

**APERTURE COMPENSATED RADON AND FOURIER TRANSFORMS**

by

**Mauricio Dino Sacchi**

**Geofísico, Universidad Nacional de La Plata**

**A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF  
DOCTOR OF PHILOSOPHY**

in

**THE FACULTY OF GRADUATE STUDIES  
GEOPHYSICS AND ASTRONOMY**

**We accept this thesis as conforming  
to the required standard**

**THE UNIVERSITY OF BRITISH COLUMBIA**

**January 1996**

**© Mauricio Dino Sacchi, 1996**

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Geophysics and Astronomy  
The University of British Columbia  
129-2219 Main Mall  
Vancouver, Canada  
V6T 1Z4

Date:

JANUARY / 30 / 1996

## Abstract

In seismic data analysis, recorded data often are transformed to various domains to discriminate against coherent and incoherent noise. For instance, by mapping a shot record from time-space domain to frequency-wavenumber domain, coherent linear noise can be attenuated. Similarly, by mapping a common-midpoint gather from time-space to time-velocity domain (velocity stacks) multiples are separated from primaries based on moveout discrimination. In these procedures the correct identification of seismic events with similar moveout can be severely affected by the aperture of the array and the discrete sampling of the wavefield.

Economic and/or logistic reasons usually dictate the cable length and spatial sampling of the seismic experiment. This thesis examines how the resolution (the ability to distinguish close events) of slant stack and parabolic stack operators deteriorates under limited aperture. An algorithm is developed to increase the resolution of the aforementioned operators. This procedure constructs an operator that collapses each seismic signal in the transform domain, thus diminishing truncation artifacts. The overall procedure is equivalent to the simulation of a longer array of receivers.

Slant stacks and the parabolic stacks are linear operations used to map the seismic data into another domain, the transform domain ( $\tau - p$  or  $\tau - q$ ). In this thesis an inverse problem is posed. This is accomplished by considering the data as the result of a linear operation onto the transform domain. This approach permits one to incorporate prior information into the problem which is utilized to attenuate truncation artifacts.

The prior information is incorporated into the inverse problem by means of the Bayesian formalism. The observational errors and the prior information are combined

through Bayes' rule using the likelihood function and a long tailed distribution, respectively. The posteriori probability is then used to induce the objective function of the problem. Finally, minimizing the objective function leads to the solution of the inverse problem. The advantage of incorporating a long tailed distribution to model the transform domain is that the solution is constrained to be sparse which is a desired feature for highly resolved models.

The method is also used to design an artifacts-reduced 2-D discrete Fourier transform. A by-product of the method is a high resolution periodogram. This periodogram coincides with the periodogram that would have been computed with a longer array of receivers if the data consist of a limited superposition of linear events.

## Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>List of Tables</b>	<b>vii</b>
<b>List of Figures</b>	<b>viii</b>
<b>Acknowledgement</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Motivation . . . . .	3
1.3 Scope and contribution of this work . . . . .	5
1.4 Thesis outline . . . . .	5
<b>2 Slant Stacks</b>	<b>7</b>
2.1 Introduction . . . . .	7
2.2 Slant stacks . . . . .	10
2.2.1 The slant stack operator (conventional definition) . . . . .	10
2.2.2 The inverse slant stack operator . . . . .	14
2.2.3 The sampling theorem for slant stacks . . . . .	16
2.3 Discrete slant stacks . . . . .	18
2.3.1 The discrete slant stack operator (conventional definition) . . . . .	20
2.3.2 The minimum norm discrete inverse slant stack operator . . . . .	21
2.3.3 Minimum norm solution with quadratic data constraints . . . . .	23

2.3.4	On the stability of the operator $\mathbf{G}$ . . . . .	26
2.3.5	The data and model resolution matrices . . . . .	35
2.3.6	Resolution analysis by means of the spectral expansion . . . . .	39
2.4	Synthetic data examples . . . . .	42
2.5	Concluding remarks . . . . .	43
<b>3</b>	<b>Velocity stacks</b>	<b>48</b>
3.1	Introduction . . . . .	48
3.2	Velocity stacks . . . . .	51
3.3	Inverse velocity stacks . . . . .	53
3.4	Discrete parabolic Radon transform . . . . .	55
3.5	Sampling considerations . . . . .	57
3.6	Example . . . . .	58
3.7	The validity of the parabolic approximation . . . . .	59
<b>4</b>	<b>High Resolution Slant Stack and Parabolic Stacks Operators</b>	<b>63</b>
4.1	Introduction . . . . .	63
4.2	Why sparse models? . . . . .	66
4.3	Inverse problems and Bayes' rule . . . . .	69
4.3.1	Bayes' Rule . . . . .	69
4.3.2	Bayesian approach to inverse problems . . . . .	70
4.3.3	Assigning a prior to the model . . . . .	73
4.3.4	Assigning a probability to the noise . . . . .	76
4.3.5	Maximum a Posteriori Solution (MAP) . . . . .	78
4.3.6	Algorithm complexity . . . . .	81
4.3.7	Convergence . . . . .	82
4.3.8	Convexity of the objective function . . . . .	83

4.3.9	Derivation of the damped least squares solution . . . . .	85
4.4	Application to slant stacks . . . . .	86
4.4.1	Slant stacking non-aliased data . . . . .	86
4.4.2	Slant stacking beyond the alias . . . . .	86
4.4.3	High resolution wavefield decomposition of VSP data . . . . .	90
4.5	Application to velocity processing . . . . .	98
4.5.1	Identification of multiples . . . . .	98
4.5.2	Offset space reconstruction from parabolic stacks . . . . .	102
4.5.3	Field data example . . . . .	107
4.6	Concluding remarks and discussion . . . . .	115
<b>5</b>	<b>Aperture Compensated Discrete Fourier Transform and Applications</b>	<b>116</b>
5.1	Introduction . . . . .	116
5.2	The discrete Fourier transform as an inverse problem . . . . .	118
5.3	Zero order quadratic regularization (damped least squares) . . . . .	120
5.4	Regularization by the Cauchy-Gauss model . . . . .	122
5.5	Hybrid two-dimensional estimator of the DFT . . . . .	124
5.5.1	First example: spatio-temporal spectrum narrow band signals . .	125
5.5.2	Second example: broad band applications . . . . .	126
5.6	Application to field data. Vertical Seismic Profiling (VSP) . . . . .	129
5.7	Discussion . . . . .	133
<b>6</b>	<b>Summary</b>	<b>137</b>
	<b>Appendices</b>	<b>148</b>
	<b>A Scalar products</b>	<b>148</b>

## List of Tables

2.1	Eigenvalues of $\phi(l - l')$ , $l, l' = 0, \dots, N - 1$ for $W = 0.2, 0.3, 0.4$ . . . . .	31
4.1	Parameters used to generate the synthetic data. . . . .	86

## List of Figures

2.1	(a) values of $\lambda_k(11, W)$ and (b) $\lambda_k(25, W)$ for $W = 0.1, 0.2, 0.3, 0.4$ . . . .	30
2.2	Synthetic data. . . . .	44
2.3	(a) $\tau - p$ panel computed with the conventional slant stack operator. (b), (c) and (d) $\tau - p$ space computed with the minimum norm slant stack operator. . . . .	45
2.4	(a) Power spectrum of the 25 Hz Ricker wavelet used in the synthetic example. (b) Condition number of the matrix $g(l - l'), l, l' = 0, \dots, 10$ . . .	46
2.5	Minimum and maximum eigenvalues values of $fg(l - l'), l, l' = 0, \dots, 10$ . . . . .	47
3.1	Synthetic data. . . . .	59
3.2	Velocity processing. (a) Velocity stack computed using the conventional operator (parabolic moveout and stacking). (b) Velocity panel computed using the minimum norm inverse operator. . . . .	60
4.1	Linear Radon transform. Conventional and inverse operators . . . . .	87
4.2	Synthetic example. . . . .	90
4.3	Slant stack computed with the conventional and with inverse operator. . . . .	91
4.4	Contour plot of the normalized envelope of the panels showed in Figure 4.3 . . . . .	92
4.5	Blowup of figure 4.4. . . . .	93
4.6	(a) The synthetic wavefield. Noisy case. . . . .	94
4.7	The $\tau - p$ panel of the noisy wavefield. . . . .	94
4.8	The $\tau - p$ space computed with inverse slant stack operator derived from the Cauchy prior . . . . .	95

4.9	Misfit function versus frequency. . . . .	96
4.10	VSP data and $\tau - p$ panels derived using damped least squares and the IMRLS. . . . .	97
4.11	Velocity processing with the conventional and inverse parabolic stack operator. . . . .	99
4.12	A synthetic CMP gather with three primaries events and multiples . . . .	100
4.13	Velocity processing with the conventional parabolic stack and with the IMRLS algorithm. . . . .	101
4.14	Parabolic processing after a $t^2$ transformation. . . . .	103
4.15	Offset space reconstruction. . . . .	104
4.16	Parabolic processing after a NMO correction. . . . .	105
4.17	Offset space reconstruction. . . . .	106
4.18	CSP gather with missing traces . . . . .	109
4.19	Parabolic transform processing after a $t^2$ transformation. . . . .	110
4.20	Parabolic transform processing after a $t^2$ transformation. IMRLS solution. . . . .	111
4.21	CSP gather with missing traces before and after NMO correction. . . . .	112
4.22	Parabolic transform processing after a NMO correction. . . . .	113
4.23	Parabolic transform processing after a NMO correction. IMRLS solution . . . . .	114
5.1	2-D spectrum of three narrow band signals . . . . .	127
5.2	a) Synthetic wavefield. . . . .	129
5.2	b-g) Spatio-temporal spectrum using the windowed DFT and the high resolution DFT. . . . .	130
5.2	i-h) Wavefield extrapolation. . . . .	131
5.3	a) Synthetic wavefield contaminated with random noise . . . . .	131

5.3	b-g) Spatio-temporal spectrum using the windowed DFT and the high resolution DFT. . . . .	132
5.4	Segment of a VSP. . . . .	133
5.5	2-D spectral signature of the VSP in the interval 500 – 800m. . . . .	134
5.6	2-D spectral signature of the VSP in the interval 800 – 1200m. . . . .	134
5.7	2-D spectral signature of the VSP in the interval 1200 – 1600m. . . . .	135

## Acknowledgement

*A miei genitori, ricordando con riconoscenza quanto amore mi venne dalla sua comprensione.*

*A Luciana y a Federico, gracias por estar siempre a mi lado durante estos años de estudio.*

I am indebted to many people. In particular, to my supervisor, Tad Ulrych, for his friendship, encouragement, and for sharing his compulsory passion for time series analysis.

My thanks also to Drs. M. Bostock, and L. Kirlin for their considerable input during the preparation of this thesis. My thanks also to Drs. G. K. C. Clarke, and L. Smith for their participation as committee member and university examiner, respectively. Moreover, I thank Drs. O. Yilmaz (external examiner) and C. Kostov for their very constructive remarks. An appreciation is given to Dr. Doug Oldenburg. His clear inverse theory viewpoint always came to my mind during the preparation of this thesis.

The friendship of Xingong Li has been very important since my very first day at UBC. The valuable input of Ken Matson in real seismic exploration problems is also appreciated.

The research that leads to this thesis was supported by an NSERC grant, a University Research Grant from Imperial Oil Ltd., and by a University Graduate Fellowship from the University of British Columbia.

## Chapter 1

### Introduction

*It was the furthest point of navigation and the culminating point of my experience.*

Joseph Conrad: *Heart of Darkness*

#### 1.1 Background

Limited aperture is a long standing problem in exploration seismology. Geophysicists are constantly dealing with a finite amount of data and consequently with truncation artifacts. An important part of the seismic technology for oil exploration still relies on two basic procedures: velocity analysis and stacking. With the development of new technologies and with the necessity of revealing more complicated seismic structures, high resolution techniques for velocity processing and multiple attenuation have become an important requirement.

Thorson and Claerbout (1985) were the first to note a need for high resolution techniques for velocity processing to cope with finite amount of data in CMP processing. They address the problem by developing an inverse technique to retrieve velocity panels from CMP data. In another scenario, the pioneering work of John P. Burg (1975) on maximum entropy spectral analysis (MESA), shows how to estimate spectral estimators without imposing any implicit assumption about the samples of the time series that were

not recorded. Burg's technique assumes that a finite length time series is only a circumstantial outcome of a longer process and therefore, it is not convenient to assume that values outside the time series gate are zero (missing samples). The Maximum Entropy principle (Jaynes, 1967) is the elegant mechanism utilized by Burg to avoid drawing any assumption about the missing information.

Slant stacking and parabolic stacking entail two-dimensional operators routinely used to process seismic data. The resolution of these operators is affected by the aperture of the seismic array. In this context, the aperture plays a role equivalent to the length of the time series in time series analysis. Slant stack and parabolic stack operators are two members of the generalized discrete Radon transform (Beylkin, 1987).

Slant stacking is usually used to decompose and/or filter signals that exhibit linear moveout. An example is vertical seismic profiling (VSP), where the composite seismogram is a superposition of seismic events with constant ray parameter. Slant stacks also perform an important role in wave equation theory, since they may be related to plane wave decomposition (Treitel et al., 1982).

The parabolic transform maps a parabola into a point in the transform domain, although this is only true when the aperture of the array is infinite. If the seismic data consist of a superposition of parabolas, then the parabolic transform will decompose the seismogram into a superposition of isolated points. However, limited aperture introduces amplitude smearing into the transform domain and therefore the resolution or focusing power of the transform is deteriorated. The same reasoning is applicable in time series analysis where truncation introduces sidelobes and, consequently, difficulty arises in the identification of close spectral peaks.

In CMP or shot gathers the seismic signal is organized in hyperbolic paths. In order to apply the parabolic transform to decompose these hyperbolas, the parabolic model

must be first satisfied. Hampson (1986) proposed to validate the parabolic approximation by means of a normal moveout (NMO) correction. After NMO, there is a range of hyperbolic events that may be accurately approximated by parabolic paths and therefore the conditions to successfully apply the parabolic transformation are satisfied. Yilmaz (1989) proposed to model CMP gathers after a  $t^2$  transformation of the temporal axis. Hampson (1986) and Yilmaz (1989) posed the problem in the frequency-offset ( $\omega - h$ ) domain. The latter permits the problem to be broken down into several smaller inverse problems, one at each frequency in the seismic band. These researchers have adopted zero order quadratic regularization or damped least squares (Lines and Treitel, 1984; Titterton, 1985) to design the parabolic operator. In this thesis, it is shown that damped least squares is not necessarily the best means by which to solve the problem. In fact, damped least squares imposes smoothness onto the transform domain and consequently, instead of sharpening up the image (enhance the focusing power of the transform), it tends to blur and introduce smearing into the transform domain.

## 1.2 Motivation

My interest in this topic arose after investigating together with T.J. Ulrych a technique to estimate a side-lobe-free 1-D discrete Fourier transform, a problem originally suggested to us by Colin Walker. There are many approaches to estimate high resolution spectral estimates based on the idea of side lobe suppression. However, none of these approaches deals with the problem of estimating a high resolution discrete Fourier transform, and hence none is capable of estimating the spectral amplitude as well as the phase of the signal. The spectral estimation problem can be posed as a linear inverse problem where the target is the power spectral density that honors a few lags of the autocorrelation function (Priestly, 1982). However, since the autocorrelation is a phase-less function, only

the spectral density can be recovered. To retrieve a high resolution estimate of amplitude and phase, instead of using autocorrelation constraints, we considered data constraints. A non-parametric method to estimate the high resolution DFT was developed by posing the problem as an inverse problem and using a regularization approach derived from a Bayesian description of the problem in probability space. Two dimensional applications were developed to estimate the spatial spectral signature of narrow and broad band seismic data (Sacchi and Ulrych, 1995a).

My attention then became focused on slant stacks and parabolic stacks. In the frequency-offset space, the Radon operator looks very similar to the transformation matrix used to retrieve the DFT. The regularization scheme developed to estimate the side-lobes free DFT was re-examined to compute slant stacks and parabolic stacks free of truncation artifacts.

The primary results of the regularization scheme devised to cope with limited aperture were presented by myself at the 64th Annual meeting of the Society of Exploration Geophysicists in Los Angeles, USA (Sacchi and Ulrych, 1994), later expanded and published (Sacchi and Ulrych, 1995b and c). These works were primarily focused on inverse velocity stacking, a name usually reserved for Thorson's (Thorson and Claerbout, 1985) technique.

Although the ultimate goal of my research involves the estimation of aperture compensated slant stack and parabolic stack operators, there are other areas where I found that these algorithms may be applied. As geophysicists we live in a truncated world. We are always faced with the problem of data coverage. A simple deconvolution problem may be posed as an aperture problem where from a few samples of the central band of the seismic signal we attempt to reconstruct the missing or deteriorated low and high frequencies. It is in these types of situations that the correct regularization of the inverse problem plays a crucial role. In particular, I would like to mention the problem of

impedance reconstruction from band-limited seismic data, extensively studied by many researchers (Oldenburg et al., 1983; Walker and Ulrych, 1983).

### 1.3 Scope and contribution of this work

The purpose of this thesis is to investigate new means of estimating Radon operators in situations where the physical size of the seismic survey (cable length) inhibits the correct identification of close signals. The proposed methodology does not replace the existing techniques to compute 2-D linear transforms but, rather, complements them.

This thesis provides a novel inverse procedure to compute Radon operators. The Bayesian framework is adopted to derive the regularization strategy that is used to solve the inverse problem. The regularization term is derived by assuming a long tailed pdf to model the prior distribution of parameters.

The inverse procedure is also used to estimate a high resolution 2-D discrete Fourier transform. The goal is to estimate from a limited number of receivers the 2-D spectral signature of a group of events that are recorded on a linear array of receivers.

### 1.4 Thesis outline

The linear Radon transform or the slant stack transform is presented in Chapter 2. The stability/resolution tradeoff is also studied. This chapter details some pivotal concepts that are used throughout this thesis.

Parabolic stacks as an alternative procedure for velocity processing are presented in Chapter 3. The validity of the parabolic approximation by means of a time stretching transformation and a normal moveout correction is also studied.

Chapter 4 introduces a new procedure to enhance the resolution of slant stack and parabolic stack operators. The feasibility of the technique to process seismic data is

examined with synthetic and real field examples.

In Chapter 5, the regularization scheme presented in chapter 4 is used to estimate 2-D spectral estimators. This problem may appear rather disconnected from the rest of the thesis. However, the aperture problem has a significant impact in spectral analysis problems. The reader will find that many of the ideas developed to treat aperture limitations in Radon operators may also be applied in a spectral analysis scenario.

Finally, in Chapter 6, the core of the thesis is reviewed. Two important aspects are examined. First, the potential strength of the techniques to process real data when dealing with severe aperture limitations. Secondly, I provide a discussion on the main limitations of the techniques.

## Chapter 2

### Slant Stacks

*Y aquí me despido yo  
Que he relatao a mi modo,  
Males que conocen todos  
Pero que naidés contó*

José Hernández: *Martín Fierro*

#### 2.1 Introduction

Different techniques have been devised to identify and/or filter linear events. Generally, they have the following common framework. First, they assume that a set of linear events are recorded on an array with discrete and limited coverage. Secondly, they assume that the noise is uncorrelated with the signals. In geophysics, linear event identification has been an active field of research. Two classic examples are vertical seismic profiles (VSP) (Hardage, 1985) and slowness vector estimation in seismographic arrays for earthquake detection and location (Goldstein and Archuleta, 1987). In VSP processing, linear event detection-estimation is used to identify and separate the principal components of the VSP data: the up-going and the down-going waves.

A general strategy for event identification-estimation involves the following approach. First, the data are transformed to a new domain where each component may be isolated. Then, after masking the undesired components, the data are mapped back to the original

domain retaining only the desired information.

The so-called covariance methods, spectral matrix methods or eigenstructure methods, that have been borrowed from the field of sonar processing (Bienvenu and Kopp, 1983; Wax et al., 1984), exploit the separability of the signal and noise covariance matrices. Although they have been developed to process narrow band signals, covariance methods can be applied to broad band applications by carrying out the processing at each narrow band frequency that comprises the broad band. These techniques have been successfully applied in VSP processing (Mari and Glangeaud, 1990; Mari and Gavin, 1990) and in wavefield decomposition of multi-component records (Rutty and Jackson, 1992). To fully separate two signals of different slowness by means of the spectral matrix approach, the signals must be uncorrelated. In other words, the inner product of their steering vectors must vanish. This is a very restrictive assumption that usually does not hold. This short-coming, however, can be overcome by un-correlating the signals by means of a frequency-spatial averaging procedure. The latter implies that some type of a priori information about the dips and the extension of the frequency smoothing must be provided (Mari and Gavin, 1990). Freire and Ulrych (1988) have applied the singular value decomposition (SVD) to separate down and up-going waves in VSP processing. Their technique is another way of interpreting the Karhunen-Loève (KL) transform introduced in geophysics by Hemon and Mace (1978) and extensively studied by Jones (1985). Similar to the spectral matrix approach, the SVD or KL method requires prior knowledge of the dip of the target signal. Recently, Ulrych et al. (1995), have developed a hybrid technique to cope with the aforesaid problem. First, they apply an eigenvalue based metric (Pisarenko, 1972) to identify the dip of each primary wavefield: up-going and down-going waves. Then, this information is used to rotate the wavefield target to simulate a signal with infinite apparent velocity. Finally, the packing property for flat events of the SVD or KL transform (Jones, 1985) is used to isolate the target wavefield.

The Radon transform does not require any a priori knowledge of the apparent velocity of each linear component, but the performance of the transform is severely affected by the aperture of the array and by spatial undersampling (alias). In seismic processing, the Radon transform is commonly known as the  $\tau - p$  ( $\tau$  denotes time and  $p$  ray parameter) or slant stack transform. The original idea developed by Radon in 1917 (Deans, 1983), has provided a basic framework for many problems of image reconstruction in physics, astronomy, medicine, optics, nondestructive testing, and geophysics. In image processing, it is also called the Hough transform (Pratt, 1991), which may be regarded as a transformation of a line in Cartesian coordinate space to a point in polar coordinate space.

In geophysics, the properties of the Radon transform were examined by Phinney et al. (1981), Durrani and Bisset (1984) and Tatham (1984). Chapman (1981) developed exact formulas for a point source in cartesian or spherical coordinates, and for a line source in cylindrical coordinates. The relationship between the Radon transform and the plane wave decomposition is also well established (Stoffa et al., 1981; Treitel et al., 1982). Least squares procedures to compute the Radon transform were investigated by Thorson and Claerbout (1985), Beylkin (1987) and Kostov (1990). These authors showed how to mitigate the smearing caused by the finite aperture. Recently, Zhou and Greenhalgh (1994) linked the least squares solution to  $p$ -dependent Wiener filters. These researchers derived the slant stack formulas in the continuous domain, but the resulting algorithms are identical to those obtained by other researchers (Beylkin, 1987; Kostov, 1990).

In order to avoid the inversion of prohibitively large matrices the problem may be posed in the frequency-space domain ( $f - x$ ). This technique was adopted by Beylkin (1987), Kostov (1990), Foster and Mosher (1992), and recently by Zhou and Greenhalgh (1994). This allows us to solve several small problems in the band that comprises the signal. Some stability concerns arise when the problem is tackled in this manner.

Particularly, a least squares solution can be extremely unstable at low frequencies. In addition, it is interesting that slant stacks can be also computed in the time-space domain. Thorson and Claerbout (1985) and, recently Yilmaz and Tanner (1994), have presented high resolution least squares slant stack operators designed in time-space domain. Their procedures use an iterative inversion scheme especially devised to solve large linear sparse operators. Thorson and Claerbout (1985) have also shown how to update in each iteration the variances of the model to drive the solution to minimum entropy. Yilmaz and Taner (1994) have also developed an interesting scheme based on fuzzy logic to mitigate the alias.

This chapter is organized as follows. First, I provide the basic definitions of the continuous Radon transform in the  $x-t$  and  $\tau-p$  planes. Then, two different approaches are used to formulate the Radon pair. The first definition, the conventional slant stack operator, uses the forward mapping to compute the Radon space and a deconvolution operator to map the  $\tau-p$  space into the  $x-t$  space. The second definition, the inverse slant stack operator, uses a deconvolution operator to compute the Radon space and a simple mapping to recover the  $x-t$  space. Discrete formulas for the aforesaid operators are also derived. In the discrete case, I analyse two crucial problems that constantly arise in geophysics: stability and resolution. These problems are formally examined using prolate spheroidal discrete sequences (Slepian and Sonenblick, 1965). Finally, the theoretical analysis is illustrated with numerical simulations.

## 2.2 Slant stacks

### 2.2.1 The slant stack operator (conventional definition)

Let  $u(h, t)$  represent a seismic signal. Throughout this thesis variable  $t$  designates the time and  $h$  the offset or range. For a continuous array we define the slant stack by means

of the following transformation

$$v(p, \tau) = (\mathcal{L}u)(p, \tau) = \int_{-\infty}^{\infty} u(h, t = \tau + hp) dh. \quad (2.1)$$

Where  $p$  and  $\tau$  denote the slope or ray parameter and the intercept time, respectively.  $v(p, \tau)$  is used to designate the signal in the  $\tau - p$  domain. The adjoint transform  $\mathcal{L}^*$  is given by

$$\tilde{u}(h, t) = (\mathcal{L}^*v)(p, \tau) = \int_{-\infty}^{\infty} v(p, t = \tau - hp) dp. \quad (2.2)$$

In the frequency domain, the pair of transformations are given by,

$$V(p, \omega) = \int_{-\infty}^{\infty} U(h, \omega) e^{i\omega p h} dh, \quad (2.3)$$

$$\tilde{U}(h, \omega) = \int_{-\infty}^{\infty} V(p, \omega) e^{-i\omega p h} dp, \quad (2.4)$$

substituting, (2.3) into (2.4) yields

$$\tilde{U}(h, \omega) = \int_{-\infty}^{\infty} U(h', \omega) \int_{-\infty}^{\infty} e^{-i\omega p(h-h')} dp dh' \quad (2.5)$$

which may be written as follows

$$\tilde{U}(h, \omega) = U(h, \omega) * \rho(h, \omega), \quad (2.6)$$

where  $*$  denotes convolution and the function  $\rho$  is given by

$$\rho(h, \omega) = \int_{-\infty}^{\infty} e^{-i\omega p h} dp, \quad (2.7)$$

making the substitution  $z = -\omega p$  equation (2.7) becomes

$$\rho(h, \omega) = \int_{-\infty}^{\infty} \frac{1}{|\omega|} e^{ihz} dz = \frac{2\pi}{|\omega|} \delta(h). \quad (2.8)$$

The convolution operator is a delta function with respect to the variable  $h$ . Using the property of the  $\delta$  function,

$$\begin{aligned} \tilde{U}(h, \omega) &= \frac{2\pi}{|\omega|} U(h, \omega) * \delta(h) \\ &= \frac{2\pi}{|\omega|} U(h, \omega) \end{aligned} \quad (2.9)$$

the inversion formula becomes,

$$U(h, \omega) = \frac{|\omega|}{2\pi} \tilde{U}(h, \omega). \quad (2.10)$$

The inverse is computed in two steps. First, the adjoint is used to evaluate  $\tilde{U}(h, \omega)$ . Then,  $\tilde{U}(h, \omega)$  is multiplied by the frequency response of the  $\rho$  filter. The conventional slant stack pair in the frequency domain results in,

$$\begin{aligned} V(p, \omega) &= \int_{-\infty}^{\infty} U(h, \omega) e^{i\omega p h} dh, \\ U(h, \omega) &= \frac{|\omega|}{2\pi} \int_{-\infty}^{\infty} V(p, \omega) e^{-i\omega p h} dp. \end{aligned} \quad (2.11)$$

Now, consider that the range of  $p$  is a finite interval,  $p \in [-P, P]$ . This case leads to the following  $\rho$  filter,

$$\rho(h, \omega) = \int_{-P}^P e^{-i\omega p h} dp = 2P \frac{\sin(\omega P h)}{\omega P h}. \quad (2.12)$$

Substituting (2.12) in (2.6),

$$\tilde{U}(h, \omega) = 2P \int_{-\infty}^{\infty} U(h', \omega) \frac{\sin(\omega P(h - h'))}{\omega P(h - h')} dh'. \quad (2.13)$$

It is evident that the data may be recovered after solving a deconvolution problem. Spatial deconvolution is required since the infinite range of the variable  $p$  is truncated to a finite range. The wavenumber response of the  $\rho$  filter has the following expression:

$$\begin{aligned} \rho(k, \omega) &= \int_{-\infty}^{\infty} \rho(h, \omega) e^{ikh} dh \\ &= \int_{-\infty}^{\infty} \int_{-P}^P e^{-i(\omega p - k)h} dh dp \\ &= \int_{-P}^P \delta(\omega p - k) dp \\ &= \frac{1}{\omega} \int_{-\omega P}^{\omega P} \delta(k' - k) dk' \\ &= \begin{cases} \frac{1}{\omega} & k \leq |\omega P| \\ 0 & , \text{ otherwise} \end{cases} \end{aligned} \quad (2.14)$$

According to the last equation, the spatial deconvolution will be unstable if the wavenumbers in the data lie outside the range  $[-\omega P, \omega P]$ . Equation (2.14) also shows that the deconvolution is unstable at low frequencies.

### 2.2.2 The inverse slant stack operator

The definition of the forward slant stack operator and its adjoint may be changed to construct another slant stack pair,

$$u(h, t) = (\mathcal{L}^* v)(p, \tau) = \int_{-\infty}^{-\infty} v(h, t = \tau - hp) dp \quad (2.15)$$

$$\tilde{v}(p, \tau) = (\mathcal{L} u)(p, \tau) = \int_{-\infty}^{\infty} u(h, t = \tau + hp) dh, \quad (2.16)$$

the pair of transformations can be posed in the frequency-offset domain,

$$U(h, \omega) = \int_{-\infty}^{\infty} V(p, \omega) e^{-i\omega p h} dp, \quad (2.17)$$

$$\tilde{V}(p, \omega) = \int_{-\infty}^{\infty} U(h, \omega) e^{i\omega p h} dh. \quad (2.18)$$

Substituting, (2.17) into (2.18) yields,

$$\tilde{V}(p, \omega) = \int_{-\infty}^{\infty} V(p', \omega) \int_{-\infty}^{\infty} e^{-i\omega h(p-p')} dh dp', \quad (2.19)$$

where now, the convolution is with respect to the variable  $p$ , and the convolutional operator is given by

$$\gamma(p, \omega) = \int_{-\infty}^{\infty} \frac{1}{|\omega|} e^{i\omega p h} dh = \frac{2\pi}{|\omega|} \delta(p). \quad (2.20)$$

The  $\gamma$  filter is a delta function with respect to the variable  $p$ . Therefore, equation (2.19) becomes,

$$\tilde{V}(p, \omega) = \frac{2\pi}{|\omega|} V(p, \omega) \quad (2.21)$$

or equivalently

$$V(p, \omega) = \frac{|\omega|}{2\pi} \tilde{V}(p, \omega). \quad (2.22)$$

From the above derivation, it is clear that the  $\rho$  and the  $\gamma$  filters have the same frequency response. Finally, the slant stack pair becomes,

$$\begin{aligned} V(p, \omega) &= \frac{|\omega|}{2\pi} \int_{-\infty}^{\infty} U(h, \omega) e^{i\omega p h} dh, \\ U(h, \omega) &= \int_{-\infty}^{\infty} V(p, \omega) e^{-i\omega p h} dp. \end{aligned} \quad (2.23)$$

Assuming that  $h \in [-H, H]$  (finite aperture), the  $\gamma$  filter has the following structure

$$\gamma(p, \omega) = \int_{-H}^H e^{i\omega p h} dh = 2H \frac{\sin(\omega H p)}{\omega H p}. \quad (2.24)$$

Hence,  $V(p, \omega)$  may be calculated by solving the following integral equation,

$$\tilde{V}(p, \omega) = 2H \int_{-\infty}^{\infty} V(p', \omega) \frac{\sin(\omega H(p - p'))}{\omega H(p - p')} dp'. \quad (2.25)$$

After a comparison of the slant stacks pairs, equations (2.11) and (2.23), it is clear that a deconvolution procedure is required in both cases. In the conventional slant stack transform, the deconvolution is necessary to recover the data from the  $\tau - p$  space. In the inverse slant stack operator the deconvolution process is required to estimate the  $\tau - p$  space. The truncation effect of the variable  $p$  may be alleviated by choosing the proper region of support of the transform. The truncation of the variable  $h$  is associated with the resolution of the transform and cannot be alleviated by simple means. Generally, both the variables  $h$  and  $p$  are truncated. Thus, deconvolution should be carried out in both the forward and inverse transform (Zhou and Greenhalgh, 1994). However, the range of  $p$  may be chosen in such a way that most of the energy in the signal lies within this range.

### 2.2.3 The sampling theorem for slant stacks

Assuming that the wavefield is evenly sampled according to  $U(n\Delta h, \omega)$ ,  $n = 0, \pm 1, \pm 2, \dots$ , the relationship between the  $\tau - p$  and the  $h - t$  spaces is given by

$$U(n\Delta h, \omega) = \frac{|\omega|}{2\pi} \int_{-\infty}^{\infty} V(p, \omega) e^{-i\omega p n \Delta h} dp, \quad (2.26)$$

where  $V(p, \omega)$  denotes the slant stack corresponding to a continuous wavefield  $U(h, \omega)$ .

The integration domain can be decomposed into small subdomains as follows,

$$\begin{aligned} U(n\Delta h, \omega) &= \frac{|\omega|}{2\pi} \sum_{k=-\infty}^{\infty} \int_{-(2k-1)\frac{\pi}{\omega\Delta h}}^{(2k+1)\frac{\pi}{\omega\Delta h}} V(p, \omega) e^{-i\omega p n \Delta h} dp \\ &= \frac{|\omega|}{2\pi} \sum_{k=-\infty}^{\infty} \int_{-\frac{\pi}{\omega\Delta h}}^{\frac{\pi}{\omega\Delta h}} V\left(p + 2k\frac{\pi}{\omega\Delta h}, \omega\right) e^{-i\omega\left(p + 2k\frac{\pi}{\omega\Delta h}\right)n\Delta h} dp \end{aligned} \quad (2.27)$$

since  $e^{-i2\pi nk} = 1 \forall n k$ , the last equation may be written in the following form

$$U(n\Delta h, \omega) = \frac{|\omega|}{2\pi} \int_{-\frac{\pi}{\omega\Delta h}}^{\frac{\pi}{\omega\Delta h}} V_d(p, \omega) e^{-i\omega p n \Delta h} dp, \quad (2.28)$$

where the relationship between the slant stack of the continuous signal and the one corresponding to the sampled wavefield,  $V_d(p, \omega)$ , is given by

$$V_d(p, \omega) = \sum_{k=-\infty}^{\infty} V(p + 2k \frac{\pi}{\omega\Delta h}, \omega). \quad (2.29)$$

Thus, the discrete signal has an  $\omega - p$  representation with support in the range  $p \in [-\frac{\pi}{\omega\Delta h}, \frac{\pi}{\omega\Delta h}]$ . The components with slope  $p - 2\frac{\pi}{\omega\Delta h}, p + 2\frac{\pi}{\omega\Delta h}, p - 4\frac{\pi}{\omega\Delta h}, p + 4\frac{\pi}{\omega\Delta h}, \dots$  will appear to have slope  $p$  and every slope outside the range  $(-\frac{\pi}{\omega\Delta h}, \frac{\pi}{\omega\Delta h})$  will have an alias inside this range. If the continuous signal has all the components inside that range, the aliased components do not exist and therefore we can write  $V_d(p, \omega) = V(p, \omega)$ . It is clear from the above discussion that spatial sampling must be chosen so as to avoid the aliasing effect. If  $P = P_{max} = -P_{min}$ , the following relationship guarantees the absence of alias,

$$\Delta h \leq \frac{1}{2P f_{max}}, \quad (2.30)$$

where  $f_{max} = \omega_{max}/2\pi$  is the maximum temporal frequency of the seismic signal. The product  $P f_{max}$  is also the maximum wavenumber. Similarly, if  $\Delta h$  is given, the maximum ray parameter that can be retrieved without alias is given by

$$P_{max} = \frac{1}{2\Delta h f_{max}}. \quad (2.31)$$

For a non-symmetric slant stack,  $P_{max} \neq -P_{min}$ , equation (2.30) is modified as follows (Turner, 1990),

$$\Delta h \leq \frac{1}{P' f_{max}}, \quad (2.32)$$

where  $P' = |P_{max} - P_{min}|$ .

### 2.3 Discrete slant stacks

Discrete versions of equations (2.1) and (2.2) are obtained by replacing integrals by summations and imposing finite limits. First, assume that the seismogram contains  $N = L_f - L_n$  traces, where the indices  $L_f$  and  $L_n$  denote far and near offset traces respectively.

$$v(p, \tau) = (\mathcal{L}u)(p, \tau) = \sum_{l=L_n}^{L_f} u(h_l, \tau + h_l p) \Delta h_l, \quad (2.33)$$

where  $\Delta h_l = (h_{l+1} - h_l)$  for  $l = L_n, \dots, L_f - 1$ . Similarly, we approximate (2.2) by the following expression

$$\tilde{u}(h, t) = (\mathcal{L}^*v)(\tau, p) = \sum_{j=J_{min}}^{J_{max}} v(h, t - hp) \Delta p_j \quad (2.34)$$

where  $\Delta p_j = (p_{j+1} - p_j)$  for  $j = J_{min}, \dots, J_{max} - 1$ . Taking the Fourier transform of equations (2.31) and (2.32) yields

$$V(p, f) = \sum_{l=L_n}^{L_f} U(h_l, f) e^{2\pi i f h_l p} \Delta h_l \quad (2.35)$$

$$\tilde{U}(h, f) = \sum_{j=J_{min}}^{J_{max}} V(p, f) e^{-2\pi i f h p_j} \Delta p_j. \quad (2.36)$$

Using matrix notation it is possible to rewrite the slant stack and its adjoint as follows ( $f$  is omitted to avoid notational clutter),

$$\mathbf{v} = \mathbf{F} \mathbf{W}_u \mathbf{u} = \mathbf{L} \mathbf{u} \quad (2.37)$$

$$\tilde{\mathbf{u}} = \mathbf{F}^H \mathbf{W}_v \mathbf{v} = \mathbf{L}^* \mathbf{v} \quad (2.38)$$

where  $\mathbf{W}_u = \text{diag}[\Delta h_{L_n}, \dots, \Delta h_{L_f}]$  and  $\mathbf{W}_v = \text{diag}[\Delta p_{J_{min}}, \dots, \Delta p_{J_{max}}]$  account for the irregular geometry of the array and for irregular sampling of the variable  $p$ , respectively. The matrix  $\mathbf{F}$  has dimension  $(J_{max} - J_{min} + 1) \times (L_f - L_n + 1)$ , and elements given by

$$F_{j,l} = e^{-i2\pi f h_l p_j}. \quad (2.39)$$

The operators  $\mathbf{L}$  and  $\mathbf{L}^*$  form an adjoint pair. The matrix  $\mathbf{L}$  is the forward operator and  $\mathbf{L}^*$  denotes the adjoint operator. The following useful relationship is proved in Appendix A

$$\mathbf{L}^* = \mathbf{W}_u^{-1} \mathbf{L}^H \mathbf{W}_v. \quad (2.40)$$

This formula shows that, apart from a constant factor, the Hermitian transpose of the forward mapping coincides with the adjoint when the discretization is regular in both spaces.

### 2.3.1 The discrete slant stack operator (conventional definition)

The slant stack operator, equation (2.37), maps the  $t - x$  space into the  $\tau - p$  domain; the adjoint, equation (2.38), maps the  $\tau - p$  domain into the  $t - x$  domain. It is clear that since  $\mathbf{L}$  is non-orthogonal  $\mathbf{L}$  and  $\mathbf{L}^*$  do not constitute an inverse pair. Given  $\mathbf{v} = \mathbf{L}\mathbf{u}$ , the problem is how to recover  $\mathbf{u}$ . A relationship between  $\mathbf{u}$  and  $\tilde{\mathbf{u}}$  is obtained after substituting (2.37) into (2.38)

$$\tilde{\mathbf{u}} = \mathbf{L}^*\mathbf{L}\mathbf{u}. \quad (2.41)$$

Equation (2.41) is uniquely invertible in  $f \in B$  provided that  $\det(\mathbf{L}^*\mathbf{L}) \neq 0$  in the band  $B$ ,

$$\begin{aligned} \mathbf{u} &= (\mathbf{L}^*\mathbf{L})^{-1}\tilde{\mathbf{u}} \\ &= \mathbf{G}^{-1}\tilde{\mathbf{u}}. \end{aligned} \quad (2.42)$$

The  $N \times N$  matrix  $\mathbf{G} = \mathbf{L}^*\mathbf{L}$  represents a discrete version of the  $\rho$  filter. The pair of transformations which map a signal from  $f - h$  to  $f - p$  and vice-versa is given by

$$\begin{aligned} \mathbf{v} &= \mathbf{L}\mathbf{u} \\ \mathbf{u} &= \mathbf{G}^{-1}\mathbf{L}^*\mathbf{v}. \end{aligned} \quad (2.43)$$

The vector  $\mathbf{v}$  always exists since it is obtained by means of a simple mapping. Both expressions constitute an inverse pair when the inverse of  $\mathbf{G}$  exists. The forward and inverse pair does not permit to adequately model the signal when additive noise is present. If the data are contaminated with noise, the noise is mapped to the Radon domain.

### 2.3.2 The minimum norm discrete inverse slant stack operator

Thus far I have described the reconstruction of the data from the slant stack domain by means of Radon's inversion formula for discrete signals (2.11). The resolution of the Radon operator is mainly affected by the aperture of the array. This problem, however, can be diminished by recasting the problem as an inverse problem. Assuming that the data are the result of some transformation on the Radon domain, the problem is how to solve  $\mathbf{L}^*\mathbf{v} = \mathbf{u}$ . This is the discrete case associated with the continuous pair discussed in section 2.2.2:

$$\mathbf{u} = \mathbf{F}^H \mathbf{W}_v \mathbf{v} = \mathbf{L}^* \mathbf{v} \quad (2.44)$$

$$\tilde{\mathbf{v}} = \mathbf{F} \mathbf{W}_u \mathbf{u} = \mathbf{L} \mathbf{u}. \quad (2.45)$$

Two cases may arise depending of the number of unknowns and observations (traces). First, I will consider the underdetermined case (more unknowns than observations). Oversampling the Radon parameter  $p$  may lead to an underdetermined problem. The latter may help to remove ambiguities due to the discretization. Oversampling, however, as it is very well known, does not improve the resolution of the slant stack operator (Yilmaz, 1994). The underdetermined problem is solved by selecting a regularization strategy. A common means of computing a regularized solution is by the concept of the minimum norm. In effect, the solution chosen from among all the possible solutions that

fit the data is the one that has minimum energy (minimum Euclidean norm). This strategy is also called zero-order quadratic regularization (Titterton, 1985). The regularized solution is calculated after solving the following problem

$$\begin{aligned} &\text{minimize } \mathbf{v}^H \mathbf{W}_v \mathbf{v}, \\ &\text{subject to } \mathbf{L}^* \mathbf{v} = \mathbf{u}. \end{aligned} \tag{2.46}$$

Incorporating the vector of Lagrange multipliers,  $\boldsymbol{\lambda}$ , the objective function of the problem becomes

$$J = \mathbf{v}^H \mathbf{W}_v \mathbf{v} + \boldsymbol{\lambda}^H \mathbf{W}_u (\mathbf{L}^* \mathbf{v} - \mathbf{u}). \tag{2.47}$$

and minimizing  $J$  subject to data constraints leads to

$$\mathbf{v} = \mathbf{L}(\mathbf{L}^* \mathbf{L})^{-1} \mathbf{u}. \tag{2.48}$$

One finds that the operator  $\mathbf{L}$  (conventional definition) has been replaced by  $\mathbf{L}(\mathbf{L}^* \mathbf{L})^{-1}$ . The forward and inverse operators become

$$\begin{aligned} \mathbf{v} &= \mathbf{L} \mathbf{G}^{-1} \mathbf{u}, \\ \mathbf{u} &= \mathbf{L}^* \mathbf{v}. \end{aligned} \tag{2.49}$$

The Radon operator exists provided that the inverse of  $\mathbf{G}$  exists. The operator  $\mathbf{G}$  plays an important role in the definition of the Radon pair. In the conventional definition,  $\mathbf{G}^{-1}$

is used to recover the data from the Radon domain (2.43), while in the minimum norm solution,  $\mathbf{G}^{-1}$  is used to construct the Radon operator (2.49).

### 2.3.3 Minimum norm solution with quadratic data constraints

In the previous section the Radon operator was computed using the concept of minimum norm. In this section, I will show the relationship that exists between the minimum norm solution for underdetermined problems and the technique called damped least squares.

First, I will assume that the problem is overdetermined,  $M < N$ . This may appear rather misleading; the observation space has finite dimension (only  $N$  receivers are located in the field) while the model space, before discretization, lies in an infinite dimensional space. The least squares solution to the problem,  $\mathbf{L}^*\mathbf{v} = \mathbf{u}$ , is computed by minimizing the squared error  $S$ ,

$$\begin{aligned} S &= \mathbf{e}^H \mathbf{W}_u \mathbf{e} \\ &= (\mathbf{L}^*\mathbf{v} - \mathbf{u})^H \mathbf{W}_u (\mathbf{L}^*\mathbf{v} - \mathbf{u}). \end{aligned} \quad (2.50)$$

The minimization with respect to the complex vector  $\mathbf{v}$  yields,

$$\mathbf{L}^{*H} \mathbf{W}_u \mathbf{L}^* \mathbf{v} = \mathbf{L}^{*H} \mathbf{W}_u \mathbf{u}. \quad (2.51)$$

Combining the result proved in the Appendix A,

$$\mathbf{L}^{*H} = \mathbf{W}_v \mathbf{L} \mathbf{W}_u^{-1} \quad (2.52)$$

with equation (2.51) results in the following expression

$$\mathbf{LL}^*\mathbf{v} = \mathbf{Lu} . \quad (2.53)$$

Finally, assuming that the inverse of  $\mathbf{LL}^*$  exists, the least squares solution becomes

$$\mathbf{v} = (\mathbf{LL}^*)^{-1}\mathbf{Lu} . \quad (2.54)$$

Equation (2.54) is the unconstrained least squares solution. To avoid singularities the solution is usually stabilized by adding a small perturbation to the diagonal of  $\mathbf{LL}^*$ . This perturbation is called the damping factor or the ridge regression parameter. The statistical literature often refers to  $\beta$  as a smoothing parameter or a hyper-parameter. Formally, the damped least squares solution may be derived by minimizing the following objective function (Lines and Treitel, 1984),

$$J_{ls} = \beta \mathbf{v}^H \mathbf{W}_v \mathbf{v} + S , \quad (2.55)$$

after taking derivatives and equating to zero

$$\mathbf{v} = (\beta \mathbf{I} + \mathbf{LL}^*)^{-1}\mathbf{Lu} . \quad (2.56)$$

The parameter  $\beta^{-1}$  is also the Lagrange multiplier of a minimum norm problem with a single quadratic constraint given by  $S$ .

Now, considering the following identities,

$$\begin{aligned}\beta\mathbf{L} + \mathbf{L}\mathbf{L}^*\mathbf{L} &= \mathbf{L}(\beta\mathbf{I} + \mathbf{L}^*\mathbf{L}) \\ &= (\beta\mathbf{I} + \mathbf{L}\mathbf{L}^*)\mathbf{L},\end{aligned}\tag{2.57}$$

and since the matrices  $(\beta\mathbf{I} + \mathbf{L}^*\mathbf{L})$  and  $(\beta\mathbf{I} + \mathbf{L}\mathbf{L}^*)$  are positive definite provided that  $\beta > 0$ , the following identity is valid

$$(\beta\mathbf{I} + \mathbf{L}\mathbf{L}^*)^{-1}\mathbf{L} = \mathbf{L}(\beta\mathbf{I} + \mathbf{L}^*\mathbf{L})^{-1},\tag{2.58}$$

and when combined with equation (2.56) yields

$$\mathbf{v} = \mathbf{L}(\beta\mathbf{I} + \mathbf{L}^*\mathbf{L})^{-1}\mathbf{u}.\tag{2.59}$$

Equations (2.56) and (2.59) may be used to either solve underdetermined and/or overdetermined problems. In general, it is more convenient to adopt equation (2.56) for overdetermined problems and equation (2.59) for the underdetermined case. This last assessment can be verified by considering the dimension of the matrices  $\mathbf{L}^*\mathbf{L}$  and  $\mathbf{L}\mathbf{L}^*$ .

If the matrix  $\mathbf{G} = \mathbf{L}^*\mathbf{L}$  has full rank (underdetermined problem), it is possible to choose  $\beta = 0$ . Equation (2.59) becomes the minimum norm solution with exact data constraints. It is interesting to note that a single quadratic constraint,  $S$ , may be used to obtain the minimum norm solution with  $N$  exact constraints. It appears that the problem formulated in (2.46) might be solved using the objective function  $J_{ls}$ , equation (2.55), with  $\beta = 0$ . This is not true. The trick is valid only when applied to the solution of the problem and not the objective function

At this point one would think that if the problem is underdetermined, we may be able to use equation (2.59) with  $\beta = 0$ . However, the matrix  $\mathbf{G} = \mathbf{L}^*\mathbf{L}$  may have eigenvalues that are very small and may even fall below machine precision. This effect is particularly drastic at low frequencies. Any attempt to invert  $\mathbf{G}$ , if it is possible, will lead to an erroneous result.

### 2.3.4 On the stability of the operator $\mathbf{G}$

Further insight into the stability of the slant stack pair is gained by considering the stability of the operator  $\mathbf{G}$ . The entire analysis is focused on two approaches to construct the slant stack pair. In the first definition of the Radon operator, the existence and the stability of Radon's inversion formula is subject to the existence and stability of the operator  $\mathbf{G}$ . In fact, only when  $\mathbf{G}^{-1}$  exists, are both transformations inverses of each other. This is also applicable to the alternative definition given in the preceding section using the concept of minimum norm. In the former situation, the stability of the operator  $\mathbf{G}$  must be carefully considered to reconstruct the data from the  $\tau - p$  space. In the latter, the stability of the operator must be taken into account to design the slant stack operator.

The elements of  $\mathbf{G}$  are given by

$$g_{ll'} = \Delta h_l \sum_{J=J_{\min}}^{J_{\max}} \Delta p_j e^{2\pi i f' p_j (h_l - h_{l'})}, \quad (2.60)$$

when  $\Delta h$  is constant the relative offset is given by  $h_l - h_{l'} = \Delta h(l - l')$ . Under this condition, the matrix  $\mathbf{G}$  reduces to a Toeplitz structure. It is natural therefore to write  $g_{l,l'}$  as

$$g_{ll'} = g(l - l'). \quad (2.61)$$

If  $\Delta h$  and  $\Delta p$  are constant, equation (2.60) corresponds to a geometric series with the following sum,

$$g_{ll'} = \Delta h \Delta p \frac{\sin\left(\frac{2\pi f \Delta p \Delta h (l-l')(J_{max}-J_{min}+1)}{2}\right)}{\sin\left(\frac{2\pi f \Delta p \Delta h (l-l')}{2}\right)} e^{i \frac{2\pi f \Delta p \Delta h (l-l')(J_{max}+J_{min})}{2}} \quad (2.62)$$

Letting  $J = J_{max} = -J_{min}$  and  $M = 2J + 1$ , where  $M$  is the number of slopes or unknowns and introducing the variable,  $\theta = 2\pi f \Delta p \Delta h (l - l')$ , equation (2.62) becomes

$$g_{ll'} = 2\pi \Delta h \Delta p D_J(\theta), \quad (2.63)$$

where the function  $D_J(\theta)$  is known as the Dirichlet kernel,

$$D_J(\theta) = \frac{1}{2\pi} \frac{\sin\left[\left(J + \frac{1}{2}\right)\theta\right]}{\sin(\theta/2)}. \quad (2.64)$$

The function  $D_J(\theta)$  plays an important role in Fourier analysis (Priestley, 1982). The discrete Fourier transform (DFT) of a finite length sequence is the convolution of the Dirichlet kernel with the DFT of the infinite sequence. Thus, the transform of the observed finite sequence is a distorted version of the infinite sequence transform. The kernel  $D_J(\theta)$  has a large peak at  $\theta = 0$  of magnitude  $(2J+1)/2\pi$ , together with secondary peaks at approximately  $\theta = \pm 5\pi/(2J+1), \pm 9\pi/(2J+1), \pm 13\pi/(2J+1), \dots$

Since  $\Delta p$  and  $(2J+1)$  are not independent, the following manipulation is convenient to continue with the analysis. Letting  $\Delta p = 2P/(2J+1)$  where  $2P$  denotes the region of support of the ray parameter, equation (2.62) becomes,

$$g(l-l') = \frac{2p\Delta h}{(2J+1)} \frac{\sin(2\pi f P \Delta h (l-l'))}{\sin(2\pi f P \Delta h \frac{(l-l')}{(2J+1)})}. \quad (2.65)$$

When the number of slopes  $2J+1$  goes to infinity, while  $P$  is bounded, the elements of  $\mathbf{G}$  reduce to

$$\begin{aligned} g(l-l') &= \frac{1}{f} \phi(l-l') \\ &= \frac{1}{f} \frac{\sin(2\pi W(l-l'))}{\pi(l-l')}, \end{aligned} \quad (2.66)$$

where  $W = Pf\Delta h$ . The matrix  $\phi(l-l')$ , which arises in maximum time-frequency concentration problems, has  $N$  eigenvectors which conform to the discrete prolate spheroidal sequences (Slepian, 1977). In time-frequency maximum concentration problems  $WN$  is also called the time-bandwidth product (Slepian and Pollack, 1961).

When  $W < 1/2$ , the eigenvalues of  $g(l-l')$  have the following properties:

**Property 1.** *The matrix  $\phi(l-l')$ ,  $l, l' = 0, \dots, N-1$  has only  $N$  non-zero eigenvalues. They are distinct, real and positive and we order them so that*

$$1 > \lambda_0(N, W) > \lambda_1(N, W) > \dots > \lambda_{N-1}(N, W) > 0.$$

**Property 2.** *The eigenvalues obey the symmetry law*

$$\lambda_k(N, \frac{1}{2} - W) = 1 - \lambda_{N-1-k}(N, W)$$

**Property 3.** For any arbitrarily small  $\epsilon > 0$  there exists a constant  $C_\epsilon$  so that

$$\lambda_k \geq 1 - \epsilon \text{ for } k \leq 2WN - C_\epsilon \ln \pi NW$$

$$1 - \epsilon \geq \lambda_k \geq \epsilon \text{ for } k \leq 2C_\epsilon \ln \pi NW$$

**Property 4.** [Slepian, 1977] As  $N \rightarrow \infty$ ,  $\lambda_k(N, W) \rightarrow 1$  if  $k = 2WN(1 - \epsilon)$ , while  $\lambda_k(N, W) \rightarrow 0$ , if  $k = 2WN(1 + \epsilon)$ . This is true for any  $\epsilon$  satisfying  $1 > \epsilon > 0$ . In fact, this is a consequence of property 3.

Note that all the eigenvalues are positive and different from zero which is a property of any Hermitian matrix (Strang, 1976). Strictly speaking, this does not guarantee the stability of the inverse, since several eigenvalues can be close to zero and even fall below machine precision. This is particularly true for large values of  $N$ . Table 2.1 shows the eigenvalues of the  $11 \times 11$   $\phi$  matrix for  $W = 0.2$ ,  $W = 0.3$  and  $W = 0.4$ . Figure 2.1 shows the dependence of the eigenvalues on  $W$  for  $N = 11$  and  $N = 25$ . For  $N > 25$  the eigenvalues cannot be computed by direct solution of an eigen-problem. In fact, asymptotic expressions must be used to cope with this situation (Slepian, 1977).

Property 4 highlights some interesting features of the eigen-spectra of the matrix  $\phi(l-l')$ . First, the eigenvalues stay close to 1 for small  $k$ , plunge to zero near the threshold  $2WN$ , and stay close to zero afterwards. Secondly, it is also interesting to note, that the width of the plunge region is proportional to  $\ln(\pi WN)$ . Since  $\lim_{x \rightarrow \infty} x^{-1} \ln(x) = 0$ , the width of the plunge region becomes small when compared with the threshold  $2NW$  when  $N$  is large.

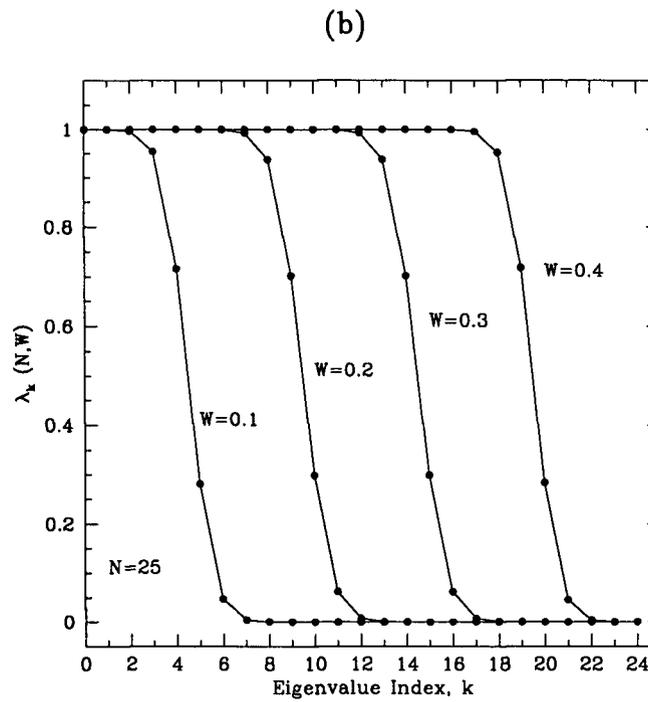
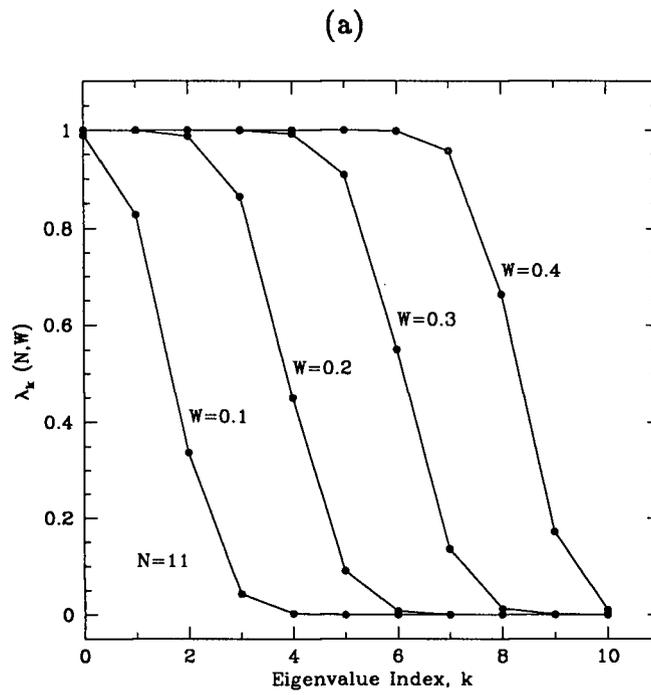


Figure 2.1: (a) values of  $\lambda_k(11, W)$  and (b)  $\lambda_k(25, W)$  for  $W = 0.1, 0.2, 0.3, 0.4$ .

$k$	$\lambda_k(11, 0.2)$	$\lambda_k(11, 0.3)$	$\lambda_k(11, 0.4)$
0	0.999992610514931D+00	0.999999999370405D+00	0.999999999999994D+00
1	0.999518220506777D+00	0.99999892249586D+00	0.99999999995432D+00
2	0.987710935929995D+00	0.999992005714193D+00	0.999999998709646D+00
3	0.864071477590623D+00	0.999677701763095D+00	0.999999993851705D+00
4	0.449739331222626D+00	0.992516733902020D+00	0.999999209925891D+00
5	0.911537572343498D-01	0.908846242765650D+00	0.999947990244767D+00
6	0.748326609798023D-02	0.550260668777374D+00	0.997947349605215D+00
7	0.322298227725263D-03	0.135928522409377D+00	0.957209444589136D+00
8	0.799428603213715D-05	0.122890640700014D-01	0.662847424914145D+00
9	0.104983211827608D-06	0.481779493220158D-03	0.171958069924268D+00
10	0.627678555715553D-09	0.738948506877286D-05	0.100905182398011D-01

Table 2.1: Eigenvalues of  $\phi(l - l')$ ,  $l, l' = 0, \dots, N - 1$  for  $W = 0.2, 0.3, 0.4$ 

The aforementioned properties are valid for  $W < 1/2$ . When  $W = n/2$ ,  $n \in \mathbb{Z}$  the matrix  $\phi(l - l')$  degenerates to the identity matrix,

$$\phi(l - l') = 2W\delta_{l,l'}. \quad (2.67)$$

In addition, when  $W > 1/2$  the following property is valid:

**Property 5.** When  $\frac{n}{2} > W' > \frac{(n+1)}{2}$ , with  $n \in \mathbb{Z}$

$$\lambda_k(N, W') = n + \lambda(N, W), \quad W = W' - \frac{n}{2}.$$

So far I have described some relevant properties of the eigen-spectra of the matrix  $\phi(l - l')$ . One can use these properties to infer a stability criterion for the inversion of the matrix  $\mathbf{G}$ . In particular, the condition number,  $C_g$ , of the matrix  $\mathbf{G}$  may be assessed by means of the condition number of the matrix  $\phi(l - l')$ ,  $C_\phi$ . The condition number is

defined as the ratio of the largest eigenvalue to the smallest eigenvalue. For a stability analysis, it is preferable to estimate the condition number since it is a scale-free criterion. Since the maximum eigenvalue is always bounded, it is also possible to draw similar conclusions by analysing the smallest eigenvalue of  $\phi(l - l')$ .

Three different situations may arise depending on the value of  $W$ . Using Properties 1, 2, 3, and 4, it immediately follows that

$$C_{\phi}(N, W) = \frac{\lambda_0(N, W)}{\lambda_{N-1}(N, W)} = \frac{1 - \lambda_{N-1}(N, \frac{1}{2} - W)}{\lambda_{N-1}(N, W)}, \quad W > \frac{1}{2}, \quad (2.68)$$

$$C_{\phi}(N, W) = 1, \quad W = \frac{n}{2}, \quad (2.69)$$

$$C_{\phi}(N, W') = \frac{n + \lambda_0(N, W)}{n + \lambda_{N-1}(N, W)}, \quad W = W' - \frac{n}{2}, \quad \frac{n}{2} < W' < \frac{n+1}{2}. \quad (2.70)$$

Accordingly to Property 3, when  $N$  is large, the condition number becomes

$$C_{\phi}(N, W) \rightarrow \infty, \quad W > \frac{1}{2}, \quad (2.71)$$

$$C_{\phi}(N, W) = 1, \quad W = \frac{n}{2}, \quad (2.72)$$

$$C_{\phi}(N, W') = \frac{n+1}{n}, \quad W = W' - \frac{n}{2}, \quad \frac{n}{2} < W' < \frac{n+1}{2}. \quad (2.73)$$

The condition number of  $\phi(l - l')$  may be computed by means of the condition number of the Toeplitz form  $g(l - l')$ . Since the condition number is a ratio of eigenvalues it is easy to see that

$$C_g(N, W) = C_\phi(N, W). \quad (2.74)$$

At this point, some comments follow immediately. First, the condition number can be prohibitively large when  $W < 1/2$ . Secondly, for values of  $W \geq 1/2$  the condition number stays close to 1. The latter may appear an advantage, but recalling the definition of the variable  $W$

$$W \geq \frac{1}{2} \Rightarrow Pf \geq \frac{1}{2\Delta h}, \quad (2.75)$$

the right hand side of the second inequality is the spatial wavenumber, while the left hand side is the maximum admissible wavenumber (Nyquist wavenumber). It is clear that when  $W \geq 1/2$  the transform is aliased. Consequently,  $W = 1/2$  is the upper limit that guarantees the absence of alias. The lower frequency limit is related to the stability of the inverse. Obviously, the inverse of  $\mathbf{G}$  does not exist at zero frequency. The latter implies that the DC level cannot be recovered (it is clear that the DC level lies in the null space of the problem.) Using continuity arguments, it follows that the inverse of  $g(l-l')$  is unstable at low frequencies. This situation is well reflected in the condition number.

At this point, some comments on the computation of the condition number are necessary. The most serious problem arises when large eigenvalues begin to cluster very close to 1. Since the values  $1 - \lambda_k$  fall below machine precision, eigen-problem routines cannot resolve the difference between two consecutive eigenvalues. This is particularly important when  $N$  is large. In this case the lower order eigenvalues cannot be computed by means of any eigenvalue routine. Slepian (1976) provided an asymptotic expansion for the eigenvalues of the matrix  $\phi(l-l')$  for fixed  $k$ , large  $N$  and  $W < 1/2$ . I am not

going to derive the following expression since it is beyond the scope of this thesis,

$$\begin{aligned}
 1 - \lambda_k(N, W) &= \pi^{\frac{1}{2}} (k!)^{-1} 2^{(14k+9)/4} [2 - \alpha]^{-(k+\frac{1}{2})} N^{k+\frac{1}{2}} e^{-\gamma N} \\
 \alpha &= 1 - \cos(2\pi W), \\
 \gamma &= \ln\left[1 + \frac{2\sqrt{\alpha}}{\sqrt{2}-\sqrt{\alpha}}\right].
 \end{aligned} \tag{2.76}$$

Thus, the condition number becomes

$$C_\phi(N, W) = \frac{1 - \pi^{\frac{1}{2}} 2^{9/4} [2 - \alpha]^{-(\frac{1}{2})} N^{\frac{1}{2}} e^{-\gamma N}}{\pi^{\frac{1}{2}} 2^{9/4} [2 - \tilde{\alpha}]^{-(\frac{1}{2})} N^{\frac{1}{2}} e^{-\tilde{\gamma} N}} \tag{2.77}$$

where  $\tilde{\alpha}$  and  $\tilde{\gamma}$  may be computed with  $\tilde{W} = 1/2 - W$ . The fit with  $\lambda_0$  is very good for  $N \geq 6$  and  $W > 0.1$ . In expression (2.77) I have replaced  $\lambda_{N-1}(N, W)$  by  $\lambda_0(N, \tilde{W})$  according to property 2. This function can be tabulated for different values of  $N, W$  to estimate the lower limit of  $W$  and hence, the low frequency cutoff,  $f_{min}$ , that guarantees the stability of the transform.

The inverse of  $\mathbf{G}$  may be stabilized using one of the following tricks:

- by shifting the signal to higher frequencies where the invertibility of  $g(l - l')$  is guaranteed (Beylkin, 1987). Unfortunately, some wavenumbers may be moved above the Nyquist limit.
- by adding a small perturbation to the principal diagonal of  $g(l - l')$ .

In this thesis, I have adopted the second strategy. Generally, a small perturbation to the diagonal of  $\mathbf{G}$  is sufficient to guarantee the invertibility of the matrix. If  $\beta$  denotes a positive scalar, the perturbed matrix is written as

$$g(l-l') = \frac{1}{f}\phi(l-l') + \beta\delta_{l,l'}, \quad l, l' = 0, N-1. \quad (2.78)$$

The condition number of the perturbed matrix becomes,

$$C_g(N, W) = \frac{\lambda_o(N, W) + f\beta}{\lambda_{N-1}(N, W) + f\beta}. \quad (2.79)$$

Note that, according to Property 5, when  $N$  is large the condition number reduces to

$$C_g = (1 + f\beta)/f\beta, \quad (2.80)$$

it is clear that  $C_g$  increases when  $f$  decreases and that the effect of  $\beta$  is dominant at low frequencies.

### 2.3.5 The data and model resolution matrices

Consider an inverse operator that solves the problem  $\mathbf{u} = \mathbf{L}^*\mathbf{v}$ . An estimate of the model parameters can be written as  $\mathbf{v}^{est} = \mathbf{L}^+\mathbf{u}$ , where  $\mathbf{L}^+$  denotes the inverse operator. Substituting  $\mathbf{v}^{est}$  into the equation  $\mathbf{L}^*\mathbf{v} = \mathbf{u}$  leads to the following relationship between the predicted data,  $\mathbf{u}^{pred}$  and the observed data,  $\mathbf{u}$ ,

$$\mathbf{u}^{pred} = \mathbf{L}^*\mathbf{v}^{est} = \mathbf{L}^*\mathbf{L}^+\mathbf{u} = \mathbf{R}_u\mathbf{u}. \quad (2.81)$$

The data resolution matrix,  $\mathbf{R}_u$ , describes how close the predicted data are to the actual observations. When  $\mathbf{R}_u = \mathbf{I}$ , the data are perfectly resolved.

Similarly, the model resolution matrix is defined by  $\mathbf{R}_v$ , where

$$\mathbf{v}^{est} = \mathbf{L}^+ \mathbf{L}^* \mathbf{v}^{true} = \mathbf{R}_v \mathbf{v}. \quad (2.82)$$

When  $\mathbf{R}_v = \mathbf{I}$  each model parameter is perfectly resolved. On the other hand, when  $\mathbf{R}_v \neq \mathbf{I}$  the estimated parameters are weighted averages of the "true model" parameters. In fact, one may define an effective averaging distance,  $d$ , that quantifies the departure of  $\mathbf{R}_v$  from the identity. This may be done by inspecting the width of the sidelobes of the matrix  $\mathbf{R}$ . As an example, consider two monochromatic linear events with slopes  $p_1$  and  $p_2$ . If  $d < |p_2 - p_1|$ , the slant stack operator can identify both events. Therefore,  $\mathbf{v}_{est}$  will exhibit two narrow peaks corresponding to the events with slopes  $p_1$  and  $p_2$ . Conversely, when this condition is not met, the slant stack operator cannot resolve both events, and the resulting model will show a broad single peak located somewhere between  $p_1$  and  $p_2$ .

Different measures of resolution have been studied in the context of time series and spectral analysis (Marple, 1986). I have found the idea of spreading functions to be very attractive, not only because it highlights some important aspects concerning the resolution of linear operators but because of its historical importance. The concept of spreading functions and averaging distance form part of the first formal approach to geophysical inverse problems (Backus and Gilbert, 1968). The spreading function may be defined as

$$\text{Spread}(\mathbf{R}_v) = \|\mathbf{I} - \mathbf{R}_v\|_F^2 = \sum_i \sum_j (I_{i,j} - R_{v,ij})^2, \quad (2.83)$$

where the subscript  $F$  stands for the Frobenious norm and  $\mathbf{I}$  for the identity matrix. Equation (2.83) is also called the Dirichlet spread function. Notice that when the model

resolution matrix approaches the identity matrix, the spreading function tends to zero. The foregoing discussion leads to two cases. First, consider the unconstrained least squares solution for the overdetermined, equation (2.54),

$$\begin{aligned} \mathbf{R}_v &= \mathbf{L}^+\mathbf{L}^* \\ &= (\mathbf{L}\mathbf{L}^*)^{-1}\mathbf{L}\mathbf{L}^* \\ &= \mathbf{I}, \end{aligned} \tag{2.84}$$

and consequently the spreading function becomes

$$\text{Spread}(\mathbf{R}_v) = 0. \tag{2.85}$$

In the underdetermined case, equation (2.48), the model resolution matrix becomes

$$\begin{aligned} \mathbf{R}_v &= \mathbf{L}^+\mathbf{L}^* \\ &= \mathbf{L}(\mathbf{L}^*\mathbf{L})^{-1}\mathbf{L}^*. \end{aligned} \tag{2.86}$$

It is straightforward to verify that the matrix  $\mathbf{I} - \mathbf{R}_v$  is idempotent, hence, the following properties may be used to derive an expression for the spreading function:

- If  $\mathbf{A}$  is idempotent,  $\mathbf{A}^2 = \mathbf{A}$ , (Strang, 1976). The Frobenious norm of  $\mathbf{A}$  is given by:

$$\|\mathbf{A}\|_F^2 = \text{Trace}(\mathbf{A}^H\mathbf{A}). \tag{2.87}$$

- For any idempotent matrix,

$$\text{Trace}(\mathbf{A}) = \text{Rank}(\mathbf{A}). \tag{2.88}$$

Letting  $\mathbf{A} = \mathbf{I} - \mathbf{R}_v$ , the spreading function becomes

$$\begin{aligned} \text{Spread}(\mathbf{R}_v) &= \text{Trace}(\mathbf{I}) - \text{Trace}(\mathbf{R}_v) \\ &= M - N. \end{aligned} \tag{2.89}$$

In the underdetermined problem the spreading function is equal to the difference between the number of unknowns (slopes) and the number of observations (traces). The preceding analysis suggests that the overdetermined option is better than the underdetermined one. However, any assessment based on spreading functions must be carefully considered. The model resolution matrix gives an input/output relationship between two vectors. However, it is obvious that the so called "true model" lies in a physical space of infinite dimension. The vectors  $\mathbf{v}$  and  $\mathbf{v}^{true}$  are a simple discrete representation of a physical space. A simple example may help to clarify the problem. Assume that a single linear event is recorded on an array of  $N = 3$  receivers. Then, according to the model resolution matrix, only  $M = 3$  slopes are sufficient to perfectly resolve the model. Obviously, this is only true if one of the tentative slopes  $p_i, i = 1, 2, 3$  coincides with the actual slope of the linear event. If this is not the case, the vector  $\mathbf{v}$  will have non-zero contributions in all of its entries.

A similar problem is encountered in harmonic retrieval from a finite length time series. In this problem  $M$  frequencies and the associated amplitudes and phases are retrieved from a complex time series of  $N$  samples ( $M < N$ .) The problem is usually solved in two steps. First the frequencies are estimated, then the amplitudes and phases using a least squares procedure or any other method that minimizes some measure of closeness between the observed and the predicted time series. The first step is crucial since an incorrect estimation of the frequencies will incorporate a bias in the estimation of the amplitudes and phases. It is clear that the first problem, the retrieval of the frequencies, is an

underdetermined problem (the frequency is a continuous variable) which may be solved by means of an estimate of the power spectral density (PSD). The key is that the PSD is computed by solving a linearly underdetermined problem where a few samples of the autocorrelation function are the data constraints. Two examples that come to my mind are the periodogram, that is the PSD estimate with minimum energy (minimum norm solution), and the maximum entropy PSD which maximizes the entropy of a Gaussian process (Burg, 1975).

### 2.3.6 Resolution analysis by means of the spectral expansion

First I will address the resolution problem for the conventional slant stack operator. To this end, I will assume that a monochromatic plane wave  $\mathbf{u}$  is recorded with a symmetric array of  $N = 2L + 1$  receivers. The ray parameter of the wave is denoted by  $p$ ,

$$u_l = e^{-i2\pi f p \Delta h (l-1)}. \quad (2.90)$$

Substituting the last equation into the expression for the conventional slant stack,  $\mathbf{L}\mathbf{v} = \mathbf{u}$ , we have

$$v_k = \Delta p \Delta h \sum_{l=-L}^L e^{i2\pi f (p-p_k) \Delta h l}. \quad (2.91)$$

This equation has been already met (equation (2.63)). The sum of the geometric series is replaced by a Dirichlet kernel,  $D_L(\theta)$ ,

$$v_k = 2\pi \Delta p \Delta h D_L(\theta) \quad (2.92)$$

where  $\theta = 2\pi f(p - p_k)\Delta h$ . Expression (2.92) has a peak at  $p = p_k$ , which corresponds to the slope of the plane wave. In the presence of many plane waves,  $v_k$  is a superposition of Dirichlet kernels. Assuming two plane waves with a slope difference of  $\delta p$ , one may ask the following question: what is the minimum  $\delta p$  that is discernible by the slant stack operator? One measure of resolution may be the first zero crossing,  $\theta = \pm 2\pi/(2L + 1)$ , of the Dirichlet kernel. Using this measure, it follows that

$$\begin{aligned}\delta p &= \frac{1}{f\Delta h(2L + 1)} \\ &= \frac{1}{f\text{Aperture}}.\end{aligned}\tag{2.93}$$

This relationship is also known in optics as ‘‘Rayleigh’s Criterion’’, (Born and Wolf, 1980), and simply states that the resolution is proportional to the inverse of the aperture.

When the slant stack is computed using the minimum norm criterion (2.49) the resolution analysis may be tackled as follows. When  $M$  is large, the following expression is valid (section 2.3.4)

$$\mathbf{G} = \frac{1}{f}\boldsymbol{\phi}.\tag{2.94}$$

The eigenvectors of  $\boldsymbol{\phi}$  are the prolate spheroidal discrete sequences. The eigenvalues of  $\boldsymbol{\phi}$  have been extensively studied in a preceding section. The eigenvalue problem is written as,

$$\frac{1}{f}\mathbf{G}\mathbf{r}_k(N, W) = \lambda_k(N, W)\mathbf{r}_k(N, W).\tag{2.95}$$

A certain economy in notation is provided by defining,

$$\frac{1}{f}\mathbf{G} = \mathbf{R}\mathbf{\Lambda}\mathbf{R}^H, \quad (2.96)$$

where  $\mathbf{R}$  is an orthogonal matrix where each column is an eigenvector  $\mathbf{r}_k(N, W)$ , and  $\mathbf{\Lambda}$  is a diagonal matrix with elements given by  $\Lambda_{i,i} = \lambda_i(N, W)$ . Equation (2.96) is the spectral expansion of the Hermitian matrix  $f^{-1}\mathbf{G}$ . Since  $\mathbf{R}$  is orthogonal, the inverse of  $\mathbf{G}$  may be computed as

$$\mathbf{G}^{-1} = \mathbf{R}\mathbf{\Lambda}^{-1}\mathbf{R}^H. \quad (2.97)$$

Now, take for instance the minimum norm solution (2.49)

$$\begin{aligned} \mathbf{v} &= \mathbf{L}\mathbf{G}^{-1}\mathbf{u} \\ &= \frac{1}{f}\mathbf{L}\mathbf{R}\mathbf{\Lambda}^{-1}\mathbf{R}^H\mathbf{u} \\ &= \frac{1}{f}\mathbf{L}\sum_k \frac{1}{\lambda_k}\mathbf{r}_k(\mathbf{r}_k^H\mathbf{u}) \\ &= \mathbf{L}\sum_k \gamma_k\mathbf{r}_k, \end{aligned} \quad (2.98)$$

where the coefficients  $\gamma_k$  are given by

$$\gamma_k = \frac{\mathbf{r}_k^H\mathbf{u}}{f\lambda_k}. \quad (2.99)$$

The key to understanding resolution is understanding the relationship between the magnitude of  $\lambda_k(N, W)$  and  $W$  and  $N$ . It has been shown that the eigenvalues are bounded

above by one for all  $k$ , and plunge to zero for values of  $k \geq 2WN$ . Beyond  $k = 2WN$ , the eigenvalues fall quickly at a fairly constant rate, such that  $\lambda_k/\lambda_{k+1}$  is approximately equal to ten (Slepian and Sonnenblick, 1965). Therefore, one may adopt  $k' = \text{int}[2WN]$  as the maximum eigenvalue to be used in (2.98). An interesting feature of the prolate spheroidal sequence,  $\mathbf{r}_k$ , is that it may be approximated by a sinusoidal function in terms of the average distance between zero crossings.

## 2.4 Synthetic data examples

A linear event impinging on a uniform array of receivers is used to study the resolution-stability tradeoff. The array is composed of 15 receivers located 10 m apart. The near and far offsets are located at  $-70$  and  $70$  m, respectively. The apparent velocity of the seismic event is  $2000$  m/s. The seismic source corresponds to a zero phase Ricker wavelet with a central frequency of  $25$  Hz. Most of the spectral energy is contained in the interval  $0 - 50$  Hz. Spectral components which are located above  $50$  Hz are negligible. The data are shown in Figure 2.2. The  $\tau - p$  plane computed with the conventional definition is illustrated in Figure 2.3a. Similarly, the minimum norm slant stack operator is used to transform the data to the  $\tau - p$  domain. In both cases  $M = 41$  slopes were computed. The minimum norm slant stack panel was computed with three different values of pre-whitening,  $\beta = 1 \times 10^{-8}$ ,  $1 \times 10^{-3}$ ,  $1 \times 10^{-1}$ . The results are shown in Figures 2.3b, c and d. It is clear that the minimum norm solution provides more resolution of the seismic signal than the conventional slant stack procedure. It is also true that damping makes the minimum norm solution behave like the conventional definition. A significant part of the spectral energy is contained at low frequencies where the operator  $\mathbf{G}$  is unstable. This feature is shown in Figure 2.4 where the condition numbers computed using the minimum and maximum eigenvalue of the matrix  $\mathbf{G}$  before and after pre-whitening are

shown. It is important to notice that the behaviour of this curve may be predicted using asymptotic values derived from the prolate spheroidal discrete sequences. The cutoff frequency is  $f = 50\text{Hz}$  ( $W = 1/2$ ) that corresponds to the upper bound to avoid spatial alias.

To complete the analysis I plotted the minimum and maximum eigenvalues of the matrix  $g(l - l') \times f$ . In the limiting case when  $N$  is large, these eigenvalues are approximately equal to those of  $\phi(l - l')$ . The result is portrayed in Figure 2.5. The curve resembles the eigen-spectra of the matrix  $\phi(l - l')$ .

## 2.5 Concluding remarks

This section was devoted to the presentation of two methods of computing the discrete slant stack operator. The first one involves the conventional definition of the slant stack operator followed by the inversion of the operator  $\mathbf{G}$  to recover the data. The second solution involves interchanging the definition of the forward transform of the problem. In this case the forward transform maps the  $\tau - p$  panel into the data space. The problem may be posed as an underdetermined linear inverse problem which may be solved using the concept of minimum norm. It must be stressed that other regularization schemes may be used to tackle the problem.

It has been proven that the resolution of the conventional slant stack is proportional to the inverse of the aperture of the array (Rayleigh limit). In the minimum norm solution, the deconvolution of the matrix  $\mathbf{G}$  may help to improve the resolution. However, because of the rather peculiar eigen-spectra of  $\mathbf{G}$ , damping to stabilize the inverse is necessary. The stability is improved at the expense of resolution.

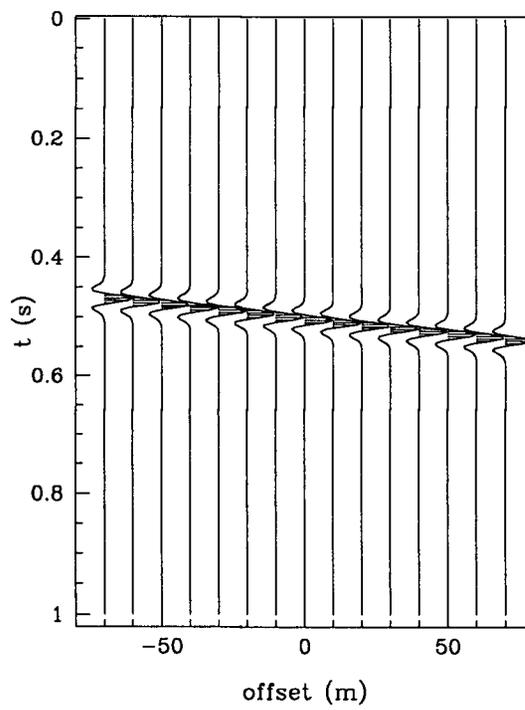


Figure 2.2: Synthetic data.

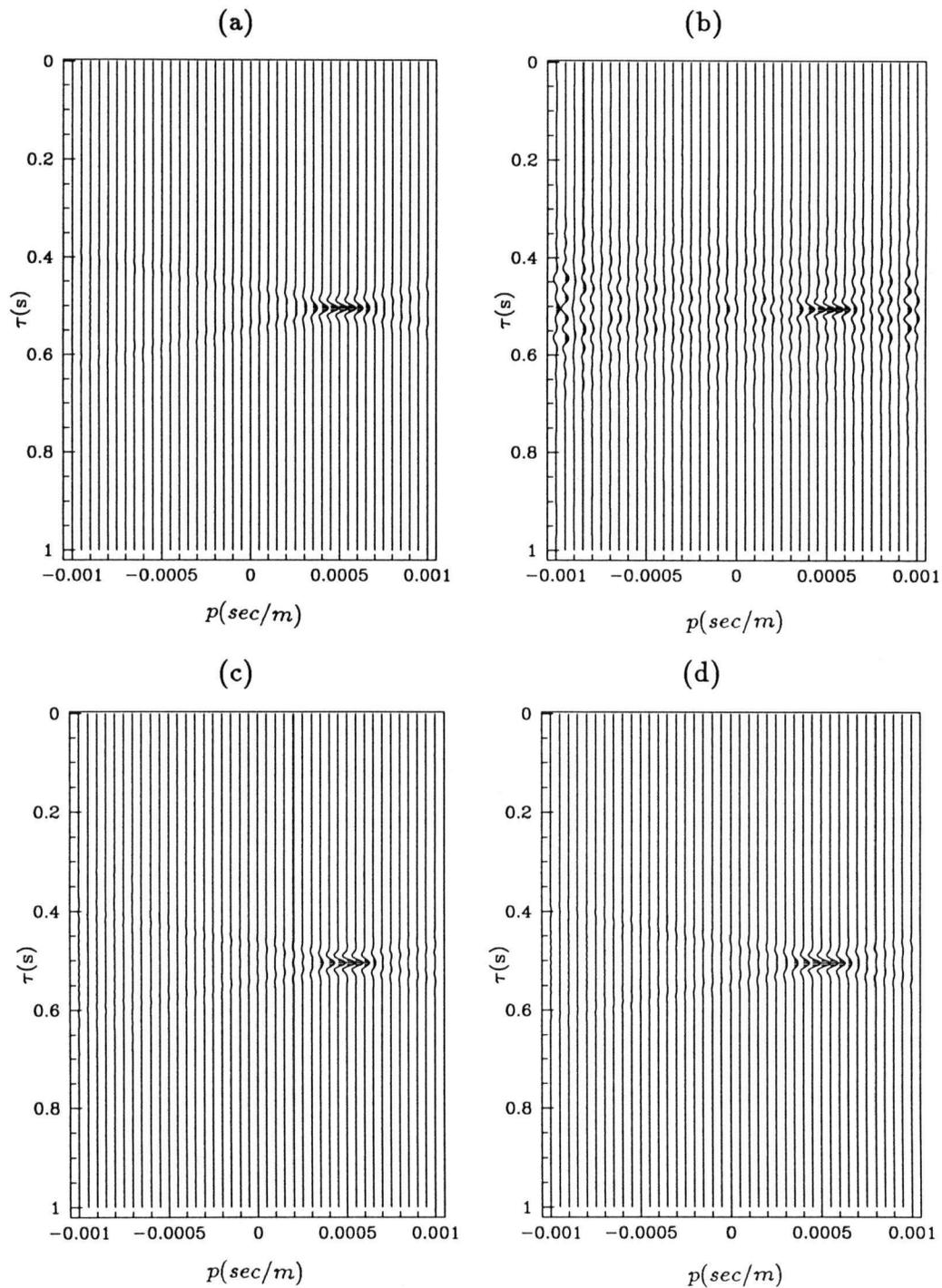


Figure 2.3: (a)  $\tau - p$  panel computed with the conventional slant stack operator. (b), (c) and (d)  $\tau - p$  space computed with the minimum norm slant stack operator with  $\beta = 1 \times 10^{-9}, 1 \times 10^{-3}$ , and  $1 \times 10^{-1}$ .

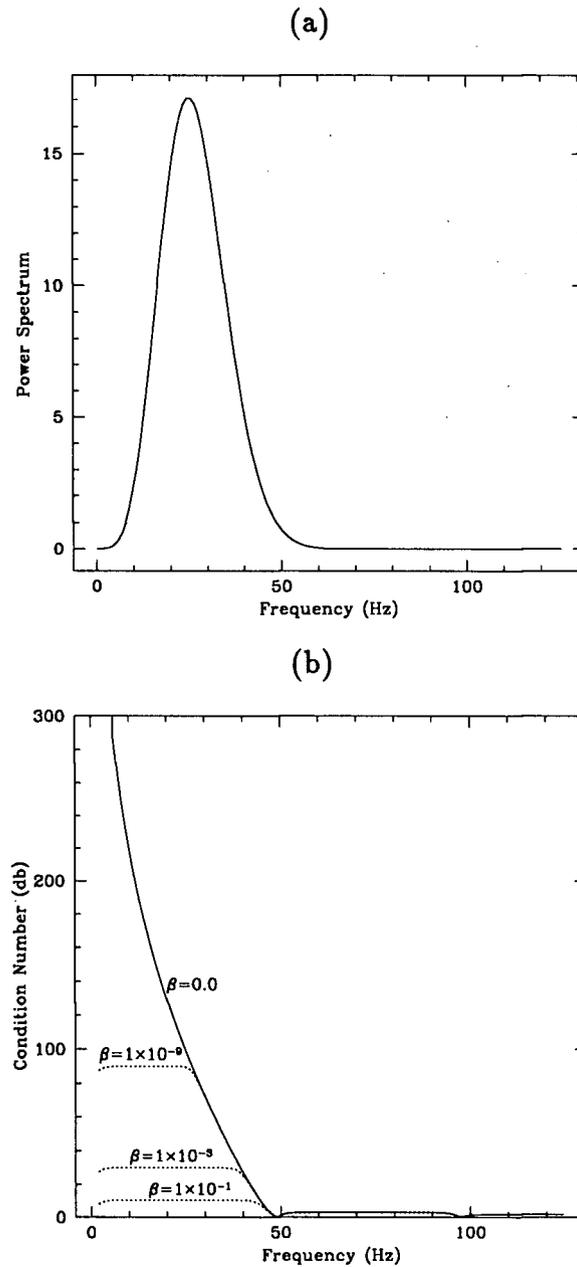


Figure 2.4: (a) Power spectrum of the 25 Hz Ricker wavelet used in the synthetic example. (b) Condition number of the matrix  $g(l-l')$ ,  $1, l'=0, \dots, 10$ . The continuous line corresponds to the condition number without damping. The dashed lines correspond to the condition number when different amounts of damping were applied. Frequencies above 50 Hz are spatially aliased.

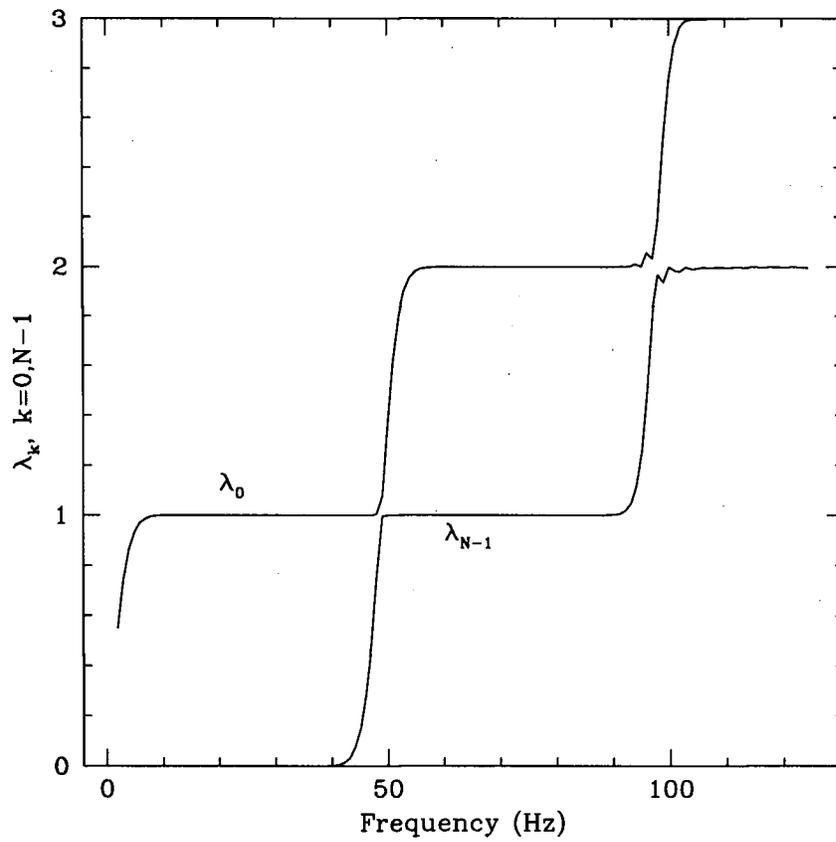


Figure 2.5: Minimum and maximum eigenvalues values of  $fg(l-l')$ ,  $l, l' = 0, \dots, 10$ . Note the strong similarity to the eigenvalues of the matrix  $\phi(l-l')$ .

## Chapter 3

### Velocity stacks

#### 3.1 Introduction

Conventional velocity analysis is performed by measuring energy along hyperbolic paths for a set of tentative velocities. The analysis of the results in the  $\tau - v$  (two way zero offset time and velocity) plane serves to estimate the stacking velocity that is later used to construct the zero offset section. The semblance (Neidell and Taner, 1971) is one of the most popular measures of coherent energy along hyperbolic trajectories in CMP gathers. The semblance measures the ratio of the signal energy within a window to the total energy in the window. Noise with non-zero mean and closely-spaced events in the same window deteriorates the velocity resolution when using this measure. The poor resolution of the semblance has led to more sophisticated techniques based on the eigenstructure of the data covariance matrix (Biondi and Kostov, 1989; Key and Smithson, 1990). In these techniques, the data covariance matrix is decomposed into signal and noise space contributions. Different metrics based on the eigenvector of the signal space are then used to measure coherent energy along hyperbolic paths.

The semblance, or any other velocity measure can be displayed as a contour map where each maximum corresponds to the kinematic pair  $\tau - v$ . Mapping the original data back from this space is not possible since these energy measures do not contain phase information. This short-coming may be overcome by using constant-velocity stack gathers that consist of constant-velocity CMP-stacked traces (Thorson and Claerbout,

1984). For an infinite-aperture array, the summation along hyperbolas should map into a point in the velocity space (for a spike-like wavelet). Limited aperture, however, spreads the information and not only makes the identification of each seismic event difficult, but also degrades any reconstruction of the offset space beyond the original aperture. Hence, there is a practical importance for estimating high-resolution or aperture-compensated velocity gathers.

To reduce amplitude smearing in the velocity space, Thorson and Claerbout (1985) performed the inversion of a set of constant-velocity stacks in  $t - h$  space. The problem involves the inversion of large matrices that is a difficult task. Thorson and Claerbout develop a stochastic inversion scheme that converges to a solution with minimum entropy. The latter is achieved by defining a Gaussian prior pdf with variable variance. To avoid the inversion of very large matrices, Yilmaz (1989) posed the problem in the  $f - h$  domain. Since velocity stacks are not time-invariant, he proposed a  $t^2$  transformation to force time-invariance. Time invariant operators can be easily posed in the frequency-offset space. Thus, instead of solving a large inverse problem, we can solve several small inverse problems at each frequency within the seismic band. After a  $t^2$  transformation summation along hyperbolas is replaced by summation along parabolas. Consequently, the hyperbolic Radon operator is replaced by an operator with integration along parabolic paths.

The parabolic transform was originally applied in a somewhat different manner by Hampson (1986) to attenuate multiple energy in CMP gathers. Hampson applied the parabolic transform to normal moveout (NMO) corrected CMP gathers. The residual moveout of a reflection, after NMO, can be approximated by a parabola. The multiples are isolated in the transform domain and after being transformed back to the offset domain they are subtracted from the original data. A similar transformation has been

proposed by Foster and Mosher (1992) who showed that the correct suppression of multiples is particularly important in some prestack processes like amplitude versus offset (AVO) analysis. Their transformation is well suited for target oriented problems. The focusing power of the transform depends on a tuning parameter that may be adjusted to isolate a particular hyperbolic event.

It is clear that the parabolic transform applied to NMO corrected CMP gathers is appropriate to model only energy that is organized along a parabolic path. At this point some remarks are in order. The focusing power of the method critically depends on how well the data match the Radon integration path. In parabolic stacking after NMO (Hampson, 1986) the hyperbolas can be accurately approximated by parabolas only for a limited range of time-velocity pairs. In other words, the focusing power of the transform is not guaranteed for the complete gather. On the other hand, if the parabolic path is validated by means of a  $t^2$  transformation any hyperbola can be converted into a parabola. Unfortunately, the  $t^2$  transformation introduces a non-linear compression of the time axis before 1 s and an expansion after 1 s. It is clear that, after the  $t^2$  transformation, the frequency decomposition is not completely valid.

This chapter is organized as follows. First, I provide analytical expressions for the forward and adjoint mappings. I assume that parabolic paths arise as a consequence of applying a  $t^2$  transformation to the CMP or CSP gather. The discrete forms are used, as in Chapter 2, to set the inverse problem in the frequency-offset space. Finally, I discuss under which conditions the NMO correction may be a good alternative to validate the parabolic approximation.

### 3.2 Velocity stacks

On CMP or CSP gathers seismic events are more likely to be hyperbolic than linear. The  $\tau - p$  transform maps hyperbolas into ellipses. This does not help much for reducing the support of the original signal in the  $\tau - p$  domain. Instead of a linear slant stack it is better to define a transform with a hyperbolic integration path. Let  $u(h, t)$  be the CMP or CSP gather and  $v(q, \tau)$  the velocity stack. For a continuous array the velocity stack is defined by the following transformation

$$v(q, \tau) = (\mathcal{L}u)(q, \tau) = \int_{-\infty}^{\infty} u(h, t = \sqrt{\tau^2 + h^2q})dh, \quad (3.1)$$

the adjoint transform  $\mathcal{L}^*$  is defined by

$$\tilde{u}(h, t) = (\mathcal{L}^*v)(q, \tau) = \int_{-\infty}^{\infty} v(h, \tau = \sqrt{t^2 - h^2q})dq. \quad (3.2)$$

The pair of transformations (3.1) and (3.2) is not time-invariant. This short-coming, however, may be overcome by introducing the following transformation,

$$t' = t^2 \quad \tau' = \tau^2, \quad (3.3)$$

this transformation compresses the time axis for samples before 1 s and stretches the samples after 1 s. The main problem of a  $t^2$  transformation is that shallow events can be severely aliased. However, this can be avoided by oversampling the time axis. A simple substitution shows that hyperbolic traveltimes are converted into parabolic traveltimes,

$$t^2 = \tau^2 + h^2 q \longrightarrow t' = \tau' + h^2 q. \quad (3.4)$$

It is possible, now, to write a new pair of transformations in stretched coordinates

$$v(q, \tau') = (\mathcal{L}u)(q, \tau') = \int_{-\infty}^{\infty} u(h, t' = \tau' + h^2 q) dh. \quad (3.5)$$

$$\tilde{u}(h, t') = (\mathcal{L}^*v)(q, \tau') = \int_{-\infty}^{\infty} v(h, \tau' = t' - h^2 q) dq. \quad (3.6)$$

The notation of  $u$ ,  $v$  and  $\mathcal{L}$  in expressions (3.1)-(3.2) and (3.5)-(3.6) has not been changed, but it is straightforward to see that they do not represent the same functions after the  $t^2$  transformation. The benefit of the  $t^2$  transform is that the pair of transformations (3.5) and (3.6) is time-invariant and therefore, the transformations can be expressed in the frequency-space domain,

$$V(q, \omega) = \int_{-\infty}^{\infty} U(h, \omega) e^{i\omega q h^2} dh, \quad (3.7)$$

$$\tilde{U}(h, \omega) = \int_{-\infty}^{\infty} V(q, \omega) e^{-i\omega q h^2} dq. \quad (3.8)$$

Substituting equation (3.7) into (3.8) leads to

$$\tilde{U}(h, \omega) = \int_{-\infty}^{\infty} U(h', \omega) \int_{-\infty}^{\infty} e^{-i\omega p(h^2 - h'^2)} dq dh'. \quad (3.9)$$

The last equation involves a particular type of convolution integral. The kernel in equation (3.9) is given by

$$\psi(h, \omega) = \int_{-\infty}^{\infty} e^{-i\omega h^2 q} dq. \quad (3.10)$$

If the parameter  $q$  is truncated  $q \in [-Q, Q]$ , the kernel  $\psi$  has the following form

$$\psi(h, \omega) = 2Q \frac{\sin(\omega h^2 Q)}{\omega h^2 Q}. \quad (3.11)$$

A special type of convolution integral must be solved in order to recover  $u(h, t')$ ,

$$\tilde{U}(h, \omega) = 2Q \int_{-\infty}^{\infty} U(h', \omega) \frac{\sin(\omega Q(h^2 - h'^2))}{\omega Q(h^2 - h'^2)} dh'. \quad (3.12)$$

### 3.3 Inverse velocity stacks

In chapter 2, I showed that the forward/inverse slant stack pair may be derived using two approaches by interchanging the definition of forward and adjoint operators. Similarly, we may want to define the following pair of forward and adjoint operators for the parabolic transform

$$u(h, t') = (\mathcal{L}^* v)(q, \tau') = \int_{-\infty}^{\infty} v(q, t' = \tau' - h^2 q) dq, \quad (3.13)$$

$$\tilde{v}(q, \tau') = (\mathcal{L} u)(q, \tau') = \int_{-\infty}^{\infty} u(h, t' = \tau' + h^2 q) dh. \quad (3.14)$$

In the frequency-offset space, these transformations become

$$U(h, \omega) = \int_{-\infty}^{\infty} V(q, \omega) e^{-i\omega q h^2} dq, \quad (3.15)$$

$$\tilde{V}(q, \omega) = \int_{-\infty}^{\infty} U(h, \omega) e^{i\omega q h^2} dh. \quad (3.16)$$

The substitution of equation (3.15) into (3.16) leads to

$$\tilde{V}(q, \omega) = \int_{-\infty}^{\infty} V(q', \omega) \int_{-\infty}^{\infty} e^{-i\omega h^2(q-q')} dh dq' \quad (3.17)$$

or equivalently,

$$\tilde{V}(q, \omega) = V(q, \omega) * \sigma(q, \omega). \quad (3.18)$$

Zhou and Greenhalgh (1994) showed that

$$\begin{aligned} \sigma(q, \omega) &= \int_{-\infty}^{\infty} e^{i\omega q h^2} dh \\ &= [1 + i \operatorname{sgn}(q)] \sqrt{\frac{\pi}{2\omega|q|}}. \end{aligned} \quad (3.19)$$

Deconvolution with respect to the variable  $q$  is also required when the aperture is infinite. This is an important distinction with respect to the linear Radon transform. Unfortunately, when  $h$  is truncated there is no analytical expression for  $\sigma(q, \omega)$ .

### 3.4 Discrete parabolic Radon transform

Discrete versions of equations (3.5) and (3.6) are obtained replacing the integral by summation and imposing finite limits. First assume that the CMP gather contains  $L_f - L_n + 1$  traces, where the indices  $L_f$  and  $L_n$  denote far and near offset traces respectively

$$v(q, \tau') = (\mathcal{L}u)(q, \tau') = \sum_{l=L_n}^{L_f} u(h_l, \tau' + h_l^2 q) \Delta h_l, \quad (3.20)$$

where  $\Delta h_l = h_{l+1} - h_l$  for  $l = L_n, \dots, L_f$ . Similarly, equation (3.6) is approximated by the following expression

$$\tilde{u}(h, t') = (\mathcal{L}^*v)(\tau', q) = \sum_{j=J_{min}}^{J_{max}} v(h, t' - h^2 q_j) \Delta q_j. \quad (3.21)$$

where  $\Delta q_j = q_{j+1} - q_j$  for  $j = J_{min}, \dots, J_{max}$ . Taking the Fourier transforms of (3.20) and (3.21) yields

$$V(q, f') = \sum_{l=L_n}^{L_f} U(h_l, f') e^{2\pi i f' h_l^2 q} \Delta h_l \quad (3.22)$$

$$\tilde{U}(h, f') = \sum_{j=J_{min}}^{J_{max}} V(q_j, f') e^{-2\pi i f' h^2 q_j} \Delta q_j. \quad (3.23)$$

Using matrix notation and suppressing the frequency dependence, equations (3.22) and (3.23) can be written as

$$\mathbf{v} = \mathbf{F}\mathbf{W}_u \mathbf{u} = \mathbf{L}\mathbf{u}, \quad (3.24)$$

$$\tilde{\mathbf{u}} = \mathbf{F}^H \mathbf{W}_v \mathbf{v} = \mathbf{L}^* \mathbf{v}, \quad (3.25)$$

where  $\mathbf{W}_u = \text{diag}[\Delta h_{L_n}, \dots, \Delta h_{L_f}]$  and  $\mathbf{W}_v = \text{diag}[\Delta q_{J_{min}}, \dots, \Delta q_{J_{max}}]$ . The matrix  $\mathbf{W}_u$  accounts for the irregular geometry of the array, while  $\mathbf{W}_v$  accounts for the irregular sampling of the inverse squared slowness  $q$ . The matrix  $\mathbf{F}$  has dimension  $(J_{max} - J_{min} + 1) \times (L_f - L_n + 1)$ , and elements given by

$$F_{l,j} = e^{i2\pi f' h_l^2 q_j}. \quad (3.26)$$

The reader will realize that an important part of the analysis presented in Chapter 2 also applies here. In fact, it is easy to derive inverse operators, by defining a new pair of forward and adjoint mappings as

$$\mathbf{u} = \mathbf{F}^H \mathbf{W}_v \mathbf{v} = \mathbf{L}^* \mathbf{v}, \quad (3.27)$$

$$\tilde{\mathbf{v}} = \mathbf{F} \mathbf{W}_u \mathbf{u} = \mathbf{L} \mathbf{u}. \quad (3.28)$$

The conventional operator computes the  $f'-q$  space after parabolic moveout and stacking (3.24). The data are recovered using the adjoint, equation (3.25), followed by the operator  $\mathbf{G}$ , as

$$\mathbf{u} = \mathbf{G}^{-1} \tilde{\mathbf{u}}. \quad (3.29)$$

In the second formulation, the  $f' - q$  space is obtained by solving equation (3.27). This equation may be solved with the zero-order quadratic regularization (damped least squares) procedures outlined in section 2.3.3. In contrast to the linear Radon transform, it is not possible to find an asymptotic formulation to assess the stability and the resolution of the operator. The elements of the operator  $\mathbf{G}$ , assuming constant  $\Delta q$  and  $\Delta h$ , are given by

$$g_{l'l'} = \Delta h \Delta q \sum_{j=J_{min}}^{J_{max}} e^{i2\pi f' \Delta q j \Delta h^2 (l^2 - l'^2)} \quad (3.30)$$

which has the following sum

$$g_{l'l'} = \Delta h \Delta q \frac{\sin\left(\frac{2\pi f' \Delta q \Delta h (l^2 - l'^2)(J_{max} - J_{min} + 1)}{2}\right)}{\sin\left(\frac{2\pi f' \Delta q \Delta h (l^2 - l'^2)}{2}\right)} e^{i\frac{2\pi f' \Delta q \Delta h (l^2 - l'^2)(J_{max} + J_{min})}{2}} \quad (3.31)$$

Gulunay (1988) has pointed out that equation (3.30) behaves like a Fresnel integral (Born and Wolf, 1980).

### 3.5 Sampling considerations

In section 2.2.3, I derived the sampling theorem for slant stacks. Unfortunately, it is not possible to derive an exact analytical expression for the parabolic stack (Hugonnet and Canadas, 1995). An approximate expression has been reported by Kabir and Verschuur (1995)

$$\Delta h \leq \frac{1}{2|q_{max} - q_{min}|f_{max}H}, \quad (3.32)$$

which has been obtained by making an analogy with the linear Radon transform. For the linear Radon transform we have,

$$\Delta h' \leq \frac{1}{|p_{max} - p_{min}| f_{max}}. \quad (3.33)$$

If we assume that the parabolic transform can be computed using the linear Radon transform after the variable transformation  $h' = h^2$ ,

$$\Delta h' = 2h \Delta h. \quad (3.34)$$

After replacing equation (3.34) into (3.33), equation (3.32) follows immediately.

### 3.6 Example

The following steps summarize the procedures involved in the computation of velocity stacks by means of the parabolic transform:

1. Stretch the time axis of each seismic trace.
2. Transform each trace to the frequency domain.
3. Apply the parabolic stack operator (conventional or inverse operator).
4. Apply the inverse Fourier transform to each trace.
5. undo the stretching of the time axis.

Figure 3.1 shows a hyperbolic event simulated with a velocity of 3000 m/sec. In figures 3.2a and 3.2b, the conventional velocity stack and the minimum norm velocity stack are portrayed. It is clear that the minimum norm velocity stack panel exhibits

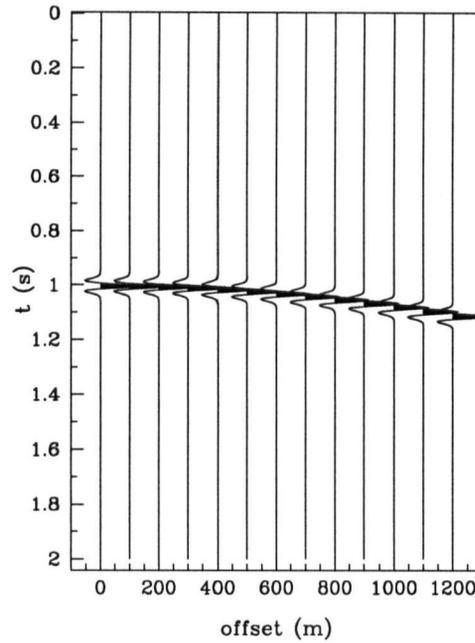


Figure 3.1: Synthetic data.

better resolution of the seismic event than the conventional panel computed with the adjoint operator.

### 3.7 The validity of the parabolic approximation

To ensure the improvement of the resolution of the parabolic transform, it is important to examine under what conditions the parabolic approximation is valid. Bear in mind that the NMO correction and the  $t^2$  transformation are tools to make the reflection paths time invariant. In other words, while a hyperbola will change if the time axis is translated, a parabola will exhibit the same moveout after a translation of the time axis. This important concept was used to set up our problem in the  $f - x$  space (see the transition from equations (3.20)-(3.21) to (3.22)-(3.23)).

The  $t^2$  transformation maps a hyperbola in  $t - x$  into a parabola in  $t^2 - x$ , regardless

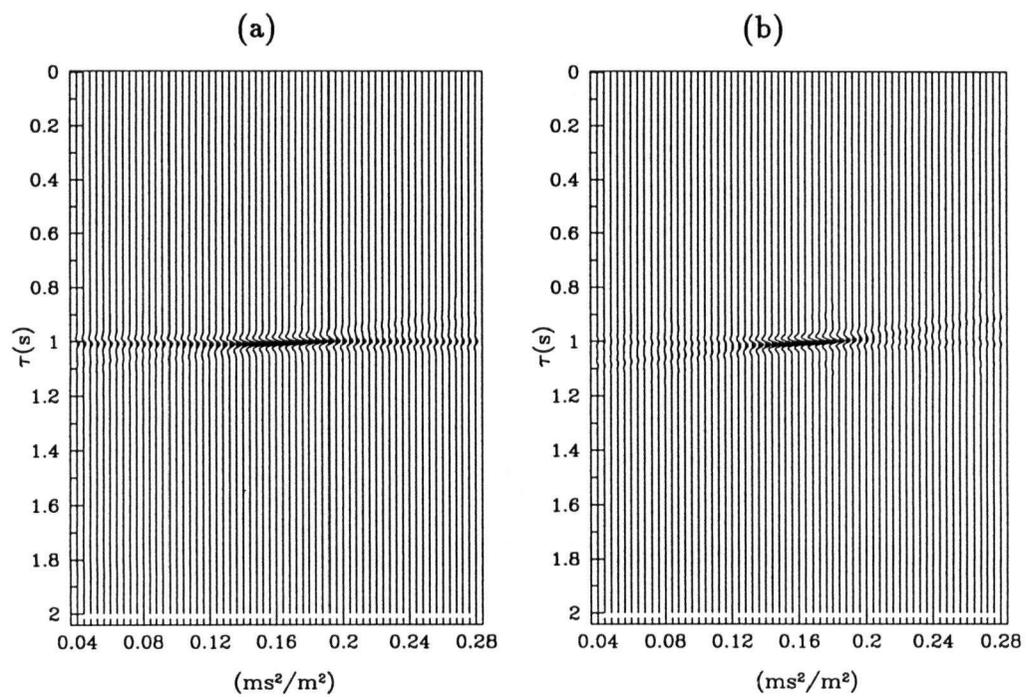


Figure 3.2: Velocity processing. (a) Velocity stack computed using the conventional operator (parabolic moveout and stacking). (b) Velocity panel computed using the minimum norm inverse operator.

of the position and velocity of the hyperbola. In other words, the transformation is valid for any  $t - v$  pair. Unfortunately, the frequency decomposition is not valid since the spectrum of the wavelet becomes time dependent. The latter was neglected in the formulation of the problem. The frequency distortion of shallow events that is present in CMP gathers reconstructed from velocity panels (Yilmaz, 1989; Sacchi and Ulrych, 1995a) can be explained as follows. To compute the velocity gather in  $\tau - q$  we must apply a  $\tau^{1/2}$  transformation to undo the original  $t^2$  transformation. The data before 1 s are stretched and the data after 1 s are compressed. Because of the alias introduced by the  $t^2$  transformation before 1 s, the reconstruction is not exact and leads to the aforementioned distortion.

The parabolic approximation can be also validated by means of an NMO correction (Hampson, 1986). Assuming that the stretching at far offsets after the correction is negligible, the spectral distortion of the signal may be very mild. In this case, the validity of the parabolic approximation depends on the position of each reflection in the velocity spectrum. If  $V$  denotes the velocity of a reflection and  $V_c$  the velocity used to perform the NMO correction, the corrected moveout can be approximated by

$$t(h) = t(0) + a_1 h^2 + a_2 h^4 + \dots \quad (3.35)$$

where the parameters  $a_1, a_2$  are given by

$$a_1 = \frac{1}{2t(0)} \left( \frac{1}{V^2} - \frac{1}{V_c^2} \right), \quad (3.36)$$

$$a_2 = \frac{1}{8t(0)^3} \left( \frac{1}{V^4} - \frac{1}{V_c^4} \right). \quad (3.37)$$

The parabolic approximation is valid provided that

$$\frac{h}{t(0)} \left( \frac{1}{V^2} - \frac{1}{V_c^2} \right)^{\frac{1}{2}} \ll 1. \quad (3.38)$$

If the last expression is satisfied, we can neglect the term  $a_2 h^4$  in equation (3.35) and consequently the parabolic approximation is valid. Such an approximation justifies identifying the Radon parameter  $q$  with the variable  $a_1$ .

It is easy to see that the approximation may fail for shallow reflections,  $t(0) \rightarrow 0$ . The correction,  $V_c$ , can be chosen to minimize the last inequality, but if there are two hyperbolas with different velocities but with similar  $t(0)$ 's, the approximation may not be completely satisfied for both hyperbolas.

Integration paths, other than the parabolic, may be more adequate to model hyperbolas after NMO correction, an example is  $\phi(h) = a_1 h^2 + a_2 h^4$ . However, the computational cost of a procedure to decompose CMP or CSP gathers in a  $\tau, a_1, a_2$  space must be carefully considered.

## Chapter 4

### High Resolution Slant Stack and Parabolic Stacks Operators

*Entia non sunt multiplicanda praeter necessitatem.* William of Occam

#### 4.1 Introduction

This chapter presents an inverse procedure to enhance the resolution of slant stacks and parabolic stacks. In particular, I replace the conventional slant stack and the parabolic stack operators by operators derived from the solution of an inverse problem that is solved using a regularization that serves to improve the focusing power of the operator. These operators seek a solution that resembles the one that might have been computed with the conventional one when the aperture of the array is infinite. Hence, the overall procedure is equivalent to simulating a longer array. Specifically, two sparseness criteria are proposed to regularize the inverse problem. These criteria reduce the support of each signal in the transform domain, and hence, the resolution or focusing power of the Radon operator is enhanced (Sacchi and Ulrych, 1994, 1995b).

It has already been mentioned that the problem of estimating a highly resolved Radon operator may be posed as an inverse problem. Inverse problems are naturally ill-posed since they do not satisfy the conditions of existence, uniqueness and stability of the solution (see, for example, Tikhonov and Goncharsky, 1987). The technique which permits the construction of a unique and stable solution by introducing some type of prior information is called regularization. In fact, a particular regularization by means of the minimum norm has already been discussed. The prior information can either be given

in a deterministic form or in a stochastic form. Positivity is a common deterministic constraint useful for solving a variety of inverse problems i.e. magnetic susceptibility inversion, density inversion, etc. A stochastic constraint assumes some relevant information about the model in terms of moments of corresponding distributions.

In many problems the regularization is carried out without consideration of the nature of the model we are seeking. That is the case of the widely used zero-order quadratic regularization. This method has the advantage of imposing smoothness to the model which is the common way to avoid the amplification of random errors associated with each observation. The well known pre-whitening technique used in spiking and predictive deconvolution is an example of quadratic regularization (Robinson and Treitel, 1980). However, in many situations one may wish to use other types of regularization which permit some relevant information about the model to be incorporated. If the additional information is in the form of a pdf it may be combined with the data likelihood using Bayes' rule.

The Radon operator is obtained by incorporating prior information in terms of a family of sparse priors modelled by a generalized Gaussian. In addition, the Cauchy density is used to induce another type of sparse prior. Both priors lead to solutions consisting of isolated wavelets in the transform domain while remaining consistent with the data. When the regularization is imposed by the Gaussian probability distribution the operator is identical to the so-called called stochastic inverse (Aki and Richards, 1980). The latter may be derived using arguments others than those provided by Bayes' rule. An example is the damped least square solution (Lines and Treitel, 1984), which may be either derived considering a probabilistic description of the problem or by modifying the eigen-spectra of the problem to make the inversion less sensitive to observational errors.

When prior information is introduced throughout a sparse regulariser not only is the focusing power of the transform improved, but also the alias may be drastically mitigated.

Therefore, this type of algorithm may be used to process real data where severe aperture limitation and undersampling inhibit the correct separation of seismic events with close moveout.

The first type of regularization discussed here is derived minimizing the  $l_p$  norm of the model subject to data constraints. From a Bayesian point of view, one can generate this regularization by assuming that the unknown parameters may be modelled by a generalized Gaussian distribution (Johnson and Kotz, 1970). In this case, two parameters are used to properly tune the inversion. A shape parameter that controls the tails of the distribution and a scale parameter that may be used to regulate the degree of fitting of the reconstructed data with the original observations. The shape parameter is crucial in the inversion of sparse models. An attractive feature of the generalized Gaussian distribution is that it may be used to provoke sparse ( $p \approx 1$ ) or smooth models ( $p = 2$ ) according to the selection of the shape parameter.

A second regularization is derived using the Cauchy pdf. This pdf has been used in robust inversion of seismic data by Crase et al. (1990) and by Amundsen (1991). It is important to point out that like the  $l_p$  regularization, the Cauchy regularization is used to diminish the influence of outlier or gross errors that may corrupt the result of the inversion. In this thesis, it is shown that these distributions lead to a regularization procedure that converges to a sparse solution. In summary, long tailed distributions are used to impose features onto the model and not to damp outliers as proposed by several authors (Gersztenkorn et al., 1986; Scales et al., 1988; Crase et al., 1990; Amundsen, 1991).

The prior pdf is combined with the data likelihood (pdf of the observational errors) using Bayes's rule. This permits the construction of the so-called Maximum Posteriori (MAP) solution of the inverse problem. The observational errors are modelled by means of a Gaussian distribution. The latter has some advantages. First, it provides a simple

description of the noise process. Secondly, a Gaussian likelihood leads to a MAP solution that resembles the least squares solution. In fact, the MAP solution is retrieved with an iteratively model re-weighted least squares (IMRLS) procedure. The latter is a least squares solution where the model is re-weighted using a non-linear functional that depends on the previous iteration. There is a strong conceptual similarity between this algorithm and iteratively re-weighted least squares (IRLS) (Scales et al., 1988). Still, it is important to mention that in IRLS the errors are re-weighted to damp the influence of outliers. In the IMRLS, the model parameters are re-weighted to impose sparseness on the solution.

Finally, I would like to point out that there are strong similarities between the regularization strategies proposed in this thesis and the problem of band-limited extrapolation (Cabrera and Parks, 1991). In this technique, a weighted norm is minimized honoring data constraints (e.g., samples of a time series). The idea is that, theoretically, a time-limited signal cannot have a band-limited spectrum and therefore the norm is used to iteratively band limit and shape the spectrum. The weighted norm incorporates the a priori information about the shape and the spectral support of the signal. The procedure involves the computation of the weights from the previous iteration. Recently, Kabir and Verschuur (1995) applied a similar algorithm to interpolation/extrapolation of missing traces. Their scheme uses the parabolic Radon transform as a band limiting device which is iteratively applied until the missing traces are restored.

## 4.2 Why sparse models?

In chapters 2 and 3, I showed that one of the principal shortcomings of using the conventional definition of the Radon operator is that amplitude smearing may mask events with similar intercept time and moveout. The inversion of the Radon operator by means

of the concept of minimum norm may help to diminish the smearing but, in situations where a highly resolved Radon panel is desired, the smoothing introduced by the minimum norm constraint does not guarantee enough resolution or focusing power. Assuming that the data are composed of a finite superposition of linear events or hyperbolic events, the Radon operator (for an ideal delta-like wavelet) would map each seismic event into a point. It is evident that a model with smearing is less sparse than a model without smearing and thus, minimizing a sparseness criterion, subject to data constraints, may help to reduce aperture artifacts.

Sparse models were widely used in another geophysical scenario. The work of Wiggins (1978) on minimum entropy deconvolution (MED) shows how a sparseness criterion may be used to simultaneously retrieve the reflectivity and the source wavelet from the seismic trace. The deconvolution problem, when the wavelet is unknown, is a severely ill-posed problem. This problem may be tackled by assuming that the reflectivity is a sparse time series. Wiggins' technique maximizes the varimax norm, but similar results may be obtained by maximizing other norms (Ooe and Ulrych, 1979; Poetic et. al, 1980). These norms measure departure from the Gaussian assumption. In other words, the MED looks for the least Gaussian reflectivity that fits the data. Projection Pursuit (PP) provides an elegant framework to understand MED (Friedman and Tukey, 1974; Huber, 1985; Jones and Sibson, 1986). PP is a statistical technique used to reduce the dimension of multivariate data. The technique operates as follows. First, a measure that defines some interesting configuration of the data is proposed <sup>1</sup>. Then, the data are projected into a 1-D or 2-D space where the criterion is maximized. This procedure enables us to project the data into a new system of coordinates where some interesting configuration is manifested. In MED deconvolution the data matrix (composed of the original time

---

<sup>1</sup>As pointed out by Huber (1985) "We cannot expect universal agreement on what constitutes an "interesting" projection."

series properly padded with zero to represent discrete convolution) is projected into a space where the sparseness is maximized. The projection operator is the MED filter and the projection is an estimate of the reflectivity.

It is believed that MED tends to retrieve a reflectivity that does not represent a real sequence of primary reflections. Some authors proposed more elaborated models to describe the reflectivity and use MED-style algorithms to compute estimates of the reflectivity sequence (Walden, 1985). I do not think that this approach provides an important advance with respect to the original MED idea. I suspect that the MED failures are not only due to the statistical assumptions. The degrees of freedom of the problem are also a crucial aspect to be taken into account. In other words, since the source wavelet is unknown, the link between the reflectivity and the trace is very weak and, thus the MED norm enters into the problem as a strong constraint. The norms that I have adopted to retrieve sparse solutions may be related to minimum entropy norms. However, there is an important distinction with respect to MED techniques. The model and the data are linked by the Radon operator. The data are the primary piece of information, and consequently the sparseness constraints enter into the problem as a secondary effect. In other words, the model is well anchored to the data and, therefore, the regularization serves to improve the model by reducing smearing. In MED, the reflectivity and the data are related by another unknown: the source wavelet or the inverse source wavelet.

I hope that this small digression on sparse models and MED illuminates the differences between MED-style techniques and the sparse regularization scheme proposed in this thesis.

### 4.3 Inverse problems and Bayes' rule

#### 4.3.1 Bayes' Rule

To clarify concepts, I will start with Bayes's rule. In what follows  $p(A|I)$  denotes the probability of proposition  $A$  being true conditional on  $I$  being true and,  $\bar{A}$  means that  $A$  is false. The rules for Bayesian parameter estimation are simply the product and sum rules of probability theory which are

$$p(AB|I) = p(A|I)p(B|AI) = p(B|I)p(A|BI) \quad (4.1)$$

and

$$p(A|B) + p(\bar{A}|B) = 1. \quad (4.2)$$

From (4.1), assuming that  $p(B|I) \neq 0$ , we derive Bayes' rule

$$p(A|BI) = p(A|I) \frac{p(B|AI)}{p(B|I)}. \quad (4.3)$$

Bayes' rule establishes the way of updating  $p(A|I)$  when additional information,  $B$ , is incorporated. In equation (4.3) the proposition  $I$  is the prior information and  $p(A|I)$  is the prior probability of  $A$  conditional only on the information given by  $I$  (Loredo, 1990). The term on the left-hand-side of equation (4.3) is called the posteriori probability. For Bayesians, probability is a measure of the degree of plausibility of a proposition. Basically, Bayes' rule serves to update the plausibility of a proposition when our state of

information changes because new data are acquired. The last assessment establishes the radical difference with respect to the “frequentist” view of probability. For frequentists, probabilities arise from the relative frequencies of outcomes of an infinite number of identical repetitions of an experiment. This definition restricts  $A$  to be a proposition about a random variable. The Bayesian viewpoint, on the other hand, expands the domain of probability theory to include any type of proposition.

### 4.3.2 Bayesian approach to inverse problems

The relationship between the Radon space ( $\omega - p$  or  $\omega - q$ ) and the data space when noise is considered is given by

$$\mathbf{L}^* \mathbf{v} + \mathbf{n} = \mathbf{u}, \quad (4.4)$$

where  $\mathbf{n}$  stands for the noise term and  $\mathbf{L}$  is the slant or parabolic stack operator depending on the problem at hand. Letting  $\mathbf{v} = A$  be the proposition we want to assess and  $\mathbf{u} = B$  represent our data, Bayes’ rule can be written as

$$p(\mathbf{v}|\mathbf{u}) = \frac{p(\mathbf{u}|\mathbf{v})p(\mathbf{v})}{p(\mathbf{u})} \quad (4.5)$$

where for simplicity we have omitted  $I$  in the notation. To clarify notation,

- $p(\mathbf{u}|\mathbf{v})$  is the probability or likelihood of obtaining the data  $\mathbf{u}$  assuming that the model  $\mathbf{v}$  is true.
- $p(\mathbf{v})$  is the prior probability of the model.
- $p(\mathbf{u})$  is the data likelihood and enters into the problem as a normalization factor.

- $p(\mathbf{v}|\mathbf{u})$  is the posterior probability of the model.

To utilize the Bayesian formalism we must consider two remaining problems. First, given  $p(\mathbf{v}|\mathbf{u})$ , how to determine  $\mathbf{v}$ , and second, how to determine  $p(\mathbf{v})$ . The first problem can be solved by an appropriate choice of a decision rule. For example, one could use the maximum a posteriori (MAP) solution,  $\mathbf{v}_{MAP}$ , which maximizes  $p(\mathbf{v}|\mathbf{u})$ . This solution has maximum posteriori probability, which is the same (up to a normalization factor) as seeking the maximum joint probability. The second problem, which is probably one of the most cumbersome problems in inverse theory, is how to translate our prior appraisal about the model into a probability density function.

If the prior probability and the likelihood functions have the forms

$$p(\mathbf{v}) \propto e^{-J_v}, \quad (4.6)$$

$$p(\mathbf{u}|\mathbf{v}) \propto e^{-J_u}. \quad (4.7)$$

Then, the posteriori distribution is given by,

$$\begin{aligned} p(\mathbf{v}|\mathbf{u}) &\propto p(\mathbf{v}) \times p(\mathbf{u}|\mathbf{v}) \\ &\propto e^{-(J_v+J_u)}. \end{aligned} \quad (4.8)$$

The MAP solution will coincide with the minimum of  $J = -\ln[p(\mathbf{v}|\mathbf{u})]$ , where  $J$  is usually called the objective or cost function.

$$J = J_v + J_u. \quad (4.9)$$

If the posterior distribution is multi-modal the objective function is not a strictly convex function, therefore a global optimization problem is required to find the MAP solution. If the prior and the likelihood are modelled with normal distributions, the objective function becomes the classical quadratic form widely used in inverse theory (Tarantola, 1987).

It is important to point out that other estimators of  $\mathbf{v}$  may be used. For instance, the mean  $\bar{\mathbf{v}}$  of the posteriori pdf

$$\bar{\mathbf{v}} = \int p(\mathbf{v}|\mathbf{u})d\mathbf{v}. \quad (4.10)$$

or the median,  $\mathbf{v}_{\frac{1}{2}}$ ,

$$P(\mathbf{v} \leq \mathbf{v}_{\frac{1}{2}}) = \frac{1}{2}. \quad (4.11)$$

From a computational point of view the MAP is usually preferred, since the computation of the mean and/or the median require the integration of the whole model space. For a symmetrical unimodal posterior pdf, the mean, the median and the MAP estimator coincide. The normalization constants of the pdf's are not required to minimize the objective function, and therefore they can be omitted in the analysis.

Bayesian inference considers  $J_v$  as additional information which may help to determine the most likely set of parameters. From a non-Bayesian perspective (rather, a classical inverse theory approach) this term serves to stabilize the inversion (Tikhonov and Goncharsky, 1987). The classical school uses the concept of regularization to obtain a stable inverse of an ill-posed problem. Typically, probabilities are not mentioned and attention is focused on formulating an objective function. In the Bayesian approach the

aim is to incorporate probabilistic information into the formulation of the inverse problem. Moreover, in the Bayesian approach the objective function is an intermediate step which is required to construct a solution.

There are two approaches to derive a cost function. The Bayesian approach, which is what was just explained, entails the construction of the posteriori probability and then the objective function. In inverse theory a more direct approach is usually adopted. In the latter the objective function is designed by combining a misfit criterion with a regularization function that is derived from a smoothing criterion. I feel that in many situations the Bayesian approach may highlight some formal aspects of the inversion. However, when an objective function is derived by means of Bayes' rule, a common question arises: why that pdf and not another?. On the other hand, in classic inverse theory the objective function is accepted without any problem, i.e., smoothing with derivatives (Constable et. al, 1985). Yet we have to realize, that both approaches may lead to similar or even to the same objective function. In fact, smoothing strategies based on derivatives may be related to Gaussian models with a particular covariance structure (Wang and Braile, 1994).

### 4.3.3 Assigning a prior to the model

Bayesian inference has been historically divided by the degree of objectivity in the selection of priors. Various attempts have been made to find prior probabilities which represent a state of total ignorance about the model (Jeffreys, 1961; Jaynes, 1968). These priors are derived from mathematical arguments of symmetry and invariance. An objective prior may be also derived using the principle of maximum entropy (Jaynes, 1968). Consider a random variable,  $x$ , with pdf  $f(x)$ . The entropy  $h$  given by

$$h = - \int_{-\infty}^{\infty} f(x) \ln[f(x)] dx \quad (4.12)$$

expresses the uncertainty associated with the distribution  $f(x)$ . Therefore, the distribution that most honestly describes the data, given only what it is known without assuming anything else, is the one with maximum entropy (Jaynes, 1968). It is natural, therefore, that when we attempt to make inferences based on incomplete information, that we draw them from that probability distribution that has the maximum entropy allowed by the available information.

The entropy is maximized subject to the following constraints

$$\int_a^b f(x) dx = 1 \quad (4.13)$$

and

$$\int_a^b f(x) g_k(x) dx = a_k, \quad k = 1, 2, \dots, m, \quad (4.14)$$

where the first constraint is a normalization constraint, the second is a set of moments that are assumed to be known. Using the Euler-Lagrange equation of the calculus of variations the maximum entropy pdf is given by

$$f(x) = e^{[-\lambda_0 - \lambda_1 g_1(x) - \lambda_2 g_2(x) - \lambda_m g_m(x)]}. \quad (4.15)$$

The Lagrange parameters  $\lambda_0, \lambda_1, \dots, \lambda_m$  are estimated by replacing  $f(x)$  into the constraints (Kapur and Kesavan, 1992) and solving a nonlinear problem.

Many well known pdf's may be obtained using the principle of maximum entropy. An example is the generalized Gaussian distribution. In this case, the prescribed constraints are equation (4.13) and the dispersion measure associated with the  $l_p$  norm,

$$(\sigma_p)^p = \int_{-\infty}^{\infty} |x|^p f(x) dx. \quad (4.16)$$

The pdf that maximizes the entropy is

$$f(x, \sigma_p, p) = \frac{p^{1-1/p}}{2\sigma_p \Gamma(p^{-1})} e^{-\frac{|x|^p}{\sigma_p^p}}. \quad (4.17)$$

The generalized Gaussian distribution with shape parameter  $p$  and scale parameter  $\sigma_p$  describes a family of distributions. In particular when  $p = 2$ , equation (4.17) corresponds to the normal or Gaussian distribution, when  $p = 1$  to the Laplace double exponential, and when  $p \rightarrow \infty$  to a uniform distribution. The metric induced by the generalized Gaussian distribution is the  $l_p$  norm. Laplace's double exponential is a pdf that has longer tails than the normal distribution. This property is exploited in robust statistic to damp the influence of outliers (Huber, 1981).

Another typical long-tailed probability density is the Cauchy distribution,

$$f(x, \sigma) = \frac{1}{\pi \sigma_c (1 + \frac{x^2}{\sigma^2})}. \quad (4.18)$$

This distribution provides an attractive prior to model a sparse distribution of parameters. In addition, it leads to a regularization function similar to Burg's entropy (Burg, 1975). Equation (4.18) can be derived using the principle of maximum Entropy when  $E[\ln(1 + x^2/\sigma_c^2)]$  is prescribed (Kapur and Kesavan, 1992) ( $E$  denotes expectation). In that case,  $E[\ln(1 + x^2/\sigma_c^2)]$ , plays a role similar to a dispersion measure.

The inverse problems studied in this thesis are posed in the frequency-offset space. Hence, the priors must be modified to cope with complex variables. Neglecting the normalizing term (not needed to compute the MAP estimator), the priors may be formally written down as

$$f(V, \sigma_p, p) \propto e^{\frac{-1}{p} \frac{|V|^p}{\sigma_p^p}}, \quad (4.19)$$

$$f(V, \sigma_c) \propto \frac{1}{1 + \frac{VV^*}{\sigma_c^2}}. \quad (4.20)$$

It is evident that we can identify the pdf of the bivariate variable  $\text{Re}(V) + i\text{Im}(V)$  with the univariate pdf of the complex variable  $V$ .

#### 4.3.4 Assigning a probability to the noise

To start with, I will assume that the second moment of the noise is known. Then, the principle of maximum entropy gives rise to the Gaussian pdf as the least informative probability distribution. Throughout this analysis I will consider that the seismic noise is Gaussian with zero mean and variance  $\sigma_n^2$ .

The real and imaginary parts of the discrete Fourier transform are defined by

$$\begin{aligned} N_r(\omega_k) &= \frac{1}{\sqrt{N}} \sum_{t=1}^N n_t \cos(\omega_k t) \\ N_i(\omega_k) &= \frac{1}{\sqrt{N}} \sum_{t=1}^N n_t \sin(\omega_k t) \end{aligned} \quad (4.21)$$

where  $\omega_k = 2\pi k/N$ ,  $k = 0, 1, 2, \dots, N-1$ . It follows that since the real and imaginary parts are linear combinations of a normally distributed variable  $n_t$ , they are also normally distributed with zero mean and variance given by (Priestly, 1981)

$$\text{Var}(N_r(\omega_k)) = \begin{cases} \sigma_n^2 & , k \neq 0, N/2 \\ 2\sigma_n^2 & , k = 0, N/2 \end{cases} \quad (4.22)$$

$$\text{Var}(N_i(\omega_k)) = \begin{cases} \sigma_n^2 & , k \neq 0, N/2 \\ 0 & , k = 0, N/2 \end{cases} \quad (4.23)$$

In addition, the covariance is given by

$$\text{Cov}(N_r(\omega_k), N_i(\omega_l)) = 0, \forall k, l \quad (4.24)$$

similarly,

$$\text{Cov}(N_r(\omega_k), N_r(\omega_l)) = \text{Cov}(N_i(\omega_k), N_i(\omega_l)) = 0, \forall k, l \quad (4.25)$$

From the above discussion it is clear that the real and imaginary parts have a bivariate distribution

$$p(N_r(\omega_k), N_i(\omega_k)) = \frac{1}{2\pi\sigma_n^2} e^{-\frac{1}{2} \frac{(N_r(\omega_k))^2 + (N_i(\omega_k))^2}{\sigma_n^2}} \quad k \neq 0, N/2 \quad (4.26)$$

Now, it is possible to regard the joint distribution of the real and imaginary part as the distribution of a complex variable  $N_r + iN_i$ ,

$$p(N(\omega_k)) = \frac{1}{2\pi\sigma_n^2} e^{-\frac{1}{2} \frac{N(\omega_k)N^*(\omega_k)}{\sigma_n^2}} \quad k \neq 0, N/2 \quad (4.27)$$

#### 4.3.5 Maximum a Posteriori Solution (MAP)

When we combine the generalized Gaussian prior and the Gaussian likelihood, the following objective function is obtained

$$J_p = \mathbf{v}^H \mathbf{Q}_p \mathbf{v} + (\mathbf{L}^* \mathbf{v} - \mathbf{u})^H \mathbf{C}_n^{-1} (\mathbf{L}^* \mathbf{v} - \mathbf{u}) \quad (4.28)$$

where the diagonal matrix  $\mathbf{Q}_p$  has the following elements

$$Q_{p;i} = \frac{p}{2\sigma_p^2} \left| \frac{v_i}{\sigma_p} \right|^{p-2} \quad (4.29)$$

When  $p = 2$ , the elements of the diagonal matrix  $\mathbf{Q}_c$  are equal. The cost function  $J_p$  corresponds to a mixed  $l_p$  and  $l_2$  norm problem (Alliney and Ruzinsky, 1994). Note, however, that this is a special mixed norm problem since the argument of the norm is the amplitude of a complex variable,  $|V_j|$ , and not the complex variable itself. The matrix  $\mathbf{C}_n$  stands for the noise covariance matrix which is assumed to be known. The minimization of  $\Phi_p$  gives rise to the following solution,

$$\mathbf{v} = (\mathbf{Q}_p + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)^{-1}\mathbf{L}\mathbf{C}_n^{-1}\mathbf{u}. \quad (4.30)$$

The diagonal matrix  $\mathbf{Q}_p$  in equation (4.30) is non-linearly related to the model. Equation (4.30) can be rewritten using the following manipulations

$$\begin{aligned} \mathbf{L} + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*\mathbf{Q}_p^{-1}\mathbf{L} &= \mathbf{L}\mathbf{C}_n^{-1}(\mathbf{C}_n + \mathbf{L}^*\mathbf{Q}_p^{-1}\mathbf{L}) \\ &= (\mathbf{Q}_p + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)\mathbf{Q}_p^{-1}\mathbf{L}. \end{aligned} \quad (4.31)$$

Since  $(\mathbf{C}_n + \mathbf{L}^*\mathbf{Q}_p^{-1}\mathbf{L})$  and  $(\mathbf{Q}_p + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)$  are positive definite and thus, invertible, the following identity holds

$$(\mathbf{Q}_p + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)^{-1}\mathbf{L}\mathbf{C}_n^{-1} = \mathbf{Q}_p^{-1}\mathbf{L}(\mathbf{C}_n + \mathbf{L}^*\mathbf{Q}_p^{-1}\mathbf{L})^{-1} \quad (4.32)$$

and expression (4.30) is equivalent to

$$\mathbf{v} = \mathbf{Q}_p^{-1}\mathbf{L}(\mathbf{C}_n + \mathbf{L}^*\mathbf{Q}_p^{-1}\mathbf{L})^{-1}\mathbf{u}. \quad (4.33)$$

It must be stressed that for theoretical assessment regarding uniqueness and convergence, the operators (4.30) and (4.33) are equivalent.

The  $l_p$  regularization may lead to division by zero when  $v_i \rightarrow 0$ . To avoid this inconsistency the following modification is included

$$Q_{p;i} = \frac{p}{2\sigma_p^p} \begin{cases} |V_i|^{p-2} & \text{if } |V_i| > \epsilon, \\ \epsilon^{p-2} & \text{if } |V_i| \leq \epsilon. \end{cases} \quad (4.34)$$

With this modification, all samples smaller than  $\epsilon$  are weighted in the same manner. This technique was also used by Huber (1981) in robust statistical procedures to assign a constant weight to errors below a given threshold.

If the Cauchy distribution is adopted, the objective function derived from Bayes theorem is given by

$$J_c = \sum_i \ln\left(1 + \frac{|V_i|^2}{\sigma^2}\right) + (\mathbf{L}^*\mathbf{v} - \mathbf{u})^H \mathbf{C}_n^{-1} (\mathbf{L}^*\mathbf{v} - \mathbf{u}). \quad (4.35)$$

It is straightforward to see that in equations (4.30) and (4.33) the matrix  $\mathbf{Q}_p$  is replaced by  $\mathbf{Q}_c$ ,

$$Q_{c:ii} = \frac{1}{\sigma_c^2 \left(1 + \frac{|V_i|^2}{\sigma_c^2}\right)}. \quad (4.36)$$

Crase et. al (1990) and Amundsen (1991) used this criterion to improved the inversion of seismic data in the presence of outliers. In a time series analysis scenario, the Cauchy regularization has been applied to retrieve the spectral signature of seismic wavefields (Sacchi and Ulrych, 1995b).

It is evident that some iterative strategy is required to solve equation (4.30) or (4.33). Letting  $\mathbf{Q}$  denote  $\mathbf{Q}_p$  or  $\mathbf{Q}_c$ , the algorithm follows

$$\mathbf{v}_{k+1} = (\mathbf{Q}_k + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)^{-1}\mathbf{L}\mathbf{C}_n^{-1}\mathbf{u}, \quad (4.37)$$

similarly we can operate with (4.33),

$$\mathbf{v}_{k+1} = \mathbf{Q}_k^{-1} \mathbf{L} (\mathbf{C}_n + \mathbf{L}^* \mathbf{Q}_k^{-1} \mathbf{L})^{-1} \mathbf{u}. \quad (4.38)$$

The algorithm is initiated with either a flat model or with the following approximation:  $\mathbf{v} = \mathbf{L}\mathbf{u}$ . The number of iterations necessary to reach the solution will depend on the adopted regularization. Usually 5 iterations yield a good approximation to the minimum of  $J_p$  ( $p \approx 1$ ) and/or  $J_c$ . One advantage of the  $l_p$  regularization, is that the degree of sparseness may be modified by selecting the shape parameter  $p$ . One inconvenience, however, is that trimming of small amplitudes is required.

The effect of the non-linear weights  $\mathbf{Q}_p$  or  $\mathbf{Q}_c$  can be summarized as follows. In each iteration, the non-linearity produces a model that, while consistent with the data, has the minimum amount of structure or maximum sparseness. In fact, the algorithm drives the solution to a state in which most samples are equal to zero with only few samples different from zero. The algorithm is referred to as the Iteratively Model Re-weighted Least Squares (IMRLS) algorithm. The re-weighting function may be also derived using other priors than those proposed in this chapter.

#### 4.3.6 Algorithm complexity

The operators (4.30) and (4.33) are equivalent, however, it is important to consider the computational advantages of one over the other. For underdetermined problems ( $M > N$ ), clearly, it is more convenient to solve equation (4.33) than equation (4.30).

The following cases arise depending on the discretization in model and data space:

- **Slant Stack operator:** If the ray parameter axis is evenly discretized, the matrix  $\mathbf{L}\mathbf{L}^*$  is Toeplitz Hermitian. If a zero order quadratic regularization is adopted (damped least squares), the elements of the diagonal matrix  $\mathbf{Q}$  are equal, consequently,

$\mathbf{Q} + \mathbf{L}\mathbf{L}^*$  is also Toeplitz. In this situation, fast solvers like Levinson recursion may be used (Kostov, 1990). When the weighting function is derived from the Cauchy prior or the generalized Gaussian ( $p \neq 2$ ), the elements of the diagonal matrix  $\mathbf{Q}$  are unequal. Hence, the Toeplitz structure is destroyed. In this case, the solution may be computed with any solver for Hermitian matrices. I have adopted a Cholesky decomposition (Press et. al, 1992) but other techniques may be utilized. Since only a few iterations are required to reach a feasible solution, the computational cost of Cholesky decomposition is reasonable.

- Parabolic stacks: The operator  $\mathbf{L}\mathbf{L}^*$  is Toeplitz Hermitian providing that the Radon parameter is evenly discretized. When zero order regularization is adopted (damped least squares) equation (4.30) may be solved using the Levinson recursion. Any other case is solved using Cholesky decomposition.

#### 4.3.7 Convergence

In this section, I show that the proposed algorithm minimizes the cost function  $J_p$  and/or  $J_c$ . The change of the objective function between two consecutive iterations may be approximated by a first order Taylor expansion

$$\begin{aligned} \Delta J_k &= J_{k+1} - J_k \\ &= \nabla J_k^H \Delta \mathbf{v}_k. \end{aligned} \tag{4.39}$$

The solution at iteration  $k + 1$  satisfies,

$$\begin{aligned} \mathbf{L}\mathbf{C}_n \mathbf{u} &= (\mathbf{Q}_k + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)\mathbf{v}_{k+1} \\ &= (\mathbf{Q}_k + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)(\mathbf{v}_k + \Delta \mathbf{v}_k) \end{aligned} \tag{4.40}$$

thus,

$$\nabla J_k = -(\mathbf{Q}_k + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)\Delta\mathbf{v}_k. \quad (4.41)$$

Finally, combining the above expressions

$$\Delta J_k = -\Delta\mathbf{v}_k^H(\mathbf{Q}_k + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)\Delta\mathbf{v}_k. \quad (4.42)$$

The objective function is decreased providing that  $(\mathbf{Q}_k + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)$  is positive definite. This is proved in the next section.

#### 4.3.8 Convexity of the objective function

If the objective function is strictly convex and possesses bounded derivatives, then the problem

$$J = \min \quad (4.43)$$

or

$$\nabla J = 0 \quad (4.44)$$

has a unique solution. The convexity assumption serves to guarantee equivalence between (4.43) and (4.44). If we are satisfied with local uniqueness, strict convexity can be

replaced by convexity (Rockafellar, 1970). The function  $J$  is convex provided that the Hessian matrix,  $\mathbf{H} = \nabla\nabla J$ , is positive definite

$$\mathbf{x}^H \nabla\nabla J \mathbf{x} > 0 \text{ for all } \mathbf{x} \text{ other than } \mathbf{x} = \mathbf{0}, \quad (4.45)$$

where the entries of the Hessian matrix are

$$\{\nabla\nabla J\}_{ij} = \frac{\partial^2 J}{\partial x_i \partial x_j^*}. \quad (4.46)$$

If the regularization is derived from the generalized Gaussian, the Hessian becomes

$$\nabla\nabla J_p = \frac{p}{2} \mathbf{Q}_p + \mathbf{L} \mathbf{C}_n^{-1} \mathbf{L}^*. \quad (4.47)$$

The quadratic form  $\mathbf{v}^H \mathbf{Q}_p \mathbf{v}$  is always positive definite, while  $\mathbf{v}^H \mathbf{L} \mathbf{C}_n^{-1} \mathbf{L}^* \mathbf{v}$  is symmetric and positive semi-definite (positive definite if  $\mathbf{L}^*$  has full rank), hence the sum is positive definite. It might happen that some elements of the matrix  $\mathbf{Q}_p$  become unbounded when  $V_j \rightarrow 0$ . The threshold operator in equation (4.34) serves to avoid this degeneracy.

If the Cauchy criterion is adopted as the prior of the problem, the Hessian of the objective function is given by

$$\nabla\nabla J_c = \mathbf{D} + \mathbf{L} \mathbf{C}_n^{-1} \mathbf{L}^*, \quad (4.48)$$

where  $\mathbf{D}$  is a diagonal matrix with the following elements

$$D_{ii} = \frac{1}{\sigma_c^2} \frac{1}{1 + \frac{|v_i|^2}{\sigma_c^2}} \left(1 - \frac{\frac{|v_i|^2}{\sigma_c^2}}{1 + \frac{|v_i|^2}{\sigma_c^2}}\right).$$

Since  $\sigma_c^2 > 0$ , the Hessian is a positive form. Thus, the convexity of  $J$  is guaranteed for both regularisers. The existence of the solution is also guaranteed in both cases. Finally, note that the equivalence between equations (4.30) and (4.33) is also satisfied.

### 4.3.9 Derivation of the damped least squares solution

What is the relationship between the well known damped least squares approach and the sparse regulariser? There is an easy way to answer this question when the regularization is induced by the generalized Gaussian. We simply set  $p = 2$  in (4.30) to end up with

$$\mathbf{v} = (\sigma_2^{-2}\mathbf{I} + \mathbf{L}\mathbf{C}_n^{-1}\mathbf{L}^*)^{-1}\mathbf{L}\mathbf{C}_n^{-1}\mathbf{u}. \quad (4.49)$$

If the noise covariance matrix is diagonal,  $\mathbf{C}_n = \sigma_n^2\mathbf{I}$ , then the damping factor reduces to the ratio of variances.

The situation is quite different when we use the Cauchy density to model the prior distribution of parameters. If the parameter  $\sigma_c$  in (4.20) is chosen according to  $\sigma_c \gg |V_i|, i = 1, \dots, M$  then,  $\mathbf{Q}_c \approx \sigma^{-2}\mathbf{I}$ . Thus, for large values of  $\sigma$  compared with the amplitude of the model, the Cauchy density also leads to the damped least squares solution (4.49). This should not be a surprise, since the Cauchy density represents a curve quite similar to the Gaussian, although with large tails.

## 4.4 Application to slant stacks

### 4.4.1 Slant stacking non-aliased data

The synthetic example presented in section 2.4 is used to test the performance of the sparse regularization scheme. First, I computed the slant stack using the high resolution slant stack operator derived with the generalized Gaussian prior with  $p = 1$ . Similarly, I also used the operator derived from the Cauchy prior. The results are shown in Figures 4.1c and 4.1d. In Figures 4.1a and 4.1d, I have also plotted the  $\tau - p$  space computed with the conventional slant stack operator and with the damped least squares solution. The sparse regularization leads to a well focused  $\tau - p$  panel. Notice, that the smearing in the first two panels has been significantly diminished.

### 4.4.2 Slant stacking beyond the alias

A synthetic data set was generated by superimposing five linear events with slowness and intercept times given in table (4.1).

<i>Event</i>	$\tau$ [s]	$p$ [s/m]	$v = p^{-1}$ [m/s]
2	0.50	$1.82 \times 10^{-4}$	5500.
3	0.60	$2.17 \times 10^{-4}$	4600.
3	0.80	$2.17 \times 10^{-4}$	4600.
4	0.82	$1.82 \times 10^{-4}$	5500.
5	0.90	$-1.80 \times 10^{-4}$	-5550.

Table 4.1: Parameters used to generate the synthetic data.

The composite seismogram is illustrated in figure 4.2. The receivers are uniformly distributed 100m apart. The spatial Nyquist frequency is  $5 \times 10^{-3} \text{m}^{-1}$ . The seismic signature is simulated with a Ricker wavelet of 35Hz. The problem is solved in the band

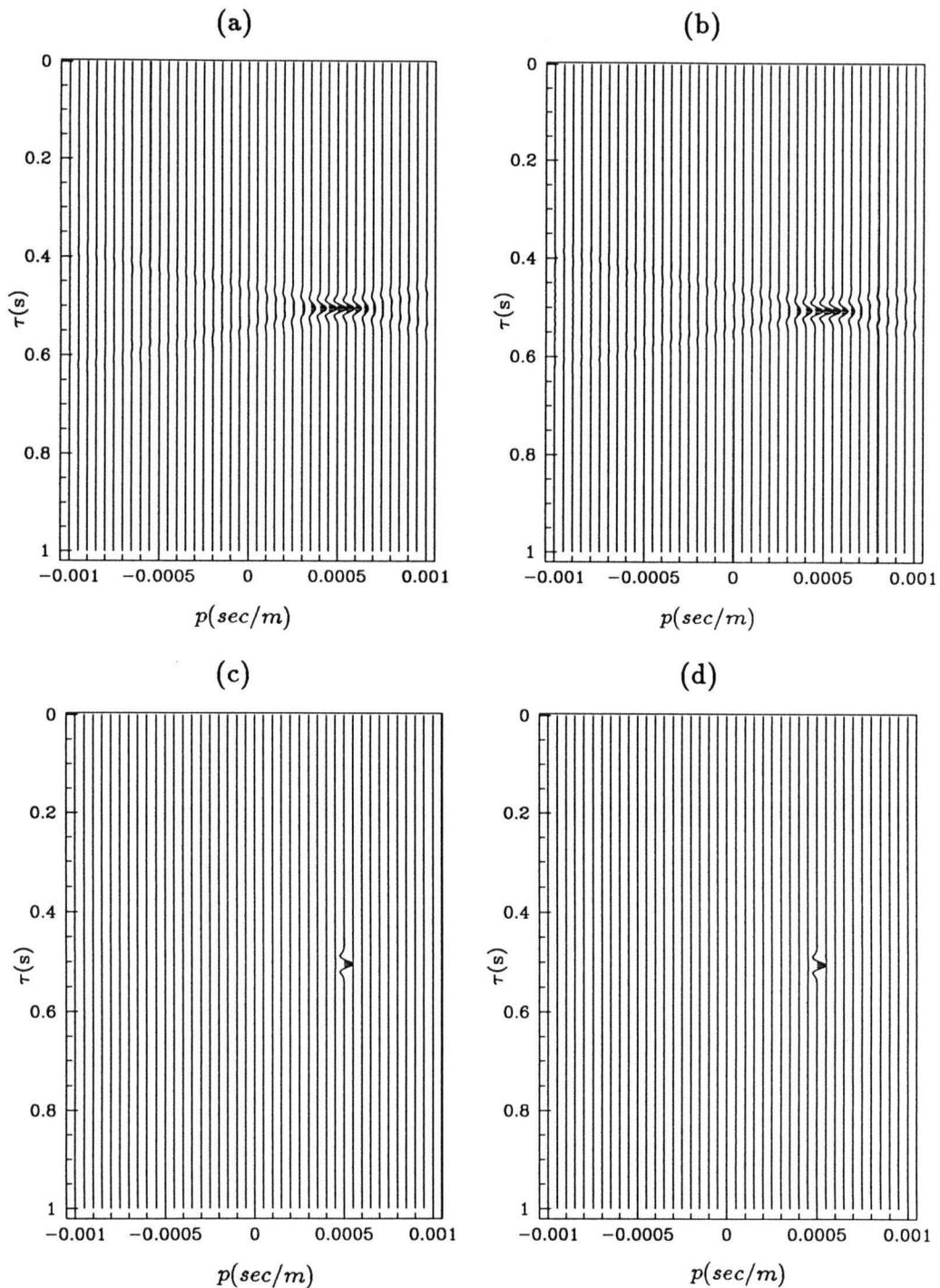


Figure 4.1: (a)  $\tau - p$  transform computed with the conventional Radon operator (linear moveout and stacking). Panel (b) shows the Radon transform computed with the damped least squares algorithm, (c) and (d) were computed with operators derived from the sparse regularization. The generalized Gaussian prior,  $p = 1$ , was used in (c) and the Cauchy prior (d).

1 – 125Hz. The maximum resulting wavenumber is  $k_{max} = f_{max} \times p_{max} = 0.022s^{-1}$ . The  $\tau - p$  panels computed with different operators are portrayed in Figures 4.3a, b, c and d. The first panel corresponds to the conventional slant stack operator, linear moveout and stacking. The remaining panels to the damped least squares solution  $p = 2$ , (figure 4.3b) and to the IMRLS solution with  $p = 1.1$  (figure 4.3c). In addition figure 4.3d shows the  $\tau - p$  domain computed using the regularization derived from Cauchy prior. The aliasing effect is clearly presented in the  $\tau - p$  panels portrayed in Figures 4.3a and b. The sparse regularization has produced not only very well focused  $\tau - p$  panels but also the aliasing effect has been significantly attenuated. This is clearly shown in figures 4.3c and d. For completeness, in Figures 4.4a, b, c and d a contour plot of the temporal envelope of the  $\tau - p$  panels is shown. The envelope was computed by means of a discrete Hilbert transform (Oppenheim and Schaffer, 1975). The maximum contour corresponds to 0db, while the minimum to -30db. An enlarged portion of these figures containing the fourth and fifth linear event is shown in Figures 4.5a, b, c and d.

Thus far I have described a noise-free problem. To continue with the analysis, the data were contaminated with Gaussian noise with standard deviation,  $\sigma_n = 0.15$ . The original wavefield, the noise and the composite seismogram are plotted in figure 4.6a, b and c respectively. First, the conventional slant stack operator is displayed in Figure 4.7a. Then, the adjoint operator is applied to the  $\tau - p$  space to map the data space (Figure 4.7b). It is clear that since the forward and the adjoint transformation do not constitute an inverse pair, the deconvolution of the operator  $\mathbf{G}$  is required to properly map the data (section 2.3.1). The data space after deconvolution is shown in Figure 4.7c. It is understood that the conventional operator can not separate signal from noise. This can be alleviated, however, using inverse operators. Figure 4.8a shows the  $\tau - p$  domain computed with the Cauchy regularization. The resolution of the operator is comparable with the resolution achieved in the noise free case (Figure 4.3d). The reconstructed data

space is portrayed in Figures 4.8b and the estimated noise space in Figure 4.8c.

There is no doubt that a critical aspect of the technique is the selection of the proper tradeoff to avoid over or under-fitting the data. The tradeoff between reliability of the estimate and resolution is governed by the parameter  $\sigma_c$ . A line search routine based on Brent's method (Press et al., 1992) is adopted to estimate the value of  $\sigma_c$  that provokes the misfit to lie in a previously defined interval. It is convenient to use a broad interval centered at the expected value of the  $\chi^2$  statistics,  $E(\chi^2) = 2N$ . This strategy significantly reduces the number of line searches. Figure 4.9 shows the average power spectrum of the data  $|\mathbf{u}/\sigma_n|^2$  and the misfit function at each frequency:

$$\chi(f)^2 = \left| \frac{\mathbf{L}^* \mathbf{v} - \mathbf{u}}{\sigma_n} \right|^2.$$

If  $\mathbf{v} = \mathbf{0}$  for all frequencies, the misfit function is proportional to the average power spectrum of the data. If we define power as

$$P(f) = \left| \frac{\mathbf{u}}{\sigma_n} \right|^2 = \frac{\mathbf{u}^H \mathbf{u}}{\sigma_n^2},$$

it is possible to compare the power spectrum of the data with the misfit function. In addition, notice that at high frequencies where  $\mathbf{v} = \mathbf{0}$  the misfit must coincide with the power spectrum. In other words at high frequencies there is nothing to invert, therefore the most reasonable vector of parameters is  $\mathbf{v} = \mathbf{0}$ . The same analysis is valid at low frequencies where the signal to noise ratio is usually very small. From this perspective the tradeoff parameter behaves as a noise rejection device.

Generally, line searching may be avoided. A trial tradeoff can be used to perform the inversion. If the  $\tau - p$  panel looks too noisy, the tradeoff may be increased to reject the

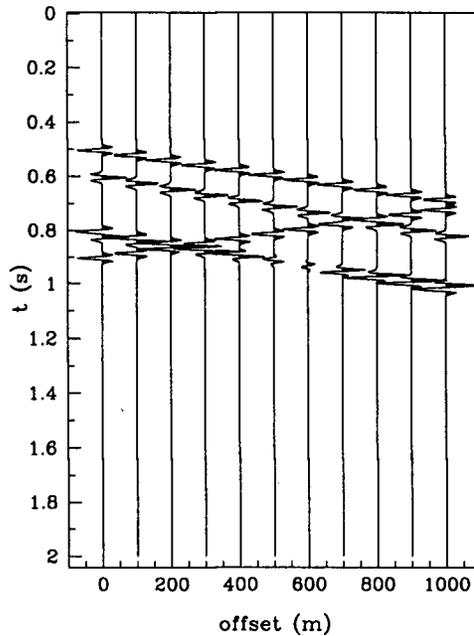


Figure 4.2: Synthetic profile with 5 linear events. The apparent velocities and intercept times are displayed in Table 4.1.

noise. Truncating the number of iterations is another strategy to avoid over-fitting the data. A useful minimum of the objective function can be reached in a few iterations and, in this case, the  $\chi^2$  test may be used as a stopping criterion.

#### 4.4.3 High resolution wavefield decomposition of VSP data

This section is devoted to the demonstration of the capability of the model re-weighted least squares algorithm to compute highly focused  $\tau - p$  panels in the context of VSP processing. Figure 4.10a shows a VSP window composed of 26 traces of 250 samples each. Figures 4.10b and 4.10c illustrate a main goal in VSP processing that of the correct separation of up-going and down-going waves. Figure 4.10b portrays the  $\tau - p$  space computed using a least squares solution with zero order regularization (damped least squares). We can distinguish two wavefields located at  $\approx \pm 1.6 \times 10^{-4}$  s/m. The  $\tau - p$  domain was

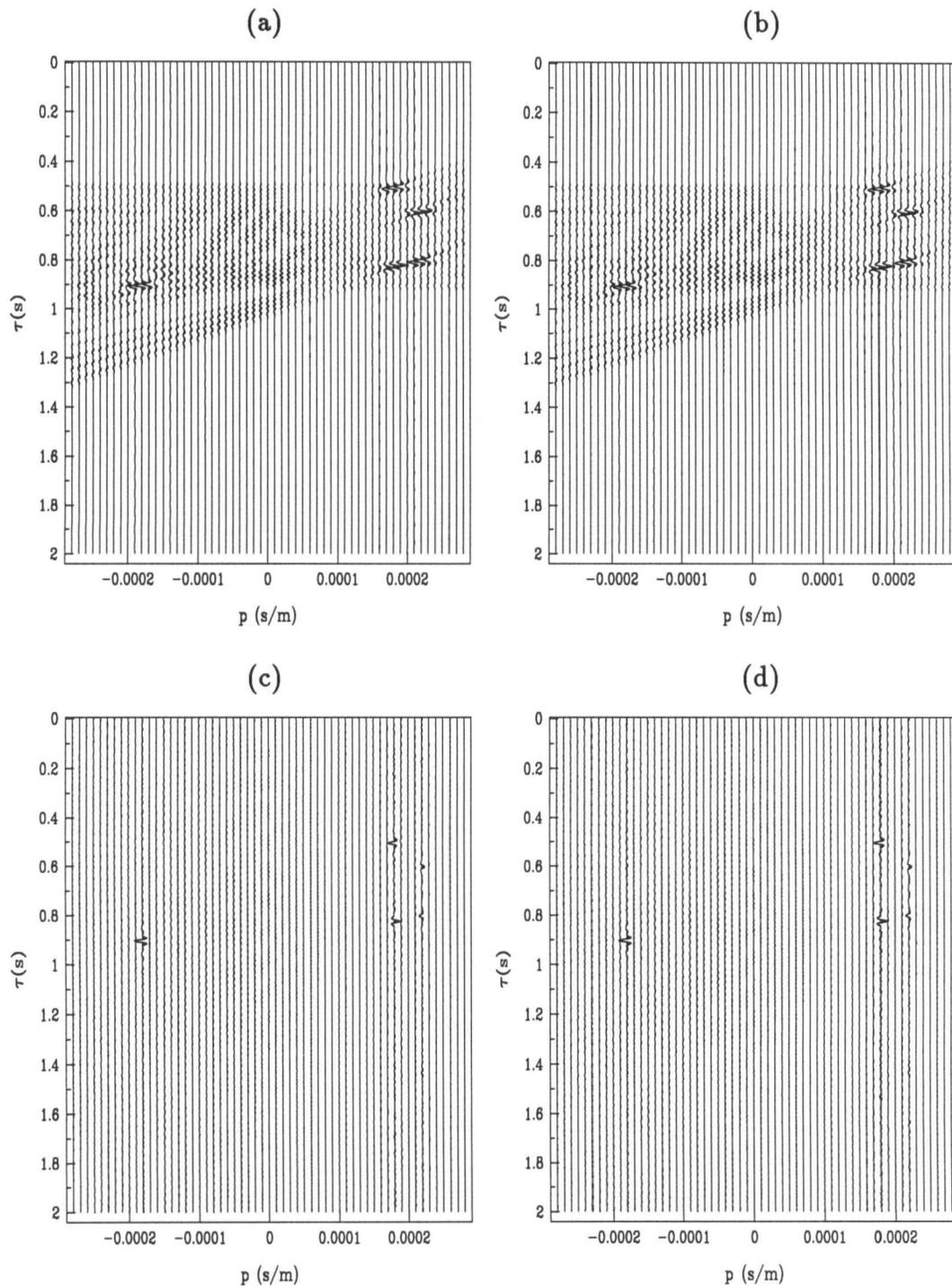


Figure 4.3: (a) Slant stack computed with the conventional operator. (b)-(c) Slant stacks computed with the inverse operator derived from the generalized Gaussian density with shape parameter  $p = 2$  and  $p = 1.1$  respectively.  $p = 2$  corresponds to the damped least squares solution. In panel (d) the slant stack was derived from the Cauchy regularization. Alias artifacts are diminished when the sparse regularization is applied (Panels (c) and (d)).

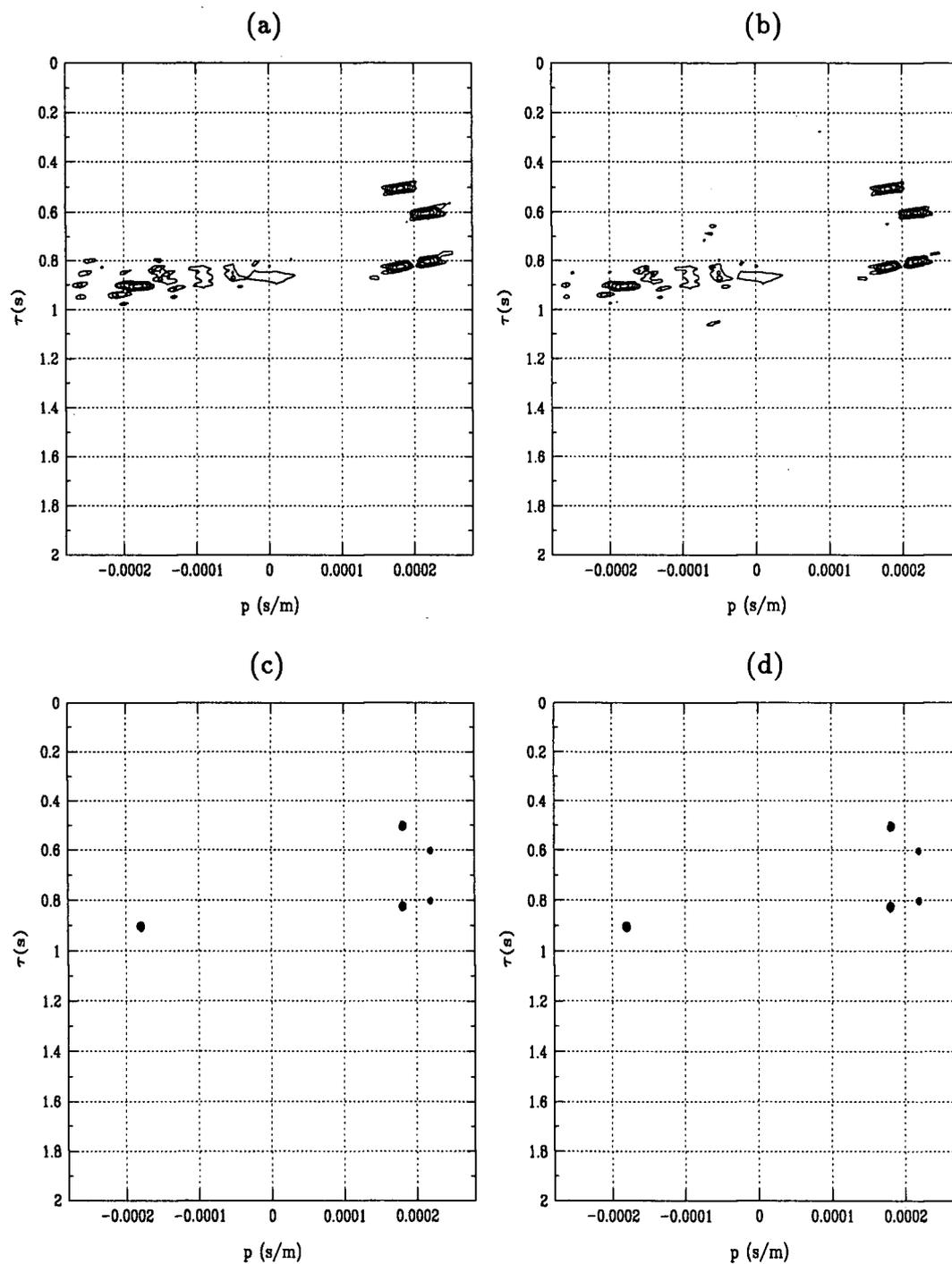


Figure 4.4: Contour plot of the normalized envelope of the panels showed in Figure 4.3.

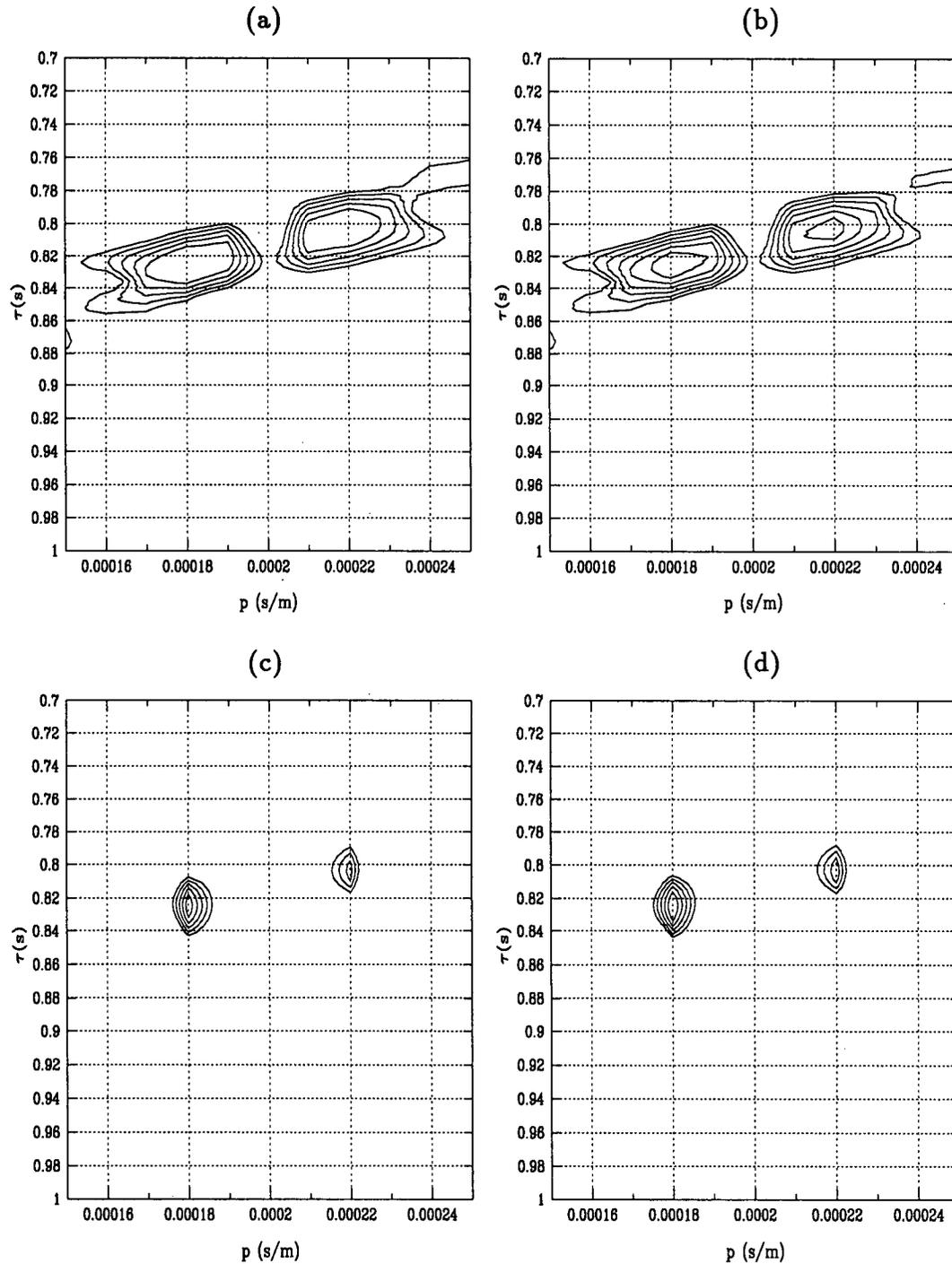


Figure 4.5: Blowup of the fourth and fifth events (Table 4.1) portrayed in figure 4.4. The events are perfectly resolved in (c) and (d). The contour lines corresponds to 0, -5, -10, -15, -20, -25 and -30db.

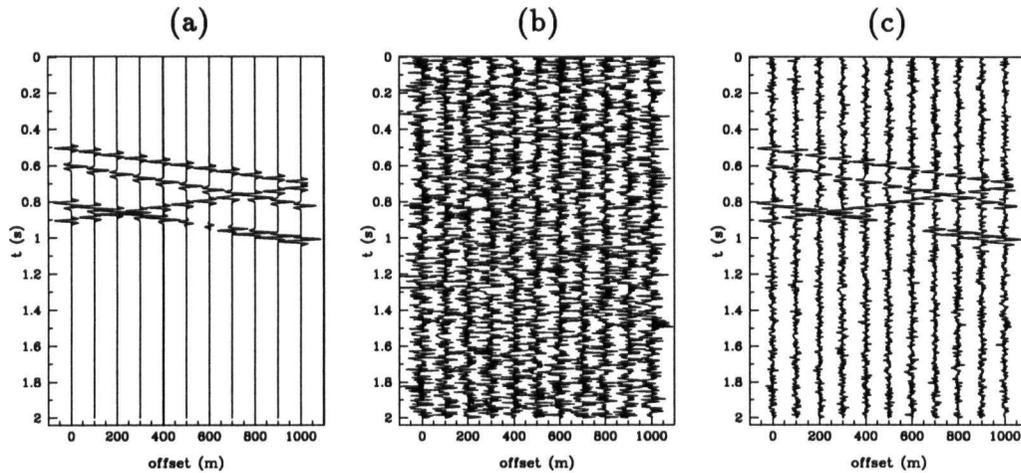


Figure 4.6: (a) The synthetic wavefield. (Figure 4.2). (b) Gaussian noise ( $\sigma_N = 0.15$ ). The noise has been normalized for plotting. (c) The composite seismogram derived from (a) and (b). The noise standard deviation ( $\sigma_n = 0.15$ ) represents 15% of the maximum amplitude encountered in the wavefield.

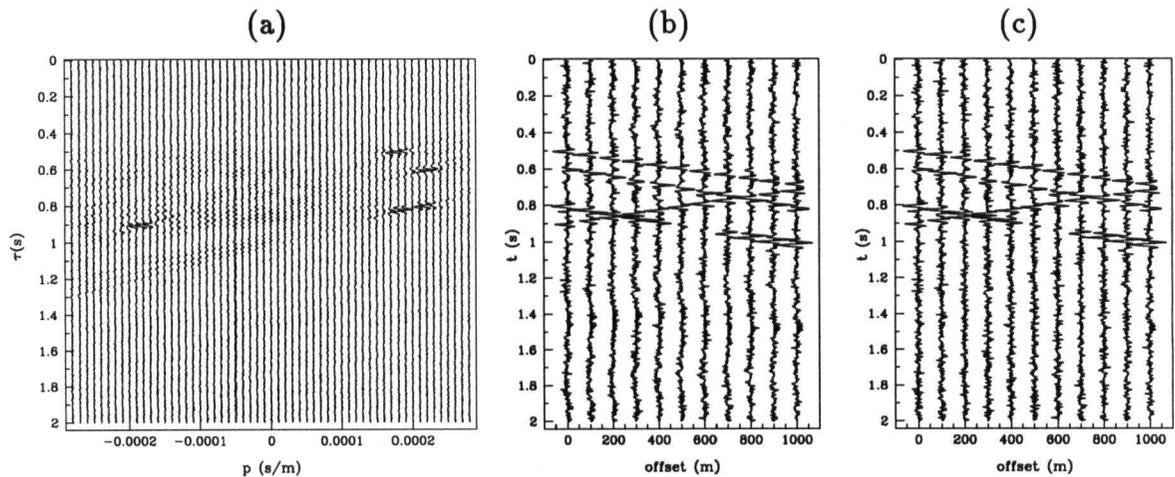


Figure 4.7: (a) The  $\tau - p$  space of the noisy synthetic wavefield displayed in Figure 4.6. (b) Data space estimate with the conventional slant stack. (c) Reconstruction of the data after the deconvolution of the operator  $G^{-1}$ . The noise is completely mapped back to the data space.

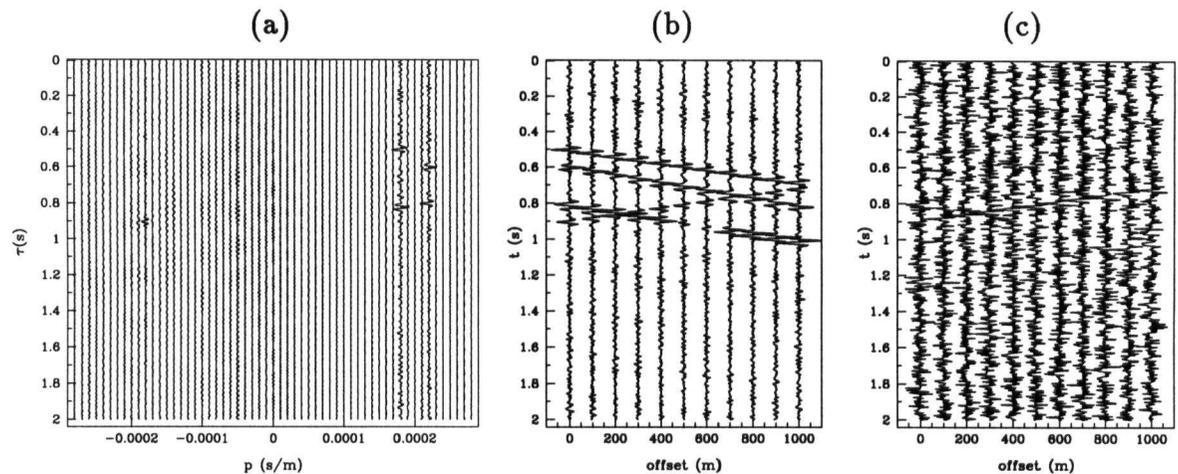


Figure 4.8: (a) The  $\tau - p$  space computed with the slant stack operator derived from the Cauchy regularization. (b) The reconstructed data space. (c) Estimate of the noise space. The noise is rejected by adjusting the tradeoff parameter at each frequency (Figure 4.9).

also computed using the  $l_p$  regularization with shape parameter  $p = 1$  (Figure 4.10c). A comparison of this figures leads to the following comments. The sparse regularization yields a panel where the finite aperture artifacts (smearing) are substantially diminished. Another striking feature is that the down-going wavefield almost collapses into a single trace, suggesting that the algorithm has managed to retrieve a resolution comparable with the one we might expect from a longer array. Filtering in  $\tau - p$  becomes an easy task since each wavefield can be clearly identified and extracted from the data.

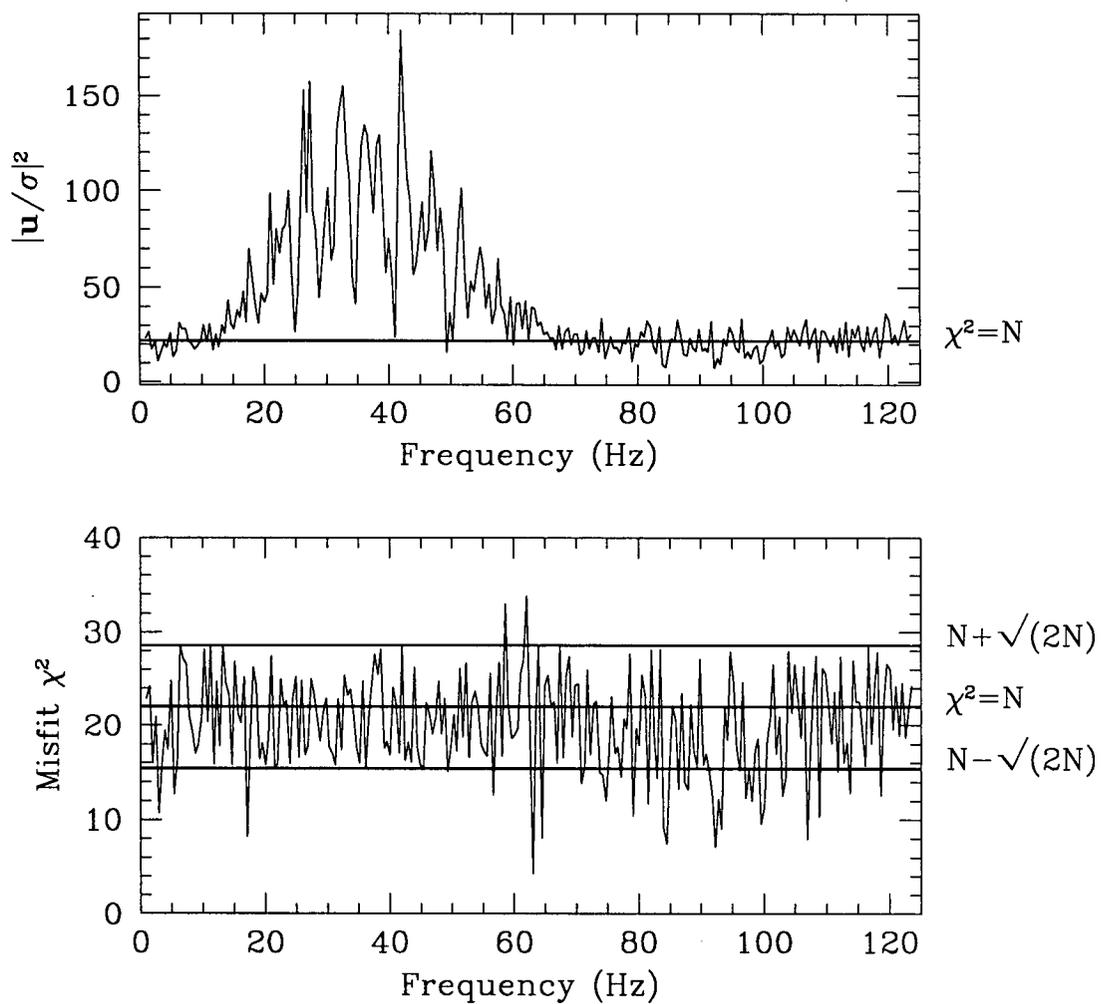


Figure 4.9: Misfit function versus frequency. The upper diagram show the power spectrum density of the data. The lower diagram shows the data misfit function,  $\chi^2$ , versus frequency. The expected value of the variable  $\chi^2$  together with its standard deviation are also plotted.

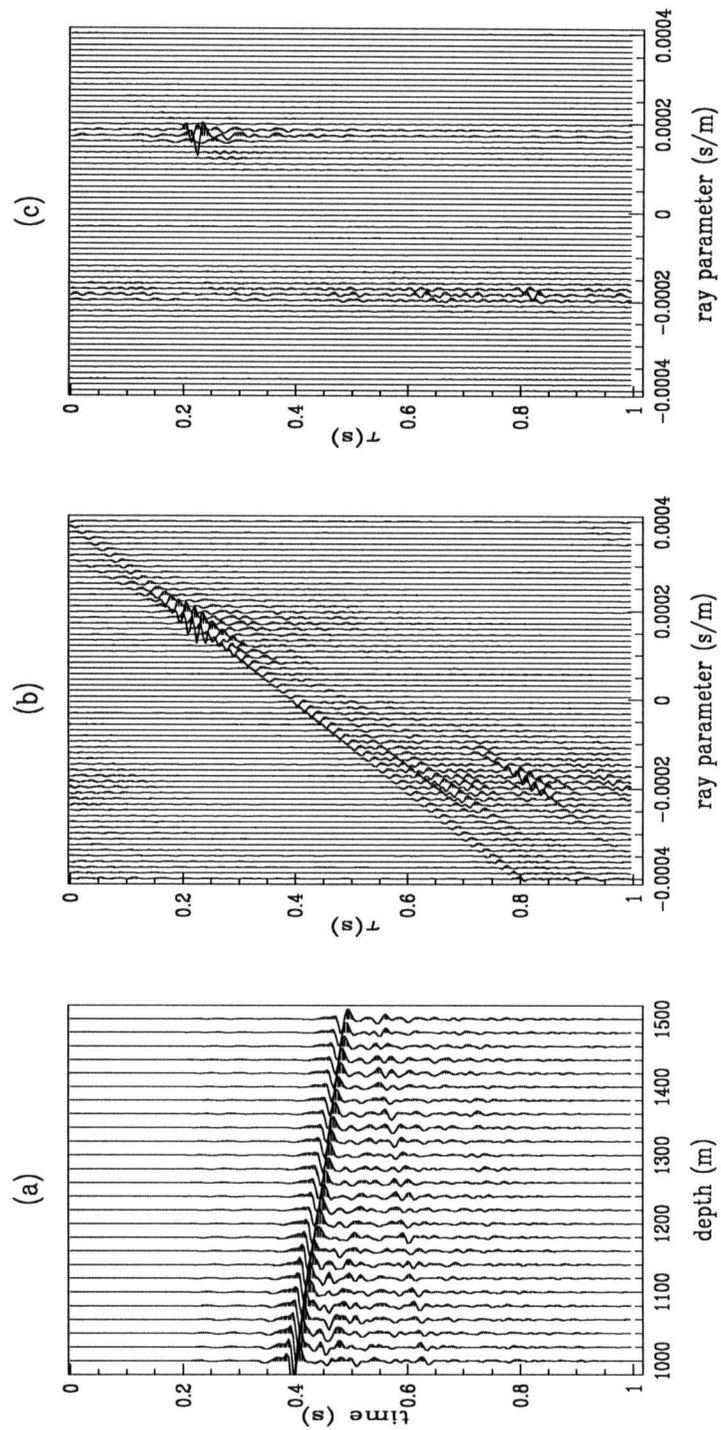


Figure 4.10: (a) VSP segment. (b)  $\tau - p$  panel computed with the damped least squares solution. (c) IMRLS solution using the  $l_p$  prior with shape parameter  $p = 1.1$ .

## 4.5 Application to velocity processing

In what follows, I devise a synthetic example to explore the focusing power of the sparse regularization procedure when applied to parabolic stacking. The synthetic example was already presented in Section 3.4. As in the previous example, a  $t^2$  transformation is used to validate the parabolic approximation. The data correspond to a single hyperbolic event of velocity  $v = 2500\text{m/s}$ . The square ray parameter of this event is  $q = 0.16\text{ms}^2/\text{m}^2$ . Figures 4.11a and 4.11b portray the velocity panel computed with the conventional parabolic stack operator and the damped least squares solution. In addition I used the Cauchy prior density and the generalized Gaussian ( $p = 1$ ) to induce sparse solutions. The results are displayed in Figures 4.11c and 4.11d. The maximum number of iterations was set to 5 in both cases.

### 4.5.1 Identification of multiples

Three primary reflections corresponding to velocities of  $3300\text{m/s}$  and with intercept times of  $0.4\text{s}$ ,  $0.8\text{s}$  and  $1.2\text{s}$  were used to simulate a CMP gather. This example intends to reproduce the synthetic example created by Yilmaz (1989) to test his algorithm. Superimposed on these events there is a primary reflection corresponding to a velocity of  $3000\text{m/s}$  with its resulting multiples. Near and far offset traces are located at  $0\text{m}$  and  $2500\text{m}$  respectively. The spatial sampling in the offset space is  $100\text{m}$ . The seismogram is displayed in Figure 4.12. Similarly to the previous examples, the  $\tau - q$  plane was first obtained with the conventional parabolic transform and with damped least squares. The results are portrayed in Figures 4.13a and 4.13b. Smearing introduces ambiguity into the discrimination of the reflections. This feature is particularly striking at  $\tau = 0.4, 0.8, 1.2\text{s}$  where primary and multiples interfere. When the data are processed with the IMRLS algorithm with weights derived from the Cauchy density and the Generalized Gaussian

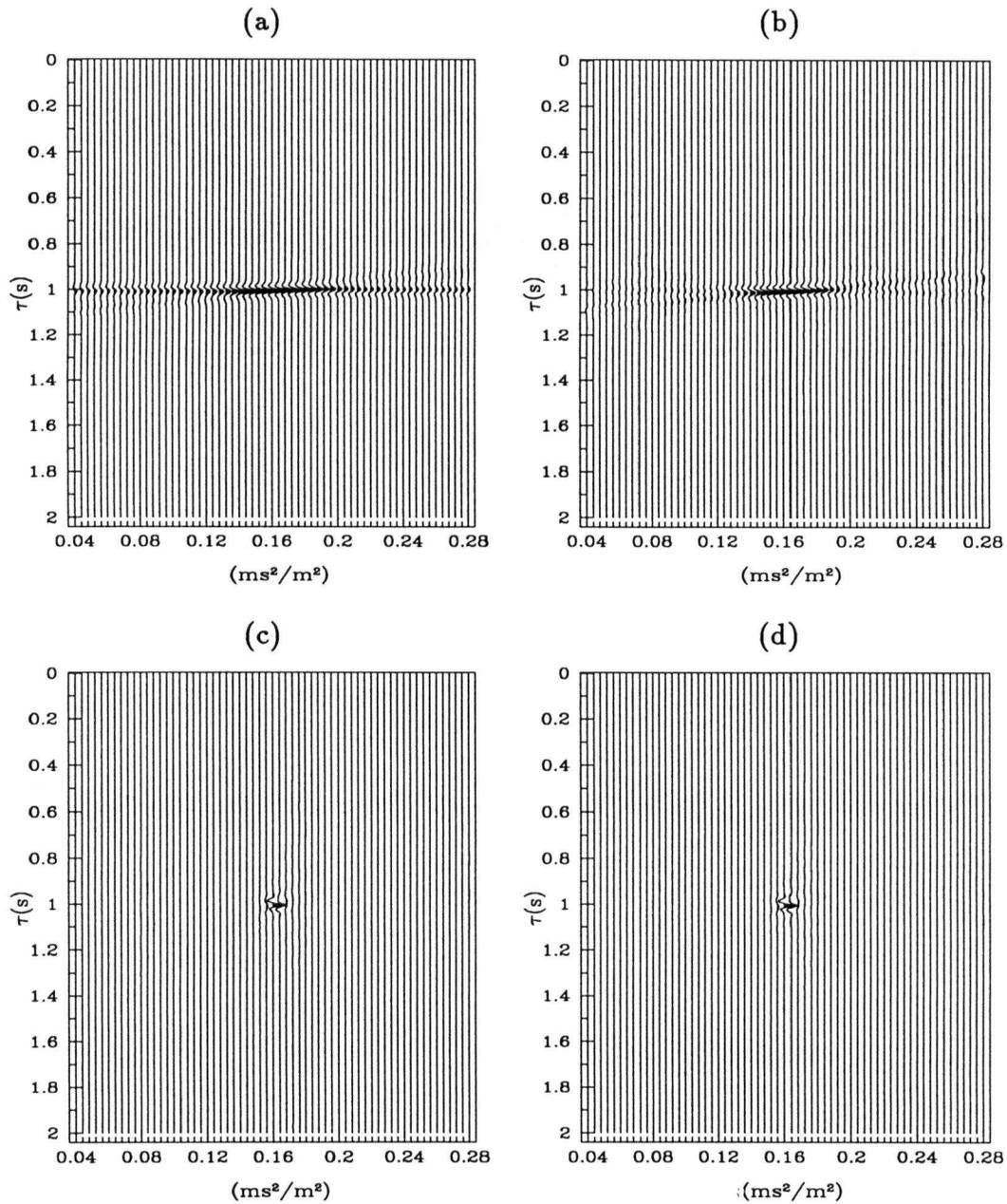


Figure 4.11: (a) Velocity processing with the conventional parabolic stacking operator. (b) shows the  $\tau - q$  panel computed with the damped least squares procedure, (c) and (d) were computed with the IMRLS algorithm with the generalized Gaussian prior ( $p = 1$ ) and the Cauchy prior, respectively.

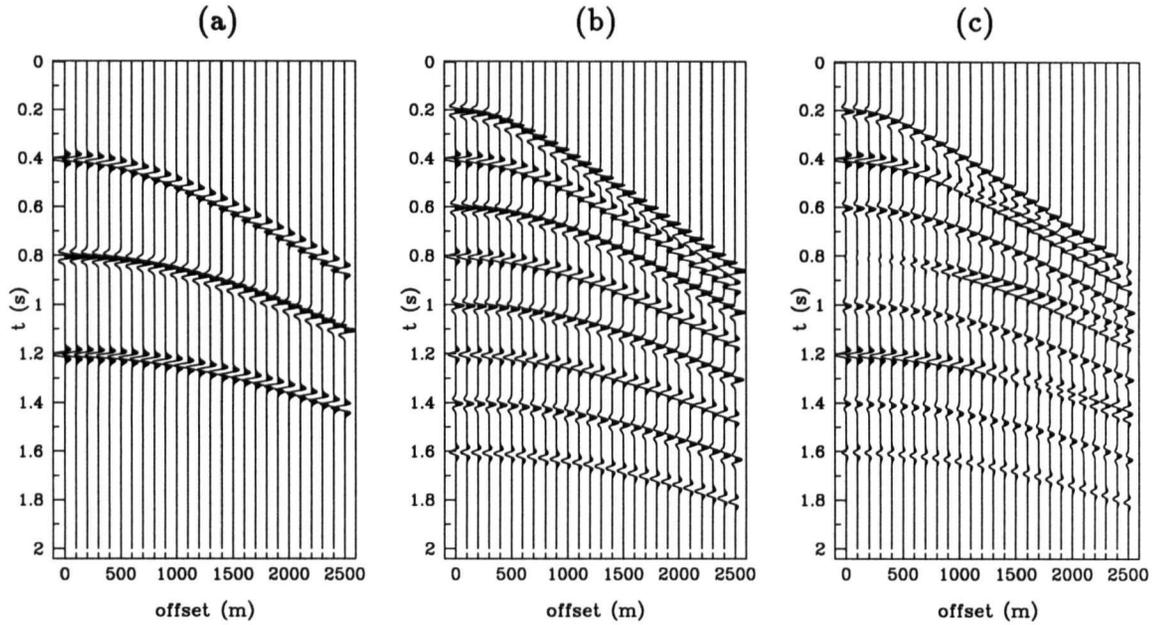


Figure 4.12: (a) A synthetic CMP gather with three primary events; (b) a primary reflection and its multiples; (c) composite CMP gather (primaries and multiples).

( $p = 1$ ), the smearing is drastically reduced. The resulting velocity panels are displayed in Figures 4.13c and 4.13d.

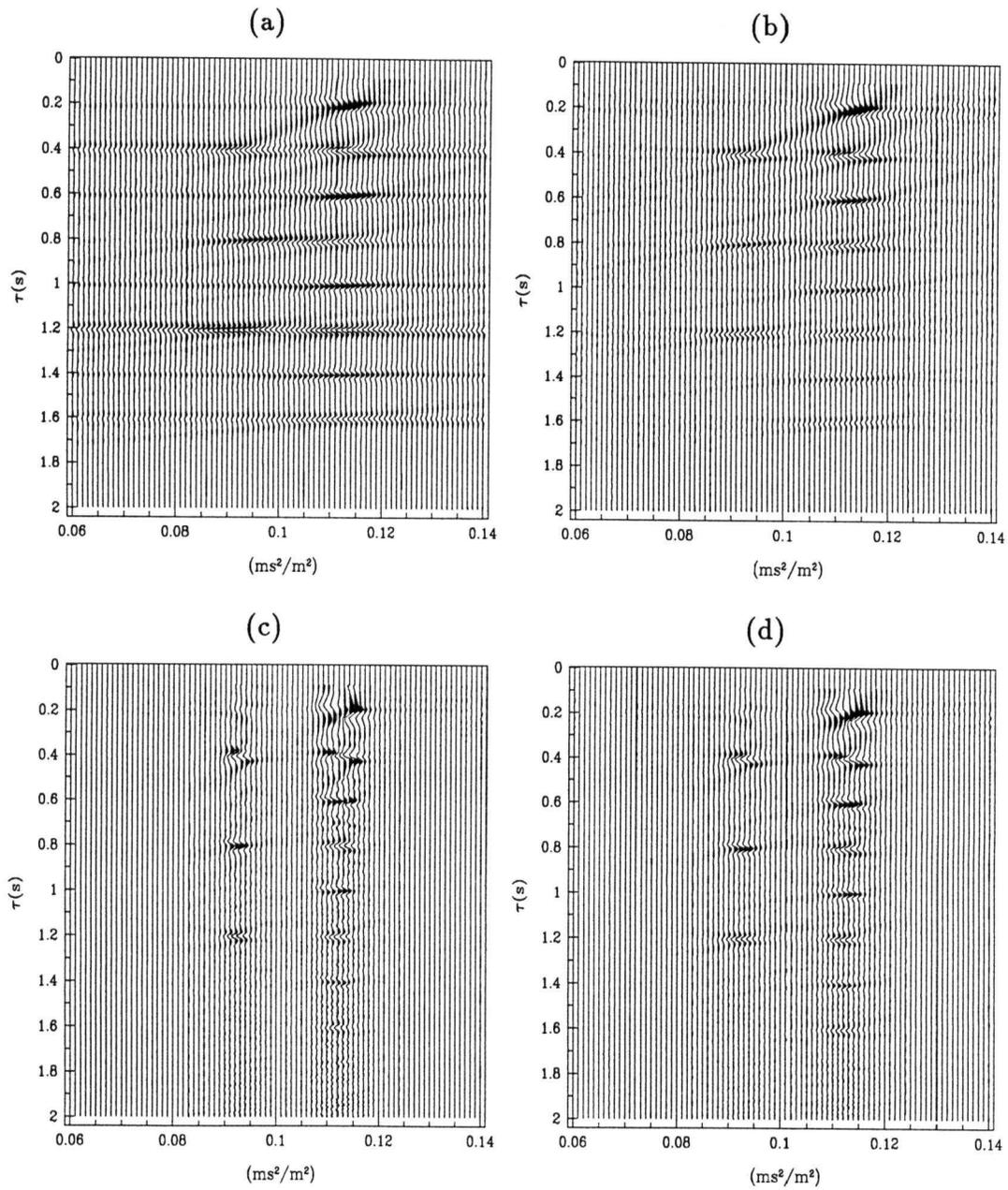


Figure 4.13: (a) Velocity processing with the conventional parabolic stack operator. Panel (b) shows the  $\tau - q$  panel computed with the damped least squares solution. (c)-(d) IMRLS solutions, generalized Gaussian prior ( $p = 1$ ) and the Cauchy prior, respectively.

### 4.5.2 Offset space reconstruction from parabolic stacks

The simulations presented in this section are primarily intended to examine the focusing power of the parabolic transform when applied to velocity processing. Figure 4.14a shows two seismic reflections of 1800 and 2200m/sec. The central frequency of the seismic source is 20 Hz. The record was contaminated with band-pass Gaussian noise with  $\sigma_n = 0.05$  (relative to the signal energy). Four traces have been removed from the gather to simulate a gap. The  $\tau - q$  space is first computed using the damped least-squares and then the IMRLS algorithm. It should also be noted that only the available traces are used in the computations. This is the principal difference with the technique proposed by Kabir and Verschuur (1995), where the gaps are loaded into the procedure as zero traces which are iteratively restored. Figure 4.14b and 4.14c show the Radon space computed using damped least squares and the IMRLS procedure after 5 iterations, respectively. Figures 4.15a and 4.15b show the reconstructed offset space corresponding to the damped least squares solution and to the IMRLS algorithm with the Cauchy prior. Figure 4.15c portrays the traces that were extracted from the data to produce the gap, together with the recovered traces. The central four traces correspond to Figure 4.15a while the rightmost segment to Figure 4.15b. It is evident from Figure 4.15c, that diminishing smearing in the Radon space improves the reconstruction of missing traces.

In the next example, we applied the parabolic transform to the same synthetic CMP, however, in this example the parabolic approximation is validated by means of a NMO correction. The data were corrected with a constant velocity  $V_c = 2640\text{m/s}$ . Figures 4b and 4c show the  $\tau - q$  panels computed using damped least squares and the IMRLS algorithm. The reconstructed data and a blow up of the missing traces are portrayed in Figures 4.17a, 4.17b and 4.17c, respectively. It appears that the parabolic transform can better model the reflections after a NMO correction than after a  $t^2$  transformation. In

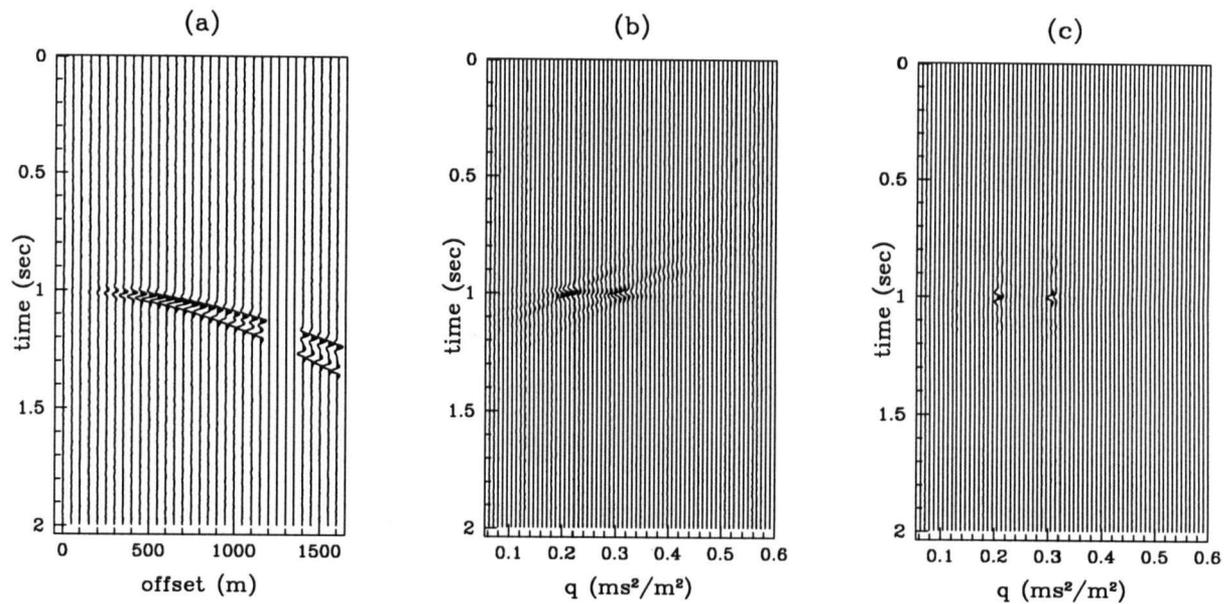


Figure 4.14: Parabolic processing after a  $t^2$  transformation. a) Synthetic CMP gather. b)  $\tau - q$  domain computed with damped least squares. c)  $\tau - q$  panel computed with model re-weighted least squares procedure. The parameter  $q$  is the squared slowness.

the latter case, the reconstruction does not strongly depend on the resolution in model space. In fact, Figure 4b is not very focused but it does give a good reconstruction of the missing traces.

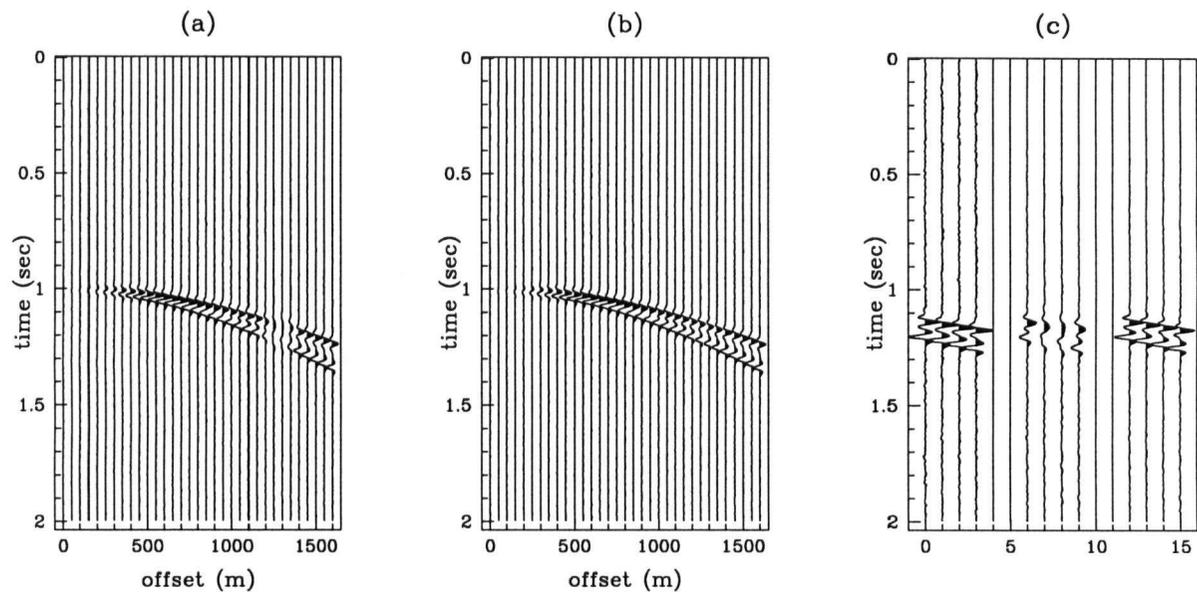


Figure 4.15: Offset space reconstruction after parabolic transform processing. The parabolic approximation is validated using a  $t^2$  transformation. a) Reconstructed offset from the  $\tau - q$  panel portrayed in Figure 4.14b. b) Reconstructed offset from the  $\tau - q$  panel portrayed in Figure 4.14c. c) Original missing traces, reconstructed missing traces from the  $\tau - q$  space: damped least squares (center), model re-weighted least squares (right).

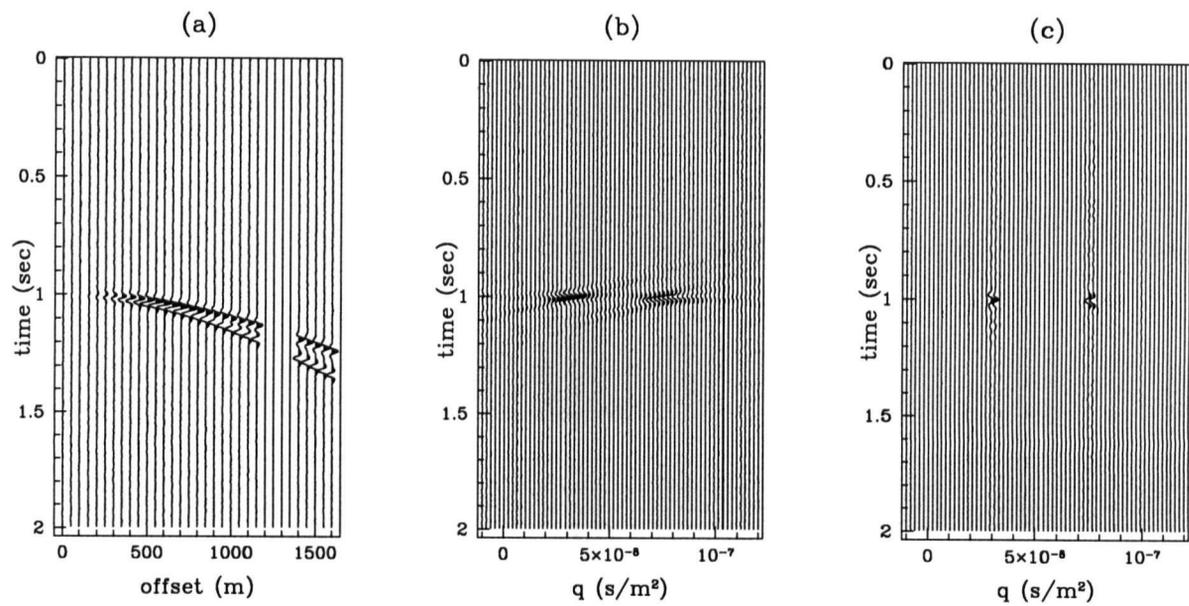


Figure 4.16: Parabolic processing after a NMO correction. a) Synthetic CMP gather. b)  $\tau - q$  domain computed with damped least squares. c)  $\tau - q$  panel computed with model re-weighted least squares procedure.

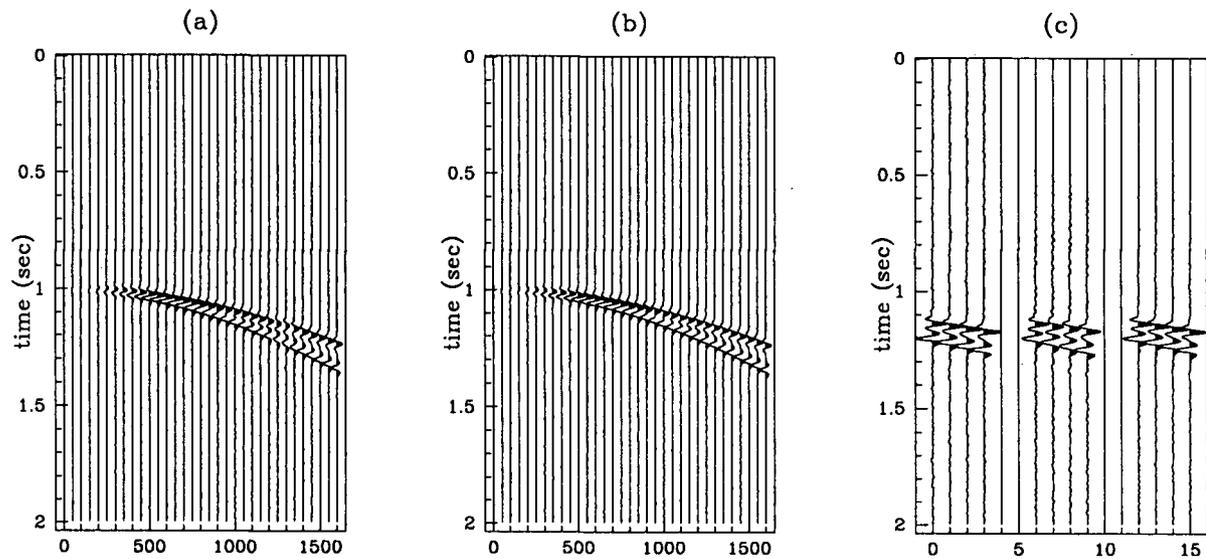


Figure 4.17: Offset space reconstruction after parabolic processing. A NMO correction is applied to validate the parabolic approximation. a) Reconstructed offset from the  $\tau - q$  panel portrayed in Figure 4.16b. b) Reconstructed offset from the  $\tau - q$  panel portrayed in Figure 4.16c. c) Original missing traces, reconstructed missing traces from the  $\tau - q$  space: damped least squares (center), model re-weighted least squares (right).

### 4.5.3 Field data example

Figure 4.18a illustrates a shot record acquired in deep water which consists of 42 traces gathered every 50 m. The inner offset is located at 230 m. The shot has four missing traces located at 1830, 1880, 1930, 1980 m. We consider the parabolic stacking after a  $t^2$  transformation and after the NMO correction. Figure 4.18b portrays the shot gather after the  $t^2$  transformation. Figure 4.19a shows the  $\tau - q$  space computed using zero order regularization or damped least squares.

The parameter  $q$  in this case is equal to the squared slowness of each event. The  $\tau - q$  panel was used to restore the shot gather (Figure 4.19b). The new near offset trace is located at 30m. The near offset observations are restored but the gap at far offsets is not properly restored. The data were also processed with the IMRLS algorithm using the Cauchy prior. The results are portrayed in Figure 4.20a ( $\tau - q$  panel) and 4.20b (restored data). The missing far offset traces are partly restored. I found that complete restoration of all the events is very difficult. The noise and the imperfections of the hyperbolic paths inhibit a perfect localization of each event and therefore not all the events are properly modelled.

To continue with the analysis, I consider the effect of parabolic stacking after a NMO correction. Figures 4.21a and 4.21b portray the original shot record and the record after the NMO correction, respectively. The  $\tau - q$  panel computed with the damped least squares solution is plotted in Figure 4.22a. The restored seismic record is plotted in Figure 4.22b. In this case the missing inner offset as well as the far offset traces are partly recovered. After the NMO correction the curvature of the reflections is small and hence, the operator can properly localize most of the events. The IMRLS solution is portrayed in Figure 4.23a. A comparison of Figures 4.23a and 4.22a shows some interesting features. It is clear that the Cauchy regularization improves the focusing power of the parabolic

transform, which is particularly clear from a comparison of the reflection at  $\approx 1.4$ sec. Yet, we have to recognize that there is a resolution tradeoff. It is evident that the Cauchy regularization enhances the resolution but it is also true that the noise is not properly attenuated.

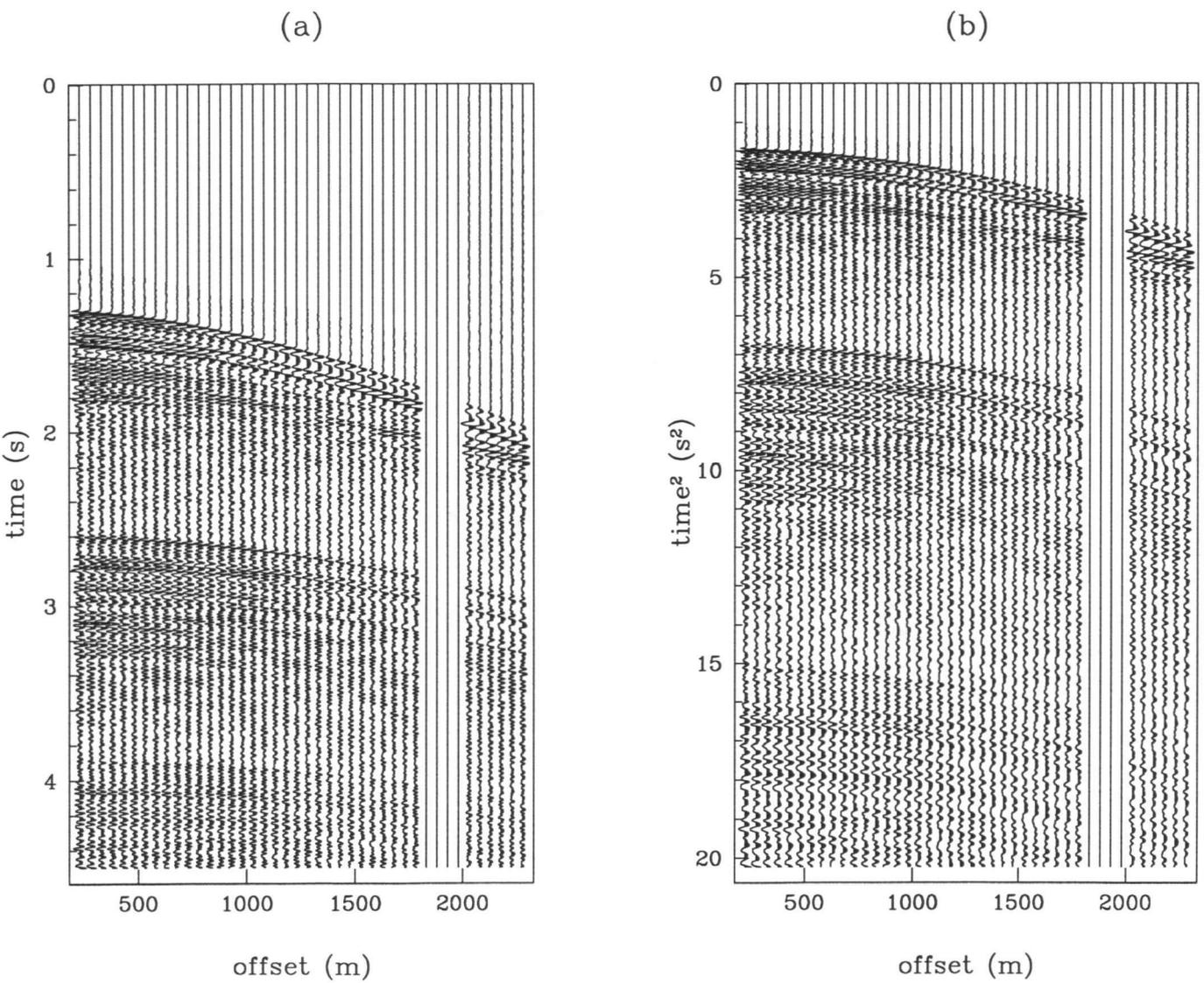


Figure 4.18: a) Shot gather with missing observations. The near offset trace is located at 230m. b) Shot gather after a  $t^2$  transformation to validate the parabolic approximation.

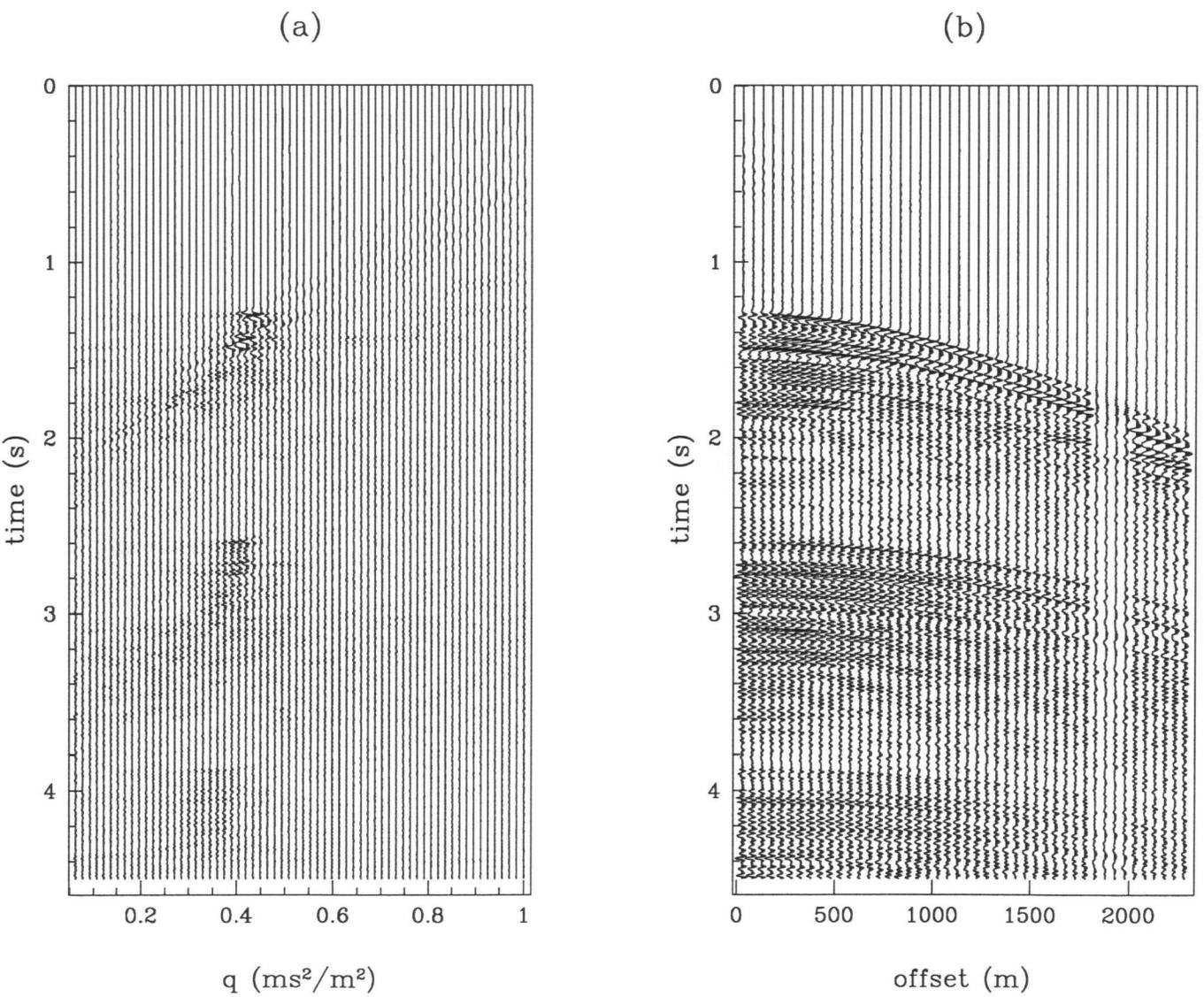


Figure 4.19: Parabolic transform processing after a  $t^2$  transformation. a)  $\tau - q$  panel obtained using damped least squares. b) Reconstructed offset space, the new inner offset is located at 30m.

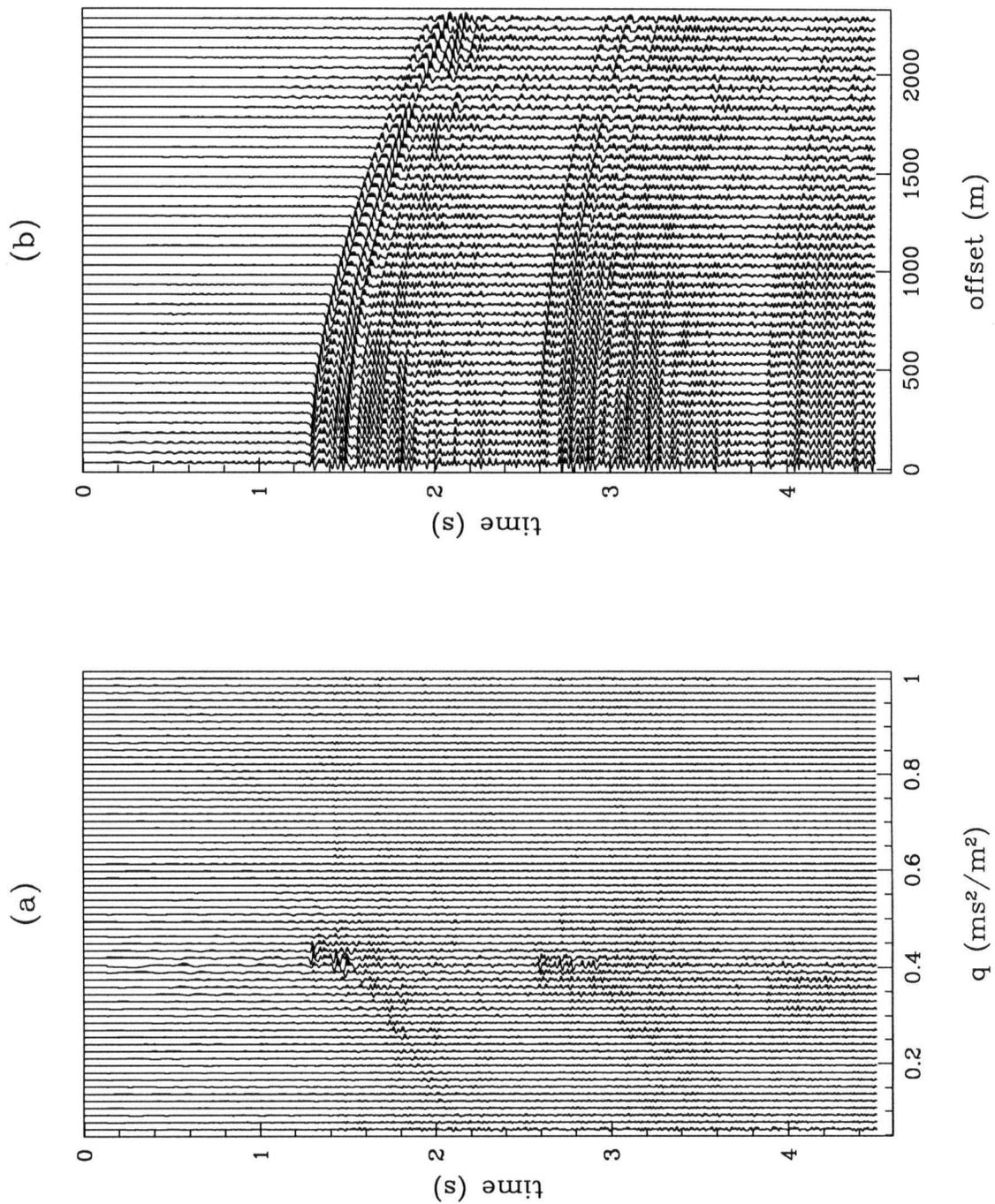


Figure 4.20: Parabolic processing after a  $t^2$  transformation. a)  $\tau$ - $q$  panel obtained using model re-weighted least squares (Cauchy prior). b) Reconstructed offset space, the new inner offset is located at 30 m.

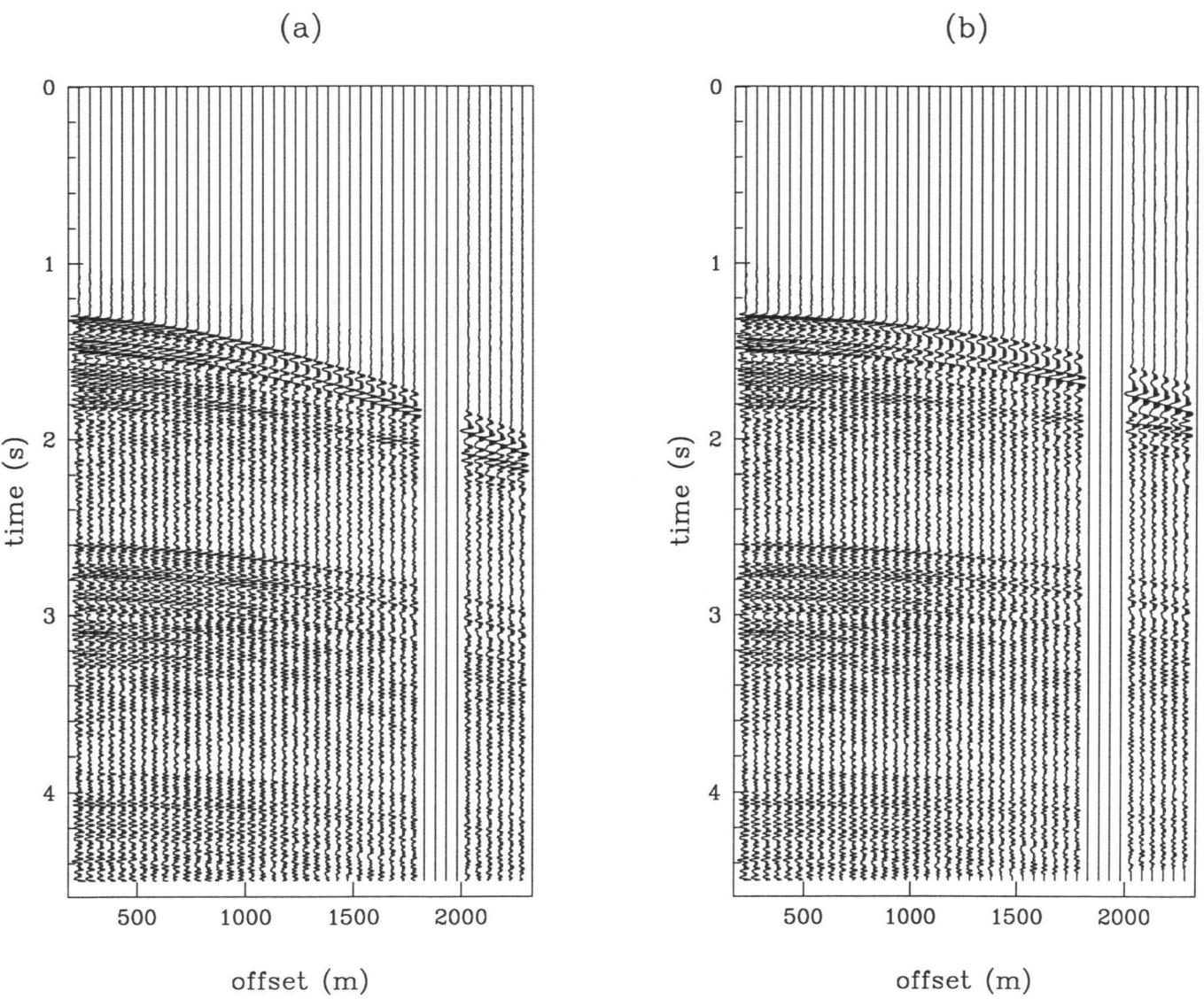


Figure 4.21: a) Shot gather with missing observations. b) Shot gather after a NMO correction to validate the parabolic approximation.

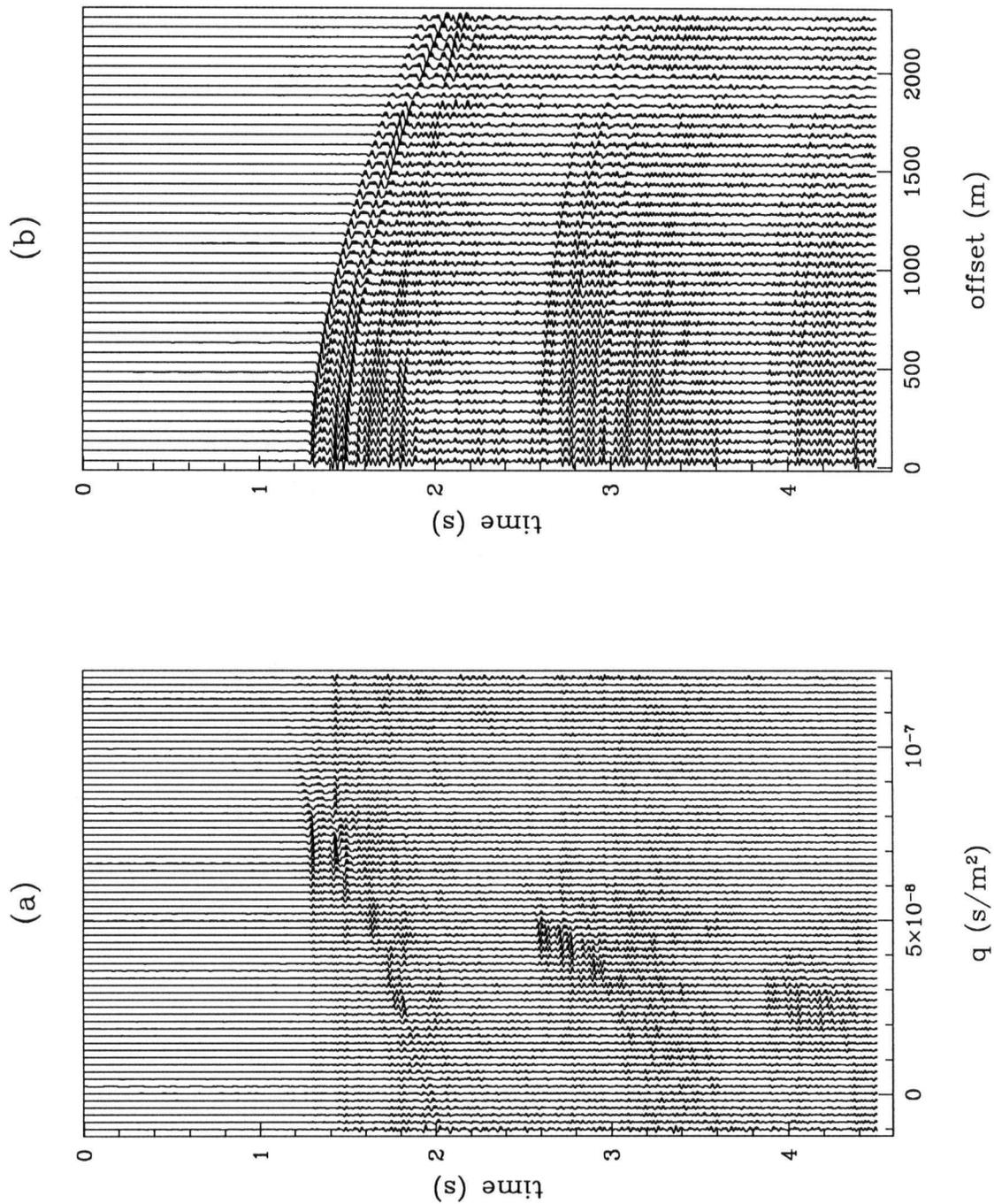


Figure 4.22: Parabolic transform processing after a NMO correction. a)  $\tau - q$  panel obtained using damped least squares. b) Reconstructed offset space, the new inner offset is located at 30m.

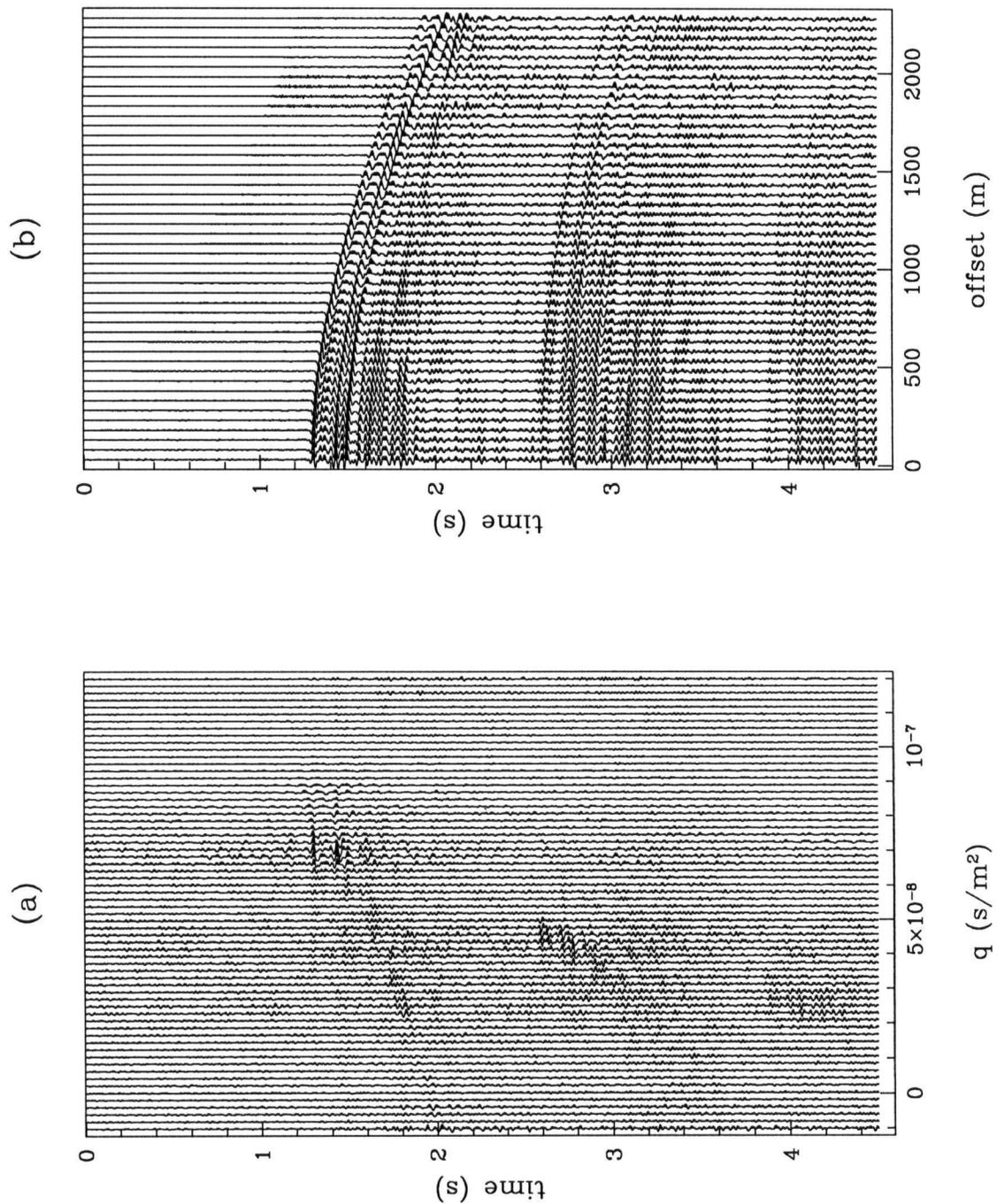


Figure 4.23: Parabolic processing after NMO correction. a)  $\tau - q$  panel was obtained with model re-weighted least squares (Cauchy prior). b) Reconstructed offset space, the new inner offset is located at 30m.

#### 4.6 Concluding remarks and discussion

It has been shown that amplitude smearing can be considerably diminished using a sparse prior to derive slant stack and parabolic stack operators. Two different criteria, which are related to well studied probability distributions, were presented. I have not found substantial differences between the  $l_p$  ( $p = 1$ ) or the Cauchy regularization. It appears that any prior that drives the solution to sparseness is adequate to diminish amplitude smearing.

The algorithm has also proved to be capable of restoring missing traces. The latter is subject to the validity of the parabolic approximation after NMO correction or to the relative absence of the effect of wavelet stretching following the  $t^2$  transformation. The validity of the parabolic approximation cannot be satisfied for any time-velocity pair of the CMP or CSP gather. Besides, we can generate highly resolved solutions for a particular range of the velocity spectrum of the data. On the other hand, the parabolic transform after a  $t^2$  transformation maps each hyperbola into a parabola. It must be realized, however, that shallow reflections can suffer a non-negligible frequency distortion due to the  $t^2$  mapping.

## Chapter 5

### Aperture Compensated Discrete Fourier Transform and Applications

*...porque supone la imposible adición del instante presente y los pretéritos.*

Jorge Luis Borges: *Tlön, Uqbar, Orbis Tertius*

#### 5.1 Introduction

Spatio-temporal analysis of seismic records is of particular relevance in many geophysical applications, e.g., vertical seismic profiles and plane wave slowness estimation in seismology. The goal is to estimate from a limited number of receivers the 2-D spectral signature of a group of events which are recorded on a linear array of receivers. The conventional analysis based on 2-D Fourier transformation might result in poorly resolved spectral panels due to the presence of sidelobes which tend to mask the signals. This is more noticeable when the spatial aperture of the signal is small compared with the range of wavenumbers we are seeking. The combined effect of the sidelobes and noise make the problem even more severe.

Many of the 1-D high resolution spectral analysis techniques which are cited in the literature can be easily extended to the 2-D case. A review of these methods can be found in Kay and Marple (1981). Two dimensional extensions of these procedures are given by Marple (1987). The power spectrum estimate is computed by honoring a few lags of the autocorrelation function (1-D or 2-D depending the problem). Examples of these approaches are 2-D parametric spectral analysis (Lim and Malik, 1982) and the

Capon maximum likelihood method (Capon, 1969). Hybrid techniques can be designed by applying any 1-D high-resolution spectral method to the rows of an auxiliary array composed of column vectors obtained with the discrete Fourier transform (DFT). Since the auxiliary array is complex, high resolution algorithms in complex form are necessary (Marple, 1987). The hybrid scheme serves to improve the resolution of only one of the variables (temporal frequency,  $f$ , or wavenumber,  $k$ .) In array processing, the spatial resolution is limited by the small number of receivers relative to the number of time samples of each trace. In other words, the spatial coverage dictates the requirement of high resolution methods.

Ulrych and Walker (1981) and more recently Swingler and Walker (1989) applied linear prediction to extrapolate the data and simulate a longer array which can be analyzed using conventional spectral analysis. The idea of creating a synthetic aperture (a longer array) is also explored here.

Subspace methods can certainly provide high resolution but their performance is severely affected when the number of signals are over or underestimated. In fact, these methods are based on a 2-D extension of Pisarenko's harmonic spectral analysis (Pisarenko, 1973) which, as is well known, is very sensitive to the number of harmonics that compose the process. A similar problem is also encountered in parametric spectral estimation, where a knowledge of the number of parameters is vital in obtaining reliable estimates. This is the usual tradeoff between resolution and statistical reliability.

Since the techniques described above provide a high resolution power spectral density (PSD) estimate, only the autocorrelation function of the data can be reconstructed. A different approach to the problem of determining a high resolution 2-D PSD estimate is explored in this chapter. First, the constraints are the data rather than the autocorrelation function. Secondly, the unknown is an un-windowed DFT, or in other words a DFT where truncation artifacts (sidelobes) are mitigated. The regularization strategy

developed in chapter 4 is applied to compute the high resolution DFT. A by-product of the method is a high-resolution periodogram (Sacchi and Ulrych, 1995a). This periodogram coincides with the periodogram that would have been computed with a longer array of receivers if the data consist of a limited superposition of planes waves. This assumption may be validated by using small spatio-temporal windows. Windowing is a common manner to satisfy stationarity requirements.

The technique is useful in array processing for two reasons. First, it provides spatial extrapolation of the array (subject to the above data assumption) and secondly, missing receivers within and outside the aperture are treated as unknowns rather than as zeros.

Synthetic and field data is used to assess the applicability of the technique. Finally, I would like to stress that this chapter expands the idea of sparse regularization to the problem of spectral estimation. Moreover, this chapter serves to place the spectral estimation problem within the linear inverse theory scenario.

## 5.2 The discrete Fourier transform as an inverse problem

For simplicity I will start with the 1-D DFT since extensions to higher dimensions are straightforward. Consider a  $N$ -sample time or spatial series  $x_0, x_1, x_2, \dots, x_{N-1}$ . The DFT of the discrete series is given by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi nk/N} \quad k = 0, \dots, N-1, \quad (5.1)$$

and similarly, the inverse DFT is given by

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{i2\pi nk/N} \quad n = 0, \dots, N-1. \quad (5.2)$$

Now, assume that we want to estimate  $M$  spectral samples where  $M > N$ . A standard approach to solving this problem is by means of zero padding. Defining a new time series consisting of the original series plus a zero extension for  $n = N, \dots, M - 1$ , one can estimate  $M$  spectral samples using the DFT. This procedure helps to remove ambiguities due to discretization of the Fourier transform but, as is well known, it does not reduce the sidelobes created by the temporal/spatial window or improve the resolution. Estimation of  $M$  spectral samples without zero padding may be posed as a linear inverse problem where the target is a DFT that is consistent only with the available information. At this point, it is interesting to note that the underlying philosophy is similar to Burg's maximum entropy method (MEM) (Burg, 1975). However, in Burg's MEM the target is a PSD estimate.

Rewriting equation (5.2) as

$$x_n = \frac{1}{M} \sum_{k=0}^{M-1} X_k e^{i2\pi nk/M} \quad n = 0, \dots, N - 1, \quad (5.3)$$

gives rise to a linear system of equations

$$\mathbf{d} = \mathbf{Fm}, \quad (5.4)$$

where the vector  $\mathbf{d} \in \mathbf{R}^N$  and  $\mathbf{m} \in \mathbf{C}^M$  denote the available information and the unknown DFT, respectively. Equation (5.4) is a linearly underdetermined problem which, as is well known, can be satisfied by many different solutions. Uniqueness is imposed by defining a regularized solution,  $\hat{\mathbf{m}}$ , which may be obtained by minimizing the following cost function

$$J(\mathbf{m}) = \Phi(\mathbf{m}) + \|\mathbf{d} - \mathbf{Fm}\|_2^2. \quad (5.5)$$

The regulariser  $\Phi(\mathbf{m})$  serves to impose a particular character on the solution.

### 5.3 Zero order quadratic regularization (damped least squares)

To start with, I will assume that the data are contaminated with Gaussian noise and that the samples of the DFT can be modelled with a Gaussian distribution. For simplicity, I will assume that the discretization of both spaces is regular and that the covariance matrices of the model and noise are diagonal matrices. Under this assumption, one may use Bayes' rule to estimate a MAP estimator of the DFT. This assumption leads to the minimization of the following objective function

$$J_{gg} = \lambda \mathbf{m}^H \mathbf{m} + (\mathbf{d} - \mathbf{Fm})^H (\mathbf{d} - \mathbf{Fm}). \quad (5.6)$$

As usual, the first term represents the model norm while the second term is the misfit function. Taking derivatives and equating to zero yield

$$\mathbf{m} = (\mathbf{F}^H \mathbf{F} + \lambda \mathbf{I}_M)^{-1} \mathbf{F}^H \mathbf{d}. \quad (5.7)$$

The last expression can be rewritten using the following identity

$$(\mathbf{F}^H \mathbf{F} + \lambda \mathbf{I}_M)^{-1} \mathbf{F}^H = \mathbf{F}^H (\mathbf{F} \mathbf{F}^H + \lambda \mathbf{I}_N)^{-1}, \quad (5.8)$$

where  $\mathbf{I}_M$  and  $\mathbf{I}_N$  represent  $M \times M$  and  $N \times N$  identity matrices, respectively. Recalling equation (5.8) and that  $\mathbf{F} \mathbf{F}^H = \frac{1}{M} \mathbf{I}_N$ , we end up with the following expression

$$\mathbf{m} = \left(\frac{1}{M} + \lambda\right)^{-1} \mathbf{F}^H \mathbf{d}. \quad (5.9)$$

The result is nothing else than the DFT of  $x_n$  modified by a scale factor. The solution expressed by equation (5.9) becomes

$$\hat{X}_k = \frac{1}{1 + \lambda M} \sum_{n=0}^{N-1} x_n e^{-i2\pi nk/(M-1)}, \quad (5.10)$$

which is the DFT of the windowed time series (equation (5.1)) and is equivalent to padding with zeros in the range  $n = N, \dots, M - 1$ . It is easy to see that this regularization yields a scaled version of the DFT. Hence, the associated periodogram exhibits a resolution that is proportional to the inverse of the length of the time series (the usual resolution criterion for the periodogram.) The periodogram becomes,

$$P_k = \hat{X}_k \hat{X}_k^*, \quad k = 0, \dots, M - 1. \quad (5.11)$$

It is interesting to point out that Oldenburg (1976) arrived to the same conclusion using the Backus and Gilbert (B&G) formalism which is a completely different approach to inversion. Specifically, when the first Dirichlet criterion of B&G theory is used the resulting expression leads to an expression equivalent to equation (5.10). This is a simple consequence of using this criterion; other spreading functions lead to different results. An important conclusion which may be drawn from Oldenburg's paper is that the resolution is fixed and cannot be modified. An intuitive explanation is as follows; since the resolution of the PSD is controlled by the width of the sidelobes which are independent of the scaling

factor, the DFT computed with the Gauss-Gauss model will lead to a PSD estimator that is equivalent to the periodogram of the truncated time series.

#### 5.4 Regularization by the Cauchy-Gauss model

In the previous section I showed that the Gauss-Gauss regularization does not serve the purpose of obtaining a high resolution PSD estimator. The sparse regularization scheme used to compute Radon operators may be used here to estimate a high resolution DFT. The reason is simple and can be interpreted as follows; in general, the interpretation of spatio-temporal data is involved with a finite number of events which exhibit spatial continuity. If the data show great complexity, time and/or spatial windows are chosen such that particular events may be more easily distinguished and mapped. Because of the limited number of such events, it is appropriate to require a model which consists of the minimum number of events that can satisfy the data. This is the classic idea of simplicity or parsimony and quite opposite in philosophy to norms which describe the “structure” of a surface by means of, for example, the smallest, flattest or smoothest models. A “long tailed” distribution, like the Cauchy pdf, will induce a model consisting of only few elements different from zero. The objective function of the problem becomes,

$$J_{cg} = J_c(\mathbf{m}) + \frac{1}{\sigma_n^2}(\mathbf{d} - \mathbf{Fm})^H(\mathbf{d} - \mathbf{Fm}), \quad (5.12)$$

The function,  $J_c(\mathbf{x})$ , which is expressed by

$$J_c(\mathbf{x}) = \sum_{k=0}^{M-1} \ln\left(1 + \frac{X_k X_k^*}{\sigma_c^2}\right)$$

is the regulariser imposed by the Cauchy distribution and is a measure of the sparseness

of the vector of spectral powers  $P_k = X_k X_k^*$ ,  $k = 0, \dots, M - 1$ . The constant  $\sigma_c$  controls the amount of sparseness that can be attained by the inversion. The sparseness of the estimate will also depend on the noise level since  $\sigma_n$  may inhibit a reliable sparse solution.

Taking derivatives of  $J_{cg}(\mathbf{x})$  and equating to zero yields the following result

$$\mathbf{m} = (\lambda \mathbf{Q}^{-1} + \mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H \mathbf{d}, \quad (5.13)$$

where  $\lambda = \sigma_n^2 / \sigma_c^2$  and  $\mathbf{Q}$  is a  $M \times M$  diagonal matrix with elements given by

$$Q_{ii} = 1 + \frac{X_i X_i^*}{\sigma_c^2}, \quad i = 0, \dots, M - 1. \quad (5.14)$$

Although expression (5.13) resembles the damped least squares solution, we note that  $\mathbf{Q}$  is nonlinearly related to the DFT of the data. Equation (5.13) can be rewritten as

$$\mathbf{m} = \mathbf{Q} \mathbf{F}^H (\lambda \mathbf{I}_N + \mathbf{F} \mathbf{Q} \mathbf{F}^H)^{-1} \mathbf{d}. \quad (5.15)$$

From the theoretical point of view of uniqueness and convergence, the operators given by equations (5.13) and (5.15) are equivalent. However, from the point of view of computational advantages, the following observations apply

- Whereas equation (5.13) demands the inversion of a  $M \times M$  matrix, equation (5.15) requires the inversion of a  $N \times N$  matrix.
- The operator  $(\lambda \mathbf{I}_N + \mathbf{F} \mathbf{Q} \mathbf{F}^H)$  in equation (5.15) is Toeplitz Hermitian, provided that the time series is uniformly discretized, and a fast solver like Levinson's recursion can be used.

- In the case of nonuniform discretization, a Cholesky decomposition is appropriate.

This type of algorithm is also appropriate in the case of gapped data.

As I have already mentioned, the Cauchy-Gauss model leads to an algorithm that resembles the damped least squares solution. This is particularly true when  $\sigma_c$  is large compared to the spectral amplitudes that we are seeking. In this case the functional  $S(\mathbf{x}) \approx K + \mathbf{x}^H \mathbf{x} / (\sigma_c^2)$  where  $K$  is a constant. Thus, minimizing  $J_{cg}(\mathbf{x})$  is equivalent to minimizing  $J_{gg}(\mathbf{x})$ . In the contrary case, the algorithm will seek a DFT with a sparse distribution of spectral amplitudes leading to an enhancement of the spectral peaks and reducing windowing effects or sidelobes. In the Gauss-Gauss regularization, the scale parameters  $\sigma_n$  and  $\sigma_x$  reduce to a single hyper-parameter,  $\lambda$ , which completely defines the distribution of the unknown model. On the other hand, in the Cauchy-Gauss regularization, we have two independent hyper-parameters,  $\sigma_c, \sigma_n$  or  $\lambda, \sigma_c$ . When the power of the noise is known, the optimum  $\sigma_c$  is computed using any fitting criterion, e.g., the  $\chi^2$  criterion. The simultaneous estimation of  $\sigma_c$  and  $\sigma_n$  is not an easy task since it demands the computation of the marginal joint pdf of  $\sigma_c$  and  $\sigma_n$  for a given solution,  $\hat{\mathbf{x}}$ .

The DFT is iteratively retrieved starting with the windowed DFT. Since the Hessian matrix of  $J_{cg}$  is positive definite, the uniqueness of the solution is guaranteed (section 4.3.8)

## 5.5 Hybrid two-dimensional estimator of the DFT

I present a hybrid procedure based on standard Fourier analysis in the temporal variable while, for the spatial variable, I will carry out the inversion using the Cauchy-Gauss regularization using the IMRLS algorithm. Usually the length of the temporal window is sufficient to achieve high resolution with simple methods based on standard Fourier analysis while the aperture of the array limits the spatial resolution. The two dimensional

algorithm works as follows:

- Each record is transformed to the frequency offset domain using the FFT.
- High resolution analysis is performed at each frequency that comprises the signal using the Cauchy-Gauss regularization.
- The amplitude in the  $f-k$  space is plotted to identify the spatio-temporal structure of each source.
- Alternatively, the data outside the original aperture may be extrapolated to simulate a longer array, and any 2-D spectral technique may be used in conjunction with the extended data set.

Undesired components can be masked before mapping back the un-windowed DFT to space-time. This is demonstrated in detail in the broad band example below.

### 5.5.1 First example: spatio-temporal spectrum narrow band signals

The IMRLS algorithm is applied to estimate the spatio-temporal spectrum of a signal received by a passive array of receivers. This problem frequently arises in radar and sonar processing (Bienvenu and Kopp, 1983). The goal is to estimate the direction of arrival and the temporal spectral signature of a set of sources impinging from different angles on a uniform array of  $N$  receivers. In seismology, the problem has been particularly studied to detect plane-wave signals and estimate the slowness vector. The array of receivers is linear but the method can be easily generalized to any distribution of receivers.

Three narrow band linear events with the following features were generated. The data consist of two sinusoids with unit amplitude and with normalized wavenumbers of 0.30 and 0.25 units and normalized frequency of 0.20 units, respectively. The third wave

( $f = 0.35, k = -0.25$ ) has an amplitude which is 25% below the amplitude of the first and second waves. The temporal extension of each channel is 150 samples, which represents one order of magnitude above the aperture of the array (15 receivers). Gaussian noise with standard deviation,  $\sigma_n = 0.1$  was added to the composite record. The noise represents 40% of the amplitude of the third wave. Each channel was tapered with a Hamming window. The spatio-temporal spectrum computed using the periodogram is illustrated in Figure 5.1a. The contour lines correspond to normalized amplitudes ranging from 0 to  $-40\text{db}$ , with an interval of  $-5\text{db}$ . The  $f - k$  plane is dominated by sidelobes due to truncation in space and time. This is more noticeable for the wavenumber, since the aperture of the array is one order smaller than the length of the time series. The data were processed with the hybrid procedure based on the Cauchy-Gauss model. The parameters  $\sigma_n$  and  $\sigma_c$  were chosen to reject the noise. The number of iterations varies from frequency to frequency, usually  $\approx 10$  iterations are sufficient to reach the minimum of the objective function,  $J_{cg}$ . The resulting high resolution  $f - k$  panel is portrayed in Figure 5.1b. Contour lines range from 0 to  $-40\text{db}$  as in Figure 5.1a. There is a clear enhancement of the spatial resolution and a suppression of the background noise. Since the contour lines exhibit similar width in the  $k$  and  $f$  direction, we can infer that the normalized aperture is of the same order as the length of the time series. In other words, the  $f - k$  panel portrayed in Figure 5.1b corresponds to an equivalent array of approximately 150 receivers.

### 5.5.2 Second example: broad band applications

The algorithm was tested with two broad-band linear events impinging an array of 25 receivers. The source is modelled with a Ricker wavelet with central frequency,  $f = 20\text{Hz}$ . The first event has slowness  $1.2 \times 10^{-5}\text{s/m}$  and the second  $-1.2 \times 10^{-5}\text{s/m}$ . Since I used a different polarity for each wave, there is destructive interference at near offset traces

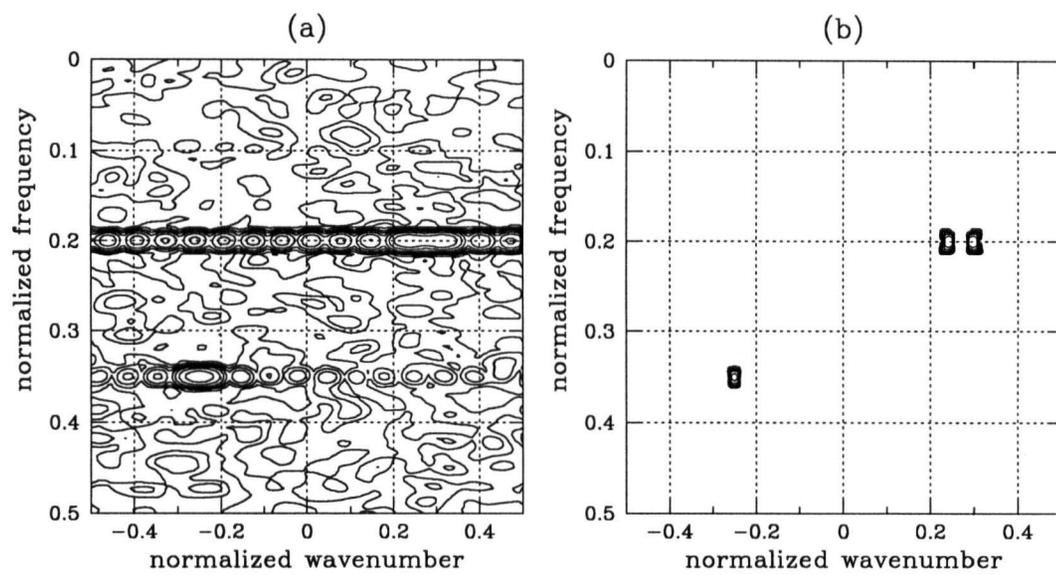


Figure 5.1: 2-D spectrum of three narrow band signals of normalized frequency-wavenumber pairs:  $(0.2, 0.3)$ ,  $(0.2, 0.25)$  and  $(0.35, -0.25)$ . a) Conventional 2-D estimator obtained with the DFT. b) 2-D estimator obtained with the Cauchy-Gauss model.

(Figure 5.2a.) I changed the sign convention of the temporal DFT in order to obtain positive wavenumbers for waves propagating with positive slowness.

First, I examine the noiseless case. Figure 5.2a shows the data while Figure 5.2b illustrates the conventional PSD computed with the DFT. Since the slowness of each event is nearly identical, the conventional  $f - k$  panel cannot distinguish the existence of two signals. The positive and negative wavenumber quadrants were masked to estimate the wave with negative slowness (Figure 5.2c) and the wave with positive slowness (Figure 5.2d). Although both wavefields were decomposed, the decomposition is not correct as is clearly shown by the character of the wavelet which varies with offset in both figures.

The same procedure was carried out using the Cauchy-Gauss inversion scheme. The noise variance was set to zero,  $\sigma_n = 0$ , to fit the data exactly. The parameter  $\sigma_c$  is 0.1% of the maximum amplitude encountered in the raw periodogram which is computed from the finite length DFT. The PSD computed with the Cauchy-Gauss DFT is shown in Figure 5.2e. We note that there is a clear enhancement of the spatial resolution. The wavefields with negative and positive slowness are portrayed in Figures 5.2f and 5.2g, respectively. These panels were computed after masking the corresponding wave number quadrant. Comparing Figures 5.2f and 5.2g with Figures 5.2c and 5.2d it is clear that the high resolution scheme enables us to correctly discriminate both events. Figures 5.2f and 5.2d also show that the wavelet does not suffer substantial changes with offset. Figures 5.2h and 5.2i show the original aperture plus the extrapolated aperture computed with the conventional DFT and with the high resolution DFT, respectively. It is evident that the high resolution DFT corresponds to a longer array of receivers.

Finally, the data were contaminated with Gaussian noise ( $\sigma_n = 0.1$ ). Figure 5.3a portrays the data. Figure 5.3b shows the PSD computed using the DFT (conventional  $f - k$  analysis). Figures 5.3c and 5.3d show the wavefield separation using the conventional DFT analysis. Not only is the character of the wavelet changing with offset, but also the

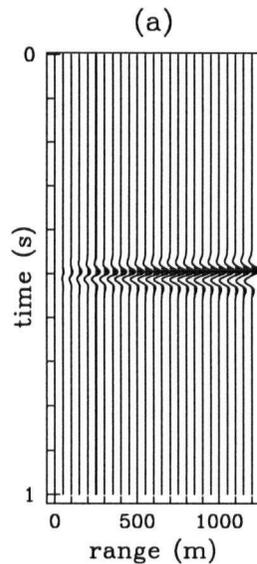


Figure 5.2: a) The synthetic wavefield used to study the performance of the high resolution DFT.

signal-to-noise-ratio has not been improved.

The Cauchy-Gauss PSD is illustrated in Figure 5.3e. When compared with Figure 5.3b, we note that the resolution has been improved and the noise is substantially attenuated. It must be pointed out that, like in the broad-band case, the parameters  $\sigma_c$  and  $\sigma_n$  are chosen to suppress the noise. The Cauchy-Gauss DFT was used to reconstruct the  $t - x$  space after masking the right and left quadrant of the  $f - k$  domain and the results are portrayed in Figures 5.3f and 5.3g. These figures show an accurate separation of each wavefield and an important signal-to-noise-ratio enhancement.

## 5.6 Application to field data. Vertical Seismic Profiling (VSP)

The VSP data are composed of two principal linear wavefields: the down-going and the up-going waves. The down-going waves have the higher amplitude and tend to mask the up-going waves. The data are portrayed in Figure 5.4. As in the previous examples, the

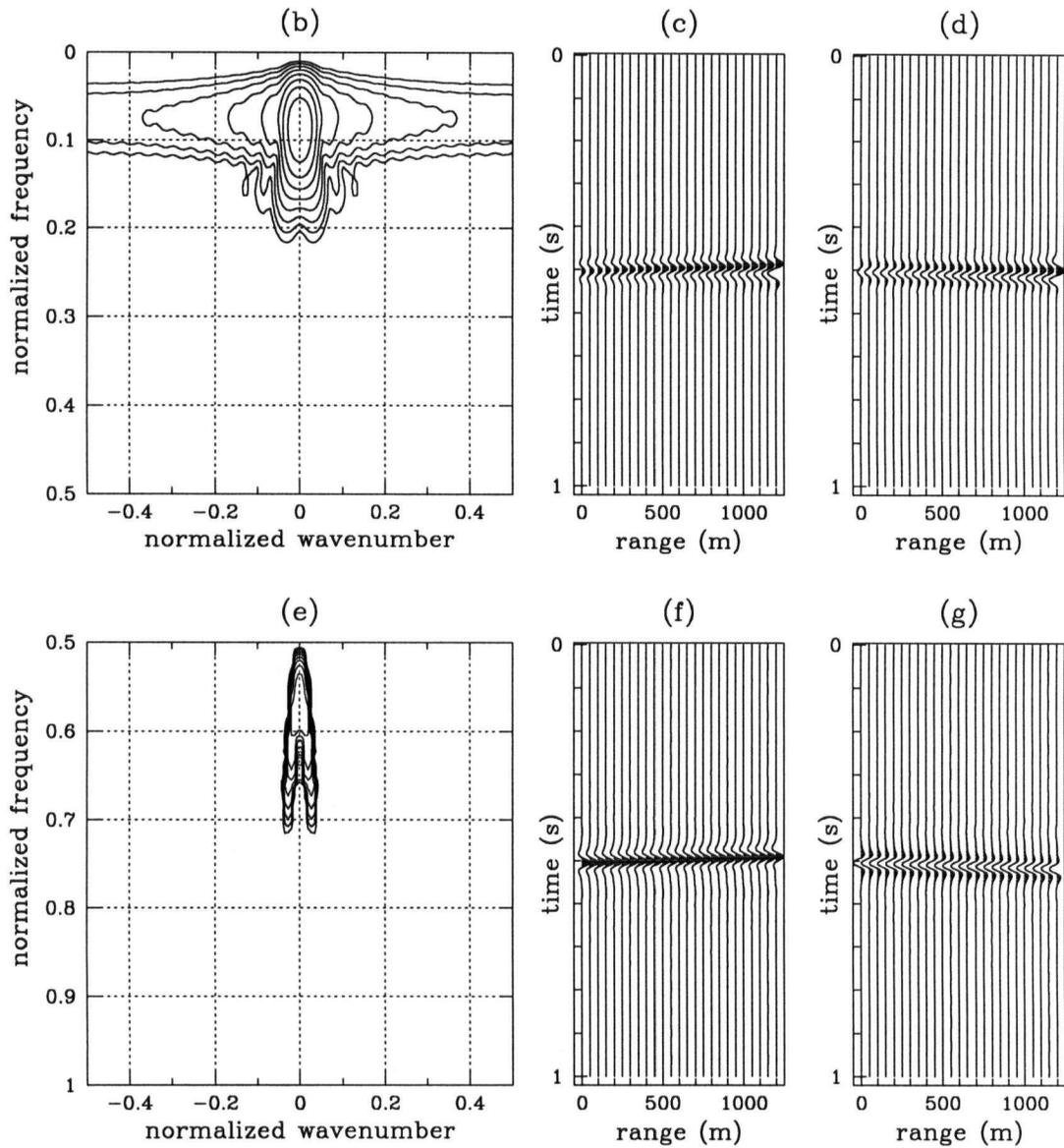


Figure 5.2: b) Spatio-temporal spectrum using the windowed DFT. c-d) Wavefield separation. e) High resolution spectrum computed using the Cauchy regularization. f-g) Wavefield separation based on the high resolution DFT.

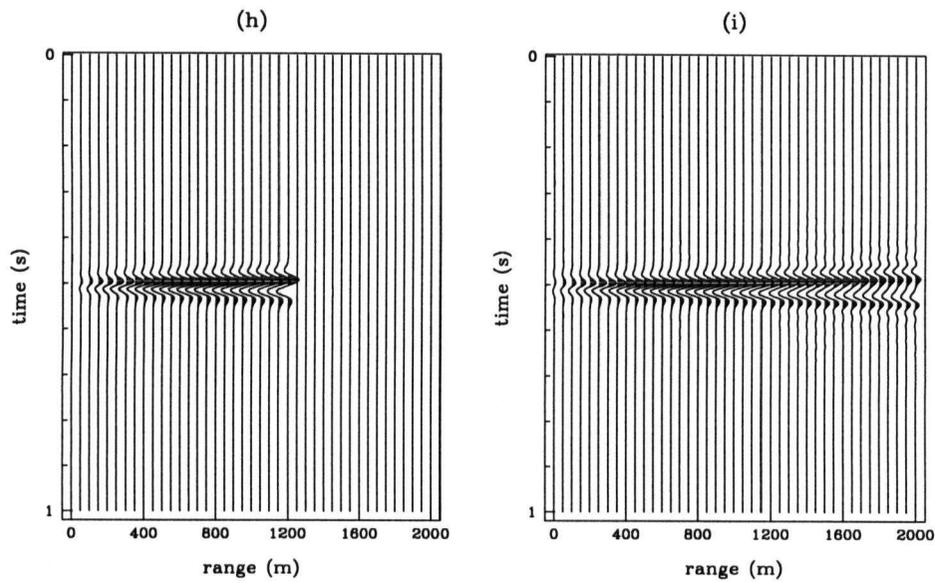


Figure 5.2: i) Wavefield extrapolation using the conventional DFT, h) Wavefield extrapolation using the DFT computed using the sparse regularization

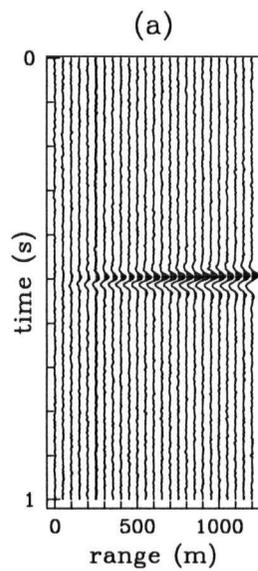


Figure 5.3: a) Synthetic wavefield contaminated with random noise.

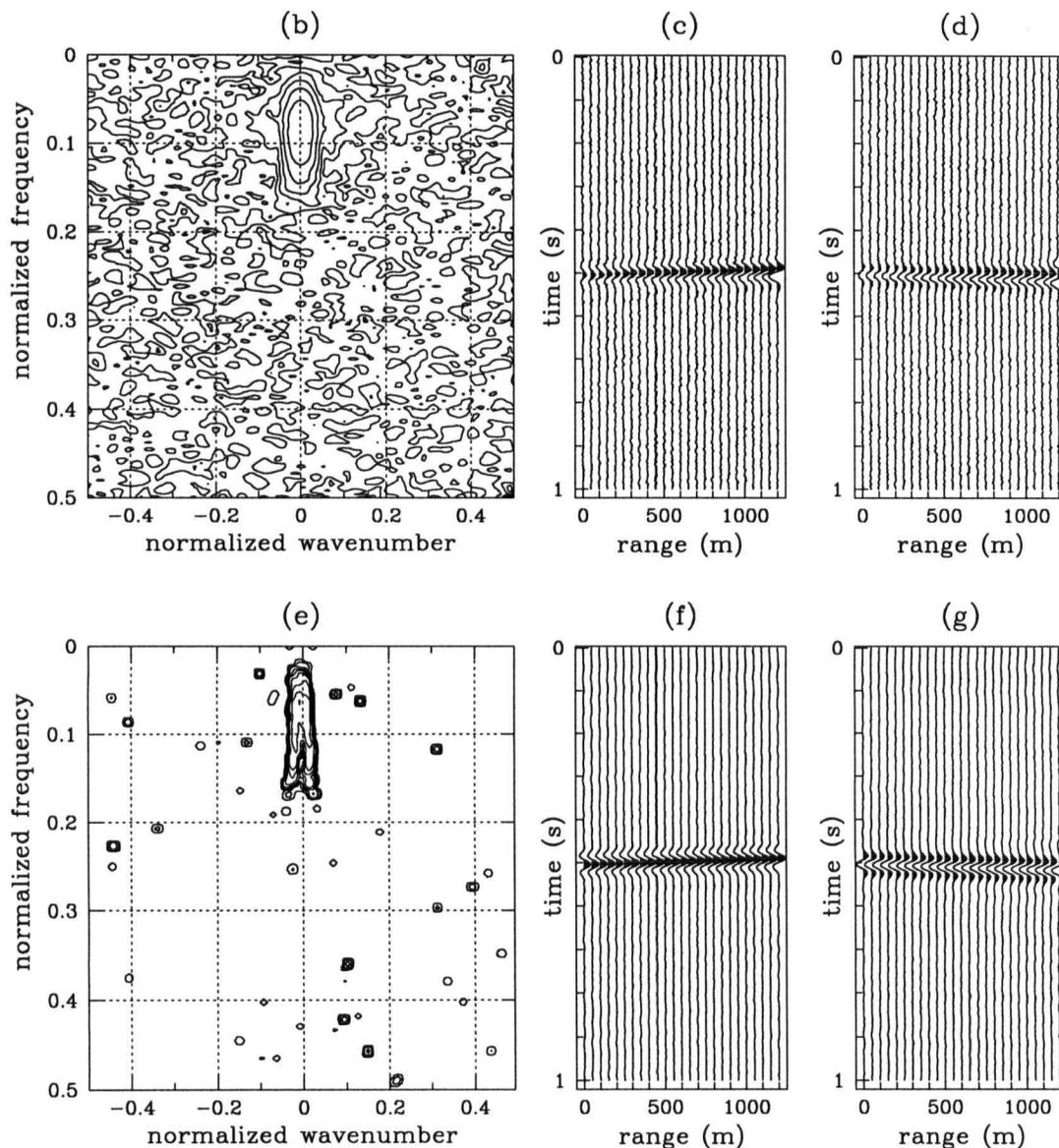


Figure 5.3: b) Spatio-temporal spectrum using the windowed DFT. c-d) Wavefield separation. e) High resolution spectrum computed using the Cauchy regularization. f-g) Wavefield separation based on the high resolution DFT.

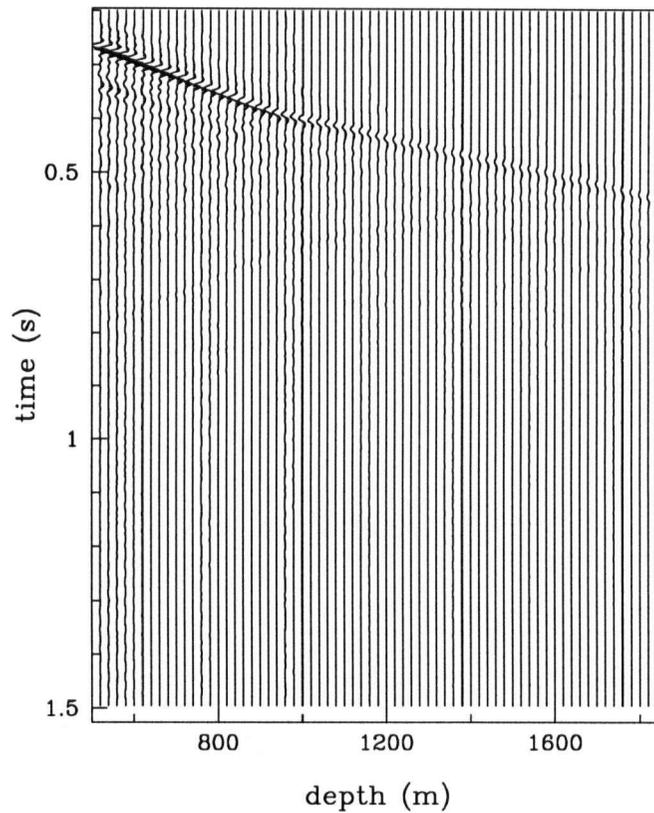


Figure 5.4: Segment of a VSP used to test the high resolution spectral analysis algorithm.

2-D PSD was computed with the conventional DFT and with the high-resolution DFT. The results are portrayed for different windows in Figures 5.5, 5.6 and 5.7. It is clear from a comparison of the  $f - k$  results that the Cauchy-Gauss  $f - k$  result is considerably superior not only in allowing a better identification of different wavefields, but also from the point of view of the design of suitable filters for surgical removal of events.

## 5.7 Discussion

The high resolution technique for the estimation of the power spectrum presented in this chapter is based on the application of an algorithm that seeks a sparse solution to the

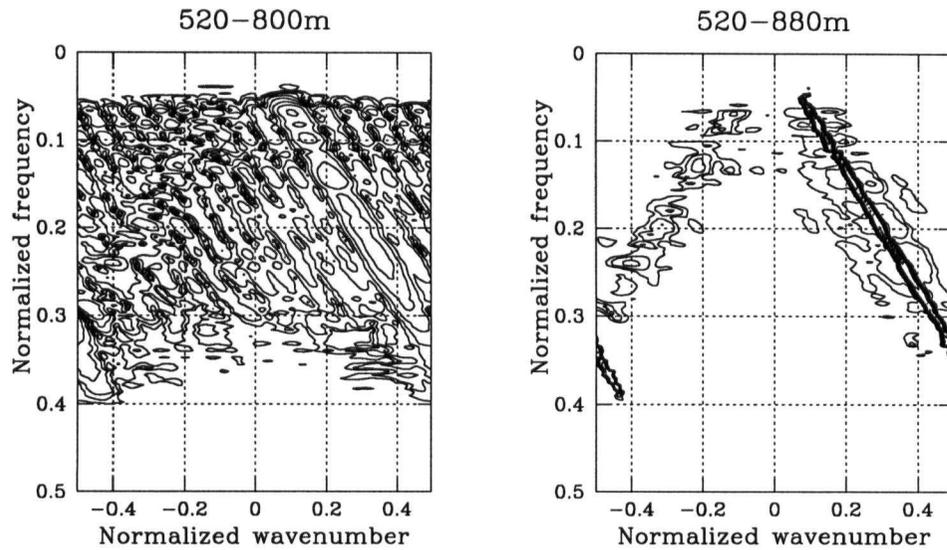


Figure 5.5: 2-D spectral signature of the VSP in the interval 500 – 800m. The left panel illustrates the spectrum computed using the windowed DFT. The right panel is the spectrum computed with the high resolution DFT.

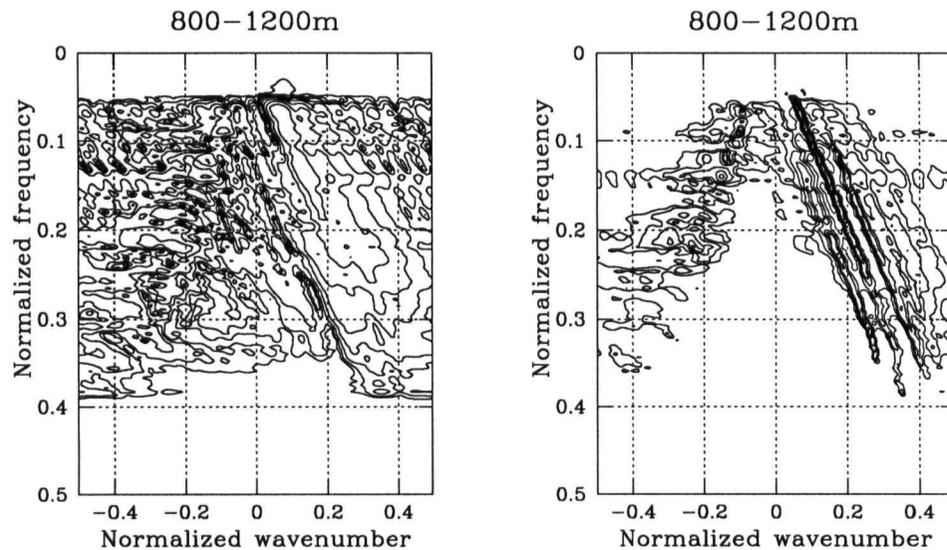


Figure 5.6: 2-D spectral signature of the VSP in the interval 800 – 1200m. The left panel illustrates the spectrum computed using the windowed DFT. The right panel is the spectrum computed with the high resolution DFT.

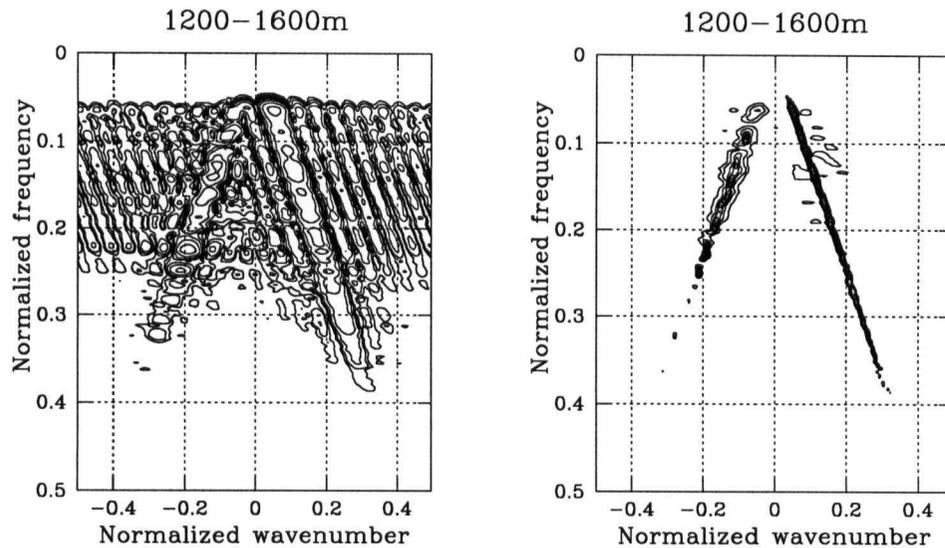


Figure 5.7: 2-D spectral signature of the VSP in the interval 1200 – 1600m. The left panel illustrates the spectrum computed using the windowed DFT. The right panel is the spectrum computed with the high resolution DFT.

ubiquitous problem of spectrum estimation from a finite set of data. What makes the algorithm very attractive is that, since the sparseness measure is minimized subject to data constraints, phase information is also recovered and allows the extrapolation of the signal outside the original window, or aperture, depending on the problem. The latter is consistent with the idea of simulating a longer array and then estimating the PSD using the DFT.

Another attractive feature of the method is that the background noise may be considerably attenuated after tuning the hyper-parameters. The synthetic examples and the VSP example show the potential of the technique in the processing of real data. The technique has wide applicability and a wide range of problems suggest themselves.

It is important to mention that the algorithm is not intended as an alternative to the pivotal FFT algorithm. However, it may be applied in cases where effects related

to aperture limitation give rise to difficulties in the identification and interpretation of closely spaced events.

Finally, I would like to stress the importance of using prior information to improve the resolution of 2-D spectral estimators. It is important to recognize, however, that there is a resolution tradeoff. If the data do not consist of a limited number of plane waves, the sparse assumption is doomed to failure. This facet is constantly present in any inverse problem.

## Chapter 6

### Summary

*I shall use the adjective 'naïve' for any theory, whether realistic or idealistic, that maintains that inferences beyond the original data are made with certainty, and 'critical' for one that admits that they are not, but nevertheless have validity.*

Sir Harold Jeffreys: *Theory of probability*

This thesis explores an inverse approach to the computation of slant stack and parabolic stack operators. In a similar fashion, I have also examined the problem of spectral estimation.

I have demonstrated the applicability of sparse regularization schemes to enhance the resolution of some linear operators that are widely used in geophysics. How to treat missing and truncated information is an important problem in exploration seismology. This thesis contributes to a better understanding of the relationship that exists between resolution/aperture and the regularization of the Radon operator.

In Chapter 2, I have analysed two different approaches to design slant stack operators. The first corresponds to the conventional definition of the Radon pair where a simple linear mapping is used to obtain the  $\tau-p$  space and inversion is needed to recover the data. The second approach, which is conceptually richer, entails the definition of an inverse problem where the target domain is the  $\tau-p$  space. When the problem is posed as an underdetermined linear inverse problem, a minimum norm regularization may be used to enhance the resolution of the slant stack operator. However, the achieved enhancement

may not be enough to correctly separate signals that exhibit similar moveout. An eigen-structure analysis of the slant stack operator based on prolate spheroidal sequences leads to several interesting results associated with stability and resolution. I have outlined a procedure to predict the condition number of the slant stack operator that may be used to optimally design the operator. Similar arguments may be applied to the parabolic stack operator. However, in this case, the eigen-structure of the operator cannot be assessed by analytical means.

I have investigated the development of non-traditional regularization procedures for linear inverse problems. The regularization of the Radon operator is derived using Bayes' rule to incorporate a long-tailed distribution into the inverse problem. Two different priors, which are related to well studied probability densities, were used to derive the regularization strategy: the generalized Gaussian distribution and the Cauchy distribution. When the generalized Gaussian distribution is adopted, we can construct models with different degrees of sparseness by selecting  $p$ , the shape parameter of the distribution. In particular, when  $p = 2$ , the solution corresponds to damped least squares which may be derived using zero-order quadratic regularization or by considering a Gaussian model and a Gaussian likelihood function. The Cauchy prior leads to a cost function that resembles Burg's entropy (Burg, 1975). In particular, it is important to stress that the algorithm (IMRLS) produces a sparse solution which, according to Burg's definition of entropy, also corresponds to a minimum entropy solution.

The Radon transform is basically a dip decomposition scheme. The inverse procedure presented in this thesis, assumes that the decomposition should be in terms of a few events. This assumption is no longer applicable if the data consist of a dense superposition of linear events or of many hyperbolas depending on the problem.

I have shown that the inversion of the Radon operator permits resolution enhancement of the linear and/or parabolic Radon transform. As Thorson and Claerbout (1985) have

pointed out, this is a global strategy that cannot handle amplitude variations. Even in this case, the technique may, up to a degree, enhance the resolution of the Radon operator. Of course, part of the residual smearing that is observed in real  $\tau - p$  or  $\tau - q$  panels is not completely due to limited aperture, but probably to amplitude versus offset variations and the imperfections of the traveltimes curves due to lateral heterogeneity.

As is well known, the solution of an inverse problem strongly reflects the assumptions that were used to pose the problem. If sparseness is the correct feature to look for, then the Cauchy or the generalized Gaussian (with shape parameter  $p = 1$ ) pdf's are good candidates to derive the regularization. I have not found substantial differences between using the generalized Gaussian ( $p = 1$ ) or the Cauchy prior. It is probable that any prior that leads to a sparse solution is adequate to diminish smearing in the Radon domain.

In Chapter 5, I applied the Cauchy regularization to retrieve a high resolution 2-D power spectrum estimator. The algorithm is best suited to the analysis of plane waves (undamped harmonics in the 1-D case). In a classical spectral estimation context, the power spectrum is retrieved by honoring a few autocorrelation constraints. I prefer to use data constraints to compute a high resolution discrete Fourier transform and then, as a by product, compute the power spectrum. As I have already mentioned, resolution enhancement is not always possible. If the data do not consist of a limited superposition of plane waves, or if strong amplitude variations are present, a more conservative approach may yield more reliable results. In this case, a simple periodogram rather than any high resolution spectral method is more appropriate to analyse the data. This is simply a consequence of using a norm that mimics a sparse distribution of parameters. Hopefully, further work on aperture compensation will be incorporated in many industrial applications associated with seismic imaging.

## References

- Aki, K., and Richards, P. G., 1980, Quantitative Seismology. Theory and methods: W. H. Freeman and Co.
- Alliney S., and Ruzinsky, S. A., 1994, An algorithm for the minimization of mixed  $l_1$  and  $l_2$  norms with application to Bayesian estimation: IEEE Trans. on Signal Processing, **42**, 618-627.
- Amundsen, L., 1991, Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion: Geophysics, **56**, 2027-2035.
- Backus, G. and Gilbert, F., 1967. Numerical applications of a formalism for geophysical inverse problems: Geophys. J. R. Astron. Soc., **13**, 247-276.
- Beylkin, G., 1987, Discrete radon transform: IEEE Trans. Acoust., Speech, Signal Processing., **ASSP-35**, 162-172.
- Biondi, B. L., and Kostov, C., 1989, High-resolution velocity spectra using eigenstructure methods: Geophysics, **54**, 832-842.
- Bienvenu, G., and Kopp, L., 1983, Optimality of high resolution array processing using the eigensystem approach: IEEE, Trans. Acoust., Speech, Signal Processing., **ASSP-31**, 1235-1248.
- Born, M., and Wolf, E., 1980, Principles of optics: Pergamon Press.
- Burg, J. P., 1975, Maximum entropy spectral analysis: Ph.D. Thesis, Stanford University.
- Cabrera, S. D., and Parks, T. W., Extrapolation and Spectral Estimation with iterative weighted norm modification: IEEE Trans. Signal Processing, **ASSP-39**, 842-851.

- Capon, J., 1969, High resolution frequency-wavenumber spectrum analysis: *Proc. IEEE*, **57**, 1408-1418.
- Chapman, C. H., 1981, Generalized Radon transforms and slant stacks: *Geophys. J. Roy. Astr. Soc.*, **54**, 481-518.
- Claerbout, J. F., 1992, *Earth soundings analysis, Processing versus inversion*: Blackwell Scientific Publications, Inc.
- Constable, S. C., Parker, R. L., and Constable, C. G., 1987, Occam's inversion: A practical algorithm for generating smooth model from electromagnetic sounding data: *Geophysics*, **52**, 289-300.
- Cruse, E., Pica, A., Noble, M., McDonald, J., and Tarantola, A., 1990, Robust elastic nonlinear waveform inversion: Application to real data: *Geophysics*, **55**, 527-538.
- Deans, S. R., 1983, *The Radon transform and some of its applications*: J. Wiley and Sons, Inc.
- Durrani, T. S., and Bisset, D., 1984, *The Radon transform and its properties*: *Geophysics*, **49**, 1180-1187.
- Freire, S. L. M., and Ulrych, T. J., 1988, Application of singular value decomposition to vertical seismic profiling: *Geophysics*, **53**, 778-785.
- Friedman, J. H., and Tukey, J. W., 1974, A projection pursuit algorithm for exploratory data analysis: *IEEE Trans. Comput.*, **23**, 881-889.
- Foster, J. D., and Mosher, C. C., 1992, Suppression of multiples reflection using the Radon transform: *Geophysics*, **57**, 386-395.
- Gersztenkorn, A., Bednar, J. B., and Lines, L. R., 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: *Geophysics*, **51**, 357-368.
- Goldstein P. and Archuleta, R. J., 1987, Array analysis of seismic signals: *Geophysical*

Research Letters, **14**, 13-16.

Gulunay, N., 1990, F-X domain least-squares tau-p and tau-q: 60th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1607-1610.

Hampson, D., 1986, Inverse velocity stacking for multiple estimation: 56th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 422-424.

Hardage, B. A., 1985, Vertical Seismic Profiling: A. Principles, *in* K. Helbig and S. Treitel, Eds., Handbook of Geophysical Exploration, **14A**, : Geophysical Press, Amsterdam.

Hemon, C. H., and Mace, D., 1978, Essai d'une application de la transformation de Karhunen-Loève au traitement sismique: Geophys. Prosp., **26**, 600-626.

Huber, P. J., 1981, Robust Statistics: John Wiley and Sons, Inc.

Huber, P. J., 1985, Projection pursuit: Ann. Statist., **13**, 435-525.

Hugonnet, P., and Canadas, G., 1995, Aliasing in the parabolic transform: 65th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1366-1369.

Jaynes, E. T., 1968, Prior probabilities: IEEE Trans. Syst. Sci. Cybern. SSC-4, 227-241.

Jaynes, E. T., 1985, Where do we go from here?, *in* Ray Smith, C., and Grandy, W. T. Jr., Eds., Maximum-entropy and Bayesian methods in inverse problems, D. Reidel Publ. Co., 21-58.

Jeffreys, H., 1961 , Theory of Probability: Oxford Univ. Press (earlier editions 1939, 1948).

Johnson, N. L., and Kotz, S., 1970, Distributions in statistics: 1 and 2: John Wiley and Sons, Inc.

Jones, I. F., 1984, Application of the Karhunen-Loève transform in reflection seismic

- processing: Ph.D. thesis, University of British Columbia, Vancouver, Canada.
- Jones, M. C., and Sibson, R., 1987, What is projection pursuit: *J. R. Statist. Soc.*, **150**, 1-36.
- Kabir, M. M. N., and Verschuur, D. J., 1995, Restoration of missing offsets by parabolic Radon transform: *Geophysical Prospecting*, **43**, 347-368.
- Kay, S. M., and Marple, L. M., 1981, Spectrum analysis - A modern perspective: *Proc. IEEE*, **69**, 1380-1419.
- Key, S. C, and Smithson, S. B., 1990, New approach to seismic-reflection event detection and velocity determination: *Geophysics*, **55**, 1057-1069.
- Kirlin, R. L., 1992, The relationship between the semblance and the eigenstructure velocity estimators, *Geophysics*: **57**, 1027-1033.
- Kostov, C., 1990, Toeplitz structure in slant-stack inversion: 60th Annual Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1618-1621.
- Lim, J. S., and Malik, N. A., 1982, A new algorithm for two-dimensional maximum entropy power estimation: *IEEE, Trans. Acoust. Speech, Signal Processing.*, vol. **ASSP-30**, 788-797.
- Lines, L. R., and Treitel, S., 1984, Tutorial: A review of least-squares inversion and its applications to geophysical problems: *Geophys. Prosp.*, **32**, 159-186.
- Loredo, T. J., 1990, From Laplace to supernova SN 1987A: Bayesian inference in astrophysics, *in* Fougère, P., Ed., *Maximum entropy and Bayesian methods*: Kluwer Academic Publishers, 81-142.
- Marple, S. M., 1987, *Digital spectral analysis, with applications*: Prentice-Hall, Inc.
- Mari, J. L., and Glangaud, F., 1990, Spectral matrix filtering applied to VSP filtering: *Rev. Institut Franç. du Pétrole*, **45**, 417-434.

- Mari, J. L., and Gavin, P., 1990, Séparation des ondes P et S a l'aide de la matrice spectrale avec information a priori: *Rev. Institut Franç. du Pétrole*, **45**, 587-603.
- Neidell, N. S., and Taner, M. T., 1971, Semblance and other coherency measures for multichannel data: *Geophysics*, **36**, 482-497.
- Oldenburg, D. W., 1976, Calculation of Fourier transforms by the Backus-Gilbert method: *Geophys. J. Roy. Astr. Soc.*, **44**, 413-431.
- Oldenburg, D. W., Scheuer, T., and Levy, S., 1983, Recovery of the acoustic impedance from reflection seismograms: *Geophysics*, **48**, 1318-1337.
- Oppenheim, A. V., and Schaffer, R. W., 1975, *Digital signal processing*: Prentice Hall Inc.
- Parker, R. L., 1977, Understanding inverse theory: *Ann. Rev. Earth Planet. Sci*, **5**, p.35-64.
- Phinney, R. A., Chowdhury, K. R., and Frazer, L. N., 1981 Transformation and analysis of record sections: *J. Geophys. Res.*, **86**, 359-377.
- Pisarenko, V. P., 1973. The retrieval of harmonics from a covariance function: *Geophys. J. Roy. Astr. Soc.*, **33**, 347-366.
- Postic, A., Fourmann, J., and Claerbout, J., 1980, Parsimonious deconvolution: Presented at the 50th Ann. Mtg., Soc. Expl. Geophys.
- Pratt, W. K., 1991, *Digital image processing*: John Wiley and Sons, Inc.
- Press, W. H, Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P., 1992, *Numerical recipes, the art of scientific computing*: Cambridge Univ. Press (earlier edition 1986).
- Priestley, M. B., 1982, *Spectral Analysis and time Series*: Academic Press Inc.
- Rao, C. R., 1973. *Linear statistical inference and its applications*: John Wiley and

Sons, Inc.

Robinson, E. A., and Treitel, S., 1980, Geophysical signal processing: Prentice Hall Inc.

Rockafellar, R. T., 1970, Convex analysis: Princeton University Press.

Rutty, M. J., and Jackson, G. M., 1992, Wavefield decomposition using spectral matrix techniques: *Exploration Geophysics*, **23**, 293-298.

Sacchi, M. D., and Ulrych, T. J., 1994, High resolution velocity gathers and offset space reconstruction: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1469-1472.

Sacchi, M.D., and Ulrych, T.J., 1995a, Aperture compensated discrete Fourier transform and applications: 57th Meeting and Technical Exhibition of the European Association of Exploration geophysicists, Expanded Abstracts, P41.

Sacchi, M. D., and Ulrych, T. J., 1995b, High resolution velocity gathers and offset space reconstruction: *Geophysics*, **60**, 1169-1177.

Sacchi, M.D., and Ulrych, T.J, 1995c, Improving resolution of Radon operators using a model re-weighted least squares procedure : *Journal of Seismic Exploration*, **4**, 315-328.

Scales, J. A., Gersztenkorn, A. and Treitel, S., 1988, Fast  $l_p$  solution of large, sparse, linear systems: Application to seismic travel time tomography: *Journal of Comp. Phys.*, **75** , 314-333.

Slepian D., and Sonenblick E., 1965, Eigenvalues associated with prolate spheroidal wave functions of zero order: *Bell Syst. Tech. J.* , **44**, 1745-1758.

Slepian D., 1978, Prolate spheroidal wave functions, Fourier analysis, and uncertainty - V: The discrete case: *Bell Syst. Tech. J.*, **57**, 1371-1430.

Stoffa, P. L., Buhl, P., Diebold, J. B., and Wenzel, F., 1981, Direct mapping of seismic

- data to the domain of intercept and ray parameter - A plane-wave decomposition: *Geophysics*, **46**, 255-267.
- Strang, G., 1976, *Linear algebra and its applications*: Academic Press.
- Swingler, D. and Walker, R., 1989, Line-array beam-forming using linear prediction for aperture interpolation and extrapolation: *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-37**.
- Taner, M. T., Cook, E. E., and Neidell, N. S., 1970, Limitations of the reflection seismic method; Lessons from computer simulations: *Geophysics*, **35**, 551-573.
- Tatham, R. H., 1984, Multidimensional filtering of seismic data: *Proc. IEEE*, **72**, 1357-1369.
- Tarantola, A., 1987, *Inverse problem theory: Methods for data fitting and model parameter estimation*: Elsevier Science Publ. Co., Inc.
- Thorson, J. R., and Claerbout, J. F., 1985, Velocity-stack and slant-stack stochastic inversion: *Geophysics*, **50**, 2727-2741.
- Tikhonov, A. H. and Goncharsky, A. V., 1987, *Ill-posed problems in the natural sciences*: Mir Publ.
- Titterton, D. M., 1985, General structure of regularization procedures in image reconstruction: *Astron. Astrophys.*, **144**, 381-387.
- Treitel, S., Gutowski, P. R., and Wagner, D. E., 1982, Plane-wave decomposition of seismograms: *Geophysics*, **47**, 1375-1401.
- Turner, G., 1990, Aliasing in the tau-p transform and the removal of spatial aliased coherent noise: *Geophysics*, **55**, 1496-1503.
- Ulrych, T. J., and Walker, C. J., 1981, High resolution two-dimensional power spectral estimation, *in* Findley, D. F., Ed., *Applied Time Series II*: Academic Press Inc.

- Ulrych, T. J., Freire, S. and Sacchi, M. D, 1995, Eigenimage processing of seismic sections, in Kirilin L. and Done D, Eds., Application of covariance analysis to seismic signal processing: SEG publications. To appear.
- Yilmaz, O., 1989, Velocity-stack processing: *Geophys. Prosp.*, **37**, 357-382.
- Yilmaz, O., and Taner, M. T., 1994, Discrete plane-wave decomposition by least-mean-square-error method: *Geophysics*, **59**, 973-982.
- Van Trees, H., 1968, *Detection, estimation, and modulation theory*: John Wiley and Sons, Inc.
- Walden, A. T., 1985, Non-Gaussian reflectivity, entropy and deconvolution: *Geophysics*, **50**, 2862-2888.
- Walker, C., and Ulrych, T. J., 1983, Autoregressive recovery of acoustic impedance: *Geophysics*, **48**, 1338-1350.
- Wang, B. and Braile, L., 1994, Stochastic view of damping and smoothing, and applications to seismic tomography: 64th Ann. Internat. Mtg., Soc. Expl. Geophys., Expanded Abstracts, 1342-1346.
- Wax, M., Shan, T., and Kailath, T., 1984, Spatio-temporal spectral analysis by eigenstructure methods: *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-32**, 817-827.
- Wiggins, R. A., 1978, Minimum entropy deconvolution: *Geoexpl.*, **16**, p.21-35.
- Zhou, B., and Greenhalgh, S. A., 1994, Linear and parabolic  $\tau - p$  revisited: *Geophysics*, **59**, 1133-1149.

## Appendix A

### Scalar products

Let  $L$  be a linear operator mapping  $V$  into  $U$  and let  $(\cdot, \cdot)_V$  and  $(\cdot, \cdot)_U$  represent the scalar product in  $V$  and  $U$  respectively. The adjoint of  $L$ ,  $L^*$ , is a linear operator mapping  $V$  into  $U$ . Then for any  $v \in V$  and  $u \in U$

$$(L^*v, u)_U = (v, Lu)_V. \quad (A.1)$$

Defining the following scalar products

$$(L^*v, u)_U = v^H L^* W_u u, \quad (A.2)$$

$$(v, Lu)_V = v^H W_v Lu. \quad (A.3)$$

Using equations (A.1), (A.2), and (A.3) we have

$$\begin{aligned} (L^*v, u)_U &= (v, Lu)_V \\ &= v^H W_v Lu \\ &= v^H W_v L W_u^{-1} W_u u \\ &= (W_u^{-1} L^H W_v v, u)_U. \end{aligned} \quad (A.4)$$

From the above equality it follows that

$$\mathbf{L}^* = \mathbf{W}_u^{-1} \mathbf{L}^H \mathbf{W}_v, \quad (\text{A.5})$$

which is exactly equation (2.40).