University of Alberta

Sparsity and Group Sparsity Constrained Inversion for Spectral Decomposition of Seismic Data

by

David Bonar

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

Master of Science
in
Geophysics

Department of Physics

Spring 2012
Edmonton, Alberta

Examining Committee

Mauricio Sacchi, Physics

Doug Schmitt, Physics

Jeff Gu, Physics

Csaba Szepesvari, Computing Science

Abstract

Local time-frequency analysis, also known as spectral decomposition, allows for a more detailed interpretation of time-series by providing the evolution of the frequency spectrum through time and has proven to be a useful seismic attribute for exercises such as reservoir characterization. This thesis explores posing the spectral decomposition problem through sparsity promoting inversion techniques to obtain a high resolution local time-frequency representation for a seismic trace. By requiring a high resolution local time-frequency representation for each individual seismic trace, increased noise variability is obtained between the local time-frequency representations of neighbouring seismic traces. To help attenuate this noise, information from nearby seismic traces can be incorporated during the inversion process for the spectral decomposition of an individual seismic trace. A similar strategy, called group sparsity, can also be incorporated for the simultaneous denoising of multicomponent seismic traces. A new method for the noise attenuation of seismic data is presented as well.

Acknowledgements

First and foremost, I thank my supervisor, Dr. Mauricio Sacchi. His ideas on how to attack problems and knack for simplifying even the most complicated concepts has played an integral role during my thesis work and as helped form how I, myself, view many scientific problems. In my MSc. studies, I had the privilege of being a member of the SAIG consortium and would like to thank my supervisor for forming this group as well as the consortium sponsors for providing financial support. During my time as a graduate student at the University of Alberta I have met and shared numerous ideas with many other fellow students and would like to thank them all. In particular, I would like to thank Ismael Vera Rodriguez for taking an interest in my work for de-noising and initiating our collaborative efforts. I also thank my examining committee (Dr. Doug Schmitt, Dr. Jeff Gu, and Dr. Csaba Szepesvari) for taking the time out of their schedules to critically read my thesis and for providing helpful suggestions on how to improve upon it. Finally, I thank my family for their continued support in all of my endeavours.

Contents

List of Tables

List of Figures

List of Symbols

Symbols are listed with the page number of their first appearance. Vectors and matrices are denoted in bold font if their variables are not explicitly defined.

ψ	mother wavelet of the Continuous Wavelet Transform	18
Ψ	mother wavelet in the Fourier domain	18
$\mathcal{D}$	wavelet dictionary of the Continuous Wavelet Transform	18
k	wavelet scale or index number within a wavelet dictionary	18
f_c	central frequency, or scale, of a Ricker wavelet	18
$\mathbf{W}$	convolutional matrix for a wavelet $\mathbf{w}$	29
$\mathbf{D}$	wavelet dictionary containing all $\mathbf{W}$ for inversion	29
$\mathbf{m}$	local time-frequency representation in terms of a collection of wavelet-dependent reflectivity series	29
$P_l(\omega)$	prediction filter at the channel l in the Fourier domain	60
$\mathbf{P}$	prediction filtering operator, i.e. $P_l(\omega)$ for all ω	62
$\mathbf{A}$	$f-x-y$ spectral decomposition operator	63
$\mathbf{v_{MMV}}$	multiple measurement vector	75
$\mathbf{c}_i$	a single component of a multicomponent seismic trace	75
$\mathbf{g}$	a vector group utilized in group sparsity	77
g	index for groups in the mixed $\ell_2-\ell_1$ norm	76
ϱ	index for elements of a group in the mixed $\ell_2-\ell_1$ norm	76
J	cost or objective function	30
$\mathcal{R}(\mathbf{m})$	Regularization term on the model $\mathbf{m}$	31
λ	trade-off parameter	31
$\mathbf{I}$	identity matrix	31
$\mathbf{Q}$	weighting matrix for Iteratively Reweighted Least Squares	33
ϵ	a small number to avoid division by zero	33
α	parameter controlling the step-length for gradient descent methods	34
$\mathbf{y}_i$	model update for the Fast Iterative Soft Thresholding Algorithm	36
ξ_i	update parameter for the Fast Iterative Soft Thresholding Algorithm	36
$\text{sgn}(\cdot)$	sign function	35
$\mathcal{T}_\lambda(\cdot)$	soft thresholding function	35
$\mathcal{G}_\lambda(\cdot)$	multidimensional thresholding operator	76
$O(\cdot)$	rate of convergence/algorithm cost	36

CHAPTER 1

Introduction

1.1 Background

One of the major goals of explorational geophysics is to provide an image of the subsurface to aid in the the mapping and characterization of subsurface structures. This endeavour has proven to be quite valuable in a number of different activities such as petroleum and mineral exploration, hydrology, hydrothermal exploration, and monitoring CO_2 sequestration. However, there are numerous methods in which one could generate an image of the subsurface. For example, microgravity surveys measure local gravity differences, resistivity surveys measure the apparent resistivity of the subsurface by injecting current into the ground, and seismic surveys measure acoustical energy after it has propagated through the subsurface. The data acquired from all of these methods characterize and map the subsurface, but they do so in different ways as each method is sensitive to different properties within the subsurface. Although this thesis deals with reflection seismic data, the concepts that are explained could also be applied to any of these other data sets.

To understand the subsurface image that seismic data provides, a brief explanation of the seismic experiment is required. The acquisition of seismic data consists of generating sources of seismic energy that travel through the subsurface to receivers, commonly called geophones, where they are subsequently recorded. Ideally, a source and receiver pair should occupy the same physical location on the surface, as depicted in Figure 1.1a, making the seismic experiment similar to a depth sounding experiment. However, this is difficult to accomplish in application and, therefore, a lateral distance between sources and receivers is utilized. The spatial location of the seismic recording or trace from a source-receiver pair is often assigned as the midpoint between the source and receiver. Due to the design nature of seismic

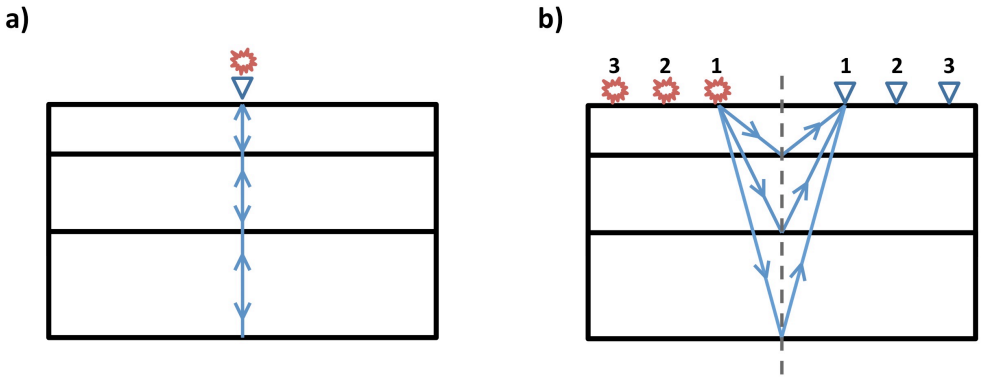


Figure 1.1: Seismic acquisition for a) ideal circumstances and b) actual circumstances (common midpoint gather). Red stars denote sources, blue triangles denote receivers, blue lines with arrows denote seismic wave propagation, and the grey dashed line denotes the common midpoint for the source-receiver pairs.

acquisition, multiple traces often have similar spatial midpoints which are subsequently organized into common midpoint gathers as illustrated by the common midpoint gather for pairs of source and receiver locations depicted in Figure 1.1b.

By spatially separating the source and receiver, seismic energy can propagate to the receiver through several different methods that would not be present under the ideal conditions of Figure 1.1a. Governed by changes in the acoustical impedance (the product of density and propagation velocity) of the subsurface, seismic energy can not only reflect back to the surface but can now also refract along these layer interfaces or travel along the surface via the direct wave or through ground roll (surface waves). All of these wavefronts are evident within an example shot gather[1] illustrated in Figure 1.2. Since reflected seismic energy can travel deeper and consequently sample more of the subsurface than other types of recorded waves, common midpoint gathers are often processed to try enhance and maintain only this portion of the data.

During this processing stage, reflection seismic energy is enhanced through several different methods. Some of these methods include random noise attenuation, ground roll removal, static corrections to help mitigate near surface effects, and multiple attenuation to leave only seismic energy that has reflected off of one interface. Another important procedure during the processing stage of common midpoint gathers is the normal moveout (NMO) correction. The NMO correction removes the curvature from reflection events, which can

[1]For a simple one dimensional Earth model (i.e., no dip), shot and common midpoint gathers look similar provided that the source does not vary significantly from shot to shot.

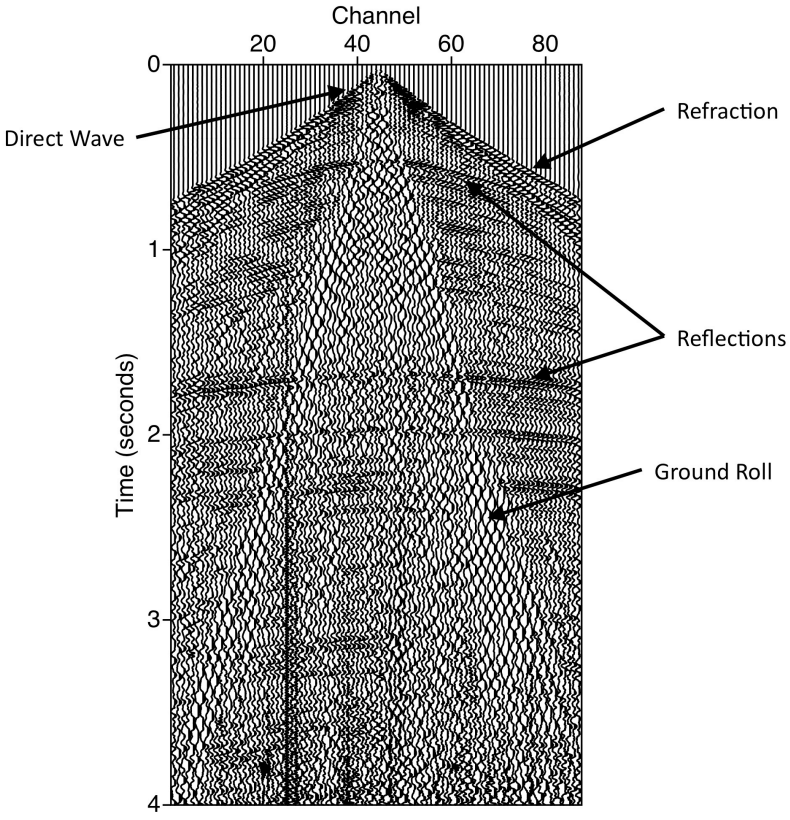
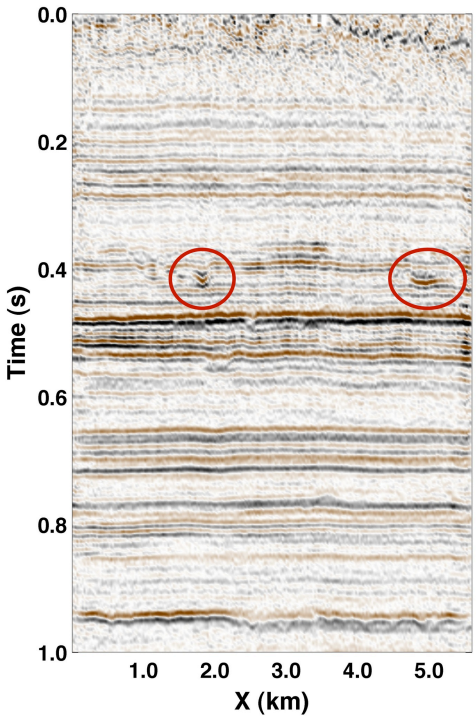


Figure 1.2: Example of a shot gather containing different waveforms from Yilmaz (2001) Shot Gather 25.

be seen in Figure 1.2, such that all single reflection events, also known as primaries, are flat. At this stage, the common midpoint gathers have been processed to contain only flat primary events which can be used for pre-stack analysis such as amplitude versus offset analysis (AVO) or stacked together to produce a single trace for each common midpoint gather. This stacked trace now represents the seismic trace that would have been recorded under the ideal conditions of Figure 1.1a.

Depending upon the spatial acquisition of the seismic data, stacked data can be a two dimensional line (time and space) or a three dimensional cube (time and two spatial coordinates). This stacked data, whether it is a two dimensional line or a three dimensional data cube, provides the image of the subsurface which can be obtained from reflection seismic experiments. These images help to provide a more continuous spatial sampling of the subsurface, unlike the sparse spatial sampling of well-logs, and aid in its geological interpretation. A real

a)



b)

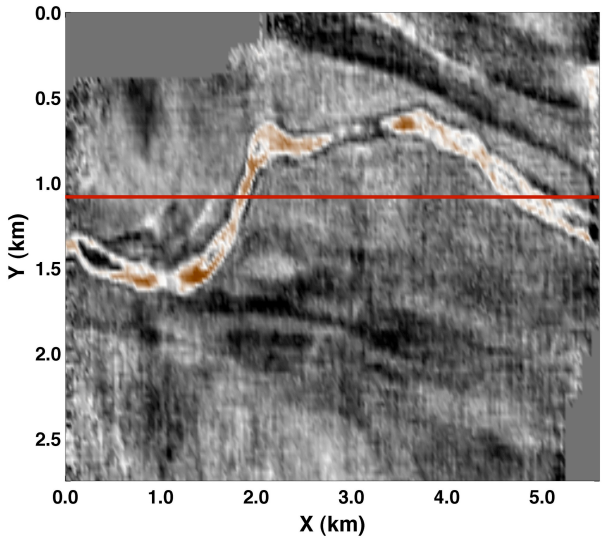


Figure 1.3: Real poststack 3D seismic data depicting a channel in a) cross section (red circles depict paleochannel) and on a b) time slice (red line denotes location of cross section).

data example of a three dimensional poststack dataset is shown in Figure 1.3 demonstrating the type of images that can be obtained from seismic data. The three dimensional poststack data is viewed in a cross section (a slice where one spatial coordinate is held constant) and on a time slice (a slice where the time coordinate is held constant) and depicts a paleochannel. If well-logs were the only dataset available to characterize this paleochannel, multiple wells would be required to achieve enough spatial sampling to properly locate and map its presence. By providing a more continuous spatial sampling, poststack reflection seismic data can help to illuminate previously unknown regions of the subsurface in a much more efficient manner. This property significantly aids in the mapping of localized subsurface features, such as paleochannels.

For a more thorough quantitative analysis of seismic data, the generation and analysis of seismic attributes is often performed. Seismic attributes are simply a quantitative measure of a seismic characteristic of interest (Chopra and Marfurt, 2005), and include measurements such as the degree of local curvature or coherence. For the real data example in Figure 1.3, seismic attributes could be used to help define the exact position and define subtle variations and structures within the paleochannel. This thesis will deal specifically with generating and studying the seismic attribute of local frequency content (Taner et al., 1979; Bodine, 1984). To aid in the understanding of the localized time-frequency analysis of a time series, a brief introduction to the problem is provided in the following section.

1.2 Local Time-Frequency Analysis

A common practice within signal processing is to represent the time series of interest in another domain. One of the most well known transforms to accomplish this goal is the Fourier transform which converts a signal from the time domain into the frequency domain. The Fourier transform achieves this transformation through representing the time series as a summation of complex exponentials. Since a complex exponential can also be written as a summation of the cosine and sine functions according to Euler's formula, the group of complex exponentials utilized by the Fourier transform is chosen to vary based upon the frequency of the cosine and sine functions. Figure 1.4 highlights the application of the Fourier transform on an example time series composed from a 15Hz cosine function and a 40Hz sine function. As the basis or set of functions that the Fourier transform utilizes are complex numbers, signals within the Fourier domain can become complex as well. Therefore, only the amplitude of the signal in the Fourier domain is displayed in Figure 1.4. The phase of this example time series in the Fourier domain is 0 at 15Hz to represent the cosine function and $\pi/2$ at 40Hz to represent the sine function.

This form of time-frequency analysis has found many applications in signal processing such

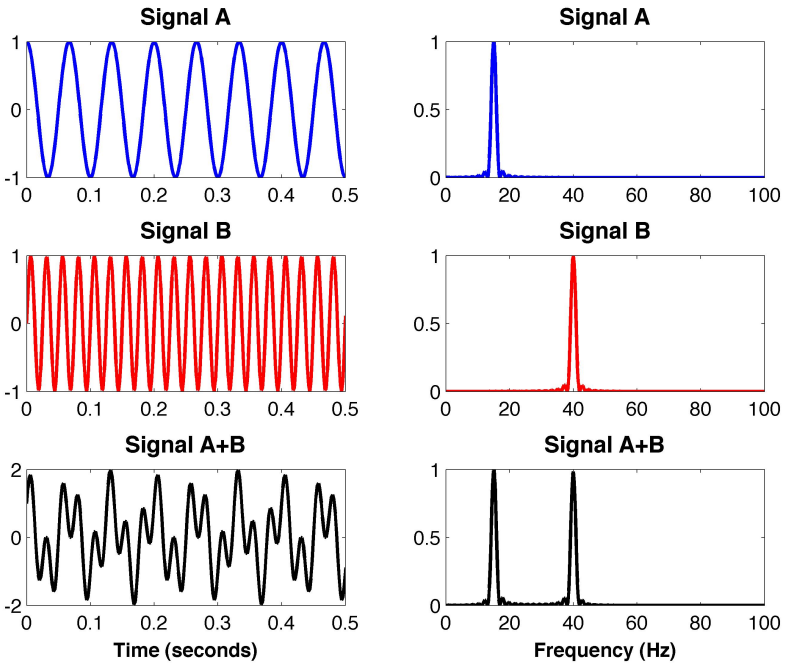


Figure 1.4: The time and frequency representation of a 15Hz cosine function, a 40Hz sine function, and their combination. (The normalized amplitude of the frequency representation is displayed.)

as seismic data analysis. Through fast computational algorithms such as the Fast Fourier Transform (FFT) and properties such as the convolutional property where convolutions in time become multiplications in frequency, the Fourier transform has been found to greatly increase the efficiency of many computational problems. In the case of seismic data analysis, the Fourier transform has provided a powerful tool to not only work in a more efficient manner but also analyze seismic traces in a different domain. For example, the Fourier transform has provided an effective domain for noise attenuation through methods such as bandpass filtering, $f-k$ filtering, and $f-x$ deconvolution and for the regularization and interpolation of seismic data to produce a more consistent spatial sampling.

One of the attributes that the Fourier transform does not provide is the time locations of specific frequencies within the signal. This problem is of little consequence for stationary signals, such as the example signal in Figure 1.4, but if the signal is non-stationary or time-varying it could prove beneficial to determine where specific frequencies are either present or absent. For example, a musical score denotes particular frequencies (i.e. notes) to be played at specific times. The non-stationary signal depicted in Figure 1.5, where the 40Hz sine function does not persist over the entire time interval, helps to further present the

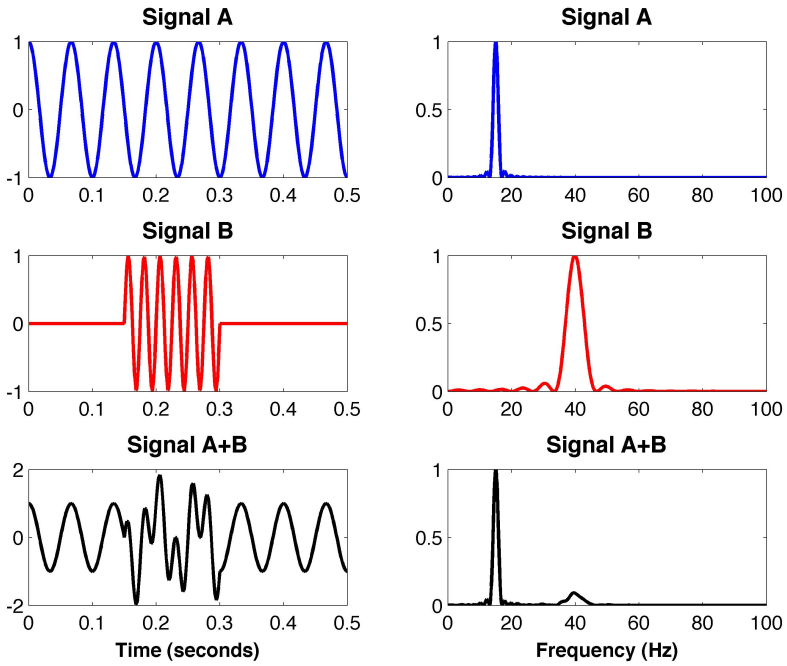


Figure 1.5: The time and frequency representation of a 15Hz cosine function, a truncated 40Hz sine function, and their combination. (The normalized amplitude of the frequency representation is displayed.)

problem of utilizing the classical Fourier transform for local time-frequency analysis. Since the Fourier transform utilizes complex exponentials and hence cosine and sine functions to convert a time series into the frequency domain, signals such as the truncated version of the 40Hz sine function require more coefficients to be represented in the frequency domain even though the actual frequency content of the this signal is 40Hz. By only analyzing the frequency domain representation of the signal, it becomes impossible to determine where certain frequencies are present or absent in the time domain. Rather than analyzing the signal globally with the traditional Fourier transform, local time-frequency analysis solves this time localization of frequency content by considering only variations of the signal on a local scale.

1.3 Thesis Overview

The organization of this thesis is as follows:

- Chapter 2 provides an overview of commonly applied local time-frequency transforms

for seismic data. A particular emphasis is placed upon the Continuous Wavelet Transform (CWT) since it represents one of the most general local time-frequency transforms. In fact, most other local time-frequency transforms can, in some degree, be explained by the CWT. The advantages of using complex wavelets for the CWT is also discussed within this chapter.

- Chapter 3 explains how local time-frequency analysis, or spectral decomposition, can be posed as an inverse problem to produce a high resolution local time-frequency representation. Through utilizing a data representation strategy that is based upon the CWT, the ℓ_1 norm is chosen as a regularization function to enforce sparsity upon the obtained local time-frequency model. The ℓ_1 norm regularized inverse problem is solved using the Fast Iterative Soft Thresholding Algorithm (FISTA).

- Chapter 4 highlights the application of the high resolution spectral decomposition method developed in Chapter 3 on the real seismic data set shown in Figure 1.3. The frequency behaviours of the paleochannels within this dataset is discussed as well as an efficient viewing strategy for spectrally decomposed data based upon RGB image blending is presented.

- Chapter 5 presents a method to deal the variable trace to trace noise that becomes more prevalent when viewing constant frequency sections obtained from high resolution spectral decomposition. This method incorporates the prediction filtering concepts of $f - x$ deconvolution as a preconditioner to the spectral decomposition inverse problem and allows for the recovery of a smoother, high resolution local time-frequency representation.

- Chapter 6 deals with the denoising of the reconstructed signal from local time-frequency analysis. Through posing the sparse spectral decomposition problem through inversion, random noise can be attenuated for the signal in the time domain. These results can become more powerful with the incorporation of group sparsity constraints rather than sparsity constraints if a multi-component signal is considered.

- Chapter 7 introduces a relatively new noise attenuation filter called the Non-Local Means (NLM) algorithm to be applied with seismic data. Although its original formulation can be computationally costly, it has the ability to preserve certain features within the data which other more commonly utilized filters for seismic data cannot.

- Chapter 8 provides a summary of the work within this thesis.

CHAPTER 2

Common Local Time-Frequency Analysis Methods

2.1 Introduction

The frequency content of a time series, such as a seismic signal, is commonly found using the Fourier transform. However, the frequency information obtained from this method relates to the entire time series and contains no information about local variations. Knowledge of how the frequency content of a signal varies in time can be significant, as highlighted by the common example of a musical score which denotes particular frequencies (i.e. notes) to be played at specific times. These local variations of frequency content in a signal can be acquired through local time-frequency analysis, also known as spectral decomposition, which decomposes a one dimensional signal (time) into a two dimensional space (time and frequency). Resolving how the frequency content of a signal changes with time has been studied in a large range of fields such as geophysics, quantum mechanics, engineering, sound analysis and speech recognition, radar, and musicography (Gardner and Magnasco, 2006).

In an effort to create the optimal time-frequency representation of a seismic signal, the geophysics community has employed several methods for spectral decomposition. Unlike instantaneous attributes such as the instantaneous frequency (Taner et al., 1979; Cohen, 1993), spectral decomposition attempts to describe the complete frequency spectrum at each time sample rather than generate a scalar parameter. This type of local time-frequency analysis allows for a greater wealth of information to be studied and aid in exercises such as reservoir characterization (Partyka et al., 1999). Methods for local time-frequency analysis that

have been employed on seismic signals generally fall under two categories, those based upon the Short Time Fourier Transform (STFT) or those based upon the Continuous Wavelet Transform (CWT).

The STFT generates a localized time-frequency representation of a time series by applying the Fourier transform within a sliding window. However, the length of the sliding window has a strong effect on the resolution of the time-frequency representation; a small window is well resolved in time but poorly resolved in frequency, whereas, a larger window is poorly resolved in time but well resolved in frequency. To help minimize the tradeoff between time-frequency resolution and the effects of applying the Fourier transform within a window, several other transforms similar to the STFT have been created such as the S transform (Stockwell et al., 1996; Pinnegar and Mansinha, 2003), and the Wigner-Ville distribution (Steeghs and Drijkoningen, 2001).

The other general category of spectral decomposition methods that have been commonly employed on seismic signals are those based upon the CWT (Chakraborty and Okaya, 1995; Sinha et al., 2005). The CWT operates by generating a representation of the seismic signal in terms of a wavelet family or group of pre-selected wavelets. For local time-frequency analysis, the wavelet family or group of pre-selected wavelets will vary based upon the frequency content of each individual wavelet. Other methods generally based upon the CWT that have been used with seismic signals include the sine curve fitting method of Hardy et al. (2003), the wavelet matching algorithm of Bradford and Wu (2007), and spectral decomposition using matching pursuit (Castagna et al., 2003; Wang, 2007) or other inversion methods (Portniaguine and Castagna, 2004; Bonar and Sacchi, 2010). In this chapter, the most commonly applied local time-frequency methods for seismic data will be reviewed with a particular emphasis on the CWT.

2.2 Instantaneous Frequency

A common method to discern information about the local frequency content of a time series is through the analysis of its analytic signal. The analytic signal, $\tilde{x}(t)$, of a real-valued time series, $x(t)$, is simply a mathematical manipulation which enables the real-valued time series to be rewritten as a signal of complex numbers that have an amplitude, $A(t)$, and a phase, $\theta(t)$. The imaginary component of this analytic signal is defined by the Hilbert transform, $\mathrm{H}[x(t)]$, which is a $\pi/2$ phase shift of the original time series,

$$\tilde{x}(t) = x(t) + i\mathrm{H}[x(t)] = A(t)e^{i\theta(t)}. \tag{2.1}$$

The computation of this analytic signal, $\tilde{x}(t)$, can easily be accomplished in the Fourier domain from the original real-valued time series, $x(t)$ (Bendat and Piersol, 2000),

$$\tilde{X}(f) = \begin{cases} 2X(f) & \text{if } f > 0 \\ X(f) & \text{if } f = 0 \\ 0 & \text{if } f < 0 \end{cases}, \tag{2.2}$$

where $\tilde{X}(f)$ and $X(f)$ denote $\tilde{x}(t)$ and $x(t)$ in the Fourier domain. Essentially, only half of the Fourier frequency content is maintained from the original real-valued signal to create the analytic signal. Due to the symmetries present within the Fourier domain for real-valued time series, the discarding of negative frequencies results in no loss of information since the original real-valued time series is simply the real component of the analytic signal. A detailed explanation of the meaning of the complex analytic signal as well as a physical illustration with a seismic trace is provided by Barnes (1998).

Once the analytic signal is obtained, a variety of instantaneous or time localized attributes can be calculated (Cohen, 1993). For instance, the instantaneous amplitude, $A(t)$, and the instantaneous phase, $\theta(t)$, can be obtained directly from forming the analytic signal. The

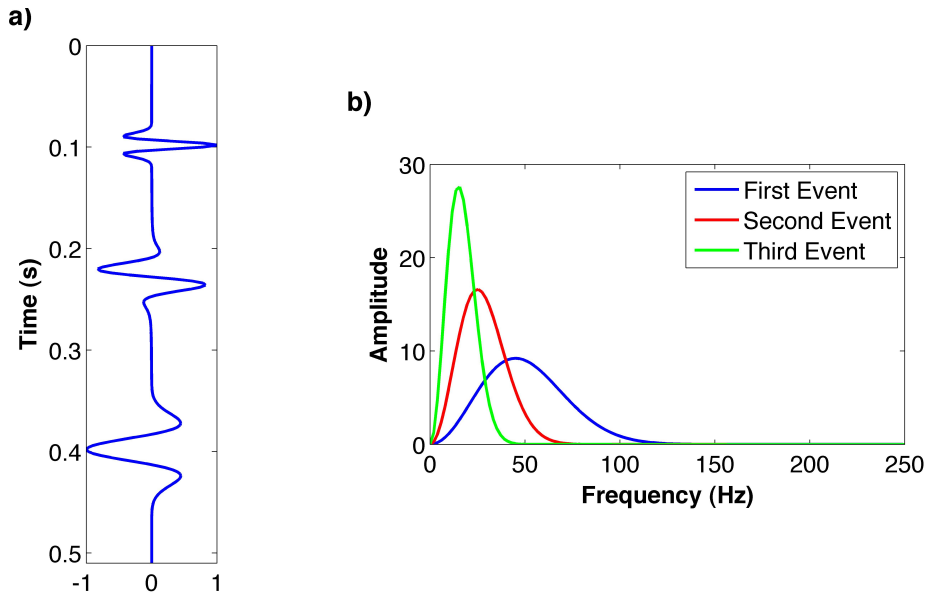


Figure 2.1: a) A synthetic time series composed from a 45Hz central frequency Ricker wavelet, a 90° phase rotated 25Hz central frequency Ricker wavelet, and a 15Hz central frequency Ricker wavelet of negative amplitude and b) the associated amplitude spectrum for each event.

instantaneous frequency, which represents the average frequency at a particular time (Cohen and Lee, 1989), is obtained by taking the derivative of the instantaneous phase with respect to time,

$$f_{inst}(t) = \frac{1}{2\pi}\frac{\mathrm{d}\theta(t)}{\mathrm{d}t}. \tag{2.3}$$

The normalization factor in Equation 2.3 is simply the conversion factor between angular frequency ω and frequency f. A thorough analysis of the significance of the instantaneous frequency and related instantaneous attributes to seismic data is detailed in Taner et al. (1979).

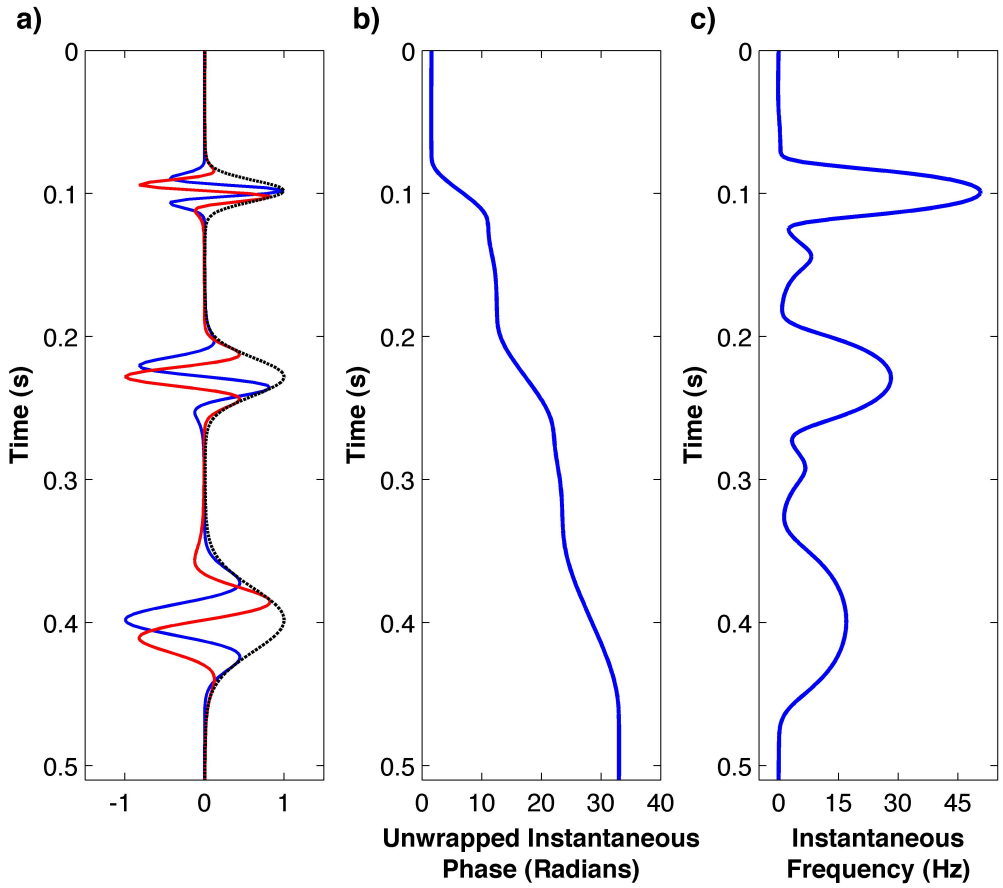


Figure 2.2: Instantaneous attributes from the synthetic time series in Figure 2.1. a) The instantaneous amplitude (black dotted line) with the synthetic time series (blue) and the imaginary component of the analytic signal (red). b) The unwrapped instantaneous phase. c) The instantaneous frequency.

To illustrate the application of the various local time-frequency analysis methods the synthetic time series in Figure 2.1 was created. This time series is composed from a 45Hz central frequency Ricker wavelet located around 0.1s, a 90° phase rotated 25Hz central frequency Ricker wavelet located around 0.2s, and a 15Hz central frequency Ricker wavelet with negative amplitude located around 0.4s. The instantaneous amplitude, phase, and frequency of this synthetic time series is displayed in Figure 2.2. Typically, the phase of a complex number is constrained to be its principal value or constrained to the region $(-\pi, \pi]$. However, this would produce a discontinuous function for the instantaneous phase, and thus instantaneous frequency. To avoid this problem and produce a continuous function, the instantaneous phase is unwrapped or adjusted to move outside the region $(-\pi, \pi]$ when large jumps of 2π are encountered. As expected, the instantaneous frequency that is derived from the unwrapped instantaneous phase becomes larger with fast changes in the instantaneous phase and vice versa. Although it adequately localizes and recovers the mean frequency at each event, the instantaneous frequency does not provide any information about the spread of frequency content or its relative intensity (Cohen and Lee, 1989) which can be potentially useful in the analysis of time series.

2.3 Short Time Fourier Transform

Probably the most well-known and intuitive local time-frequency transform is the Short Time Fourier Transform (STFT). Originally proposed in Gabor (1946), the STFT localizes the frequency content of a time-series by performing the Fourier transform within small sliding windows according to,

$$F_{STFT}(t, \omega) = \int_{-\infty}^{\infty} x(\tau) h_{STFT}(t - \tau) e^{-i\omega\tau} \mathrm{d}\tau, \tag{2.4}$$

where F_{STFT} is the local time-frequency representation obtained from the STFT, $h_{STFT}(t)$ is a real-valued and symmetric windowing function, and $e^{-i\omega t}$ is the Fourier kernel. Essentially, the STFT computes the Fourier transform of a portion of the time series around a particular time sample to localize the frequency content and then repeats the process for all time samples. Therefore, the selection of the windowing function becomes crucial to the applicability of this method. To avoid ringing effects caused by the Gibbs phenomenon, optimal window choices for the STFT tend to favour tapered windowing functions such as Hamming or Hann windows. In fact, the Gaussian function has been proven to provide the most optimal windowing function to minimize the uncertainty in the resolution of local time-frequency analysis based upon the STFT (Mallat, 2008).

The time-frequency resolution of the STFT not only depends upon the type of windowing

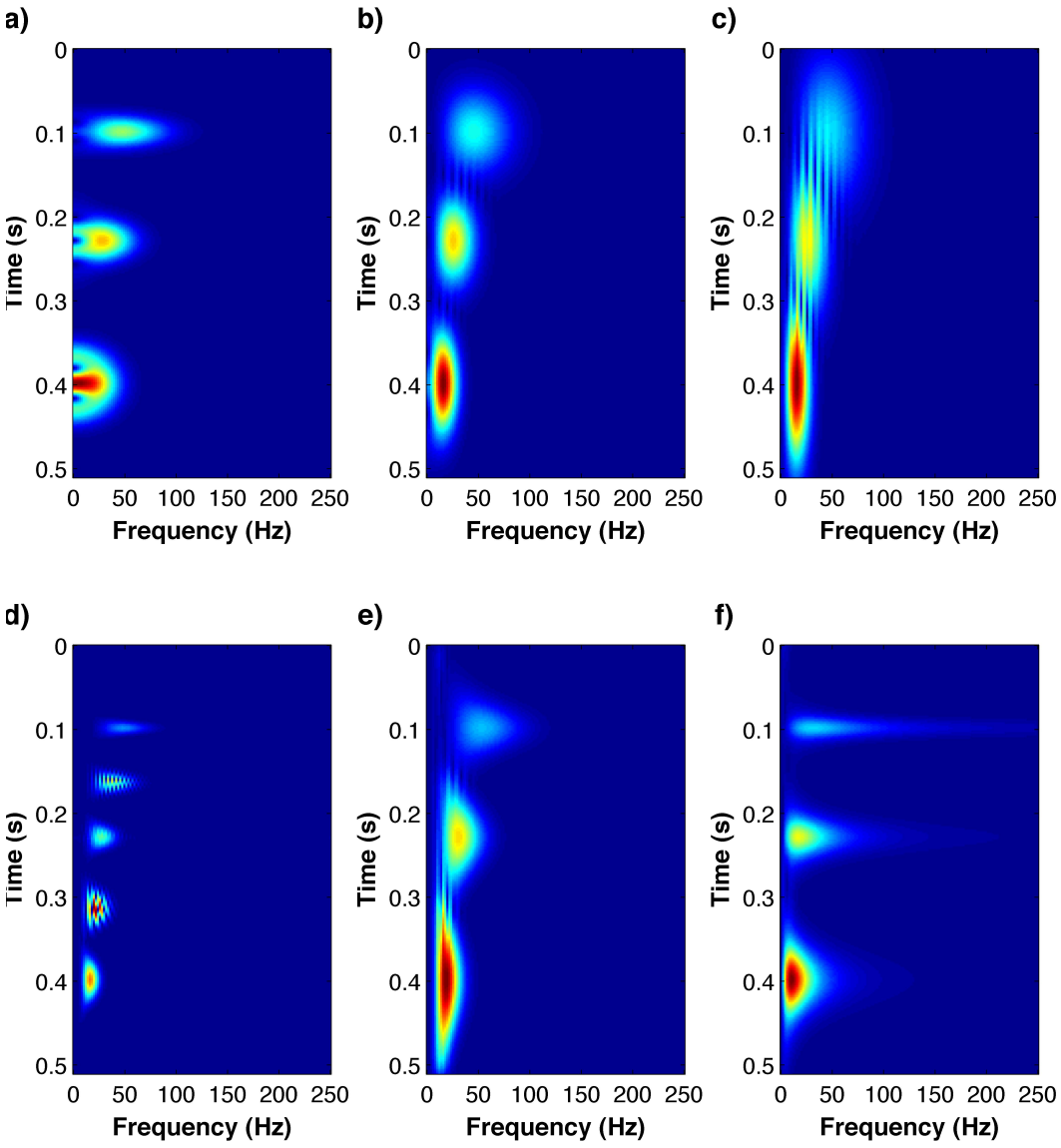


Figure 2.3: The amplitude of the local time-frequency representation of the example time series in Figure 2.1 obtained from a) STFT with a 60ms window, b) STFT with a 200ms window, c) STFT with a 360ms window, d) Wigner-Ville with τ set to be half of the total signal length, e) S-transform, and f) CWT with complex Ricker wavelets (scale has been converted to frequency).

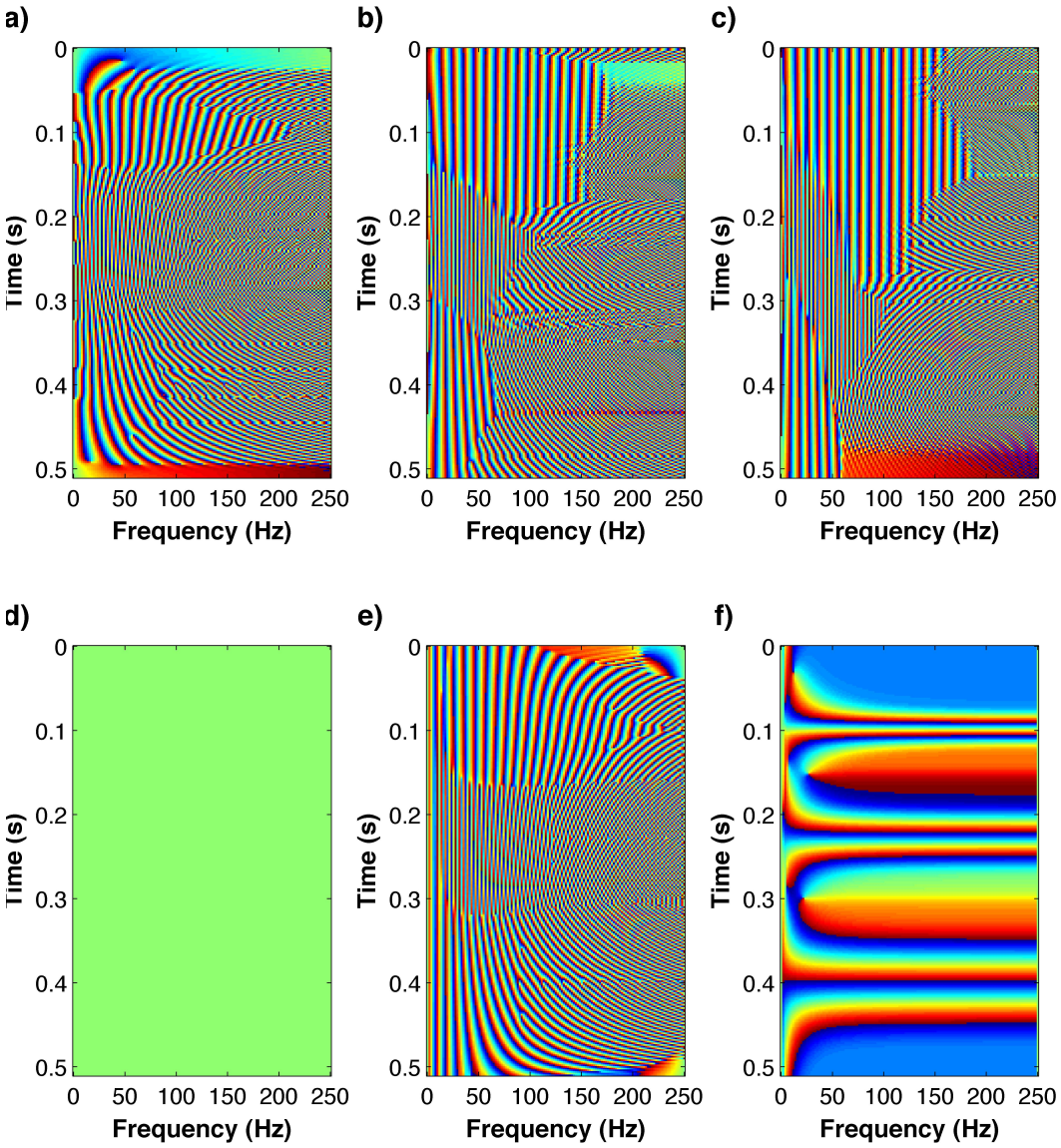


Figure 2.4: The phase of the local time-frequency representation of the example time series in Figure 2.1 obtained from a) STFT with a 60ms window, b) STFT with a 200ms window, c) STFT with a 360ms window, d) Wigner-Ville with τ set to be half of the total signal length, e) S-transform, and f) CWT with complex Ricker wavelets (scale has been converted to frequency). The colourbar for all plots is defined as blue $= -\pi$, green $= 0$, and red $= \pi$.

function implemented but also the windowing functions length in time. This property is displayed in Figure 2.3 a), b), and c) where three different length Gaussian functions are employed to compute the STFT of the synthetic time series in Figure 2.1. By increasing the time length of the window, increased resolution is obtained in the frequency dimension at the expense of the resolution in the time dimension. This trade-off property between time and frequency resolution has often been compared to the Heisenberg uncertainty principle (i.e. Mallat (2008)). The phase of the STFT local time-frequency representation also becomes more difficult to interpret than the phase of the standard Fourier transform. Not only can the STFT phase be referenced to zero time or the center of each sliding window, it is also influenced by the length of the windowing function as evidenced by Figure 2.4 a), b), and c).

2.4 Wigner-Ville Distribution

The Wigner-Ville distribution for time-frequency analysis within signal processing was originally introduced in Ville (1948), although it was originally studied by Wigner (1932), is defined by,

$$F_{WV}(t,\omega) = \int_{-\infty}^{\infty} \tilde{x}(t+\tau/2)\tilde{x}^{H}(t-\tau/2)e^{-i\tau\omega}d\tau, \tag{2.5}$$

where F_{WV} is the local time-frequency representation obtained from the Wigner-Ville distribution and $[\,]^H$ denotes the Hermitian or complex conjugate transpose. Since the analytic signal is used twice during the creation of the local time-frequency representation, the Wigner-Ville distribution is often called a quadratic or bilinear representation. By utilizing the time reversed analytic signal, $\tilde{x}^H(t-\tau/2)$, as the windowing function, the Wigner-Ville distribution can be considered as a special case of the STFT where the windowing effects have been minimized (Steeghs and Drijkoningen, 2001). However, despite its higher time-frequency resolution, the application of the Wigner-Ville distribution is limited by the existence of interference or cross-product terms (Mallat, 2008). These properties are highlighted in Figure 2.3 d) where the superior time-frequency resolution is contaminated with extraneous events caused by cross-product terms around 0.15s and 0.3s. The influence of these interference terms can be reduced by smoothing during the generation of the Wigner-Ville distribution, but this occurs at the expense of time-frequency resolution, see Wu and Liu (2009) and Steeghs and Drijkoningen (2001) for example. Another interesting property of the Wigner-Ville distribution is that its local time-frequency representation contains only real-values as evidenced by a constant zero phase in Figure 2.4.

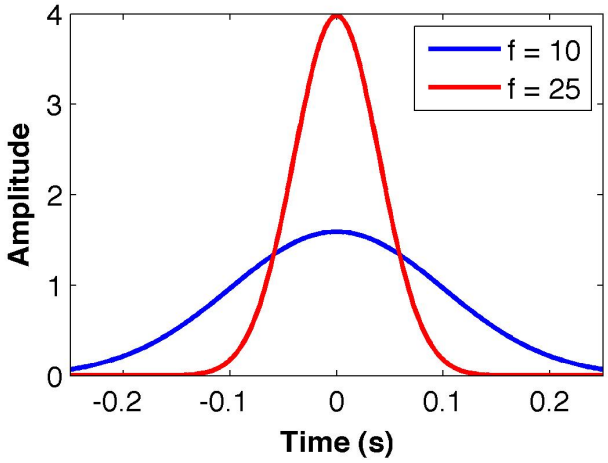


Figure 2.5: Frequency dependent Gaussian windows utilized by the S-transform.

2.5 S-transform

Proposed by Stockwell et al. (1996), the S-transform can be considered a special version of the STFT utilizing a Gaussian window. By scaling the Gaussian windowing function differently for each frequency, it attempts to combat the inherent time resolution versus frequency resolution trade-off problem that is inherent with the STFT. This frequency scaled Gaussian windowing function, defined by,

$$h_S(t) = \frac{|f|}{\sqrt{2\pi}} e^{-t^2 f^2/2}, \tag{2.6}$$

enables smaller time windows when analyzing higher frequencies to help with time localization and longer time windows for lower frequencies to help with frequency localization as depicted in Figure 2.5. The formal definition of the S-transform is given by,

$$F_S(t,f) = \int_{-\infty}^{\infty} x(\tau) \frac{|f|}{\sqrt{2\pi}} e^{-(t-\tau)^2 f^2/2} e^{-i\omega\tau} \mathrm{d}\tau, \tag{2.7}$$

where the windowing function $g(t)$ has been written out explicitly and F_S represents the local time-frequency representation obtained from the S-transform. Essentially, the S-transform's goal is to try and incorporate the positive properties of long and short time windows from the STFT together and eliminate the extra requirement of setting a window length for local time-frequency analysis. Figure 2.3 e) demonstrates the variable time-frequency resolution of the S-transform with low frequencies having a similar resolution to long time windows

of the STFT (Figure 2.3 c) and high frequencies having a similar resolution to short time windows of the STFT (Figure 2.3 a). Like the STFT, the phase of the local time-frequency representation obtained from the S-transform could be referenced to zero time or the center of each particular windowing function. The actual phase obtained from the S-transform becomes a mixture of the phases that would have been obtained from varying length windows of the STFT (Figure 2.4 e). In an effort to improve the resolution obtained from the S-transform, various scaling modifications to the windowing function have been attempted, i.e., Pinnegar and Mansinha (2003).

2.6 Continuous Wavelet Transfrom

Originating from Morlet et al. (1982) and Goupillaud et al. (1984), the Continuous Wavelet Transform (CWT) became popular within the signal processing community when its connections to discrete signal processing were established through the work of Daubechies (1988) and Mallat (1989). Unlike the STFT which applies the Fourier transform in small windows to achieve a local time-frequency representation, the CWT fits time-frequency atoms to the time series of interest. These time-frequency atoms often resemble tiny waves that grow and decay in short periods of time which has lead to their name "wavelets" (Chakraborty and Okaya, 1995). By choosing a specific wavelet which is often called the mother wavelet, the CWT generates the dictionary or group of daughter wavelets, $\mathcal{D}$, to be fit to the time series through scaled dilations and translations of the mother wavelet,

$$\mathcal{D} = \left[\psi_{\tau,k}(t) = \frac{1}{\sqrt{k}} \psi \left(\frac{t-\tau}{k} \right) \right], \tag{2.8}$$

where k corresponds to the scale of the daughter wavelet. To be considered as a possible mother wavelet, a function must satisfy the admissibility condition (Daubechies, 1992),

$$\int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} \mathrm{d}\omega < \infty, \tag{2.9}$$

where Ψ represents the mother wavelet, ψ, in the Fourier domain. As a result of satisfying the admissibility condition, the mother wavelet for a CWT also satisfies the conditions,

$$\int_{-\infty}^{\infty} \psi(t) \mathrm{d}t = 0, \tag{2.10a}$$

$$\int_{-\infty}^{\infty} |\psi(t)|^2 \mathrm{d}t < \infty, \tag{2.10b}$$

or in other words the mother wavelet has zero mean and a finite energy. There are many different functions that can be classified as wavelets, such as the Morlet wavelet, but for the purposes of this thesis only the Ricker wavelet was utilized to demonstrate the application of the CWT. The Ricker wavelet, also known as the Mexican hat wavelet, is simply the second derivative of the Gaussian function and can be uniquely defined by its central, or peak, frequency, f_c, in either the time or Fourier domain according to (Sheriff, 2002),

$$w(t) = \left(1 - 2\pi^2 {f_c}^2 t^2\right) e^{-\pi^2 {f_c}^2 t^2}, \tag{2.11a}$$

and,

$$w(f) = \frac{2}{\sqrt{\pi}} \frac{f^2}{{f_c}^3} e^{-f^2/{f_c}^2}. \tag{2.11b}$$

When implemented as the mother wavelet for the wavelet dictionary of the CWT, the Ricker wavelet's central frequency corresponds to its scale, k.

Once a suitable function is chosen as the mother wavelet for a CWT, a scalogram or time-scale $(t-k)$ map (Sinha et al., 2005) is generated via,

$$F_{CWT}(t,k) = \int_{-\infty}^{\infty} x(t) \frac{1}{\sqrt{k}} \psi^H\left(\frac{\tau - t}{k}\right) \mathrm{d}\tau, \tag{2.12}$$

where F_{CWT} represents the local time-frequency representation obtained from the CWT. Although the scalogram is a local-time frequency representation, since different scales of the wavelet have different frequency contents, it cannot be directly compared to other local time-frequency analysis methods such as the STFT because its measure of frequency is not the traditional frequency obtained from the Fourier transform. However, the Fourier spectrum of each of the scaled or daughter wavelets can be computed and consequently provide a mapping of $F_{CWT}(t,k)$ to $F_{CWT}(t,f)$. Figure 2.6 displays the results of mapping the time-scale map obtained from the CWT utilizing complex Ricker wavelets of the synthetic time series of Figure 2.1 to a local time-frequency representation. Despite not being exactly equivalent, the scale definition of the central frequency of Ricker wavelets generates a time-scale map that is almost identical to its mapped local time-frequency representation.

If the time-scale map in Figure 2.6 was generated using a dictionary of real-valued Ricker wavelets instead of complex Ricker wavelets, the time-scale map in Figure 2.7 a) would be obtained. Even though a localization of frequency content with time is still achieved, polarity effects overprint the frequency content for specific events. This is most evident in the second event of the synthetic time-series where the amplitude of the frequency content mirrors that of the time-series, changing from a negative polarity to a positive polarity. Intuitively, the frequency content at the central zero crossing of this event should be nonzero, which is not observed in the time-scale map of Figure 2.7 a). To combat this issue, a complex

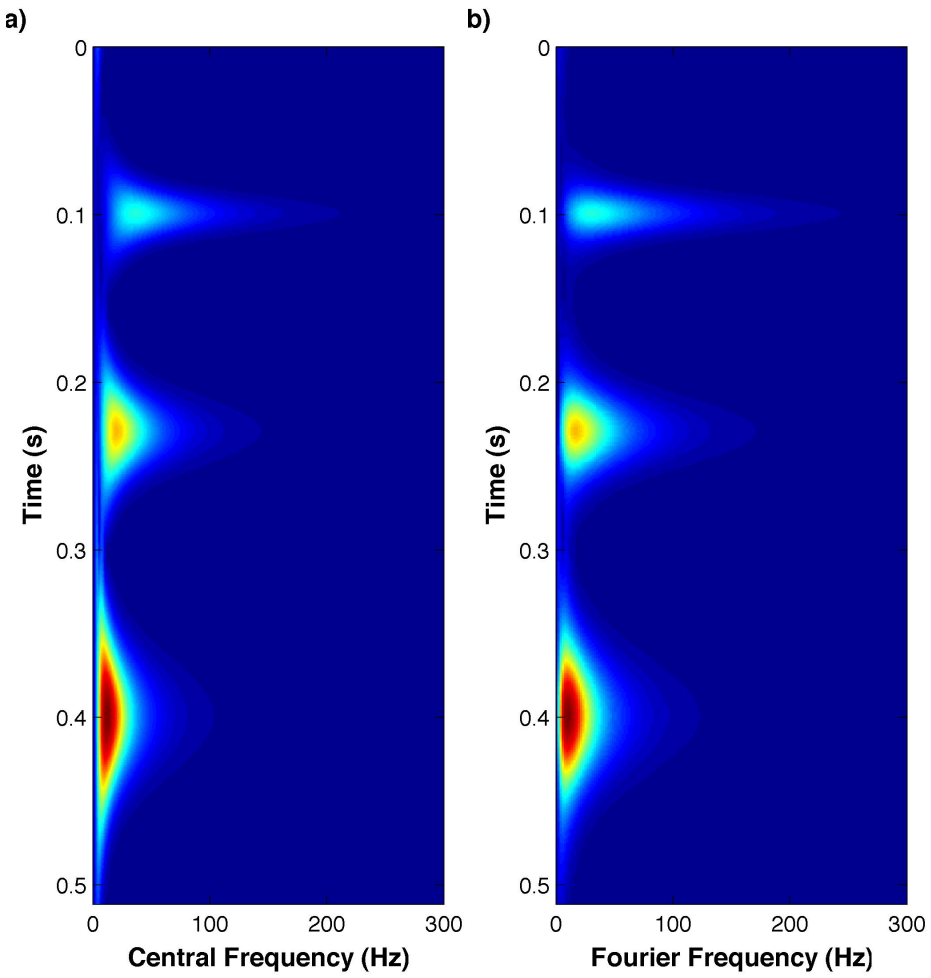


Figure 2.6: The amplitude of the a) time-scale map and b) time-frequency map obtained from the CWT using complex Ricker wavelets of the synthetic time-series in Figure 2.1. The scale of a complex Ricker wavelet is defined by its central frequency.

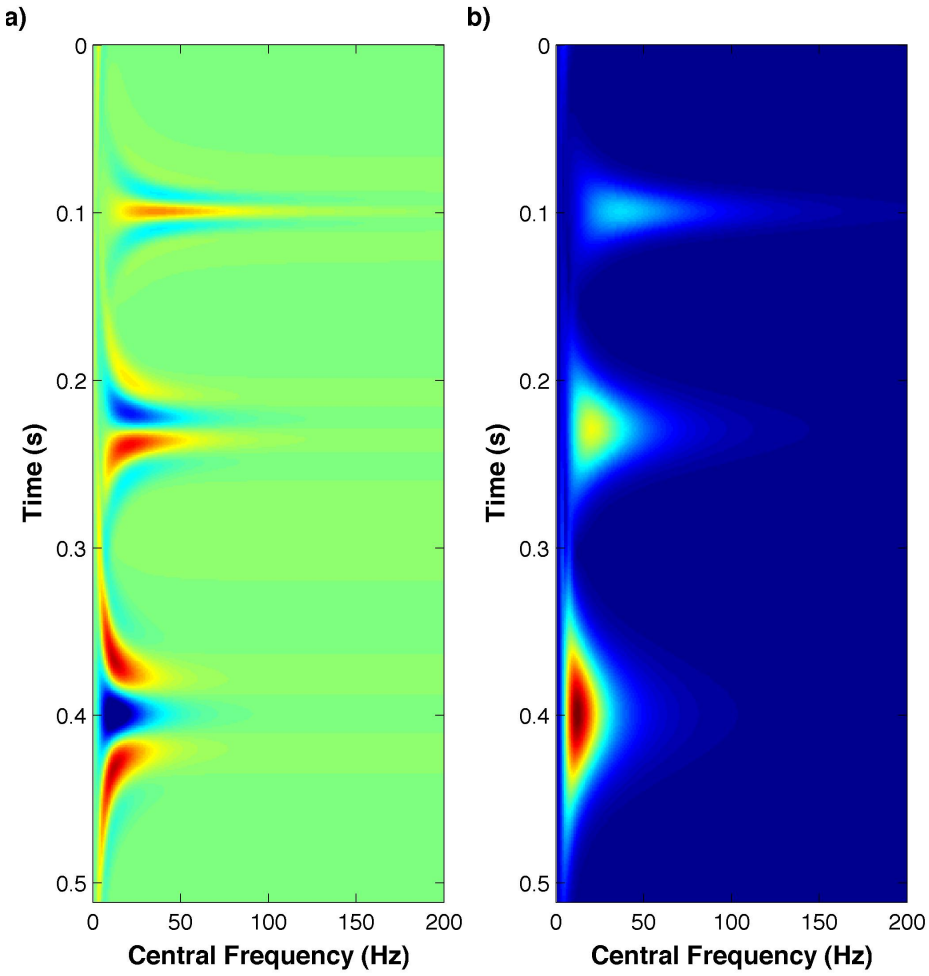


Figure 2.7: The time-scale map of the synthetic time series in Figure 2.1 obtained from the CWT using a) real Ricker wavelets and b) complex Ricker wavelets.

or analytic Ricker wavelet dictionary was created via Equation 2.2. The incorporation of complex Ricker wavelets for the wavelet dictionary forces the resultant time-scale map to be complex as well, providing an amplitude and phase for each coefficient of the time-scale representation (Mallat, 2008). The amplitude of the complex time-scale map, Figure 2.7 b), does not suffer from the polarity overprint as it relates to the envelope of the time-scale map obtained using the real-valued Ricker wavelet dictionary, while also having the added benefit of being associated with phase information. Essentially, the inclusion of complex wavelets in the wavelet dictionary implicitly adds a dramatic increase in the number of wavelets within the dictionary as each scale of the wavelet is no longer restricted to a particular phase.

In comparison to the other local time-frequency methods discussed, the amplitude of the local time-frequency representation obtained from mapping the scalogram of the CWT using complex Ricker wavelets is most similar to that of the S-transform as both methods exhibit a lack of interference terms and variable time-frequency resolution. In particular, the CWT utilizing complex Ricker wavelets provides higher time resolution at the cost of a more compact frequency resolution when compared to the S-transform, as exhibited in Figure 2.3. The phase of the local time-frequency representation obtained from the CWT with complex wavelets, Figure 2.4 f), is also easier to interpret than the other discussed local time-frequency methods. For the CWT, the phase simply represents the phase of the wavelet required to represent the time-series at that particular time sample.

The utilization of a complex Ricker wavelet dictionary not only enables phase to be associated with the time-scale map generated from the CWT but also enables the analysis of complex signals. Figure 2.8 depicts a complex signal composed of the example time-series from Figure 2.1 as the real component of the complex signal and three Ricker wavelets of various phases and central frequencies, which are unrelated to the example time-series, as the imaginary component and the amplitude of its local-time frequency representation obtained by the CWT. Since the signal to be analyzed is no longer contained to the real number domain, the entire frequency range (i.e., both positive and negative frequencies) must be displayed to fully visualize the local time-frequency representation of the complex signal. This is due to the lack of symmetries in the Fourier domain that complex signals have in comparison to real-valued time-series[1]. However, if the complex Ricker wavelet dictionary is computed to contain only positive frequencies, as suggested by Equation 2.2, it becomes impossible to solve for the negative frequencies of the signal if it is complex and symmetries cannot be used. To combat this issue, the complex Ricker wavelet dictionary must be expanded to incorporate wavelets which maintain only their negative frequencies. Figures 2.9 and 2.10 illustrate a complex Ricker wavelet containing only positive frequencies or only negative frequencies. The effect of maintaining the negative frequencies of a signal

[1]Real-valued signals in the Fourier domain have their real component defined by an even function and their imaginary component defined by an odd function.

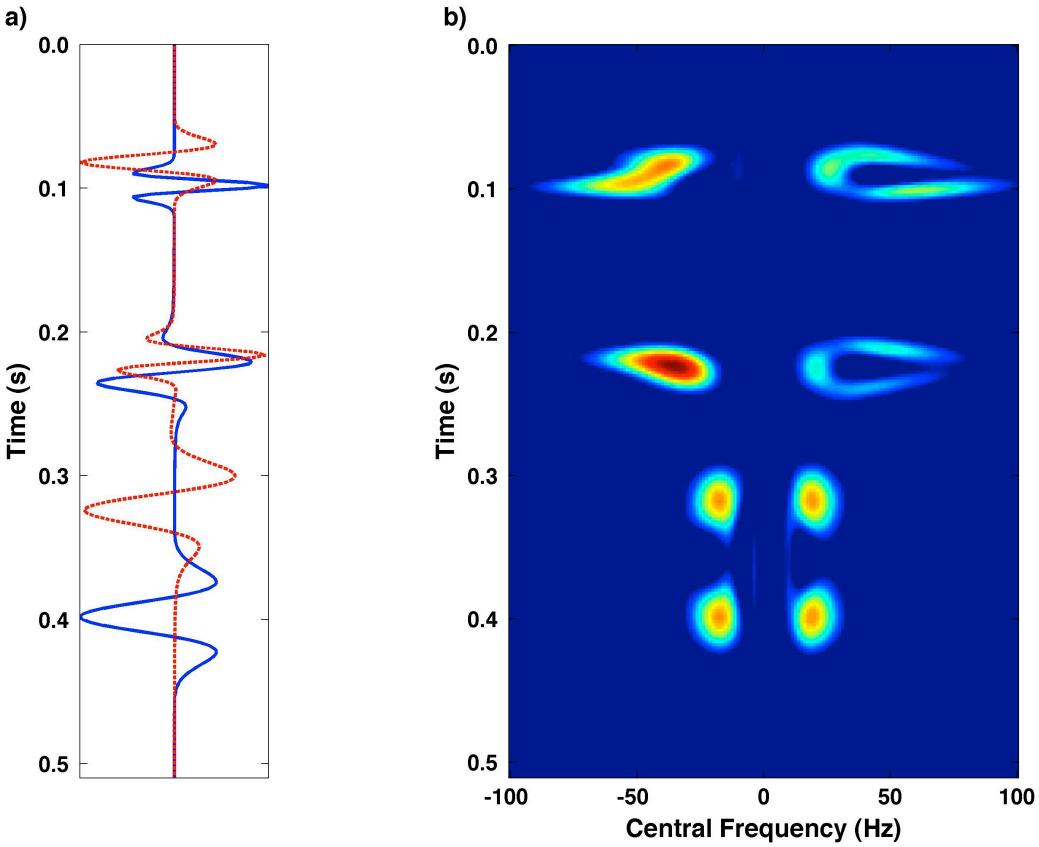


Figure 2.8: Local time-frequency analysis of a complex signal. a) The complex signal in which the real component (blue) is the example time-series in Figure 2.1 and the imaginary component (red) is composed of three phase shifted Ricker wavelets of different central frequencies. b) The amplitude of the time-scale map obtained from the CWT with complex Ricker wavelets.

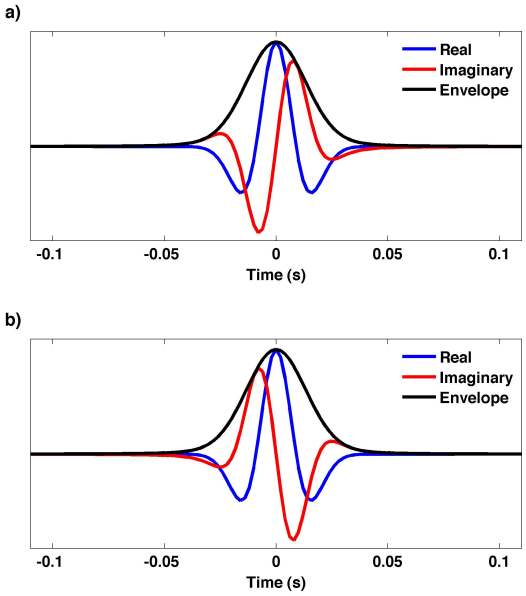


Figure 2.9: A complex Ricker wavelet containing a) positive frequencies and b) negative frequencies.

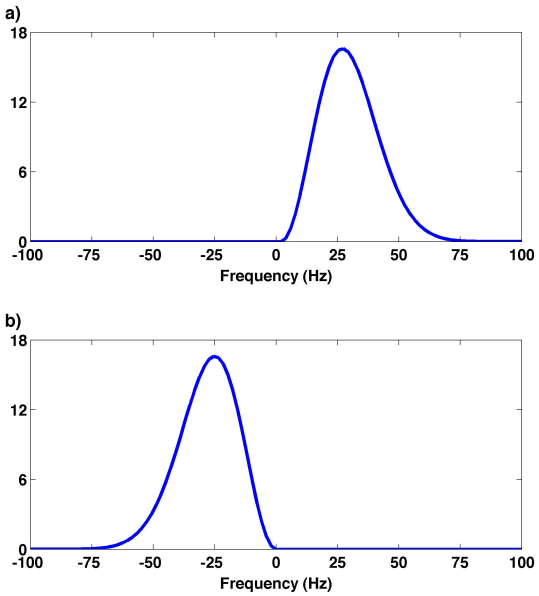


Figure 2.10: The amplitude of the Fourier spectrum for a) Figure 2.9 a) and b) Figure 2.9 b).

is that the imaginary component becomes a $-\pi/2$ phase shift of the real component, unlike a $+\pi/2$ phase shift associated with only maintaining the positive frequencies. The ability to apply spectral decomposition to complex signals may be useful in studying problems such as the motion of the Chandler Wobble through time (motion in the North-South direction could potentially become the real component of the signal while motion in the East-West direction could become the imaginary component) or two-component seismic data where the complex signal is composed by placing one of the components onto the imaginary axis. It should be noted that the ability to analyze complex signals is not unique to the CWT as other local time-frequency methods such as STFT and S-transform are capable as well.

2.7 Summary and Discussion

In this chapter, a review of commonly applied local time-frequency transforms that have been utilized for seismic data was provided. Unlike the global frequency analysis of a time-series which can be achieved through the Fourier transform, several methods for local time-frequency analysis have arisen due to the inherent uncertainty of knowing exactly which frequencies occur at particular time samples of this non-unique process. With the exception of the instantaneous frequency which generates a scalar parameter, all of the discussed local time-frequency analysis, or spectral decomposition, methods generate a frequency spectrum at each time sample and thus map a one-dimensional signal in time into a two-dimensional space of time and frequency. To accomplish this goal, the STFT applies the Fourier transform in small sliding windows, the Wigner-Ville distribution utilizes the time reversed analytic signal as a special windowing function, the S-transform applies the Fourier transform in frequency scaled windows, and the CWT utilizes time-frequency atoms called wavelets to be fit to the time-series. However, the resulting local time-frequency representation generated from each of these techniques varies. The STFT is quite sensitive to the size and choice of the windowing function producing a resolution trade-off between time and frequency. The Wigner-Ville distribution perhaps provides the highest resolution of both time and frequency but suffers from interference terms which generates time-frequency anomalies where there were no events in the time-series. The S-transform and the CWT utilizing complex Ricker wavelets both generate a local time-frequency representation with varying resolution where low frequencies are well resolved at the expense of time resolution and high frequencies are less resolved to promote higher time resolution. In direct comparison, the S-transform provides a higher frequency resolution whereas, the CWT utilizing complex Ricker wavelets provides a higher time resolution. It should be noted that the CWT actually produces a time-scale map, or scalogram, that can be mapped to a local time-frequency representation. However, for the case of complex Ricker wavelets, the scalogram is quite similar to the actual local time-frequency representation.

For all of the local time-frequency methods discussed, with the exception of the instantaneous frequency, the localization of the frequency content of a time-series with time is accomplished by studying the time-series within windows, either explicitly like the STFT or implicitly like the wavelets of the CWT. Thus, at some level all of these local time-frequency methods are connected. For instance, the STFT could be considered a special case of the CWT where the wavelets are defined by the product of the Fourier kernel, $e^{-i\omega t}$, and the windowing function, $g(t)$ (note that these particular wavelets may not satisfy the admissibility condition). In fact, the S-transform was originally explained in terms of the CWT. For this reason, an alternative approach to creating the optimal time-frequency representation of a time-series is studied in the next chapter. Rather than defining a new approach to solve for the time-frequency map, the local time-frequency analysis problem is posed as an inverse problem.

CHAPTER 3

High-Resolution Local Time-Frequency Inversion

3.1 Introduction

In an effort to obtain the optimal local time-frequency representation of a time-series, not only have various methods been developed, as discussed in Chapter 2, but also different techniques have been implemented to try and optimize the local time-frequency representations that are produced from these methods. One such optimization technique which operates after the local time-frequency representation has been created is the reassignment method (Kodera et al., 1976; Auger and Flandrin, 1995; Pedersen et al., 2003; Gardner and Magnasco, 2006). The reassignment method creates a modified version of the local time-frequency representation, regardless of the method by which it was obtained, by relocating values away from where they are computed to help produce a better localization of the signal components (Wu and Liu, 2009). Another technique to optimize the local time-frequency representation is to pose the local time-frequency analysis problem as a data representation problem that can be solved through inversion. The advantage of posing the local time-frequency analysis problem through inversion is that it generates a solution which satisfies certain conditions rather than altering a local time-frequency representation after it has been obtained. It is this optimization method which will be discussed within this chapter.

Through posing the spectral decomposition or local time-frequency analysis problem as an inverse problem, regularization can be incorporated to impose additional constraints on

the estimated local time-frequency representation for the recovery of a unique and stable solution. This is an advantageous characteristic due to the non-uniqueness of the local time-frequency analysis problem which allows for different local time-frequency representations to theoretically represent the signal of interest. The ill-posed nature of this problem is a major contributor to the many methods that have been developed for local time-frequency analysis. The goal of regularization is to make estimation problems well-posed (where a unique solution exists and depends continuosly on the data) by adding indirect constraints on the estimated model (Aster et al., 2005). The act of solving problems through regularized inversion has been commonly applied throughout geophysics for problems such as travel time tomography (Bube and Langan, 2008; Clapp et al., 2004), high-resolution Radon transforms (Sacchi and Ulrych, 1995; Trad et al., 2003), AVO inversion (Buland and Omre, 2003; Alemie and Sacchi, 2011), migration (Kuehl and Sacchi, 2003; Kaplan et al., 2010), and full waveform inversion (Plessix and Cao, 2011; Anagaw and Sacchi, 2011). Specifically for the problem of spectral decomposition or local time-frequency analysis, local time-frequency representations of a compact nature are desired to help emphasize small frequency changes within the signal.

A compact or high resolution local time-frequency representation essentially requires the local time-frequency distribution to be sparse or contain relatively few significant (i.e., non-zero) values. Several functions such as the Cauchy and Huber norms (Sacchi and Ulrych, 1995; Sacchi, 1997) can be utilized as regularization terms for robust or sparsity constrained inversion, however, this chapter will only specifically study sparsity constrained inversion based upon the ℓ_1 norm. Although sparsity constrained inversion has been a topic of study in the past (Taylor et al., 1979), a renewed vigor of its study has been undertaken with the introduction of compressive sensing (Candes et al., 2006; Candes and Tao, 2006; Donoho, 2006a). The compressive sensing theory asserts that the recovery of certain signals and images from far fewer samples or measurements than traditional methods utilize is possible based upon the principles of sparsity and incoherence (Candes and Wakin, 2008). Despite not posing the spectral decomposition problem in terms of compressive sensing, the recent study of efficient algorithms to solve sparsity constrained inversion problems have become extremely helpful for high resolution transforms.

Before sparsity constrained inversion is presented within this chapter, a data representation method for spectral decomposition which essentially provides a more physical meaning to the CWT is introduced. This data representation method provides an operator for mapping a time-series to a local time-frequency representation and a local time-frequency representation to a time-series. The implementation of this operator into an inversion problem is then demonstrated through the common damped least squares inversion problem. Finally, a high resolution spectral decomposition method is defined based upon ℓ_1 norm (sparsity) constrained inversion.

3.2 Spectral Decomposition Through Inversion

Before a model can be employed to represent a data set, a method must be defined which predicts the data to be observed from a specific model. This mapping from model to data is often called the forward operator and for spectral decomposition relates to how a local time-frequency representation can model a time-series. Working from a seismic data point of view, a forward problem for spectral decomposition is described from the work of Bonar and Sacchi (2010).

A seismic trace is often represented using a convolutional model which states that the seismic trace $s(t)$ is composed from the convolution of a wavelet $w(t)$ and a reflectivity series $r(t)$ with additive noise $n(t)$ (Robinson and Treitel, 1980),

$$s(t) = w(t) * r(t) + n(t). \tag{3.1}$$

Based upon this representation, a multi-wavelet convolutional model is defined to allow for the utilization of multiple wavelets to describe the seismic trace,

$$s(t) = \sum_{k}^{K} [w(t,k) * r(t,k)] + n(t). \tag{3.2}$$

The index k defines the particular wavelet $w(t,k)$ within the set of K wavelets and its associated coefficient series $r(t,k)$. Fundamentally, the defined multi-wavelet convolutional model generates $s(t)$ from the summation of K wavelet specific seismic traces $s(t,k)$, since the argument of the summation operator in Equation 3.2 is simply the convolutional model stated in Equation 3.1. Using this analogy, the coefficient series for each $w(t,k)$ can be considered as a wavelet-dependent reflectivity series. Although not physically realistic since a seismic source does not generate multiple wavelets, this method allows for an alternative representation of the seismic trace in terms of a set, or dictionary, of wavelets and their corresponding wavelet-dependent reflectivity series. The advantage of using multiple wavelets to represent the seismic trace is that the associated wavelet-dependent reflectivity series highlight areas within $s(t)$ which can be represented from its wavelet, $w(t,k)$. If the dictionary of wavelets is chosen to change only in a certain property, such as frequency content, the set of wavelet-dependent reflectivity series will highlight how this property changes in $s(t)$. Thus, the set of wavelet-dependent reflectivity series for a dictionary of frequency varying wavelets can be considered as a local time-frequency representation.

Equation 3.2 describes the forward problem of how a seismic trace is composed from a dictionary of wavelets and their corresponding wavelet-dependent reflectivity series. In

terms of linear algebra, this forward problem could also be written as,

$$\mathbf{s} = \begin{pmatrix} \mathbf{W}_1 & \mathbf{W}_2 & \dots & \mathbf{W}_K \end{pmatrix} \begin{pmatrix} \mathbf{r}_0 \\ \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_K \end{pmatrix} + \mathbf{n} = \mathbf{D}\mathbf{m} + \mathbf{n}, \tag{3.3}$$

where $\mathbf{W_k}$ denotes the convolutional matrix for $w(t,k)$, $\mathbf{D}$ denotes the wavelet dictionary, and $\mathbf{m}$ denotes the collection of wavelet-dependent reflectivity series. If noise is not considered, the local time-frequency representation or collection of wavelet-dependent reflectivity series can be obtained directly from taking the inverse of $\mathbf{D}$,

$$\mathbf{m} = \mathbf{D^{-1}s}. \tag{3.4}$$

However, this method of solving for the local time-frequency representation may not be attainable since $\mathbf{D}$ might not be invertible. To help avoid this problem, Equation 3.3 can be multiplied by the Hermitian transpose or adjoint of $\mathbf{D}$ and, if noise is not considered, the solution for the local time-frequency representation becomes,

$$\mathbf{m} = (\mathbf{D}^H\mathbf{D})^{-1}\mathbf{D}^H\mathbf{s}. \tag{3.5}$$

By multiplying Equation 3.3 by the adjoint of $\mathbf{D}$, the inverse of $\mathbf{D}^H\mathbf{D}$ is more likely to exist since its dimensions are ensured to be square. Obtaining the local time-frequency representation in this manner is equivalent to minimizing the cost function J,

$$J = ||\mathbf{s} - \mathbf{D}\mathbf{m}||_{\mathbf{2}}^{\mathbf{2}}, \tag{3.6}$$

where $||\cdot||_2^2$ defines the ℓ_2 norm, $\sum_k {x_k}^2$, and is called the minimum norm solution. The cost function defined in Equation 3.6 simply measures the error between the observed data $\mathbf{s}$ and the predicted data $\mathbf{Dm}$ through the ℓ_2 norm.

Solving for the minimum norm solution of the spectral decomposition problem requires the definition of the adjoint operator $\mathbf{D}^H$. Specifically, this means that a mapping from the data into the projected model space $\widehat{\mathbf{m}}$,

$$\widehat{\mathbf{m}} = \mathbf{D}^H\mathbf{s}, \tag{3.7}$$

is required. Since the adjoint operation of convolution is known to be cross-correlation (Claerbout, 1992), the operation described by $\mathbf{D}^H$ simply becomes the cross-correlation of each wavelet $w(t,k)$ in the wavelet dictionary with the seismic trace $s(t)$. To ensure the

operator $\mathbf{D}^H$ is truly the adjoint operation of $\mathbf{D}$, it must pass the dot product test (Claerbout, 1992). Essentially, $\mathbf{D}$ and $\mathbf{D}^H$ represent the discrete operators for the forward and inverse CWT, however, when they are implemented in an inversion process the solution for the local time-frequency representation can be constrained to exhibit certain characteristics. Like the CWT, local time-frequency representations generated through inversion with these operators are actually time-scale maps but, as discussed in Chapter 2, if the wavelet dictionary is chosen as a set of complex Ricker wavelets the resultant time-scale map can serve as an adequate proxy for the local time-frequency representation.

3.3 Damped Least Squares Inversion

The proposed method for spectral decomposition defines an inherently underdetermined problem as a one dimensional seismic trace is generating a two dimensional local time-frequency representation. With many more unknown variables (model) than known variables (data), an infinite number of local time-frequency representations could theoretically characterize the seismic trace. The minimum norm solution selects a specific solution for the local time-frequency representation that minimizes the error between the predicted and observed data, but it may be unstable in the presence of noise since the inverse of $\mathbf{D}^H\mathbf{D}$ might be poorly conditioned (has a determinant close to zero). This means that small changes in the observed data, whether noise related or not, could potentially create large changes in the local time-frequency representation. Through adding an additional constraint on the desired local time-frequency representation, the inversion problem becomes more well-posed with small changes in the observed data correlating to small changes in the resulting local time-frequency representation. The act of adding additional constraints on the solution of the inversion problem is known as regularization (Aster et al., 2005) and is accomplished by incorporating an additional term to the cost function J,

$$J = ||\mathbf{s} - \mathbf{D}\mathbf{m}||_2^2 + \lambda\mathcal{R}(\mathbf{m}). \tag{3.8}$$

The strength of the regularization term, $\mathcal{R}(\mathbf{m})$, in comparison to the data misfit term is balanced through the trade-off parameter λ. Although many different choices for the regularization term exist, its selection is based upon certain assumptions about the desired model. A common choice for $\mathcal{R}(\mathbf{m})$ is the ℓ_2 norm of the model parameters which minimizes their amplitude. Constraining the model to have a minimum amplitude in the ℓ_2 norm sense generates a smooth solution and prevents small variations within the data from having large effects on the model. Thus, the incorporation of the ℓ_2 norm of the model as the

regularization term alters the cost function J to become,

$$J = ||\mathbf{s} - \mathbf{D}\mathbf{m}||_2^2 + \lambda||\mathbf{m}||_2^2. \tag{3.9}$$

The minimization[1] of the cost function in Equation 3.9 results in a closed form solution known as the damped least squares solution (Claerbout, 1992),

$$\mathbf{m} = (\mathbf{D}^H\mathbf{D} + \lambda\mathbf{I})^{-1}\mathbf{D}^H\mathbf{s}, \tag{3.10}$$

where $\mathbf{I}$ represents the identity matrix of equivalent dimensions to $\mathbf{D}^H\mathbf{D}$. The addition of the ℓ_2 norm of the model parameters for the regularization term effectively adds a constant value along the diagonal of $\mathbf{D}^H\mathbf{D}$ to ensure its determinant is non-zero and stabilizes the inversion process. The selection of the trade-off parameter can be chosen through the χ^2 test if the noise level is known or heuristically, such as utilizing an L-curve, if the noise level is unknown as is often the case with real data (Hansen, 1998). Specifically, a larger value for λ puts greater emphasis on the regularization term, generates a smoother model, and assumes a greater amount of noise within the data while a smaller value for λ puts greater emphasis on the data misfit term and assumes a smaller amount of noise within the data. The trade-off parameter, therefore, controls the degree of fit between the predicted data and observed data through constraining the model solution and should be set such that the noise is not being fit within the model and that the predicted data adequately fits the observed data.

Although the damped least squares solution can be computed directly from the closed form solution given by Equation 3.10, it may not always be optimal for large scale problems since computational difficulties can arise if the inversion matrix becomes too large (Oldenburg et al., 1993). Alternatively iterative methods, such as the conjugate gradient method (Hestenes and Stiefel, 1952), can be employed to solve the damped least squares solution indirectly. The advantage of solving for the damped least squares solution iteratively is that the matrices for the forward and adjoint operator are not always required to be stored in memory. Instead these methods only require the result of applying the forward and adjoint operators. To increase the computational efficiency of applying forward and adjoint operators of the proposed spectral decomposition method, they can be computed in the Fourier domain with the aid of the Fast Fourier Transform (FFT) where convolution and cross-correlation simply become multiplications and multiplications by the complex conjugate, respectively.

[1]The minimization of a function is accomplished by taking its derivative and setting it equal to zero.

3.4 Sparse Inversion

Although the damped least squares solution generates a stable solution in the presence of noise, it does not generate a high resolution or sparse local time-frequency representation. Ideally, the regularization term for sparsity constrained inversion should be a direct measure of sparsity. This can be accomplished with the ℓ_0 norm which measures the number of non-zero elements within a function. Unfortunately, the only known method to solve the regularized inversion problem with the ℓ_0 norm is combinatorial and NP hard (Donoho, 2006b) leading it to be computationally inefficient. Therefore, other regularization terms such as the Cauchy and Huber norms (Sacchi, 1997) are often utilized to produce models which mimic the solution that would have been obtained with the ℓ_0 norm (i.e., the sparsest solution). Through the recent studies of compressive sensing (Candes and Tao, 2006), increased attention has been devoted to sparsity constrained inversion where the ℓ_0 norm has often been approximated by the ℓ_1 norm, $\sum_k |x_k|$, which under certain conditions has been found to also produce the sparsest solution (Donoho, 2006b). The sparse optimization problem involving the ℓ_1 norm is known as Basis Pursuit (Chen et al., 1998) and can be posed through the minimization of the cost function,

$$J = \frac{1}{2}||\mathbf{s} - \mathbf{D}\mathbf{m}||_2^2 + \lambda||\mathbf{m}||_1. \tag{3.11}$$

When minimized, the following solution for the local time-frequency representation is achieved,

$$\mathbf{m} = \left(\mathbf{D}^H\mathbf{D} + \lambda\mathbf{Q}\right)^{-1}\mathbf{D}^H\mathbf{s}, \tag{3.12}$$

where the matrix $\mathbf{Q}$ is a diagonal matrix whose elements are the reciprocal of the absolute value of $\mathbf{m}$. In other words, $Q_{jj} = \frac{1}{|m_j|}$ and is zero everywhere else. However, the direct presence of $\mathbf{m}$ in $\mathbf{Q}$ causes the solution given by Equation 3.11 to become non-linear since the model itself is required in the inversion. This property is a result of the non-quadratic nature of the ℓ_1 norm which does not allow a closed form solution, such as the damped least squares solution, and makes ℓ_1 norm constrained inversion more computationally expensive than ℓ_2 norm constrained inversion. Since a closed form solution does not exist, iterative algorithms must be employed to minimize the cost function of Equation 3.11.

One method for iteratively solving for the solution of ℓ_1 norm regularized inversion is the Iteratively Reweighted Least Squares (IRLS) method (Daubechies et al., 2010). In essence, the IRLS method assumes a starting model, such as a model full of zeros, and solves for the constrained model solution via,

$$\mathbf{m}_{i+1} = \left(\mathbf{D}^H\mathbf{D} + \lambda\mathbf{Q}_i\right)^{-1}\mathbf{D}^H\mathbf{s}, \tag{3.13}$$

where the index i represents the iteration number. For large scale problems, the right hand side of Equation 3.13 can be solved through methods such as conjugate gradients. The IRLS method can also be used with other regularization terms such as the Cauchy and Huber norms and has often been utilized for inverse problems within geophysics (Sacchi, 1997; Trad et al., 2003). Particularly, the IRLS method with a pseudo ℓ_0 norm, $\sum_k \frac{|m_k|^2}{|m_k|^2+\epsilon^2}$, has been applied for sparsity constrained spectral decomposition on seismic data which utilizes the operators described in Section 3.2 (Portniaguine and Castagna, 2004).

For the purposes of this thesis, another iterative solver which minimizes Equation 3.11 was utilized to help introduce different sparse solving methods involving the ℓ_1 norm within the geophysics community. The Fast Iterative Soft Thresholding Algorithm (FISTA) (Beck and Teboulle, 2009) is an adaptation of the Iterative Soft Thresholding Algorithm (ISTA) (Daubechies et al., 2004) which is a proximal gradient descent method. The gradient descent method, also known as the Landweber iteration method (Landweber, 1951), iteratively solves the minimization of cost functions through moving negatively along their gradient via,

$$\mathbf{m}_{i+1} = \mathbf{m}_i - \frac{1}{\alpha}\nabla J\left(\mathbf{m}_i\right), \tag{3.14}$$

which for the case of Equation 3.6 becomes,

$$\mathbf{m}_{i+1} = \mathbf{m}_i + \frac{1}{\alpha}\mathbf{D}^H\left(\mathbf{s} - \mathbf{D}\mathbf{m}_i\right). \tag{3.15}$$

The parameter α in Equations 3.14 and 3.15 controls the iteration step length and thus the convergence rate. Through moving negatively along the gradient of the cost function, the gradient descent method iteratively estimates models which become closer to the minimum of the actual cost function as illustrated by Figure 3.1. The gradient descent method can also be equivalently posed through the proximal regularization of the cost function at $\mathbf{m_{i-1}}$ according to (Beck and Teboulle, 2009),

$$\mathbf{m}_i = \min\left\{J\left(\mathbf{m}_{i-1}\right) + \langle \mathbf{m} - \mathbf{m}_{i-1}, \nabla J\left(\mathbf{m}_{i-1}\right)\rangle + \frac{\alpha}{2}||\mathbf{m} - \mathbf{m}_{i-1}||_2^2\right\}. \tag{3.16}$$

Effectively, a non-negative function of $\mathbf{m}$ is added to the cost function to create a new cost function whose values are always greater than the original cost function except at the point $\mathbf{m}_{i-1}$ where the original and new cost functions are equal. The goal is that this new cost function is much simpler to minimize than the original cost function, which is often the case if the original cost function is non-linear. Through continually minimizing these new cost functions, $\mathbf{m}_i$ iteratively moves closer to the minimum of the original cost function. If the constant terms within the minimization of Equation 3.16 are ignored, it can be more

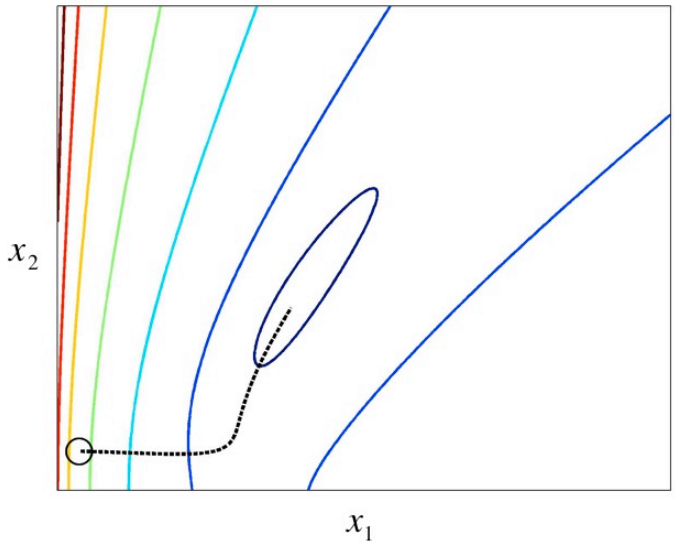


Figure 3.1: An example of the gradient descent method. The coloured contour lines represent the cost function for a model $\mathbf{x}$ that has two variables, x_1 and x_2. The initial estimate of the solution is marked by the open black circle and the dotted line represents the path of the gradient descent algorithm towards the minimum of the cost function.

elegantly rewritten as,

$$\mathbf{m}_i = \min \left\{ \frac{\alpha}{2} ||\mathbf{m} - [\mathbf{m}_{i-1} - \frac{1}{\alpha} \nabla J(\mathbf{m}_{i-1})]||_2^2 \right\}. \tag{3.17}$$

Since the ℓ_1 norm for the cost function in Equation 3.11 is separable, its minimization is not affected by the minimization of the data misfit term. Therefore, the minimization of a slightly easier cost function,

$$J = ||\mathbf{s} - \mathbf{m}||_2^2 + \lambda ||\mathbf{m}||_1, \tag{3.18}$$

is studied to serve as an analogy for minimizing Equation 3.11 (Selesnick, 2009). For this particular example, it is assumed that $\mathbf{s}$ and $\mathbf{m}$ have the same dimensions. Taking the derivative of Equation 3.18 and setting it equal to zero results in the expression,

$$\mathbf{s} = \mathbf{m} + \frac{\lambda}{2} \text{sgn}(\mathbf{m}), \tag{3.19}$$

where the function sgn$(\cdot)$ produces $+1$ for positive numbers and -1 for negative numbers.

The solution for $\mathbf{m}$ from Equation 3.19 is defined by,

$$\mathbf{m} = \mathcal{T}_{\lambda/2}\left(\mathbf{s}\right), \tag{3.20}$$

where $\mathcal{T}_{\lambda}(\cdot)$ is the soft thresholding function which is also known as the shrinkage thresholding function. For a complex number $Ae^{i\theta}$, the soft thresholding function is defined as,

$$\mathcal{T}_{\lambda}\left(Ae^{i\theta}\right) = \begin{cases} \left(A-\lambda\right)e^{i\theta} & \text{if } A > \lambda \\ 0 & \text{if } A \leq \lambda \end{cases}. \tag{3.21}$$

Through this analogy, the minimization of Equation 3.11 becomes,

$$\mathbf{m} = \mathcal{T}_{\lambda/2}\left(\nabla\left(||\mathbf{s}-\mathbf{D}\mathbf{m}||_2^2\right)\right). \tag{3.22}$$

Since $\nabla\left(||\mathbf{s}-\mathbf{D}\mathbf{m}||_2^2\right)$, or essentially the minimization of the cost function in Equation 3.6, can be accomplished through the gradient descent method, the iterative solution for the minimization of Equation 3.11 provided by ISTA becomes,

$$\mathbf{m}_{i+1} = \mathcal{T}_{\lambda/2\alpha}\left(\mathbf{m}_i + \frac{1}{\alpha}\mathbf{D}^H\left(\mathbf{s}-\mathbf{D}\mathbf{m}_i\right)\right), \tag{3.23}$$

or simply the soft threshold of each gradient step. Since ISTA does not require the explicit matrices for $\mathbf{D}$ and $\mathbf{D}^H$, just their application, the utilization of operators as discussed in Section 3.3 can be incorporated into the implementation of this algorithm. To ensure the convergence of this algorithm the parameter α must be constrained to be equal to or greater than the maximum eigenvalue of $\mathbf{D}^H\mathbf{D}$. However, since the explicit matrices for $\mathbf{D}$ and $\mathbf{D}^H$ are not required to be defined, the estimation of α can become difficult. To help combat this problem, the power iteration method (Golub and Loan, 1996) can be utilized to estimate the largest eigenvalue of $\mathbf{D}^H\mathbf{D}$.

Despite the simplicity of this first order method[2], ISTA has generally been recognized as a slow method with a sub-linear global rate of convergence (Beck and Teboulle, 2009). The simplicity of applying this algorithm has lead to several alterations to help increase its global rate of convergence such as the TWIST (Two step Iterative Shrinkage Thresholding) (Bioucas-Dias and Figueiredo, 2007) and FISTA (Fast Iterative Shrinkage Thresholding Algorithm) (Beck and Teboulle, 2009) algorithms. As per Beck and Teboulle (2009), FISTA globally converges at a faster rate than the ISTA and TWIST algorithms and is shown to experimentally take significantly fewer iterations to reach an acceptable level of accuracy. FISTA accomplishes this task by utilizing a carefully chosen update to $\mathbf{m}_i$ before $\mathbf{m}_{i+1}$ is

[2]First order methods are optimization methods that are based on function values and gradient evaluations (Beck and Teboulle, 2009).

computed. This update, $\mathbf{y}_i$, is a particular combination of the previous two iterations given by,

$$\mathbf{y}_i = \mathbf{m}_i + \left(\frac{\xi_i - 1}{\xi_{i+1}}\right)(\mathbf{m}_i - \mathbf{m}_{i-1}), \tag{3.24}$$

where ξ_i is an update parameter determined by,

$$\xi_{i+1} = \frac{1 + \sqrt{1 + 4\xi_i^2}}{2}\,. \tag{3.25}$$

By incorporating $\mathbf{y}_i$ rather than $\mathbf{m}_i$ to solve for $\mathbf{m}_{i+1}$, FISTA alters ISTA to become,

$$\mathbf{m}_{i+1} = \mathcal{T}_{\lambda/2\alpha}\left(\mathbf{y}_i + \frac{1}{\alpha}\mathbf{D}^H(\mathbf{s} - \mathbf{D}\mathbf{y}_i)\right). \tag{3.26}$$

This clever update, $\mathbf{y}_i$, enables FISTA to improve the rate of convergence of ISTA from approximately $O(1/i)$ to approximately $O(1/i^2)$ without sacrificing the simplicity of calculating one gradient step per iteration.

3.5 Discussion

Although the two regularization terms that have been discussed in this chapter attempt to measure the same property, the amplitude of the model parameters, the resultant solutions are quite different. This effect is demonstrated in Figures 3.2 and 3.3 where the proposed inverse spectral decomposition method based upon the multi-wavelet convolutional model was applied to the synthetic time-series depicted in Figure 2.1. Similar to the application of the CWT in Chapter 2, the wavelet dictionary was chosen to be composed of complex Ricker wavelets which are uniquely defined by their central frequency and the resulting local time-frequency representation obtained is actually a time-scale map. However, as shown in Figure 2.6, the time-scale map obtained with a complex Ricker wavelet dictionary serves as an adequate proxy to its mapped local time-frequency representation. By imposing the ℓ_1 norm as the regularization term, the amplitude of the obtained local time-frequency representation was more efficiently localized, or compacted, around the actual events of the time-series than the local time-frequency representation obtained with the ℓ_2 norm. The stronger localization of the local time-frequency representation did not strongly influence the phase around these areas, however, a difference is seen in areas where the signal should not have a strong influence on the local time-frequency representation. This difference arises from the sparse model that is created through the use of the ℓ_1 norm as the regularization term, since values of zero amplitude are assigned a zero phase. Likewise, the regularization with the ℓ_2 norm generates a phase for the entire local time-frequency representation since its model contains mostly non-zero coefficients which are associated with a phase, no matter

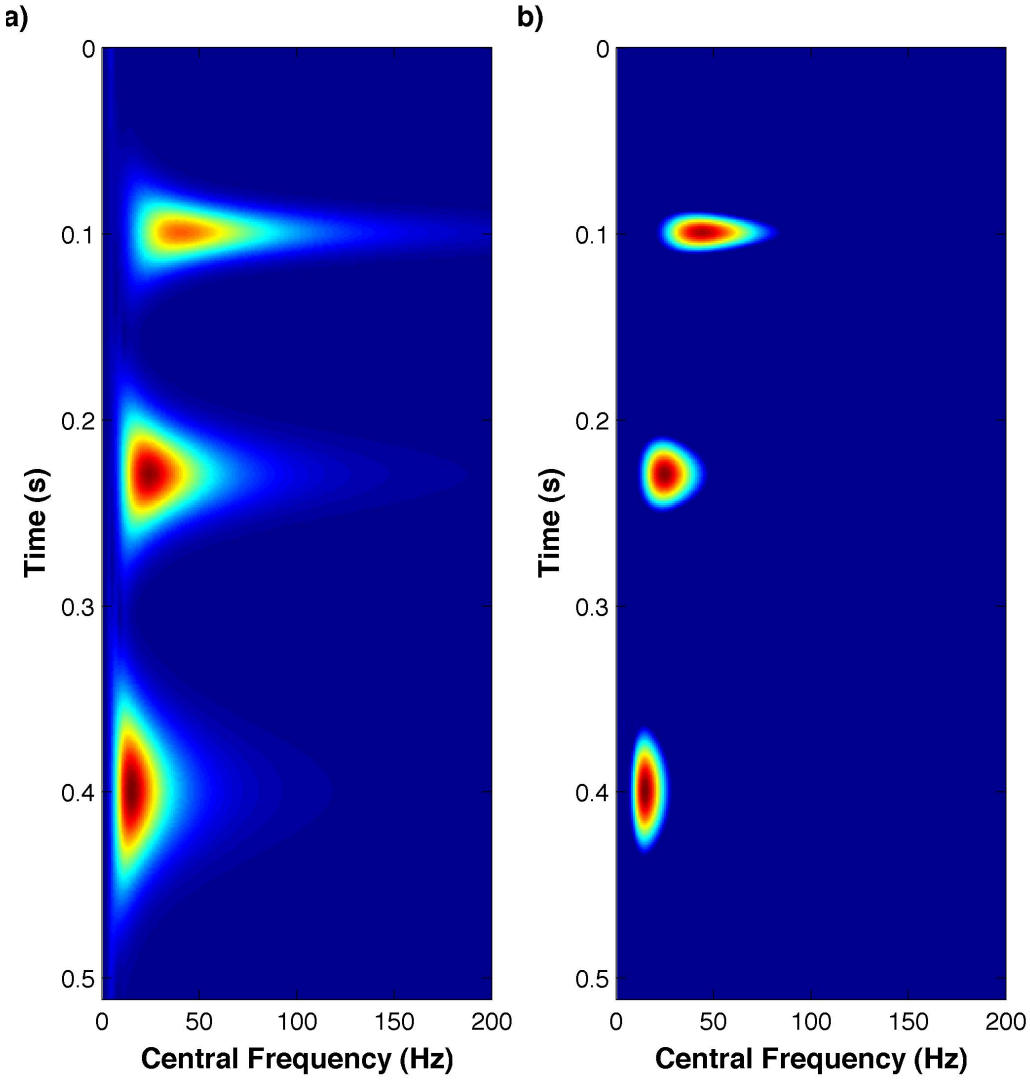


Figure 3.2: The amplitude of the local time-frequency representation obtained through inversion using a) the ℓ_2 norm and b) the ℓ_1 norm as the regularization term.

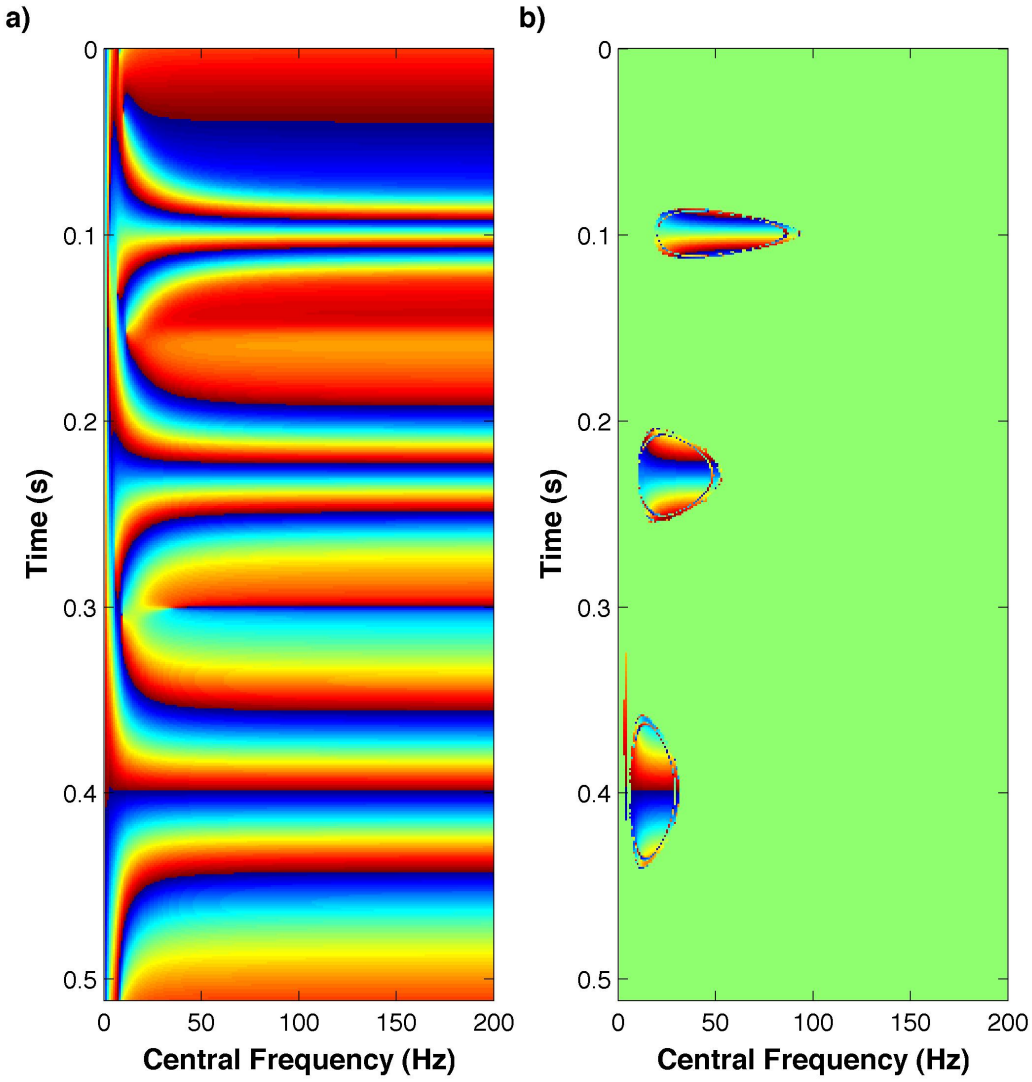


Figure 3.3: The phase of the local time-frequency representation obtained through inversion using a) the ℓ_2 norm and b) the ℓ_1 norm as the regularization term.

how minutely small they are.

The reason for the large difference in the solution from ℓ_2 norm and ℓ_1 norm regularization can be explained through Figure 3.4. Since the ℓ_2 norm utilizes the square of the model parameters, its minimization has the effect of penalizing larger values of the model more than smaller values and thus creates a smooth solution to help avoid large amplitudes. Conversely, the ℓ_1 norm utilizes the absolute value of the model parameters and tends to promote larger values and fewer smaller values in the model than the ℓ_2 norm when minimized. This behaviour of the ℓ_1 norm causes its solution to become more sparse since a larger coefficient value for a single model parameter has the ability to represent a larger proportion of the signal. These properties can be explicitly seen in Figure 3.4 where the dashed line representing the absolute value, or ℓ_1 norm, lies above the solid line representing the squared value, or ℓ_2 norm, for small values ($\mathbf{x} < 1$) to create a stronger penalty of small values for the ℓ_1 norm and below the solid line for large values ($\mathbf{x} > 1$) to create a stronger penalty of large values for the ℓ_2 norm.

It should be noted that the direct generation of a sparse local time-frequency representation using the operators in Section 3.2 can also be accomplished using the Matching Pursuit algorithm. Matching Pursuit is a greedy algorithm introduced by Mallat and Zhang (1993)

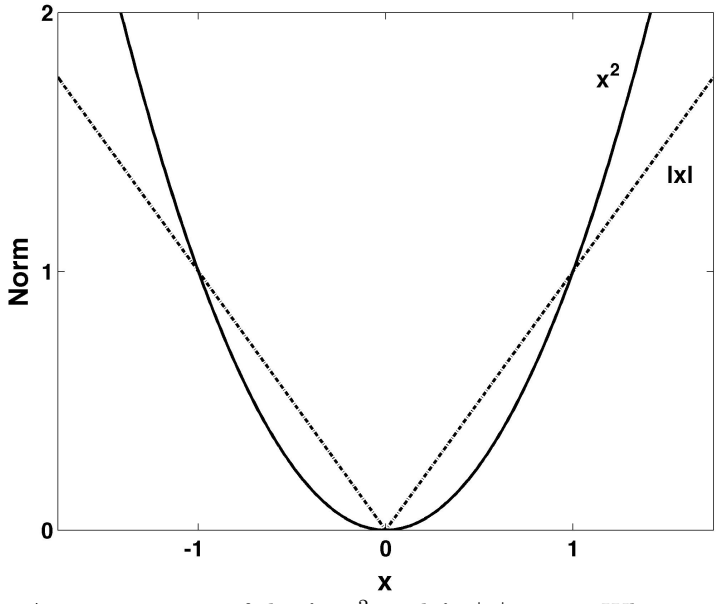


Figure 3.4: A representation of the ℓ_2, $\mathbf{x}^2$, and ℓ_1, $|\mathbf{x}|$, norm. When minimized, the ℓ_1 norm will support fewer small coefficients compared to the ℓ_2 norm causing it to appear more sparse.

that was originally proposed for fitting time-frequency atoms, or wavelets, to a time-series. This heuristic or empirical method generates a model to fit the data by selecting the single best time-frequency atom which represents the signal. Once this atom is subtracted from the signal, Matching Pursuit then selects the single best time-frequency atom which represents the residual (the part of the signal not represented by the model). Through iteratively continuing this approach, the newly constructed model will eventually fit the original signal. By populating the model one coefficient at a time, Matching Pursuit builds the model by fitting the most major elements of the signal first. This has the effect of making the model look sparse, although sparsity is not a specific goal of Matching Pursuit algorithm. Matching Pursuit's generation of a sparse model is in direct contrast to Basis Pursuit, or the minimization of the cost function in Equation 3.11, whose specific goal is to generate a sparse model based upon the minimization of the ℓ_1 norm. The difference between the goals of Matching Pursuit and Basis Pursuit means that Basis Pursuit can theoretically yield representations that are more sparse since it globally optimizes the model during each iteration rather than Matching Pursuit's local optimization (Mallat, 2008) . However, the Matching Pursuit method has shown the ability to generate sparse local time-frequency representations for seismic data (Chakraborty and Okaya, 1995; Castagna et al., 2003).

3.6 Summary

This chapter poses the spectral decomposition problem as an inverse problem through utilizing a multi-wavelet convolutional model. The forward problem, which maps the model (local time-frequency representation) to the data (time-series), requires the summation of each wavelet within a wavelet dictionary after it has been convolved with its respective wavelet-dependent reflectivity series. If the wavelet dictionary is chosen strategically, such a dictionary of complex Ricker wavelets which are uniquely defined by their central frequency, the set of wavelet-dependent reflectivity series can represent a local time-frequency representation. For the adjoint problem or the mapping from the data to the model, the adjoint operator is defined as the cross-correlation of each wavelet with the data. By defining the spectral decomposition problem in this manner, the forward and adjoint operators simply become the inverse and forward operators for the CWT, respectively.

To create a high resolution local time-frequency representation, the inverse spectral decomposition problem is regularized by the ℓ_1 norm. Garnering increased popularity through the introduction of compressive sensing, ℓ_1 norm regularized inversion constrains the solution or local time-frequency representation to be sparse and thus at a higher resolution. This is in direct contrast to the more common and less computationally expensive ℓ_2 norm regularized inversion which generates a smooth solution by deterring large amplitudes within

the model. To solve the ℓ_1 norm regularized inversion problem, iterative techniques such as Iteratively Reweighted Least Squares (IRLS) or the Fast Iterative Soft Thresholding Algorithm (FISTA) are required since the solution to the problem is non-linear. Since IRLS has already been applied within the geophysics community, FISTA was utilized within this thesis to solve all ℓ_1 regularized inverse problems to help introduce a new sparse solver. One particular advantage with the utilization of FISTA in comparison with IRLS is the reduced number of estimated parameters (trade-off parameter and number of iterations vs. trade-off parameter, number of iterations to update $\mathbf{Q}$, and number of iterations for conjugate gradients). The application of this strategy to generate a high resolution local time-frequency representation on seismic data is described in the next chapter.

CHAPTER 4

Spectral Decomposition of Seismic Data

4.1 Introduction

Local time-frequency analysis has proven to be a powerful interpretational tool for seismic data. By generating the complete frequency spectrum at each time sample rather than just generating a scalar parameter with the instantaneous frequency, spectral decomposition has provided a greater wealth of information to be studied and analyzed for exercises such as reservoir characterization (Partyka et al., 1999). Of particular interest, spectral decomposition has been utilized for thin bed analysis (Partyka et al., 1999; Marfurt and Kirlin, 2001), the identification of subsurface structures such as paleochannels (Giroldi and Alegria, 2005; Liu and Marfurt, 2007) and reefs (Li and Zheng, 2008), for measuring seismic attenuation (Reine et al., 2009), and as a possible indicator for hydrocarbon presence based upon low frequency shadows (Castagna et al., 2003) or high frequency signatures (Li et al., 2011) depending upon the local geology. Localized time-frequency representations have also contributed to more detailed interpretations when used in conjunction with AVO analysis (Odebeatu et al., 2006) and more recently, the phase residues from spectral decomposition have been utilized to determine the location of stratigraphic discontinuities (de Matos et al., 2011). Spectral decomposition has also found applications in other areas of geophysics such as global seismology (Vasudevan and Cook, 2001), potential field data such as gravity or magnetic data[1] (Hornby et al., 1999; Martelet et al., 2001), and ground penetrating radar

[1]Note that local time-frequency analysis changes to local space-wavenumber analysis for potential field data.

(Bradford and Wu, 2007; Bradford, 2007).

Utilizing the real data set in Figure 1.3, the application of the inverse spectral decomposition method described in Chapter 3 is demonstrated. Through studying the local time-frequency representation of a cross section of this data set, the different frequency behaviour of the two large paleochannels can be seen. A comparison of ℓ_2 norm regularized inversion and ℓ_1 norm regularized inversion of this cross section is also conducted to highlight the advantages and disadvantages of both techniques. The local time-frequency representation of a time slice through these paleochannels further corroborates these findings. Visualization of a smaller paleochannel on a time slice is also enhanced through the high resolution spectral decomposition method. Finally, a technique to help deal with viewing the large data volume that is created through local time-frequency analysis is presented.

4.2 Cross Section Analysis

An example cross section shown in Figure 4.1 through the example real 3D data set displays two large paleochannels located at approximately 0.540s around 2.25km and 0.425s around 0.5km and 1.5km. As spectral decomposition adds another dimension, frequency, to the data set, local time-frequency representations become position-time-frequency cubes for a seismic

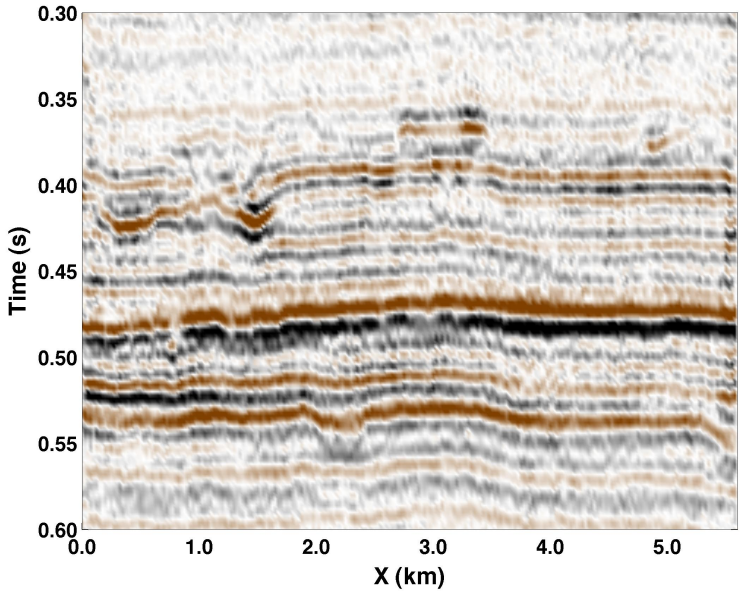


Figure 4.1: An example cross section through the a real 3D data set displaying paleochannels around 0.425s and 0.540s.

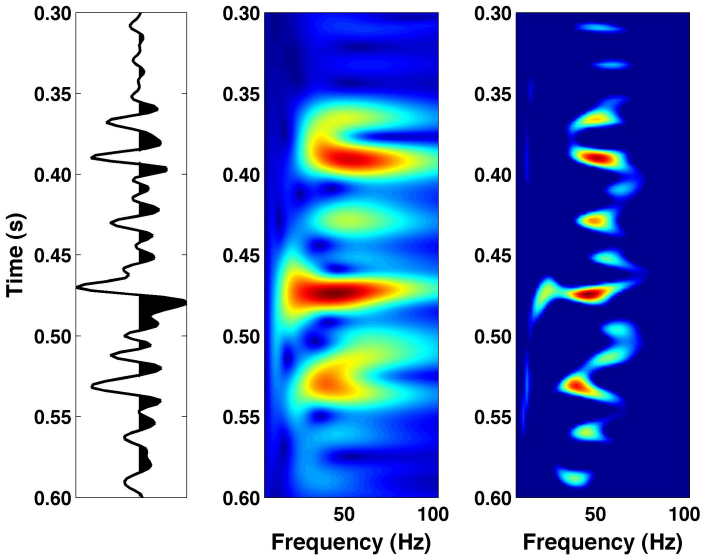


Figure 4.2: Local time-frequency representation from the seismic trace located at 2.2km on the cross section in Figure 4.1 (left) using ℓ_2 norm regularized inversion (center) and ℓ_1 norm regularized inversion (right).

cross section. Therefore, to visualize the spectrally decomposed data, slices of constant time, position, or frequency must be taken. Figure 4.2 depicts the local time-frequency representation of this seismic cross section for both ℓ_2 norm and ℓ_1 norm regularized inversion by maintaining the position dimension to be constant at 2.2km. Further comparison of ℓ_2 norm and ℓ_1 norm regularized inversion is provided through two constant frequency[2] slices of the local time-frequency representation depicted in Figures 4.3 and 4.4. As with the synthetic time-series example in Chapter 3, the amplitude of the local time-frequency representation contains a much higher resolution in the ℓ_1 norm solution which allows for greater variability between different frequency slices when compared to the ℓ_2 norm solution. Although the ℓ_2 norm solution is not as sparse or compact as the ℓ_1 norm solution, its global phase image is more complete since almost none of the coefficients in the local time-frequency representation are zero. This means that phase information can be calculated across the entire local time-frequency representation because phase information is not sensitive to the magnitude of the coefficient. In fact, the localized phase information of the ℓ_1 norm solution for the local time-frequency representation can be exactly found with the more global phase information of the ℓ_2 norm solution's local time-frequency representation. The advantages and disadvantages for both the ℓ_1 norm regularized inversion solution and ℓ_2 norm regularized

[2]Note, as in previous chapters, this frequency actually corresponds to the central frequency of a complex Ricker wavelet.

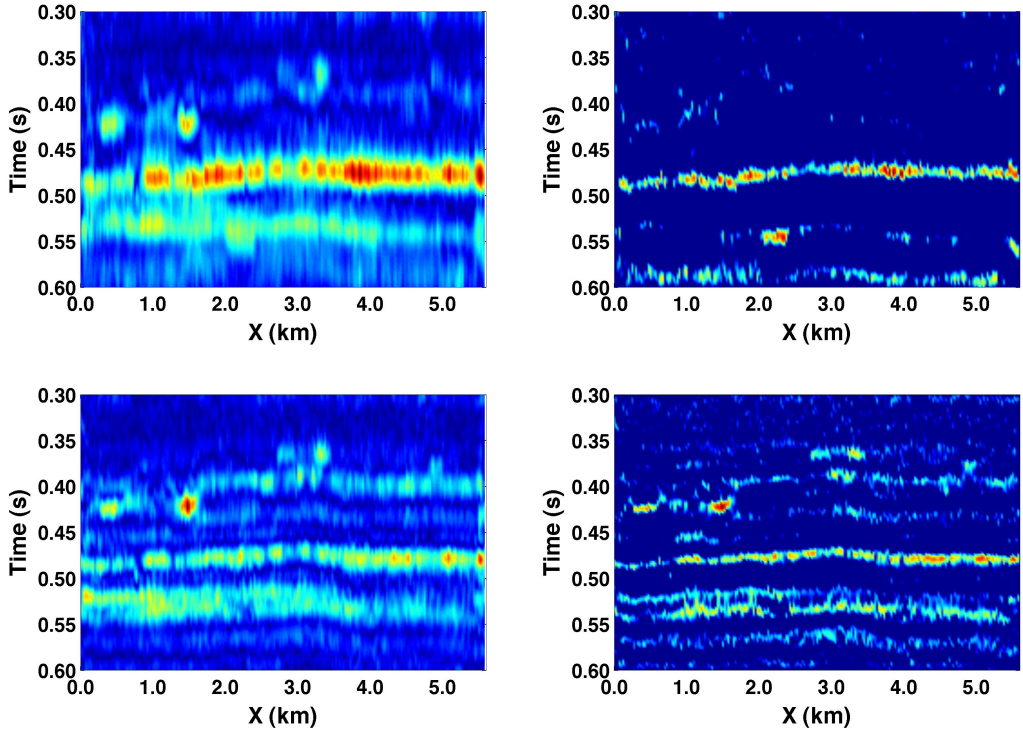


Figure 4.3: Comparison of local time-frequency representation for the cross section in Figure 4.1 obtained using the ℓ_2 norm (left) and the ℓ_1 norm (right) for regularization term. The amplitude of the constant 25Hz frequency slice (top) and the constant 45Hz frequency slice (bottom) are displayed.

inversion solution for the local time-frequency representation is summarized in Table 4.1.

A set of four constant frequency slices are highlighted in Figures 4.5 and 4.6 to display the change in frequency content of the example cross section. The amplitude of the constant frequency slices shown in Figure 4.5 were obtained with the ℓ_1 norm solution and depict two different frequency behaviours for the large paleochannels. The upper paleochannel around 0.425s contains a much stronger lower frequency content than its neighbouring reflectors but relatively maintains its higher frequency content. This lower frequency shadow behaviour is not displayed in the lower paleochannel which is dominated by a lower frequency content and an absence of higher frequency content when compared to its neighbouring reflectors. The contrast of frequency content is especially highlighted by the 25Hz and 45Hz sections. Figure 4.6 displays the ℓ_2 norm solution's phase and characterizes different scales of stratigraphic layering. This property may improve interpretation in regions where thin bed tuning effects are dominant such as the pinch out event located around 0.5s in the example cross

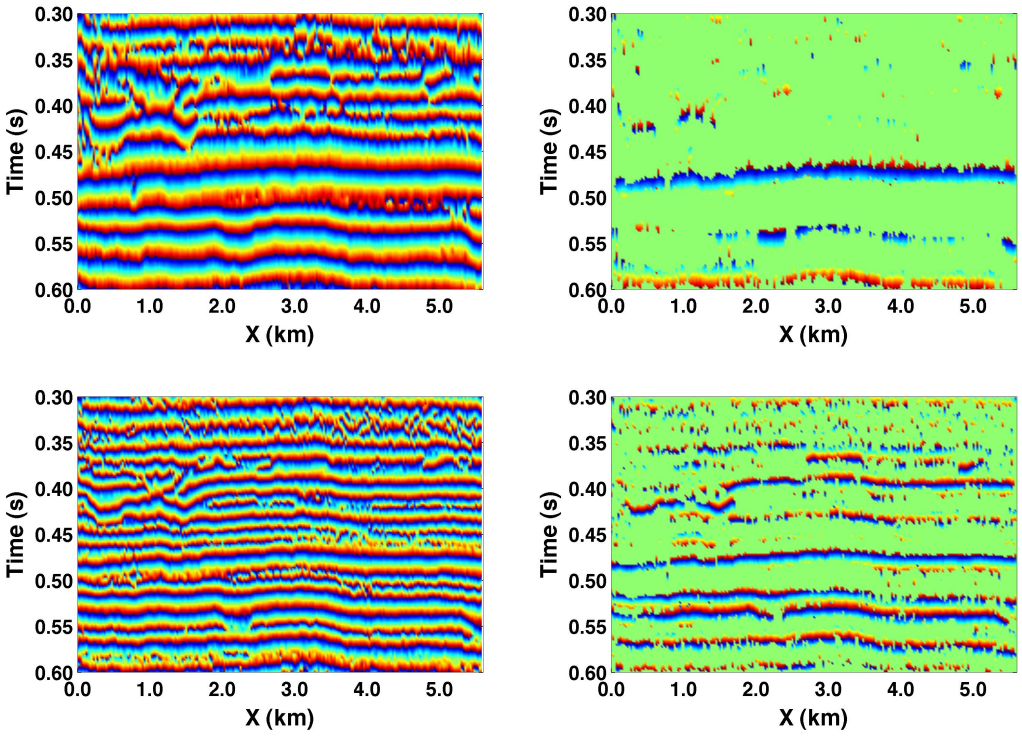


Figure 4.4: Comparison of local time-frequency representation for the cross section in Figure 4.1 obtained using the ℓ_2 norm (left) and the ℓ_1 norm (right) for regularization term. The phase of the constant 25Hz frequency slice (top) and the constant 45Hz frequency slice (bottom) are displayed.

	Amplitude	**Phase**
ℓ_2 **Norm Solution**	Low resolution with slight variability between constant frequency slices	Continuous phase distribution enables global and local interpretation
ℓ_1 **Norm Solution**	High resolution with increased variability between constant frequency slices	Localized phase distribution enables only local phase interpretation

Table 4.1: Comparison of the local time-frequency representation obtained from ℓ_2 norm regularized inversion and ℓ_1 norm regularized inversion for the interpretation of its amplitude and phase.

section which occurs around 4.0km in the 25Hz frequency slice and around 2.0km in the 45Hz frequency slice. By generating a local time-frequency representation, the different frequency behaviours of these large paleochannels, which could not be clearly interpreted on the original cross section, become apparent.

4.3 Time Slice Analysis

In actuality, the three dimensional data set containing the paleochannels becomes a four dimensional local time-frequency representation after undergoing local time-frequency analysis. Therefore, not only does the previous section maintain a constant frequency to produce an image but also fixed one of the spatial coordinates. Instead of fixing one of the spatial coordinates, the time coordinate can be fixed to provide an analysis through constant time sections, or time slices, of the data such as Figure 4.7 which depicts the lower large paleochannel that was seen in the example cross section. The amplitude of the ℓ_1 norm regularized inversion local time-frequency representation for this time slice is shown in Figure 4.8 for different constant frequency sections. The dominant frequency content of this paleochannel is found to be around 25Hz whereas the energy from its neighbouring reflectors is more dispersed among the higher frequencies, supporting the results from the example cross section. The phase of the ℓ_2 norm regularized inversion local time-frequency represenation, shown in Figure 4.9, defines the paleochannel across all of the constant frequency sections, although different portions of the paleochannel are more clearly defined in particular constant frequency sections. For example, the upper portion of the paleochannel is well

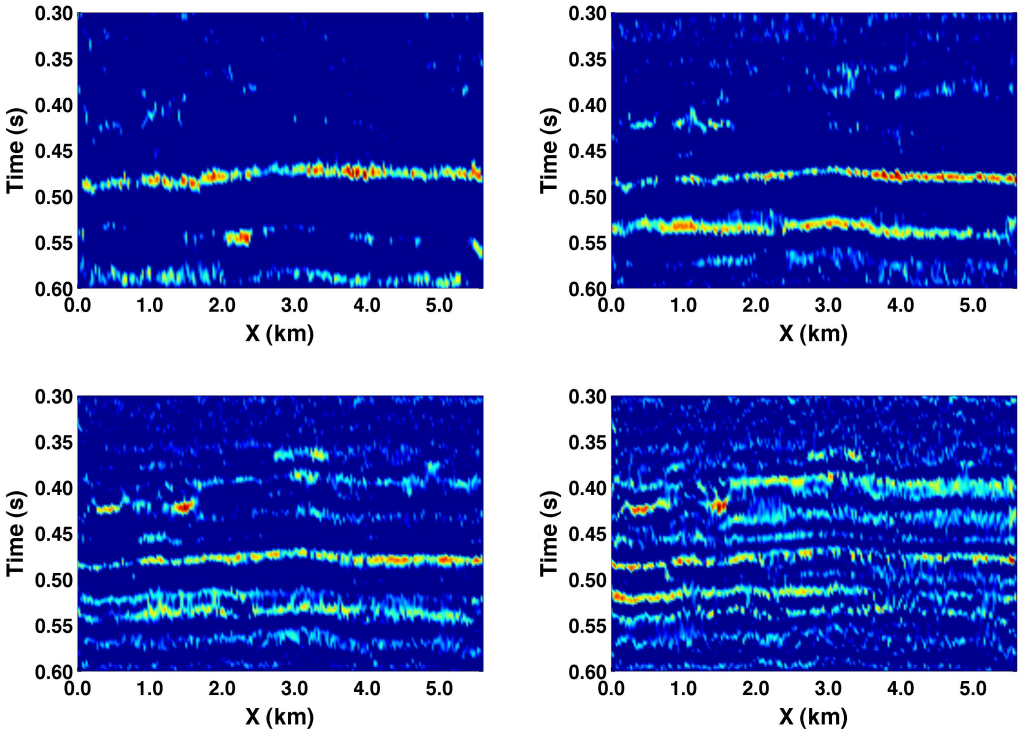


Figure 4.5: Amplitude of the ℓ_1 norm local time-frequency representation for the example cross section in Figure 4.1. Top left: 25Hz slice, Top right: 35Hz slice, Bottom left: 45Hz slice, Bottom right: 55Hz slice.

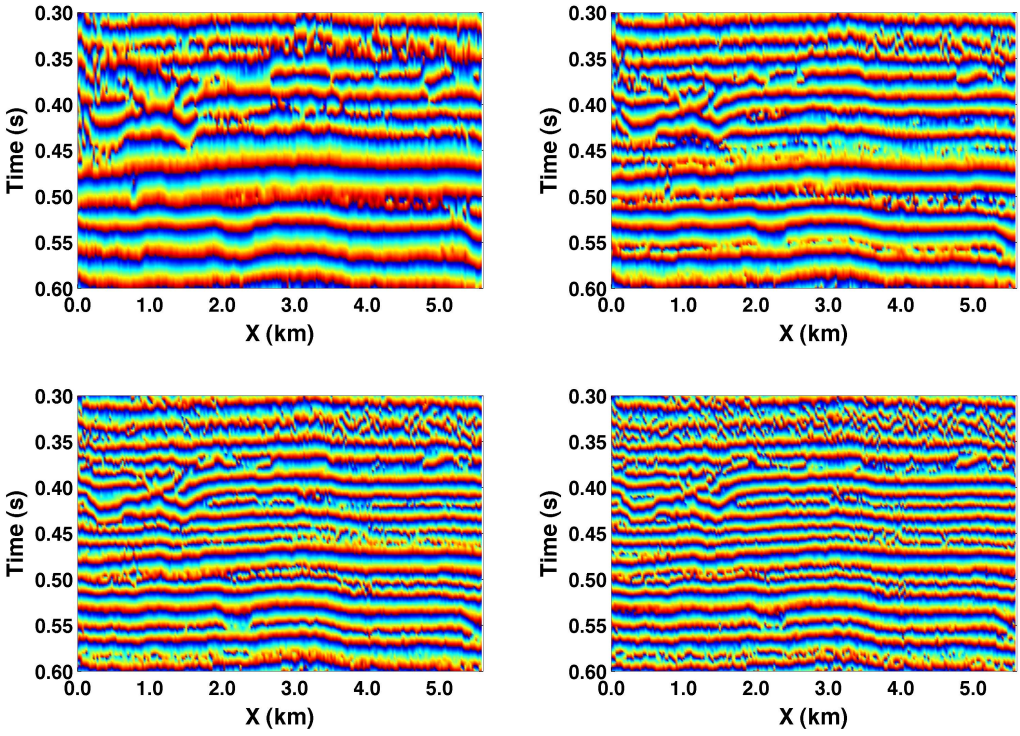


Figure 4.6: Phase of the ℓ_2 norm local time-frequency for the example cross section in Figure 4.1. Top left: 25Hz slice, Top right: 35Hz slice, Bottom left: 45Hz slice, Bottom right: 55Hz slice.

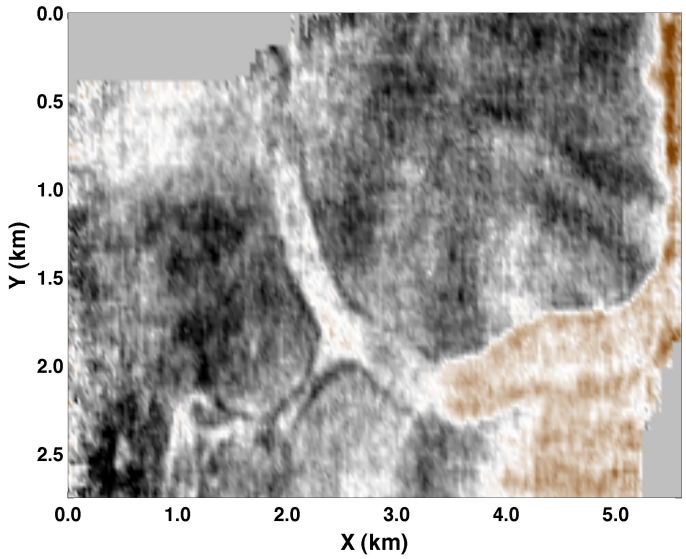


Figure 4.7: Time slice of the 3D data set displaying the lower channel (0.544s).

defined in the constant 25Hz frequency section as compared to the 45Hz frequency section, whereas the 45Hz frequency slice defines the lower lefthand portion of the paleochannel more clearly in comparison with the 25Hz frequency section. The generally sharp phase contrast found at the paleochannel boundaries of the different constant frequency sections may lead to improved edge detection results for the paleochannel when compared to edge detection utilizing only the seismic amplitude volume.

The final example with this real 3D data set illustrates how spectral decomposition can be used to enhance the imaging of small scale features. Figure 4.10 is a time slice between the two large paleochannels in the example cross section of Figure 4.1 and vertically displays a small channel-like structure. This structure has the base of the upper large paleochannel, which runs horizontally, overprinting it and also suffers from a polarity reversal of its surroundings near the bottom of the image. The small channel appears with better resolution in Figure 4.11 by an apparent lack of energy in the amplitude of the ℓ_1 norm regularized inversion local time-frequency representation's constant 45Hz frequency section while different portions of the channel also appear with the same lack of energy in the amplitude of the constant 25Hz and 35Hz frequency sections. Unlike the improved visualization from a lack of energy in the amplitude of the ℓ_1 norm regularized inversion local time-frequency representation, the phase of the ℓ_2 norm regularized inversion local time-frequency representation solution in Figure 4.12, does not provide any significant improvements for imaging this small scale feature.

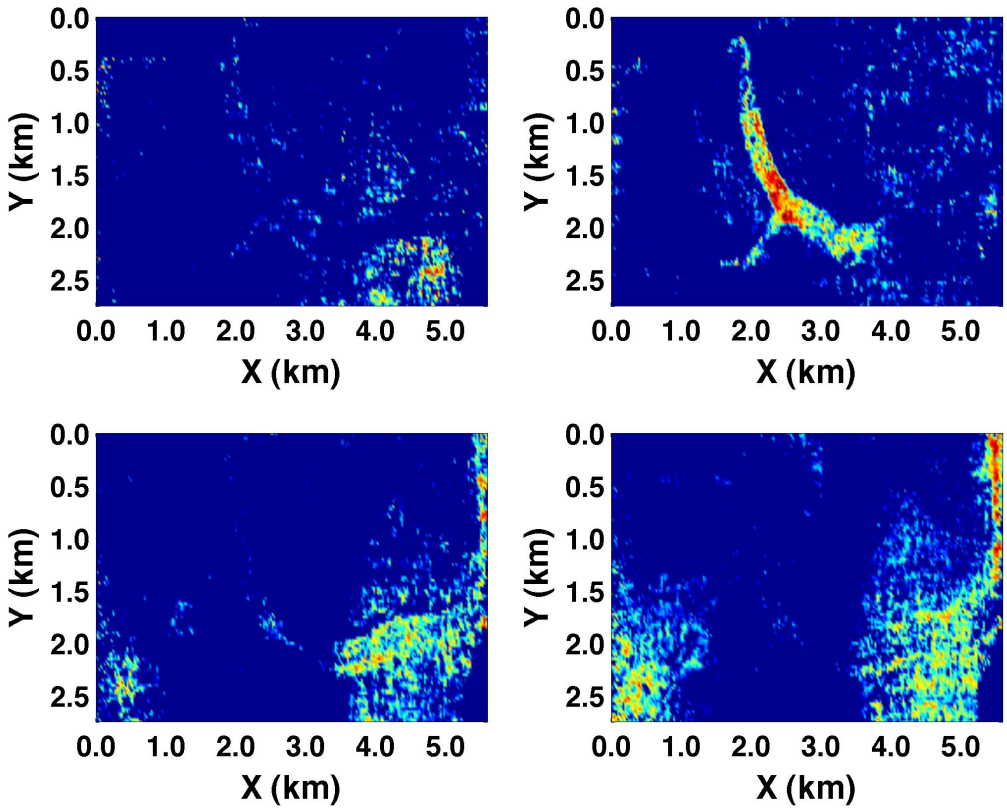


Figure 4.8: Amplitude of the ℓ_1 norm local time-frequency representaion for the time slice in Figure 4.7. Top left: 15Hz slice, Top right: 25Hz slice, Bottom left: 35Hz slice, Bottom right: 45Hz slice.

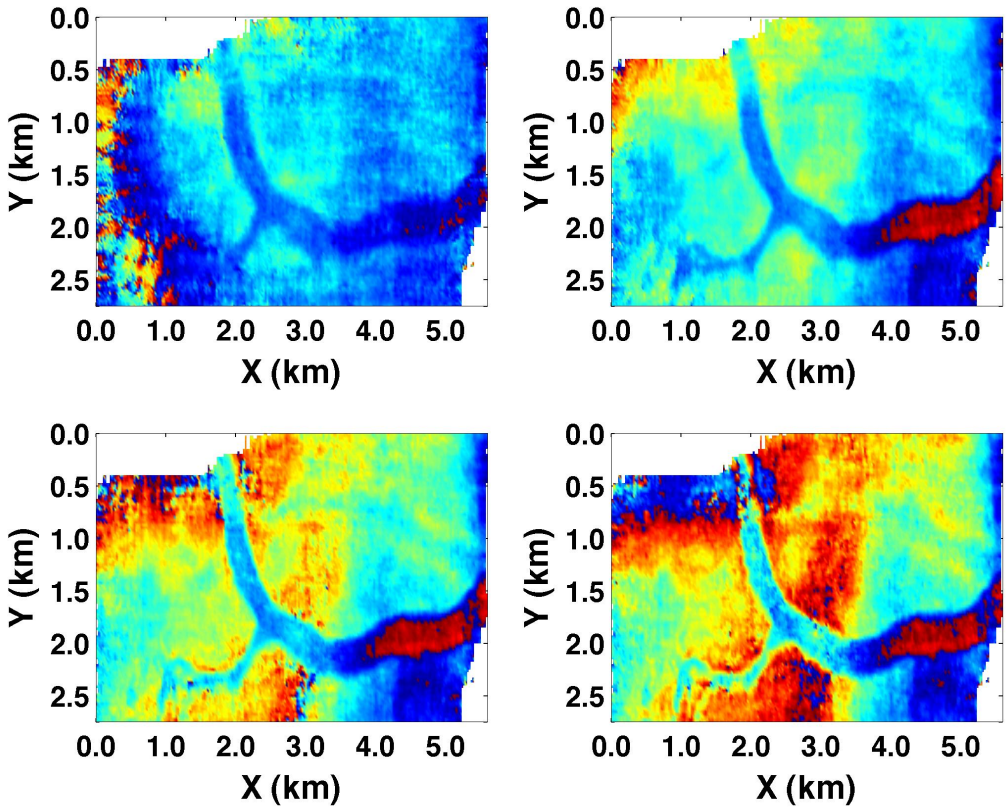


Figure 4.9: Phase of the ℓ_2 norm local time-frequency representation for the time slice in Figure 4.7. Top left: 15Hz slice, Top right: 25Hz slice, Bottom left: 35Hz slice, Bottom right: 45Hz slice.

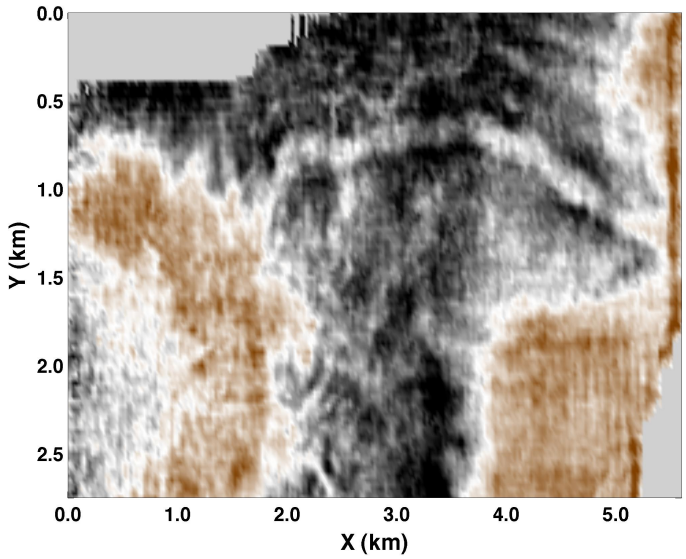


Figure 4.10: Time slice of the 3D dataset displaying a small channel-like structure travelling vertically through the image (0.478s).

4.4 RGB Image Blending

To fully image the small paleochannel from the ℓ_1 norm regularized inversion local time-frequency representation, one must look at each individual frequency section which may be a difficult task as a large amount of different frequency sections are generated during the spectral decomposition process. The issue of trying to compress all of the spectral information obtained through local time-frequency analysis can be partially addressed through the creation of a blended RGB image (Henderson et al., 2007). In an RGB image each pixel is uniquely defined by three parameters, the intensity of the colours red, green, and blue which each have a range of [0 1]. Through the summation of the amplitude from all of the constant frequency sections within a specific frequency band and its subsequent normalization, the local time-frequency representation can be condensed into the three RGB parameters where the red intensity of a pixel represents the intensity of low frequencies, the green intensity represents the intensity of intermediate frequencies, and the blue intensity represents the intensity of high frequencies. Although this procedure will generate a blended RGB image of low, intermediate, and high frequencies for a local time-frequency representation, the resultant RGB image might not exhibit significant colour variation due to most pixel values having a similar intensity. To help enhance the contrast of intensities for the pixels on an RGB image, a contrast stretch can be applied to each individual colour component

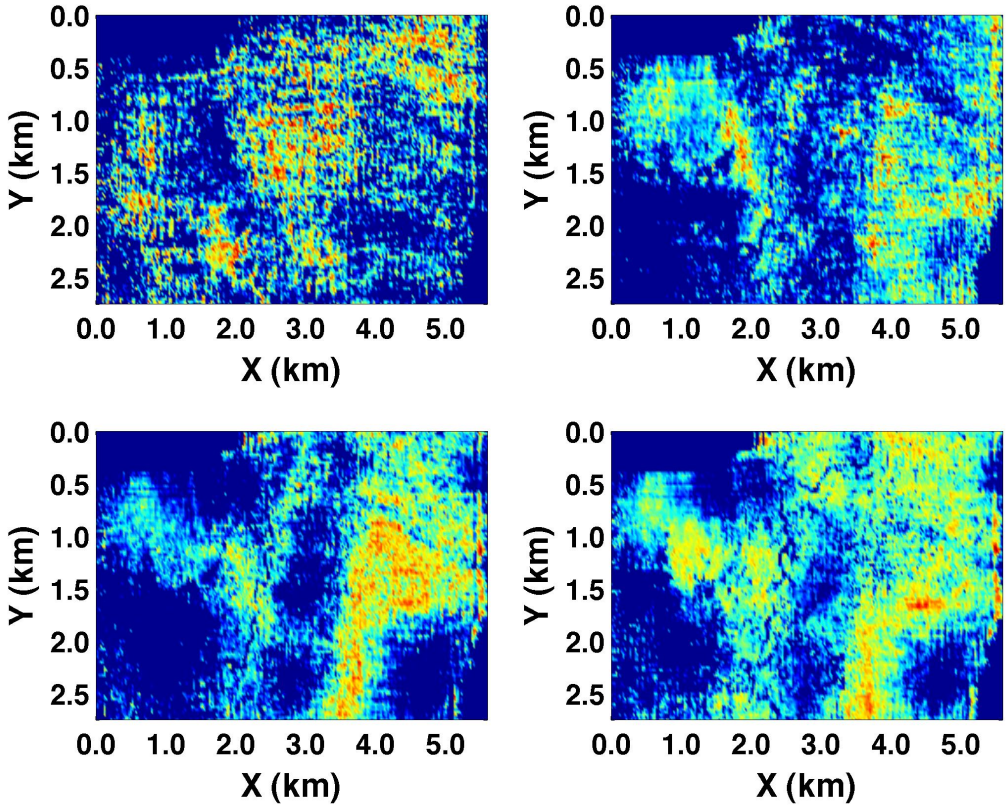


Figure 4.11: Amplitude of the ℓ_1 norm local time-frequency representation for the time slice in Figure 4.10. Top left: 15Hz slice, Top right: 25Hz slice, Bottom left: 35Hz slice, Bottom right: 45Hz slice. Note the apparent lack of energy in the 45Hz slice which displays the position of the channel.

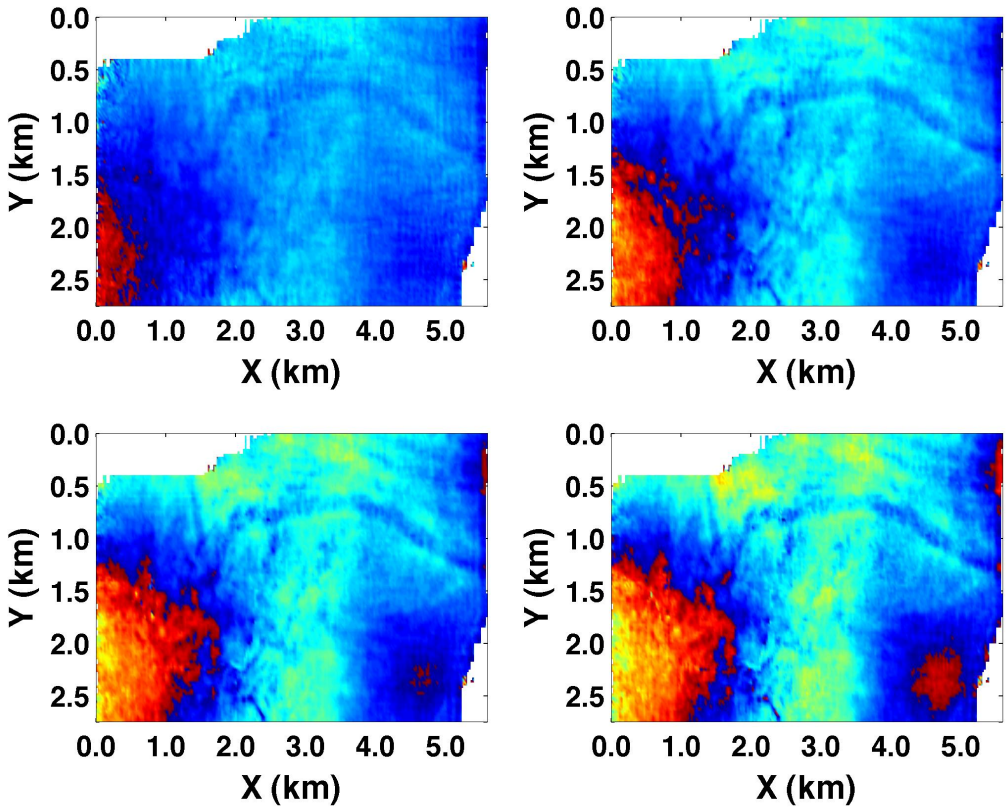


Figure 4.12: Phase of the ℓ_2 norm local time-frequency representation for the time slice in Figure 4.10. Top left: 15Hz slice, Top right: 25Hz slice, Bottom left: 35Hz slice, Bottom right: 45Hz slice.

according to,

$$u_s(i) = (u(i) - u_{min}) \frac{1}{u_{max} - u_{min}}, \tag{4.1}$$

where $u(i)$ and $u_s(i)$ represent an image pixel before and after a contrast stretch and u_{min} and u_{max} represent the minimum and maximum pixel value of u. The process of contrast stretching simply expands information that is clustered together to encompass the entire data range available as illustrated by Figure 4.13. The contrast stretched blended RGB image of the ℓ_1 norm regularized inversion local time-frequency representation is shown in Figure 4.14. Through creating a blended RGB image, the resolution of the small paleochannel from Figure 4.10 is significantly enhanced.

4.5 Summary

Utilizing the inverse method for spectral decomposition developed in Chapter 3, the advantages and disadvantages of both ℓ_1 norm and ℓ_2 norm regularization for application to seismic data was discussed. Through constraining the local time-frequency representation to become sparse, ℓ_1 norm regularization's higher resolution local time-frequency representation provided more variability between the amplitude of constant frequency sections. This allowed for the different frequency behaviours of the two large paleochannels found in the example cross section of Figure 4.1 to become more evident. When analyzing the seismic amplitude volume alone, the difference between the low frequency shadows of the upper paleochannel and the dominant low frequency energy of the lower paleochannel would have been extremely difficult to ascertain. However, the sparseness of the local time-frequency representation obtained through ℓ_1 norm regularization does not allow for a global interpretation of the phase like the local time-frequency representation obtained from ℓ_2 norm regularization. Although not commonly utilized for interpretation, the phase obtained from ℓ_2 norm regularization based upon the multi-wavelet convolutional model may prove useful for studying thin bed tuning effects such as stratigraphic pinch out events. By compressing all of the spectral information obtained through spectral decomposition into a blended RGB image, the information of how the frequency content changes in a cross section or time slice can be displayed in a single image. This endeavour proves useful with large data sets since local time-frequency analysis generates large volumes of data which can become difficult to interpret. The creation of a blended RGB image from local time-frequency representations can also aid in the visualization of features such as the small paleochannel in Figure 4.10.

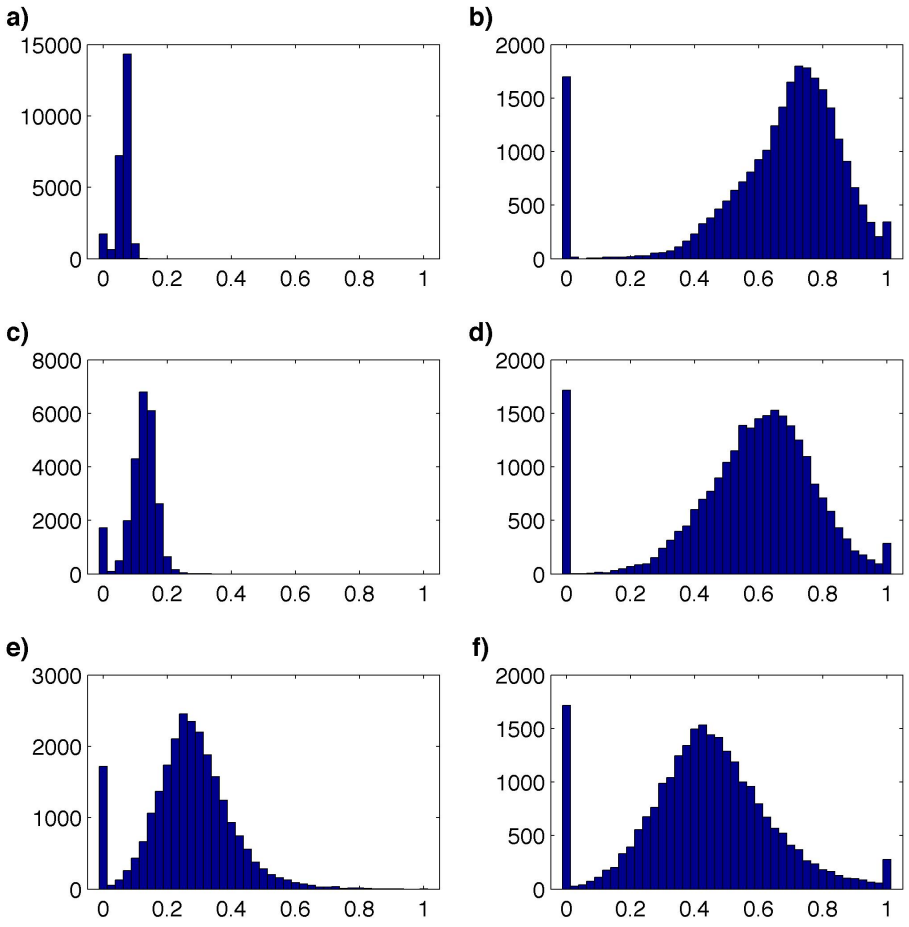


Figure 4.13: Histograms for the red pixel intensity a) before and b) after contrast stretching, the green pixel intensity c) before and d) after contrast stretching, and the blue pixel intensity e) before and f) after contrast stretching for the blended RGB image of Figure 4.14.

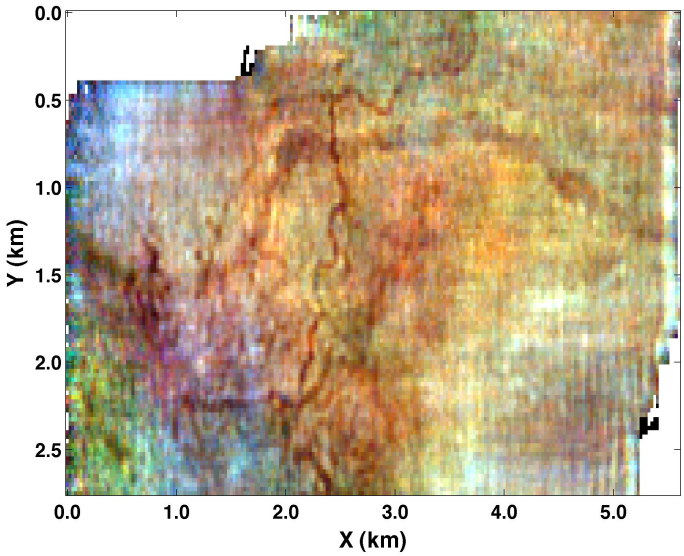


Figure 4.14: A contrast stretched blended RGB image of the ℓ_1 norm regularized inversion local time-frequency representation for the time slice in Figure 4.10 utilizing the frequency bands: below 25Hz, between 25Hz and 45Hz, and above 45Hz. By combining all of the spectral decomposition information into a single image, visualization of the small paleochannel was greatly improved.

CHAPTER 5

Spectral Decomposition with $f - x - y$ Preconditioning[1]

5.1 Motivation

Through the pursuit of creating a high resolution local time-frequency representation of the seismic trace, noise becomes a greater problem as small changes within the data will have a more profound effect on the local time-frequency representation. Although the denoising capabilities of sparse local time-frequency transforms have been well studied for the reconstructed signal (Chen et al., 1998), the recovery of a denoised version of the local time-frequency representation has not been as extensively studied. Typically, the spectral decomposition problem for seismic data has been viewed as a transformation on a one dimensional seismic trace into a two dimensional local time-frequency representation. Meanwhile, seismic data is generally known to change slowly in the lateral direction, meaning neighbouring seismic traces often have a similar structure. This general idea is applied in Wang (2010) where a linearity assumption is implemented into a matching pursuit algorithm to try to produce a more robust local time-frequency representation for a seismic cross section and is the only published attempt to produce a denoised version of the local time-frequency representation for seismic data.

In this chapter, the spectral decomposition problem is viewed as a multidimensional problem to allow for the incorporation of neighbouring traces during the construction of the local time-frequency representation of a seismic volume, i.e., a two dimensional seismic cross

[1] A version of this chapter has been submitted for publication (Bonar and Sacchi, 2012b).

section to a three dimensional time-frequency representation. As discussed in Chapter 3, representing the seismic trace as a multi-wavelet convolutional process allows its local time-frequency representation to be considered as a set of wavelet dependent reflectivities. This enables concepts such as $f-x$ or $f-x-y$ deconvolution to be incorporated into the spectral decomposition algorithm via the preconditioning of the inversion problem similar to the structure-and-amplitude-preserving multichannel deconvolution algorithm proposed in Wang and Sacchi (2009). The incorporation of $f-x$ or $f-x-y$ deconvolution into the spectral decomposition algorithm allows for the structure of neighbouring seismic traces to influence the time-frequency representation of a seismic trace during its generation, helping to eliminate the effects of noise. The application of the proposed $f-x-y$ preconditioned spectral decomposition algorithm is demonstrated with a synthetic data set and tested on the real 3D data set containing the small paleochannel in Figure 4.10 where it is compared with k_x-k_y, or bandpass, filtering for denoising after the generation of the local time-frequency representation using sparse inversion.

5.2 Theory

The generation of a denoised time-frequency representation can be achieved by incorporating the idea of $f-x$ or $f-x-y$ deconvolution within the sparse inversion approach described previously for spectral decomposition. To accomplish this goal, a similar derivation to the structure-and-amplitude-preserving multi-channel deconvolution method proposed in Wang and Sacchi (2009) is followed. At the channel i of a seismic cross section, the convolutional model of a seismic trace can be described in the Fourier domain by

$$S_i(\omega) = W(\omega)R_i(\omega) + N_i(\omega), \tag{5.1}$$

where S_i represents the seismic trace, W represents the wavelet, R_i represents the reflectivity series, and N_i represents random noise. Utilizing the assumption that the reflectors of the subsurface are laterally continuous (i.e. low trace to trace variability), the reflectivity at the channel i can be predicted from the previous L channels since the reflectivity in the $f-x$ domain becomes a sum of complex exponentials at each frequency ω (Canales, 1984). This assumption leads to

$$R_i(\omega) = P_1(\omega)Z_{i-1}(\omega) + P_2(\omega)Z_{i-2}(\omega) + \ldots + P_L(\omega)Z_{i-L}(\omega), \tag{5.2}$$

where P_l refers to the prediction filter and Z_l can be considered as the reflectivity series prior to spatial smoothing. Once the prediction filter is found for a given frequency ω, the

convolutional model can be rewritten as

$$S_i(\omega) = W(\omega) \sum_{l}^{L} [P_l(\omega) Z_{i-l}(\omega)] + N_i(\omega), \tag{5.3}$$

or more commonly the model for $f-x$ deconvolution. The actual estimation of the spatial prediction filter is a topic unto itself, however, examples of its estimation can be found in Yilmaz (2001).

To expand this concept to the multi-wavelet convolutional model, Equation 3.2 must be transformed into the Fourier domain,

$$S_i(\omega) = \sum_{k}^{K} [W(\omega, k) R_i(\omega, k)] + N_i(\omega). \tag{5.4}$$

Presuming that the assumption of reflectivity being laterally continuous also holds true for the wavelet dependent reflectivity series, the time-frequency representation at the channel i can also be predicted from the previous L channels. At a specific Fourier frequency ω, this assumption leads to

$$R_i(\omega, k) = P_1(\omega) Z_{i-1}(\omega, k) + P_2(\omega) Z_{i-2}(\omega, k) + \ldots + P_L(\omega) Z_{i-L}(\omega, k). \tag{5.5}$$

Since the estimation of the unknown prediction filter, P_l, from the unknown wavelet dependent reflectivity series is a non-linear process and difficult to compute, it is often estimated from the original data, as discussed in Wang and Sacchi (2009), and will be considered constant for all k. Combining Equations 5.4 and 5.5 results in,

$$S_i(\omega) = \sum_{k}^{K} \{W(\omega, k) \sum_{l}^{L} [P_l(\omega) Z_{i-l}(\omega, k)]\} + N_i(\omega), \tag{5.6}$$

or the forward model for spectral decomposition with $f-x$ preconditioning. In this formulation, the concept of a reflectivity series is expanded from being a one dimensional object in time to a three dimensional object that varies not only in time, but also in space and frequency. Figure 5.1 highlights this concept where spatial preconditioning is incorporated into spectral decomposition. Instead of calculating each time-frequency representation separately, as in Figure 5.1a, they are computed as a group to enforce spatial continuity, as in Figure 5.1b.

The extension of this method to incorporate multiple spatial dimensions is similar to the extension of $f-x$ deconvolution to $f-x-y$ deconvolution where rather than incorporating only one spatial dimension to predict the reflectivity series, two spatial dimensions are considered (Chase, 1992; Gulunay et al., 1993). Thus, a two dimensional prediction filter

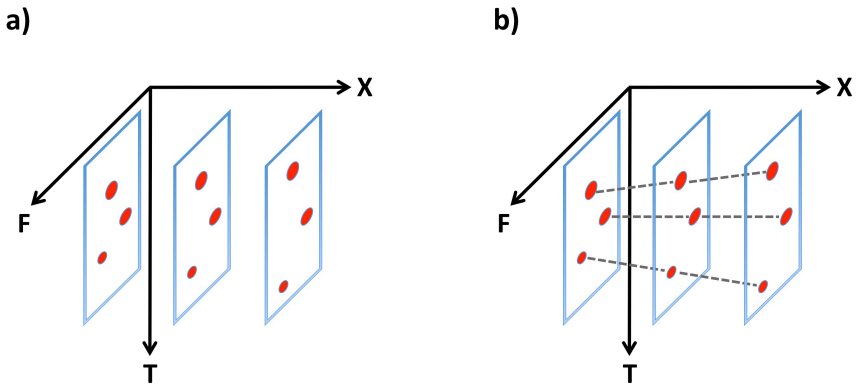


Figure 5.1: An example of spectral decomposition a) without and b) with spatial preconditioning.

$P_{l,m}$ is required instead of a one dimensional prediction filter P_l to allow for spatial prediction among multiple azimuths. The forward model for $f-x-y$ preconditioned spectral decomposition then becomes,

$$S_{i,j}(\omega) = \sum_{k}^{K}\{W(\omega,k)\sum_{l,m}^{L,M}[P_{l,m}(\omega)Z_{i-l,j-m}(\omega,k)]\} + N_{i,j}(\omega), \tag{5.7}$$

where the sub-indices i and j represent the channel number in the x and y directions.

To solve for the spatially preconditioned time-frequency representation, the forward (Equation 5.8) and adjoint (Equation 5.9) problems are written in terms of linear algebra,

$$\mathbf{s} = \mathbf{DPz} + \mathbf{n} \tag{5.8}$$

$$\widehat{\mathbf{z}} = \mathbf{P}^H\mathbf{D}^H\mathbf{s}. \tag{5.9}$$

Here, the variable $\mathbf{z}$ represents the time-frequency representation prior to spatial smoothing and the operator $\mathbf{P}$ denotes the predictive filtering process. Note that the variables $\mathbf{s}, \mathbf{z}, \mathbf{n}$ no longer represent a single seismic trace but rather a group of traces that have been organized into a vector. The determination of the spatially preconditioned time-frequency representation is now achieved by minimizing the altered cost function J,

$$J = ||\mathbf{s} - \mathbf{DPz}||_{\mathbf{2}}^{\mathbf{2}} + \lambda||\mathbf{z}||_{\mathbf{1}}, \tag{5.10}$$

and then applying the prediction filtering process on the solution for $\mathbf{z}$ to obtain the spatially preconditioned time-frequency representation $\mathbf{m}_{EST}$,

$$\mathbf{m}_{EST} = \mathbf{P}\mathbf{z}. \tag{5.11}$$

The forward operator required for FISTA to minimize the cost function in Equation 5.10 now becomes,

$$\mathbf{A} = \mathbf{D}\mathbf{P}, \tag{5.12}$$

whose adjoint operator is given by,

$$\mathbf{A}^H = \mathbf{D}^H\mathbf{P}^H. \tag{5.13}$$

As previously stated, the prediction filters were estimated from the original data and kept constant for all k prior to the inversion process and were also not updated during the iterative process of FISTA.

5.3 Results

To test the capabilities of the proposed preconditioned spectral decomposition algorithm, a synthetic seismic cube was created with three linear events of varying frequency content and a signal to noise ratio, defined by the ratio of the maximum amplitude of the signal to the maximum amplitude of the noise, of 2 as depicted in Figure 5.2. The linear events were created using Ricker wavelets with central frequencies ranging from 15Hz to 40Hz and approximately 30% of the traces were time-shifted a small amount to simulate statics within the data. Studying the local time-frequency representation obtained for a single trace within the synthetic seismic cube (Figure 5.3), the benefits of preconditioning spectral decomposition with $f-x-y$ deconvolution can be seen. Although adding $f-x$ or $f-x-y$ preconditioning to spectral decomposition creates more non-zero coefficients in the local time-frequency representation, the local time-frequency representation does not lose resolution when compared with a high resolution spectral decomposition. Instead, the preconditioning simply spreads the energy from the noisy artifacts in the high resolution spectral decomposition local time-frequency representation making them less pronounced while promoting the resolution of the actual events in the data. By extending the preconditioning to include two spatial dimensions at once, these effects become stronger as information from multiple azimuths are incorporated. This statement is further illustrated through RGB composite frequency volumes of a cross section through the data (Figure 5.4). A composite RGB frequency volume allows the information from several frequency bandwidths to be displayed in the same plot. In this specific instance, the RGB image was

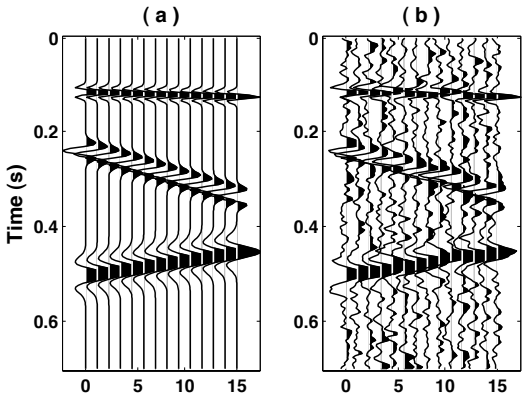


Figure 5.2: A cross section of a synthetic data cube containing three linear events a) without and b) with noise.

created with the amplitude of the low frequencies corresponding to the red pixel magnitude, the amplitude of the intermediate frequencies corresponding to the green pixel magnitude, and the amplitude of the high frequencies corresponding to the blue pixel magnitude. Similar to viewing the local time-frequency representation for a single trace, the artifacts caused by noise are less apparent when the spectral decomposition algorithm is preconditioned with $f-x$ or $f-x-y$ deconvolution. The incorporation of $f-x$ or $f-x-y$ deconvolution within spectral decomposition also enables the algorithm to negate the effects of statics within seismic data, regardless of the events frequency content. This is most evident with the $f-x-y$ preconditioned spectral decomposition data as the composite RGB image for $f-x-y$ preconditioned spectral decomposition contains the least deviation from the perfectly linear events of the the noise free data.

When applied to a real 3D data set from Figure 1.3, the spectral decomposition with $f-x-y$ preconditioning method was applied in small windows to meet the assumptions inherent within $f-x-y$ deconvolution. Specifically, the data was divided into overlapping windows whose spatial size was approximately 15 traces by 15 traces and the prediction filter for $f-x-y$ preconditioning was chosen to be a non-causal 7 x 7 prediction filter that was estimated from the original seismic amplitude volume. The results of spectral decomposition with $f-x-y$ preconditioning for this data set are displayed in Figures 5.5 and 5.6. An improvement in the lateral continuity of seismic reflectors is achieved in the $f-x-y$ preconditioned constant frequency sections of the seismic cross section when compared to constant frequency sections obtained from sparse spectral decomposition (Figure 7.2), while a decrease in the amount of noise in the $f-x-y$ preconditioned constant frequency sections of the seismic time slice is apparent when compared to the sparse spectral decomposition constant frequency sections (Figure 7.3). Most importantly, the structural features that

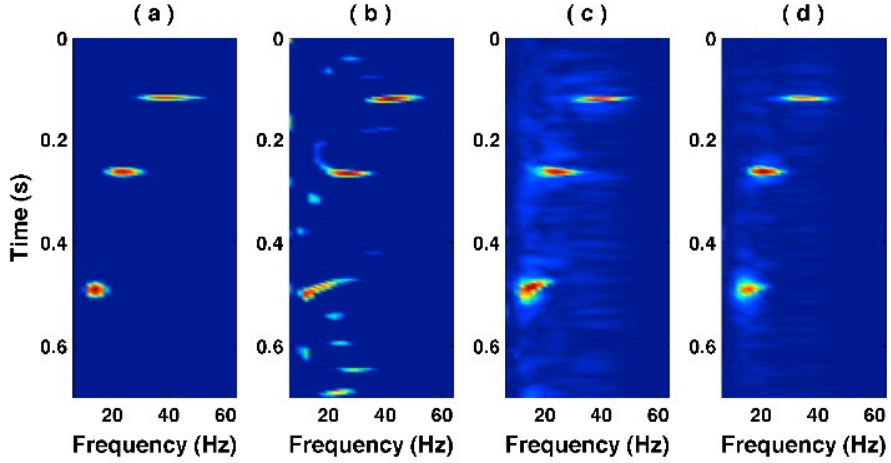


Figure 5.3: The local time-frequency representation for a single trace within the synthetic data set in Figure 5.2 for a) sparse spectral decomposition on the noise free data, b) sparse spectral decomposition on the noisy data, c) spectral decomposition with $f-x$ preconditioning on the noisy data, and d) spectral decomposition with $f-x-y$ preconditioning on the noisy data.

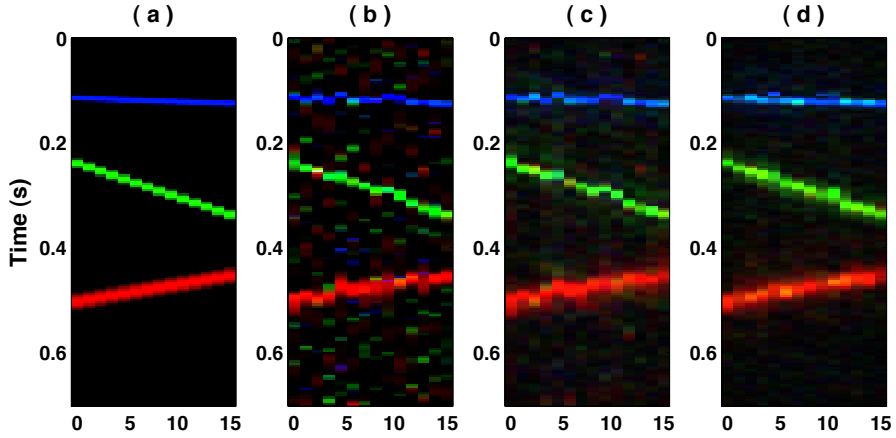


Figure 5.4: A composite RGB image for the local time-frequency representation of the cross section of the synthetic data represented in Figure 5.2 for a) sparse spectral decomposition on the noise free data, b) sparse spectral decomposition on the noisy data, c) spectral decomposition with $f-x$ preconditioning on the noisy data, and d) spectral decomposition with $f-x-y$ preconditioning on the noisy data. The blue pixel magnitude represents the amplitude of the high frequencies, the green pixel magnitude represents the amplitude of the intermediate frequencies, and the red pixel magnitude represents the amplitude of the low frequencies.

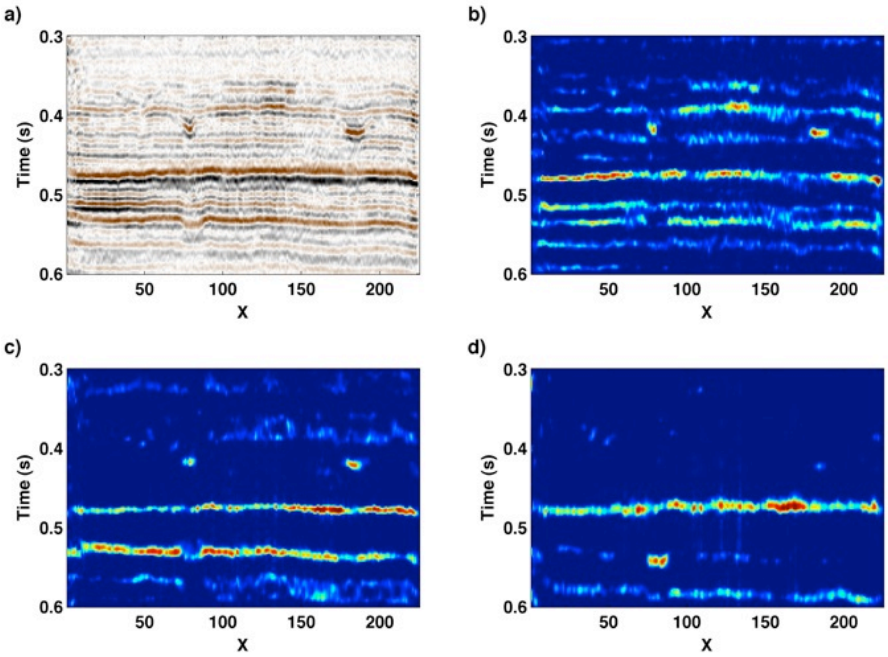


Figure 5.5: a) A seismic amplitude cross section through a real 3D poststack data set. b) A constant 50Hz section of a) using spectral decomposition with $f-x-y$ preconditioning. c) A constant 35Hz section of a) using spectral decomposition with $f-x-y$ preconditioning. d) A constant 25Hz section of a) using spectral decomposition with $f-x-y$ preconditioning.

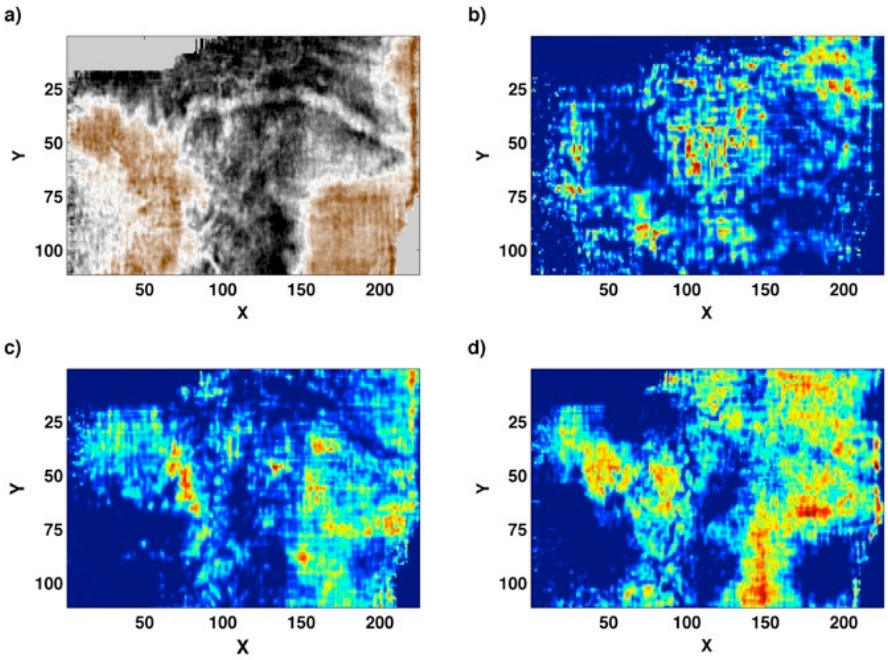


Figure 5.6: a) A seismic amplitude time slice through a real 3D poststack data set. b) A constant 15Hz section of a) using spectral decomposition with $f-x-y$ preconditioning. c) A constant 25Hz section of a) using spectral decomposition with $f-x-y$ preconditioning. d) A constant 45Hz section of a) using spectral decomposition with $f-x-y$ preconditioning.

were seen in the sparse spectral decomposition have been preserved. The difference between the frequency behaviours of the lower and upper large channels is still identifiable in the $f-x-y$ preconditioned constant frequency sections of the cross section and the definition of the small channel within the $f-x-y$ preconditioned constant 45Hz section of time slice has improved.

The denoising of a local time-frequency representation of the seismic volume can also be performed after the local time-frequency representation has been generated, unlike $f-x-y$ preconditioning which denoises while the local time-frequency representation is being generated. To denoise the local time-frequency representation after it is generated, k_x-k_y or bandpass filtering was performed on each time slice of the local time-frequency representation to remove large wave numbers representing noise within the data. A comparison of denoising the local time-frequency representation after or while it was generated is depicted in Figures 5.7 and 5.8. While the results of denoising the spectrally decomposed seismic cross section are quite comparable (Figure 5.7), the difference between these local time-frequency

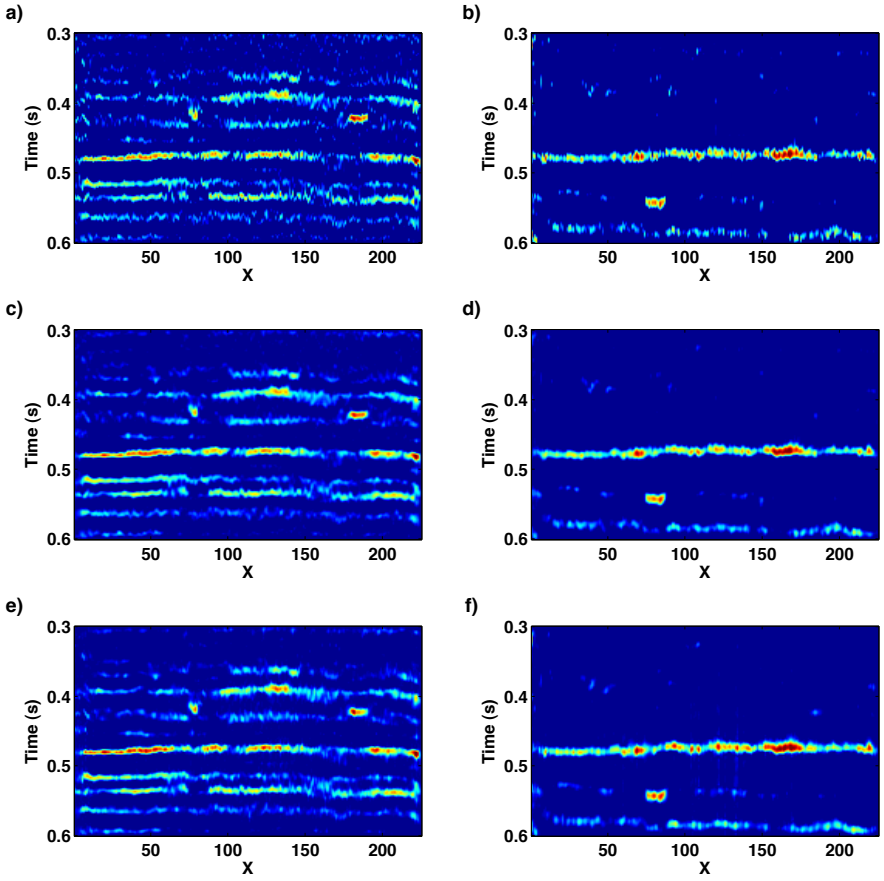


Figure 5.7: Comparison of local time-frequency representation denoising techniques for the seismic amplitude cross section. a), c), and e) represent the sparse spectral decomposition, k_x-k_y filtering, and $f-x-y$ preconditioning spectral decomposition for the constant 50Hz frequency section while b), d), and f) represent the sparse spectral decomposition, $k_x - k_y$ filtering, and $f - x - y$ preconditioning spectral decomposition for the constant 25Hz frequency section.

representation denoising methods is highlighted through the spectrally decomposed seismic time slice depicted in Figure 5.8. By always trying to fit the data exactly, spectral decomposition with $f-x-y$ preconditioning helps to prevent the smoothing out of small geological structures, such as the small channel, since the data is always being fit during the denoising process. Likewise, k_x-k_y filtering is altering the local time-frequency representation after it is generated, meaning it changes the local time-frequency representation after it is fit to the data. This makes methods which denoise after the local time-frequency representation is generated, such as k_x-k_y filtering, more likely to lose the resolution or smooth out small scale geological structures since the denoising process is not specifically trying to reconstruct the seismic data. However, as in every signal processing problem there are parameters to be set before optimal results can be acquired with high resolution spectral decomposition with $f-x-y$ preconditioning. Specifically, the trade-off parameter λ, the prediction filter length, and the window size must all be specified through user input and can be found in an ad hoc way similar to the prediction filter length of $f-x$ deconvolution. In other words, a solution is found to predict the data and the residuals or data misfit is studied. For instance, if the residuals contain signal it means that the model is underfitting the data and λ, the prediction filter length, and/or the window size needs to be decreased. Likewise, if the residuals do not contain enough noise, i.e., the local time-frequency representation still contains a significant amount of noise, λ, the prediction filter length, and/or the window size needs to be increased.

5.4 Conclusions

Spectral decomposition, or localized time-frequency analysis, can be a powerful tool for interpreting seismic data. However, through the pursuit of generating high resolution local time-frequency representations using sparsity constraints, noise that was not visible in the original seismic amplitude volume becomes more noticeable in the spectrally decomposed data. To combat this issue, denoising efforts can be carried out after the local time-frequency representation is generated or during the creation of the local time-frequency representation. By posing the spectral decomposition problem through a multi-wavelet convolutional model, the concepts of $f-x$ and $f-x-y$ deconvolution were able to be incorporated into an inversion method for generating a local time-frequency representation that is denoised while it is created. A comparison with the post processing denoising technique of k_x-k_y filtering on a sparse spectrally decomposed real data set demonstrates the advantages of denoising the local time-frequency representation while it is being generated. Although large scale features are relatively unaffected by the choice of when denoising occurs, small scale features are more likely to be preserved when the denoising of the local time-frequency representation occurs during its creation since spectral decomposition with $f-x-y$ preconditioning attempts to

always fit the original data set unlike post processing denoising techniques which alters the local time-frequency representation after it is fit to the data.

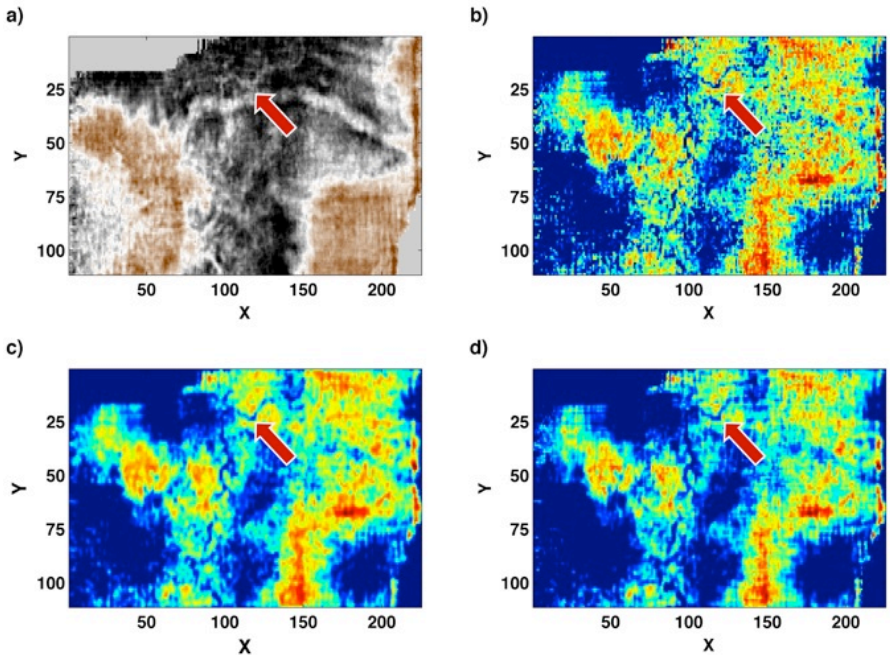


Figure 5.8: Comparison of local time-frequency representation denoising techniques for the seismic amplitude time slice. (a) seismic amplitude time slice, b) sparse spectral decomposition constant 45Hz section, c) $k_x - k_y$ filtering of b), d) $f-x-y$ preconditioned spectral decomposition constant 45Hz section.) Note the enhanced resolution of the small channel feature, highlighted with the red arrow, in d) when compared to the other plots.

CHAPTER 6

Group Sparsity Local Time-Frequency Denoising

6.1 Introduction

As shown in previous chapters, local time-frequency analysis can be employed as an interpretational tool for seismic data. In addition to this powerful trait, spectral decomposition can also be applied as a serviceable tool during its processing. Through transforming and expanding the seismic signal to display its frequency spectrum at each time sample, spectral decomposition has been deployed to identify and attenuate undesirable signals such as ground roll (Deighan and Watts, 1997; Askari and Siahkoohi, 2008) and has provided an alternative domain for regularization and interpolation (Naghizadeh and Innanen, 2011). Another application of local time-frequency analysis for processing seismic data is the attenuation of random noise, either through strategic thresholding of the local time-frequency representation (Parolai, 2009; Elboth et al., 2010) or directly through inversion (Vera Rodriguez et al., 2011). The latter of these random noise attenuation methods is more formally known as the Basis Pursuit Denoising problem (Chen et al., 1998). As briefly discussed in Chapter 3, regularization stabilizes the inversion process in the presence of noise and, for the case of ℓ_1 norm regularization, has been shown to also produce an adequate denoised version of the data. In contrast to user defined strategic muting or thresholding strategies of the local time-frequency representation to generate a denoised time-series, Basis Pursuit Denoising simply fits a signal model up to the noise level of the signal.

Similar to how information from nearby spatial traces were incorporated to generate a denoised local time-frequency representation in Chapter 5, the simultaneous utilization of

information from the different components of a multicomponent seismic signal can aid in the signal denoising process. As a seismic wave travels in three dimensional space, it is often recorded with a multicomponent geophone which contains three orthogonally oriented recorders, one vertical and two horizontal, to capture the three dimensional nature of the signal. When the seismic wave arrives and is recorded on the multicomponent geophone, the frequency content at a particular time should be identical on all three components even though it may have a different amplitude and phase within each component. Therefore, sparsity constraints might not be the optimal choice for regularization since this commonality between the components will not necessarily be preserved. Instead it would be more beneficial to consider the three components together before sparsity constraints are applied. This goal can be accomplished through quaternions (the extension of complex numbers, see Stanton and Sacchi (2011) for an example with seismic data) or through the consideration of group sparsity constraints (Yuan and Lin, 2006; Eldar and Bolcskei, 2009). The concept of group sparsity has been implemented in problems such as linear regression (Yuan and Lin, 2006), brain activity imaging (Cotter et al., 2005) and the determination of the direction of arrival in antennas (Hyder and Mahata, 2010). Specifically within geophysics, group sparsity constraints have been used in compressive imaging by wavefield inversion (Herrmann, 2009) and the joint determination of the origin time, hypocenter location and moment tensor of microseismic events (Vera Rodriguez et al., 2010a,b). Before the concepts of group sparsity are explored in this chapter, a review of the parameter selection of ℓ_1 norm regularized inversion for spectral decomposition is provided to illustrate the denoising capabilities of local time-frequency analysis.

6.2 Denoising Through Sparsity Constrained Inversion

Formally, the cost function described in Equation 3.11 defines the Basis Pursuit Denoising problem, not the Basis Pursuit problem. However, in the limit when the trade-off parameter λ approaches zero this cost function does describe the Basis Pursuit problem since the goal of Basis Pursuit is to reconstruct the signal exactly with a minimum ℓ_1 norm model. This goal is in slight contrast to the goal of Basis Pursuit Denoising which is to approximately reconstruct the signal with a minimum ℓ_1 norm model (Chen et al., 1998). Therefore, λ controls the degree of fit between the observed signal and the reconstructed or predicted signal. Since observed data is always influenced by the presence of noise, this parameter should in effect always be larger than zero which requires the solution to Basis Pursuit Denoising problem.

To illustrate the denoising properties of the ℓ_1 norm regularized inverse method for spectral decomposition, a synthetic signal containing two separate events of different phase and

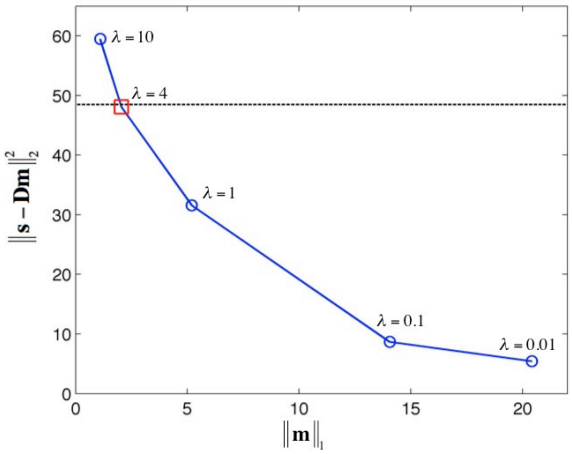


Figure 6.1: Pareto curve for ℓ_1 norm regularized inversion of the noisy synthetic time series in Figure 6.2. The dotted black line denotes the ℓ_2 norm of the noise and the red box represents the optimal choice for λ.

frequency content was contaminated with random noise from the Matlab function **randn.m**. The signal to noise ratio (SNR) of this noise contaminated synthetic time-series is 2, as defined by the maximum amplitude of the signal divided by the maximum amplitude of the noise. To ensure that over fitting or under fitting of the noisy signal does not occur, the Pareto curve (Hennenfent et al., 2008) was employed. The Pareto curve is simply a comparison of the data misfit term (i.e., the predicted noise) and the regularization term defined by the ℓ_1 norm and allows for a quantitative analysis of how λ influences the amount of noise that is fit within the model. As depicted in Figure 6.1, increasing λ provides more influence to the regularization term meaning that a greater value for the data misfit term is permitted. In other words, a larger value for λ signifies a greater estimated level of noise within the signal. For the case of synthetic data where the noise level is known, the optimal selection of λ simply becomes the point on the Pareto curve that corresponds to the actual noise level (i.e., the ℓ_2 norm of the noise). However, when applied with real data the noise level is often unknown so heuristic methods are required. Once the optimal choice for λ is determined, the denoised version of the synthetic signal, $\mathbf{s}_{\ell_1}$, can be obtained from mapping the solution of the ℓ_1 norm minimization, $\mathbf{m}_{\ell_1}$, back into the data domain with the forward operator,

$$\mathbf{s}_{\ell_1} = \mathbf{D}\mathbf{m}_{\ell_1}. \tag{6.1}$$

Figure 6.2 displays the result of denoising through ℓ_1 norm regularized inversion and, for comparison, the result of denoising through ℓ_2 norm regularized inversion. It is apparent that although ℓ_2 norm regularization attenuates some of the random noise, a cleaner esti-

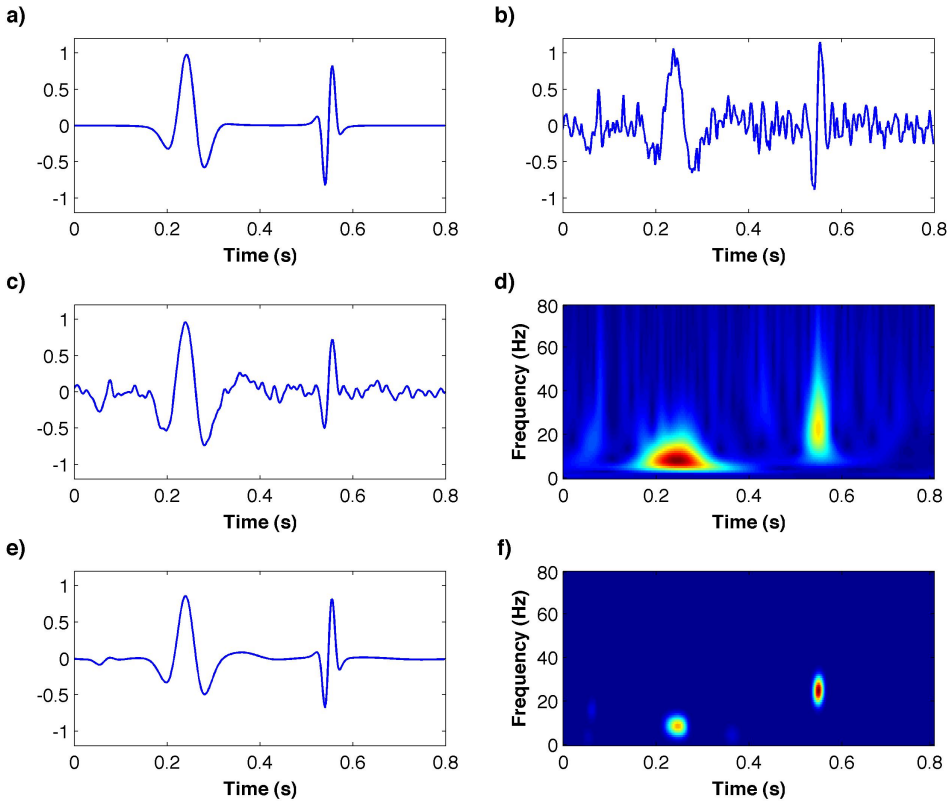


Figure 6.2: A comparison of denoising with spectral decomposition with ℓ_2 norm and ℓ_1 norm regularized inversion: a) synthetic time-series, b) noisy synthetic time-series (SNR = 2), c) ℓ_2 norm denoised time-series, d) local time-frequency representation of c), e) ℓ_1 norm denoised time-series, f) local time-frequency representation of e).

mated signal is obtained through constraining the local time-frequency representation to be sparse. This comparison illustrates that the choice for the regularization term in the spectral decomposition inversion method has a strong influence on its denoising capabilities. Real data examples with microseismic data for denoising through ℓ_1 norm regularized inversion based on local time-frequency analysis are presented in Vera Rodriguez et al. (2011).

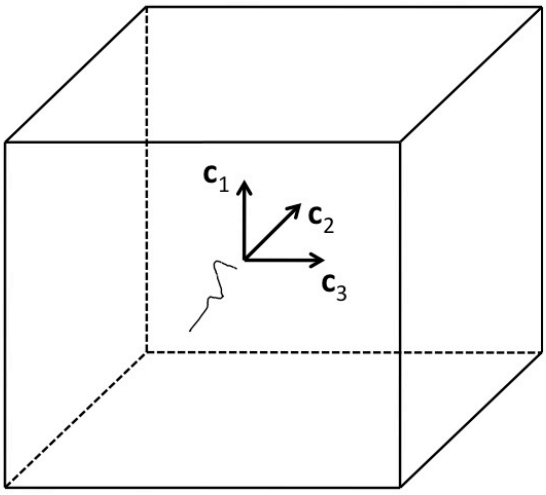


Figure 6.3: An example of a seismic signal travelling through a three dimensional space and being recorded by a multicomponent geophone.

6.3 Denoising Through Group Sparsity Constrained Inversion

To fully capture a seismic wavefront that travels through the three dimensional space of the subsurface, it can be recorded on a multicomponent geophone as illustrated in Figure 6.3. Since all three components of a multicomponent geophone record the same signal, it can be assumed that the frequency content at one particular time are similar for all components as depicted in Figure 6.4, regardless of the different amplitude and phase that the signal might have on each individual component. Through considering these three components as a multiple measurement vector (MMV) (Cotter et al., 2005; Hyder and Mahata, 2010), the concept of group sparsity can be incorporated as the regularization term for the spectral decomposition inversion method. The MMV, $\mathbf{v_{MMV}}$, for a multicomponent seismic trace can be defined as,

$$\mathbf{v_{MMV}} = \begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{bmatrix}, \tag{6.2}$$

where $\mathbf{c}_1$, $\mathbf{c}_2$, and $\mathbf{c}_3$ correspond to the seismic trace for each of the three components recorded from a single multicomponent geophone. To impose the constraint for groups to be sparse, such as a particular frequency and time for all components, a different metric other than the ℓ_1 norm to measure sparsity must be defined. A metric that can handle

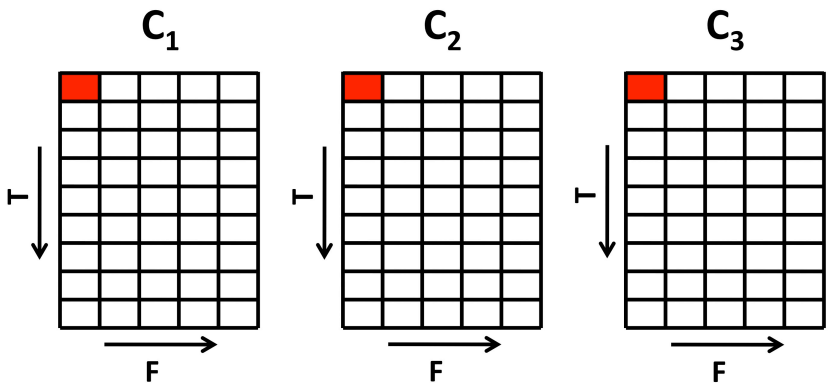


Figure 6.4: The local time-frequency representation obtained from a multicomponent seismic trace. The highlighted red box represents the assumption that if the signal has a frequency content at a particular time, it must be present on all components.

this constraint comes from the study of mixed p, q norms (Kowalski and Torrésani, 2009). Specifically, this chapter will focus on the mixed $\ell_2 - \ell_1$ norm defined by,

$$||\mathbf{x}||_{2,1} = \left(\sum_{g} \left(\sum_{\varrho} \mathbf{x}_{g,\varrho}{}^2 \right)^{1/2} \right), \tag{6.3}$$

where $||\cdot||_{2,1}^2$ denotes the mixed $\ell_2 - \ell_1$ norm, g is the index for groups, and ϱ is the index for the elements within a group. By incorporating group sparsity constraints for the inversion method of spectral decomposition, the cost function becomes,

$$J = ||\mathbf{v_{MMV}} - \mathbf{Dm}||_2^2 + \lambda ||\mathbf{m}||_{2,1}, \tag{6.4}$$

where $\mathbf{D}$ and $\mathbf{m}$ have been modified to incorporate multicomponent data according to,

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}(\mathbf{c}_1), & \mathbf{D}(\mathbf{c}_2), & \mathbf{D}(\mathbf{c}_3) \end{pmatrix}, \tag{6.5a}$$

and

$$\mathbf{m} = \begin{pmatrix} \mathbf{m}(\mathbf{c}_1) \\ \mathbf{m}(\mathbf{c}_2) \\ \mathbf{m}(\mathbf{c}_3) \end{pmatrix}. \tag{6.5b}$$

Similar to the ℓ_1 norm in the cost function of Equation 3.11 being separable, the mixed $\ell_2 - \ell_1$ norm in the cost function of Equation 6.4 is also separable. Therefore, the minimization of the cost function in Equation 6.4 can be computed in a similar manner to how FISTA solves

	Frequency	Amplitude	Phase
First event in $\mathbf{c}_1$	16 Hz	0.5	-30°
First event in $\mathbf{c}_2$	16 Hz	0.25	-10°
First event in $\mathbf{c}_3$	16 Hz	0.25	-110°
Second event in $\mathbf{c}_1$	36 Hz	1	90°
Second event in $\mathbf{c}_2$	36 Hz	0.5	60°
Second event in $\mathbf{c}_3$	36 Hz	0.125	180°

Table 6.1: Construction of multicomponent synthetic time-series to test group sparsity denoising.

the minimization of the cost function in Equation 3.11. Rather than taking the soft threshold of each gradient descent step as in FISTA, a new multidimensional thresholding operator is required (Fornasier and Rauhut, 2008; Puig et al., 2011). This new multidimensional thresholding operator, $\mathcal{G}_\lambda$ is defined by,

$$\mathcal{G}_\lambda(\mathbf{g}) = \begin{cases} \left(||\mathbf{g}||_2^2 - \lambda\right) \frac{\mathbf{g}}{||\mathbf{g}||_2^2} & \text{if } ||\mathbf{g}||_2^2 > \lambda \\ 0 & \text{if } ||\mathbf{g}||_2^2 \leq \lambda \end{cases}, \tag{6.6}$$

where $\mathbf{g}$ is a pre-defined vector group. Since $\mathcal{G}_\lambda$ does not affect the model parameter update of FISTA, it can be directly incorporated into a FISTA-like algorithm (Fornasier and Rauhut, 2008),

$$\mathbf{m}_{i+1} = \sum_{\mathbf{g}} \mathcal{G}_{\lambda/2\alpha} \left(\mathbf{y}_i(\mathbf{g}) - \frac{1}{\alpha} \mathbf{D}^H \left[\mathbf{v}(\mathbf{g}) - \mathbf{D}\mathbf{y}_i(\mathbf{g}) \right] \right), \tag{6.7}$$

which in this thesis will be called Group FISTA. In essence, the multidimensional thresholding operator is applied to all pre-defined groups within $\mathbf{v}$ and $\mathbf{y}_i$ during each iteration to determine $\mathbf{m}_{i+1}$.

To test if group sparsity constraints provide additional benefits when compared with sparsity constraints for denoising multicomponent seismic signals through local time-frequency analysis, a multicomponent synthetic time-series was created according to the parameters in Table 6.1. This multicomponent synthetic time-series was then contaminated with random noise from the Matlab function **rand.m** to obtain three different noise contaminated signals with a SNR of 10, 4, and 2. As explained in Section 6.2, the optimal choice for λ

was determined with the aid of Pareto curves and the known noise levels. The denoised results for the three different noise levels are displayed in Figure 6.5 for both sparsity and group sparsity constraints. Although similar results are obtained for the denoising of $\mathbf{c}_1$ with perhaps a slightly cleaner signal obtained from sparsity constraints for the SNR = 2 signal, the advantages of applying group sparsity constraints to the inversion process are evident. These advantages are probably best highlighted with the noise attenuated signals of $\mathbf{c}_3$ where sparsity constrained inversion was unable to recover the relatively low amplitude second event for all SNR's unlike group sparsity constrained inversion. For $\mathbf{c}_2$, group sparsity constrained inversion was able to recover the first event for the SNR = 2 signal in contrast to sparsity constrained inversion and was also able to preserve the phase information of the second event more efficiently. These differences can be directly seen in the local time-frequency representations obtained from the noisy synthetics as depicted in Figures 6.6

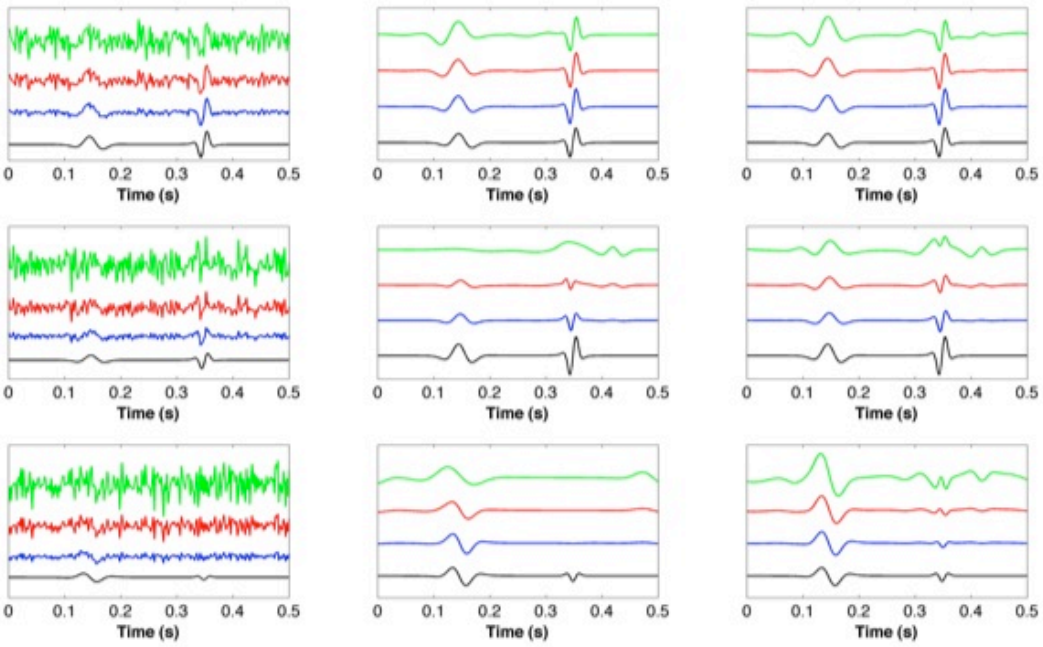


Figure 6.5: A comparison of sparsity and group sparsity denoising for the multicomponent synthetic time-series described in Table 6.1. The top row represents $\mathbf{c}_1$, the middle row represents $\mathbf{c}_2$, and the bottom row represents $\mathbf{c}_3$. In the left column the multicomponent synthetic time-series is displayed with no noise (black), a SNR = 10 (blue), a SNR = 4 (red), and a SNR = 2 (green). The middle column displays the denoising results based upon sparsity constraints with the noise free synthetic signal (black). The right column displays the denoising results based upon group sparsity constraints with the noise free synthetic signal (black).

and 6.7. Despite the more coefficients in the local time-frequency representation obtained with group sparsity, similar frequency content at particular time values are seen across all components. This characteristic enables group sparsity constrained inversion to maintain smaller amplitude information, even when it is below the noise level, more efficiently if another component has a larger amplitude event of similar frequency content at the same particular time. When sparsity constraints are utilized, information from other components have no influence in the generation of the local time-frequency representation leading to a greater possibility of smaller amplitude events being eliminated as illustrated by the lack of recovery of the second event in $\mathbf{c}_2$. It should be noted that the quality of the noise attenuation efforts of both sparsity constrained and group sparsity constrained inversion degrades as the noise level increases. Real data tests using group sparsity constraints for denoising with inverse spectral decomposition were performed on microseismic data by Ismael Vera Rodriguez in (Vera Rodriguez et al., 2012).

6.4 Summary

Through regularizing the inverse spectral decomposition method developed in Chapter 3, the potential exists to attenuate random noise within the observed data. By creating a sparse local time-frequency representation, ℓ_1 norm regularization tries to represent the signal in as few coefficients as possible. A byproduct of this characteristic is that random noise is not easily fit since it requires a greater number of coefficients and therefore it will be attenuated in the predicted or reconstructed data. The degree to which the predicted data fits the observed data is controlled by the trade-off parameter λ and represents the estimated noise in the signal. By increasing λ, more influence is given to the regularization term meaning that a greater amount of noise is expected in the observed data. The selection of λ can be accomplished through the use of Pareto curves if the noise level is known, otherwise, it can be selected through the use of heuristic methods.

However, sparsity constraints may not be the most optimal choice for multicomponent seismic signals. It can be assumed that as a seismic wave is recorded on a multicomponent geophone its frequency content is similar on all components at the same particular time. To preserve this commonality between the components, the mixed $\ell_2 - \ell_1$ norm can be utilized as the regularization term to constrain the local time-frequency representation to exhibit group sparsity. Constraining the local time-frequency representation to be group sparse provides the benefit that small amplitude events can take advantage from information in the other components for denoising. This enables small amplitude events that are below the noise level, which would have been eliminated with denoising through sparsity constraints, to have the possiblity of being preserved if the other components are above the noise level.

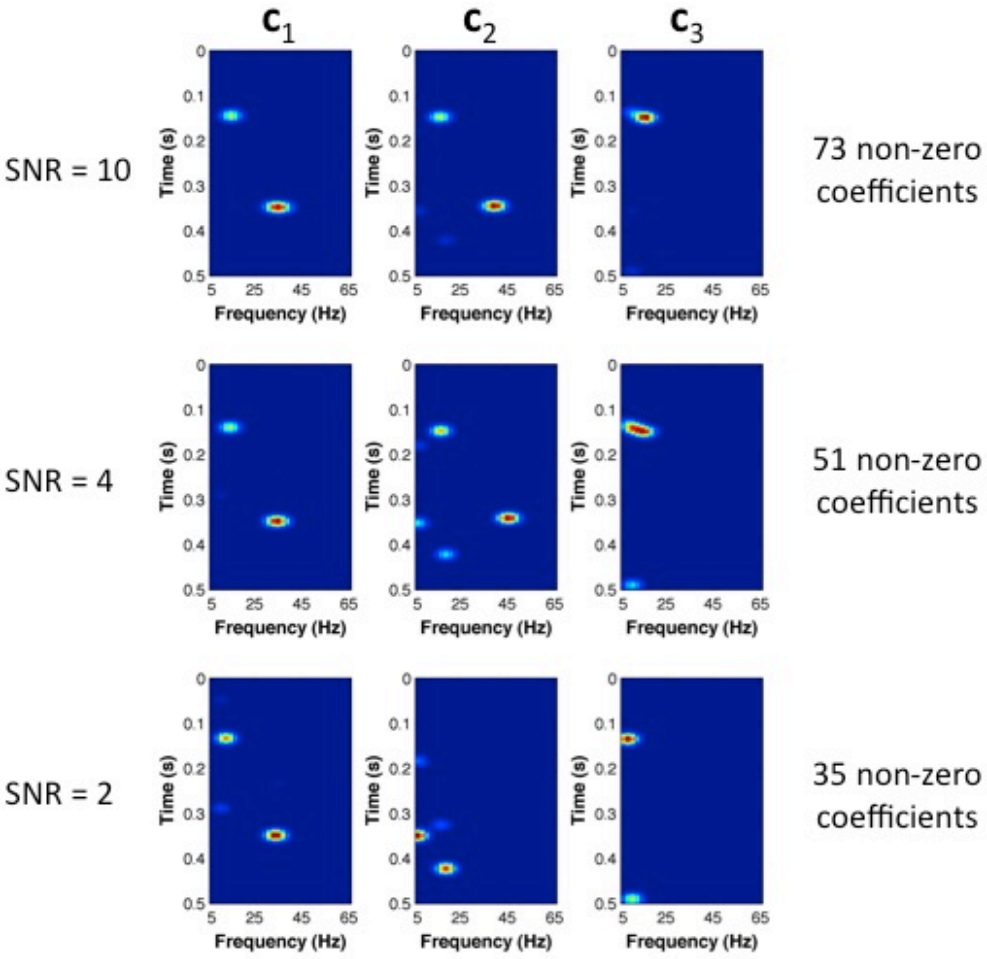


Figure 6.6: The local time-frequency representation of the noise contaminated multicomponent synthetic time-series obtained from sparsity constraints, i.e., the local time-frequency representation of the middle column of Figure 6.5.

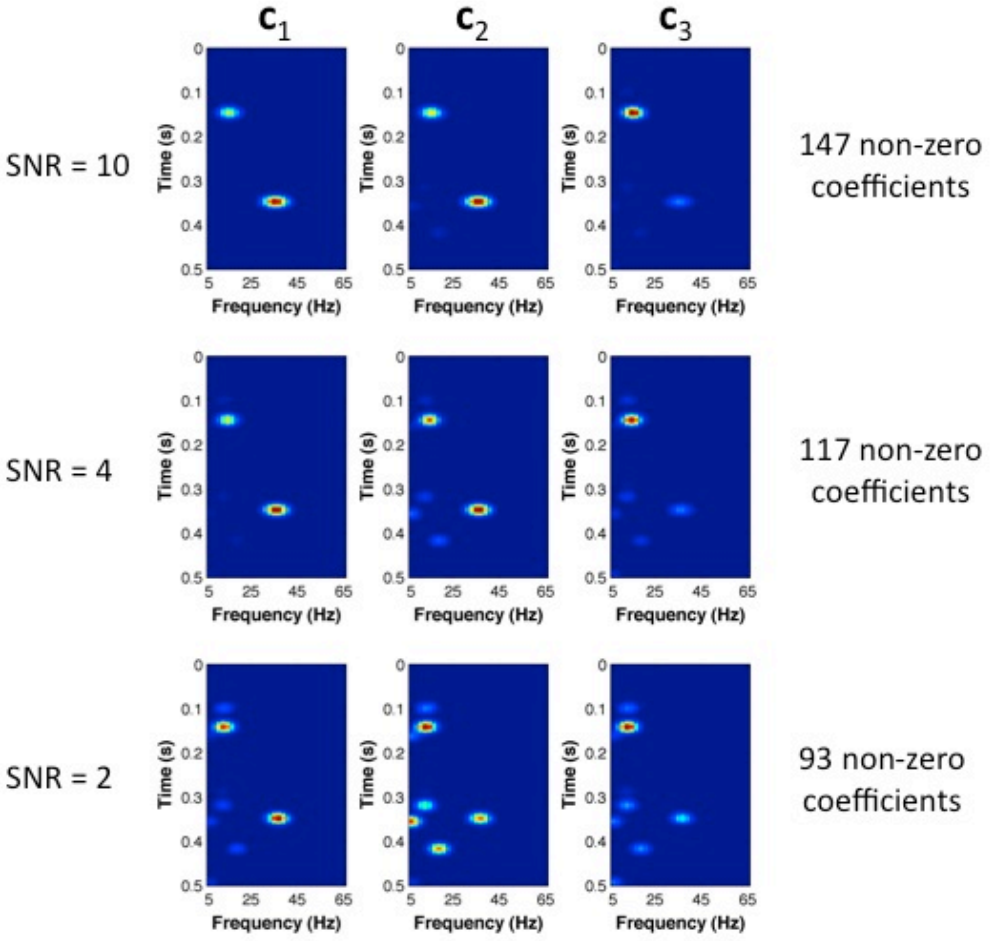


Figure 6.7: The local time-frequency representation of the noise contaminated multicomponent synthetic time-series obtained from group sparsity constraints, i.e., the local time-frequency representation of the right column of Figure 6.5.

CHAPTER 7

Denoising seismic data using the Non-Local Means algorithm[1]

7.1 Introduction

As mentioned in Chapters 5 and 6, the removal or attenuation of noise remains a crucial step for processing seismic data. To effectively mitigate the influence of this noise, one must be specific about what type of noise is desired to be removed. Generally noise within seismic data can be considered to fit into two categories, coherent or incoherent noise. For the purposes of this chapter, the removal of incoherent or random noise is considered.

Random noise attenuation within seismic data has generally been performed using different methods based on different assumptions about the data. For example, band-pass, $f-k$, and k_x-k_y filtering (Yilmaz, 2001) all transform seismic data into the Fourier domain to mute undesirable portions of the signal, i.e., noise, based on the assumption that the signal and noise are separated in this new domain. Other methods such as $f-x$ deconvolution (Canales, 1984), $t-x$ prediction filtering (Abma and Claerbout, 1995), and Cadzow filtering (Trickett, 2008) or Singular Spectrum Analysis (Oropeza and M.Sacchi, 2011) attempt to remove random noise based on assumptions such as the linearity of seismic events.

Originally developed for image processing (Buades et al., 2005), the Non-Local Means (NLM) algorithm is a random noise attenuation filter that assumes the degree of redundancy present within an image can be utilized to reinforce the structures within any small window, or neighborhood, within the image from the many similar windows that are also

[1] A version of this chapter has been published (Bonar and Sacchi, 2012a).

within that image. Thus, the data itself is employed for denoising. Since the NLM algorithm uses the entire image or data set to denoise a single pixel or location, it can become computationally demanding (Buades et al., 2010). Therefore, several variations of the algorithm have arisen to decrease the computational time. Of particular interest is the NLM algorithm's highly parallelizable nature which enables it to be implemented on GPUs (graphical processing units) (Sheng et al., 2009). Other attempts to decrease the computational time for a single pixel or location, such as only considering a portion of the image for denoising, have also been undertaken (Coupé et al., 2008; Mahmoudi and Sapiro, 2005; Brox et al., 2008). Most recently, claims have been made to have decreased the computational time by 15 to 180 times from previous methods by using hashed structures (Dowson and Salvado, 2011) and to have adapted the NLM algorithm to become real-time, or near real-time (de Fontes et al., 2011). The NLM method has been shown to successfully denoise medical data (such as MRI, Coupé et al., 2008), radar data (Deledalle et al., 2011), speech and audio data (Zoican, 2010), and microscopy images (Wei and Yin, 2010). We propose to adopt the NLM algorithm for attenuating the random noise within seismic data. In particular, we use the traditional NLM algorithm originally developed by Buades et al. (2005) for random noise attenuation in a time slice extracted from a spectrally decomposed volume depicting a small channel, in a synthetic data set containing curved events and sharp discontinuities, and in a noisy seismic cross section. It should be noted that the NLM algorithm is not only limited to denoising in the time domain, but also has the potential to be implemented in the frequency domain, like $f - x$ deconvolution, or even integrated into a POCS (projection onto convex sets) like algorithm where each iteration from POCS is denoised using NLM (Huang et al., 2011).

7.2 Theory

We describe the NLM denoising algorithm based on the description provided by Buades et al. (2010). Let a discrete noise contaminated image v be defined by

$$v = u + n, \tag{7.1}$$

or simply the summation of the original noise free image u with random noise n. At the pixel i, the non local means denoised pixel, $\hat{v}(i)$, is simply the weighted average of all of the pixels within the image v,

$$\hat{v}(i) = \sum_j w_{NLM}(i,j)v(j), \tag{7.2}$$

where the weights $w_{NLM}(i,j)$ depend upon the similarity between the pixels i and j and must satisfy the conditions $0 \leq w_{NLM}(i,j) \leq 1$ and $\sum_j w_{NLM}(i,j) = 1$. Note that each

pixel i of the image has its own independent weights of the other j pixels within the image. To quantify the similarity between the pixels i and j, a neighborhood or window, $\mathcal{N}_i$, around the pixel of interest is defined to allow for information about local structures and textures to be incorporated. The neighborhood of a pixel is generally chosen to be a square or cube, depending upon the dimensionality of the image, with a dimension size of 3 to 9 (Awate and Whitaker, 2006; Coupé et al., 2008; Dowson and Salvado, 2011) centered upon the pixel of interest, however, the size and shape of the neighborhood can vary. The similarity between the pixels i and j is then computed using a Gaussian weighted Euclidean distance, $D_G(i,j)$, between the neighborhood around the pixel i, $v(\mathcal{N}_i)$, and the neighborhood around the pixel j, $v(\mathcal{N}_j)$,

$$D_G{}^2(i,j) = \sum_{l}^{nl} \left[G_a(l)\left(v(\mathcal{N}_i(l)) - v(\mathcal{N}_j(l))\right)\right]^2, \tag{7.3}$$

where G_a represents the Gaussian kernel with standard deviation a, and l represents one of the total nl elements within a neighborhood. For a two dimensional image, the Gaussian kernel, G_a, can be defined by,

$$G_a(x,y) = \exp\left(-\frac{(x-x_o)^2 + (y-y_o)^2}{2a}\right), \tag{7.4}$$

where x_o and y_o denote the center of the Gaussian kernel with x and y corresponding to the coordinates of the element l in equation 7.3. By weighting the Euclidean distance between the neighborhoods of pixels i and j with a Gaussian kernel, smaller weights are assigned to distant pixels within a neighborhood allowing structures closest to the pixel of interest to be more likely preserved. Given the Gaussian weighted Euclidean distance, $D_G(i,j)$, between the pixels i and j, the weights $w_{NLM}(i,j)$ are computed according to

$$w_{NLM}(i,j) = \frac{1}{Z_{NLM}(i)} \exp\left(\frac{-D_G{}^2(i,j)}{h^2}\right), \tag{7.5}$$

where $Z_{NLM}(i)$ is the normalizing factor defined by

$$Z_{NLM}(i) = \sum_{j} \exp\left(\frac{-D_G{}^2(i,j)}{h^2}\right) \tag{7.6}$$

to ensure $\sum_j w_{NLM}(i,j) = 1$. The parameter h is a constant which controls the decay of the exponential function as a function of the Euclidean distance. For example, a large value for h will provide similar weight for all j pixels in the image while a small value for h will provide a significant weight for only a few of the j pixels in the image. In Figure 7.1, a grid has been generated to display the NLM image weights calculated with equation 7.5 for each of the coloured boxes. This image clearly illustrates how the weights depend upon the

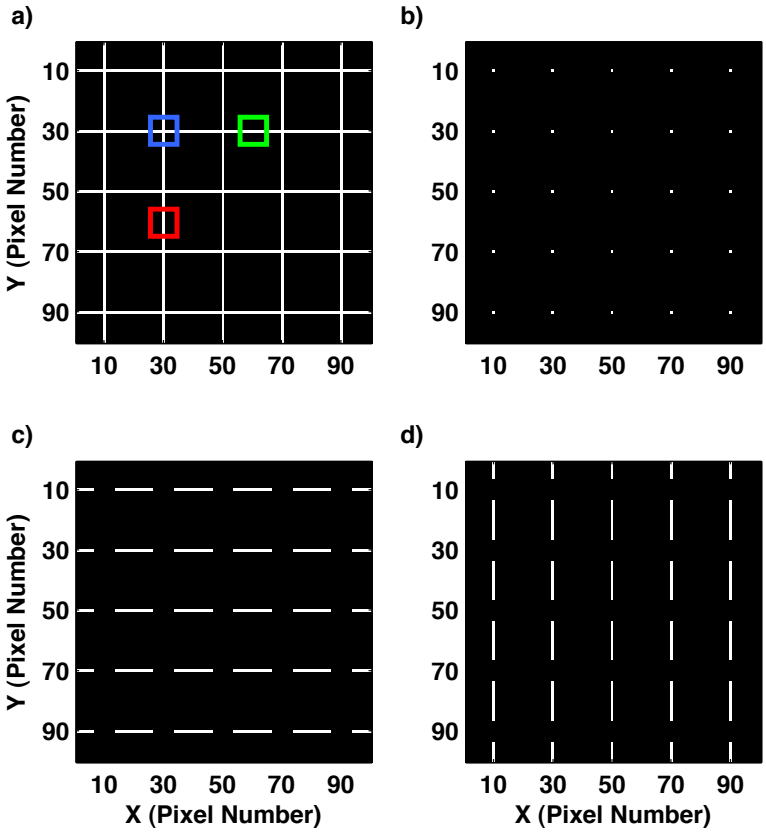


Figure 7.1: a) Example grid to display the pixel weights used by NLM for three different regions of the synthetic image. b) NLM pixel weights for the blue box. c) NLM pixel weights for the green box. d) NLM pixel weights for the red box.

local structure or neighborhood of the pixel of interest and can yet be spatially non-local, as evidenced by the weights for the blue box (Figure 7.1b) which highlight all of the points within the image that contain an intersection between the horizontal and vertical lines, the weights for the green box (Figure 7.1c) which highlight all of the areas that only contain a horizontal line, and the weights for the red box (Figure 7.1d) which highlight all of the areas that contain only a vertical line.

7.3 Examples

To demonstrate how the NLM denoising algorithm can be applied to seismic data, we first look at a time slice extracted from a spectrally decomposed volume of a real data set at

45Hz (Partyka et al., 1999) as seen in Figure 7.2. Although the small channel (only a few traces wide) running vertically through the time slice is more clearly defined in the constant 45Hz section, the amount of noise and residual acquisition footprint degrades the quality of the image. The NLM denoising algorithm, low pass $k_x - k_y$ filtering, and $x - y$ prediction filtering (Abma and Claerbout, 1995) were applied to the noisy constant frequency section to try to attenuate and reduce the effect of this noise. For the purposes of this article, the standard deviation of the Gaussian kernel in the NLM method was kept at a constant value of $a = 0.25$ for all examples and the window or neighbourhood size was chosen to be an 11**x**11 square for this particular example. Of these three denoised images, the NLM image preserves the small channel most effectively since its goal is to explicitly attenuate random noise by reinforcing the structures already within an image, unlike the other two methods whose goals are to limit the frequencies/wavenumbers within the data or to predict the data, attenuating random noise as a byproduct. Although the small channel only occurs once within the data, its structure within a small neighborhood can be replicated several times. This redundancy of structures within a neighborhood allows for the attenuation of random noise using the NLM algorithm, enhancing the visualization of the channel's structure.

A comparison of Figures 7.2e and 7.2f, NLM denoising with $h = 40$ and $h = 100$, depicts the effect of choosing the parameter h to be too large. By increasing h to the extreme, the NLM algorithm essentially assigns a constant value to all of the weights, as stated previously, which produces a constant intensity image. Likewise, if h is chosen to be too small, the NLM algorithm is incapable of denoising images since the weight assigned to the most similar pixel, i.e. the actual pixel of interest, will be too large. Proper estimation of the parameter h can be carried out using methods such as the χ^2 criterion for the goodness of fit (Aster et al., 2005) if the standard deviation of the noise is known or heuristically, like the estimation of the prediction filter length in $f - x$ deconvolution, if little is known about the noise. For consistency, the parameter h was chosen heuristically throughout this article. In other words, we visually inspected the denoised image for both noise attenuation and structure preservation to determine the optimal value for the parameter h. Typically this value was found to be around one order of magnitude less than the largest amplitude present within the data.

One of the specific advantages that the NLM algorithm can have for attenuating random noise in seismic data is its ability to handle sharp discontinuities, such as faults, and events with curvature. To illustrate these capabilities, a synthetic data set containing linear, curved, and crossing events with sharp discontinuities was created and contaminated with random noise such that the signal to noise ratio (SNR) was 0.8. Here, the SNR is defined as the maximum amplitude of the data divided by the maximum amplitude of the noise. The denoising of this synthetic data set, as shown in Figure 7.3, was performed using NLM with

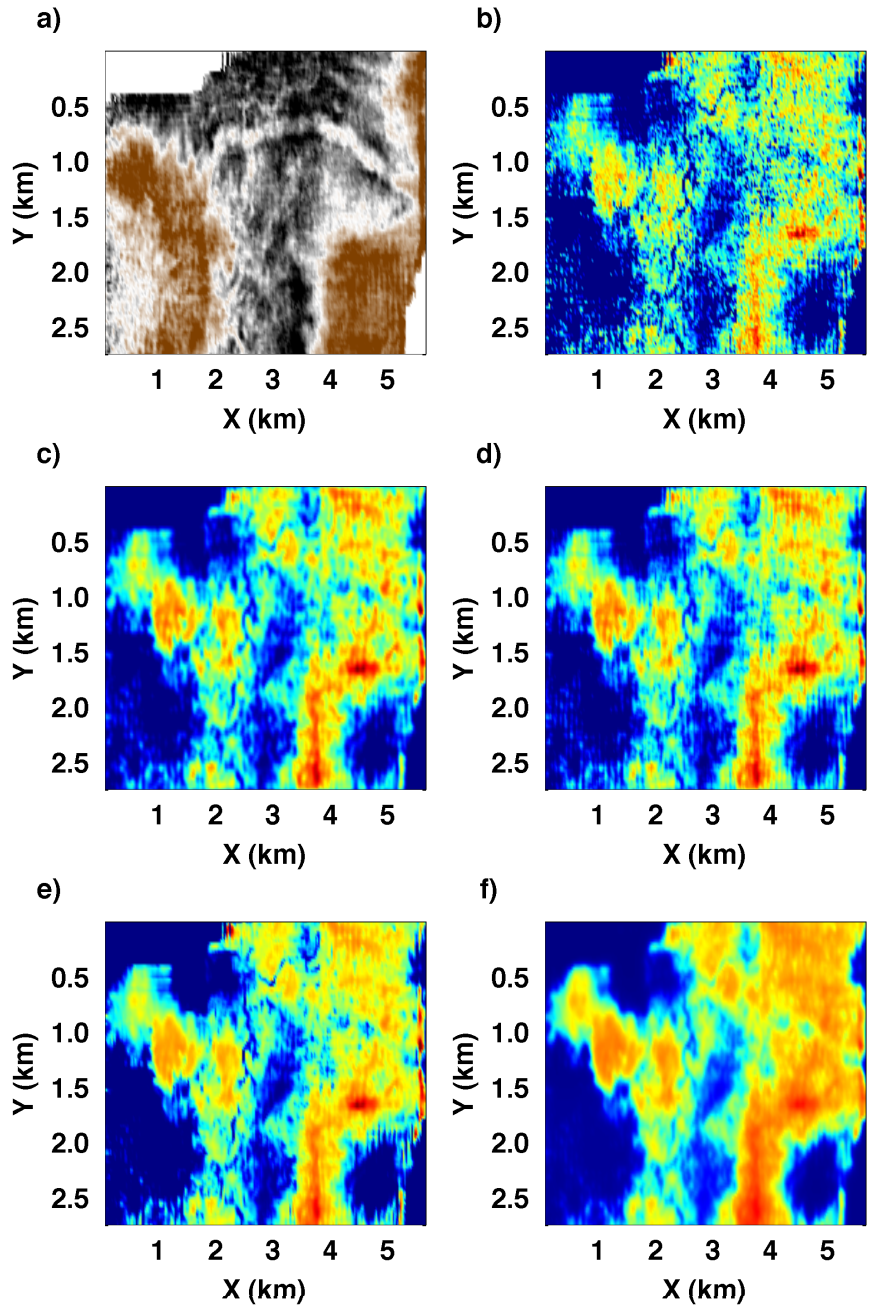


Figure 7.2: a) Real data time slice depicting a small channel running vertically. b) Constant 45Hz slice of a) obtained from spectral decomposition. c) Denoised version of b) using low pass $k_x - k_y$ filtering. d) Denoised version of b) using 2D prediction filtering. e) Denoised version of b) using NLM with $h = 40$. f) Denoised verion of b) using NLM with $h = 100$.

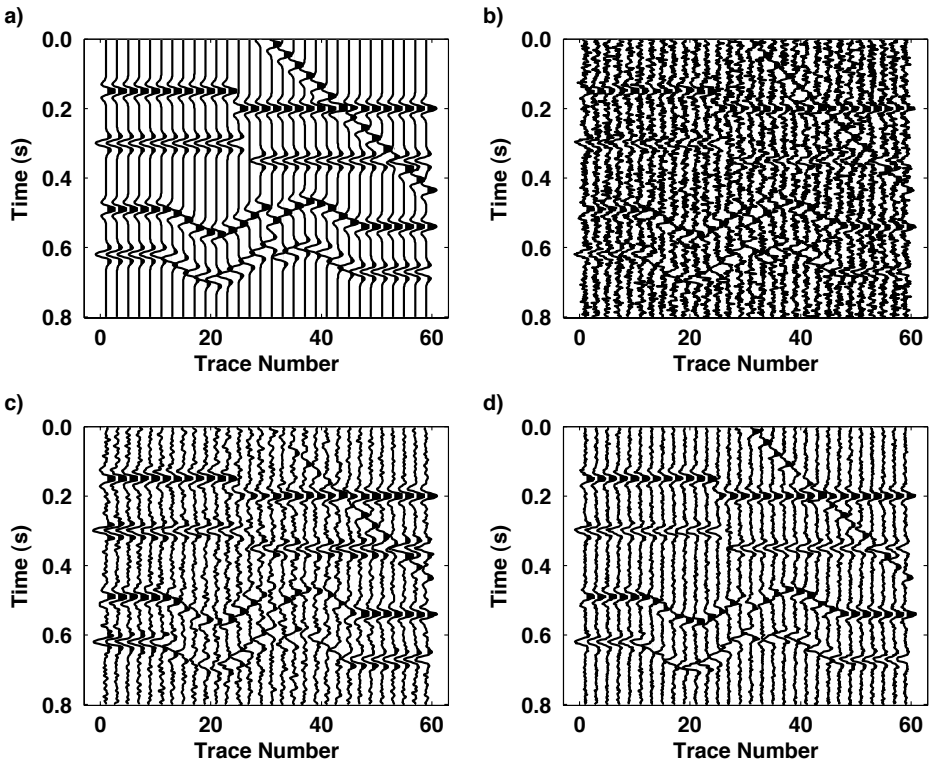


Figure 7.3: a) Synthetic data with linear, curved, and crossing events with sharp discontinuities. b) Synthetic data contiminated with random noise to a SNR = 0.8. c) Denoising with $f - x$ deconvolution. d) Denoising with NLM. Note how the amplitude of the curved events are more effectively conserved and the lack of energy smeared across the sharp discontinuities with NLM.

$h = 0.15$ and an 11**x**11 square window and $f - x$ deconvolution that was applied in small spatial windows, 11 traces wide, to help validify its plane wave assumption. Unlike $f - x$ deconvolution, the NLM algorithm preserves the amplitude of the curved events and does not smear seismic energy across the sharp discontinuities. Figure 7.4 depicts the Fourier amplitude spectrum of the synthetic data and its denoised versions and highlights the ability of NLM to adequately maintain high frequencies as compared to $f - x$ deconvolution, even though it is an averaging process.

As a final example, the NLM algorithm was applied to a real poststack seismic cross section as seen in Figure 7.5. For optimal results, the parameter h was chosen to be 0.2 with a window size of 17**x**17. The larger window size was required to fully encompass a peak and a trough signature within the seismic data (i.e., the structures that are trying to be denoised). The results from denoising the cross section using NLM allows for the character of

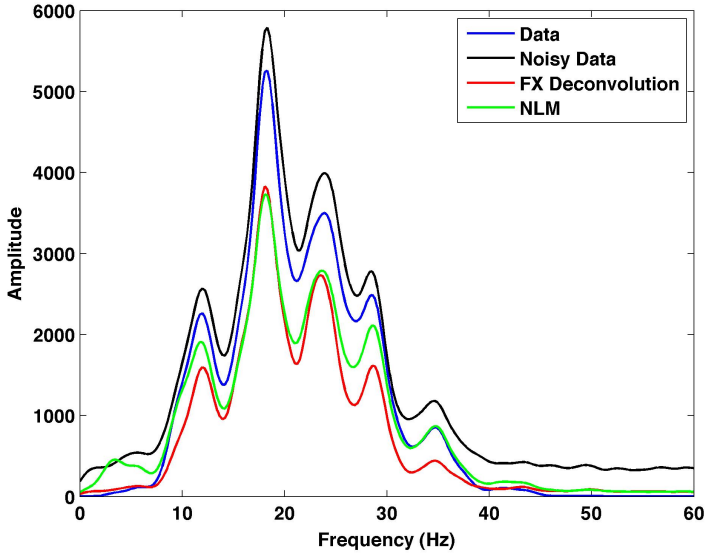


Figure 7.4: The Fourier amplitude spectrum of the data in Figure 7.3.

the reflectors to become more apparent while not losing any of their energy to the estimated noise. Because of the adaptive nature of this method, the amount of noise which is removed varies as evidenced from the greater amount of estimated noise at later times where the original data appears more noisy. Like the synthetic data example, the NLM algorithm maintains higher frequencies comparably well to other denoising methods, such as $f - x$ deconvolution, as shown in Figure 7.6. Although this dataset lacks sharp discontinuities or significant events with curvature, it demonstrates that the NLM algorithm can be just as effective for attenuating random noise in seismic data as more commonly used denoising methods such as $f - x$ deconvolution highlighted by the amplitude comparison shown in Figure 7.7. However, it should be noted that to achieve these results, the simple implementation of the NLM algorithm had a computational time that was greater than two orders of magnitude than that of $f - x$ deconvolution. To become more feasible for application with seismic data, further research into decreasing the computational time of NLM needs to be conducted.

7.4 Conclusions

The NLM algorithm is a random noise attenuation filter that utilizes the redundancy of structures within a data set to denoise each location within it. Unlike other common denoising methods for seismic data, the concepts governing NLM makes it less likely to smooth

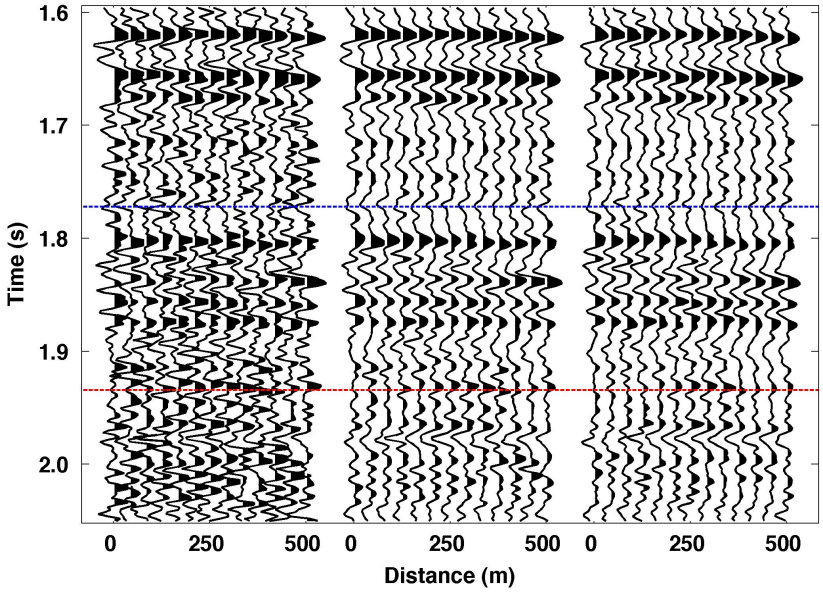


Figure 7.5: Real poststack data set denoised using NLM and $f - x$ deconvolution (from left to right: noisy data, NLM denoised data, $f - x$ deconvolution denoised data). The dotted blue and red lines indicate the horizons utilized by Figure 7.7.

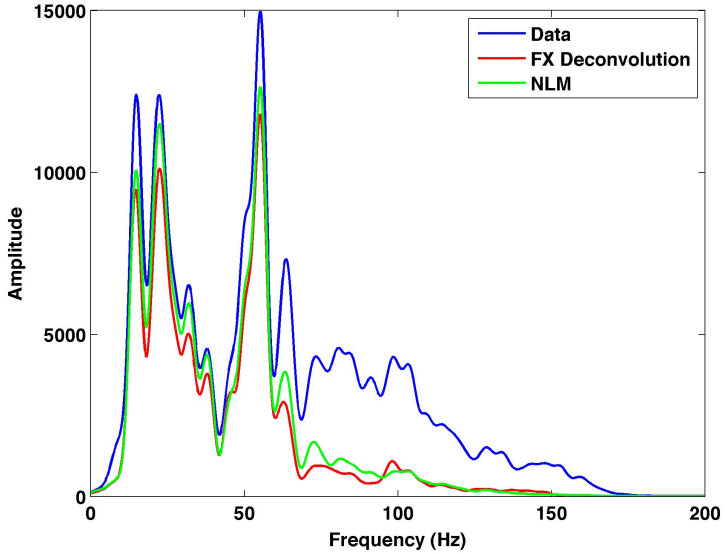
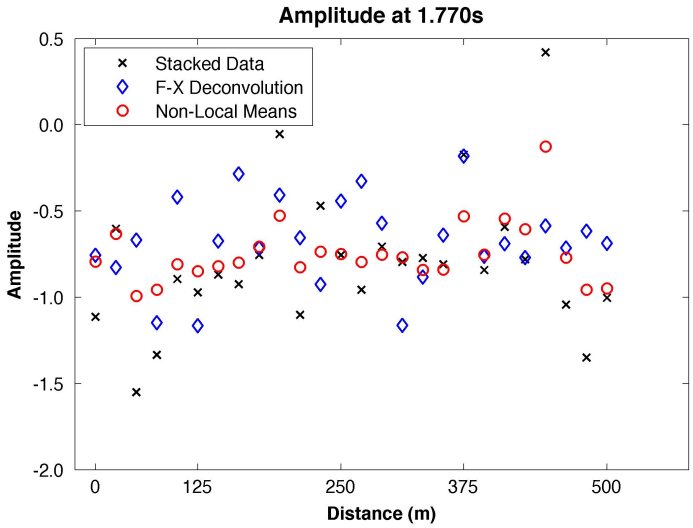


Figure 7.6: The Fourier amplitude spectrum of the data in Figure 7.5.

a)



b)

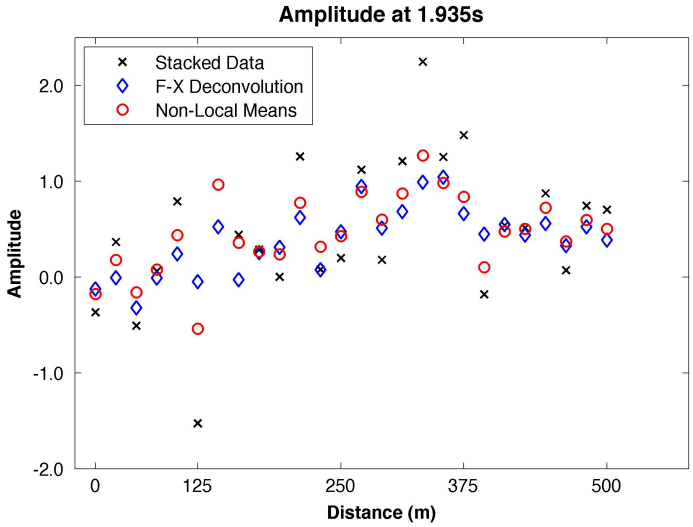


Figure 7.7: Amplitude comparison for the two highlighted horizons in Figure 7.5. The NLM algorithm compares quite favourably to $f-x$ deconvolution at removing the random noise across these horizons.

over small features, such as small channels, while enabling it to handle curvature and faults without losing the resolution of these features. In its basic implementation there are three parameters to be controlled: the window or neighborhood size, the standard of deviation a for the Gaussian kernel, and the parameter h. While the algorithm is relatively insensitive to the parameter a since it only controls the weights at the edges of a neighborhood, the neighborhood size must be chosen such that it is large enough to encapsulate the structures of interest within the data. For reflection seismic data, this means that the neighborhood size, in the time dimension, should be larger than the wavelet length. The parameter h controls the level of denoising the NLM algorithm performs and for the examples provided was found to be optimally set to be about one order of magnitude less than the maximum amplitude of the data. Comparisons on real seismic data illustrate that this denoising method can produce similar results to more commonly used denoising methods such as $f - x$ deconvolution. The NLM algorithm can be easily expanded to incorporate multiple dimensions and has the potential to become a common denoising method in either prestack or poststack seismic data because of its abilities to not smear seismic energy at termination points or sharp discontinuities (i.e., faults).

CHAPTER 8

Conclusions

Time-frequency analysis has traditionally been accomplished through the use of the Fourier transform through converting a signal from the time domain into the frequency domain. However when studying signals that vary their frequency content with time, this type of time-frequency analysis has difficulties locating where specific frequencies are present or absent within the time domain signal. To solve this non-unique problem, local time-frequency analysis also known as spectral decomposition must be employed. Chapter 2 provides a review of the general strategies most commonly applied within geophysics. Some of these strategies include the analysis of the complex signal (instantaneous frequency), the application of the Fourier transform within small time windows (STFT), and the representation of the signal with time-frequency atoms (CWT). Perhaps the most generalized of these concepts is the CWT which can be manipulated to express other local time-frequency transforms, such as the STFT, through an appropriate choice of the time-frequency atoms known as wavelets.

In an effort to create a high resolution local time-frequency representation, Chapter 3 presents an inversion method based upon a multi-wavelet convolutional model that was regularized with the ℓ_1 norm of the model parameters. The forward and adjoint operator, or the mapping from model space to data space and the mapping from data space to model space, of this multi-wavelet convolutional model simply become the CWT and inverse CWT, respectively. Through regularizing the inverse problem with the ℓ_1 norm of the model parameters a sparse local time-frequency representation is obtained. This is in direct contrast to the smooth solution, known as the damped least squares solution, that would have been obtained through regularization utilizing the ℓ_2 norm of the model solutions. However, the drawback of utilizing the ℓ_1 norm for regularization is that it is much more computationally expensive in comparison to ℓ_2 norm regularization. A modification to the

iterative soft thresholding strategy for solving ℓ_1 norm regularized inversion called FISTA (Fast Iterative Soft Thresholding Algorithm) was utilized to help address this problem. As with all regularized inverse problems, a trade-off parameter must be selected to balance the emphasis placed upon the measurements for the data misfit and regularization terms.

The application of the high resolution spectral decomposition inversion method to a 3D poststack seismic data set is shown in Chapter 4. Through applying sparsity constraints on the local time-frequency representation, increased variability can be seen between constant frequency sections of the local time-frequency representation. This property helped to highlight the different frequency behaviours of two large paleochannels within the data set. Although sparsity provides high resolution local time-frequency representations, it cannot provide a continuous phase like the damped least squares solution for the local time-frequency representation since many of the coefficients of the sparse local time-frequency representation are zero. Despite a general lack of published interpretation of the phase from local time-frequency representations, the continuous phase definition of the damped least squares solution for the local time-frequency representation may help provide further insight into different scales of stratigraphic layering.

By obtaining a high resolution local time-frequency representation through sparsity constrained inversion, it was noticed that noise has a more profound effect on the local time-frequency representation. Chapter 5 addresses this issue through preconditioning the ℓ_1 norm regularized inversion method developed in Chapter 3 with concepts obtained from $f-x-y$ deconvolution. Instead of considering the spectral decomposition problem as a transformation of a one dimensional signal into a two dimensional local time-frequency representation, Chapter 5 views it is as a multidimensional problem to allow for information from neighbouring seismic traces to influence the local time-frequency representation obtained for a particular seismic trace. This information is incorporated through predictive filtering in a manner similar to $f-x-y$ deconvolution which assumes that the subsurface does not contain large lateral variations. The preconditioning process helps to generate a denoised version of the local time-frequency representation while maintaining the integrity of small scale geological features such as small paleochannels. This is in contrast to denoising after the local time-frequency representation has been obtained which tends to also remove these small scale geological features from the local time-frequency representation.

Through posing the spectral decomposition problem as an ℓ_1 norm regularized inversion, random noise can be attenuated from the time domain signal. Chapter 6 explores this problem, more formally known as Basis Pursuit Denoising, which is solely controlled by the trade-off parameter. The selection of the optimal trade-off parameter based upon the noise level of the signal is highlighted through the utilization of Pareto curves. Chapter 6 also explores how the information from different components of a multicomponent signal

can aid in the denoising process. By regularizing the inverse problem with group sparsity constraints rather than sparsity constraints, the preservation of events well below the noise level is possible if the event has a larger amplitude in another component.

Chapter 7 presents the application of a random noise attenuation method called the Non-Local Means (NLM) algorithm, which was originally developed for image processing, for seismic data to aid the data processing procedure before local time-frequency analysis occurs. Through weighted averaging, this algorithm provides promising random noise attenuation results in areas of curvature and sharp discontinuities where traditional random noise attenuation methods, such as $f - x$ deconvolution, have difficulties. However, it should be noted that these improvements come at a much greater computational cost. The various acceleration modifications that have been provided in the literature for NLM were not investigated within this thesis but may allow this algorithm to become more practical for application to large seismic data sets.

Future endeavours to be explored from the works of this thesis include a comparison of ℓ_1 norm regularized inversion to the reassignment method for generating high resolution local time-frequency representations, a study of the advantages and disadvantages of group sparsity and quaternions for vector processing of multicomponent seismic data, the computational acceleration of NLM to make it more feasible for application with seismic data, and the application of the multi-wavelet convolutional model for thin bed reflectivity inversion following the concepts developed in Puryear and Castagna (2008).

Bibliography

Abma, R. and J. Claerbout. "Lateral prediction for noise attenuation by $t - x$ and $f - x$ techniques." *Geophysics* 60 (1995): 1887–1896.

Alemie, W. and M. Sacchi. "High-resolution three-term AVO inversion by means of a Trivariate Cauchy probability distribution." *Geophysics* 76 (2011): R43–R55.

Anagaw, A. and M. Sacchi. "Regularized 2D acoustic full waveform inversion." *EAGE Expanded Abstracts* (2011).

Askari, R. and H. Siahkoohi. "Ground roll attenuation using the S and $x-f-k$ transforms." *Geophysical Prospecting* 56 (2008): 105–114.

Aster, R., C. Thurber, and B. Borchers. *Parameter Estimation and Inverse Problems.* International Geophysics Series. Elsevier Academic Press, 2005.

Auger, F. and P. Flandrin. "Improving the readability of time-frequency and time-scale representations by the reassignment method." *IEEE Transactions on Signal Processing* 43 (1995): 1068–1089.

Awate, S. and R. Whitaker. "Unsupervised, information-theoretic, adaptive image filtering for image restoration." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2006): 364 –376.

Barnes, A. "The complex seismic trace made simple." *The Leading Edge* 17 (1998): 473–476.

Beck, A. and M. Teboulle. "A fast iterative shrinkage-thresholding algorithm for linear inverse problems." *SIAM Journal on Imaging Sciences* 2 (2009): 183–202.

Bendat, J. and A. Piersol. *Random Data - Analysis and Measurment Procedures.* Third edition. Wiley-Interscience Publication, 2000.

Bioucas-Dias, J. and M. Figueiredo. "A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration." *IEEE Transactions on Image Processing* 16 (2007): 2992–3004.

Bodine, J. "Waveform analysis with seismic attributes." *SEG Expanded Abstracts* 3 (1984): 505–509.

Bonar, D. and M. Sacchi. "Complex spectral decomposition via inversion strategies." *SEG Expanded Abstracts* 29 (2010): 1408–1412.

Bonar, D. and M. Sacchi. "Denoising seismic data using the nonlocal means algorithm." *Geophysics* 77 (2012): A5–A8.

Bonar, D. and M. Sacchi. "Spectral decomposition with $f - x - y$ preconditioning." Submitted for publication. *Geophysical Prospecting* (2012).

Bradford, J.. "Frequency-dependent attenuation analysis of ground-penetrating radar data." *Geophysics* 72 (2007): J7–J16.

Bradford, J. and Y. Wu. "Instantaneous spectral analysis: time-frequency mapping via wavelet matching with application to contaminated-site characterization by 3D GPR." *The Leading Edge* 26 (2007): 1018–1023.

Brox, T., O. Kleinschmidt, and D. Cremers. "Efficient nonlocal means for denoising of textural patterns." *IEEE Transactions on Image Processing* 17 (2008): 1083 –1092.

Buades, A., B. Coll, and J. Morel. "A review of image denoising algorithms, with a new one." *Multiscale Modeling and Simulation* 4 (2005): 490–530.

Buades, A., B. Coll, and J. Morel. "Image denoising methods. A new nonlocal principle." *SIAM Review* 52 (2010): 113–147.

Bube, K. and R. Langan. "A continuation approach to regularization of ill-posed problems with application to crosswell-traveltime tomography." *Geophysics* 73 (2008): VE337–VE351.

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

Canales, L. "Random noise reduction." *SEG Expanded Abstracts* 3 (1984): 525–527.

Candes, E., J. Romberg, and T. Tao. "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information." *IEEE Transactions on Information Theory* 52 (2006): 489–509.

Candes, E. and T. Tao. "Near-optimal signal recovery from random projections: universal encoding strategies?." *IEEE Transactions on Information Theory* 52 (2006): 5406–5425.

Candes, E. and M. Wakin. "An introduction to compressive sensing." *IEEE Signal Processing Magazine* 25 (2008): 21–30.

Castagna, J., S. Sun, and R. Siegfried. "Instantaneous spectral analysis: Detection of low-frequency shadows associated with hydrocarbons." *The Leading Edge* 22 (2003): 22.

Chakraborty, A. and D. Okaya. "Frequency-time decomposition of seismic data using wavelet-based methods." *Geophysics* 60 (1995): 1906–1916.

Chase, M. "Random noise reduction by 3-D spatial prediction filtering." *SEG Expanded Abstracts* 11 (1992): 1152–1153.

Chen, S., D. Donoho, and M. Saunders. "Atomic decomposition by basis pursuit." *SIAM Journal on Scientific Computing* 20 (1998): 33–61.

Chopra, S. and K. Marfurt. "Seismic attributes — A historical perspective." *Geophysics* 70 (2005): 3SO–28SO.

Claerbout, J. *Earth Soundings Analysis: Processing Versus Inversion*. Stanford Exploration Project. Blackwell Scientific Publications, 1992.

Clapp, R., B. Biondi, and J. Claerbout. "Incorporating geologic information into reflection tomography." *Geophysics* 69 (2004): 533–546.

Cohen, L. "Instantaneous 'anything'." *IEEE International Conference on Acoustics, Speech, and Signal Processing* 4 (1993): 105–108.

Cohen, L. and C. Lee. "Instantaneous frequency and time-frequency distributions." *IEEE International Symposium on Circuits and Systems* 2 (1989): 1231–1234.

Cotter, S., B. Rao, K. Engan, and K. Kreutz-Delgado. "Sparse solutions to linear inverse problems with multiple measurement vectors." *IEEE Transactions on Signal Processing* 53 (2005): 2477–2488.

Coupé, P., P. Yger, S. Prima, P. Hellier, C. Kervrann, and C. Barillot. "An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images." *IEEE Transactions on Medical Imaging* 27 (2008): 425–441.

Daubechies, I. "Orthonormal bases of compactly supported wavelets." *Communications on Pure and Applied Mathematics* 41 (1988): 909–996.

Daubechies, I. *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics, 1992.

Daubechies, I., M. Defrise, and C. Mol. "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint." *Communications on Pure and Applied Mathematics* 57 (2004): 1413–1457.

Daubechies, I., R. DeVore, M. Fornasier, and C. Güntürk. "Iteratively reweighted least squares minimization for sparse recovery." *Communications on Pure and Applied Mathematics* 63 (2010): 1–38.

de Fontes, F., G. Barroso, P. Coupé, and P. Hellier. "Real time ultrasound image denoising." *Journal of Real-Time Image Processing* 6 (2011): 15–22.

de Matos, M., O. Davogustto, K. Zhang, and K. Marfurt. "Detecting stratigraphic discontinuities using time-frequency seismic phase residues." *Geophysics* 76 (2011): P1–P10.

Deighan, A. and D. Watts. "Ground-roll suppression using the wavelet transform." *Geophysics* 62 (1997): 1896–1903.

Deledalle, C., L. Denis, and F. Tupin. "NL-InSAR: Nonlocal interferogram estimation." *IEEE Transactions on Geoscience and Remote Sensing* 49 (2011): 1441 –1452.

Donoho, D. "Compressed sensing." *IEEE Transactions on Information Theory* 52 (2006): 1289–1306.

Donoho, D. "For most large underdetermined systems of linear equations, the minimal ell-1 norm solution is also the sparsest solution." *Communications on Pure and Applied Mathematics* 59 (2006): 797–829.

Dowson, N. and O. Salvado. "Hashed nonlocal means for rapid image filtering." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (2011): 485–499.

Elboth, T., I. Presterud, and D. Hermansen. "Time-frequency seismic data de-noising." *Geophysical Prospecting* 58 (2010): 441–453.

Eldar, Y. and H. Bolcskei. "Block-sparsity: Coherence and efficient recovery." *IEEE International Conference on Acoustics, Speech, and Signal Processing* (2009): 2885–2888.

Fornasier, M. and H. Rauhut. "Recovery algorithms for vector-valued data with joint sparsity constraints." *SIAM Journal on Numerical Analysis* 46 (2008): 577–613.

Gabor, D. "Theory of communication." *Journal of the Institute of Electrical Engineers* 93 (1946): 429–457.

Gardner, T. and M. Magnasco. "Sparse time-frequency representations." *Proceedings of the National Academy of Sciences* 103 (2006): 6094–6099.

Giroldi, L. and F. Alegria. "Using spectral decomposition to identify and characterize glacial valleys and fluvial channels within the Carboniferous section in Bolivia." *The Leading Edge* 24 (2005): 1152–1159.

Golub, G. and C. Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, 1996.

Goupillaud, P., A. Grossmann, and J. Morlet. "Cycle-octave and related transforms in seismic signal analysis." *Geoexploration* 23 (1984): 85–102.

Gulunay, N., V. Sudhakar, C. Gerrard, and D. Monk. "Prediction filtering for 3-D poststack data." *SEG Expanded Abstracts* 12 (1993): 1183–1186.

Hansen, P. *Rank-Deficient and Discrete Ill-posed Problems - Numerical Aspects of Linear Inversion*. SIAM, 1998.

Hardy, H., R. Beier, and J. Gaston. "Frequency estimates of seismic traces." *Geophysics* 68 (2003): 370–380.

Henderson, J., S. Purves, and C. Leppard. "Automated delineation of geological elements from 3D seismic data through analysis of multichannel, volumetric spectral decomposition data." *First Break* 25 (2007): 87–93.

Hennenfent, G., E. van den Berg, M. Friedlander, and F. Herrmann. "New insights into one-norm solvers from the Pareto curve." *Geophysics* 73 (2008): A23–A26.

Herrmann, F. "Compressive imaging by wavefield inversion with group sparsity." *SEG Expanded Abstracts* 28 (2009): 2337–2341.

Hestenes, M. and E. Stiefel. "Methods of conjugate gradients for solving linear systems." *Journal of Research of the National Bureau of Standards* 49 (1952): 409–436.

Hornby, P., F. Boschetti, and F. Horowitz. "Analysis of potential field data in the wavelet domain." *Geophysical Journal International* 137 (1999): 175–196.

Huang, J., J. Ma, N. Liu, H. Zhang, Z. Bian, Ya. Feng, Q. Feng, and W. Chen. "Sparse angular CT reconstruction using non-local means based iterative-correction POCS." *Computers in Biology and Medicine* 41 (2011): 195–205.

Hyder, M. and K. Mahata. "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation." *IEEE Transactions on Signal Processing* 58 (2010): 4646–4655.

Kaplan, S., M. Naghizadeh, and M. Sacchi. "Data reconstruction with shot-profile least-squares migration." *Geophysics* 75 (2010): WB121–WB136.

Kodera, K., C. de Villedary, and R. Gendrin. "A new method for the numerical analysis of non-stationary signals." *Physics of the Earth and Planetary Interiors* 12 (1976): 142–150.

Kowalski, M. and B. Torrésani. "Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients." *Signal, Image and Video Processing* 3 (2009): 251–264.

Kuehl, H. and M. Sacchi. "Least-squares wave-equation migration for AVP/AVA inversion." *Geophysics* 68 (2003): 262–273.

Landweber, L. "An iteration formula for Fredholm integral equations of the first kind." *American Journal of Mathematics* 73 (1951): 615–624.

Li, Y. and X. Zheng. "Spectral decomposition using Wigner-Ville distribution with applications to carbonate reservoir characterization." *The Leading Edge* 27 (2008): 1050–1057.

Li, Y., X. Zheng, and Y. Zhang. "High-frequency anomalies in carbonate reservoir characterization using spectral decomposition." *Geophysics* 76 (2011): V47–V57.

Liu, J. and K. Marfurt. "Instantaneous spectral attributes to detect channels." *Geophysics* 72 (2007): P23–P31.

Mahmoudi, M. and G. Sapiro. "Fast image and video denoising via nonlocal means of similar neighborhoods." *IEEE Signal Processing Letters* 12 (2005): 839–842.

Mallat, S. "A theory for multiresolution signal decomposition: the wavelet representation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11 (1989): 674 –693.

Mallat, S. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 2008.

Mallat, S. and Z. Zhang. "Matching pursuits with time-frequency dictionaries." *IEEE Transactions on Signal Processing* 41 (1993): 3397–3415.

Marfurt, K. and R. Kirlin. "Narrow-band spectral analysis and thin-bed tuning." *Geophysics* 66 (2001): 1274–1283.

Martelet, G., P. Sailhac, F. Moreau, and M. Diament. "Characterization of geological boundaries using 1-D wavelet transform on gravity data: Theory and application to the Himalayas." *Geophysics* 66 (2001): 1116–1129.

Morlet, J., G. Arens, E. Fourgeau, and D. Glard. "Wave propagation and sampling theory - Part I: Complex signal and scattering in multilayered media." *Geophysics* 47 (1982): 203–221.

Naghizadeh, M. and K. Innanen. "Seismic data interpolation using a fast generalized Fourier transform." *Geophysics* 76 (2011): V1–V10.

Odebeatu, E., J. Zhang, M. Chapman, E. Liu, and X. Li. "Application of spectral decomposition to detection of dispersion anomalies associated with gas saturation." *The Leading Edge* 25 (2006): 206–210.

Oldenburg, D., P. Mcgillivray, and R. Ellis. "Generalized subspace methods for large-scale inverse problems." *Geophysical Journal International* 114 (1993): 12–20.

Oropeza, V. and M.Sacchi. "Simultaneous seismic data denoising and reconstruction via multichannel singular spectrum analysis." *Geophysics* 76 (2011): V25–V32.

Parolai, S. "Denoising of seismograms using the S transform." *Bulletin of the Seismological Society of America* 99 (2009): 226–234.

Partyka, G., J. Gridley, and J. Lopez. "Interpretational applications of spectral decomposition in reservoir characterization." *The Leading Edge* 18 (1999): 353–360.

Pedersen, H., J. Mars, and P. Amblard. "Improving surface-wave group velocity measurements by energy reassignment." *Geophysics* 68 (2003): 677–684.

Pinnegar, C. and L. Mansinha. "The S-transform with windows of arbitrary and varying shape." *Geophysics* 68 (2003): 381–385.

Plessix, R. and Q. Cao. "A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium." *Geophysical Journal International* 185 (2011): 539–556.

Portniaguine, O. and J. Castagna. "Inverse spectral decomposition." *SEG Expanded Abstracts* 23 (2004): 1786–1789.

Puig, A., A. Wiesel, G. Fleury, and A. Hero. "Multidimensional shrinkage-thresholding operator and group LASSO penalties." *IEEE Signal Processing Letters* 18 (2011): 363 –366.

Puryear, C. and J. Castagna. "Layer-thickness determination and stratigraphic interpretation using spectral inversion: Theory and application." *Geophysics* 73 (2008): R37–R48.

Reine, C., M. van der Baan, and R. Clark. "The robustness of seismic attenuation measurements using fixed- and variable-window time-frequency transforms." *Geophysics* 74 (2009): WA123–WA135.

Robinson, E. and S. Treitel. *Geophysical Signal Analysis*. Prentice-Hall, Inc., 1980.

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

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

Selesnick, I. "Sparse Signal Restoration." *Connexions* (2009): <http://cnx.org/content/m32168/1.3/>.

Sheng, B., P. Li, and H. Sun. "Image-based material restyling with fast non-local means filtering." *International Conference on Image and Graphics* (2009): 841–846.

Sheriff, R. *Encyclopedic Dictionary of Applied Geophysics*. Fourth edition. Society of Exploration Geophysicists, 2002.

Sinha, S., P. Routh, P. Anno, and J. Castagna. "Spectral decomposition of seismic data with contiuous-wavelet transform." *Geophysics* 70 (2005): P19–P25.

Stanton, A. and M. Sacchi. "Multicomponent seismic data reconstruction using the quaternion Fourier transform and POCS." *SEG Expanded Abstracts* 30 (2011): 1267–1272.

Steeghs, P. and G. Drijkoningen. "Seismic sequence analysis and attribute extraction using quadratic time-frequency representations." *Geophysics* 66 (2001): 1947–1959.

Stockwell, R., L. Mansinha, and R. Lowe. "Localization of the complex spectrum: the S transform." *IEEE Transactions on Signal Processing* 44 (1996): 998–1001.

Taner, M., F. Koehler, and R. Sheriff. "Complex seismic trace analysis." *Geophysics* 44 (1979): 1041–1063.

Taylor, H., S. Banks, and J. McCoy. "Deconvolution with the ℓ_1 norm." *Geophysics* 44 (1979): 39–52.

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

Trickett, S. "$f - x - y$ Cadzow noise suppression." *SEG Expanded Abstracts* 27 (2008): 2586–2590.

Vasudevan, K. and F. Cook. "Time-frequency analysis of deep crustal reflection seismic data using Wigner-Ville distributions." *Canadian Journal of Earth Sciences* 38 (2001): 1027–1035.

Vera Rodriguez, I., D. Bonar, and M. Sacchi. "Microseismic record de-noising using a sparse time-frequency transform." *SEG Expanded Abstracts* 30 (2011): 1693–1698.

Vera Rodriguez, I., D. Bonar, and M. Sacchi. "Microseismic data denoising using a 3C group sparsity constrained time-frequency transform." *Geophysics* 77 (2012): V21–V29.

Vera Rodriguez, I., M. Sacchi, and Y. Gu. "Continuous hypocenter and source mechanism inversion via a Green's function-based matching pursuit algorithm." *The Leading Edge* 29 (2010): 334–337.

Vera Rodriguez, I., M. Sacchi, and Y. Gu. "Toward a near real-time system for event hypocenter and source mechanism recovery via compressive sensing." *SEG Expanded Abstracts* 29 (2010): 2140–2145.

Ville, J. "Theorie et applications de la notion de signal analytique." *Cables et Transmission* 1 (1948): 61–74.

Wang, J. and M. Sacchi. "Noise reduction by structure-and-amplitude-preserving multichannel deconvolution." *CSEG Recorder* 34 (2009): 24–27.

Wang, Y. "Seismic time-frequency spectral decomposition by matching pursuit." *Geophysics* 72 (2007): V13–V20.

Wang, Y. "Multichannel matching pursuit for seismic trace decomposition." *Geophysics* 75 (2010): V61–V66.

Wei, D. and C. Yin. "An optimized locally adaptive non-local means denoising filter for cryo-electron microscopy data." *Journal of Structural Biology* 172 (2010): 211 – 218.

Wigner, E. "On the quantum correction for thermodynamic equilibrium." *Physical Review* 40 (1932): 749–759.

Wu, X. and T. Liu. "Spectral decomposition of seismic data with reassigned smoothed pseudo Wigner-Ville distribution." *Journal of Applied Geophysics* 68 (2009): 386–393.

Yilmaz, Ö. *Seismic Data Analysis: Processing, Inversion, and Interpretation of Seismic Data*. Second edition. Society of Exploration Geophysicists, 2001.

Yuan, M. and Y. Lin. "Model selection and estimation in regression with grouped variables." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (2006): 49–67.

Zoican, S. "Speech de-noising system with non local means algorithm." *International Symposium on Electronics and Telecommunications* 9 (2010): 315 –318.