Seismic applications of the Radon transform and a new second-order traveltime local transform

by

Ehsan Soleymani

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science
in
Geophysics

Department of Physics
University of Alberta

Abstract

Radon transforms play an essential role in seismic data processing. They have been extensively used to separate signals from coherent noise, reconstruct seismic data, and estimate the arrival direction of waves impinging on arrays. This thesis focuses on applying Radon transforms to seismic signal processing, including multiple suppression, denoising and seismic data regularization.

After reviewing classical Radon transforms that operate on the full aperture of a seismic gather, I propose a new local Radon transform parameterized by two adjustable parameters. The proposed Radon transform is derived using a second-order Taylor approximation of the reflection traveltime around a central seismogram. The proposed local Radon transform enables processing data composed of a superposition of complex waveforms with spatially variant dips.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Mauricio D. Sacchi, for his supervision, guidance and patience through each stage of the process of this thesis and for teaching me how to see the forest for the trees. I am honoured to be one of his students in the research group. I want to thank members of my committee for their time and review of my thesis. I take this opportunity to acknowledge my colleagues and friends in the Signal Analysis and Imaging Group (SAIG). Last but not least, I am grateful to my wife, family and friends for their support and encouragement during my studies.

Contents

Appendices

List of Figures

List of symbols and abbreviations

$t - x$	Time-space domain
$\tau - p$	Intercept-ray parameter domain
$f - x$	Frequency-space domain, f is frequency in Hz
$\omega - x$	Frequency-space domain, $\omega = 2\pi f$ in frequency in radians/s
p	Ray Parameter
q	Curvature of the parabolic Radon transform
v	Velocity
τ	Intercept Time
μ	Trade-off Parameter
A	Linear component of the two-term Radon approximation
B	Quadratic component of the two-term Radon approximation
$\mathbf{L}$	Matrix
$\mathbf{L}^H$	Hermitian Transpose of $\mathbf{L}$
$\mathcal{L}$	Linear operator
$\mathcal{L}^*$	Adjoint operator of $\mathcal{L}$
$\mathcal{F}$	Fourier operator
$\mathcal{F}^{-1}$	Inverse Fourier operator
$\mathbf{d}$	Vector indicating data
$d(t, x)$	Data as a function of time and spatial variable x
$m(\tau, p)$	Linear Radon Transform
$m(\tau, p)$	Parabolic Radon Transform
$m(\tau, v)$	Hyperbolic Radon Transform
2D	Two Dimensions
3D	Three Dimensions
AVO	Amplitude Variation with Offset
CMP	Common Midpoint

NMO	Normal Moveout
SNR	Signal-to-Noise Ratio
CGLS	Conjugate Gradient Least-Squares
FISTA	Fast Iterative Shrinkage-Thresholding Algorithm
IRLS	Iterative Reweighed Least-Squares

CHAPTER 1

Introduction

Geophysicists aim to quantitatively estimate subsurface geological structures and properties. Furthermore, they are also interested in knowing about processes that happen in the interior of the earth (Fowler, 1990). Subsurface studies are often carried out via indirect sensing methods. In other words, subsurface properties are estimated from data recorded on the earth's surface. Geophysicists adopt two families of methods for obtaining these properties. The first of these, the seismic method, uses traveltime and amplitude variations of waves propagating in the subsurface to estimate geological interfaces and elastic properties. Artificial explosive sources emit seismic waves into the subsurface. Subsurface interfaces reflect these waves; they propagate upwards and are finally recorded by arrays of receivers. The data (seismograms, also called seismic traces) can outline geological structures after applying signal processing and inversion techniques. In global seismology, naturally occurring sources (earthquakes) also produce waves that travel through the earth's interior. The recorded waveforms permit to map discontinuities at global scales and elucidate processes associated with them.

The second class of methods is entirely different in principle. They are often called potential field methods because they measure the distortion of a scalar or vector potential field caused by perturbations of a physical property in the subsurface. An example of the latter is the gravity prospection method that studies the distortion of the gravitational potential generated by variations in subsurface density distribution. Similarly, electric and magnetic methods measure the distortion of the electric and magnetic fields due to subsurface variations of resistivity and susceptibility (Telford

et al., 1990).

Seismic, magnetic, gravity, electrical, and electromagnetic methods are generally used to explore resources. Some of these techniques are mainly used for hydrocarbon exploration, and others are employed principally to prospect minerals. For example, the seismic method is often used for oil and gas exploration. Gravity and magnetic methods can also be used for oil and gas exploration during an early reconnaissance of basins. More recently, electromagnetic methods in conjunction with seismic methods have also been adopted for oil and gas exploration (Mehta et al., 2005). Despite efforts to combine and utilize several techniques cooperatively, oil and gas exploration is predominately dominated by seismic methods.

As I mentioned, the seismic method uses artificially generated waves to determine geological structures containing natural resources such as oil and gas. The seismic method provides data that allows geophysicists to estimate subsurface images combined with other geophysical and geological data. Civil engineering and hydrogeology are other areas in which the seismic method can play a significant role. For instance, in the cases mentioned above, the seismic method is often used to map the depth to bedrock for construction purposes like towers, dams, highways, or delineate aquifers.

The seismic method can be divided into two sub-categories. One can adopt refractions or head waves to determine the depth of layers and the propagation speed of waves in the subsurface or adopt reflections to define interfaces that reflect energy back to the subsurface (Yilmaz, 2001). The former is called the Seismic Refraction Method, and the latter is called the Seismic Reflection Method. This thesis will concentrate on one particular aspect of the Seismic Reflection Method. Concretely, it focuses on one processing tool called the Radon transform. The Radon transform is a mathematical tool that allows decomposing seismic waveforms into simple components. This decomposition is used to filter undesired signals such as multiple reflections or to perform trace interpolation.

The seismic reflection method can be divided into four important steps that can be summarized as follows:

1. **Data acquisition:** Different acquisition layouts are used in the seismic reflection method. These layouts are based on exploration purposes, target location,

cost, and operational environment. The goal of land and marine data acquisition is to collect data, that after processing, can image subsurface structures serving as a deposit of resources (Vermeer, 1990).

In a land operation, a line of fixed receiver stations records data after a source is activated. The process is repeated by changing the position of the source, giving rise to a data volume of seismic traces. Each seismic trace corresponds to one source-receiver pair. Figure 1.1a shows a shot gather corresponding to a 2D acquisition where receivers and source are deployed along a line. Figure 1.1b represents the response for one layer. The theoretical traveltime curve for such an experiment is a hyperbola. Multiple shots are then sorted in common mid-point (CMP) gathers to produce the acquisition configuration shown in Figure 1.2a and b. The CMP configuration is obtained via sorting data acquired as shot gathers. The CMP configuration is needed to carry out processes such as velocity analysis and normal moveout correction (Yilmaz, 2001).

Because the geological structures are three-dimensional, a 2D data acquisition layout cannot accurately image these structures. In this case, a 2D seismic survey will record data that will miss-position geological structures. Therefore, modern 3D seismic exploration uses an areal pattern of receivers and sources (in other words, receivers and sources are not deployed along a single line). A typical 3D land acquisition involves using multiple lines of receives and numerous lines of sources perpendicular to each other. The latter is often called an orthogonal survey (Vermeer, 2002). In marine data acquisition, a vessel tows one or more energy sources and hydrophone streamers. Another option is to use ships that tow receivers and sources producing narrow and wide-azimuth 3D seismic survey configurations.

2. **Data processing, also called data preconditioning:** The elevation of source and receiver are corrected to a common datum. During this stage, we also apply incoherent and coherent noise removal techniques and source equalization methods (deconvolution) (Yilmaz, 2001). The main idea is to process the data so that each common-source gather approximates the acoustic or elastic Green function corresponding to an experiment where a wave produced by a point source propagates in the earth. Included in the preconditioning stage are methods for signal-to-noise ratio enhancement and data reconstruction.

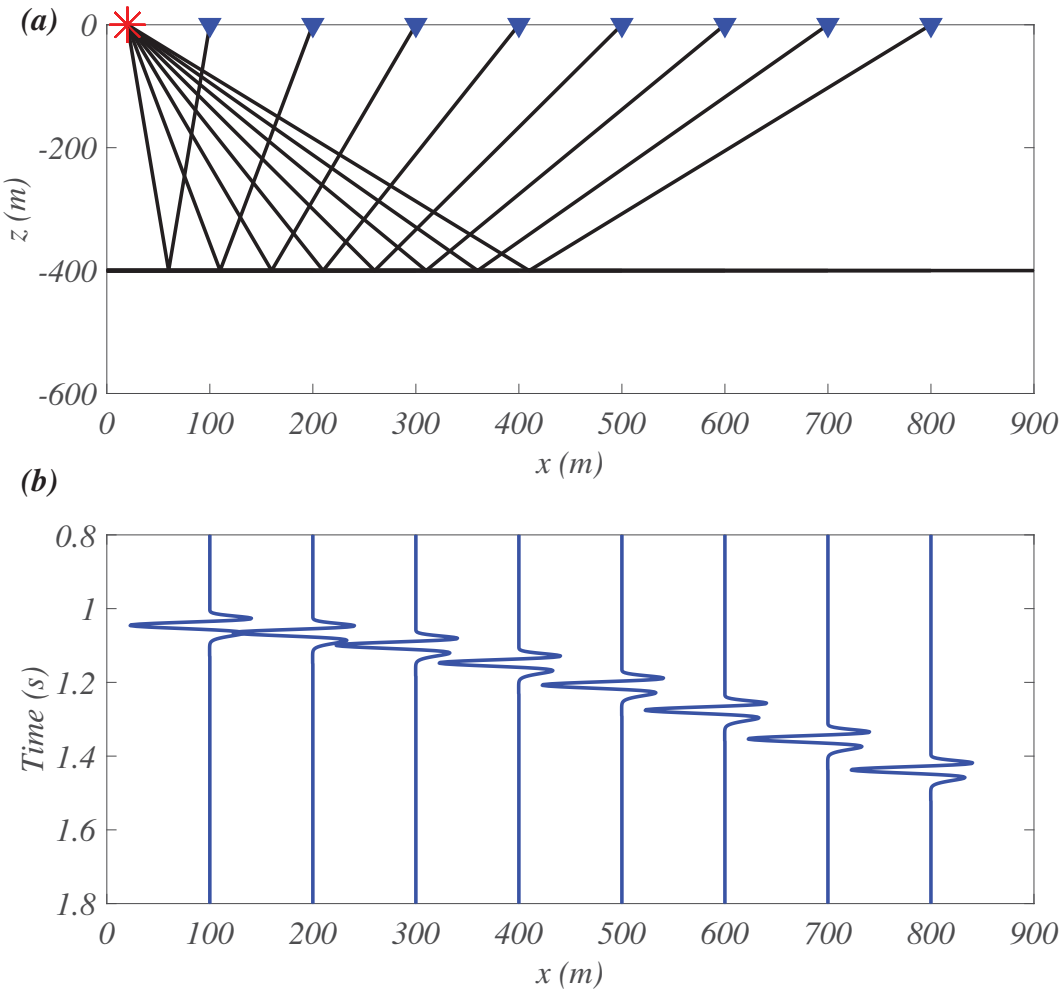

Figure 1.1: a) A shot gather. b) Recorded data via a shot gather. The red star is the source and the triangles represent receivers. The variable x represents the position of the receiver.

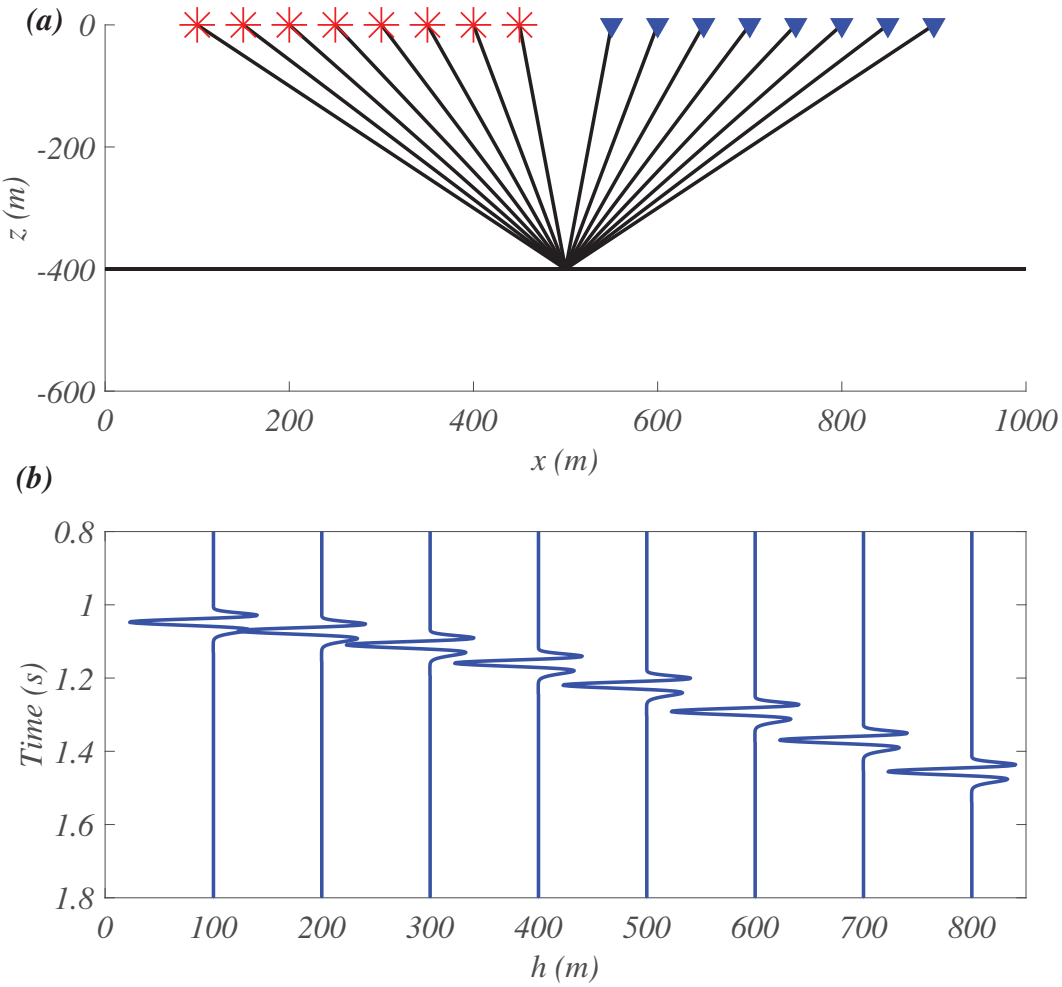

Figure 1.2: a) A common midpoint (CMP) gather. b) Recorded data for CMP gather. Red stars are sources and triangles are receivers.The variable h represents source-receiver distance for each source receiver pair that composes the CMP gather.

This thesis concentrates on the application of Radon transforms for coherent and incoherent noise reduction and reconstruction. The Radon transform is an indispensable part of the data processing stage of the reflection seismology method (Hampson, 1986a,b; Trad et al., 2003), and it can also be used for data interpolation (Kabir and Verschuur, 1995).

3. **Imaging and inversion:** Imaging and inversion are similar processes. In general, imaging is used to refer to methods to determine the boundaries of subsurface structures. Inversion refers to approaches to estimate subsurface parameters such as elastic moduli or propagation speed of compressional and shear waves. Both imaging and inversion are based on the solution of classical math-physics problems associated with wave propagation phenomena (Claerbout, 1985). Historically, imaging techniques are referred to as migration algorithms. Early migration algorithms were based on the Huygens principle and on Kirchhoff's integral theorem (Gray et al., 2001). Modern imaging and inversion algorithms operate with one-way and two-way acoustic or elastic wave equations, which are often linearized via the Born approximation (Sava and Hill, 2009).

4. **Seismic and geological interpretation:** At this stage, subsurface images are interpreted with the help of regional geological information, core samples, and formation properties estimated by well logging (Brown, 2011). In other words, images of the subsurface provide boundaries and, after inversion, one can also estimate material (rock) properties. However, these images do not tell us how sediments have deposited during the geological time or how structures (faults and folds) have emerged; neither they tell us how fluids in the crust have migrated and accumulated in reservoirs. Geologists and geophysicists use seismic images in conjunction with geological information to discover areas with an accumulation of hydrocarbons.

1.1 General considerations about transform-based processing

In signal processing, we often transform acquired data to a new domain where signal separation or filtering processes become easy tasks. Consider a transformation that

maps a generic signal d into a new signal called m. The transform can be a Fourier transform, a wavelet transform, a Radon transform, etc. Many choices of transforms exist, and we merely assume one has access to the transformation, which we can write in operator form as follows $\mathcal{T}$. Similarly, let us assume we also have access to the transform named $\mathcal{T}^{-1}$ that allows returning to the original space. Let us also assume that the transformation $\mathcal{T}$ is linear. In other words, if the original signal is composed of, for instance, three signals $d = d_1 + d_2 + d_3$, then the data after being transformed is given by $m = m_1 + m_2 + m_3$. Where d_i maps to m_i, $i = 1, 2, 3$. Symbolically, one can summarize the process of transforming the data d into m as

$$m = \mathcal{T} d \,. \tag{1.1}$$

The main goal of transforming the data d into m is to find a simple signal representation. This representation permits, for instance, to isolate the elements of m in the transform domain. This can be understood through a simple cartoon given in Figure 1.3. In this simple example, we assume a signal composed of three elements or waveforms d_1, d_2, d_3. As an example, let us assume that one would like to isolate d_3. The latter is a difficult task because d_3 overlaps with d_1 and d_2. The data mapped to the new domain is now composed of signals m_1, m_2, m_3 which do not overlap. Therefore, it is a simple task to filter (mute) m_1 and m_2 without affecting m_3. After the process of filtering, one can use the transform $\mathcal{T}^{-1}$ to map back m_3 into d_3. The idealized cartoon shows the process that is generally utilized to separate waveforms via the Radon transform. It is important to mention that the isolation of events in the transformed domain also entails some level of compressibility of the signals d_i when they are transformed to m_i. This concept will become clear when applying it with examples associated with the Radon transform in conjunction with sparse inversion techniques.

1.2 Radon transforms

Transformations used in seismic data processing such as Fourier, Wavelet, and Radon are mathematical linear mappings representing data in a new domain. In the new domain, the data properties can be employed for signal-to-noise ratio enhancement and waveform separation. Among the various transforms used in seismic data processing and imaging, the Radon transform has particular significance.

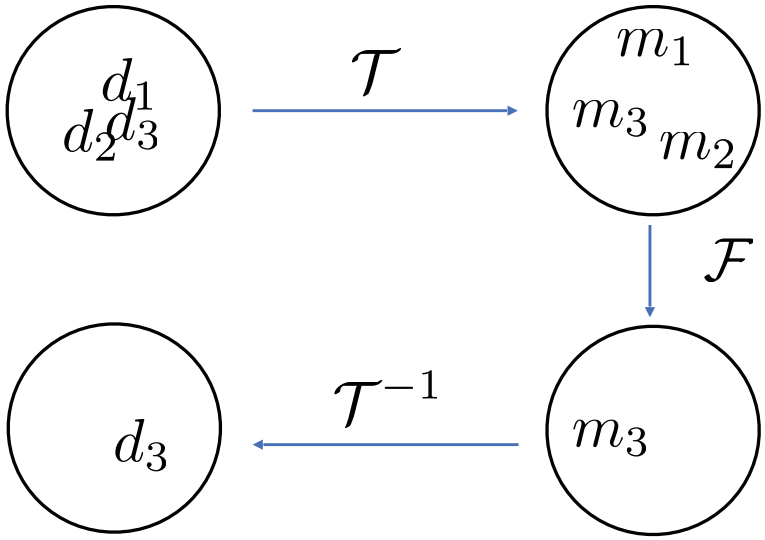

Figure 1.3: a) Pictorial representation showing the process of filtering after transforming data to a new domain. Data d is composed of 3 elements or waveforms d_1, d_2, d_3 which after applying the transform $\mathcal{T}$, they map to m_1, m_2, m_3. Data in the new domain are isolated and components m_1, m_2, m_3 are separated. Therefore, filtering ($\mathcal{F}$) becomes an easy task. In this example, in the transformed domain it is possible to eliminate m_1 and m_2 and then return to data space via the transform $\mathcal{T}^{-1}$ to estimate d_3.

The Radon transform owns its name to Johan Radon, who introduced it in 1917. The Radon transform integrates a physical attribute along a specific trajectory (Deans, 1983; Durrani and Bisset, 1984) and has numerous applications in seismology and medical imaging.

In exploration seismology, the Radon transform has been used in different forms. For instance, the linear Radon transform also called the Slant Stack transform, has been used to synthesize plane waves (Treitel et al., 1982). It can be shown that a point-source seismic shot gather can be decomposed into a set of plane-wave seismograms for arbitrary angles of incidence via the Slant Stack transform. Such plane-wave seismograms are an attractive domain for processes such as one-dimensional inversion and noise attenuation via predictive deconvolution (Corning

et al., 1987).

In its simplest form, the Radon transform involves mapping data from the offset (x) and time (t) to intercept time (τ) and ray parameter (p). In this case, the Slant Stack or Linear Radon transform is defined as an integration of the seismic data $d(x,t)$ along lines

$$u(\tau,p)=\sum_{x} d(t=\tau+px,x) \tag{1.2}$$

where x indicates offset or source-receiver distance. For illustrative purposes, I consider the synthetic data in Figure 1.4a composed of two events of constant dip. The data could correspond to a small spatiotemporal window of a seismic two-dimensional record. The corresponding representation in τ,p is illustrated in Figure 1.4b. As it can be seen by this figure, the transformed data consists of two focused events in τ,p. This property will be used to separate events and filter them. In Chapters 2 and 3, I will study methods to guarantee that one can map back the transformed data $u(\tau,p)$ to data space. I will also investigate the estimation of $u(\tau,p)$ via an inversion process that focuses the seismic events in the transformed domain. The latter facilitates the reconstruction of the data from the transformed signal and attenuation of noise in the reconstructed data.

1.2.1 Previous contributions

The Radon transform and its numerical solution has received significant attention in exploration seismology. Initial efforts posed the Radon transform computation as an inverse problem in the $\omega-x$ domain. The latter is the classical approach introduced by Hampson (1986b), and Beylkin (1987). It is important to mention that Beylkin (1987) studies a linear Radon transform with integration path given by $t=\tau+px$ while Hampson (1986b) investigates the Parabolic Radon transform. In other words, a Radon transform with a parabolic integration path $t=\tau+px^2$. The parabolic approximation is used to model the traveltime of reflections after normal moveout correction (Yilmaz, 2001). Hence, the parabolic Radon transform is often used to discriminate and filter multiples in common-min-point gathers. The assumption is that after NMO correction, primaries are flat events and multiples exhibit parabolic residual moveout. This property is exploited by the parabolic

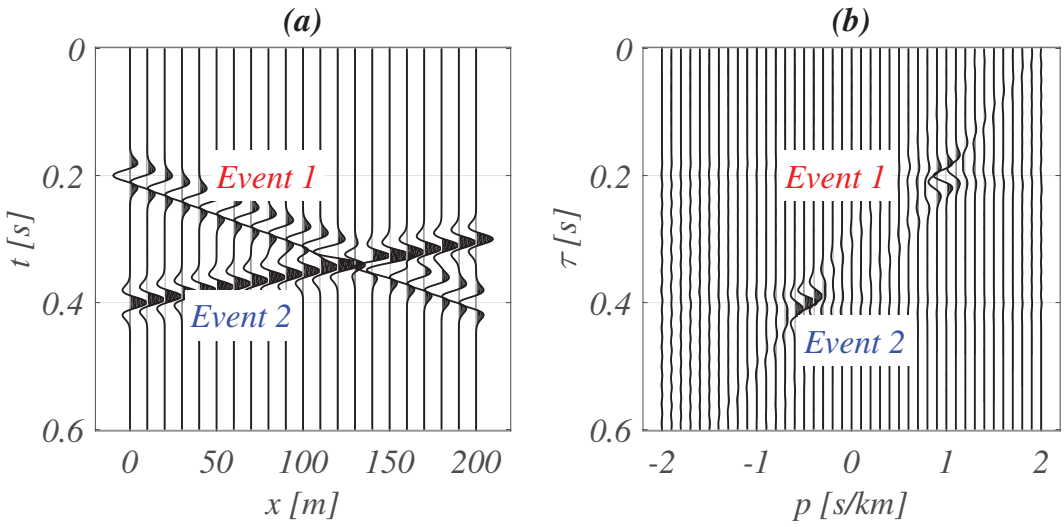

Figure 1.4: a) Two waveforms of ray parameter $p = -0.5$ s/km and $p = 1.0$ s/km. b) Slant Stack transform. The first event maps to $\tau = 0.2$s, $p = 1.0$s/km. The second event maps to $\tau = 0.4$s, $p = -0.5$s/km. Artifacts are caused by the finite aperture of the data and sampling of both the offset h and ray-parameter p axes.

Radon transform to separate multiples from primaries in the Radon domain. For instance, one can mute the contribution of primaries in the Radon domain and then return to data space with a model of the multiple reflections, which can be subtracted from the original data. By this means, it is possible to obtain a CMP gather that only contains primary reflections. This type of filter has been investigated by many authors, including Kabir and Verschuur (1995) who adopted the parabolic Radon transform to also reconstruct near offset traces, and Darche (1990) who applied the parabolic Radon transform for trace interpolation. Similarly, Kostov (1990) proposed a fast algorithm to compute linear and parabolic Radon transforms that exploits the Toeplitz structure of the Radon least-squares solution in the $\omega - x$ domain. The above contributions correspond to least-squares solutions of the Radon transform in the $\omega - x$ domain. They all use explicit form operators. In other words, they are based upon transforming data from $t-x$ domain to $\omega-x$ domain and setting the Radon transform problem as one that entails matrix-times-vector multiplications with explicit form matrices. We often call this type of Radon transform, frequency domain Radon transform or $\omega - x$ domain Radon transforms to differentiate them from Radon transforms that operate directly in the $t - x$ space with implicit form linear operators.

Yilmaz (1989) proposed to adopt the parabolic Radon transform also for multiple suppression. However, rather than adopting the NMO correction to validate the parabolic traveltime assumption, he offered to use a t^2 transformation of the time axis of the CMP gather to map hyperbolic traveltimes to their parabolic approximation. This technique is also adopted by Sacchi and Ulrych (1995) to obtain high-resolution parabolic velocity gathers. In Sacchi and Ulrych (1995), a new regularization technique for the solution of the Radon transform problem is also proposed. Rather than adopting the classical damped least-squares solution (Hampson, 1986b; Beylkin, 1987; Zhou and Greenhalgh, 1994), these authors adopted a Cauchy regularization term that provokes sparsity in the solution of the Radon transform coefficients. The approach has been improved by many authors, including Herrmann et al. (2000) who developed an anti-alias Radon transform, Schonewille and Duijndam (2001) who investigate computational solvers and sampling considerations for optimal application of the Radon transform and Sacchi and Porsani (2005) who combine a fast solver for Toeplitz forms with the conjugate gradient method to obtain a high-resolution Radon transform where matrix-times-vector multiplications

are efficiently executed via a Fast Fourier Transform.

The Radon transform can also be computed in the $t-x$ domain via implicit form operators (Thorson and Claerbout, 1985). That is, in different words, by adopting linear operator that behave like a matrix-times-vector multiplication but where the matrix is never physically built. Thorson and Claerbout (1985) employed the Radon transform with a hyperbolic integration path to compute velocity stacks which can be used for velocity analysis, CMP denoising, and data regularization. The approach has also been studied by Trad et al. (2003) that provides several iterative algorithms for efficient computation of time-domain Radon transforms and mixed domain Radon transforms.

In general, frequency domain solutions are faster than time-domain solutions. However, for the computation of the high-resolution Radon transform, several authors have concluded solutions are more stable for the time domain solvers rather than for frequency domain Radon transforms (Cary, 1998; Trad et al., 2003).

Often Radon transforms are applied with time-invariant integration paths that can be linear or parabolic or with time-variant integration paths that can be parameterized with hyperbolas or apex-shifted hyperbolas (Ibrahim and Sacchi, 2014). Time invariant paths lead to Radon transforms computed in the $\omega-x$ domain or in the $t-x$ domain. In general, as I have mentioned, frequency-domain algorithms are more efficient than time-domain algorithms. Conversely, time-variant paths such as those used for the hyperbolic Radon transform and the apex-shifted hyperbolic Radon transform lead to an algorithm that operates in the $t-x$ domain. In other words, time-variant integration paths do not lead to simple algorithms in $\omega-x$ where the Radon operators are estimated via simple matrix manipulations. In the case of time-invariant integration paths (linear and parabolic Radon transforms), we can also use mixed domain algorithms where the optimization problem associated with the computation of the Radon is done in the time domain. Still, the forward and adjoint operators are speed-up by transforming the data to $\omega-x$ and $\omega-p$. This mixed domain approach is prevalent today as it permits to incorporate sparsity constraints directly on $\tau-p$ rather domain than in $\omega-p$ (Cary, 1998).

It is essential to mention that a more flexible Radon transform can be obtained by defining a two-term second-order approximation to the traveltime in small spatio-temporal windows. The latter permits the integration path to adapt to the traveltime of the waveform locally, and it is not restricted to applications of the Radon

transform that operate in specific domains (e.g., CMP gathers) on the full aperture of the gather. This idea of the locality of the Radon transform in combination with a second-order approximation of the traveltime integration path is the main contribution of this thesis. Basically, the traditional linear and parabolic Radon transforms are adapted to operate with integration paths that are second-order accurate in small windows. This local approach to the computation of the Radon transform will permit us to process data that does not conform to typical hyperbolic traveltime signatures or parabolic traveltimes after NMO correction. This flexibility is gained by producing a significantly more expensive problem where the Radon parameters now depend on τ plus two kinematic attributes that we named A and B rather than one (p). In essence we aim to map data from $t-x$ to a domain described by $\tau - A - B$, where A and B are two kinematic parameters describing a local approximation to the traveltime of the seismic reflection. For this purpose, in this this, I first need to review classical tools and methods used for computing both $\omega - x$ ad $t - x$ Radon transforms and then introduce novel aspects of this thesis that are associated to the proposed new Radon transform.

1.3 Organization of the thesis

The structure of this thesis is as follows:

- Chapter 1 offers a succinct description of previous contributions, contributions, and overall basic background about Radon transforms.
- Chapter 2 describes frequency domain solvers for the linear and parabolic Radon transforms. These solvers are posed in the $f - x$ domain and use explicit matrix-times-vector multiplication and the inversion of matrices to compute the Radon transform in the $\omega - p$ domain. Moreover, we review the classical approach proposed by Hampson (1986b) and the high-resolution (sparse) frequency domain Radon transform proposed by Sacchi and Ulrych (1995).
- Chapter 3 presents solvers that compute the Radon transform in the time domain. These solvers are the Conjugate Gradient method, Iterative Reweighted Least-squares method (IRLS), and the recently proposed Fast Iterative Shrinkage-Thresholding Algorithm (FISTA). Both IRLS and FISTA permit to estimate

high-resolution Radon transforms by imposing sparsity constraints in the $\tau - p$ domain. Also, this chapter explains the adoption of implicit form Radon operators. In other words, the forward and adjoint Radon operators are treated as linear operators in implicit form (matrix-free form), and no effort is made to construct these operators explicitly as matrices in the $t - x$ and $\tau - p$ domains.

- Chapter 4 introduces a new Radon transform. This Radon transform is local and adopts an integration path that is a second-order expansion of the traveltime curve of the seismic reflection. The main goal is to develop a flexible tool for sparsity promoting denoising and reconstruction of seismic data with waveforms that can not be represented by hyperbolas or parabolas. Given that the parametrization is local (as opposed to one that spans the full aperture of a seismic gather), the proposed Radon operator can cope with spatially variant dips. The local Radon coefficients found by this method can synthesize denoised data and interpolate seismic data.

- Chapter 5 In the final chapter, conclusions, and recommendations for future work are offered.

1.4 Contribution of this thesis to seismic data processing

Traditionally, Radon transforms use integration paths given by $t = \tau + px$ (linear Radon), $t = \tau + px^2$ (parabolic Radon) or $t = \sqrt{(}\tau^2 + px^2$ (hyperbolic Radon). I have proposed to use a new integration path that locally adapts to any waveform. The path is a second order approximation to reflection travel-times in any domain and is given by $t = \tau + A(x - x_0) + B(x - x_0)^2$ where x_0 is the central trace to which we attribute the kinematic attributes A, B. This leads to a new Radon transform where the the signal is mapped from $t - x$ space to amplitudes that depend on τ, A, B, x_0. As one can imagine, this is equivalent to solving for a multi-parameter Radon transform. Not fully explore in this thesis, the propose local multi-parameter Radon transform has a significant resemblance to the non-linear beam-forming method proposed by Bakulin et al. (2020) for signal-to-noise ratio enhancement. However, while Bakulin et al. (2020) compute enhanced traces by summing along second-order travel times derived from parameters computed via

semblance analysis, the proposed method constructs the denoised traces by the solution of an inverse problem that entails finding the sparse solution of a multi-parameter local Radon transform. Some advantages of the proposed method are

- adaptability to spatially variant dips,
- applicability of the method in any domain (not restricted to CMP gathers such as the classical parabolic or hyperbolic Radon transforms), and
- the local behavior of the transform makes it more adaptable to cope with amplitude-versus-offset variations and could be an improvement to previously reported AVO-preserving Radon operators (Wang et al., 2011; Gholami and Farshad, 2019; Kazemi and Sacchi, 2021).

CHAPTER 2

Frequency domain Radon transforms

2.1 Introduction

In this chapter, I describe the classical algorithm used to estimate the Radon transform as initially proposed by Hampson (1986a). The algorithm operates in the frequency-space ($\omega - x$) domain. It is an explicit form solution to the problem of obtaining the Radon transform. An explicit form solver is one where the Radon transform is expressed as a matrix-times-vector multiplication. The matrix is also created and stored in memory. In the next chapter, I will describe implicit form Radon operators. The latter are Radon transforms implemented via matrix-free operators and iterative methods for linear systems of equations that replace explicit matrix inversion techniques.

2.2 Linear Radon transform

We first define the Radon transform allowing mapping data from the $t - x$ (time-space) domain to the $\tau - p$ domain where τ is intercept-time, and p is dip or ray parameter. We denote the seismic gather or a window of it as $d(t, x)$ where t indicates time and x indicates the spatial coordinate of the seismogram. For instance, x could refer to the receiver-source distance (offset) of the seismic trace or any other spatial variable associated with our data.

We first write the general expression of the Linear Radon transform as a simple mapping from data in $t-x$ to $\tau-p$

$$\hat{m}(\tau,p) = \mathcal{L}^*[d(t,x)] = \int_X d(\tau+px,x)\,dx\,, \tag{2.1}$$

where X indicates the range of integration of the spatial variable x. Similarly, we can define a forward Radon transform via the following expression

$$d(t,x) = \mathcal{L}[m(\tau,p)] = \int_P m(t-px,p)\,dp\,. \tag{2.2}$$

The symbol P denotes the integration range of the ray-parameter p. The operator $\mathcal{L}^*$ is called the Radon adjoint operator. Similarly, the operator $\mathcal{L}$ will be called the forward Radon operator. To simplify the notation, the short operator form expressions $\hat{m} = \mathcal{L}^* d$ and $d = \mathcal{L}m$ are adopted for equations 2.1 and 2.2, respectively.

Unlike the Fourier transform, the Radon transform is not an orthogonal transform. In other words, one can substitute 2.2 into 2.1 and verify

$$\hat{m} = \mathcal{L}^*\mathcal{L}m \tag{2.3}$$

where $\mathcal{L}^*\mathcal{L} \neq \mathcal{I}$ (identity). The latter is the reason, one must used the symbol $\hat{m}(\tau,p)$ rather than $m(\tau,p)$ in equation 2.1. One can interpret the adjoint Radon operator $\mathcal{L}^*$ as an analysis operator that maps the signal from $t-x$ to $\tau-p$. Similarly, the operator $\mathcal{L}$ is interpreted as forward modelling operator capable of synthesizing data in the $t-x$ domain from its representation in the $\tau-p$ domain. In general, $\hat{m}(\tau,p)$ resembles $m(\tau,p)$ but clearly, one needs to compute $m(\tau,p)$ in order to synthesize the data via equation 2.2. Therefore, we are in the presence of an inverse problem where given the seismic observations $d(t,x)$ one need to estimate $m(\tau,p)$. The inverse problem is linear and expressible via a simple matrix-times-vector operation in the frequency $(\omega-x)$ space.

2.3 Obtaining $m(\tau,p)$ via inversion in the $\omega-x$ domain

It is clear that we will be using discrete data, and therefore, equation 2.1 will be written as a sum

$$\hat{m}(\tau, p_k) = \sum_{n=1}^{n_x} d(t = \tau + p_k\, x_n, x_n)\,, \quad k = 1 \ldots n_p \tag{2.4}$$

where we understand that x and p are all discrete variables. For instance x can be replaced by x_n, $n = 1 \ldots n_x$. Similarly, the ray parameter p can also be discretized via the expression $p_k = p_{min} + (k-1)\Delta p$, $k = 1, \ldots n_p$, such that $p_{n_p} = p_{max}$ and $p_1 = p_{min}$. The increment Δp can be defined following criteria provided by Turner (1990). Similarly, n_p defines the number of τ, p traces that one would like to estimate. The parameters p_{min} and p_{max} are the minimum and maximum dips in the data that must be provided by the user. These parameters must contain all possible dips in the data.

Following Hampson (1986b), it is preferable to rewrite the estimation of the Radon coefficients via the solution of an inverse problem. We reformulate the problem where the data $d(t, x)$ is synthesized from the Radon $\tau - p$ domain. To this end, one can adopt the following formula to synthesize the data

$$d(t, x_n) = \sum_{k=1}^{n_p} m(\tau = t - p_k\, x_n, p_k)\,, \quad n = 1 \ldots n_x\,. \tag{2.5}$$

We often call equation 2.4 the adjoint operator or analysis operator. In contrast, equation 2.5 is called the forward or synthesis operator. Equation 2.5 can be simplified by expressing it in the frequency domain. In other words, we first apply the Fourier transform to both sides of equation 2.5 and, then, we use the Fourier delay theorem (Bracewell, 1978) to obtain the following expression

$$D(\omega, x_n) = \sum_{k=1}^{n_p} M(\omega, p_k)\, e^{-i\omega p_k x_n}\,, \quad n = 1 \ldots n_x\,. \tag{2.6}$$

The variable ω is the temporal frequency in radians/sec. In the last expression $D(\omega, x_n)$ indicates the signal $d(t, x_n)$ transformed to the frequency domain. In other words, $d(t, x_n) \leftrightarrow D(\omega, x_n)$. Similarly, $m(\tau, p_j) \leftrightarrow M(\omega, p_j)$. Equation 2.6 represents a linear system of equation given by

$$\begin{pmatrix} D(\omega, x_1) \\ D(\omega, x_2) \\ D(\omega, x_3) \\ \vdots \\ D(\omega, x_{n_x}) \end{pmatrix} = \begin{pmatrix} e^{-i\omega x_1 p_1} & e^{-i\omega x_1 p_2} & \dots & e^{-i\omega x_1 p_{n_p}} \\ e^{-i\omega x_2 p_1} & e^{-i\omega x_2 p_2} & \dots & e^{-i\omega x_2 p_{n_p}} \\ \vdots & \vdots & \dots & \vdots \\ e^{-i\omega x_{n_x} p_1} & e^{-i\omega x_{n_x} p_2} & \dots & e^{-i\omega x_{n_x} p_{n_p}} \end{pmatrix} \cdot \begin{pmatrix} M(\omega, p_1) \\ M(\omega, p_2) \\ M(\omega, p_3) \\ \vdots \\ M(\omega, p_{n_p}) \end{pmatrix} \quad (2.7)$$

The system of equations 2.7 can be represented in compact form as follows

$$\mathbf{D}(\omega) = \mathbf{L}(\omega)\,\mathbf{M}(\omega)\,, \quad (2.8)$$

where $\mathbf{D}(\omega)$ is the vector of size $n_x \times 1$ with the spatial data amplitudes for frequency ω. Similarly, $\mathbf{M}(\omega)$ is the vector of size $n_p \times 1$ with the amplitudes of the Radon transform for each ray parameter p_k at frequency ω. Clearly, the elements of the $n_x \times np$ matrix $\mathbf{L}(\omega)$ are

$$[\mathbf{L}(\omega)]_{n,k} = e^{-i\omega x_n\, p_k}\,, \quad n = 1 \dots n_x\,,\ k = 1 \dots n_p\,. \quad (2.9)$$

One can also show that the adjoint operator given by equation 2.4 can be written in the $\omega - x$ domain as follows

$$\hat{\mathbf{M}}(\omega) = \mathbf{L}^H(\omega)\,\mathbf{D}(\omega) \quad (2.10)$$

where $\mathbf{L}^H(\omega)$ is used to indicate the Hermitian transpose of the matrix $\mathbf{L}(\omega)$. Again, we can insert equation 2.8 into 2.10 to obtain

$$\hat{\mathbf{M}}(\omega) = \mathbf{L}^H(\omega)\mathbf{L}(\omega)\mathbf{M}(\omega)\,. \quad (2.11)$$

The operator $\mathbf{L}^H(\omega)\mathbf{L}(\omega)$ is not an identity, showing that inversion is needed to estimate $\mathbf{M}(\omega)$.

2.3.1 Radon transform solution via the damped least-squares method

So far two operators were defined. These operators entail mapping data from $\omega - p$ to $\omega - x$ (Forward) and from $\omega - x$ to $\omega - p$ (adjoint):

$$\text{Forward Transform} \quad \mathbf{D}(\omega) = \mathbf{L}(\omega)\mathbf{M}(\omega) \tag{2.12}$$

$$\text{Adjoint Transform} \quad \hat{\mathbf{M}}(\omega) = \mathbf{L}(\omega)^H \mathbf{D}(\omega)\,. \tag{2.13}$$

We further assume data contaminated with noise and therefore, we will write the following expression

$$\mathbf{D}(\omega) = \mathbf{L}(\omega)\,\mathbf{M}(\omega) + \mathbf{E}(\omega) \tag{2.14}$$

where $\mathbf{E}(\omega)$ is the $n_x \times 1$ vector representing noise or errors. Our task is to compute $\mathbf{M}(\omega)$ from noisy observations represented via equation 2.14. For this purpose we define a cost function and minimize the error with respect to the unknown vector $\mathbf{M}(\omega)$. We define our cost function to minimize using the least-squares criterion plus a regularization term

$$J(\omega) = \|\mathbf{E}(\omega)\|_2^2 + \mu\|\mathbf{M}(\omega)\|_2^2 \tag{2.15}$$

$$= \|\mathbf{L}(\omega)\mathbf{M}(\omega) - \mathbf{D}(\omega)\|_2^2 + \mu\|\mathbf{M}(\omega)\|_2^2 \tag{2.16}$$

where the first term is the squared l_2 norm of the error and the second term is the squared l_2-norm of the model parameters. The squared l_2-norm of the vector, for instance, $\mathbf{M}(\omega)$ is given by

$$\|\mathbf{M}(\omega)\|_2^2 = \sum_{k=1}^{n_p} |\, M(\omega, p_k)\,|^2\,, \tag{2.17}$$

where $|\cdot|$ is the absolute value of a complex number.

The scalar $\mu > 0$ is the tradeoff parameter that controls the relative balance between the error and model norm. The solution can be obtained by minimizing J with respect to the unknown parameters $\mathbf{M}(\omega)$. In this case, we use the condition for the minimum of the cost function $J(\omega)$ which is given by

$$\frac{\partial J(\omega)}{\partial \mathbf{M}(\omega)} = \mathbf{0}\,. \tag{2.18}$$

The last equation leads to the so-called damped least-squares solution (DLS) (Menke,

1989)

$$\mathbf{M}_s(\omega) = [\,\mathbf{L}^H(\omega)\mathbf{L}(\omega) + \mu\mathbf{I}\,]^{-1}\mathbf{L}^H(\omega)\mathbf{D}(\omega) \tag{2.19}$$

$$= \mathbf{L}^H(\omega)[\,\mathbf{L}(\omega)\mathbf{L}^H(\omega) + \mu\mathbf{I}\,]^{-1}\mathbf{D}(\omega)\,. \tag{2.20}$$

It is easy to show that equations 2.19 and 2.20 are equivalent expressions. In the case where $n_p > n_x$ one should adopt equation 2.20 because it involves the inversion of matrix of size $n_x \times n_x$. Similarly, when $n_p < n_x$ it is better to adopt 2.19 because it requires the inversion of a matrix of size $n_p \times n_p$.

The cost function $J(\omega)$ has two terms. The first term is the misfit, which is the data fidelity term. The second term is the model norm $\|\mathbf{M}(\omega)\|_2^2$ also called the regularization term. Determining the tradeoff parameter μ is very important. Different values of the parameter μ lead to solutions with different degrees of fitting and stability. Practically, speaking μ is chosen by monitoring the predicted data, and it is adjusted to avoid overfitting or underfitting the data. Given the solution computed via the damped least-squares method, $\mathbf{M}_s(\omega)$, we can use the latter to predict the data via expression:

$$\mathbf{D}_p(\omega) = \mathbf{L}(\omega)\mathbf{M}_s(\omega)\,. \tag{2.21}$$

The computation of the Radon transform using expressions 2.19 or 2.20 and 2.21 require the solution of one matrix inversion per frequency ω. In general, the tradeoff parameter μ should be different for each frequency. However, for practical purposes, one can choose the same parameter μ for all the frequencies representing the signal in the $\omega - x$ space.

Figure 2.1 illustrates a synthetic example. In Figure 2.1a we present two linear events. In 2.1b the two liner events are represented in $\tau - p$. The latter was obtained via the damped least-squares solution with trade-off parameter $\mu = 0.1$. The panel is $\tau - p$ was then used to obtain the predicted data in Figure 2.1c

A pseudo-code for the $\omega - x$ domain Radon transform algorithm is presented in Algorithm 1. Similarly, Algorithm 2 shows the pseudo-code for the forward Radon $\omega - x$ transform.

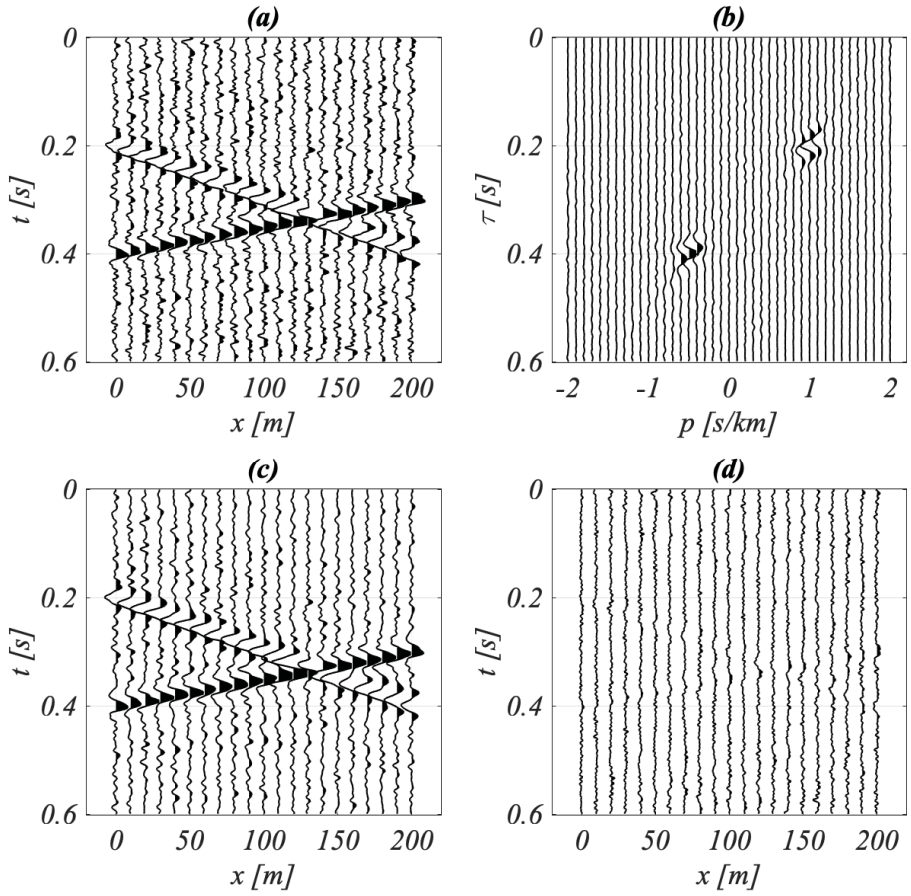

Figure 2.1: a) Two waveforms of ray parameter $p = -0.5$ s/km and $p = 1.0$ s/km. This is also the input data. b) Linear Radon transform computed via the damped least-squares method with tradeoff parameter $\mu = 0.1$. c) The predicted data synthesized from the Radon $\tau - p$ panel in c. d) The error or difference between the input data in a) and the predicted data in c).

Algorithm 1 Frequency domain solution of the Radon transform via the damped least-squared method. Given the seismic data, this code estimates the Radon transform.

Data: $d(t,x)$

offset:$\mathbf{x}$, dips: $\mathbf{p}$ both given as column vectors

min and max frequencies: $\omega_{min}, \omega_{max}$, tradeoff: μ

$D(\omega,x) = \texttt{fft}[d(t,x)]$

for $\omega \in [\omega_{min}, \omega_{max}]$ **do**

$\mathbf{d} = D(\omega,:)$

$\mathbf{L} = e^{-i\omega\mathbf{x}\mathbf{p}^T}$

$\mathbf{m} = (\mathbf{L}^H\mathbf{L} + \mu\mathbf{I})^{-1}\mathbf{L}^H\mathbf{d}$

$M(\omega,:) = \mathbf{m}$

end for

Honor the Fourier domain Hermitian symmetry

$M(-\omega,:) = M^*(\omega,:)$

$m(\tau,p) = \texttt{ifft}[M(\omega,p)]$

Algorithm 2 Frequency domain Forward Radon transform. Given the Radon transform $m(\tau,p$, this code synthesizes the seismic data.

Radon domain traces $m(t,x)$

offset:$\mathbf{x}$, dips: $\mathbf{p}$ both given as column vectors

min and max frequencies: $\omega_{min}, \omega_{max}$

$M(\omega,x) = \texttt{fft}[m(\tau,x)]$

for $\omega \in [\omega_{min}, \omega_{max}]$ **do**

$\mathbf{m} = M(\omega,:)$

$\mathbf{L} = e^{-i\omega\mathbf{x}\mathbf{p}^T}$

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

$D(\omega,:) = \mathbf{d}$

end for

Honor Fourier Domain Hermitian symmetry

$D(-\omega,:) = D^*(\omega,:)$

$d(t,x) = \texttt{ifft}[D(\omega,x)]$

2.3.2 High-resolution Radon Transforms

Rather than using a simple zero-order quadratic regularization (damping) in the solution of the Radon transform in $\omega - x$ space, Sacchi and Ulrych (1995) propose the so-called high-resolution Radon transform where the cost function to minimize is given by

$$J(\omega) = \|\mathbf{L}(\omega)\mathbf{M}(\omega) - \mathbf{D}(\omega)\|_2^2 + \mu\mathcal{R}[\mathbf{M}(\omega)] \tag{2.22}$$

where the Cauchy criterion for sparsity was adopted

$$\mathcal{R}[\mathbf{M}(\omega)] = \sum_k \ln\left(1 + \frac{|M(\omega, p_k)|^2}{\sigma_c^2}\right). \tag{2.23}$$

This norm provokes sparsity in $M(\omega, p_k)$. Therefore, it was adopted as a means to increase the focusing power of the Radon transform. Several algorithms have been proposed to minimize equation 2.22 which now is a non-quadratic cost function and hence, its condition for a minimum does not lead to a linear system of equations. Taking the derivatives of $J(\omega)$ with respect to the unknown model parameters and setting them equal to zero leads to the solution of the high-resolution $\omega - p$ Radon transform

$$\mathbf{M}(\omega) = [\,\mathbf{L}^H(\omega)\mathbf{L}(\omega) + \mu\mathbf{Q}(\omega)\,]^{-1}\mathbf{L}^H(\omega)\mathbf{D}(\omega)\,. \tag{2.24}$$

In the above expression, the matrix $\mathbf{Q}(\omega)$ is diagonal, and its diagonal elements are given by

$$[\mathbf{Q}(\omega)]_{k,k} = \frac{\sigma_c^2}{\sigma_c^2 + |\,M(\omega, p_k)\,|^2}\,. \tag{2.25}$$

The problem needs to be solved iteratively because the matrix of weights $\mathbf{Q}(\omega)$ depends on the solution $\mathbf{M}(\omega)$. Therefore, the problem can be solved iteratively adopting the method named Iteratively Reweighted Least Squares (IRLS) (Scales et al., 1988; Sacchi and Ulrych, 1995; Daubechies et al., 2010). Take for instance equation 2.24 which in iterative form becomes

$$\mathbf{M}^n(\omega) = [\,\mathbf{L}^H(\omega)\mathbf{L}(\omega) + \mu\mathbf{Q}^{n-1}(\omega)\,]^{-1}\mathbf{L}^H(\omega)\mathbf{D}(\omega) \tag{2.26}$$

where n is now used to indicate iteration number. In this algorithm one needs to set two parameters μ and σ_c. The parameter μ controls the degree of fitting to

the data. The parameter σ_c is a scale parameter that has to be properly chosen by trial and error in order to obtain sparse solutions. The solution for this method $\mathbf{M}_s(\omega) = \mathbf{M}^K(\omega)$, where K is the final iteration at which IRLS has converged. In general, convergence is achieved in a small number of iterations, typically less than 10 iterations.

Figure 2.2 illustrates a synthetic example. In Figure 2.2a we present two linear events. In 2.2b the two liner events are represented in $\tau - p$. The latter was obtained via the high-resolution solution with trade-off parameter $\mu = 1.0$ and $K = 5$ iterations. The panel in $\tau - p$ was then used to obtain the predicted data in Figure 2.2c. A comparison of Figures 2.1b and 2.2b shows that the solution with the Cauchy norm solution has more resolution and generates fewer artifacts than the solution obtained via the damped least-squares method.

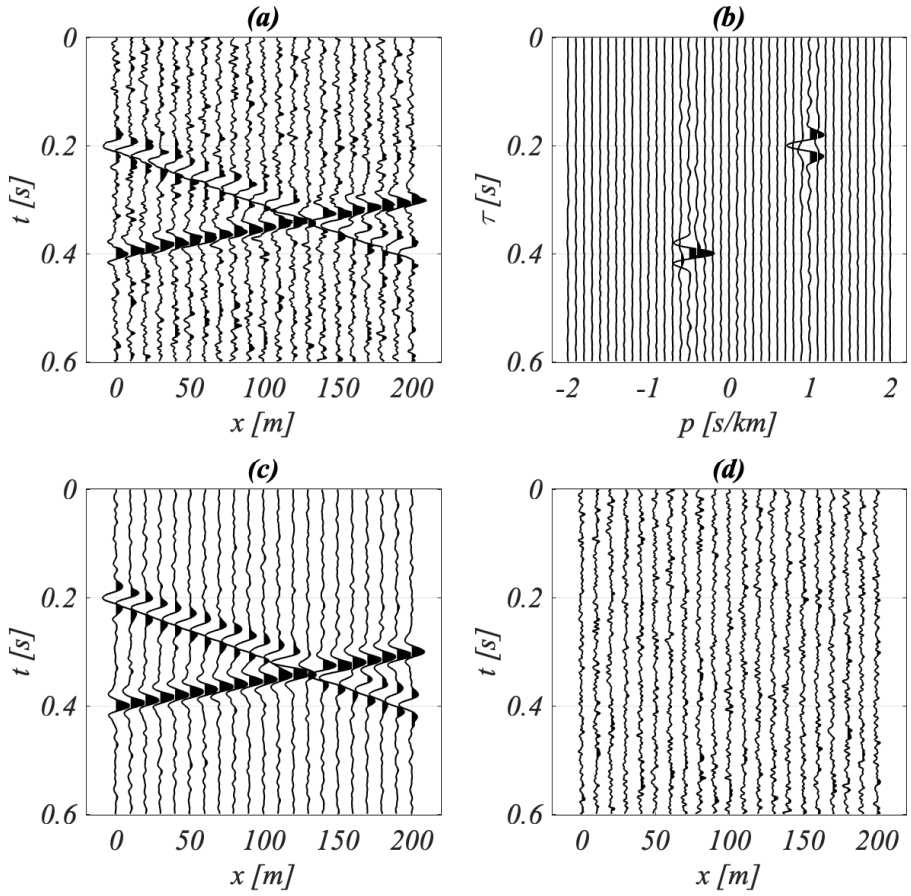

Figure 2.2: a) Two waveforms of ray parameter $p = -0.5$ s/km and $p = 1.0$ s/km. This is also the input data. b) Linear Radon transform computed via the High-resolution Radon transform with tradeoff parameter $\mu = 1.0$. c) The predicted data synthesized from the Radon $\tau - p$ panel in c. d) The error or difference between the input data in a) and the predicted data in c).

2.4 Parabolic Radon transforms

The process called multiple attenuation, or multiple removal is a challenging step in seismic exploration. Multiples are coherent noise that destructively interferes with primary reflections. Multiples are reflections that have undergone more than one reflection in the subsurface. Parabolic Radon transforms have been one of the most widely used techniques to suppress multiples. The Parabolic Radon transform pair is given by

$$\hat{m}(\tau, p) = \mathcal{L}'[d(t, x)] = \int_X d(\tau + px^2, x)\, dx\,, \tag{2.27}$$

where X indicates the range of integration of the spatial variable x. The spatial variable x for a CMP gather is the offset or source-receiver distance. Similarly, we can define a forward Radon transform via the following expression

$$d(t, x) = \mathcal{L}[m(\tau, p)] = \int_P m(t - px^2, p)\, dp\,. \tag{2.28}$$

The last two expression can be discretized and expressed in the $\omega - p$ and $\omega - x$ as done for the Linear Radon transform. However, in this case, the elements of the matrix of the $\omega - x$ Radon operator are given by

$$[\mathbf{L}(\omega)]_{n,k} = e^{-i\omega x_n^2\, p_k}\,, \quad n = 1 \ldots n_x\,,\ k = 1 \ldots n_p\,. \tag{2.29}$$

The matrix $\mathbf{L}(\omega)$ is now

$$\mathbf{L}(\omega) = \begin{pmatrix} e^{-i\omega x_1^2 p_1} & e^{-i\omega x_1^2 p_2} & \ldots & e^{-i\omega x_1^2 p_{n_p}} \\ e^{-i\omega x_2^2 p_1} & e^{-i\omega x_2^2 p_2} & \ldots & e^{-i\omega x_2^2 p_{n_p}} \\ \vdots & \vdots & \ldots & \vdots \\ e^{-i\omega x_{n_x}^2 p_1} & e^{-i\omega x_{n_x}^2 p_2} & \ldots & e^{-i\omega x_{n_x}^2 p_{n_p}}\,. \end{pmatrix} \tag{2.30}$$

and the problem again entails solving $\mathbf{D}(\omega) = \mathbf{L}(\omega)\mathbf{M}(\omega)$ to obtain the $\omega - p$ vector of complex amplitudes $\mathbf{M}(\omega)$.

2.4.1 Traveltime of reflections after the NMO correction

The Parabolic Radon transform is often applied to CMP gathers. In CMP gathers, primary and multiple reflections are approximated by seismic events with hyperbolic traveltime signatures that are symmetric $t(x) = t(-x)$ where x is the offset. The NMO-correction (Yilmaz, 1989) is applied to CMP gathers to flatten hyperbolas. For this purpose, we use the velocity of the primary reflections. Since the velocity of multiples is slower than the velocity of primaries, after NMO correction, primaries are flattened, and multiples will show residual moveout. Residual moveout after NMO correction can be represented via parabolic traveltimes. Hence, the Parabolic Radon transform can model a superposition of constant amplitude parabolas representing primaries and multiple reflections. The latter will enable us to map primaries and reflections to the Radon domain, where it is easy to separate them.

The traveltime equation for a reflection before NMO is represented via a hyperbola

$$t(x) = \sqrt{t_0^2 + \frac{x^2}{v^2}} \tag{2.31}$$

where t_0 is zero-offset two-way time and v is the RMS velocity of the reflection. The NMO correction is given by

$$\Delta t_{NMO} = t(x) - t_0 \tag{2.32}$$

or, equivalently

$$\Delta t_{NMO} = \sqrt{t_0^2 + \frac{x^2}{v_{NMO}^2}} - t_0 \tag{2.33}$$

where v_{NMO} is the NMO velocity that flattens the hyperbolas associated to primary reflections. Hence, the traveltime of an event of RMS velocity v after NMO correction is given by

$$t(x) = \sqrt{t_0^2 + \frac{x^2}{v^2}} - \sqrt{t_0^2 + \frac{x^2}{v_{NMO}^2}} + t_0 \tag{2.34}$$

using Taylor series approximation and retaining up to the second order term, equa-

tion (2.34) can be written as

$$t(x) \approx t_0 + \frac{x^2}{2t_0 v_{res}^2} \tag{2.35}$$

where v_{res} is

$$\frac{1}{v_{res}} = \frac{1}{v} - \frac{1}{v_{NMO}}\,. \tag{2.36}$$

We define p, the curvature of the reflections after the NMO correction

$$p = \frac{1}{2t_0}(\frac{1}{v^2} - \frac{1}{v_{NMO}^2})\,. \tag{2.37}$$

Consequently, one can write the parabolic traveltime

$$t(x) = t_0 + px^2\,. \tag{2.38}$$

The traveltimes for a primary reflection after NMO correction becomes $t(x) = t_0$ because we choose $v_{NMO} = v$; in other words, the NMO correction is performed with the velocity of the primary. Multiple reflections, in general, have a velocity that is lower than the velocity of primary reflections and hence $v < v_{NMO}$. For this reason, multiples after NMO correction will show parabolic residual moveout. The parametrization in terms of p is usually replaced by the following parametrization

$$t(x) = t_0 + q\left(\frac{x}{x_f}\right)^2\,. \tag{2.39}$$

With the parametrization given above, q represents residual moveout at far offset x_f (maximum absolute offset). The parametrization 2.39 is more convenient that the one given by expression 2.38. The latter is because q is residual moveout in seconds, a parameters much easier to interpret than p in equation 2.38 which has units of s/m^2.

It is important to mention that the parabolic approximation could also be validated via a t^2 time stretching (Yilmaz, 1989). In this case, hyperbolic traveltimes in a CMP gather are transform via the following change of variables $t' = t^2$ which leads to

$$t^2 = t_0^2 + \frac{x^2}{v^2} \longrightarrow t' = t_0' + \frac{x^2}{v^2}\,. \tag{2.40}$$

If we call $p = \frac{1}{v^2}$, then hyperbolic traveltimes in time-stretched coordinates become

parabolic traveltimes. In my research, I have not used time-stretching to validate the parabolic approximation because it produces an undesired time-variant wavelet deformation (Yilmaz, 1989). All my examples involving Parabolic Radon transforms adopt the NMO correction to validate the parabolic approximation.

Figure 2.3 illustrates a synthetic example. In Figure 2.3a, we present two NMO-corrected events. One event mimics a primary and the other a multiple with residual moveout. In 2.3b the events are represented in $\tau - q$. The latter was obtained via the damped-least squares solution with tradeoff parameter $\mu = 0.7$. The panel in $\tau - q$ was then used to obtained the predicted data in Figure 2.3c and the error panel in Figure 2.3d.

Figure 2.4 illustrates the same synthetic example shown in Figure 2.2. However, in this example, I used the high-resolution Radon transform as proposed by Sacchi and Ulrych (1995) with Cauchy regularization term. Heuristically, the best tradeoff parameter was found to be $\mu = 10$. It is clear that the high-resolution Radon transform has produced a Radon panel with more focused events than the panel obtained via the damped least-squares solution.

2.4.2 Real data example

Figures 2.5 and 2.6 compare multiple attenuation via the Parabolic Radon transform with the classical solution that uses damped least-squares inversion and for the high-resolution Radon transform obtained via Cauchy regularization. Figure 2.5a shows a CMP gather after NMO correction. The data corresponds to a marine survey carried out in the Gulf of Mexico by Western Geophysical. Generally known as the Mississippi Canyon dataset, the data was released to researchers and used to test multiple attenuation algorithms by several research groups (Verschuur, 1999; Li, 2001; Sabbione and Sacchi, 2017). Figure 2.5b shows the Radon solution via damped least-squares. The Radon panel was filter with a cutoff curvature $q_c = 0.05$s. In other words, the primaries were eliminated by removing all Radon coefficients with residual moveout $q < q_c$. Then the remaining coefficients were mapped back to data space to synthesize the multiple reflections. Finally, multiples are removed from the original data by simple subtraction to yield Figure 2.5c. As one can see, the curved events corresponding to under-corrected multiples were eliminated from the CMP gather. The process is repeated in Figure 2.6 but using the high-resolution

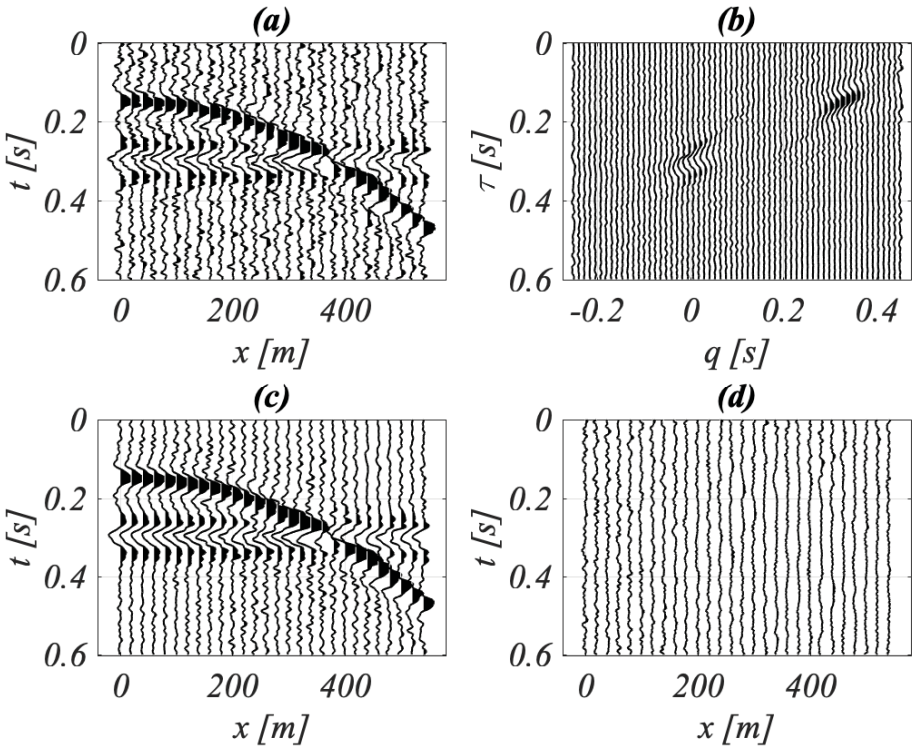

Figure 2.3: a) The input data are two NMO-corrected events. b) The Parabolic Radon transform computed via the damped least-squares method with tradeoff parameter $\mu = 0.70$. c) The predicted data synthesized from the Radon $\tau - q$ panel in c. d) The error or difference between given by the input data in minus the predicted data.

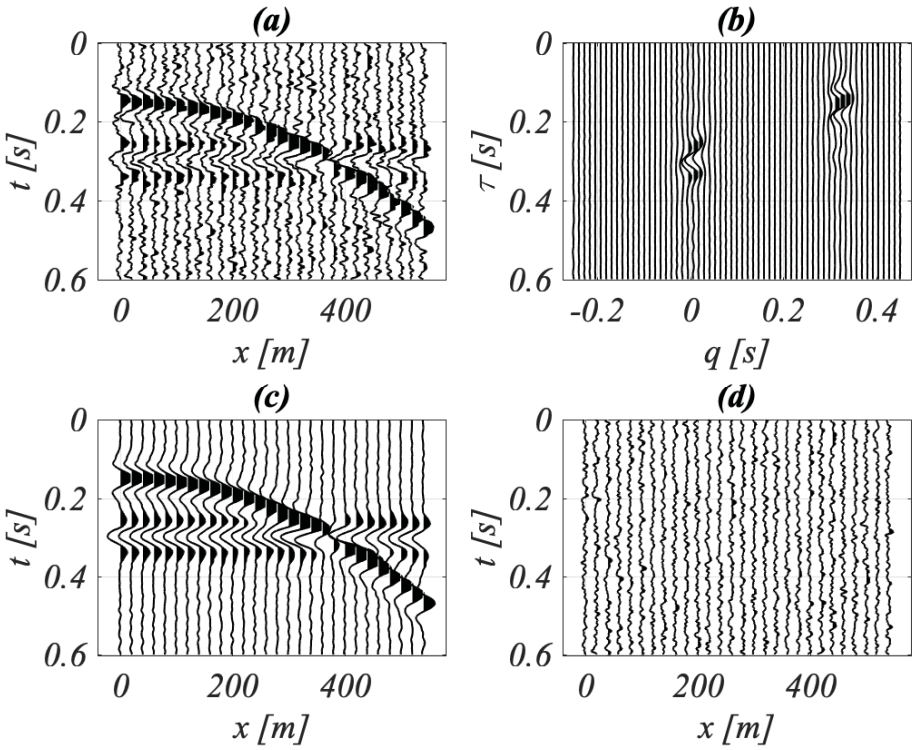

Figure 2.4: a) The input data are two NMO-corrected events. b) The Parabolic Radon transform computed via the damped high-resolution Parabolic Radon transform with tradeoff parameter $\mu = 10$. c) The predicted data synthesized from the Radon $\tau - q$ panel in c. d) The error or difference between given by the input data in minus the predicted data.

Parabolic Radon transform implemented via a Cauchy regularization term. If we compare Figure 2.5b and 2.6b, it is clear that the high-resolution Radon transform attempts to collapse the smearing of events in the $\tau - q$ domain. In other words, events are more focused when adopting the high-resolution Radon transform than using the classical Radon transform with the damped least-squares solver.

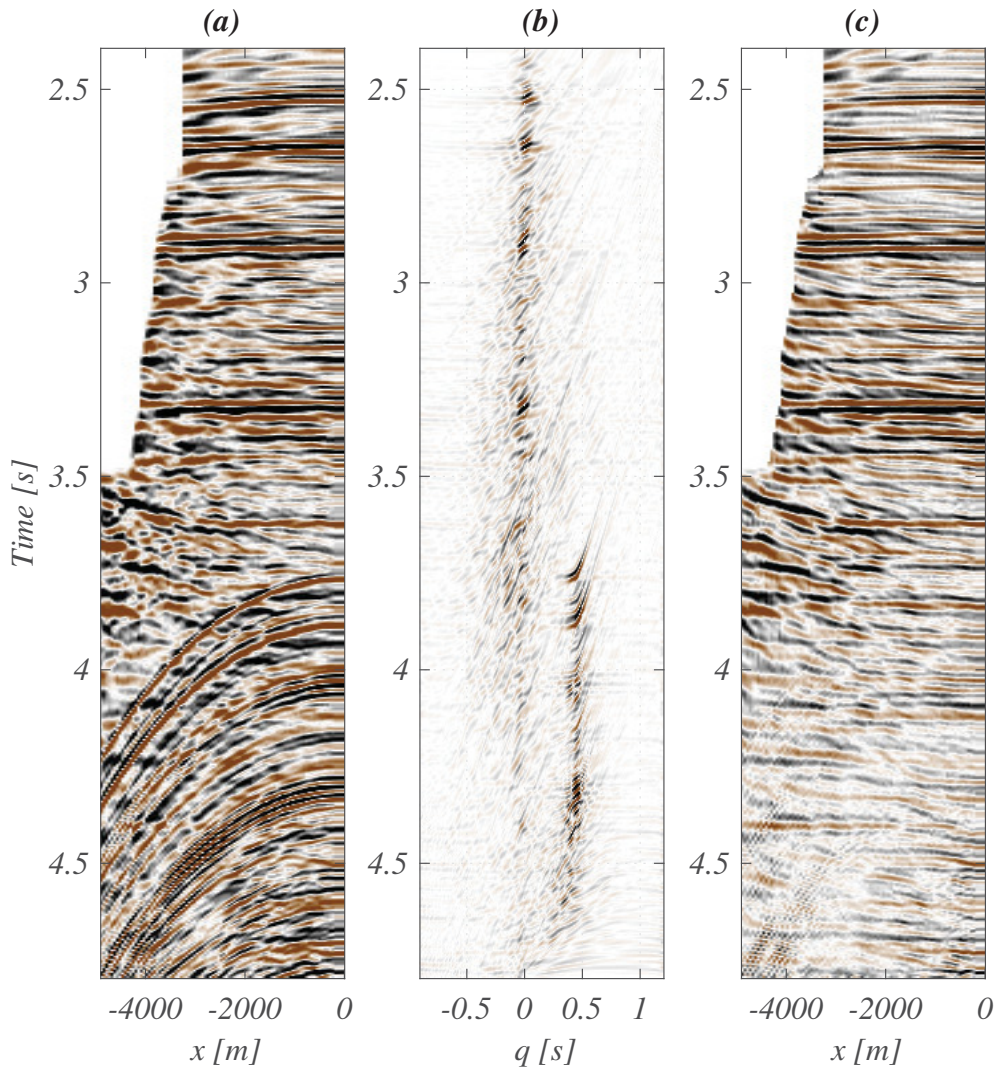

Figure 2.5: a) CMP gather from the Gulf of Mexico (Mississippi Canyon dataset). b) Parabolic Radon transform (Damped least-squares solution). c) The estimated primaries.

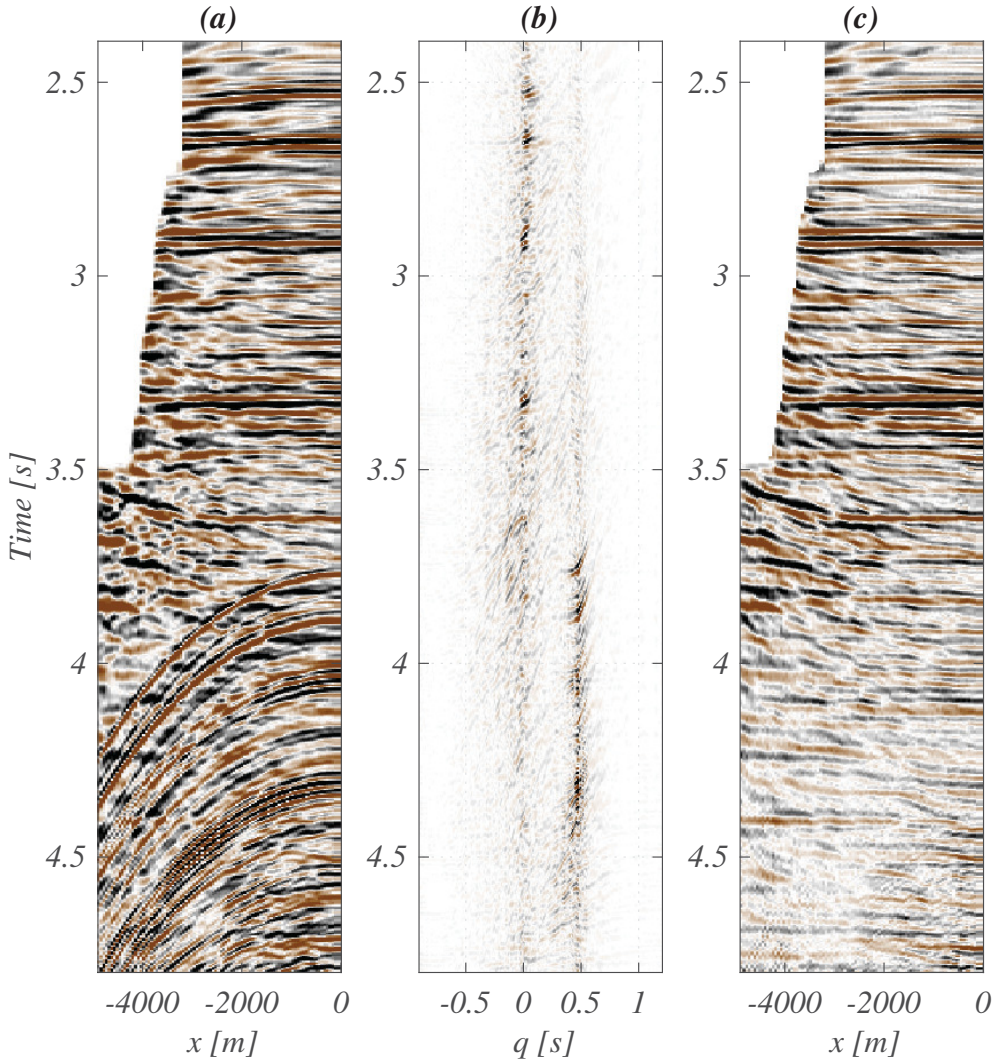

Figure 2.6: a) CMP gather from the Gulf of Mexico (Mississippi Canyon dataset). b) Parabolic Radon transform (High-resolution solution). c) The estimated primaries.

CHAPTER 3

Time domain Radon transforms

3.1 Introduction

Chapter 2 examines the classical damped least-squares, and high-resolution Radon transforms (Hampson, 1986a; Beylkin, 1987; Sacchi and Ulrych, 1995) with algorithms that are implemented in the $\omega - x$ domain with explicit solvers. These algorithms are very efficient because the Radon transform is computed by solving small inverse problems in the $\omega - x$ domain. Moreover, as pointed out by Kostov (1990) and further explored by Sacchi and Porsani (2005), one can exploit the Toeplitz structure of the $n_p \times n_p$ matrix $\mathbf{L}^H(\omega)\mathbf{L}(\omega)$ in the $\omega - x$ solutions discussed in Chapter 2. For instance, systems of equations with Toeplitz structure can be efficiently solved via Levinson recursion in $\mathcal{O}(n_p^2)$ operations (Kostov, 1990).

This chapter explores Radon domain solutions computed in the $t - x$ domain via implicit form operators. This idea originates in the paper by Thorson and Claerbout (1985) that develops linear and hyperbolic Radon transforms with $t - x$ operators. The latter also allows the development of Radon transforms for time-variant integration paths, such as in the case of the Hyperbolic Radon transform. It also provides algorithms where regularization is directly imposed on $\tau - p$ rather than in the $\omega - p$ domain, as discussed in Chapter 2. In essence, both the misfit function that controls data fit and the model norm regulating the smoothness or spareness of the solution is posed in the $t - x$ and $\tau - p$ domain. This type of method has been investigated by Cary (1998) who was the first to recognize that time-domain solutions are more

stable than $\omega - x$ solutions at the time of computing sparse (high-resolution) Radon transforms. The problem has also been studied in detail by Trad et al. (2003) who discussed different iterative solvers to obtain time-domain Radon transforms.

3.2 Hyperbolic Radon transform

The hyperbolic Radon transformation is a time-variant Radon operator that was first introduced by Thorson and Claerbout (1985) who coined the name Velocity Stacks for it. The integration is now across hyperbolas of the form

$$t = \sqrt{\tau^2 + \frac{x^2}{v^2}} = \sqrt{\tau^2 + px^2}\,, \tag{3.1}$$

where τ is the intercept (zero offset) time, and v is the velocity of the hyperbola. In other words, we are after a mapping from $t - x$ to $\tau - p$ or more specific to $\tau - v$ where $p = 1/v^2$. This type of transform is an excellent operator to process seismic data in the CMP domain, where one can consider a superposition of waveforms with hyperbolic traveltimes. Notice that the integration path is time-variant, and therefore, one cannot use the Fourier delay theorem to map the transform to the $\omega - x$ domain as it was done in Chapter 2.

As in the Linear and Parabolic Radon transform, we express the adjoint and forward operators as follows

$$\hat{m}(\tau, v) = \int_X d(t = \sqrt{(\tau^2 + \frac{x^2}{v^2}}, x)\, dx \tag{3.2}$$

$$d(t, x) = \int_V m(\tau = \sqrt{(t^2 - \frac{x^2}{v^2}}, v)\, dv\,. \tag{3.3}$$

Now, our task is estimating from the data $d(t, x)$ the associated Hyperbolic Radon transform $m(\tau, v)$. Clearly, we first need to discretize the operators

$$\hat{m}(\tau, v_j) = \sum_{i=1}^{n_x} d(t = \sqrt{(\tau^2 + \frac{x_i^2}{v_j^2}}, x_j) \quad j = 1 \ldots n_v \tag{3.4}$$

and

$$d(t, x_i) = \sum_{j=1}^{n_v} m(\tau = \sqrt{(t^2 - \frac{x_i^2}{v_j^2}}, v_j) \quad i = 1 \dots n_x \tag{3.5}$$

where we have discretized the velocity axis v in n_v velocities. The offset x is also discrete and corresponds to n_x seismic traces. The time axes t and τ are also discrete. For simplicity let us assume they have been discretize at the same sampling rate Δt following $t_n = (n-1)\Delta t$ and $\tau_n = (n-1)\Delta t$ with $n = 1 \dots n_t$. Then one can use operator notation to discribe equations 3.4 and 3.5 as follows

$$\hat{\mathbf{m}} = \mathcal{L}'\mathbf{d} \tag{3.6}$$

and

$$\mathbf{d} = \mathcal{L}\mathbf{m}\,, \tag{3.7}$$

where $\mathbf{m}$ is the vector of length $(n_t \times n_v) \times 1$ of Radon coefficients and $\mathbf{d}$ is the vector of length $(n_t \times n_x) \times 1$ containing the CMP gather. Clearly, the action of the hyperbolic Radon adjoint operator on $\mathbf{d}$ is the operator $\mathcal{L}'$ and the action of the forward Radon operator on $\mathbf{m}$ is the operator $\mathcal{L}$. We use calligraphic fonts to indicate operators and we stress that these two operators $\mathcal{L}$ and $\mathcal{L}'$ are not stored as matrices (e.g. $\mathbf{L}$ and $\mathbf{L}^T$). One can think that $\mathcal{L}$ and $\mathcal{L}'$ are functions or subroutines whose action is equivalent to that of applying matrix-times-vector multiplications. It is clear that the matrix $\mathbf{L}$ cannot be constructed in explicit form[1]. because such a matrix will have an unmanageable size of $(n_t \times n_x) \times (n_t \times n_v)$. The latter is forbiddingly large for applications involving CMPs, for instance, of typical size $n_x = 100$ traces $n_v = 100$ velocities and $n_t = 1000$ time samples.

Clearly, given that $\mathcal{L}$ and $\mathcal{L}'$ are operators (not matrices), one cannot find, for instance, the least-squares solution via explicit form expressions that involve $(\mathcal{L}'\mathcal{L})^{-1}$. Hence, we will resort to iterative solvers where the operators $\mathcal{L}$ and $\mathcal{L}'$ are applied without constructing their matrix-form expression. In essence, we will use an algorithm to compute either $\mathcal{L}$ or $\mathcal{L}'$ and then resort on the dot-product test to guarantee that these algorithms behave as if they were mimicking matrix-times-vector multiplications (see, for instance, Claerbout (1992), for a clear exposition of the dot-product test.) Before discussing the inverse problem, we also provide pseudo-codes for the operators $\mathcal{L}$ and $\mathcal{L}'$ in Algorithm 3. Bear in mind, that Algorithm 3 is a

[1]The opposite of explicit form is an implicit form or matrix-free method.

simplification and the actual algorithm is more complicated because it involves an interpolation step to assign times t or τ to samples of $d(t_n, x)$ or $m(\tau_n, v)$ where t_n and τ_n are discrete time samples, respectively.

Algorithm 3 Time-domain implicit Radon operators for Hyperbolic traveltimes. The algorithm applies the forward $\mathcal{L}$ or the adjoint $\mathcal{L}'$ operator.

```
if adj = true      Input is d (do m̂ = L'd)
if adj = false     Input is m (do d = Lm)
for τ do
    for h do
        for v do
            t = sqrt(τ^2 + (x/v)^2)
            if adj = true      m̂(τ,v) = m̂(τ,v) + d(t,x)
            if adj = false     d(t,x) = d(t,x) + m(τ,v)
        end for
    end for
end for
if adj = true     Return m̂
if adj = false    Return d
```

3.3 The inverse problem

We now are in the condition of estimating $\mathbf{m}$ from $\mathbf{d}$ to estimate the hyperbolic Radon transform. For this purpose, we will use iterative solvers (See Appendix) to find the solution $\mathbf{m}_s$ that minimize the cost function

$$\mathbf{m}_s = \underset{\mathbf{m}}{\operatorname{argmin}} \|\mathbf{d} - \mathcal{L}\mathbf{m}\|_2^2 + \mu\mathcal{R}(\mathbf{m}) \tag{3.8}$$

where $\mathcal{R}$ is the regularization term. Once $\mathbf{m}_s$ is estimated, it can be used to predict he seismic data via $\mathbf{d}_{pred} = \mathcal{L}\mathbf{m}_s$.

3.3.1 Least-squares solution with damping via the Conjugate Gradient Least-squares (CGLS) method

If $\mathcal{R}$ is the squared l_2 norm of the model parameters then we have

$$\mathbf{m}_s = \underset{\mathbf{m}}{\operatorname{argmin}} \|\mathbf{d} - \mathcal{L}\mathbf{m}\|_2^2 + \mu\|\mathbf{m}\|_2^2 \tag{3.9}$$

where $\|\mathbf{m}\|_2^2 = \sum_{n,j} |m(\tau_n, v_j)|^2$. The solution that minimizes J can be found via the method of conjugate gradient (CG) (Hestenes et al., 1952) or more specifically, the Conjugate Gradient Least-squares method (CGLS) which operates directly with operator $\mathcal{L}$ and $\mathcal{L}'$ (Golub and Van Loan, 1996). One important aspect of CGLS that one must mention is that the algorithm finds the minimum of J via a series of steps where one has to only evaluate the action of $\mathcal{L}$ and $\mathcal{L}'$. In other words, one needs to know how to perform the rules $\mathcal{L}\mathbf{u}$ and $\mathcal{L}'\mathbf{v}$ to any arbitrary vector $\mathbf{u}$ and $\mathbf{v}$ of model and data space, respectively. If one Knows these two rules, the minimum of J is attained in a total of iterations equal to the number of unknowns. However, it is important to mention that an approximate solution can be found in fewer steps (iterations). The solution corresponds to the damped least-squares solution of the Hyperbolic Radon transform that, for simplicity, we will name the CGLS solution. Trad et al. (2003) has also discussed the possibility of using preconditioning and iteration truncation to find time-domain Radon transforms with CGLS.

3.3.2 Least-squares solution with a sparsity constraint (High-resolution Radon transform)

One can find high-resolution solutions by invoking a sparsity-promoting solution. In this case, we adopt an l_1 regularization term that forces solutions that are sparse. In other words, we solve the problem

$$\mathbf{m}_s = \underset{\mathbf{m}}{\operatorname{argmin}} \|\mathbf{d} - \mathcal{L}\mathbf{m}\|_2^2 + \mu\|\mathbf{m}\|_1 \,. \tag{3.10}$$

The l_1 norm of the Radon coefficients is given by $\|\mathbf{m}\|_1 = \sum_{n,j} |m(\tau_n, v_j)|$. The cost function to minimize in equation 3.10 is non-quadratic, and therefore, one cannot use the conjugate gradient method. Different methods have been proposed to solve this type of problem. In this thesis, two methods have been adopted. One is Iterative

reweighted least-squares (IRLS) with a CGLS inner solver (Trad et al., 2003), and the other is the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) proposed by Beck and Teboulle (2009). Both IRLS and FISTA only require the evaluation of $\mathcal{L}$ and $\mathcal{L}'$ in each iteration. As such, these algorithms are incredibly flexible for situations where one has access to implicit form operators (See Appendix).

3.3.3 Incorporation of a wavelet into the time-domain Radon transform

We have found the convergence of CGLS, IRLS, and FISTA can be accelerated by incorporating a wavelet in the implicit-form formulation of the Radon problem. For instance, in the codes, $\mathcal{L}$ is replaced by $\mathcal{W}\mathcal{L}$ where $\mathcal{W}$ is an operator that applies convolution with a wavelet. Similarly, the adjoint operator becomes $\mathcal{L}'\mathcal{W}'$ where now $\mathcal{W}'$ means crosscorrelation with a wavelet (Claerbout, 1992). This modification helps find a solution that requires fewer coefficients in the Radon domain. In other words, we are using a more accurate forward operator given by $\mathbf{d} = \mathcal{W}\mathcal{L}\mathbf{m}$. The addition of the wavelet does not add significantly to the computational cost of the solution. In cases when the wavelet is unknown, one can adopt a simple band-pass zero-phase wavelet with a band estimated from the amplitude spectrum of the data. This approximated wavelet is often sufficient to obtain a reasonable solution $\mathbf{m}_s$.

3.4 Synthetic data examples

Algorithm 3 can also be used to compute the forward and adjoint operators for the parabolic Radon transform by replacing the hyperbolic integration path with $t = \tau + q(x/x_f)^2$ (the parabolic approximation in terms of residual moveout at far offset). This section uses synthetic experiments to test the $t - x$ inversion of the parabolic, and hyperbolic Radon transforms and compares inversions computed via CGLS, IRLS and FISTA.

Figure 3.1a show and example with two primaries and four multiples after NMO correction. Figure 3.1b shows the $\tau - q$ model estimated by CGLS. Figure 3.1c is the reconstructed data from the $\tau - q$ domain. Similarly, Figure 3.1d is the estimated Radon $\tau - q$ data obtained via the FISTA solver and the associated reconstructed data in Figure 3.1e. Finally, Figures 3.1f and g are the $\tau - q$ panel and

the reconstructed data obtained via the IRLS algorithm. Notice that both FISTA and IRLS have produced similar $\tau - q$ transform of superior resolution than the $\tau - q$ domain estimated via CGLS. This is due to the having adopted the l_1 constraint in the estimation of Figure 3.1d (FISTA) and Figure 3.1g (IRLS). In Figure 3.2 we show modelled multiples after applying muting in $\tau - q$ domain and transforming back the multiples in $\tau - q$ to $t - x$. Then the multiples were subtracted from the data to obtain the primaries. Figures 3.2 a, b and c correspond to the solution obtained via CGLS. Similarly, Figures 3.2d, e and f and 3.1g, h and i correspond to sparse Radon transforms using FISTA and IRLS, respectively.

Figure 3.3a show and example with primaries and multiples in a synthetic CMP gather. Contrary to the previous example, we did not apply NMO correction, and this experiment is concerned with testing the hyperbolic Radon transform. Figure 3.3b shows $\tau - v$ model estimated by CGLS. Figure 3.3c is the reconstructed data from the $\tau - v$ domain. Similarly, Figures 3.3d is the estimated Radon $\tau - v$ data obtained via the FISTA solver and the associated reconstructed data in Figure 3.3e. Finally, Figures 3.3f and g are the $\tau - v$ panel and the reconstructed data obtained via the IRLS algorithm. Notice that both FISTA and IRLS have produced similar $\tau - v$ transforms of improved resolution than the $\tau - v$ domain estimated via CGLS. This is due to the having adopted the l_1 constraint in the estimation of Figure 3.3d (FISTA) and Figure 3.3g (IRLS). In Figure 3.4 we show modelled multiples after applying muting in $\tau - v$ domain and transforming back the multiples in $\tau - v$ to $t - x$. Then the multiples were subtracted from the data to obtain the primaries. Figures 3.4 a, b and c correspond to the solution obtained via CGLS. Similarly, Figures 3.4d,e and f and 3.3g,h and i correspond to sparse Radon transforms employing FISTA and IRLS, respectively.

3.5 Offset regularization via the hyperbolic Radon transform

For completeness, the last example shows an application of the hyperbolic Radon transform to offset (source-receiver) regularization. Figure 3.5a shows a CMP gather from an onshore survey. The offset is irregularly sampled. This often happens when the sources and receivers are not regularly distributed on the acquisition line. We use the hyperbolic Radon transform to estimate the $\tau - v$ representation of the

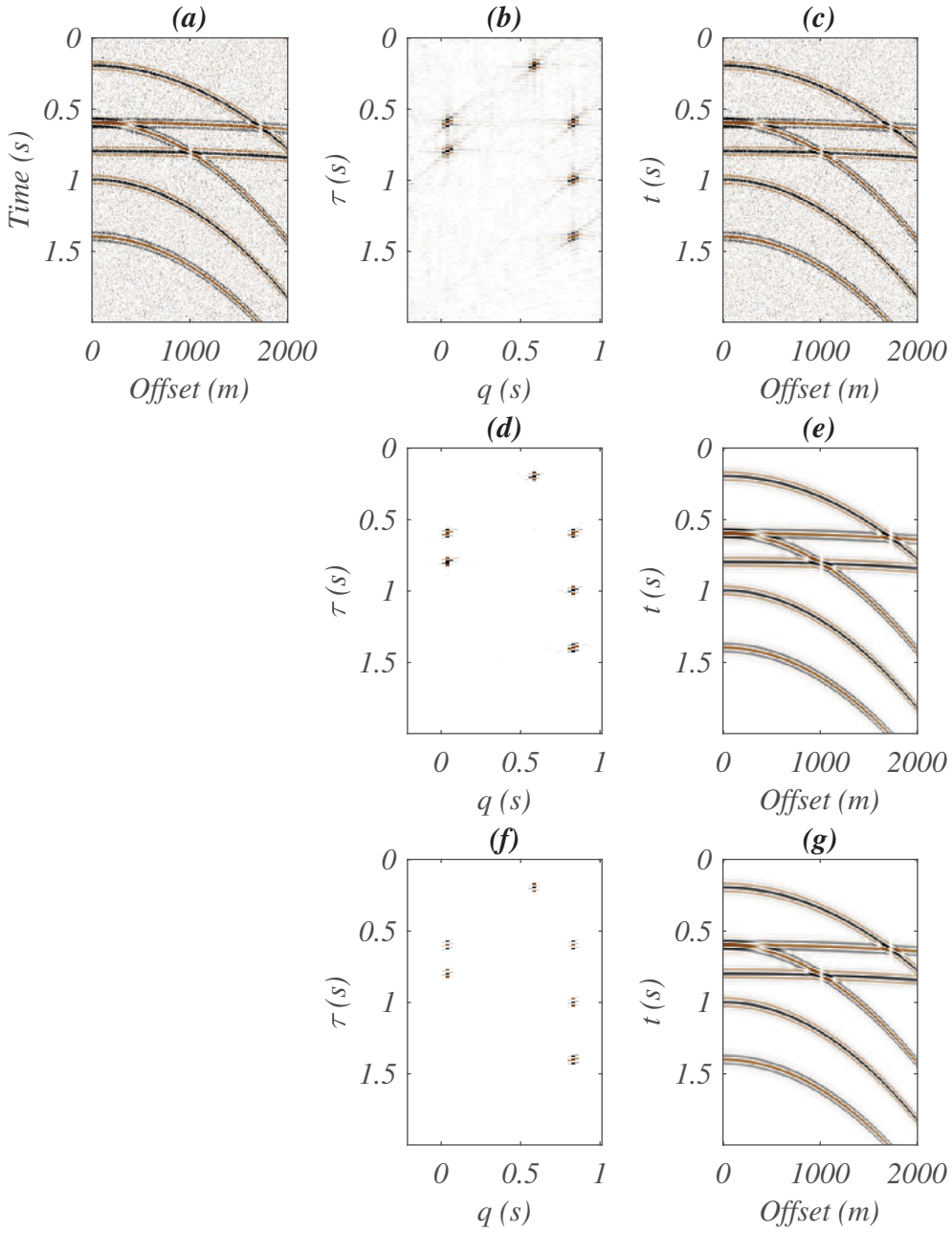

Figure 3.1: a) NMO corrected data composed of primary and multiple events. b) $\tau - q$ domain estimated via CGLS. c) Predicted data from b). d) Sparse $\tau - q$ domain estimated via FISTA. e) Predicted data from d). f) Sparse $\tau - q$ domain estimated via IRLS. g) Predicted data from f).

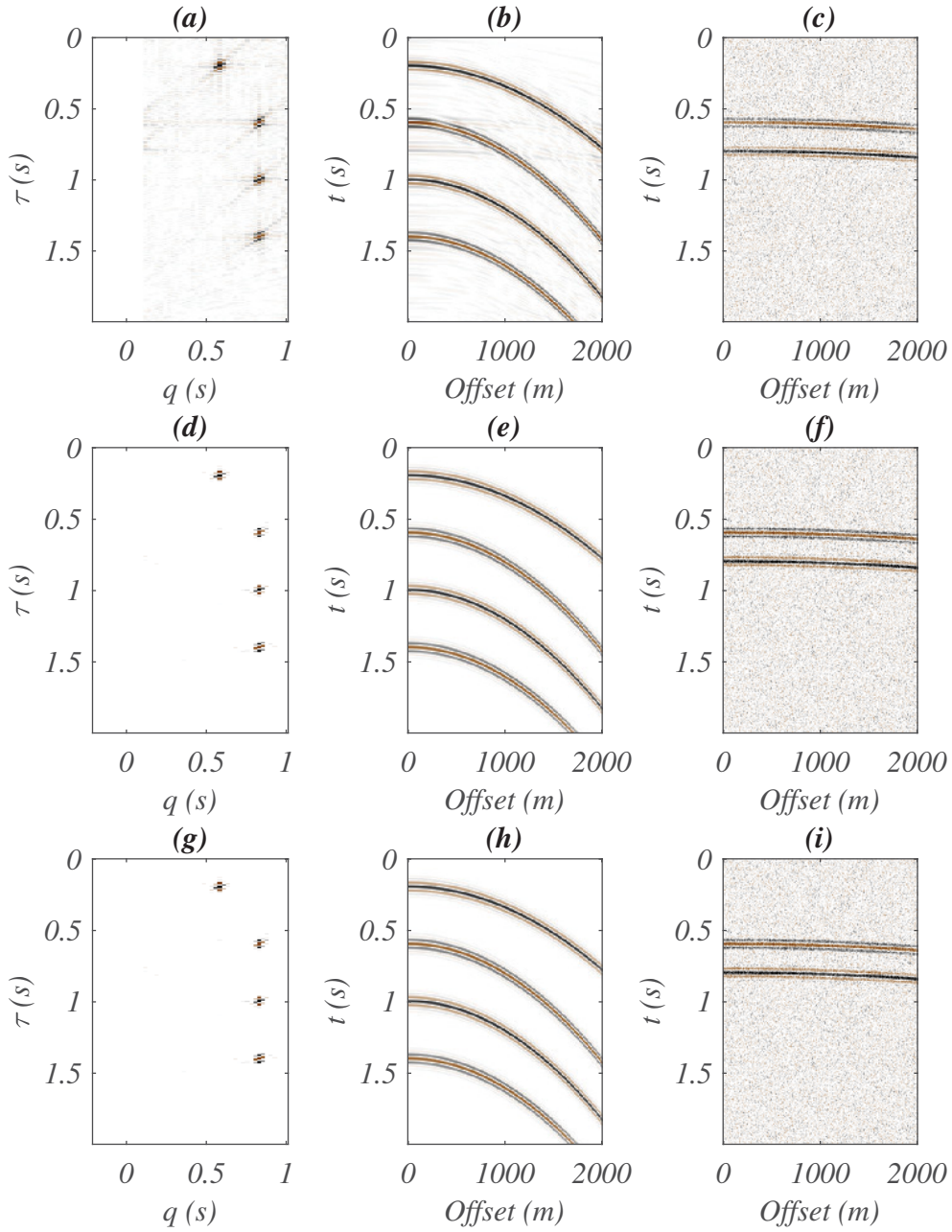

Figure 3.2: a), b) and c) $\tau - q$ domain, predicted multiples and primaries via CGLS. d), e) and f) Sparse $\tau - q$ domain, predicted multiples and primaries via FISTA. g), h) and i) Sparse $\tau - q$ domain, predicted multiples and primaries via IRLS.

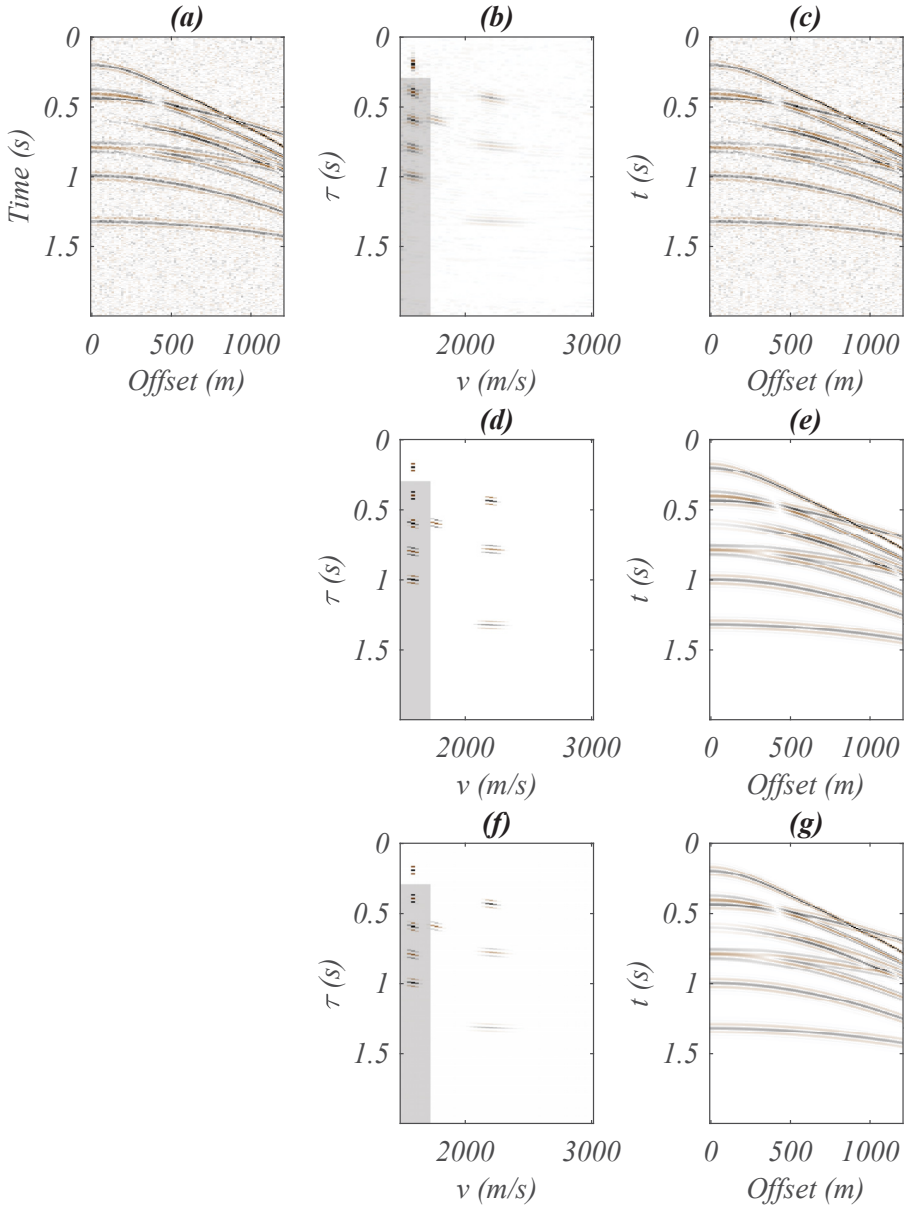

Figure 3.3: a) Data composed of primary and multiple events. b) $\tau - v$ domain estimated via CGLS. c) Predicted data from b). d) Sparse $\tau - v$ domain estimated via FISTA. e) Predicted data from d). f) Sparse $\tau - v$ domain estimated via IRLS. g) Predicted data from f). The shaded area corresponds to the multiples in $\tau - v$ space.

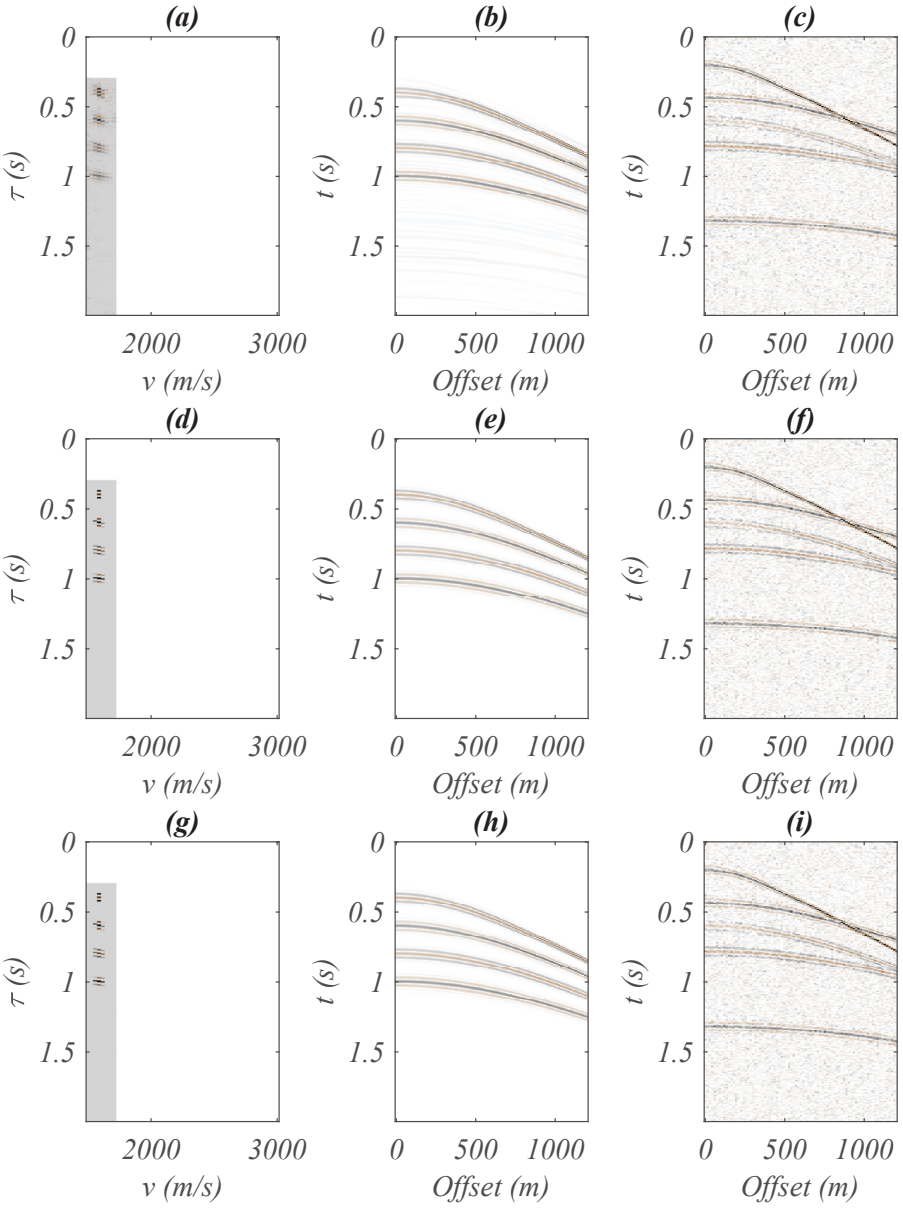

Figure 3.4: a), b) and c) $\tau - v$ domain, predicted multiples and primaries via CGLS. d), e) and f) Sparse $\tau - v$ domain, predicted multiples and primaries via FISTA. g), h) and i) Sparse $\tau - v$ domain, predicted multiples and primaries via IRLS.

data. Then, from the $\tau - v$ result, one can synthesize data at new spatial regular coordinates. Figures 3.5b and c show the $\tau - v$ panels computed with the Radon transforms computed via CGLS and FISTA (sparse), respectively. Figure 3.5d and e are the reconstructed data from the $\tau - v$ panels in Figure 3.5b and c, respectively. This technique can be utilized to regularize the offset, for instance, before seismic migration. Last, Figure 3.6 shows the velocity spectra (Semblance) for the data before regularization and after regularization from the $\tau - v$ solutions obtained via the FISTA and IRLS algorithms.

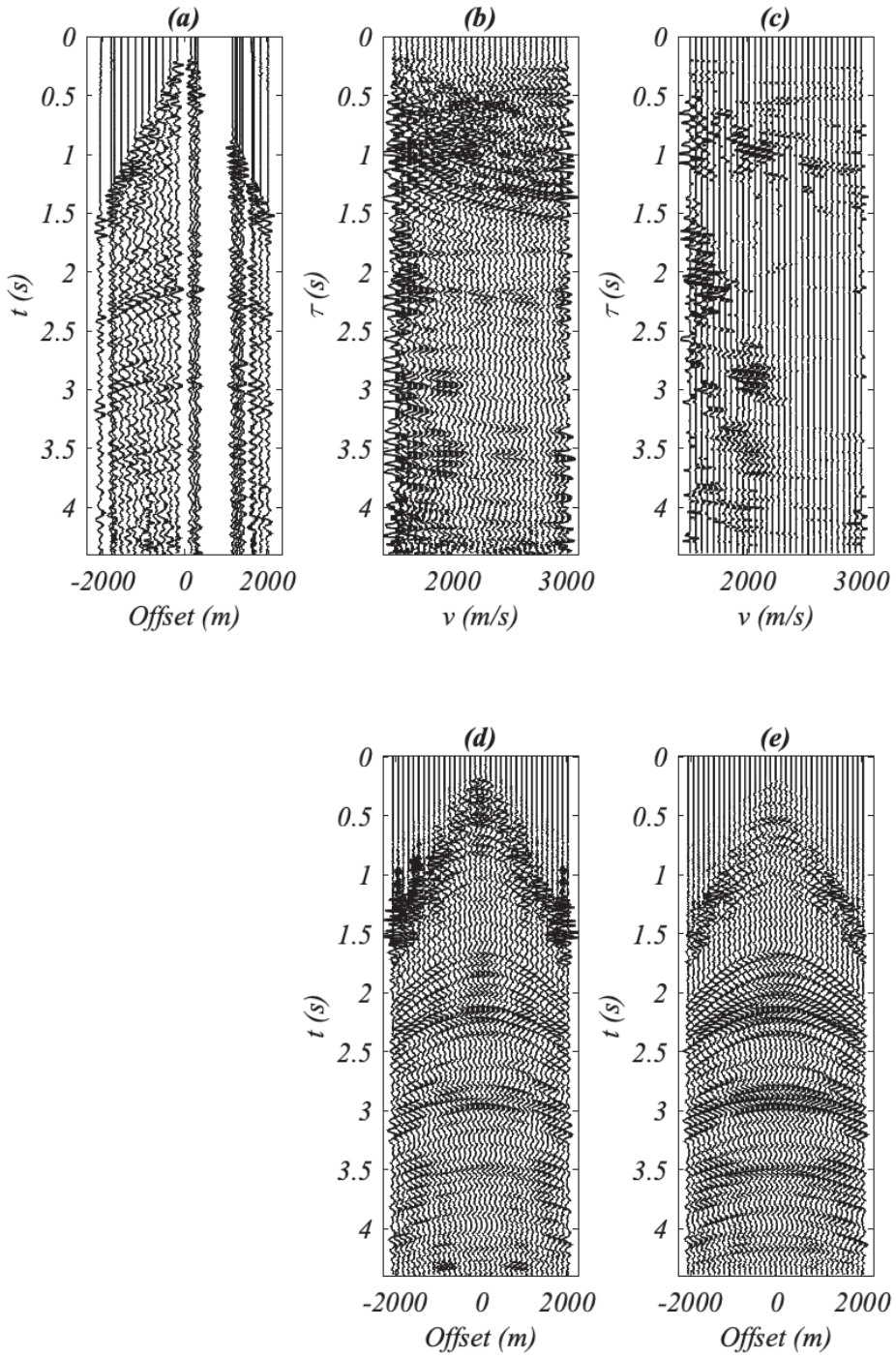

Figure 3.5: a) Onshore CMP gather with irregular distribution of offset. b) $\tau - v$ transform obtained via CGLS. c) Sparse $\tau - v$ transform obtained via the FISTA algorithm. d) Reconstructed from b). e) Reconstructed from c).

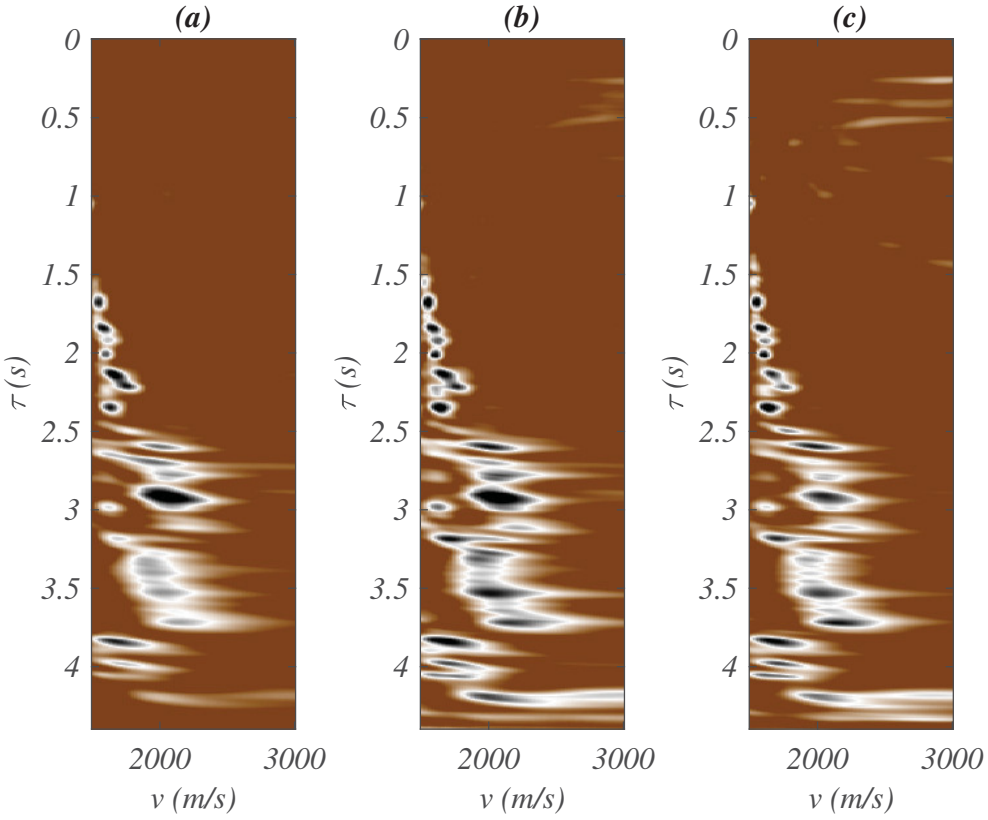

Figure 3.6: Semblance panels showing velocity spectra. a) Velocity spectra of the original CMP gather in Figure 3.5a. b) Velocity spectra of the reconstructed data in Figure 3.5d. c) Velocity spectra of the reconstructed data in 3.5e. Notice the superior resolution of events at about 3.5s in b) and c).

CHAPTER 4

Local Radon transform with a second-order traveltime approximation

4.1 Introduction

This chapter will introduce a new transform named the local Radon transformation. Many studies have given much attention to optimized Radon transformations for high-resolution solutions and lower computational costs. However, these studies are often concerned with the application of linear, parabolic and hyperbolic Radon transforms. These transforms can not handle complex waveforms well because they have been explicitly designed to cope with linear, parabolic, and hyperbolic moveout events. Moreover, they are often designed to operate on the whole aperture of a seismic gather where waveforms with specific traveltime trajectories are expected to be found. Contrary to the latter, seismic data can have complexity not captured by linear, parabolic or hyperbolic traveltimes.

Radon transform can be characterized into three categories based on the specific integration paths to stack the energy along the designated integral lines. The most common lines for Radon transform applied in seismic processing are straight lines, parabolic and hyperbolic curves. However, the travel time equation is not always hyperbolic for reflections, and they also depend on which data domain reflections are analyzed. For instance, hyperbolic events are correct for many applications when the data are processed in the CMP domain. However, simple hyperbolic traveltimes might not be optimal in common-shot or common-receiver domains in the presence

of structural dip. Also, in complex structures such as overthrust tectonics, salt domes and basalt, the assumption of hyperbolic trajectories is disrupted because of the complexity of the subsurface and propagation of waves in anisotropic media. Therefore, classical Radon transforms suffer from inaccuracy, particularly at large offsets. Other moveout parametrization could be used as generalized moveout approximations such as those proposed by Stovas and Fomel (2017), the method of Ravve and Koren (2017) and the extended generalized moveout approximation of Abedi and Stovas (2019) to cope, for instance, with traveltimes signatures caused by anisotropy.

In this thesis, I propose a local Radon transform that can adapt to complex waveforms. By making the Radon transform local, one can use Radon domain coefficients to synthesize complex waveform with laterally varying dips. Many applications of this proposed local Radon transform come to mind. For instance, a local Radon transform can be used for seismic data denoising and seismic data interpolation. Moreover, following the work of Lin and Sacchi (2020) a local Radon transform could serve for robust simultaneous source processing.

4.2 Local Radon transforms

Opposite to the conventional local Radon transform parameterized in slope, curvature (residual moveout) or velocity, I have developed a multi-parameter parameterization for the new transform using a second-order Taylor expansion to approximate the traveltime of reflections in any domain. The main idea of this approach is to use dip and curvature information of seismic events in local windows of neighbouring traces to extract the corresponding Radon domain coefficients that synthesize the data. As shown in Chapters 2 and 3, I will solve an inverse problem where the data is mapped from $t - x$ to the Radon domain. The local Radon domain now depends on two kinematic parameters described in the following sections, and the inverse problem becomes rather cumbersome for explicit form solutions as those proposed in Chapter 2. Hence, I will resort to implicit operators following the route already described in Chapter 3.

We first consider a local approximation of the traveltime of a reflection in terms of

a second-order Taylor expansion.

$$T(x) \approx T_{app}(x) = T_0 + A(x)(x - x_0) + B(x)(x - x_0)^2 \,. \tag{4.1}$$

In the above, T_0 is the intercept time at $x = x_0$, x is offset, and x_0 is the position of a central seismic trace. The parameters A and B correspond to derivatives evaluated at x_0. Clearly, A is also dip or local ray parameter and B is curvature

$$T(x) \approx T_{app}(x) = T_0 + \frac{\partial T(x_0)}{\partial x}(x - x_0) + \frac{1}{2}\frac{\partial^2 T(x_0)}{\partial x^2}(x - x_0)^2 \,. \tag{4.2}$$

Once could divide the spatial aperture of the data into spatially localized windows, and the Radon transform can be implemented for each window independently for a specific number of traces placed. However, it is preferable to use a single transform where the window is part of the Radon operator. For this purpose, the proposed adjoint and forward operators for the local Radon transform are as follows

$$\hat{m}(\tau, A, B, x_0) = \sum_{x=x_0-L}^{x_0+L} d(t = \tau + A(x - x_0) + B(x - x_0)^2, x) \tag{4.3}$$

and

$$d(t, x) = \sum_{x_0} \sum_{A,B} m(\tau = t - (A(x - x_0) + B(x - x_0)^2), A, B, x) \,. \tag{4.4}$$

The number of traces placed in each window depends on the geological structures, controlled by the parameter L. The total integration aperture of the transform for each central trace x_0 is $2L + 1$. Smaller windows should be applied to complicated areas with low SNR, while larger windows should be used for relatively simple geological features. However, the increase of L will cause an increase in computation time. So the size of the window is a trade-off between cost and accuracy. These two transforms are speeded-up by discretizing x_0 at a different rate than x. As in Chapters 2 and 3, we can simply designate the forward mapping as $\mathcal{L}$ and the adjoint operator as $\mathcal{L}'$. Moreover, in other to increase computational efficiency, we can adopt a hybrid approach where the inverse problem to determine $m(\tau, A, B, x_0)$ from equation 4.3 is written in the frequency domain as follows

$$\hat{M}(\omega, A, B, x_0) = \sum_{x=x_0-L}^{x_0+L} D(\omega, x) e^{i\omega(A(x-x_0)+B(x-x_0)^2)} \tag{4.5}$$

and

$$D(\omega, x) = \sum_{x_0} \sum_{A,B} m(\omega, A, B, x_0) e^{-i\omega(A(x-x_0)+B(x-x_0)^2)} . \tag{4.6}$$

We now move the operators to time domain, in a way that expressions corresponds to mapping from $t - x$ to $\tau - A - B - x_0$ and vicervesa, yet with the cost of sums in frequency domain

$$\hat{m}(\tau, A, B, x_0) = \mathcal{F}^{-1}\hat{M}(\omega, A, B, x_0) = \mathcal{F}^{-1} \sum_{x=x_0-L}^{x_0+L} e^{i\omega(A(x-x_0)+B(x-x_0)^2)} \mathcal{F} d(t, x) \tag{4.7}$$

and

$$d(t, x) = \mathcal{F}^{-1} D(\omega, x) = \mathcal{F}^{-1} \sum_{x_0} \sum_{A,B} e^{-i\omega(A(x-x_0)+B(x-x_0)^2)} \mathcal{F} m(\tau, A, B, x_0) . \tag{4.8}$$

where $\mathcal{F}$ is the Fourier operator that maps traces in time t to frequency ω. Similarly, $\mathcal{F}^{-1}$ is the inverse Fourier operator. Equations 4.7 and 4.8 can be written in operator form as follows

$$\hat{\mathbf{m}} = \mathcal{L}'\mathbf{d} \tag{4.9}$$

and

$$\mathbf{d} = \mathcal{L}\mathbf{m} \tag{4.10}$$

where $\mathbf{m}$ is the vector containing all coefficients of $m(\tau, A, B, x_0)$ and $\mathbf{d}$ is the data vector with elements $d(t, x)$. In essence we gave packed the operators required in equations 4.7 and 4.8 into $\mathcal{L}$ and $\mathcal{L}'$. Notice that the Hermitian Fourier symmetry property for real signals is used to speed up the computation of $\mathcal{L}$ and $\mathcal{L}'$. In other words, only positive frequencies are used when evaluating the sum. Clearly, in the numerical implementation, $\mathcal{F}$ and $\mathcal{F}'$ are time-domain Discrete Fourier Transforms are executed in the implicit form via Fast Fourier Transforms.

The problem now is reduced to estimating $\mathbf{m}$ by minimizing the cost function of the problem

$$J = \|\mathbf{d} - \mathbf{L}\mathbf{m}\|_2^2 + \mu\mathcal{R}\mathbf{m} \tag{4.11}$$

where, as in Chapter 2, we will adopt the CGLS method minimize J when $\mathcal{R}$ is the l_2-norm of $\mathbf{m}$. Whereas, we will use IRLS or FISTA when solving the problem

with R given by the l_1 norm. Again, it is important to mention that the l_1-norm regularization will induce sparse (high-resolution) Radon transforms. The CGLS, IRLS and FISTA Methods are given in Appendix A.

4.3 Local Radon using travel time in a CMP gather given by Dix's hyperbola

This section discusses the kinematic meaning of the parameters A and B via a simple example. We adopt a hyperbola to describe a reflection, for instance, in the CMP domain and study how the second-order traveltime approximation approximates the hyperbola. We start with the hyperbolic traveltime for a reflection

$$T = \sqrt{\tau^2 + \frac{x^2}{v^2}} \tag{4.12}$$

where x is the offset, v is the velocity of the reflection, and τ is the two-way time. Expanding this travel time using the Taylor series in the vicinity of the arbitrary position, for example, x_0, results in the following second-order approximation

$$T(x_0 + \Delta x) = T_0 + \frac{\partial T(x_0)}{\partial x}(\Delta x) + \frac{1}{2}\frac{\partial^2 T(x_0)}{\partial x^2}(\Delta x)^2 \tag{4.13}$$

where the coefficients A and B can be calculated by taking the first and second derivative of Dix's equation precisely as

$$A = \frac{x}{v^2\sqrt{\tau^2 + \frac{x^2}{v^2}}} \tag{4.14}$$

and

$$B = (\frac{1}{2v^2\sqrt{\tau^2 + \frac{x^2}{v^2}}})(1 - \frac{Ax}{\sqrt{\tau^2 + \frac{x^2}{v^2}}})\,. \tag{4.15}$$

Figure 4.1 and Figure 4.2 show the moveout curves, the blue ones, and $T_{app}(x)$, red ones obtained by the Taylor expansion. The red curves are for $x_0 = 0$m and $x_0 = 100$m. Figure 4.1(a) and Figure 4.2(a) display the approximated trajectory only using the first term of Taylor expansion, while 4.1(b) and 4.2(b) are the approximated curves based on the two-term expansion. As they show, the approximated

curves do not match the traveltime very well if constructed with only one term. However, $T_{app}(x)$ and the true traveltime overlap reasonably for a small window, $(x_0 - \Delta x \leq x \leq x_0 + \Delta x)$ using two-term approximation.

Figure 4.3 shows different plots for coefficients A and B calculated using equations (4.14) and (4.15) for different velocities and $\tau = 1.2$s. As shown for relatively small offsets, the coefficient A (the slope) is small, increasing for far offsets. The value of B or the curvature is maximum for near offsets, decreasing for far offsets.

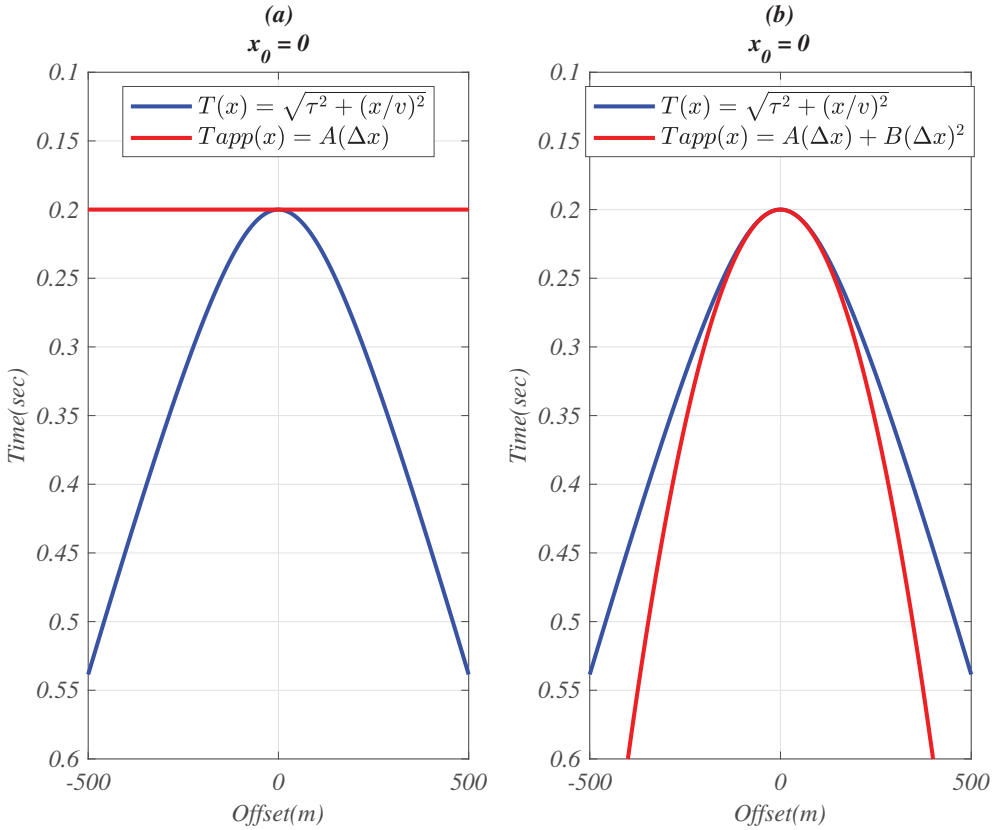

Figure 4.1: (a) Traveltime, the blue curve and the first term approximated Taylor expansion, the red line for $x_0 = 0$m, (b) Traveltime, the blue curve and the second terms approximated Taylor expansion, the red line, for $x_0 = 0$m.

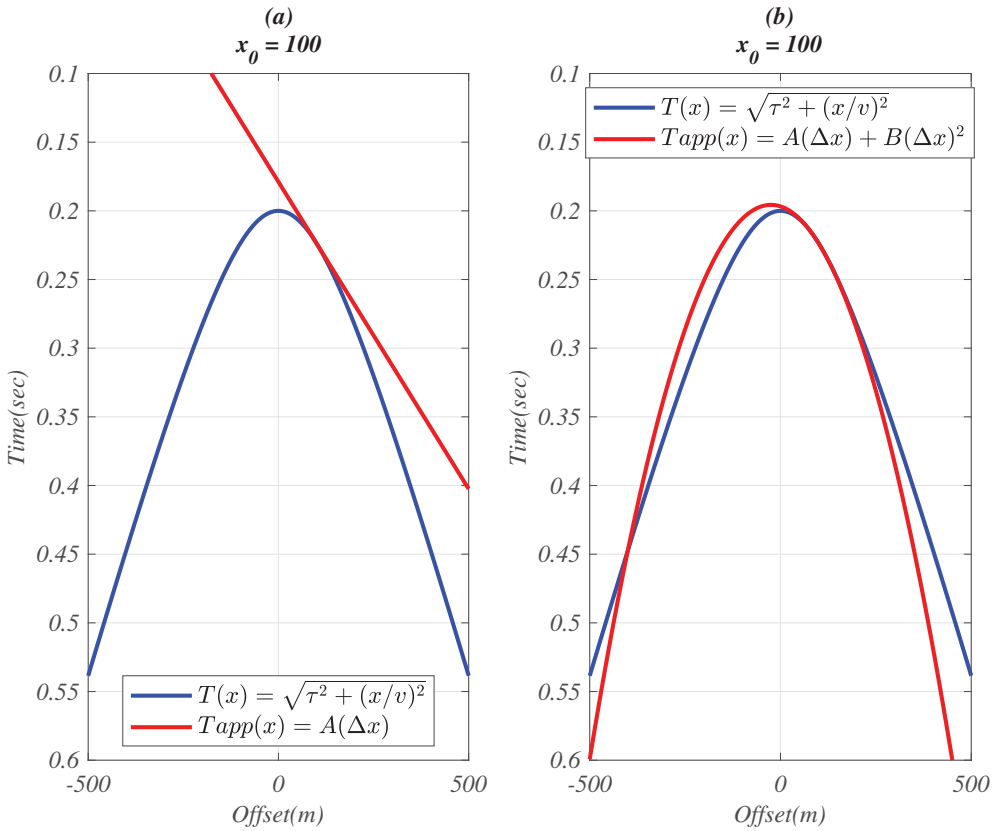

Figure 4.2: (a) Traveltime, the blue curve and the first term approximated Taylor expansion, the red line, for $x_0 = 100$m, (b) Traveltime, the blue curve and the second terms approximated Taylor expansion, the red line, for $x_0 = 100$m.

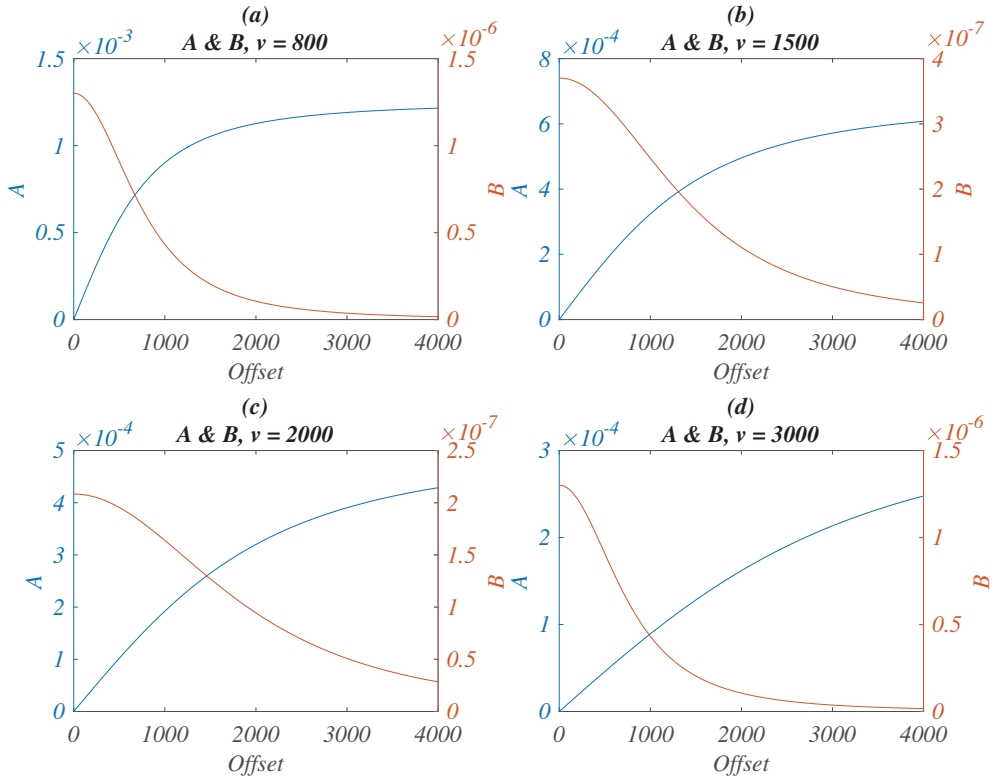

Figure 4.3: Coefficients A and B for different velocities and offsets, $t_0 = 1.2$ ms.

4.4 Synthetic examples

We examine the local Radon transform with a numerically synthetic dataset to show the proposed local transform. To obtain the sparse model in the local Radon domain, we used the FISTA and IRLS algorithms. The synthetic dataset corresponds to 91 traces or 251 samples each. The window length is 325 meters, and the interval of x_0 is 157 meters. The wavelet used in the data set is Ricker 25 Hz.

Figure 4.4 shows the results obtained via FISTA algorithm after 55 iterations. The figures consist of four panels. The left panel (a) is the free noise data, (b) is observed data with $SNR = 2$, panel (c) is predicted data (reconstructed from the model), and (d) is the difference between observed and reconstructed data. Figure 4.5 shows six cube models computed for six different x_0 by the FISTA algorithm.

Figure 4.6 shows the results obtained via the IRLS algorithm using $\mu = 5$ as a trade-off parameter and $SNR = 2$ after ten iterations and consists of four panels: the left panel (a) is the free noise data, (b) is observed data, panel (c) is predicted data (reconstructed from the model), and (d) is the difference between observed and reconstructed data. Figure 4.7 shows different reconstructed data with varying trade-offs μ and the difference between the reconstructed data and original data as an error. To show the enhancement of the data using the local Radon algorithm, we compare the input and output SNR. Figure 4.8 to Figure 4.12 show the input data for $SNR = -5, -2, 0, 2, 4$, respectively. Each figure has four panels that are a) the noise-free data, b) the noisy data, c) the reconstructed data, and d) is the error panel.

Figure 4.13 shows the output SNR (reconstruction quality) versus the input SNR for the previous figures. This experiments portray the denoising capability of the local Radon transform. In general, the method is sensitive to the selection of the trade-off parameter μ and monitoring the residuals is necessary to obtain an optimal trade-off parameter for proper data denoising.

Besides the enhancement, the local Radon transformation can be used for the reconstruction of missing data. In this case, we run the algorithm with a sampling operator. Once we estimated the local Radon transform from the sampled data, one can synthesize all the data (missing and given observation) and obtain the reconstructed data. Figure 4.14 is consists of four panels. Panel a) is the synthetic data

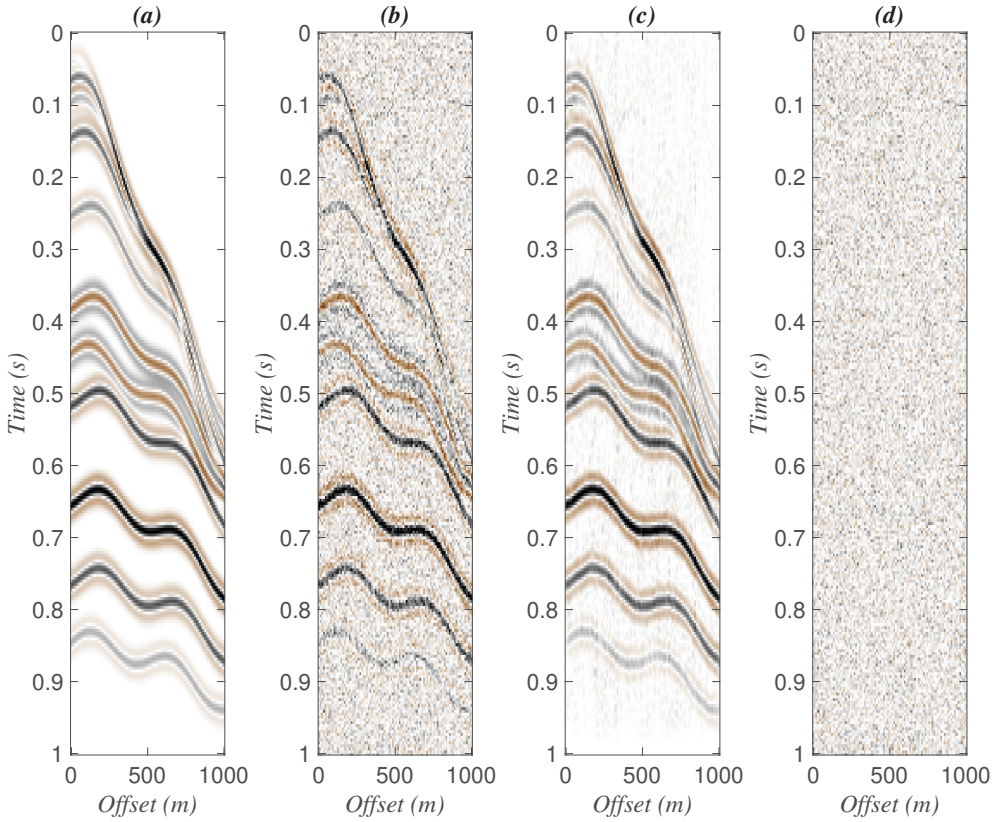

Figure 4.4: Estimated results by the FISTA algorithm after 55 iterations. (a) The free noise data. (b) The observed data. (c) Predicted data (reconstructed from the Radon domain). (d) The difference between the observed and reconstructed data.

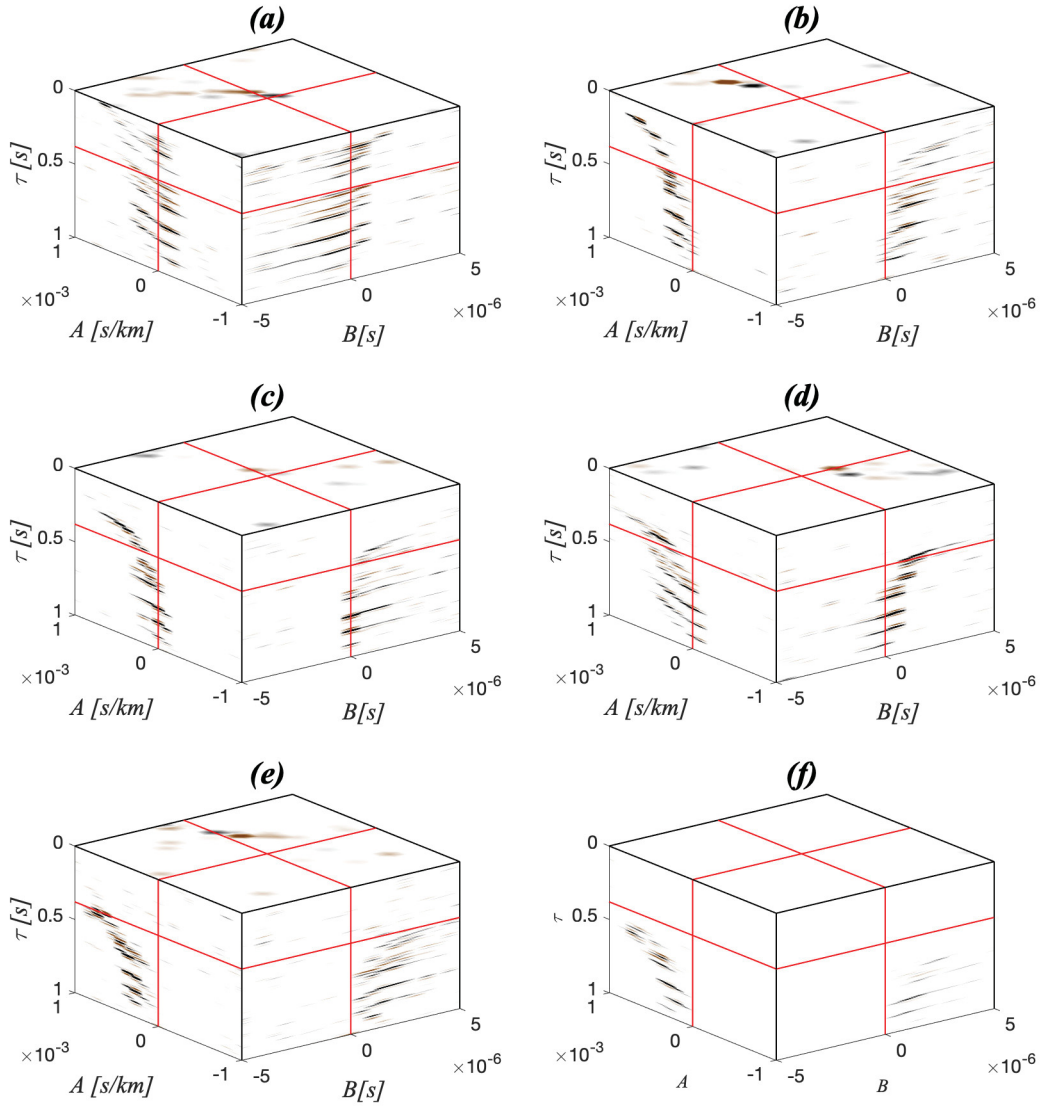

Figure 4.5: Cube models computed for six different x_0s. These results were obtained by inverting the Radon coefficients via the FISTA algorithm.

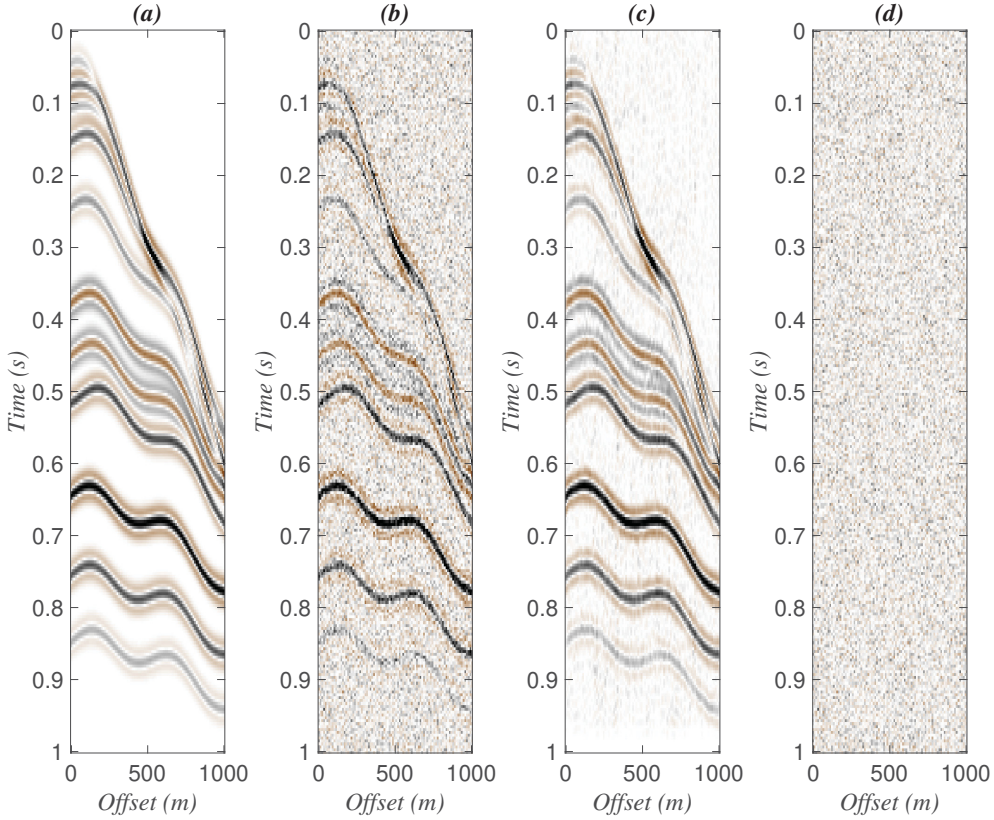

Figure 4.6: Computed results via the IRLS algorithm using $\mu = 5$ after ten iterations. (a) The free noise data. (b) The observed data. (c) Predictied data (reconstructed from the model). (d) The difference between the observed and reconstructed data.

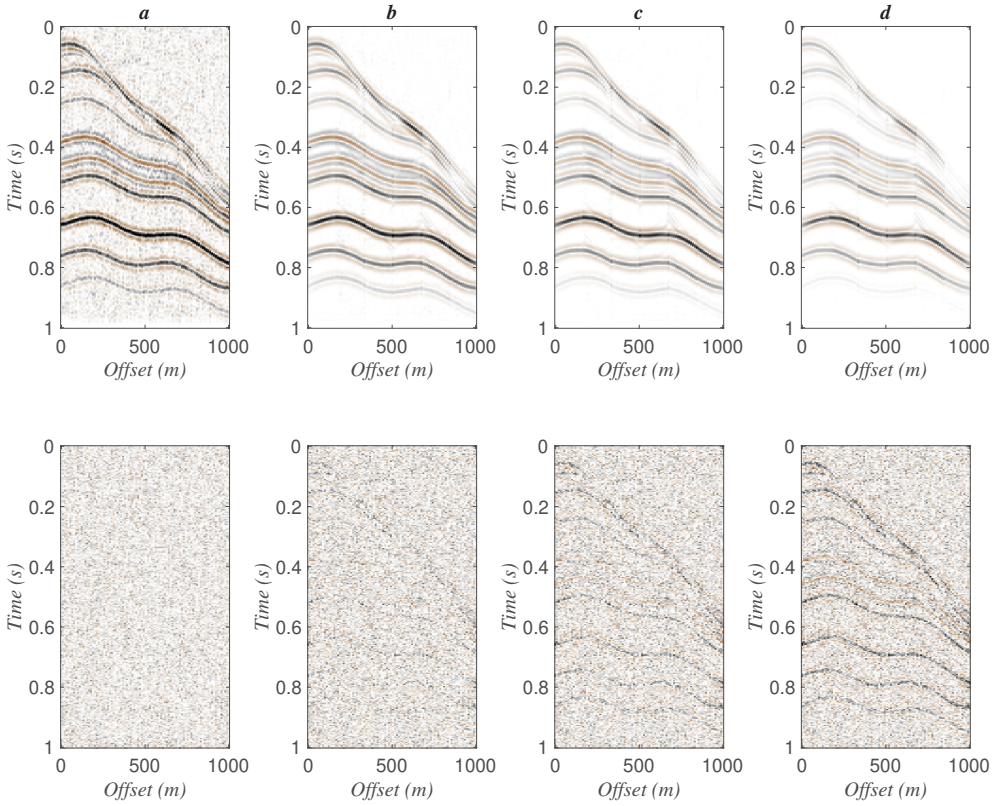

Figure 4.7: Computed results by the IRLS algorithm using different values of μ after ten iterations. In this experiment the first row is the predicted data from the Radon model. The second row is the estimated noise. From left to right, the tradeoff parameter from small to large ($\mu = 0.1, 1, 10, 100$). Excessive overfitting is observed in a). Excessive under-fitting is observed in d).

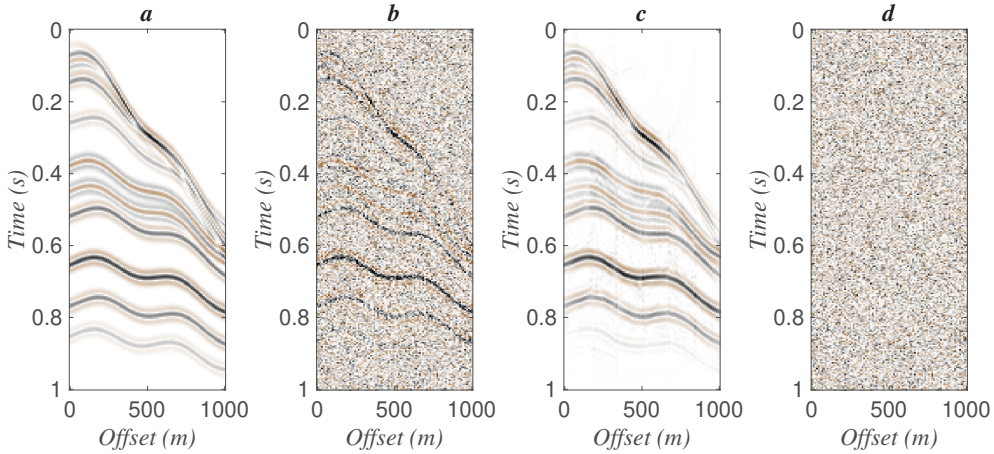

Figure 4.8: (a) The noise free data. (b) Noisy data $SNR = -5$, (c) Reconstructed data. (d) The error.

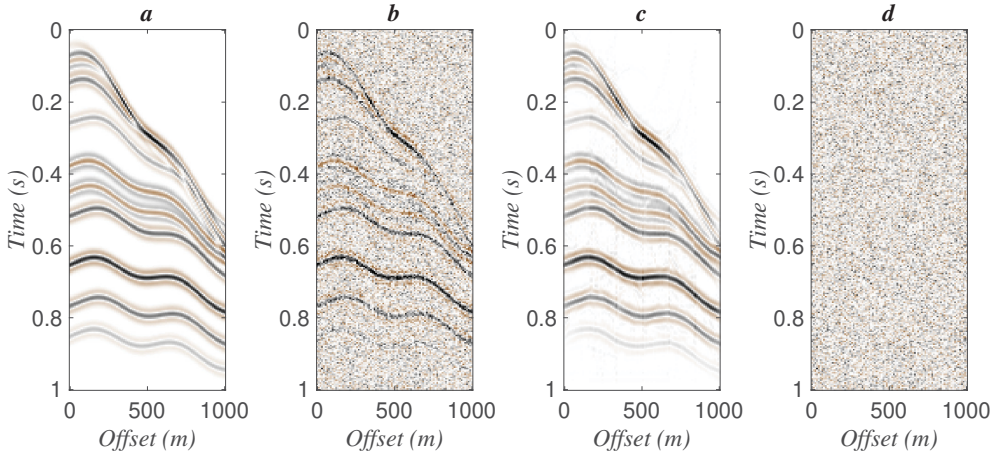

Figure 4.9: (a) The noise free data. (b) Noisy data $SNR = -2$. (c) Reconstructed data. (d) The error.

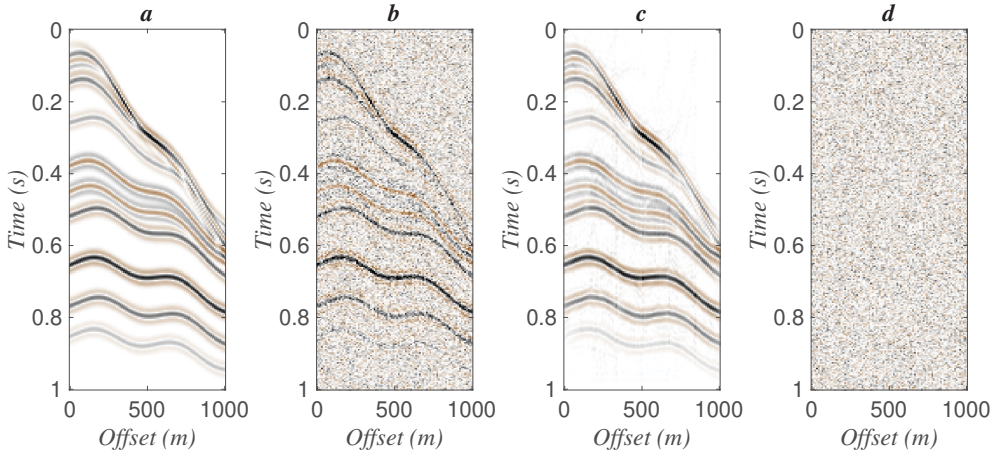

Figure 4.10: (a) The noise free data. (b) Noisy data $SNR = 0$. (c) Reconstructed data. (d) The error.

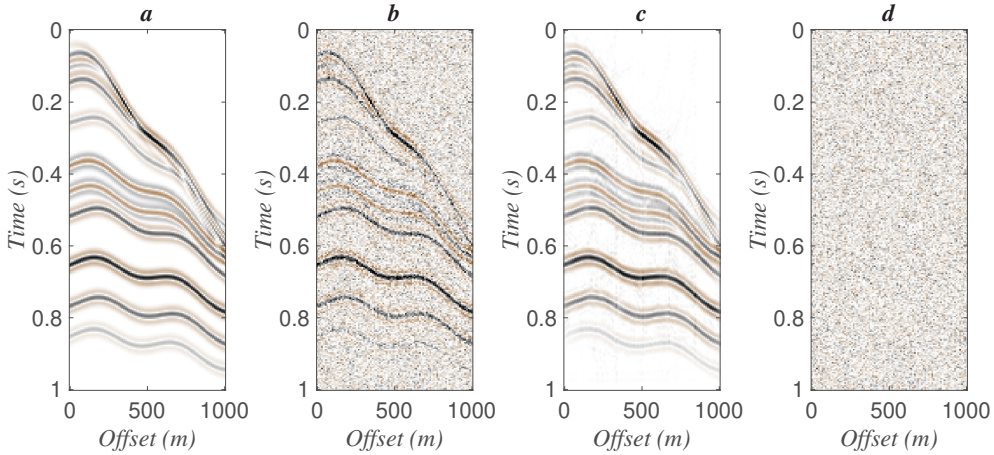

Figure 4.11: (a) The noise free data. (b) Data with additive noise $SNR = 2$. (c) Reconstructed data. (d) The error panel.

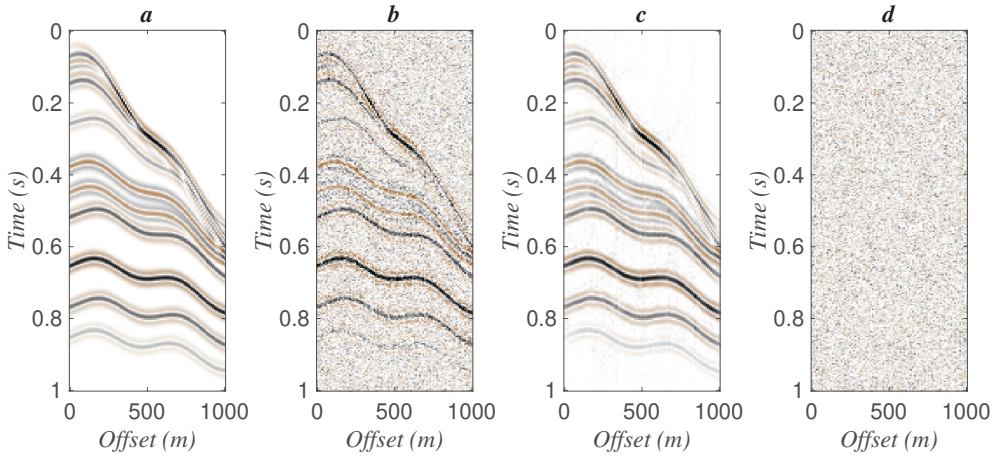

Figure 4.12: (a) The noise free data. (b) Noisy data $SNR = 4$. (c) Reconstructed data. (d) The error.

set with $SNR = 15$, b) is the decimated data from panel a), panel c) shows the interpolated data using local Radon transformation, and panel d) is the difference between Figure 4.14a and Figure 4.14 b. This example is challenging for interpolation because dips vary quickly across offset. Despite the latter, the Local Radon transform was able to recover the missing observations.

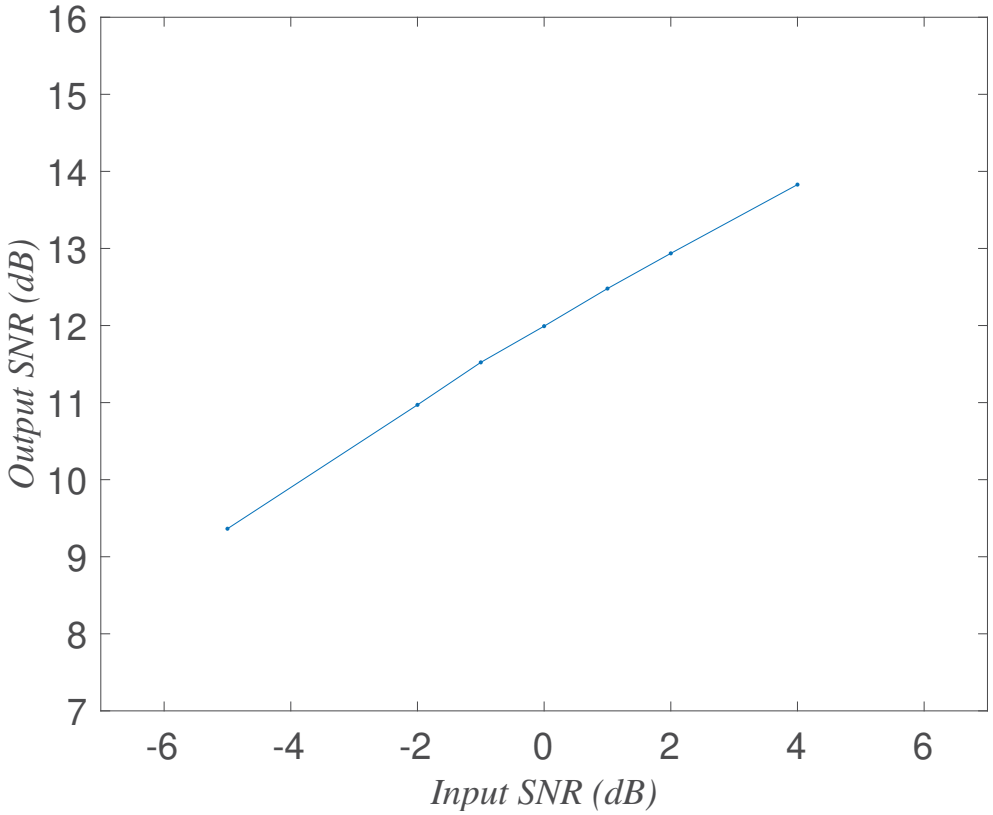

Figure 4.13: The graph shows the output SNR (quality of the reconstitution) versus the SNR of the input data for the experiments.

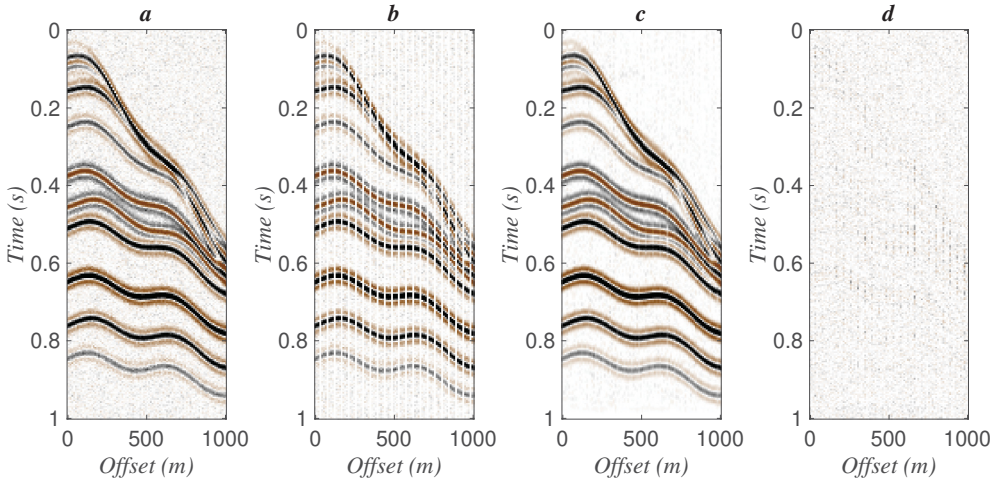

Figure 4.14: (a) Synthetic data $SNR = 15$. (b) Decimated data. (c) Interpolated data using local Radon transformation. (d) The difference between (a) and (b).

4.5 Real data example

This section tests the local Radon transformation on a real dataset (Gulf of Mexico, Mississippi Canyon data). The data corresponds to a near offset section.

Figure 4.15 shows three panels which are (a) the input data, (b) is the reconstructed data using local Radon transform from the decomposed data and (c) is the difference between input and reconstructed, (a) and (b). This experiment shows that the proposed local Radon transform can synthetize data composed of complex laterally varying dips.

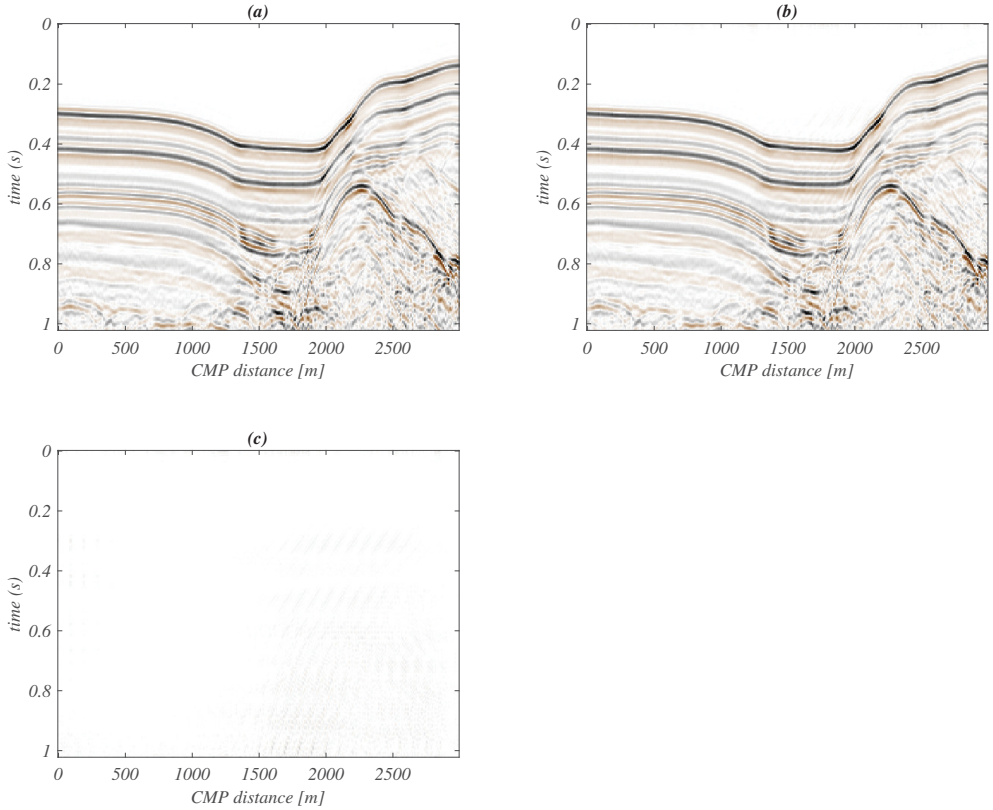

Figure 4.15: (a) Real dataset, a near offset section from the Mississippi Canyon survey. (b) The reconstructed data using local Radon transform. (c) The difference between panels (a) and (b).

Figure 4.16 shows the reconstruction of missing data for the above real data using the local Radon transform. Figure 4.16a is the input data, Figure 4.16b is the decimated data, 4.16c is the reconstructed data using local Radon transform from the decomposed data and panel (d) is the difference between the input (a) and reconstructed (c). Figure 4.17 repeats the experiment but for random sampling. The interpolation experiments show that the local Radon transform could also be used for compressive acquisition where data are regularly or randomly subsampled for saving acquisition costs.

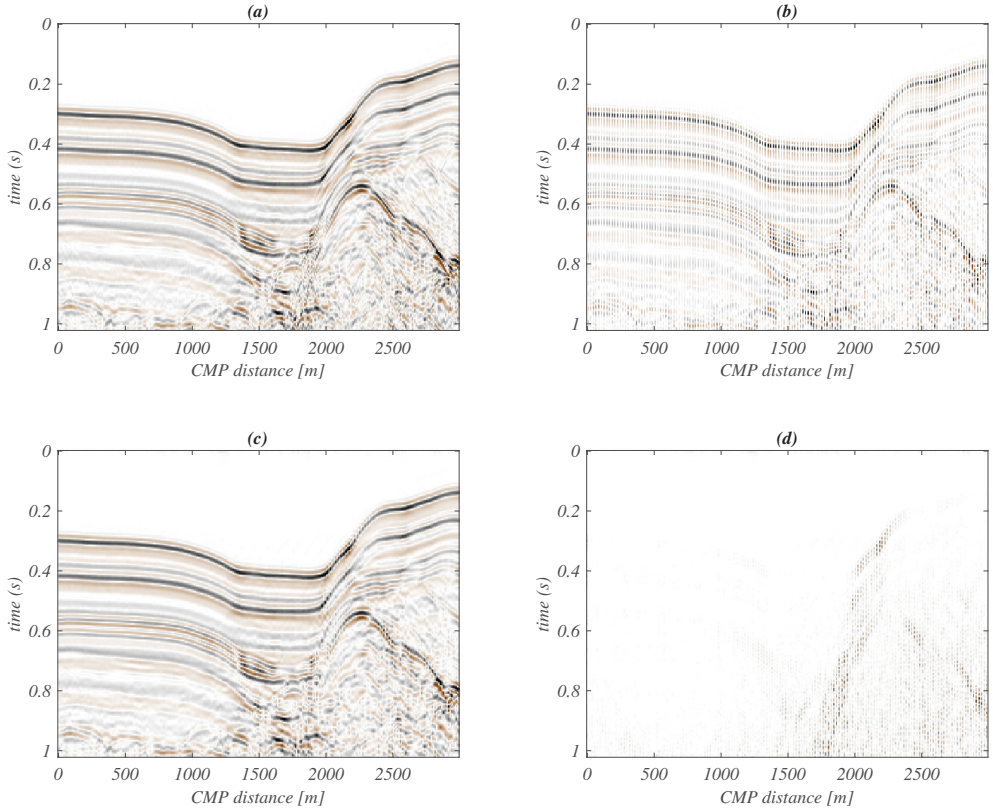

Figure 4.16: (a) A near offset marine dataset corresponding to a survey in the Gulf of Mexico (Mississippi Canyon data). (b) Regularly decimated data to test interpolation. c) Interpolated data via the proposed local Radon Transform. d) Difference panel corresponding to the true data in a) minus the interpolated data in c). As one can see, only steep dips have been poorly reconstructed.

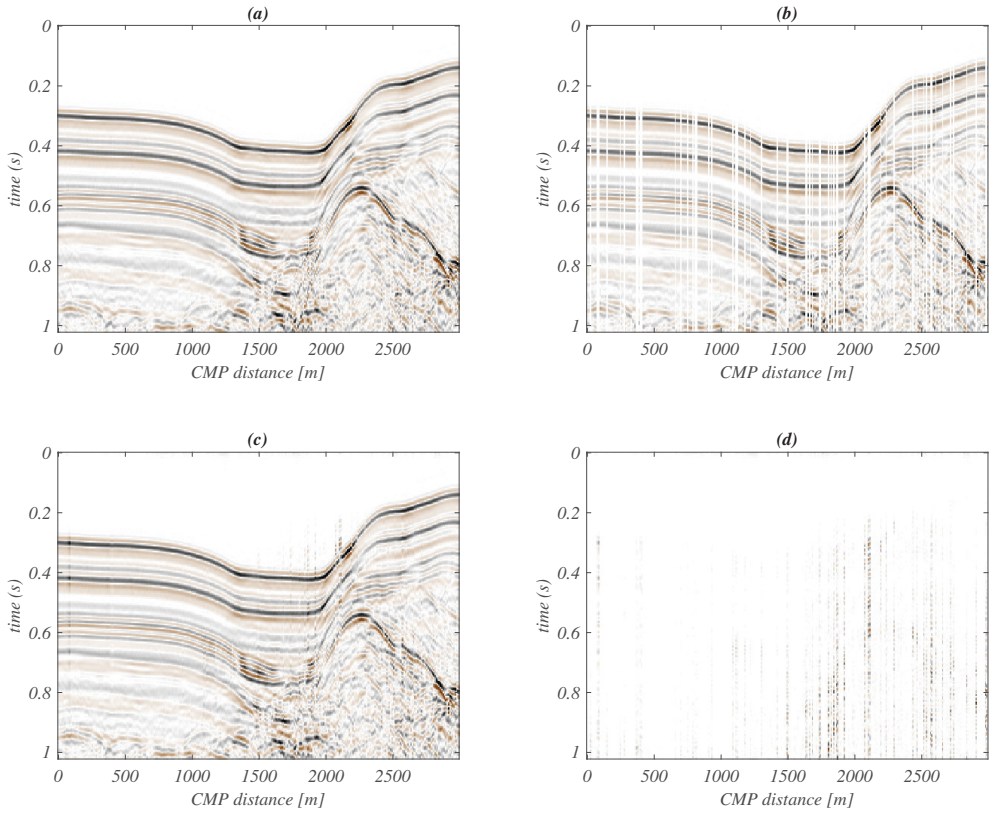

Figure 4.17: (a) A near offset marine dataset corresponding to a survey in the Gulf of Mexico (Mississippi Canyon data). (b) Randomly decimated data to test interpolation about 80% of traces are missing. c) Interpolated data via the proposed local Radon Transform. d) Difference panel corresponding to the true data in a) minus the interpolated data in c). As one can see, only steep dips have been poorly interpolated.

CHAPTER 5

Conclusions

I have reviewed different Radon transforms used in exploration seismology and paid particular attention to understanding solvers for estimating sparse Radon transforms (also called high-resolution Radon transforms). These solvers include CGLS, FISTA and IRLS, and, as other authors, I have shown that FISTA and IRLS provide a solution that has more resolution than classical Radon transforms solved via damped-least squares and CGLS. I also studied different operators that can be categorized as explicit and implicit. For practical purposes, I preferred to concentrate on implicit form operators where operators replace matrices, and iterative solvers permit to estimate the Radon transform. In general, Parabolic and Hyperbolic Radon transforms are potential tools to process seismic data in the CMP domain. To process data in other domains or structurally complicated data sets, one needs to develop a local Radon transform where the operator adapts to spatially varying dips. My main contribution is the development of a local Radon transform that adopts a second-order traveltime approximation. This new transform was inspired in the work mainly presented in Chapter 3 on implicit operators for Parabolic, and Hyperbolic Radon transforms, where I showed how to use different solvers to map data from $t - x$ to $\tau - q$ or $\tau - v$. These operators were modified to work locally with the second-order travel time approximation that adapts to different reflection traveltime signatures, in general, with applications not restricted to CMP gathers. This flexible local Radon transform offers a new operator for seismic data denoising and seismic data reconstruction. However, more work is required to move the current research to process data from 3D surveys where the data depends on two or

more spatial dimensions. At the current stage, the operators I have developed are too time-consuming for porting them to processing 3D surveys, and more research is needed to find even more efficient ways to compute local Radon transforms for dataset of the form $d(t, x, y)$ which represent gathers of a typical modern 3D survey.

As the main conclusion for this work, I developed a 2D local Radon transform that is very versatile in terms of applicability. I have also found that adopting FISTA or IRLS solvers for computing the sparse Radon transform produces similar results with no real computational difference between these two methods. In essence, the main block for advancing this research is to find fast ways of implementing forward and adjoint operators $\mathcal{L}$ and $\mathcal{L}'$. At this stage, I conclude that the hybrid implementation of the local Radon transform in $t - x$ and $\omega - x$ provides a more computationally efficient way of computing the transform than using an approach fully developed in the time domain as in Chapter 3.

Future work should not only concentrate on finding more efficient algorithms but also on in-depth studies of local Radon transforms for application such as simultaneous source processing and multidimensional compressive reconstructions.

Bibliography

Abedi, M. M., and A. Stovas, 2019, Extended generalized non-hyperbolic moveout approximation: Geophysical Journal International, **216**, 1428–1440.

Bakulin, A., I. Silvestrov, M. Dmitriev, D. Neklyudov, M. Protasov, K. Gadylshin, and V. Dolgov, 2020, Nonlinear beamforming for enhancement of 3D prestack land seismic data: Geophysics, **85**, V283–V296.

Beck, A., and M. Teboulle, 2009, A fast iterative shrinkage-thresholding algorithm for linear inverse problems: SIAM journal on imaging sciences, **2**, 183–202.

Beylkin, G., 1987, Discrete Radon transform: IEEE Transactions on Acoustics, Speech, and Signal Processing, **35**, 162–172.

Bracewell, R., 1978, The Fourier transform and its applications, second ed.: McGraw-Hill Kogakusha, Ltd.

Brown, A. R., 2011, Interpretation of Three-Dimensional Seismic Data: American Association of Petroleum Geologists.

Cary, P. W., 1998, The simplest discrete Radon transform, *in* SEG Technical Program Expanded Abstracts 1998: Society of Exploration Geophysicists, 1999–2002.

Claerbout, J. F., 1985, Imaging the Earth's interior: Blackwell Scientific Publications Oxford.

———, 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications London.

Corning, P., R. Goodrum, N. Blais, J.-L. Pennacchioni, and H. Brysk, 1987, Multiple removal in the tau-p domain: OTC Offshore Technology Conference.

Darche, G., 1990, Spatial interpolation using a fast parabolic transform: 60th Annual International Meeting, SEG, Expanded Abstracts, 1647–1650.

Daubechies, I., R. DeVore, M. Fornasier, and C. S. Güntürk, 2010, Iteratively reweighted least squares minimization for sparse recovery: Communications on

Pure and Applied Mathematics, **63**, 1–38.

Deans, S. R., 1983, The Radon transform and some of its applications: Wiley-Blackwell.

Durrani, T., and D. Bisset, 1984, The Radon transform and its properties: Geophysics, **49**, 1180–1187.

Fowler, C. M. R., 1990, The solid Earth: an introduction to global geophysics: Cambridge University Press.

Gholami, A., and M. Farshad, 2019, The Shuey-Radon transform: Geophysics, **84**, V197–V206.

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

Gray, S. H., J. Etgen, J. Dellinger, and D. Whitmore, 2001, Seismic migration problems and solutions: Geophysics, **66**, 1622–1640.

Hampson, D., 1986a, Inverse velocity stacking for multiple elimination: SEG Technical Program Expanded Abstracts 1986, 422–424.

———, 1986b, Inverse velocity stacking for multiple elimination: Canadian Journal of Exploration Geophysics, **22**, 44–55.

Herrmann, P., T. Mojesky, M. Magesan, and P. Hugonnet, 2000, De-aliased, high-resolution Radon transforms: SEG Technical Program Expanded Abstracts 2000, 1953–1956.

Hestenes, M. R., E. Stiefel, et al., 1952, Methods of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards, **49**.

Ibrahim, A., and M. D. Sacchi, 2014, Simultaneous source separation using a robust Radon transform: Geophysics, **79**, V1–V11.

Kabir, M. N., and D. Verschuur, 1995, Restoration of missing offsets by parabolic Radon transform: Geophysical Prospecting, **43**, 347–368.

Kazemi, N., and M. D. Sacchi, 2021, Offset-extended sparse Radon transform: application to multiple suppression in the presence of AVO: Geophysics, **86**, 1–41.

Kostov, C., 1990, Toeplitz structure in slant-stack inversion: SEG Technical Program Expanded Abstracts 1990, 1618–1621.

Li, Q., 2001, High-resolution hyperbolic Radon transform multiple removal: MSc thesis, University of Alberta.

Lin, R., and M. D. Sacchi, 2020, Separation of simultaneous sources acquired with a high blending factor via coherence pass robust radon operators: Geophysics, **85**, V269–V282.

Mehta, K., M. Nabighian, and Y. Li, 2005, Controlled source electromagnetic (csem) technique for detection and delineation of hydrocarbon reservoirs: an evaluation: SEG Technical Program Expanded Abstracts 2005, 546–549.

Menke, W., 1989, Geophysical data analysis: discrete inverse theory, second ed.: Academic Press.

Ravve, I., and Z. Koren, 2017, Traveltime approximation in vertical transversely isotropic layered media: Geophysical Prospecting, **65**, 1559–1581.

Sabbione, J. I., and M. D. Sacchi, 2017, Attenuating multiples with the restricted domain hyperbolic Radon transform: 15th International Congress of the Brazilian Geophysical Society and EXPOGEF, Rio de Janeiro, 603–608.

Sacchi, M. D., and M. Porsani, 2005, Fast high resolution parabolic Radon transform: SEG Technical Program Expanded Abstracts 1999, 1477–1480.

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

Sava, P., and S. J. Hill, 2009, Overview and classification of wavefield seismic imaging methods: The Leading Edge, **28**, 170–183.

Scales, J. A., A. Gersztenkorn, and S. Treitel, 1988, Fast Lp solution of large, sparse, linear systems: Application to seismic travel time tomography: Journal of Computational Physics, **75**, 314–333.

Schonewille, M. A., and A. J. W. Duijndam, 2001, Parabolic Radon transform, sampling and efficiency: Geophysics, **66**, 667–678.

Stovas, A., and S. Fomel, 2017, The modified generalized moveout approximation: A new parameter selection: Geophysical Prospecting, **65**, 687–695.

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

Thorson, J. R., and J. F. Claerbout, 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, **50**, 2727–2741.

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

Treitel, S., P. R. Gutowski, and D. E. Wagner, 1982, Plane-wave decomposition of seismograms: Geophysics, **47**, 1375–1401.

Turner, G., 1990, Aliasing in the tau-p transform and the removal of spatially aliased coherent noise: Geophysics, **55**, 1496–1503.

Vermeer, G., 1990, Seismic wavefield sampling: Society of Exploration Geophysicists.

Vermeer, G. J., 2002, 3-D seismic survey design: Society of Exploration Geophysicists.

Verschuur, J., 1999, Multiple removal results from Delft University: The Leading Edge, **18**, 86–91.

Wang, B., M. Sacchi, and X. Yin, 2011, Avo-preserving sparse parabolic Fourier transform: 73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011, cp–238.

Yilmaz, Ö., 1989, Velocity stack processing: Geophysical Prospecting, **34**, 357 – 382.

———, 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: Society of exploration geophysicists.

Zhou, B., and S. A. Greenhalgh, 1994, Linear and parabolic $\tau - p$ transforms revisited: Geophysics, **59**, 1133–1149.

APPENDIX A

Solvers

A.1 Conjugate Gradient Least-squares Method (CGLS)

The CG algorithm (Hestenes et al., 1952) minimizes the cost function

$$J = \|\mathbf{Lm} - \mathbf{d}\|_2^2 + \mu\|\mathbf{d}\|_2^2 \tag{A.1}$$

which is equivalent to solving the system of equations

$$(\mathbf{L}^T\mathbf{L} + \mu\mathbf{I})\mathbf{m} = \mathbf{g}\,. \tag{A.2}$$

If we let $\mathbf{A} = \mathbf{L}^T\mathbf{L} + \mu\mathbf{I}$ and $\mathbf{g} = \mathbf{L}^T\mathbf{d}$, one gets

$$\mathbf{Am} = \mathbf{g}\,. \tag{A.3}$$

The symmetric matrix $\mathbf{A}^T = \mathbf{A}$ is positive-definitive ($\mathbf{x}^T\mathbf{Ax} > 0$) and of size $M \times M$. The classical CG algorithm solves A.3 in M steps. The multiplication of $\mathbf{A}$ times a vector is executed in each step. If the algorithm converges in $K < N$ iterations, then its cost is about $\mathcal{O}(N^2K)$. Moreover, if we have a way of implementing fast matrix-times vector multiplications by exploiting a special structure[1] of $\mathbf{A}$, the cost $\mathcal{O}(N^2K)$ further decreases.

CGLS is a modification of the CG method where in its steps, multiplication of $\mathbf{A}$

[1]For instance, if the matrix has a Toeplitz structure, products of the form $\mathbf{Au}$, can be executed via the Fast Fourier Transform (Sacchi and Porsani, 2005).

times a vector are avoided. Each step directly requires the application of matrix-time-vector products of matrices $\mathbf{L}$ and $\mathbf{L}^T$ (Golub and Van Loan, 1996). Algorithm 4 shows the classical CGLS algorithm, the matrix $\mathbf{L}$ is replaced by the linear operator $\mathcal{L}$ and $\mathbf{L}^T$ by $\mathcal{L}'$. In Chapter 3 $\mathcal{L}$ and $\mathcal{L}'$ are time-domain Radon operators. In Chapter 4, they are time-domain local Radon operators but partially implemented in the frequency domain.

Algorithm 4 CGLS

Given `Forward` $\mathcal{L}$ and `Adjoint` $\mathcal{L}'$ operators and initial point $\mathbf{m}_0$,
this algorithm minimizes $J = \|\mathcal{L}\,\mathbf{m} - \mathbf{d}\|_2^2 + \mu\|\mathbf{m}\|_2^2$
$\mathbf{m} = \mathbf{m_0} \quad \& \quad k = 0$
$\mathbf{r} = \mathbf{b} - \mathcal{L}\,\mathbf{m}$ `Forward`
$\mathbf{s} = \mathcal{L}'\mathbf{r} - \mu\mathbf{m}$ `Adjoint`
$\mathbf{p} = \mathbf{s}$
$\gamma = \|\mathbf{s}\|_2^2$
for $k < k_{max}$ **do**
  $\mathbf{q} = \mathcal{L}\,\mathbf{p}$ `Forward`
  $\delta = \|\mathbf{q}\|_2^2 + \mu\|\mathbf{p}\|_2^2$
  $\alpha = \gamma/\delta$
  $\mathbf{m} = \mathbf{m} + \alpha\mathbf{p}$
  $\mathbf{r} = \mathbf{r} - \alpha\mathbf{q}$
  $\mathbf{s} = \mathcal{L}'\,\mathbf{r} - \mu\mathbf{m}$ `Adjoint`
  $\gamma_{new} = \|\mathbf{s}\|_2^2$
  $\beta = \gamma_{new}/\gamma$
  $\gamma = \gamma_{new}$
  $\mathbf{p} = \mathbf{s} + \beta\mathbf{p}$
  $k = k + 1$
end for

A.2 Iterative Reweighted least-squares (IRLS)

The IRLS (Scales et al., 1988; Daubechies et al., 2010) algorithm finds sparse solutions of the Radon transform by minimizing the cost function

$$J = \|\mathbf{Lm} - \mathbf{d}\|_2^2 + \mu\|\mathbf{m}\|_1 \,. \tag{A.4}$$

The cost function J can be replaced by the following quadratic form

$$J = \|\mathbf{Lm} - \mathbf{d}\|_2^2 + \mu\|\mathbf{Wm}\|_2^2 \,, \tag{A.5}$$

where

$$\|\mathbf{m}\|_1 \approx \|\mathbf{Wm}\|_2^2 \tag{A.6}$$

with $\mathbf{W}$ a diagonal matrix with elements given by $W_{i,i} = \frac{1}{|m_i|+\epsilon}$ and ϵ is a mall number to avoid division by zero.The IRLS solution is expressed by

$$\mathbf{m}^k = \underset{\mathbf{m}}{\operatorname{argmin}} \|\mathbf{Lm} - \mathbf{d}\|_2^2 + \mu\|\mathbf{W}^{k-1}\mathbf{m}\|_2^2 \,, \tag{A.7}$$

where k is the IRLS iteration and $W_{i,i}^{k-1} = \frac{1}{|m_i^{k-1}|+\epsilon}$. We let $\mathbf{Wm} = \mathbf{u}$ and hence, $\mathbf{Pu} = \mathbf{m}$, with $P_{i,i} = |m_i| + \epsilon$, then

$$\mathbf{u}^k = \underset{\mathbf{u}}{\operatorname{argmin}} \|\mathbf{LP}^{k-1}\mathbf{u} - \mathbf{d}\|_2^2 + \mu\|\mathbf{u}\|_2^2 \,, \tag{A.8}$$

where the last expression is equivalent to A.1. Hence, each quadratic problem A.8 is solved via the CGLS algorithm described in the proceeding section but now the operator $\mathcal{L}$ in Algorithm 4 must be replaced by $\mathcal{L}\mathbf{P}$. In this way, we have an external IRLS iteration and the internal iteration of the CGLS method. In the IRLS implementation used in this thesis, the internal CGLS iteration is set to a maximum of 50 iterations and the norm of the gradient of J is used as a stopping criterion. The external IRLS iteration is fixed to a maximum of 5 iterations unless convergence is achieved in fewer iterations. A pseudo-code for IRLS is given in Algorithm 5.

Algorithm 5 IRLS

Given `Forward` $\mathcal{L}$ and `Adjoint` $\mathcal{L}'$ operators and initial point $\mathbf{m}_0$,
this algorithm minimizes $J = \|\mathcal{L}\,\mathbf{m} - \mathbf{d}\|_2^2 + \mu\|\mathbf{m}\|_1$
$\mathbf{m} = \mathbf{m_0}$ & $k = 0$
Initialize $\mathbf{P}$
for $k < k_{max}$ **do**
  use `cgls` to solve $\mathbf{u} = \text{argmin}_{\mathbf{u}}\, \|\mathcal{L}\mathbf{P}\mathbf{u} - \mathbf{d}\|_2^2 + \mu\|\mathbf{u}\|_2^2$
  $\mathbf{m} = \mathbf{P}\mathbf{u}$
  update $\mathbf{P}$
  $k = k + 1$
end for

A.3 Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)

FISTA is also used to minimize A.4 (Beck and Teboulle, 2009). The algorithm derives from the Iterative Shrinkage-Thresholding Algorithm (ISTA) (Daubechies et al., 2010). FISTA is basically ISTA in conjunction with Nesterov's accelerated gradient descent (Beck and Teboulle, 2009). The algorithm FISTA implemented in this thesis follows Beck and Teboulle (2009) with step length computed via the power-iteration method (Golub and Van Loan, 1996) to guarantee its convergence. Algorithm 6 is the pseudo-code for FISTA. The operator $\mathcal{S}_T$ in the pseudo-code is the shrinkage operator defined as

$$\mathcal{S}_T(x) = (|x| - T)_+ sign(x)\,. \tag{A.9}$$

Algorithm 6 FISTA

Given `Forward` $\mathcal{L}$ and `Adjoint` $\mathcal{L}'$ operators and initial point $\mathbf{m}_0$,
this algorithm minimizes $J = \|\mathcal{L}\,\mathbf{m} - \mathbf{d}\|_2^2 + \mu\|\mathbf{m}\|_1$
$\mathbf{m} = \mathbf{m_0}$ & $k = 0$
$\mathbf{u} = \mathbf{m}$
use `power iteration` to compute α =`max eigenvalue of` $\mathcal{L}'\mathcal{L}$
$T = \mu/(2\alpha)$
for $k < k_{max}$ **do**
    $\mathbf{m}_{old} = \mathbf{m}$
    $\mathbf{v} = \mathcal{L}\mathbf{u}$
    $\mathbf{m} = \mathcal{S}_T[\mathbf{u} + (1/\alpha)\mathcal{L}'(\mathbf{d} - \mathbf{v})]$
    $t' = t$
    $t = (1 + \sqrt{1 + 4t^2})/2$
    $\mathbf{u} = \mathbf{m} + [(t' - 1)/t](\mathbf{m} - \mathbf{m}_{old})$
end for