University of Alberta

Library Release Form

Name of Author: Mingyu Zhang

Title of Thesis: Simultaneous Inversion of Time-lapse Data

Degree: Master of Science

Year this Degree Granted: 2006

Mingyu Zhang
Department of Physics
University of Alberta
Edmonton, AB
Canada T6G 2J1

Date: _____

University of Alberta

SIMULTANEOUS INVERSION OF TIME-LAPSE DATA

by

Mingyu Zhang

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
Spring 2006

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 **Simultaneous Inversion of Time-lapse Data** submitted by Mingyu Zhang in partial fulfillment of the requirements for the degree of **Master of Science** in *Geophysics*.

_____

Dr. M. D. Sacchi (co-supervisor)

_____

Dr. D. R. Schmitt (co-supervisor)

_____

Dr. R. Chalaturnyk

_____

Dr. J. Gu (chair)

Date: _____

To my Parents

Abstract

In the past few years, time-lapse seismic technology has been a rapid developing area in the oil exploration industry. It helps us to understand fluid movements within the reservoir and optimize the production of oil fields.

Time-lapse inversion is a two-stage inversion. First stage is to estimate seismic attributes, for example impedance, velocity etc., from the seismic data. The second stage is to estimate the reservoir physical parameters such as porosity and saturation . In this thesis, we focus on the first stage. The research is divided into two categories: regularization techniques and inversion work-flows for time-lapse data.

We review the rock physical basis for the time-lapse technologies. To attenuate the noise effects, which are caused by non-repeatability, in the inversion, a structural constraint is introduced. This constraint is combined with sparseness and impedance constraints and applied to time-lapse data. Finally, we investigate the three inversion schemes for time-lapse data.

Acknowledgements

I would like to thank my supervisors, Dr. Mauricio D. Sacchi and Dr. Douglas R. Schmitt for their guidance to the world of inversion and rock physics. They opened the door of the research world for me.

I also want to thank Juefu Wang, Ulrich Theune and Yajun Zhang for sharing their experience of inversion, rock physics and programming. Thanks to other colleagues in SAIG and SHOC group for

their friendship and valuable help on my research. Thanks to the friends in physics department for the happy time we spend together. I would like to thank Tiewei He for providing the experimental data for the rock physics research in my thesis .

Finally, thanks to my family, especially my parents, for their understanding, patience and confidence to me.

Contents

List of Tables

List of Figures

List of Symbols

∇ Laplacian operator

ω temporal frequency

τ traveltime

ρ density

K bulk compression modulus

$\mathbf{L}$ modelling or forward operator

$\mathbf{L}^{-1}$ inverse operator

$\mathbf{L}^{T}$ adjoint operator

$\mathbf{m}$ the earth model

$\mathbf{d}$ seismic data

$\mathbf{n}$ additive noise

$\mathbf{D}_1$ first order derivative operator

$\mathbf{D}_2$ second order derivative operator

λ trade-offset parameter for inversion

δ delta function or scaling factor

ξ acoustic impedance

$\mathbf{W_d}$ weighting matrix for data

$\mathbf{W_m}$ weighting matrix for model

$\mathbf{C_d}$ data covariance matrix

$\mathbf{C_m}$ model covariance matrix

List of Abbreviations

CG Conjugate Gradients

SNR Signal-to-noise Ratio

FFT Fast Fourier Transform

IFFT Inverse Fast Fourier Transform

IRLS Iterative Reweighted Least-Squares

CGLS Conjugate Gradient Least Squares

SAGD Steam-Assisted Gravity Drainage

Chapter 1

Introduction

1.1 Defining the Problem

The study of a physical system can be stated as a three-step process (Tarantola, 1987): first, parameterize the system to obtain discrete model parameters; second, apply the physical laws in order to predict the measured data (forward modeling); third, use the measured data to estimate the model parameters (the inverse modeling or inversion). We can treat a seismic problem as the following mathematical problem,

$$\mathbf{d} = \mathbf{L}\mathbf{m} + \mathbf{n}, \tag{1.1}$$

where $\mathbf{d}$ denotes the observed data such as post-stack seismic data, $\mathbf{m}$ denotes the model parameters such as the reflectivity series and $\mathbf{L}$ denotes the forward modeling operator such as the convolution operator and $\mathbf{n}$ means the noise such as random noise or multiples. The conventional seismic inversion is to estimate the reflectivity $\mathbf{m}$ from the recorded seismogram $\mathbf{d}$ where $\mathbf{L}$ represents the convolution of a source wavelet with the reflectivity $\mathbf{m}$. This is the standard linear convolution model which is often used in applied seismology to process and invert data (Robinson, 1967). It is clear that when the parametrization of the earth model is made in terms of quantities such as impedance or velocity perturbations, $\mathbf{L}$ symbolizes a modeling operator like a de-migration operator (Santos et al., 2000) or an AVO forward modeling operator (Castagna, 1993). In other words, according to the type of parametrization and the data used to solve the problem $\mathbf{L}$, $\mathbf{m}$ and $\mathbf{d}$ will have different meanings. It is important to stress that in our particular study $\mathbf{L}.\mathbf{m}$ symbolizes convolution of the source wavelet with the earth reflectivity.

Therefore, our inversion will attempt to retrieve reflectivity estimates of the subsurface.

As the development of reservoir characterization and monitoring, the potential of repeated seismic acquisition in the same location at different calendar times, which is called time-lapse seismic, attracts more and more interests in the seismic industry. This time-lapse seismic problem is different from other seismic problems: two or more data sets are observed, the goal is not the model parameters but the model parameters difference. For simplicity, we assume that two data sets are acquired. This time-lapse seismic problem can be stated as,

$$\begin{aligned} \mathbf{d}_1 &= \mathbf{L}_1\mathbf{m}_1 + \mathbf{n}_1, \\ \mathbf{d}_2 &= \mathbf{L}_2\mathbf{m}_2 + \mathbf{n}_2, \end{aligned} \tag{1.2}$$

where the subscripts denote the different calendar time and $\mathbf{d}$, $\mathbf{m}$, $\mathbf{L}$ and $\mathbf{n}$ follows the definitions in equation 1.1. The time-lapse inversion is to estimate $\mathbf{m}_1 - \mathbf{m}_2$ from $\mathbf{d}_1$ and $\mathbf{d}_2$.

From the comparison of the equation 1.1 and 1.2, it may look like there is nothing new in the time-lapse inversion. However, since the seismic inversion cannot achieve an unique result, the inversion of different data sets may introduce unwanted model differences due to inconsistences of the inversion procedure. Therefore, it is important to investigate the time-lapse inversion problem separately.

1.2 Review of Time-lapse Inversion

From equation 1.2, we can find that theoretically the time-lapse inversion can be investigated from the following views: first, what kind of seismic attributes, that is $\mathbf{m}$, are inverted ; second, what kind of data, that is $\mathbf{d}$, have we inverted; third, is the approximation of the forward operator linear or nonlinear and finally how to integrate the two data sets in the inversion, that is the inversion scheme.

Many seismic attributes have been used in time-lapse research (Lines, 2002), seismic reflectivity (Winthaegen and Verschuur, 2001), seismic traveltimes, seismic acoustic impedance, AVO and $\frac{V_p}{V_s}$ ratio changes. Chambers (2002) reviewed the selection and use of seismic attributes in reservoir characterization. Kalkomey (1997) discussed the

potential risks of using seismic attributes and suggested using those seismic attributes that have a justifiable physical relationship with the reservoir properties. Cooke (1999) investigated post-stack attributes and concluded that absolute impedance is the best attribute but relative impedance is more practical for reservoir characterization. Lewis (1997) proposed a workflow for modeling and analyzing seismic attributes for reservoir monitoring. Galikeev (2004) proposed the use of the reservoir-based seismic attributes such as volume frequency decomposition to delineate the time-lapse changes and discussed whether the a difference of attributes or an attributes of a difference should be studied.

In the time-lapse case, the most often used forward operators are: Born operator (Rickett et al., 1996; Biondi et al., 1998; Abubakar et al., 2001), Kirchhoff operator (Lumley and Beydoun, 1994) and convolution operator (Sarkar et al., 2003). Common approaches to time-lapse inversion are sparse spike inversion (Herawati, 2002; Malaver, 2004) and model-based inversion (Herawati, 2002). Galikeev (2004) proposed a method called Peseduo-Geostatistical Cascaded Inversion to increase the resolution of the estimated impedance. Abubakar (2003) described the use of Multiplicative Regularized Contrast Source Inversion (MR-CSI) method for time-lapse inversion. Tennebo (1998) proposed an inversion algorithm based on a technique called Best Feasible Approximation. Sarkar (2003) compared three time-lapse inversion schemes which he called uncoupled inversion, coupled inversion and inversion of difference and concluded the inversion of difference is often more stable. In his thesis, the difference of the uncoupled inversion and coupled inversion is the latter one use the estimated model of the base survey as the initial model for the estimation of the monitor inversion which is similar to the scheme proposed by Tennebo (1998). However, Abubakar (2001) concluded that the inversion of difference has a serious limitation and inverting the base and monitor survey separately is more preferred.

Besides the research mentioned above, much research have been done to estimate the reservoir parameters such as porosity, pressure and saturation from seismic properties such as impedance, reflectivity. We omit the review of this part because the main focus of this thesis is on the first stage of the time-lapse inversion.

1.3 Scope of the Thesis

This thesis discusses the rock physics basis for the time-lapse seismic technology and the inversion methods for time-lapse data. The scope of the thesis is the study of the forward modeling of time-lapse data and the corresponding inversion scheme. Structural, sparseness and impedance constraint are combined to regularize the inversion. The inversion scheme is tested with the synthetic data to get a better control of this method.

1.4 Thesis Outline

The structure of the thesis is as follows:

- In chapter 1 I first define the time-lapse inversion problem and provide the motivation to investigate this problem. Then I briefly review other people's work. Finally, the scope of this thesis is provided.
- In chapter 2 rock physics basis for the time-lapse research is reviewed. The methods which relate the rock parameters to the seismic data are presented in detail.
- In chapter 3 I review the basics of geophysical inverse theory. Regularization techniques and several commonly used numerical methods for inversion are the main contents.
- In chapter 4 sparseness constraints, impedance constraint and structural constraints for time-lapse inversion are investigated.
- In chapter 5 I focus on the three inversion schemes and their comparison. The advantages and possible limitations are discussed.
- In chapter 6 I summarize my research and provide a discussion of the most relevant findings. I will also discuss possible future research directions. Finally, I will discuss the advantages and weaknesses of the simultaneous inversion scheme.

Chapter 2

Rock Physics

Travel-time and reflection amplitudes of seismic data provide information about subsurface. Seismic data is often used to image these structure that might trap hydrocarbons. In recent years, the potential of seismic data to get lithology information of the reservoirs have been explored. The basis for this advancement is the rock physics which relates seismic data, reservoir parameters and properties.

This chapter consists of three parts: reservoir parameters, fluid substitution and rock physics for time-lapse case.

2.1 Reservoir Parameters

Many factors such as pressure, temperature, saturation, fluid type, porosity, porosity type, mineralogical composition, permeability and effective stress can affect seismic properties such as velocity and density. These factors are usually called reservoir parameters. The effects of these factors are very complex and often not independent.

To relate the seismic properties to the reservoir parameters, several steps are needed: first, the properties of various constituent minerals and pore fluids are calculated through the empirical equations provided, for instance by Batzle (1992); second, the elastic properties of composite fluid and composite rock are calculated by effective media theory and empirical methods; third, the elastic properties of the saturated porous medium are calculated by fluid substitution theory; and finally, we calculate the P-wave and S-wave velocity from the elastic properties of the composite material (Lay and Wallace, 1995). In this section, we will focus on the first two steps.

2.1.1 Pore Fluid and Rock Properties

To estimate the elastic properties of the saturated rock, we need to know the following concepts which are summarized by Mavko et al. (1998): the volume fractions of the various phases, the elastic properties of each phase and the geometric details of how the phases are arranged to each other. The various phases refer to the rock components which consist of the minerals and the pore fluids. The volume fractions mean the porosity of the rock and fluids and the saturation of each phase of the fluids. The geometric details are usually unknown. Therefore, only upper and lower bounds of the elastic properties of the composite are estimated. The properties of the minerals, which make up the rock, and the pore fluid are usually measured in laboratory and are related to the reservoir parameters by empirical equations. Usually, people measure the density, P-wave velocity and S-wave velocity, which are often called seismic properties, instead of measuring the elastic properties directly. Batzle (1992) summarized seismic properties of the common pore fluids: hydrocarbon gases, oil, brine. As we can see from the empirical equations, the most important parameters involved in the calculation of the seismic properties of pore fluids are temperature, pressure, porosity and saturation. Once we have the seismic properties of the various phases of the saturated rock, we can calculate the elastic moduli by the following formula,

$$K = \rho(V_P^2 - \frac{4}{3}V_S^2), \tag{2.1}$$

$$\mu = \rho V_S^2. \tag{2.2}$$

2.1.2 Elastic properties of the Composite Material

The density of the composite material ρ is determined by the weighted average sum,

$$\rho = (1 - \phi)\rho_s + \phi\rho_f, \tag{2.3}$$

where ϕ denotes the porosity and ρ_f and ρ_s are calculated by

$$\begin{aligned} \rho_f &= \sum_i S_i \rho_i, \\ \rho_s &= \sum_j f_j \rho_j, \end{aligned} \tag{2.4}$$

where S_i denotes the saturation of the ith phase of the fluids and ρ_i denotes the corresponding density. f_j is the volume fraction of the ith component of the solid rock and ρ_j is its density.

The bulk modulus of the mixture of the fluids are calculated by Wood's formula,

$$\frac{1}{K_f} = \sum_i \frac{S_i}{K_{f_i}}, \tag{2.5}$$

here S_i is the saturation of the ith fluid component and K_{f_i} is its bulk modulus.

Several theoretical models exist to predict the bounds of the saturated rocks. The Voigt (1910) and Reuss (1929) models are the simplest ones. The Hashin-Shtrikman (1963) model provides the best bounds. The Voigt model provides the upper bound of the elastic modulus M :

$$M = \sum_i^N f_i M_i \tag{2.6}$$

where f_i is the volume fraction of the ith component, which can be mineral or pore fluid, of the saturated rock. The parameter M_i is its elastic modulus which can be bulk modulus K or shear modulus μ. The parameter N means the number of the components of the saturated rock. This bound is referred to as the isostrain average. The Reuss model provides the lower bound of the elastic modulus,

$$\frac{1}{M} = \sum_i^N \frac{f_i}{M_i}, \tag{2.7}$$

this bound is referred to as the isostress average. The mathematical average of the Voigt and Reuss bounds is called Voigt-Reuss-Hill model (1952). This average which has no physical meaning is useful when we want to estimate the elastic moduli instead of a range. This idea can be extended to the Hashin-Shtrikman model and get an estimated elastic moduli. The Hashin-Shtrikman bound is the narrowest possible range without knowing any geometric details of the saturated rock,

$$
\begin{aligned}
K^{HS\pm} &= K_1 + \frac{f_2}{(K_2 - K_1)^{-1} + f_1(K_1 + \frac{4}{3}\mu_1)^{-1}}, \\
\mu^{HS\pm} &= \mu_1 + \frac{f_2}{(\mu_2 - \mu_1)^{-1} + \frac{2f_1(K_1+2\mu_1)}{5\mu_1(K_1+\frac{4}{3}\mu_1)}},
\end{aligned} \tag{2.8}
$$

where K, μ denote the bulk and shear modulus respectively. The parameter f means the volume fraction. Number $1, 2$ denote $1st$ and $2nd$ phase of the material. The upper and lower bounds are calculated by interchanging the phase 1 and 2. The $+$ means the upper bound and $-$ means the lower bound.

2.2 Fluid Substitution

When various fluids scenarios are provided, the seismic attributes may be quite different. The approach which relates the fluid scenario to the seismic attributes is called fluid substitution . In this sense, fluid substitution is part of the seismic attribute work.

A variety models and theories exist for fluid substitutions. However the most commonly used approach is Gassmann's equation (Gassmann, 1951). In this section, we will focus on the Gassmann's equation after reviewing the various existing models and theories.

2.2.1 Review of Fluid Substitutions

The work on fluid substitutions can be categorized as two groups: the empirical approaches and the theoretical approaches.

The common empirical approaches are Wyllie's (1956) time average equation and Raymer-Hunt-Gardner (1980) relations . Other empirical approaches are: Geertsma's (1961) empirical relations; Nur's modified Voigt average (Nur et al., 1991; Nur et al., 1995). The disadvantages of these approaches are: first, only porosity are taken into consideration; second, they cannot be justified theoretically.

The main theoretical approaches are: Biot theory (Biot, 1956b; Biot, 1956a), Gassmann's (1951) equation, Brown and Korringa's (1975) relations, Mavko-Jizba (1991) Squirt theory and Biot-squirt model (Dvorkin and Nur, 1993; Dvorkin et al., 1994). These theoretical approaches are designed for various scenarios, for example considering the anisotropy of

the rocks. However, the Gassmann's equation is most often used because of its simplicity and accuracy for the seismic frequency bandwidth.

Other factors which affect the elastic moduli are: porosity geometry (Johansen et al., 2002) and pore fluid distributions (Packwood, 1997), clay content, overburden pressure and textures of rocks, etc. (Wang, 2001). These effects can be important, therefore, the relations between seismic attributes and reservoir parameters are complicated in practice. For theoretical studies, we can mainly focus on the most commonly used approach — Gassmann's equation.

2.2.2 Gassmann Fluid Substitution

Gassmann's equation relates the bulk modulus of the saturated rock ,K_{sat}, to the bulk modulus of the rock matrix frame, K_d, the bulk modulus of the solid material, K_S, the bulk modulus of the fluid, K_f, and the porosity, ϕ. It assumes the shear modulus of the saturated rock does not depend on the pore fluids,

$$\begin{aligned} K_{sat} &= K_d + \frac{(1 - K_d/K_s)^2}{(1 - K_d/K_s - \phi)/K_s + \phi/K_f}, \\ \mu_{sat} &= \mu_d, \end{aligned} \tag{2.9}$$

where μ_{sat} is the shear modulus of the saturated rock and μ_d denotes the shear modulus of the rock matrix frame. Comparing to the empirical models such as Voigt and Reuss models, Gassmann's equation considers the effect of the stiffness of the rock frame besides the effect of the porosity. Wang (2001) summarized the assumptions for the Gassmann's equation. The most important requirements implied in these assumptions are: source wave function should be low-frequency and the pore connectivity should be high.

From equation 2.9, the input parameters for the Gassmann's equation are: porosity of the rock; elastic properties of the pore fluids and solid material; and elastic properties of the rock matrix frame. The porosity can be calculated from core samples or log data. The process to calculate the elastic moduli of the pore fluids and solid are described in the first section of this chapter. The elastic moduli of the various phases are usually obtained from the core sample, log data and laboratory measurements. Other approaches to get pore fluids properties include: measuring the pore fluids directly (at the temperatures and

pressures of the reservoir from which the pore fluids are obtained) and calculating the properties from the equations of the state. The elastic moduli of the rock matrix frame are hard to measure directly. There are three indirect approaches: first, measure the velocity of the dry rock (the dry rock for the laboratory measurements should be slightly wet or moist instead of too dry) and then use the equation 2.1 and 2.2 to calculate the elastic moduli of the rock matrix frame; second, apply the empirical relationships or effective medium theory; and third, calculated from log data. The last approach contains mainly two steps: the first step is to get saturated bulk modulus of the rock, K_{sat} using the equation 2.1 and K_s, K_f and ϕ using the process shown above. Second step is to rewrite the Gassmann's equation in the following form,

$$K_d = \frac{K_{sat}(\phi K_S/K_f + 1 - \phi) - K_S}{\phi K_s/K_f + K_{sat}/K_S - 1 - \phi}, \tag{2.10}$$

then we substitute the output parameters of the first step into the right side of the equation 2.10 and calculate the bulk modulus of the rock matrix frame. This equation also shows that the bulk modulus of the rock matrix frame partly depends on porosity. Therefore, it is incorrect to use Gassmann's equation without adjusting the value of the K_d for different porosity. After that, we can estimate the P-wave and S-wave velocity using the following equations, which are the backward form of the equations 2.1 and 2.2 ,

$$\begin{aligned} V_P &= \sqrt{\frac{K_{sat} + 4/3\mu_{sat}}{\rho}}, \\ V_s &= \sqrt{\frac{\mu_{sat}}{\rho}}, \end{aligned} \tag{2.11}$$

where ρ is calculated using the equation 2.3.

2.3 Rock Physics for Time-lapse Case

In time-lapse case, the volume fractions of the fluids, the temperatures and pressures in a reservoir are changed due to the reservoir production. For example, for a field using steam-assisted gravity drainage (Butler, 1994), hot steam is injected into the reservoir from a horizontal injection well. After a chamber of hot steam is formed in the reservoir, the temperature of the material adjacent to this chamber increases due to the conduction

of heat. With the production of oil, the hot steam replaces the oil in the space and gradually cools down. On the other hand, we assume that the background lithology does not change during the time-lapse acquisition process. Therefore, pressure, temperature and fluid saturation are the main changing factors in time-lapse seismics if we assume constant porosity.

From the viewpoint of Gassmann's equation, the seismic properties of the reservoir changes in time-lapse case are mainly determined by stiffness of the rock matrix frame, pore fluids properties and fluid saturation. In this sense, we can treat the effects of temperature, pore pressure on seismic properties by considering their effects on the rock matrix frame and pore fluid properties.

2.3.1 Pressure, Temperature and Saturation

There are two kinds of pressure in a reservoir: overburden pressure P_c (also called confining pressure) and pore pressure P_p (also called fluid pressure or reservoir pressure). The overburden pressure is the pressure due to the overburden rock. This pressure is usually assumed constant in time-lapse case. The reservoir pressure is the pressure due to the fluid mass. This pressure is changed during the time-lapse acquisition process. The difference between the two pressures P_d is called net overburden pressure or differential pressure. The net overburden pressure is sometimes called effective pressure P_e, but the two pressures are different in the strict definition (Wang, 2001). The definitions for P_d and P_e are,

$$\begin{aligned} P_d &= P_c - P_p, \\ P_e &= P_c - nP_p, \end{aligned} \tag{2.12}$$

where n is less than 1. The net overburden pressure affects the stiffness of the rock matrix frame and thus controls the seismic properties of the reservoir instead of the overburden or pore pressure alone. However, since the overburden pressure is assumed constant in time-lapse case, we can consider only pore pressure effects in time-lapse case. According to the definition of the net overburden pressure, we can find that the change of the pore pressure will cause the opposite change of the net overburden pressure. The effects of the pressure on the rock matrix frame is illustrated in Figure 2.1.

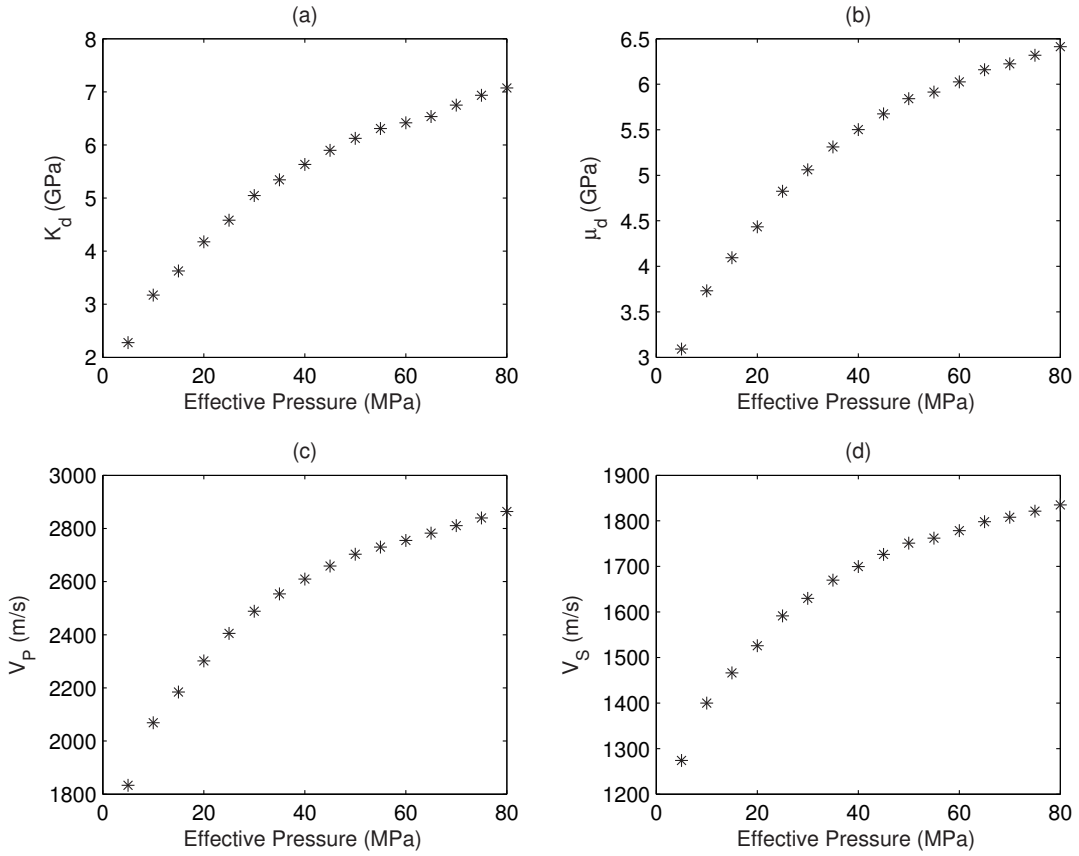

Figure 2.1: The effects of effective pressure on: (a) the bulk modulus of the rock matrix frame, (b) the shear modulus of the rock matrix frame, (c) the P wave velocity of the saturated rock and (d) the S wave velocity of the saturated rock. [Data Coutersy of Tiewei He and Dr. Doug Schmitt]

Besides the pressure effects on the rock matrix frame, the pressure also affect the densities and bulk moduli of the pore fluids. The equations relating the pore pressure and the elastic properties of the pore fluids are given by Batzle (1992). Since the compressibility of the gas is more easily affected by the pressure, we can know the pressure effects on pore fluids are mainly determined by the amount of gas existing in the reservoir.

Generally, the seismic properties such as velocity increase in a nonlinear way with the increase of the net overburden pressure: for low net overburden pressure regions, the seismic properties change faster (Wang, 2001). As a result, we need to know the pressure regime besides the pressure change in time-lapse seismic.

The effects of temperatures are mainly due to changes of bulk modulus and density of the pore fluids (Bell and Shirley, 1980). The equations relating the temperature and elastic properties of pore fluids have been given by Batzle (1992). For gas or water saturated rocks, the seismic velocities slightly decrease when the temperature increases. For oil saturated rocks, the seismic P-wave velocity decrease greatly when the temperature increases (Wang and Nur, 1988). But the S-wave velocity only slightly depends on the temperature. This dependence is caused by the pore pressure changes due to the thermal expansion.

Saturation mainly changes the densities and bulk modulus and thus affect the velocity. S-wave velocity changes caused saturation effects are mainly due to the density changes (Domenico, 1977). P-wave velocity changes caused by saturation effects are the combination effects of the density and bulk modulus changes. Packwood (1997) discussed the effect of patchy saturation distribution and claimed this effect cannot be ignored for seismic monitoring (i.e. time-lapse seismic) study.

Since the change of pore pressure results in the changes of stiffness of the rock matrix frame, pore fluids and the temperature affects the pore fluids, we can see from the Gassmann's equation that the resulting seismic property changes are the combined effects of pore pressure, temperature and saturation.

2.3.2 Application

Rock physics plays two roles in time-lapse seismic. First, it helps us forward model the seismic data difference due to reservoir changes. This is part of the feasibility study of time-lapse seismic. Second, it aids the interpretation of the measured time-lapse data.

According to current rock physical studies: first, not only the reservoir changes but also the properties of the experimental reservoir itself are important for the feasibility of the time-lapse seismic (Theune, 2004); second, the time-lapse seismic difference are the combined effects of saturation, pore pressure and temperature. Therefore, we need to decouple these factors according to their characteristics in the process of the interpretation of time-lapse data.

For the step of inverting the seismic attributes from the time-lapse seismic data, the rock physical forward modeling is one of the most important tools for determining which

seismic attributes and type of seismic data are most appropriate for a certain reservoir.

2.4 An Example

This section contains an illustrative example of what kinds of changes might occur in a very simple "reservoir" due to injection of water and the subsequent decrease in the effective pressure with distance from the well.

2.4.1 Defining the Model

The model here is very simple and essentially consists of layer of sandstone bounded above by a shale and below a carbonate. This model is loosely based on the geologic structure encountered in the Western Canada Sedimentary Basin although the sandstone thickness here is much thicker than is usually encountered in order to produce a substantial signal. The compressional wave velocities and densities of this material are assumed to be those of the weak sandstone encountered above. The velocities will be varied by changing the effective pressure on the material. That is, injection of water to this reservoir will lower the effective pressure and consequently the P and S wave velocities.

Typical velocities of the upper and lower layers are taken from well logs in the Senlac area of western Saskatchewan as provided in Theune (2004). The physical properties of the materials for this simple layered model are provided in Table 2.1.

The sandstone is assumed to be a clean sand with a high porosity of 30% and with complete water saturation. Again, to assist the simulation, the overburden pressure is assumed to be $80MPa$ and the pore pressure is assumed to be zero, that is, the initial effective pressure is $80MPa$. It must be noted that this value is quite high and would correspond in a sedimentary basin to depths of about $4000m$. However, the purpose here is not to simulate an actual situation but to illustrate the kind of changes that might be possible due to changing pressures in a reservoir. One should not take this model to be realistic as the changes provided will be large relative to those that can occur during the substitution of gas for liquid.

The changes in the effective pressure is assumed to range from $80MPa$ to $5MPa$ with the variations of the velocity being described by that from the lab experiment of He

and Schmitt (2004) which is shown in Figure 2.1. This variation in effective pressure is produced by changing the pore fluid pressure from zero to $75MPa$. Again, this is not a realistic change but the purpose of this exercise is to provide a model data set for testing the inversion parameters.

2.4.2 Determination of the Pore Fluids' Elastic Parameters

In this example, we assume the reservoir only contains water and the elastic properties of water are given. The bulk modulus of water is $2.2 \times 10^9 N/m^2(Pa)$ and the density of water is $1000kg/m^2$. These values are not exact but are appropriate for the illustration purposes. In a real study, one would have to include the pressure and temperature variations of the water's compressibility and density. The concentration and type of dissolved solids within water will also be important.

2.4.3 Determination of the Other Elastic Parameters

The effects of the effective pressure on the bulk modulus of the rock matrix frame are given by the experimental data which is shown in Figure 2.1(a) (assumed to be quartz grains) and provided in detail in Table 2.2. The bulk modulus of the solid is given as $36.5GPa$ and the density is $2670kg/m^3$.

The shear modulus of the rock matrix frame is determined from the shear sonic and density logs by equation 2.2.

2.4.4 Calculation of Seismic Velocity

The bulk modulus of the rock matrix frame, solid material, pore fluids and porosity are substituted into the Gassmann's equation 2.9 in order to obtain the bulk modulus of the saturated rock. The shear modulus of the saturated rock is obtained from the shear modulus of the rock matrix frame by equation 2.9. The density of the saturated rock is obtained by substituting the density of pore fluids, solid material and porosity into equation 2.3.

Finally, P-wave and S-wave Velocities are calculated using equation 2.11.

2.4.5 Synthetic time-lapse seismic data

After getting the seismic velocity and density of the reservoir, we can calculate the synthetic seismogram by seismic forward modeling techniques. The properties of the base survey model is given in the following table 2.1.

Layer	Thickness (m)	Interval Velocity(m/s)	Density(kg/m^3)
First	100	3000	2400
Second	100	3033	2169
Third	100	5000	2600

Table 2.1: Model Parameters for the Base Survey

P_p (MPa)	P_d (MPa)	K_d(GPa)	μ_d (GPa)	V_p (m/s)	V_s (m/s)
75	5	2.279	3.0915	2363.4	1193.9
70	10	3.1734	3.7321	2503.4	1311.7
65	15	3.6282	4.0951	2576.3	1374.1
55	25	4.5843	4.8243	2719.7	1491.4
50	30	5.0465	5.0599	2773.7	1527.4
45	35	5.3423	5.3113	2818.7	1564.8
40	40	5.632	5.502	2856.2	1592.7
35	45	5.8973	5.6755	2890.1	1617.6
30	50	6.1248	5.8411	2920.6	1641
25	55	6.308	5.914	2938.6	1651.2
20	60	6.4176	6.0253	2956.4	1666.7
15	65	6.5329	6.1595	2976.8	1685.2
10	70	6.7485	6.2248	2995.6	1694.1
5	75	6.9352	6.3188	3015.6	1706.8
0	80	7.0739	6.4128	3032.8	1719.5

Table 2.2: Elastic Parameters for Different Effective Pressure

The model for the monitor survey only changes the properties of the second layer which has been calculated by the above process. We assume the source wavelet is a Berlage wavelet (Aldridge, 1990). This analytical form of this wavelet is defined as:

$$w(t) = AH(t)t^n e^{-\alpha t} cos(2\pi f_0 t + \phi_0), \tag{2.13}$$

where $H(t)$ is the Heaviside unit step function, ϕ_0 is the initial phase angle, A is the maximum amplitude of this wavelet, α is the exponential decay factor and the time exponent

is n.

This wavelet and its amplitude spectrum are shown in Figure 2.2. This wavelet is a minimum phase wavelet which possesses a certain degree of differentiability. It is suitable for the accurate simulation of many physical wave propagation phenomena. The forward modeling results are shown in the following figures 2.3, 2.4, 2.5. No amplitude scaling is applied to these figures. The velocity model of the base survey (left panel of Figure 2.3) contains three horizontal layers whose elastic properties are provided in table 2.1. The pore pressure in the second layer of the monitor survey is decreased from the center trace to the near and far trace which follows the values in table 2.1. This gives us the velocity model of the monitor survey (right panel of Figure 2.3). Figure 2.4 shows the noise free synthetic results for the two surveys. After adding some random noise, we obtain the noisy synthetic seismograms (Figure 2.5). The two figures show that the synthetic seismogram for the monitor survey have a pull down effect. This can be explained by the results shown in Figure 2.1(c). The overburden pressure is fixed, so the decrease in pore pressure will increase the effective pressure and thus increase the reservoir P-wave velocities. Since we assume the thickness of the reservoir is constant, the increase of the velocities will cause the decrease of the travel-time.

2.5 Summary

This chapter can be divided into two parts. In the first part, the concept of reservoir parameters and the calculation of pore fluids and rock properties are first introduced. Then I give the equations for the estimation of the composite material. After that, the fluid substitution theory is reviewed and the Gassmann's equation is described in detail. In the second part, I discussed the most common reservoir parameters which can affect the time-lapse seismic data. The effects of effective pressure on the reservoir velocities are given by experimental data. Based on these experimental data, a simple synthetic time-lapse seismic data set is generated by assuming only the effective pressure is changed during the time-lapse acquisition.

In this thesis, we follow the most-often used rock physical knowledge for time-lapse research. However, in the real world, many factors we ignored are important and cannot

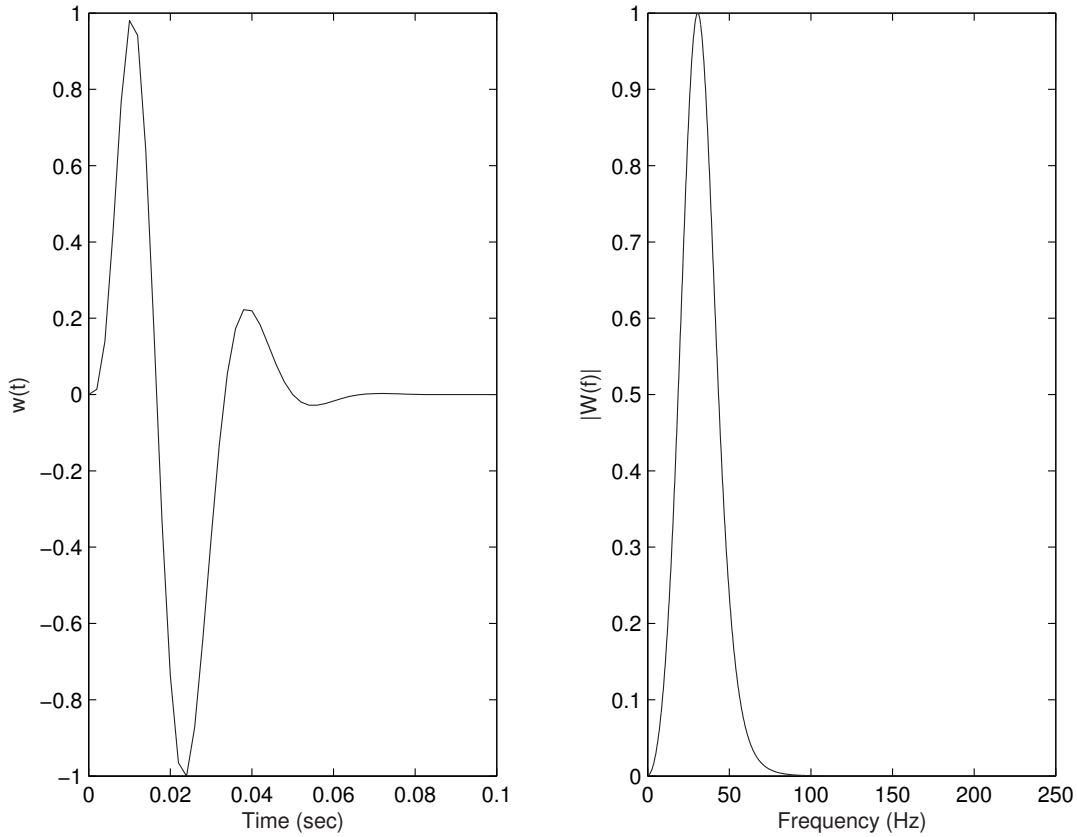

Figure 2.2: Left panel: a berlage wavelet with constants $A = 10^6 mm/s^4$, $n = 3$, $\alpha = 188.5 rad/s$, $f_0 = 30 Hz$, $\phi_0 = -\pi/2$. Right panel: the amplitude spectrum of this Berlage wavelet.

be neglected. Thus for a better result, more factors such as porosity geometry, pore fluid distributions, clay content, overburden pressure and textures of rocks should be taken into consideration and other fluid substitution theory should be applied.

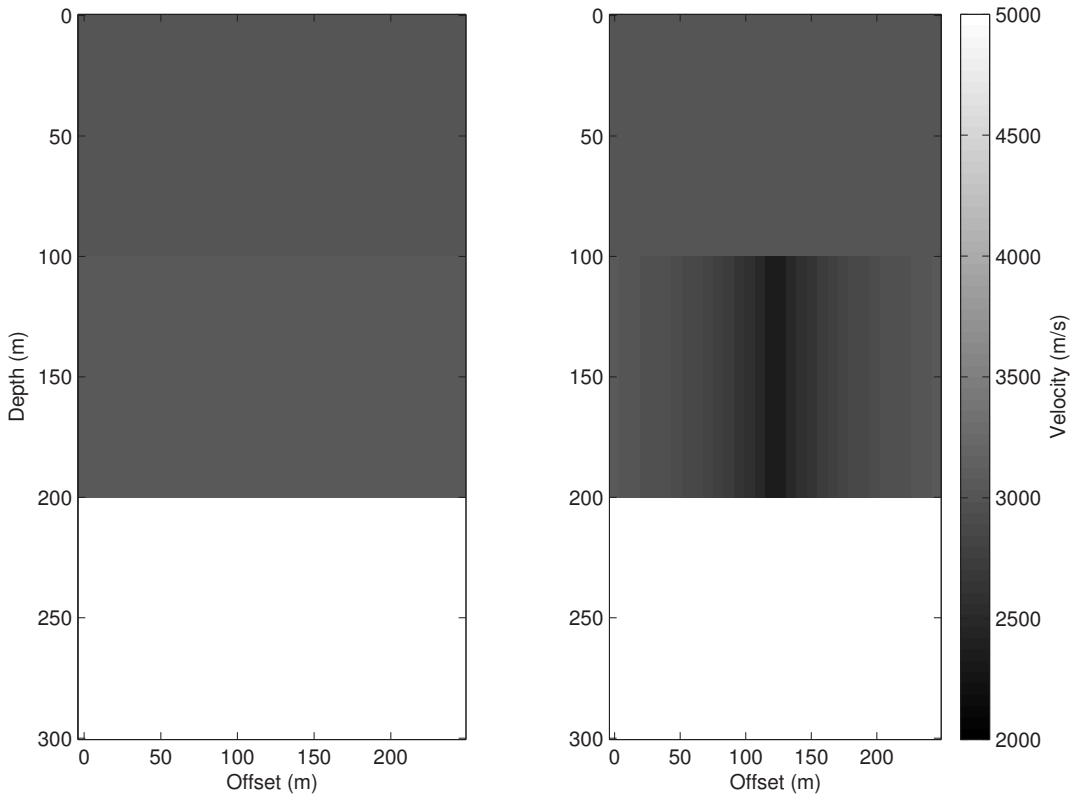

Figure 2.3: Left panel: the velocity model of the base survey. Right panel: the velocity model of the monitor survey.

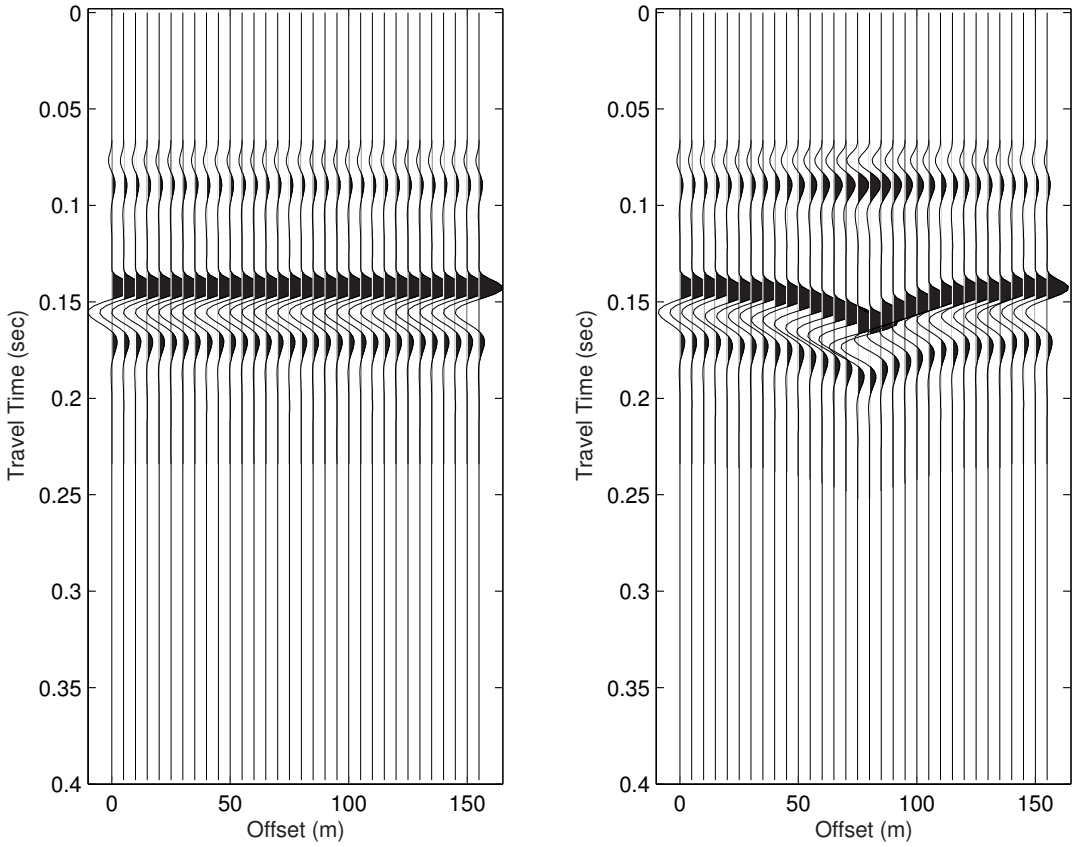

Figure 2.4: Left panel: noise free seismic data of the base survey. Right panel: noise free seismic data of the monitor survey.

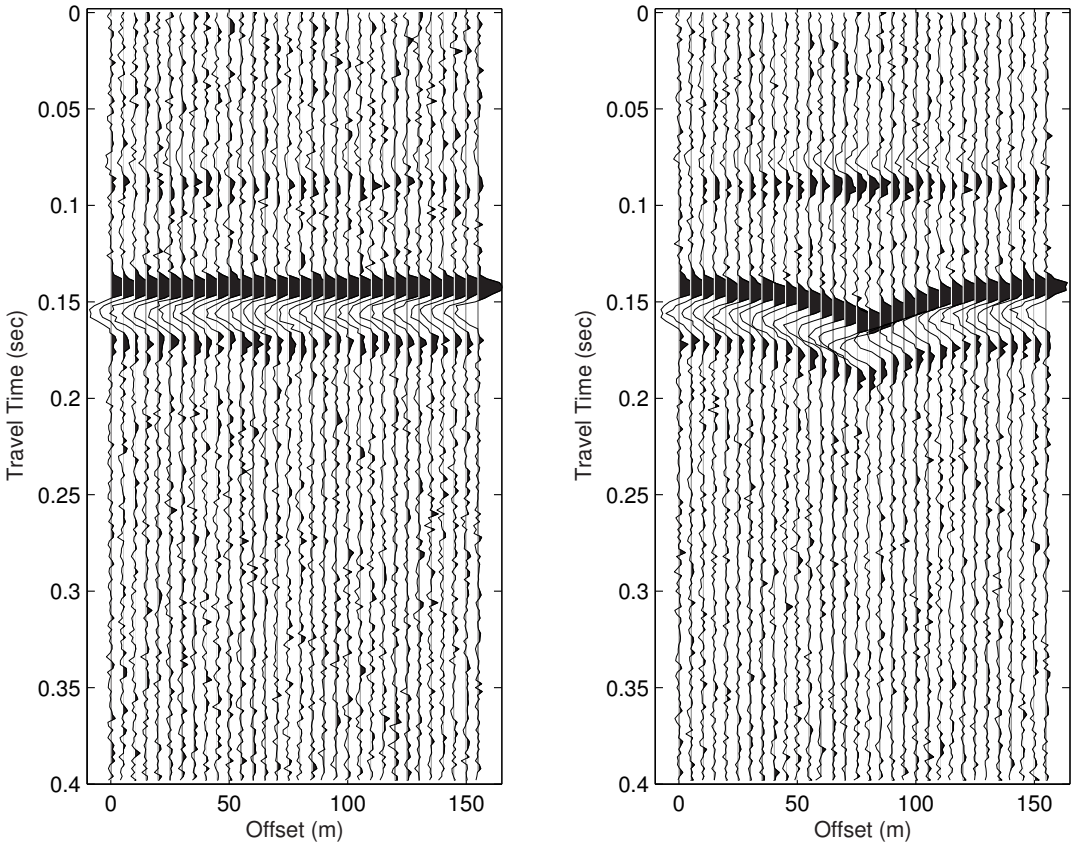

Figure 2.5: Left panel: noisy seismic data of the base survey. Right panel: noisy seismic data of the monitor survey.

Chapter 3

Fundamentals of Inversion

3.1 Ill-posed and Well-posed Problems

The concept of ill-posed and well-posed problems are introduced by Hadamard (1923). He claims that a problem is ill-posed if the solution does not exist or is not unique or stable. On the contrary, a problem is well-posed if the solution exists and is unique and stable. Hadamard believed the ill-posed problems were not physically or mathematically meaningful. However, this idea is mistaken. Today, many applications in science and engineering lead to ill-posed problems in the form of inverse problems. The solutions to these ill-posed problems have well-defined physical meaning.

An inverse problem and its corresponding forward problem is a pair. In general, we call the problem which is well-posed as the forward problem and the one which is ill-posed as the inverse problem. Therefore, inverse problems in this thesis is equivalent to ill-posed problems. Mathematically, ideal (noise free) forward and inverse problems can be written as,

$$\begin{aligned} \mathbf{d} &= \mathbf{L}\mathbf{m}, \\ \mathbf{m} &= \mathbf{L}^{-1}\mathbf{d}, \end{aligned} \tag{3.1}$$

where $\mathbf{L}$ is the forward operator, $\mathbf{L}^{-1}$ is the inverse of the forward operator or called inverse operator. The vector $\mathbf{m}$ denotes the solution or model parameters, $\mathbf{d}$ means the data. If $\mathbf{L}$ is a linear operator, the inverse problem is called a linear inverse problem.

In practice, the data, model parameters and operators are discretized. The data $\mathbf{d}$ are assumed known while $\mathbf{m}$ is unknown. If we have same number of known and unknown,

a problem is an even-determined problem. If we have more observations than unknowns, this problem is a over-determined problem. If we have more unknowns than known, this problem is under-determined problem. An under-determined problem does not have an unique solution, thus it is an ill-posed problem. Practical ill-posed problems are not purely under-determined or over-determined. They may have more known than unknowns but the known data are not independent (Menke, 1984). Therefore, some model parameters are over-determined but some model parameters are under-determined.

The most important property of an inverse problem is the stability. If a problem is not stable, it means any small perturbation of the data can cause large perturbation on the solution. Since the data of a problem are always perturbed by noise and computation errors always exist, the computed solution will be far from the true solution if the problem lacks stability. In a mathematical viewpoint, a problem is unstable if the forward operator $\mathbf{L}$ has a large condition number or the kernel matrix is singular (Golub and Van Loan, 1996).

The seismic exploration problem, which often infers the subsurface structures or lithology from the data measured at the surface of the earth, should have certain solutions because the underlying models exist in the real world. However, there are no unique answers since the observed data are finite while the real earth is a continuous distribution of physical property (Scales and Smith, 1994). Meanwhile, the measured data are always contaminated by random and systematic noise (Jackson, 1972). If the noise level is larger than the data responses due to the model parameters, it is impossible to get a stable solution.

3.2 Regularization

The fundamental difficulty of the inverse problems is the lack of information. The non-uniqueness of the solution is because of the lack of the information of the solution (Franklin, 1970; Jackson, 1979). The stability problem of the solution is due to the lack of information (Gray and Symes, 1985). This important idea is stated by Lanczos (1961) :"A lack of information cannot be remedied by any mathematical trickery! "

In 1960s, Tikhonov developed the theoretical foundations of the solution of ill-posed

problems. He introduced the technique called regularization. The idea of regularization is to incorporate additional information about the inverse problems to obtain a meaningful and stable solution. Mathematically, regularization technique is equivalent to constraining the model null space to a desired pattern.

3.2.1 Definition of the Null Space

A linear seismic inversion problem is to seek a solution to a problem in the following form,

$$\mathbf{d} = \mathbf{L}\mathbf{m},$$

where $\mathbf{L}$ is a linear operator that maps a model $\mathbf{m}$ to the data $\mathbf{d}$. The definition of the null space of $\mathbf{L}$ is given by

$$\mathbf{0} = \mathbf{L}\mathbf{m}_0, \tag{3.2}$$

where $\mathbf{m}_0$ is a vector which belongs to a space $\mathbf{N}$. This space $\mathbf{N}$ is the null space in which any vector satisfies the above equation (Menke, 1984; Nichols, 1994).

It is obvious that the null space is not empty if a problem is under-determined or mixed-determined and a vector in the null space does not change the data misfit. Therefore infinite solutions can be obtained by adding some amount of the null space to a valid solution because the additional null space vector does not harm the data misfit.

3.2.2 Model and Data Regularization

In real world, two different sources of information are needed to constrain the null space: first, the noise information of the data; second, the characteristics of the desired model. Mathematically, for a practical forward problem $\mathbf{d} \approx \mathbf{L}\mathbf{m}$, the inverse problem which incorporates additional noise and model information, can be formulated as a problem to minimize the cost function,

$$\mathbf{J} = \mathcal{F}(\mathbf{d} - \mathbf{L}\mathbf{m}) + \epsilon\mathcal{R}(\mathbf{m}), \tag{3.3}$$

where, $\mathcal{F}$ is a functional of data residual $\mathbf{d} - \mathbf{L}\mathbf{m}$ and $\mathcal{R}$ is a functional of model $\mathbf{m}$. The first part of the right-hand side of this equation is called data term and the second part of the right-hand side of this equation is called model term, ϵ is the trade-off (regularization) parameter which adjusts the weights of data term and model term (Zhou et al., 2002).

The most common functionals $\mathcal{F}$ is the l_p norm (where $1 \leq p \leq 2$). The typical functionals $\mathcal{R}$ try to damp, smooth the model parameters or make them sparse. The typical methods to choose ϵ are: discrepancy principle; L-curve criterion and generalized cross-validation (GCV) (Hansen, 1998; Trad, 2003).

It is obvious that we can obtain a high-fidelity inversion result if good forms for $\mathcal{F}$, $\mathcal{R}$ and a suitable value of ϵ are selected. Generally, the choice of data and model terms depend on a *priori* knowledge of the model and data noise. The choice of ϵ depends on the noise level.

3.2.3 Model-Space and Data-Space Regularization

Model and data regularization design the functionals of the model and data residual in a cost function for an inverse problem. After obtaining this cost function, we can adjust the form of the cost function in an equivalent way. Depending on the way adjusting the cost function, the formulation of the regularized inverse problem can be called model-space or data-space regularization (Fomel, 1997). For simplicity, we just consider a linear inverse problem here.

Suppose we have a linear inverse problem $\mathbf{d} = \mathbf{Lm}$, model-space regularization incorporates a *priori* information about the model into this problem. One possible cost function for this linear inverse problem can be written as,

$$\mathbf{J} = ||\mathbf{d} - \mathbf{Lm}||_2^2 + \epsilon||\mathbf{Dm}||_2^2, \tag{3.4}$$

$\mathbf{D}$ acts like a filter which pass the undesired components of a estimated model. The linear system of this cost function can be formulated as,

$$\left[\begin{array}{c} \mathbf{L} \\ \sqrt{\epsilon}\mathbf{D} \end{array} \right] \mathbf{m} = \left[\begin{array}{c} \mathbf{d} \\ \mathbf{0} \end{array} \right]. \tag{3.5}$$

In this equation, the forward operator are augmented by adding more columns and the data vector are augmented by adding a zero vector. The analytical solution to this problem is,

$$\hat{\mathbf{m}} = (\mathbf{L}^T\mathbf{L} + \epsilon\mathbf{D}^T\mathbf{D})^{-1}\mathbf{L}^T\mathbf{d}, \tag{3.6}$$

where $\hat{\mathbf{m}}$ means the estimated solution.

If we use data-space regularization, the trial case (that is $\mathbf{D}$ is an identity operator) of the cost function 3.5 can be reformulated as,

$$J = ||\mathbf{d} - \tilde{\mathbf{L}}\tilde{\mathbf{m}}||_2^2, \tag{3.7}$$

where

$$\begin{aligned} \tilde{\mathbf{L}} &= \begin{bmatrix} \mathbf{L} & \sqrt{\epsilon}\mathbf{I_d} \end{bmatrix} \\ \tilde{\mathbf{m}} &= \begin{bmatrix} \mathbf{m} \\ \mathbf{r} \end{bmatrix} \\ \epsilon\mathbf{r} &= \mathbf{d} - \mathbf{L}\mathbf{m}, \end{aligned} \tag{3.8}$$

where $\mathbf{I_d}$ denotes the data-space identity operator. For "non-trivial" case, a preconditioning operator $\mathbf{D}$ is used to make an equality

$$\mathbf{x} = \mathbf{D}\mathbf{m}. \tag{3.9}$$

Now, the problem becomes non-trivial. Data-space regularization can formulate the following linear system,

$$\begin{bmatrix} \mathbf{L}\mathbf{D}^{-1} & \sqrt{\epsilon}\mathbf{I_d} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{r} \end{bmatrix} = \mathbf{d}, \tag{3.10}$$

where $\mathbf{r}$ follows equation 3.8. The analytical solution to this problem is

$$\mathbf{m} = \mathbf{D}^{-1}\hat{\mathbf{x}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{L}^T(\mathbf{L}(\mathbf{D}^T\mathbf{D})^{-1}\mathbf{L}^T + \epsilon\mathbf{I_d})^{-1}\mathbf{d}. \tag{3.11}$$

Comparing the solutions shown in the equations 3.11 and 3.6, we can find the two solutions are identical. It means that the model-space and data-space regularization give the same final result. However, the effect of preconditioning will generates different results when we use semi-iterative methods such as conjugate gradient (CG) method. We will discuss this later in this chapter.

3.2.4 Bayesian Framework for Inversion

As we have said before, the regularization technique is to incorporate a *priori* information into the inverse problem in order to obtain a stable and meaningful solution. An important tool to incorporate the *priori* information is Bayes' theorem which originates

from probability theory (Box and Tiao, 1973; Jackson, 1985; Sivia, 1996; Tarantola, 1987). Bayes' rule can be put in the following form,

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}. \tag{3.12}$$

The left-hand side of the above equation is called a posteriori probability density function (pdf) which is a conditional pdf of model parameters $\mathbf{m}$ given a data vector $\mathbf{d}$. $P(\mathbf{d}|\mathbf{m})$ is the likelihood function which gives the probability of the data given the model parameters. It contains information about the noise and relations between the data and the model, $\mathbf{m}$ is a *priori* information about the model parameters. This information was provided by general knowledge of the model or information given by other source of data. The denominator $P(\mathbf{d})$ does not depend on the model parameters so that it can regarded as a constant scaling factor. Once we obtain the posteriori pdf, all the information about the model , which is actually the solution to the inverse problem, given by the data is known.

Bayes' theorem allows more emphasis on scientific interests instead of mathematical convenience. The difficulty of the Bayesian framework is that there is no good way to obtain the exact *priori* information about the model and likelihood information about the noise. However, Bayesian framework for inversion is still valuable if we can use it diligently (Ulrych et al., 2001).

3.3 Numerical Techniques for Inversion

Since the practical inverse problems are dealing with discrete numbers containing the noise, it is important to investigate the numerical techniques for inversion. In this section, we will discuss the most common ones: adjoint operator, Iteratively Re-weighted Least Squares (IRLS) method (Scales, 1987) and conjugate gradient method (Hestenes and Steifel, 1952).

3.3.1 Forward and Adjoint Operator

An operator $\mathbf{L}^T$ is called the adjoint operator to the operator $\mathbf{L}$, when the two operators satisfy the following equation,

$$(\mathbf{Lm}, \mathbf{d}) = (\mathbf{m}, \mathbf{L}^T\mathbf{d}), \tag{3.13}$$

where $(,)$ means dot product. Theoretically, the inverse operator $\mathbf{L}^{-1}$ will provide a better solution than the adjoint operator. However, the adjoint operator tolerates the noise and lack of information better than the inverse operators (Claerbout, 1992). Therefore the adjoint operator works better than the inverse operator in practice. An adjoint operator can be done by a matrix transpose operation. Two common approaches to find the adjoint operator are: formulate a continuous forward operator, obtain the adjoint operator to this forward operator, then discretize the adjoint operator; formulate a continuous operator, discretize the forward operator and take the Hermitian adjoint of this discretized forward operator (Ji, 1994). The distinct feature of the first approach is that it has a Jacobian weighting factor whose effects are still not clear. The adjoint operator does not adequately approximate the inverse operator in many situations. Therefore, we usually iteratively use the forward operator and adjoint operator to optimize the cost function for the inverse problem.

In real world applications, the size and complexity of an operator make a matrix operator not practical. Thus forward and adjoint operators are often done by a subroutine instead of a big matrix (Claerbout, 1992; Claerbout, 2004). To test if two subroutines are forward and adjoint pairs, the equation 3.13 is suitable and usually used. This test using the equation 3.13 is called a dot product test.

In practice, an adjoint operator is often scaled to obtain a better solution and many empirical scaling factors have been studied by researchers (Claerbout, 1997a; Claerbout, 1997b). However, there is no straight forward theory for the scaling of an adjoint operator.

3.3.2 Iteratively Re-weighted Least Squares

A data or model weighting matrix is often designed to keep the cost function of a problem in quadratic form when a different norm for the data residual or model is desired (i.e. the case of model or data regularization). The reason is that a quadratic form cost function is simple to solve by methods like conjugate gradient. However, after adding a weighting matrix, this problem becomes nonlinear. Iteratively Re-weighted Least Squares (IRLS) method is often applied to this type nonlinear optimization. Given a simple example cost function $\mathbf{J} = ||\mathbf{W_d}(\mathbf{d} - \mathbf{Lm})||_2^2 + ||\mathbf{W_m}\mathbf{m}||_2^2$, it does not need IRLS method if $\mathbf{W_d}$

and $\mathbf{W_m}$ are known. If $\mathbf{W_d}$ is a function of the data $\mathbf{d}$ or $\mathbf{W_m}$ is a function of the model $\mathbf{m}$, the algorithm of IRLS for this cost function is:

1. Fixing the weighting matrix $\mathbf{W_d}$ and $\mathbf{W_m}$ so that the nonlinear problem turns out to be a linear problem, and using a direct inversion method or conjugate gradient method for solving this linear problem. This process is called an inner iteration.

2. Using the result from the previous iteration to update the weighting matrix $\mathbf{W_d}$ and $\mathbf{W_m}$.

3. Solving the linear problem using the new weighting matrix $\mathbf{W_d}$ and $\mathbf{W_m}$. The iteration which is from step one to step three is called an external iteration.

The number of inner iterations is determined by an automatic stopping criterion and the number of external iterations is determined by trial and error method.

3.3.3 Conjugate Gradient Method

Conjugate gradient (CG) method (Hestenes and Steifel, 1952) is widely used in optimization problems. It does not require any explicit matrix operation. This property is crucial for seismic inversion because the operators in seismic inverse problems are coded as subroutines instead of matrices.

Given a simple example problem $\mathbf{Lm} \approx \mathbf{d}$, we can obtain the solution $\mathbf{m}$ by minimizing the l_2 norm the data residual. If the forward operator $\mathbf{L}$ is equivalent to a symmetric positive definite (SPD) matrix (Golub and Van Loan, 1996), the conjugate gradient (CG) method can be applied. If $\mathbf{L}$ is not equivalent to a SPD matrix, the CG method can be applied to a modified form of this equation $\mathbf{L^T Lm} = \mathbf{L}^T\mathbf{d}$. Here, $\mathbf{L}^T$ is the adjoint operator. One CG algorithm called conjugate gradient least squares (CGLS) for solving this equation can avoid the explicit calculation of $\mathbf{L^T L}$. When data or model regularization is applied, the operator $\mathbf{L}$ and model and data vectors are augmented in the form similar to the equations 3.5 and 3.10. In this case, we should substitute augmented data vector, model vectors and operator $\mathbf{L}$ into the simple form CGLS algorithm. The simple form CGLS (Scales, 1987) algorithm can be stated as follows: Choose the starting model $\mathbf{m}_0$ to

be zero. Put $\mathbf{s}_0 = \mathbf{d} - \mathbf{L}\mathbf{m}_0$, $\mathbf{r}_0 = \mathbf{p}_0 = \mathbf{L}^T\mathbf{s}_0$ and $\mathbf{q}_0 = \mathbf{L}\mathbf{p}_0$. Then for $k = 0, 1, 2, \ldots$

$$\begin{aligned}
\alpha_{k+1} &= \frac{(\mathbf{r}_k, \mathbf{r}_k)}{(\mathbf{q}_k, \mathbf{q}_k)}, \\
\mathbf{m}_{k+1} &= \mathbf{m}_k - \alpha_{k+1}\mathbf{p}_k, \\
\mathbf{s}_{k+1} &= \mathbf{s}_k - \alpha_{k+1}\mathbf{q}_k, \\
\mathbf{r}_{k+1} &= \mathbf{L}^T\mathbf{s}_{k+1}, \\
\beta_{k+1} &= \frac{(\mathbf{r}_{k+1}, \mathbf{r}_{k+1})}{(\mathbf{q}_k, \mathbf{q}_k)}, \\
\mathbf{p_{k+1}} &= \mathbf{r}_{k+1} + \beta_{k+1}\mathbf{p}_k, \\
\mathbf{q}_{k+1} &= \mathbf{L}\mathbf{p}_{k+1},
\end{aligned} \tag{3.14}$$

where $\mathbf{L}$ and $\mathbf{L}^T$ are the forward and adjoint operator, respectively. Operator $(,)$ denotes the dot product. Each iteration of this algorithm create an approximation to the solution. Therefore, we can stop or resume the algorithm at any stage before completion (Hanke and Hansen, 1993; Hansen, 1998). Theoretically, we can obtain an accurate solution after N, where N is the number of data points, iteration. If we stop the iteration before it reach the final solution, we say the optimization is not completed. However, this intermediate solution is still useful.

3.4 Example

In this section, we will illustrate the effects of regularization and compare the model-space and data-space regularization by some simple examples.

Suppose that data point d_j and the model $\mathbf{m}$ are related by the following equation,

$$d_j = \int_0^L e^{-\alpha(x-r_j)^2} m(x)dx, \tag{3.15}$$

where r_j is the position of the jth observation. We can discretize the above equation as,

$$d_j = \textstyle\sum_{k=0}^{M-1} \triangle x \; e^{-\alpha(r_j - x_k)^2} m(x_k) \quad j = 0, 1, 2, \ldots N-1, \tag{3.16}$$

where N indicates the number of observation, M denotes the number of model parameters and x_k is the position of kth model parameter. For simplicity, we assume r_j is given

by

$$r_j = jL/(N-1). \tag{3.17}$$

In our examples, we assume $\alpha = 150$, $L = 1$, $N = 10$, $M = 60$, noise is Gaussian noise and the noise level is one percent. Figure 3.1 shows the minimum norm solution to this problem. The model regularization for this case is a damping regularization term. As we

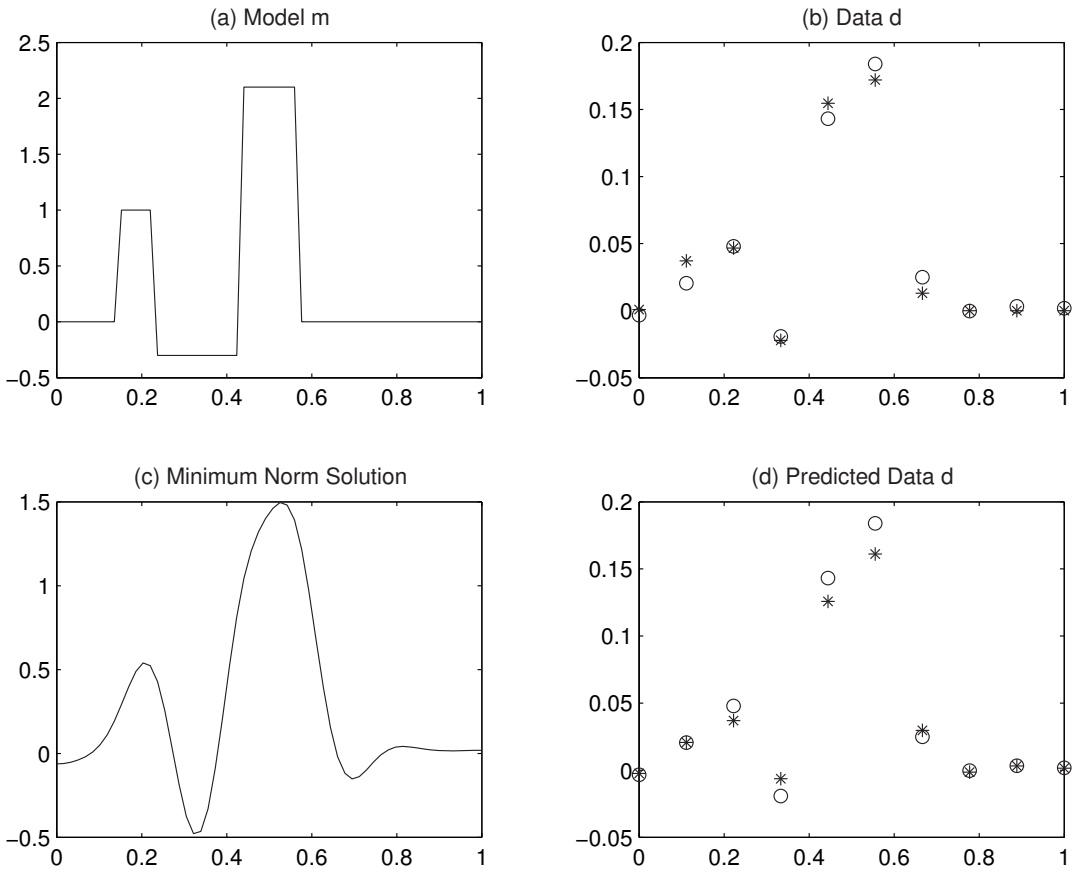

Figure 3.1: a) Model. b) Observed data and noise free data. c) Minimum norm solution. d) Predicted data and observed data. The circle in b) and d) denotes the observed data. x denotes the model parameter position and r means the observation position. The trade-off parameter is selected by χ^2 test.

can see from this figure, the predicted data does not fit the noise due to the effects of the damping regularization term. The trade-off parameter μ is determined by χ^2 test (Press et al., 1986) since the noise is Gaussian noise. The χ^2 test for this minimum norm solu-

tion is shown in the following Figure 3.2. Other model regularization terms include the

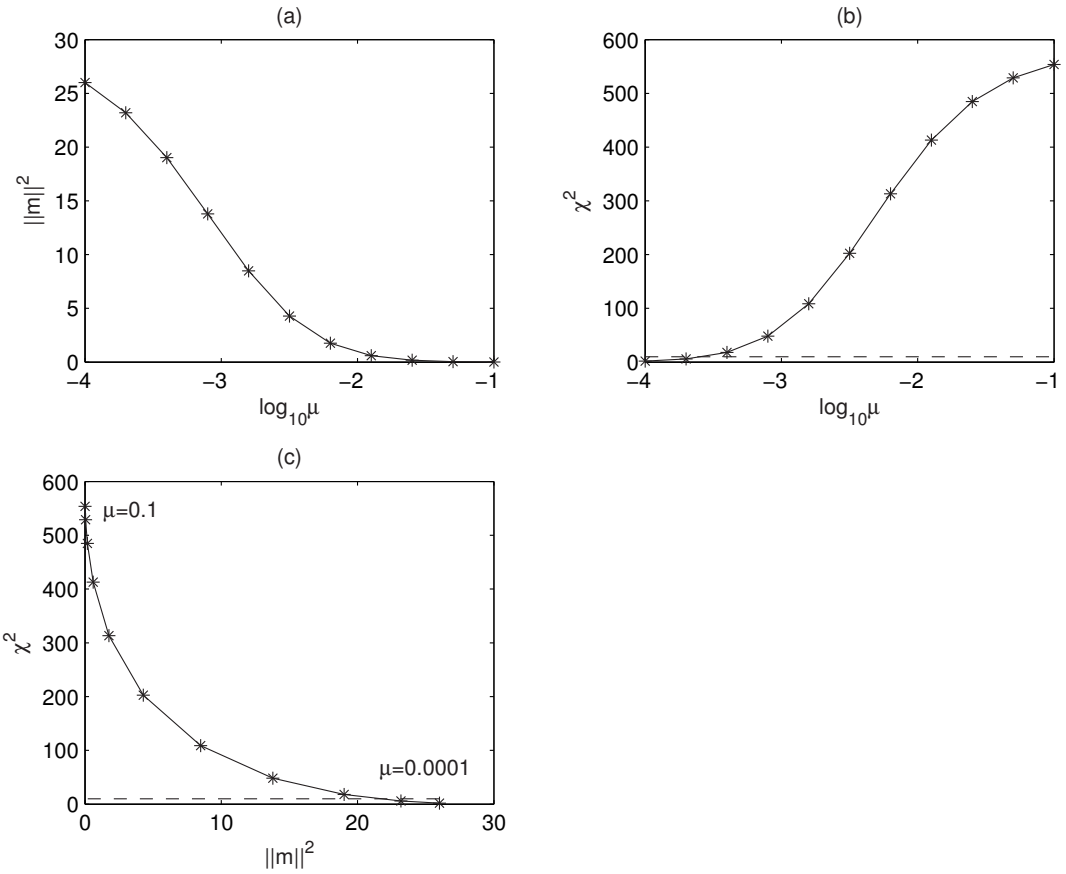

Figure 3.2: a) Model norm as a function of the logarithm of the trade-off parameter $log_{10}\mu$. b) The χ^2 test for the minimum norm solution. c) The trade-off curve for the misfit and model norm. The dashed line in b) and c) indicates the desired χ^2 value.

first and second derivative of the model parameters. The solution using the first derivative regularization term $\mathbf{D}_1$ is shown in Figure 3.3. Figure 3.4 illustrates the solution using the second derivative regularization term $\mathbf{D}_2$. These figures illustrate that damping regularization term forces the solution has the smallest magnitude, $\mathbf{D}_2$ regularization term make the solution smoothly varying and $\mathbf{D}_1$ regularization term make the solution tends to be flat. This is because the $\mathbf{D}_2$ term minimizes variation of the model parameters changes and the $\mathbf{D}_1$ term minimizes variation of the adjacent model parameters. The χ^2 test for the two regularization terms $\mathbf{D}_1$, $\mathbf{D}_2$ are shown in Figure 3.5 and 3.6 respec-

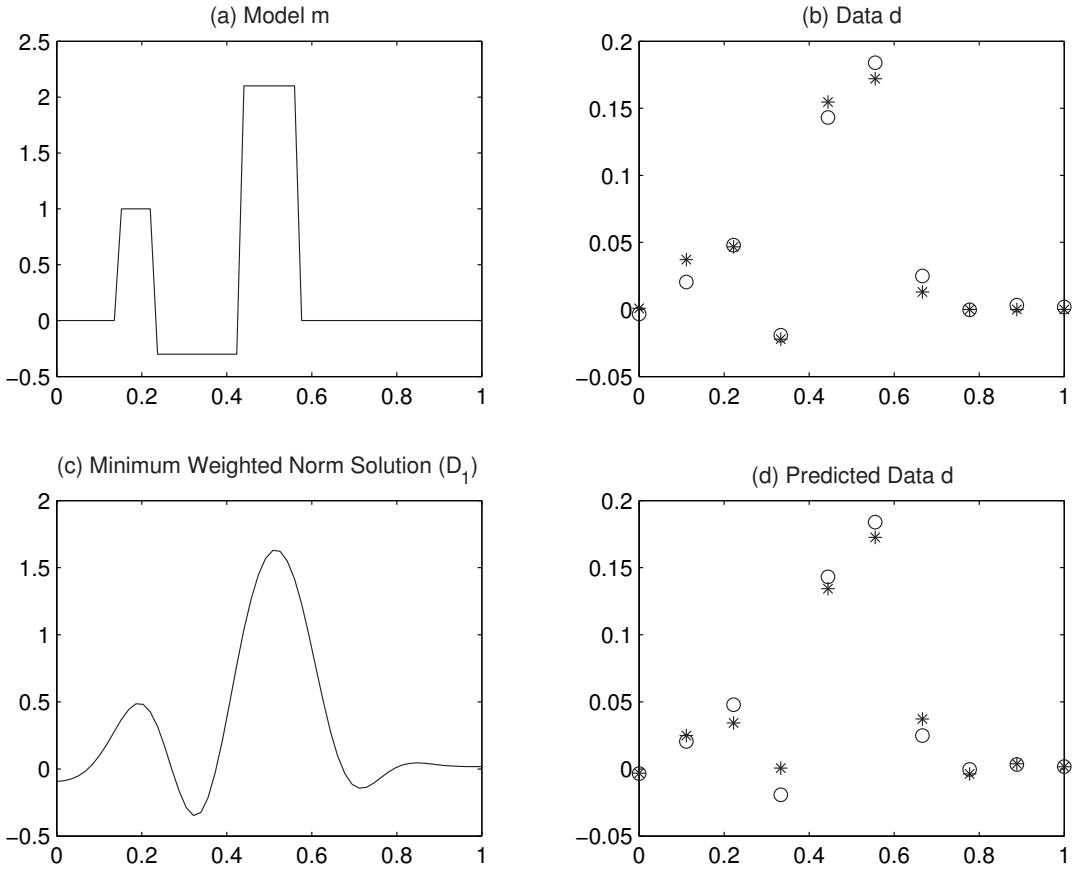

Figure 3.3: a) Model. b) Observed data and noise free data. c) Minimum weighted norm solution $\mathbf{D}_1$. d) Predicted data and observed data. The circle in b) and d) denotes the observed data. x denotes the model parameter position and r means the observation position. The trade-off parameter is selected by the χ^2 test.

tively. It is observed that the trade-off parameter is different for different regularization term. This indicates that we should determine the trade-off parameter by χ^2 test or other methods individually for different regularization terms.

The above examples show the effects of the regularization terms and one method of choosing the trade-off parameters. In the following part, we will compare the model-space and data-space regularization. The first example use a damping regularization term in model-space and data-space respectively. The results is shown in Figure 3.7. As we can see from Figure 3.7, the model-space regularization and data-space regularization

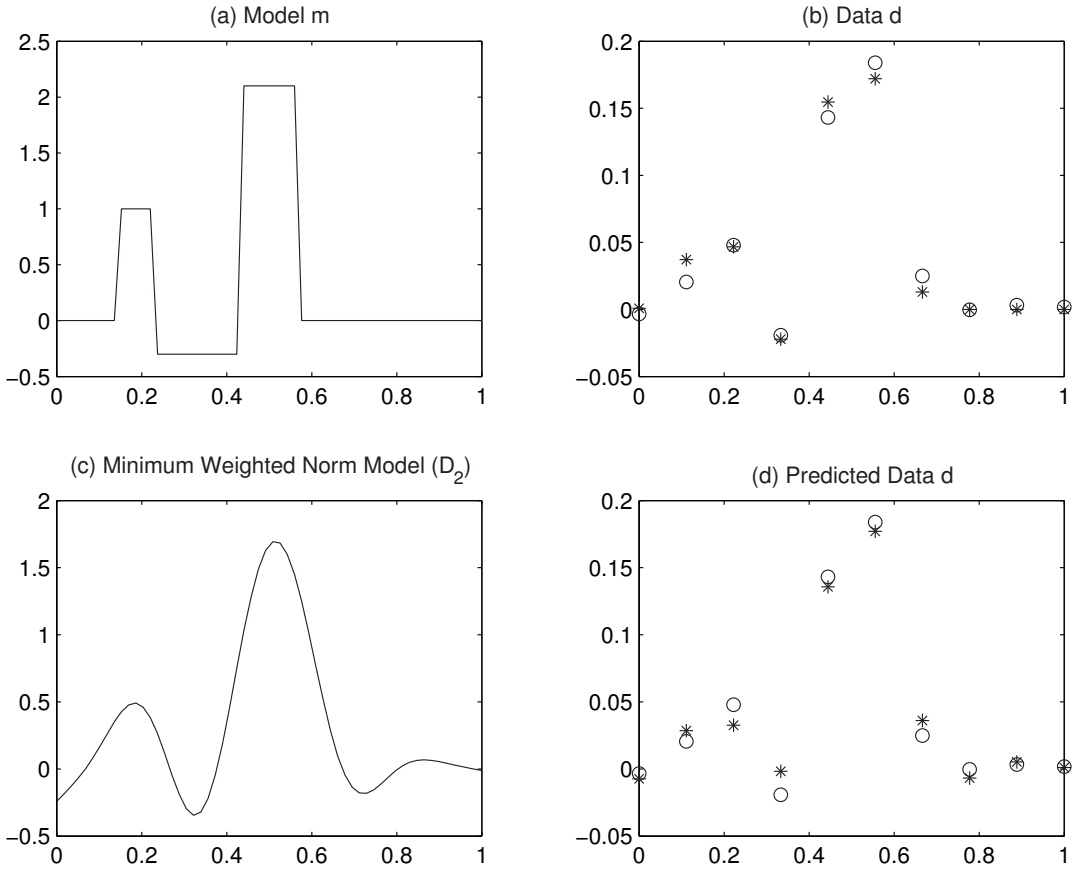

Figure 3.4: a) Model. b) Observed data and noise free data. c) Minimum weighted norm solution $\mathbf{D}_2$. d) Predicted data and observed data. The circle in b) and d) denotes the observed data. x denotes the model parameter position and r means the observation position. The trade-off parameter is selected by the χ^2 test.

give the same results even when the optimization is not completed. If we regularized the solution by $\mathbf{D}_1$, we need to introduce a preconditioning operator to solve the data-space regularization case. Figure 3.8 shows the results after different iteration numbers. It is obvious data-space and model-space regularization give us different solutions after 5 iterations. Results in 3.8 (c) shows that the two regularization methods give us the same results after 20 iterations. The results when we use $\mathbf{D}_1$ regularization terms indicate that the final results of the model-space and data-space regularization are identical but the results when the optimization is no completed is different. Comparing to the results

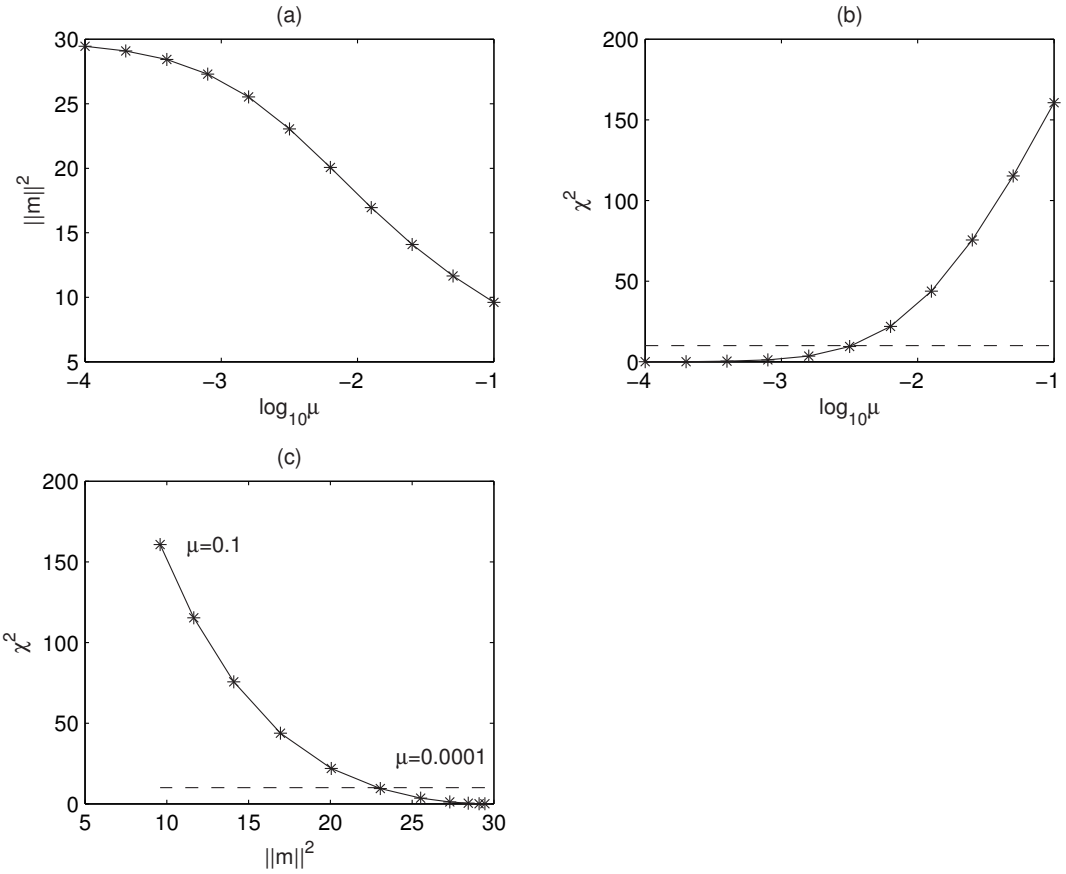

Figure 3.5: a) Model norm as a function of the logarithm of the trade-off parameter $log_{10}\mu$, b) the χ^2 test for the minimum weighted norm solution $\mathbf{D}_1$, c) the trade-off curve for the misfit and model norm. The dashed line in b) and c) indicates the desired χ^2 value.

shown in Figure 3.7 and 3.8, we can find the difference between the data-space and model-space regularization is due to the effects of preconditioning.

All these simple examples support the discussions on the regularization, model-space and data-space regularization in the above sections.

3.5 Summary

In this chapter, I reviewed the basics of inverse theory in three parts. The concepts of ill-posed and well-posed problem are introduced in the first part. The second part reviewed the regularization techniques: the definition of regularization, the concept of null space,

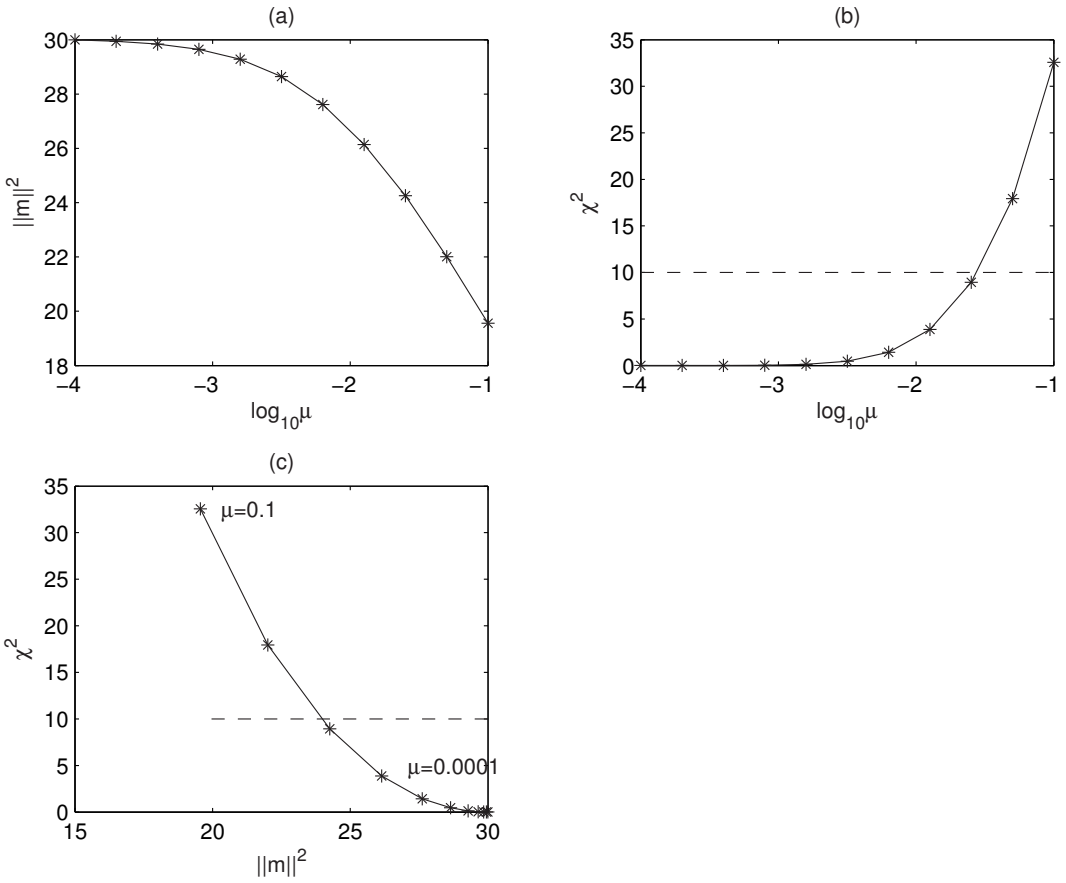

Figure 3.6: a) Model norm as a function of the logarithm of the trade-off parameter $log_{10}\mu$, b) the χ^2 test for the minimum weighted norm solution ($\mathbf{D}_2$), c) the trade-off curve for the misfit and model norm. The dashed line in b) and c) indicates the desired χ^2 value.

the model and data regularization, the model-space and data-space regularization and then the Bayesian framework for inversion. In the third part, I briefly introduced the most common numerical techniques for inversion: forward and adjoint operators, IRLS and CGLS method and the Conjugate Gradient algorithm. Finally, we test the regularization techniques on some toy examples.

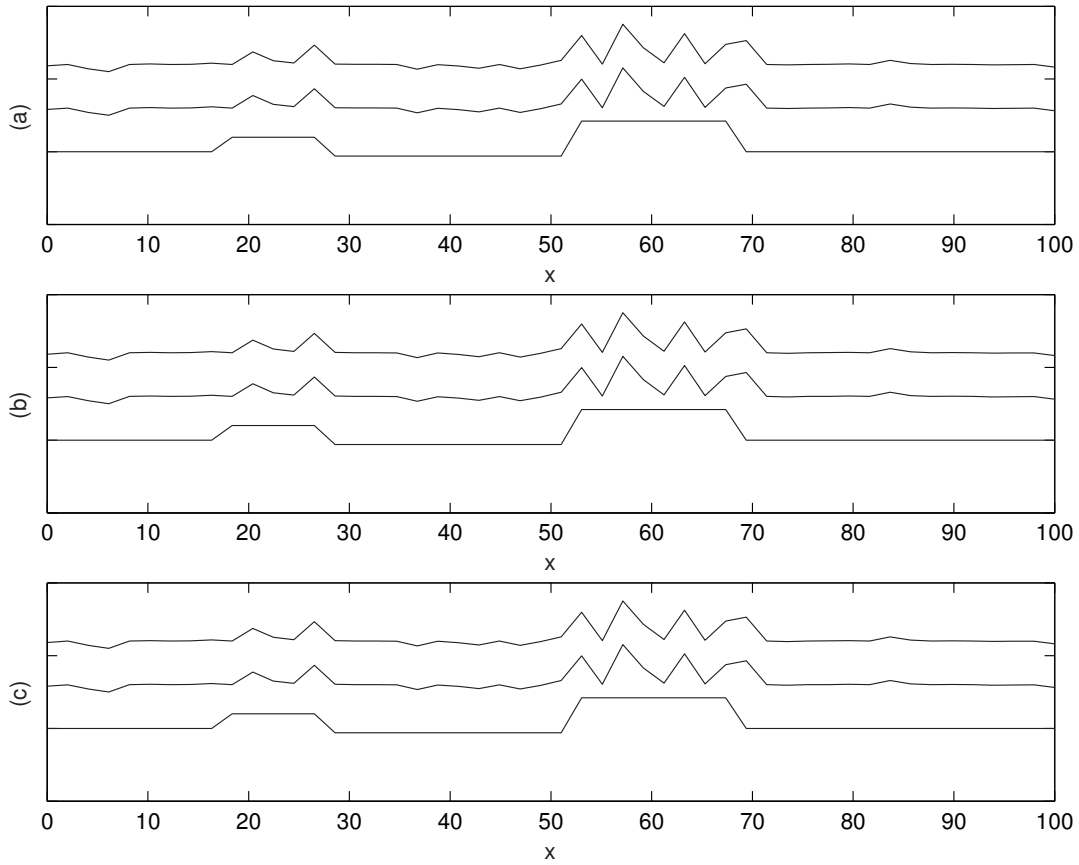

Figure 3.7: In each figure, the top curve is the solution by model-regularization (damping regularization term), the middle curve is the one by data-regularization and the bottom one is the true model. (a) Results after 5 iterations. (b) Results after 10 iterations. (c) Results after 20 iterations.

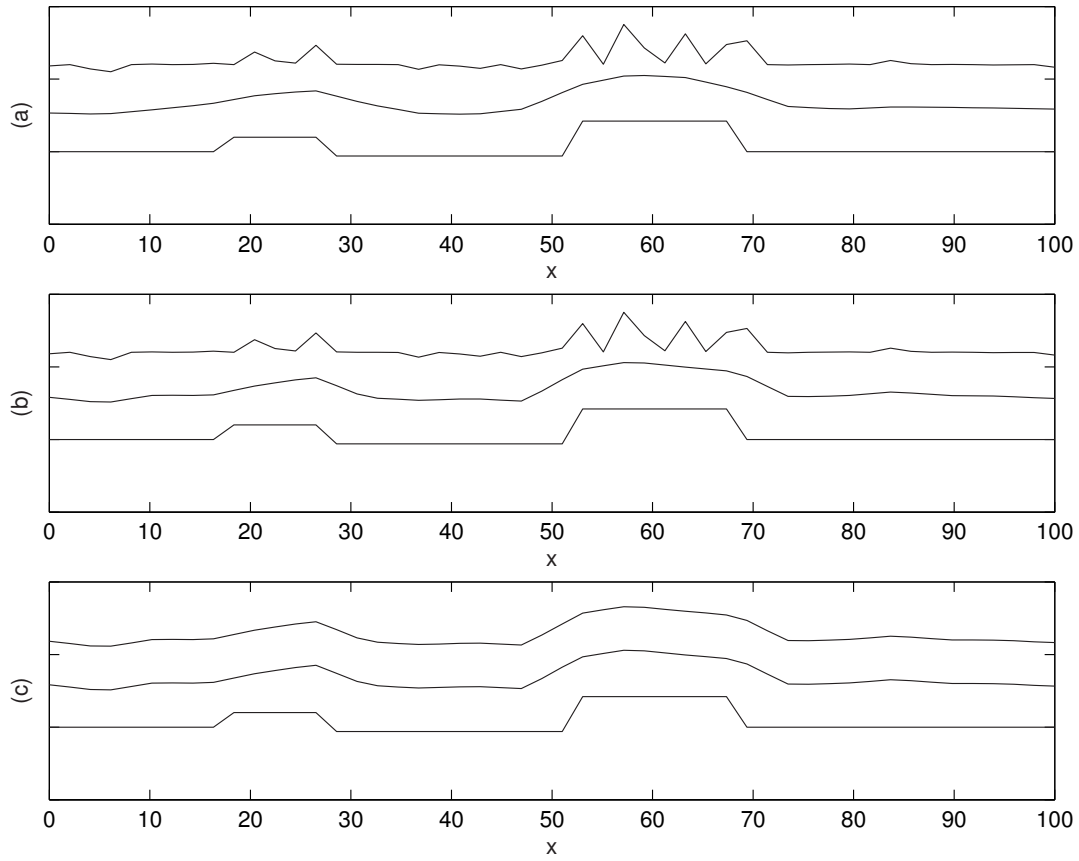

Figure 3.8: In each figure, the top curve is the solution by model-regularization (first derivative $\mathbf{D}_1$ regularization term), the middle curve is the one by data-regularization and the bottom one is the true model. (a) Results after 5 iterations. (b) Results after 10 iterations. (c) Results after 20 iterations.

Chapter 4

Constraints for Time-lapse Inversion

4.1 Limitations of Seismic Data

The advantages of an inversion framework is that inversion methods can obtain a better solution by integrating all information in the form of constraints. Therefore, it is necessary to investigate the limitations of seismic data in order to impose proper constraints. In practice, constraints are often given in a matrix form and coded as subroutines (Youzwishen, 2001; Claerbout, 1992; Claerbout, 2004) .

The first limitation of seismic data is that the observed data is finite while the real earth should be described as a continuous distribution of physical parameters (Scales and Smith, 1994). This limitation is fundamental because it means that the solution is always non-unique. In practice, we overcome this problem by assuming the model parameters are discrete which is equivalent to multiply the real earth model by a comb function. However, some errors are introduced and can be regarded as systematic errors.

The second limitation is that the seismic data is band-limited (Ghosh, 2000). This limitation can be explained by the convolution model. Following the convolution model, the discrete seismogram $\mathbf{s_k}$ can be formulated as,

$$s_k = r_k * w_k, \tag{4.1}$$

where r_k represents kth component of the primary reflectivity, w_k is kth component of the seismic wavelet and $*$ denotes the convolution operator. The band-limited nature is best illustrated in frequency domain. By forward Fourier transform, the above equation

is re-written as,

$$S(\omega_k) = R(\omega_k)W(\omega_k). \tag{4.2}$$

From the above equation, it is very clear that $S(\omega_k)$ will equal to zero when $R(\omega_k)$ or $W(\omega_k)$ is zero. Usually reflectivity series are broad-band, but source wavelet is band-limited. Thus, seismic data is band-limited. This band-limited nature is illustrated in Fig 4.1, where the source wavelet is a Berlage wavelet (Aldridge, 1990). It is clear that at least the zero frequency component of the data is lost. And the resolution of the data is limited because the highest frequency of the data is a finite number. Therefore, additional infor-

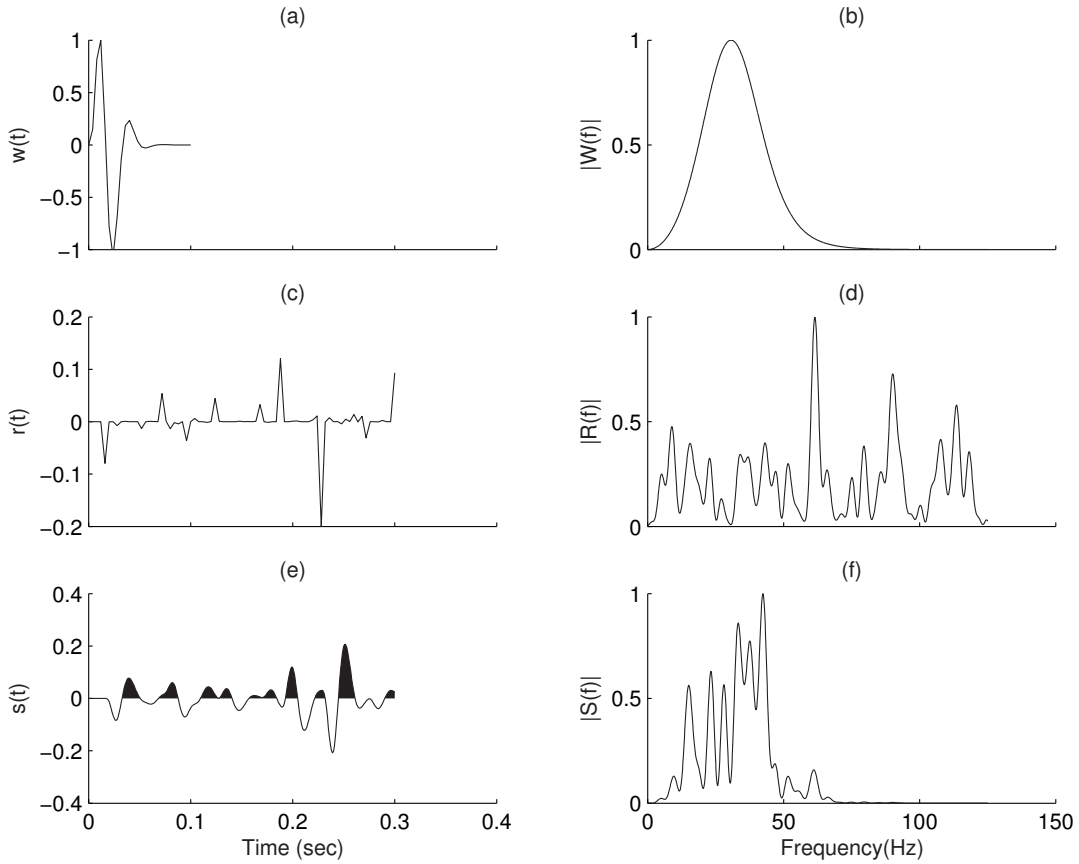

Figure 4.1: Band-limited nature of seismic data: (a) Berlage source wavelet, (c) a random sparse reflectivity series, (e) the corresponding seismogram, (b), (d) and (f) are the amplitude spectrum of (a), (c) and (e), respectively.

mation about the lost frequency component is needed. One approach is to assume some

characteristics of the model (i.e. model regularization methods) so that the lost frequency component can be recovered (Sacchi, 1997). One typical method for this approach is imposing a sparseness constraint. The sparseness constraint can be designed in two ways: first way is that we apply Bayes' theorem after assuming some probability distribution of the model parameters; second way is that we apply some norm to the model parameters directly.

The third limitation is that the seismic data are contaminated by noise. And the equation 4.1 can be rewritten as,

$$s_k = r_k * w_k + n_k, \tag{4.3}$$

where n_k is the noise. In the frequency domain, the equation 4.3 can be written as,

$$S(\omega) = R(\omega)W(\omega) + N(\omega). \tag{4.4}$$

From the equation 4.4, it is obvious that only dividing both sides by $W(\omega)$ will introduce noise component into the estimated model. The noise effect is shown in Fig. 4.2. Illustrated in this figure, it is clear that even a little noise will make the estimated solution unstable. The constraints imposed on the model parameters can stablize the solution. However, due to the noise, the value of the model parameters may be distorted. Therefore, an impedance constraint is often applied. Besides these methods, we can apply different norm of the data residual (i.e. the data regularization methods) by assuming the characteristics of the noise. The norm of the data residual can be given directly or derived in a Bayesian framework by assuming the probability distribution of the noise.

4.2 Sparseness Constraints

Different regularization strategies can be applied to improve the inversion (e.g. deconvolution) of seismic data. For simplicity, we only discuss the model regularization methods here. As we have said in the above section, there are two ways to design a sparseness constraint: first one is from a Bayes' viewpoint (Ulrych et al., 2001) and another one is from robust statistics' viewpoint (Sacchi, 1997). In this section, we will show the two ways and discuss the applications of sparseness constraint.

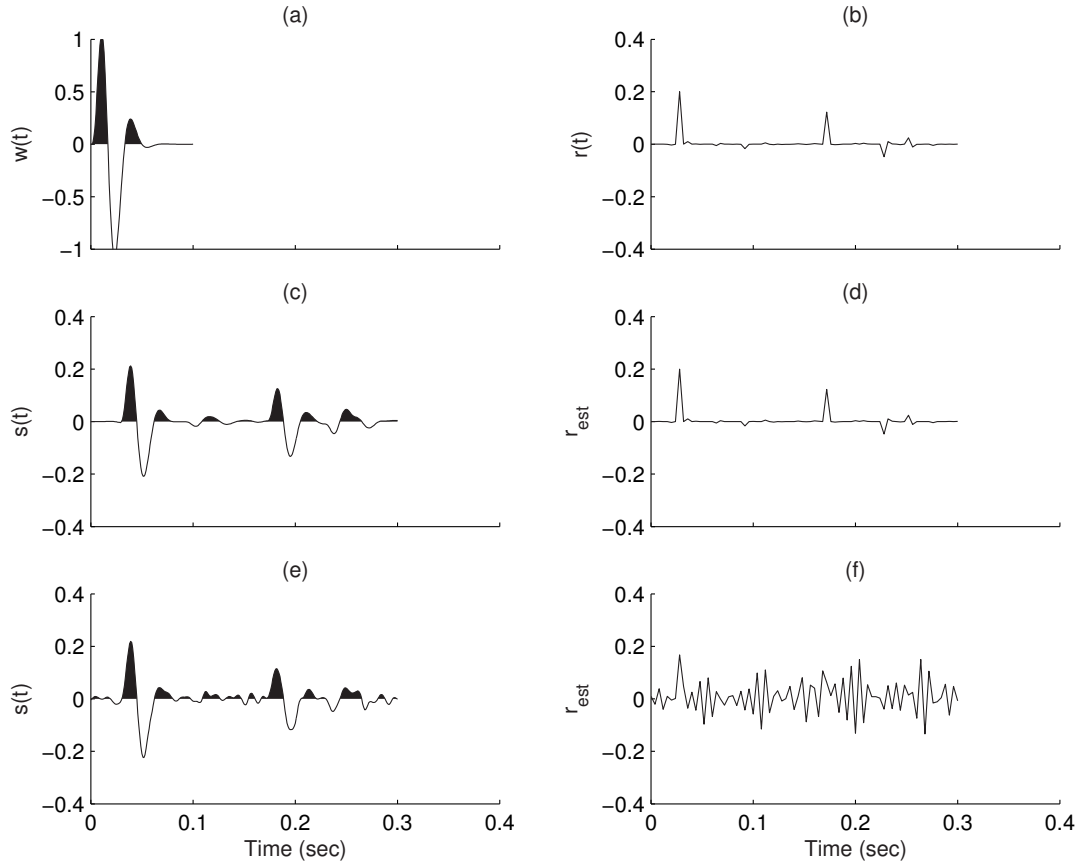

Figure 4.2: Noise effects on the inversion of seismic data: (a) a Berlage source wavelet; (b) a random sparse reflectivity series; (c) the corresponding noise free seismogram; (e) the corresponding seismogram with one percent Gaussian noise; (d) and (f) is the estimated reflectivity from the data shown in (c) and (e) , respectively.

4.2.1 Design of Sparseness Constraints

Bayes' Viewpoint

Recalling the Bayes' theorem 3.12 , we often use the form without the normalization factor. It can be written as,

$$P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{m})P(\mathbf{d}|\mathbf{m}). \tag{4.5}$$

For a noisy data set,

$$\begin{aligned} \mathbf{d}_{obs} &= \mathbf{d} + \mathbf{n} \\ &= \mathbf{Lm} + \mathbf{n}, \end{aligned} \tag{4.6}$$

where $\mathbf{n}$ is the noise vector. It is obvious the likelihood function $P(\mathbf{d}|\mathbf{m})$ follows the distribution of noise. This relationship can be stated as the following equation,

$$P(\mathbf{d}_{obs}|\mathbf{m}) = P(\mathbf{n}). \tag{4.7}$$

Many assumptions can be applied to the noise vector. The most common one is that the noise has a Gaussian distribution. The probability function for this case is

$$P(n_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}}\, e^{\frac{-(n_i-\mu n)^2}{2\sigma_i^2}}, \tag{4.8}$$

where i denotes the i_{th} component of the noise vector. The parameter μ_n is the mean value of the noise vector (we assume zero-mean noise, therefore $\mu_n = 0$) and σ^2 is the variance. Suppose the noise vector has N components which are uncorrelated, we can write the likelihood function in the following form,

$$\begin{aligned} P(\mathbf{n}) &= P(n_1)P(n_2)P(n_3)\dots P(n_N) \\ &= \left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}}\right)\, e^{\sum_{i=1}^{N} -\frac{(n_i-\mu n)^2}{2\sigma_i^2}} \\ &= \left((2\pi)^N \det \mathbf{C_d}\right)^{-\frac{1}{2}}\; e^{-\frac{1}{2}(\mathbf{d}-\mathbf{Lm})^T \mathbf{C_d}^{-1}(\mathbf{d}-\mathbf{Lm})}, \end{aligned} \tag{4.9}$$

where $\mathbf{C_d}$ is the data covariance operator. If the noise is uncorrelated, the covariance operator is given by

$$\mathbf{C_d} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & \dots & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_i^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & \dots & \sigma_N^2. \end{pmatrix} \tag{4.10}$$

The model vector $\mathbf{m}$ has M components. If we assume the model parameters are uncorrelated and follow a Gaussian distribution, we can follow the derivation shown above

and obtain the probability function of the model vector. This *priori* probability can be written as,

$$\begin{aligned} P(\mathbf{m}) &= P(m_1)P(m_2)P(m_3)\dots P(m_M) \\ &= \left(\prod_{i=1}^{M} \frac{1}{\sqrt{2\pi\sigma_{m_i}^2}}\right) e^{\sum_{i=1}^{M} -\frac{(m_i-\mu m)^2}{2\sigma_{m_i}^2}}, \\ &= \left((2\pi)^M \det \mathbf{C_m}\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\mathbf{m}^T {\mathbf{C_m}}^{-1}\mathbf{m}}, \end{aligned} \tag{4.11}$$

where, μ_m is the mean value of the model vector (we assume that μ_n is zero) and $\sigma_{m_i}^2$ is the variance. The matrix $\mathbf{C_m}$ is the model covariance operator and given by the following equation

$$\mathbf{C_m} = \begin{pmatrix} \sigma_{m_1}^2 & 0 & \dots & \dots & \dots & 0 \\ 0 & \sigma_{m_2}^2 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \sigma_{m_i}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & \dots & \sigma_{m_M}^2. \end{pmatrix} \tag{4.12}$$

According to Bayes' theorem, the posterior distribution will follow the equation shown below,

$$P(\mathbf{m}|\mathbf{d}) \propto \left((2\pi)^{M+N} \det \mathbf{C_d} \det \mathbf{C_m}\right)^{-\frac{1}{2}} e^{-\frac{1}{2}\left((\mathbf{d}-\mathbf{Lm})^T {\mathbf{C_d}}^{-1}(\mathbf{d}-\mathbf{Lm})+\mathbf{m}^T {\mathbf{C_m}}^{-1}\mathbf{m}\right)}. \tag{4.13}$$

The common way to choose a solution is to find a model vector which maximize the posterior probability $P(\mathbf{m}|\mathbf{d})$. This solution is called maximum a posterior(MAP) solution. To obtain MAP solution, we first take natural logarithm on both sides of the equation 4.13,

$$\begin{aligned} -\ln\left(P(\mathbf{m}|\mathbf{d})\right) &\propto \frac{1}{2}\left((\mathbf{d}-\mathbf{Lm})^T {\mathbf{C_d}}^{-1}(\mathbf{d}-\mathbf{Lm}) + \mathbf{m}^T {\mathbf{C_m}}^{-1}\mathbf{m}\right) \\ &+ \frac{1}{2}\ln\left((2\pi)^{M+N} \det \mathbf{C_d} \det \mathbf{C_m}\right). \end{aligned} \tag{4.14}$$

When $\mathbf{C_d}$ and $\mathbf{C_m}$ are given, the second term in the equation 4.14 is constant. Therefore, maximizing $P(\mathbf{m}|\mathbf{d})$ is equivalent to minimizing

$$\mathbf{J} = (\mathbf{d}-\mathbf{Lm})^T {\mathbf{C_d}}^{-1}(\mathbf{d}-\mathbf{Lm}) + \mathbf{m}^T {\mathbf{C_m}}^{-1}\mathbf{m}. \tag{4.15}$$

This equation can be reformulated in the form of the following cost function,

$$\mathbf{J}(\mathbf{m}) = (\mathbf{d} - \mathbf{Lm})^T {\mathbf{C_d}}^{-1} (\mathbf{d} - \mathbf{Lm}) + \mathbf{m}^T {\mathbf{C_m}}^{-1} \mathbf{m}. \tag{4.16}$$

For simplicity, we assume, for any i, $\sigma_i = \sigma_n$ and $\sigma_{m_i} = \sigma_m$. Therefore, we can rewrite the above equation as,

$$\mathbf{J}(\mathbf{m}) = (\mathbf{d} - \mathbf{Lm})^T (\mathbf{d} - \mathbf{Lm}) + \lambda \mathbf{m}^T \mathbf{m}, \tag{4.17}$$

where $\lambda = \frac{\sigma_n^2}{\sigma_m^2}$. This cost function is the same one for the common l_2 norm solution. From the above derivation, we can see that the l_2 norm solution is the one obtained by assuming the noise vector and model vector are the Gaussian distribution function whose mean value is zero and the variance is σ_n and σ_m respectively.

4.2.2 Sparse Solution

To obtain a sparse solution, two common prior distributions, which are exponential distribution and Cauchy distribution (Youzwishen, 2001), are used. The differences between the two distributions and the Gaussian distribution is shown in Figure 4.3. It is clear that exponential and Cauchy distributions have narrower peak than the Gaussian distribution and decline to zero more slowly. This kind of distribution is called long-tailed distribution (Lupton, 1993). Due to the narrower peak, most of the solution following the long-tailed distribution will be zero value. Meanwhile, in this case, the non-zero value will have bigger range of values than the Gaussian distribution. Therefore, comparing to Gaussian distribution, a long-tailed distribution such as the exponential and Cauchy distributions will force a solution in which most values are zeros and a few are large non-zero values (i.e. the sparse solution).

To derive the sparseness constraints for exponential and Cauchy distribution cases, we just follow the same procedure for the Gaussian distribution case (Zhang and Schmitt, 2004). For simplicity, the likelihood function $P(\mathbf{d}_{obs}|\mathbf{m})$ is always given by the equation 4.9.

First, let us suppose the model parameters follow the Cauchy distribution (Sacchi, 1997),

$$P(m_i) = \frac{1}{\pi \sigma_i} \frac{1}{1 + (m_i - \mu)^2 / \sigma_i^2}, \tag{4.18}$$

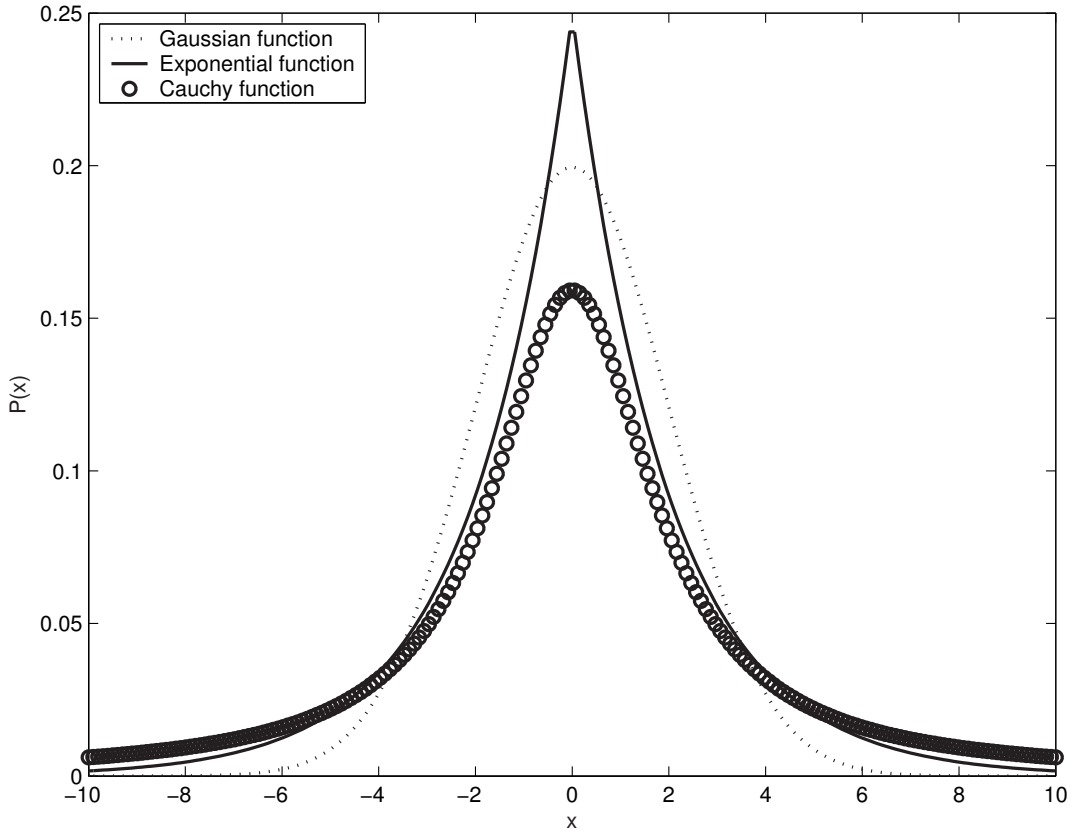

Figure 4.3: Gaussian, Exponential and Cauchy Distributions: mean value is 0 and the variance is 4

where μ is the mean value (we assume it to be zero here) and σ_i^2 is the variance for the ith component of the model vector m_i. Then we can derive the prior function $P(\mathbf{m})$,

$$P(\mathbf{m}) = \left(\pi^M \prod_{i=1}^{N} \sigma_i\right)^{-1} \prod_{i=1}^{N} \left(\frac{1}{1 + m_i^2/\sigma_i^2}\right). \tag{4.19}$$

The posteriori distribution then is

$$\begin{aligned} P(\mathbf{m}|\mathbf{d}) &\propto \left(\pi^M \prod_{i=1}^{N} \sigma_i\right)^{-1} \prod_{i=1}^{N} \left(\frac{1}{1+m_i^2/\sigma_i^2}\right) \\ &\times \left((2\pi)^N \det \mathbf{C_d}\right)^{-\frac{1}{2}} \; e^{-\frac{1}{2}(\mathbf{d}-\mathbf{Lm})^T \mathbf{C_d}^{-1}(\mathbf{d}-\mathbf{Lm})} \\ &\propto \left(\pi^M \prod_{i=1}^{N} \sigma_i\right)^{-1} \left((2\pi)^N \det \mathbf{C_d}\right)^{-\frac{1}{2}} \\ &\quad e^{\left(-\frac{1}{2}(\mathbf{d}-\mathbf{Lm})^T \mathbf{C_d}^{-1}(\mathbf{d}-\mathbf{Lm}) + \sum_{i=1}^{N}\left(-\ln(1+m_i^2/\sigma_i^2)\right)\right)}. \end{aligned} \tag{4.20}$$

For simplicity, we assume the components of diagonal matrix $\mathbf{C_d}$ are constant value and $\sigma_i = \sigma_m$ for all the components of the model vector. Therefore, by taking the negative natural log of the both sides, the MAP solution is to find the minimum of the following cost function (Youzwishen, 2001; Tarantola, 1987)

$$\mathbf{J} = (\mathbf{Lm} - \mathbf{d})^T (\mathbf{Lm} - \mathbf{d}) + \lambda \mathbf{R}(\mathbf{m}), \tag{4.21}$$

where λ is the trade-off parameter and the regularization term $\mathbf{R}(\mathbf{m})$ is

$$\mathbf{R}(\mathbf{m}) = \sum_{i=1}^{M} \ln(1 + m_i^2/\sigma_m^2). \tag{4.22}$$

Take the derivative of the cost function to each component of the model vector, we can obtain

$$\nabla \mathbf{J} = \mathbf{L}^T \mathbf{Lm} - \mathbf{L}^T \mathbf{d} + \lambda \nabla \mathbf{R}(\mathbf{m}), \tag{4.23}$$

where $\nabla \mathbf{R}(\mathbf{m})$ is given by

$$\nabla \mathbf{R}(\mathbf{m}) = \begin{pmatrix} \frac{\partial \mathbf{R}(\mathbf{m})}{\partial m_1} \\ \frac{\partial \mathbf{R}(\mathbf{m})}{\partial m_2} \\ \frac{\partial \mathbf{R}(\mathbf{m})}{\partial m_3} \\ \vdots \end{pmatrix}, \tag{4.24}$$

and we know that ith component of the above matrix is given by

$$\frac{\partial}{\partial m_i} \mathbf{R}(\mathbf{m}) = \frac{2}{\sigma_m^2} m_i \frac{1}{1 + m_i^2/\sigma_m^2}. \tag{4.25}$$

If we put the constant term $1/\sigma_m^2$ into the trade-off parameter λ, equation 4.24 can be

reformulated as

$$\nabla \mathbf{R}(\mathbf{m}) = \begin{pmatrix} 2m_1 \frac{1}{1+m_1^2/\sigma_m^2} \\ 2m_2 \frac{1}{1+m_2^2/\sigma_m^2} \\ 2m_3 \frac{1}{1+m_3^2/\sigma_m^2} \\ \vdots \end{pmatrix} \tag{4.26}$$

We can rewrite the above equation into the following matrix form

$$\nabla \mathbf{R}(\mathbf{m}) = 2\ \mathbf{Q}\ \mathbf{m}, \tag{4.27}$$

where $\mathbf{m}$ is the model vector and $\mathbf{Q}$ is a diagonal matrix whose ith diagonal element is given by

$$Q_{ii} = \frac{1}{1 + m_i^2/\sigma_m^2}. \tag{4.28}$$

If we let $\nabla \mathbf{J} = 0$ and thus the cost function is minimized, the solution would be the MAP solution . Therefore, it turns out to solve the following equation

$$\mathbf{L}^T\mathbf{L}\mathbf{m} - \mathbf{L}^T\mathbf{d} + \lambda\mathbf{Q}\mathbf{m} = 0. \tag{4.29}$$

Solving this equation, it is clear the estimated solution is

$$\hat{\mathbf{m}} = \left(\mathbf{L}^T\mathbf{L} + \lambda\mathbf{Q}\right)^{-1}\mathbf{L}^T\mathbf{d}, \tag{4.30}$$

where $\hat{\mathbf{m}}$ denotes the estimated solution and $\mathbf{Q}$ is the diagonal matrix whose diagonal elements is defined by the equation 4.28. This sparse solution corresponds to the assumption that the prior distribution of the model parameters is Cauchy distribution.

Another common sparseness constraint is l_1 norm (O'Brien et al., 1994; Taylor et al., 1979). This constraint can be derived by assuming the model vector follows a double exponential distribution. This distribution can be described mathematically as,

$$P(m_i) = \frac{1}{2\sigma_i}\ e^{-|m_i-\mu|/\sigma_i}, \tag{4.31}$$

where μ is the mean value and σ_i^2 is the variance of the ith component of the model vector. For simplicity, we assume the mean value is zero and the variance is a constant σ_m. Then the *priori* function $P(\mathbf{m})$ is

$$\begin{aligned} P(\mathbf{m}) &= P(m_1)P(m_2)P(m_3)\ldots P(m_M) \\ &= \frac{1}{(2\sigma_m)^M}\ e^{\left(-\frac{1}{\sigma m}\sum_{i=1}^{M}|m_i|_1\right)}. \end{aligned} \tag{4.32}$$

If we assume the noise follows a Gaussian distribution with a mean value of zero and the variance is a constant σ_n^2, then the posterior distribution will be defined as

$$P(\mathbf{m}|\mathbf{d}) \propto \frac{1}{(2\pi\sigma_N^2)^{N/2}} \frac{1}{(2\sigma_m)^M} \; e^{-\frac{1}{2\sigma_n^2}(\mathbf{Lm}-\mathbf{d})^T(\mathbf{Lm}-\mathbf{d}) - \frac{1}{\sigma m}\sum_{i=1}^{M}|m_i|_1}. \tag{4.33}$$

The MAP solution maximizes this posterior function. This maximization is equivalent to find the minimum of $-\ln P(\mathbf{m}|\mathbf{d})$, for given σ_n and σ_m, which is given by

$$-\ln P(\mathbf{m}|\mathbf{d}) = \frac{1}{2\sigma_n^2}(\mathbf{Lm}-\mathbf{d})^T(\mathbf{Lm}-\mathbf{d}) + \frac{1}{\sigma_m}\sum_{i=1}^{M}|m_i|_1. \tag{4.34}$$

This equation can also be written in the form of a cost function

$$\mathbf{J} = (\mathbf{Lm}-\mathbf{d})^T(\mathbf{Lm}-\mathbf{d}) + \lambda\mathbf{R}(\mathbf{m}), \tag{4.35}$$

where λ is the trade-off parameter. The regularization term $\mathbf{R}(\mathbf{m})$ is

$$\mathbf{R}(\mathbf{m}) = \sum_{i=1}^{M}|m_i|_1. \tag{4.36}$$

Minimizing this cost function is equivalent to find the solution to $\nabla J = 0$. We know the derivative of the regularization term is

$$\frac{\partial}{\partial m_i}\mathbf{R}(\mathbf{m}) = \frac{m_i}{|m_i|}. \tag{4.37}$$

Thus

$$\nabla\mathbf{R}(\mathbf{m}) = \begin{pmatrix} \frac{m_1}{|m_1|} \\ \frac{m_2}{|m_2|} \\ \frac{m_3}{|m_3|} \\ \vdots \end{pmatrix}. \tag{4.38}$$

This gradient term can also be expressed in a matrix form

$$\nabla\mathbf{R}(\mathbf{m}) = \mathbf{Q}\;\mathbf{m}, \tag{4.39}$$

where $\mathbf{Q}$ is a diagonal matrix whose diagonal element is defined by

$$Q_{ii} = \frac{1}{|m_i|}. \tag{4.40}$$

Considering the stability of $\mathbf{Q}$, this matrix is modified to

$$Q_{ii} = \begin{cases} |m_i|^{-1} & \text{if } |m_i| > \epsilon \\ \epsilon^{-1} & \text{if } |m_i| \leq \epsilon \end{cases} \tag{4.41}$$

where ϵ is a small, positive number. So when m_i is very small, the corresponding element of Q_{ii} is a certain number which approximate the exact number good enough instead of an unstable number. Solving this cost function, the solution $\hat{\mathbf{m}}$ is

$$\hat{\mathbf{m}} = \left(\mathbf{L}^T\mathbf{L} + \lambda\mathbf{Q}\right)^{-1}\mathbf{L}^T\mathbf{d}, \tag{4.42}$$

where $\mathbf{Q}$ is defined by the equation 4.41. We find this equation 4.42 and 4.30 is the same except the matrix $\mathbf{Q}$ is different. Therefore, both cost functions can be solved using the IRLS algorithm discussed in the chapter 3.

Robust Statistics' Viewpoint

Considering the problem $\mathbf{d} \approx \mathbf{Lm}$, the cost function, in the least square sense, can be rewritten as (Sacchi, 1997)

$$\mathbf{J}_r = \sum_{i=1}^{N} \rho_1\left(\frac{r_i}{\sigma_i}\right), \tag{4.43}$$

where r_i is the ith component of the residual vector $\mathbf{r}$ which is defined as

$$\mathbf{r} = \mathbf{d} - \mathbf{Lm}, \tag{4.44}$$

and the function $\rho_1(u)$ is

$$\rho_1(u) = \frac{1}{2}u^2, \tag{4.45}$$

where u is the weighted residual r_i/σ_i. For simplicity we assume that σ_i is equal to a constant σ_m for $i = 1, 2, \ldots, N$. To minimize this cost function, we solve the following equation,

$$\nabla\mathbf{J}_r = \sum_{i=1}^{N} \psi(r_i/\sigma_n) = 0, \tag{4.46}$$

where $\psi(u) = \partial\rho_1(u)/\partial u$. This function measure the influence of the residuals on the model estimation. In robust statistics, it is called influence function. In practice, another regularization term $\mathbf{J_m}$ is needed besides the cost function $\mathbf{J}_r$ to stabilize the solution.

$$\mathbf{J} = \sum_{i=1}^{N} \rho_1\left(\frac{r_i}{\sigma_i}\right) + \mathbf{J_m}, \tag{4.47}$$

where $\mathbf{J_m}$ is defined as

$$\mathbf{J_m} = \sum_{i=1}^{M} \rho(m_i/\sigma_{m_i}), \tag{4.48}$$

where $\sigma_{m_i}^2$ is the variance of the ith component of the model vector. If we assume the function $\rho(u) = \rho_1(u)$, then the solution to this cost function is the common l_2 norm solution.

In robust statistics, a common approach to re-evaluate the influence of the outlier is to redesign the function $\mathbf{J}_r$. A similar approach can be used to redesign the regularization term $\mathbf{J_m}$. To enforce a sparse solution, four regularization functions are often used: l_p norm, Sech norm, Cauchy norm and Huber norm (Sacchi et al., 2003). The four functions are defined as,

$$\rho_p(u) = |u|^p, \tag{4.49}$$

$$\rho_{Sech}(u) = \ln\left(cosh(u)\right), \tag{4.50}$$

$$\rho_{Cauchy}(u) = \ln\left(1 + u^2/2\right), \tag{4.51}$$

$$\rho_{Huber}(u) = \begin{cases} u^2/2 & \text{if } |u| \leq a \\ a|u| - a^2/2 & \text{if } |u| > a, \end{cases} \tag{4.52}$$

where $cosh(u) = \frac{e^u + e^{-u}}{2}$ and a is a given positive number. The influence functions for the four norm are,

$$\psi_p(u) = p\ u/|u|^{(2-p)}, \tag{4.53}$$

$$\psi_{Sech}(u) = \frac{e^u - e^{-u}}{e^u + e^{-u}}, \tag{4.54}$$

$$\psi_{Cauchy}(u) = \frac{u}{(u^2/2) + 1}, \tag{4.55}$$

$$\psi_{Huber}(u) = \begin{cases} u & \text{if } |u| \leq a \\ asign(u) & \text{if } |u| > a. \end{cases} \tag{4.56}$$

The new system of equations after regularization can be written in the following form,

$$(\mathbf{L}^T\mathbf{L} + \lambda\mathbf{Q})\mathbf{m} = \mathbf{L}^T\mathbf{d}. \tag{4.57}$$

The diagonal matrix $\mathbf{Q}$ is different for different regularization functions. For simplicity, we try to let the constant value such as p, σ_m and σ_n absorbed in the trade-off parameter

λ. The diagonal elements of $\mathbf{Q}$ are

$$Q_{p_{ii}} = 1/|m_i|^{(2-p)}, \tag{4.58}$$

$$Q_{Sech_{ii}} = \frac{e^{m_i/\sigma_m} + e^{-m_i/\sigma_m}}{m_i(e^{m_i/\sigma_m} + e^{-m_i/\sigma_m})}, \tag{4.59}$$

$$Q_{Cauchy_{ii}} = \frac{1}{m_i^2 + \epsilon}, \tag{4.60}$$

$$Q_{Huber_{ii}} = \begin{cases} 1 & \text{if } |\frac{m_i}{\sigma_m}| \leq a \\ \frac{a}{|m_i|} & \text{if } |\frac{m_i}{\sigma_m}| > a \end{cases} \tag{4.61}$$

where a and ϵ are given positive number. If we compare the Cauchy norm and l_p norm(here $p = 1$) to the results derived above by assuming *priori* distribution is Cauchy or double exponential distribution, we can find both viewpoints give the same result. Therefore, Bayesian viewpoint and robust statistics viewpoint are the two different ways of presenting the same problem.

4.2.3 Example of Sparseness Constraints

As we have discussed in the above subsection, an inverse problem constrained by a sparseness constraint can be formulated in the form of the equation 4.57 and its solution is given by the equation 4.42. This equation can be solved using IRLS algorithm. However, there are two problems: first, the selection of the trade-off parameter λ; and second, the selection of the regularization functions.

Figure 4.4 (a) shows a sparse reflectivity series, its corresponding noisy synthetic trace and the source wavelet. The noise is Gaussian noise and the signal to noise ratio (snr) is ten. First, we do the deconvolution using damping (also named as zero-order) regularization. The result is shown in Figure 4.4 (b). Then we applied the Cauchy norm and l_1 norm to estimate the reflectivity. Figure 4.4 (c)and (d) illustrates the results for the two regularization terms respectively. As shown in the above figure, sparseness constraints such as Cauchy norm and l_1 norm give better results than that by means of zero-order regularization. This shows that a sparseness constraint will improve the result if the model follow the sparseness assumption. On the other hand, the estimated result using Cauchy norm looks a little better than that using l_1 norm. This shows different sparseness constraints will give different results. Therefore, it is necessary to consider

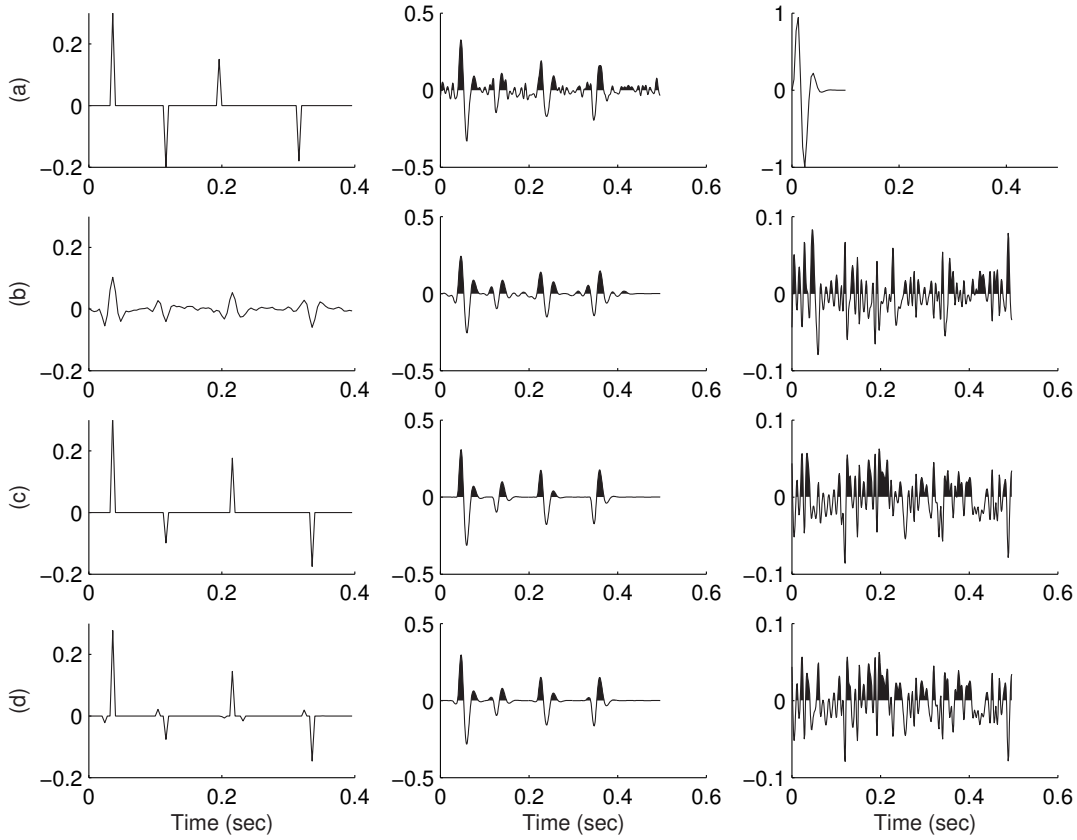

Figure 4.4: (a) A sparse reflectivity (left), synthetic trace (center) and a Berlage source wavelet (right). (b) The estimated reflectivity series using zero-order regularization (left), predicted trace (center) and residuals (noisy synthetic trace minus the predicted trace). (c) Results using the Cauchy norm regularization and (d) results using the l_1 norm.

the selection of the sparseness constraint for different data sets. From Figure 4.4, we can also find that the estimated reflectivity series by means of sparseness constraints do not give us correct relative amplitude for these reflectivities. This can be partially explained by the design of the sparseness constraint. For example, we know the diagonal element of weighting matrix $\mathbf{Q}_{Cauchy}$ is $\frac{1}{m_i^2+\epsilon}$. Theoretically, the weights for the two spiking elements, for example m_k, m_j, $\frac{1}{m_j^2+\epsilon}$ and $\frac{1}{m_k^2+\epsilon}$ should be equal in the sense of physical meaning because the two model parameters are the exact solutions. However, it is obvious that for the two spike model parameter m_k, m_j, the corresponding weights for the two elements $\frac{1}{m_j^2+\epsilon}$ and $\frac{1}{m_k^2+\epsilon}$ will not be equal unless $|m_j| = |m_k|$.

An advantage of the sparseness constraint is that it can recover the lost high and low frequency component of the original model. This effect is best illustrated in Figure 4.5.

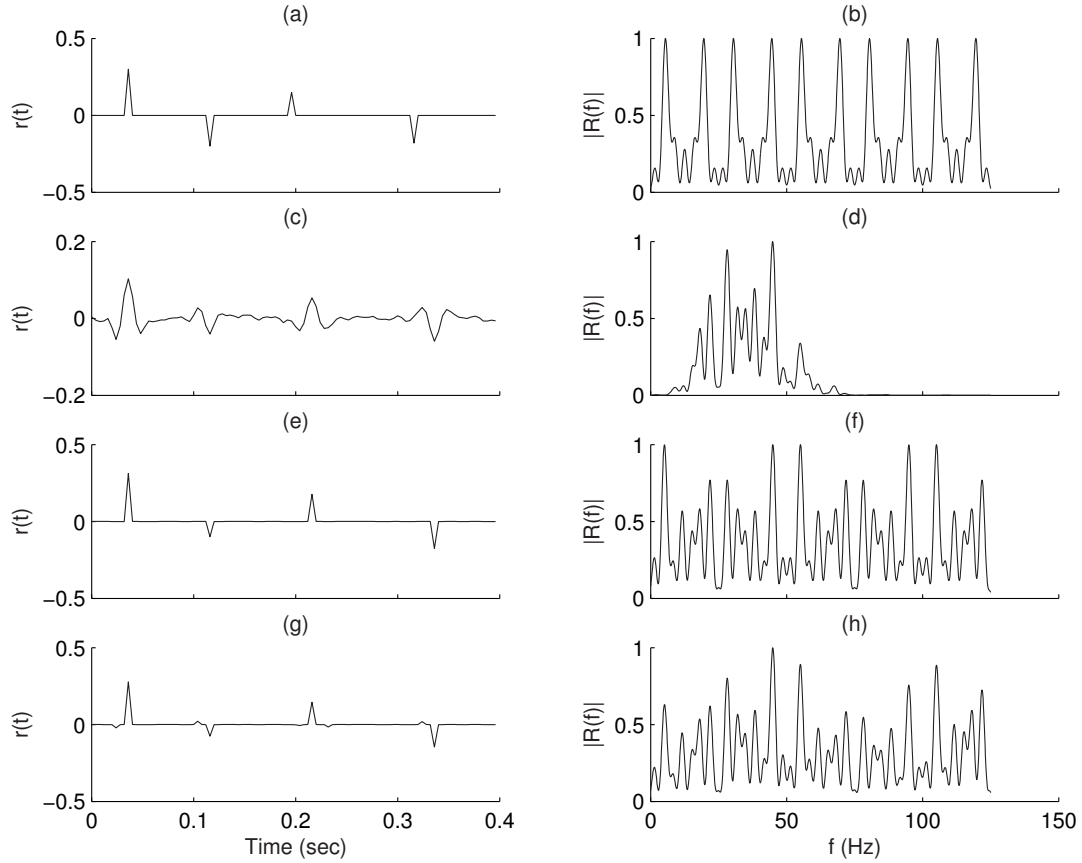

Figure 4.5: (a) True reflectivity series, (b) the power spectrum of the true model, (c) estimated reflectivity series using zero-order regularization, (e) and (f) show the estimated reflectivity series using the Cauchy norm and the l_1 norm respectively. (d), (f) and (h) are corresponding power spectrum to (b), (e) and (g) respectively.

4.3 Impedance Constraints

Well-log data is an important source of subsurface information. The impedance derived from the well-log data can be used to constrain the estimated subsurface seismic properties. In this section, we will first relate the impedance obtained from well-log to the reflectivity and construct the impedance constraint. Then we will show that an impedance

constraint can improve the inversion (Oldenburg et al., 1983; Zhang and Schmitt, 2004).

4.3.1 Derivation of Impedance Constraints

For simplicity, we assume the earth is consisted of k parallel layers and each layer's properties are constant as shown in the Figure 4.6. The acoustic impedance ξ_k (Telford et al., 1990) is defined as

$$\xi_k = \rho_k v_k. \tag{4.62}$$

Thus we can calculate the acoustic impedance from the sonic and density well-log data.

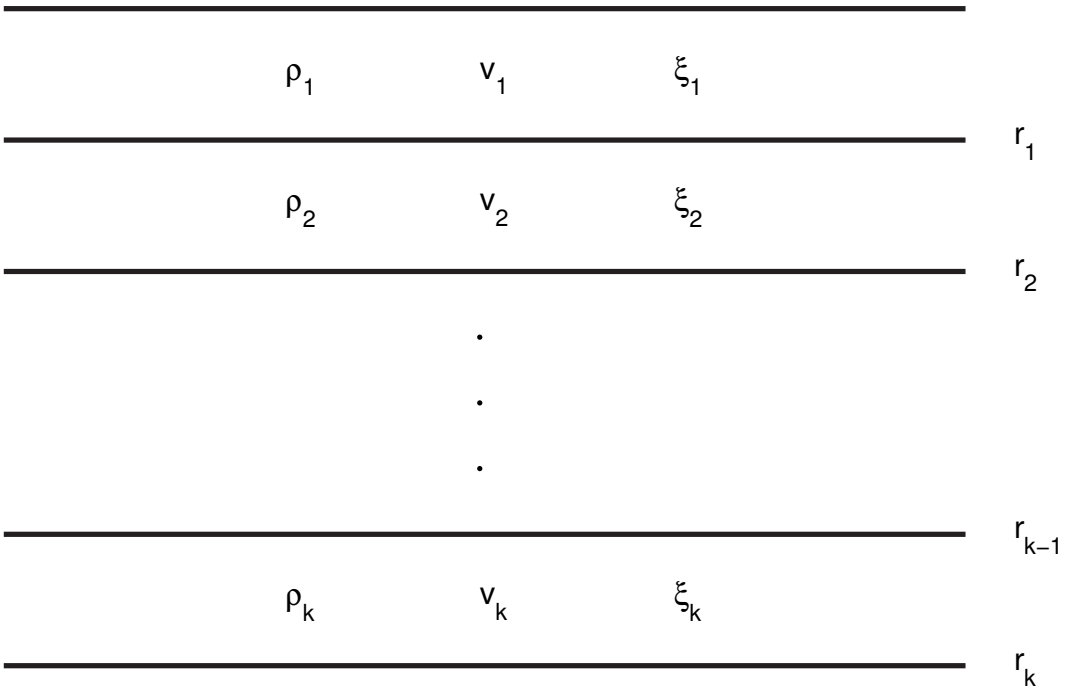

Figure 4.6: The layered earth model. r_k denotes the reflection coefficient of the bottom of the kth layer. ρ_k, v_k and ξ_k represents the density, velocity and acoustic impedance of the the kth layer respectively.

By a convolution model $\mathbf{s} = \mathbf{Wr}$ (where $\mathbf{s}$ which represents the seismic data $\mathbf{d}$ is a seismic trace, $\mathbf{W}$ which denotes the forward operator $\mathbf{L}$ is a convolution matrix and $\mathbf{r}$ which denotes the model $\mathbf{m}$ is a reflectivity series.), we can relate the seismic data to the reflectivity coefficient. Meanwhile, for normal incidence case, the reflection coefficient r_k

(Telford et al., 1990) is given by

$$r_k = \frac{\xi_{k+1} - \xi_k}{\xi_{k+1} + \xi_k}. \tag{4.63}$$

Rearranging the above equation, we get

$$\xi_{k+1} = \xi_k \left(\frac{1 + r_k}{1 - r_k} \right) = \xi_1 \prod_{i=1}^{k} \left(\frac{1 + r_i}{1 - r_i} \right). \tag{4.64}$$

By this equation, we can relate the acoustic impedance to the reflectivity coefficients. In practice, the reflection coefficients and the acoustic impedance are related by an approximation (Peterson et al., 1955),

$$r(t) = \frac{1}{2} \frac{d[\ln \xi(t)]}{dt}. \tag{4.65}$$

The equation 4.65 is valid if the impedance $\xi(t)$ is continuous. However, the error can be neglected if the discrete jumps (reflection coefficients) in $\xi(t)$ are less than 0.3 (Oldenburg et al., 1983). The equation 4.65 can be rewritten as

$$\ln[\xi(t)/\xi(0)] = 2 \int_0^t r(u) du, \tag{4.66}$$

where $\xi(0) = \xi_1$ is the impedance of the first layer and r is the reflectivity. In real world, r is replaced by the estimated $\hat{r}$ and a summation is used instead of the integral. The equation 4.66 is reformulated as

$$\sum_{i=1}^{k} r_i = \frac{1}{2} \ln[\xi_i/\xi_1], \tag{4.67}$$

where ξ_i is the acoustic impedance of the ith layer. We can rewrite this equation in a matrix form

$$\mathbf{Cr} = \beta, \tag{4.68}$$

where $\mathbf{C}$ is a summation matrix and β is a column vector whose element is given by the right-hand side of equation 4.67. In other words, equation 4.67 represents one row of equation 4.68. The information provided by equation 4.68 is imposed into the inverse problem as the impedance constraint. If we assume the variance of the well-log derived impedance is a constant, the impedance constraint in the form of a cost function is

$$\mathbf{J}_\xi = ||\mathbf{Cr} - \beta||_2^2, \tag{4.69}$$

and the final cost function becomes

$$\begin{aligned} \mathbf{J} &= \mathbf{J}_r + \lambda \mathbf{J}_\xi & (4.70) \\ &= ||\mathbf{s} - \mathbf{W}\mathbf{r}||_2^2 + \lambda ||\mathbf{C}\mathbf{r} - \beta||_2^2. & (4.71) \end{aligned}$$

In this cost function 4.71, other constraints such a sparseness constraint are not included.

4.3.2 Example of Impedance Constraints

As we discussed in the above section, a sparseness constraint cannot ensure a correct relative amplitude of the solution. Therefore, an impedance constraint is often applied to force a correct relative magnitude of the solution.

The comparison between the results by means of a Cauchy norm constraint and a combination of Cauchy norm and an impedance constraint is illustrated in the following Figure 4.7. It is obvious that the impedance constraint greatly improve the accuracy of the estimated impedance. The comparison between the power spectrum of the estimated model is shown in Figure 4.8. We can easily find that the impedance constraints improved the results.

4.4 Structural Constraint

One of the most important characteristics of the time-lapse data is that the underlying model background does not change outside the reservoir. Therefore, a structural constraint can be designed for the time-lapse data to constrain the model difference caused by the reservoir changes. In this section, we will discuss the design of structural constraint and then the difficulties related to this structural constraint.

4.4.1 Derivation of Structural Constraint

For simplicity we assume the time-lapse seismic data have two data sets which are

$$\begin{aligned} \mathbf{d}_b &= \mathbf{L}_b \mathbf{m}_b + \mathbf{n}_b \\ \mathbf{d}_s &= \mathbf{L}_s \mathbf{m}_s + \mathbf{n}_s, & (4.72) \end{aligned}$$

where subscript b means the base survey and s denotes the second survey. $\mathbf{d}$, $\mathbf{L}$, $\mathbf{m}$, $\mathbf{n}$, represents the seismic data, forward operator, model parameters and noise respectively.

In time-lapse seismic inversion case, we need to clarify two things: first, each data set of time-lapse data is not different from ordinary seismic data set; second, the objective of time-lapse inversion is to find the model changes due to the reservoir changes instead of the model parameters. The first thing implies that we can use all the existing seismic inversion methods for time-lapse data. The second thing tells us that extra efforts are needed to attenuate the unwanted model changes which can be caused by inconsistent inversion of time-lapse data and non-repeatability of the time-lapse data sets. To attenuate the none-repeatability effects, people usually apply cross-equalization techniques. It must be emphasised the importance of the uniform processing for time-lapse data. The basic idea of the cross-equalization technique is: the matching of the data above the reservoir will force the forward operators to be the same because the corresponding models do not change during time-lapse acquisition process. The problems of cross-equalization are: first, the matching the high resolution data to the low resolution data and thus high frequency information is lost; second, the noise in the two data sets are different so that the cross-equalization might match the noise. Since seismic processing can be regarded as the first iteration of an seismic inversion, one idea is to construct some constraint which acts like the cross-equalization technique. This constraint is called the structural constraint which constraints the model differences. The advantages of the structural constraint are: first, it does not require the same forward operator for the time-lapse data sets; second, it is more flexible to handle the inconsistent noise of the data sets.

The structural constraint can be designed in a Bayesian framework. For simplicity, we assume the noise and the model for the two time-lapse data sets both follow a Gaussian distribution whose mean value is zero and the variance is σ_n and σ_m respectively. From a Bayes' viewpoint, one approach of the time-lapse inversion is try to find the probability function $P(\mathbf{m}_b, \mathbf{m}_s|\mathbf{d}_b, \mathbf{d}_s)$.

$$P(\mathbf{m}_b, \mathbf{m}_s|\mathbf{d}_b, \mathbf{d}_s) \propto P(\mathbf{d}_b, \mathbf{d}_s|\mathbf{m}_b, \mathbf{m}_s)P(\mathbf{m}_b, \mathbf{m}_s), \tag{4.73}$$

where $P(,)$ means the joint probability. Since we know the likelihood function is equivalent to the probability function of the noise and the noise of the two time-lapse data sets

are independent, $P(\mathbf{d}_b, \mathbf{d}_s|\mathbf{m}_b, \mathbf{m}_s)$ can be given by

$$\begin{aligned} P(\mathbf{d}_b, \mathbf{d}_s|\mathbf{m}_b, \mathbf{m}_s) &= P(\mathbf{d}_b|\mathbf{m}_b, \mathbf{m}_s)P(\mathbf{d}_s|\mathbf{m}_b, \mathbf{m}_s) \\ &= P(\mathbf{d}_b|\mathbf{m}_b)P(\mathbf{d}_s|\mathbf{m}_s). \end{aligned} \tag{4.74}$$

However, the models for the two data sets are related by

$$\mathbf{m}_s = \mathbf{m}_b + \delta\mathbf{m}, \tag{4.75}$$

where $\delta\mathbf{m}$ follows some distribution. For simplicity, we assume it is a Gaussian distribution with a mean value zero and the variance is $\sigma_{\delta mi}$ for the ith component of $\delta\mathbf{m}$. We can know easily that the condition probability $P(\mathbf{m}_b|\mathbf{m}_s)$ or $P(\mathbf{m}_s|\mathbf{m}_b)$ is equivalent to the probability function $P(\delta\mathbf{m})$. After defining the relationship between the two models, the joint probability $P(\mathbf{m_b}, \mathbf{m_s})$ is

$$P(\mathbf{m}_b, \mathbf{m}_s) = P(\mathbf{m}_b)P(\mathbf{m}_s|\mathbf{m}_b) = P(\mathbf{m}_s)P(\mathbf{m}_b|\mathbf{m}_s). \tag{4.76}$$

The equation 4.76 shows two expressions for $P(\mathbf{m}_b, \mathbf{m}_s)$. Thus, the posterior probability distribution can be written in two ways

$$P(\mathbf{d}_b, \mathbf{d}_s|\mathbf{m}_b, \mathbf{m}_s) \propto P(\mathbf{d}_b|\mathbf{m}_b)P(\mathbf{d}_s|\mathbf{m}_s)P(\mathbf{m}_b)P(\mathbf{m}_s|\mathbf{m}_b), \tag{4.77}$$

$$P(\mathbf{d}_b, \mathbf{d}_s|\mathbf{m}_b, \mathbf{m}_s) \propto P(\mathbf{d}_b|\mathbf{m}_b)P(\mathbf{d}_s|\mathbf{m}_s)P(\mathbf{m}_s)P(\mathbf{m}_b|\mathbf{m}_s). \tag{4.78}$$

The MAP solution is the solution that maximizes $P(\mathbf{m}_b, \mathbf{m}_s|\mathbf{d}_b, \mathbf{d}_s)$, that is, minimizes $-\ln(P(\mathbf{m}_b, \mathbf{m}_s|\mathbf{d}_b, \mathbf{d}_s))$. Following the derivation process for the sparseness constraint, we can obtain the two cost functions for the time-lapse inversion,

$$\begin{aligned} \mathbf{J_b} &= ||\mathbf{d}_b - \mathbf{L}_b\mathbf{m}_b||_2^2 + ||\mathbf{d}_s - \mathbf{L}_s\mathbf{m}_s||_2^2 \\ &+ \lambda_1||\mathbf{m}_b||_2^2 + \lambda_2||\mathbf{W}_c(\mathbf{m}_b - \mathbf{m}_s)||_2^2, \end{aligned} \tag{4.79}$$

$$\begin{aligned} \mathbf{J_s} &= ||\mathbf{d}_b - \mathbf{L}_b\mathbf{m}_b||_2^2 + ||\mathbf{d}_s - \mathbf{L}_s\mathbf{m}_s||_2^2 \\ &+ \lambda_3||\mathbf{m}_s||_2^2 + \lambda_4||\mathbf{W}_c(\mathbf{m}_b - \mathbf{m}_s)||_2^2, \end{aligned} \tag{4.80}$$

where λ_1, λ_2, λ_3 and λ_4 are trade-off parameters. The term $||\mathbf{W}_c(\mathbf{m}_b - \mathbf{m}_s)||_2^2$ is the structural constraint for time-lapse seismic inversion. The matrix $\mathbf{W}_c$ is a diagonal matrix whose diagonal elements is

$$\mathbf{W}_{c_{ii}} = \frac{1}{\sigma_{\delta mi}}. \tag{4.81}$$

It is clear that this matrix $\mathbf{W}_c$ determines how the structural constraint works. Therefore the most important part of the design of the structural constraints is to design the matrix $\mathbf{W}_c$. One simple $\mathbf{W}_c$ can be designed as

$$diag(\mathbf{W}_{c_{ii}}) = \begin{cases} \beta & \text{Area of no change.} \\ 1 & \text{Area of changes.} \end{cases}, \tag{4.82}$$

where β is a positive big number. Depending on the matrix $\mathbf{W}_c$, the final cost function can be $\mathbf{J_b}$, $\mathbf{J_s}$ or $\mathbf{J_b} + \mathbf{J_s}$.

Similar to the design of the sparseness constraint, different structural constraint can be constructed by assuming different distribution of model differences or directly changing the norm of the model difference. However, the weighting matrix $\mathbf{W}_c$ is the most critical part for all kinds of structural constraint. Therefore, it is important to investigate the characteristics of the model difference, which depends on the domain in which the inversion applied and the processing of the data before the inversion, to design a suitable $\mathbf{W}_c$.

4.4.2 The Difficulties of the Structural Constraint

The weighting matrix $\mathbf{W}_c$ in the structural constraint tells us which model parameter is most probably changed. Therefore a very accurate $\mathbf{W}_c$ will give us a better result. The difficulties of the structural constraint mainly comes from the design of this weighting matrix.

The first and fundamental difficulty is that there is no theoretical approach to calculate the weighting matrix. The good thing of time-lapse case is that we know a blur range of the area where the model does not change. Therefore, the advantage of the time-lapse case is that at least we know a blur picture of where the variance should be higher. However, we do not know the variance for each component of the wanted model difference and thus this difficulty still exists.

After we know the first difficulty, we may have an idea whether we can design the weighting matrix in such a way that the variance of the wanted model difference can be iteratively improved by the estimated result after attenuating the unwanted model difference. However, it is hard to design the weighting matrix so that the attenuation of unwanted model difference can effectively improve the inversion of the model parame-

ters which do changed. This difficulty means that the predictability of this constraint is pretty weak and thus greatly limit the usefulness of the structural constraint.

The third difficulty comes from the fact that the model parameters below the reservoir will have a time-shift due to the reservoir change in time domain (Pepper et al., 1997; Watson, 2004). Generally the reservoir changes have two influences on the time-lapse data: amplitude changes and time-shift. The amplitude changes are caused by the impedance (which includes the effects of the velocity and density) changes. The time-shift of the model parameters below the reservoir is caused by the velocity change of the above layers. Theoretically both quantities can be used to estimate the reservoir changes. This time-shift effects is illustrated in Figure 4.4.2. From this figure, we can find that even a small time-shift will cause a normalized root mean square amplitude (NRMS) difference (Kragh and Christie, 2002) which is bigger than common time-lapse NRMS change. Therefore, it is important to know the exact time-shift to remove the data difference caused by the time-shift. However, it is very difficult to obtain an accurate estimation of the time-shift because many factors can affect the time-shift. Therefore, in time domain, we use $\mathbf{m}_b(i) - \mathbf{m}_s(i)$ instead of $\mathbf{m}_b(i) - \mathbf{m}_s(i+\tau_i)$ (where τ_i is the time-shift for the ith component of the model, m_i) for the structural constraint. This cause another problem: the constraint will affect the result less because all the model parameters below the reservoir have a time-shift and there is no way to obtain accurate weights for those model parameters.

If we do the inversion in depth domain, we will find that the uncertainty of the input velocity model will limit the usage of this structural constraint. The example shown in Figure 4.4.2 and Figure 4.4.2 illustrates that how the design of the structural constraint affects the final inversion results. In Figure 4.4.2, the data difference and the real reflectivity difference are given. The inversion results illustrated in Figure 4.4.2 demonstrate that the more accurate structural constraint the better the results.

4.5 Constraints for the Time-lapse Case

For the time-lapse data, there are two problems to consider when we apply constraints. First, should we apply the constraints to separate data sets or data differences? Second,

will the constraint be different for different inversion schemes?

We can apply sparseness constraint and impedance constraint to the case of separate data sets and the data difference case. The difference between the two cases for the sparseness is that whether we assume the model or the model differences follow some probability distribution or some characteristic. For impedance constraint, there is no difference between the two cases. As to the structural constraint, it is always applied to the data difference.

In the following chapter, we will discuss three time-lapse inversion schemes: first, invert the data sets separately and then subtract them to obtain the model differences; second, invert the data differences to obtain the model difference and third, simultaneously invert the data sets to obtain the model differences. For each scheme, the sparseness constraint and impedance constraint is designed according to whether the data sets or data differences are inverted and the form of the two constraints will be the same. If we apply the structural constraint for the three inversion schemes, the structural constraint will be given in the following cost function form

$$\mathbf{J_{str}} = \begin{cases} ||\mathbf{W}_c(\mathbf{m}_b - \mathbf{m}_s)||_2^2 & \text{The scheme that simultaneous invert the data sets} \\ ||\mathbf{W}_c(\mathbf{m}_s - \hat{\mathbf{m}}_b)||_2^2 & \text{The scheme that invert the data sets separately} \\ ||\mathbf{W}_c\delta\mathbf{m}||_2^2 & \text{The scheme that invert the data differences} \end{cases}, \tag{4.83}$$

where $\hat{\mathbf{m}}_\mathbf{b}$ is the estimated base model without using a structural constraint and $\delta\mathbf{m}$ is the model differences. From the above equation, it is clear the key point of the structural constraint is the weighting matrix $\mathbf{W}_c$. Figure 4.5 illustrates the results using these constraints under different schemes. We can find the structural constraint can attenuate the unwanted model difference where the prior information tells us that no wanted model difference should exist. The impedance constraint can give a better relative magnitude of the model parameters.

4.6 Summary

In this chapter, I first reviewed the limitations of the seismic data: the data is finite while the earth reflectivity is continuously distributed, the data is band-limited and the data contain noise. Then I provide most common constraints to improve the inversion of

the seismic data: sparseness constraints and impedance constraints. In time-lapse case, another constraint called the structural constraint is introduced.

Two ways to derive the sparseness constraints are provided: the Bayesian framework and the robust statistics theory. Both ways give us the same regularization terms for the same *priori* information. However, the sparseness constraints cause amplitude scale problems and thus the impedance constraint is introduced. Based on the relationship between the reflectivity series and the impedance log, the impedance constraint is derived. Then the structural constraint is derived in the Bayesain framework for time-lapse data. The simple examples show that the design of the structural constraint is an important problem unsolved. Finally, I test the constraints and their combination on simple time-lapse data example and the results show that the combination of the three constraints provide the best result comparing to other combination.

Discussions in this chapter imply that a good time-lapse inversion requires more research on the constraints applied to the time-lapse seismic data. On the one hand, the three constraints still have many things to investigate to improve their performance in the time-lapse case.

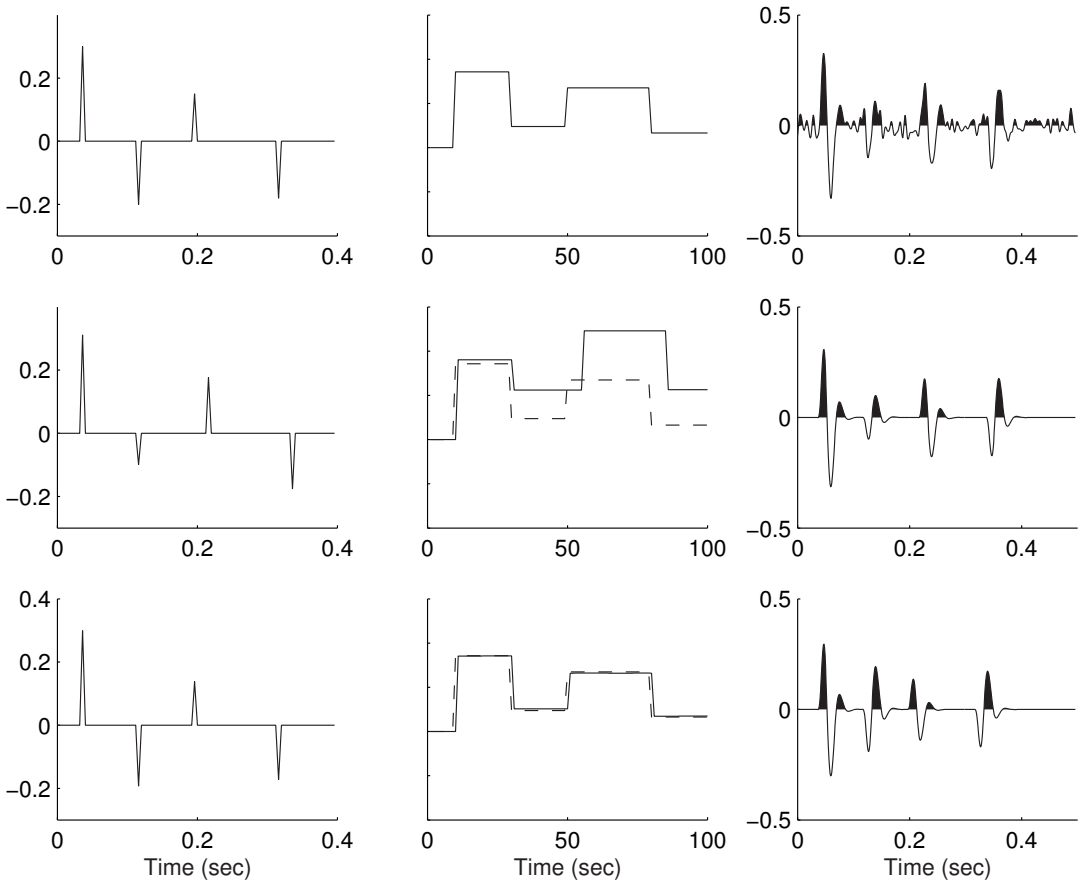

Figure 4.7: (a) A sparse reflectivity (left), true impedance (centre), noisy synthetic trace (snr= 10, right). (b) The estimated reflectivity series using Cauchy norm regularization (left), estimated impedance (centre) and predicted data. (c) The estimated reflectivity series using Cauchy norm regularization and an impedance constraint (left), estimated impedance (centre) and predicted data. The dashed line in (b) and (c) indicates the true impedance.

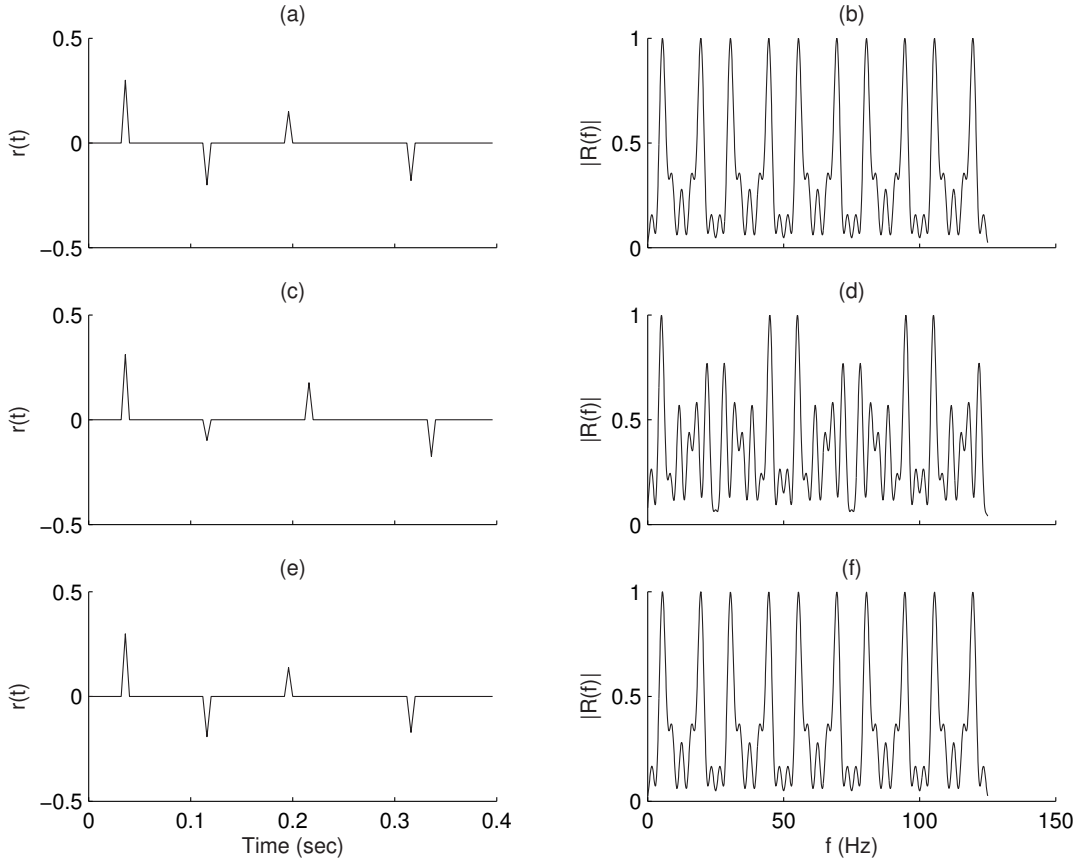

Figure 4.8: (a) A sparse reflectivity . (b) The power spectrum of the true reflectivity (c) The estimated reflectivity series using Cauchy norm regularization and (d) is its corresponding power spectrum, (e) the estimated reflectivity series using Cauchy norm regularization and an impedance constraint, (f) the power spectrum of the reflectivity series in (e).

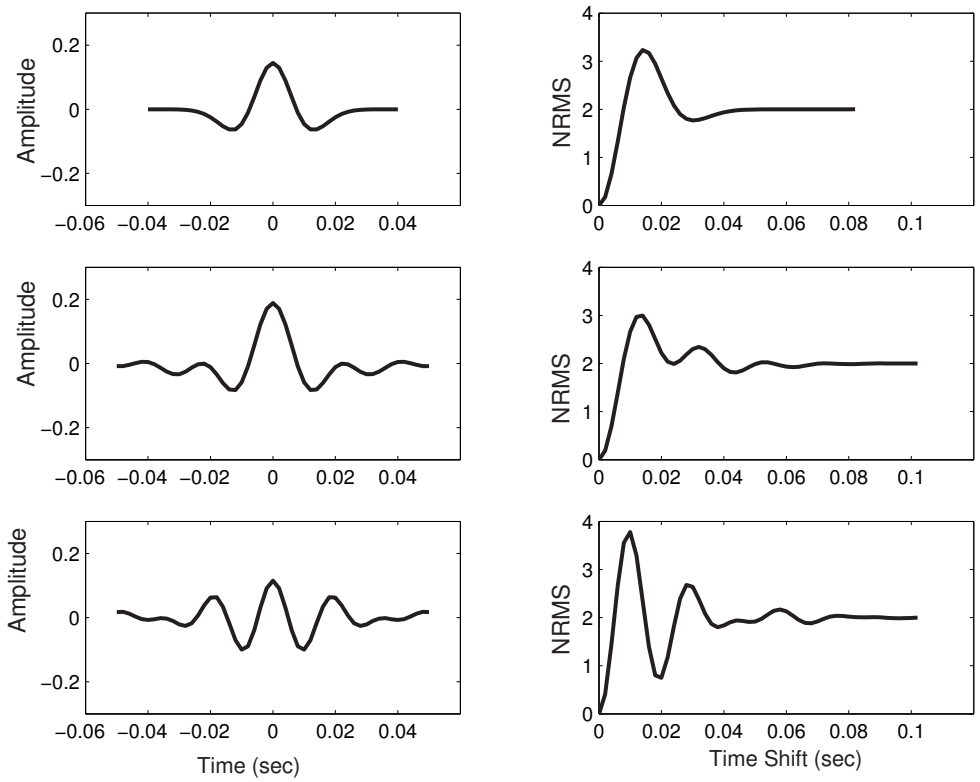

Figure 4.9: Left column are different source wavelets and right column are the corresponding difference curves as a function of time-shift. The NRMS is defined in Appendix A.

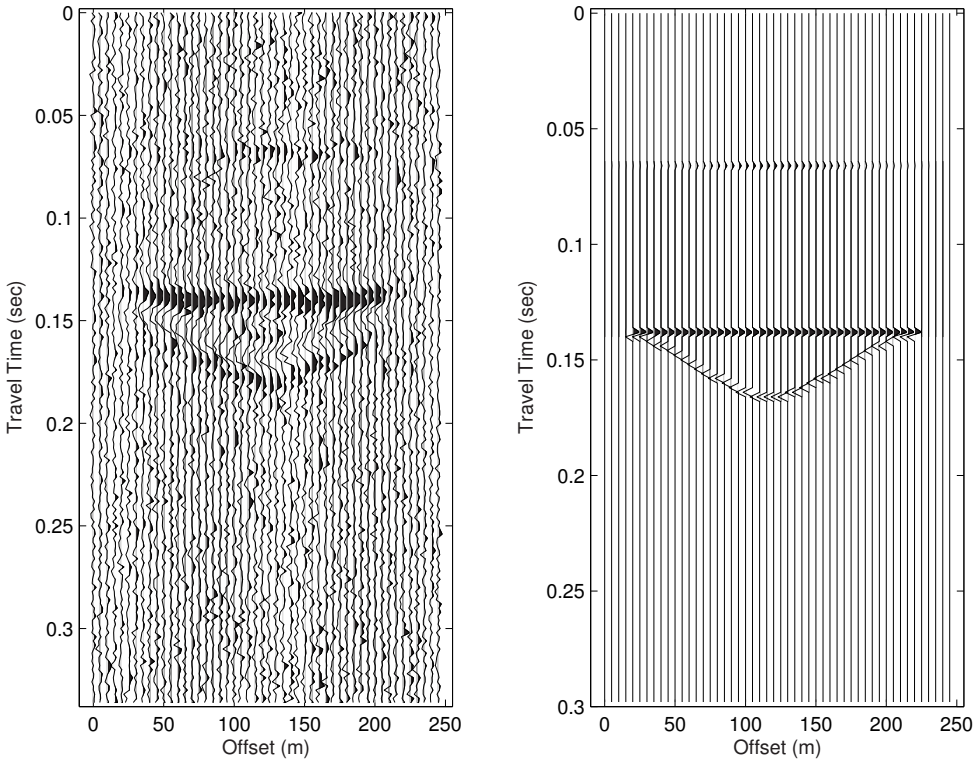

Figure 4.10: Left: time-lapse data difference. Right: the real reflectivity difference due to the reservoir change.

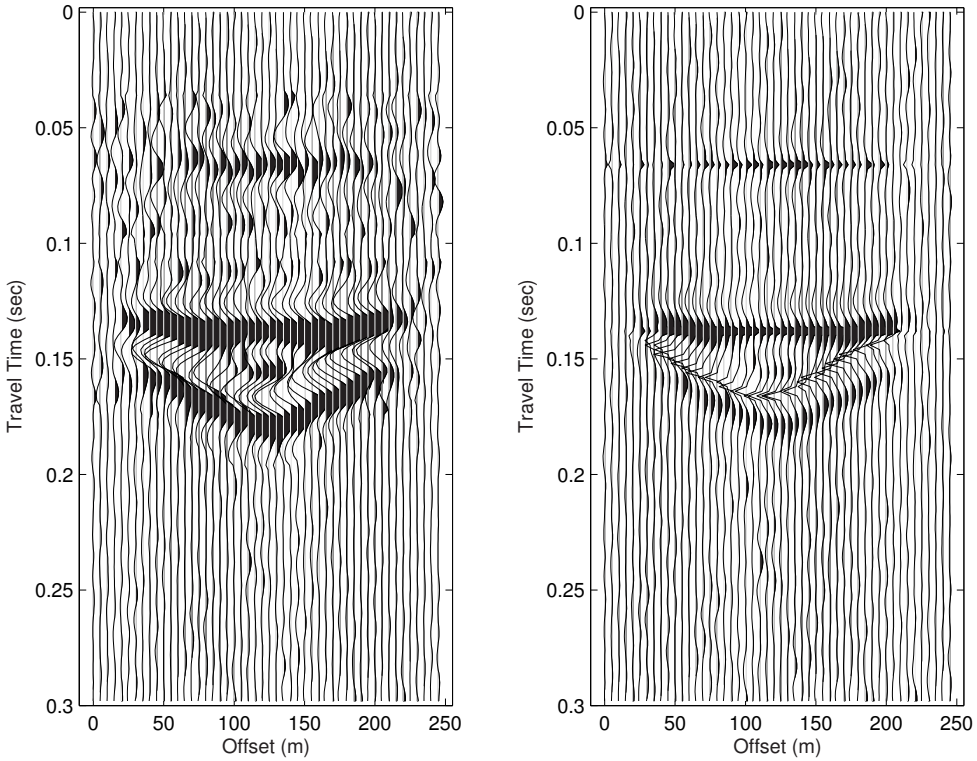

Figure 4.11: Left: inversion results using a wider structural constraint. Right: inversion results using an accurate structural constraint.

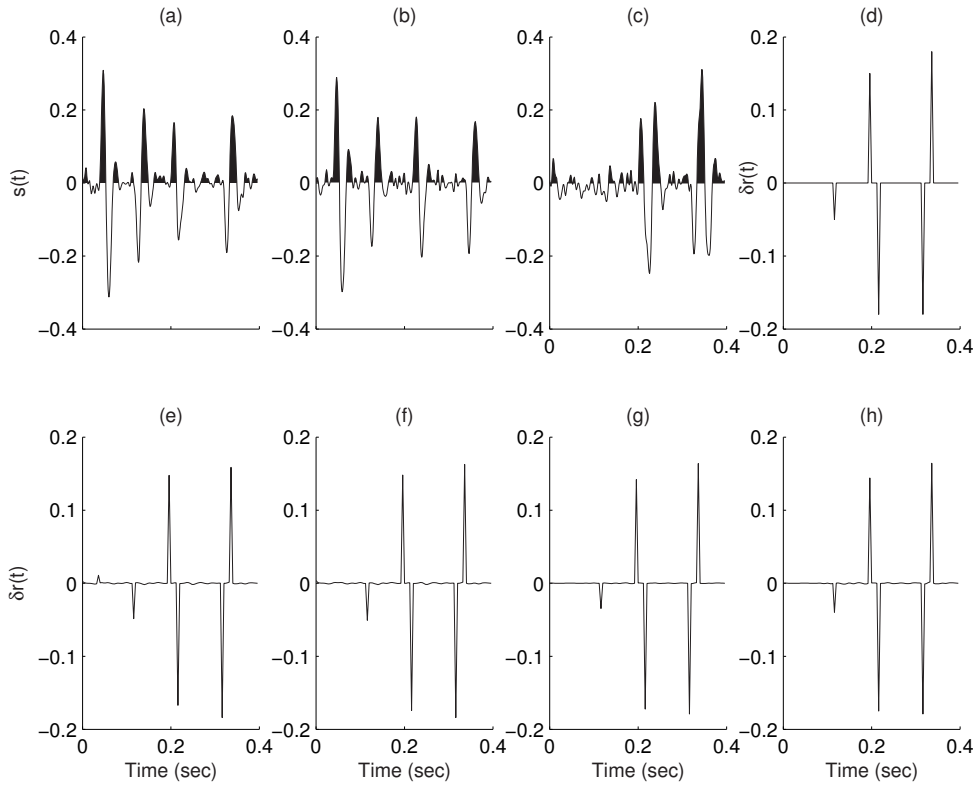

Figure 4.12: Comparison between the inversion results: (a) base seismogram, (b) monitoring seismogram, (c) data difference between the monitor and base survey, (d) the real model difference, from (e) to (f) are all estimated model difference using different schemes. (e) is invert survey by survey without impedance constraint, (f) is invert survey by survey with impedance constraint, (g) is structurally constrained simultaneous inversion without impedance constraint and (h) is structurally constrained simultaneous inversion with impedance constraint.

Chapter 5

Three Inversion Schemes

Time-lapse seismic data contains two or more seismic data sets. Thus there exists different schemes for inverting time-lapse seismic data which is different from conventional seismic inversion. From a mathematical point view, it is natural to divide the schemes into three category: first, invert the data sets separately and then subtract the estimated models to obtain the model difference; second, subtract the data and then invert the data difference for the model difference and third, simultaneously invert the data sets and then obtain the model difference by subtracting the estimated models.

In this chapter, we will first state the three schemes in a mathematical form and derive the cost function for them. Then, we will compare the three schemes and discuss their advantages and disadvantages by some simple examples. Finally, the resolution limit of the time-lapse inversion is discussed.

5.1 Mathematical Formulation

For simplicity, we assume the time-lapse data only have two data sets. Assume we have two data sets which satisfy the convolution model

$$\begin{aligned} \mathbf{s}_b &= \mathbf{W}_b\mathbf{r}_b + \mathbf{n}_b \\ \mathbf{s}_s &= \mathbf{W}_s\mathbf{r}_s + \mathbf{n}_s \end{aligned} \tag{5.1}$$

where $\mathbf{s}$ is the data, $\mathbf{W}$ is the convolution operator, $\mathbf{r}$ denotes the reflectivity, $\mathbf{n}$ means the noise and the subscript b, s mean the base survey and monitoring survey respectively.

The first time-lapse inversion scheme is to invert the two data sets separately and then estimate the model difference. In this case, we need to solve the following two cost functions:

$$\begin{aligned} \mathbf{J}_b &= ||\mathbf{s}_b - \mathbf{W}_b\mathbf{r}_b||_2^2 + \lambda_1\mathbf{R}(\mathbf{r}_b) + \lambda_2||\mathbf{C}_1\mathbf{r}_b - \beta_b||_2^2 \\ \mathbf{J}_s &= ||\mathbf{s}_s - \mathbf{W}_s\mathbf{r}_s||_2^2 + \lambda_3\mathbf{R}(\mathbf{r}_s) + \lambda_4||\mathbf{C}_2\mathbf{r}_s - \beta_s||_2^2 \end{aligned} \tag{5.2}$$

where λ denotes the trade-off parameter, $\mathbf{C}$ means the impedance constraint matrix, $\mathbf{beta}$ are the impedance constraints and $\mathbf{R}(\mathbf{m})$ denotes a sparseness constraint. If we take the structural constraint into the consideration, the second equation in the 5.2 needs an extra term $||\mathbf{W}_c(\mathbf{r}_s - \hat{\mathbf{r}}_b)||_2^2$, where $\mathbf{W}_c$ is the weighting matrix in the structural constraint.

Another time-lapse inversion scheme is to invert the data differences. This scheme require the forward operators in the equation 5.1 to be the same, that is $\mathbf{W}_b = \mathbf{W}_s = \mathbf{W}$. If this requirement is satisfied, this scheme turns out to solve the cost function:

$$\mathbf{J} = ||(\mathbf{s}_b - \mathbf{s}_s) - \mathbf{W}(\mathbf{r}_b - \mathbf{r}_s)||_2^2 + \lambda_1\mathbf{R}(\mathbf{r}_b, \mathbf{r}_s) + \lambda_2||\mathbf{C}(\mathbf{r}_b - \mathbf{r}_s) - (\beta_b - \beta_s)||_2^2 \tag{5.3}$$

where the sparseness constraint $\mathbf{R}$ can be applied to the two model $\mathbf{r}_b, \mathbf{r}_s$ separately or the model difference $\mathbf{r}_b - \mathbf{r}_s$. The form of this cost function is similar to the cost function for one data set shown in the equation 5.2. We can add a term $||\mathbf{W}_c(\mathbf{r}_b - \mathbf{r}_s)||_2^2$ into the above equation 5.3 to absorb the structural prior information.

The last time-lapse inversion scheme is to invert the two data sets simultaneously and then obtain the model difference. The cost function for this scheme is:

$$\begin{aligned} \mathbf{J} &= ||\mathbf{s}_b - \mathbf{W}_b\mathbf{r}_b||_2^2 + ||\mathbf{s}_s - \mathbf{W}_s\mathbf{r}_s||_2^2 + \lambda_1||\mathbf{r}_b||_2^2 + \lambda_2||\mathbf{W}_c(\mathbf{r}_b - \mathbf{r}_s)||_2^2 \\ &+ \lambda_3||\mathbf{r}_s||_2^2 + \lambda_4||\mathbf{C}_1\mathbf{r}_b - \beta_b||_2^2 + \lambda_5||\mathbf{C}_2\mathbf{r}_s - \beta_s||_2^2 \end{aligned} \tag{5.4}$$

If we assume the model parameters are sparse, then this cost function can be changed to

$$\begin{aligned} \mathbf{J} &= ||\mathbf{s}_b - \mathbf{W}_b\mathbf{r}_b||_2^2 + ||\mathbf{s}_s - \mathbf{W}_s\mathbf{r}_s||_2^2 + \lambda_1\mathbf{R}(\mathbf{r}_b) + \lambda_2||\mathbf{W}_c(\mathbf{r}_b - \mathbf{r}_s)||_2^2 \\ &+ \lambda_3\mathbf{R}(\mathbf{r}_s) + \lambda_4||\mathbf{C}_1\mathbf{r}_b - \beta_b||_2^2 + \lambda_5||\mathbf{C}_2\mathbf{r}_s - \beta_s||_2^2 \end{aligned} \tag{5.5}$$

In practice, we can invert the time-lapse seismic data by any of the three schemes. However, the three schemes will give different inversion results. The comparison among the three schemes is shown in the following section.

5.2 Comparison of Three Inversion Schemes

In this section, we applied the three scheme to synthetic $2D$ post-stack time-lapse data sets. The synthetic data is generated by the methods shown in the chapter 2. The model parameters are the same as the one used in the chapter 2 except that the thickness of the reservoir are given three different values in this chapter. The elastic properties of the reservoir is given by the table 2.2. In Figure 5.1, the thickness of the reservoir is $100m$. In Figure 5.2, the thickness of the reservoir is $25m$. The thickness of the reservoir is $10m$ in Figure 5.3. All the data sets contains three percent Gaussian noise. The corresponding reflectivity difference for the three examples are shown in Figure 5.4 (a), (b) and (c) respectively. In the following part, we will invert the three different time-lapse seismic data by different inversion schemes. The impedance constraint, sparseness constraint and structural constraint are applied to achieve a better result.

5.2.1 Comparing the three schemes

The inversion results in the case that the reservoir is $100m$ thick are illustrated in Figure 5.5. Figure 5.6 shows the results in the case that the reservior thickness is $25m$ and the results when the reservoir thickness is $10m$ is shown in Figure 5.7.

As we can see from these figures, the second scheme provides very bad results. The possible reasons are: the noise level of the data difference is not the same one of the separate data and thus the χ^2 criterion is not satisfied; the reservoir changes is not so big and thus the data differences generate a problem similar to the inversion of a thin layer. The comparison among Figures 5.5, 5.6 and 5.7 demonstrates that the results become worse as the thickness of the reservoir decreases for all the three inversion schemes. While, the comparison between the results of inverting data separately and simultaneous inversion shows that the simulatenous inversion scheme performs better when the thickness of the reservoir decreases.

5.3 Summary

In this chapter, I provided the formulation of the three inversion schemes for time-lapse seismic data. Then I generated three synthetic examples based on the rock physics and

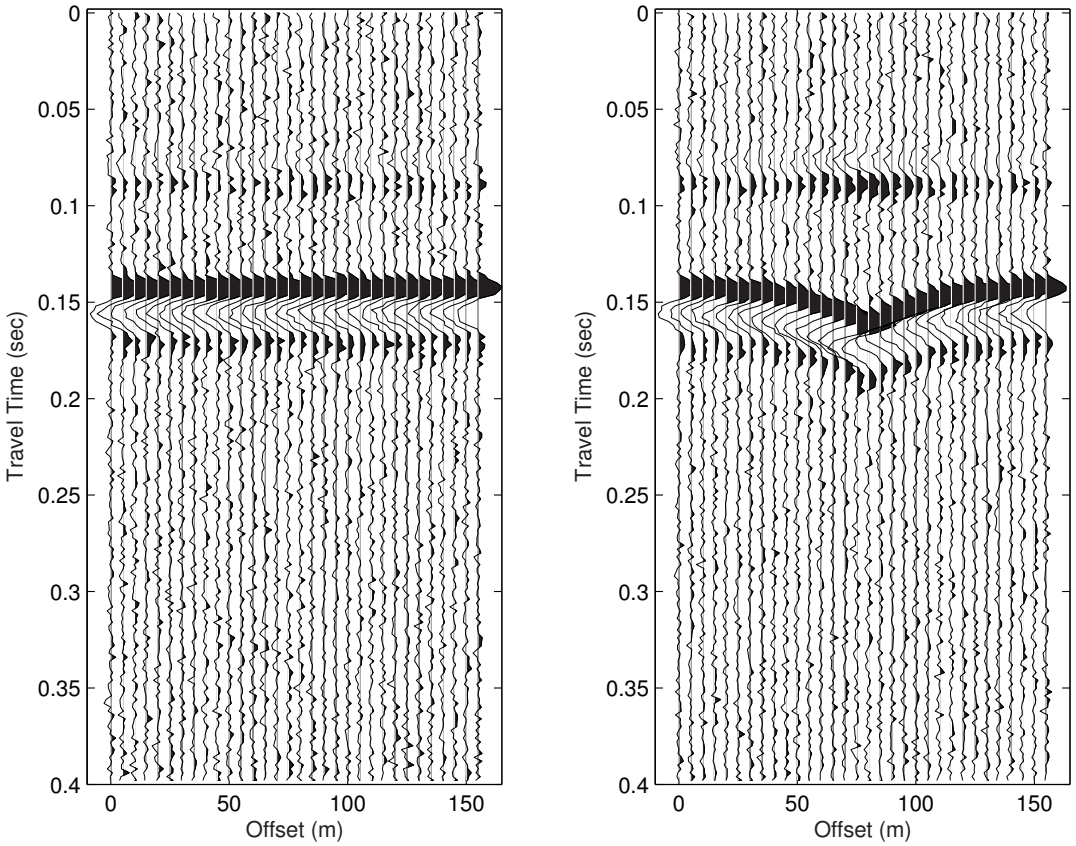

Figure 5.1: Left: base survey data; Right: monitor survey data. The thickness of the reservoir is $100m$.

experimental data provided in the chapter 2. The three schemes are tested on these toy examples. These examples indicate the following things: (a) as the thickness of the reservoir decrease, the results of all the three inversion schemes become worse, (b) since the reservoir change is not large, the scheme that inverts the data difference give the worst results, (c) when the reservoir is very thin, the simultaneous inversion scheme performs better than other schemes.

In the real world, the reservoir is usually thin and the noise factors are more complicated than the synthetic examples. Therefore, it's necessary to consider how to design a better scheme for the thin layer case and how to improve the performance of the constraints in these schemes.

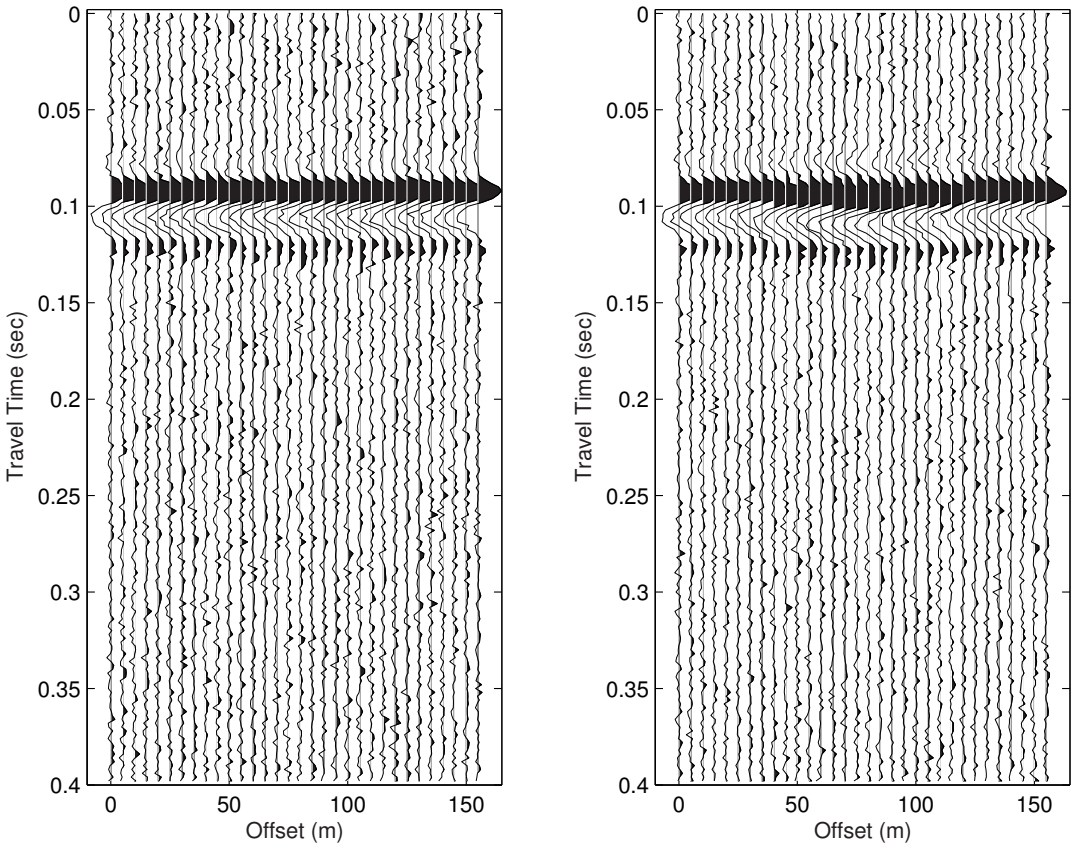

Figure 5.2: Left: base survey data; Right: monitor survey data. The thickness of the reservoir is $25m$.

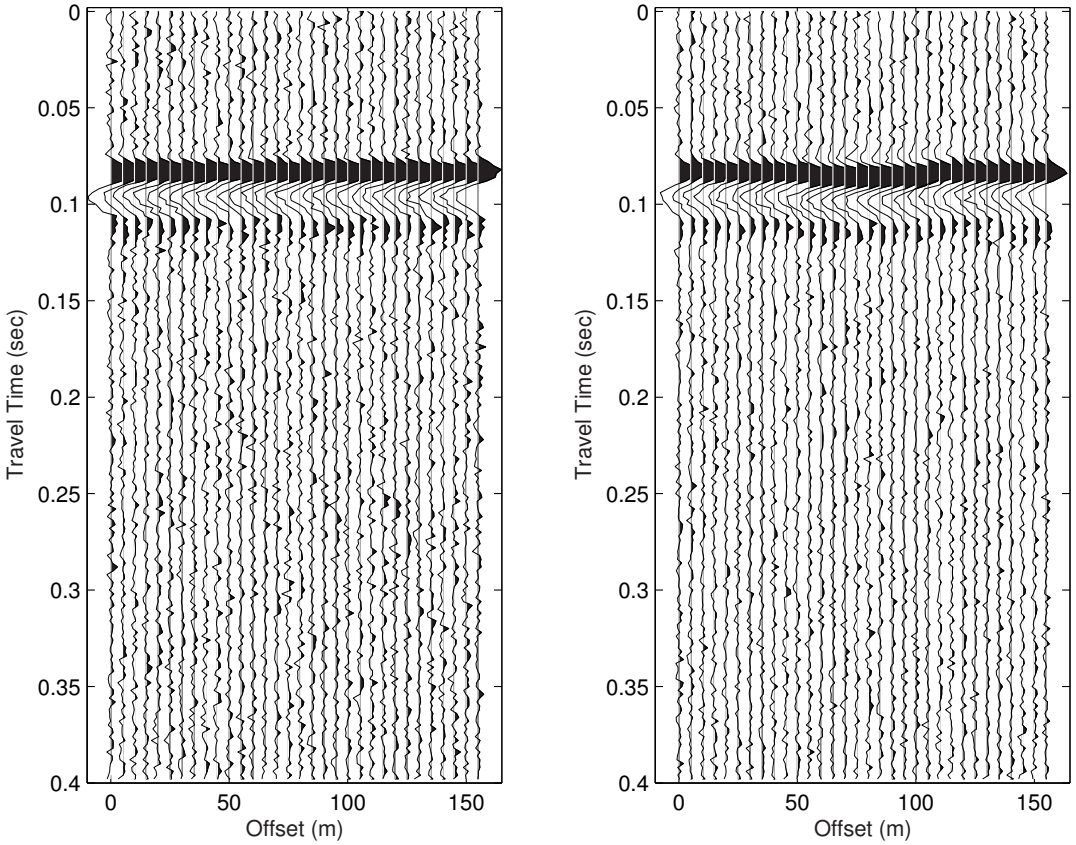

Figure 5.3: Left: base survey data; Right: monitor survey data. The thickness of the reservoir is $10m$.

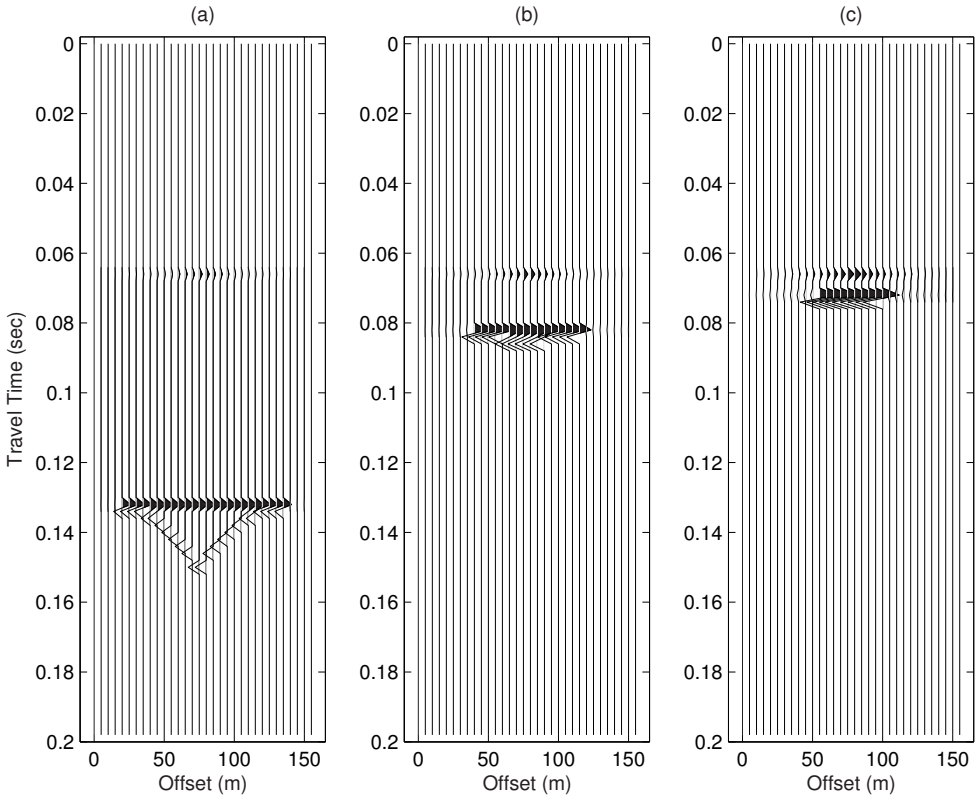

Figure 5.4: The reflectivity difference between the base survey and the monitor survey: (a) the case that the thickness of the reservoir is $100m$, (b) the case that the thickness of the reservoir is $25m$ and (c) the case that the thickness of the reservoir is $10m$.

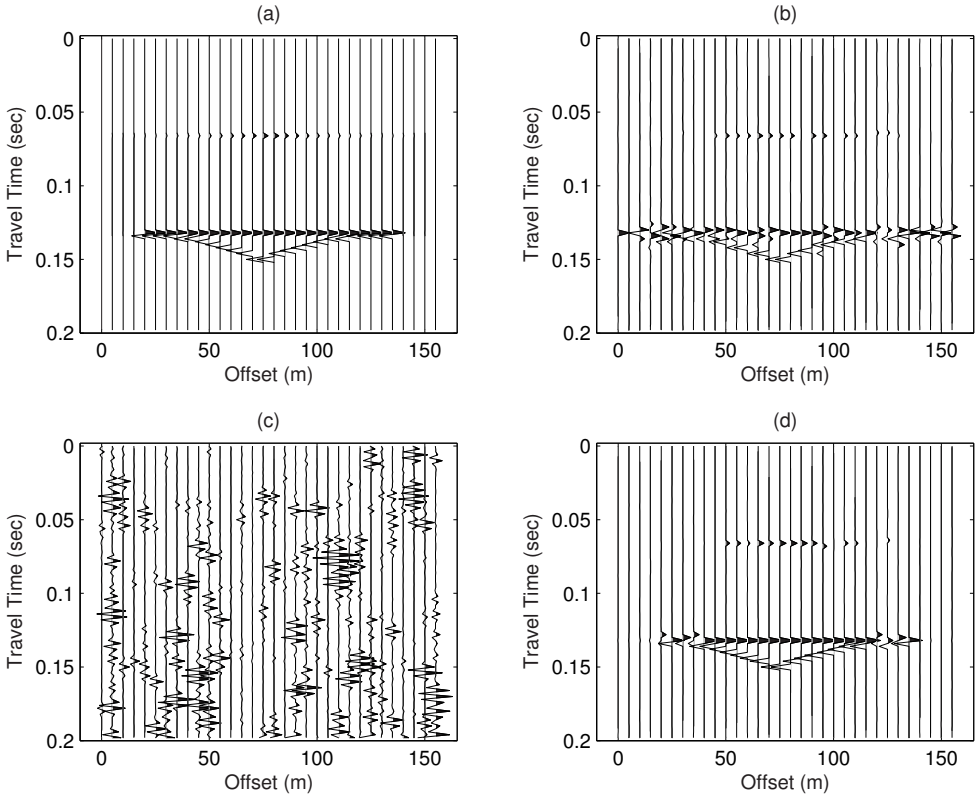

Figure 5.5: The thickness of the reservoir is $100m$: (a) The true reflectivity difference between the base survey and the monitor survey. (b) The estimated reflectivity difference between the base survey and the monitor survey by inverting the time-lapse data separately. (c)The estimated reflectivity difference between the base survey and the monitor survey by inverting the data difference. (d) The estimated reflectivity difference between the base survey and the monitor survey using the simultaneous inverson scheme.

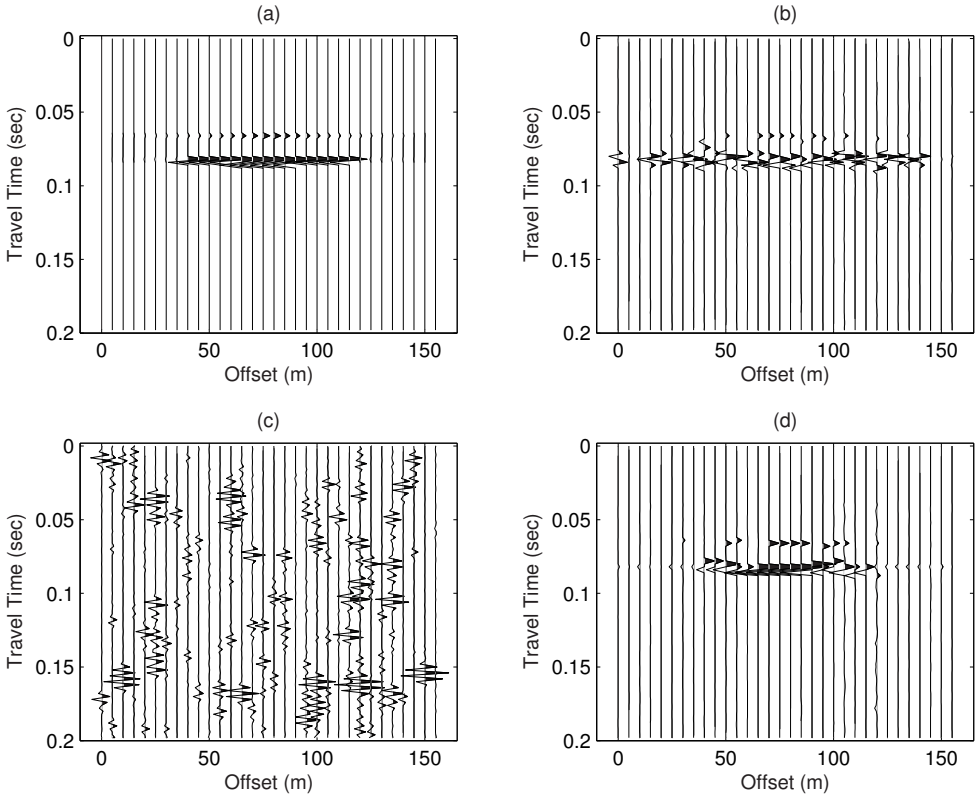

Figure 5.6: The thickness of the reservoir is $25m$: (a) The true reflectivity difference between the base survey and the monitor survey. (b) The estimated reflectivity difference between the base survey and the monitor survey by inverting the time-lapse data separately. (c)The estimated reflectivity difference between the base survey and the monitor survey by inverting the data difference. (d) The estimated reflectivity difference between the base survey and the monitor survey using the simultaneous inverson scheme.

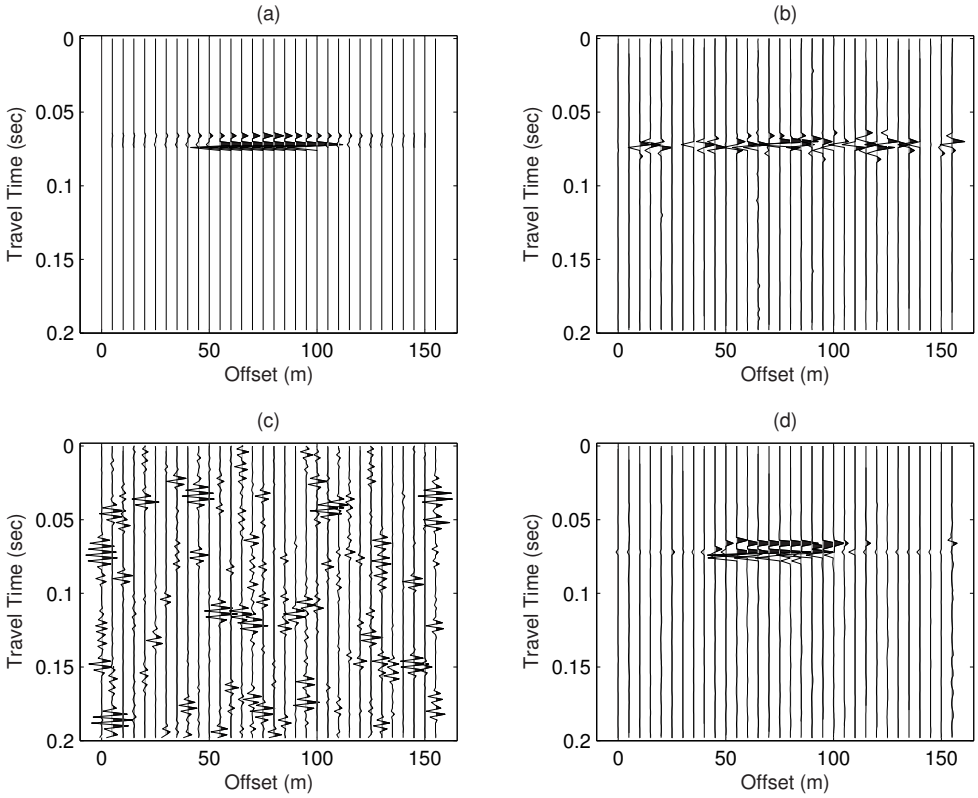

Figure 5.7: The thickness of the reservoir is $10m$: (a) The true reflectivity difference between the base survey and the monitor survey. (b) The estimated reflectivity difference between the base survey and the monitor survey by inverting the time-lapse data separately. (c)The estimated reflectivity difference between the base survey and the monitor survey by inverting the data difference. (d) The estimated reflectivity difference between the base survey and the monitor survey using the simultaneous inverson scheme.

Chapter 6

Conclusions and Discussions

This thesis explores the inverse problem for time-lapse seismic data. I have also reviewed the rock physics basis for time-lapse seismic research.

I have discussed the sparse, impedance and structural constraints and their applicability in time-lapse case. How to attenuate the model changes due to non-repeatable noise is an important issue in time-lapse inversion. This thesis investigates the possibility of three inversion schemes which also applied sparse, impedance and structural constraints for time-lapse inversion.

In chapter 2, I have reviewed the rock physics basis for the forward modeling of the time-lapse seismic data. I first introduced the concept of reservoir parameters and the calculation of pore fluids and rock properties. Then I discussed the estimation of the composite material's elastic properties. After that, I reviewed the fluid substitution theory, especially the Gassmann's equation. Finally I discussed the most common reservoir parameters which can affect time-lapse seismic data.

In order to understand the time-lapse inverse problems, I reviewed the concepts of ill-posed and well-posed problem and reviewed the regularization techniques. The definition of regularization, the concept of null space, the model and data regularization, the model-space and data-space regularization and then the Bayesian framework which is used to design the regularization terms are discussed in details. Then the most common numerical techniques for inversion which are forward and adjoint operators, IRLS and CGLS method and the Conjugate Gradient algorithm are mainly discussed. The examples illustrate that regularization terms implies the restriction by the *priori* information

and the difference between data-space regularization and the model-space regularization is mainly due to the introduction of a pre-conditioner.

The seismic data has its own limitations. It is band-limited, always noisy and the data points are finite while the earth model parameters are continuously distributed. Three constraints which are sparseness constraints, impedance constraints and structural constraints are derived. Bayesian theory plays an important roll for the design of the sparseness and structural constraints. The impedance constraint helps to maintain right amplitude scale. The sparseness constraint is to recover the lost low and high frequency components of the seismic data. The structural constraints are designed especially for time-lapse case and attenuate the model changes outside the reservoir area.

In chapter 5, three inversion schemes for time-lapse seismic data are mathematically formulated. The three schemes are tested on three synthetic examples. These examples indicate the following things: (a) as the thickness of the reservoir decrease, the results of all the three inversion schemes become worse, (b) since the reservoir change is not large, the scheme that inverts the data difference give the worst results, (c) when the reservoir is very thin, the simultaneous inversion scheme performs better than other schemes.

Potential future directions are: (a) the design of a better structural constraint, (b) the resolution limit of time-lapse inversion, (c) other methods to recover the lost frequency information, (d) improve the inversion schemes for coherent noise case, (e) methods to identify whether the estimated changes are the real reservoir changes.

References

Abubakar, A., van den Berg, P., and Fokkema, J., 2001, A feasibility study on nonlinear inversion of time-lapse seismic data:, *in* 71st Ann. Internat. Mtg Soc. of Expl. Geophys., 1664–1667.

Abubakar, A., van den Berg, P. M., and Fokkema, J., 2003, Towards non-linear inversion for characterization of time-lapse phenomena through numerical modeling.: Geophysical Prospecting, **51**, 285–293.

Aldridge, D. F., 1990, The Berlage wavelet (short note): Geophysics, **55**, 1508–1511.

Batzle, M., and Wang, Z., 1992, Seismic properties of pore fluids: Geophysics, **57**, 1396–1408.

Bell, D. W., and Shirley, D. J., 1980, Temperature variation of the acoustical properties of laboratory sediments: Journal of the Acoustical Society of America, **68**, 227–231.

Biondi, B., Mavko, G., Mukerji, T., Rickett, J., Lumley, D., Deutsch, C., Gundeso, R., and Thiele, M., 1998, Reservoir monitoring: A multidisciplinary feasibility study: The Leading Edge, **17**, 1404–1414.

Biot, M. A., 1956a, Theory of propagation of elastic waves in a fluid-saturated porous solid (i. higher frequency range): J. Acous. Soc. Am., **28**, 179–191.

——— 1956b, Theory of propagation of elastic waves in a fluid-saturated porous solid (i. low frequency range): J. Acous. Soc. Am., **28**, 168–178.

Box, G. E. P., and Tiao, G. C., 1973, Bayesian inference: Addison-Wesley.

Brown, R., and Korringa, J., 1975, On the dependence of the elastic properties of a porous rock on the compressibility of the pore fluid: Geophysics, **40**, 608–616.

Butler, R. M., 1994, Steam-assisted gravity drainage: concept, development, performance and future: The Journal of Canadian Petroleum Techology, **33**, 44–50.

Castagna, J. P., 1993, Avo analysis - tutorial and review, *in* Backus, M. M., Ed., Offset-dependent reflectivity - theory and practice of AVO analysis: Soc. of Expl. Geophys., 3–36.

Chambers, R., Yarus, J., Alexeev, V., and Sudhakar, V., 2002, Quantitative use of seismic attributes for reservoir characterization: Recorder, **27**, 14–25.

Claerbout, J. F., 1992, Earth sounding analysis: Processing versus inversion: Blackwell Science Inc.

Claerbout, J. F., 1997a, Diagonal weighting: An elementary challenge to mathematicians: SEP-Report, **94**, 133–136.

——— 1997b, Preconditioning and scaling: SEP-Report, **94**, 137–140.

Claerbout, J. F., 2004, Image estimation by example: Geophysical soundings image construction: multidimensional autoregression: Stanford Exploration Project. URL http://sepwww.stanford.edu/sep/prof/.

Cooke, D., Ball, V., Muryanto, T., O'Donnell, G., and Sena, A., 1999, What is the best seismic attribute for quantitative seismic reservoir characterization?:, *in* 69th Ann. Internat. Mtg Soc. of Expl. Geophys., 1588–1591.

Domenico, S. N., 1977, Elastic properties of unconsolidated porous sand reservoirs: Geophysics, **42**, 1339–1368.

Dvorkin, J., and Nur, A., 1993, Dynamic poroelasticity: A unified model with the squirt and the biot mechanism: Geophysics, **58**, 524–533.

Dvorkin, J., Nolen-Hoeksema, R., and Nur, A., 1994, Squirt-flow mechanism: Macroscopic description: Geophysics, **59**, 428–438.

Fomel, S., 1997, On model-space and data-space regularization: A tutorial: SEP-Report, **94**, 141–164.

Franklin, J. N., 1970, Well-posed stochastic extensions of ill-posed linear problems: Journal of Mathematical Analysis and Applications, **31**, 682–716.

Galikeev, T., 2004, Application of new seismic attributes to reservoir mointoring: Ph.D thesis, Stanford University.

Gassmann, F., 1951, Elasticity of porous media: Uber die elastizitat poroser medien: Vierteljahrsschrift der Naturforschenden Gesselschaft in Zurich, **96**, 1–23.

Geertsma, J., 1961, Velocity-log interpretation: The effect of rock bulk compressibility: Soc. Pet. Eng. J., **26**, 169–181.

Ghosh, S. K., 2000, Limitations on impedance inversion of band-limited reflection data: Geophysics, **65**, 951–957.

Golub, G. H., and Van Loan, C. F., 1996, Matrix computations: Johns Hopkins University Press, third edition.

Gray, S. H., and Symes, W. W., 1985, Stability consideration for one-dimensional inverse problems: Geophysical Journal of the Royal Astronomical Society, **80**, 149–163.

Hadamard, J., 1923, Lectures on the cauchy problem in linear partial differential equations: Yale University Press.

Hanke, M., and Hansen, P. C., 1993, Regularization methods for large-scale problems: Survey on Mathematics for Industry, **49**, 253–315.

Hansen, P. C., 1998, Rank-deficient & discrete ill-posed problems: Numerical aspects of linear inversion: SIAM.

Hashin, Z., and Shtrikman, S., 1963, A variational approach to the elastic behavior of multiphase materials: J. Mech. Phys. Solids, **11**, 127–140.

He, T., and Schmitt, D. R., 2004, Measurement of the elastic frame properties on weakly consolidated sandstone in support of fluid substitution studies:, *in* Annual Convention Can. Soc. Expl. Geophys., 4 pages.

Herawati, I., 2002, The use of time-lapse P-wave impedance inversion to monitor a CO_2 flood at Weyburn field, Saskatchewan: MSc. thesis, Colorado School of Mines.

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

Hill, R. W., 1952, The elastic behavior of a crystalling aggregate: Proc. Phys. Soc. London, **65 A**, 349–354.

Jackson, D. D., 1972, Interpretation of inaccurate, insufficient and inconsistent data: Geophysical Journal of the Royal Astronomical Society, **28**, 97–109.

Jackson, D. D., 1979, The use of a priori data to resolve non-uniqueness in linear inversion: Geophys. J. R. astr. Soc., **57**, 137–157.

Jackson, D. D., 1985, A bayesian approach to nonlinear inversion: Journal of Geophysical Research, **90**, 581–591.

Ji, J., 1994, Toward an exact adjoint: Semicircle versus hyperbola: SEP-Report, **80**, 499–512.

Johansen, T., Jakobsen, M., and Ruud, B., 2002, Estimation of the internal structure and anisotropy of shales from borehole data: Journal of Seismic Exploration, **11**, 363–381.

Kalkomey, C. T., 1997, Potential risks when using seismic attributes as predictors of reservoir properties: The Leading Edge, **16**, 247–251.

Kragh, E., and Christie, P., 2002, Seismic repeatability, normalized rms, and predictability: The Leading Edge, **21**, 640–647.

Lanczos, C., 1961, Linear differential operators: D. Van Nostrand Company Ltd.

Lay, T., and Wallace, T. C., 1995, Modern global seismology: Academic Press.

Lewis, C., 1997, Seismic attributes for reservoir monitoring: A feasibility study using forward modeling: The Leading Edge, **16**, 459–469.

Lines, L., 2002, Seismic monitoring of hot and cold heavy oil production:, *in* 72nd Ann. Internat. Mtg Soc. of Expl. Geophys., 1666–1668.

Lumley, D., and Beydoun, W. B., 1994, Angle-dependent reflectivity estimation by kirchhoff migration/inversion: Theory: SEP-Report, **79**, 217–240.

Lupton, R., 1993, Statistics in theory and practice: Princenton University Press.

Malaver, C. H., 2004, Statistical applications to quantitative seismic reservoir characterization and monitoring at west pearl queen field, Lea County, New Mexico: Ph.D thesis, Stanford University.

Mavko, G., and Jizba, D., 1991, Estimating grain-scale fluid effects on velocity dispersion in rocks: Geophysics, **56**, 1940–1949.

Mavko, G., Mukerji, T., and Dvorkin, J., 1998, The rock physics handbook: Cambridge University Press.

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

Nichols, D., 1994, A simple example of a null space and how to modify it: SEP-Report, **82**, 141–164.

Nur, A., Marion, D., and Yin, H., 1991, Wave velocities in sediments, *in* Hovem, J. M., Richardson, M. D., and Stoll, R. D., Eds., Shear Waves in Marine Sediments: Kluwer Academic Publisher, Dordrecht, The Netherland, 131–140.

Nur, A. M., Mavko, G., Dvorkin, J., and Gal, D., 1995, Critical porosity: The key to relating physical properties to porosity in rocks: 65th Ann. Internat. Mtg, pages 878–881.

O'Brien, M. S., Sinclair, A. N., and Kramer, S., 1994, Recovery of a sparse spike time series by l_1 norm deconvolution: IEEE Trans. Signal Proccessing, **42**, 3353–3365.

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

Packwood, J. L., 1997, Rock physics for hydrocarbon recovery: Ph.D Thesis, Stanford University.

Pepper, R., van Bemmel, P., Rusdinadar, S., and Waite, M., 1997, 4-d seismic analysis to measure fluid induced reflection time changes:, *in* 67th Ann. Internat. Mtg Soc. of Expl. Geophys., 906–909.

Peterson, R. A., Fillipone, W. R., and Coker, F. B., 1955, The synthesis of seismograms from well log data: Geophysics, **20**, 516–538.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T., 1986, Numerical recipes: The art of scientific computing: Cambridge University Press, New York, NY.

Raymer, L. L., Hunt, E. R., and Gardner, J. S., 1980, An improved sonic transit time-to-porosity transform: Trans. Soc. Prof. Well Log Analysts, 21st Annual Logging Symposium, page Paper P.

Reuss, A., 1929, Berechnung der fliegrenze von mischkristallen: Angew. Mathem. u. Mech., **9**, 49–58.

Rickett, J., Biondi, B., and Lumley, D., 1996, Modeling heterogeneous reservoirs using the first order Born approximation: SEP-Report, **92**, 75–82.

Robinson, E. A., 1967, Predictive decomposition of time series with application to seismic exploration: Geophysics, **32**, 418–484.

Sacchi, M. D., Constantinescu, C. M., and Feng, J., 2003, Enhancing resolution via non-quadratic regularization-next generation of imaging algorithms:, *in* CSPG/CSEG Joint Convention Can. Soc. Expl. Geophys., 5 pages.

Sacchi, M. D., 1997, Reweighting strategies in seismic deconvolution: Geophysical Journal International, **129**, 651–656.

Santos, L. T., Schleicher, J., Tygel, M., and Hubral, P., 2000, Seismic modeling by demigration: Geophysics, **65**, 1281–1289.

Sarkar, S., Gouveia, W., and Johnston, D., 2003, On the inversion of time-lapse seismic data:, *in* 73rd Ann. Internat. Mtg. Soc. of Expl. Geophys., 1489–1492.

Scales, J. A., and Smith, M., 1994, Introductory geophysical inverse theory: Samizdat Press.

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

Sivia, D. S., 1996, Data analysis: A bayesian tutorial: Clarendon Press - Oxford.

Tarantola, A., 1987, Inverse problem theory-methods for data fitting and model parameter estimation: Elsevier.

Taylor, H. L., Banks, S. C., and McCoy, J. F., 1979, Deconvolution with the l_1 norm: Geophysics, **44**, 39–52.

Telford, W., Geldart, L., and Sheriff, R., 1990, Applied geophysics: Cambridge University Press, Cambridge.

Tennebo, P. O., Veire, H. H., Sonneland, L., Signer, C., and Reymond, B., 1998, Inversion of 4-d seismic with best feasible approximation:, *in* 68th Ann. Internat. Mtg Soc. of Expl. Geophys., 52–53.

Theune, U., 2004, Seismic monitoring of heavy oil reservoirs : rock physics and finite element modelling: Ph.D thesis, the University of Alberta.

Trad, D., 2003, Implementations and applications of the high resolution radon transform: Ph.D thesis, University of British Columbia.

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

Voigt, W., 1910, Lehrbuch der kritallphysik: Teubner-Verlag, Leipzig.

Wang, Z., and Nur, A., 1988, Effect of temperature on wave velocities in sands and sandstones with heavy hydrocarbons: Seismic and Acoustic Velocities in Reservoir Rocks, vol. 1 of Geophysical reprinted series, pages 188–194.

Wang, Z., 2001, Fundamentals of seismic rock physics: Geophysics, **66**, 398–412.

Watson, I. A., 2004, Integrated geological and geophysical analysis of a heavy-oil reservoir at pikes peak, saskatchewan: MSc. thesis, University of Calgary.

Winthaegen, P., and Verschuur, D., 2001, Cfp-approach to time-lapse angle-dependent reflectivity analysis:, *in* 71st Ann. Internat. Mtg Soc. of Expl. Geophys., 1580–1583.

Wyllie, M. R. J., Gregory, A. R., and Gardner, L. W., 1956, Elastic wave velocities in heterogeneous and porous media: Geophysics, **21**, 41–70.

Youzwishen, C. F., 2001, Non-linear sparse and blocky constrants for seismic inverse problems: MSc. thesis, the University of Alberta.

Zhang, Y., and Schmitt, D. R., 2004, Time-lapse impedance inversion using hybrid data transformation and the spike deconvolution method: 74th Ann. Internat. Mtg., pages 2303–2306.

Zhou, H., Zhang, Y., Gray, S., and Zhang, G., 2002, Regularization algorithms for seismic inverse problem:, *in* 72nd Ann. Internat. Mtg Soc. of Expl. Geophys., 942–945.

Appendix A

RMS Value for time-lapse case

Suppose we have a time window (start from t_1 to t_2) on two traces, and $a(t)$ and $b(t)$ are the data in this window for the two traces respectively. For the general case, $a(t)$ and $b(t)$ can be written as:

$$a(t) = r_a(1)\, w_a(t) + r_a(2)\, w_a(t-\tau_a) + n_a(t) \tag{A.1}$$

$$b(t) = r_b(1)\, w_b(t) + r_b(2)\, w_b(t-\tau_b) + n_b(t), \tag{A.2}$$

where, number 1 and 2 in the bracket means the upper and lower bound respectively, a and b means trace a and b respectively, w means the source wavelet, r means the reflectivity, n means the noise and τ is the time delay caused by the thickness of the reservoir.

The RMS value of the $(a(t)-b(t))$ is defined as:

$$RMS(a(t)-b(t)) = \sqrt{\frac{\sum_{t=t_1}^{t_2}(a(t)-b(t))^2}{N}}, \tag{A.3}$$

substitute $a(t)$, $b(t)$ into the above equation, we can get

$$\begin{aligned} RMS^2(a(t)-b(t)) \times N = & \sum_{k=1}^{N}(r_a(1) * w_a(k) + r_a(2) * w_a(k-\tau_a) + n_a(k) \\ & -r_b(1) * w_b(k) - r_b(2) * w_b(k-\tau_b) - n_b(k))^2. \end{aligned} \tag{A.4}$$

For simplicity, we will assume $w_a(t) = w_b(t) = w(t)$ in the following part of the appendix.

First, let us consider a simple case: only energy from the upper bound are contained in the time-window, which means the reservoir is thick. So the terms which contain τ can

be neglected. This equation is turned out to be:

$$RMS^2(a(t)-b(t)) \times N = \sum_{k=1}^{N} (\triangle r(1)w(k) + \triangle n(k))^2$$

$$RMS(a(t)-b(t)) = \sqrt{\frac{\sum_{k=1}^{N} (\triangle r(1)w(k) + \triangle n(k))^2}{N}}. \quad (A.5)$$

Here, $\triangle r(1) = r_a(1) - r_b(1)$ and $\triangle n(k) = n_a(k) - n_b(k)$. If $\triangle r(1) = 0$,which means no reservoir change, we will know

$$RMS(a(t)-b(t)) = \sqrt{\frac{\sum_{k=1}^{N} \triangle^2 n(k)}{N}}. \quad (A.6)$$

If $|\triangle r(1) * w(k)| \gg |\triangle n(k)|$, which means we can observe the reservoir change, it is reasonable to use the following approximate equation

$$RMS(a(t)-b(t)) \approx |\triangle r(1)| \sqrt{\frac{\sum_{k=1}^{N} w^2(k)}{N}}. \quad (A.7)$$

In this case, the normalized RMS (NRMS) value can be written as:

$$\begin{aligned} NRMS &= \frac{200 \times RMS(a(t)-b(t))}{RMS(a(t)) + RMS(b(t))} \\ &= 100 \times |\frac{\triangle r(1)}{\overline{r(1)}}|. \end{aligned} \quad (A.8)$$

Here,

$$\overline{r(1)} = \frac{r_a(1) + r_b(1)}{2},$$

and NRMS is expressed in percentage. Equation(10) means that, in this scenario, the normalized RMS will give us the relative change in reflectivity.

Now, suppose the reservoir is thin. In this case, energy from the lower bound can not be neglected, and we have to consider the terms which contain τ. We can obtain the following equation:

$$RMS^2(a(t)-b(t)) \times N = \sum_{k=1}^{N} ((\triangle r(1)w(k) + \triangle n(k)) + r_a(2)w(k-\tau_a) - r_b(2)w(k-\tau_b))^2. \quad (A.9)$$

If we can observe the reservoir change, the above equation can be rewritten as:

$$RMS^2(a(t)-b(t)) \times N \approx \sum_{k=1}^{N} (\triangle r(1)w(k) + r_a(2)w(k-\tau_a) - r_b(2)w(k-\tau_b))^2. \quad (A.10)$$

Since the reservoir is very thin, we can assume $\tau_a = \tau_b = \tau$ for simplicity, then we can get

$$RMS^2(a(t) - b(t)) \times N \approx \triangle^2 r(1) \sum_{k=1}^{N} w^2(k) + \triangle^2 r(2) \sum_{k=1}^{N} (w^2(k - \tau)) + 2\triangle r(1)\triangle r(2) \sum_{k=1}^{N} (w(k)w(k - \tau)). \quad \text{(A.11)}$$

This equation show that the RMS value is a function of the autocorrelation of the wavelet and the time-lagged wavelet plus a term which is the cross-correlation of the two wavelets. The latter term shows the interference or tuning effects of a thin reservoir.

To make the RMS more realistic, we may let $\tau_a \neq \tau_b$ and $w_a(t) \neq w_b(t)$. The equations shown above will become more complicated. But we will find the numerical value of RMS will change but the qualitative characteristics of the RMS value will not change.

Appendix B

Resolution Limitation of Time-lapse Inversion

When we consider the problem of resolution of time-lapse inversion , we need to answer following questions:

- How big the time-lapse amplitude changes needed to be observed?
- In what circumstances, our inversion scheme will work?

Basically, there are two ways to measure the time-lapse amplitude differences. One the absolute the amplitude difference. Another is the relative amplitude changes. To make the results stable, we usually compare the difference in one time window instead of just one point.

Kragh and Christie (2002) proposed a normalized root mean squares (NRMS) difference as a measure of amplitude difference between two traces, a_t and b_t, within a given time window $t1 - t2$. The formula to calculate the difference is expressed as a percentage:

$$NRMS = \frac{200 \times RMS(a(t) - b(t))}{RMS(a(t)) + RMS(b(t))}, \tag{B.1}$$

where the RMS operator is defined as

$$RMS(x_i) = \sqrt{\frac{\sum_{t_1}^{t_2}(x_i)^2}{N}}, \tag{B.2}$$

and N is the number of samples in the interval $t_1 - t_2$.

In time domain, the lower bound of the reservoir may change its position while the upper bound is always fixed. For simplicity, we will only compare the time-lapse amplitude change of the upper bound.If we assume the two data sets have the same source wavelet $w(t)$ and let the length of the source wavelet be the window length,which is the ideal case, we will get the following results:

$$\begin{aligned} a(t) &= r_a \times w(t) \\ b(t) &= r_b \times w(t) \\ RMS(a(t)) &= r_a \times \sqrt{\frac{\sum_{t_1}^{t_2} w^2(t)}{N}} \\ RMS(b(t)) &= r_b \times \sqrt{\frac{\sum_{t_1}^{t_2} w^2(t)}{N}} \\ RMS(a(t)-b(t)) &= (r_a - r_b) \times \sqrt{\frac{w^2(t)}{N}} \\ NRMS &= \frac{200 \times RMS(a(t)-b(t))}{RMS(a(t)) + RMS(b(t))} \\ &= 100 \times |\frac{r_a - r_b}{\frac{r_a + r_b}{2}}|, \end{aligned} \tag{B.3}$$

where, $a(t)$ and $b(t)$ are the traces obtained at t_1, t_2 respectively, and r_a,r_b are the reflection coefficients corresponding to the $a(t)$ and $b(t)$ respectively.

Relative difference can help us understand how big the time-lapse change,but it may mislead us when the noise level is low. In other words, we may get high percentage difference when there is no reservoir change. So, we also need absolute amplitude difference. For convenience, I use the value of $RMS(a(t) - b(t))$.

We have tested the two measures on a simple three layer model: the velocity of first layer is $3000m/s$, the second layer is $2800m/s$ and the third layer is $5000m/s$. We change the velocity of the second layer to $2720m/s$ and $2640m/s$ respectively. The source wavelet is Ricker wavelet. The time interval of sample is $2ms$. Then we make the dominant frequency of the source wavelet to be $20, 30, 40Hertz$ respectively.

First, we assume two traces which have no reservoir change, signal to noise ratio is 10. Then we calculate the NRMS and RMS difference using the source wavelet with different central frequency. The results is shown in figure B.1. From this figure, we find, if there is no reservoir change, the relative amplitude difference will be affected by the

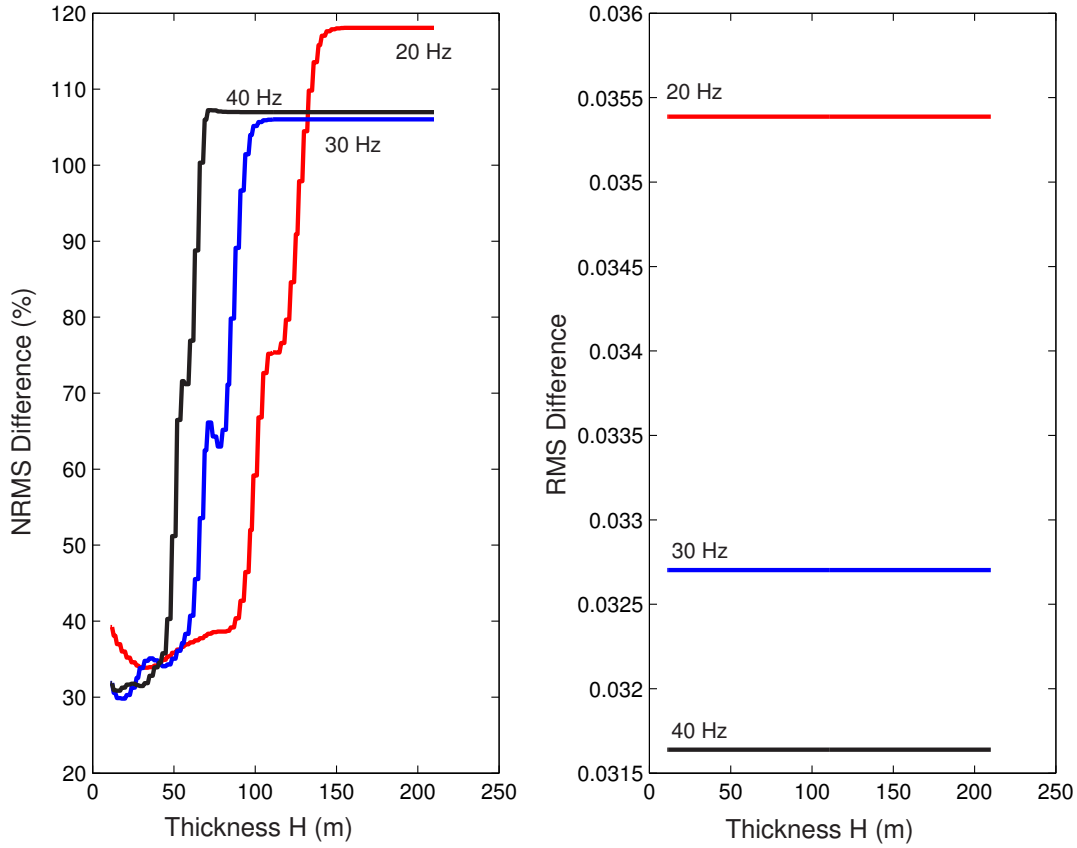

Figure B.1: Relative amplitude difference (NRMS) and absolution amplitude difference (RMS) (no reservoir change).

thickness of the reservoir while the relative amplitude difference will not. The reason for this difference is because of the interference of the reflections from upper bound and lower bound of the reservoir. Since the length of our time window is the same as the length of the source wavelet, this interference is only affected by the thickness of the reservoir and the frequency content of the source wavelet. In this case, the formula B.3 we have derived is not valid because some part of the reflections from the lower bound is also in the time window we use. So, the formula need to be modified as:

$$a(t) = r_a \times w(t) + n_a \qquad b(t) = r_b \times w(t) + n_b,$$

where n_a and n_b are the energy from the lower bound. Since there is no reservoir change, n_a will equal to n_b. So it is easy to know the value of $RMS(a(t) - b(t))$ will not change but the value of $RMS(a(t))$ and $RMS(b(t))$ will change and thus the value $NRMS$ will be different (see Appendix for more details). For different frequency, the energy from lower bound may be different and thus it will only change the value of NRMS. So, if we use wavelets with different central frequency, our conclusion will be the same except the numerical value will be different. Since the most common central frequency of the seismic source wavelet is around 30Hz, we will only use the source wavelet whose central frequency is 30Hz in the following example. We should notice that the tuning effects show in the right panel of the figure B.1 is not guaranteed to increase the change, the exact value should be predicted using equations in the appendix A. Theoretically, it shows that the tuning effect may decrease the amplitude change cause by production. But I cannot find a good example shown in figures by now because this decrease is pretty small. More work is needed to find when this tuning effect will increase or decrease the amplitude change caused by production.

Now, we decrease the velocity and calculate the NRMS and RMS value. We can see these results in figure B.2. From figure B.2, we can know the following things:a. for thick reservoir, the amplitude change is stable and proportional to the reservoir change (see appendix A for more details); b. if the reservoir change is small, it is difficult to identify the change in both NRMS and RMS curve; c. when the reservoir is thin, the RMS and NRMS value is oscillating, but it is easier to identify the change by comparing the RMS value than NRMS value; d. for thin reservoir, if the reservoir change is big enough, the amplitude change can be observed in most cases, except very thin reservoir and the thickness of the reservoir is in the transition zone (from thin to thick zone).

By now, we can get a preliminary answer to our first question in the introduction section: to observe a time-lapse change, the reservoir change must be bigger than a certain value and the thickness of the reservoir cannot be some value.

The resolution limit in time-lapse case is different from other cases because it is the limit to find the difference between the two data sets. From the discussion above, we can get two interesting results. First, a thin layer is easier to detect in time-lapse data. The only requirement is the reservoir change is big enough. Second, the resolution limit

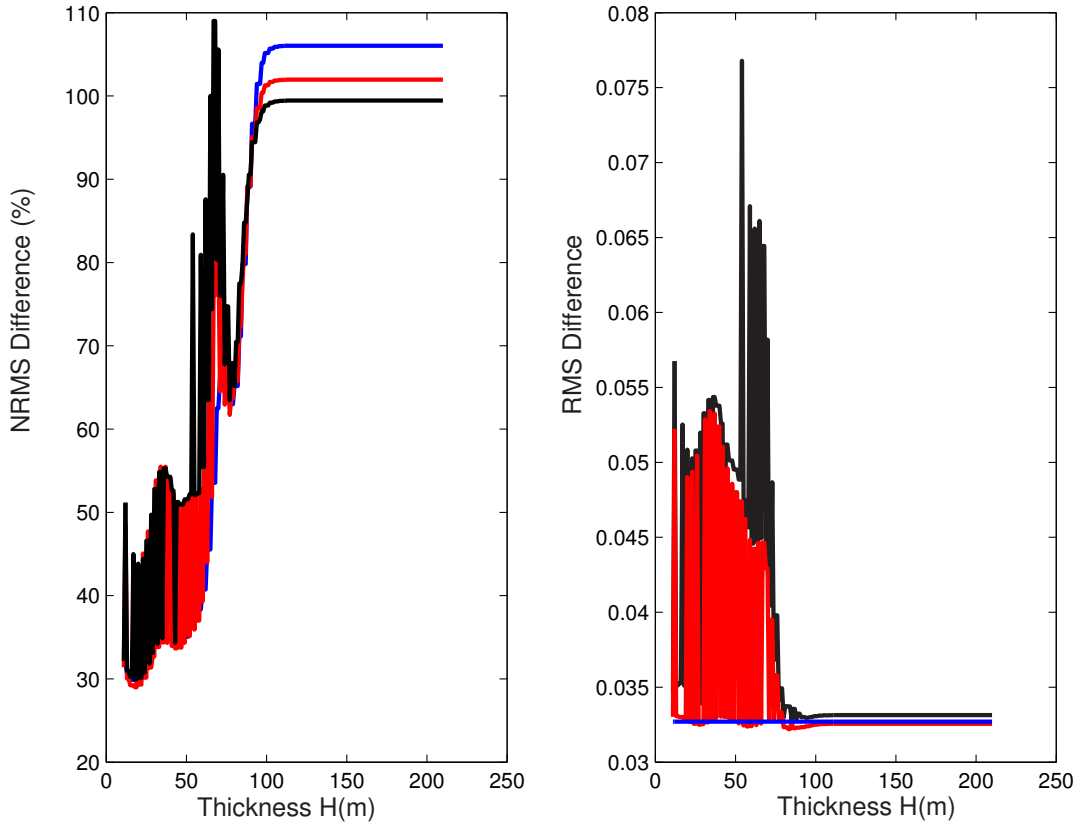

Figure B.2: Comparison of NRMS and RMS of different reservoir change: the red line means the velocity change is $80m/s$, the dark line means the velocity change is $160m/s$, the blue line is for the case without reservoir change and signal to noise ration (SNR) is 10.

to time-lapse inversion may be smaller than the usual limit. This may be because of the non-uniqueness introduced by tuning effects of thin layers and the effects caused by small time-shift. We should notice if the source wavelet is different, for thin layers, things will be more unpredictable. More analytical work needs to be done for the time-lapse inversion of thin layers.