RICE UNIVERSITY

Approximate Inverse Scattering Using Pseudodifferential Scaling

by

Rami Nammour

A Thesis Submitted
in Partial Fulfillment of the
Requirements for the Degree

Master of Arts

Approved, Thesis Committee:

Dr. William Symes, Chairman
Professor of Computational and Applied Mathematics

Dr. Liliana Borcea
Professor of Computational and Applied Mathematics

Dr. Mark Embree
Professor of Computational and Applied Mathematics

Houston, Texas

October, 2008

Abstract

Approximate Inverse Scattering Using Pseudodifferential Scaling

by

Rami Nammour

This thesis proposes a computationally efficient method for approximating the inverse of the normal operator arising in the linearized inverse problem for reflection seismology.

The inversion of the normal operator using direct matrix methods is computationally infeasible. Approximate inverses estimate the solution of the inverse problem or precondition iterative methods. Application of the normal operator requires an expensive solution of large scale PDE problems. However, the normal operator approximately commutes with pseudodifferential operators, hence shares their near diagonality in a frame of localized monochromatic pulses. Estimation of a diagonal representation in this frame encoded in the symbol of the normal operator:

- follows from its application to a single input vector;

- suffices to approximate its inverse.

I use an efficient algorithm to apply pseudodifferential operators, given their symbol, to construct a rapidly converging optimization algorithm that estimates the symbol of an inverse for the normal operator, thereby approximately solving the inverse problem.

Acknowledgements

First and foremost, I am indebted to my academic advisor Dr William Symes. His guidance and insight are palpable in all parts of this thesis and his enthusiasm for the field of reflection seismology sparked my interest in the field in the first place.

The credit of making this manuscript readable goes to Dr Jan Hewitt. Her patience and perseverance in requiring a better standard of writing were invaluable in bringing my writing effort into fruition.

I also would like to thank Dr Eric Dussaud for his mentorship in my summer internship at Total where I extended my work and gained insight into the method I had developed. It was a great experience to work with someone who shares your interest and is willing to put so much effort to elevate the work to a new level.

I am forever grateful to my thesis committee, Dr Liliana Borcea and Dr Mark Embree for overlooking this work, their contribution enriches my academic career.

To my friends, I owe a huge thank you for guiding me through the initiation rituals into this department and making my transition and my stay here $C^\infty([2006, 2008])$.

Finally, I would like to thank those who are not with me here, but because of whom I am here ...

Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis treats the subject of linearized inverse scattering. The linearized acoustics scattering operator $\mathbf{A}$ maps the model $\mathbf{m}$ to the measured data $\mathbf{d}$. The model consists of an interesting physical property of the earth, typically the velocity, density, bulk modulus, shear modulus or some combination of such properties. These properties are fields, i.e., they depend on the position within the earth. The measured data is usually recorded near the earth's surface by an array of appropriate sensors. Since the linearized (Born) scattering operator is an approximation, we write

$$\mathbf{Am} \approx \mathbf{d}. \tag{1.1}$$

The inverse problem aims at recovering the model from the measured data. The approximate equality is inevitable, due to the nature of physical experiments and the uncertainties associated with any measurement coupled with the Born approximation and the inadequacy of any continuous mechanical model. The fundamental

relationship (1.1) is interpreted in a least squares sense, yielding the normal equations

$$\mathbf{A}^T\mathbf{A}\mathbf{m} = \mathbf{A}^T\mathbf{d} := \mathbf{b}, \tag{1.2}$$

where $\mathbf{A}^T\mathbf{A}$ is the normal operator (Hessian). In seismology $\mathbf{A}^T$ is called the migration or imaging operator, and $\mathbf{b} := \mathbf{A}^T\mathbf{d}$ is an image, known as the migrated image.

Equation (1.2) should, in theory, be inverted for the model $\mathbf{m}$. In practice, the high dimensionality of the normal operator makes its inversion via dense numerical linear algebra methods (e.g., Gaussian Elimination) numerically infeasible. The model $\mathbf{m}$ requires GBytes to store and the operator $\mathbf{A}$ Pflops to compute, ruling out the use of direct methods. Moreover, the structure of the normal operator is such that its application is expensive, limiting the number of numerically tractable iterations of the CG (Conjugate Gradient) algorithm for example, since each iteration requires an application of the normal operator.

In the face of this difficulty, a number of authors have sought an easily invertible approximation to the normal operator. These approximations might be used to approximate the solution directly, or to precondition iterative methods. I shall refer to such methods as *scaling methods*, and the approximations as *scaling factors*. The term *scaling methods* refers to the approximation of the inverse of the normal operator by a space dependent multiplier, a *scaling factor*, to correct the amplitudes of seismic images. Scaling methods have two variants depending on the nature of the approximation: diagonal approximations (Claerbout and Nichols, 1994; Rickett, 2003; Shin et al., 2001; Nemeth et al., 1999; Valenciano et al., 2006; Clément and Chavent, 1993),

and nearly diagonal approximations (Guitton, 2004; Chavent and Plessix, 1999). A detailed discussion of these methods constitutes the subject of chapter 3.

The aforementioned scaling methods derive from empirical observations about the normal operator, and are driven by their cheap computational cost. However, the normal operator is a pseudodifferential operator (Beylkin, 1985; Rakesh, 1988; Stolk, 2000), and is nearly diagonal in phase space in a basis (really, a frame) of localized monochromatic pulses. Accurate and efficient scaling methods based on this observation have been introduced by Symes (2008) and Herrmann et al. (2008b). These methods are grounded in the theory of pseudodifferential operators and are faithful to the theoretical underpinnings of linearized inverse scattering.

I propose a scaling method that leverages the near diagonality of the normal operator in phase space to devise an efficient algorithm to approximate the scale factor. Herrmann et al. (2008b) use an explicit basis that approximately diagonalizes the normal operator and express the data in this basis. The essential motivation behind this approach is obtaining a way to efficiently apply the normal operator to data and ultimately invert it. The "eigenvalues" of a pseudodifferential operator are encoded in its symbol, a function in phase space with certain characteristic asymptotic behavior. I observe that approximate inversion may still be accomplished as long as we are given an algorithm to efficiently apply pseudodifferential operators to data, given their symbols, without explicit use of the localized monochromatic pulses frame.

Bao and Symes (1996) proposed an efficient algorithm to approximate the action

of a pseudodifferential operator in terms of a spherical harmonics expansion of its symbol. This thesis uses this algorithm to approximate the action of the normal operator and its inverse, and formulates the recovery of the scale factor as an optimization problem. The efficiency of the Bao and Symes (1996) approach is pivotal, since optimization scheme requires the application of the approximation to the normal operator at each step. Diagonal scaling methods (spacial multipliers) work well only when the support of the data's Fourier transform is localized near a single direction. The method I present in this manuscript does not suffer from that limitation. This scaling method yields an approximate solution to the linearized inverse problem, and may be used to precondition iterative methods.

I discuss the theory underlying the linearized inverse problem in reflection seismology in the second chapter, and the motivation behind the formulation of the recovery of the scaling factor as an optimization problem. The third chapter concerns a discussion of different scaling methods from the literature. The methods I develop in this thesis and a discussion of the Bao and Symes (1996) algorithm constitute the subject of the fourth chapter. The inversion results on the Marmousi synthetic data set occupies the fifth chapter, along with a "plaid" model that tests the ability of the method to resolve areas in the image where the Fourier transform has multiple directions. Chapter six draws the conclusions from this thesis and is the final chapter.

Chapter 2

Theory

2.1 Introduction

This chapter describes how the normal operator arises from linearization of the forward map and discusses its properties. The asymptotic expansion lemma of pseudodifferential operators makes precise the near diagonal property of the normal operator. A presentation of a linear algebra analogue to the adopted method clarifies the approach of this work, and concludes the chapter.

2.2 Linearization

The acoustic wave equation is the simplest model that describes the reaction of the earth (variation in the pressure field) due to an acoustic excitation (explosion, air

guns ...), and may be written as

$$\frac{1}{\rho(x)c^2(x)}\frac{\partial^2 p}{\partial t^2}(x,t) - \nabla \cdot \frac{1}{\rho(x)}\nabla p(x,t) = f(x,t), \tag{2.1}$$

where $\rho(x)$ is the density field, $c(x)$ the velocity field, and $p(x,t)$ the pressure field varying as a function of time; $f(x,t)$ is the source of acoustic energy. This thesis will treat the case where $x \in \mathbb{R}^2$ (two spatial dimensions); however, (2.1) applies equally well to three spatial dimensions.

Assuming the earth was at equilibrium before the forcing is put to effect (causal source), we complement (2.1) with:

$$\begin{aligned} p(x,t) &\equiv 0, \quad t \ll 0 \\ f(x,t) &\equiv 0, \quad t \ll 0. \end{aligned} \tag{2.2}$$

The physical setting of the experiment will invariably enforce some boundary conditions, at the sea surface for example in the case of a marine geophysical experiment.

An abstraction of the wave equation (2.1) regards it as a "rule" that associates a model (given density and velocity field) to the pressure field $p(x,t)$ sampled at various spacial positions on the surface. The appeal of this abstraction lies in its possible generalization to any equation modeling the behavior of the earth (acoustic wave equation, linear elasticity ...). In other words, we define the forward map S associating the model $\mathbf{m} = \{c(x), \rho(x), \dots\}$ to the measured data p at the surface according to the acoustic wave equation (2.1),

$$S[\mathbf{m}] = p|_{surface}. \tag{2.3}$$

The inverse problem consists of recovering the model $\mathbf{m}$ from the measurement of the pressure at the surface $S^{obs} := p|_{surface}$. The structure of S makes the inverse problem large scale. Moreover, the nonlinearity of S amplifies the complexity of the problem. Most standard seismic processing is based on *linearization*, and I shall discuss only the resulting linear inverse problem.

Given a background model $\mathbf{m}_0$ and a perturbation $\delta\mathbf{m}$ to $\mathbf{m}_0$, write

$$\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}. \tag{2.4}$$

The linearized forward map $\mathbf{A}$ ("Born modeling") is defined by

$$\mathbf{A}[\mathbf{m}_0]\delta\mathbf{m} = \delta p|_{surface}, \tag{2.5}$$

in which δp is the perturbation of the pressure field. So formally: $\mathbf{A}[\mathbf{m}_0] = DS[\mathbf{m}_0]$.

Again, (2.5) is an abstraction of the linearization of the acoustic wave equation. An explicit linearization of the acoustic wave equation for example yields:

$$\begin{gathered} \frac{1}{\rho_0 c_0^2}\frac{\partial^2 \delta p}{\partial t^2} - \nabla\cdot\frac{1}{\rho_0}\nabla\delta p = \frac{2\delta c}{\rho_0 c_0^3}\frac{\partial^2 p}{\partial t^2} - \frac{1}{\rho_0}\nabla\frac{\delta\rho}{\rho_0}\cdot\nabla p_0 \\ \delta p \equiv 0, \quad t \ll 0, \end{gathered} \tag{2.6}$$

where ρ_0 and c_0 are the background density and velocity fields, respectively. The first order perturbations to ρ_0 and c_0 are $\delta\rho$ and δc, respectively. The dependence of the fields on spatial and time variables was dropped for simplicity.

The notation $\mathbf{A}[\mathbf{m}_0]$ stresses the dependence of the linearized forward map on the background model, apparent in (2.6). This dependence is repressed throughout this manuscript but we will allude to this property when necessary. The inverse problem

reduces to a linear subproblem: given $\mathbf{d}$ and $\mathbf{m}_0$, find $\delta\mathbf{m}$ such that

$$\mathbf{A}\delta\mathbf{m} \approx S^{obs} - S[\mathbf{m}_0] := \mathbf{d}, \tag{2.7}$$

where $\mathbf{d}$ is defined to be the data. The approximate equality is a consequence of noisy data and the uncertainty of measurements inherent in physical experiments and error in the physical model (acoustics) and linearization. Interpreting the linear subproblem in a least square sense, yields the normal equations

$$\mathbf{A}^T\mathbf{A}\delta\mathbf{m} = \mathbf{A}^T\mathbf{d}. \tag{2.8}$$

$\mathbf{A}^T\mathbf{A}$ is known as the *normal operator* or *Hessian*. In what follows, the model $\mathbf{m}$ denotes the perturbation $\delta\mathbf{m}$ and the δ notation is dropped. Though the nonlinearity has been alleviated or ignored, the system (2.8) is large scale and cannot be inverted by direct matrix methods. The models are typically sampled on rectangular grids, with a spacing of about 10 m to cover an extent of 10 Km in each direction. Therefore, the typical size of the model $\mathbf{m}$ is $\mathcal{O}(10^3)$ in each spatial direction, so the size of $\mathbf{m}$ in 2D is $\mathcal{O}(10^6)$ and in 3D $\mathcal{O}(10^9)$, hence the normal operator would be of size $\mathcal{O}(10^6 \times 10^6)$ in 2D and $\mathcal{O}(10^9 \times 10^9)$ in 3D. Moreover, the application of the normal operator requires solving large scale PDE problems. The expense of applying the normal operator limits the number of affordable iterations of an iterative method (CG, for example) which require at least one application per iteration.

To deal with this difficulty, a number of authors have sought an approximation $\mathbf{V}^2$ (usually diagonal or nearly diagonal) to the normal operator that is easily invertible.

These approximation methods are known as *scaling methods*. The approximation is in the sense that $\mathbf{V}^2$ acts on a reference vector $\mathbf{m}_{ref}$ in the same way the normal operator does:

$$\mathbf{V}^2\mathbf{m}_{ref} \approx \mathbf{A}^T\mathbf{A}\mathbf{m}_{ref}. \tag{2.9}$$

The normal operator is, however, nearly diagonal in phase space, a manifestation of its pseudodifferential nature when the background model is smooth and under generic conditions in 2D (Beylkin, 1985; Rakesh, 1988; Stolk, 2000). "Nearly diagonal in phase space" means that it acts approximately as a multiplication by a number when applied to localized monochromatic pulses. This property follows from a variant of the asymptotic expansion lemma for pseudodifferential operators (Taylor, 1981). Let $\chi(x)$ be a smooth function compactly supported inside a small ball, and $\Psi(x)$ a smooth function with non-vanishing gradient inside the same ball. I call a function of the form $\chi(x)e^{i\omega\Psi(x)}$ a *localized monochromatic pulse*. Then

$$\mathbf{A}^T\mathbf{A}\chi(x)e^{i\omega\Psi(x)} = q_m(x,\omega\nabla\Psi(x))\chi(x)e^{i\omega\Psi(x)} + O(\omega^{m-\beta}), \tag{2.10}$$

where $\beta > 0$, ω is the frequency and $q_m(x,\omega\nabla\Psi(x))$ is the principal symbol of the normal operator positively homogeneous of order m:

$$q_m(x,\omega\nabla\Psi(x)) = |\omega|^m q_m(x,\nabla\Psi(x)). \tag{2.11}$$

So if the support of $\chi(x)$ is small with respect to the smoothness of q and Ψ, and x_0 is in the support of $\chi(x)$,

$$q_m(x,\omega\nabla\Psi(x))\chi(x) \approx q_m(x_0,\omega\nabla\Psi(x_0))\chi(x). \tag{2.12}$$

whence,

$$\mathbf{A}^T\mathbf{A}\chi(x)e^{i\omega\Psi(x)} \approx q_m(x_0, \omega\nabla\Psi(x_0))\chi(x)e^{i\omega\Psi(x)} + O(\omega^{m-\beta}), \tag{2.13}$$

i.e., $\mathbf{A}^T\mathbf{A}$ acts like multiplication by a number, and $\chi(x)e^{i\omega\Psi(x)}$ is an asymptotic "eigenvector". Moreover, (2.13) reveals that the symbol q encodes the eigenvalues.

The order of the symbol and hence the pseudodifferential operator follows from the underlying theory: in space dimension n the order is $n-1$ (Rakesh, 1988; Stolk, 2000).

Equation (2.10) gains utmost importance in view of the following facts: Any seismic image can be resolved locally into oscillatory factors like $\chi(x)e^{i\omega\Psi(x)}$ by constructing *frames* (redundant bases) from these localized monochromatic pulses. Fourier analysis shows one way to effect this expansion, and other choices exist (wavelets, curvelets,...). In seismic images, the interface between two volumes of space having different physical properties (impedance, reflectivity, ...) constitutes the *reflector*, this discontinuity accounts for the high frequency components (rapid changes) in the expansion. The gradient $\nabla\Psi(x)$ of the phase function $\Psi(x)$ is normal to the level curves of $\Psi(x)$, hence normal to the tangent plane of the reflector at each point. This normal vector is known as the "reflector dip".

2.3 Linear Algebra Analogue

Casting the method in the language of linear algebra clarifies the approach. Given a vector $\mathbf{b}$, we seek to solve a system

$$\mathbf{Bm} = \mathbf{b}, \tag{2.14}$$

for $\mathbf{m}$, in which $\mathbf{B}$ is symmetric positive semi-definite and represents the normal operator. Though possible, application of $\mathbf{B}$ is computationally expensive. Suppose, in addition, that we are given a unitary operator $\mathbf{U}$ that approximately diagonalized the pseudo-inverse $\mathbf{B}^\dagger$, i.e., for some diagonal Λ

$$\mathbf{B}^\dagger \approx \mathbf{U}^T \Lambda \mathbf{U}. \tag{2.15}$$

In other words, $\mathbf{B}$ and $\mathbf{B}^\dagger$ are members of the family of approximately commuting operators, namely those diagonalized by $\mathbf{U}$.

Then, using $\mathbf{b}$ and $\mathbf{Bb}$, formulate the recovery of eigenvalues as an optimization scheme

$$\Lambda_{\mathbf{b}} = \underset{\Lambda}{\operatorname{argmin}} \|\mathbf{b} - \mathbf{U}^T \Lambda \mathbf{U}\, \mathbf{Bb}\|^2. \tag{2.16}$$

If an iterative optimization algorithm is used to estimate $\Lambda_{\mathbf{b}}$, then its cost will depend on the efficiency of applying $\mathbf{U}^T\Lambda\mathbf{U}$: no further applications of $\mathbf{B}$ will be required. As we shall see, application of $\mathbf{U}^T\Lambda\mathbf{U}$ may be much more efficient than application of $\mathbf{B}$. In that case, the main cost of the method is the formation of $\mathbf{Bb}$.

Notice that the recovered eigenvalues are those corresponding to eigendirections that make a nontrivial contribution to $\mathbf{b}$, and $\Lambda_{\mathbf{b}}$ will be the diagonal matrix contain-

ing these eigenvalues. More precisely, if $\mathbf{b}_k = (\mathbf{U}\mathbf{b})_k = 0$, that is $\mathbf{b}$ has no component along the kth eigen-direction, then $\Lambda_{bk} = 0$ regardless of Λ_k. However, in this case

$$\mathbf{m}_k = (\mathbf{U}\mathbf{m})_k = (\mathbf{U}\mathbf{U}^T\Lambda\mathbf{U}\mathbf{b})_k = (\Lambda\mathbf{U}\mathbf{b})_k = \Lambda_k(\mathbf{U}\mathbf{b})_k = 0,$$

i.e., the contribution of the kth eigendirection in the original model is trivial, so will the contribution of the kth eigendirection to the inverted model.

Another interesting case occurs when $\Lambda_k = 0$ for some k, this occurs because the normal operator is not invertible since some areas of the image may not be illuminated for example. We deal with this case by considering the pseudoinverse of $\mathbf{B}$. In this case, it is easy to see that $(\mathbf{B}\mathbf{b})_k = 0$. However, $\mathbf{b}_k = 0$ too; this fact is a direct consequence of the nature of the right hand side $\mathbf{b}$:

$$\mathbf{b} = \mathbf{A}^T\mathbf{d} \in \mathcal{R}(\mathbf{A}^T) \perp \mathcal{N}(\mathbf{A}^T\mathbf{A}) = \mathcal{N}(\mathbf{B}).$$

$\mathcal{R}$ and $\mathcal{N}$ denote the Range and Null spaces of an operator, respectively. Also, the eigenvalue decomposition in this case coincides with the singular value decomposition of $\mathbf{B}$, since $\mathbf{B}$ is symmetric positive semi-definite. Hence, the eigenvectors corresponding to zero eigenvalues constitute a basis for the Null space of $\mathbf{B}$. Therefore, $\mathbf{b}_k = (\mathbf{U}\mathbf{b})_k = 0$ and we are back to the case discussed above.

Note also that it is only the application of $\mathbf{U}^T\Lambda\mathbf{U}$, not of $\mathbf{U}$ itself, that is involved: that is, the algorithm does not require expansion in the basis of $\mathbf{U}$, only efficient application of operators which can be diagonalized by $\mathbf{U}$.

The approximate solution then follows from

$$\mathbf{m} \approx \mathbf{m}_{inv} = \mathbf{U}^T \Lambda_{\mathbf{b}} \mathbf{U} \mathbf{b}. \tag{2.17}$$

The connection between the recovery of eigenvalues of $\mathbf{B} = \mathbf{A}^T\mathbf{A}$ and the recovery of the symbol of the normal operator is closely related to the asymptotic expansion lemma (2.10), which we restate here:

$$\mathbf{A}^T\mathbf{A}\chi(x)e^{i\omega\Psi(x)} = q_m(x, \omega\nabla\Psi(x))\chi(x)e^{i\omega\Psi(x)} + O(\omega^{m-\beta}). \tag{2.18}$$

Recall that $\chi(x)$ is localized in a small ball; therefore, $\chi(x)e^{i\omega\Psi(x)}$are localized monochromatic pulses (eigenvectors of the normal operator). Moreover, $q_m(x, \omega\nabla\Psi(x))$ is smooth and slowly varying in its arguments and is thus approximately constant on that ball. Putting everything together, (2.18) is an asymptotic eigenvalue equation where *the symbol encodes the eigenvalues*. It is therefore natural for the adopted method to formulate the recovery of the symbol of the normal operator as an optimization problem, given the Bao and Symes (1996) algorithm which efficiently applies a pseudodifferential operator given its symbol. This algorithm will play the role mentioned above, of an efficient method to apply the operators diagonalized by a change-of-frame operator $\mathbf{U}$ without explicit access to $\mathbf{U}$.

Chapter 3

Scaling Methods

3.1 Introduction

Scaling methods provide approximate inverses to the normal operator to approximate the solution of the inverse problem or to precondition iterative methods. The scaling factors (approximate inverses to the normal operator) may be used to correct the amplitudes of seismic images as these usually suffer from illumination artifacts. Scaling methods are efficient because they rely on approximating the action of the normal operator from its application on a single vector. In this chapter I discuss previously suggested scaling methods and how they relate to the scaling method I develop as part of this thesis work.

3.2 Scaling Methods

Claerbout and Nichols (1994) proposed a diagonal approximation to the inverse of the Hessian, motivated by empirical observations and personal communications with W.W. Symes. The explicit scale factor suggested by Claerbout and Nichols amounts to the ratio between the *migrated image* $\mathbf{m}_{mig} = \mathbf{A}^T\mathbf{d}$ and the *remigrated image* $\mathbf{m}_{remig} = \mathbf{A}^T\mathbf{A}\mathbf{m}_{mig}$ (Claerbout and Nichols, 1994), more explicitly the reference model in (2.9) used is the migrated image. This scaling method proposed by Claerbout and Nichols (1994) leads to a series of related works discussed here. Also, this thesis shares the same setting with the Claerbout and Nichols (1994) method, as it also uses the migrated image as a reference vector and tries to approximate a scaling factor between the migrated and the remigrated image.

Rickett (2003) generalizes the setting of Claerbout and Nichols (1994) to the one presented in the previous chapter. The diagonal approximation was refined by Rickett (2003). He showed the rationale behind the choice of the approximation and addressed some of the technical difficulties in the implementation of the method (Rickett, 2003). The success of this approximation relies on the accuracy of the reference model, more precisely the similarity between the reverence model and the real model, so that the application of the normal operator scales both in the same way. Rickett quantifies this sensible fact by illustrating the failure of the method for a random reference model (Rickett, 2003). Rickett also asserts that one adequate choice for the reference model is the adjoint image, as proposed by Symes (Claerbout and Nichols, 1994).

This work adopts the migrated image as the reference model, making this previous assertion pivotal for our purposes.

Alternatively, Chavent and Plessix (1999) propose a diagonal approximation of the Hessian by mass lumping, i.e., adding the near diagonal components to the diagonal and regarding that as the scale factor (Chavent and Plessix, 1999). The mass lumping idea seems arbitrary and can only be made to work after an a posteriori correction factor, revealing the rather qualitative nature of the similarity between the Hessian and its approximation. The first choice for a diagonal approximation to the Hessian would be the diagonal itself, and Shin et al. (2001) derive a way to approximate the diagonal of the Hessian and use it to correct the amplitudes of the images (Shin et al., 2001). However, an important observation exposes the limitation of this approximation: the diagonal of the Hessian cannot account for the dependence of the illumination on local reflector dip (Rickett, 2003).

The previously presented methods (Claerbout and Nichols, 1994; Rickett, 2003) are data dependent and hence take account of the dip, though diagonal. Guitton (2004) suggested another approximation capable of resolving the local reflector dip, who proposes a near diagonal approximation to the inverse of the Hessian and uses it directly to estimate the material properties. He proves that this method is comparable to linearized least squares inversion with fewer artifacts and better computational cost. Guitton's approach is widely applicable and is completely data driven, and uses the migrated image as a reference model (Guitton, 2004). Though the previously

mentioned methods depend on a good reference model for their success, the migrated image may be used as a reference model as suggested by Claerbout and Nichols and validated by Rickett, which then makes them also completely data driven. Guitton's method may be thought of as a generalization to the Claerbout and Nichols and Rickett methods, in that the scaling factor is nearly diagonal and allows for more degrees of freedom in correcting the amplitudes than a diagonal scaling factor (Guitton, 2004). In the special case where only a limited area is of interest (a reservoir, for example), a target oriented variant of Guitton's method is presented by Valenciano et al. (2006). The restriction to a target area decreases the scale of the Hessian relative to the entire model, and its sparsity renders its application cheap enough to make least squares inversion numerically feasible and appropriate for obtaining the solution (Valenciano et al., 2006). One drawback of Guitton's scaling method, hence Valenciano's too, is that the "filters" used to derive the approximation are arbitrary integral operators supported near the diagonal, and Guitton's method does not specify them completely.

All the methods mentioned rely on empirical observations about the near diagonality of the normal operator. The theoretical basis for this observation gives more insight about the success and failure of these methods. The near diagonality of the normal operator in phase-space is a manifestation of its pseudodifferential nature when the background parameters are smooth. Moreover, the pseudodifferential nature of the normal operator explains which scaling methods should produce accurate approximate inverses. In particular, the normal operator acts as a multiplication by a

function when applied to a localized monochromatic pulse see (2.10) (Taylor, 1981). However, this simple scaling is only possible after a filtering step that renders the scaling factor an order zero pseudodifferential operator (Symes, 2008). The Symes (2008) paper thus devises a data adaptive filtering/scaling method, the scale factor depends on the model (the migrated image) which depends on the data. Symes uses the migrated image as reference model which parallels the framework of the previous methods in that respect. Rickett's method may be regarded as a special case where filtering is ignored, however the approximate inverse so produced is not as accurate (Symes, 2008). This method may also be regarded as a variant of Guitton's method in which the filter is a completely specified power of the Laplacian predicted by the underlying theory composed with a zero order pseudodifferential operator. Though the method is particularly natural and simple, it assumes a unique local dip, which presents a limitation on its domain of applicability. The method I propose may be regarded as a generalization of this method where no unique local dip is assumed, thereby allowing for more degrees of freedom.

Herrmann et al. (2008b) adopt a different route also based on the asymptotic expansion lemma for pseudodifferential operators (2.10). They use a curvelet frame (Candes and Demanet, 2005), motivated by the sparsity of seismic images in the curvelet domain and the approximate invariance of the curvelet frame under the normal operator (Herrmann et al., 2008b), a direct consequence of (2.10) and the fact that curvelets are localized monochromatic pulses. Herrmann et al. (2008b) develop

a scaling method in which amplitude recovery is achieved by solving a nonlinear optimization problem where the "sparsity in the curvelet domain and the continuity along the imaged reflectors" is imposed as a regularization to the optimization problem. Once again the method depends on a reference model and the migrated image is used as an initial guess to obtain satisfactory images showing fewer artifacts and recovering the amplitudes (Herrmann et al., 2008b).

The use of the curvelet frame requires advanced theoretical tools and an intricate implementation. The curvelet frame explicitly diagonalizes the normal operator rendering its application efficient. However, the efficient application of any pseudo-differential operator is also achieved by approximating its symbol (which encodes the eigenvalues asymptotically) and completely bypassing the eigenvectors; this is the content of the Bao and Symes (1996) algorithm to efficiently apply pseudodifferential operators. I discuss the Bao and Symes (1996) algorithm in the next chapter and use it to derive a scaling method that formulates the recovery of the scaling factor as an optimization scheme without explicitly diagonalizing the normal operator.

Chapter 4

Methods

4.1 Introduction

This chapter discusses the Bao and Symes (1996) algorithm and uses it to formulate the recovery of the approximate inverse to the normal operator as an optimization problem. Also, I present a description of the code implementing the scaling method proposed in this thesis.

4.2 The Algorithm

The pseudodifferential nature of the normal operator is predicted by the underlying theory, hence the need for an algorithm that applies pseudodifferential operators.

Bao and Symes describe an original algorithm for applying a pseudodifferential operator on a function in two dimensions (Bao and Symes, 1996). Though the pre-

sentation is in 2D, the algorithm may be generalized to 3D.

Pseudodifferential operators are defined in terms of their symbols $q_m(\mathbf{x}, \xi)$,

$$q_m(\mathbf{x}, \xi) : \Omega \times \mathbb{R}^n \backslash \{\mathbf{0}\} \to \mathbb{R},$$

where $\Omega \subset \mathbb{R}^n$ is an open set and $n = 2$ or 3 (the dimension of the space).

The symbols of interest $q_m(\mathbf{x}, \xi)$ are smooth and homogeneous of order m, and for any compact set $K \subset \mathbb{R}^n$, and real α,β, there exists constants $C_{K,\alpha,\beta}$, such that

$$|D_x^\alpha D_\xi^\beta q_m(\mathbf{x}, \xi)| \leq C_{K,\alpha,\beta}(1 + |\xi|)^{m-|\beta|}, \tag{4.1}$$

for all $\mathbf{x} \in K$ and $\xi \in \mathbb{R}^n$. Homogeneity means that, given $r \in \mathbb{R}$,

$$q_m(\mathbf{x}, r\xi) = r^m q_m(\mathbf{x}, \xi). \tag{4.2}$$

Homogeneous symbols satisfy (4.1); however, it should be noted that (4.1) is satisfied by a more general class of symbols not treated in this thesis. The applications of this thesis only include symbols of order 0, i.e., $m = 0$; however, the discussion is simple enough for general m and it is therefore explicitly shown in the formulae.

The rest of this discussion is restricted to 2D, so we may write $\mathbf{x} = (x, z)$. The pseudodifferential operator is then characterized by its symbol and defined by

$$Q_m u(x, z) = \int\int q_m(x, z, \xi, \eta)\hat{u}(\xi, \eta)e^{i(x\xi + z\eta)}\, d\xi\, d\eta, \tag{4.3}$$

where $q_m(x, z, \xi, \eta)$ is the principal symbol, homogeneous of degree m, and $\hat{u} = \mathcal{F}[u]$ is the Fourier transform of u.

Thus writing $\xi = \omega \cos\theta$, $\eta = \omega \sin\theta$, and using the homogeneity of q_m, we have

$$q_m(x, z, \xi, \eta) = \omega^m \tilde{q}_m(x, z, \theta). \tag{4.4}$$

Notice that $\tilde{q}_m(x, z, \theta) = q_m(x, z, \cos\theta, \sin\theta)$ is periodic and smooth in θ, and hence it admits a rapidly converging Fourier expansion. We thus truncate the Fourier series, approximating the symbol by its first $K + 1$ Fourier modes:

$$\tilde{q}_m(x, z, \theta) \approx \sum_{l=-K/2}^{l=K/2} c_l(x, z) e^{il\theta} = \sum_{l=-K/2}^{l=K/2} \omega^{-l} c_l(x, z)(\xi + i\eta)^l. \tag{4.5}$$

Plugging (4.5) into (4.3) we obtain

$$Q_m u(x, z) \approx \sum_{l=-K/2}^{l=K/2} c_l(x, z) \mathcal{F}^{-1}[\omega^{m-l}(\xi + i\eta)^l \hat{u}(\xi, \eta)]. \tag{4.6}$$

Fourier transform theory identifies ω^{m-l} as the symbol of $(-\nabla)^{\frac{m-l}{2}}$, and ξ and η are respectively the symbols of $D_x = -i\partial_x$ and $D_z = -i\partial_z$.

Sampling the field $u(x, z)$ and the symbol $\tilde{q}_m(x, z, \theta)$,

$$U_{ij} = u(x_0 + (i-1)\Delta x, z_0 + (j-1)\Delta z),$$

$$Q_{ijk} = \tilde{q}_m(x_0 + (i-1)\Delta x, z_0 + (j-1)\Delta z, k\Delta\theta),$$

$$i = 1, \cdots, M,\ j = 1, \cdots, N,\ k = -K/2, \cdots, K/2.$$

Choosing $\Delta\xi = \frac{1}{(M-1)\Delta x}$ and $\Delta\eta = \frac{1}{(N-1)\Delta z}$ yields the unaliased discretizations of the symbols of the square root of the negative Laplacian, D_x and D_z

$$\Omega_{pr} = 2\pi\sqrt{(p\Delta\xi)^2 + (r\Delta\eta)^2}$$

$$\Xi_{pr} = 2\pi p \Delta\xi$$

$$Z_{pr} = 2\pi r \Delta\eta$$

$$p = -M/2, \cdots, M/2, \quad r = -N/2, \cdots, N/2$$

Equation (4.6) suggests the following algorithm to estimate $Q_m u$ (Bao and Symes, 1996). All Fourier transforms refer to a discrete Fourier transform.

1. Compute $\hat{U}_{pr} = \mathcal{F}[U_{ij}]$.

2. For each $i \in [1, M]$ and $j \in [1, N]$,

   compute $\hat{Q}_{ij} = \{\hat{Q}_{ijl}\}_{l=-K/2}^{K/2}$ the discrete Fourier transform of $Q_{ij} = \{Q_{ijk}\}_{k=-K/2}^{K/2}$.

3. Initialize $(QU)_{ij} = 0$, for $i \in [1, M]$, $j \in [1, N]$,

   For $l = -K/2 : K/2$

   (a) compute $\{R_{ij}^l\}_{i=1,j=1}^{M,N} = \mathcal{F}^{-1}[\Omega_{pr}^{m-l}(\Xi_{pr} + iZ_{pr})^l \hat{U}_{pr}]$

   for $p = -M/2, \cdots, M/2$ and $r = -N/2, \cdots, N/2$

   (b) accumulate

   $$(QU)_{ij} = (QU)_{ij} + \hat{Q}_{ijl} R_{ij}^l$$

   End

A straightforward discretization of (4.3) has a computational complexity of $O(N^4 \log(N))$. The algorithm described above uses *FFT* (Fast Fourier Transform), and thus exhibits a complexity of $O(KN^2(\log(N) + \log(K)))$. The appeal of this approach is that K is independent of N. In fact, applications to reflection seismology

require that the symbol be smooth and slowly varying in θ, thus may be captured accurately by a modest number of Fourier modes or, more explicitly, a small K.

4.3 Inversion

Recall that the problem at hand is that of obtaining the true model $\mathbf{m}_{true}$ from

$$\mathbf{A}^T\mathbf{A}\mathbf{m}_{true} = \mathbf{A}^T\mathbf{d}. \tag{4.7}$$

The following discussion inherits the notation and the procedure of Symes (2008).

The theory predicts that $\mathbf{A}^T\mathbf{A} \approx \mathbf{L}^{\frac{n-1}{2}}\mathbf{V}^2$, where $\mathbf{L} = -\nabla$ and $\mathbf{V}^2$ is an order zero symmetric positive semidefinite pseudodifferential operator.

Define $\mathbf{W}^2 = (\mathbf{V}^2)^\dagger$, the pseudoinverse of $\mathbf{V}^2$, thus,

$$\mathbf{m}_{inv} = \left(\mathbf{A}^T\mathbf{A}\right)^\dagger \mathbf{A}^T\mathbf{d} \approx \mathbf{W}^2\mathbf{L}^{-\frac{n-1}{2}}\mathbf{A}^T\mathbf{d}. \tag{4.8}$$

Using the properties of pseudoinverses, we obtain an expression for $\mathbf{W}^2$ independent of $\mathbf{m}_{true}$

$$\mathbf{W}^2\mathbf{L}^{-\frac{n-1}{2}}(\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{d} \approx \mathbf{A}^T\mathbf{d} \tag{4.9}$$

Given the migrated image $\mathbf{m}_{mig} = \mathbf{A}^T\mathbf{d}$, and $\mathbf{m}_{remig} = \mathbf{A}^T\mathbf{A}\mathbf{m}_{mig}$, or even better $\mathbf{m}_{filt} = \mathbf{L}^{-\frac{n-1}{2}}\mathbf{m}_{remig}$.

Restating (4.9)

$$\mathbf{W}^2\mathbf{m}_{filt} \approx \mathbf{m}_{mig}. \tag{4.10}$$

The notation $\mathbf{W}^2$ stresses the fact that the scaling operator is positive definite or

at least semi-definite, which then requires its symbol to be positive or at least non-negative; moreover the filtering step ensures that $\mathbf{W}^2$ is of order zero.

The aim is to estimate $\mathbf{m}_{true} \approx \mathbf{m}_{inv}$, where

$$\mathbf{m}_{inv} = \mathbf{W}^2 \mathbf{L}^{-\frac{n-1}{2}} \mathbf{m}_{mig}. \tag{4.11}$$

If $\mathbf{W}^2$ is known, the last step requires a direct application of the algorithm described above. The real problem lies in determining $\mathbf{W}^2$, more precisely its symbol $\tilde{q}_m(x, z, \theta)$, here $m = 0$. I carry m along in the rest of this discussion because a variant of this discussion can skip the explicit filtering step and use a pseudodifferential scaling factor $\mathbf{W}^2$ of order -1.

The algorithm defined above for fixed input $u = \mathbf{m}_{filt}$ and known output $Q_m u = \mathbf{m}_{mig}$ may be regarded as a function of the symbol $\tilde{q}$,

$$Q_m[\tilde{q}]\mathbf{m}_{filt} \approx \mathbf{m}_{mig}. \tag{4.12}$$

The principal symbol may be recovered by an optimization scheme

$$q_m(x, z, \theta) = \underset{\tilde{q} \geq 0}{\operatorname{argmin}} \|Q_m[\tilde{q}]\mathbf{m}_{filt} - \mathbf{m}_{mig}\|^2. \tag{4.13}$$

The feasible set consists of non-negative symbols because $\mathbf{W}^2$ is symmetric positive semidefinite. Additional regularization terms (e.g., Tichonov regularization) are added to (4.13) if needed.

The problem admits multiple minimizers since the system is underdetermined. The degree of underdeterminism may be seen more transparently from

$$Q_m u \approx \sum_{l=-K/2}^{l=K/2} c_l(x, z) \mathcal{F}^{-1}[\omega^{m-l}(\xi + i\eta)^l \hat{u}(\xi, \eta)].$$

For each (x, z) the system consists of one equation in $K + 1$ unknowns. The requirement that $\tilde{q}$ be slowly varying in angle, is enforced by limiting K directly.

The system is rendered determined by enforcing the continuity of $\tilde{q}$ using a parsimonious basis technique. Let $\{\psi_j(x, z)\}_{j=1}^{J}$ be a set of smooth shape functions (cubic b-splines for example). Write

$$c_l(x, z) = \sum_{j=1}^{J} c_l^j \psi_j(x, z) \Rightarrow \tilde{q}_m(x, z, \theta) = \sum_{l=-K/2}^{K/2} \sum_{j=1}^{J} c_l^j \psi_j(x, z) e^{il\theta}. \tag{4.14}$$

The system now consists of N^2 equations in $(K + 1)J$ unknowns.

Finally, enforcing positivity of the symbol may be achieved by letting $\tilde{q}(x, z, \theta) = q(x, z, \theta)^2$ with $q(x, z, \theta)$ given by (4.14):

$$q(x, z, \theta) = \sum_{l=-K/2}^{K/2} c_l(x, z) e^{il\theta} \Rightarrow q^2(x, z, \theta) = \sum_{l=-K}^{K} \sum_{n=-K/2}^{l+K/2} c_{l-n}(x, z) c_n(x, z) e^{il\theta}, \tag{4.15}$$

where $c_l(x, z) \equiv 0$ when $l \notin [-K/2, K/2]$.

Notice that

$$a_l(x, z) := \sum_{n=-K/2}^{l+K/2} c_{l-n}(x, z) c_n(x, z) = (c_n * c_n)_l(x, z), \tag{4.16}$$

i.e., the convolution of the vector of Fourier coefficients with itself, and the sum is implemented as such. Really, (4.15) is nothing but the convolution theorem for Fourier transforms in its discrete form. The coefficients $c_l(x, z)$ are given by (4.14). This approach enforces the positivity of the symbol directly.

4.3.1 Gradient Calculation

In this section, the dependence of the symbol on its Fourier coefficients needs to be written explicitly:

$$q_m(x,z,\theta) = \sum_{l=-K}^{K} c_l'(x,z)e^{il\theta}. \tag{4.17}$$

I choose bi-cubic splines for the smooth shape function $\psi_j(x,z) = B_i(x)B_j(z)$ referenced in (4.14), with this bi-cubic splines choice $c_l(x,z)$ given by

$$c_l(x,z) = \sum_{i,j} c_l^{ij} B_i(x)B_j(z). \tag{4.18}$$

Equation (4.18) gives the explicit dependence of the Fourier coefficients on the bi-cubic basis functions. Finally, $c_l'(x,z)$ is the auto-convolution of $c_l(x,z)$,

$$c_l'(x,z) = \sum_{n=-K/2}^{l+K/2} c_{l-n}(x,z)c_n(x,z). \tag{4.19}$$

Recall that the action of a pseudodifferential operator is given by

$$Q_m u(x,z) = \sum_{l=-K}^{K} c_l'(x,z)\mathcal{F}^{-1}\{\omega^{m-l}(\xi+i\eta)^l \hat{u}(\xi,\eta)\}. \tag{4.20}$$

The objective function with the norm interpreted in the least squares sense is regarded as a function of the coefficients c_l^{ij},

$$J(c_l^{ij}) = \|Q_m u(x,z) - d(x,z)\|^2 = \sum_{x_i,z_j} |Q_m u(x_i,z_j) - d(x_i,z_j)|^2. \tag{4.21}$$

The gradient of the objective function follows from (4.21):

$$\frac{\partial J}{\partial c_{l'}^{i'j'}} = 2\sum_{x_i,z_j} \mathsf{Re}\left\{ \frac{\partial Q_m u}{\partial c_{l'}^{i'j'}} (Q_m u - d)^* \right\}. \tag{4.22}$$

Thus the important element in this calculation is $\frac{\partial Q_m u}{\partial c_{l'}^{i'j'}}$:

$$\frac{\partial Q_m u}{\partial c_{l'}^{i'j'}} = \frac{\partial}{\partial c_{l'}^{i'j'}} \left\{ \sum_{l=-K}^{K} \sum_{n=-K/2}^{l+K/2} \sum_{i,j} c_{l-n}^{ij} B_i(x) B_j(z) . \sum_{i,j} c_n^{ij} B_i(x) B_j(z) . \mathcal{F}^{-1}[\dots] \right\}, \tag{4.23}$$

since $\mathcal{F}^{-1}[\omega^{m-l}(\xi + i\eta)^l \hat{u}(\xi, \eta)]$ does not depend on c_l^{ij}, it will be carried through the calculation and denoted by $\mathcal{F}^{-1}[\dots]$.

Expanding (4.23),

$$\begin{aligned}
\frac{\partial Q_m u}{\partial c_{l'}^{i'j'}} &= \sum_{l=-K}^{K} \sum_{n=-K/2}^{l+K/2} \sum_{i,j} \delta_i^{i'} \delta_j^{j'} \delta_{l-n}^{l'} B_i(x) B_j(z) . \sum_{i,j} c_n^{ij} B_i(x) B_j(z) . \mathcal{F}^{-1}[\dots] \\
&\quad + \sum_{l=-K}^{K} \sum_{n=-K/2}^{l+K/2} \sum_{i,j} c_{l-n}^{ij} B_i(x) B_j(z) . \sum_{i,j} \delta_i^{i'} \delta_j^{j'} \delta_n^{l'} B_i(x) B_j(z) . \mathcal{F}^{-1}[\dots] \\
&= 2 B_{i'}(x) B_{j'}(z) \sum_{l=-K}^{K} \sum_{i,j} c_{l-l'}^{ij} B_i(x) B_j(z) . \mathcal{F}^{-1}[\dots] \\
&= 2 B_{i'}(x) B_{j'}(z) \sum_{l=-K}^{K} c_{l-l'}(x, z) \mathcal{F}^{-1}[\omega^{m-l}(\xi + i\eta)^l \hat{u}(\xi, \eta)].
\end{aligned}$$

This last result is worth restating:

$$\frac{\partial Q_m u}{\partial c_{l'}^{i'j'}} = 2 B_{i'}(x) B_{j'}(z) \sum_{l=-K}^{K} c_{l-l'}(x, z) \mathcal{F}^{-1}[\omega^{m-l}(\xi + i\eta)^l \hat{u}(\xi, \eta)], \tag{4.24}$$

because it has a natural interpretation: the calculation of the gradient of $Q_m u$ requires K applications of the Bao and Symes (1996) algorithm each time with shifted coefficients $c_{l-l'(x,z)}$ and evaluation of the splines at fixed points.

4.4 Description Of The Code

This section discusses the specific implementation of the scaling method and the choice I adopted to enforce the positivity on the symbol.

The symbol in the Bao and Symes (1996) algorithm is naturally parametrized by its Fourier coefficients and thus any property of the symbol translates to a property of its Fourier coefficients, specifying its implementation.

The first requirement that the implementation must satisfy is the continuity and smoothness of the symbol. A parametrization of the Fourier coefficients of the symbol in terms of the coefficients of a bi-cubic spline expansion ensures the smoothness of the symbol on the domain (parsimonious basis technique):

$$c_l(x,z) = \sum_{i,j} c_l^{i,j} B_i(x) B_j(z).$$

It turns out that the positivity requirement on the symbol is the most intricate of all and to solve this, we choose the symbol to be the square of a real smooth function (its square root). Restating the positivity requirement in terms of the Fourier coefficients amounts to implementing the symbol as an auto-convolution of the Fourier coefficients of the real square root. Reality is enforced on the Fourier coefficients by requiring conjugate symmetry. To summarize the steps:

1. Provide c_l for $l \geq 0$.

2. Symmetrize, $c_{-l} = \bar{c}_l$, to ensure that q is real.

3. Autoconvolve, $q = \sum_l c_l * c_l e^{il\theta}$ to ensure positivity.

The code consists of the algorithm for application of a pseudodifferential operator `PsiDO` described extensively in the first section. I implement the objective function and the gradient calculation in `objective.m`. `QmakSq` ensures the positivity of the symbol emphasized earlier, by constructing it as the square of a real function (symmetric Fourier coefficients).

A walk through `objective.m` describes the entirety of the code:

- The input consists of $\mathbf{m}_{filt}$ and $\mathbf{m}_{mig}$ and the coefficients of the b-splines.
- The coefficients yield a positive symbol in $q =$ `QmakSq`.
- The objective function (4.21) is computed, with q as input to `PsiDO`.
- The gradient (4.22) uses (4.24).

The implementation accepts complex coefficients, parametrized by their real and imaginary parts. Only half of the coefficients are supplied by the user, since the code internally symmetrizes the coefficients in `QmakSq` to ensure reality before autoconvolving them to enforce positivity.

As a result of this specific parametrization $\frac{\partial Q_m u}{\partial c_{l'}^{i'j'}}$ must pick up a few more terms:

$$\frac{\partial Q_m u}{\partial Re(c_{l'}^{i'j'})} = 2B_{i'}(x)B_{j'}(z)\sum_{l=-K}^{K}(c_{l-l'}+c_{l+l'})(x,z)\mathcal{F}^{-1}[\omega^{m-l}(\xi+i\eta)^l\hat{u}(\xi,\eta)], \quad (4.25)$$

$$\frac{\partial Q_m u}{\partial Im(c_{l'}^{i'j'})} = 2B_{i'}(x)B_{j'}(z)\sum_{l=-K}^{K}(ic_{l-l'}-ic_{l+l'})(x,z)\mathcal{F}^{-1}[\omega^{m-l}(\xi+i\eta)^l\hat{u}(\xi,\eta)]. \quad (4.26)$$

The primary results of the method suggested that only the even Fourier modes should be implemented to obtain a real inverted image. This observation motivated the following lemma.

Lemma 4.4.1 *If u is real , $c_{-l} = c_l^*$, and $c_{2l+1} = 0$ for all l, then $Q_m u$, as defined by (4.6), is real.*

The proof of the continuous case is easier than the proof on the discrete algorithm and follows from the definition (4.3) and arguments presented here. Recall that:

$$Q_m u \approx \sum_{l=-K/2}^{l=K/2} c_l(x,z)\mathcal{F}^{-1}[\omega^{m-l}(\xi+i\eta)^l \hat{u}(\xi,\eta)], \tag{4.27}$$

since the reality of u is equivalent to $\hat{u}(-\xi,-\eta) = \hat{u}(\xi,\eta)^*$ (conjugate symmetry). Now if $c_{2l+1} = 0$ we need to only consider the even terms:

$$\begin{aligned} c_{-2l}\mathcal{F}^{-1}[\omega^{m+2l}(\xi+i\eta)^{-2l}\hat{u}(\xi,\eta)] + c_{2l}\mathcal{F}^{-1}[\omega^{m-2l}(\xi+i\eta)^{2l}\hat{u}(\xi,\eta)] = \\ \mathcal{F}^{-1}[c_{2l}\omega^{m-2l}(\xi+i\eta)^{2l}\hat{u}(\xi,\eta) + c_{2l}^*\omega^{m+2l}(\xi+i\eta)^{-2l}\hat{u}(\xi,\eta)]. \end{aligned} \tag{4.28}$$

Then let $\hat{f}(\xi,\eta) = c_{2l}\omega^{m-2l}(\xi+i\eta)^{2l}\hat{u}(\xi,\eta) + c_{2l}^*\omega^{m+2l}(\xi+i\eta)^{-2l}\hat{u}(\xi,\eta)$.

Then,

$$\begin{aligned} \hat{f}(-\xi,-\eta) &= c_{2l}\omega^{m-2l}(-\xi-i\eta)^{2l}\hat{u}(-\xi,-\eta) + c_{2l}^*\omega^{m+2l}(-\xi-i\eta)^{-2l}\hat{u}(-\xi,-\eta) \\ &= c_{2l}\omega^{m-2l}(\xi+i\eta)^{2l}\hat{u}(\xi,\eta)^* + c_{2l}^*\omega^{m+2l}(\xi+i\eta)^{-2l}\hat{u}(\xi,\eta)^* \\ &= c_{2l}\omega^m e^{2il\theta}\hat{u}(\xi,\eta)^* + c_{2l}^*\omega^m e^{-2il\theta}\hat{u}(\xi,\eta)^* \\ &= \left[c_{2l}\omega^m e^{2il\theta}\hat{u}(\xi,\eta) + c_{2l}^*\omega^m e^{-2il\theta}\hat{u}(\xi,\eta)\right]^* \\ &= \hat{f}(\xi,\eta)^* \end{aligned} \tag{4.29}$$

Hence, by the properties of the Fourier transform, $\mathcal{F}^{-1}[\hat{f}]$ is real and therefore $Q_m u$ is real. This implies that $\mathbf{m}_{inv} = Q_m[q^\dagger]\mathbf{m}_{mig}$ is real since $\mathbf{m}_{mig}$ is real.

This property allows us to cut the search space in half and only implement symbols with even Fourier coefficients.

Chapter 5

Results

5.1 Introduction

This section investigates the accuracy of the proposed method. Results on the 2D Marmousi benchmark model (Versteeg and Grau, 1991) corroborate the validity of the method in approximating the real model. The intrinsic difference between $K = 1$ and $K > 1$ enables the latter to resolve multiple dip events and hinders the former from doing so; in fact $K = 1$ reduces to the Symes (2008) method. The Marmousi model is incapable of revealing the difference between the two cases, since the difference between the result for $K = 1$ and $K = 5$ is marginal, mainly because multiple dip events are limited in this model or the dependence of the symbol of the normal operator on dip is weak. To test the ability of the method to resolve multiple dip events, I construct a "plaid" model consisting of multiple dip events in most areas of

the model. The results on the "plaid"' example reveal the superiority of $K > 1$ while giving a lot of insight into the behavior of the method.

The "plaid" data example uses the Bao and Symes (1996) algorithm to generate the data, whereas the Marmousi *migrated image* $\mathbf{m}_{mig} = \mathbf{A}^T\mathbf{d}$ and the *remigrated image* $\mathbf{m}_{remig} = \mathbf{A}^T\mathbf{A}\mathbf{m}_{mig}$ are computed using finite difference wave equation solvers. In terms of the inverse problem setting, the Marmousi simulation and inversion algorithms are separate whereas the simulation algorithm is the same as the inversion algorithm for the "plaid" data example. The "plaid" example is an "inverse crime" where the same method is used to generate data and to invert. A wave equation example similar to the "plaid" data is coming soon, and it would be interesting to investigate the results in that case, too.

5.2 Results

5.2.1 Marmousi

The 2D Marmousi benchmark model (Versteeg and Grau, 1991) is a synthetic example that retains some of the same challenges of real data. The model is smoothed and a perturbation to the model $\mathbf{m}_{true}$ (Figure 5.1) is obtained as the difference between the full model and the smoothed model. The model is windowed and tapered to the window of interest. The migrated and remigrated images (Figures 5.2 and 5.3) reveal the distortion in the amplitudes of the image when compared to the real

model (Figure 5.1) after the application of the migration operator, even worse after application of the normal operator. The amplitudes differ by orders of magnitude and the distortion is nonuniform in depth, in fact it tends to attenuate the amplitudes in the deeper parts of the image (the regions of interest) to the point where events become invisible. Amplitude correction becomes therefore a mandatory procedure. The pseudodifferential scaling method with $K = 1$ and $K = 5$ (see Figures 5.4 and 5.5) corrects the amplitudes. At first glance the amplitudes are recovered to the right order of magnitude. More importantly, the correction reinstates the events in the deeper part of the image that were hidden in the migrated image from the amplitude distortion. Both results are satisfactory for this example. Note that the scaling results are shown on the window of interest rather than the full model for emphasis.

The difference between scaling with $K = 5$ (Fig 5.5) and scaling with $K = 1$ (Fig 5.4) is displayed in (Fig 5.6). Although the difference is small in most parts of the image, it is apparent that the largest amplitude differences between the two scaling methods occur exactly at the areas of the image where multiple dip events are. The amplitude difference is largest at places where two reflectors intersect (faults) and the intersection points tend to show the largest differences in amplitudes (either brighter or dimmer). The scaling with $K = 5$ in fact performs marginally better than $K = 1$. However, the Marmousi example is not the most suitable to divulge the difference between the $K = 1$ and $K > 1$ cases, mainly because multiple dip events are not "abundant" in this model. A model that investigates the difference between scaling

with $K = 1$ and $K > 1$ constitutes the subject of the next section.

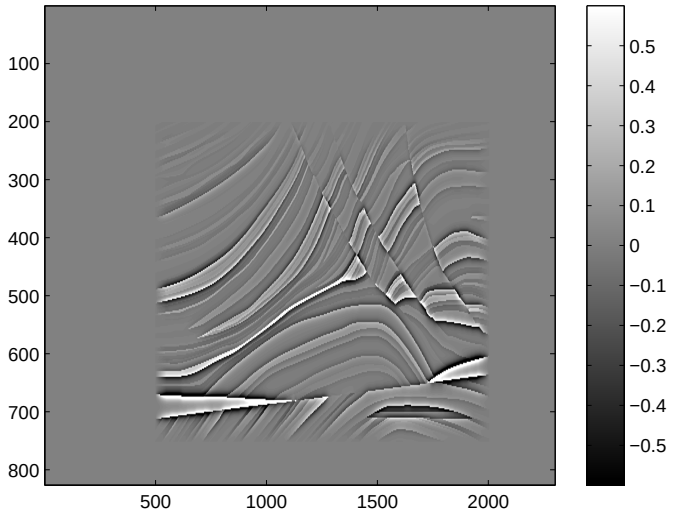

Figure 5.1: $\mathbf{m}_{true}$

5.2.2 Plaid data

One important feature of the method is the intrinsic difference between $K = 1$ and $K > 1$ since the method will only be able to capture the dependence of model or symbol on dip if $K > 1$. This feature is not stressed in the Marmousi example since the model has well defined dip almost everywhere and thus the method did not fail for $K = 1$. To test this feature, I create a set of "plaid" data $\mathbf{b}$ (Fig 5.7) where multiple dip events are abundant. I apply a pseudodifferential $\mathbf{A} = Q[q]$ operator with symbol $q = \cos(\theta)^2$ twice thus creating $\mathbf{Ab}$ (Fig 5.8) and $\mathbf{A}^2\mathbf{b}$ (Fig 5.9). A scale factor is fit between $\mathbf{Ab}$ and $\mathbf{A}^2\mathbf{b}$; the scale factor should be an approximation of $\mathbf{A}$

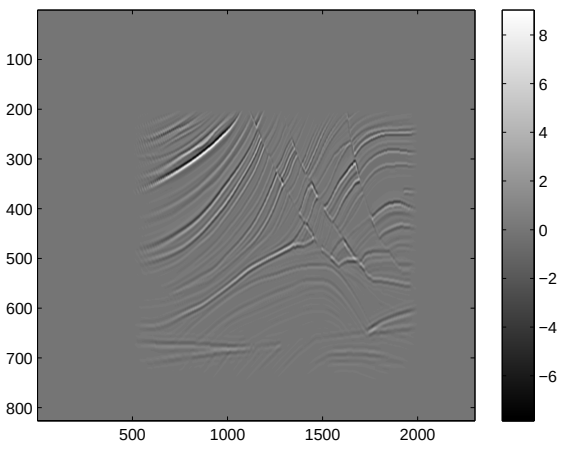

Figure 5.2: $\mathbf{m}_{mig} = A^T d$

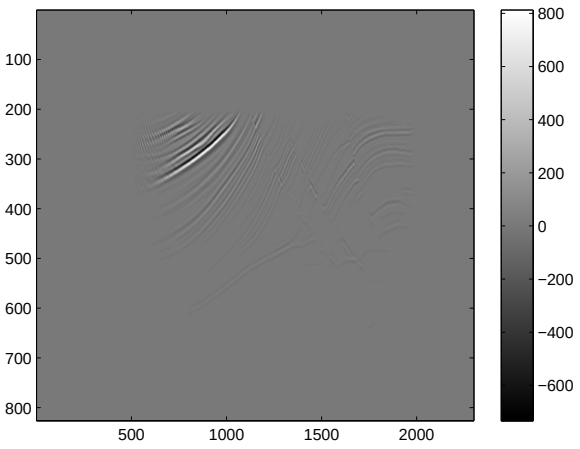

Figure 5.3: $\mathbf{m}_{remig} = A^T A \mathbf{m}_{mig}$

($Q[\tilde{q}] \approx \mathbf{A}$) which is then applied to $\mathbf{b}$ in hope of recovering $\mathbf{Ab}$. I thus call $\mathbf{Ab}$ the "true image" and $Q[\tilde{q}]\mathbf{b}$ the "inverted image".

The "inversion" results are displayed in for $K = 1$ in Fig 5.11 and for $K = 5$ in Fig 5.13. The result for $K = 1$ already shows a weakness in the resolution of the amplitudes in different parts of the image whereas the result for $K = 5$ is more successful in approximating the true image. A more insightful investigation on the errors (misfit between the inverted and true image) in the two cases reveals the true story. The error for $K = 1$ (Fig 5.14) is highly non-isotropic. It shows the same structure of the image in one of the directions. $K = 1$ struggles with the multiple dip events and can only resolve one direction. In comparison, the error for $K = 5$ (Fig 5.15) is isotropic: it shows no structure from the image. In fact the method resolves multiple dip events and altogether fits the image to less than 0.5% root mean square relative error.

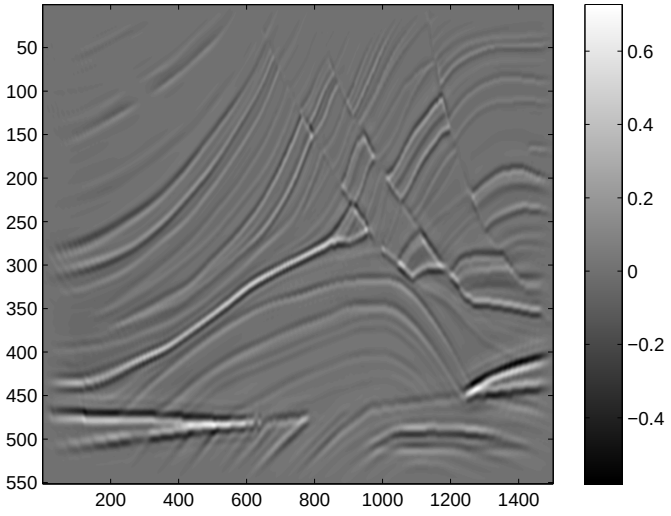

Figure 5.4: Scaling with $K = 1$

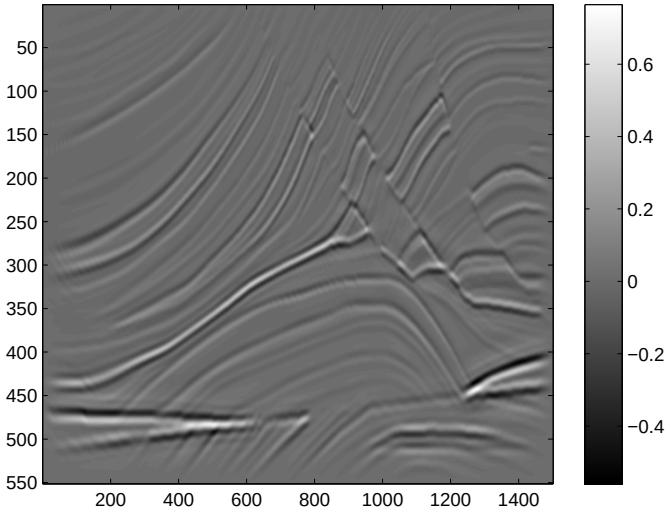

Figure 5.5: Scaling with $K = 5$

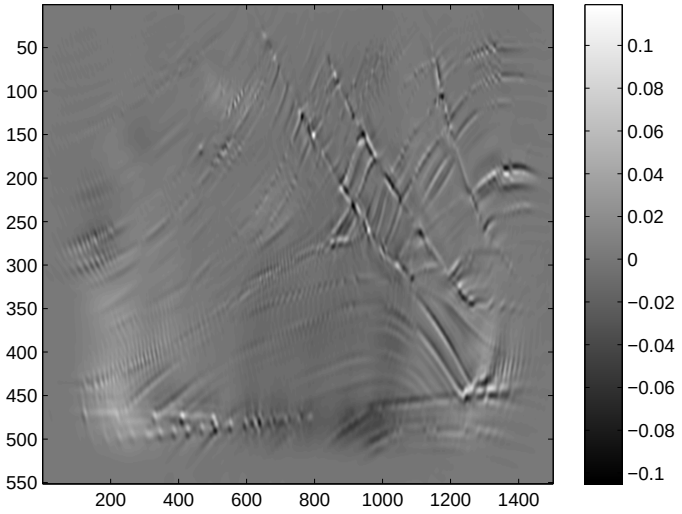

Figure 5.6: Difference between scaling with $K = 5$ and $K = 1$

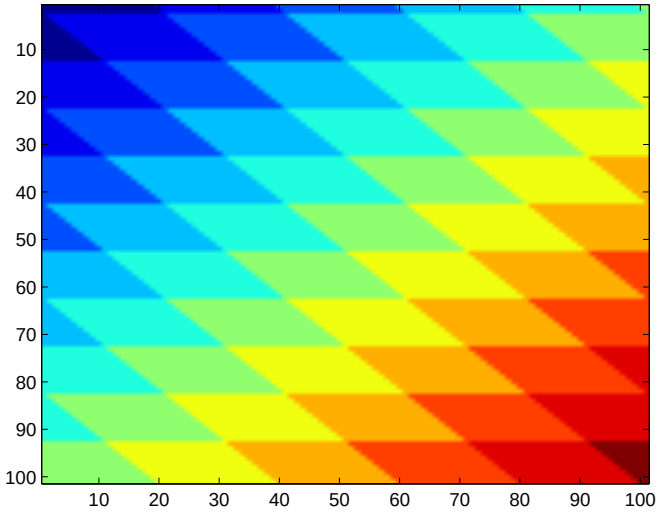

Figure 5.7: Plaid Data = **b**

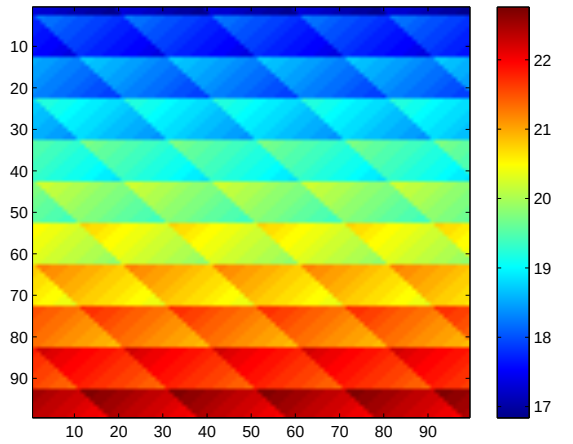

Figure 5.8: $\mathbf{Ab} = Q[\cos^2(\theta)]\mathbf{b}$

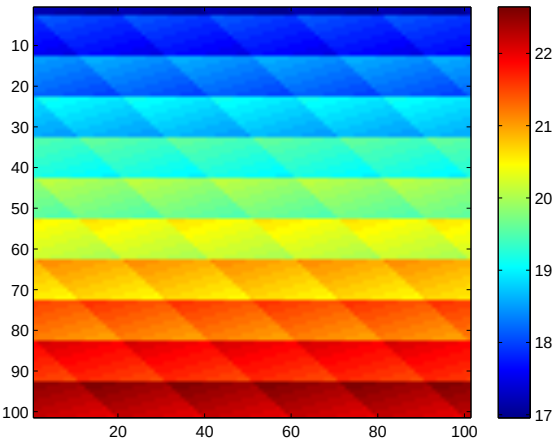

Figure 5.9: $\mathbf{A}^2\mathbf{b} = Q[\cos^2(\theta)]\mathbf{Ab}$

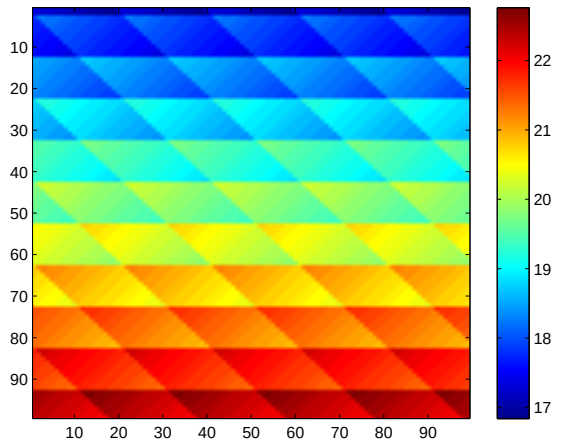

Figure 5.10: True Image: $\mathbf{Ab}$

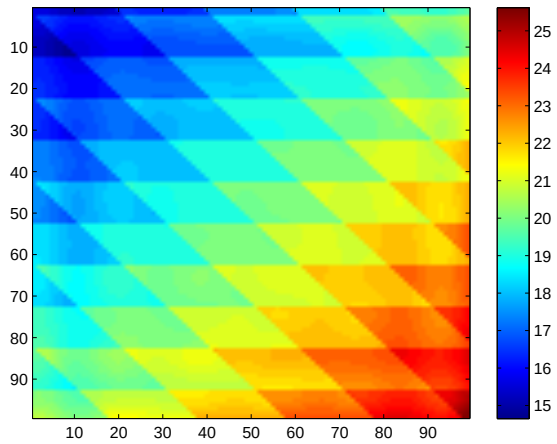

Figure 5.11: Inverted Image $K = 1$: $Q[\tilde{q}]\mathbf{b}$

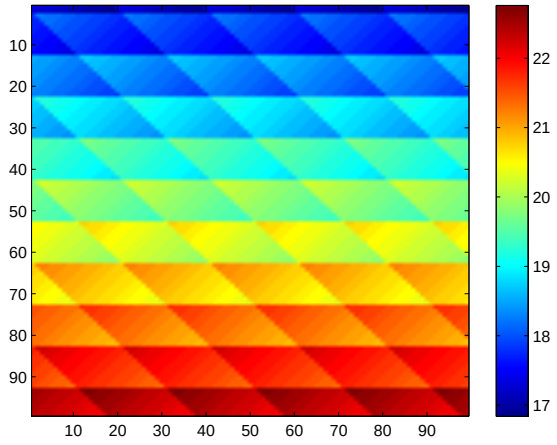

Figure 5.12: True Image: $\mathbf{Ab}$

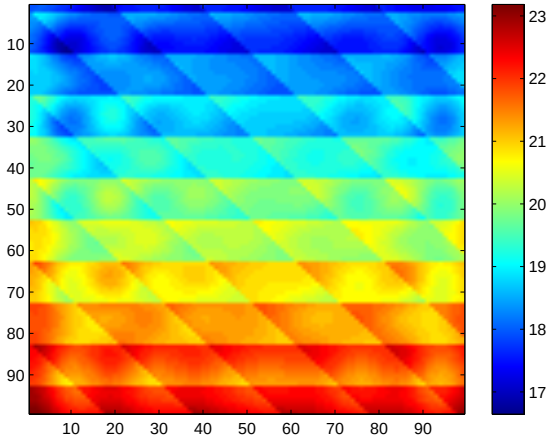

Figure 5.13: Inverted Image $K = 5$: $Q[\tilde{q}]\mathbf{b}$

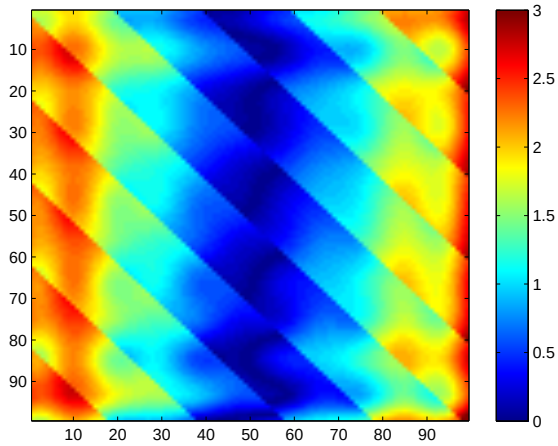

Figure 5.14: Error for $K = 1$

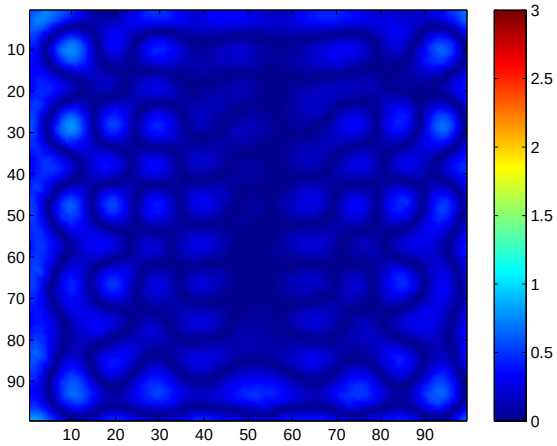

Figure 5.15: Error for $K = 5$

Chapter 6

Conclusions

This thesis leverages the near diagonality of the normal operator in a basis of localized monochromatic pulses to develop a pseudodifferential scaling method. The recovery of the scaling factor is cast as an optimization scheme that scales the migrated image, i.e., $\mathbf{A}^T\mathbf{d}$, to the remigrated image, i.e., $(\mathbf{A}^T\mathbf{A})\mathbf{A}^T\mathbf{d}$, and recovers a positive pseudodifferential scaling factor, i.e., the approximate pseudoinverse of the normal operator of the right order (predicted by the theory). The procedure is efficient since it uses the Bao and Symes (1996) algorithm for the action of a pseudodifferential operator.

Tests on the Marmousi benchmark model validate the method: the amplitudes of the inverted image resemble those of the real image and the distortions from the application of the migration operator become less prevalent. An example in which multiple dip events are abundant highlights the ability of the method to resolve these events.

The method may be used to correct the amplitudes of seismic images at the cost of one resimulation (application of $\mathbf{A}$) and one remigration (application of $\mathbf{A}^T$). These procedures tend to be extremely expensive and thus overwhelm the cost of applying the scaling method. The need for these extra simulation and migration operations is not unreasonable, since iterative least squares inversion requires such applications at each iteration. The method's ability to precondition least squares inversion and accelerate its convergence when the background model is not a good approximation to the real model still needs to be tested. Herrmann has carried out a version of this program (Herrmann et al., 2008a) for the linearized inversion using the results of Herrmann et al. (2008b). However, the results on the Marmousi data set suggest that the output of the method is a satisfactory inversion result and least squares inversion may not be even needed when the background model is an appropriate smooth approximation to the real model.

This method cannot escape the sensitive dependence on the background model and the quality of the migrated and remigrated images. These parameters are more controllable in synthetic examples. The real challenge lies in the application of the method to real data, where the background model is a priori unknown and the quality of the images is, at best, hard to evaluate.

The setting of the method lends itself to generalizations to three dimensions as well as to the multi-parameter case (variable density acoustics, for example). These extensions present themselves as natural directions to pursue; in fact, preliminary

results on the variable density acoustics extension are promising.

Bibliography

Bao, G. and Symes, W. (1996). Computation of pseudo-differential operators. *SIAM J. Sci. Comput.*, 17(2):416–429.

Beylkin, G. (1985). Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *Journal of Mathematical Physics*, 26:99–108.

Candes, E. and Demanet, L. (2005). The curvelet representation of wave propagators is optimally sparse. *Communications on Pure and Applied Mathematics*, 58:1472–1528.

Chavent, G. and Plessix, R.-E. (1999). An optimal true-amplitude least-squares prestack depth-migration operator. *Geophysics*, 64:508–515.

Claerbout, J. and Nichols, D. (1994). Spectral preconditioning. Technical Report 82, Stanford Exploration Project, Stanford University, Stanford, California, USA.

Clément, F. and Chavent, G. (1993). Waveform inversion through MBTT formulation. In Kleinman, E., Angell, T., Colton, D., Santosa, F., and Stakgold, I., editors,

Mathematical and Numerical Aspects of Wave Propagation. Society for Industrial and Applied Mathematics, Philadelphia.

Guitton, A. (2004). Amplitude and kinematic corrections of migrated images for nonunitary imaging operators. *Geophysics*, 69:1017–1024.

Herrmann, F., Brown, C., Erlangga, Y., and Moghaddam, P. (2008a). Curvelet-based migration preconditioning. Technical Report 7, The University of British Columbia.

Herrmann, F., Moghaddam, P., and Stolk, C. (2008b). Sparsity- and continuity-promoting seismic image recovery with curvelet frames. *Applied and Computational Harmonic Analysis*, 24:150–173.

Nemeth, T., Wu, C., and Schuster, G. (1999). Least-squares migration of incomplete reflection data. *Geophysics*, 64:208–221.

Rakesh (1988). A linearized inverse problem for the wave equation. *Communications on Partial Differential Equations*, 13(5):573–601.

Rickett, J. E. (2003). Illumination-based normalization for wave-equation depth migration. *Geophysics*, 68:1371–1379.

Shin, C., Jang, S., and Min, D.-J. (2001). Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, 49:592–606.

Stolk, C. (2000). *On the modeling and inversion of seismic data*. PhD thesis, Universiteit Utrecht.

Symes, W. W. (2008). Approximate linearized inversion by optimal scaling of prestack depth migration. *Geophysics*, 73:R23–R35.

Taylor, M. (1981). *Pseudodifferential Operators*. Princeton University Press, Princeton, New Jersey.

Valenciano, A., Biondi, B., and Guitton, A. (2006). Target-oriented wave-equation inversion. *Geophysics*, 71:A35–A38.

Versteeg, R. and Grau, G. (1991). Practical aspects of inversion: The Marmousi experience. Proceedings of the EAEG, The Hague.