RICE UNIVERSITY

Approximate Multi-Parameter Inverse Scattering Using Pseudodifferential Scaling
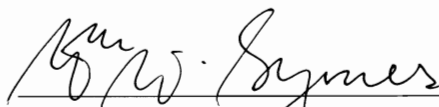
by

Rami Nammour

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
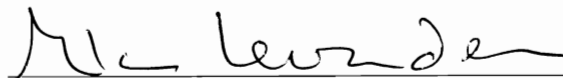REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

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

Dr. Liliana Borcea
Professor of Computational and Applied
Mathematics

Dr. Alan Levander
Professor of Geology and Geophysics

Dr. Mark Embree
Professor of Computational and Applied
Mathematics

HOUSTON, TEXAS

MARCH, 2011

Abstract

Approximate Multi-Parameter Inverse Scattering Using Pseudodifferential Scaling

by

Rami Nammour

I propose a computationally efficient method to approximate the inverse of the normal operator arising in the multi-parameter linearized inverse problem for reflection seismology in two and three spatial dimensions.

Solving the inverse problem using direct matrix methods like Gaussian elimination is computationally infeasible. In fact, the application of the normal operator requires solving large scale PDE problems. However, under certain conditions, the normal operator is a matrix of pseudodifferential operators. This manuscript shows how to generalize Cramer's rule for matrices to approximate the inverse of a matrix of pseudodifferential operators. Approximating the solution to the normal equations proceeds in two steps:

- First, a series of applications of the normal operator to specific permutations of the right hand side. This step yields a phase-space scaling of the solution.

Phase space scalings are scalings in both physical space and Fourier space.

- Second, a correction for the phase space scaling. This step requires applying the normal operator once more.

The cost of approximating the inverse is a few applications of the normal operator (one for one parameter, two for two parameters, six for three parameters).

The approximate inverse is an adequately accurate solution to the linearized inverse problem when it is capable of fitting the data to a prescribed precision. Otherwise, the approximate inverse of the normal operator might be used to precondition Krylov subspace methods in order to refine the data fit.

I validate the method on a linearized version of the Marmousi model for constant density acoustics for the one-parameter problem. For the two parameter problem, the inversion of a variable density acoustics layered model corroborates the success of the proposed method. Furthermore, this example details the various steps of the method. I also apply the method to a 1D section of the Marmousi model to test the behavior of the method on complex two-parameter layered models.

Acknowledgements

First and foremost, I am indebted to my academic advisor Dr William Symes. For four years, he balanced careful and close mentorship, and left enough space for me to become an independent thinker: a meticulous advisor! Not to mention his effort to make this manuscript readable.

I am forever grateful to my thesis committee, Dr Liliana Borcea, Dr Mark Embree and Dr Alan Levander for overlooking this work, their contribution enriches my academic career. I am also grateful to the CAAM department for their academic support, and the sense of community that this department provides.

I also would like to thank Dr Eric Dussaud, Dr Fuchun Gao, Dr Paul Williamson and Dr Uwe Albertin for their mentorship during my summer internships where I developed invaluable professional experience.

To my friends, I owe a huge thank you for guiding me through the initiation rituals into CAAM and Houston and making my transition and my stay here $C^{\infty}([2006, 2011])$. A special thank you for the people who shared the road with me these five years, and stayed with me through all the unexpected turns.

To my family in Lebanon: without your transatlantic support, this work would not be possible.

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

This Doctorate dissertation is dedicated to my mother, her love wills me to surpass myself every day.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this thesis, I review a method to approximately solve the linearized inverse problem of constant density acoustics, work I have accomplished as part of my masters thesis. I propose a generalization for variable density acoustics and other multi-parameter inverse problems. Finally, I propose how the method to solve the linearized inverse problem accelerates the convergence of iterative methods aimed at solving the full nonlinear inverse problem (referred to as full waveform inversion methods). Previous work was limited to 2D (two spatial dimensions); however, I also propose extensions to 3D and discuss the similarities between the two approaches and the challenges specific to 3D wave propagation.

The model problem for this work is the variable density acoustic wave equation, the simplest model describing the reaction of the earth to acoustic excitation (explosions,

air gun ...):

$$\frac{1}{\rho(x)c^2(x)}\frac{\partial^2 p}{\partial t^2}(x,t) - \nabla \cdot \frac{1}{\rho(x)}\nabla p(x,t) = f(x,t), \tag{1.1}$$

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; $f(x,t)$ represents the source of acoustic energy. Note that the formulation is the same in 2D and 3D.

Assuming the earth was at equilibrium before the forcing is put to effect (causal source), the pressure field satisfies:

$$p(x,t) \equiv 0, \quad t \ll 0$$
$$\tag{1.2}$$
$$f(x,t) \equiv 0, \quad t \ll 0.$$

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

It is advantageous to think of equation (1.1) as defining a map that associations the earth properties (density and velocity fields), to the measurements of the pressure at the surface. Group density and velocity parameters together to form the parameter $m = [\rho, c]$. The map I have in mind is:

$$S[m] = p|_{surface}. \tag{1.3}$$

I refer to S as the nonlinear forward map, it maps the model to the measurement of the pressure at the surface $p|_{surface}$. The pressure at the surface is measured using a recording device, like geophones or hydrophones.

The advantage of this abstraction is that it describes any model of the earth in the same way. The generalization to elasticity, for example. falls under the same rubric,

with m standing for the elastic coefficients and density. The methods proposed in this thesis often extend to models of seismic wave propagation other than variable density acoustics (1.1). The behavior of these methods depends not so much on the specific equations chosen to model seismic waves, but rather on properties of the forward map shared by various models.

The inverse problem aims at solving for the model parameters, given the measurements of the pressure at the surface: Given d, solve for m such that $S[m] = d$. Note that though the wave equation itself is linear, the dependence of the solution on the model parameters is nonlinear. The inverse problem is therefore nonlinear.

The linearization of the inverse problem assumes a natural splitting of the model parameters into a background m_0 and a perturbation δm,

$$m = m_0 + \delta m. \tag{1.4}$$

The background is given and it is required to solve for the the perturbation δm. The formal derivative of the nonlinear forward map at m_0 maps the perturbation δm to δp:

$$F[m_0]\delta m = \delta p. \tag{1.5}$$

The linearized forward map F is referred to as Born Modeling. The linearized forward problem is: given δd, find δm such that $F\delta m = \delta d$. I discuss the details of the linearized inverse problem in the chapter devoted to linearization.

The acoustic wave equation models the dependence of the response of the earth on two parameters (density and velocity). One simplification assumes a constant

density field, constant density acoustics. The constant density approximation is valid when the acoustic waves are reflected from a region that exhibits negligible density contrast relative to velocity contrasts. Allowing for multi-parameter descriptions of the earth models allows for a better description of the underlying physics. However, it comes at the expense of introducing ill-conditioning in the inverse problem, thus making it more difficult to solve accurately. This manuscript begins by describing one parameter inversions and then multi-parameter extensions.

Finally, an approximate solver for the linearized inverse problem may yield one way to accelerate the convergence of iterative methods for the nonlinear inverse problem (referred to as full waveform inversion methods or FWI in short). I explain this in the section on full waveform inversion.

I summarize the structure of this thesis in the following diagram.
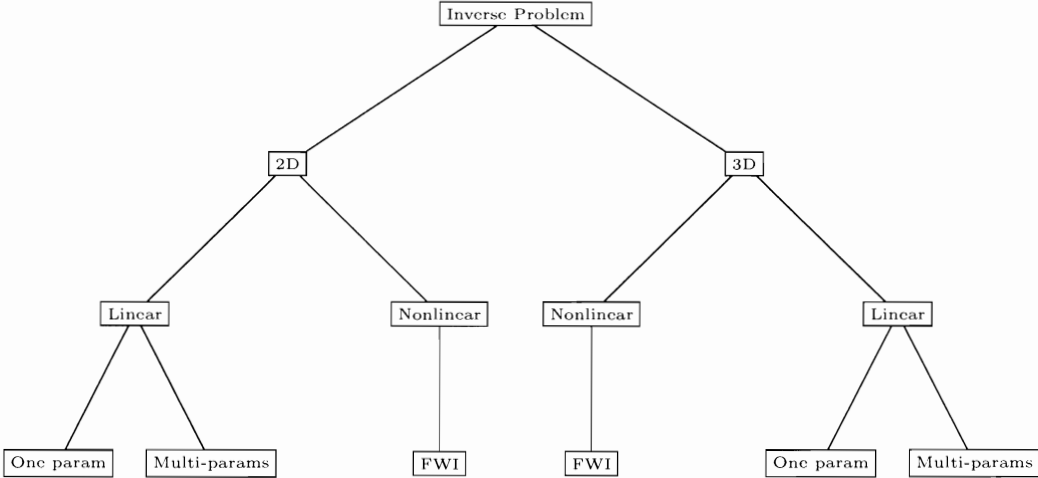
1.1 Organization of the Thesis

This thesis comprises six chapters. The first chapter is a general introduction. A brief review of the theory of the linearized inverse problem and the existing literature constitute the second chapter. The third chapter introduces the proposed method. The results chapter, chapter four, comprises the numerical tests that validate the method presented in chapter three and test its limitations. Possible future developments constitute the subject of chapter five. Chapter six presents the conclusions.

Chapter 2

Theory and Literature Review

2.1 Introduction

This chapter describes the linearization of the inverse problem in the abstract setting, with emphasis on the case of variable density acoustics. Solving the linearized inverse problem requires solving the normal equations. The efficient approximation of the inverse of the normal operator relies on its pseudodifferential nature. For one parameter inversion, the application of the normal operator on one input vector suffices to infer its action and represent its approximate inverse from a class of pseudodifferential operators. The extension to p-parameters generalizes this result and requires more applications of the normal operator.

2.2   Linearization of the Inverse Problem

The linearization of the inverse problem splits the model parameters into a smooth part m_0 and a rough perturbation δm,

$$m = m_0 + \delta m. \tag{2.1}$$

We assume that the background reference model m_0 is given and the inverse problem is to recover δm.

The nonlinear Forward map is thus approximated by,

$$S[m] = S[m_0 + \delta m] \approx S[m_0] + F[m_0]\delta m. \tag{2.2}$$

The linear operator F is known as Born modeling. Formally F is the derivative of the nonlinear map at m_0, it measures the sensitivity of S to small variations in the model. The conditions under which the right hand side of equation (2.2) provides a good approximation to its left hand side are discussed in Stolk (2000). The fact that the background velocity is smooth (and some suitable conditions on m_0) implies that F is generically asymptotically invertible (Stolk, 2000), in 2D (an invertible Fourier integral operator). The same result is conjectured for 3D. The smooth part of the velocity field models the kinematics in the problem; it controls the large scale behavior of the propagation of the wave: travel times, positioning of reflectors .... The rough part describes the nature of the reflection decided by the amplitude and nature of the discontinuities in the earth parameters.

The linearized inverse subproblem: Given d, m_0 find δm so that,

$$F[m_0]\delta m \approx d - S[m_0] := \delta d. \tag{2.3}$$

The linearized subproblem is an approximation due to the linearization process, and equation (2.3) is interpreted in a least squares sense to arrive at the normal equations:

$$F^*[m_0]F[m_0]\delta m = F^*[m_0]\delta d. \tag{2.4}$$

The operator F^* is adjoint to F, and is known as the migration operator. The operator F^*F is called the normal operator or the Hessian. The right hand side of (2.4) is the migrated image $m_{mig} = F^*\delta d$.

An explicit linearization of the acoustic wave equation for example yields:

$$\frac{1}{\rho_0 c_0^2}\frac{\partial^2 \delta p}{\partial t^2} - \nabla \cdot \frac{1}{\rho_0}\nabla \delta p = \frac{2\delta c}{\rho_0 c_0^3}\frac{\partial^2 p_0}{\partial t^2} - \frac{1}{\rho_0}\nabla\frac{\delta\rho}{\rho_0}\cdot\nabla p_0$$

$$\delta p \equiv 0, \quad t \ll 0, \tag{2.5}$$

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 linear forward map is therefore,

$$F\delta m = F\left(\frac{\delta c}{c}, \frac{\delta\rho}{\rho}\right)^T = \delta p, \tag{2.6}$$

where δp is obtained by solving (2.5). Note that the solution δp is linear in the model perturbation δm.

The solution of the linearized inverse problem requires solving the normal equations (2.4), equivalently inverting the normal operator. In reality, the application of

the normal operator as seen from equation (2.5) and its adjoint requires the solution of large scale PDE problems. The process of applying the normal operator amounts to modeling followed by a migration. These processes typically require computations that can take days or weeks on computer clusters. Also the problem is large scale: in 2D the fields are length $\approx 10^6$, giving the normal operator a $10^6 \times 10^6$ matrix representation. These numbers prohibit explicitly storing the normal operator to invert it using direct matrix methods like Gaussian elimination. Krylov subspace methods are used to solve (2.4), but the expensive application of F^*F limits the number of affordable iterations since these methods require at least one application of the normal operator per iteration.

The properties of the normal operator have been studied in the literature on the subject (Beylkin, 1985; Rakesh, 1988). For one parameter inversion (constant density acoustics), the normal operator is a pseudodifferential operator under specific conditions when the background velocity field is smooth, and no ray multipathing occurs (Beylkin, 1985; Rakesh, 1988). Stolk (2000) proves that the normal operator is the sum of a pseudodifferential operator and a non-microlocal part. He discusses the conditions under which the non-microlocal part is a Fourier integral operator and can be analyzed as such. Stolk (2000) concludes by proving that the normal operator is generically a pseudodifferential operator plus a smoother Fourier integral operator correction in 2D. The smoother error is of lower frequency order in high frequency asymptotics. In the more general case, the normal operator is a $p \times p$ matrix

of pseudodifferential operators for p-parameter inversion, when scattering preserves polarization as in P to P or S to S scattering (Beylkin and Burridge, 1990). The limitation to polarization preserving scattering is necessary, otherwise the normal operator maps one reflector to multiple reflectors in different places (one for each polarization). The normal operator would not preserve singularities in such case, and is therefore not a pseudodifferential operator. Symes (1998) provides an explicit proof that the normal operator is a two by two matrix of pseudodifferential operators for the variable density acoustics case, under some restrictions on ray geometry. Namely, the simple ray geometry condition, which excludes refracted rays from the sources to the receivers.

Pseudodifferential operators provide a generalization of differential operators. They are defined by their action on a function $u \in C_0^\infty(\mathbb{R}^n)$:

$$Qu(x) = \int q(x, \xi)\hat{u}(\xi)e^{ix.\xi}\,d\xi, \tag{2.7}$$

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

is the *symbol* of the pseudodifferential operator, and $\hat{u} = \mathcal{F}[u]$ is the Fourier transform of u.

The correspondence between the pseudodifferential operator and its symbol is stressed using the notation $Q = op(q)$, which reads: Q is the the pseudodifferential operator whose symbol is q.

Symbols are required to obey the following set of estimates: There exists $m \in \mathbb{R}$

(referred to as the order of the operator) so that for any compact set $K \subset \mathbb{R}^n$, and

α, β nonnegative multi-indices, there exist constants $C_{K,\alpha,\beta}$, such that

$$|D_x^\alpha D_\xi^\beta q(\mathbf{x}, \xi)| \leq C_{K,\alpha,\beta}(1 + |\xi|)^{m-|\beta|}, \tag{2.8}$$

for all $\mathbf{x} \in K$ and $\xi \in \mathbb{R}^n$. Such estimates are satisfied by smooth $q(\mathbf{x}, \xi)$ that are

positively homogeneous of order m in ξ. Positive homogeneity means that, given

$r \in \mathbb{R}$, $r > 0$,

$$q(\mathbf{x}, r\xi) = r^m q(\mathbf{x}, \xi). \tag{2.9}$$

Homogeneous symbols satisfy (2.8); however, it should be noted that (2.8) is satisfied

by a more general class of symbols not treated in this thesis. By allowing such general

classes of symbols pseudodifferential operators generalize differential operators.

In fact, we can allow for more general symbols that admit a polyhomogeneous

expansion. That is, there exist homogeneous $q_k \in C^\infty(\mathbb{R}^n \times \mathbb{R}^n \backslash \{\mathbf{0}\})$ of degree $k \leq m$,

for which

$$q \sim \sum_{j \geq 0} q_{m-j} \quad , \text{in the sense,} \quad q - \sum_{j=0}^{N-1} q_{m-j} \quad \text{is a symbol of order } m - N. \tag{2.10}$$

Pseudodifferential operators of order m generalize a differential operator of order m

as operators between Sobolev spaces:

$$Q : H_0^s(\mathbb{R}^n) \to H_{loc}^{s-m}(\mathbb{R}^n).$$

The first term in the polyhomogeneous expansion q_m is called the principal symbol

of q, and the remainder of the expansion maps H_0^s to a smoother space H_{loc}^{s-m+1}.

Properties of pseudodifferential operators are thus defined up to smoother error, which in the frequency domain corresponds to a lower order in frequency, at high frequency.

In order to be able to compose pseudodifferential operators, we will assume that supports are compact, that is, the set $\{x \in \mathbb{R}^n : (x, \xi) \in supp(q) \ for \ some \ \xi \in \mathbb{R}^n \backslash \{0\}\}$ is compact.

For a complete account on pseudodifferential operators and their applications in solutions of PDEs, please consult (Taylor, 1981).

Pseudodifferential operators act in phase-space. They are determined to leading order in frequency by their principal symbol $q_m(x, \xi)$ depending on both the spatial variable x and the momentum or Fourier variable ξ. The action of pseudodifferential operators preserves the singularities of the distributions on which it acts. However the amplitudes and Fourier spectra of these singularities will be modified depending on:

- Spatial position of the singularity,

- Orientation of the singularity (referred to as dip),

- The order m of the pseudodifferential operator.

We say that pseudodifferential operators act by *scaling* the input vector. To understand this behavior it is enough to compare δm and $F^* d = F^* F \, \delta m$ from equation (2.4) (Figures 2.1 and 2.2), derived from called the Marmousi model (Versteeg and Grau, 1991). Recall that $F^* F$ is a pseudodifferential operator under some circum-

stances, which include this example. Note that the inverse and the migrated image are related by the normal operator. The original Marmousi model is smoothed and split into a smooth part and a residual δm. The normal operator uses wave equation methods to apply F (Born modeling) and F^* (reverse time migration). It is obvious that the positions of the reflectors or discontinuities are preserved. The normal operators acts by scaling the amplitudes of δm.
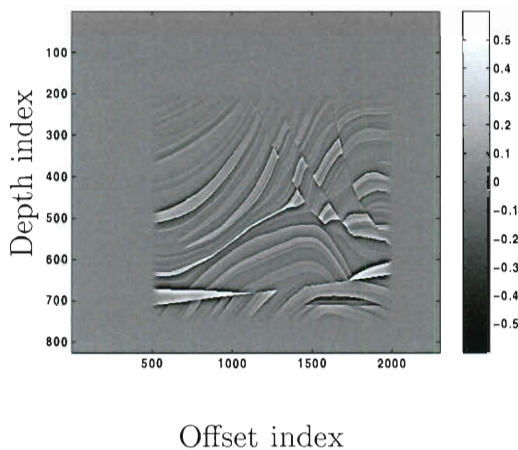
Figure 2.1: δm, input to normal operator
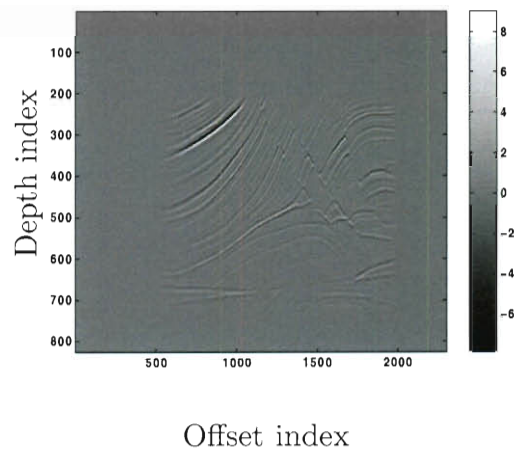
Figure 2.2: $m_{mig} = F^*\delta d = (F^*F)\delta m$, normal operator applied to the input.

The sense in which a pseudodifferential operator (and hence the normal operator) acts, by locally scaling the amplitudes of the discontinuities while preserving their position, is explained by the asymptotic expansion lemma for pseudodifferential operators (Taylor, 1981) (page 184). Let $\chi(x)$ be a smooth function compactly supported inside a 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 for a pseudodifferential operator Q of order m:

$$Q\,\chi(x)e^{i\omega\Psi(x)} = q_m(x, \omega\nabla\Psi(x))\chi(x)e^{i\omega\Psi(x)} + O(\omega^{m-1}), \tag{2.11}$$

where ω is the frequency and q_m is the principal symbol of Q, positively homogeneous of order m:

$$q_m(x, \omega\xi) = \omega^m q_m(x, \xi), \qquad \omega > 0. \tag{2.12}$$

We can even see how the scaling is related to the symbol of the pseudodifferential operator from (2.11).

In seismic images, reflectors are interfaces between two regions of space that exhibit discontinuities or high contrasts in physical parameters. These discontinuities account for the high frequency components (rapid changes). The relevant functions are of the form $a(x)f(\Psi(x))$, an amplitude modulated function of the phase. The high frequency components of these functions correspond to the discontinuities/reflectors. The vector $\nabla\Psi(x)$ is normal to the level sets of the phase function and represents dip as it points to a direction normal to the reflector identified with the level set of the phase function. This normal vector fails to be well defined in multiple dip events (fault, point reflector ...).

2.3 Scaling Methods

The realization that the normal operator acts by scaling the amplitudes of the right hand side of equation (2.4) called the migrated image (2.4), lead to the idea that the

action of the normal operator may be approximated from its application on a single input vector. In the same way diagonal matrices scale all vectors in the same way, the normal operator is approximately diagonal in a basis of localized monochromatic pulses (justified by equation (2.11)). I call methods that rely on one application of the normal operator to approximate it or to approximate its inverse as *scaling methods*, and I refer to the approximations they yield as *scaling factors*. The choice of the migrated image as the input vector was suggested by Symes to Claerbout and Nichols as they developed an early scaling method in 1994. The choice of this input vector is motivated by the fact that the migrated image contains all relevant directions or reflectors.

Claerbout and Nichols (1994) propose approximating the normal operator and its inverse as a multiplication by a smooth function, using the migrated image as input vector. The method was then refined by Rickett (2003).

Guitton (2004) proposes a more general near diagonal approximation of the normal operator: near diagonal integral operators that are not completely specified by Guitton.

Symes (2008) proposes a correction to the Claerbout and Nichols method: he proves that the normal operator is approximated by multiplication by a smooth function after application of a Laplacian filter of a specific power, the power is predicted by the underlying theory and the filter is completely specified, in contrast to Guitton's method. However, the method cannot approximate the inverse of the normal

operator in places of the image that admit multiple dip events (faults, point reflectors ...). The method requires dip to be well defined in all parts of the image thus limiting its applicability. This method therefore fails locally in places that admit a fault or a point reflector, for example.

Herrmann et al. (2008b) approximately diagonalize the normal operator in a frame of approximate localized monochromatic pulses, namely *curvelets*. They rely on the asymptotic expansion lemma (2.11) to justify their method. This method is capable of resolving multiple dip events.

Demanet et al. (2011) propose an approximation of the inverse of the normal operator, through a number of applications of the normal operator to randomized trial functions in curvelet space. In fact, the paper compares the result of the proposed method to the pseudodifferential scaling methods presented in the next chapter. The authors assert that the pseudodifferential scaling method with one application of the normal operator (only the one parameter case was treated), outperforms the randomized trial functions method no matter how many applications of the normal operator are effected (Demanet et al., 2011). However, the application to multiple trial functions yields an approximation of the inverse of normal operator that can be applied to any vector, the authors refer to this property by calling the inverse generalizable. The method I propose in this manuscript, approximates the action of the inverse of the normal operator to one vector, the right hand side of the normal equations (the migrated image). The paper also shows how the performance of scaling

methods degrades as the smoothness of the background model is lost (Demanet et al., 2011).

I propose a generalization to the method introduced by Symes, a scaling method that resolves multiple dip events and skips the explicit diagonalization of the normal operator and the use of curvelets altogether. The method relies on a truncated spherical harmonics expansion of the symbol to approximate its action. The method reduces to the method proposed by Symes when the expansion of the symbol consists of the first term. I present this method in more detail in the next chapter.

2.4 Amplitude Versus Offset (AVO)

The Zoeppritz equations specify how waves are transmitted and reflected at an interface. The study is conducted for the elastic wave equation, and the results specify the reflection and transmission coefficients in terms of the offset angle (Aki and Richards, 1980). These equations are nonlinear in the physical parameters of the earth. Nonetheless, they can be solved if one is willing to invest the computational cost.

The complexity of the Zoeppritz equations led to attempts to simplify these equations while preserving the qualitative predictions of the full Zoeppritz equations. Usually these simplifications are linearizations in the relative differences of the physical parameters from the two sides of the interface. Aki and Richards (1980) present one of these linearizations (p. 153). Another simplification widely used in the AVO study

is that presented by Shuey (1985). Shuey presents a formulation that demarcates the behavior for normal incidence, from intermediate angles (about 30 degrees), to wide angles (approach to critical angle) (Shuey, 1985).

While zero-offset reflection contains information about the acoustic impedance only, the variation of the reflection coefficients with offset angle contain information about all the elastic parameters (Lörtzer and Berkhout, 1989). In AVO, the Zoeppritz equations and their various simplifications are used in an inverse problem sense to infer information about the elastic parameters from the variation of the amplitudes of the reflection coefficients as a function of offset (or offset angle).

Rutherford and Williams (1989) classify different reflectors according to the qualitative variation of the reflection coefficient as a function of offset angles. This classification consists of three classes of reflectors that exhibit qualitatively different AVO behavior. Rutherford and Williams (1989) deal with gas sands encountered in exploration and split them into:

1. Class 1: High impedance sands,

2. Class 2: Near zero impedance contrast sands, and

3. Class 3: Low impedance sands.

The use of AVO data to approximate the material parameters abounds in the literature. Lörtzer and Berkhout (1989) presents a statistically Bayesian based approach to predict a combination of physical parameters that predicts the AVO variation. The

paper stresses the need for multi-component data to alleviate the ill-conditioning of the problem and pin down certain parameter combinations that are difficult to resolve (Lörtzer and Berkhout, 1989). It also quantifies how the resolution of different parameters varies as a function of the angle range, and how sensitive the recovery of parameters is to calibration parameters used in the statistical approach.

The method presented in this manuscript is an alternative to AVO analysis in that the inverse contains all the info that an AVO study can possibly yield about the different parameters. In fact, after approximating the Zoeppritz equations and taking into account the uncertainty in the data, AVO analysis yields information about anomalies in the physical parameters (usually the Poisson ratio) rather than quantitative measures of these parameters. Linearized inversion aims at recovering the physical parameters quantitatively.

2.5   Linearized Multi-Parameter Inversion

The attempts for multi-parameter inversion are limited in the literature. Bourgeois et al. (1989) study the linearized multi-parameter inversion and conclude that the success of this procedure relies on an accurate background velocity model. With an accurate background model, Bourgeois et al. (1989) conclude that the recovery of the impedance is possible in that its inversion yields significant corrections to seismic images produced by migration. The results of this paper confirm the well known fact that the recovery of the acoustic impedance is a well conditioned problem for

variable density acoustics. The recovery of the other parameters remains challenging, an aspect of this inverse problem that is alluded to in the subsequent chapters.

Santosa and Symes (1988) present a study of the inverse problem for a layered acoustic fluid. They parametrize the problem in terms of density and incompressibility. They conclude that the problem is well conditioned away from critical angles of reflection and with enough aperture, in the absence of low velocity zones. In the presence of low velocity zones, the degradation of the conditioning of the inverse problem is unavoidable. The study of the conditioning in this paper resembles the conditioning study I present on the normal operator in this manuscript. The inverse problem studied in (Santosa and Symes, 1988) is effectively one dimensional (layered). The method presented here is not limited to 1D; in fact, the formulation is independent of dimension. The conditioning study, in this manuscript and the paper, is restricted to layered models; these models allow for an analytical study of the conditioning of the inverse problem.

Multi-parameter linearized inversion constitutes part of Minkoff and Symes (1997). In fact, they show that the success of linearized multi-parameter inversion relies on pinning down other aspects of the inverse problem accurately: source estimation, background field approximation, modeling the physics accurately (including elasticity, attenuation ...). Minkoff and Symes (1997) succeed in fitting field data by including all the parameters mentioned above. Linearized inversion constitutes one of these steps, but relies heavily on all the others. In this thesis, I assumed that the seismic

source is known, the background model given and the data were generated using the model used to fit them. These steps ensure that linearized multi-parameter inversion has the potential of succeeding.

Charara et al. (1996) invert for P-velocity, S-velocity and density in the linear elasticity inverse problem. Though the inversion is nonlinear, the estimation of these parameters is done in the linear regime (when the background fields are accurate enough). Charara et al. (1996) incorporate constraints on both the data space and the model space and use a least squares inversion approach by formulating these constraints through covariance matrices, as advocated in (Tarantola, 1987). Some of these covariance matrices rely on a priori knowledge, like well data. The inversion relies on starting at a good initial model, stressing the role of the accurate background model in successful inversion.

The works presented above are all known as wave equation methods, as they rely on a numerical solution of the wave equation to obtain seismic images and data, and inversion is formulated in terms of these quantities. Another approach derives explicit formulas for the inverse of the linearized forward map, which involves calculating geometric optics quantities derived under an asymptotic regime. This work follows the approach of Beylkin (1985). These computations usually involve conditions on the medium of propagation for the geometric optics quantities to be well defined, and this is reflected as an instability in computing these quantities in complex media. The method proposed in this manuscript is closer to wave equation methods, in that it

does not require any computation of geometric optics quantities.

Foss et al. (2005) go through the necessary computations to derive the asymptotic inverse to the linearized map for anisotropic elastic media. The authors present a numerical example of the recovery of a linear combination of the density and two other elastic parameters, and they claim that their "framework applies with decent accuracy" (Foss et al., 2005). Foss et al. (2005) do not show synthetic examples so it is difficult to judge the accuracy of the framework.

Virieux et al. (1992) take a mixed approach to invert for the P and S impedance in linear elasticity. They use geometric optics computations to calculate the forward map and an approximation of the Hessian, and a Gauss-Newton iterative method for the inversion. The approximation of the Hessian accelerates the convergence of the inversion. Virieux et al. (1992) also study the conditioning of the inverse problem, and conclude that it is ill-conditioned for single component data; multi-component data is necessary for a successful inversion. The authors note that this is in accordance with the study of Santosa and Symes (1988) in the special case of a layered fluid. The inversion for P and S impedances is successful in the linear regime; the authors stress the importance of an accurate background model (Virieux et al., 1992).

The attempts for linear multi-parameter inversion in the literature that use wave equation methods rely on an iterative approach to minimize the least square misfit between the measured data and the data predicted by linearized modeling. The method proposed here, to use a few applications of the normal operator to produce

an approximate inverse, is novel.

Chapter 3

Methods

3.1 Introduction

This chapter presents the scaling method to approximate the inverse of the normal operator efficiently for one parameter inversion, and its generalization to multi-parameter inversion. Cramer's rule for reduces the multi-parameter inversion problem to one-parameter inversion. The inverse of the normal operator in the one parameter problem is represented from a class of pseudodifferential operators defined by a truncated Fourier expansion of their symbol. The efficient approximation relies on an algorithm derived by Bao and Symes (1996) to approximate the action pseudodifferential operators. I refer to this algorithm as the PsiDO algorithm. This algorithm was derived in 2D but extends to 3D. The presentation of the method is independent of the space dimensions. I derive the generalization to 3D and show the implementa-

tion of the PsiDO algorithm in 3D. I end this chapter with a discussion of the errors committed in each step of the method, and ways to control them.

3.2 One Parameter Inversion: Pseudodifferential Scaling

Recall that the aim of this manuscript is to solve,

$$N\delta m = b, \tag{3.1}$$

where $N = F^*F$ is the normal operator and $b = F^*\delta d \in Range(N)$ is the migrated image.

Given b and Nb (we refer to Nb as the remigrated image), we seek a scaling factor c that minimizes the following objective function:

$$c = \operatorname*{argmin}_{c \in \Psi DO} \|b - cNb\|^2. \tag{3.2}$$

The advantage of obtaining c lies in the ability to derive an approximate inverse δm_{inv}, given c:

$$\delta m = N^\dagger b \approx N^\dagger cNb \approx cN^\dagger Nb = cb := \delta m_{inv}. \tag{3.3}$$

The first equation expresses the solution of equation (3.1), where $N^\dagger$ is the pseudoinverse (regularized inverse) of N.

One way to construct this approximate inverse is by considering a protected inverse of the principal symbol q of the normal operator. That is $N^\dagger = op(q^\dagger)$. With $q^\dagger = \frac{1}{q+\lambda}$ where λ is a small positive number. This is a protected inverse that avoids problems arising from areas where the symbol is zero.

The second approximation relies on the quality of the fit in equation (3.2). The third uses the property that pseudodifferential operators approximately commute. The successive approximations thus yield an approximate inverse $\delta m_{inv} = cb$.

The sense in which the scaling factor c approximates the inverse of the normal operator is specifically in the sense of (3.3): c scales the amplitudes of the migrated image b in the same way the inverse of the normal operator $N^\dagger$ does. The scaling factor approximates the action of the inverse of the normal operator on one right hand side of the normal equations, namely the migrated image.

3.3 Multi-Parameter Inversion: Cramer's Rule for Pseudodifferential Operators

The generalization to p parameters formally tries to solve the same problem:

$$N\delta m = b. \tag{3.4}$$

The model m is a collection of p parameters, and the normal operator is therefore a $p \times p$ matrix of pseudodifferential operators in polarization preserving scattering. The p migrated images are contained in b.

The theory of pseudodifferential operators introduces a powerful concept: the algebraic relationships between symbols and matrices of symbols can be mapped to asymptotic properties for pseudodifferential operators. This concept is powerful since matrices of symbols are matrices of scalar functions, with a plethora of identities and theorems from linear algebra to choose from. This work uses a version of Cramer's rule to devise an inversion scheme for multi-parameter inversion.

For this end, we recall the definition of the adjugate of a matrix A, denoted by $Adj(A)$ (Strang, 1988). Defined as the transpose of the matrix of cofactors of A. In our case, the transpose may be ignored, as the matrix A and thus its adjugate, are symmetric positive definite. When the matrix A is invertible the adjugate may be defined as

$$Adj(A) := det(A)\, A^{-1}, \tag{3.5}$$

where A^{-1} is the inverse of the matrix A and $det(A)$ is the determinant of the matrix A. More generally, the adjugate is defined to be the matrix which satisfies:

$$Adj(A)\, A = A\, Adj(A) = det(A)\, I, \tag{3.6}$$

where I is the identity matrix. If $N = op(A)$ (N has matrix symbol A), and we define the adjugate of N to be $Adj(N) = op(Adj(A))$, with slight abuse of notation. We can map the property (3.6) on matrices to a property on matrices of pseudodifferential operators:

$$Adj(N)\, N \approx N\, Adj(N) \approx det(N)\, I. \tag{3.7}$$

The equation above features another abuse of notation, with $det(N) := op(det(A))$.

The power of (3.7) is revealed when applied to (3.1):

$$Adj(N) \, b = Adj(N) \, N \, \delta m \approx det(N) \, \delta m. \tag{3.8}$$

Equation (3.8) recovers the inverse up to the pseudodifferential operator $det(N)$, after the application of the adjugate. The restriction of this approach to 1-parameter inversion and 2-parameter inversion is particularly simple and elegant. The extension to general p-parameters is more involved as shown by the case $p = 3$, which we discuss in the results chapter.

The problem of recovering δm is not solved yet. While δm is recovered up to the pseudodifferential operator $det(N)$, the inverse of this factor needs to be approximated to complete the inversion scheme. For this end, we resort to a method similar to the one we previously developed for $p = 1$.

First apply the normal operator again, to form:

$$N \, det(N) \, \delta m \approx det(N) \, N \, \delta m = det(N) \, b. \tag{3.9}$$

Where we have used the fact that scalar pseudodifferential operators approximately commute with matrices of pseudodifferential operators, to commute N and $det(N)$.

Now, given b and $det(N) \, b$, approximate the scaling factor c:

$$c = \underset{c \in \Psi DO}{argmin} \|b - c \, det(N) \, b\|^2. \tag{3.10}$$

To minimize this objective function I use a quasi-Newton method: limited memory BFGS (lBFGS). Such methods only require a user defined gradient, and approximate

the Hessian of the objective function. In my masters thesis, I discuss in detail the specific choice to parametrize the scaling factor c, which leads to a derivation of the gradient. For details on this part of the method, please consult (Nammour, 2009).

Approximate the solution of normal equations by:

$$\delta m = N^\dagger b \approx N^\dagger\, c\, det(N)\, b \approx c\, det(N)\, N^\dagger b$$
$$\approx c\, det(N)\, \delta m := \delta m_{inv}. \tag{3.11}$$

Thus the scaling factor c is an approximation of the inverse of $det(N)$, in that it is applied to $det(N)\, \delta m$ (obtained previously), to approximate δm.

$N^\dagger$ is an approximate inverse, one way to define this inverse is by noting that the adjugate matrix is a matrix of symbols. Also the determinant of N may be inverted in a protected manner, as discussed in the previous section. Putting these two pieces together we may define:

$$N^\dagger = op\left((det(A))^\dagger\, Adj(A)\right).$$

It is also straightforward to see that the one-parameter case is a restriction of this general approach to $p = 1$.

The efficiency of this approach relies on the efficiency of the optimization in (3.10). Any optimization scheme will require the application of the pseudodifferential scaling factors at each iteration. It is therefore pivotal to use an algorithm that applies pseudodifferential operators efficiently. Bao and Symes (1996) develop an algorithm to efficiently approximate the action of pseudodifferential operators, which relies on

a spherical harmonics expansion of their symbols. We refer to this algorithm as the PsiDO algorithm.

The advantage of using the PsiDO algorithm is twofold: first its efficiency, and second the ability to represent pseudodifferential operators that act by spatial, frequency and dip dependent scaling capable of resolving multiple dip events.

3.4   The PsiDO Algorithm

I present the PsiDO algorithm that allows the efficient representation and approximation of the scaling factors. As developed by Bao and Symes (1996), the algorithm is presented explicitly in 2D. The extension of the PsiDO code to 3D is shown in the next section.

This discussion is restricted to 2D, so we may write $\mathbf{x} = (x, z)$. Recall that a pseudodifferential operator is characterized by its symbol and defined by

$$Q_m u(x, z) = \int \int q_m(x, z, \xi, \eta)\hat{u}(\xi, \eta)e^{i(x\xi + z\eta)}\, d\xi\, d\eta, \tag{3.12}$$

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

$$q_m(x, z, \xi, \eta) = \omega^m \tilde{q}_m(x, z, \theta). \tag{3.13}$$

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:

$$\tilde{q}_m(x, z, \theta) \approx \sum_{l=-K/2}^{l=K/2} c_l(x, z) e^{il\theta} = \sum_{l=-K/2}^{l=K/2} \omega^{-l} c_l(x, z) (\xi + i\eta)^l. \tag{3.14}$$

Plugging (3.14) into (3.12) we obtain

$$Q_m u(x, z) \approx \sum_{l=-K/2}^{l=K/2} c_l(x, z) \mathcal{F}^{-1}[\omega^{m-l}(\xi + i\eta)^l \hat{u}(\xi, \eta)]. \tag{3.15}$$

An error results from truncating the Fourier series. The error analysis that specifies how this error propagates to the action of a pseudodifferential operator is a possible future development prospect.

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 (3.15) 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 (3.12) has a computational complexity of $O(N^4 \log(N))$. The algorithm described above uses the *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; it depends on the smoothness of the symbol. 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. In the examples I show in this

manuscript K does not exceed 10. Special care is taken to make sure that increasing K results in no significant change in the fit of the objective function (3.10).

The dependence on dip is captured in the angle variable θ, and the method allows us to capture multiple dip events by increasing $K > 1$.

3.5 Extension to 3D

The only part that restricts the application of the method to 2D is the PsiDO algorithm written explicitly for 2D. An extension of the PsiDO algorithm to 3D effectively opens up the possibility of applying the approximate inversion method to 3D models.

The 3D extension proceeds in the same fashion as Bao and Symes (1996). First write the symbol as:

$$q_m(x, y, z, \xi, \zeta, \eta) = \omega^m \tilde{q}_m(x, y, z, \theta, \phi). \tag{3.16}$$

The relationship between ω, ϕ and θ to ξ, ζ, and η is given by the spherical coordinates transformation:

$$\xi = \omega \cos(\phi) \sin(\theta)$$

$$\zeta = \omega \sin(\phi) \sin(\theta) \tag{3.17}$$

$$\eta = \omega \cos(\theta)$$

The truncated spherical harmonics expansion of $\tilde{q}$ is then given by

$$\tilde{q}(x, y, z, \theta, \phi) = \sum_{l=0}^{K} \sum_{n=-l}^{l} c_{ln}(x, y, z) Y_l^n(\theta, \phi), \tag{3.18}$$

where Y_l^n are the spherical harmonics basis functions. The coefficients can be calculated as,

$$c_{ln}(x, y, z) = \int_\Omega \tilde{q}(x, y, z, \theta, \phi) Y_l^{n*} \, d\Omega = \int_0^{2\pi} d\phi \int_0^\pi d\theta \sin(\theta) \tilde{q}(x, y, z, \theta, \phi) Y_l^{n*}.$$

(3.19)

In order to evaluate the action of the PsiDO algorithm, we need to express $Y_l^n(\theta, \phi) = Y_l^n(\xi, \zeta, \eta)$. In fact,

$$Y_l^n(\theta, \phi) = \sqrt{\frac{(2l+1)(l-n)!}{4\pi(l+n)!}} P_l^n(\cos(\theta)) e^{in\phi},$$

(3.20)

where P_l^n are the associated Legendre polynomials, which may be calculated by recursion formulas or using an identity known as the Rodrigues' formula. Thus,

$$\omega = \sqrt{\xi^2 + \zeta^2 + \eta^2}$$

$$\cos(\theta) = \frac{\eta}{\omega}$$

$$e^{in\phi} = e^{in \arctan(\zeta/\xi)}$$

(3.21)

Thus we can define $Y_l^n(\xi, \zeta, \eta) := Y_l^n(\arccos(\eta/\omega), \arctan(\eta/\xi))$. Plugging into the action of a pseudodifferential operator,

$$Q_m \, u(x, y, z) = \sum_{l=0}^{K} \sum_{n=-l}^{l} c_{ln}(x, y, z) \mathcal{F}^{-1} \left\{ \omega^m Y_l^n(\xi, \zeta, \eta) \hat{u}(\xi, \zeta, \eta) \right\},$$

(3.22)

The cost of this algorithm is $(K+1)^2$ applications of the inverse Fourier transform; if we use the FFT, the cost would be $(K+1)^2 N^3 \log(N)$.

If we only use this algorithm to approximate scaling factors as discussed in the methods chapter, then we only need to represent the associated Legendre polynomials

explicitly. The symbol is then parametrized by $c_{ln}(x, y, z)$. In fact, we do not require any spherical harmonics transforms. Such transforms are needed if we are given a symbol and asked to apply its action, which is not the problem at hand since we never have access to the symbol of the normal operator, and we can use directly the parametrization (3.18) in forming c. The algorithm for 3D, which I implemented in MATLAB, becomes:

Sample the field $u(x, y, z)$ and the coefficients $c_{ln}(x, y, z)$,

$$U_{ikj} = u(x_0 + (i - 1)\Delta x, y_0 + (k - 1)\Delta y, z_0 + (j - 1)\Delta z),$$

$$C_{ikjln} = c_{ln}(x_0 + (i - 1)\Delta x, y_0 + (k - 1)\Delta y, z_0 + (j - 1)\Delta z),$$

$$i = 1, \cdots, M, \ k = 1, \cdots, P, \ j = 1, \cdots, N \ .$$

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

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

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

$$Z_{pqr} = 2\pi q\Delta\zeta$$

$$E_{pqr} = 2\pi r\Delta\eta$$

$$\Phi_{pqr} = \arctan(\frac{E_{pqr}}{\Xi_{pqr}})$$

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

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

2. Initialize $(QU)_{ikj} = 0$, for $i \in [1, M]$, $k \in [1, P]$, $j \in [1, N]$,

   For $l = 0 : K$

   For $n = -l : l$

   (a) compute $(Y_l^n)_{pqr} = \sqrt{\frac{(2l+1)(l-n)!}{4\pi(l+n)!}}(P_l^n)(E_{pqr}/\Omega_{pqr})e^{in\Phi_{pqr}}$

   (b) compute $\{R_{ikj}^{ln}\}_{i=1,k=1,j=1}^{M,P,N} = \mathcal{F}^{-1}[\Omega_{pqr}^m (Y_l^n)_{pqr} \hat{U}_{pqr}]$

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

   (c) accumulate

   $$(QU)_{ikj} = (QU)_{ikj} + C_{ikjln} R_{ikj}^{ln}$$

   End

One application of this algorithm is dip filtering. Dip filtering consists of filtering out discontinuities with a specific orientation. Pseudodifferential operators achieve dip filtering since they act in phase space. Moreover, they allow for spatially dependent dip filtering if one wants to target only a certain area of the model. I show an example where the Marmousi model (Figure 3.1) is extended to a 3D model by spreading along the y-direction. I design a specific pseudodifferential operator to filter the vertical events. This is achieved by making the symbol vanish along the vertical direction. The symbol $\sin^2(\theta)$ is one example, easily parametrized by its spherical harmonics expansion. The dip-filtered Marmousi model is shown in Figure 3.2. Note

how vertical discontinuities are smoothed out to zero, while horizontal discontinuities
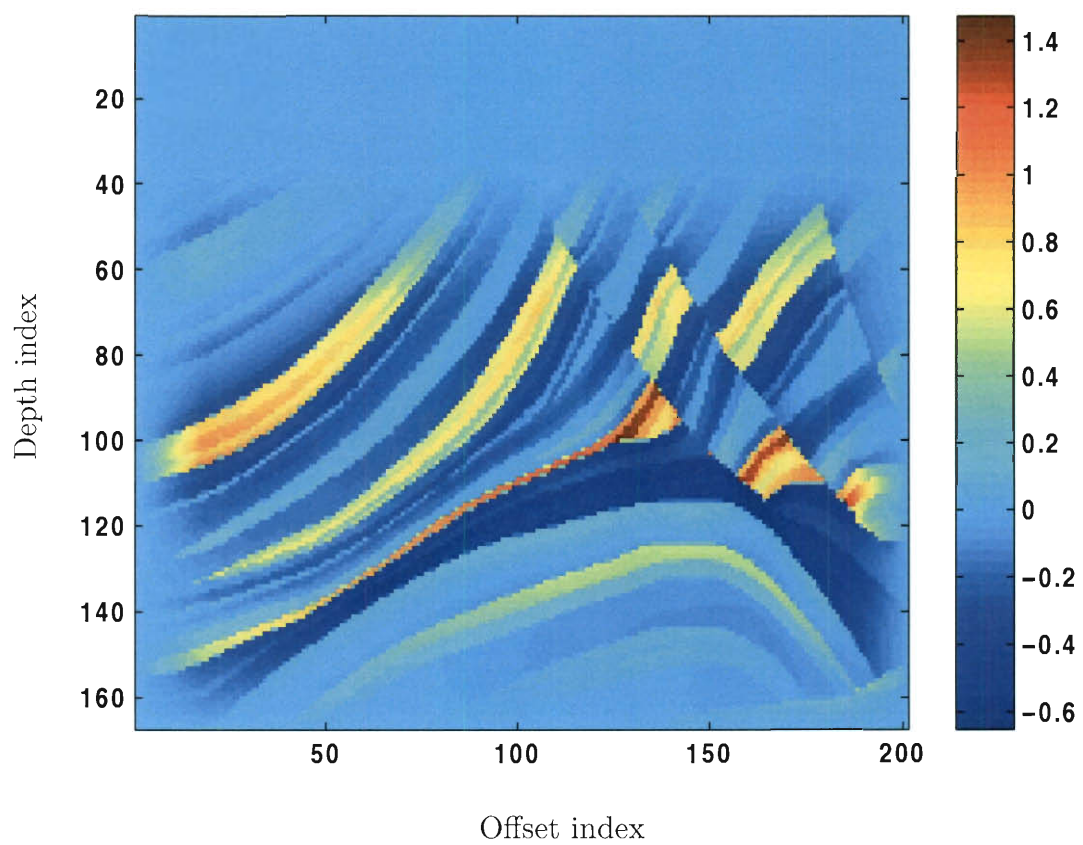
remain untouched.

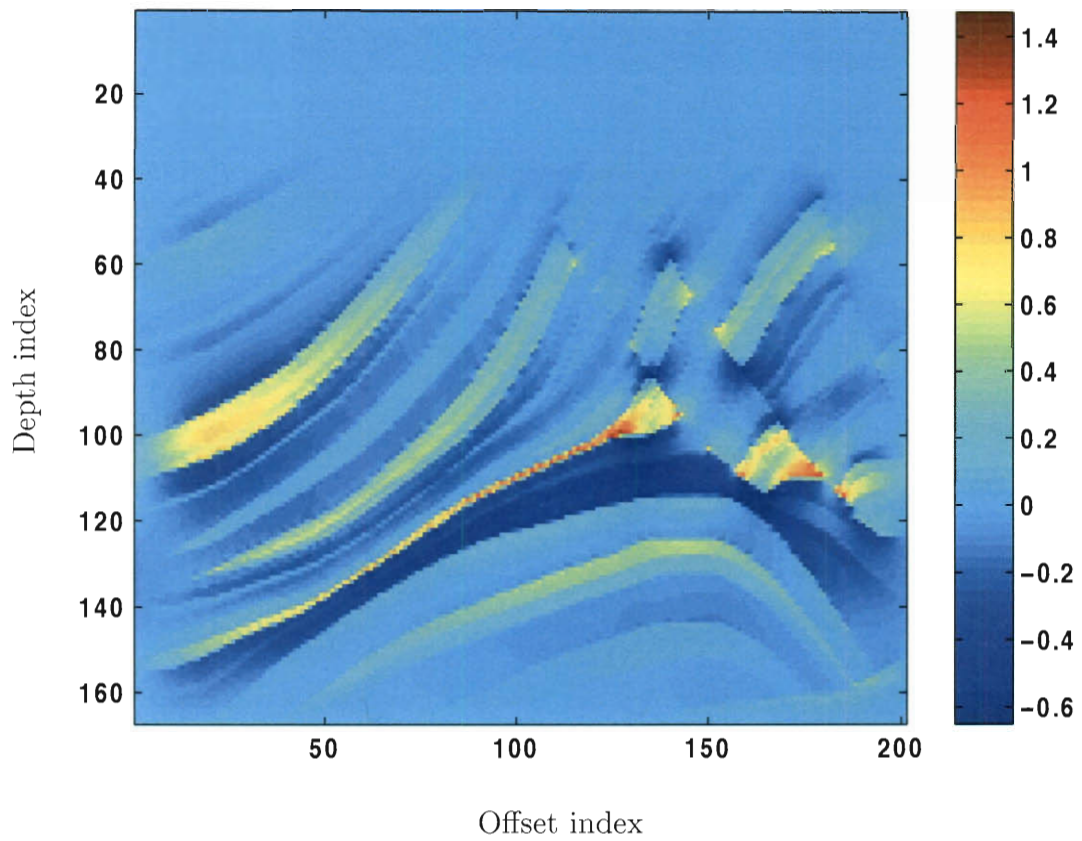Figure 3.1: A 2D section of the Marmousi model.

Figure 3.2: A 2D section of the Marmousi model with smooth filtering of the vertical dips.

3.6   Summary

I summarize the procedure that constitutes the scaling method.

To solve $N\delta m = b$,

- Apply $Adj(N)$ on b to form: $Adj(N)\, b \approx det(N)\, x$

- Apply N to the result to get: $N\, det(N)\, x \approx det(N)\, b$

- Represent the scaling factor using the PsiDO algorithm $c = Q_m[q]$

- Compute the scaling factor c:

$$c = \underset{c \in \Psi DO}{argmin} \|b - c\, det(N)\, b\|^2.$$

- Approximate the inverse: $x_{inv} := c\, det(N)\, x \approx x$

The method applies in 2D and 3D.

3.7   Discussion of the error in various steps

This section discusses the errors committed in each step of the method summarized above.

First, the error committed in applying the adjugate. The adjugate of the symbol matrix in equation (3.6) is exact. This is due to the fact that scalar functions commute. The adjugate of the normal operator in equation (3.7) is approximate, since

scalar pseudodifferential operators commute approximately. The error committed is a commutator error. The commutator of two operators a, b is defined by :

$$[a, b] = ab - ba.$$

I will explicitly discuss the error for the two parameter case. This discussion generalizes for any p. For two parameters:

$$Adj(N)\,N = \begin{pmatrix} N_{22} & -N_{12} \\ -N_{12} & N_{11} \end{pmatrix} \begin{pmatrix} N_{11} & N_{12} \\ N_{12} & N_{22} \end{pmatrix}$$

$$= (N_{22}N_{11} - N_{12}^2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & [N_{22}, N_{12}] \\ [N_{11}, N_{12}] & [N_{11}, N_{22}] \end{pmatrix} \quad (3.23)$$

$$= det(N)\,I + \begin{pmatrix} 0 & [N_{22}, N_{12}] \\ [N_{11}, N_{12}] & [N_{11}, N_{22}] \end{pmatrix}.$$

The error committed in approximating $Adj(N)\,N \approx det(N)\,I$ is due to commutators of scalar pseudodifferential operators. The theory of pseudodifferential operators predicts that:

$$ord([a, b]) \leq ord(a) + ord(b) - 1.$$

Hence the commutators are lower order in the high frequency asymptotic limit. The same commutator error appears in $N\,det(N)\,x \approx det(N)\,b$.

Note that the PsiDO algorithms in 2D and 3D introduce two types of errors. First, a discretization error due to approximating the continuous Fourier transform

by a discrete Fourier transform. If the data is un-aliased, this error is smaller than the other errors due to commutators, for example. Second, is the error introduced by approximating a symbol by a truncated spherical harmonics expansion. We do not have access to the real symbol of the normal operator or its inverse, so we cannot measure this error directly.

Second, the error in fitting the objective function (3.10), is due to the choice of parameters in the optimization and whether $det(N)$ is indeed a pseudodifferential operator. The latter fact is enforced by respecting the underlying theory (smooth background, polarized scattering, bandpass source ...). This error is hard to predict a priori; how well the objective function (3.10) is minimized provides a control over this error ex post facto. The fact that we only approximate the principal symbol in representing the scaling factor c ignores, again, a lower order frequency error. We test the error discussed in the previous paragraph, resulting from truncating the spherical harmonics expansion of the symbol, by increasing the number of terms in the spherical harmonics expansion until the objective function ceases to decrease further.

Third, the error in approximating the inverse is

$$\delta m = N^\dagger b \approx N^\dagger \, c \, det(N) \, b = c \, N^\dagger \, det(N) \, b + [N^\dagger, c] \, det(N) \, b$$

$$= c \, det(N) \, N^\dagger \, b + c \, [N^\dagger, det(N)] \, b + [N^\dagger, c] \, det(N) \, b$$

$$= \delta m_{inv} + c \, [N^\dagger, det(N)] \, b + [N^\dagger, c] \, det(N) \, b.$$

$$(3.24)$$

Again, the error is due to commutators. Note that we have abused notation where we

have implicitly defined the commutator between a scalar pseudodifferential operator and a matrix of pseudodifferential operators, to be the matrix whose entries are the commutators between the scalar pseudodifferential operator and the respective entries of the matrix.

In addition to the error being of lower order in frequency, the error committed in this step is unavoidable. We never represent the symbol of N explicitly, and thus the symbol of $N^\dagger$ is not accessible. The commutator errors involving $N^\dagger$ are unavoidable, which justifies throwing away lower order frequency error in all the previous steps as they are of the same order as these commutators.

Some numerical representations of pseudodifferential operators go beyond the principal symbol representation used in this thesis. Notably, Demanet and Ying (2011) in their paper on discrete symbol calculus, propose a numerical approximation of the full symbol of a pseudodifferential operator. Their algorithm may be used to parametrize the scaling factor c, instead of using the PsiDO algorithms in 2D and 3D. Such a parametrization avoids the error in approximating only the principal symbol. However, the error of the same order in frequency shown in equation (3.24) is unavoidable, for the reasons discussed above. Therefore, the use of an algorithm to represent the scaling factor c that goes beyond the principal symbol, only delays the error and does not necessarily decrease it. It is for that reason that the PsiDO algorithms are sufficient for the purposes of this thesis.

Chapter 4

Results

4.1   Introduction

In this chapter I summarize the results of the masters thesis concerning constant density acoustics inversion. I validate the scaling method for one parameter inversion on the Marmousi benchmark model. The two-parameter case is tested on a set of layered examples ranging from simple two layer models, to complex 1D sections of the Marmousi model with homogeneous and smooth background models. The numerical experiments validate the method. A final example shows the limitations of this method when a condition for the normal operator to be pseudodifferential (smoothness of the background model) is not respected. This chapter also presents an analytical example for the three parameter case. The last section is a conditioning study of the linearized acoustic inverse problem for the recovery of impedance and

44

density.

4.2 One parameter inversion: Constant density acoustics

In this case the adjugate is particularly easy, $Adj(N) = I$, and $det(N) \equiv N$. This approach boils down to approximating c such that

$$c = \underset{c \in \Psi DO}{argmin} \|b - c\,N\,b\|^2. \tag{4.1}$$

I validate one parameter inversion on the 2D Marmousi synthetic benchmark model (Versteeg and Grau, 1991). Details of this example and other tests for the one parameter case constitute the bulk of my Masters thesis (Nammour, 2009). The model is smoothed to construct a background model, and the residual is the perturbation $\mathbf{m}_{true}$ (Figure 4.1). The images are windowed and tapered to the window of interest. The true model is Born modeled and then migrated to obtain the migrated image and the process is repeated to obtain the remigrated image (Figures 4.2 and 4.3). These images show the amplitude distortion resulting from the application of the normal operator. It is obvious that the amplitudes in the deeper part of the image are attenuated, making these regions invisible without amplitude correction. The pseudodifferential scaling method with $K = 1$ and $K = 5$ yields scaling factors that I apply to the migrated image to obtain the approximate inverses (Figures 4.4 and 4.5). The amplitudes are recovered to the right order of magnitude. Moreover, the

amplitudes are uniform in depth and compare better to those of the real image. Both these results are successful approximate inversions.

The scaling method with $K = 1$ cannot resolve multiple dip events, in contrast with $K = 5$. I plot the difference between the two inversion results in figure 4.6 to study this feature. It is apparent that the amplitude difference is greatest at the locations of multiple dip events, where two reflectors intersect (faults). The high amplitude difference appears as brighter or dimmer spots in figure 4.6.

Please consult (Nammour, 2009), where I present a test of the ability of the scaling method to resolve multiple dip events.
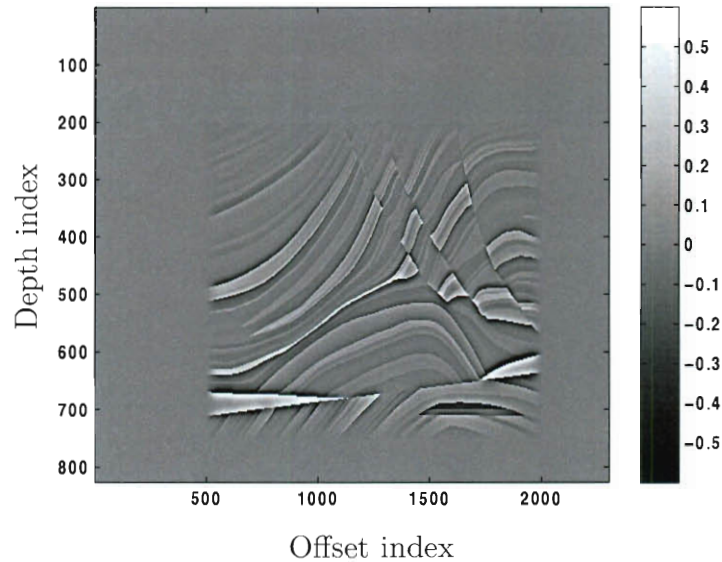
Figure 4.1: $\mathbf{m}_{true}$, true model.
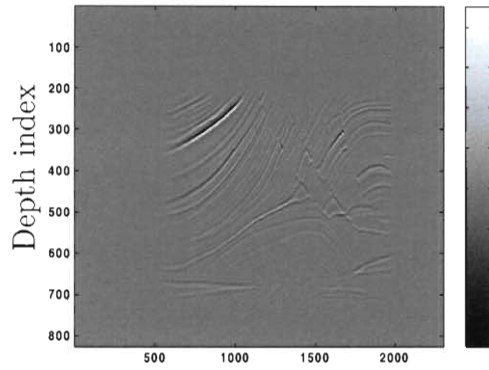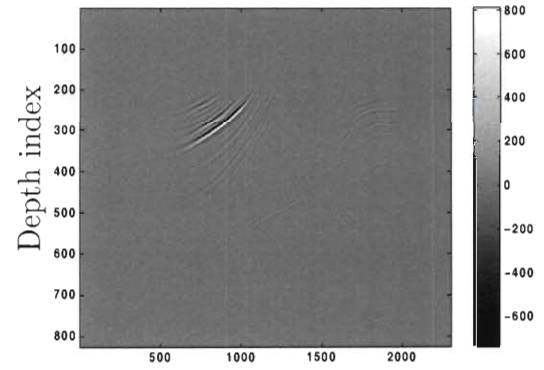
Figure 4.2: $\mathbf{m}_{mig} = F^*d$, migrated image.

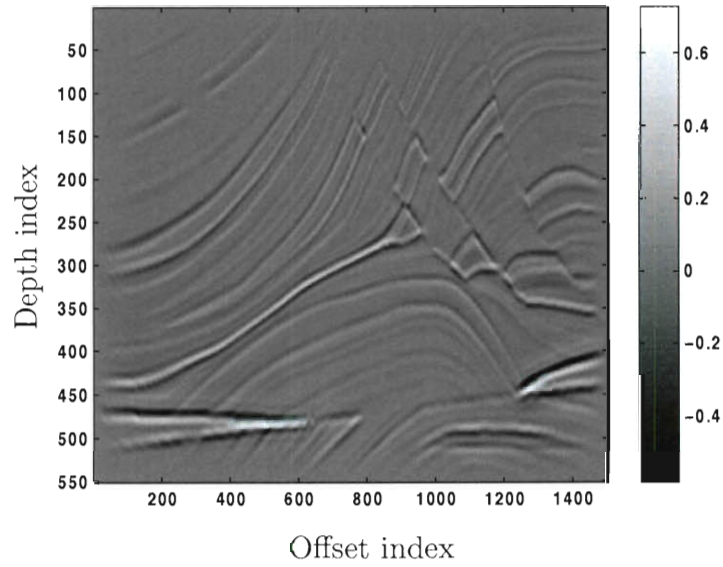Figure 4.3: $\mathbf{m}_{remig} = F^*F\mathbf{m}_{mig}$, remigrated image.

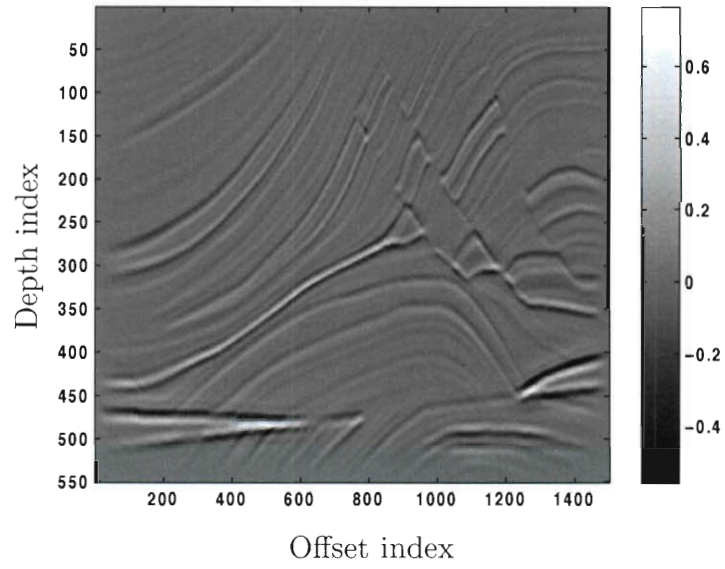Figure 4.4: Inverted model using scaling with $K = 1$

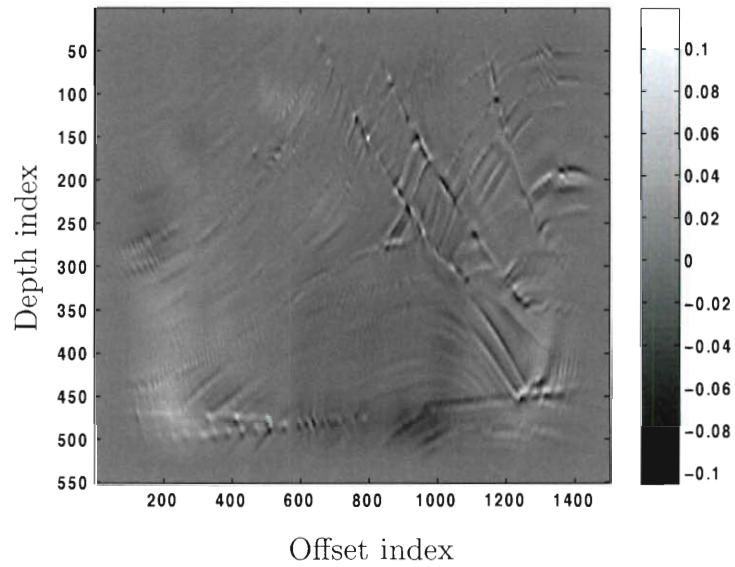Figure 4.5: Inverted model using scaling with $K = 5$

Figure 4.6: Difference between scaling with $K = 5$ and $K = 1$

4.3 Two-parameter case: p=2

The restriction to two-parameters is appealing because it is the first instance of multi-parameter inversion. It turns out that it is also particularly simple and elegant.

In this case, denote

$$N = \begin{pmatrix} N_{11} & N_{12} \\ N_{12} & N_{22} \end{pmatrix}.$$

Its adjugate is then given by

$$Adj(N) = \begin{pmatrix} N_{22} & -N_{12} \\ -N_{12} & N_{11} \end{pmatrix},$$

and

$$Adj(N)\,b = \begin{pmatrix} N_{22}b_1 - N_{12}b_2 \\ -N_{12}b_1 + N_{11}b_2 \end{pmatrix}. \tag{4.2}$$

The aim is to express (4.2) as a combination of permutations of indices of b and applications of N. We will here introduce a notation that allows the manipulation of expressions like (4.2). While it seems too involved for $p = 2$, it will simplify the manipulation for $p \geq 3$. It is mostly an exercise in representation theory.

Denote,

$$N_{ij}b_k e_l = e_l e_i^T N e_j e_k^T b := lijk. \tag{4.3}$$

Note that the symmetry of N allows us to commute i and j if needed.

The power of this representation becomes apparent in the following manipulation:

$$Adj(N)\,b := 1221 - 1122 - 2121 + 2112$$

$$= 1221 - 1212 - 2121 + 2112$$

$$= (12 - 21)(21) + (21 - 12)(12)$$

$$= (12 - 21)(21 - 12)$$

(4.4)

In the first step we commute indices, and afterwards we are using matrix factorization. The end result is interpreted as:

$$Adj(N)\,b := (12 - 21)(21 - 12)$$

$$= (e_1 e_2^T - e_2 e_1^T)N(e_2 e_1^T - e_1 e_2^T)b$$

$$= -JNJ\,b = J^T NJ\,b,$$

(4.5)

where

$$J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Equation (4.5) implies that the application of the adjugate on the migrated image requires one application of N and another to approximate the scaling factor. Which brings the cost of approximating the inverse for $p = 2$ to two applications of the normal operator.

4.4   Setup for numerical experiments

The experiments all start with a header file specifying the geometry, background fields and perturbations. The header file is common to all experiments. Since the examples are layered, only one shot is required. The model extends around 1.7 km in depth and 6.5 km horizontally. One source is put in the middle and receivers are laid out to created an offset ranging $[-2.7$ km, 2.7 km$]$.

The isotropic point source wavelet is a 2.5-5 15-20 trapezoidal bandpass filter, used in the first set of experiments. Another high frequency 5-10 40-50 trapezoidal bandpass filter is used for the set of Marmousi examples.

The models are on a 181 by 331 grid with grid spacing 10 m by 20 m for the low frequency source. For the high frequency source the models are re-sampled to a 361 by 801 grid with a 4 m by 8 m grid spacing to limit grid dispersion.

PML absorbing boundary conditions are imposed on all four sides of the model.

When a homogeneous background is used, the velocity is $vp = 2km/s$, and density is $dn = 2000kg/m^3$. Otherwise a smooth version of the Marmousi model is used (see figures 4.15 and 4.16).

To apply the linearized forward modeling operator F and the migration operator F^* (reverse time migration) required to apply the normal operator $N = F^*F$, we use a (2,4) staggered grid finite difference code based on IWAVE (Sun and Symes, 2010b,a; Levander, 1988).

4.4.1 Application: Layered variable density acoustics

As a first application to the two parameter inversion, we construct a variable density acoustics model perturbation consisting of a velocity layer and a density layer in a different place (see Figures 4.7 and 4.8). The choice of oscillatory layers rather than jump discontinuities in the material parameter in this example makes it resemble the underlying theory, which is derived for the action of pseudodifferential operators on oscillatory wave packets (see the asymptotic expansion lemma (2.11), for example). The background model is homogeneous, with $vp = 2\,km/s$ and $dn = 2000\,kg/m^3$.
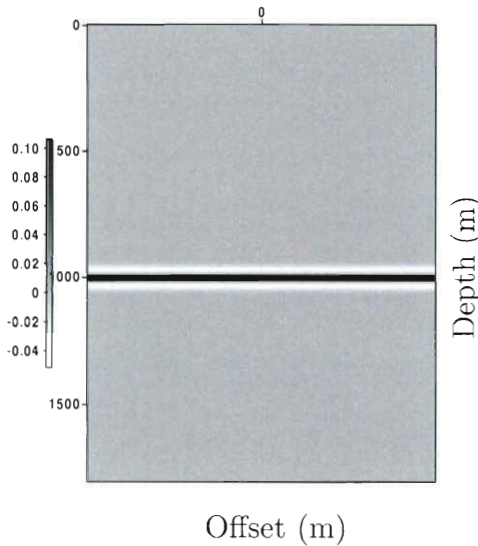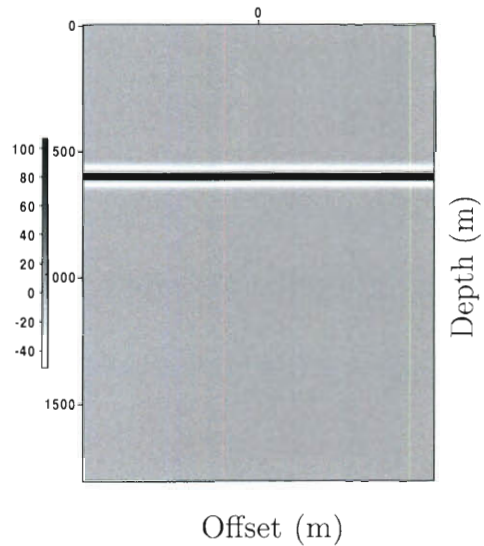

Figure 4.7: vp, velocity perturbation

Figure 4.8: dn, density perturbation

Migrating the model perturbation shows how migration mixes the effects of the two models in the two components of the migrated images (Figure 4.9). We shall refer to the migrated images as b_1 and b_2, to remain consistent with our notation

where the vector of migrated images is b. This example, albeit simple, stresses a new challenge of multi-parameter inversion: For one parameter inversion, the events in the migrated image corresponded to events in the true model. In multi-parameter inversion, events in the migrated images may correspond to an event in one or more of the components of the model. It is virtually impossible to tell that these migrated images correspond to a model with separate events for velocity and density without successful inversion. Applying the scheme outlined above, we form
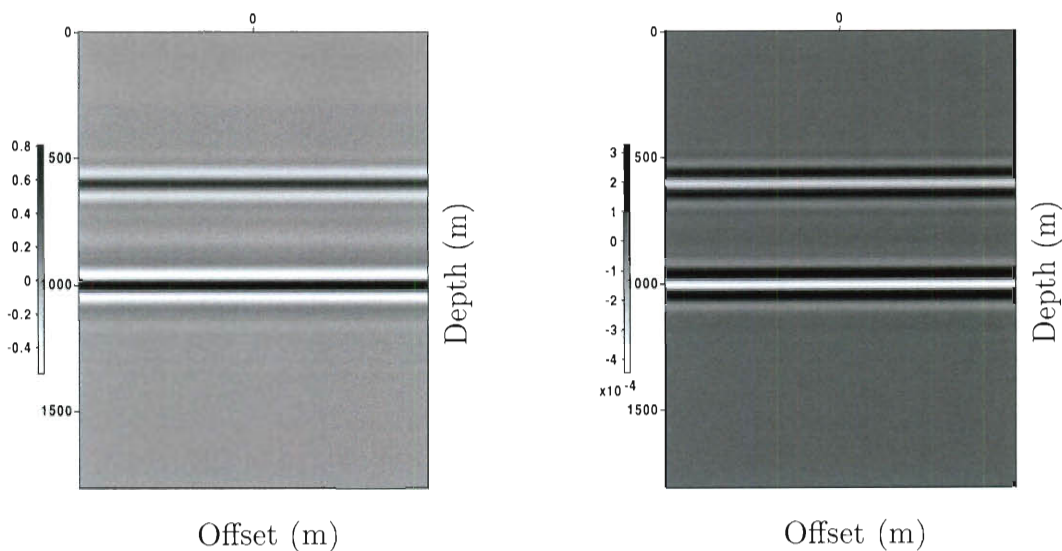
b_1, velocity component of the migrated image.

b_2, density component of the migrated image.

Figure 4.9: Migrated images mixing the contributions from density and velocity, and effecting a phase space scaling.

$$J^T N J b \approx det(N)\, x.$$

The result is shown in Figure 4.10, and shows how one application of the normal operator effectively separated the contributions of the velocity and density events. It remains to effect an amplitude correction, by approximating an inverse to $det(N)$. For this end, we are required to form $N\, det(N)\, x \approx det(N)\, b$, shown in Figure 4.11.
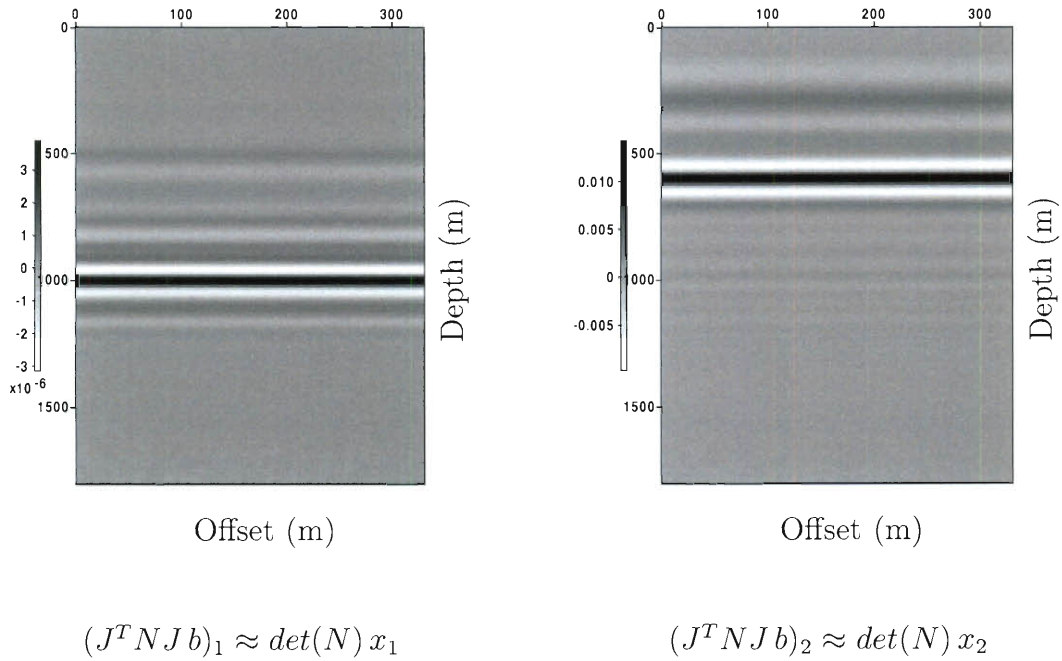

$$(J^T N J b)_1 \approx det(N)\, x_1 \qquad\qquad (J^T N J b)_2 \approx det(N)\, x_2$$

Figure 4.10: The application of the adjugate separates the velocity and density contributions. This result is a phase space scaling of the true model.

The final step corrects the amplitudes of $det(N)\, x$ by undoing the effect of $det(N)$, which yields an approximate inverse. This final step complements the separation we obtained earlier with an amplitude correction. Figure 4.12 shows that the approx-
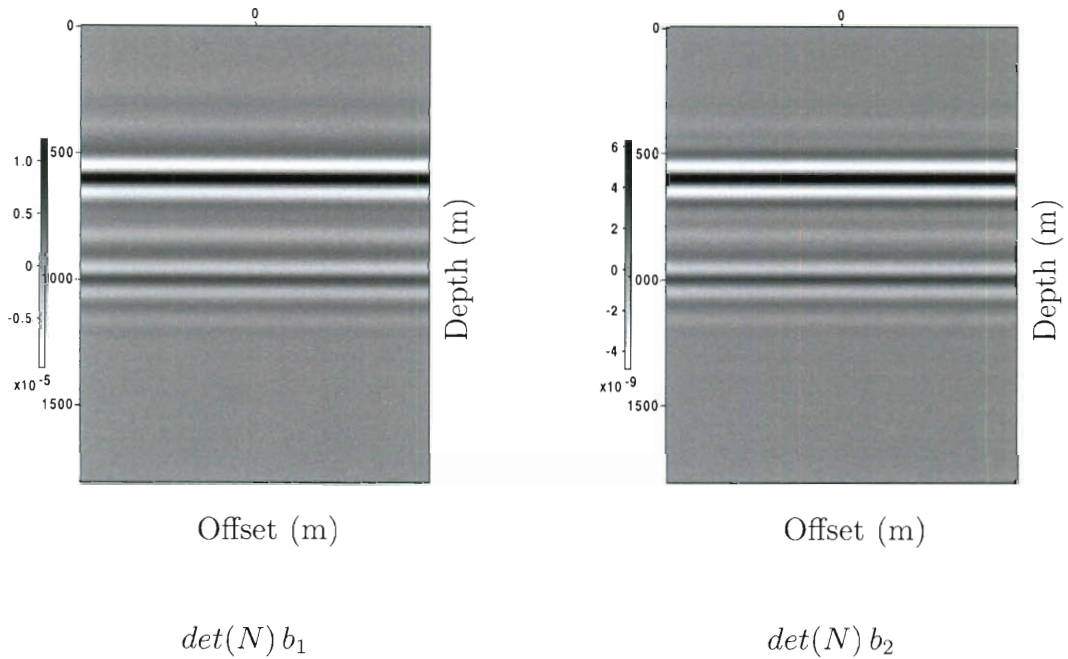
Figure 4.11: Scaling of the migrated images by $det(N)$, used to undo the determinant

imate inverse compares favorably with the true model. An interesting observation about this result is the fact that the velocity model is better recovered than the density model: traces of the velocity event in the density model are more apparent than that the density event in the velocity model. This observation is in accordance with the theoretical fact that the recovery of velocity in variable density acoustics is better conditioned than the recovery of density. Please consult the section about the conditioning of the normal operator at the end of this chapter for an explanation of this fact.
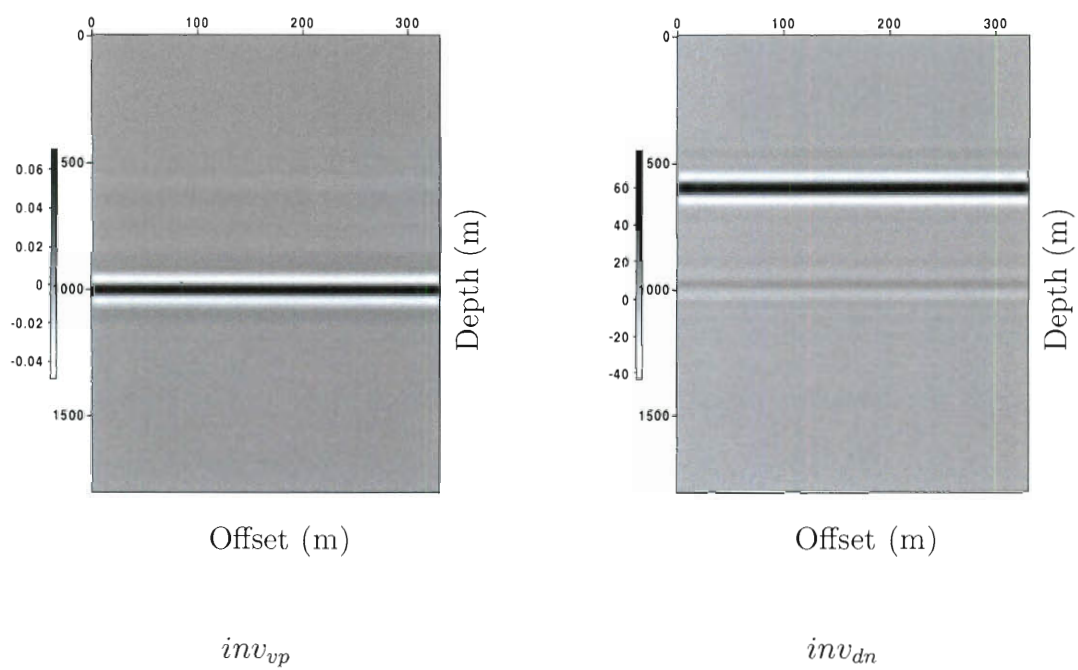
inv_{vp} inv_{dn}

Figure 4.12: The approximate inverse. The contributions from velocity and density are separated and the amplitudes are corrected.

4.4.2 Extended layers with smooth non-homogeneous background

To investigate the effect of finite bandwidth on the method, I create a model with two extended layers (Figures 4.13 and 4.14). This time the background model is not homogeneous; it is rather a smoothed out version of a 1D trace of the Marmousi model, shown in Figures 4.15 and 4.16.

The migrated images only detect the edges of the constant layers as expected, which is due to the finite bandwidth of the source. However, the application of the adjugate still succeeds in separating the effect of the velocity and density contributions. The approximate inverse obtained compares well with the model, there is significant loss of amplitude due to the finite bandwidth (see figure 4.19).

The source used to generate all the numerical experiments up to this one is the low frequency source described in the numerical experiments setup. If instead we use a "high frequency source", a trapezoidal bandpass filter [5 Hz, 10 Hz, 40 Hz, 50 Hz], the amplitude loss is decreased, with the high frequency source as seen in Figures 4.22 and 4.23. This explains a better fit of the data for the high frequency source than the low frequency source (see figures 4.24 and 4.25). The high frequency source fits 70% of the data, whereas the low frequency source only fits 50%.
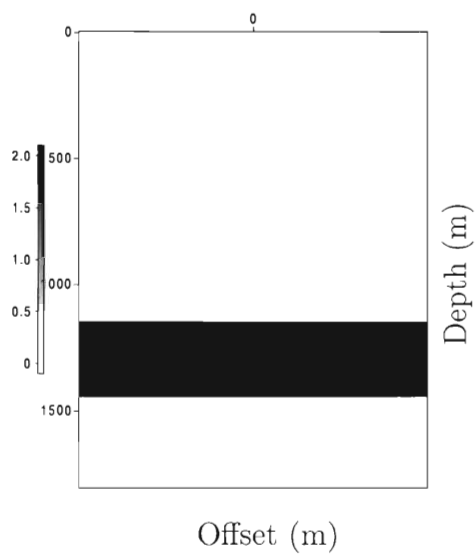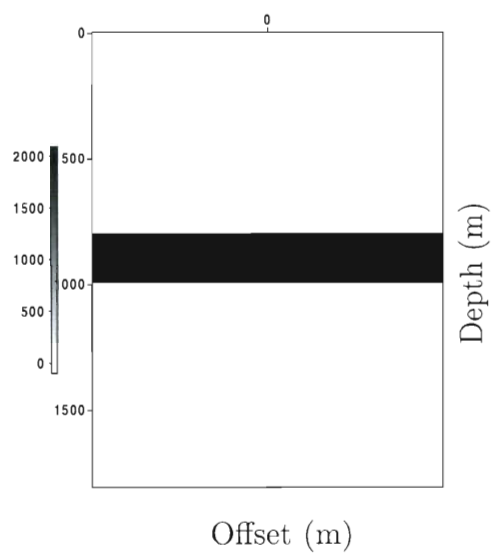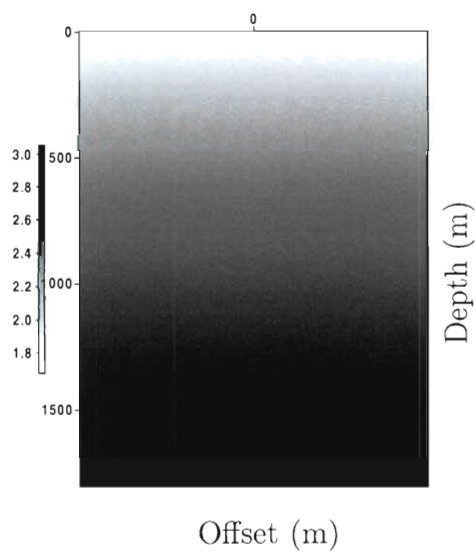
58

Figure 4.13: *vp*, velocity perturbation

Figure 4.14: *dn*, density perturbation
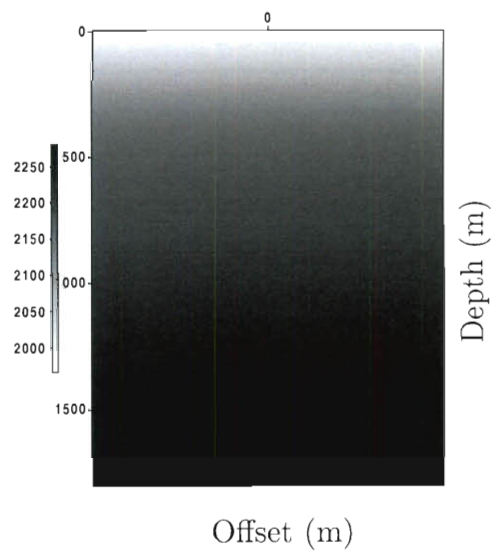

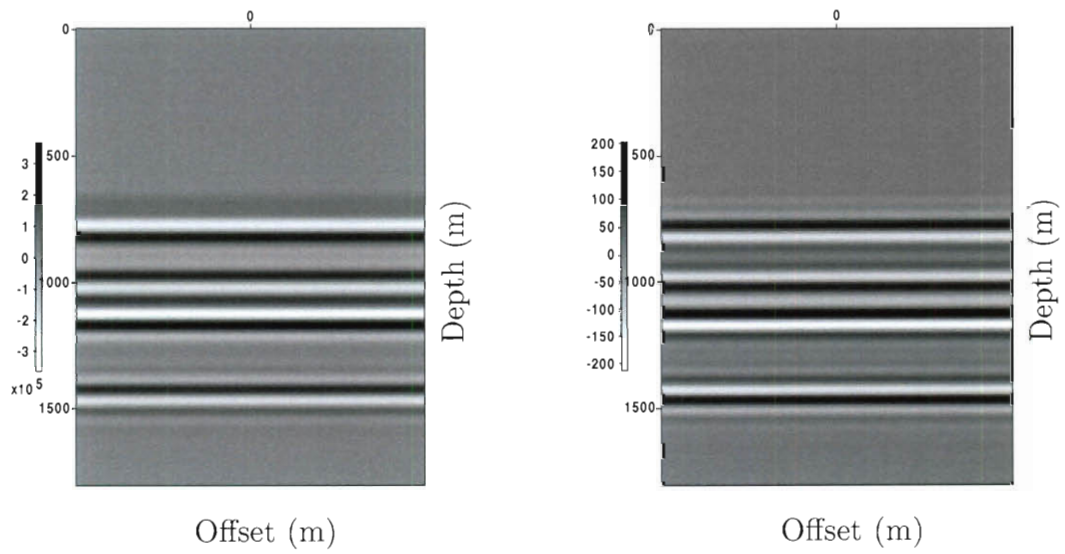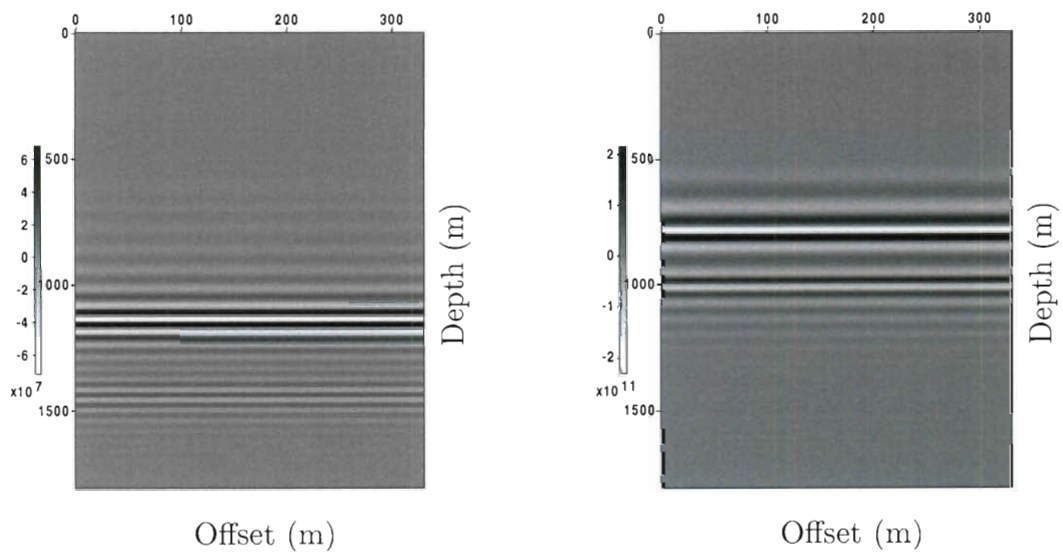Figure 4.15: smooth background velocity

Figure 4.16: smooth background density

b_1, velocity component of the migrated image.

b_2, density component of the migrated image.

Figure 4.17: Migrated images mixing the contributions from density and velocity, and effecting a phase space scaling. The migrated images only detect the edges of the layers.

$$(J^T N J b)_1 \approx det(N)\, x_1 \qquad\qquad (J^T N J b)_2 \approx det(N)\, x_2$$

Figure 4.18: The application of the adjugate separates the velocity and density contributions. This result is a phase space scaling of the true model.

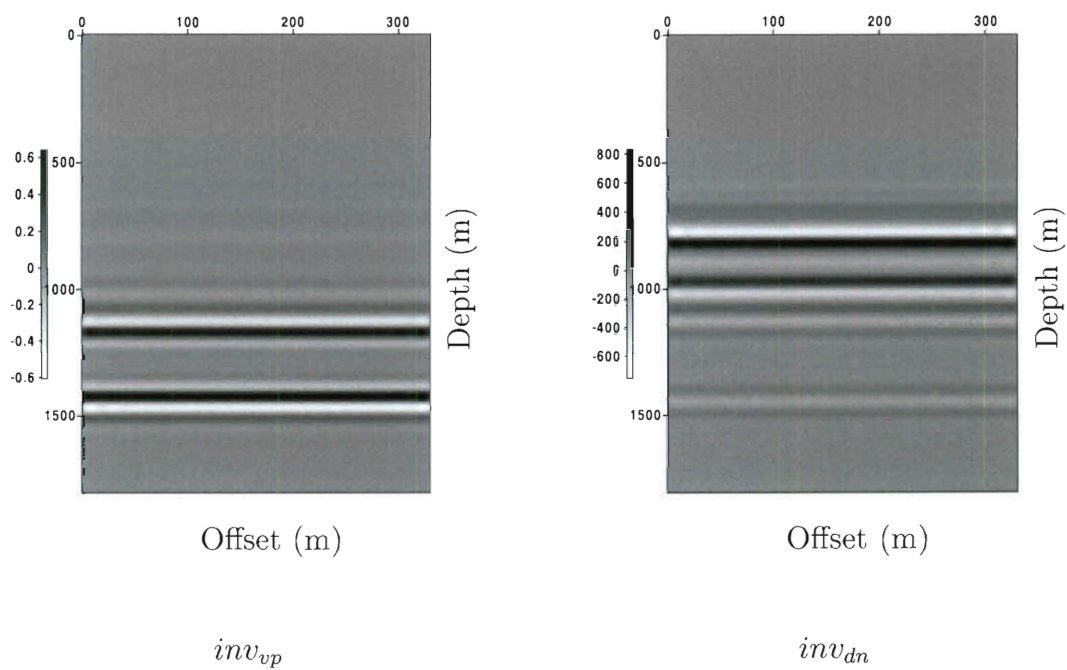inv_{vp} $\qquad\qquad\qquad\qquad\qquad\qquad$ inv_{dn}

Figure 4.19: The approximate inverse. Shows how the contributions from velocity and density are separated. The loss of amplitude is due to the low frequency source.
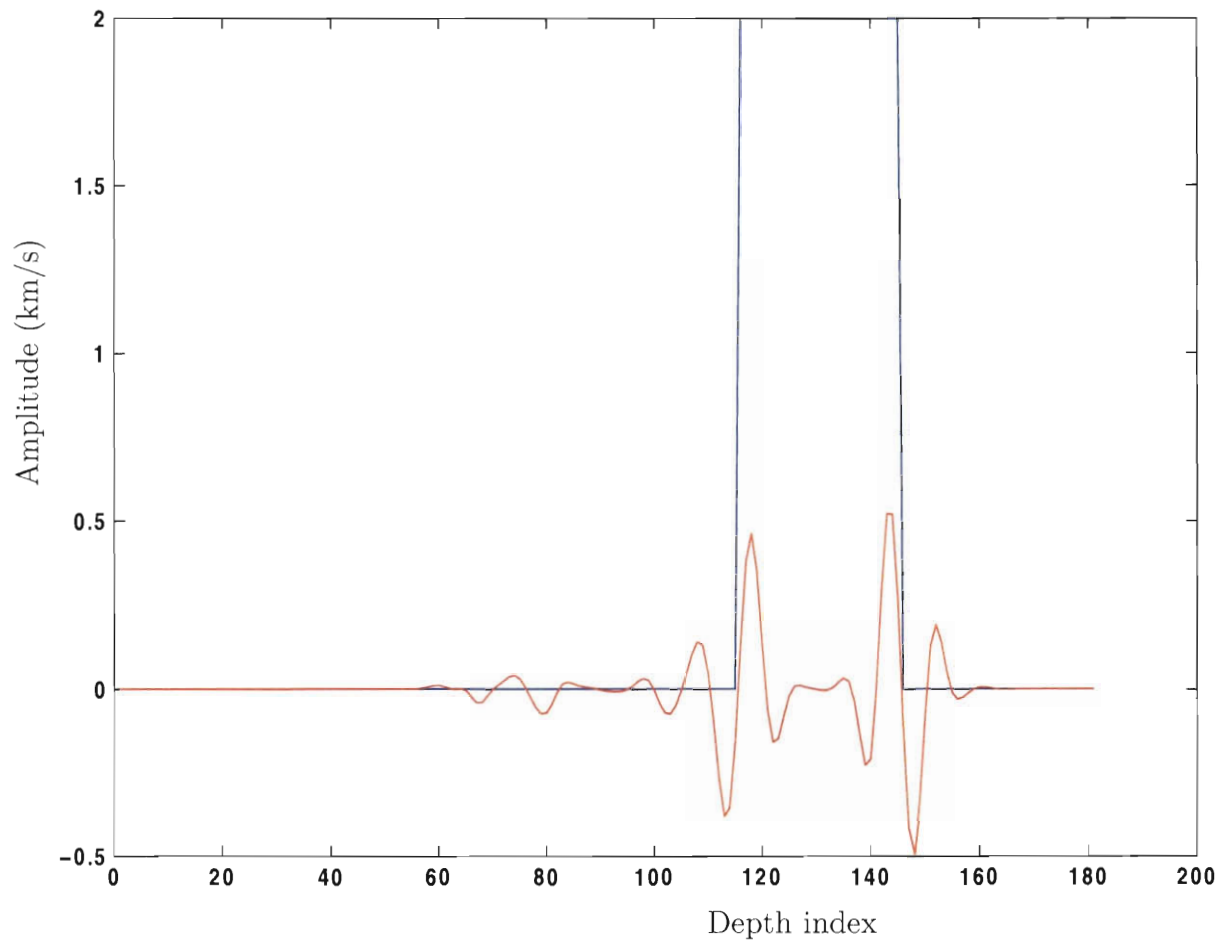
Figure 4.20: Inverted velocity versus real velocity; note the amplitude loss due low frequency source.
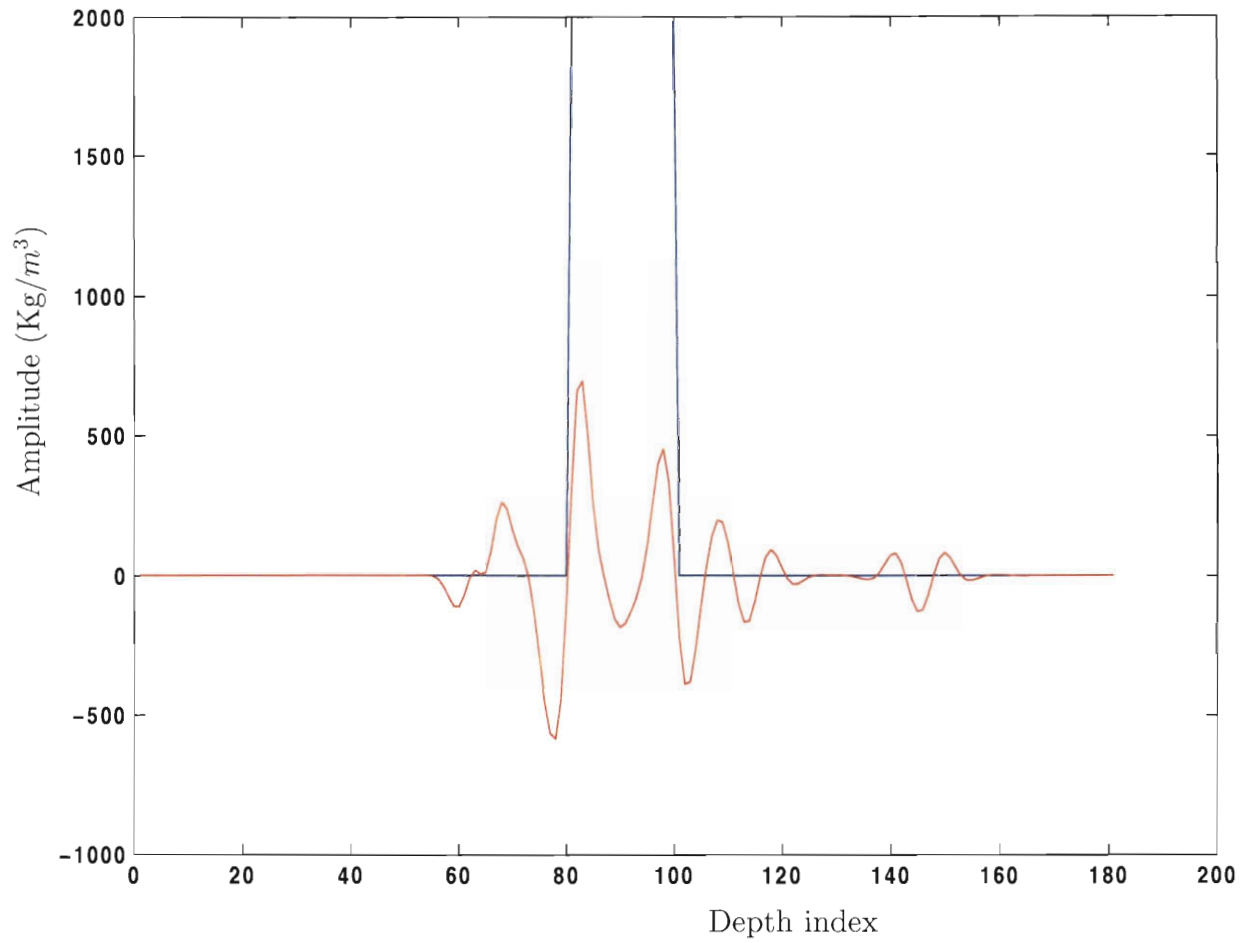
Figure 4.21: Inverted density versus real density; note the amplitude loss due to low frequency source.

Figure 4.22: Inverted velocity versus real velocity; note the reduced amplitude loss with a high frequency source.
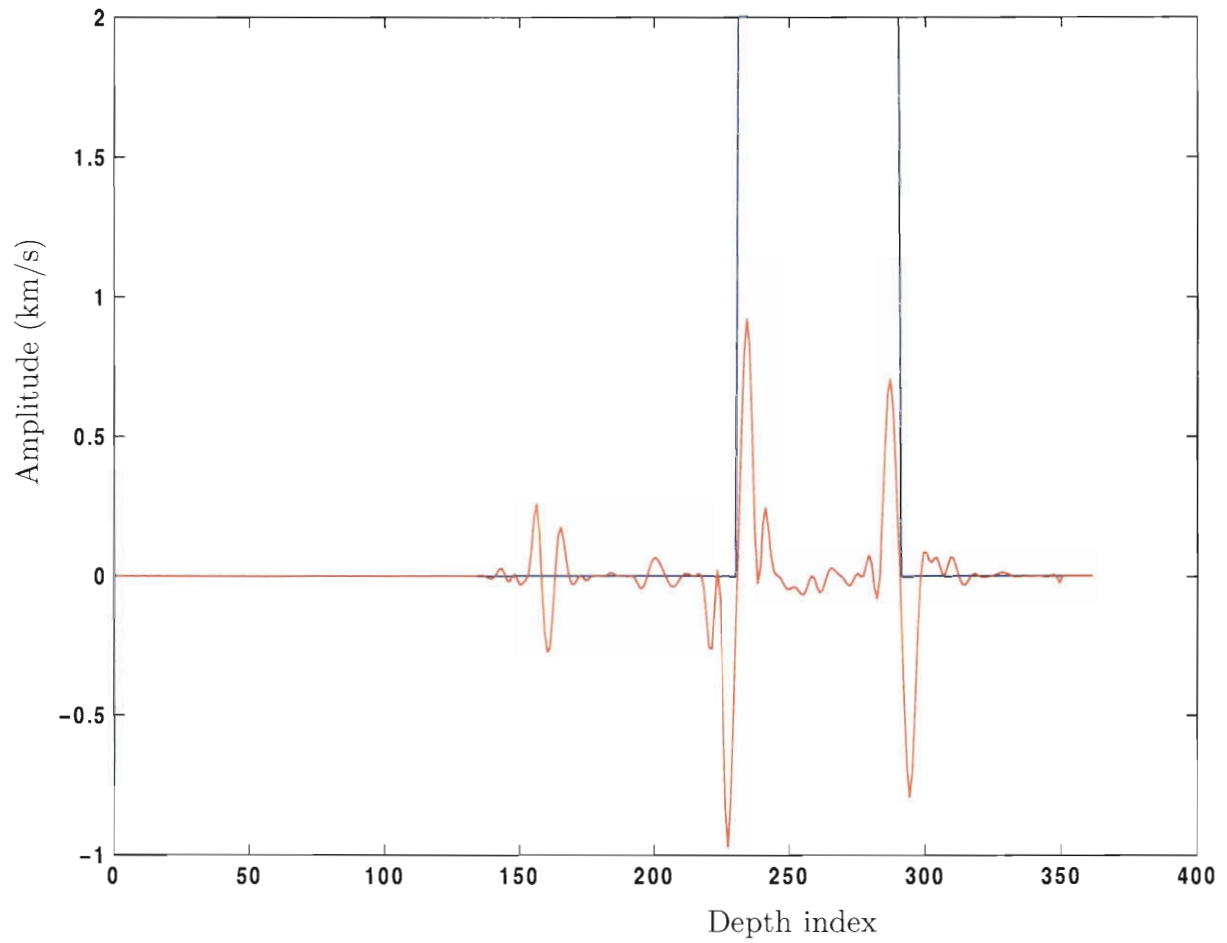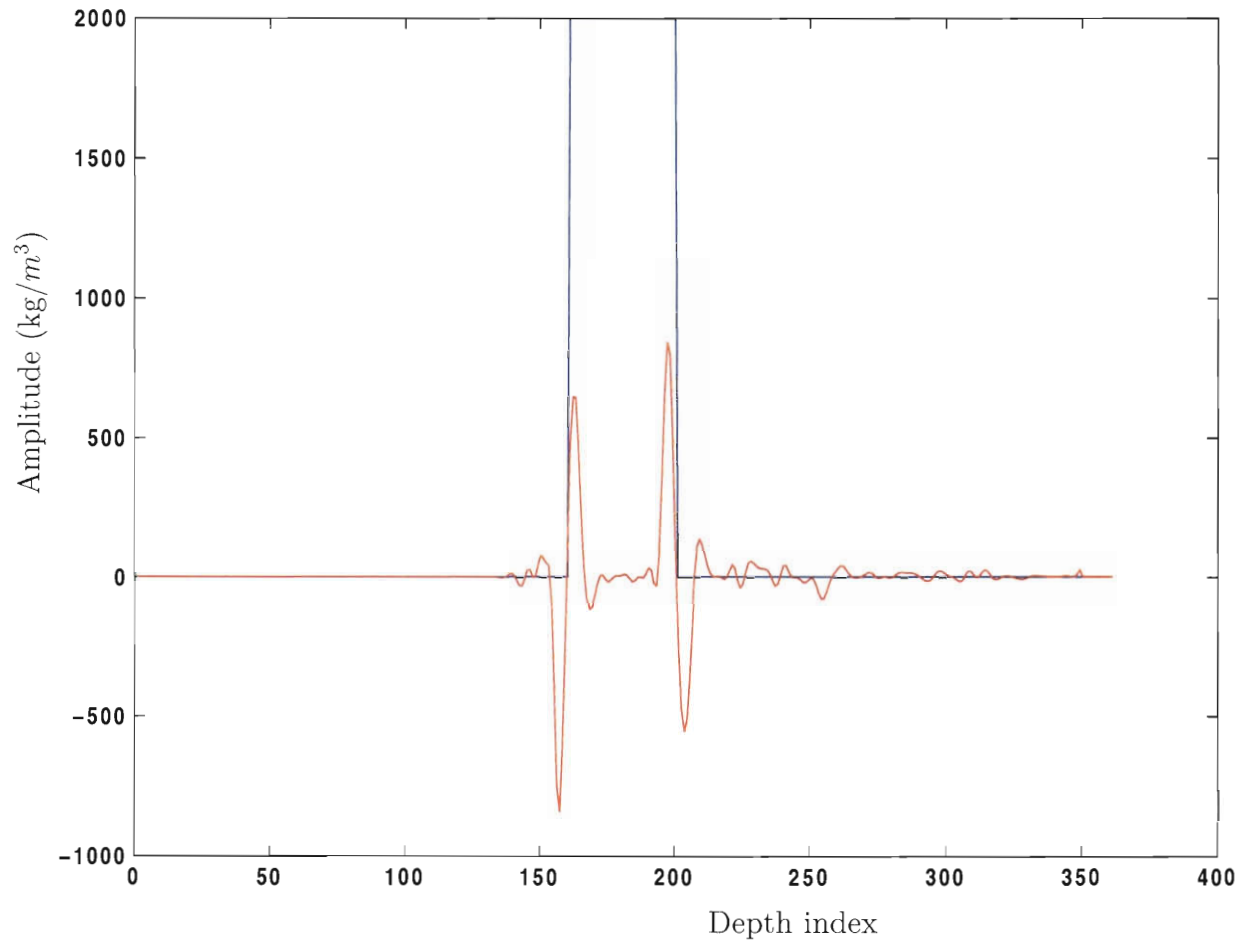
Figure 4.23: Inverted density versus real density; note the reduced amplitude loss with a high frequency source.
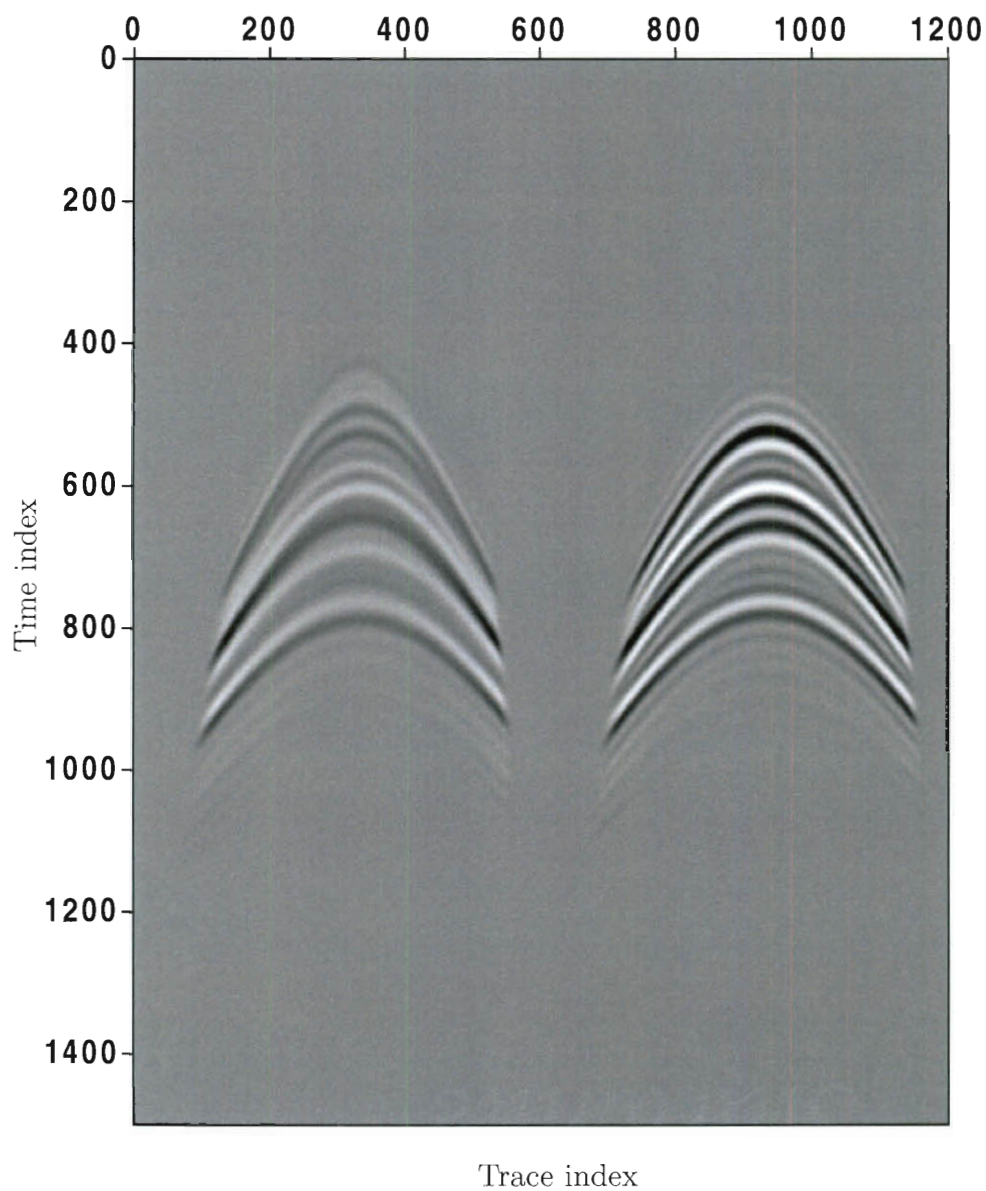
Figure 4.24: The difference between the predicted data and the actual data for low frequency source (left), plotted on the same scale as the data (right) (50% data fit). The data fit is not as good as the high frequency source, explaining the amplitude

Figure 4.25: The difference between the predicted data and the actual data for high frequency source (left), plotted on the same scale as the data (right) (70% data fit). This gives a better data fit than the low frequency source, explaining the reduced amplitude loss in the inverted model.

4.4.3   Marmousi model with homogeneous background

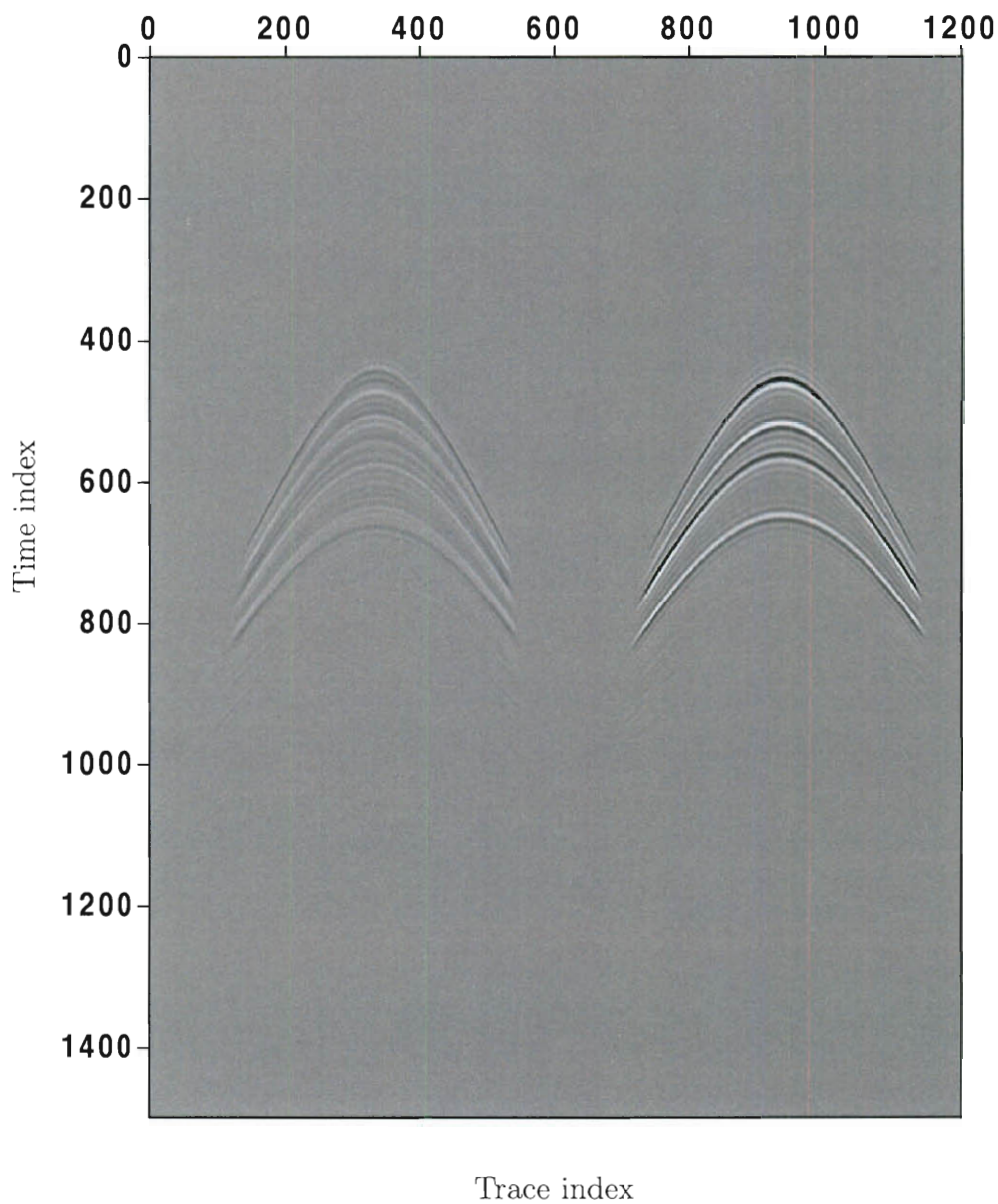The next test involves a complex model, a 1D section of the Marmousi model with a constant background. I zero out part of the velocity perturbation around 1000 m depth and the density perturbation around 700 m (see figures 4.26 and 4.27, to obtain a qualitative test like the separation test discussed above. The complexity of the model requires a high frequency source to be able to resolve the details and obtain an accurate result.

The inverted velocity and density results are shown and compared to the real velocity and density in figures 4.31 and 4.32. The inverted velocity is better than density, as expected. The ill-conditioning of density inversion is unavoidable for complex models, as explained in the last section of this chapter. The inverted velocity passes our qualitative test and is zero around the zero inclusion at 1000 m; the density tends to do the same but with considerably less accuracy.

The inverted result predicts the data well; the difference between the predicted data and the actual data are plotted with the same scale in figure 4.33. The difference is considerably fainter than the data, showing that the predicted data fits the bulk of the data. The inverted model fits 60% of the data.
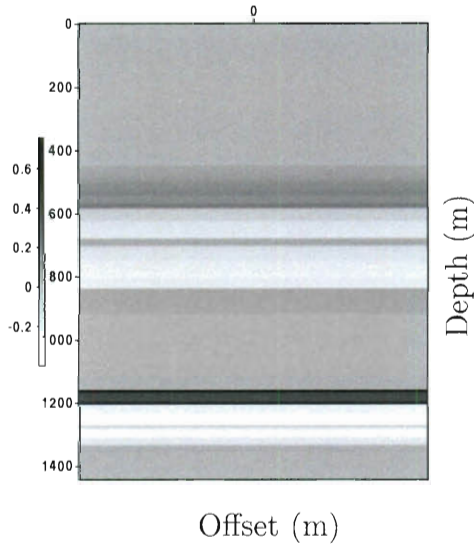
Figure 4.26: vp, velocity perturbation
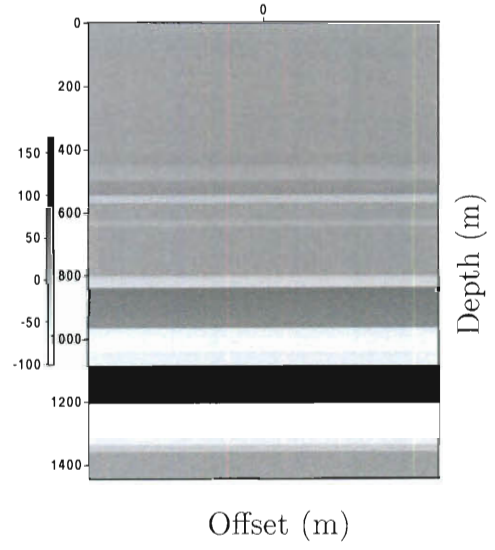
Figure 4.27: dn, density perturbation
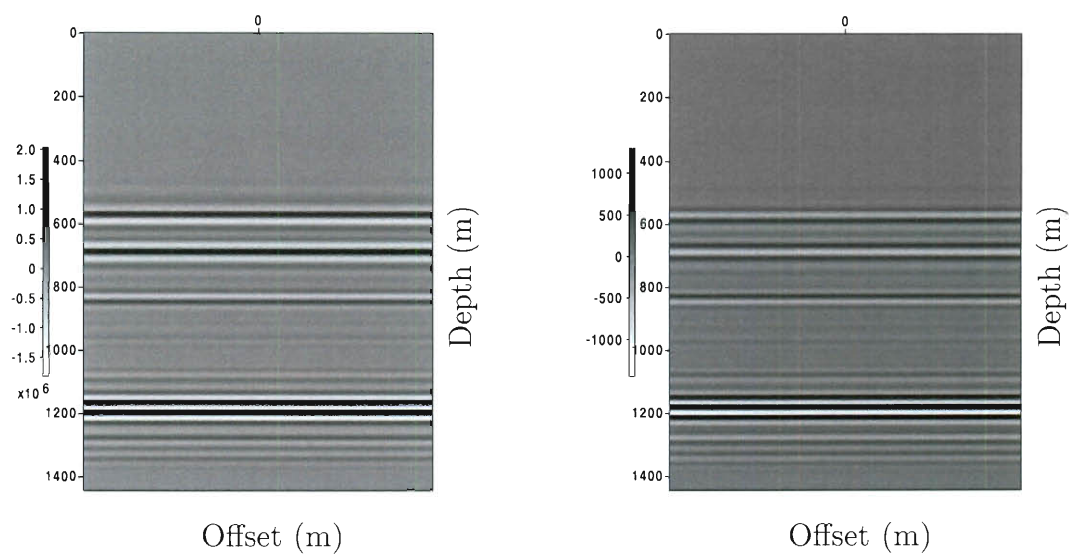
b_1, velocity component of the migrated image.

b_2, density component of the migrated image.

Figure 4.28: Migrated images mixing the contributions from density and velocity, and effecting a phase space scaling. The migrated images only detect the edges of the layers.
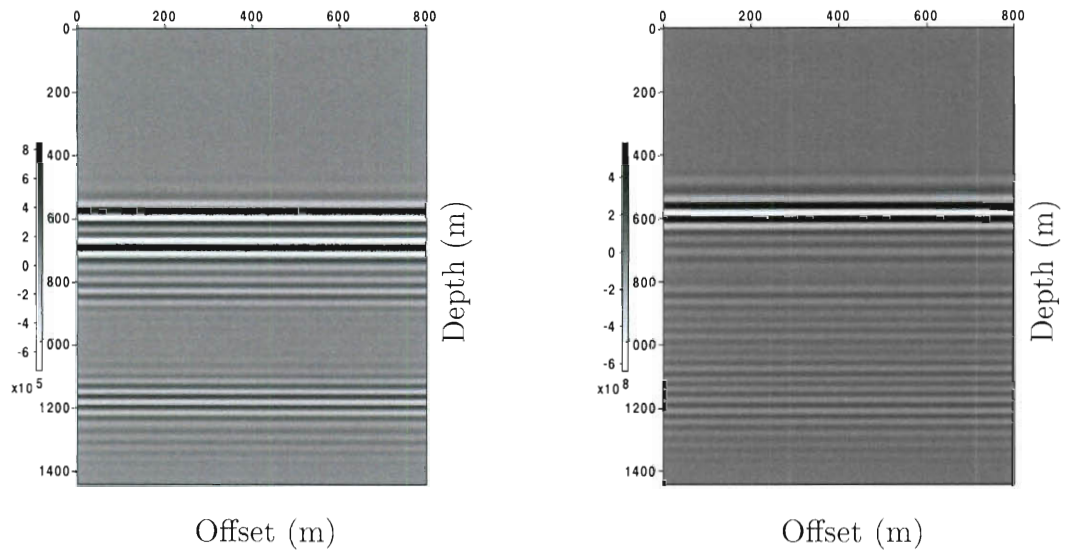
$$(J^T N J b)_1 \approx det(N)\, x_1 \qquad\qquad (J^T N J b)_2 \approx det(N)\, x_2$$

Figure 4.29: The application of the adjugate separates the velocity and density contributions reasonably well for the velocity; the density result is worse. This result is a phase space scaling of the true model.

inv_{vp} inv_{dn}
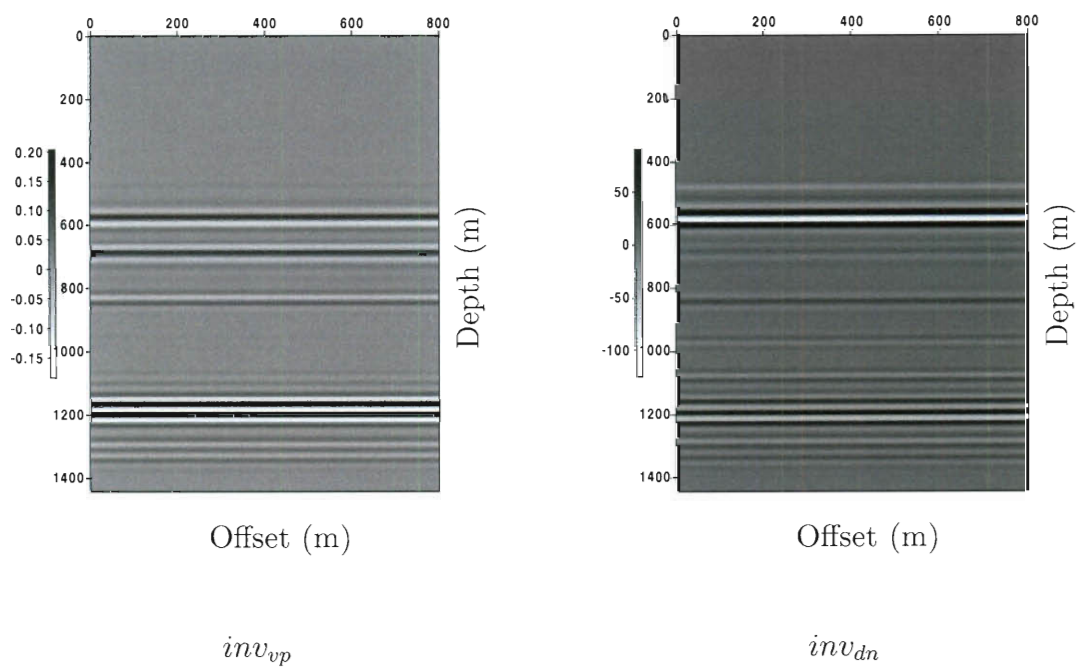
Figure 4.30: The approximate inverse.
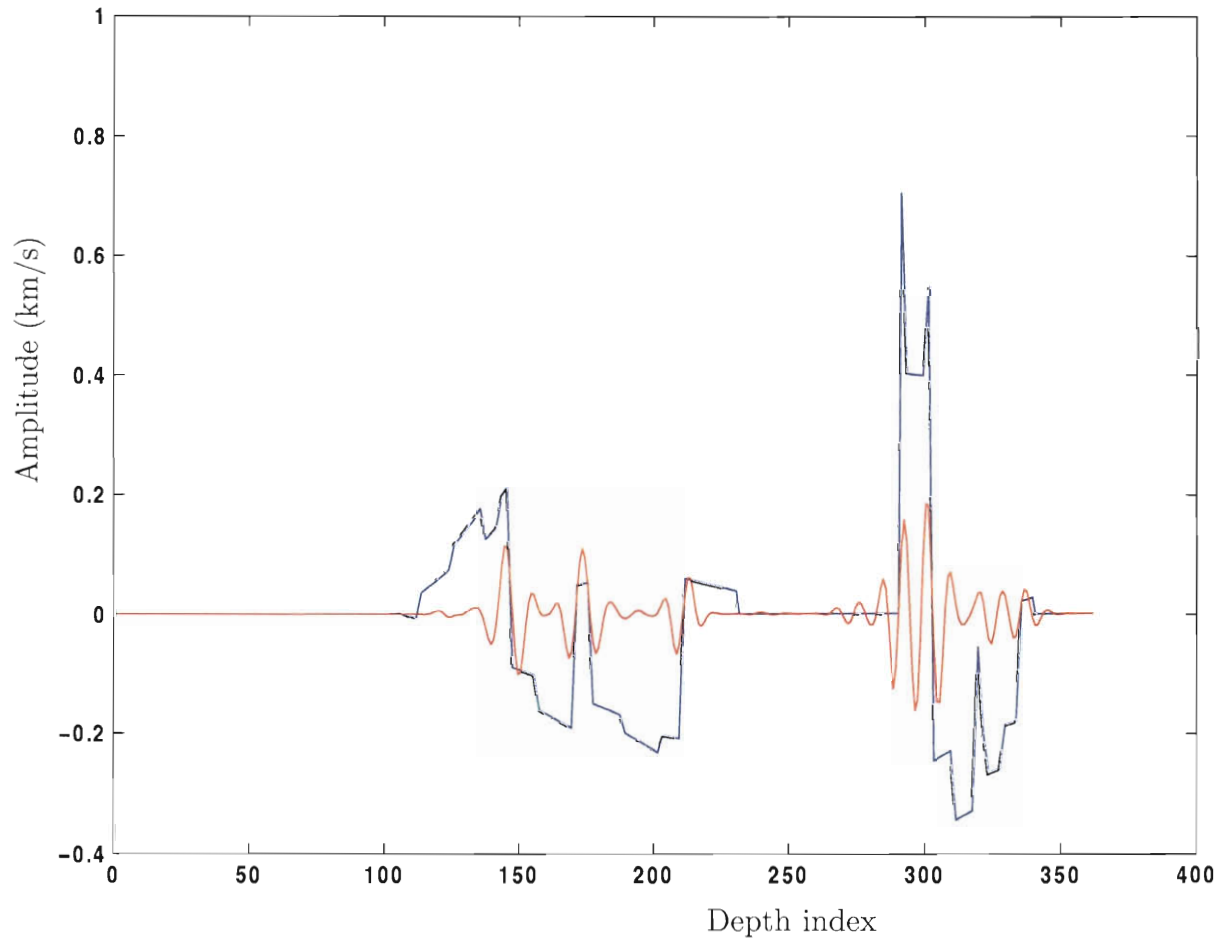
Figure 4.31: Inverted velocity versus real velocity. Notice how the inverse compares well to the real velocity and achieves separation around the inclusion around depth index 250 (1000 m).
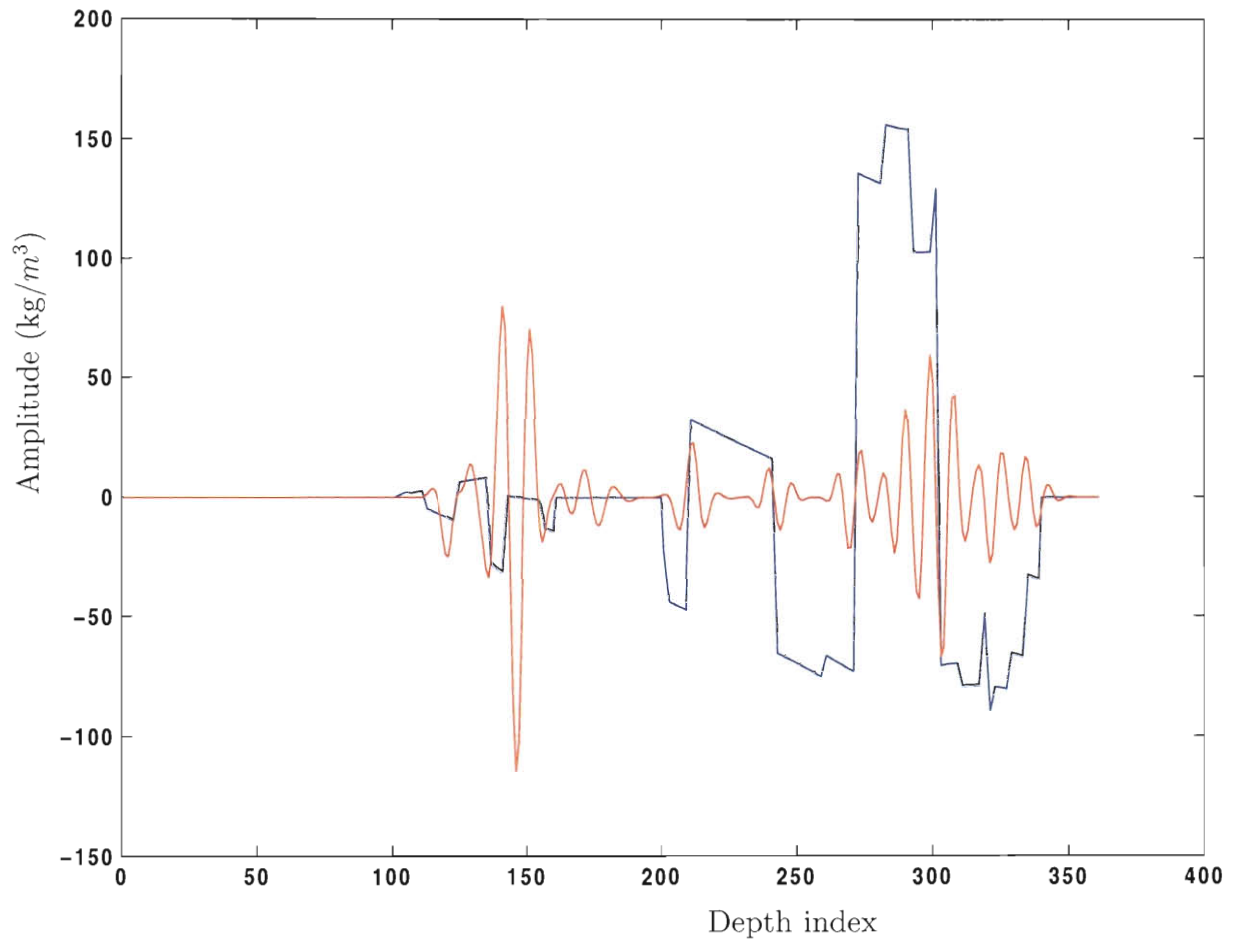
Figure 4.32: Inverted density versus real density. The result is worse than the velocity result; the ill-conditioning in the density inversion is unavoidable for complex models.
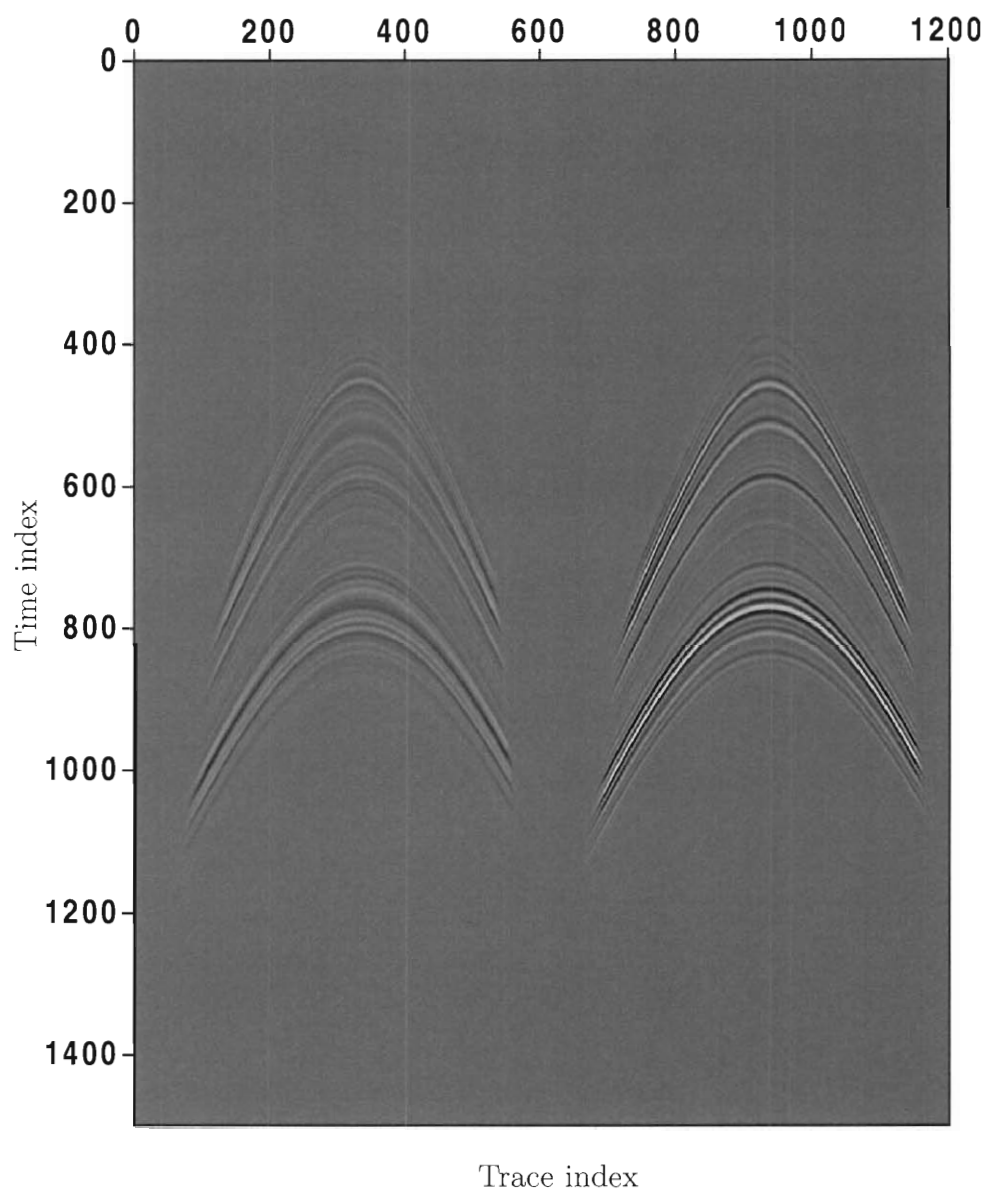
Figure 4.33: The difference between the predicted data and the actual data (left), plotted on the same scale as the data (right) (60% data fit). Note how the difference is considerably fainter; the inverted model succeeds in predicting the bulk of the data.

4.4.4   Marmousi model with smooth background

The same velocity and density perturbations used in the previous section are used here. This time, however, the background velocity and density fields are smooth rather than homogeneous (see figures 4.15 and 4.16). The background velocity increases linearly, which causes the production of diving waves (refracted waves). The existence of these waves is forbidden by the underlying theory: the normal operator is pseudodifferential for purely reflective data. The parts of the data which contain diving waves have to be removed carefully to obtain sensible results, which we do as part of the pre-processing.

One can see that the velocity inversion (figure 4.39) is comparable to the previous section. The inverted density, however, is less accurate (figure 4.40). The data fit is still satisfactory for the near offset data. It becomes worse for the far offset data (see figure 4.41, residual error in the data versus the data, plotted on the same scale). This is due to the density fit. The velocity events control the amplitude of small offsets, whereas the density events predict the amplitude of large offset data. The inverted model in this case is only able to fit 40% of the data.
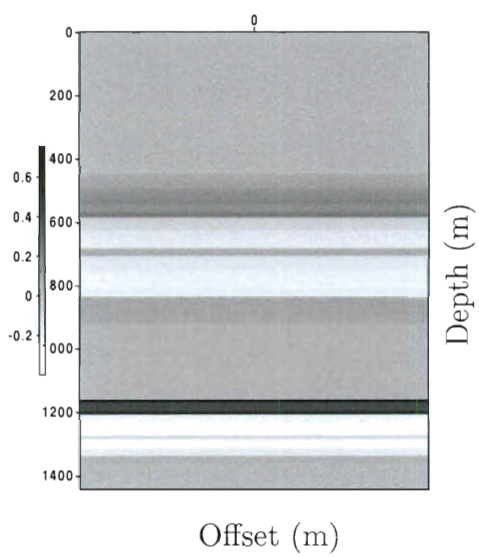
Figure 4.34: *vp*, velocity perturbation
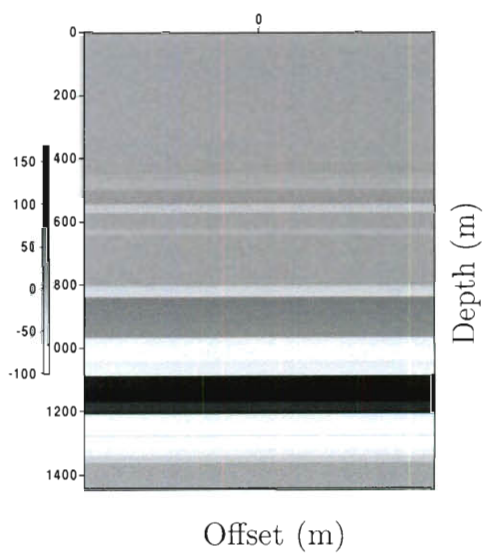
Figure 4.35: *dn*, density perturbation
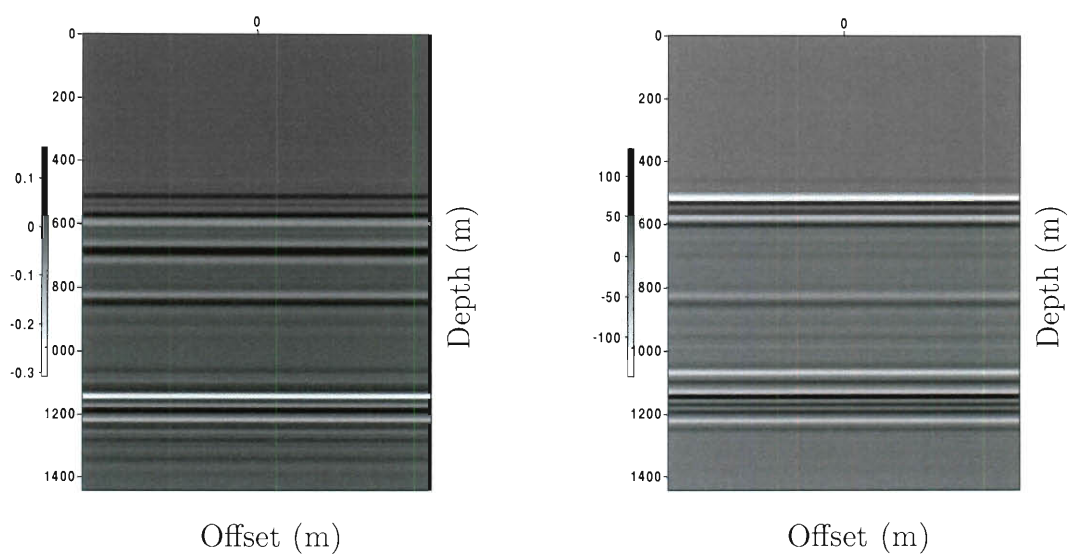
b_1, velocity component of the migrated image.

b_2, density component of the migrated image.

Figure 4.36: Migrated images mixing the contributions from density and velocity, and effecting a phase space scaling. The migrated images only detect the edges of the layers.
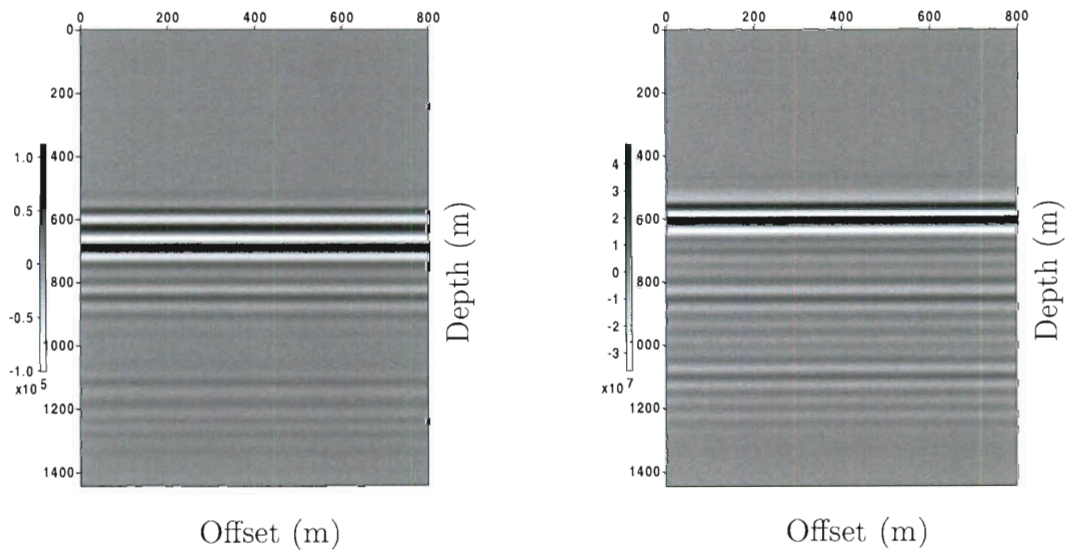
$$(J^T N J b)_1 \approx det(N)\, x_1 \qquad\qquad (J^T N J b)_2 \approx det(N)\, x_2$$

Figure 4.37: The application of the adjugate separates the velocity well; the density result is not as good. This result is a phase space scaling of the true model.

inv_{vp} $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ inv_{dn}
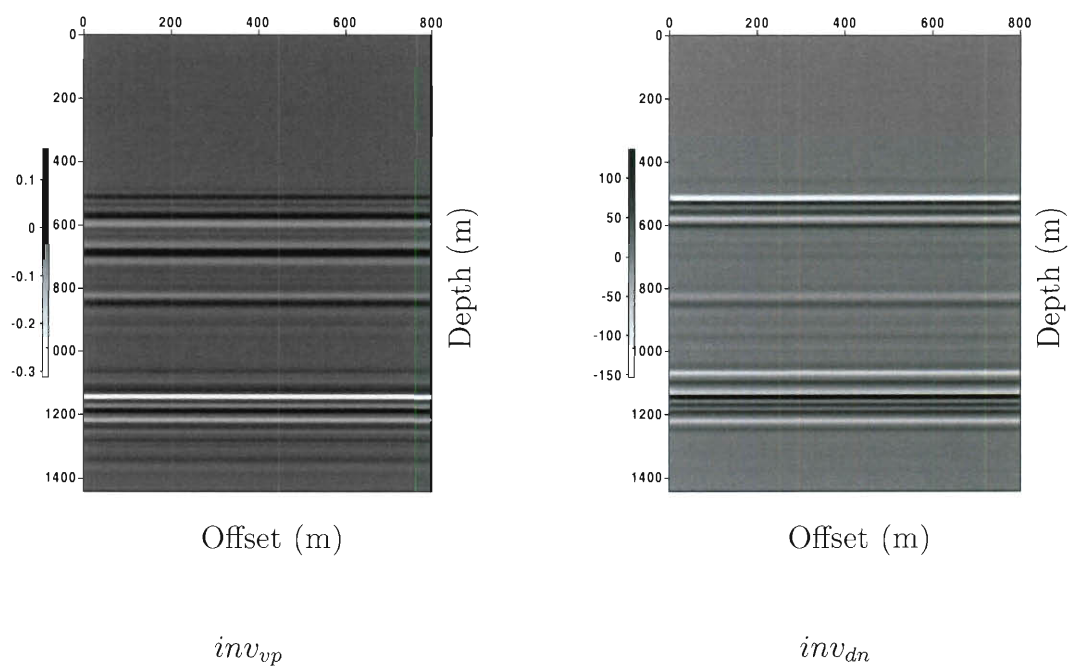
Figure 4.38: The approximate inverse: A good inverted velocity field, the density inversion is considerably worse.
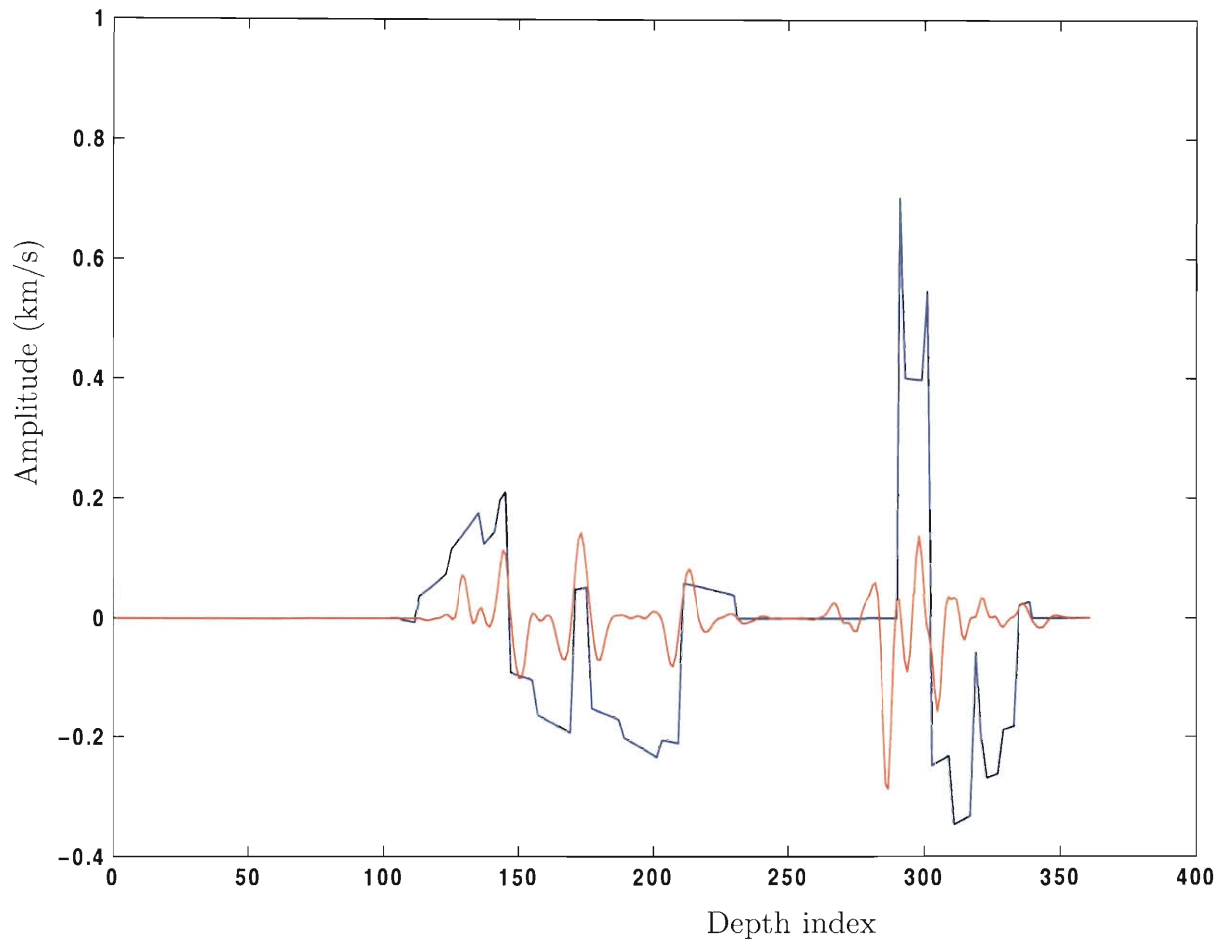
Figure 4.39: Inverted velocity versus real velocity, the method manages to pinpoint the inclusion and is good overall.
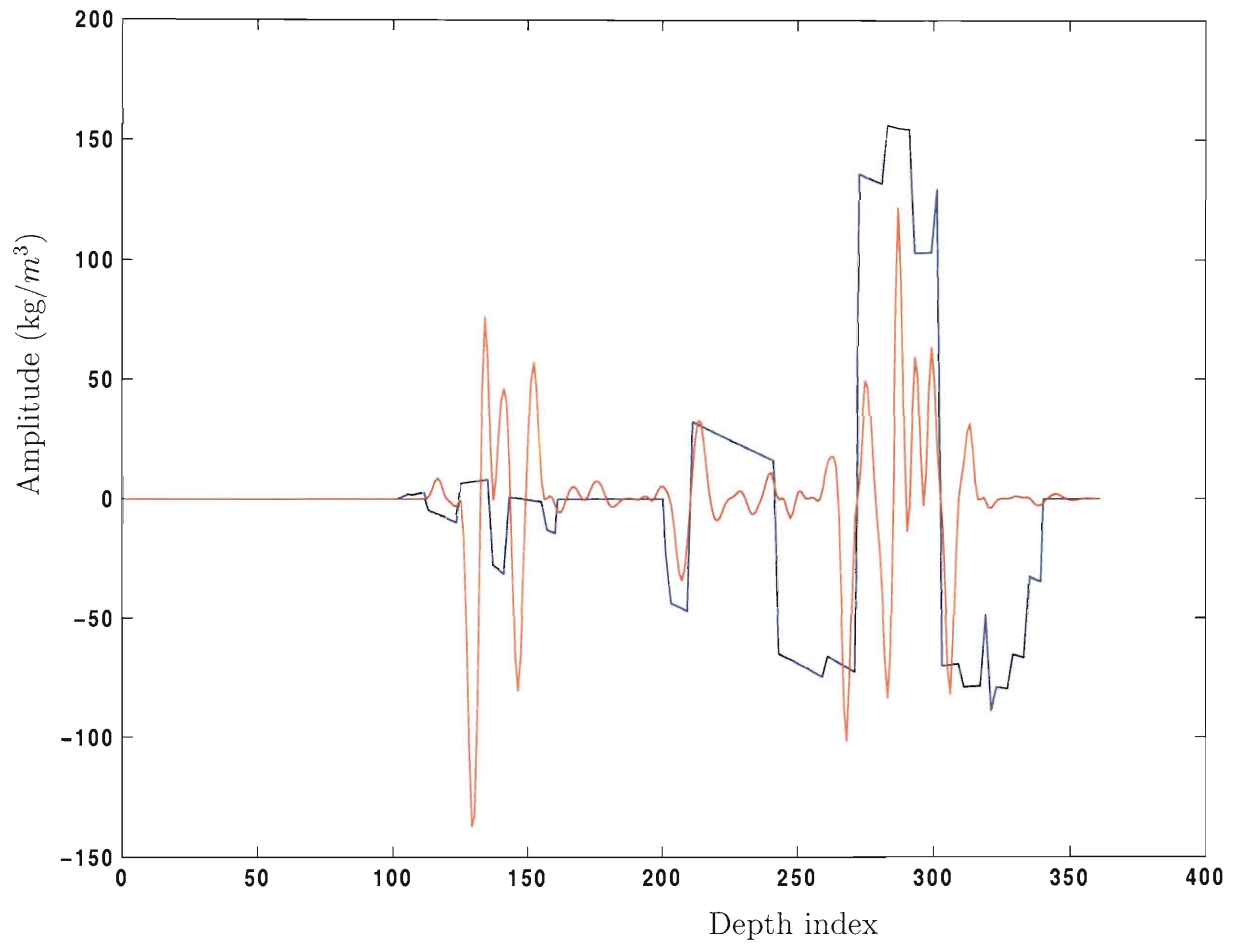
Figure 4.40: Inverted density versus real density; the result shows some high oscillations that do not conform with the real density.
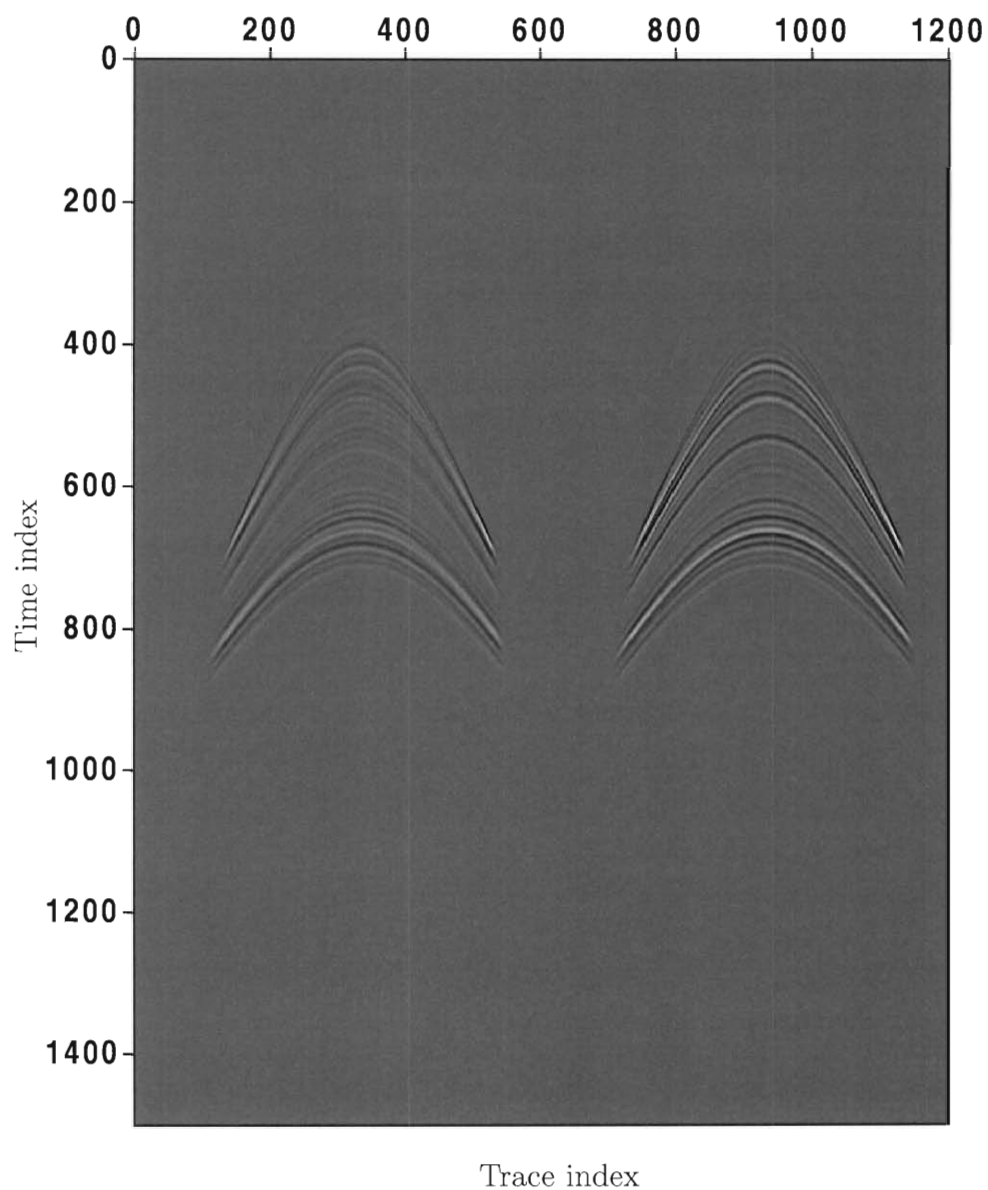
Figure 4.41: The difference between the predicted data and the actual data (left), plotted on the same scale as the data (right) (40% data fit). The inverted model explains the bulk of small offset data; the bad density fit leads to a poor fit of large

4.4.5 Layers with non-smooth background

I modify the layered example by using a non-smooth velocity background that features a reflective event around 800 m depth (see figure 4.44). Within the bandwidth considered, introducing a fast oscillation in the background model, renders it non-smooth. The background wave field thus contains a reflection from that event. The normal operator is not a pseudodifferential operator for a non-smooth background, violating an assumption of the method. The commutator error discussed in (3.23) becomes in fact larger than the $det(N) I$ term. There is no control on the commutator of general linear operators (not even matrices), that states that it is small with respect to the product of the operators. This is corroborated by the result of the application of the adjugate numerically (see figure 4.47). It is hard to infer anything from the application of the adjugate. The result is no longer a scaling of the target model. The commutator error now becomes sizable.

This test rules out the use of this method for a non-smooth background. Even for the one parameter case, the relationship between b and $N b$ ceases to be pseudodifferential and the scaling factor is not an approximate inverse. This is why the velocity has to be separated into a smooth background and perturbation at each step, as discussed in the section about full waveform inversion.

The background model being non-smooth for band limited data means that it has variations on the order of a wavelength. These experiments are done using the low frequency source. Using the background velocity field and the frequency band of

the source, we can predict a characteristic wavelength. For this specific model, with a minimum frequency of 5 Hz and an average velocity of 2300 m/s, we can predict that the characteristic wavelength is about 460 m. If the variations of the background model occur at a distance greater than that characteristic wavelength, the background is considered smooth and the method applies. If not, an assumption of the method is violated and the method fails. I repeat the same experiment as above with 4 different background velocity fields. These background fields are ramps increasing from 2 km/s to 2.6 km/s over distances ranging from 1000 m to 250 m, I refer to those functions as ramp 1 to ramp 4 (see figures 4.48, 4.49, 4.50 and 4.51). The background density is constant $dn = 2000 kg/m^3$. The variation in the background velocity field for ramp 1 occurs at a distance larger than the characteristic wavelength, while the variation for ramp 4 occurs at a distance shorter than the characteristic wavelength. The method succeeds for ramp 1 background model, degrades gradually, and fails for ramp 4. This is captured by the results for the application of the adjugate shown in figures 4.52, 4.53, 4.54 and 4.55. The significant difference happens for a transition zone of 500 m (ramp 3), close to the characteristic wavelength as discussed above.
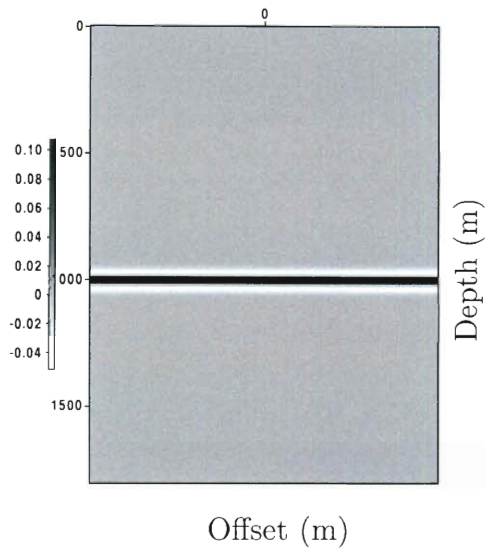
Figure 4.42: *vp*, velocity perturbation
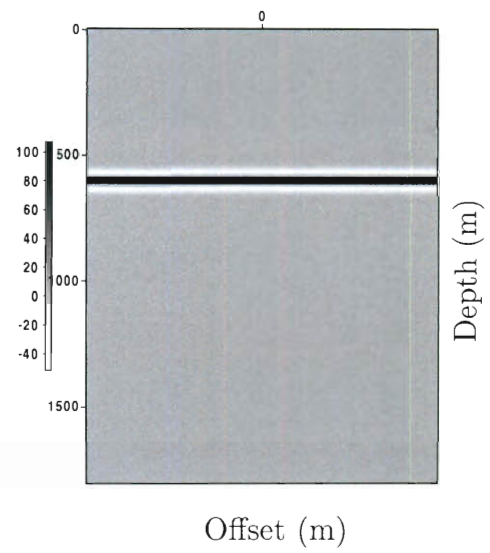
Figure 4.43: *dn*, density perturbation
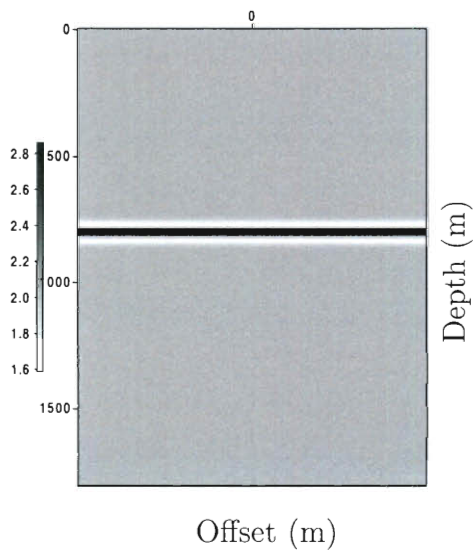
Figure 4.44: non smooth background velocity
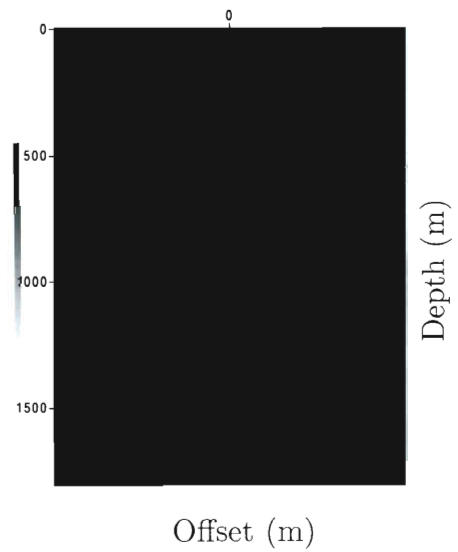
Figure 4.45: homogeneous background density
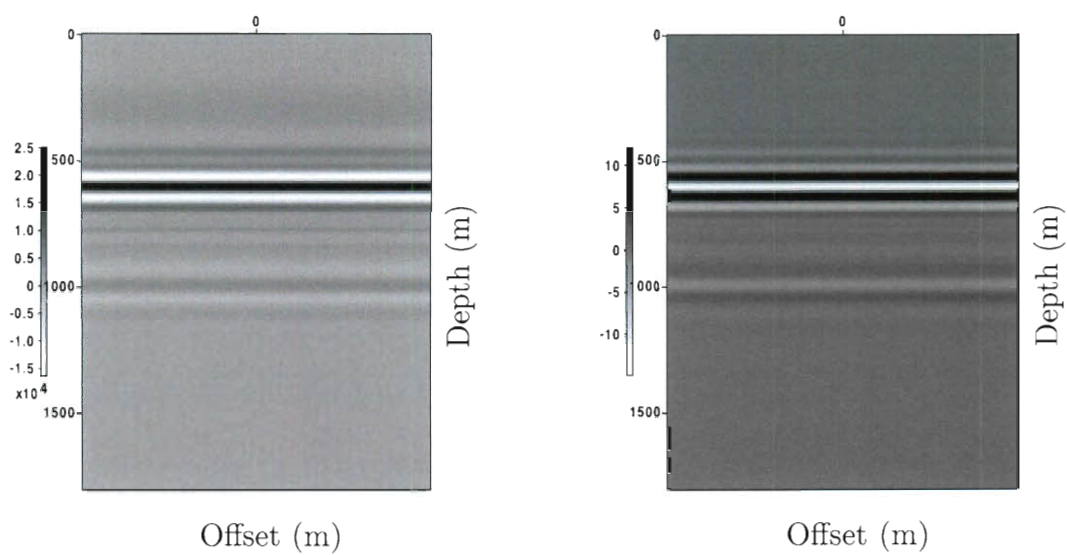
b_1, velocity component of the migrated image.

b_2, density component of the migrated image.

Figure 4.46: Migrated images. Traces of the background reflector start showing up faintly (since the perturbations are zero around 800 m depth). The lower event is considerably fainter.

$$(J^T N J b)_1 \approx det(N) x_1 \qquad\qquad (J^T N J b)_2 \approx det(N) x_2$$
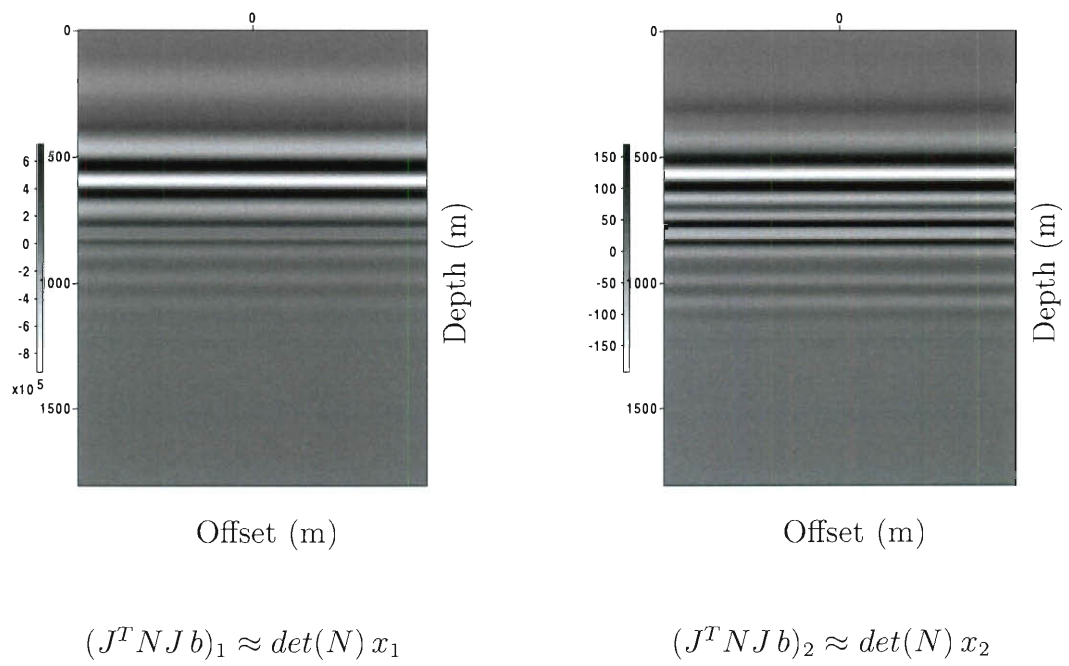
Figure 4.47: The normal operator is not a matrix of pseudodifferential operators and the application of the adjugate fails!
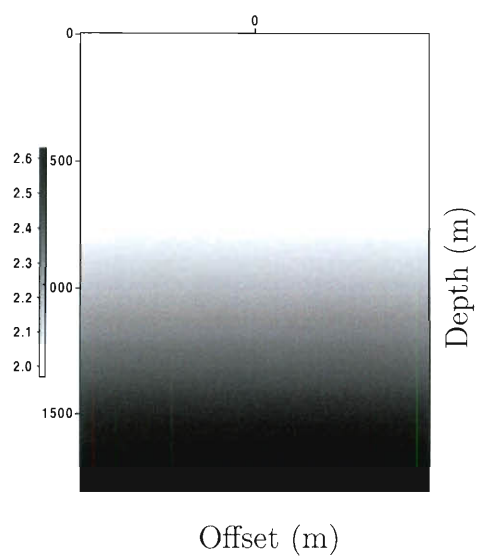
Figure 4.48: Background velocity: ramp 1
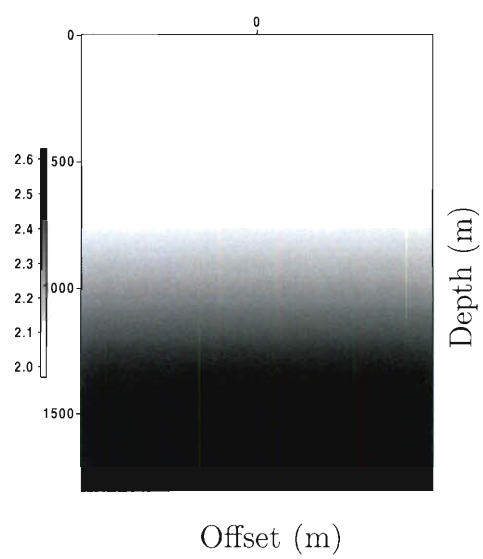
Figure 4.49: Background velocity: ramp 2
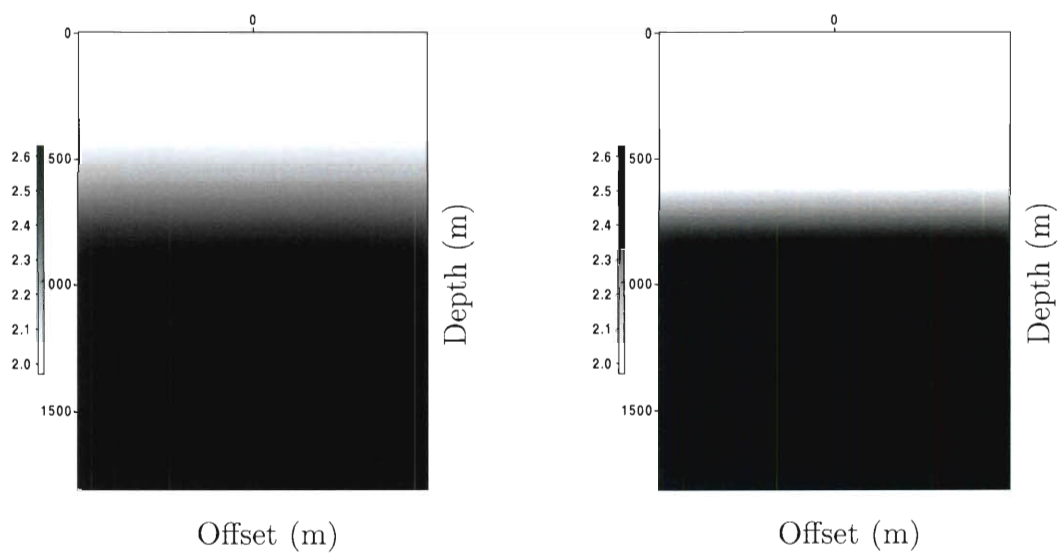
Figure 4.50: Background velocity: ramp 3

Figure 4.51: Background velocity: ramp 4

$$(J^T N J b)_1 \approx det(N)\, x_1 \qquad\qquad (J^T N J b)_2 \approx det(N)\, x_2$$
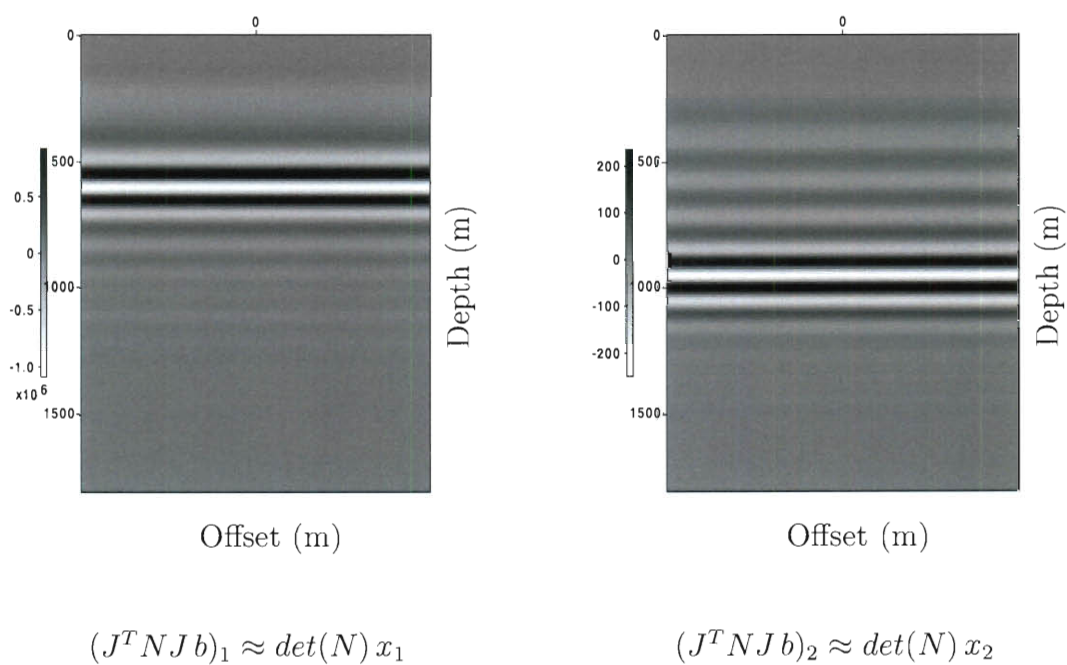
Figure 4.52: Application of adjugate. Ramp 1 result; the method succeeds.

$$(J^T N J \, b)_1 \approx det(N) \, x_1 \qquad\qquad (J^T N J \, b)_2 \approx det(N) \, x_2$$
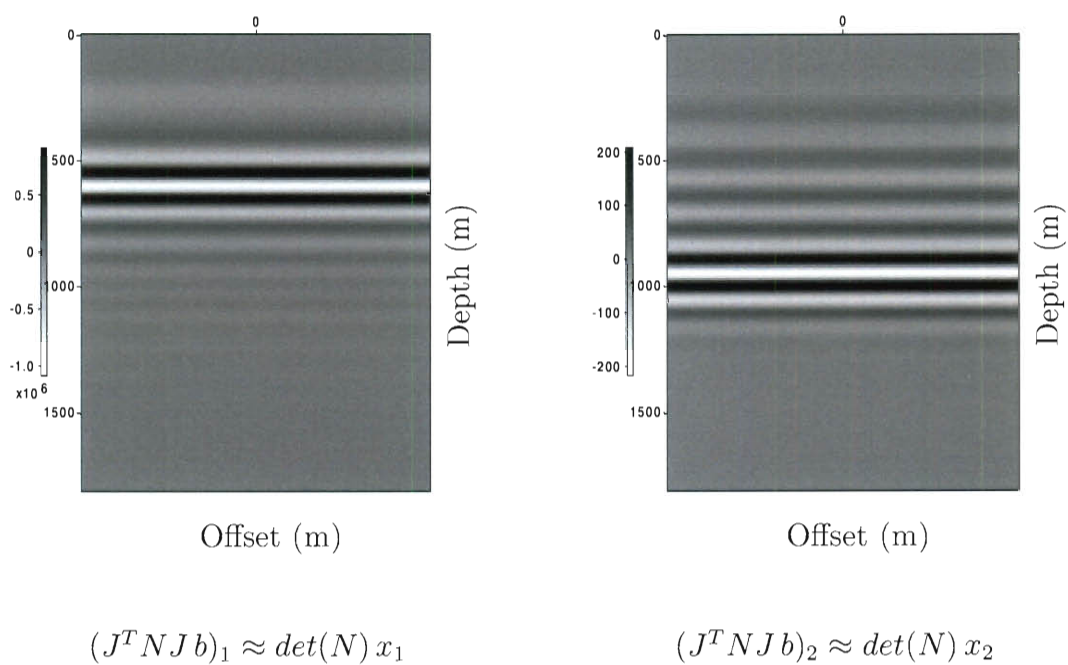
Figure 4.53: Application of adjugate. Ramp 2 result; the method succeeds.
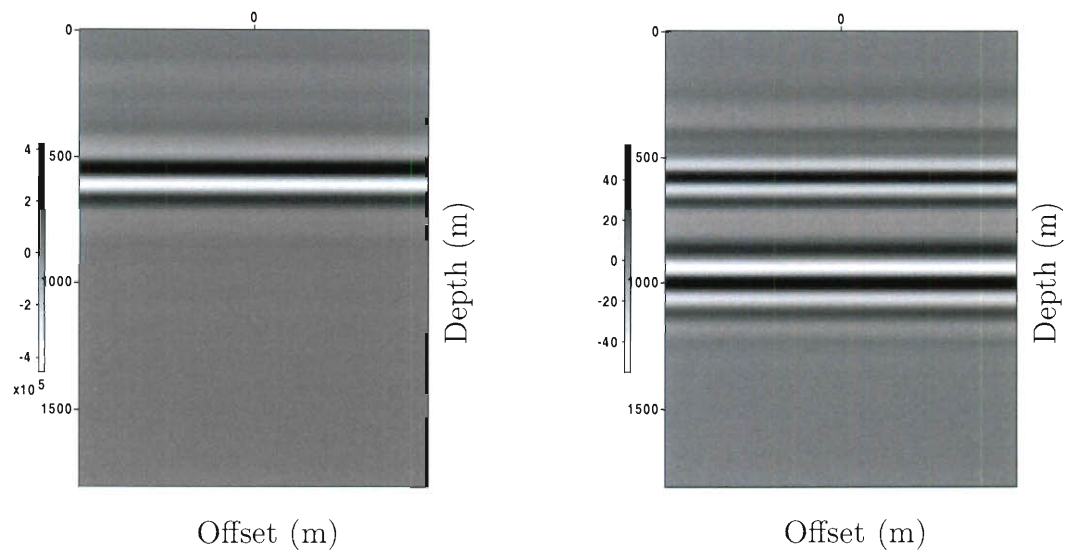
$$(J^T N J b)_1 \approx det(N)\, x_1 \qquad\qquad (J^T N J b)_2 \approx det(N)\, x_2$$

Figure 4.54: Application of adjugate. Ramp 3 result; the quality of the result degrades. As the transition zone comes close to the characteristic wavelength, the separation of the two events is lost. The deeper event in the figure on the right is still stronger than the shallow event.
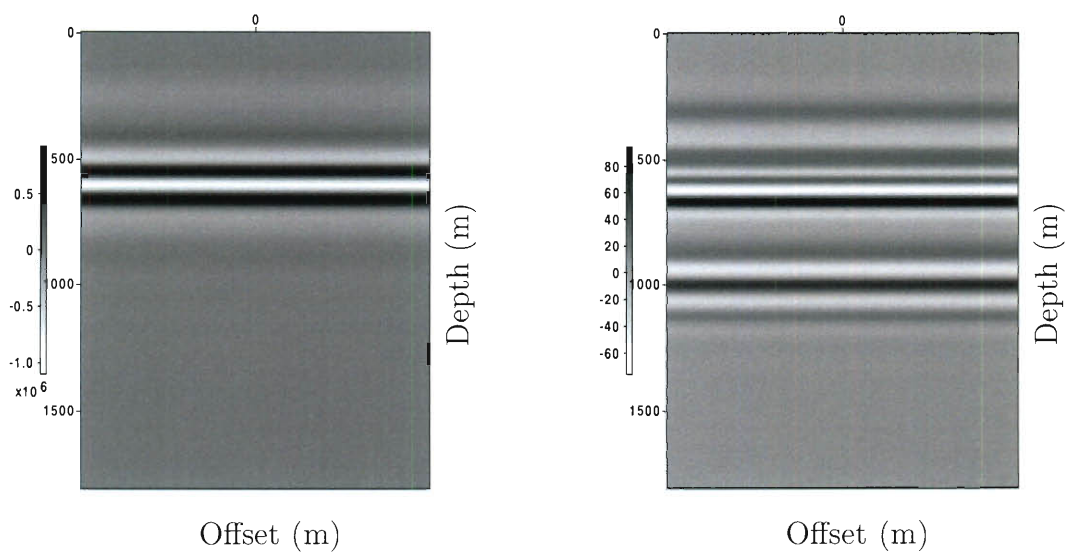
$$(J^T N J b)_1 \approx det(N)\, x_1 \qquad\qquad (J^T N J b)_2 \approx det(N)\, x_2$$

Figure 4.55: Application of adjugate. Ramp 4 result; the method fails. The transition zone is significantly shorter than the characteristic wavelength. The shallow event in the figure on the right is stronger than the deeper event.

4.5   Conditioning of the normal operator

This section studies the conditioning of the normal operator. The general form of the symbol of the normal operator for variable density acoustics in 2D derived in (Symes, 1998). These notes explain in detail how to represent the symbol of the normal operator as a function of the opening angle θ (the angle between the incident and reflected ray) defined in terms of position variables and Fourier variables. We use this form since it is adequate for the analysis of the conditioning of the normal operator. For a detailed derivation, please consult (Symes, 1998).

The symbol A of the normal operator N for variable density acoustics is of the form,

$$A = f(\theta) \begin{pmatrix} 1 & \sin^2(\frac{\theta}{2}) \\ \sin^2(\frac{\theta}{2}) & \sin^4(\frac{\theta}{2}) \end{pmatrix} |\xi| \tag{4.6}$$

The opening angle θ depends on source position x_s, receiver position x_r and spatial position x. ξ is the Fourier variable. I suppress the explicit form of the function $f(\theta)$ as the ill-conditioning of the matrix is due to the matrix part of equation (4.6).

In what follows, we will search for the weight function $f(\theta)$ that minimizes the condition number of the normal operator (the original weight factor predicted by the theory may be absorbed into this weight function). This approach yields an optimal weight $f(\theta)$ that renders the normal operator better conditioned.

To be explicit, we study the conditioning of matrices of the form:

$$N = \int_0^{\theta_{max}} d\theta \; f(\theta) \begin{pmatrix} 1 & \sin^2(\frac{\theta}{2}) \\ \sin^2(\frac{\theta}{2}) & \sin^4(\frac{\theta}{2}) \end{pmatrix}.$$

Denote the eigenvalues of N by $0 < \lambda_{min} \leq \lambda_{max}$ (since the matrix is positive definite). Minimize the condition number:

$$\kappa = \frac{\lambda_{max}}{\lambda_{min}}, \quad s.t. \; f \geq 0, \quad \int_0^{\theta_{max}} f(\theta) \, d\theta = 1.$$

If the condition number is parametrized in terms of $S = \lambda_{max} + \lambda_{min} = trace(N)$ and $P = \lambda_{max}\lambda_{min} = det(N)$, then

$$\kappa = \frac{S + \sqrt{S^2 - 4P}}{S - \sqrt{S^2 - 4P}}. \tag{4.7}$$

As a reference, we study the condition number κ_r as a function of θ_{max} for $f(\theta) = \frac{1}{\theta_{max}}$ (correct normalization). Figure 4.56 is a logarithmic plot of the condition number of N as a function of θ_{max}. It shows how the condition number increases as the maximum offset angle decreases. The main source of ill-conditioning in this problem is the maximum offset angle.

A candidate weighting to ameliorate the condition number of N is one that amounts to a low offset/large offset stack:

$$f(\theta) = (1 - \alpha)\delta(\theta) + \alpha\delta(\theta - \theta_{max}), \quad 0 \leq \alpha \leq 1. \tag{4.8}$$

This type of stack that puts emphasis on large offset and small offset separately has been used, since it is known that the different offset ranges give different information about the underlying physical parameters.
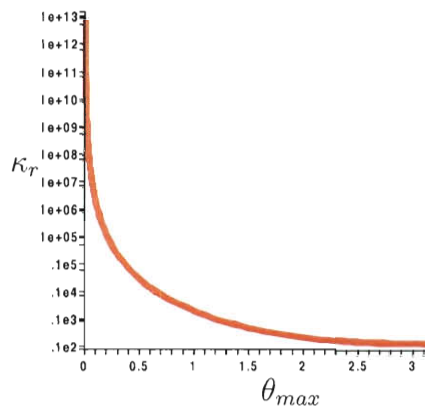
Figure 4.56: Condition number as a function of θ_{max}

Minimizing κ with the weight given in (4.8), gives

$$\alpha = \frac{1}{2 + \beta},$$

$$\kappa_{min} = \frac{\beta + 1 + \sqrt{1 + \beta}}{\beta + 1 - \sqrt{1 + \beta}},$$

where $\beta = \sin^4(\frac{\theta_{max}}{2})$.

It is interesting to note that the result depicted above predicts that for large offset ($\theta_{max} \to \pi$), small offsets are weighted double.

To compare the condition number obtained with this weight to the reference case, I plot the ratio $\frac{\kappa_{min}}{\kappa_r}$ as a function of θ_{max} in Figure 4.57. The reduction in condition number is about 33% of its reference value for low offsets, and 45% for large offsets.

The reduction in the condition number is not dramatic, which warrants a closer look at the asymptotics of the condition number. For small θ_{max}:

- Reference case:
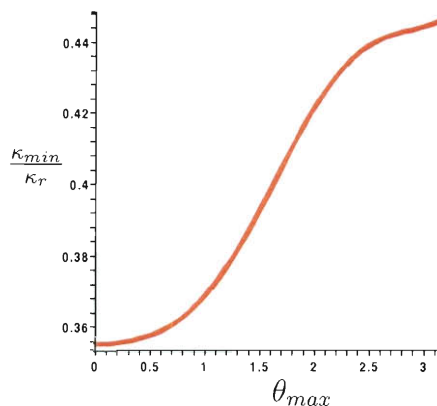
Figure 4.57: Ratio of optimal condition number to reference

$$- \; \lambda_{max} = 1 + \mathcal{O}(\theta_{max}^4)$$

$$- \; \lambda_{min} = \frac{\theta_{max}^4}{180} + \mathcal{O}(\theta_{max}^6)$$

$$- \; \kappa_r = \frac{180}{\theta_{max}^4} + \mathcal{O}(\theta_{max}^{-2})$$

- Optimal stacks:

$$- \; \lambda_{max} = 1 + \mathcal{O}(\theta_{max}^4)$$

$$- \; \lambda_{min} = \frac{\theta_{max}^4}{64} + \mathcal{O}(\theta_{max}^8)$$

$$- \; \kappa_{min} = \frac{64}{\theta_{max}^4} + \mathcal{O}(1)$$

The two cases exhibit the same asymptotics, which explains why the reduction in the condition number is not orders of magnitude. Interestingly, the result above also predicts that an adequately weighted small offset/large offset stack is better conditioned than using the entirety of the offset range!
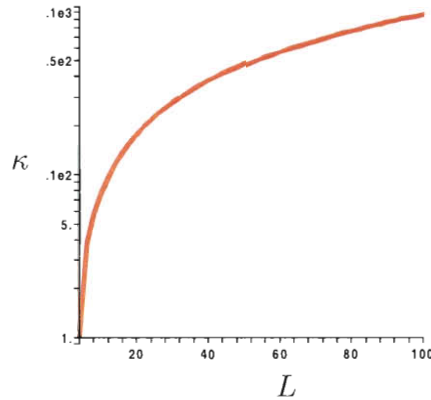
Figure 4.58: Condition number as a function of $L = \frac{S^2}{P}$

The first order conditions for this problem turn out to emphasize an interesting property of this optimization problem. The first order conditions are,

$$\delta\kappa = 0 \Rightarrow 2\frac{\delta S}{S} = \frac{\delta J}{J}.$$

The solution to this equation yields a different parametrization of the optimization problem in terms of L, with $L \geq 4$:

$$S^2 = LP \Rightarrow \frac{S^2}{P} = L,$$

which we can note also algebraically,

$$\kappa = \frac{S + \sqrt{S^2 - 4P}}{S - \sqrt{S^2 - 4P}} = \frac{1 + \sqrt{1 - \frac{4}{L}}}{1 - \sqrt{1 - \frac{4}{L}}}.$$

Minimizing the condition number amounts to minimizing $L = \frac{S^2}{P}$. To move along the curve that describes the condition number as a function of L, see Figure 4.58.

Another important piece of information that we can obtain from the reference case is the eigenvector with the largest eigenvalue, which we can calculate analytically. The result is not shown here, and the calculations are done using MAPLE. However, this calculation reveals the conditioning of the recovery of the interesting physical parameters: velocity, density and impedance. The angle that the vector representing each of these parameters makes with the eigenvector corresponding to the largest eigenvalue, describes the conditioning of the recovery of the parameter at hand. The results are shown in Figures 4.59, 4.60 and 4.61. In fact these figures explain the assertion made earlier about the recovery of the density being more ill conditioned compared to the recovery of the velocity. Figure 4.60 shows that, for small offset angle, the density is almost perpendicular to the optimal eigendirection, and therefore is aligned with the eigenvector with the smallest eigenvalue. The impedance is the best conditioned physical parameter for recovery, as it is aligned with the optimal eigendirection for small coverage angle. The velocity is in between, starting off with an angle of 45 degrees, and remaining in the mid-range. It is in this sense that the recovery of the density is the most ill-conditioned of the three physical parameters chosen usually.
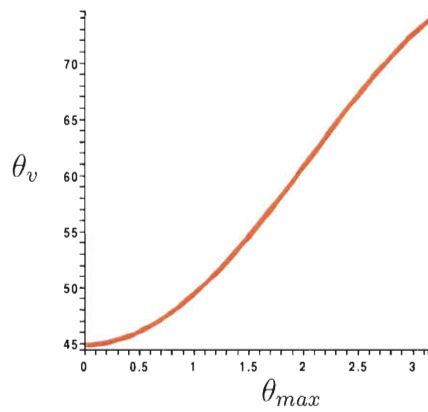
Figure 4.59: Angle in degrees that the velocity vector makes with the eigenvector corresponding to the largest eigenvalue as a function of maximum offset angle
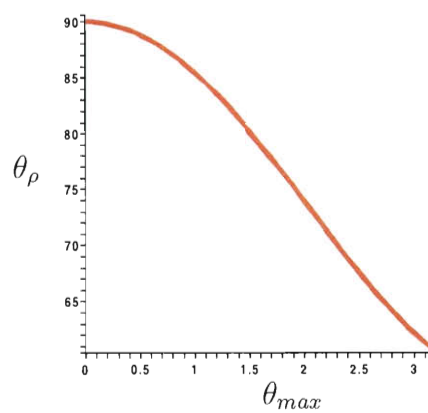
Figure 4.60: Angle in degrees that the density vector makes with the eigenvector corresponding to the largest eigenvalue as a function of maximum offset angle
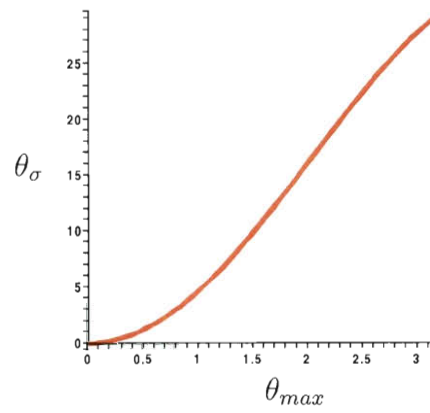
Figure 4.61: Angle in degrees that the impedance vector makes with the eigenvector corresponding to the largest eigenvalue as a function of maximum offset angle

4.6 Multi-component data

Another way to ameliorate the conditioning of the normal operator is measuring multi-component data. The data considered classically are the measurement of the pressure field at the surface. When additional data are available, they may be used to better pinpoint the material parameters sought in the inverse problem and accordingly improve the condition number of the normal operator.

We consider the case where the linearized forward map F has the relative perturbations in the impedance and density to pressure perturbations and averaged vertical derivatives of such perturbations. The second set of measurement is available and is known as over under cable data, a practical method to measure the vertical gradient of the pressure field at the surface.

$$F : \begin{pmatrix} \frac{\delta\sigma}{\sigma}(x) \\ \frac{\delta\rho}{\rho}(x) \end{pmatrix} \rightarrow \begin{pmatrix} \delta p(x_s, x_r, t_r) \\ c_0 \int_0^{t_r} \frac{\partial p}{\partial z_r}(x_s, x_r, t_r') \, dt_r' \end{pmatrix}. \tag{4.9}$$

As usual x_s, x_r, x are respectively the source, receiver, and spatial positions. t_r is time sampled at the receiver location, and z_r is the vertical component of x_r. A procedure similar to the one used in (Symes, 1998) yields an expression for the symbol of the normal operator for this specific case of multicomponent data. Surprisingly, it takes the form:

$$A' = f(\theta) \begin{pmatrix} a(\theta)^2 & -a(\theta)^2 \sin^2(\frac{\theta}{2}) \\ -a(\theta)^2 \sin^2(\frac{\theta}{2}) & a(\theta)^2 \sin^4(\frac{\theta}{2}) \end{pmatrix}. \tag{4.10}$$

We can write out $a(\theta) > 0$ explicitly. However, it is more interesting to note that the effect of the multi-component data improves the condition number by changing

the weight in front of the matrix studied in the previous section (the trace is scaled by a^2 and the determinant is scaled by a^4 , hence the eigenvalues are scaled by a^2 and the condition number remains unchanged). In the case of variable density acoustics and for this type of multicomponent data, we can improve the conditioning by optimally weighting the matrix as discussed in the previous section. In fact, adequately weighting the symbol matrix, as shown in the previous section, emulates multi-component data with single-component data!

4.7   Three-parameter case: $p = 3$

When formulating the method explicitly for more than two parameters, the algebra becomes a bit more involved; this is why we only show the case $p = 3$. However the procedure is generalizable to any p, albeit tediously.

The normal operator is denoted by

$$N = \begin{pmatrix} N_{11} & N_{12} & N_{13} \\ N_{12} & N_{22} & N_{23} \\ N_{13} & N_{23} & N_{33} \end{pmatrix}.$$

Its adjugate is,

$$Adj(N) = \begin{pmatrix} (N_{22}N_{33} - N_{23}^2) & -(N_{12}N_{33} - N_{13}N_{23}) & (N_{12}N_{23} - N_{13}N_{22}) \\ -(N_{12}N_{33} - N_{23}N_{13}) & (N_{11}N_{33} - N_{13}^2) & -(N_{11}N_{23} - N_{13}N_{12}) \\ (N_{12}N_{23} - N_{22}N_{13}) & -(N_{11}N_{23} - N_{13}N_{12}) & (N_{11}N_{22} - N_{12}^2) \end{pmatrix}.$$

We will again introduce special notation that will facilitate subsequent manipulations, so we can write $Adj(N)\,b$ as a series of swap operations on the entries of b followed by applications of N. The entries of $Adj(N)\,b$ are of the form:

$$N_{ij}N_{i'j'}b_k e_l = e_l e_i^T N e_j e_{i'}^T N e_{j'} e_k^T b := liji'j'k, \tag{4.11}$$

where again the adopted notation only shows the indices. Symmetry of N and approximate commutativity of its entries, allows us to deduce identities like $liji'j'k = ljii'j'k = lijj'i'k = li'j'ijk$. Using this notation and the previous identities we can write:

$$
\begin{aligned}
Adj(N)\,b &= 122331 - 123231 + 123132 - 121332 + 121233 - 122133 \\
&\quad + 231231 - 233121 + 233112 - 231132 + 231123 - 232113 \\
&\quad + 312231 - 313221 + 313122 - 311232 + 311223 - 312123 \\
&= (12)[2331 - 3231 + 3132 - 1332 + 1233 - 2133] \\
&\quad + (21)[3231 - 2331 + 1332 - 3132 + 2133 - 1233] \\
&\quad + (31)[2231 - 3221 + 3122 - 1232 + 1223 - 2123] \\
&= (-12 + 21)[33(21 - 12) + 23(13 - 31) + 13(32 - 23)] \\
&\quad + (31)[22(31 - 13) + 32(12 - 21) + 12(23 - 32)]
\end{aligned}
\tag{4.12}
$$

Equation (4.12) is interpreted as follows:

1. Form $N(e_2 e_1^T - e_1 e_2^T)b$, $N(e_1 e_3^T - e_3 e_1^T)b$ and $N(e_3 e_2^T - e_2 e_3^T)b$

2. Form $(e_2 e_1^T - e_1 e_2^T)N[e_3 e_3^T N(e_2 e_1^T - e_1 e_2^T) + e_2 e_3^T N(e_1 e_3^T - e_3 e_1^T) + e_1 e_3^T N(e_3 e_2^T - e_2 e_3^T)]b$

3. Form $-(e_3 e_1^T)N[e_3 e_2^T N(e_2 e_1^T - e_1 e_2^T) + e_2 e_2^T N(e_1 e_3^T - e_3 e_1^T) + e_1 e_2^T N(e_3 e_2^T - e_2 e_3^T)]b$

4. Sum the last two images to obtain $Adj(N)\, b \approx det(N)\, x$

The above procedure amounts to 5 applications of the normal operator N, followed by one extra application to approximate $det(N)$, bringing the total cost to 6 applications of N for $p = 3$.

In order to validate the method for this case, I construct an analytical example. I choose u to be a vector of oscillatory functions localized at different points, shown in figure 4.62:

$$u = \begin{pmatrix} \sin(30(x-z))e^{((-(x-1.5)^2 - z^2)/0.1)} \\ \sin(30(z+2x))e^{((-x^2 - (z-1)^2)/0.1)} \\ \sin(30(z+3x))e^{((-x^2 - (z+1.5)^2)/0.1)} \end{pmatrix}$$

The normal operator is also constructed analytically to be a matrix of pseudodifferential operators given its symbol matrix. The angle θ refers to the angle that the Fourier vector makes with the horizontal axis. This choice ensures that N is symmetric positive definite.

$$N = Op \begin{pmatrix} 1 & \cos^2(\theta) & \sin^2(\theta) \\ \cos^2(\theta) & \cos^4(\theta) + (x^2+1)^2 & \cos^2(\theta)\sin^2(\theta) + (x^2+z^2)(x^2+1) \\ \sin^2(\theta) & \cos^2(\theta)\sin^2(\theta) + (x^2+z^2)(x^2+1) & \sin^4(\theta) + (x^2+z^2)^2 + (z^2+1)^2 \end{pmatrix}.$$

The right hand side $b = N\,u$ mixes the three oscillatory packets (see figure 4.63).

ω	$\frac{\|N_{22}N_{23}u_1 - N_{23}N_{22}u_1\|}{\|N_{22}N_{23}u_1\|}$
5	16 %
10	9 %
20	5 %
40	2.6 %

Table 4.1: The relative size of the commutator as a function of the frequency. Results corroborate the ω^{-1} decrease predicted by the theory.

Applying the adjugate, by applying N to the specific permutations of b described above, yields the result shown in figure 4.64. Once again, applying the adjugate succeeds in separating the contributions from the different wave packets in b. It is obvious that the separation is not perfect. The adjugate of N satisfies equation (3.7) approximately, since scalar pseudodifferential operators commute approximately. The theory of pseudodifferential operators predicts that relative error is of the order of ω^{-1}. I have checked this theoretical fact numerically by calculating $\frac{\|N_{22}N_{23}u_1 - N_{23}N_{22}u_1\|}{\|N_{22}N_{23}u_1\|}$ for different ω and the error decreases like ω^{-1}. The results of this numerical experiment are shown in Table 4.1. Roughly speaking, in this case $\omega = 30$ (the frequency of the pulse). The result degrades if $\omega = 10$; see figure 4.65.
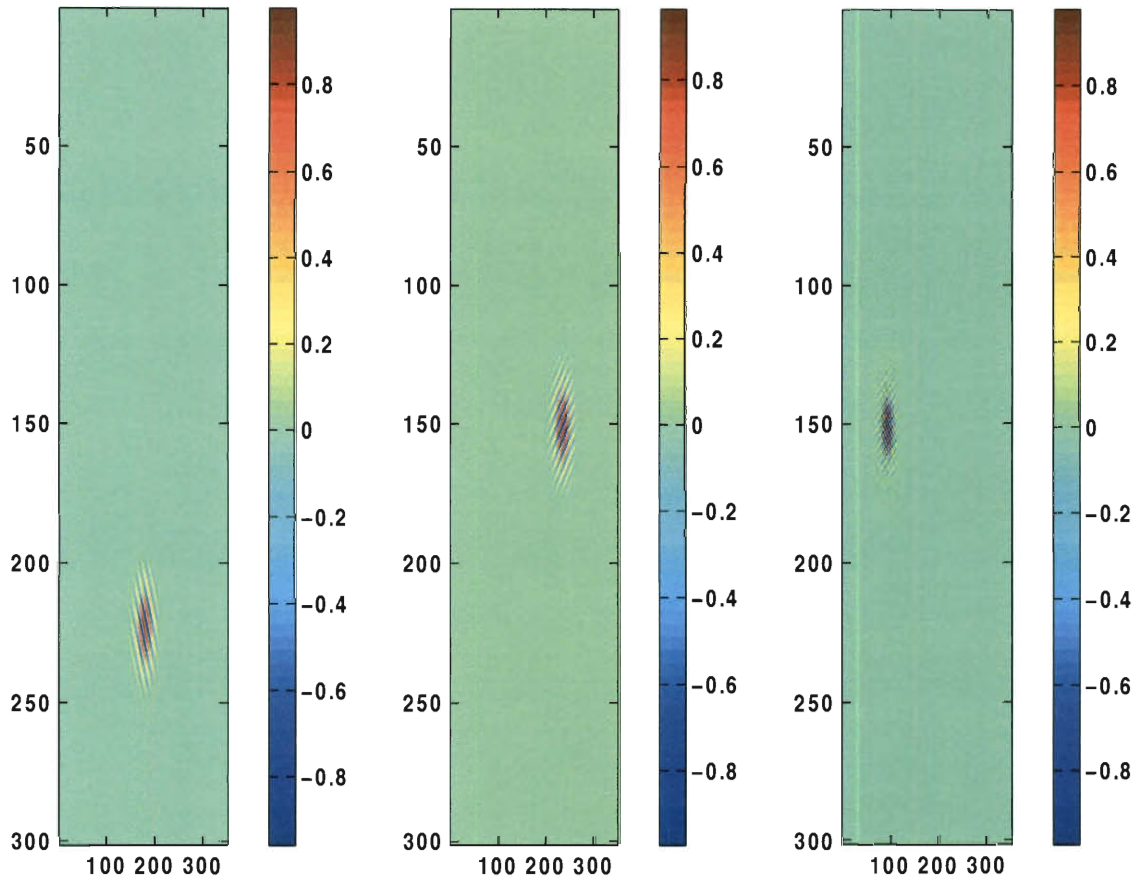
Figure 4.62: u, input vector consisting of three wave packets with different orientations in different places
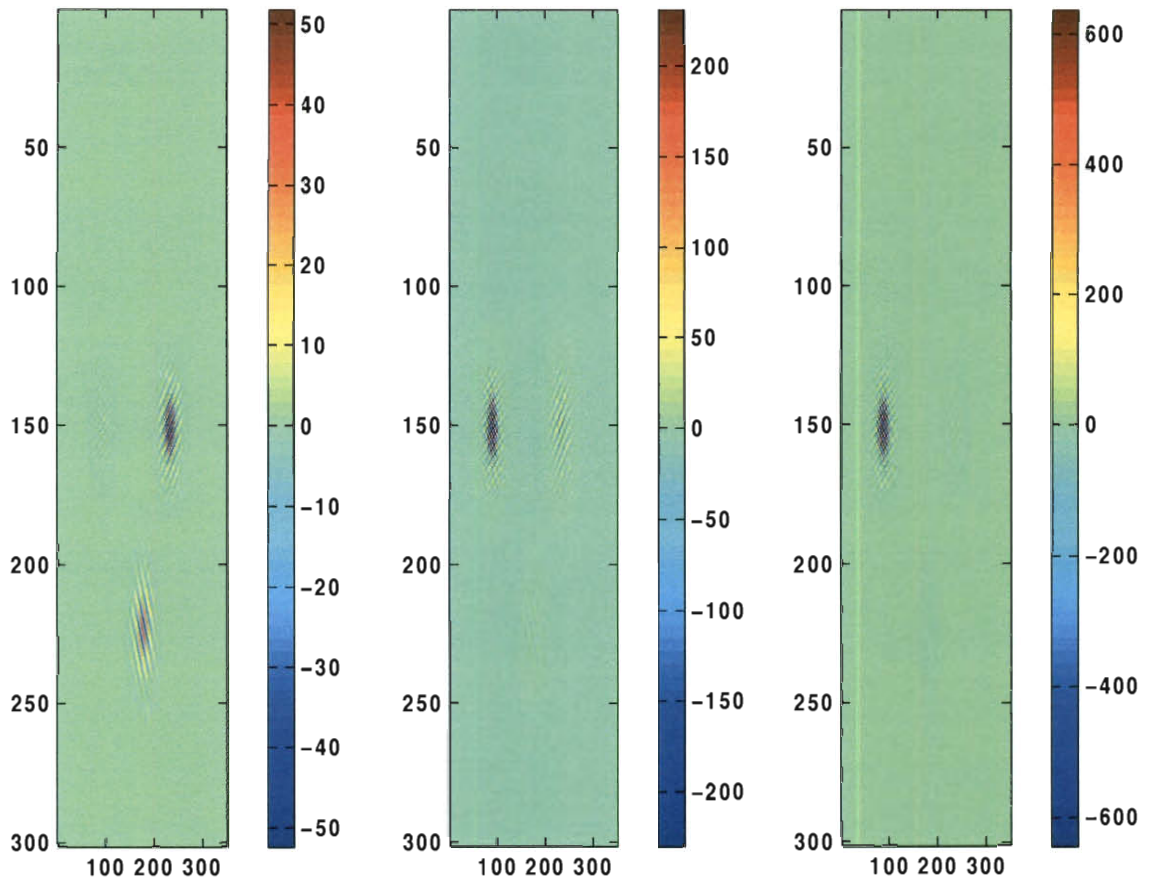
Figure 4.63: $b = Nu$, result after application of the matrix of pseudodifferential operators. The three wave packets are mixed and scaled in phase space
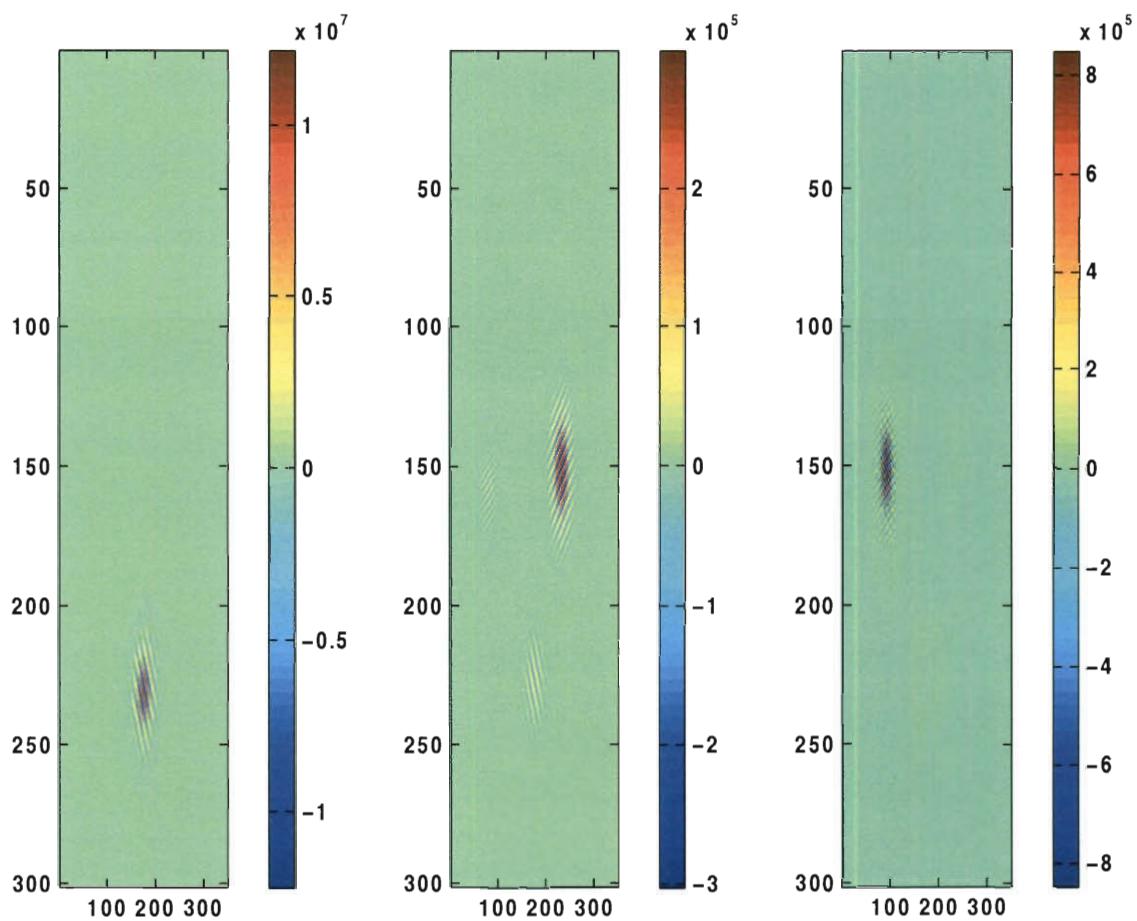
Figure 4.64: $Adj(N)\,b$ with $\omega = 30$. At high frequency the adjugate accurately separates the different wave packets from b. This result is a phase space scaling of the input u.
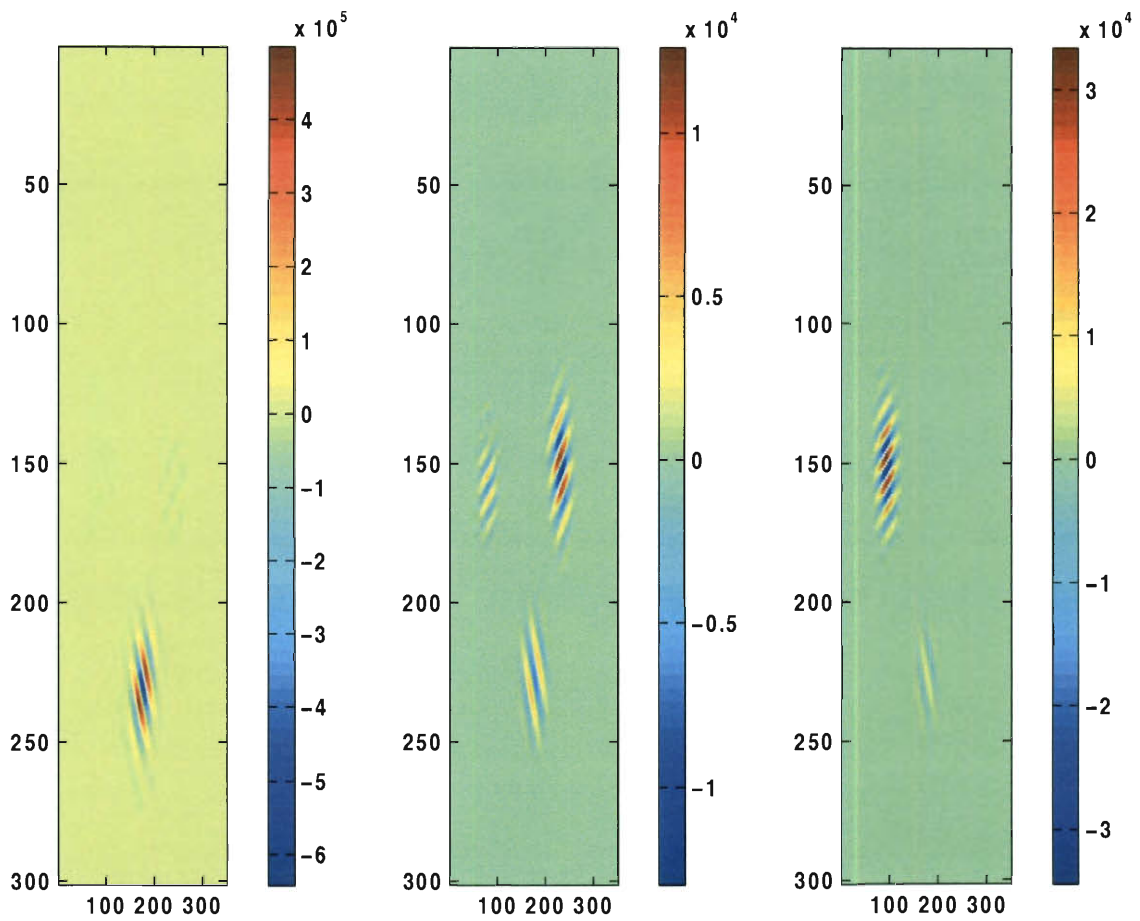
Figure 4.65: $Adj(N)\,b$ with $\omega = 10$. As the frequency decreases, the error committed is of order ω^{-1} and the same procedure fails to accurately separate the three wave packets.

Chapter 5

Possible Further Developments

The first item on the list is an application of the method to multi-parameter non-layered models. A 2D Marmousi variable density acoustics model is one possibility. The rest of this chapter discusses untested recent developments and a proposal to use the method proposed in this manuscript to precondition full waveform inversion methods.

5.1   Rotations

This section introduces an approach to limit the number of applications of the normal operator to p times for p parameters.

The requirement to apply the normal operator six times for $p = 3$ and the trend of growth with p becomes costly very fast. It is important to limit the number of the applications of the normal operator. The following idea limits the number of these

applications to p when the number of parameters is p, and this approach would be optimal.

Recall that we want to solve:

$$N\,x = b.$$

Where N is a $p \times p$ symmetric matrix of pseudodifferential operators.

Let $b_\perp$ be the vector that satisfies $b_\perp^T b = 0$ pointwise (b is a p vector of scalar functions). Then

<u>Claim 1:</u> $(Nb_\perp)^T x \approx 0$.

I will present a proof and an exact interpretation of this claim shortly. I begin by showing how to develop a method to approximate x which only requires p applications of the normal operator.

If $\{b_\perp^1, \ldots, b_\perp^{p-1}\}$ are $p-1$ linearly independent vectors perpendicular to b, i.e. $(b_\perp^i)^T b = 0$ for all $i \in [1, p-1]$. Then $\{N\,b_\perp^1, \ldots, N\,b_\perp^{p-1}\}$ are $p-1$ linearly independent vectors perpendicular to x, by the claim presented above. The fact that these vectors are linearly independent follows from the requirement that N is full rank, as a matrix of operators.

Therefore, $x \in Ker(span\{N\,b_\perp^1, \ldots, N\,b_\perp^{p-1}\})$.

Explicitly, $b_\perp = R_b\,b$. $b_\perp$ is a rotation of 90 degrees of the right hand side. $R_{b_\perp}\,N\,b_\perp = R_{b_\perp}\,N\,R_b\,b = x'$.

<u>Claim 2:</u> $x' = R_{b_\perp}\,N\,R_b\,b = \lambda x$, where λ is a pseudodifferential operator of order

$2 \times ord(N)$.

To see this, it is enough to note that $b = Nx$ and the rotation are operations in space and can be arranged to have zero order; they do not act in the frequency domain at all.

This leads to the following program to approximate x:

1. Form: $\{b_\perp^1, \ldots, b_\perp^{p-1}\} \in Ker(b^T)$

2. Apply N to form: $\{Nb_\perp^1, \ldots, Nb_\perp^{p-1}\}$

3. Compute $x' = \lambda x \in Ker(span\{Nb_\perp^1, \ldots, Nb_\perp^{p-1}\})$

4. Apply N again, to get $Nx' = N\lambda x \approx \lambda Nx = \lambda b$

5. Compute a scaling factor c,

$$c = \underset{c \in \Psi DO}{argmin} \|b - c\lambda b\|^2$$

6. Approximate x by $x_{inv} = c\lambda x$

The cost of these steps is p applications of the normal operator. Note that this method is equivalent to Cramer's rule for $p = 2$; the change in the approach only affects $p \geq 3$.

To complete this section we still have to justify claim 1. The first justification uses the asymptotic expansion lemma (2.11). Assume that $x_i = \sum_j \chi_{ji} e^{i\omega\Psi_{ji}}$ (a linear

combination of wavepackets). We have dropped the dependence of χ and Ψ on the spatial variables for brevity.

Then, using (2.11) to explicitly approximate the action of the normal operator on the wave packets, with $ord(N) = m$, we denote the symbol of N_{ij} as q_{ij}. The ith entry of b is given by

$$
\begin{aligned}
b_i &= \sum_k N_{ik} x_k \\
&= \sum_{kj} q_{ik} \chi_{jk} e^{i\omega \Psi_{jk}} + \mathcal{O}(\omega^{m-1})
\end{aligned}
\tag{5.1}
$$

The ith entry of the $b_\perp$ is given by $(b_\perp)_i = \sum_{kj} q'_{ik} \chi_{jk} e^{i\omega \Psi_{jk}}$, satisfying:

$$
\begin{aligned}
0 &= b_\perp^T b \\
&= \sum_i \left(\sum_{kj} q'_{ik} \chi_{jk} e^{i\omega \Psi_{jk}} \right) \left(\sum_{k'j'} q_{ik'} \chi_{j'k'} e^{i\omega \Psi_{j'k'}} \right) \\
&= \sum_{ikjk'j'} q'_{ik} q_{ik'} \chi_{jk} \chi_{j'k'} e^{i\omega \Psi_{jk}} e^{i\omega \Psi_{j'k'}} .
\end{aligned}
\tag{5.2}
$$

We are now ready to address the claim that $(N b_\perp)^T x \approx 0$:

$$(N\,b_\perp)^T x = \sum_i (Nb_\perp)_i x_i$$

$$= \sum_{ij} N_{ij}(b_\perp)_j x_j$$

$$= \sum_i \left(\sum_j N_{ij} \sum_{k'j'} q'_{jk'} \chi_{j'k'} e^{i\omega\Psi_{j'k'}} \right) \left(\sum_l \chi_{li} e^{i\omega\Psi_{li}} \right) \quad (5.3)$$

$$\approx \sum_{ijk'j'l} q_{ij} q'_{jk'} \chi_{j'k'} \chi_{li} e^{i\omega\Psi_{j'k'}} e^{i\omega\Psi_{li}}$$

$$= \sum_{kikj'j} q_{ki} q'_{ik'} \chi_{j'k'} \chi_{jk} e^{i\omega\Psi_{j'k'}} e^{i\omega\Psi_{jk}}$$

Using the fact that $q_{ki} = q_{ik}$ since the normal operator is symmetric, we can identify this last expression with the orthogonality condition between b and $b_\perp$ and conclude:

$$(N\,b_\perp)^T x \approx 0.$$

The approximate equality means to leading order in frequency. We can use the fact that wave packets constitute a tight frame for seismic images to leverage these calculations into a proof of Claim 1, when dip $\nabla\Psi$ is well defined everywhere.

In general, denote by $\langle .,. \rangle$ the L^2 inner product and, let q be a smooth function

(a test function). Then since $q\,(b_\perp)^T b = 0$,

$$0 = \sum_i \langle q(b_\perp)_i, b_i \rangle$$

$$= \sum_i \langle q(b_\perp)_i, (Nx)_i \rangle$$

$$= \sum_{ij} \langle q(b_\perp)_i, N_{ij}x_j \rangle$$

$$= \sum_{ij} \langle N_{ji}q(b_\perp)_i, x_j \rangle \qquad (5.4)$$

$$= \sum_{ij} \langle qN_{jk}(b_\perp)_i, x_j \rangle + \sum_{ij} \langle [q, N_{jk}](b_\perp)_i, x_j \rangle$$

$$= \sum_j \langle q(Nb_\perp)_j, x_j \rangle + \sum_{ij} \langle [q, N_{jk}](b_\perp)_i, x_j \rangle$$

$$= \langle q, (Nb_\perp)^T x \rangle + \sum_{ij} \langle [q, N_{jk}](b_\perp)_i, x_j \rangle.$$

The fact that $[q, N_j k](b_\perp)_i$ is smoother than $qN_{jk}(b_\perp)_i$ allows us to assume that the term involving the commutator is a smoother error, and to leading order:

$$\langle q, (Nb_\perp)^T x \rangle \approx 0,$$

for any smooth function q. So we interpret Claim 1 in a weak sense to say that $(Nb_\perp)^T x$ is the zero distribution.

5.2 Full Waveform Inversion

In this section we take a step back to the original nonlinear inverse problem. I point out how the pseudodifferential scaling method developed for the linearized inverse

problem accelerates the convergence of the optimization for the model m. I present the method in the case of one parameter inversion; the generalization to multi-parameters follows suit.

I developed this work with the help of Dr Fuchun Gao in an attempt to precondition his FWI code.

One way to recover the model m without linearizing around an a-priori known background model is trying to optimize for m by trying to fit the data through the nonlinear forward map. The objective function is given by

$$J = \frac{1}{2} \left\| S[m] - d \right\|^2 . \tag{5.5}$$

The gradient of the objective function is,

$$g = F^*(S[m] - d), \tag{5.6}$$

and the Hessian is,

$$H = F^*F + \frac{\partial F^*}{\partial m}(S[m] - d). \tag{5.7}$$

Newton's method to minimize the objective function in an effort to recover m will have the following updates:

$$m_{k+1} = m_k - H^{-1}g. \tag{5.8}$$

Gauss-Newton's method neglects the second term in equation (5.7) $H \approx F^*F$, under assumptions of small residual or mild nonlinearity. Even after this approximation, it is too expensive to invert the Hessian. The approximation to the inverse of the normal operator obtained by the pseudodifferential scaling method serves as

a substitute for H^{-1}. The inverse of the Hessian is usually replaced by a constant and the Gauss-Newton method reduces to steepest descent with line search; the scaling factor preconditions the problem and introduces curvature information about the objective function, and is expected to accelerate the convergence of the optimization.

Full waveform inversion does not split the model into a smooth and rough component. The approximation of good scaling factor however relies on the fact that the velocity field is split into these two parts. We therefore proceed at each step with splitting the velocity,

$$m_k = m_{k0} + \delta m_k.$$

Apply the normal operator on the rough part to obtain a scaling factor c,

$$c = \underset{c \in \Psi DO}{\operatorname{argmin}} \| \delta m_k - c\, F^*[m_{k0}] F[m_{k0}] \delta m_k \|^2. \tag{5.9}$$

Use the scaling factor thus derived to approximate the inverse of the Hessian, $H^{-1} \approx c$:

$$m_{k+1} = m_k - \alpha_k c\, g, \tag{5.10}$$

where α_k is a line search parameter to ensure decrease in the objective function.

This development is independent of the space dimension.

The justification for using the scaling factor as an approximation to the inverse Hessian is somewhat a posteriori. The approximation of the scaling factor requires the application of the normal operator and is thus comparable to the expense of one iteration of the FWI. The use of the scaling factor is justified if the overall

number of applications of the modeling and migration operators is decreased when using the scaling factor, as opposed to simple steepest descent. Also, the scaling factor is dependent on the smooth part of the model, and we may be able to skip its approximation at each step if the smooth part remains constant. This last point constitutes part of the experiments that we need to conduct as part of the future work.

Also, the scaling factor is an approximation of the inverse only in the part of the data that is explained by the linearized theory (no multiples, . . .). It remains to be seen how well of an approximation it is on the data that cannot be explained by the linearized part exclusively.

Herrmann et al. (2008a) apply part of this program on the linearized least squares problem with satisfactory results on the acceleration of the inversion process. Jang et al. (2009) also use the scaling factor they derive from the method of virtual sources to precondition full waveform inversion and accelerate the conversion of the nonlinear optimization problem.

Chapter 6

Conclusion

I have proposed a method to approximate the inverse of the normal operator arising in the linearized multi-parameter inverse problem for reflection seismology in two and three spatial dimensions. Under some conditions, the normal operator is a matrix of pseudodifferential operators. This manuscript shows how to generalize Cramer's rule to solve equations involving matrices of pseudodifferential operators.

The method applies the normal operator to permutations of the right hand side to produce a phase space scaling of the solution. It then proceeds to correct for the phase space scaling.

The use of Cramer's rule reduces the multi-parameter problem to the one parameter problem. Its advantage is that it uses the application of the normal operator as a black box. Therefore, it applies to various multi-parameter linearized inverse problems (variable density acoustics, linear elasticity ...). It only relies on the fact that

the normal operator is a matrix of pseudodifferential operators with a given order.

The tests shown in the results section validate the success of the method, specifically in separating the influences from different material parameters from a right hand side that features a mix of all these contributions. These examples range from simple two layer models with constant background to the more complex layered models with smooth background. In principle the same method should work for laterally homogeneous models at a higher computational cost to apply the normal operator. I have verified this in my masters thesis for the one-parameter case.

A final example uses a non-smooth background model, for which the normal operator is no longer a matrix of pseudodifferential operators. It shows how the method fails as one of the conditions for its success is violated.

This thesis also proposes a method to limit the number of required applications of the normal operator in the possible developments section. This section also contains a discussion about using the approximate inverse of the normal operator to precondition the nonlinear inverse problem.

Bibliography

Aki, K. and Richards, P. (1980). *Quantitative Seismology: Theory and Methods*. Freeman, San Francisco.

Bao, G. and Symes, W. (1996). Computation of pseudo-differential operators. *SIAM Journal on Scientific Computing*, 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.

Beylkin, G. and Burridge, R. (1990). Linearized inverse scattering problem of acoustics and elasticity. *Wave Motion*, 12:15–22.

Bourgeois, A., Jiang, B., and Lailly, P. (1989). Linearized inversion: A significant step beyond pre-stack migration. *Geophysics J. Int.*, 99:435–445.

Charara, M., Barnes, C., and Tarantola, A. (1996). Constrained waveform inversion of seismic well data. In Jacobsen, B., Mosegaard, K., and Sibani, P., editors, *Inverse*

Methods, volume 63 of *Lecture Notes in Earth Sciences*, pages 98–112. Springer Berlin / Heidelberg. 10.1007/BFb0011767.

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

Demanet, L., Létourneau, P., Boumal, N., Calandra, H., Chiu, J., and Snelson, S. (2011). Matrix probing: a randomized preconditioner for the wave-equation Hessian. Preprint (2011) available at `http://math.mit.edu/icg/papers/matrix-probing-seismic.pdf`.

Demanet, L. and Ying, L. (2011). Discrete symbol calculus. *SIAM Review*, 53(1):71–104.

Foss, S.-K., de Hoop, M. V., and Ursin, B. (June 2005). Linearized 2.5-dimensional parameter imaging inversion in anisotropic elastic media. *Geophysical Journal International*, 161:722–738(17).

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.

Jang, U., Min, D., and Shin, C. (January 2009). Comparison of scaling methods for waveform inversion. *Geophysical Prospecting*, 57:49–59(11).

Levander, A. (1988). Fourth order finite difference P-SV seismograms. *Geophysics*, 53:1425–1434.

Lörtzer, G. J. M. and Berkhout, A. J. (1989). Linear AVO inversion of multicomponent seismic data. *SEG Technical Program Expanded Abstracts*, 8(1):967–972.

Minkoff, S. E. and Symes, W. W. (1997). Full waveform inversion of marine reflection data in the plane-wave domain. *Geophysics*, 62(2):540–553.

Nammour, R. (2009). Approximate inverse scattering using pseudodifferential scaling. Technical Report 09-09, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

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.

Rutherford, S. R. and Williams, R. H. (1989). Amplitude-versus-offset variations in gas sands. *Geophysics*, 54:680–688.

Santosa, F. and Symes, W. W. (1988). High-frequency perturbational analysis of the surface point-source response of a layered fluid. *Journal of Computational Physics*, 74:318–381.

Shuey, R. T. (1985). A simplification of Zoeppritz equations. *Geophysics*, 50:609–614.

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

Strang, G. (1988). *Linear Algebra and Its Applications.* Brooks Cole.

Sun, D. and Symes, W. W. (2010a). IWAVE implementation of adjoint state method. Technical Report 10-06, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

Sun, D. and Symes, W. W. (2010b). IWAVE implementation of Born simulation. Technical Report 10-05, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

Symes, W. (1998). Mathematics of reflection seismology. Available at `http://www.trip.caam.rice.edu/downloads/preamble.pdf`.

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

Tarantola, A. (1987). *Inverse Problem Theory.* Elsevier.

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

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

Virieux, J., Jin, S., Madariaga, R., and Lambaré, G. (1992). Two dimensional asymptotic iterative elastic inversion. *Geophysical Journal International*, 108:575–588.