University of Alberta

Library Release Form

NAME OF AUTHOR: Carrie F. Youzwishen

TITLE OF THESIS: Non-linear sparse and blocky con-
straints for seismic inverse problems

DEGREE: Master of Science

YEAR THIS DEGREE GRANTED: 2001

Carrie F. Youzwishen
Department of Physics
University of Alberta
Edmonton, AB
Canada T6G 2J1

September 25, 2001

We can do no great things; only small things with great love.

Mother Teresa

University of Alberta

Non-linear sparse and blocky constraints for seismic inverse problems

by

Carrie F. Youzwishen

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Geophysics

Department of Physics

Edmonton, Alberta

Fall 2001

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **Non-linear sparse and blocky constraints for seismic inverse problems** submitted by **Carrie F. Youzwishen** in partial fulfillment of the requirements for the degree of Master of Science.

_____

Dr. Mauricio D. Sacchi

_____

Dr. Frances R. Fenrich

_____

Dr. Lawrence H. Le

_____

Dr. Bruce R. Sutherland

DATE:

To

Oliver

Abstract

The seismic exploration problem can be expressed as a linear integral relating the Earth model and the recorded data. An estimated Earth model may be recovered from observed data by inversion. A 2D acoustic, constant density inversion algorithm is presented.

Standard smooth constraints deal well with noise, but blur material boundaries. In contrast, a blocky solution will encourage sharp discontinuities. This is enforced by placing sparse constraints on the model gradient. A non-linear sparse constraint found to be especially suited for blocky medical imaging by Charbonnier et al. (1997) is tested for exploration seismology.

The non-linear sparse/blocky constraint is tested for 1D seismic impedance inversion. It is then applied to the 2D migration/inversion problem. In both instances, it consistently recovers accurate, piecewise continuous profiles. Higher resolution images are obtained than those recovered by standard solutions.

Preface

Acknowledgements

I would like to thank my supervisor, Dr. Mauricio D. Sacchi, for all of his help and effort, and for his guidance into the world of research. He has changed the way I think about geophysics. I now truly understand when he says that everything in life is an inverse problem.

I would like to acknowledge my colleagues Veronica Martinez and Henning Kuehl for their encouragement, and many conversations. I am especially grateful to my good friend, Jennifer Wickman, who practically lived each day of graduate school with me. The support and humour of my friends are very much appreciated.

I would also like to thank Kristen Beaty who has been patient and helpful with the thesis document preparation, and formatting.

My family has played an instrumental role in getting me to where I am today. I would like to thank them for their constant support, and for their strong example.

Finally, I would like to recognize my husband, Oliver, for his unending encouragement and dedication. He has been involved in all aspects of this thesis, and in his characteristic generosity, has taken as much interest in my work as in his own.

Table of Contents

TABLE OF CONTENTS

List of Tables

List of Figures

LIST OF FIGURES

Chapter 1

Introduction

1.1   The Earth's response

Within the discipline of geophysics, probing the earth is a large area of study. This is because most of the science deals with describing, interpreting, and understanding earth structures that are hidden beneath the surface. Exploration seismology deals with discovering the earth's geology in order to exploit natural resources, or for environmental purposes. Material properties are inferred from their response to an induced excitation.

The reflection and refraction of seismic waves occurs when there is a difference in impedances between two media (Beck, 1991). In the acoustic case (where only a pressure or longitudinal wave is considered), the acoustic impedance is defined as the product of the density and acoustic velocity. The density variation is usually very small with respect to the velocity variation. Because of this, a change in acoustic impedance is generally associated with a change in velocity. The magnitude of this discontinuity is described by its reflection coefficient, or reflectivity. Thus, the earth is characterized by the impedances of its layers, or the reflectivity of its layer boundaries. The recovery of these material properties

is the aim of seismic imaging.

1.2 Seismic tomography and the generalized Radon transform

The word tomography comes from the Greek root *tomos*, meaning slice. Tomography is defined as the reconstruction of a function from line integrals through it (Claerbout, 1992). Diffraction tomography refers to the reconstruction of a material perturbation that results in wavefield scattering using the reflected, or scattered part of the wavefield (Song et al., 1995). In general, tomography is characterized by the division of the unknown medium into a grid of cells whose material properties are to be determined.

Diffraction tomography is a common technique of medical imaging, where x-rays or ultrasonic rays are passed through the body to surrounding receivers. In geophysics, seismic tomography is called thus because of its mathematical similarities to medical imaging (Lay and Wallace, 1995). However, the receiver coverage is not as easily manipulated, and therefore, oftentimes not complete. Seismic tomography includes cross-well experiments, and can even be extended to experiments where a dense array of receivers are positioned on the earth's surface.

In geophysical and medical applications, a wavefield is passed though the material and recorded. The assumption is made that the wavefield is composed of two parts: one created by the background or known material properties, and one created by the perturbation from this background state. The Born approximation presumes that the scattered wavefield is very small with respect to the background one, and produces in a linear scattering integral (Beylkin, 1985).

The linearized scattering equation of the diffraction tomography problem is a

filtered version of the generalized Radon transform (GRT) (Beylkin, 1985; Miller et al., 1987). By using an asymptotic inverse expression of the GRT, material properties may be recovered from the recorded data, and the tomographic problem solved. This method has been used in computed tomography (Melamed et al., 1999), ultrasonic imaging (Mast, 1999), and imaging of the Earth's crust using teleseismic waves (Bostock and Rondenay, 1999). Applications to environmental problems using ground penetrating radar (GPR) have also been explored successfully (Cui and Weng, 2000; Wang and Oristaglio, 2000).

Within seismic exploration, solving the inverse scattering problem using the GRT was first introduced by Beylkin (1985) for a 2D acoustic, constant density medium. The technique was extended to exploration geophysics for the purpose of oil and gas exploration (Miller et al., 1987; Bleistein, 1987). Further refinements include the extension of the inversion to the elastic case (Beylkin and Burridge, 1990), and to anisotropic media (de Hoop and Bleistein, 1997). All of these methods are used in the field of seismic imaging.

1.3   Seismic imaging

Seismic imaging, as the name implies, is the imaging of the earth using seismic wavefield data. There are two types of seismic imaging (Berkhout, 1984; Jin et al., 1992): imaging accomplished by direct or discrete inversion. Direct inversion recovers the earth image using a direct relation between the data and model. It is customized by an analytic solution, and works only with particular source-receiver geometries. This type of imaging is also referred to as migration because it reverses the seismic experiment. The wavefield is propagated (migrated) back in space and time to the original scattering points that created it. An example of direct inversion is solving the linearized scattering equation using an asymptotic inverse to the GRT (Beylkin, 1985).

The second type of imaging uses a discrete, numerical method to approximate the inverse. It minimizes the difference between observed data and synthetic data computed from the Earth image. This technique of solving the problem numerically is called discrete inverse theory, and is often referred to simply as inversion. The synthetic data can be computed using ray theoretic approaches (Beydoun and Mendez, 1989), or finite difference schemes (Crase et al., 1990). Discrete inversion can handle arbitrary source-receiver geometries because of its numerical rather than analytical approach, however it is computationally expensive. It is also necessary to start with a reference Earth model or in some way impose a constraint to recover a unique, stable solution.

Jin et al. (1992) were the first to combine direct and discrete inversion to take advantage of their individual strengths. By using a numerical version of the elastic Born modeling/direct inversion technique (Beylkin and Burridge, 1990), they were able to include the linear, less computationally expensive algorithm as a means to compute synthetic data. This was implemented in the framework of discrete inversion to retrieve the final image.

This combined technique is commonly known as a migration/inversion scheme. It has been used to recover 2D acoustic velocity and impedance perturbations (Thierry et al., 1999), and adapted to retrieve reflectivity information (Duquet et al., 2000). However, the inverse problem remains non-unique, and a constraint must be enforced to select a stable, unique solution.

1.4 Constraints in discrete inversion

To date, the majority of migration/inversion algorithms use a background or prior model to constrain the solution (Tarantola, 1984; Jin et al., 1992; Thierry et al., 1999). The constraint is applied by minimizing the error between the updated and prior models. This error is defined in terms of the distances of the

solutions from one another, and is commonly chosen to be an L_2 norm. However, other definitions of error length are possible.

Constraints can also be thought of in terms of probability distributions, or from a Bayesian perspective (Menke, 1984; Tarantola, 1987). These probability distributions influence the type of solution obtained. The use of a Gaussian prior probability distribution enforces a smooth solution, while a longer-tailed exponential prior distribution yields a sparse one (Tarantola, 1987). These two constraints can impose the most common types of solutions desired in geophysical exploration.

A flat solution is created by enforcing a smooth gradient of the model. In this way, the difference between adjacent material properties is minimized (Constable et al., 1987; Duquet et al., 2000). When the second derivatives of the model parameters are smoothed, a continuously varying, smooth solution is retrieved (Menke, 1984). Finally, if a sparse constraint is placed on the gradient of a model, the model solution is piecewise continuous, or blocky, rather than smooth.

This work will focus on sparse solutions encouraged by the longer-tailed probability distribution constraints. The application of sparse constraints are found in two areas: sparse-spike inversion and blocky inversion.

1.4.1 Sparse-spike inversion

Reflectivity information can be used to define the relative impedances of the layers above and below a discontinuity (Beck, 1991). The standard impedance inversion problem is represented in this way: data are assumed to come from a one dimensional model, where the wavefield travels straight down into the earth and is reflected directly back by discontinuities to a receiver. In this simple model, the seismic data can be defined as a convolution of the source with the

Earth's reflectivity. The discrete inverse problem seeks to recover the reflectivity from the data.

A sparse constraint will return the simplest earth model possible: the one with the least number of layers. The relative impedance amplitudes can then be recovered from the reflectivity by an integral or summation (Waters, 1978; Santosa and Symes, 1988). This technique, standard in the exploration geophysics, is known as sparse-spike inversion, despite the fact that the recovered impedances are blocky, and it is the reflectivity that is sparse (Pendrel and van Riel, 2000).

The sparseness constraint was first imposed via an L_1 error norm on the reflectivity series (Oldenburg et al., 1983; Santosa and Symes, 1988; Djikpesse and Tarantola, 1999). The technique is very successful, and is so common that in exploration seismology inversion is frequently accepted to mean sparse-spike or impedance inversion. The sparseness constraint is not limited solely to the L_1 norm, and other long-tailed probability distributions can easily be used (Sacchi, 1997).

1.4.2 Blocky inversion techniques

The idea behind blocky inversion techniques is simple: if the first order derivative is sparse, a function will become piecewise constant. This is seen in blocky impedance inversion: a sparse reflectivity will result in an impedance, or integral of the reflectivity, that is blocky. Blocky inversion is not only applicable in areas where the model parameters are actually blocky in form, but can be used anywhere sharp edges divided by planar features are present. It has the effect of sharpening and focusing the image, reducing the effects of noise and blurring.

The use of blocky inversion began with a constraint composed of the L_1 norm of the gradient of the model parameters. This function is also known as the total variation (TV) or bounded variation (BV) function because as well as

encouraging sparseness, it penalizes very large values (Acar and Vogel, 1994). In this way, the total variation of the model is limited.

The total variation function is useful, but is not differentiable at zero. It also tends to create ringing in the solution due to its sharp truncation. In an effort to circumvent these difficulties, a modified function is introduced (Acar and Vogel, 1994) that adds a small parameter. The updated total variation function is smoother, and defined at all values. This constraint function has been applied to imaging for medical physics (Dobson and Santosa, 1994), illumination by electromagnetic waves (van den Berg and Kleinman, 1995), and satellite imaging (Vogel and Oman, 1998).

Concurrently, the exploration of constraints to enforce sparseness was introduced from the Bayesian perspective in the field of computer vision and tomography (Geman and Geman, 1984). The idea was to assign a different prior distribution on the model parameters than to the normally distributed noise. The function of choice was a long-tailed Gibbs distribution, which would enforce sparseness.

Until this point, the majority of sparse constraints resulted in a non-linear inverse problem. The evolvement of these techniques lead to the realization that solving the non-linear problem was computationally expensive, and not guaranteed to converge to the true solution (Geman and Reynolds, 1992). Further work revealed that though theoretically less accurate, linearization of the non-linear problem was a faster, more dependable way to solve the problem. The series of papers that followed saw a refinement and comparison of different sparse constraint functions that could be easily linearized (Geman and Reynolds, 1992; Geman and Yang, 1995; Charbonnier et al., 1997).

The application of blocky inversion techniques to geophysics has been limited. Most of the activity has been in the area of gravity and magnetic inversion. One

application involves placing a sparse constraint on the volume rather than the gradient of an image, to limit the smearing or blurring of body of interest (Last and Kubik, 1983; Guillen and Menichetti, 1984). With the advances in computer imaging, the concept of blocky gravity inversion using the model gradient has been introduced (Portniaguine and Zhdanov, 1999). Other geophysical applications include using a Bayesian method to encourage strong variations in solutions to geophysical tomographic traveltime data (Clippard et al., 1995). To date, the author has found no evidence of application of these techniques to seismic or GPR data.

1.5 The scope of this thesis

This work will study the applicability of sparse and blocky constraints for applications in exploration geophysics. I choose to work with the function introduced for medical imaging by Charbonnier et al. (1997), because of its simple form that is easily manipulated into a linear inverse problem. As well, this particular constraint has been found to be extremely well suited to enforcing blocky solutions. The function is tested against the other commonly used sparse functions in the context of impedance inversion for the 1D convolutional model of the earth.

The constraint is then extended to a 2D acoustic wavefield migration/inversion. A modified version of the inverse GRT is used to compute synthetic data. By enforcing a sparse constraint on the gradient of the model, blocky or piecewise continuous images of the acoustic velocity are retrieved. I show that blocky inversion recovers sharper, cleaner images of the earth model than the standard inversion procedures.

1.6   Thesis outline

The structure of the thesis is as follows:

- Chapter 1 provides the motivation and scope of my research.

- Chapter 2 is an introduction to Born linear scattering theory, and direct inversion using the asymptotic inverse of the GRT. The modification of the inverse GRT for the discrete, numerical migration/inversion algorithm is then presented.

- In Chapter 3, discrete inverse theory and the common quadratic solution constraints are introduced.

- Chapter 4 continues the discussion of constraints from a Bayesian perspective, and introduces sparse constraints. The computational methods of solving the discrete inverse problem are explored. Finally, my sparseness constraint is compared to other common constraints in the 1D impedance inversion problem.

- Chapter 5 describes the blocky inversion algorithm using a sparseness constraint, and illustrates its effectiveness for the 2D acoustic migration/inversion problem.

- Chapter 6 summarizes my research and the most relevant findings. Finally, the advantages and disadvantages of the sparse constraint are discussed.

Chapter 2

The linear scattering problem

In geophysical exploration, wavefield data created by an induced source are recorded to collect information about the subsurface. This seismic experiment can be considered a tomographic problem, described by the linear scattering integral (Miller et al., 1987). The integral directly relates data and model space. Forward modeling is achieved by predicting the data created by a known Earth model. Because the geophysical experiment records the resulting wavefield, inverse modeling, or recovering the earth image from known data, is of more interest.

The inverse can be difficult to recover, as it requires perfectly complete and accurate data (Claerbout, 1992). As well, it is generally unstable. Unlike the inverse, the adjoint can deal with incomplete, inaccurate data. Fortunately, in geophysical problems the adjoint retrieves an imperfect form of the inverse and can be used to approximate inverse modeling.

Seismic imaging refers to the creation of an image of the Earth's geology using inversion. There are two ways to create a seismic image: direct or discrete inversion (Berkhout, 1984; Jin et al., 1992). Direct inversion uses an analytical method to find the inverse. The inverse may then be computed numerically. Al-

ternatively, discrete inversion uses a discrete, numerical method to approximate the inverse, relying on adjoint modeling.

The linear scattering integral can be defined in terms of the generalized Radon transform (GRT), to directly relate the data and model space. The linear scattering problem may then be solved using an asymptotic inverse to the GRT (Beylkin, 1985; Miller et al., 1987). This is a direct inversion method. It has the advantage of being relatively computationally inexpensive, and the disadvantage of working only with particular source-receiver geometries (Berkhout, 1984). Direct inversion is also known as migration, as it projects or migrates data back into model space.

In order to render the imaging technique more flexible, a computational rather than analytical method of solving the linear scattering problem is presented. Discrete inversion recovers the Earth model by minimizing differences between the observed and synthetic data. Methods of modeling synthetic data such as ray theoretic approaches (Beydoun and Mendez, 1989), or finite difference algorithms (Crase et al., 1990) are computationally taxing. To circumvent this, computational methods for forward and adjoint modeling are adapted from direct inversion via the GRT. This enables the two inversion methods to be combined in the manner of Thierry et al. (1999) to solve acoustic, constant density Earth models. The resulting algorithm has the benefits of being computationally efficient, and the ability to adapt to any source-receiver geometry.

2.1   Linearizing the scattering problem

The linear scattering relation is a standard representation for problems in optics and geophysics (Miller et al., 1987; Thierry et al., 1999). In three dimensions, the acoustic wave equation of a constant density medium is denoted by the scalar wave equation (for derivation see Appendix A)

$$\nabla^2 u(\mathbf{x}, \mathbf{s}, \omega) + \frac{\omega^2}{c^2(\mathbf{x})} u(\mathbf{x}, \mathbf{s}, \omega) = -\delta(\mathbf{x} - \mathbf{s}) \,. \qquad (2.1.1)$$

In this equation, $u(\mathbf{x}, \mathbf{s}, \omega)$ is the acoustic pressure field or displacement, $\mathbf{s}$ and $\mathbf{x}$ are the positions of the source and an arbitrary point within the 3-D model space, and $c(\mathbf{x})$ is the acoustic velocity. The acoustic scattering potential is defined as

$$f(\mathbf{x}) = \frac{1}{c^2(\mathbf{x})} - \frac{1}{c_0^2(\mathbf{x})} \,, \qquad (2.1.2)$$

where $f(\mathbf{x})$ is the perturbation from the known background velocity field $c_0(\mathbf{x})$. The acoustic wave equation can be rewritten in terms of the unknown acoustic perturbation as

$$\nabla^2 u(\mathbf{x}, \mathbf{s}, \omega) + \frac{\omega^2}{c_0^2(\mathbf{x})} u(\mathbf{x}, \mathbf{s}, \omega) = -\delta(\mathbf{x} - \mathbf{s}) - \omega^2 f(\mathbf{x}) u(\mathbf{x}, \mathbf{s}, \omega) \,. \qquad (2.1.3)$$

The wave equation of the background wavefield can be written in terms of Green's functions as

$$\nabla^2 G_0(\mathbf{x}, \mathbf{y}, \omega) + \frac{\omega^2}{c_0^2(\mathbf{x})} G_0(\mathbf{x}, \mathbf{y}, \omega) = -\delta(\mathbf{x} - \mathbf{y}) \,. \qquad (2.1.4)$$

The Lippman-Schwinger equation of quantum mechanics (Taylor, 1972) combines expressions (2.1.3) and (2.1.4) to relate the background and total wavefields. Thus, one can obtain the expression for the wavefield at receiver position $\mathbf{r}$

$$u(\mathbf{r}, \mathbf{s}, \omega) = \underbrace{G_0(\mathbf{r}, \mathbf{s}, \omega)}_{Background\ wavefield} + \underbrace{\omega^2 \int G_0(\mathbf{r}, \mathbf{x}, \omega) f(\mathbf{x}) u(\mathbf{x}, \mathbf{s}, \omega) d^3 \mathbf{x}}_{Scattered\ wavefield} \,. \qquad (2.1.5)$$

The total wavefield can be broken into two parts: the background wavefield and the scattered wavefield. By rearranging this definition, one can state

$$u_{sc}(\mathbf{r}, \mathbf{s}, \omega) = u(\mathbf{r}, \mathbf{s}, \omega) - G_0(\mathbf{r}, \mathbf{s}, \omega) \,, \tag{2.1.6}$$

and therefore that

$$u_{sc}(\mathbf{r}, \mathbf{s}, \omega) = \omega^2 \int G_0(\mathbf{r}, \mathbf{x}, \omega) f(\mathbf{x}) u(\mathbf{x}, \mathbf{s}, \omega) d^3\mathbf{x} \,. \tag{2.1.7}$$

As it is written, this expression is a non-linear equation; the scattered wave-field is a function of the total wavefield. In order to create a linear expression the Born approximation is used. This approximation states that $u_{sc} << u$, which leads to

$$u(\mathbf{r}, \mathbf{s}, \omega) \cong G_0(\mathbf{r}, \mathbf{s}, \omega) \,. \tag{2.1.8}$$

The linearized scattering problem is written as

$$u_{sc}(\mathbf{r}, \mathbf{s}, \omega) = \omega^2 \int G_0(\mathbf{r}, \mathbf{x}, \omega) f(\mathbf{x}) G_0(\mathbf{x}, \mathbf{s}, \omega) d^3\mathbf{x} \,, \tag{2.1.9}$$

where the acoustic potential must be small to satisfy the approximation.

The Green's functions for an arbitrary background velocity field can be expressed using the first-order asymptotic expansion of the Hankel function given by geometrical optics (Beylkin, 1985). In this way, the Green's functions are established to be

$$G_0(\mathbf{x}, \mathbf{y}, \omega) = A(\mathbf{x}, \mathbf{y}) e^{i\omega\tau(\mathbf{x}, \mathbf{y})} \,, \tag{2.1.10}$$

where $A(\mathbf{x}, \mathbf{y})$ represents the amplitude of the ray that travels from $\mathbf{x}$ to $\mathbf{y}$, and τ is the traveltime between the two points. The amplitude term must satisfy the transport equation

$$A(\mathbf{x},\mathbf{y})\nabla_{\mathbf{x}}^2\tau(\mathbf{x},\mathbf{y}) + 2\,\nabla_{\mathbf{x}}A(\mathbf{x},\mathbf{y}) \cdot \nabla_{\mathbf{x}}\tau(\mathbf{x},\mathbf{y}) = 0\,, \qquad (2.1.11)$$

and the traveltime function, $\tau(\mathbf{x},\mathbf{y})$ is defined by the eikonal equation

$$[\nabla_{\mathbf{x}}\tau(\mathbf{x},\mathbf{y})]^2 = \frac{1}{c_0^2(\mathbf{x})}\,. \qquad (2.1.12)$$

Using the definition of Green's functions, the scattered field becomes

$$u_{sc}(\mathbf{r},\mathbf{s},\omega) = \omega^2 \int A(\mathbf{r},\mathbf{x})e^{i\omega\tau(\mathbf{r},\mathbf{x})}f(\mathbf{x})A(\mathbf{x},\mathbf{s})e^{i\omega\tau(\mathbf{x},\mathbf{s})}d^3\mathbf{x}\,. \qquad (2.1.13)$$

New amplitude and traveltime functions are defined, composed of terms for the incident and scattered wavefields:

$$A(\mathbf{r},\mathbf{x},\mathbf{s}) = A(\mathbf{r},\mathbf{x})A(\mathbf{x},\mathbf{s})\,,$$

and

$$\tau(\mathbf{r},\mathbf{x},\mathbf{s}) = \tau(\mathbf{r},\mathbf{x}) + \tau(\mathbf{x},\mathbf{s})\,.$$

Thus, the scattering equation can be rewritten

$$u_{sc}(\mathbf{r},\mathbf{s},\omega) = \omega^2 \int A(\mathbf{r},\mathbf{x},\mathbf{s})e^{i\omega\tau(\mathbf{r},\mathbf{x},\mathbf{s})}f(\mathbf{x})d^3\mathbf{x}\,. \qquad (2.1.14)$$

Using the Fourier transform to express this in the time domain results in

$$u_{sc}(\mathbf{r},\mathbf{s},t) = -\frac{\partial^2}{\partial t^2} \int A(\mathbf{r},\mathbf{x},\mathbf{s})\delta[t - \tau(\mathbf{r},\mathbf{x},\mathbf{s})]f(\mathbf{x})d^3\mathbf{x}\,. \qquad (2.1.15)$$

Either expression (2.1.14) or (2.1.15) may be referred to as the linear scattering integral.

2.1.1 The linear scattering problem for a source wavelet

When the source is not a delta function, the scattered wavefield will include information of the source wavelet. This is written as

$$u_{sc}(\mathbf{r}, \mathbf{s}, \omega) = \omega^2 \int A(\mathbf{r}, \mathbf{x}, \mathbf{s}) e^{i\omega\tau(\mathbf{r}, \mathbf{x}, \mathbf{s})} f(\mathbf{x}) W(\omega) d^3\mathbf{x} \,, \tag{2.1.16}$$

or as a convolution in the time domain

$$u_{sc}(\mathbf{r}, \mathbf{s}, t) = \left[-\frac{\partial^2}{\partial t^2} \int A(\mathbf{r}, \mathbf{x}, \mathbf{s}) \delta[t - \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})] f(\mathbf{x}) d^3\mathbf{x} \right] \circledast w(t) \,. \tag{2.1.17}$$

$W(\omega)$ and $w(t)$ represent the source wavelet in the frequency and time domains respectively. The time derivative can be shifted onto the wavelet (Tarantola, 1984) to create the new expression

$$u_{sc}(\mathbf{r}, \mathbf{s}, t) = \left[\int A(\mathbf{r}, \mathbf{x}, \mathbf{s}) \delta[t - \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})] f(\mathbf{x}) d^3\mathbf{x} \right] \circledast \frac{-\partial^2 w(t)}{\partial t^2} \,. \tag{2.1.18}$$

2.1.2 The linear scattering problem in two dimensions

The linear scattering problem can be easily adapted to two dimensions. In this context, the Green's functions are approximate. This occurs because the first-order asymptotic expansion of the Hankel function is exact only for the 3D problem (Beylkin, 1985). Using this expansion, the Green's functions are expressed as

$$G_0(\mathbf{x}, \mathbf{y}, \omega) = (-i\omega)^{\frac{-1}{2}} A(\mathbf{x}, \mathbf{y}) e^{i\omega\tau(\mathbf{x}, \mathbf{y})} \,. \tag{2.1.19}$$

The scattering equation for a delta function as the source is:

$$u_{sc}(\mathbf{r}, \mathbf{s}, \omega) = i\omega \int A(\mathbf{r}, \mathbf{x}, \mathbf{s}) e^{i\omega\tau(\mathbf{r}, \mathbf{x}, \mathbf{s})} f(\mathbf{x}) d^2\mathbf{x} \,. \qquad (2.1.20)$$

If the source has a signature wavelet, $W(\omega)$, the expression becomes

$$u_{sc}(\mathbf{r}, \mathbf{s}, \omega) = i\omega \int A(\mathbf{r}, \mathbf{x}, \mathbf{s}) e^{i\omega\tau(\mathbf{r}, \mathbf{x}, \mathbf{s})} f(\mathbf{x}) W(\omega) d^2\mathbf{x} \,, \qquad (2.1.21)$$

or in the time domain

$$u_{sc}(\mathbf{r}, \mathbf{s}, t) = \int A(\mathbf{r}, \mathbf{x}, \mathbf{s}) \delta[t - \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})] f(\mathbf{x}) d^2\mathbf{x} \circledast \frac{\partial w(t)}{\partial t} \,. \qquad (2.1.22)$$

2.2   An asymptotic inverse operator

The true inverse of the linear scattering integral will recover the original earth model from scattered acoustic wavefield data. Unfortunately, a general expression of this inverse does not exist, and so an approximation must be made. This approximation begins by identifying the linear scattering integral as a modified form of the generalized Radon transform (Beylkin, 1985). The GRT does not have a true inverse, but does have an adjoint expression. The adjoint recovers a distorted version of the original model. The asymptotic inverse to the scattering problem is created by applying a weighting function to the adjoint of the GRT. An analytic expression of the weighting function for common source-receiver geometries is then used to recover the original earth model (Miller et al., 1987; Bleistein, 1987).

2.2.1   The generalized Radon transform

The GRT is defined to be (Ramm and Katsevich, 1996)

$$f(\xi, p) = \int g(p, \xi, \mathbf{x}) \delta[p - \xi \cdot \mathbf{x}] f(\mathbf{x}) d^3 \mathbf{x}. \tag{2.2.1}$$

Here $g(p, \xi, \mathbf{x})$ is a weighting function, ξ specifies a family of "parallel" isochron surfaces passing through point $\mathbf{x}$, and p fixes one of these isochrons. Thus the GRT is an integral over a fixed isochron surface. There is no inverse to this transform, so the adjoint is used to recover $f(\mathbf{x})$. The function at model point $\mathbf{y_0}$, is found by evaluating the following integral at $p = \xi \cdot \mathbf{y_0}$

$$\widetilde{f}(\mathbf{y_0}) = \frac{-1}{8\pi^2} \int dw(p, \xi, \mathbf{y_0}) \frac{\partial^2}{\partial p^2} \int g(p, \xi, \mathbf{x}) \delta[p - \xi \cdot \mathbf{x}] f(\mathbf{x}) d^3 \mathbf{x}. \tag{2.2.2}$$

This equation introduces an additional weighting function, $dw(p, \xi, \mathbf{y_0})$, in the expression of the recovered acoustic potential, $\widetilde{f}(\mathbf{y_0})$.

Adjoint modeling using the generalized Radon transform

Recall that:

$$u_{sc}(\mathbf{r}, \mathbf{s}, t) = -\frac{\partial^2}{\partial t^2} \int A(\mathbf{r}, \mathbf{x}, \mathbf{s}) \delta[t - \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})] f(\mathbf{x}) d^3 \mathbf{x}. \tag{2.2.3}$$

By using the definition of the GRT, the linearized scattering problem can be written as

$$u_{sc}(\mathbf{r}, \mathbf{s}, t) = -\frac{\partial^2}{\partial t^2} f^{GRT}(\mathbf{r}, \mathbf{s}, t), \tag{2.2.4}$$

where $f^{GRT}(\mathbf{r}, \mathbf{s}, t)$ is the GRT of $f(\mathbf{x})$. This reveals that our forward modeling is simply a filtered version of the generalized Radon transform. The adjoint operator can be recovered analytically using the adjoint of the GRT

$$\widetilde{f}(\mathbf{y_0}) = \frac{-1}{8\pi^2} \int dW(\mathbf{r}, \mathbf{y_0}, \mathbf{s}) \frac{\partial^2}{\partial t^2} \int A(\mathbf{r}, \mathbf{x}, \mathbf{s})\delta[t - \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})]f(\mathbf{x})d^3\mathbf{x}, \quad (2.2.5)$$

where $dW(\mathbf{r}, \mathbf{y_0}, \mathbf{s})$ is a weighting function. The integral is evaluated at $t = \tau(\mathbf{r}, \mathbf{y_0}, \mathbf{s})$.

Conceptually, the traveltime function $\tau(\mathbf{r}, \mathbf{y_0}, \mathbf{s})$ specifies a family of "parallel" isochron surfaces. When the background velocity is constant, the traveltime isochrons will become concentric ellipsoids with foci at the source and receiver positions. Specifying one source-receiver pair and a traveltime, t, will isolate an isochron surface. Integration then proceeds over this surface. The previous formula can be reduced to:

$$\widetilde{f}(\mathbf{y_0}) = \frac{1}{8\pi^2} \int dW(\mathbf{r}, \mathbf{y_0}, \mathbf{s}) u_{sc}(\mathbf{r}, \mathbf{s}, t). \quad (2.2.6)$$

By substituting in the definition of the scattered wavefield (2.1.18), this becomes

$$\widetilde{f}(\mathbf{y_0}) = \frac{1}{8\pi^2} \int dW(\mathbf{r}, \mathbf{y_0}, \mathbf{s}) \times$$
$$\left[\left(\int A(\mathbf{r}, \mathbf{x}, \mathbf{s})\delta[t - \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})]f(\mathbf{x})d^3\mathbf{x} \right) \circledast \frac{-\partial^2 w(t)}{\partial t^2} \right]. \quad (2.2.7)$$

The inverse solution involves solving the weighting function to recover $f(\mathbf{x})$. The weighting function undoes the convolution with the source wavelet, and removes the amplitude terms for the source-receiver geometry. Specific solutions for common source-receiver geometries are discussed by Miller et al. (1987), and Bleistein (1987).

2.3 Modeling the linear scattering problem

In discrete inversion, the seismic image is adjusted by comparing the observed data to the synthetic data calculated from the Earth model. Thus, methods of computing the synthetic data from a model, and then projecting the data back into model space are necessary. These steps are referred to as forward and adjoint modeling respectively (Claerbout, 1992).

2.3.1 Forward modeling

The linear scattering equation can be used to calculate the scattered wavefield of a seismic experiment for a known acoustic potential. The integral scattering equation in time (2.1.18) can be expressed as a linear Born operator. The Born operator, denoted $\mathcal{B}$, can be broken into two distinct parts: an integral operator $\mathcal{L}$, and a convolution operator $\mathcal{C}$

$$
\begin{aligned}
d(\mathbf{r}, \mathbf{s}, t) &= \mathcal{B}f(\mathbf{x}) \\
&= \mathcal{C}\mathcal{L}f(\mathbf{x}) .
\end{aligned}
\tag{2.3.1}
$$

In this equation, $f(\mathbf{x})$ is the acoustic potential, and $d(\mathbf{r}, \mathbf{s}, t)$ are the synthetic data. Both are continuous functions. The integral operator denotes

$$
(\mathcal{L}f)(\mathbf{r}, \mathbf{s}, t) = \int A(\mathbf{r}, \mathbf{x}, \mathbf{s})\delta[t - \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})]f(\mathbf{x})d^3\mathbf{x} .
\tag{2.3.2}
$$

The convolution operator is defined as

$$
(\mathcal{C}r)(t) = \frac{-\partial^2 w(t)}{\partial t^2} \circledast r(t) ,
\tag{2.3.3}
$$

where $r(t)$ is a dummy variable. Both operators apply to continuous functions, and may be expressed in the frequency domain. When the Fourier transform is applied, they become

$$(\mathcal{L}f)(\mathbf{r}, \mathbf{s}, \omega) = \int A(\mathbf{r}, \mathbf{x}, \mathbf{s})e^{i\omega\tau(\mathbf{r}, \mathbf{x}, \mathbf{s})}f(\mathbf{x})d^3\mathbf{x}, \qquad (2.3.4)$$

and

$$(\mathcal{C}R)(\omega) = \omega^2 W(\omega)R(\omega). \qquad (2.3.5)$$

2.3.2 The discrete forward operator

The earth's properties may be discrete, such as the Earth's mass, or continuous, like its density. Continuous properties of the earth may be approximated as discrete by assigning an average value of the property to each cell of a tomographic grid. To approximate this realistically, a large number of cells are necessary. Geophysical data are already discrete, because they are recorded at regularly or irregularly sampled spatial locations.

When working with discrete rather than continuous properties, the linear operators can be thought of as matrices, and the parameters that they act on as vectors. The discrete forward problem is expressed as

$$\mathbf{d} = \mathbf{CLf}, \qquad (2.3.6)$$

where $\mathbf{d}$ and $\mathbf{f}$ are vectors of the data and acoustic potential, and the matrices $\mathbf{C}$ and $\mathbf{L}$ are the convolution and integrator operators. The convolutional operator, $\mathbf{C}$, can easily be placed in matrix form using the method described in Appendix C. The operator matrix=, $\mathbf{L}$, is constructed by placing an amplitude term in the

matrix in order to weight the magnitude of the acoustic perturbations. These weighted terms are placed at time samples in the data space that match the traveltime between the source, receiver, and current point in model space. If this sounds complicated and non-intuitive, you are reading correctly. This is why it is preferable to use subroutines as operators, discussed further on in this chapter.

2.3.3 Adjoint modeling

The opposite of forward modeling is adjoint modeling. Adjoint modeling is based on adjoint operators, which are analogous to taking the transpose of a matrix. In certain problems, the transpose is a good approximation to the direct inverse, and the linearized scattering problem falls into this category (Claerbout, 1992). Discrete inverse theory uses forward and adjoint modeling to numerically approximate the inverse.

The adjoint modeling operator is thus

$$\widetilde{f}(\mathbf{x}) = [\mathcal{C}\mathcal{L}]^T d(\mathbf{r}, \mathbf{s}, t) \,, \tag{2.3.7}$$

where $\widetilde{f}(\mathbf{x})$ is recovered acoustic potential, and $d(\mathbf{r}, \mathbf{s}, t)$ are scattered wavefield data. By using the definition of a transpose, one can write

$$\widetilde{f}(\mathbf{x}) = \mathcal{L}^T \mathcal{C}^T d(\mathbf{r}, \mathbf{s}, t) \,. \tag{2.3.8}$$

This states that the adjoint to the forward modeling operator involves first undoing the convolution with the time derivative of the source signature, and then reversing the summation over the traveltime isochron. Suppose that $d' = \mathcal{C}^T d$

$$(\mathcal{C}^T d)\,(\mathbf{r}, \mathbf{s}, \omega) = d(\mathbf{r}, \mathbf{s}, \omega) W^*(\omega)\, \omega^2 \,, \tag{2.3.9}$$

where the conjugate of the wavelet is denoted $W^*(\omega)$. The integral expression for $\mathcal{L}^T d'$ becomes

$$\left(\mathcal{L}^T d'\right)(\mathbf{y_0}) = \int \int \int G_0^*(\mathbf{r}, \mathbf{y_0}, \omega) \, G_0^*(\mathbf{y_0}, \mathbf{s}, \omega) \, d'(\mathbf{r}, \mathbf{s}, \omega) \, d\mathbf{r} \, d\mathbf{s} \, d\omega, \qquad (2.3.10)$$

where all source and receiver combinations are integrated over, and $\mathbf{y_0}$ is an arbitrary point within the model space. $G_0^*(\mathbf{r}, \mathbf{y_0}, \omega)$ indicates the adjoint or conjugate Green's functions, defined as

$$G_0^*(\mathbf{r}, \mathbf{y_0}, \omega) = A(\mathbf{r}, \mathbf{y_0}) \, e^{-i\omega\tau(\mathbf{r},\mathbf{y_0})}. \qquad (2.3.11)$$

The forward linear Born operator transforms information from the model space to the data space. The adjoint operator reverses this transformation. This movement in time and space is why an adjoint operator is often referred to as a backprojection or operator: it projects the information moved by the forward operator back to where it originated from. This idea is evident in the change of sign that occurs in the adjoint Green's functions.

Using the definition of the scattered wavefield expressed in the frequency domain (2.1.16), and the notation of Green's functions, the model recovered by adjoint modeling can be rewritten as

$$\widetilde{f}(\mathbf{y_0}) = \left(\mathcal{L}^T d'\right)(\mathbf{y_0}) = \int \int \int \int \omega^4 \, G_0^*(\mathbf{r}, \mathbf{y_0}, \omega) \, G_0^*(\mathbf{y_0}, \mathbf{s}, \omega) \, \times$$
$$G_0(\mathbf{r}, \mathbf{x}, \omega) \, G_0(\mathbf{x}, \mathbf{s}, \omega) W(\omega) W^*(\omega) \, f(\mathbf{x}) \, d^3\mathbf{x} \, d\mathbf{r} \, d\mathbf{s} \, d\omega. \qquad (2.3.12)$$

Integrating the Green's functions and wavelet of equation (2.3.12) over source-receiver geometry and frequency results in a single kernel $\mathcal{K}(\mathbf{y_0}, \mathbf{x})$. The integral can then be written as

$$\widetilde{f}(\mathbf{y_0}) = \int \mathcal{K}(\mathbf{y_0}, \mathbf{x})\, f(\mathbf{x})\, d^3\mathbf{x}\,. \qquad (2.3.13)$$

Equation (2.3.13) would become the inverse operator if the kernel were a delta-like kernel. Comparing equations (2.2.6) and (2.3.13) indicates that the weighting function in the asymptotic inverse collapses the kernel to a delta-like function that ensures the correct magnitude of the acoustic potential. However, when using adjoint operators, the kernel does not completely collapse to a delta function. This occurs because even though Green's functions are defined as the impulse response of a linear system (Snieder, 1998), integration takes place over a finite region for a limited source-receiver geometry. As well, the adjoint of the wavelet is not the same as its inverse, and further inaccuracy results. Therefore, the kernel is an approximation to a delta function, resulting in a solution that is a blurred version of the original image (Figure 2.1).

2.3.4   The discrete adjoint operator

When working with discrete expressions of the model and data, the adjoint operators are implemented as matrices. In keeping with previous notation, the discrete adjoint operator is expressed as

$$\widetilde{\mathbf{f}} = [\mathbf{L}^T\mathbf{C}^T]\mathbf{d}. \qquad (2.3.14)$$

It is interesting to note that the adjoint of the convolution operator is cross-correlation (see Appendix C).

2.3.5   Subroutines as operators

In simple examples it is easy to work with matrices. However, as the problem becomes larger and more complicated, the matrices are not so simply defined. It
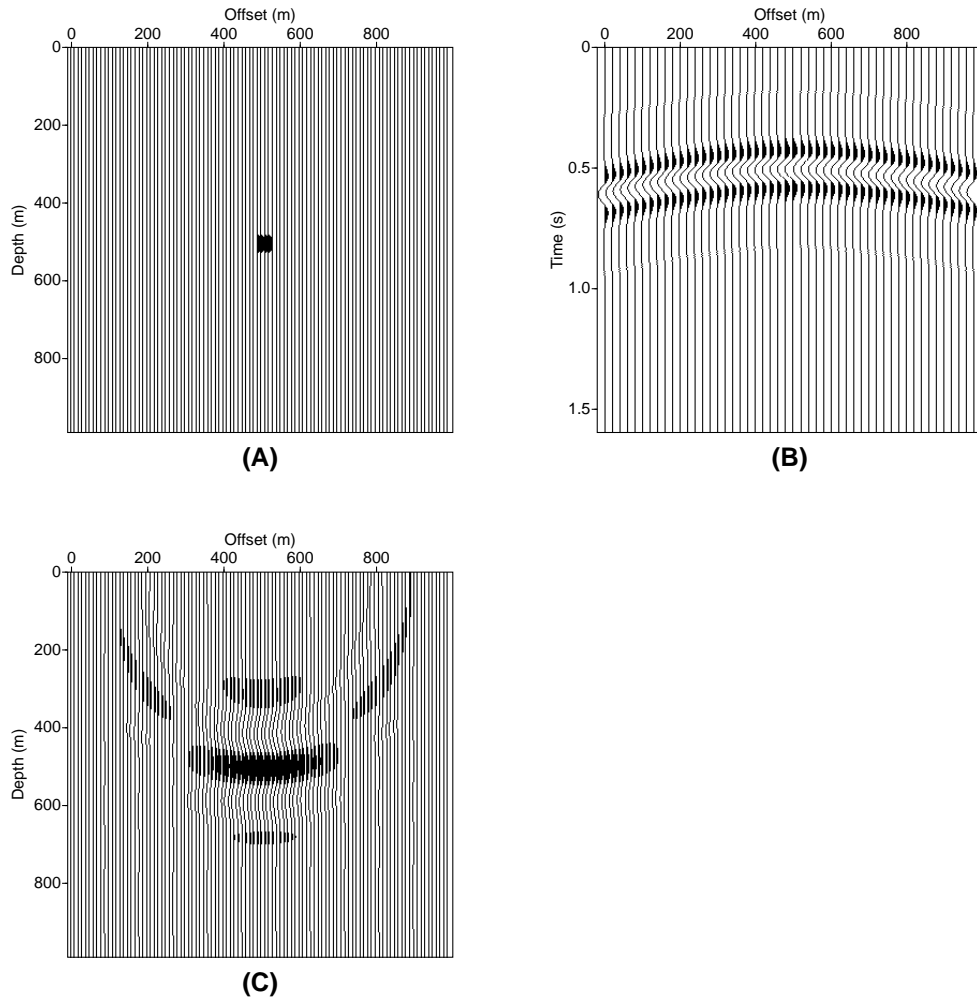
Figure 2.1: (A) The original acoustic potential $f(\mathbf{x})$. (B) Synthetic data created by the forward operator $\mathbf{d} = \mathbf{CLf}$. (C) The smeared acoustic potential recovered by the adjoint operator $\widetilde{\mathbf{f}} = \mathbf{L}^T \mathbf{C}^T \mathbf{d}$. Note that the adjoint recovers a blurred version of the true perturbation. This is because the adjoint is not an inverse operator, only an approximation to it.

is possible to circumvent the need for matrices by simply writing subroutines that perform the same action. For instance, instead of building a matrix convolution operator $\mathbf{C}$, the forward and adjoint operations can be performed by convolution and cross-correlation subroutines.

In order to check that two subroutines are true adjoints of each other, the dot-product test is used. Suppose $\mathbf{y} = \mathbf{Gx}$, where $\mathbf{G}$ and $\mathbf{G}^T$ are the forward and adjoint operators respectively. One can state

$$\mathbf{y}^T(\mathbf{Gx}) = (\mathbf{y}^T\mathbf{G})\mathbf{x}, \qquad (2.3.15)$$

or

$$\mathbf{y}^T(\mathbf{Gx}) = (\mathbf{G}^T\mathbf{y})^T\mathbf{x}. \qquad (2.3.16)$$

The dot-product test, as described by Claerbout (1992) contains these steps:

- Create two vectors of random numbers, $\mathbf{x}_1$ and $\mathbf{y}_2$.

- Compute $\mathbf{y}_1 = \mathbf{Gx}_1$ and $\mathbf{x}_2 = \mathbf{G}^T\mathbf{y}_2$ using the programs for the forward and adjoint operators.

- Equation (2.3.16) states that $\mathbf{y}_1^T\mathbf{y}_2 = \mathbf{x}_1^T\mathbf{x}_2$.

If this last relation is satisfied, $\mathbf{G}^T$ is a true adjoint of $\mathbf{G}$.

2.4   Implementation of Modeling Operators

To retain the flexibility to handle arbitrary source-receiver geometries, the solution is constructed using operators rather than the asymptotic inverse. The more complex nature of the forward and adjoint operators means that subroutines as

operators are preferable to creating matrices. An algorithm is written to create

forward and adjoint models for a 2D acoustic, constant density earth. Data are

represented in the time domain. The pseudo-code for the subroutine is given

below. The algorithm can also be thought of as a mapping function which moves

between model and data space, as illustrated in Figure 2.2.

```
#  Subroutine to perform forward/adjoint modeling operators
if adjoint operator
   then correlate wavelet with data:   d_temp= C^T d
endif
for itraces = all receivers for each source
   for x = all (x,z) positions in 2D earth model
      distance = |x-r| + |x-s|
      # where r and s are 2D position vectors of receivers and sources
      time = distance/velocity

      if forward operator
         then create data from potential:   Lf = d_temp
         d_temp(itrace,time) = d_temp(itrace,time) + A(r,x,s)*f(x)
      elseif adjoint operator
         then create potential from data:   L^T d_temp = f̃
         f̃(x) = f̃(x) + A(r,x,s)*d_temp(itrace,time)
      endif

   endfor
endfor
```
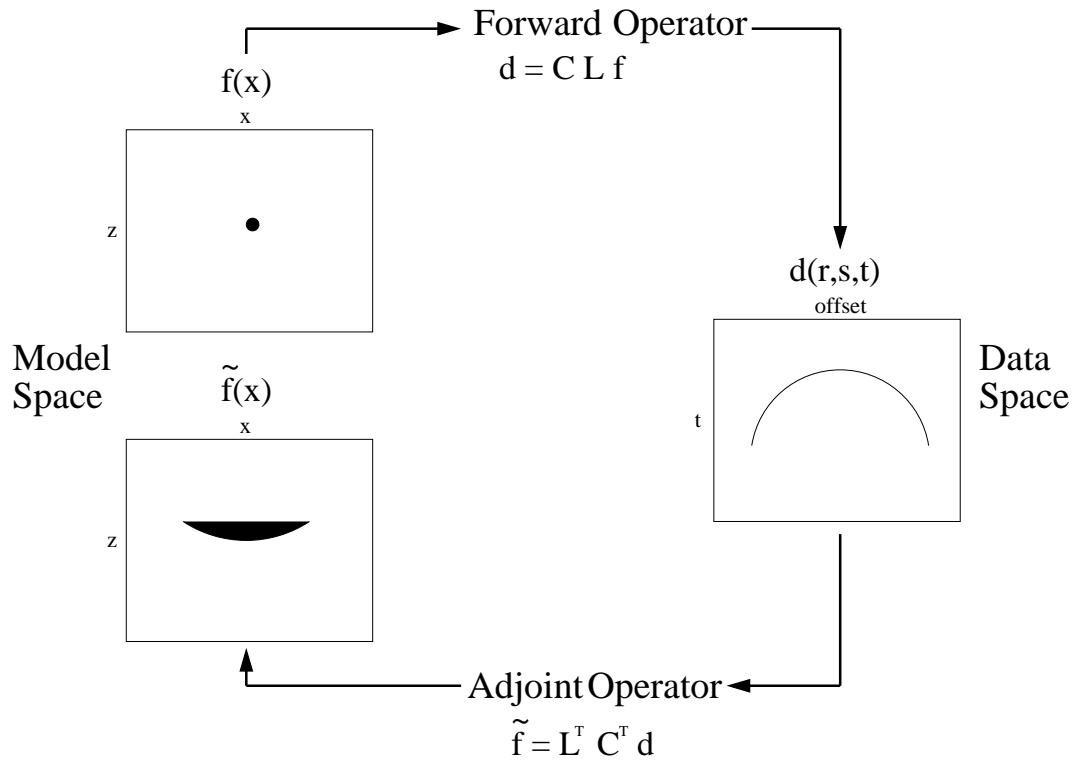
Figure 2.2: A pictorial representation of the modeling subroutine that maps between model and data space.

```
if forward operator
    then convolve wavelet and data:   d = Cd_temp
endif
```

2.5 Summary

In this chapter, the linear scattering integral arising from the Born approximation was derived. This integral can be expressed as a filtered form of the generalized Radon transform. Taking advantage of this, the inverse to the linear scattering problem is found using an asymptotic inverse of the GRT. This is a method of direct inversion.

The direct inversion solution is only applicable for specific source-receiver geometries, so that a new analytic formula must be used each time these geometries change. The technique of discrete inversion does not have this limitation because the problem is solved numerically. The forward and adjoint discrete modeling operators are derived from the linear scattering integral. This enables a numerical approximation to direct inversion to be placed within the framework of discrete inversion.

Chapter 3

Discrete inverse theory

Discrete geophysical inverse theory is the study of computational techniques to solve the linear inverse problem. The solution to the inverse problem is found by minimizing the difference between the observed data and synthetic data for an Earth model. Within this chapter, the synthetic data are computed using the discrete operators described in Chapter 2. These operators are adapted from the direct inversion of the linear scattering problem using the GRT. This will combine the computational advantages of direct inversion with ability to adapt to any source-receiver geometry possessed by the numerical, discrete approach.

The error between the observed and synthetic data is defined in terms of the L_2 norm. This definition of error is stated in the cost function: a mathematical expression of the desired characteristics of the solution. Constraints on the solution are also included in the cost function as terms known as regularization terms.

Most geophysical problems fall into the category of ill-posed problems. These require constraints on the Earth properties to retrieve a stable, unique solution. The most common constraint is to enforce a smooth solution (Constable et al., 1987; Duquet et al., 2000). This is desirable because a smoothly varying earth

with the simplest possible structure is retrieved.

The solution to the discrete inverse problem can be found using either direct or indirect computation (Strang, 1986). Direct computation is unrelated to direct inversion methods, and involves direct matrix inversion to find a numerical solution. This method requires the discrete operators to be in matrix form. Indirect computation finds the solution in an iterative manner, and works with operators in matrix or subroutine form. The indirect method is the most common technique for solving large geophysical inverse problems because of its computational advantages.

3.1 Classifications of discrete inverse problems

Inverse theory makes no assumption as to the completeness and accuracy of the data. The classification of the problem will depend on the number of known data and unknown model parameters, N and M respectively (Menke, 1984). A simple example of an inverse problem is fitting a line to multiple points, where the data points are designated as a vector $\mathbf{d} = [d_1, d_2, d_3, ...]^T$. The model parameters to be recovered are slope, s, and intercept, b, so that $\mathbf{m} = [s, b]^T$. The statement of the discrete inverse problem is given by

$$d_i = \sum_{j=1}^{M} G_{ij} m_j , \quad i = 1, 2, \ldots N \tag{3.1.1}$$

where G_{ij} is the ij^{th} element of the matrix operator $\mathbf{G}$. The matrix operator linearly relates data and model space.

3.1.1 The even-determined problem

The inverse problem will have one definite solution when there is exactly enough information to determine the problem uniquely. In the case of fitting a line, this

occurs when there are two data points without noise, or $N = M$. This situation never occurs in geophysical inverse problems, because it is impossible to gather noise-free data, and to evaluate all of the earth's material properties within every cell of a tomographic grid.

3.1.2 The over-determined problem

The over-determined problem occurs where there are more observations than unknowns $(N > M)$. The linear problem could possibly have three or more points that do not lie along one straight line. There is no perfect solution to this problem, so inverse theory seeks to find the best solution. The least squares approach defines the best solution as one where the error, or distance between each point and the best fit line, is minimized. Using the definition of the L_2 norm to define distance, the cost or objective function to be minimized is

$$\begin{aligned} J(\mathbf{m}) &= \sum_{i=1}^{N} |d_i^{obs} - d_i^{pred}|^2 \\ &= (\mathbf{d} - \mathbf{Gm})^T (\mathbf{d} - \mathbf{Gm}) \,. \end{aligned} \qquad (3.1.2)$$

Within this expression, the predicted or synthetic data are defined $\mathbf{d}^{pred} = \mathbf{Gm}$.

3.1.3 The under-determined problem

An under-determined problem has less observations than unknowns $(N < M)$. An example of this type of problem is fitting a line through one data point. There are an infinite number of solutions that fit the data perfectly. To solve the problem, one solution must be defined as more desirable than the others by imposing a constraint. Within the cost function, the extra term that describes the constraint is known as the regularization term.

A popular constraint is to enforce a minimum norm solution, or one that has a minimum Euclidean length. At this point, the assumption is that the data are free of noise and inaccuracy so that the solution must satisfy $\mathbf{Gm} - \mathbf{d} = 0$. The cost function becomes

$$J(\mathbf{m}) = \lambda^T (\mathbf{d} - \mathbf{Gm}) + \mathbf{m}^T \mathbf{m}, \qquad (3.1.3)$$

where λ^T is a vector of Lagrange multipliers.

3.1.4 The ill-posed problem

Most geophysical problems lie in the category of ill-posed problems (Menke, 1984). This type of problem is neither purely under- or over-determined. It characteristically has more observations than unknowns, but these observations are not linearly independent. Thus, they give redundant information about the problem. An example of this would be a scattered wavefield travelling through the same tomographic cells for repeated experiments, while others are never crossed. Some areas of the grid have no information, while others may have multiple observations for particular model parameters.

Within this particular example of the ill-posed problem, the observations are known to be noisy and inaccurate. The inverse problem is governed by the relation

$$\mathbf{d} = \mathbf{Gm} + \mathbf{n}, \qquad (3.1.4)$$

where $\mathbf{n}$ is Gaussian noise.

The cost function of an ill-posed problem contains a data misfit term and a constraint term to satisfy its mixed-determined nature. In this example, an L_2 measure of misfit is chosen, and the constraint is a minimum norm solution

$$J(\mathbf{m}) = (\mathbf{d} - \mathbf{Gm})^T(\mathbf{d} - \mathbf{Gm}) + \mu\mathbf{m}^T\mathbf{m}\,. \qquad (3.1.5)$$

The parameter μ defines the amount of weight or priority given to satisfying the minimum norm constraint versus the data misfit term. This parameter is appropriately named the weighting or trade-off parameter. The first term of the cost function is known as the data misfit term, while the second term is called the regularization term. Neither term can be perfectly satisfied, so the best solution is a compromise between the two.

The optimum weighting parameter is different for each inverse problem. However, both the data misfit and constraint term must be satisfied to a certain degree to retrieve a realistic solution. If the weighting parameter is set to zero, the solution will fit the observations completely: both the data and the noise. The minimum norm constraint will be ignored completely.

Conversely, as the weighting parameter approaches infinity, all efforts will be put towards minimizing the model norm. This will create a null solution that does not fit the data. A larger weighting parameter also creates a smoother solution, as more effort is made to cluster the values near zero.

To find the minimum of the cost function, the derivative of the function with respect to the model parameters is set to zero

$$\nabla J(\mathbf{m}) = \frac{\partial}{\partial\mathbf{m}}[\mathbf{d}^T\mathbf{d} - \mathbf{d}^T\mathbf{Gm} - \mathbf{m}^T\mathbf{G}^T\mathbf{d} + \mathbf{m}^T\mathbf{G}^T\mathbf{Gm} + \mu\mathbf{m}^T\mathbf{m}]\,, \qquad (3.1.6)$$

so that

$$\nabla J(\mathbf{m}) = 2\mathbf{G}^T\mathbf{Gm} - 2\mathbf{G}^T\mathbf{d} + \mu 2\mathbf{m} = 0\,. \qquad (3.1.7)$$

For details on the derivative of a cost function with respect to a vector see Appendix D. This reduces to the damped least squares solution

$$\widetilde{\mathbf{m}} = (\mathbf{G}^T\mathbf{G} + \mu I)^{-1}\mathbf{G}^T\mathbf{d}\,. \tag{3.1.8}$$

The matrix $\mathbf{G}^T\mathbf{G}$ is in general not invertible for ill-posed problems (Strang, 1986). Adding the diagonal term contributed by the weighting parameter prevents the instability associated with inverting zero terms. This addition is known as damping or pre-whitening, hence the damped least squares (DLS) solution (Claerbout, 1992).

3.2 Quadratic constraints

Constraints on the solution are expressed as regularization terms in the cost function. The constraints discussed here result in a quadratic cost function, and therefore, a linear inverse problem. In geophysical inverse problems, the most common type of constraint is one that enforces smoothness. This is because the properties in the earth's layers often change along a continuous gradient in depth. As well, a solution with minimum variation is the simplest one: overly complicated structures that have a lesser chance of being realistic are avoided (Constable et al., 1987).

The minimum norm regularization term forces the solution values to cluster around zero. In this way, the damping term acts as a smoothing operator. Other forms of constraints to find simple Earth models are implemented by minimizing variation between adjacent cells in a tomographic grid. These constraints can be implemented as weighting functions within the DLS solution.

3.2.1 Weighting functions

When there exists prior knowledge about the solution or experiment, it may be desirable to give more credence to certain observations than others. Commonly,

data with a great amount of error or noise are given less weight in the solution. This is expressed within a weighting matrix in the cost function. The new cost function will be

$$J(\mathbf{m}) = (\mathbf{W_d}\,[\mathbf{d} - \mathbf{Gm}]\,)^T\,(\mathbf{W_d}\,[\mathbf{d} - \mathbf{Gm}]\,) + \mu(\mathbf{W_m}\mathbf{m})^T(\mathbf{W_m}\mathbf{m})\,, \qquad (3.2.1)$$

where $\mathbf{W_d}$ and $\mathbf{W_m}$ are the weighting matrices on the data misfit and model parameters respectively. The resulting solution is

$$\widetilde{\mathbf{m}} = (\mathbf{G}^T\mathbf{W_d}^T\mathbf{W_d}\mathbf{G} + \mu\mathbf{W_m}^T\mathbf{W_m})^{-1}\mathbf{G}^T\mathbf{W_d}^T\mathbf{W_d}\,\mathbf{d}\,. \qquad (3.2.2)$$

Within the previous equation, the combined data weighting matrices are equivalent to the inverse of the covariance matrix of the data, $\mathbf{W_d}^T\mathbf{W_d} = \mathbf{C}_d^{-1}$ (Menke, 1984). The diagonal of the covariance matrix corresponds to the variance of the recorded data. Therefore, values with large variance or limited accuracy receive little weight, and those with small variance or greater accuracy are heavily weighted. The model weighting terms are related in the same way to the covariance of the prior model parameters.

Flat and smooth constraints using weighting matrices

Weighting matrices can be used to enforce flat or smooth solutions. Minimizing the first derivatives of the model norm will create a solution that has little change between adjacent parameters: a flat solution. In the same vein, minimizing the second derivatives of the model norm will create a solution with minimum abrupt changes between sets of parameters. This results in a smooth solution. Both of these constraints limit the variation of the solution, and construct simple Earth models.

The derivatives can be written as weighting matrices that act on the vector of model parameters. These are constructed using a finite difference approximations to first and second order derivatives. For example

$$\mathbf{D}_1 = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and

$$\mathbf{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Within the solution to the cost function, the model weighting matrix of equation 3.2.2 is replaced by the derivative matrix. A flat solution will be retrieved by

$$\widetilde{\mathbf{m}} = (\mathbf{G}^T\mathbf{G} + \mu\mathbf{D}_1^T\mathbf{D}_1)^{-1}\mathbf{G}^T\mathbf{d}, \tag{3.2.3}$$

and a smooth solution by

$$\widetilde{\mathbf{m}} = (\mathbf{G}^T\mathbf{G} + \mu\mathbf{D}_2^T\mathbf{D}_2)^{-1}\mathbf{G}^T\mathbf{d}. \tag{3.2.4}$$

Smoothing can be enforced in both the horizontal and vertical directions in two dimensional problems. In these cases, the matrix of model parameters will be

placed into a vector using lexicographic notation. The derivative matrices for such cases will be determined by which model parameters are adjacent in the vector $\mathbf{m}$.

3.3 Solving the discrete inverse problem

3.3.1 The direct algorithm

A direct approach to solving the inverse problem involves two steps. First, the analytic solution is found by minimizing the cost function. This gives an expression for the estimated model parameters, such as the DLS solution

$$\widetilde{\mathbf{m}} = (\mathbf{G}^T\mathbf{G} + \mu I)^{-1}\mathbf{G}^T\mathbf{d}. \tag{3.3.1}$$

The solution is found numerically by directly recovering the inverse of the matrix $(\mathbf{G}^T\mathbf{G} + \mu\mathbf{I})$. The inverse can be found using any number of computational schemes, such as Gaussian elimination. See Strang (1986) and Press et al. (1988) for a further discussion of numerical techniques. It is not efficient to solve an inverse problem directly when the matrix to be inverted is large or sparse. In both cases, less time consuming methods exist.

3.3.2 Minimization of the cost function

In situations where the direct solution is not the most effective approach, one can use the cost function itself to find a solution. The cost function associated with the damped least squares solution,

$$J(\mathbf{m}) = \mathbf{m}^T\mathbf{G}^T\mathbf{G}\mathbf{m} - \mathbf{d}^T\mathbf{G}\mathbf{m} - \mathbf{m}^T\mathbf{G}^T\mathbf{d} + \mathbf{d}^T\mathbf{d} + \mu\mathbf{m}^T\mathbf{m}, \tag{3.3.2}$$

is an example of a quadratic equation. Rather than finding the solution directly, one can minimize the cost function using a gradient algorithm. These methods are known as gradient methods because they rely on the gradients of the cost function to guide them toward the minimum. The simplest of these approaches is the steepest descent algorithm.

Steepest Descent

An example of a quadratic function is the cost function 3.3.2. When the matrix $\mathbf{G}$ is positive definite, this function creates a paraboloid (Figure 3.1). The steepest descent algorithm begins at a point $\mathbf{m}_0$, and moves in the direction opposite the largest gradient, effectively sliding down to the minimum of the function as illustrated in Figures 3.2 and 3.3. The drawback to this simple method is that it is only able to change directions at finite steps. In essence, it is unable to turn to follow the curves and slopes of the function. Strang (1986) describes the algorithm "like a skier who can't turn; he goes forward until his path begins to climb". In certain situations, like when in a narrow valley, this causes the algorithm to take many repetitive steps to correct its path. Because of this disadvantage, the more sophisticated conjugate gradient method is preferred.

Conjugate gradients

The conjugate gradient (CG) algorithm works in much the same way as the steepest descent method. The difference is in the direction of descent: to prevent tracking back and forth, this algorithm will proceed in a direction orthogonal to all previous steps. In this way, the CG scheme arrives at the minimum more quickly. Scales (1987) describes an algorithm to solve least squares problems, adapting the classic CG algorithm of Hestenes and Stiefel (1952). This method is known as conjugate gradients for least squares (CGLS), and its structure is
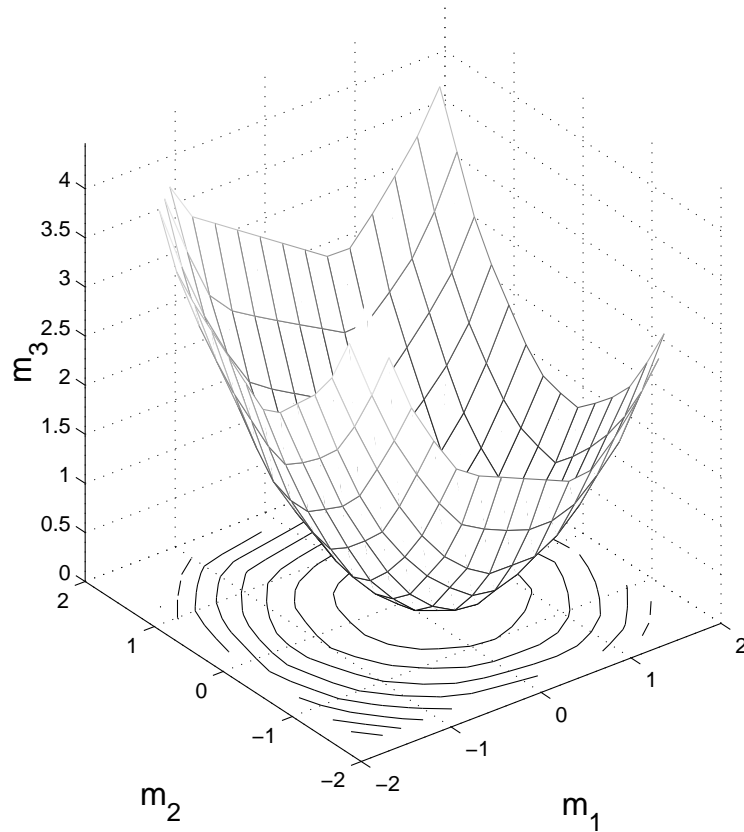
Figure 3.1: The cost function in equation 3.3.2 containing a positive definite matrix $\mathbf{G}$ will be a paraboloid. Contours drawn beneath the function indicate constant $J(\mathbf{m})$.
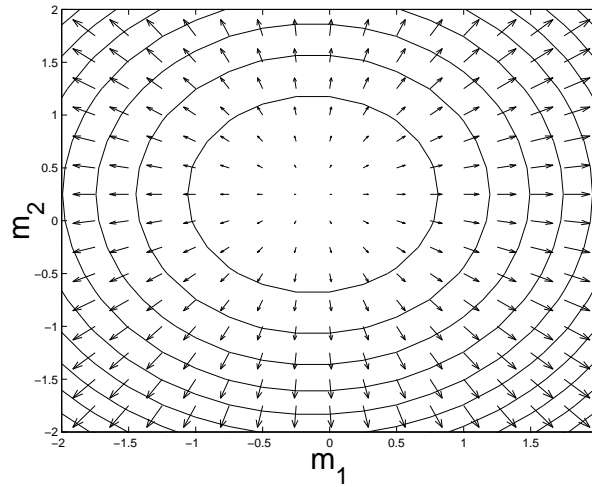
Figure 3.2: The gradient of the quadratic function indicated by the direction and magnitude of quivers superimposed on the contour plot. Steepest descent moves in the direction opposite to the largest gradients to find the minimum point where the gradient is zero.
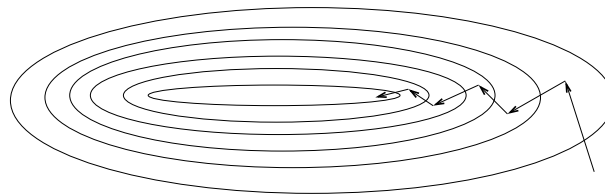
Figure 3.3: The directions of steepest descent as it enters a narrow valley.

shown below.

CGLS Subroutine

```
Initilization
```

$\mathbf{s}_0 = \mathbf{d} - \mathbf{G}\mathbf{m}_0$

$\mathbf{r}_0 = \mathbf{p}_0 = \mathbf{G}^T(\mathbf{d} - \mathbf{G}\mathbf{m}_0)$

where r_0 is the residual and p_0 the gradient in model space

$\mathbf{q}_0 = \mathbf{G}\mathbf{p}_0$

```
for k = 0,1, ...
```

$\texttt{alpha}_{k+1} = (\mathbf{r}_k^T\mathbf{r}_k)/(\mathbf{q}_k^T\mathbf{q}_k)$, # *to determine step size*

$\mathbf{m}_{k+1} = \mathbf{m}_k + \texttt{alpha}_{k+1}\mathbf{p}_k$,

arrive at new point m_{k+1} in model space by moving along gradient direction

$\mathbf{s}_{k+1} = \mathbf{s}_k - \texttt{alpha}_{k+1}\mathbf{q}_k$,

$\mathbf{r}_{k+1} = \mathbf{G}^T\mathbf{s}_{k+1}$, # *update residual*

$\texttt{beta}_{k+1} = (\mathbf{r}_{k+1}^T\mathbf{r}_{k+1})/(\mathbf{r}_k^T\mathbf{r}_k)$,

$\mathbf{p}_{k+1} = \mathbf{r}_{k+1} + \texttt{beta}_{k+1}\mathbf{p}_k$,

update direction p_{k+1} orthogonal to previous directions

$\mathbf{q}_{k+1} = \mathbf{G}\ \mathbf{p}_{k+1}$.

```
end
```

Augmented matrices

The CGLS algorithm contains only one matrix operator, $\mathbf{G}$. To obtain the DLS solution a system of augmented matrices is used. The inverse problem is written in terms of the augmented operator matrix and data vector

$$\mathbf{G}_A\,\mathbf{m} = \mathbf{d}_A\,. \tag{3.3.3}$$

The augmented operator is defined as

$$\mathbf{G}_A = \begin{pmatrix} \mathbf{G} \\ \sqrt{\mu}\,\mathbf{I} \end{pmatrix} , \tag{3.3.4}$$

where $\mathbf{G}$ is size $N \times M$, μ is a scalar, and $\mathbf{I}$ is an identity matrix size $M \times M$. The augmented data vector is

$$\mathbf{d}_A = \begin{pmatrix} \mathbf{d} \\ \mathbf{z} \end{pmatrix} , \tag{3.3.5}$$

where $\mathbf{d}$ is a vector length $N \times 1$, and $\mathbf{z}$ is a vector of zeros, length $M \times 1$.

The CGLS algorithm will recover the least solution

$$\begin{aligned} \widetilde{\mathbf{m}} &= (\mathbf{G}_A^T \mathbf{G}_A)^{-1} \mathbf{G}_A^T \mathbf{d}_A \\ &= \left[\begin{pmatrix} \mathbf{G}^T & \sqrt{\mu}\,\mathbf{I}^T \end{pmatrix} \begin{pmatrix} \mathbf{G} \\ \sqrt{\mu}\,\mathbf{I} \end{pmatrix} \right]^{-1} \begin{pmatrix} \mathbf{G}^T & \sqrt{\mu}\,\mathbf{I}^T \end{pmatrix} \begin{pmatrix} \mathbf{d} \\ \mathbf{z} \end{pmatrix} \\ &= (\mathbf{G}^T \mathbf{G} + \mu \mathbf{I})^{-1} \mathbf{G}^T \mathbf{d} . \end{aligned} \tag{3.3.6}$$

Following the same procedure, the augmented matrix of a flat solution (3.2.3) is

$$\mathbf{G}_A = \begin{pmatrix} \mathbf{G} \\ \sqrt{\mu}\,\mathbf{D}_1 \end{pmatrix} , \tag{3.3.7}$$

where $\mathbf{G}$ is size $N \times M$, μ is a scalar, and $\mathbf{D}_1$ is the first derivative matrix, size $M \times M$. The augmented data vector remains (3.3.5). The augmented matrix of a smooth solution (3.2.4) simply replaces $\mathbf{D}_1$ with $\mathbf{D}_2$.

3.3.3 Computation time

The CGLS algorithm of the previous section recovers the solution in an indirect way. Rather than performing a direct matrix inversion, the solution is found by iteratively performing forward and adjoint modeling. Computation time may be minimized by exploiting this characteristic of the indirect algorithm. Subroutines that perform the forward and adjoint modeling actions, $\mathbf{Gm}$ and $\mathbf{G}^T\mathbf{d}$, can replace matrices. In this scheme, the matrix operator is not stored directly, so memory is saved (Scales, 1987). Additionally, when the matrix $\mathbf{G}$ is sparse, the operator subroutines will reach the solution more quickly than matrix multiplication, which performs unnecessary computations.

Lastly, because the CG algorithm moves successively toward the solution, it can be stopped at any time. The interim solution is often acceptable, so computing time may be minimized in this way (Scales and Smith, 1994).

3.3.4 Determining weighting parameters

If there is enough knowledge about the data, a simple numerical test can be done to determine weighting parameters. To perform the χ^2 test, the noise must be Gaussian, have a zero mean, and a known variance of σ_n (Tarantola, 1987). The noise estimate for a specific value of the trade-off parameter μ will be

$$\mathbf{n}(\mu) = \mathbf{G}\widetilde{\mathbf{m}}(\mu) - \mathbf{d}\,. \qquad (3.3.8)$$

The χ^2 value is defined as

$$\chi^2(\mu) = \frac{\mathbf{n}(\mu)^T\mathbf{n}(\mu)}{\sigma_n^2}\,. \qquad (3.3.9)$$

For Gaussian distributed noise, the expected χ^2 value lies in a range determined by the number of observations, N. Any value $\chi^2(\mu) \sim N \pm \sqrt{N}$ will be a

good choice. Thus, using this information, an appropriate weighting parameter may be chosen.

3.4   The behavior of quadratic constraints

In Chapter 2, the numerical forward and adjoint modeling operators for the linear scattering problem were described. In this section, these operators are placed into the framework of discrete inverse theory. This creates the hybrid iterative inversion method referred to as migration/inversion (Jin et al., 1992; Thierry et al., 1999). The integral and the convolutional operators described in Chapter 2 are combined into the Born operator, $\mathbf{B} = \mathbf{CL}$, to create the inverse problem

$$J(\mathbf{m}) = (\mathbf{d} - \mathbf{Bm})^T(\mathbf{d} - \mathbf{Bm}) + \mu R(\mathbf{m}), \qquad (3.4.1)$$

where $R(\mathbf{m})$ is a regularization term. Three methods of achieving solutions with minimum variation are compared: the minimum norm, flat, and smooth constraints. All of the examples are performed on acoustic, constant density media. The model parameters consist of acoustic velocity perturbations, and the data of seismic traveltime and amplitude information.

In the first example, a 1D case is created by allowing the model to vary only with depth, and by placing the source and receiver at the same horizontal position at the earth's surface (a zero-offset experiment). Because the problem is a simple one, the modeling operators are put in matrix form. A direct solution is found by inverting the matrices using a subroutine provided by Matlab.

Figure 3.4 shows the velocity model and source wavelet. The background velocity model is set to be a homogeneous 2000 m/s, from which the acoustic perturbation is calculated, and used to create the synthetic data. One percent Gaussian noise is added to the data. Three different weighting parameters are

compared for the damped least squares solution. It is apparent that when the weighting parameter is too large, most of the effort is directed towards minimizing the model norm. This results in a solution with a small magnitude that does not fit the data well. Conversely, as the trade-off parameter becomes too small, all of the emphasis is placed on fitting the observed data, including the noise. This creates artifacts in the solution. The best solution determined by the χ^2 test is the one corresponding a misfit value of 250, or approximately to $\mu = 0.9 \times 10^{-6}$.

The behavior of the solution response to the weighting parameter, μ, is further explored in Figure 3.5. The χ^2 test demonstrates that the data misfit and chi-squared values will increase with the trade-off parameter. This follows from the simple definition that a larger weighting parameter results in less emphasis on fitting the data, and more on minimizing the model norm. It then makes sense that the model norm is inversely proportional to the trade-off parameter. The final curve in Figure 3.5 is the trade-off curve: illustrating that one can have either a low value of misfit or model norm, but not both.

Two dimensional examples are larger, and so direct matrix inversion is no longer feasible. Therefore, the 2D examples are created using a CG algorithm with forward and adjoint modeling subroutines. The source-receiver geometry of the next two examples is seen in Figure 3.6.

The first inversion is performed on a simple perturbation of one pixel (Figure 3.7). Three percent Gaussian noise is added to the synthetic data, and solutions for the three constraints are recovered. Figure 3.8 compares the cross sections of the solutions. This figure illustrates that the DLS solution puts the most emphasis on fitting a solution near to zero, and has the smallest magnitude. The D_2 smoothing creates a smoothly varying solution opposed to the flat one retrieved by the D_1 constraint. This is because it minimizes variation over a larger number of adjacent velocities.
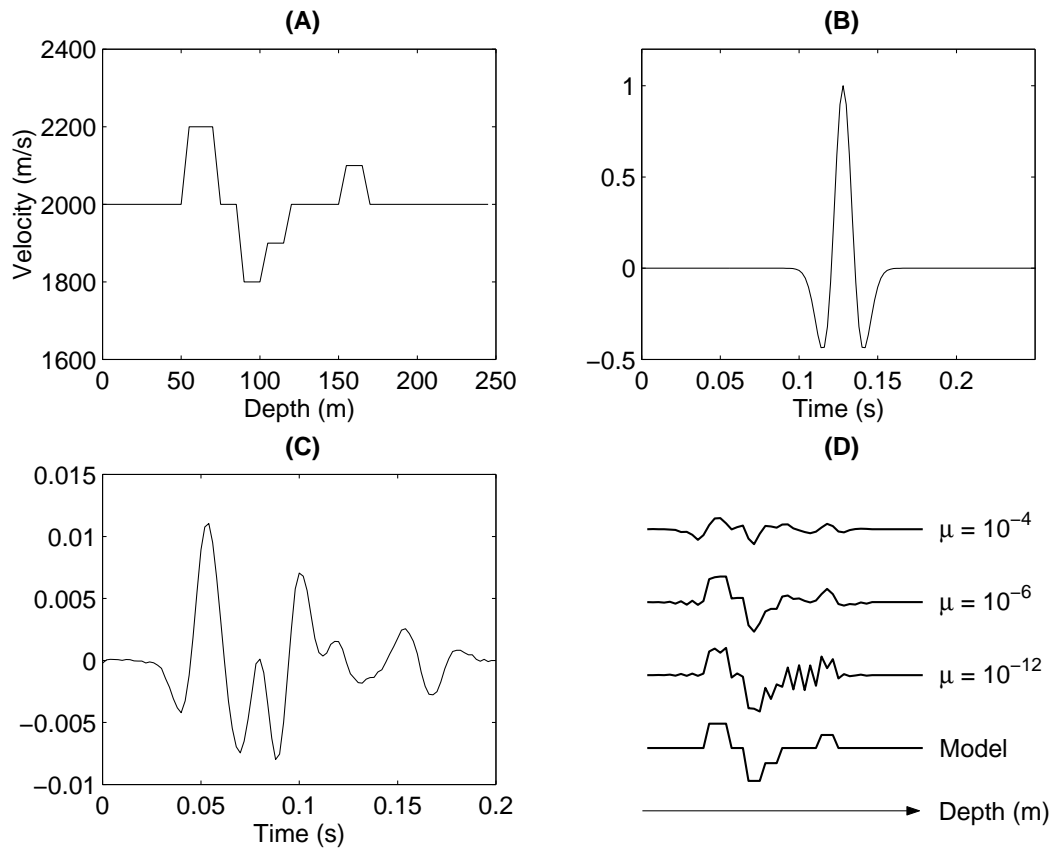
Figure 3.4: The velocity model (A), and source wavelet (B) are used to generate seismic data (C) with the forward operator. One percent Gaussian noise is added to the synthetic data. The damped least squares solutions for three different values of the weighting parameter are compared in (D). Both the source and receiver are located at the Earth's surface, where depth = 0m.
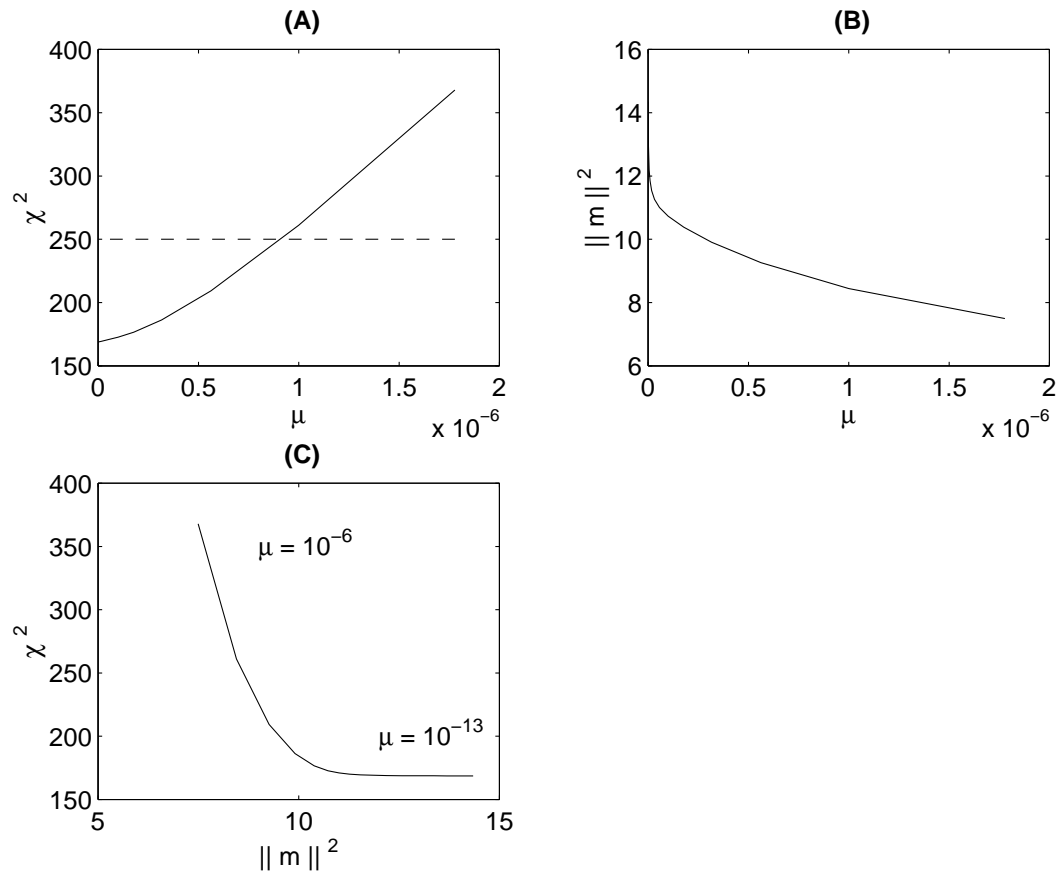
Figure 3.5: The χ^2 test, where the dashed line indicates the desired misfit value found by the number of observations (A). Also shown are the model norm as a function of the weighting parameter, μ, (B), and the trade-off curve for the misfit and model norm (C).
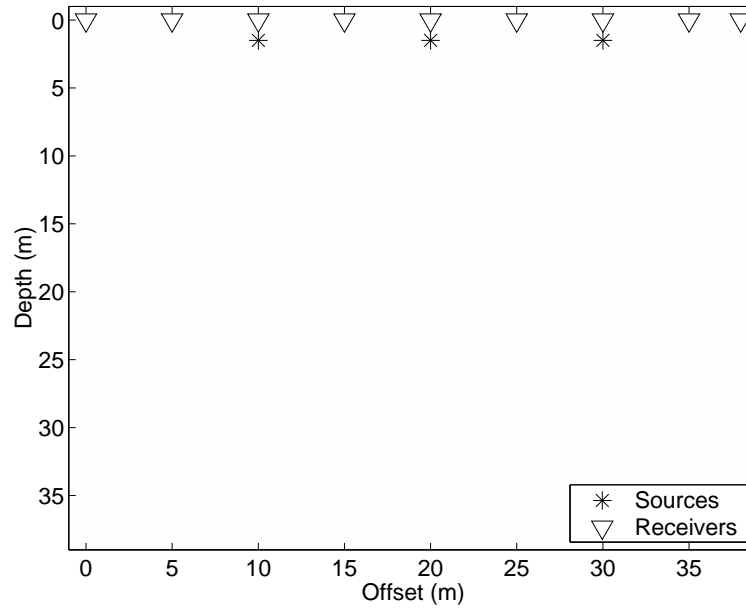
Figure 3.6: The source-receiver geometry of the 2D examples.

The inverse problem is solved using the conjugate gradient algorithm to find the minimum of the cost function. Theoretically, the algorithm will recover the exact solution in n iterations, where n is the largest dimension of the matrix operator. However, due to round-off error, this does not occur. It is standard practise to pick a stopping criterion that analyzes the changes in the solutions as the algorithm proceeds (Scales and Smith, 1994). Often, the solution obtained before n iterations is very close to the final one. In this way, an acceptable solution may be obtained more rapidly using a less stringent stopping criterion. The minimization of the cost function for the DLS solution is shown in Figure 3.9. After the algorithm has reached 100 iterations, the solution changes very little. A comparison of four different solutions retrieved as the algorithm converges (Figure 3.10) shows that as little as 50 iterations could produce an acceptable solution.

Figure 3.7: A 2D inversion using the iterative approach of conjugate gradients. Shown here are the acoustic perturbation (A), the source wavelet (B), synthetic data (C), the solutions to the minimum norm (D), flat (E), and smooth (F) constraints. The weighting parameters are $\mu = 10$ (D), $\mu_x = \mu_z = 2$ (E), and $\mu_x = \mu_z = 2$ (F).

Figure 3.8: A comparison of the cross sections at 20 m depth of the three solutions in Figure 3.7.

The second inversion is done for a 2D step function illustrated in Figure 3.11. Again, the inversion is performed for minimum norm, flat and smooth constraints. A comparison of the cross sections at 30 m depth illustrates the difference in the recovered images (Figure 3.12). Similar to the last example, the DLS solution has the smallest magnitude, as it seeks to be near zero. The emphasis on keeping continuity between adjacent pixels is seen in flat constraint, while the smooth constraint allows more variation.

3.5 Summary

Within this chapter, an overview of discrete inverse theory is given. This technique relies on forward and adjoint modeling to approximate numerically the inverse. The numerical modeling operators have been derived from the analytic
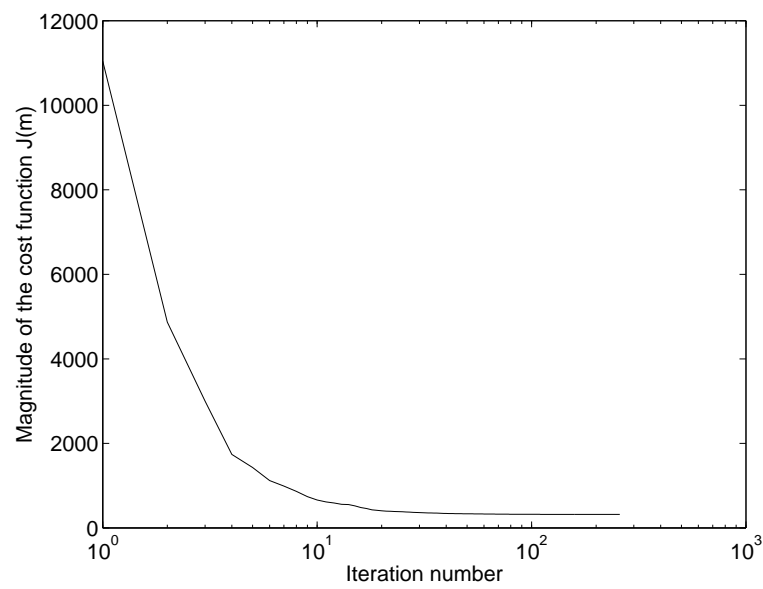
Figure 3.9: The magnitude of the cost function plotted against the number of iterations performed by the CG algorithm. The algorithm is recovering the DLS solution for the problem shown in Figure 3.7. The solution converges at 250 iterations.
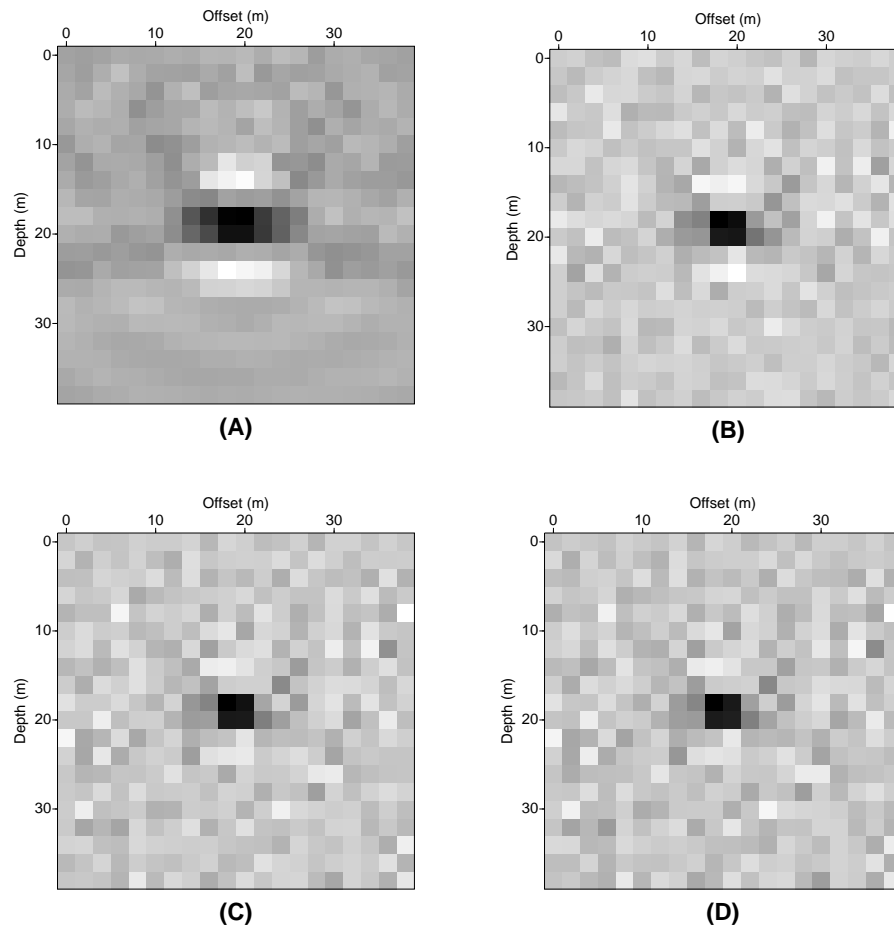
Figure 3.10: The solution recovered by the conjugate gradient algorithm at: (A) 10, (B) 50, (C) 100, and (D) 250 iterations.
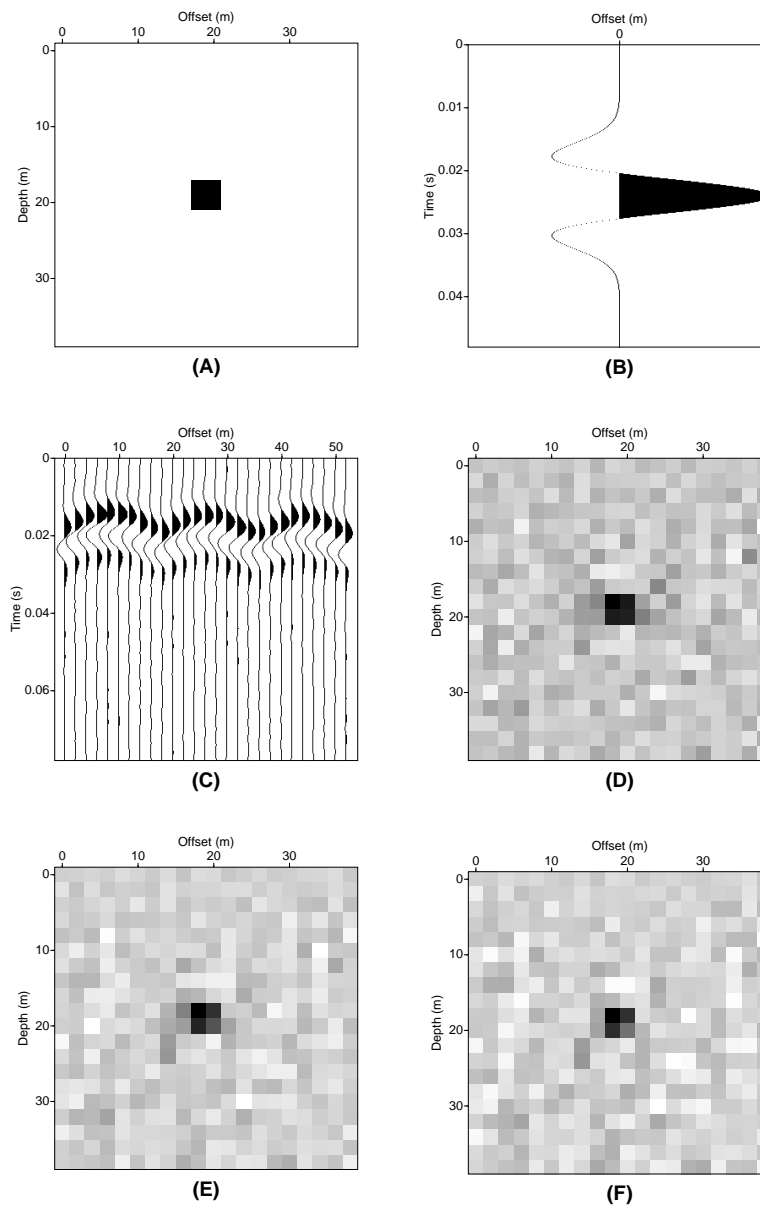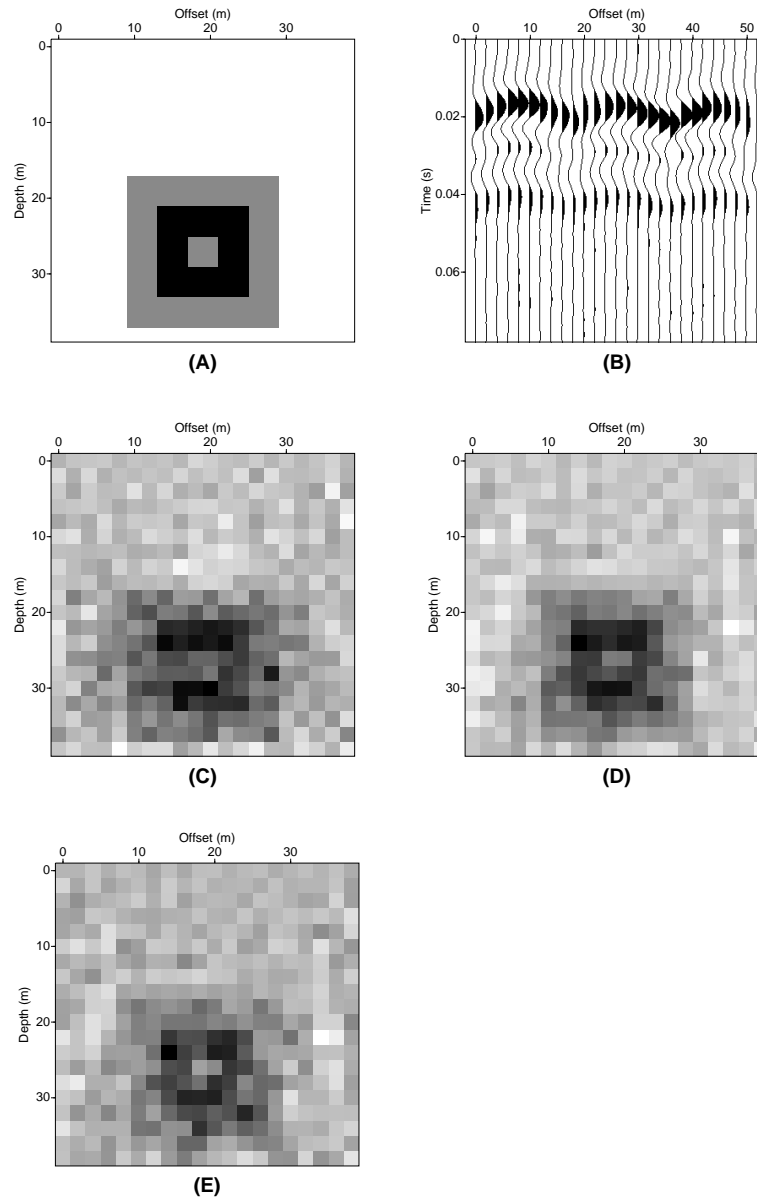
Figure 3.11: The second 2D inversion using the iterative approach of conjugate gradients. Shown here are: the acoustic perturbation (A), synthetic data (B), and solutions to the minimum norm (C), flat (D), and smooth (E) constraints. The weighting parameters are $\mu = 20$ (D), $\mu_x = \mu_z = 15$ (E), and $\mu_x = \mu_z = 15$ (F).
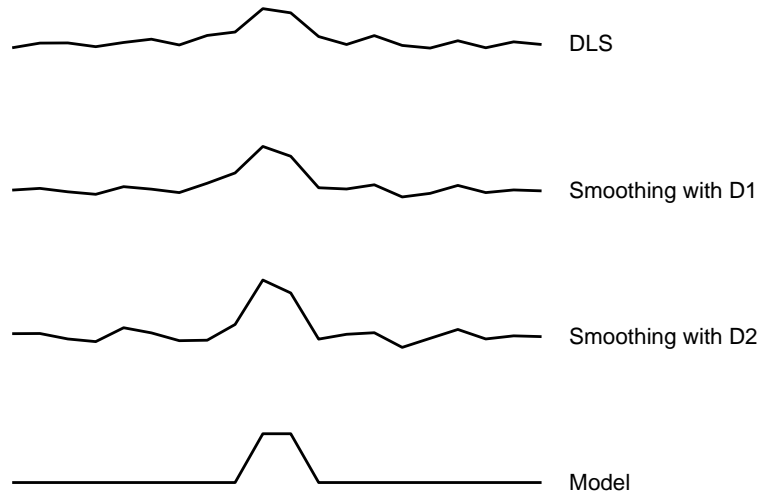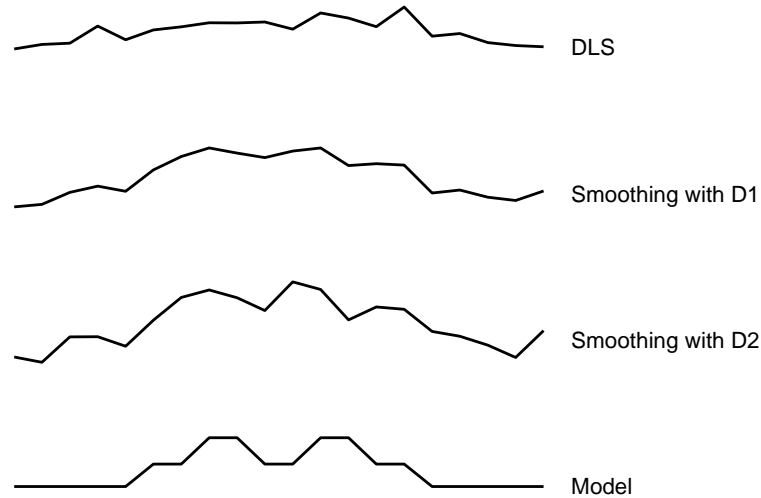
Figure 3.12: A comparison of the cross sections at 30 m depth of the three solutions in Figure 3.11.

migration technique using the GRT. This results in a hybrid migration/inversion algorithm that combines the advantages of direct and discrete inversion.

It is explained that the majority of geophysical inverse problems fall into the category of ill-posed problems. This type of inverse problem requires a constraint to retrieve a stable, unique solution. The regularization terms of common constraints have been introduced, and compared in the examples. I show how the DLS solution is encouraged to cluster around zero, and is comparable to a smoothing function. Minimizing the first and second order derivatives of the model norm both minimize variation in the solution, but have different effects. The former creates a flat solution, while the latter, a smoothly varying one.

Finally, the numerical algorithms to solve the discrete inverse problem are discussed. It is pointed out that direct numerical techniques are only useful in small problems. For larger problems, an iterative algorithm that searches

the cost function for its minimum is more effective. Iterative schemes have the added advantage of being able to implement the forward and adjoint modeling operators as subroutines, saving memory. As well, the number of iterations may be truncated, while still retrieving a close approximate to the exact solution.

Chapter 4

Sparse inversion techniques

Solutions of limited variation have the advantage of avoiding irregular or large deviations in the recovered Earth model (Constable et al., 1987). These do not often make sense, and are not encouraged. However, there are situations where more variation is desired, and a smooth solution is not appropriate. This occurs in the geophysical problem of inverting for reflectivity information. Like smoothness, sparseness in the solution is enforced through the regularization terms of the cost function.

Once a reflectivity series is known, the relative impedance profile can be recovered by integrating the reflection coefficients (Waters, 1978; Walker and Ulrych, 1983). When sparseness is enforced on the first derivatives of a model parameter, the resulting solution will be blocky, or piecewise continuous. This is evident in the blocky impedance profile.

Within this thesis, the non-linear sparse regularization term introduced by Charbonnier et al. (1997) for medical imaging will be tested for geophysical applications. This constraint has the advantage of being especially suited for blocky inversion problems. It is also easy to linearize.

In this chapter, sparse constraints are introduced as prior probability density

functions in the Bayesian framework. The constraint of Charbonnier et al. (1997) is referred to as a modified Cauchy prior, due to its similarity to a Cauchy probability distribution. The exponential and Cauchy distributions, common in sparse inversion techniques, are also introduced.

All three sparse constraints result in non-linear inverse problems. Because they are weakly non-linear, the technique of iteratively re-weighted least squares (IRLS) can be applied to solve them in a linear, iterative manner (Scales and Smith, 1994). The effectiveness of these constraints is compared in the 1D impedance inversion problem. This is an ideal problem to test with, because both the recovered sparse reflectivity and blocky impedance profiles reveal properties of the solution.

4.1   Bayes Theorem

Bayes theorem combines prior knowledge of the model and data, expressed as probability distributions, in order to make judgements about the recovered model (Scales and Smith, 1994). Bayes theorem states that

$$P(\mathbf{m}, \mathbf{d}) = \frac{P(\mathbf{m})\, P(\mathbf{d}, \mathbf{m})}{P(\mathbf{d})}\,. \tag{4.1.1}$$

$P(\mathbf{m})$ represents the probability distribution of the model parameters, $\mathbf{m}$, known prior to collecting data. It is referred to as the prior distribution of $\mathbf{m}$. The likelihood function, $P(\mathbf{d}, \mathbf{m})$, describes how knowledge of the data modifies the prior knowledge. The distribution of the data, $P(\mathbf{d})$, plays the role of a scaling function. Finally, the posterior distribution term, $P(\mathbf{m}, \mathbf{d})$, indicates what is known about the model parameters given the information provided by the observed data (Box and Tiao, 1992). In other words, it is the probability that the model is correct given a certain set of observations (Ulrych et al., 2001).

Bayes theorem is commonly expressed without the normalization factor as

$$P(\mathbf{m}, \mathbf{d}) \propto P(\mathbf{m}) \, P(\mathbf{d}, \mathbf{m}) \,. \tag{4.1.2}$$

A more thorough exploration of Bayesian methods is discussed by Box and Tiao (1992), or Ulrych et al. (2001).

4.1.1 Derivation of the DLS solution using Bayes theorem

In the case of noisy data,

$$\mathbf{Gm} - \mathbf{d} = \mathbf{n} \,, \tag{4.1.3}$$

where $\mathbf{n}$ is the noise vector. The likelihood function is assigned to be the distribution of the noise

$$P(\mathbf{d}, \mathbf{m}) = P(\mathbf{n}) \,. \tag{4.1.4}$$

It is assumed that the noise has a Gaussian distribution, a mean value $\overline{n}$ and a variance σ^2. The probability function in this case is

$$P(n_i) = \frac{1}{\sqrt{2\pi\sigma^2}} \, exp \left(\frac{-1}{2\sigma^2} \, (n_i - \overline{n})^2 \right). \tag{4.1.5}$$

Assuming the mean value of the noise, $\overline{n}$, is zero, this term can be neglected. If the noise is uncorrelated and the variance of each noise sample remains constant, the total probability of the noise vector is

$$
\begin{aligned}
P(\mathbf{n}) &= P(n_1)\,P(n_2)\,P(n_3)\,\ldots P(n_N) \\
&= \frac{1}{(2\pi\sigma^2)^{N/2}}\,exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{N}n_i^2\right) \\
&= K_1\,exp\left(\frac{-1}{2\sigma^2}\,\mathbf{n}^T\mathbf{n}\right).
\end{aligned}
\tag{4.1.6}
$$

The model parameters are represented by the vector composed of elements $m_i,\ i = 1, 2, \ldots M$. These parameters are also assumed to have Gaussian distribution, be uncorrelated, and have a variance σ_m^2 and a mean of zero. The prior probability is then

$$
\begin{aligned}
P(\mathbf{m}) &= \frac{1}{(2\pi\sigma_m^2)^{M/2}}\,exp\left(\frac{-1}{2\sigma_m^2}\,\mathbf{m}^T\mathbf{m}\right) \\
&= K_2\,exp\left(\frac{-1}{2\sigma_m^2}\,\mathbf{m}^T\mathbf{m}\right).
\end{aligned}
\tag{4.1.7}
$$

The likelihood probability function has been set to the probability of the noise. Then, by substituting the likelihood (4.1.6) and prior (4.1.7) distributions into Bayes rule (4.1.2), the posteriori distribution will become

$$
P(\mathbf{m},\mathbf{d}) \propto K_1\,K_2\,exp\left(\frac{-1}{2\sigma_m^2}\,\mathbf{m}^T\mathbf{m}\right)exp\left(\frac{-1}{2\sigma^2}\,\mathbf{n}^T\mathbf{n}\right).
\tag{4.1.8}
$$

This can be rewritten, using the relation $\mathbf{n} = \mathbf{Gm} - \mathbf{d}$

$$
P(\mathbf{m},\mathbf{d}) \propto K_1\,K_2\,exp\left(\frac{-1}{2\sigma_m^2}\,\mathbf{m}^T\mathbf{m}\right)exp\left(\frac{-1}{2\sigma^2}\,(\mathbf{Gm}-\mathbf{d})^T(\mathbf{Gm}-\mathbf{d})\right).
$$
$$
\tag{4.1.9}
$$

One way to choose a solution is to find a set of model parameters such that the posterior probability, $P(\mathbf{m},\mathbf{d})$, is maximized. The solution is known as the

maximum *a posteriori* estimator or MAP. Maximizing the posterior function is the same as finding the minimum of

$$-lnP(\mathbf{m}, \mathbf{d}) \propto -lnK_1\,K_2 + \frac{1}{2\sigma_m^2}\,\mathbf{m}^T\mathbf{m} + \frac{1}{2\sigma^2}\,(\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d})\,. \quad (4.1.10)$$

The term $-lnK_1\,K_2$ is a constant, and will not change the solution that minimizes $-lnP(\mathbf{m}, \mathbf{d})$. Therefore, solution is found by minimizing

$$(\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d}) + \frac{\sigma^2}{\sigma_m^2}\,\mathbf{m}^T\mathbf{m}\,. \quad (4.1.11)$$

This takes the form of the following cost function

$$J(\mathbf{m}) = (\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d}) + \mu\,\mathbf{m}^T\mathbf{m}\,. \quad (4.1.12)$$

where the weighting parameter is defined by the variance of the data and model parameters

$$\mu = \frac{\sigma^2}{\sigma_m^2}\,. \quad (4.1.13)$$

The Bayesian MAP solution reduces to the same cost function as the minimum norm problem (3.1.5), which leads to the damped least squares solution. Thus, Bayesian inversion gives another perspective on the same problem, where minimizing an L_2 norm of the model parameters corresponds to a Gaussian prior probability distribution.

4.1.2   Gaussian prior probability functions

The constraints enforced to stabilize the inverse problem can be considered prior information. By viewing the prior probability distributions, one can more intuitively understand the constraints placed on the solution.

The probability of an L_2 norm gives a Gaussian distribution

$$P(\mathbf{m}) = \frac{1}{(2\pi\sigma_m^2)^{M/2}} \, exp \, \left(\frac{-1}{2\sigma_m^2} \sum_{i=1}^{M}(m_i - \overline{m}) \right), \qquad (4.1.14)$$

where σ_m^2 is the variance and $\overline{m}$ is the mean. Gaussian probability functions are shown in Figure 4.1. Equation (4.1.13) states that the weighting parameter in the DLS solution is inversely proportional to the variance of the model parameters. Therefore, in the Bayesian framework, a larger trade-off parameter is the same as having a prior distribution with less variance. This implies that the solution will be more strongly constrained to zero, and this is evident in the narrower probability distribution. When the mean of the model parameters is not equal to zero, this is equivalent to a minimum norm solution subject to a reference parameter. A solution near to the mean or reference parameter is enforced.

4.1.3 Probability functions to enforce sparse solutions

Thus far, only Gaussian probabilities have been considered. As one might suspect, different norms will have associated probability functions and characteristic solutions. Another type of solution is a sparse solution. It is characterized by a series that is primarily zeros, with a few non-zero spikes (Figure 4.2). When a sparse solution is required, the L_2 norm is no longer appropriate. The corresponding Gaussian probability function has a distribution tightly centered around the mean, which will constrain the solution to smoothly vary around zero. Two probability distribution which do create sparse solutions are the exponential and Cauchy distributions.

The exponential prior distribution, when used to construct a MAP solution, corresponds to the minimization of the L_1 norm of the model parameters (Tarantola, 1987). The L_1 norm is defined as the absolute value of a function
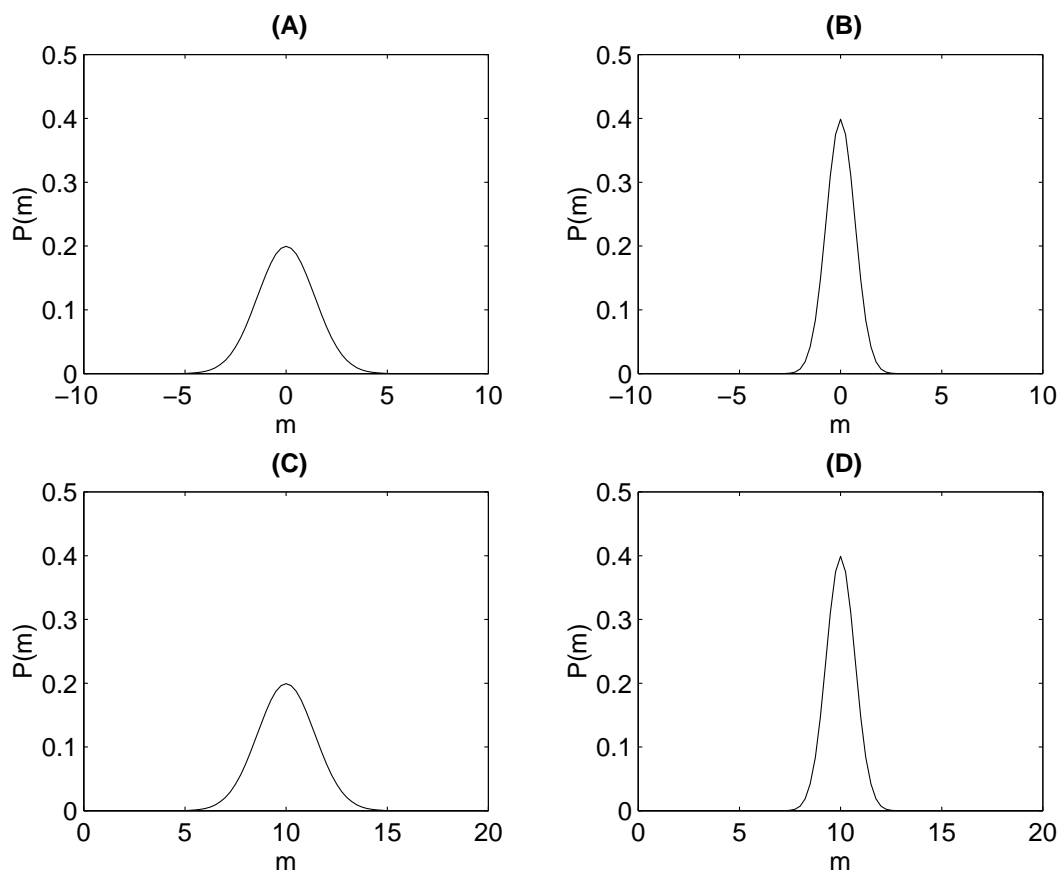
Figure 4.1: The Gaussian probability distribution of a function with variance $\sigma_m^2 = 4$ and mean $\overline{m} = 0$ (A) and $\overline{m} = 10$ (C). The distributions corresponding to $\sigma_m^2 = 1$ for the same mean values are shown in (B) and (D).
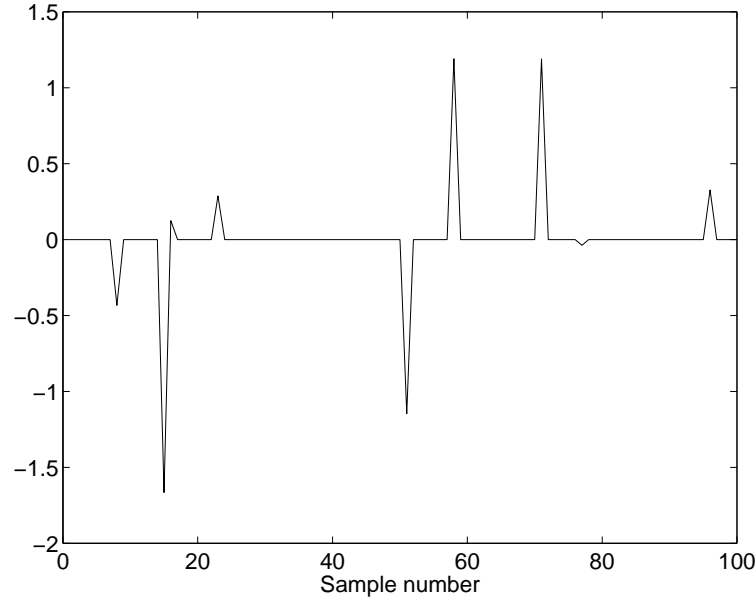
Figure 4.2: An example of a 1D sparse series.

$$L_1(\mathbf{m}) = \sum_i |m_i - \overline{m}| . \qquad (4.1.15)$$

The exponential probability distribution is defined

$$P_{exp}(\mathbf{m}) = \frac{1}{2\sigma_m} \, exp \, \left(\frac{-1}{\sigma_m} \sum_{i=1}^{M} |m_i - \overline{m}|_1 \right) . \qquad (4.1.16)$$

Another distribution used to enforce sparseness is the Cauchy function (Tarantola, 1987; Sacchi, 1997; Sacchi et al., 1998). The Cauchy probability distribution function is given by

$$P_{Cauchy}(\mathbf{m}) = \frac{1}{(\pi\sigma_m)^M} \prod_{i=1}^{M} \left[\frac{1}{1 + (m_i - \overline{m})^2/\sigma_m^2} \right] . \qquad (4.1.17)$$

It is easily seen that the exponential and Cauchy functions differ from the Gaussian one in that they have narrower distributions at their peaks, and approach
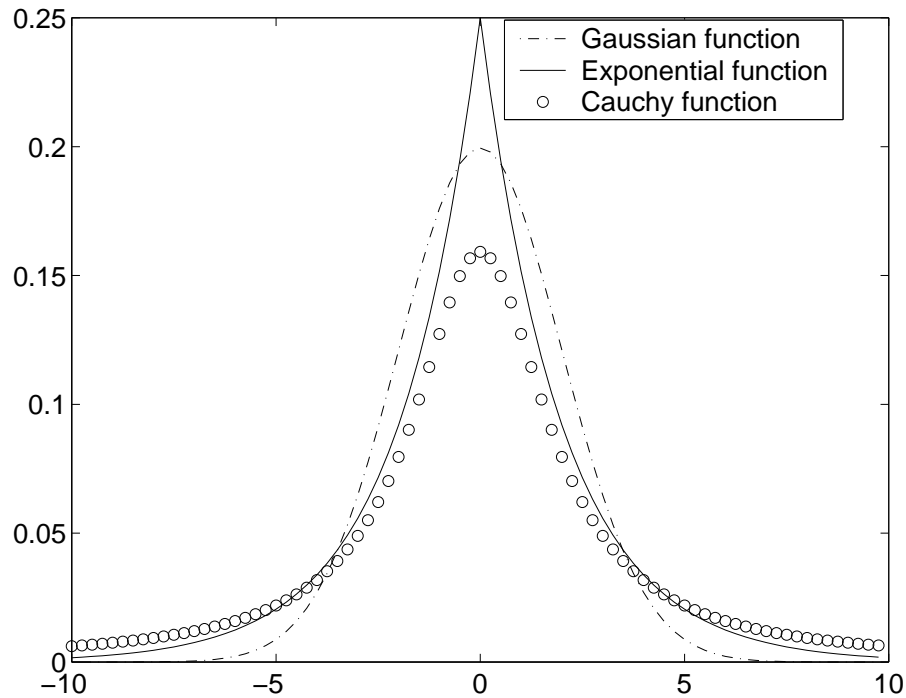
Figure 4.3: The probability distributions for Gaussian, exponential, and Cauchy functions that have a mean, $\overline{m} = 0$, and variance $\sigma^2 = 4$.

zero more slowly away from the mean (Figure 4.3). This is referred to in the literature as a probability function having a long-tailed distribution. These distributions are more tightly centered around zero, and so less non-zero values are present in the solution. Further, because of the long-tailed shape, the non-zero values have more freedom to deviate from zero than in the Gaussian distribution. This results in a solution of mostly zeros, with a few larger values interspersed: a sparse solution.

MAP solution to an exponential prior distribution

Within this section, the MAP solution to an ill-posed problem using an exponential prior is derived. As explained previously, the likelihood function $P(\mathbf{d}, \mathbf{m})$ is equal to the probability of the noise. Assuming that the noise has a Gaussian

distribution with a mean of zero, the likelihood function is

$$P(\mathbf{d}, \mathbf{m}) = P(\mathbf{n}) = K_1 \, exp \left(\frac{-1}{2\sigma^2} (\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) \right). \qquad (4.1.18)$$

The prior is set to be an exponential distribution, with a mean of zero

$$P(\mathbf{m}) = K_2 \, exp \left(\frac{-1}{\sigma_m} \sum_{i=1}^{M} |m_i| \right). \qquad (4.1.19)$$

The posteriori term combines the two, becoming

$$P(\mathbf{m}, \mathbf{d}) \propto K_1 \, K_2 \, exp \left(\frac{-1}{2\sigma^2} (\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) \right) exp \left(\frac{-1}{\sigma_m} \sum_{i=1}^{M} |m_i| \right). \tag{4.1.20}$$

Finding the maximum of the posterior function is equivalent to finding the minimum of $-lnP(\mathbf{m}, \mathbf{d})$, or

$$\frac{1}{2\sigma^2} (\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) + \frac{1}{\sigma_m^2} \sum_{i=1}^{M} |m_i|, \qquad (4.1.21)$$

which can be written as the cost function

$$J(\mathbf{m}) = (\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) + \mu \, R(\mathbf{m}), \qquad (4.1.22)$$

where the weighting parameter, μ combines the constant variance terms. The regularization term, $R(\mathbf{m})$, is defined

$$R(\mathbf{m}) = \sum_{i=1}^{M} |m_i|. \qquad (4.1.23)$$

Minimizing this cost function results in a non-linear problem. The minimum is found by taking the gradient

$$
\begin{aligned}
\nabla J(\mathbf{m}) &= \frac{\partial}{\partial \mathbf{m}} \left[(\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) + \mu\, R(\mathbf{m}) \right] \\
&= 2\mathbf{G}^T \mathbf{Gm} - 2\mathbf{G}^T \mathbf{d} + \mu\, \frac{\partial}{\partial \mathbf{m}} R(\mathbf{m})\,.
\end{aligned}
\tag{4.1.24}
$$

The derivative of the regularization term is

$$
\frac{\partial}{\partial m_k}\, R(\mathbf{m}) = sign(m_k) = \frac{m_k}{|m_k|}\,.
\tag{4.1.25}
$$

Subsequently,

$$
\nabla R(\mathbf{m}) =
\begin{pmatrix}
\frac{\partial R(\mathbf{m})}{\partial m_1} \\
\frac{\partial R(\mathbf{m})}{\partial m_2} \\
\frac{\partial R(\mathbf{m})}{\partial m_3} \\
\vdots
\end{pmatrix}
=
\begin{pmatrix}
\frac{m_1}{|m_1|} \\
\frac{m_2}{|m_2|} \\
\frac{m_3}{|m_3|} \\
\vdots
\end{pmatrix}.
\tag{4.1.26}
$$

In order to write the equation in weighted DLS form, the regularization gradient is expressed

$$
\nabla R(\mathbf{m}) =
\begin{pmatrix}
\frac{1}{|m_1|} & 0 & 0 & 0 \\
0 & \frac{1}{|m_2|} & 0 & 0 \\
0 & 0 & \frac{1}{|m_2|} & 0 \\
0 & 0 & 0 & \ddots
\end{pmatrix}
\begin{pmatrix}
m_1 \\
m_2 \\
m_3 \\
\vdots
\end{pmatrix}.
\tag{4.1.27}
$$

Thus, the gradient of the regularization term is

$$
\nabla R(\mathbf{m}) = \mathbf{Qm}\,,
\tag{4.1.28}
$$

where the diagonal matrix $\mathbf{Q}$ is defined

$$Q_{ii} = \frac{1}{|m_i|} \, . \tag{4.1.29}$$

To prevent instability when small model parameters are present, the matrix $\mathbf{Q}$ is usually modified to

$$\mathbf{Q}_{ii} = \begin{cases} |m_i|^{-1} & \text{if } |m_i| > \epsilon \\ \epsilon^{-1} & \text{if } |m_i| \leq \epsilon \, , \end{cases}$$

where ϵ is a small, positive parameter (Scales and Smith, 1994).

When all the constants are collected together in the weighting parameter, the solution to the cost function becomes

$$\mathbf{G}^T \mathbf{G} \mathbf{m} - \mathbf{G}^T \mathbf{d} + \mu \, \mathbf{Q} \mathbf{m} = 0 \, . \tag{4.1.30}$$

Note that since $\mathbf{Q}$ is a function of the model parameters, $\mathbf{m}$, the solution is found by solving a non-linear equation.

MAP solution to a Cauchy prior distribution

The prior distribution is set to be a Cauchy distribution (4.1.17) with zero mean, and the likelihood distribution remains Gaussian (4.1.18), as in the previous section. The posterior distribution then becomes

$$\begin{aligned} P(\mathbf{m}, \mathbf{d}) \;\propto\; K_1 \, K_2 \left[\prod_{i=1}^{M} \frac{1}{1 + m_i^2/\sigma_m^2} \right] \times \\ exp \left(\frac{-1}{2\sigma^2} \, (\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) \right) . \end{aligned} \tag{4.1.31}$$

Using the definition that $1/a = exp\,(\,ln(1/a)\,) = exp\,(\,-ln(a)\,)$, the posterior distribution can be defined as

$$
\begin{aligned}
P(\mathbf{m}, \mathbf{d}) \quad \propto \quad & K_1\,K_2\,exp\left(\sum_{i=1}^{M} -ln(1+m_i^2/\sigma_m^2)\right) \times \\
& exp\left(\frac{-1}{2\sigma^2}\,(\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d})\right).
\end{aligned}
\tag{4.1.32}
$$

Finding the maximum of the posterior function is equivalent to finding the minimum of

$$
\frac{1}{2\sigma^2}(\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d}) + \sum_{i=1}^{M} ln(1+m_i^2/\sigma_m^2),
\tag{4.1.33}
$$

which can be written as the cost function

$$
J(\mathbf{m}) = (\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d}) + \mu R(\mathbf{m}).
\tag{4.1.34}
$$

The regularization term is defined

$$
R(\mathbf{m}) = \sum_{i=1}^{M} ln(1+m_i^2/\sigma_m^2),
\tag{4.1.35}
$$

and its derivative is

$$
\frac{\partial}{\partial m_k}R(\mathbf{m}) = \frac{2}{\sigma_m^2}\,m_k\,\frac{1}{(1+m_k^2/\sigma_m^2)}.
\tag{4.1.36}
$$

The constant term, $\frac{1}{\sigma_m^2}$, is absorbed into the weighting parameter μ, and is no longer included as part of the regularization gradient. Thus, it can be stated that

$$\nabla R(\mathbf{m}) = \begin{pmatrix} \frac{\partial R(\mathbf{m})}{\partial m_1} \\[6pt] \frac{\partial R(\mathbf{m})}{\partial m_2} \\[6pt] \frac{\partial R(\mathbf{m})}{\partial m_3} \\[6pt] \vdots \end{pmatrix} = \begin{pmatrix} 2\, m_1\, \frac{1}{1+m_1^2/\sigma_m^2} \\[6pt] 2\, m_1\, \frac{1}{1+m_2^2/\sigma_m^2} \\[6pt] 2\, m_1\, \frac{1}{1+m_3^2/\sigma_m^2} \\[6pt] \vdots \end{pmatrix}. \tag{4.1.37}$$

This is rearranged into

$$\nabla R(\mathbf{m}) = 2 \begin{pmatrix} \frac{1}{1+m_1^2/\sigma_m^2} & 0 & 0 & 0 \\[6pt] 0 & \frac{1}{1+m_2^2/\sigma_m^2} & 0 & 0 \\[6pt] 0 & 0 & \frac{1}{1+m_3^2/\sigma_m^2} & 0 \\[6pt] 0 & 0 & 0 & \ddots \end{pmatrix} \begin{pmatrix} m_1 \\[6pt] m_2 \\[6pt] m_3 \\[6pt] \vdots \end{pmatrix}. \tag{4.1.38}$$

Finally, the gradient of the regularization term is

$$\nabla R(\mathbf{m}) = 2\, \mathbf{Q}\mathbf{m}, \tag{4.1.39}$$

where the diagonal matrix $\mathbf{Q}$ is defined

$$Q_{ii} = \frac{1}{1 + m_i^2/\sigma_m^2}. \tag{4.1.40}$$

The solution to the cost function is again given by a non-linear equation

$$\mathbf{G}^T\mathbf{G}\mathbf{m} - \mathbf{G}^T\mathbf{d} + \mu\, \mathbf{Q}\mathbf{m} = 0. \tag{4.1.41}$$

MAP solution to a modified Cauchy prior distribution

The focus of this work is the sparse constraint used by Charbonnier et al. (1997). Their regularization term is presented within the cost function, rather than as

a Bayesian prior distribution. These two have been shown different perspectives of the same problem, and so the derivation of the solution is continued from the statement of the cost function. The cost function to be minimized is again

$$J(\mathbf{m}) = (\mathbf{Gm} - \mathbf{d})^T (\mathbf{Gm} - \mathbf{d}) + \mu R(\mathbf{m}) \,. \tag{4.1.42}$$

The regularization term is defined

$$R(\mathbf{m}) = \sum_{i=1}^{M} \frac{(m_i/\delta)^2}{1 + (m_i/\delta)^2} \,, \tag{4.1.43}$$

where δ is a scaling parameter. The derivative of the regularization term is

$$\frac{\partial}{\partial m_k} R(\mathbf{m}) = \frac{1}{\delta} \frac{2\,(m_k/\delta)}{[1 + (m_k/\delta)^2]^2} \,. \tag{4.1.44}$$

This can be rearranged into

$$\frac{\partial}{\partial m_k} R(\mathbf{m}) = 2 \frac{1}{\delta^2} \frac{m_k}{[1 + (m_k/\delta)^2]^2} \,. \tag{4.1.45}$$

The constant term, $\frac{1}{\delta^2}$, becomes part of the weighting parameter μ, and is no longer included as part of the regularization gradient. Thus, the gradient is defined

$$\nabla R(\mathbf{m}) = \begin{pmatrix} \frac{\partial R(\mathbf{m})}{\partial m_1} \\ \frac{\partial R(\mathbf{m})}{\partial m_2} \\ \frac{\partial R(\mathbf{m})}{\partial m_3} \\ \vdots \end{pmatrix} = \begin{pmatrix} 2\,m_1 \frac{1}{[1+(m_1/\delta)^2]^2} \\ 2\,m_2 \frac{1}{[1+(m_2/\delta)^2]^2} \\ 2\,m_3 \frac{1}{[1+(m_3/\delta)^2]^2} \\ \vdots \end{pmatrix} \,. \tag{4.1.46}$$

This is rewritten

$$\nabla R(\mathbf{m}) = 2 \begin{pmatrix} \frac{1}{[1+(m_1/\delta)^2]^2} & 0 & 0 & 0 \\ 0 & \frac{1}{[1+(m_2/\delta)^2]^2} & 0 & 0 \\ 0 & 0 & \frac{1}{[1+(m_3/\delta)^2]^2} & 0 \\ 0 & 0 & 0 & \ddots \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \end{pmatrix}. \qquad (4.1.47)$$

Finally, the gradient of the regularization term is

$$\nabla R(\mathbf{m}) = 2\,\mathbf{Q}\mathbf{m}, \qquad (4.1.48)$$

where the diagonal matrix $\mathbf{Q}$ is

$$Q_{ii} = \frac{1}{[1+(m_i/\delta)^2]^2}. \qquad (4.1.49)$$

The similarity to the previously discussed Cauchy prior distribution constraint,

$$\mathbf{Q}_{(Cauchy)\ ii} = \frac{1}{[1+(m_i/\sigma_m)^2]}, \qquad (4.1.50)$$

is apparent. For this reason, this sparse constraint will be referred to as the modified Cauchy prior distribution. The non-linear solution to the cost function again reduces to

$$\mathbf{G}^T\mathbf{G}\mathbf{m} - \mathbf{G}^T\mathbf{d} + \mu\,\mathbf{Q}\mathbf{m} = 0. \qquad (4.1.51)$$

4.2 Solving the non-linear problem

All the sparse regularization terms introduced result in non-linear expressions. The degree of a non-linearity in a problem will determine the optimum solution

method. There are four general techniques of solving non-linear problems (Tarantola, 1987). The first method is known as systematic exploration, or the exhaustive search technique. It consists of systematically trying all combinations of model parameters to find the solution. This method will work equally well for all degrees of non-linear problems. However, as one might suspect, it is extremely time consuming, not very realistic, and not recommended except for very small problems.

The second class of algorithms to solve non-linear problems are the stochastic ones. These are useful for strongly non-linear problems, where the cost function is multimodal, because they search a wide portion of the model space. Included in this category are Monte Carlo techniques, Simulated Annealing (Kirkpatrick et al., 1983), and Genetic Algorithms (Scales and Smith, 1994).

Gradient methods are most suited for solving quasi-linear problems. As the name suggests, these techniques involve using the gradient of the cost function to search for the minimum. Examples include the Newton method and its modifications, steepest descent and the conjugate gradients method. They can be used for all degrees of non-linear problems, but run the risk of becoming stuck in a local extremum in multi-modal functions.

Finally, if the non-linear problem is quasi-linear, one can approximate it as a linear problem and solve it iteratively. This method is known as iteratively re-weighted least squares (IRLS) for reasons which will be discussed further on. Following the example of Scales and Smith (1994), I demonstrate how the inverse problem subject to the three sparse constraints may be solved using IRLS.

4.2.1   Solving sparse inverse problems with IRLS

The solution to a cost function containing an non-linear sparse regularization term is

$$\widetilde{\mathbf{m}} = \left[\mathbf{G}^T \mathbf{G} + \mu \, \mathbf{Q} \right]^{-1} \mathbf{G}^T \mathbf{d} \,. \tag{4.2.1}$$

The matrix $\mathbf{Q}$ is a diagonal matrix defined in terms of the vector $\mathbf{m}$. The definitions of this weighting matrix for the exponential, Cauchy, and modified Cauchy prior distributions were derived in the previous section. This solution may be found iteratively by setting

$$\widetilde{\mathbf{m}}^k = \left[\mathbf{G}^T \mathbf{G} + \mu \mathbf{Q}^{k-1} \right]^{-1} \mathbf{G}^T \mathbf{d} \,, \tag{4.2.2}$$

where

$$\mathbf{Q}^{k-1} = \mathbf{Q}(\mathbf{m}^{k-1}) \,. \tag{4.2.3}$$

The index k indicates the iteration number. The algorithm follows the steps listed below, outlined by Sacchi (1997).

1. Assume an initial vector of model parameters, $\mathbf{m}$. This can be set to a default vector of zeros if there is not enough information to begin elsewhere.

2. Select the trade-off parameter.

3. Compute the matrix $\mathbf{G}^T \mathbf{G}$, and the value of $\mathbf{Q}^0$ from the initial vector of model parameters.

4. Iteratively solve the non-linear equation (4.2.2).

5. Proceed until the following tolerance criterion is satisfied

$$\frac{\mid J^k - J^{k-1} \mid}{\left(\mid J^k \mid + \mid J^{k-1} \mid \right)/2} \le Tolerance \,, \tag{4.2.4}$$

where J^k is the cost function evaluated at iteration k.

6. Evaluate the χ^2 criterion using the data misfit to determine if the trade-off parameter should be adjusted.

It can be seen that the iterative solution resembles the weighted damped least squares solution, as described by equation (3.2.2). The weighting matrix $\mathbf{Q}$ is updated, and hence the solution is re-weighted at each iteration. It is from these properties that the iteratively re-weighted least squares (IRLS) algorithm derives its name.

The weighting matrices for each of the three sparse constraints award a weight inversely proportional to the corresponding model parameters. Very small or zero values are awarded resulting in a solution that is primarily zeros, except where it is necessary to deviate from this range to satisfy the data.

4.3 Applications to acoustic impedance inversion

4.3.1 Reflection coefficients and acoustic impedance

According to Snell's Law, reflected and refracted (transmitted) waves will occur at boundaries between two different media (Beck, 1991). As an incident wave encroaches on a boundary, its energy will be partitioned into reflected and refracted waves. The magnitude of energy that each wave receives is indicated in their amplitudes, and depends on the contrast in material properties at the boundary. An extreme boundary will transmit very little energy, and reflect almost all of it. This is obvious in the simple thought experiment of sending a transverse wave along a rope to a fixed point on a wall. As the wave comes to the interface between air and solid, it will be reflected back at approximately the same magnitude.

This behavior can be described mathematically, and gives rise to a definition of the ratios of the amplitudes of the incident, transmitted, and reflected waves (see Appendix B). For a 2D acoustic medium, where the incident wavefield approaches perpendicular to the interface dividing 2 materials

$$\frac{A_{trans}}{A_{in}} = \frac{2\,\rho_1 c_2}{\rho_1 c_1 + \rho_2 c_2}\,. \tag{4.3.1}$$

The subscripts on density ρ, and acoustic velocity c indicate which material the property relates to. Note that the wave travels from material 1 into material 2. The ratio in equation (4.3.1) is also known as the transmission coefficient. In a similar vein, the reflection coefficient is found by

$$\frac{A_{refl}}{A_{in}} = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_1 c_1 + \rho_2 c_2}\,. \tag{4.3.2}$$

The reflection coefficient, or reflectivity, can also be written in terms of acoustic impedance, $I = \rho c$, as

$$\mathcal{R} = \frac{I_2 - I_1}{I_1 + I_2}\,. \tag{4.3.3}$$

In geophysics, one of the most common representations of the earth is as a layered model, where ρ_k and c_k represent the density and velocity of the kth layer. One can relate the reflection coefficient and acoustic impedance of each layer, assuming a normal incidence wave

$$r_k = \frac{I_{k+1} - I_k}{I_{k+1} + I_k}\,. \tag{4.3.4}$$

The impedance values are calculated from wavefield data recorded in time, so it is more convenient to express this relation in time

$$r(t) = \frac{I(t+1) - I(t)}{I(t+1) + I(t)} \,. \tag{4.3.5}$$

Assuming that the difference between the two impedance values is small, it can be approximated that $I(t+1) \approx I(t)$

$$r(t) = \frac{I(t+1) - I(t)}{I(t+1) + I(t)} \approx \frac{I(\Delta t)}{2\,I(t)} \,. \tag{4.3.6}$$

This relation can be written as a differential when the reflection coefficient is relatively small ($r < 0.3$) (Waters, 1978; Walker and Ulrych, 1983). The reflectivity is approximated

$$r(t) \approx \frac{dI(t)}{2\,I(t)} \approx \frac{\partial}{\partial t} \frac{ln[\,I(t)\,]}{2} \,. \tag{4.3.7}$$

When this equation is integrated with respect to time, an expression for the relative impedances occurs

$$ln\,\frac{I(t)}{I(t_0)} = \int_{t_0}^{t} 2\,r(\eta)d\eta \,, \tag{4.3.8}$$

or

$$\frac{I(t)}{I(t_0)} = exp\left[2\int_{t_0}^{t} r(\eta)d\eta\right] \,. \tag{4.3.9}$$

Assuming that $t_0 = 0$, the impedance at time t becomes

$$I(t) = I(0)\,exp\left[2\int_{0}^{t} r(\eta)d\eta\right] \,, \tag{4.3.10}$$

where $I(0)$ is the initial impedance value.

4.3.2 The convolutional model

The relationship between reflectivity and seismic information is expressed in the convolutional model

$$s(t) = I_r(t) \circledast w(t),\qquad(4.3.11)$$

that includes terms for the seismogram, $s(t)$; source waveform, $w(t)$; and the earth impulse response, $I_r(t)$. The earth's impulse response includes the reflection coefficients as well as all other physical responses. These other effects include geometrical divergence, anelastic absorption, wavelet dispersion, transmission losses across boundaries, multiple reflections and mode conversions (Oldenburg et al., 1983; Margrave, 1998). By combining many of these terms with the source waveform term, as well as assuming that the data has been processed to remove as many of the earth's other effects as possible, the convolutional model becomes

$$s(t) = w_e(t) \circledast r(t) + n(t).\qquad(4.3.12)$$

In this expression, $n(t)$ is Gaussian noise, $r(t)$ is the reflectivity series, and $w_e(t)$ is the earth filter that includes the source wavelet. All effects accounted for in the earth filter term are those considered stationary, or those that do not change as the wave propagates forward in time and depth.

A fundamental problem in retrieving the reflectivity series from the seismogram is that the source wavelet and resulting seismic data are bandlimited. This creates a null space in the frequency domain, and the recovered Earth model will be bandlimited. To demonstrate this, a simple example is created (Figure 4.4). The earth filter is set as a 30 Hz Ricker wavelet. A synthetic seismogram is created by convolving the wavelet with a random reflectivity model. The frequency spectrum of this model is displayed in Figure 4.5.
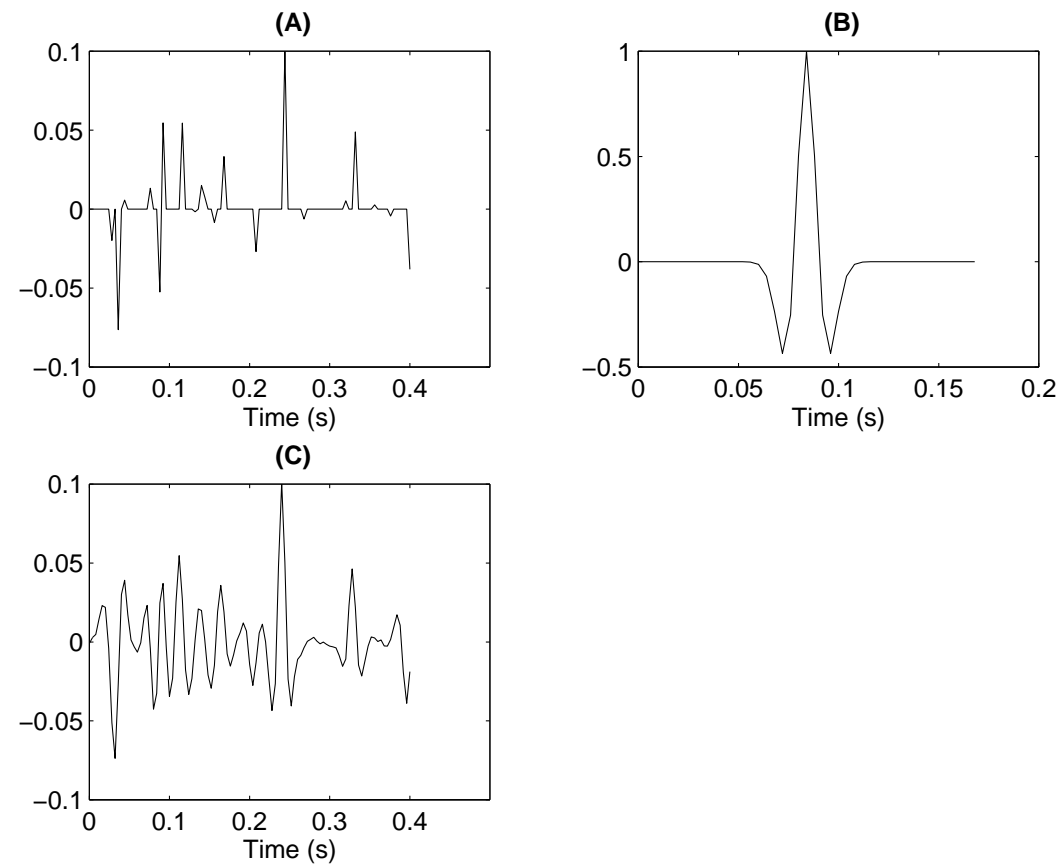
Figure 4.4: The reflectivity series (A) and wavelet (B) used to create the synthetic seismogram (C). One percent Gaussian noise is added to the synthetic data.
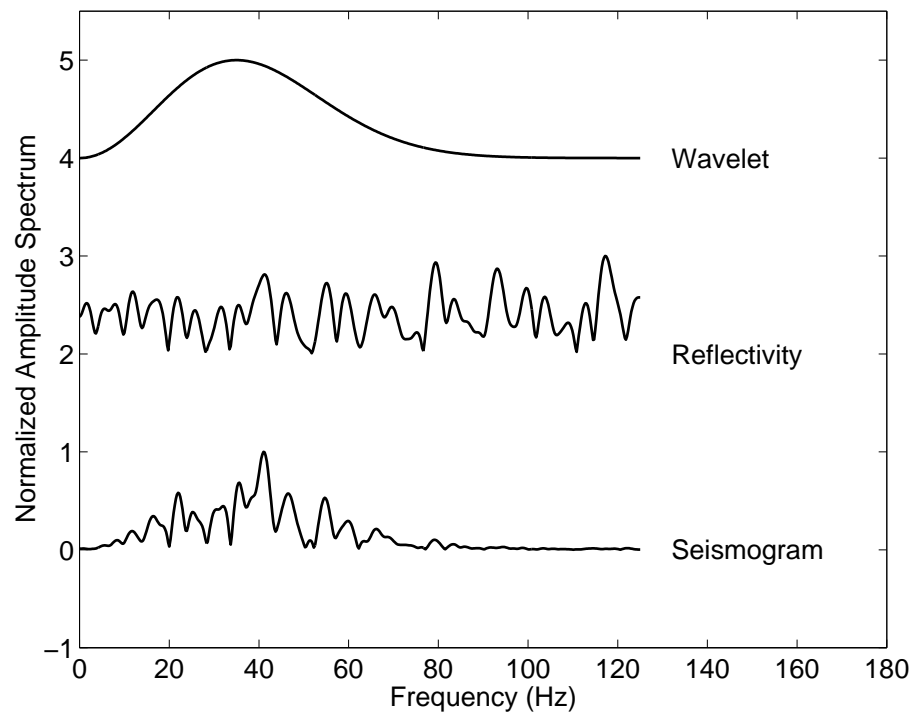
Figure 4.5: The frequency spectrum of the wavelet, reflectivity and seismogram of the simple model. It is clear how the bandlimited wavelet truncates the spectrum of the seismic data.

4.3.3 The discrete inverse problem

The convolutional model can be posed as a discrete inverse problem

$$\mathbf{s} = \mathbf{Wr} + \mathbf{n}\,. \tag{4.3.13}$$

The matrix $\mathbf{W}$ is the earth filter presented in convolutional form (see Appendix C). This problem is non-unique, due to the missing frequency information in the data. Therefore, it must be regularized to retrieve a stable and unique solution. The most common constraint is a sparse constraint on the reflectivity.

A sparse solution creates the most simple earth model possible: the one with the least number of layers. In addition, enforcing isolated spikes in the reflectivity series approximates a superposition of delta functions. This will increase the frequency content, following from the fact that a delta function has a white frequency spectrum. The two most common methods of enforcing sparseness in impedance inversion are prior Cauchy or exponential probability distributions (Oldenburg et al., 1983; Sacchi, 1997; Djikpesse and Tarantola, 1999).

The cost function is

$$J(\mathbf{r}) = (\mathbf{Wr} - \mathbf{s})^T(\mathbf{Wr} - \mathbf{s}) + \mu\,R(\mathbf{r})\,, \tag{4.3.14}$$

where $R(\mathbf{r})$ is a function to enforce sparseness. Once the reflectivity solution is found, the impedance series can be recovered using equation (4.3.10). The DLS solution is compared to solutions of three sparse constraints: the exponential, Cauchy and modified Cauchy prior distributions. The non-linear problems are solved using IRLS. Finally, because the 1D problem is relatively small, the operators are kept as matrices, and inversion is performed using a Matlab function.

The recovered reflectivity is displayed in Figures 4.6. The narrower peaks within the reflectivity series recovered by the sparse constraints indicate a broader

Figure 4.6: The reflectivity series recovered by DLS (A), and the exponential (B), Cauchy (C), and modified Cauchy (D) prior distributions.

frequency content than the DLS solution. The frequency spectrums of the reflectivity (Figure 4.7) show that though all the sparse solutions increase the bandwidth beyond the 80 Hz of the data. The Cauchy and modified Cauchy priors recover the most accurate reflectivity estimates and frequency spectrums.

The recovered impedance profiles are displayed in Figure 4.8. The best solution is recovered by the modified Cauchy distribution. The piecewise continuous solution manages to retain sharp discontinuities while preserving smooth, planar areas. This agrees with previous studies that find the modified Cauchy distribution to be especially suited for blocky inversion (Charbonnier et al., 1997).

Within the geophysical industry, it is a general rule that further constraints on the reflectivity problem are necessary to recover a suitable solution. The most common technique of adding constraints is to use prior information of

Figure 4.7: The frequency content of the reflectivity recovered by DLS (A), and the exponential (B), Cauchy (C), and modified Cauchy (D) prior distributions. Note that each frequency spectrum is scaled to a maximum of one for display purposes.
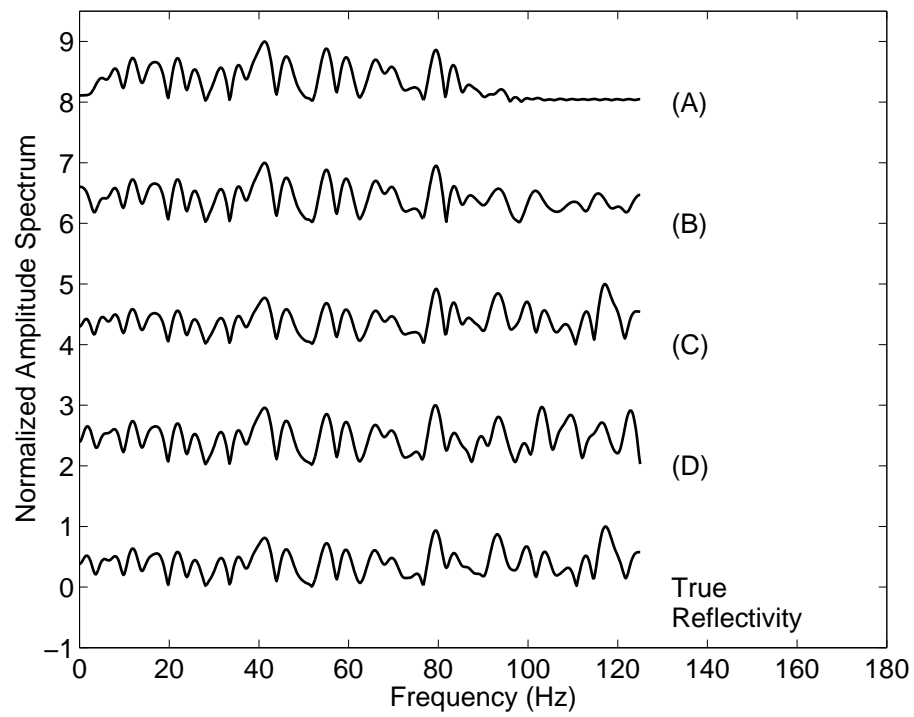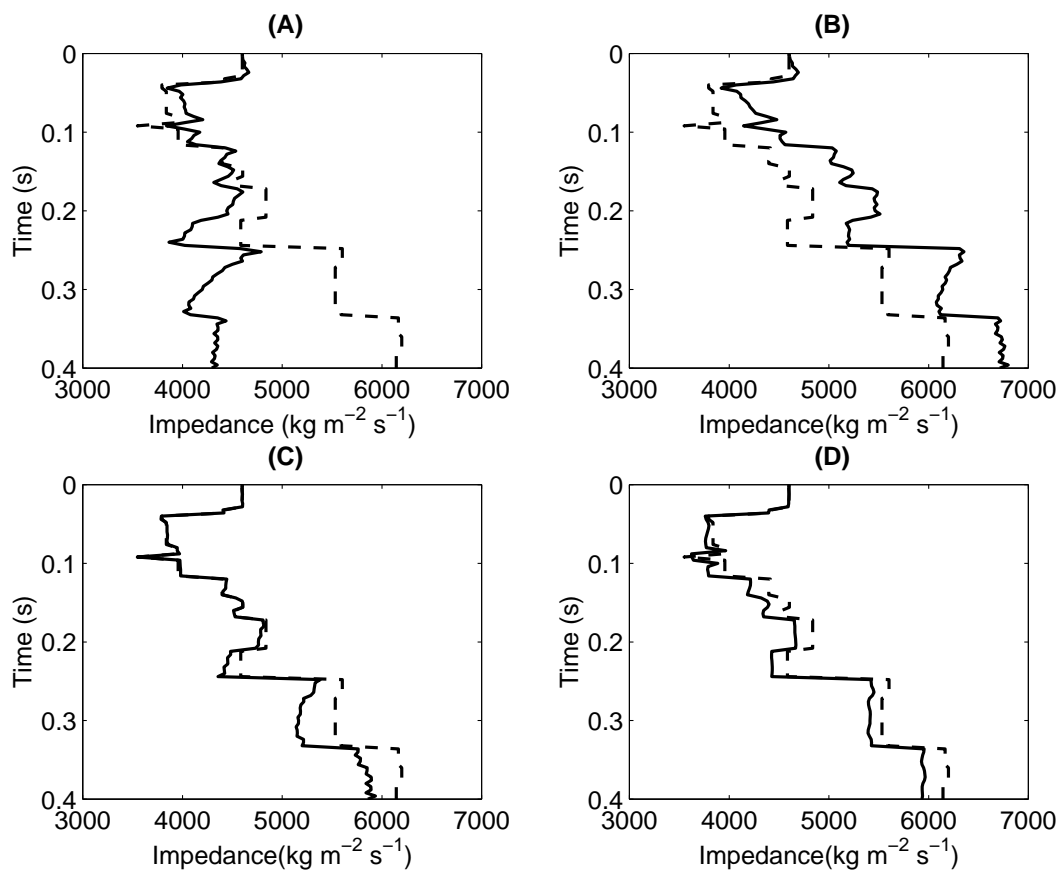
Figure 4.8: The impedance series recovered by DLS (A), and the exponential (B), Cauchy (C), and modified Cauchy (D) prior distributions. The reference impedance series of the model is represented by the dashed line.

the geological model. When an oil or gas well is drilled, information about the velocity and density of the surrounding material is recorded. Impedance information calculated from these measurements will constrain the model, and assist in increasing frequency bandwidth. Knowing that the impedance is a function of the integral of the reflectivity, constraints can be directly placed on the reflectivity inversion. The new cost function will become

$$J(\mathbf{r}) = (\mathbf{Wr} - \mathbf{s})^T (\mathbf{Wr} - \mathbf{s}) + \alpha \, (\mathbf{Cr} - \epsilon)^T (\mathbf{Cr} - \epsilon) + \mu \, R(\mathbf{r}) \,. \qquad (4.3.15)$$

The matrix $\mathbf{C}$ is a simple integrator operator, and α is the associated weighting parameter. The vector of impedance constraints, ϵ, is in the form of the left-hand side of

$$\frac{1}{2} \, ln \left[\frac{I(t)}{I(t_0)} \right] = \int_{t_0}^{t} r(\eta) d\eta \,. \qquad (4.3.16)$$

The same synthetic data are used in the inversion. The three sparse prior distributions are tested, and 4 impedance constraints are given. The recovered reflectivities and impedances are displayed in Figures 4.9 and 4.10. Finally, the frequency spectrums of the three solutions are compared in Figure 4.11. It is apparent that the solutions match our earth model more closely, and that frequency content is increased. Again, the modified Cauchy prior distribution selects the best solution. Not only does it allow the sharp discontinuities in the impedance profile, but the constant layers are relatively smooth and artifact free.

Figure 4.9: The reflectivity series recovered by the exponential (A), Cauchy (B), and modified Cauchy (C) prior distributions, subject to 4 impedance constraints.

Figure 4.10: The impedance series recovered by the exponential (A), Cauchy (B), and modified Cauchy (C) prior distributions, subject to 4 impedance constraints. The reference impedance series of the model is represented by the dashed line, and the constraints by stars.

Figure 4.11: The frequency content of the reflectivity recovered by the exponential (A), Cauchy (B), and modified Cauchy 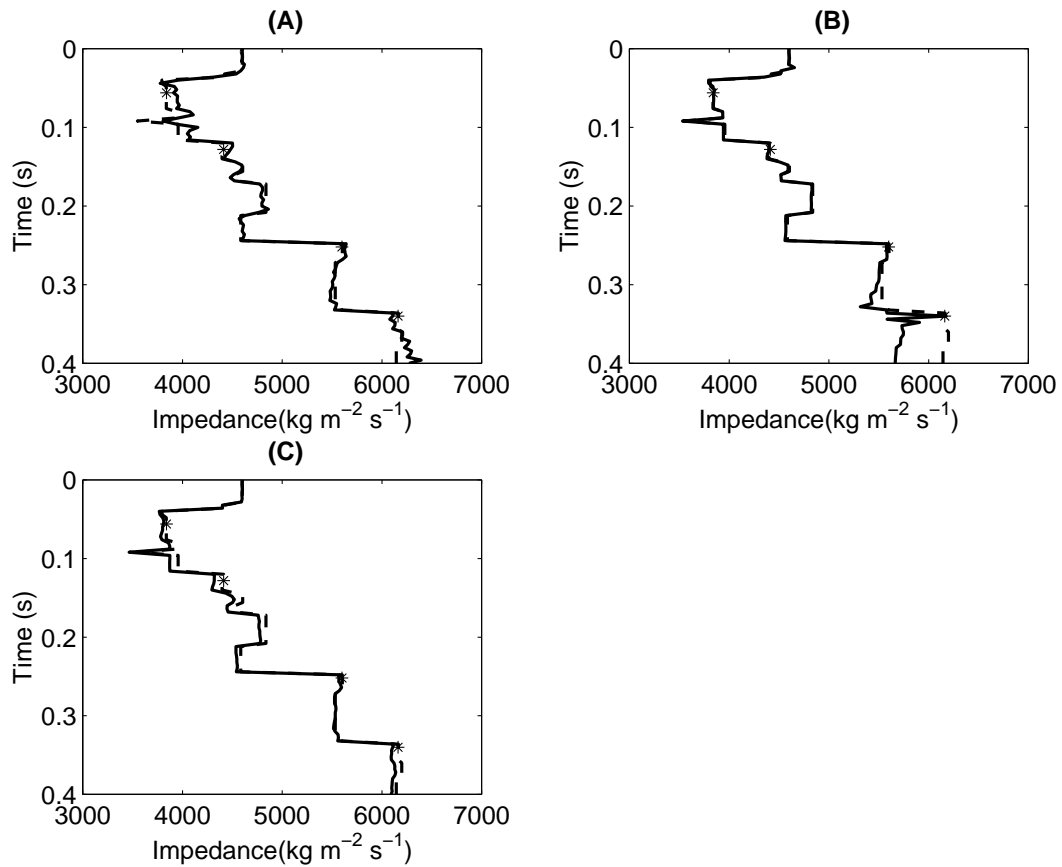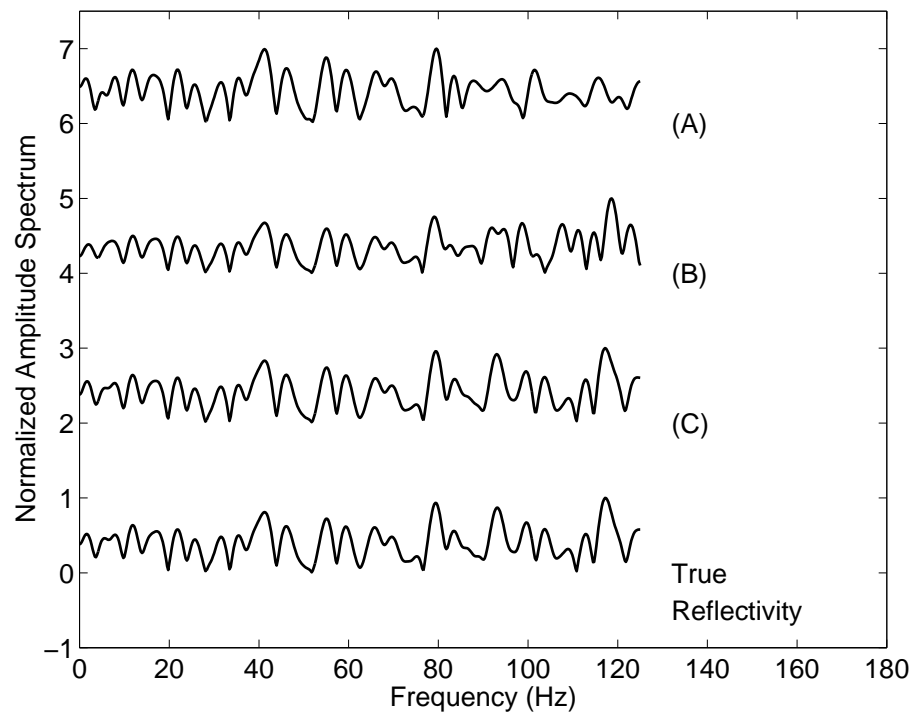(C) prior distributions, subject to 4 impedance constraints. Note that each frequency spectrum is scaled to a maximum of one for display purposes.

4.3.4 Behavior of the hyperparameters of the modified Cauchy prior

Recall that the modified Cauchy prior results in a solution

$$\widetilde{\mathbf{m}}^k = \left[\mathbf{G}^T\mathbf{G} + \mu\mathbf{Q}^{k-1} \right]^{-1} \mathbf{G}^T\mathbf{d}, \qquad (4.3.17)$$

where

$$\mathbf{Q}_{ii}^{k-1} = \frac{1}{[\, 1 + (m_i^{k-1}/\delta)^2\,]^2}\,. \qquad (4.3.18)$$

This particular prior distribution has a greater flexibility in specifying model constraints because of its two hyperparameters: μ and δ. In this section, the interaction and behavior of the weighting terms are explored. In order to isolate the parameters solely in the sparse regularization term, an impedance constraint is not added to the cost function.

A series of tests on synthetic data are performed to understand the roles that δ and μ play in determining the solution. A seismogram is created by convolving a 35 Hz Ricker wavelet with a known reflectivity series, and adding one percent Gaussian noise (Figure 4.12). The first step is to confirm that our understanding of the limiting cases of the regularization term is correct. To test this, μ is held constant, while the values of the scaling parameter δ are varied. The recovered values for reflectivity are shown in Figure 4.13.

It is shown that when δ is too large, sparseness is not enforced. Correspondingly, when δ is small, the solution becomes very sparse. This follows from what one would expect from the regularization term: that δ is inversely proportional to the amount of weight placed on the regularization term.

The next test is to determine how the two parameters affect the data misfit of the solution. Using the χ^2 test, it is found that there are multiple pairs of
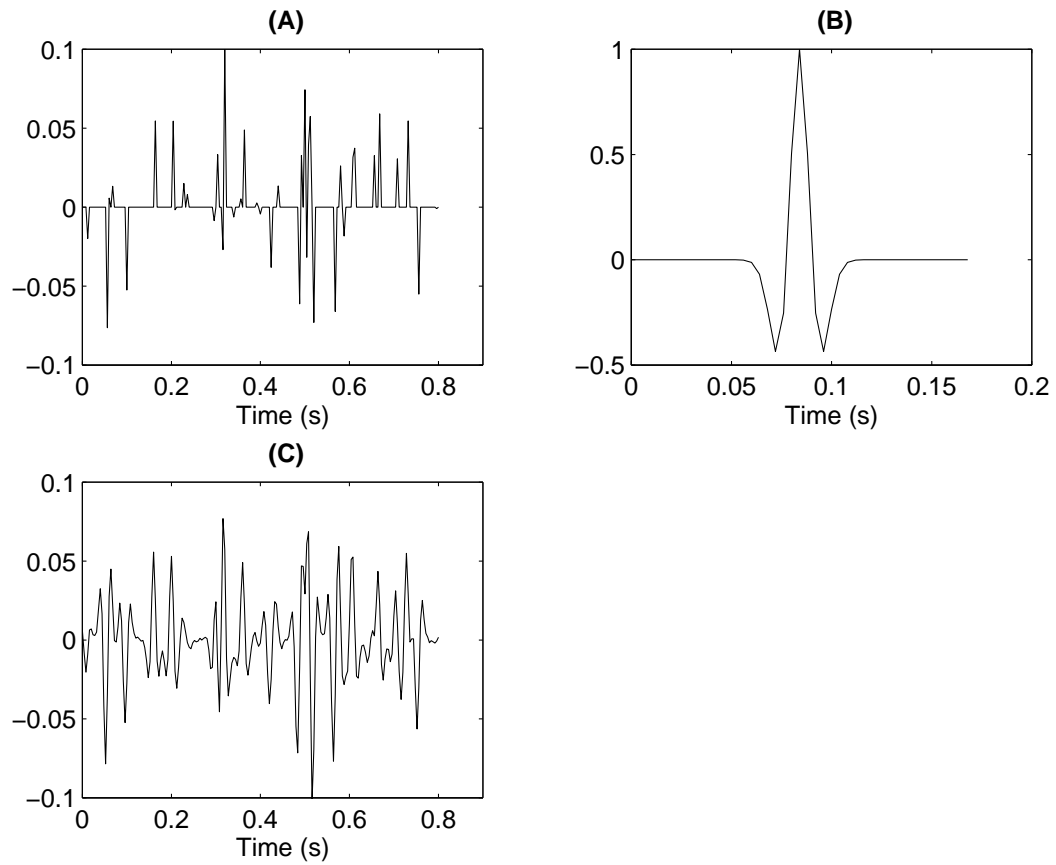
Figure 4.12: The reflectivity series (A) and wavelet (B) used to create the synthetic seismogram (C). One percent Gaussian noise is added to the synthetic data.
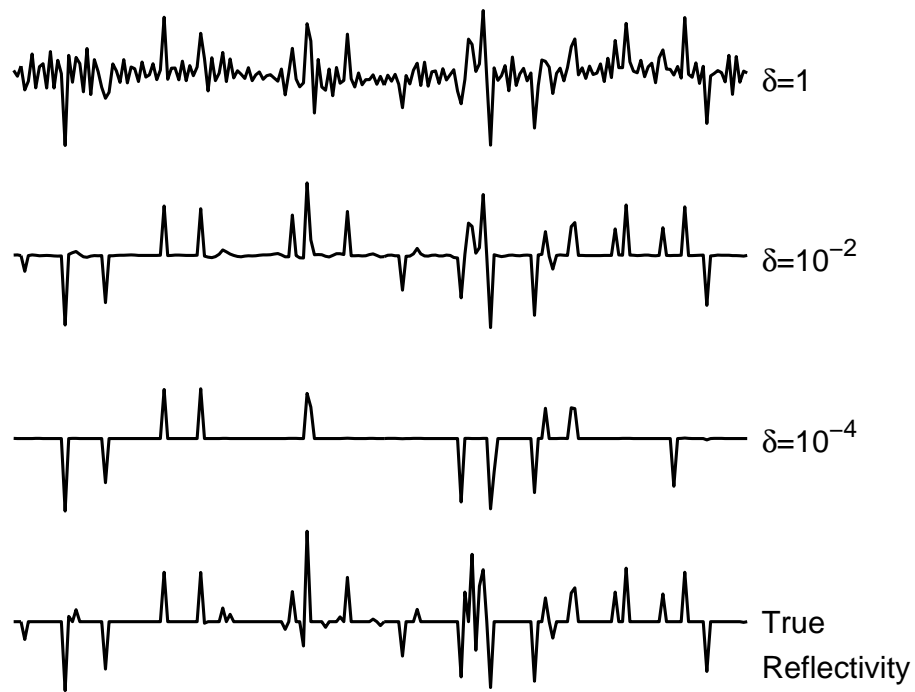
Figure 4.13: The recovered reflectivity for 3 different values of the scaling parameter, δ, when the trade-off parameter, μ, is held at a constant 10^{-4}.
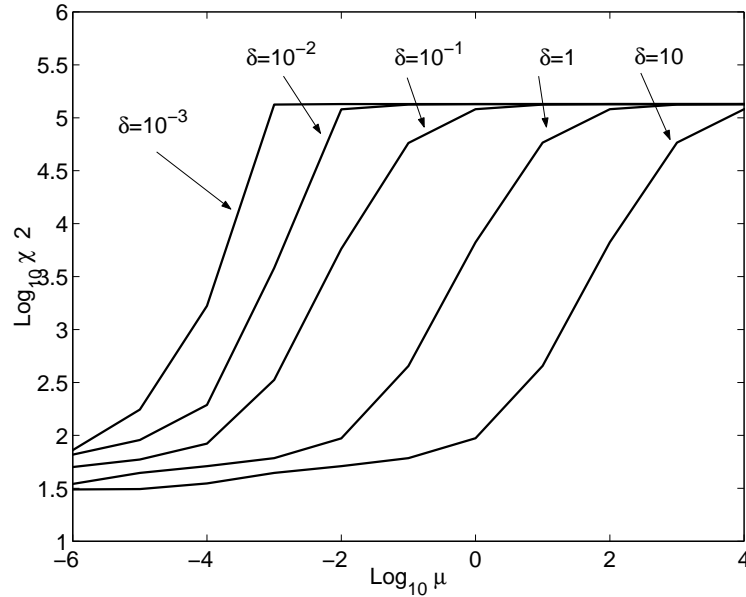
Figure 4.14: A logarithmic plot of χ^2 versus the trade-off parameter μ, for varying values of the scaling parameter δ.

parameters that will give the same misfit (Figure 4.14). As expected, larger values of the trade-off parameter or smaller values of the scaling parameter δ will increase misfit.

To better understand the multiple solutions, three different solutions that have the desired χ^2 value are examined. The recovered values for reflectivity are displayed in Figure 4.15. There is a single solution that is most accurate. When the scaling parameter is small with respect to the reflectivity, the solution is very spiky, and so a small μ is necessary preserve fidelity to the data. If δ is large with respect to r, the solution approaches a damped least squares solution. In order to preserve a small misfit, μ becomes large to keep the entire solution close to zero, compensating for the lack of sparseness.

The final test is to explore the behavior of the cost function as the parameters are changed. The cost function is redefined as:

Figure 4.15: Solutions for the reflectivity that all have the same desired value of $\chi^2 = 200$.

$$J = M + \mu R, \tag{4.3.19}$$

where M is the misfit term, $(\mathbf{Wr} - \mathbf{s})^T(\mathbf{Wr} - \mathbf{s})$, and R is the regularization function, $R(\mathbf{r})$. Figures 4.16 (A) and (D) show again that th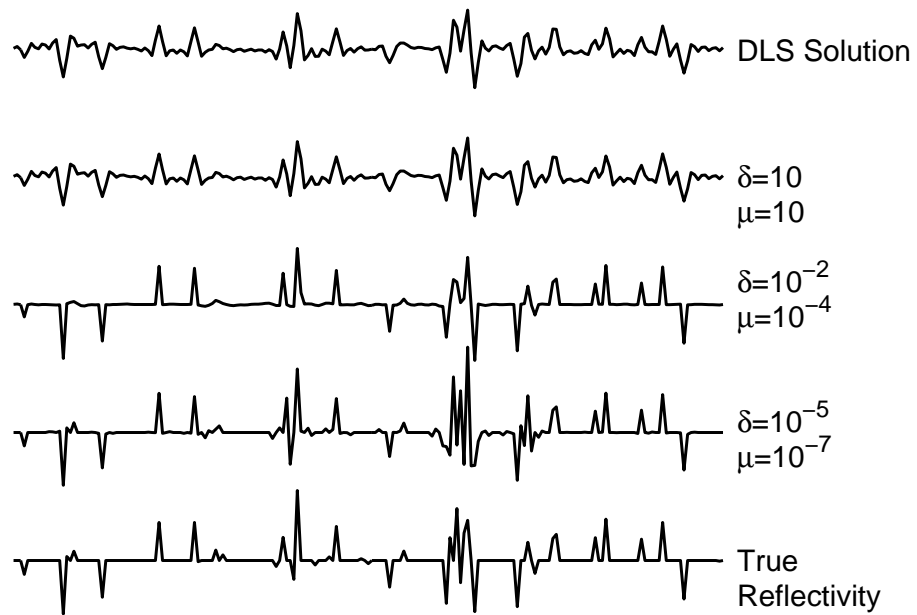e data misfit, and the magnitude of the cost function, are proportional to μ and inversely proportional to δ. The regularization function decreases for larger μ values, as one would expect (Figure 4.16(B)). Like the misfit term, it is inversely proportional to δ. For larger values of δ, the denominator of the re-weighting matrix in equation 4.3.18 will approach 1, and the weight on the regularization function will decrease. As δ decreases, the denominator will become large, the fraction will approach unity, and the weight on the regularization function will increase.

Figure 4.16 (C) illustrates the regularization function multiplied by the trade-off parameter. One might be tempted to believe that this plot describes the symmetry of the inverse problem, but it does not. The figure simply shows that the magnitude of the regularization function is scaled by the corresponding trade-off parameter.

4.4 Summary

In this chapter, the Bayesian definition of constraints as prior probability distributions has been introduced. A Gaussian prior probability distribution will enforce a smooth solution that clusters around zero. Sparse constraints correspond to long-tailed distributions that have a greater area tightly centered around zero, but allow larger magnitudes in the few samples that deviate from zero. The exponential, Cauchy, and modified Cauchy probability distributions are all examples of sparse constraints. These sparse constraints result in non-linear inverse problems. Fortunately, the problems are quasi-linear, and so IRLS
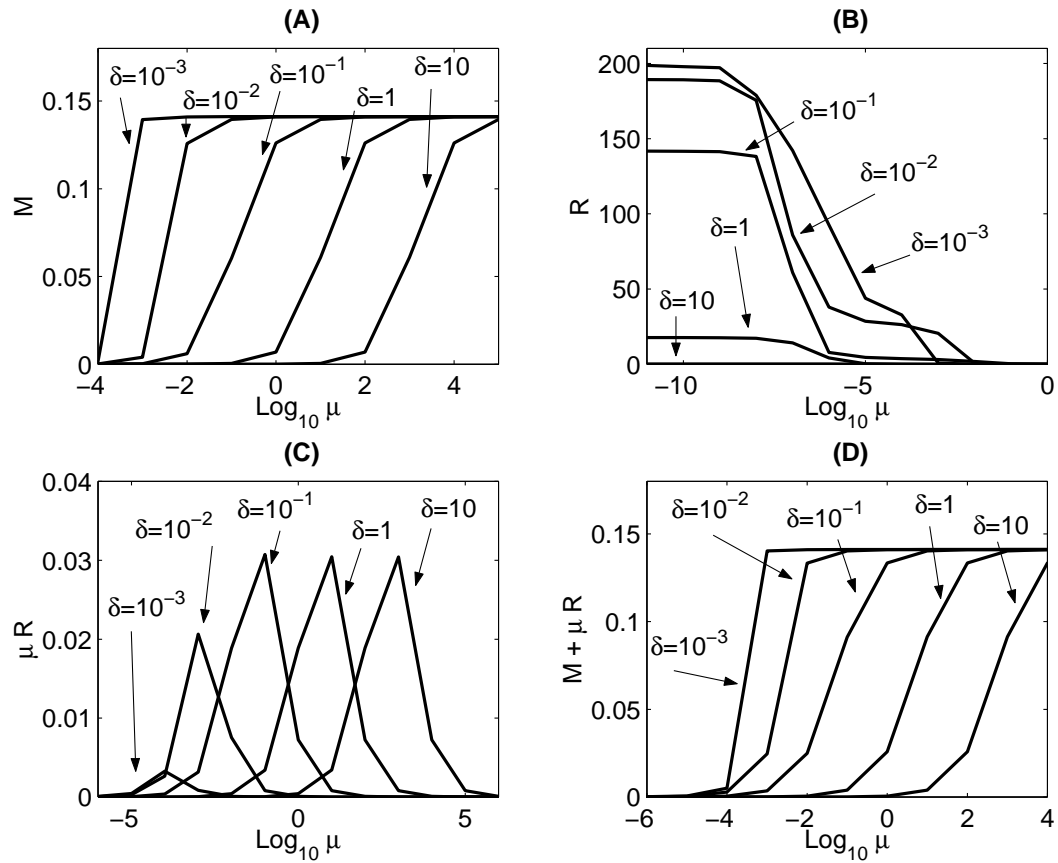
Figure 4.16: Different components of the cost function plotted against μ for varying values of the scaling parameter δ: the data misfit term (A), the regularization term (B), the regularization term multiplied by the weighting parameter (C), and the complete cost function (D).

can be used to find the solutions in an iterative, linear manner.

The modified Cauchy function is compared to Cauchy and exponential prior distributions within impedance inversion. It successfully recovers sparse reflectivity series, and is especially suited for recovering blocky profiles. The examples show that the modified Cauchy prior is able to recover sharp discontinuities separated by flat, planar areas more successfully than the other distributions. Finally, the behavior of the two hyperparameters of the modified Cauchy prior is explored.

Chapter 5

Blocky inversion techniques

5.1 Introduction

The idea behind blocky inversion techniques is simple: a function will become piecewise constant if its first order derivatives are sparse. This is seen in blocky impedance inversion. A sparse reflectivity will result in an impedance, $I_k \propto \int_k r_k dk$, that is blocky. Blocky inversion is not only applicable in areas where the model image consists of blocks, but can be used anywhere sharp edges divided by planar features are present. It has the effect of sharpening and focusing the image, and in this way reducing the effects of noise and blurring.

The blocky inversion constraint explored here has been introduced in the previous chapter as the modified Cauchy prior distribution. In the context of blocky inversion it will be referred to as the Edge-Preserving Regularization (EPR) function. This regularization technique has been successfully applied to gravity data (Portniaguine and Zhdanov, 1999), magnetic resonance imaging for medical applications (Charbonnier et al., 1997; Barone, 1999) and radio astronomy (Molina et al., 2001).

Within this chapter, the applicability of the EPR function for 2D acoustic

migration/inversion problems will be tested. I show the half-quadratic technique of linearizing this function is equivalent to using IRLS (Geman and Yang, 1995; Charbonnier et al., 1997). The linearization of the inverse problem creates a DLS solution weighted with first order derivative flatness constraints. The flatness constraint can be thought of as a form of smoothing, and so this is referred to as a weighted, smoothed solution. The re-weighting matrices, or auxiliary variables in the half-quadratic linearization, detect the presence of edges. Once detected, the smoothing terms are turned off at these locations. In this way, the algorithm allows large variations separated by continuous planes.

To date, the standard constraints for migration/inversion schemes are a damping or smoothing term (Jin et al., 1992; Thierry et al., 1999; Duquet et al., 2000). Therefore, all tests of the EPR function will be compared against the DLS solution.

5.2 The Edge-Preserving Regularization (EPR) function

The EPR function has the same form as the modified Cauchy regularization function. Its different behavior occurs because it acts not on the model parameters, but on their derivatives

$$J(\mathbf{m}) = (\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d}) + \mu_x\,R(\mathbf{D_x m}) + \mu_z\,R(\mathbf{D_z m})\,, \qquad (5.2.1)$$

where μ_x and μ_z are the weighting parameters placed on the regularization terms in the horizontal and vertical directions. $\mathbf{D_x}$ and $\mathbf{D_z}$ are matrix forms of the derivative operators defined

$$
\begin{aligned}
(D_x m)_{ij} &= (m_{i+1,j} - m_{i,j})/\delta \\
(D_z m)_{ij} &= (m_{i,j+1} - m_{i,j})/\delta \,.
\end{aligned}
\tag{5.2.2}
$$

The indices, $i = 1, 2, \ldots M_x$ and $j = 1, 2, \ldots M_z$, encompass the total number of horizontal and vertical cells in the 2D Earth model, size M_x by M_z. Note that the constant widths of the tomographic cells are absorbed into the scaling parameter present in the derivative function. The regularization functions enforced on the derivatives are defined as

$$
\begin{aligned}
R(\mathbf{D_x m}) &= \sum_{l=1}^{M_x * M_z} \Phi([D_x m]_l) \\
R(\mathbf{D_z m}) &= \sum_{l=1}^{M_x * M_z} \Phi([D_z m]_l) \,,
\end{aligned}
\tag{5.2.3}
$$

where the 2D model parameter derivatives are placed into vector form using lexicographic notation. The model parameters within the data misfit term of the cost function are also expressed in this form. The function $\Phi(t)$ is chosen to be the modified Cauchy prior distribution

$$
\Phi(t) = \frac{t^2}{1 + t^2} \,.
\tag{5.2.4}
$$

The solution to this function found using IRLS is

$$
\widetilde{\mathbf{m}}^k = \left[\mathbf{G}^T \mathbf{G} + \mu_x \mathbf{D_x}^T \mathbf{Q_x}^{k-1} \mathbf{D_x} + \mu_z \mathbf{D_z}^T \mathbf{Q_z}^{k-1} \mathbf{D_z} \right]^{-1} \mathbf{G}^T \mathbf{d} \,,
\tag{5.2.5}
$$

where $\mathbf{Q_x}$ and $\mathbf{Q_z}$ are diagonal matrices with elements from the recovered solution at iteration $k - 1$

$$(Q_x)_{ll}^{k-1} = \frac{1}{[\,1 + (\,[D_x m]_l^{k-1})^2\,]^2}\,, \tag{5.2.6}$$

and

$$(Q_z)_{ll}^{k-1} = \frac{1}{[\,1 + (\,[D_z m]_l^{k-1})^2\,]^2}\,. \tag{5.2.7}$$

5.2.1   Half-quadratic regularization and IRLS

The solution to the non-linear cost function using IRLS intuitively makes sense. However, a more formal proof of the soundness of this type of solution has been presented using the Legendre transformation, called half-quadratic regularization (Geman and Yang, 1995; Charbonnier et al., 1997). The purpose is to introduce a new function having the same minimum as the non-linear one. This new function has the advantage that it may be solved linearly (Geman and Yang, 1995). In order to do this, a new variable is introduced, such that the regularization term is expressed

$$\Phi(t) = min_w[\Phi^*(t, w)]\,. \tag{5.2.8}$$

Subsequently, the cost function can be rewritten as

$$
\begin{aligned}
J(\mathbf{m}) \;=\; & min_{\mathbf{q_x},\mathbf{q_z}}[J^*(\mathbf{m}, \mathbf{q_x}, \mathbf{q_z})] \\
=\; & min_{\mathbf{q_x},\mathbf{q_z}}[\,(\mathbf{Gm} - \mathbf{d})^T(\mathbf{Gm} - \mathbf{d}) \\
& +\; \mu_x \sum_{l=1}^{M_x * M_z} \Phi^*((D_x m)_l, (q_x)_l) + \mu_z \sum_{i=1}^{M_x * M_z} \Phi^*((D_z m)_l, (q_z)_l)\,]
\end{aligned} \tag{5.2.9}
$$

In the 2D problem, two auxiliary variable vectors, $\mathbf{q_x}$ and $\mathbf{q_z}$, are introduced. The new regularization function Φ^* becomes quadratic with respect to $\mathbf{m}$ when

the auxiliary variables are fixed, and therefore a linear solution occurs. This is demonstrated in the next section.

5.2.2 The Legendre Transformation and auxiliary variables

The Legendre transformation is defined as

$$F^*(y) = min_x[yx - F(x)]. \qquad (5.2.10)$$

The transformation will find the y-intercept $F^*(y)$ of the line $yx - d$ that is tangent to the function $F(x)$. It will map a function from $[F(x), x]$ (function vs. x) space to $[F^*(y), y]$ (y-intercept vs. slope) space. The transform also has the property

$$F^*(y) = \mathcal{LT}[F(x)], \qquad (5.2.11)$$

and

$$F(x) = \mathcal{LT}[F^*(y)], \qquad (5.2.12)$$

where $\mathcal{LT}$ represents the Legendre transformation. Using this transform, $\Phi(t)$ and $\Phi^*(t, w)$ are defined to be a Legendre pair

$$\begin{aligned} \Phi(t) &= min_w[\Phi^*(t, w)] \\ \Phi(t) &= min_w[w\, t^2 - \Psi(w)]. \end{aligned} \qquad (5.2.13)$$

By using the definition

$$\Psi(w) = min_t[t^2\, w - \Phi(t)]\,, \qquad (5.2.14)$$

and equation 5.2.4, it is simple to calculate that

$$\Psi(w) = 2\sqrt{(w)} - w - 1\,. \qquad (5.2.15)$$

The minimum of $[t^2\, w - \Phi(t)]$ occurs at

$$w = \frac{1}{(1+t^2)^2} = \frac{\Phi'(t)}{2t}\,. \qquad (5.2.16)$$

From this it can be written that

$$\Phi^*(t, w) = [w\, t^2 - \Psi(w)]\,, \qquad (5.2.17)$$

or in terms of the previously defined variables

$$
\begin{aligned}
\Phi^*((D_x m)_l, (q_x)_l) &= [(q_x)_l\, (D_x m)_l^2 - \Psi((q_x)_l)] \\
\Phi^*((D_z m)_l, (q_z)_l) &= [(q_z)_l\, (D_z m)_l^2 - \Psi((q_z)_l)]\,.
\end{aligned} \qquad (5.2.18)
$$

The solution to the cost function containing the new regularization terms (equation 5.2.9) will become quadratic with respect to $\mathbf{m}$ when the auxiliary variables are fixed. The solution to the original cost function is easily found, because

$$
\begin{aligned}
\Phi(\mathbf{D_x m}) &= min_{\mathbf{q_x}}[\Phi^*(\mathbf{D_x m}, \mathbf{q_x})] \\
\Phi(\mathbf{D_z m}) &= min_{\mathbf{q_z}}[\Phi^*(\mathbf{D_z m}, \mathbf{q_z})]\,,
\end{aligned} \qquad (5.2.19)
$$

and the minimum values of the vectors $\mathbf{q_x}$ and $\mathbf{q_z}$ have already been defined analytically as

$$min \ (q_x)_l = \frac{\Phi'(D_x m)_l}{2(D_x m)_l} = \frac{1}{(1 + (D_x m)_l^2)^2}$$
$$min \ (q_z)_l = \frac{\Phi'(D_z m)_l}{2(D_z m)_l} = \frac{1}{(1 + (D_z m)_l^2)^2} \, . \tag{5.2.20}$$

The solution to the modified cost function will be solved iteratively: alternatively minimizing the cost function with respect to the model parameters $\mathbf{m}$ and the auxiliary variables, $\mathbf{q_x}$ and $\mathbf{q_z}$. The solution then becomes

$$\widetilde{\mathbf{m}}^k = \left[\mathbf{G}^T\mathbf{G} + \mu_x\mathbf{D_x}^T\mathbf{Q_x}^{k-1}\mathbf{D_x} + \mu_z\mathbf{D_z}^T\mathbf{Q_z}^{k-1}\mathbf{D_z} \right]^{-1} \mathbf{G}^T\mathbf{d} \, , \tag{5.2.21}$$

where the weighting matrices are defined in terms of the minimum auxiliary variables (5.2.20): $\mathbf{Q_x}^{k-1} = diag((q_x)_l^{k-1})$, and $\mathbf{Q_z}^{k-1} = diag((q_z)_l^{k-1})$. This matches the IRLS solution exactly. Therefore, the Legendre transformation is found to be a more mathematically formal way to justify the approach of an IRLS solution.

5.2.3 Edge detection and preservation using EPR

As discussed in the Chapter 4, the solution to the modified Cauchy cost function takes the form of a weighted minimum norm solution. In the case of the EPR cost function, derivative terms are introduced, and the solution is one of a DLS solution weighted by the first order derivatives to yield a flat solution. Because enforcing flatness can be considered a form of smoothing, this will be referred to as a smooth solution. The weighting matrices, $\mathbf{Q_x}$ and $\mathbf{Q_z}$, locate the presence of edges, and adjust the amount of smoothing applied to the solution. By examining equation 5.2.20, it is easily seen that if the derivative of the model

is small (no edge), the auxiliary variable will approach 1, and full smoothing is applied. Conversely, if the derivative of the model is large (an edge is present), the auxiliary variable will approach zero, and the smoothing term is turned off. In this way, the algorithm will mark the location of the edges, and adjust the amount of smoothing in order to preserve them.

Which discontinuities are preserved depends on the threshold value chosen to specify an edge as opposed to noise. This threshold value is controlled by the scaling parameter δ (see Figure 5.1). In the background areas of the model, the algorithm is able to apply homogeneous smoothing and encourage a piecewise smooth solution. This is an advantage over other sparse regularization methods, where smoothing is weak, and implemented by enforcing limited variation (Portniaguine and Zhdanov, 1999).

The EPR algorithm will first be tested on a 1D model. Assuming a zero-offset experiment, or one where the source and receiver are in the same location, the traveltime data of a horizontally homogeneous earth are measured. The modeling and migration operators are implemented as matrices in Matlab. The velocity model, source wavelet, and synthetic traveltime data are displayed in Figure 5.2. The background velocity is chosen to be a homogeneous 2000 m/s, and the acoustic perturbation is then computed to create synthetic data.

This simple experiment again reveals that multiple pairs of the parameters μ and δ will give the same value of χ^2, or data misfit (Figure 5.3 A). Because the interaction of these parameters is identical to that of the modified Cauchy function shown in Chapter 4, the discussion will not be repeated. The second part of the figure illustrates three solutions sharing the same χ^2 value. It is clear that only one solution is appropriate. Only the correct combination of parameters will yield a solution that is piecewise continuous, and so even without prior information of the model, one should be able to pick out the correct solution.
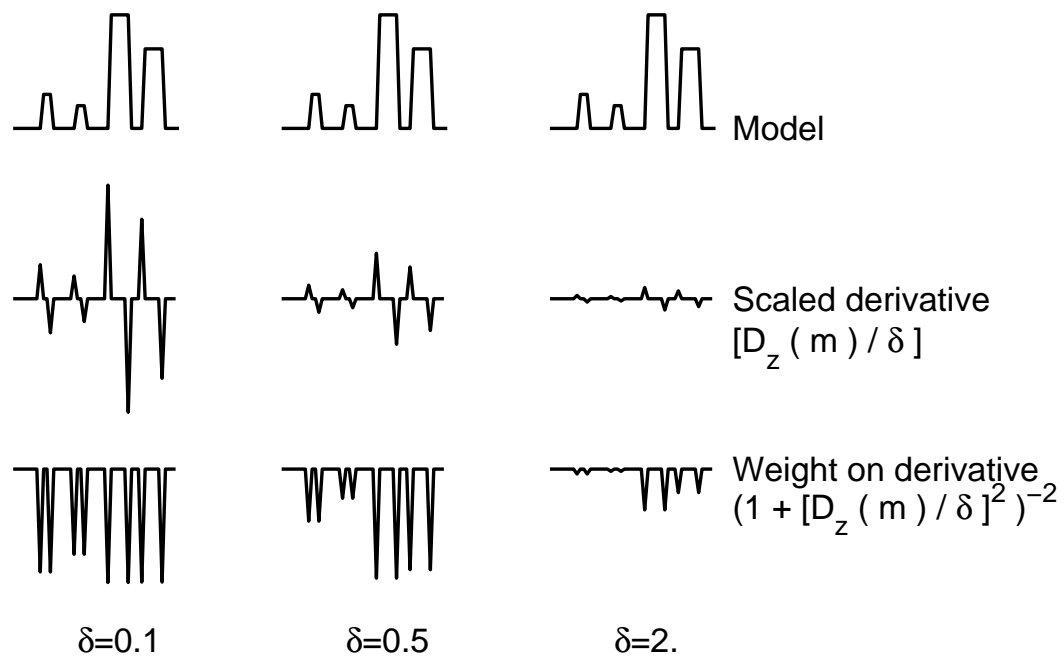
Figure 5.1: A 1D depth model, the scaled derivative, and weight on the smoothing term for three different values of the scaling parameter, δ. As δ becomes small, the derivatives are magnified, and all discontinuities are classified as edges. Smoothing at these locations is minimized, indicated by the near-zero weights. As the scaling parameter becomes large, none of the discontinuities qualify as edges, and the weights are near 1 indicating full smoothing.

Figure 5.2: The velocity model (A), source wavelet (B), and synthetic data with one percent Gaussian noise added (C).

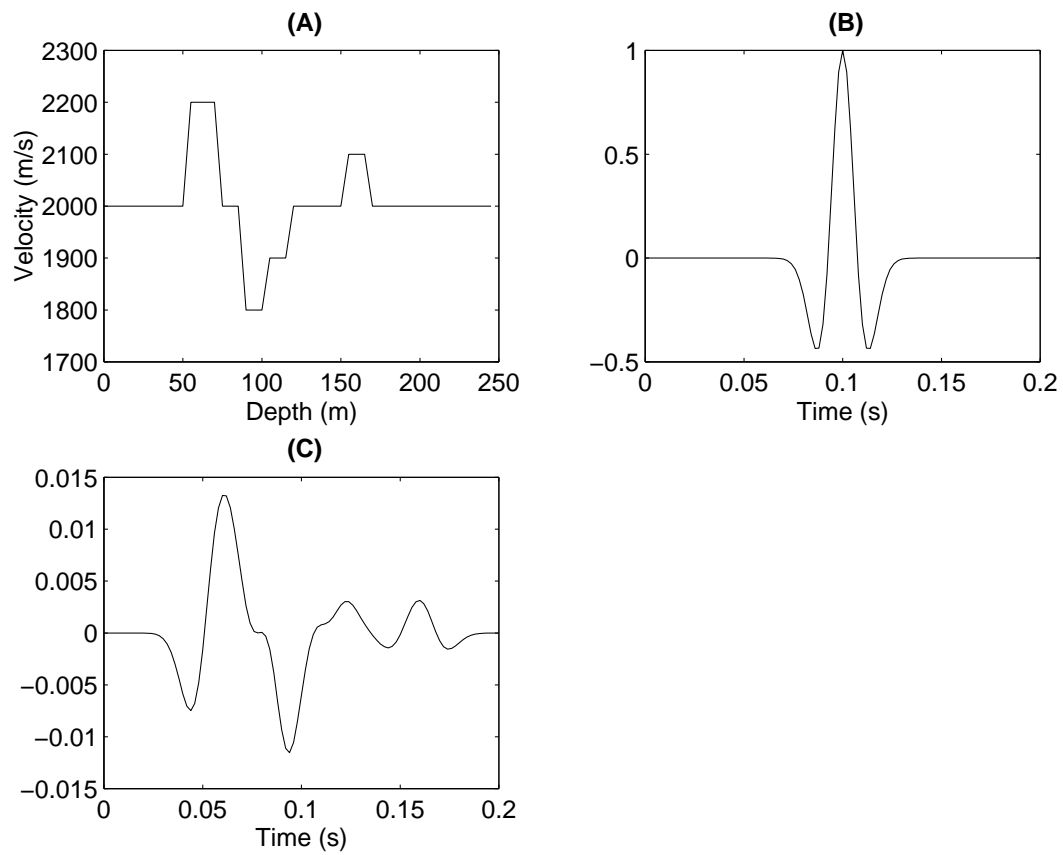Finally, Figure 5.3 C displays the effect that the scaling parameter δ has on the solution when μ is held constant. A large scaling parameter will result in small values for the scaled derivative. The weighting term will approach one, and the smoothing operator receives near full weight. Therefore, the solution of a large δ value will approach a smoothed solution. On the other hand, when the scaling parameter is small, the scaled derivative will be magnified. The weighting term will approach zero to turn off smoothing, and preserve the discontinuities. A very edgy solution is expected for the small scaling parameter. However, the solution displayed is edgy only in one area. The explanation for this is the interaction of the two parameters.
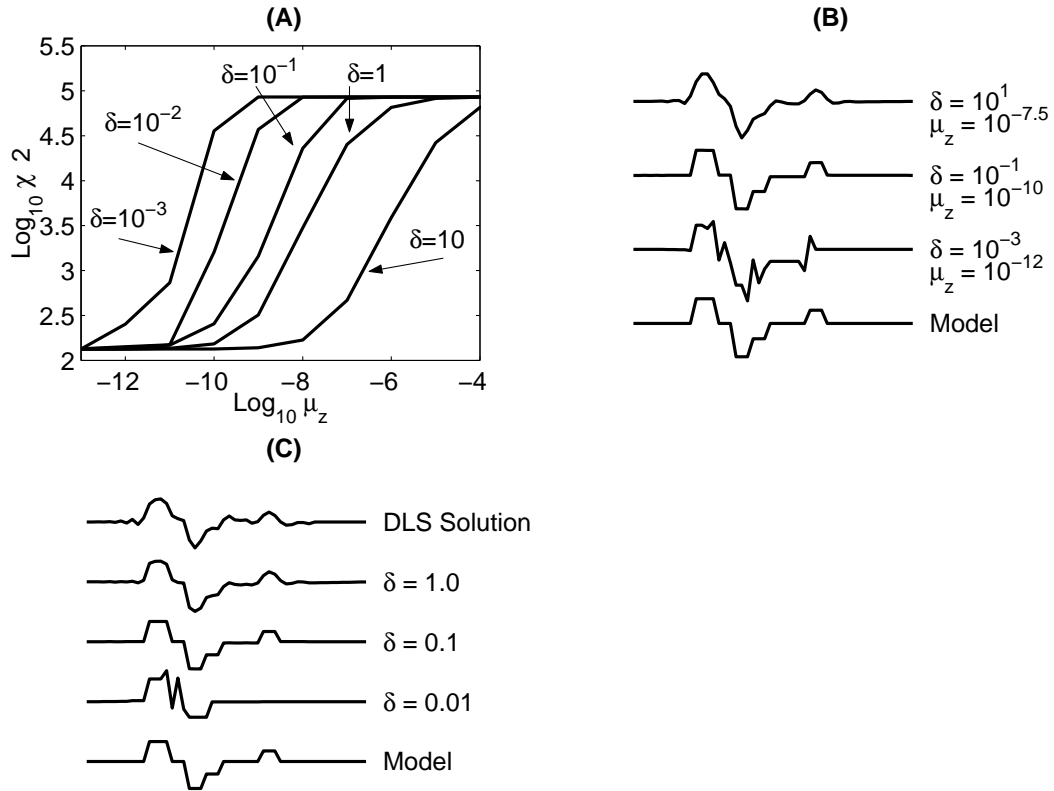
Figure 5.3: (A) A logarithmic plot of $\chi 2$ versus μ for different values of δ. (B) Solutions for different parameter pairs of δ, and μ, all sharing the optimum $\chi 2$ value for this problem. (C) The initial model, solutions for 3 values of δ, and the DLS solution, where $\mu = 10^{-10}$.

5.2.4 The interaction of the weighting and scaling parameters

The last example demonstrated that the discontinuities in the solution are not controlled solely by the scaling parameter, δ. They are also influenced by the weighting parameter, μ. The reason for the strong influence is that in the first iteration of the algorithm, no edges are detected, because the initial model is set to zeros. Therefore, full smoothing is applied to the model, and the weighting parameter for this iteration becomes a combination of the two parameters

$$\mu_z^1 = \frac{\mu_z}{\delta} \, , \tag{5.2.22}$$

where μ_z^1 denotes the vertical weighting parameter for the first iteration of the algorithm. It is apparent that a large value of μ_z, or a small value of δ will create a large weighting parameter that will enforce a very smooth solution. It is this combination of parameters that creates the initial solution that the edges are detected and classified from.

Figure 5.4 shows the progression of both the solution and the weighting matrix as the algorithm progresses for the edgy solution ($\delta = 0.01$) obtained in Figure 5.3 C. The small value of the scaling parameter combines with the weighting parameter to enforce a very smooth solution in the first iteration. Thus, when the edges are identified and preserved, there are not as many as one would expect. The progression of the algorithm for the smooth solution ($\delta = 1$) and the best solution ($\delta = 0.1$) of the same problem are shown in Figures 5.5 and 5.6.

Figure 5.4: The solution, and values of the weighting matrix (dashed line) as the EPR algorithm progresses from the second (A) to the fifth (D) iteration of the edgy solution. The weighting terms have a maximum value of one for full smoothing, and a minimum of zero to turn off smoothing. Note that the first iteration is not shown because the solution is set to zero, and the weighting terms are all unity to enforce full smoothing.

Figure 5.5: The solution, and values of the weighting matrix (dashed line) as the EPR algorithm progresses from the second (A) to the fifth (D) iteration of the smooth solution.
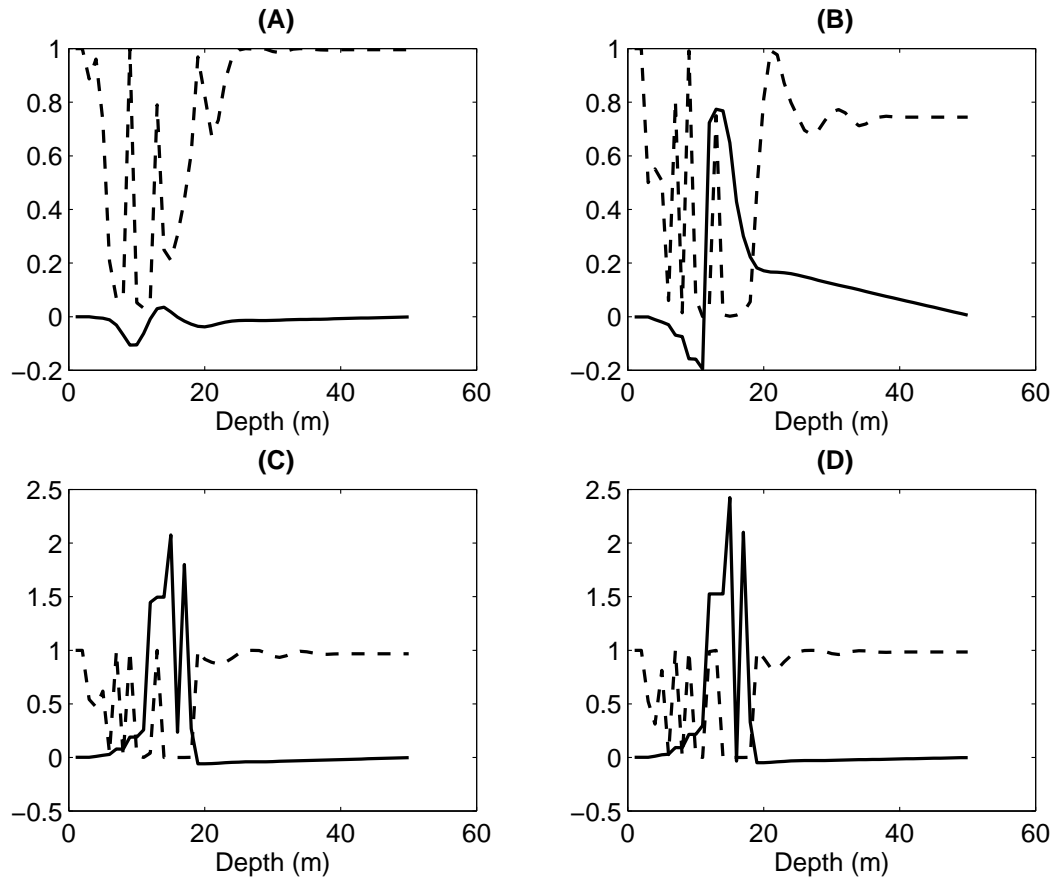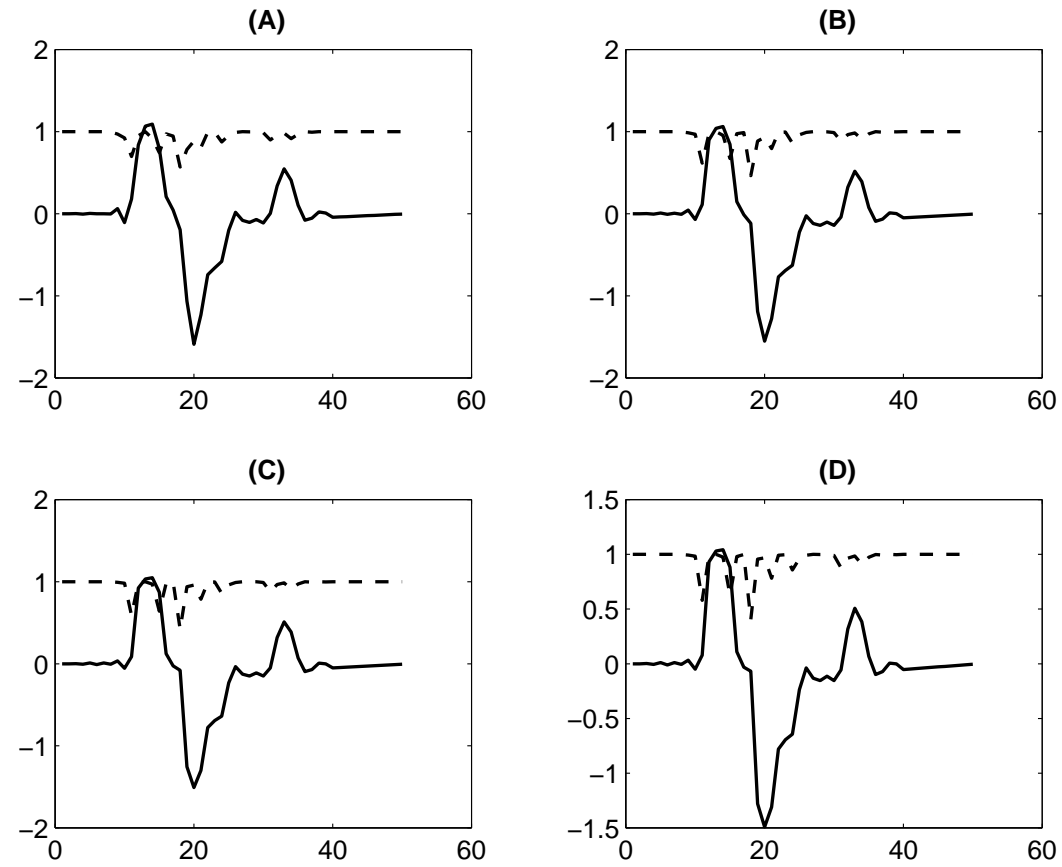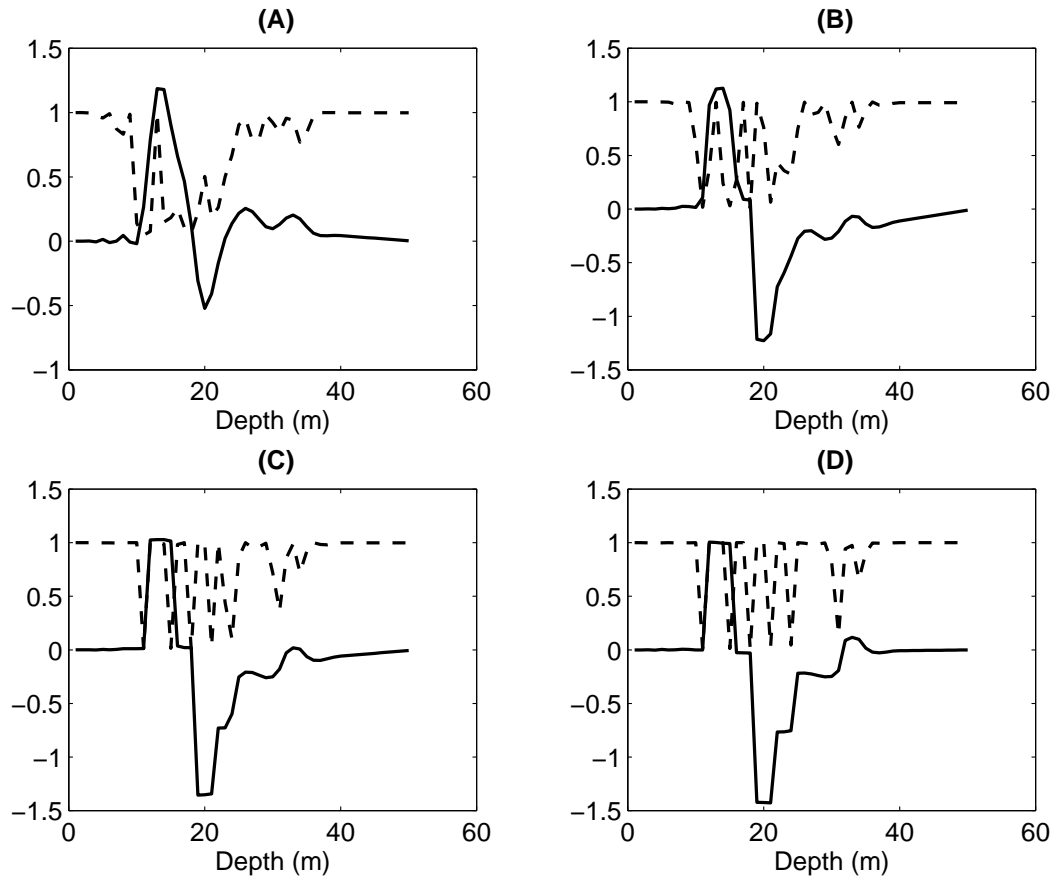
Figure 5.6: The solution, and values of the weighting matrix (dashed line) as the EPR algorithm progresses from the second (A) to the fifth (D) iteration of the best solution.

5.3 Application of EPR to synthetic data

5.3.1 2D examples

To begin testing the application of EPR to 2D models, the simple examples introduced in Chapter 2 are used. All of the examples in this section will be compared with the DLS solution, as that is the standard inverse solution in migration/inversion schemes (Jin et al., 1992; Thierry et al., 1999; Duquet et al., 2000). The geometry, source signature, and synthetic data can be found in Figures 3.6, 3.7, and 3.11. A comparison between the DLS and EPR solutions for the point scatterer and the step perturbation are seen in Figures 5.7 and 5.8 respectively. Both solutions converge within 5 iterations. Cross section of the solutions are portrayed in Figures 5.9 and 5.10. It is clear that the blocky inversion technique creates a more piecewise continuous solutions, while still being able to smooth the noise.

To further test the algorithm, a more complicated example is tested. A larger grid, and a more geologically realistic perturbation are included. Synthetic data are created using the source-receiver geometry in Figure 5.11, and 5 percent Gaussian noise is added. The DLS and EPR solutions to the inverse problem are displayed in Figure 5.12. The EPR solution converged within 5 iterations of the algorithm. Figure 5.13 shows the progression of the updated model as it is refined by the algorithm. The weighting matrices, $\mathbf{Q_x}$ and $\mathbf{Q_z}$, clearly mark the positions of the edges, and turn off the smoothing at these locations. The EPR technique recovers an almost perfect solution, the exception being made at the horizontal edges of the model. At these locations there is not enough data coverage to resolve the model, due to the source-receiver geometry.

Lastly, the minimization of the cost function is illustrated in Figure 5.14. Each minimization of the newly updated IRLS solution is achieved within 100

Figure 5.7: The acoustic perturbation (A), the DLS solution (B), and the solution using EPR (C). The parameters corresponding to these solutions are: $\mu = 10$ (B), and $\delta_x = \delta_z = 0.5$ and $\mu_x = \mu_z = 1$ (C).

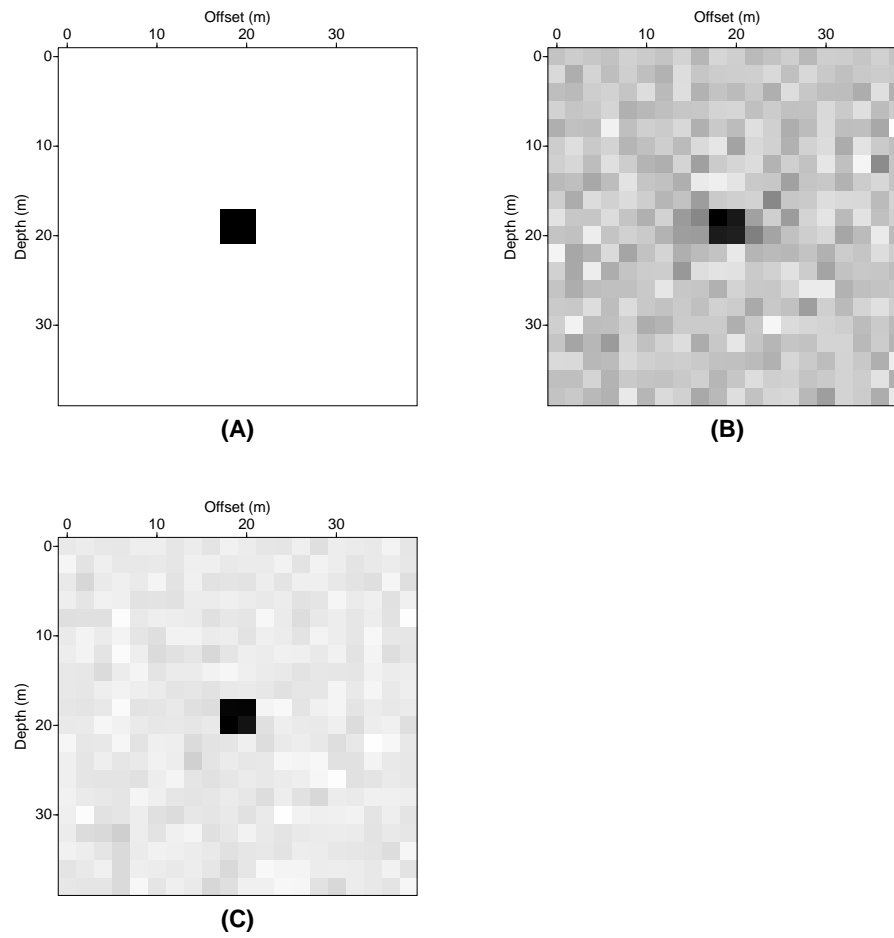Figure 5.8: The acoustic perturbation (A), the DLS solution (B), and the solution using EPR (C). The parameters corresponding to these solutions are: $\mu = 20$ (B), and $\delta_x = \delta_z = 0.1$ and $\mu_x = \mu_z = 2$ (C).
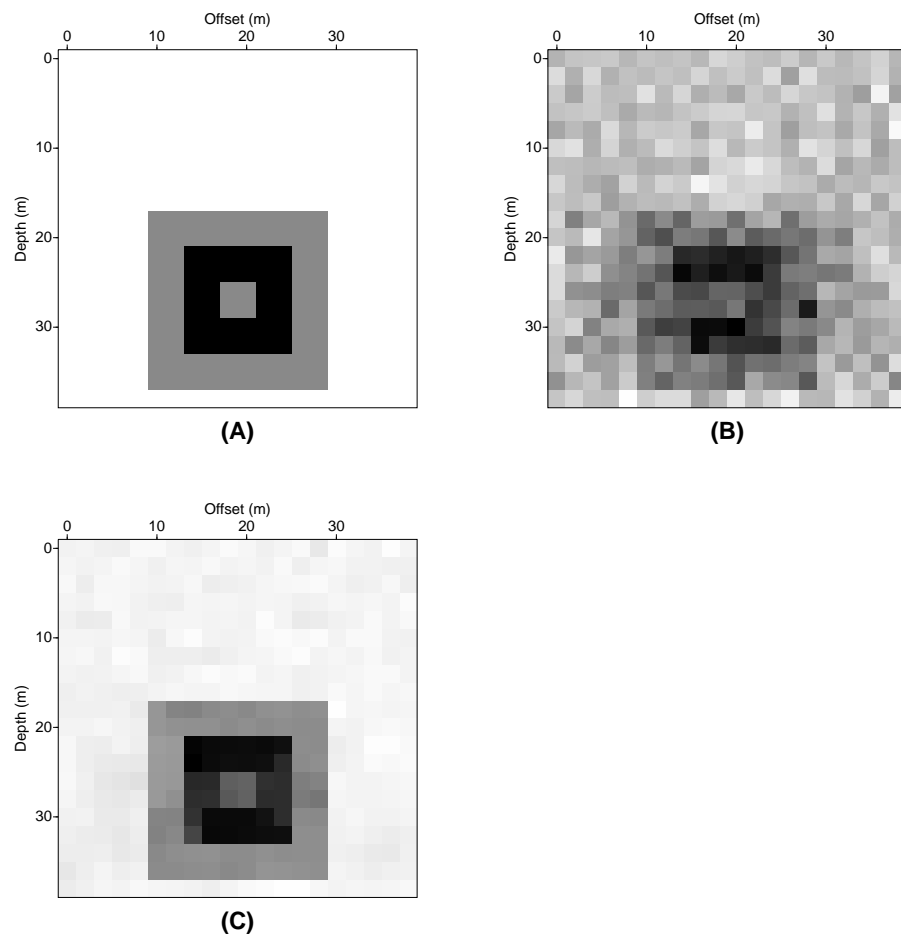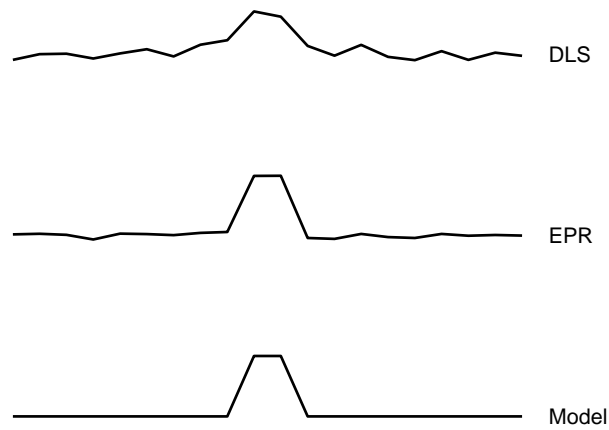
Figure 5.9: A comparison of the cross sections at 20 m depth of the two solutions in Figure 5.7.
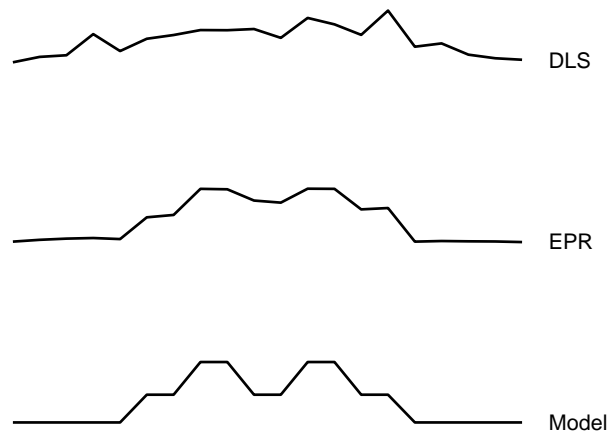
Figure 5.10: A comparison of the cross sections at 30 m depth of the two solutions in Figure 5.8.
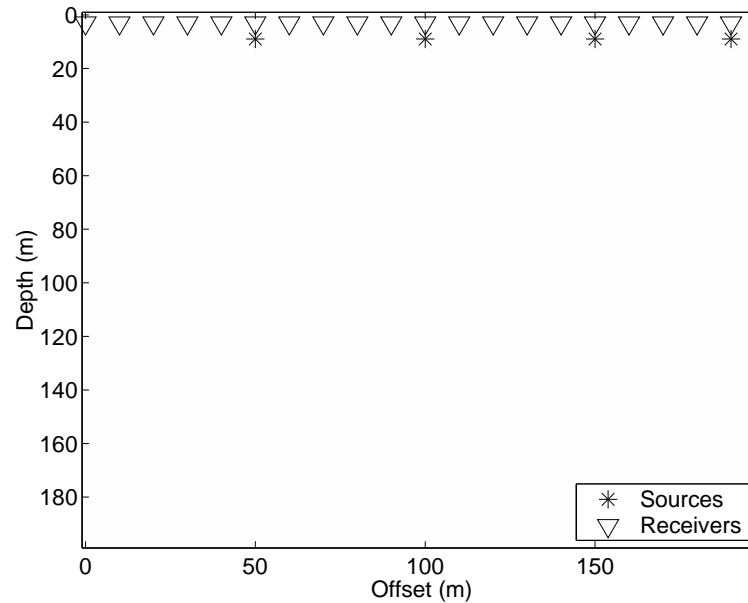
Figure 5.11: The source-receiver geometry of the dipping layer model.

iterations of the conjugate gradient algorithm. It is easily seen that these updated solutions help to accelerate the solution towards convergence. As well, the first iteration of the algorithm is the "work-horse" of the algorithm, and those after it are simply refinements on the algorithm. Because of this, it is important to select parameters that will yield an optimum solution in the first step.

5.3.2 Pitfalls to avoid

As one might suspect, finding the correct combination of two hyperparameters can involve some effort. The χ^2 test can be used to assist, but this technique is only helpful if the noise level in the data is known, or can be estimated with accuracy. As previously demonstrated, there are multiple combinations of the hyperparameters that will yield the same data misfit. Even with the assistance of the χ^2 test, a discriminating user is necessary. There are techniques of estimating the hyperparameters (Geman and Reynolds, 1992), but they are not used in this
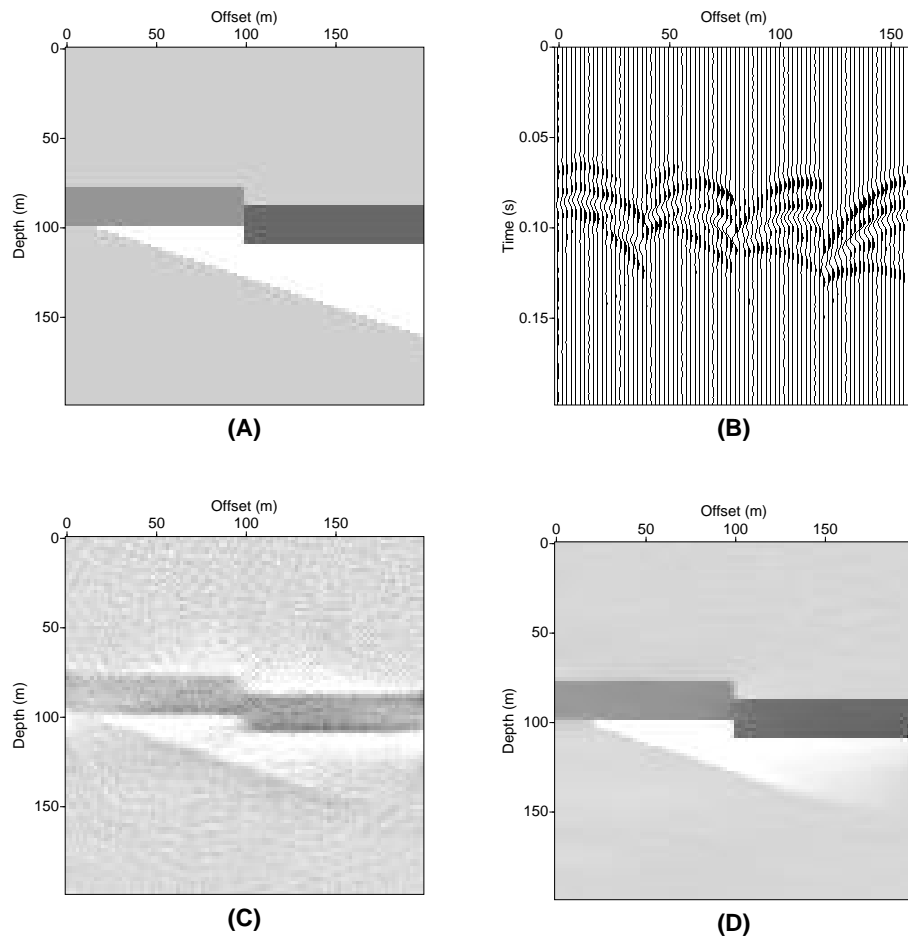
Figure 5.12: The acoustic perturbation (A), synthetic data (B), the DLS solution (C), and the EPR solution (D).The parameters corresponding to these solutions are: $\mu = 200$ (C), and $\delta_x = \delta_z = 0.35$ and $\mu_x = \mu_z = 8$ (D).
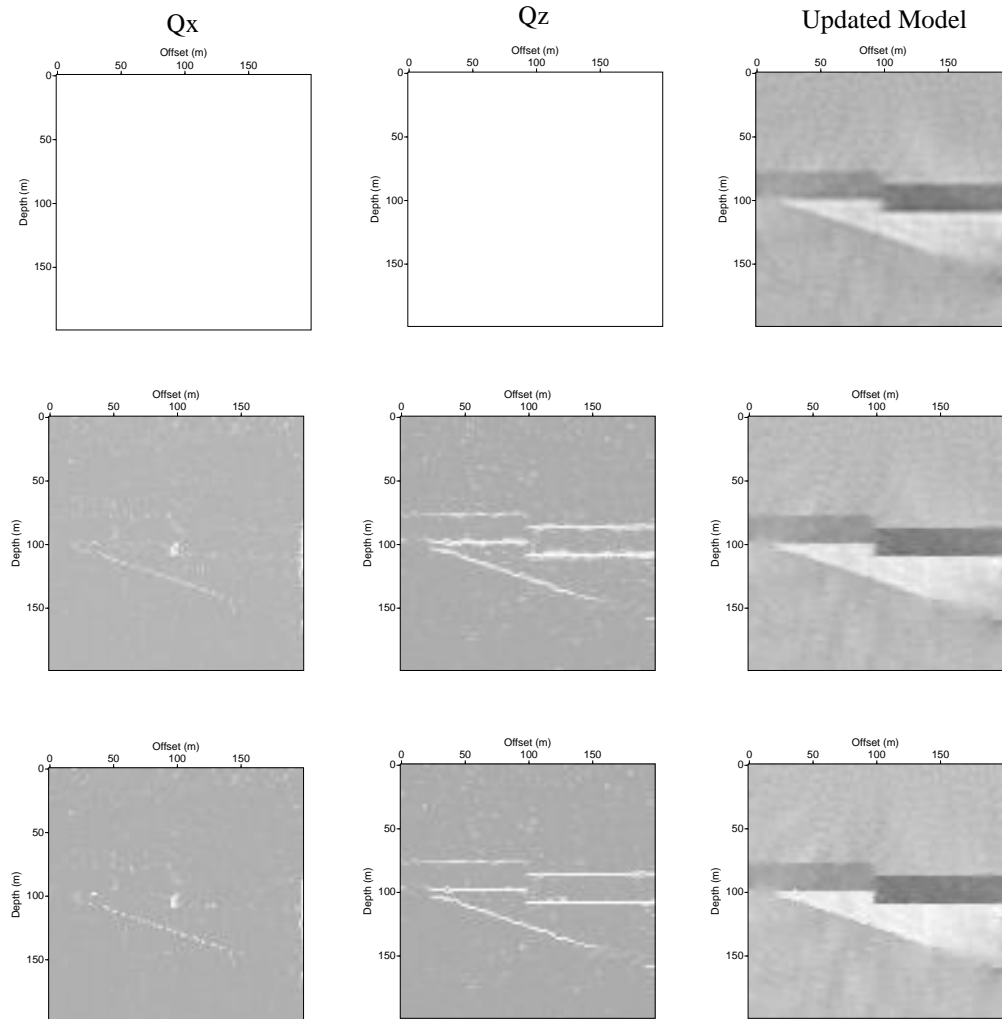
Figure 5.13: The weighting matrices $\mathbf{Q_x}$ and $\mathbf{Q_z}$, and the updated model at iterations 1, 3, and 5 of the EPR algorithm. At the first iteration, the model is homogeneous, so the weighting matrices apply full smoothing by default.
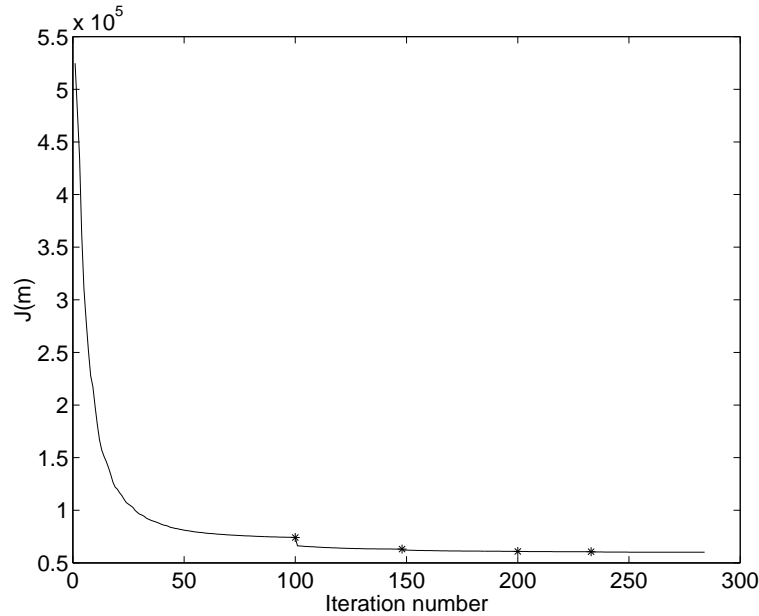
Figure 5.14: The magnitude of the cost function plotted against iteration number. The iterations at which the weighting matrices are updated are marked by stars.

study.

However, hyperparameter selection can be reasonably achieved using visual evaluation. Previous examples have shown that even though multiple pairs of parameters will give the same data misfit, only one will retrieve a piecewise continuous solution. The wrong selection of hyperparameters is immediately obvious, but viewing the updated solutions and the weighting matrices at each iteration assists in evaluating which parameter is incorrect. If the final solution is the only evaluation, it can be difficult to decide which parameter to adjust. The effects of choosing wrong hyperparameter pairs are shown in the following examples.

The first example is one where both the weighting parameter μ, and the scaling parameter δ are the wrong magnitude. Figure 5.15 displays the progression of the algorithm through 5 iterations. In the first iteration, the combination of

the two parameters does not yield a large enough weight to recover a solution
that smooths the noise. In the following iterations, the smaller scaling parameter
indiscriminately classifies the discontinuities due to noise as edges, as well as the
desired ones. This has the effect of not only preserving these noisy patches in
the solution, but amplifying them as well. The final solution, known as the noisy
solution, is obviously contaminated by noise. Here it is clear that even if the
earth model is completely unknown, this combination of parameters is incorrect.
To fix this solution, one would increase δ until only large scale edges are marked
and preserved. At the same time, μ should be increased until the first solution
is a smooth one, with as few discontinuities due to noise as possible, while still
retaining a clear image.

The second common problem in choosing the parameters is finding a scaling
parameter which will encompass all of the edges in the solution. There can
be a wide range of magnitudes of discontinuities. If the noise in the solution
creates discontinuities larger than the smaller magnitude edges, the user must
choose between a noisy solution, or one that may smooth some of the features.
Fortunately, because the first iteration is a smoothing operator, most problems
of these types are avoided. However, one must be careful to make sure the scaling
parameter will mark all of the edges. In the second example (Figure 5.16) the
final solution itself does not indicate any problems. Further scrutiny of the
weighting matrices, $\mathbf{Q_x}$ and $\mathbf{Q_z}$ reveal that only some of the edges are being
marked and preserved. This solution is referred to an the under-marked solution.
The algorithm is able to recover a fair estimate of the initial model by simply
satisfying data misfit term of the cost function, but this is not the best solution
possible.

Figure 5.17 compares the DLS solution with the three models recovered by
the EPR algorithm so far: the optimum, noisy, and under-marked solutions.

Figure 5.15: The weighting matrices $\mathbf{Q_x}$ and $\mathbf{Q_z}$, and the updated model at iterations 1, 3, and 5 of the EPR algorithm.
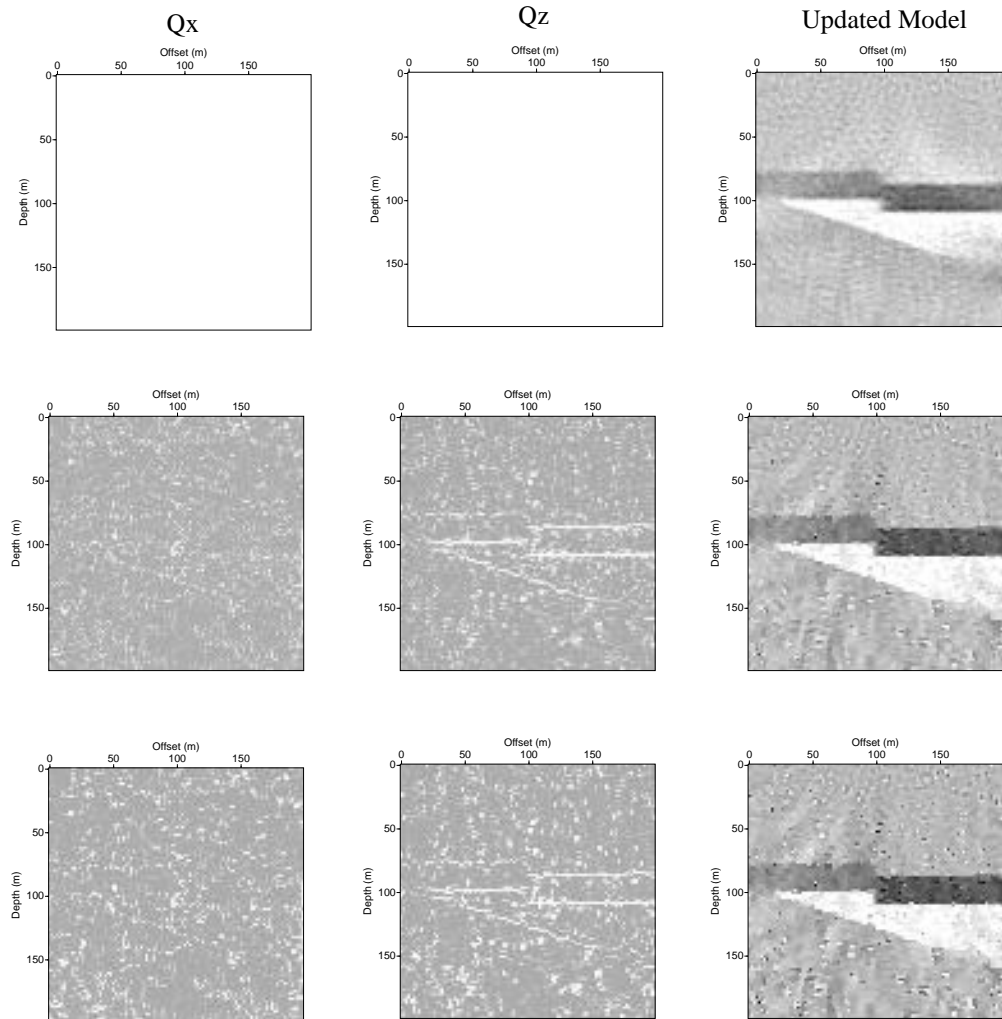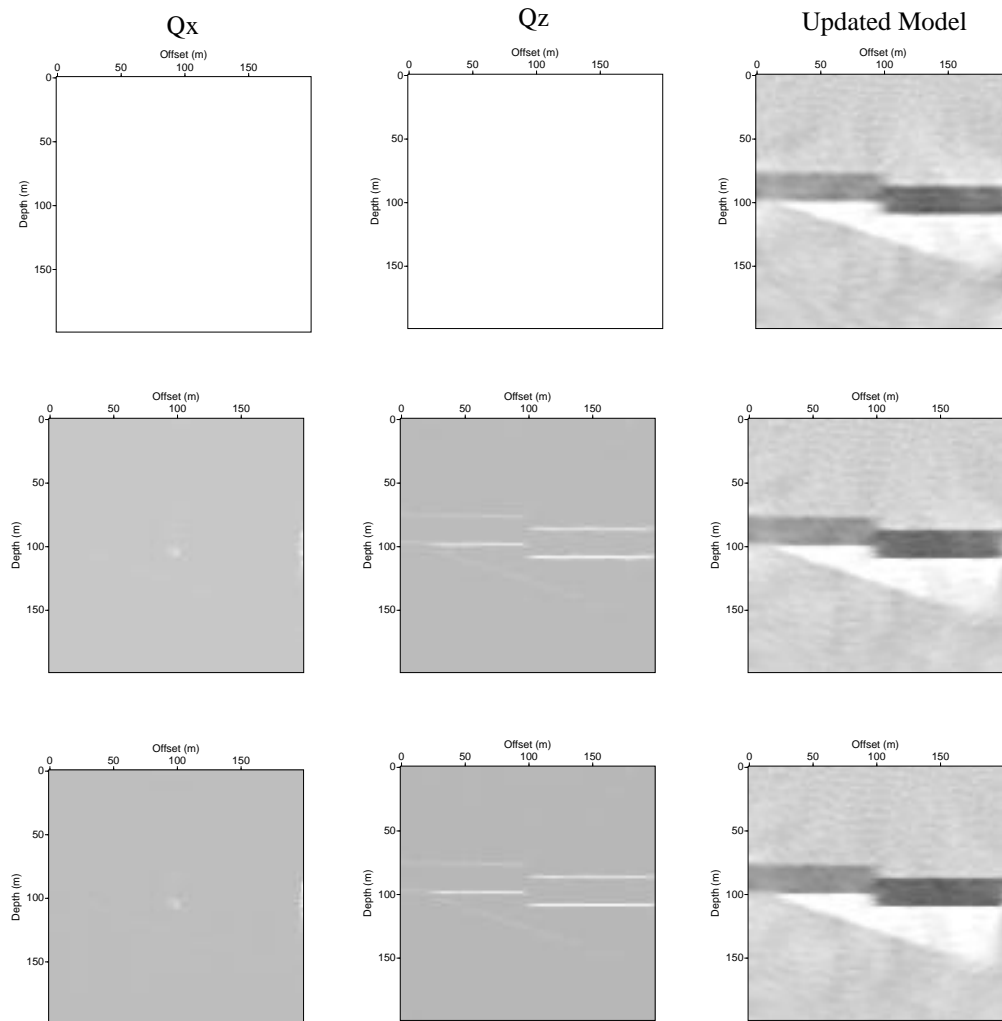
Figure 5.16: The weighting matrices $\mathbf{Q_x}$ and $\mathbf{Q_z}$, and the updated model at iterations 1, 3, and 5 of the EPR algorithm.

What is dramatically striking is that even erroneous solutions are retrieving a model comparable or superior to the DLS solution. In the case of the noisy solution, it is obvious that the parameters must be adjusted to minimize the noise. Contrary to this, the under-marked solution does not have any glaring indicators, except for the fact that the boundaries are not razor sharp. This is a subtle change, and without the comparison of a better solution, it might be unrealistic to notice. The algorithm is robust though, and the solution, even with the wrong combination of scaling parameters, is better than the DLS solution.

5.3.3   A non-homogeneous background model

Thus far, only homogeneous background models have been considered. The algorithm is easily adapted for non-homogenous background models: all that is necessary are traveltime and amplitude tables in order to calculate the background Green's functions. Traveltime tables are calculated using a simple ray-shooting algorithm implemented in Matlab. A four layer background velocity model is created, and the acoustic perturbation used in the previous examples is superimposed on top. The inversion is for the same acoustic perturbation, and this solution is added to the background velocity model to retrieve the final solution. It should be noted that in theory, a table of amplitude data should be computed as well. The approximation of constant velocity amplitude calculations is used for simplicity. This does not create a problem with synthetic data, but might if real data were to be inverted.

The velocity model, synthetic data, and its solutions are displayed in Figure 5.18. The EPR solution converged within 5 iterations of the algorithm. Again, the superiority of the algorithm for recovering high resolution images is shown.
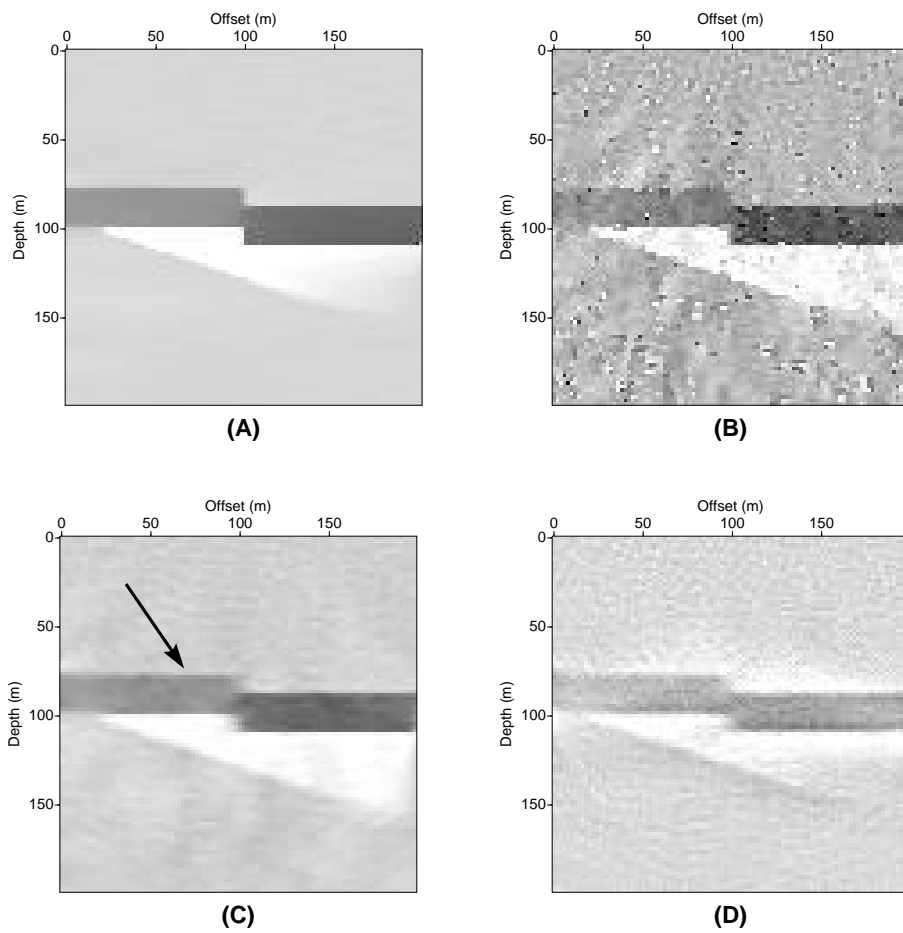
Figure 5.17: The optimum (A), noisy (B), under-marked (C) solutions of the EPR algorithm compared with the DLS solution (D). The arrow indicates the fuzzy boundary due to an incorrect scaling parameter in the under-marked solution.
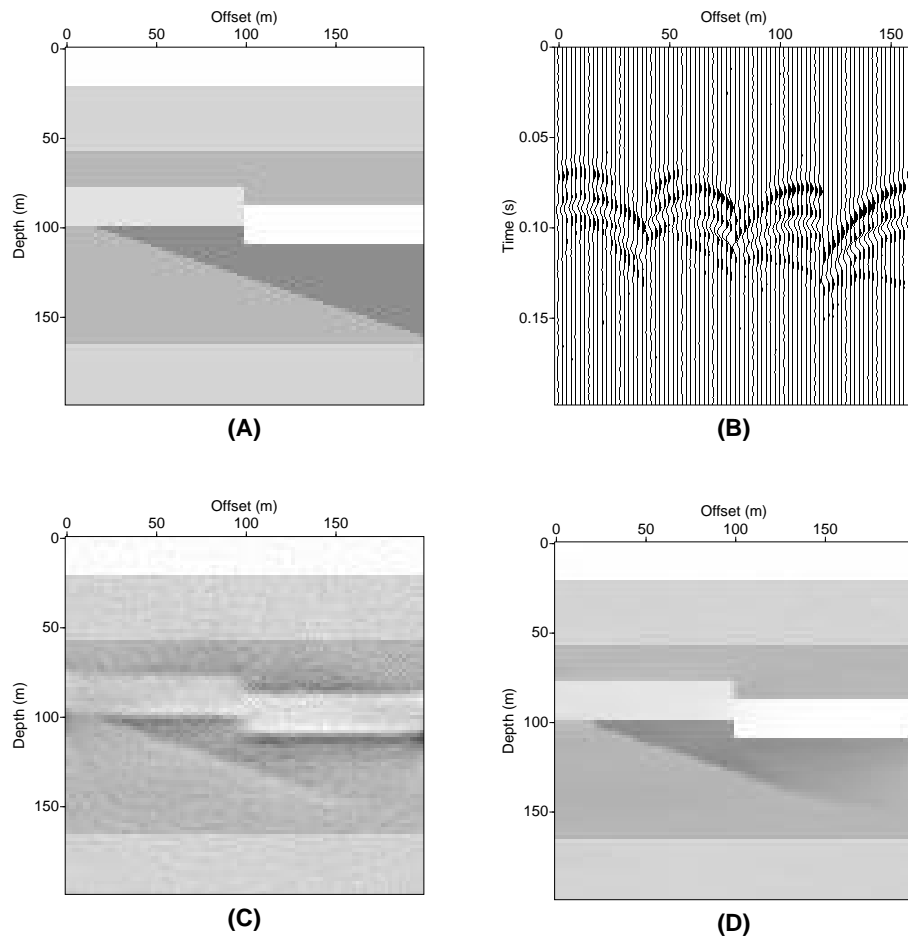
Figure 5.18: The layered velocity model (A), synthetic data (B), DLS solution (C) and EPR solution (D). The parameters corresponding to these solutions are: $\mu = 100$ (C), and $\delta_x = \delta_z = 0.5$ and $\mu_x = \mu_z = 50$ (D).

5.4   Summary

In this chapter, the EPR function has been evaluated for applications to 2D acoustic inversion. It has been shown that the half-quadratic method of linearizing the problem is equivalent to the IRLS method. This linearization creates a re-weighted flat solution. Because the flatness constraint can also be thought of a type of smoothness constraint, this is also referred to as a re-weighted smoothed solution. The re-weighting matrices detect edges that are then preserved by turning off the smoothing at appropriate locations.

The disadvantage of the EPR function is that two parameters must be chosen to create a constraint. However, it is shown that it is not difficult to visually evaluate the parameters. The algorithm should generate a piecewise continuous solution, and even if there is no knowledge of the earth model, it is obvious when the parameter combination is incorrect. In other words, the artifacts due to incorrect parameter selection are easily identified, and are very difficult to mistake as part of the earth model. Finally, even when incorrect parameter combinations are chosen, the recovered image has a higher resolution than the DLS solution.

Traditional smoothing constraints must compromise between smoothing noise and preserving edges. In contrast, the EPR algorithm can apply a large degree of smoothing without smearing the edges of the solution. In this way, the EPR function recovers high resolution images, superior to those found by traditional migration/inversion schemes.

Chapter 6

Conclusions

Within this thesis, the linear scattering problem has been posed as an inverse problem. In Chapter 2, discrete forward and adjoint modeling operators were adapted from the direct inverse solution that uses the inverse GRT. These operators were then implemented in the framework of discrete numerical inversion in Chapter 3. This combined technique is known as migration/inversion (Thierry et al., 1999). It has the advantage of being computationally efficient and the ability to solve problems containing any source-receiver geometry (Bleistein, 1987; Jin et al., 1992).

The inverse scattering problem remains ill-posed, and so constraints must be enforced to retrieve a stable and unique solution. To date, the standard constraint for migration/inversion schemes is a damping or smoothing term (Jin et al., 1992; Thierry et al., 1999; Duquet et al., 2000). This technique tends to blur the boundaries of acoustic properties within the layered earth. In contrast, a blocky inversion would preserve these edges. This type of inversion constrains the first derivatives of a function to be sparse, resulting in a piecewise continuous solution. The regularization term introduced in medical imaging by Charbonnier et al. (1997), is particularly attractive because of its ability to apply any level of

smoothing to planar areas without destroying the resolution of edges.

This regularization term is found to be very similar to the Cauchy probability distribution, and is referred to as a modified Cauchy function. In Chapter 4, it is first tested against the common sparseness constraints enforced by the Cauchy and exponential distributions in the 1D impedance inversion problem. All three prior probability functions result in quasi-linear problems. These problems may be solved in a linear, iterative manner using IRLS. It is shown that the modified Cauchy function recovers sparse reflectivity series as well as the standard constraints, but is especially suited for recovering blocky impedance profiles. The examples show that it is able to recover sharp discontinuities separated by flat, planar areas more successfully than the other prior distributions.

The modified Cauchy function is then applied to the 2D acoustic, constant density migration/inversion problem in Chapter 5. Within this context it is referred to as the Edge-Preserving Regularization (EPR) function. This name arises from the fact that in its IRLS form the weighting matrices of the algorithm detect the position and magnitude of edges. The amount of smoothing applied by the flatness constraint is then varied according to the classification of the discontinuities. In this way, the algorithm allows a solution containing large variations separated by continuous planes.

Traditional smooth constraints must compromise between smoothing noise, and preserving edges. In contrast, the EPR algorithm can apply a large degree of smoothing, without smearing the edges of the solution. In this way, the EPR function recovers high resolution images, superior to those found by traditional smooth constraints. As well, testing has shown that the EPR algorithm is robust. Even when incorrect parameter pairs are chosen, the recovered solution is higher resolution than the standard smoothed image.

The first disadvantage of EPR is that two parameters must be selected.

Proper selection depends on a good understanding of the interaction of these parameters. This occurs when not only the solution is evaluated, but also its progression in conjunction with the weighting matrices. Further, EPR is computationally more expensive than other migration/inversion schemes. The conventional minimum norm, smooth solution can be retrieved by solving the linear inverse problem once to retrieve the DLS solution. In contrast, the EPR scheme converges within $2 - 6$ iterations of a weighted, DLS solution.

This algorithm has a higher computational cost than those used to recover standard smooth solutions. It also requires high frequency data so that the asymptotic approximation to the Green's functions remains valid. For these two reasons this technique is not recommended for large scale seismic surveys. It is instead more applicable to small, detailed studies, such as cross-well and environmental imaging. In these applications the need for a precise image will justify the cost.

In conclusion, the modified Cauchy prior distribution has been found to recover very accurate solutions to both 1D impedance inversion and 2D migration/inversion problems. I recommend that further study be done on the suitability of the algorithm for seismic imaging. Specifically, the EPR method should be tested on suitably processed finite difference data, and then on real data. Automatic hyperparameter selection should also be explored. Finally, applications to other geophysical areas of EM, gravity and GPR imaging may be of interest.

References

Acar, R., and Vogel, C. R., 1994, Analysis of bounded variation penalty methods for ill-posed problems: Inverse Problems, **10**, 1217–1229.

Barone, P., 1999, Fast deconvolution by a two-step method: SIAM J. Sci. Comput., **21**, 883–899.

Beck, A. E., 1991, Physical Principles of Exploration Methods: Wuerz Pulishing Ltd.

Berkhout, A. J., 1984, Multidimensional linearized inversion and seismic migration: Geophysics, **49**, 1881–1895.

Beydoun, W., and Mendez, M., 1989, Elastic ray-Born L2 migration/inversion: Geophys. J. Int., **97**, 151–160.

Beylkin, G., and Burridge, R., 1990, Linearized inverse scattering problems in acoustics and elasticity: Wave Motion, **12**, 15–52.

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a casual generalized Radon transform: J. Math. Phys., **26**, 99–108.

Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, **52**, 931–942.

Bostock, M. G., and Rondenay, S., 1999, Migration of scattered teleseismic body waves: Geophys. J. Int., **137**, 732–746.

Box, G. E., and Tiao, G. C., 1992, Bayesian inference in statistical analysis: Addison-Wesley Pub. Co.

Charbonnier, P., Blanc-Feraud, L., Aubert, G., and Barlaud, M., 1997, Deterministic edge-preserving regularization in computed imaging: IEEE Trans. Image Proc., **6**, 298–311.

Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications.

Clippard, J. D., Christensen, D. H., and Rechtien, R. D., 1995, Composite distribution inversion applied to crosshole tomography: Geophysics, **60**, 1283–1294.

Constable, S. C., Parker, R. L., and Constable, C. G., 1987, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data: Geophysics, **52**, 289–300.

Crase, E., Pica, A., Noble, M., MacDonald, J., and Tarantola, A., 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538.

Cui, T. J., and Weng, C. C., 2000, Novel diffraction tomographic alogrithm for imaging two-dimensional targets buried under a lossy earth: IEEE Trans. Geosci. Remote Sensing, **38**, 2033–2041.

de Hoop, M. V., and Bleistein, N., 1997, Generalized Radon transform inversions for reflectivity in anisotropic elastic media: Inverse Problems, **13**, 669–690.

Djikpesse, H. A., and Tarantola, A., 1999, Multiparameter l1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms: Geophysics, **64**, 1022–1035.

Dobson, D. C., and Santosa, F., 1994, An image-enhancement technique for electrical impedance tomography: Inverse Problems, **10**, 317–334.

Duquet, B., Marfurt, K. J., and Dellinger, J. A., 2000, Kirchoff modeling, inversion for reflectivity, and subsurface illumination: Geophysics, **65**, 1195–1209.

Geman, S., and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images: IEEE Trans. Pattern Anal. Machine Intell., **6**, 721–741.

Geman, D., and Reynolds, G., 1992, Constrained restoration and the recovery of discontinuities: IEEE Trans. Pattern Anal. Machine Intell., **14**, 367–383.

Geman, D., and Yang, C., 1995, Nonlinear image recovery of half-quadratic regularization: IEEE Trans. Image Proc., **4**, 932–945.

Guillen, A., and Menichetti, V., 1984, Gravity and magnetic inversion with minimization of a specific function: Geophysics, **49**, 1354–1360.

Hestenes, M., and Steifel, E., 1952, Methods of conjugate gradients for solving linear systems: Nat. Bur. Standards J. Res., **49**, 403–436.

Jin, S., Madariaga, R., Virieux, J., and Lambare, G., 1992, Two-dimensional asymptotic iterative elastic inversion: Geophys. J. Int., **108**, 575–588.

Kirkpatrick, S., Gelatt, C. D. J., and Vecchi, M. P., 1983, Optimization by simulated annealing: Science, **220**, 671–680.

Last, B. J., and Kubik, K., 1983, Compact gravity inversion: Geophysics, **48**, 713–721.

Lay, T., and Wallace, T. C., 1995, Modern Global Seismology: Academic Press, Inc.

Margrave, G. F., 1998, Methods of Seismic Data Processing: Course lecture notes: University of Calgary.

Mast, T. D., 1999, Wideband quantitative ultrasonic imaging by time-domain diffraction tomography: J. Acoust. Soc. Am., **106**, 3061–3071.

Melamed, T., Heyman, E., and Felsen, L. B., 1999, Local spectral analysis of short-pulse excited scattering from weakly inhomogeneous media - part ii: Inverse scattering: IEEE Trans. Antennas Propagat., **47**, 1218–1227.

Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press, Inc.

Miller, D., Oristaglio, M., and Beylkin, G., 1987, A new slant on seismic imaging: Migration and integral geometry: Geophysics, **52**, 943–964.

Molina, R., Nunez, J., Cortijo, F. J., and Mateos, J., 2001, Image restoration in astronomy: A Bayesian perspective: IEEE Signal Proc. Mag., **18**, 11–29.

Oldenburg, D. W., Scheuer, T., and Levy, S., 1983, Recovery of the acoustic impedance from reflection seismograms: Geophysics, **48**, 1318–1337.

Pendrel, J., and van Riel, P., 2000, Effective well control on constrained sparse spike seismic inversion: CSEG Recorder, **25**, 18–26.

Portniaguine, O., and Zhdanov, M. S., 1999, Focusing geophysical inversion images: Geophysics, **64**, 874–887.

Press, W. H., Teukolsky, S., Vetterling, W. T., and Flannery, B., 1988, Numerical recipes in C: Cambridge University Press.

Ramm, A. J., and Katsevich, A. I., 1996, The Radon Transform and Local Tomography: CRC Press, Inc.

Sacchi, M. D., Ulrych, T. J., and Walker, C. J., 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: IEEE Trans. Signal Processing, **46**, 31–38.

Sacchi, M. D., 1997, Reweighting strategies in seismic deconvolution: Geophys. J. Int., **129**, 651–656.

Santosa, F., and Symes, W. W., 1988, Reconstruction of blocky impedance profiles from normal-incidence reflection seismograms which are band-limited and miscalibrated: Wave Motion, **10**, 209–230.

Scales, J. A., and Smith, M. L., 1994, Introductory Geophysical Inverse Theory: Samizdat Press.

Scales, J. A., 1987, Tomographic inversion via the conjugate gradient method: Geophysics, **52**, 179–185.

Snieder, R., 1998, A guided tour of mathematical physics: Samizdat Press.

Song, Z., Willianson, P. R., and Pratt, R. G., 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data: Part II-inversion method, synthetic experiments and real-data results: Geophysics, **60**, 796–809.

Strang, G., 1986, Introduction to Applied Mathematics: Wellesley-Cambridge Press.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Tarantola, A., 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier Science Publishing Company Inc.

Taylor, J. R., 1972, Scattering Theory: John Wiley and Sons, Inc.

Thierry, P., Operto, S., and Lambare, G., 1999, Fast 2-D ray+Born migration/inversion in complex media: Geophysics, **64**, 162–181.

Ulrych, T. J., Sacchi, M. D., and Woodbury, A., 2001, Tutorial: A Bayes tour of inversion: Geophysics, **66**, 55–69.

van den Berg, P. M., and Kleinman, R. E., 1995, Total variation enhanced modified gradient algorithm for profile reconstruction: Inverse Problems, **11**, L5–L10.

Vogel, C. R., and Oman, M. E., 1998, Fast, robust total variation-based reconstruction of noisy, blurred images: IEEE Trans. Image Proc., **7**, 813–824.

Walker, C., and Ulrych, T. J., 1983, Autoregressive recovery of the acoustic impedance: Geophysics, **45**, 1338–1350.

Wang, T., and Oristaglio, M. L., 2000, GPR imaging using the generalized Radon transform: Geophysics, **65**, 1553–1559.

Waters, K. H., 1978, Reflection seismology, a tool for energy resource exploration: John Wiley and Sons, Inc.

Appendix A

The scalar wave equation

This derivation follows the explanation and terminology of the second chapter of Lay and Wallace (1995).

A.1 Strain

The motions within a solid can be described by a vector field, $\mathbf{u}(\mathbf{x}, t)$. This field describes the motion of every point in a medium of a continuous distribution of particles. It is assumed that the motion undergone is a straining, or internal deformation. The deformations can be classified into two types: normal strains and shear strains. Normal strains are measures of elongation only, while shear strains include information of angular deformation.

The three dimensional strain matrix includes nine term:

$$\begin{pmatrix} \epsilon_{11} & \epsilon_{12} & \epsilon_{13} \\ \epsilon_{21} & \epsilon_{22} & \epsilon_{23} \\ \epsilon_{31} & \epsilon_{32} & \epsilon_{33} \end{pmatrix} . \tag{A.1.1}$$

The normal strains, ϵ_{11}, ϵ_{22}, and ϵ_{33} give the relative length changes in the

coordinate directions. The remaining six shear strains give the angular changes of each coordinate direction with respect to the other directions.

A.2 The relation between strain and displacement

The strain is related to displacement derivatives by the relation

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right) \tag{A.2.1}$$

$$= \frac{1}{2}(u_{i,j} + u_{j,i}), \tag{A.2.2}$$

where the final expression is in indicial notation. This relationship holds if both the strain and displacement derivatives are small. The trace of the strain tensor is called the cubic dilatation, and also relates to the displacement vector

$$\theta = \epsilon_{ii} = \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = \nabla \cdot \mathbf{u} \,. \tag{A.2.3}$$

A.3 Stress

Two types of forces can act upon a continuum: body forces and contact forces. These forces will depend either on the volume (body force) or surface area (contact force) of the material. Stress, σ, is a contact force, and can be expressed in tensor form as

$$\sigma_{ij} = lim_{\Delta A_i \to 0} \frac{\Delta F_j}{\Delta A_i} \,. \tag{A.3.1}$$

Here, F_j is a small force acting in the jth direction on a surface area A_i that has a normal in the ith direction. Combining information from all directions in the 3D system results in the stress tensor

$$
\begin{pmatrix}
\sigma_{11} & \sigma_{12} & \sigma_{13} \\
\sigma_{21} & \sigma_{22} & \sigma_{23} \\
\sigma_{31} & \sigma_{32} & \sigma_{33}
\end{pmatrix} .
\tag{A.3.2}
$$

As in the case of strain, diagonal terms describe normal stresses, and all others describe shear stresses. The equilibrium equations require a balance of spatial gradients of the stresses in a medium for that medium to be in stable equilibrium (Lay and Wallace, 1995). These equations are summarized

$$
\frac{\partial \sigma_{ij}}{\partial x_i} = 0 .
\tag{A.3.3}
$$

To maintain equilibrium, the moments of the body must also be zero. This results in a symmetric stress tensor, where

$$
\sigma_{ij} = \sigma_{ji} .
\tag{A.3.4}
$$

A.4   Equation of motion

The force per unit volume of a body of density ρ is

$$
\rho \frac{\partial^2 u_i}{\partial t^2} .
\tag{A.4.1}
$$

The force per unit volume is also equal to the combined action of the body and contact forces

$$f_i + \frac{\partial \sigma_{ij}}{\partial x_j}, \tag{A.4.2}$$

where f_i is the body force per unit volume. The contact force is defined in terms of stress, because stress is defined as force per unit area A.3.1. Combining these two definitions of force with Newton's Law yields the equation of motion

$$\rho \frac{\partial^2 u_i}{\partial t^2} = f_i + \frac{\partial \sigma_{ij}}{\partial x_j}. \tag{A.4.3}$$

The homogeneous equation of motion describes the case where body forces are not considered

$$\rho \frac{\partial^2 u_i}{\partial t^2} = \frac{\partial \sigma_{ij}}{\partial x_j}. \tag{A.4.4}$$

Hooke's law describes the relation between stress and strain, and therefore, can be used to define a relation between displacement and stress. The general form of Hooke's law is

$$\sigma_{ij} = C_{ijkl}\, \epsilon_{kl}. \tag{A.4.5}$$

The elastic moduli of the tensor C_{ijkl} describe the medium's material properties. An isotropic elastic substance has only two independent elastic moduli, called the Lame constants, λ, and μ (Lay and Wallace, 1995). These are related to the general elastic tensor

$$C_{ijkl} = \lambda\, \delta_{ij}\, \delta_{kl} + \mu\, (\, \delta_{ik}\, \delta_{jl} + \delta_{il}\, \delta_{jk}\,), \tag{A.4.6}$$

where δ represents the Kronecker delta function. Including this in Hooke's law (A.4.5) results in

$$\sigma_{ij} = \left[\lambda \, \delta_{ij} \, \delta_{kl} + \mu \, (\, \delta_{ik} \, \delta_{jl} + \delta_{il} \, \delta_{jk} \,) \, \right] \epsilon_{kl} \,. \tag{A.4.7}$$

Using the identity that $\delta_{kl}\epsilon_{kl} = \epsilon_{kk}$, this can be rewritten as

$$\sigma_{ij} = \lambda \, \epsilon_{kk} \, \delta_{ij} + 2 \, \mu \, \epsilon_{ij} = \lambda \, \theta \, \delta_{ij} + 2 \, \mu \, \epsilon_{ij} \,. \tag{A.4.8}$$

The equations of strain-displacement (A.2.2), Hooke's law (A.4.5), and the homogeneous equation of motion (A.4.4) can be combined to yield an equation of motion for an isotropic linear elastic medium not subject to body forces. Considering only the x direction in the homogeneous equation of motion

$$\rho \frac{\partial^2 u_1}{\partial t^2} = \frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{12}}{\partial x_2} + \frac{\partial \sigma_{13}}{\partial x_3} \,. \tag{A.4.9}$$

Hooke's law and the strain-displacement relations for this direction yield expressions for the strain

$$\sigma_{11} = \lambda \, \theta + 2 \, \mu \, \epsilon_{11} = \lambda \left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} \right) + 2 \, \mu \, \frac{\partial u_1}{\partial x_1} \,, \tag{A.4.10}$$

$$\sigma_{12} = 2 \, \mu \, \epsilon_{12} = \mu \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) \,, \tag{A.4.11}$$

and

$$\sigma_{13} = 2 \, \mu \, \epsilon_{13} = \mu \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right) \,. \tag{A.4.12}$$

Combining equations (A.4.10), (A.4.11), and (A.4.12), and assuming that λ and μ are constant throughout the medium yields

$$\rho\frac{\partial^2 u_1}{\partial t^2} = \lambda\frac{\partial\theta}{\partial x_1} + \mu\frac{\partial}{\partial x_1}\left(\frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3}\right) + \mu\left(\frac{\partial^2 u_1}{\partial x_1^2} + \frac{\partial^2 u_2}{\partial x_2^2} + \frac{\partial^2 u_3}{\partial x_3^2}\right).$$

(A.4.13)

Using the identities of cubic dilatation, θ, and the Laplacian, $\nabla^2 u_1$, gives

$$\rho\frac{\partial^2 u_1}{\partial t^2} = (\lambda + \mu)\frac{\partial\theta}{\partial x_1} + \mu\,\nabla^2 u_1\,. \qquad \text{(A.4.14)}$$

The same relations in the y and z directions yield

$$\rho\frac{\partial^2 u_2}{\partial t^2} = (\lambda + \mu)\frac{\partial\theta}{\partial x_2} + \mu\,\nabla^2 u_2\,, \qquad \text{(A.4.15)}$$

and

$$\rho\frac{\partial^2 u_3}{\partial t^2} = (\lambda + \mu)\frac{\partial\theta}{\partial x_3} + \mu\,\nabla^2 u_3\,. \qquad \text{(A.4.16)}$$

In vector form, these three expressions combine to create a three dimensional homogeneous equation of motion

$$\rho\frac{\partial^2\mathbf{u}}{\partial t^2} = (\lambda + \mu)\nabla(\nabla\cdot\mathbf{u}) + \mu\,\nabla^2\mathbf{u}\,. \qquad \text{(A.4.17)}$$

Using the vector equation

$$\nabla^2\mathbf{u} = \nabla(\nabla\cdot\mathbf{u}) - (\nabla\times\nabla\times\mathbf{u})\,, \qquad \text{(A.4.18)}$$

the equation of motion subsequently becomes

$$\rho\frac{\partial^2\mathbf{u}}{\partial t^2} = (\lambda + 2\,\mu)\nabla(\nabla\cdot\mathbf{u}) - (\mu\,\nabla\times\nabla\times\mathbf{u}) \qquad \text{(A.4.19)}$$

A.5   Helmholtz's theorem

Helmholtz's theorem states that any vector field $\mathbf{u}$ can be written in terms of a vector potential Ψ, and a scalar potential Φ as

$$\mathbf{u} = \nabla\Phi + \nabla \times \Psi\,, \qquad\qquad\text{(A.5.1)}$$

if

$$\nabla \times \Phi = 0\,, \qquad\qquad\text{(A.5.2)}$$

and

$$\nabla \cdot \Psi = 0\,. \qquad\qquad\text{(A.5.3)}$$

Substituting Helmholtz's theorem (A.5.1) into the vector equation of motion (A.4.19), and using the identity (A.4.18) gives

$$\nabla\left[(\lambda + 2\,\mu)\nabla^2\Phi - \rho\frac{\partial^2\Phi}{\partial t^2}\right] + \nabla \times \left[\mu\,\nabla^2\Psi - \rho\frac{\partial^2\Psi}{\partial t^2}\right] = 0\,. \qquad\text{(A.5.4)}$$

Thus, the equation will be satisfied if each bracketed term goes to zero. By letting

$$\alpha = \sqrt{\frac{\lambda + 2\,\mu}{\rho}}\,, \qquad\qquad\text{(A.5.5)}$$

and

$$\beta = \sqrt{\frac{\mu}{\rho}}\,, \qquad\qquad\text{(A.5.6)}$$

equation (A.5.4) will be solved if

$$\nabla^2 \Phi - \frac{1}{\alpha^2} \frac{\partial^2 \Phi}{\partial t^2} = 0 \,, \tag{A.5.7}$$

and

$$\nabla^2 \Psi - \frac{1}{\beta^2} \frac{\partial^2 \Psi}{\partial t^2} = 0 \,. \tag{A.5.8}$$

This yields a scalar wave equation for the potential Φ, having a velocity, α, that is known as the P wave velocity for pressure, or longitudinal waves. The vector wave equation for potential Ψ has a velocity β, that is known as the S wave velocity for shear, or transverse waves. Both of these potentials describe a wavefield, or a system of waves. The displacement due to the P and S waves are found by computing the gradient or curl of the respective potentials.

Appendix B

The reflection coefficient

Following the derivation of Lay and Wallace (1995), the incident P (longitudinal) wave is described by a 2D plane wave solution to the potential discussed in Appendix A

$$\Phi_{in} = A_{in}exp^{i\omega(px_1+\eta_1x_3-t)} \,. \tag{B.0.1}$$

Here $p = sin(i)/c$ is the horizontal slowness and $\eta_1 = cos(i)/c$ is the vertical slowness in medium 1, where i is the incident angle of the wave. The velocity c is P-wave velocity, x_1 and x_3 are the horizontal and vertical directions, and t is time.

The reflected and transmitted waves are described in the same manner

$$\Phi_{refl} = A_{refl}exp^{i\omega(px_1-\eta_1x_3-t)} \,, \tag{B.0.2}$$

and

$$\Phi_{trans} = A_{trans}exp^{i\omega(px_1+\eta_2x_3-t)} \,. \tag{B.0.3}$$

Therefore, the waves in the 2 different mediums are summarized as

$$\Phi_1 = \Phi_{in} + \Phi_{refl} , \tag{B.0.4}$$

and

$$\Phi_2 = \Phi_{trans} . \tag{B.0.5}$$

The next step is to enforce boundary conditions. Firstly, we ask that displacement be continuous across the boundary, or that at $x_3 = 0$:

$$\frac{\partial \Phi_1}{\partial x_3} = \frac{\partial \Phi_2}{\partial x_3} . \tag{B.0.6}$$

Substituting equations B.0.4 and B.0.5 into this expression results in

$$\eta_1 (A_{in} - A_{refl}) = \eta_2 A_{trans} . \tag{B.0.7}$$

The second condition we impose is that stress be continuous across the boundary or that

$$\sigma_{33}^- = \lambda \nabla \Phi + 2\mu \epsilon_{33} = \sigma_{33}^+ . \tag{B.0.8}$$

The variables λ and μ are the Lame parameters, σ represents stress, and ϵ represents strain. Because this derivation is for acoustic waves solely, it is assumed that $\mu = 0$. However, it should be noted that the final answer is valid for solids as well. Hence it is imposed that

$$\lambda_1 \nabla^2 \Phi_1 = \lambda_2 \nabla^2 \Phi_2 . \tag{B.0.9}$$

Using the definition that $\nabla^2 \Phi = -\omega^2/c^2 \, \Phi$, this becomes

$$\frac{\lambda_1}{c^2} \left(A_{in} + A_{refl} \right) = \frac{\lambda_2}{c^2} A_{trans} \, . \tag{B.0.10}$$

Next, using the definition that $\lambda_n = \rho_n \, c_n^2$ where ρ is density, and combining equations B.0.7 and B.0.10 we arrive at

$$\frac{A_{refl}}{A_{in}} = \frac{\rho_2 \eta_1 - \rho_1 \eta_2}{\rho_1 \eta_2 + \rho_2 \eta_1} \, . \tag{B.0.11}$$

This equation defines the reflection coefficient or reflectivity $\mathcal{R}$. In the vertical incidence case, $(\eta_1 = 1/c_1$ and $\eta_2 = 1/c_2)$

$$\mathcal{R} = \frac{\rho_2 c_2 - \rho_1 c_1}{\rho_1 c_1 + \rho_2 c_2} \, . \tag{B.0.12}$$

The reflection coefficient can also be written in terms of impedance, where $I = \rho c$

$$\mathcal{R} = \frac{I_2 - I_1}{I_1 + I_2} \, . \tag{B.0.13}$$

Appendix C

The convolutional operator

The convolutional of two continuous series, c, and x is defined (Margrave, 1998)

$$y(t) = \int c(\tau)\, x(t - \tau)\, d\tau\,.$$

When these series are discrete, the convolution is expressed as

$$y_j = \sum_k c_k x_{j-k}\,.$$

The example below demonstrates how easily convolution can be achieved by placing the vector c_k into a matrix operator $\mathbf{C}$

$$\begin{pmatrix} c_1 & 0 & 0 & 0 \\ c_2 & c_1 & 0 & 0 \\ 0 & c_2 & c_1 & 0 \\ 0 & 0 & c_2 & c_1 \\ 0 & 0 & 0 & \ddots \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ \vdots \end{pmatrix} = \begin{pmatrix} x_1 c_1 \\ x_1 c_2 + x_2 c_1 \\ x_2 c_2 + x_3 c_1 \\ x_3 c_2 + x_4 c_1 \\ \vdots \end{pmatrix}\,.$$

The conjugate, or adjoint of the conjugate operator is defined as $\mathbf{C}^T \mathbf{y} = \widetilde{\mathbf{x}}$. By again using the simple example we find

$$\begin{pmatrix} c_1 & c_2 & 0 & 0 & 0 \\ 0 & c_1 & c_2 & 0 & 0 \\ 0 & 0 & c_1 & c_2 & 0 \\ 0 & 0 & 0 & c_1 & c_2 \\ 0 & 0 & 0 & 0 & \ddots \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ \vdots \end{pmatrix} = \begin{pmatrix} y_1 c_1 + y_2 c_2 \\ y_2 c_1 + y_3 c_2 \\ y_3 c_1 + y_4 c_2 \\ y_4 c_1 + y_5 c_2 \\ \vdots \end{pmatrix} . \qquad \text{(C.0.4)}$$

This can be rewritten as the summation

$$x_j = \sum_k c_k y_{j+k} . \qquad \text{(C.0.5)}$$

In continuous form, the summation becomes the integral

$$x(\tau) = \int b(t)\, y(t + \tau)\, dt , \qquad \text{(C.0.6)}$$

which is the definition of cross-correlation (Margrave, 1998). Hence, the conjugate or adjoint of convolution is cross-correlation. Both the forward and adjoint discrete operators may be placed into matrix form.

Appendix D

The DLS solution

Inverse theory solutions requiring finding the minima of quadratic functions. It is necessary to take the derivatives of matrix and vector combinations. The cost or objective function of the minimum norm solution is

$$J(\mathbf{m}) = (\mathbf{d} - \mathbf{G}\mathbf{m})^T(\mathbf{d} - \mathbf{G}\mathbf{m}) + \mu\mathbf{m}^T\mathbf{m}\,. \tag{D.0.1}$$

The cost function can also be written in summation notation as

$$J(\mathbf{m}) = \sum_{i=1}^{N}\left[d_i - \sum_{j=1}^{M}G_{ij}\,m_j\right]\left[d_i - \sum_{k=1}^{M}G_{ik}\,m_k\right] + \mu\sum_{j=1}^{M}\sum_{k=1}^{M}m_j\,m_k\,,$$
$$\tag{D.0.2}$$

where j and k are dummy variables. Following the derivation of Menke (1984), the terms are multiplied out to yield

$$\begin{aligned}
J(\mathbf{m}) \;=\; & \sum_{j=1}^{M}\sum_{k=1}^{M}m_j\,m_k\sum_{i=1}^{M}G_{ij}\,G_{ik} - 2\sum_{j=1}^{M}m_j\sum_{i=1}^{N}G_{ij}\,d_i + \\
& \sum_{i=1}^{N}d_i\,d_i + \mu\sum_{j=1}^{M}\sum_{k=1}^{M}m_j\,m_k\,.
\end{aligned} \tag{D.0.3}$$

The derivative of the first term is

$$
\frac{\partial}{\partial m_q} \left[\sum_{j=1}^{M} \sum_{k=1}^{M} m_j \, m_k \sum_{i=1}^{M} G_{ij} \, G_{ik} \right] = \sum_{j=1}^{M} \sum_{k=1}^{M} \left[\delta_{jq} m_k + m_j \delta_{kq} \right] \sum_{i=1}^{N} G_{ij} \, G_{ik}
$$

$$
= 2 \sum_{k=1}^{M} m_k \sum_{i=1}^{N} G_{iq} \, G_{ik} \, . \tag{D.0.4}
$$

Note that the derivatives of the form $\partial m_i / \partial m_j$ reduce to the Kronecker delta δ_{ij}. The derivative of the second term is

$$
-2 \frac{\partial}{\partial m_q} \left[\sum_{j=1}^{M} m_j \sum_{i=1}^{N} G_{ij} \, d_i \right] = -2 \sum_{j=1}^{M} \delta_{jq} \sum_{i=1}^{N} G_{ij} \, d_i
$$

$$
= -2 \sum_{i=1}^{N} G_{iq} \, d_i \, . \tag{D.0.5}
$$

The derivative of the third term is zero, since it does not depend on the model parameters, but only on the data.

$$
\frac{\partial}{\partial m_q} \left[\sum_{i=1}^{N} d_i \, d_i \right] = 0 \, . \tag{D.0.6}
$$

Finally, the derivative of the last term, the minimum norm term is

$$
\frac{\partial}{\partial m_q} \left[\mu \sum_{j=1}^{M} \sum_{k=1}^{M} m_j \, m_k \right] = \mu \sum_{j=1}^{M} \sum_{k=1}^{M} \left[\delta_{jq} m_k + m_j \delta_{kq} \right]
$$

$$
= 2\mu \sum_{k=1}^{M} m_k \tag{D.0.7}
$$

The minimum of the cost function is found by combining all the derivatives and setting the resulting equation to zero

$$\frac{\partial}{\partial m_q} J(\mathbf{m}) \;=\; 2 \sum_{k=1}^{M} m_k \sum_{i=1}^{N} G_{iq}\, d_i - 2 \sum_{i=1}^{N} G_{iq}\, d_i + 2\mu \sum_{k=1}^{M} m_k$$

$$= \; 0\,. \tag{D.0.8}$$

When written in matrix notation, this reduces to the linear equation

$$\mathbf{G}^{T}\mathbf{G}\mathbf{m} + \mu\, \mathbf{m} - \mathbf{G}^{T}\mathbf{d} = 0\,. \tag{D.0.9}$$

The solution to this equation yields the well-known DLS solution

$$\widetilde{\mathbf{m}} = (\mathbf{G}^{T}\mathbf{G} + \mu\, I)^{-1}\mathbf{G}^{T}\mathbf{d}\,. \tag{D.0.10}$$