Adaptive Eigenspace for Inverse Problems in the Frequency Domain

Inauguraldissertation

zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Uri Nahum
aus Yavne, Israel

Basel, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Marcus J. Grote
Prof. Dr. Eldad Haber

Basel, den 20.9.2016

Prof. Dr. Jörg Schibler
Dekan

Abstract

Inverse scattering problems are used in a vast number of applications, such as geophysical exploration and medical imaging. The goal is to recover unknown media using wave propagation. The inverse problem is designed to minimize simulated data with observation data, using partial differential equations (PDE) as constrains. The resulting minimization problem is often severely ill-posed and contains a large number of local minima. To tackle ill-posedness, several optimization and regularization techniques have been explored. However, the applications are still asking for improvement and stability.

In this thesis, a nonlinear optimization method is proposed for the solution of inverse scattering problems in the frequency domain, when the scattered field is governed by the Helmholtz equation. The time-harmonic inverse medium problem is formulated as a PDE-constrained optimization problem and solved by an inexact truncated Newton-type method. Instead of a grid-based discrete representation, the unknown wave speed is projected to a particular finite-dimensional basis, which is iteratively adapted during the optimization. Truncating the adaptive eigenspace (AE) basis at a (small and slowly increasing) finite number of eigenfunctions effectively introduces regularization into the inversion and thus avoids the need for standard Tikhonov-type regularization. We actually show how to build an AE from the gradients of Tikhonov-regularization functionals.

Both analytical and numerical evidence underpin the accuracy of the AE representation. Numerical experiments demonstrate the efficiency and robustness to missing or noisy data of the resulting adaptive eigenspace inversion (AEI) method. We also consider missing frequency data and apply the AEI to the multi-parameter inverse scattering problem.

Contents

1 Introduction **11**
1.1 Applications of the Inverse Scattering Problem 11
1.1.1 PDE-Constrained Optimization 12
1.2 Numerical Methods 13
1.2.1 Qualitative Methods 13
1.2.2 Quantitative Methods 14
1.2.3 Optimization Strategies for the Inverse Helmholtz Problem 15
1.3 Regularization Techniques 15
1.3.1 Regularization by Filtering and Discretization 16
1.3.2 Tikhonov Regularization 16
1.3.3 Size Reduction and Adaptive Regularizations 17
1.4 Adaptive Eigenspace (AE) 18
1.5 Multi-Parameter Inverse Problem 18
1.6 Thesis Outline 19
1.7 Main Contributions 20

I Optimization Techniques for Inverse Scattering Problems in the Frequency Domain **21**

2 Inverse Scattering in the Frequency Domain **23**
2.1 Forward Problem 23
2.1.1 Finite Element Discretization 24
2.1.2 Finite Differences Discretization 25
2.2 Inverse Problem 27
2.2.1 Full-Space Approach 28
2.2.2 Reduced-Space Approach 29

2.3 Inexact Newton Methods for Inverse Helmholtz Problems 34
2.4 Inexact (Quasi)-Newton Algorithm 39
2.5 Regularization Problems 40

3 Regularization of the Inverse Problem 41
3.1 Tikhonov Regularization for Inverse Helmholtz Problems 41
3.2 Numerical Experiments 44

II Adaptive Eigenspace for the Inverse Helmholtz Equation 47

4 Adaptive Eigenspace Inversion (AEI) 49
4.1 Adaptive Eigenspace Basis 49
4.2 Adaptive Eigenspace Inversion Algorithm 51

5 Approximation Properties of the AE Basis 53
5.1 One-Dimensional Case 53
5.2 Two-Dimensional Case 55

6 Numerical Experiments 61
6.1 Adaptive Eigenspace vs. Nodal Basis 62
6.2 Adaptive Finite Element Discretization 64
6.3 Adaptive vs. Harmonic Background $\mathbf{u_0}$ 64
6.4 Constant vs. Adaptive Dimension of V_K 65
6.5 Sparse AEI Method 67
6.6 Noisy Data 68
6.7 Sample Average Approximation 70
6.8 Computational Cost 71

7 AEI for the Limited Frequency Data 73
7.1 Single Frequency AEI (SF-AEI) 73
7.2 Constant Background Numerical Experiments 74
7.3 Single Frequency AEI Algorithm 76
7.4 Numerical Experiments 77

8 AE of Gradients of Penalty Functionals 81
8.1 AE from other Penalty Functionals 81

8.2 Numerical Experiments 84

III Multiparameter Inverse Helmholtz Problem 89

9 AEI with Multi-Parameter 91

9.1 The Multi-Parameter Inverse Helmholtz Problem 92

9.2 Cross-Talk 93

9.3 AEI for the Multi-Parameter Problem 98

9.4 Numerical Experiments 99

Conclusions 105

Bibliography 106

Acknowlegdment

I would like to mention the people without whom the writing of this thesis would not have been possible and thank them from the bottom of my heart for taking part in this journey.

First, I would like to thank Prof. Dr. Marcus J. Grote for giving me the opportunity to do my PhD under his supervision, for his guidance, support, advice, and useful interesting meetings. It is a real pleasure and very inspiring to work with you. I would also like to thank Prof. Dr. Eldad Haber for agreeing to act as a co-referee in my thesis committee, for his talk in Ascona 2013 and our meeting in Zurich 2015 which inspired me in my thesis.

Great thanks go to my colleagues at the Department of Mathematics and Computer Science at the University of Basel, it was great working with you all: Dr. Loredana Gaudio, Dr. Teodora Mitkova, Dr. Gabriele Mancini, Jet Hoe Tang and Tobias Arnold. Special thanks goes to my office mate Dr. Marie Graff-Kray and for my colleague Dr. Michaela Mehlin for the good work together, for the helpful discussions and emotional support. Thank you Dr. Johannes Huber for the many fruitful discussions.

Many thanks go to Dr. Marie Graff-Kray, Jürgen Dölz, Dr. Johannes Huber, and Dr. Michaela Mehlin for proofreading this thesis.

A great thank you goes to my wife Noemi, for her love, support and understanding. Thank you for being there for me with patience and a big warm hug, whenever it is needed. You are always there for me and I am forever grateful for that. Another thanks goes to my children Ella and Alon, who are able to make me smile even in the most stressful moments. A big thanks goes to my sisters Nira Nahum-Cohen and Heli Nahum-Shimoni for their encouragement. I would also like to thank my parents-in-law Annkathrin Zwygart, Andreas Müller, Daniel Zwygart, and Susanne Härri for their support, especially in the time of writing this thesis. A last special thanks goes to my parents Avner and Shoshana Nahum:

כל הישג בחיי הוא תוצר החינוך שהקניתם לי. אני מוקיר לכם תודה על כל מה שעשיתם עבורי בחיים.
אני אוהב ומעריץ אתכם.

This thesis was written at the Department of Mathematics and Computer Science at the University of Basel and was partly supported by the Swiss National Science Foundation.

Chapter 1

Introduction

1.1 Applications of the Inverse Scattering Problem

Inverse scattering problems occur in a wide range of applications such as radar [79, 2] and sonar technology [54], non-destructive testing, geophysical exploration [80, 36, 86], and medical imaging [1, 5, 91, 61, 85]. By illuminating an unknown body, the scatterer, with waves of various directions or wavelengths, one attempts to obtain information about that body from the scattered waves recorded at some distance. The scatterer is a penetrable, bounded inhomogeneity inside the medium characterized by one or several varying physical parameters and the inverse problem consists in estimating these parameters from scattering data. This problem is also known as *inverse medium problem* or *inverse scattering problem*.

In general, inverse medium problems are solved by the following strategy: A known acoustic source produces a wave which travels through the medium. The wave collects information about the medium by propagating through its layers and features or by reflecting from them. The wave and its reflections are then measured at sensors outside the medium. After collecting the data, the measurements are compared with a simulation of the wave propagating through an estimated medium, which is chosen as an initial guess. The misfit between the simulated data and the true measurements is minimized to reveal the properties and structure of the medium.

Inverse medium problems occur in oil and gas exploration in geophysics or breast tumor detection in medical imaging. In Fig. 1.1 we illustrate the inverse medium problem for both applications. The left picture of Fig. 1.1 shows sound waves being shot into the ocean from a boat and the measurements of the reflected waves being collected by hydrophones behind it[1]. The inverse scattering problem is also known as full wave inversion (FWI) in the geophysical community. On the right of Fig. 1.1 we illustrate the inverse medium problem for the detection of a tumor in a human breast[2]. The breast is composed mainly of fatty tissue, muscles and lymph nodes. The different kinds of tissues inside the breast have

[1] `www.rigzone.com`

[2] `www.cancer.gov`

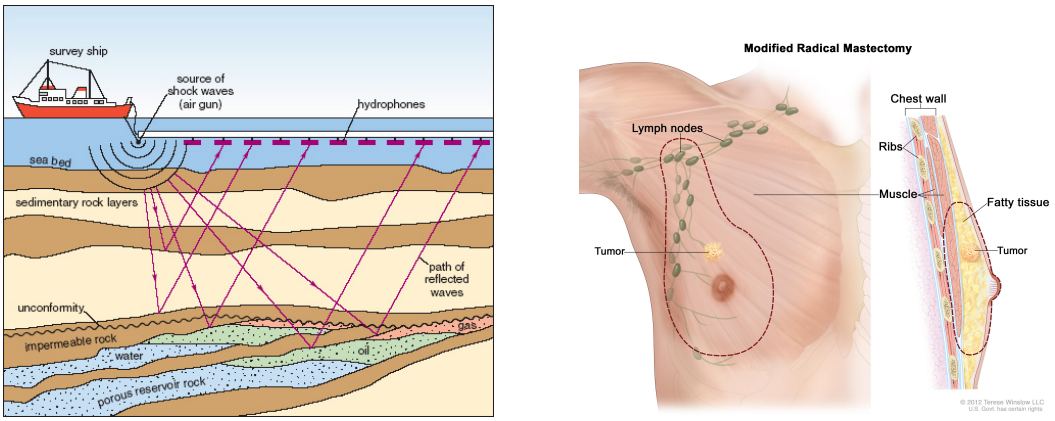

Figure 1.1: Illustration of the inverse medium problem. Seismology survey reveals the layers of the Earth and searches for oil or gas (left). Inverse medium problem detects possible tumor in a human breast, where different tissues inside the breast have different wave velocities (right).

different wave velocities. This can be used to identify a possible tumor by sending acoustic waves into the human breast and measuring their response outside of it, for example with a medical ultrasound machine. Those measurements are compared with simulated measurements made on an estimated model of the breast without a tumor. If the breast differs from the estimated model, we obtain a misfit in the measurements. Then, we start a process of fitting between the simulated and the true measurements by changing the estimation of the model breast in such a way that the misfit between the measurements will be as small as possible. The goal of the minimization process of the misfit is not only to be able to reveal the existence of a possible tumor, but also its position, form and size.

1.1.1 PDE-Constrained Optimization

The propagation of waves through a medium can be described by the acoustic wave equation

$$y_{tt}(x,t) - \nabla \cdot (u(x)\nabla y(x,t)) = f(x,t)\,, \tag{1.1}$$

where $u(x) > 0$ represents the squared medium velocity, $f(x,t)$ the source function, and $y(x,t)$ the pressure variation, i.e. the wave field. Hence, for a given medium and an acoustic source, we get a solution $y(x,t)$ which represents the wave at location x and time t.

Now, we assume that $u(x)$ is unknown and would like to reconstruct the medium using acoustic waves. Therefore, we perform an illumination of the medium with source term f and denote by y the corresponding solution of (1.1). Given the measurements y^{obs} at sensor positions outside of the medium, we seek a reconstruction of the unknown squared wave speed u, such that the solution y of (1.1) with f coincides with the measurements y^{obs}. To solve the inverse medium problem, we formulate it as a PDE-constrained optimization

problem:

$$\text{For } \mathcal{F}(u,y) = \frac{1}{2}\left\|Py - y^{obs}\right\|_{L^2}^2 , \tag{1.2}$$

$$\begin{gathered}\text{minimize } \mathcal{F}(u,y)\,, \\ \text{such that } y(u) \text{ satisfies (1.1) for a source } f\,,\end{gathered} \tag{1.3}$$

where P is the projection at the sensor positions.

The amount of data in inverse medium problems tends to be large, especially when data is available for all time. The acoustic source is known and by using Fourier transform of the time variable, the source and the wave field can be represented by a sum of time harmonic waves. This allows us to separate time and space dimensions and write f and y as

$$y(x,t) = \hat{y}(x)e^{-i\omega t} \quad \text{and} \quad f(x,t) = \hat{f}(x)e^{-i\omega t}\,, \tag{1.4}$$

where $\omega > 0$ denotes the time frequency. Using the ansatz (1.4) in (1.1) leads to the Helmholtz equation

$$-\omega^2\hat{y}(x) - \nabla\cdot(u(x)\nabla\hat{y}(x)) = \hat{f}(x)\,, \tag{1.5}$$

which allows us to work in the frequency domain. By using (1.5) as the PDE constraint in (1.3) instead of the time dependent counterpart (1.1), we avoid large-scale time-dependent data. Still, the discovery of smaller features in the medium involves solving the Helmholtz equation at high frequencies, which is notoriously difficult, especially in three space dimensions.

1.2 Numerical Methods

Numerical methods for the solution of inverse scattering problems essentially fall into two classes: qualitative and quantitative methods. Qualitative methods can estimate the location and the shape of a scatterer, while quantitative methods are able to recover also some physical properties of it.

1.2.1 Qualitative Methods

Qualitative methods [11] generally require little a priori knowledge about the scatterer and allow to estimate its location and shape efficiently. Examples are the MUSIC (MUltiple SIgnal Classification) algorithm [15, 51], the DORT (Decomposition of the Time Reversal Operator) method [69, 12], the linear sampling method [16, 40], the probe method [68], and the factorization method [51, 52]. These methods provide a criterion for deciding whether a point lies inside or outside the scatterer only on the basis of far-field measurements of the scattered field. Hence, they effectively determine the support of a scatterer and thus permit to quickly detect material defects or obstacles, but do not provide any further quantitative information about its physical characteristics such as the local speed of sound.

To better understand the functionality of the qualitative methods, let us briefly describe the linear sampling method [16, 40]. For a given domain, which contains an (unknown) obstacle D, a grid of "sampling points" on the domain is chosen. The method uses the far field operator to define a far field equation with a solution g_z for a given grid point z. In [10], it is shown and proven that g_z satisfies different inequalities for points inside D and for points outside of it. Hence, the method uses the inequalities to select a cut-off constant C and uses it to decide if a grid point is inside D or not. If $\|g_z\| \leq C$, the grid point z is inside D, otherwise z is outside. Moreover, from the far-field data it is possible to distinguish between two different scatterers D_1 and D_2 in the domain.

1.2.2 Quantitative Methods

Quantitative methods, which we shall consider in this thesis, typically reformulate the inverse scattering problem as a PDE-constrained optimization problem, where the unknown physical parameters are determined by minimizing an appropriate objective functional [81], see for example (1.2)-(1.3). As the waves propagate through the medium and collect information about it, we formulate the problem as an optimization problem which minimizes the misfit between the solution of the simulated data and the true measurements. The minimization process starts from an initial guess of the medium and updates it during the optimization as close as possible to the exact parameter.

The optimization problem uses a metric to minimize the misfit of the data. In this thesis, we choose the L^2-norm (nonlinear least squares), but other metrics may be considered, for example, the Wasserstein metric [24], which is convex with respect to shifted or delayed data. This is an advantage of the Wasserstein metric with respect to the minimization over the L^2-norm, whereas the latter is strongly non-convex. Recently, this metric was applied successfully to the inverse time dependent problem in [25] and it may be an interesting subject for future work.

Considering PDE-constrained optimization, at least two strategies are then available: the *full-space* approach and the *reduced-space* approach. In the full-space approach the parameter u and the state variable y are sought simultaneously and their dependency is described by the constraints. This method is also denoted in the literature as "all-at-once" approach. In the reduced-space approach, which we shall consider in this thesis, the state variable y is eliminated from the objective functional as $\mathcal{F}(u) = \mathcal{F}(y(u), u)$, taking advantage of the linearity of the underlying Helmholtz equation and uniqueness of its solution under appropriate boundary conditions [35, 59, 32, 86].

To solve the optimization problem numerically, we have to discretize it at some point. Here again, two approaches are possible: *optimize-then-discretize*, where the continuous optimality conditions are formulated and then discretized in order to be solved. The second approach is *discretize-then-optimize*, where we discretize the problem first, which results in a finite dimensional optimization problem [46]. In the discretize-then-optimize approach it is easier to apply advanced algorithms from state-of-the-art optimization methods.

1.2.3 Optimization Strategies for the Inverse Helmholtz Problem

The inverse scattering problem as a PDE constrained optimization, using the Helmholtz equation, has been an area of research for over 30 years now [80]. The problem is known to be severely ill-posed. In the last three decades many applications [70, 37, 87], optimization methods [35, 71, 27] and algorithms [59, 45] have been applied to it. However, the problem remains challenging and optimization techniques do not always yield the expected result.

Solving the inverse problem using standard optimization techniques, for instance standard Newton or quasi-Newton methods from nonlinear optimization [65, 44], usually leads to unreasonable solutions and likely ends in a false local minimum. For inverse medium problems, however, the exact solution of the Newton equations at each iteration may be prohibitively expensive due to the very large number of (unknown) parameters. In recent years, inexact truncated Newton methods [22, 19], where at each (outer) iteration the (quasi-)Newton equations are solved by using only a few (inner) Krylov subspace iterations, have proved to be particularly effective for large-scale inverse medium problems [59]. Moreover, appropriate safeguards [22, 63] guarantee convergence.

At higher frequencies, waves detect and carry more detailed information about the scatterer [14]. Yet, solving the inverse problem directly with a high frequency, may dramatically increase the number of local minima. To prevent the optimization process from converging to a false local minimum, a frequency continuation strategy [14, 6, 71] is applied. This strategy is denoted in the literature also as frequency stepping. The principle of frequency stepping is to solve the inverse problem for a sequence of increasing frequencies ω, initializing each optimization run with the solution obtained from the previous lower frequency. In doing so, we reduce the risk of converging to a false local minimum by obtaining improving solutions for the optimization problem and we are still able to use highest frequency to reveal more details about the medium.

The number of linear systems to be solved in each optimization step grows linearly with the number of sources. For some applications the number of sources can reach several thousands. Thus, the cost of solving such a high number of linear systems may become prohibitive. To limit the computational cost without ignoring any of the available data, the sample average approximation (SAA) approach [38] uses "super-shots" as sources for the optimization. "Super-shots" are linear combinations of simultaneous sources determined by random vectors with zero mean and identity covariance. This yields a stochastic method, which needs to evaluate the mean of the misfit and uses the Monte Carlo approximation to do so. In this thesis, we combine our method with the SAA approach in several numerical experiments.

1.3 Regularization Techniques

By implementing all strategies, optimization methods and safeguards mentioned above, the solution of the inverse problem is improved, but often remains unsatisfactory as the

numerical solution usually contains perturbation and artifacts, especially on the source and receiver locations. To tackle this issue, the problem needs to be regularized. The term *regularization method* refers to any method which yields a stable approximate solution to an ill-posed problem [11]. In the following, we introduce some important regularization methods.

1.3.1 Regularization by Filtering and Discretization

In regularization by filtering [88, 41] the idea is to filter the information and to extract only useful information from an ill-posed problem. This regularization is based on the Singular Value Decomposition (SVD) of a matrix H, whose inversion is needed for the solution of an ill-posed inverse problem. Here, H typically denotes the Hessian matrix. As we invert H, the smallest singular values σ_i cause instabilities of order σ_i^{-1} in the solution. The regularization method then filters the right and left singular vectors related to those small values. More generally, the concept of reducing the dimension of the solution space is also referred to as regularization by discretization [50].

For very large and ill-posed systems one cannot always implement regularization by filtering, because it requires the computation of the Hessian and its singular value decomposition, which can be prohibitively expensive for large systems.

1.3.2 Tikhonov Regularization

Tikhonov regularization [82, 31, 23] can tackle ill-posedness with a small computational cost, even for very large systems. Several Tikhonov regularization functionals exist and they can be easily applied to the inverse problem by adding a penalty functional to (1.2). Among the Tikhonov-regularization functionals, we have the L^2 and the Sobolev H^1-penalty functionals [31, 88]. The *L^2-penalty* term penalizes directly the quantity of the reconstructed medium. The *Sobolev H^1-penalty* functional penalizes strong variations in the solution and is mainly suitable for smooth media. Unlike the H^1-penalty functional, *Total Variation* (TV) regularization [74], which was originally introduced for noise removal in digital image processing, can reconstruct media with discontinuities; it is able to remove unwanted noise, while preserving important details. TV can reconstruct nearly piecewise constant images with high quality [20] and has been implemented for inverse scattering problems successfully, see for example [55, 28].

Tikhonov regularization functionals are usually multiplied with a positive scalar α, which controls the trade-off between the misfit of the data and the regularization functional. If α is too high, the solution u will essentially minimize only the penalty function and if α is too small, the optimization process will essentially ignore the penalty function. In penalized TV-regularization [89] or in other regularization approaches [21], one adds another parameter ε to the penalty functional, which yields a differentiable functional that allows solutions with strong variability or discontinuities. Here, as for α, the parameter

ε must be well chosen to enable the penalty functional to regularize the inverse problem properly. If not, this may lead to unreasonable solutions or solutions whose discontinuities are smoothed out. Finding the right parameters, which are suitable for the reconstructions of all profiles, is not always easy [4].

At least two strategies are then available for the choice of the parameters: the *a posteriori* parameter selection and the *a priori* parameter selection. In the a posteriori case, the parameter depends on the misfit data, but not on any preceding information about the optimal solution. A standard a posteriori parameter choice for α is called Morozov's discrepancy principle [62]. A priori parameter selection includes prior information on the parameter u (or on the noise level), which cannot be recovered from the misfit data. For example, if we know in advance that the profile contains discontinuities, we can choose the parameter α in order to preserve them in the reconstruction, as in multiplicative regularization [84].

1.3.3 Size Reduction and Adaptive Regularizations

In [13], *regularization by size reduction* of admissible parameters is introduced. The set of admissible parameters can be set in two ways: the first tightens existing constraints while the second limits the space of admissible parameters to a subspace, spanned by global functions. For example, if we know a priori that our solution is smooth, we may choose a subspace V of smooth global functions and restrict the search space to V.

Another regularization, the *adaptive regularization*, is introduced in [13]. When a priori information is available but contradicts the information coming from measurements, the regularization shifts the parameter far from its desired value to accommodate both. Adaptive regularization progressively agrees with the measurements as much as possible, taking into account a priori information about u and the model. According to [13], adaptive regularization is the most desirable regularization. This area of research is very active and new papers introduce new adaptive regularization techniques. In [83], a two stage adaptive regularization is introduced for inverse Helmholtz problems, where frequency-stepping is available only at high frequencies. There, even if a priori information on discontinuities in the profile is given, the choice of regularization considers the frequency data limitations and promotes a very smooth solution at the first one or two frequencies. After obtaining a smooth solution, the penalty functional is adapted in order to resolve sharp features in the model.

In this thesis, we combine the space reduction regularization with the adaptive regularization. We call this combination the *Adaptive Eigenspace (AE)*: Starting the optimization at the lowest frequency with an initial guess far from the true solution, we first restrict the search to a smooth initial solution. In doing so, we restrict the set of admissible parameters to a subspace spanned by the eigenfunctions of the Laplacian that correspond to the smallest eigenvalues since they are smooth and not highly oscillatory. Using available a priori information about the parameter, we choose the most suitable Tikhonov penalty

functional $\mathcal{R}(u)$. Through the variation of $\mathcal{R}(u)$ we compute the related gradient $\nabla_u \mathcal{R}(u)$. Next, we compute the first K eigenfunctions of $\nabla_u \mathcal{R}(u)$ as a basis for the admissible parameters subspace V_K. The dimension K adapts itself at every frequency step, such that for increasing frequency ω the subspace V_K has a larger dimension, which depends on ω. This is called the AE basis and we will introduce it in the next section.

1.4 Adaptive Eigenspace (AE)

We consider the adaptive eigenspace inversion (AEI) as formerly presented in [18] for the time-dependent visco-elasticity equation and then for the time-dependent wave equation in [17]. In the AEI method the unknown wave speed is projected onto a particular finite-dimensional basis of eigenfunctions instead of a grid-based discrete representation. This basis is iteratively adapted during the optimization and reduces the number of control variables dramatically. Instead of a standard nodal basis, in [18, 17] a basis of eigenfunctions $\{\phi_m\}_{m=1}^K$ is being used to represent the parameter $u(x)$, where the eigenfunctions ϕ_m are given by

$$\begin{cases} -\nabla \cdot \left(\dfrac{1}{\max\{|\nabla u(x)|, \varepsilon\}} \nabla \phi_m(x) \right) &= \lambda_m \phi_m(x), & \forall\, x \in \Omega, \\ \phi_m(x) &= 0, & \forall\, x \in \Gamma. \end{cases} \tag{1.6}$$

The parameter $\varepsilon > 0$, which ensures that the denominator of μ does not vanish, is typically set to a very small value.

1.5 Multi-Parameter Inverse Problem

In recent years, research on multi-parameter inverse problems has become very popular, even though it was already introduced in 1984 [80]. The major interest in this area of study is in geophysics [66, 72, 90, 30], where the reconstruction of the two parameters density and bulk-modulus is crucial in oil and gas exploration.

The multi-parameter inverse Helmholtz has two further major challenges. The first one is the "cross-talk" between parameters: optimizing on several parameters appearing in the same equation can create an undesirable side-effect, so-called "cross-talk", i.e. the parameters create artifacts in each other [66]. The gradients of the parameters are coupled, meaning that each parameter appears and affects the gradient in the direction of the other parameter [67]. Even if a parameter reached the true solution and thus its gradient was zero, a perturbation in the other parameter may cause a non-zero gradient in the direction of the first parameter. The second problem is related to the ill-posedness, which is more extreme than in the one-parameter inverse problem. The two challenges mentioned above lead to a very difficult inverse problem. In this thesis, we illustrate the "cross-talk" and the coupled gradients and use the AEI to tackle the ill-posedness and reduce non-physical side-effects.

1.6 Thesis Outline

This thesis is organized in three parts. In Part I, we introduce optimization techniques and issues regarding the inverse scattering problem. In Chapter 2, we first introduce the Helmholtz equation and its discretization with Finite Element (FE) [49, 9] and Finite Difference (FD) [78] methods. Afterwards, we give the formulation for the full-space and reduced-space approaches for the inverse problem and present the Newton method and some quasi-Newton methods for their solution. We illustrate how solving the problem using standard optimization techniques usually leads to non-reasonable solutions and we thus present relevant optimization-techniques, such as inexact truncated Newton-like methods, safeguards and frequency-stepping. In Chapter 3, we illustrate the ill-posedness of the problem and propose some tools to reduce the instability of the problem. We implement Tikhonov penalty functionals and show some numerical examples illustrating their improvement over the solutions.

In Part II, we introduce the AE basis as a regularization technique. In Chapter 4, we describe our AEI approach, which combines state-of-the-art techniques from large-scale nonlinear optimization, such as inexact truncated Newton-like methods and frequency stepping [14, 6, 59], with an adaptive eigenspace representation of the parameter u for regularization. Next, in Chapter 5, we present both analytical and numerical evidence which underpins the remarkable accuracy of our particular choice of basis functions. In particular, we show how adapting the dimension of the eigenspace basis effectively builds regularization into the inversion. We subject our AEI method in Chapter 6 to a series of numerical tests that demonstrate not only its accuracy and robustness with respect to missing or noisy data, but also its versatility by combining it with a sample averaging approximation [38]. In Chapter 7, we solve the AEI with a single frequency and for limited frequency data. We show the ability of the AE to deal with missing data and to regularize the problem even starting with an high frequency. In Chapter 8, we introduce some other AE bases, which result from eigenspaces of gradients of typical penalty-terms. We show how to use the new bases in order to solve a seismic profile which is based on the Marmousi-profile [48, 77, 92, 39].

In Part III, we introduce the multiparameter inverse Helmholtz problem. We illustrate the main difficulties in solving it: the coupled gradients [67], the ill-posedness and the cross-talk between parameters [72, 90]. Next, we show some numerical results, applying the AE bases to each parameter separately and comparing the results to a nodal basis quasi-Newton method. The numerical results demonstrate the ability of the AEI to build different regularization for each parameter and reduce the cross-talk between the parameters.

1.7 Main Contributions

The main contributions of this thesis are as follows:

1. We extend the AE approach to the inverse Helmholtz problem and combine it with state-of-the-art techniques from large-scale nonlinear optimization, such as inexact truncated Newton-like methods. We deal with information from different frequencies by applying a frequency continuation strategy and adapting the dimension of the basis with respect to the frequency.

2. We show how truncating the adaptive eigenspace (AE) basis at a small and slowly increasing finite number of eigenfunctions effectively introduces regularization into the inversion and thus avoids the need for standard Tikhonov-type regularization.

3. We establish the relationship between standard Tikhonov penalty functionals and the elliptic partial differential equation satisfied by the AE basis functions. This new understanding of the AE basis enables us to consider other AE bases which are derived from different regularization functionals.

4. Our AEI method may be used for other ill-posed problems, where the frequency continuation strategy is unavailable. In this thesis, we show how to deal with a single frequency and how to choose the basis for even more complex problems, such as multi-parameter inverse scattering problems.

Part I

Optimization Techniques for Inverse Scattering Problems in the Frequency Domain

Chapter 2

Inverse Scattering in the Frequency Domain

We consider a time-harmonic problem in unbounded space with applications for example in seismic or medical imaging. We use the frequency counterpart of the wave equation through the Fourier transform, which yield the Helmholtz equation. We would like to discover what is inside a medium using acoustic waves. To do so, we formulate the inverse medium problem as PDE-constrained optimization problem.

2.1 Forward Problem

We formulate the inverse medium problem as PDE-constrained optimization problem, where the wave field solves the time independent Helmholtz equation. To solve the inverse problem we need to solve several times the forward problem. The model problem for time dependent acoustic waves is given by (1.1). Using the ansatz of (1.4) in (1.1) leads to the Helmholtz equation, as in (1.5). For the sake of simplicity, we now drop the hat notation from the Fourier transform:

$$-\omega^2 y(x) - \nabla \cdot \big(c^2(x)\nabla y(x)\big) = f(x)\,. \tag{2.1}$$

Here, $c > 0$ represents the medium velocity, f the source function, y the pressure variation, i.e. the wave field, and ω the time frequency. In the following, we will also denote by u the squared medium velocity, $u = c^2$.

For numerical simulations on unbounded domains, the infinite exterior must be truncated by an artificial boundary. Then, special boundary conditions are implemented to mimic the unbounded exterior. Two well-known approaches are Perfect Matched Layers [34, 8, 76] and absorbing boundary conditions [26, 7]. Here, we opt for the Sommerfeld

boundary condition, which is among the second kind.

$$\begin{cases} -\omega^2 y - \nabla \cdot (c^2 \nabla y) = f, & \text{in } \Omega, \\ \dfrac{\partial y}{\partial n} - iky = 0, & \text{on } \Gamma = \partial\Omega, \end{cases} \tag{2.2}$$

where $k(x) = \dfrac{\omega}{c(x)}$ is the wavenumber at time frequency ω.

Equation (2.2) can be discretized by different kinds of methods, such as Finite Volumes, Finite Differences, Finite Elements, Spectral Elements and so forth. The discretization of the continuous problem yields a problem with a finite number of unknowns, which can be solved on a computer. In this thesis we concentrate on Finite Element (FE) and Finite Difference (FD) methods. In either discretization method, for a given $c(x)$, we get a linear system for the forward problem $A(c^2)\, y = f$, which corresponds to the Helmholtz problem.

2.1.1 Finite Element Discretization

The Finite Element discretization uses an equivalent variational formulation of the problem. In the FE method, we divide the domain into smaller simpler finite parts which are called finite elements. In 2D, one usually uses squares or triangles, see Fig. 2.1. We then discretize the variational formulation of the problem on a finite dimensional subspace of piecewise polynomial functions. The resulting finite-dimensional problem can be represented in a large linear system. The FE method is very flexible and efficient for complicated geometries and for mesh adaptation [49].

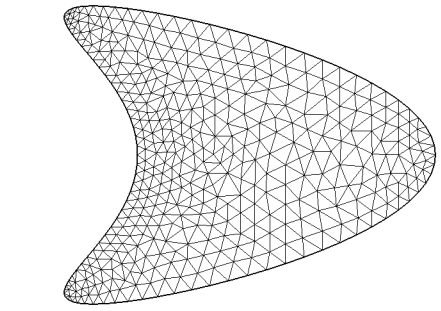

Figure 2.1: A Finite Element triangular mesh for a kite-shaped domain.

To obtain the semi-discrete problem of (2.2), we start from its weak formulation. Multiplying (2.2) with the complex conjugate $\bar{v}$ of a test function $v \in H^1(\Omega)$, integrating over Ω and using integration by parts, we find the following variational formulation:

Find $u \in H^1(\Omega)$ such that

$$a(y, v) - \omega^2 (y, v)_{L^2} - i\omega\, b(y, v) = (f, v)_{L^2} \tag{2.3}$$

for all $v \in H^1(\Omega)$, where $(\cdot,\cdot)_{L^2}$ denotes the $L^2(\Omega)$ hermitian product, the sesquilinear form $a(\cdot,\cdot)$ is given by

$$a(y,v) = \int_\Omega c^2(x)\nabla y(x)\nabla \bar{v}(x)dx\,.$$

and the sesquilinear form $b(\cdot,\cdot)$ is given by

$$b(y,v) = \int_\Gamma c(x)y(x)\bar{v}(x)dx\,.$$

The weak problem (2.3) has a unique solution [57, 29]. Let $V_h \subset H^1(\Omega)$ be a finite dimensional subspace of piecewise polynomial functions of degree p. Restricting (2.3) to V_h leads to the semi-discrete Galerkin formulation of (2.2): Find $y_h \in V_h$ such that

$$a(y_h,v) - \omega^2(y_h,v)_{L^2} - i\omega\, b(y_h,v) = (f,v)_{L^2} \tag{2.4}$$

for all $v \in V_h$. Let now $\{\varphi_i\}_{i=1}^N$ denote a nodal Lagrangian basis of V_h. For $y_h = \sum_{j=1}^N Y_j\varphi_j$ the semi-discrete formulation (2.4) is equivalent to

$$\sum_{j=1}^N Y_j a(\varphi_j,\varphi_i) - \omega^2 \sum_{j=1}^N Y_j(\varphi_j,\varphi_i)_{L^2} - i\omega \sum_{j=1}^N Y_j b(\varphi_j,\varphi_i) = (f,\varphi_i)_{L^2}$$

for all $i = 1,\dots,N$.
We finally get to the following system

$$\boldsymbol{K}\,\boldsymbol{Y} - \omega^2 \boldsymbol{M}\,\boldsymbol{Y} - i\omega \boldsymbol{B}\,\boldsymbol{Y} = \boldsymbol{R}, \tag{2.5}$$

where the vector $\boldsymbol{R} \in \mathbb{R}^{N\times 1}$ and the matrices $\boldsymbol{M} \in \mathbb{R}^{N\times N}$, $\boldsymbol{K} \in \mathbb{R}^{N\times N}$ and $\boldsymbol{B} \in \mathbb{R}^{N\times N}$ are given by

$$R_i = (f,\varphi_i)\,, \quad M_{i,j} = (\varphi_j,\varphi_i)\,, \quad K_{i,j} = a(\varphi_j,\varphi_i)\,, \quad B_{i,j} = b(\varphi_j,\varphi_i)\,, \tag{2.6}$$

for $i,j = 1,\dots,N$. The matrix $\boldsymbol{M}$ is sparse, symmetric and positive definite, the matrix $\boldsymbol{K}$ is sparse, symmetric and positive semi-definite and the matrix $\boldsymbol{B}$ is sparse, symmetric and has low rank.

2.1.2 Finite Differences Discretization

Another method which is used for the discretization of partial differential equations is the Finite Differences method (FD). In the FD method, derivatives in the partial differential equations are approximated using linear combinations of function values at grid points. The discretization with the FD method is less flexible for complicated geometries and mesh-adaptation in comparison with the FE method, but the discretization of the problem

(2.2) can be very efficiently computed, which reduces remarkably the computational time of the program. For the FD formulation, we use the strong formulation of the problem.

For simplicity, we separate the real and imaginary parts of the solution y and the source f and write them as

$$y = y^r + iy^i\,, \quad f = f^r + if^i\,.$$

We compute the finite differences for both parts of y and obtain two real valued systems instead of one complex-valued system. For example in the 2D case, the real part y^r of y satisfies

$$-\left(c^2_{i+\frac{1}{2},j}\frac{y^r_{i+1,j}-y^r_{i,j}}{h}\right)_{x_1} - \left(c^2_{i,j+\frac{1}{2}}\frac{y^r_{i,j+1}-y^r_{i,j}}{h}\right)_{x_2} - \omega^2 y^r_{i,j} = f^r_{i,j}\,, \tag{2.7}$$

where $\frac{y^r_{i+1,j}-y^r_{i,j}}{h}$ is the second order approximation of the derivative of y^r w.r.t. x_1 at $x_{i+\frac{1}{2},j}$ and $\frac{y^r_{i,j+1}-y^r_{i,j}}{h}$ is the second order approximation of the derivative of y^r w.r.t. x_2 at the grid point $x_{i,j+\frac{1}{2}}$, see Fig 2.2. We continue with the discretization for the second derivative and get:

$$\begin{aligned} &-\left(c^2_{i+\frac{1}{2},j}\frac{y^r_{i+1,j}-y^r_{i,j}}{h^2} - c^2_{i-\frac{1}{2},j}\frac{y^r_{i,j}-y^r_{i-1,j}}{h^2}\right) \\ &-\left(c^2_{i,j+\frac{1}{2}}\frac{y^r_{i,j+1}-y^r_{i,j}}{h^2} - c^2_{i,j-\frac{1}{2}}\frac{y^r_{i,j}-y^r_{i,j-1}}{h^2}\right) - \omega^2 y^r_{i,j} = f^r_{i,j} \end{aligned} \tag{2.8}$$

or

$$\begin{aligned} &-\frac{1}{h^2}\left[c^2_{i+\frac{1}{2},j}y^r_{i+1,j} + c^2_{i-\frac{1}{2},j}y^r_{i-1,j} + c^2_{i,j+\frac{1}{2}}y^r_{i,j+1} + c^2_{i,j-\frac{1}{2}}y^r_{i,j-1}\right] \\ &+\frac{1}{h^2}\left[c^2_{i+\frac{1}{2},j} + c^2_{i-\frac{1}{2},j} + c^2_{i,j+\frac{1}{2}} + c^2_{i,j-\frac{1}{2}}\right] y^r_{i,j} - \omega^2 y^r_{i,j} = f^r_{i,j}\,. \end{aligned} \tag{2.9}$$

For the imaginary part y^i the system is similar. Note that y and u are discretized on different grid points. This kind of grid is called staggered grid and is illustrated in Fig. 2.2. We notice that on $\Omega\backslash\Gamma$ the systems for y^i and y^r are not coupled.

For the Sommerfeld boundary conditions we use again second order FD. The boundary condition reads

$$\frac{\partial y}{\partial n} \;-\; iky = 0, \quad \text{on } \Gamma\,, \tag{2.10}$$

and is discretized on the upper boundary of Γ by

$$\frac{y_{N+1,j}-y_{N,j}}{h} \;-\; i\frac{\omega}{c_{N+\frac{1}{2},j}}y_{N,j} = 0, \quad \text{on } \Gamma\,. \tag{2.11}$$

We separate y in its imaginary and real parts on the boundary as well and get

$$\begin{cases} y^r_{N+1,j} &= y^r_{N,j} - \dfrac{\omega\, h}{c_{N+\frac{1}{2},j}}y^i_{N,j}, & \text{on } \Gamma\,, \\ y^i_{N+1,j} &= y^i_{N,j} + \dfrac{\omega\, h}{c_{N+\frac{1}{2},j}}y^r_{N,j}, & \text{on } \Gamma\,. \end{cases} \tag{2.12}$$

All computations on the other boundaries are analogue.

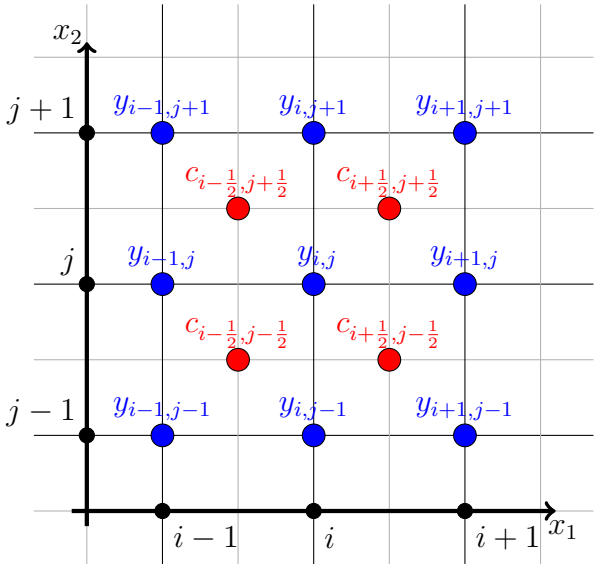

Figure 2.2: Staggered grid used in the FD method.

We assume that the velocity $c(x)$ on the boundary is given and hence, the parameter $c_{N+\frac{1}{2},j}$ is known. On Γ the system (2.12) is coupled and the real part of y is dependent on its imaginary part and vice versa. Hence, the discretized forward problem cannot be solved for y_i and y_r independently. In order to simulate the Sommerfeld boundary conditions, we apply (2.12) in (2.9) for all boundary points of the FD mesh discretization.

2.2 Inverse Problem

Now, we would like to reconstruct the medium using acoustic waves. Therefore, we perform N_s illuminations of the medium inside Ω with source term $f = f_\ell$, $\ell = 1, \ldots, N_s$, and denote by y_ℓ the corresponding solutions of (2.2). Given the measurements y_ℓ^{obs} on Γ, or part of it, we seek to reconstruct the unknown squared wave speed $u = c^2$ inside Ω such that every solution y_ℓ of (2.2) with $f = f_\ell$ coincides at Γ with the measurements y_ℓ^{obs}, $\ell = 1, \ldots, N_s$. In doing so, we assume that the wave speed c is known on the boundary Γ. To solve the inverse medium problem, we formulate it as a PDE-constrained optimization problem:

$$\text{For } \mathcal{F}(u,y) = \frac{1}{2}\sum_{\ell=1}^{N_s} \left\| Py_\ell - y_\ell^{obs} \right\|_{L^2}^2 , \tag{2.13}$$

$$\begin{gathered} \text{minimize } \mathcal{F}(u,y)\,, u \in S\,, y_\ell \in V \\ \text{such that } y_\ell(u) \text{ satisfies (2.2) for } f = f_\ell, \quad \ell = 1, ..., N_s. \end{gathered} \tag{2.14}$$

The minimization problem above is strongly non-convex and severely ill-posed. To tackle this issue, a regularization term $\mathcal{R}(u)$ is typically added and multiplied by a scalar

$\alpha \in \mathbb{R}$. This method of adding an extra term to an ill-posed problem is called Tikhonov regularization [88] and will be further discussed in Chapter 3. We then solve the inverse problem (2.14) for

$$\mathcal{F}(u, y) = \frac{1}{2} \sum_{\ell=1}^{N_s} \left\| P y_\ell - y_\ell^{obs} \right\|_{L^2}^2 + \alpha \mathcal{R}(u) \,. \tag{2.15}$$

Given measurements y_ℓ^{obs}, $\ell = 1, \ldots, N_s$, different methods from PDE-constrained optimization can be used to retrieve u by minimizing the misfit, stated either in the full-space or in the reduced-space approach [32, 35].

2.2.1 Full-Space Approach

In the full-space approach all parameters are optimized in parallel. We consider the full-space optimization problem

$$\begin{aligned} &\text{minimize } \mathcal{F}(u, y) \,, u \in S \,, y_\ell \in V \\ &\text{such that } y_\ell \text{ satisfies (2.2) for } f = f_\ell, \quad \ell = 1, ..., N_s \,, \end{aligned} \tag{2.16}$$

with the Lagrangian $\mathcal{L}$ of the optimization problem (2.16)

$$\mathcal{L}(u, y, \lambda) = \frac{1}{2} \sum_{\ell=1}^{N_s} \left\| P y_\ell - y_\ell^{obs} \right\|_{L^2}^2 + \sum_{\ell=1}^{N_s} \lambda_\ell^T \left(A(u) y_\ell - f_\ell \right) + \alpha \mathcal{R}(u) \,, \tag{2.17}$$

where we recall that $A(u)$ is the operator of the forward Helmholtz problem (2.2).

To optimize (2.16), the first derivatives of (2.17) must vanish. From the Karush-Kuhn-Tucker conditions (KKT), we seek y_ℓ, u and λ_ℓ, $\ell = 1, , \ldots, N_s$ such that

$$\begin{cases} \mathcal{L}_{y_\ell} = P^T (P y_\ell - y_\ell^{obs}) + A^T(u) \lambda_\ell = 0 \,, \\ \mathcal{L}_u = \alpha \mathcal{R}_u + \sum_{\ell=1}^{N_s} G_\ell^T \lambda_\ell = 0 \,, \\ \mathcal{L}_{\lambda_\ell} = A(u) y_\ell - f_\ell = 0 \,, \end{cases} \tag{2.18}$$

with

$$G_\ell = \frac{\partial (A(u) y_\ell)}{\partial u} \quad \text{and} \quad \mathcal{R}_u = \frac{\partial \mathcal{R}(u)}{\partial u} \,. \tag{2.19}$$

Newton's method yields the search direction of the optimization using the Hessian of the Lagrangian $\mathcal{L}$ (2.17). For this purpose we introduce also the terms:

$$\mathcal{R}_{uu} = \frac{\partial^2 \mathcal{R}(u)}{\partial u^2} \,, \quad W_\ell = \frac{\partial (A^T(u) \lambda_\ell)}{\partial u} \,, \quad Q = \sum_{\ell=1}^{N_s} \frac{\partial (G_\ell^T \lambda_\ell)}{\partial u} \,, \quad \ell = 1, , \ldots, N_s.$$

We now find the search directions $\delta y = (\delta y_1, \ldots, \delta y_{N_s})$, δu, $\delta\lambda = (\delta\lambda_1, \ldots, \delta\lambda_{N_s})$ of the optimization by solving the linear system:

$$\begin{pmatrix} \tilde{P} & \tilde{W} & \tilde{A}^T \\ \tilde{W}^T & \alpha\mathcal{R}_{uu} + Q & \tilde{G}^T \\ \tilde{A} & \tilde{G} & 0 \end{pmatrix} \begin{pmatrix} \delta y \\ \delta u \\ \delta\lambda \end{pmatrix} = - \begin{pmatrix} \mathcal{L}_y \\ \mathcal{L}_u \\ \mathcal{L}_\lambda \end{pmatrix}, \tag{2.20}$$

where $\tilde{P}$, $\tilde{A}$ are the block-diagonal matrices with N_s blocks of the matrix P^TP and the matrix A respectively

$$\tilde{P} = \begin{pmatrix} P^TP & & & \\ & P^TP & & \\ & & \ddots & \\ & & & P^TP \end{pmatrix}, \qquad \tilde{A} = \begin{pmatrix} A(u) & & & \\ & A(u) & & \\ & & \ddots & \\ & & & A(u) \end{pmatrix},$$

the matrices $\tilde{W}$ and $\tilde{G}$ are given by

$$\tilde{W} = \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_{N_s} \end{bmatrix}, \qquad \tilde{G} = \begin{bmatrix} G_1 \\ G_2 \\ \vdots \\ G_{N_s} \end{bmatrix}$$

and the vectors $\mathcal{L}_y = \left[\mathcal{L}_{y_1}, \mathcal{L}_{y_2}, \ldots, \mathcal{L}_{y_{N_s}}\right]^T$, $\mathcal{L}_\lambda = \left[\mathcal{L}_{\lambda_1}, \mathcal{L}_{\lambda_2}, \ldots, \mathcal{L}_{\lambda_{N_s}}\right]^T$.

The full-space approach optimizes a vast number of control variables, which can reach several millions. Hence the search space is very large and solving (2.20) usually requires efficient direct solvers. An attractive compromise is the reduced space approach, which reduces the search space of the optimization to the medium variable u only by eliminating y_ℓ, λ_ℓ [45].

2.2.2 Reduced-Space Approach

Following [35, 47], we reduce the search space by using the reduced-space approach. This approach takes advantage of the dependence of the adjoint and state variables on the control variable u. The state variable y_ℓ can be expressed as a function of u in the following way

$$y_\ell = A(u)^{-1} f_\ell \,. \tag{2.21}$$

Here again, $A(u)$ is the operator of the forward Helmholtz problem (2.2). By the elimination of y_ℓ the optimization problem for the reduced space approach reads

$$\text{minimize } \mathcal{F}^r(u),\ u \in S\,, \text{ where} \tag{2.22}$$

$$\mathcal{F}^r(u) = \frac{1}{2}\sum_{\ell=1}^{N_s} \left\| PA(u)^{-1} f_\ell - y_\ell^{obs} \right\|_{L^2}^2 + \alpha \mathcal{R}(u) \tag{2.23}$$

without any additional constraint. To solve (2.22), we now need to compute the gradient of $\mathcal{F}^r$ and its Hessian. The corresponding gradient reads

$$\nabla_u \mathcal{F}^r(u) = \sum_{\ell=1}^{N_s} \left(\frac{\partial y_\ell(u)}{\partial u} \right)^T P^T \left(PA(u)^{-1} f_\ell - y_\ell^{obs} \right) + \alpha \mathcal{R}_u. \tag{2.24}$$

Using the chain rule we get

$$0 = \frac{\partial f_\ell}{\partial u} = \frac{\partial (A(u) y_\ell(u))}{\partial u} = \frac{\partial (A(u) y_\ell)}{\partial u} + \frac{\partial (A(u) y_\ell)}{\partial y_\ell} \frac{\partial y_\ell}{\partial u} = G_\ell + A(u) \frac{\partial y_\ell}{\partial u}\,, \tag{2.25}$$

where G_ℓ is as defined in (2.19). This yields

$$\left(\frac{\partial y_\ell(u)}{\partial u} \right)^T = -G_\ell^T A(u)^{-T}, \quad \ell = 1, \ldots, N_s\,. \tag{2.26}$$

We denote then by λ_ℓ the following term

$$\lambda_\ell = A(u)^{-T} P^T \left(PA(u)^{-1} f_\ell - y_\ell^{obs} \right), \quad \ell = 1, \ldots, N_s\,, \tag{2.27}$$

which appears in the computation of the gradient (2.24). The equations (2.21) and (2.27) are compatible with the full-space approach and satisfy the third and first equations of (2.18) for state and adjoint variables respectively.

From the Hessian in (2.20), we can derive the reduced-space Hessian and find an equation for the search direction δu. By considering the third row of (2.20), we get

$$\delta y_\ell = -A(u)^{-1} (\mathcal{L}_{\lambda_\ell} + G_\ell \delta u)\,, \quad \ell = 1, \ldots, N_s \tag{2.28}$$

and the first row of (2.20), using (2.28), yields for $\ell = 1, \ldots, N_s$

$$\delta \lambda_\ell = (A(u)^{-T} P^T P A(u)^{-1} G_\ell - A(u)^{-T} W_\ell) \delta u + A(u)^{-T} P^T P A(u)^{-1} \mathcal{L}_{\lambda_\ell} - A(u)^{-T} \mathcal{L}_{y_\ell}\,. \tag{2.29}$$

Finally by using the second row of (2.20) and (2.29), we get a linear system for the reduced search direction δu

$$\begin{aligned} \left(\alpha \mathcal{R}_{uu} + Q + \sum_{\ell=1}^{N_s} J_\ell^T J_\ell - W_\ell^T A(u)^{-1} G_\ell - G_\ell^T A(u)^{-T} W_\ell \right) \delta u = \\ -\alpha \mathcal{R}_u - \sum_{\ell=1}^{N_s} J_\ell^T (PA(u)^{-1} f_\ell - y_\ell^{obs})\,, \end{aligned} \tag{2.30}$$

where the Jacobian J_ℓ is given by $J_\ell = -PA(u)^{-1} G_\ell$ and $y_\ell = A(u)^{-1} f_\ell$ for $\ell = 1, \ldots, N_s$.

In this thesis we shall consider only the reduced space approach and hence for simplicity, we denote **from now on $\mathcal{F} = \mathcal{F}^r$ as the reduced space objective function** (2.23).

The solution of the Hessian involves solving a large number of linear systems. To reduce the computational cost but still get useful results, we consider quasi-Newton methods. They are based on the Newton method, but do not require the Hessian, only an approximation of it. The approximations of the Hessian in the quasi-Newton methods are cheaper than the full Hessian needed for Newton's method. In the following, some quasi-Newton methods are presented.

The Gauss-Newton Method

The Gauss-Newton (G-N) method is a quasi-Newton method, which keeps the first derivatives appearing in the Hessian, but ignores the second derivative terms. In many situations the G-N approximation is very close to the full Hessian and thus holds a similar convergence rate as the Newton method [65].

Here we describe the G-N method for the optimization problem (2.22). Actually, if we aim for the G-N approximation in the reduced-space approach, the second-order part of (2.30) can be neglected: instead of solving (2.30), in the G-N method we write

$$\sum_{\ell=1}^{N_s} J_\ell^T J_\ell \, \delta u = -\sum_{\ell=1}^{N_s} J_\ell^T \, r_\ell \,, \tag{2.31}$$

with the Jacobian J_ℓ defined by $J_\ell = -PA(u)^{-1}G_\ell$ and the residual $r_\ell = (PA(u)^{-1}f_\ell - y_\ell^{obs})$, $\ell = 1, \ldots, N_s$. The Hessian can be approximated by $J_\ell^T J_\ell$ in (2.30), together with the regularization term. Hence we solve the linear system

$$\left(\alpha\mathcal{R}_{uu} + \sum_{\ell=1}^{N_s} G_\ell^T A(u)^{-T} P^T P A(u)^{-1} G_\ell\right) \delta u = -\alpha\mathcal{R}_u - \sum_{\ell=1}^{N_s} J_\ell^T (PA(u)^{-1}f_\ell - y_\ell^{obs}) \,. \tag{2.32}$$

Note that we do not approximate the Hessian of the regularization, which appears in (2.30), because it is cheap to compute.

Equation (2.32) involves the solution of fewer linear systems than in (2.30), and hence is interesting in terms of computational cost. In [65], the advantages of the Gauss-Newton method are discussed:

1. This approximation saves the computational effort by neglecting the second-order part of (2.30) and thus, per iteration step, less linear systems need to be solved. In the inverse Helmholtz problem, the use of the Gauss-Newton approximation can save hundreds to several millions of linear system computations.

2. In many situations, especially close to the solution, the approximation $J_\ell^T J_\ell$ is a very good approximation of the Hessian and so the Gauss-Newton method holds a similar convergence rate as the full-Newton method.

3. For a non-zero $\nabla_u \mathcal{F}$ and whenever the matrices J_ℓ have full rank, the search direction δu provided by the G-N method, given by (2.31), is a descent direction and fulfills:

$$\begin{aligned}\delta u^T \nabla_u \mathcal{F} &= \delta u^T \left(\sum_{\ell=1}^{N_s} J_\ell^T r_\ell \right) = \delta u^T \left(-\sum_{\ell=1}^{N_s} J_\ell^T J_\ell \, \delta u \right) \\ &= -\sum_{\ell=1}^{N_s} \delta u^T J_\ell^T J_\ell \, \delta u = -\sum_{\ell=1}^{N_s} \|J_\ell \delta u\|^2 \leq 0 \,.\end{aligned}$$

Actually, the inequlity is strict, otherwise $J_\ell \delta u = 0$ for $\ell = 1, \ldots, N_s$ and since we assumed that the matrices J_ℓ have a full rank, we get $J_\ell^T J_\ell \, \delta u = 0$ for $\ell = 1, \ldots, N_s$ and hence $\nabla_u \mathcal{F} = 0$.

4. Global convergence can be proved under some assumptions:

Assuming that there is a constant $a > 0$ such that

$$\|J_\ell(u) z\| \geq a \|z\| \tag{2.33}$$

and for all u in a neighborhood $\mathcal{N}$ of the level set

$$U = \{u \,|\, \mathcal{F}(u) \leq \mathcal{F}(u_0)\} \,, \tag{2.34}$$

where u_0 is the starting value for the algorithm and U is bounded, the following Theorem holds.

Theorem 2.2.1. *Let k denote the iteration number. Suppose each residual function $r_{\ell,k}$ is Lipschitz continuously differentiable in a neighborhood $\mathcal{N}$ of a bounded set* (2.34)*, and that the Jacbobians $J_{\ell,k}(u)$ satisfy the uniform full-rank condition* (2.33) *on $\mathcal{N}$. Then if the u_k are generated by the Gauss-Newton method with the Wolfe line search, we have*

$$\lim_{k \to \infty} J_{\ell,k}^T r_{\ell,k} = 0 \,.$$

Theorem 2.2.1 and its proof can be found in [65].

The BFGS Method

Another very popular quasi-Newton method is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method. The BFGS method is a quasi-Newton method as it avoids the computation of the Hessian. It computes iteratively low-rank approximations of the Hessian at a low computational cost.

The idea of this method is based on the fact that

$$\nabla f(x_k + s_k) = \nabla f(x_k) + \nabla^2 f(x_k) s_k + \mathcal{O}(\|s_k\|^2) \,, \quad \text{for } \|s_k\| \to 0.$$

If we choose

$$s_k = x_{k+1} - x_k \quad \text{and} \quad g_k = \nabla f(x_{k+1}) - \nabla f(x_k) \tag{2.35}$$

it holds

$$\nabla^2 f(x_k)\, s_k = g_k + \mathcal{O}(\|s_k\|^2)\,.$$

Let B_k be the approximation of the Hessian $\nabla^2 f(x_k)$. We search for B_{k+1} the approximation of $\nabla^2 f(x_{k+1})$, such that the following equation holds

$$B_{k+1}\, s_k = g_k\,.$$

To reduce the cost in computing B_{k+1}, we search for

$$B_{k+1} = B_k + U_k\,,$$

where U_k is a low rank matrix. Effectively we update B_{k+1} iteratively through the following formula

$$B_{k+1} = B_k - \frac{B_k\, s_k s_k^T\, B_k}{s_k^T\, B_k\, s_k} + \frac{g_k g_k^T}{g_k^T s_k}\,,$$

with an initial approximation of the Hessian B_0. For the full derivation of the BFGS method and the choice of B_{k+1} we refer to [65].

The BFGS method has two key advantages. The first one is its convergence: if we use the Wolfe line search conditions and under some assumptions on f, we get a superlinear convergence. The second advantage is the computational cost: each iteration can be performed at a cost of $\mathcal{O}(N_u^2)$ operations [65].

In the inverse Helmholtz problem, we have to take into account that the cost of the evaluations of the gradient $\nabla_u \mathcal{F}(u_k)$ is added to the total cost of the BFGS method. Each evaluation of a gradient involves the solution of N_s linear systems. Those computations appear not only in the BFGS algorithm, but also in the Wolfe line search conditions, which are essential for the success of the method. Another disadvantage of the method is its large memory requirement. If only limited memory is available, we instead turn to the limited-memory BFGS presented in the next subsection.

Limited-Memory BFGS

A variation of the BFGS method, which is very common for large scale problems, is the Limited-Memory BFGS (L-BFGS). To save storage, the method uses only partial information of the BFGS method, i.e. the newest which is likely more relevant, instead of the complete one. At iteration k, the method approximates the inverse Hessian using the $j \ll N_u$ previous steps $\{s_k,\, s_{k-1}, \ldots,\, s_{k-j}\}$ and gradient differences $\{g_k,\, g_{k-1}, \ldots,\, g_{k-j}\}$ given in (2.35). For inverse problems using L-BFGS method, it is common to choose $B_0 = \mathcal{R}_{uu}$. Here, as in the BFGS method, the Wolfe conditions are used for the line search [64].

In this thesis there will be no need to use this method because we will use a special basis which requires little memory as described in Chapter 4. Note that for a nodal basis discretization, such as FE, FD and others, this method is very popular, see for example [3, 59, 60].

2.3 Inexact Newton Methods for Inverse Helmholtz Problems

We now introduce our model-problem and consider the profile $u(x_1, x_2)$ shown in Fig. 2.3. The profile u mimics a layered material with regions of different wave speed.

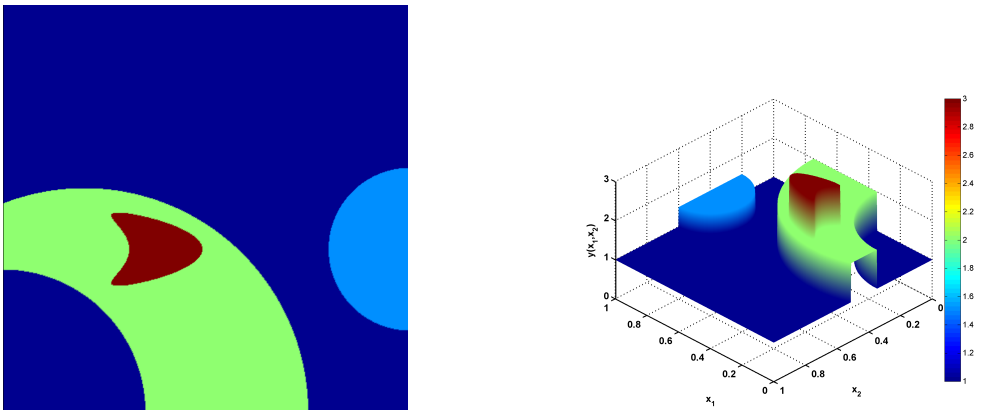

Figure 2.3: Exact profile u: two-dimensional view (left) and three-dimensional view (right).

We would like to reconstruct u using, naively, the standard optimization tools as described in Section 2.2. For this reconstruction, we use the following settings: nine equispaced Gaussian sources are located along the upper boundary at $(0.1, 0.8), \ldots, (0.9, 0.8)$, whereas the receivers are located on the four lateral boundaries of $\Omega = (0, 1) \times (0, 1)$. We use second-order finite differences on a 200×200 mesh for the discretization of (2.2) and optimize at a single frequency $\omega = 8$. In each optimization step, we compute δu using the reduced space Gauss-Newton (G-N) approximation for the Hessian, which is given in (2.31). In Fig. 2.4 we present the numerical result for the G-N reconstruction. Although the minimization problem reaches a minimum, $\mathcal{F}(u) = 1.45 \cdot 10^{-9}$, the reconstruction is poor and has a relative L^2-error of more than two million percent. To get a more appropriate solution, the optimization needs to be stabilized. For this purpose we apply in the following several relevant strategies.

In the inverse Helmholtz problem, computing the Hessian, or approximating it, can be very inefficient or even impossible in terms of computational costs. In addition, as shown in Fig. 2.4, the use of the exact Hessian or its G-N approximation can lead to an unstable optimization. To overcome these difficulties, we turn to inexact Newton or quasi-Newton

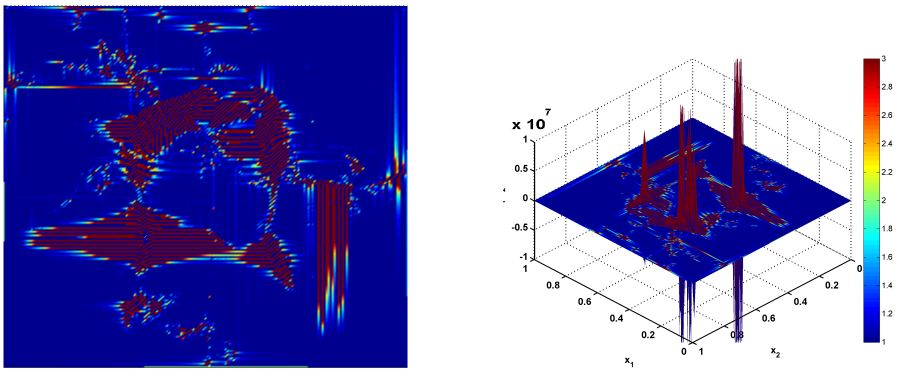

Figure 2.4: Ill-posedness of the inverse problem. Reconstruction of the profile described in Fig. 2.3 without regularization with a relative L^2-error of more than two million percent. Two dimensional view (left) and three dimensional view (right).

methods. The robustness of inexact Newton or quasi-Newton methods depends on their *forcing term* whose choice is critical for the optimization. In this section, we shall explain what a forcing term is, what is its importance and how to choose it. In addition we will show why the Hessian cannot be stored and how to overcome this problem.

We focus on the reduced-space approach and we wish to optimize (2.23) efficiently using Newton or quasi-Newton methods. Therefore, we have to solve at each iteration

$$H\delta u = -\nabla_u \mathcal{F}(u)\,, \tag{2.36}$$

where H represents the Hessian or the quasi-Newton approximation of it, δu the search direction and $\nabla_u \mathcal{F}(u)$ the reduced space gradient from (2.24). In the inverse Helmholtz problem, computing the exact Hessian or its G-N approximation may be very costly and involves solutions of a large number of linear systems, which is in many cases impossible. For example, if we compute the G-N Hessian's approximation, given N_s sources we need to compute the term

$$A^{-1}G_\ell \quad \text{for } \ell = 1, \ldots, N_s\,, \tag{2.37}$$

which appears in both (2.30) and (2.32), with $A \in \mathbb{R}^{N_y \times N_y}$ the discretization of (2.2), $G_\ell \in \mathbb{R}^{N_y \times N_u}$ from (2.19), N_y and N_u are the degrees of freedom of the state variable y and the parameter u respectively. To solve (2.37) we need to solve N_u linear systems for each source $\ell = 1, \ldots, N_s$. This involves solutions of a large number of linear systems of size $Ny \times Ny$ per optimization step

$$\mathcal{O}(N_u \cdot N_s)\,. \tag{2.38}$$

The term in (2.37) is just one of the terms we need to compute in (2.30) or in (2.32). Actually, even if N_u and N_s are not large, we need to compute several millions of linear systems per optimization step. For example, in the previous experiment we needed to compute $1,447,200$ linear systems of size 400×400 per optimization step.

Hence, we turn to inexact Newton methods. We use the ability of Krylov-methods to solve a linear system without needing to store the matrix explicitly; for a given vector p, we only need to compute the product Hp. We consider now the G-N approximation of the Hessian H (2.32). Its multiplication with a vector p, yields

$$\left(G_\ell^T A^{-T} P^T P A^{-1} G_\ell + \alpha \mathcal{R}_{uu}\right) p. \tag{2.39}$$

Since the multiplication $G_\ell\, p$ is a matrix-vector multiplication, it yields a vector. Hence, the cost of solving $A^{-1}\left(G_\ell\, p\right)$ is actually solving a single linear system per source. The second linear system which we solve is $A^{-T}\left(P^T P A^{-1} G_\ell\, p\right)$. Here, $P^T P A^{-1} G_\ell\, p$ is a vector as well. Hence, using the G-N approximation, this involves solving of two linear systems per source. A similar computation can be done for the full Hessian (2.20), where three linear systems per source need to be solved. Algorithms which avoid the explicit computation of a matrix are also denoted in the literature as matrix-free algorithms.

In the matrix-free inverse Helmholtz problem, we use the *truncated* Newton or *truncated* quasi-Newton method, which is an inexact Newton method. The term "truncated" comes from the fact that we solve the linear system (2.36) inexactly using conjugate-gradient (CG) method and stop the inner iterations of the CG method before the exact solution for δu is found.

Performing a large number of CG iterations may be expensive and result in a solution which is close to the exact solution of the (quasi-)Newton method, but does not correspond to the expected solution, see Fig. 2.4. On the other hand, performing a very small number of iterations may yield inefficient search direction. Hence, we shall decide how many CG iterations we should perform in order to get an efficient and inexpensive search direction δu. To do so, we need to find a criterion which stops the CG iterations when a sufficient δu is found. Such a criterion is called a stopping criterion. Note that we do not try to reach the convergence of the CG method in order to compute δu exactly, but convergence according to this criterion, as will be explained in the following.

Next, we choose the stopping criterion which is the most suitable for our inverse Helmholtz problem and leads to a robust and efficient solution. We turn to the Eisenstat-Walker stopping criteria [22] and its safeguards [22, 63], to find a suitable stopping criterion for the inner CG iterations. The robustness of the inexact Newton or inexact quasi-Newton methods depends on the chosen stopping criterion, and is essential for the success of the optimization process.

A truncated Newton method is, according to [63], effective if

1. a small number of inner iterations is sufficient to produce a "good" step,
2. each inner iteration can be performed efficiently,
3. safeguards are implemented to guarantee convergence.

Now, we go through all previous conditions and show that our algorithm satisfies all three of them:

1. To optimize (2.23), we opt for a quasi-Newton method and use a standard truncated G-N method. We solve (2.36) inexactly using the CG method and take advantage of the ability of the CG method to solve the linear system without storing the matrix H. This allows efficient inner iterations. Combining the CG method together with the Walker-Eisenstat stopping criteria for inexact Newton methods [22] ensures a sufficient step at a low cost. In [22], the linear inner iterations are stopped when

$$\|H(u)p + \nabla_u \mathcal{F}(u)\| \leq \eta_k \|\nabla_u \mathcal{F}(u)\|\,, \tag{2.40}$$

where η_k, with k the iteration number, is called the forcing term and refers to one of the following choices

$$\begin{cases} \eta_k^{(1)} = \dfrac{\|\nabla_u \mathcal{F}(u_k) - \nabla_u \mathcal{F}(u_{k-1}) - \alpha_{k-1} H(u_{k-1})p\|}{\|\nabla_u \mathcal{F}(u_{k-1})\|}\,, \\ \eta_k^{(2)} = \dfrac{\|\nabla_u \mathcal{F}(u_k)\| - \|\nabla_u \mathcal{F}(u_{k-1}) + \alpha_{k-1} H(u_{k-1})p\|}{\|\nabla_u \mathcal{F}(u_{k-1})\|}\,, \\ \eta_k^{(3)} = \gamma_1 \left(\dfrac{\|\nabla_u \mathcal{F}(u_k)\|}{\|\nabla_u \mathcal{F}(u_{k-1})\|} \right)^{\gamma_2}\,, \quad \gamma_1 \in [0,1) \text{ and } \gamma_2 \in (1,2]\,. \end{cases} \tag{2.41}$$

For $k = 0$ and for each forcing term, we pick $\eta_0^{(i)} \in [0,1]$, for $i = 1, 2, 3$. In this thesis, we choose experimentally $\eta_0^{(i)} = 0.7$ for all three choices in (2.41). The first two forcing terms are related to the approximation of the gradient

$$\nabla_u \mathcal{F}(u_k) \simeq \nabla_u \mathcal{F}(u_{k-1}) + \alpha_{k-1} \nabla_{uu} \mathcal{F}(u_{k-1})p\,,$$

where $\alpha_{k-1} \in [0,1]$ is the line-search step in the previous iteration. Therefore $\eta_k^{(1)}$ and $\eta_k^{(2)}$ are getting smaller when the approximation of the gradient is getting better. In (2.40), the smaller η_k, the more inner iterations are needed, in particular, when the optimization is close to the expected minimum. We observe the same phenomenon for $\eta_k^{(3)}$: the smaller the gradient, the smaller $\eta_k^{(3)}$ and the more CG iterations must be performed [59]. However, this third forcing term requires the adjustment of parameters γ_1 and γ_2. A poor choice of the last two parameters can lead to an unsatisfactory stopping criterion and affect the robustness of the method, thus, we prefer to work with $\eta_k^{(1)}$ or $\eta_k^{(2)}$.

In [59], the forcing terms are tested for the FWI problem. There, they conclude that the forcing term $\eta_k^{(1)}$ is the most suitable one to the FWI problem, especially for the truncated Newton method. Numerical experiments with each forcing term in (2.41) suggested that all three are suitable for the truncated G-N. We choose to use the first term $\eta_k^{(1)}$ or second one $\eta_k^{(2)}$, which prove to give good results and do not have the inconvenience of the two extra parameters like in $\eta_k^{(3)}$.

Theorem 2.3.1. *Let $\mathcal{F}(u) : \mathbb{R}^{N_u} \to \mathbb{R}$ be the objective functional with a gradient $\nabla_u \mathcal{F}(u)$, which is continuously differentiable in a neighborhood of $u_\star$ and for which $\nabla_u \mathcal{F}(u_\star) = 0$. Let the Hessian $H(u_\star)$ be nonsingular, Lipschitz continuous at $u_\star$. If u_0 is sufficiently near $u_\star$, then $\{u_k\}$ with $\{\eta_k\}$ given by $\eta_k^{(1)}$ or $\eta_k^{(2)}$ in (2.41) converges to $u_\star$ at least superlinear.*

Theorem 2.3.1 and its proof can be found in [65]. Note that for $\eta_k^{(3)}$ given in (2.41), the convergence rate depends on the choice of γ_1 and γ_2.

2. Using the Krylov-methods, for example CG method, the linear system (2.36) can be solved without storing the matrix explicitly. We showed for the G-N approximation of the Hessian, using (2.39), that it involves solving two linear systems per source times number of CG iterations. Similarly, using the full Hessian (2.30), we need to solve three linear systems per source times number of CG iterations. In practice, the extra linear system added in the case of the full-Hessian (2.20) might be expensive in terms of computational costs (it may reach several thousands of linear systems). Hence, we use the G-N approximation of the Hessian to solve less linear systems and to speed up our optimization. In [59], it is concluded that the Gauss-Newton method with the Eisenstat-Walker stopping criteria gives the best results in terms of cost-performance ratio.

3a. To prevent the forcing terms from being too small or being larger than one, we implement safeguards to modify $\eta_k^{(i)}$ for $i = 1, 2, 3$ described in (2.41) as recommended in [22, 63].

For the first two choices, $\eta_k^{(1)}$ and $\eta_k^{(2)}$ in (2.41), the following modifications will be applied

$$\begin{cases} \eta_k^{(i)} = \eta_0^{(i)} & \text{if } \eta_k^{(i)} > 1 \qquad \text{for } i = 1, 2\,, \\ \eta_k^{(i)} = \max\left(\eta_k^{(i)}, \left(\eta_{k-1}^{(i)}\right)^{(1+\sqrt{5})/2}\right) & \text{if } \left(\eta_{k-1}^{(i)}\right)^{(1+\sqrt{5})/2} > 0.1 \qquad \text{for } i = 1, 2\,. \end{cases} \tag{2.42}$$

For the third choice $\eta = \eta_k^{(3)}$, since its condition in (2.41) is based only on the gradient values of $\mathcal{F}$, we implement another safeguard. The modification of $\eta_k^{(3)}$ reads:

$$\begin{cases} \eta_k^{(3)} = \eta_0^{(3)} & \text{if } \eta_k^{(3)} > 1 \\ \eta_k^{(3)} = \max\left(\eta_k^{(3)}, \gamma_1 \left(\eta_{k-1}^{(3)}\right)^{\gamma_2}\right) & \text{if } \gamma_1 \left(\eta_{k-1}^{(3)}\right)^{\gamma_2} > 0.1\,, \end{cases} \tag{2.43}$$

for γ_1 and γ_2 as in (2.41).

3b. At higher frequencies, waves detect and carry more detailed information about the scatterer [14]. Yet the number of local minima of the full or reduced-space objective functions in (2.16) or (2.23) may also increase. To minimize the risk of converging

to a (false) local minimum, we apply frequency continuation strategy [14, 6, 71]. For the lowest frequency we solve the inverse problem, then increase the frequency, solve again, and so on, initializing each optimization process at each new frequency from the previous optimal u at the lower frequency and obtain improving solutions for the optimization problem. This strategy is also denoted in the literature as *frequency stepping*.

2.4 Inexact (Quasi)-Newton Algorithm

We summarize the information above in the following algorithm.

Inexact (Quasi-)Newton Algorithm.
Input: initial guess $u_0 = 1$, Output: u^*.

1. For $\omega = \omega_1, \ldots, \omega_n$
   (a) Compute $\mathcal{F}(u_0)$, $\nabla_u \mathcal{F}(u_0)$
   (b) set $\eta_0 = 0.7$
   (c) for $k = 0, 1, 2 \ldots$ UNTIL $\|\nabla_u \mathcal{F}(u_k)\| \leq tol$
      i. Solve $H\delta u_k = -\nabla_u \mathcal{F}(u_k)$ using CG UNTIL
      $\|H\delta u_k + \nabla_u \mathcal{F}(u_k)\| \leq \eta_k \|\nabla_u \mathcal{F}(u_k)\|$
      ii. Find step α_k s.t. $u_{k+1} := u_k + \alpha_k p_k$
      iii. Update $\mathcal{F}(u_{k+1})$, $\nabla_u \mathcal{F}(u_{k+1})$
      iv. compute η_{k+1} using (2.41)
      v. apply safeguards on η_{k+1} using (2.42) or (2.43)
   (d) $u_0 = u_{k+1}$
2. $u^* = u_{k+1}$

Remarks:

- Step 1(c)i is computed without storing the matrix H explicitly (matrix-free).
- In step 1(c)ii we can use the Armijo line-search method (Newton, Gauss-Newton methods) or the Wolfe conditions (BFGS).
- In step 1(c)iv we use $\eta_k^{(1)}$ or $\eta_k^{(2)}$, but it is also possible to use $\eta_k^{(3)}$ and get very good results. The values of γ_1 and γ_2 then need to be chosen carefully. For inverse Helmholtz problems, we recommend the values of $\gamma_1 = 0.9$ and $\gamma_2 = 1.1$. In addition, if we choose to use $\eta_k^{(3)}$, we need to use in step 1(c)v the safeguards described in (2.43).

2.5 Regularization Problems

Even when using the criteria of Eisenstat-Walker [22], their safeguards [22, 63] and frequency stepping [14, 6, 71], we are still confronted with regularization problems. The regularization problems are reflected in unwanted and non-reasonable perturbations in the reconstruction of the medium. We repeat the simulation at the beginning of this section and reconstruct u displayed in Fig. 2.3 at the frequencies $\omega = 8, 10, 20, \ldots, 80, 90$, using frequency stepping. Here, we use $\eta_k^{(2)}$ as forcing term and the corresponding safeguards (2.42). The results are shown in Fig. 2.5. Through the criteria, the safeguards and the frequency stepping, we get a great improvement over the solution. Still, it is not entirely satisfactory, since we observe non negligible artifacts around the sources and on the boundary where the receivers are.

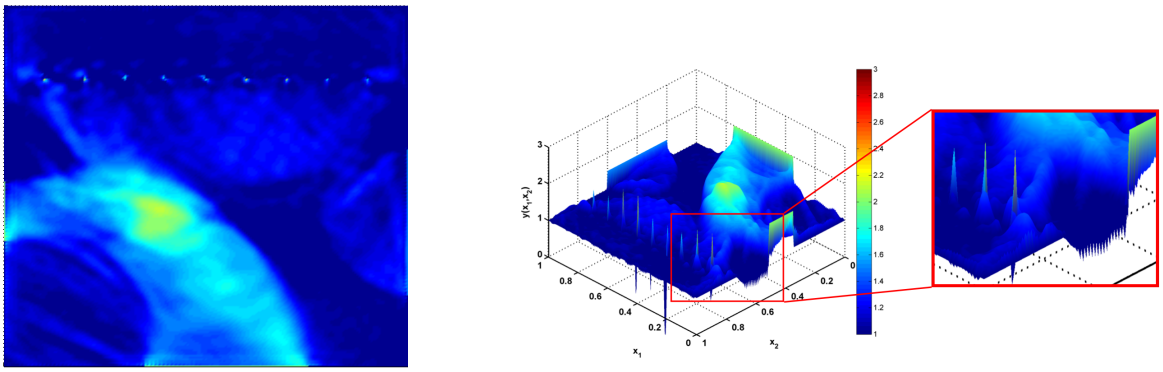

Figure 2.5: Eisenstat-Walker criteria and frequency stepping. Reconstruction of the profile described in Fig. 2.3 with relative L^2-error= 24.55%. Two-dimensional view (left) and three-dimensional view (right).

Unfortunately, the inexact Newton method and the frequency stepping are not able to remove all artifacts of the ill-posed problem. The last example illustrates typical artifacts and perturbations in the solution and the need for extra regularization when solving the inverse problem. For this purpose, we show in the next chapter how to regularize the optimization problem to prevent unreasonable solutions.

Chapter 3

Regularization of the Inverse Problem

In Section 2.3, we introduced inexact Newton or quasi-Newton methods and used forcing terms, safeguards and frequency stepping to enable an efficient and stable optimization. Since the inverse problems (2.17) and (2.23) are generally ill-posed, an artificial regularization term is typically added. The term regularization method refers to any method constructing a stable approximate solution to an ill-posed problem [11]. For such large scale of ill-posed inverse problems, several efficient methods exist. In Section 1.3, we presented some regularization methods for the FWI. In the following, we focus on an important regularization-tool, the Tikhonov regularization [82, 31, 23]. We suggest some additional a priori information on the control variable and add a regularization functional $\mathcal{R}(u)$ to the minimization problem. The additional functional penalizes the optimization problem in a case where the control u does not fit with the available information. For example, if we have a priori information that the profile u is smooth, one may choose $\mathcal{R}(u)$ as a functional which penalizes high derivatives in u. The Tikhonov regularization can tackle the ill-posedness with a small computational cost and will be the regularization tool in this chapter.

3.1 Tikhonov Regularization for Inverse Helmholtz Problems

In this section, we consider Tikhonov regularization [82, 31, 23] and introduce some Tikhonov regularization functionals. Tikhonov regularization is a very powerful regularization tool, which requires only small computing cost. When a priori information about the control variable is available, we may add to the misfit functional a regularization functional $\mathcal{R}(u)$ to penalize parameters u, which do not fit with the information. The minimization

problem

$$\text{Find } u \in S\,,\ \text{s.t. } u = \underset{v \in S}{\operatorname{argmin}}\, \mathcal{F}(v)\,, \tag{3.1}$$

then reads

$$\text{Find } u \in S\,,\ \text{s.t. } u = \underset{v \in S}{\operatorname{argmin}} \{\mathcal{F}(v) + \alpha \mathcal{R}(v)\}\,, \tag{3.2}$$

where $\alpha \in \mathbb{R}_{\geq 0}$ is chosen to weigh the functional $\mathcal{R}(v)$ to the misfit functional $\mathcal{F}(v)$. The choice of α is critical for the optimization (3.2): if α is too high, such that $\mathcal{F}(v) \ll \alpha\, \mathcal{R}(v)$, the solution u of (3.2) will essentially minimize only $\mathcal{R}(v)$. If α is too small, such that $\mathcal{F}(v) \gg \alpha\, \mathcal{R}(v)$ the optimization process will essentially ignore the additional information in $\mathcal{R}(v)$. Hence, the parameter α must balance between the two functionals $\mathcal{F}(v)$ and $\mathcal{R}(v)$ [84], namely

$$\mathcal{F}(v) \approx \alpha\, \mathcal{R}(v)\,.$$

It is important to keep the term $\alpha\, \mathcal{R}(v)$ slightly smaller than $\mathcal{F}(v)$, because our final goal is to minimize $\mathcal{F}(v)$, which needs to carry a higher weight in the optimization. Automatic techniques for the choice of α exist. For example we can mention the L-curve criterion [42], Morozov's discrepancy principle [62], dynamic prior regularization parameter [4] and more.

Tikhonov Regularization Functionals

For Tikhonov regularization, we can choose different kinds of penalty functionals $\mathcal{R}(u)$ [88]. The standard Tikhonov *L^2-penalty* functional is

$$\mathcal{R}_{L^2}(u) = \frac{1}{2}\|u\|_{L^2}^2\,. \tag{3.3}$$

Another penalty functional is the *Sobolev H^1-penalty* functional

$$\mathcal{R}_{\nabla u}(u) = \frac{1}{2}\int_\Omega \sum_{i=1}^{d} \left(\frac{\partial u}{\partial x_i}\right)^2 dx\,, \tag{3.4}$$

which penalizes strong variation in the solution. These two penalty functionals, can be generalized [31] by

$$\mathcal{R}_D(u) = \frac{1}{2}\|D\,u\|_{L^2}^2\,. \tag{3.5}$$

Usually, D is a first or second space derivative operator, but other choices are possible. If D is the identity operator, we have standard L^2-Tikhonov regularization. The operator D in the functional $\mathcal{R}(u)$ can give us information about the model. For example, in seismic models, the profile is a layered material, which is represented by a non-smooth u in the vertical direction, but smooth in the horizontal one. We can use the operator D to penalize the derivatives only in one direction.

Another regularization functional, for example, is the Gaussian regularization operator [21], which is non-convex

$$R_{Gauss}(u) = \frac{1}{2}\int_\Omega 1 - \exp\left(-\frac{|\nabla u|^2}{\sigma^2}\right) dx\,, \quad \sigma > 0\,. \tag{3.6}$$

Another non-convex regularization functional is the Lorentzian penalty term

$$R_{Lorentz}(u) = \frac{1}{2}\int_\Omega \frac{\gamma|\nabla u|^2}{1+\gamma|\nabla u|^2}\, dx\,, \quad \gamma > 0\,. \tag{3.7}$$

Both penalty functionals in (3.6) and (3.7) penalize strong variation in the solution and contain an extra parameter to allow discontinuities.

An important penalty functional is the penalized Total Variation (TV) [74, 89]. The penalized TV regularization was used originally in digital image processing, for noise removal. The method was able to remove unwanted noise, while preserving important details. In inverse Helmholtz problems, the penalized TV regularization is a good choice for discontinuous and noisy data. Unlike the group of regularizations given by (3.5), the penalized TV regularization uses the L^1 norm and reads

$$R_{TV}(u) = \frac{1}{2}\int_\Omega \sqrt{|\nabla u|^2 + \varepsilon^2}\, dx\,, \quad \varepsilon \neq 0\,. \tag{3.8}$$

Note that the penalized TV-regularization functional is differentiable. This regularization penalizes the gradient of u and allows discontinuities in the profile through the penalization parameter ε as well.

Gradient of the Regularization Functionals

In the solution of the optimization of (3.2) through Newton-like methods, we need the gradient of the functional $\mathcal{R}(u)$. We calculate $\nabla_u \mathcal{R}(u)$ through the variation of $\mathcal{R}(u)$:

$$\frac{d}{dt}\mathcal{R}(u + t\varphi)\Big|_{t=0}\,, \tag{3.9}$$

where φ is the direction of the derivation and $t \in \mathbb{R}$. Using (3.9) and the definition of a gradient of an operator $\mathcal{R}$, we calculate the gradients of the functionals (3.3), (3.4), (3.6), (3.7) and (3.8):

$$\nabla_u \mathcal{R}_{L^2}(u) = u\,, \tag{3.10}$$

$$\nabla_u \mathcal{R}_{\nabla u}(u) = -\Delta u\,, \tag{3.11}$$

$$\nabla_u \mathcal{R}_{Gauss}(u) = -\nabla \cdot \left(\frac{\nabla u}{\sigma^2 \exp\left(-|\nabla u|^2/\sigma^2\right)}\right)\,, \quad \sigma > 0\,, \tag{3.12}$$

$$\nabla_u \mathcal{R}_{Lorentz}(u) = -\nabla \cdot \left(\frac{\gamma \nabla u}{(1+\gamma|\nabla u|^2)^2} \right), \quad \gamma > 0\,, \tag{3.13}$$

$$\nabla_u \mathcal{R}_{TV}(u) = -\nabla \cdot \left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \varepsilon^2}} \right), \quad \varepsilon \neq 0\,. \tag{3.14}$$

In all calculations of the gradients (3.10)-(3.14), we assume Neumann boundary conditions for u.

The gradients (3.10)-(3.14) of the regularization functionals play a main role in parts II and III of this thesis.

3.2 Numerical Experiments

Now, we present numerical results for the inverse problem (2.22) with the objective functional (2.23) using the inexact (quasi-)Newton algorithm from Section 2.4. As regularisation we consider three different Tikhonov penalty functionals (3.3), (3.4) and (3.8). In all cases we choose α proportional to the misfit functional [84] as

$$\alpha = \tilde{\alpha}\, \frac{\|\mathcal{F}(u_k)\|}{\|\mathcal{R}(u_k)\|}\,, \quad 0 < \tilde{\alpha} < 1\,. \tag{3.15}$$

The model problem and the parameters settings for the following numerical experiments are the same as in Section 2.5. The results for the L^2-penalty functional, the Sobolev H^1-penalty functional and the penalized TV-regularization are shown in Fig. 3.1.

The L^2-penalty functional is not able to remove most of the artifacts and the inverse problem is not well regularized. In contrast, the Sobolev H^1-penalty functional and the penalized TV-regularization are able to remove the artifacts almost completely. However, the discontinuities in the true profile u are smoothed out in the reconstructions.

An important set-up for Tikhonov regularization is the parameter α appears in (3.2). As explained in Section 3.1, the parameter α is chosen to weigh the penalty functional $\mathcal{R}$ to the minimization problem and its choice is critical for the reconstruction. Unfortunately, finding the right parameter α for all profiles is difficult [4]. In penalized TV-regularization (3.8), an additional parameter ε must be determined. The parameter ε allows solutions with strong variability or discontinuities. Hence, a bad choice of ε might disturb the reconstruction to resolve the discontinuities in the parameter.

In the bottom of Fig. 3.1, we see the reconstruction of the profile u with TV-regularization using the optimal α and ε ($\tilde{\alpha} = 0.3$ in (3.15) and $\varepsilon = 10$). In Fig. 3.2 we illustrate four results for TV-regularization with unsuitable choices for α and ε. Note however that those values are a priori not unrealistic choices and they are fairly close to the optimal ones used in Fig. 3.1, as we slightly change $\tilde{\alpha}$ in (3.15).

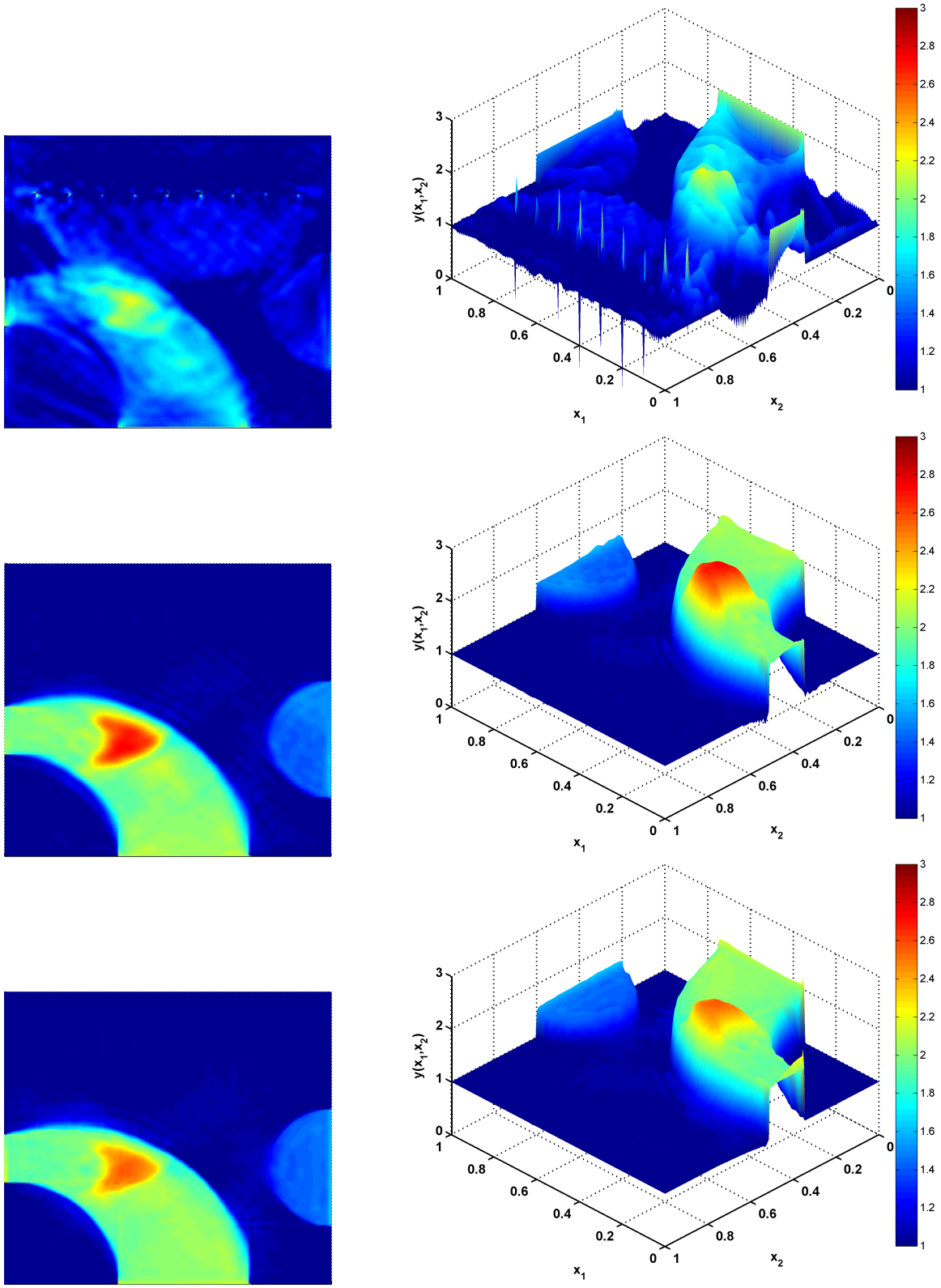

Figure 3.1: Reconstruction of the profile u with Tikhonov regularization. Two-dimensional view (left) and three-dimensional view (right). From top to bottom: L^2-penalty functional with relative L^2-error= 24.16%, Sobolev H^1-penalty functional with relative L^2-error= 9.25%, penalized TV-regularization with $\varepsilon = 10$ and relative L^2-error= 10.02%.

In all cases illustrated in Fig. 3.2 we get solutions where the discontinuities in the profile u are not well reconstructed or smoothed out. In summary these numerical experiments show that Tikhonov regularization is quite sensitive to the precise choice of α and ε.

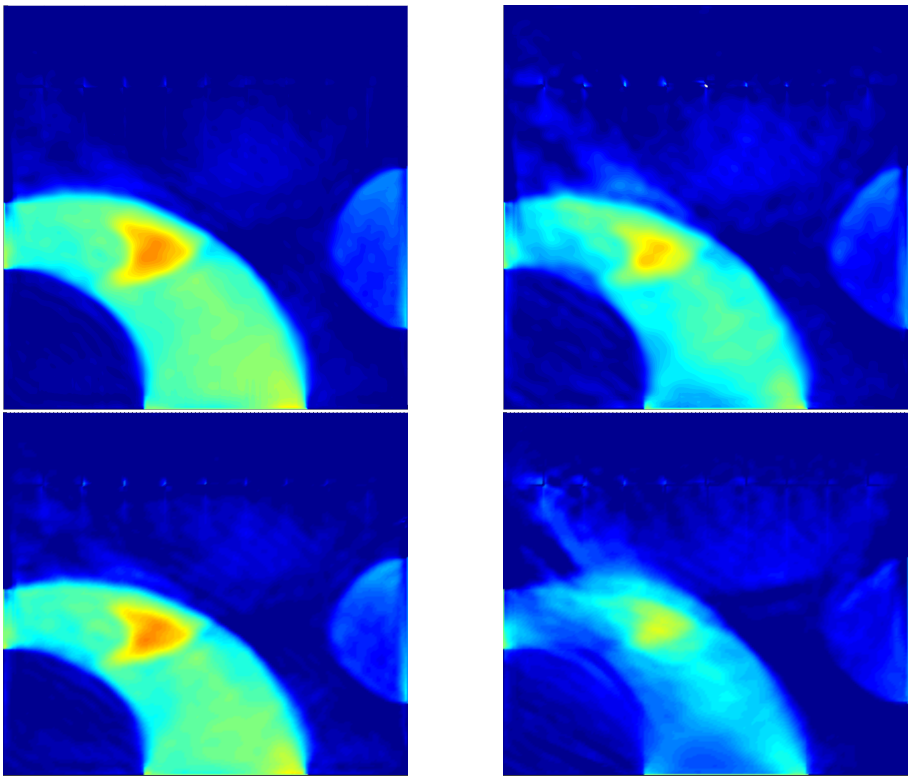

Figure 3.2: Reconstruction of the profile u with penalized TV-regularization. Top: too high value of α with $\tilde{\alpha} = 0.4$ in (3.15) (left) and too small value of α with $\tilde{\alpha} = 0.1$ (right). Bottom: too large $\varepsilon = 50$ (left) and too small $\varepsilon = 1$ (right).

Hence, we aim for another way to regularize, where it is easier to choose the regularization parameters. In Part II we use efficient and a priori determined regularization parameters which avoid the need for standard Tikhonov-type regularization.

Part II

Adaptive Eigenspace for the Inverse Helmholtz Equation

Chapter 4

Adaptive Eigenspace Inversion (AEI)

In this chapter, we present the Adaptive Eigenspace Inversion (AEI) method for inverse medium problems in the frequency domain. The method is a combination between two regularization types: the first one is regularization by size reduction of the set of admissible parameters and the second is adaptive regularization [13], see Section 1.3.3. In the AE method the unknown wave speed is projected to a particular finite-dimensional basis of eigenfunctions, which is iteratively adapted during the frequency stepping. Depending on the frequency we truncate the adaptive eigenspace (AE) basis at a small and slowly increasing finite number of eigenfunctions. This effectively introduces regularization into the inversion, which is an alternative to the standard Tikhonov-type regularization. Hence, in the following chapters, we consider the reduced space objective function (2.23) without any added regularization, i.e. $\alpha = 0$.

Combining these two regularizations, we introduce the adaptive eigenspace representation for u, where both the dimension and the basis functions themselves are adapted during the optimization process. Chapters 4, 5 and 6 can be essentially found in [33].

4.1 Adaptive Eigenspace Basis

Instead of a standard nodal basis (FD, FE grid-based discretization), we shall use a basis of (global) eigenfunctions $\{\phi_m\}_{m\geq 1}$ to represent the parameter $u(x)$ as

$$u(x) \;=\; u_0(x) \;+\; \sum_{m\geq 1} \beta_m \phi_m(x). \tag{4.1}$$

Here the "background" $u_0(x) \in H^1(\Omega)$ solves the elliptic problem:

$$\begin{cases} -\nabla\cdot(\mu(x)\nabla u_0(x)) &= 0, & \forall\, x\in\Omega,\\ u_0(x) &= c^2(x), & \forall\, x\in\Gamma, \end{cases} \tag{4.2}$$

where $\mu(x)$ is defined by

$$\mu(x) \;=\; \frac{1}{\max\{|\nabla u(x)|, \varepsilon\}}, \qquad \forall\, x \in \Omega,\ \varepsilon > 0. \tag{4.3}$$

The parameter $\varepsilon > 0$, which ensures that the denominator of μ does not vanish, is typically set to a very small value such as $\varepsilon = 10^{-6}$. While the primary role of u_0 is to accommodate the (known) inhomogeneous boundary values of u, we shall demonstrate in Chapter 5, that u_0 also captures much of the behavior of u in the interior.

Following [17], we choose for the functions ϕ_m the orthonormal basis of eigenfunctions $\phi_m \in H_0^1(\Omega)$ of the elliptic operator,

$$\left\{ \begin{array}{rcll} -\nabla \cdot (\mu(x) \nabla \phi_m(x)) & = & \lambda_m \phi_m(x), & \forall\, x \in \Omega, \\ \phi_m(x) & = & 0, & \forall\, x \in \Gamma, \end{array} \right. \tag{4.4}$$

with corresponding eigenvalues $0 < \lambda_m \leq \lambda_{m+1}$, $m \geq 1$. Clearly, at higher λ_m, the eigenfunctions ϕ_m in (4.4) will be increasingly oscillatory.

In Chapter 5, we shall provide analytical and numerical evidence which underpins the remarkable accuracy of this basis for representing any given $u(x)$. In our AEI approach, the eigenfunctions $\{\phi_m\}_m$ and the background u_0 are repeatedly recomputed as the underlying control $u(x)$ varies during the optimization. Hence, we call $\{u_0\} \cup \{\phi_m\}_{m \geq 1}$ an adapted eigenspace (AE) basis.

Since $u(x)$ is precisely the quantity we seek, and thus unknown, we always use in (4.3) the value from the previous optimization step. At the first step, when no information about u inside Ω is available yet, we simply set $\mu(x) \equiv 1$. Then, u_0 is a harmonic prolongation of c^2 from Γ into Ω while the basis $\{\phi_m\}_{m \geq 1}$ simply corresponds to the eigenfunctions of the Laplacian operator in Ω.

Remark 1. The elliptic operator in (4.2) and (4.4) essentially coincides with the gradient of the penalized total variation (TV) regularization term [74, 89],

$$R_{TV}(u) \;=\; \frac{1}{2} \int_\Omega \sqrt{|\nabla u|^2 + \varepsilon^2}\, dx,$$

given by

$$\begin{aligned} \nabla R_{TV}(u) \;&=\; -\nabla \cdot \left(\frac{1}{\sqrt{|\nabla u|^2 + \varepsilon^2}} \nabla u \right) \\ &\simeq\; -\nabla \cdot (\mu(x) \nabla u)\,. \end{aligned}$$

Penalized TV-regularization is well-known in image processing for noise removal, while preserving sharp interfaces. Instead of adding a Tikhonov regularization term to the objective functional, the AEI approach projects u to the basis of eigenfunctions of the gradient of the penalized TV-regularization functional; hence, the AE basis inherits similar properties. The relation between the gradient of Tikhonov-regularization functionals and the AEI will be discussed in detail in Chapter 8.

4.2 Adaptive Eigenspace Inversion Algorithm

In practice, we truncate the infinite sum in (4.1) at a finite number $K \geq 1$:

$$u(x) \;=\; u_0(x) \;+\; \sum_{m=1}^{K} \beta_m \phi_m(x). \tag{4.5}$$

To keep both the memory requirements and the computational effort low, it is imperative to keep the number K of eigenfunctions minimal. The truncation of the eigenfunction expansion is also crucial for numerical stability, as it builds regularization into the AEI approach – see Remark 1 above but also Remark 3 below.

At higher frequencies, waves detect and carry more detailed information about the scatterer, yet the number of local minima of $\mathcal{F}(u)$ may also increase. To minimize the chance of converging to a (false) local minimum, we also apply a standard frequency continuation procedure [14, 6]. First, we solve the inverse problem (2.22) at the lowest frequency ω_1. Then we progressively increase $\omega = \omega_2, \ldots, \omega_n$ while re-initializing the optimization at every ω_j from the previous lower frequency ω_{j-1}. In doing so, we assume that the measurements are available through a range of frequencies, for instance via Fourier transform of a time-dependent signal.

AEI Algorithm.
Input: initial guess $u = 1$, observations y_ℓ^{obs}. Output: u^*.

1. Choose $K \geq 1$ and compute $\{\phi_m\}_{m=1}^K$ from (4.4) and u_0 from (4.2) with $\mu \equiv 1$
2. Expand $u(x) = u_0(x) \;+\; \sum_{m=1}^K \beta_m \phi_m(x)$
3. For $\omega = \omega_1, \ldots, \omega_n$
   (a) Compute $\mathcal{F}(u)$ and $\nabla\mathcal{F}(u)$, set H
   (b) STOP: if $\|\nabla\mathcal{F}(u)\| \leq$ Tol
      i. Solve $Hp = -\nabla\mathcal{F}(u)$
      ii. Determine step size α and set $u := u + \alpha p$
      iii. Update $\mathcal{F}(u)$, $\nabla\mathcal{F}(u)$ and H
   (c) Set μ from (4.3) with ∇u
   (d) Update K, compute $\{\phi_m\}_{m=1}^K$ from (4.4) and u_0 from (4.2)
   (e) Expand $u(x) = u_0(x) \;+\; \sum_{m=1}^K \beta_m \phi_m(x)$
4. $u^* = u$

The AEI approach applies regardless of the underlying optimization method used. Here we consider truncated Newton-like methods [19, 63] and denote by H either the true Hessian or some approximation of it, depending on the method (Newton, BFGS, or Gauss-Newton)

used. In all cases the linear system in 3(b)i is solved by a truncated CG-iteration with the Eisenstat-Walker criterion [22]. In 3(b)ii, the step size α of the search direction p is determined either by Armijo (Newton, Gauss-Newton) or Wolfe (BFGS) step-size control, depending on the underlying method [65]. In steps 1 and 3d of the AEI Algorithm, we compute the first K eigenfunctions in (4.4) by using a standard restarted Lanczos iteration [53].

Chapter 5

Approximation Properties of the AE Basis

In the AEI method presented in Chapter 4, the (unknown) parameter u is expanded as in (4.5) in the L^2-orthogonal basis of eigenfunctions $\{\phi_m\}_{m=1,\dots,K}$ defined by (4.4) together with u_0 defined by (4.2). In this entire section, we shall assume that u is known and shall now provide some analytical and some numerical evidence which underpins the remarkable accuracy provided by our particular choice of u_0 and the AE basis.

5.1 One-Dimensional Case

In one space dimension, (4.2) reduces to

$$\begin{cases} -\dfrac{d}{dx}\left(\mu(x)\dfrac{d}{dx}u_0(x)\right) = 0 \qquad \forall\, x \in (a,b), \\ \qquad u_0(a) = c^2(a)\,, \quad u_0(b) = c^2(b), \end{cases} \tag{5.1}$$

where

$$\mu(x) = \frac{1}{\max\{|u'(x)|, \varepsilon\}}\,, \quad \forall x \in (a,b). \tag{5.2}$$

The following result is immediate.

Proposition 1. *Let $u : \overline{\Omega} \to \mathbb{R}$, $\Omega = (a,b)$, $a < b$, u continuous, piecewise differentiable and with $C \geq u'(x) \geq \varepsilon > 0$ or $-C \leq u'(x) \leq -\varepsilon < 0$, at every differentiability point $x \in \Omega$, $u(a) = c^2(a)$ and $u(b) = c^2(b)$. If u_0 solves (5.1) then $u_0(x) = u(x)$, $\forall x \in \overline{\Omega}$.*

Proof. We first assume that u is strictly increasing and therefore $u'(x) \geq \varepsilon > 0$. Since $\mu(x) = 1/\max\{|u'(x)|, \varepsilon\} = 1/u'(x)$, u itself also satisfies (5.1) with $u_0(a) = c^2(a)$ and $u_0(b) = c^2(b)$. As μ is strictly positive through Ω, the (weak) solution of (5.1) is unique and therefore $u_0(x) = u(x)$, $\forall x \in \overline{\Omega}$.
If u is strictly decreasing, the proof is similar for $\mu(x) = -1/u'(x) > 0$. □

Remark 2. From Proposition 1 we conclude that our choice for u_0 is in fact *optimal in one space dimension* when u' is strictly positive (or negative) throughout Ω, since it automatically yields u itself. The same conclusion immediately holds in higher dimensions, if the background medium is layered, that is if $u(x_1, x_2, \ldots, x_n)$ depends on a single variable. In general, however, u_0 will not equal u, in particular when u is not monotonic.

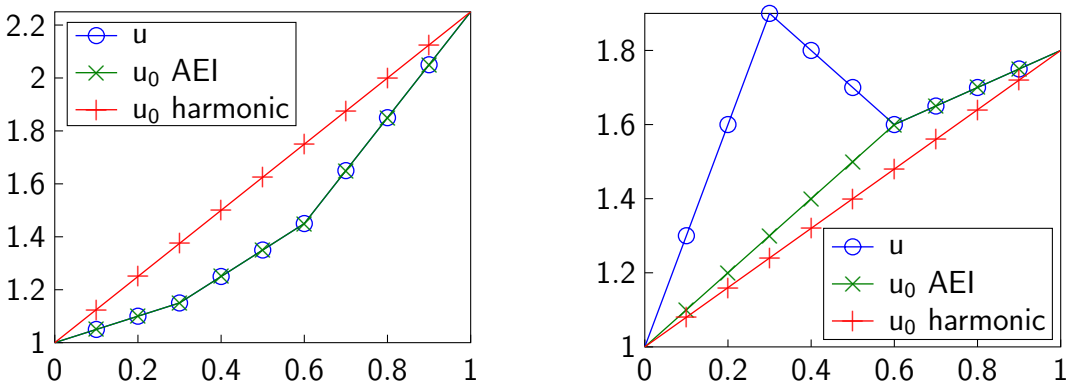

Figure 5.1: One-dimensional case. Comparison of adaptive and harmonic choices for u_0. Left: u monotonically increasing; right: u arbitrary.

First, we illustrate the usefulness of our particular choice for u_0 in (5.1)-(5.2) by comparing it to a straightforward choice where u_0 is harmonic and satisfies (5.1) with $\mu(x) \equiv 1$. If u is strictly increasing (or decreasing) over $\Omega = (0, 1)$, u_0 defined by (5.1)-(5.2) automatically coincides with u – see Proposition 1. Clearly, if u_0 is merely harmonic over Ω, it does not coincide with u, as shown in Fig. 5.1. If u both increases and decreases over Ω, neither the harmonic nor the adaptive u_0 will coincide with u. However, the adaptively computed u_0 will in general better approximate u inside Ω, as shown in Fig. 5.1; in fact, the adaptive u_0 coincides with u over $[0.6, 1]$.

Next, we approximate the difference $u - u_0$ in $\text{span}\{\phi_1, \phi_2, \ldots, \phi_K\}$, where all ϕ_m satisfy (4.4) in one space dimension, that is

$$\begin{cases} -\dfrac{d}{dx}\left(\mu(x)\dfrac{d}{dx}\phi_m(x)\right) = \lambda_m \phi_m(x) & \forall\, x \in (a, b), \\ \phi_m(a) = 0\,, \quad \phi_m(b) = 0, \end{cases} \tag{5.3}$$

with μ defined in (5.2).

Clearly, the local behavior of ϕ_m in the neighborhood of any fixed $x_0 \in \Omega$ is essentially determined by the magnitude of $C = |u'(x_0)|$. If $|u'(x)| \simeq C > \varepsilon$ in a neighborhood of x_0, ϕ_m essentially behaves like

$$\phi_m(x) \simeq A_m \sin(\sqrt{C\lambda_m}\, x) + B_m \cos(\sqrt{C\lambda_m}\, x)$$

near x_0 and hence is strongly oscillatory at higher m. However, if u is essentially constant near x_0, $|u'(x)| \simeq 0$, then $\mu = 1/\varepsilon$ there and ϕ_m essentially behaves like

$$\phi_m(x) \simeq A_m \sin(\sqrt{\varepsilon\lambda_m}\, x) + B_m \cos(\sqrt{\varepsilon\lambda_m}\, x).$$

Since ε is very small, ϕ_m will also remain essentially constant there for moderate values of m.

To illustrate this behavior, we now consider the piecewise linear profile u shown in Fig. 5.2, where $u'(x) = 3, 0$ and -9 on the interval $(0, 0.3)$, $(0.3, 0.9)$ and $(0.9, 1)$, respectively. Since u' is piecewise constant, we can immediately determine all eigenfunctions as:

$$\phi_m(x) = \begin{cases} A_{m,1} \sin(\sqrt{3\lambda_m}\, x) + B_{m,1} \cos(\sqrt{3\lambda_m}\, x), & x \in [0, 0.3), \\ A_{m,2} \sin(\sqrt{\varepsilon\lambda_m}\, x) + B_{m,2} \cos(\sqrt{\varepsilon\lambda_m}\, x), & x \in [0.3, 0.9), \\ A_{m,3} \sin(\sqrt{9\lambda_m}\, x) + B_{m,3} \cos(\sqrt{9\lambda_m}\, x), & x \in [0.9, 1]. \end{cases} \tag{5.4}$$

In Fig. 5.2, we show u together with some of the eigenfunctions from (5.3). On every subinterval, ϕ_m has a different frequency determined by the local value of $\sqrt{C\lambda_m}$. In $[0.3, 0.9)$, the frequency $\varepsilon\lambda_m$ is very small, as $\varepsilon = 10^{-4}$, and ϕ_m appears essentially constant. As λ_m further increases, the frequency $\sqrt{\varepsilon\lambda_m}$ increases as well and oscillations appear. Clearly, the smaller ε, the more eigenfunctions ϕ_m essentially behave as constants wherever u is essentially constant. We remark that ϕ_1 nearly coincides with u up to a scaling factor.

Finally, we consider for u the smooth polynomial profile

$$u(x) = 1 - 89.95\, x^6 + 281.35\, x^5 - 339.51\, x^4 + 199.56\, x^3 - 60.85\, x^2 + 9.40\, x.$$

It is shown in Fig 5.3 together with the first eigenfunction ϕ_1 obtained from (5.3) either with $\mu \equiv 1$ or μ as in (5.2). Again, we observe that the adaptively determined eigenfunction ϕ_1 captures well the main features of u up to a scaling factor, unlike the first eigenfunction of the Laplacian. To reach a relative error below 1% for the best approximation with respect to the L^2-norm, only the first six eigenfunctions $\phi_1, \ldots, \phi_6$ are needed in the adaptive case. In contrast, the first six eigenfunctions of the Laplacian yield a seven times larger relative L^2-error, which drops below 1% only once thirteen eigenfunctions are included in the approximation; hence, the AE basis better captures the essential information about u.

5.2 Two-Dimensional Case

To illustrate the remarkable approximation properties of the AE basis in two space dimensions, we now consider the profile $u(x)$, $x = (x_1, x_2)$, shown in Fig. 2.3. Next, we compute u_0 from (4.2) with $\varepsilon = 10^{-6}$ and μ as in (4.3). In Fig. 5.4, we observe that u_0 matches with remarkable accuracy the background medium but misses the embedded kite-shaped obstacle. The first eigenfunction ϕ_1 from (4.4), however, ignores the background and capture precisely the remaining obstacle. Using u_0 and ϕ_1, we expand u as in (4.5) with $K = 1$ and compute its best L^2-approximation. Shown in Fig. 5.4, it is hardly distinguishable from the true u with well-defined sharp contours and a relative L^2-error below 2%.

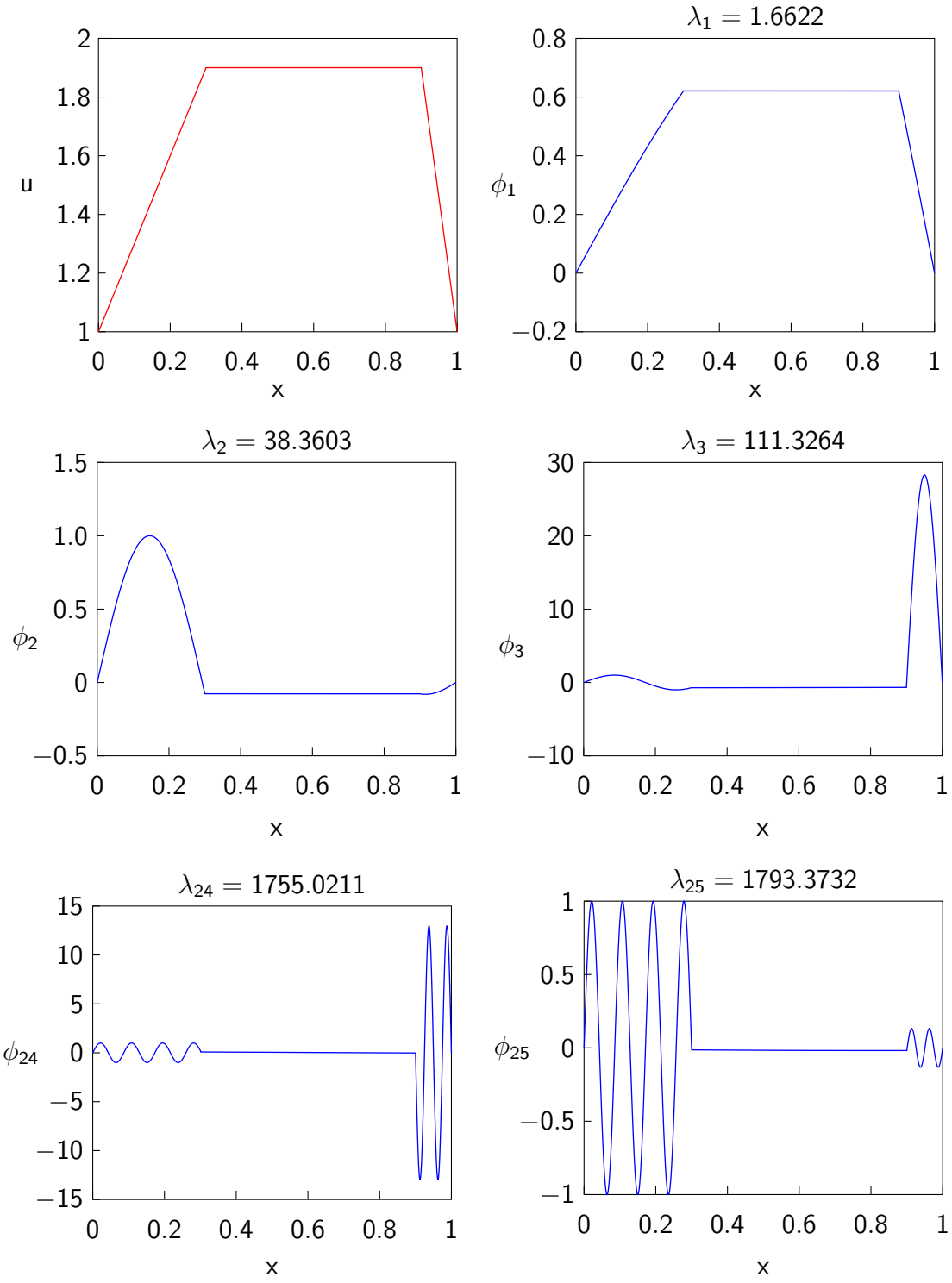

Figure 5.2: One-dimensional case. The true profile u (top left), together with the eigenfunctions ϕ_1, ϕ_2, ϕ_3, ϕ_{24} and ϕ_{25} from (5.3) with $\varepsilon = 10^{-4}$.

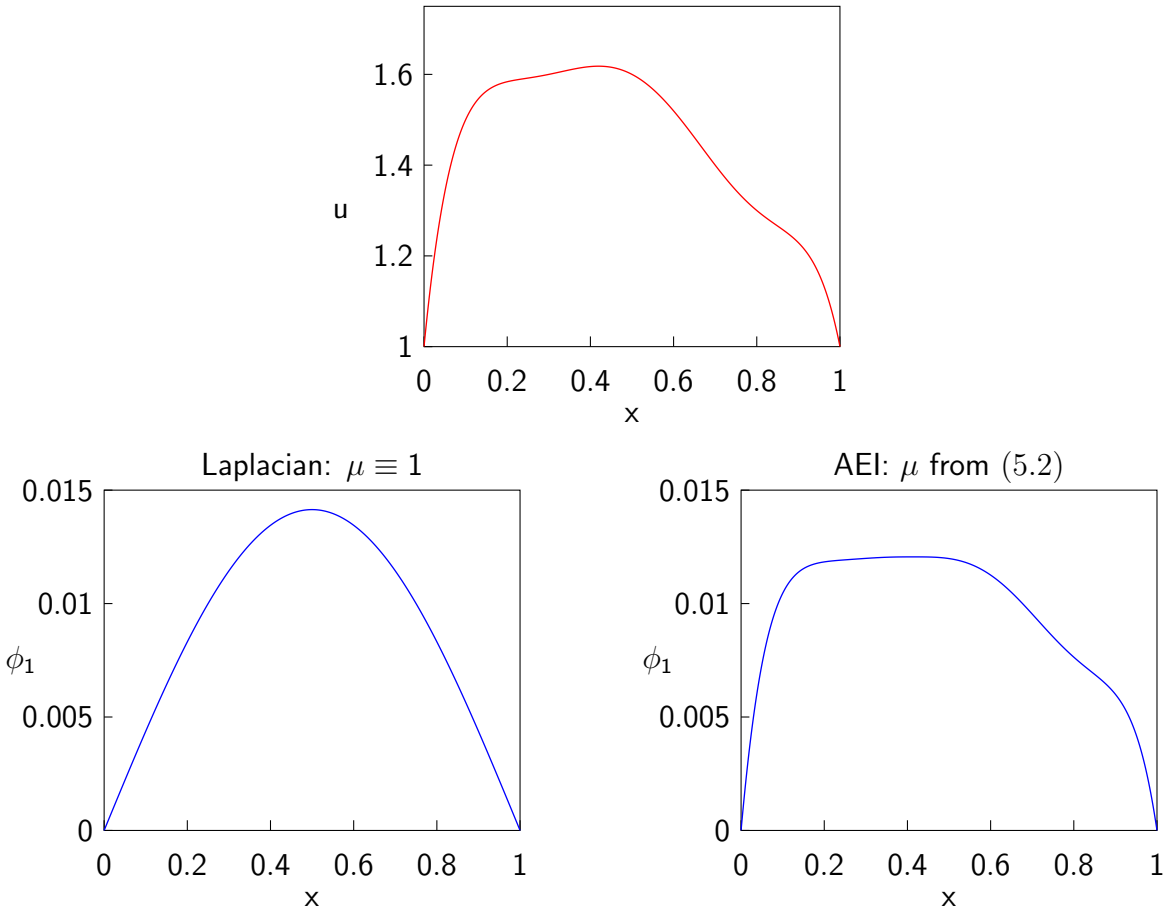

Figure 5.3: One-dimensional case. The true smooth profile u (top) and the first eigenfunction ϕ_1 from (5.3): $\mu \equiv 1$ (left), and μ as in (5.2) (right).

In contrast, if we repeat the same experiment with $\mu \equiv 1$, we observe in Fig. 5.4 how u_0 indeed matches the boundary values of u but fails to capture any additional features inside Ω. Similarly, the first eigenfunction of the Laplacian is independent of $u(x)$ and thus, as expected, carries no information about it. Again, we expand u as in (4.5) but now use the first 1000 eigenfunctions of the Laplacian instead. Although the L^2 best approximation with 1,000 Laplacian eigenfunctions now yields a reasonable approximation of u with 6% relative L^2-error, the contours are blurred while small high-frequency ripples appear due to the well-known Gibbs-phenomenon. These results illustrate the remarkable accuracy even of but a few eigenfunctions of the AE basis.

Remark 3. For a given profile u, the corresponding AE basis $V_K = \{u_0, \phi_1, \ldots, \phi_K\}$ of relatively small dimension usually yields a remarkably accurate representation of u. At higher eigenvalues λ_m, however, the (mutually orthogonal) AE basis functions ϕ_m become increasingly oscillatory and no longer carry useful information about u – see Fig. 5.5. Truncating the expansion in (4.5) at a finite value, K, thus effectively builds regularization into the AEI approach.

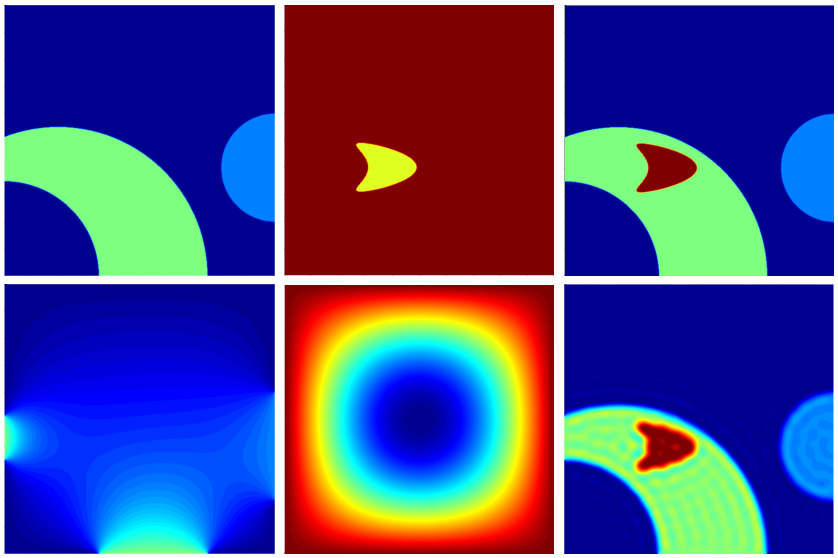

Figure 5.4: Two-dimensional case. Top, from left to right: the background u_0 from (4.2), the first AE eigenfunction ϕ_1 from (4.4), the L^2 best approximation of u using $\{u_0, \phi_1\}$. Bottom, from left to right: the harmonic u_0, the first Laplace eigenfunction ϕ_1, the L^2 best approximation of u using the first 1,000 Laplace eigenfunctions $\{u_0, \phi_1, \ldots, \phi_{1000}\}$.

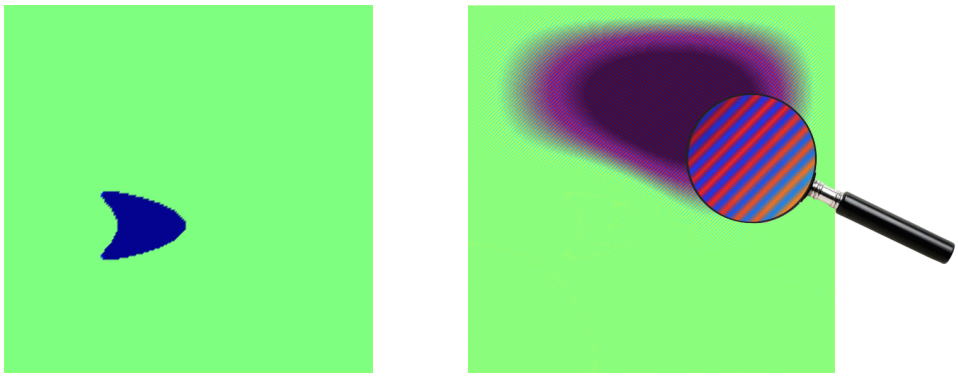

Figure 5.5: First AEI eigenfunction ϕ_1 (left) and last AEI eigenfunction ϕ_N (right) on a 200×200 grid.

Specification of the Number of Eigenfunctions K

We consider once more the true profile u described in Fig. 2.3 and observe in Fig. 5.5 the eigenfunction ϕ_1, respectively ϕ_N, corresponding to the smallest, respectively the largest, eigenvalue. We see that ϕ_1 is relevant for the reconstruction whereas ϕ_N contains only irrelevant information and is highly oscillating. The challenge is to find a way to choose the number of eigenfunctions K, that insures only relevant information. If we choose truncate (4.5) in this way, we insure no extra regularization term such as Tikhonov in

our formulation (2.23). Note that for this section, we focus on the reconstruction of the kite-shaped obstacle only.

Our aim is to sort out all ϕ_m that do not contain information about u. If u is constant almost everywhere, then we have $|\nabla u| = 0$ and thus we can rewrite (4.4), as

$$-\nabla \cdot \left(\frac{1}{\varepsilon} \nabla \phi_m(x) \right) = \lambda_m \phi_m(x),$$

or, since ε is constant, as

$$-\Delta \phi_m(x) = (\varepsilon \lambda_m) \phi_m(x)\,,$$

where the resulting ϕ_m actually do not contain information on u.

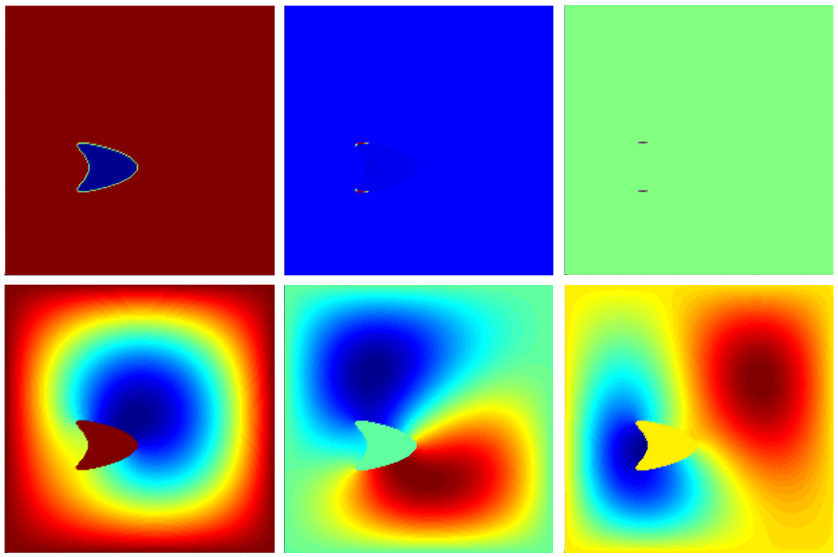

Figure 5.6: From top left to bottom right: ϕ_1, ϕ_3, ϕ_{131}, ϕ_{132}, ϕ_{133} and ϕ_{134}.

In Fig. 5.6, we display some eigenfunctions for the exact profile u. In the upper row, we see relevant eigenfunctions for the reconstruction which contain the variations of u and are constant outside the reconstructed parameter. On the contrary, the lower row Fig. 5.6 shows eigenfunctions after $m_0 = 132$, which contain information only outside the kite-shaped obstacle and are zero inside of it. The fundamental change in behavior of the eigenfunctions ϕ_m occurs at $m = 132$: at $m \leq 131$, ϕ_m clearly encodes relevant information about the obstacle whereas beyond $m \geq 132$ the eigenfunctions essentially behave like Laplace eigenfunctions orthogonal to the first 131.

Our aim is to automatically detect this transition. To sort out the eigenfunctions ϕ_m which are not relevant for the reconstruction we can solve the eigenvalue problem (4.4), with μ in (4.3) twice, once with ε and once with $\tilde{\varepsilon} = \varepsilon/\gamma$, $\gamma \geq 1$. For a certain $m_0 \geq 1$, we then observe that

$$\frac{\tilde{\lambda}_m}{\lambda_m} = \gamma\,, \qquad \text{for } m \geq m_0, \tag{5.5}$$

where $\tilde{\lambda}_m$ are the corresponding eigenvalues of (4.4) with $\tilde{\varepsilon}$. This is well illustrated on the left picture of Fig. 5.7. From this observation, we can set $K = m_0 - 1$ and we use all eigenfunctions ϕ_m, for $m \leq K$ in the reconstruction. In this way to determine K we need to solve two eigenvalue problems, which can be very costly especially for large systems.

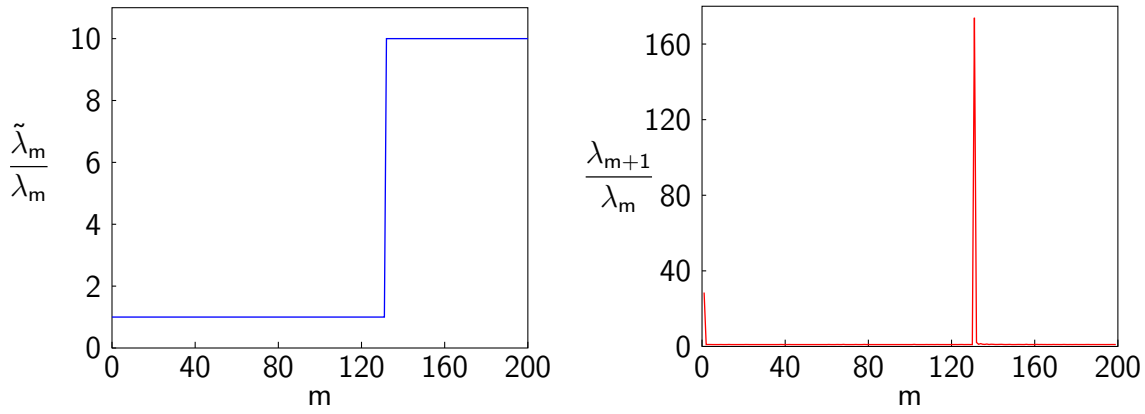

Figure 5.7: Left: ratio $\dfrac{\tilde{\lambda}_m}{\lambda_m}$ of (5.5), for $\gamma = 10$. Right: ratio $\dfrac{\lambda_{m+1}}{\lambda_m}$ of (5.6).

Another way to determine K by solving only one eigenvalue problem is to compute the following ratio

$$\frac{\lambda_{m+1}}{\lambda_m}. \tag{5.6}$$

From the observation above, we expect the ratio to be larger at one m_0, which corresponds to the m_0 from (5.5). On the right picture of Fig. 5.7, we display the ratio (5.6) and see the jump at m_0. Now, we can once more set $K = m_0 - 1$ by solving only once the eigenvalue problem (4.4).

In practice, when the profile u is unknown, we use $K \leq m_0 - 1$ as a regularization parameter, which allows us to avoid any additional regularization term in (2.23). We start with a small number K of eigenfunctions and when ω increases during the frequency stepping process, we also increase K. In doing so, we dramatically reduce the number of control variables and regularize the solution. Also, as K is chosen for each frequency ω, we avoid any additional computation of useless eigenfunctions. The choice of K is discussed in Section 6.4 of the next chapter.

Chapter 6

Numerical Experiments

We shall now illustrate the usefulness and versatility of the AEI method through a series of numerical experiments. Clearly, the squared velocity u of the medium is now unknown and we shall attempt to recover it from boundary measurements by solving (2.22) with the objective functional (2.23), where $\alpha = 0$.

First, we compare the adaptive eigenspace representation to a standard grid-based nodal representation of the control u to demonstrate the resulting significant reduction in degrees of freedom. Next we add yet another level of adaptivity by solving the auxiliary elliptic eigenvalue problem with an adaptive finite element method. We also present numerical experiments that underpin our choices for the background profile u_0 and the varying dimension of V_K with increasing frequency. Then, we show that the AE basis functions ϕ_m are highly localized, and hence easily sparsified; thus, the memory requirement of the adaptive eigenspace basis is kept to a minimum. Finally, we demonstrate the robustness of the AEI approach in the presence of noise or missing data and combine it with the sample average approximation (SAA) approach [38].

In the entire section, we consider the profile u displayed in Fig. 2.3 and reminded in Fig. 6.1, which mimics a layered material with regions of different wave speed. Unless specified otherwise, the typical parameter settings in the numerical experiments are the following: nine equispaced Gaussian sources are located along the upper boundary at $(0.1, 0.8), \ldots, (0.9, 0.8)$, whereas the receivers are located on the four lateral boundaries of $\Omega = (0, 1) \times (0, 1)$. We use second-order staggered finite differences on a 500×500 Cartesian mesh for the discretization of (2.2), (4.2)-(4.4). To avoid any inverse crime, the reference solution is computed on a separate finer mesh, which does not contain the coarser computational mesh.

In the AEI algorithm described in Section 4.2, we always set the initial guess to $u(x) \equiv 1$. Starting at the lowest frequency $\omega = 8$, we progressively increase the frequency $\omega = 10, 12, 14, \ldots, 90$. The number of eigenfunctions K starts at $K = 16$ or 32 and increases linearly with the frequency ω. For the optimization, we use a standard truncated Gauss-Newton method [35, 65] without extra regularization term. The search

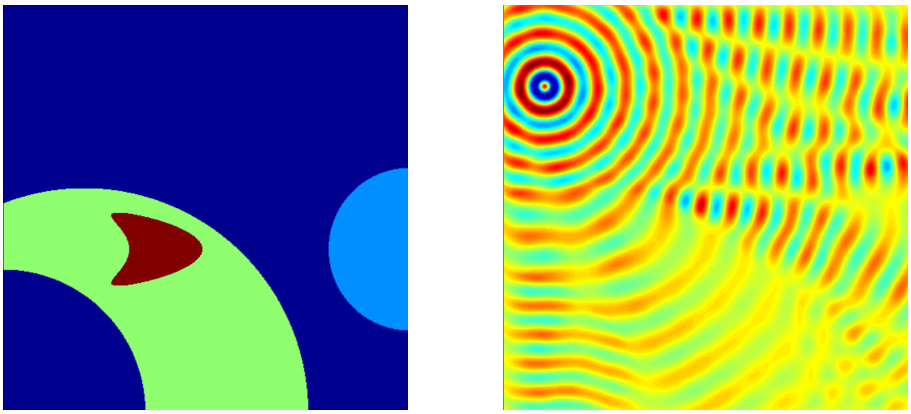

Figure 6.1: Left: true profile u; Right: real part of y_ℓ with f_ℓ at $(0.1, 0.8)$ and $\omega = 90$

direction is by a truncated Conjugate-Gradient iteration with the Eisenstat-Walker stopping criterion [22, 59, 63, 19] and the step-size by a standard Armijo rule. In the definition of μ in (4.3), we always set $\varepsilon = 10^{-6}$.

In the following numerical experiments, we either use finite differences for the *discretize-then-optimize* approach, or finite elements for the *optimize-then-discretize* approach. We shall not emphasize any particular choice for the discretization as it did not affect the results.

6.1 Adaptive Eigenspace vs. Nodal Basis

The use of an adaptive eigenspace (AE) basis for the control variable u instead of a standard grid-based nodal basis is the distinguishing feature of the AEI method. Thus, we now compare the AE vs. a standard nodal representation for the reconstruction of u. In both cases, we omit extra Tikhonov-type regularization.

First, we include an additional tenth source located at $(0.15, 0.15)$. The two corresponding reconstructed profiles are shown in Fig. 6.2. Although both methods recover the essential features of the medium, the AEI method clearly yields much crisper boundaries but also higher accuracy inside the various subregions. Moreover, the AEI method achieves the higher accuracy with fewer than $N_u = 360$ degrees of freedom vs. $N_u = 501,000$ for the standard nodal representation – see Table 6.1. Clearly, adding regularization would certainly remove some of the artifacts in the grid-based approach and thus yield a smoother, but not necessarily more accurate, reconstruction. See Fig. 3.1.

Next, we repeat the previous experiment but now omit all the receivers at the lower boundary of Ω together with the tenth source located at $(0.15, 0.15)$. Hence, much less information about the lower part of the medium is available in the data. Nevertheless, as shown in Fig. 6.3, the AEI method is still able to recover u everywhere inside Ω, unlike the standard grid-based approach. Indeed, as shown in Table 6.1, the relative L^2-error for the

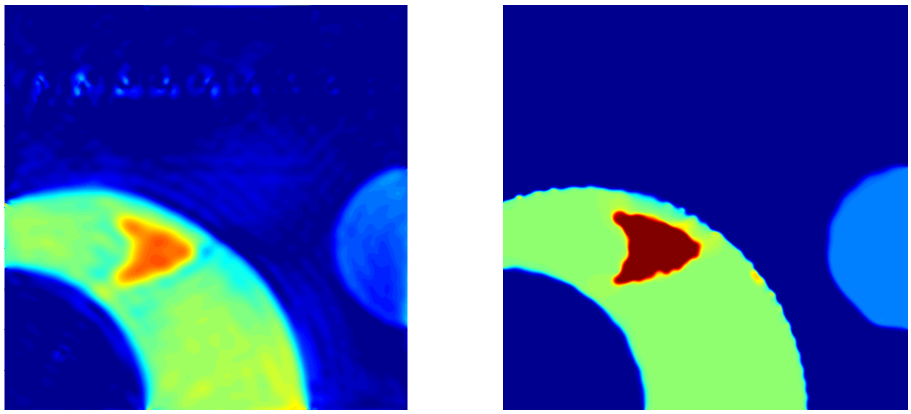

Figure 6.2: Adaptive eigenspace vs. nodal basis. Full boundary data: reconstruction with a nodal basis (left) or an AE basis (right).

	Full data		Partial data	
Basis	L^2-Error	N_u	L^2-Error	N_u
Nodal	15.91%	501,000	30.24%	501,000
AEI	4.65%	≤ 360	4.80%	≤ 360

Table 6.1: Adaptive eigenspace vs. nodal basis. Relative L^2-error and number of degrees of freedom for u.

nodal approach has now almost doubled whereas the error for the AEI method has hardly changed. Although the AEI method uses much fewer control variables than the grid-based approach, i.e. $N_u = K$, the reconstructions appear remarkably accurate and tolerant to missing data.

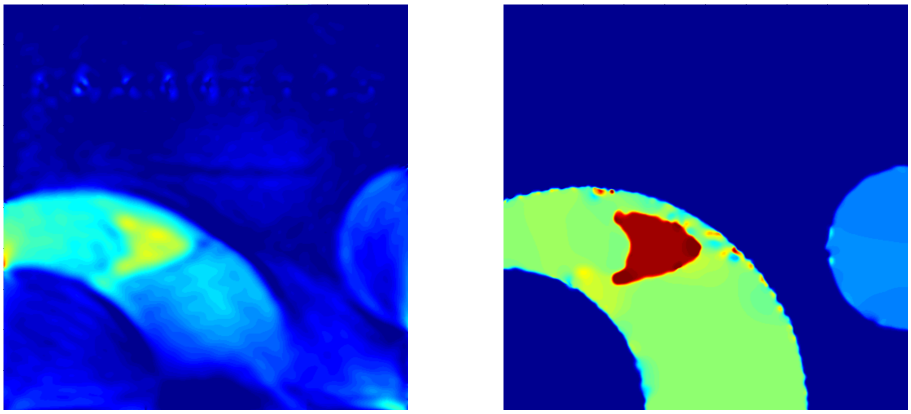

Figure 6.3: Adaptive eigenspace vs. nodal basis. Missing boundary data: reconstruction with a nodal basis (left) or an AE basis (right).

6.2 Adaptive Finite Element Discretization

The AEI method uses two separate computational meshes, the first for the forward and adjoint problems (2.21), (2.27) and the second for the auxiliary elliptic eigenvalue problems (4.2), (4.4). So far both meshes were spatially uniform. Here we include yet another level of adaptivity by adapting the finite element (FE) mesh in the solution of (4.2) and (4.4) to better capture small-scale features of u.

In the AEI Algorithm, mesh adaptation is performed after the while loop in step 3b and before the update in step 3c. We use a standard a priori FE adaptive mesh strategy based on the Hessian of the current u [73], which is available in the open source software FreeFem++ [43]. Again, the mesh to solve the forward problem (2.21) and the adjoint problem (2.27) remains fixed and uniform (200×200 Cartesian mesh) throughout the entire computation. For the optimization, we now use a truncated Quasi-Newton (BFGS) method with Wolfe step-size control.

In Fig. 6.4, we present the numerical results for the AEI method with and without mesh adaptation. The reconstruction on a uniform triangular mesh yields 4.86% relative L^2-error, whereas the reconstruction with adaptive FE strategy yields 4.17% relative L^2-error, yet with a ten times smaller number of vertices. Hence, we have not only reduced the error in the reconstruction with even crisper and smoother edges, but also greatly reduced the number of degrees of freedom in the control thereby dividing the overall execution time by 2.5. As shown in Fig. 6.4, the adapted mesh automatically concentrates the degrees of freedom along variations of u and is refined only where it is needed.

6.3 Adaptive vs. Harmonic Background $\mathbf{u_0}$

As shown in Section 5.2, the background state u_0 defined through (4.2) allows the AEI method to accomodate varying boundary data. Here we compare the accuracy in the reconstruction for two different choices for u_0:

- harmonic: u_0 is computed only once at the beginning of the AEI Algorithm by solving (4.2) with $\mu \equiv 1$
- adaptive: u_0 is recomputed at each frequency step by solving (4.2) with μ as in (4.3)

In Fig. 6.5, we compare the recovered u for the above two different background states u_0, both either with or without the adaptive mesh strategy from Section 6.2. Clearly, the adaptive background u_0 improves the accuracy of the reconstruction as the relative L^2-errors are approximately halved – see Table 6.2. Moreover, the artifacts near the exterior boundary, visible in the harmonic background approach, are now completely absent.

Hence the more accurate background state u_0 enables the AEI method to better approximate the remainder $u - u_0$ with the AE basis. Still, the adaptive strategy is also slightly more expensive since it adaptively recomputes u_0 for every frequency. These conclusions

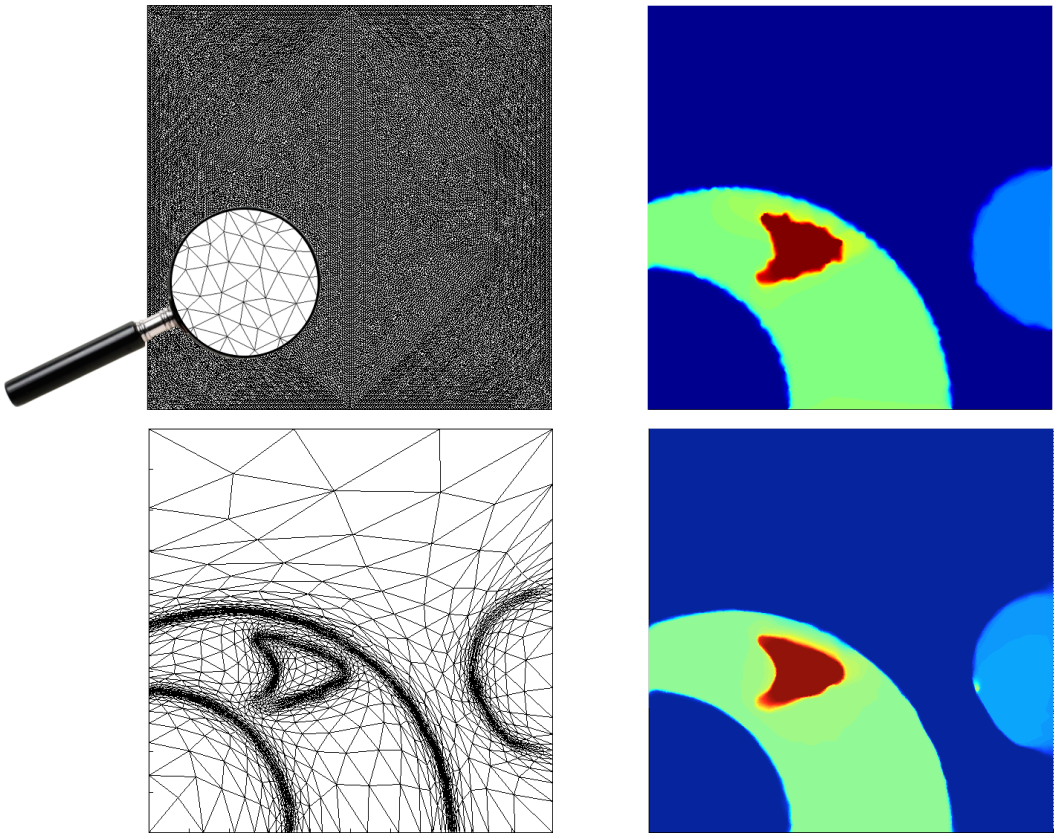

Figure 6.4: Adaptive FE discretization. Top: uniform triangular mesh with 30,534 vertices (left) and recovered u (right). Bottom: with mesh adaptation, final mesh for $\omega = 90$ with 2,783 vertices (left) and recovered u (right).

also appear to hold if an adaptive FE strategy is included in the solution of (4.2), (4.4), which in fact seems to have little impact on the overall accuracy here.

background state u_0	harmonic	adaptive
without mesh adaptation	10.73%	4.86%
with mesh adaptation	10.90%	4.17%

Table 6.2: Adaptive vs. harmonic background. Relative L^2-error in u.

6.4 Constant vs. Adaptive Dimension of V_K

Our AEI method does not require extra regularization term, such as standard Tikhonov regularization. Instead it restricts the optimization to the AE subspace $V_K = \text{span}\{u_0, \phi_1, \ldots, \phi_K\}$, which effectively acts as inherent regularization of the inverse problem. Here we focus on the choice of K to understand the regularization effect of the AEI method. To prevent any

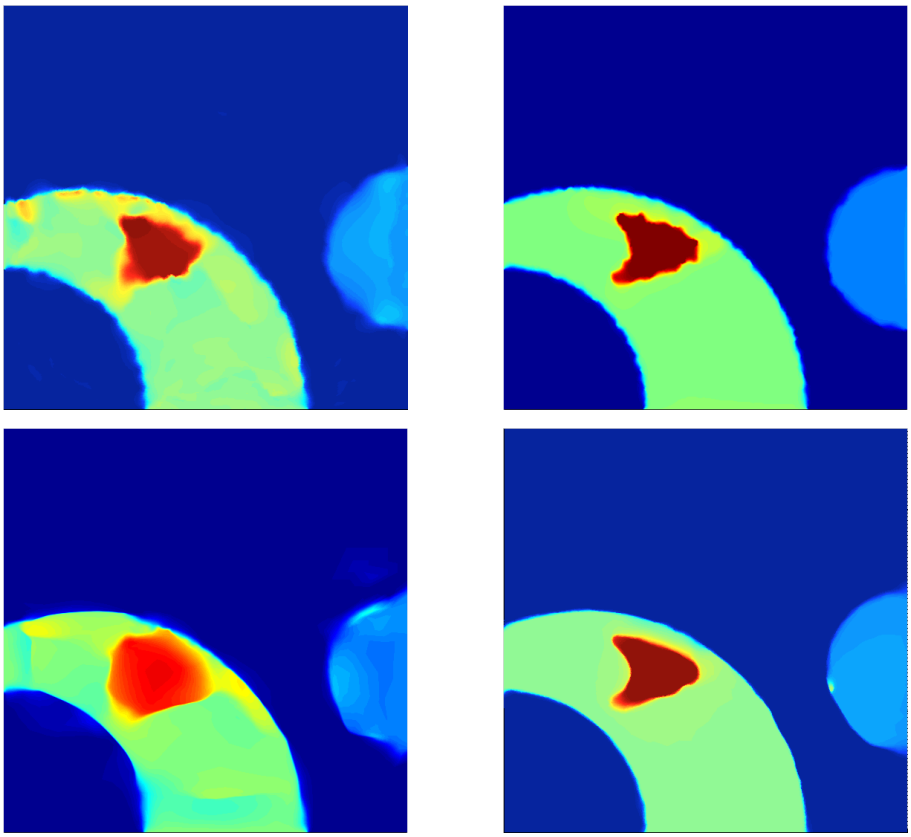

Figure 6.5: Adaptive vs. harmonic background. Top: without mesh adaptation, harmonic u_0 (left) and adaptive u_0 (right). Bottom: with mesh adaptation, harmonic u_0 (left) and adaptive u_0 (right).

intrinsic regularization from a very fine mesh or over-abundant data, we omit the receivers at the lower boundary of Ω (missing data) and use a coarser 200×200 Cartesian mesh.

First, we keep the number of eigenfunctions constant through the entire frequency stepping process. In Fig. 6.6, we present the numerical results for constant $K = 100$ using either a truncated full-Newton or Gauss-Newton method. For the full-Newton method, the regularization is not sufficient to reduce the perturbation at the top of the computational domain, although the reconstruction is more accurate than with the nodal basis. In contrast, for the Gauss-Newton method, the regularization is too strong and the method has difficulty reconstructing the kite. Similar conclusions hold for other constant values of K, which underlines the need for K to vary with the frequency ω.

Next, we let K vary linearly with ω, starting at a small number of eigenfunctions ($K = 32$) to reduce the risk of converging to a false local minimum. As ω increases during frequency continuation, we slowly increase K to capture smaller details of the scatterer. In Fig. 6.6, both reconstructions are now quite accurate with an L^2-error of 6.50% for the Newton and 6.72% for the Gauss-Newton method, respectively.

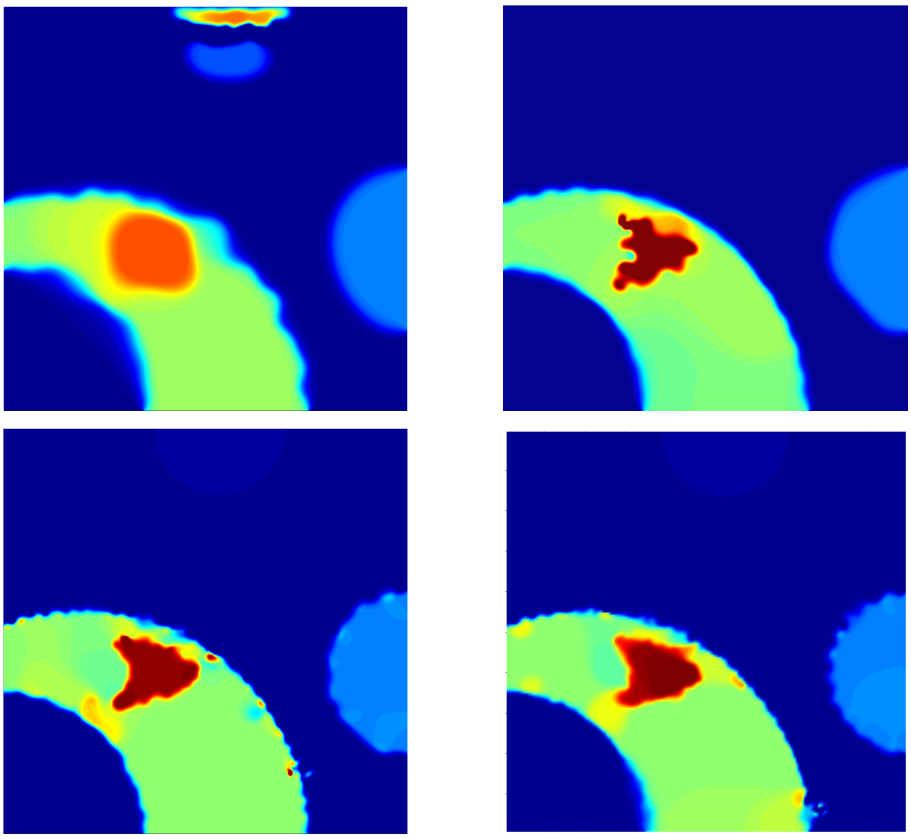

Figure 6.6: Constant vs. adaptive dimension of V_K. Top: for constant $K = 100$, truncated full-Newton method with relative L^2-error = 14.79% (left) and truncated Gauss-Newton method with relative L^2-error = 7.95% (right). Bottom: for linearly varying K, truncated full-Newton method with relative L^2-error = 6.50% (left) and truncated Gauss-Newton method with relative L^2-error = 6.72% (right).

Remark 4. The number of eigenfunctions K controls the regularization, similarly to the parameter α in a standard Tikhonov regularization term $\alpha R[u]$. For Tikhonov regularization, α is initially large but then gradually decreases to zero as the nonlinear iteration approaches the desired minimum. In contrast for our AEI approach, K is initially small but then gradually increases during optimization so that the AE basis includes a higher number of eigenfunctions for better accuracy.

6.5 Sparse AEI Method

The AEI method greatly reduces the number of parameter values by restricting the optimization to the subspace $V_K = \text{span}\{u_0, \phi_1, \ldots, \phi_K\}$ of much smaller dimension. Since the discretized version of the eigenvalue problem (4.4) leads to a sparse, symmetric and positive definite matrix, the first K eigenfunctions can be efficiently computed via a standard

Lanczos iteration [53]. Still, the storage of the first K eigenfunctions, which are global functions in Ω, may at first appear quite large, especially in three space dimensions.

In fact, much of the information contained in those eigenfunctions is highly localized in space and essentially negligible in most of Ω. Again we consider the example described at the beginning of Chapter 6. Next, for each eigenfunctions ϕ_m, we monitor all entries smaller than $\eta \cdot \|\phi_m\|_\infty$ in magnitude. In Fig. 6.7, we display the percentage of small entries in ϕ_m for each frequency averaged over all $m \leq K$; recall that K increases linearly with ω. As ω increases, the percentage of small entries also increases and quickly saturates above 90% at higher frequencies; hence, more than 90% of all entries are in fact negligible. To save memory space, we can therefore set to zero all those small values and simply replace the eigenfunctions ϕ_m by their sparse approximations.

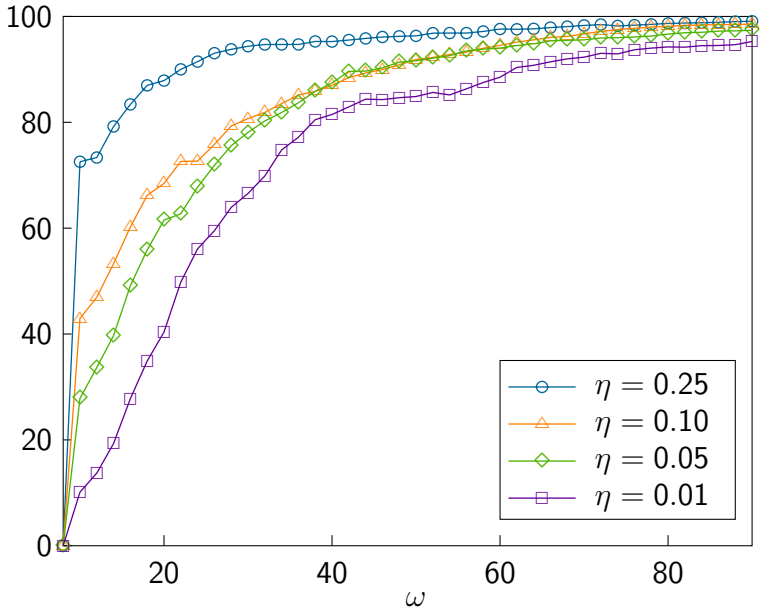

Figure 6.7: Sparse AEI method. Percentage of entries smaller than $\eta \cdot \|\phi_m\|_\infty$ in magnitude averaged over all $m \leq K$ at each frequency.

In Fig. 6.8, we show the resulting reconstruction with the sparse AEI approach, where all small entries in the eigenfunctions below $\eta\|\phi_m\|_\infty$ are set to zero for $\eta = 0.1,\ 0.05,$ or 0.01. Remarkably, the reconstruction hardly changes and, in fact, is even slightly better (L^2-errors below 4%), while saving more than 90% of memory.

6.6 Noisy Data

To illustrate the robustness of the AEI method with respect to noise, we now add multiplicative noise to the observations for each observation point x_i and frequency ω_j:

$$y^{obs,\delta}(x_i, \omega_j) = y^{obs}(x_i, \omega_j)(1 + \delta \cdot \xi_{i,j}),$$

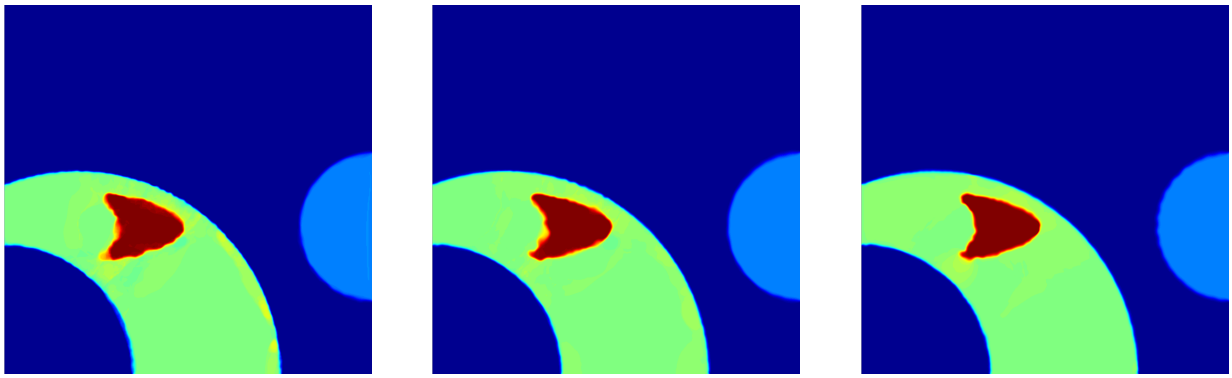

Figure 6.8: Sparse AEI method. Reconstruction of u with sparsified eigenfunctions. Left: with $\eta = 0.1$ (error = 3.83%). Center: $\eta = 0.05$ (error = 3.81%). Right: $\eta = 0.01$ (error = 3.90%)

where $\xi_{i,j}$ are i.i.d. Gaussian random variables with mean zero and variance equal to one. The level of noise is denoted by δ.

Again, we consider the parameter settings described at the beginning of this chapter, but use the adaptive finite element strategy from Section 6.2. In Fig. 6.9 we present the reconstruction results for two different levels of noise: for $\delta = 10\%$ (left) we obtain a relative L^2-error of 4.01% and for $\delta = 20\%$ (right) we obtain a relative L^2-error of 5.22%. As shown in Fig. 6.9, the AEI method is still able to reconstruct the profile without any added regularization and without artifacts due to noise. In fact, at the smaller noise level $\delta = 10\%$, the relative L^2-error is even slightly better than that without noise, 4.01% vs. 4.17%; each case, however, leads to (slightly) different meshes and eigenfunctions due to the adaptive finite element strategy.

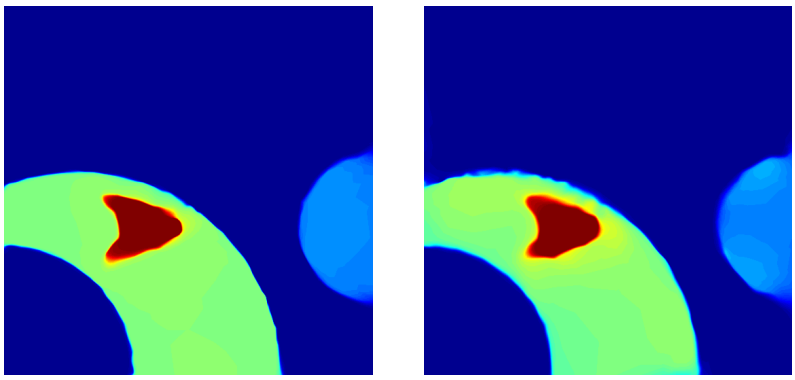

Figure 6.9: Noisy data for varying noise level δ. Left: $\delta = 10\%$ (L^2-error = 4.01%). Right: $\delta = 20\%$ (L^2-error = 5.22%).

6.7 Sample Average Approximation

As the number of sources N_s in (2.23) increases, the cost of computing N_s forward and adjoint solutions in (2.21), (2.27) may become prohibitive. To limit the computational cost without ignoring any of the available data, we consider the sample average approximation (SAA) approach from [38], which replaces the sources f_ℓ by N_r "super-shots", $N_r \ll N_s$,

$$F_j \;=\; \sum_{\ell=1}^{N_s} \xi_\ell^j \; f_\ell \,, \qquad j = 1, \ldots, N_r, \tag{6.1}$$

where the ξ_ℓ^j are i.i.d. random variables with zero mean and unit variance and corresponding observations

$$Y_j^{obs} \;=\; \sum_{\ell=1}^{N_s} \xi_\ell^j \; y_\ell^{obs} \,, \qquad j = 1, \ldots, N_r. \tag{6.2}$$

During frequency stepping, we choose for each ω a different realization of ξ_ℓ^j – here $\xi_\ell^j = \pm 1$ with probability 0.5.

Again, we consider the parameter settings described at the beginning of Chapter 6, but now with $N_s = 201$ Gaussian sources located at $(0.1, 0.8)$, $(0.11, 0.8)$, $\ldots$, $(0.89, 0.9)$, $(0.9, 0.8)$. The SAA approach [38] with only a single "super-shot", $N_r = 1$, yields the reconstruction shown in Fig. 6.10. For comparison, we also display the reconstruction without SAA with a single source located at $(0.5, 0.8)$. Although the computational cost of both approaches is identical, the SAA approach yields better accuracy, remarkably so, given that only a single ("super-shot") source is used for the approximation.

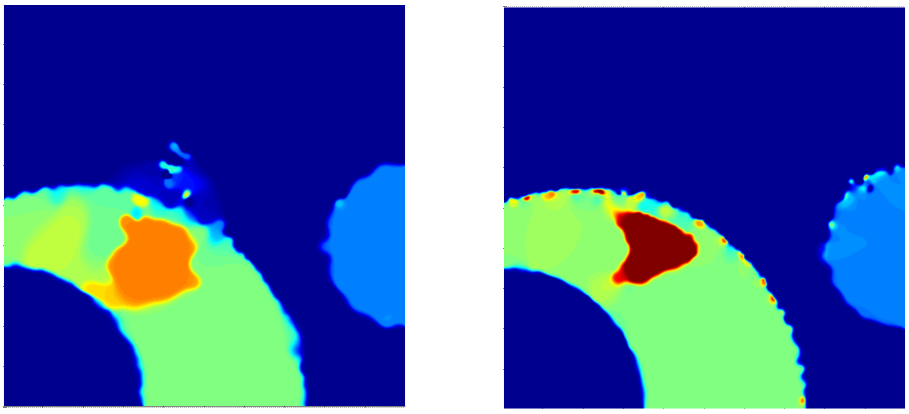

Figure 6.10: SAA approach. Reconstruction of u with a single source. Left: without SAA and with $N_s = 1$ (10.05% relative error). Right: with a single SAA "super-shot", i.e. $N_r = 1$, from $N_s = 201$ sources (5.79% relative error). Note that the computational effort is identical.

6.8 Computational Cost

We now compare the computational cost of the AEI method to the standard nodal basis with penalized TV-regularization, where the regularization parameters are adjusted to the problem. We reconstruct again the profile shown in Fig. 6.1 with the parameter settings described at the beginning of Chapter 6, using frequency stepping at $\omega = 8$, 10, 30, 50, 70, 90, on a 200×200 FD staggered grid. We stop the iterations at each frequency step when the relative residuum of the reduced-space gradient (2.24) is smaller than 10^{-1}. The CG algorithm is stopped when the CG iterations has converged according to (2.40) with $\eta_k = \eta_k^{(2)}$ in (2.41).

The bottleneck of the optimization arises from the number of linear systems (discretized forward operators) to solve, whose number grows with the number of outer G-N iterations, inner CG iterations and step-size determination. Hence, in Table 6.3 we compare the number of linear systems to solve, the number of CG iterations and the number of G-N iterations. Since the AEI requires the solution of an extra eigenspace problem for each frequency ω, the runtime of the eigenspace calculation is given as well. The simulations are computed on a desktop computer with an Intel Xeon E3-1270 V2 with 16 GB RAM, on a MATLAB R2015a program.

	$\omega = 8$		$\omega = 10$		$\omega = 30$	
Basis	Nodal	AE	Nodal	AE	Nodal	AE
# CG-Iter	120	41	61	13	1,421	25
# GN-Iter	3	15	4	4	9	7
# linear systems	6,642	2,124	3,501	621	77,436	1,188
Basis comp. runtime	-	1.91 s.	-	2.18 s.	-	8.77 s.
	$\omega = 50$		$\omega = 70$		$\omega = 90$	
Basis	Nodal	AE	Nodal	AE	Nodal	AE
# CG-Iter	61	20	82	12	93	12
# GN-Iter	4	6	4	4	5	4
# linear systems	3,474	963	4,617	585	5,238	585
Basis comp. runtime	-	16.53 s.	-	29.95 s.	-	47.11 s.

Table 6.3: Adaptive eigenspace vs. nodal basis with penalized TV-regularization. Comparison of runtime, number of CG and G-N iterations and basis computations.

In our computational cost test the AEI converges faster in terms of the forcing term (2.40) and needs less CG iterations per non-linear optimization step. In addition, at 85% of the G-N iterations for the AEI approach, the step size of the Armijo step-size control shows a full step, thus reducing extra computations of linear systems in order to find the step size. For the last example under the above described conditions, the AEI method solves 3 – 65 times less linear discretized Forward operators per frequency. We remark that a preconditioned G-N iteration might further reduce by about 35% the number of CG iterations, as it did in [59] for another application.

Chapter 7

AEI for the Limited Frequency Data

In some applications, especially in medical imaging, only limited frequency data are available. In fact, many medical imaging devices operate only at a single frequency. Hence, frequency stepping is not available. As described in Section 2.3, frequency stepping is an important tool, which helps preventing converging to a false local minimum. Thus, if we start the optimization with a too high frequency, it is likely that the optimization process will end in a false local minimum.

As we cannot take into account frequency stepping, we need to apply another strategy, which guides the algorithm towards a good minimum. We now introduce the single frequency adaptive eigenspace (SF-AEI) approach.

7.1 Single Frequency AEI (SF-AEI)

In Chapter 4, the AEI-algorithm updates the AE basis together with the number of eigenfunctions K, when the frequency ω increases. As explained in Chapter 5, the number of eigenfunctions K is linear in the frequency ω, i.e.

$$K = C\omega\,,$$

for some constant C. In the case of a single frequency, we propose to update the basis as follows: first, we optimize the problem with $\mu \equiv 1$, as we do not have information on the profile u. After the convergence of the optimization process, we actually adapt the basis to the AE basis with μ from (4.3) and optimize again. Instead of stopping after one adaptation for ω, as we do with frequency stepping, we propose to adapt repeatedly the AE basis from the previous optimal u for the same frequency ω.

We now have to deal with the choice of the number K of eigenfunctions. If we consider $K = C\omega$, where ω is our constant single frequency, which might be high, the regularization of the inverse problem might be inappropriate. One way to apply stronger regularization is to reduce the number of eigenfunctions used in the reconstruction.

7.2 Constant Background Numerical Experiments

As explained in Section 2.3, frequency stepping strategy is essential to the success of the optimization process. Hence, we want to draw your attention on the fact that when only a single frequency is available and in the absence of a good initial guess, we cannot recover very complicated media: if the single frequency is low, waves carry less information about the scatterer and the reconstruction may be unsatisfactory. If the available frequency is high, frequency stepping is needed to prevent the optimization process from converging to a false local minimum.

Hence, we start with a simplified version of the parameter u to illustrate the regularization process at a single frequency. The profile, shown in Fig. 7.1 contains a kite, centered at (0.4,0.4), and the value of u in the kite is two times higher than the value in the background. Next, we test several constant values of K for the optimization. The

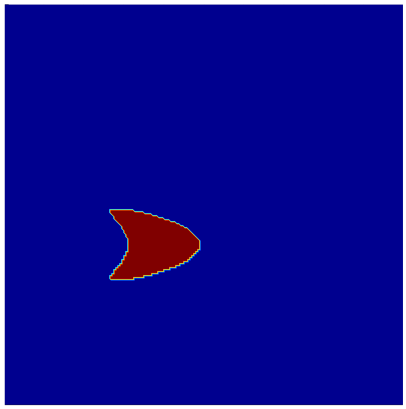

Figure 7.1: The true profile u to reconstruct with a single frequency.

settings for the numerical experiments using the AEI algorithm are as follows: Nine equispaced Gaussian sources are located along the upper boundary at $(0.1, 0.8), \ldots, (0.9, 0.8)$, whereas the receivers are located on the four lateral boundaries of $\Omega = (0, 1) \times (0, 1)$. We use second-order finite differences on a 200×200 mesh for the discretization of (2.2), (4.2)-(4.4). We optimize at the single frequency $\omega = 60$ and use the truncated GN-method for the optimization.

We decide to take a constant number of eigenfunctions K between $15 - 240$ and optimize the problem eight times, while adapting the basis between the optimizations. In Fig. 7.2, we present the numerical results for $K = 15$, $K = 30$ and $K = 240$ left to right respectively. Several tests using several values of constant K between 15 and 240 were performed, and the relative L^2-error for each value of K is plotted in Fig. 7.3. As shown in Fig. 7.2, if we take $K = 15$ the problem is over regularized and the relative L^2-error 11.06% is higher as with $K = 30$, 7.88%. If we use a too high constant number of eigenfunctions, e.g. $K = 240$, the problem is not well regularized and cannot prevent the pertubation in the computational domain.

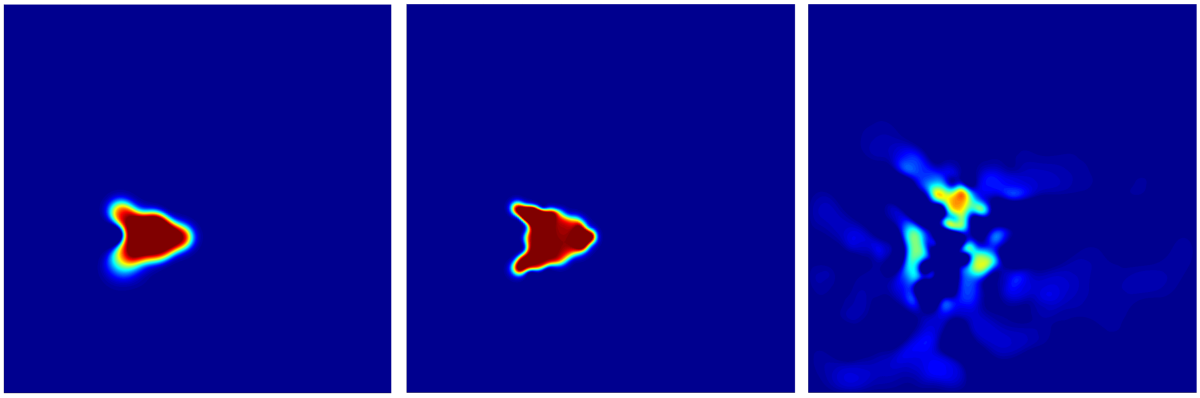

Figure 7.2: Constant dimension of V_K for constant frequency. Reconstruction with AEI of parameter u for $K = 15$ with relative L^2-error=11.06% (left), for $K = 30$ with relative L^2-error=7.88% (center) and for $K = 240$ with relative L^2-error=32.77% (right).

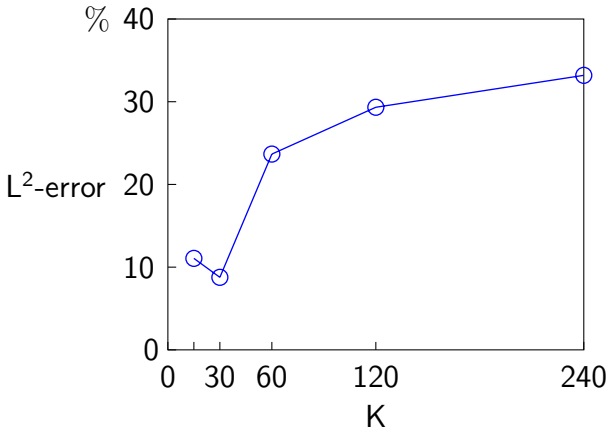

Figure 7.3: Relative L^2-error of the optimization w.r.t. the number of eigenfunctions K participating in the reconstruction of u.

Varying Number of Eigenfunctions

If we have more complicated media to reconstruct, the low constant number of eigenfunctions might not be able to capture all the details of the reconstructed medium u. Hence, we consider another way to regularize the problem by varying the number of eigenfunction K starting from a small value up to $C\omega$. The aim is to optimize the problem (2.23) N_K times for an increasing K. Once more we solve the inverse problem (2.23) first with $\mu \equiv 1$. Then we adapt the basis with μ from (4.3), increase K and solve the inverse problem again. We increase K and adapt the basis repeatedly, until $K = C\omega$, and take the previous optimal u as initial guess in each new optimization process.

We consider two strategies to increase K:

$$K = \left\lfloor \frac{C\omega}{N_K} \right\rfloor, \left\lfloor \frac{C\omega}{N_k - 1} \right\rfloor, \ldots, \left\lfloor \frac{C\omega}{2} \right\rfloor, C\omega\,, \tag{7.1}$$

or

$$K = \left\lfloor \frac{C\omega}{N_K} \right\rfloor, \left\lfloor \frac{2C\omega}{N_k} \right\rfloor, \ldots, \left\lfloor \frac{(N_k - 1)C\omega}{N_K} \right\rfloor, C\omega\,. \tag{7.2}$$

Considering again the example from Fig. 7.1, we optimize the problem eight times, i.e. $N_K = 8$. The values of K with respect to the optimization number is shown in Fig. 7.4.

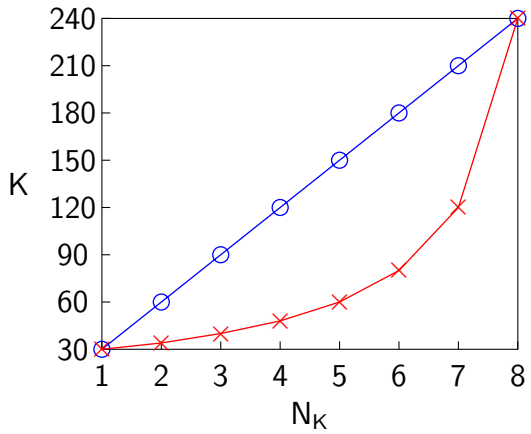

Figure 7.4: Number of eigenfunctions K participating in the reconstruction of u w.r.t. the optimization number for $N_K = 8$.

Before we conclude this chapter with numerical examples we state the single frequency AEI algorithm.

7.3 Single Frequency AEI Algorithm

We consider again the reduced-space functional $\mathcal{F}$ from (2.23) and the corresponding gradient $\nabla\mathcal{F}$ from (2.24) and Hessian (or approximation of it) H from (2.30).
To solve the the inverse problem at a single frequency, we introduce the following algorithm:

SF-AEI Algorithm.
Input: initial guess $u = 1$, observations y_ℓ^{obs} for one frequency. Output: u^*.

1. Choose $N_K \geq 2$
2. compute $\{\phi_m\}_{m=1}^K$ from (4.4) and u_0 from (4.2) with $\mu \equiv 1$, where $K = \left\lfloor \frac{C\omega}{N_K} \right\rfloor$
3. Expand $u(x) = u_0(x) \ + \ \sum_{m=1}^K \beta_m \phi_m(x)$
4. For $j = 1, 2, \ldots, N_K$
   (a) Compute $\mathcal{F}(u)$ and $\nabla\mathcal{F}(u)$, set H
   (b) STOP: if $\|\nabla\mathcal{F}(u)\| \leq$ Tol
      i. Solve $Hp = -\nabla\mathcal{F}(u)$ using inexact Newton
      ii. Determine step size α_k and set $u := u + \alpha_k p$
      iii. Update $\mathcal{F}(u)$, $\nabla\mathcal{F}(u)$ and H
   (c) Set μ from (4.3) with ∇u
   (d) Update K by the chosen strategy from (7.1) or (7.2)
   (e) Compute $\{\phi_m\}_{m=1}^K$ from (4.4) and u_0 from (4.2)
   (f) Expand $u(x) = u_0(x) \ + \ \sum_{m=1}^K \beta_m \phi_m(x)$
5. $u^* = u$

7.4 Numerical Experiments

We now illustrate the regularization ability of the SF-AEI, starting directly at a high frequency $\omega = 60$. We use the SF-AEI method with increasing K and $N_K = 8$. Fig. 7.5 illustrates the results using strategies (7.1) and (7.2) in the SF-AEI. Both reconstructions come with very similar L^2-errors (about 8.6%) and are comparable to the reconstruction with a constant $K = 30$ for eight iterations. However, we shall now illustrate that the choice of K, especially when K varies, can capture more details in the case of a complicated medium.

More precisely, we want to illustrate the ability of the SF-AEI to reconstruct a complicated medium with a good initial guess and to compare it to the fixed number of eigenfunctions. We reconstruct our original profile u shown from Chapters 2, 3 and 6 illustrated once more in Fig. 7.6.

Since this profile is more complicated and the considered frequency $\omega = 60$ is too high for a starting frequency, we use the initial guess shown in Fig 7.6. The settings for the

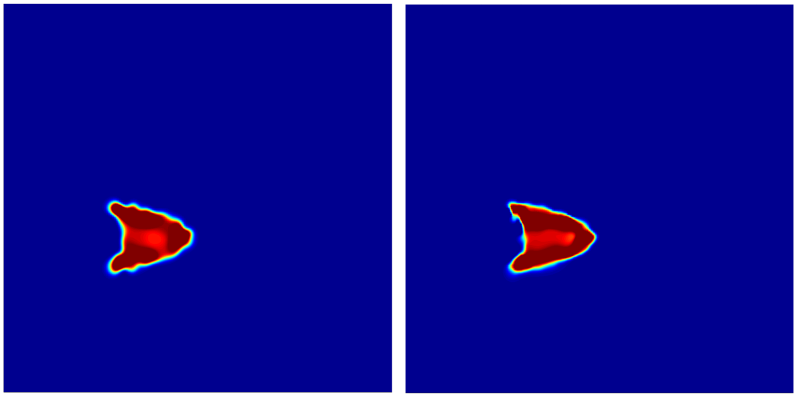

Figure 7.5: SF-AEI with $N_K = 8$ and varying K in two strategies: strategy (7.1) with relative L^2-error=8.67% (left), strategy (7.2) with relative L^2-error=8.68% (right).

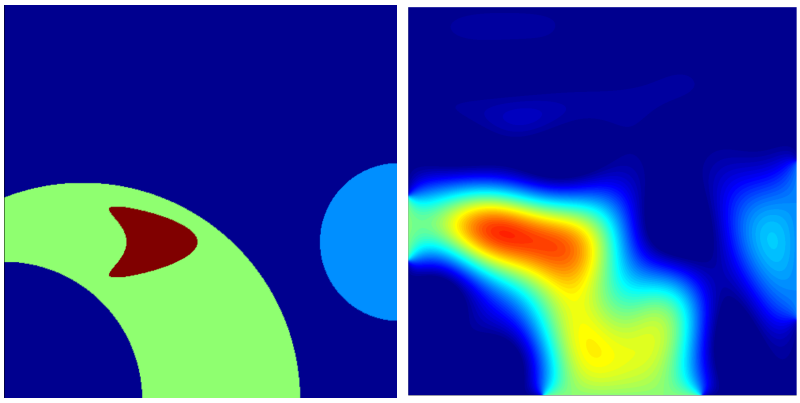

Figure 7.6: The true profile u (left), together with the initial guess for the reconstruction (right).

numerical experiments using the SF-AEI algorithm are as in the previous experiments. Since we start with a good initial guess, we can consider a smaller number of optimizations $N_K = 4$.

The results for the SF-AEI are shown in Fig. 7.7, for the strategy (7.1) and for the strategy (7.2) and for the best choice of constant number of eigenfunctions $K = 60$. We observe that the strategy (7.1) gives the better results whereas the strategy (7.2) does not regularize strong enough during the optimization process. In the case of the constant number of eigenfunction $K = 60$, the solution is highly regularized and it is not able to reconstruct the kite shaped feature in the medium.

In the numerical experiments of the previous chapters, we used K in linear dependence of ω. In the single frequency, we opt for a choice, which increases K more slowly. This can be explained as follows: when only one high frequency is available, it seems that the number of local minima is raising and a stronger regularization must be applied. For more complicated media, a constant number of K might be insufficient. Hence, when only one

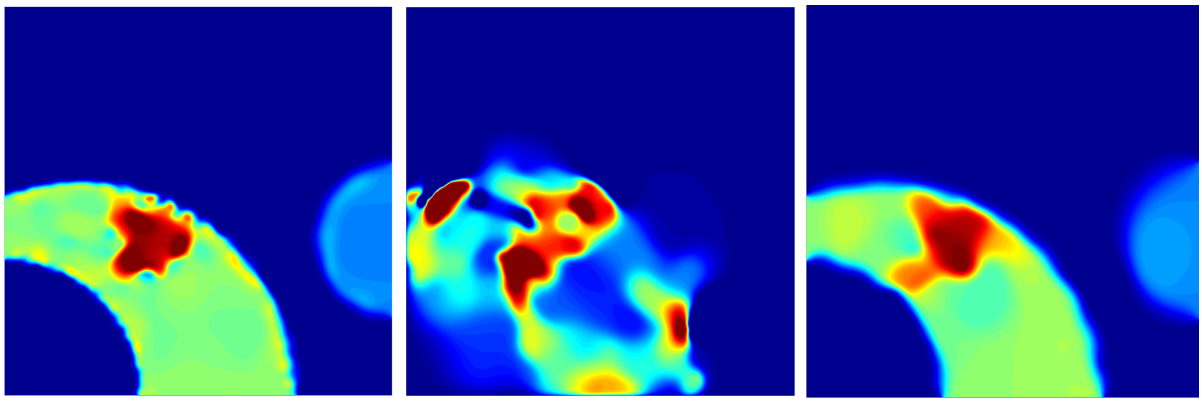

Figure 7.7: Reconstruction of parameter u with SF-AEI: with increasing K from (7.1) with L^2-error=6.40% (left), with increasing K from (7.2) with L^2-error=30.18% (center) and for constant $K = 60$ with L^2-error=8.78% (right).

frequency is available, the SF-AEI with the strategy (7.1) of varying K seems to be the most suitable choice.

As we have seen in Section 7.2, the AEI is able to deal with restricted frequency data. In the last numerical experiment, whose result is shown on the left of Fig. 7.7, the SF-AEI was able to reconstruct the medium using very limited frequency data. Actually, the initial guess for the previous example, shown in Fig. 7.6 is the reconstruction of the AEI algorithm at $\omega = 8$ and $\mu \equiv 1$. That shows that using only two frequencies, $\omega = 8$ and $\omega = 60$, we can reach an acceptable reconstruction with relative L^2-error of 6.40%.

Chapter 8

AE of Gradients of Penalty Functionals

In Chapters 4-7, we used the adaptive eigenspace (AE) derived from the penalized TV-regularization. To build the AE, we considered the regularization functional $\mathcal{R}_{TV}$ (3.8) and compute its gradient (3.14). The resulting AE problem reads (see Remark 1)

$$\begin{cases} -\nabla \cdot (\mu(x)\nabla\phi_m(x)) &= \lambda_m\phi_m(x), & \forall\, x \in \Omega, \\ \phi_m(x) &= 0, & \forall\, x \in \Gamma, \end{cases} \tag{8.1}$$

where μ is given from the gradient of the TV-regularization by

$$\mu(x) \;=\; \frac{1}{\sqrt{|\nabla u(x)|^2 + \varepsilon^2}}\,, \qquad \forall\, x \in \Omega,\ \varepsilon > 0\,. \tag{8.2}$$

8.1 AE from other Penalty Functionals

In this chapter, we would like to derive new adaptive eigenspaces from other penalty functionals. In Chapter 3 we introduced various Tikhonov regularization functionals (3.3)-(3.8), whose gradients are given by (3.10)-(3.14). Each gradient of a regularization functional can deliver an adaptive eigenspace. In der following, we replace the gradient of the TV-regularization by a gradient of another regularization functional and get an AE with new properties: the resulting eigenspace, together with its eigenfunctions, has similar properties as the corresponding standard added Tikhonov regularization.

AE from the H^1-Penalty Term

The first example is the H^1-regularization penalty functional $\mathcal{R}_{\nabla u}$ (3.4). The aim of this functional is to penalize non-smooth solutions. The gradient of $\mathcal{R}_{\nabla u}$ is, as in (3.11):

$$\nabla_u \mathcal{R}_{\nabla u}(u) = -\Delta u\,,$$

and the corresponding AE solves

$$\begin{cases} -\Delta\phi_m(x) &= \lambda_m\phi_m(x), & \forall\, x \in \Omega, \\ \phi_m(x) &= 0, & \forall\, x \in \Gamma. \end{cases} \tag{8.3}$$

The eigenspace (8.1) and the Laplacian eigenspace (8.3) were already discussed in Chapter 5. There, we have shown that the AEI with the eigenfunctions basis of the TV-regularization allows discontinuities in the reconstructed profiles. On the contrary, the eigenspace of the H^1-regularization (the Laplacian eigenspace) is only suitable for smooth profiles. This indicates that each eigenspace inherits its properties from the corresponding regularization functional.

AE from the Gaussian Penalty Term

We now consider Gaussian-regularization [21]. The gradient corresponding to the Gauss-regularization functional (3.6) is, as already given in (3.12):

$$\nabla_u \mathcal{R}_{Gauss}(u) = -\nabla \cdot \left(\frac{\nabla u}{\sigma^2 \, \exp\left(|\nabla u|^2/\sigma^2\right)} \right), \quad \sigma > 0\,.$$

We derive the associated eigenspace problem in a similar way as for TV-regularization and get (8.1) with

$$\mu(x) = \frac{1}{\sigma^2 \, \exp\left(|\nabla u|^2/\sigma^2\right)}, \qquad \forall\, x \in \Omega,\ \sigma > 0\,. \tag{8.4}$$

Note that μ here also depends on the control variable u as it does in AE of the gradient of penalized TV-regularization (8.2). As we have seen for the AE of the penalized TV-regularization, the weight function μ which depends on u adapts the eigenspace to the reconstruction at each frequency step and yields relevant eigenfunctions.

AE from the Lorentzian Penalty Term

Next, we consider the Lorentzian-regularization [21], which is used to penalize strong variation in the solution and contains an extra parameter to allow discontinuities. We calculate the gradient corresponding to the Lorentzian-regularization functional (3.7), as computed in (3.13):

$$\nabla_u \mathcal{R}_{Lorentz}(u) = -\nabla \cdot \left(\frac{\gamma \nabla u}{\left(1 + \gamma|\nabla u|^2\right)^2} \right), \quad \gamma > 0\,.$$

Again, we derive the associated eigenspace problem as before and get (8.1) with

$$\mu(x) = \frac{\gamma}{\left(1 + \gamma|\nabla u|^2\right)^2}, \qquad \forall\, x \in \Omega,\ \gamma > 0\,. \tag{8.5}$$

Note that μ also depends on the control variable u as for the penalized TV-regularization and the Gaussian AE.

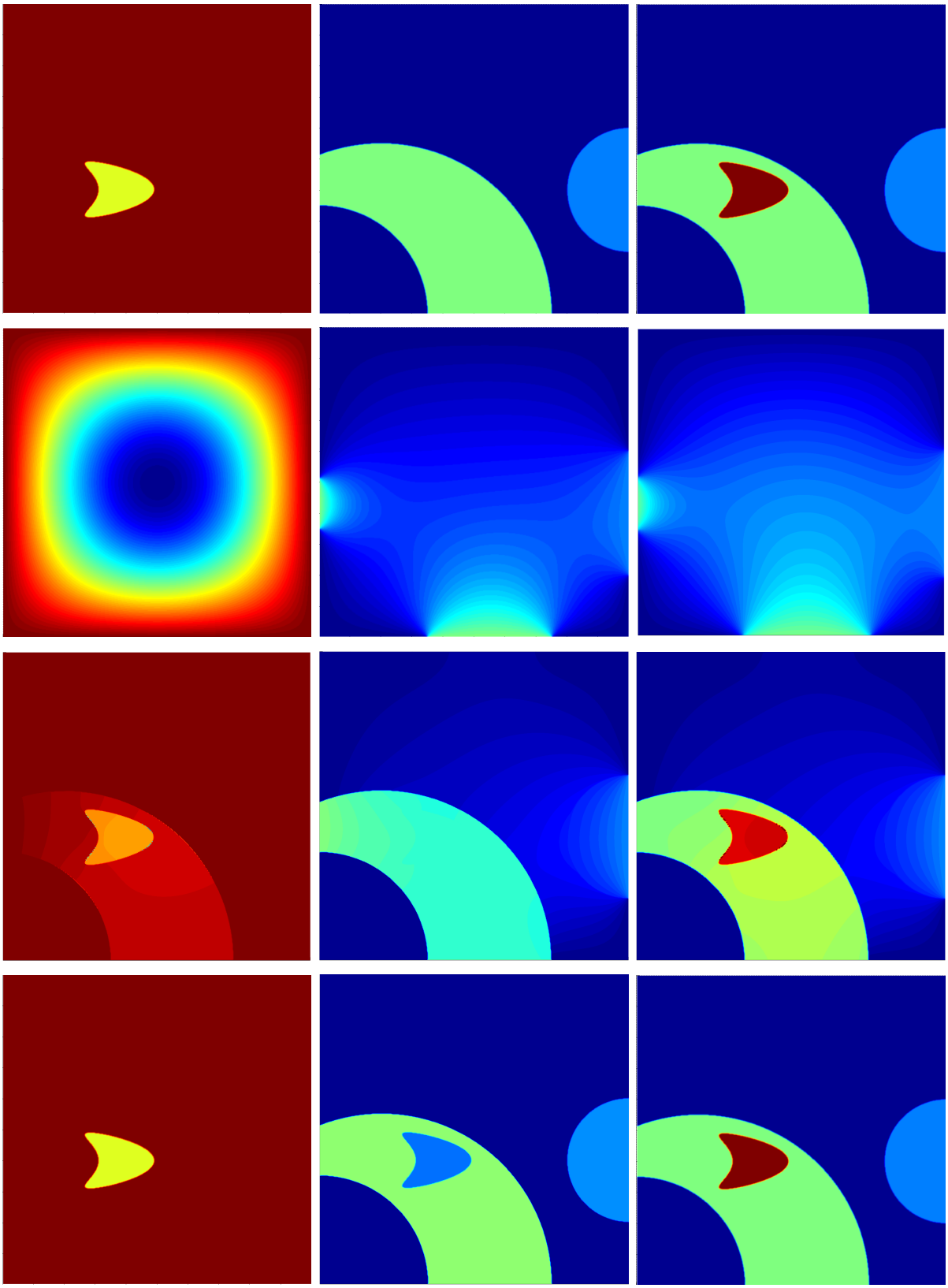

Figure 8.1: Adaptive eigenspaces from Tikhonov regularization. From left to right: the first eigenfunction ϕ_1, the boundary contribution u_0, the expansion of u using ϕ_1 and u_0 for the respective AE. From top to bottom: AE of the TV-regularization, AE of the H^1-regularization, AE of the Gaussian-regularization and AE of the Lorentzian regularization.

About the Background-Problem and Illustration

Using μ, we can also compute u_0 as described in Chapter 4 by solving

$$\begin{cases} -\nabla \cdot (\mu(x)\nabla u_0(x)) &= 0, & \forall\, x \in \Omega, \\ u_0(x) &= c^2(x), & \forall\, x \in \Gamma\,, \end{cases} \tag{8.6}$$

where c is the true velocity on the boundary. Actually, all settings of the adaptive eigenspaces described in Chapter 4 can also be applied here, including for u_0 and for the number K of considered eigenfunctions.

We have seen how to define a new AE from an existing regularization. We can actually choose a relevant AE depending on the properties of the profile to reconstruct. from Fig. 2.3 and Chapter 6: the TV-based AE with μ from (8.2), the Laplacian AE (8.3), the Gauss-based AE with μ from (8.4) and the Lorentzian-based AE with μ from (8.5).

Of course, there are many other possible regularization functionals that may be considered. As shown for the AE of the penalized TV regularization, the H^1-penalty regularization, the Gauss-regularization and the Lorentzian-regularization, taking each Tikhonov regularization functional and deriving an AE from it gives many new adaptive eigenspaces. However, Tikhonov regularization functionals may also produce irrelevant AE as well, for example, the AE for the L^2-regularization penalty gives the identity operator:

$$\begin{cases} \phi_m(x) &= \lambda_m \phi_m(x), & \forall\, x \in \Omega, \\ \phi_m(x) &= 0, & \forall\, x \in \Gamma\,. \end{cases} \tag{8.7}$$

8.2 Numerical Experiments

We now apply the various regularizations to the more realistic profile u, shown in Fig. 8.2. The profile mimics a layered material with regions of different wave speed and is based on a well-known geophysical model: the Marmousi model.

The settings in the numerical experiments are as follows: we use six realizations of 201 samples, which yield six sources ("super-shots") located then at $(0.1, 0.9)$, ..., $(0.9, 0.9)$. The receivers are located on the north, east and west lateral boundaries of the computational domain $\Omega = (0,1) \times (0,1)$. We use a 300×300 FD mesh and frequency stepping, starting at the lowest frequency $\omega = 8$ with a step of 4 up to $\omega = 200$. The initial guess is the four-layered profile shown in Fig. 8.3.

We now reconstruct the profile u under the settings mentioned above with the bases from (8.1) using four different choices of μ: from (8.2), $\mu \equiv 1$ (the Laplace eigenfunctions (8.3)), from (8.4) and from (8.5). In Fig. 8.4, we see that in all four reconstructions we get a very smooth solution. Hence, the adaptive eigenspaces regularize the solution too strongly. This illustrates the need for an AE which can be suitable for seismic imaging.

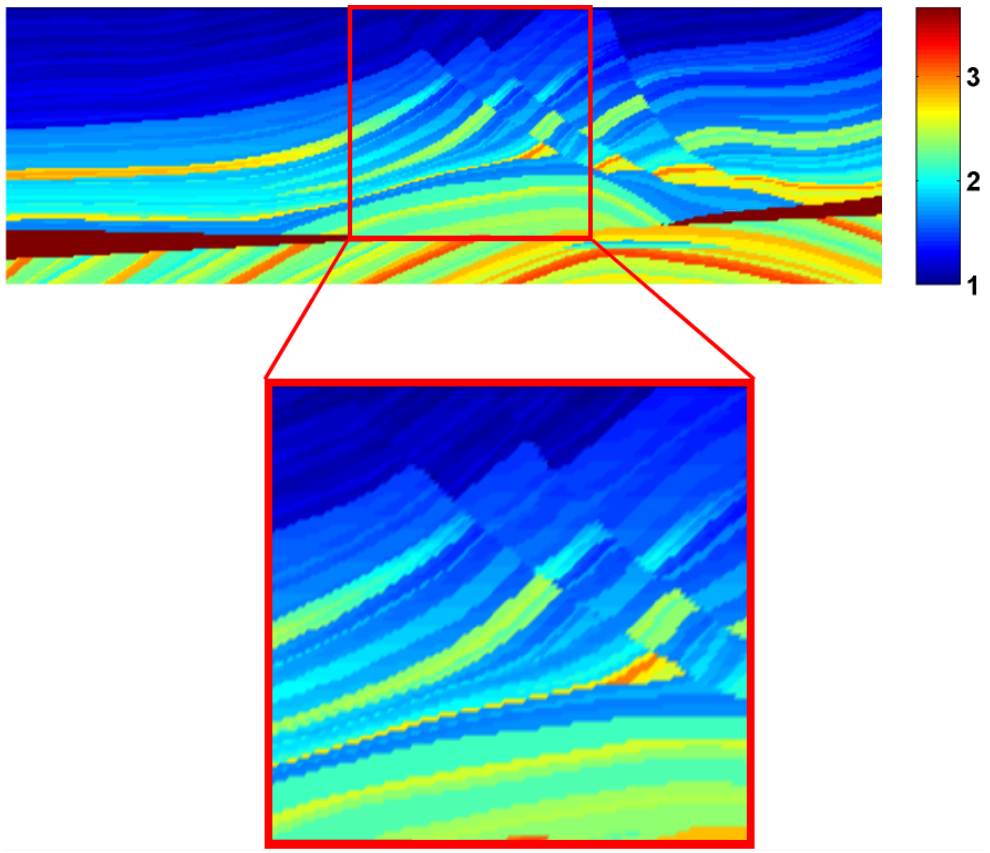

Figure 8.2: A Marmousi-based profile: A cut from the Marmousi profile.

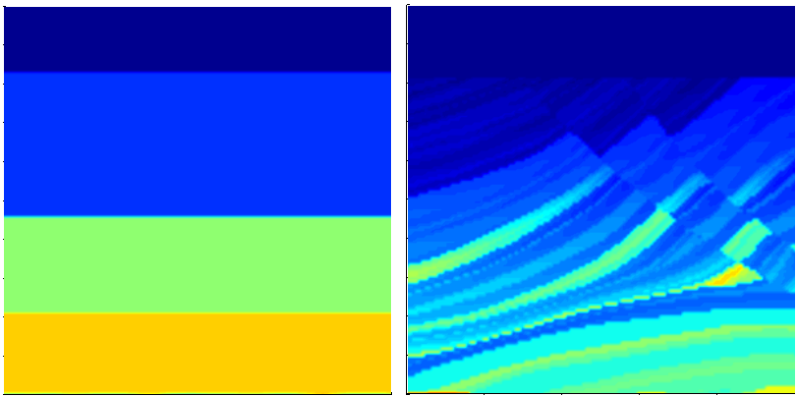

Figure 8.3: The Marmousi-based profile: the initial guess (left) and the true profile (right).

AE for Seismic Imaging

In seismic models, the medium profiles are layered materials, which are typically represented by a smoothly varying parameter u in the horizontal direction, but non-smoothly in the vertical one. Since we understand how to build an AE adapted to the properties of

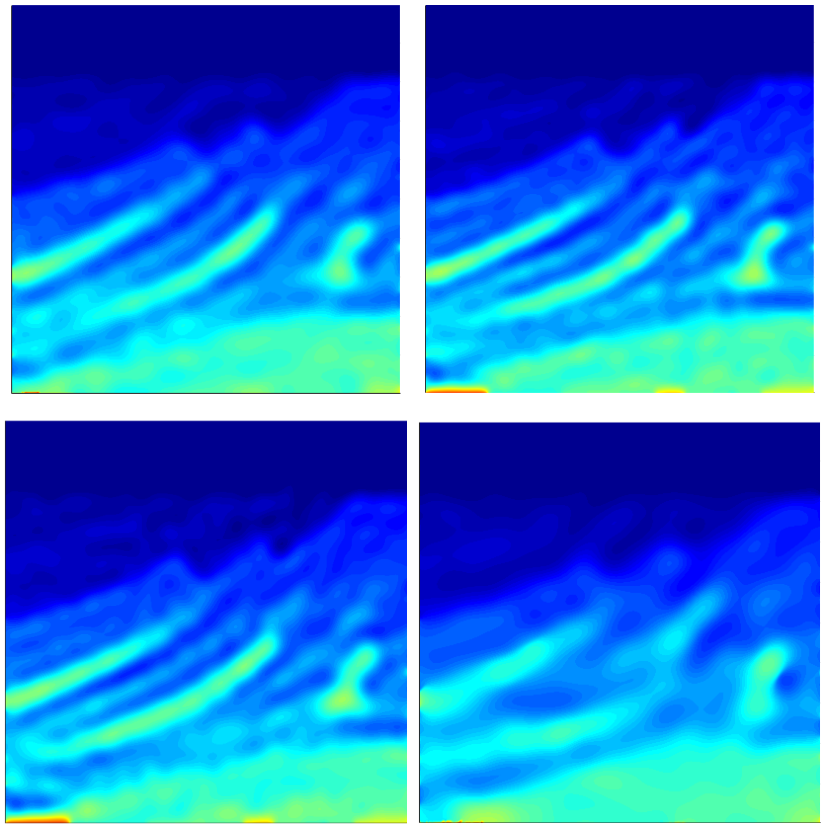

Figure 8.4: AEI using four different adaptive eigenspaces for the Marmousi-based profile: Top: TV-regularization AE with relative L^2 error = 6.42% (left), H^1-regularization AE with relative L^2 error = 6.43% (right). Bottom: Gaussian-regularization AE with relative L^2 error = 7.12% (left), Lorentzian-regularization AE with relative L^2 error = 8.83%.

the profile, we can now construct an AE for typical profiles from seismic imaging. For this anisotropic profile, we consider a penalized TV-regularization in the vertical direction and Sobolev H^1-penalty functional in the horizontal one:

$$\mathcal{R}_{d\nabla u}(u) = \frac{1}{2}\int_\Omega \left(\frac{\partial u}{\partial x_1}\right)^2 dx\,, \quad \mathcal{R}_{dTV}(u) = \frac{1}{2}\int_\Omega \sqrt{\left|\frac{\partial u}{\partial x_2}\right|^2 + \varepsilon^2}\, dx\,. \tag{8.8}$$

As for the former regularizations in this chapter, we build a new AE using the gradient of the anisotropic functional $\mathcal{R}_{dTV}$ in (8.8)

$$\nabla_u \mathcal{R}_{dTV}(u) = -\frac{\partial}{\partial x_2}\left(\frac{1}{\sqrt{\left|\frac{\partial u}{\partial x_2}\right|^2 + \varepsilon^2}}\frac{\partial u}{\partial x_2}\right)\,. \tag{8.9}$$

Again, we derive the associated eigenspace problem

$$\begin{cases} -\dfrac{\partial}{\partial x_2}\left(\dfrac{1}{\sqrt{\left|\frac{\partial u(x)}{\partial x_2}\right|^2+\varepsilon^2}}\dfrac{\partial \phi_m(x)}{\partial x_2}\right) &= \lambda_m \phi_m(x), & \forall\, x \in \Omega, \\ \phi_m(x) &= 0, & \forall\, x \in \Gamma\,. \end{cases} \tag{8.10}$$

Let ϕ_m be an eigenfunction of (8.10) and $\varphi = \varphi(x_1)$ be any function depending only on x_1 and $\varphi(x_1) = 0$, $\forall\, x \in \Gamma$. Then $\varphi\phi_m$ is also an eigenfunction of (8.10) and thus we have no control on the eigenfunctions in the direction of x_1, however, they should respect the functional $\mathcal{R}_{d\nabla u}$ (8.8) in this direction. Hence, we replace the partial derivative $\frac{\partial}{\partial x_2}$ by a standard full divergence and gradient operators and get the analogue to (8.1):

$$\begin{cases} -\nabla\cdot\left(\dfrac{1}{\sqrt{\left|\frac{\partial u(x)}{\partial x_2}\right|^2+\varepsilon^2}}\nabla \phi_m(x)\right) &= \lambda_m \phi_m(x), & \forall\, x \in \Omega, \\ \phi_m(x) &= 0, & \forall\, x \in \Gamma\,. \end{cases} \tag{8.11}$$

Note that the resulting eigenspace (8.11) takes into account both functionals in (8.8): if $\partial u/\partial x_2 = 0$ on a subregion inside Ω, then $\mu = 1/\varepsilon$ on that subregion. Since μ is constant in that subregion the eigenspace problem can be written as

$$-\Delta\phi_m(x) = (\varepsilon\lambda_m)\phi_m(x)\,.$$

This yields locally Laplacian eigenfunctions, which automatically corresponds to the functional $\mathcal{R}_{d\nabla u}$ in (8.8).

Next, we repeat the previous experiment but now use the AE (8.11). The solution is smooth only in horizontal direction and the discontinuity in the vertical direction is well reconstructed. The resulted AE is thus adapted to the geophysical profile and the results are shown in Fig. 8.5.

AE as Initial Estimation

For the reconstruction of complex media such as the Marmousi-based model problem, one typically starts from a relatively close initial guess, for example a smoother representation of the target profile. In contrast, our initial guess uses almost no a priori information about the profile, yet we get an acceptable reconstruction. If we want to get an even more detailed reconstruction we can use the reconstruction of the AEI as initial estimate for a standard nodal basis approach. The main benefit of this approach is that the AEI can recover the main structure of the Marmousi based model, starting from a poor initial estimation. The nodal basis is then able to recover the finest details, once the intermediate reconstruction

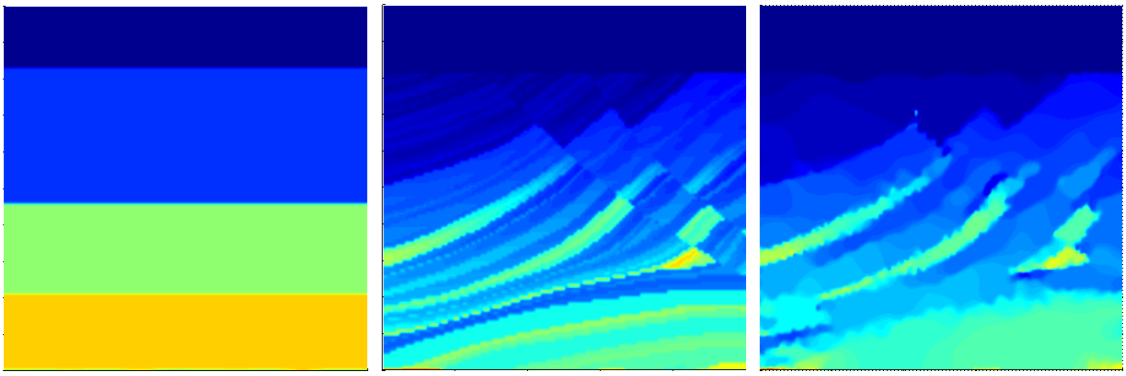

Figure 8.5: Reconstruction of the Marmousi-based profile using AEI with μ as described in (8.11). The initial guess (left), the true profile (center) and the reconstruction (right).

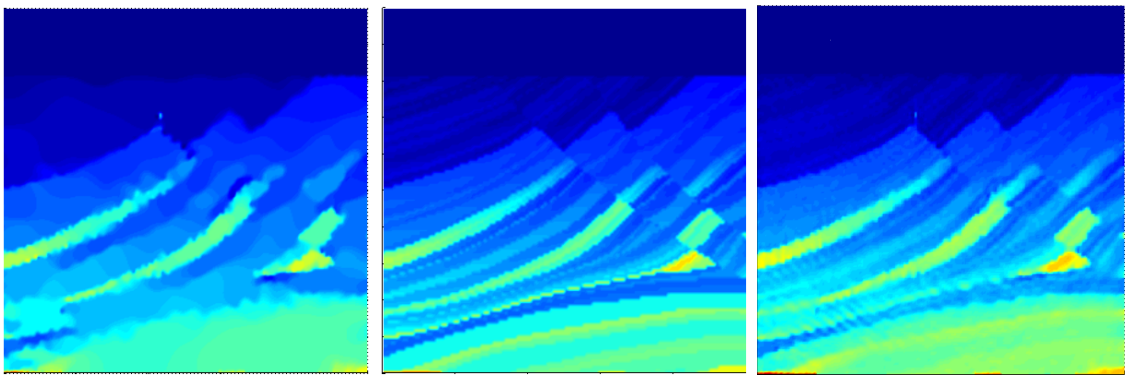

Figure 8.6: Reconstruction of the Marmousi-based profile using AEI with μ as described in (8.11) as initial value for the optimization with the nodal basis. The initial starting value obtained with the AE (left), the true profile (center) and the final reconstruction (right).

is close enough. By combining both approaches, we achieve an excellent reconstruction as shown in Fig. 8.6, even in the absence of a good initial guess.

The different bases will play an important role in part III, where several different regularization are considered. There, we will not only see how the AEI can regularize the optimization, but also how to avoid unphysical artifacts.

Part III

Multiparameter Inverse Helmholtz Problem

Chapter 9

AEI with Multi-Parameter

Multi-parameter inverse problems have been an area of research since 1984 [80]. In the last couple of years this area of study become very popular, especially in geophysics [66, 72, 90, 30]. The two distinct parameters, density and bulk-modulus, are crucial in oil and gas exploration. Hence, large number of recent geophysical papers discuss this problem. Several issues make multi-parameter inverse problems very difficult:

1. **Cross-talk** - optimizing over several parameters appearing in the same equation often results in so-called cross-talk, i.e. the parameters create undesired artifacts by influencing each other [66]. The gradients of the different parameters are coupled and a small change in one parameter can cause spurious effects and regularization problems in the other parameters [67].

2. **Ill-posedness** - another challenge of the multi-parameter inverse problem is its ill-posedness [72], which is even more extreme than the one-parameter problems.

In the papers mentioned above, some methods are introduced to tackle those problems: in [72] the full-Newton method for the multi-parameter and the data driven strategy are presented. The data driven strategy defines the expected decrease in the misfit function and together with a separate Tikhonov regularization for each parameter, reduces the cross-talk between parameters. In [67], a sparse version of the Gauss-Newton Hessian is used for the penalty formulation and in [30] an a priori information on the similarity of the structures of the parameter is used. In [90], random sequential sources for multi-parameter inversion with multiple right hand-sides are proposed.

Here, we propose the AEI to solve the multi-parameter inverse problem, based on our former study for one-parameter AE. In Chapter 4, we used the eigenspace of the TV-regularization gradient as an adaptive eigenspace and we were thus able to regularize the problem through the slowly increasing number of eigenfunctions. In Chapter 8 we extended the AE approach to different regularization gradients. We shall now extend the AE approach to the multi-parameter inverse problem by using for each parameter its proper basis, which may be adapted to the parameter properties.

Before we further discuss the AEI approach for multi-parameter inverse problems, we shall define the particular problem considered.

9.1 The Multi-Parameter Inverse Helmholtz Problem

The propagation of waves through a medium in time can be described by the acoustic wave equation

$$\frac{1}{K(x)}\, y_{tt}(x,t) - \nabla \cdot \left(\frac{1}{\rho(x)} \nabla y(x,t) \right) = f(x,t)\,, \tag{9.1}$$

where $K(x) > 0$ represent the bulk modulus, $\rho(x) > 0$ the density, $f(x,t)$ the source of the acoustic waves and $y(x,t)$ the pressure variation [80].

In the literature we often find the acoustic wave equation in the form of (1.1). However, since we are not always confronted with a piecewise constant density or bulk modulus, we now would like to consider equation (9.1).

As for the one-parameter (velocity) wave equation (1.1), we use Fourier transform of the time variable and write

$$y(x,t) = \hat{y}(x)e^{-i\omega t} \quad \text{and} \quad f(x,t) = \hat{f}(x)e^{-i\omega t}\,, \tag{9.2}$$

where ω is again the time frequency. We hence get the following two-parameter Helmholtz equation:

$$-\frac{\omega^2}{K(x)}\hat{y}(x) - \nabla \cdot \left(\frac{1}{\rho(x)} \nabla \hat{y}(x) \right) = \hat{f}(x)\,. \tag{9.3}$$

For the sake of simplicity, we rewrite (9.3) without the Fourier transform notation and after variable transformation as

$$-\omega^2 v(x) y(x) - \nabla \cdot (u(x) \nabla y(x)) = f(x)\,, \tag{9.4}$$

where $v = 1/K > 0$ and $u = 1/\rho > 0$. As in the one-parameter inverse problem of Chapter 2, we impose a Sommerfeld boundary condition to model the unbounded medium

$$\frac{\partial y}{\partial n} \; - \; iky = 0, \quad \text{on} \;\; \Gamma = \partial\Omega\,. \tag{9.5}$$

Here $k(x) = \omega/c(x)$ is the wavenumber at the time frequency ω and the velocity of the acoustic waves $c(x)$ is given by

$$c(x) = \sqrt{\frac{u(x)}{v(x)}}\,. \tag{9.6}$$

Again, we use FD to discretize (9.4). For a given u and v the forward operator is

$$A(u,v)y = f\,.$$

In the multi-parameter AEI, we again opt for the reduced-space approach. The optimization problem then reads

$$\begin{gathered} \text{minimize } \mathcal{F}(u,v)\,, u \in U\,, v \in V \text{ where} \\ \mathcal{F}(u,v) = \tfrac{1}{2}\sum_{\ell=1}^{N_s} \left\| PA(u,v)^{-1} f_\ell - \hat{y}_\ell \right\|^2 + \alpha_1 \mathcal{R}_1(u) + \alpha_2 \mathcal{R}_2(v)\,. \end{gathered} \tag{9.7}$$

Since the multi-parameter inverse problem is very ill-posed, the regularization functionals $\mathcal{R}_1(u)$, $\mathcal{R}_2(v)$ are added to the misfit. The functionals $\mathcal{R}_1(u)$, $\mathcal{R}_2(v)$ are usually chosen as Tikhonov regularization functionals, see Chapter 3, for example from (3.4), (3.6), (3.7) or (3.8).

Through the choice of $\mathcal{R}_1(u)$ and $\mathcal{R}_2(v)$ we can regularize each parameter separately according to a priori information. To penalize the difference between the geophysical structures of two parameters and thus avoid cross-talk, an additional constraint is sometimes added to the optimization, see [30].

For the optimization we use the inexact (quasi-)Newton method with the Eisenstat-Walker stopping criteria described in Section 2.3. Since the modifications to the optimization method are fairly straightforward, we will omit here the details.

9.2 Cross-Talk

In this section, we illustrate the cross-talk between parameters through numerical examples and simulations, and show how the parameters cause undesired artifacts in each other.

The Gradients of the Multi-Parameter Inverse Problem

The multi-parameter gradients are coupled and given in the directions of u and v as

$$\nabla_u \mathcal{F}(u,v) \;=\; \sum_{\ell=1}^{N_s} \left(\frac{\partial y_\ell(u,v)}{\partial u} \right)^\top P^\top \left(PA(u,v)^{-1} f_\ell - y_\ell^{obs} \right)\,, \tag{9.8}$$

$$\nabla_v \mathcal{F}(u,v) \;=\; \sum_{\ell=1}^{N_s} \left(\frac{\partial y_\ell(u,v)}{\partial v} \right)^\top P^\top \left(PA(u,v)^{-1} f_\ell - y_\ell^{obs} \right)\,. \tag{9.9}$$

In equations (9.8) and (9.9), we see that one parameter appears in the gradient of the other. Hence a small change in one parameter can perturb the gradient of the other [67]. Clearly, if we set for both parameters the true value, both gradients (9.8) and (9.9) vanish. A small perturbation in one of the parameters, however, may cause a non-zero gradient in both parameters, even if the other parameter already equals the true value.

We illustrate the issue of coupled gradients for the multi-parameter inverse problem through two slightly different test examples. The examples are chosen such that in both of

them, one parameter shows the true value and the other one shows a value which is very close to the true one. In the first example, we modify the variable u slightly and fix the parameter v at the true solution. In the second example, we make the same modification, this time in the parameter v, and let the parameter u show the true solution. Hence, we would expect a zero gradient in the direction of this parameter, and non-zero gradient in the other. In the following examples this is, unfortunately, not the case. In both cases, a perturbation occurs in both parameter's gradients.

Note that we are not optimizing or reconstructing in any of the two examples below, we are just interested in their reduced-space gradients.

Numerical Examples

We consider u and v from (9.4). For the calculation of the reduced-space gradient, we consider the following settings:

1. A two dimensional bounded region $\Omega = (0,1) \times (0,1)$.

2. We use nine Gaussian sources located at $(0.1, 0.8)$, $(0.2, 0.8)$, $\ldots$, $(0.9, 0.8)$.

3. Time frequency $\omega = 8$.

4. All parameters are discretized with second order staggered finite differences on a 200×200 Cartesian mesh (i.e. no AE representation).

Example 1:

In Fig. 9.1, we present a two-parameter profile in Ω. Both parameters, u and v, are piecwise constant inside a circle O centered at $(0.5, 0.5)$ and with radius 0.1.

We set the value of the *true parameters* as follows:

$$\begin{cases} u(x) = 1.2 & x \in O\,, \\ u(x) = 1 & x \in \Omega \backslash O\,, \end{cases} \qquad \begin{cases} v(x) = 1.5 & x \in O\,, \\ v(x) = 1 & x \in \Omega \backslash O\,. \end{cases} \tag{9.10}$$

For the initial guess, we slightly perturb the true value of v inside O by 0.01 but keep u at the true value:

$$\begin{cases} u(x) = 1.2 & x \in O\,, \\ u(x) = 1 & x \in \Omega \backslash O\,, \end{cases} \qquad \begin{cases} v(x) = 1.49 & x \in O\,, \\ v(x) = 1 & x \in \Omega \backslash O\,. \end{cases} \tag{9.11}$$

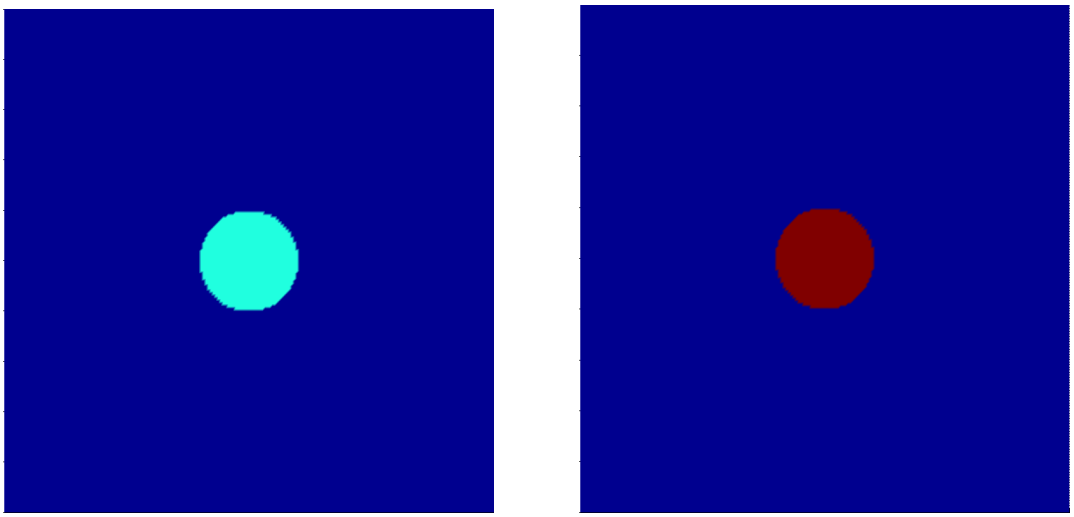

Figure 9.1: Two parameter profile: the true value of u (left), the true value of v (right)

Example 2:

For the second example, we use the same true profile (9.10) and consider the following initial guess where the value of u inside O is slightly perturbed, wheras v is kept at the true value:

$$\begin{cases} u(x) = 1.19 & x \in O\,, \\ u(x) = 1 & x \in \Omega\backslash O\,, \end{cases} \qquad \begin{cases} v(x) = 1.5 & x \in O\,, \\ v(x) = 1 & x \in \Omega\backslash O\,. \end{cases} \tag{9.12}$$

The Reduced Gradient:

The last two examples correspond to two initial guesses for the same true profile, where the initial value is exact for one parameter and slightly modified for the other. Now, we compute the reduced-space gradients for both parameters and initial guesses. In the upper row of Fig. 9.2 we show the gradients for u and v of the first example. In the lower row of Fig. 9.2 we show the gradients for u and v of the second example. Note that the gradients for both parameter in both experiments are non-zero (they are shown on the same scale).

A zero gradient at the optimal value of any parameter is essential for the success of the optimization. If a small change in one parameter perturbs the reduced-space gradient of the other parameter, even if this parameter is at the optimal value, we can expect significant cross-talk between the parameters, as will be illustrated in the following.

A small modification in one parameter causes not only perturbations in the other parameter's gradient, it causes also regularization problems. In both examples shown in Fig. 9.2, unwanted artifacts can be seen at the boundary and at the source locations. This illustrates not only the cross-talk between parameters, but also the ill-posedness of the problem: even when we start that close to the true profile, the gradients may point to a false solution.

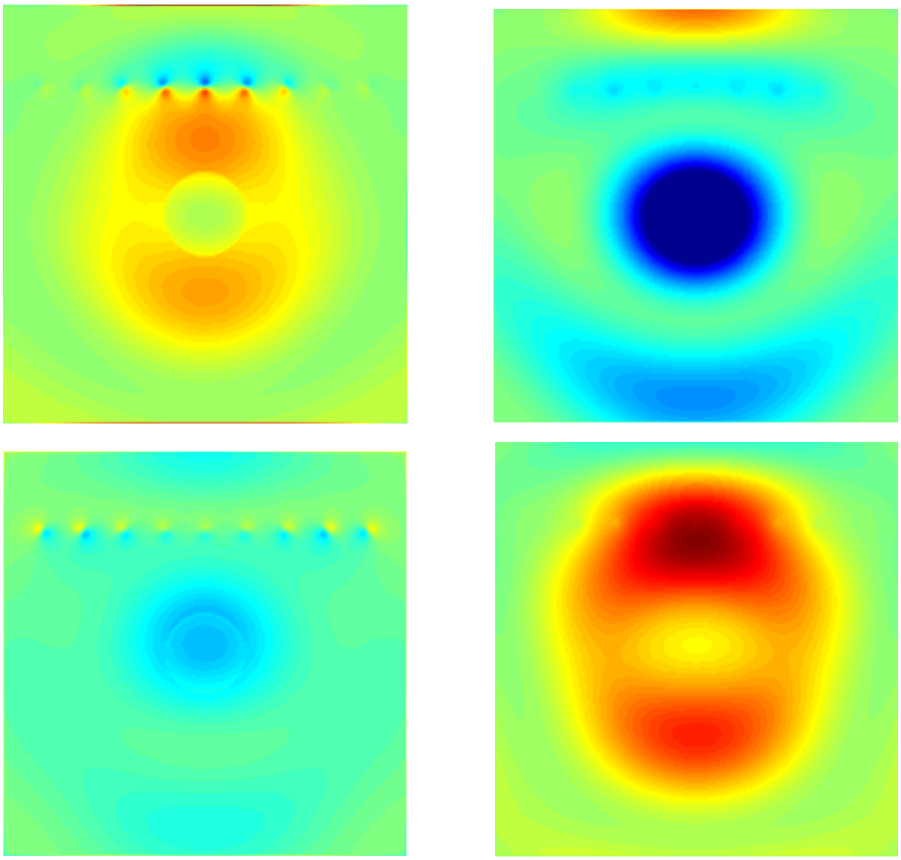

Figure 9.2: Reduced gradient of each parameter for the profile (9.10): the gradients of u (left) and v (right), top: for the initial guess (9.11), bottom: for the initial guess (9.12).

Numerical Simulation

To illustrate the problem of cross-talk in inverse problems, we consider again (9.4) in the unit square $\Omega = (0,1) \times (0,1)$. To reconstruct u and v, we opt for the reduced-space approach with the inexact Newton method using the full Hessian. We adapt the inexact (quasi-)Newton algorithm in Section 2.4 to the two-parameter case and set H to the exact Hessian. Here, we consider the standard nodal basis representation for u and v.

In the upper row of Fig. 9.3, we show the two-parameter target profile. The parameters u and v are given by

$$\begin{cases} u(x) = 1.5 & x \in D \\ u(x) = 1 & x \in \Omega \backslash D \end{cases} \qquad \begin{cases} v(x) = 1.4 & x \in Q \\ v(x) = 1 & x \in \Omega \backslash Q\,, \end{cases} \tag{9.13}$$

where D is the kite and Q a square with a side length of 0.4 centered at $(0.6, 0.4)$.

We now reconstruct the parameters u and v by solving problem (9.7) using frequency stepping at $\omega = 8,\ 10,\ \ldots,\ 198,\ 200$. The wave field y and the control variables u and v are

discretized with second order FD. The reconstructed u and v are shown in the lower row of Fig. 9.3. Both the shapes of D and Q are nicely reconstructed and their values very close to the true ones. However, we also observe the shadows of one parameter appearing in the other: there is cross-talk between the parameters. Still in this example, the cross-talk does not severely affect the reconstruction of the parameters.

Indeed, in the example presented in Fig. 9.3, the objects do not overlap, which induces only weak coupling artifacts. We now repeat the last simulation, but we place the center of the kite D at $(0.6, 0.35)$, such that the kite and the square now overlap - see Fig. 9.4. Now, we reconstruct the profiles again under the same conditions as before. The reconstruction of u and v shown in the lower row of Fig. 9.4 clearly displays cross-talk which perturbs the reconstruction, unlike in the previous example. Even when using Newton's method with the exact Hessian, the cross-talk creates spurious effects in the reconstruction.

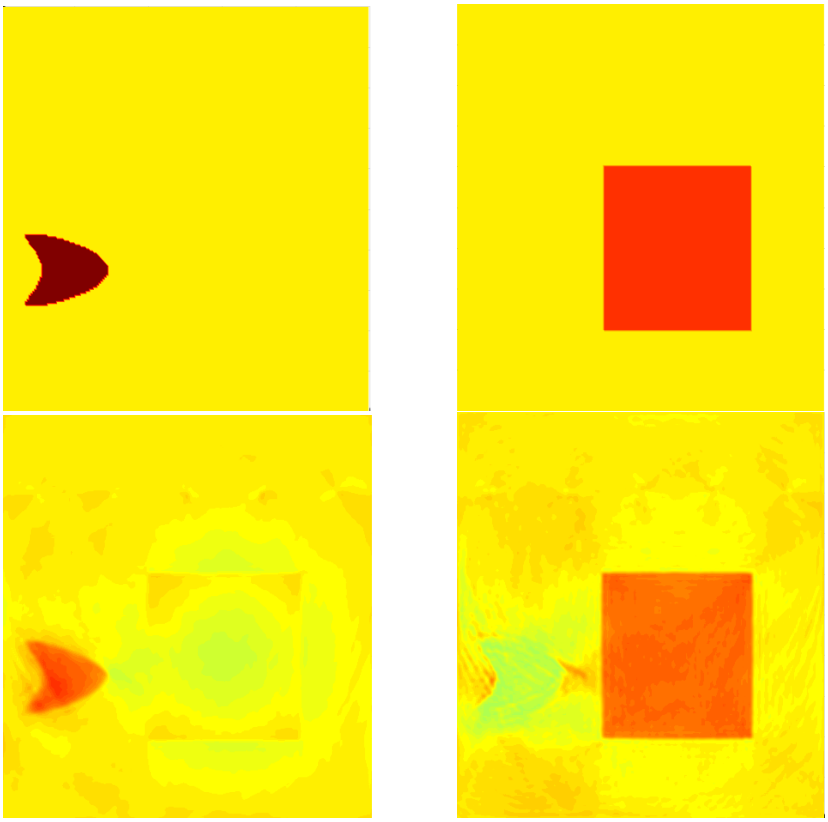

Figure 9.3: Cross-talk on non-overlapping parameters. Top: true parameter u (left) and v (right). Bottom: reconstruction for u (left) and v (right).

The same experiments with other quasi-Newton methods yield similar results: the cross-talk between u and v disturbs the reconstruction just like for the Newton method, see Fig. 9.3 and Fig. 9.4. Our previous example, illustrated in Fig. 9.2, showed that gradient-based methods cannot handle the multi-parameter inverse problem. The current numerical experiment shows that Newton methods alone cannot solve the cross-talk problem.

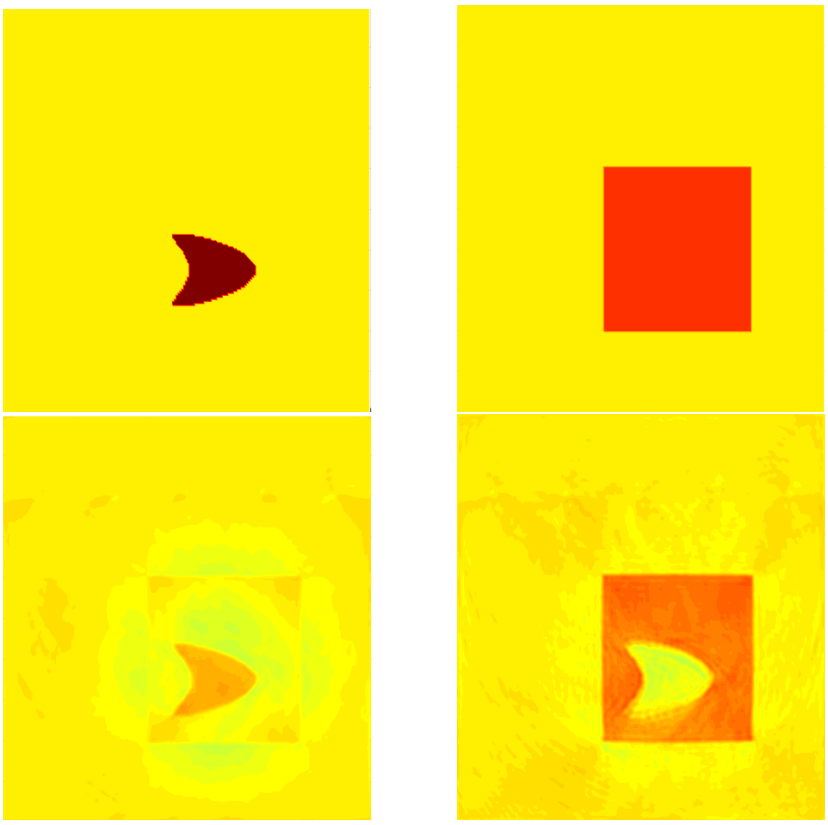

Figure 9.4: Cross-talk on overlapping parameters. Top: true parameter u (left) and v (right). Bottom: reconstruction for u (left) and v (right).

To tackle the cross-talk, two additional strategies are introduced in [58]: a data driven strategy, which defines the expected reduction in the misfit function, or a separate Tikhonov regularization for each parameter. Here, we propose the AEI approach for multi-parameter inverse problems to tackle the cross-talk.

9.3 AEI for the Multi-Parameter Problem

As concluded in Section 9.2, a regularization term must be added to prevent cross-talk. Again, we opt for the AEI approach and illustrate the ability of the AE to reduce cross-talk artifact through numerical examples. In Fig. 9.5 we introduce a two-parameter profile with u and v, where both parameters are piecewise constant. We build for each parameter a separate eigenbasis

$$u(x) \;=\; u_0(x) \;+\; \sum_{m=1}^{K_1} \beta_m \, \phi_m(x) \quad \text{and} \quad v(x) \;=\; v_0(x) \;+\; \sum_{l=1}^{K_2} \gamma_l \, \varphi_l(x), \tag{9.14}$$

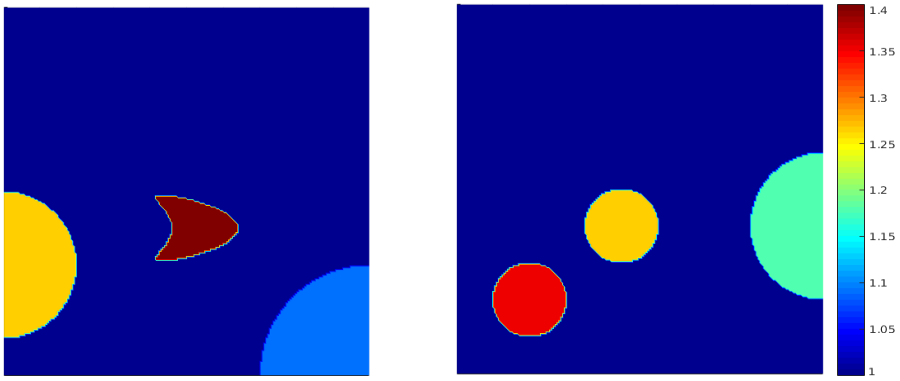

Figure 9.5: The true parameters u (left) and v (right).

where ϕ_m and φ_l are eigenfunctions of two separate bases given by (8.1) with μ in (8.2) from the penalized TV-regularization gradient. The parameters K_1 and K_2 are usually equal, but leave the option to increase either one more slowly than the other.

9.4 Numerical Experiments

We now illustrate the usefulness of the AEI method in preventing cross-talk through two numerical experiments. First, we perform a simulation using a standard grid-based nodal representation of the parameters to demonstrate again the cross-talk problem. Second, we consider AE to reduce the cross-talk.

The parameter settings in the numerical experiments are the following: six super-shot sources (with SAA) are located along the upper boundary at $(0.1, 0.8), \ldots, (0.9, 0.8)$, whereas the receivers are located on the four lateral boundaries of $\Omega = (0, 1) \times (0, 1)$. We discretize (9.4), (4.2)-(4.4) with second-order FD on a 200×200 mesh. We set the initial profiles $u \equiv 1$ and $v \equiv 1$ and perform frequency stepping, starting at $\omega = 8$ and progressively increasing the frequency $\omega = 10, 20, 30, \ldots, 90$. A standard truncated G-N method is used for the optimization, see Section 2.3. For the nodal approach we choose TV-regularization for $\mathcal{R}_1$ and $\mathcal{R}_2$ in (9.7) and for the AEI we set $\alpha_1 = \alpha_2 = 0$ in (9.7), i.e. no added regularization.

For the multi-parameter AEI, we consider separate bases for each parameter. Each basis is obtained by solving (8.1), with μ from (8.2) and $\varepsilon = 10^{-6}$. The number of eigenfunctions K_1, and K_2 respectively, starts at $K_1 = K_2 = 16$ and increases linearly with the frequency ω. In Fig. 9.6, we display the reconstructions for the standard grid-based nodal representation and for the AEI approach.

Although both methods recover the position of the essential features of the medium, the AEI method clearly achieves higher accuracy with minimal cross-talk between the two parameters. In the nodal basis representation, cross-talk appears inside the various subregions, especially where the parameters overlap. Moreover, the value of the kite-shaped

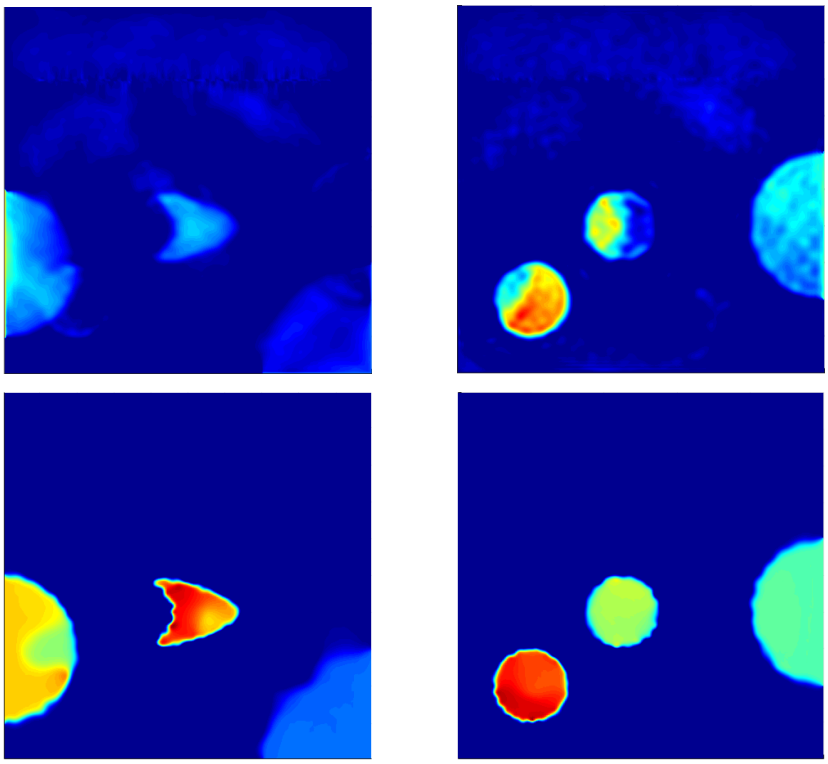

Figure 9.6: Two-parameter profile. Top: the reconstructed parameters u (left) and v (right) for the nodal approach. Bottom: the reconstructed parameters for the AEI approach.

obstacle in the nodal basis reconstruction of u is far from the true value.

Next, we repeat the last numerical experiment, but this time set the initial profiles u and v as follows

$$\begin{cases} u(x) = 1.4 & x \in O\,, \\ u(x) = 1 & x \in \Omega \backslash O\,, \end{cases} \qquad \Big\{ v(x) = 1 \quad x \in \Omega\,,$$

where O is a circle centered at $(0.55, 0.4)$ and with radius 0.2. The initial values for u and v are plotted at the top of Fig. 9.7 together with the reconstructions for the standard grid-based nodal representation (middle) and for the AEI approach (bottom).

The initial value of u helps in both cases to get a more accurate value for the reconstructed kite-shaped obstacle. However, it increases the cross-talk between parameters in both AEI and nodal approaches. In the case of the nodal basis, the shape of O is seen in both parameters and the quarter circle at the bottom right has almost vanished. The cross-talk of the kite-shaped obstacle can be clearly seen in the parameter v.

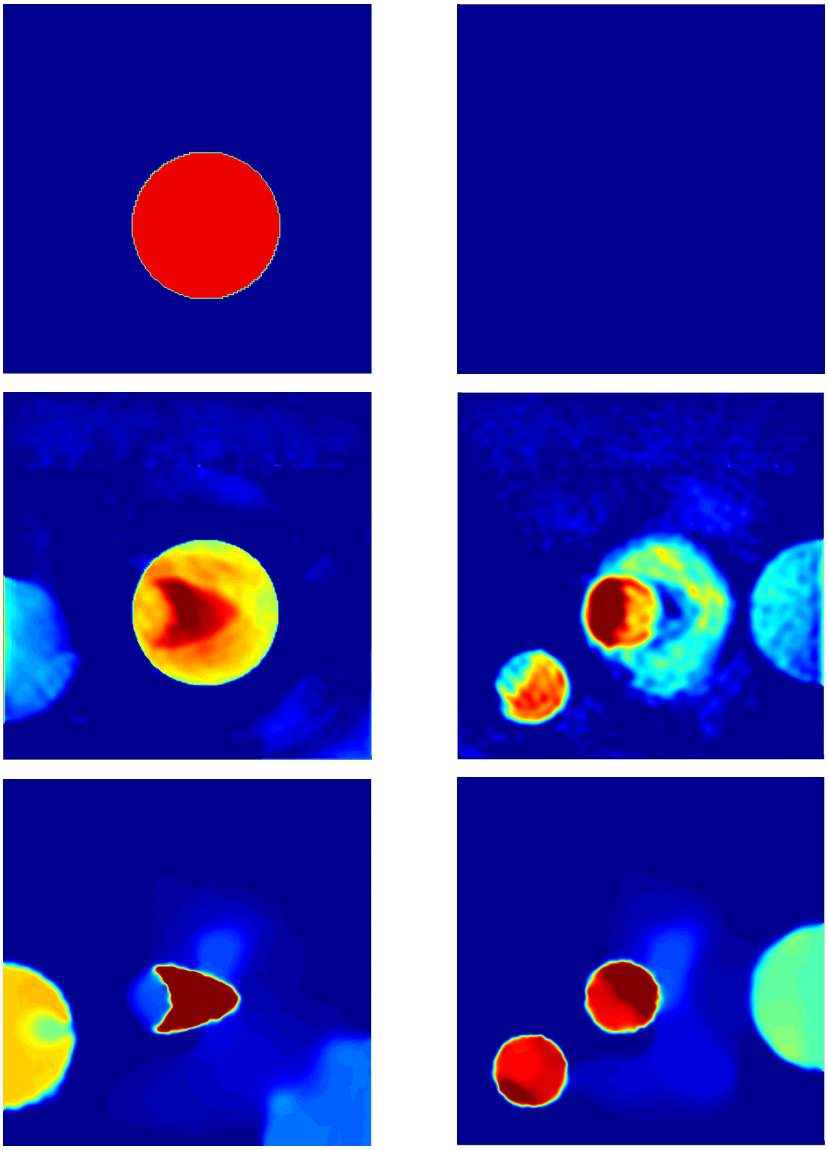

Figure 9.7: Two-parameter profile. Top: the initial value for u (left) and v (right). Middle: the reconstructed parameters for the nodal basis. Bottom: the reconstructed parameters for the AEI approach.

Hence, in the case of the nodal approach, the reconstruction become even worse in comparison to the previous example and thus unsatisfactory. On the other hand, in the case of the AEI approach, light perturbations in the subregion around the kite appear in both parameters u and v. Additionally, the value of the circle-shaped obstacle which is placed in the middle of the parameter v is higher than in the true profile and thus, the perturbations in the AEI approach are comparatively negligible.

AE from Different Penalty Functionals

In Chapter 8, several AE have been introduced. Here, we can choose for each parameter a different AE. The choice of a separate AE for each parameter not only allows to include different a priori information for each parameter, but also reduces the cross-talk. We illustrate the choice of different AE for each parameter through an example. In Fig. 9.8 we introduce a two-parameter profile with u and v.

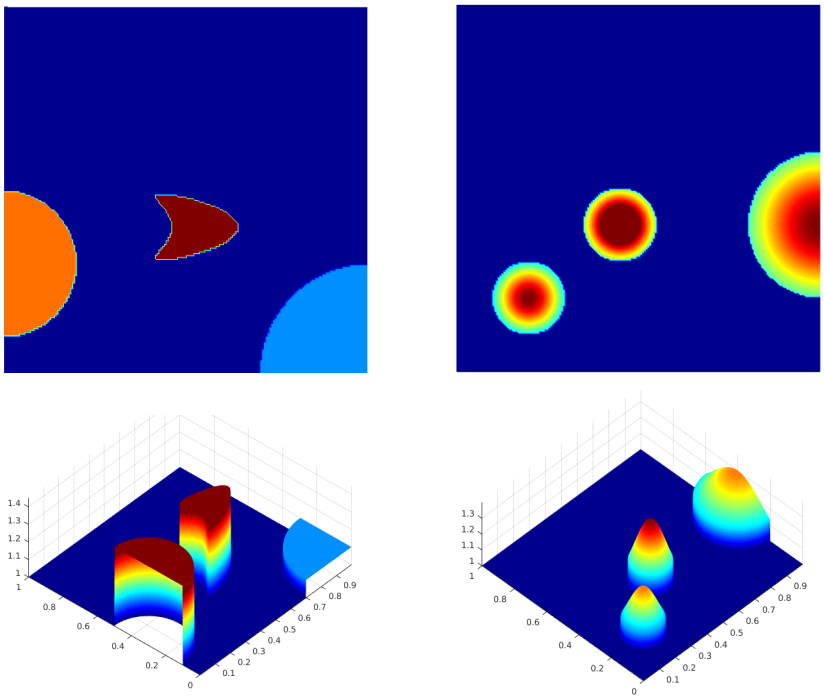

Figure 9.8: The true parameters u (left) and v (right). Two-dimensional view (top) and three-dimensional view (bottom).

For the chosen example, the following a priori information is available: the profile u is piecewise constant and the obstacles in v are smooth. We shall now compare the reconstructions with both the nodal basis approach and the multi-parameter AEI.

The parameter settings for the numerical experiments are similar to the first numerical experiment of this section. Here, we use second-order FD on a 500×500 mesh for the discretization and set the initial profiles $u \equiv 1$ and $v \equiv 1$. From the a priori information on the parameters, we choose judicious bases and expand u and v in (9.14), where ϕ_m are the penalized TV-regularization AE functions, given by (8.1), with μ from (8.2) and $\varepsilon = 10^{-6}$, and the eigenfunctions φ_l are the H^1-regularization AE functions given by (8.3). The number of eigenfunctions K_1 and K_2 starts at $K_1 = K_2 = 16$ and increases linearly with the frequency ω. Similarly, for the nodal basis approach, we choose penalized TV-regularization for $\mathcal{R}_1$ and H^1-penalty term for $\mathcal{R}_2$ in (9.7).

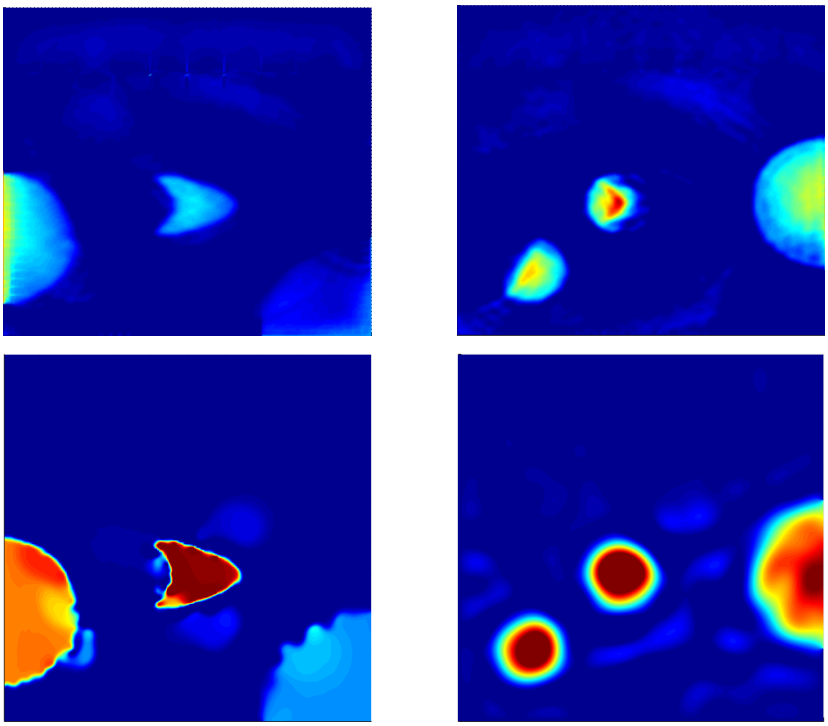

Figure 9.9: Reconstruction of the two-parameter profile with different regularizations for each parameter u (left) and v (right). Top: for the nodal basis representation. Bottom: for the multi-parameter AEI.

In Fig. 9.9, we display the reconstructions for the standard grid-based nodal representation and for the AEI approach. Again, both methods recover the position of the essential features of the medium, but the AEI method clearly achieves higher accuracy with minimal cross-talk. In the last numerical experiments, the different eigenspaces for each parameter are not only efficiently regularizing the problem, but are also an efficient tool to reduce the cross-talk between the parameters.

Conclusions

We have presented a nonlinear optimization method for the solution of inverse scattering problems in the frequency domain, when the scattered field is governed by the Helmholtz equation with one or two parameters. Instead of a standard (FD or FE) grid-based representation, the unknown (squared) sound speed u is projected to the finite-dimensional subspace $V_K = \text{span}\{u_0, \phi_1, \ldots, \phi_K\}$ of much smaller dimension (regularization by space reduction). The "background" u_0 is determined by solving (4.2) whereas the remaining orthonormal basis functions ϕ_m are determined by computing the first K eigenfunctions in (4.4). The time-harmonic inverse medium problem is formulated as a PDE-constrained optimization problem and solved by an inexact truncated Newton or quasi-Newton iteration. During the optimization process, which may include frequency continuation, both the basis and the dimension of V_K are repeatedly adapted to the current iterate (adaptive regularization). The full Adaptive Eigenspace Inversion (AEI) Algorithm is given in Section 4.2.

For monotonic one-dimensional or layered media, we have proved that our choice for u_0 is in fact optimal. For arbitrary media, our numerical results suggest that it is clearly superior to a straightforward harmonic extension from the known boundary values. Together with but a few eigenfunctions, the adaptive eigenspace basis yields a remarkably accurate representation of u. At higher eigenvalues, the eigenfunctions become increasingly oscillatory while no longer carrying useful information about u. Hence adapting the dimension of the eigenspace basis in (4.5), effectively builds a priori regularization into the inversion, so that no additional Tikhonov regularization is needed – see Remark 4. As the frequency ω increases, smaller scale features of the scatterer become visible and the dimension of V_K thus ought to increase accordingly.

In contrast to a standard grid-based nodal representation, the AEI method uses much fewer control variables for u. Still the reconstructions are remarkably accurate, display less artifacts and prove more tolerant to partial or missing data. Moreover, our AEI method leads to a significant reduction in execution time and proves robust with respect to added noise. By combining it with the Sample Average Approximation (SAA) approach from [38], it also efficiently handles large numbers of sources.

Since the discrete version of the eigenvalue problem (4.4) leads to a sparse, symmetric and positive definite matrix, the first K eigenfunctions can be efficiently computed via a standard Lanczos iteration. If finite element mesh adaptation is used for the numerical so-

lution of (4.4), small-scale features and interfaces are captured with even greater accuracy in the reconstruction, without increasing the computational effort. Although the eigenfunctions are global, their information content is highly localized in space so that most entries are in fact negligible.

Future Work

As summarized above, the AEI method is a remarkable tool, where both analytical and numerical evidence underpins the accuracy and versatility of it. Numerical experiments demonstrate the efficiency and robustness to missing or noisy data of the resulting adaptive eigenspace method. It is able to regularize efficiently the optimization process for one or two-parameter inverse problems, where in the latter case it is an efficient tool to reduce the cross-talk between the parameters. However, there are still open questions.

In Chapter 5, we showed some analytical evidence for the choice of the basis. We would like in the future to prove the optimality of this choice for two and three dimensions. In the AE approach, truncating the expansion of the parameter u at a finite value K effectively builds regularization into the AEI approach. In the future we would like to find an analytical proof, which supports our automatic determination of the parameter K.

In the numerical experiments in Chapter 6, we showed the versatility of the AE approach and in Chapter 7 we showed how to deal with single frequency data, hence we would like to expand our approach to other inverse problems and PDE constraints, for example the inverse Maxwell problem. In Section 6.8, we discussed the computational cost of the AEI approach, we would like to improve and parallelize our program using a package for PDE parameter estimation (jInv) [75] implemented in Julia[1]. Then we may also implement our method in three space dimensions.

Finally, we have seen in Chapter 9, how the AE is able to remarkably reduce cross-talk between parameters. We would like to extend our AE approach to more realistic multi-parameter problems using for example the Marmousi II profile [56] and be able to remove cross-talk regardless of the starting value or the target profiles.

[1] `http://julialang.org`

Bibliography

[1] G. S. Abdoulaev, K. Ren, and A. H. Hielscher. Optical tomography as a pde-constrained optimization problem. *Inverse Problems*, 21(5):1507, 2005.

[2] M. G. Amin. *Through-the-wall radar imaging*. CRC press, 2016.

[3] A. Y. Aravkin and T. Van Leeuwen. Estimating nuisance parameters in inverse problems. *Inverse Problems*, 28(11):115016, 2012.

[4] A. Asnaashari, R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux. Regularized seismic full waveform inversion with prior model information. *Geophysics*, 78(2):R25–R36, 2013.

[5] K. Baganas, B. B. Guzina, A. Charalambopoulos, and G. D. Manolis. A linear sampling method for the inverse transmission problem in near-field elastodynamics. *Inverse Problems*, 22(5):1835, 2006.

[6] G. Bao and J. Lin. Imaging of local surface displacement on an infinite ground plane: The multiple frequency case. *SIAM Journal on Applied Mathematics*, 71(5):1733–1752, 2011.

[7] A. Bayliss and E. Turkel. Radiation boundary conditions for wave-like equations. *Comm. Pure Appl. Math.*, 33(6):707–725, 1980.

[8] J.-P. Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200, Oct. 1994.

[9] D. Braess. *Finite elements: Theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, 2007.

[10] F. Cakoni and D. Colton. On the mathematical basis of the linear sampling method. *Georgian Mathematical Journal*, 10(3):411–425, 2003.

[11] F. Cakoni and D. L. Colton. *Qualitative methods in inverse scattering theory: an introduction*. Interaction of mechanics and mathematics series. Springer, 2006.

[12] M. Cassier and C. Hazard. Multiple scattering of acoustic waves by small sound-soft obstacles in two dimensions: mathematical justification of the Foldy–Lax model. *Wave Motion*, 50(1):18–28, 2013.

[13] G. Chavvent. *Nonlinear Least Squares for Inverse Problems*. Springer Science & Business Media, 1996.

[14] Y. Chen. Inverse scattering via Heisenberg's uncertainty principle. *Inverse Problems*, 13(2):253, 1997.

[15] M. Cheney. The linear sampling method and the MUSIC algorithm. *Inverse Problems*, 17(4):591–595, 2001.

[16] D. Colton and A. Kirsch. A simple method for solving inverse scattering problems in the resonance region. *Inverse Problems*, 12(4):383–393, 1996.

[17] M. de Buhan and M. Kray. A new approach to solve the inverse scattering problem for waves: combining the TRAC and the Adaptive Inversion methods. *Inverse Problems*, 29(8):085009, 2013.

[18] M. de Buhan and A. Osses. Logarithmic stability in determination of a 3D viscoelastic coefficient and a numerical example. *Inverse Problems*, 26(9):95006, 2010.

[19] R. S. Dembo and T. Steihaug. Truncated-Newton algorithms for large-scale unconstrained optimization. *Mathematical Programming*, 26(2):190–212, 1983.

[20] D. C. Dobson and F. Santosa. Recovery of blocky images from noisy and blurred data. *SIAM Journal on Applied Mathematics*, 56(4):1181–1198, 1996.

[21] S. Durand and M. Nikolova. Stability of the minimizers of least squares with a non-convex regularization. part i: Local behavior. *Applied Mathematics and Optimization*, 53(2):185–208, 2006.

[22] S. C. Eisenstat and H. F. Walker. Choosing the forcing terms in an inexact Newton method. *SIAM J. Sci. Comput.*, 17(1):16–32, 1996.

[23] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[24] B. Engquist and B. D. Froese. Application of the Wasserstein metric to seismic signals. *arXiv preprint arXiv:1311.4581*, 2013.

[25] B. Engquist, B. D. Froese, and Y. Yang. Optimal transport for seismic full waveform inversion. *arXiv preprint arXiv:1602.01540*, 2016.

[26] B. Engquist and A. J. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Math. Comp.*, 31(139):629–651, 1977.

[27] I. Epanomeritakis, V. Akçelik, O. Ghattas, and J. Bielak. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion. *Inverse Problems*, 24(3):034015, 2008.

[28] E. Esser, L. Guasch, T. van Leeuwen, A. Y. Aravkin, and F. J. Herrmann. Total variation regularization strategies in full waveform inversion for improving robustness to noise, limited data and poor initializations. 2015.

[29] S. Esterhazy and J. M. Melenk. On stability of discretizations of the Helmholtz equation. In *Numerical analysis of multiscale problems*, pages 285–324. Springer, 2012.

[30] A. L. Gallardo and M. A. Max. Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints. *Journal of Geophysical Research*, 109(B3).

[31] G. H. Golub, P. C. Hansen, and D. P. O'Leary. Tikhonov regularization and total least squares. *SIAM Journal on Matrix Analysis and Applications*, 21(1):185–194, 1999.

[32] M. J. Grote, J. Huber, D. Kourounis, and O. Schenk. Inexact interior-point method for PDE-constrained nonlinear optimization. *SIAM J. Sci. Comp.*, 36(3):A1251–A1276, 2014.

[33] M. J. Grote, M. Kray, and U. Nahum. Adaptive eigenspace method for inverse scattering problems in the frequency domain. preprint hal-01342694, July 2016.

[34] M. J. Grote and I. Sim. Perfectly matched layer for the second-order wave equation. In *Waves 09*, number EPFL-CONF-173740, pages 370–371, 2009.

[35] E. Haber, U. M. Ascher, and D. Oldenburg. On optimization techniques for solving nonlinear inverse problems. *Inverse Problems*, 16:1263, 2000.

[36] E. Haber, U. M. Ascher, and D. W. Oldenburg. Inversion of 3D electromagnetic data in frequency and time domain using an inexact all-at-once approach. *Geophysics*, 69(5):1216–1228, 2004.

[37] E. Haber and M. Chung. Simultaneous source for non-uniform data variance and missing data. available on arXiv:1404.5254, submitted in 2014.

[38] E. Haber, M. Chung, and F. Herrmann. An effective method for parameter estimation with PDE constraints with multiple right-hand sides. *SIAM J. Optim.*, 22:739–757, 2012.

[39] E. Haber and S. MacLachlan. A fast method for the solution of the Helmholtz equation. *Journal of Computational Physics*, 230(12):4403–4418, 2011.

[40] H. Haddar, A. Lechleiter, and S. Marmorat. An improved time domain linear sampling method for Robin and Neumann obstacles. *Applicable Analysis*, 0:1–22, 2013.

[41] P. C. Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, volume 4. Siam, 1998.

[42] P. C. Hansen and D. P. O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.

[43] F. Hecht. New development in FreeFem++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[44] R. Herzog and K. Kunisch. Algorithms for PDE-constrained optimization. *GAMM-Mitt.*, 33(2):163–176, 2010.

[45] J. Hicken and J. Alonso. Comparison of reduced- and full-space algorithms for PDE-constrained optimization. In *51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, volume No. AIAA–2013–1043, Grapevine, Texas, United States, 2013.

[46] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich. *Optimization with PDE Constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*. Springer Netherlands, first edition, 2009.

[47] J. Huber. *Interior-Point Methods for PDE-Constrained Optimization*. PhD thesis, University of Basel, 2013.

[48] T. Irons. Marmousi model. www.reproducibility.org/RSF/book/data/marmousi/ paper.pdf.

[49] C. Johnson. *Numerical solution of partial differantial equations by the finite elements method*. Cambridge University Press, first edition, 1994.

[50] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Springer, 1996.

[51] A. Kirsch. The MUSIC-algorithm and the factorization method in inverse scattering theory for inhomogeneous media. *Inverse Problems*, 18(4):1025–1040, 2002.

[52] A. Lechleiter. The factorization method is independent of transmission eigenvalues. *Inverse Problems and Imaging*, 3(1):123–138, 2009.

[53] R. B. Lehoucq and D. C. Sorensen. Deflation techniques for an implicitly re-started Arnoldi iteration. *SIAM J. Matrix Anal. Appl*, 17:789–821, 1996.

[54] A. K. Louis and E. T. Quinto. Local tomographic methods in sonar. In *Surveys on solution methods for inverse problems*, pages 147–154. Springer, 2000.

[55] M. Maharramov and B. Biondi. Robust joint full-waveform inversion of time-lapse seismic data sets with total-variation regularization. *arXiv preprint arXiv:1408.0645*, 2014.

[56] G. S. Martin, R. Wiley, and K. J. Marfurt. Marmousi 2: An elastic upgrade for Marmousi. *The Leading Edge*, 25(2):156–166, 2006.

[57] J. Melenk and S. Sauter. Convergence analysis for finite element discretizations of the Helmholtz equation with Dirichlet-to-Neumann boundary conditions. *Mathematics of Computation*, 79(272):1871–1914, 2010.

[58] L. Métivier, F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux. Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures. *Geophysical Prospecting*, 62(6):1353–1375, 2014.

[59] L. Métivier, R. Brossier, J. Virieux, and S. Operto. Full waveform inversion and the truncated Newton method. *SIAM J. Sci. Comput.*, 35(2):B401–B437, 2013.

[60] P. P. Moghaddam, F. J. Herrmann, et al. Randomized full-waveform inversion: a dimenstionality-reduction approach. In *2010 SEG Annual Meeting*. Society of Exploration Geophysicists, 2010.

[61] P. Mojabi and J. LoVetri. Microwave biomedical imaging using the multiplicative regularized Gauss-Newton inversion. *Antennas and Wireless Propagation Letters, IEEE*, 8:645–648, 2009.

[62] V. A. Morozov. The error principle in the solution of operational equations by the regularization method. *USSR Computational Mathematics and Mathematical Physics*, 8(2):63–87, 1968.

[63] S. G. Nash. A survey of truncated-Newton methods. *J. Comput. Appl. Math.*, 124(1-2):45–59, 2000.

[64] J. Nocedal. Updating quasi-Newton matrices with limited storage. *Math. Comp.*, 35:773–782, 1980.

[65] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 2 edition, 2006.

[66] W. Pan and K. Innanen. Suppress parameter cross-talk for elastic full-waveform inversion: Parametrization and acquisition geometry. *GeoConvension*, 2016.

[67] B. Peters and F. J. Hermann. A sparse reduced Hessian approximation for multi-parameter wavefield reconstruction inversion. *SEG Denver Annual Meeting*, 2014.

[68] R. Potthast. A survey on sampling and probe methods for inverse problems. *Inverse Problems*, 22(2):R1, 2006.

[69] C. Prada, S. Manneville, D. Spoliansky, and M. Fink. Decomposition of the time reversal operator: Application to detection and selective focusing on two scatterers. *J. Acoust. Soc. Am.*, 99(4):2067–2076, 1996.

[70] R. G. Pratt. Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *GEOPHYSICS*, 64(3):888–901, 1999.

[71] R. G. Pratt, C. Shin, and G. J. Hicks. Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophys. J. Int.*, 133(2):341–362, 1998.

[72] V. Prieux, R. Brossier, S. Operto, and J. Virieux. Multiparameter full waveform inversion of multicomponent Ocean-Bottom-Cable data from Valhall field. Part 1: imaging compressional wave speed, density and attenuation. *Geophys. J. Int.*, 194(3):1640–1664, 2013.

[73] A. Quarteroni. *Numerical Models for Differential Problems*, volume 8 of *MS&A*. Springer, second edition, 2014.

[74] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[75] L. Ruthotto, E. Treister, and E. Haber. jInv–a flexible Julia package for PDE parameter estimation. *arXiv preprint arXiv:1606.07399*, 2016.

[76] I. Singer and E. Turkel. A perfectly matched layer for the Helmholtz equation in a semi-infinite strip. *Journal of Computational Physics*, 201(2):439–465, 2004.

[77] L. Sirgue and R. G. Pratt. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248, 2004.

[78] G. D. Smith. *Numerical solution of partial differential equations: finite difference methods*. Oxford university press, 1985.

[79] L.-P. Song, C. Yu, and Q. H. Liu. Through-wall imaging (TWI) by radar: 2-D tomographic results and analyses. *IEEE Transactions on Geoscience and Remote Sensing*, 43(12):2793–2798, 2005.

[80] A. Tarantola. Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266, 1984.

[81] A. Tarantola. *Inverse Problem Theory And Methods For Model Parameter Estimation*. Society for Industrial and Applied Mathematics, 2005.

[82] A. N. Tikhonov. On the stability of inverse problems. *Dokl. Akad. Nauk SSSR*, 39(5):195–198, 1943.

[83] E. Treister and E. Haber. Full waveform inversion guided by travel time tomography. *arXiv preprint arXiv:1607.00968*, 2016.

[84] P. M. van den Berg, A. Abubakar, and J. T. Fokkema. Multiplicative regularization for contrast profile inversion. *Radio Science*, 38(2):n/a–n/a, 2003. 8022.

[85] J. F. M. Van Doren, P. M. J. Van den Hof, J. D. Jansen, and O. H. Bosgra. Parameter identification in large-scale models for oil and gas production. In *Proc. 18th IFAC World Congress*, pages 10857–10862, Milano, Italy, 2011.

[86] T. van Leeuwen and F. J. Herrmann. 3D frequency-domain seismic inversion with controlled sloppiness. *SIAM J. Sci. Comput.*, 36(5):S192–S217, 2014.

[87] J. Virieux and S. Operto. An overview of full-waveform inversion in exploration geophysics. *GEOPHYSICS*, 74(6):WCC1–WCC26, 2009.

[88] C. Vogel. *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics, 2002.

[89] C. R. Vogel and E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17(1):227–238, 1996.

[90] C. Wang, D. Yingest, J. Brittan, and J. Leveille. Fast multi-parameter anisotropic full waveform inversion with irregular shot sampling. *Society of Exploration Geophysicists*, pages 1147–1151, 2014.

[91] K. Wang, T. Matthews, F. Anis, C. Li, N. Duric, and M. A. Anastasio. Breast ultrasound computed tomography using waveform inversion with source encoding. In *SPIE Medical Imaging*, pages 94190C–94190C. International Society for Optics and Photonics, 2015.

[92] S. Xu and G. Lambare. Fast migration/inversion with multivalued ray fields: Part 1-method, validation test, and application in 2D to Marmousi. *Geophysics*, 69(5):1311–1319, 2004.