University of Alberta

Regularization of the AVO inverse problem by means of a multivariate
Cauchy probability distribution

by

Wubshet M. Alemie

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

Examining Committee

Dr. Mauricio D. Sacchi (Supervisor), Physics

Dr. V. Kravchinsky (Chair and Examiner), Physics

Dr. J. Gu (Examiner), Physics

Dr. O. Leuangthong (Examiner), Faculty of Engineering

Abstract

Amplitude Variation with Offset (AVO) inversion is one of the techniques that is being used to estimate subsurface physical parameters such as P-wave velocity, S-wave velocity, and density or their attributes. AVO inversion is an ill-conditioned problem which has to be regularized in order to obtain a stable and unique solution. In this thesis, a Bayesian procedure that uses a Multivariate Cauchy distribution as a prior probability distribution is introduced. The prior includes a scale matrix that imposes correlation among the AVO attributes and induces a regularization that provokes solutions that are sparse and stable in the presence of noise. The performance of this regularization is demonstrated by both synthetic and real data examples using linearized approximations to the Zoeppritz equations.

Acknowledgements

Contents

List of Tables

List of Figures

List of Symbols

Symbols Description	
s	Second
ms	Millisecond
m	Meter
kg	Kilogram
α	P-wave velocity
β	S-wave velocity
ρ	Density
σ	Standard deviation
A	P-wave reflectivity
B	S-wave reflectivity
C	Density reflectivity
$\mathbf{r}_\alpha$	P-wave impedance
$\mathbf{r}_\beta$	S-wave impedance
λ	First Lamé parameter
μ	Bulk modulus
μ_h	Hyper-parameter
γ	S-wave to P-wave velocity ratio
p	Ray parameter
h or X	Offset
V_{rms}	Root mean square velocity
ν	Degree of freedom
J	Objective function
mg	Index for Multivariate Gaussian
uc	Index for Univariate Cauchy
bc	Index for Bivariate Cauchy
tc	Index for Trivariate Cauchy
Q	Regularization term
Ψ	Scale matrix
$\mathbf{C}_f$	Covariance matrix for Fatti's parameters
$\mathbf{C}_a$	Covariance matrix for Aki and Richard's parameters
S/N	Signal to noise ratio
CDP	Common depth point
$RMSE$	Root mean square error

CHAPTER 1

Introduction

The reflection seismic method is one of the geophysical tools that is used to explore subsurface physical structures and properties. It is extensively used in exploration for hydrocarbons (i.e.., petroleum, natural gas) and to lesser extent for hydrogeology (Haeni, 1986) and mineral exploration (Eaton et al., 2003). The basic procedure in seismic exploration uses seismic sources such as explosives, vibrators or air guns to generate seismic waves and record the amplitudes to build an image of the subsurface. The recorded seismic signal shows the variation in amplitude and phase as a function of the source-receiver distance (offset) and two-way travel time. Amplitude Variation with Offset (AVO) is one of the techniques which study the change in the amplitude of a seismic signal with respect to offset. If the offset information is mapped into angle of incidence, the method is called Amplitude Variation with Angle (AVA). The amplitude and the total travel time depends on the angle of incidence, the properties and the structure of the underlying rocks. Usually, due to the limitations in acquisition and noise, the seismic data has to be processed to extract the signal. After routine processing steps, the seismic data are used to analyze and infer the nature of the subsurface. The first information that can be extracted from a given seismic data are about the geological structures of the subsurface. A direct detection of hydrocarbon from the amplitude of a seismic section was first suggested by Ostrander (1982). In other words, detailed investigation of the amplitude of reflections on a seismic section can provide important information about the presence of hydrocarbons. The quality of the interpretation depends on the processing steps carried out to accurately represent the amplitudes of the recorded waveforms (Chopra and Castagna, 2007). For instance, the presence of gas or oil in a pore space of rocks lowers the velocity as compared to water. Although the change in density is small, gas and oil also lower density which in turn has a great effect in lowering the acoustic impedance. Depending on the local circumstances, the presence of the hydrocarbon

may show different anomalies on seismic section such as bright-spot, flat-spot and dim-spot (McQuillin et al., 1984). 'Bright-spot' appears where rocks-filled with gas and results in an increased contrast i.e. it is associated with strong amplitude contrast across lithology and their fluid content. On the contrary, 'dim-spots' are associated with weak amplitude and 'flat-spot' occur at fairly sharp gas-liquid contact. These anomalies on the seismic section can be used as direct indicators of hydrocarbon. However, such anomalies can exist due to other structural changes; for example local buildups of carbonate (old oyster reefs) (Taner et al., 1979). This indicates that detailed information about the physical properties of the rocks that causes the amplitude anomaly should be estimated. Those physical properties help to predict fluid content (oil, gas, or water) and avoid the ambiguity in the interpretation of the amplitude anomalies. In other words, they give additional information about potential reservoirs or help to identify new hydrocarbon reservoirs. However, extracting the physical model parameters from seismic data is not a simple task. It involves extensive mathematical modeling and algorithms for seismic data inversion. In the next two sections, I will discuss about forward and inverse problems.

1.1 Forward Problems

The method to derive information about a given system from known physical parameters is what is called forward (direct) problem. The concern in this kind of problem is describing a behavior of a given system in terms of the measured parameters i.e. constructing the physics of the problem. It is an easy problem if the physical model is known. A simplified diagram which shows a relationship between the model and the data in forward modeling is shown in Figure 1.1.

1.2 Inverse Problems

Inverse problems are problems which arises when data measurements are obtained and one wants to estimate unobservable property. In other words, it is a method of inferring model parameters that characterize a system under study (see Figure 1.1). Inverse problems are common in applied sciences such as geophysical exploration, medical tomography, remote sensing, and astronomy. The basic step in formulating inverse problems is to write down a system of equations (linear or non-linear) that fully describe the relationship between the observed data, the physics model, and the unknown model parameters. Depending on the number of equations and the unknown model parameters, the system of equations can be classified as underdetermined, even-determined or overdetermined. For certain defined number of equations and unknowns, the problem is said to be overdetermined if the system of
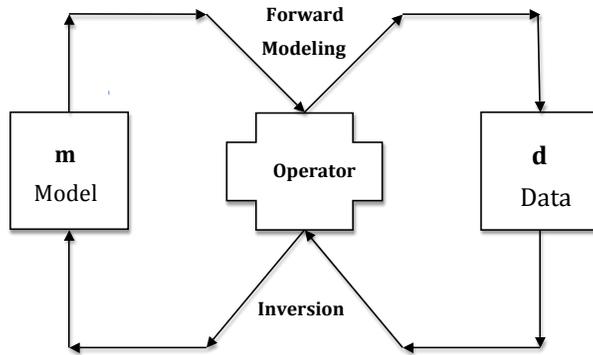
Figure 1.1: A simple diagram which depicts the relationtionship between the model and data both for forward and inverse problem.

equations outnumbers the number of unknown parameters. In this case, the system of equations contain too much information which may lead to contradictions. In even-determined problems, the number of equations is equal to the number of unknowns. This implies that there is enough information to determine the model parameters. But this does not mean that the system of equation can estimate the parameters accurately. Underdetermined problems arise when the number of equations is fewer than the number of unknowns i.e. there is less information to retrieve the parameters of interest. In general, an inverse problem could be well-posed or an ill-posed problem. A problem is called well-posed if

- the solution exists,

- the solution is unique,

- the solution continuously depends on the data.

If it is not well-posed, it is an ill-posed problem i.e. the solution may not exist or the solution is not unique or a small perturbation in the data produce large change in the solution (Tikhonov and Arsenin, 1987). In many geophysical problems, most often, the solutions are non-unique and so the problems turn out to be ill-posed. Such kind of problems needs prior information (regularization). The main topic of this thesis is to investigate the regularization of the AVO inverse problem via Bayesian approach.

1.3 AVO inversion

As defined above, the amplitude variation with offset (AVO) describes the change in amplitude of a seismic signal incident on an interface between two adjacent layers of different contrast as a function offset. AVO inversion is one of the inversion techniques that is becoming most popular in oil and gas industries. It uses pre-stack seismic data that has been processed in amplitude preserving fashion. The inversion helps to extract information about subsurface physical parameters such as P-wave velocity, S-wave velocity, and density or their attributes. The combination of these parameters is a powerful tool to distinguish hydrocarbon reservoirs from other subsurface materials. The basis for forward modeling in AVO inversion are the Zoeppritz equations. These equations provide a non-linear link between elastic parameters and amplitudes of waves transmitted and reflected at a given interface (Aki and Richards, 1980). The non-linearity of this equation makes it complicated for inversion. The first approximation to the Zoeppritz equation was provided by Bortfeld (1962). This equation shows clear relationship between the reflection coefficient and the earth model parameters but non-linear with respect to those parameters. Following the first approximation, other researchers have made linearized approximations including Aki and Richards's, Shuey's, and Smith and Gidlow's, Fatii's approximations. Using these approximations as a physical model, one can construct an inverse problem in order to retrieve the model parameters associated to the physical model. Combining the information from the processed data with the results of the inversion, makes it possible to produce accurate information about the subsurface. AVO inversion is an ill-posed problem which makes the solution of the inversion unreliable and therefore it needs regularization. Bayesian method is a statistical approach that can be used for many geophysical problems including AVO to formulate inverse problems with prior information (Tarantola, 1987; Ulrych et al., 2001). It combines the observed data with the knowledge we have about the model parameters before measuring the data. This could increase stability of the inversion and leads to reliable result. Before I go into the details of prior information, I would like to explain some of the empirical relationships among P-wave velocity, S-wave velocity, and density.

1.4 Empirical relationship between AVO parameters

It has been shown that P-wave velocity and S-wave velocity, P-wave velocity and density, and S-wave velocity and density are correlated. These also imply strong correlation among their corresponding adjacent layer contrast parameters; P-wave reflectivity, S-wave reflectivity, and density reflectivity.

1.4.1 Relationship between P-wave velocity and density

A well-known empirical relationship between P-wave velocity and density was first derived by Gardner (1974). This equation is given by

$$\rho = a\alpha^b,$$ (1.1)

where ρ is density, α is the P-wave velocity, a and b are empirical constants which are directly related to the type of rock. For clastic sedimentary rocks, Gardner found that $a = 0.23$ and $b = 0.25$. From this equation, it is also possible to derive the relationship between P-wave reflectivity ($\frac{\Delta\alpha}{\alpha}$) and density reflectivity ($\frac{\Delta\rho}{\rho}$) as

$$\frac{\Delta\rho}{\rho} = b\frac{\Delta\alpha}{\alpha}$$ (1.2)

which is a linear relationship.

1.4.2 Relationship between P-wave velocity and S-wave velocity

An empirical relationship between P-wave velocity and S-wave velocity was derived by Castagna et al. (1985). This is a linear relationship given by

$$\alpha = c\beta + d,$$ (1.3)

where the velocity is in km/s. The values of c and d is different for different rocks. For $c = 1.16$ and $d = 1.36$, it is called a mudrock line. For P-wave and S-wave velocities, it is possible to find the values of those two constants using linear fitting to bore-hole measurements. It is also possible to derive a relationship between P-wave reflectivity and S-wave reflectivity ($\frac{\Delta\beta}{\beta}$) from the above linear relationship. One can easily show that

$$\frac{\Delta\alpha}{\alpha} = \gamma c\frac{\Delta\beta}{\beta}.$$ (1.4)

where γ is the S-wave velocity to P-wave velocity ratio.

1.4.3 Relationship between S-wave velocity and density

A relationship between S-wave velocity and density was derived empirically by Potter et al. (1998) . It was found that S-wave velocity and density also follow a similar equation to the Gardener's equation (P-wave velocity and density) which suggests the S-wave velocity can

be used to predict density. This equation is given by

$$\rho = 0.37\beta^{0.22}.$$ (1.5)

Using this relationship, a relationship between density reflectivity and S-wave reflectivity takes the form

$$\frac{\Delta\rho}{\rho} = 0.22\frac{\Delta\beta}{\beta}.$$ (1.6)

The above three empirical relationships may not hold all the time but suggest a strong correlation between density and velocity. In AVO inversion, the correlation information can be incorporated via a prior probability distribution in Bayesian frame-work (Buland and Omre, 2003; Downton and Lines, 2004). It can be represented by a matrix and takes different form for different probability distribution used as prior. Well-log data or the above empirical relationships can be used as source of this information (Downton and Lines, 2004). The prior distributions and the correlation information matrices will be discussed in detail in Chapters 3, 4, and 5.

1.5 Scope and Goals

AVO inversion can be summarized as the process of retrieving model parameters from pre-stack seismic data. As mentioned above, the AVO inverse problem is ill-conditioned i.e. a small change in noise results large uncertainty in the estimates. This makes it difficult to produce reliable estimated parameters. For instance, the density attribute which provides information for reservoir characterization is difficult to retrieve by AVO inversion techniques (Downton and Lines, 2004; Li, 2005; Mahmoudian and Margrave, 2007). One way to increase reliability and resolution of the parameters is to appropriately regularize the AVO inverse problem. The Bayesian approach needs the appropriate choice of a probability distribution function to formulate the inverse problem. As explained above, the correlation informations can be incorporated in the inversion, most commonly, using a Gaussian probability distribution. On the other hand, the Cauchy probability distribution has long tails as compared to the Gaussian probability distribution. This implies that it gives relatively higher probability for large parameters. Therefore, using Cauchy distribution as a prior results in a regularization method which can play a role in of sparsity in the solution and thereby increase resolution of the estimated parameters (Sacchi and Ulrych, 1995; Downton and Lines, 2004). Downton and Lines (2004) used Gaussian probability distribution for the

three-term model parameters: the P-wave reflectivity, S-wave reflectivity and density reflectivity which are derived from the Aki and Richards's approximation. These parameters are correlated and the correlation relation is represented by a covariance matrix. In the same work, Univariate Cauchy probability distribution was used by diagonalizing the covariance matrix and applying the change of variables to treat the variables independent.

The focus of this thesis is to treat the prior distribution in a more formal way by constructing it from a single probability distribution called Multivariate Cauchy probability distribution. It is a special case of Multivariate **t** distribution (Appendix A.1) with one degree of freedom (Johnson and Kotz, 1972; Chauanhai, 1994). In this prior, the parameters are correlated at a time sample but statistically independent from one time sample to another. This prior distribution has long tail nature like the Univariate Cauchy and at the same time it allows to incorporate the correlation information via scale matrix. This prior imposes correlation among the AVO attributes and induces a regularization that provokes solutions that are sparse and stable in the presence of noise. The calculation of this correlation rely on the maximum likelihood estimation, EM algorithm (Moon, 1996; Chauanhai, 1994). In particular, the Bivariate Cauchy and the Trivariate Cauchy probability distribution functions are proposed for a two-term and three-term AVO inversion respectively. These are demonstrated by both synthetic and real data examples in comparison with the Multivariate Gaussian and Univariate Cauchy prior distributions.

1.6   Organization of the thesis

This thesis is organized as follows:

In Chapter 2, I will discuss modeling aspects of the AVO problem. Starting from the concept of wave propagation in a two-layered model, the Zoeppritz equation for the P-wave reflection coefficient and S-wave reflection coefficient are derived. In order to implement the Zoeppritz equations for AVO inversion, linearized approximation are needed. Those approximations are compared and their effect in modeling is investigated. Since ray tracing is one of the crucial steps in mapping offset into angle, two ray tracing methods are given in the last section of this chapter. The first ray tracing algorithm is an approximation that assumes straight line paths from source to the reflection point and then to the receiver using the root mean square velocity (rms velocity). The second method is an exact ray tracing which follows the change in the assumed background velocity.

In Chapter 3, the Bayesian approach and inversion algorithms for typical inverse problem (for AVO inversion) are discussed. The nature of two prior probability distributions, the Multivariate Gaussian and Multivariate **t** distribution, and their advantage one on another

are explained. The inversion algorithms for different possible scenarios are discussed. These include ordinary least squares solution (OLS), weighted least squares solution (WLS), iterative re-weighted least squares solution (IRLS), and the target oriented or window by window AVO inversion via IRLS. In the same section, two possible ways of selecting a regularization parameter (or hyper-parameter), the chi-square test and trade-off-diagram, are also discussed.

In Chapter 4, a two-term AVO inversion using the Fatti's approximation on Zoeppritz equation as forward modeling is investigated. Three prior distributions are used independently to formulate the inversion via Bayesian approach. The first prior distribution is the Gaussian probability distribution which leads to a quadratic regularization term. The other two are family of the Multivariate **t** distribution, Univariate Cauchy and Bivariate Cauchy distributions. The later two prior distribution lead to non-linear regularizations i.e. regularization term which depend on the model parameters. Finally, using the methods of solving the inversion given in chapter 3, the three regularizations are demonstrated by synthetic examples.

In Chapter 5, a three-term AVO inversion using the Aki and Richards's approximation as forward modeling is investigated. Three prior distributions are independently used to formulate the inversion via Bayesian approach. These include Multivariate Gaussian, Univariate Cauchy, and Trivariate Cauchy probability distributions. Similar to two-term AVO, regularization via Multivariate Gaussian is the inverse of parameter covariance matrix and the other two result non-linear regularizations. In the last section of this chapter, the regularizations are demonstrated by synthetic and real data examples.

In the last chapter, the summary of all chapters are briefly described. It focuses on the aim and the findings of this research.

CHAPTER 2

AVO Modeling

2.1 Introduction

In seismology, the structure and property of the Earth's interior is studied by analyzing the propagation of elastic waves (Aki and Richards, 1980). From a given source, parts of the generated waves travel through the earth's interior and part of the wave travels under the surface. Those elastic waves which propagate in the interior of the Earth are called body waves (P-waves and S-waves) and those which travel close to the surface are called surface waves (Love and Rayleigh waves). Seismic exploration mainly focus on body waves which contain much information about deeper structure of the earth.

P-waves

These type of waves are generated due to alternate compression and relaxation of the media. The motion of the waves are in the direction of the motion of the particles of the underground for this reason they are longitudinal waves. Most seismic experiments in seismic exploration use compressional wave sources in order to generate P-waves.

S-waves

Contrary to P-waves, S-waves (shear waves) are transverse waves i.e the particle motion of the underground is perpendicular to the direction of the motion of the waves. The most interesting features of these waves is they travel only through solids because liquids and gases do not support shear stress (Stein and Wysession, 2002).

The transformation of the seismic waves upon incident on two homogeneous isotropic elastic half spaces in welded contact at a plane interface is given by Zoeppritz equation (Aki and Richards, 1980). This equation is the base for AVO inversion as forward model. The next section is devoted to show the theoretical derivation for transmission and reflection
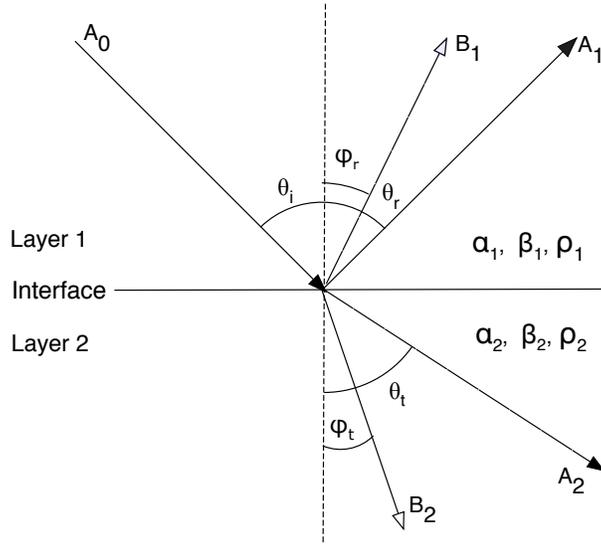
Figure 2.1: A two layer model in welded contact shows the transformation of an incident P-wave with initial amplitude A_0 into P and S-waves with their corresponding amplitudes .

coefficients for P and S-waves.

2.2 Zoeppritz Equation

Zoeppritz equation describe the amplitudes of the reflected and transmitted P-wave and S-waves which are incident at boundary between the two media. It is derived under the assumption that incident angle at the boundary is below a critical angle. The critical angle of incidence is determined by the velocity of the upper and lower layers. Let us denote the compressional and shear wave potentials by ϕ and ψ respectively.

The potentials ϕ and ψ satisfy a two dimensional wave equation in the cartesian coordinate system (x, z) i.e

$$\frac{\partial^2 \phi}{\partial x^2} + \frac{\partial^2 \phi}{\partial z^2} = \frac{1}{\alpha^2} \frac{\partial^2 \phi}{\partial t^2} \tag{2.1}$$

and

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} = \frac{1}{\beta^2} \frac{\partial^2 \psi}{\partial t^2}. \tag{2.2}$$

Let the incident and reflected P-wave potential field, the converted wave upon reflection

(S-wave reflected), the transmitted P-wave, and converted wave upon transmission (S-wave transmitted) be represented by ϕ_1, ψ_1, ϕ_2, and ψ_2 respectively. Indices 1 and 2 indicate the upper and lower layers respectively. The plane wave solutions to equations (2.1) and (2.2) according to the transformation shown in Figure 2.1, we have

$$
\begin{aligned}
\phi_1 &= A_0 e^{i\omega(px+\frac{\cos\theta_i}{\alpha_1}z-t)} + A_1 e^{i\omega(px-\frac{\cos\theta_r}{\alpha_1}z-t)}, \\
\psi_1 &= B_1 e^{i\omega(px-\frac{\cos\varphi_r}{\beta_1}z-t)}, \\
\phi_2 &= A_2 e^{i\omega(px+\frac{\cos\theta_t}{\alpha_2}z-t)}, \\
\psi_2 &= B_2 e^{i\omega(px+\frac{\cos\varphi_t}{\beta_2}z-t)},
\end{aligned}
\tag{2.3}
$$

where A_0, A_1, B_1, A_2 and B_2 are the wave amplitudes for the incident P-wave, reflected P-wave, reflected S-wave (converted upon reflection), transmitted P-wave and transmitted S-wave (converted P-wave upon transmission) respectively. The parameters p and ω are ray parameter and frequency respectively. The angles θ_i, θ_r, θ_t are the incidence, reflection and transmission angles of the P-wave respectively. The remaining angles φ_r and φ_t are the reflection and transmission angles for S -waves respectively (refer Figure 2.1).

To proceed further, we need to write Snell's law in terms of the angles and velocities of the media. The apparent velocity along the interface between the two layers is constant

$$
c_x = \frac{1}{p} = \frac{\alpha_1}{\sin\theta_i} = \frac{\beta_1}{\sin\varphi_r} = \frac{\alpha_2}{\sin\theta_t} = \frac{\beta_2}{\sin\varphi_t}.
\tag{2.4}
$$

The corresponding displacement vectors of the above equations can be written as

$$
\begin{aligned}
u_{x1} &= \frac{i\omega}{\alpha_1}(A_0+A_1)\sin\theta_i e^{i\omega(px+\frac{\cos\theta_i}{\alpha_1}z-t)} + \frac{\omega}{\beta_1}B_1\cos\varphi_r e^{i\omega(px-\frac{\cos\varphi_r}{\beta_1}z-t)}, \\
u_{x2} &= \frac{i\omega}{\alpha_2}A_2\sin\theta_t e^{i\omega(px+\frac{\cos\theta_i}{\alpha_2}z-t)} + \frac{\omega}{\beta_2}B_2\cos\varphi_t e^{i\omega(px+\frac{\cos\varphi_r}{\beta_2}z-t)}, \\
u_{z1} &= \frac{i\omega}{\alpha_1}(A_0-A_1)\cos\theta_i e^{i\omega(px+\frac{\cos\theta_i}{\alpha_1}z-t)} + \frac{\omega}{\beta_1}B_1\sin\varphi_r e^{i\omega(px-\frac{\cos\varphi_r}{\beta_1}z-t)}, \\
u_{z2} &= \frac{i\omega}{\alpha_2}A_2\cos\theta_t e^{i\omega(px+\frac{\cos\theta_i}{\alpha_1}z-t)} - \frac{\omega}{\beta_2}B_2\sin\varphi_t e^{i\omega(px+\frac{\cos\varphi_r}{\beta_1}z-t)},
\end{aligned}
\tag{2.5}
$$

where u_{x1}, u_{x2}, u_{z1}, and u_{z2} are the tangential component in the first medium, the tangential component in the second medium, the normal component in the first medium and the normal component in the second medium respectively. In order to arrive to the final Zoeppritz equation, we have to impose boundary conditions on equation (2.3). Some of the boundary conditions are from the definition of stress and strain. Therefore, before we deal with the the boundary conditions, it is worth writing down the important relationship between stress

and strain components. According to the usual geometry depicted in Figure 2.1.

$$\sigma_{ij} = \lambda \Delta \delta_{ij} + 2\mu e_{ij}, \tag{2.6}$$

where σ_{ij} is the stress component, e_{ij} is the strain component and Δ is the dilation which is the sum of the normal strain components. the indices ij refers to the various combination x, y and z directions. the constants λ and μ (shear modulus) are the Lamé parameters. The following boundary conditions holds at $z = 0$.

Boundary Condition 1: The tangential displacement component is continuous across the boundary,

$$u_{x1} = u_{x2}. \tag{2.7}$$

Using the first two equations of equation (2.5) and setting z=0, we can easily see that

$$\frac{\omega}{\alpha_1} A_0 \cos \theta_i - \frac{\omega}{\alpha_1} A_1 \cos \theta_i + \frac{\omega}{\beta_1} B_1 \sin \varphi_r = \frac{\omega}{\alpha_2} A_2 \cos \theta_t - \frac{\omega}{\beta_2} B_2 \sin \varphi_t. \tag{2.8}$$

Boundary Condition 2: The normal displacement component is continuous across the boundary,

$$u_{z1} = u_{z2}. \tag{2.9}$$

Using the third and the fourth equations of equation (2.5) and setting z=0, we have

$$\frac{\omega}{\alpha_1} A_0 \sin \theta_i + \frac{\omega}{\alpha_1} A_1 \sin \theta_r + \frac{\omega}{\beta_1} B_1 \cos \varphi_r = \frac{\omega}{\alpha_2} A_2 \sin \theta_t + \frac{\omega}{\beta_2} B_2 \cos \varphi_t. \tag{2.10}$$

Boundary Condition 3: The normal stress component is continuous across the boundary. According to equation (2.6), the normal stress component can be written as

$$\sigma_{zz} = \lambda(e_{xx} + e_{zz}) + 2\mu e_{zz}, \tag{2.11}$$

where

$$
\begin{aligned}
e_{xx} &= \frac{\partial u_x}{\partial x}, \\
e_{zz} &= \frac{\partial u_z}{\partial z}.
\end{aligned}
\tag{2.12}
$$

We can rewrite equation (2.11) as

$$\sigma_{zz} = (\lambda + 2\mu)\frac{\partial u_z}{\partial z} + \lambda\frac{\partial u_x}{\partial x}. \tag{2.13}$$

At this point, we need to express the Lamé parameters in terms of density, P-wave and S-wave velocities in order to express the final result solely in terms of density and velocities of the media. To this end, the following relationships are useful

$$
\begin{aligned}
\alpha &= (\frac{\lambda + 2\mu}{\rho})^{\frac{1}{2}}, \\
\beta &= (\frac{\mu}{\rho})^{\frac{1}{2}}.
\end{aligned}
\tag{2.14}
$$

Using these relations in equation (2.13), we have

$$
\sigma_{zz} = \rho\alpha^2\frac{\partial u_z}{\partial z} + \rho\beta^2\frac{\partial u_x}{\partial x}.
\tag{2.15}
$$

From which follows,

$$
\begin{aligned}
(\sigma_{zz})_1 &= (\sigma_{zz})_2 \\
\rho_1\alpha_1^2\frac{\partial u_{z1}}{\partial z} + \rho_1\beta_1^2\frac{\partial u_{x1}}{\partial x} &= \rho_2\alpha_2^2\frac{\partial u_{z2}}{\partial z} + \rho_2\beta_2^2\frac{\partial u_{x2}}{\partial x}.
\end{aligned}
\tag{2.16}
$$

Substituting equations (2.5) into equation (2.13) and applying differentiation, one can easily show that

$$
\begin{aligned}
\rho_1\alpha_1(1 - 2\beta_1^2 p^2)\frac{\omega}{\alpha_1}A_0 &+ \rho_1\alpha_1(1 - 2\beta_1^2 p^2)\frac{\omega}{\alpha_1}A_1 - 2\rho_1\beta_1^2 p\cos\varphi_r\frac{\omega}{\beta_1}B_1 \\
&= \rho_2\alpha_2(1 - 2\beta_2^2 p^2)\frac{\omega}{\alpha_2}A_2 - 2\rho_2\beta_2^2 p\cos\varphi_t\frac{\omega}{\beta_2}B_2.
\end{aligned}
\tag{2.17}
$$

Boundary Condition 4: The tangential stress component is continuous across the boundary i.e According to the relation given by equation (2.6), the tangential stress component can be written as

$$
\sigma_{xz} = 2\mu e_{xz}.
\tag{2.18}
$$

But e_{xz} can be expressed as

$$
e_{xz} = \frac{1}{2}(\frac{\partial u_z}{\partial x} + \frac{\partial u_x}{\partial z}).
\tag{2.19}
$$

Using this expression and equation (2.14) in equation (2.18), we can write the boundary condition as

$$
\rho_1\beta_1^2(\frac{\partial u_{z1}}{\partial x} + \frac{\partial u_{x1}}{\partial z}) = \rho_2\beta_2^2(\frac{\partial u_{z2}}{\partial x} + \frac{\partial u_{x2}}{\partial z}).
\tag{2.20}
$$

Employing equations (2.5) in equation (2.20) and applying differentiation, we have

$$2\rho_1\beta_1^2 p \cos\theta_i \frac{\omega}{\alpha_1} A_0 - 2\rho_1\beta_1^2 p \cos\theta_i \frac{\omega}{\alpha_1} A_1 - \rho_1\beta_1(1 - 2\beta_1^2 p^2)\frac{\omega}{\beta_1} B_1$$

$$= \rho_2\beta_2^2 p \cos\varphi_t \frac{\omega}{\beta_2} A_2 + \rho_2\beta_2(1 - 2\beta_2^2 p^2)\frac{\omega}{\alpha_2} B_2. \tag{2.21}$$

The four equations (2.8) ,(2.10), (2.17), and (2.21) are the final result after imposing the boundary conditions. But they are expressed in terms of the potential amplitudes. As the reflection and transmission coefficients are defined as ratios of the displacement amplitude, we can rewrite the four equations as

$$-\alpha p R_{pp} - \cos\varphi_r R_{ps} + \alpha_2 p T_{pp} + \cos\varphi_t T_{ps} = \alpha p,$$

$$\cos\theta_i R_{pp} - \beta_1 p R_{ps} + \cos\theta t T_{pp} - \beta_2 p T_{ps} = \cos\theta_i,$$

$$-\rho_1\alpha_1(1 - 2\beta_1^2 p^2) R_{pp} - 2\rho_1\beta_1^2 p \cos\varphi_r R_{ps} + \rho_2\alpha_2(1 - 2\beta_2^2 p^2) T_{pp}$$

$$-2\rho_2\beta_2^2 p \cos\varphi_t T_{pp} = \rho_1\alpha_1(1 - 2\beta_1^2 p^2),$$

$$2\rho_1\beta_1^2 p \cos\theta_i R_{pp} + \rho_1\beta_1(1 - 2\beta_1^2 p^2) R_{ps} + \rho_2\beta_2^2 p \cos\varphi_t T_{pp}$$

$$+\rho_2\beta_2(1 - 2\beta_2^2 p^2) T_{ps} = 2\rho_1\beta_1^2 p \cos\theta_i, \tag{2.22}$$

where

$$R_{pp} = \frac{A_1}{A_0}, R_{ps} = \frac{\alpha_1}{\beta_1}\frac{B_1}{A_0}, T_{pp} = \frac{\alpha_1}{\alpha_2}\frac{A_2}{A_0}, T_{ps} = \frac{\alpha_1}{\beta_2}\frac{B_2}{A_0},$$

which are the reflection and transmission coefficients. The elegant way of expressing the Zoeppritz equation is in matrix form. In matrix form, equation (2.22) takes the form

$$MX = Y, \tag{2.23}$$

where

$$M = \begin{pmatrix} -\alpha p & -\cos\varphi_r & \alpha_2 p & \cos\varphi_t \\ \cos\theta_i & -\beta_1 p & \cos\theta t & -\beta_2 p \\ -\rho_1\alpha_1(1 - 2\beta_1^2 p^2) & -2\rho_1\beta_1^2 p \cos\varphi_r & \rho_2\alpha_2(1 - 2\beta_2^2 p^2) & -2\rho_2\beta_2^2 p \cos\varphi_t \\ 2\rho_1\beta_1^2 p \cos\theta_i & \rho_1\beta_1(1 - 2\beta_1^2 p^2) & \rho_2\beta_2^2 p \cos\varphi_t & 2\rho_2\beta_2^2 p \cos\varphi_t \end{pmatrix},$$

$$X = \begin{pmatrix} R_{pp} \\ R_{ps} \\ T_{pp} \\ T_{ps} \end{pmatrix}, \quad Y = \begin{pmatrix} \alpha p \\ \cos\theta_i \\ \rho_1\alpha_1(1 - 2\beta_1^2 p^2) \\ 2\rho_1\beta_1^2 p \cos\theta_i \end{pmatrix}.$$

The resulting equation is a specific case in which we have a down-going P-wave (incident), reflected P-wave, reflected S-wave, transmitted P-wave and transmitted S-wave. Following a similar procedure given above, it is also possible to generalize for all possible incidence, reflected and transmitted waves. For AVO analysis, the R_{pp} and R_{ps} are the most exploited reflection coefficients i.e in typical seismic experiment using compressional wave sources , P-waves, and receiving the P-wave component and/ or S-wave component by receivers. In this thesis, only the R_{pp} of equation (2.23) part is used.

2.3 Approximation of the Zoeppritz Equations

Solving for R_{pp} from Zoeppritz equation, the exact PP wave reflection coefficients in terms of the ray parameter p, velocity and density of the media can be expressed as

$$R_{pp} = \frac{(b\frac{\cos\theta_i}{\alpha_1} - c\frac{\cos\theta_t}{\alpha_2})F - (a + d\frac{\cos\theta_i}{\alpha_1}\frac{\cos\theta_t}{\alpha_2})Hp^2}{D} \tag{2.24}$$

where

$$
\begin{aligned}
a &= \rho_2(1 - 2\beta_2 p^2) - \rho_1(1 - 2\beta_1 p^2), \\
b &= \rho_2(1 - 2\beta_2 p^2) + 2\rho_1\beta_1 p^2, \\
c &= \rho_1(1 - 2\beta_1 p^2) + 2\rho_2\beta_2 p^2, \\
d &= 2(\rho_2\beta_2 - \rho_1\beta_1^2), \\
F &= b\frac{\cos\varphi_r}{\beta_1} + c\frac{\cos\varphi_t}{\beta_2}, \\
H &= a - d\frac{\cos\theta_t}{\alpha_2}\frac{\cos\varphi_r}{\beta_1}, \\
D &= \frac{\det M}{\alpha_1\alpha_2\beta_1\beta_2}.
\end{aligned}
\tag{2.25}
$$

Although the above equation shows the exact transformation of the reflection coefficient, it is non-linear and complicated with respect to the the model parameters of the medium. For this reason, under the assumption of small property contrasts, expanding and retaining only linear terms, a number of papers has been published on linearized approximations which can be directly used for AVO inversion. The next section is devoted in revisiting the approximations

2.3.1 Bortfeld's approximation

The first attempt in making an approximation on Zoeppritz equation to make it more useful for AVO analysis was done by Bortfeld (1962). This equation is given by

$$R_{pp}(\theta_i) = \frac{1}{2} \ln \frac{\alpha_2 \rho_2 \cos \theta_i}{\alpha_1 \rho_1 \cos \theta_t} + \frac{\sin^2 \theta_i}{\alpha_1^2} (\beta_1^2 - \beta_2^2)(2 + \frac{\ln \frac{\rho_2}{\rho_1}}{\ln \frac{\alpha_2}{\alpha_1} - \ln \frac{\alpha_2 \beta_1}{\alpha_1 \beta_2}}). \qquad (2.26)$$

This approximation is easier to understand how the reflection coefficient is related to the angles and the earth model parameters than the exact equation (2.24). But it still non-linear with respect to the Earth model parameters which makes it not suitable for AVO inversion.

2.3.2 Aki and Richards's approximation

The Bortfield's approximation was revisited by Richards and Frasier (1976) and later re-defined by Aki and Richards (1980) which is linear in three model parameters, the P-wave reflectivity, S-wave reflectivity, and density reflectivity. This equation is given by,

$$R_{PP}(\theta_i) = \frac{1}{2}[1 + \tan^2(\theta)]\frac{\Delta \alpha}{\alpha} - 4\gamma^2 \sin^2(\theta)\frac{\Delta \beta}{\beta} + \frac{1}{2}[1 - 4\gamma^2 \sin^2(\theta)]\frac{\Delta \rho}{\rho}, \qquad (2.27)$$

where

$$
\begin{aligned}
\Delta \alpha &= \alpha_2 - \alpha_1, \\
\Delta \beta &= \beta_2 - \beta_1, \\
\Delta \rho &= \rho_2 - \rho_1, \\
\alpha &= \frac{1}{2}(\alpha_2 + \alpha_1), \\
\beta &= \frac{1}{2}(\beta_2 + \beta_1), \\
\rho &= \frac{1}{2}(\rho_2 + \rho_1), \\
\gamma &= \frac{\beta}{\alpha}, \\
\theta_t &= \sin^{-1}(\frac{\alpha_2}{\alpha_1} \sin(\theta_i)), \\
\theta &= \frac{1}{2}(\theta_i + \theta_t).
\end{aligned}
\qquad (2.28)
$$

This is a very elegant approximation in the sense that it is expressed in terms of three model parameters independently. It is even the starting equation for other approximations which are described below. The parameters defined in the Aki and Richards's approximation are used in the next sections.

2.3.3   Shuey's approximation

This approximation is modified version of the Aki and Richards's approximation. But in this case the equation was rearranged by Shuey (1985) to incorporate the Poisson's ratio as parameter instead of the S-wave velocity. The equation has the from

$$R(\theta_i) = R_p + R_g \sin^2 \theta + \frac{1}{2}\frac{\Delta\alpha}{\alpha}(\tan^2 \theta - \sin^2 \theta) \tag{2.29}$$

where

$$R_p = \frac{1}{2}(\frac{\Delta\alpha}{\alpha} + \frac{\Delta\rho}{\rho}),$$

$$R_g = [R_p A_0 + \frac{\Delta\sigma}{(1-\sigma)^2}],$$

$$\sigma = \frac{\frac{1}{\gamma^2} - 2}{2(\frac{1}{\gamma^2} - 1)},$$

$$\Delta\sigma = \sigma_2 - \sigma_1,$$

$$A_0 = B - 2(1 + B)\frac{1 - 2\sigma}{1 - \sigma},$$

$$B = \frac{2\frac{\Delta\alpha}{\alpha}}{R_p}.$$

Note that α, β and ρ has the same definition as in the Aki and Richards's approximation. The parameter σ is called the Poisson's ratio. This approximation is very useful if the analysis is in terms of the Poisson's ratios. It is an interesting property of rocks which is directly related to their elasticity.

2.3.4   Verm and Hilterman's approximation

Further simplification on Shuey's equation was by ignoring the last term and setting $\gamma = 4$ (Verm and Hilterman, 1995). The simplified two term R_{pp} has the form,

$$R_{pp}(\theta_i) = R_p + G_p \sin^2 \theta, \tag{2.30}$$

where

$$G_p = \frac{9}{4}\Delta\sigma - R_p.$$

The first term in equation (2.30), R_p, is called the Intercept and G_p is called the Gradient. This is often used in AVO analysis and is called Intercept-Gradient analysis.

2.3.5 Smith and Gidlow's approximation

Smith and Gidlow (2003) rearranged the Aki and Richards's approximation as

$$R_{pp}(\theta_i) = \frac{1}{2}(\frac{\Delta\alpha}{\alpha} + \frac{\Delta\rho}{\rho}) - 2\gamma^2\sin^2\theta(2\frac{\Delta\beta}{\beta} + \frac{\Delta\rho}{\rho}) + \frac{1}{2}\tan^2\theta\frac{\Delta\alpha}{\alpha}. \tag{2.31}$$

Further, under the assumption that density follows Gardner's relationship, equation (1.1), they put R_{pp} as only two terms where the model parameters are the P-wave reflectivity and S-wave reflectivity. This equation is given by

$$R_{pp}(\theta_i) = \frac{1}{2}(1 + a + \tan^2\theta - 4a\gamma^2\sin^2\theta)\frac{\Delta\alpha}{\alpha} - 4\gamma^2\sin^2\theta\frac{\Delta\beta}{\beta} \tag{2.32}$$

where $a = 0.25$ is constant in Gardner's equation, equation (1.1). In addition to the above given approximation, they also define what is called fluid factor making use of Castagna's equation, equation (1.3). The fluid factor is defied as

$$\Delta F = \frac{\Delta\alpha}{\alpha} - c\gamma\frac{\Delta\beta}{\beta} \tag{2.33}$$

where $c = 1.16$ is gradient of the Castagna's equation. The equality in this equation don't hold for hydrocarbon reservoirs and it is useful indicator for hydrocarbon indicator by analyzing the deviation from this relationship (Castagna et al., 1998).

2.3.6 Fatti's approximation

This approximation was proposed by Fatti et al. (1994) by rearranging the Aki and Richards's approximation and expressing it in terms of P-wave impedance, S-wave impedance, and density reflectivity. The resulting equation has the form

$$R_{pp}(\theta_i) = \frac{1}{2}(1 + \tan^2\theta)\frac{\Delta I_p}{I_p} - 4\gamma^2\sin^2\theta\frac{\Delta I_s}{I_s} - \frac{1}{2}(\tan^2\theta - 4\gamma^2\sin^2\theta)\frac{\Delta\rho}{\rho}, \tag{2.34}$$

where

$$\frac{\Delta I_p}{I_p} = \frac{\alpha_2\rho_2 - \alpha_1\rho_1}{\alpha_2\rho_2 + \alpha_1\rho_1},$$
$$\frac{\Delta I_s}{I_s} = \frac{\beta_2\rho_2 - \beta_1\rho_1}{\beta_2\rho_2 + \beta_1\rho_1}. \tag{2.35}$$

Assuming very small change in density, the last term in equation (2.34) cab be ignored and results in the following expression

$$R_{pp}(\theta_i) = \frac{1}{2}(1 + \tan^2\theta)\frac{\Delta I_p}{I_p} - 4\gamma^2\sin^2\theta\frac{\Delta I_s}{I_s}. \tag{2.36}$$

The model parameters in this equation are the P-wave and S-wave impedance. This approximation is used in Chapter 4 for the two-term AVO inversion problem.

2.4   Comparison of approximations

In this section, the various approximations are compared to the exact equation. Using the P-wave velocity, S-wave velocity and density models in Table 2.1, the reflection coefficients are calculated and plotted against the angle of incidence up to the critical angle.

Model	Layer	P-wave velocity (m/s)	S-wave velocity (m/s)	density (kg/m^3)
1	1	3000	1800	2200
	2	3200	2000	2250
2	1	3000	1800	2200
	2	4000	2500	2400

Table 2.1: Two layered models which are used for comparison of linear approximations to Zoeppritz equation (R_{pp}).

Equations (2.27), (2.31), and (2.34) are the same: the later two are rearranged forms of Aki and Richards's approximation. Figure 2.2 shows the comparison of equations (2.24), (2.27), (2.29), (2.30), (2.32), and (2.36). For small layer contrast (model 1), all the approximations are in a good agreement with the exact equation up to angle range from 30^0 to 40^0. Beyond this angle, some deviations are seen near the critical angle particularly for Shuey's and Verm and Hilterman's equations while the Fatti's approximation extends quite well up to 50^0. The rest of the approximations (Aki and Richards's and Smith and Gidlow's approximation) exactly coincides with the exact equation up to the critical angle. A similar comparison using model 2 shows a significant deviation. The Shuey's approximation is in a good agreement up to only 25^0. Aki and Richards's and Smith and Gidlow's approximation are the same equations. One is the rearrangement of the other so they exactly coincides with each other but show some deviations from the exact solution. The Fatti's approximation is almost exact as the later two up to 35^0. The response of the reflection coefficient may vary from model to model i.e we may observe different behavior for each approximation. In general, those comparison show that at larger angle of incidence and/or larger model contrast, the deviation from the exact is significant.
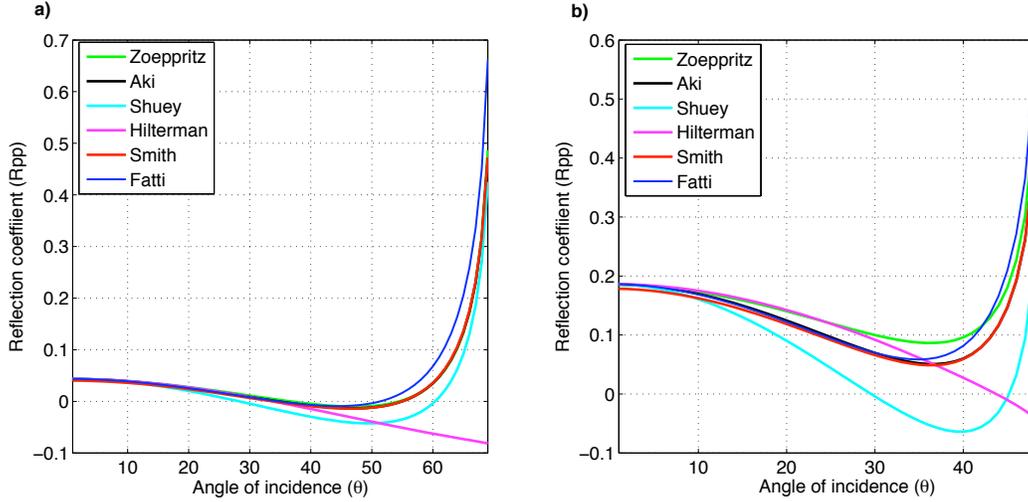
Figure 2.2: Comparison of approximations using the two models given in Table 2.1.
(**a**) using model 1, and (**b**) using model 2.

2.5   Effect of $\gamma = \frac{\beta}{\alpha}$ on forward modeling

All approximations but Shuey's and Verm and Hilterman's involve the S-wave velocity to
P-wave velocity ratio, γ, for forward modeling. This parameter is calculated as part of the
the modeling with the available velocity models. So, it is important to see the effect of
this parameter in the forward modeling. The easy way to see this effect is to calculate the
reflection coefficients for a given set of incidences angles by changing γ. In Figure 2.3a,
the reflection coefficient for the actual value of γ are calculated using Zoeppritz equation,
equation (2.24), and the Aki and Richards's approximation, equation (2.27) using model 1
in Table 2.1. In addition, three different values of γ which are off from the actual value were
used to calculate the reflection coefficients using the Aki and Richards's approximation. It
is clear that using the actual value of γ for this particular model coincides with the exact
reflection coefficient calculated using the Zoeppritz equation. The wrong values of γ lead
to small deviation from the exact solution for this particular model. Using model 2, the
deviations become larger as compared to model 1 (Figure 2.3b). This implies that as the
contrast between the layer model parameter is greater, the parameter γ has more influence
on forward modeling. Therefore, available background velocity (well-log information) should
be used to reduce this effect on forward modeling.
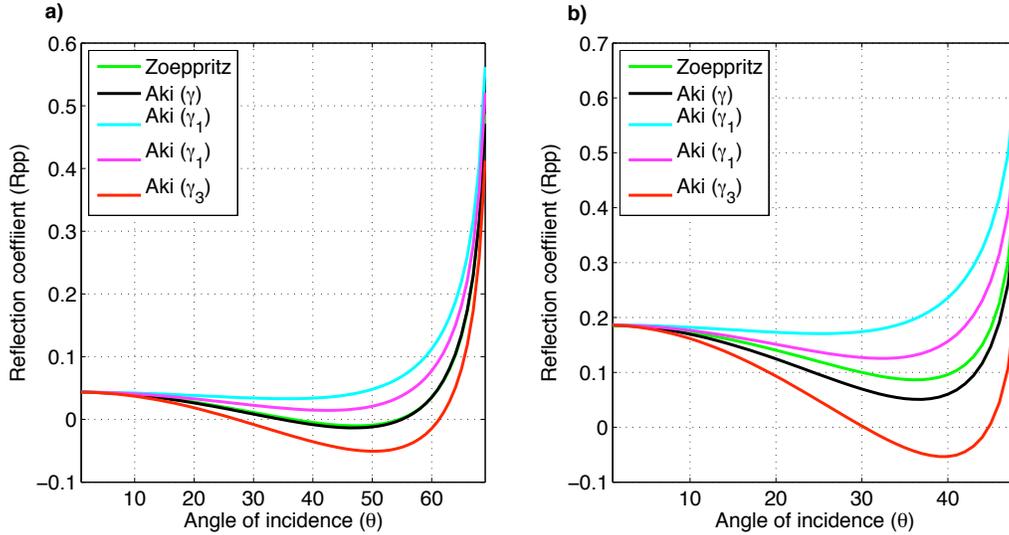
Figure 2.3: The effect of γ on forward modeling using the two models given in Table 2.1. (**a**) using model 1 for ($\gamma = 0.6129$, $\gamma_1 = 0.4129$, $\gamma_2 = 0.5129$, and $\gamma_3 = 0.7129$), and (**b**) using model 2 for $\gamma = 0.6143$, $\gamma_1 = 0.4143$, $\gamma_2 = 0.5143$, and $\gamma_3 = 0.7143$.

2.6 Ray Tracing Algorithms

The forward model for AVO inversion is based on the approximations provided in section 2. These approximations solely depend on the angle of incidence (the average of the angle of incidence and angle of transmission to be exact) at each layer where reflection occurs. In the actual seismic records, the data are recorded as a function of offset and two way travel time. The offset information should be mapped to angle using ray tracing methods. By assuming a velocity model, one can parametrize the layers either in terms of a constant depth (Δz) or a constant zero offset travel time (Δt). Using this parameterization, two ray tracing algorithms are given in the next subsections.

2.6.1 Ray tracing method 1

This ray tracing method is an approximation by assuming a straight line ray path from a source to a reflection point and then to a receiver (Walden, 1991). The ray parameter for a ray path reflected at n^{th} layer is defined as

$$p = \frac{\sin \theta_n}{\alpha_n}, \tag{2.37}$$
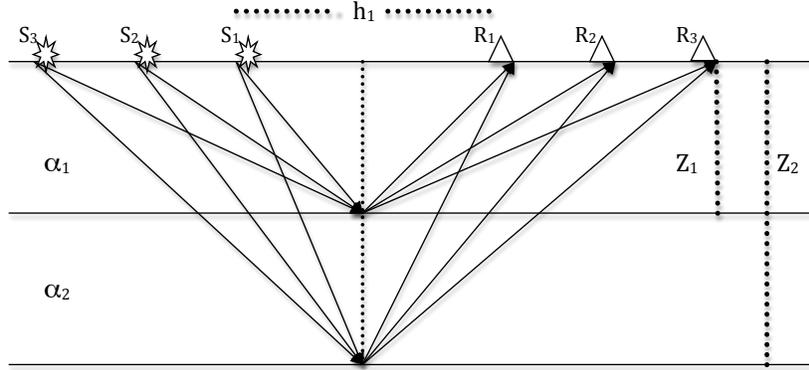
Figure 2.4: The source-receiver geometry for CDP gathers assuming straight line ray path from a source to a reflection point and from reflection point to the receiver. The symbols S_1, S_2, S_3 are sources and R_1, R_2, and R_3 are receivers. The depths of each layer represented by Z_1 and Z_2 with corresponding interval velocity α_1 and α_2. This geometry is used for ray tracing method 1.

where θ_n the angle of incidence on that layer and α_n is the interval velocity of the layer. For zero offset two-way travel time, $\tau_n = n\Delta t$, and offset, h. According to the geometry shown in Figure 2.4, the approximate two-way travel time, t_n, has the form

$$t_n^2 = \tau_n^2 + \left(\frac{h}{V_{rms}^n}\right)^2, \tag{2.38}$$

$$t_n = \sqrt{\tau_n^2 + \left(\frac{h}{V_{rms}^n}\right)^2}, \tag{2.39}$$

where V_{rms}^n is the root mean square velocity down to the nth layer. The ray parameter can also be derived from the two-way travel as

$$p = \frac{\partial t_n}{\partial h} = \frac{h}{t_n(V_{vrms}^n)^2}. \tag{2.40}$$

Combing equations (2.37) and (2.40), the angle of incidence at the nth layer takes the form

$$\sin\theta(t_n, h) = \frac{h\alpha_n}{t_n(V_{vrms}^n)^2}. \tag{2.41}$$

2.6.2   Ray tracing method 2

Unlike the first method, this method tracks the ray path from the source to the receiver according to the layer velocity to determine the angle of incidence and calculate the total
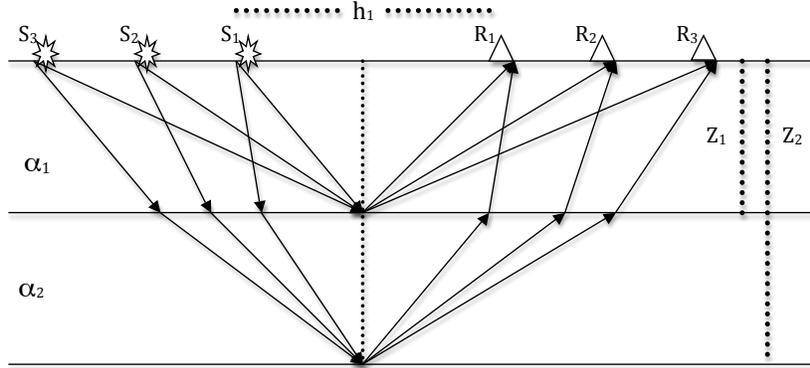
Figure 2.5: The exact ray tracing for CDP gathers. It has similar source-reciever geometry as Figure 2.4 except the ray paths which depend on change in the interval velocity of the layers. This geometry is used for the ray tracing method 2.

travel time for each path (Dahl and Ursin, 1991). I prefer to work by choosing an emerging angle instead of a ray parameter. A simplified geometry of the ray paths for 3 source-receiver and two-layered model is shown in Figure 2.5. The first step in this algorithm is to determine the angle by which the ray emerges from the source or the ray parameter for a given ray path that matches the given offset. For a given ray path, let the angle of incidences at each layer, down to a given layer (n^{th} layer) be $\theta_1, \theta_2, ..., \theta_i, ...\theta_N$. Using snell's law, the angle of incidence and the velocity model are related as

$$p = \frac{\sin \theta_1}{\alpha_1} = \frac{\sin \theta_2}{\alpha_2} = ... = \frac{\sin \theta_i}{\alpha_i} = ... = \frac{\sin \theta_n}{\alpha_n}. \tag{2.42}$$

Choosing an emerging angle, θ_e, and using the relationship given by equation (2.42), we have

$$\sin \theta_i = \frac{\alpha_i}{\alpha_1} \sin \theta_e. \tag{2.43}$$

Using the chosen emerging angle, the sum of the horizontal distance traveled by the ray in each layer should give the offset for that ray path,

$$X(\tau_n, \theta_e) = \sum_{i=1}^{n} \alpha_i^2 \Delta t \frac{\sin \theta_e,}{\sqrt{\alpha_1^2 - \alpha_i^2 \sin \theta_e^2}}, \tag{2.44}$$

where θ_e is the emerging angle, α_i velocity of the ith layer and X is the offset for the chosen emerging angle. To find the offset that approximate the actual recorded offset, h, we can

use the following search direction

$$\Delta\theta_e = \frac{\partial\theta_e}{\partial X}\Delta X. \tag{2.45}$$

From which follows

$$\theta_e^{j+1} = \theta_e^j + (h - X(\theta_e^j))f^n(\theta_e^j), \tag{2.46}$$

where

$$f^n(\theta_e^j) = \left(\frac{\partial X}{\partial\theta_e}\right)^{-1}\Bigg|_{\theta_e=\theta_e^j} = \left[\sum_{i=1}^n \alpha_i^4 \Delta t \frac{\cos\theta_e}{(\alpha_1^2 - \alpha_i^2\sin\theta_e^2)^{3/2}}\right]^{-1}\Bigg|_{\theta_e=\theta_e^j}. \tag{2.47}$$

and j refers to the iteration level which runs until the convergence of the calculated offset X to the actual value h within certain level of accuracy. Taking the final emerging angle (θ_h) that best approximate the offset, the total two-way travel time for the given ray path has the form

$$t_n(h) = t(\tau_n, h) = \sum_{i=1}^n \frac{\alpha_1 \Delta t}{\sqrt{\alpha_1^2 - \alpha_i^2\sin^2\theta_h}} \tag{2.48}$$

where $\tau_n = n\Delta t$ is the zero offset two-way travel time. Finally, the angle of incidence, $\theta(t_n, h)$, can be calculated using the relation

$$\sin\theta(t_n, h) = \frac{\alpha_n}{\alpha_1}\sin\theta_h. \tag{2.49}$$

Note that θ_h is the actual emerging angle at a given offset h and θ_n, the sin inverse of the right side of equation (2.49), is the angle of incidence at the nth layer.

2.7 Summary

In this chapter, the theoretical background of AVO modeling has been reviewed. The Zoeppritz equations were derived using two fundamental boundary conditions (continuity of the displacement and stress components across the boundary). The non-linearity and complexity of the Zoeppritz equations and the linear approximations were discussed. The various approximations were compared with the exact equation using two two-layered models which have small and larger layer contrasts. The larger the angle of incidence and/or the layer contrasts, the larger the deviation is from the exact. In this situation, the inversion result may lead to a wrong prediction. Similarly, the effect of γ should also be taken under consideration as it has influence on the forward modeling. Such types of modeling problems

can be reduced by having additional geological information about the region under study (well-log velocity) and limiting the maximum angle according to the chosen approximation. In the last section of this chapter, two-ray tracing algorithms have been discussed. The first method is an approximation which assumes a straight ray path but with root mean square velocity. The second method is nearly exact with certain level of accuracy as it tracks the change in the interval velocity of the layers. But it is more computationally expensive than the first method. The first method is used for synthetic data examples and second method for real data examples in the following chapters.

CHAPTER 3

Bayesian Inversion Approach and Algorithms

3.1 Introduction

Ill-posed inverse problems like AVO inversion are very common in many geophysical problems. Such kind of problems may have infinitely many solutions that can fit a given observed data or the problem may be ill-conditioned. The Bayesian method has become most popular to formulate AVO inverse problem with prior information (Buland and Omre, 2003; Downton and Lines, 2004). This helps to stabilize the inversion and thereby increase the reliability of the estimated parameters. This chapter is devoted in general mathematical formulation of inverse problems with some prior distributions in Bayesian framework and to discuss inversion algorithms. In addition, a target oriented or window by window AVO inversion is presented.

3.2 Bayes' Theorem

Bayes' theorem is a probabilistic theorem that can be used to calculate conditional probabilities (Sivia and Skilling, 2006). Mathematically, it can be expressed as

$$P(\mathbf{m}|\mathbf{d}) = \frac{P(\mathbf{d}|\mathbf{m})P(\mathbf{m})}{P(\mathbf{d})}, \tag{3.1}$$

where $P(\mathbf{m}|\mathbf{d})$ is the posterior distribution, $P(\mathbf{d}|\mathbf{m})$ is the likelihood of the observed data, $P(\mathbf{d})$ is a non-vanishing marginal probability of the observed data, and $P(\mathbf{m})$ is the prior distribution which is a knowledge one has before observation. This theorem describes how the conditional probability of $\mathbf{m}$ given $\mathbf{d}$ is related to the converse probability of $\mathbf{d}$ given $\mathbf{m}$. Since the marginal probability of $\mathbf{d}$ is a constant, equation (3.1) is equivalent to

$$P(\mathbf{m}|\mathbf{d}) \propto P(\mathbf{d}|\mathbf{m})P(\mathbf{m}). \tag{3.2}$$

This equation indicates that Bayesian inference mainly depends on constructing or modeling the prior distribution of the problem. Therefore, a reasonable choice of the probability distributions is mandatory such that the results of inversion using this prior is agreeable.

3.3 Prior distributions

From Bayesian point of view, the additional information which can be added to constraint a given problem is called *a priori*. The choice of the prior distributions is the most controversial part of the bayesian analysis (Scales and Tenorio, 2001; Ulrych et al., 2001). This is because incorrect choice of the *a priori* leads to wrong inferences about the problem under consideration. It should be chosen on the basis of the type of problem, the target and kind of system under consideration. When prior distributions are used to formulate inverse problems, different prior distributions lead to different regularizations. For instance, regularizations are incorporated for stability, smoothing via Gaussian or l_2-norm (Tikhonov and Arsenin, 1987), sparsity via Cauchy norm (Sacchi and Ulrych, 1995), and edge preserving and image de-noising via total variation regularization (Vogel and Oman, 1998). The later two regularizations are non-quadratic regularizations which help to reinforce relatively larger parameters and suppress small fluctuations (noise). Discussing all the existing regularization methods is beyond the scope of this thesis. The focus of this thesis is to use Multivariate Gaussian and Multivariate $\mathbf{t}$ distribution to regularize AVO inversion. In the next sections, the behavior of these distributions are discussed.

Multivariate Gaussian distribution

Gaussian distribution function is a smooth and continuous function which tells us clusters of a given data around the average or mean(Johnson and Kotz, 1972). The Multivariate Gaussian distribution for n model parameters, $\mathbf{x}$, is given by

$$P(\mathbf{x}|\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{n/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) \right\}, \tag{3.3}$$

where $\boldsymbol{\Sigma}$ is a parameter covariance matrix. The univariate version of equation (3.3) has the form

$$P(x|\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \qquad (3.4)$$

where σ is the standard deviation of the model parameter, x.

Multivariate t distribution

The Multivariate **t** distribution is generalization of classical univariate Student **t** distribution which is another type of distribution useful for multivariate analysis (Chauanhai, 1994). It is known by a typical nature having heavy tails as compared to exponential distributions and they are more realistic in real world (Nadarajah and Korz, 2008). The general Multivariate **t** distribution is given by,

$$P(\mathbf{x}^i|\mu, \boldsymbol{\Psi}, \nu) = \frac{\Gamma(\frac{\nu+p}{2})|\boldsymbol{\Psi}|^{-1/2}}{(\pi\nu)^{p/2}\Gamma(\frac{\nu}{2})[1 + \frac{1}{\nu}(\mathbf{x}^i - \boldsymbol{\mu})^T\boldsymbol{\Psi}^{-1}(\mathbf{x}^i - \boldsymbol{\mu})]^{\frac{(\nu+p)}{2}}}. \qquad (3.5)$$

The model $\mathbf{x}^i$ is a p dimensional vector whose elements are in the range $-\infty < x_j < \infty$ for $j = 1, 2, 3, ..., p$ and having a p dimensional vector center, $\boldsymbol{\mu}$. The degree of correlation is represented by a $p \times p$ dimensional scale matrix, $\boldsymbol{\Psi}$. The parameter ν is the degree of freedom or the shaping parameter which controls the shape of the distribution. Before discussing the multivariate case, it is easier to see the Univariate **t** distribution in comparison with the Univariate Gaussian distribution. For $p = 1$, the Univariate form of equation (3.6) has the form .

$$P(x|\mu, \sigma, \nu) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})[1 + \frac{1}{\nu\sigma^2}(x-\mu)^2]^{\frac{(\nu+1)}{2}}}, \qquad (3.6)$$

where σ is the scale parameter. For $\nu = 1$, this probability distribution is called Univariate Cauchy probability distribution. In Figure (3.1), the probability density function is plotted to investigate the heavy tail of the Cauchy distribution and its similarity and difference as compared to the Gaussian distribution. It is easy to observe that the Cauchy probability distribution has heavy tails that allow larger probability to larger parameters (x). In addition, increasing the degree of freedom, the Univariate **t** distribution approaches the Gaussian distribution. In this sense, the Univariate **t** distribution can be considered as a generalization of Gaussian distribution and this makes it more flexible to model parameters of interest. It is also possible to see how the degree of freedom changes the shape of the Multivariate **t** distribution. Setting $p = 2$ in equation (3.6) results in Bivariate t distribution
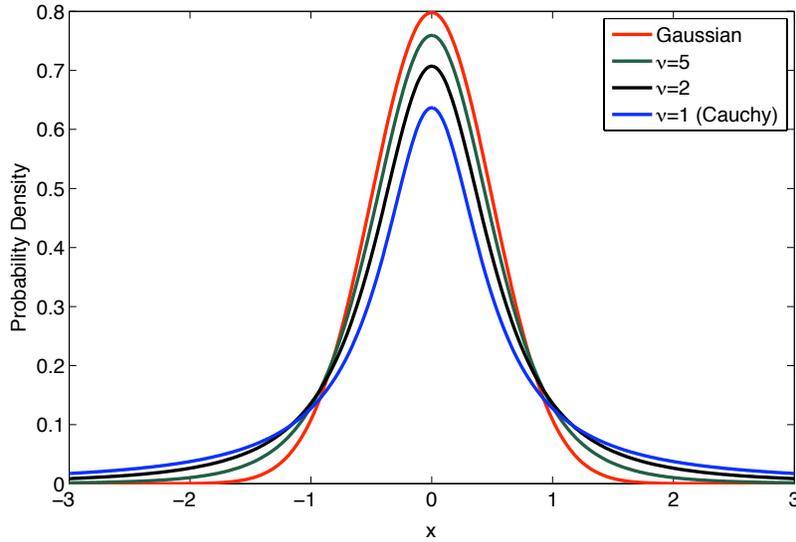
Figure 3.1: Probability density of the Univariate Gaussian (red), Univariate Cauchy (blue) for $\nu = 1$, and Univariate t for $\nu = 2$ and $\nu = 5$

given by

$$P(\mathbf{x}^i|\mu, \boldsymbol{\Psi}, \nu) = \frac{\Gamma(\frac{\nu+2}{2})|\boldsymbol{\Psi}|^{-1/2}}{\pi\nu\Gamma(\frac{\nu}{2})[1 + \frac{1}{\nu}(\mathbf{x}^i - \boldsymbol{\mu})^T \boldsymbol{\Psi}^{-1}(\mathbf{x}^i - \boldsymbol{\mu})]^{\frac{(\nu+2)}{2}}}. \tag{3.7}$$

where $\mathbf{x}$ is a 2×1 vector and $\boldsymbol{\Psi}$ is a 2×2 scale matrix. Figures 3.2 and 3.3 show plots of the probability density for different values of degree of freedom in order to investigate the heavier-tail nature of the Bivariate $\mathbf{t}$ distribution. In Figure 3.2, the peak is sharper for smallest ν and the peak starts to spread at larger values. Figure 3.3 is only the tail part of the Bivariate $\mathbf{t}$ distribution. It is easy to see that it has heavier tail at the smallest degree of freedom ($\nu = 1$, Bivariate Cauchy Distribution). The degree of freedom can therefore be used as a very good tuning parameter depending on the behavior of the model parameters. It has been shown that the when a heavier-tail prior distribution is used, the sparser is the solution in the estimated parameters(Sacchi and Ulrych, 1995; Downton and Lines, 2004). In this thesis, $\nu = 1$ is chosen in order to have sparsity in the solution.

3.4 Inversion Formulation

According to equation (3.2), given the likelihood function and the prior distribution, the solution of the problem is the model parameters that maximize the posterior distribution. In other words, the combination of the likelihood and the prior information leads to an
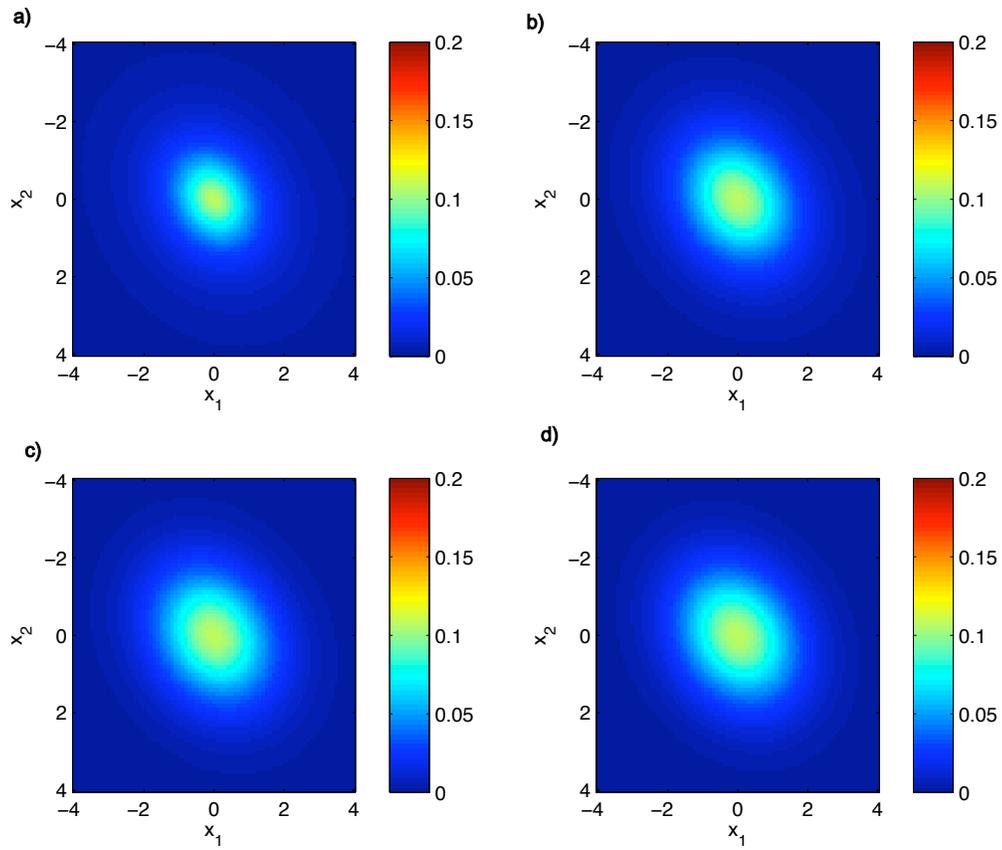
Figure 3.2: The peak of the (**a**) the Bivariate Cauchy Probability distribution ($\nu = 1$), and the Bivariate **t** distributions for (**b**) $\nu = 2$, **c**) $\nu = 5$, and **d**) $\nu = 10$.
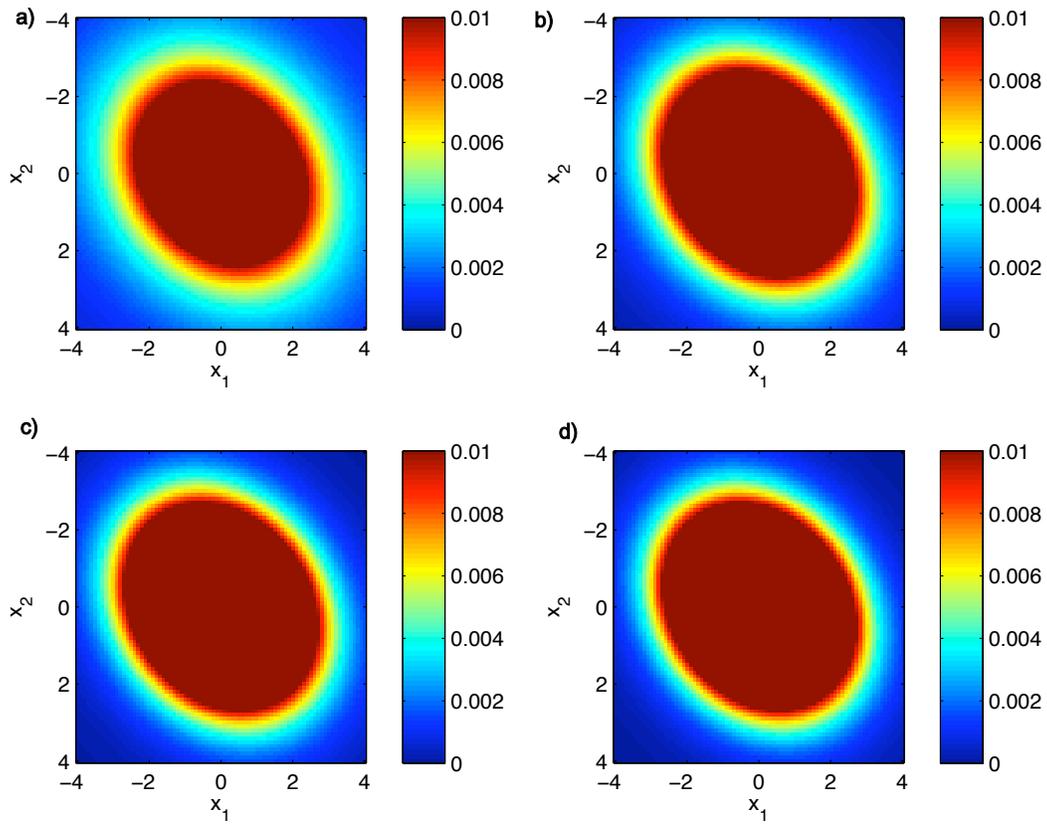
Figure 3.3: The tail of **a**) the Bivariate Cauchy distribution ($\nu = 1$), and the Bivariate **t** distributions for **b**) $\nu = 2$, **c**) $\nu = 5$, and **d**) $\nu = 10$.

objective function that has to be minimized such that it maximizes the posterior distribution. Usually, the objective function of inverse problem has two terms. One that takes care of the observation and the other the prior information. This can be written as

$$J(\mathbf{m}) = \phi_d + \mu_h \phi_m \tag{3.8}$$

where J is the objective function, $\mathbf{m}$ represents the model parameters, and μ_h is called the hyper-parameter to be determined such that it honors both the data and the prior. The functions ϕ_d and ϕ_m are results of the likelihood function and the prior distribution respectively. Lets take a typical inverse problem in which the relationship between the data and the model parameters are expressed as

$$\mathbf{d} = \mathbf{Lm} + \mathbf{n}, \tag{3.9}$$

where $\mathbf{d}$ is the observed data, $\mathbf{L}$ is a linear operator which depend on the physics model that governs the system, and $\mathbf{n}$ is noise in the data and some theoretical error. Assuming the noise terms are independent and Gaussian, the noise can be modeled using Multivariate Gaussian probability distribution function given by

$$P = P_o \exp\{-\frac{1}{2}\mathbf{n}^T \mathbf{C}_d^{-1} \mathbf{n}\}, \tag{3.10}$$

where $\mathbf{C_d}$ is the noise covariance matrix. Using equation (3.9) and (3.10), the likelihood function of the data can be expressed as

$$P(\mathbf{d}|\mathbf{m}) = P_o \exp\{-\frac{1}{2}(\mathbf{d} - \mathbf{Lm}))^T \mathbf{C_d}^{-1}(\mathbf{d} - \mathbf{Lm})\}, \tag{3.11}$$

where

$$P_o = \frac{1}{(2\pi)^{(n)/2}|\mathbf{C_d}|^{1/2}}, \tag{3.12}$$

and n is the number of data points. Assuming a certain prior distribution, the posterior distribution takes the form

$$P(\mathbf{m}|\mathbf{d}) \propto \exp\{-\frac{1}{2}(\mathbf{d} - \mathbf{Lm})^T \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{Lm}) - \mu_h R(\mathbf{m})\}, \tag{3.13}$$

Maximizing the posterior distribution leads to an objective function which has similar structure as equation (3.8)

$$J(\mathbf{m}) = \frac{1}{2}(\mathbf{d} - \mathbf{Lm})^T \mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{Lm}) + \mu_h R(\mathbf{m}), \tag{3.14}$$

where $R(\mathbf{m})$ is the regularization from *a prior* distribution. In the next two chapters, the construction of the objective functions for two-term and three-term AVO inversion is given in detail. Minimizing the objective function is equivalent to differentiating the function with respect to $\mathbf{m}$ and setting the resulting equation to zero which leads to

$$(\mathbf{L}^T \mathbf{C}_d^{-1} \mathbf{L} + \mu_h \mathbf{Q})\mathbf{m} = \mathbf{L}^T \mathbf{C}_d^{-1} \mathbf{d}, \tag{3.15}$$

where the matrix $\mathbf{Q}$ is a regularization term which depend on the choice of the prior distribution. Although the original relationship between the data and the the model is linear, there is a possibility for the final solution to be either linear or non-linear. If the regularization term is quadratic in $\mathbf{m}$, $\mathbf{Q}$ is independent of the the model $\mathbf{m}$. If the regularization term is non-quadratic in $\mathbf{m}$, $\mathbf{Q}$ is dependent on the model $\mathbf{m}$. The algorithms to solve this two possibilities are given in the next sections.

3.4.1   Parameter selection

The parameter μ_h can be estimated in two ways. By assuming a range of μ_h values and doing the inversion for each values of the selected parameters.

1. Using the Chi-square test (Sacchi et al., 1998): it uses the Chi-square distribution given by

   $\chi^2 = \sum_{j=1}^{N} \frac{e_j^2}{\sigma_j^2}$

   where σ_j is the noise standard deviation, e_j is the trade-off between the observed data and predicted data from a solution. Plotting the χ^2 versus μ and selecting $\mu = \mu_h$ value that correspond to the number of data points on the χ^2 axis. In this method, the noise variance should be known.

2. Using trade-off-diagram (Sacchi, 1998): plotting ϕ_d (misfit) versus ϕ_m (model norm) according to equation (3.8) and then selecting a value of $\mu = \mu_h$ at which large fluctuation occurs on a plotted curve.

Figure (3.4) is simple typical plots which show the relationship between the data misfit and the hyper-parameter. This parameter should be selected such that it honors both the data and the prior (model norm). Both methods are expensive as they need large number of trials for the selected range of μ values. Beside this, the misfit and model norm should be calculated. Nevertheless, they are still important for small data sets. In AVO inversion, if we have large number of CDP gathers for 2D or 3D profile of the model parameters, it is very difficult to implement any of methods for parameter selection.
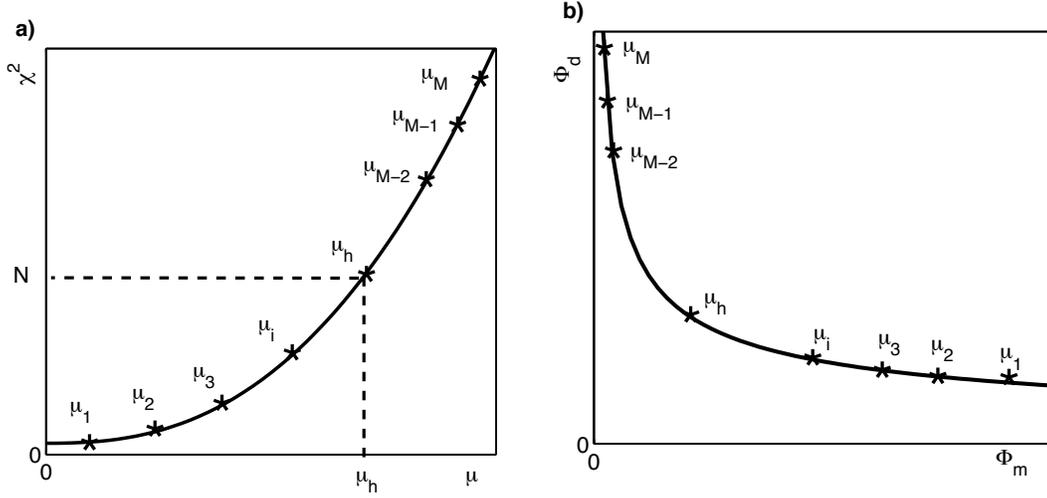
Figure 3.4: (**a**) The Chi-square test and (**b**) the trade-off curve for N observation data points and M number of μ_h trials

3.4.2 The least squares method

Least squares (LS) method refers to a technique of minimizing the square of the distance between the observed data and fitted response using a linear model. Equation (3.15) is an ordinary least squares problem, if the regularization term is turned off and the noise (or error) covariance matrix, $\mathbf{C}_d$, is an identity matrix (assuming uncorrelated noise and same variance). In other words, the parameters are just estimated from the observed data only (no prior information). The solution of such kind of problem is given by

$$\mathbf{m} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{d}. \tag{3.16}$$

Given the data and the linear operator, the solution is just a simple inversion. But it doesn't grant the exactness of the solution unless the problem is a well-posed problem.

Table 3.1: LS Algorithm

1: Input data, $\mathbf{d}$.
2: Construct the operator $\mathbf{L}$,
3: Solve,
$\mathbf{m} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{d}.$
4: Output, $\mathbf{m}$.

3.4.3 Weighted Least squares method

The weighted least squares (WLS) method is used when some sort of weight (or prior information) is put into the problem. According to equation (3.15), it said to be weighted least squares if the hyper-parameter is different from zero and the regularization term $\mathbf{Q}$ is independent of the model parameter $\mathbf{m}$. But it has no restriction in the noise covariance matrix. If $\mathbf{Q}$ is an identity matrix the solution to the problem is called damped least square solution which can be written as

$$\mathbf{m} = (\mathbf{L}^T\mathbf{C}_d^{-1}\mathbf{L} + \mu_h\mathbf{I})^{-1}\mathbf{L}^T\mathbf{C}_d^{-1}\mathbf{d}. \tag{3.17}$$

Such regularization method also called Tikhonov regularization (Tikhonov and Arsenin, 1987). If $\mathbf{Q}$ is different from $\mathbf{I}$, the solution has the form

$$\mathbf{m} = (\mathbf{L}^T\mathbf{C}_d^{-1}\mathbf{L} + \mu_h\mathbf{Q})^{-1}\mathbf{L}^T\mathbf{C}_d^{-1}\mathbf{d}. \tag{3.18}$$

where $\mathbf{Q}$ is independent of the $\mathbf{m}$ and also different from $\mathbf{I}$. For instance, in quadratic regularization, $\mathbf{Q}$ can take inverse of the covariance matrix of the model parameters i.e assuming Gaussian prior distribution.

Table 3.2: WLS Algorithm

1: Input data; $\mathbf{d}$, covariance, $\mathbf{C}_d$.
2: Estimate the regularization, $\mathbf{Q}$.
3: Construct the operator $\mathbf{L}$.
4: Select μ_h.
5: Solve,
$\mathbf{m} = (\mathbf{L}^T\mathbf{C}_d^{-1}\mathbf{L} + \mu_h\mathbf{Q})^{-1}\mathbf{L}^T\mathbf{C}_d^{-1}\mathbf{d}.$
6: Output, $\mathbf{m}$.

3.4.4 Iterative re-weighed least squares method

Iterative re-weighted least square(IRLS) algorithm is an algorithm where the hyper-parameter in equation (3.15) is non-zero and the matrix $\mathbf{Q}$ depend on $\mathbf{m}$ i.e a non-linear least square. There are cases in which the weight can be imposed on the linear model while the hyper-parameter is zero (Trad et al., 2003). In other words, if any weight that depend on the model parameters is imposed as weight, it is solved by IRLS algorithm. In this case, the

solution is given by

$$\mathbf{m} = (\mathbf{L}^T \mathbf{C}_d^{-1} \mathbf{L} + \mu \mathbf{Q}(\mathbf{m}))^{-1} \mathbf{L}^T \mathbf{C}_d^{-1} \mathbf{d}. \tag{3.19}$$

This problem can be solved by setting a certain number of iterations to update the weighting term from previous iteration and treating the problem as a weighted least square at each iteration.

Table 3.3: IRLS Algorithm

1: Input data, $\mathbf{d}$.
2: Construct the operator, $\mathbf{L}$.
3: Select μ_h.
4: Initialize the solution, $\mathbf{m}$.
5: Set the maximum iteration to up-date $\mathbf{Q}$ ($\cong$ 3 to 5 iterations).
6: Calculate $\mathbf{Q}(\mathbf{m}^k)$,
7: use the result in step 6 to find,
$\mathbf{m}^{k+1} = (\mathbf{L}^T \mathbf{L} + \mu \mathbf{Q}(\mathbf{m}^k))^{-1} \mathbf{L}^T \mathbf{C}_d^{-1} \mathbf{d}$,
8: repeat step 6 and 7 until the maximum iteration is reached.

In the above three possible least square problems, all involves solving the inverse of a matrix. If the size of the matrices is small to be manipulated easily, one can set up the problem in matrix form and use any matrix inverse solver such as Gauss-Seidel or Gaussian-Elimination or matrix form of the Conjugate Gradient algorithm shown in Table 3.4 (Shewchuk, 1994). If the size of the problem is large, solving the problem by setting the operators in matrix form needs a lot of computer memory. This kind of problems can be solved using iterative inverse solvers such as Conjugate Gradient algorithms which uses forward and adjoint operators in function form to replace the matrix operators.

3.5 Target oriented AVO inversion

In most geophysical problems, the targets could be identified from the acquired data. If this is the case, target oriented inversion is advantageous to reduce computational time of the inversion process and to clearly understand the result at region of the target. If inversion is important through out the available data set, window by window inversion can be done. This is same as breaking up a large problem in to pieces of small problems which can be easily manipulated in matrix form. One advantage of using window by window inversion is to avoid problems resulting from treatment of very large and small model parameters in

Table 3.4: Conjugate Gradient Algorithm

Input:	$iter_{max}$
	ϵ (Tolerance)
	$\mathbf{m}_0 =$ Initial solution
	$\mathbf{b} = \mathbf{L}^T\mathbf{d}$
	$\boldsymbol{\Omega} = \mathbf{L}^T\mathbf{L} + \mu\mathbf{Q}(\mathbf{m}_0),$.
CG:	$\mathbf{r}_0 = \mathbf{b} - \boldsymbol{\Omega}\mathbf{m}_0$
	$\mathbf{m}_i = \mathbf{m}_0$
	$\mathbf{r}_i = \mathbf{r}_0$
	$\mathbf{p}_i = \mathbf{r}_i$
	$\sigma_{new} = \mathbf{r}_i^T\mathbf{r}_i$
	$\sigma_0 = \sigma_{new}$
	$i = 1$
	$while\ (i < iter_{max})\ and\ (\sigma_{new} < \epsilon\sigma_0)$
	$\qquad \mathbf{q}_i = \boldsymbol{\Omega}\mathbf{p}_i$
	$\qquad \alpha_i = \frac{\sigma_{new}}{\mathbf{p}_i^T\mathbf{q}_i}$
	$\qquad \mathbf{m}_{i+1} = \mathbf{m}_i + \alpha_i\mathbf{p}_i$
	$\qquad \mathbf{r}_{i+1} = \mathbf{r}_i - \alpha_i\mathbf{q}_i$
	$\qquad \sigma_{old} = \sigma_{new}$
	$\qquad \sigma_{new} = \mathbf{r}_{i+1}^T\mathbf{r}_{i+1}$
	$\qquad \beta_{i+1} = \frac{\sigma_{new}}{\sigma_{old}}$
	$\qquad \mathbf{p}_{i+1} = \mathbf{m}_i + \beta_{i+1}\mathbf{p}_i$
	$\qquad i = i + 1$
	$end\ while$
Output:	$\mathbf{m} = \mathbf{m}_{i+1}$

the same window. The first step is defining the target on the data or dividing the data set into several windows. As the impulse response of a seismic signal is represented by the convolution of a wavelet with reflectivity, there is mixing of events just above and below each window. The mixing of events can be avoided by setting an overlapping window above and below each window. In order to map the offset information into angle, ray tracing method is used for each window. Then linear operator that relates the observed data and the model parameters is constructed using the convolution model for that given window. Likewise, the regularization term also calculated for each window according to type of regularization used to constraint the inverse problem. Following this, we setup the operators for inversion. The brief algorithm to find the solution in a given target window is given in Table 3.5. If we have more than a window in a given data set, we repeat this algorithm successively for each window.

Table 3.5: Target Oriented AVO inversion Via WLS/IRLS

Entire time window
1: Input data, $\mathbf{d}$ and well-log data, α, β and ρ.
2: Estimate wavelet/s from the data.
3: Select a target window.
4: Set overlap parameters.
5: Define offset information and background velocity.
New window
6: Select the data in the new window (target + overlap).
7: Ray tracing (mapping offset into angle).
8: Construct the operator, $\mathbf{L}$, from steps 2 and 7.
9: Calculate the correlation information (matrix) from the well-log data in the given window.
10: Use steps 7-9 via WLS or IRLS Algorithm to estimate the model parameters.
Target window
11: Take the result of the step 10 in the target window.
12: Output, model parameters in the target window.

3.6   Summary

This chapter was devoted to the Bayesian formulation of inverse problems. The first section started with the famous statistical theorem which is base for the formulation of the Bayesian inversion. In order to implement this theorem, it needs the likelihood function and the prior distribution. The Multivariate Gaussian and Multivariate $\mathbf{t}$ prior distributions to model AVO parameters were discussed. These distributions were described fully by figures by taking some special cases which determine the shapes and nature of the distributions. Along this line, the appropriate choice of prior distribution for model parameters of a given problem was emphasized. In addition, a typical type of inverse problem was discussed in different possible scenarios with their corresponding inversion algorithms. If the prior distribution imposes a regularization which depends on the model parameters, it uses the IRLS algorithms. Otherwise, it uses either the ordinary or weighted least square methods. In the final section, the target oriented AVO inversion via WLS/IRLS has been discussed. In the next two chapters, this inversion algorithm will be used for two-term and three-term AVO inversion.

CHAPTER 4

Two-term AVO Inversion

4.1　Introduction

In chapter two, the various approximations on Zoeppritz equation were discussed by plotting the reflection coefficient versus angle of incidences. The two-term approximations include Verm and Hilterman, Smith and Gidlow, and Fatti. The dimensionality of the model parameters in AVO inversion has influence on the stability of the inversion (Ursenbach and Stewart, 2008; Mallick, 2001). For instance, a two term AVO inversion is more stable than a three term AVO inversion (three-term AVO in the next chapter). The Sheuy's approximation consist of two parameters; the intercept and gradient (the coefficient of the angle dependent term). This approximation is in a very good agreement with Zoeppritz equation for small layer contrast and angles range from 0 to 30^0, but it doesn't hold for large layer contrast and angles. The two terms in Fatti's approximation are the P-wave impedance and S-wave impedance. According to the comparison, this approximation is a better approximation as compared to any of other two term approximations which is relatively less sensitive for larger layer contrasts and angles. Nevertheless, for near offset gathers, depending on target of interpretation, any of the two term approximations are useful tools to extract AVO parameters either by forward modeling or inversion. These AVO parameters are used by oil and gas industry for hydrocarbon identification purposes by cross-plotting the two terms one against another (Castagna and Swan, 1997; Whitcombe et al., 2004). In this chapter, the Fatti's approximation is used as a forward model for a two-term AVO inversion. Bivariate Cauchy distribution is proposed as prior distribution. The performance of this prior is investigated by synthetic data examples in comparison with two other prior distributions.

4.2 Two-term AVO inversion formulation

Fatti's Approximation

For convenience, the Fatti's approximation, equation (2.36), can be rewritten as

$$R_{pp}(\bar{\theta}) = F_{r_\alpha}(\bar{\theta})r_\alpha + F_{r_\beta}(\bar{\theta})r_\beta, \tag{4.1}$$

where

$$
\begin{aligned}
r_\alpha &= \frac{\Delta I_\alpha}{I_\alpha}, \\
r_\beta &= \frac{\Delta I_\beta}{I_\beta},
\end{aligned}
\tag{4.2}
$$

and the coefficients are represented by F_{r_α} and F_{r_α}.

Convolution model

Equation (4.1) represents the earth's impulse response of a seismic signal. In geophysics, seismic signal is modeled by a mathematical tool called convolution. It is used to represent a seismic trace by convolving a source wavelet (source signal) with the impulse response (reflectivity). Parameterizing layers with a constant zero-offset time sample, Δt, and using the convolution model, the relationship between the observed data and reflectivity model at a given offset, X, can be written as

$$\mathbf{d}(X,t) = \mathbf{w}(t) * \mathbf{R}_{pp}^f(X,t) + \mathbf{n}(t), \tag{4.3}$$

For M number of receivers, we have a common depth point (CDP) gathers which can be put in matrix form as

$$
\begin{pmatrix} \mathbf{d}_1 \\ . \\ . \\ . \\ \mathbf{d}_M \end{pmatrix} = \begin{pmatrix} \mathbf{W}_1 & & & \\ & . & & \\ & & . & \\ & & & . \\ & & & \mathbf{W}_M \end{pmatrix} \begin{pmatrix} \mathbf{F}_{r_\alpha 1} & \mathbf{F}_{r_\beta 1} \\ . & . \\ . & . \\ . & . \\ \mathbf{F}_{r_\alpha M} & \mathbf{F}_{r_\beta M} \end{pmatrix} \begin{pmatrix} \mathbf{r}_\alpha \\ \mathbf{r}_\beta \end{pmatrix} + \begin{pmatrix} \mathbf{n}_1 \\ . \\ . \\ . \\ \mathbf{n_M} \end{pmatrix}, \tag{4.4}
$$

where $\mathbf{d}_i$ is a data, $\mathbf{W}_i$ is a wavelet matrix, $\mathbf{F}_{r_\alpha i}$ and $\mathbf{F}_{r_\beta i}$ are block diagonal matrices, $\mathbf{n}_i$ is noise associated with the data $\mathbf{d}_i$ at the ith offset, and $\mathbf{r}_\alpha$ and $\mathbf{r}_\beta$ are the model parameters each of them are vectors with N elements, and $i = 1, 2, 3, ..., M$. It should be noted the angles are obtained by mapping the offset information using the available velocity model. Equation (4.4) can be put in a simple linear form as,

$$\mathbf{d} = \mathbf{W}\mathbf{G}_f\mathbf{m}_f + \mathbf{n}$$

or

$$\mathbf{d} = \mathbf{L}_f\mathbf{m}_f + \mathbf{n}, \tag{4.5}$$

where $\mathbf{L}_f = \mathbf{W}\mathbf{G}_f$ and

$$\mathbf{m}_f = \begin{pmatrix} \mathbf{r}_\alpha \\ \mathbf{r}_\beta \end{pmatrix}. \tag{4.6}$$

4.2.1 Likelihood function of the data

Theoretically, the uncertainty between the observed data and synthetic data (data generated using convolution model) is the noise in the observed data. Assuming the noise terms are independent and Gaussian, the noise can be modeled using Multivariate Gaussian probability distribution given by

$$P(\mathbf{n}|\mathbf{C}_d) = \exp\{-\frac{1}{2}\mathbf{n}^T\mathbf{C}_d^{-1}\mathbf{n}\}, \tag{4.7}$$

where $\mathbf{C_d}$ is the data covariance matrix having the form

$$\mathbf{C}_d = \begin{pmatrix} \sigma_{d1}^2 & 0 & . & . & . & 0 \\ 0 & \sigma_{d2}^2 & & & & 0 \\ . & . & . & & & . \\ . & . & & . & & . \\ . & . & & & . & . \\ 0 & 0 & . & . & . & \sigma_{d_{MN}}^2 \end{pmatrix}. \tag{4.8}$$

Using equation (4.5) and (4.7), the likelihood function of the data can be expressed as

$$P(\mathbf{d}|\mathbf{m}_f) = P_o\exp\{-\frac{1}{2}(\boldsymbol{\Upsilon}(\mathbf{d} - \mathbf{L}_f\mathbf{m}_f))^T\mathbf{C_d}^{-1}(\boldsymbol{\Upsilon}(\mathbf{d} - \mathbf{L_f}\mathbf{m_f}))\}, \tag{4.9}$$

where,

$$P_o = \frac{1}{(2\pi)^{(NM)/2}|\mathbf{C_d}|^{1/2}}. \tag{4.10}$$

A diagonal matrix, $\boldsymbol{\Upsilon}$, is also introduced for muting i.e to protect the algorithm from trying to fit with the input data with zero entries.

4.2.2 Multivariate Gaussian prior

In this section, Gaussian probability distribution is used *a priori* for the model parameters. It is given by

$$P(\mathbf{m}_f) = \frac{1}{\pi^N |\mathbf{C}_f|^{1/2}} \exp\{-\frac{1}{2}\mathbf{m}_f^T \mathbf{C}_f^{-1} \mathbf{m}_f\}, \tag{4.11}$$

where $\mathbf{C}_f$ is parameter covariance matrix which shows the correlation between the two model parameters; P-wave impedance and S-wave impedance. Assuming the model parameters are correlated at a given time sample but independent from one time sample to another, the covariance matrix for $2N$ model parameters can be expressed as

$$\mathbf{C}f = \begin{pmatrix} \sigma_{r_\alpha}^2 & & & \sigma_{r_\alpha r_\beta} & \\ & \ddots & & & \ddots \\ & & \sigma_{r_\alpha}^2 & & & \sigma_{r_\alpha r_\beta} \\ \sigma_{r_\beta r_\alpha} & & & \sigma_{r_\beta}^2 & & \\ & \ddots & & & \ddots & \\ & & \sigma_{r_\beta r_\alpha} & & & \sigma_{r_\beta}^2 \end{pmatrix} \tag{4.12}$$

which is $(2N) \times (2N)$ symmetric matrix. In this matrix, the diagonal elements are the variances of each of the model parameter elements and the off-diagonal matrices shows the degree of correlation of various possible combination between a pair of model elements. In real data inversion, this matrix can be constructed from available bore-hole data (Downton and Lines, 2004).

Objective function

The objective function of the problem can be constructed by combining the likelihood function and the prior model using Bayes' theorem. Thus, using equations (3.2), (4.9) and (4.11), we have

$$P(\mathbf{m}_f|\mathbf{d}) \propto \exp\{-\frac{1}{2}(\mathbf{d} - \mathbf{L}_f \mathbf{m}_f)^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_f \mathbf{m}_f) - \frac{1}{2}\mathbf{m}_f^T \mathbf{C}_f^{-1} \mathbf{m}_f)\}. \tag{4.13}$$

Maximizing the posterior distribution, equation (4.13), is equivalent to minimizing the following objective function

$$J^{mg}(\mathbf{m}_f) = \frac{1}{2}(\mathbf{d} - \mathbf{L}_f \mathbf{m}_f)^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_f \mathbf{m}_f) + \frac{1}{2}\mathbf{m}_f^T \mathbf{C}_f^{-1} \mathbf{m}_f)\}. \tag{4.14}$$

Differentiating (4.14) with respect to $\mathbf{m}_f$ and setting the resulting expression to zero, we

get

$$(\mathbf{L}_f^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} \mathbf{L}_f + \mathbf{C}_f^{-1}) \mathbf{m}_f = \mathbf{L}_f^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} \mathbf{d}. \tag{4.15}$$

For practical purpose we assume the same noise variance for each of the data elements. This assumption reduces the data covariance matrix to

$$\mathbf{C}_d = \sigma_d^2 \mathbf{I}, \tag{4.16}$$

where $\mathbf{I}$ is an identity matrix which has same size as $\mathbf{C}_d$. Substituting equation (4.16) into equation (4.15), we get

$$(\mathbf{L}_f^T \mathbf{\Upsilon}^T \mathbf{\Upsilon} \mathbf{L}_f + \mu^{mg} \mathbf{C}_f^{-1}) \mathbf{m}_f = \mathbf{L}_f^T \mathbf{\Upsilon}^T \mathbf{\Upsilon} \mathbf{d}, \tag{4.17}$$

where

$$\mu^{mg} \sim \sigma_d^2. \tag{4.18}$$

Theoretically, the parameter μ^{mg} is in the order of the noise variance but can be estimated depending on the weight we need to impose to the data and the prior information. Equation (4.17) is the final least square solution for two-term AVO using Gaussian probability distribution to model both the noise, and the model parameters.

4.2.3 Univariate Cauchy prior

Univariate Cauchy probability distribution belongs to the Multivariate $\mathbf{t}$ distributions for particular choice of parameters, $\nu = 1$ and $p = 1$. This kind of prior has become a powerful tool in high resolution geophysical analysis: Radon transform and AVO inversion. Modeling the parameters via Univariate Cauchy probability distribution means that the model parameters are treated as uncorrelated. Therefore, assuming the center (location parameter) to be zero for the two term models, the prior has the form

$$P(\mathbf{m}_f) = \frac{1}{(\pi\sigma)^{2N}} \exp(-\sum_{k=1}^{2N} \ln(1 + (\frac{m_f^k}{\sigma})^2)), \tag{4.19}$$

where σ is a scale parameter which is the dispersion of the models from center of the distribution.

Objective function

Combining equations (4.9) and (4.19) using Bayes' theorem, the posterior distribution for

this particular prior becomes

$$P(\mathbf{m}_f|\mathbf{d}) \propto \exp\{-\frac{1}{2}(\mathbf{d} - \mathbf{L}_f\mathbf{m}_f)^T\mathbf{\Upsilon}^T\mathbf{C}_d^{-1}\mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_f\mathbf{m}_f) - \sum_{i=1}^{2N}\ln(1 + (\frac{m_f^i}{\sigma})^2)\}. \quad (4.20)$$

From which follows, the objective function

$$J^{uc}(\mathbf{m}_f) = \frac{1}{2}(\mathbf{d} - \mathbf{L}_f\mathbf{m}_f)^T\mathbf{\Upsilon}^T\mathbf{C}_d^{-1}\mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_f\mathbf{m}_f) + R^{uc}(\mathbf{m}_f)\} \quad (4.21)$$

where

$$R^{uc}(\mathbf{m}_f) = \sum_{i=1}^{2N}\ln(1 + (\frac{m_f^i}{\sigma})^2) \quad (4.22)$$

which is the regularization that comes from the Univariate Cauchy prior. The next step is to minimize the objective function. Differentiating J^{uc} with respect to $\mathbf{m}_f$, we get

$$\frac{\partial J^{uc}(\mathbf{m}_f)}{\partial \mathbf{m}_f} = \mathbf{L}_f^T\mathbf{\Upsilon}^T\mathbf{C}_d^{-1}\mathbf{\Upsilon}\mathbf{L}_f\mathbf{m}_f - \mathbf{L}_f^T\mathbf{\Upsilon}^T\mathbf{C}_d^{-1}\mathbf{\Upsilon}\mathbf{d} + \frac{\partial R^{uc}(\mathbf{m}_f)}{\partial \mathbf{m}_f}, \quad (4.23)$$

where

$$\frac{\partial R^{uc}(\mathbf{m}_f)}{\partial \mathbf{m}_f} = \frac{2}{\sigma^2}\mathbf{Q}_f^{uc}\mathbf{m}_f \quad (4.24)$$

and the matrix $\mathbf{Q}_f^{uc}$ is a $(2N) \times (2N)$ diagonal matrix whose elements are

$$[(\mathbf{Q}_f^{uc})_{kk}] = \left(1 + \left(\frac{m_f^k}{\sigma}\right)^2\right)^{-1}. \quad (4.25)$$

The detail derivation for generalized Multivariate $\mathbf{t}$ distribution and the regularization is given in Appendix (A.2). Substituting equations (4.16) and (4.24) into equation (4.23), and setting the resulting expression to zero, we finally have

$$(\mathbf{L}_f^T\mathbf{\Upsilon}^T\mathbf{\Upsilon}\mathbf{L}_f + \mu^{uc}\mathbf{Q}_f^{uc})\mathbf{m}_f = \mathbf{L}_f^T\mathbf{\Upsilon}^T\mathbf{\Upsilon}\mathbf{L}_f\mathbf{d}, \quad (4.26)$$

where

$$\mu^{uc} \sim \frac{2\sigma_d^2}{\sigma^2}$$

which means the hyper-parameter μ^{uc} is in the order of ratio of the noise variance and the square of the scale parameter. The dependence of $\mathbf{Q}_f^{uc}$ on $\mathbf{m}_f$ makes the inverse problem non-linear unlike using the Gaussian probability distribution as *a prior*. The IRLS algorithm

is used to solve this problem as outlined in the previous chapter.

4.2.4   Bivariate Cauchy prior

Bivariate Cauchy probability distribution is also a family of Multivariate $\mathbf{t}$ distribution for $\nu = 1$ and $p = 2$. This kind of prior has never been used to constraint AVO inversion. In this thesis, this prior is proposed to constraint a two term AVO inversion for two basic reasons. First, the probability distribution has a long tailed nature to sparsity. Secondly, it allows to add some geological information (well-log information) via a 2×2 scale matrix (correlation information matrix). Assuming the model parameters are correlated at a given time sample but independent from one to another time sample, the joint Bivariate Cauchy probability distribution can be written as

$$P(\mathbf{m}_f) = P_{m0} \exp[-\frac{3}{2} \sum_{i=1}^{N} \ln(1 + \mathbf{m}_f^T (\mathbf{\Phi}_f^{bc})^i \mathbf{m}_f)], \tag{4.27}$$

where

$$(\mathbf{\Phi}_f^{bc})^i = (\mathbf{D}^i)^T \mathbf{\Psi}^{bc} \mathbf{D}^i, \tag{4.28}$$

$$\mathbf{\Psi}^{bc} = \begin{pmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{pmatrix}, \tag{4.29}$$

and $\mathbf{D}^i$ is a $3 \times 3N$ matrix all zero except the first row at the ith column, and the second row at the $(i + N)^{th}$ column i.e

$$[\mathbf{D}_{nl}^i] = \begin{cases} 1, & if & n = 1 & and & l = i \\ 1, & if & n = 2 & and & l = i + N, \\ 0, & otherwise, & & i = 1, 2, 3, .., N. \end{cases} \tag{4.30}$$

This matrix is incorporated for convenience to write the probability distribution in terms of all the $2N$ model parameters in column form i.e it selects two elements from $\mathbf{r}_\alpha$ and $\mathbf{r}_\beta$ at a given time sample i. The normalization constant, P_{m0}, is as defined in Appendix (A.2) and the objective function is independent this constant.

Objective function

In a similar way, combining equations (4.9) and (4.27), the objective function takes the form

$$J^{bc}(\mathbf{m}_f) = \frac{1}{2}(\mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_f \mathbf{m}_f))^T \mathbf{C}_d^{-1} \mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_f \mathbf{m}_f)) + R^{bc}(\mathbf{m}_f), \tag{4.31}$$

where R^{bc} is the regularization term given by

$$R^{bc}(\mathbf{m}_f) = \frac{3}{2} \sum_{i=1}^{N} \ln(1 + \mathbf{m}_f^T (\boldsymbol{\Phi}_f^{bc})^i \mathbf{m}_f).$$ (4.32)

The solution of the inverse problem is a model vector $\mathbf{m}_f$ that minimizes the objective function J^{bc}. Thus, differentiating equation (4.31) with respect to $\mathbf{m}$ and setting the resulting expression to zero, we have

$$(\mathbf{L}^T \boldsymbol{\Upsilon}^T \mathbf{C}_d^{-1} \boldsymbol{\Upsilon} \mathbf{L} + \frac{3}{2} \mathbf{Q}_f^{bc}) \mathbf{m}_f = \mathbf{L}_f^T \boldsymbol{\Upsilon}^T \mathbf{C}_d^{-1} \boldsymbol{\Upsilon} \mathbf{d}.$$ (4.33)

where $\mathbf{Q}_f^{bc}$ is a $(2N) \times (2N)$ matrix whose elements are defined as

$$[(\mathbf{Q}_f^{bc})_{kn}] = \sum_{i=1}^{N} \frac{2(\boldsymbol{\Phi}_f^{bc})_{kn}^i}{1 + \mathbf{m}_f^T (\boldsymbol{\Phi}_f^{bc})^i \mathbf{m}_f} \qquad k, n = 1, 2, 3, ..., 2N.$$ (4.34)

The derivation of the regularization is given in Appendix (A.2). By similar assumption we made about the noise covariance matrix in equation (4.16), equation (4.33) becomes

$$(\mathbf{L}_f^T \boldsymbol{\Upsilon}^T \boldsymbol{\Upsilon} \mathbf{L}_f + \mu^{bc} \mathbf{Q}_f^{bc}) \mathbf{m}_f = \mathbf{L}_f^T \boldsymbol{\Upsilon}^T \boldsymbol{\Upsilon} \mathbf{d}.$$ (4.35)

where

$$\mu^{bc} \sim \frac{3}{2} \sigma_d^2$$ (4.36)

which means the hyper-parameter μ^{bc} is in the order of the variance of the noise terms (square of the standard deviation of the noise).

4.3 Synthetic data examples

To assess the performance of the proposed regularization methods, synthetic data are generated by convolving the reflectivity series constructed from the P-wave, S-wave and density reflectivities derived from the true P-wave velocity, S-wave velocity and density (Figure 4.1) with a 35 Hz ricker wavelet. The generated synthetic data in the entire time window is shown in Figure 4.2a without noise Figure 4.2b with noise S/N=4. The data consist of 31 traces with offset ranges from 10 m to 2410 m and angle of incidence designed in the algorithm to have not more than 45^0. The sample covariance matrix for Gaussian and the scale matrix for Bivariate Cauchy (using the EM algorithm given in Appendix (B.1) are constructed by smoothing the true reflectivities of the model parameters. The inversion is

run in a arbitrarily selected time window 0.2 s to 0.636 s with an overlap of a length of the wavelet above and below the selected time window (the target oriented inversion algorithm given at the end of chapter 3). Figures 4.3, 4.5, 4.7, 4.9, 4.11, and 4.13 are results of the inversion by regularizing the inverse problem using Multivariate Gaussian, Univariate Cauchy, and Bivariate Cauchy probability distributions. Three of them are results from noiseless (S/N=10^4) and three of them are from noisy data (S/N=4). The true (blue) and the inverted (red) are plotted in each case one on the top of the other on the same panel to see how much the true and the inverted result matches. The root mean square errors are given in Table 4.1. Figures 4.4, 4.6, 4.8, 4.10, 4.12, and 4.14 are the corresponding input data, predicted data and the difference between the input and the predicted data which show the degree of data fitting.

	Univariate Cauchy		Multivariate Gaussian		Trivariate Cauchy	
S/N	$RMSE_{r_\alpha}$	$RMSE_{r_\beta}$	$RMSE_{r_\alpha}$	$RMSE_{r_\beta}$	$RMSE_{r_\alpha}$	$RMSE_{r_\beta}$
10^4	0.0039	0.0109	0.0311	0.0472	0.0042	0.0068
4	0.0244	0.0331	0.0354	0.0536	0.0070	0.0108

Table 4.1: The calculated root mean square errors; $RMSE_{r_\alpha}$ and $RMSE_{r_\beta}$ for P and S-wave impedances respectively. See Appendix C for RMSE calculation.

All the conditions in the data are same except the noise level. The inverted results from data with very small noise (the ideal condition) meet expectations. From Table 4.1, we can see that the Univariate Cauchy and Bivariate Cauchy have smaller root mean square error than the Multivariate Gaussian prior. This is because they have a power of rejecting the small fluctuations. Failure in data fitting at large offset (angles) are observed irrespective of the regularization methods. This is also expected as the two-term approximation deviates from the exact model at larger angles (offset). In the noisy case (S/N=4), the inverted results using the Multivariate Gaussian and the Univariate Cauchy priors show greater root mean square errors as compared to Bivariate Cauchy. This implies that unless we include both the correlation and sparsity at the same time, it is very difficult to get accurate result from noisy data. These examples demonstrate clearly the influence of the noise in the inversion result. The effectiveness of the Bivariate Cauchy can be viewed in two ways: the possibility of adding correlation information and sparsity at the same time.
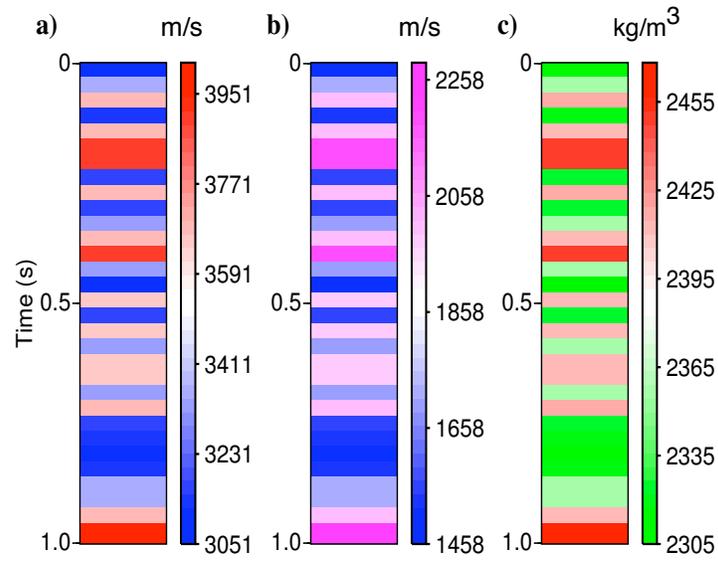
Figure 4.1: (**a**) P-wave velocity (**b**) S-wave velocity and (**c**) density. These models are used for the synthetic example.
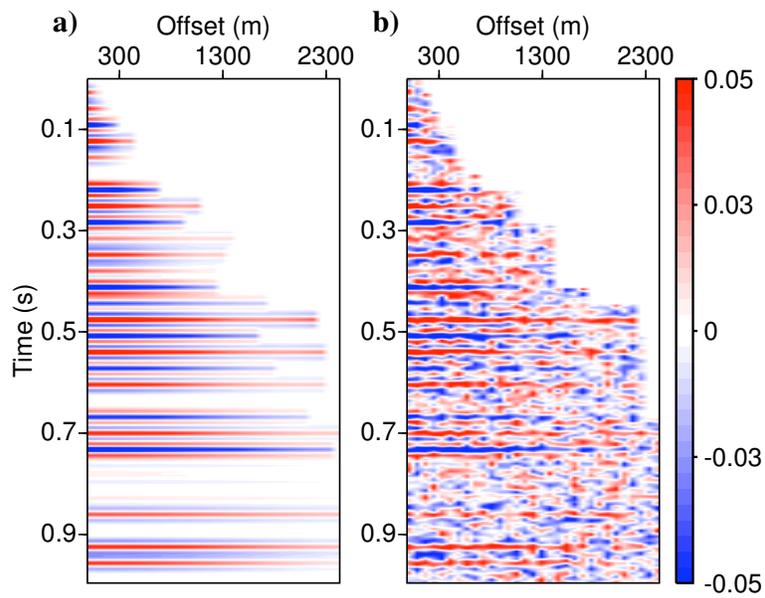
Figure 4.2: Synthetic data (**a**) noiseless, and (**b**) with noise, S/N=4.
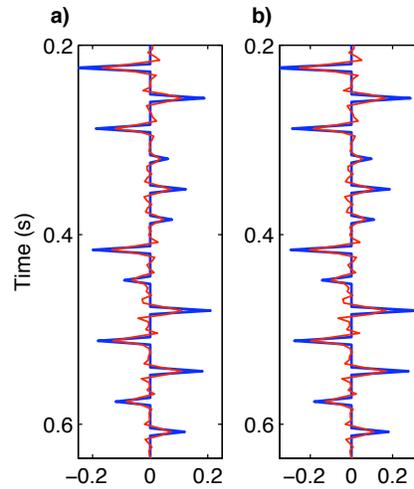
Figure 4.3: (**a**) P-wave impedance ($RMSE_{r_\alpha} = 0.0311$), and (**b**) S-wave impedance ($RMSE_{r_\beta} = 0.0472$) retrieved from noiseless data (S/N=10^4) using Multivariate Gaussian as *a priori*.
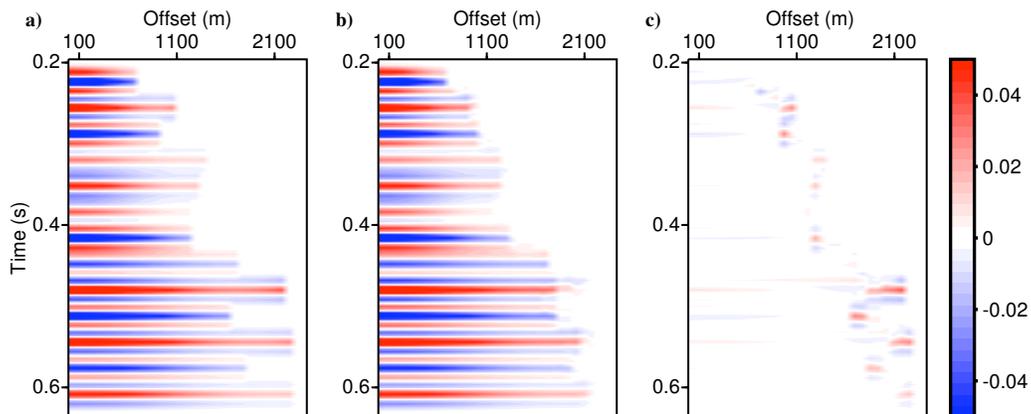
Figure 4.4: (**a**) The input data (S/N=10^4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Multivariate Gaussian as *a priori*.
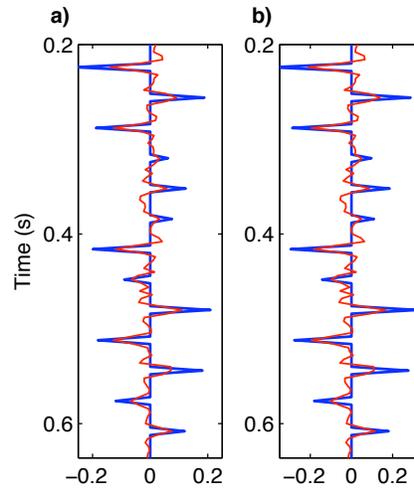
Figure 4.5: (**a**) P-wave impedance ($RMSE_{r_\alpha} = 0.0354$), and (**b**) S-wave impedance ($RMSE_{r_\beta} = 0.0536$) retrieved from data with S/N=4 using Multivariate Gaussian as *a priori*.
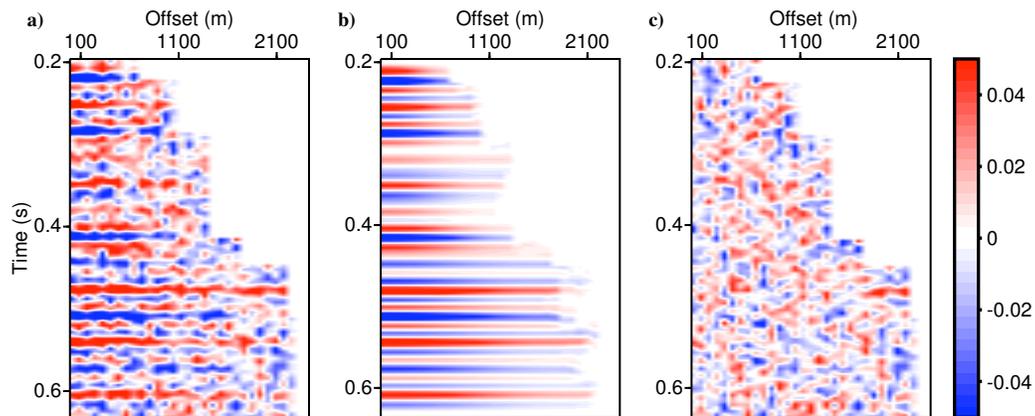
Figure 4.6: (**a**) The input data (S/N=4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Multivariate Gaussian as *a priori*.
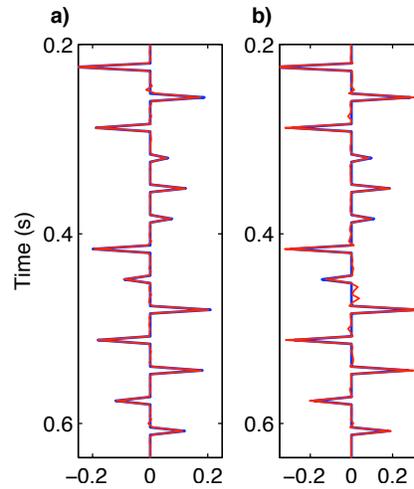
Figure 4.7: (**a**) P-wave impedance ($RMSE_{r_\alpha} = 0.0039$), and (**b**) S-wave impedance ($RMSE_{r_\beta} = 0.0109$) retrieved from noiseless data (S/N=10^4) using Univariate Cauchy as *a priori*.
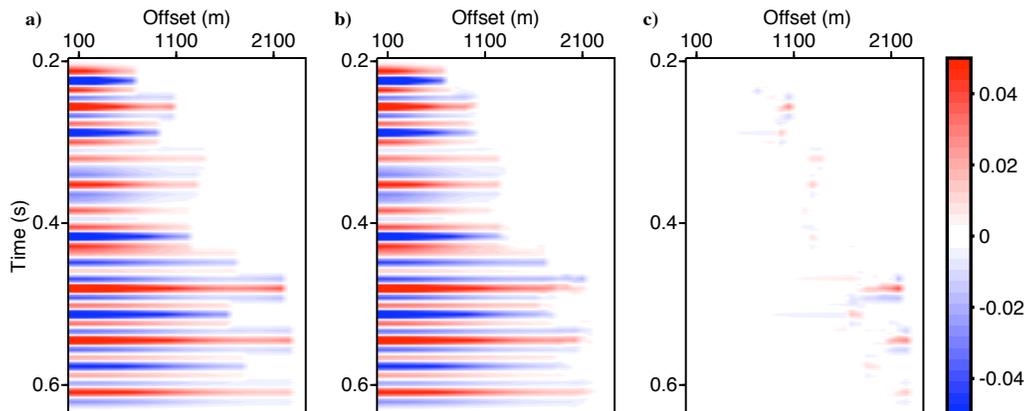
Figure 4.8: (**a**) The input data (S/N=10^4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Univariate Cauchy as *a priori*.
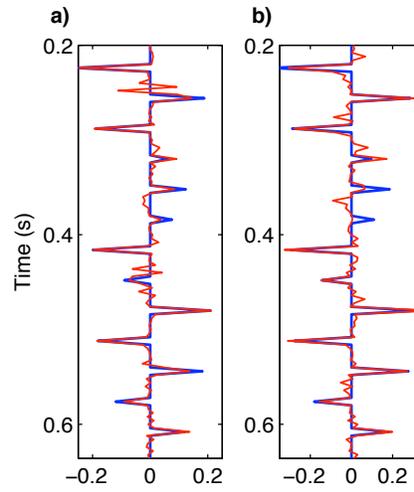
Figure 4.9: (**a**) P-wave impedance ($RMSE_{r_\alpha} = 0.0244$), and (**b**) S-wave impedance ($RMSE_{r_\beta} = 0.0331$) retrieved from data with S/N=4 using the Univariate Cauchy as *a priori*.
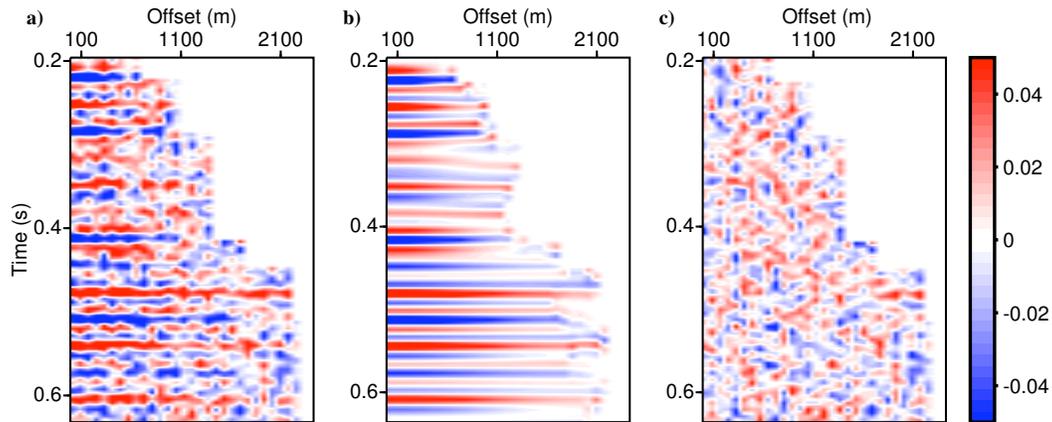
Figure 4.10: (**a**) The input data (S/N=4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Univariate Cauchy as *a priori*.
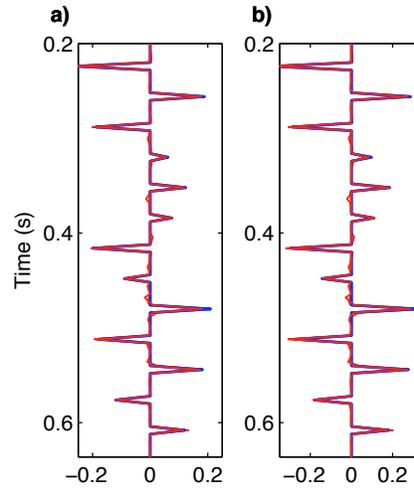
Figure 4.11: (**a**) P-wave impedance ($RMSE_{r_\alpha} = 0.0042$), and (**b**) S-wave impedance ($RMSE_{r_\beta} = 0.0068$) retrieved from noiseless data (S/N=10^4) using Bivariate Cauchy as *a priori*.
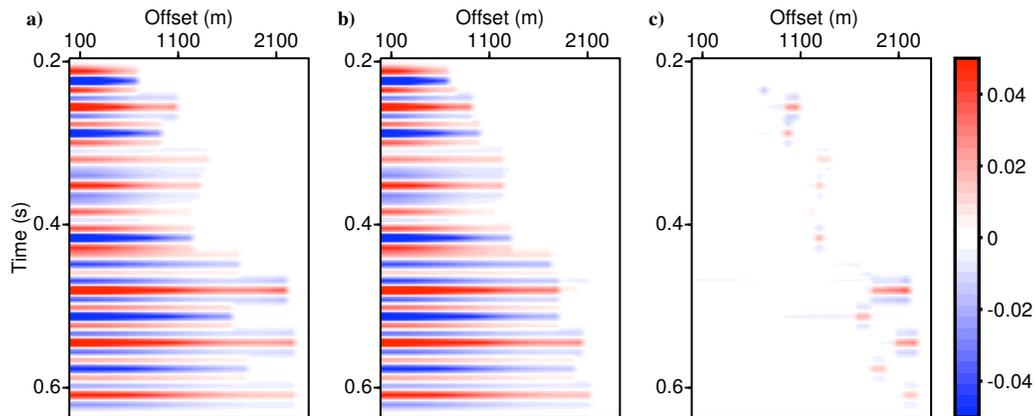
Figure 4.12: (**a**) The input data (S/N=10^4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Bivariate Cauchy as *a prior*.
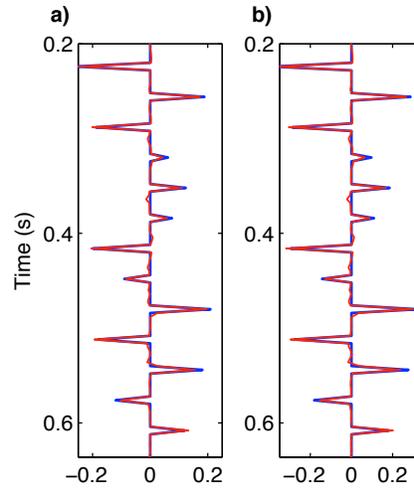
Figure 4.13: (**a**) P-wave impedance ($RMSE_{r_\alpha} = 0.007$), and (**b**) S-wave impedance ($RMSE_{r_\alpha} = 0.0108$) retrieved from data with S/N=4 using Bivariate Cauchy as *a priori*.
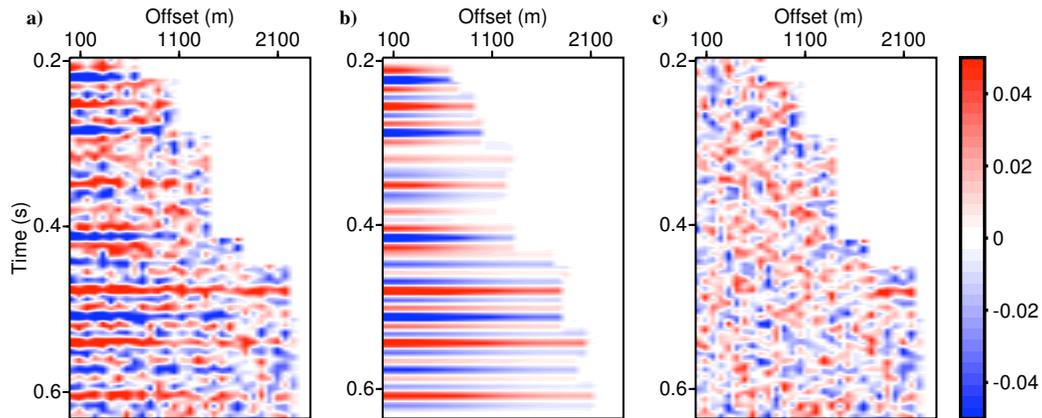
Figure 4.14: (**a**) The input data (S/N=4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Bivariate Cauchy as *a priori*.

CHAPTER 5

Three-term AVO Inversion

5.1 Introduction

Three-term AVO inversion uses three-term approximations on Zoeppritz equation as a forward model. It is the most favored inversion method for two main basic reasons. First, the three-term approximations like Aki and Richards's are considered as the best approximation to the exact reflection coefficient (R_{pp}). In addition, the three term approximation allows to retrieve one more model parameter i.e direct density information which is useful parameter to characterize physical property of rocks (subsurface geology). Including the third term provides information which can be used to separate gas sands from wet sands, tight sands, and shales or to study gas saturation (Kabir et al., 2006; Sun et al., 2005; Li, 2005). However, as pointed out in the introduction part of chapter 4, using three-term approximations as inversion model is more ill-conditioned than a two-term approximation. This suggests that more attention should be given to stabilizing the inversion although both cases need regularization in order to make the solution more stable. Following a similar procedure given in the previous chapter, the three term AVO regularizations are constructed via Bayesian approach. This chapter is mainly to address a new way of regularizing the three-term AVO inversion using Trivariate Cauchy probability distribution in comparison with other two possible regularization methods. This is demonstrated by both synthetic and real data examples.

5.2   Three-term inversion formulation

Aki and Richards's Approximation

The Aki and Richards's approximation is expressed in terms of the three model parameters the P-wave reflectivity, S-wave reflectivity and density reflectivity. These approximation allows to retrieve the velocity and density information independently and then the parameters can be rearranged and interpreted into other attributes like impedance or fluid factor. For convenience, equation (2.27) can be rewritten as

$$R_{PP}(\bar{\theta}) = F_A(\bar{\theta})A + F_B(\bar{\theta})B + F_C(\bar{\theta})C, \tag{5.1}$$

where F_A, F_B, and F_C are corresponding coefficients of the three model parameters. Convolving the reflection coefficient with a source wavelet for N number of layers parametrized with a constant zero-offset travel time, dt, the relationship between observed data and the reflection coefficient is expressed as

$$\mathbf{d}(X, t) = \mathbf{w}(t) * \mathbf{R}_{pp}^a(X, t) + \mathbf{n}(t), \tag{5.2}$$

where $\mathbf{d}_i$ is the observed data, $\mathbf{w}$ is the wavelet. For M number of receivers, the AVO gathers can be put in matrix form as

$$
\begin{pmatrix} \mathbf{d}_1 \\ . \\ . \\ . \\ \mathbf{d}_M \end{pmatrix}
=
\begin{pmatrix} \mathbf{W}_1 & & & \\ & . & & \\ & & . & \\ & & & . \\ & & & \mathbf{W}_M \end{pmatrix}
\begin{pmatrix} \mathbf{F}_{A1} & \mathbf{F}_{B1} & \mathbf{F}_{C1} \\ . & . & . \\ . & . & . \\ . & . & . \\ \mathbf{F}_{AM} & \mathbf{F}_{BM} & \mathbf{F}_{CM} \end{pmatrix}
\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \end{pmatrix}
+
\begin{pmatrix} \mathbf{n}_1 \\ . \\ . \\ . \\ \mathbf{n_M} \end{pmatrix}, \tag{5.3}
$$

where $\mathbf{d}_i$ is a data, $\mathbf{W}_i$ is a wavelet matrix, $\mathbf{F}_{Ai}$, $\mathbf{F}_{Bi}$ and $\mathbf{F}_{Ci}$ are block diagonal matrices, $\mathbf{n}_i$ is noise associated with the data $\mathbf{d}_i$ at the ith offset, and $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are the model parameters each of them are vectors with N elements, and $i = 1, 2, 3, ..., M$ stands for offset index. The above matrix equation has the same from as in the two-term case except that the reflection coefficient has now three-term model parameters $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ and therefore, changes the size of the $\mathbf{L}_a$ matrix. Equation (5.3) can be put in a simple linear form as,

$$\mathbf{d} = \mathbf{W}\mathbf{G}_a\mathbf{m}_a + \mathbf{n}$$

or

$$\mathbf{d} = \mathbf{L}_a\mathbf{m}_a + \mathbf{n}, \tag{5.4}$$

where $\mathbf{L}_a = \mathbf{WG}_a$ and

$$\mathbf{m}_a = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \\ \mathbf{C} \end{pmatrix}. \tag{5.5}$$

Having setup the physics model that relates the the observed data and the model parameters, the next step is to construct the inverse problem with *a priori* via Bayes' theorem.

5.2.1 Likelihood function of the data

The likelihood function of the data is modeled by assuming the noise is Gaussian. It takes a similar functional form as equation (4.9). For this particular problem, we have

$$P(\mathbf{d}|\mathbf{m}) = P_o \exp\{-\frac{1}{2}(\mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_a\mathbf{m}_a))^T \mathbf{C_d}^{-1}(\mathbf{\Upsilon}(\mathbf{d} - \mathbf{L_a m_a}))\}, \tag{5.6}$$

where $\mathbf{C}_d$ and P_o are already defined in the previous chapter. A diagonal matrix $\mathbf{\Upsilon}$ is a muting operator.

5.2.2 Multivariate Gaussian prior

The Multivariate Gaussian prior distribution for the three model parameters, equation (5.5), is given by

$$P(\mathbf{m}_a) = \frac{1}{(2\pi)^{(3N)/2}|\mathbf{C_a}|^{1/2}} \exp\{-\frac{1}{2}\mathbf{m}_a^T \mathbf{C}_a^{-1}\mathbf{m}_a\}, \tag{5.7}$$

where $\mathbf{C}_a$ is parameter covariance matrix which shows the correlation among the three model parameters. Assuming the model parameters are correlated at a given time sample but independent from one time sample to another for $3N$ model parameters, the covariance

matrix has the form

$$
\mathbf{C}_a = \begin{pmatrix}
\sigma_A^2 & & & \sigma_{AB} & & & \sigma_{AC} & & \\
& \ddots & & & \ddots & & & \ddots & \\
& & \sigma_A^2 & & & \sigma_{AB} & & & \sigma_{AC} \\
\sigma_{BA} & & & \sigma_B^2 & & & \sigma_{BC} & & \\
& \ddots & & & \ddots & & & \ddots & \\
& & \sigma_{BA} & & & \sigma_B^2 & & & \sigma_{BC} \\
\sigma_{CA} & & & \sigma_{CB} & & & \sigma_C^2 & & \\
& \ddots & & & \ddots & & & \ddots & \\
& & \sigma_{CA} & & & \sigma_{CB} & & & \sigma_C^2
\end{pmatrix}
\tag{5.8}
$$

which is $(3N) \times (3N)$ symmetric matrix. In this matrix, the diagonal elements are the variances of each of the model parameter and the off-diagonal matrices show the degree of correlation of various possible combination between a pair of model elements.

Objective function

Combining equations (5.6) and (5.7), the posterior distribution for the three-term inversion via Gaussian prior distribution can be shown as

$$
P(\mathbf{m}_a|\mathbf{d}) \propto \exp\{-\frac{1}{2}(\mathbf{d} - \mathbf{L}_a\mathbf{m}_a)^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} (\mathbf{d} - \mathbf{L}_a\mathbf{m}_a) - \frac{1}{2}\mathbf{m}_a^T \mathbf{C}_a^{-1} \mathbf{m}_a)\}.
\tag{5.9}
$$

Maximizing the posterior distribution, equation (5.9), is equivalent to minimizing the following objective function

$$
J^{mg}(\mathbf{m}_a) = \frac{1}{2}(\mathbf{d} - \mathbf{L}_a\mathbf{m}_a)^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} (\mathbf{d} - \mathbf{L}_a\mathbf{m}_a) + \frac{1}{2}\mathbf{m}_a^T \mathbf{C}_a^{-1} \mathbf{m}_a)\}.
\tag{5.10}
$$

Differentiating (5.10) with respect to $\mathbf{m}_a$ and setting the resulting expression to zero, we get

$$
(\mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} \mathbf{L}_a + \mathbf{C}_a^{-1})\mathbf{m}_a = \mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} \mathbf{d}.
\tag{5.11}
$$

Using the same assumption we made in the previous chapter, the data variance is the same as equation (4.16). Employing this equation in equation (5.11), we have

$$
(\mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{\Upsilon} \mathbf{L}_a + \mu^{mg} \mathbf{C}_a^{-1})\mathbf{m}_a = \mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{\Upsilon} \mathbf{d},
\tag{5.12}
$$

where

$$
\mu^{mg} \sim \sigma_d^2.
\tag{5.13}
$$

The right parameter μ^{mg} is the one that honors both the data and prior information. It should be chosen either using chi-square test (data misfit versus μ^{mg}) or using trade-off curve (data misfit versus the model norm). Equation (5.12) is the final weighted least square solution for three-term AVO via Gaussian prior.

5.2.3   Univariate Cauchy prior

Following a similar procedure outlined in a two-term case in Chapter 4, the Univariate Cauchy prior for three-term inversion has the form

$$P(\mathbf{m}_a) = \frac{1}{(\pi\sigma)^{3N}} \exp(-\sum_{k=1}^{3N} \ln(1 + (\frac{m_a^k}{\sigma})^2), \tag{5.14}$$

where σ is a scale parameter which tells us the dispersion of the models from its center. In this case, there is no correlation between the two model parameters namely P-wave impedance and S-wave impedance. This kind of treatment for correlated parameters is not recommended. But if there is no correlation information to constrain the inversion, it can be used as prior especially for noise attenuation and impose sparsity on the estimated model parameters. This prior performed very well with little discrepancy in a two-term case (Chapter 4).

Objective function

Combining equations (5.6) and (5.14) using Bayes' theorem, the posterior distribution for this particular prior becomes

$$P(\mathbf{m}_a|\mathbf{d}) \propto \exp\{-\frac{1}{2}(\mathbf{d} - \mathbf{L}_a\mathbf{m}_a)^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_a\mathbf{m}_a) - \sum_{i=1}^{3N} \ln(1 + (\frac{m_a^i}{\sigma})^2\}. \tag{5.15}$$

From which follows, the objective function

$$J^{uc}(\mathbf{m}_a) = \frac{1}{2}(\mathbf{d} - \mathbf{L}_a\mathbf{m}_a)^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_a\mathbf{m}_a) + R^{uc}(\mathbf{m}_a)\}, \tag{5.16}$$

where

$$R^{uc}(\mathbf{m}_a) = \sum_{k=1}^{3N} \ln(1 + (\frac{m_a^k}{\sigma})^2) \tag{5.17}$$

which is the regularization that comes from the Univariate Cauchy prior for $3N$ model parameters. The next step is to minimize the objective function. Differentiating J^{uc} with

respect to $\mathbf{m}_a$, it takes the form

$$\frac{\partial J^{uc}(\mathbf{m}_a)}{\partial \mathbf{m}_a} = \mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} \mathbf{L}_a \mathbf{m}_a - \mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} \mathbf{d} + \frac{\partial R^{uc}(\mathbf{m}_a)}{\partial \mathbf{m}_a}, \qquad (5.18)$$

where

$$\frac{\partial R^{uc}(\mathbf{m}_a)}{\partial \mathbf{m}_a} = \frac{2}{\sigma^2} \mathbf{Q}_a^{uc} \mathbf{m}_a. \qquad (5.19)$$

The term $\mathbf{Q}_a^{uc}$ is a $(3N) \times (3N)$ diagonal matrix whose elements are

$$[(\mathbf{Q}_a^{uc})_{kk}] = \left(1 + \left(\frac{m_a^k}{\sigma} \right)^2 \right)^{-1}. \qquad (5.20)$$

This regularization has the same form as in the two-term except the dimension of the matrices. See Appendix A for the derivation of the regularization term for Multivariate $\mathbf{t}$ distribution. Substituting equation (5.19) into equation (5.18) and setting the resulting expression to zero, we finally have

$$(\mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{\Upsilon} \mathbf{L}_a + \mu^{uc} \mathbf{Q}_a^{uc}) \mathbf{m}_a = \mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{\Upsilon} \mathbf{d}, \qquad (5.21)$$

where

$$\mu^{uc} \sim \frac{2\sigma_d^2}{\sigma^2} \qquad (5.22)$$

This is a non-linear inversion due to the dependence of $\mathbf{Q}_a^{uc}$ on $\mathbf{m}_a$ which should be solved using IRLS algorithm.

5.2.4 Trivariate Cauchy prior

The Multivariate $\mathbf{t}$ distribution can also reduce to the Trivariate Cauchy probability distribution for $\nu = 1$ and $p = 3$. This probability distribution can be used as prior distribution for three term AVO inversion which leads to a non-linear regularization term. In this thesis, this regularization is proposed to constraint three-term AVO inversion for similar reasons given for Bivariate Cauchy prior in the previous chapter. First, the probability distribution has a long tail to play a role of sparsity and thereby increase resolution. Secondly, it allows to add some geological information (well-log information) via 3×3 scale matrix (correlation information matrix) to stabilize the solution. Assuming the model parameters ($\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$) are correlated at a given time sample but independent from one to another time sample,

the joint Trivariate Cauchy probability distribution can be written as

$$P(\mathbf{m}_a) = P_{m0} \exp[-2 \sum_{i=1}^{N} \ln(1 + \mathbf{m}_a^T (\mathbf{\Phi}_a^{tc})^i \mathbf{m}_a)], \tag{5.23}$$

where

$$(\mathbf{\Phi}_a^{tc})^i = (\mathbf{D}^i)^T (\mathbf{\Psi}_a^{tc})^{-1} \mathbf{D}^i, \tag{5.24}$$

$$\mathbf{\Psi}_a^{tc} = \begin{pmatrix} \psi_{11} & \psi_{12} & \psi_{13} \\ \psi_{21} & \psi_{22} & \psi_{23} \\ \psi_{31} & \psi_{32} & \psi_{33} \end{pmatrix}, \tag{5.25}$$

and $\mathbf{D}^i$ is a $3 \times 3N$ matrix all zero except the first row at the ith column, the second row at the $(i + N)^{th}$ column, and the third row at $(i + 2N)^{th}$ column which have values of one i.e

$$[\mathbf{D}_{nl}^i] = \begin{cases} 1, & if & n = 1 & and & l = i \\ 1, & if & n = 2 & and & l = i + N \\ 1, & if & n = 3 & and & l = i + 2N \\ 0, & otherwise. \end{cases} \tag{5.26}$$

This matrix is incorporated for convenience to write the probability distribution in terms of all the $3N$ model parameters in column form i.e it selects three elements from $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ at a given time sample i. The constant P_{m0} is the normalization constant.

Objective function

Combining equations (5.6) and (5.23), the objective function takes the form

$$J^{tc}(\mathbf{m}_a) = \frac{1}{2}(\mathbf{d} - \mathbf{L}_a \mathbf{m}_a)^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon}(\mathbf{d} - \mathbf{L}_a \mathbf{m}_a) + R^{tc}(\mathbf{m}_a), \tag{5.27}$$

where R^{tc} is the regularization term given by

$$R^{tc}(\mathbf{m}_a) = 2 \sum_{i=1}^{N} \ln(1 + \mathbf{m}_a^T (\mathbf{\Phi}_a^{tc})^i \mathbf{m}_a). \tag{5.28}$$

The solution of the inverse problem is a model vector $\mathbf{m}_a$ that minimizes the objective function J^{tc}. Thus, differentiating equation (5.27) with respect to $\mathbf{m}_a$ and setting the

resulting expression to zero, we get

$$(\mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} \mathbf{L}_a + 2\mathbf{Q}_a^{tc})\mathbf{m}_a = \mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{C}_d^{-1} \mathbf{\Upsilon} \mathbf{d}, \tag{5.29}$$

where $\mathbf{Q}_a^{tc}$ is a $(3N) \times (3N)$ matrix whose elements are defined as

$$[(\mathbf{Q}_a^{tc})_{kn}] = \sum_{i=1}^{N} \frac{2[(\mathbf{\Phi}_a^{tc})_{kn}^i]}{1 + \mathbf{m}_a^T (\mathbf{\Phi}_a^{tc})^i \mathbf{m}_a}, \qquad k, n = 1, 2, 3, ..., 3N. \tag{5.30}$$

Equation (5.29) can be reduced to

$$(\mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{\Upsilon} \mathbf{L}_a + \mu^{tc} \mathbf{Q}_a^{tc})\mathbf{m}_a = \mathbf{L}_a^T \mathbf{\Upsilon}^T \mathbf{\Upsilon} \mathbf{d}, \tag{5.31}$$

where

$$\mu^{tc} \sim 2\sigma_d^2 \tag{5.32}$$

which means the hyper-parameter μ^{tc} for Trivariate Cauchy regularization is in the order of the variance of the noise terms (square of the standard deviation of the noise).

5.3 Synthetic data examples

In this section, synthetic data examples are given to demonstrate the three regularization strategies discussed above. The velocity and density models (Figure 4.1) and the synthetic data (Figure 4.2) are used. The sample covariance matrix (equation (5.8)) for Gaussian and the scale matrix (equation (5.25)) for Trivariate Cauchy distribution are constructed by smoothing the true reflectivities of the model parameters. The inversion is run in an arbitrarily selected time window 0.2 s to 0.636 s with an overlap of a length of the wavelet above and below the selected time window. Figures 5.1, 5.3, 5.5, 5.7, 5.9, and 5.11 are results of the inversion by regularizing the inverse problem using Multivariate Gaussian, Univariate Cauchy, and Trivariate Cauchy probability distributions. Three of them are results from noiseless ($S/N=10^4$) and three of them are from noisy data ($S/N=4$). The true (blue) and the inverted (red) are plotted in each case one on the top of the other on the same panel to see how much the true and the inverted result match. Along with this, the root mean square errors for single realization are given in each case so that one can easily compare the results using the three methods. Figures 5.2, 5.4, 5.6, 5.8, 5.10, and 5.12 are the input data, predicted data and the difference between the input and the predicted data which show the degree of data fitting. Tables 5.1, 5.2, 5.3 show Monte Carlo Simulation i.e. root mean square errors and their corresponding standard deviation which are calculated by taking 20 realizations at each S/N. Figure 5.13 (**a**) - (**i**) are plots of the root mean square errors versus the S/N for the three model parameters and the three methods. The standard deviations are shown by error bars at various S/N. See Appendix C for the calculation of the RMSE and standard deviations.

All the inverted results shown are at nearly the same level of data misfit using the chi-square test. In the case in which the data has very small noise ($S/N=10^4$), all the three regularizations give good result except the density reflectivity using the Univariate Cauchy which is a little bit noisy. Using the data with $S/N=4$, Univariate Cauchy plays the role of sparsity in the solution but the result is unstable. In this particular example, the noise dominates the true amplitudes in the density reflectivity and affects the P-wave and S-wave reflectivity as well. This may be because the model parameters are treated uncorrelated while they are correlated. In other words, this regularization does not provide stability to the solution. Multivariate Gaussian gives a better solution than the Univariate Cauchy but lacks resolution especially for small layer contrasts and does not preserve the amplitudes. The Trivariate Cauchy allows to retrieve the reflectivities perfectly (with high resolution and also preserves the amplitudes) which shows the stability of the inversion. The Trivariate Cauchy has therefore the advantage over the Univariate Cauchy and Multivariate Gaussian prior distributions combined i.e the role of sparsity and incorporating correlation.

The stability and the uncertainty in the inversion using the various probability distribution functions are also analyzed via Monte Carlo simulation for various S/N. Increasing the noise level corresponds to large uncertainty in the inverted results. Especially, the instability of the result using the Univariate Cauchy regularization is more clearly observed. This can be explained by the absence of the correlation information in this regularization. Adding only the correlation information does not grant the accuracy of the result as seen by the Gaussian prior. The density reflectivity usually has very small magnitude as compared to the P-wave and S-wave reflectivity. Therefore, one has to expect smaller root mean square error as compared to the P-wave and S-wave reflectivity. But the smaller is the magnitude the more likely appears to be dominated by the noise. This characteristic is also seen in the simulation. The relative mean square error in the density reflectivity is larger than the P-wave reflectivity and S-wave reflectivity. Nevertheless, using the Trivariate Cauchy regularization has the least error than the other two for all the three parameters.

5.4 Real data examples

The real data set is provided by Petrobras, Argentina. The data consist of 61 NMO corrected CDP gathers offset ranges from 240 m - 3210 m in a difference of 240 m. The time ranges from 0 to 1.502 s with sampling interval of 2 ms. The wavelet is estimated from each trace and an average wavelet is used for each CDP. The inversion is done only in a time window of 0.75 s -1.35 s which is subdivided into two window each with an overlapping window of 0.146 s (the time length of the wavelet) below and above each window. Separate inversion is done using the target oriented inversion algorithm for each window and the results are put together. The inversion is constrained by including a correlation information matrix which is obtained from well-log data. Understanding the ineffectiveness of the Univariate Cauchy for correlated parameters, the real data inversion is done for Gaussian and Trivariate Cauchy only. Figure 5.14 and 5.15 are the inversion results using the Gaussian and Trivariate Cauchy probability distributions as prior respectively. Figure 5.16a shows the stacked real data. Figures 5.16 b and c shows the data predicted from the results of the inverted reflectivities from each CDP gather and then stacked using the Multivariate Gaussian and Trivariate Cauchy priors respectively.

The real data results show similar behaviors like the synthetic case. The same trend is observed both in the Gaussian and Trivariate Cauchy prior distributions which reflects they do similar job incorporating the well-log information thereby stabilizing the inversion. As in the synthetic case, the resolution enhancement is evident in the Trivariate Cauchy prior. The stacked sections (Figures 5.16 (b) and (c)) are quite similar to the original stacked data (Figure 5.16 a). This implies that the inversion honors the original input data.

Figure 5.1: (a) P-wave reflectivity ($RMSE_A$ = 0.0070), (b) S-wave reflectivity ($RMSE_B$ = 0.0119), and (c) density reflectivity ($RMSE_C$ = 0.0018) retrieved from data with S/N=10^4 using Multivariate Gaussian as *a priori*.

Figure 5.2: (a) The input data (S/N=10^4), (b) the predicted data, and (c) the difference between the input and the predicted data using Multivariate Gaussian as *a priori*.

Figure 5.3: (**a**) P-wave reflectivity ($RMSE_A$ = 0.0237), (**b**) S-wave reflectivity ($RMSE_B$ = 0.0390), and (**c**) density reflectivity ($RMSE_C$ = 0.0059) retrieved from data with S/N=4 using Multivariate Gaussian as *a priori*.

Figure 5.4: (**a**) The input data (S/N=4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Multivariate Gaussian as *a priori*.
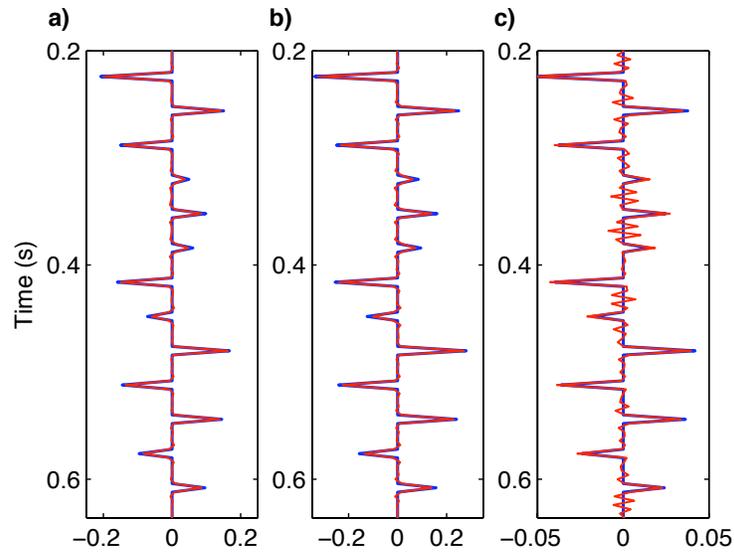
Figure 5.5: (**a**) P-wave reflectivity ($RMSE_A = 0.0055$), (**b**) S-wave reflectivity ($RMSE_B = 0.0104$), and (**c**) density reflectivity ($RMSE_C = 0.0035$) retrieved from data with S/N=10^4 using Univariate Cauchy as *a priori*.
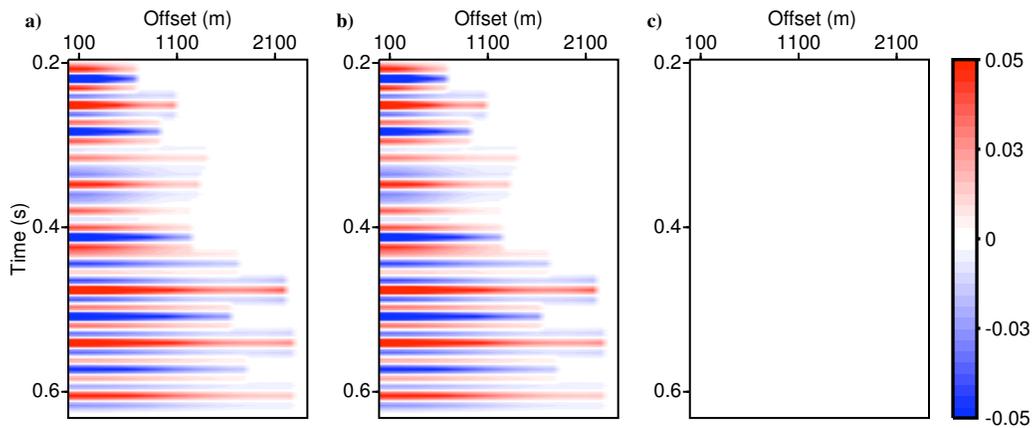
Figure 5.6: (**a**) The input data (S/N=10^4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Univariate Cauchy as *a priori*.
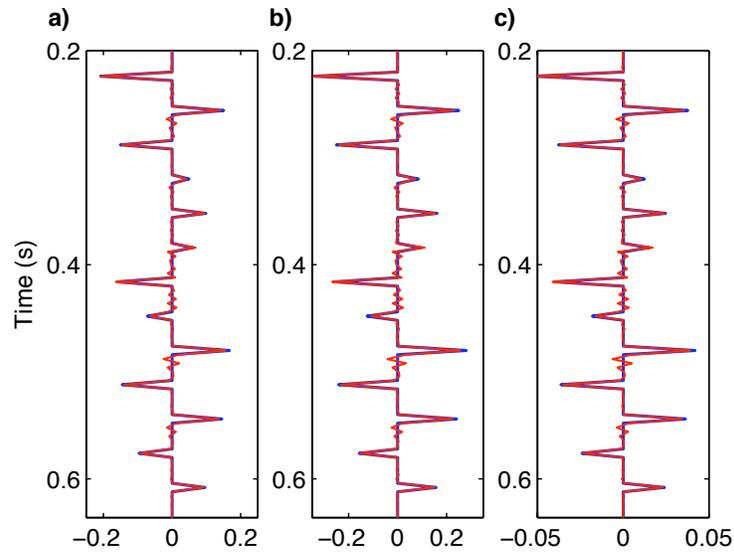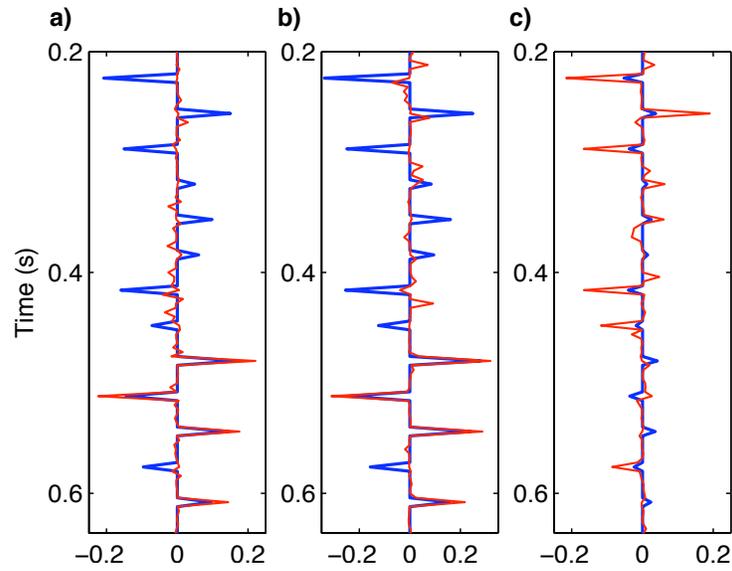
Figure 5.7: (**a**) P-wave reflectivity ($RMSE_A = 0.0391$), (**b**) S-wave reflectivity ($RMSE_B = 0.0611$), and (**c**) density reflectivity ($RMSE_C = 0.0328$) retrieved from data with S/N=4 using Univariate Cauchy as *a priori*.
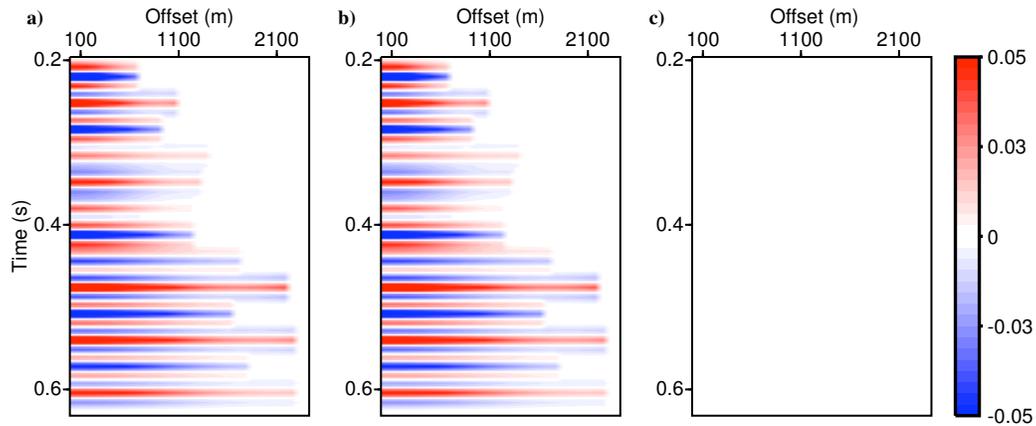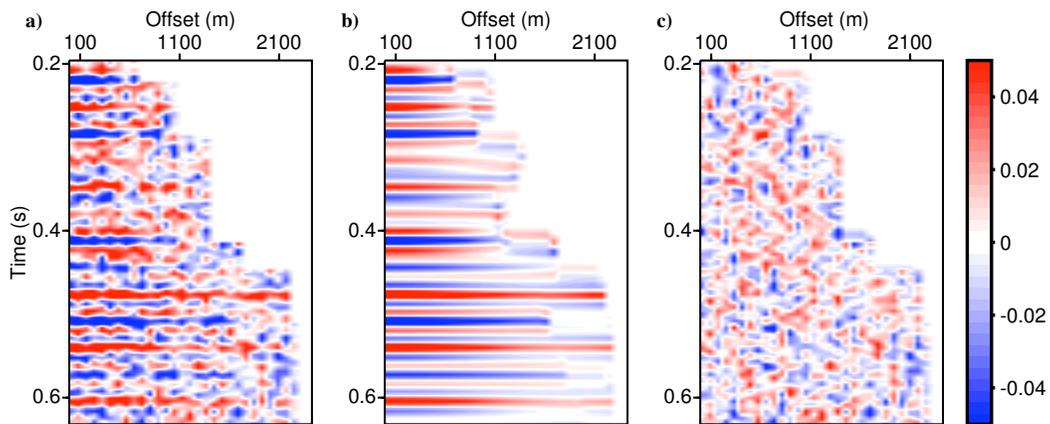

Figure 5.8: (**a**) The input data (S/N=4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Univariate Cauchy as *a priori*.
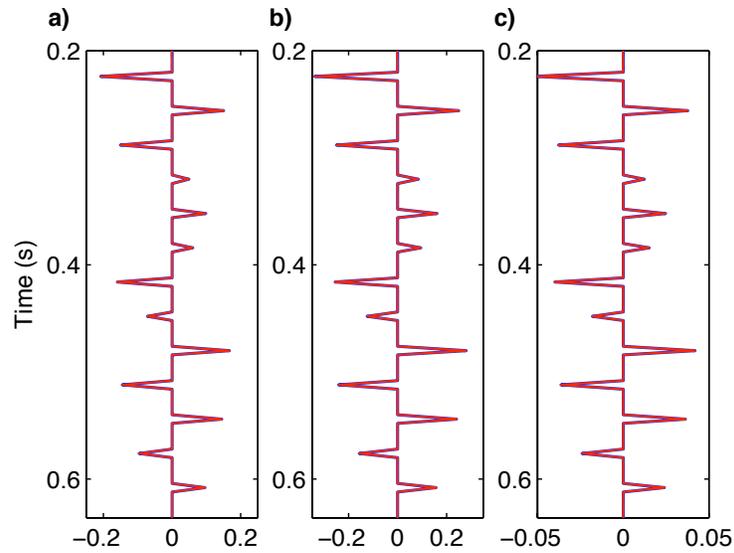
Figure 5.9: (**a**) P-wave reflectivity ($RMSE_A = 2.33 \times 10^{-4}$), (**b**) S-wave reflectivity ($RMSE_B = 4.16 \times 10^{-4}$), and (**c**) density reflectivity ($RMSE_A = 5.66 \times 10^{-5}$) retrieved from data with S/N=10^4 using Trivariate Cauchy as *a priori*.
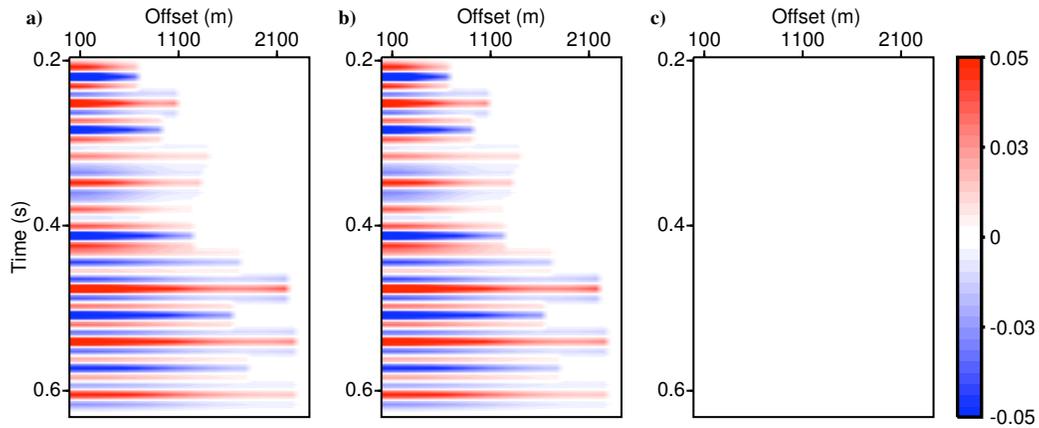
Figure 5.10: (**a**) The input data (S/N=10^4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Trivariate Cauchy as *a priori*.
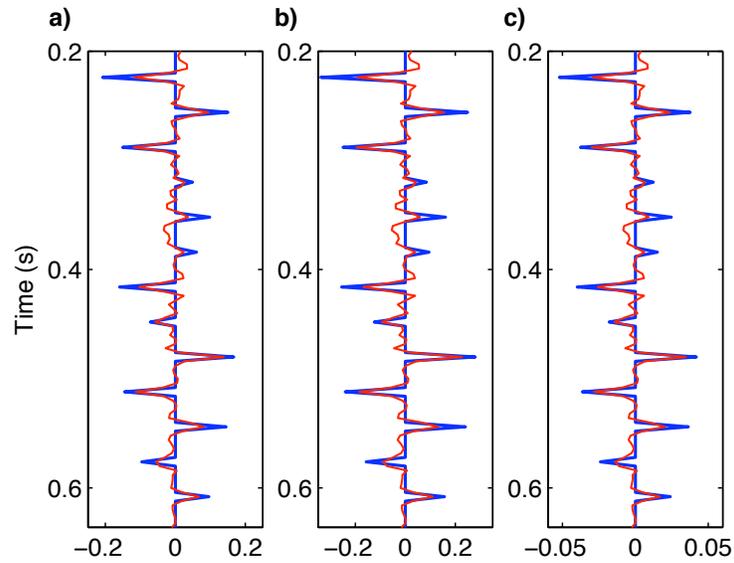
Figure 5.11: (**a**) P-wave reflectivity ($RMSE_A = 0.0044$), (**b**) S-wave reflectivity ($RMSE_B = 0.0073$), and (**c**) density reflectivity ($RMSE_C = 0.0011$) retrieved from data with S/N=4 using Trivariate Cauchy as *a priori*.
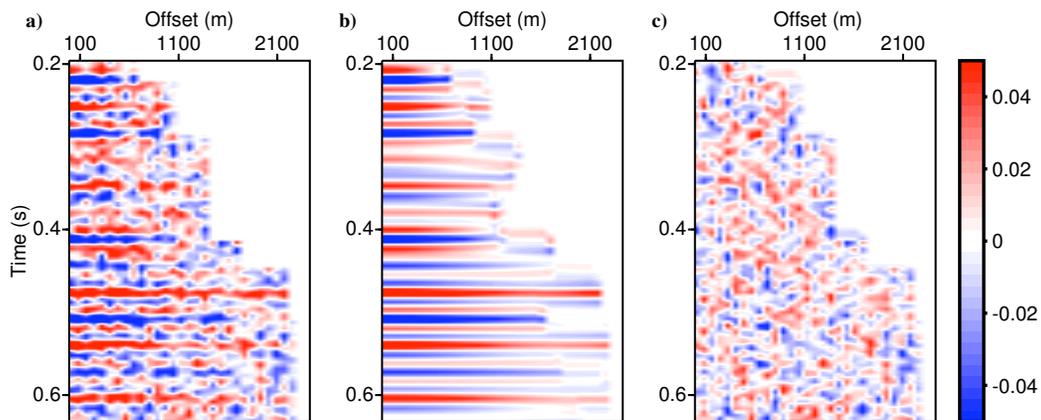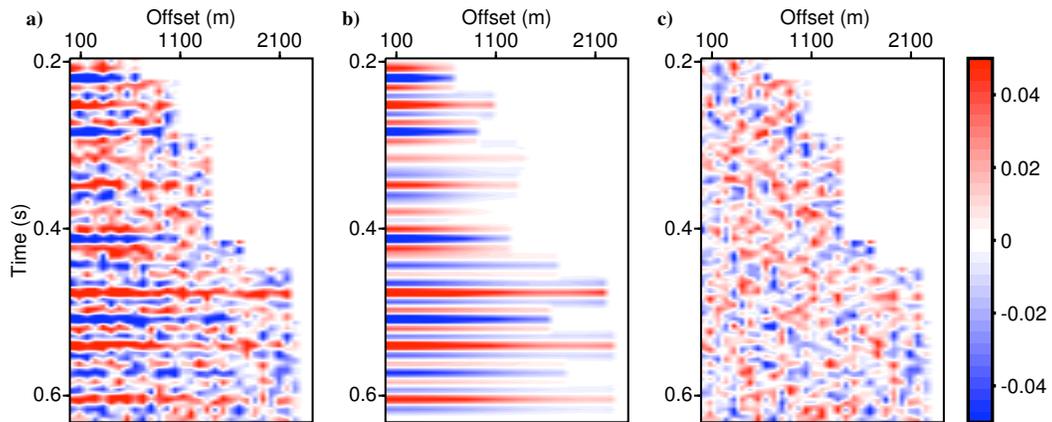

Figure 5.12: (**a**) The input data (S/N=4), (**b**) the predicted data, and (**c**) the difference between the input and the predicted data using Trivariate Cauchy pdf as *a priori*.
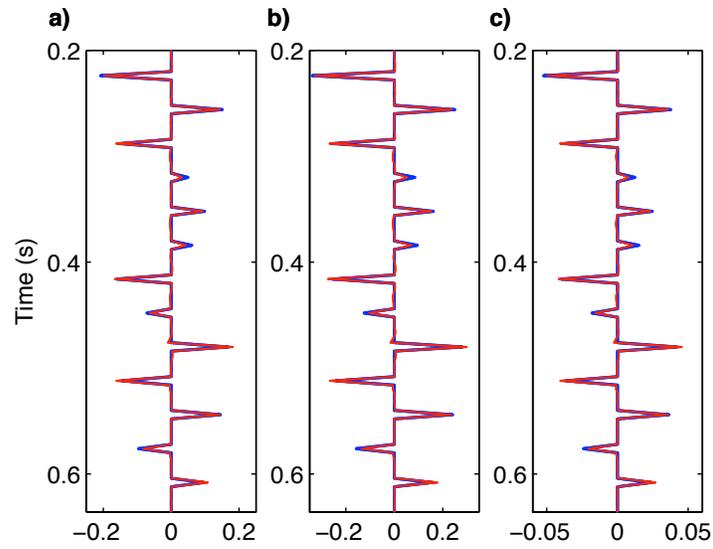
S/N	Univariate Cauchy		Multivariate Gaussian		Trivariate Cauchy	
	$RMSE_A$	$\pm\sigma_A$	$RMSE_A$	$\pm\sigma_A$	$RMSE_A$	$\pm\sigma_A$
100	0.0158	0.0141	0.0069	0.0027	0.0008	0.0004
60	0.0136	0.0130	0.0076	0.0029	0.0011	0.0005
30	0.0333	0.0125	0.0105	0.0038	0.0014	0.0006
15	0.0496	0.0091	0.0154	0.0052	0.0016	0.0008
10	0.0530	0.0118	0.0182	0.0064	0.0021	0.0011
6	0.0542	0.0128	0.0223	0.0067	0.0033	0.0023
4	0.0547	0.0164	0.0257	0.0073	0.0050	0.0039

Table 5.1: The root mean square errors ($RMSE_A$) and their corresponding standard deviations (σ_A) for P-wave reflectivity using the three prior distributions.

S/N	Univariate Cauchy		Multivariate Gaussian		Trivariate Cauchy	
	$RMSE_B$	$\pm\sigma_B$	$RMSE_B$	$\pm\sigma_B$	$RMSE_B$	$\pm\sigma_B$
100	0.0193	0.0172	0.0117	0.0044	0.0016	0.0007
60	0.0175	0.0159	0.0128	0.0048	0.0021	0.0010
30	0.0415	0.0154	0.0175	0.0062	0.0028	0.0012
15	0.0645	0.0136	0.0254	0.0085	0.0032	0.0016
10	0.0710	0.0185	0.0300	0.0105	0.0038	0.0021
6	0.0724	0.0141	0.0368	0.0110	0.0056	0.0039
4	0.0742	0.0044	0.0424	0.0120	0.0084	0.0063

Table 5.2: The root mean square errors ($RMSE_B$) and their corresponding standard deviations (σ_B) for S-wave reflectivity using the three prior distributions.

S/N	Univariate Cauchy		Multivariate Gaussian		Trivariate Cauchy	
	$RMSE_C$	$\pm\sigma_C$	$RMSE_C$	$\pm\sigma_C$	$RMSE_C$	$\pm\sigma_C$
100	0.0153	0.0136	0.0017	0.0007	0.0002	0.0001
60	0.0132	0.0126	0.0019	0.0007	0.0003	0.0001
30	0.0321	0.0123	0.0026	0.0009	0.0004	0.0001
15	0.0474	0.0107	0.0039	0.0013	0.0004	0.0002
10	0.0497	0.0130	0.0045	0.0016	0.0005	0.0003
6	0.0502	0.0132	0.0056	0.0017	0.0008	0.0006
4	0.0500	0.0207	0.0064	0.0018	0.0013	0.0010

Table 5.3: The root mean square errors ($RMSE_C$) and their corresponding standard deviations (σ_C) for density reflectivity using the three prior distributions.
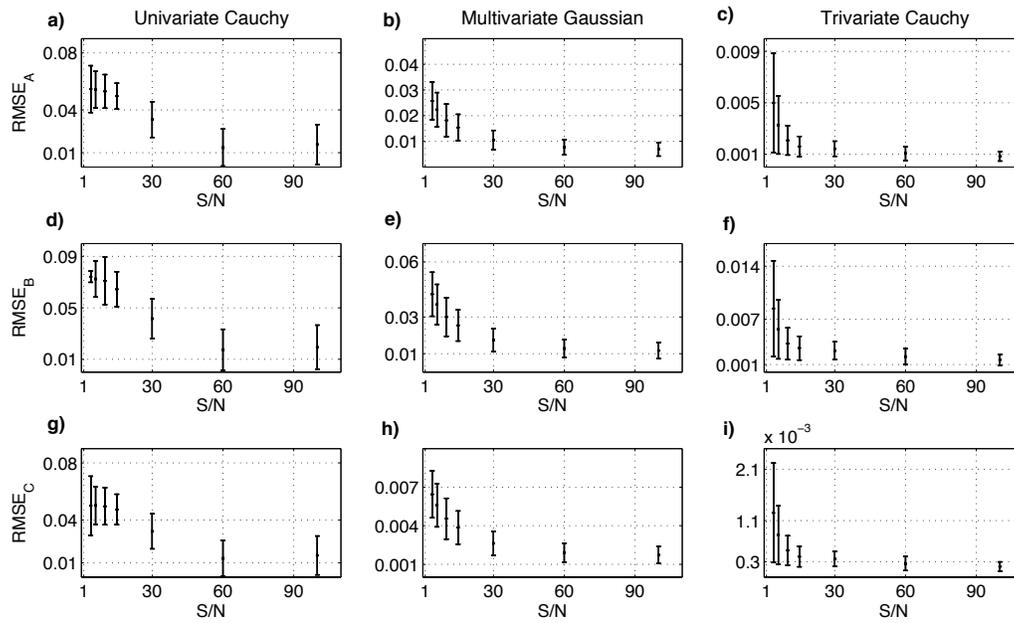
Figure 5.13:   Each figure shows the root mean square errors ($RMSE$) versus the signal to noise ratios (S/N). The error bars are standard deviations of the root mean square errors at the various signal to noise ratios . The column in this figure show the methods and the rows show type of parameter. Indices A, B and C are used for P-wave reflectivity, S-wave reflectivity, and density reflectivity respectively. Note that errors and standard deviations are calculated by taking 20 realizations at each S/N. See Tables 5.1, 5.2, and 5.3.
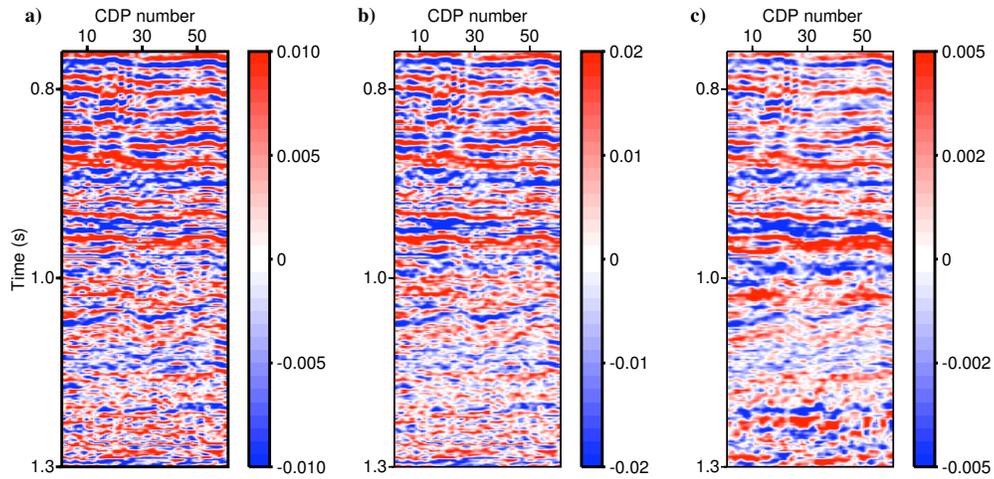
Figure 5.14: (**a**) P-wave reflectivity, (**b**) S-wave reflectivity, and (**c**) density reflectivity using Multivariate Gaussian as *a priori* .
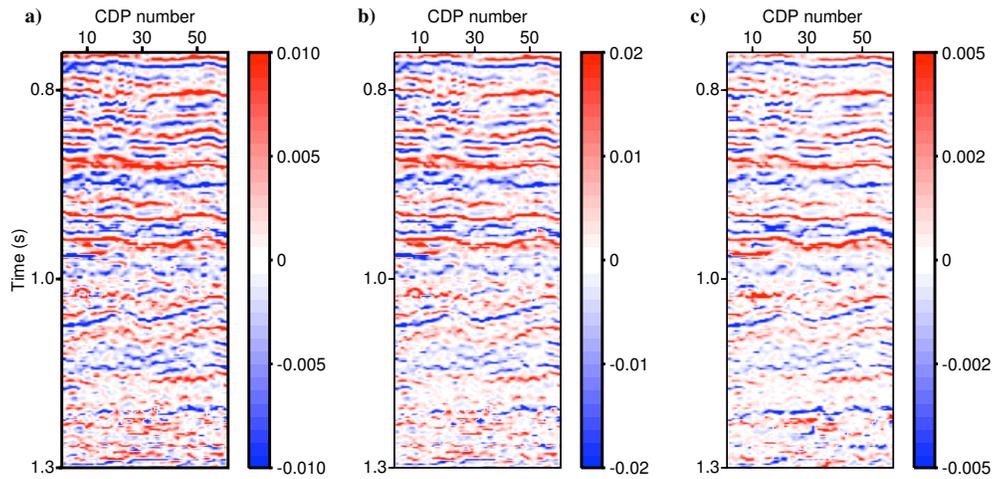
Figure 5.15: (**a**) P-wave reflectivity, (**b**) S-wave reflectivity, and (**c**) density reflectivity using Trivariate Cauchy as *a priori* .
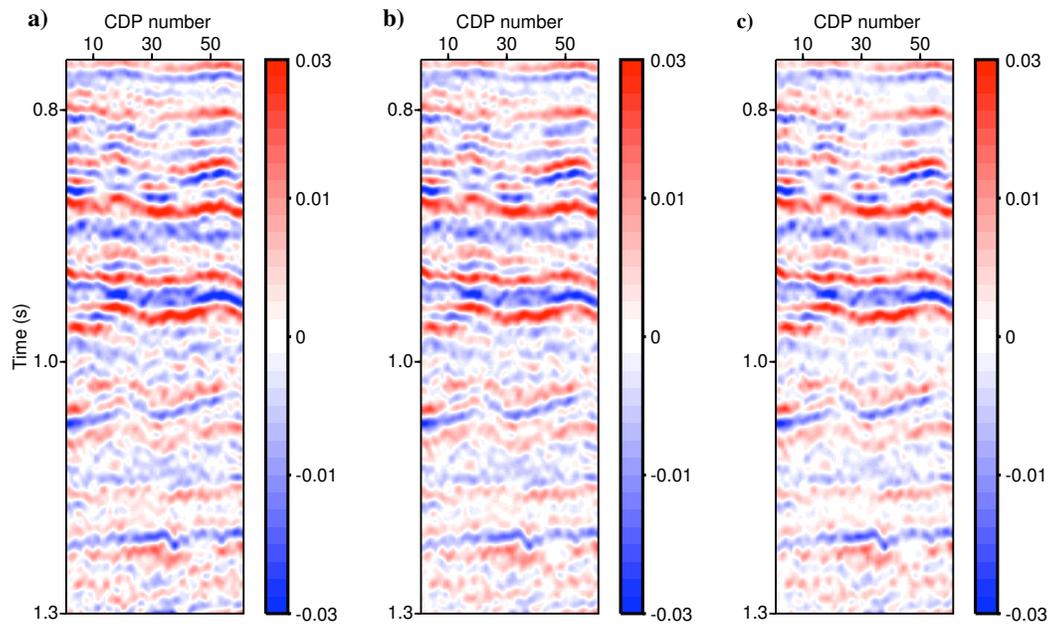
Figure 5.16: (**a**) Stacked section of the real data , (**b**) data generated using the inverted reflectivities (Figure 5.14) and then stacked, and (**c**) data generated using the inverted reflectivities (Figure 5.15) and then stacked.

5.5 Summary

In this chapter, the Bayes' theorem was used to formulate the inversion using three possible prior distributions which could be used to regularize AVO inversion. Each regularization method was investigated with synthetic and real data examples. The synthetic data inversion via Univariate Cauchy prior does a good job in the two-term inversion but the result is unstable when this prior is used for the three-term inversion. This implies that the two-term is more stable than three-term. In other words, the regularization with correlation information is very important to stabilize the inverse problem. This was demonstrated by the Multivariate Gaussian and Trivariate Cauchy regularizations although the result in Multivariate Gaussian prior has no sparsity in the solution. The Trivariate Cauchy regularization has two advantages combined (introducing correlation and also its role for sparsity). In nutshell, it plays a good role in stabilizing the inversion and increase the resolution of the inverted parameters. This makes it promising for better prediction of subsurface physical parameters by avoiding discrepancy which usually arises for decision making.

CHAPTER 6

Conclusions

6.1 Conclusions

The main topic of this thesis was to address the problem of defining the prior distribution for the AVO parameters. Along with this, two issues were investigated for both the two-term and the three-term AVO inversion. These are stability of the inversion and at the same time the vertical resolution of the estimated parameters.

First, the main concern in AVO inversion is instability of the inversion due to the ill-posedness of the problem. The stability of ill-posed inverse problems is resolved by incorporating regularization terms in the least squares inversion. We have seen that regularization can be introduced into an inverse problem using a Bayesian inversion method. It is a probabilistic treatment that allows to choose a solution with certain features.

Second, the resolution of the inverted model parameters is also an important quality measure i.e identifying layer contrasts with preserved amplitudes. An increased resolution of the model parameters can be achieved by including sparsity constraint that have high noise rejection ability. In this thesis, long-tailed probability distribution was used as a prior distribution which leads to a regularization that promotes sparsity.

In order to construct the forward physical model for AVO inversion, the Zoeppritz equations were derived. These equations govern the propagation of plane waves from one medium to another with different properties. They show the relationship between reflection and transmission coefficients in terms of the angle of incidence and the three model parameters (P-wave velocity, S-wave velocity and density). Due to the non-linearity of the equations, they are barely used directly as a forward model for AVO inversion. The linearized approximations which can be used directly for AVO inversion were discussed: some of them contain

two terms and some of them contain three terms. These approximations were compared by plotting the reflection coefficient versus the angle of incidence in two-layered models with small and large layer contrasts. For a given two-layered model, it was demonstrated that as the angle of incidence increases, the deviation from the exact reflection coefficient increases. Comparing two two-layered models which have small and large contrasts, the deviation becomes much more significant in the model with large layer contrast. Nevertheless, all those approximations are useful for AVO analysis. One has to fix the maximum angle to be used for a given approximation equation to decrease discrepancy which may arise due to the inexactness of the model.

Setting a relationship between observed data and model parameters using any of those approximations leads to a linear equation. In order to retrieve those parameters, the Bayesian approach was used to formulate the inverse problem. It makes use of the famous statistical theorem (Bayes' theorem) which is the base for the formulation of the Bayesian inversion. This theorem has three probability distribution to deal with: the posterior distribution, the likelihood function, and *a prior* distribution. In most cases, the likelihood function is built from a Gaussian probability distribution function by assuming Gaussian noise. Therefore, the main concern using Bayesian inversion is to choose a prior distribution. For AVO inversion, two potential prior distributions were investigated in detail: the Multivariate Gaussian and the Multivariate **t** distributions. It has been shown that the Multivariate **t** distributions has a long tails. Another interesting characteristics for the two prior distribution is that as the degree of freedom increases, the Multivariate **t** distributions approaches the Multivariate Gaussian distribution. Therefore, modeling the model parameters such as P-wave velocity, S-wave velocity, and density or their attributes using the Multivariate **t** distributions is still realistic. In this thesis, two prior distribution which belongs to the Multivariate **t** distribution family are proposed. These include the Bivariate Cauchy and Trivariate Cauchy probability distributions which were used to derive regularization terms for two-term and three-term AVO inversion respectively. The main advantage of this kind of regularization is to answer the problems I mentioned above. In other words, the two proposed regularizations allows to incorporate correlation information of the model parameters to stabilize the inversion and thereby increase stability. Furthermore, they also promote sparse solutions (high resolution).

Two other prior distribution were also used to constrain the inversion: the Multivariate Gaussian and Univariate Cauchy distributions. The Multivariate Gaussian prior gives a quadratic regularization. The AVO inverse problem with quadratic regularizations is solved by weighted least squares method. When Univariate Cauchy is used as prior distribution for AVO parameters, it means that the parameters are treated as independent (or uncorrelated). But any of the Cauchy prior distributions (Univariate or Multivariate) result in a non-quadratic regularization term i.e the regularization that depend on the model parameters.

Therefore, IRLS algorithm is used to solve the AVO inverse problem.

The results of the inversion in both synthetic and real data examples indicated that the correlation information is very important to stabilize the inverse problem. This was demonstrated by the Multivariate Gaussian and Multivariate Cauchy regularizations. This especially is evident in the three-term AVO inversion where we have density term. In the two-term case, even Univariate Cauchy may perform good as the inversion is less ill-conditioned than the three-term AVO inversion. But it still shows a little discrepancy due to the treatment of the parameters as independent. In the three-term AVO inversion, Monte Carlo Simulation for inverted results was done by calculating the root mean square errors for a number of realizations at each signal to noise ratio. This helps to understand which part of the model parameters can be resolved reliably as the noise level increases. We have seen that increasing the noise level affects the density reflectivity as compared to the other two parameters (P-wave reflectivity and S-wave reflectivity). Furthermore, for all the three parameters, the Trivariate Cauchy has least uncertainty as compared to the Univariate Cauchy and Multivariate Gaussian regularizations. The over all investigation of the prior distributions used to constrain the inverse problem showed that the Multivariate Cauchy distribution are promising for better prediction of subsurface physical parameters.

Finally, solving inverse problems in general involves an hyper-parameter selection. Large data sets should have an automatic hyper-parameter selection method that can be implemented in the inversion algorithm. More work is required in order to select hyper-parameters automatically.

Bibliography

Aki, K. and P. G. Richards. *Quantitative seismology : theory and methods.* W. H. Freeman, San Francisco, 1980.

Bortfeld, R. "Exact solution of the reflection and refraction of arbitrary spherical compressional waves at liquid-liquid interfaces and at solid-solid interfaces with equal shear velocities and equal densities." *Geophysical Prospecting* 10 (1962): 35–67.

Buland, A. and H. Omre. "Bayesian linearized AVO inversion." *Geophysics* 68 (2003): 185–198.

Castagna, J. P., H.W. Swan, and J. F. Douglas. "Framework for AVO gradient and intercept interpretation." *Geophysics* 63 (1998): 948–956.

Castagna, J.P., M.L. Batzle, and R.L. Eastwood. "Relationships between compressional and shear-wave velocities in clastic silicate rocks." *geophysics* 50 (1985): 334.

Castagna, J.P. and H.W. Swan. "Principles of AVO crossplotting." *The Leading Edge* 16 (1997): 337–344.

Chauanhai, L. "Statistical analysis using the Multivariate t distribution." *Phd thesis, Harvard University* (1994).

Chopra, S. and J. P. Castagna. "Introduction to this special section—AVO." *The Leading Edge* 26 (2007): 1506–1507.

Dahl, T. and B. Ursin. "Parameter estimation in a one-dimensional anelastic medium." *J. Geophys. Res.* 96 (1991).

Downton, J.E. and L.R. Lines. "Three term AVO waveform inversion." *SEG Technical Program Expanded Abstracts* 23 (2004): 215–218.

Eaton, D.W., B. Milkereit, and M. Salisbury. "Seismic methods for deep mineral exploration: mature technologies adapted to new targets." *The Leading Edge* 22 (2003): 580–585.

Fatti, J.L., G.C. Smith, P.J. Vail, P.J. Strauss, and P.R. Levitt. "Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the geostack technique." *Geophysics* 59 (1994): 1362–1376.

Gardner, G. H. F. "Formation velocity and density: the diagnostic basics for stratigraphic traps." *Geophysics* 39 (1974): 770–+.

Haeni, F. P. "Application of continuous seismic reflection methods to hydrologic Studies." *Ground Water* 24 (1986): 23–31.

Johnson, N.L. and S. Kotz. *Distribution in statistics: continuous multi-variate distribution.* New York: John Willey and Sons., 1972.

Kabir, N., R. Crider, R. Ramkhelwan, and C. Baynes. "Can hydrocarbon saturation be estimated using density contrast parameter?:." *CSEG Recorder* (2006): 31–37.

Li, Y. "A study on applicability of density inversion in defining reservoirs." *SEG Technical Program Expanded Abstracts* 24 (2005): 1646–1649.

Mahmoudian, F. and G.F. Margrave. "P-wave impedance, S-wave impedance and density from linear AVO inversion: Application to VSP data from Alberta." *CSPG CSEG convention* (2007).

Mallick, S. "AVO and elastic impedance." *The Leading Edge* 20 (2001): 1094–1104.

McQuillin, R., M. Bacon, and W. Barclay. *An introduction to seismic interpretation.* Graham and Trotman, London, 1984.

Moon, T.K. "The expectation-maximization algorithm." *Signal Processing Magazine, IEEE* 13 (1996): 47–60.

Nadarajah, S. and S. Korz. "Estimation methods for the multivariate t distribution.." *Acta Applicandae Mathematica* 102 (2008): 99 – 118.

Ostrander, W. J. "Plane wave reflection coefficients for gas sands at non-normal angles of incidence." *SEG technical program expanded abstracts* 1 (1982): 216–218.

Potter, C.C., A.K. Dey, and R.R. Stewart. "Density prediction using P-and S-wave sonic velocities." *Geotriad, CSPG, CSEG, CWLS joint convention* (1998).

Richards, P.G. and W. Frasier. "Scattering of elastic waves from depth-depenendent inhomogeneities." *Geophyiscs* 41 (1976): 441–458.

Sacchi, Mauricio D. "Inversion of large ill-posed seismic imaging problems." *SEG Technical Program Expanded Abstracts* 17 (1998): 1657–1660.

Sacchi, M.D. and T.J. Ulrych. "High-resolution velocity gathers and offset space reconstruction." *Geophyiscs* 60 (1995): 1169–1177.

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

Scales, J.A. and L. Tenorio. "Prior information and uncertainty in inverse problems." *Geophysics* 66 (2001): 389–397.

Shewchuk, J.R. An introduction to the conjugate gradient method without the agonizing pain. Technical report, Pittsburgh, PA, USA, 1994.

Shuey, R.T. "A simplication of the Zoeppritz equations." *Geophyiscs* 50 (1985): 609–614.

Sivia, D.S. and J. Skilling. *Data analysis: a Bayesian tutorial.* Oxford University Press, 2006.

Smith, G.C. and P. Gidlow. "The fluid factor angle and the crossplot angle." *73rd Ann. Internat. Mgg: Soc. of Exp. Geophys* (2003): 185–188.

Stein, S. and M. Wysession. *An introduction to Seismology, Earthquakes, and Earth structure.* Blackwell Publishers, 2002.

Sun, P., Y. Li, and X. Lu. "Robust three-term AVO analysis and elastic parameters inversion." *SEG Technical Program Expanded Abstracts* 24 (2005): 289–292.

Taner, M.T., F. Koehler, and R.E. Sheriff. "Complex trace analysis." *geophysics* 44 (1979): 1041–63.

Tarantola, A. *Inverse problem theory, method for data fitting and model parameter estimation.* Elsevier Science Publishers, 1987.

Tikhonov, A.N. and V.Y. Arsenin. *Ill-posed problems in natural sciences.* Oxford University Press, 1987.

Trad, D., T. Ulrych, and M. Sacchi. "Latest views of the sparse Radon transform." *Geophysics* 68 (2003): 386–399.

Ulrych, T.J., M.D. Sacchi, and A. Woodbury. "A Bayes tour of inversion: a tutorial." *Geophyiscs* 66 (2001): 55–69.

Ursenbach, C.P. and R.R. Stewart. "Two-term AVO inversion: equivalences and new methods." *Geophysics* 73 (2008): C31–C38.

Verm, R. and F. Hilterman. "Lithology color-coded seismic sections: the calibration of AVO crossplotting to rock properties." *The Leading Edge* 14 (1995): 847–853.

Vogel, C. R. and M. E. Oman. "Fast, robust total variation-based reconstruction of noisy, blurred images." *IEEE transactions on image processing* 7 (1998): 813–824.

Walden, A. T. "Making AVO section more robust." *Geophysical Prospecting* 39 (1991): 915–942.

Whitcombe, D.N., M. Dyce, C. J. S. McKenzie, and H. Hoeber. "Stabilizing the AVO gradient." *SEG Technical Program Expanded Abstracts* 23 (2004): 232–235.

APPENDIX A

Multivariate t distribution and Regularizations

A.1   Multivariate t distribution

The general central multivariate $\mathbf{t}$ distribution is given by,

$$P(\mathbf{x}^i|\mu, \mathbf{\Psi}, \nu) = \frac{\Gamma(\frac{\nu+p}{2})|\mathbf{\Psi}|^{-1/2}}{(\pi\nu)^{p/2}\Gamma(\frac{\nu}{2})[1 + \frac{1}{\nu}(\mathbf{x}^i - \boldsymbol{\mu})^T\mathbf{\Psi}^{-1}(\mathbf{x}^i - \boldsymbol{\mu})]^{\frac{(\nu+p)}{2}}}, \tag{A.1}$$

The model $\mathbf{x}^i$ is a p dimensional vector whose elements are in the range $-\infty < x_j < \infty$ for $j = 1, 2, 3, ..., p$ which are correlated and having a p dimensional vector center, μ. The degree of correlation is represented by a $p \times p$ dimensional scale matrix, $\mathbf{\Psi}$. The parameter ν is the degree of freedom. For N independent samples $(\mathbf{x}^1, \mathbf{x}^2, \mathbf{x}^3, .., \mathbf{x}^N)$, the joint probability density distribution is the product of each density which can be written as

$$P(\mathbf{X}|\mu, \mathbf{\Psi}, \nu) = P_o(\mathbf{\Psi}, p, \nu)\prod_{i=0}^{N} \frac{1}{[1 + \frac{1}{\nu}(\mathbf{x}^i - \boldsymbol{\mu})^T\mathbf{\Psi}^{-1}(\mathbf{x}^i - \boldsymbol{\mu})]^{\frac{(\nu+p)}{2}}}. \tag{A.2}$$

where

$$P_o(\mathbf{\Psi}, p, \nu) = p_o = \prod_{i=0}^{N} \frac{\Gamma(\frac{\nu+p}{2})|\mathbf{\Psi}|^{-1/2}}{(\pi\nu)^{p/2}\Gamma(\frac{\nu}{2})}$$

and the model $\mathbf{X}$ is now has p times by N elements which can be put in a vector form as

$$\mathbf{X} = [\mathbf{x}_1^1, \mathbf{x}_1^2, ..., \mathbf{x}_1^N, \mathbf{x}_2^1, \mathbf{x}_2^2, ..., \mathbf{x}_2^N, ..., \mathbf{x}_p^1, \mathbf{x}_p^2, ..., \mathbf{x}_p^N]^T.$$

For simplicity, we define p by (pN) matrix $\mathbf{D}^i$ which selects elements in vector $\mathbf{X}$ at a given sample i

$$\mathbf{x}^i = \mathbf{D}^i \mathbf{X}, \qquad i = 1, 2, 3, .., N. \tag{A.3}$$

After substituting equation $(A.3)$ into equation (A.2), we obtain

$$P(\mathbf{X}|\mu, \mathbf{\Psi}, \nu) = P_o \prod_{i=0}^{N} \frac{1}{[1 + \frac{1}{\nu}(\mathbf{D}^i\mathbf{X} - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1}(\mathbf{D}^i\mathbf{X} - \boldsymbol{\mu})]^{\frac{(\nu+p)}{2}}}. \tag{A.4}$$

We can rewrite equation (A.4) in exponential form as follows

$$P(\mathbf{X}|\mu, \mathbf{\Psi}, \nu) = P_o \exp\left\{ -\frac{(\nu+p)}{2} \sum_{i=0}^{N} \ln\left[1 + \frac{1}{\nu}(\mathbf{D}^i\mathbf{X} - \boldsymbol{\mu})^T \mathbf{\Psi}^{-1}(\mathbf{D}^i\mathbf{X} - \boldsymbol{\mu})\right]\right\}. \tag{A.5}$$

This equation was used in Chapters 3 and 4 to derive the prior distributions of parameters for our AVO problem.

A.2   Differentiation of the Regularization $R(\mathbf{m})$

The regularization given by equation (5.28) is derived from the joint probability distribution, equation (A.5), under the assumption that the center for the model parameters is zero ($\boldsymbol{\mu} = 0$). Furthermore, $p = 3$ and $\nu = 1$ which leads to Trivariate Cauchy distribution. In order to minimize the objective function, we need to differentiate the regularization with respect to $\mathbf{m}$. Thus,

$$\frac{\partial R^{tc}(\mathbf{m})}{\partial \mathbf{m}} = \left(\frac{\partial R^t c(\mathbf{m})}{\partial m_1} ... \frac{\partial R^{tc}(\mathbf{m})}{\partial m_k} ... \frac{\partial R^{tc}(\mathbf{m})}{\partial m_{3N}} \right)^T. \tag{A.6}$$

Taking the derivative of $R^{tc}(\mathbf{m})$ with respect to m_k where $k = 1, 2, 3, ..., 3N$, we have

$$\frac{\partial R^{tc}(\mathbf{m})}{\partial m_k} = 2 \sum_{i=1}^{N} \left(\frac{1}{1 + \mathbf{m}^T \mathbf{\Phi}^i \mathbf{m}} \right) \frac{\partial}{\partial m_k} \left(\mathbf{m}^T \mathbf{\Phi}^i \mathbf{m} \right). \tag{A.7}$$

Expanding the second term in equation (A.7), we get

$$\frac{\partial R^{tc}(\mathbf{m})}{\partial m_k} = 2 \sum_{i=1}^{N} \left(\frac{1}{1 + \mathbf{m}^T \mathbf{\Phi}^i \mathbf{m}} \right) \frac{\partial}{\partial m_k} \left(\sum_{l=1}^{3N} \sum_{n=1}^{3N} m_l m_n \mathbf{\Phi}_{ln}^i \right). \tag{A.8}$$

Applying the chain rule and using the fact that

$$\frac{\partial m_l}{\partial m_k} = \delta_{lk}, \tag{A.9}$$

$$\frac{\partial m_n}{\partial m_k} = \delta_{nk}, \tag{A.10}$$

where δ_{lk} and δ_{lk} are Kronecker-delta functions, equation (A.8) becomes

$$\frac{\partial R^{tc}(\mathbf{m})}{\partial m_k} = 2 \sum_{i=1}^{N} \left(\frac{\sum_{n=1}^{3N} 2 m_n \mathbf{\Phi}_{kn}^i}{1 + \mathbf{m}^T \mathbf{\Phi}^i \mathbf{m}} \right). \tag{A.11}$$

After interchanging the order of summation in the above equation, we get

$$\begin{aligned}
\frac{\partial R^{tc}(\mathbf{m})}{\partial m_k} &= 2 \sum_{n=1}^{3N} \left(\sum_{i=1}^{N} \frac{2 \mathbf{\Phi}_{kn}^i}{1 + \mathbf{m}^T \mathbf{\Phi}^i \mathbf{m}} \right) m_n \\
&= 2 \sum_{n=1}^{3N} Q_{kn}^{tc} m_n,
\end{aligned} \tag{A.12}$$

where

$$[Q_{kn}^{tc}] = \sum_{i=1}^{N} \frac{2 \mathbf{\Phi}_{kn}^i}{1 + \mathbf{m}^T \mathbf{\Phi}^i \mathbf{m}}. \tag{A.13}$$

A similar procedure can lead to the above expression but in matrix from

$$\frac{\partial R^{tc}(\mathbf{m})}{\partial \mathbf{m}} = 2 \mathbf{Q^{tc}} \mathbf{m}. \tag{A.14}$$

The matrix $\mathbf{Q}^{tc}$ is $(3N)$ by $(3N)$ non-diagonal matrix whose elements are given by equation (A.13). By same procedure given above the regularization terms for Univariate Cauchy and Bivariate Cauchy distributions can also be derived. The Univariate Cauchy regularization has the form

$$R^{uc}(\mathbf{m}_a) = \sum_{i=1}^{3N} \ln(1 + (\frac{m^i}{\sigma})^2) \tag{A.15}$$

and the differentiation of the latter results in a regularization term $\mathbf{Q}^{uc}$ of size $(3N) \times (3N)$ with elements given by

$$[Q_{ii}^{uc}] = \left(1 + \left(\frac{m^i}{\sigma}\right)^2\right)^{-1}. \tag{A.16}$$

The Bivariate Cauchy regularization is given by

$$R^{bc}(\mathbf{m}) = \frac{3}{2}\sum_{i=1}^{N}\ln(1 + \mathbf{m}^T(\mathbf{\Phi}^{bc})^i\mathbf{m}). \tag{A.17}$$

and the differentiation of this regularization term results a regularization matrix $\mathbf{Q}^{bc}$ of size $(3N) \times (3N)$ with elements are given by

$$[Q_{kn}^{bc}] = \sum_{i=1}^{N}\frac{2(\mathbf{\Phi}^{bc})_{kn}^i}{1 + \mathbf{m}^T(\mathbf{\Phi}^{bc})^i\mathbf{m}} \qquad k, n = 1, 2, 3, ..., 2N. \tag{A.18}$$

Notice that now equation (A.17) contains non-diagonal elements.

APPENDIX B

EM Algorithm

B.1 EM-Algorithm to estimate the scale matrix $\mathbf{\Psi}$

EM algorithm can be used to estimate the scale matrix, ($\mathbf{\Psi}$), and the location parameter, ($\boldsymbol{\mu}_l$). If the center (the location parameter) for the model parameters to be zero, we only need to estimate scale matrix (a matrix that contains the correlation information about the model parameter). Taking the natural log of equation (5.23), the prior distribution for model parameters $\mathbf{m}$, we have

$$L(\mathbf{\Psi}) = \sum_{i=1}^{N}[-\frac{1}{2}\ln|\mathbf{\Psi}| - 2\ln(\pi) - 2\ln(1 + \boldsymbol{\Delta}_i^T\mathbf{\Psi}^{-1}\boldsymbol{\Delta}_i), \tag{B.1}$$

where

$$\boldsymbol{\Delta}_i = \mathbf{D}^i\mathbf{m}. \tag{B.2}$$

Differentiating equation (B.1) with respect to $\mathbf{\Psi}^{-1}$ keeping the other parametes constant, we have

$$\frac{\partial L}{\partial(\mathbf{\Psi}^{-1})} = N\mathbf{\Psi} - \frac{N}{2}diag\{\mathbf{\Psi}\} - \sum_{i=1}^{N}\frac{4}{1 + \boldsymbol{\Delta}_i^T\mathbf{\Psi}^{-1}\boldsymbol{\Delta}_i}\boldsymbol{\Delta}_i\boldsymbol{\Delta}_i^T - \frac{1}{2}diag\{\boldsymbol{\Delta}_i\boldsymbol{\Delta}_i^T\}] \tag{B.3}$$

From equation (B.3), the maximum likelihood estimates of $\boldsymbol{\Psi}$ satisfies

$$\sum_{i=1}^{N} \omega_i \boldsymbol{\Delta}_i = 0 \tag{B.4}$$

and

$$\boldsymbol{\Psi} = \frac{1}{N} \sum_{i=1}^{N} \omega_i \boldsymbol{\Delta}_i \boldsymbol{\Delta}_i^T, \tag{B.5}$$

where

$$\omega_i = \frac{4}{1 + \boldsymbol{\Delta}_i^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Delta}_i}. \tag{B.6}$$

Table B.1: EM Algorithm

---

1. Calculate the model parameters, **m**, from bore-hole data. This is an input for the algorithm.
2. Intialize $\boldsymbol{\Psi}$ (3×3). An identity or sample covariance matrix.
3. Set maximum number of iteration (Max) or a convergence criteria

E-step,

4. For j $<= Max$, use the results of the jth iteration to calculate ω_i^{j+1} as defied by equation (B.6) for all i,
5. then calculate the following

$M^{j+1} = \sum_{i=1}^{N} \omega_i^{j+1} \boldsymbol{\Delta}_i \boldsymbol{\Delta}_i^T,$

M-step,

6. then Calculate

$\boldsymbol{\Psi}^{j+1} = \frac{1}{N} M^{j+1},$

7. repeat the E and M steps until convergence or a maximum number of iteration reached.

APPENDIX C

Calculation of root mean square error

In this section, the calculation of root mean square error (RMSE) which is used for the Monte Carlo Simulation is given. Let the true (t) and the inverted (i) parameters are represented by vectors $\mathbf{x}_j^t$ and $\mathbf{x}_j^i$ respectively at the jth realization. For N number of time samples, each vector has N number of elements. Then the root mean square error at the jth realization, R_x^j, is defined as

$$R_x^j = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(x_{jk}^t - x_{jk}^i)^2}. \tag{C.1}$$

Therefore, for n realizations, the root mean square error can be calculated using

$$\bar{R}_x = \frac{1}{n}\sum_{j=1}^{n}R_x^j. \tag{C.2}$$

The corresponding standard deviation for $\bar{R}_x$ is given by

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(R_x^j - \bar{R}_x^j)^2}. \tag{C.3}$$

From equations (C.2) and (C.3), we can easily see that

$$\bar{R}_x - \sigma_x < RMSE_x < \bar{R}_x + \sigma_x, \tag{C.4}$$

Note that this calculation is for n realizations at a specific S/N. The x represents an AVO parameter; P-wave reflectivity, S-wave reflectivity or density reflectivity.