

**University of Alberta**

**Library Release Form**

NAME OF AUTHOR: Qiang Li

TITLE OF THESIS: High Resolution Hyperbolic Radon  
Transform Multiple Removal

DEGREE: Master of Science

YEAR THIS DEGREE GRANTED: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as hereinbefore provided neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

---

Qiang Li  
Department of Physics  
University of Alberta  
Edmonton, AB  
Canada T6G 2J1

June, 2001

**University of Alberta**

**High Resolution Hyperbolic Radon Transform Multiple  
Removal**

by

**Qiang Li**

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

**Department of Physics**

Edmonton, Alberta

Fall 2001

**University of Alberta**

**Faculty of Graduate Studies and Research**

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled **High Resolution Hyperbolic Radon Transform Multiple Removal** submitted by **Qiang Li** in partial fulfillment of the requirements for the degree of Master of Science.

---

Dr. Mauricio D. Sacchi

---

Dr. Andrew B. G. Bush

---

Dr. Frank Hegmann

---

Dr. Moritz Heimpel

DATE:

*To*  
*Yuyan*

## **Abstract**

Multiple suppression is a long-standing problem in seismic data processing. This thesis focuses on the development of a multiple suppression method based on the hyperbolic Radon transform.

The application of the Radon transform to the problem of multiple suppression is reviewed. The classic multiple attenuation tool, the parabolic Radon transform, is introduced and compared with the hyperbolic Radon transform. An algorithm was developed to compute high resolution velocity gathers using an inversion procedure capable of increasing the focusing power of the hyperbolic Radon transform.

The algorithm is tested with synthetic data and with a marine data set from the Gulf of Mexico.

## Acknowledgements

Bona fida appreciations to my academic supervisor, Dr. Mauricio D. Sacchi, for his kindly guidance and propelling during the process of this research project.

Thanks to Mr. Robert Zschuppe for his kind help with  $\text{\LaTeX}$  and proof reading.

Special thanks to my family for their understanding and encouragement during my studies at the University of Alberta.

# Table of Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Historical background . . . . .	1
1.1.1	Multiple elimination of CMP data . . . . .	6
1.1.2	Velocity analysis and resolution enhancement . . . . .	7
1.2	Scope of the thesis . . . . .	8
1.3	Thesis outline . . . . .	8
<b>2</b>	<b>The problem of multiple elimination</b>	<b>10</b>
2.1	Introduction . . . . .	10
2.2	Basic concepts of reflection seismology . . . . .	11
2.2.1	Root Mean Square (RMS) velocity . . . . .	14
2.2.2	Normal move-out correction . . . . .	17
2.2.3	Velocity analysis . . . . .	18
2.3	Radon transforms and multiple suppression . . . . .	20
2.3.1	Linear Radon ( $\tau - p$ ) transform . . . . .	20
2.3.2	Parabolic Radon transform . . . . .	21
2.3.3	Hyperbolic Radon Transform . . . . .	21
2.3.4	Frequency domain algorithms . . . . .	22
2.3.5	Validity of the Parabolic Radon transform to process seismic data . . . . .	27

## TABLE OF CONTENTS

2.3.6	High resolution parabolic Radon transform . . . . .	29
2.4	Velocity discrimination in the $f - k$ domain . . . . .	33
2.5	Velocity discrimination based on the periodicity of multiples . . .	33
2.6	Summary . . . . .	34
<b>3</b>	<b>Multiple suppression with the hyperbolic Radon transform</b>	<b>36</b>
3.1	Introduction . . . . .	36
3.2	The hyperbolic Radon transform and velocity stacks . . . . .	37
3.2.1	Forward and conjugate mapping . . . . .	37
3.2.2	The inverse problem . . . . .	39
3.2.3	Near and far offset effects . . . . .	41
3.3	Inversion of hyperbolic Radon transform using the method of Con- jugate Gradients . . . . .	43
3.3.1	The Quadratic form . . . . .	46
3.3.2	The method of Steepest Descent . . . . .	51
3.3.3	The method of Conjugate Gradients . . . . .	53
3.4	Examples . . . . .	54
3.5	Summary . . . . .	55
<b>4</b>	<b>Resolution enhancement</b>	<b>58</b>
4.1	Introduction . . . . .	58
4.2	High resolution hyperbolic Radon transform . . . . .	60
4.2.1	Solution of the system of equations . . . . .	63
4.2.2	Hyper-parameter selection . . . . .	64
4.3	Source wavelet deconvolution . . . . .	66
4.3.1	Convolution and Correlation . . . . .	67
4.4	Examples . . . . .	68
4.4.1	Synthetic data examples . . . . .	68



*TABLE OF CONTENTS*

4.4.2	Real data examples . . . . .	71
4.5	Summary . . . . .	74
<b>5</b>	<b>Field data experiments</b>	<b>75</b>
5.1	Introduction . . . . .	75
5.2	Real data processing procedures . . . . .	77
5.3	Final stack after multiple removal . . . . .	78
5.4	Techniques used to improve the computational efficiency of the algorithm . . . . .	78
5.5	Summary . . . . .	79
<b>6</b>	<b>Conclusions</b>	<b>85</b>

# List of Tables

5.1	Field acquisition parameters for the Gulf of Mexico data set. . . .	76
-----	---	----

# List of Figures

1.1	Different types of multiples. . . . .	2
2.1	Seismic data field acquisition geometry. . . . .	14
2.2	The relationship between offset and time. . . . .	15
2.3	Root mean square velocity for multiple layer formations . . . . .	15
2.4	A CMP gather and its velocity spectrum. . . . .	19
2.5	Stacking chart showing a field geometry. . . . .	19
2.6	Examples of parabolic Radon transforms. . . . .	31
2.7	Examples of high resolution parabolic Radon transforms. . . . .	32
3.1	A CMP gather from the Gulf of Mexico data set. . . . .	39
3.2	The velocity gather solved by least squares. . . . .	39
3.3	The data from the velocity gather shown in Figure 3.2. . . . .	40
3.4	One data point at the near offset. . . . .	42
3.5	The forward mapping of Figure 3.4. . . . .	42
3.6	One data point at far offset. . . . .	43
3.7	The inverted velocity gather from Figure 3.6. . . . .	43
3.8	One hyperbolic event in data space. . . . .	44
3.9	The inverted velocity gather. . . . .	44
3.10	2D linear system example. . . . .	46
3.11	Contour plot of the quadratic form. . . . .	47

*LIST OF FIGURES*

3.12	Gradient $f'(\mathbf{x})$ of the quadratic form. . . . .	48
3.13	Contour plot of the quadratic form. . . . .	50
3.14	The search path. . . . .	52
3.15	The solution seeking path of the steepest descent. . . . .	53
3.16	The basic multiple suppression procedure. . . . .	56
3.17	Velocity analysis. . . . .	57
4.1	Regularization terms. . . . .	61
4.2	The operator $\mathbf{Q}$ values for different model elements. . . . .	62
4.3	The operator $\mathbf{Q}$ versus model parameters. . . . .	63
4.4	A synthetic example of the Huber regularization. . . . .	69
4.5	A synthetic example of least square inversion. . . . .	70
4.6	Field data example. . . . .	72
4.7	Field data example. . . . .	73
5.1	Original stacked section from the Gulf of Mexico data set . . . . .	80
5.2	Original data, primaries and associated velocity analysis. . . . .	81
5.3	NMO corrected CMP gather. . . . .	82
5.4	The final stack section after multiple removal. . . . .	83
5.5	Choosing $\tau$ dependent velocities. . . . .	84

# Chapter 1

## Introduction

### 1.1 Historical background

Seismic exploration is widely used in the oil and gas industry. Seismic data are collected in the field, processed and then interpreted to provide a picture of the subsurface geology. Improving subsurface imaging is a constant goal of these steps, and this often depends on how well the data are processed.

The application of seismic imaging techniques has played a significant role in hydrocarbon exploration. Seismic imaging techniques, however, assume that the input data are free of multiples. It is well known, of course, that this is not true. The presence of multiple reflections, which are reflections that have been reflected back into the Earth more than once, leaves us with incorrect seismic images. Figure 1.1 is a diagram with the terminology often used in exploration seismology to describe multiple reflections.

It is also important to stress that, as imaging techniques increase in complexity, multiple attenuation methods become important at the time of imaging data acquired in complex environments. Sub-salt imaging is a dramatic example (Hill et al., 1999).

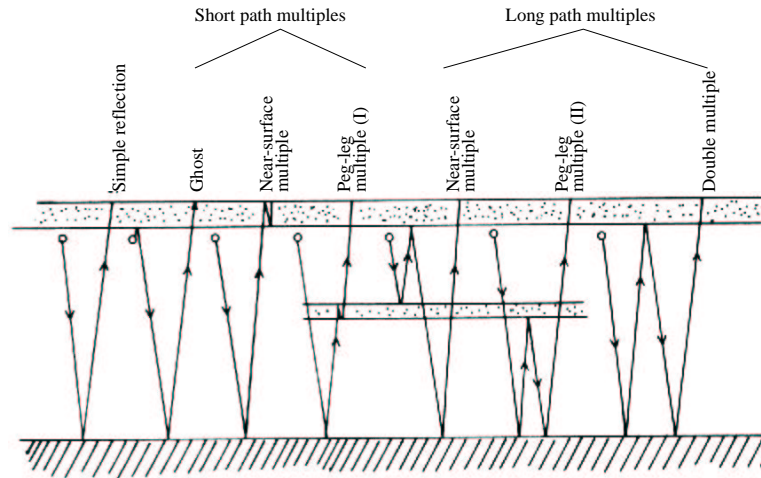


Figure 1.1: Different types of multiples (After Sheriff, 1991).

In addition to sub-salt imaging, multiples also hamper the proper interpretation of seismic data in amplitude versus offset (AVO) analysis (Foster and Mosher, 1992). In this case the amplitude variation of primary reflections can be contaminated by the presence of multiples. The interpretation of this type of data can be quite cumbersome if multiples are not properly removed.

Multiple attenuation techniques have their roots in the early days of the development of seismic processing methods. The introduction of common depth point (CDP) acquisition and the development of common mid-point (CMP) stacking method is recognized as the first attempt for multiple attenuation in stacked sections (Hill et al., 1999). Another early technique is the use of predictive deconvolution to eliminate multiples generated in the water column during marine acquisition (Yilmaz, 1987).

In seismic processing it is assumed that reflection data are free of multiples. If the primaries are contaminated with multiples, they can be misinterpreted as, or interfere with, multiples. This is a problem with a long history, and not totally solved yet. There are many methods that can be used to remove multiples,

but they are only valid when their assumptions and prerequisites are met. In many cases these assumptions are violated or the prerequisites are difficult or impossible to attain; hence, multiple removal remain a problem.

Applied geophysicists have been working on multiple suppression problems for a long time. With the industry trend toward more complex, costly, and challenging exploration targets multiple removal solutions have become more and more advanced. For example, deep water exploration with a dipping ocean-bottom, sub-salt and sub-basalt processing are environments where traditional methods can fail and highly sophisticated multiple elimination methods are required. The growing economic risk and complexity of these exploration and production objectives demand more advanced techniques and methods that can accommodate less *a priori* information, fewer restrictions and unrealistic assumptions (Weglein, 1999).

In general, methods that can satisfy the demand of advanced exploration techniques increase processing costs and require a dense sampling of the seismic wave field. However, such surveys can still be accepted if they enable the removal of multiples while preserving primaries. The extra cost of processing is not really expensive compared with that of a dry hole (Weglein, 1999).

Multiple suppression methods can be classified in methods based on move-out discrimination, methods that predicts the multiples using the wave equation (Weglein et al., 1992; Verschuur, 1999), and stochastic methods where the prediction of the multiples is done using prediction error filters (Taner, 1980). In this thesis we will concentrate on methods that extract the multiple using the concept of move-out discrimination. In particular, we will develop a linear transform capable of isolating and discriminating multiples.

In general, multiple removal methods based on the concept of move-out discrimination can be summarized as follows: CMP gathers are modeled as a super-

position of waveforms with hyperbolic travel-time curves. The fact that multiples and primaries exhibit different curvature permits one to use 2D linear transforms to discriminate, isolate and filter the energy that corresponds to multiple reflections. Techniques that follows this scheme can be easily described in terms of Radon transforms.

The Radon transform is a mathematical technique that has seen popular usage in seismic data processing and image analysis in recent years (Zhou, 1994). It is more commonly known as the slant-stack technique or the  $\tau - p$  transform, which is a discrete Radon transform (Turner, 1990).

There are three types of Radon transforms: linear, parabolic and hyperbolic. In general, one can say, that the linear Radon transform is a linear operator that maps linear events in  $t - x$  (time-offset) domain into points in the transformed domain (often called the  $\tau - p$  domain). Similarly, the parabolic and hyperbolic Radon transforms are capable of mapping parabolic events and hyperbolic events to points, respectively.

The mathematical theory of the Radon transform can be found in the book by Deans (1983). The fundamental properties were examined by Durrani and Bisset (1984). For the linear  $\tau - p$  transform, Chapman (1981) presented the exact forms of the generalized Radon transform pairs for a point source in Cartesian or spherical coordinates, and for a line source in cylindrical coordinates.

This thesis focuses on theoretical and practical aspects of the hyperbolic Radon transform. The latter has been proposed by Thorson and Claerbout (1985) as a means to improve the signal to noise ratio (S/N) of CMP gathers and as tool for velocity analysis. These researchers have developed a non-linear algorithm where the least squares error constraint on the reconstructed data is used to estimate the Radon panel (designated by them as velocity-stack). They have also imposed a minimum entropy constraint that is used to enhance the



focusing power (resolution) of the transform.

Hampson (1986) proposed to implement the Thorson-Claerbout method in the  $f$ - $x$  (frequency-offset) domain after adopting the parabolic approximation. The residual move-out (curvature) of seismic reflections after normal move-out (NMO) correction can be approximated by a parabolic travel-time curve. Yilmaz (1989), on the other hand, employed a  $t^2$  stretching to validate the parabolic approximation. In other words, the hyperbolic events in CMP gathers are transformed into parabolic events and, consequently, the more efficient parabolic Radon transform can be adopted. Fast versions of the Parabolic Radon Transform were proposed by Kostov (1990) and Gulunay (1990). These researchers realized that the Toeplitz structure of the parabolic Radon transform permits to use fast solvers like Levinson recursion (Marple, 1987).

Sacchi and Ulrych (1995) introduced a frequency-domain sparse inversion algorithm that improves the velocity resolution and can be used to reconstruct near and far offset missing traces in a cost effective manner. This approach parallels the one of Thorson and Claerbout (1985), however, Sacchi and Ulrych have adopted a parabolic Radon Transform rather than a hyperbolic one. Recently, Sacchi and Porsani (1999) proposed a new method to further reduce the computational cost of the Radon Transform. In this approach the conjugate gradient method is used in conjunction with fast algorithms that operate with circulant matrices to efficiently compute the Radon operator.

The relationship between the CMP data and the corresponding transformed gather can be obtained as the solution of an inverse problem. The relationship between them can be written as

$$\mathbf{L}\mathbf{m} = \mathbf{d}, \quad (1.1.1)$$

where  $\mathbf{d}$  denotes the CMP gather,  $\mathbf{m}$  denotes Radon gather and  $\mathbf{L}$  denotes the

Radon operator.

The linear and parabolic Radon transforms can be written in the  $f-x$  domain and, consequently, break down the problem into many small problems one at each frequency in the signal band-width. This is not true for the hyperbolic Radon transform, in this case we need to devise techniques to operate with large sparse operators in the time-offset domain. Since  $\mathbf{L}$  is a large operator, its inversion can be quite cumbersome. In this thesis we will examine the problem of inverting equation (1.1.1) using a new regularization scheme based on the Huber norm (Huber, 1981). We will show that the Huber norm yields to high resolution transforms.

### 1.1.1 Multiple elimination of CMP data

Removing reverberations from reflection seismograms has been a long-standing problem in exploration geophysics. Multiple reflections often destructively interfere with the primary reflections of interest. The most robust and effective way to suppress multiples is stacking normal move-out corrected seismic gathers. Unfortunately, stacking does not eliminate all multiples. Also, stacking attenuates multiples only in stacked seismograms (Yilmaz, 1987).

Traditional velocity-stack gathers consist of constant-velocity CMP stacked traces. It emphasizes the amplitudes that follow hyperbolic travel-time trajectories in the CMP gather. Ideally, a hyperbolic event on a CMP gather maps onto a point on a velocity-stack gather. Because a CMP gather only includes a cable length portion of a hyperbolic trajectory, this mapping is not exact. The finite cable length, discrete sampling along the offset axis and the closeness of hyperbolic summation paths at near-offsets cause smearing of the stacked amplitudes along the velocity axis. Unless the smearing is removed, inverse mapping from velocity space back to offset space does not reproduce the original amplitudes in

the original CMP gather.

Usually, parabolic Radon transforms are used to obtain velocity-stack gathers where one can suppress the multiples. The method involves a  $t^2$ -stretching or NMO correction. Time stretching and/or NMO correction maps reflection amplitudes along hyperbolic move-out curves to those along parabolic move-out curves. The CMP gather is Fourier transformed along the time axis. Each Fourier component is then used in the least-squares minimization to compute the corresponding Fourier component of the proper velocity-stack gather. Finally, inverse transforming and undoing the stretching or NMO correction yield the proper velocity-stack gather. During this inverse mapping multiples, primaries, or all of the hyperbolic events can be modeled.

When the hyperbolic Radon transform is adopted there is no need of pre-processing the data with the  $t^2$  stretching or the NMO correction, Hyperbolic Radon transforms operate directly on CMP gathers (after conventional processing, i.e., muting, gain, etc).

### 1.1.2 Velocity analysis and resolution enhancement

There are three fundamental seismic data processes: deconvolution, CMP stacking and migration. Velocity analysis plays a very important role in seismic data processing. The velocities affect the last two procedures.

By using redundancy of information in CMP gathers, stacking can significantly suppress uncorrelated noise, thereby increasing the S/N. It also can attenuate a large part of the coherent noise in the data, such as guided waves and multiples. The normal move-out correction before stacking is done using the primary velocity function. Because multiples have larger move-out than primaries, they are under-corrected and, hence attenuated during stacking (Yilmaz, 1987). Without correct velocities, the CMP data can not be stacked as expected, the

events can not be enhanced, and the S/N can not be improved.

In traditional processing the velocities are obtained from velocity compute using semblance analysis (Yilmaz, 1987). The resolution of the velocity spectrum computed using semblance is often not good enough to discriminate primaries from multiples. The high resolution hyperbolic Radon transforms can overcome the aforementioned problem.

## 1.2 Scope of the thesis

This thesis describes the methods associated with Radon transforms in the time and frequency domains. The scope of the thesis is a detailed study of the hyperbolic Radon transform in terms of inversion. The inversion is regularized by the Huber norm, which is capable of leading to high resolution velocity panels. Synthetic and real data from the Gulf of Mexico are used to test the algorithm. In the real data example, I have successfully managed to remove multiple reflections in pre-stack marine data. The multiple attenuated seismic section displays an important S/N enhancement.

## 1.3 Thesis outline

The structure of the thesis is as follows:

- In Chapter 1 I provide the motivation and scope of my research.
- In Chapter 2 I provide the background information required to understand current technologies in multiple suppression.
- Chapter 3 focuses on the hyperbolic Radon transform in terms of traditional least squares, and conjugate gradients (a semi-iterative algorithm often used to solve large system of equations).

- Chapter 4 introduces multiple suppression by means of a high resolution hyperbolic Radon transform.
- Chapter 5 describes real data experiments. A marine data set from the Gulf of Mexico is used to illustrate the results.
- Chapter 6 summarizes my research and provide a discussion of the most relevant findings. I will also discuss possible future research directions. Finally, I will discuss the advantages and weaknesses of de-multiple using hyperbolic Radon transforms.

## Chapter 2

# The problem of multiple elimination

### 2.1 Introduction

Multiple suppression methods can be divided in three categories:

1. Methods that exploit move-out difference between primaries and multiples (velocity discrimination).
2. Methods that exploit the periodicity of multiples.
3. Wave equation methods.

This thesis, as I have already mentioned, focuses on de-multiple methods based on move-out discrimination. Primary reflections and multiples exhibit different curvature in their travel-time curves. This feature is used to stack out multiples, by summing along the travel-time curves associated to primaries reflections and/or to design Radon transforms capable of mapping multiples and primaries to different areas of the transformed domain.

The travel-time curves of primaries and multiples are given in terms of normal moveout (NMO) velocities. When the NMO correction is applied using the velocities of the primaries, as is normally done to generate stack section, the primaries are aligned while the multiples are under-corrected. This suggests that CMP stacking itself is a viable method of multiple suppression (Yilmaz, 1987). It is well known that stacking far offsets works well to suppress multiples. However, stacking near offsets works against multiple suppression since the move-out difference between primaries and multiples is negligible. The simplest way around this problem is to apply an inside mute to the CMP gathers before stacking. Another problem then emerges: the outside mute. The severity of this mute governs the amount of far-offset data left at early times for velocity discrimination. If there is a severe multiple problem, an effort must be made to preserve the maximum amount of far-offset data associated with the target events.

## 2.2 Basic concepts of reflection seismology

In this section I will review some basic concepts and terminology utilized in exploration seismology. Readers familiar with reflection seismology can skip this section.

Seismic data are collected in the field by many pairs of sources and receivers. Logistic and processing requirements often dictate the extent and sampling of a seismic survey. In this section I will discuss some common terminology used in reflection seismology. In particular, I will provide a review of common-mid-point gathers and normal-move-out corrections.

### Common Shot Gathers (CSG)

Figure 2.1(a) shows the common shot gather. This type of acquisition geometry is often used in the field to collect 2D data. Sources and receivers are moved along the seismic line in order to sample the subsurface geology with seismic waves. A CSG is a collection of seismograms with common source position, mathematically one can express a CSG as follows:

$$D(x_s = x_0, x_r, t),$$

where  $x_s$  (constant) is the source position along the seismic line,  $x_r$  (variable) is the receiver position and  $t$  the time. The constant  $x_0$  is used to indicate the fix position of the source. In general, one obtains a collection of CSG by making  $x_0$  a variable. A collection of CSG can be reorganized in common receiver gathers. The latter can be indicated as follows

$$D(x_s, x_r = x_0, t),$$

please note that in this case we collect all the seismograms with the same receiver position and different source position.

### Common Midpoint Gather (CMP)

Each seismogram has associated a source and receiver position. In general, it is more convenient to define the data volume in terms of two new variables: midpoint and offset. We define offset as the distance between source and receiver for a given seismogram

$$h = x_r - x_s .$$



Now, we define the midpoint coordinate as the distance half away the source and the seismogram

$$x_m = \frac{x_r + x_s}{2}.$$

A common midpoint (CMP) gather is defined as a collection seismograms sharing the same midpoint position

$$D(x_m = x_0, h, t).$$

In the last expression we have represented the data volume in terms of the  $h - x_m$  variables. Similarly, we can define a common offset gather (COG) as a collection of seismograms sharing the same offset

$$D(x_m, h = h_0, t).$$

Seismic data are acquired using common shot gathers (CSG) and then re-organized (using sorting algorithms) into other type of gathers (i.e., common mid point gathers).

Figure 2.1(b) shows a CMP gather and the associated source-receiver positions. In Figure 2.1(b) we also illustrate the CDP (common depth point position). For a horizontal interface the horizontal CDP and CMP positions are equal. This is not true for dipping interfaces.

Finally, it is important to mention that one may think that there is only one geophone for each receiver point; this is not true. Usually, in order to eliminate surface waves and ground noise, several geophones are placed at each receiver position, this is often called a group or an array. A group of geophones can be designed to operate like an analog filter to reject surface waves (Telford et al., 1990).

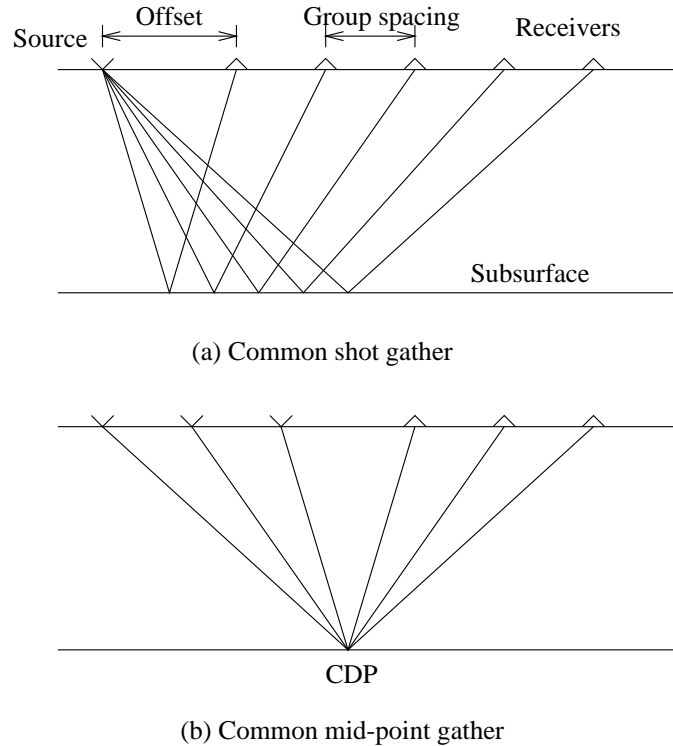


Figure 2.1: Seismic data field acquisition geometry. (a) A common shot gather. (b) A common mid-point gather.

### 2.2.1 Root Mean Square (RMS) velocity

The energy generated by the seismic source propagates down into the Earth until being reflected back by geological interfaces. Figure 2.2 is used to derive the relationship that exists between the travel-time and offset for a single horizontal layer. The travel-time curve for a perturbation traveling from source  $O$  to receiver  $d$  can be expressed as

$$t = \frac{OA + Ad}{V} = \frac{\sqrt{(2d_0)^2 + h^2}}{V} = \sqrt{\frac{4d_0^2}{V^2} + \frac{h^2}{V^2}} = \sqrt{t_0^2 + \frac{h^2}{V^2}}, \quad (2.2.1)$$

where  $t_0$  is the two way travel-time at zero offset and the velocity  $V$  is the medium velocity. Equation (2.2.1) is the equation of a hyperbola. This is only true for

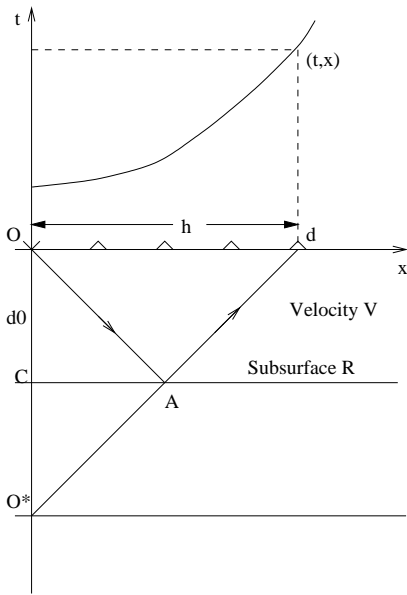


Figure 2.2: The relationship between offset and time.

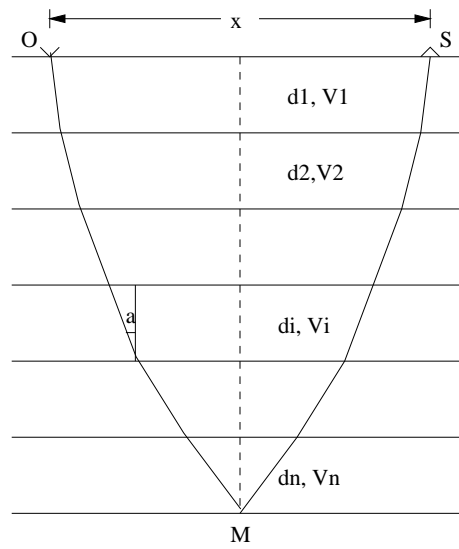


Figure 2.3: Root mean square velocity for multiple layer formations

the one layer case, what about multiple layers?

Let us look at the multi-horizontal-layer case illustrated in Figure 2.3. In this case  $O$  is the source, the reflection from layer  $M$  is recorded at point  $S$ . The travel-time along the ray path  $OMS$  can be expressed as

$$t = 2 \sum_{i=1}^n \frac{d_i}{\cos \alpha_i V_i}, \quad (2.2.2)$$

the corresponding offset is

$$x = 2 \sum_{i=1}^n d_i \tan \alpha_i. \quad (2.2.3)$$

Equation (2.2.2) and (2.2.3) express the time offset relationship for seismic rays in the multi-horizontal-layer case. Usually, we express the relationship in terms

of the ray parameter  $p$ , which is defined from Snell's law as

$$p = \frac{\sin \alpha_1}{V_1} = \frac{\sin \alpha_2}{V_2} = \dots = \frac{\sin \alpha_i}{V_i} = \dots = \frac{\sin \alpha_n}{V_n}. \quad (2.2.4)$$

After combining equations (2.2.2), (2.2.3) and (2.2.4), one has

$$t = \sum_{i=1}^n \frac{t_i}{\sqrt{1 - p^2 V_i^2}}, \quad (2.2.5)$$

$$x = \sum_{i=1}^n \frac{p t_i V_i^2}{\sqrt{1 - p^2 V_i^2}}, \quad (2.2.6)$$

where  $t_i$  is the vertical travel-time in the  $i$ th layer. The above equations are exact expressions for the travel-time and offset relationships in a multi-horizontal-layer case. For practical purposes a simplified form of the last two equations is needed. When the following condition is satisfied

$$x < 2 \sum_{i=1}^n \frac{V_i d_i}{(V_M^2 + V_i^2)^{\frac{1}{2}}}, \quad (2.2.7)$$

where  $V_M$  is the maximum velocity in the  $n$  layers. Equation (2.2.5) and (2.2.6) can be expanded into a geometric series (Telford et. al, 1990):

$$t^2 = t_0^2 + \sum_{j=1}^{\infty} \beta_j x^{2j}. \quad (2.2.8)$$

where  $t_0 = 2 \sum_{i=1}^n t_i$ . After some mathematical manipulations equation (2.2.8) can be written as

$$t^2 = t_0^2 + \frac{x^2}{\frac{\sum_{i=1}^n t_i V_i^2}{\sum_{i=1}^n t_i}} - \frac{1}{4} \left[ \frac{(\sum_{i=1}^n t_i)(\sum_{i=1}^n t_i V_i^4)}{(\sum_{i=1}^n t_i V_i^2)^4} - \frac{1}{(\sum_{i=1}^n t_i V_i^2)^2} \right] x^4 + \dots, \quad (2.2.9)$$

if terms with rank higher than  $x^2$  are neglected, and after defining the following new variable

$$V_R^2 = \frac{\sum_{i=1}^n t_i V_i^2}{\sum_{i=1}^n t_i}, \quad (2.2.10)$$

then equation (2.2.9) can be written as

$$t^2 = t_0^2 + \frac{x^2}{V_R^2}. \quad (2.2.11)$$

We have arrived an equation that relates offset and travel-time in the multi-layer case. This expression is equivalent to the equation derived for the single horizontal layer case. However, it is important to note that now the velocity is not the velocity of the first layer (equation (2.2.1)) but  $V_R$ . The latter is called the root mean squared velocity (RMS velocity). That is to say, when we consider the relationship between travel-time and offset in a multiple-layer media as a hyperbola, the correspondent velocity is the root mean squared velocity; this is a special average of the velocities of the layers above the reflection point.

It is important to note that all of the above explanations are valid for the horizontal case. For dipping interfaces, the relationship between travel-time and offset can still be expressed via hyperbolic curves. However, the definition of  $V_R$  also involves the dip of each layer.

### 2.2.2 Normal move-out correction

We have already shown that travel-times in a CMP gather can be approximated by hyperbolic curves parametrized by the root mean squared velocity. If the velocity of the reflection is known, one can removing the offset dependence from the data. In other words, seismic reflections can be turned into horizontal events.

This correction is called the normal move-out corrections (NMO). Mathematically, we can say that if the time of the reflection before NMO was given by

$$t^2 = t_0^2 + \frac{x^2}{V_R^2}.$$

then, after NMO, we have

$$t^2 = t_0^2.$$

The latter is equivalent to remove the offset dependence from the data or, in other words, to simulate an experiment where sources and receivers in the CMP gather share the same position (Sacchi, 2001). This is an important step towards the construction of a seismic section. The NMO corrected CMP gather can be stacked (horizontal summation) to produce an enhanced trace. A collection of stacked traces is called a seismic section.

### 2.2.3 Velocity analysis

Velocity analysis is the basis for determining the velocities required to apply the NMO correction. In order to obtain velocity information, velocity analysis is performed on selected CMP gathers or groups of gathers. The output of the velocity analysis process is often given as a velocity spectrum, this is a panel displaying velocities versus two way zero-offset time. The energy clusters in the velocity spectrum represent a measure of signal coherency along the hyperbolic trajectories (Yilmaz, 1987). Figure 2.4 shows a CMP and its velocity spectrum. The vertical axis are the two way zero-offset travel-time. Primary velocities in general increase with time (depth). On the other hand, multiples generated by interactions in shallow layers exhibit low velocities.

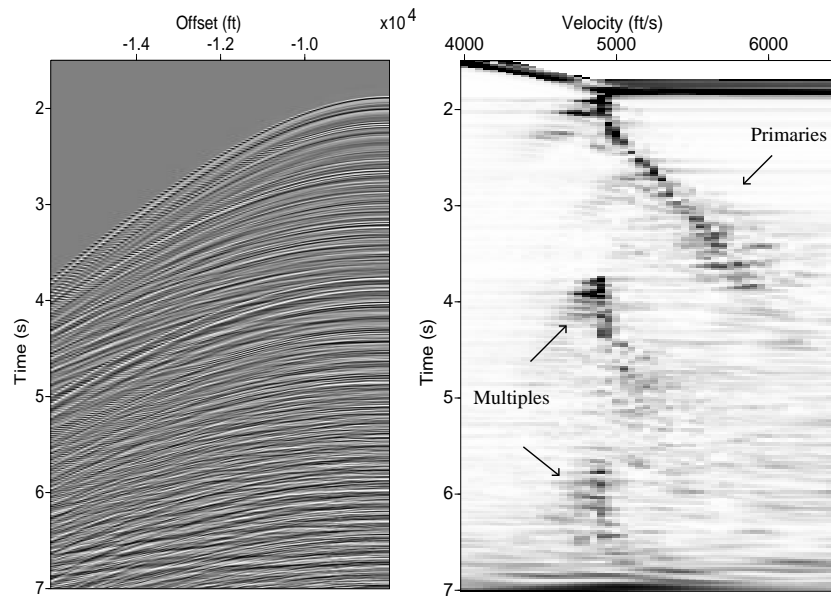


Figure 2.4: A CMP gather and its velocity spectrum.

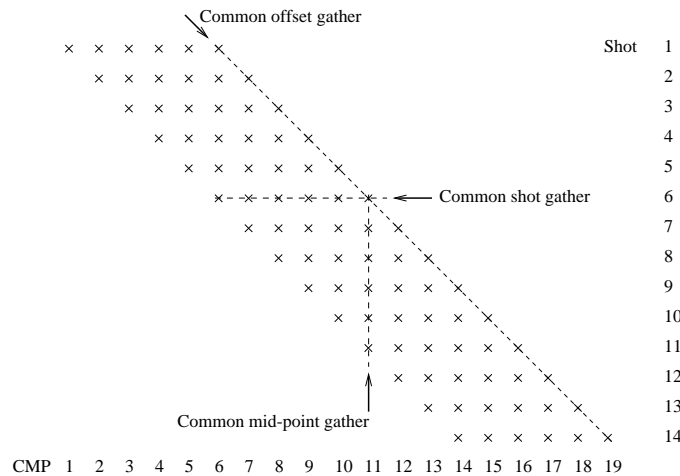


Figure 2.5: Stacking chart showing a field geometry. The right column indicates the shot numbers. The symbol “x” stands for shot points.

## 2.3 Radon transforms and multiple suppression

In the next section, two time-invariant Radon transforms will be described: the linear Radon transform or  $\tau - p$  transform and the parabolic Radon transform. In addition, we will present a third transform, the Hyperbolic Radon transform. The latter is a time-variant transform. We will also present an algorithm to compute time-invariant transforms in the frequency-space domain.

### 2.3.1 Linear Radon ( $\tau - p$ ) transform

Using the linear Radon transform a wavefront with linear moveout in  $t - h$  can be mapped to a point in the  $\tau - p$  domain. This process entails summation along lines parametrized with the parameter  $p$  (slope or ray parameter). The linear Radon or  $\tau - p$  transform is given by the following expression

$$\tilde{v}(\tau, p) = \int_{-\infty}^{\infty} d(t = \tau + ph, h)dh, \quad (2.3.1)$$

where  $d(t, h)$  denotes the seismic signal,  $t$  is time, and  $h$  is range of offset. It is clear that this is an integration along lines defined by the formula  $t = \tau + ph$ . The transformed data, also called the  $\tau - p$  panel, is denoted by  $\tilde{v}(\tau, p)$ . A transform like this can be used to isolate linear events in  $t - h$ . However, it is important to stress that reflections in CMP gathers are described via hyperbolic events and therefore, a linear Radon transform will not be capable of collapsing hyperbolic events into clusters in the  $\tau - p$  panel. The linear Radon transform can be utilized to remove ground roll in common shot gathers (Yilmaz, 1989).

A remaining problem is to define an inverse transform, this is a transform that allows us to map  $\tilde{v}(\tau, p)$  to the original  $t - h$  space. For that purpose we define a forward Radon transform of the form



$$d(t, h) = \int_{-\infty}^{\infty} v(\tau = t - ph, p) dp, \quad (2.3.2)$$

this is a transform that maps a point in  $\tau - p$  domain into a linear event in  $t - h$ . At this time it is important to stress that equations (2.3.1) and (2.3.2) constitute a forward-adjoint pair. In other words, if equation (2.3.2) is a forward operator, equation (2.3.1) is its adjoint. Some transforms, e.g., the Fourier transform, are unitary transforms, that is the adjoint of the transformation is also the inverse transform. This is not true for the Radon transform. In this case a procedure to find an inversion formula is required. We will come back to this point in section 2.3.4.

### 2.3.2 Parabolic Radon transform

One can interchange summation along lines by summation along parabolic events,

$$\tilde{v}(\tau, q) = \int_{-\infty}^{\infty} d(t = \tau + qh^2, h) dh, \quad (2.3.3)$$

where  $h$  denotes offset, we will see that a transform of this type can be used to remove multiple reflections. Similarly, one can obtain a transform to map data in  $\tau - q$  to  $t - h$ ,

$$d(t, h) = \int_{-\infty}^{\infty} v(\tau = t - qh^2, q) dq, \quad (2.3.4)$$

Again, it is important to stress that equations (2.3.3) and (2.3.4) are not an inverse pair; they constitute a forward/adjoint pair.

### 2.3.3 Hyperbolic Radon Transform

The hyperbolic Radon transform is defined in terms of the following expressions:

$$\tilde{u}(v, \tau) = \int_{-\infty}^{\infty} d(h, t = \sqrt{\tau^2 + h^2/v^2}) dh, \quad (2.3.5)$$

$$d(h, t) = \int_{-\infty}^{\infty} u(v, \tau = \sqrt{t^2 - h^2/v^2}) dv, \quad (2.3.6)$$

where  $t$  is the two-way travel-time,  $\tau$  represents the two-way zero offset time,  $h$  denotes offsets, and  $v$  is the NMO velocity. This transform was originally proposed by Thorson and Claerbout (1985) to improve the S/N of CMP gathers. It is clear that the Hyperbolic Radon transform is the kind of operator that one needs to process reflections in CMP gathers. In this case, the travel-time of the reflections (hyperbolic travel-times) are matched by the integration curve of the transformation. Unfortunately, the hyperbolic Radon transform is a time-variant operator, and consequently, a fast implementation of the forward/adjoint operators in the frequency-space domain is not possible. We will come back to this point during the description of the frequency domain implementation of the parabolic Radon transform.

### 2.3.4 Frequency domain algorithms

In this section we will present an algorithm to compute the linear and parabolic Radon transforms in the frequency domain. I will derive the algorithm for the parabolic Radon transform, but bear in mind that the same procedure is applicable to the linear Radon transform.

We first define the discrete adjoint and forward parabolic Radon operators:

$$\tilde{u}(\tau, q_j) = \sum_{k=1}^N d(t = \tau + q_j h_k^2, h_k), \quad (2.3.7)$$

$$d(t, h_k) = \sum_{j=1}^M u(\tau = t - q_j h_k^2, q_j). \quad (2.3.8)$$

By applying the Fourier transform to equation (2.3.7) with respect to  $t$ , we get

$$\tilde{U}(\omega, q_j) = \sum_{k=1}^N D(\omega, h_k) e^{i\omega h_k^2 q_j^2}. \quad (2.3.9)$$

Similarly, we can apply the Fourier transform to both sides of equation (2.3.8) to obtain the following expression

$$D(\omega, h_k) = \sum_{j=1}^M U(\omega, q_j) e^{-i\omega h_k^2 q_j^2}, \quad (2.3.10)$$

Equations (2.3.9) and (2.3.10) can be written in matrix form as follows:

$$\tilde{\mathbf{U}}(\omega) = \mathbf{L}^H(\omega) \mathbf{D}(\omega). \quad (2.3.11)$$

$$\mathbf{D}(\omega) = \mathbf{L}(\omega) \mathbf{U}(\omega). \quad (2.3.12)$$

where  $\tilde{\mathbf{U}}(\omega)$  and  $\mathbf{U}(\omega)$  are  $M \times 1$  vectors containing the elements  $\tilde{u}(\omega, q_j)$  and  $u(\omega, q_j)$   $j = 1 : M$ , respectively. The  $N \times 1$  vector  $\mathbf{D}(\omega)$  contains the elements  $d(\omega, h_k)$ ,  $k = 1 : N$ . The matrix  $\mathbf{L}(\omega)$  is given by

$$\mathbf{L}(\omega) = \begin{pmatrix} e^{-i\omega h_1^2 q_1} & e^{-i\omega h_1^2 q_2} & \dots & e^{-i\omega h_1^2 q_M} \\ e^{-i\omega h_2^2 q_1} & e^{-i\omega h_2^2 q_2} & \dots & e^{-i\omega h_2^2 q_M} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i\omega h_N^2 q_1} & e^{-i\omega h_N^2 q_2} & \dots & e^{-i\omega h_N^2 q_M} \end{pmatrix}, \quad (2.3.13)$$

The elements of the matrix  $\mathbf{L}$  depend on the geometry of the input data gather and the range and sampling of the parameter  $q$ . At this point it is important to mention that the adjoint operator (2.3.11) and the forward operator (2.3.12) do not constitute an inverse pair. To compute the inverse operator we can define two alternative pairs of transforms.

### Radon pair I

In this case we first assume that we have the data in the frequency domain. In order to avoid notational clutter we also define  $\mathbf{D} = \mathbf{D}(\omega)$ . One can use the operator  $\mathbf{L}^H$  to compute  $\tilde{\mathbf{U}}$  using equation (2.3.11)

$$\tilde{\mathbf{U}} = \mathbf{L}^H \mathbf{D}. \quad (2.3.14)$$

Similarly, we can use the operator  $\mathbf{L}$  to map back  $\tilde{\mathbf{U}}$  to data space,

$$\mathbf{L} \tilde{\mathbf{U}} = \mathbf{L} \mathbf{L}^H \mathbf{D}. \quad (2.3.15)$$

It is clear that if one wants to recover  $\mathbf{D}$  the Hermitian form  $\mathbf{L} \mathbf{L}^H$  needs to be inverted.

$$\mathbf{D} = (\mathbf{L} \mathbf{L}^H)^{-1} \mathbf{L} \tilde{\mathbf{U}}. \quad (2.3.16)$$

To summarize, we have found an operator to compute the Radon transform and its inversion formula,

$$\begin{cases} \tilde{\mathbf{U}} = \mathbf{L}^H \mathbf{D} \\ \mathbf{D} = (\mathbf{L} \mathbf{L}^H)^{-1} \mathbf{L} \tilde{\mathbf{U}} \end{cases} \quad (2.3.17)$$

The last two equations constitute a forward/inverse pair. Notice that we have not discussed the problem of invertibility of  $(\mathbf{L}\mathbf{L}^H)$ , we have assumed that the inverse exists. If the inverse does not exist we can always use the pseudo-inverse (Strang, 1986). In this case, the inversion formula provides an approximation of the form

$$\mathbf{D} \approx (\mathbf{L}\mathbf{L}^H)^\dagger \mathbf{L} \tilde{\mathbf{U}}. \quad (2.3.18)$$

where in the last equation the symbol  $\dagger$  is used to indicate the pseudo-inverse.

### Radon pair II

We can think that the data are obtained as the result of a transformation  $\mathbf{L}$  operating on  $\mathbf{U}$ ,

$$\mathbf{D} = \mathbf{L} \mathbf{U}. \quad (2.3.19)$$

in this case,  $\mathbf{U}$  becomes the unknown of our problem. We can design a Radon forward operator by treating last equation as an inverse problem: given  $\mathbf{D}$ , find  $\mathbf{U}$ . If the problem is under-determined ( $M > N$ ) one solution to the last equation is given by the minimum norm solution. This is the solution  $\mathbf{U}$  that among all possible solutions is the one with minimum norm. The minimum norm solution is given by

$$\mathbf{U}_{min\ norm} = \mathbf{L} (\mathbf{L}\mathbf{L}^H)^{-1} \mathbf{D}, \quad (2.3.20)$$

in this case, the transform pair is given by

$$\begin{cases} \mathbf{U}_{min\ norm} = \mathbf{L} (\mathbf{L}\mathbf{L}^H)^{-1} \mathbf{D}, \\ \mathbf{D} = \mathbf{L} \mathbf{U}. \end{cases} \quad (2.3.21)$$

When the problem is over-over-determined ( $N > M$ ), the inverse operator can be obtained by solving a least-squares problem. In this case we minimize the following error function:

$$J = (\mathbf{D} - \mathbf{L}\mathbf{U})^T(\mathbf{D} - \mathbf{L}\mathbf{U}). \quad (2.3.22)$$

Minimization of  $J$  with respect to  $\mathbf{u}$  yields the desired least squares solution

$$\mathbf{U}_{ls} = (\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T\mathbf{D}, \quad (2.3.23)$$

where  $(\mathbf{L}^T\mathbf{L})^{-1}\mathbf{L}^T$  is the least squares inverse of  $\mathbf{L}$ . To avoid singularities or near-singularities in the matrix  $\mathbf{L}^T\mathbf{L}$ , the solution is constrained by incorporating Marquardt's damping factor into equation (2.3.23)

$$\mathbf{U}_{ls} = (\mathbf{L}^T\mathbf{L} + \beta\mathbf{I})^{-1}\mathbf{L}^T\mathbf{D}. \quad (2.3.24)$$

In this case the pair of transforms is given by:

$$\begin{cases} \mathbf{U}_{ls} = (\mathbf{L}^T\mathbf{L} + \beta\mathbf{I})^{-1}\mathbf{L}^T\mathbf{D}, \\ \mathbf{D} \approx \mathbf{L}\mathbf{U}. \end{cases} \quad (2.3.25)$$

So far, we have presented the forward, adjoint, and inverse parabolic Radon transforms in the frequency domain. We have also defined two different approaches to obtain the forward/inverse pair (type I and II). It is important to clarify, however, that in general we will use the parabolic Radon pair of type II. The advantage of this approach is that the Radon panel can be obtained as the solution of an inverse problem and, therefore, one can include constraints and/or penalty terms in the formulation of the problem. This is important at the time

of designing a Radon transform capable of focusing seismic events with similar moveout (high resolution Radon transform).

The frequency domain algorithm to solve the Radon transform of type II is extremely fast. First, one inversion is needed at each frequency in the band that contains the seismic signal. In addition, the matrix  $\mathbf{L}^H \mathbf{L}$  is a Toeplitz form (Kostov, 1990; Darche, 1990), therefore, fast inversion algorithms like Levinson's recursion (Marple, 1987) can be used to compute its inverse. This type of algorithm can solve the system of normal equation (2.3.25) in a number of operations that is proportional to  $M^2$ , this is an important improvement with respect to standard methods for a general Hermitian matrix with a cost proportional to  $M^3$  (i.e., Cholesky decomposition).

### 2.3.5 Validity of the Parabolic Radon transform to process seismic data

We have already mentioned that seismic reflections (primaries and multiples) are organized in a CMP gather according to seismic events that follows hyperbolic moveout curves. It is clear that a transformation that involves integration along parabolic trajectories will not provide an optimal solution to our problem. Rather than using a time-variant hyperbolic Radon transform, Hampson (1986) proposed to modify the input data in order to transform hyperbolic events into parabolic events. This is achieved by applying a NMO correction to the CMP gather using a velocity law that is close to the velocity of the primaries. The residual moveout of the reflections after NMO can be approximated by a parabola. Another method to transform hyperbolic events into parabolic events is the  $t^2$  stretching proposed by Yilmaz (1989).

In the following two points we describe the techniques proposed by Hampson (1986) and Yilmaz (1989) to pre-process the seismic data before applying the

parabolic Radon transform.

### Parabolic Radon Transform after $t^2$ (Yilmaz 1989)

In CMP gathers, the hyperbolic events are defined by

$$t^2 = \tau^2 + \frac{h^2}{V^2}. \quad (2.3.26)$$

If we redefine the time axis by using the following transformation

$$t' = t^2, \quad \tau' = \tau^2,$$

equation (2.3.26) then takes the form

$$t = \tau + \frac{h^2}{V^2}. \quad (2.3.27)$$

Letting  $q = \frac{1}{V}$  we end up with the expression of a parabola in stretched time coordinates.

### Parabolic residual moveout (Hampson, 1986)

If one applies NMO correction to the CMP gather with a velocity  $V_{nmo}$  to a reflection parametrized with a velocity  $V$ , then the travel-time curve of the reflection after NMO becomes

$$t = t_0 + \sqrt{t_0^2 + \frac{h^2}{V^2}} - \sqrt{t_0^2 - \frac{h^2}{V_{nmo}^2}}. \quad (2.3.28)$$

After expanding equation (2.3.28) in Taylor series and keeping terms of order  $h^2$ , one has

$$t = t_0 + \frac{h^2}{2t_0} \left( \frac{1}{V^2} - \frac{1}{V_{nmo}^2} \right). \quad (2.3.29)$$



Letting  $q = \frac{1}{2t_0}(\frac{1}{V^2} - \frac{1}{V_{nmo}^2})$  we have, again, an expression for a parabola. Sometimes, it is more convenient to parameterize the parabola in terms of residual moveout at far offset, in this case we can write

$$t = t_0 + q \frac{h^2}{h_{max}^2}, \quad (2.3.30)$$

now the parameter  $q$  has units of time and it simply defines the time difference of the parabola between  $h = 0$  and  $h = h_{max}$ .

It is important to stress that we often prefer to use the NMO correction with an approximate velocity law rather than the  $t^2$  stretching. In fact, the  $t^2$  transformation causes compression of data before 1 second and stretching after 1 second. Therefore there can be a potential problem of aliasing near  $t = 0$ , causing frequency distortion for shallow events. This can be avoided by using a finer sampling along the  $t^2$  axis (Yilmaz, 1989).

The range and sampling interval of the parameter  $q$  is quite important. In general,  $q$  should span the velocities associated with primary and multiple reflections. A practical choice of the  $q$  sampling increment is such that the number of  $q$  parameters is set equal to the number of traces in offset space.

### 2.3.6 High resolution parabolic Radon transform

The parabolic Radon transform is widely accepted technique for multiple suppression. The technique can be implemented in the frequency domain via a fast algorithm that exploits the Toeplitz structure of the least squares Radon operator (Kostov, 1990; Darche, 1990). Unfortunately, the resolution of the least-squares Parabolic Radon transform is not often adequate to properly distinguish events with similar moveout curves (i.e., short period multiples). In order to overcome this problem, Sacchi and Ulrych (1995) proposed a high resolution algorithm to

increase the ability of the transform to distinguish events with similar move-out. This algorithm is based on a procedure that attempts to find a sparse representation of the reflections in the parabolic Radon domain.

The high resolution parabolic Radon transform can be used to isolate multiples within a few milliseconds of residual move-out at far offsets. This is a problem frequently encountered when dealing with short period multiple reflections in the Western Canadian Basin (Hunt et al., 1996).

One of the advantages of the high resolution parabolic Radon transform is that the focusing power of the transform is considerably increased with respect to the classical least squares parabolic Radon transform. Unfortunately, the high resolution parabolic Radon transform leads to the inversion of an operator that is Hermitian but does not exhibit a Toeplitz structure. The resulting Hermitian operator can be inverted using Cholesky decomposition, in this case the computational cost of the algorithm can be prohibitively high. Sacchi and Porsani (1999) proposed a method to invert the Hermitian operator via the method of Conjugate Gradients with the addition of a fast matrix times vector multiplication using circulant matrix.

Figure 2.6 and 2.7 compare the inversion results from least squares and high resolution parabolic Radon transform. Figure 2.6(a) shows the synthetic data, and 2.6(b) shows the inverted model by least squares. Figures 2.6(c) and 2.6(d) show the inverted data and difference between the inverted data and original data respectively. Figure 2.7 gives the results from high resolution parabolic Radon transforms.

The fast high resolution parabolic Radon transform operates at a speed that is comparable to the least squares Radon operator obtained by the Levinson recursion. The efficiency of the algorithm is improved by an order of magnitude with respect to the original algorithm.

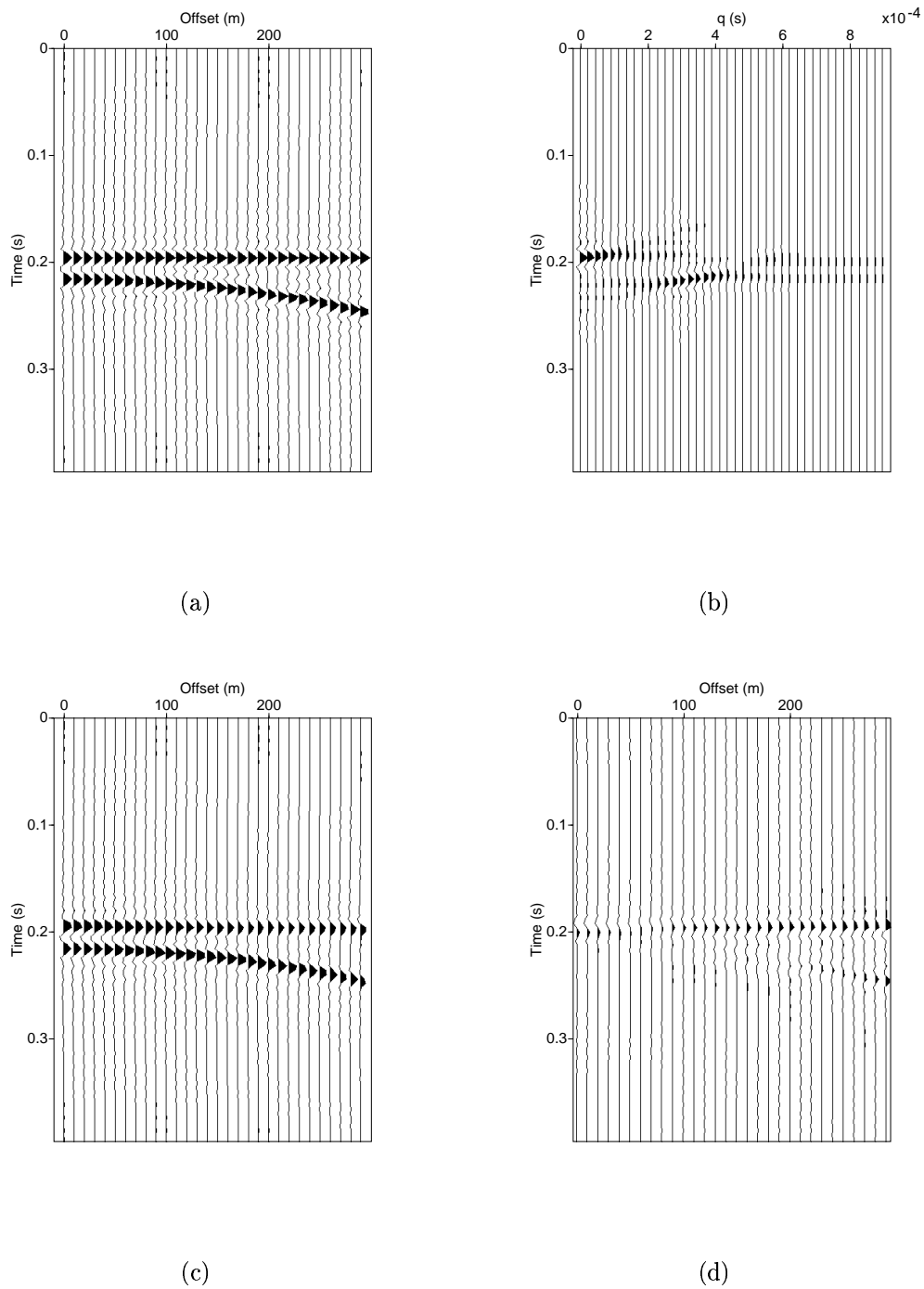


Figure 2.6: (a) Synthetic data. (b) Radon panel of the data obtained by least squares. (c) Reconstructed data from Figure 2.6(b). (d) The difference between original data (Figure 2.6(a)) and inverted data (Figure 2.6(c)).

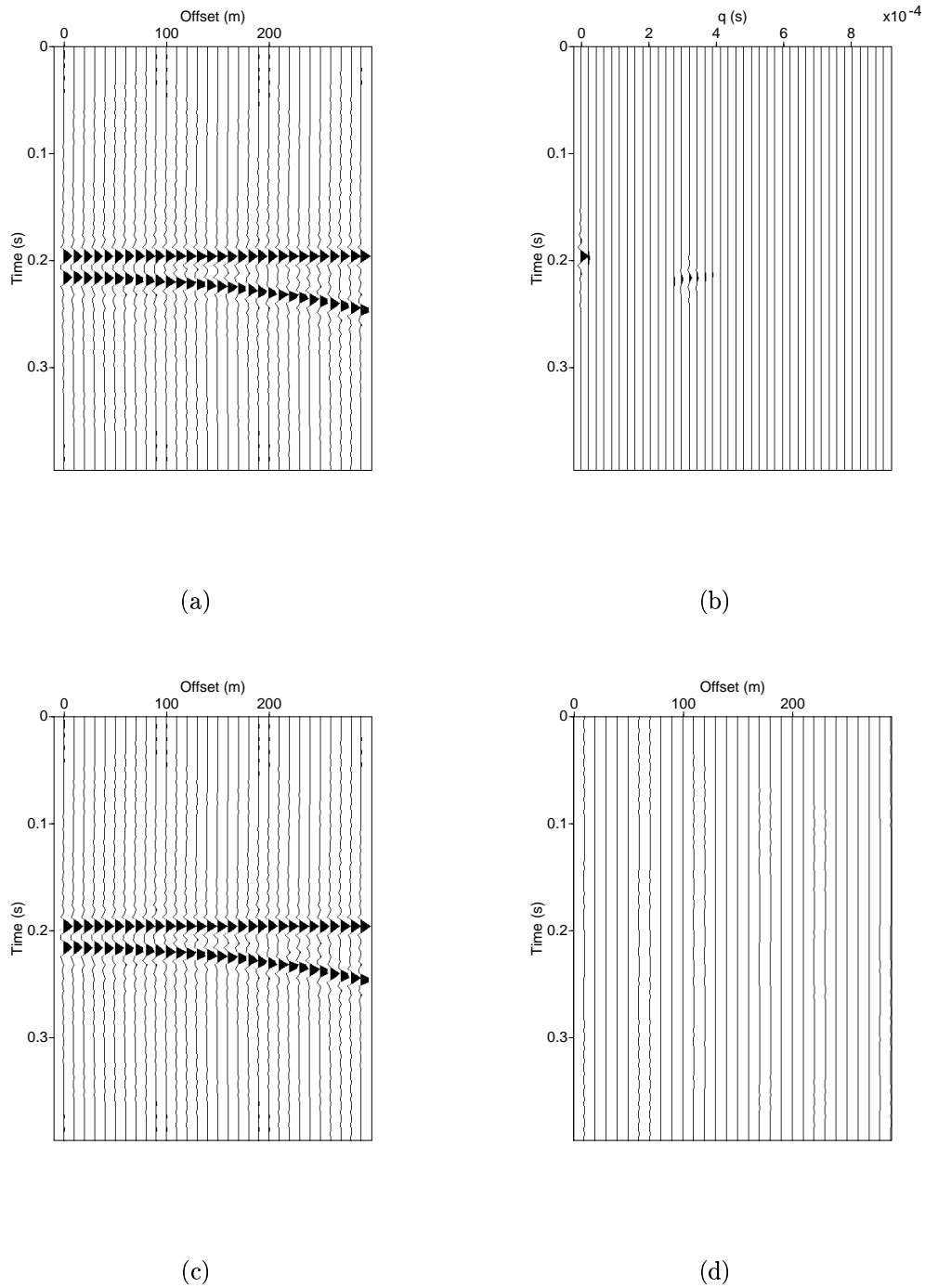


Figure 2.7: (a) Synthetic data. (b) Radon panel obtained by high resolution parabolic Radon transforms. (c) Reconstructed data from Figure 2.7(b). (d) The difference between original (Figure 2.7(a)) and inverted data (Figure 2.7(c)).

## 2.4 Velocity discrimination in the $f - k$ domain (Ryu, 1982)

If we apply a 2-D fast Fourier transform (FFT) to the CMP data with multiples, the primary and multiple energy can be separated into two quadrants in the  $f-k$  (frequency-wavenumber) domain. This is achieved by NMO correcting the gather using a velocity function that is between the primary and the multiple velocities. The multiples are under-corrected, while the primaries are over-corrected. On the  $f-k$  plane, the multiples and the primaries are mapped into two different quadrants. The exception to this separation is the near-offset energy (both primaries and multiples), that almost entirely maps along the frequency axis. This is because multiples and primaries have no significant move-out difference at near offsets.

Multiples can be suppressed by zeroing the quadrant corresponding to multiple energy in the  $f-k$  domain. After applying an inverse NMO correction (using the same intermediate velocity function) one can restore the original move-out of the primary reflections.

## 2.5 Velocity discrimination based on the periodicity of multiples

The slant stack multiple suppression technique, which is based on predictive criterion is discussed here. Alam and Austin (1981) and Treitel et al. (1982) investigated predictive deconvolution application in the slant stack domain for multiple suppression, this has been used in the case of vertical incidence and zero offset data (Yilmaz, 1987). However, multiples are not predicted in time for a given nonzero offset. The time separation between the multiple arrivals at a

particular offset  $h$  are only equal if  $h = 0$ .

Taner(1980) first recognized that the time separation between arrivals is equal along a radial direction. A trace can be constructed by pulling out the samples along one of these radial directions. This trace, along which the angle of propagation is constant, is called a radial trace. A radial trace in a layered medium is called a Snell trace (Claerbout, 1985). In a layered medium, the Snell trace would not follow a straight path, since its angle of propagation changes at layer boundaries according to Snell's Law.

Taner (1980) applied predictive deconvolution along radial traces to successfully eliminate long-period multiples. The magnitude of the time separations between multiples is different from one radial trace to another. However, the time separations are equal along each of the slanted paths of summation. Therefore, a predictive deconvolution operator can be designed from the autocorrelogram of each  $p$  trace and applied to suppress multiples.

Since slant stack is a plane-wave decomposition, and since plane waves do not have spherical divergence, input to the slant stack must not be compensated for by geometric spreading. Preserving correct amplitude relationships is essential for the effectiveness of slant-stack multiple suppression. The geometric spreading correction is applied to offset data by using a primary velocity function. This enhances the multiples in the data and destroys the amplitude ratio between them. Predictive deconvolution in the offset domain would then not suppress these multiples effectively.

## 2.6 Summary

In this chapter we have focused our discussion on the parabolic Radon transform as a tool to discriminate primaries from multiples in CMP gathers. While the technique seems to have a good conceptual basis, its performance on field data

can be disappointing. There are several possible explanations for this. First, for velocity discrimination techniques to be effective, significant move-out differences must exist between the primaries and multiples. However, the inability to exploit the large move-out between the primaries and multiples in the mute zone works against the methods based on velocity discrimination. There is also a problem caused by the application of geometric spreading correction, which is applied using the primary velocity function. This type of correction usually results in enhancement of multiple amplitudes.

One of the advantages of using parabolic Radon transforms is the possibility of using fast algorithms in the frequency domain to invert the operator  $\mathbf{L}^T \mathbf{L}$ . This makes the parabolic Radon transform a very effective method to process marine data sets where we usually need to process thousands of CMP gathers.

## Chapter 3

# Multiple suppression with the hyperbolic Radon transform

### 3.1 Introduction

An important step in normal move-out (NMO) and stacking for seismic reflection data involves the summation of data over paths represented by a family of hyperbolic events. This summation process is a linear transform and maps the data into velocity space. Examination of data in the velocity space is used for analysis of subsurface velocities and filtering of undesired coherent events, in our case multiples, but the filtering step is useful only if an approximate inverse to the NMO and stack operation is available. One method of effective velocity filtering is to use the operator  $\mathbf{L}^T$  (defined for NMO and stacking) and its adjoint  $\mathbf{L}$  as a transform pair, but this leads to unacceptable filtered output. Designing a better estimated inverse to  $\mathbf{L}$  than  $\mathbf{L}^T$  is a generalization of the inversion problem. The inversion process is complicated by missing data because surface seismic data are recorded only within a finite spatial aperture on the Earth surface.



## 3.2 The hyperbolic Radon transform and velocity stacks

The hyperbolic Radon transform involves mapping seismic reflections along hyperbolic trajectories in the time-offset domain to points in the time-velocity (or slowness) domain. Then the conjugate (or transpose) transform maps the points in the time-velocity domain to hyperbolic trajectories in the time-offset domain. For conjugate mapping, the hyperbolic Radon transform is similar to velocity analysis, which converts seismic traces in the time domain to velocity clusters in the time-velocity domain. However, the conjugate hyperbolic Radon transform can map points in velocity space back to the time-offset domain.

### 3.2.1 Forward and conjugate mapping

For typical seismic data, a CMP gather consists of hyperbolic events that represent reflections from the subsurface. According to the Dix equation formula (see discussion in Section 2.2.1), the reflections in a CMP gather can be written as

$$t^2 = \tau^2 + \frac{h^2}{v^2}, \quad (3.2.1)$$

where  $\tau$  is the two way travel-time at zero offset, and  $v$  is the RMS velocity. Equation (3.2.1) shows that the reflections in a CMP gather are hyperbolic. Next, let's look at the transformations that map the offset space into velocity space and vice-versa. Let  $d(h, t)$  be the data in the time-offset space and  $u(v, \tau)$  be the transformed data in the time-velocity space. The mapping from the offset space to the velocity space is achieved by summing over offset

$$\tilde{u}(v, \tau) = \int_{h_{min}}^{h_{max}} d(h, t = \sqrt{\tau^2 + h^2/v^2}) dh, \quad (3.2.2)$$

where  $u(v, \tau)$  represents the velocity gather. The variables  $h, v, t$  and  $\tau$  represent the offset, velocity, time and zero offset travel-time respectively. Similarly, the conjugate transformation (mapping from velocity space to offset space) involves summation along the velocity axis

$$d(h, t) = \int_{v_{min}}^{v_{max}} u(v, \tau = \sqrt{t^2 - h^2/v^2}) dv, \quad (3.2.3)$$

where  $d(h, t)$  is the reconstructed CMP gather. Discretizing the equations (3.2.2) and (3.2.3), one has

$$\tilde{u}(v, \tau) = \sum_h d(h, t = \sqrt{\tau^2 + h^2/v^2}), \quad (3.2.4)$$

and

$$d(h, t) = \sum_v u(v, \tau = \sqrt{t^2 - h^2/v^2}). \quad (3.2.5)$$

Figure 3.1 shows a CMP gather from the Gulf of Mexico data set, the offset starts from  $-15993$  ft ( $4874.7$  m), the space interval is  $87.5$  ft ( $26.7$  m). There are  $1751$  time samples, the time sampling interval is  $0.004$  s. Figure 3.2 shows the mapping of this CMP in velocity space by equation (3.2.4).

The result in velocity space shows artifacts arising from the limited aperture of the CMP gather. The horizontal lines crossing the velocity panel are the effect of near offset truncation. The reason for this effect will be shown later in section 3.2.3. Figure 3.3 shows the result of forward mapping. This is done by solving the inverse problem which is discussed in next section. From the figure one can not tell the difference between the original data and the inverted data.

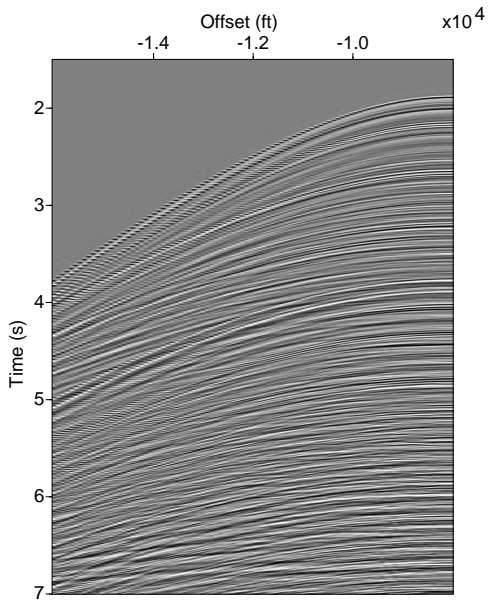


Figure 3.1: A CMP gather from the Gulf of Mexico data set.

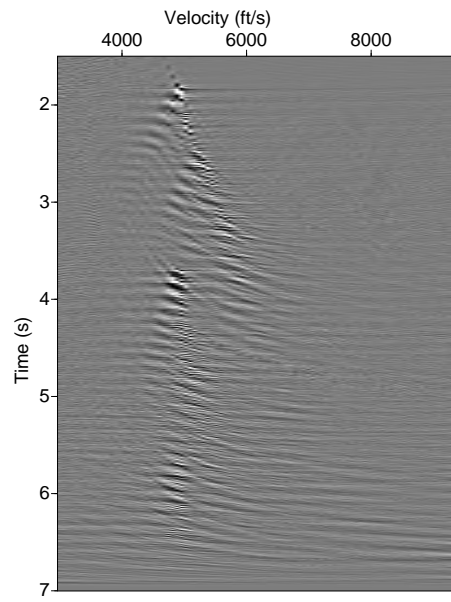


Figure 3.2: The velocity gather solved by least squares.

### 3.2.2 The inverse problem

Velocity estimation from a CMP gather can be regarded as linear inverse problems. Equation (3.2.5) can be written in matrix form as

$$\mathbf{d} = \mathbf{Lm}, \tag{3.2.6}$$

where  $\mathbf{d}$  is a column vector of size  $nt \times nh$ , and  $\mathbf{m}$  is a row vector of size  $nt \times nv$ . These panels (matrices) can be rewritten in lexicographic order, for example

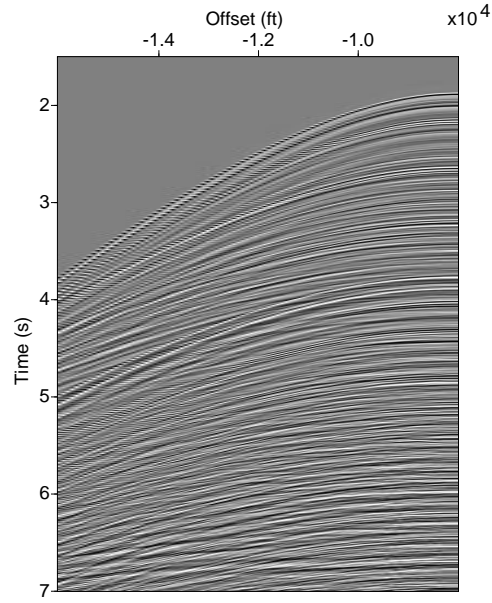


Figure 3.3: The data from the velocity gather shown in Figure 3.2.

$$\begin{bmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{m1} & d_{m2} & \cdots & d_{mn} \end{bmatrix} \rightarrow \begin{bmatrix} d_{11} \\ d_{21} \\ \vdots \\ d_{m1} \\ d_{12} \\ d_{22} \\ \vdots \\ d_{m2} \\ \vdots \\ d_{mn} \end{bmatrix}. \quad (3.2.7)$$

Thus the dimension of  $\mathbf{L}$  is  $(nt \times nh \times nt \times nv)$ . The operator  $\mathbf{L}$  is the time-variant forward mapping operator. Basically, the forward operator converts points in time-velocity space into hyperbolic events in the time-offset domain.

The transpose operator  $\mathbf{L}^T$  is a superposing operator. It sums the hyperbolic trajectories in the time-offset domain and puts the energy into corresponding points in velocity space. The transpose mapping can be written as

$$\tilde{\mathbf{m}} = \mathbf{L}^T \mathbf{d}. \quad (3.2.8)$$

In order to compute  $\mathbf{m}$  we need to solve a least-squares problem, in other words, we need to minimize the sum of the squares of the error vector given by

$$\mathbf{e} = \mathbf{L}\mathbf{m} - \mathbf{d}. \quad (3.2.9)$$

This is equivalent to minimize the following objective or cost function:

$$J = \mathbf{e}^T \mathbf{e} = (\mathbf{L}\mathbf{m} - \mathbf{d})^T (\mathbf{L}\mathbf{m} - \mathbf{d}). \quad (3.2.10)$$

The least squares solution can be written as

$$\mathbf{m}_{ls} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{d}. \quad (3.2.11)$$

For a small data set of 100 traces, 5 s of data with a 0.002 s sampling interval in time, the size of operator  $\mathbf{L}$  can go up to  $(250000 \times 250000)$ . A matrix of this size cannot be inverted by conventional methods. The best way is to employ an algorithm to solve it in an iterative manner. The method of Conjugate Gradients is a very powerful tool to solve large system of equations where the matrix does not need to be explicitly provided.

### 3.2.3 Near and far offset effects

Near and far offset artifacts due to limited aperture are the main source of resolution degradation in hyperbolic Radon transforms. In order to illustrate these

effects, one data point at the very near offset is used to get the corresponding velocity gather. Figure 3.4 shows the synthetic data in a CMP gather, Figure 3.5 shows the inverted velocity from Figure 3.4. The near offset effect looks like a horizontal event across the velocity panel.

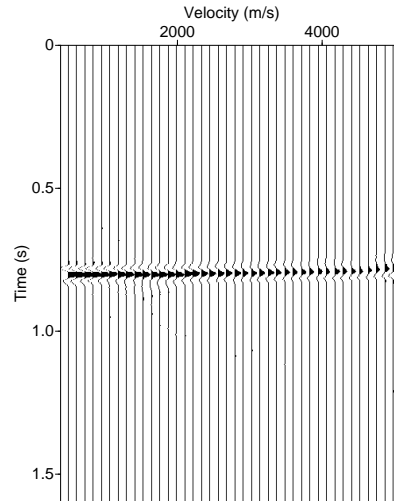
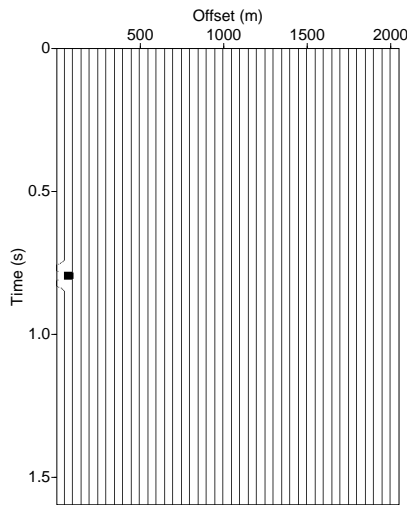


Figure 3.4: One data point at the near offset.

Figure 3.5: The forward mapping of Figure 3.4.

One can do the same thing to investigate far offset artifacts. A far offset data point is used to obtain the corresponding velocity panel (Figure 3.6 and 3.7). The velocity trajectory for this far offset point curves from the top to the right of the velocity panel.

Consequently, a hyperbolic event in the time offset domain is inverted to a point with horizontal and curvaceous artifacts. This is illustrated in Figures 3.8 and 3.9. To overcome the near and far offset artifacts, an offset proportional weighting function can be used. The function can be designed in many ways. One way is to allow far offset data to contribute more to the stack than the

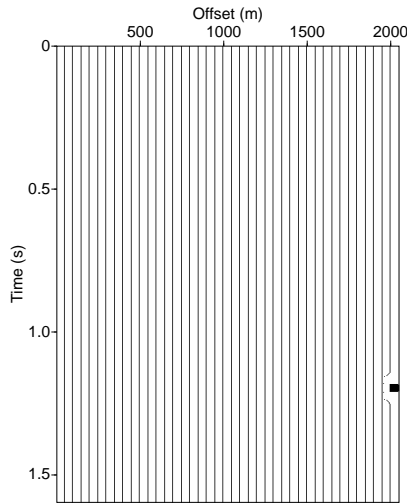


Figure 3.6: One data point at far off-set.

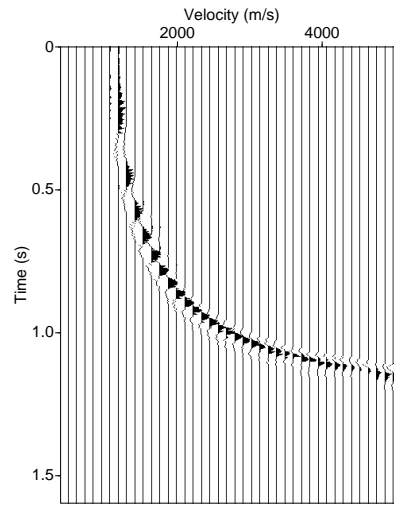


Figure 3.7: The inverted velocity gather from Figure 3.6.

near offset data (Thorson and Claerbout, 1985). For far offsets, tapering can effectively reduce the far offset artifacts (Kabir and Marfurt, 1999). However, in order to prevent damaging the data, the weighting function should be considered instead of tapering.

### 3.3 Inversion of hyperbolic Radon transform using the method of Conjugate Gradients

The hyperbolic Radon leads to the problem of solving the following system of equations

$$\mathbf{Lm} = \mathbf{d}, \tag{3.3.1}$$

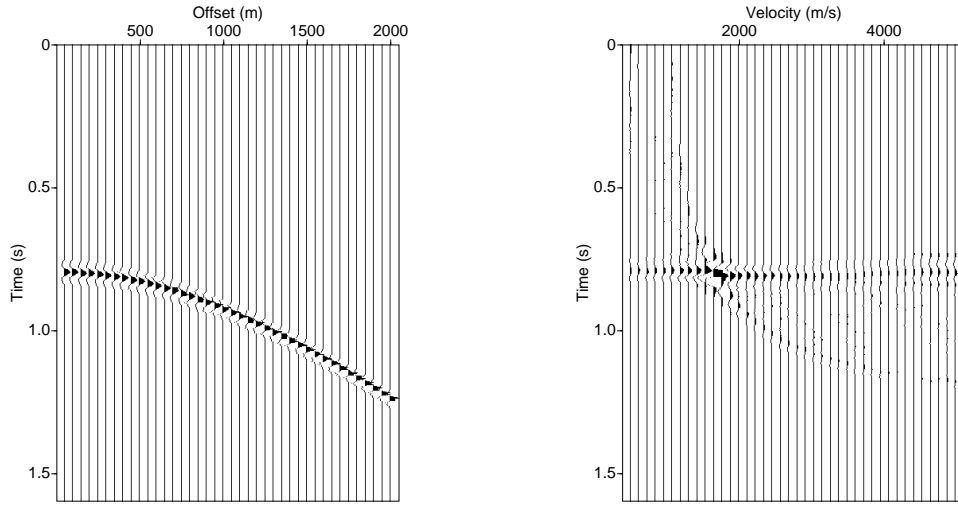


Figure 3.8: One hyperbolic event in Figure 3.9: The inverted velocity data space. gather.

the problem needs to be solved with an iterative algorithm like CG. The size of  $\mathbf{L}$  makes the problem intractable for non-iterative solvers. We have already seen that the system of normal equations that we need to solve has the following form:

$$\mathbf{L}^T \mathbf{L} \mathbf{m} = \mathbf{L}^T \mathbf{d}, \tag{3.3.2}$$

where the operator  $\mathbf{L}^T \mathbf{L}$  is interpreted as the application of two operators:  $\mathbf{L}$  and  $\mathbf{L}^T$ . If the right side is denoted  $\mathbf{b} = \mathbf{L}^T \mathbf{d}$ . Then we have a new problem of the form  $\mathbf{A} \mathbf{m} = \mathbf{b}$ , where  $\mathbf{A} = \mathbf{L}^T \mathbf{L}$

The method of CG is the most prominent method for solving sparse systems of linear equations. CG is effective for systems of the form



$$\mathbf{Ax} = \mathbf{b}, \quad (3.3.3)$$

where  $\mathbf{x}$  is an unknown vector,  $\mathbf{b}$  is a known vector, and  $\mathbf{A}$  is a known, square, symmetric, positive-definite matrix.

CG is suited for the use with sparse matrices. If  $\mathbf{A}$  is dense, the best course of action is probably to factor  $\mathbf{A}$  and solve the equation by back substitution. The time spent factoring a dense  $\mathbf{A}$  is roughly equivalent to the time spent solving the system iteratively; and once  $\mathbf{A}$  is factored, the system can be back-solved quickly for multiple values of  $\mathbf{b}$ . One can compare this dense matrix with a large sparse matrix that fills the same amount of memory. The triangular factors of a sparse  $\mathbf{A}$  usually have many more non-zero elements than  $\mathbf{A}$  itself. Factoring may be impossible due to limited memory, and will be time-consuming as well; even the back-solving step may be slower than the iterative solution. On the other hand, most iterative methods are memory-efficient and run quickly with sparse matrices.

Let  $\mathbf{A}$  be an  $n \times n$  matrix, and  $\mathbf{x}$  and  $\mathbf{b}$  are vectors, that is,  $n \times 1$  matrix,

$$\begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}. \quad (3.3.4)$$

The inner product of two vectors is written  $\mathbf{x}^T \mathbf{y}$ , and represents the scalar-sum  $\sum_{i=1}^n \mathbf{x}_i \mathbf{y}_i$ . Note that  $\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$ . If  $\mathbf{x}$  and  $\mathbf{y}$  are orthogonal, then  $\mathbf{x}^T \mathbf{y} = 0$ . In general, expressions that reduce to  $1 \times 1$  matrix, such as  $\mathbf{x}^T \mathbf{y}$  and  $\mathbf{x}^T \mathbf{Ax}$ , are treated as scalar values.

A matrix  $\mathbf{A}$  is positive-definite, if for every nonzero vector  $\mathbf{x}$

$$\mathbf{x}^T \mathbf{A} \mathbf{x} > 0. \quad (3.3.5)$$

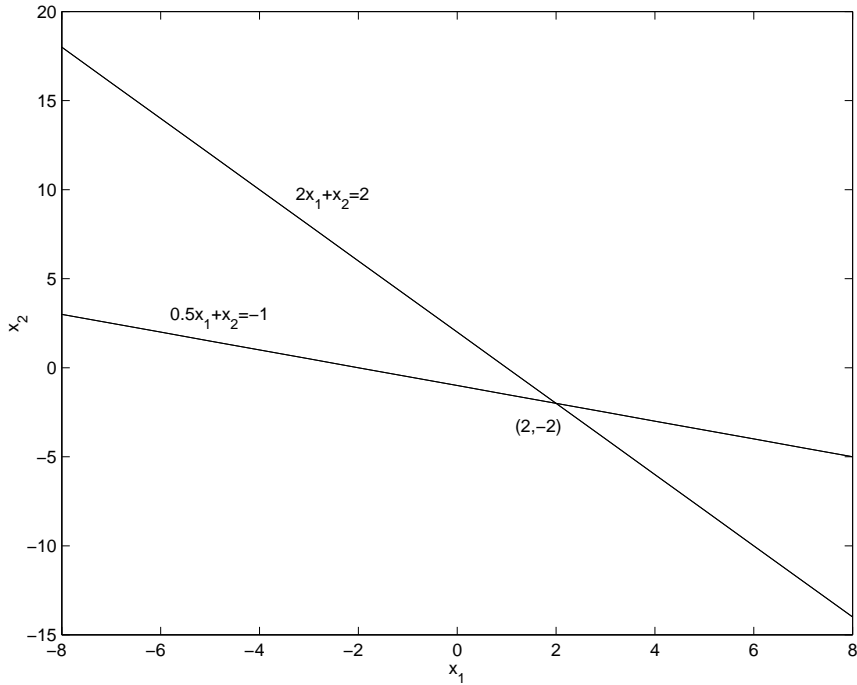


Figure 3.10: 2D linear system example, the solution lies at the intersection of the lines.

### 3.3.1 The Quadratic form

The quadratic form is simply a scalar, quadratic function of a vector with the form

$$f(\mathbf{x}) = \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x} + c, \quad (3.3.6)$$

where  $\mathbf{A}$  is a matrix,  $\mathbf{x}$  and  $\mathbf{b}$  are vectors, and  $c$  is a scalar constant. If  $\mathbf{A}$  is symmetric and positive definite,  $f(\mathbf{x})$  is minimized by the solution to  $\mathbf{A} \mathbf{x} = \mathbf{b}$ .

The ideas here will be demonstrated with the simple sample problem

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0.5 & 1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, c = 0. \quad (3.3.7)$$

The system  $\mathbf{Ax} = \mathbf{b}$  is shown in Figure 3.10. Generally the solution  $\mathbf{x}$  lies at the intersection point of  $n$  hyper planes, each has dimension  $n - 1$ . It is only possible to show the case when the size of  $\mathbf{A}$  is less or equal to  $2 \times 2$ . For this specific problem, the solution is  $\mathbf{x} = [2, -2]^T$ . The corresponding contour plot of  $f(\mathbf{x})$  is illustrated in Figure 3.11.

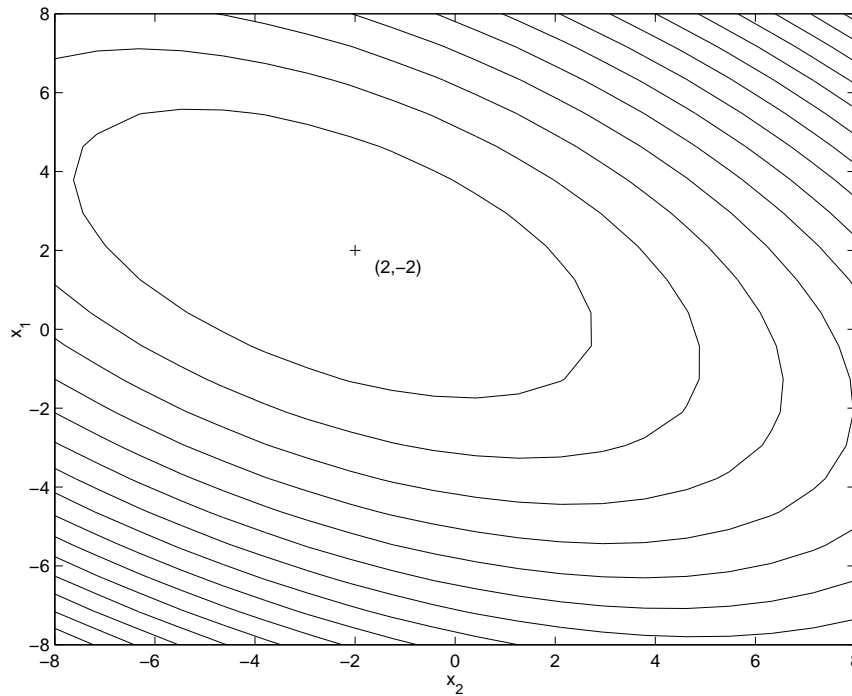


Figure 3.11: Contour plot of the quadratic form. Each contour has constant  $f(\mathbf{x})$ .

Because  $\mathbf{A}$  is positive-definite, the surface defined by  $f(\mathbf{x})$  is shaped like a

paraboloid bowl. The gradient of a quadratic form is defined to be

$$f'(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(\mathbf{x}) \\ \frac{\partial}{\partial x_2} f(\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n} f(\mathbf{x}) \end{bmatrix}. \quad (3.3.8)$$

The gradient is a vector field that, for a given point  $\mathbf{x}$ , points in the direction of greatest increase of  $f(\mathbf{x})$ . Figure 3.12 shows the gradient vectors for equation (3.3.8).

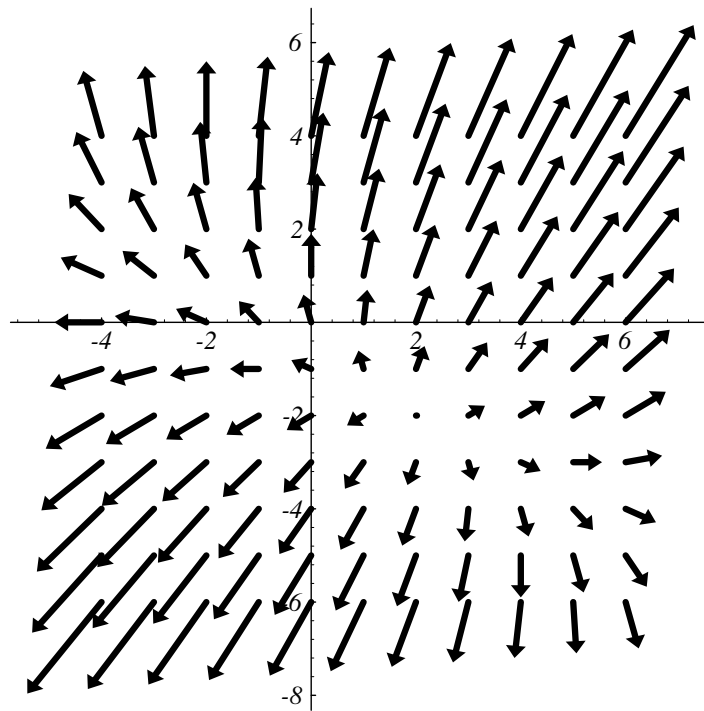


Figure 3.12: Gradient  $f'(\mathbf{x})$  of the quadratic form. For every  $\mathbf{x}$ , the gradient points in the direction of steepest increase of  $f(\mathbf{x})$ , and is orthogonal to the contour lines (After Shewchuk, 1994).

At the bottom of the paraboloid bowl, the gradient is zero. One can minimize  $f(\mathbf{x})$  by setting  $f'(\mathbf{x})$  equal to zero. Applying equation (3.3.6) to (3.3.8), one gets

$$f'(\mathbf{x}) = \frac{1}{2}\mathbf{A}^T\mathbf{x} + \frac{1}{2}\mathbf{A}\mathbf{x} - \mathbf{b}. \quad (3.3.9)$$

If  $\mathbf{A}$  is symmetric, this equation can be written as

$$f'(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}, \quad (3.3.10)$$

next, setting the gradient to zero, equation (3.3.3) is obtained, which the linear system needed to solve. Therefore, the solution to  $\mathbf{A}\mathbf{x} = \mathbf{b}$  is a critical point of  $f(\mathbf{x})$ . If  $\mathbf{A}$  is positive-definite as well as symmetric, then this solution is a minimum of  $f(\mathbf{x})$ , so  $\mathbf{A}\mathbf{x} = \mathbf{b}$  can be solved by finding an  $\mathbf{x}$  that minimizes  $f(\mathbf{x})$ . If  $\mathbf{A}$  is not symmetric, then equation (3.3.9) implies that CG will find a solution to the system  $\frac{1}{2}(\mathbf{A}^T + \mathbf{A})\mathbf{x} = \mathbf{b}$ , where  $\frac{1}{2}(\mathbf{A}^T + \mathbf{A})$  is symmetric.

Why do symmetric positive-definite matrices have this property? Consider the relationship between  $f$  at some arbitrary point  $p$  and at the solution point  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$ . From equation (3.3.6) one can show that

$$\begin{aligned} f(\mathbf{x} + \mathbf{e}) &= \frac{1}{2}(\mathbf{x} + \mathbf{e})^T \mathbf{A}(\mathbf{x} + \mathbf{e}) - \mathbf{b}^T(\mathbf{x} + \mathbf{e}) + c && \text{(by equation (3.3.6))} \\ &= \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{e}^T \mathbf{A}\mathbf{x} + \frac{1}{2}\mathbf{e}^T \mathbf{A}\mathbf{e} - \mathbf{b}^T \mathbf{x} - \mathbf{b}^T \mathbf{e} + c && \text{(by symmetry of } \mathbf{A} \text{)} \\ &= \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} - \mathbf{b}^T \mathbf{x} + c + \mathbf{e}^T \mathbf{b} + \frac{1}{2}\mathbf{e}^T \mathbf{A}\mathbf{e} - \mathbf{b}^T \mathbf{e} \\ &= f(\mathbf{x}) + \frac{1}{2}\mathbf{e}^T \mathbf{A}\mathbf{e}. \end{aligned} \quad (3.3.11)$$

If  $\mathbf{A}$  is positive-definite, then the latter term is positive for all  $\mathbf{e} \neq 0$ ; therefore  $\mathbf{x}$  minimizes  $f$ . But if  $\mathbf{A}$  is symmetric and not necessary positive-definite, one has

$$f(\mathbf{p}) = f(\mathbf{x}) + \frac{1}{2}(\mathbf{p} - \mathbf{x})^T \mathbf{A}(\mathbf{p} - \mathbf{x}). \quad (3.3.12)$$

If  $\mathbf{A}$  is positive-definite as well, then from inequality (3.3.5) the latter term is positive for all  $\mathbf{p} \neq \mathbf{x}$ . It follows that  $\mathbf{x}$  is a global minimum of  $f$ . The fact that  $f(\mathbf{x})$  is a paraboloid is our best intuition of what it means for a matrix to be positive-definite. If  $\mathbf{A}$  is not positive-definite, there are several other possibilities.  $\mathbf{A}$  could be negative-definite, the result of negating a positive-definite matrix, see Figure 3.13, but held upside down.  $\mathbf{A}$  might be singular, in which case no solution is unique; the set of solutions is a line or hyper plane having a uniform value for  $f$ . If  $\mathbf{A}$  is none of the above, then  $\mathbf{x}$  is a saddle point, and techniques like Steepest Descent and CG will likely fail. Figure 3.13 demonstrates the possibilities.

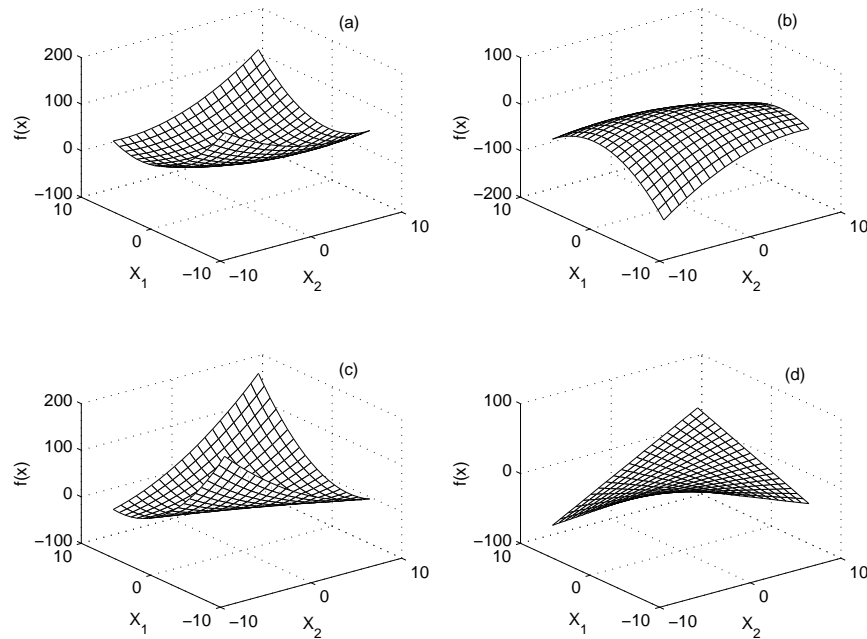


Figure 3.13: Contour plot of the quadratic form. Each contour has constant  $f(\mathbf{x})$ .

The values of  $\mathbf{b}$  and  $c$  determine where the minimum point of the paraboloid lies, but do not affect the paraboloid's shape.

### 3.3.2 The method of Steepest Descent

In the method of Steepest Descent, one starts at an arbitrary point  $\mathbf{x}_{(0)}$ , and then slides down to the bottom of the paraboloid. In other words, one takes a series of steps  $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots$  until one is close enough to the solution  $\mathbf{x}$ .

When taking a step, the direction is chosen in which  $f$  decreases most quickly, which is the direction opposite  $f'(\mathbf{x}_i)$ . According to equation (3.3.10), the direction is  $-f'(\mathbf{x}_{(i)}) = \mathbf{b} - \mathbf{A}\mathbf{x}_{(i)}$ .

It is important to introduce some definitions. The error  $\mathbf{e}_{(i)} = \mathbf{x}_{(i)} - \mathbf{x}$  is a vector that indicates how far one is from the solution and the residual  $\mathbf{r}_{(i)} = \mathbf{b} - \mathbf{A}\mathbf{x}_{(i)}$  indicates how far one is from the correct value of  $\mathbf{b}$ . It is easy to see that  $\mathbf{r}_{(i)} = -\mathbf{A}\mathbf{e}_{(i)}$ , and the residual can be thought of being the error transformed by  $\mathbf{A}$  into the same space  $\mathbf{b}$ . More importantly,  $\mathbf{r}_{(i)} = -f'(\mathbf{x}_{(i)})$ , and also the residual should be considered as the direction of steepest descent.

Suppose one starts at  $[-2, -2]^T$ . Following the steepest descent direction gives the solid line in Figure 3.14(a).

Figure 3.14(b) shows the intersection of the surfaces. Along the direction shown in Figure 3.14(a) choose a point

$$\mathbf{x}_{(1)} = \mathbf{x}_{(0)} + \alpha \mathbf{r}_{(0)}. \quad (3.3.13)$$

The purpose is to choose  $\alpha$  to minimize  $f$  along a line. Figure 3.14(c) shows the parabola defined by the intersection of the surfaces shown in Figure 3.14(b). The next step is to take a direction which is orthogonal to the gradient at the bottom of the hyperbola. The new gradient is also orthogonal to the gradient of the previous step, see Figure 3.14(d). The method of steepest descent is

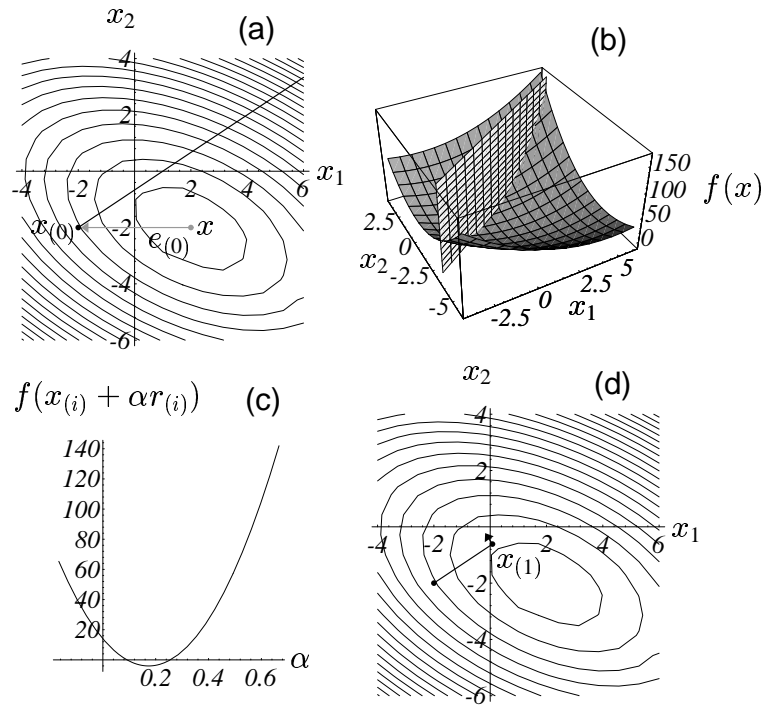


Figure 3.14: (a) From at  $[-2, -2]^T$ , take a step in the direction of steepest descent of  $f$ . (b) Find the point on the intersection of these two surfaces that minimizes  $f$ . (c) The parabolic intersection of the surfaces. (d) The gradient at the bottom point is orthogonal to that of the previous step (After Shewchuk, 1994).

summarized as follows

$$\mathbf{r}(i) = \mathbf{b} - \mathbf{A}\mathbf{x}(i), \quad (3.3.14)$$

$$\alpha(i) = \frac{\mathbf{r}(i)^T \mathbf{r}(i)}{\mathbf{r}(i)^T \mathbf{A} \mathbf{r}(i)}, \quad (3.3.15)$$

$$\mathbf{x}(i+1) = \mathbf{x}(i) + \alpha \mathbf{r}(i). \quad (3.3.16)$$

The solution-finding path for the steepest descent is shown in Figure 3.15. For each step, the gradient is orthogonal to that of the previous step.



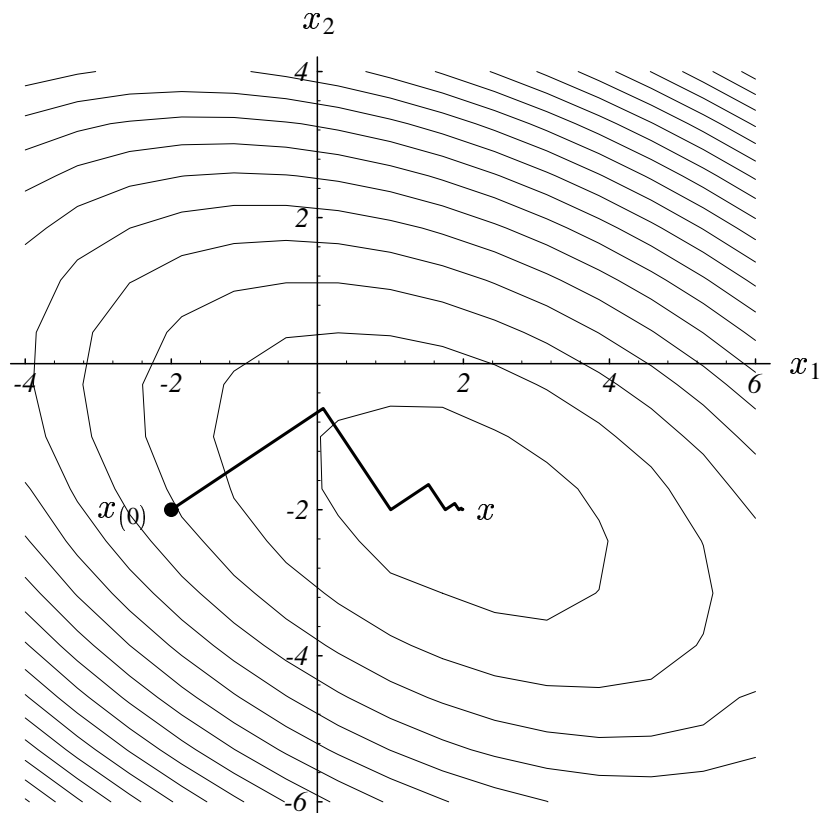


Figure 3.15: The solution seeking path of the steepest descent (After Shewchuk, 1994).

### 3.3.3 The method of Conjugate Gradients

The method of steepest descent often finds itself taking steps in the same direction as earlier steps (see Figure 3.15). It would be better if it got it right the first time, which is why the method of CG is used. Pick a set of orthogonal search directions  $\mathbf{d}_{(0)}, \mathbf{d}_{(1)}, \dots, \mathbf{d}_{(n-1)}$ . In each search direction take exactly one step, and make that step just long enough to reach  $\mathbf{x}$ . After  $n$  steps, the solution is reached.

The solution is to make the search directions  $\mathbf{A}$ -orthogonal instead of orthog-

onal. Two vectors  $\mathbf{d}_{(i)}$  and  $\mathbf{d}_{(j)}$  are  $\mathbf{A}$ -orthogonal, or conjugate, if

$$\mathbf{d}_{(i)}^T \mathbf{A} \mathbf{d}_{(j)} = 0. \quad (3.3.17)$$

Replacing the definition of  $\alpha_{(i)}$  in steepest descent by

$$\alpha_{(i)} = \frac{\mathbf{d}_{(i)}^T \mathbf{r}_{(i)}}{\mathbf{d}_{(i)}^T \mathbf{A} \mathbf{d}_{(i)}}, \quad (3.3.18)$$

the CG are defined as

$$\mathbf{d}_{(0)} = \mathbf{r}_{(0)} = \mathbf{b} - \mathbf{A} \mathbf{x}_{(0)}, \quad (3.3.19)$$

$$\alpha_{(i)} = \frac{\mathbf{r}_{(i)}^T \mathbf{r}_{(i)}}{\mathbf{d}_{(i)}^T \mathbf{A} \mathbf{d}_{(i)}}, \quad (3.3.20)$$

$$\mathbf{x}_{(i+1)} = \mathbf{x}_{(i)} + \alpha_{(i)} \mathbf{d}_{(i)}, \quad (3.3.21)$$

$$\mathbf{r}_{(i+1)} = \mathbf{r}_{(i)} - \alpha_{(i)} \mathbf{A} \mathbf{d}_{(i)}, \quad (3.3.22)$$

$$\beta_{(i+1)} = \frac{\mathbf{r}_{(i+1)}^T \mathbf{r}_{(i+1)}}{\mathbf{r}_{(i)}^T \mathbf{r}_{(i)}}, \quad (3.3.23)$$

$$\mathbf{d}_{(i+1)} = \mathbf{r}_{(i+1)} + \beta_{(i+1)} \mathbf{d}_{(i)}. \quad (3.3.24)$$

The later is the algorithm we have used to invert the hyperbolic Radon transform.

### 3.4 Examples

The goal for multiple suppression is to remove as many multiples as possible, leaving the primaries untouched. The de-multiple algorithm can be summarized as follows:

- Invert the hyperbolic Radon operator to compute the velocity panel  $\mathbf{m}$ .
- Filter the primaries in the velocity panel and retain the multiples.
- Map the multiple panel obtained in the previous step to the offset-time space using the operator  $\mathbf{L}$ . This is our estimate of the multiples.

- The multiple free data are then obtained by subtracting the estimated multiples from the original data.

Figure 3.16(a) shows the muting operator that I have designed to separate multiples from primaries in the velocity gather. Figure 3.16(b) shows the previous panel after muting. Figure 3.16(c) shows the multiples obtained by transforming back the velocity gather to data space using the operator  $\mathbf{L}$ . Figure 3.16(d) shows the data after de-multiple, in this case the multiple were subtracted from the original data.

To decide if the multiples have been successfully removed or not, a quality control measure is needed. The latter can be done by plotting the velocity spectrum after and before de-multiple. A velocity analysis program was used to compute the velocity spectrum (semblance) as discussed in section 2.2.7.

Figure 3.17 shows the velocity analysis for both the original data and the data after multiple removal. In Figure 3.17(d), one can see that the multiples have been properly removed, but the result is not perfect. There is still a trace of residual multiple energy leaking in the primaries.

### 3.5 Summary

In this chapter we discussed a numerical procedure to invert the time-variant hyperbolic Radon operator. Since this type of operators cannot be posed in the frequency domain a strategy to solve large systems of equations is required. In our research, we have found that the method of conjugate gradients provides an efficient manner to invert the hyperbolic Radon operator. One of the advantages of CG is that the matrix does not need to be defined. In our algorithms, the CG requires only to know how to compute inner products of the form  $\mathbf{L} \mathbf{x}$  and  $\mathbf{L}^T \mathbf{y}$  without knowing the actual structure of the matrix.

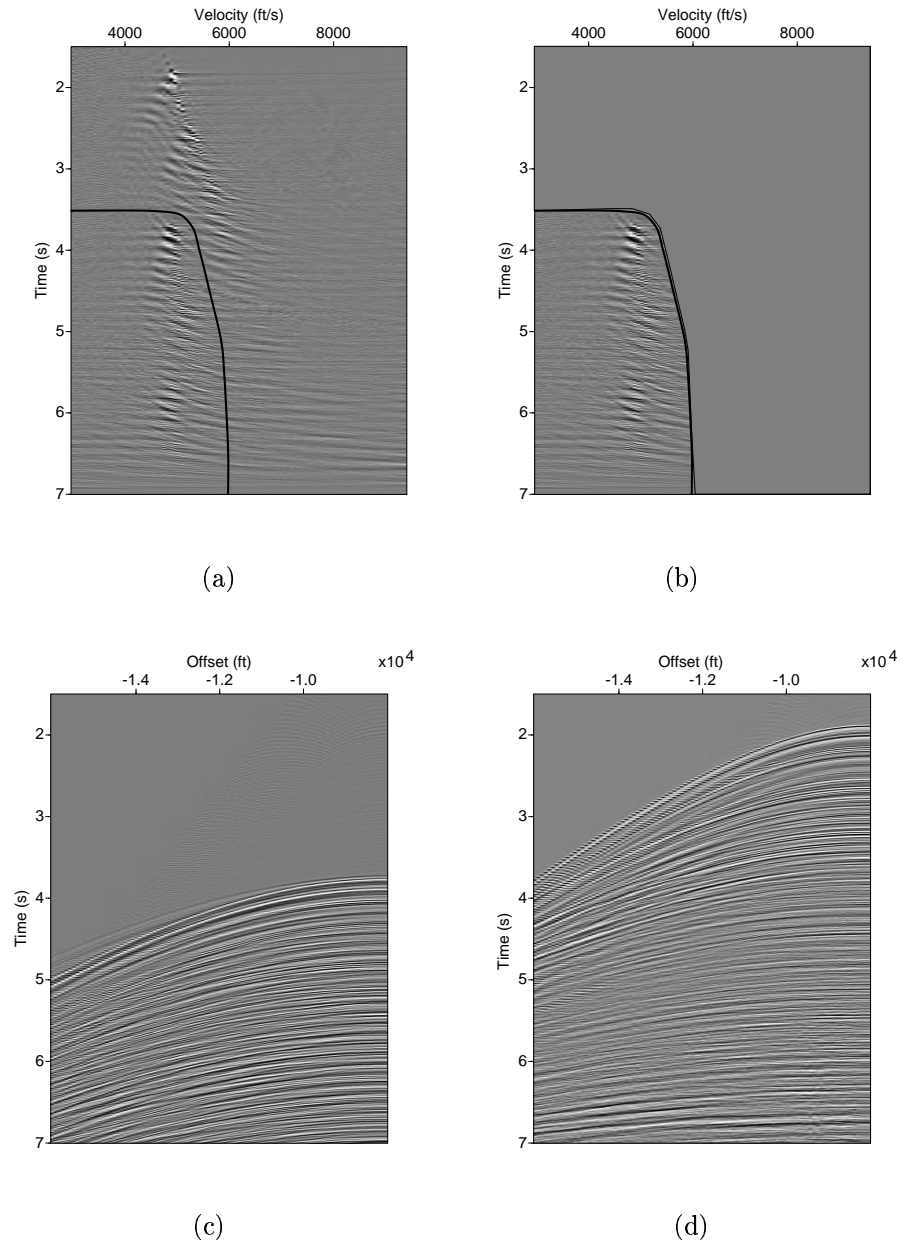


Figure 3.16: (a) The cut off line in the velocity panel for multiple suppression. (b) The primary part of the velocity is cut off, leaving the multiple part. (c) The multiple velocities mapped back into the data space. (d) the primaries obtained by subtracting multiples from the original data.

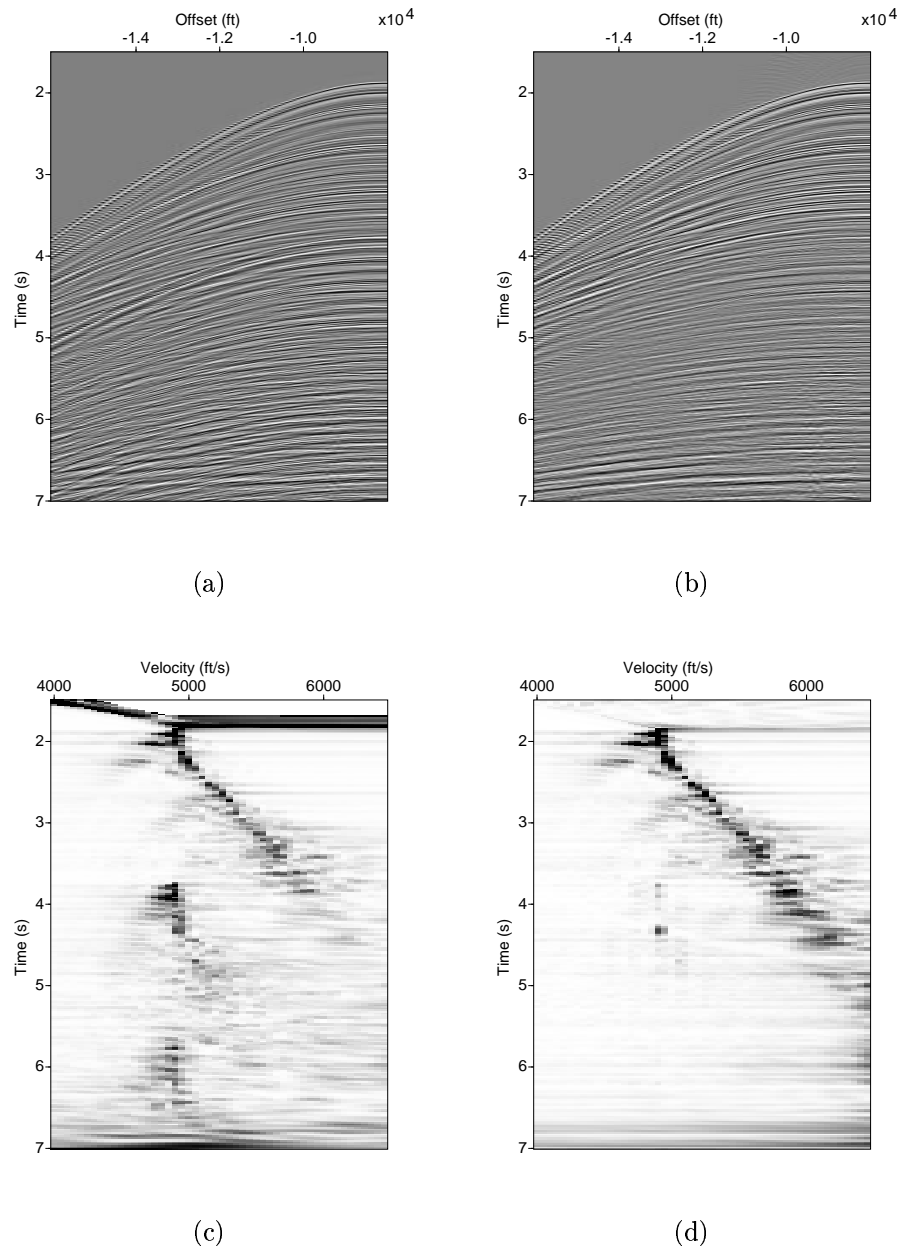


Figure 3.17: (a) Original data. (b) The data after multiple suppression. (c) Velocity analysis of Figure 3.17(a). (d) Velocity analysis of Figure 3.17(b).

# Chapter 4

## Resolution enhancement

### 4.1 Introduction

In the previous chapter we discussed the problem of inverting the hyperbolic Radon operator using least squares. In particular, we have centered our discussion around the problem of using the method of conjugate gradients to solve the system of normal equations that arises in the least-squares solution.

It is important to mention that to guarantee the stability of the inversion we often use an extension of least squares called penalized least-squares or damped least squares. In least squares we minimize the following cost function

$$J = (\mathbf{L} \mathbf{m} - \mathbf{d})^T (\mathbf{L} \mathbf{m} - \mathbf{d}). \quad (4.1.1)$$

The minimization of  $J$  leads to the least squares solution

$$\mathbf{m} = (\mathbf{L}^T \mathbf{L})^{-1} \mathbf{L}^T \mathbf{d}. \quad (4.1.2)$$

It is clear that when writing the last formula we have assumed that the matrix (operator)  $\mathbf{L}^T \mathbf{L}$  can be inverted. Unfortunately, this is a large operator where

some of its eigenvalues are quite small (often below machine precision). A technique to stabilize the inverse entails the definition of a penalized cost function of the form

$$J = (\mathbf{L} \mathbf{m} - \mathbf{d})^T (\mathbf{L} \mathbf{m} - \mathbf{d}) + \mu \mathbf{m}^T \mathbf{m}. \quad (4.1.3)$$

The second term in the right hand side of the last equation is the penalty term. This term forces the solution to be small, or in other words, forces the solution to be of finite size. Minimizing the penalized cost function with respect to  $\mathbf{m}$  leads to the damped least squares solution

$$\mathbf{m} = (\mathbf{L}^T \mathbf{L} + \mu \mathbf{I})^{-1} \mathbf{L}^T \mathbf{d}. \quad (4.1.4)$$

If the  $\mu > 0$  the new matrix  $(\mathbf{L}^T \mathbf{L} + \mu \mathbf{I})$  is a positive definite form, and therefore, invertible. The parameter  $\mu$  is often called the trade-off parameter of the problem.

This is a very simple way of stabilizing the inverse problem. However, damping the solution can lead to resolution degradation (Sacchi and Ulrych, 1995). In order to overcome this problem, we will redefine the penalty term in terms of a norm capable of estimating solutions that exhibit certain degree of sparseness. This concept has been originally proposed by Thorson and Claerbout (1985) to invert velocity stacks. In their approach the damping term  $\mu \mathbf{m}^T \mathbf{m}$  is replaced by a minimum entropy norm. This is a norm that enables one to retrieve sparse solution.

In this thesis, the quadratic penalty term is replaced by a Huber norm. The Huber norm has the property of behaving like a standard quadratic norm for small values of its argument while under-emphasizing large amplitudes. We will see that the Huber norm leads to an inversion strategy useful at the time of computing high resolution hyperbolic Radon transforms.

## 4.2 High resolution hyperbolic Radon transform

When data and velocity gather are written in lexicographic order, the hyperbolic Radon transform becomes

$$\mathbf{L} \mathbf{m} \simeq \mathbf{d}. \quad (4.2.1)$$

In the previous chapter we have used the method of conjugate gradients to solve the system of normal equations of the form

$$\mathbf{m}_{ls} = (\mathbf{L}^T \mathbf{L} + \mu \mathbf{I})^{-1} \mathbf{A}^T \mathbf{d}. \quad (4.2.2)$$

Rather than using quadratic regularization we introduce a new regularization term of the form

$$J = \|\mathbf{A} \mathbf{m} - \mathbf{d}\|^2 + \mu \mathfrak{R}(\mathbf{m}), \quad (4.2.3)$$

where  $\mathfrak{R}(\mathbf{m})$  can be defined by the Huber norm or Cauchy criteria as

$$\mathfrak{R}_H(\mathbf{m}) = \sum_i \rho_H(m_i/\sigma) \quad (4.2.4)$$

$$\mathfrak{R}_C(\mathbf{m}) = \sum_i \rho_C(m_i/\sigma) \quad (4.2.5)$$

where subscript “*C*” to denote functions associated with the Cauchy criteria, and “*H*” for the Huber norm.

### Cauchy norm regularization

The Cauchy norm is given by the function

$$\rho_c(x) = \ln\left(\frac{x^2}{2} + 1\right). \quad (4.2.6)$$



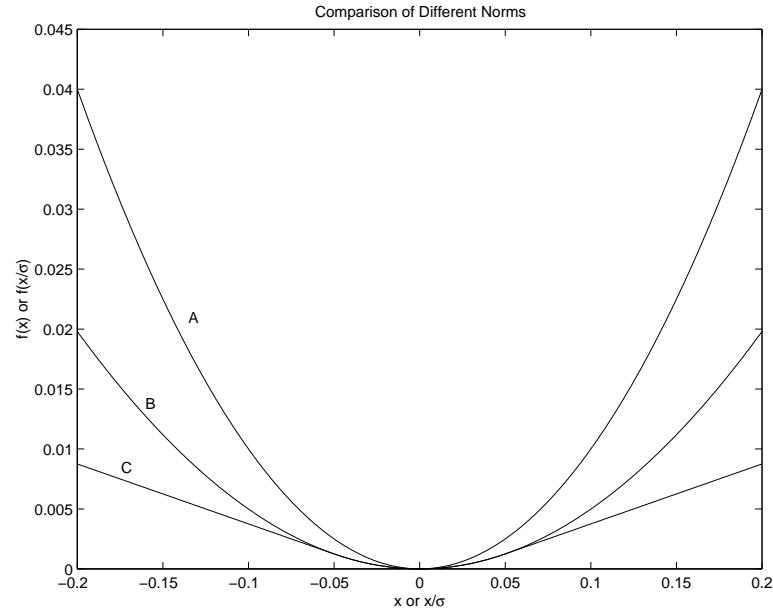


Figure 4.1: A: Euclidean norm. B: Cauchy norm. C: Huber norm ( $a = 0.01$ ).

Minimizing equation (4.2.3), with  $\rho$  defined by equation (4.2.6) yields

$$\mathbf{m}_C = (\mathbf{L}^T \mathbf{L} + \mu \mathbf{Q}_C)^{-1} \mathbf{L}^T \mathbf{d}, \quad (4.2.7)$$

where  $\mathbf{Q}_C$  is a diagonal matrix with elements given by:

$$Q_{C,ii} = \frac{2}{m_i^2 + 2\sigma^2}. \quad (4.2.8)$$

### Huber norm regularization

The Huber norm is given by

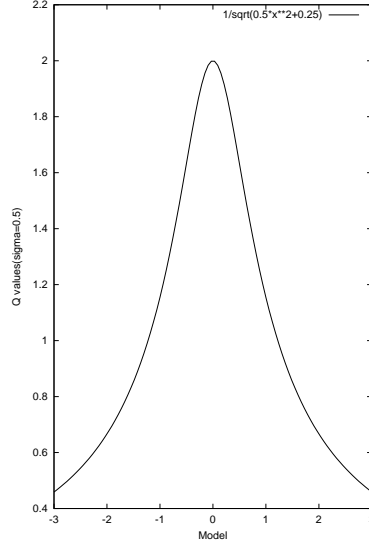


Figure 4.2:  $\mathbf{Q}$  values for different model elements. For smaller model elements,  $\mathbf{Q}$  is big, for bigger model elements,  $\mathbf{Q}$  is smaller.  $\sigma = 0.5$ .

$$\rho_h(x) = \begin{cases} x^2/2 & |x| \leq a \\ a|x| - a^2/2 & |x| > a \end{cases}, \quad (4.2.9)$$

where  $a$  is a threshold value that needs to be defined. After minimizing equation (4.2.3), with the regularization term given by the Huber norm (equation (4.2.9)) we obtain

$$\mathbf{m}_H = (\mathbf{A}^T \mathbf{A} + \mu \mathbf{Q}_H)^{-1} \mathbf{A}^T \mathbf{d}, \quad (4.2.10)$$

where the elements of  $\mathbf{Q}_H$  are

$$Q_{H,ii} = \begin{cases} \frac{1}{\sigma^2} & \frac{|m_i|}{\sigma} \leq a \\ \frac{a}{\sigma^2} \frac{|m_i|}{\sigma} & \frac{|m_i|}{\sigma} > a \end{cases}, \quad (4.2.11)$$

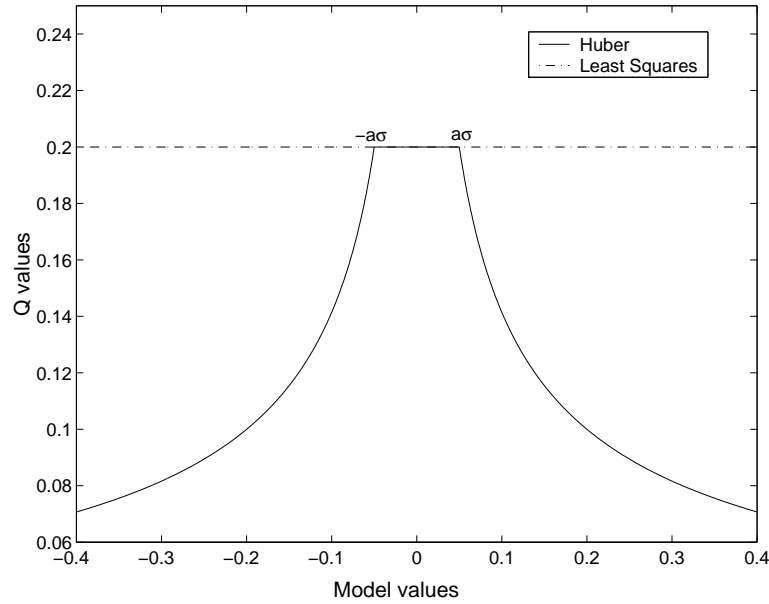


Figure 4.3: Elements of the matrix  $\mathbf{Q}$  versus model elements  $m$ .

### 4.2.1 Solution of the system of equations

When the Cauchy or Huber norm are adopted as regularization terms we need to solve a non-linear problem of the form:

$$(\mathbf{L}^T \mathbf{L} + \mu \mathbf{Q}(\mathbf{m})) \mathbf{m} = \mathbf{L}^T \mathbf{d}, \quad (4.2.12)$$

where the  $\mathbf{Q}(\mathbf{m})$  is the diagonal matrix that arises after using the Cauchy or the Huber penalty term. The last equation can be solved using iteratively re-weighted least-squares (IRLS) (Scales, 1987). Let us assume that the solution at iteration  $k$  is  $\mathbf{m}^k$ , then IRLS solves the following problem:

$$(\mathbf{L}^T \mathbf{L} + \mu \mathbf{Q}(\mathbf{m}^k)) \mathbf{m}^{k+1} = \mathbf{L}^T \mathbf{d}. \quad (4.2.13)$$

The system at iteration  $k + 1$  is solved using the method of conjugate gradients.

It is clear that now we have two iteration loops; one to solve the CG problem; the other to update the nonlinear term  $\mathbf{Q}(\mathbf{m})$ . In general, we have found that only 5 – 6 updates are required to find the solution. As per the CG internal iterations we have used the following stopping criterion:

$$\frac{|J^k - J^{k-1}|}{|J^k + J^{k-1}|} < \epsilon$$

where  $\epsilon = 10^{-4} - 10^{-6}$ . In general, one can start the algorithm with the null solution,  $\mathbf{m}^0 = \mathbf{0}$ .

### 4.2.2 Hyper-parameter selection

In this section we discuss the problem of estimating the hyper-parameters  $\mu$ ,  $\sigma$  and  $a$  required by the Huber norm. In general, the Cauchy norm and the Huber norm give similar results. Moreover, the convergence of the inversion is quite similar in both cases. Therefore, we will concentrate on the analysis of the Huber norm. This is the norm that we have adopted in our numerical simulations.

From Figure 4.2 and 4.3, we can see that  $a\sigma$  is a threshold parameter. Before going to the next step, the significance of  $\mathbf{Q}$  has to be explained in detail. In our inversion we are simultaneously constraining the solution to honor the following two equations:

$$\begin{cases} \mathbf{L}\mathbf{m} \approx \mathbf{d} \\ \mathbf{Q}^{1/2}\mathbf{m} \approx \mathbf{0} \end{cases} . \quad (4.2.14)$$

The first equation states that we want a solution that fits (within some error bounds) the data; whereas the second equation states that the weighted average  $\mathbf{Q}^{1/2}\mathbf{m}$  should be small. Let us expand the term  $\mathbf{Q}^{1/2}\mathbf{m}$

$$\begin{pmatrix} \sqrt{Q_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{Q_2} & 0 & \cdots & 0 \\ 0 & 0 & \sqrt{Q_3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sqrt{Q_n} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \vdots \\ m_n \end{pmatrix} \approx \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \quad (4.2.15)$$

Equation (4.2.15) suggests that the elements of  $\mathbf{Q}$  should be defined in such a way that a large weight should be given to model parameters ( $m_i$ ) that should be small, and a small weight to model parameters that should be large. One can interpret these weights as the inverse of a variance; where a small variance means that the parameter is not allowed to fluctuate. The elements  $q_i$  are a function of the model parameter  $m_i$ . The weights  $q_i$  will serve to enhance large model parameters and to de-emphasize small values of  $m_i$ . This is consistent with the idea of finding a sparse solution. In other words, a solution that can represent the data with the minimum number of clusters in velocity space.

The criteria to choose the parameters  $a$ ,  $\sigma$  and  $\mu$  is described as follows. Equation (4.2.11) can be re-written as

$$Q_{H,ii} = \begin{cases} \frac{1}{\sigma^2} & |m_i| \leq a \sigma \\ \frac{a}{\sigma^2 \left| \frac{m_i}{\sigma} \right|} & |m_i| > a \sigma \end{cases}. \quad (4.2.16)$$

Introducing a new variable  $m_c = a\sigma$ , where  $m_c$  is chosen as some percentage of the maximum expected value of  $m_i$ , i.e.,

$$m_c = 0.01 \max[m_i]$$

then the regularization term becomes:

$$Q_{H,ii} = \frac{1}{\sigma^2} \begin{cases} 1 & |m_i| \leq m_c \\ \frac{m_c}{|m_i|} & |m_i| > m_c \end{cases}, \quad (4.2.17)$$

The unknown term  $\sigma^2$  can be absorbed by the tradeoff parameter  $\mu$  in equation (4.2.10). This new parameter is estimated by trial and error after analyzing the data misfit. In other words, we looked at the error panel (original data minus the reconstructed data) and select  $\mu$  in such a way that data are properly reproduced.

### 4.3 Source wavelet deconvolution

Our original operator  $\mathbf{L}$  is modified in order to contain a source wavelet term. The new hyperbolic Radon operator can be written down as follows:

$$\mathbf{d} = \mathbf{W} \mathbf{L} \mathbf{m} \quad (4.3.1)$$

If  $\mathbf{m}$  is a collection of spikes in the  $\tau - v$  space (velocity space), then  $\mathbf{L} \mathbf{m}$  is a superposition of full band wavelets with hyperbolic move-out in data space. The operator  $\mathbf{W}$  is a temporal convolution with a source wavelet that makes the seismic data band-limited. This is quite important since it helps to fit the data which is naturally band-limited.

In our algorithm we need to specify the adjoint operator of  $\mathbf{W} \mathbf{L}$ . The adjoint is given by  $\mathbf{L}^T \mathbf{W}^T$ . We have already seen that  $\mathbf{L}^T$  is a stacking along hyperbolic events. The operator  $\mathbf{W}^T$  is the adjoint of convolution, this is the correlation of a vector in model space with the wavelet (Claerbout, 1992).

The source wavelet in our algorithm is a band-limited zero phase finite length filter.

### 4.3.1 Convolution and Correlation

In the following section, I will show that convolution and correlation are a forward/adjoint pair.

For two vectors,  $\mathbf{b} = (b_1, b_2)$  and  $\mathbf{x} = (x_1, x_2, x_3)$ , the convolution of the two vectors are defined by

$$\begin{aligned} \mathbf{y} &= \mathbf{b} * \mathbf{x} \\ &= (b_1, b_2) * (x_1, x_2, x_3) \\ &= (b_1x_1, b_2x_1 + b_1x_2, b_2x_2 + b_1x_3, b_2x_3), \end{aligned} \quad (4.3.2)$$

where  $\mathbf{y}$  present the output vector, and  $*$  represents convolution. The convolution can also be written in matrix form,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} b_1 & 0 & 0 \\ b_2 & b_1 & 0 \\ 0 & b_2 & b_1 \\ 0 & 0 & b_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad (4.3.3)$$

where the multiplication can be shortened as

$$\mathbf{Y} = \mathbf{B}\mathbf{x}, \quad (4.3.4)$$

the operation  $\mathbf{B}\mathbf{x}$  convolves vector  $\mathbf{b}$  with  $\mathbf{x}$ , whereas the operation  $\mathbf{B}^T\mathbf{y}$  cross correlates  $\mathbf{b}$  with  $\mathbf{y}$ , which can be written in matrix form,

$$\begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & 0 & 0 \\ 0 & b_1 & b_2 & 0 \\ 0 & 0 & b_1 & b_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix}, \quad (4.3.5)$$

this is the same as the definition of cross-correlation of two vectors,

$$\begin{aligned}
 \mathbf{x}' &= \mathbf{b} \cdot \mathbf{y} \\
 &= (b_1, b_2) \cdot (y_1, y_2, y_3, y_4) \\
 &= (b_2 y_1, b_1 y_1 + b_2 y_2, b_1 y_2 + b_2 y_3, b_1 y_3 + b_2 y_4, b_1 y_4), \quad (4.3.6)
 \end{aligned}$$

where “.” denote cross-correlation.

## 4.4 Examples

The validity of the Huber norm regularization is demonstrated using synthetic and real data examples.

### 4.4.1 Synthetic data examples

Figure 4.4 shows the synthetic example used to test Huber regularization algorithm. The near offset is at  $h_0=40$  m, the offset increment is  $\Delta h=50$ m, the total number of samples is  $nt=800$ , and the temporal sampling rate is  $\Delta t=0.004$ s. Figure 4.4(a) and 4.5(a) show a primary at time 0.8 s, and its multiple at 1.6 s. The horizontal axis is offset, the vertical axis is time. In these examples, there are 800 time samples and the sampling interval ( $\Delta t$ ) is 4 ms. The offset starts at 40 m. The offset interval is 50 m. The velocity axis starts at 500 m/s and ends at 5000 m/s. There are 40 velocity samples.

Figure 4.4(b) shows the inverted velocity using the Huber norm. By comparison with Figure 4.5(b), it is obvious that the Huber inverted velocity has a more focused energy distribution than that of damped least squares. Next, the inverted velocities are mapped back into data space. The velocity gather is first muted to eliminate the primaries. Figure 4.4(c) and figure 4.5(c) show inverted



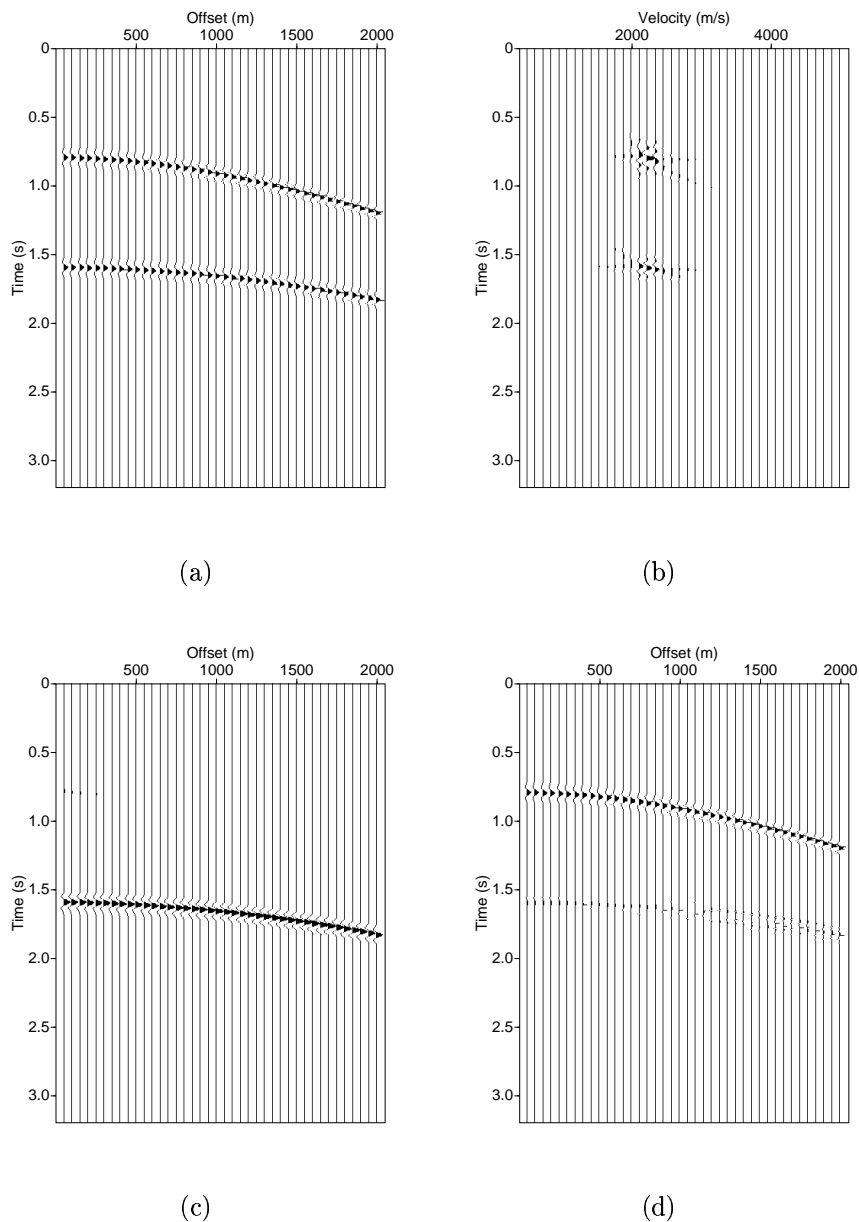


Figure 4.4: (a) Synthetic data. (b) Velocity gather obtained by inverting the hyperbolic Radon transform using the Huber norm regularization. The parameters adopted in this simulation were  $\mu=500$ ,  $m_c=0.08$ ,  $v_0=500m/s$ ,  $v_{max}=5000m/s$ , and a total number of velocity traces  $nv = 40$ . (c) Multiples obtained after muting figure (b). (d) Primaries obtained after subtracting the multiples from the data.

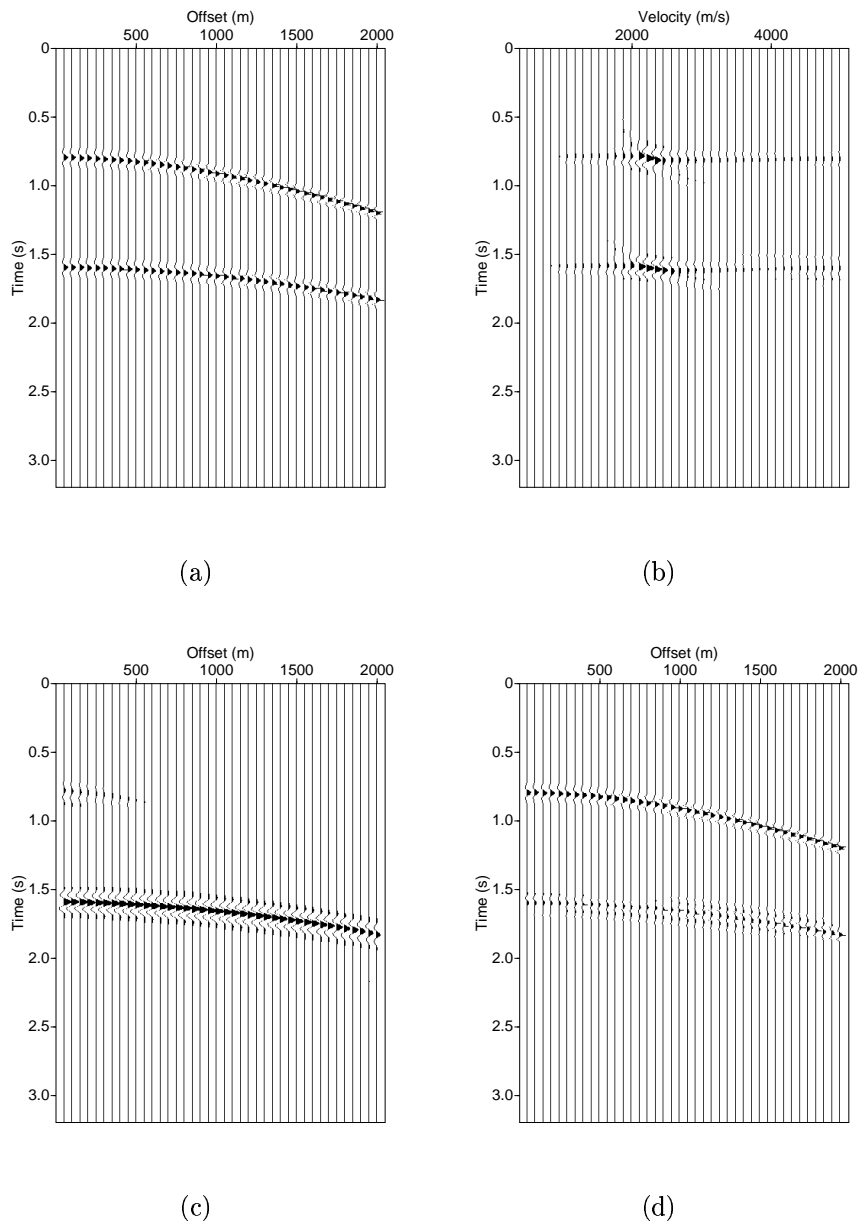


Figure 4.5: (a) Synthetic data with parameter given by  $h_0=40$ ,  $\Delta h=50$ ,  $nt=800$ ,  $dt=0.004s$ . (b) Damped least-squares solution obtained using the method of conjugate gradients. The parameters used in the inversion were  $\mu=500$ ,  $v_0=500m/s$ ,  $v_{max}=5000m/s$ ,  $nv=40$ . (c) Inverted model of multiples using damped least squares. (d) Result of subtracting multiples from the data.

velocity gathers obtained with the Huber norm regularization and damped least squares respectively. In the next step, the multiples are subtracted from the data. The estimated primaries obtained with the Huber norm regularization and damped least squares respectively are shown in Figure 4.4(d) and Figure 4.5(d).

## 4.4.2 Real data examples

In this section a CMP from a sub-salt data set is used to illustrate the multiple removal. The field geometry information is listed in Table 5.1 at the beginning of the next chapter. The CMP is processed using the Huber regularization and least squares, the results of which are shown in Figure 4.6 and 4.7, respectively. Figure 4.6(a) shows the original data of CMP 1000. Then the data are mapped to velocity space by inverting the hyperbolic Radon transform using the Huber regularization, as shown in Figure 4.6(b). The resolution of the velocity space is quite good, and the multiples and primaries are well separated. A mask is applied to the velocity panel to cut off the primaries, leaving only the multiples. In the next step, the multiples are mapped back to data space. The result is shown in Figure 4.6(c). In this figure, there is no event above 4 s because the very first multiple, the water bottom multiple, appears at 4 s. After subtracting the multiples from the original data, we obtained the primaries. Figure 4.6(d) shows the primaries obtained using the Huber norm inversion procedure.

By comparison, Figure 4.7 shows the results obtained using damped least-squares. Figure 4.7(a) shows the original data, CMP 1000. Figure 4.7(b) is the velocity obtained by damped least-squares. Note that the resolution is not quite good, amplitude smearing introduced by finite aperture is the major source of resolution degradation. Figure 4.7(c) and 4.7(d) are the multiples and primaries obtained by muting 4.7(b).

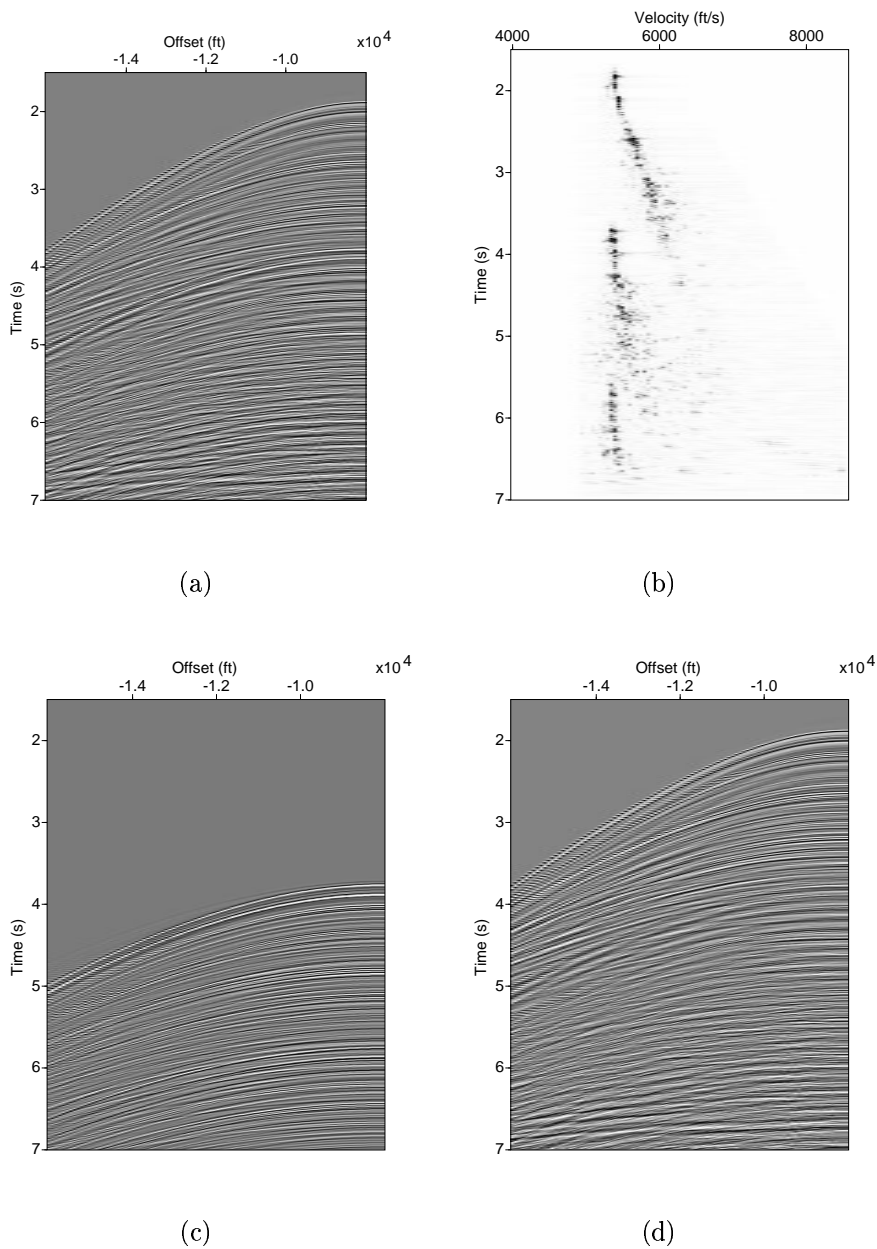


Figure 4.6: Field data example. The CMP gather was inverted using the Huber norm regularization method described in this chapter. (a) CMP 1000 from a Gulf of Mexico marine data set. (b) Inverted velocity panel. (c) Multiples obtained by muting the primaries in velocity space. (d) Primaries, this is the result of subtracting multiples from the original data shown in Figure 4.6(a).

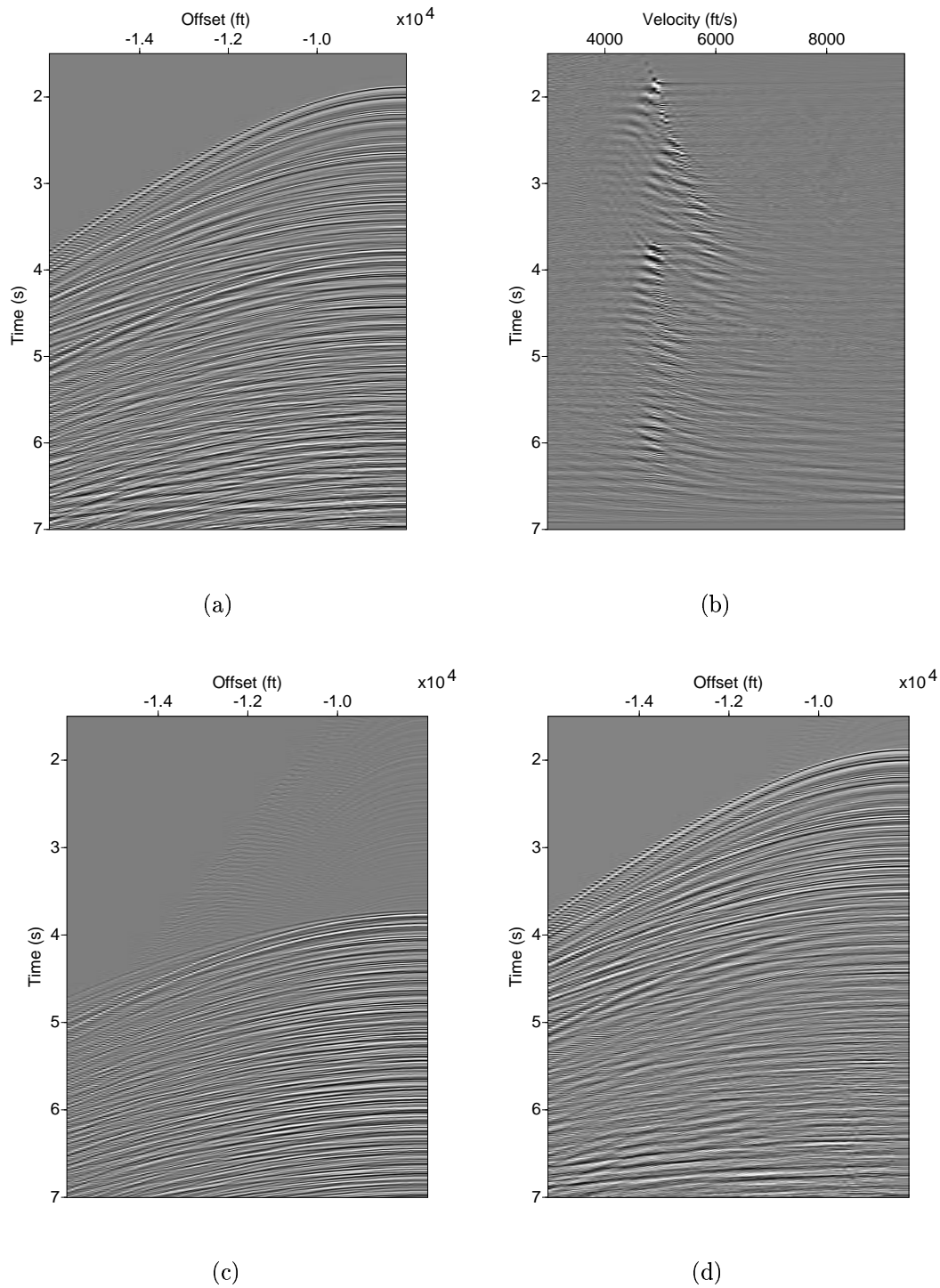


Figure 4.7: (a) CMP #1000 from the Gulf of Mexico. (b) Damped least-squares inversion. (c) Multiples obtained. (d) Model of primaries.

It is important to stress that the high resolution velocity gather obtained with the Huber regularization allows us to design a better mute than the velocity gather obtained using damped least-squares.

## 4.5 Summary

In this chapter I have proposed a new strategy to invert the hyperbolic Radon transform that is used to compute velocity gathers for de-multiple. In this new method a Huber norm regularization term is included in the formulation of the inverse problem to construct a velocity gather that is sparse.

The method of conjugate gradients plays a major role in obtaining the solution of our inverse problem. The hyperbolic Radon operator, as we have already discussed in the previous chapter, cannot be inverted by direct methods. In the presence of a non-quadratic regularization term (i.e., Huber and Cauchy norms) iterative re-weighted least-squares has proven to be an effective algorithm to solve our problem.

# Chapter 5

## Field data experiments

### 5.1 Introduction

In this chapter, we use the hyperbolic Radon transform to process a marine data set from the Gulf of Mexico. These data have been provided by Western Geophysical to several academic and industrial research groups to test new multiple attenuation technologies.

The Gulf of Mexico data set has severe sub-salt imaging problems. Multiples are more problematic than usual despite the fact that they are weak. This is because the primaries below the salt are weak and the primary-multiple ratio is usually low. Therefore, removing the multiples without touching primaries is a critical concern at the time of processing these data. The depth of the water column is approximately 1500 *m*. First and second order multiples can be easily identified because of their periods which is about 2 *s*. The data acquisition parameters are given in Table 5.1.

Figure 5.1 shows the stacked section of the data set before multiple suppression. Please note that the time axis starts at 1.5 second. In this figure, the most significant first order multiples below the salt body are indicated with arrows.

CDP numbers	818-2618
number of traces per shot	183
shot interval	87.5 ft
group spacing	87.5 ft
farthest offset	-15993 ft
nearest offset	-68 ft
number of samples per trace	1751
time sampling interval	4 ms
starting time	1.5 s

Table 5.1: Field acquisition parameters for the Gulf of Mexico data set.

Note that a large part of the multiple problem is caused by surface multiples that are not related to the water bottom. For example, multiples that bounce upward twice from the the top of the salt, or once each from top and bottom of the salt. From 6.0 second downward, a series of second order multiples can be observed. A common effect in this environment are zones with weak multiple energy, such as in the areas around CDP 1400 and 2550, which are right next to zones with strong, focused multiple energy around CDP 1300 and 2300. These effects are due to the flanks in the water-bottom and sub-bottom reflectors which act as a lens. Note also that since the top and the bottom are not very continuous many diffractions can be observed. This effect is more obvious in the multiples generated in the salt boundaries (Verschuur, 1999).

## 5.2 Real data processing procedures

The following procedures are applied in order to remove the multiples from the data set.



- Sort shot gathers into CDP gathers.
- Apply gain function to compensate for geometric spreading. The gain used in this study is a  $t^2$  function.
- Compute the velocity gather for each CDP by inverting the hyperbolic Radon transform using the Huber regularization strategy outlined in chapter 4.
- Design a mask or muting function to retain the multiples in velocity space. This mask is assumed to be variant from CDP to CDP.
- Map the muted velocity gather to data space to generate a model of multiples.
- For each CDP, subtract the multiples from the original data to generate a model of primaries.
- Use normal moveout correction and stacking to generate a multiple free seismic section.
- Perform velocity analysis at selected CDP locations to verify that the algorithm has been able to remove the multiples.

Figure 5.2 shows a CMP gather from the real data set. This is CMP #1000 located at the right hand side of the stack in Figure 5.1. Figure 5.2(b) shows the primaries in this CMP (multiples removed). Figures 5.2(c) and 5.2(d) show the velocity spectra computed using velocity analysis for Figures 5.2(a) and 5.2(b), respectively. The later illustrates that the multiples have been properly removed.

Figure 5.3 shows the estimated multiples and the associated semblance panel. After multiple removal the primaries are NMO-corrected. Figures 5.3(c) and 5.3(d) show the NMO-corrected original data and the NMO-corrected primaries.

In the Figure 5.3(c) the under-corrected multiples are clearly visible. After multiple removal the under-corrected multiples have been eliminated. The only events left are the primaries, see Figure 5.3(d).

### 5.3 Final stack after multiple removal

After the multiples are removed from the CDPs, each CDP is summed along the offset to compute a stacked trace at each CDP location. Figure 5.4 shows the result of the final stack. The final stack shows that the reflections below the salt body are more visible than before de-multiple. Also, the diffraction patterns at CDP 1350 are clearly imaged. By comparing with Figure 5.1, it is obvious that the multiples from 6.0 to 6.5 s, and CDP 1000 to 1300 are well removed and that the reflections are clear. The salt body in Figure 5.1 is also imaged much better in Figure 5.4.

### 5.4 Techniques used to improve the computational efficiency of the algorithm

When dealing with real data it is extremely important to optimize the de-multiple algorithm as much as possible in order to minimize turn-around times. This is a summary of the techniques used to speed up the de-multiple algorithm:

**Select  $\tau$  dependent velocities** When solving the inverse problem, we adopt  $\tau$  dependent velocities to decrease the size of the hyperbolic Radon transform operator. The shaded area in Figure 5.5 defines the range of velocities used by the algorithm.

**Windowing the CMP gather** The size of the hyperbolic Radon operator can be further decreased by avoiding areas of the CDP that have been muted.

## 5.5 Summary

In this chapter, the high resolution hyperbolic Radon transform was successfully applied to remove multiples from a marine data set from the Gulf of Mexico.

The hyperbolic Radon transform was inverted using the Huber regularization method proposed in Chapter 4. It is important to stress that the high resolution velocity gather obtained with this method permits one to easily define a muting operator to isolate multiples from primaries. Semblance panels computed via velocity analysis were used as a quality control tool.

For large data sets an simple implementation in a PC cluster can substantially decrease computing time. In other words, each processor in the cluster can process one CDP at the time. Since there is no communication between processor the scalability of the program is guaranteed. The latter is quite important when dealing with large marine data sets where thousands of CDPs have to be processed during multiple removal.

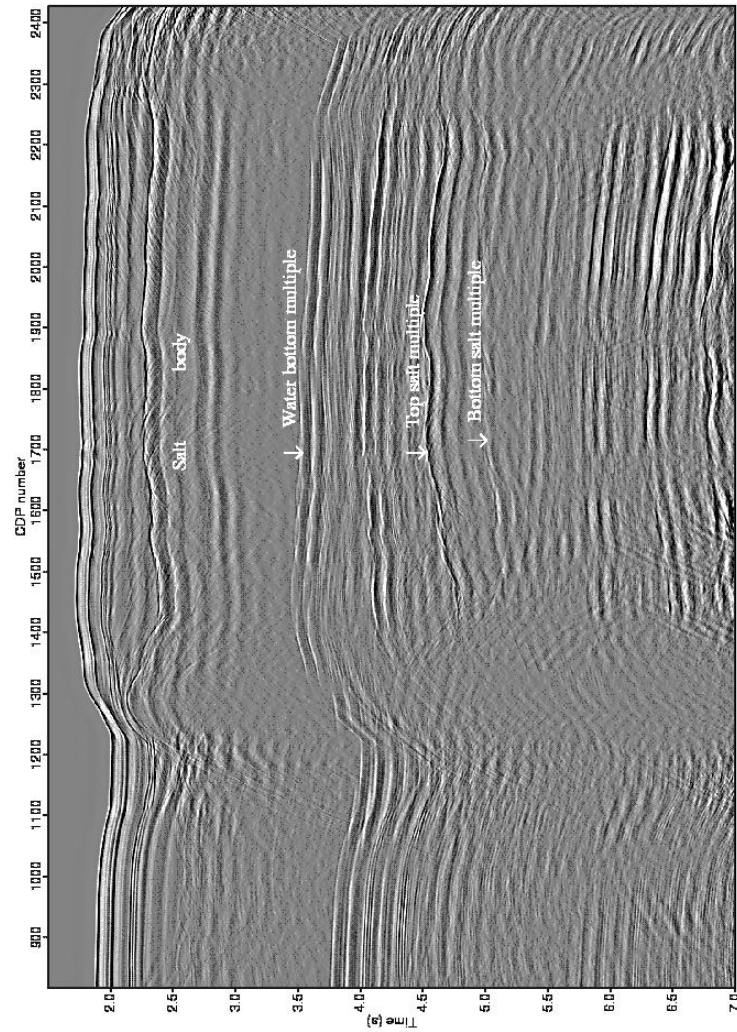


Figure 5.1: Stacked section from the Gulf of Mexico data set. The data are severely contaminated by multiples.

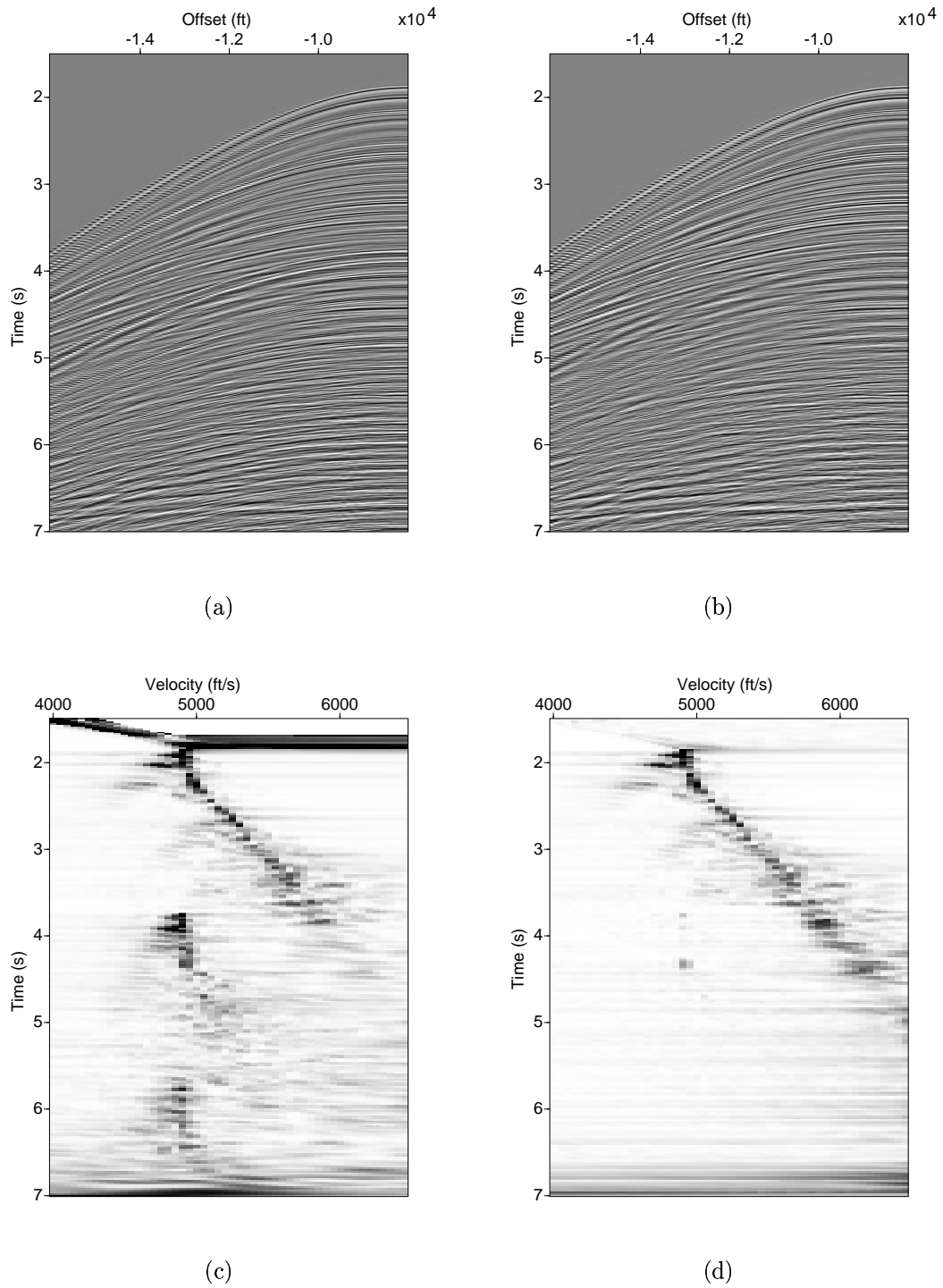


Figure 5.2: (a) Original data. (b) The estimated primaries. (c) Velocity analysis of Figure 5.2(a). (d) Velocity analysis of Figure 5.2(a).

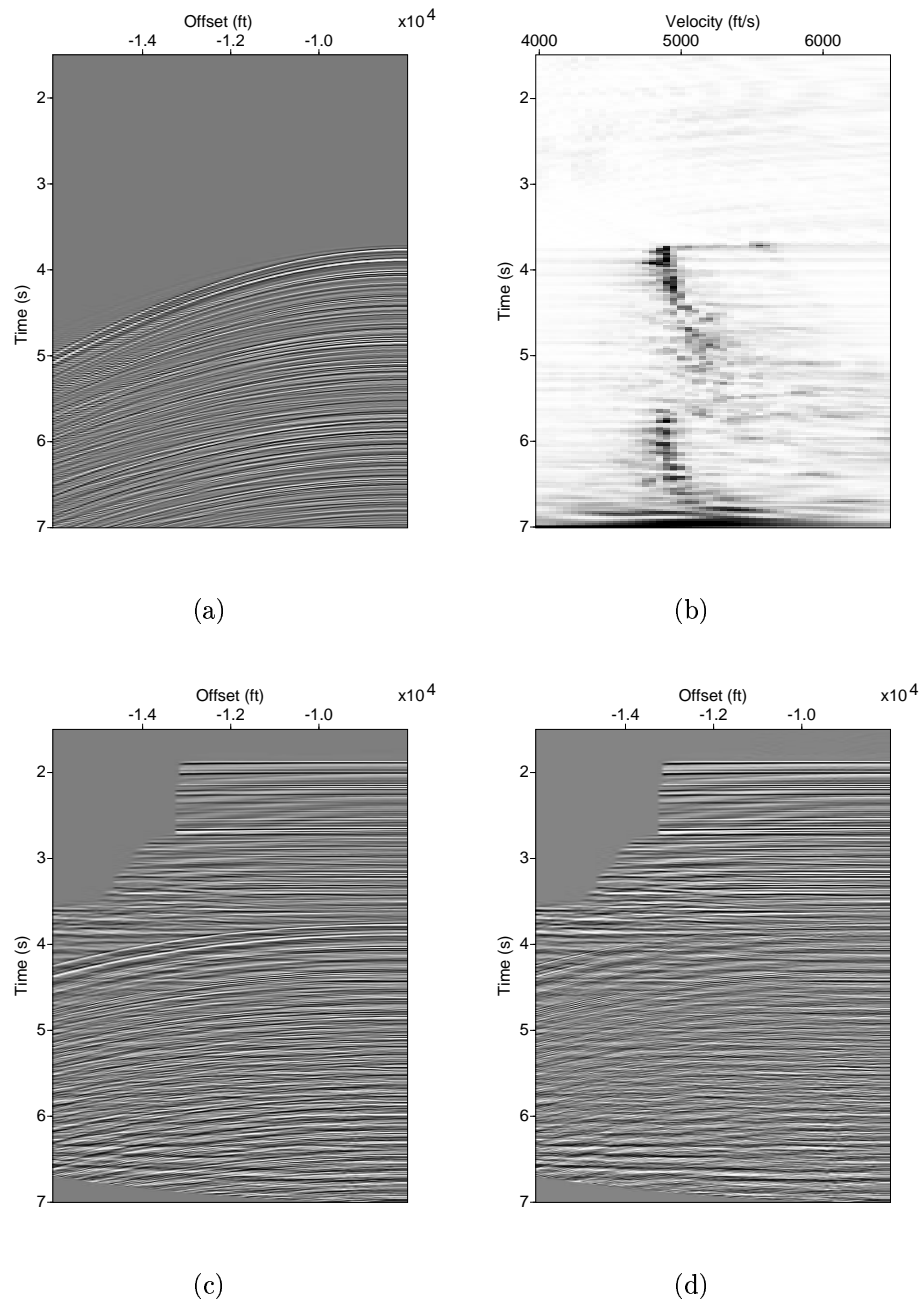


Figure 5.3: (a) Multiples separated from the original data. (b) Velocity analysis of Figure 5.3(a). (c) NMO-corrected original data. (d) NMO-corrected primaries.

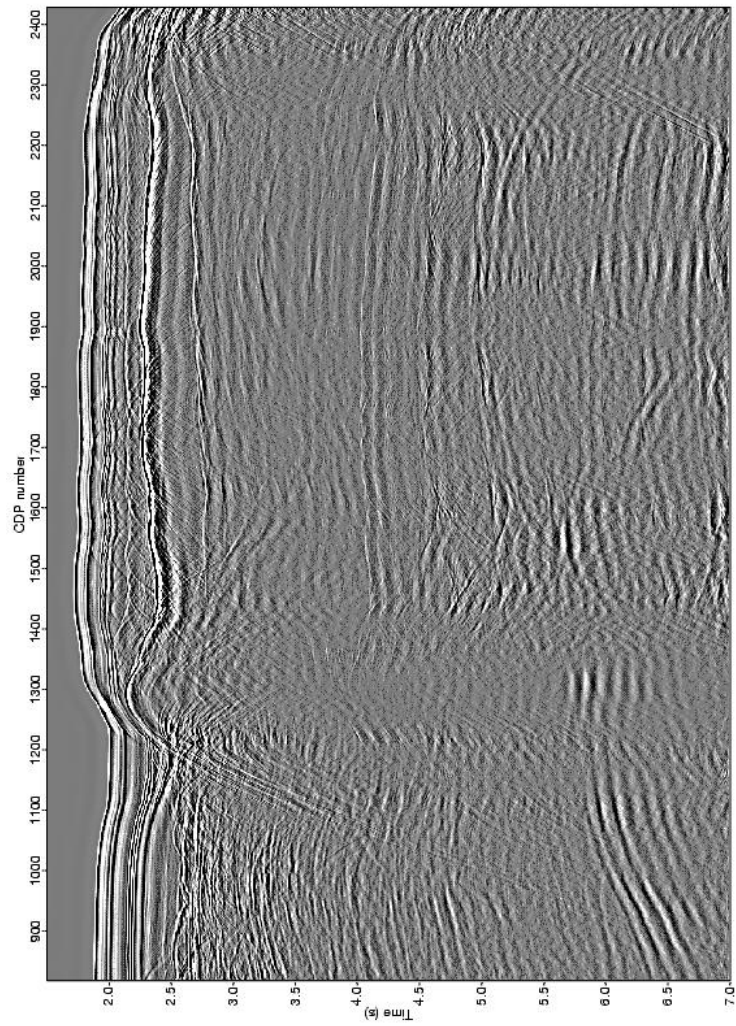


Figure 5.4: The final stack section after multiple removal.

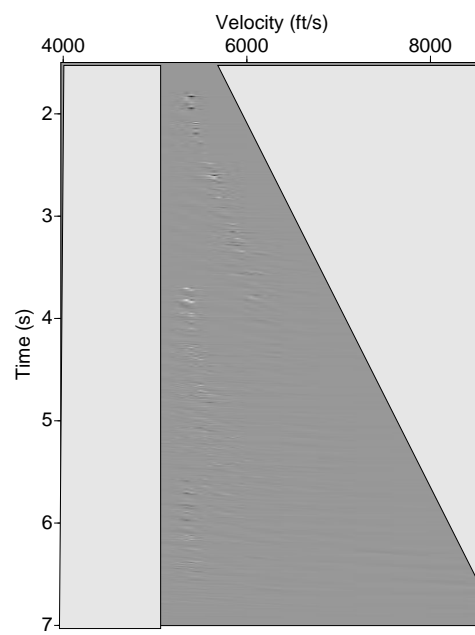


Figure 5.5:  $\tau$  dependent velocities were selected to decrease the size of the hyperbolic Radon operator.



## Chapter 6

# Conclusions

In this thesis I have discussed the problem of multiple removal in reflection seismology using the Radon transform. In particular, I have focused my research on the problem of inverting the hyperbolic Radon transform. This is an alternative to the popular time invariant parabolic Radon transform often used in the oil/gas exploration industry.

One of the advantages of the Parabolic Radon transform is that there exist very fast algorithms that operate in the frequency domain. This is not true for the hyperbolic Radon transform. It needs to be implemented in the time-offset space. Therefore, its inversion becomes a difficult task. We have developed an inversion methodology that uses the method of conjugate gradients. The method of conjugate gradients is capable of inverting any linear operator as far as the inner products  $\mathbf{L}\mathbf{x}$  and  $\mathbf{L}^T\mathbf{y}$  are defined. In other words, the operator  $\mathbf{L}$  does not need to be explicitly defined as a matrix.

In Chapter 4 we explored the problem of designing a high resolution velocity gather by introducing the Huber norm regularization term in the formulation of our inverse problem. It is important to mention that there are other norms (i.e., Cauchy) that can be used to improve resolution in velocity space. However, the

simplicity of the Huber norm and its similarity to the classical quadratic norm makes it very attractive for the task of inverting sparse model.

In Chapter 5 I have successfully applied the hyperbolic Radon transform to a field data de-multiple problem. I have processed a marine data set from the Gulf of Mexico. These particular data have become a benchmark to test de-multiple algorithms. It is important to stress that the results obtained in Chapter 5 coincide quite well with those obtained by other researchers using wave equation de-multiple methods (see for instance, the special issue of *The Leading Edge* of January 1999).

At this point it is important to mention that more research needs to be done in order to further improve the computational efficiency of the hyperbolic Radon de-multiple algorithm. The computational cost of the hyperbolic Radon transform is about 4 – 5 times the cost of the parabolic Radon transform. Reducing the computational cost of computing/inverting the hyperbolic Radon transform is an interesting research avenue to explore in the future.

# Bibliography

- [1] Alam, A., and Austin, J., 1981, Multiple suppression using slant stacks: Tech. Rep., Western Geophysical Company.
- [2] Cary, W., P., 1998, The simplest discrete Radon transform: 68th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts.
- [3] Chapman, C. H., 1981, Generalized Radon transforms and slant stacks: Geophys. J. Roy. Astr. Soc., **66**, 445-453.
- [4] Claerbout, J. F., 1992, Earth sounding analysis: processing versus inversion: Blackwell Scientific Publications.
- [5] Claerbout, J. F., 1985, Imaging the Earth's interior: Blackwell Scientific Publications.
- [6] Deans, S. R., 1983, The Radon transform and some of its applications: J. Wiley & Sons, Inc.
- [7] Durrani, T. S., and Bisset, D., 1984, The Radon transform and its properties: Geophysics, **49**, 1180-1187.
- [8] Foster, D. J., and Mosher, C. C., 1992, Suppression of multiple reflections using the Radon transform: Geophysics, **57**, 386-395.

- [9] Guitton, A., and Symes, W. W., 1998, Robust and stable velocity analysis using the Huber function: 68th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts.
- [10] Gulunay, N., 1990, F-X domain least-squares  $\tau$ - $p$  nad  $\tau$ - $q$ : 60th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1607-1610.
- [11] Hampson, D., Dec. 1986, Inverse velocity stacking for multiple elimination: J. Can. Soc. Expl. Geophys., **22**, 44-55.
- [12] Hill, S., Dragoset, B., and Weglein, A., 1999, The new world of multiple attenuation: The Leading Edge, **18**, No.1, 38-38.
- [13] Huber, P. J, 1981, Robust Statistics: John Wiley and Sons, Inc.
- [14] Hunt, L, Cary, P., and Upham, W., 1996, An improved Radon transform for short period multiple attenuation: CSEG 23rd Annual Mtg., Expanded Abstracts, 58-59.
- [15] Kabir, M. M. N., and Marfurt, K. J., 1999, Towards true amplitude multiple removal: The Leading Edge, **18**, No. 1, 66-73.
- [16] Kostov, C., 1990, Toplitz structure in slant stack inversion: 60th Annual Internat. Mtg., Soc. Expl. geophys., Expanded Abstracts, 1618-1621.
- [17] Marple, S. M., 1987, Digital spectral analysis, with applications: Prentice-Hall.
- [18] Neidell, N. S., and Taner, M. T., 1971, Semblance and other coherency measures for multichannel data: Geophysics, **36**, 482-497.
- [19] Robinson, E. A., and Treitel, S., 1980, Geophysical signal processing: Prentice Hall.

- [20] Ryu, J. V., 1982, Decomposition (DECOM) approach applied to wave-field analysis with seismic reflection records: *Geophysics*, 47, 869-883.
- [21] Sacchi, M. D., and Ulrych, T. J., 1995, High-resolution velocity gathers and offset space reconstruction: *Geophysics*, **60**, 1169-1177.
- [22] Sacchi, M. D., 1997, Reweighting strategies in seismic deconvolution: *Geophys. J. Internat.*, **129**, 651-656.
- [23] Sacchi, M. D., and Porsani, M.J., 1999, Fast high resolution Radon transform: 69th Annual Intern. Mtg. Soc. Expl. Geophys., Expanded Abstracts, 1657-1660.
- [24] Sacchi, M. D., 2001, GEOPH-326 Refraction and Reflection Seismology, Course Notes, Dept. of Physics, University of Alberta.
- [25] Scales, A., J., 1987, Tomographic inversion via the conjugate gradient method: *Geophysics*, **52**,179-185.
- [26] Sheriff, R. E., 1991, Encyclopedic dictionary of exploration geophysics: Soc. Expl. Geophys.
- [27] Shewchuk, R., J., 1994, An introduction to the conjugate gradient method without the agonizing pain: <http://www.cs.cmu.edu/~jrs/>.
- [28] Strang, G., 1986, Introduction to applied mathematics: Welley-Cambridge Press.
- [29] Taner, M. T., 1980, Long-period sea-floor multiples and their suppression: *Geophys., Prosp.*,**28**, 30-48.
- [30] Telford, W., M., Geldart, L. P., and Sheriff, R. E., 1990, Applied Geophysics, 2nd edition, Cambridge University Press.

- [31] Thorson, J. R., and Claerbout, J. F., 1985, Velocity-stack and slant stack stochastic inversion: *Geophysics*, **50**, 2727-2741.
- [32] Treitel, S., Gutowski, P. R., and Wagner, D. E., 1982, Plane wave decomposition of seismograms: *Geophysics*, **47**, 1375-1401.
- [33] Turner, G., 1990, Aliasing in the  $\tau$ - $p$  transform and the removal of spatial aliased coherent noise: *Geophysics*, **55**, 1496-1503.
- [34] Verschuur, J., D., 1999, Multiple removal results from Delft University: *The Leading Edge*, **18**, No.1, 86-91.
- [35] Weglein, A. B., Carvalho, P. M. and Stolt, R. H., 1992, Nonlinear inverse scattering for multiple suppression: Application to real data. Part I, 62nd Ann. Internat. Mtg: Soc. Expl. Geophys., 1093-1095.
- [36] Weglein, A., 1999, Multiple attenuation: an overview of recent advances and the road ahead (1999): *The Leading Edge*, **18**, No.1, 40-44.
- [37] Yilmaz, Ö., 1989, Velocity-stack processing: *Geophys. Prosp.*, **37**, 357-382.
- [38] Yilmaz, Ö., 1987, *Seismic data processing*: Soc. Expl. Geophys.
- [39] Zhou, B., and Greenhalgh, S. A., 1994, Linear and parabolic  $\tau$ - $p$  transforms revisited: *Geophysics*, **59**, 1133-1149.