Karlsruhe Institute of Technology

Fakultät für Mathematik

The Method of Multiple Scales for nonlinear Klein-Gordon and Schrödinger Equations

Diploma Thesis of

Patrick Krämer

Duration: 06. June 2013 – 06. December 2013

Supervisor:	Prof. Dr. Marlis Hochbruck
Second advisor:	JProf. Dr. Katharina Schratz

At the Department of Mathematics

Institute of Applied
and Numerical Mathematics

Karlsruhe Institute of Technology

ACKNOWLEDGEMENTS

My thanks go to Marlis Hochbruck for supervising this thesis and for many valuable suggestions. I thank Katharina Schratz very much for her helpful feedback.
Thanks to David Hipp, Julian Ott, Mareike Rieger and Simone Buchholz for proof-reading and their suggestions on style issues.
I especially thank my parents and grandparents and all my family and friends for their support during my studies.

CONTENTS

CHAPTER

ONE

INTRODUCTION

Abstract

The main objective of this thesis is the derivation of an approximate solution to the nonlinear Klein-Gordon equation via the *method of multiple scales*. We may sometimes also call this technique the *multiple scale analysis*. This method follows the concept of expanding the solution into a perturbation series and including multiple temporal and spatial scales. The resulting approximation, proposed in [5] for a modified version of this equation, has the simple form of a plane wave which is modulated by some envelope function A, i.e.

$$\varepsilon A e^{i(kx-\omega t)} + \varepsilon A^* e^{-i(kx-\omega t)} \tag{1.1}$$

with small parameter $0 < \varepsilon \ll 1$. We will see that A satisfies the nonlinear Gross-Pitaevskii equation, a nonlinear Schrödinger equation (NLS) which plays an important role in physics, for example in nonlinear optics (see [6]). Hence, in order to obtain an approximate solution to the Klein-Gordon equation, it is sufficient to solve the Gross-Pitaevskii equation numerically. This can be done very efficiently by using the splitting method, extensively discussed in [16] by C. Lubich, which we want to present later in this thesis. This enables us to compute a feasible approximation in less time than by using methods to solve the Klein-Gordon equation directly instead. Thus it makes sense to use the multiple scale approximation if we want to obtain a good approximation to the nonlinear wave equation in short time. Throughout the thesis we call (1.1) the *NLS approximation*.

Furthermore we will present a powerful splitting integrator which we use to solve the Klein-Gordon equation numerically. This integrator is very efficient and features nice numerical properties such as norm and energy conservation even for large time step sizes.

Outline

In **chapter 2** we want to gain insight into the method of multiple scales and explain it on the basis of two ordinary differential equations. We will discuss resonance effects which lead to so called *secular* terms, i.e. terms that become unbounded as time advances. Dealing with these terms is a main part in the method of multiple scales, therefore we will encounter them at many points in this thesis.

Chapter 3 treats the application of the multiple scale method to the nonlinear Klein-Gordon equation. Furthermore we will give some basic properties of the Gross-Pitaevskii equation and prove an error bound for the NLS approximation.

The tools for the numerical solution of the Gross-Pitaevskii and the Klein-Gordon equation are introduced in **chapter 4**. We present Strang splitting methods for both equations and adapt the implicit midpoint rule and the leapfrog method to the Klein-Gordon equation.

Using these integrators we do some numerical tests in **chapter 5** in order to verify the NLS approximation numerically. Furthermore we show the norm and energy conservation properties of the integrators and their numerical order.

Some Notational Remarks for this Thesis

For easier reading we may omit the spatial argument and just write $v(t)$ instead of $v(x,t)$ for some function $v : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Furthermore the L^2 norm is always understood with respect to the spatial variable, i.e.

$$\|v(\cdot,t)\|_{L^2} = \|v(t)\|_{L^2}\,.$$

The complex conjugate of a number $z \in \mathbb{C}$ is denoted by $\overline{z}$, where we may sometimes also use the notation A^* for $A \in \mathbb{C}$ instead of $\overline{A}$. We abbreviate the complex conjugate of a preceding term by *c.c.*, e.g.

$$\varepsilon A e^{i(kx-\omega t)} + c.c. \quad = \quad \varepsilon A e^{i(kx-\omega t)} + \varepsilon A^* e^{-i(kx-\omega t)}.$$

Furthermore whenever we consider terms f_ε depending on a small parameter $0 < \varepsilon \ll 1$, we are interested in the limit case $\varepsilon \to 0$.

In the caption of some figures in this thesis we may find a reference to the MATLAB files, which we have used to create the figure.

CHAPTER

TWO

MULTIPLE SCALE ANALYSIS

The notion of *multiple scale analysis* or the *method of multiple scales* describes a method to analyze the influence of weak nonlinearities in ordinary and partial differential equations on the corresponding solution. In this chapter we want to give a short introduction to multiple scale analysis and present the method on the basis of some ordinary differential equations.

For illustrating how a multiple scale analysis is done, we will have a look at the ordinary differential equation

$$\ddot{y} + 2\varepsilon\dot{y} + y = 0, \quad y(0) = \alpha \in \mathbb{R}, \dot{y}(0) = 0,$$

that describes the movement of a lightly damped harmonic oscillator with small damping parameter $0 < \varepsilon \ll 1$, and afterwards at the weak nonlinear oscillator problem

$$\ddot{y} + y + \varepsilon y^3 = 0, \quad y(0) = 1, \qquad \dot{y}(0) = 0,$$

where the weak nonlinearity is given by εy^3 with small $0 < \varepsilon \ll 1$.

The aim is to write down an approximation to the solution as a truncated and hence finite perturbation series with small $0 < \varepsilon \ll 1$ such as

$$y(t) \coloneqq\approx \varepsilon^M Y_M(t, \varepsilon t, \varepsilon^2 t, \dots) + \sum_{k=0}^{M-1} \varepsilon^k Y_k(t, \varepsilon t, \varepsilon^2 t, \dots), \ M \in \mathbb{N}.$$

Making this ansatz for the solution of our problem will lead to a sequence of differential equations for the functions Y_k, which sheds light on how to choose the functions Y_k and the index M to truncate the series.

Furthermore we are going to encounter secular terms, that arouse resonance effects and unbounded growth, so they are undesirable. They play a major role throughout this thesis. Hence we will describe these terms first in the successive section on basis of [17], chapter 4.1, [4] chapter 11.1 and 11.2 and with the help of [21] chapter 7.

Please note that there is no generally valid error analysis for this method. Therefore it has to be analyzed anew when applying it to a different problem. Fortunately the procedure is always the same.

2.1. Resonance and Secular Behaviour

If we speak about oscillatory problems, resonance effects are a crucial topic that can not be ignored. They may occur for instance when a driving force is present in our system. Think of pushing a child on a swing. If your pushing at the same frequency as the child swings, its amplitude will become larger and larger. This effect has to be avoided.

In order to understand what this is about mathematically, let us consider the following example of a harmonic oscillator of natural frequency ω_0 that is taken from chapter 11.1 of [4].

Example 2.1.

The movement of a harmonic oscillator, e.g. a pendulum or a spring with a punctual weight, can be represented by the homogeneous ordinary differential equation

$$\ddot{y}(t) + \omega_0^2 y(t) = 0, \tag{2.1}$$

where $y(t)$ shall be the displacement of the oscillator at time t and ω_0 the natural frequency of the system.

Its general solution is well known and given by

$$y_h(t) = A\cos(\omega_0 t) + B\sin(\omega_0 t)$$

with arbitrary constants $A, B \in \mathbb{R}$. Since for all $t \in \mathbb{R}$ $|\cos(\omega_0 t)| \leq 1$ and $|\sin(\omega_0 t)| \leq 1$ we can bound

$$|y_h(t)| \leq A + B.$$

Now if we include a driving force, that periodically puts energy into the system at frequency ω, we get a right hand side inhomogeneity in (2.1) and the oscillation can be represented by

$$\ddot{y}(t) + \omega_0^2 y(t) = \cos(\omega t). \tag{2.2}$$

Its general solution depends on the relation between the driving frequency ω and the system's natural frequency ω_0.

This driving force can be for example an electromagnet, that is placed under the spring and that creates a periodically varying field to influence the oscillation of a magnetic mass hanging on the spring.

But what happens if $|\omega| \to |\omega_0|$, i.e. if the driving frequency gets close to the natural frequency of the system?

In the case $|\omega| \neq |\omega_0|$ we find the general solution of (2.2):

$$y(t) = A\cos(\omega_0 t) + B\sin(\omega_0 t) + \frac{\cos(\omega t)}{\omega_0^2 - \omega^2}. \tag{2.3}$$

As we can see, the denominator of the last summand gets close to 0 as $|\omega| \to |\omega_0|$. Since $\cos(\omega t)$ is bounded, the amplitude of the oscillation thus increases more and more. This can be explained in physikcal terms by the system absorbing more and more energy from the external force when the driving frequency ω gets close to the natural frequency ω_0 of the system.

But nevertheless we observe that for all fixed $|\omega| \neq |\omega_0|$ the solution remains bounded for all times t, since the oscillation is out of phase with the driving force.

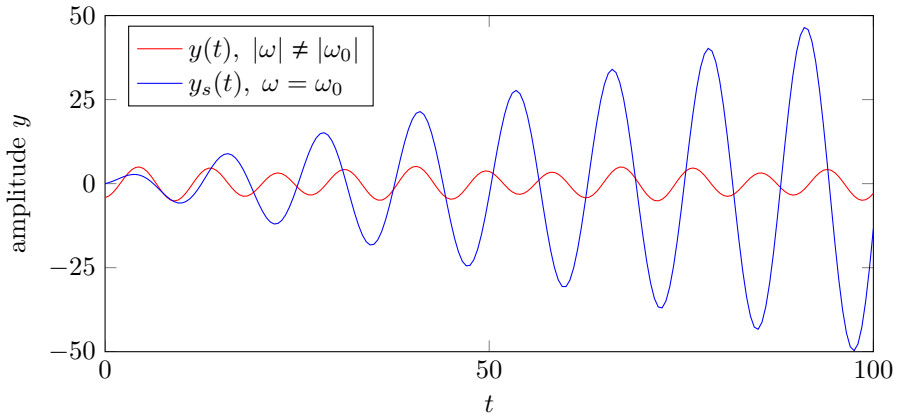


Figure 2.1. (resonance.m)
We chose $A = 0$, $B = 1$ and $\omega_0 = 0.5$ to plot the nonsecular solution $y(t)$ (red line) with driving frequency $\omega = 0.7$ and the secular solution $y_s(t)$ (blue line) with $\omega = \omega_0$. One observes that y stays bounded over all t whereas y_s is unbounded.

However, if $|\omega| = |\omega_0|$ the solution is given by

$$y_s(t) = A\cos(\omega_0 t) + B\sin(\omega_0 t) + \frac{1}{2}t\sin(\omega t), \quad |\omega| = |\omega_0|. \tag{2.4}$$

and therefore grows with t. Hence y_s is unbounded as $t \to \infty$ (cf. figure 2.1).

In this case the system can continually absorb energy from the periodic external force, so the amplitude of the oscillation of (2.2) increases without any bound. We say the system is in *resonance* with the external force.

The term $\frac{1}{2}t\sin(\omega t)$, which appears in (2.4), is called a *secular* term or just *secularity*, i.e. its amplitude grows algebraically with t. This secularity has appeared because the right hand side of (2.2), i.e. $\cos(\omega t)$ with $|\omega| = |\omega_0|$, itself is a solution of the homogeneous equation (2.1).

◇

As we have seen in this example secular terms are a problem of big concern whenever we are interested in bounded solutions of differential equations. Whenever an inhomogeneity in a differential equation is itself a solution of the associated homogeneous *constant coefficient* problem, those unwanted secularities appear. They always grow faster than the solution of the homogeneous problem by at least a factor of t.

The word *secular* itself means "pertaining to an age, a long period of time" and is used in this mathematical context because of the study of planetary motions where a secular term might not be noticeable over hundreds of years because the time scale is so long, but eventually will become unbounded.

Since secular terms play a major role in multiple scale analysis, let us give a proper definition of secular terms:

Definition 2.2 (secular terms)**.** *We call $f(t)$ a secular term if it becomes unbounded as $t \to \infty$, i.e.*

$$f \quad \textit{secular} \quad :\Longleftrightarrow\ \forall T > 0\, \forall C > 0\, \exists t > T : |f(t)| > C.$$

In the context of differential equations secular terms cause unboundedness of an inhomogeneous solution, even though the homogeneous solution itself is bounded.

2.2. Elimination of Secular Terms

In this section we want to give a possible solution on how to deal with arising secularities. But since the procedure described in this section is not very elegant as we will see, we will discuss the successive nonlinear example again in section 2.3.2 in order to illustrate how the method of multiple scales can be applied to a nonlinear initial value problem.

Therefore let us consider a nonlinear oscillator, i.e. a mass suspended from the ceiling by a spring, that is not perfectly ideal, an example that can also be found in [4], chapter 11.1.

Then Hooke's law, which depicts a linear relation between the displacement $y(t)$ and the restoring force of an **ideal** spring, applies only for quite small displacements y of the mass from the equilibrium, where the **linear** differential equation

$$\ddot{y}(t) + y(t) = 0, \quad y(0) = 1, \dot{y}(0) = 0$$

represents the movement of the oscillator (compare (2.1) in Example 2.1 with natural frequency $\omega_0 = 1$).

However, larger displacements from the equilibrium evoke the linearly assumed oscillation to become **nonlinear**. Having $0 < \varepsilon \ll 1$, the corresponding perturbed equation of motion changes to

$$\ddot{y}(t) + y(t) + \varepsilon y(t)^3 = 0, \quad y(0) = 1,\ \dot{y}(0) = 0, \tag{2.5}$$

which is known as a so called *Duffing's equation* (see [17], section 2.2 for more details on the derivation of this equation from the physical point of view).

In order to understand how the nonlinearity εy^3 affects the corresponding solution of (2.5), we proceed in 3 steps:

(1) Make the ansatz of expanding y as a power series in ε

$$y(t) \coloneqq \sum_{n=0}^{\infty} \varepsilon^n y_n(t), \tag{2.6}$$

$$\begin{cases} y_0(0) = 1, \dot{y}_0(0) = 0, \\ y_n(0) = 0, \dot{y}_n(0) = 0, \quad n \geq 1 \end{cases} \tag{2.7}$$

(2) Substitute ansatz (2.6) into the differential equation (2.5)

(3) Set the terms associated with equal powers of ε equal to zero to obtain a sequence of **linear** differential equations.

This leads to

$$\begin{aligned} 0 &= \sum_{n=0}^{\infty} \varepsilon^n \ddot{y}_n + \sum_{n=0}^{\infty} \varepsilon^n y_n + \varepsilon \left(\sum_{n=0}^{\infty} \varepsilon^n y_n \right)^3 \\ &= (\ddot{y}_0 + y_0) + \varepsilon \left(\ddot{y}_1 + y_1 + y_0^3 \right) + \mathcal{O}(\varepsilon^2). \end{aligned}$$

In our example this yields the first two differential equations

$$\ddot{y}_0 + y_0 = 0 \tag{2.8a}$$

$$\ddot{y}_1 + y_1 = -y_0^3, \tag{2.8b}$$

where the second one (and also all further equations in the sequence) is inhomogeneous, so here the nonlinearity affects the solution for the first time.

A solution to (2.8a) sufficing the initial conditions given in (2.7) is

$$y_0(t) = \cos(t).$$

Now if we use the addition theorems for trigonometric functions we obtain

$$\cos^3(t) = \frac{1}{4}\cos(3t) + \frac{3}{4}\cos(t),$$

such that the right hand side of (2.8b) contains a harmless component $\frac{1}{4}\cos(3t)$ and a component $\frac{3}{4}\cos(t)$, whose frequency equals the natural frequency $\omega_0 = 1$ of the unperturbed oscillator. Writing down (2.8b) explicitely gives

$$\ddot{y}_1 + y_1 = \underbrace{-\frac{1}{4}\cos(3t)}_{=:g_n(t)} \underbrace{-\frac{3}{4}\cos(t)}_{=:g_s(t)}.$$

We solve the following two differential equations seperately by using (2.3) and (2.4), i.e.

$$\ddot{y}_{1a} + y_{1a} = g_n(t) \overset{(2.3)}{\Rightarrow} y_{1a}(t) = A\cos(t) + B\sin(t) - \frac{1}{4}\cdot\frac{\cos(3t)}{1-3^2} \quad \textit{(nonsecular)}$$

$$\ddot{y}_{1b} + y_{1b} = g_s(t) \overset{(2.4)}{\Rightarrow} y_{1b}(t) = C\cos(t) + D\sin(t) - \frac{3}{4}\cdot\frac{1}{2}t\sin(t) \quad \textit{(secular)}$$

Then we sum their solutions y_{1a} and y_{1b} up to get the general solution y_1 of (2.8b), i.e.

$$y_1(t) = y_{1a} + y_{1b} = \tilde{A}\cos(t) + \tilde{B}\sin(t) + \frac{1}{32}\cos(3t) - \frac{3}{8}t\sin(t). \tag{2.9}$$

Together with the initial conditions $y_1(0) = \dot{y}_1(0) = 0$ given in (2.7) this provides the particular solution

$$y_1(t) = -\frac{1}{32}\cos(t) + \frac{1}{32}\cos(3t) - \frac{3}{8}t\sin(t).$$

We can see that y_1 contains a secular term of the form $t\sin(t)$. Hence y_1 features linear growth in t.

Now that we have found functions y_0 and y_1 which satisfy (2.8a) and (2.8b), we can write down an approximation to the solution y of (2.5) in the form of the ansatz that we made in (2.6):

$$\begin{aligned} y(t) &= y_0(t) &&+\varepsilon y_1(t) && &&+\mathcal{O}(\varepsilon^2) \\ &= \cos(t) &&+\varepsilon\left[-\frac{1}{32}\cos(t) + \frac{1}{32}cos(3t) - \frac{3}{8}t\sin(t)\right] && &&+\mathcal{O}(\varepsilon^2). \end{aligned}$$

Here $\mathcal{O}(\varepsilon^2)$ means that for fixed t the error

$$err(t) \coloneqq |y(t) - (y_0(t) + \varepsilon y_1(t))|$$

is at most of order ε^2. But this is only true as long as $t \ll \varepsilon^{-1}$. For these t we have $y_1(t) = \mathcal{O}(\varepsilon^{-1})$ because of the term $\frac{3}{8}t\sin(t)$ in (2.9). If we allow t to be of order ε^{-1} or larger, this secular term gains more and more weight, since the amplitude of the oscillation grows with t, such that the error estimate $err(t) = \mathcal{O}(\varepsilon^2)$ does not hold any more.

However, the exact solution $y(t)$ to (2.5) remains bounded for all t as C. Bender and A. Orszag show in [4], chapter 11.

Thus, the fact that our approximation has secular terms seems to be a paradox. But actually the resolution of this paradox can be found in the summation of the perturbation series in (2.6), that converges to $y(t)$ for each fixed t as $\varepsilon \to 0+$. So even though each term y_n in the series may contain secular terms, these secularities must be caught by successive terms in the series, i.e. they must disappear by summation. We can illustrate this by an easy example:

Consider the function

$$\begin{aligned} f(t) \coloneqq e^{-\varepsilon t} &= \sum_{n=0}^{\infty} \frac{(-1)^n \varepsilon^n t^n}{n!} \\ &= 1 - \varepsilon t + \frac{1}{2}\varepsilon^2 t^2 - \frac{1}{6}\varepsilon^3 t^3 + \cdots \end{aligned} \quad , \quad \varepsilon \to 0+$$

and observe that each term of the series for itself is secular when t is of order ε^{-1} or larger, but $f(t)$ remains bounded for all $t > 0$.

So how can we achieve to construct a series that is a good approximation to the exact solution y of (2.5) and that remains bounded for all t?

Let us analyze the more complicated series of ansatz (2.6), i.e.

$$y(t) = \sum_{n=0}^{\infty} \varepsilon^n y_n(t).$$

Considering the sequence of differential equations with inhomogeneity $\mathcal{I}_n$

$$\ddot{y}_n + y_n = -\mathcal{I}_n, \quad y_n(0) = \alpha_n,\ \dot{y}_n(0) = \beta_n, \quad n \geq 0$$

for some $\alpha_n, \beta_n \in \mathbb{R}$, one can show by an inductive argument, that the most secular terms of a summand $y_n(t)$ are growing like t^n, cf. [4], end of section 11.1. They are of the form

$$A_n t^n e^{it} + A_n^* t^n e^{-it},$$

where

$$A_n = \frac{1}{2}\frac{1}{n!}\left(\frac{3i}{8}\right)^n$$

and A_n^* is the complex conjugate of A_n. Less secular terms, i.e. terms growing like t^k, $(k < n)$ can be ignored in the following, because for $t \in \mathcal{O}(\varepsilon^{-1})$ they are negligible compared to at least one of the most secular terms included in the following equation.

Constructing a series of the most secular terms using A_n defined as above yields

$$\sum_{n=0}^{\infty} \frac{1}{2}\varepsilon^n t^n \left[\frac{1}{n!}\left(\frac{3i}{8}\right)^n e^{it} + \frac{1}{n!}\left(-\frac{3i}{8}\right)^n e^{-it}\right] = \frac{1}{2}\left(e^{i(t+\frac{3}{8}\varepsilon t)} + e^{-i(t+\frac{3}{8}\varepsilon t)}\right) = \cos\left[t\left(1+\frac{3}{8}\varepsilon\right)\right], \quad (2.10)$$

which is not secular and remains bounded for all t.

One can see in figure 2.2 that at $t = 160$ this approximation is still nearly "in phase" with the exact solution y for $\varepsilon = 0.1$ of the nonlinear equation, whereas the solution of the linear oscillator $\cos(t)$ is already about a whole period "out of phase". For $\varepsilon = 0.3$ this effect can be seen even better. Then $\cos(t)$ is already 3 periods out of phase and also the better approximation loses synchrony with y more and more.

The result of (2.10) can be interpreted as a phase shift of the harmonic oscillator, caused by the nonlinearity εy^3 in our differential equation.

2.3. Method of Multiple Scales: The "How to" in two Examples

The main result of the previous section is the fact, that perturbation theory in powers of ε is invalid when t gets larger than $\mathcal{O}\varepsilon^{-1}$, because secular terms appear in all orders of ε and lead to unboundedness of a truncated perturbation series although the exact solution y may be bounded.

We have seen, that we can get rid of the most secular terms of a perturbation series by summing them to all orders of ε such that we hopefully obtain a convergent series, in the best case convergent to the exact solution.

This calculation can be very lengthy. It can be bypassed by using a more elegant method, the method of multiple scales. As the method's name suggests we do not only use one time scale t but also several further time scales εt, $\varepsilon^2 t, \ldots$, such that we can deal with times t of order $\mathcal{O}(\varepsilon^{-k})$. Then it will provide

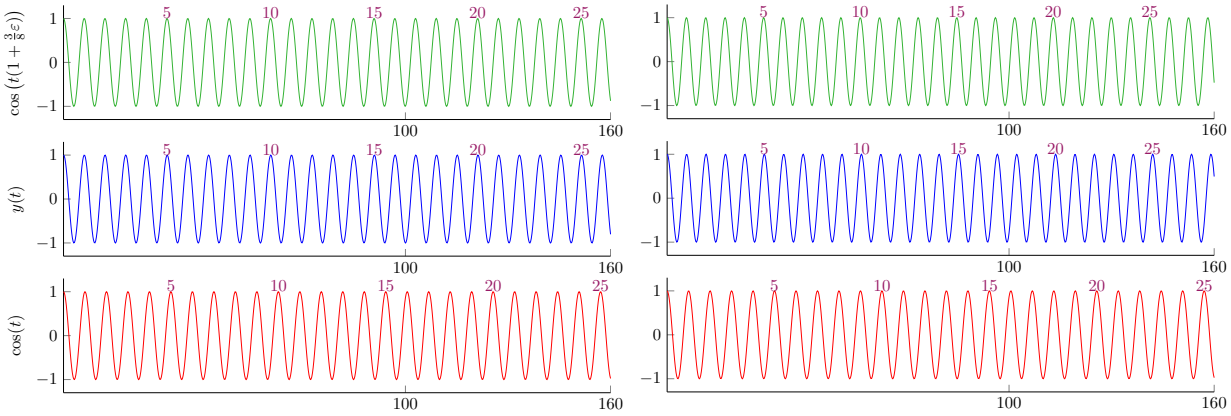


Figure 2.2. (duffing.m)
Comparison of the exact solution $y(t)$ of the Duffing's equation to the approximations $\cos(t)$ and $\cos\left(t(1+\frac{3}{8}\varepsilon)\right)$, for $\varepsilon = 0.1$ (left) and $\varepsilon = 0.3$ (right).

a pretty nice way to eliminate secular terms in the approximate solution as we will see in the following examples.

Firstly in order to get a bit into this method we are about to treat a linear example and afterwards we will come back to the Duffing's equation that we already studied. For the latter the method of multiple scales results in the same approximation $\tilde{y}(t) = \cos\left(t(1+\frac{3}{8})\right)$ that we already derived from a different point of view.

Before we proceed to the examples let us give the definition of long and short time scales.

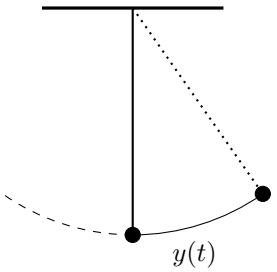
Definition 2.3. *Let t be the argument of a function $y \in \mathcal{C}^2(\mathbb{R})$, $0 < \varepsilon < 1$, $k \in \mathbb{N}$.*

*Then $\varepsilon^k t$ is called a **long / slow** variable, whereas t itself is called a **short / fast** variable. If t represents the time variable we may also call it slow or respectively fast time scale.*

We may use the terms of the "long" and respectively "short" variable interchangeably with "slow" and "fast" respectively.

2.3.1. A linear Example: The Damped Oscillator

Now let us explain the method of multiple scales on the basis of a linear example first. We consider a damped oscillator with small damping, caused by external forces such as friction. An oscillator of that kind can be for instance a simple pendulum. For example think of a child sitting on a swing. If we deflect it from the equilibrium and then leave it the swing will oscillate, but its amplitude will decrease. The mathematical example of this section is taken from [21], chapter 7.B.



If we do not respect the friction of the string against the hanging and of the whole swing against the air, then its oscillation at frequency ω_0 with small amplitude α can be described as before by

$$\ddot{y}(t) + \omega_0^2 y(t) = 0, \quad y(0) = \alpha,\ \dot{y}(0) = 0,$$

and will go on forever, where $y(t)$ shall denote the deflection from the equilibrium at time t. But as we know from everyday life the amplitude of such an oscillator will decay over time t, thus we should respect the external forces described above. These can be approximated by the damping term $2\zeta\omega_0\dot{y}(t)$, since the damping depends on the current velocity of the pendulum. Therefore the differential equation, which characterizes this behaviour reads

$$\ddot{y}(t) + 2\zeta\omega_0\dot{y}(t) + \omega_0^2 y(t) = 0, \quad y(0) = \alpha,\ \dot{y}(0) = 0.$$

Here we have to distinguish three different cases for $\zeta > 0$:

(1) $\zeta > 1$: The system is overdamped, i.e. the amplitude of the oscillator decays exponentially to 0 without oscillating.
(2) $\zeta = 1$: The system is critically damped, i.e. it will return to the equilibrium as fast as possible without oscillating.
(3) $0 < \zeta < 1$: The system is underdamped, i.e. we observe an oscillation, where the amplitude is exponentially decaying to 0.

The last case is the one we are interested in, since we only want to consider small damping in our multiple scale method. Therefore for $\varepsilon \in (0, 1)$ we set $\zeta = \varepsilon$ and for simplicity we choose a pendulum whose natural frequency is $\omega_0 = 1$. This leads to the differential equation

$$\ddot{y}(t) + 2\varepsilon\dot{y}(t) + y(t) = 0, \quad y(0) = \alpha,\ \dot{y}(0) = 0, \tag{2.11}$$

which we want to solve approximately by the method of multiple scales, even though we can solve it exactly as

$$y(t) = \alpha e^{-\varepsilon t}\left[\cos\left(\sqrt{1-\varepsilon^2}t\right) + \frac{\varepsilon}{\sqrt{1-\varepsilon^2}}\sin\left(\sqrt{1-\varepsilon^2}t\right)\right]. \tag{2.12}$$

One can easily check that y is bounded for all $t \geq 0$.

The method to get a multiple scale approximation is pretty technical and goes as follows:

The First Step

Firstly we introduce the new variables

$$\tau := \varepsilon t, \quad \sigma := \varepsilon^2 t$$

that define long time scales. These time scales are called *long* since τ and σ are not negligible when t is of order ε^{-1} or larger.

The solution $y(t)$ of the problem (2.11) clearly depends only on the *short* time t. Nevertheless the method of multiple scales seeks solutions that are functions of both the fast time scale t and the slow time scales τ and σ. Even though in the actual solution t, τ and σ are in correlation to each other, this method treats them as independent variables. This artifice is a means to remove secular effects in an elegant way as we will see. We want to emphasize again that t and τ are ultimately **not** independent.

The Second Step

Then secondly we make the ansatz

$$y(t) = \sum_{n=0}^{\infty} \varepsilon^n Y_n(t, \tau, \sigma, \varepsilon^3 t, \dots)$$

of y as a so called perturbation series, where this representation is only formal and where Y_n is allowed to depend on as many time scales as needed. For our purpose the variables t, τ and σ are sufficient, i.e. $Y_n = Y_n(t, \tau, \sigma)$.

If we want to use this representation of y in the damped equation (2.11) we have to expand the notion of the derivative with respect to t by a differential operator A_t that satifies

$$\frac{d}{dt} y(t) = A_t \sum_{n=0}^{\infty} \varepsilon^n Y_n(t, \tau, \sigma)$$

for all t. Having $\frac{d}{dt}\tau = \varepsilon$ and $\frac{d}{dt}\sigma = \varepsilon^2$, we obtain the formal differential operator

$$A_t \coloneqq \left(\frac{\partial}{\partial t} + \varepsilon \frac{\partial}{\partial \tau} + \varepsilon^2 \frac{\partial}{\partial \sigma} \right),$$

by using the chain rule for partial differentiation. Here we assume each Y_n to be at least two times continuously differentiable with respect to all variables.

Hence we find the first and second derivative of y

$$\frac{d}{dt} y = \frac{\partial}{\partial t} Y_0 + \varepsilon \left(\frac{\partial}{\partial \tau} Y_0 + \frac{\partial}{\partial t} Y_1 \right) + \varepsilon^2 \left(\frac{\partial}{\partial \sigma} Y_0 + \frac{\partial}{\partial \tau} Y_1 + \frac{\partial}{\partial t} Y_2 \right) + \mathcal{O}(\varepsilon^3)$$

and

$$\frac{d^2}{dt^2} y = \frac{\partial^2}{\partial t^2} Y_0 + \varepsilon \left(2 \frac{\partial^2}{\partial \tau \partial t} Y_0 + \frac{\partial^2}{\partial t^2} Y_1 \right) + \varepsilon^2 \left(\frac{\partial^2}{\partial \tau^2} Y_0 + 2 \frac{\partial^2}{\partial \sigma \partial t} Y_0 + 2 \frac{\partial^2}{\partial \tau \partial t} Y_1 + \frac{\partial^2}{\partial t^2} Y_2 \right) + \mathcal{O}(\varepsilon^3). \tag{2.13}$$

The Third Step

In the third step we insert (2.13) into the differential equation (2.11) and sort the terms by powers of ε. This yields

$$\begin{aligned} &\left\{ \frac{\partial^2}{\partial t^2} Y_0 + Y_0 \right\} + \varepsilon \left\{ \frac{\partial^2}{\partial t^2} Y_1 + Y_1 + 2 \frac{\partial^2}{\partial \tau \partial t} Y_0 + 2 \frac{\partial}{\partial t} Y_0 \right\} \\ &\qquad + \varepsilon^2 \left\{ \frac{\partial^2}{\partial t^2} Y_2 + Y_2 + \frac{\partial^2}{\partial \tau^2} Y_0 + 2 \frac{\partial^2}{\partial \sigma \partial t} Y_0 + 2 \frac{\partial}{\partial \tau} Y_0 + 2 \frac{\partial^2}{\partial \tau \partial t} Y_1 + 2 \frac{\partial}{\partial t} Y_1 \right\} \\ &\qquad + \mathcal{O}(\varepsilon^3) = 0. \end{aligned}$$

Comparing the left hand side with the right hand side of this equation, we see that it makes sense to set the terms in front of each power of ε to 0, which yields a sequence of differential equations. In a similar way we obtain the corresponding initial values. In this example we have $y(0) = \alpha$ and $\frac{d}{dt} y(0) = 0$. By

our ansatz for $y(t)$ we have

$$\alpha = Y_0(0,0,0) + \varepsilon Y_1(0,0,0) + \varepsilon^2 Y_2(0,0,0) + \dots$$

and from (2.13) we can conclude

$$\begin{aligned} 0 =& \frac{\partial}{\partial t} Y_0(0,0,0) + \varepsilon \left(\frac{\partial}{\partial \tau} Y_0(0,0,0) + \frac{\partial}{\partial t} Y_1(0,0,0) \right) \\ &+ \varepsilon^2 \left(\frac{\partial}{\partial \sigma} Y_0(0,0,0) + \frac{\partial}{\partial \tau} Y_1(0,0,0) + \frac{\partial}{\partial t} Y_2(0,0,0) \right) + \dots \end{aligned}$$

and if we again compare the terms in orders of ε we get the initial values associated to the corresponding differential equations.

Thus we have to solve

$$\frac{\partial^2}{\partial t^2} Y_0 + Y_0 = 0, \tag{2.14}$$

$$Y_0(0,0,0) = \alpha, \ \frac{\partial}{\partial t} Y_0 = (0,0,0)$$

$$\frac{\partial^2}{\partial t^2} Y_1 + Y_1 = -\left(2 \frac{\partial^2}{\partial \tau \partial t} Y_0 + 2 \frac{\partial}{\partial t} Y_0 \right), \tag{2.15}$$

$$Y_1(0,0,0) = 0, \ \frac{\partial}{\partial t} Y_1(0,0,0) = -\frac{\partial}{\partial \tau} Y_0(0,0,0)$$

$$\frac{\partial^2}{\partial t^2} Y_2 + Y_2 = -\left(\frac{\partial^2}{\partial \tau^2} Y_0 + 2 \frac{\partial^2}{\partial \sigma \partial t} Y_0 + 2 \frac{\partial}{\partial \tau} Y_0 + 2 \frac{\partial^2}{\partial \tau \partial t} Y_1 + 2 \frac{\partial}{\partial t} Y_1 \right), \tag{2.16}$$

$$Y_2(0,0,0) = 0, \ \frac{\partial}{\partial t} Y_2(0,0,0) = -\frac{\partial}{\partial \tau} Y_1(0,0,0) - \frac{\partial}{\partial \sigma} Y_0(0,0,0).$$

We can solve (2.14) with the ansatz

$$Y_0(t,\tau,\sigma) = A(\tau,\sigma)\cos t + B(\tau,\sigma)\sin t, \quad A(0,0) = \alpha, \ B(0,0) = 0 \tag{2.17}$$

with yet to specify functions A and B. We calculate the partial derivatives of Y_0 with respect to t, τ and σ, treating those as independent, uncorrelated variables, and insert them into (2.15) . This will give some first conditions, that A and B have to satisfy. We obtain

$$\frac{\partial^2}{\partial t^2} Y_1 + Y_1 = 2 \left(\frac{\partial}{\partial \tau} A(\tau,\sigma) + A(\tau,\sigma) \right) \sin t - 2 \left(\frac{\partial}{\partial \tau} B(\tau,\sigma) + B(\tau,\sigma) \right) \cos t. \tag{2.18}$$

We see that the right hand side is in resonance with the solution of the homogeneous equation

$$\frac{\partial^2}{\partial t^2} Y_1 + Y_1 = 0,$$

if the terms in front of $\cos t$ and $\sin t$ are nonzero. Because we treat t, τ and σ as independent, these terms are constant with respect to t. Therefore in order to avoid secular terms in differential equations, A and B have to satisfy

$$\begin{aligned} \frac{\partial}{\partial \tau} A &= -A, \quad A(0,0) = \alpha, \\ \frac{\partial}{\partial \tau} B &= -B, \quad B(0,0) = 0, \end{aligned} \tag{2.19}$$

where the initial conditions are from (2.17). This gives

$$\begin{aligned} A(\tau,\sigma) &= \gamma(\sigma) e^{-\tau}, \quad \gamma(0) = \alpha, \\ B(\tau,\sigma) &= \delta(\sigma) e^{-\tau}, \quad \delta(0) = 0. \end{aligned}$$

At this point we simplified (2.18) in such a way that the right hand side vanishes and we obtain a homogeneous equation again for Y_1:

$$\frac{\partial^2}{\partial t^2}Y_1 + Y_1 = 0, \quad Y_1(0,0) = 0, \ \frac{\partial}{\partial t}Y_1(0,0) = \alpha. \tag{2.20}$$

Analogously to (2.17), we make the ansatz

$$Y_1(t,\tau,\sigma) = C(\tau,\sigma)\cos t + D(\tau,\sigma)\sin t, \quad C(0,0) = 0, \ D(0,0) = \alpha. \tag{2.21}$$

for the solution of (2.20). When we wanted to determine the functions $A(\tau,\sigma), B(\tau,\sigma)$ in the ansatz for Y_0, we had to consider the equation for Y_1 and had to arrange that no secular terms arouse in its right hand side. The same is done now in order to determine C and D by analyzing the differential equation for Y_2 and especially its right hand side. The latter turns out again to be in resonance with the solution of the homogeneous equation

$$\frac{\partial^2}{\partial t^2}Y_2 + Y_2 = 0.$$

Thus we have to set the coefficients of the secular terms equal to zero and solve another ordinary differential equation for C and D.

But first let us write down again the equation for Y_2:

$$\begin{aligned}\frac{\partial^2}{\partial t^2}Y_2 + Y_2 &= -\left(\frac{\partial^2}{\partial\tau^2}Y_0 + 2\frac{\partial^2}{\partial\sigma\partial t}Y_0 + 2\frac{\partial}{\partial\tau}Y_0 + 2\frac{\partial^2}{\partial\tau\partial t}Y_1 + 2\frac{\partial}{\partial t}Y_1\right)\\ &= \left[(2\gamma'(\sigma) + \delta(\sigma))\,e^{-\tau} + 2\left(\frac{\partial}{\partial\tau}C + C\right)\right]\sin(t)\\ &+ \left[(-2\delta'(\sigma) + \gamma(\sigma))\,e^{-\tau} - 2\left(\frac{\partial}{\partial\tau}D + D\right)\right]\cos(t).\end{aligned}$$

Again we see that we have to set the terms in front of $\sin(t)$ and $\cos(t)$ equal to zero to avoid secularities. This gives

$$\begin{aligned}2\gamma'(\sigma) + \delta(\sigma) &= -2e^{\tau}\left(\frac{\partial}{\partial\tau}C + C\right)\\ -2\delta'(\sigma) + \gamma(\sigma) &= 2e^{\tau}\left(\frac{\partial}{\partial\tau}D + D\right)\end{aligned} \tag{2.22}$$

and because the left hand side is only depending on σ we conclude that the right hand side has to be constant with respect to τ. Therefore we write the equations for $C(\tau,\sigma)$ and $D(\tau,\sigma)$ as

$$\begin{aligned}\frac{\partial}{\partial\tau}C + C &= -\frac{1}{2}e^{-\tau}f(\sigma), \quad C(0,0) = 0\\ \frac{\partial}{\partial\tau}D + D &= \ \ \frac{1}{2}e^{-\tau}g(\sigma), \quad D(0,0) = \alpha.\end{aligned}$$

Taking a closer look one recognizes the secular terms $e^{-\tau}$ on the right hand side of both equations. In order to assure that Y_1 stays bounded we therefore have to set $f(\sigma) \equiv 0 \equiv g(\sigma)$ and find

$$C(\tau,\sigma) \equiv 0 \qquad \text{and} \qquad D(\tau,\sigma) = \alpha e^{-\tau}.$$

Then (2.22) becomes

$$\begin{aligned}2\gamma'(\sigma) + \delta(\sigma) &= 0\\ -2\delta'(\sigma) + \gamma(\sigma) &= 0\end{aligned}, \quad \gamma(0) = \alpha, \ \delta(0) = 0.$$

As one can easily check, the functions

$$\gamma(\sigma) = \alpha\cos\left(\frac{1}{2}\sigma\right) \qquad \text{and} \qquad \delta(\sigma) = \alpha\sin\left(\frac{1}{2}\sigma\right)$$

solve this system.

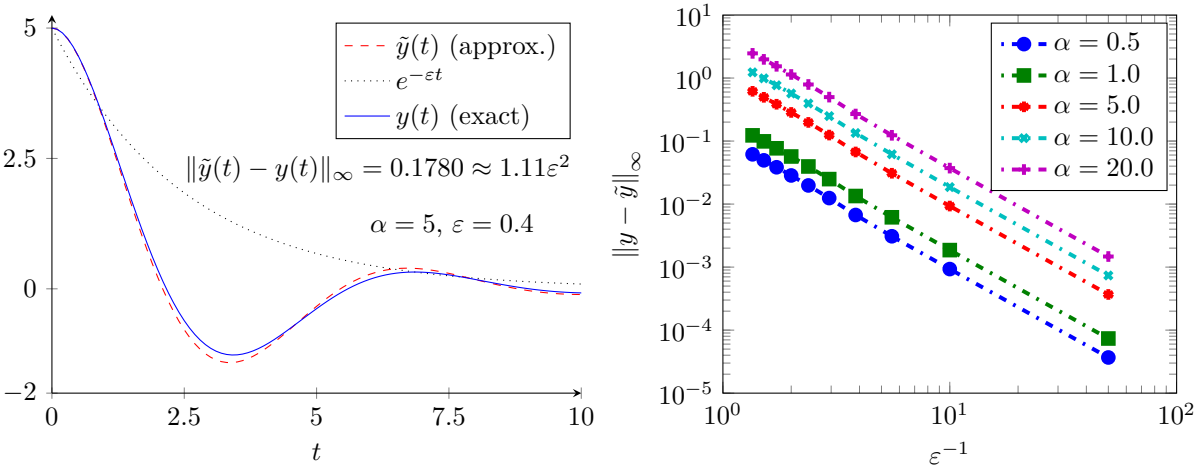


Figure 2.3. (damped_oscillator.m)
Left: Evolution of the damped oscillator's amplitude over time. The blue solid line shows the exact solution $y(t)$ (see (2.12)) of the damped oscillator equation (2.11) with $\alpha = 5$ and $\varepsilon = 0.4$. The red dashed line shows the multiple scale approximation $\tilde{y}(t)$. The amplitude goes to zero over time like $e^{-\varepsilon t}$ (dashed line).
Right: Logarithmic error plot of the maximal error of the approximation $\tilde{y}$ in the time interval $t \in [0, 200]$ for various $\alpha > 0$. The logarithmic error decreases linearly with gradient 2. Therefore the order of the error is $\mathcal{O}(\varepsilon^2)$.

Thus altogether

$$\begin{aligned} Y_0(t,\tau,\sigma) &= \alpha e^{-\tau}\left[\cos\left(\frac{1}{2}\sigma\right)\cos(t) + \sin\left(\frac{1}{2}\sigma\right)\sin(t)\right] \\ &= \alpha e^{-\tau}\cos\left(t - \frac{1}{2}\sigma\right) \\ Y_1(t,\tau,\sigma) &= \alpha e^{-\tau}\sin(t) \end{aligned}$$

Finally, if we have a look at the initial conditions of the (now homogeneous) equation (2.16) for Y_2 we find $Y_2(0,0,0) = 0$ and $\frac{\partial}{\partial t}Y_2(0,0,0) = 0$ such that $Y_2(t,\tau,\sigma) = 0,\ \forall t,\tau,\sigma$.

The Result

Overall we obtain the approximation $\tilde{y} \coloneqq Y_0 + \varepsilon Y_1$, i.e.

$$\tilde{y}(t) = \alpha e^{-\varepsilon t}\left[\cos\left((1 - \frac{\varepsilon^2}{2})\,t\right) + \varepsilon\sin(t)\right].$$

If we have a look at the error

$$\sum_{n=0}^{\infty}\varepsilon^n Y_n - \tilde{y} = \varepsilon^3 Y_3 + \mathcal{O}(\varepsilon^5)$$

we find that $\tilde{y}$ has an error of order $\mathcal{O}(\varepsilon^2)$ as long as $t \in \mathcal{O}(\varepsilon^{-1})$ or smaller. We have to set up such a bound to the time interval, since we did not fetch more information on the ε^3 term Y_3 in the series of our ansatz. This term may contain secularities of order $\mathcal{O}(t)$ or even worse such that Y_3 becomes $\mathcal{O}(1)$ if we exceed a certain time barrier. Therefore we set up the *validity interval* $[0, T_0/\varepsilon]$ for our approximation for some $T_0 > 0$.

The multiple scale approximation $\tilde{y}(t)$ and the exact solution $y(t)$ for $\varepsilon = 0.4$ and $\alpha = 5$ are depicted in figure 2.3 and from the error plot one can conclude that as expected the error is of order $\mathcal{O}(\varepsilon^2)$.

2.3.2. A nonlinear Example: The Duffing's Equation

To gain more insight into how this method helps to analyze the influence of nonlinear effects let us again have a look at the nonlinear differential equation from the beginning of this chapter, the Duffing's equation

$$\ddot{y}(t) + y(t) + \varepsilon y(t)^3 = 0, \quad y(0) = 1, \dot{y}(0) = 0. \tag{2.23}$$

The Formal Procedure

We proceed in this nonlinear example as we have already seen in the previous section:

Firstly we assume a perturbation expansion of the solution y, that is of the form:

$$y(t) = Y_0(t,\tau) + \varepsilon Y_1(t,\tau) + \cdots = \sum_{n=0}^{\infty} \varepsilon^n Y_n(t,\tau), \tag{2.24}$$

where $Y_n(t,\tau)$ are the functions we are looking for. For our purpose this time the functions Y_n only depend on the two time scales t and $\tau = \varepsilon t$. Of course we can extend this procedure to as many time scales as we like, but since some upcoming equations would become unnecessarily difficult without any further insight into the method itself we leave it at two time scales.

Again we obtain the derivatives of $y(t)$ by using the chain rule for partial differentiation, which yields

$$\frac{d}{dt}y(t) = \frac{\partial Y_0}{\partial t} + \varepsilon\left(\frac{\partial Y_0}{\partial \tau} + \frac{\partial Y_1}{\partial t}\right) + \mathcal{O}(\varepsilon^2), \tag{2.25a}$$

$$\frac{d^2}{dt^2}y(t) = \frac{\partial^2 Y_0}{\partial t^2} + \varepsilon\left(2\frac{\partial^2 Y_0}{\partial \tau \partial t} + \frac{\partial^2 Y_1}{\partial t^2}\right) + \mathcal{O}(\varepsilon^2). \tag{2.25b}$$

By substituting the series (2.24) and equations (2.25) into our differential equation (2.23) we obtain

$$\frac{\partial^2 Y_0}{\partial t^2} + Y_0 + \varepsilon\left(\frac{\partial^2 Y_1}{\partial t^2} + Y_1 + 2\frac{\partial^2 Y_0}{\partial \tau \partial t} + Y_0^3\right) + \mathcal{O}(\varepsilon^2) = 0.$$

In the same way as in the previous section we obtain a sequence of differential equations. The first two equations of this sequence are given by

$$\frac{\partial^2 Y_0}{\partial t^2} + Y_0 = 0, \qquad Y_0(0,0) = 1, \quad \frac{\partial}{\partial t}Y_0(0,0) = 0, \tag{2.26a}$$

$$\frac{\partial^2 Y_1}{\partial t^2} + Y_1 = -Y_0^3 - 2\frac{\partial^2 Y_0}{\partial \tau \partial t}, \quad Y_1(0,0) = 0, \quad \frac{\partial}{\partial t}Y_1(0,0) = -\frac{\partial}{\partial \tau}Y_0(0,0). \tag{2.26b}$$

To include the dependence of the solutions to these differential equations on the independently treated variables t and τ we make the following ansatz for Y_0

$$Y_0(t,\tau) = A(\tau)e^{it} + A^*(\tau)e^{-it}, \tag{2.27}$$

which indeed solves (2.26a) and even is its most general real solution.

Here, $A(\tau)$ shall be a yet arbitrary complex function of τ, that we want to determine more precisely in the following steps and $A^*(\tau)$ denotes its complex conjugate.

We differentiate $Y_0(t,\tau)$ with respect to t and τ and obtain its partial derivatives

$$\frac{\partial Y_0}{\partial t}(t,\tau) = iA(\tau)\ e^{it} - iA^*(\tau)\ \ e^{-it} \tag{2.28a}$$

$$\frac{\partial Y_0}{\partial \tau}(t,\tau) = \frac{dA}{d\tau}(\tau)e^{it} + \frac{dA^*}{d\tau}(\tau)e^{-it}, \tag{2.28b}$$

again treating t and τ as independent variables.

The function $A(\tau)$ is determined by setting up the condition that secular terms do *not* appear in equation (2.26b). We have

$$\frac{\partial^2 Y_0}{\partial\tau\partial t} = i\frac{dA}{d\tau}(\tau)e^{it} - i\frac{dA^*}{d\tau}(\tau)e^{-it}$$

and

$$Y_0^3 = A^3e^{3it} + 3A^2A^*e^{it} + 3A\left(A^*\right)^2e^{-it} + \left(A^*\right)^3e^{-3it}$$

such that finally

$$\begin{aligned} -Y_0^3 - 2\frac{\partial^2 Y_0}{\partial\tau\partial t} = \quad & -A^3e^{3it} + e^{it}\underbrace{\left[-3A^2A^* - 2i\frac{dA}{d\tau}\right]}_{=:s_1(\tau)} \\ & -\left(A^*\right)^3e^{-3it} + e^{-it}\underbrace{\left[-3A\left(A^*\right)^2 + 2i\frac{dA^*}{d\tau}\right]}_{=:s_2(\tau)}. \end{aligned} \tag{2.29}$$

One can easily check that $e^{\pm it}$ is itself a solution of the homogeneous equation

$$\frac{\partial^2}{\partial t^2}Y_1 + Y_1 = 0$$

corresponding to (2.26b).

So if the terms $s_1(\tau), s_2(\tau)$ in front of $e^{\pm it}$ in (2.29) are nonzero, $Y_1(t,\tau)$ will be secular in t. But that is exactly what we want to avoid. Hence we set s_1 and s_2 equal to zero:

$$-3A^2A^* - 2i\frac{dA}{d\tau} = 0 \tag{2.30a}$$

$$-3A\left(A^*\right)^2 + 2i\frac{dA^*}{d\tau} = 0, \tag{2.30b}$$

and observe that (2.30b) is just the complex conjugate equation of (2.30a) and can be omitted. If we have A satisfying these conditions, Y_1 will not contain secular terms and at least no secularities appear in the first two terms of the series representation of y in (2.24) , but we have to be aware that we don't have any information on further terms.

So we have to restrict our time interval to $t \in [0, T_0/\varepsilon]$ such that the error of the approximation

$$\tilde{y}(t) \coloneqq Y_0(t,\tau) + \varepsilon Y_1(t,\tau)$$

is of order $\mathcal{O}(\varepsilon^2)$, i.e. $y(t) - \tilde{y}(t) = \mathcal{O}(\varepsilon^2)$, similar as before.

To achieve this objective let us try to solve (2.30a) by writing $A(\tau)$ in polar coordinates:

$$A(\tau) \coloneqq R(\tau)e^{i\theta(\tau)}, \quad R, \theta : \mathbb{R} \mapsto \mathbb{R}.$$

By (2.30a) we get the differential equation

$$\left(-3R^3(\tau) - 2i\frac{dR}{d\tau}(\tau) + 2R(\tau)\frac{d\theta}{d\tau}(\tau)\right)e^{i\theta(\tau)} = 0,$$

which actually gives two equations, one for the real part, the other for the imaginary part of the equation. Since $Y_0(0,0) = 2\,\mathrm{Re}\,A(0) = 1$ we can assume $R(0) \neq 0$. Thence

$$\frac{dR}{d\tau} = 0 \qquad \Rightarrow R(\tau) = R(0) \quad \forall\tau > 0 \tag{2.31}$$

$$\frac{d\theta}{d\tau} = \frac{3}{2}R^2 \quad \overset{(2.31)}{\Rightarrow} \theta(\tau) = \theta(0) + \frac{3}{2}R^2(0)\tau \tag{2.32}$$

and altogether

$$A(\tau) = R(0)e^{i\left(\theta(0) + \frac{3}{2}R^2(0)\,\tau\right)}.$$

If we now insert this result into our ansatz (2.27) this yields the zeroth order solution, i.e. the solution for Y_0 in front of ε^0

$$\begin{aligned} Y_0(t,\tau) &= A(\tau)e^{it} + A^*(\tau)e^{-it} \\ &= \mathrm{Re}\left(R(0)e^{i\left(\theta(0)+\frac{3}{2}R^2(0)\ \tau+t\right)}\right) \\ &= 2R(0)\cos\left[\theta(0)+\frac{3}{2}R^2(0)\ \tau+t\right]. \end{aligned}$$

We still have to determine $R(0)$ and $\theta(0)$ by using the initial conditions of our original problem (2.23), i.e. $y(0)=1, \dot{y}(0)=0$. We have to translate them into initial conditions for $Y_0, Y_1, \ldots$. Recall that $y(t) = Y_0(t,\tau) + \varepsilon Y_1(t,\tau) + \ldots$ and remember that these conditions read

$$\begin{array}{lll} y(0)=1 & \Rightarrow & Y_0(0,0)=1, \quad Y_1(0,0)=0, \quad Y_2(0,0)=0, \ \ldots \\ \dfrac{dy}{dt}(0)=0 & \overset{(2.25a)}{\Rightarrow} & \dfrac{\partial Y_0}{\partial t}(0,0)=0, \quad \dfrac{\partial Y_1}{\partial t}(0,0) = -\dfrac{\partial Y_0}{\partial \tau}(0,0), \ldots \end{array}$$

So we have the conditions for $R(0)$ and $\theta(0)$:

$$\begin{aligned} \frac{\partial Y_0}{\partial t}(0,0) = -2R(0)\sin\theta(0) = 0 \qquad &\Rightarrow \quad \theta(0)=0, \\ Y_0(0,0) = 2R(0)\cos\theta(0) \overset{\theta(0)=0}{=} 2R(0) = 1 \quad &\Rightarrow \quad R(0)=\frac{1}{2}. \end{aligned}$$

Therefore we obtain the result

$$Y_0(t,\tau) = \cos\left(t+\frac{3}{8}\tau\right)$$

and hence we finally have for the solution of the nonlinear oscillator problem (2.23), recalling that $\tau=\varepsilon t$

$$y(t) = \cos\left(t\left(1+\frac{3}{8}\varepsilon\right)\right) + \mathcal{O}(\varepsilon),$$

where $t \in \mathcal{O}(\varepsilon^{-1})$.

Observe that we have reproduced the approximation to the solution of (2.10) by simply setting up the condition that secularities do not appear in the solution Y_1 to (2.26b), corresponding to order ε, without actually solving this equation. Therefore Y_1 stays bounded for all t, τ, σ, i.e. $\varepsilon Y_1 \in \mathcal{O}(\varepsilon)$ and we only have to worry about the term $\varepsilon^2 Y_2$ which we did not investigate. This term may contain a secular part which grows like $\mathcal{O}(t)$. That's why we have to set the validity interval to $t \in [0, T_0/\varepsilon]$ for some $T_0 > 0$.

CHAPTER

THREE

A MULTIPLE SCALE METHOD APPLIED TO THE NONLINEAR KLEIN-GORDON EQUATION

Now that we have seen a few examples on how to use the method of multiple scales we are ready to apply it to the a bit more complicated problem of nonlinear waves in dispersive matter. In this chapter the content is mostly taken from [18], chapter 2c, [6], chapter 5, and [13]

Therefore we will apply the method to the nonlinear Klein-Gordon equation and encounter the Gross-Pitaevskii equation, a nonlinear Schrödinger equation which will provide a condition for the envelope of the wave packet to avoid secularities in the truncated multiple scale ansatz. The consequence will be a modulation in space and time of the solution of the wave equation (see figure 3.1), i.e. we obtain an approximation

$$\varepsilon\psi_{NLS}(x,t) \coloneqq \varepsilon A e^{i(kx-\omega t)} + \varepsilon A^* e^{-i(kx-\omega t)},$$

where its amplitude A obeys the Gross-Pitaevskii equation.

But before we begin, we want to introduce the *residual* of a nonlinear partial differential equation. If we want to know how accurate a multiple scale method is, applied to an equation of the form

$$\mathcal{L}(u) = \mathcal{N}(u), \tag{3.1}$$

we have to set up a means to measure the goodness of an approximation to the solution. Here $\mathcal{L}$ denotes a linear differential operator and $\mathcal{N}$ a nonlinearity in the solution u. For that purpose let us define the residual of such a differential equation.

Definition 3.1 (the residual)**.** *The residual of a nonlinear differential equation is given by*

$$\mathrm{Res}(\tilde{u}) = \mathcal{L}(\tilde{u}) - \mathcal{N}(\tilde{u}).$$

It gives information about the goodness of an approximation $\tilde{u}$.

If $\mathrm{Res}(\tilde{u}) = 0$ *then $\tilde{u}$ is the exact solution of the differential equation.*

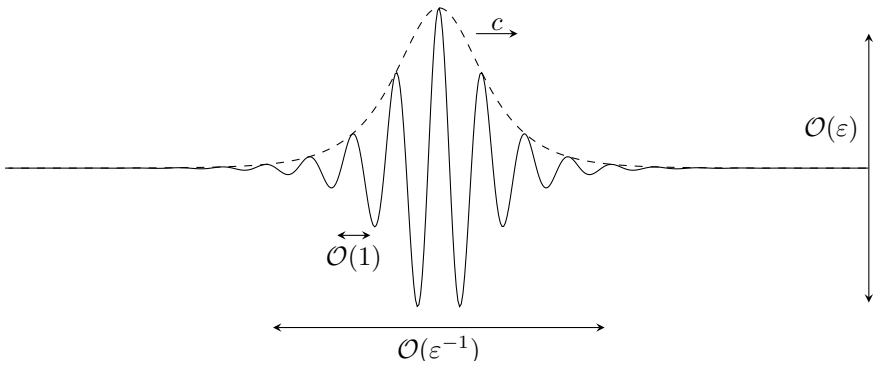


Figure 3.1. (RUN_wave_shape.m)
A wave packet (solid line) modulated by an envelope function, which is a solution of the nonlinear Schrödinger equation (dashed line). The envelope and the wave packet move with group velocity c. The amplitude and the width of the pulse are defined in terms of the small parameter ε, whereas the oscillation width of the wave itself is $\mathcal{O}(1)$.

3.1. The nonlinear Klein-Gordon Equation

When applying a multiple scale analysis to a partial differential equation our aim is to gain an approximation $\tilde{u}$ with small residual. Now we want to apply a multiple scale analysis to the nonlinear wave equation with cubic nonlinearity, i.e.

$$\begin{aligned} \frac{\partial^2}{\partial t^2}u(x,t) &= \alpha\frac{\partial^2}{\partial x^2}u(x,t) - \beta u(x,t) + \lambda u^3(x,t), \quad \alpha,\beta \geq 0,\ \lambda \in \mathbb{R}, \\ u(\,\cdot\,,0) &= u_0, \qquad \frac{\partial}{\partial t}u(\,\cdot\,,0) = u_1 \end{aligned} \tag{3.2}$$

also called nonlinear *Klein-Gordon equation*, and proceed as in [18], chapter 2c. This equation describes the movement of a wave packet at constant group velocity c. As we will see the envelope of the wave packet is modulated by some function A in the spatial order ε^{-1}, see figure 3.1. Therefore we will additionally to some long time scales also introduce a long spatial scale $X = \varepsilon x$, which enables us to investigate this modulation function in an elegant way.

Metaphorically speaking the long time scale means that we have a look at the movement of the solution in "slow motion". Likewise the long spatial scale eneables us to "zoom" into the solution in order to investigate its spatial properties.

Before we start applying the method of multiple scales let us fix some notation:

- We denote the short (or resp. fast) spatial variable by $x \in \mathbb{R}$ shall and the corresponding large (or slow) spatial variable by $X := \varepsilon x$.
- We set t the fast and $\tau := \varepsilon t$, $\sigma := \varepsilon^2 t$ the corresponding slow time scales.

Again $0 < \varepsilon \ll 1$ is a small parameter. Like in the case of ordinary differential equations we can make the ansatz of a perturbation series for the solution $u(x,t)$ of the partial differential equation (3.2) such that

$$u(x,t) = \varepsilon\sum_{n=0}^{\infty}\varepsilon^n\mathcal{U}_n(x,X,t,\tau,\sigma), \qquad \varepsilon \to 0+. \tag{3.3}$$

Inserting this ansatz into the wave equation leads to a sequence of partial differential equations, which gives more information about how to choose the functions $\mathcal{U}_n$ such that we get a feasible approximation

$$\tilde{u}(x,t) = \varepsilon\sum_{n=0}^{M}\varepsilon^n\mathcal{U}_n, \qquad \varepsilon \to 0+,$$

for some $M \in \mathbb{N}$. The aim is to get a small residual $\mathrm{Res}(\tilde{u})$. To achieve this goal we have to take care of secular terms and if necessary eliminate them. As for the examples in chapter 2 we can identify a term of the right hand side of an equation as a (possible) secularity, when it is a solution to the homogeneous differential equation by itself.

Note that as before x, X, t, τ and σ are treated independently when we differentiate $\mathcal{U}_n$ with respect to them, even though they are actually not independent. It will be quite the same procedure as for ordinary differential equations, such that we interpret the differentiations ∂_x and ∂_t applied to u as partial differential operators

$$A_x \coloneqq \left(\frac{\partial}{\partial x} + \varepsilon\frac{\partial}{\partial X}\right)$$

and

$$A_t \coloneqq \left(\frac{\partial}{\partial t} + \varepsilon\frac{\partial}{\partial \tau} + \varepsilon^2\frac{\partial}{\partial \sigma}\right)$$

applied to the series in (3.3) such that

$$\begin{aligned}\frac{\partial}{\partial x}u(x,t) &= A_x\left(\varepsilon\sum_{n=0}^{\infty}\varepsilon^n\mathcal{U}_n(x,X,t,\tau,\sigma)\right)\\ &= \varepsilon\left(\frac{\partial}{\partial x}\mathcal{U}_0 + \varepsilon\left(\frac{\partial}{\partial x}\mathcal{U}_1 + \frac{\partial}{\partial X}\mathcal{U}_0\right) + \varepsilon^2\left(\frac{\partial}{\partial x}\mathcal{U}_2 + \frac{\partial}{\partial X}\mathcal{U}_1\right) + \varepsilon^3\ldots\right)\end{aligned} \tag{3.5a}$$

and respectively

$$\begin{aligned}\frac{\partial}{\partial t}u(x,t) &= A_t\left(\varepsilon\sum_{n=0}^{\infty}\varepsilon^n\mathcal{U}_n(x,X,t,\tau,\sigma)\right)\\ &= \varepsilon\left(\frac{\partial}{\partial t}\mathcal{U}_0 + \varepsilon\left(\frac{\partial}{\partial t}\mathcal{U}_1 + \frac{\partial}{\partial \tau}\mathcal{U}_0\right) + \varepsilon^2\left(\frac{\partial}{\partial t}\mathcal{U}_2 + \frac{\partial}{\partial \tau}\mathcal{U}_1 + \frac{\partial}{\partial \sigma}\mathcal{U}_0\right) + \varepsilon^3\ldots\right).\end{aligned} \tag{3.5b}$$

Since $\frac{\partial^2}{\partial t^2} = A_t^2$ and $\frac{\partial^2}{\partial x^2} = A_x^2$ and therefore

$$\begin{aligned}\frac{\partial^2}{\partial x^2}u &= \left(\frac{\partial^2}{\partial x^2} + 2\varepsilon\frac{\partial^2}{\partial X\partial x} + \varepsilon^2\frac{\partial^2}{\partial X^2}\right)\ \varepsilon\sum_{n=0}^{\infty}\varepsilon^n\mathcal{U}_n,\\ \frac{\partial^2}{\partial t^2}u &= \left(\frac{\partial^2}{\partial t^2} + 2\varepsilon\frac{\partial^2}{\partial \tau\partial t} + 2\varepsilon^2\frac{\partial^2}{\partial \sigma\partial t} + \varepsilon^2\frac{\partial^2}{\partial \tau^2} + 2\varepsilon^3\frac{\partial^2}{\partial \sigma\partial \tau} + \varepsilon^4\frac{\partial^2}{\partial \sigma^2}\right)\ \varepsilon\sum_{n=0}^{\infty}\varepsilon^n\mathcal{U}_n\end{aligned}$$

we have

$$\begin{aligned}\frac{\partial^2}{\partial x^2}u = \varepsilon\frac{\partial^2}{\partial x^2}\mathcal{U}_0 &+ \varepsilon^2\left\{\frac{\partial^2}{\partial x^2}\mathcal{U}_1 + 2\frac{\partial^2}{\partial X\partial x}\mathcal{U}_0\right\}\\ &+ \varepsilon^3\left\{\frac{\partial^2}{\partial x^2}\mathcal{U}_2 + 2\frac{\partial^2}{\partial X\partial x}\mathcal{U}_1 + \frac{\partial^2}{\partial X^2}\mathcal{U}_0\right\} + \mathcal{O}(\varepsilon^4)\end{aligned} \tag{3.6}$$

and

$$\begin{aligned}\frac{\partial^2}{\partial t^2}u = \varepsilon\frac{\partial^2}{\partial t^2}\mathcal{U}_0 &+ \varepsilon^2\left\{\frac{\partial^2}{\partial t^2}\mathcal{U}_1 + 2\frac{\partial^2}{\partial \tau\partial t}\mathcal{U}_0\right\}\\ &+ \varepsilon^3\left\{\frac{\partial^2}{\partial t^2}\mathcal{U}_2 + 2\frac{\partial^2}{\partial \tau\partial t}\mathcal{U}_1 + 2\frac{\partial^2}{\partial \sigma\partial t}\mathcal{U}_0 + \frac{\partial^2}{\partial \tau^2}\mathcal{U}_0\right\} + \mathcal{O}(\varepsilon^4).\end{aligned} \tag{3.7}$$

Applying the Cauchy product formula to the series (3.3) yields

$$u^3(x,t) = \varepsilon^3\sum_{n=0}^{\infty}\varepsilon^n\sum_{k=0}^{n}\sum_{l=0}^{k}\mathcal{U}_{n-k}\mathcal{U}_{k-l}\mathcal{U}_l. \tag{3.8}$$

In the following let us define the linear differential operator $\mathcal{L}$ for the Klein-Gordon equation (3.2) as

$$\mathcal{L}(\varphi) \coloneqq \frac{\partial^2}{\partial t^2}\varphi - \alpha\frac{\partial^2}{\partial x^2}\varphi + \beta\varphi$$

and its nonlinearity as

$$\mathcal{N}(\varphi) = \lambda\varphi^3.$$

If we insert these representations of the derivatives into the wave equation (3.2) this gives

$$\begin{aligned} 0 = \quad & \frac{\partial^2}{\partial t^2}u(x,t) - \alpha\frac{\partial^2}{\partial x^2}u(x,t) + \beta u(x,t) - \lambda u^3(x,t) \\ = \quad & \varepsilon\,\mathcal{L}\left(\mathcal{U}_0\right) + \varepsilon^2\left\{\mathcal{L}\left(\mathcal{U}_1\right) + 2\frac{\partial^2}{\partial\tau\partial t}\mathcal{U}_0 - 2\alpha\frac{\partial^2}{\partial X\partial x}\mathcal{U}_0\right\} \\ & + \varepsilon^3\left\{\mathcal{L}\left(\mathcal{U}_2\right) + 2\frac{\partial^2}{\partial\tau\partial t}\mathcal{U}_1 - 2\alpha\frac{\partial^2}{\partial X\partial x}\mathcal{U}_1 + 2\frac{\partial^2}{\partial\sigma\partial t}\mathcal{U}_0 + \frac{\partial^2}{\partial\tau^2}\mathcal{U}_0 - \alpha\frac{\partial^2}{\partial X^2}\mathcal{U}_0 - \lambda\mathcal{U}_0^3\right\} \\ & + \mathcal{O}(\varepsilon^4) \end{aligned} \tag{3.9}$$

such that we get a sequence of partial differential equations for the $\mathcal{U}_n$ by equating the terms in front of powers of ε to 0. As before in the two examples of the previous chapter we can translate the initial and boundary conditions for $u(x,t)$ into initial and boundary conditions for the functions $\mathcal{U}_n$ by recalling the representation (3.3) of u and its derivatives in (3.5). Then comparing the the terms in front of each order of ε leads to the subsequent differential equations and its corresponding initial and boundary values.

The term in front of ε gives the equation

$$\mathcal{L}(\mathcal{U}_0) = 0,$$

which is equivalent to the equation

$$\begin{aligned} \frac{\partial^2}{\partial t^2}\mathcal{U}_0 - \alpha\frac{\partial^2}{\partial x^2}\mathcal{U}_0 + \beta\mathcal{U}_0 &= 0 \\ \mathcal{U}_0(\,\cdot\,,\varepsilon\,\cdot\,,0,0,0) &= \varepsilon^{-1}u_0, \\ \partial_t\mathcal{U}_0(\,\cdot\,,\varepsilon\,\cdot\,,0,0,0) &= \varepsilon^{-1}u_1. \end{aligned}$$

It is solved by a function of the form

$$\mathcal{U}_0(x,X,t,\tau,\sigma) = A(X,\tau,\sigma)e^{i(kx-\omega t)} + A^*(X,\tau,\sigma)e^{-i(kx-\omega t)},$$

where the relation between ω and k has to satisfy the *dispersion relation*

$$\omega(k) = \sqrt{\alpha k^2 + \beta}. \tag{3.10}$$

We call ω the frequency of the wave and k the corresponding wave number. $A(X,\tau,\sigma)$ is a yet unspecified amplitude function dependent on the slow variables, which we want to derive in the following steps, A^* denotes its complex conjugate.

Inserting this ansatz into the differential equation given by the ε^2 term, i.e.

$$\begin{aligned} \mathcal{L}(\mathcal{U}_1) &= \underbrace{-2\frac{\partial^2}{\partial\tau\partial t}\mathcal{U}_0 + 2\alpha\frac{\partial^2}{\partial X\partial x}\mathcal{U}_0}_{=:(*)}, \\ \mathcal{U}_1(\,\cdot\,,\varepsilon\,\cdot\,,0,0,0) &= 0, \\ \partial_t\mathcal{U}_1(\,\cdot\,,\varepsilon\,\cdot\,,0,0,0) &= -\partial_\tau\mathcal{U}_0(\,\cdot\,,\varepsilon\,\cdot\,,0,0,0), \end{aligned} \tag{3.11}$$

yields the right hand side to be

$$(*) = 2i\left(\omega\frac{\partial}{\partial\tau}A + \alpha k\frac{\partial}{\partial X}A\right)e^{i(kx-\omega t)} - 2i\left(\omega\frac{\partial}{\partial\tau}A^* + \alpha k\frac{\partial}{\partial X}A^*\right)e^{-i(kx-\omega t)}.$$

It is easy to see, that $(*)$ has the same x-t-structure as the solution to the homogeneous problem and even is a solution to this by itself. So $(*)$ represents a secularity that needs to be eliminated, since it would lead to unbounded growth in $\mathcal{U}_1$ over long periods of time. The term *x-t-structure* describes the relation between the frequency $\tilde{\omega}$ and the wave number $\tilde{k}$ of a propagating wave $e^{i(\tilde{k}x-\tilde{\omega}t)}$. So we get the first conditions to the amplitude function A and A^* respectively:

$$\omega\frac{\partial}{\partial\tau}A = -\alpha k\frac{\partial}{\partial X}A \qquad \text{and} \qquad \omega\frac{\partial}{\partial\tau}A^* = -\alpha k\frac{\partial}{\partial X}A^*, \tag{3.12}$$

which are basically only one condition, since one equation is the complex conjugate of the other.

We observe that this condition is equivalent to the wave propagating at group velocity

$$c := \frac{\alpha k}{\omega} = \omega'(k)$$

in the slow variables. We can make A satisfy this demand if we perform a coordinate transform and introduce the variable $\xi := X - c\tau$. This is pretty intuitive since we assume the wave packet to move at group velocity c.

Then $A(\xi,\sigma) = A(X - c\tau,\sigma)$ satisfies (3.12) since $\frac{\partial}{\partial\xi}A = \frac{\partial}{\partial X}A$ holds and by (3.11) the equation for $\mathcal{U}_1$ reads

$$\mathcal{L}(\mathcal{U}_1) = 0.$$

Its solution is again of the form

$$\mathcal{U}_1 = B(X,\tau,\sigma)e^{i(kx-\omega t)} + B^*(X,\tau,\sigma)e^{-i(kx-\omega t)}$$

for another function B and bounded.

Collecting the terms of order ε^3 in (3.9) gives

$$\begin{aligned}\mathcal{L}(\mathcal{U}_2) &= -2\frac{\partial^2}{\partial\sigma\partial t}\mathcal{U}_0 - \frac{\partial^2}{\partial\tau^2}\mathcal{U}_0 + \alpha\frac{\partial^2}{\partial X^2}\mathcal{U}_0 + \lambda\mathcal{U}_0^3 + 2\frac{\partial^2}{\partial\tau\partial t}\mathcal{U}_1 - 2\alpha\frac{\partial^2}{\partial X\partial x}\mathcal{U}_1\\ \mathcal{U}_2(\,\cdot,\varepsilon\,\cdot,0,0,0) &= 0,\\ \partial_t\mathcal{U}_2(\,\cdot,\varepsilon\,\cdot,0,0,0) &= -\partial_\tau\mathcal{U}_1(\,\cdot,\varepsilon\,\cdot,0,0,0) - \partial_\sigma\mathcal{U}_0(\,\cdot,\varepsilon\,\cdot,0,0,0)\end{aligned} \tag{3.13}$$

where the right hand side is equal to

$$\left(2i\omega\frac{\partial}{\partial\sigma}A + (\alpha - c^2)\frac{\partial^2}{\partial\xi^2}A + 3\lambda|A|^2A + 2i\left(\omega\frac{\partial}{\partial\tau}B + \alpha k\frac{\partial}{\partial X}B\right)\right)e^{i(kx-\omega t)} + \lambda A^3e^{3i(kx-\omega t)} + c.c.$$

We see that another secular term arises here if the term in the brackets in front of $\exp(i(kx-\omega t))$ is nonzero. Since we want to make sure that $\mathcal{U}_2$ stays bounded we have to set it equal to zero. Therefore we demand that, just like A, also B satisfies

$$\omega\frac{\partial}{\partial\tau}B + \alpha k\frac{\partial}{\partial X}B = 0$$

and set $B(X,\tau,\sigma) = B(\xi,\sigma)$ then this term vanishes. Furthermore we demand that $A(\xi,\sigma)$ and its complex conjugate satisfy

$$2i\omega\frac{\partial}{\partial\sigma}A + \omega\omega''\frac{\partial^2}{\partial\xi^2}A + 3\lambda|A|^2A = 0 \tag{3.14}$$

and

$$-2i\omega\frac{\partial}{\partial\sigma}A^* + \omega\omega''\frac{\partial^2}{\partial\xi^2}A^* + 3\lambda|A|^2A^* = 0$$

respectively, which is again just the complex conjugate equation of the other and thus can be omitted. Here we used that $\alpha - c^2 = \omega\omega''$.

This demand on A represents the nonlinear Schrödinger equation, which is very important in mathematics and physics. It will be treated more extensive in the subsequent section. The remainder term

$$\lambda A^3 e^{3i(kx-\omega t)} + \lambda A^{*3} e^{-3i(kx-\omega t)}$$

is harmless since $\omega(3k) \neq 3\omega(k)$ for almost all k and thus it is not in resonance with the homogeneous solution.

Hence we can assume that $\mathcal{U}_2$ is bounded. For any of the further functions $\mathcal{U}_l$, $l \geq 3$ we do not want to make any statement and we set the approximation

$$\varepsilon\psi_{NLS}(x,t) := \varepsilon A(\xi,\sigma)e^{i(kx-\omega t)} + c.c.$$

to the solution u. Recalling that we made the ansatz (3.3) for u we have

$$u - \varepsilon\psi_{NLS} = \varepsilon^2\mathcal{U}_1 + \varepsilon^3\mathcal{U}_2 + \varepsilon\sum_{n=3}^{\infty}\varepsilon^n\mathcal{U}_n = \mathcal{O}(\varepsilon^2), \qquad \forall t \in [0, T_0/\varepsilon^2]$$

for some $T_0 > 0$ in $\mathcal{O}(1)$. This restriction on the time t can be explained as follows:

Even though $\mathcal{U}_1$ and $\mathcal{U}_2$ are bounded for all times t the function $\mathcal{U}_3$ belonging to the formal order $\mathcal{O}(\varepsilon^4)$ may contain secular terms which grow linearly with t. This means that for $t \in \mathcal{O}(\varepsilon^{-2})$ there holds $\varepsilon^4\mathcal{U}_3 \in \mathcal{O}(\varepsilon^2)$ but for even larger $t \in \mathcal{O}(\varepsilon^{-3})$ this approximation loses validity because then $\varepsilon^4\mathcal{U}_3 \in \mathcal{O}(\varepsilon)$. Therefore we have to restrict the time validity interval to $t \in [0, T_0/\varepsilon^2]$.

3.2. The Gross-Pitaevskii Equation

As we have seen, we encounter the nonlinear Schrödinger equation (NLS), also called *Gross-Pitaevskii equation*,

$$\begin{aligned} \frac{\partial}{\partial\sigma}A &= i\nu_1\frac{\partial^2}{\partial\xi^2}A + i\nu_2|A|^2A, \quad \xi,\ \sigma \in \mathbb{R} \quad A(\xi,\sigma) \in \mathbb{C} \\ A(\,\cdot\,,0) &= A_0, \end{aligned} \tag{3.15}$$

with real coefficients ν_1, ν_2 by applying a multiple scale method to the nonlinear Klein-Gordon equation

$$\frac{\partial^2}{\partial t^2}u = \alpha\frac{\partial^2}{\partial x^2}u - \beta u + \lambda u^3. \tag{3.16}$$

Now we want to collect some properties of the NLS equation, which are taken mostly from [6]. The NLS equation serves as an amplitude equation for the amplitude $A(\xi,\sigma)$ of the so called *NLS approximation*

$$\varepsilon\psi_{NLS} = \varepsilon A(\xi,\sigma)e^{i(kx-\omega t)} + \varepsilon A^*(\xi,\sigma)e^{-i(kx-\omega t)}$$

to the exact solution u of the wave equation, that we obtained by applying a multiple scale method. In our example we have found

$$\nu_1 = \frac{\alpha - c^2}{2\omega}, \qquad \nu_2 = \frac{3\lambda}{2\omega}. \tag{3.17}$$

Here the dispersion relation $\omega^2 = \alpha k^2 + \beta$ has to be satisfied and $c := \omega'(k)$ is the group velocity of the wave packet.

The NLS equation can be transformed into the form

$$\frac{\partial}{\partial\sigma}A = -i\frac{\partial^2}{\partial\xi^2}A + i\alpha|A|^2A, \quad \alpha = \pm 1 \tag{3.18}$$

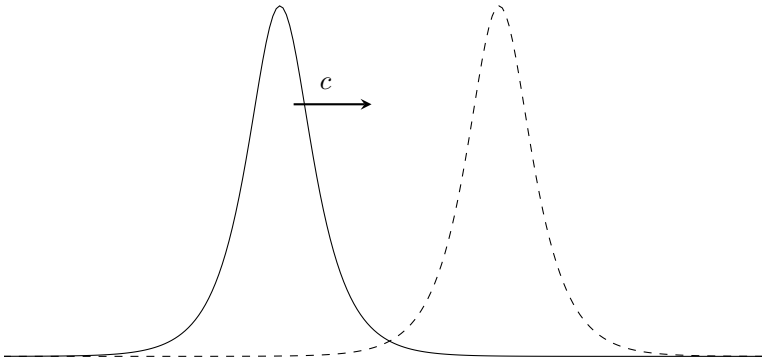


Figure 3.2. (RUN_pulse.m)
A localized pulse (or *soliton*) solution propagates at constant velocity c without changing shape.

by rescaling A, σ and ξ, where we distinguish between the *defocusing* case $\alpha = 1$ and the *focusing* case $\alpha = -1$, whose difference can be found in the special form of the solutions. Generally we have the focusing case if ν_1 and ν_2 have both like algebraic signs.

If we firstly have a look at the linear problem

$$\frac{\partial}{\partial\sigma}A = -i\frac{\partial^2}{\partial\xi^2}A,$$

we see that it shows dispersion, that is given by the relation $\omega = k^2$ for solutions $e^{ikx+i\omega t}$. That means that the group velocity $c_g = \omega'(k)$ is dependent on k, such that waves with different frequencies travel at different speed and a travelling wave packet will dissolve. Its solution is given by

$$A(\xi,\sigma) = \frac{1}{\sqrt{4\pi i\sigma}}\int_{-\infty}^{\infty} e^{\frac{-(\xi-Y)^2}{4i\sigma}}A(Y,0)dY$$

and we immediately obtain the estimate

$$\sup_{\xi\in\mathbb{R}}|A(\xi,\sigma)| \leq \frac{1}{\sqrt{4\pi\sigma}}\int_{-\infty}^{\infty}|A(\xi,0)|d\xi, \quad \forall\sigma > 0.$$

Therefore for spatially localised initial conditions the solutions decay uniformly towards zero with a rate $\sigma^{-\frac{1}{2}}$. Furthermore dispersion conserves energy but spreads it all over the real axis.

If we consider the focusing case $\alpha = -1$, the focusing of energy by the nonlinearity and the defocusing of energy by dispersion are in some equilibrium, such that so called *pulse* or *soliton* solutions of the form

$$A(\xi,\sigma) = B(\xi)e^{i\omega\sigma}$$

can exist, where B has to satisfy $0 = B'' + \omega B - \alpha B^3$. That means, that the dissolving of the wave packet by dispersion is just evened out by the nonlinear effect.

A soliton is a stable *solitary wave*, which is a localized pulse that propagates at constant velocity without changing shape over time, cf. figure 3.2. For instance think of a burning candle, whose flame maintains its shape while it advances into the wax at constant velocity. Here the diffusion of heat from the flame into the wax and the nonlinear energy release of the vaporizing wax cancel out. For more details on soliton solutions and their properties we refer to [8], chapter 4.

This effect can also be seen for example if optical waves travel through a so called *Kerr medium* such as a glass fiber. The refraction index of the material causes the dispersion and therefore the divergence

of the wave packet, whereas at the same time a nonlinear effect is responsible for convergence of waves with different frequencies such that the result is a (nearly) not changing wave packet.

Let us give a result on the local existence and uniqueness of solutions in Sobolev spaces. Therefore let the space H^m be equipped with the norm $\|u\|_{H^m} = \max\limits_{j\in\{0,\dots,m\}} \|\partial_x^j u\|_{L^2}$.

Theorem 3.2. *Let $m \geq 1$ and $A_0 \in H^m(\mathbb{R})$ a complex function. Then there exists a time $T_0 = T_0(\|A_0\|_{H^m}) > 0$ and a unique solution $A \in C([0,T_0], H^m)$ of the NLS (3.18) with initial value A_0.*

Proof (of theorem 3.2). *In lemma A.22 we proved that $L = -i\frac{\partial^2}{\partial\xi^2}$ generates the C_0 contraction semigroup $e^{-i\sigma\frac{\partial^2}{\partial\xi^2}}$. Furthermore one can show that the mapping $F(u) \coloneqq i\alpha|u|^2u$ is locally Lipschitz continuous in $H^m(\mathbb{R})$ by modifying the proof of Lemma 1.2 in chapter 8.1 in Pazy's book [19], which was given for the case $H^2(\mathbb{R}^2)$. We assume the solution A to be continuous in σ and ξ, therefore it is also $F(A)$ and theorem A.23 applies. Set $T_0 \coloneqq \left(4|\alpha|\,\|A_0\|^2_{H^m}\right)^{-1}$. Then together with the previous results a solution of the NLS equation (3.18) is given by the variation of constants formula*

$$A(\,\cdot\,,\sigma) = e^{-i\sigma\frac{\partial^2}{\partial\xi^2}}A(\,\cdot\,,0) + \int_0^\sigma e^{-i(\sigma-s)\frac{\partial^2}{\partial\xi^2}} i\alpha|A(\,\cdot\,,s)|^2A(\,\cdot\,,s)ds,$$

where the right hand side is a contraction in the space

$$\mathscr{A} \coloneqq \left\{ A \in C([0,T_0],H^m) \quad | \sup_{\sigma\in[0,T_0]} \|A(\,\cdot\,,\sigma)\|_{H^m} \leq 2\|A_0\|_{H^m} \right\}$$

equipped with the norm $\|A\|_{\mathscr{A}} = \sup_{\sigma\in[0,T_0]}\|A(\,\cdot\,,\sigma)\|_{H^m}$. The local uniqueness is obtained by the contraction mapping principle for the fixed point equation

$$A(\,\cdot\,,\sigma) = F(A)(\,\cdot\,,\sigma),$$

where

$$F(A)(\,\cdot\,,\sigma) = e^{-i\sigma\frac{\partial^2}{\partial\xi^2}}A(\,\cdot\,,0) + \int_0^\sigma e^{-i(\sigma-s)\frac{\partial^2}{\partial\xi^2}} i\alpha|A(\,\cdot\,,s)|^2A(\,\cdot\,,s)ds.$$

Let $A_1, A_2 \in \mathscr{A}$ with $A_1(\,\cdot\,,0) = A_2(\,\cdot\,,0) = A_0$ then

$$\begin{aligned}
\|F(A_1) - F(A_2)\|_{\mathscr{A}} &\leq \|A_1(0) - A_2(0)\|_{\mathscr{A}} + \int_0^{T_0} |\alpha| \sup_{s\in[0,T_0]} \left\| \,|A_1(\,\cdot\,,s)|^2A_1(\,\cdot\,,s) - \,|A_2(\,\cdot\,,s)|^2A_2(\,\cdot\,,s)\right\|_{H^m} ds \\
&\leq \int_0^{T_0} |\alpha| \sup_{s\in[0,T_0]} \left(\max\left(\|A_1(\,\cdot\,,s)\|^2_{H^m}, \|A_2(\,\cdot\,,s)\|^2_{H^m}\right) \| \,A_1(\,\cdot\,,s) - \,A_2(\,\cdot\,,s)\|_{H^m}\right) ds \\
&\leq 4|\alpha|T_0\|A_0\|^2_{H^m}\,\|A_1 - A_2\|_{\mathscr{A}}\,.
\end{aligned}$$

Hence for our choice of T_0 we see that $\|F(A_1) - F(A_2)\|_{\mathscr{A}} \leq \|A_1 - A_2\|_{\mathscr{A}}$, which yields the local existence and uniqueness of a solution. □

Lemma 3.3. *The solution A of (3.18) conserves the L^2 norm and energy:*

$$\begin{aligned}
\frac{d}{d\sigma}\|A\|^2_{L^2} &= 0 \\
\frac{d}{d\sigma}E(A(\,\cdot\,,\sigma)) &= 0,
\end{aligned}$$

where

$$E\left(u(\,\cdot\,,t)\right) \coloneqq \frac{1}{2}\int |\partial_x u|^2 dx + \frac{1}{4}\alpha \int |u|^4 dx \tag{3.19}$$

for a function $u(\,\cdot\,,t) \in C\left([0,T_0], H^1(\mathbb{R})\right)$.

Proof. Applying integration by parts we have

$$\begin{aligned}\frac{d}{d\sigma}\|A\|_{L^2}^2 = \frac{d}{d\sigma}\int AA^* d\xi &= 2\,\mathrm{Re}\int (\partial_\sigma A)\,A^* d\xi = 2\,\mathrm{Re}\int \left(-i(\partial_\xi^2 A - \alpha|A|^2 A)\right) A^* d\xi \\ &= -2\,\mathrm{Im}\int \left(|\partial_\xi A|^2 + \alpha|A|^4\right) d\xi = 0\end{aligned}$$

and thus have the conservation of the L^2 *norm.*

By definition

$$\begin{aligned}\frac{d}{d\sigma}E(u) = \frac{d}{d\sigma}\left(\frac{1}{2}\int |\partial_\xi A|^2 d\xi + \frac{1}{4}\alpha\int |A|^4 d\xi\right) &= \int \mathrm{Re}\left((\partial_\xi A)\partial_\xi\left(\frac{\partial}{\partial\sigma}A^*\right)\right))d\xi + \alpha\int \mathrm{Re}\left(|A|^2 A\frac{\partial}{\partial\sigma}A^*\right) d\xi \\ = \mathrm{Re}\int\left(-\frac{\partial^2}{\partial\xi^2}A + \alpha|A|^2A\right)\frac{\partial}{\partial\sigma}A^* d\xi = \mathrm{Re}\int i\frac{\partial}{\partial\sigma}A\frac{\partial}{\partial\sigma}A^* d\xi &= \mathrm{Re}\int i|\frac{\partial}{\partial\sigma}A|^2 d\xi = 0\end{aligned}$$

and therefore $E\left(A(\,\cdot\,,\sigma)\right) = E\left(A_0\right)\ \forall\sigma\in[0,T_0]$. □

Global existence and uniqueness of solutions then follows from the results in chapter 10 of the lecture notes [12].

3.3. The Residual of the NLS approximation

To get an estimate on the error that we make by using the NLS ansatz

$$\tilde{u}(x,t) \coloneqq \varepsilon A\left(\xi,\sigma\right)e^{i(kx-\omega t)} + \varepsilon A^*\left(\xi,\sigma\right)e^{-i(kx-\omega t)}, \quad \xi = \varepsilon(x-ct),\ \sigma = \varepsilon^2 t$$

as an approximation to the solution of the nonlinear wave equation, we abbreviate $E = e^{i(kx-\omega t)}$ and have a look at its residual

$$\begin{aligned}\mathrm{Res}(\tilde{u}) = &\quad \mathcal{L}(\tilde{u}) - \mathcal{N}(\tilde{u}) \\ = &\quad \frac{\partial^2}{\partial t^2}\tilde{u} - \alpha\frac{\partial^2}{\partial x^2}\tilde{u} + \beta\tilde{u} - \lambda\tilde{u}^3 \\ = &\quad \varepsilon E\left((\alpha k^2 + \beta - \omega^2)A\right) \\ &\quad + \varepsilon^2 E\left((2i\omega c - 2i\alpha k)\frac{\partial}{\partial\xi}A\right) \\ &\quad + \varepsilon^3 E\left(-(2i\omega\frac{\partial}{\partial\sigma}A + (\alpha - c^2)\frac{\partial^2}{\partial\xi^2}A + 3\lambda|A|^2A\right) \\ &\quad + \varepsilon^3 E^3 A^3 + \varepsilon^4 E\left(-2c\frac{\partial^2}{\partial\xi\partial\sigma}A\right) + \varepsilon^5 E\frac{\partial^2}{\partial\sigma^2}A + c.c.\end{aligned} \tag{3.20}$$

Provided that A satisfies the NLS equation (3.14) it follows from the dispersion relation $\omega^2 = \alpha k^2 + \beta$ and the group velocity $c = \frac{\alpha k}{\omega}$ that the first terms vanish and formally $\mathrm{Res}(\tilde{u}) = \mathcal{O}(\varepsilon^3)$.

We can improve these formal orders of the residual to estimates in norms. Therefore we introduce the subsequent function spaces.

Definition 3.4 (cf. chapter 5.1 in [6])**.** *We define the space of uniformly continuous and uniformly bounded functions by*

$$C_b^0(\mathbb{R},\mathbb{R}) \coloneqq \left\{u : \mathbb{R}\to\mathbb{R} \mid u \text{ is uniformly continuous and uniformly bounded}\right\},$$

equipped with the norm

$$\|u\|_{C_b^0} = \sup_{x\in\mathbb{R}} |u(x)|.$$

Furthermore, the space of m times differentiable functions $u : \mathbb{R} \to \mathbb{R}$ with uniformly continuous and uniformly bounded derivatives $\frac{\partial^j}{\partial x^j} u$ is given by

$$C_b^m(\mathbb{R},\mathbb{R}) := \left\{ u : \mathbb{R} \to \mathbb{R} \mid \frac{\partial^j}{\partial x^j} u \in C_b^0 \quad \forall j = 1,\dots,m \right\}$$

equipped with the norm

$$\|u\|_{C_b^m} = \max_{j\in\{1,\dots,m\}} \|\frac{\partial^j}{\partial x^j} u\|_{C_b^0}.$$

Omitting the spatial argument we already found

$$\mathrm{Res}(\tilde{u}(t)) = \varepsilon^3 E^3 A^3(\sigma) + \varepsilon^4 E \left(-2c \frac{\partial^2}{\partial\xi\partial\sigma} A(\sigma) \right) + \varepsilon^5 E \frac{\partial^2}{\partial\sigma^2} A(\sigma) + c.c.,$$

cf. (3.20). The estimate in the C_b^0 norm is

$$\|\mathrm{Res}(\tilde{u})\|_{C_b^0} \le s_1 + s_2 + s_3,$$

where

$$\begin{aligned}
s_1 &= 2 \left\| \varepsilon^3 E^3 \, A^3(\sigma) \right\|_{C_b^0} \le 2\varepsilon^3 \|A(\sigma)\|_{C_b^0}^3, \\
s_2 &= 4 \left\| \varepsilon^4 E \, c \frac{\partial^2}{\partial\xi\partial\sigma} A(\sigma) \right\|_{C_b^0} \le 4\varepsilon^4 c \left\| \frac{\partial}{\partial\sigma} A(\sigma) \right\|_{C_b^1}, \\
s_3 &= 2 \left\| \varepsilon^5 E \, \frac{\partial^2}{\partial\sigma^2} A(\sigma) \right\|_{C_b^0} \le 2\varepsilon^5 \left\| \frac{\partial^2}{\partial\sigma^2} A(\sigma) \right\|_{C_b^0}.
\end{aligned}$$

The right hand side of the NLS equation (3.14), i.e.

$$\frac{\partial}{\partial\sigma} A = \frac{1}{2\omega} \left(i(\alpha - c^2) \frac{\partial^2}{\partial\xi^2} A + 3i\lambda |A|^2 A \right),$$

yields the estimates for $\|\frac{\partial}{\partial\sigma} A\|_{C_b^1}$ and $\|\frac{\partial^2}{\partial\sigma^2} A\|_{C_b^0}$, i.e.

$$\left\| \frac{\partial}{\partial\sigma} A \right\|_{C_b^1} \le \frac{1}{2\omega} \left((\alpha - c^2) \left\| \frac{\partial^2}{\partial\xi^2} A \right\|_{C_b^1} + 3\lambda \|A\|_{C_b^1}^3 \right) = \mathcal{O}\left(\|A\|_{C_b^3} \right)$$

and analogously

$$\left\| \frac{\partial^2}{\partial\sigma^2} A \right\|_{C_b^0} = \mathcal{O}\left(\|A\|_{C_b^4} \right).$$

This gives the following result.

Proposition 3.5 (cf. [6], Proposition 5.3.1)**.** *Let $A \in C([0,T_0], C_b^4)$ be a solution of the NLS equation. Then for all $\varepsilon_0 \in (0,1]$ there exists a $C > 0$ such that for all $\varepsilon \in (0,\varepsilon_0)$:*

$$\sup_{t\in[0,T_0/\varepsilon^2]} \|\mathrm{Res}(\tilde{u}(\,\cdot\,,t))\|_{C_b^0} \le C\varepsilon^3.$$

It is worth to mention that measuring the norm of the residual in the Sobolev spaces H^m is more suitable to get estimates on the long time scale $\mathcal{O}(1/\varepsilon^2)$ than doing so in the C_b^m spaces. But sadly we have to be aware of the following remark.

Remark 3.6. *Due to the spatial scaling we lose a factor of $\varepsilon^{-\frac{1}{2}}$ when evaluating the L^2 norm in our estimate, such that we obtain*

$$\| \operatorname{Res}(\tilde{u}(\,\cdot\,,t))\|_{H^s} = \mathcal{O}\left(\varepsilon^{5/2}\|A\|_{H^{s+4}}\right)$$

with $A \in C([0,T_0], H^{s_A})$ and $s_A \geq 0$ sufficiently large.

Unfortunately estimates for the residual are only a necessary but not sufficient condition that the NLS equation predicts correctly the behaviour of the original wave equation. Errors may sum up in time such that correctly derived amplitude equations make wrong predictions. Therefore we do not want to get into more detail at this point. More information on this topic can be found in chapter 5 of [6].

3.4. An Error Estimate on the NLS Approximation

So far we only made statements about the formal order of the residual of the NLS approximation

$$\tilde{u}(x,t) = \varepsilon A(\xi,\sigma)e^{i(kx-\omega t)} + c.c.$$

to the exact solution of the nonlinear wave equation

$$\frac{\partial^2}{\partial t^2}u = \alpha\frac{\partial^2}{\partial x^2} - \beta u + \lambda u^3.$$

Again the slow variables ξ and σ are given by $\varepsilon(x-ct)$ and $\varepsilon^2 t$ respectively and $A = A(\xi,\sigma)$ satisfies the NLS equation

$$2i\omega\frac{\partial}{\partial\sigma}A + (\alpha - c^2)\frac{\partial^2}{\partial\xi^2}A + 3\lambda|A|^2A = 0.$$

In order to prove that the actual error of the NLS approximation $\tilde{u}$ stays small for a certain period of time, we define the so called *energy space*

$$Y \coloneqq H^1(\mathbb{R}) \times L^2(\mathbb{R})$$

equipped with the norm $\|\cdot\|_Y$ defined by

$$\|(u,v)\|_Y \coloneqq \left(\|u\|_{H^1}^2 + \|v\|_{L^2}^2\right)^{1/2} = \left(\int_{\mathbb{R}} (\partial_x u)^2 + u^2 + v^2 dx\right)^{1/2},$$

which is a Banach space, as one can show. Due to the following theorem the error of the NLS approximation goes to 0 as $\varepsilon \to 0$.

Theorem 3.7 (cf. Theorem 5.3.6 in [6], and Theorem 3.1 in[13])**.** *Let $A = A(\xi,\sigma)$ be a solution of the NLS equation such that $\partial_\xi^n\partial_\sigma^l A \in C([0,T_1], L^2(\mathbb{R}))$ with $n+l \leq 2$ and let $\tilde{u}$ be the NLS approximation from above. Then for every $T_0 \leq T_1$ and every $d > 0$ there exist ε_0, $C > 0$ such that for all $\varepsilon \in (0,\varepsilon_0)$ there holds:*

Let $u = u(x,t)$ be a solution of the nonlinear wave equation (3.2) such that

$$\|\,(u(0),\partial_t u(0)) - (\tilde{u}(0),\partial_t\tilde{u}(0))\,\|_Y \leq d\varepsilon^{3/2}$$

then $\tilde{u}$ satisfies the estimate

$$\|\,(u(t),\partial_t u(t)) - (\tilde{u}(t),\partial_t\tilde{u}(t))\,\|_Y \leq C\varepsilon^{3/2} \qquad \forall\, t \in [0, T_0/\varepsilon^2].$$

Proof (mostly taken from the proofs to theorems 2.1 and 3.1 in [13]). *We want to show that the error* $u(\cdot,t)-\tilde{u}(\cdot,t)$ *remains of order* $\mathcal{O}(\varepsilon^{3/2})$ *for times* $t\leq\frac{T_0}{\varepsilon^2}$. *However we have already seen that inserting* $\tilde{u}$ *into the wave equation leaves a residual term* $\varepsilon^3A^3E^3$. *Unfortunately this leads to an error* $\mathcal{O}(\varepsilon)$ *when integrating over* $[0,T_0/\varepsilon^2]$. *As a start we derive all the estimates for arbitrary* $t\geq 0$, *but later on in this proof we will see that we have to bound* t *by* $\varepsilon^2t\leq T_0$.

The proof is divided into three parts: Firstly we derive a differential equation for the error, which we solve in the second part using the variation-of-constants formula. In the third part we use Gronwall's inequality to obtain an estimate on the error.

a) ***Derive a differential equation for the error:*** *To handle this problem we make an improved ansatz*

$$v_A(x,t)=\tilde{u}+\left(\varepsilon^3A_3(\xi,\sigma)e^{3i(kx-\omega t)}+c.c.\right)$$

and since A *solves the NLS equation we obtain the residual*

$$\operatorname{Res}(v_A)=\varepsilon^3E^3\left(A^3-(9\omega^2-9\alpha k^2+\beta)A_3\right)+c.c.+\mathcal{O}(\varepsilon^4)$$

which is of order $\mathcal{O}(\varepsilon^4)$ *if we set* $A_3:=\frac{1}{9\omega^2-9\alpha k^2+\beta}A^3(\xi,\sigma)$. *Therefore*

$$v_A(x,t)=\varepsilon A(\xi,\sigma)e^{i(kx-\omega t)}+\frac{1}{9\omega^2-9\alpha k^2+\beta}\varepsilon^3A^3(\xi,\sigma)e^{3i(kx-\omega t)}+c.c.$$

and because of the scaling $\xi=\varepsilon(x-ct)$ *and lemma A.17 the* L^2 *norm of* v_A *can be estimated as*

$$\|v_A(\cdot,t)\|_{L^2}\leq 2\varepsilon^{\frac{1}{2}}\|A(\cdot,\sigma)\|_{L^2}+2\varepsilon^{\frac{5}{2}}\|A(\cdot,\sigma)\|_{H^1}^3$$

and thus v_A *is bounded. Similarly, since the leading term of the residual* $\operatorname{Res}(v_A)$ *is of the form* $\varepsilon^4\sum_{n+l\leq 2}c_{n,l}\,\partial_\xi^n\partial_\sigma^lAe^{i(kx-\omega t)}$ *for some constants* $c_{n,l}$, *we have* $\|\operatorname{Res}(v_A)\|_{L^2}=\mathcal{O}(\varepsilon^{\frac{7}{2}})$. *For the error we make the ansatz*

$$\varepsilon^{\frac{3}{2}}R=u-v_A, \tag{3.21}$$

which yields the partial differential equation

$$\begin{aligned}\varepsilon^{\frac{3}{2}}\frac{\partial^2}{\partial t^2}R=\;&\varepsilon^{\frac{3}{2}}\alpha\frac{\partial^2}{\partial x^2}R-\varepsilon^{\frac{3}{2}}\beta R+\lambda\left(\varepsilon^{\frac{3}{2}}R\right)^3-\underbrace{\left(\frac{\partial^2}{\partial t^2}v_A-\alpha\frac{\partial^2}{\partial x^2}v_A+\beta v_A-\lambda v_A^3\right)}_{=\operatorname{Res}(v_A)}\\&+3\lambda\left(\varepsilon^3R^2v_A+\varepsilon^{\frac{3}{2}}Rv_A^2\right)\end{aligned}$$

or equivalently

$$\frac{\partial^2}{\partial t^2}R=\alpha\frac{\partial^2}{\partial x^2}R-\beta R-\varepsilon^{-\frac{3}{2}}\operatorname{Res}(v_A)+\lambda\left(\varepsilon^3R^3+3\varepsilon^{\frac{3}{2}}R^2v_A+3Rv_A^2\right).$$

b) ***Solution of the error equation via the variation-of-constants formula:***

We rewrite this second-order equation as a first order system for (R,∂_tR) *in the Banach space* Y *and define the operator matrix* $B:=\begin{bmatrix}0 & I\\ \alpha\frac{\partial^2}{\partial x^2}-\beta & 0\end{bmatrix}$ *on* Y. *Hence we are now looking for solutions* $y:=\begin{bmatrix}R\\S\end{bmatrix}$ *of the system*

$$\frac{\partial}{\partial t}y=By+\varepsilon^{\frac{3}{2}}\begin{bmatrix}0\\a(\varepsilon,t)R\end{bmatrix}+\varepsilon^{\frac{5}{2}}N(\varepsilon,t,R)-\varepsilon^2\rho(\varepsilon,t),$$

with

$$y(0) = \begin{bmatrix} R(\,\cdot\,,0) \\ \partial_t R(\,\cdot\,,0) \end{bmatrix} = \varepsilon^{-\frac{3}{2}} \begin{bmatrix} u(\,\cdot\,,0) - v_A(\,\cdot\,,0) \\ \partial_t u(\,\cdot\,,0) - \partial_t v_A(\,\cdot\,,0) \end{bmatrix} \Longrightarrow \|y(0)\|_Y \leq 2d,$$

where

$$a(\varepsilon,t) \coloneqq 3\lambda\varepsilon^{-\frac{3}{2}} v_A^2,$$

$$N(\varepsilon,t,R) \coloneqq \lambda \begin{bmatrix} 0 \\ \varepsilon^{\frac{1}{2}} R^3 + 3\varepsilon^{-1} v_A R^2 \end{bmatrix},$$

and

$$\rho(\varepsilon,t) \coloneqq \varepsilon^{-\frac{7}{2}} \begin{bmatrix} 0 \\ \mathrm{Res}(v_A) \end{bmatrix}.$$

This system can be formally solved by the variation of constants formula

$$y(t) = G(t)y(0) + \varepsilon^{\frac{3}{2}} \int_0^t G(t-s) \left(\begin{bmatrix} 0 \\ a(\varepsilon,s)R(s) \end{bmatrix} + \varepsilon N\left(\varepsilon, s, R(s)\right) - \varepsilon^{\frac{1}{2}} \rho(\varepsilon,s) \right) ds,$$

where $(G(t))_{t\geq 0} = \left(e^{tB}\right)_{t\geq 0}$ *denotes the* C_0 *semigroup generated by* B *on the space* Y*. For more details we refer to [12], chapter 5 and 6, and [19], chapter 7.4. Let* $C > 0$ *be large enough such that it sates the subsequent estimates. Because* $G(\,\cdot\,)$ *is a* C_0 *semigroup it is bounded, such that for* $t \in [0,T]$*,* $T > 0$ *and every initial value* $y_0 \coloneqq (R_0, S_0)$ *there holds* $\|G(t)y_0\|_Y \leq C\|y_0\|_Y$*.*

c) ***Estimation of the error via Gronwall's lemma:***

Therefore we get the estimate

$$\|y(t)\|_Y \leq C\|y(0)\|_Y + \varepsilon^{\frac{3}{2}} C \int_0^t \left(\|a(\varepsilon,s)R(s)\|_{L^2} + \lambda\varepsilon \left\| \varepsilon^{\frac{1}{2}} R^3(s) + 3\varepsilon^{-1} v_A R^2(s) \right\|_{L^2} + \varepsilon^{\frac{1}{2}} \|\rho(\varepsilon,s)\|_Y \right) ds.$$

Now having $\tilde{u}(0) - v_A(0) = \mathcal{O}(\varepsilon^3)$ *we can bound* $\|y(0)\|_Y \leq 2d$*. Since* $\varepsilon^{-1} v_A = \mathcal{O}(1)$ *and from the assumption on the NLS solution* A *together with* $\|\mathrm{Res}(v_A)\|_{L^2} = \mathcal{O}(\varepsilon^{\frac{7}{2}})$ *we obtain* $\|\rho(\varepsilon,t)\|_Y \leq C$ *and* $\|a(\varepsilon,t)\|_\infty \leq \varepsilon^{\frac{1}{2}} C$*. Furthermore* $\|R(t)\|_{L^2} \leq \|R(t)\|_{H^1} = \mathcal{O}(1)$ *and* $\|R(t)\|_{L^2} \leq C\,\|y(t)\|_Y$ *by definition.*

Applying again lemma A.17, this enables us to bound the nonlinear term for each $D > 0$ *and for all* y *with* $\|y\|_Y \leq D$ *like*

$$\|\varepsilon^{\frac{1}{2}} R^3(s) + 3\varepsilon^{-1} v_A R^2(s)\|_{L^2} \leq M, \quad M > 0.$$

Altogether

$$\|y(t)\|_Y \leq 2Cd + \int_0^t \varepsilon^2 C^2 \|y(s)\|_Y ds + \varepsilon^2 tC \left(\varepsilon^{\frac{1}{2}} \lambda M + C \right),$$

as long as $y(t)$ *stays in the ball of radius* D*. Now define*

$$r(t) \coloneqq \|y(t)\|_Y, \qquad b(t) \coloneqq 2Cd + \varepsilon^2 tC \left(\varepsilon^{\frac{1}{2}} \lambda M + C \right) \qquad \text{and} \qquad \gamma(t) \coloneqq \varepsilon^2 C^2.$$

Since $\|R(t)\|_{L^2} \leq \|y(t)\|_Y$ *we obtain the integral inequality*

$$r(t) \leq b(t) + \int_0^t \gamma(s) r(s) ds.$$

Obviously if we demand $\varepsilon^2 t \leq T_0$*, all requirements to apply Gronwall's lemma A.24 are satisfied and since* b *is monotonically increasing and continuous, there holds for* $t \in [0, T_0/\varepsilon^2]$

$$r(t) \leq e^{\Lambda(t)} b(t),$$

where $\Lambda(t) \coloneqq \int_0^t \gamma(\tau) d\tau = \varepsilon^2 C^2 t$*. Therefore*

$$\|y(t)\|_Y \leq \tilde{C} e^{\varepsilon^2 C^2 t}, \quad \tilde{C} \coloneqq 2Cd + T_0 C \left(\varepsilon^{\frac{1}{2}} \lambda M + C \right). \tag{3.22}$$

Let $\hat{C} = 2Cd + T_0 C(C + C)$ *and* $D = \hat{C} e^{C^2 T_0}$ *and make* ε_0 *smaller such that* $\varepsilon^{\frac{1}{2}} \lambda M \leq C$ *and thus* $\tilde{C} \leq \hat{C}$*. Then inequality (3.22) shows that* y *indeed satisfies* $\|y(t)\|_Y \leq D$ *for all* $t \leq T_0/\varepsilon^2$*. Now if we reconsider the definition of* R *in (3.21) and* v_A *we obtain the desired result from*

$$\begin{bmatrix} u(x,t) \\ \partial_t u(x,t)) \end{bmatrix} - \begin{bmatrix} \tilde{u}(x,t) \\ \partial_t \tilde{u}(x,t) \end{bmatrix} = \varepsilon^{\frac{3}{2}} \begin{bmatrix} R(x,t) \\ \partial_t R(x,t) \end{bmatrix} + \begin{bmatrix} w(x,t) \\ \partial_t w(x,t) \end{bmatrix},$$

where $w(x,t) = \frac{1}{9\omega^2 - 9\alpha k^2 + \beta} \varepsilon^3 A^3(\xi, \sigma) e^{3i(kx - \omega t)} + c.c.$

□

CHAPTER

FOUR

SPATIAL DERIVATIVES AND TIME INTEGRATION

We have seen that the method of multiple scales yields a feasible approximation to the solution of the Klein-Gordon equation. In order to numerically verify this approximation we want to present numerical methods in this chapter which enable us to obtain a good numerical solution of both the Gross-Pitaevskii equation (3.15) and the nonlinear Klein-Gordon equation (3.2).

In both equations the Laplacian appears on the right hand side. Therefore we need methods to approximate this second spatial derivative as efficient as possible with good accuracy. We will explain *spectral methods* which we use to discretize the equations in space (cf. [25], chapter 3 and 4).

Since the Gross-Pitaevskii equation is a highly oscillatory problem, the norm of the derivatives of its solution can become very large such that estimates on approximations based on Taylor expansion do not hold any more. Therefore we choose a *Strang-Splitting method* to do the time integration, which we want to present in this chapter. Furthermore we will do an error analysis of this integrator applied to the Gross-Pitaevskii equation.

Amongst other time integrators for the Klein-Gordon eqution, we present another powerful splitting method. This will be used to compute a reference solution of the Klein-Gordon equation that we want to compare with our NLS approximation in chapter 5.

In this chapter we mostly made use of the content in [11] to introduce the discrete Fourier transform and its techniques. Now let us start with an introduction to spatial derivatives.

4.1. Spatial Derivatives Based on Spectral Methods

In this section we want to give a short introduction into two popular approaches to calculate spatial derivatives numerically. First we will have a look at the method of *finite differences* developed in the 1950s. Then we will discuss *spectral methods* of the 1970s that make use of the discrete Fourier transform and therefore can be implemented very efficiently as we will see. Before we start let us fix some notation: Instead of

$$\frac{d^k}{dx^k} \quad \text{or} \quad \frac{\partial^k}{\partial x^k}$$

we may also write just

$$d_x^k \quad \text{or} \quad \partial_x^k.$$

When we want to solve partial differential equations such as the linear non dispersive wave equation with periodic boundary conditions

$$\begin{aligned} \frac{\partial^2}{\partial t^2}u(x,t) &= c^2\,\Delta u(x,t), \\ u(x,0) &= u_0(x), \qquad & \partial_t u(x,0) &= u_1(x) \quad \forall x \\ u(-L,t) &= u(L,t) \qquad & \partial_x u(-L,t) &= \partial_x u(L,t) \quad \forall t \end{aligned} \tag{4.1}$$

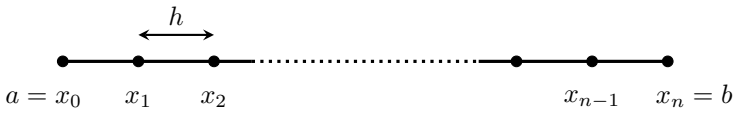
with high accuracy, it is important to have a good spatial discretization and an efficient method to calculate spatial derivatives $\partial_x^k u(\cdot,t)$.

Here and in the following $\Delta = \sum_{j=1}^{m} \partial_{x_j}^2$ denotes the Laplacian in the corresponding space dimension m. Similarly Δ_h denotes the discrete version of the Laplacian Δ.

Let $m = 1,\ n \in \mathbb{N}$ and assume that we have already discretized the interval $X = [a,b] \subset \mathbb{R}$ into a discrete interval $\tilde{X}_h = [x_0, x_1, \ldots, x_n] \subset \mathbb{R}^{n+1}$ like

$$a = x_0 < x_1 < \cdots < x_n = b,$$

where $x_j = x_0 + jh,\ j = 1,\ldots,n$ with spatial step size h, see picture below, and let $u : [a,b] \times [0,\infty) \to \mathbb{R}$ be a smooth solution of (4.1).



Finite Differences:

One possibility of approximating spatial derivatives is to use a finite differences scheme that is obtained by just expanding the smooth funcion $v : X \to \mathbb{R}$ into its Taylor series around some $x \in (a,b)$. We have

$$\begin{aligned} v(x_k + h) &= v(x_k) + h\frac{d}{dx}v(x_k) + \frac{1}{2}h^2\frac{d^2}{dx^2}v(x_k) + \frac{h^3}{6}\frac{d^3}{dx^3}v(x_k) + \mathcal{O}(h^4) \\ v(x_k - h) &= v(x_k) - h\frac{d}{dx}v(x_k) + \frac{1}{2}h^2\frac{d^2}{dx^2}v(x_k) - \frac{h^3}{6}\frac{d^3}{dx^3}v(x_k) + \mathcal{O}(h^4), \end{aligned} \tag{4.2}$$

which gives the central difference quotient as an approximation to the first spatial derivative of $v(\cdot)$ at some $x_k \in (a,b)$ just by subtracting the lower equation from the upper one, and an approximation to the second spatial derivative just by adding them up.

We define $v_j \coloneqq v(x_j)$ and thus have

$$\begin{aligned} \frac{d}{dx}v(x_k) &\approx \frac{v_{k+1} - v_{k-1}}{2h} \\ \frac{d^2}{dx^2}v(x_k) &\approx \frac{v_{k+1} - 2v_k + v_{k-1}}{h^2}, \end{aligned} \tag{4.3}$$

whose errors are both of order h^2 for smooth v, which can be seen easily by having a closer look at the Taylor series (4.2).

Remark 4.1. *The same approximations as for d_x^k can also be found for ∂_x^k if we replace $v(x)$ by some function $u(x,t)$, $u : X \times [0,\infty) \to \mathbb{R}$, which is smooth in x, and the derivatives d_x^k by the partial derivatives ∂_x^k in the equations (4.2):*

$$\begin{aligned} \frac{\partial}{\partial x}u(x_k,t) &\approx \frac{u_{k+1}(t) - u_{k-1}(t)}{2h}, \\ \frac{\partial^2}{\partial x^2}u(x_k,t) &\approx \frac{u_{k+1}(t) - 2u_k(t) + u_{k-1}(t)}{h^2}, \end{aligned} \tag{4.4}$$

where the $u_k : [0, \infty) \to \mathbb{R}$ *are still functions of the time* t.

Furthermore let us assume that $u(x,t)$ is periodic on X, i.e. for all t there holds $u(b,t) = u_n(t) = u_0(t) = u(a,t)$ which is a natural assumption on the solution of the wave equation (4.1) with periodic boundary conditions. Then $\tilde{X}_h \in \mathbb{R}^{n+1}$ reduces to an interval $X_h = [x_1, \dots, x_n] \in \mathbb{R}^n$, such that this scheme involves matrix vector multiplications of some derivative matrix $A_h \in \mathbb{R}^{n \times n}$ containing the finite difference coefficients by some vector $v(t) = [u_1(t), \dots, u_n(t)]^T \in \mathbb{R}^n$ containing the evaluations of $u(x,t)$ at each grid point x_j at time t.

Hence using the finite differences scheme (4.4) in discretized space, the partial differential equation (4.1) breaks down to an ordinary differential equation in t

$$\frac{\partial^2}{\partial t^2} v(t) = c^2 \, \Delta_h v(t), \quad \Delta_h \coloneqq \frac{1}{h^2} \begin{bmatrix} -2 & 1 & 0 & 1 \\ 1 & -2 & \ddots & 0 \\ 0 & \ddots & \ddots & 1 \\ 1 & 0 & 1 & -2 \end{bmatrix},$$

that can be solved by using an adequate numerical time integrator.

Spectral Methods

Another popular approach to calculate spatial derivatives numerically is based on the discrete Fourier transform of some 2π-periodic function $v : \mathbb{R} \to \mathbb{C}$, i.e. of a function v satisfying $v(m \cdot 2\pi) = v(0) \; \forall m \in \mathbb{Z}$, which enables us to differentiate v in its spectral space of Fourier modes instead of real space.

This will be done by using the *Fast Fourier Transform* (FFT) and its inverse (IFFT). The big advantage of this approach over the finite differences method is, that we only do multiplications of the Fourier transformed $\hat{v}_k$ by the corresponding Fourier mode k instead of costly matrix vector multiplications. So we will only need $\mathcal{O}(N \log N)$ operations instead of $\mathcal{O}(N^2)$ if we choose N to be a power of 2, cf. [11] Satz 5.11.

If we want to understand how spectral methods work, we have to introduce the discrete Fourier transform and define an interpolant which is based on it.

4.1.1. The Discrete Fourier Transform

In this section we want to recall the technique of the discrete Fourier transform. The framework is mostly taken from [11], chapter 5.

The Fourier series of a 2π-periodic function $f : \mathbb{C} \to \mathbb{C}$ reads

$$f(x) \approx \sum_{k \in \mathbb{Z}} \widehat{f}(k) e^{ikx}, \tag{4.5}$$

where the Fourier coefficients are defined by

$$\widehat{f}(k) = \frac{1}{2\pi} \int_0^{2\pi} f(x) e^{-ikx} dx. \tag{4.6}$$

Under certain conditions f agrees with its Fourier series (4.5), cf. [20], section 5.4 ff.

Let f be 2π-periodic, $N \in \mathbb{N}$ and assume that we know the values of $f(x_j)$ at equidistant points

$$x_j = \frac{2\pi j}{N}, \quad j = 0, 1, \dots, N.$$

Because of the periodicity of f, we have $f(x_N) = f(x_0)$ such that the trapezoid rule to approximate (4.6) has the representation

$$\widehat{f}_N(k) = \frac{1}{N}\sum_{j=1}^{N} f(x_j)e^{-ikx_j} = \frac{1}{N}\sum_{j=1}^{N} f(x_j)\omega_N^{kj}. \tag{4.7}$$

Here $\omega_N := e^{-i\frac{2\pi}{N}}$ denotes the N-th primitive root of unity. Sometimes we may also write ω instead of ω_N if N is clear from the context. Likewise if necessary we extend sequences $x = [x_1\ x_2\ \dots\ x_N] \in \mathbb{C}^N$ to arbitrary integer indices, where

$$x_k = x_m \qquad \text{for} \qquad k \equiv m \mod N.$$

Remark 4.2. *In this section we restrict ourselves to functions that are periodic on the interval $[0, 2\pi]$. But all results of this section are also valid for functions that are periodic on an arbitrary interval $[a, b] \subset \mathbb{R}$ with length $L = b - a$ if we do the transform $T\colon [a, b] \to [0, 2\pi]$, $T(\tilde{x}) = \frac{2\pi}{L}(\tilde{x} - a)$. Then, given a periodic grid $\tilde{x}_j = a + j\frac{L}{N}$, $j = 1, \dots, N$ on $[a, b]$ we get the grid on $[0, 2\pi]$ by evaluating $x_j = T(\tilde{x}_j)$.*

Two criterions for the Fourier series (4.5) converging to f are stated in the following theorems, that can be found in [2], chapter 7 and [14], chapter 9.

Theorem 4.3 (normal convergence of the Fourier series, see [2], Theorem 7.21.)**.** *Let $f\colon \mathbb{R} \to \mathbb{C}$ be 2π -periodic, continuous and piecewise continuously differentiable. Then the Fourier series (4.5) converges normally to f and hence $\sum_{k=-\infty}^{\infty} |\widehat{f}(k)| < \infty$.*

Theorem 4.4 (see [14], Theorem 9.1.)**.** *Let $f : \mathbb{R} \to \mathbb{C}$ be 2π -periodic and let $\sum_{k=-\infty}^{\infty} |\widehat{f}(k)| < \infty$, then the Fourier series $\sum_{k=-\infty}^{\infty} \widehat{f}(k)e^{ik\,\cdot}$ converges uniformly to f.*

Theorem 4.5 (Parseval's formula,see [14], Theorem 34.2.)**.** *Let $f : \mathbb{R} \to \mathbb{C}$ be 2π-periodic and continuous, then*

$$\|f\|_{L^2}^2 = \sum_{k=-\infty}^{\infty} |\widehat{f}(k)|^2.$$

Now let us define the discrete Fourier transform.

Definition 4.6 (discrete Fourier transform)**.** *The linear mapping $\mathcal{F}_N\colon \mathbb{C}^N \to \mathbb{C}^N$ defined by*

$$\mathcal{F}_N x = \widehat{x} \quad \text{with} \quad \widehat{x}_k = \sum_{j=1}^{N} \omega_N^{kj} x_j, \qquad k = 1, \dots, N$$

*is called **discrete Fourier transform**. The corresponding matrix $F_N := \left(\omega_N^{kj}\right)_{k,j=1}^{N}$ is called **Fourier matrix**.*

The following lemma yields the representation of the inverse mapping $\mathcal{F}_N^{-1}$ to the discrete Fourier transform.

Lemma 4.7 (orthogonality, see [11], Satz 5.2.)**.** *There holds $F_N\overline{F_N} = NI_N$ or equivalent*

$$\sum_{j=1}^{N} \omega_N^{kl}\overline{\omega}_N^{km} = \begin{cases} N, & l \equiv m \mod N \\ 0, & \text{else} \end{cases}$$

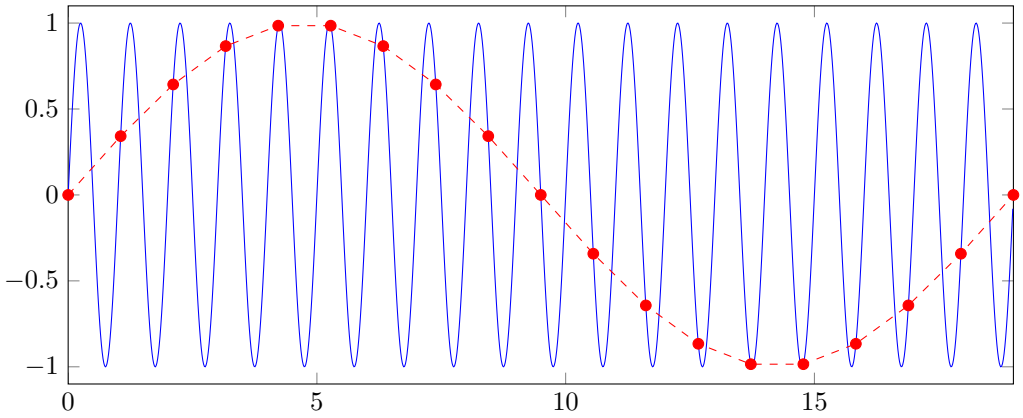


Figure 4.1. (aliasing .m)
Aliasing phenomena: $\sin(19t)$ (blue solid line) and the trigonometric interpolant in 19 points (red dashed line).

Theorem 4.8 (inverse discrete Fourier transform, [11], Satz 5.3)**.** *One has* $F_N^{-1} = \frac{1}{N}\overline{F_N} = \frac{1}{N}F_N^H$ *such that*

$$x_j = \frac{1}{N}\sum_{k=1}^{N} \omega_N^{-kj}\widehat{x}_k = \left(\mathcal{F}_N^{-1}\widehat{x}\right)_j \qquad j = 1, \ldots, N.$$

Proof. Follows from Lemma 4.7.

Theorem 4.9 (aliasing, see [11], Satz 5.6.)**.** *Let the series* $\sum_{k\in\mathbb{Z}} \widehat{f}(k)$ *be absolutely convergent, then*

$$\widehat{f}_N(k) = \sum_{j\in\mathbb{Z}} \widehat{f}(k + jN). \tag{4.8}$$

An illustration of the aliasing phenomena can be found in figure 4.1.

Corrolary 4.10 (cf. [11], Korollar 5.7)**.** *Let* $p \geq 2$ *and* $f \in C^p(\mathbb{R}, \mathbb{C})$ 2π*-periodic, then there holds*

$$\widehat{f}_N(k) - \widehat{f}(k) = \mathcal{O}(N^{-p}) \qquad \text{for} \qquad |k| \leq \frac{N}{2}.$$

Especially for $h = \frac{2\pi}{N}$ *we have*

$$h\sum_{j=1}^{N} f(x_j) - \int_0^{2\pi} f(x)dx = \mathcal{O}(h^p).$$

Proof. Applying integration by parts to (4.6) we can show

$$\widehat{f}(k) = (ik)^{-p}\,\frac{1}{2\pi}\int_0^{2\pi} f^{(p)}(x)e^{-ikx}dx \tag{4.9}$$

and hence we can bound $\widehat{f}$ *by*

$$|\widehat{f}(k)| \leq Ck^{-p}, \quad C = \frac{1}{2\pi}\int_0^{2\pi} |f^{(p)}(x)|dx.$$

For $|k| \leq \frac{N}{2}$ *and* $j \neq 0$ *there holds* $|k + jN| \geq \left(|j| - \frac{1}{2}\right) N$ *and we get from Theorem (4.9)*

$$|\widehat{f}_N(k) - \widehat{f}(k)| \leq \sum_{\substack{j\in\mathbb{Z}\\ j\neq 0}} |\widehat{f}(k + jN)| \leq \sum_{\substack{j\in\mathbb{Z}\\ j\neq 0}} C|k + jN|^{-p} \leq 2\sum_{j\geq 1} C(j - \frac{1}{2})^{-p}N^{-p} = \tilde{C}N^{-p},$$

because the series converges for $p > 1$. The additional claim follows from applying equations (4.6) and (4.7) to $k = 0$, then it is just the first statement with this k. □

4.1.2. The Trigonometric Interpolation Polynomial

At the end of this section we will present a method to calculate approximately the derivatives of a function f on a grid using the discrete Fourier transform. Therefore we want to introduce the trigonometric interpolation polynomial to given points (x_j, y_j) on a periodic grid, i.e. a linear combination of functions e^{ikx} that has the value y_j at position $x_j = j\frac{2\pi}{N}$, $j = 1, \dots, N$. This discretization is used to obtain the following results.

Theorem 4.11 (cf. [11], Satz 5.8)**.** *Given pairs (x_j, y_j), $j = 1, \dots, N$ with N even,*

$$t_N(x) = \frac{1}{N}\sum_{k=-\frac{N}{2}}^{\frac{N}{2}}{}' \widehat{y}_k e^{ikx} := \frac{1}{2N}\left(\widehat{y}_{-\frac{N}{2}} e^{-ix\frac{N}{2}} + \widehat{y}_{\frac{N}{2}} e^{ix\frac{N}{2}}\right) + \frac{1}{N}\sum_{|k|<\frac{N}{2}} \widehat{y}_k e^{ikx}, \qquad x \in [0, 2\pi] \tag{4.10}$$

represents the trigonometric interpolant with $t_N(x_j) = y_j$, $j = 1, \dots, N$.

Proof. Let j be arbitrary but fixed. Because of periodicity we have

$$\widehat{y}_{-\frac{N}{2}} e^{-ix_j\frac{N}{2}} = \widehat{y}_{\frac{N}{2}} e^{ix_j\frac{N}{2}}$$

and hence

$$t_N(x_j) = \frac{1}{N}\sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} \widehat{y}_k e^{ikx_j} = \frac{1}{N}\sum_{k=0}^{N-1} \widehat{y}_k e^{ikx_j} = \frac{1}{N}\sum_{k=1}^{N} \widehat{y}_k e^{ikx_j} = \left(\mathcal{F}_N^{-1}\widehat{y}\right)_j = \left(\mathcal{F}_N^{-1}\mathcal{F}_N y\right)_j = y_j,$$

where $\widehat{y}_N = \widehat{y}_0$. □

t_N is a trigonometric polynomial of degree at most $\frac{N}{2}$. This means that it can be written as a linear combination of the functions 1, $\sin(x)$, $\cos(x)$, $\sin(2x)$, $\dots$, $\sin(\frac{N}{2}x)$, $\cos(\frac{N}{2}x)$, where acually the $\sin(\frac{N}{2}x)$ term is not needed.

Remark 4.12. *Later we make use of this trigonometric interpolant in points $(x_j, y_j(t) = f(x_j, t))$ to calculate the approximate spatial derivatives of a function $f(x, t)$.*

For simplicity we omit the t-dependence of y_j and $\widehat{y}_k$ in the following. The reason why we defined $\sum'$ and t_N in the way we did can be found by having a look at the inverse discrete Fourier transform for some fixed $j = 1, \dots, N$

$$y_j = \frac{1}{N}\sum_{k=0}^{N-1} e^{ikx_j}\widehat{y}_k = \frac{1}{N}\sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} e^{ikx_j}\widehat{y}_k = \frac{1}{N}\left(\widehat{y}_0 + \sum_{k=1}^{\frac{N}{2}-1}\left[e^{ikx_j}\widehat{y}_k + e^{-ikx_j}\widehat{y}_{-k}\right] + \widehat{y}_{\frac{N}{2}} e^{ix_j\frac{N}{2}}\right), \tag{4.11}$$

so the highest wave number is treated asymmetrically. Replace x_j in (4.11) by x gives a term $e^{i\frac{N}{2}x}$ with derivative $(i\frac{N}{2})e^{i\frac{N}{2}x}$. But since $e^{i\frac{N}{2}x}$ represents a real sawtooth wave on the grid, see figure 4.2, its derivative should be zero at the grid points. This problem can be fixed by defininig $\widehat{y}_{-\frac{N}{2}} = \widehat{y}_{\frac{N}{2}}$ and setting t_N as it stands in (4.10).
For more details see [25], chapter 3.

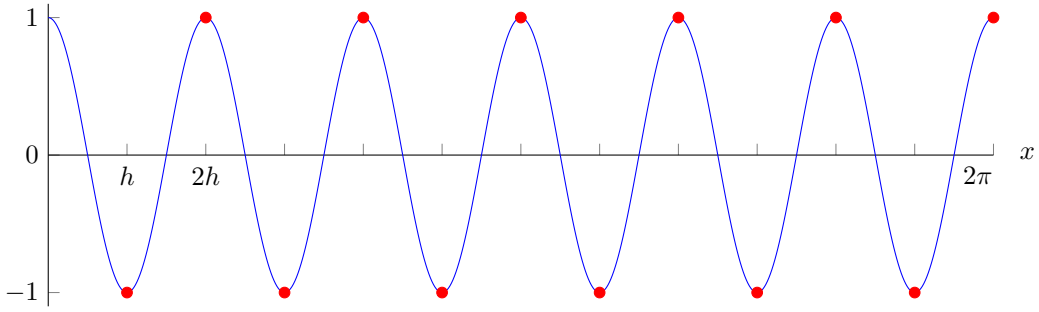


Figure 4.2. (sawtooth.m)
The sawtooth grid function and its interpolant $t_N(x) = \cos(\pi x/h) = \frac{1}{2}\left(e^{i\pi x/h} + e^{-i\pi x/h}\right)$. The proper spectral derivative is zero at every grid point, which would not be true if we used the interpolant $e^{i\pi x/h}$.

Theorem 4.13 (error of the trigonometric interpolation, cf. [11], Satz 5.9)**.** *Let $\sum_{n\in\mathbb{Z}} \widehat{f}(n)$ be absolutely convergent, then*

$$|t_N(x) - f(x)| \leq \sum_{|n|\geq\frac{N}{2}}{}' |\widehat{f}(n)| \qquad \forall x.$$

Proof. *Theorem 4.4 yields that $f(x) = \sum_{k=-\infty}^{\infty} \widehat{f}(k)e^{ikx}$. Define $y = [y_1, \ldots, y_N]$ by $y_j = f(x_j)$ and $\widehat{y}_k = \sum_{j=0}^{N-1} f(x_j)\omega_N^{kj} = N\widehat{f}_N(k)$. Together with Theorem 4.9 we then have*

$$\begin{aligned}
|t_N(x) - f(x)| &= |\sum_{n=-\frac{N}{2}}^{\frac{N}{2}}{}' \widehat{f}_N(n)e^{inx} - \left(\sum_{|n|<\frac{N}{2}} \widehat{f}(n)e^{inx} + \sum_{|n|>\frac{N}{2}} \widehat{f}(n)e^{inx} + \frac{1}{2}\left[2\widehat{f}(-\frac{N}{2})e^{-i\frac{N}{2}x} + 2\widehat{f}(\frac{N}{2})e^{i\frac{N}{2}x}\right]\right)| \\
&= |\sum_{n=-\frac{N}{2}}^{\frac{N}{2}}{}' \left(\widehat{f}_N(n) - \widehat{f}(n)\right)e^{inx} - \sum_{n\geq\frac{N}{2}}{}' \widehat{f}(n)e^{inx}| \\
&\leq |\sum_{n=-\frac{N}{2}}^{\frac{N}{2}}{}' \sum_{\substack{j\in\mathbb{Z}\\ j\neq 0}} \widehat{f}(n+jN)e^{inx}| + |\sum_{n\geq\frac{N}{2}}{}' \widehat{f}(n)e^{inx}| \\
&\leq 2\sum_{n\geq\frac{N}{2}}{}' |\widehat{f}(n)|.
\end{aligned}$$

□

As we have seen in the last theorem, the sum of all the higher truncated frequencies of the function f is a bound to the error of the trigonometric interpolation. Hence, we can control the error by choosing N bigger and thus taking more Fourier modes to evaluate the interpolant.

Now let us define some notation and then have a look at the derivatives of t_N.

Definition 4.14. *Let $a, b \in \mathbb{C}^N$, $m \in \mathbb{N}$. Then we define the pointwise multiplication $a \bullet b = (c_k)_{k=1}^N$ by*

$$c_k = (a \bullet b)_k := a_k b_k, \qquad k = 1, \ldots, N$$

and $a^m := \underbrace{a \bullet a \bullet \cdots \bullet a}_{m\text{-times}}$. Likewise we define $a^{-1} := \left(a_k^{-1}\right)_k$ and $e^a := \left(e^{a_k}\right)_k$.

Lemma 4.15 (derivatives of the interpolant in the grid points)**.** *For given (x_j, y_j), $j = 1, \dots, N$, the pth derivative of t_N in the grid points is given as follows*

$$t_N^{(p)}(x_j) = \left\{ \mathcal{F}_N^{-1} \left((i\tilde{k})^p \bullet \widehat{y} \right) \right\}_j,$$

where $\widehat{y} = (\widehat{y}_m)_m$, $m = -\frac{N}{2} + 1, \dots, \frac{N}{2}$ and

$$\tilde{k} \coloneqq \begin{cases} [-\frac{N}{2} + 1, \dots, \frac{N}{2} - 1, \frac{N}{2}], & p \text{ even} \\ [-\frac{N}{2} + 1, \dots, \frac{N}{2} - 1, \ 0\ ], & p \text{ odd} \end{cases}$$

Proof. Because of periodicity we have $\widehat{y}_{-\frac{N}{2}} = \widehat{y}_{\frac{N}{2}}$ and thus

$$t_N(x) = \frac{1}{N} \widehat{y}_{\frac{N}{2}} \cos\left(\frac{N}{2} x\right) + \frac{1}{N} \sum_{|k| < \frac{N}{2}} \widehat{y}_k e^{ikx}.$$

Differentiating gives

$$t_N'(x) = -\frac{1}{N} \frac{N}{2} \widehat{y}_{\frac{N}{2}} \sin\left(\frac{N}{2} x\right) + \frac{1}{N} \sum_{|k| < \frac{N}{2}} ik \widehat{y}_k e^{ikx}$$

$$t_N''(x) = -\frac{1}{N} \left(\frac{N}{2}\right)^2 \widehat{y}_{\frac{N}{2}} \cos\left(\frac{N}{2} x\right) + \frac{1}{N} \sum_{|k| < \frac{N}{2}} (ik)^2 \widehat{y}_k e^{ikx}$$

and using an inductive argument yields

$$t_N^{(p)}(x) = \frac{1}{N} \sum_{|k| < \frac{N}{2}} (ik)^p \widehat{y}_k e^{ikx} + \frac{1}{N} \left(\frac{N}{2}\right)^p \widehat{y}_{\frac{N}{2}} \cdot \begin{cases} (-1)^{\frac{p+1}{2}} \sin\left(\frac{N}{2} x\right), & p \text{ odd} \\ (-1)^{\frac{p}{2}} \cos\left(\frac{N}{2} x\right), & p \text{ even} \end{cases}$$

Because $x_j = j\frac{2\pi}{N}$, $j = 1, \dots, N$ there hold

$$\sin\left(\frac{N}{2} x_j\right) = \sin(j\pi) = 0 \ \forall j$$

and

$$\cos\left(\frac{N}{2} x_j\right) = \cos(j\pi) = (-1)^j = e^{ij\pi} = e^{i\frac{N}{2} x_j}$$

such that we have for $p = 2l + 1$, $l \in \mathbb{N}$

$$t_N^{(p)}(x_j) = \frac{1}{N} \sum_{|k| < \frac{N}{2}} (ik)^p \widehat{y}_k e^{ikx_j} + 0 \cdot \widehat{y}_{\frac{N}{2}} e^{i\frac{N}{2} x_j} = \left\{ \mathcal{F}_N^{-1} \left((i\tilde{k})^p \bullet \widehat{y} \right) \right\}_j$$

and for $p = 2l$, $l \in \mathbb{N}$

$$t_N^{(p)}(x_j) = \frac{1}{N} \sum_{|k| < \frac{N}{2}} (ik)^p \widehat{y}_k e^{ikx_j} + \frac{1}{N} \left(i\frac{N}{2}\right)^p \widehat{y}_{\frac{N}{2}} e^{i\frac{N}{2} x_j} = \left\{ \mathcal{F}_N^{-1} \left((i\tilde{k})^p \bullet \widehat{y} \right) \right\}_j.$$

□

Let us denote $y_j^{(\nu)} \coloneqq t_N^{(\nu)}(x_j)$, $j = 1, \dots, N$ for some $\nu \in \mathbb{N}$. The next theorem will show that $y_j^{(\nu)}$ is a good approximation to the ν-th derivative of f in the grid points.

Theorem 4.16 (cf. [25], Theorem 4 (a))**.** *Let $f \in C^p(\mathbb{R})$, 2π-periodic and let $\nu \leq p - 2$, $h = \frac{2\pi}{N}$. Then*

$$|y_j^{(\nu)} - f^{(\nu)}(x_j)| = \mathcal{O}\left(N^{-(p-\nu)}\right) = \mathcal{O}\left(h^{p-\nu}\right)$$

Proof. Applying integration by parts to the kth Fourier coefficient of $f^{(\nu)}$ as in the proof of corollary 4.10 and using (4.9) gives

$$\widehat{f^{(\nu)}}(k) = \frac{1}{2\pi}\int_0^{2\pi} f^{(\nu)}(x)e^{-ikx}dx = (ik)^\nu \frac{1}{2\pi}\int_0^{2\pi} f(x)e^{-ikx}dx = (ik)^\nu\, \widehat{f}(k)$$

such that we can go on as in the proof of theorem 4.13

$$\begin{aligned}
|y_j^{(\nu)} - f^{(\nu)}(x_j)| = & \; |{\sum_{|k|\le\frac{N}{2}}}' (ik)^\nu\, \widehat{f}_N(k)e^{ikx_j} - \sum_{k=-\infty}^{\infty} f^{(\nu)}(k)e^{ikx_j}| \\
= & \; |{\sum_{|k|\le\frac{N}{2}}}' (ik)^\nu\, \widehat{f}_N(k)e^{ikx_j} - \left({\sum_{|k|\le\frac{N}{2}}}' (ik)^\nu\, \widehat{f}(k)e^{ikx_j} + {\sum_{|k|\ge\frac{N}{2}}}' (ik)^\nu\, \widehat{f}(k)e^{ikx_j}\right)| \\
= & \; |{\sum_{|k|\le\frac{N}{2}}}' (ik)^\nu \left(\widehat{f}_N(k) - \widehat{f}(k)\right) e^{ikx_j} - {\sum_{|k|\ge\frac{N}{2}}}' (ik)^\nu\, \widehat{f}(k)e^{ikx_j}| \\
\overset{(4.8)}{=} & \; |{\sum_{|k|\le\frac{N}{2}}}' (ik)^\nu \sum_{\substack{j\in\mathbb{Z}\\ j\neq 0}} \widehat{f}(k+jN)e^{ikx_j} - {\sum_{|k|\ge\frac{N}{2}}}' (ik)^\nu\, \widehat{f}(k)e^{ikx_j}| \\
\le & \; {\sum_{|k|\le\frac{N}{2}}}' |k|^\nu \underbrace{\sum_{\substack{j\in\mathbb{Z}\\ j\neq 0}} |\widehat{f}(k+jN)|}_{\overset{\text{prf. thm. 4.10}}{\Longrightarrow} \le C_1 N^{-p}} + {\sum_{|k|\ge\frac{N}{2}}}' |k|^\nu \underbrace{|\widehat{f}(k)|}_{\overset{\text{prf. thm. 4.10}}{\Longrightarrow} \le C_2|k|^{-p}} \\
\le & \; 2C_1 \sum_{k=1}^{\frac{N}{2}} k^\nu N^{-p} + C_2\left(\frac{N}{2}\right)^{-(p-\nu)} + 2C_2 \underbrace{\sum_{k=\frac{N}{2}+1}^{\infty} k^{-(p-\nu)}}_{\le C_3} \qquad \le \tilde{C}N^{-(p-\nu)},
\end{aligned}$$

since $\sum_{k=\frac{N}{2}+1}^{\infty} k^{-(p-\nu)}$ converges to some constant less than $\frac{\pi^2}{6}$ for $p-\nu \ge 2$. □

Remark 4.17. *If we are working on arbitrary intervals $[a,b]$ the last theorem about the derivatives still holds, but we have to multiply $\tilde{k}$ by $\frac{2\pi}{b-a}$, see remark 4.2.*

All the calculations we did in this section to evaluate the discrete Fourier transform $\mathcal{F}_N y$ and its inverse, which for arbitrary N takes generally $\mathcal{O}(N^2)$ operations, can be done in a very more efficient way, namely the *fast Fourier tranform* (FFT) and the *inverse fast Fourier tranform* (IFFT) if we choose $N = 2^L$ as a power of 2. Then it takes only $\mathcal{O}(N\log_2 N)$ operations to get $\mathcal{F}_N y$. For more details on the FFT we refer to [11], chapter 5.

4.1.3. Differentiation by Spectral Methods: A Summary

Let $\tilde{v}\colon \mathbb{R}\to\mathbb{C}$ be 2π periodic and let $v_j = \tilde{v}(x_j)$, $j = 1,\dots N$ with $x_j = j\frac{2\pi}{N}$. Then the discrete Fourier transform $\mathcal{F}_N$ and its inverse $\mathcal{F}_N^{-1}$ are given by $\widehat{v} = \mathcal{F}_N v$ and respectively $v = \mathcal{F}_N^{-1}\widehat{v}$, where

$$\begin{aligned}
\widehat{v}_k = \quad & (\mathcal{F}_N v)_k = \sum_{j=1}^{N} e^{-ikx_j} v_j, \qquad k = -\frac{N}{2}+1,\dots,\frac{N}{2} \\
v_j = \quad & \left(\mathcal{F}_N^{-1}\widehat{v}\right)_j = \frac{1}{N}\sum_{k=-\frac{N}{2}+1}^{\frac{N}{2}} e^{ikx_j}\widehat{v}_k, \qquad j = 1,\dots,N.
\end{aligned}$$

By theorem 4.11 we know that the trigonometric interpolation polynomial t_N in the pairs (x_j, v_j) is given by equation (4.10),

$$t_N(x) = \frac{1}{2N}\left(\widehat{v}_{-\frac{N}{2}} e^{-ix\frac{N}{2}} + \widehat{v}_{\frac{N}{2}} e^{ix\frac{N}{2}}\right) + \frac{1}{N}\sum_{|k|<\frac{N}{2}} \widehat{v}_k e^{ikx}, \qquad x \in [0, 2\pi],$$

such that $t_N(x_j) = v_j$. From lemma 4.15 we know about the derivatives of the trigonometric interpolation polynomial

$$t_N^{(\nu)}(x_j) = \left\{\mathcal{F}_N^{-1}\left((i\tilde{k})^\nu \bullet \widehat{v}\right)\right\}_j, \qquad \tilde{k} \coloneqq \begin{cases} [-\frac{N}{2}+1, \ldots, \frac{N}{2}-1, \frac{N}{2}], & \nu \text{ even} \\ [-\frac{N}{2}+1, \ldots, \frac{N}{2}-1, \ 0\ ], & \nu \text{ odd} \end{cases},$$

which gives an approximation to the νth derivative of v in the grid points x_j

$$\frac{d^\nu}{dx^\nu} v(x_j) \approx v_j^{(\nu)} \coloneqq t_N^{(\nu)}(x_j),$$

as we have seen in theorem 4.16. If we assume v to be p times continuously differentiable the error of this spectral differentiation is of order $p-\nu$ as long as $\nu \leq p-2$, which we proved in theorem 4.16 such that

$$\frac{d^\nu}{dx^\nu} v(x_j) = v_j^{(\nu)} + \mathcal{O}(h^{p-\nu}), \quad j = 0, \ldots, N-1.$$

An example of how good the spectral differentiation method is for the non smooth "hat" function f and a smooth function $g(x) = e^{\sin(x)}$ is illustrated in the figures 4.3 and 4.4 respectively. We can see that for non smooth functions spectral methods fail but for smooth functions they yield a far better approximation than the finite difference method.

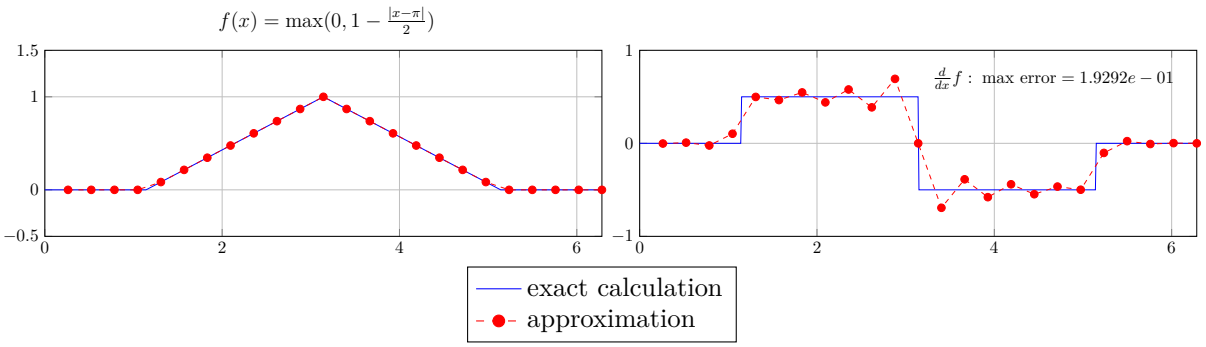


Figure 4.3. (spectraldiff .m)
Differentiation of the "hat function" f using spectral differentiation on the grid with step size $h = \frac{2\pi}{24}$.

If we again consider the wave equation (4.1) and v as a function of x and t we have to replace all evaluations v_j of this section by $v_j(t)$ and respectively $\widehat{v}_k$ by $\widehat{v}_k(t)$ to transform the wave equation into an ordinary equation using partial differentiation $\frac{\partial^\nu}{\partial x^\nu}$ instead of $\frac{d^\nu}{dx^\nu}$ such that

$$\frac{\partial^2}{\partial t^2} v(t) = \mathcal{F}_N^{-1}\left((ik)^2 \bullet \mathcal{F}_N v(t)\right), \qquad k = [-\frac{N}{2}+1, \ldots, \frac{N}{2}-1, \frac{N}{2}].$$

This can be solved again choosing a suitable time marching theme.

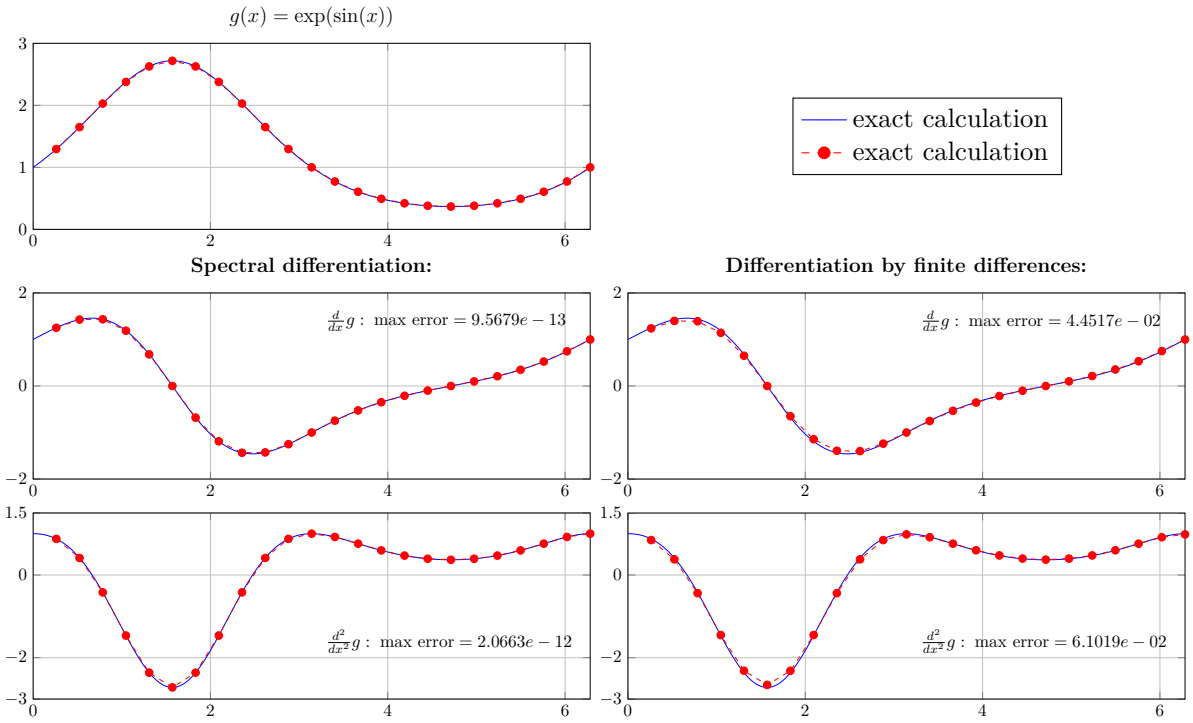


Figure 4.4. (spectraldiff .m)
Differentiation of the smooth 2π-periodic function g on the interval $[0, 2\pi]$ using spectral differentiation on the grid with step size $h = \frac{2\pi}{24}$ and comparison to differentiation by finite differences .

4.2. A Strang Splitting Method for the Gross-Pitaevskii Equation

Splitting methods are often used in numerical mathematics when we have to deal with a differential equation of the form

$$\dot{y} = (A + B)y, \quad y(0) = y_0,$$

where in case of ordinary differential equations A and B are matrices and in the case of partial differential equations A and B are (not necessarily linear) operators. They are very useful when the computational effort of solving the equation in its entirety is too high and the two separate equations

$$\dot{v} = Av, \quad v(0) = v_0, \qquad \dot{w} = Bw, \quad w(0) = w_0$$

can be solved by lower cost. Therefore we have *split* the equation into two parts. Often one can solve one of these problems even exactly although the original equation can only be solved approximately due to nonlinearities or similar hurdles.

After solving the separate equations we combine the results to obtain an approximation to the exact solution of the original problem. We will see that such a method works good for the nonlinear Schrödinger equation

$$i\dot{\psi} = \nu_1 \Delta\psi + \nu_2|\psi|^2\psi, \quad \psi(\cdot, t = 0) = \psi_0$$

with real coefficients ν_1 and ν_2, since we can exploit that, assuming periodic boundary conditions for practical implementation, the Laplacian can be diagonalized by using FFT and that the potential part can be solved exactly. But before we start analyzing these properties we want to introduce some theory on splitting methods that helps to get a better understanding of what splitting methods are about. The theory for section 4.2.1 is mostly taken from [10], chapter II.3 to II.5.

4.2.1. Theory on Splitting Methods

Recall the definition of the flow of a differential equation

Definition 4.18 (flow of a differential equation)**.** *The flow φ_t of a differential equation*

$$\dot{y} = f(y), \quad y(t_0) \quad \text{given}, \tag{4.12}$$

over time t is the mapping which associates to any y_0 the exact solution $y(t)$ with initial value $y(t_0) = y_0$. This map is thus defined by

$$\varphi_t(y_0) = y(t), \quad \text{if} \quad y(t_0) = y_0.$$

It satisfies $\varphi_{-t}^{-1} = \varphi_t$, a property that does not necessarily hold for a numerical time stepping method. So let us introduce the notion of the adjoint of a method.

Definition 4.19 (see [10], Definition II.3.1)**.** *The adjoint method Φ_h^* of a one-step method Φ_h is the inverse map of the original method with reversed time step $-h$, i.e.*

$$\Phi_h^* := \Phi_{-h}^{-1}.$$

In other words, $y_1 = \Phi_h^*(y_0)$ is implicitly defined by $\Phi_{-h}(y_1) = y_0$.

A method is called symmetric if $\Phi_h^* = \Phi_h$.

For instance the implicit Euler method is the adjoint method of the explicit Euler method and vice versa. Without proof there hold $(\Phi_h^*)^* = \Phi_h$ and

$$(\Phi_h \circ \Psi_h)^* = \Psi_h^* \circ \Phi_h^*. \tag{4.13}$$

Definition 4.20 (order of consistency)**.** *We say that a numerical method Φ_h to solve (4.12) is (consistent) of order p, if there holds*

$$\Phi_h(y(t)) = \varphi_h(y(t)) + \mathcal{O}(h^{p+1}),$$

for a smooth function $y(t)$.

Theorem 4.21 (cf. [10] Theorem II.3.2)**.** *Let φ_t be the exact flow of (4.12) and let Φ_h be a one-step method of order p satisfying*

$$\Phi_h(y_0) = \varphi_h(y_0) + C(y_0)h^{p+1} + \mathcal{O}(h^{p+2}).$$

The the adjoint method Φ_h^* has the same order p and we have

$$\Phi_h^*(y_0) = \varphi_h(y_0) + (-1)^p C(y_0)h^{p+1} + \mathcal{O}(h^{p+2}).$$

If the method is symmetric, its maximal order is even.

Proof. From a given initial value y_0 we compute $\varphi_h(y_0)$ and $y_1 = \Phi_h^*(y_0)$ with the local error $e^* = y_1 - \varphi_h(y_0)$ of Φ_h^*. We use the method Φ_{-h} to project this error back to become $e = \Phi_{-h}(y_1) - \Phi_{-h}(\varphi_h(y_0))$, where $-e$ is the local error of Φ_{-h}, i.e. by the hypothesis on $\Phi_h(y_0)$

$$e = (-1)^p C(\varphi_h(y_0))h^{p+1} + \mathcal{O}(h^{p+2}).$$

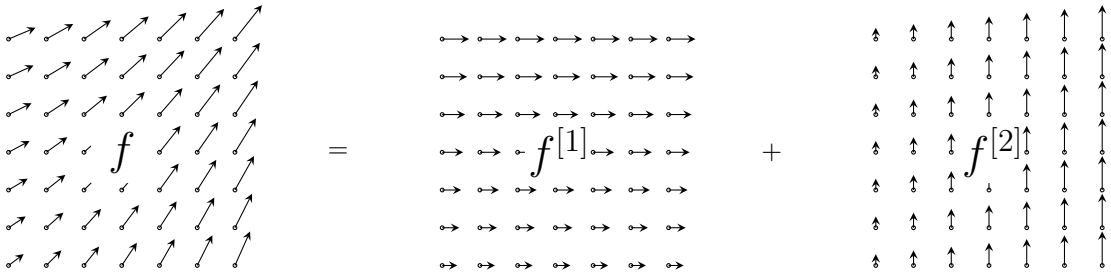


Figure 4.5. Splitting of a vector field.

Now if we use that $\Phi_{-h}(x) = x + (-h)f(x) + \mathcal{O}(h^2)$ *we can write*

$$\begin{aligned} e &= \Phi_{-h}(y_1) - \Phi_{-h}(\varphi_h(y_0)) = y_1 - \varphi_h(y_0) + (-h)\left(f(y_1) - f(\varphi_h(y_0))\right) + \mathcal{O}(h^2) \\ &= e^* + hCe^* + \mathcal{O}(h^2). \end{aligned}$$

Therefore together with $\varphi_h(y_0) = y_0 + \mathcal{O}(h)$ *we obtain*

$$e^* = (-1)^p C(y_0) h^{p+1} + \mathcal{O}(h^{p+2}).$$

For symmetric methods there holds $\Phi_h = \Phi_h^*$ *which implies* $C(y_0) = (-1)^p C(y_0)$. *Thus* $C(y_0) \neq 0$ *only for even* p. □

Now consider an arbitrary system

$$\dot{y} = f(y)$$

in $\mathbb{R}^n$, and suppose that, as illustrated in figure 4.5, the vector field is split as

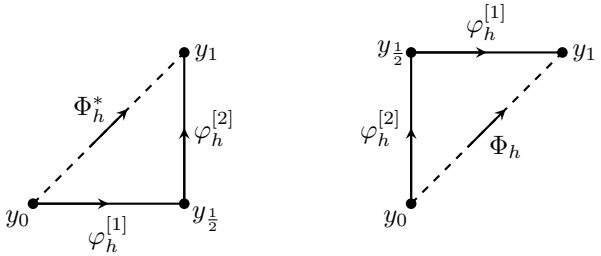
$$\dot{y} = f^{[1]}(y) + f^{[2]}(y). \tag{4.14}$$

Let us assume that the exact flows $\varphi_t^{[1]}$ and $\varphi_t^{[2]}$ of the systems

$$\dot{y} = f^{[1]}(y) \qquad \text{and} \tag{4.15a}$$

$$\dot{y} = f^{[2]}(y) \tag{4.15b}$$

can be calculated explicitely. Then we can first solve the first system to obtain a value $y_{\frac{1}{2}}$ and from this value integrate the second system to obtain y_1. This procedure is summarized in the numerical methods



$$\Phi_h^* = \varphi_h^{[2]} \circ \varphi_h^{[1]}$$
$$\Phi_h = \varphi_h^{[1]} \circ \varphi_h^{[2]},$$

that are known as the *Lie-Trotter splitting*.

Lemma 4.22. *The Lie-Trotter splitting method and its adjoint*

$$\Phi_h^* = \varphi_h^{[2]} \circ \varphi_h^{[1]}, \qquad \Phi_h = \varphi_h^{[1]} \circ \varphi_h^{[2]},$$

are of order p=1.

Proof. The Taylor expansion of $\varphi_h(y_0)$ reads

$$\varphi_h(y_0) = y_0 + hf(y_0) + \frac{h^2}{2} f'(y_0) + \mathcal{O}(h^3).$$

We have

$$\begin{aligned}
\left(\varphi_h^{[1]} \circ \varphi_h^{[2]}\right)(y_0) &= \varphi_h^{[1]}\left(y_0 + hf^{[2]}(y_0) + \frac{h^2}{2} f^{[2]\prime}(y_0) f^{[2]}(y_0) + \mathcal{O}(h^3)\right) \\
&= \left(y_0 + hf^{[2]}(y_0) + \frac{h^2}{2} f^{[2]\prime}(y_0) f^{[2]}(y_0)\right) + hf^{[1]}\left(y_0 + hf^{[2]}(y_0) + \mathcal{O}(h^2)\right) \\
&\quad + \frac{h^2}{2} f^{[1]\prime}(y_0) f^{[1]}(y_0) + \mathcal{O}(h^3) \\
&= y_0 + h\left(f^{[1]}(y_0) + f^{[2]}(y_0)\right) + \frac{h^2}{2}\left(f^{[1]\prime}(y_0) + f^{[2]\prime}(y_0)\right)\left(f^{[1]}(y_0) + f^{[2]}(y_0)\right) \\
&\quad + \frac{h^2}{2}\left(f^{[1]\prime}(y_0) f^{[2]}(y_0) - f^{[2]\prime}(y_0) f^{[1]}(y_0)\right) + \mathcal{O}(h^3)
\end{aligned}$$

and therefore

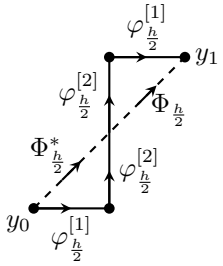
$$\left(\varphi_h^{[1]} \circ \varphi_h^{[2]}\right)(y_0) = \varphi_h(y_0) + \frac{h^2}{2}\left(f^{[1]\prime}(y_0) f^{[2]}(y_0) - f^{[2]\prime}(y_0) f^{[1]}(y_0)\right) + \mathcal{O}(h^3). \tag{4.16}$$

Thus the difference is $\varphi_h(y_0) - \left(\varphi_h^{[1]} \circ \varphi_h^{[2]}\right)(y_0) = \mathcal{O}(h^2)$ and we have that the *Lie-Trotter splitting* has order $p = 1$. □

Another splitting method with higher order is known as the *Strang splitting*, where the flows $\varphi^{[1]}$ and $\varphi^{[2]}$ are split symmetrically. This method is given by

$$\Phi_h^{[S]} = \varphi_{\frac{h}{2}}^{[1]} \circ \varphi_h^{[2]} \circ \varphi_{\frac{h}{2}}^{[1]} \tag{4.17}$$

and if we split the flow $\varphi_h^{[2]} = \varphi_{\frac{h}{2}}^{[2]} \circ \varphi_{\frac{h}{2}}^{[2]}$ we see that



$$\Phi_h^{[S]} = \Phi_{\frac{h}{2}} \circ \Phi_{\frac{h}{2}}^* \tag{4.18}$$

is the composition of the Lie-Trotter method and its adjoint with halved step sizes $\frac{h}{2}$. Hairer and Wanner showed in chapter II.4 of their book [10] that

Theorem 4.23 (cf. [10], Theorem II.4.1 and following pages)**.** *For a one-step method Φ_h of order p there holds that the composition*

$$\Psi_h = \Phi_{\alpha_s h} \circ \Phi_{\beta_s h}^* \circ \cdots \circ \Phi_{\beta_2 h}^* \circ \Phi_{\alpha_1 h} \circ \Phi_{\beta_1 h}^* \tag{4.19}$$

has order $p + 1$, if

$$\begin{aligned}
\sum_{j=1}^{s} \alpha_j + \beta_j &= 1 \\
\sum_{j=1}^{s} \alpha_j^{p+1} + (-1)^p \beta_j^{p+1} &= 0.
\end{aligned} \tag{4.20}$$

Proof. We proceed as in the proof of theorem 4.21. The local errors e_j and e_j^* of the methods $\Phi_{\alpha_j h}$ and $\Phi^*_{\beta_j h}$ respectively are given by

$$\begin{aligned} e_j &= \quad C(y_0)\cdot \alpha_j^{p+1} h^{p+1} + \mathcal{O}(h^{p+2}),\\ e_j^* &= (-1)^p \quad C(y_0)\cdot \beta_j^{p+1} h^{p+1} + \mathcal{O}(h^{p+2}). \end{aligned}$$

Now if $\sum \alpha_j + \beta_j = 1$ we have similar as before

$$\varphi_h(y_0) - \Psi_h(y_0) = C(y_0)\left(\alpha_1^{p+1} + (-1)^p\beta_1^{p+1} + \cdots + \alpha_s^{p+1} + (-1)^p\beta_s^{p+1}\right) h^{p+1} + \mathcal{O}(h^{p+2}),$$

whose $\mathcal{O}(h^{p+1})$ term vanishes if the conditions of the theorem are satisfied. □

Since the Lie-Trotter method has order $p = 1$ and $\alpha_1 = \beta_1 = \frac{1}{2}$, the Strang splitting method (4.18) therefore has order 2. The symmetry of the method is obvious from (4.13).

Combining Exact and Numerical Flows

All this theory on splitting so far is based on the assumption that $\dot y = f(y)$ can be split as in (4.14) such that both the flow $\varphi_t^{[1]}$ and $\varphi_t^{[2]}$ can be calculated explicitely. But it may happen that only the flow $\varphi_t^{[1]}$ can be computed exactly, such that we have to think about combining exact and numerical flows. If $f^{[1]}$ is the dominant part of the vector field it is desirable to construct integrators that retain this information.

Let $f^{[1]}$ such that the flow $\varphi_t^{[1]}$ of (4.15a) can be calculated exactly and let $\Phi_h^{[2]}$ be a numerical first-order one-step method to solve the system (4.15b).

We define

$$\Phi_h = \varphi_h^{[1]} \circ \Phi_h^{[2]}, \qquad \Phi_h^* = \Phi_h^{[2]*} \circ \varphi_h^{[1]}$$

as the basis for (4.19) and obtain

$$\Psi_h = \varphi^{[1]}_{\alpha_s h} \circ \Phi^{[2]}_{\alpha_s h} \circ \Phi^{[2]*}_{\beta_s h} \circ \varphi^{[1]}_{(\beta_s+\alpha_{s-1})h} \circ \Phi^{[2]}_{\alpha_{s-1}h} \circ \cdots \circ \Phi^{[2]*}_{\beta_1 h} \circ \varphi^{[1]}_{\beta_1 h}.$$

One can show that, using $\Phi_h^{[2]}(y_0) = y_0 + hf^{[2]}(y_0) + \mathcal{O}(h^2)$ and replacing $\varphi_h^{[2]}$ by $\Phi_h^{[2]}$ in (4.16), Φ_h is also a first-order method. Thus, if α_j and β_j, $j = 1,\dots,s$ satisfy (4.20) then the resulting method Ψ_h is again of second order by theorem 4.23.

4.2.2. Application to the nonlinear Schrödinger Equation

This section is mostly taken from [15], chapter II.1.3.

When it comes to partial differential equations, where the solution possesses high-frequency oscillations, such as for example the nonlinear Schrödinger equation

$$i\frac{\partial}{\partial t}\psi(x,t) = \nu_1\,\Delta\psi(x,t) + \nu_2|\psi(x,t)|^2\psi(x,t), \quad \psi(x,0) = \psi_0(x), \tag{4.21}$$

also called *Gross-Pitaevskii-equation* with coefficients ν_1 and ν_2 given in (3.17), time marching schemes based on Taylor expansion are not a very good choice any more because higher order time derivatives will now have large norms. So we have to rather revert to other methods. One approach is applying a splitting method to the problem. The main reason is that the equations

$$i\dot\theta = \nu_1\,\Delta\theta \qquad \text{and} \qquad i\dot\phi = \nu_2|\phi|^2\phi$$

can be solved more easily and a lot more efficient than the full Schrödinger equation. We will solve the linear equation, the so called *kinetic* part, by a Fourier technique and as we will see the nonlinear equation, the so called *potential part* can fortunately be solved exactly.

Some notational remarks for this section: In the following we suppress the spatial variable in the differential equation and set

$$u(t) = \psi(\cdot, t)$$

because for our purpose only the time dependence is important. Let $\psi(\cdot, t) \in H_0^2(\mathbb{R}) \coloneqq \left\{v \in H^2(\mathbb{R}) \mid v(x) \to 0,\ |x| \to \infty\right\}$, where the index 0 indicates that all elements in this function space decay to zero as $x \to \infty$. Let the Lebesgue space L^2 be endowed with the scalar product

$$\langle f, g\rangle_{L^2} = \int_{\Omega^d} f(x)\overline{g(x)}dx, \quad f, g \in L^2$$

and corresponding norm.

Then the Schrödinger equation (4.21) turns into the initial value problem

$$\frac{\partial}{\partial t}u(t) = Au(t) + B\left(u(t)\right)u(t), \quad u(t_0) = u_0 \quad \text{given},$$

where

$$A \coloneqq -i\,\nu_1\,\Delta$$

denotes a linear differentiation operator and

$$B\left(u(t)\right) \coloneqq -i\,\nu_2|u(t)|^2$$

is a nonlinear operator. The idea of a splitting method now is to decompose the right hand side of the differential equation such that we obtain two separate differential equations

$$\frac{\partial}{\partial t}v(t) = Av(t), \qquad v(t_0) = v^0 \quad \text{given} \tag{4.22a}$$

$$\frac{\partial}{\partial t}w(t) = B\left(w(t)\right)w(t), \quad w(t_0) = w^0 \quad \text{given} \tag{4.22b}$$

4.2.3. Fourier Pseudo-Spectral Method for the Kinetic Part

Similar to the paragraph on spatial derivatives we will derive a method for the linear part (4.22a) of the Schrödinger equation (4.21), the so called *Fourier pseudospectral method*. For practical implementation issues we restrict our spatial interval to $x \in [-L, L]$ and assume periodic boundary conditions of our solution on this interval, i.e.

$$\psi(-L, \cdot) = \psi(L, \cdot), \qquad \partial_x\psi(-L, \cdot) = \partial_x\psi(L, \cdot),$$

where we choose L that large that we can neglect the effect of these boundary conditions and still can apply the results of the case $x \in \mathbb{R}$. Since we can transform any real interval $[a, b]$ onto the interval $I \coloneqq [-\pi, \pi]$ using a simple rescaling and shift $x \mapsto \alpha x + \beta$ it's enough to work on the interval I. The linear part of the NLS equation with periodic boundary conditions reads

$$\begin{aligned} i\frac{\partial}{\partial t}\psi(x,t) &= \nu_1\frac{\partial^2}{\partial x^2}\psi(x,t), \qquad x \in [-\pi, \pi], \\ \psi(\cdot, 0) &= \psi^0 \\ \psi(-\pi, \cdot) &= \psi(\pi, \cdot). \end{aligned} \tag{4.23}$$

Now we want to pick up the idea, which C. Lubich presented in [15], chapter III.1.3. We approximate $\psi(x, t)$ by a trigonometric interpolation polynomial at every time t on a periodic grid $x_j = -\pi + jh,\ j =$

$1, \dots N$ with $h = \frac{2\pi}{N}$, such that

$$\psi(x,t) \approx \psi_N(x,t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} c_k(t) e^{ikx}, \qquad x \in [-\pi, \pi],$$

and $\psi(x_j,t) = \psi_N(x_j,t)$, where $N \in 2\mathbb{N}$.

The Fourier coefficients $c_k(t)$ are determined via a so called collocation method which requires that $\psi_N(x,t)$ satisfies (4.23) in each grid point x_j, $j = 1, \dots N$, i.e.

$$i\frac{\partial}{\partial t}\psi_N(x_j,t) = \nu_1 \frac{\partial^2}{\partial x^2}\psi_N(x_j,t), \quad \psi_N(x_j,0) = \psi_N^0(x_j), \qquad j = -\frac{N}{2}, \dots, \frac{N}{2} - 1, \tag{4.24}$$

where

$$\psi_N^0(x) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} c_k^0 e^{ikx}.$$

Inserting the interpolation polynomial gives

$$i \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \dot{c}_k(t) e^{ikx_j} = \nu_1 \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} -k^2 c_k(t) e^{ikx_j}$$

and each c_k has to satisfy the ordinary differential equation

$$\dot{c}_k(t) = i\nu_1 k^2 c_k(t). \tag{4.25}$$

Hence $c_k(t) = e^{i\nu_1 k^2 t} c_k(0)$. Here $c_k(0) \coloneqq \left(\mathcal{F}_N \left(\psi_N(x_j,0)\right)_{j=1}^N\right)_k$ are the Fourier coefficients of the interpolation polynomial $\psi_N(x,0)$ corresponding to the initial value $\psi^0(x)$. Furthermore we observe that

$$\left(\psi_N(x_j,t)\right)_{j=1}^N = \mathcal{F}_N^{-1}\left(c_k(t)\right)_{k=-\frac{N}{2}}^{\frac{N}{2}-1}.$$

Defining $v_j(t) \coloneqq \psi_N(x_j,t)$, $v(t) \coloneqq \left(v_j(t)\right)_{j=1}^N$ and $K_N \coloneqq \left[-\frac{N}{2}, \dots, \frac{N}{2} - 1\right]$ we have

$$v(t) = \mathcal{F}_N^{-1}\left(e^{i\nu_1 K_N^2 t} \bullet \mathcal{F}_N v(0)\right).$$

Since we want to use numerical methods to solve (4.23) we also have to discretize the time interval $[t_0, T]$ for some $T > t_0$ and set $t_m = t_0 + m\delta_t$, $m = 0, 1, 2, \dots$ with time step size δ_t. Now defining $v_j^m \coloneqq v_j(t_m)$ and $v^m \coloneqq \left(v_j(t_m)\right)_{j=1}^N$ enables us to write down a fully discrete approach of solving (4.22a), which reads

$$v^m = \mathcal{F}_N^{-1}\left(e^{i\,\nu_1 K_N^2 t_m} \bullet \mathcal{F}_N v^0\right), \quad \mu = (\mu_k)_{k \in K_N}.$$

Assuming v^m already to be computed, we observe that

$$e^{i\,\nu_1 K_N^2 t_m} \bullet \mathcal{F}_N v^0 = \mathcal{F}_N \mathcal{F}_N^{-1}\left(e^{i\,\nu_1 K_N^2 t_m} \bullet \mathcal{F}_N v^0\right) = \mathcal{F}_N v^m$$

and thus obtain

$$v_j^{m+1} = \mathcal{F}_N^{-1}\left(e^{i\,\nu_1 K_N^2 \delta_t} \bullet \mathcal{F}_N v^m\right). \tag{4.26}$$

Before we go on with the nonlinear part let us have a look at the error that we make by using the scheme (4.26) to discretize space. In the following we denote the interpolation polynomial to some function f by $\mathcal{I}_N f$, where

$$\mathcal{I}_N f(x) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} c_k e^{ikx} \quad \text{with} \quad \left(c_k\right)_{k=-\frac{N}{2}}^{\frac{N}{2}-1} = \mathcal{F}_N\left(f(x_j)\right)_{k=-\frac{N}{2}}^{\frac{N}{2}-1}.$$

Theorem 4.24 (interpolation error, cf. [15], theorem III.1.7)**.** *Let f be 2π-periodic and $f \in H^s$ for some $s \geq 1$. Then the interpolation error is bounded in L^2 by*

$$\|f - \mathcal{I}_N f\|_{L^2} \leq CN^{-s} \|\frac{\partial^s}{\partial x^s} f\|_{L^2}.$$

Furthermore

$$\left\| \frac{\partial^m}{\partial x^m}(f - \mathcal{I}_N f) \right\|_{L^2} \leq CN^{-s} \left\| \frac{\partial^{s+m}}{\partial x^{s+m}} f \right\|_{L^2}. \tag{4.27}$$

The constant C depends only on s.

Proof. *The proof is very similar to the proof of theorem 4.13 and theorem 4.16:*

We write down f and $\mathcal{I}_N f$ as

$$f(x) = \sum_{k=-\infty}^{\infty} a_k e^{ikx}, \qquad \mathcal{I}_N f(x) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} c_k e^{ikx}.$$

The interpolation property then gives the aliasing formula from theorem 4.9

$$c_k = \sum_{l=-\infty}^{\infty} a_{k+lN}.$$

Parseval's formula (theorem 4.5) and the Cauchy-Schwarz inequality yield similar to the proof of theorem 4.16

$$\begin{aligned}
\|f - \mathcal{I}_N f\|_{L^2}^2 &= \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \left(\left| \sum_{l\neq 0} a_{k+lN} \right|^2 + \sum_{l\neq 0} |a_{k+lN}|^2 \right) \\
&\leq \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \left(\sum_{l\neq 0} (k+lN)^{-2s} \cdot \sum_{l\neq 0} (k+lN)^{2s} |a_{k+lN}|^2 \right. \\
&\qquad \left. + \sum_{l\neq 0} (k+lN)^{-2s} \cdot (k+lN)^{2s} |a_{k+lN}|^2 \right) \\
&\leq C^2 N^{-2s} \sum_{k=-\infty}^{\infty} |k^s a_k|^2 = C^2 N^{-2s} \left\| \frac{\partial^s}{\partial x^s} f \right\|_{L^2}^2.
\end{aligned}$$

The second part of the theorem can be proved analogously. □

Theorem 4.25 (collocation error, cf. [15], theorem III.1.8)**.** *Let the exact solution $\psi(t) = \psi(\cdot, t)$ of (4.23) satisfy $\psi(t) \in H^{s+2}$ for every $t \geq 0$, some $s \geq 1$. Then the error of the Fourier pseudospectral method using the collocation (4.24) with initial value $\psi_N(x,0) = \mathcal{I}_N \psi(x,0)$ is bounded in L^2 by*

$$\|\psi_N(t) - \psi(t)\|_{L^2} \leq CN^{-s}(1+t) \max_{0\leq r\leq t} \left\| \frac{\partial^{s+2}}{\partial x^{s+2}} \psi(r) \right\|_{L^2}, \quad C = C(s).$$

Proof. *We begin by reformulating (4.24) as an equation with continuous argument on both sides, i.e. by interpolating both sides such that*

$$i\frac{\partial}{\partial t}\psi_N(x,t) = \nu_1 \frac{\partial^2}{\partial x^2}\psi_N(x,t), \qquad x \in [-\pi, \pi].$$

On the other hand since $\mathcal{I}_N \frac{\partial}{\partial t}\psi = \frac{\partial}{\partial t}\mathcal{I}_N \psi$ we have that the interpolant to the solution satisfies

$$i\frac{\partial}{\partial t}\mathcal{I}_N\psi(x,t) = \nu_1 \frac{\partial^2}{\partial x^2}\mathcal{I}_N\psi(x,t) + \delta_N(x,t),$$

where the defect $\delta_N(x,t)$ *is given by*

$$\delta_N \coloneqq \nu_1 \left(\mathcal{I}_N \frac{\partial^2}{\partial x^2}\psi - \frac{\partial^2}{\partial x^2}\mathcal{I}_N\psi \right).$$

Therefore the error $\varepsilon_N = \psi_N - \mathcal{I}_N\psi$ *satisfies*

$$i\frac{\partial}{\partial t}\varepsilon_N = \nu_1 \frac{\partial^2}{\partial x^2}\varepsilon_N - \delta_N.$$

In terms of the Fourier coefficients $e = (e_k)$ *and* $d = (d_k)$ *given by*

$$\varepsilon_N(x,t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} e_k(t)e^{ikx}, \quad \delta_N(x,t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} d_k(t)e^{ikx},$$

this gives equations for the e_k *similar to (4.25),*

$$\dot{e}_k = i\nu_1 k^2 e_k - d_k.$$

Using the variation of constants formula, we obtain

$$e_k(t) = e^{i\nu_1 k^2 t}e_k(0) + \int_0^t e^{i\nu_1 k^2 (t-r)}d_k(r)dr$$

and thus an estimate in the euclidean norm

$$\|e(t)\| \leq \|e(0)\| + \int_0^t \|d(r)\|dr.$$

Parseval's formula yields that this is equivalent to

$$\|\varepsilon_N(t)\|_{L^2} \leq \|\varepsilon_N(0)\|_{L^2} + \int_0^t \|\delta_N(r)\|_{L^2}dr.$$

Using $f(x,r) \coloneqq \frac{\partial^2}{\partial x^2}\psi(\,\cdot\,,r)$ *we can rewrite*

$$\begin{aligned}\delta_N &= \nu_1 \left(\mathcal{I}_N f - \frac{\partial^2}{\partial x^2}\mathcal{I}_N\psi + \frac{\partial^2}{\partial x^2}\psi - f \right) \\ &= \nu_1 \left(\mathcal{I}_N f - f + \frac{\partial^2}{\partial x^2}\left(\psi - \mathcal{I}_N\psi\right) \right),\end{aligned}$$

such that we can apply theorem 4.24 and equation (4.27) to f *and* ψ *respectively with* $m = 2$*. Then*

$$\|\delta_N(r)\|_{L^2} \leq CN^{-s} \left\| \frac{\partial^{s+2}}{\partial x^{s+2}}\psi(r) \right\|_{L^2}.$$

Since $\varepsilon_N = \psi_N - \mathcal{I}_N\psi$*, there holds* $\varepsilon_N(\,\cdot\,,0) = 0$ *and*

$$\psi_N - \psi = \varepsilon_N + \mathcal{I}_N\psi - \psi.$$

Furthermore again theorem 4.24 gives

$$\|\mathcal{I}_N\psi - \psi\|_{L^2} \leq CN^{-s} \left\| \frac{\partial^{s+2}}{\partial x^{s+2}}\psi \right\|_{L^2}.$$

Having

$$\|\psi_N(t) - \mathcal{I}_N\psi(t) + \mathcal{I}_N\psi(t) - \psi(t)\|_{L^2} \leq \|\varepsilon_N(t)\|_{L^2} + \|\mathcal{I}_N\psi(t) - \psi(t)\|_{L^2},$$

we obtain altogether

$$\|\psi_N(t) - \psi(t)\|_{L^2} \leq CN^{-s}(1+t) \max_{0\leq r\leq t} \left\| \frac{\partial^{s+2}}{\partial x^{s+2}}\psi(r) \right\|_{L^2}.$$

□

4.2.4. Solution of the nonlinear Part

Now let us continue solving equation (4.22b). From

$$\frac{\partial}{\partial t} w(t) = -i\nu_2 |w(t)|^2 w(t), \quad w(t_0) = w^0$$

and respectively

$$\frac{\partial}{\partial t} \overline{w}(t) = i\nu_2 |w(t)|^2 \overline{w}(t), \quad \overline{w}(t_0) = \overline{w^0}$$

we can conclude that

$$\frac{\partial}{\partial t} |w(t)|^2 = \frac{\partial}{\partial t} \left(w(t) \cdot \overline{w}(t) \right) = \left(-i\nu_2 |w(t)|^2 w(t) \right) \overline{w}(t) + w(t) \left(i\nu_2 |w(t)|^2 \overline{w}(t) \right) = 0.$$

This means $|w(t)|^2 = |w(t_0)|^2 \; \forall t > 0$ such that we can solve this problem exactly as

$$w(t) = e^{-i\nu_2 |w(t_0)|^2 (t - t_0)} w^0. \tag{4.28}$$

To derive an iteration for this equation we evaluate w at the grid points x_j, $j = 1, \ldots, N$ at time $t_m = t_0 + m\delta_t$, $m \in \mathbb{N}_0$ and set $w_j^m = w(x_j, t_m)$. Assuming that we already computed w_j^m then we obtain w_j^{m+1} by

$$w_j^{m+1} = e^{-i\nu_2 |w(t_0)|^2 \delta_t} w_j^m.$$

4.2.5. Composition of the Partial Solutions

The idea of getting a solution for the nonlinear Schrödinger equation (4.21) now is to compose the partial solutions in the manner of Strang splitting that we already introduced in section 4.2.1. Without loss of generality we set $t_0 = 0$ and in the notation of that section define $f^{[1]}(u) \coloneqq Au$ and $f^{[2]}(u) \coloneqq B(u)u$. The corresponding flows for the systems

$$\dot{v} = f^{[1]}(v), \quad v(0) = v_0 \qquad \text{and} \qquad \dot{w} = f^{[2]}(w), \quad w(0) = w_0$$

are given by φ_t^A and φ_t^B respectively, where

$$A \coloneqq -i\,\nu_1\,\Delta \quad \text{and} \quad B\left(u(t)\right) \coloneqq -i\,\nu_2 |u(t)|^2.$$

This gives

$$\frac{\partial}{\partial t} u = f^{[1]}(u) + f^{[2]}(u), \quad u(0) = u_0,$$

such that we obtain a second order approximation to the solution by applying the Strang splitting method (4.17) to the problem for some time step δ_t and some approximation $u^n \approx u(t_n)$:

$$u^{n+1} = \left(\varphi^A_{\frac{1}{2}\delta_t} \circ \varphi^B_{\delta_t} \circ \varphi^A_{\frac{1}{2}\delta} \right) (u^n)$$

Because for some given u^0 the flow φ_t^B is given by

$$\varphi_t^B(u^0) = u(t) \overset{(4.28)}{=} e^{-i\nu_2 |u^0|^2\, t} u^0,$$

it can be calculated explicitely and exactly, thus we make numerical errors only when we approximate φ_t^A by the Fourier pseudo spectral method that we explained in section 4.2.3. We combine the numerical flow $\Phi_h^A \approx \varphi_h^A$ with the exact flow φ_h^B and obtain

$$u^{n+1} = \left(\Phi^A_{\frac{1}{2}\delta_t} \circ \varphi^B_{\delta_t} \circ \Phi^A_{\frac{1}{2}\delta} \right) (u^n),$$

see algorithm 4.1.

Algorithm 4.1: Split step method for the NLS equation

given $u^0 = u(t_0)$;

$M = \lfloor \frac{T-t_0}{\delta_t} \rfloor$;

for $m = 0 : M-1$ **do**

apply Fourier spectral method Φ_h^A with step size $h = \frac{\delta_t}{2}$ to u^m:

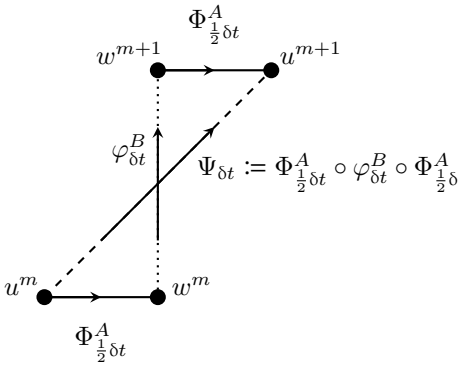
$$w^m = \Phi^A_{\frac{1}{2}\delta_t}(u^m)$$

calculate w^{m+1} using $\varphi^B_{\delta_t}$:

$$w^{m+1} = e^{-i\nu_2|w^m|^2\,\delta_t} w^m$$

obtain u^{m+1} by applying the Fourier spectral method with step size $\frac{\delta_t}{2}$ to w^{m+1}:

$$u^{m+1} = \Phi^A_{\frac{1}{2}\delta_t}(w^{m+1})$$



end

$u^M \approx u(t_M)$;

4.2.6. Error Bounds for the Strang Splitting Method

This section is mostly taken from [16], where C. Lubich presented some error estimates on the Strang Splitting method for the nonlinear Schrödinger equation. We adapted some parts of the proofs for this section.

The split step integration scheme for the nonlinear Schrödinger equation is given by

$$\begin{aligned} \psi^-_{n+1/2} &= e^{-i\frac{\nu_1}{2}\tau\,\Delta}\psi_n, \\ \psi^+_{n+1/2} &= e^{-i\nu_2\tau|\psi^-_{n+1/2}|^2}\psi^-_{n+1/2}, \\ \psi_{n+1} &= e^{-i\frac{\nu_1}{2}\tau\,\Delta}\psi^+_{n+1/2}, \end{aligned} \tag{4.29}$$

as derived above. To obtain results on the error bounds of this scheme we proceed as C. Lubich in [16] and adapt them to the 1D case $x \in \mathbb{R}$. At this point we want to have a look at some auxiliary results that will enable us to establish an error bound on the local and global error of the Strang splitting method applied to the cubic nonlinear Schrödinger equation.

As we have already shown in lemma 3.3, the solution of the NLS equation conserves the L^2 norm for all times, such that we can assume

$$\|\psi(\,\cdot\,,t)\|_{L^2} = \|\psi(\,\cdot\,,0)\|_{L^2} = 1. \tag{4.30}$$

For simplicity without loss of generality we set the coefficients of the NLS $\nu_1 = -1$, $\nu_2 = 1$ in this section. The proofs work analogously for other values of ν_1 and ν_2.

Theorem 4.26 (global error, cf. [16], Theorem 7.1)**.** *Let the solution of the cubic nonlinear Schrödinger equation (4.21) satisfy*

$$m_4 := \max_{0\le t\le T} \|\psi(t)\|_{H^4} < \infty, \quad T > 0.$$

Then the numerical solution ψ_n given by the splitting scheme above with step size $\tau > 0$ obeys the following error bounds in H^2 and L^2

$$\begin{aligned} \|\psi_n - \psi(t_n)\|_{H^2} &\leq C(m_4, T)\,\tau \\ \|\psi_n - \psi(t_n)\|_{L^2} &\leq C(m_4, T)\,\tau^2 \end{aligned} \qquad \text{for} \quad t_n = n\tau \leq T.$$

The statement of this main theorem of this section is also shown in figure 5.8. In order to prove it we will state some auxiliary results on the stability and the local error of the splitting theme (4.29) and introduce the concept of the Lie derivatives. We write one step of the splitting scheme (4.29) with step size τ briefly as

$$\psi_{n+1} = \Phi_\tau(\psi_n).$$

Lemma 4.27. *The operator $e^{it\,\Delta}$ is an isometry in the space $L^2([-\pi,\pi])$ and the sobolev spaces $H^s([-\pi,\pi])$, $\forall s \in \mathbb{N}$, i.e. for $u_0 \in H^s$ there holds*

$$\left\|e^{it\,\Delta}u_0\right\|_{H^s} = \|u_0\|_{H^s} \qquad \forall s \in \mathbb{N}_0.$$

Proof. Consider the free Schrödinger equation

$$\begin{aligned} i\frac{\partial}{\partial t}u = -\Delta u, \qquad u(x,0) = u_0(x), \\ u(-\pi,\,\cdot) = u(\pi,\,\cdot), \\ \partial_x u(-\pi,\,\cdot) = \partial_x u(\pi,\,\cdot) \end{aligned}$$

Its equivalent formulation in Fourier space with $\hat{u}(k,t) = \mathcal{F}(u(x,t))$ reads

$$i\frac{\partial}{\partial t}\mathcal{F}(u) = -\mathcal{F}(\Delta u), \qquad \hat{u}(k,0) = \mathcal{F}(u_0(x))(k) =: \hat{u}_0(k),$$

where the Fourier transform induces the periodic boundary conditions naturally. Since $\mathcal{F}(\Delta u) = -k^2\mathcal{F}(u)$, as we know from lemma A.12, we can give the solution in Fourier space as

$$\mathcal{F}(u)(k,t) = e^{-ik^2t}\hat{u}_0(k).$$

Therefore

$$u(x,t) = \mathcal{F}^{-1}e^{-ik^2t}\mathcal{F}(u_0(x)).$$

Generally each derivative $\partial_x^m u,\ m \geq 0$ can be expressed in Fourier space as

$$\mathcal{F}\left(\frac{\partial^m}{\partial x^m}u\right) = i^m k^m \mathcal{F}(u),$$

which yields

$$\frac{\partial^m}{\partial x^m}u = \mathcal{F}^{-1}i^m k^m \mathcal{F}\left(\mathcal{F}^{-1}e^{-ik^2t}\mathcal{F}(u_0)\right) = \mathcal{F}^{-1}i^m k^m e^{-ik^2t}\mathcal{F}(u_0).$$

From theorem A.13 we know that the Fourier transform and its inverse is an isometry with respect to the L^2 norm such that

$$\left\|\frac{\partial^m}{\partial x^m}u\right\|_{L^2} = \|k^m\mathcal{F}(u_0)\|_{L^2} = \left\|\frac{\partial^m}{\partial x^m}u_0\right\|_{L^2}.$$

Hence

$$\left\|e^{it\,\Delta}u_0\right\|_{H^s}^2 = \|u(\,\cdot\,,t)\|_{H^s}^2 = \sum_{m=0}^{s}\left\|\frac{\partial^m}{\partial x^m}u(\,\cdot\,,t)\right\|_{L^2}^2 = \|u_0\|_{H^s}^2, \qquad \forall s \in \mathbb{N}_0.$$

□

We write one step of the splitting scheme (4.29) with step size τ briefly as

$$\psi_{n+1} = \Phi_\tau(\psi_n).$$

Proposition 4.28 (H^1- conditional L^2- and H^1-stability, H^2- conditional H^2-stability, cf. [16], Proposition 7.3)**.** *Let $\psi, \phi \in H^1$ with*

$$\|\psi\|_{H^1} \le M_1, \quad \|\phi\|_{H^1} \le M_1,$$

then

$$\begin{aligned} a) \quad & \|\Phi_\tau(\psi) - \Phi_\tau(\phi)\|_{L^2} \le e^{c_0\tau} \|\psi - \phi\|_{L^2}, \\ b) \quad & \|\Phi_\tau(\psi) - \Phi_\tau(\phi)\|_{H^1} \le e^{c_1\tau} \|\psi - \phi\|_{H^1}, \end{aligned}$$

where c_0, c_1 only depend on M_1.

If we assume instead $\psi, \phi \in H^2$ with

$$\|\psi\|_{H^2} \le M_2, \quad \|\phi\|_{H^2} \le M_2,$$

there holds

$$c) \quad \|\Phi_\tau(\psi) - \Phi_\tau(\phi)\|_{H^2} \le e^{c_2\tau} \|\psi - \phi\|_{H^2},$$

where c_2 only depends on M_2.

Proof. *a) Since $e^{i\tau\Delta}$ conserves both the L^2 and the H^1 norm, we only need to compare $e^{-i\tau|\psi^2|}\psi$ and $e^{-i\tau|\phi^2|}\phi$, which are solutions at time τ of the linear initial value problems*

$$\begin{aligned} i\dot\theta &= |\psi|^2\theta, \qquad \theta(0) = \psi, \\ i\dot\eta &= |\phi|^2\eta, \qquad \eta(0) = \phi. \end{aligned}$$

We use the identity $|z|^2 = z\overline{z}$ for some complex number z and rewrite

$$|\psi|^2\theta - |\phi|^2\eta = (\psi - \phi)\,\overline{\psi}\theta + \phi\left(\overline{\psi} - \overline{\phi}\right)\theta + \phi\overline{\phi}\,(\theta - \eta).$$

By lemma A.17 we then have

$$\begin{aligned} \left\||\psi|^2\theta - |\phi|^2\eta\right\|_{L^2} \le \quad & K_0 \, \|\psi - \phi\|_{H^1} \, \|\psi\|_{H^1} \, \|\theta\|_{L^2} \\ + \quad & K_0 \, \|\psi - \phi\|_{H^1} \, \|\phi\|_{H^1} \, \|\theta\|_{L^2} \\ + \quad & K_0 \, \|\phi\|_{H^1}^2 \, \|\theta - \eta\|_{L^2}. \end{aligned}$$

Recalling that by (4.30) $\|\theta\|_{L^2} = 1$ we obtain by integration of the right hand side

$$\|\theta(t) - \eta(t)\|_{L^2} \le \|\psi - \phi\|_{L^2} + 2K_0M_1t\,\|\psi - \phi\|_{L^2} + \int_0^t K_0M_1^2\,\|\theta(s) - \eta(s)\|_{L^2}\,ds.$$

Now we set $a(t) := \|\theta(t) - \eta(t)\|_{L^2}$ and $b(t) := (1 + 2K_0M_1t)\,\|\psi - \phi\|_{L^2}$. Then we have

$$a(t) \le b(t) + \int_0^t \lambda a(s)ds, \quad \lambda = K_0M_1^2$$

and obtain by Gronwall's lemma A.24 with c_0 only depending on M_1

$$\left\|e^{-i\tau|\psi|^2}\psi - e^{-i\tau|\phi|^2}\phi\right\|_{L^2} = \|\theta(\tau) - \eta(\tau)\|_{L^2} \le e^{c_0\tau}\,\|\psi - \phi\|_{L^2},$$

which yields the desired first inequality.

b) We proceed in the same way for the H^1 estimate. Firstly, using again lemma A.17 we get the estimates

$$\left\| |\psi|^2\theta - |\phi|^2\eta \right\|_{H^1} \leq K_2 \left\|\psi - \phi\right\|_{H^1} \left(\|\psi\|_{H^1} + \|\phi\|_{H^1}\right) \|\theta\|_{H^1} + K_2 \left\|\phi\right\|_{H^1}^2 \|\theta - \eta\|_{H^1} .$$

and

$$\left\|\dot{\theta}\right\|_{H^1} = \left\| |\psi|^2\theta \right\|_{H^1} \leq K_2 \left\|\psi\right\|_{H^1}^2 \|\theta\|_{H^1} .$$

Secondly Gronwall's inequality gives a bound for $\|\theta\|_{H^1}$:

$$\|\theta(t)\|_{H^1} \leq \|\psi\|_{H^1} + \int_0^t K_2 M_1^2 \|\theta(s)\|_{H^1} \, ds$$
$$\overset{Lemma A.24}{\Rightarrow} \|\theta(t)\|_{H^1} \leq e^{a_1 t} \|\psi\|_{H^1} ,$$

where a_1 only depends on M_1. So altogether

$$\begin{aligned} \|\theta(t) - \eta(t)\|_{H^1} \leq \quad & \|\psi - \phi\|_{H^1} + \int_0^t 2K_2 M_1^2 e^{a_1 s} \|\psi - \phi\|_{H^1} \, ds \\ & + \int_0^t K_2 M_1^2 \|\theta(s) - \eta(s)\|_{H^1} \, ds. \end{aligned}$$

And another time Gronwall's lemma A.24 provides the desired estimate

$$\left\| e^{-i\tau|\psi|^2}\psi - e^{-i\tau|\phi|^2}\phi \right\|_{H^1} = \|\theta(\tau) - \eta(\tau)\|_{H^1} \leq e^{c_1\tau} \|\psi - \phi\|_{H^1} ,$$

where c_1 only depends on H^1.

c) The proof of c) works exactly as the proof of b). We only have to replace each H^1 norm by a H^2 norm (see the last inequality in lemma A.17) and K_2 by K_3. Likewise we have to replace the estimates $\|\psi\|_{H^1} \leq M_1$ by $\|\psi\|_{H^2} \leq M_2$, for ϕ analogously, and a_1 by some a_2 only depending on M_2. Then the rest is exactly the same and we get the desired estimate.

□

Proposition 4.29 (local error in L^2 and H^2, cf. [16] Proposition 7.3 and 7.4)**.** *If $\psi_0 \in H^4$ with $\|\psi_0\|_{H^4} \leq M_4$, then the error after one step of the method (4.29) is bounded with a constant C_4 only depending on M_4*

a) in the H^2 norm by

$$\|\psi_1 - \psi(\tau)\|_{H^2} \leq C_4 \tau^2,$$

b) in the L^2 norm by

$$\|\psi_1 - \psi(\tau)\|_{L^2} \leq C_4 \tau^3.$$

Before we can prove this proposition we have to give the definition of the *Lie-brackets*, also called *commutator* of two vector fields, and explain the concept of the *Lie-derivative*.

Definition 4.30 (Lie derivative, [16], sect. 4.3)**.** *Let F and G be vector fields on H^m, $m \in \mathbb{N}$, and let $v_0 \in H^m$. Then the Lie derivative D_F of G in direction of F is defined by*

$$\left(D_F G\right)(v_0) \coloneqq \left.\frac{d}{dt}\right|_{t=0} G\left(\varphi_F^t(v_0)\right) = G'(v_0)F(v_0).$$

Here v_0 can be understood as the initial value of a differential equation

$$\dot{v} = F(v), \qquad v(0) = v_0,$$

and $\varphi_F^t(v_0)$ is the corresponding flow to the initial value v_0.

Here F and G can be operators such as the Laplacian or a multiplication operator. Furthermore we have

$$\big(\exp(tD_F)G\big)(v) = G\left(\varphi_F^t(v)\right).$$

In particular if we apply this concept to the identity operator Id we have

$$\exp(tD_F)\,\mathrm{Id}(v) = \varphi_F^t(v).$$

such that

$$\frac{d}{dt}\exp(tD_F)G(v) = \big(D_F\exp(tD_F)G\big)(v) = \big(\exp(tD_F)D_FG\big)(v).$$

Definition 4.31 (Lie-Brackets, [10], III.5.2)**.** *Let F and G be vector fields on H^m, $m \in \mathbb{N}$, and let $v_0 \in H^m$. Then the commutator of F and G is defined as*

$$\big([F,G]\big)(v) \coloneqq F'(v)G(v) - G'(v)F(v).$$

One can easily show that for two vector fields F and G there holds

$$[D_F, D_G] = D_{[G,F]},$$

i.e. the commutator of the Lie derivatives of F and G is the Lie derivative of the commutator of the vector fields in reversed order.

We define the differentiation operator T by $T(\psi) \coloneqq i\,\Delta\psi$ and the multiplication operator V by $V(\psi) \coloneqq -i|\psi|^2\psi$ and its sum $H \coloneqq T + V$ on dense subsets of H^2. The commutator of the vector fields $T(\psi)$ and $V(\psi)$ can be calculated as

$$\begin{aligned}
[T,V](\psi) &= T'(\psi)V(\psi) - V'(\psi)T(\psi)\\
&= i\,\Delta\left(-i(\psi\overline{\psi})\psi\right) - (-i)\left[\left(i\,\Delta\psi\overline{\psi}\right)\psi + \left(\psi\overline{i\,\Delta\psi}\right)\psi + \left(\psi\overline{\psi}\right)i\,\Delta\psi\right]\\
&= 2\left(\nabla\psi\cdot\nabla\overline{\psi}\right)\psi + 2\left(\nabla\psi\overline{\psi}\right)\cdot\nabla\psi + 2\left(\psi\nabla\overline{\psi}\right)\cdot\nabla\psi + 2\left(\psi\overline{\Delta\psi}\right)\psi
\end{aligned} \tag{4.31}$$

and the following lemma gives a H^2 bound for it.

Lemma 4.32 (cf. [16] chapter 8)**.** *The commutator $[T,V]$ is bounded in H^2 by*

$$\|[T,V](\psi)\|_{H^2} \le C\,\|\psi\|_{H^4}^3$$

and the double commutator $[T,[T,V]]$ in L^2 by

$$\|[T,[T,V]](\psi)\|_{L^2} \le C\,\|\psi\|_{H^4}^3.$$

Proof. The first inequality follows from (4.31) by just applying the H^2 inequality from lemma A.17. The second one can be proved analogously. □

Now let us continue with the proof of proposition 4.29.

Proof (of proposition 4.29). *The proof is divided into three parts. In part a) we derive the error* $\psi_1 - \psi(\tau)$ *and afterwards in parts b) and c) we set up bounds for it in* H^2 *and* L^2 *respectively.*

a) *We consider the Lie derivatives* D_T, D_V, D_H *of the operators* T, V *and* H *respectively and write down the exact solution* $\psi(\tau)$ *of the NLS using the Lie derivatives and the nonlinear variation-of-constants formula as*

$$\begin{aligned}\psi(\tau) &= \exp(\tau D_H)\,\mathrm{Id}(\psi_0)\\ &= \exp(\tau D_T)\,\mathrm{Id}(\psi_0) + \int_0^\tau \exp\left((\tau - s)D_H\right) D_V \exp(sD_T)\,\mathrm{Id}(\psi_0)ds\\ &= \exp(\tau D_T)\,\mathrm{Id}(\psi_0) + \int_0^\tau \exp\left((\tau - s)D_T\right) D_V \exp(sD_T)\,\mathrm{Id}(\psi_0)ds + r_1\end{aligned}$$

with the remainder term

$$r_1 = \int_0^\tau \int_0^{\tau - s} \exp\left((\tau - s - \sigma)D_H\right) D_V \exp(\sigma D_T) D_V \exp(sD_T)\,\mathrm{Id}(\psi_0)d\sigma ds.$$

On the other hand in the notation of the Lie derivatives the numerical Strang splitting solution can be written as

$$\psi_1 = \exp(\frac{1}{2}\tau D_T)\exp(\tau D_V)\exp(\frac{1}{2}\tau D_T)\,\mathrm{Id}(\psi).$$

With the Taylor expansion

$$\exp(\tau D_V) = I + \tau D_V + \tau^2 \int_0^1 (1-\theta)\exp(\theta\tau D_V) D_V^2 d\theta$$

we obtain

$$\psi_1 = \exp(\tau D_T)\,\mathrm{Id}(\psi_0) + \tau \exp(\frac{1}{2}\tau D_T) D_V \exp(\frac{1}{2}\tau D_T)\,\mathrm{Id}(\psi_0) + r_2,$$

where the remainder r_2 *reads*

$$r_2 = \tau^2 \int_0^1 (1-\theta)\exp(\frac{1}{2}\tau D_T)\exp(\theta\tau D_V) D_V^2 \exp(\frac{1}{2}\tau D_T)\,\mathrm{Id}(\psi_0)d\theta.$$

The error now becomes

$$\begin{aligned}\psi_1 - \psi(\tau) = \tau \exp(\frac{1}{2}\tau D_T) D_V \exp(\frac{1}{2}\tau D_T)\,\mathrm{Id}(\psi_0)\\ - \int_0^\tau \exp\left((\tau - s)D_T\right) D_V \exp(sD_T)\,\mathrm{Id}(\psi_0)ds + (r_2 - r_1).\end{aligned} \tag{4.32}$$

b) ***The*** H^2 ***error bound****:*

If we define

$$f(s) \coloneqq \exp\left((\tau - s)D_T\right) D_V \exp(sD_T)\,\mathrm{Id}(\psi_0),$$

we can express the principal error term, which can be identified as the quadrature error of the midpoint rule applied to the integral over $[0, \tau]$*, in first order Peano form (see [11], chapter 1):*

$$\tau f(\frac{1}{2}\tau) - \int_0^\tau f(s)ds = \tau^2 \int_0^1 \kappa_1(\tau) f'(\theta\tau)d\theta,$$

where κ_1 *is the scalar and bounded Peano kernel* κ_1 *of the midpoint rule.*

We have

$$\begin{aligned} f'(s) &= -\exp\left((\tau-s)D_T\right)\left(D_T D_V - D_V D_T\right)\exp(sD_T)\,\mathrm{Id}(\psi_0) \\ &= -\exp\left((\tau-s)D_T\right)\left[D_T, D_V\right]\exp(sD_T)\,\mathrm{Id}(\psi_0) \\ &= \exp\left((\tau-s)D_T\right) D_{[T,V]}\exp(sD_T)\,\mathrm{Id}(\psi_0) \\ &= e^{is\,\Delta}\,[T,V]\left(e^{i(\tau-s)\,\Delta}\psi_0\right) \end{aligned}$$

and therefore since $e^{is\,\Delta}$ preserves the H^2 norm, lemma 4.32 yields the quadrature error bound

$$\left\| \tau f(\frac{1}{2}\tau) - \int_0^\tau f(s)ds \right\|_{H^2} \leq C\tau^2 \left\|\psi_0\right\|_{H^4}^3 \tag{4.33}$$

Finally we have to estimate the remainder terms r_1 and r_2. For $\|\psi_0\|_{H^2} \leq M_1$ we show that

$$\|r_1\|_{H^2} + \|r_2\|_{H^2} \leq C_1\tau^2, \tag{4.34}$$

where C_1 only depends on M_1. From the definitions we have

$$\begin{aligned} \exp(\rho D_H)D_V\exp(\sigma D_T)D_V\exp(sD_T)\,\mathrm{Id}(\psi_0) &= e^{is\,\Delta}V'\left(e^{i\sigma\,\Delta}\psi(\rho)\right)e^{i\sigma\,\Delta}V(\psi(\rho)) \\ \exp(\frac{1}{2}\tau D_T)\exp(\theta\tau D_V)D_V^2\exp(\frac{1}{2}\tau D_T)\,\mathrm{Id}(\psi_0) &= e^{i\tau\,\Delta/2}V'(\eta)V(\eta), \end{aligned}$$

where $\eta = e^{-i\theta\tau|\phi|^2}\phi$ with $\phi = e^{i\tau\,\Delta/2}\psi_0$ and $\|\eta\|_{H^2} \leq e^{a_2\tau}\|\psi_0\|_{H^2}$ by the proof of proposition 4.28. Again lemma A.17 gives the bounds

$$\|V(\psi)\|_{H^2} \leq \|\psi\|_{H^2}^3 \quad \text{and} \quad \|V'(\psi)\phi\|_{H^2} \leq C\|\psi\|_{H^2}^2\|\phi\|_{H^2} \tag{4.35}$$

and we obtain the bound (4.34), which together with (4.33) yields the result

$$\|\psi_1 - \psi(\tau)\|_{H^2} \leq C_4\tau^2.$$

c) **The L^2 error bound**

The proof of the local error in L^2 works pretty similar to the estimate in H^2. We consider again the error formula (4.32) and this time we rewrite it in second-order Peano form

$$\tau f(\frac{1}{2}\tau) - \int_0^\tau f(s)ds = \tau^3\int_0^1 \kappa_2(\tau)f''(\theta\tau)d\theta,$$

with the Peano kernel κ_2 of the midpoint rule. The second derivative of f reads

$$\begin{aligned} f''(s) &= \exp\left((\tau-s)D_T\right)\left(D_T\left[D_T,D_V\right] - \left[D_T,D_V\right]D_T\right)\exp(sD_T)\,\mathrm{Id}(\psi_0) \\ &= \exp\left((\tau-s)D_T\right)\left[D_T,\left[D_T,D_V\right]\right]\exp(sD_T)\,\mathrm{Id}(\psi_0) \\ &= \exp\left((\tau-s)D_T\right)D_{[T,[T,V]]}\exp(sD_T)\,\mathrm{Id}(\psi_0) \\ &= e^{is\,\Delta}[T,[T,V]]\left(e^{i(\tau-s)\,\Delta}\psi_0\right) \end{aligned}$$

such that once more lemma 4.32 implies

$$\left\| \tau f(\frac{1}{2}\tau) - \int_0^\tau f(s)ds \right\|_{L^2} \leq C\tau^3\left\|\psi_0\right\|_{H^4}^3. \tag{4.36}$$

Defining the function

$$g(s,\sigma) \coloneqq \exp\left((\tau-s-\sigma)D_T\right)D_V\exp(\sigma D_T)D_V\exp(sD_T)\,\mathrm{Id}(\psi_0)$$

we can express the remainder as

$$\begin{aligned} r_2 - r_1 &= \tau^2 \int_0^1 (1-\theta) g(\frac{\tau}{2}, 0) d\theta - \int_0^\tau \int_0^{\tau - s} g(s, \sigma) d\sigma ds + \widetilde{r_2} - \widetilde{r_1} \\ &= \frac{\tau^2}{2} g(\frac{\tau}{2}, 0) - \int_0^\tau \int_0^{\tau - s} g(s, \sigma) d\sigma ds + \widetilde{r_2} - \widetilde{r_1}, \end{aligned}$$

where

$$\begin{aligned} \widetilde{r_1} = \int_0^\tau \int_0^{\tau - s} \int_0^{\tau - s - \sigma} &\exp\left((\tau - s - \sigma - \xi) D_H\right) D_V \exp(\xi D_T) \\ &D_V \exp(\sigma D_T) D_V \exp(s D_T) \operatorname{Id}(\psi_0) d\xi d\sigma ds \end{aligned}$$

and

$$\widetilde{r_2} = \tau^2 \int_0^1 (1-\theta) \exp(\frac{1}{2}\tau D_T) \left[\exp(\theta \tau D_V) - I\right] D_V^2 \exp(\frac{1}{2}\tau D_T) \operatorname{Id}(\psi_0) d\theta.$$

Considering $u' = V(u)$ with $u(0) = u_0$ we can write

$$\left[\exp(\theta \tau D_V) - I\right](u_0) = \int_0^{\theta\tau} u'(s) ds$$

and set up the bound

$$\|\widetilde{r_1}\|_{L^2} + \|\widetilde{r_2}\|_{L^2} \leq \widetilde{C}_2 \tau^3$$

similarly as before for the norms of r_1 and r_2 with $\widetilde{C}_2$ depending only on $\|\psi_0\|_{H^2}$. The norm of the part which is left in $r_2 - r_1$ can be estimated using the quadrature error of a first order two-dimensional quadrature formula:

$$\left\| \frac{\tau^2}{2} g(\frac{\tau}{2}, 0) - \int_0^\tau \int_0^{\tau - s} g(s, \sigma) d\sigma ds \right\|_{L^2} \leq C \tau^3 \left(\max_m \left\| \frac{\partial}{\partial s} g \right\|_{L^2} + \max_m \left\| \frac{\partial}{\partial \sigma} g \right\|_{L^2} \right),$$

with $m \coloneqq \{(s, \sigma) \mid 0 \leq s \leq \tau,\ 0 \leq \sigma \leq \tau - s\}$. The partial derivatives of g only contain V and the commutator $[T, V]$ and derivatives of them such that their L^2 norms can be bounded in terms of the $\|\psi_0\|_{H^2}$ using again lemma 4.32 and the same arguments as for the estimates in (4.35).

This yields

$$\|r_2 - r_1\|_{L^2} \leq C_2 \tau^3, \tag{4.37}$$

with C_2 only depending on $\|\psi_0\|_{H^2}$. Altogether

$$\begin{aligned} \|\psi_1 - \psi(\tau)\|_{L^2} &\leq \left\| \tau f(\frac{1}{2}\tau) - \int_0^\tau f(s) ds \right\|_{L^2} + \|r_2 - r_1\|_{L^2} \\ &\overset{(4.36),(4.37)}{\leq} \left(C \|\psi_0\|_{H^4}^3 + C_2 \right) \tau^3 \leq C_4 \tau^3. \end{aligned}$$

□

Now we have set up all tools that we need to prove theorem 4.26:

Proof (of theorem 4.26). *The results of proposition 4.28 on the H^1- and H^2- conditional stability and the local error bounds from proposition 4.29 yield the result on the global error with the standard argument of Lady Windermere's fan (cf. [11], chapter 8.3).* □

4.3. Time Integrators for the Klein Gordon Equation

Since we want to compare the NLS approximation $\varepsilon\psi_{NLS}$ in (3.16) with the solution u of the Klein-Gordon equation (3.2) it is important not only to have a reliable integrator for the nonlinear Schrödinger equation but also to choose an adequate solver for the Klein-Gordon equation. Therefore in this section I present three different integrators for this nonlinear wave equation: I will describe at first a pretty young splitting integrator, then a well known Runge-Kutta method, namely the implicit midpoint rule, and finally the well-proven Leapfrog method. But instead of doing a lengthy error analysis as we did for the splitting method applied to the Gross-Pitaevskii equation in section 4.2 we will later give some numerical results for this methods considering the numerical order and the preservation of the energy and the norm. Now let us have a look at the following splitting method.

4.3.1. A Splitting Method for the Klein-Gordon Equation

In this section I want to present a splitting method that we want to use to solve the nonlinear Klein Gordon equation from chapter 3. The framework is basically taken from [9], chapter 2.

If we use a slightly different notation for the nonlinear part of equation (3.2) we have

$$\frac{\partial^2}{\partial t^2} z = a_1\,\Delta z + a_2 z + a_3|z|^2 z, \qquad z(\,\cdot\,,0) = z_0, \quad \partial_t z(\,\cdot\,,0) = w_0 \tag{4.38}$$

with $a_1 \geq 0$, $a_2 \leq 0$ and $a_3 \in \mathbb{R}$. Furthermore for practical implementation issues we introduce periodic boundary conditions

$$z(-L,\,\cdot\,) = z(L,\,\cdot\,), \qquad \partial_x z(-L,\,\cdot\,) = \partial_x z(L,\,\cdot\,),$$

where we choose $L > 0$ large enough, such that the boundary conditions are neglectable.

For simplicity we transform this periodicity interval to $[-\pi,\pi]$, whereas all statements in this section are also valid for the interval $[-L,L]$ considering this transformation. To a given number $N \in \mathbb{N}$ we choose the spatial step size $h = \frac{2\pi}{N}$ and discretize space by $x_j = -\pi + jh,\ j = 1,\dots N$.

In order to get a well-arranged notation let us define the operator

$$\langle\Delta\rangle \coloneqq (-a_2 - a_1\,\Delta)^{1/2},$$

and consider its representation in Fourier space. To a given function $g \in H^1([-L,L])$ set $g^j = g(x_j),\ j = 1,\dots,N$ and $\tilde{g} = \left(g^j\right)_{j=1}^N$. Then we have

$$\left(\langle\Delta\rangle\, g(x_j)\right)_{j=1}^N \approx \mathcal{F}_N^{-1}\left\{\sqrt{a_1k^2 - a_2}\,\left(\mathcal{F}_N\tilde{g}\right)_k\right\}_{k=-\frac{N}{2}+1}^{\frac{N}{2}}.$$

Please note that due to the demands for a_1 and a_2 the square root $\sqrt{a_1k^2 - a_2}$ is well defined. For notational simplicity we define

$$\mu_k = \sqrt{a_1k^2 - a_2}.$$

If we make the ansatz

$$\begin{aligned} u &= z - i\,\langle\Delta\rangle^{-1}\,\partial_t z, & u(\,\cdot\,,0) &= z_0 - i\,\langle\Delta\rangle^{-1}\,w_0 \eqqcolon u_0,\\ v &= \overline{z} - i\,\langle\Delta\rangle^{-1}\,\overline{\partial_t z}, & v(\,\cdot\,,0) &= \overline{z_0} - i\,\langle\Delta\rangle^{-1}\,\overline{w_0} \eqqcolon v_0, \end{aligned} \tag{4.39}$$

to get a first order differential equation in time we find that

$$z = \frac{1}{2}(u + \overline{v}). \tag{4.40}$$

Differentiating u and v with respect to t we can use by (4.39) that

$$i\partial_t z = \langle\Delta\rangle\,(z - u), \qquad i\overline{\partial_t z} = \langle\Delta\rangle\,(\overline{z} - v)$$

and replace $\partial_t^2 z$ with the right hand side of (4.38). This yields

$$\begin{aligned} i\partial_t u &= -\langle\Delta\rangle u + \frac{a_3}{8}\langle\Delta\rangle^{-1}|u+\overline{v}|^2(u+\overline{v}),\\ i\partial_t v &= -\langle\Delta\rangle v + \frac{a_3}{8}\langle\Delta\rangle^{-1}|\overline{u}+v|^2(\overline{u}+v) \end{aligned}, \quad u(\cdot,0)=u_0, \quad v(\cdot,0)=v_0, \tag{4.41}$$

which we can split into the two subproblems

$$\begin{aligned} i\partial_t u &= -\langle\Delta\rangle u\\ i\partial_t v &= -\langle\Delta\rangle v \end{aligned} \qquad u(\cdot,0)=\tilde{u}_0, \quad v(\cdot,0)=\tilde{v}_0, \tag{I}$$

and

$$\begin{aligned} i\partial_t u &= \frac{a_3}{8}\langle\Delta\rangle^{-1}|u+\overline{v}|^2(u+\overline{v}),\\ i\partial_t v &= \frac{a_3}{8}\langle\Delta\rangle^{-1}|\overline{u}+v|^2(\overline{u}+v) \end{aligned} \qquad u(\cdot,0)=\tilde{u}_0, \quad v(\cdot,0)=\tilde{v}_0. \tag{II}$$

For the rest of this section we omit the x dependence and just write $u(t)$ instead of $u(x,t)$. Furthermore we assume the functions already to be discretized, such that $u(t)=(u(x_j,t))_{j=1}^N$.

In Fourier space subproblem (I) can be solved exactly except for the spatial discretization error by

$$\begin{aligned} \mathcal{F}_N u(t) &= e^{i\mu_k t}\mathcal{F}_N\tilde{u}_0,\\ \mathcal{F}_N v(t) &= e^{i\mu_k t}\mathcal{F}_N\tilde{v}_0, \end{aligned} \tag{4.42}$$

cf. section 4.2.3.

An easy calculation shows that in subproblem (II)

$$\partial_t\left(u(t)+\overline{v}(t)\right)=0, \qquad \text{i.e.} \qquad u(t)+\overline{v}(t)=\tilde{u}_0+\overline{\tilde{v}}_0 \qquad \forall t.$$

Therefore with $\tilde{z}_0 \coloneqq \frac{1}{2}\left(\tilde{u}_0+\overline{\tilde{v}}_0\right)$ we can solve (II) also exactly in Fourier space by

$$\begin{aligned} \mathcal{F}_N u(t) &= \mathcal{F}_N\tilde{u}_0 - i\,t\,\frac{a_3}{\mu_k}\mathcal{F}_N\left(|\tilde{z}_0|^2\,\tilde{z}_0\right),\\ \mathcal{F}_N v(t) &= \mathcal{F}_N\tilde{v}_0 - i\,t\,\frac{a_3}{\mu_k}\mathcal{F}_N\left(|\overline{\tilde{z}_0}|^2\,\overline{\tilde{z}_0}\right). \end{aligned} \tag{4.43}$$

Since we can solve both subproblems exactly in Fourier space it is obvious to use the same strategy as in the section about the nonlinear Schrödinger equation to get an approximate solution of (4.38), i.e. we apply once more a Strang splitting method to our wave equation.

Let us discretize time with time step size δ_t such that to a given time t_0 and a number $M\in\mathbb{N}$ we have

$$t_m = t_0 + m\delta_t,\ m=1,\dots,M$$

and we shortly write $u_m \coloneqq\approx u(t_m)$

We solve subproblem (I) at time t by the mapping

$$\Phi_1^t(u_0,v_0)=(u_1,v_1)$$

which gives the solution at time t for given initial values u_0, v_0 according to (4.42). Similarly based on the scheme (4.43) we denote the mapping which gives the solution of subproblem (II) at time t for initial values u_0, v_0 by

$$\Phi_2^t(u_0,v_0)=(u_2,v_2).$$

Then the Strang splitting method applied to equation (4.41) performing one step δ_t reads as follows: For given initial values u_0, v_0 we have

$$(u_1,v_1)=\Phi_1^{\delta_t/2}\circ\Phi_2^{\delta_t}\circ\Phi_1^{\delta_t/2}(u_0,v_0) =: \Psi^{\delta_t}(u_0,v_0). \tag{4.44}$$

Then we obtain an approximate solution to the wave equation (4.38) after one time step by using the identity (4.40). Therefore

$$z(t_1) \approx z_1 \coloneqq \frac{1}{2}(u_1 + \overline{v}_1).$$

Analogously in order to obtain an approximation z_m to $z(t_m)$ we apply Ψ^{δ_t} m-times to given initial values u_0, v_0, i.e.

$$(u_m, v_m) = \underbrace{\Psi^{\delta_t} \circ \Psi^{\delta_t} \circ \cdots \circ \Psi^{\delta_t}}_{m-times}(u_0, v_0) = \left(\Psi^{\delta_t}\right)^m (u_0, v_0), \tag{4.45}$$

Then

$$z_m = \frac{1}{2}(u_m + \overline{v}_m).$$

Considering efficiency of this solver it is advisable to combine two succeeding half steps of the mapping Φ_1 in one full step,i.e.

$$\begin{aligned}\left(\Psi^{\delta_t}\right)^m (u_0, v_0) &= \Phi_1^{\delta_t/2} \circ \Phi_2^{\delta_t} \circ \underbrace{\Phi_1^{\delta_t/2} \circ \Phi_1^{\delta_t/2}}_{=\Phi_1^{\delta_t}} \circ \cdots \circ \underbrace{\Phi_1^{\delta_t/2} \circ \Phi_1^{\delta_t/2}}_{=\Phi_1^{\delta_t}} \circ \Phi_2^{\delta_t} \circ \Phi_1^{\delta_t/2}(u_0, v_0) \\ &= \Phi_1^{\delta_t/2} \circ \Phi_2^{\delta_t} \circ \left(\Phi_1^{\delta_t} \circ \Phi_2^{\delta_t}\right)^{m-1} \circ \Phi_1^{\delta_t/2}(u_0, v_0).\end{aligned} \tag{4.46}$$

We can summarize this section in algorithm 4.2.

Algorithm 4.2: Strang Splitting method for the nonlinear wave equation (4.38)

Having the coefficients a_1, a_2, a_3 from (4.38)we set $\mu = \left(\sqrt{a_1 k^2 - a_2}\right)_{k=-\frac{N}{2}+1}^{\frac{N}{2}}$.

begin

given the initial values $z_0 = (z(x_j, 0))_{j=1}^n$, $w_0 = (\partial_t z(x_j, 0))_{j=1}^n$

we set

$$\tilde{u}_0 \coloneqq \mathcal{F}_N z_0 - i\,\mu^{-1}\mathcal{F}_N w_0, \qquad \tilde{v}_0 \coloneqq \mathcal{F}_N \overline{z_0} - i\,\mu^{-1}\mathcal{F}_N \overline{w_0}$$

given time step size δ_t we compute

$u_1 = \mathcal{F}_N^{-1}\, e^{i\mu\delta_t/2}\, \tilde{u}_0$,

$v_1 = \mathcal{F}_N^{-1}\, e^{i\mu\delta_t/2}\, \tilde{v}_0$

and start the iteration

for $m = 1, 2, \ldots, M$ **do**

set $z_m = \frac{1}{2}(u_m + \overline{v_m})$

$\tilde{u}_m = \mathcal{F}_N u_m - i\, a_3 \mu^{-1}\mathcal{F}_N\left(|z_m|^2 z_m\right)$,

$\tilde{v}_m = \mathcal{F}_N v_m - i\, a_3 \mu^{-1}\mathcal{F}_N\left(|z_m|^2 \overline{z_m}\right)$

note that $\tilde{u}_m, \tilde{v}_m$ are in Fourier space, then

$u_{m+1} = \mathcal{F}_N^{-1} e^{i\mu\delta_t}\tilde{u}_m$,

$v_{m+1} = \mathcal{F}_N^{-1} e^{i\mu\delta_t}\tilde{v}_m$

end

$u_M = \mathcal{F}_N^{-1} e^{-i\mu\delta_t/2}\mathcal{F}_N u_M$,

$v_M = \mathcal{F}_N^{-1} e^{-i\mu\delta_t/2}\mathcal{F}_N v_M$,

then we have

$z_M = \frac{1}{2}(u_M + \overline{v_M}) \approx z(t_M)$.

end

4.3.2. Time Integration by the Implicit Midpoint Rule

In this section we want to derive an algorithm which allows us to apply the implicit midpoint rule, a well known Runge-Kutta method, efficiently to the Klein-Gordon equation by applying the technique of spectral discretization of space, that we introduced in section 4.1. Having the time discretization

$t_n = t_0 + n\delta_t$, $n = 0, 1, \ldots, M$ for some $M \in \mathbb{N}$ and $t_0 = 0$ we can write down the exact solution of the differential equation

$$y' = f(y), \qquad y(t_0) = y_0$$

at time t_1 as

$$y(t_1) = y_0 + \delta_t \int_0^1 f(y(s\delta_t))ds.$$

This integral can be approximated by a quadrature formula as

$$\int_0^1 f(y(s\delta_t))ds \approx f\left(y\left(\frac{1}{2}\delta_t\right)\right),$$

which is known as the so called *midpoint rule* (cf. [11], chapter 1). Therefore if we use $\frac{1}{2}(y(t_0) + y(t_1)) \approx y(t_0 + \delta_t/2)$ we can approximate $y(t_1)$ by the Gauß collocation method

$$y_1 = y_0 + \delta_t f\left(\frac{y_0 + y_1}{2}\right),$$

simply called the implicit midpoint rule.

Then if we discretize space by a spectral method and introduce periodic boundary conditions with $L > 0$ large enough as in section 4.2.3 and 4.3 we have

$$\begin{aligned} \frac{\partial^2}{\partial t^2}u = \alpha\frac{\partial^2}{\partial x^2}u - \beta u + \lambda u^3, \quad u(\cdot, 0) = u_0, \quad \partial_t u(\cdot, 0) = v_0, \\ u(-L, \cdot) = u(L, \cdot) \qquad \partial_x u(-L, \cdot) = \partial_x u(L, \cdot), \end{aligned} \tag{4.47}$$

which we can reformulate as a first order system

$$\begin{bmatrix} u(t) \\ v(t) \end{bmatrix}' = \underbrace{\begin{bmatrix} 0 & 1 \\ \alpha\frac{\partial^2}{\partial x^2} - \beta & 0 \end{bmatrix}}_{:=A} \begin{bmatrix} u(t) \\ v(t) \end{bmatrix} + \begin{bmatrix} 0 \\ \lambda u^3(t) \end{bmatrix} =: f(y(t)),$$

where $v(t) = \frac{\partial}{\partial t}u(t)$ and $y(t) = \begin{bmatrix} u(t) \\ v(t) \end{bmatrix}$. Denoting $u_n \approx u(t_n)$, $v_n \approx v(t_n)$ and $y_n = \begin{bmatrix} u_n \\ v_n \end{bmatrix}$ the implicit midpoint rule yields

$$\begin{aligned} & & y_1 = \;& y_0 + \frac{\delta_t}{2}A(y_0 + y_1) + \lambda\delta_t \begin{bmatrix} 0 \\ ((u_0 + u_1)/2)^3 \end{bmatrix} \\ \iff & & (I - \frac{\delta_t}{2}A)y_1 = \;& (I + \frac{\delta_t}{2}A)y_0 + \lambda\delta_t \begin{bmatrix} 0 \\ ((u_0 + u_1)/2)^3 \end{bmatrix} \\ \iff & & 0 = \;& (I - \frac{\delta_t}{2}A)y_1 - (I + \frac{\delta_t}{2}A)y_0 - \lambda\delta_t \begin{bmatrix} 0 \\ ((u_0 + u_1)/2)^3 \end{bmatrix} \\ & & =: \;& F(y_1; y_0), \end{aligned} \tag{4.48}$$

where y_0 is treated as a parameter only.

At this point assuming that we have discretized the 2π periodic function $u(t) = u(x, t)$ in space at N discrete points, a simplified Newton's method can be applied to obtain y_1: Recalling the definition of y_1 the derivative of F is given by

$$F'(y_1; y_0) \coloneqq \frac{d}{dy_1}F(y_1; y_0) = (I - \frac{\delta_t}{2}A) + \lambda\delta_t \begin{bmatrix} 0 & 0 \\ g(u_1) & 0 \end{bmatrix} \approx (I - \frac{\delta_t}{2}A),$$

where $g(u_1)$ is a nonlinear function in u_1. Since we want to solve the differential equation efficiently and we want to avoid the necessity of solving nonlinear equation systems, nonlinearities in the derivative of

F are not desirable. Therefore we leave out the nonlinear term and just perform a simplified Newton's method in which we use $B \coloneqq I - \frac{\delta_t}{2}A$ as an approximation to $F'(y_1; y_0)$ in each Newton iteration step.

Denoting the initial value for the Newton iteration by $X^0 \coloneqq y_0$ we get the next value by $X^{m+1} = X^m + \zeta_m$, where

$$B\zeta_m = -F(X^m; X^0),$$

i.e. the update vector ζ_m is the solution of a linear system of equations, which is, recalling that B contains a differentiation operator, not that trivial at first. But using the spectral methods from section 4.1, we can diagonalize B with FFT such that we can solve this efficiently.

Let us write $\zeta_k = \begin{bmatrix} h_x^m \\ h_y^m \end{bmatrix}$ and $F(X^m; X^0) = \begin{bmatrix} F_1^m \\ F_2^m \end{bmatrix}$, then the equation above becomes

$$\begin{bmatrix} I & -\frac{\delta_t}{2}I \\ -\frac{\delta_t}{2}\left(\alpha\frac{\partial^2}{\partial x^2} - \beta\right) & I \end{bmatrix} \begin{bmatrix} h_x^m \\ h_y^m \end{bmatrix} = - \begin{bmatrix} F_1^m \\ F_2^m \end{bmatrix}$$

and by Gauß elimination

$$\begin{bmatrix} I & -\frac{\delta_t}{2}I \\ 0 & I - \frac{\delta_t^2}{4}\left(\alpha\frac{\partial^2}{\partial x^2} - \beta\right) \end{bmatrix} \begin{bmatrix} h_x^m \\ h_y^m \end{bmatrix} = - \begin{bmatrix} F_1^m \\ F_2^m + \frac{\delta_t}{2}\left(\alpha\frac{\partial^2}{\partial x^2} - \beta\right)F_1^m \end{bmatrix},$$

i.e.

$$\left(I - \frac{\delta_t^2}{4}\left(\alpha\frac{\partial^2}{\partial x^2} - \beta\right)\right) h_y^m = -F_2^m - \frac{\delta_t}{2}\left(\alpha\frac{\partial^2}{\partial x^2} - \beta\right)F_1^m,$$

which is equivalent to the equation in Fourier space

$$\left(1 - \frac{\delta_t^2}{4}\left(\alpha(-k^2) - \beta\right)\right)\left(\mathcal{F}_N h_y^m\right)_k = -\left(\mathcal{F}_N F_2^m\right)_k - \frac{\delta_t}{2}\left(\alpha(-k^2) - \beta\right)\left(\mathcal{F}_N F_1^m\right)_k$$

for the Fourier numbers $k = -\frac{N}{2} + 1, \dots, \frac{N}{2}$. Therefore we have

$$h_y^m = -\mathcal{F}_N^{-1}\left\{ \left(\mu_k^{-1}\left[\left(\mathcal{F}_N F_2^m\right)_k + \left(-\frac{\delta_t}{2}\alpha k^2 - \frac{\delta_t}{2}\beta\right)\left(\mathcal{F}_N F_1^m\right)_k\right] \right)_{k=-\frac{N}{2}+1}^{\frac{N}{2}} \right\}$$

with $\mu_k^{-1} = \left(1 + \frac{\delta_t^2}{4}\alpha k^2 + \frac{\delta_t^2}{4}\beta\right)^{-1}$, $k = -\frac{N}{2} + 1, \dots \frac{N}{2}$ and

$$h_x^m = -F_1^m + \frac{\delta_t}{2}h_y^m.$$

We stop the Newton iteration if the eucledian norm $\|\zeta_m\| \leq \text{tol}$ for some given tolerance tol. Altogether the implicit midpoint rule for the nonlinear wave equation can be summarized in algorithm 4.3. In chapter 5 we will do some numerical tests which approve the quality of this integrator.

4.3.3. The Leapfrog Method

At this point I want to explain another very popular time marching scheme which has arisen from the Störmer-Verlet scheme for ordinary differential equations. The literature for this section can be found in [10], chapter I.1.4. It is called the *leapfrog method* and has proved itself over decades as a reliable explicit time marching scheme for ordinary and partial differential equations with good energy and norm conservation properties. We want to consult it as a comparative solver in order to confirm the good behaviour of both the implicit midpoint rule and the Strang splitting solver for the nonlinear wave equation (4.47). Denoting its right hand side by $a(u)$ and setting $v(x,t) \coloneqq \partial_t u(x,t)$, we shortly write it down as

$$\frac{\partial^2}{\partial t^2}u = a(u), \quad u(\,\cdot\,, 0) = u_0, \quad v(\,\cdot\,, 0) = v_0, \tag{4.49}$$

Algorithm 4.3: The implicit midpoint rule for the nonlinear wave equation (4.47)

Let $u_0 = (u(x_j,0)_{j=1}^N$, $v_0 = (\partial_t u(x_j,0))_{j=1}^N$
begin
set $y_0 = \begin{bmatrix} u_0 & v_0 \end{bmatrix}^T$,
$\mu \coloneqq (\mu_k)_{k=-\frac{N}{2}+1}^{\frac{N}{2}}$, $\quad \mu_k \coloneqq 1 + \frac{\delta_t^2}{4}\alpha k^2 + \frac{\delta_t^2}{4}\beta, \quad k = -\frac{N}{2}+1, \dots \frac{N}{2}$,
$\eta \coloneqq (\eta_k)_{k=-\frac{N}{2}+1}^{\frac{N}{2}}$, $\quad \eta_k \coloneqq -\frac{\delta_t}{2}\alpha k^2 - \frac{\delta_t}{2}\beta, \quad k = -\frac{N}{2}+1, \dots \frac{N}{2}$,
and set $F(z;w)$ as in (4.48)
for $n = 0, 1, 2, \dots, M-1$ **do**
set $X^0 \coloneqq y_n$
$m = 0$
while $\|\zeta_m\| >$ tol **do**
$\begin{bmatrix} F_1^m & F_2^m \end{bmatrix}^T = F(X^m; X^0)$, where F_1^m and F_2^m are both of length N
$h_y^m = -\mathcal{F}_N^{-1}\left\{\mu^{-1} \bullet (\mathcal{F}_N F_2^m + \eta \bullet \mathcal{F}_N F_1^m)\right\}$
$h_x^m = -F_1^m + \frac{\delta_t}{2} h_y^m$
$\zeta_m = \begin{bmatrix} h_x^m \\ h_y^m \end{bmatrix}$
$X^{m+1} = X^m + \zeta_m$
$m = m + 1$
end
$y_{n+1} = X^m$
end
$u_M = y_M(1:N)$
therefore
$u(x_j, t_m) \approx u_M(j)$.
end

having in mind that for practical implementation we still assume periodic boundary conditions

$$u(-L, \cdot) = u(L, \cdot) \qquad \partial_x u(-L, \cdot) = \partial_x u(L, \cdot). \tag{4.50}$$

In this notation $v(x,t)$ can be interpreted as the velocity and

$$a(u) \coloneqq \alpha\, \Delta u - \beta u + \lambda u^3$$

as the acceleration at position x and time t. We transform the spatial domain $[-L, L]$ onto the interval $[-\pi, \pi]$ such that we can discretize space once more by $x_j = -\pi + jh, j = 1, \dots, N$ with spatial step size $h = \frac{2\pi}{N}$ and time by $t_m = t_0 + m\delta_t$, $m = 1, \dots, M$ for given time t_0 and time step size δ_t.

This method is called *leapfrog method* since the positions $u_m \approx (u(x_j, t_m))_{j=1}^N$ and the velocities $v_m \approx (v(x_j, t_m))_{j=1}^N$ are leapfrogged over each other in half time steps, i.e. for given position u_m and velocity v_m we obtain the scheme

$$\begin{aligned} v_{m+1/2} &= v_m &&+ \frac{1}{2}\delta_t a(u_m), \\ u_{m+1} &= u_m &&+ \delta_t v_{m+1/2}, \\ v_{m+1} &= v_{m+1/2} &&+ \frac{1}{2}\delta_t a(u_{m+1}). \end{aligned} \tag{4.51}$$

Here we can apply a spectral method to calculate $a(u_m)$ in each step. One can show that for ordinary differential equations this scheme is of order 2. We will see in some numerical experiments that this is also true for the application of this method to our wave equation.

CHAPTER

FIVE

NUMERICAL EXPERIMENTS

In this chapter we numerically verify the NLS approximation

$$\varepsilon\psi_{NLS}(x,t) = \varepsilon A(\xi,\sigma)e^{i(kx-\omega t)} + \varepsilon A^*(\xi,\sigma)e^{-i(kx-\omega t)}. \tag{5.1}$$

For practical implementation issues we introduce periodic boundary conditions for our solution, where we choose the spatial domain large enough such that the boundary conditions are negelectable. Therefore (3.2) becomes

$$\begin{aligned} \frac{\partial^2}{\partial t^2}u(x,t) &= a_1\frac{\partial^2}{\partial x^2}u(x,t) + a_2u(x,t) + a_3u^3(x,t), \\ u(\,\cdot\,,0) &= u_0, \qquad\qquad \partial_t u(\,\cdot\,,0) = u_1, \\ u(-L,\,\cdot\,) &= u(L,\,\cdot\,), \qquad\quad \partial_x u(-L,\,\cdot\,) = \partial_x u(L,\,\cdot\,). \end{aligned} \tag{5.2}$$

for some coefficients $a_1 > 0$, $a_2 < 0$ and $a_3 \in \mathbb{R}$, that we obtained in chapter 3.

The wave number k and the frequency ω obey the dispersion relation

$$\omega^2 = a_1k^2 - a_2$$

and the group velocity of the wave packet is given by

$$c = \frac{a_1k}{\omega}.$$

As in chapter 3, $\xi = \varepsilon(x-ct)$ and $\sigma = \varepsilon^2 t$ denote the slow variables in space and time. We have found that the envelope A in (5.1) has to satisfy the nonlinear Schrödinger equation

$$\begin{aligned} i\frac{\partial}{\partial\sigma}A = \nu_1\frac{\partial^2}{\partial\xi^2}A + \nu_2|A|^2A, \quad A(\cdot,0) \quad &= A_0, \\ A(-\varepsilon L,\,\cdot\,) &= A(\varepsilon L,\,\cdot\,), \\ \partial_x A(-\varepsilon L,\,\cdot\,) &= \partial_x A(\varepsilon L,\,\cdot\,), \end{aligned} \tag{5.3}$$

where

$$\nu_1 = \frac{c^2-a_1}{2\omega}, \qquad \nu_2 = -\frac{3a_3}{2\omega}.$$

Here the boundary conditions arise from the spatial scaling $\xi = \varepsilon(x-ct)$. Note that for $a_2 < 0$ we have

$$c^2 = \frac{a_1^2k^2}{a_1k^2-a_2} = a_1\left(1+\frac{a_2}{\omega^2}\right) < a_1 \quad \Rightarrow \quad \nu_1 < 0,$$

likewise $\nu_2 < 0$ if $a_3 > 0$.

In our simulations we restrict ourselves to the coefficients

$$a_1 = 1, \qquad a_2 = -1 \quad \text{and} \quad a_3 = 1$$

and the wave number $k = 0.3$. Therefore $\nu_1, \nu_2 < 0$ and thus soliton solutions of our Schrödinger equations can exist, cf. chapter 3.2.

Note that $\varepsilon\psi_{NLS}$ is valid as an approximation to u as long as $t \in [0, T/\varepsilon^2]$ in the sense that

$$\sup_{t\in[0,T/\varepsilon^2]} \|u(\cdot,t) - \varepsilon\psi_{NLS}(\cdot,t)\|_{H^1} \leq C\varepsilon^{3/2}$$

for some $T > 0$, cf. theorem 3.7.

We use spectral methods to solve both the Klein-Gordon and the Gross-Pitavskii equation. Therefore we have to give the count N of Fourier numbers, that we use to discretize space such that we obtain the discrete interval associated to the interval $[0, 2\pi]$,

$$\tilde{x}_j = jh_{2\pi},\ j = 1, \dots N, \quad \text{with step size} \quad h_{2\pi} = \frac{2\pi}{N}.$$

Since we want to discretize the interval $[-L, L]$ we therefore apply the simple transform $T(\tilde{x}) = -L + \frac{2L}{2\pi}\cdot\tilde{x}$ to the grid $\tilde{x}_1, \dots, \tilde{x}_N$ and get the spatial discretization

$$x_j = -L + jh,\ j = 1, \dots N, \quad \text{with step size} \quad h = \frac{2L}{2\pi}h_{2\pi} = \frac{2L}{N}.$$

Note that we integrate over a long time period up to times $t \in \mathcal{O}(\varepsilon^{-2})$, whereas $\varepsilon \to 0$. Since we want to compare the numerical solution of the nonlinear Klein-Gordon equation with the NLS approximation we need to have a neglectable error in the time integration. Therefore we choose small time steps, i.e.

$$\delta_t = h_{2\pi}/4 = \frac{\pi}{2}N^{-1},$$

as proposed in cf. [25], chapter 10. This yields the time discretization

$$t_m = m\delta_t, \quad m = 0, 1, 2, 3, \dots$$

Furthermore we want to avoid additional errors in approximating the H^1 norm. Hence we consider the error estimate in the l^∞ sense and define

$$s(t_m) = \sup_{j=1,\dots,N} |u_j(t_m) - \varepsilon\psi_{NLS}(x_j, t_m)|, \tag{5.4}$$

the maximal error of the NLS approximation at time t, where

$$u_j(t_m) \coloneqq u(jh, t_m) \approx u(x_j, t_m)$$

is the approximation of the numerical to the exact solution at time t_m. We will find that

$$\sup_{t_m\in[0,T/\varepsilon^2]} s(t_m) \in \mathcal{O}(\varepsilon^2). \tag{5.5}$$

But why not like $C\varepsilon^{3/2}$? The answer is that we lose the factor $\varepsilon^{1/2}$ if we take the L^2 or H^1 norm because of the special scaling of the slow space variable, compare to remark 3.6. But since the l^∞ norm is not subject to it we will end up with a power of 2 rather than 3/2, (cf. introduction to chapter 3 in [5]).

5.1. The Results on the NLS Approximation

In this section we want to discuss briefly some numerical simulations which shall fortify the theoretical results that we achieved in chapter 3. The results of this section can be reproduced by running the

RUN_Simulation.m. The error analysis is done by RUN_ERRORS.m. We do the simulations on the spatial domain

$$x \in [-L, L], \qquad L = \pi \cdot \frac{50}{k} \approx 524 \tag{5.6}$$

with $k = 0.3$ and in the time interval

$$t \in [0, 5/\varepsilon^2].$$

But before we start we have to clarify the initial data which we use to solve (5.3)and (5.2). Firstly we have a look at the numerical solution of the Gross-Pitaevskii equation (5.3).

One can show that

$$\psi_s(\xi, \sigma) = d_1 \ \cosh^{-1}(d_2\xi)e^{id_3\sigma} \tag{5.7}$$

with

$$d_1 = \sqrt{-\frac{2\gamma}{\nu_2}}, \quad d_2 = \sqrt{-\frac{\gamma}{\nu_1}}, \quad d_3 = \gamma$$

for some $\gamma \in \mathbb{R}$, is a soliton solution of the Gross-Pitaevskii equation. It provides the initial data for (5.3) in our first simulation, i.e. we set

$$A_0^s(\xi) = \psi(\xi, 0). \tag{5.8}$$

Here be aware that at $t = 0$ we have $\xi = \varepsilon x$. Furthermore we choose $\gamma = 0.5$. Recall that soliton solutions are solutions which do not change their shape over time, cf. the paragraph before theorem 3.2 in chapter 3.2. In a second simulation we repeat all the calculations with a different initial value, which has an almost rectangular shape. It is given by

$$A_0^r(\xi) = \frac{d_1}{4}\left[1 - \tanh\left(d_2 \cdot (\xi - 11d_3)\right)\tanh\left(d_2 \cdot (\xi + 11d_3)\right)\right]. \tag{5.9}$$

Figure 5.1 depicts how the solutions to (5.3) evolutes over time using the initial values A_0^s and A_0^r. Note that the solution $A(\xi, \varepsilon^2 t)$ is evaluated at times t instead of $\varepsilon^2 t$.

This solution can be used to compute the NLS approximation

$$\varepsilon\psi_{NLS}(x,t) = \varepsilon A\big(\varepsilon(x - ct), \varepsilon^2 t\big)e^{i(kx-\omega t)} + \varepsilon A^*\big(\varepsilon(x - ct), \varepsilon^2 t\big)e^{-i(kx-\omega t)}, \tag{5.10}$$

which we want to compare with the numerical solution of the Klein-Gordon equation. Therefore we use the splitting method from chapter 4.3.1 and the implicit midpoint rule from chapter 4.3.2 and choose the initial values of (5.2) as follows. Taking into account the chain rule for computing $\partial_t\psi_{NLS}(x,t)\big|_{t=0}$ and the representation of $\partial_\sigma A$ given by (5.3), we can set the initial data as

$$\begin{aligned} u(x,0) &= u_0(x) \coloneqq \varepsilon\psi_{NLS(x,0)} \\ \partial_t u(x,0) &= u_1(x) \coloneqq \varepsilon\left[e^{ikx}\left(-i\omega A_0 - c\varepsilon\partial_\xi A_0 - i\nu_1\partial_\xi^2 A_0 - i\nu_2|A_0|^2 A_0\right) + c.c.\right], \end{aligned} \tag{5.11}$$

where $A_0 \in \{A_0^s, A_0^r\}$ is the initial value for the nonlinear Schrödinger equation.

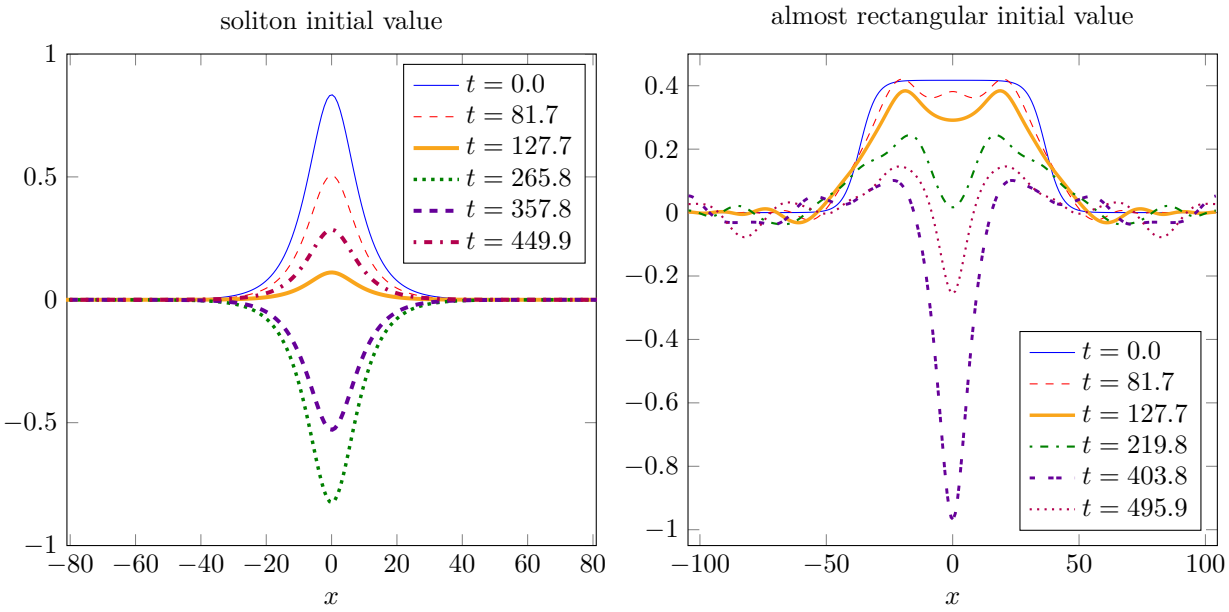


Figure 5.1. (RUN_schroed.m)
Solution of the Gross-Pitaevskii equation with a soliton initial value (left) and an almost rectangular initial value (right). We used $\varepsilon = 0.15$ and in order to get a reasonable illustration we have set $c = 0$. Having $c > 0$ the same evolution can be observed but each curve is spatially shifted by εct, due to the spatial scaling. Observe that the shape of the solution does not change over time for the soliton initial value A_0^s in contrast to the nonsoliton initial value A_0^r.

Parameter	Value	Parameter	Value
$(a_1 \quad a_2 \quad a_3)$	$(1 \quad -1 \quad 1)$	ν_1	$c^2 - a_1/(2\omega)$
L	$50\pi/k \approx 524$	ν_2	$-3a_3/(2\omega)$
k	0.3	d_1	$\sqrt{-2\gamma/\nu_2}$
ω	$\sqrt{a_1 k^2 - a_2}$	d_2	$\sqrt{-\gamma/\nu_1}$
c	k/ω	d_3	γ
γ	0.5	A_0	A_0^s, A_0^r

Table 5.1. Overview of the parameters used in our simulations

All the parameters that we have used in our simulations are listed in table 5.1. We repeated the simulation with the initial values corresponding to the soliton A_0^s as well as with the ones according to the "almost recatangular" envelope A_0^r for various $0 < \varepsilon < 1$ and for various $N \in \mathbb{N}$ and investigated the maximal error over all times in the sense of (5.4). In order to get an idea how the numerical solution develops over time we have a look at figure 5.2 and 5.3, where we see the numerical solution $\tilde{u}$ of the Klein-Gordon equation and the NLS approximation $\varepsilon\psi_{NLS}$. One finds that after some time the envelope of the NLS approximation still represents more are less the envelope of the numerical solution $\tilde{u}$ and that $\tilde{u}$ almost completely overlaps with $\varepsilon\psi_{NLS}$.But if we look closer we observe that $\tilde{u}$ has propagated a bit further than $\varepsilon\psi_{NLS}$, see figure 5.2 on the lower right. Therefore a small error between the numerical solution and the NLS approximation is already visible without rigorous error analysis.

Another interesting thing is the change of shape of the envelope in figure 5.3 for nonsoliton initial values, i.e. $A_0 = A_0^r$. The shape changes from an almost rectangular form to a completely different outline, but the numerical solution is still bordered by the Solution A of the Schrödinger equation, i.e. the envelope of the NLS approximation.

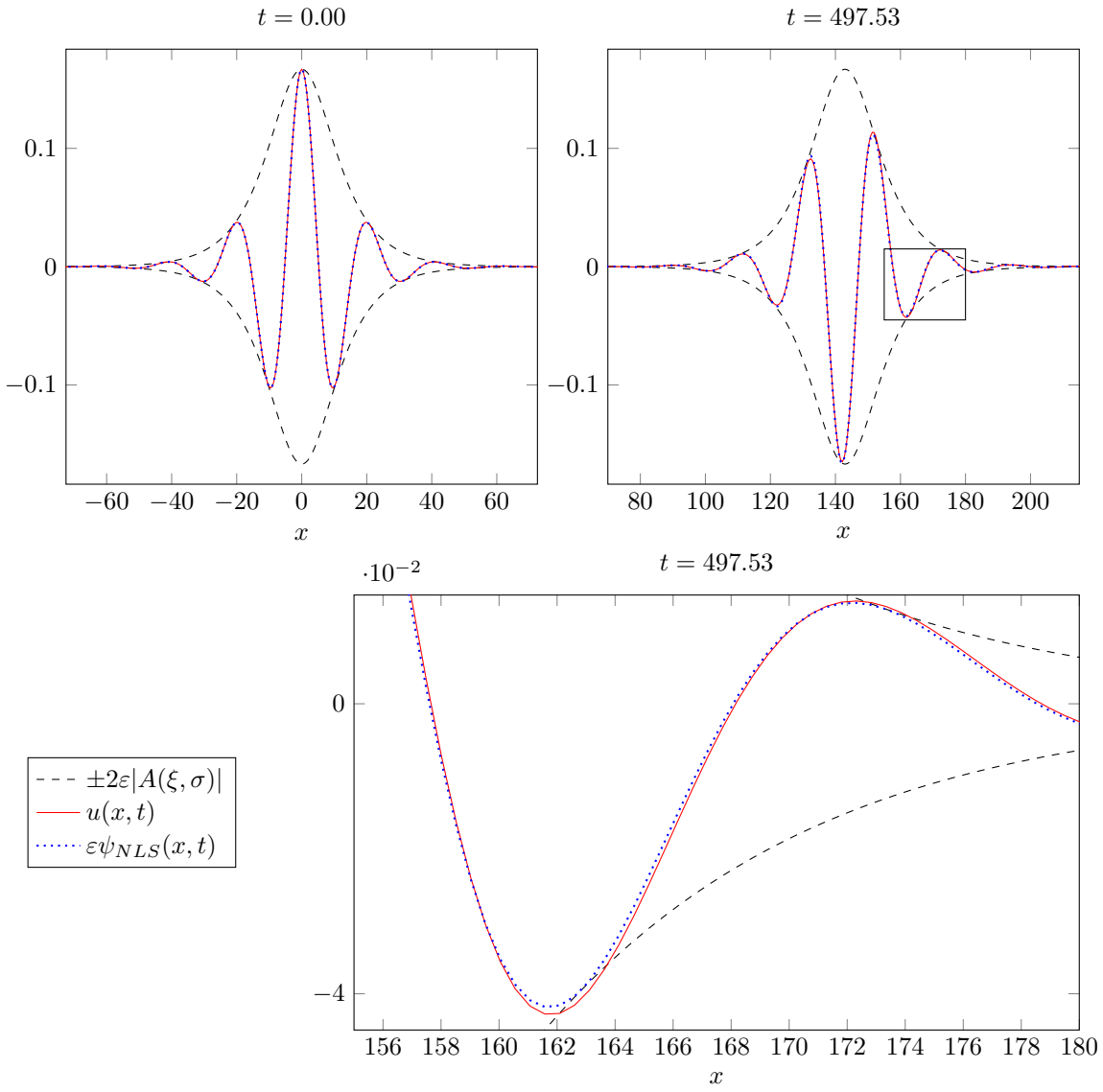


Figure 5.2. The initial pulse u_0 and the envelope given by A_0^s at $t = 0$ (upper left) and the numerical solution of the Klein Gordon equation for $\varepsilon = 0.1$ at $t = 497.53$ (upper right). The figure on the lower left shows the magnified aperture which is outlined by the rectangle in the figure on the upper right. The dashed black line depicts the shape of the envelope A, the blue dotted line shows the NLS approximation $\varepsilon\psi_{NLS}$.

The Error Bounds

In order to discuss the error of the NLS approximation we have a look at figure 5.4. It shows pretty well that the error

$$\sup_{m,\, t_m \in [0,5/\varepsilon^2]} \sup_{j=1,\ldots,N} |u(x_j, t_m) - \varepsilon\psi_{NLS}(x_j, t_m)| \tag{5.12}$$

of the NLS approximation compared to the numerical solution behaves like $C\varepsilon^2$, which fortifies statement (5.5). In our experiments we found $C \in (0.24, 0.28)$. The evolution of the l^∞-error in the interval $[0, 5\varepsilon^2]$ is presented in figure 5.5 for soliton type initial values and in figure 5.6 for "almost rectangular" type initial values. One can see that for both types of integrators and initial values and for each ε the error starts to

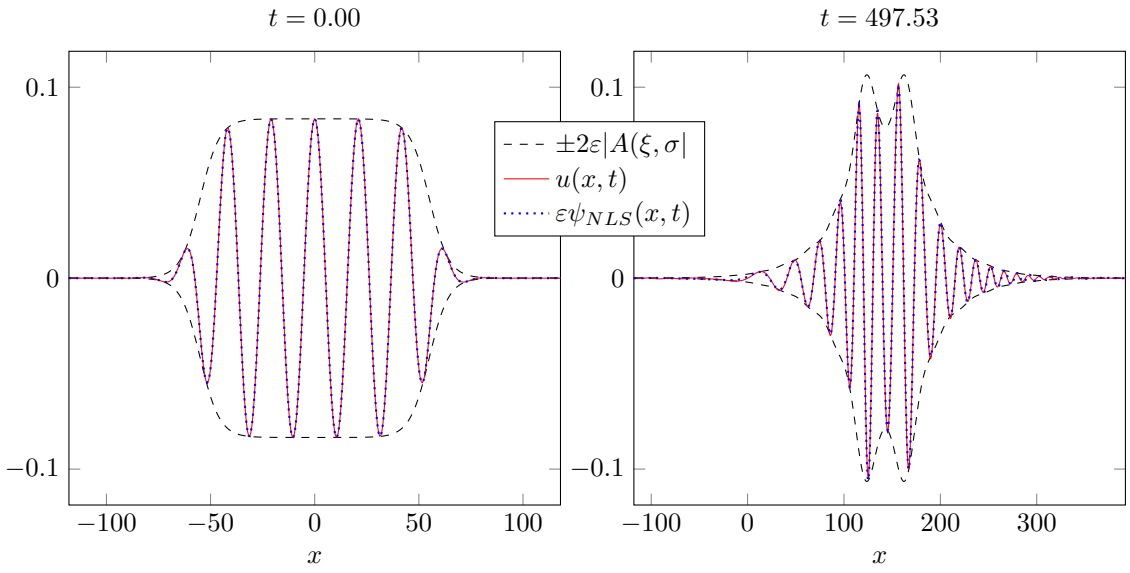


Figure 5.3. Evolution of the solution to the Klein-Gordon equation with almost rectangular initial values. Note that the shape of the envelope $2\varepsilon|A|$ changes over time. $\varepsilon = 0.1$.

grow linearly with time as $t \in \mathcal{O}(\varepsilon^2)$. This effect can be observed for the initial values corresponding to A_0^r even earlier.

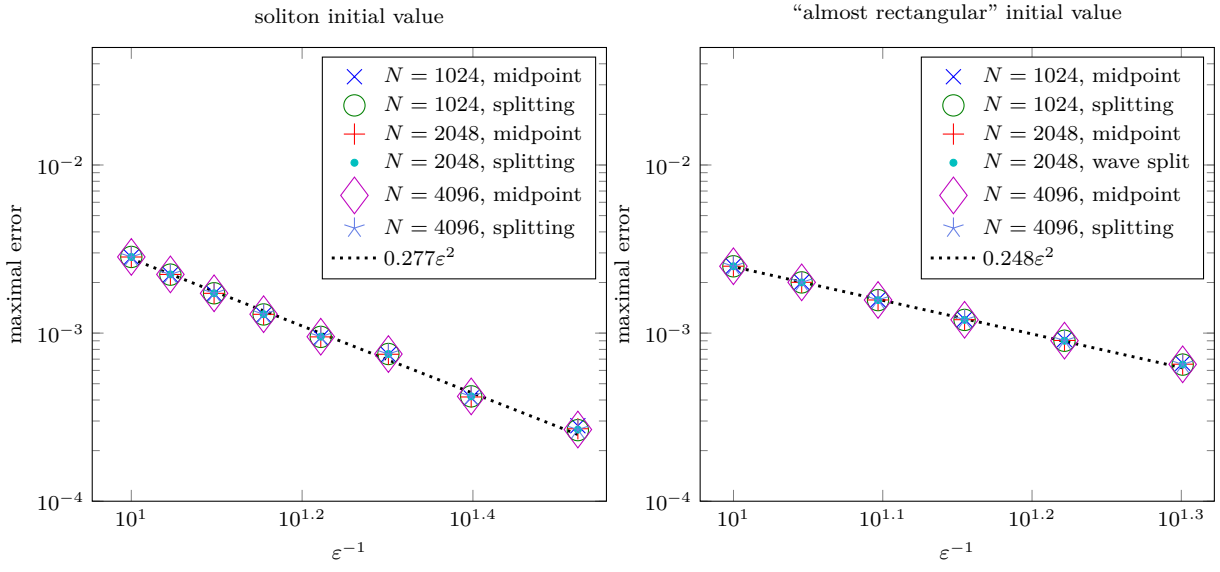


Figure 5.4. (PLOT_MAXERRORS.m)
Order plot of the maximal error $\sup_{m,\, t_m \in [0,5/\varepsilon^2]} \|u(\cdot,t_m) - \varepsilon\psi_{NLS}(\cdot,t_m)\|_{l^\infty}$ of the NLS approximation in terms of ε^{-1} using Strang splitting and the implicit midpoint rule. The error behaves like $C\varepsilon^2$ for various N, for both integrators and for both types of initial values.

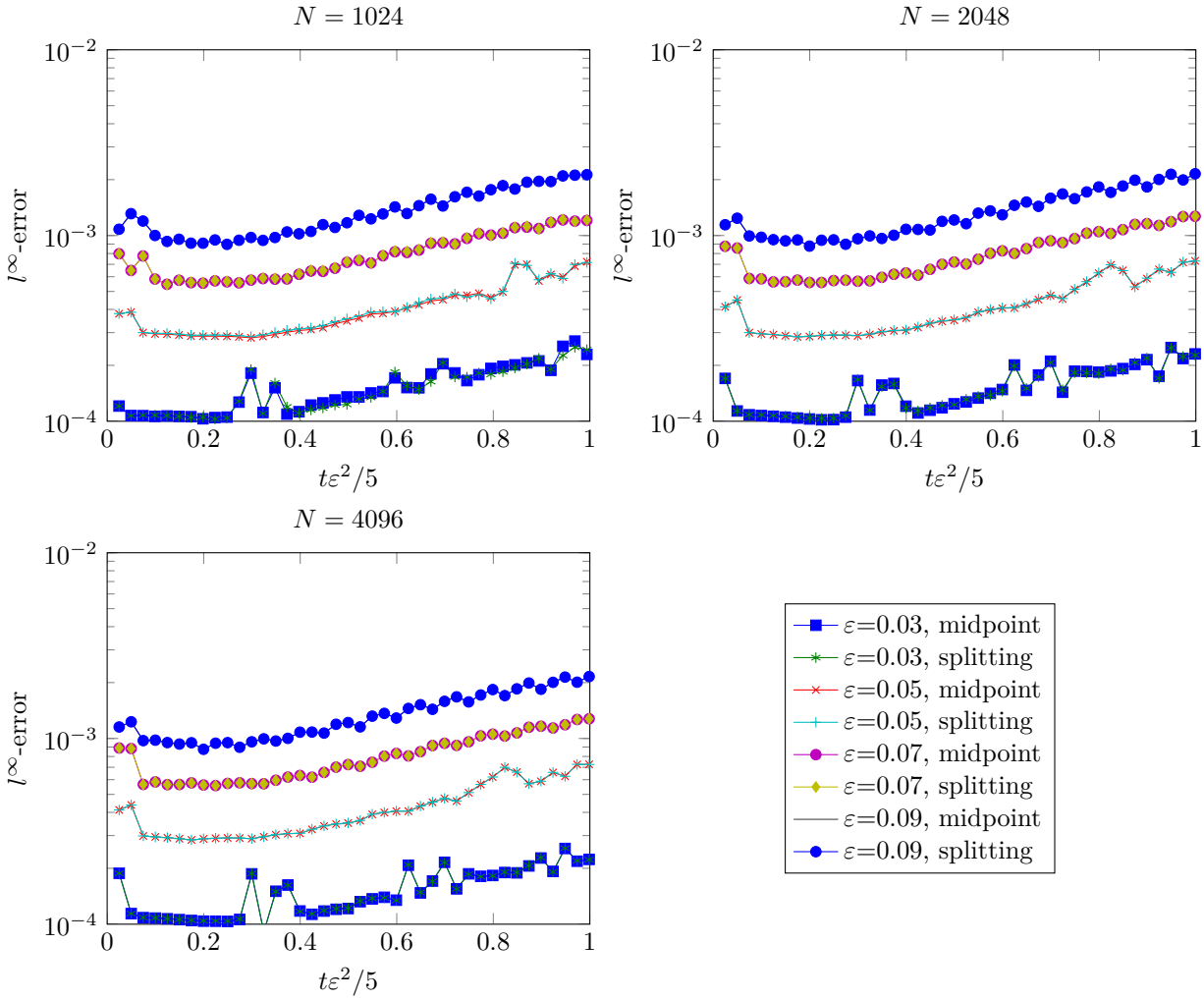


Figure 5.5. (PLOT_ERROR_EVOLUTION.m)
Evolution of the l^∞-error in terms of the "normalized" time $t\varepsilon^2/5$ for soliton type initial values.

The Reason why to use the NLS Approximation

We have already seen that the maximal error of the NLS approximation in the time interval $t \in [0, T/\varepsilon^2]$ is small for some $T > 0$ and small $\varepsilon > 0$. However, even if we reduce the step size for the Schrödinger splitting integrator and therefore the error of the amplitude A, we cannot improve the error of the NLS approximation for fixed $\varepsilon > 0$ more than up to order $\mathcal{O}(\varepsilon^2)$.

So why should we use such an approximation if we can obtain a more accurate solution to the Klein-Gordon equation by solving it directly with one of the integrators from chapter 4.3?

The answer is CPU time!

As we will see in section 5.3.3, among our three integrators for the Klein-Gordon equation, the splitting integrator is the most efficient one with respect to CPU time and accuracy. But it still needs a lot more CPU time to solve this equation in the time interval $[0, T/\varepsilon^2]$ with step size δ_t than the Schrödinger splitting integrator needs to solve the Gross-Pitavskii equation in the time interval $[0, T]$ with step size $\delta_\sigma = \varepsilon^2\delta_t$, see figure 5.7. Therefore if the order of the error of our NLS approximation is

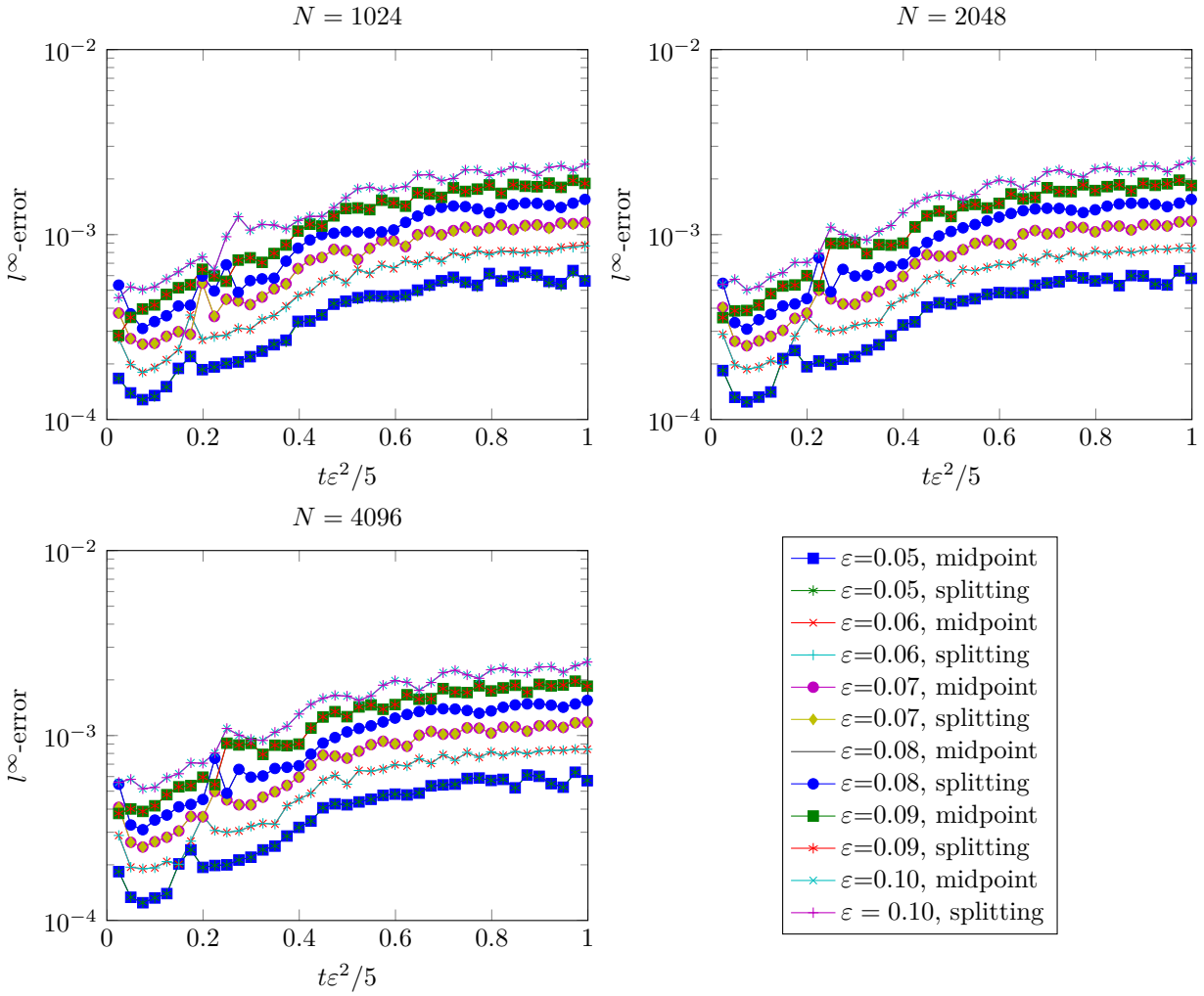


Figure 5.6. (PLOT_ERROR_EVOLUTION.m)
Evolution of the l^∞-error in terms of the "normalized" time $t\varepsilon^2/5$ for "almost rectangular" type initial values.

small enough for our purpose it makes sense to use it instead of computing a numerical solution of the Klein Gordon equation directly.

5.2. The Numerical Properties of the Schrödinger Splitting Integrator

Now let us investigate some properties of the solution to the Gross-Pitaevskii equation (5.3). Note that we did all the integration with respect to the slow variable $\sigma = \varepsilon^2 t$ and that the spatial scaling $\xi = \varepsilon(x - ct)$ causes a scaling of the L^2 norms by $\varepsilon^{1/2}$ in this section, i.e. for a function $f(\xi, \sigma)$ we have

$$\|f(\cdot, \sigma)\|_{L^2} = \varepsilon^{1/2} \|f(\varepsilon\cdot, \sigma)\|_{L^2} .$$

To a given time step size δ_t corresponding to the fast time scale t we define the slow time step size

$$\delta_\sigma := \varepsilon^2 \delta_t,$$

corresponding to the slow time scale $\sigma = \varepsilon^2 t$.

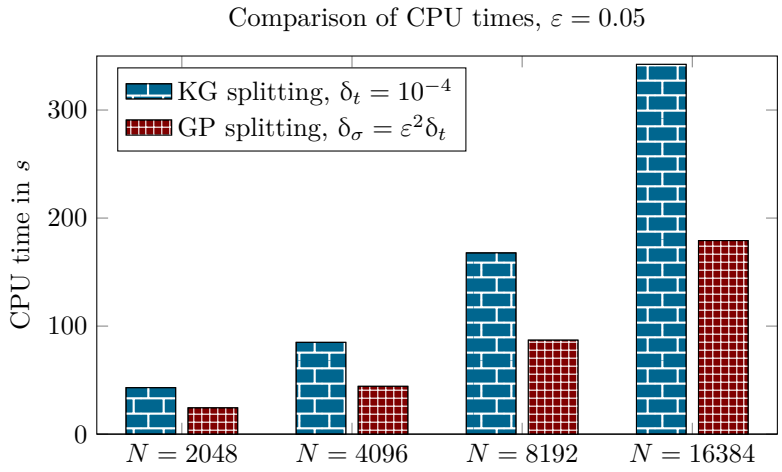


Figure 5.7. (RUN_SCHROED_ERRORS.m)
Comparison of the CPU times needed to solve the Klein-Gordon equation (KG splitting) in the time interval $t \in [0, 10]$ with step size $\delta_t = 10^{-4}$ and the Gross-Pitaevskii equation (GP splitting) in the time interval $\sigma \in [0, 10\varepsilon^2]$ with step size $\delta_\sigma = \varepsilon^2 \delta_t$ using the corresponding splitting solvers for various N. The GP splitting only needs about half the CPU time of the KG splitting. $\varepsilon = 0.05$, $A_0 = A_0^s$.

Firstly we regard the numerical order of the splitting integrator. In theorem 4.26 we have shown that for the time step size δ_σ the global error in the L^2 norm is bounded by $C\delta_\sigma^2$. We reproduced these results in our numerical tests for the initial value $A_0 = A_0^s$ and plotted the maximal error

$$\sup_{\sigma \in [0, 10000\varepsilon^2]} \|A(\sigma) - \psi_s(\sigma)\|_{L^2}$$

for various step sizes δ_σ in figure 5.8, where $A(\xi, \sigma)$ is the numerical and $\psi_s(\xi, \sigma)$ is the exact solution corresponding to the initial value $A_0 = A_0^s$.

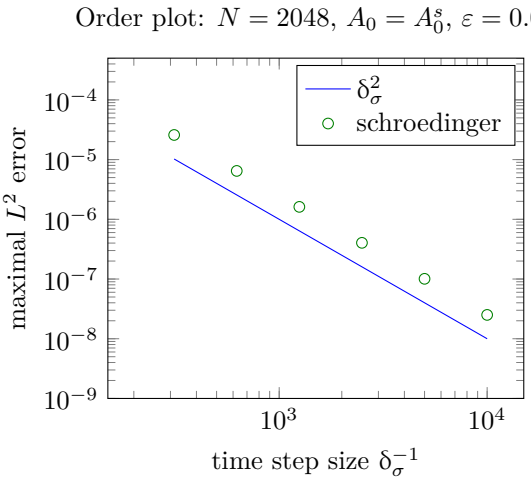


Figure 5.8. (RUN_SCHROED_ERRORS.m)
Order plot of the splitting method for the Gross-Pitavskii equation. We applied the method on the time interval $\sigma \in [0, \varepsilon^2 \cdot 10000]$. Reference solution is $\psi_s(\xi, \sigma)$ in (5.7). One can see that the results from theorem 4.26 are approved by the plot.

A very nice numerical result is the conservation of the L^2 norm and the energy, which is according to (3.19) given by

$$\mathcal{E}(A,\sigma) \coloneqq \frac{1}{2}\nu_1 \|A(\sigma)\|_{L^2}^2 + \frac{1}{4}\nu_2 \left\|A(\sigma)^2\right\|_{L^2}^2 . \tag{5.13}$$

Both the norm $\|A(\sigma)\|_{L^2}$and the energy $\mathcal{E}(A,\sigma)$ stay very close to the initial norm $\|A_0\|_{L^2}$ and energy $\mathcal{E}(A,0)$. The numerical results are depicted in figure 5.9.

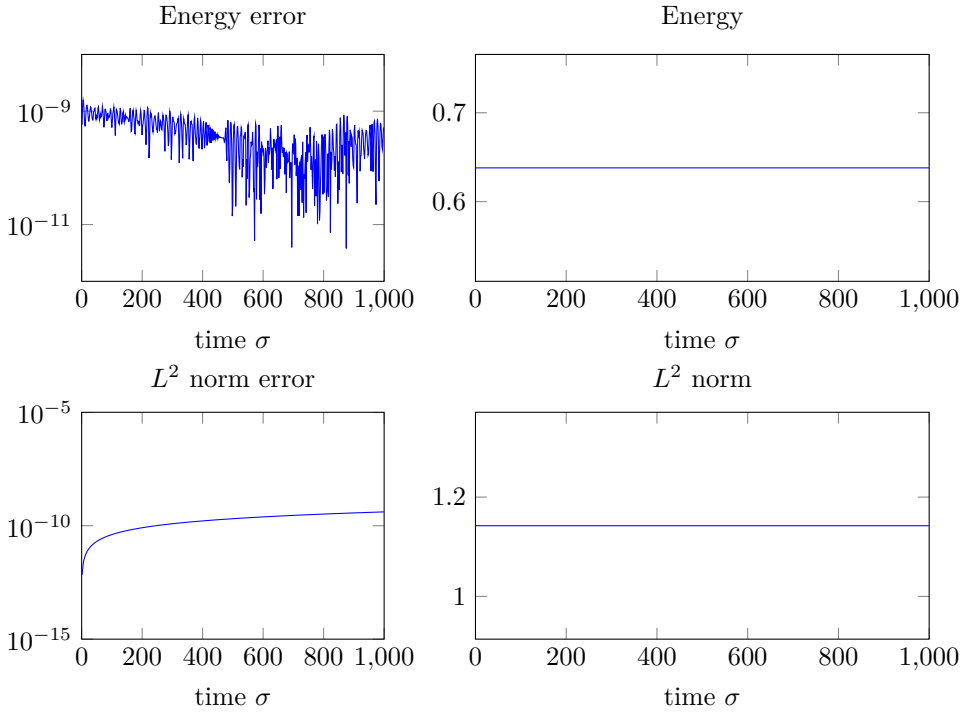


Figure 5.9. Energy and norm of the solution $A(\xi,\sigma)$ of the Gross-Pitaevskii equation and the corresponding errors using the splitting integrator in algorithm 4.1. Both the norm and energy are conserved by the splitting integrator. Parameters: $N = 2048$, $\varepsilon = 0.05$, $\delta_\sigma = 10^{-4}$, $A_0 = A_0^s$.

5.3. Numerical Properties of the Klein-Gordon Splitting Integrator and the Implicit Midpoint Rule

Since we want to make sure that the integrators that we used in section 5.1 to solve the Klein-Gordon equation provide an acceptable reference solution to compare the NLS approximation with, we will have a look at some numerical properties of the splitting integrator and the implicit midpoint rule in this section and compare them to the results by the leapfrog method. In this section we will denote the numerical solution of the Klein-Gordon equation (5.2) that we obtained by one of the three integrators from chapter 4.3 by $\tilde{u}$. If we do not explicitly mention something else, the initial values for the simulations are constructed by (5.11) with $A_0 = A_0^s$ and the parameters are taken from table 5.1. Furthermore we set the tolerance for the Newton iteration in algorithm 4.3 to $\text{tol} = 10^{-8}$. The results and all the figures of this section can be reproduced by running RUN_ORDERPLOT.m.

5.3.1. The Numerical Order

At first we investigate the numerical order of the methods. But since we do not know the exact solution to (5.2) we have to compute a reference solution by another numerical method. For ordinary

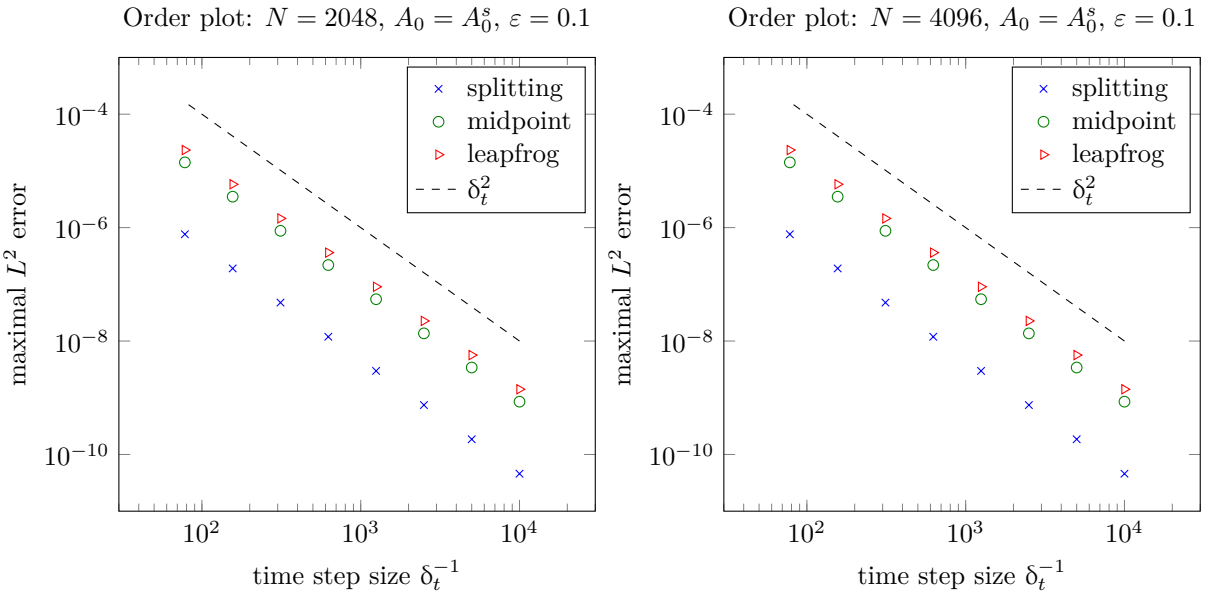


Figure 5.10. Order plot for the leapfrog and the splitting method and the implicit midpoint rule in the time interval $[0, 1]$. One can see that the splitting method works most precisely. All three integrators are of numerical order 2.

differential equations we know that all three integrators are of order 2, i.e. the global error is of order $\mathcal{O}(\delta_t^2)$, cf. [10]. But due to additional spatial discretization errors we can not be sure if our methods conserve this property if we apply them to partial differential equations such as our nonlinear wave equation.

Therefore the integrator which we want to use to compute the reference solution has to be at least of order 2 or higher. For our purpose we choose the 4th order integrator which was presented by Haruo Yoshida in [27] in 1990 and use the same splitting ansatz as we did in section 4.3.1.

For fixed N and therefore fixed spatial step size h we apply each integrator to the wave equation in the time interval $[0, 1]$ and compare the maximal errors in the L^2 norm. We normalize the initial values from (5.11) and set

$$\tilde{u}_0 = u_0 / \|u_0\|_{L^2} , \qquad \tilde{u}_1 = u_1 / \|u_0\|_{L^2} . \tag{5.14}$$

the new initial values. We repeated the calculation for various time step sizes δ_t and obtained the result

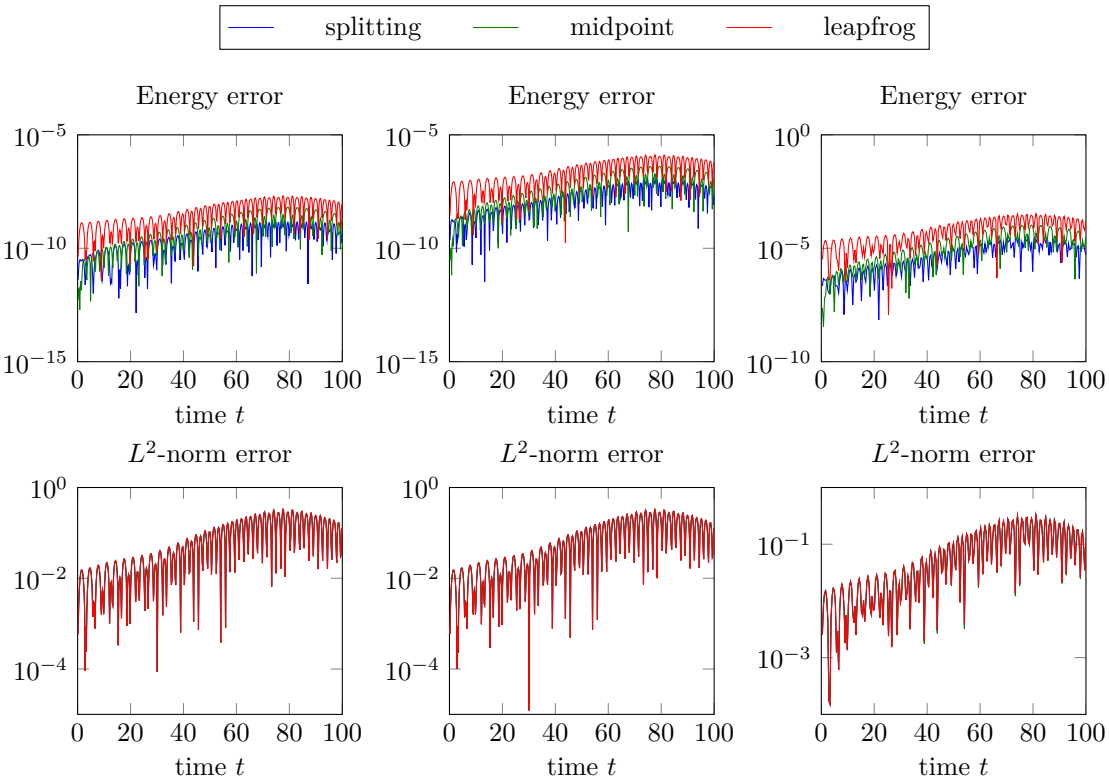
$$\|u_{ref}(t) - \tilde{u}(t)\|_{L^2} \leq C\delta_t^2, \qquad \forall t \in [0, 1] \tag{5.15}$$

with some constant $C > 0$, where $\tilde{u}$ is the numerical solution and u_{ref} is the reference solution, cf. figure 5.10. This numerically underlines the second order convergence of the splitting method and the implicit midpoint rule.

5.3.2. Energy and Norm Conservation

But in our problem it is not only important that a time integrator is very accurate but also that it conserves energy and the norm of our system over a long period of time, i.e. if we define the energy of system (5.2) as (cf. [9], chapter 5)

$$\mathcal{E}(u, t) \coloneqq \|\partial_t u(t)\|_{L^2}^2 + \|\partial_x u(t)\|_{L_2}^2 + \|u(t)\|_{L^2}^2 - \frac{1}{2} \left\|u^2(t)\right\|_{L^2}^2 \tag{5.16}$$



(a) $\delta_t = 2.0 \cdot 10^{-4}$ **(b)** $\delta_t = 1.6 \cdot 10^{-3}$ **(c)** $\delta_t = 2.56 \cdot 10^{-2}$

Figure 5.11. Error plot for the energy error $|\mathcal{E}(\tilde{u}, t) - \mathcal{E}(\tilde{u}, 0)|$ and the error of the L^2 norm $\left| \|\tilde{u}(\cdot, t)\|_{L^2} - \|\tilde{u}(\cdot, 0)\|_{L^2} \right|$ using various step sizes δ_t. For all three integrators the energy error stays small. The best energy conservation is observed using our spliting integrator. Regarding the L^2 norm error we can not see any difference between the integrators: for all three it stays small. Simulation parameters: $\varepsilon = 0.1$, $N = 4096$.

we want to have ideally

$$\begin{aligned} \|\tilde{u}(t)\|_{L^2} &= \|\tilde{u}(0)\|_{L^2}\,, \\ \mathcal{E}(\tilde{u}, t) &= \mathcal{E}(\tilde{u}, 0) \end{aligned} \qquad \forall t \in [0, T] \tag{5.17}$$

for some $T > 0$. Therefore we tested the integrators for these properties and found that even for large step sizes $\delta_t = 0.05$ the energy and the L^2 norm are conserved over a long period of time, see figure 5.12. For smaller step sizes we investigated the energy error

$$|\mathcal{E}(\tilde{u}, t) - \mathcal{E}(\tilde{u}, 0)|$$

and the L^2 norm error

$$|\|\tilde{u}(t)\|_{L^2} - \|\tilde{u}(0)\|_{L^2}|$$

with respect to the start time and found that they also stay small for a long time span, see figure 5.11. Our splitting integrator keeps the energy error smaller than the implicit midpoint rule or the leapfrog method. It also conserves the L^2 norm and the corresponding error in pretty much the same quality as using one of the other integrators, see figure 5.11. Hence our splitting method seems to be very suitable for our problem.

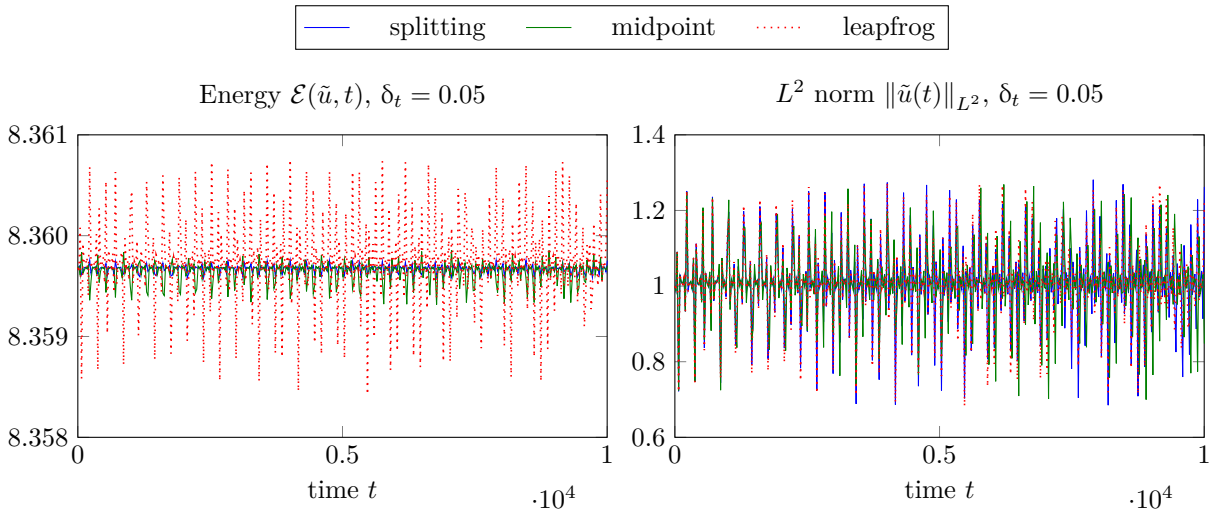


Figure 5.12. Evolution of the energy and the L^2 norm of $\tilde{u}$ in the time interval $[0, 10000]$ with a pretty large step size $\delta_t = 0.05$. All our integrators conserve the energy and the norm pretty well, no drift is visible. Note that the oscillation width of the energy is smallest for the splitting integrator. Simulation parameters: $\varepsilon = 0.1$, $N = 4096$.

5.3.3. Efficiency

Now we want to see for ourselves that our splitting integrator is not only very accurate but also very efficient compared to the leapfrog method or the implicit midpoint rule. Therefore we applied the three methods to the Klein Gordon equation (5.2) and compared the CPU times they needed to solve the problem on the time interval $[0, 2]$ for various time step sizes $\delta_t^{[j]}$ and the in the process evolving L^2 errors

$$\max_{t \in [0,2]} \|u_{ref}(t) - \tilde{u}(t)\|_{L^2}, \tag{5.18}$$

see figure 5.13. One can see that the splitting integrator produces a smaller error at less CPU time than the leapfrog method or the implicit midpoint rule.

By applying a power fit to the calculated errors with respect to the spent CPU time we calculated the following relation between the CPU time t and the corresponding error $r(t)$ with some constant c depending on the discretization and the method we used, i.e.

$$r(t) \approx ct^{-2}.$$

For $N = 8192$ we obtain the constant $c_s = 4.46 \cdot 10^{-8}$ using the splitting method and the constant $c_l = 3.34 \cdot 10^{-6}$ using the leapfrog method. Therefore for fixed error r this yields the relation

$$\frac{t_s}{t_l} = \sqrt{\frac{c_s}{c_l}} \approx 0.11$$

of the CPU time t_s by using the splitting method and the CPU time t_l by using the Leapfrog method. Hence the splitting method is about 9 times faster than the leapfrog method. Considering the discretization with $N = 4096$ gives the same result.

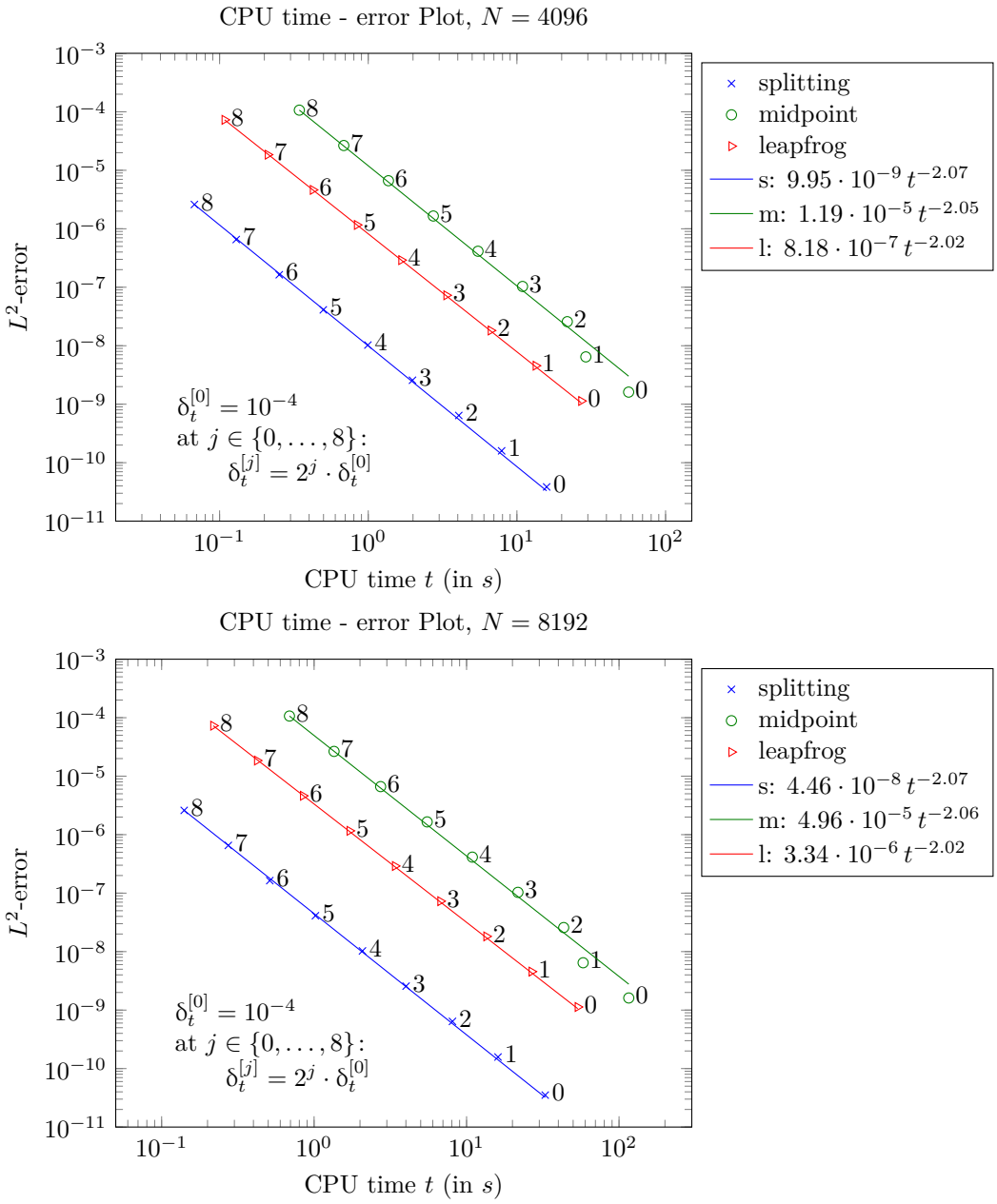


Figure 5.13. Maximal L^2 error against CPU time. Among our three methods, the splitting integrator is the most efficient one.

CONCLUSION

A Short Summary

We used the method of multiple scales to derive the so called NLS approximation to the solution of the nonlinear Klein-Gordon equation. Its amplitude function solves the Gross-Pitaevskii equation, a nonlinear Schrödinger equation that can be solved very efficiently using the splitting method presented in chapter 4.2.

We found that the error bounds on the NLS approximation depend on a small parameter ε. They are valid as long the time t is of order $\mathcal{O}(\varepsilon^{-2})$. We could underline this theoretical result of chapter 3 in chapter 5 numerically.

The main advantages of this approximation are its simple form and that we can obtain its amplitude very efficiently by the solution of a nonlinear Schrödinger equation using a splitting method. We need less time to compute this solution than by using a time integration method to solve the Klein-Gordon equation directly.

In order to get a reasonable reference solution of the Klein-Gordon equation, we analyzed amongst others a splitting integrator for this equation numerically. This splitting method has very nice properties, such as norm and energy conservation and a practical numerical order. Moreover it is very efficient.

Furthermore we learned how to use spectral methods based on the fast Fourier transform to calculate spatial derivatives and found that for smooth functions they produce a much smaller error than for example finite difference methods.

Future Work

We intend to do a rigorous error analysis for the Klein-Gordon splitting integrator in the near future. Furthermore we want to prove its conservation properties that we observed in this thesis.

A related problem to our Klein-Gordon equation is treated in [5], where additionally a quasilinear term is present. The ansatz for an approximation is also done by an NLS approximation of the same form as we derived during this thesis. The authors have been able to numerically state error bounds, which are of the same order as the error bounds for our NLS approximation. But a stringent error analysis for that problem remains to be done.

APPENDIX

A

FUNCTIONAL ANALYTICAL BACKGROUND

Let $U \subseteq \mathbb{R}^d$ be an open set and set

$$L^1_{loc}(U) \coloneqq \left\{ f : U \to \mathbb{C} \colon f \text{ is measurable, } f|_K \in L^1(K)\ \forall K \subseteq U \text{ compact} \right\}$$

and

$$C_c^\infty(U) \coloneqq \{\varphi \in C^\infty(U) \colon \varphi \text{ has compact support}\}.$$

Definition A.1 (Lebesgue spaces, see [1], chapter 2)**.** *For $n \in \mathbb{N}$, $p \in [1, \infty]$ we define the Banach spaces*

$$\begin{cases} L^p(\mathbb{R}^n) \coloneqq \{f : \mathbb{R}^n \to \mathbb{C}, \quad \|f\|_p < \infty\}, & p \in [1, \infty) \\ L^\infty(\mathbb{R}^n) \coloneqq \{f : \mathbb{R}^n \to \mathbb{C}, \quad \|f\|_\infty < \infty\}, & p = \infty \end{cases} \tag{A.1}$$

with the corresponding norms

$$\begin{cases} \|f\|_p \coloneqq \left(\int\limits_{\mathbb{R}^n} |f(x)|^p dx \right)^{\frac{1}{p}}, & p \in [1, \infty) \\ \|f\|_\infty \coloneqq \operatorname*{ess\,sup}\limits_{x \in \mathbb{R}^n} |f(x)|, & p = \infty, \end{cases}$$

where

$$\operatorname*{ess\,sup}_{x \in \mathbb{R}^n} |f(x)| \coloneqq \inf \{\alpha > 0;\ |f(x)| \leq \alpha \quad \text{for almost all} \quad x \in \mathbb{R}^n\}.$$

We may sometimes also write just L^p and L^∞ or $L^p(\mathbb{R}^n)$ and $L^\infty(\mathbb{R}^n)$ for the spaces (A.1).

A.1. The continuous Fourier transform

Let $J = (a, b) \subset \mathbb{R}$ be an interval.

Definition A.2. *$f : J \to \mathbb{C}$ is said to be **admissible** if the restriction $f\big|_K$ is jump continuous for every compact $K \subset J$.*

Definition A.3. *Let $f : J \to \mathbb{C}$ be admissible.*
*Then we call f **absolutely integrable** if the integral $\int\limits_J |f(x)|dx \in \mathbb{R}$ exists.*

Proposition A.4 (see [2], chapter VI, Proposition 8.7.)**.** *Let $f : J \to \mathbb{C}$ be absolutely integrable, then the integral $\int\limits_J f(x)dx$ exists in $\mathbb{C}$.*

Proof. See [2], Proposition 8.7. □

Theorem A.5 (majorant criterion,see [2], chapter VI, Theorem 8.8.)**.** *Let $f : J \to \mathbb{C}$ and $g : J \to \mathbb{R}^+$ be admissible with*

$$|f(x)| \leq g(x) \quad \text{for} \quad x \in J.$$

If g is integrable, then f is absolutely integrable.

Remark A.6. *The function $f : \mathbb{R} \to \mathbb{C}$ being absolutely integrable is equivalent to $f \in L^1(\mathbb{R})$.*

Remark A.7. *Theorem A.5 also holds if $J \subset \mathbb{R}^n$. Then the statement is: if g is integrable, then $f \in L^1(J)$.*

Definition and Lemma A.8. *Let $f \in L^1(\mathbb{R}^n, \mathbb{C})$, $k \in \mathbb{R}^n$. Then the function $x \mapsto e^{-i\langle k,x\rangle} f(x)$ is also in L^1 and the Fourier integral*

$$\mathcal{F}f(k) := \widehat{f}(k) := \int_{\mathbb{R}^n} e^{-i\langle k,x\rangle} f(x) dx \in \mathbb{C}$$

*exists. We call the operator $\mathcal{F} : f \mapsto \widehat{f}$ the (continuous) **Fourier transform** and $\widehat{f}$ is called the Fourier transformed of f.*

Proof. There holds $|e^{-i\langle k,x\rangle}| = 1 \; \forall x \in \mathbb{R}^\varkappa$. Set $g(x) := |e^{-i\langle k,x\rangle} f(x)|$. Then the claim folows directly from remark A.7 (majorant criterion for $\mathbb{R}^n$) and proposition A.4. □

Theorem A.9. *Let $f \in L^1(\mathbb{R}^n)$. Then $\mathcal{F}f \in C_0(\mathbb{R}^n) := \{f \in C(\mathbb{R}^n), f \to 0, |x| \to infty\}$. Furthermore is $\mathcal{F} : L^1(\mathbb{R}^n) \to C_0(\mathbb{R}^n)$ a linear continuous operator with operator norm $\|\mathcal{F}\| \leq 1$.*

Proof. See [26], Satz V.2.2. □

Definition A.10. *Let $\alpha := (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}_0^n$, f a smooth function. Then we define the differential operator D^α formally by*

$$D^\alpha f := \frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}} f.$$

*α is called a **multiindex**.*

Definition A.11 (see [26], Definition V.2.3)**.** *A function $f : \mathbb{R}^n \to \mathbb{C}$ is called **fast decreasing**, if*

$$\lim_{|x| \to \infty} x^\alpha f(x) = 0 \quad \forall \alpha \in \mathbb{N}_0^n,$$

where $x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$. The space

$$\mathscr{S}(\mathbb{R}^n) := \left\{ f \in C^\infty(\mathbb{R}^n) : \; D^\beta f \text{ fast decreasing } \forall \beta \in \mathbb{N}_0^n \right\}$$

*is called **Schwartz space** and its elements **Schwartz functions**.*

Lemma A.12 (see [26], Lemma V.2.4 and V.2.5.)**.** *Let $f \in \mathscr{S}(\mathbb{R}^n)$ and α a multiindex. Then there hold*

a) $\mathcal{F}f \in C^\infty(\mathbb{R}^n)$ and $D^\alpha(\mathcal{F}f) = (-i)^{|\alpha|}\mathcal{F}(x^\alpha f)$.
b) $\mathcal{F}(D^\alpha f)(k) = i^{|\alpha|} k^\alpha \mathcal{F}f$
c) $\mathcal{F}f \in \mathscr{S}(\mathbb{R}^n)$

Theorem A.13 (see [26], Satz V.2.8.)**.** *The Fourier transformation $\mathcal{F}$ is a bijection from $\mathscr{S}(\mathbb{R}^n)$ onto $\mathscr{S}(\mathbb{R}^n)$ and its inverse operator is given by*

$$(\mathcal{F}^{-1}f)(x) = \int_{\mathbb{R}^n} f(k)e^{i\langle x,k\rangle}dk.$$

Furthermore there holds

$$\langle \mathcal{F}f, \mathcal{F}g\rangle_{L^2} = \langle f, g\rangle_{L^2} \qquad \forall f, g \in \mathscr{S}(\mathbb{R}^n).$$

Hence $\mathcal{F}$ is an isometry with respect to the norm $\|\cdot\|_{L^2}$.

So the operator $\mathcal{F}$ is well-defined on the subset $\mathscr{S}(\mathbb{R}^n) \subset L^2(\mathbb{R}^n)$, bijective and isometric with respect to $\|\cdot\|_{L^2}$. One can show that $\mathscr{S}(\mathbb{R}^n)$ is dense in $L^2(\mathbb{R}^n)$ and thus we can extend $\mathcal{F}$ to an isometric operator $\mathcal{F}_2\colon L^2(\mathbb{R}^n) \to L^2(\mathbb{R}^n)$ that satifies

Theorem A.14 (see [26], Satz V.2.9.)**.** *a) For $f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n)$ there holds*

$$(\mathcal{F}_2 f)(k) = (\mathcal{F}f)(k) \qquad \text{almost everywhere}$$

b) Let $f \in L^2(\mathbb{R}^n)$, $B_R := \{x \in \mathbb{R}^n\colon |x| \le R\}$ and define

$$g_R(k) = \int_{B_R} f(x)e^{-i\langle x,k\rangle}dx,$$

then

$$\mathcal{F}_2 f = \lim_{R\to\infty} g_R, \qquad \text{w.r.t.} \quad \|\cdot\|_{L^2}.$$

A.2. Sobolev spaces and operator theory

Definition A.15 (weak derivatives and Sobolev spaces, see Schnaubelt [23], Definition 3.1)**.** *Let $f \in L^1_{loc}(U)$ and $\alpha \in \mathbb{N}_0^d$ a multiindex. If there is a function $g \in L^1_{loc}(U)$ such that*

$$\int_U g\varphi dx = (-1)^{|\alpha|}\int_U f\partial^\alpha\varphi dx, \quad \forall\varphi \in C_c^\infty(U),$$

then $g =: \partial^\alpha f$ is called weak derivative of f. If f possesses weak derivatives for all $|\alpha| \le k$, then we write $f \in W^k(U)$.
For $k \in \mathbb{N}, 1 \le p \le \infty$, we define the Sobolev spaces by

$$W_p^k(U) := \left\{f \in L^p(U) \cap W^k(U)\colon \partial^\alpha f \in L^p(U) \forall |\alpha| \le k\right\}$$

endowed with the norm

$$\|f\|_{k,p} = \begin{cases} \left(\sum\limits_{0\le|\alpha|\le k} \|\partial^\alpha f\|_p^p\right)^{\frac{1}{p}}, & 1 \le p < \infty, \\ \max\limits_{0\le|\alpha|\le k} \|\partial^\alpha f\|_\infty, & p = \infty, \end{cases}$$

where $\partial^0 f := f$ and $W_p^0(U) = L^p(U)$.
In the case of $p = 2$ we write

$$H^k(U) := W_2^k(U).$$

Remark A.16. *Instead of the norm $\|\cdot\|_{k,2}$ the space $H^k(U)$ may be endowed with the equivalent norm $|||\cdot|||_{H^k}$,*

$$|||f|||_{H^k} := \max_{0\le|\alpha|\le k} \|\partial^\alpha f\|_2$$

for $f \in H^k(U)$. The equivalence of these norms follows from

$$|||f|||_{H^k}^2 \leq \|f\|_{k,2}^2 = \sum_{0\leq|\alpha|\leq k} \|\partial^\alpha f\|_2^2 \leq C\,|||f|||_{H^k}^2.$$

Lemma A.17. *Let u, v, w be functions of one of the spaces $L^2(\mathbb{R})$, $H^1(\mathbb{R})$ or $H^2(\mathbb{R})$ such that the corresponding norms exist. Then there hold the following inequalities*

$$\begin{aligned}
\|uvw\|_{L^2} &\leq K_0\ \|u\|_{L^2}\,\|v\|_{H^1}\,\|w\|_{H^1}\,,\\
\|uvw\|_{L^2} &\leq K_1\ \|u\|_{L^2}\,\|v\|_{H^2}\,\|w\|_{H^2}\,,\\
\|uvw\|_{H^1} &\leq K_2\ \|u\|_{H^1}\|v\|_{H^1}\,\|w\|_{H^1}\,,\\
\|uvw\|_{H^2} &\leq K_3\ \|u\|_{H^2}\|v\|_{H^2}\,\|w\|_{H^2}\,.
\end{aligned}$$

with constants $K_0, K_1, K_2, K_3 > 0$.

Proof. The proof is essentially based on the Sobolev embedding $H^1(\mathbb{R}) \hookrightarrow L^\infty(\mathbb{R})$ and the estimates $\|\cdot\|_{L^2} \leq \|\cdot\|_{H^1} \leq \|\cdot\|_{H^2}$, such that from

$$\|uvw\|_{L^2} \leq \|u\|_{L^2}\,\|v\|_\infty\,\|w\|_\infty$$

there follow the first and the second estimates immediately. Having

$$\frac{\partial}{\partial x}(uvw) = (\frac{\partial}{\partial x}u)vw + u(\frac{\partial}{\partial x}v)w + uv(\frac{\partial}{\partial x}w)$$

and $\|\cdot\|_{H^1}^2 = \|\cdot\|_{L^2}^2 + \left\|\frac{\partial}{\partial x}\cdot\right\|_{L^2}^2$, we obtain

$$\|uvw\|_{H^1}^2 \leq \|uvw\|_{L^2}^2 + \left\|(\frac{\partial}{\partial x}u)vw\right\|_{L^2}^2 + \left\|u(\frac{\partial}{\partial x}v)w\right\|_{L^2}^2 + \left\|uv(\frac{\partial}{\partial x}w)\right\|_{L^2}^2.$$

Using the first inequality, where the L^2 norm of the right hand side is applied to the differentiated function respectively, finally yields the third inequality. The last inequality is shown in a similar way, using that $\|\cdot\|_{H^1} = \|\cdot\|_{L^2} + \left\|\frac{\partial}{\partial x}\cdot\right\|_{L^2} + \left\|\frac{\partial^2}{\partial x^2}\cdot\right\|_{L^2}$. □

Definition A.18. *Let X and Y be Hilbert spaces with scalar product $\langle\cdot,\cdot\rangle_X$ and $\langle\cdot,\cdot\rangle_Y$ respectively. Let $T \in \mathcal{L}(X,Y)$, i.e. linear and bounded, and $A : X \to Y$ a linear densely defined operator with domain $D(A) \subset X$. We define the Hilbert space adjoints T' and A' by*

$$\begin{aligned}
&\langle x, T'y\rangle_X = \langle Tx, y\rangle_Y \qquad \forall x \in X,\ y \in Y\\
&D(A') = \{y \in Y : \exists z \in X \forall x \in D(A) : \langle Ax, y\rangle_Y = \langle x, z\rangle_X\}, \quad A'y := z.
\end{aligned}$$

If $X = Y$, we call T self-adjoint if $T' = T$ and respectively A if $A' = A$, in particular $D(A) = D(A')$. T is called unitary if it is invertible with $T' = T^{-1}$.

Definition A.19 (operator semigroups, Werner[26], Definition VII.4.1)**.** *A strongly continuous semigroup of operators (C_0 semigroup) is a one parameter family $T(t) : X \to X$, $t \geq 0$ of bounded linear operators on a Banach space X with the following properties:*

(i) $T(0) = Id$,
(ii) $T(s+t) = T(s)T(t), \quad \forall s, t \geq 0$
(iii) $\lim_{t\to 0} T(t)x = x, \quad \forall x \in X.$

The operator defined by

$$Ax = \lim_{h\to 0} \frac{T(h)x - x}{h}, \qquad D(A) = \left\{ x \in X \mid \lim_{h\to 0} \frac{T(h)x - x}{h} \text{ exists} \right\}$$

is called the (infinitesimal) generator of $(T(t))_{t\geq 0}$.

Theorem A.20 (Stone's theorem, Pazy [19], Theorem 1.10.8)**.** *A is an infinitesimal generator of a C_0-semigroup of unitary operators on a Hilbert space H if and only if iA is self-adjoint.*

Theorem A.21 (Schnaubelt [23], Theorem 4.18)**.** *Let H be a Hilbert space and A be a densely defined, closed and symmetric operator, then*

$$A \quad \text{is self adjoint} \quad \Longleftrightarrow \quad \text{the spectrum} \quad \sigma(A) \subseteq \mathbb{R}$$

Lemma A.22. *The operator $L \coloneqq -i\,\Delta$ is generator of the C_0 contraction semigroup $\left(e^{-i\sigma\,\Delta}\right)_{\sigma\geq 0}$ in the space $H^m(\mathbb{R})$, $m \geq 1$.*

Proof. *We show that the operator $iL = \Delta$ is self-adjoint such that theorem A.21 and therefore Stone's theorem A.20 applies and we are done.*

(1) The Laplacian Δ is defined on the dense subset $C_c^\infty \subseteq H^m(\mathbb{R})$ of $L^2(\mathbb{R})$. Therefore iL is densely defined in $L^2(\mathbb{R})$.

(2) Furthermore iL is a closed operator with spectrum $\sigma(iL) = (-\infty, 0] \subseteq \mathbb{R}$ as shown in example 4.11 and 4.22 of the lecture notes [23].

(3) $\sigma(iL) = (-\infty, 0]$ also yields that $\|e^{\sigma L}\| \leq 1$ and therefore we have a contraction.

(4) Since

$$\langle \Delta u, \varphi \rangle_{L^2} = \int_{\mathbb{R}^d} \Delta u \overline{\varphi} dx \int_{\mathbb{R}^d} u \overline{\Delta \varphi} dx = \langle u, \Delta \varphi \rangle_{L^2} \quad \forall \varphi \in C_c^\infty(\mathbb{R}),$$

the operator iL is symmetric and all conditions of theorem A.21 are satisfied.

Hence $iL = \Delta$ is self-adjoint and Stone's theorem gives the result. □

Theorem A.23 (Pazy[19], Theorem 6.1.4)**.** *Let X be a Hilbert space and let $f\colon [0,\infty) \times X \to X$ be continuous in $t \geq 0$ and locally Lipschitz continuous in u, uniformly Lipschitz continuous in t on bounded intervals. If $-A$ is the infinitesimal generator of a C_0 semi-group $T(t)$ on X then for every $u_0 \in X$ there is a $t_{max} \leq \infty$ such that the initial value problem*

$$\frac{\partial}{\partial t} u(t) = -Au(t) + f\left(t, u(t)\right), \quad t \geq 0, \quad u(0) = u_0$$

has a unique solution in the form

$$u(t) = T(t)u_0 + \int_0^t T(t-s) f(s, u(s)) ds, \quad 0 \leq t \leq t_{max}.$$

Lemma A.24 (Gronwall's lemma in integral form)**.** *Let $T \in (0,\infty]$, $a, b \in L^\infty(0,T)$ and $\lambda \in L^1(0,T)$, $\lambda(t) \geq 0$ for almost all $T \in [0,T]$. Then,*

$$a(t) \leq b(t) + \int_0^t \lambda(s)a(s)ds \qquad \text{almost everywhere in } [0,T]$$

implies for almost all $t \in [0,T]$

$$a(t) \leq b(t) + \int_0^t e^{\Lambda(t)-\Lambda(s)}\lambda(s)a(s)ds,$$

where $\Lambda(t) \coloneqq \int_0^t \lambda(\tau)d\tau$. If $b \in W_1^1(0,T)$, it follows

$$a(t) \leq e^{\Lambda(t)}\left(b(0) + \int_0^t e^{-\Lambda(s)}b'(s)ds\right).$$

Moreover, if b is a monotonically increasing, continuous function, it holds

$$a(t) \leq e^{\Lambda(t)}b(t).$$

Proof. For the proof of this lemma, look inside Emmrich's paper [7] on Gronwall's lemma, Proposition 2.1. □

Theorem A.25 (Sobolev imbedding theorem,[1], Theorem 4.12)**.** *Let Ω be a domain in $\mathbb{R}^n$ and, for $1 \leq k \leq n$, let Ω_k be the intersection of Ω with a plane of dimension k in $\mathbb{R}^n$. (If $k = n$, then $\Omega_k = \Omega$). Let $j \geq 0$ and $m \geq 1$ be integers and let $1 \leq p < \infty$.*

If Ω satisfies the cone condition and if $mp > n$ or $m = n$ and $p = 1$, then

$$W_p^{j+m}(\Omega) \hookrightarrow C_B^j(\Omega),$$

where $C_B^j(\Omega) \coloneqq \left\{f \in C^j(\Omega) \middle| \left\|\frac{\partial^{|\alpha|}}{\partial x^\alpha}f\right\|_\infty < \infty,\ \forall |\alpha| \leq j\right\}$ Moreover, if $1 \leq k \leq n$, then

$$W_p^{j+m}(\Omega) \hookrightarrow W_q^j(\Omega_k), \qquad p \leq q \leq \infty$$

and in particular,

$$W_p^m(\Omega) \hookrightarrow L^q(\Omega), \qquad p \leq q \leq \infty.$$

BIBLIOGRAPHY

[1] Adams, R. A. & Fournier, J. J. F. (2003). *Sobolev spaces* (Second ed.)., volume 140 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam.

[2] Amann, H. & Escher, J. (2008). *Analysis. II*. Basel: Birkhäuser Verlag. Translated from the 1999 German original by Silvio Levy and Matthew Cargo.

[3] Amann, H. & Escher, J. (2009). *Analysis. III*. Basel: Birkhäuser Verlag. Translated from the 2001 German original by Silvio Levy and Matthew Cargo.

[4] Bender, C. M. & Orszag, S. A. (1978). *Advanced mathematical methods for scientists and engineers*. New York: McGraw-Hill Book Co. International Series in Pure and Applied Mathematics.

[5] Chong, C. & Schneider, G. (2012). *Numerical evidence for the validity of the NLS approximation in systems with a quasilinear quadratic nonlinearity*. University of Massachusetts Amherst, Universität Stuttgart.

[6] Dörfler, W., Lechleiter, A., Plum, M., Schneider, G., & Wieners, C. (2011). *Photonic Crystals: Mathematical Analysis and Numerical Approximation*, volume 42 of *Oberwolfach Seminars*.

[7] Emmrich, E. (July 1999). Discrete versions of gronwall's lemma and their application to the numerical analysis of parabolic problems. TU Berlin.

[8] Enns, R. H. (2011). *Its a Nonlinear World* (1 ed.). Springer Undergraduate Texts in Mathematics and TechnologySpringerLink : Bücher. New York, NY: Springer New York.

[9] Faou, E. & Schratz, K. (2013). Asymptotic preserving schemes for the klein-gordon equation in the non-relativistic limit regime. *To appear in Numerische Mathematik*, (To appear), 1–29.

[10] Hairer, E., Lubich, C., & Wanner, G. (2006). *Geometric numerical integration* (Second ed.)., volume 31 of *Springer Series in Computational Mathematics*. Berlin: Springer-Verlag. Structure-preserving algorithms for ordinary differential equations.

[11] Hochbruck, M. (2012). *Skriptum zur Vorlesung Numerik I-IV*. SS 2010 – WS 2011/12. Karlsruher Institut für Technologie: Arbeitsgruppe Numerik, Institut für Angewandte und Numerische Mathematik.

[12] Hundertmark, D., Meyries, M., Machinek, M., & Schnaubelt, R. (2013). Operator semigroups and dispersive equations: Lecture notes. http://www.math.kit.edu/iana3/~schnaubelt/media/isem16-skript.pdf, Karlsruher Institut für Technologie.

[13] Kirrmann, P., Schneider, G., & Mielke, A. (1992). The validity of modulation equations for extended systems with cubic nonlinearities. *Proc. Roy. Soc. Edinburgh Sect. A*, *122*(1-2), 85–91.

[14] Körner, T. W. (1988). *Fourier analysis*. Cambridge: Cambridge University Press.

[15] Lubich, C. (2008a). *From quantum to classical molecular dynamics: reduced models and numerical analysis*. Zurich Lectures in Advanced Mathematics. European Mathematical Society (EMS), Zürich.

[16] Lubich, C. (2008b). On splitting methods for Schrödinger-Poisson and cubic nonlinear Schrödinger equations. *Math. Comp.*, *77*(264), 2141–2153.

[17] Murdock, J. A. (1991). *Perturbations*. A Wiley-Interscience Publication. New York: John Wiley & Sons Inc. Theory and methods.

[18] Newell, A. C. (1985). *Solitons in mathematics and physics*, volume 48 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

[19] Pazy, A. (1983). *Semigroups of linear operators and applications to partial differential equations*, volume 44 of *Applied Mathematical Sciences*. New York: Springer-Verlag.

[20] Reichel, W. (WS 2009/2010). Analysis iii kurzskript. http://www.math.kit.edu/iana2/~reichel/media/ana3.pdf.

[21] Roessel, H. J. & Bowman, J. C. (2012). Math 538: Asymptotic methods. http://www.math.ualberta.ca/~bowman/m538/m538.pdf, University of Alberta, Edmonton, Canada.

[22] Romeel, D. (2011). Leapfrog integration. http://ursa.as.arizona.edu/~rad/phys305/ODE_III/node11.html.

[23] Schnaubelt, R. (2012). Lecture notes spectral theory. http://www.math.kit.edu/iana3/~schnaubelt/media/st-skript.pdf. Karlsruher Institut für Technologie.

[24] Thalhammer, M., Caliari, M., & Neuhauser, C. (2009). High-order time-splitting Hermite and Fourier spectral methods. *J. Comput. Phys.*, *228*(3), 822–832.

[25] Trefethen, L. N. (2000). *Spectral methods in MATLAB*, volume 10 of *Software, Environments, and Tools*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM).

[26] Werner, D. (2000). *Funktionalanalysis* (extended ed.). Berlin: Springer-Verlag.

[27] Yoshida, H. (1990). Construction of higher order symplectic integrators. *Phys. Lett. A*, *150*(5-7), 262–268.

ERKLÄRUNG

Hiermit erkläre ich, Patrick Krämer, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel und Quellen benutzt, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des Karlsruher Instituts für Technologie zur Sicherung guter wissenschaftlicher Praxis beachtet habe.

Patrick Krämer, Karlsruhe, den 6.12.2013.