\tau \le \frac{\pi\varepsilon^2 h}{3\sqrt{h^2 + \varepsilon^2(\pi^2 + M_3 h^2)}}, \tag{4.119}$$

one can have,

Theorem 4.9. *Let $u_M^n(x)$ be the approximation obtained from Gautschi-SP method (4.89) with (4.95). Assume $\tau \lesssim \varepsilon^2$ and $f(\cdot) \in C^3(\mathbb{R})$, under the assumption (B), there exist $h_0 > 0$ and $\tau_0 > 0$ sufficiently small and independent of ε such that, for* *any* $0 < \varepsilon \leq 1$, *when* $0 < h \leq h_0$ *and* $0 < \tau \leq \tau_0$ *and under the condition (4.119), the following error estimate holds,*

$$\|u(x,t_n) - u_M^n(x)\|_{L^2} \lesssim \frac{\tau^2}{\varepsilon^4} + h^{m_0}, \quad \|u_M^n\|_{L^\infty} \leq 1 + M_1, \tag{4.120}$$

$$\|\partial_x[u(x,t_n) - u_M^n(x)]\|_{L^2} \lesssim \frac{\tau^2}{\varepsilon^4} + h^{m_0-1}, \quad 0 \leq n \leq \frac{T}{\tau}. \tag{4.121}$$

Proof. The estimates (4.120)–(4.121) will be proved by the method of mathematical induction in the classical discrete energy framework. From the discretization of initial data, i.e. $u_M^0 = \mathcal{P}_M\phi$, one has

$$\begin{aligned}
&\|u(x,t=0) - u_M^0\|_{L^2} = \|\phi - \mathcal{P}_M\phi\|_{L^2} \lesssim h^{m_0},\\
&\|\partial_x[u(x,t=0) - u_M^0]\|_{L^2} = \|\partial_x\phi - \mathcal{P}_M\partial_x\phi\|_{L^2} \lesssim h^{m_0-1},\\
&\|u_M^0\|_{L^\infty} - M_1 \leq \|\mathcal{P}_M\phi - \phi\|_{L^\infty} + \|\phi\|_{L^\infty} - M_1 \lesssim h^{m_0-1}.
\end{aligned}$$

Thus there exists a $h_1 > 0$ sufficiently small and independent of ε such that, when $0 < h \leq h_1$, the three estimates in (4.120)–(4.121) are valid for $n = 0$.

Write the exact solution $u(x,t)$ as

$$u(x,t) = \sum_{l=1}^{\infty} \widehat{u}_l(t)\sin(\mu_l(x-a)), \quad a \leq x \leq b,$$

and denote the "error" function as

$$e^n(x) := \mathcal{P}_M u(x,t_n) - u_M^n(x) = \sum_{l=1}^{M-1} \widehat{e}_l^n \sin(\mu_l(x-a)), \quad a \leq x \leq b, \tag{4.122}$$

then one has

$$\widehat{e}_l^n = \widehat{u}_l(t_n) - \widehat{(u_M^n)}_l, \quad l = 1,2,\ldots,M-1, \quad n \geq 0, \tag{4.123}$$

with $\widehat{u}_l(t_n)$ $(l = 1,2,\ldots)$ the sine transform coefficients of $u(x,t_n)$. Using the triangle inequality and Parseval's equality,

$$\begin{aligned}
\|u(x,t_n) - u_M^n(x)\|_{L^2} &\leq \|u(x,t_n) - \mathcal{P}_M u(x,t_n)\|_{L^2} + \|e^n(x)\|_{L^2}\\
&\lesssim h^{m_0} + \sqrt{\sum_{l=1}^{M-1} |\widehat{e}_l^n|^2}, \quad 0 \leq n \leq \frac{T}{\tau}.
\end{aligned} \tag{4.124}$$

Similarly,

$$\|\partial_x[u(x,t_n)-u_M^n(x)]\|_{L^2}\lesssim h^{m_0-1}+\sqrt{\sum_{l=1}^{M-1}\mu_l^2|\widehat{e}_l^n|^2},\quad 0\le n\le\frac{T}{\tau}. \tag{4.125}$$

Thus, only the last terms in the above two inequalities need be estimated.

Similar to the derivation in (4.82)–(4.88), for $l=1,2,\dots,$

$$\widehat{u}_l(\tau)=\widehat{\phi}_l\cos\left(\tau\beta_l^0\right)+\widehat{\gamma}_l\frac{\sin\left(\tau\beta_l^0\right)}{\varepsilon^2\beta_l^0}-\int_0^\tau\widehat{G}_l^0(s)\frac{\sin\left(\beta_l^0(\tau-s)\right)}{\varepsilon^2\beta_l^0}ds, \tag{4.126}$$

$$\begin{aligned}\widehat{u}_l(t_{n+1})&=-\widehat{u}_l(t_{n-1})+2\cos(\tau\beta_l^n)\widehat{u}_l(t_n)\\&\quad-\int_0^\tau\left[\widehat{G}_l^n(-s)+\widehat{G}_l^n(s)\right]\frac{\sin\left(\beta_l^n(\tau-s)\right)}{\varepsilon^2\beta_l^n}ds,\quad n\ge1,\end{aligned} \tag{4.127}$$

where

$$\widehat{G}_l^n(s)=\widehat{(f(u))}_l(t_n+s)-\alpha^n\widehat{u}_l(t_n+s),\quad s\in\mathbb{R},\quad n\ge0. \tag{4.128}$$

For each $l=1,2,\dots,M-1$, subtracting (4.91) and (4.90) from (4.127) and (4.126), respectively, one can obtain the equation for "error" function $\widehat{e}_l^n$ as

$$\widehat{e}_l^{n+1}=-\widehat{e}_l^{n-1}+2\cos\left(\beta_l^n\tau\right)\widehat{e}_l^n+\widehat{\xi}_l^n,\quad 1\le n\le\frac{T}{\tau}-1, \tag{4.129}$$

$$\widehat{e}_l^0=0,\quad \widehat{e}_l^1=\widehat{\xi}_l^0, \tag{4.130}$$

where

$$\widehat{\xi}_l^n=\frac{1}{\varepsilon^2\beta_l^n}\int_0^\tau\widehat{W}_l^n(s)\sin\left(\beta_l^n(\tau-s)\right)ds,\quad 0\le n\le\frac{T}{\tau}-1, \tag{4.131}$$

with, for $0\le s\le\tau$,

$$\widehat{W}_l^n(s)=\begin{cases}\widehat{f(\phi)}_l-\alpha^0\widehat{\phi}_l-\widehat{G}_l^0(s), & n=0,\\ 2\widehat{f(u_M^n)}_l-2\alpha^n\widehat{(u_M^n)}_l-\widehat{G}_l^n(-s)-\widehat{G}_l^n(s), & 1\le n\le\frac{T}{\tau}-1.\end{cases} \tag{4.132}$$

Combining (4.117), (4.95) and (4.120)–(4.121) with $n=0$, noticing (4.84), under the condition (4.119),

$$0\le\alpha^0\le M_3,\quad \varepsilon^2\beta_l^0\ge1,\quad 0<\tau\beta_l^0\le\frac{\pi}{3},\quad \frac12\le\cos\left(\beta_l^0\tau\right)<1,$$

$$0\le\sin\left(\beta_l^0(\tau-s)\right)\le\sin\left(\beta_l^0\tau\right)<1,\quad 0\le s\le\tau.$$

From (4.131) with $n = 0$, using the Hölder inequality,

$$
\begin{aligned}
\left|\widehat{\xi}_l^0\right|^2 &= \left|\frac{1}{\varepsilon^2\beta_l^0}\int_0^\tau \widehat{W}_l^0(s)\sin\left(\beta_l^0(\tau-s)\right)\mathrm{d}s\right|^2 \\
&\le \int_0^\tau \sin\left(\beta_l^0(\tau-s)\right)\mathrm{d}s \cdot \int_0^\tau \left|\widehat{W}_l^0(s)\right|^2 \sin\left(\beta_l^0(\tau-s)\right)\mathrm{d}s \\
&\le \tau\left[1-\cos\left(\beta_l^0\tau\right)\right]\frac{\sin\left(\beta_l^0\tau\right)}{\beta_l^0\tau}\int_0^\tau \left|\widehat{W}_l^0(s)\right|^2 \mathrm{d}s \\
&\le \tau\left[1-\cos\left(\beta_l^0\tau\right)\right]\int_0^\tau \left|\widehat{W}_l^0(s)\right|^2 \mathrm{d}s. \qquad (4.133)
\end{aligned}
$$

Summing the above inequality for $l = 1, 2, \dots, M-1$, noticing (4.130) and (4.133),

$$
\begin{aligned}
\|e^1\|_{L^2}^2 &= \frac{b-a}{2}\sum_{l=1}^{M-1}\left|\widehat{e}_l^1\right|^2 = \frac{b-a}{2}\sum_{l=1}^{M-1}\left|\widehat{\xi}_l^0\right|^2 \\
&\le \frac{\tau(b-a)}{2}\sum_{l=1}^{M-1}\int_0^\tau \left|\widehat{W}_l^0(s)\right|^2 \mathrm{d}s.
\end{aligned}
$$

Plugging (4.132), (4.128) and (4.84) into the above inequality, using the triangle inequality and Parseval's equality,

$$
\begin{aligned}
\|e^1\|_{L^2}^2 &\le \frac{\tau(b-a)}{2}\sum_{l=1}^{M-1}\int_0^\tau \left|\widehat{(f(\phi))}_l - \widehat{(f(u))}_l(s) + \alpha^0\left(\widehat{u}_l(s)-\widehat{\phi}_l\right)\right|^2 \mathrm{d}s \\
&\le \tau(b-a)\int_0^\tau \sum_{l=1}^{M-1}\left[\left|\widehat{(f(\phi))}_l - \widehat{(f(u))}_l(s)\right|^2 + (\alpha^0)^2\left|\widehat{u}_l(s)-\widehat{\phi}_l\right|^2\right]\mathrm{d}s \\
&= \tau\int_0^\tau \left(\|\mathcal{P}_M\left[f(u(\cdot,s))-f(\phi)\right]\|_{L^2}^2 + (\alpha^0)^2\|\mathcal{P}_M\left[u(\cdot,s)-\phi\right]\|_{L^2}^2\right)\mathrm{d}s \\
&\le \tau\int_0^\tau \left(\|f(u(\cdot,s))-f(\phi)\|_{L^2}^2 + M_3^2\|u(\cdot,s)-\phi\|_{L^2}^2\right)\mathrm{d}s. \qquad (4.134)
\end{aligned}
$$

Under the assumption on u, using the Hölder inequality,

$$
\begin{aligned}
\|u(\cdot,s)-\phi\|_{L^2}^2 &= \int_a^b |u(x,s)-u(x,0)|^2\,\mathrm{d}x = \int_a^b\left|\int_0^s \partial_w u(x,w)\,\mathrm{d}w\right|^2 \mathrm{d}x \\
&\le \int_a^b s\int_0^s |\partial_w u(x,w)|^2\,\mathrm{d}w\,\mathrm{d}x = s\int_0^s \|\partial_w u(\cdot,w)\|_{L^2}^2\,\mathrm{d}w \\
&\le s^2\|\partial_t u(\cdot,t)\|_{L^\infty([0,T];L^2)}^2 \lesssim \frac{s^2}{\varepsilon^4}, \quad 0 \le s \le \tau. \qquad (4.135)
\end{aligned}
$$

Similarly, under the assumption on u and f,

$$
\|f\left(u(\cdot,s)\right)-f(\phi)\|_{L^2}^2 \le s^2M_2^2\|\partial_t u(\cdot,t)\|_{L^\infty([0,T];L^2)}^2 \lesssim \frac{s^2}{\varepsilon^4}, \quad 0 \le s \le \tau. \quad (4.136)
$$

Plugging (4.135) and (4.136) into (4.134), noticing (4.124) with $n = 1$, we obtain

$$\|e^1\|_{L^2}^2 \lesssim \tau \int_0^\tau \frac{s^2}{\varepsilon^4}\,ds \lesssim \frac{\tau^4}{\varepsilon^4} \lesssim \frac{\tau^4}{\varepsilon^8} \quad \Rightarrow \quad \|u(x,t_1) - u_M^1(x)\|_{L^2} \lesssim h^{m_0} + \frac{\tau^2}{\varepsilon^4}.$$

Similarly, one can get

$$\|\partial_x[u(x,t_1) - u_M^1(x)]\|_{L^2} \lesssim h^{m_0-1} + \frac{\tau^2}{\varepsilon^4}.$$

This, together with the triangle inequality and inverse inequality, implies

$$\begin{aligned}
\|u_M^1\|_{L^\infty} - M_1 &\le \|u_M^1\|_{L^\infty} - \|u(x,t_1)\|_{L^\infty} \le \|u_M^1 - u(x,t_1)\|_{L^\infty} \\
&\le \|\mathcal{P}_M u(x,t_1) - u(x,t_1)\|_{L^\infty} + \|u_M^1(x) - \mathcal{P}_M u(x,t_1)\|_{L^\infty} \\
&\lesssim \|u(x,t_1) - \mathcal{P}_M u(x,t_1)\|_{L^\infty} + \|u_M^1(x) - \mathcal{P}_M u(x,t_1)\|_{H^1} \\
&\lesssim h^{m_0-1} + \|e^1\|_{H^1} \\
&\lesssim h^{m_0-1} + \frac{\tau^2}{\varepsilon^4}. \qquad (4.137)
\end{aligned}$$

Thus under the assumption $\tau \lesssim \varepsilon^2$, there exist $h_2 > 0$ and $\tau_2 > 0$ sufficiently small and independent of ε, such that when $0 < h \le h_2$ and $0 < \tau \le \tau_2$,

$$\|u_M^1\|_{L^\infty} \le 1 + M_1.$$

Therefore, the three estimates in (4.120)–(4.121) are valid when $n = 1$.

Now, assume that (4.120)–(4.121) are valid for all $1 \le n \le m-1 \le \frac{T}{\tau} - 1$, then one needs to show that they are still valid when $n = m$. Denote

$$\mathcal{E}^n = \sum_{l=1}^{M-1} \widehat{\mathcal{E}}_l^n, \quad \widehat{\mathcal{E}}_l^n = \left|\widehat{e}_l^{n+1}\right|^2 + |\widehat{e}_l^n|^2 + \frac{\cos(\beta_l^n\tau)}{1-\cos(\beta_l^n\tau)}\left|\widehat{e}_l^{n+1} - \widehat{e}_l^n\right|^2. \qquad (4.138)$$

For each $l = 1, 2, \ldots, M-1$ and $1 \le n \le m-1$, noticing (4.84), under the condition (4.119),

$$\begin{aligned}
&0 \le \alpha^{n-1} \le \alpha^n \le M_3, \quad 1 \le \varepsilon^2\beta_l^{n-1} \le \varepsilon^2\beta_l^n, \quad 0 < \tau\beta_l^{n-1} \le \tau\beta_l^n \le \frac{\pi}{3}, \\
&\frac{1}{2} \le \cos\left(\beta_l^n\tau\right) \le \cos\left(\beta_l^{n-1}\tau\right) < 1, \quad \frac{\cos(\beta_l^n\tau)}{1-\cos(\beta_l^n\tau)} \le \frac{\cos(\beta_l^{n-1}\tau)}{1-\cos(\beta_l^{n-1}\tau)}, \\
&0 \le \sin\left(\beta_l^n(\tau - s)\right) \le \sin\left(\beta_l^n\tau\right) < 1, \quad 0 \le s \le \tau.
\end{aligned}$$

Then similar to (4.133),

$$\left|\widehat{\xi}_l^n\right|^2 \le \tau\left[1-\cos\left(\beta_l^n\tau\right)\right]\int_0^\tau \left|\widehat{W}_l^n(s)\right|^2 \mathrm{d}s.$$

Multiplying both sides of (4.129) by $\widehat{e}_l^{n+1}-\widehat{e}_l^{n-1}$ and dividing by $1-\cos(\beta_l^n\tau)$, one can have

$$\begin{aligned}
\widehat{\mathcal{E}}_l^n-\widehat{\mathcal{E}}_l^{n-1} &\le \frac{1}{1-\cos(\beta_l^n\tau)}\left|\widehat{\xi}_l^n\right|\cdot\left|\widehat{e}_l^{n+1}-\widehat{e}_l^{n-1}\right| \\
&\le \frac{1}{1-\cos(\beta_l^n\tau)}\left(2\tau\left|\widehat{e}_l^{n+1}-\widehat{e}_l^n\right|^2+2\tau\left|\widehat{e}_l^n-\widehat{e}_l^{n-1}\right|^2+\frac{1}{\tau}\left|\widehat{\xi}_l^n\right|^2\right) \\
&\le \frac{4\tau\cos(\beta_l^n\tau)}{1-\cos(\beta_l^n\tau)}\left(\left|\widehat{e}_l^{n+1}-\widehat{e}_l^n\right|^2+\left|\widehat{e}_l^n-\widehat{e}_l^{n-1}\right|^2\right)+\int_0^\tau\left|\widehat{W}_l^n(s)\right|^2\mathrm{d}s \\
&\le 4\tau\left(\widehat{\mathcal{E}}_l^n+\widehat{\mathcal{E}}_l^{n-1}\right)+\int_0^\tau\left|\widehat{W}_l^n(s)\right|^2\mathrm{d}s.
\end{aligned}$$

Summing the above inequality for $l=1,2,\ldots,M-1$,

$$\mathcal{E}^n-\mathcal{E}^{n-1}\le 4\tau\left(\mathcal{E}^n+\mathcal{E}^{n-1}\right)+\int_0^\tau\sum_{l=1}^{M-1}\left|\widehat{W}_l^n(s)\right|^2\mathrm{d}s,\quad 0\le n\le m-1.$$

Summing the above inequality for $n=1,2,\ldots,m-1$, when $\tau\le 1/8$,

$$\mathcal{E}^{m-1}\le 2\mathcal{E}^0+8\tau\sum_{n=0}^{m-2}\mathcal{E}^n+2\sum_{n=1}^{m-2}\int_0^\tau\sum_{l=1}^{M-1}\left|\widehat{W}_l^n(s)\right|^2\mathrm{d}s,\quad 2\le m\le\frac{T}{\tau}.$$

Using the discrete Gronwall's inequality,

$$\mathcal{E}^{m-1}\le C\left[\mathcal{E}^0+\sum_{n=1}^{m-1}\int_0^\tau\sum_{l=1}^{M-1}\left|\widehat{W}_l^n(s)\right|^2\mathrm{d}s\right],\quad 2\le m\le\frac{T}{\tau}, \tag{4.139}$$

where the constant C is independent of h (or l), τ (or m), and ε. Combining (4.139), (4.138) and (4.122), one can obtain

$$\begin{aligned}
\|e^m\|_{L^2}^2 &= \frac{b-a}{2}\sum_{l=1}^{M-1}|\widehat{e}_l^m|^2\le\frac{b-a}{2}\mathcal{E}^{m-1} \\
&\le \frac{C(b-a)}{2}\left[\mathcal{E}^0+\sum_{n=1}^{m-1}\int_0^\tau\sum_{l=1}^{M-1}\left|\widehat{W}_l^n(s)\right|^2\mathrm{d}s\right].
\end{aligned}\tag{4.140}$$

From (4.138) with $n=0$, noticing (4.130), (4.133), (4.134) (4.135) and (4.136),

$$\begin{aligned}
\mathcal{E}^0 &= \sum_{l=1}^{M-1}\frac{1}{1-\cos\left(\beta_l^0\tau\right)}|\widehat{e}_l^1|^2=\sum_{l=1}^{M-1}\frac{1}{1-\cos\left(\beta_l^0\tau\right)}|\widehat{\xi}_l^0|^2 \\
&= \tau\sum_{l=1}^{M-1}\int_0^\tau\left|\widehat{W}_l^0(s)\right|^2\mathrm{d}s\lesssim\frac{\tau^4}{\varepsilon^4}\lesssim\frac{\tau^4}{\varepsilon^8}.
\end{aligned}\tag{4.141}$$

From (4.132), (4.128) and (4.84), using the triangle inequality,

$$\begin{aligned}\sum_{l=1}^{M-1}\left|\widehat{W}_l^n(s)\right|^2 &= \sum_{l=1}^{M-1}\Big|2\widehat{f(u_M^n)}_l-\widehat{(f(u))}_l(t_n-s)-\widehat{(f(u))}_l(t_n+s)\\ &\quad +\alpha^n\left[\widehat{u}_l(t_n-s)+\widehat{u}_l(t_n+s)-2\widehat{(u_M^n)}_l\right]\Big|^2\mathrm{d}s\\ &\le \frac{2}{b-a}\|2f(u_M^n)-f(u(\cdot,t_n-s))-f(u(\cdot,t_n+s))\|_{L^2}^2\\ &\quad +\frac{2M_3^2}{b-a}\|u(\cdot,t_n-s)+u(\cdot,t_n+s)-2u_M^n\|_{L^2}^2. \qquad (4.142)\end{aligned}$$

Under the regularity assumption on u, using the triangle inequality and Hölder inequality, noticing (4.121),

$$\begin{aligned}&\|u(\cdot,t_n-s)+u(\cdot,t_n+s)-2u_M^n\|_{L^2}^2\\ &\le \|u(\cdot,t_n-s)+u(\cdot,t_n+s)-2u(\cdot,t_n)\|_{L^2}^2+4\|u(\cdot,t_n)-u_M^n\|_{L^2}^2\\ &\le \int_a^b\left|\int_0^s\int_{-w}^{w}\partial_{qq}u(x,t_n+q)\,\mathrm{d}q\,\mathrm{d}w\right|^2\mathrm{d}x+4\|u(\cdot,t_n)-u_M^n\|_{L^2}^2\\ &\le \int_0^s s\int_a^b\left|\int_{-w}^{w}\partial_{qq}u(x,t_n+q)\,\mathrm{d}q\right|^2\mathrm{d}x\,\mathrm{d}w+4\|u(\cdot,t_n)-u_M^n\|_{L^2}^2\\ &\le \int_0^s 2sw\int_{-w}^{w}\|\partial_{qq}u(\cdot,t_n+q)\|_{L^2}^2\,\mathrm{d}q\,\mathrm{d}w+4\|u(\cdot,t_n)-u_M^n\|_{L^2}^2\\ &\le \frac{4s^4}{3}\|\partial_{tt}u(\cdot,t)\|_{L^\infty([0,T];L^2)}^2+4\|u(\cdot,t_n)-u_M^n\|_{L^2}^2\\ &\lesssim \frac{\tau^4}{\varepsilon^8}+h^{2m_0}+\|e^n\|_{L^2}^2,\quad 0\le s\le\tau,\quad 1\le n\le m-1. \qquad (4.143)\end{aligned}$$

Similarly, under the assumption on u and f,

$$\begin{aligned}&\|f(u(\cdot,t_n-s))+f(u(\cdot,t_n+s))-2f(u_M^n)\|_{L^2}^2\\ &\le \frac{8s^4M_2^2}{3}\left[\|\partial_t u(\cdot,t)\|_{L^\infty([0,T];L^4)}^4+\|\partial_{tt}u(\cdot,t)\|_{L^\infty([0,T];L^2)}^2\right]\\ &\quad +4M_2^2\cdot\|u(\cdot,t_n)-u_M^n\|_{L^2}^2\\ &\lesssim \frac{\tau^4}{\varepsilon^8}+h^{2m_0}+\|e^n\|_{L^2}^2,\quad 0\le s\le\tau,\quad 1\le n\le m-1. \qquad (4.144)\end{aligned}$$

Plugging (4.144), (4.143), (4.142) and (4.141) into (4.140) leads to

$$\begin{aligned}\|e^m\|_{L^2}^2 &\lesssim \frac{\tau^4}{\varepsilon^8} + \tau(m-1)\left[\frac{\tau^4}{\varepsilon^8} + +h^{2m_0}\right] + \tau\sum_{n=1}^{m-1}\|e^n\|_{L^2}^2 \\ &\lesssim \frac{\tau^4}{\varepsilon^8} + T\left[\frac{\tau^4}{\varepsilon^8} + +h^{2m_0}\right] + \tau\sum_{n=1}^{m-1}\|e^n\|_{L^2}^2.\end{aligned}$$

Then, using the discrete Gronwall's inequality again, together with (4.124), one can get the first estimate in (4.120) for $n = m$. Similar to the above procedure by defining

$$\mathcal{S}^n = \sum_{l=1}^{M-1}\mu_l^2\widehat{\mathcal{E}}_l^n, \quad n \geq 0,$$

and noticing

$$\begin{aligned}\sum_{l=1}^{M-1}\mu_l^2\left|\widehat{W}_l^n(s)\right|^2 &\lesssim \|\partial_x\left[2f(u_M^n) - f(u(\cdot,t_n-s)) - f(u(\cdot,t_n+s))\right]\|_{L^2}^2 \\ &\quad + \|\partial_x\left[u(\cdot,t_n-s) + u(\cdot,t_n+s) - 2u_M^n\right]\|_{L^2}^2,\end{aligned}$$

one can obtain (4.121). In addition, similar to the proof towards (4.137),

$$\|u_M^m\|_{L^\infty} - M_1 \lesssim h^{m_0-1} + \frac{\tau^2}{\varepsilon^4}.$$

Again under the assumption $\tau \lesssim \varepsilon^2$, there exist $h_3 > 0$ and $\tau_3 > 0$ sufficiently small and independent of $2 \leq m \leq T/\tau$, such that when $0 < h \leq h_3$ and $0 < \tau \leq \tau_3$,

$$\|u_M^m\|_{L^\infty} \leq 1 + M_1.$$

Thus the second estimate in (4.120) is valid when $n = m$ too. Therefore, the proof of (4.120)–(4.121) is completed by the method of mathematical induction under the choice of $h_0 = \min\{h_1, h_2, h_3\}$ and $\tau_0 = \min\{1/8, \tau_2, \tau_3\}$. □

Similar to the proof in above, for Gautschi-FD with (4.95), assume that $f(\cdot) \in C^2(\mathbb{R})$, $u \in C^2([0,T];W^{1,\infty}) \cap C^1([0,T];W^{1,\infty}) \cap C([0,T];W^{5,\infty} \cap H_0^1)$, and

$$\begin{aligned}&\|u(x,t)\|_{L^\infty([0,T];W^{5,\infty})} \lesssim 1, \\ &\|\partial_t u(x,t)\|_{L^\infty([0,T];W^{1,\infty})} \lesssim \frac{1}{\varepsilon^2}, \quad \|\partial_{tt} u(x,t)\|_{L^\infty([0,T];W^{1,\infty})} \lesssim \frac{1}{\varepsilon^4},\end{aligned}$$

then one can prove the following error estimate for Gautschi-FD with (4.95),

Theorem 4.10. *Let u_j^n be the approximation obtained from Gautschi-FD with (4.95). Assume $\tau \lesssim \varepsilon^2$, under the above assumptions on exact solution u and the nonlinear function f, there exist $h_0 > 0$ and $\tau_0 > 0$ sufficiently small and independent of ε such that, for any $0 < \varepsilon \le 1$, when $0 < h \le h_0$ and $0 < \tau \le \tau_0$ and under the condition (4.119),*

$$\|e^n\|_{l^2} \lesssim \frac{\tau^2}{\varepsilon^4} + h^2, \quad \|\delta_x^+ e^n\|_{l^2} \lesssim \frac{\tau^2}{\varepsilon^4} + h^2, \quad 0 \le n \le \frac{T}{\tau}, \tag{4.145}$$

where

$$e^n = (e_0^n, e_1^n, \ldots, e_M^n)^T, \quad \textit{with} \quad e_j^n = u(x_j, t_n) - u_j^n, \quad j = 0, 1, \ldots, M, \quad n \ge 0.$$

Proof. Follow the analogous proofs of Theorems 4.9 and 4.8, and the details are omitted here for brevity. □

Based on Theorems 4.9 and 4.10, for the nonlinear KG equation in the nonrelativistic limit regime, the ε-scalability of Gautschi-SP and Gautschi-FD is

$$\tau = O\left(\sqrt{\delta}\varepsilon^2\right) = O(\varepsilon^2), \quad h \le O\left(\sqrt{\delta}\right) = O(1), \quad 0 < \varepsilon \ll 1. \tag{4.146}$$

Remark 4.3. *The estimates in Theorems 4.9 and 4.10 can be directly extended to 2D and 3D, without changing the convergence rates, by replacing the condition $\tau \lesssim \varepsilon^2$ by $\tau \lesssim \varepsilon^2\sqrt{C_d(h)}$ $(d = 2,\ 3)$, where $C_d(h)$ is given in Remark 4.1.*

Remark 4.4. *A final remark is made here that if one considers the periodic boundary conditions or homogeneous Neumann boundary conditions, similar numerical methods discussed in Section 4.2 and 4.3 can be easily designed. For example, one can readily construct Gautschi-type integrator Fourier pseudospectral discretization for periodic boundary conditions, and respectively, Gautschi-type integrator cosine pseudospectral method for homogeneous Neumann boundary conditions. Also, the numerical analysis results obtained in Section 4.2 and 4.3 remain valid in both cases.*

Table 4.1: Temporal discretization errors of Impt-EC-FD at time $t = 0.4$ in nonlinear case with $h = 1/128$ for different ε and τ under ε-scalability $\tau = O(\varepsilon^3)$: (i) l^2-error (upper 4 rows); (ii) discrete H^1-error (middle 4 rows); (iii) l^∞-error (lower 4 rows).

ε-scalability	τ =1.00E-3	τ =5.00E-4	τ =2.50E-4	τ =1.25E-4	τ =6.25E-5
$\varepsilon = 0.1,\ \tau$	4.6484E-2	1.1063E-2	2.7344E-3	6.8472E-4	1.7533E-4
$\varepsilon/2,\ \tau/2^3$	4.9171E-2	1.2912E-2	3.2712E-3	8.2486E-4	2.1197E-4
$\varepsilon/4,\ \tau/4^3$	4.6831E-2	1.1162E-2	2.7597E-3	6.9083E-4	1.7681E-4
$\varepsilon/8,\ \tau/8^3$	3.6900E-2	9.6406E-3	2.4426E-3	6.1784E-4	1.6129E-4
$\varepsilon = 0.1,\ \tau$	7.5093E-2	1.8650E-2	4.6877E-3	1.2087E-3	2.8179E-4
$\varepsilon/2,\ \tau/2^3$	9.2221E-2	2.3739E-2	6.0030E-3	1.5347E-3	3.8945E-4
$\varepsilon/4,\ \tau/4^3$	7.0780E-2	1.7431E-2	4.3724E-3	1.1292E-3	2.9825E-4
$\varepsilon/8,\ \tau/8^3$	7.7202E-2	1.9840E-2	5.0233E-3	1.2937E-3	3.1687E-4
$\varepsilon = 0.1,\ \tau$	2.9725E-2	7.7927E-3	1.8177E-3	4.5897E-4	1.2252E-4
$\varepsilon/2,\ \tau/2^3$	4.3783E-2	1.1543E-2	2.9273E-3	7.3938E-4	1.9031E-4
$\varepsilon/4,\ \tau/4^3$	2.8754E-2	6.9944E-3	1.7321E-3	4.2850E-4	1.1127E-4
$\varepsilon/8,\ \tau/8^3$	2.9213E-2	7.9193E-3	2.0227E-3	5.1313E-4	1.3350E-4

4.4 Numerical results

In this section, numerical results are reported to support the error estimates and demonstrate the superiority of Gautschi-type integrator over finite difference in time resolution when $0 < \varepsilon \ll 1$. In order to do so, in the KG equation (4.8)–(4.10), we choose

$$f(u) = \lambda u^{p+1}, \quad \phi(x) = \frac{2}{\mathrm{e}^{x^2} + \mathrm{e}^{-x^2}}, \quad \gamma(x) = 0, \quad x \in \mathbb{R}. \tag{4.147}$$

The computational interval $[a, b]$ is chosen large enough such that the homogeneous Dirichlet boundary conditions do not introduce a significant aliasing error relative to the problem in the whole space. Let $u(x, t)$ be the "exact" solution which is

Table 4.2: Temporal discretization errors of SImpt-FD at time $t = 0.4$ in nonlinear case with $h = 1/128$ for different ε and τ under ε-scalability $\tau = O(\varepsilon^3)$: (i) l^2-error (upper 4 rows); (ii) discrete H^1-error (middle 4 rows); (iii) l^∞-error (lower 4 rows).

ε-scalability	τ =1.00E-3	τ =5.00E-4	τ =2.50E-4	τ =1.25E-4	τ =6.25E-5
$\varepsilon = 0.1,\ \tau$	4.4395E-2	1.0598E-2	2.6213E-3	6.5674E-4	1.6847E-4
$\varepsilon/2,\ \tau/2^3$	4.8329E-2	1.2678E-2	3.2113E-3	8.0980E-4	2.0824E-4
$\varepsilon/4,\ \tau/4^3$	4.6690E-2	1.1131E-2	2.7521E-3	6.8896E-4	1.7635E-4
$\varepsilon/8,\ \tau/8^3$	3.6864E-2	9.6301E-3	2.4399E-3	6.1716E-4	1.6113E-4
$\varepsilon = 0.1,\ \tau$	7.2046E-2	1.7868E-2	4.4913E-3	1.1605E-3	2.9095E-4
$\varepsilon/2,\ \tau/2^3$	9.0650E-2	2.3316E-2	5.8955E-3	1.5080E-3	3.8325E-4
$\varepsilon/4,\ \tau/4^3$	7.0580E-2	1.7381E-2	4.3599E-3	1.1261E-3	2.9257E-4
$\varepsilon/8,\ \tau/8^3$	7.7123E-2	1.9818E-2	5.0178E-3	1.2923E-3	3.1356E-4
$\varepsilon = 0.1,\ \tau$	2.8663E-2	7.0385E-3	1.7558E-3	4.4435E-4	1.1932E-4
$\varepsilon/2,\ \tau/2^3$	4.2804E-2	1.1276E-2	2.8593E-3	7.2231E-4	1.8604E-4
$\varepsilon/4,\ \tau/4^3$	2.8674E-2	6.9762E-3	1.7277E-3	4.2753E-4	1.1111E-4
$\varepsilon/8,\ \tau/8^3$	2.9168E-2	7.9066E-3	2.0194E-3	5.1231E-4	1.3329E-4

obtained numerically by using Gautschi-SP with very fine mesh size and small time step, e.g. $h = 1/1024$ and $\tau =$ 1E-8. In order to quantify the convergence, we define three error functions, l^2-error, l^∞-error and discrete H^1-error as

$$e_{l^2} = \|u(\cdot, t_n) - u^n\|_{l^2}, \quad e_{l^\infty} = \max_j |u(x_j, t_n) - u_j^n|, \tag{4.148}$$

$$e_{H^1} = \sqrt{\|u(\cdot, t_n) - u^n\|_{l^2}^2 + \|\delta_x^+ (u(\cdot, t_n) - u^n)\|_{l^2}^2}. \tag{4.149}$$

CASE I. A nonlinear case, where we choose $\lambda = 4$ and $p = 2$ in (4.147) and solve the KG equation (4.8)–(4.10) on the interval $[-8, 8]$. In order to study the temporal resolution or ε-scalability in time of different methods, a very small mesh size $h = 1/128$ is chosen such that the discretization error in space is negligible.

Table 4.3: Temporal discretization errors of Gautschi-SP at time $t = 0.4$ in nonlinear case with $h = 1/128$ for different ε and τ under ε-scalibility $\tau = O(\varepsilon^2)$: (i) l^2-error (upper 4 rows); (ii) discrete H^1-error (middle 4 rows); (iii) l^∞-error (lower 4 rows).

ε-scalability	τ =5.00E-3	τ =2.50E-3	τ =1.25E-3	τ =6.25E-4	τ =3.125E-4
$\varepsilon = 0.1,\ \tau$	2.4902E-3	6.1124E-4	1.5208E-4	3.7957E-5	9.4697E-6
$\varepsilon/2,\ \tau/2^2$	3.1009E-3	7.6212E-4	1.8973E-4	4.7384E-5	1.1845E-5
$\varepsilon/4,\ \tau/4^2$	2.5929E-3	6.3666E-4	1.5846E-4	3.9564E-5	9.8826E-6
$\varepsilon/8,\ \tau/8^2$	2.5965E-3	6.3757E-4	1.5862E-4	3.9563E-5	9.8072E-6
$\varepsilon = 0.1,\ \tau$	6.0409E-3	1.4857E-3	3.6976E-4	9.2230E-5	2.2948E-5
$\varepsilon/2,\ \tau/2^2$	8.6467E-3	2.1232E-3	5.2845E-4	1.3197E-4	3.2989E-5
$\varepsilon/4,\ \tau/4^2$	6.3003E-3	1.5450E-3	3.8453E-4	9.6000E-5	2.3974E-5
$\varepsilon/8,\ \tau/8^2$	7.9670E-3	1.9557E-3	4.8650E-4	1.2126E-4	3.0079E-5
$\varepsilon = 0.1,\ \tau$	1.9268E-3	4.7365E-4	1.1786E-4	2.9447E-5	7.3746E-6
$\varepsilon/2,\ \tau/2^2$	2.4770E-3	6.0895E-4	1.5161E-4	3.7863E-5	9.4650E-6
$\varepsilon/4,\ \tau/4^2$	1.9261E-3	4.7358E-4	1.1797E-4	2.9445E-5	7.3572E-6
$\varepsilon/8,\ \tau/8^2$	1.9103E-3	4.6947E-4	1.1682E-4	2.9120E-5	7.2235E-6

Tabs. 4.1 and 4.2 tabulate l^2-error, H^1-error and l^∞-error at time $t = 0.4$ of Impt-EC-FD and SImpt-FD, respectively, for various time steps τ and parameter values ε under ε-scalability $\tau = O(\varepsilon^3)$. Tabs. 4.3 and 4.4 show similar results for Gautschi-SP and Gautschi-FD, under ε-scalability $\tau = O(\varepsilon^2)$. Similarly, in order to compare errors of spatial discretization, we always choose very fine time step τ such that time discretization error is negligible. Tab. 4.5 lists l^2-errors at time $t = 0.4$ of Impt-EC-FD, SImpt-EC-FD, Gautschi-FD and Gautschi-SP with different ε and τ satisfying the required ε-scalability. Numerical experiments are also carried out for Impt-EC-FD and Expt-FD, where the results are similar to those of Impt-EC-FD and SImpt-FD, and thus we omit them for brevity.

Table 4.4: Temporal discretization errors of Gautschi-FD at time $t = 0.4$ in nonlinear case with $h = 1/128$ for different ε and τ under ε-scalability $\tau = O(\varepsilon^2)$: (i) l^2-error (upper 4 rows); (ii) discrete H^1-error (middle 4 rows); (iii) l^∞-error (lower 4 rows).

ε-scalability	τ =5.00E-3	τ =2.50E-3	τ =1.25E-3	τ =6.25E-4	τ =3.125E-4
$\varepsilon = 0.1,\ \tau$	2.4910E-3	6.1204E-4	1.5295E-4	3.9100E-5	9.9575E-6
$\varepsilon/2,\ \tau/2^2$	3.1013E-3	7.6248E-4	1.9017E-4	4.8121E-5	1.2658E-5
$\varepsilon/4,\ \tau/4^2$	2.5937E-3	6.3748E-4	1.5836E-4	4.0818E-5	1.0343E-6
$\varepsilon/8,\ \tau/8^2$	2.6106E-3	6.3814E-4	1.5927E-4	4.0565E-5	9.9140E-6
$\varepsilon = 0.1,\ \tau$	6.0467E-3	1.4916E-3	3.7655E-4	9.9249E-5	2.3124E-5
$\varepsilon/2,\ \tau/2^2$	8.6502E-3	2.1268E-3	5.3291E-4	1.3656E-4	3.6826E-5
$\varepsilon/4,\ \tau/4^2$	6.3067E-3	1.5698E-3	3.9225E-4	1.0798E-4	2.4152E-5
$\varepsilon/8,\ \tau/8^2$	7.8831E-3	1.9601E-3	4.9192E-4	1.2936E-4	3.7448E-5
$\varepsilon = 0.1,\ \tau$	1.9254E-3	4.7230E-4	1.1654E-4	2.8122E-5	7.5647E-6
$\varepsilon/2,\ \tau/2^2$	2.4755E-3	6.0746E-4	1.5013E-4	3.6437E-5	8.8545E-6
$\varepsilon/4,\ \tau/4^2$	1.9247E-3	4.7285E-4	1.1662E-4	2.9170E-5	7.6260E-6
$\varepsilon/8,\ \tau/8^2$	1.9340E-3	4.6890E-4	1.1568E-4	2.7885E-5	7.5793E-6

CASE II A linear case, where we choose $\lambda = 4$ and $p = 0$ in (4.147) and solve the KG equation (4.8)–(4.10) on the interval $[-16, 16]$. Here, only the results of Gautschi-SP and Gautschi-FD are presented to verify that there is no time discretization error of Gautschi-type integrator for the linear KG equation. Tab. 4.6 lists the l^2-error of Gautschi-SP and Gautschi-FD at time $t = 1$ for different τ, h and ε under the ε-scalability $\tau = O(1)$ and $h = O(1)$. Similar convergence patterns of the discrete H^1-error and l^∞-error were also observed and they are omitted here for simplicity. In addition, the results for FDTD methods are quite similar to those in nonlinear case and thus are omitted here too for brevity.

From Tabs. 4.1–4.6 and extensive numerical results not shown here for brevity,

Table 4.5: Spatial discretization error e_{l^2} of Impt-EC-FD/SImpt-FD (under ε-scalability $\tau = O(\varepsilon^3)$), and Gautschi-FD/Gautschi-SP (under ε-scalability $\tau = O(\varepsilon^2)$) at time $t = 0.4$ in nonlinear case with $\varepsilon_0 = 0.1$ and τ_0=2E-5 for different mesh sizes h.

		$h = 1/4$	$h = 1/8$	$h = 1/16$	$h = 1/32$
Impt-EC-FD	$\varepsilon_0,\ \tau_0$	2.0671E-2	5.5497E-3	1.4075E-3	3.5551E-4
	$\varepsilon_0/2,\ \tau_0/2^3$	2.2900E-2	6.2179E-3	1.5834E-3	4.0110E-4
	$\varepsilon_0/4,\ \tau_0/4^3$	2.2881E-2	6.2815E-3	1.6021E-3	4.0398E-4
SImpt-FD	$\varepsilon_0,\ \tau_0$	2.0671E-2	5.5497E-3	1.4075E-3	3.5541E-4
	$\varepsilon_0/2,\ \tau_0/2^3$	2.2900E-2	6.2178E-3	1.5833E-3	4.0101E-4
	$\varepsilon_0/4,\ \tau_0/4^3$	2.2881E-2	6.2815E-3	1.6021E-3	4.0398E-4
Gautschi-FD	$\varepsilon_0,\ \tau_0$	2.0668E-2	5.5462E-3	1.4041E-3	3.5182E-4
	$\varepsilon_0/2,\ \tau_0/2^2$	2.2894E-2	6.2129E-3	1.5784E-3	3.9568E-4
	$\varepsilon_0/4,\ \tau_0/4^2$	2.2878E-2	6.2790E-3	1.5996E-3	4.0120E-4
		$h = 1$	$h = 1/2$	$h = 1/4$	$h = 1/8$
Gautschi-SP	$\varepsilon_0,\ \tau_0$	1.1873E-1	3.9320E-3	3.1799E-5	1.0722E-7
	$\varepsilon_0/2,\ \tau_0/2^2$	8.3243E-2	3.2486E-3	3.3677E-5	7.6844E-8
	$\varepsilon_0/4,\ \tau_0/4^2$	1.1899E-1	3.9849E-3	2.8723E-5	8.4444E-8

we can draw the following conclusions:

(i). In the $O(1)$-speed of light regime, i.e. $0 < \varepsilon = O(1)$ fixed, the FDTD methods and Gautschi-FD are of second-order accuracy in both time and space (cf. Tabs. 4.1, 4.2, 4.4 and 4.5); and Gautschi-SP is second-order and spectral-order accurate in time and space, respectively (cf. Tabs. 4.3 and 4.5). In addition, there is no time discretization error of Gautschi-SP and Gautschi-FD for the linear KG equation (cf. Tab. 4.6). Therefore, in this regime all

Table 4.6: Temporal and spatial discretization error e_{l^2} of Gautschi-SP and Gautschi-FD in linear case at time $t = 1$ with $\tau_0 = 0.25$ and $h_0 = 0.5$ for different τ, h and ε.

			$\varepsilon = 0.02$	$\varepsilon = 0.002$	$\varepsilon = 0.0002$	$\varepsilon = 0.00002$
Gautschi-SP	τ_0	$h_0/8$	1.1239E-15	9.7781E-16	1.9602E-15	1.6371E-15
	$\tau_0/2$	$h_0/8$	1.2503E-15	1.6460E-15	1.3867E-15	1.6133E-15
	$\tau_0/4$	$h_0/8$	2.8930E-15	2.2077E-15	2.5100E-15	2.4671E-15
	τ_0	h_0	3.9029E-3	5.5134E-3	4.1445E-3	5.5276E-3
	τ_0	$h_0/2$	1.1041E-5	1.2214E-5	8.6830E-6	1.2093E-5
	τ_0	$h_0/4$	4.1894E-10	5.2825E-10	5.2296E-10	4.8898E-10
Gautschi-FD	τ_0	$h_0/32$	2.2447E-4	2.2633E-4	2.2723E-4	2.2641E-4
	$\tau_0/2$	$h_0/32$	2.2447E-4	2.2633E-4	2.2723E-4	2.2641E-4
	$\tau_0/4$	$h_0/32$	2.2447E-4	2.2633E-4	2.2723E-4	2.2641E-4
	τ_0	$h_0/4$	1.3608E-2	1.3703E-2	1.3765E-2	1.3708E-2
	τ_0	$h_0/8$	3.5636E-3	3.5923E-3	3.6069E-3	3.5934E-3
	τ_0	$h_0/16$	8.9699E-4	9.0441E-4	9.0802E-4	9.0468E-4

the methods considered are compatible in time discretization and Gautschi-SP is of higher accuracy in space than the rest. Indeed, generally Gautschi-SP performs much better in time discretization than the rest under the same time step and mesh size.

(ii). In the nonrelativistic limit regime, i.e. $0 < \varepsilon \ll 1$, for FDTD methods the "correct" ε-scalability is $\tau = O(\varepsilon^3)$ and $h = O(1)$ which confirms the analytical results (4.41); and, for Gautschi-SP and Gautschi-FD methods, the "correct" ε-scalability is $\tau = O(1)$ and $h = O(1)$ for the linear KG equation which verifies the analytical results (4.116), and respectively, $\tau = O(\varepsilon^2)$ and $h = O(1)$ for the nonlinear KG equation which again confirms the analytical results (4.146).

In view of both temporal and spatial resolution capacities, one can conclude that Gautschi-SP is the best candidate for discretizing the KG equation, especially in the nonrelativistic limit regime.

Chapter 5

Comparisons between sine–Gordon & perturbed NLS equations

This chapter is devoted to extensive numerical comparisons among the 2D *light bullets* solutions of the sine–Gordon equation (1.14), the perturbed NLS equation (1.16) and the critical cubic focusing NLS equation ($\varepsilon = 0$ in (1.16)). To this purpose, efficient numerical methods are proposed, for which rigorous error estimates are also carried out.

5.1 Sine–Gordon, perturbed NLS and their approximations

The propagation and interaction of spatially localized optical pulses (so-called *light bullets* (LBs)) with particle features in several space dimensions are of both physical and mathematical interests [6, 79]. They have been found useful as information carriers in communication [82, 119], as energy sources, switches and logic gates in optical devices [112]. Such LBs have been observed in numerical simulations of the full Maxwell system with instantaneous Kerr ($\chi^{(3)}$ or cubic) nonlinearity in 2D [70]. They are short femtosecond pulses that propagate without essentially changing shapes over a long distance and have only a few EM (electromagnetic) oscillations under their envelopes [70, 81, 82, 127, 149].

In 1D, the Maxwell system modeling light propagation in nonlinear media admits constant-speed traveling waves as exact solutions, also known as the *light bubbles* (unipolar pulses or solitons), [28,34,90,91]. The complete integrability of a Maxwell–Bloch system is shown in [7]. In several space dimensions, constant-speed traveling waves (mono-scale solutions) are harder to come by. Instead, space-time oscillating (multiple-scale) solutions are more robust [149]. The so-called LBs are of multiple-scale structures with distinct phase/group velocities and amplitude dynamics. Even though direct numerical simulations of the full Maxwell system are motivating [70], asymptotic approximation is necessary for analysis in several space dimensions [149]. The approximation of 1D Maxwell system has been extensively studied. Long pulses are well approximated via envelope approximation by the cubic focusing nonlinear Schrödinger (NLS) for $\chi^{(3)}$ medium [119]. A comparison between Maxwell solutions and those of an extended NLS [81, 82, 89] also showed that the cubic NLS approximation works reasonably well on short stable 1D pulses. Mathematical analysis on the validity of NLS approximation of pulses and counter-propagating pulses of 1D sine–Gordon equation has been carried out [93, 125]. However, in 2D, the envelope approximation with the cubic focusing NLS breaks down [8], because critical collapse of the cubic focusing NLS occurs in finite time ([33, 38, 60, 61, 137, 142] and references therein). On the other hand, due to the intrinsic physical mechanism or material response, Maxwell system itself typically behaves fine beyond the cubic NLS collapse time. One example is the semi-classical two level dissipationless Maxwell–Bloch system where smooth solutions persist forever [50]. It is thus a very interesting question how to modify the cubic NLS approximation to capture the correct physics for modeling the propagation and interaction of light signals in 2D Maxwell type systems. One approach will be discussed in the following.

Considering the transverse electric regime, after taking a distinguished asymptotic limit of the two level dissipationless Maxwell–Bloch system studied in [70], Xin [149] found that the well-known sine-Gordon (SG) equation (1.14)–(1.15) also admits 2D LBs solutions. In the SG equation (1.14)–(1.15), it is well-known that the *energy*

$$E^{\mathrm{SG}}(t) := \int_{\mathbb{R}^2} \left[(\partial_t u)^2 + c^2 \left|\nabla u\right|^2 + 2G(u)\right] \mathrm{d}\mathbf{x}, \quad t \geq 0, \tag{5.1}$$

with

$$G(u) = \int_0^u \sin(s)\mathrm{d}s = 1 - \cos(u), \tag{5.2}$$

is conserved. Direct numerical simulations of the SG equation in 2D were performed in [127,149], which are much simpler tasks than simulating the full Maxwell system. Moving pulse solutions being able to keep the overall profile over a long time were observed, just like those in Maxwell system [70,81,82,127,149]. See also [113,114] for related breather-type solutions of the SG equation in 2D based on a modulation analysis in the Lagrangian formulation.

Also, as derived in [149], with the SG-LBs as starting point one can look for a modulated planar pulse solution of the SG equation (1.14) in the form:

$$u(\mathbf{x},t) = \varepsilon A(\varepsilon(x-\nu t), \varepsilon y, \varepsilon^2 t)\mathrm{e}^{i(kx-\omega(k)t)} + \mathrm{c.c.} + \varepsilon^3 u_2, \quad \mathbf{x} = (x,y) \in \mathbb{R}^2,\ t \geq 0, \tag{5.3}$$

where $0 < \varepsilon \ll 1$, $\omega = \omega(k) = \sqrt{1+c^2k^2}$, $\nu = \omega'(k) = c^2k/\omega$, the group velocity, and c.c. refers to the complex conjugate of the previous term. Plugging (5.3) into (1.14), setting $X = \varepsilon(x-\nu t)$, $Y = \varepsilon y$ and $T = \varepsilon^2 t$, calculating derivatives, expressing the sine function in series and removing all the resonance terms, one can obtain the following complete perturbed NLS equation (see details in [149]):

$$\begin{aligned} -2i\omega\partial_T A + \varepsilon^2\partial_{TT} A = {} & \frac{c^2}{\omega^2}\partial_{XX} A + c^2\partial_{YY} A + 2\varepsilon\nu\partial_{XT} A \\ & + |A|^2 A \sum_{l=0}^{\infty} \frac{(-1)^l(\varepsilon|A|)^{2l}}{(l+1)!(l+2)!}, \quad T > 0, \end{aligned} \tag{5.4}$$

where $A := A(\mathbf{X}, T)$, $\mathbf{X} = (X,Y) \in \mathbb{R}^2$, is a complex-valued function. This new equation is second order in space-time and contains a nonparaxiality term, a mixed derivative term, and a novel nonlinear term which is saturating for large amplitude.

Introducing the scaling variables $\widetilde{X} = (\omega/c)X$, $\widetilde{Y} = Y/c$ and $\widetilde{T} = T/(2\omega)$, substituting them into (5.4) and then removing all $\widetilde{\ }$, one gets a standard perturbed NLS equation, as already introduced in Section 1.2,

$$i\partial_T A - \frac{\varepsilon^2}{4\omega^2}\partial_{TT}A = -\Delta A - \frac{\varepsilon ck}{\omega}\partial_{XT}A + f_\varepsilon\big(|A|^2\big)A, \qquad T > 0, \tag{5.5}$$

with initial conditions,

$$A(\mathbf{X}, 0) = A^{(0)}(\mathbf{X}), \quad \partial_T A(\mathbf{X}, 0) = A^{(1)}(\mathbf{X}), \quad \mathbf{X} \in \mathbb{R}^2, \tag{5.6}$$

where,

$$\rho = |A|^2, \quad f_\varepsilon(\rho) = \sum_{l=0}^{\infty} \frac{(-1)^{l+1}\varepsilon^{2l}\rho^{l+1}}{(l+1)!(l+2)!}. \tag{5.7}$$

In fact, equation (5.5) can be viewed as a perturbed cubic NLS equation with both a saturating nonlinearity (series) term and nonparaxial terms (the A_{TT} and A_{XT} terms). As proven in [149], it conserves the *energy*, i.e.,

$$E^{\text{PNLS}}(T) := \int_{\mathbb{R}^2} \left[\frac{\varepsilon^2}{4\omega^2}\left|A_T\right|^2 + |\nabla A|^2 + F_\varepsilon\left(|A|^2\right)\right] \mathrm{d}\mathbf{X} \equiv E^{\text{PNLS}}(0),\ T \geq 0, \tag{5.8}$$

with

$$F_\varepsilon(\rho) = \int_0^\rho f_\varepsilon(s)\,\mathrm{d}s = \sum_{l=0}^{\infty} \frac{(-1)^{l+1}\varepsilon^{2l}\rho^{l+2}}{(l+1)!(l+2)!(l+2)}, \tag{5.9}$$

and has the *mass* balance identity

$$\frac{\mathrm{d}}{\mathrm{d}T}\left(\int_{\mathbb{R}^2} |A|^2\,\mathrm{d}\mathbf{X} - \frac{\varepsilon^2}{2\omega^2}\mathrm{Im}\int_{\mathbb{R}^2} A_T A^*\,\mathrm{d}\mathbf{X}\right) = \frac{2\varepsilon\nu}{c}\mathrm{Im}\int A_X {A_T}^*\,\mathrm{d}\mathbf{X}. \tag{5.10}$$

In addition, the perturbed NLS equation (5.5) is globally well-posed and does not have finite-time collapse [149], i.e., for any given initial data $A^{(0)}(\mathbf{X}) \in H^2(\mathbb{R}^2)$ and $A^{(1)}(\mathbf{X}) \in H^1(\mathbb{R}^2)$, the initial value problem of (5.5) with initial conditions (5.6) has a unique global solution $A \in C([0,\infty]; H^2(\mathbb{R}^2))$, $A_T \in C([0,\infty]; H^1(\mathbb{R}^2))$, and $A_{TT} \in C([0,\infty]; L^2(\mathbb{R}^2))$.

In practice, the infinite series of nonlinearity in (5.5) could be truncated to finite terms with focusing-defocusing cycles. Denote

$$f_\varepsilon^N(\rho) = \sum_{l=0}^{N} \frac{\varepsilon^{4l}\rho^{2l+1}}{(2l+1)!(2l+2)!}\left[-1 + \frac{\varepsilon^2\rho}{(2l+2)(2l+3)}\right], \tag{5.11}$$

then the perturbed NLS equation (5.5) can be approximated by the following truncated NLS equation:

$$i\partial_T A - \frac{\varepsilon^2}{4\omega^2}\partial_{TT}A = -\Delta A - \frac{\varepsilon c k}{\omega}\partial_{XT}A + f_\varepsilon^N\left(|A|^2\right)A, \qquad T > 0. \tag{5.12}$$

Similar to the proof in [149] for the perturbed NLS equation (5.5), one can show that the truncated NLS equation (5.12) with the initial conditions (5.6) also conserves the *energy*, i.e.,

$$E_N^{\text{PNLS}}(T) := \int_{\mathbb{R}^2}\left[\frac{\varepsilon^2}{4\omega^2}\left|A_T\right|^2 + \left|\nabla A\right|^2 + F_\varepsilon^N\left(|A|^2\right)\right]d\mathbf{X} \equiv E_N^{\text{PNLS}}(0),\ T \geq 0, \tag{5.13}$$

with

$$F_\varepsilon^N(\rho) = \int_0^\rho f_\varepsilon^N(s)\,ds = \sum_{l=0}^N \frac{\varepsilon^{4l}\rho^{2l+2}}{(2l+1)!(2l+2)!(2l+2)}\left[-1 + \frac{\varepsilon^2\rho}{(2l+3)^2}\right], \tag{5.14}$$

and has the *mass* balance identity (5.10).

When $\varepsilon = 0$, the perturbed NLS equation (5.5) and its approximation (5.12) collapse to the well-known critical cubic focusing NLS equation:

$$i\partial_T A = -\Delta A - \frac{1}{2}|A|^2 A, \quad T > 0, \tag{5.15}$$

with initial condition,

$$A(\mathbf{X}, 0) = A^{(0)}(\mathbf{X}), \quad \mathbf{X} \in \mathbb{R}^2. \tag{5.16}$$

It is well-known that this cubic NLS equation conserves the energy, i.e.,

$$E^{\text{CNLS}}(T) := \int_{\mathbb{R}^2}\left[|\nabla A|^2 - \frac{1}{4}\left|A\right|^4\right]d\mathbf{X} \equiv \int_{\mathbb{R}^2}\left[\left|\nabla A^{(0)}\right|^2 - \frac{1}{4}\left|A^{(0)}\right|^4\right]d\mathbf{X}, \tag{5.17}$$

and collapses in finite-time when the initial energy $E^{\text{CNLS}}(0) < 0$ [33, 38, 142], which motivates different choices of initial data in (5.16) and (5.6) for numerical experiments.

By closing this section, it would be desired to point out some numerical challenges in order to perform extensive comparisons among the LBs solutions of the SG equation, perturbed NLS and critical focusing NLS equations. The computation challenge involved in SG simulation is that the disparate time scales between the SG and perturbed NLS equations require a long-time simulation of the SG equation. To illustrate this, noting (5.3), the disparate time scales for the perturbed NLS equation (5.12) and the SG equation (1.14) are $T = O(1)$ and $t = O(\varepsilon^{-2})$, respectively, which immediately implies that it requires a much-longer-time simulation for the SG equation (1.14) if the time regime beyond the collapse time of the critical NLS equation (5.15) is of interest, when ε is small. Also, the computation domain for SG simulation needs to be extended if the interested time point turns out to be further away due to the propagating property of the SG-LBs (cf. (5.3)). On the other hand, for perturbed NLS simulation the challenge is that high spatial resolution is required to capture the focusing-defocusing mechanism which prevents the critical NLS collapse. In what follows, in order to balance the stability and efficiency, instead of using those fully implicit conservative methods [1, 2, 48, 75, 130, 141], semi-implicit sine pseudospectral discretizations are proposed, which can be explicitly solved in phase space and are of spectral order accuracy in space.

5.2 Numerical methods for SG and perturbed NLS equations

Since the finite-time propagation of the LBs is of interests in its right, noting the inherent far-field vanishing property of the LBs solutions of the SG and NLS equations, in practice, one can always truncate the whole space problems on a bounded computational domain Ω, e.g. $\Omega = [a, b] \times [c, d]$, with homogeneous Dirichlet boundary conditions, i.e., consider

$$\partial_{tt} u - c^2 \Delta u + \sin(u) = 0, \quad \mathbf{x} \in \Omega, \quad t > 0, \tag{5.18}$$

$$u(\mathbf{x}, 0) = u^{(0)}(\mathbf{x}), \quad \partial_t u(\mathbf{x}, 0) = u^{(1)}(\mathbf{x}), \quad u(\mathbf{x}, t)|_{\partial\Omega} = 0, \quad t \geq 0. \tag{5.19}$$

and a similar initial-boundary-value problem for the truncated perturbed NLS equation (5.12).

Let $\Delta t > 0$ be the time step and denote time steps as $t_n = n\Delta t$, $n = 0, 1, \dots$; choose spatial mesh sizes $\Delta x = \frac{b-a}{J}$ and $\Delta y = \frac{d-c}{K}$ with J, K being two positive even integers, and denote the grid points be

$$x_j := a + j\Delta x, \quad j = 0, 1, \dots, J; \quad y_k := c + k\Delta y, \quad k = 0, 1, \dots, K.$$

Let

$$Y_{JK} = \text{span}\left\{\phi_{lm}(\mathbf{x}),\ l = 1, 2, \dots, J-1,\ m = 1, 2, \dots, K-1\right\},$$

where

$$\phi_{lm}(\mathbf{x}) := \sin\left(\mu_l(x-a)\right)\sin\left(\lambda_m(y-c)\right), \quad \mathbf{x} = (x, y) \in \mathbb{R}^2,$$
$$\mu_l = \pi l/(b-a), \quad \lambda_m = \pi m/(d-c), \quad l = 1, 2, \dots, J-1, \quad m = 1, 2, \dots, K-1.$$

For a function $\xi(\mathbf{x}) \in L_0^2(\Omega) = \{v(\mathbf{x}) \mid v \in L^2(\Omega),\ v|_{\partial\Omega} = 0\}$ and a matrix $\varphi := \{\varphi_{jk}\}_{j,k=0}^{J,K} \in \mathbb{C}_0^{(J+1)(K+1)} = \{w \in \mathbb{C}^{(J+1)(K+1)} \mid w_{0k} = w_{Jk} = w_{j0} = w_{jK} = 0,\ j = 0, 1, \dots, J,\ k = 0, 1, \dots, K\}$, denote $\mathcal{P}_{JK} : L_0^2(\Omega) \to Y_{JK}$ and $\mathcal{I}_{JK} : \mathbb{C}_0^{(J+1)(K+1)} \to Y_{JK}$ be the standard projection and trigonometric interpolation operators [133,148], respectively, i.e.,

$$(\mathcal{P}_{JK}\xi)(\mathbf{x}) = \sum_{l=1}^{J-1}\sum_{m=1}^{K-1}\widehat{\xi}_{lm}\phi_{lm}(\mathbf{x}), \quad (\mathcal{I}_{JK}\varphi)(\mathbf{x}) = \sum_{l=1}^{J-1}\sum_{m=1}^{K-1}\widetilde{\varphi}_{lm}\phi_{lm}(\mathbf{x}), \quad \mathbf{x} \in \Omega, \tag{5.20}$$

where

$$\widehat{\xi}_{lm} = \frac{4}{(b-a)(d-c)}\int_\Omega \xi(\mathbf{x})\phi_{lm}(\mathbf{x})\mathrm{d}\mathbf{x}, \quad \widetilde{\varphi}_{lm} = \frac{4}{JK}\sum_{j=1}^{J-1}\sum_{k=1}^{K-1}\varphi_{jk}\phi_{lm}(x_j, y_k), \tag{5.21}$$

$$\widetilde{\xi}_{lm} = \frac{4}{JK}\sum_{j=1}^{J-1}\sum_{k=1}^{K-1}\xi(x_j, y_k)\phi_{lm}(x_j, y_k), \quad l = 1, \dots, J-1,\ m = 1, \dots, K-1. \tag{5.22}$$

5.2.1 Method for the SG equation

A semi-implicit sine pseudospectral method is discussed here for solving the SG equation. Let $u_{JK}^n(\mathbf{x})$ be the approximation of $u(\mathbf{x}, t_n)$ ($\mathbf{x} \in \Omega$), and respectively, u_{jk}^n

be the approximation of $u(x_j, y_k, t_n)$ ($j = 0, 1, \ldots, J,\ k = 0, 1, \ldots, K$) and denote u^n be the matrix with components u^n_{jk} at time $t = t_n$. Choose $u^0_{JK}(\mathbf{x}) = \mathcal{P}_{JK}(u^{(0)})$ for $\mathbf{x} \in \Omega$, by applying the sine spectral method for spatial derivatives, and second-order implicit and explicit schemes for linear and nonlinear terms respectively in time discretization for the SG equation (5.18), one can get the semi-implicit sine spectral discretization as:

Find $u^{n+1}_{JK}(\mathbf{x}) \in Y_{JK}$, i.e.,

$$u^{n+1}_{JK}(\mathbf{x}) = \sum_{l=1}^{J-1}\sum_{m=1}^{K-1} \widehat{(u^{n+1}_{JK})}_{lm}\phi_{lm}(\mathbf{x}), \quad \mathbf{x} \in \Omega, \quad n \geq 0, \tag{5.23}$$

such that for $\mathbf{x} \in \Omega$ and $n \geq 1$,

$$\frac{u^{n+1}_{JK} - 2u^n_{JK} + u^{n-1}_{JK}}{(\Delta t)^2} - \frac{c^2}{2}\left(\Delta u^{n+1}_{JK} + \Delta u^{n-1}_{JK}\right) + \mathcal{P}_{JK}\left(\sin(u^n_{JK})\right) = 0, \tag{5.24}$$

and the initial data in (5.19) is discretized as

$$\frac{u^1_{JK} - u^0_{JK}}{\Delta t} = \mathcal{P}_{JK}(u^{(1)}) + \frac{\Delta t}{2}\left[c^2\Delta u^0_{JK} - \mathcal{P}_{JK}\left(\sin(u^{(0)})\right)\right]. \tag{5.25}$$

Plugging (5.23) into (5.25) and (5.24) and noticing the orthogonality of sine functions, for $l = 1, 2, \ldots, J-1$ and $m = 1, 2, \ldots, K-1$,

$$\widehat{(u^{n+1}_{JK})}_{lm} = \begin{cases} \left[1 - \frac{c^2}{2}(\Delta t)^2\left(\mu_l^2 + \lambda_m^2\right)\right]\widehat{(u^{(0)})}_{lm} + \Delta t\,\widehat{(u^{(1)})}_{lm} - \frac{(\Delta t)^2}{2}\widehat{(\sin(u^{(0)}))}_{lm}, & n = 0; \\ \frac{2}{2 + c^2(\Delta t)^2(\mu_l^2 + \lambda_m^2)}\left[2\widehat{(u^n_{JK})}_{lm} - (\Delta t)^2\widehat{(\sin(u^n_{JK})}_{lm}\right] - \widehat{(u^{n-1}_{JK})}_{lm}, & n \geq 1. \end{cases}$$

The above discretization scheme (5.24)–(5.25) is spectral order accurate in space and second-order accurate in time; in fact, one can have the following error estimate,

Theorem 5.1. *Let $t^* > 0$ be a fixed time and suppose the exact solution $u(\mathbf{x}, t)$ of problem (5.18)–(5.19) satisfies $u(\mathbf{x}, t) \in C^4\left([0, t^*]; L^2\right) \cap C^3\left([0, t^*]; H^1\right) \cap C^2\left([0, t^*]; H^2\right) \cap C\left([0, t^*]; H^m \cap H^1_0\right)$ for some $m \geq 2$. Let $u^n_{JK}(\mathbf{x})$ be the approximations obtained from (5.24)–(5.25), then there exist two positive constants k_0 and h_0, such that for any $0 < \Delta t \leq k_0$ and $0 < h := \max\{\Delta x,\, \Delta y\} \leq h_0$,*

$$\|e^n(\mathbf{x})\|_{L^2} \lesssim (\Delta t)^2 + h^m, \quad \|e^n(\mathbf{x})\|_{H^1} \lesssim (\Delta t)^2 + h^{m-1}, \quad 0 \leq n \leq \frac{t^*}{\Delta t}, \tag{5.26}$$

where $e^n(\mathbf{x}) = u(\mathbf{x}, t_n) - u^n_{JK}(\mathbf{x})$.

Proof. From the regularity of exact solution, one has

$$\max_{0\le t\le t^*} \left\{\left\|\partial_t^4 u(\cdot,t)\right\|_{L^2}, \left\|\partial_t^3 u(\cdot,t)\right\|_{H^1}, \|\partial_{tt}u(\cdot,t)\|_{H^2}, \|u(\cdot,t)\|_{H^m}\right\} \lesssim 1. \qquad (5.27)$$

Denote

$$u_{_{JK}}(\mathbf{x},t_n) := \mathcal{P}_{JK}u(\mathbf{x},t_n), \quad \eta^n(\mathbf{x}) := u_{_{JK}}(\mathbf{x},t_n) - u^n_{JK}(\mathbf{x}), \quad \mathbf{x}\in\Omega, \quad n\ge 0, \ (5.28)$$

then $\eta^n(\mathbf{x}) \in Y_{JK}$, and define the local truncation errors

$$\tau^0(\mathbf{x}) = \frac{u(\mathbf{x},t_1) - u^{(0)}(\mathbf{x})}{\Delta t} - u^{(1)}(\mathbf{x}) - \frac{\Delta t}{2}\left[c^2\Delta u^{(0)}(\mathbf{x}) - \sin\left(u^{(0)}(\mathbf{x})\right)\right], \qquad (5.29)$$

$$\begin{aligned}\tau^n(\mathbf{x}) &= \frac{u(\mathbf{x},t_{n+1}) - 2u(\mathbf{x},t_n) + u(\mathbf{x},t_{n-1})}{(\Delta t)^2} - \frac{c^2}{2}\left[\Delta u(\mathbf{x},t_{n+1}) + \Delta u(\mathbf{x},t_{n-1})\right]\\ &\quad + \sin(u(\mathbf{x},t_n)), \quad 1\le n\le \frac{t^*}{\Delta t} - 1, \quad \mathbf{x}\in\Omega. \qquad (5.30)\end{aligned}$$

Applying Taylor expansions to (5.29), noticing (5.18), (5.19) and (5.27), using the Hölder's inequality, one can get

$$\begin{aligned}\|\tau^0\|^2_{L^2} &= \int_\Omega \left[\int_0^{\Delta t} \frac{t^2}{2(\Delta t)}\partial_t^3 u(\mathbf{x},t)\,dt\right]^2 d\mathbf{x}\\ &\le \int_\Omega \left[\int_0^{\Delta t} \frac{t^4}{4(\Delta t)^2}\,dt \cdot \int_0^{\Delta t} \left|\partial_t^3 u(\mathbf{x},t)\right|^2 dt\right] d\mathbf{x}\\ &\le \frac{(\Delta t)^3}{20}\int_0^{\Delta t}\int_\Omega \left|\partial_t^3 u(\mathbf{x},t)\right|^2 d\mathbf{x}\,dt = \frac{(\Delta t)^3}{20}\int_0^{\Delta t} \|\partial_t^3 u(\cdot,t)\|^2_{L^2}\,dt\\ &\le \frac{(\Delta t)^4}{20}\max_{0\le t\le \Delta t}\|\partial_t^3 u(\cdot,t)\|^2_{L^2} \lesssim (\Delta t)^4. \qquad (5.31)\end{aligned}$$

Similarly,

$$\|\nabla\tau^0\|^2_{L^2} = \int_\Omega \left[\int_0^{\Delta t} \frac{t^2}{2(\Delta t)}\partial_t^3 \nabla u(\mathbf{x},t)\,dt\right]^2 d\mathbf{x} \lesssim (\Delta t)^4. \qquad (5.32)$$

From (5.30), (5.18) and (5.27),

$$\begin{aligned}
\|\tau^n\|_{L^2}^2 &\leq \int_\Omega \left\{ \int_{t_n}^{t_{n+1}} \left[\frac{(t-t_n)^3}{12(\Delta t)^2} \partial_t^4 u(\mathbf{x},t) + \frac{c^2}{2}(t-t_n)\partial_t^2 \Delta u(\mathbf{x},t) \right] dt \right. \\
&\quad \left. + \int_{t_{n-1}}^{t_n} \left[\frac{(t_n-t)^3}{12(\Delta t)^2} \partial_t^4 u(\mathbf{x},t) + \frac{c^2}{2}(t_n-t)\partial_t^2 \Delta u(\mathbf{x},t) \right] dt \right\}^2 d\mathbf{x} \\
&\leq (\Delta t)^4 \left[\frac{1}{126} \max_{t_{n-1}\leq t\leq t_{n+1}} \|\partial_t^4 u(\cdot,t)\|_{L^2}^2 + \frac{2c^4}{3} \max_{t_{n-1}\leq t\leq t_{n+1}} \|\partial_t^2 \Delta u(\cdot,t)\|_{L^2}^2 \right] \\
&\lesssim (\Delta t)^4, \quad 1 \leq n \leq \frac{t^*}{\Delta t} - 1. \qquad (5.33)
\end{aligned}$$

Applying the projection operator $\mathcal{P}_{JK}$ to (5.29) and (5.30), noticing (5.28), one can obtain

$$\mathcal{P}_{JK}\tau^0(\mathbf{x}) = \frac{u_{JK}(\mathbf{x},t_1) - u_{JK}^0}{\Delta t} - \mathcal{P}_{JK}u^{(1)} - \frac{\Delta t}{2}\left[c^2 \Delta u_{JK}^0 - \mathcal{P}_{JK}\sin\left(u^{(0)}\right)\right], \qquad (5.34)$$

$$\begin{aligned}
\mathcal{P}_{JK}\tau^n(\mathbf{x}) &= \frac{u_{JK}(\mathbf{x},t_{n+1}) - 2u_{JK}(\mathbf{x},t_n) + u_{JK}(\mathbf{x},t_{n-1})}{(\Delta t)^2} + \mathcal{P}_{JK}\sin(u(\mathbf{x},t_n)) \\
&- \frac{c^2}{2}\left[\Delta u_{JK}(\mathbf{x},t_{n+1}) + \Delta u_{JK}(\mathbf{x},t_{n-1})\right], \quad \mathbf{x}\in\Omega, \quad 1 \leq n \leq \frac{t^*}{\Delta t} - 1. \qquad (5.35)
\end{aligned}$$

Subtracting (5.24) and (5.25) from (5.35) and (5.34), respectively, noting (5.28), for $1 \leq n \leq \frac{t^*}{\Delta t} - 1$,

$$\begin{aligned}
&\frac{\eta^{n+1}(\mathbf{x}) - 2\eta^n(\mathbf{x}) + \eta^{n-1}(\mathbf{x})}{(\Delta t)^2} - \frac{c^2}{2}\left[\Delta\eta^{n+1}(\mathbf{x}) + \Delta\eta^{n-1}(\mathbf{x})\right] \\
&\quad = g^n(\mathbf{x}) - \mathcal{P}_{JK}(\tau^n(\mathbf{x})), \qquad (5.36) \\
&\eta^0(\mathbf{x}) = 0, \quad \frac{\eta^1(\mathbf{x}) - \eta^0(\mathbf{x})}{\Delta t} = \mathcal{P}_{JK}(\tau^0(\mathbf{x})), \quad \mathbf{x}\in\Omega, \qquad (5.37)
\end{aligned}$$

where,

$$g^n(\mathbf{x}) = \mathcal{P}_{JK}\left[\sin\left(u_{JK}^n\right) - \sin(u(\mathbf{x},t_n))\right], \quad \mathbf{x}\in\Omega, \quad 1 \leq n \leq \frac{t^*}{\Delta t}. \qquad (5.38)$$

From (5.38), using Poincaré inequality, one can get

$$\begin{aligned}
\|g^n\|_{L^2} &\leq \|\sin\left(u_{JK}^n(\mathbf{x})\right) - \sin(u(\mathbf{x},t_n))\|_{L^2} \leq \|\cos(\cdot)\|_{L^\infty} \cdot \|u_{JK}^n(\mathbf{x}) - u(\mathbf{x},t_n)\|_{L^2} \\
&\leq \|e^n\|_{L^2} \leq \|\eta^n\|_{L^2} + \|u(\mathbf{x},t_n) - \mathcal{P}_{JK}u(\mathbf{x},t_n)\|_{L^2} \\
&\lesssim \|\nabla\eta^n\|_{L^2} + h^m, \quad 1 \leq n \leq \frac{t^*}{\Delta t}. \qquad (5.39)
\end{aligned}$$

Define the energy for the error function η^n as

$$\mathcal{E}^n = \left\| \frac{\eta^{n+1}-\eta^n}{\Delta t} \right\|_{L^2}^2 + \frac{c^2}{2}\left(\|\nabla \eta^{n+1}\|_{L^2}^2 + \|\nabla \eta^n\|_{L^2}^2 \right), \quad 0 \le n \le \frac{t^*}{\Delta t} - 1. \tag{5.40}$$

Using (5.37), (5.31) and (5.32),

$$\begin{aligned} &\left\| \frac{\eta^1-\eta^0}{\Delta t} \right\|_{L^2}^2 \le \|\mathcal{P}_{JK}(\tau^0)\|_{L^2}^2 \le \|\tau^0\|_{L^2}^2 \lesssim (\Delta t)^4, \quad \|\nabla \eta^0\|_{L^2}^2 = 0, \\ &\|\nabla \eta^1\|_{L^2}^2 = (\Delta t)^2 \|\mathcal{P}_{JK}(\nabla \tau^0)\|_{L^2}^2 \le (\Delta t)^2 \|\nabla \tau^0\|_{L^2}^2 \lesssim (\Delta t)^6. \end{aligned} \tag{5.41}$$

Plugging (5.41) and (5.41) into (5.40) with $n = 0$,

$$\mathcal{E}^0 \lesssim (1 + (\Delta t)^2)(\Delta t)^4. \tag{5.42}$$

Multiplying both sides of (5.36) by $\eta^{n+1} - \eta^{n-1}$, integrating over Ω and using integration by parts, noticing (5.40), (5.33) and (5.39), one can have

$$\begin{aligned} \mathcal{E}^n - \mathcal{E}^{n-1} &\le \int_\Omega \left(|\mathcal{P}_{JK}(\tau^n)| + |g^n| \right) \left| \eta^{n+1} - \eta^{n-1} \right| d\mathbf{x}, \\ &= \Delta t \int_\Omega \left(|\mathcal{P}_{JK}(\tau^n)| + |g^n| \right) \left| \frac{\eta^{n+1}-\eta^n}{\Delta t} + \frac{\eta^n - \eta^{n-1}}{\Delta t} \right| d\mathbf{x} \\ &\le \Delta t \left[\|\mathcal{P}_{JK}(\tau^n)\|_{L^2}^2 + \|g^n\|_{L^2}^2 + \left\| \frac{\eta^{n+1}-\eta^n}{\Delta t} \right\|_{L^2}^2 + \left\| \frac{\eta^n-\eta^{n-1}}{\Delta t} \right\|_{L^2}^2 \right] \\ &\lesssim \Delta t \left[(\Delta t)^4 + h^{2m} + \mathcal{E}^n + \mathcal{E}^{n-1} \right], \quad 1 \le n \le \frac{t^*}{\Delta t} - 1. \end{aligned} \tag{5.43}$$

Then, there exists a positive constant $k_0 \le 1$, such that for $0 < \Delta t \le k_0$,

$$\mathcal{E}^n - \mathcal{E}^{n-1} \lesssim \Delta t \left[(\Delta t)^4 + h^{2m} + \mathcal{E}^{n-1} \right], \quad 1 \le n \le \frac{t^*}{\Delta t} - 1. \tag{5.44}$$

Summing up for $n \ge 1$, and noticing (5.42),

$$\mathcal{E}^n \lesssim (\Delta t)^4 + h^{2m} + \Delta t \sum_{r=0}^{n-1} \mathcal{E}^r, \quad 1 \le n \le \frac{t^*}{\Delta t} - 1. \tag{5.45}$$

Using the discrete Gronwall's inequality,

$$\mathcal{E}^n \lesssim (\Delta t)^4 + h^{2m}, \quad 0 \le n \le \frac{t^*}{\Delta t} - 1. \tag{5.46}$$

Thus the desired result (5.26) follows from (5.46) and (5.40), as well as the following triangle inequality

$$\begin{aligned} \|\nabla e^n\|_{L^2} &\leq \|\nabla \eta^n\|_{L^2} + \|\nabla\left(u(\mathbf{x},t_n) - \mathcal{P}_{JK}u(\mathbf{x},t_n)\right)\|_{L^2} \\ &\lesssim \|\nabla \eta^n\|_{L^2} + h^{m-1}, \quad 0 \leq n \leq \frac{t^*}{\Delta t}. \end{aligned}$$

□

The scheme (5.24)–(5.25) is not suitable in practice due to the difficulty in computing the integrals in (5.24), (5.25) and (5.21). Similar to previous chapters, an efficient implementation is achieved via approximating the integrals in (5.24), (5.25) and (5.21) by a quadrature rule on the grids $\{(x_j,y_k),\ j=0,1,\ldots,J,\ k=0,1,\ldots,K\}$. Choose $u^0_{jk} = u^{(0)}(x_j,y_k)$ ($j=0,1,\ldots,J,\ k=0,1,\ldots,M$), for $n=0,1,\ldots,$ the semi-implicit sine pseudospectral discretization for the problem (5.18)–(5.19) reads

$$u^{n+1}_{jk} = \sum_{l=1}^{J-1}\sum_{m=1}^{K-1} \widetilde{(u^{n+1})}_{lm}\phi_{lm}(x_j,y_k), \quad j=0,1,\ldots,J, \quad k=0,1,\ldots,K, \tag{5.47}$$

where

$$\begin{aligned} &\widetilde{(u^{n+1})}_{lm} \\ &= \begin{cases} \left[1-\frac{c^2}{2}(\Delta t)^2\left(\mu_l^2+\lambda_m^2\right)\right]\widetilde{(u^{(0)})}_{lm} + \Delta t\widetilde{(u^{(1)})}_{lm} - \frac{(\Delta t)^2}{2}\widetilde{(\sin(u^{(0)}))}_{lm}, & n=0; \\ \frac{2}{2+c^2(\Delta t)^2(\mu_l^2+\lambda_m^2)}\left[2\widetilde{(u^n)}_{lm} - (\Delta t)^2\widetilde{(\sin(u^n))}_{lm}\right] - \widetilde{(u^{n-1})}_{lm}, & n\geq 1. \end{cases} \end{aligned}$$

Again, this scheme is spectral order accurate in space and second-order accurate in time. It is explicitly solvable in phase space, the memory cost is $O(J\,K)$ and computation cost per time step is $O(J\,K\ln(JK))$ thanks to fast discrete sine transform (FST), thus it is very efficient in computation.

5.2.2 Method for the perturbed NLS equation

A semi-implicit sine pseudospectral method is discussed here for solving the perturbed NLS equation. Let $\Delta T > 0$ be the time step and denote time steps as $T_n = n\Delta T$, $n = 0,1,\ldots$; and choose spatial mesh sizes ΔX and ΔY and grid points

X_j $(j = 0, 1, \ldots, J)$ and Y_k $(k = 0, 1, \ldots, K)$ in a similar manner to Δx and Δy as well as x_j and y_k. Let $A^n_{JK}(\mathbf{X})$ be the approximation of $A(\mathbf{X}, T_n)$ $(\mathbf{X} \in \Omega)$, and respectively, A^n_{jk} be the approximation of $A(X_j, Y_k, T_n)$ $(j = 0, 1, \ldots, J,\ k = 0, 1, \ldots, K)$ and denote A^n be the matrix with components A^n_{jk} at time $T = T_n$. Choose $A^0_{JK}(\mathbf{X}) = \mathcal{P}_{JK}(A^{(0)})$ for $\mathbf{X} \in \Omega$, by applying the sine spectral method for spatial derivatives, and second-order implicit and explicit schemes for linear and nonlinear terms respectively in time discretization for the perturbed NLS equation (5.12), one gets the semi-implicit sine spectral discretization as:

Find $A^{n+1}_{JK}(\mathbf{X}) \in Y_{JK}$, i.e.,

$$A^{n+1}_{JK}(\mathbf{X}) = \sum_{l=1}^{J-1}\sum_{m=1}^{K-1} \widehat{(A^{n+1}_{JK})}_{lm}\phi_{lm}(\mathbf{X}), \quad \mathbf{X} \in \Omega, \quad n \geq 0, \tag{5.48}$$

such that for $\mathbf{X} \in \Omega$ and $n \geq 1$

$$\begin{aligned} i\frac{A^{n+1}_{JK} - A^{n-1}_{JK}}{2\Delta T} =& \frac{\varepsilon^2}{4\omega^2}\frac{A^{n+1}_{JK} - 2A^n_{JK} + A^{n-1}_{JK}}{(\Delta T)^2} - \frac{\varepsilon ck}{2\omega\Delta T}\left(\partial_X A^{n+1}_{JK} - \partial_X A^{n-1}_{JK}\right) \\ & - \frac{1}{2}\left(\Delta A^{n+1}_{JK} + \Delta A^{n-1}_{JK}\right) + \mathcal{P}_{JK}\left(f^N_\varepsilon\left(|A^n_{JK}|^2\right)A^n_{JK}\right), \end{aligned} \tag{5.49}$$

and the initial data in (5.6) is discretized as

$$\begin{aligned} \frac{A^1_{JK} - A^0_{JK}}{\Delta T} =& \mathcal{P}_{JK}(A^{(1)}) + \frac{4\omega^2\Delta T}{2\varepsilon^2}\Big[i\mathcal{P}_{JK}(A^{(1)}) + \Delta A^0_{JK} \\ & + \frac{\varepsilon ck}{\omega}\partial_X\mathcal{P}_{JK}(A^{(1)}) - \mathcal{P}_{JK}\left(f^N_\varepsilon(|A^{(0)}|^2)A^{(0)}\right)\Big]. \end{aligned} \tag{5.50}$$

Plugging (5.48) into (5.50) and (5.49) and noticing the orthogonality of sine functions, for $l = 1, 2, \ldots, J-1$ and $m = 1, 2, \ldots, K-1$,

$$\begin{aligned} & \widehat{(A^{n+1}_{JK})}_{lm} \\ =& \begin{cases} \alpha_{lm}\widehat{(A^{(0)})}_{lm} + \beta_{lm}\widehat{(A^{(1)})}_{lm} - \dfrac{2\omega^2(\Delta T)^2}{\varepsilon^2}\widehat{(g^0)}_{lm}, & n = 0; \\ \dfrac{i - \gamma_{lm}}{i + \gamma_{lm}}\widehat{(A^{n-1}_{JK})}_{lm} - \dfrac{\varepsilon^2}{\omega^2\Delta T\,(i + \gamma_{lm})}\widehat{(A^n_{JK})}_{lm} + \dfrac{2\Delta T}{i + \gamma_{lm}}\widehat{(g^n)}_{lm}, & n \geq 1, \end{cases} \end{aligned}$$

where

$$\begin{aligned}
&\alpha_{lm} = 1 - \frac{2\omega^2(\Delta T)^2}{\varepsilon^2}\left(\mu_l^2+\lambda_m^2\right), \quad \beta_{lm} = \Delta T + \frac{i2\omega^2(\Delta T)^2}{\varepsilon^2} + \frac{i2\omega ck\mu_l(\Delta T)^2}{\varepsilon}\\
&\gamma_{lm} = -\Delta T\left(\mu_l^2+\lambda_m^2\right) - \frac{\varepsilon^2}{2\omega^2\Delta T} + \frac{i\varepsilon ck\mu_l}{\omega}, \quad 1 \le l \le J-1,\ 1 \le m \le K-1,\\
&g^0(\mathbf{X}) = f_\varepsilon^N\left(|A^{(0)}(\mathbf{X})|^2\right)A^{(0)}(\mathbf{X}), \qquad (5.51)\\
&g^n(\mathbf{X}) = f_\varepsilon^N\left(|A_{JK}^n(\mathbf{X})|^2\right)A_{JK}^n(\mathbf{X}), \quad n \ge 1, \quad \mathbf{X}\in\Omega.
\end{aligned}$$

Similarly, the above discretization scheme (5.49)–(5.50) is spectral order accurate in space and second-order accurate in time; in fact, one can have the following error estimate,

Theorem 5.2. *Let $\varepsilon = \varepsilon_0$ be a fixed constant in (5.12) and $T^* > 0$ be any fixed time, suppose the exact solution $A(\mathbf{X},T)$ of an initial-boundary-value problem of (5.12) satisfies $A(\mathbf{X},T) \in C^4\left([0,T^*];L^2\right) \cap C^3\left([0,T^*];H^1\right) \cap C^2\left([0,T^*];H^2\right) \cap C\left([0,T^*];H^m\cap H_0^1\cap L^\infty(\Omega)\right)$ for some $m \ge 2$. Let A_{JK}^n be the approximations obtained from (5.49) and (5.50) at time $T = T_n$, then there exist two positive constants k_0 and h_0, such that for any $0 \le \Delta T \le k_0$ and $0 < h := \max\{\Delta X, \Delta Y\} \le h_0$, satisfying $\Delta T \lesssim 1/\left|\ln(h)\right|$,*

$$\|e^n(\mathbf{X})\|_{L^2} \lesssim (\Delta T)^2 + h^m, \quad \|e^n(\mathbf{X})\|_{H^1} \lesssim (\Delta T)^2 + h^{m-1}, \quad 0 \le n \le \frac{T^*}{\Delta T}, \quad (5.52)$$

where $e^n(\mathbf{X}) = A(\mathbf{X},T_n) - A_{JK}^n(\mathbf{X})$.

Proof. The proof proceeds by means of mathematical induction, and without loss of generality one can assume $\Delta X = \Delta Y$. From the regularity of exact solution,

$$\begin{aligned}
\max_{0\le T\le T^*}&\left\{\left\|\partial_T^4 A(\mathbf{X},T)\right\|_{L^2}, \left\|\partial_T^3 A(\mathbf{X},T)\right\|_{H^1},\right.\\
&\left.\|\partial_{TT}A(\mathbf{X},T)\|_{H^2}, \|A(\mathbf{X},T)\|_{H^m}, \|A(\mathbf{X},T)\|_{L^\infty}\right\} \lesssim 1, \qquad (5.53)
\end{aligned}$$

and by the smoothness of f_ε^N,

$$\max_{0\le T\le T^*}\left\{\left\|f_\varepsilon^N\left(|A(\mathbf{X},T)|^2\right)\right\|_{L^\infty}, \left\|\left(f_\varepsilon^N\right)'\left((|A(\mathbf{X},T)|+1)^2\right)\right\|_{L^\infty}\right\} \lesssim 1. \qquad (5.54)$$

Denote

$$A_{JK}(\mathbf{X},T_n) := \mathcal{P}_{JK}A(\mathbf{X},T_n), \quad \eta^n(\mathbf{X}) := A_{JK}(\mathbf{X},T_n) - A_{JK}^n(\mathbf{X}), \quad \mathbf{X}\in\Omega, \quad n \ge 0,$$

$$(5.55)$$

then $\eta^n(\mathbf{X}) \in Y_{JK}$, and define the local truncation errors as

$$\begin{aligned}
\tau^0(\mathbf{X}) &= \frac{A(\mathbf{X},T_1)-A^{(0)}(\mathbf{X})}{\Delta T} - A^{(1)}(\mathbf{X}) - \frac{4\omega^2\Delta T}{2\varepsilon^2}\Big[i\,A^{(1)}(\mathbf{X}) + \Delta A^{(0)}(\mathbf{X}) \\
&\quad + \frac{\varepsilon ck}{\omega}\partial_X A^{(1)}(\mathbf{X}) - \mathcal{P}_{JK}\left(f_\varepsilon^N\left(|A^{(0)}(\mathbf{X})|^2\right)A^{(0)}(\mathbf{X})\right)\Big], \quad \mathbf{X}\in\Omega, \qquad (5.56)\\
\tau^n(\mathbf{X}) &= i\,\frac{A(\mathbf{X},T_{n+1})-A(\mathbf{X},T_{n-1})}{2\Delta T} - \frac{\varepsilon^2}{4\omega^2}\frac{A(\mathbf{X},T_{n+1})-2A(\mathbf{X},T_n)+A(\mathbf{X},T_{n-1})}{(\Delta T)^2} \\
&\quad + \frac{1}{2}\left(\Delta A(\mathbf{X},T_{n+1}) + \Delta A(\mathbf{X},T_{n-1})\right) - f_\varepsilon^N\left(|A(\mathbf{X},T_n)|^2\right)A(\mathbf{X},T_n) \\
&\quad + \frac{\varepsilon ck}{2\omega\Delta T}\left(\partial_X A(\mathbf{X},T_{n+1}) - \partial_X A(\mathbf{X},T_{n-1})\right), \quad 1\le n\le \frac{T^*}{\Delta T}-1, \qquad (5.57)
\end{aligned}$$

then via similar arguments to (5.31)–(5.33), one can get

$$\|\tau^n\|_{L^2}^2 \lesssim (\Delta T)^4, \quad 0\le n\le \frac{T^*}{\Delta T}-1, \quad \left\|\nabla\tau^0\right\|_{L^2}^2 \lesssim (\Delta T)^4. \qquad (5.58)$$

Similar to Theorem 5.1, the error function η^n satisfies

$$\begin{aligned}
&i\frac{\eta^{n+1}(\mathbf{X})-\eta^{n-1}(\mathbf{X})}{2\Delta T} = \frac{\varepsilon^2}{4\omega^2}\frac{\eta^{n+1}(\mathbf{X})-2\eta^n(\mathbf{X})+\eta^{n-1}(\mathbf{X})}{(\Delta T)^2} \\
&\quad - \frac{\varepsilon ck}{2\omega\Delta T}\left(\partial_X\eta^{n+1}(\mathbf{X}) - \partial_X\eta^{n-1}(\mathbf{X})\right) - \frac{1}{2}\left(\Delta\eta^{n+1}(\mathbf{X}) + \Delta\eta^{n-1}(\mathbf{X})\right) \\
&\quad + q^n(\mathbf{X}) + \mathcal{P}_{JK}\left(\tau^n(\mathbf{X})\right), \quad 1\le n\le \frac{T^*}{\Delta T}-1, \qquad (5.59)\\
&\eta^0(\mathbf{X}) = 0, \quad \frac{\eta^1(\mathbf{X})-\eta^0}{\Delta T} = \mathcal{P}_{JK}\left(\tau^0(\mathbf{X})\right), \quad \mathbf{X}\in\Omega, \qquad (5.60)
\end{aligned}$$

where for $1\le n\le T^*/\Delta T-1$,

$$q^n(\mathbf{X}) = \mathcal{P}_{JK}\left[f_\varepsilon^N\left(|A(\mathbf{X},T_n)|^2\right)A(\mathbf{X},T_n) - f_\varepsilon^N\left(|A_{JK}^n(\mathbf{X})|^2\right)A_{JK}^n(\mathbf{X})\right]. \quad (5.61)$$

Define the energy for error function η^n as

$$\mathcal{E}^n = \frac{\varepsilon^2}{4\omega^2}\left\|\frac{\eta^{n+1}-\eta^n}{\Delta T}\right\|_{L^2}^2 + \frac{1}{2}\left(\left\|\nabla\eta^{n+1}\right\|_{L^2}^2 + \|\nabla\eta^n\|_{L^2}^2\right), \quad 0\le n\le \frac{T^*}{\Delta T}-1. \quad (5.62)$$

Then, similar to (5.41) and (5.42),

$$\mathcal{E}^0 \lesssim \left(1+(\Delta T)^2\right)(\Delta T)^4. \qquad (5.63)$$

Multipling both sides of (5.59) by $(\eta^{n+1})^* - (\eta^{n-1})^*$, integrating over Ω and taking the real part, with similar argument to (5.43), one can have for $1 \le n \le T^*/\Delta T - 1$,

$$\begin{aligned} &\mathcal{E}^n - \mathcal{E}^{n-1} \\ &\le \ \Delta T \left[\|q^n\|_{L^2}^2 + \|\mathcal{P}_{JK}(\tau^n)\|_{L^2}^2 + \left\| \frac{\eta^{n+1} - \eta^n}{\Delta T} \right\|_{L^2}^2 + \left\| \frac{\eta^n - \eta^{n-1}}{\Delta T} \right\|_{L^2}^2 \right]. \end{aligned} \tag{5.64}$$

Note that

$$\left\|\eta^1\right\|_{L^2} = \Delta T \left\|\mathcal{P}_{JK}(\tau^0)\right\|_{L^2} \lesssim (\Delta T)^3, \quad \left\|\nabla\eta^1\right\|_{L^2} = \Delta T \left\|\mathcal{P}_{JK}(\nabla\tau^0)\right\|_{L^2} \lesssim (\Delta T)^3,$$

then,

$$\begin{aligned} &\left\|e^1\right\|_{L^2} \le \left\|\eta^1\right\|_{L^2} + \|\mathcal{P}_{JK}A(\mathbf{X}, T_1) - A(\mathbf{X}, T_1)\|_{L^2} \lesssim (\Delta T)^3 + h^m, \\ &\left\|\nabla e^1\right\|_{L^2} \lesssim \left\|\nabla\eta^1\right\|_{L^2} + \|\nabla\left(\mathcal{P}_{JK}A(\mathbf{X}, T_1) - A(\mathbf{X}, T_1)\right)\|_{L^2} \lesssim (\Delta T)^3 + h^{m-1}, \end{aligned} \tag{5.65}$$

which results in the estimate (5.52) for $n = 1$.

Since $\eta^n \in Y_{JK}$ and noticing $\Delta T \lesssim 1/\left|\ln(h)\right|$, by the inverse inequality one can have

$$\left\|\eta^1\right\|_{L^\infty} \lesssim \left|\ln(h)\right| \left\|\eta^1\right\|_{H^1} \lesssim (\Delta T)^2, \tag{5.66}$$

and then,

$$\left\|e^1\right\|_{L^\infty} \le \left\|\eta^1\right\|_{L^\infty} + \|\mathcal{P}_{JK}A(\mathbf{X}, T_1) - A(\mathbf{X}, T_1)\|_{L^\infty} \lesssim (\Delta T)^2 + h^{m-1}. \tag{5.67}$$

Choose $k_0' > 0$ and $h_0' > 0$ such that

$$\left\|A_{JK}^1\right\|_{L^\infty} \le \|A(\mathbf{X}, T_1)\|_{L^\infty} + \left\|e^1\right\|_{L^\infty} \le \|A(\mathbf{X}, T_1)\|_{L^\infty} + 1, \quad \Delta T \le k_0', \quad h \le h_0'. \tag{5.68}$$

Now one can estimate $\mathcal{E}^1$. At $T = T_1$, noticing (5.68) and (5.54),

$$\begin{aligned} &\left\|q^1\right\|_{L^2} \le \left\|f_\varepsilon^N(|A(\mathbf{X}, T_1)|^2)A(\mathbf{X}, T_1) - f_\varepsilon^N(|A_{JK}^1|^2)A_{JK}^1\right\|_{L^2} \\ &\le \ \left\|f_\varepsilon^N(|A(\mathbf{X}, T_1)|^2)\right\|_{L^\infty} \left\|e^1\right\|_{L^2} + \left\|\left(f_\varepsilon^N\left(|A(\mathbf{X}, T_1)|^2\right) - f_\varepsilon^N\left(|A_{JK}^1|^2\right)\right) A_{JK}^1\right\|_{L^2} \\ &\lesssim \ \left\|e^1\right\|_{L^2} \left[1 + \left(2\left\|A(\mathbf{X}, T_1)\right\|_{L^\infty} + 1\right)^2 \left\|(f_\varepsilon^N)'\left((|A(\mathbf{X}, T_1)| + 1)^2\right)\right\|_{L^\infty}\right] \\ &\lesssim \ \left\|e^1\right\|_{L^2} \lesssim \|\eta^1\|_{L^2} + h^m \lesssim \|\nabla\eta^1\|_{L^2} + h^m. \end{aligned} \tag{5.69}$$

Plugging (5.58) and (5.69) into (5.64) and noticing (5.62),

$$\mathcal{E}^1 - \mathcal{E}^0 \leq C_1 \Delta T \left[(\Delta T)^4 + h^{2m} + \left(\mathcal{E}^1 + \mathcal{E}^0 \right) \right]. \tag{5.70}$$

Then when $\Delta T \leq \frac{1}{2C_1}$,

$$\mathcal{E}^1 \leq \mathcal{E}^0 + 4C_1 \Delta T \left[(\Delta T)^4 + h^{2m} + \mathcal{E}^0 \right]. \tag{5.71}$$

Noticing the estimate of $\mathcal{E}^0$ (5.63), for $h \leq h_0'$ and $\Delta T \leq \min\left\{\frac{1}{2C_1}, k_0'\right\}$,

$$\begin{aligned} \mathcal{E}^1 &\leq (C_2 + 4C_1\Delta T)\left((\Delta T)^4 + h^{2m}\right) \mathrm{e}^{4C_1 \Delta T} \\ &\leq (C_2 + 4C_1 T^*)\left((\Delta T)^4 + h^{2m}\right) \mathrm{e}^{4C_1 T^*}. \end{aligned} \tag{5.72}$$

In view of (5.62) for $n = 1$, with the above estimate on $\mathcal{E}^1$, thanks to Poincaré inequality and inverse inequality, one can have for $h < 1$,

$$\left\|\eta^2\right\|_{L^2} \lesssim \left\|\nabla \eta^2\right\|_{L^2} \lesssim (\Delta T)^2 + h^m, \tag{5.73}$$

$$\left\|\eta^2\right\|_{L^\infty} \lesssim |\ln(h)| \left\|\eta^2\right\|_{H^1} \lesssim \Delta T + |\ln(h)|\, h^m \lesssim \Delta T + h^{m-1}. \tag{5.74}$$

So,

$$\left\|e^2\right\|_{L^2} \lesssim (\Delta T)^2 + h^m, \quad \left\|\nabla e^2\right\|_{L^2} \lesssim (\Delta T)^2 + h^{m-1}, \quad \left\|e^2\right\|_{L^\infty} \lesssim \Delta T + h^{m-1}, \tag{5.75}$$

which establishes (5.52) for $n = 2$. Again, there exist $k_0'' > 0$ and $1 > h_0'' > 0$, such that

$$\left\|e^2\right\|_{L^\infty} \leq 1, \tag{5.76}$$

if $\Delta T \leq k_0''$ and $h \leq h_0''$.

Choose

$$k_0 = \min\left\{\frac{1}{2C_1}, k_0', k_0''\right\}, \quad h_0 = \min\{h_0', h_0''\}, \tag{5.77}$$

where k_0' and h_0' are chosen such that (5.68) holds, and k_0'' and h_0'' are chosen such that (5.76) is valid. Noting that k_0 and h_0 only depend on the regularity of exact solution and smoothness of f_ε^N, i.e. (5.53) and (5.54), as well as the finial computation time T^*, the rest justification is due to induction.

For $1 \le n \le T^*/\Delta T - 1$ and $\Delta T \le k_0$ and $h \le h_0$, satisfying $\Delta T \lesssim 1/|\ln(h)|$, assume

$$\left\|e^l\right\|_{L^2} \lesssim (\Delta T)^2 + h^m, \quad \left\|\nabla e^l\right\|_{L^2} \lesssim (\Delta T)^2 + h^{m-1}, \quad \left\|e^l\right\|_{L^\infty} \le 1, \quad 2 \le l \le n. \tag{5.78}$$

For $l = 1$, one already has (5.65) and (5.68). Then

$$\left\|A^l_{JK}(\mathbf{X})\right\|_{L^\infty} \le \|A(\mathbf{X}, T_l)\|_{L^\infty} + 1, \quad 1 \le l \le n. \tag{5.79}$$

With similar argument to (5.69),

$$\left\|q^l\right\|_{L^2} \lesssim \left\|\nabla \eta^l\right\|_{L^2} + h^m, \quad 1 \le l \le n. \tag{5.80}$$

Noticing (5.64) and (5.58), similar to the proof of Theorem 5.1, when $\Delta T \le \frac{1}{2C_1}$,

$$\mathcal{E}^n \le \mathcal{E}^0 + 4C_1 n\Delta T \left[(\Delta T)^4 + h^{2m}\right] + 4C_1\Delta T \sum_{l=0}^{n-1} \mathcal{E}^l. \tag{5.81}$$

Since $n \le T^*/\Delta T - 1$, when $\Delta T \le \min\left\{\frac{1}{2C_1}, k_0'\right\}$, one can obtain, by using the discrete Gronwall's inequality and noting (5.63),

$$\mathcal{E}^n \le (C_2 + 4C_1 n\Delta T)\left((\Delta T)^4 + h^{2m}\right) \lesssim (\Delta T)^4 + h^{2m}. \tag{5.82}$$

In view of (5.62), similar to (5.73)–(5.74) and (5.75), one can obtain

$$\left\|e^{n+1}\right\|_{L^2} \lesssim (\Delta T)^2 + h^m, \quad \left\|\nabla e^{n+1}\right\|_{L^2} \lesssim (\Delta T)^2 + h^{m-1}, \tag{5.83}$$

$$\left\|e^{n+1}\right\|_{L^\infty} \lesssim \Delta T + |\ln(h)|\, h^m. \tag{5.84}$$

Noticing k_0 and h_0 are chosen as (5.77), when $\Delta T \le k_0$ and $h \le h_0$, one has

$$\left\|e^{n+1}\right\|_{L^\infty} \le 1. \tag{5.85}$$

In above estimates, the constants C_1 and C_2 are independent of mesh size h and time step ΔT as well as time steps $0 \le n \le \frac{T}{\Delta T}$, therefore k_0', k_0'', h_0' and h_0'' are the same as before and they can be chosen such that they are independent of mesh size h and time step ΔT as well as time steps $0 \le n \le \frac{T}{\Delta T}$. Hence, (5.83), (5.84) and (5.85) prove (5.78) for $l = n + 1$, and the claim in Theorem 5.2 follows by mathematical induction. □

Again, the scheme (5.49) and (5.50) is not suitable in practice due to the difficulty in computing the integrals in (5.49), (5.50) and (5.21). Similarly, one can apply a pseudospectral method in implementation. Choose $A^0_{jk} = A^{(0)}(X_j, Y_k)$ ($j = 0, 1, \ldots, J,\ k = 0, 1, \ldots, M$), for $n = 0, 1, \ldots,$ the semi-implicit sine pseudospectral discretization for the problem (5.12) and (5.6) reads

$$A^{n+1}_{jk} = \sum_{l=1}^{J-1}\sum_{m=1}^{K-1} \widetilde{(A^{n+1})}_{lm}\phi_{lm}(X_j, Y_k), \quad j = 0, 1, \ldots, J, \quad k = 0, 1, \ldots, K, \tag{5.86}$$

where

$$\begin{aligned}&\widetilde{(A^{n+1})}_{lm}\\ &= \begin{cases} \alpha_{lm}\widetilde{(A^{(0)})}_{lm} + \beta_{lm}\widetilde{(A^{(1)})}_{lm} - \dfrac{2\omega^2(\Delta T)^2}{\varepsilon^2}\widetilde{(g^0)}_{lm}, & n = 0; \\ \dfrac{i-\gamma_{lm}}{i+\gamma_{lm}}\widetilde{(A^{n-1})}_{lm} - \dfrac{\varepsilon^2}{\omega^2\Delta T\,(i+\gamma_{lm})}\widetilde{(A^n)}_{lm} + \dfrac{2\Delta T}{i+\gamma_{lm}}\widetilde{(g^n)}_{lm}, & n \geq 1; \end{cases}\end{aligned}$$

where

$$g^n_{jk} = f^N_\varepsilon\left(|A^n_{jk}|^2\right)A^n_{jk}, \quad 0 \leq j \leq J, \quad 0 \leq k \leq K, \quad n \geq 0.$$

Again, this scheme is spectral order accurate in space and second-order accurate in time. It is explicitly solvable in phase space, the memory cost is $O\,(J\,K)$ and computation cost per time step is $O\,(J\,K\ln(JK))$ thanks to FST, thus it is very efficient in computation.

5.3 Numerical results

In this section, the SG equation (1.14), the perturbed NLS equation (5.12) with different N, and the cubic NLS equation (5.15) are numerically studied for modeling the LBs. Numerical comparisons are made among them, and the propagating pulses are investigated via solving the perturbed NLS equation (5.12) with N adequately large. The SG and perturbed NLS equations are solved by the efficient methods proposed before, and the cubic NLS equation is solved by the efficient and accurate time-splitting pseudospectral method [19–21]. In simulation, $c = 1$ in (1.14) and the initial data $A^{(0)}(\mathbf{X})$ in (5.6) and (5.16) is chosen such that it decays to zero sufficiently fast as $|\mathbf{X}| \to \infty$. In order to make the perturbed NLS equation (5.12) be consistent with the cubic NLS equation (5.15) at $T = 0$ when $\varepsilon \to 0$, the initial data $A^{(1)}(\mathbf{X})$ ($A^{(1)}(\mathbf{X})$ appears in the coefficient before $O(\varepsilon^3)$ term in the ansatz (5.3) for initial data of the SG equation) in (5.6) is chosen as

$$A^{(1)}(\mathbf{X}) = i\left[\Delta A^{(0)}(\mathbf{X}) + \frac{1}{2}\left|A^{(0)}(\mathbf{X})\right| A^{(0)}(\mathbf{X})\right], \quad \mathbf{X} \in \mathbb{R}^2. \tag{5.87}$$

From the ansatz (5.3) with $t = 0$ and omitting all $O(\varepsilon^3)$ terms, the initial data in (1.15) for the SG equation can be chosen as

$$u^{(0)}(\mathbf{x}) = \varepsilon\left[\cos(kx)\left(A^{(0)} + \left(A^{(0)}\right)^*\right) + i\sin(kx)\left(A^{(0)} - \left(A^{(0)}\right)^*\right)\right], \quad \mathbf{x} \in \mathbb{R}^2, \tag{5.88}$$

$$\begin{aligned} u^{(1)}(\mathbf{x}) = {} & \varepsilon\omega\left[i\cos(kx)\left(\left(A^{(0)}\right)^* - A^{(0)}\right) + \sin(kx)\left(\left(A^{(0)}\right)^* + A^{(0)}\right)\right] \\ & - \varepsilon^2 k\left[\cos(kx)\partial_X\left(A^{(0)} + \left(A^{(0)}\right)^*\right) + i\sin(kx)\partial_X\left(A^{(0)} - \left(A^{(0)}\right)^*\right)\right], \end{aligned} \tag{5.89}$$

where

$$A^{(0)} = A^{(0)}(\mathbf{X}) = A^{(0)}(\varepsilon\omega x, \varepsilon y), \quad X = \varepsilon\omega x, \quad Y = \varepsilon y, \quad \mathbf{X} \in \mathbb{R}^2.$$

With the solution A^n_{jk} of the perturbed NLS equation (5.12) or the cubic NLS equation (5.15), one can construct the envelope solution of NLS-type equations as

$$u^{\text{nls}}(\mathbf{x}, t) = \varepsilon A\left(\frac{\omega\varepsilon(x - \nu t)}{c}, \frac{\varepsilon y}{c}, \frac{\varepsilon^2 t}{2\omega}\right) \mathrm{e}^{i(kx - \omega t)} + \text{c.c.}, \quad \mathbf{x} \in \mathbb{R}^2, \quad t \geq 0. \tag{5.90}$$

Computations are always carried out on a domain large enough such that the zero boundary conditions do not introduce a significant aliasing error relative to the problem in whole space. Also, in all the results below, there is no substantial improvement by refining the mesh sizes and time steps. The studies mainly focus on the regime beyond the critical collapse in cubic NLS, but some results in the regime that no blow-up occurs in cubic NLS will be reported first.

5.3.1 Comparisons for no blow-up in cubic NLS

Take the initial data in (5.6) and (5.16) as

$$A^{(0)}(X,Y) = ia_0 \exp\left(-\frac{X^2+Y^2}{\sigma^2}\right), \quad (X,Y) \in \mathbb{R}^2, \tag{5.91}$$

with $a_0 = 1.6$ and $\sigma^2 = 1$ such that $E^{\text{CNLS}}(0) > 0$ and thus no finite-time collapse occurs in the cubic NLS equation (5.15). Plugging (5.91) into (5.88) and (5.89), one can immediately get the initial conditions in this case for the SG equation (1.14) as

$$u^{(0)}(x,y) = -2\varepsilon a_0 \varepsilon \exp\left(-\frac{\varepsilon^2(\omega^2x^2+y^2)}{\sigma^2}\right)\sin(kx), \quad (x,y) \in \mathbb{R}^2, \tag{5.92}$$

$$u^{(1)}(x,y) = 2a_0\varepsilon\omega \exp\left(-\frac{\varepsilon^2(\omega^2x^2+y^2)}{\sigma^2}\right)\left(\cos(kx) - \frac{2\varepsilon^2 kx}{\sigma^2}\sin(kx)\right), \tag{5.93}$$

with $\omega = \sqrt{1+k^2}$.

Here, numerical results are reported for $\varepsilon = 0.1$ and $k = 1$. Fig. 5.1 shows the surface plots of u^{sg} of the SG equation (1.14) with $\varepsilon = 0.1$ and u^{nls} of the perturbed NLS equations with $N = 0, 1$ as well as the cubic NLS equation at $t = 40$ in the SG time scales which corresponds to $T = 0.1414$ in the NLS time scale. Fig. 5.2 depicts the slice plots of u^{sg} and u^{nls} at $t = 40$ along x- and y-axis.

5.3.2 Comparisons when blow-up occurs in cubic NLS

Take the initial data in (5.6) and (5.16) as

$$A^{(0)}(\mathbf{X}) = ia_0 \text{sech}\left(\frac{X^2+Y^2}{\sigma^2}\right), \quad \mathbf{X} \in \mathbb{R}^2, \tag{5.94}$$

with $a_0 = 5.2$ and $\sigma^2 = 0.8$ such that $E^{\text{CNLS}}(0) < 0$ and thus finite-time collapse occurs in the cubic NLS equation (5.15). Again, plugging (5.94) into (5.88) and (5.89), one can immediately get the initial conditions in this case for the SG equation (1.14) as

$$u^{(0)}(\mathbf{x}) = -2a_0\varepsilon \text{sech}\left(\frac{\varepsilon^2(\omega^2x^2+y^2)}{\sigma^2}\right)\sin(kx), \quad \mathbf{x} \in \mathbb{R}^2, \tag{5.95}$$

$$u^{(1)}(\mathbf{x}) = -\omega u^{(0)}(\mathbf{x})\left[\cot(kx) - \frac{2\varepsilon^2 kx}{\sigma^2}\tanh\left(\frac{\varepsilon^2(\omega^2x^2+y^2)}{\sigma^2}\right)\right], \tag{5.96}$$

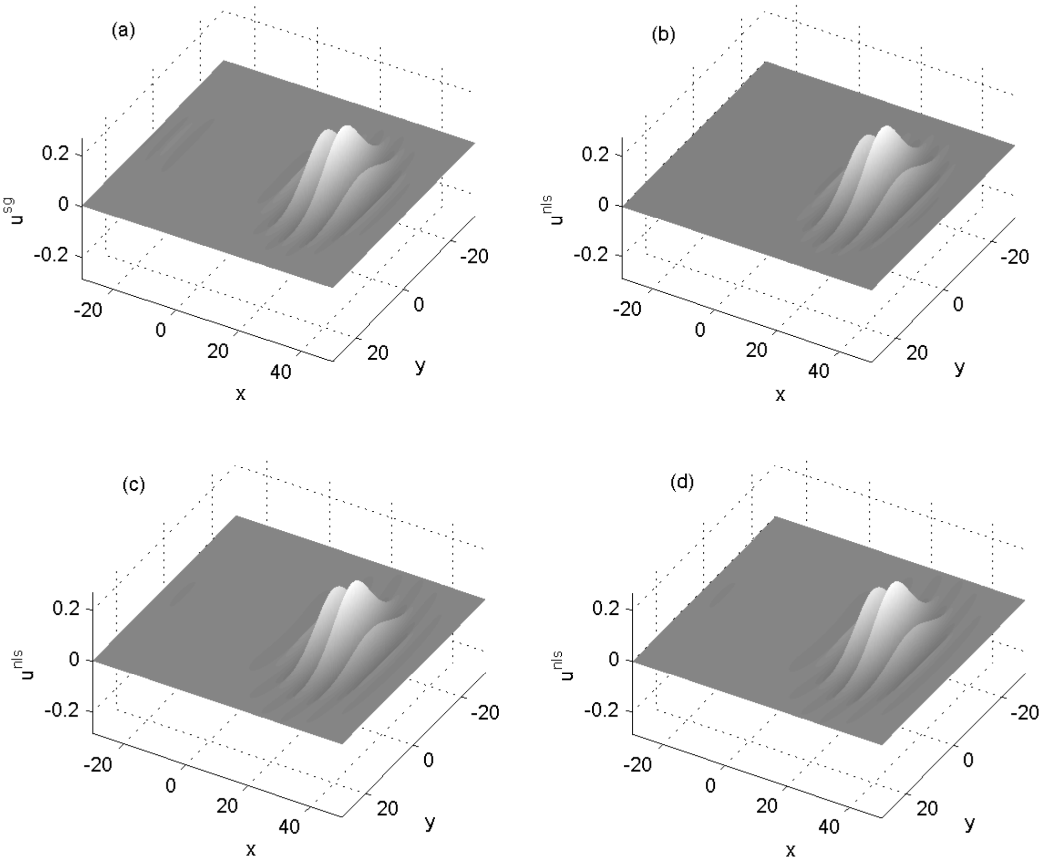


Figure 5.1: Surface plots of the numerical solutions of $u^{\rm sg}$ and $u^{\rm nls}$ at $t = 40$ in the SG time scale which corresponds to $T = 0.1414$ in the NLS time scale for $\varepsilon = 0.1$ and $k = 1$, in the case that no finite time collapse occurs in the cubic NLS. (a) SG solution; (b) cubic NLS solution; (c) perturbed NLS solution with $N = 0$; and (d) perturbed NLS solution with $N = 1$.

with $\omega = \sqrt{1+k^2}$.

Here, numerical results are reported for $\varepsilon = 0.1$ and $\varepsilon = 0.05$ ($k = 1$ in both cases), with comparing the approximated LBs solutions of the SG, the perturbed NLS and the cubic NLS equations at three typical time regimes, i.e. before, near and after the collapse time $T = T^c \approx 0.1310$ of the cubic NLS equation. Here, T^c is numerically found by looking at the evolution of either center density $|A(0,0,T)|^2$ or kinetic energy $K^{\rm cnls} := \int_\Omega \frac{1}{2} \|\nabla A(\mathbf{X},T)\|^2 \, d\mathbf{X}$; see Fig. 5.3.

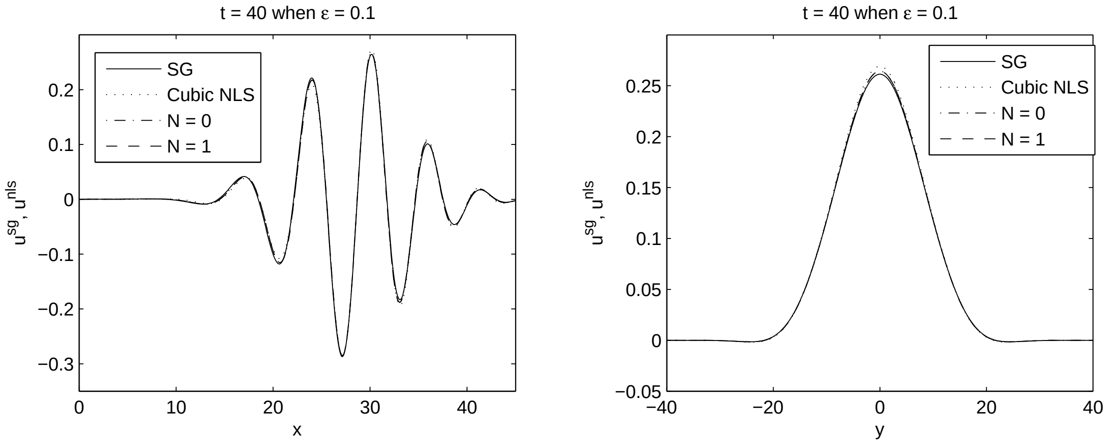


Figure 5.2: Slice plots of the numerical solutions of u^{sg} and u^{nls} at $t = 40$ for $\varepsilon = 0.1$ and $k = 1$, in the case that no finite time collapse occurs in the cubic NLS. Left column: along x-axis at $y = 0$; right column: along y-axis at $x = 30$.

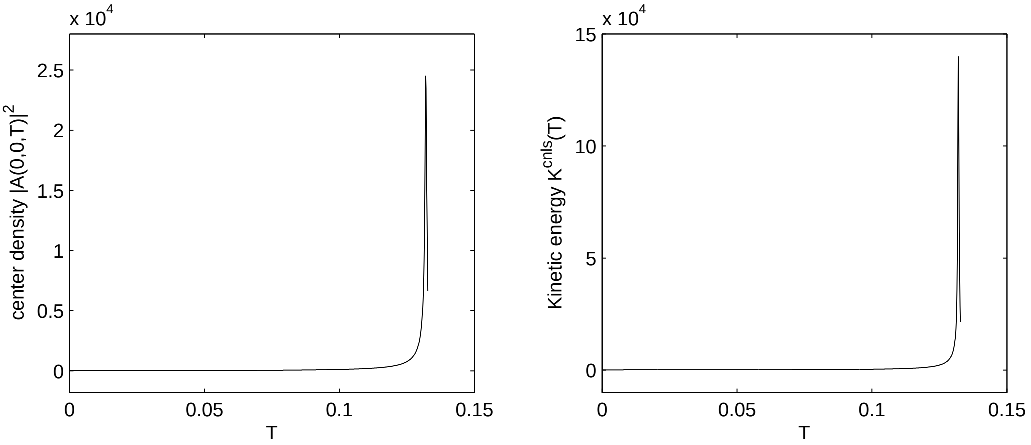


Figure 5.3: Evolution of center density $|A(0, 0, T)|^2$ and kinetic energy $K^{\mathrm{cnls}}(T)$ for cubic NLS with initial data chosen as (5.94) and $a_0 = 5.2$, numerically implying blow-up happens at $T^c \approx 0.1310$.

(i). *Numerical results well before collapse time of cubic NLS*, Fig. 5.4 shows the surface plots of u^{sg} of the SG equation (1.14) with $\varepsilon = 0.1$, and u^{nls} of the perturbed NLS equation with $N = 0, 1$ as well as cubic NLS at $t = 27.12$ in the SG time scale which corresponds to $T = 0.0950 < T^c$ in the NLS time scale (before collapse time of cubic NLS). Similar results for $\varepsilon = 0.05$ are shown in

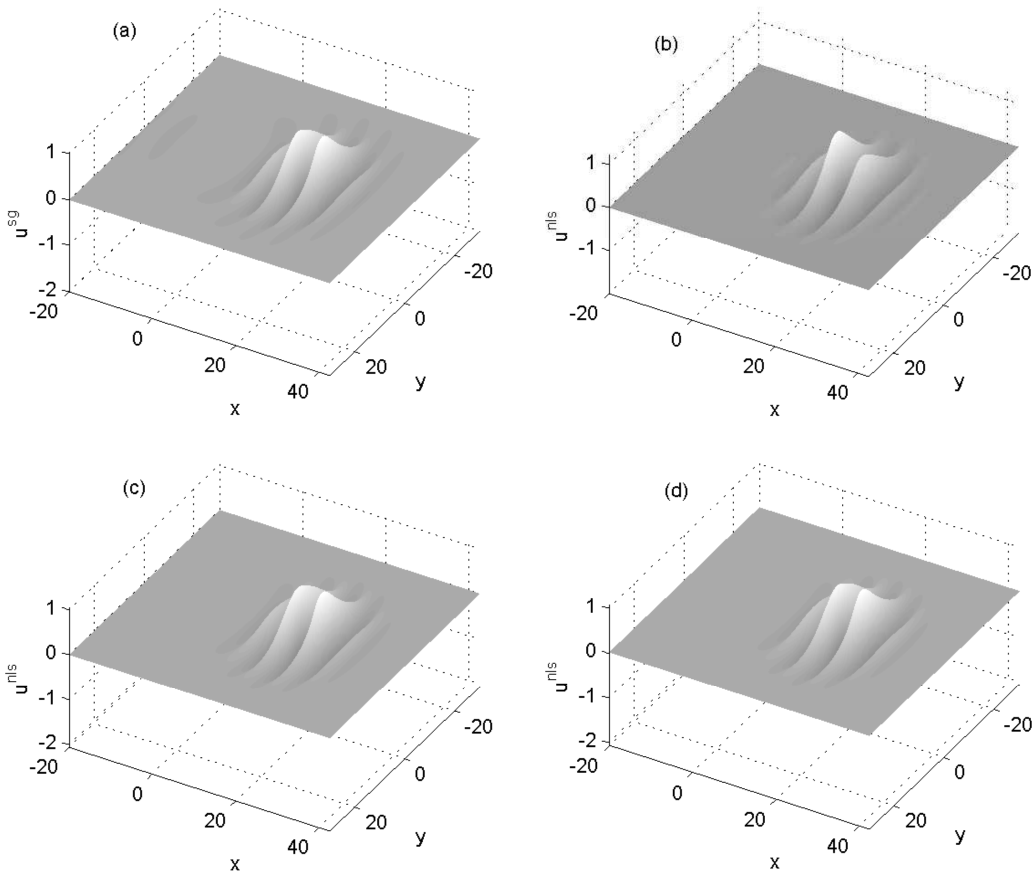


Figure 5.4: Surface plots of the numerical solutions of $u^{\rm sg}$ and $u^{\rm nls}$ at $t = 27.12$ in the SG time scale which corresponds to $T = 0.095 < T^c$ (well before collapse of cubic NLS) in the NLS time scale for $\varepsilon = 0.1$ and $k = 1$. (a) SG solution; (b) cubic NLS solution; (c) perturbed NLS solution with $N = 0$; and (d) perturbed NLS solution with $N = 1$.

Fig. 5.5. Fig. 5.6 plots $u^{\rm sg}$ and $u^{\rm nls}$ along the x-axis with $y = 0$ in this case.

(ii). *Numerical results near collapse time of cubic NLS*, Fig. 5.7 shows the surface plots of $u^{\rm sg}$ of the SG equation (1.14) with $\varepsilon = 0.1$, and $u^{\rm nls}$ of the perturbed NLS equation with $N = 0, 1$ as well as cubic NLS at $t = 37.04$ in the SG time scale which corresponds to $T = 0.1310 \approx T^c$ in the NLS time scale (near collapse time of cubic NLS). The similar results for $\varepsilon = 0.05$ are also shown in Fig. 5.8, and $u^{\rm sg}$ and $u^{\rm nls}$ along the x-axis with $y = 0$ are plotted out in Fig.

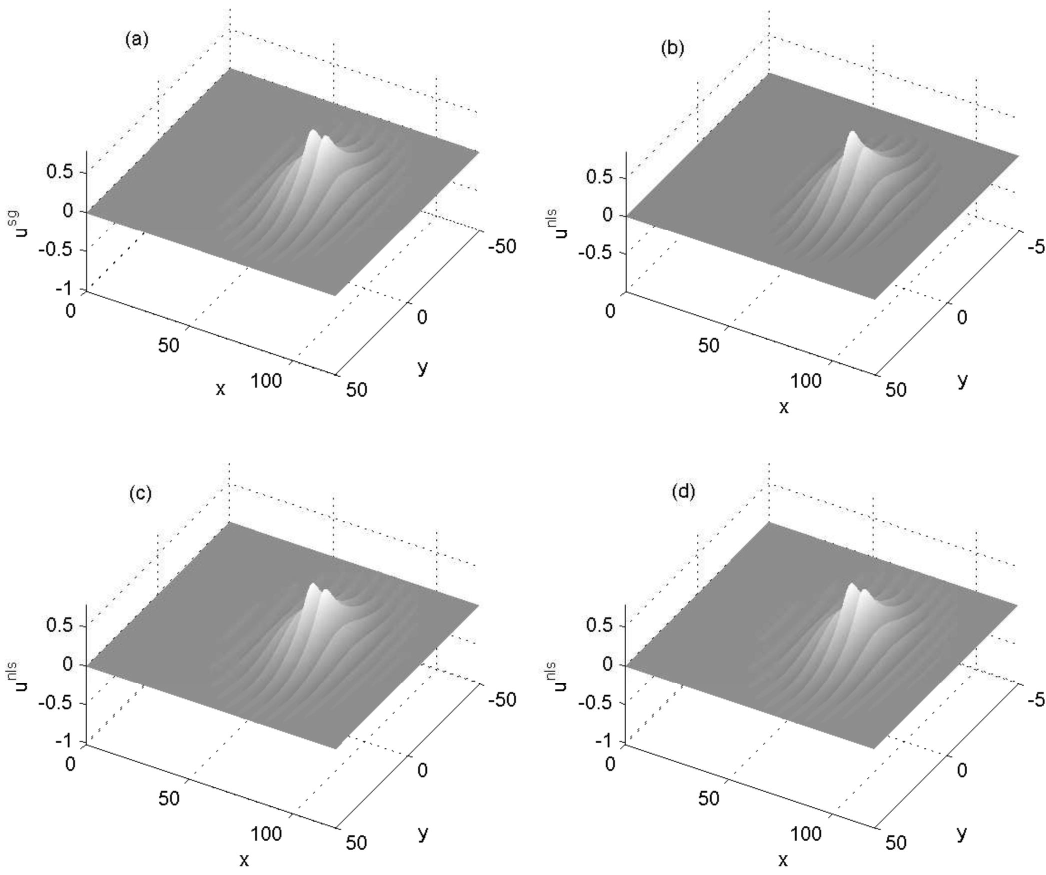


Figure 5.5: Surface plots of the numerical solutions of u^{sg} and u^{nls} at $t = 115.2$ in the SG time scale which corresponds to $T = 0.095 < T^c$ (well before collapse of cubic NLS) in the NLS time scale for $\varepsilon = 0.05$ and $k = 1$. (a) SG solution; (b) cubic NLS solution; (c) perturbed NLS solution with $N = 0$; and (d) perturbed NLS solution with $N = 1$.

5.9.

(iii). *Numerical results well after collapse time of cubic NLS*, Fig. 5.10 shows the surface plots of u^{sg} of the SG equation (1.14) with $\varepsilon = 0.1$, and u^{nls} of the perturbed NLS equation with $N = 0, 1, 2$ at $t = 64$ in the SG time scale which corresponds to $T = 0.2263 > T^c$ in the NLS time scale (after collapse time of cubic NLS). The similar results for $\varepsilon = 0.05$ are shown in Fig. 5.11. Fig. 5.12 plots u^{sg} and u^{nls} along the x-axis with $y = 0$ in this case.

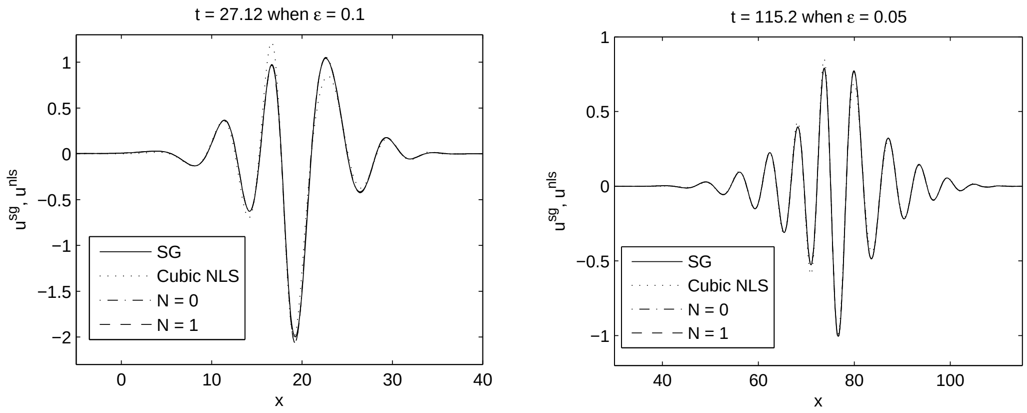


Figure 5.6: Slice plots of the numerical solutions of u^{sg} and u^{nls} along x-axis with $y = 0$ for $k = 1$. Left column: for $\varepsilon = 0.1$ at $t = 27.12$; right column: for $\varepsilon = 0.05$ at $t = 115.2$.

It can be seen that the results in the case, in which no finite-time collapse occurs in cubic NLS, are quite similar to the results before collapse time. From Figs. 5.1–5.2, 5.4–5.12 and additional numerical results (for refined meshes, different ε as well as various k) not shown here for brevity, one can draw the following conclusions for the propagation of the LBs:

(i). In the time regime well before the collapse time of the cubic NLS, or cubic NLS without blow-up, both cubic NLS equation (5.15) and the perturbed NLS equation (5.12) with $N \geq 0$ agree qualitatively and quantitatively, when ε is reasonably small, with the SG equation (1.14) (cf. Figs. 5.1, 5.4, and 5.5).

(ii). In the time regime near the collapse time of the cubic NLS, cubic NLS (5.15) fails to approximate the SG equation (1.14) neither quantitatively nor qualitatively (cf. Figs. 5.7 a & b, 5.8 a & b, and 5.9 “top row”); the perturbed NLS equation (5.12) with $N \geq 0$ agrees qualitatively and quantitatively, when ε is reasonably small, with the SG equation (1.14) (cf. Figs. 5.7 a, c & d, 5.8 a, c & d, and 5.9 “bottom row”).

(iii). In the time regime beyond the collapse time of the cubic NLS, cubic NLS (5.15) is no longer valid for the approximation of the SG equation (1.14);

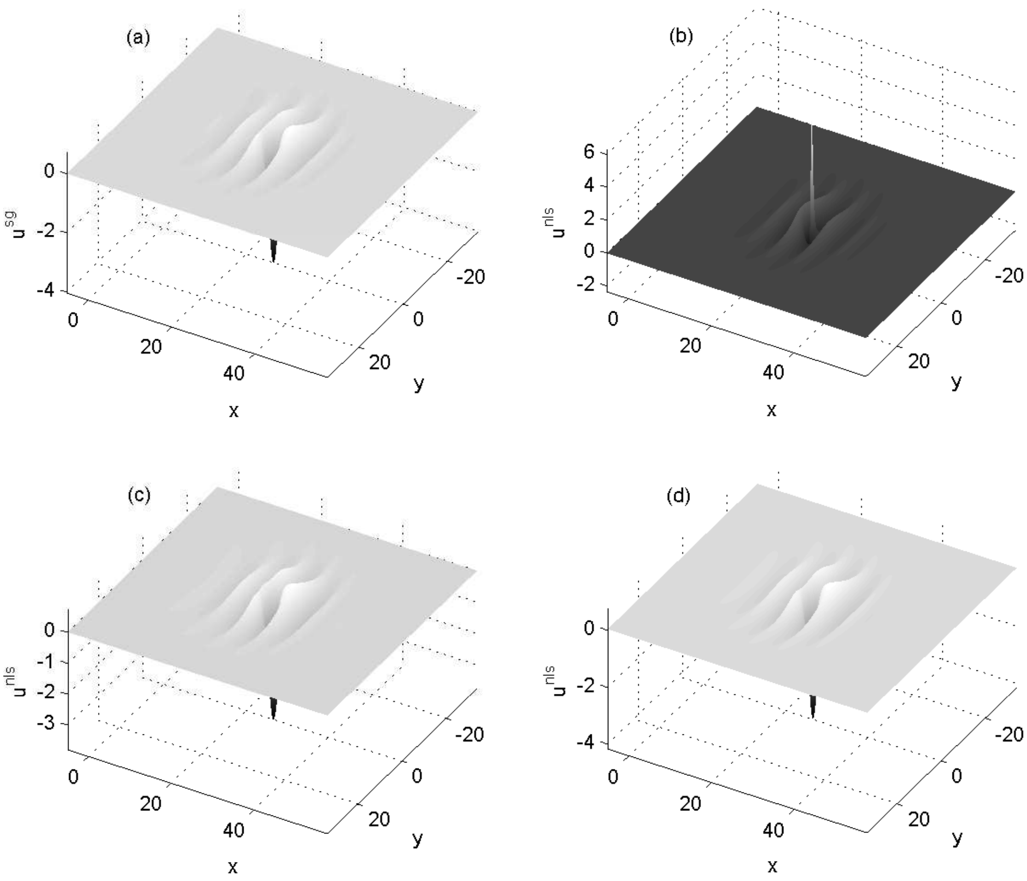


Figure 5.7: Surface plots of the numerical solutions of $u^{\rm sg}$ and $u^{\rm nls}$ at $t = 37.04$ in the SG time scale which corresponds to $T = 0.1310 \approx T^c$ (near collapse of cubic NLS) in the NLS time scale for $\varepsilon = 0.1$ and $k = 1$. (a) SG solution; (b) cubic NLS solution; (c) perturbed NLS solution with $N = 0$; and (d) perturbed NLS solution with $N = 1$.

the perturbed NLS equation (5.12) with $N = 0$ agrees qualitatively but not quantitatively with the SG equation (1.14) (cf. Figs. 5.10 a & b, 5.11 a & b, and 5.12 "top row"); and the perturbed NLS equation (5.12) with $N \geq 1$ agrees qualitatively and quantitatively, when ε is reasonably small, with the SG equation (1.14) (cf. Figs. 5.10 a, c & d, 5.11 a, c & d, and 5.12 "bottom row").

(iv). In general, for fixed time t, the smaller ε is and the larger N is, the better the

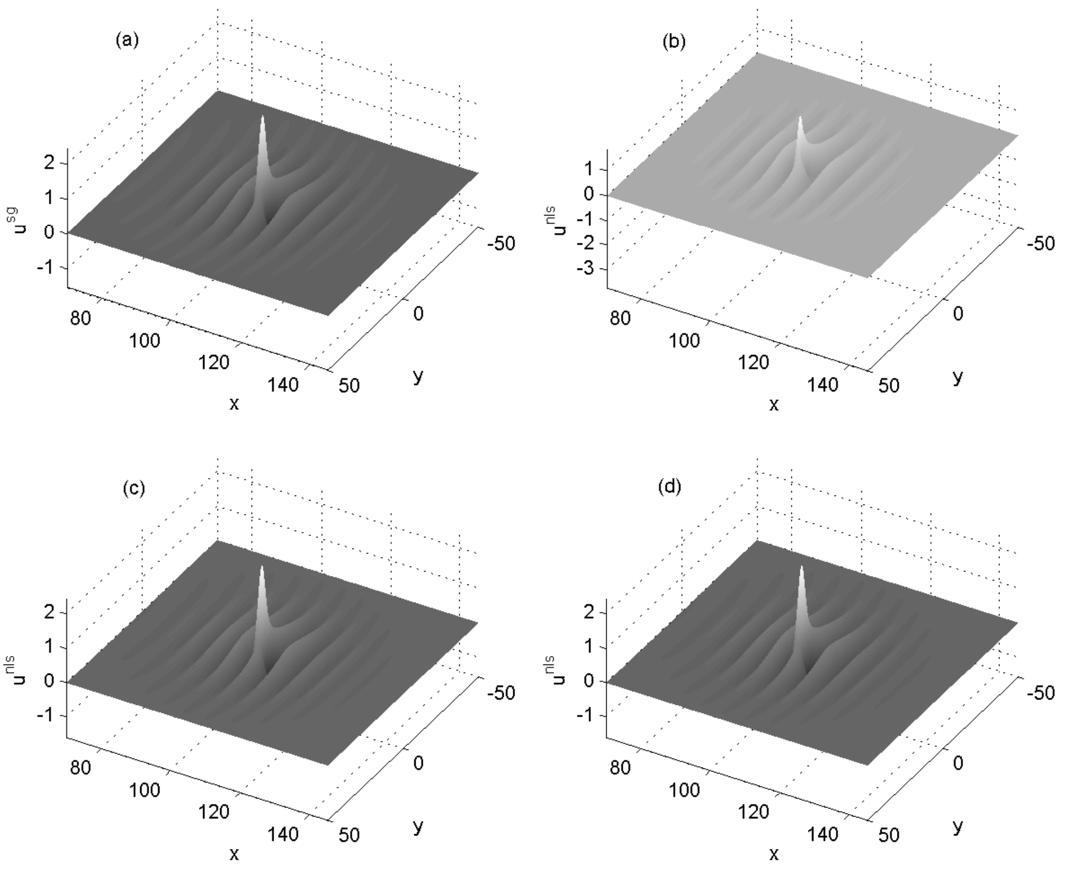


Figure 5.8: Surface plots of the numerical solutions of $u^{\rm sg}$ and $u^{\rm nls}$ at $t = 148.16$ in the SG time scale which corresponds to $T = 0.1310 \approx T^c$ (near collapse of cubic NLS) in the NLS time scale for $\varepsilon = 0.05$ and $k = 1$. (a) SG solution; (b) cubic NLS solution; (c) perturbed NLS solution with $N = 0$; and (d) perturbed NLS solution with $N = 1$.

approximation is (cf. Figs. 5.6, 5.9 “bottow row”, and 5.12 “bottom row”).

The above observations validate what are normally expected, i.e., cubic NLS fails to match SG well before and beyond its collapse time, but the perturbed NLS still agrees with SG beyond the critical collapse.

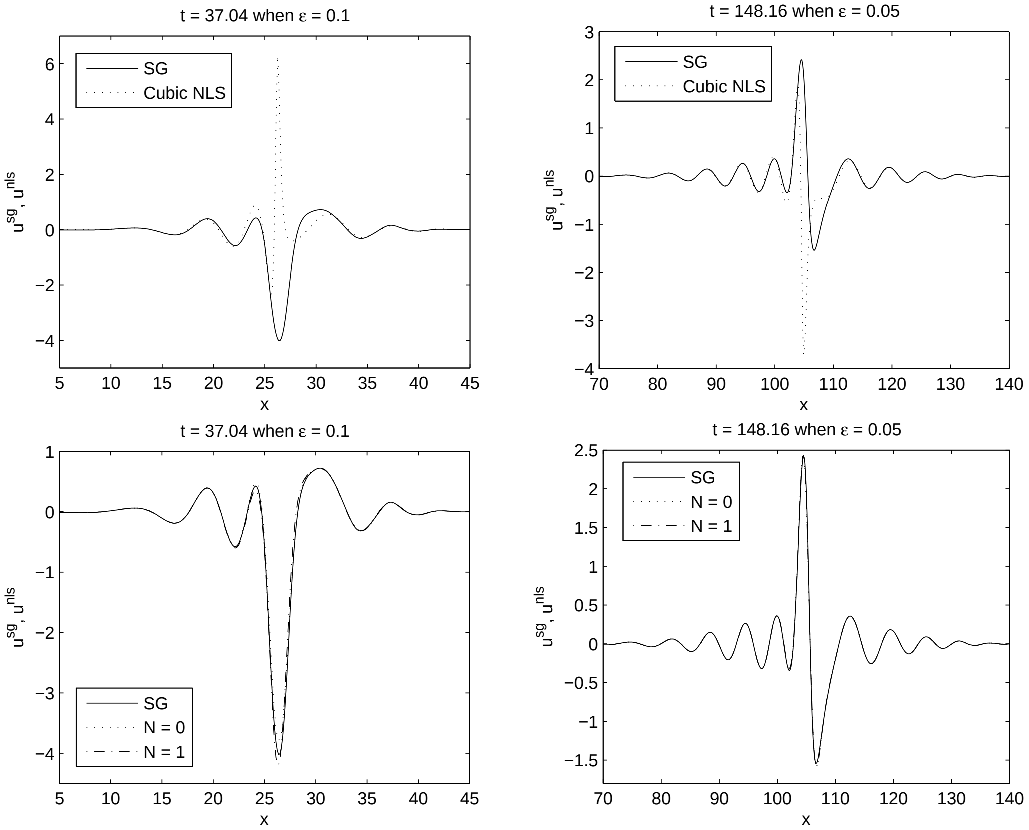


Figure 5.9: Slice plots of the numerical solutions of u^{sg} and u^{nls} along x-axis with $y = 0$ for $k = 1$. Top row: comparison of SG and cubic NLS; Bottom row: comparison of SG and perturbed NLS with different N.

5.3.3 Study on finite-term approximation

To understand how good the finite-term approximation (5.11) to (5.7) in the perturbed NLS equation (5.12) is, we solve (5.12) with initial data (5.6) for different N and ε. Fig. 5.13 plots time evolution of $\|A(\mathbf{X}, T)\|_\infty$ when initial data $A^{(0)}(\mathbf{X})$ in (5.6) is chosen as (5.94) with initial amplitude $a_0 = 5.2$, i.e., initial data leading to the occurrence of finite-time collapse in the cubic NLS, for different N and ε; and Fig. 5.14 shows similar results when $N = 50$ for different ε.

From Figs. 5.13, 5.14 and additional numerical results (for different initial data in (5.6) and different ε and N) not shown here for brevity, we can draw the following conclusions:

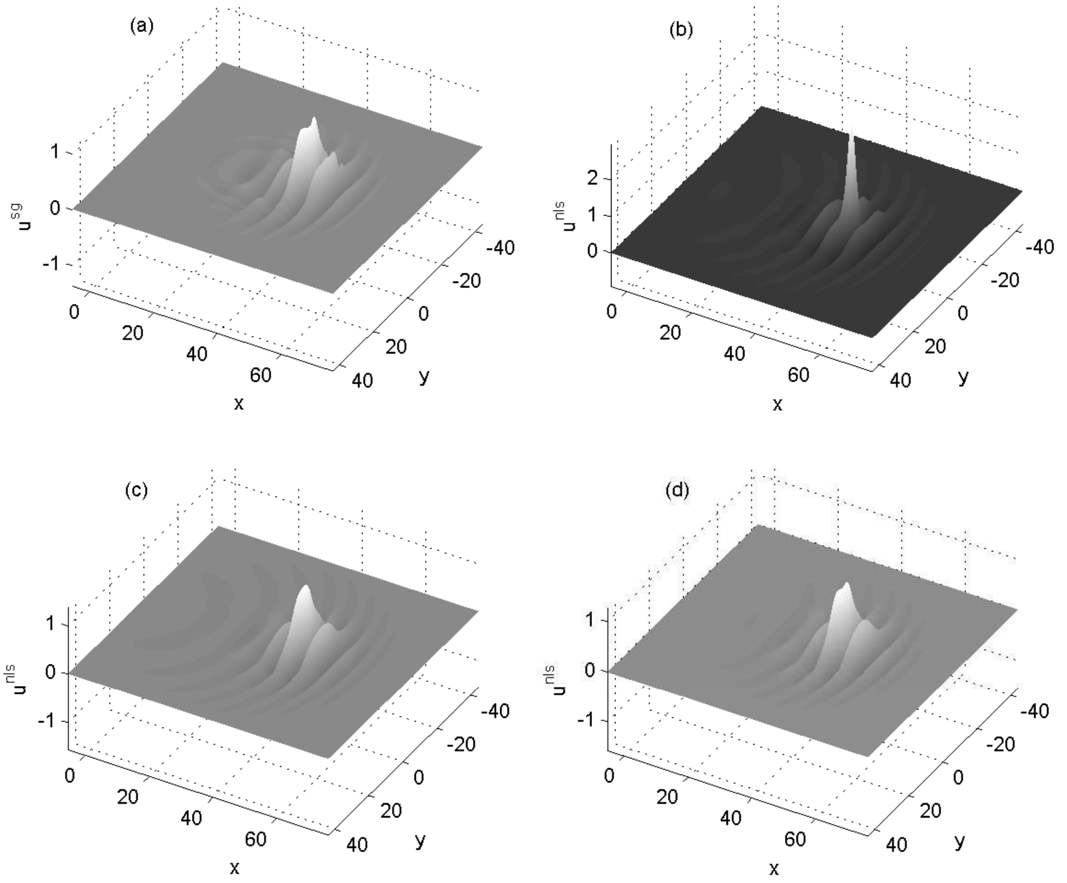


Figure 5.10: Surface plots of the numerical solutions of u^{sg} and u^{nls} at $t = 64$ in the SG time scale which corresponds to $T = 0.2263 > T^c$ (after collapse of cubic NLS) in the NLS time scale for $\varepsilon = 0.1$ and $k = 1$. (a) SG solution; (b) perturbed NLS solution with $N = 0$; (c) perturbed NLS solution with $N = 1$; and (d) perturbed NLS solution with $N = 2$.

(i). For initial data in (5.6) such that cubic NLS has no finite-time collapse, $\|A(\mathbf{X}, T)\|_\infty$ of either the cubic NLS equation (5.15) or the perturbed NLS equation (5.12) is uniformly bounded for $T \geq 0$, $N \geq 0$ and $0 < \varepsilon \leq \varepsilon_0$, for some ε_0.

(ii). For initial data in (5.6) such that cubic NLS has finite-time collapse, in the time regime $0 \leq T \leq T_0 < T^c$, i.e. well before the collapse time of cubic NLS,

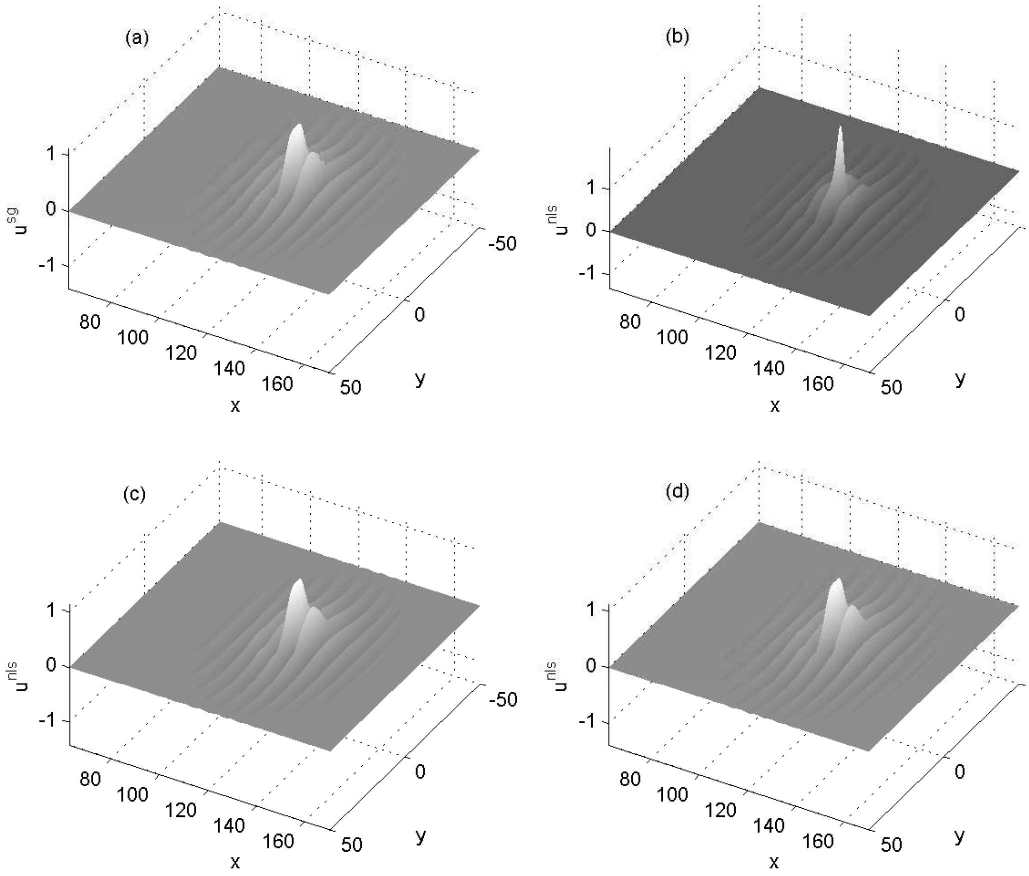


Figure 5.11: Surface plots of the numerical solutions of $u^{\rm sg}$ and $u^{\rm nls}$ at $t = 179.2$ in the SG time scale which corresponds to $T = 0.1584 > T^c$ (after collapse of cubic NLS) in the NLS time scale for $\varepsilon = 0.05$ and $k = 1$. (a) SG solution; (b) perturbed NLS solution with $N = 0$; (c) perturbed NLS solution with $N = 1$; and (d) perturbed NLS solution with $N = 2$.

$\|A(\mathbf{X},T)\|_\infty$ of the cubic NLS equation (5.15) and the perturbed NLS equation (5.12) is again uniformly bounded for $N \geq 0$ and $0 < \varepsilon \leq \varepsilon_0$; however, in the time regimes $T \approx T^c$ and $T > T^c$, i.e. near and after the collapse time of cubic NLS, $\|A(\mathbf{X},T)\|_\infty$ of cubic NLS goes to ∞ when $T \to T^c$; for fixed $\varepsilon > 0$, $\|A(\mathbf{X},T)\|_\infty$ of the perturbed NLS equation (5.12) is uniformly bounded for $N \geq 0$ and $T \geq T^c$ but the peak values of $\|A(\mathbf{X},T)\|_\infty$ increases linearly as $O\left(\varepsilon^{-1}\right)$ (cf. Fig. 5.14) which implies $\varepsilon\|A(\mathbf{X},T)\|_\infty$ is uniformly bounded (cf.

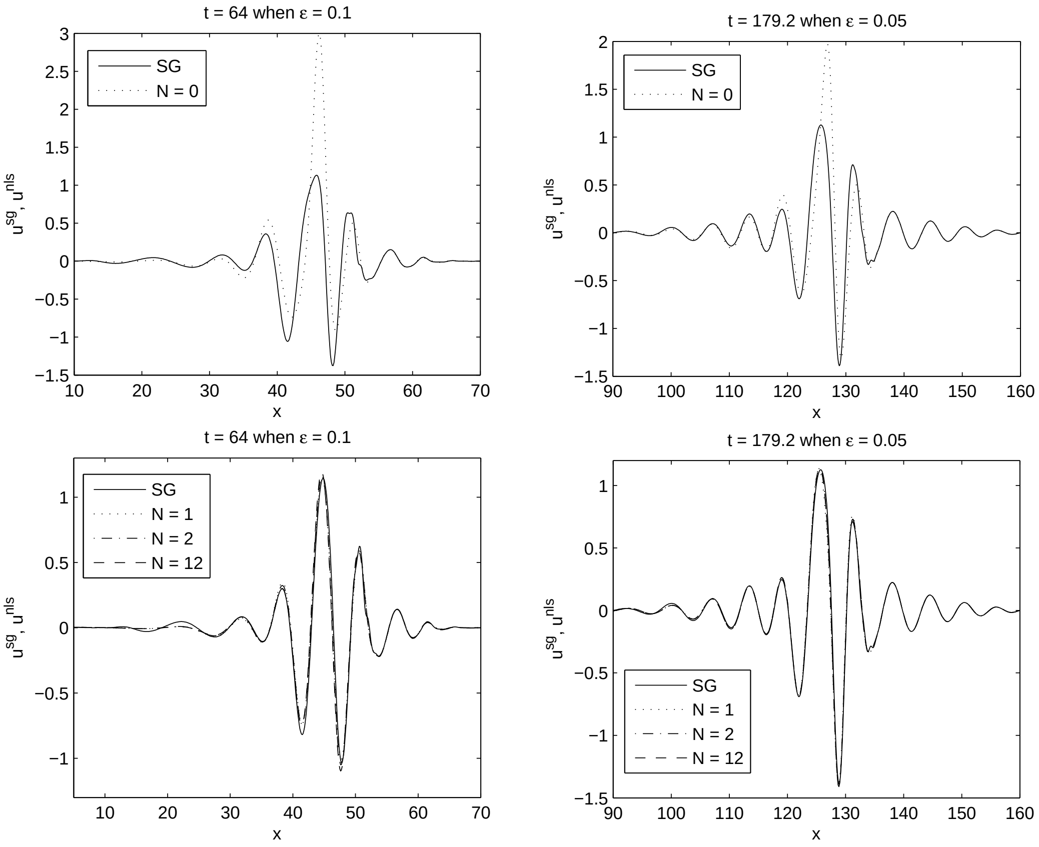


Figure 5.12: Slice plots of the numerical solutions of u^{sg} and u^{nls} along x-axis with $y = 0$ for $k = 1$. Top row: comparison between SG and perturbed NLS with $N = 0$; Bottom row: comparison between SG and perturbed NLS with $N = 1, 2, 12$.

Figs. 5.13 and 5.14), and such bound depends on the initial amplitude. The linear increase of $\|A(\mathbf{X}, T)\|_\infty$ with ε^{-1} agrees with the modulation analysis of perturbed NLS equation. Recall from (5.4) in [149] that $|A(\mathbf{X}, T)| \sim L_\varepsilon^{-1} R$, where R is the bounded Townes soliton profile, and L_ε undergoes oscillation with minimum value of order $O(\varepsilon)$; see conclusion I(1) and (5.24) on p. 358 in [149]. It follows that in the regime of focusing-defocusing (breathing) cycle, $\|A(\mathbf{X}, T)\|_\infty = O(\varepsilon^{-1})$.

(iii). When $N \geq N_0$ for some N_0, e.g. $N_0 = 3$ for initial data (5.94), there is no substantial difference in the dynamics of $\|A(\mathbf{X}, T)\|_\infty$ (cf. Fig. 5.13), and such an adequate N_0 also depends on the initial amplitude (cf. Fig. 5.15).

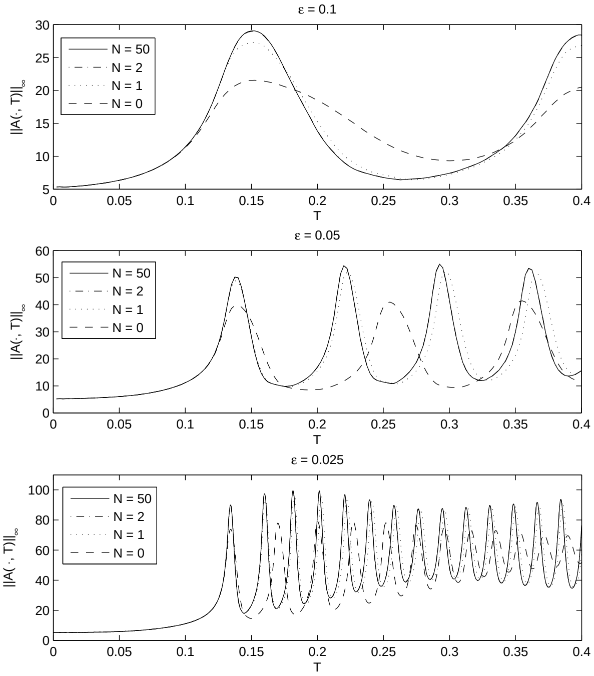


Figure 5.13: Time evolution of $\|A(\mathbf{X},T)\|_\infty$ for the perturbed NLS (5.12) with initial data (5.94) for different N and ε.

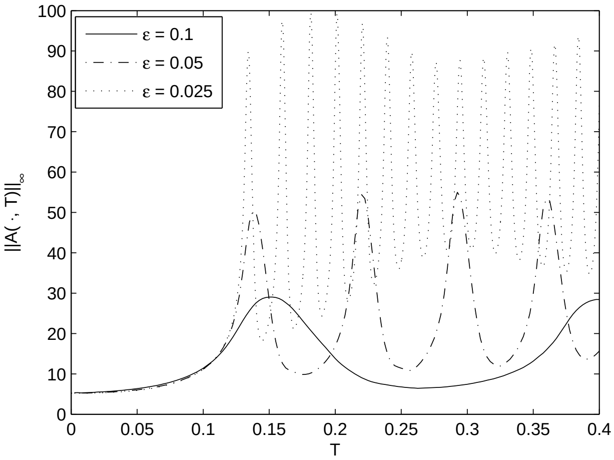


Figure 5.14: Time evolution of $\|A(\mathbf{X},T)\|_\infty$ for the perturbed NLS (5.12) with initial data (5.94) when $N = 50$ for different ε.

(iv). For fixed $N \geq 0$ and ε, the dynamics of $\|A(\mathbf{X},T)\|_\infty$ undergoes focusing-defocusing cycles (cf. Fig. 5.13).

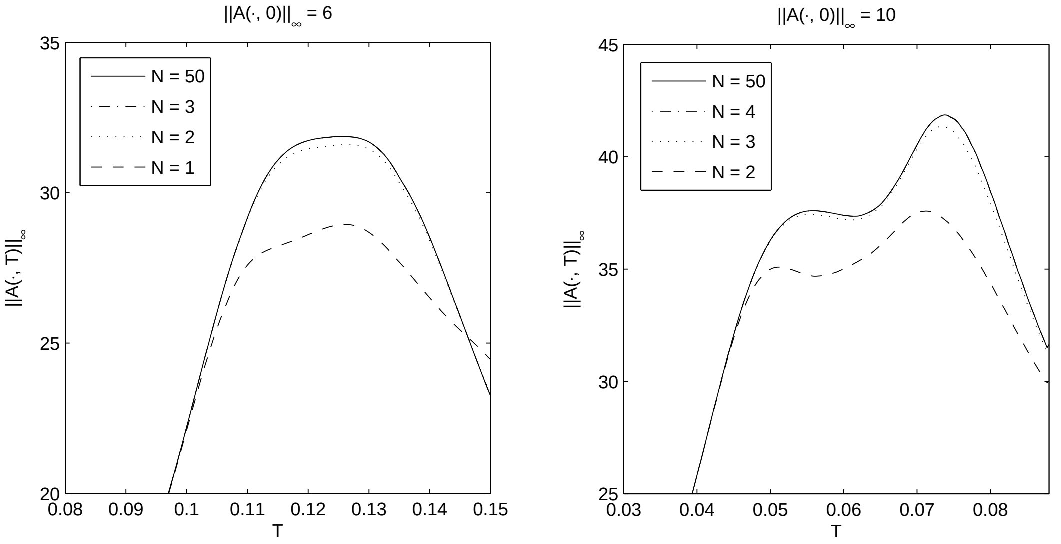


Figure 5.15: Time evolution of $\|A(\mathbf{X},T)\|_\infty$ for the perturbed NLS (5.12) with initial data (5.94) for different N and $\varepsilon = 0.1$.

5.3.4 Propagation of light bullets in perturbed NLS

From the above results, one can conclude that the perturbed NLS equation (5.12) with reasonably large N (at this point we take $N = 5$ in view of the initial amplitude considered here) agrees with the SG equation very well for modeling propagating pulses. Noticing that solving the perturbed NLS equation requires much less computation load than the SG equation due to the disparate scales involved and propagating property of the SG-LBs, we hence solve the perturbed NLS equation here to study the propagation of LBs instead of simulating the SG equation. The initial data in (5.6) is chosen as

$$A^{(0)}(\mathbf{X}) = ia_0 \exp\left(-\frac{X^2}{\sigma_x^2} - \frac{Y^2}{\sigma_y^2}\right), \quad \mathbf{X} \in \mathbb{R}^2, \tag{5.97}$$

with $\sigma_x = \omega$ and $\sigma_y = 1$. Note that such initial data has been extensively used in previous studies [127, 149] via solving the SG equation directly. The results below are reported for $a_0 = 3.5$, $\varepsilon = 0.2$, and $k = 2,\ 5$. The results for other sets of parameters are quite similar and omitted here for brevity.

Fig. 5.16 shows the top view of pulse for $k = 2$, propagating far beyond the

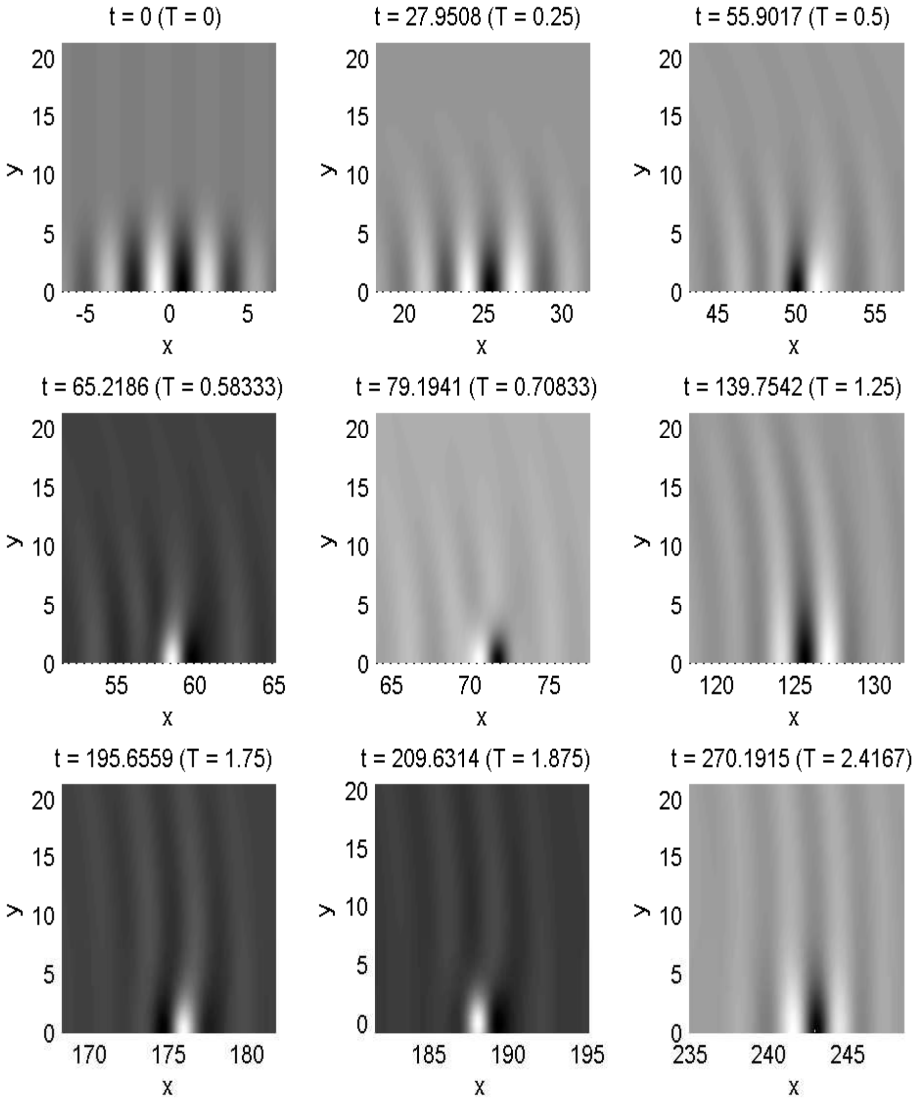


Figure 5.16: Top view of u^{nls} in perturbed NLS (5.12) with reasonable large $N = 5$, for $\varepsilon = 0.2$ and initial data (5.97) with $a_0 = 3.5$ and $k = 2$: propagation far beyond critical NLS collapse time $T^c \approx 0.6980$.

critical NLS collapse time and Fig. 5.17 depicts similar results for $k = 5$, which indicate that: (i) over time, the envelope tends to expand along y-axis slightly; (ii) before the collapse time the outside edge moves at a slower velocity than the

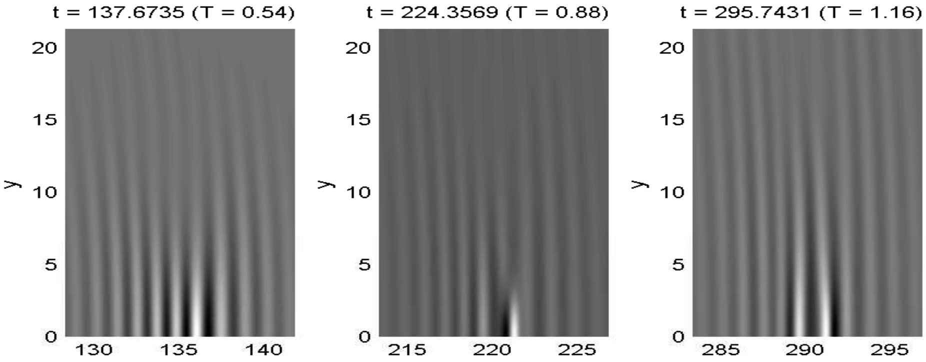


Figure 5.17: Top view of u^{nls}, same parameters as Fig. 5.16, except that $k = 5$ and critical NLS collapse time is $T^c \approx 0.7280$.

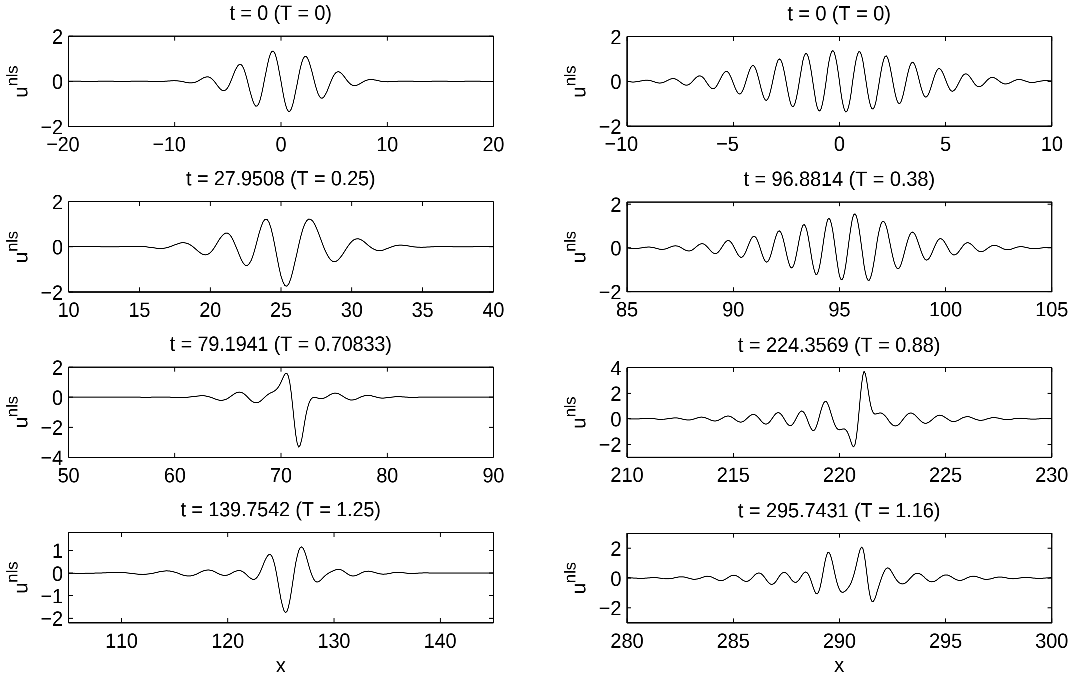


Figure 5.18: Slice plots of u^{nls} along x-axis with $y = 0$: (1) left column, pulses in Fig. 5.16, i.e. $k = 2$; (2) right column, pulses in Fig. 5.17, i.e. $k = 5$.

centerline of the envelope; and (iii) close to the collapse time, the envelope turns to be unstable (observed better in Figs. 5.18 and 5.9) and focus along x-axis, and the central part tends to delay which can be explained by the focusing mechanism taking effect in the perturbed NLS or observed in the profile of solution A of the perturbed NLS (this phenomena can also be observed in Fig. 5.8); (iv) after the collapse time and before the focusing takes strong effect again in NLS, the envelope moves in a similar pattern to that before the collapse time, except that most pulse energy concentrates at the central part (cf. Fig. 5.18, where pulse profiles along x-axis are plotted, and also Fig. 5.12). Changing the envelope wave number k, we observe similar results.

Chapter 6

Concluding remarks and future work

This thesis is devoted to numerical methods, their analysis and their applications, for some classes of nonlinear dispersive equations, namely the Schrodinger–Poisson–Slater (SPS) equation (1.1)–(1.2), the nonlinear relativistic Hartree equation (1.8), the nonlinear Klein–Gordon (KG) equation (1.12) in the nonrelativistic limit regime, the sine–Gordon (SG) equation (1.14) and perturbed NLS equation (1.16) for modeling 2D light bullets. In the sequel, results obtained for these subjects will be summarized, and possible topics for future work will be also discussed.

On the SPS equation

In the first part of Chapter 2, the numerics of the SPS equation (1.1)–(1.2) in all space dimensions (1D, 2D and 3D) were considered. To compute the ground states and dynamics of the SPS equation, a backward Euler sine/Fourier pseudospectral method and a time-splitting sine/Fourier pseudospectral method were proposed and applied with different approaches approximating the Hartree potential. The approaches considered here include: (1) fast convolution algorithms to evaluate the convolution of Laplacian kernel with density, with the help of FFT in 1D and fast multipole method (FMM) in higher dimensions; (2) a sine pseudospectral method to discretize a Poisson equation with homogeneous Dirichlet boundary conditions;

and (3) a Fourier pseudospectral method to discretize a Poisson equation with periodic boundary conditions. For the third approach, due to the inconsistency in 0-mode after taking Fourier transformation, the error from the truncated computation domain dominates the whole process, and the approximation converges as the domain is chosen larger. This can be illustrated from the formulation of method as well as numerical experiment results. Detailed numerical comparisons also showed that in 1D the fast convolution and sine pseudospectral approaches are compatible, both achieving spectral order of accuracy in space, while in 3D the fast convolution based on FMM, where linear interpolation is applied, is second-order accurate and the Fourier pseudospectral approach is better than it in both efficiency and accuracy. Therefore, the sine pseudospectral approach is the best choice among all the ones discussed here. As a benefit of such observations, the backward Euler and time-splitting time integrations with sine pseudospectral spatial discretization were applied to compute the ground states and dynamics of 3D SPS equation in various setups.

In the second part of Chapter 2, the focus was put on the scenario where the SPS equation is of spherical symmetry. In the spherically symmetric case, the sine pseudospectral discretizations, proposed for general external potential and initial condition in 3D, were simplified. The simplification is achieved by introducing a proper change of variables into the reduced quasi-1D model. The simplified methods still admit spectral order of accuracy in space, with significantly less demand on memory and computational load, and are more efficient in implementation than the standard finite difference approaches for the spherically symmetric case.

It should be commented that the methods proposed in Chapter 2 cannot be extended to the governing Schrödinger–Poisson (SP) type systems with discontinuous coefficients, for example, the relevant systems arising in semiconductor area. This is because the derivation and high-order accuracy of the spectral-type spatial discretization exclusively depend on the high regularity of functions. When discontinuity occurs, other methods, like finite difference, finite element or spectral element, would be the potential alternatives, and the detailed investigation is of course an interesting topic for further studies. However, the results in Chapter 2 still shed some light on the choice of spatial discretization in further numerical studies of the coupled SP type systems arising in quantum physics area; for example, the study in Chapter 3 serves as one application of these results.

As discussed in Chapter 2, the sine pseudospectral discretization, which solves the Poisson equation in SP type systems to approximate the nonlocal Hartree potential, is merely applicable in 1D and 3D. This is because the derivation of sine pseudospectral discretization in space highly relies on the boundary conditions put after the truncation, i.e. the homogeneous Dirichlet boundary conditions, which cannot be approximately assumed to be true in 2D. It is thus interesting to find appropriate artificial boundary conditions of the Hartree potential in 2D, and derive efficient numerical methods. Another interesting topic is about the numerical analysis in 1D, 2D and 3D. In practical computation, the time-splitting pseudospectral method has shown its high efficiency and accuracy for the SP type equation; thus, it is favorable to carry out error estimates for the time-splitting methods. Noting that all the convergence results reported in the literature (cf. [29, 105]) deal with the semi-discretization, the next step forward would be to understand the full discretization.

On the relativistic Hartree equation

In Chapter 3, efficient and accurate numerical methods were proposed for computing the ground states and dynamics of the nonlinear relativistic Hartree equation (1.8), which also refers to the relativistic SP equation, with both general and spherically symmetric solutions. The main challenge in the numerics lies in discretizing the pseudodifferential kinetic operator in 3D, which arises in special relativity. In general, the usual finite difference spatial discretizations cost much more memory load and/or computation time. In the proposed methods, the sine pseudospectral approach is applied in spatial discretization, and the kinetic operator is then approximated by multiplying its eigenvalue in phase space, which is analogous to its definition in continuous level. With this spatial discretization, a backward Euler sine pseudospectral (BESP) method was proposed to discretize a gradient flow with discrete normalization for computing the ground states. And similar to Chapter 2, in particular, when the system has spherical symmetry, a BESP method was given based on a reduced quasi-1D problem. For dynamics, a time-splitting sine pseudospectral discretization was proposed for general and spherically symmetric solutions. Numerical tests demonstrated that the methods are spectrally accurate in space, less demanding on memory and efficiently solvable. Applications of the methods in various setups were also reported.

It would be worthwhile to point out that in numerical experiments, some intriguing properties of boson stars, which can be modeled by the relativistic Hartree equation in its mean-field limit, were observed. For example, the monotone of each component in the energy in ground states with respect to single particle mass, similar monotonic property in "gravitational collapse" time, and the damping phenomena in the dynamics of the center of mass were observed. Motivated by these numerical observations, it would be highly desirable to carry out mathematical analysis which can give rigorous explanations to these interesting properties. On the other hand, rigorous error estimates for the time-splitting methods solving the SP equation involving relativistic effects are of great interests; however, there are few results on this topic in the literature. Although similar issues, in both mathematical and numerical analysis aspects, have been settled for nonrelativistic problems, many challenges remain when relativistic considerations are included, due to, for example, the appearance of the pseudodifferential kinetic operator in (1.8).

On the nonlinear KG equation

In Chapter 4, two classes of numerical methods with different time integrations were analyzed rigorously and compared numerically for solving the KG equation

(1.12) in the nonrelativistic limit regime, i.e. if $0 < \varepsilon \ll 1$ or the speed of light goes to infinity. The first class are the standard second-order finite difference time domain (FDTD) methods. For FDTD methods, including energy conservative/non-conservative and implicit/semi-implicit/explicit ones, error estimates were rigorously carried out, which show that their ε-scalability is $\tau = O(\varepsilon^3)$ with ε-independent h. The second class are based on applying the Gautschi-type exponential wave integrator for time discretization, which is combined with either sine pseudospectral (Gautschi-SP) or finite difference (Gautschi-FD) discretization in space. For the linear KG equation, the Gautschi-type time integration does not introduce error in time discretization. In addition, rigorous error estimates suggest that the ε-scalability of Gautschi-SP and Gautschi-FD is improved to $\tau = O(1)$ and $\tau = O(\varepsilon^2)$ for the linear and nonlinear KG equations, respectively. Comparison between Gautschi-SP and Gautschi-FD also indicate that this temporal resolution competence of Gautschi-type methods is independent of the spatial discretization it combines with. Hence, Gautschi-SP performs the best among all the methods discussed here in both non-relativistic limit regime and $O(1)$-speed of light regime.

All the numerical methods discussed in Chapter 4 solve the KG equation involving the highly oscillatory scaling in a standard flow, i.e. exactly the same manner as they are applied for non-oscillatory problems. It is thus quite imperative to propose more sophisticated numerical methods for this subject, which are expected to be based on the insight into the asymptotic behavior of solutions as taking the nonrelativistic limit. In future, we will investigate some multiscale methods, which would be based on suitable frequency-decomposition (scale-separation), in the spirit of some recent theoretical advances on this subject [106, 107, 110]. It is expected that the new multiscale methods would achieve higher resolution capacity for the oscillation.

On the SG and perturbed NLS equations

In Chapter 5, numerical comparisons were carried out among the solutions of the SG equation (1.14), the perturbed NLS equation (1.16) with its finite-term nonlinearity approximations, and the critical cubic NLS equation ($\varepsilon = 0$ in (1.16)), for the propagation of 2D light bullets (LBs) in nonlinear optical media. This was achieved by efficient semi-implicit sine pseudospectral methods, which can be rigorously proved to be spectrally accurate in space, second-order in time, and are very efficient in practical implementation. Based on extensive numerical comparison results, the conclusions are summarized as follows, provided that ε is reasonably small:

(i). If there is no finite time collapse in the cubic NLS equation, both the cubic NLS-LBs and the perturbed NLS-LBs agree with the SG-LBs qualitatively and quantitatively.

(ii). If the cubic NLS equation collapses in finite time, then in the time regime well before the collapse time, both the cubic NLS-LBs and the perturbed NLS-LBs again agree with the SG-LBs qualitatively and quantitatively; in the time regime near the collapse time, the cubic NLS-LBs fail to approximate the SG-LBs neither quantitatively nor qualitatively whereas the perturbed NLS-LBs agree with the SG-LBs both qualitatively and quantitatively; and in the time regime beyond the collapse time, the LBs of the perturbed NLS equation, with finite terms in the nonlinearity, still agree with the SG-LBs both qualitatively and quantitatively.

(iii). To well approximate the SG-LBs, the number of nonlinearity terms in the perturbed NLS equation depends on the initial data yet is independent of the small parameter ε. In general, only a few terms, e.g. $N \geq 3$, are needed in the perturbed NLS equation in practical computation.

Consequently, solving the perturbed NLS equations with reasonably many nonlinear terms demands much less computational load than simulating the SG equation directly due to the disparate scales involved. The computational domain for the SG equation also needs to be adaptively extended if the propagation to a further time point is desired. Thus, the perturbed NLS equation is a more efficient model for numerically tracking the propagation of LBs in 2D. In future, we propose to investigate the propagation of 3D LBs, which can also be modeled by the NLS type equations (cf. [152] and references therein).

Bibliography

[1] M.J. Ablowitz, M.B. Herbst, C.M. Schober, *On the numerical solution of the sine–Gordon equation: I. Integrable discretizations and homoclinic manifolds*, J. Comput. Phys. 126 (1996) 299–314.

[2] M.J. Ablowitz, M.B. Herbst, C.M. Schober, *On the numerical solution of the sine–Gordon equation: II. Performance and numerical schemes*, J. Comput. Phys. 131 (1997) 354–367.

[3] M.J. Ablowitz, M.D. Kruskal, J.F. Ladik, *Solitary wave collisions*, SIAM J. Appl. Math. 36 (1979) 428–437.

[4] G. Adomian, *Nonlinear Klein–Gordon equation*, Appl. Math. Lett. 9 (1996) 9–10.

[5] A. Aftalion, Q. Du, *Vortices in a rotating Bose-Einstein condensate: Critical angular velocities and energy diagrams in the Thomas-Fermi regime*, Phys. Rev. A 64 (2001) article 063603.

[6] G.P. Agrawal, *Nonlinear Fiber Optics*, Academic Press, New York, 1989.

[7] M. Agrotis, N.M. Ercolani, S.A. Glasgow, J.V. Moloney, *Complete integrability of the reduced Maxwell–Bloch equations with permanent dipole*, Physica D 138 (2000) 134–162.

[8] N. Akhmediev, A. Ankiewicz, *Does the nonlinear Schrödinger equation correctly describe beam propagation?*, Opt. Lett. 18 (1993) 411–413.

[9] G.L. Aki, P.A. Markowich, C. Sparber, *Classical limit for semirelativistic Hartree systems*, J. Math. Phys. 49 (2008) 102–110.

[10] N. Angelescu, M. Pulvirenti, A. Teta, *Derivation and classical limit of a mean field equation for a quantum Coulomb system: Maxwell–Boltzmann statistics*, J. Stat. Phys. 74 (1994) 147–165.

[11] D.D. Baǐnov, E. Minchev, *Nonexistence of global solutions of the initial-boundary value problem for the nonlinear Klein–Gordon equation*, J. Math. Phys. 36 (1995) 756–762.

[12] C. Bardos, L. Erdös, F. Golse, N.J. Mauser, H.T. Yau, *Derivation of the Schrödinger-Poisson equation from the quantum N-particle Coulomb problem*, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 515–520.

[13] C. Bardos, F. Golse, N.J. Mauser, *Mean field dynamics of fermions and the time-dependent Hartree-Fock equation*, J. d. Mathématiques Pures et Appl. 82 (2003) 665–683.

[14] C. Bardos, A. Gottilieb, F. Golse, N.J. Mauser, *Derivation of the time-dependent Hartree-Fock equation: the Coulomb interaction case*, preprint.

[15] W. Bao, Y. Cai, *Optimal error estimates of finite difference methods for the Gross-Pitaevskii equation with angular momentum rotation*, Math. Comp, in press.

[16] W. Bao, Y. Cai, H. Wang, *Efficient numerical methods for computing ground states and dynamics of dipolar Bose-Einstein condensates*, J. Comput. Phys. 229 (2010) 7874–7892.

[17] W. Bao, I-L. Chern, F.Y. Lim, *Efficient and spectrally accurate numerical methods for computing ground and first excited states in Bose-Einstein condensates*, J. Comput. Phys. 219 (2006) 836–854.

[18] W. Bao, Q. Du, *Computing the ground state solution of Bose-Einstein condensates by a normalized gradient flow*, SIAM J. Sci. Comput. 25 (2004) 1674–1697.

[19] W. Bao, D. Jaksch, *An explicit unconditionally stable numerical method for solving damped nonlinear Schrödinger equations with a focusing nonlinearity*, SIAM. J. Numer. Anal. 41 (2003) 1406–1426.

[20] W. Bao, S. Jin, P.A. Markowich, *On time-splitting spectral approximations for the Schrödinger equation in the semiclassical regime*, J. Comput. Phys. 175 (2002) 487–524.

[21] W. Bao, S. Jin, P.A. Markowich, *Numerical studies of time-splitting spectral discretizations of nonlinear Schrödinger equations in the semiclassical regime*, SIAM J. Sci. Comput. 25 (2003) 27–64.

[22] W. Bao, X. Li, *An efficient and stable numerical method for the Maxwell–Dirac system*, J. Comput. Phys. 199 (2004) 663–687.

[23] W. Bao, N.J. Mauser, H.P. Stimming, *Effective one particle quantum dynamics of electrons: A numerical study of the Schrödinger-Poisson-Xα model*, Comm. Math. Sci. 1 (2003) 809-831.

[24] W. Bao, F.F. Sun, *Efficient and stable numerical methods for the generalized and vector Zakharov system*, SIAM J. Sci. Comput. 26 (2005) 1057–1088.

[25] W. Bao, L. Yang, *Efficient and accurate numerical methods for the Klein–Gordon-Schrödinger equations*, J. Comput. Phys. 225 (2007) 1863–1893.

[26] W. Bao, Y. Zhang, *Dynamics of the ground state and central vortex states in Bose-Einstein condensation*, Math. Models Meth. Appl. Sci. 15 (2005), 1863–1896.

[27] N. Ben Abballan, P. Degond, P.A. Markowich, *On a one-dimensional Schrödinger–Poisson scattering model*, ZAMP 48 (1997) 35–55.

[28] L. Bergé, T. Colin, *A singular perturbation problem for an envelope equation in plasma physics*, Physica D 84 (1995) 437–459.

[29] C. Besse, B. Bidégaray, S. Descombes, *Order estimates in time of splitting mehotds for the nonlinear Schrödinger equation*, SIAM J. Numer. Anal. 40 (2002) 26–40.

[30] O. Bokanowski, B. Grébert, N.J. Mauser, *Local density approximation for the energy of a periodic Coulomb model*, Math. Models Meth. Appl. Sci. 13 (2003) 1185–1217.

[31] O. Bokanowski, N.J. Mauser, *Local approximation for the Hartree-Fock exchange potential: a deformation approach*, Math. Models Meth. Appl. Sci. 9 (1999) 941–961.

[32] O. Bokanowski, J.L. López, J. Soler, *On a exchange interaction model for quantum transport: The Schrödinger-Poisson-Slater system*, Math. Models Meth. Appl. Sci. 12 (2003) 1397–1412.

[33] J. Bourgain, *Global Solutions of Nonlinear Schrödinger Equations*, AMS, 1999.

[34] T. Brabec, F. Krausz, *Nonlinear optical pulse propagation in the single-cycle regime*, Phys. Rev. Lett. 78 (1997) 3282–3286.

[35] F.E. Browder, *On nonlinear wave equations*, Math. Z. 80 (1962) 249–264.

[36] P. Brenner, W. van Wahl, *Global classical solutions of nonlinear wave equations*, Math. Z. 176 (1981) 87–121.

[37] W. Cao, B. Guo, *Fourier collocation method for solving nonlinear Klein–Gordon equation*, J. Comput. Phys. 108 (1993) 296–305.

[38] T. Cazenave, *Semilinear Schröinger Equations*, Courant Lecture Notes in Mathematics, vol. 10, New York University, AMS, 2003.

[39] C. Cheng, Q. Liu, J. Lee, H.Z. Massoud, *Spectral element method for the Schrödinger–Poisson system*, J. Comput. Electron. 3 (2004) 417–421.

[40] H. Cheng, L. Greengard, V. Rokhlin, *A fast adaptive multipole algorithm in three dimensions*, J. Comput. Phys. 155 (1999) 468–498.

[41] M.L. Chiofalo, S. Succi, M.P. Tosi, *Ground state of trapped interacting Bose-Einstein condensates by an explicit imaginary-time algorithm*, Phys. Rev. E 62 (2000) 7438–7444.

[42] Y. Cho, T. Ozawa, *On the semirelativistic Hartree-type equation*, SIAM J. Math. Anal. 38 (2006) 1060–1074.

[43] P. Choquard, J. Stubbe, M. Vuffray, *Stationary solutions of the Schrödinger–Newton model–an ODE approach*, Diff. Int. Eqns. 21 (2008) 665–679.

[44] D. Cohen, E. Hairer, Ch. Lubich, *Conservation of energy, momentum and actions in numerical discretizations of non-linear wave equations*, Numer. Math. 110 (2008) 113–143.

[45] K.T.R. Davies, H. Flocard, S. Krieger, M.S. Weiss, *Application of the imaginary time step method to the solution of the static Hartree-Fock problem*, Nucl. Phys. A 342 (1980) 111–123.

[46] A.S. Davydov, *Quantum Mechanics, 2nd Edition*, Pergamon, Oxford, 1976.

[47] E.Y. Deeba, S.A. Khuri, *A decomposition method for solving the nonlinear Klein–Gordon equation*, J. Comput. Phys. 124 (1996) 442–448.

[48] M. Delfour, M. Fortin, G. Payne, *Finite difference solution of a nonlinear Schrödinger equation*, J. Comput. Phys. 44 (1981) 277–288.

[49] P.A.M. Dirac, *Note on exchange phenomena in the Thomas-Fermi atom*, Proc. Cambridge Pilos. Soc. 26 (1931) 376–385.

[50] P. Donnat, J. Rauch, *Global solvability of the Maxwell–Bloch equations from nonlinear optics*, Arch. Rat. Mech. Anal. 136 (1996) 291–303.

[51] R.M. Dreizler, E.K.U. Gross, *Density Functional Theory*, Springer, Berlin, 1990.

[52] D.B. Duncan, *Symplectic finite difference approximations of the nonlinear Klein–Gordon equation*, SIAM J. Numer. Anal. 34 (1997) 1742–1760.

[53] M. Edwards, K. Burnett, *Numerical solution of the nonlinear Schrödinger equation for small samples of trapped neutral atoms*, Phys. Rev. A 51 (1995) 1382–1386.

[54] M. Ehrhardt, A. Zisowsky, *Fast calculation of energy and mass preserving solutions of Schrödinger–Poisson systems on unbounded domains*, J. Comput. Appl. Math. 187 (2006) 1–28.

[55] A. Elgart, B. Schlein, *Mean field dynamics of Boson stars*, Comm. Pure Appl. Math. 60 (2007) 500–545.

[56] F. Ethridge, L. Greengard, *A new fast-multipole accelerated Poisson solver in two dimensions*, SIAM J. Sci. Comput. 23 (2001) 741–760.

[57] B. Fornberg, *A Practical Guide to Pseudospectral Methods*, Cambridge University Press, Cambridge, 1998.

[58] J. Fröhlich, B.L. Jonsson, E. Lenzmann, *Effective dynamics for boson stars*, Nonlinearity 20 (2007) 1031–1075.

[59] J. Fröhlich, E. Lenzmann, *Blowup for nonlinear wave equations describing boson stars*, Comm. Pure Appl. Math. 60 (2007) 1691–1705.

[60] G. Fibich, G. Papanicolaou, *Self-focusing in the perturbed and unperturbed nonlinear Schrödinger equation in critical dimension*, SIAM J. Appl. Math. 60 (2000) 183–240.

[61] G. Fibich, G. Papanicolaou, *A modulation method for self-focusing in the perturbed critical nonlinear Schrödinger equation*, Phys. Lett. A 239 (1998) 167–173.

[62] A. Gammal, T. Frederico, L. Tomio, *Improved numerical approach for the time-independent Gross–Pitaevskii nonlinear Schrödinger equation*, Phys. Rev. E 60 (1999) 2421–2424.

[63] W. Gautschi, *Numerical integration of ordinary differential equations based on trigonometric polynomials*, Numer. Math. 3 (1961) 381–397.

[64] J. Ginibre, G. Velo, *The global Cauchy problem for the nonlinear Klein–Gordon equation*, Math. Z. 189 (1985) 487–505.

[65] J. Ginibre, G. Velo, *The global Cauchy problem for the nonlinear Klein–Gordon equation–II*, Ann. Inst. H. Poincaré Anal. Non Linéaire 6 (1989) 15–35.

[66] L. Greengard, V. Rokhlin, *A fast algorithm for particle simulations*, J. Comput. Phys. 73 (1987) 325–348.

[67] L. Greengard, V. Rokhlin, *A new version of the fast multipole method for the Laplace equation in three dimensions*, Acta Numer. 6 (1997) 229–269.

[68] R. Glassey, *On the asymptotic behavior of nonlinear wave equations*, Trans. Amer. Math. Soc. 182 (1973) 187–200.

[69] R. Glassey, M. Tsutsumi, *On uniqueness of weak solutions to semi-linear wave equations*, Commun. Partial Diff. Eqn. 7 (1982) 153–195.

[70] P. Goorjian, Y. Silberberg, *Numerical simulations of light bullets using the full-vector time-dependent nonlinear Maxwel equations*, J. Opt. Soc. Am. B (Optical Physics) 14 (1997) 3253–3260.

[71] D. Gottlieb, S.A. Orszag, *Numerical Analysis of Spectral Methods : Theory and Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 1993.

[72] V. Grimm, *A note on the Gautschi-type method for oscillatory second-order differential equations*, Numer. Math. 102 (2005) 61–66.

[73] V. Grimm, *On error bounds for the Gautschi-type exponential integrator applied to oscillatory second-order differential equations*, Numer. Math. 100 (2005) 71–89.

[74] V. Grimm, *On the use of the Gautschi-type exponential integrator for wave equation*, In: Numerical Mathematics and Advanced Applicaions (ENUMATH2005), editors: A. Bermúdez de Castro, D. Gómez, P. Quintela and P. Salgado, Springer-Verlag Berlin Heidelberg, 2006, pp. 557–563.

[75] B. Guo, P.J. Pascual, M.J. Rodriguez, L. Vázquez, *Numerical solution of the sine–Gordon equation*, Appl. Math. Comput. 18 (1986) 1–14.

[76] J. Hahm, C.M. Lieber, *Direct ultrasensitive electrical detection of DNA and DNA sequence variations using nanowire nanosensors*, Nano Lett. 4 (2004) 51–54.

[77] E. Hairer, Ch. Lubich, G. Wanner, *Geometric Numerical Integration*, Springer-Verlag, 2002

[78] R. Harrison, I.M. Moroz, K.P. Tod, *A numerical study of Schrödinger–Newton equations*, Nonlinearity 16 (2003) 101–122.

[79] A. Hasegawa, *Optical Solitons in Fibers*, Springer, New York, 1989.

[80] J.S. Hesthaven, S. Gottlieb, D. Gottlieb, *Spectral Methods for Time-Dependent Problems*, Cambridge University Press, Cambridge, New York, 2007.

[81] C.V. Hile, *Comparisons between Maxwell's equations and an extended nonlinear Schrödinger equation*, Wave Motion 24 (1996) 1–12.

[82] C.V. Hile, W. L. Kath, *Numerical solutions of Maxwell's equations for nonlinear optical pulse propagation*, J. Opt. Soc. Am. B 13 (1996) 1135–1146.

[83] M. Hochbruck, Ch. Lubich, *A Gautschi-type method for oscillatory second-order differential equations*, Numer. Math. 83 (1999) 402–426.

[84] M. Hochbruck, A. Ostermann, *Exponential integrators*, Acta Numer. 19 (2000) 209–286.

[85] T. Hrycak, V. Rokhlin, *An improved fast multipole algorithm for potential fields*, SIAM J. Sci. Comput. 19 (1998) 1804–1826.

[86] H. Holden, K.H. Karlsen, K-A. Lie, N.H. Risebro, *Splitting Methods for Partial Differential Equations with Rough Solutions: Analysis and MATLAB Programs*, European Mathematical Society, Zürich, 2010.

[87] S. Ibrahim, M. Majdoub, N. Masmoudi, *Global solutions for a semilinear, two-dimensional Klein–Gordon equation with exponential-type nonlinearity*, Comm. Pure Appl. Math. 59 (2006) 1639–1658.

[88] S. Jiménez, L. Vázquez, *Analysis of four numerical schemes for a nonlinear Klein–Gordon equation*, Appl. Math. Comput. 35 (1990) 61–94.

[89] R.M. Joseph, S.C. Hagness, A. Taflove, *Direct time integration of Maxwell's equations in linear dispersive media with absorption for scattering and propagation of femtosecond electromagnetic pulses*, Opt. Lett. 16 (1991) 1412.

[90] A.E. Kaplan, P. Shkolnikov, *Electromagnetic "bubbles" and shock waves: unipolar, nonoscillating EM solitons*, Phys. Rev. Lett., 75 (1995) 2316–2319.

[91] A.E. Kaplan, P. Shkolnikov, *Subfemtosecond high-intensity unipolar electromagnetic soliton and shock waves*, J. Nonlinear Opt. Phys. Mater. 4 (1995) 831–841.

[92] O.B. Kellogg, *Foundations of Potential Theory*, Dover, New York, 1953.

[93] P. Kirrmann, G. Schneider, A. Mielke, *The validity of modulation equations for the extended system with cubic nonlinearities*, Proc. Roy. Soc. Edin. 122A (1992) 85–91.

[94] R. Kosecki, *The unit condition and global existence for a class of nonlinear Klein–Gordon equations*, J. Diff. Eqn. 100 (1992) 257–268.

[95] L. Lehtovaara, J. Toivanen, J. Eloranta, *Solution of time-independent Schrödinger equation by the imaginary time propagation method*, J. Comput. Phys. 221 (2007) 148–157.

[96] E. Lenzmann, *Well-posedness for semi-relativistic Hartree equations of critical type*, Math. Phys. Anal. Geom. 10 (2007) 43–64.

[97] E. Lenzmann, *Uniqueness of ground states for pseudo-relativistic Hartree equations*, Analysis & PDE 1 (2009) 1–30.

[98] S. Li, L. Vu-Quoc, *Finite difference calculus invariant structure of a class of algorithms for the nonlinear Klein–Gordon equation*, SIAM J. Numer. Anal. 32 (1995) 1839–1875.

[99] E.H. Lieb, *Existence and uniqueness of the minimizing of Choquards' nonlinear equation*, Studies Appl. Math. 57 (1976/77) 93–105.

[100] E.H. Lieb, *Thomas-Fermi and related theories of atoms and molecules*, Rev. Modern Phys. 55 (1981) 603–641.

[101] E.H. Lieb, B. Simon, *The Thomas-Fermi theory of atoms, molecules, and solids*, Adv. Math. 23 (1977) 22–116.

[102] E.H. Lieb, H. Yau, *The Chandrasekhar theory of stellar collapse as the limit of quantum mechanics*, Comm. Math. Phys. 112 (1987) 147–174.

[103] F. Linares, G. Ponce, *Introduction to Nonlinear Dispersive Equations*, Springer, New York, 2009.

[104] P.L. Lions, *Solution of Hartree–Fock equations for Coulomb systems*, Comm. Math. Phys. 109 (1987) 33–97.

[105] Ch. Lubich, *On splitting methods for Schrödinger–Poisson and cubic nonlinear Schrödinger equations*, Math. Comp. 77 (2008) 2141–2153.

[106] S. Machihara, *The nonrelativistic limit of the nonlinear Klein–Gordon equation*, Funkcial. Ekvac. 44 (2001) 243–252.

[107] S. Machihara, K. Nakanishi, T. Ozawa, *Nonrelativistic limit in the energy space for nonlinear Klein–Gordon equations*, Math. Ann. 322 (2002) 603–621.

[108] G.I. Marchuk, *Splitting and alternating direction methods*, Handbook of Numerical Analysis, vol. I, North-Holland, Amsterdam, 1990, pp. 197–462.

[109] S. Masaki, *Energy solution to Schrödinger–Poisson system in the two-dimensional whole space*, preprint.

[110] N. Masmoudi, K. Nakanishi, *From nonlinear Klein–Gordon equation to a system of coupled nonlinear Schrödinger equations*, Math. Ann. 324 (2002) 359–389.

[111] N.J. Mauser, *The Schrödinger–Poisson–X^α equation*, Appl. Math. Lett. 14 (2001) 759–763.

[112] R. McLeod, K. Wagner, S. Blair, *(3+1)-dimensional optical soliton dragging logic*, Phys. Rev. A 52 (1995) 3254–3278.

[113] A. Minzoni, N. Smyth, A. Worthy, *Pulse evolution for a two dimensional sine–Gordon equation*, Physica D 159 (2001) 101–123.

[114] A. Minzoni, N. Smyth, A. Worthy, *Evolution of two-dimensional standing and traveling breather solutions for the sine–Gordon equation*, Physica D 189 (2004) 167–187.

[115] C. Morawetz, W. Strauss, *Decay and scattering of solutions of a nonlinear relativistic wave equation*, Comm. Pure Appl. Math. 25 (1972) 1–31.

[116] B. Najman, *The nonrelativistic limit of the nonlinear Klein–Gordon equation*, Nonlinear Anal. 15 (1990) 217–228.

[117] B. Najman, *The nonrelativistic limit of Klein–Gordon and Dirac equations*, Differential Equations with Applications in Biology, Physics, and Engineering (Leibnitz, 1989), pp. 291–299, Lecture Notes in Pure and Appl. Math., 133, Dekker, New York, 1991.

[118] M. Nakamura, T. Ozawa, *The Cauchy problem for nonlinear Klein–Gordon equations in the Sobolev spaces*, Publ. Res. Inst. Math. Sci. 37 (2001) 255–293.

[119] A. Newell, J. Moloney, *Nonlinear Optics*, Addison-Wesley, Redwood City, CA, 1991.

[120] B.G. Pachpatte, *Inequalities for Finite Difference Equations*, Monographs and Textbooks in Pure and Applied Mathematics, Marcel Dekker Inc., New York, 2002

[121] P.J. Pascual, S. Jiménez, L. Vázquez, *Numerical simulations of a nonlinear Klein–Gordon model*, Applications. Computational physics (Granada, 1994), pp. 211–270, Lecture Notes in Phys., 448, Springer, Berlin, 1995.

[122] J.A. Pava, *Nonlinear Dispersive Equations : Existence and Stability of Solitary and Periodic Travelling Wave Solutions*, AMS, Providence, RI. 2009.

[123] H. Pecher, *Nonlinear small data scaterring for the wave and Klein–Gordon equation*, Math. Z. 185 (1984) 261–270.

[124] R. Pego, *Origin of the KdV equation*, Notices AMS, 45 (1997) 358.

[125] R.D. Pierce, C.E. Wayne, *On the validity of mean-field amplitude equations for counter-propagating wavetrains*, Nonlinearity 8 (1995) 769–779.

[126] V.S. Popov, *Imaginary-time method in quantum mechanics and field theory*, Phys. At. Nucl. 68 (2005) 686–708.

[127] T. Povich, J. Xin, *A numerical study of the light bullets interaction in the (2+1) sine–Gordon equation*, J. Nonlinear Sci. 15 (2005) 11–25.

[128] J.J. Sakurai, *Advanced Quantum Mechanics*, Addison Wesley, New York, 1967.

[129] Ó. Sánchez, J. Soler, *Long-time dynamics of the Schrödinger–Poisson–Slater systems*, J. Statist. Phys. 114 (2004) 179–204.

[130] J.M. Sanz-Serna, *Methods for the numerical solution of the nonlinear Schrödinger equation*, Math. Comp. 43 (1984) 21–27.

[131] I.E. Segal, *The global Cauchy problem for a relativistic scalar field with power interaction*, Bull. Soc. Math. Fr. 91 (1963) 129–135.

[132] J. Shatah, *Normal forms and quadratic nonlinear Klein–Gordon equations*, Commun. Pure Appl. Math. 38 (1985) 685–696.

[133] J. Shen, T. Tang, *Spectral and High-Order Methods with Applications*, Science Press, Beijing, 2006.

[134] J.C.H. Simon, E. Taflin, *The Cauchy problem for non-linear Klein–Gordon equations*, Commun. Math. Phys. 152 (1993) 433–478.

[135] J.C. Slater, *A simplification of the Hartree-Fock method*, Phys. Rev. 81 (1951) 385–390.

[136] G.D. Smith, *Numerical Solution of Partial Differential Equations*, Oxford University Press, London, 1965.

[137] Y. Siberberg, *Collapse of optical pulses*, Opt. Lett. 15 (1990) 1282–1284.

[138] H.P. Stimming, *The IVP for the Schrödinger–Poisson–Xα equation in one dimension*, Math. Models Meth. Appl. Sci. 15 (2005) 1169–1180.

[139] G. Strang, *On the construction and comparison of difference schemes*, SIAM J. Numer. Anal. 5 (1968) 505–517.

[140] W. Strauss, *Decay and asymptotics for* $\Box\, u = f(u)$, J. Funct. Anal. 2 (1968) 409–457.

[141] W. Strauss, L. Vázquez, *Numerical solution of a nonlinear Klein–Gordon equation*, J. Comput. Phys. 28 (1978) 271–278.

[142] C. Sulem, P.-L. Sulem, *The Nonlinear Schrödinger Equation: Self-focusing and Wave Collapse*, AMS, 139; Springer-Verlag, New York, 1999.

[143] T. Tao, Nonlinear Dispersive Equations: Local and Global Analysis, CBMS Regional Conference Series in Mathematics, AMS, Providence, RI, 2006.

[144] M. Thalhammer, M. Caliari, C. Neuhauser, *High-order time-splitting Hermite and Fourier spectral methods*, J. Comput. Phys. 228 (2009) 822–832.

[145] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer, 1997.

[146] Y. Tourigny, *Product approximation for nonlinear Klein–Gordon equations*, IMA J. Numer. Anal. 9 (1990) 449–462.

[147] M. Tsutsumi, *Nonrelativistic approximation of nonlinear Klein–Gordon equations in two space dimensions*, Nonlinear Anal. 8 (1984) 637–643.

[148] L.N. Trefethen, *Spectral Methods in Matlab*, Socity of Industrial and Applied Mathematics, 2000.

[149] J. Xin, *Modeling light bullets with the two-dimensional sine–Gordon equation*, Physica D 135 (2000) 345–368.

[150] K.I. Yoshida, *Applications of Fast Multipole Method to Boundary Integral Equation Method*, Ph.D thesis, 2001.

[151] D.G. Zhao, J.F. Huang, Y. Xiang, *A new version Fast Multipole Method for evaluating the stress field of dislocation ensembles*, Modeling Simul. Mater. Sci. Eng. 18 (2010) 045006.

[152] W. Zhong, M. Belić, T. Huang, *Three-dimensional Bessel light bullets in self-focusing Kerr media*, Phys. Rev. A 82 (2010) article 033834.

List of Publications

[1] *Singular limits of Klein–Gordon–Schrödinger equations to Schrödinge–Yukawa equations* (with Weizhu Bao and Shu Wang), Multiscale Modeling and Simulation: a SIAM Interdisciplinary Journal, Vol. 8 (5), pp. 1742–1769, 2010.

[2] *Comparisons between sine–Gordon equation and perturbed nonlinear Schrödinger equations for modeling light bullets beyond critical collapse* (with Weizhu Bao and Jack Xin), Physica D: Nonlinear Phenomena, Vol. 239 (13), pp. 1120–1134, 2010.

[3] *On the computation of ground state and dynamics of Schrödinger–Poisson–Slater system* (with Yong Zhang), Journal of Computational Physics, Vol. 230 (7), pp. 2660–2676, 2011.

[4] *Numerical methods for computing ground states and dynamics of nonlinear relativistic Hartree equation for boson stars* (with Weizhu Bao), Journal of Computational Physics, Vol. 230 (10), pp. 5449–5469, 2011.

[5] *A short note on simplified pseudospectral methods for computing ground state and dynamics of spherically symmetric Schrödinger–Poisson–Slater system*, Journal of Computational Physics, Vol. 230 (22), pp. 7917–7922, 2011.

[6] *Analysis and comparison of numerical methods for the Klein–Gordon equation in the nonrelativistic limit regime* (with Weizhu Bao), Numerische Mathematik, Vol. 120 (2), pp. 189-229, 2012.

[7] *A fourth-order split-step pseudospectral scheme for the Kuramoto–Tsuzuki equation*, Communications in Nonlinear Science and Numerical Simulation, Vol. 17, pp. 3161–3168, 2012.

[8] *A trigonometric integrator pseudospectral discretization for the N-coupled nonlinear Klein–Gordon equations*, Numerical Algorithms, to appear.

[9] *An exponential wave integrator pseudospectral method for the Klein–Gordon–Zakharov system* (with Weizhu Bao), preprint.

[10] *Error estimates in the energy space for a Gautschi-type integrator spectral discretization for the coupled nonlinear Klein–Gordon equations*, preprint.

[11] *Numerical solutions of the symmetric regularized-long-wave equation by trigonometric integrator pseudospectral discretization*, preprint.