

Time Domain Full Waveform Inversion Using ADI Modeling

Bernd Klimm

Vom Fachbereich Mathematik
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation.

Datum der Disputation: 01. July 2013

Erster Gutachter: Prof. Dr. Axel Klar
Zweiter Gutachter: Prof. Dr. Thomas Bohlen
(D 386)

Abstract

Constructing accurate earth models from seismic data is a challenging task. Traditional methods rely on ray based approximations of the wave equation and reach their limit in geologically complex areas. Full waveform inversion (FWI) on the other side seeks to minimize the misfit between modeled and observed data without such approximation.

While superior in accuracy, FWI uses a gradient based iterative scheme that makes it also very computationally expensive. To reduce the costs of our two dimensional time domain algorithm, we apply an alternating direction implicit (ADI) scheme to the acoustic wave equation. The ADI scheme can be seen as an intermediate between explicit and implicit finite difference modeling schemes. Being less computationally demanding than an implicit scheme it can handle coarser discretization compared to an explicit scheme and can be efficiently parallelized.

With the modeling in place, we test and compare several inverse schemes. To avoid local minima and improve speed of convergence, we use different hierarchical approaches, for example advancing over different bands of increasing frequency content. We can then demonstrate the effectiveness of the method on the Marmousi model and the 2004 BP model.

Acknowledgment

First of all, I gratefully acknowledge Prof. Dr. Axel Klar for his interest in this topic, his advice and continuous support during this thesis.

I would like to express my sincere appreciation and gratitude to Dr. Norman Ettrich for his excellent advice, support, guidance and friendship at the Fraunhofer Institute for Industrial Mathematics ITWM.

I acknowledge Fraunhofer ITWM for my PhD scholarship and in particular Dr. Franz-Josef Pfreundt and the HPC department for providing a great work environment as well as all the resources needed for this thesis.

Many thanks go to Statoil ASA and in particular to Frank Maaø and the seismic imaging group at the Statoil research center in Trondheim for enabling two internships in their group. I am thankful for their warm help and support as well as many instructive discussions.

I am thankful to my co-referee Prof. Dr. Thomas Bohlen for his interest in this work and the discussions with his group at the Geophysical Institute at Karlsruhe Institute of Technology.

I am indebted to Dr. Leo Neseman for his feedback and suggestions to the editing of this thesis.

I am thankful to Maxim Illyasov for providing the RTM software, that provided help in the initial development of the code created for this thesis.

I would like to extend my gratitude to all my colleagues at Fraunhofer ITWM, whose friendship and support I deeply enjoyed over the past three years.

Finally, I owe very special gratitude to my wife for her unconditional support and continuous encouragement.

Contents

1	Introduction	7
1.1	The Seismic Problem	7
1.2	Imaging and Inversion	9
1.2.1	Role of Full Waveform Inversion	10
1.2.2	Probabilistic vs. Deterministic	10
1.2.3	Overview of FWI Algorithm in this Thesis	11
1.3	Outline	11
2	Wave Equation and Inverse Problem	12
2.1	The Forward Problem	13
2.1.1	2D Problem with Constant Velocity	13
2.1.2	Inhomogeneous Velocity in 1D	15
2.2	The Inverse Problem	17
2.2.1	Existence of a Solution in 1 D	17
2.2.2	Derivation of the Gradient	20
3	Modeling	23
3.1	Alternating Direction Method	23
3.1.1	Dispersion	25
3.2	Boundary Conditions	25
3.2.1	Boundary Conditions in 1D	25
3.2.2	Boundary Conditions in 2D	26
3.3	Validation of the Code	27
3.4	Comparison with Explicit Finite Difference Code	29
3.4.1	The Marmousi Model	29
3.4.2	Comparison of the Results	30
3.5	Checkpointing	33
4	Nonlinear Minimization	35
4.1	Descent Methods	36
4.1.1	Starting Model	37
4.1.2	Descent Direction	37
4.1.3	Computing the Step Size	40
4.1.4	Stopping Criteria	40
4.2	Improving Convergence	41
4.2.1	Gradient in Water	41
4.2.2	Alternative Error Functions	41
4.3	Multiscale Approaches	44
4.3.1	Multiscale Approach over Frequencies	44
4.3.2	Multiscale Approach over Time Damping	45

Contents

4.3.3	Multiscale Approach over Offsets	46
5	Test Cases	48
5.1	Marmousi	48
5.1.1	First Gradient	48
5.1.2	General Descent Scheme	50
5.1.3	L_1 Norm	57
5.1.4	Noise	59
5.1.5	Multiscale Schemes	62
5.2	2004 BP Model	73
6	Summary and Conclusion	75
	Bibliography	78

1 Introduction

Finding natural resources is an exceedingly challenging task which is becoming more and more difficult as resources are getting depleted. Oil & gas companies are drilling deeper and in geologically more complex environments than before and costs of drilling often exceed one million dollar per well. On the other side this means that there are high incentives in trying to obtain accurate models of the area of interest in order to better plan for the drilling and production and to reduce the risk of unfortunate dry wells. The process of generating these models is still a challenging task, that usually involves many disciplines from geology and geophysics to mathematics. In this thesis, we study full waveform inversion (FWI), which currently is the most advanced inversion algorithm to obtain earth properties from seismic data as part of the whole exploration process.

1.1 The Seismic Problem

The earth propagates vibrations over long distances. We observe this naturally during earthquakes, where surface waves are transmitted on continental scales. For oil and gas exploration the earth is excited artificially by seismic experiments and vibrations are measured locally. A seismic source generates vibrations (e.g., pressure waves) and receivers record the reflected data at the surface. The principle is similar to an ultrasound unit, although at much larger scales.

In marine exploration a seismic vessel creates acoustic waves inside the water with an array of airguns. The vessel tows a long array of hydrophones some distance behind the source. Hydrophones are basically special microphones in water that measure pressure and record the reflected signal at the surface as displayed in Figure 1.1. One line of hydrophones is called a streamer. The vessel then travels forward with this configuration and activates the source (i.e., shoots) in regular intervals (e.g., every 25 m). That means that when we speak of a seismic experiment, seismic acquisition, or survey, we actually mean multiple single shot experiments carried out in sequence. With the different shot and receiver positions, we observe reflections from the underground under multiple angles.

The source is designed to (ideally) produce a wavelet, a short, essentially compact signal that often has a peak frequency around or below 50 Hz. The distance between the source and the receivers is called offset. In modern acquisitions the streamer lengths are usually around 8 km or more with receivers spaced as close as every 25 m. The data is usually recorded for more than 8 seconds. Furthermore, a seismic vessel can tow several streamers simultaneously which then cover part of the 2D surface. All in all a single seismic survey can thereby produce terabytes of data. There are more complex acquisitions that include multiple seismic vessels in one survey or ocean bottom cable acquisitions, where receivers are laid out on the ocean floor.

1. Introduction

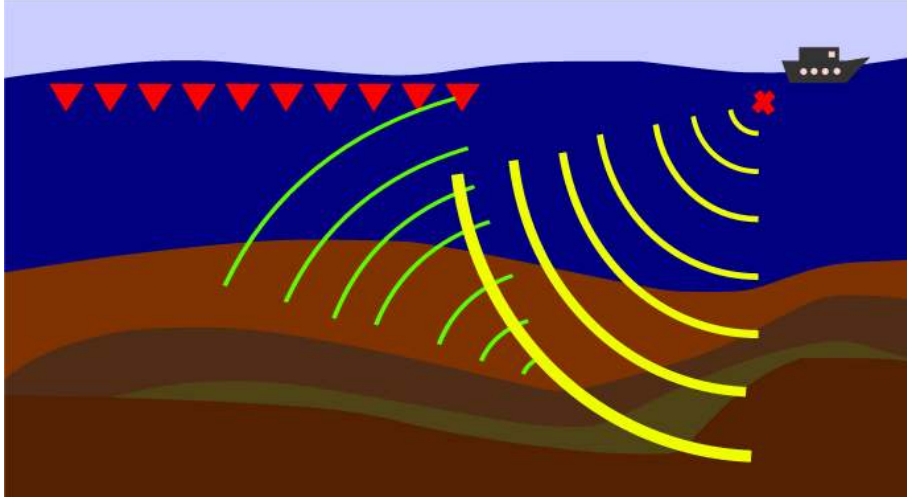


Figure 1.1: The marine seismic experiment

For this thesis, we have to limit the scope of our investigation. We will assume that we have a single two-dimensional section, that means data from a single streamer on a straight line.

The principals of land acquisitions are quite similar to marine acquisitions. The receivers (geophones) are laid out on the ground and a moving seismic truck generates vibrations. However, the geometry of the acquisition is often more complex as source and receivers need to follow the natural terrain. In this work, we do not deal with specifics of that. Furthermore, marine acquisitions are in some sense easier as the properties in the waterlayer are often well known or can easily be determined. This has some advantage in the computation as we see later. Nevertheless, apart from some rather technical differences, the general scheme and many observations we make with the marine experiment in this work could be transferred to the land survey as well.

The next step after the data acquisition is to use the data to obtain an image of the underground. For this we need to make assumptions towards the correct physical equations to use. Most general, wave propagation in media can be described with the full elastodynamic wave equation

$$\left(\frac{1}{\rho} \frac{\partial}{\partial i} C_{ijkl} \frac{\partial}{\partial j} - \delta_{kl} \frac{\partial^2}{\partial t^2} \right) u_l = \frac{1}{\rho} F_k \quad (1.1)$$

where δ is the Kronecker delta, ρ the density, F_k are body forces, and C is the stiffness tensor with up to 21 independent variables. Unfortunately, there are not nearly enough data available to solve the full problem. Even if there were, the required compute power would be completely out of range today. Thus we need to approximate the earth with a less detailed model. Research is done with the isotropic wave equation and even with certain kinds of anisotropy (e.g., transverse isotropy) but for the scope of this work we reduce the complexity even more and consider only the acoustic case with constant density

$$\Delta u - \frac{1}{c^2} \frac{d^2}{dt^2} u = f. \quad (1.2)$$

1. Introduction

Here, f is the source term, u denotes the scalar pressure in the medium, and c is the velocity or acoustic wave speed (i.e., speed of sound). In light of the computational complexity of full waveform inversion and because we are testing general properties of the inversion scheme, it seems reasonable for us to use just one parameter c .

There are several introductory books available that explain the seismic acquisition process in depth and also discuss possible effects of elasticity and anisotropy in greater detail, for example [Yil01], [Rob10], [GV10], [Jon10]. Note, that we also exclude the important step of data preprocessing from this work. Steps like noise removal, data regularization, filtering, sorting, removal of direct wave and so on are commonly performed before inversion and imaging tools are used. Please also refer to the cited introductory books for details.

1.2 Imaging and Inversion

Seismic data imaging is traditionally separated into two parts: velocity model building and migration. If a velocity model is available, migration means that a reflectivity model is generated based on the seismic data and that velocity model. Of course, a correct velocity model would already display the reflectors, however building a velocity model is much more difficult. So the common workflow is to start with a good, approximate velocity model and then use this model to migrate the seismic data to the “correct” position in the ground. The reflectivity model is then used for interpretation. The process is called imaging or migration. There are many industry-standard codes available, many of them use a ray approximation of the wave equation. Most accurate among the migration algorithms is reverse time migration (RTM), which uses the full wave equation to compute the reflectivity and is closely related to full waveform inversion, as we will see in chapter 2.2.2. An overview over different migration algorithms can be found in [Rob10].

Velocity model building, the focus of this work, is the more challenging part and is done iteratively. First, velocity models might be created based on geological reasoning and stacking analysis. Then tomography techniques are often applied, that are commonly based on ray approximations of the wave equation. Methods using the wave equation like full waveform inversion, which we investigate here, are most computational expensive and might only be used for complex parts of the whole problem, or when the simpler methods do not yield sufficiently accurate results. Please refer to [Jon10] for an overview of common velocity model building tools.

The whole process of seismic data processing normally encompasses multiple cycles of model building, imaging, quality control and interpretation and can take several months. What methods are employed and what amount of detail is required is ultimately a management decision. Common questions that guide the process might be: “Given certain amount of time and computational resources, what is the best image/ model that we can obtain?” or “What do we have to do in order to get an image/ model that can be used to make a decision (e.g., is it safe to drill)?”

1.2.1 Role of Full Waveform Inversion

In this whole process, the tomography based methods are very efficient for simple geological targets, for which the ray approximation holds, but reach their limits in complex situations, for example around salt domes, fold and thrust belts, and foothills [VO09]. In those cases even the separated process of building a velocity model and subsequential migration might no longer be applicable as even high quality imaging methods with an insufficient detailed velocity model might suffer accuracy. In those cases our full waveform inversion can lead to improvements.

Full waveform inversion goes back to Lailly in 1983 [Lai83] and Tarantola in 1984 [Tar84]. The underlying mathematical principles of least square minimization and adjoint state method are even older. Nevertheless it has not gained much attention until very recently because of its computational requirements and it is not in routine productive use up to now. Inversion schemes that use raybased approaches or combine wave equation modeling with ray theory were developed (e.g., [CC88], [SD92], [BCSJ01]) as well as schemes based on finite difference modeling (e.g., [KCL86], [IDT88], [CPN⁺90], [PDT90]) and finite element modeling [CMS08]. A second reason, why full waveform inversion was not considered successful in the past was because of the lack of long offsets in the seismic data. To start the inversion for a first (low frequency) background velocity model, diving waves are needed (i.e., waves that are not reflected, but slowly bended back upwards, when c increases with depth). To record diving waves with considerable depth, long receiver offsets are needed (or crosshole data measured from within boreholes). This data is now available in modern surveys and can be exploited by full waveform inversion [Mor87], [Mor88], [PW90], [PSWW96].

1.2.2 Probabilistic vs. Deterministic

Although modern explorations cover wider offsets and might use more shots and longer receiver data, the amount of data is always limited and the inverse problem of recovering the earth c does not have a unique solution. It is possible to use probabilistic inverse theory, where a probability function $p(c)$ is defined over the space of velocity models. The probability measures both the misfit between computed and measured data $U(c)$ and U_d and contains a priori knowledge about possible velocity models [Tar05]. However, this probability function p is not known for real world problems and could only be approximated by sampling the model space and solving the wave equation for each sample. Unfortunately the number of samples required grows exponentially with the size of the model space [Fic11]. For the seismic problem the velocity model can contain a million parameters or more, which makes this method infeasible.

Instead, we have to use deterministic inverse theory in our work. There, we define a misfit function E which will depend on the computed and observed data. We will then choose a starting model c_0 , being (presumably) close to the real model \tilde{c} , and start an iterative scheme with the goal of finding a c under which $E(c)$ becomes minimal.

1.2.3 Overview of FWI Algorithm in this Thesis

We will introduce full waveform inversion with the 2D acoustic constant-density wave equation. That is we use the wave equation 1.2 in two dimensions with a two dimensional velocity model c . We do the modeling in time domain with an alternating direction implicit (ADI) scheme [FM65], [Lee62]. Full implicit schemes are usually not applicable for full waveform inversion, since the linear equations that need to be solved are too large. Instead, explicit schemes are usually employed. These schemes however, require finer grid sizes and/ or time step sizes to achieve the same accuracy. Our ADI scheme lies in between. It solves multiple lowerdimensional systems that can be solved efficiently. Though it is still more compute intensive than an explicit scheme, it promises more accuracy on the same grid space.

With this modeling in place, we tackle the inverse problem and focus on different methods to avoid local minima. Our basic approach follows the multiscale techniques first introduced to full waveform inversion by Banks et al. [BSZC95]. We will investigate both the effects of different error functions and of different multiscale schemes on the inversion result.

1.3 Outline

The remainder of this thesis is organized as follows. In Chapter 2 we review the theoretical foundations. We develop a simple reference solution and show existence of a solution for the forward and the inverse problem in 1D. We then apply Lagrange Multiplier methods to derive the gradient, which we will need in the inversion in Chapter 4.

Chapter 3 is then devoted to the modeling. We introduce the ADI method and discuss boundary conditions. We then experimentally verify our modeling with the help of analytical solutions and compare the code on a realistic test model with an explicit finite difference code.

In Chapter 4 we discuss the gradient based inversion and different optimization schemes that should improve convergence and/ or reduce the risk of reaching local minima.

Finally we test the schemes from Chapter 4 in Chapter 5.

2 Wave Equation and Inverse Problem

In this chapter we want to define the exact problem and establish the theoretical foundations.

Let $\Omega \subset \mathbb{R}^n$ be the domain of interest ($n = 1, 2, 3$ in practice) and $T > 0$ the maximum observation time. Let $\Gamma_s \subset \Omega$ be the set of seismic source locations with source function $f : \Gamma_s \times [0, T] \rightarrow \mathbb{R}$. $c : \Omega \rightarrow \mathbb{R}$ shall denote the velocity model. Then the pressure $u : \Omega \times [0, T] \rightarrow \mathbb{R}$ is the solution of the wave equation

$$\begin{aligned} \Delta u(s) - \frac{1}{c^2} \frac{d^2}{dt^2} u(s) &= f(s) && \text{in } \Omega \times (0, T) \\ u(s) = \frac{d}{dt} u(s) &= 0 && \text{on } \Omega \times 0 \end{aligned} \tag{2.1}$$

for every source position $s \in \Gamma_s$. For every source, there is a set of receiver positions $\Gamma_{r,s}$ that record the shot from position s in the seismic experiment. That means the function u is measured in the receiver points $\Gamma_{r,s}$ and we write $U = u|_{\Gamma_{r,s}}$ for the receiver data. Note that it can sometimes be assumed that sources and receivers are at the surface, thus $\Gamma_s, \Gamma_{r,s} \subset \partial\Omega$, however, in sea acquisitions sources and receivers are usually a few meters below the water surface. Modeling, or solving the forward problem, means in our context, that for given c and f we want to compute U by solving the initial value problem (2.1).

Our final goal is to solve the inverse problem. Ideally that means that for a given data $U = U_d$ and source function f we want to obtain a velocity c satisfying Equation (2.1). This might never be possible in any real experiment, instead we can set out for minimizing the distance between U_d and our solution $U(c)$. Measuring in the L^2 norm, this means we want to find a c that minimizes

$$E(c) = \int_{\Gamma_s} \int_{\Gamma_{r,s}} (U_d - U(c))^2. \tag{2.2}$$

In practice, Γ_s and Γ_r are usually finite sets, for which we can replace the integrals in 2.2 with finite sums. To simplify notation, we will define

$$\Gamma = \Gamma_s \times \Gamma_{r,s} \tag{2.3}$$

so that we can write

$$E(c) = \int_{\Gamma} (U_d - U(c))^2. \tag{2.4}$$

There are other options for defining an misfit function E as well, which we will discuss in Chapter 4. Note that the source function f is not always known and sometimes has to be determined as well as part of the inversion.

2.1 The Forward Problem

2.1.1 2D Problem with Constant Velocity

For constant $c > 0$ and simple domain geometries, the Equation (2.1) can be solved analytically. With the help of this solution, some simple problems with other velocity profiles (e.g., one reflector) can be constructed and we can use the solution to check numerical solutions later. We briefly state the results for 2D here.

Let us first consider the homogeneous initial-value problem for $\Omega = \mathbb{R}^2$

$$\begin{aligned} \Delta u - \frac{1}{c^2} \frac{d^2}{dt^2} u &= 0 & \text{in } \Omega \times (0, \infty) \\ u &= u_0, \quad \frac{d}{dt} u = u_1 & \text{on } \Omega \times \{0\} \end{aligned} \quad (2.5)$$

Then we can state the following theorem.

Theorem 1 (Dzuik, [Dzu10], Ch. 6). *Let $u_0 \in C^2(\mathbb{R}^2)$, $u_1 \in C^1(\mathbb{R}^2)$. Then there exists a unique solution of (2.5) given by*

$$u(x, t) = \frac{1}{2\pi c} \int_{|y| \leq ct} \frac{u_1(x+y)}{\sqrt{c^2 t^2 - |y|^2}} dS(y) + \frac{d}{dt} \frac{1}{2\pi c} \int_{|y| \leq ct} \frac{u_0(x+y)}{\sqrt{c^2 t^2 - |y|^2}} dS(y). \quad (2.6)$$

We use this theorem to compute the nonhomogeneous problem with zero-initial conditions. The solution

$$\begin{aligned} \Delta u - \frac{1}{c^2} \frac{d^2}{dt^2} u &= f & \text{in } \Omega \times (0, \infty) \\ u &= 0 & \text{on } \Gamma \\ \frac{d}{dt} u &= 0 & \text{on } \Gamma \end{aligned} \quad (2.7)$$

for a given source function f can be developed from the solution of the homogeneous problem. Let $u(x, t; s)$ be the solution of

$$\begin{aligned} \Delta u(\cdot, s) - \frac{1}{c^2} \frac{d^2}{dt^2} u(\cdot, s) &= 0 & \text{in } \Omega \times (0, \infty) \\ u(\cdot; s) = 0, \quad \frac{d}{dt} u(\cdot; s) &= f(\cdot, s) & \text{on } \Omega \times \{t = s\} \end{aligned} \quad (2.8)$$

Now, we obtain the solution of (2.7) in Ω by applying Duhamel's principle (e.g., [Eva98])

$$u(x, t) = \int_0^t u(x, t; s) ds. \quad (2.9)$$

With theorem 1 we compute

$$u(x, t; s) = \frac{1}{2\pi c} \int_{|y| \leq c(t-s)} \frac{f(x+y, s)}{\sqrt{c^2(t-s)^2 - |y|^2}} dS(y). \quad (2.10)$$

2. Wave Equation and Inverse Problem

Now we introduce a point source at $x = (0, 0)$ of the form $f(x, s) = \delta(x)g(s)$ and obtain

$$u(x, t; s) = \frac{1}{2\pi c} \frac{g(s)}{\sqrt{c^2(t-s)^2 - |x|^2}} \Theta(c(t-s) - |x|) \quad (2.11)$$

where

$$\Theta(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases}$$

denotes the Heaviside function. We insert this into Equation (2.9), perform some basic variable transformations and obtain

Theorem 2. *Let $f(x, t) = \delta(x)g(t)$ be a single point source for $g \in C^1(0, \infty)$, then Equation (2.7) has the solution*

$$u(x, t) = \begin{cases} \frac{1}{2\pi} \int_{|x|}^{ct} \frac{g(t - s/c)}{\sqrt{s^2 - |x|^2}} ds, & |x| < ct \\ 0, & x \geq ct. \end{cases} \quad (2.12)$$

For a source modeled by a Ricker wavelet

$$g(t) = (1 - 2\pi^2 f_p^2 t^2) e^{-\pi^2 f_p^2 t^2} \quad (2.13)$$

with peak frequency f_p we can see the solution in Figure 2.1.

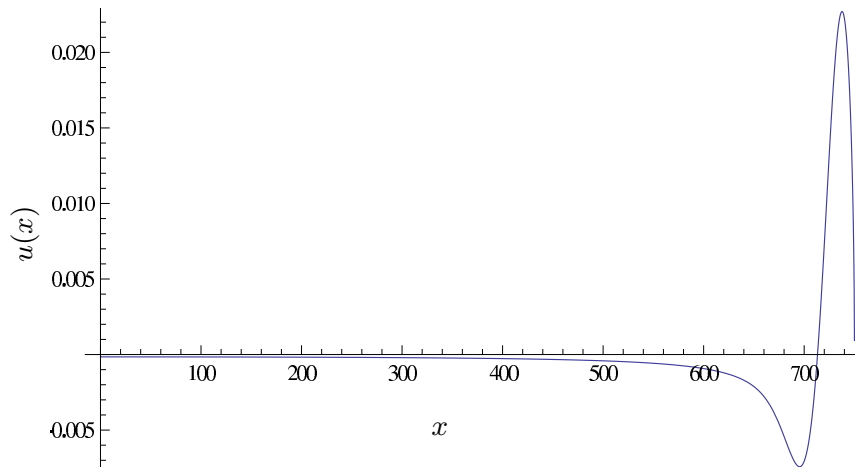


Figure 2.1: Analytical solution of Equation (2.12) for $t = 0.5$, $c = 1500$, $f_p = 15$

Differences between 2D and 3D

From theorem 1 we see already that Huygens principle is not valid in 2D. Unlike in 3D (or higher odd dimensional spaces) the solution at a point x at time t is determined by integrals over the complete ball $|x - ct| > 0$ and not only over the surface of the sphere $|x - ct| = 0$. That also means, that unlike in 3D a discontinuous source does not excite a wave in a discontinuous way. We can see that, when we

2. Wave Equation and Inverse Problem

compute the wavefront of a discontinuous signal: We compute the wavefront at $|x| = ct$ for $g(x) = \Theta(|x|)$:

$$\begin{aligned}
 \lim_{|x| \rightarrow ct, |x| < ct} u(x, t) &= \lim_{|x| \rightarrow ct, |x| < ct} \frac{1}{2\pi} \int_x^{ct} \frac{1}{\sqrt{s^2 - |x|^2}} ds \\
 &= \lim_{|x| \rightarrow ct, |x| < ct} \frac{1}{2\pi} \log \left(\frac{ct + \sqrt{c^2 t^2 - |x|^2}}{|x|} \right) \\
 &= 0.
 \end{aligned} \tag{2.14}$$

The general solutions for three and more dimensions, as well as the difference between even and odd dimensional spaces are discussed in depth in [Eva98].

When using 2D algorithms on 3D data a conversion is necessary. Of course, any such conversion must be an approximation. Crase et al. suggest a scaling U with \sqrt{t} and convolute with $\frac{1}{\sqrt{t}}$. For details we refer to [CPN⁺90] and [Rob05].

2.1.2 Inhomogeneous Velocity in 1D

We cannot state compact formulas for arbitrary velocity models, but nevertheless it is possible to prove the existence of a solution of the forward problem, as well as of the inverse problem (Chapter 2.2) in 1D.

Let us consider the wave equation in one dimension

$$\begin{aligned}
 \frac{\partial^2 u}{\partial z^2} - \frac{1}{c^2(z)} \frac{\partial^2 u}{\partial t^2} &= 0 & \forall t > 0, \forall z > 0 \\
 -\frac{\partial u}{\partial t}(0, t) &= f(t) & \forall t > 0 \\
 u(z, 0) = \frac{\partial u}{\partial t}(z, 0) &= 0 & \forall z > 0
 \end{aligned} \tag{2.15}$$

with a source function $f(t)$ placed on the boundary at $z = 0$. This is only a slight modification of the problem above, where the sources could be below the top boundary, that is in the water. (However, for constant water velocity in 1D a lower source could be replaced by an equivalent source on the boundary.)

We now follow the work of Bamberger et al. [BCL79] and give a variational formulation of the problem. Let

$$T > 0, \Omega = (0, Z) \tag{2.16}$$

be the maximum recording time and corresponding spacial domain. We also assume a maximum depth

$$Z > 0, \tag{2.17}$$

where we can introduce final conditions. We will see in the next section (Proposition 4) that this is not a real restriction. Let further

$$0 < c_- \leq c_+ \tag{2.18}$$

2. Wave Equation and Inverse Problem

be given. As in [BCL79], we define

$$\Sigma^Z = \{c \in L^\infty(0, Z) | c_- \leq c(z) \leq c_+ \text{ a.e. on } (0, Z)\} \quad (2.19)$$

to be the set of bounded velocity profiles, defined up to depth Z and

$$c \in \Sigma^Z \quad (2.20)$$

shall be given. Let us now introduce two spaces

$$V = \{v \in H^1(\Omega) | v(Z) = 0\} \quad (2.21)$$

with the scalar product

$$(u, v)_V = \int_{\Omega} \frac{du}{dz} \frac{dv}{dz} dz \quad \forall u, v \in V \quad (2.22)$$

and

$$H = L^2(\Omega) \quad (2.23)$$

with the scalar product

$$(u, v) = \int_{\Omega} uv dz. \quad \forall u, v \in H \quad (2.24)$$

$\|\cdot\|_V$ and $|\cdot|$ shall denote the corresponding norms. We may identify H with its dual, V to a part of H and therefore the dual $H' = H$ to a part of V' . Thus we have the dense inclusions and continuous injections

$$V \subset H \subset V'. \quad (2.25)$$

We further associate with c the scalar product

$$(u, v)_c = \int_{\Omega} \frac{1}{c^2(z)} u(z)v(z) dz \quad \forall u, v \in H. \quad (2.26)$$

If the source f satisfies

$$f \in W^{1,1}(0, T) \quad (2.27)$$

we can define the linear operator $L_f \in L^2((0, T); H)$ with

$$(L_f(t), v) = f(t)v(0) \text{ a.e. on } (0, T) \quad \forall v \in V. \quad (2.28)$$

Now, the variational formulation of system (2.15) can be stated as

$$\begin{aligned} u &\in L^2((0, T); V) & \frac{du}{dt} &\in L^2((0, T); H) \\ \frac{d^2}{dt^2}(u(t), v)_c + ((u(t), v)) &= (L_f(t), v) & \forall v \in V \\ u(0) = \frac{du}{dt}(0) &= 0 \end{aligned} \quad (2.29)$$

We can state the existence of the solution of the forward problem in the following theorem.

2. Wave Equation and Inverse Problem

Theorem 3 (Bamberger et al., [BCL79], Th. 1). *Given the definitions and hypotheses of (2.16) to (2.28), the system (2.29) has a unique solution*

$$\begin{aligned} u &\in L^\infty((0, T); V) \\ \frac{du}{dt} &\in L^\infty((0, T); H) \end{aligned} \tag{2.30}$$

and the mapping

$$f : W^{1,1}(0, T) \rightarrow L^\infty((0, T); V) \times L^\infty((0, T); H), \quad f \mapsto \left(u, \frac{du}{dt} \right) \tag{2.31}$$

is linear continuous.

Proof. We only restate the basic idea behind the proof of Bamberger et al. The injection of V into H is compact, therefore we can find an ascending sequence of subspaces V_m generated from orthogonal eigenfunctions with respect to $(\cdot, \cdot)_c$. Over each subspace V_m we can then construct a solution $u_m \in L^2((0, T); V_m)$ of Equation (2.29) restricted to V_m . Next, we have to show that the energy function

$$I_m(t) = \frac{1}{2} \left(\left| \frac{du_m}{dt} \right|_c^2 + \|u_m\|^2 \right) \tag{2.32}$$

is bounded and thus u_m and $\frac{du_m}{dt}$ are bounded in $L^\infty((0, T); V)$ and $L^\infty((0, T); H)$ respectively. This leads finally to the conclusion, that the limit can be passed and that both $u_m(t) \rightarrow u(t)$ and $\frac{du_m}{dt}(t) \rightarrow \frac{du}{dt}(t)$ converge strongly in their spaces V and H , respectively. We refer to [BCL79] for further details. \square

The requirements of Theorem 3 do not pose problems in practice. A real world velocity profile will always be bounded by a reasonable minimum and maximum velocity and most common source wavelets like the Ricker wavelet we use fulfill much higher regularity than required for f in Theorem 3.

2.2 The Inverse Problem

2.2.1 Existence of a Solution in 1 D

We will now continue with the 1 D study started in the previous section. After showing that for a given c , Equation (2.29) has a unique solution, we now want to check that, for given surface data $U_d(t)$, measured at the surface $z = 0$ we can find a suitable velocity c . More precisely: Let

$$U(c, t) = u(0, t) \tag{2.33}$$

be the synthetic surface signal for a velocity c , where u is the solution of (2.15). Then we want to solve the inverse problem of minimizing the misfit of observed and synthetic data in the L^2 norm, that is find $c \in \Sigma$ such that

$$E(c) = \int_0^T (U(c, t) - U_d(t))^2 dt \rightarrow \min \tag{2.34}$$

2. Wave Equation and Inverse Problem

where Σ denotes a space of admissible velocity models that satisfies

$$\Sigma \subset \Sigma_b = \{c \in L^\infty(\mathbb{R}^+) | c_- \leq c(z) \leq c_+ \text{ a.e. on } \mathbb{R}^+\}. \quad (2.35)$$

As before, we constrain the choice of c with lower and upper bounds $0 < c_- \leq c_+$, however, we will not require continuity or differentiability for c . In the previous section, we already introduced one possible subspace of Σ_b

$$\Sigma^Z = \{c \in L^\infty(0, Z) | c_- \leq c(z) \leq c_+ \text{ a.e. on } (0, Z)\} \quad (2.36)$$

that contains the velocity profiles, defined up to a certain depth Z . We introduce a second subspace

$$\Sigma^X = \{c \in L^\infty(0, Z(X)) | c_- \leq c(z) \leq c_+ \text{ a.e. on } (0, X)\} \quad (2.37)$$

where $Z(X)$ is the depth, that for the velocity c corresponds to X via

$$\int_0^{Z(X)} \frac{dz}{c(z)} = X. \quad (2.38)$$

(Cf. Bamberger et al. [BCL79].) X is called the travel time. It is the time needed for a wave in medium with wavespeed c to travel downwards to depth Z , which can be checked with the method of characteristics (e.g., [Eva98]). That also means that the surface response $U(c, t)$ for $t \leq T$ can only depend upon velocities in

$$\Sigma^X \quad \text{with } X \leq \frac{T}{2}. \quad (2.39)$$

Furthermore it offers us an easy way to add a second boundary condition to the state Equation (2.15) without harm:

Proposition 4. *For a given c , the response $U(c, t)$ of system (2.15) does not change over the interval $(0, T)$, if we introduce a supplementary boundary condition*

$$u(Z, t) = 0 \quad \forall t > 0 \quad (2.40)$$

as long as

$$\int_0^Z \frac{dz}{c(z)} \geq \frac{T}{2}. \quad (2.41)$$

Since, for a given traveltime X , $Z(X)$ depends on the velocity c , there cannot be a one to one correspondence between Σ^X and Σ^Z . The reasons mentioned above motivate Bamberger et al. ([BCL79]) to prove the existence of the solution of the inverse problem on Σ^X :

Theorem 5 (Bamberger et al., [BCL79], Th. 6). *Let $T > 0$, the length of the recorded seismogram U_d , be given. $X = \frac{T}{2}$ shall denote the corresponding travel time. Let the source function f satisfy*

$$\begin{aligned} f(0) = f'(0) = 0 \\ f \in W^{3,1}(0, T), \end{aligned} \quad (2.42)$$

then the inverse problem of finding $c \in \Sigma^X$ with

$$E(c) = \int_0^T (U(c; t) - U_d(t))^2 dt \rightarrow \min \quad (2.43)$$

has at least one solution.

2. Wave Equation and Inverse Problem

Proof. Again, we will only state the general idea of the proof from Bamberger et al. and refer to their work ([BCL79]) for the details. Although the final proof is done with the space Σ^X , we can naturally define a distance function on Σ^Z which makes E Lipschitz-continuous, but at the same time is weak enough to make Σ^Z compact. So, we start by defining a distance function

$$d^Z(c_1, c_2) = \max_{s \in [0, Z]} \left| \int_0^s \frac{1}{c_1(z)^2} - \frac{1}{c_2(z)^2} dz \right|. \quad (2.44)$$

for all $c_1, c_2 \in \Sigma^Z$. According to [BCL79], Property 6, this is a distance function and (Σ^Z, d^Z) is compact. We can then show that E is Lipschitz-continuous with

$$\|U(c_1) - U(c_2)\|_{C^0([0, T])} \leq C \|f\|_{W^{3,1}(0, T)} d^Z(c_1, c_2), \quad (2.45)$$

for all $c_1, c_2 \in \Sigma^Z$, where C depends only on c_- , Z and T ([BCL79], Th. 4). Compactness and Lipschitz-continuity guarantees that a minimum exists in Σ^Z . We now define a "pseudo norm" on Σ^X . For $c_1, c_2 \in \Sigma^X$, let Z_1, Z_2 be the corresponding depths with

$$X_i = \int_0^{Z_i} \frac{1}{c(z)} dz \quad (2.46)$$

for $i = 1, 2$. Then we define

$$d^X(c_1, c_2) = \max_{s \in [0, \min\{Z_1, Z_2\}]} \left| \int_0^s \frac{1}{c_1(z)^2} - \frac{1}{c_2(z)^2} dz \right|. \quad (2.47)$$

d^X satisfies the two properties of a norm:

$$d^X(c_1, c_2) \geq 0, \quad d^X(c_1, c_2) = 0 \Rightarrow c_1 = c_2 \quad (2.48)$$

and

$$d^X(c_1, c_2) = d^X(c_2, c_1). \quad (2.49)$$

for every $c_1, c_2 \in \Sigma^X$. It is not a norm since it does not satisfy the triangular inequality, however we can show that any d^X converging sequence has a unique limit in Σ^X ([BCL79], Property 10). Furthermore, Σ^X with the "pseudo-distance" d^X is compact in the following sense: Let (c_n) be a sequence in Σ^X , then there exists a c and a subsequence c_{n_k} so that

$$d^X(c_{n_k}, c) \rightarrow 0 \text{ for } n_k \rightarrow \infty. \quad (2.50)$$

This is also shown in [BCL79], Property 10, where the previous results for (Σ^Z, d^Z) are used in the proof. With d^X we can also show Lipschitz continuity

$$\|U_1(c_1) - U_2(c_2)\|_{C^0([0, T])} \leq C \|f\|_{W^{3,1}(0, T)} d^X(c_1, c_2). \quad (2.51)$$

for all $c_1, c_2 \in \Sigma^X$, where the constant C depends only on c_- and T ([BCL79], Th. 5). Finally, this estimate and the "compactness" of (Σ^X, d^X) are sufficient to show that a minimum exists in Equation (2.43). \square

2. Wave Equation and Inverse Problem

The requirements of Theorem 5 can be easily fulfilled in practice. We already mentioned that our source functions usually fulfill higher regularity than required. When compact source functions are used $f(0) = 0$ and $f'(0) = 0$ can always be achieved with a time shift. The Ricker wavelet that we use is not compact, however it decays so fast, that within numerical accuracy we can assume compact support. Although Σ^X is more intuitive and advantageous in theory, in practice velocity profiles are defined in depth and a constant depth is assumed before the inversion. But we can guarantee $\Sigma^X \subset \Sigma^Z$ with sufficiently large choice of Z . This could introduce (additional) non-uniqueness, however the existence result persists.

Remark 1. Bamberger et al. discuss in their work actually a more general inverse problem of finding (ρ, μ) with the wave equation

$$\frac{\partial}{\partial z} \left(\mu(z) \frac{\partial u}{\partial z} \right) - \rho(z) \frac{\partial^2 u}{\partial t^2} = 0, \quad (2.52)$$

again for $z > 0, t > 0$ with zero initial conditions and boundary excitation at $z = 0$. However, as they discuss in their work, ρ and μ cannot be uniquely determined simultaneously from surface responses, therefore the authors construct different equivalence relations that identify classes of parameters that produce the same surface observation. Then they work on the quotient space instead. Although not fundamentally different, we were able to simplify some notation and skip that step above by substituting $\mu = 1$ and $\rho = \frac{1}{c^2}$.

2.2.2 Derivation of the Gradient

After showing existence in 1 D, we will finish this chapter by showing how the gradient of the misfit function E in Equation (2.2) can be computed. The gradient is later needed in the inversion algorithms described in Chapter 4. We will further assume, that we have a numerical way to solve the forward problem, which is described in Chapter 3.

We will now again use the general wave Equation (2.1) on $\Omega \subset \mathbb{R}^n$ and for given $T > 0$. We can obtain the gradient from the adjoint state method that can be derived with the Lagrange multiplier method (cf. [Tar84], [Ple06]).

Let $F(c) = E(U(c))$, thus our goal is to minimize $F(c)$ by using the gradient $F'(c)$. We continue to write $U = u|_{\Gamma_{r,s}}$ to remember that we measure the misfit for every source experiment only in the corresponding receiver locations. To simplify notation we will assume that we have only one source in the following computations. (We only need to sum over the sources to get the general case.) We define the Lagrange function \mathcal{L} with the Lagrange parameters λ, λ_1 and λ_2 as

$$\begin{aligned} \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) = & E(u) + \int_{\Omega} \int_0^T \lambda \left(\Delta u - \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} + f \right) dt dx \\ & + \int_{\Omega} \left(\lambda_1 u + \lambda_2 \frac{\partial u}{\partial t} \right) dx \Big|_{t=0}, \end{aligned} \quad (2.53)$$

where u, c and $(\lambda, \lambda_1, \lambda_2)$ are independent functions on $\Omega \times [0, T]$. Minimization of the Lagrange function can be understood as minimization of the original misfit

2. Wave Equation and Inverse Problem

function with the wave equation for u and the corresponding boundary conditions being constraints. Without further restrictions to u and c the minimum of \mathcal{L} is obtained at a stationary point of \mathcal{L} with respect to all parameters. It can be easily seen that

$$D_{(\lambda, \lambda_1, \lambda_2)} \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) = 0 \quad (2.54)$$

returns the constraints, that is wave equation and boundary conditions, whereby we use D as a symbol for the Fréchet derivative. Calculating

$$D_u \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) = 0 \quad (2.55)$$

will enable us to get the Lagrange parameters in the stationary points. They will be obtained as the solution of the adjoint wave equation.

If we insert $u = u(c)$ (i.e., the solution of the forward wave equation) and $\lambda = \lambda(c)$, $\lambda_1 = \lambda_1(c)$, and $\lambda_2 = \lambda_2(c)$ (i.e., the solution of the adjoint wave equation) into Equation (2.53), we obtain

$$\mathcal{L}(u(c), c, \lambda(c), \lambda_1(c), \lambda_2(c)) = E(u(c)) = F(c). \quad (2.56)$$

To obtain the gradient $F'(c)$ we can first compute

$$D_c \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) \quad (2.57)$$

in general from Equation (2.53) (for arbitrary u, λ, \dots) and then insert the solution of the wave equation and the adjoint wave equation.

Now, let us compute the Lagrange parameters from Equation (2.55). Let $\partial\Omega$ be the boundary of Ω and ν be the outward pointing normal direction. We start by integrating Equation (2.53) twice by parts.

$$\begin{aligned} \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) &= E(u) + \int_{\Omega} \int_0^T u \left(\Delta \lambda - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \lambda \right) dt dx \\ &+ \int_{\Omega} \int_0^T \lambda s dt dx + \int_{\partial\Omega} \int_0^T (\lambda \partial_{\nu} u - u \partial_{\nu} \lambda) dt ds \\ &+ \int_{\Omega} \frac{1}{c^2} \left(u \frac{\partial \lambda}{\partial t} - \lambda \frac{\partial u}{\partial t} \right) dx \Big|_{t=0}^{t=T} + \int_{\Omega} \lambda_1 u + \lambda_2 \frac{\partial u}{\partial t} dx \Big|_{t=0} \end{aligned} \quad (2.58)$$

$D_u \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) = 0$ means that for all functions $h : \Omega \rightarrow \mathbb{R}$

$$D_u \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) h = 0 \quad (2.59)$$

With Equation (2.58) this becomes for sufficiently differentiable h

$$\begin{aligned} 0 &= \int_{\Omega} \int_0^T h \left(U - U_d + \Delta \lambda - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \lambda \right) dt dx + \int_{\partial\Omega} \int_0^T (\lambda \partial_{\nu} h - h \partial_{\nu} \lambda) dt ds \\ &+ \int_{\Omega} \frac{1}{c^2} \left(h \frac{\partial \lambda}{\partial t} - \lambda \frac{\partial h}{\partial t} \right) dx \Big|_{t=0}^{t=T} + \int_{\Omega} \lambda_1 h + \lambda_2 \frac{\partial h}{\partial t} dx \Big|_{t=0}. \end{aligned} \quad (2.60)$$

2. Wave Equation and Inverse Problem

In particular, this equation must hold for all h with $h = \partial_\nu h = 0$ on $\partial\Omega \times [0, T]$ and $h = \frac{\partial}{\partial t} h = 0$ on $\Omega \times \{t = 0, T\}$. With such an h all but the first integral in (2.60) vanish and the variety of possible h with that property ensures that

$$\left(\Delta - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \lambda + u - u_d = 0 \quad (2.61)$$

in a weak sense. This is the adjoint wave equation in λ for the source $u - u_d$. Note that in case of the wave equation, the differential operator is the same as its adjoint. With a similar approach we can obtain initial conditions. Let λ solve the wave equation and let $h = \partial_\nu h = 0$ on $\partial\Omega \times [0, T]$ and $h = \frac{\partial}{\partial t} h = 0$ on $\Omega \times \{t = 0\}$. With this h only the third integral in (2.60) remains for $t = T$. Variation of h within the specified boundaries provides us with

$$\left. \begin{array}{l} \lambda = 0 \\ \frac{\partial}{\partial t} \lambda = 0 \end{array} \right\}, \text{ on } \Omega \times \{t = T\} \quad (2.62)$$

Finally we obtain $D_c \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2)$ by direct computation from Equation (2.53).

$$D_c \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) = \frac{2}{c^3} \int_0^T \lambda \frac{\partial^2}{\partial t^2} u dt. \quad (2.63)$$

Considering multiple sources again, we obtain

$$D_c \mathcal{L}(u, c, \lambda, \lambda_1, \lambda_2) = \frac{2}{c^3} \int_{\Gamma_s} \int_0^T \lambda \frac{\partial^2}{\partial t^2} u dt. \quad (2.64)$$

Inserting λ and u from above yields the gradient.

Remark 2. Two modeling runs are required to compute the gradient. The solution of the adjoint equation is often more vividly referred to as the backpropagated wave of the receiver misfit. So computing the gradient means correlating the forward solution of the wave equation with the backpropagated misfit via Equation (2.63), which is called imaging condition. The imaging condition is closely related to the imaging condition in reverse time migration (RTM). In RTM, λ is the backpropagation of the receiver signal itself instead of the misfit and we use u instead of its second time derivative. With this modification the same program computing the gradient in full waveform inversion can compute an RTM image. Physically speaking, RTM shows us the reflectivity in the model that is revealed by the real data. The gradient in FWI on the other side shows us where the model needs to be updated to better fit the real data.

3 Modeling

In this section we introduce the modeling algorithm and thus the core of our FWI program. Modeling is by far the most time consuming component, so an efficient implementation is imperative. When u is defined on a rectangular grid with N_x by N_z grid points, full implicit schemes require the solution of a $N_x \cdot N_z$ dimensional linear system in every time step. Although the matrix is sparse, computation is still extremely memory demanding and time consuming. Explicit schemes on the other hand require smaller timesteps for stability. We use an Alternating Direction Implicit (ADI) method [FM65] which could be seen as a tradeoff between an implicit and an explicit scheme.

Let Ω be a rectangular domain and $\Omega_h = \{(ih, jh) | i = 0, \dots, N_x, j = 0, \dots, N_z\}$ its uniform discretization. For now, we consider the wave equation with zero boundary conditions and zero initial conditions

$$\begin{aligned}
 Lu &= \Delta u - \frac{1}{c^2} \frac{d^2}{dt^2} u = f && \text{in } \Omega \times (0, T) \\
 u &= 0, \frac{d}{dt} u = 0 && \text{on } \partial\Omega \times (0, T) \cdot \\
 u &= 0, \frac{d}{dt} u = 0 && \text{on } \Omega \times 0
 \end{aligned} \tag{3.1}$$

We discretize the problem in time on $[0, T]_k = 0, \dots, kN_t$ and want to compute the update in step m : $u_{m+1} = u((m+1)k)$ by solving a system of the form

$$\begin{aligned}
 D_x u_{m+1}^* &= f_1(u_m, u_{m-1}) \\
 D_z u_{m+1} &= f_2(u_{m+1}^*, u_m, u_{m-1})
 \end{aligned} \tag{3.2}$$

on Ω_h with matrices D_x and D_z originating only from x - and z -derivatives respectively. This means that the matrices can be decoupled in one space dimension. So instead of solving one system of dimension $N_x \cdot N_z$ we can first solve N_z equations of dimension N_x to obtain an intermediate solution u_{m+1}^* and then N_x equations of dimension N_z to obtain u_{m+1} . Each set of lower dimensional systems is decoupled and can thus be computed in parallel (e.g., using vectorization and multithreading on modern CPUs).

3.1 Alternating Direction Method

In this work we use the ADI scheme from Fairweather and Mitchell [FM65], which we briefly restate here for our purpose. We try to approximate the wave operator L

3. Modeling

with a general discrete operator

$$L_h u = u_{m+1} - 2u_m + u_{m-1} + (\delta_x^2 + \delta_y^2) (a_1 u_{m+1} + a_2 u_m + a_3 u_{m-1}) \\ + \delta_x^2 \delta_y^2 (a_4 u_{m+1} + a_5 u_m + a_6 u_{m-1}) \quad (3.3)$$

for parameters a_1, \dots, a_6 yet to be determined. δ_x^2 and δ_y^2 denote the central difference operator in x and in y direction, for example

$$\delta_x^2 u(x, y, t) = u(x + h, y, t) - 2u(x, y, t) + u(x - h, y, t), \quad (3.4)$$

where the scaling with h^2 takes place in the parameters a_i . While $u_{m+1} - 2u_m + u_{m-1}$ and $\delta_x^2 + \delta_y^2$ are the (scaled) central difference approximations of $\frac{d^2}{dt^2} u$ and the Laplace operator respectively, the mixed term with $\delta_x^2 \delta_y^2$ is artificially introduced and allows us to decouple the equation $L_h u = f$ into

$$u_{m+1}^* - 2u_m + u_{m-1} + \delta_x^2 (a_1 u_{m+1}^* + a_2 u_m + a_3 u_{m-1}) \\ + \delta_y^2 \left[\left(a_2 - \frac{a_5}{a_1} \right) u_m + \left(a_3 - \frac{a_6}{a_1} \right) u_{m-1} \right] = f \quad (3.5)$$

$$u_{m+1}^* = u_{m+1} + \delta_y^2 \left[a_1 u_{m+1} + \frac{a_5}{a_1} u_m + \frac{a_6}{a_1} u_{m-1} \right],$$

when we choose $a_4 = a_1^2 \neq 0$. The other parameters are determined so that L_h approximates L best. For this purpose the terms in (3.3) are expanded in Taylor series in u up to 6th order in space and even time derivatives are replaced with (3.1). Eliminating all terms of lower order yields the coefficients a_1, \dots, a_6 . With the coefficients determined the final operator is

$$L_h u = u_{m+1} - 2u_m + u_{m-1} \\ + (\delta_x^2 + \delta_y^2) \left[\frac{1}{12} (1 - r^2) (u_{m+1} + u_{m-1}) - \frac{1}{6} (1 + 5r^2) u_m \right] \\ + \delta_x^2 \delta_y^2 \left[\frac{1}{144} (1 - r^2)^2 (u_{m+1} + u_{m-1}) - \frac{1}{72} (1 + 10r^2 + r^4) u_m \right] = f \quad (3.6)$$

with $r = \frac{ck}{h}$ (cf. [FM65] for details on the computation). For $r = 1$ this equation reduces to the explicit formula

$$u_{m+1} - 2u_m + u_{m-1} - (\delta_x^2 + \delta_y^2) u_m - \frac{1}{6} \delta_x^2 \delta_y^2 u_m = f \quad (3.7)$$

For $r \neq 1$ we can again obtain a decoupled formula

$$u_{m+1}^* - 2u_m + u_{m-1} + \frac{1}{12} (1 - r^2) \delta_x^2 \left[u_{m+1}^* - \frac{2 + 10r^2}{1 - r^2} u_m + u_{m-1} \right] \\ + r^2 \frac{1 + r^2}{1 - r^2} \delta_y^2 u_m = f \quad (3.8)$$

$$u_{m+1}^* = u_{m+1} + \frac{1}{12} (1 - r^2) \delta_y^2 \left[u_{m+1} - \frac{2 + 20r^2 + r^4}{(1 - r^2)^2} u_m + u_{m+1} \right]$$

3. Modeling

In that case the calculation of the coefficients already proves that the operator L_h is consistent with

$$|Lu - L_h u| = O(h^2 + k^2). \quad (3.9)$$

Lemma 6 ([FM65]). *For*

$$r \leq \sqrt{3} - 1, \quad (3.10)$$

L_h as given above is stable.

Remark 3. With different coefficients a_i above it is possible to create an ADI scheme that is unconditionally stable. Lees [Lee62] created such a scheme. Although it is also second order accurate in time and space it does not approximate the wave equation operator as well as this scheme and it is more prone to causing dispersion.

3.1.1 Dispersion

Additional to fulfilling the stability requirements above, the discretization should be chosen small enough to avoid dispersion. For this ADI scheme Daudt et al. [DBNC89] showed that dispersion can be neglected when using about four to five grid points per wavelength. So we should choose

$$h < \frac{\lambda_{min}}{4} = \frac{c_{min}}{4f_{max}}. \quad (3.11)$$

This is an advantage towards second order explicit finite-difference schemes, which require a grid density of at least ten points per wavelength [AKB74].

3.2 Boundary Conditions

The real seismic experiment only has one boundary, the water-air boundary. Ignoring water waves at the surface, we model this boundary as a hard wall, that is, we use zero boundary conditions at $z = 0$. To obtain a finite modeling domain we introduce artificial boundaries at $x = x_l$, $x = x_r$ and $z = z_b$ with $x_l < x_r$ and $z_b > 0$. When the boundaries are placed sufficiently far away from the source, the seismic waves do not reach the boundary during observation time and the boundary conditions are irrelevant. While this is achievable in small test problems, it is impractical for larger problems. So in general, we need boundary conditions that reduce reflections from the boundaries as much as possible.

3.2.1 Boundary Conditions in 1D

Let us first consider the one dimensional wave equation where perfectly nonreflecting boundaries are possible. The well known general solution of

$$\frac{\partial^2}{\partial x^2} u - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} u = 0, \quad (3.12)$$

as derived by d'Alembert is

$$u(x, t) = F(x - ct) + G(x + ct). \quad (3.13)$$

3. Modeling

F and G are right and left traveling functions, i.e., the shape of F and G stays constant while being shifted to the right or left with time t at velocity c . They are solutions to the equations

$$\frac{\partial}{\partial x}u - \frac{1}{c}\frac{\partial}{\partial t}u = 0 \quad (3.14)$$

and

$$\frac{\partial}{\partial x}u + \frac{1}{c}\frac{\partial}{\partial t}u = 0, \quad (3.15)$$

respectively. An intuitive way, to arrive at above equations is the formal “factorization” of the wave operator:

$$\left(\frac{1}{c}\frac{\partial}{\partial t} - \frac{\partial}{\partial x}\right)\left(\frac{1}{c}\frac{\partial}{\partial t} + \frac{\partial}{\partial x}\right)u = 0. \quad (3.16)$$

Having no reflections at the boundaries, implies having only right traveling waves at the right boundary ($G(x_r + ct) = 0$) and only left traveling waves at the left boundary ($F(x_l - ct) = 0$). Thus (3.14) and (3.15) give us the desired boundary conditions

$$\frac{\partial}{\partial x}u(x_l, t) - \frac{1}{c}\frac{\partial}{\partial t}u(x_l, t) = 0 \quad (3.17)$$

and

$$\frac{\partial}{\partial x}u(x_r, t) + \frac{1}{c}\frac{\partial}{\partial t}u(x_r, t) = 0. \quad (3.18)$$

3.2.2 Boundary Conditions in 2D

Nonreflecting boundaries are only possible for certain idealized conditions. For our purposes we use the one dimensional boundary conditions for the side and bottom boundaries of the two dimensional problem as well. This will completely remove the components of plane waves, traveling perpendicular to the boundary in question. Parallel components however are not muted and the reflection at the boundary increases as the incident angle departs from the orthogonal. Now, the complete initial and boundary conditions are:

$$\begin{aligned} u(x, 0) = 0, \quad \frac{d}{dt}u(x, 0) = 0 \\ \left(\frac{\partial}{\partial x} - \frac{1}{c}\frac{\partial}{\partial t}\right)u(x_l, z, t) = 0 \\ \left(\frac{\partial}{\partial x} + \frac{1}{c}\frac{\partial}{\partial t}\right)u(x_r, z, t) = 0 \\ \left(\frac{\partial}{\partial z} + \frac{1}{c}\frac{\partial}{\partial t}\right)u(x, z_b, t) = 0 \end{aligned} \quad (3.19)$$

As those boundary conditions are only dependent on one space dimension, they can be easily integrated into the discrete ADI formulation above. Figure 3.1 shows a radial symmetric wave, transmitted from a single source in a homogeneous medium

3. Modeling

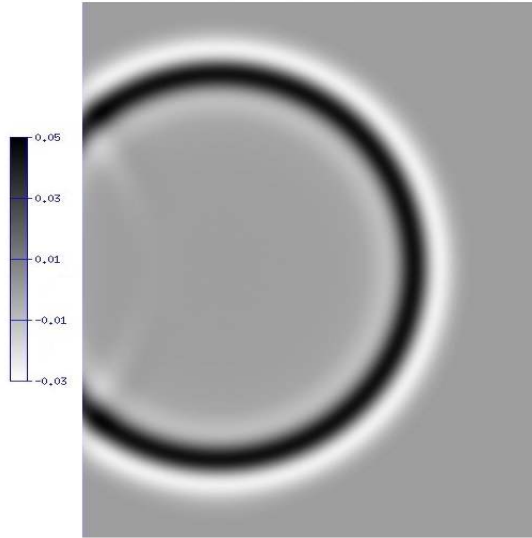


Figure 3.1: Boundary reflection at the left boundary.

that reached the left boundary of the model. There is no reflection exactly perpendicular to the boundary and only a small hardly visible reflection otherwise. The maximum amplitude at 45° is less than 5% of the original nonreflected amplitude.

Remark 4. More advanced boundary conditions trying to also reduce reflections from non orthogonal incident angles have been proposed in [CE77] and [Rey78]. However, some of them can not be immediately translated into the ADI context. Either an ADI formulation has to be found or separate treatment of the boundary is required. In the latter case computational efficiency can decrease and different discretization requirements for the non-ADI part can cause further complications. Many current codes use perfectly matched layers (PML), which are considered to be superior to other absorbing boundary conditions and can further reduce reflections. So far we could find little literature on combining PMLs with wave equation ADI methods. [WT03] and [LG00] describe PMLs for ADI methods to solve Maxwell's equations.

3.3 Validation of the Code

To validate the modeling, we consider a simple two layered model as shown in Figure 3.2 a. We assume that the model is infinitely extended in all space direction, so no boundary effects are considered. We can see a snapshot of the wave propagation in Figure 3.2 b after 0.5 s. We see a radially symmetric wave initiated from the source propagating through the upper medium. The transmitted wave, which is slightly weaker in amplitude, has travelled faster in the lower medium with the higher velocity and a weak reflection is visible coming back from the reflector.

To verify the solution quantitatively, we use the software Gar6more2D [DE11]. It uses the Cagniard de Hoop method [dH60] to compute the Green function of the wave equation for the desired configuration for a pulse source (delta distribution). The Green function is then convolved with the desired source function by means of

3. Modeling

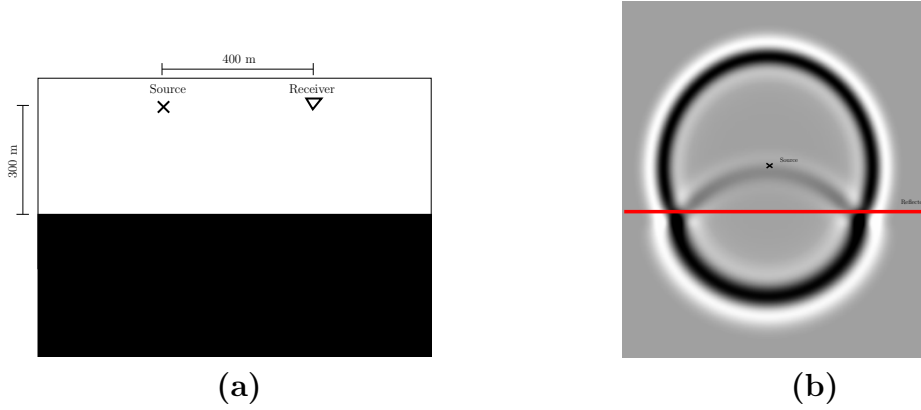


Figure 3.2: **(a)** Simple configuration with one reflector. Velocity is 1500 m/s above and 2000 m/s below the reflector. **(b)** Propagated signal in space after 0.5 s.

numerical integration. Given a sufficiently small step size during that integration we can regard this solution as a correct reference solution.

As source function we use the commonly used Ricker wavelet[Rya94]

$$f(t) = (1 - 2\pi^2 f_p^2 (t - \tau)^2) e^{-\pi^2 f_p^2 (t - \tau)^2} \quad (3.20)$$

as plotted in Figure 3.3 (a) with a peak frequency of $f_p = 8$ Hz and time shift $\tau = 0.1$ s. The frequency spectrum in Figure 3.3 contains contributions that are essentially below 25 Hz. That means to avoid aliasing we have to choose spatial discretization $h < 15$ m and time discretization $k < 0.005$ s. In Figure 3.4 we see that our numerical solution matches the reference solution very well. A 30 m spatial discretization causes slight aliasing whereas the 20 m solution already yields an almost perfect fit to the analytical solution.

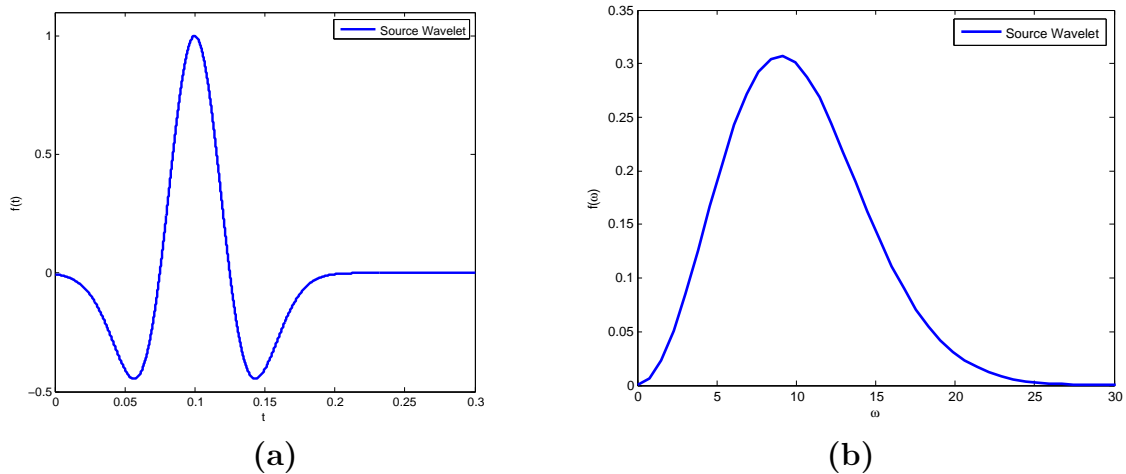


Figure 3.3: **(a)** Source wavelet. **(b)** Frequency spectrum of the source wavelet.

3. Modeling

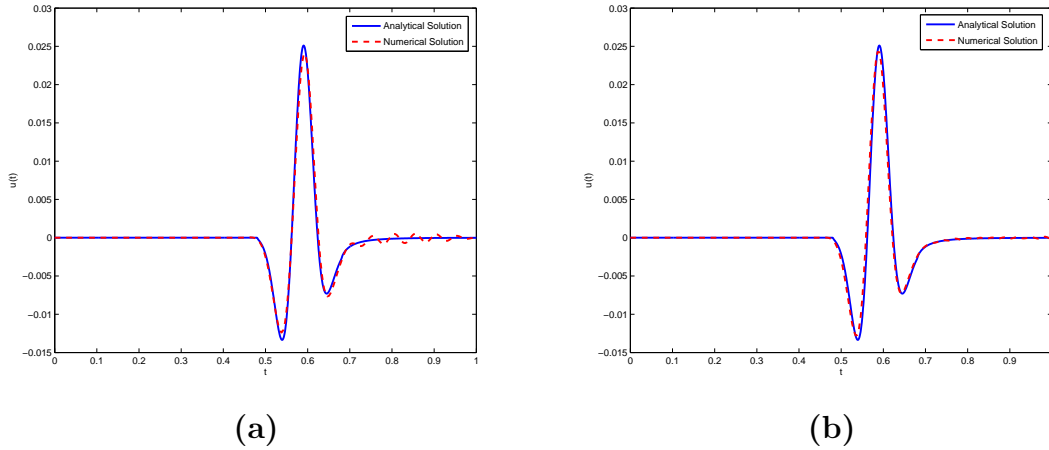


Figure 3.4: Numerical and analytical solution for $k = 0.4$ ms, (a) $h = 30$ m and (b) $h = 20$ m.

3.4 Comparison with Explicit Finite Difference Code

3.4.1 The Marmousi Model

In multiple tests in this thesis we use the Marmousi model [Ver94] shown in Figure 3.5. The model was created for a 1990 EAGE workshop [LRV90] in order to provide a common test to compare different methods of velocity estimation. It is still one of the most popular test cases today. It is a 2D synthetically created geologically plausible model, based on a profile through the North Quenguela Trough in the Cuanza Basin in Angola. Based on a model with many different layers of deposited material (marls, carbonates, sandstone etc.) a synthetical model with 160 layers was created. Velocity and density models were then derived from this model. In our thesis we set the density to one and make use only of the velocity model. Due to the many reflectors, the steep dips and the strong horizontal and vertical velocity gradients, the model is very complex. We add an additional water layer to the model, which is quite common in literature and convenient for technical reasons (easier to handle source imprint in FWI). The model is 3.5 km deep and 9.2 km wide. By modern standards this is rather small since some surveys extend over 100 km in width and beyond 10 km in depth, however it is quite suitable for testing purposes. We have the original model on a 5 m by 5 m grid available (~ 1.3 M grid points). For all other resolutions we use, we interpolate or extrapolate from this grid.

To test our modeling algorithm we now consider a simple setup with a single source at a horizontal position of 5.5 km and a depth of 30 m. Figure 3.6 illustrates with snapshots of the modeling how the pressure wave travels through the model. The waves are then recorded at a receiver line with receiver spacing of 20 m across the whole model. The receiver line is also placed at 30 m depth. Source and receiver positions are chosen so, that they coincide with grid points in all experiments and thus we can avoid spatial interpolation when we compare the results (cf. below, only for the 15 m computation grid some receiver signals are horizontally interpolated.)

3. Modeling

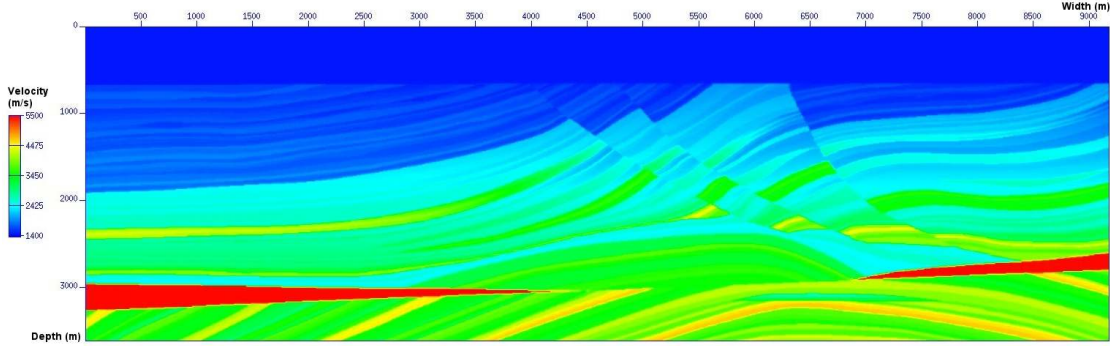


Figure 3.5: The Marmousi velocity model.

3.4.2 Comparison of the Results

For a runtime comparison, we compare our ADI scheme with the code Proteus, a parallel explicit finite difference code for 2D FWI that has been developed at the Geophysical Institute (GPI) at the Karlsruhe Institute of Technology (KIT)¹. We run the single source modeling experiment described above with both codes on a single node with four cores (dual Intel Xeon 5148LV). We first compute the receiver seismogram with our ADI scheme and a 2.5 m grid space. We can consider this solution as very accurate (both codes converge towards this solution) and compare all other solutions with this reference solution r . Then we calculate the relative error

$$e = \frac{\|u - r\|}{\|r\|}, \quad (3.21)$$

where $\|\cdot\|$ denotes the L_2 norm over time and source/ receiver configuration

$$\|u\| = \int_{\Gamma} \int_0^T u^2 dt. \quad (3.22)$$

At the time of this writing, the amplitudes of the Proteus code require scaling to produce the correct amplitudes. Therefore we do an amplitude fit of the Proteus results \tilde{u}_P to match our amplitudes

$$u_P = \frac{\int_{\Gamma} \int_0^T u_P r dt}{\|r\|^2} \tilde{u}_P. \quad (3.23)$$

Figure 3.7 visualizes how both codes converge with decreasing step size. Table 3.1 shows the errors and runtimes for different spatial grid sizes. Our code produces slightly better results at a 10 m grid spacing than the Proteus code at 5 m while only needing a two thirds of the runtime. Depending on the required accuracy even the results at 15 m gridspace might still be valuable from our code, although dispersion effects start to show. However they are considerably better than the results of the Proteus code at 10 m. And the Proteus results at 15 m with an error of 1 are no longer usable.

¹For further details on this code, contact the Department of Applied Geophysics at Karlsruhe Institute of Technology, details at <http://www.gpi.kit.edu>

3. Modeling

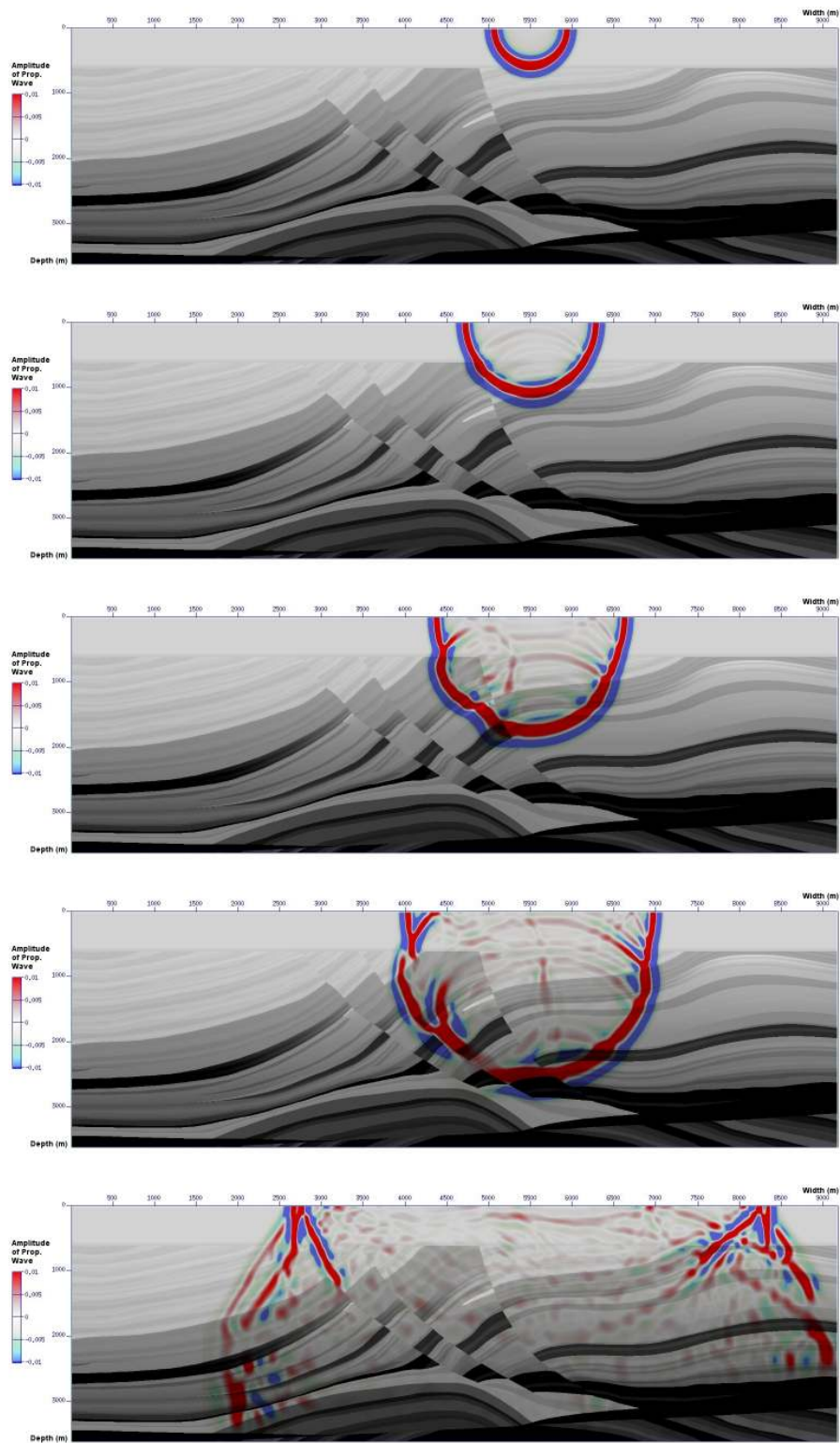


Figure 3.6: Superposition of velocity model and propagated wave in the medium. From top to bottom propagation is shown at 0.5 s, 0.7 s, 1 s, 1.2 s and 2.2 s.

3. Modeling

We should note that the Proteus code uses mpi, even for parallelization on a single compute node. On the one hand this allows the code to distribute the computation of one source onto multiple nodes on the other hand it might be slightly disadvantageous on one node as the different mpi processes do not have shared memory and have to exchange data. Our ADI scheme on the other side is implemented with pthreads. Pthreads might be advantageous on one node, as they share memory, however we cannot simply extend the code to multiple machines. Whether this is useful depends on the problem. Often the number of sources (and thus single modeling experiments) are of the order of 100 or larger and commonly exceeds the number of available compute nodes. In that case it is usually the better strategy to parallelize over the different sources and keep the computation of every single source on one machine. An exception might be when the amount of memory needed for the FWI computation exceeds the available main memory of one compute node. In that case it might be sensible to spread one modeling step across multiple nodes for memory reasons. However that usually does not occur: Even though modern acquisitions are overall much bigger than the Marmousi model, the offset used for a single source does usually not exceed 8-12 km. So even, with bigger depth and longer recording time the total amount of memory needed per source is not much more than about 2-10 times the amount of memory needed in the Marmousi example and would still fit into main memory of modern compute nodes. Unfortunately this reasoning does not hold for 3D computations. If the model extends in the third direction for 10 km, the model size already increases by a factor of 2000 when using a 5 m grid. In 3D cases the velocity model itself usually no longer fits into main memory and the computation must be distributed to several machines.

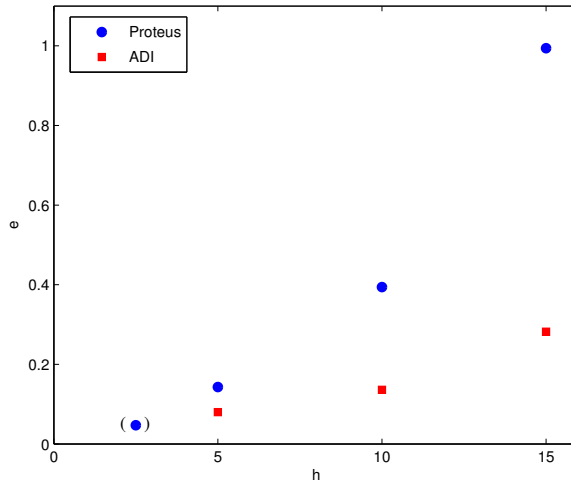


Figure 3.7: Relative error e for Proteus and ADI code for different grid sizes h . ADI Scheme at 2.5 m is used as reference solution.

3. Modeling

h [m]	k_P [s]	k_A [s]	e_P	e_A	t_P [s]	t_A [s]
2.5	4.4e-4	4.5e-4	(0.047)	-	225	1739
5	8.8e-4	9.1e-4	0.143	0.080	29.2	162
10	1.8e-3	1.8e-3	0.394	0.136	4.31	19.9
15	2.6e-3	2.7e-3	0.994	0.282	2.71	5.8

Table 3.1: Comparison of code Proteus P and our ADI scheme A . For the given spatial distance h , the time step k for both codes are chosen according to requirements of each code (based on CFL condition). The computation time is denoted by t .

3.5 Checkpointing

We mentioned in the last section, that 2D models usually fit into main memory. For the Marmousi model with 5 m grid, we only need to handle a 5 MB array to store u . Industrial sized 2D models could reach sizes of the order of 100 MB. However, to compute the gradient (Chapter 2.2.2), we need to correlate u with the adjoint λ via

$$\nabla_c E = \frac{2}{c^3} \int_{\Gamma_s} \int_0^T \lambda \frac{\partial^2}{\partial t^2} u. \quad (3.24)$$

We need to compute u before we can compute the adjoint, since the computation of λ needs U , that is u in the receiver positions, as input data. Also u is modeled forward from 0 to T while the adjoint wave equation has final conditions, so λ is modeled from T to 0. Therefore we need to store u in time to evaluate the time integral in Equation (3.24). Instead of using every timestep from the modeling algorithm, we can approximate the time integral by using a coarser time step size (e.g., the Nyquist sampling rate) but we still have to store more than 1000 snapshots of u (e.g., $T = 4s$, sampling $4ms$) and the total amount of data can easily exceed available main memory. When that is the case, we can either use the slower hard drive or employ a checkpointing scheme.

Let us assume we want to use u in a sequence of n time steps $(t_k) = \{t_1, \dots, t_n\}$ to compute (3.24). Then we select a subset of n_C checkpoints $(t_{k_i}) \subset (t_k)$ and n_b buffers $(t_{k_b}) = \{t_{n-n_b}, \dots, t_n\}$. In the first modeling run, we store u and $\frac{\partial u}{\partial t}$ in every checkpoint and only u itself in the buffer positions. Then we start the modeling of the adjoint (backward) equation and use the buffers for the correlation. Once the last buffer t_{n-n_b} is reached, we use $u(t_C)$ and $\frac{\partial u(t_C)}{\partial t}$ as initial data to recompute u up to t_{n-n_b} and fill a new buffer with the snapshots of u up to that point. The procedure then repeats with the new buffer and the remaining earlier checkpoints. The scheme is illustrated in Figure 3.8. For our tests, we only need very few checkpoints, which we select evenly spaced. For an optimal procedure of selecting checkpoints we refer to [GW00].

3. Modeling

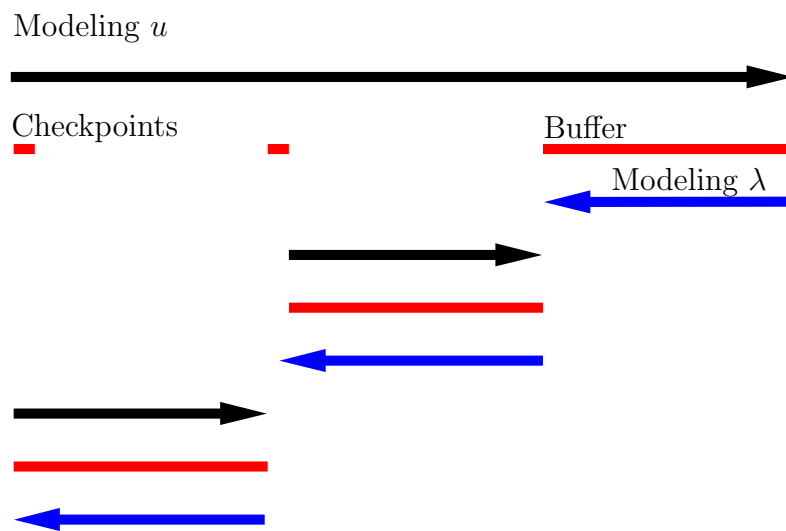


Figure 3.8: checkpointing scheme. u is only stored in checkpoints and buffers in order to save memory through recomputation.

4 Nonlinear Minimization

In this chapter we will finish the formulation of the complete inversion scheme. From previous chapters we have seen how the modeling algorithm provides a solution in the receiver positions for a given velocity profile c . We proved the existence of a minimum in one dimension and we obtained the gradient of the misfit function with respect to c with the adjoint state method. Now, we can use the method of steepest descent or other gradient based methods to find a (local) minimum. As stated before the inverse problem is non-linear in c and we must assume that the problem has many local minima. An easy way to see that is illustrated with Figure 4.1. If the model velocity c is too high compared to the real-earth velocity the synthetic wavelet arrives much earlier than the observed wavelet. When we shift the calculated wavelet towards the observed wavelet (i.e., slowly correct the kinematic error) the misfit between the two wavelets travels through a local minimum.

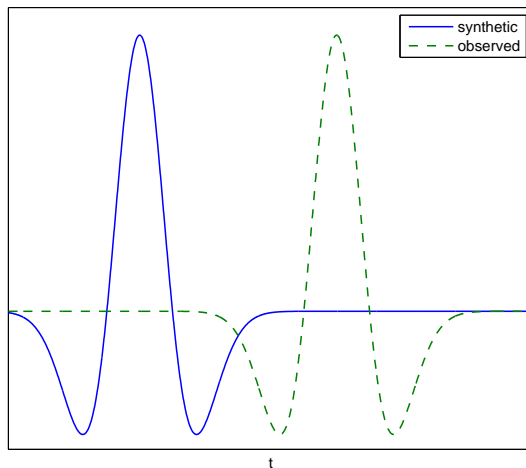


Figure 4.1: Here the time difference between synthetic and observed wavelet is so large that the misfit function would show a local minimum when the time difference is reduced.

This simple illustration also gives us a few immediate ideas on how to avoid local minima. We could

- obtain a good initial velocity profile by other (geophysical / geological) means, which produces modeling results close to the real data,
- modify the misfit function so that the computational solution is within the range of the global minimum,

4. Nonlinear Minimization

- or modify the observed data and/or input data directly so that the computational solution is within the range of the global minimum for the resulting misfit function.

The best procedure will probably be a combination of the three. Ultimately there is no practical way to prove for a real world problem that a global minimum has been reached and the solution of the inverse problem is correct. We can only outline different strategies here and test them on realistic synthetic test cases where we know the exact solution.

Let us first start by formulating the general descent methods.

4.1 Descent Methods

Let c_0 be our initial velocity model. Our goal is to iteratively construct new models c_1, \dots, c_n with $E(c_{k+1}) < E(c_k)$ so that $E(c_k)$ converges towards a minimum for $k \rightarrow \infty$. In step k we achieve this by constructing a descent direction g_k and a step length α_k and performing the update

$$c_{k+1} = c_k + \alpha_k g_k \tag{4.1}$$

so that

$$E(c_k + \alpha_k g_k) - E(c_k) < 0. \tag{4.2}$$

For a positive and sufficiently small step length α_k , this can be guaranteed if

$$g_k \cdot \nabla_c E(c_k) = \lim_{\alpha_k \rightarrow 0} \frac{1}{\alpha_k} (E(c_k + \alpha_k g_k) - E(c_k)) < 0. \tag{4.3}$$

There are different possibilities of defining g_k so that the condition $g_k \cdot \nabla_c E(c_k)$ holds. For example for a positive definite matrix A_k ,

$$g_k = -A_k \cdot \nabla_c E(c_k) \tag{4.4}$$

satisfies Equation (4.3). So the complete algorithm can be stated as:

1. Choose starting model c_0 .
2. In every step k compute update direction g_k , for example with

$$g_k = -A_k \cdot \nabla_c E(c_k). \tag{4.5}$$

3. Compute update with

$$c_{k+1} = c_k + \alpha_k g_k, \tag{4.6}$$

where the stepsize α_k is chosen to ensure $E(c_{k+1}) < E(c_k)$.

4. Iterate over k until a termination criterion is reached, which promises that the data is explained by the model sufficiently well.

Let us look at each step more closely.

4.1.1 Starting Model

The starting model must already produce reflections or at least diving waves, so that waves reach the receiver positions during the computation, therefore it is not possible to start with a constant velocity model. A velocity model linearly increasing with depth, that is with constant gradient in depth, could already work in principal. However, as we have already noticed above much care must be taken to avoid converging towards local minima in such a case. Better starting models could be obtained from other geophysical methods, based on ray theory and on geological reasoning.

4.1.2 Descent Direction

Steepest Descent

The most obvious way to construct a descent direction is to use the identity matrix for A_k in Equation (4.4). That means in every update step we locally compute the gradient and perform an update step towards that direction. This method is usually referred to as the method of steepest descent or just gradient method. It is tempting for its simplicity of formulation and implementation. However, the local approach often generates directions, which can only be used with small steplengths so that the overall rate of convergence tends to suffer.

Newton's Method

Instead of following the gradient only locally, we can take a more global view. Finding an extremum is equivalent of finding a zero of the gradient of the misfit function $\nabla_c E$. Let \tilde{c} be such a solution, then we can make a linear approximation in a step k

$$\nabla_c E(\tilde{c}) \approx \nabla_c E(c_k) + H_E(c_k) \cdot (\tilde{c} - c_k), \quad (4.7)$$

where H_E is the Hessian of E with respect to \tilde{c} . With $\nabla_c E(\tilde{c}) = 0$, solving for \tilde{c} formally gives

$$\tilde{c} \approx c_k - H_E^{-1}(c_k) \nabla_c E(c_k) \quad (4.8)$$

now using this approximation for c_{k+1} in the iterative method gives us Newton's method. In the general framework above, we obtain Newton's method by choosing

$$A_k = H_E^{-1}(c_k) \quad (4.9)$$

in the descent direction, provided that the inverse of the Hessian is positive definite. This method is of higher order, it converges quadratically while the steepest descent method only converges linearly. However, the Hessian could have negative eigenvalues and be ill-conditioned or even singular. Therefore damping and/ or regularization is usually introduced. The damped Newton's Method makes again use of a step length α_k , exactly like in the steepest descent method that has to be determined in each step (e.g., with a line search, cf. below). So the update for the damped method is

$$c_{k+1} = c_k - \alpha_k H_E^{-1}(c_k) \cdot \nabla_c E(c_k). \quad (4.10)$$

4. Nonlinear Minimization

As regularization we can add a scaled unit matrix to the Hessian, so that the update reads

$$c_{k+1} = c_k - (H_E(c_k) + \beta_k I)^{-1} \cdot \nabla_c E(c_k) \quad (4.11)$$

where β_k can be either fixed or can be determined analogous to α_k in each step. The regularized version of the Gauss-Newton method is called the Levenberg method. For large values of β_k the method approaches the steepest descent direction and will lose the advantage in convergence. An overview of the methods and regularizations can also be found in [Fic11]

Even with those improvements, Newton's method might still be infeasible in practice. First of all, we have to compute the Hessian in the first place and then solve a (generally dense) linear system in every iteration. Santosa and Symes [SS88] demonstrated the use of this method in one dimension. One way to avoid the need of computing second derivatives is an approximation of the Hessian, that is done in the Gauss-Newton method.

Gauss-Newton Method

$E = E(c)$ depends on c via the synthetic wave field $U(c)$. Let us now build the Hessian by differentiating E twice with respect to c while using the chain rule

$$H_E = (\nabla_U^2 E)(\nabla_c U \nabla_c U) + (\nabla_U E)(\nabla_c^2 U). \quad (4.12)$$

We have split the Hessian H_E into one part with first derivatives of U with respect to c and one part with second derivatives. When U behaves near linear with respect to c , we can argue that the second term with $\nabla_c^2 U$ is small compared to the first term and use the first term only to approximate

$$H_E \approx H_a = \nabla_U \nabla_U E(\nabla_c U \nabla_c U). \quad (4.13)$$

This method is called the Gauss-Newton method. A regularization term can be added in the same manner as in the full Newton method. In seismics, Pratt et al. [PSH98] demonstrated benefits of the Gauss-Newton method over the method of steepest descent. The full Newton method is usually not considered due to the high costs of computing the second derivatives of u . The first derivative can be obtained easily. U satisfies the wave equation. Taking the derivative of the wave equation for U with respect to c yields

$$\Delta \nabla_c U - \frac{1}{c^2(x)} \frac{\partial^2}{\partial t^2} \nabla_c U = -\frac{2}{c^3} \nabla_c c(x) \frac{\partial^2}{\partial t^2} U. \quad (4.14)$$

where $\nabla_c c$ is just the identity. The right hand term can be considered a source term for the wave equation for $\nabla_c U$. The term is sometimes referred to as a virtual source. It means that for every $c_j = c(x_j)$ we can obtain the derivative $\frac{\partial}{\partial c_j} U$ by solving the wave equation for a source

$$f_{virt} = -\frac{2}{c^3} \frac{\partial^2}{\partial t^2} U \quad (4.15)$$

in position x_j . In case the set of source positions coincides with the set of receiver positions, then this computation does not even require a new modeling step. The

4. Nonlinear Minimization

computation of the derivatives can be done simultaneously with the solution of the adjoint equation. However, usually the receiver positions are denser than the source positions. In that case additional effort is needed to compute the derivatives. In our test cases later, we will commonly use 10 times more receivers than sources. Thus for those tests computing even the Hessian would mean significant extra computation, which we will avoid.

After computing the first derivatives of U , we still need to compute the inverse of H_a (or solve the corresponding linear systems). When discussing the modeling algorithm in Chapter 3 we already argued that solving systems of that size is undesired and avoided them. If we want to avoid them here as well, we can further simplify the approximation of the Hessian, and use only the diagonal of the Hessian as done in [VS08]. Including regularization, we then obtain

$$A_k = (\text{diag}(H_a + \beta_k I))^{-1}. \quad (4.16)$$

Let us finally look at one further way to improve the general descent method by introducing conjugate descent directions.

Conjugate-Gradient Method

The conjugate-gradient method [HS52] constructs the descent directions g_k , so that they are orthogonal to each other. If the model space is n -dimensional, than after n iterations, all search directions g_0, \dots, g_{n-1} would form a basis of that model space. If furthermore the stepsize in every iteration is chosen optimal and if we neglect numerical errors, than the algorithm would have searched the whole model space and for quadratic problems convergence towards a minimum would be guaranteed. However, in practice we do not have a quadratic problem, the stepsize is not chosen that carefully and the model space is so large, that we could not afford that many iterations anyways. Nonetheless, in general the conjugate gradient method is known to converge faster than the simpler steepest descent method and it can be implemented at almost no extra cost.

Let

$$\tilde{g}_k = -A_k \cdot \nabla_c E(c_k) \quad (4.17)$$

be a descent direction in step k like discussed before. For $k = 0$ we define $g_0 = \tilde{g}_0$. For $k > 0$ we now obtain the conjugate direction g_k with

$$g_k = \tilde{g}_k + \gamma_k g_{k-1} \quad (4.18)$$

with

$$\gamma_k = \frac{\tilde{g}_k \cdot (\tilde{g}_k - g_{k-1})}{\|g_{k-1}\|^2}. \quad (4.19)$$

And then the velocity update is computed as before

$$c_{k+1} = c_k + \alpha_k g_k \quad (4.20)$$

To avoid problems due to inaccurate stepsize, the method can be restarted after a few iterations, so that only a limited number of orthogonal descent directions are used before the whole space is searched again.

4.1.3 Computing the Step Size

We still need to discuss how to calculate the step size α_k in the algorithm above and its variations. In general we want to have a stepsize α_k that minimizes $E(c_{k+1}) = E(c_k + \alpha_k g_k)$ with respect to the search direction g_k . Different one dimensional search strategies are possible. One of the most common ways is to solve for α_k from

$$\frac{d}{d\alpha_k} E(c_k + \alpha_k g_k) = 0. \quad (4.21)$$

For a linear expansion of E in c_k we obtain

$$g_k \cdot \nabla_c E(c_k) + \alpha_k g_k \cdot H_E(c_k) g_k = 0 \quad (4.22)$$

from where we can solve for α_k with

$$\alpha_k = -\frac{g_k \cdot \nabla_c E(c_k)}{g_k \cdot H_E(c_k) g_k}. \quad (4.23)$$

However, thereby we once again introduce the Hessian, which we might want to avoid for computational costs. A popular alternative is a line search. We calculate the misfit for several trial step lengths and compute a new step size that minimizes the polynomial function that interpolates through the misfits with those step sizes. In numerical experiments we can see that $E(c_k + \alpha_k g_k)$ is often close to quadratic in α_k so a quadratic polynomial should suffice. $E(c_k)$ is already known, so we know the misfit for $\alpha_{k,0} = 0$. We select two more points $\alpha_{k,1}$ and $\alpha_{k,2}$, compute $E(c_k + \alpha_{k,1} g_k)$ and $E(c_k + \alpha_{k,2} g_k)$, and compute α_k as the minimum of the quadratic function through those three points.

With this line search, we need to solve four forward modeling problems in total. Two, to compute the gradient and two more for the line search. Tape et al. ([TLMT10]) suggest, that some computation could be saved by not solving the whole forward problem during the line search, but instead limit the computation to certain events that influence the line search most. However, we do not discuss this approach here.

4.1.4 Stopping Criteria

Assuming our algorithm reaches a global minimum and produces an output that can explain the whole data well, then we can stop our algorithm after

$$E < \epsilon_d \quad (4.24)$$

where ϵ_d is threshold depending on the uncertainty of the data. Trying to descent even further would only mean that our algorithm would explain measurement errors with model features. In practice lack of a very good starting model and/ or lack of computational power to use very sophisticated multiscale algorithms (cf. below) can prevent us from reaching the global minimum. Therefore it might be better to add a stopping criteria that activates when the update from last to current iteration is no longer significant, for example with

$$E(c_{k-1}) - E(c_k) < \theta E(c_{k-1}) \quad (4.25)$$

for some threshold θ . Furthermore, we might want to stop the iteration after a certain number of steps, based on available compute power and reasonable time, or stop the iteration when the user is satisfied with the obtained update (e.g., based on visual criterion).

4.2 Improving Convergence

As indicated before, there are several further ways to improve convergence and lower the risk of converging towards a local minimum. Some of the methods use geophysical reasoning to justify certain modifications to the gradient, the error function, or to the data. Modifications to data or gradient are often described as pre-conditioning, although the terminology is not correct in the strict mathematical sense, where pre-conditioning usually means replacing the problem by an equivalent problem with the same solution but faster convergence.

4.2.1 Gradient in Water

In our work we will assume that the wave speed in the water layer as well as the ocean bottom line is known. This assumption is not trivial since the wave speed varies with pressure and salt content, however, different other inversion techniques can obtain this information reliably. So for our purposes, we track the ocean bottom line from the initial model and set the gradient $\nabla_c E = 0$ in the water layer.

Without this assumption, we usually see very strong contributions of the gradient below the water surface, where sources and receivers coincide. These contributions, called source imprint, are correct due to geometrical reasons of the model but significantly supersede the actually wanted contributions of the gradient in deeper levels. Fixing the water velocity during the inversion solves this problem. However this technique might not be possible for very shallow water layers or land acquisitions.

4.2.2 Alternative Error Functions

There are different ways to measure the residual between observed and synthetic data. So far we considered only the least squares (L_2) norm

$$E = \frac{1}{2} \|U - U_d\|_2^2 = \frac{1}{2} \int_{\Gamma} \int_0^T (U - U_d)^2 dt \quad (4.26)$$

with the corresponding gradient

$$\nabla_c E = \frac{2h}{c^3} \int_{\Gamma} \int_0^T \lambda^{(2)} \frac{\partial^2}{\partial t^2} u dt. \quad (4.27)$$

where $\lambda^{(2)}$ denotes the adjoint with for the L_2 norm E . Alternatively we can also use the least absolute value (L_1) norm

$$E_1 = \|U - U_d\|_1 = \int_{\Gamma} \int_0^T |U - U_d| dt. \quad (4.28)$$

4. Nonlinear Minimization

The derivation of the gradient can be done analogously to the derivation of the gradient for the L_2 error. A faster way to see, how the gradient can be calculated is to execute and compare the derivatives with respect to c directly. We already know that

$$\begin{aligned}\nabla_c E_2 &= \frac{d}{dc} \left(\frac{1}{2} \int_{\Gamma} \int_0^T (U - U_d)^2 dt \right) \\ &= \int_{\Gamma} \int_0^T (U - U_d) \frac{d}{dc} U dt \\ &= \frac{2h}{c^3} \int_{\Gamma} \int_0^T \lambda^{(2)} \frac{\partial^2}{\partial t^2} u dt,\end{aligned}\tag{4.29}$$

where λ_2 is the solution of the adjoint equation with source $U - U_d$. Now, similarly, we can see that

$$\begin{aligned}\nabla_c E_1 &= \frac{d}{dc} \left(\int_{\Gamma} \int_0^T |(U - U_d)| dt \right) \\ &= \int_{\Gamma} \int_0^T \frac{(U - U_d)}{|U - U_d|} \frac{d}{dc} U dt \\ &= \frac{2h}{c^3} \int_{\Gamma} \int_0^T \lambda^{(1)} \frac{\partial^2}{\partial t^2} u dt,\end{aligned}\tag{4.30}$$

where $\lambda^{(1)}$ is the solution of the adjoint equation with source term $\frac{U - U_d}{|U - U_d|}$.

The least absolute value norm has been used by Brossier et al. ([BOV10]) in frequency domain full waveform inversion. They found advantages over the L_2 norm in the presence of noise, especially for decimated datasets (which they have, when inverting for single frequencies only, cf. section 4.3.1). Another property of the L_1 residual is, that it reproduces sharper discontinuous features in the gradient, compared to smoother gradients of the L_2 norm, as we will see in the numerical experiments. This might be beneficial for some applications, when those discrete features are used (e.g., for interpretation) in early stages of the inversion.

Mathematically we have to keep in mind that the absolute value is not differentiable in 0, which means that in case an optimal model can explain the data completely, the gradient becomes singular in the minimum. Although this might not happen in real world applications, the L_1 norm could be problematic near the global minimum. It is possible to start with the L_1 norm and proceed with the L_2 after some iterations. More elaborate ways include the use of hybrid error functionals like the Huber criterion ([Hub73]) or the use of a hybrid L_1/L_2 criterion as proposed by Bube and Langan ([BL97]). Applications for frequency domain full waveform inversion can be found in [BOV10]. Another approach is to use regularization by using a source term

$$\frac{U - U_d}{|U - U_d| + \epsilon}\tag{4.31}$$

for the computation of the adjoint.

A third residual function we want to discuss briefly is the correlation error measure. Instead of computing the difference between synthetic data and observed data,

4. Nonlinear Minimization

we correlate both data. This can be advantageous when amplitudes in the observed or synthetic data are not considered reliable, since the correlation ignores amplitude differences and focuses on phase difference of the measured functions. The correlation error is

$$\begin{aligned} E_{corr} &= - \int_{\Gamma} \frac{\int_0^T UU_d dt}{\left(\int_0^T U^2 dt\right)^{\frac{1}{2}} \left(\int_0^T U_d^2 dt\right)^{\frac{1}{2}}} \\ &= - \int_{\Gamma} \frac{\langle U, U_d \rangle_2}{\|U\|_2 \|U_d\|_2}. \end{aligned} \quad (4.32)$$

This function is obviously not a norm, but we can still minimize it and compute the gradient as we did before.

$$\begin{aligned} \nabla_c E_{corr} &= - \frac{d}{dc} \int_{\Gamma} \frac{\langle U, U_d \rangle_2}{\|U\|_2 \|U_d\|_2} \\ &= - \int_{\Gamma} \int_0^T \left[\frac{1}{\|U\|_2 \|U_d\|_2} \left(\frac{\langle U, U_d \rangle_2}{\|U\|_2^2} U - U_d \right) \right] \frac{d}{dc} U dt \\ &= \frac{2h}{c^3} \int_{\Gamma} \int_0^T \lambda_{corr} \frac{\partial^2}{\partial t^2} U dt, \end{aligned} \quad (4.33)$$

where λ_{corr} is the solution to the adjoint equation with the source term

$$\frac{1}{\|U\|_2 \|U_d\|_2} \left(\frac{\langle U, U_d \rangle_2}{\|U\|_2^2} U - U_d \right). \quad (4.34)$$

This source term is actually the same as the source term of the least squares minimization with some additional scaling that can be understood as follows: The term

$$\frac{1}{\|U\|_2 \|U_d\|_2} \quad (4.35)$$

normalizes each receiver line. Thus far offset receivers, which receive a weaker signal due to spreading and damping are considered equal to near offset receivers. Advantages of this are also discussed below in section 4.3.3. The term

$$\frac{\langle U, U_d \rangle_2}{\|U\|_2^2} \quad (4.36)$$

is actually the solution α of the amplitude fit

$$\min_{\alpha} \|\alpha U - U_d\|. \quad (4.37)$$

So minimizing E_{corr} is equivalent to minimizing E_2 after an amplitude fit has been performed for the synthetic data towards the measured data and the difference has been scaled by weights from both the synthetic and the real data traces. This correlation measure has been specifically used together with an encoding function to do simultaneous inversion of multiple sources, details are discussed in [KAH⁺09].

4.3 Multiscale Approaches

When we look at the inverse problem with the full data, the problem usually has many local minima and even a good starting model might not be in the range of attraction for the global minimum. Now we want to look at several methods of solving the problem on different levels or scales, where the complexity of the problem in some sense is reduced and the misfit function has less local minima. We want to reduce the complexity so that the initial model is already in the basin of attraction of the minimum of the first (reduced) problem. That solution then should be in the range of attraction for the next finer problem. We proceed iteratively to the original problem. This multiscale approach is closely related to multigrid methods that are well known in mathematics (e.g., [Bra77], [BHM00], [McC88]). Those methods aim at the solution of large scale linear or nonlinear problems. They project or restrict the problem to a coarser grid (i.e., coarser scale), then perform a relaxation operator on that scale and inject the solution back onto a finer grid.

4.3.1 Multiscale Approach over Frequencies

As introduced by Bunks et al. ([BSZC95]) we can increase the range of attraction of the global minimum by looking at the problem at a larger scale, based on frequency content. Therefore we apply a lowpass filter to both the measurement data and the synthetic data. This filtered data contains only long wavelength and has much wider maxima, that is the time difference between maximum and minimum of the recorded or computed wavelet is much longer. Accordingly the misfit function has fewer extrema that are further apart. Our strategy is now to solve the problem at that low frequency and then incrementally increase the frequency content to include higher and higher frequencies and solve the problem again after each increase. We illustrate the strategy in Figure 4.2 with two scales.

The required computational grid size depends on the frequency content (cf. Chapter 3). That means, that we can start the minimization for the lower frequencies with coarser discretization of u and decrease the grid space as we increase the frequency band.

Instead of iterating over increasing bands of overlapping frequencies, it is possible to use a staggered approach, where we only iterate over single frequencies. This approach developed by Pratt et al. ([Pra99], [PS99]) uses a Fourier transformation of the problem. The Fourier transformed wave equation is

$$\Delta \tilde{u} + \frac{\omega^2}{c^2} \tilde{u} = \tilde{f}, \quad (4.38)$$

for frequencies ω , where the $\tilde{\cdot}$ denotes of the Fourier transform of the functions in time. This equation is also called Helmholtz equation. Solving this system for all frequencies ω is equivalent to solving the wave equation in the time domain. However, as the system here decouples over ω it is also possible to solve the system for selected frequencies only. This possibility is the major advantage of doing full waveform inversion in the frequency domain instead of the time domain (as we do). However, solving the Helmholtz equation for large systems is not easy and is usually done with parallel direct solvers ([SOV⁺09a], [SOV⁺09b]). Scalability concerns for

4. Nonlinear Minimization

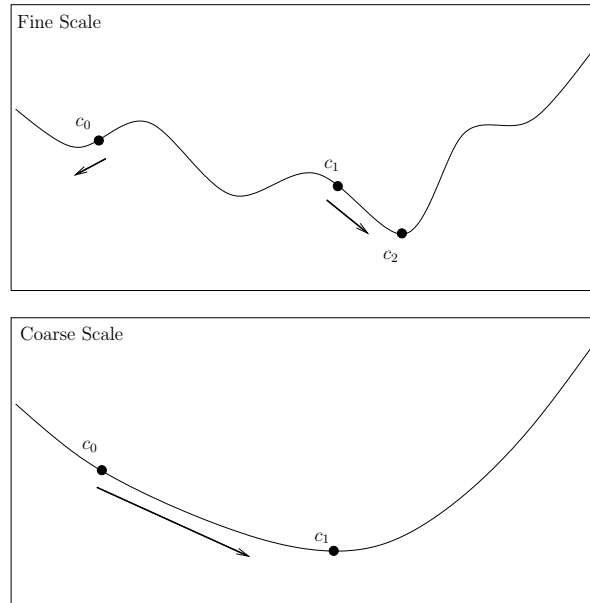


Figure 4.2: Heuristic illustration of the multiscale method on two scales. While c_0 does not lie in the range of attraction of c_2 on the fine scale, c_1 – reached on the coarse scale – does.

large data often motivates to work in the time domain instead. Notice that the differences between time domain and frequency domain are purely technical and important for implementation and performance. Mathematically the two approaches of full waveform inversion are equivalent.

One restriction of this multiscale approach over frequencies is the availability of low frequency content in the seismic acquisitions. In older surveys reliable data is often only available above 5 Hz. Obtaining lower frequency data requires higher energy seismic sources or more sensitive receivers, which rises environmental and technical problems respectively. Development into technology for obtaining lower frequencies is actively pursued and more recent studies were able to use frequency content from as low as 3 Hz [SBD⁺10]. Even lower frequencies can be obtained from passive seismologic measurements. However these surveys usually focus on surface waves and are less significant for exploration seismic.

4.3.2 Multiscale Approach over Time Damping

In general, early data in the receivers is only influenced by diving waves and first reflections from shallow reflectors. The data that arrives later at the receivers comes from wavepaths with longer traveltime, that is from waves that are reflected in deeper parts of the model or from multiples (waves that have been reflected several times before reaching the surface). Of course this description can only be a rough approximation, since the traveltimes depend on the heterogeneous velocity model. Nevertheless this heuristic suggests that the early synthetic data is more reliable than the later, since errors in shallow part of the velocity model accumulate for deeper reflections and multiple reflections. Therefore it is reasonable to focus on the early arrivals first to update the lower parts of the velocity model and then later in

4. Nonlinear Minimization

the inversion include later receiver data to improve the model accuracy in depth. An elegant way of doing this is by introducing an exponential time damping in the gradient with

$$\nabla_c E = \frac{2h}{c^3} \int_{\Gamma} \int_0^T \lambda \frac{\partial^2}{\partial t^2} u \exp(-kt) dt, \quad (4.39)$$

where k is a damping parameter. We can start by solving the inverse problem with a strong damping and then repeat for descending coefficients k until we solve the original problem for $k = 0$.

It is possible to improve the speed of computation by combining this technique with a time window. Depending on k we can approximate the gradient with an integration up to a time $T_k \leq T$ after which contributions of the solution are considered negligible due to the damping. In this case we only have to do the modeling until time T_k and use the adjoint $\lambda^{(k)}$ with final conditions at T_k to compute

$$\nabla_c E = \frac{2h}{c^3} \int_{\Gamma} \int_0^{T_k} \lambda^{(k)} \frac{\partial^2}{\partial t^2} u \exp(-kt) dt. \quad (4.40)$$

This method of introducing time damping is usually referred to ‘‘Laplace damping’’ or ‘‘Laplace inversion’’ since

$$\tilde{u}(k) = \int_0^T u(t) \exp(-kt) dt \quad (4.41)$$

is the Laplace transformation of u . In the frequency domain we have already seen, that it is possible to invert for specific frequencies instead of overlapping bands by transferring the problem into the frequency domain with a Fourier transformation. The same is possible here, where we can transform the whole problem into the Laplace domain and then invert for single Laplace components ([SC08], [LS09], [HPYS09]). The Fourier transform is equal to the Laplace transform for the complex $k = i\omega$. So after doing the complex Laplace transform, both frequency content and damping are described by a single complex Laplace coefficient k . That means a combined staggered algorithm can be set up that iterates over single frequencies and damping coefficients simultaneously ([SC09], [KSCP09]). However, like with the frequency inversion above, we have to solve the Helmholtz equation for the (complex) Laplace inversion as well. Thus the efficiency and scalability concerns raised before remain unchanged.

4.3.3 Multiscale Approach over Offsets

A last multiscale approach to refine the model iteratively makes use of different source - receiver offsets (i.e., distance between source and receiver). Usually, compared to the shorter offsets, larger offsets contain more data, that are influenced by the shallow parts of the model and data that come from shallow diving waves. Shorter offsets can contain fewer diving waves for geometrical reason. In the beginning of the inversion, a coarser background velocity (or smoothed model) might already be determined adequately, whereas individual (local) reflectors might not be accurate yet. In this case the diving waves are more reliable than the reflected

4. Nonlinear Minimization

waves. To make use of this observation, we can start the inversion by using only far offset data and then proceed to include shorter offsets in the process.

Unfortunately this approach already requires the solution of the whole problem in the early iterations, so we cannot gain computational efficiency by reducing grid size or time interval like before. However the shallow depth points are covered more by long offsets compared to short offsets, meaning for every model point (excluding near the boundaries), there are more single source-receiver experiments with long offsets that are influenced by that point than with short offsets. That suggests that it could be possible to save computational time by using only fewer sources in the beginning, where we consider only far offsets and then add more sources, when we add the nearer offsets. Unfortunately, as we will see in the test cases in Chapter 5, we will see more spherical artifacts when we limit the offsets (discussed in Ch. 5) which stands against this reasoning.

5 Test Cases

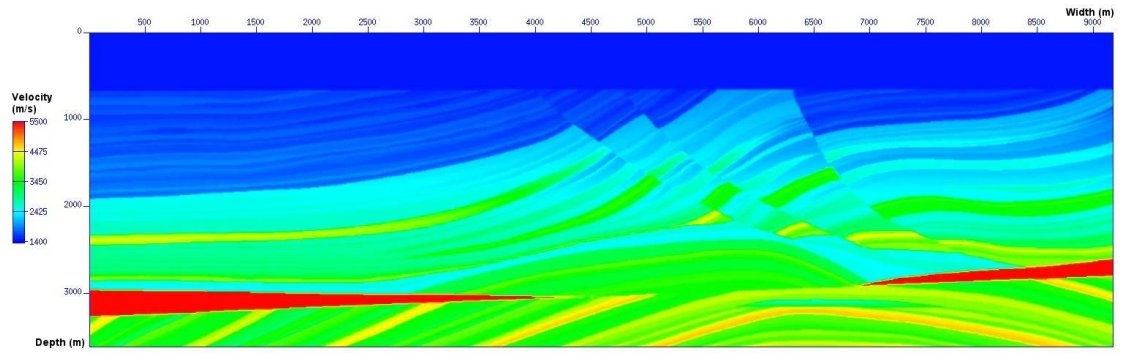
5.1 Marmousi

We already introduced the Marmousi model in Chapter 3. We will now use this model to test the inversion strategies outlined in the previous chapter. The model is 9200 m wide and 3500 m deep, including a 500 m water layer. We distribute 8 sources 1000 m apart from each other between 1000 m to 8000 m on top of the model. Receivers cover the whole surface with a 25 m spacing. In light of the discussion in chapter 4.3.1 we start with low frequencies only and apply a Butterworth bandpass filter with an upper limit of 5 Hz. The low frequencies allow us to use a rather coarse gridspacing. We use $h = 15$ m to stay consistent with later tests, but could choose even larger h . We keep in mind that in practice both higher frequencies and more sources are available. However, as we will see the configuration here is sufficient for first tests of our inversion schemes. Figure 5.1 (a) shows the original Marmousi model. We use this model to generate a synthetic response to our sources. Figure 5.1 (c) shows the response in a so called common source gather. It means that all the receiver traces are ordered by the sources. The first 280 traces belong to the first source at 1000 m, the next traces belong to the next source placed at 2000 m and so on. We now use this data instead of real measurement data as our input data in the inversion. As starting model we use a smoothed version of the real model that is displayed in Figure 5.1 (b). As we assume that the water velocity is known, the model is only smoothed underneath the ocean bottom and correct inside the water layer.

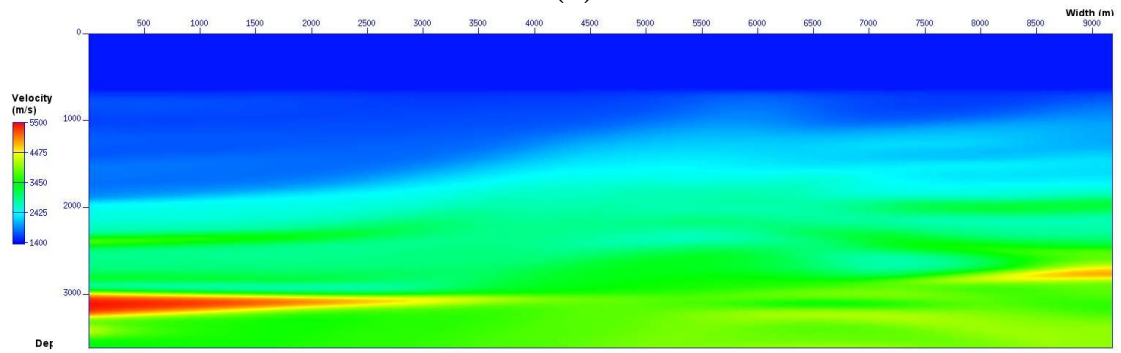
5.1.1 First Gradient

Figure 5.3 (a) shows us the simple gradient, scaled only with the step length of the first update step. We can immediately see that this gradient is not very good for a fast update. We see strong contributions in the water layer that are up to 10 times larger in amplitude than the contributions in the rock, despite the fact, that our model is correct in the water layer. The spherical artifacts happen due to limitations in the source-receiver geometry: A reflected wave that travels upwards from its point of reflection behaves almost like a plane wave after some time, when it reaches the receivers. However, due to the limited length of the receiver chain (or even due to sparse receiver locations, when using only few receivers) the adjoint (backward) wave equation does not reintroduce a plane wave solution, but rather a solution enveloped by a curved/ spherical wave front, as illustrated in Figure 5.2. Those radial contributions are then correlated with the near plane wave solutions in the gradient and produce the unwanted spherical artifacts. Those artifacts are especially strong near the source positions, where the wavefield u is strong. They dominate the velocity update there in an unrealistic way. One intuitive way to mitigate the

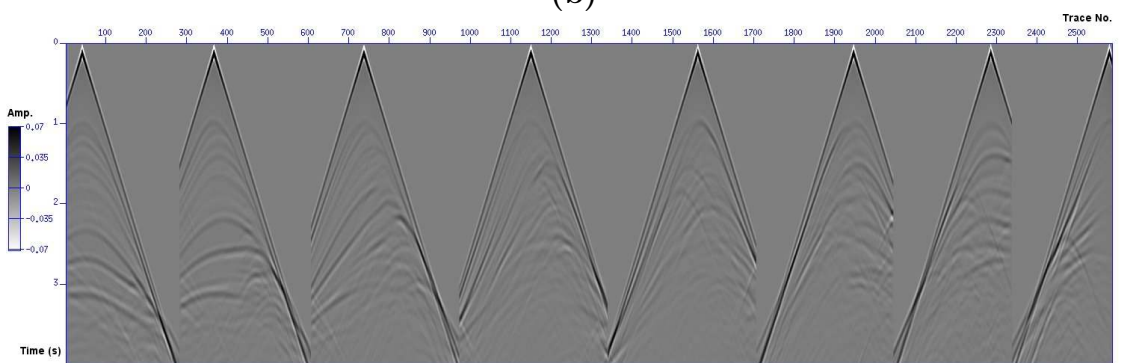
5. Test Cases



(a)



(b)



(c)

Figure 5.1: (a) Original Marmousi model. (b) Smoothed starting model. (c) Common source receiver gather.

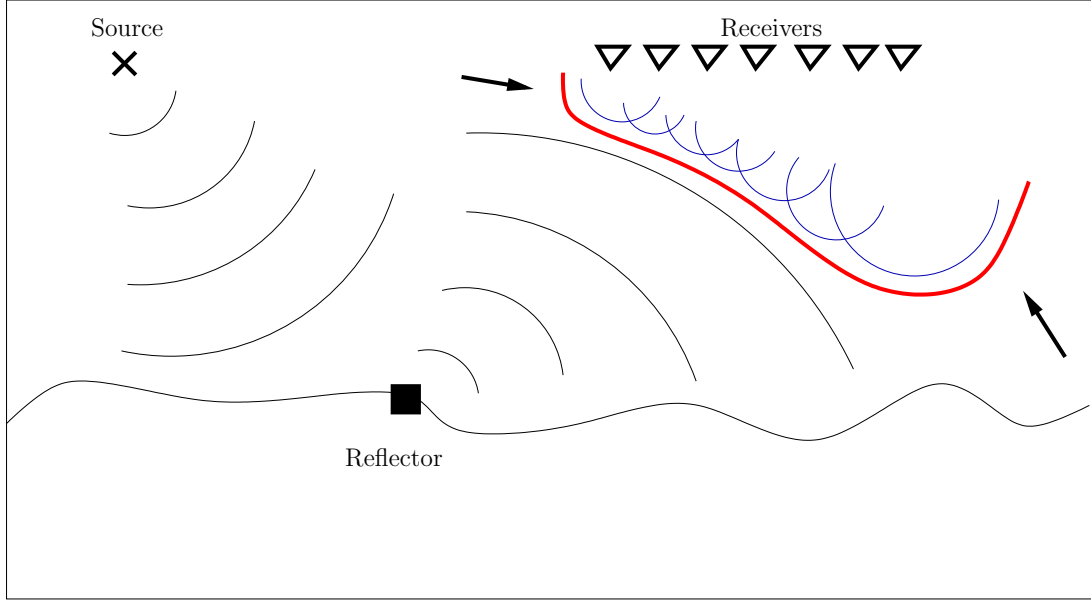


Figure 5.2: Illustration of the receiver imprint. The waves, reflected from one point are shown. The backpropagated waves (solutions of the adjoint equation) are shown in blue. Their envelope (red) has curved sections (indicated by arrows) on the sides due to the limited length of the receiver array.

problem is by scaling the gradient with the so called source illumination or receiver illumination. That means, we divide the gradient by a field

$$S(x) = \int_0^T u(x, t)^2 dt \quad (5.1)$$

or

$$R(x) = \int_0^T \lambda(x, t)^2 dt \quad (5.2)$$

or a combination of the two. This method is also known from reverse time migration [KG06]. As we can see in figure 5.3 (b), it also emphasizes parts of the gradient that are not “well illuminated”, meaning points in the underground that do not receive strong contributions from u or λ , for example due to depth or geometry of the acquisition.

An even more efficient way to improve the gradient, is just to ignore the gradient in the upper part, underneath the sources, which is easily possible as long as there is a sufficiently thick water layer in the model (cf. Figure 5.3 (c)).

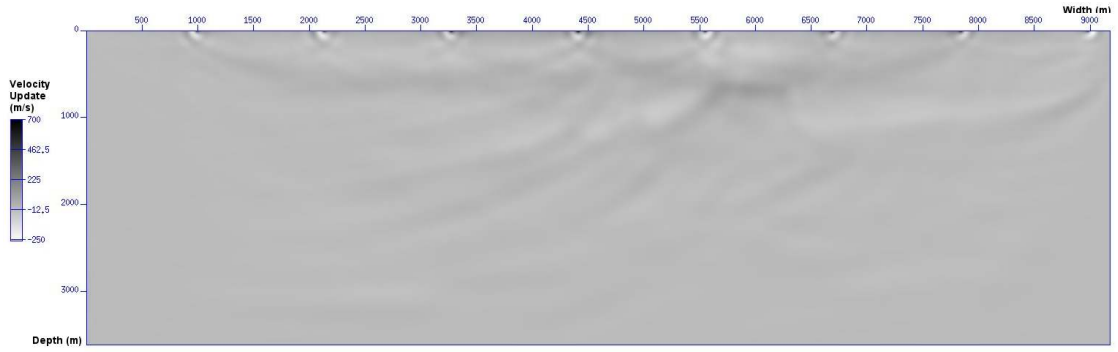
5.1.2 General Descent Scheme

First, let us check the line search. In figure 5.4 we show the true misfit function

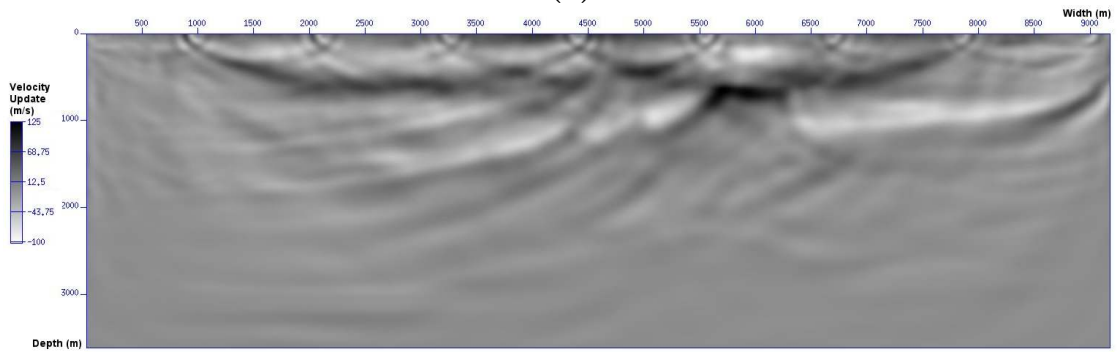
$$E(\alpha) = E(c + \alpha \nabla_c E(c)) \quad (5.3)$$

evaluated for 50 α together with a parabolic fit through three (estimated / guessed) evaluations of the first iteration. We see that the parabola fits the real misfit function

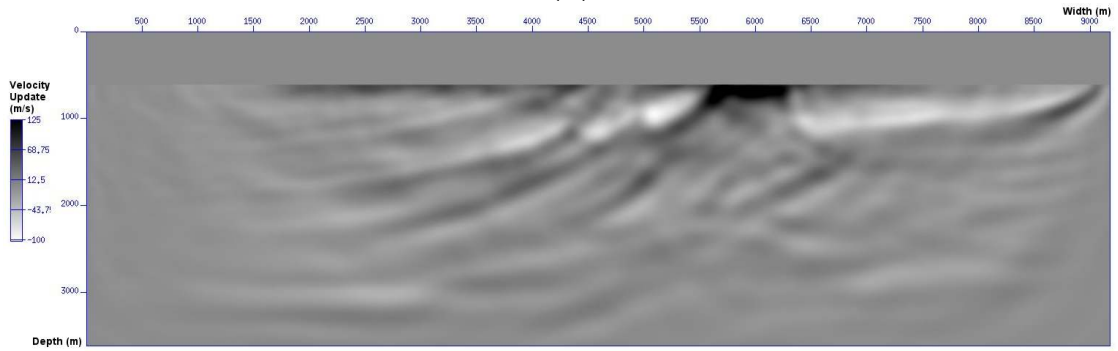
5. Test Cases



(a)



(b)



(c)

Figure 5.3: (a) First gradient, scaled by the steplength in first iteration. In (b) the gradient is additionally scaled with the source illumination. In (c) the gradient is set to zero in the water layer, but not scaled with the source illumination.

5. Test Cases

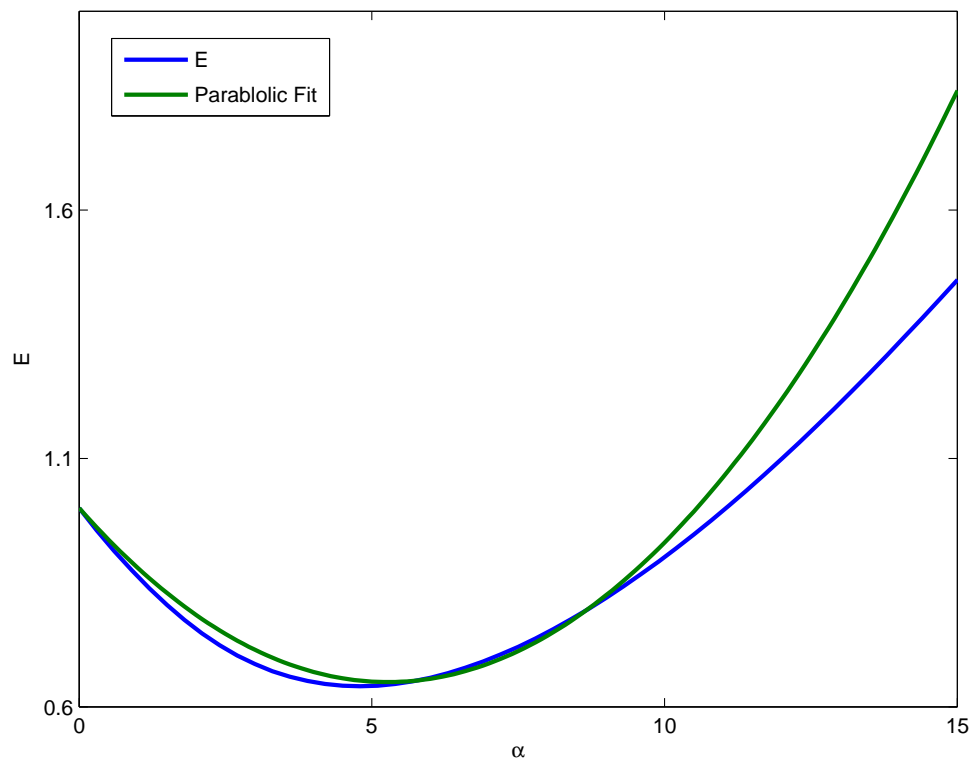


Figure 5.4: True misfit function E and parabolic fit in first iteration of the gradient descent scheme.

5. Test Cases

quite adequate near the global minimum and the minimum of the fit is close to the accurate minimum.

As stopping criterion we use

$$E(c_{k-1}) - E(c_k) < \theta E(c_{k-1}) \quad (5.4)$$

for $\theta = 0.01$. Using the simple gradient, only muted in water, and the plain steepest descent method, we see the first 30 iterations of the method in Figure 5.5, where we plotted the relative error

$$E_r(c_k) = \frac{E(c_k)}{E(c_0)}. \quad (5.5)$$

With the stopping criterion above, we would have stopped the inversion after 12 iterations. Given the small updates at that point, this seems to be a reasonable choice. We might even consider stopping earlier, when less accuracy is defensible, or if we intend to do further minimization in a multiscale scheme. For example, we could use $\theta = 0.05$ which in this instance results in stopping after the 9th iteration or even $\theta = 0.1$ which would have triggered the stop after the 5th iteration.

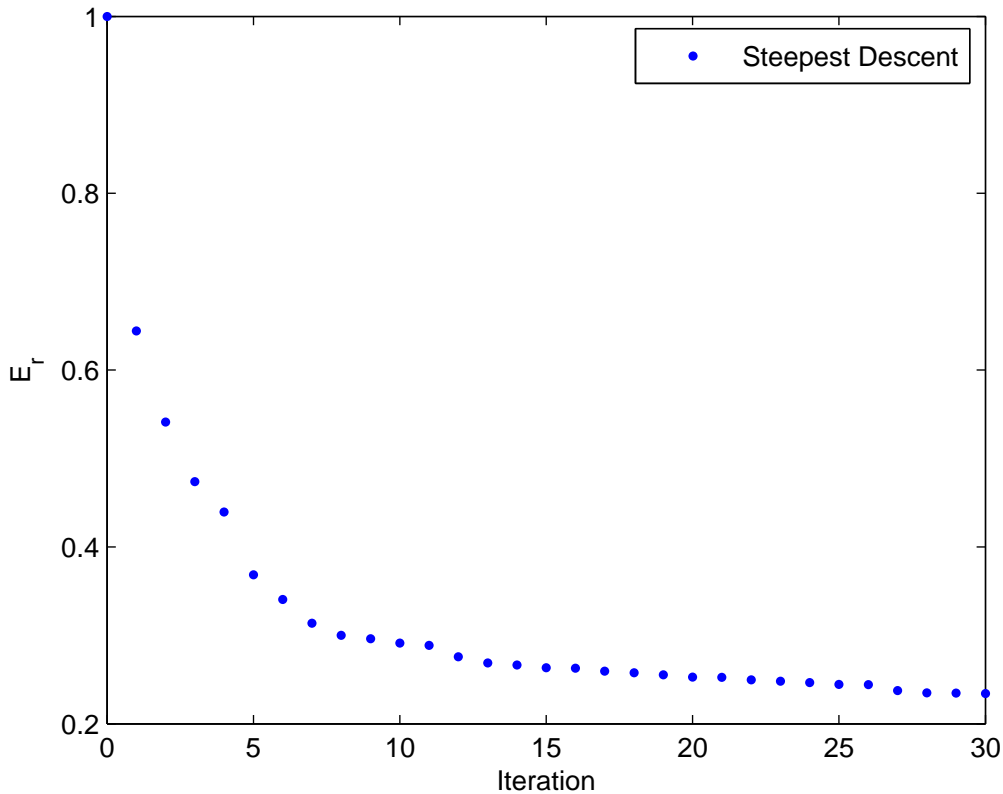


Figure 5.5: Convergence of steepest descent.

We already discussed reasons for muting the gradient in water. We can further strengthen our argument by looking at the rate of convergence. In Figure 5.6 we see the convergence in E_r for four different schemes: The unmodified gradient, the gradient scaled with the source illumination S , the gradient muted inside the water

5. Test Cases

layer, and the gradient both scaled with S and muted. All four schemes seem to converge quite similarly. The schemes with the muted gradient converge slightly faster in the beginning, but seem inferior after about 5 iterations. However, the difference can be seen in the velocity update. We compare the velocity updates directly by using the scaled Euclidean norm

$$E_c(c_k) = \frac{\|c_k - c_0\|}{\|c_0\|} \quad (5.6)$$

in Figure 5.7. In this comparison we see much larger differences in the schemes. The unmodified gradient performs worst. Scaling with the source illumination improves the result slightly, but muting the gradient in the water layer has the biggest effect. Tests with the receiver illumination produced comparable results to the tests shown. Combining the mute and the scaling does not give superior results. This confirms, that without the mute, the algorithm tries to explain data with erroneous updates near the sources in the water layer. A scaling with the source illumination that mitigates the problem slightly does unfortunately not improve the overall convergence for the deeper parts of the model. That means that whenever a sufficiently deep waterlayer is known in the model, we can mute the gradient, in other cases using source (or receiver) illumination could lead to improvement.

Remark 5. Although it is easy and intuitive to check the velocity update with the error function E_c above, we should only use it with caution. First of all, we do not and cannot minimize this function directly, so we cannot expect this function to converge to zero. Furthermore, it weighs the velocity difference equally over the whole model, which can include parts of the model, that gain little or no update at all, for example near the model boundaries or in deep parts of the model. In Figure 5.7 we see that we only gain a 5% improvement in E_c . This is because differences in the lower parts of the model and on the lower sides are very strong and dominate the whole error, even though we make considerable improvement in the top part of the model. Finally, we have to keep in mind, that the real velocity model is only available in our synthetic tests, so this function cannot be used in real world applications.

Finally, let us compare the steepest descent method with the conjugate gradient method. In figure 5.8 we see that the conjugate-gradient scheme converges slightly faster in the first iterations and then follows the steepest descent algorithm after the 6th iteration. Restarting the conjugate gradient algorithm in between did not lead to improvement. Obeying the stopping criteria stated before, the gradient algorithm would have stopped already after the 12th iteration, which for the conjugate gradient would be achieved with a stopping criterion of $\theta = 0.2$. In Figure 5.9 we also compare the length of the gradients, computed with both methods in the error function

$$E_{Grad}(c_k) = \frac{\|\nabla_c E(c_k)\|}{\|\nabla_c E(c_0)\|}. \quad (5.7)$$

As expected the gradients get smaller during the iterations when the update steps get smaller. The gradients in the conjugate gradient algorithm decrease more monotone than the gradients of steepest descent, which makes a robust line search and reliable stopping criterion easier. Keeping in mind that the conjugate gradient can

5. Test Cases

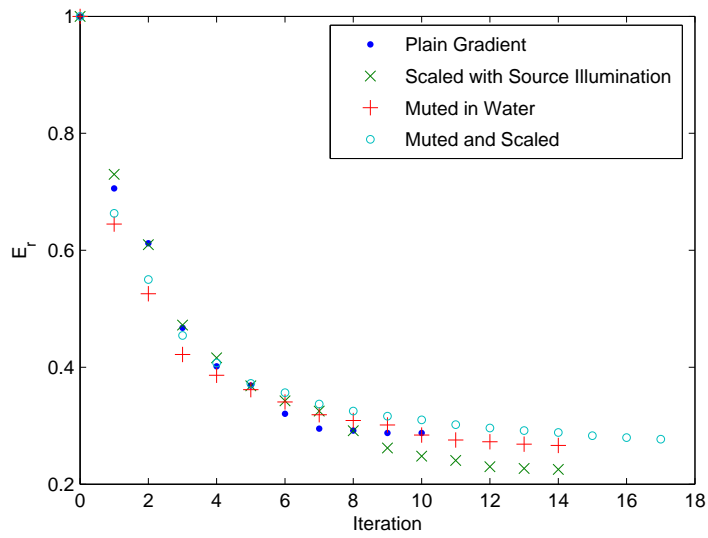


Figure 5.6: Comparison of the relative misfit E_r for different modifications of the gradient.

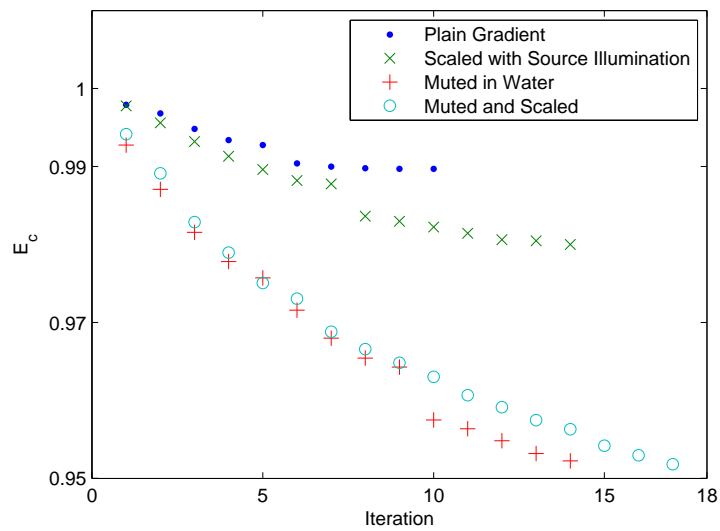


Figure 5.7: Comparison of the velocity error E_c for different modifications of the gradient.

5. Test Cases

be computed at practically no extra cost, this and the faster convergence in the beginning convinces us to give preference to the conjugate gradient method over steepest descent.

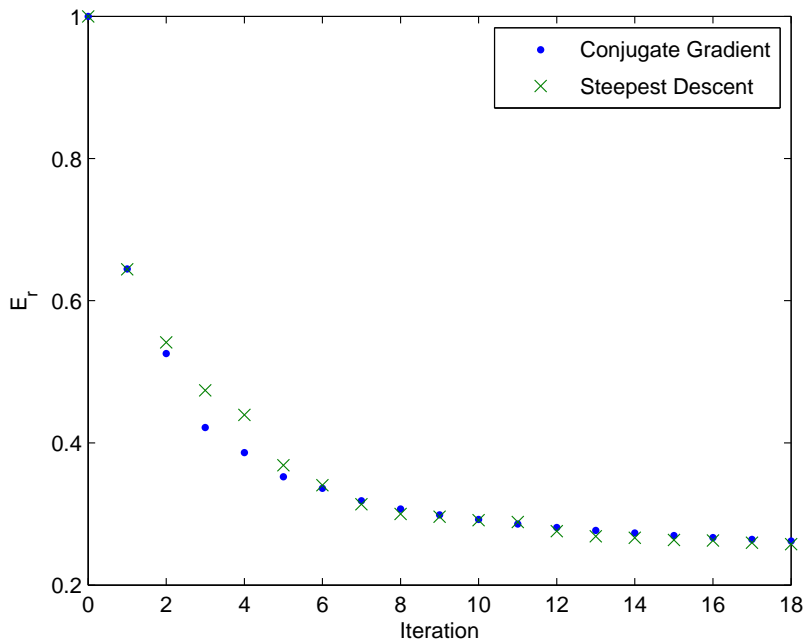


Figure 5.8: Comparison of conjugate gradient and steepest descent.

Before continuing with more advanced schemes, let us briefly examine, how much progress we made in solving the inverse problem with the method tested up to this point. We solved the inverse problem on bandpassed data and in 12 iterations we were able to explain 72% of that filtered data ($E_r = 0.28$). If necessary, more iterations can, although very slowly, still improve on that (e.g., 30 iterations: $E_r = 0.23$). We see the resulting velocity model now in Figure 5.10 (a). The first layers are now clearly visible and the mountain like top structures in the middle are now recognizable with sharp edges. We can better see the improvements by comparing the difference plots in Figure 5.10 (b) and (c). The errors in the velocity model are now very small in the top structures above 1500 m, some improvement can be seen in the middle area around 2000 m and de facto no update is performed below 2800 m. To further improve the results, we can both try to improve the methods on the current data and we can include the higher frequency data.

Without bandpass filter, the current model explains only 35% ($E_r = 0.65$) when using a Ricker signal with 15 Hz peak frequency. The effect can also be seen, when comparing a single trace from the original model and the solution in Figure 5.11. We can see that the bandlimited traces in Figure 5.11 (b) already match quite well, even for later times, which mostly correspond to deeper levels. Another observation we can make from the traces is that the phases already fit quite well, whereas the amplitudes of the solution are too small. In general, errors in the (smooth/ average) background velocity would produce phase shifts, whereas the strength of the reflectors (e.g., velocity contrast between two reflectors) produces

5. Test Cases

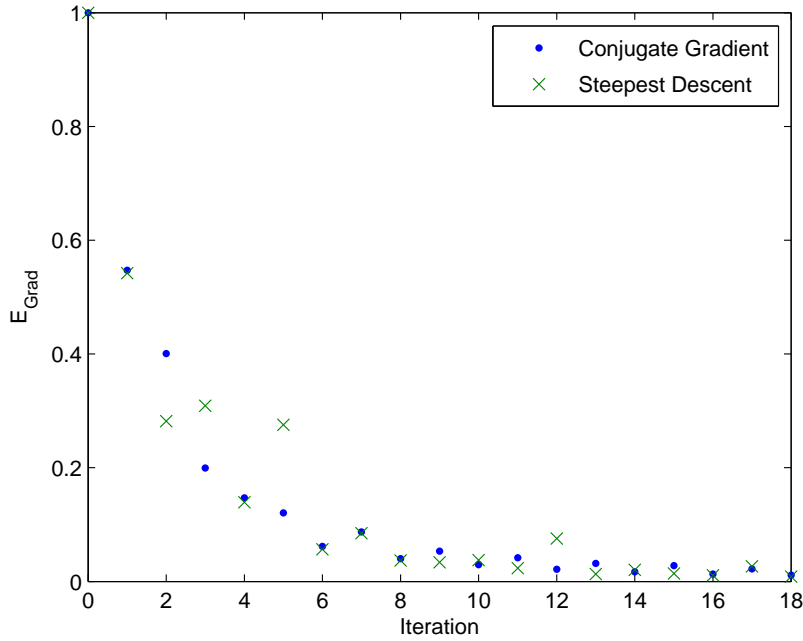


Figure 5.9: Comparison of gradients of conjugate gradient and steepest descent method.

errors in the amplitude. By implication, we need to improve local contrasts to improve convergence.

5.1.3 L_1 Norm

After the preceding analysis, we expect benefits from the least absolute value norm, since it is known to allow for sharper contrasts in the gradient. Indeed, in 5.12 we see that the first gradient from the misfit function

$$E_1 = \|U - U_d\|_1 = \int_{\Gamma} \int_0^T |U - U_d| dt \quad (5.8)$$

shows stronger contrasts than the gradient from the least square norm. We also notice that the contributions in lower parts of the model are clearer and stronger. We see that the method converges in the L_1 norm

$$E_{1,r}(c_k) = \frac{E_1(c_k)}{E_1(c_0)} \quad (5.9)$$

in 15 iterations for ($\theta = 0.005$). With a relative error of over 50%, we are probably converging towards a local minimum of that norm here. However, when we compare the error with the least squares method in Figure 5.14 (a), we see that the least absolute norm converges even slightly faster and more uniformly than the least square norm measured in E_r . The biggest advantage can be seen when comparing the velocity update in Figure 5.14 (b). When minimizing the L_1 norm we achieve in seven iterations roughly the same quantitative velocity update, that the

5. Test Cases

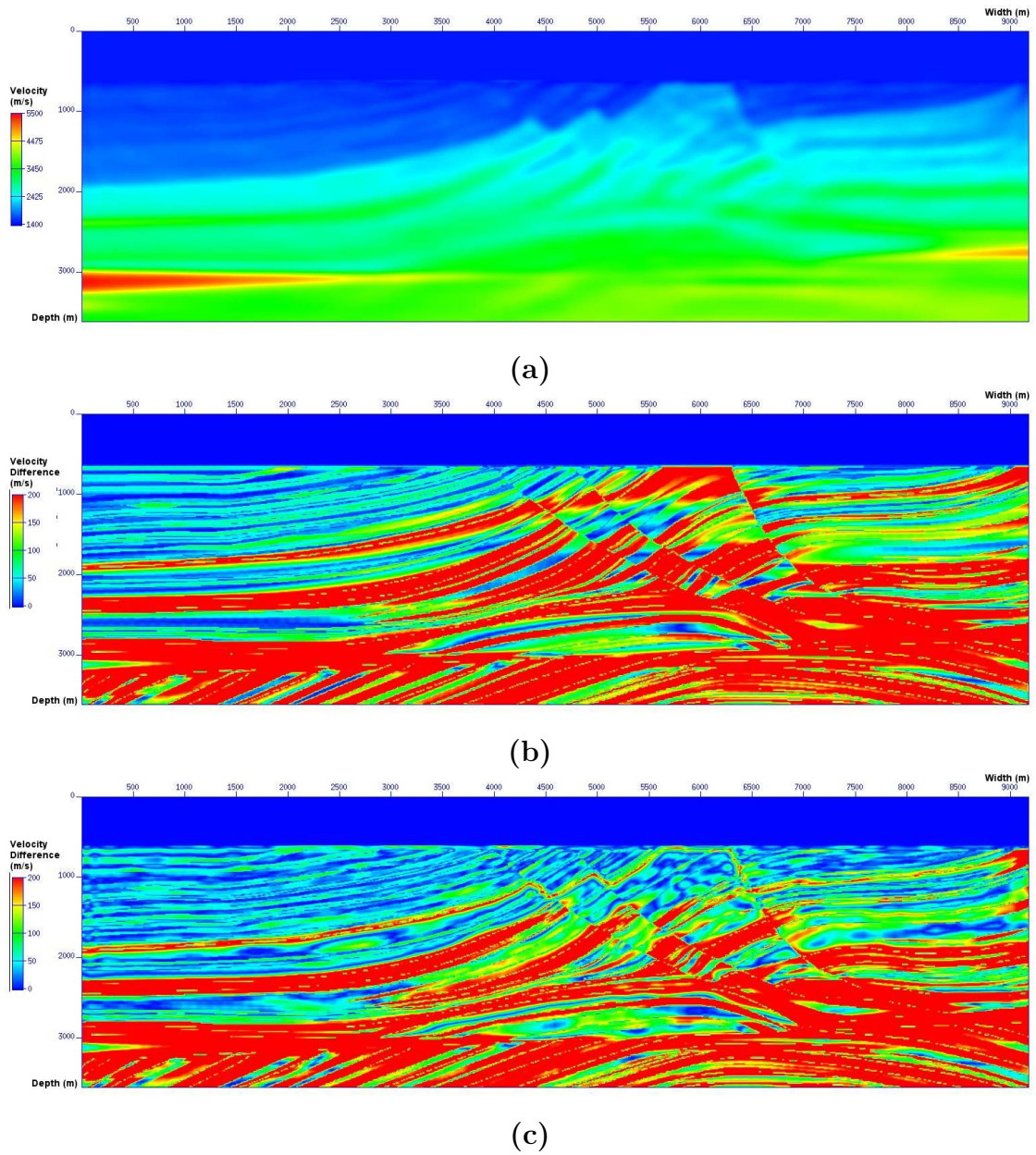


Figure 5.10: (a) Velocity model c_{12} , i.e. velocity model after 12 iterations. (b) Absolute value of difference between starting model and original model $|c_0 - c_{real}|$. (c) Difference between c_{12} and real model $|c_{12} - c_{real}|$.

5. Test Cases

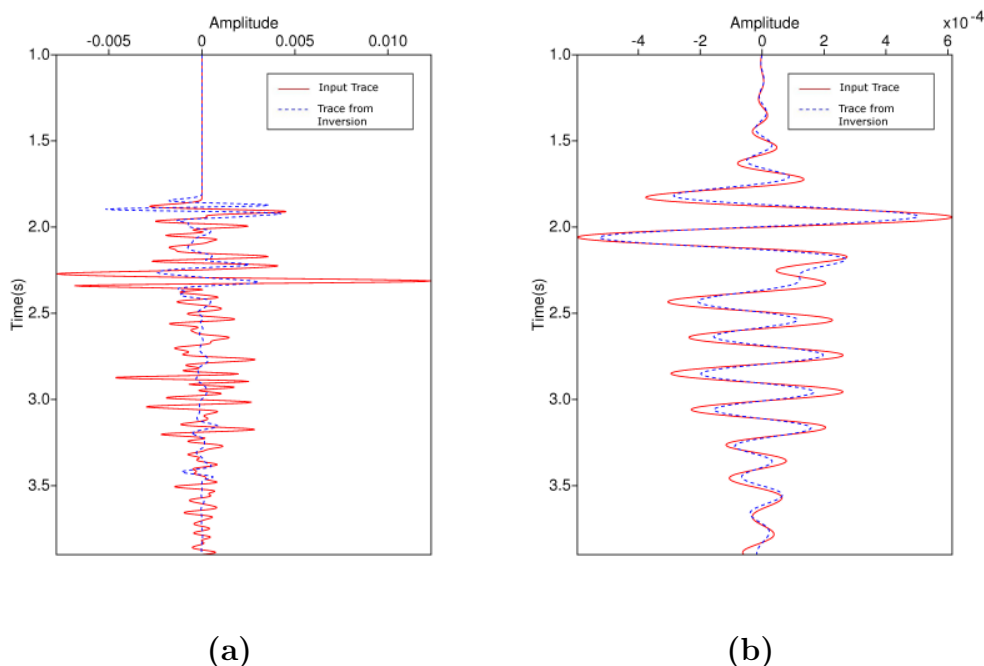


Figure 5.11: Single trace from receiver at 5500 m, source at 3000 m. Traces modeled with the real model and the inversion result from c_{12} . (a) Full frequency range, (b) bandpass filtered below 5 Hz.

L_2 method produces in 15 iterations. Furthermore the update is also more regular. The stronger gradient contributions in lower parts improve the result. When comparing the differences in Figure 5.15 we now see that some updates are performed in the lower regions of the model, around 3000 m.

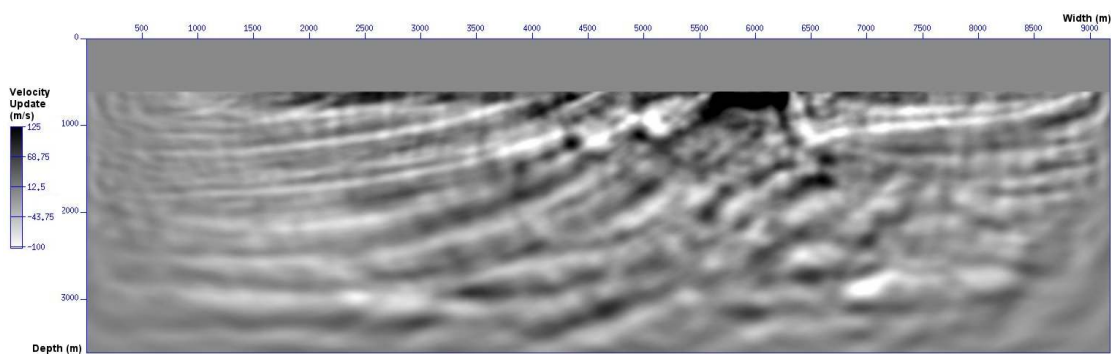


Figure 5.12: First gradient based on L_1 misfit E_1 scaled with steplength in first iteration.

5.1.4 Noise

Up to this point, we always committed the “inverse crime” by inverting with receiver data that we previously computed with the same algorithm. This method is quite common and has been instructive so far. To make our experiments more realistic

5. Test Cases

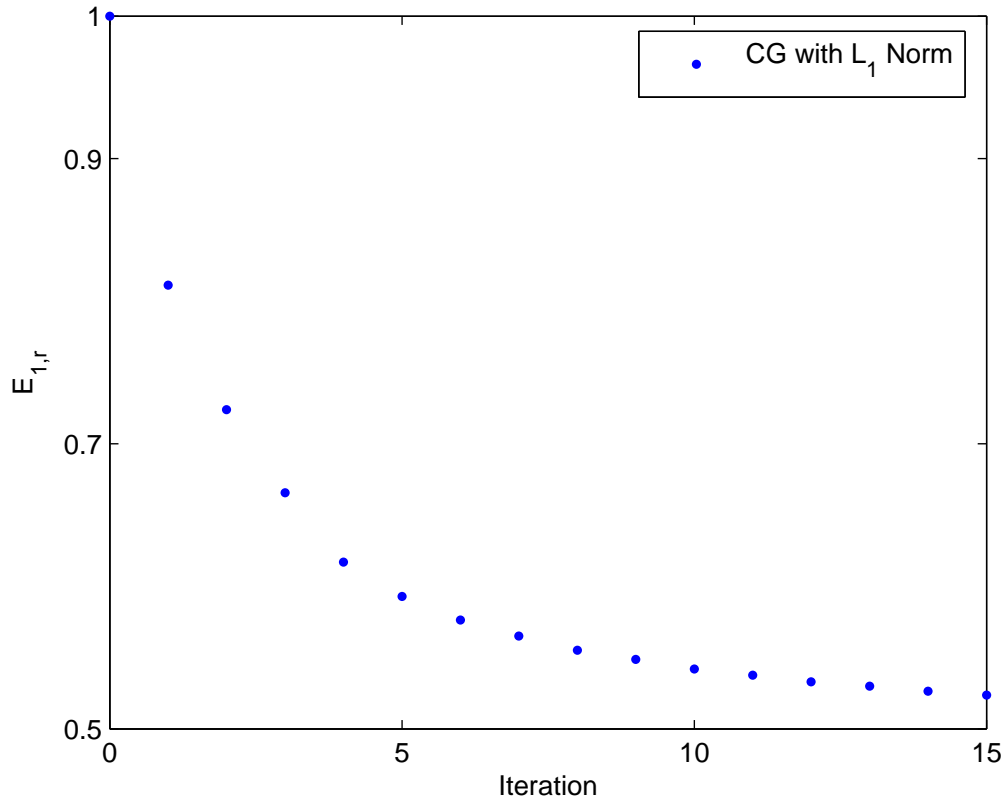


Figure 5.13: Relative L_1 error of conjugate gradient method based on L_1 norm.

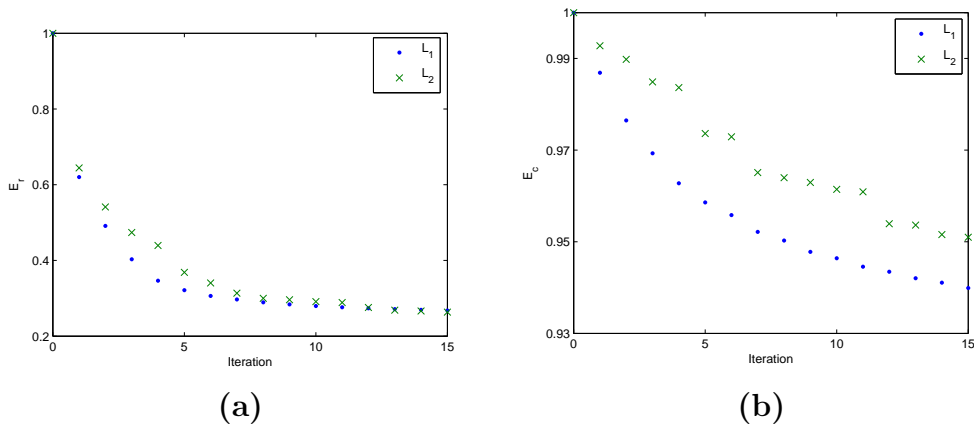


Figure 5.14: Comparison of L_1 and L_2 based conjugate gradient method. (a) Misfit measured in L_2 norm. (b) Error in the velocity field.

5. Test Cases

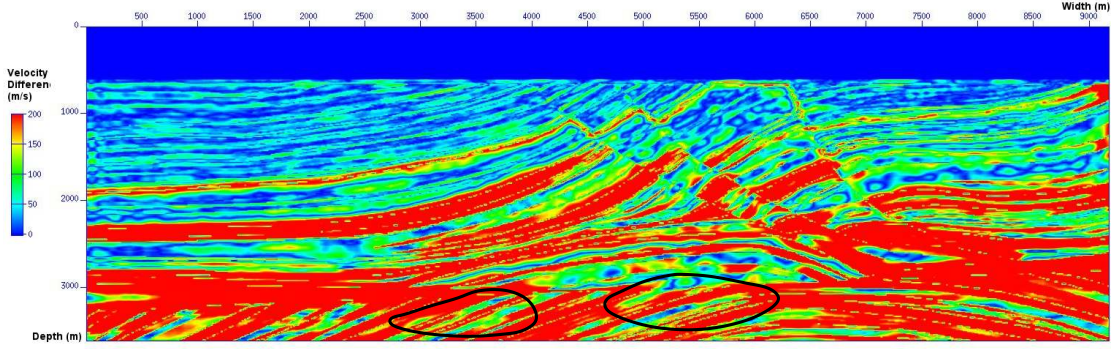


Figure 5.15: Difference between the solution of the L_1 based inversion with the real velocity model.

and to test the stability of the algorithms, we will now add random Gaussian noise to the input data. We apply noise levels of $s_n = 20$, $s_n = 10$, $s_n = 5$, where

$$s_n = \frac{\|U\|}{\|U_n\|} \quad (5.10)$$

is the signal to noise ratio for Gaussian noise U_n . (We removed the direct wave in U here, as it would completely dominate the nominator in s_n .) Figure 5.16 shows the effect of this noise on the seismograms.

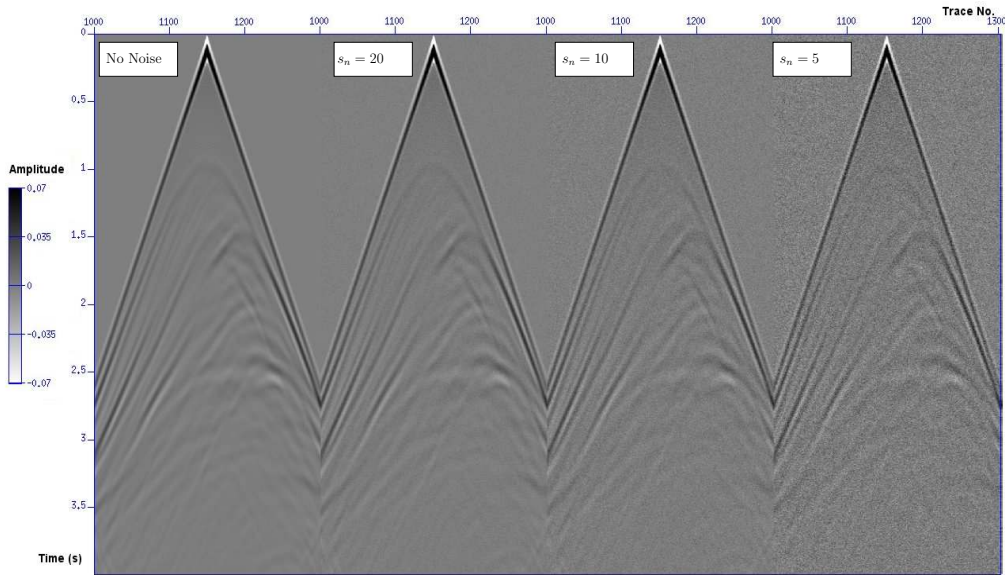


Figure 5.16: Receiver data for source at 4000 m with different noise levels.

We compare the effects of that noise with the L_1 and the L_2 norm over the first 15 iterations in Figure 5.17. We can see, that both algorithms converge for every noise level and the L_2 errors are similar for both minimization schemes for each noise level. However, there is a significant difference in the velocity error in Figure 5.18. While the velocity errors during the L_2 minimization stay close to each other for all

5. Test Cases

noise levels (Only $s_n = 5$ and no noise plotted), the velocity error for large s_n gets notably worse for the L_1 norm. However, even for $s_n = 5$ the velocity error with the L_1 norm stays below the error of the noise free L_2 scheme, therefore we can still recommend the L_1 norm over the L_2 norm in the presence of noise.

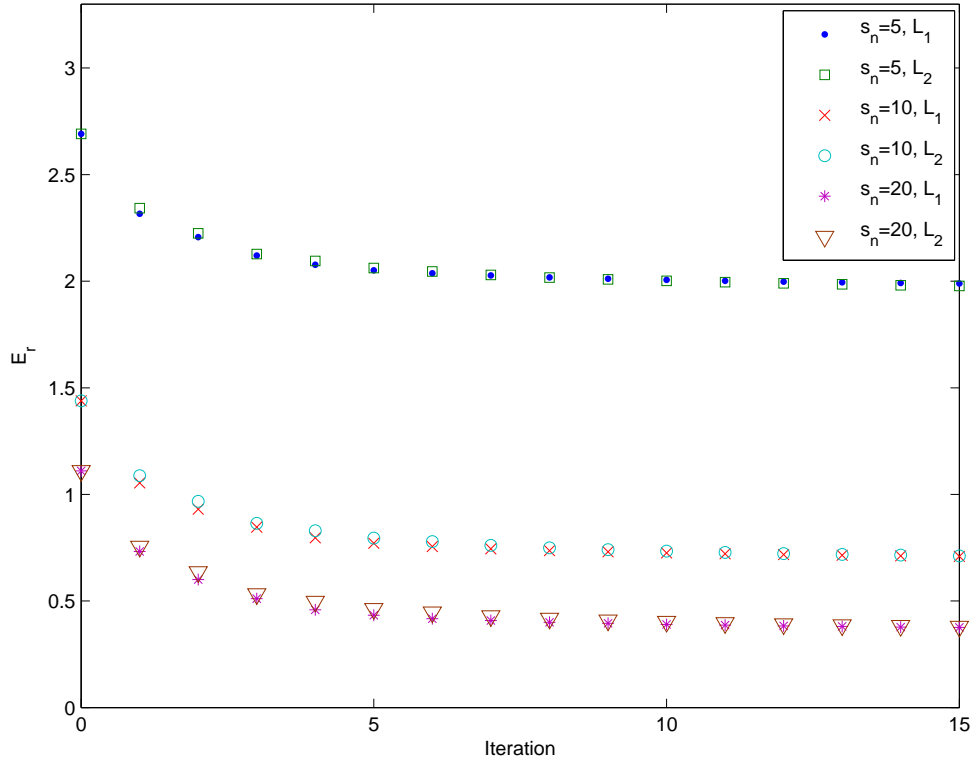


Figure 5.17: L_1 and L_2 minimization with noise. E_r is normalized with the misfit $E(c_0)$ of the noise free experiment.

5.1.5 Multiscale Schemes

Now, we want to make use of the complete available data and use the multiscale algorithms described in section 4.3 to avoid local minima.

Increasing Frequencies

We will now use the inversion schemes as we tested them before on increasing bands of frequencies. For each frequency band we run our algorithm until the stopping criterion is reached. Choosing good frequency bands for the inversion is not easy. Sirgue and Pratt developed in [SP04] a strategy to select single frequencies for their staggered frequency domain inversion based on 1D linearized assumptions. For a near offset survey with receiver offsets up to 3 km, they use 8 frequencies between 2 and 8 Hz, for a far offset survey with offsets up to 10 km, they manage with just

5. Test Cases

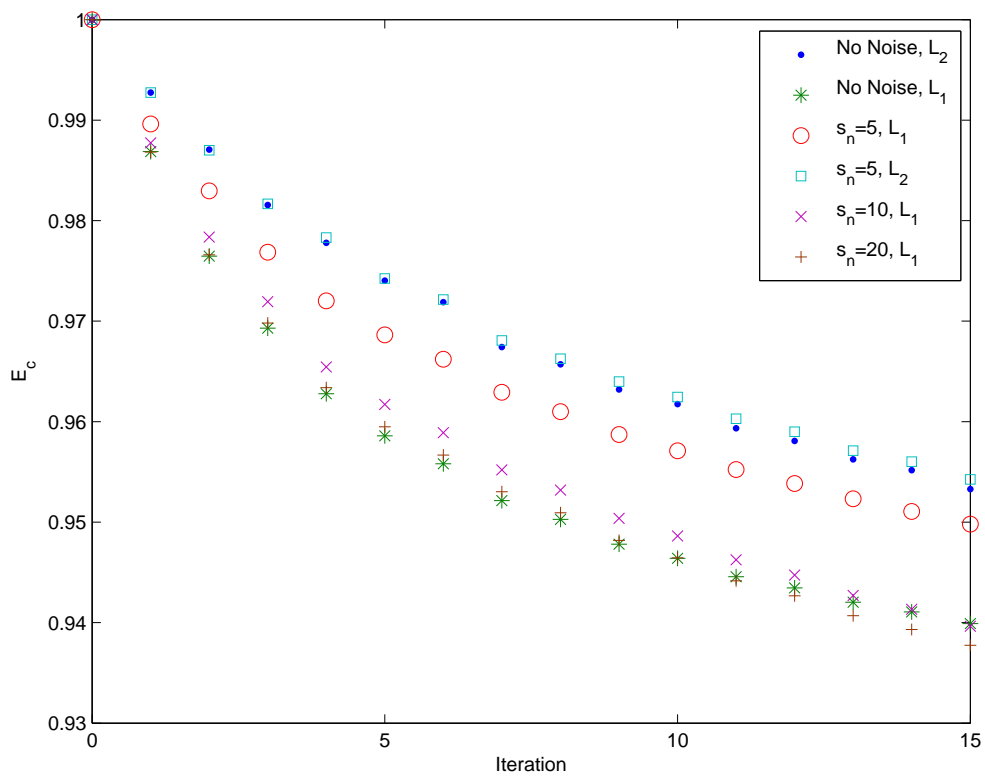


Figure 5.18: Velocity error in L_1 and L_2 minimization with noise.

5. Test Cases

three frequencies (2, 5, 8 Hz). They argue, that there is some redundancy in the data for which they can skip frequencies in their inversion algorithm.

Those results do not translate directly to our time domain approach, however it gives us an estimation for a reasonable upper limit to the number of bands needed. High and Starr ([VSK09]) argue that the redundancy in data during time domain inversion is an advantage in real data applications as it improves the signal to noise ratio. They use only three frequency bands to avoid local minima.

We have seen that the inversion with the 5 Hz frequency limit already explains the data at that range quite well, so we use our previous results as a starting model. We can determine the next higher filter frequency in a heuristic fashion. We compare several traces of the current synthetic seismogram with the real data as we did in Figure 5.11. We can then filter those seismograms to higher frequencies and visually check whether the filtered synthetic data lies within the range of attraction of the filtered real data, that is whether minima and maxima are within half a wave length next to each other. We can also do a Fourier transform and compare the frequency spectrum of the two datasets. In Figure 5.19 we can see how the inversion result departs from the input data with increasing frequency.

We now use an input signal with 15 Hz peak frequency, bandlimit it to 5, 7, 9, 12, and 15 Hz and perform the inversion successively for each frequency band. To guarantee comparability, we use the same grid spacing, adjusted to the highest frequency band for all frequencies. Table 5.1 lists the number of iterations done in each frequency band. After the 15 Hz band we perform the last iteration on the full dataset. (The 5 Hz solution and its stopping criterion varies slightly from the previous results, due to higher numerical accuracy in this test). To validate the whole scheme, we look again at a single trace and the frequency spectrum of the velocity model c_{end} after the last iteration with the L_1 norm. In Figure 5.20 we see the error in a single trace. It fits much better than before in Figure 5.11 and even the later part of the signal after 3 seconds is mostly matched well. Looking at the frequency spectrum in Figure 5.21, we also see considerable improvement. The section after 20 Hz is poorly matched, but this is not surprising since we stopped the systematic bandlimiting after 15 Hz.

The algorithm based on the absolute value norm reaches its stopping criterion faster than the least squares inversion on most frequency bands and finishes after 56 iterations, while we take 78 iterations for the least squares method. The L_1 method saves iterations in the early iterations on the lower frequency bands, where coarser grid space could be used. On the computationally more expensive higher frequency bands both methods needed the same amount of iterations, therefore the overall computation time of L_1 is only slightly faster than of the L_2 scheme. The relative error in the seismograms decreases over all multiscale steps as can be seen in Figure 5.22. In Figure 5.23 we see the velocity improvement over all iterations of the schemes. We immediately see that the L_1 norm produces better updates in fewer iterations. In some cases we see a small increase of the velocity error from one frequency band to the next. This might be a hint, that the previous velocity model was already too much refined for the first update steps in the next iteration. The problem could be reduced by either using more closer frequency bands or stopping the previous inversion earlier.

5. Test Cases

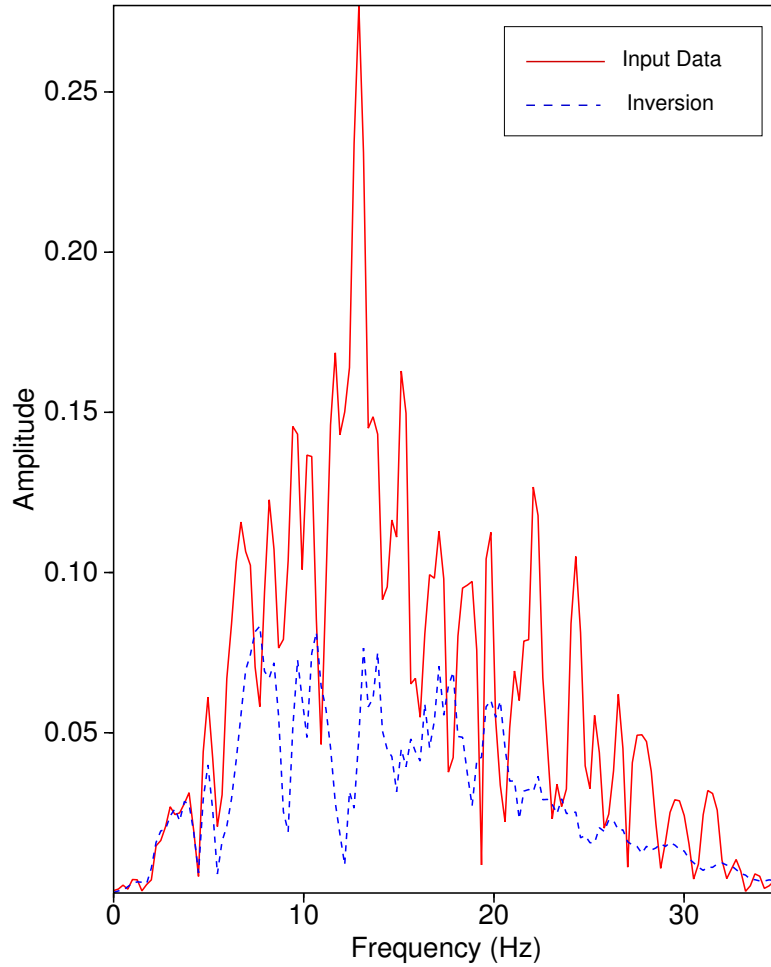


Figure 5.19: Full frequency spectrum of the traces from Figure 5.11. We see a very good match up to about 5 Hz.

Frequency	5 Hz	7 Hz	9 Hz	12 Hz	15 Hz	Full Range	Total
L_1	5	12	10	12	8	9	56
L_2	16	19	15	12	8	8	78

Table 5.1: Number of iterations at each frequency band for L_1 and L_2 based minimization.

5. Test Cases

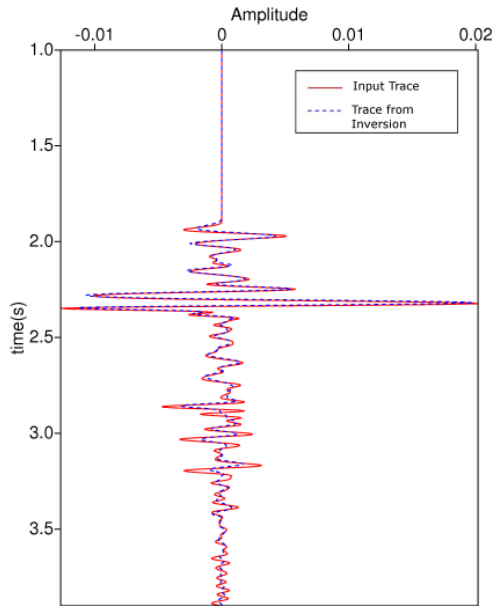


Figure 5.20: Single trace from receiver at 5500 m, source at 3000 m. Traces modeled with the real model and the inversion result from c_{end} after all iterations of the frequency multiscale L_1 scheme.

Decreasing time damping

Next, we will test the method of time damping described in 4.3.2. We do not apply a bandpass filter, instead we scale the gradient with an exponential damping term

$$\nabla_c E = \frac{2}{c^3} \int_{\Gamma} \int_0^T \lambda \frac{\partial^2}{\partial t^2} u \exp(-kt) dt, \quad (5.11)$$

and reduce the damping parameter k in the course of the inversion. Like in the previous section, we do not have an optimal algorithm to determine a set of damping parameters. Similar to the frequency approach, we could compare the input seismogram and the inversion results after applying time damping or after a Laplace transform. However, this could only indirectly give clues to the possible effects of that damping in the gradient.

In [SC08] Shin and Cha use 19 different Laplace coefficients between 20 and 2 in their staggered Laplace waveform inversion, when starting from a 1D velocity model which increases linearly with depth. With a better starting model and an overlapping approach, we decide to use fewer steps and iterate over 5 equally spaced Laplace coefficients between 10 and 0. Unlike in the previous section, where the misfit function changed with the frequency content the misfit function itself is not modified by the damping. We observed, that this makes it more difficult for us to adjust a stopping criterion that is effective for every given k . Instead, we just use a fixed set of 10 iterations that we perform per damping constant. We can see the effect of the damping on the gradient directly in Figure 5.24. As we expect, the gradient is mostly present in the upper parts of the model for strong damping and

5. Test Cases

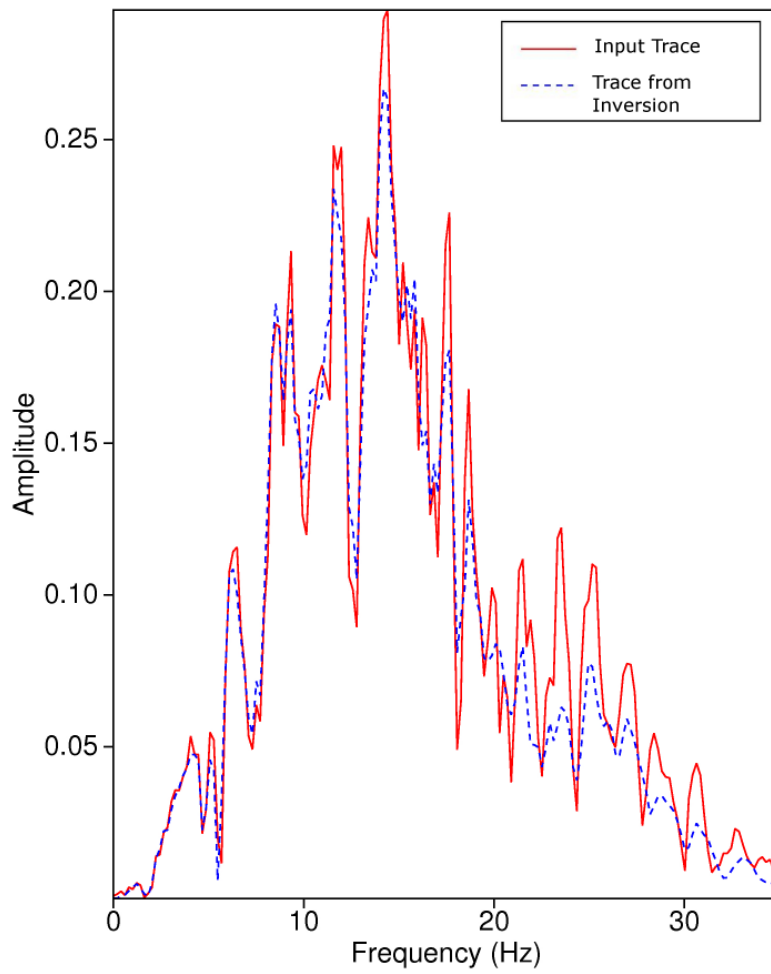


Figure 5.21: Full frequency spectrum of the traces from Figure 5.11.

5. Test Cases

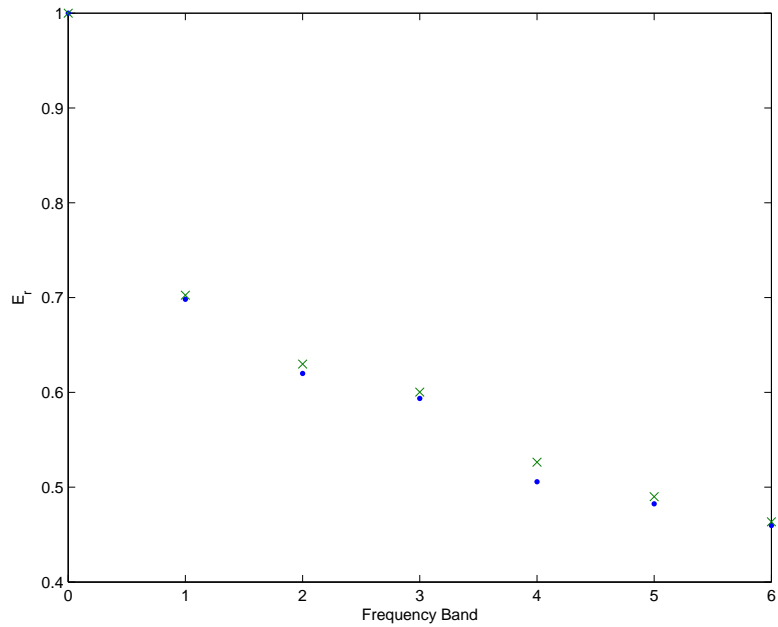


Figure 5.22: Relative L_2 error after each frequency band of the multiscale inversion computed with full frequency data. Frequency bands are 1) 5 Hz, 2) 7 Hz, 3) 9 Hz, 4) 12 Hz, 5) 15 Hz, 6) full range.

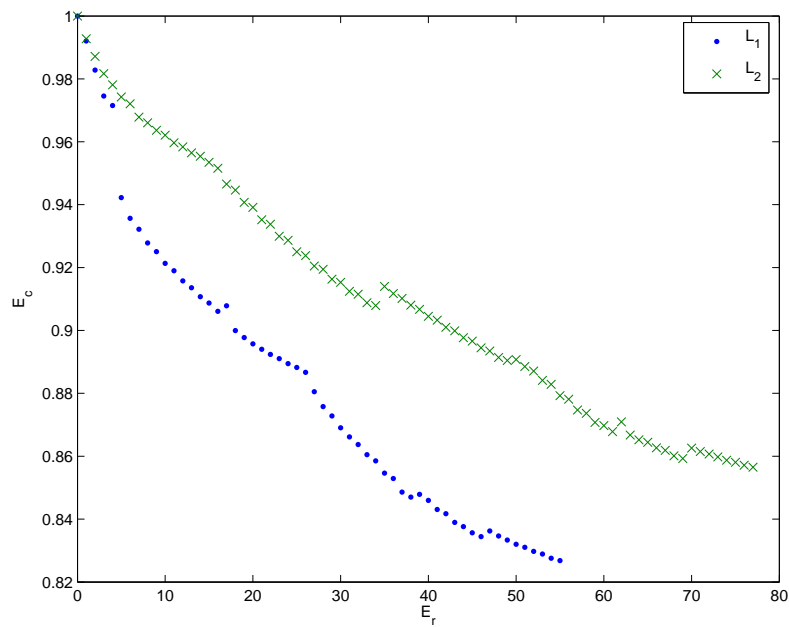
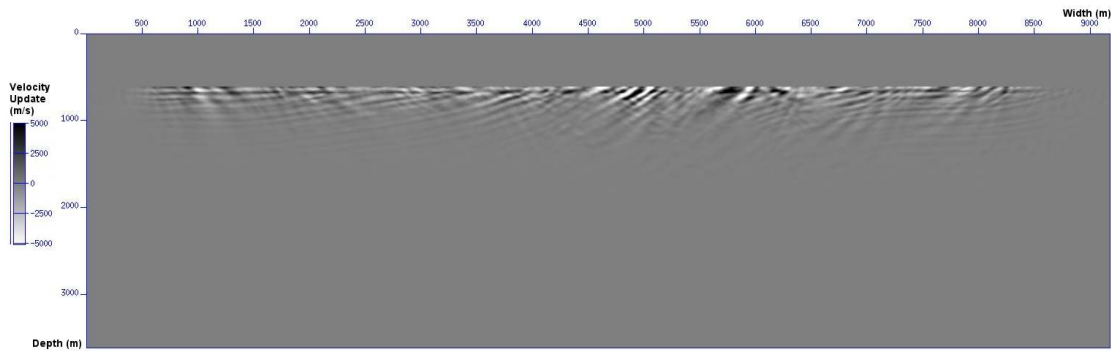


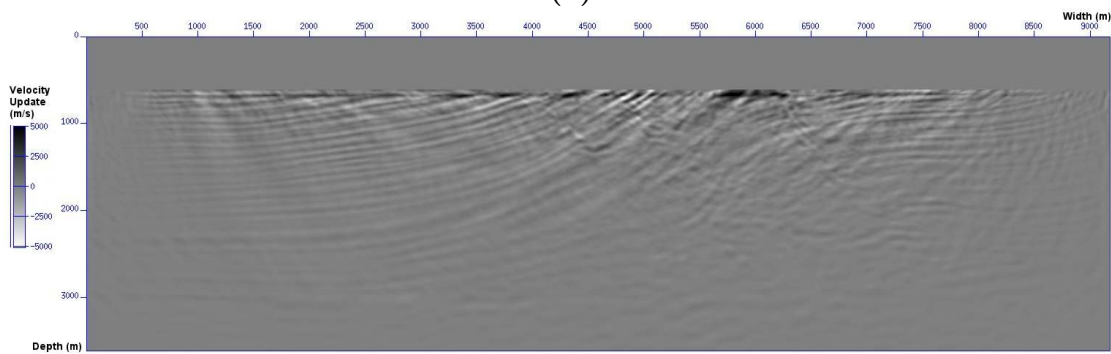
Figure 5.23: Difference between real velocity model and velocity model in each step of the multiscale inversion over frequency bands for L_1 and L_2 based inversion.

5. Test Cases

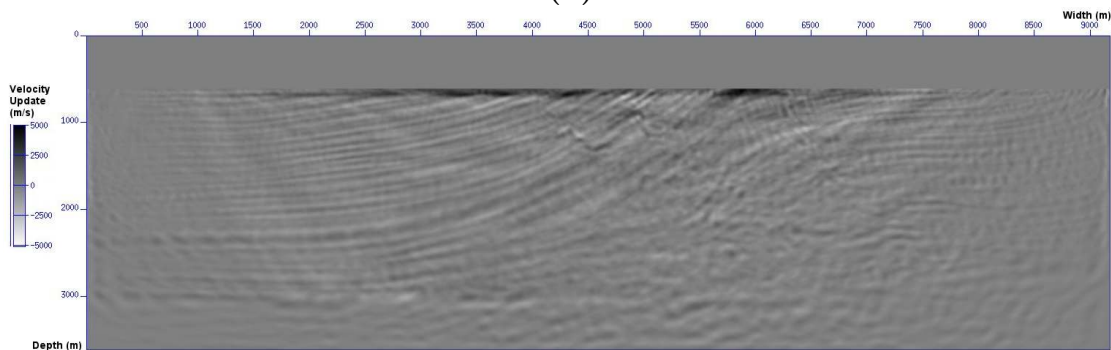
gets stronger in the lower parts of the model when we reduce k . We also note, that for the unfiltered data, there are some unusually strong contributions in the gradient that lead to velocity updates of up to 5000 m/s, which remain almost unaffected by the damping. We see in the decreasing misfit over the 50 iterations (Figure 5.25) that the method works, however the slow convergence in the beginning compared to the end suggests that a better choice of k might be possible. The total improvement in the misfit is also inferior to the update in the frequency scheme and the total velocity update we gain over the 50 iterations (measured in E_c) is only around 7%, while we had an update of almost 20% with the bandlimiting. While we can conclude that the strategy does work, we might at least want to combine it with a different strategy, like the bandlimiting to get better results.



(a)



(b)



(c)

Figure 5.24: First gradient of L_1 misfit without bandlimit. With Laplace damping of (a) $k = 10$, (b) $k = 5$ and (c) $k = 0$ (no damping).

5. Test Cases

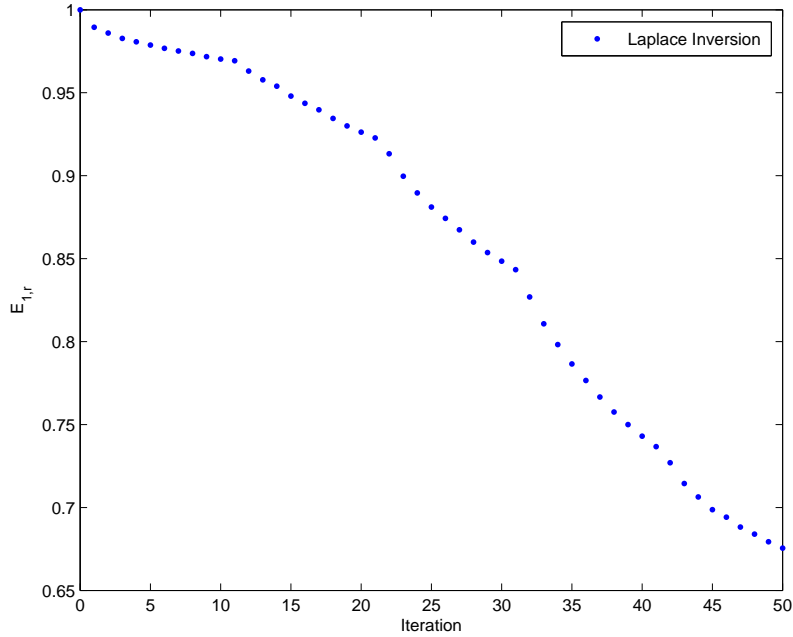


Figure 5.25: Relative L_1 error for the Laplace minimization with damping $k = -10, -7.5, -5, -2.5, 0$, decreasing every 10 iterations.

Offset Classes

As the last multiscale method we want to test, we iterate over offset classes, that is we start by using only long offsets and then add shorter offsets subsequently. Figure 5.26 shows the first gradient computed from only the offsets that are further apart than 4000 m. Unfortunately we see a lot of radial contributions in the gradient, that have nothing to do with the model. This can again be explained in light of Figure 5.2 and the discussion in the beginning of this chapter. By using only the the far offsets, we effectively reduce the streamer length. Instead of backpropagating the correct wavefield with the adjoint equation, the adjoint contains spherical approximations. Using this gradient would effectively generate updates, that worsen the velocity field, even though simultaneously decreasing the L_2 norm and thus stirring the algorithm towards a wrong minimum. Unlike in the beginning, we can see the effects deep in the model, so that any kind of mute or taper would be difficult.

Instead we can increase the number of sources. Instead of 8 sources, we now use 48 sources, distributed evenly like before. Although each new source produces similar artifacts, the overall gradient gets better, since correct contributions add up, while the artifacts are mostly uncorrelated.

We perform 15 iterations with offsets greater than 4000 m, then 15 iterations with offsets greater than 2000 m and finally 15 iterations with all offsets (L_1 error in Figure 5.27). Due to the different number of sources and data range, direct comparison with the previous tests would be unfair. However, we can see that we improve the misfit in E_1 with every offset cycle. The overall velocity update, measured in E_c is only around 6% and thus similar to our tests with the Laplace

5. Test Cases

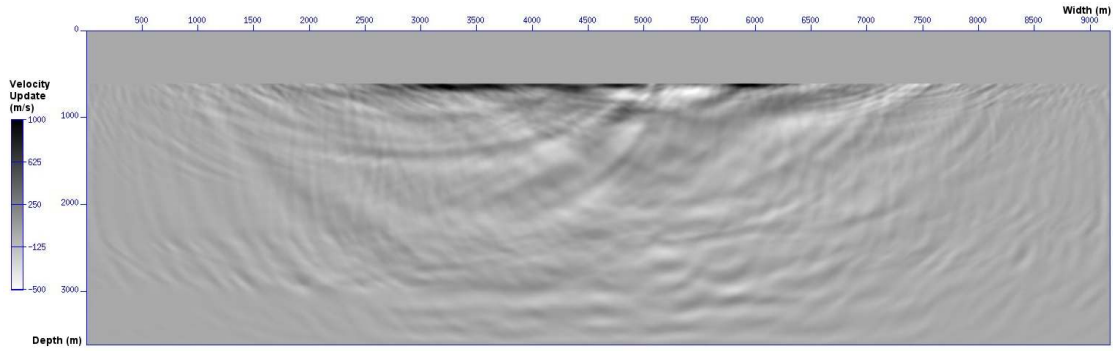


Figure 5.26: First gradient computed from 8 sources and offsets greater 4000 m.

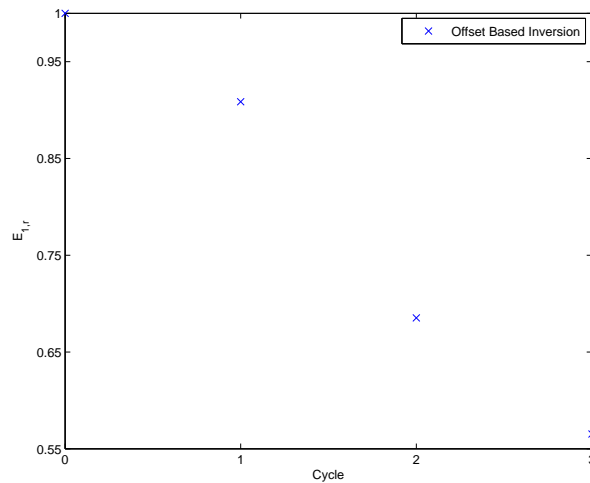


Figure 5.27: Relative L_1 error for the offset based minimization. The points are for cycles of 15 iterations with 1) $> 4000m$, 2) $> 2000m$, 3) all offsets.

5. Test Cases

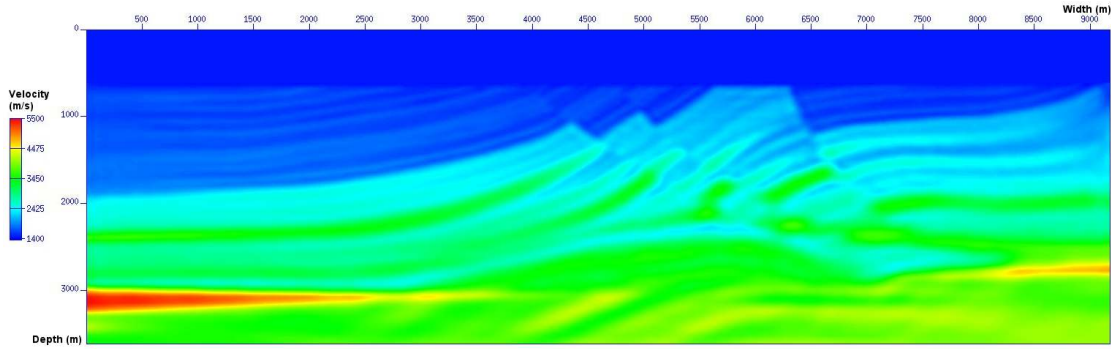


Figure 5.28: Final velocity model from combined multiscale scheme.

norm. We must assume that the artifacts, that are still visible with 48 sources, although reduced, still interfere with faster convergence.

Combined Multiscale Scheme

Finally we want to combine the three multiscale schemes described and tested above. There are many possibilities on how to combine them and it is not obvious how to proceed. We could use them in sequence, or nested, or both, we could use different convergence criteria or just use a fixed number of iterations with a certain method and save computational cost by proceeding earlier to the next step.

We noticed, that the offset based scheme produced artifacts and neither the Laplace nor the offset based scheme performed very good with high frequency content. Therefore, we use those schemes only for low frequencies. We start with 5 Hz and for each class of offsets with more than 3000 m, more than 1500 m, and all offsets, we perform a Laplace based inversion. Each Laplace inversion is done with five iterations with damping of $k = 10$, $k = 5$ and $k = 0$. After that we apply the frequency based inversion scheme with limits of 6, 7, 8, 10, and 15 Hz with stopping criterion of $\theta = 0.01$ (without additional offset reduction or time damping). For the whole process we use 48 sources, like in the offset based scheme.

Despite the additional effort, the results are not considerably better than in the frequency based scheme before. This again validates the heuristical methodology we applied in selecting the frequencies. Closer frequency bands, additional damping and more sources did not create a vastly superior model. We see the result in Figure 5.28, which simultaneously shows the best Marmousi result, we are able to create in this chapter. It resembles the original model very closely. Especially the top structures above 2000 m are clearly visible, but we also see some structure in the low areas below 3000 m. The salt flanks on the sides (the high velocity area below 3000 m at the left boundary) are not imaged very well. However, due to the boundary, this part of the model is not covered as well under the source/ receiver geometry as the rest of the model.

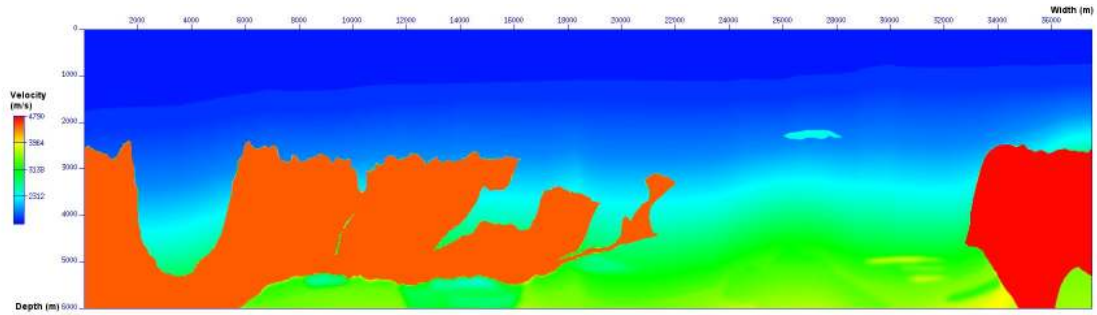
5.2 2004 BP Model

After we tested the different methods from Chapter 4 on the Marmousi model, we will now do a final test on a part of the 2004 BP model, shown in Figure 5.29 (a). The model and corresponding synthetic data was originally created by BP and provided as part of a blind test for different velocity building techniques, results of which were presented on the 2004 EAGE conference in Paris. We see a waterlayer with varying depth, a layer-based sediment background with a velocity that increases with depth and two (high velocity) salt structures. Good delineation of the salt bodies is the main challenge. Geologically this part of the model is motivated by the geology of deep water Gulf of Mexico. Further details on the whole model and the original test can be found in [BBD05].

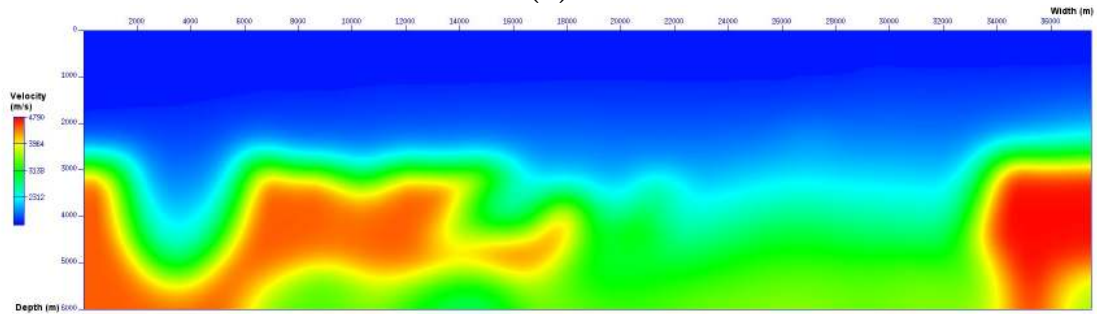
We use a 37 km by 6 km section of the original model and create our own synthetic data. Figure 5.29 (b) shows a smoothed version of the model that we use as a starting model. We do the modeling with 64 sources evenly spread over the surface and corresponding receivers with offsets up to 8 km. Both the model and our synthetic data is larger than in case of the Marmousi test case, furthermore the velocity contrasts, that we seek to recover are stronger and in deeper area, which makes the inversion more difficult.

We apply the multiscale scheme in frequency with the L_1 misfit function E_1 . Unlike in case of Marmousi, we start already at a lower frequency of 3 Hz. We iterate until the stopping criterion $\theta = 0.005$ is reached and proceed with frequency bands up to 4, 5, and 7 Hz. The stopping criterion is reached after 14, 15, 11, and 21 iterations respectively. So after a total of 61 iterations we obtain the result in Figure 5.29 (c). The salt structure in the middle is now very well visible. The top structure is delineated very well with clear edges. The velocity variations within the salt structure are less than before and even the bottom of the salt at a depth of around 5000 m was improved. Additionally there are slight improvements to the salt dome on the right, at 34 km and the small inclusion seen at 25-29 km and a depth of 2 km is now clearly visible. This inclusion has only low contrast to the background and is not visible in the smoothed initial model at all.

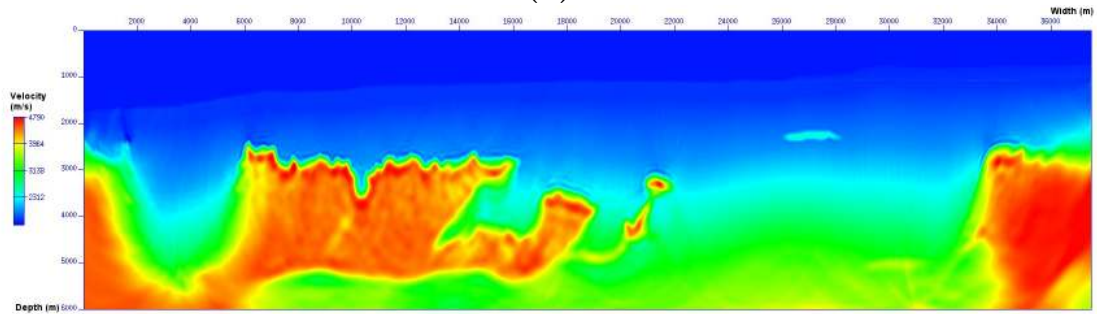
5. Test Cases



(a)



(b)



(c)

Figure 5.29: (a) 2004 BP velocity model. (b) Smoothed model from (a), that we use as starting model. (c) Final result after a total of 61 iterations with the multiscale scheme over 3, 4, 5, and 7 Hz.

6 Summary and Conclusion

Velocity model building from seismic data remains a challenging task. In this thesis, we discussed full waveform inversion, which is both the most accurate and simultaneously the most compute expensive technology under development today. We presented an efficient modeling algorithm for the two dimensional acoustic wave equation in time domain and tested different strategies for the inversion.

Modeling is by far the most time consuming part of full waveform inversion, therefore an efficient modeling algorithm is crucial. We introduced an alternating direction implicit (ADI) method to the seismic problem. ADI schemes can be seen as an intermediate between full implicit and explicit schemes. While we gain more accuracy than with explicit schemes (for given discretization) we do not have to solve as large systems of equations as implicit schemes require. Furthermore, we can easily parallelize the scheme on one compute node using threading and vectorization techniques. The stability conditions for the ADI scheme we introduced, demand similar time step sizes than the CFL condition does for explicit schemes, however, the method is less prone to dispersion so that we can use larger spatial step sizes. In tests, where an explicit scheme uses a 5 m spatial grid, we are able to obtain comparable results with a 10 m grid and a direct comparison of the two codes showed a runtime advantage on the side of our code.

In the second part of this work, we studied and compared different inversion techniques. The inverse problem is highly ill-conditioned and special care must be taken to avoid convergence towards a local minimum. In several tests we see superior results for the minimization of a missfit function based on the L_1 norm compared to the L_2 norm. We test three multiscale schemes which reduce the problem in early iterations by focussing on low frequency content, early time, or far offset. While we can demonstrate that all three techniques work, we obtained best results with the first technique, where we also have a heuristic way of choosing the steps in the multiscale scheme. Future work could focus more on how to determine parameters of those algorithms automatically and how to combine the different methods for optimal results.

We applied the discussed methods to the Marmousi model and to an extract of the 2004 BP model. Starting from a smoothed version of the original model, we were able to recover significant parts of both models. The final result of the Marmousi model shows clear structures down to a depth of below 2000 m and even some improvement near the bottom of the model. In the 2004 BP model we were able to recover both the high contrast top salt structures and a lower contrast inclusion accurately.

Notations

We use the following notations frequently in this dissertation. Additional variables are introduced, where they are used. We demonstrate the notation in two space dimensions here, which we use most often throughout this thesis. However, in some instances we explicitly restrict ourselves to one dimension or allow for the general case as noted where applicable.

Variables

$t \geq 0$ is the time.

$T \geq 0$ is the maximum recording time or maximum time for which the wave propagation is considered.

$z \geq 0$ is the depth. $z = 0$ is usually at the surface.

$Z \geq 0$ is the maximum depth considered, $Z(c, T)$ is the maximum depth a wave travels to in material c during time interval $[0, T]$.

h is the spatial discretization.

k is the discretization in time.

Ω is the domain of interest.

Ω_h is the discretization of Ω .

$\Gamma_s \subset \Omega$ is the set of source positions.

$\Gamma_{r,s} \subset \Omega$ is the set of receiver positions for a source s .

$\Gamma = \Gamma_s \times \Gamma_{r,s}$ is the set of all source-receiver pairs in the survey.

x is the horizontal distance when using components (x, z) , or in general a vector $x \in \Omega$.

$c(x, z)$ is the wavespeed in the medium.

$u(x, z, t)$ is the analytical solution to the wave equation. The physical meaning depends on the wave equation. In our case it can be seen as the pressure or excitation.

$U = u|_{\Gamma_{r,s}}$ is the response in the receiver positions.

U_d is the measured response at the surface (at receiver positions $\Gamma_{r,s}$).

$f(x, t)$ is the source function, usually at $z = 0$.

$\lambda(x, z, t)$ is the solution to the adjoint equation.

Error functions

E is the misfit between the analytical solution and the observed data in the receiver spaces. Unless otherwise noted, we use $E = \int_{\Gamma} \int_0^T (U_d - U(c))^2 dt$.

$E_r(c_k) = \frac{E(c_k)}{E(c_0)}$ is the relative error with regard to the initial velocity model in the iterative scheme.

$$E_1 = \int_{\Gamma} \int_0^T |U_d - U(c)| dt.$$

$$E_{1,r}(c_k) = \frac{E_1(c_k)}{E_1(c_0)} \text{ the relative error for the } L_1 \text{ norm.}$$

$$E_{Grad}(c_k) = \frac{\|\nabla_c E(c_k)\|}{\|\nabla_c E(c_0)\|}.$$

Further Mathematical Symbols

$\Delta u = \frac{\partial^2}{\partial x^2} u + \frac{\partial^2}{\partial z^2} u$ is the Laplace operator.

$\nabla_c u$ is the gradient/ first derivative of u with respect to function c .

$\nabla_c^2 u = \nabla_c(\nabla_c u)$ is the second derivative with respect to c .

L denotes a linear differential operator.

L_h is the discretization of L

$(a, b) = \{x \in \mathbb{R} | a < x < b\}$ is the open interval for $a, b \in \mathbb{R}$.

$[a, b] = \{x \in \mathbb{R} | a \leq x \leq b\}$ is the closed interval for $a, b \in \mathbb{R}$.

$$\|u\|_{L^p(\Omega)} = \left\langle \int_{\Omega} |u|^p \right\rangle^{\frac{1}{p}}.$$

$$L^p(\Omega) = \{u : \Omega \rightarrow \mathbb{R} | u \text{ is Lebesgue measurable, } \|u\|_{L^p(\Omega)} < \infty\}.$$

Bibliography

- [AKB74] R. M. Alford, K. R. Kelly, and D. M. Boore, *Accuracy of finite-difference modeling of the acoustic wave equation*, *Geophysics* **39** (1974), 834–842.
- [BBD05] F. J. Bilette and S. Brandsberg-Dahl, *The 2004 bp velocity benchmark*, 67th Annual EAGE Meeting, EAGE, Expanded Abstracts, 2005.
- [BCL79] A. Bamberger, G. Chavent, and P. Lailly, *About the stability of the inverse problem in 1-d wave equations—application to the interpretation of seismic profiles*, *Applied Mathematics & Optimization* **5** (1979), 1–47.
- [BCSJ01] N. Bleistein, J. K. Cohen, and J. W. Stockwell Jr., *Mathematics of multidimensional seismic inversion*, Springer, Berlin, 2001.
- [BHM00] W. L. Briggs, Van E. Henson, and S. F. McCormick, *A multigrid tutorial*, Society for Industrial and Applied Mathematics, 2000.
- [BL97] K. P. Bube and R. T. Langan, *Hybrid l_1/l_2 minimization with applications to tomography*, *Geophysics* **62** (1997), 1183–1195.
- [BOV10] R. Brossier, S. Operto, and J. Virieux, *Which data residual norm for robust elastic frequency-domain full waveform inversion?*, *Geophysics* **75** (2010), R37–R46.
- [Bra77] A. Brandt, *Multi-level adaptive solutions to boundary value problems*, *Math. Comp* **31** (1977), 333–390.
- [BSZC95] C. Bunks, F. M. Saleck, S. Zaleski, and G. Chavent, *Multiscale seismic waveform inversion*, *Geophysics* **60** (1995), 1457–1473.
- [CC88] P. Cary and C. Chapman, *Automatic 1-d waveform inversion of marine seismic refraction data*, *Geophysical Journal of the Royal Astronomical Society* **93** (1988), 527–546.
- [CE77] R. Clayton and B. Engquist, *Absorbing boundary conditions for acoustic and elastic wave equations*, *Bulletin of the Seismological Society of America* **67** (1977), 1529–1540.
- [CMS08] Y. Choi, D.-J. Min, and C. Shin, *Two-dimensional waveform inversion of multi-component data in acoustic-elastic coupled media*, *Geophysical Prospecting* **56** (2008), 863–881.
- [CPN⁺90] E. Crase, A. Pica, M. Noble, J. McDonald, and A. Tarantola, *Robust elastic nonlinear waveform inversion: Application to real data*, *Geophysics* **55** (1990), 527–238.

Bibliography

- [DBNC89] C. R. Daudt, L. W. Braile, R. L. Nowack, and C. S. Chiang, *A comparison of finite-difference and fourier method calculations of synthetic seismograms*, Bulletin of the Seismological Society of America **79** (1989), 1210–1230.
- [DE11] J. Diaz and A. Ezziani, *Documentation for gar6more2d*, May 2011.
- [dH60] A. T. de Hoop, *A modification of cagniards method for solving seismic pulse problems*, Applied Scientific Research, Section B. **8** (1960), 349–356.
- [Dzu10] G. Dzuik, *Theorie und numerik partieller differentialgleichungen*, de Gruyter, 2010.
- [Eva98] L. C. Evans, *Partial differential equations*, American Mathematical Society, 1998.
- [Fic11] A. Fichtner, *Full seismic waveform modelling and inversion*, Springer, Berlin, 2011.
- [FM65] G. Fairweather and A. R. Mitchell, *A high accuracy alternating direction method for the wave equation*, IMA Journal of Applied Mathematics **1** (1965), 309–316.
- [GV10] D. Gisolf and E. Verschuur, *The principles of quantitative acoustical imaging*, EAGE Publications, 2010.
- [GW00] A. Griewank and A. Walther, *An implementation of checkpointing for the reverse or adjoint mode of computational differentiation*, ACM Transactions on Mathematical Software **26** (2000), 19–45.
- [HPYS09] W. Ha, S. Pyun, J. Yoo, and C. Shin, *Robustness of laplace-domain waveform inversion for coherent noisy data*, SEG Expanded Abstracts **28** (2009), 2253–2258.
- [HS52] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, Journal of Research of the National Bureau of Standards **49** (1952), 409–436.
- [Hub73] P. J. Huber, *Robust regression: Asymptotics, conjectures and monte carlo*, The Annals of Statistics **1** (1973), 799–821.
- [IDT88] J. Ikelle, J. P. Diet, and A. Tarantola, *Linearized inversion of multioffset seismic reflection data in the $\omega - k$ domain: Depth-dependent reference medium*, Geophysics **53** (1988), 50–64.
- [Jon10] I. F. Jones, *An introduction to velocity model building*, EAGE Publications, 2010.
- [KAH⁺09] J. R. Krebs, John E. Anderson, D. Hinkley, A. Baumstein, and S. Lee, *Fast full wave seismic inversion using source encoding*, SEG Expanded Abstracts **28** (2009), 2273–2277.

Bibliography

- [KCL86] P. Kolb, F. Collino, and P. Lailly, *Prestack inversion of 1-d medium*, Proceedings of the IEEE (1986), 498–508.
- [KG06] B. Kaelin and A. Guitton, *Imaging condition for reverse time migration*, SEG Annual Meeting, 2006.
- [KSCP09] N.-H. Koo, C. Shin, Y. Ho Cha, and K. Park, *Sequentially ordered single-frequency 2-d acoustic waveform inversion in the laplace-fourier domain*, SEG Expanded Abstracts **28** (2009), 2248–2252.
- [Lai83] P. Lailly, *The seismic inverse problem as a sequence of pre-stack migration*, Society of Industrial and Applied Mathematics, Bednar, J.B., Redner, R., Robinson, E. and Weglein, A., Eds., Conference on inverse scattering: Theory and Applications., 1983.
- [Lee62] M. Lees, *Alternating direction methods for hyperbolic differential equations*, Journal of the Society for Industrial and Applied Mathematics **10** (1962), 610–616.
- [LG00] G. Liu and S. D. Gedney, *Perfectly matched layer media for an unconditionally stable three-dimensional adi-fdtd method*, IEEE Microwave and Guided Wave Letters **10** (2000), 261–263.
- [LRV90] P. Lailly, F. Rocca, and R. Versteeg, *The marmousi experience: synthesis*, EAEG Workshop on Practical Aspects of Seismic Data Inversion, 1990.
- [LS09] D. Lee and C. Shin, *The direct arrival removal method of laplace waveform inversion*, Korean Society for Industrial and Applied Mathematics (KSIAM) 2009 Spring Conference, 2009.
- [McC88] S. F. McCormick, *Multigrid methods: Theory, applications, and supercomputing: Lecture notes in pure and applied mathematics*, CRC, 1988.
- [Mor87] P. R. Mora, *Nonlinear two-dimensional elastic inversion of multi-off-set seismic data*, Geophysics **53** (1987), 1211–1228.
- [Mor88] ———, *Elastic wavefield inversion of reflection and transmission data*, Geophysics **53** (1988), 750–759.
- [PDT90] A. Pica, J. P. Diet, and A. Tarantola, *Nonlinear inversion of seismic reflection data in a laterally invariant medium*, Geophysics **55** (1990), 284–292.
- [Ple06] R.-E. Plessix, *A review of the adjoint-state method for computing the gradient of a functional with geophysical applications*, Geophysical Journal International **167** (2006), 495–503.
- [Pra99] R. G. Pratt, *Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model*, Geophysics **64** (1999), 888–901.

Bibliography

- [PS99] R. G. Pratt and R. M. Shipp, *Seismic waveform inversion in the frequency domain, part 2: Fault delineation in sediments using crosshole data*, *Geophysics* **64** (1999), 902–914.
- [PSH98] R. G. Pratt, C. Shin, and G. J. Hicks, *Gauss-newton and full newton methods in frequency-space seismic waveform inversion*, *Geophysical Journal International* **133** (1998), 341–362.
- [PSWW96] R. G. Pratt, Z. M. Song, P. R. Williamson, and M. Warner, *Two-dimensional velocity model from wide-angle seismic data by wavefield inversion*, *Geophysical Journal International* **124** (1996), 323–340.
- [PW90] R. G. Pratt and M. H. Worthington, *Inverse theory applied to multi-source cross-hole tomography*, *Geophysical Prospecting* **38** (1990), 287–310.
- [Rey78] A. C. Reynolds, *Boundary conditions for the numerical solution of wave propagation problems*, *Geophysics* **43** (1978), 1099–1110.
- [Rob05] M. Roberts, *A line source to point source transform for seismograms*, *Lithos Science Report* **7** (2005), 49–52.
- [Rob10] E. Robein, *Seismic imaging*, EAGE Publications, 2010.
- [Rya94] H. Ryan, *Ricker, ormsby, klauder, butterworth - a choice of wavelets*, *CSEG Recorder* (1994), 8–8.
- [SBD⁺10] L. Sirgue, O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, *Full waveform inversion: the next leap forward in imaging at valhall*, *First Break* **28** (2010), 65–70.
- [SC08] C. Shin and Y. H. Cha, *Waveform inversion in the laplace domain*, *Geophysical Journal International* **173** (2008), 922–931.
- [SC09] C. Shin and Y. H. Cha, *Waveform inversion in the laplacefourier domain*, *Geophysical Journal International* **3** (2009), 1067–1079.
- [SD92] M. Sambridge and G. Drijkoningen, *Genetic algorithms in seismic waveform inversion*, *Geophysical Journal International* **109** (1992), 323–342.
- [SOV⁺09a] F. Sourbier, S. Operto, J. Virieux, P. Amestoy, and J. Y. L'Excellent, *Fwt2d: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data—part 1: Algorithm*, *Computers & Geosciences* **35** (2009), no. 3, 487 – 495.
- [SOV⁺09b] ———, *Fwt2d: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data—part 2: Numerical examples and scalability analysis*, *Computers & Geosciences* **35** (2009), no. 3, 496–514.

Bibliography

- [SP04] L. Sirgue and R. G. Pratt, *Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies*, *Geophysics* **69** (2004), 231–248.
- [SS88] F. Santosa and W. W. Symes, *Computation of the hessian for least-squares solutions of inverse problems of reflection seismology*, *Inverse Problems* **4** (1988), 211–233.
- [Tar84] A. Tarantola, *Inversion of seismic reflection data in the acoustic approximation*, *Geophysics* **49** (1984), 1259–1266.
- [Tar05] ———, *Inverse problem theory*, SIAM, 2005.
- [TLMT10] C. Tape, Q. Y. Liu, A. Maggi, and J. Tromp, *Seismic tomography of the southern california crust based on spectral-element and adjoint methods*, *Geophysical Journal International* **180** (2010), 433–462.
- [Ver94] R. Versteeg, *The marmousi experience: Velocity model determination on a synthetic complex data set*, *The Leading Edge* **13** (1994), 927–936.
- [VO09] J. Virieux and S. Operto, *An overview of full-waveform inversion in exploration geophysics*, *Geophysics* **74** (2009), 127–152.
- [VS08] D. Vigh and E. W. Starr, *3d prestack plane-wave, full-waveform inversion*, *Geophysics* **73** (2008), 135–144.
- [VSK09] D. Vigh, E. W. Starr, and J. Kapoor, *Developing earth models with full waveform inversion*, *The Leading Edge* **28** (2009), 432–435.
- [WT03] S. Wang and F. L. Teiseira, *An efficient pml implementation for the adi-fdtd method*, *IEEE Microwave and Wireless Components Letters* **13** (2003), 72–74.
- [Yil01] Ö. Yilmaz, *Seismic data analysis*, Society of Exploration Geophysicists, 2001.

Curriculum Vitae

- 8/1992 - 6/2004 Schulausbildung. Abschluss: Allgemeine Hochschulreife
- 10/2004 - 5/2009 Studium der Mathematik mit Nebenfach Physik an der Universität zu Köln. Abschluss: Diplom Mathematiker
- 8/2007 - 4/2008 Studium an der North Carolina State University, USA
- 4/2007 - 7/2007, 5/2008 - 6/2009 Studentische Hilfskraft und Diplomarbeit am Fraunhofer Institut für Algorithmen und wissenschaftliches Rechnen SCAI
- 9/2009 - 12/2012 Promotionsstudium an der Technischen Universität Kaiserslautern mit Promotionsstipendium vom Fraunhofer Institut für Techno- und Wirtschaftsmathematik ITWM
- 5/2011 - 8/2011, 1/2012 - 3/2012 Forschungsaufenthalt am Statoil Forschungszentrum Trondheim, Norwegen

Curriculum Vitae

- 8/1992 - 6/2004 School education. Degree: Allgemeine Hochschulreife
- 10/2004 - 5/2009 Study of mathematics with minor in physics at Cologne University. Degree: Diplom Mathematiker
- 8/2007 - 4/2008 Studies at the North Carolina State University, USA
- 4/2007 - 7/2007, Student assistant and diploma thesis at the Fraunhofer Institute
5/2008 - 6/2009 for Scientific Computing and Algorithms SCAI
- 9/2009 - 12/2012 Doctoral studies at the University Kaiserslautern, scholarship
from Fraunhofer Institute for Industrial Mathematics ITWM
- 5/2011 - 8/2011, Research visit at the Statoil Research Center Trondheim, Nor-
1/2012 - 3/2012 way