

# Sparse Signal Recovery in a Transform Domain

by

Evgeniy Lebed

B.Sc., Simon Fraser University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

The Faculty of Graduate Studies

(Mathematics)

The University Of British Columbia

(Vancouver)

August, 2008

© Evgeniy Lebed 2008

# Abstract

The ability to efficiently and sparsely represent seismic data is becoming an increasingly important problem in geophysics. Over the last thirty years many transforms such as wavelets, curvelets, contourlets, surfacelets, shearlets, and many other types of '*x-lets*' have been developed. Such transform were leveraged to resolve this issue of sparse representations. In this work we compare the properties of four of these commonly used transforms, namely the shift-invariant wavelets, complex wavelets, curvelets and surfacelets. We also explore the performance of these transforms for the problem of recovering seismic wavefields from incomplete measurements.

# Table of Contents

<b>Abstract</b> . . . . .	ii
<b>Table of Contents</b> . . . . .	iii
<b>List of Tables</b> . . . . .	v
<b>List of Figures</b> . . . . .	vi
<b>Preface</b> . . . . .	viii
<b>Acknowledgments</b> . . . . .	ix
<b>Dedication</b> . . . . .	x
<b>1 Introduction</b> . . . . .	1
<b>2 The four transforms</b> . . . . .	3
2.1 Motivation . . . . .	3
2.2 Separable transforms . . . . .	4
2.2.1 The shift-invariant wavelet transform . . . . .	4
2.2.2 The dual-tree complex wavelet transform ( CWT) . . . . .	7
2.3 Non-separable transforms . . . . .	8
2.3.1 The curvelet transform . . . . .	8
2.3.2 The surfacelet transform . . . . .	10
<b>3 Sparse seismic signal recovery</b> . . . . .	13
3.1 Motivation . . . . .	13
3.2 Model . . . . .	14
3.3 Compressive sampling . . . . .	15
3.4 $\ell_1$ minimization algorithm . . . . .	17

*Table of Contents*

---

<b>4</b>	<b>Results</b>	20
4.1	Recovering from a set of missing traces	20
4.2	Recovering from a set of missing angular frequencies	22
4.2.1	A different sampling scheme	26
4.2.2	The importance of random sampling	28
4.2.3	Direct Fourier spectra interpolation	30
4.3	Compressibility	33
4.3.1	Nonlinear approximation	33
4.3.2	The Pareto curve	36
4.4	Mutual coherence	37
4.4.1	Coherence approximation	38
4.4.2	Data-dependent coherence approximation	39
<b>5</b>	<b>Conclusions and discussion</b>	42
 <b>Appendices</b>		
<b>A</b>	<b>Wavelet concepts</b>	44
A.1	Frames	44
	<b>Bibliography</b>	45

# List of Tables

4.1	Physical restriction performance summary . . . . .	22
4.2	Frequency restriction performance summary. . . . .	26
4.3	Coherence approximations . . . . .	39
4.4	Data-dependent coherence approximations using 10000 most <i>significant</i> transforms' coefficients . . . . .	40
4.5	Data-dependent coherence approximations using 10000 most <i>insignificant</i> transforms' coefficients . . . . .	40
5.1	Recovery performance summary . . . . .	42

# List of Figures

2.1	Examples of seismic data sets. (a) real data (b) synthetic data.	3
2.2	CWT filter bank (adapted from Selesnick et al. (2005)). Left: two levels of the analysis filter bank. Right: synthesis filter bank. . . . .	7
2.3	Frequency domain partitioning of the curvelet transform (adapted from Candès et al. (2006)). One curvelet is strictly localized to one angular wedge (shaded) in the frequency domain. . .	9
2.4	(Adapted from Lu and Do (2007)) Left: first two level of the analysis filter bank of the surfacelet transform. Subsequent levels of decomposition are achieved by recursively inserting the point $a_n$ into $a_{n+1}$ . Right: synthesis filter bank of the surfacelet transform . . . . .	11
3.1	A typical land seismic survey setup (adapted from <a href="http://www.iongeo.com">www.iongeo.com</a> ). The survey consists of Vibroseis trucks and an array of geophones spread around the survey area. There are physical restrictions on where sensors can be placed: e.g., one cannot put sensors on the roads or the lake. . . . .	13
3.2	Example of the soft-thresholding function $T_\lambda(x)$ with a threshold level $\lambda = 3$ . . . . .	18
4.1	Seismic data and missing traces. (a) Model - synthetic seismic wavefield. (b) Data - same seismic wavefield as in (a) but with 50% of the traces removed . . . . .	20
4.2	Recovered seismic wavefield via (a) curvelets; (b) surfacelets; (c) shift-invariant wavelets; (d) complex wavelets . . . . .	21
4.3	Fourier spectrum and missing frequencies. (a) The model - frequency spectrum of the synthetic wavefield in Figure 4.1(a). (b) The data - 50% of the temporal frequencies are missing. .	23
4.4	The data from Figure 4.3(b) in the physical domain. Aliasing artifact are present throughout the entire image. . . . .	23

*List of Figures*

---

4.5	Recovered seismic wavefield via (a) curvelets; (b) surfacelets; (c) shift-invariant wavelets; (d) complex wavelets. . . . .	24
4.6	Fourier spectra of the recovered signals via (a) curvelets; (b) surfacelets; (c) shift-invariant wavelets; (d) complex wavelets. . . . .	25
4.7	Frequency-weighted sampling scheme. (a) Averaged frequency distribution with 20% of the frequencies used to sample the signal (red lines); (b) recovered signal via shift-invariant wavelets with the frequency samples taken from (a). . . . .	27
4.8	A 1D signal and it's Fourier spectrum . . . . .	28
4.9	Randomly subsampled 1D signal and it's Fourier spectrum . . . . .	29
4.10	Uniformly subsampled 1D signal and it's Fourier spectrum . . . . .	30
4.11	Uniform sub-sampling of angular frequencies. (a) Three-fold uniformly subsampled frequency spectra of the synthetic wavefield; (b) Attempted recovery result by shift invariant wavelets with a uniform subsampling scheme from (a); (c) interpolated spectrum. . . . .	31
4.12	Subsampled signal. (a) Subsampled Fourier spectrum of a seismic signal; (b) same signal as in (a), but in the physical domain. . . . .	32
4.13	Directly interpolating the Fourier spectra. (a) Interpolated spectra from Figure 4.12(a); (b) interpolated signal in the physical domain. . . . .	32
4.14	Nonlinear approximation. (a): Randomly sampled transform coefficients plotted in magnitude-descending order. (b) MSE of the reconstruction from an incomplete set of transform coefficients . . . . .	35
4.15	Pareto curves for the different transforms. The lined in the plot represent Pareto curves for compressibility of seismic data by: Fourier (yellow), curvelets (blue), surfacelets (red), complex wavelets (green) and shift-invariant wavelets (black) . . . . .	37
4.16	Single atoms of the four transforms in the physical and frequency domains. (a): curvelet (top right), surfacelet (top right), shift-invariant wavelet (bottom left), complex wavelet (bottom right) in the physical domain. (b): Fourier transform of the atoms in the same order. . . . .	41

# Preface

This thesis was prepared with Madagascar, a reproducible research software package available at [rsf.sf.net](http://rsf.sf.net). Many results shown in this thesis are directly linked to the codes that generated them. This reproducibility allows direct access to the experiments performed in this work allowing transparency and dissemination of knowledge for the Seismic Laboratory for Imaging and Modeling (SLIM), its sponsors, and the entire research community as a whole.

The programs required to reproduce the reported results are Madagascar programs written in C/C++, Matlab, or Python. The numerical algorithms and applications are mainly written in Matlab with a few exceptions written in Python as part of SLIMpy ([slim.eos.ubc.ca/SLIMpy](http://slim.eos.ubc.ca/SLIMpy)). The external Matlab toolboxes required to reproduce some of the results in this thesis are: Curvelab, Surfbox, RWT, and Complex wavelet toolbox.

# Acknowledgments

First of all I would like to thank my supervisors, Ozgur Yilmaz and Felix Herrmann for the never-ending support and advice which they have provided me with throughout these last two years. Without their help I would not have been able to learn half the things which I have learned. Their patience with me is much appreciated. I would like to express my gratitude to Gilles Hennenfent, Ewout van den Berg, Cody Brown and the rest of the SLIM team for their help with resolving programming and other scientific issues. Just as well, I would like to thank Yue Lu for helping me unravel the thread of puzzles of the surfacelet transform.

And lastly I would like to express my deepest thanks to all of my family and my lovely fiancé Kristina. There is no question that without your guys' help I would not have been able to get to where I am today.

**To my grandfather Dr. Igor A. Karnovsky.**

# Chapter 1

## Introduction

There are many different transforms that can be used to obtain a sparse representations of signals (e.g., seismic images, natural images, and audio signals). Numerous ‘sparsifying’ transforms have been invented and extensively studied in the seismic processing literature (see, e.g., Holschneider et al. (1989), Donoho and Johnstone (1994), Candès et al. (2006)). In this work we compare some of the important properties that these transforms possess. A sparsifying transform is an invertible linear operator, mapping a signal (e.g., an image) to a sparse vector, i.e., a vector with few large components. Different types of images have sparse representations under different transforms. For example, piecewise constant images can be sparsely represented by spatial finite differences. Natural, real-life images are known to have a sparse representation in the discrete cosine transform (DCT) and wavelet transform domains (Dagher et al. (2003)). It has recently been observed that the curvelet transform yields a sparse representation of seismic data (the curvelet dictionary consists of atoms that locally behave like waves, and seismic data, consisting of solutions to the wave equation, has a good sparse representation in curvelets; see, e.g., Candès and Demanet (2005)). Of course, in any practical setting, one does not expect to have a sparse representation with a non-adaptive transform. Therefore, strictly speaking, when we say “sparse” we mean that the magnitude-sorted transform coefficients decay rapidly. In this case we also say the signal is “compressible.” More precisely, the coefficient vectors live in a bounded ball in weak  $\ell^q$  with  $0 < q < 1$ .

Surfacelets also try to achieve the same result, but with a much different construction. The complex-wavelet transform (Selesnick et al. (2005)) is of interest because it has a fixed and low redundancy factor and has the ability to give signals a multidirectional representation. Classical shift-invariant wavelets are of interest because computationally they provide a very fast multiscale representation of a signal although at a price of a high redundancy factor.

Transform design is only one of the challenges towards obtaining sparse representations for signals. The second obstacle we need to tackle is find-

ing the sparse coefficients for a given signal. This is easy if the transform corresponds to an orthonormal basis expansion as there is a unique coefficient vector in this case. However, many transforms of interest correspond to redundant expansions and therefore a given signal can be represented by infinitely many coefficient vectors. To find a sparse vector among these is highly non-trivial (Candès et al. (2006)). Several properties of the transform vectors, e.g., their coherence determine whether this task can be performed in a computationally tractable manner. Recently, our understanding in this direction has improved by the explosion in compressed sensing literature.

### Outline of this thesis

- First, we compare the construction and properties of four different transforms: the shift-invariant wavelet, the complex wavelet, the curvelet and the surfacelet transform. These all attempt to achieve a common goal - a sparse representation of all signals in a given class (e.g., all seismic datasets or all natural images). These transforms have different localization properties in the space and frequency domains, and as a result they all have unique and interesting properties.
- Second, we discuss how these transforms are tied into the theory of compressed sensing (Tsaig and Donoho (2006)).
- And last, we investigate how these transforms perform in recovering signals from incomplete measurements (see e.g., Lustig et al. (2007)). In other words, we try to answer the question of which of the four transforms is most appropriate to use when recovering a signal from different types of measurements (either from the physical or the frequency domains).

Successful recovery of a signal (modeled as a vector in  $\mathbb{R}^N$  with potentially very large  $N$ ) from incomplete measurements depends mainly on three factors: (i) how *sampling* is done (e.g., uniform or non-uniform, in the spatial domain or in the frequency domain), (ii) how *compressible* the signal is with respect to some prescribed transform (e.g., Fourier, wavelet, curvelet, etc.) and (iii) how *mutually incoherent* the measurement and sparsity domains are. In recent years the interplay between these three key factors and their effect on successful sparse recovery is being investigated extensively. Recent results from compressive sampling give us a better understanding

## *Chapter 1. Introduction*

---

in this direction. However, these results do not necessarily apply to cases where we deal with structured and heavily redundant transforms such as shift-invariant wavelet, complex wavelet, curvelet or surfacelet transforms.

## Chapter 2

# The four transforms

In this chapter we investigate the construction and properties of four different transforms. We start with two separable transforms - shift-invariant and complex wavelets, and conclude with two non-separable transforms - curvelets and surfacelets.

### 2.1 Motivation

As mentioned in the previous chapter, many different transforms were developed for a wide variety of applications. In this work we focus on one particular type of data - seismic wave fields. Some typical seismic wave fields are shown in Figure 2.1. Figure 2.1(a) shows a real seismic wavefield while Figure 2.1(b) shows a synthetic seismic wavefield.

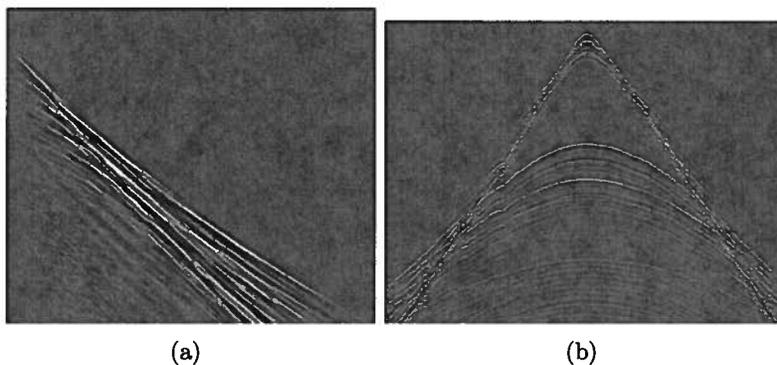


Figure 2.1: Examples of seismic data sets. (a) real data (b) synthetic data.

One can see that these datasets have a distinct structure. The wavefronts consist of hyperbolic and linear events, in some places with overlapping cusps. Furthermore, there are many singularities present in the data.

An effective transform for this type of data should have the following useful properties:

**Directional selectivity.** Typical seismic data (shown in Figure 2.1) consists of wavefronts with a clear orientation. Thus the basis functions of our transform should be able to align to these events.

**Sparsity.** The transform should be able to capture most energy of the signal with only a few coefficients.

**Small redundancy.** Real seismic datasets can be up to several gigabytes large, and even a moderate redundancy factor could significantly slow down the processing flow.

**Incoherence.** As suggested by the theory of compressive sampling, the basis functions of the transform should be as incoherent as possible with the domain from which our measurements are acquired (Candès and Wakin (2007)).

## 2.2 Separable transforms

A separable transform is constructed with separable products of a scaling function  $\phi$  and a wavelet  $\psi$ . In this section we explore two transforms that are separable: shift-invariant wavelets and complex wavelets.

### 2.2.1 The shift-invariant wavelet transform

Numerous variations of the wavelet transforms have been extensively studied and have appeared in literature as early as 1994 (Donoho and Johnstone (1994)), even though a mature mathematical theory of wavelets, as well as practical constructions of orthonormal wavelet bases was developed a decade later. Their ability to efficiently represent data that has multiscale structure has made the wavelet transform (WT) a powerful tool for researchers to use in all areas of signal processing. Perhaps the most widely implemented variation of the WT is the decimated bi-orthogonal wavelet transform (Masud and McCanny (1999)), which is utilized in the JPEG2000 compression algorithm (Dagher et al. (2003)). For other applications, such as edge detection and texture analysis, researchers continue to employ wavelet transforms with an arbitrary number of scales. This essentially means that the signal is modeled as a continuous function in some infinite dimensional function space, and one uses a wavelet basis or frame for this space. Because of computational constraints, this type of transform would only be practical for relatively small datasets. Usually these applications do not involve a reconstruction step.

In other applications, where a reconstruction from modified coefficients is required, researchers choose to use the undecimated wavelet transform (UDWT), which consists of a filter-bank construction but eliminates the decimation step in the orthogonal wavelet transform. These types of algorithms are very fast and efficient in terms of computational complexity. Starck et al. (2007) have shown that utilizing an undecimated wavelet transform for image denoising can lead to an improvement of up-to 2.5dB when compared to denoising via a decimated wavelet transform. The coefficients of the UDWT are computed using the same filter-bank as in the standard decimated bi-orthogonal WT. This construction leads to three different bands of coefficients (vertical, horizontal and diagonal) for every scale.

### One-dimensional discrete wavelet transform

The undecimated discrete wavelet transform UDWT (also called the shift-invariant wavelet transform) was independently discovered by several authors, see for example Shensa (1992). Next, we outline how the coefficients associated with UDWT are computed. We adopt the usual notation and denote the scaling and wavelet functions by  $\phi$  and  $\psi$ , respectively:

$$\phi(x/2) = \sum_k h[k]\phi(x - k), \quad (2.1)$$

$$\psi(x/2) = \sum_k g[k]\psi(x - k). \quad (2.2)$$

Here  $k \in \mathbb{Z}$ ,  $x \in \mathbb{R}$ , and  $h[k]$  and  $g[k]$  are the analysis filters. Using the filter bank  $(h, g)$ , the undecimated wavelet transform of a 1-D signal returns a set of coefficients  $W = \{w_1, \dots, w_J, c_J\}$  where  $w_j$  are the wavelet coefficients at scale  $j$  and  $c_J$  are the coefficients at the coarsest scale. One can obtain the coefficients at the next finest scale from the ones at the previous scale using Mallat's à trous algorithm (Holschneider et al. (1989), Shensa (1992), Starck et al. (2007)):

$$c_{j+1}[l] = (\bar{h}^{(j)} * c_j)[l] = \sum_k h[k]c_k[l + 2^j k], \quad (2.3)$$

$$w_{j+1}[l] = (\bar{g}^{(j)} * c_j)[l] = \sum_k g[k]c_k[l + 2^j k], \quad (2.4)$$

with

$$h^{(j)} = \begin{cases} h[l] & \text{if } l/2^j \in \mathbb{Z}, \\ 0 & \text{otherwise,} \end{cases} \quad (2.5)$$

where  $\bar{h}[n] = h[-n]$  and  $\bar{g}[n] = g[-n]$ . The reconstruction is obtained by

$$c_j[l] = \frac{1}{2}[\tilde{h}^{(j)} * c_{j+1}[l] + \tilde{h}^{(j)} * w_{j+1}[l]]. \quad (2.6)$$

In the above equation, the filters  $\tilde{h}$  and  $\tilde{g}$  refer to the synthesis of  $h$  and  $g$ , respectively. The filters are designed in such a way that the filter bank  $(h, g, \tilde{h}, \tilde{g})$  satisfies the perfect reconstruction condition:

$$H(z^{-1})\tilde{H}(z) + G(z^{-1})\tilde{G}(z) = 1, \quad (2.7)$$

where  $H, G, \tilde{H}$  and  $\tilde{G}$  are the z-transforms of  $h, g, \tilde{h}$  and  $\tilde{g}$  respectively.

### Two dimensional discrete wavelet transform

The 2-D version of UDWT is obtained by extending the à trous algorithm as follows:

$$c_{j+1}[k, l] = (\bar{h}^{(j)}\bar{h}^{(j)} * c_j)[k, l], \quad (2.8)$$

$$w_{j+1}^1[k, l] = (\bar{g}^{(j)}\bar{h}^{(j)} * c_j)[k, l], \quad (2.9)$$

$$w_{j+1}^2[k, l] = (\bar{h}^{(j)}\bar{g}^{(j)} * c_j)[k, l], \quad (2.10)$$

$$w_{j+1}^3[k, l] = (\bar{g}^{(j)}\bar{g}^{(j)} * c_j)[k, l], \quad (2.11)$$

where  $hg * c$  is the convolution of  $c$  by the separable filter  $hg$ , i.e., convolution first along the columns by  $h$  and then convolution along the rows by  $g$  (Starck et al. (2007)). The algorithm produces three wavelet images at each scale,  $w^1, w^2, w^3$  with each wavelet image being the same size of the original image. Consequently the redundancy factor of UDWT depends on the number of scales we choose to have in the decomposition, and is computed by is easy to compute; it is  $3(L - 1) + 1$  where  $L$  is the number of scales.

The key properties of the UDWT are:

1. The UDWT is shift-invariant, i.e., a small shift in the signal does not perturb UDWT's coefficient oscillation pattern around singularities.
2. In 2-D the UDWT has a redundancy factor of  $3(L - 1) + 1$  where  $L$  is the number of scales in the decomposition.

It is important to note that although computational complexity of UDWT is  $O(LN)$ , the redundancy factor of UDWT is exceedingly high for practical applications. If we are dealing with a moderately-sized dataset that we wish to analyze, perhaps with 6 scales, this already amounts to a redundancy factor of  $3(6-1)+1 = 16$ , which is unpractical for a lot of applications.

### 2.2.2 The dual-tree complex wavelet transform (CWT)

The complex dual-tree framework proposed by Selesnick et al. (2005) extends the work of Kingsbury (1999) by utilizing a filter bank implementation in conjunction with two real discrete wavelet transforms (DWTs). These DWTs use different sets of filters and satisfy the perfect reconstruction conditions, i.e., the analysis operation followed by synthesis operation would produce the original signal. One of the DWTs gives the real part of the CWT's coefficients while the other DWT gives the imaginary part of the coefficients.

If we let  $h_0(n)$  and  $h_1(n)$  be the lowpass/highpass pair for one of the filterbanks and  $g_0(n)$  and  $g_1(n)$  be the lowpass/highpass pair for the other filter-bank than we can display the analysis/synthesis for the CWT in a schematic format, as shown in Figure 2.2.

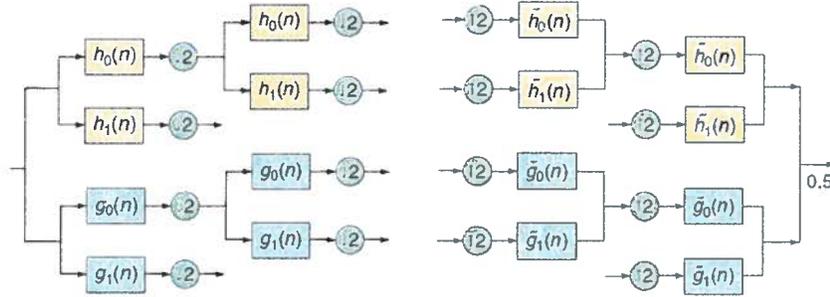


Figure 2.2: CWT filter bank (adapted from Selesnick et al. (2005)). Left: two levels of the analysis filter bank. Right: synthesis filter bank.

These filters are designed in such a way so that the wavelet  $\psi(t) := \psi_h(t) + i\psi_g(t)$ , where  $\psi_g(t)$  and  $\psi_h(t)$  are real wavelet functions associated with the two real wavelet transforms, are approximately analytic (see Lilly and Olhede (2008)). In this context, approximately analytic means that  $\psi(t)$  can be locally approximated by a convergent power series.

The implementation of CWT is simple: the filter pair  $h_0(n), h_1(n)$  is used to implement one separable 2-D wavelet transform and the filter pair  $g_0(n), g_1(n)$  is used to implement the other 2-D wavelet transform. Applying both transforms to the same 2-D image gives a total of 6 sub-bands: 2 HL, 2 LH and 2 HH sub-bands.

The key properties of the CWT are outlined below;

1. The CWT is nearly shift invariant, i.e., a small shift in the signal does

not alter the coefficients' oscillation pattern around singularities.

2. The CWT is a multiscale transform with six directionally-selective subbands in 2-D.
3. A complex wavelet is strictly localized in the spatial domain.
4. Regardless of the number of scales used in the decomposition, the redundancy factor of CWT is always  $2^n$  for  $n$  dimensions.

## 2.3 Non-separable transforms

The shift-invariant and complex wavelet transforms both provide multiscale representations of signals. In addition, the CWT provides a limited number of directions in the representation. Next we explore two non-separable transforms which provide more flexibility in choosing the number of directions at any given scale.

### 2.3.1 The curvelet transform

From a conceptual point of view a curvelet (Candès and Donoho (2004)) is a scaled, shifted, and dilated copy of a basis element  $\varphi_a$  with width approximately equal to  $a$  and length approximately equal to  $\sqrt{a}$ . A family of curvelets at scale  $a$ , location  $b$ , and orientation  $\theta$  are defined by

$$\varphi_{a,b,\theta}(x) = \varphi_a(R_\theta(x - b)), \quad (2.12)$$

where  $R_\theta$  is a rotation by  $\theta$  radians. In the continuous version of the curvelet transform, the parameter  $a \in (0, 1)$ , indicates the scale,  $\theta$  is the orientation, and  $b \in \mathbf{R}^2$  is the spatial location. In the discrete version of the curvelet transform the parameters  $a$  and  $\theta$  are given by

$$a_j = 2^{-j}, \text{ and } \theta_{j,l} = 2\pi l \cdot 2^{-\lfloor j/2 \rfloor}. \quad (2.13)$$

At any given scale a curvelet is defined by applying translations and rotations to a “mother curvelet”. In the frequency domain it is defined by

$$\hat{\varphi}_{j,0,0} = 2^{-3j/4} W(2^{-j}|\xi|) V(2^{\lfloor j/2 \rfloor} \theta), \quad (2.14)$$

where  $W$  and  $V$  are smooth windows with compact support. Since  $W$  and  $V$  are compactly supported, one can see directly from Equation 2.14 that a curvelet will be strictly localized in the frequency domain. It turns out that

in the spatial domain curvelets are also well localized with fast decay. Once can see this idea demonstrated in Figure 4.16.

The 2-D discrete curvelet transform, proposed by Candès and Donoho (2004), takes a 2-D signal of the form  $f(n_1, n_2)$ ,  $0 \leq n_1, n_2 < n$  as input and outputs a collection of coefficients that are obtained by computing inner products of the signal with curvelet window functions at different scales and directions (Candès et al. (2006)). The curvelet coefficients  $c(j, l, k)$  at scale  $j$ , orientation  $l$  and spatial location  $k$  are defined by

$$c(j, l, k) = \sum_{n_1, n_2} \overline{\varphi_{j,l,k}(n_1, n_2)} f(n_1, n_2), \quad (2.15)$$

where  $\overline{\varphi_{j,l,k}(n_1, n_2)}$  is the complex conjugate of the curvelet function  $\varphi_{j,l,k}(n_1, n_2)$ . The outline of the curvelet transform algorithm is as follows: A 2-D fast Fourier transform is applied to  $f(n_1, n_2)$  to get  $\hat{f}(\omega_1, \omega_2)$  with  $-n/2 \leq \omega_1, \omega_2 < n/2$ . Next the Fourier samples of the input image are multiplied with the curvelet window functions  $\tilde{U}_{j,l}(\omega_1, \omega_2)$  at different scales and directions to obtain the product  $s_{j,l} = \tilde{U}_{j,l}(\omega_1, \omega_2) \hat{f}(\omega_1, \omega_2)$ . This product is wrapped around the origin to produce  $W(s_{j,l})(\omega_1, \omega_2)$ . No wrapping is required at coarsest and finest scales. Lastly, a windowed 2-D inverse fast Fourier transform is applied to  $s_{j,l}$ , and the curvelet coefficients  $c(j, l, k)$  are collected.

The resulting collection of curvelet coefficients  $\varphi_{a,\theta,b}$  is a tight Parseval frame (see, e.g., Mallat (1999)) for a Hilbert space:

$$f = \sum_{a,\theta,b} \langle f, \varphi_{a,\theta,b} \rangle \varphi_{a,\theta,b}, \quad \|f\|_2^2 = \sum_{a,\theta,b} |\langle f, \varphi_{a,\theta,b} \rangle|^2. \quad (2.16)$$

See Appendix A.1 for the definition of frames for Hilbert spaces. Some of the important properties of curvelets are:

1. The curvelet transform provides a tight-frame expansion for any function  $f \in L^2(\mathbb{R}^2)$  as a series of curvelets.
2. A single curvelet is strictly localized to one angular wedge in the frequency domain and has fast decay in the spatial domain. The effective support of  $\varphi_{j,l,k}(x)$  obeys a parabolic scaling relation with  $width \propto length^2$ . At scale  $j$  a curvelet is a ‘fat needle’ of length  $2^{-j/2}$  and width  $2^{-j}$ .

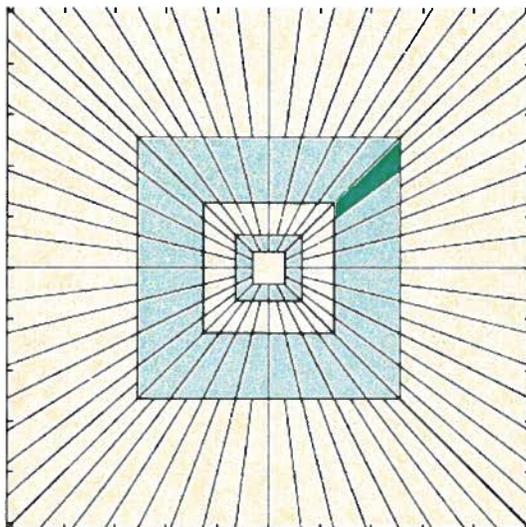


Figure 2.3: Frequency domain partitioning of the curvelet transform (adapted from Candès et al. (2006)). One curvelet is strictly localized to one angular wedge (shaded) in the frequency domain.

3. A curvelet is smooth along its major axis and is oscillatory along its minor axis (see Figure 4.16(a)).
4. The curvelet transform, as implemented in Candès et al. (2006), is approximately 8 times redundant in 2-D and 24 times redundant in 3D.

### 2.3.2 The surfacelet transform

Curvelets provided us with a multiscale and a multidirectional signal decomposition. We now explore *surfacelet transforms*, which are multiscale and multidirectional transforms that is less redundant than curvelets, and have a slightly different frequency partitioning.

The surfacelet transforms, proposed by Lu and Do (2007), conceptually begins from a different construction than the curvelet transform. The surfacelet construction starts by extending the directional filter bank (DFB) that was proposed by Bamberger and Smith (1992) to an arbitrary number ( $N \geq 2$ ) of dimensions. The DFB partitions the frequency plane into  $2^l, l \in \mathbb{N}$  triangular wedges of equal area by radial lines passing through the origin of the  $(\omega_1, \omega_2)$  plane. In the case of a general  $N$ , the  $N$ -dimensional

filter bank (NDFB) partitions the frequency spectrum into rectangular pyramids radiating from the origin.

We consider a 3D example: to achieve the desired frequency partitioning, the NDFB (now  $N = 3$ ) partitions the frequency cube into three hourglass-shaped subbands that are aligned with the  $\omega_1, \omega_2$  and  $\omega_3$  axes, respectively. Next, two iteratively resampled checkerboard filterbanks (IRC) (Lu and Do (2007)), that act along the  $(n_1, n_2)$  and  $(n_1, n_3)$  planes, are applied to each of the hourglass-shaped frequency subbands. The resulting output is  $3 \cdot 2^l$  directional subbands (Lu and Do (2006)).

Next, we describe the construction of surfacelets, as give by (Lu and Do (2007): the Fourier transform of the signal gets inputed into first level of the multiscale pyramid where a highpass filter  $H(\omega)$  and a lowpass filter  $L(\omega)$  operate on the signal. This is similar to the CWT construction, where lowpass and highpass filters act on the signal to achieve a multiscale representation (see Figure 2.2). The output of  $H(\omega)$  is inputed into the NDFB, and this returns the fine scale directional approximation to the signal. At the lowpass branch of the pyramid the signal is upsampled by a factor of two followed by an anti-aliasing filter  $S(\omega)$ . Finally, the signal is downsampled by a factor of 3. The output of this branch is a coarse approximation to the signal, which gets inputed into the next level of the multiscale pyramid. At the subsequent lowpass branches  $S(\omega)$  is no longer applied and the signal only gets downsampled by a factor of 2. Figure 2.4 describes the above process.

Surfacelets were originally designed to get similar frequency partitioning as curvelets, but the two transforms aim to achieve this result with two very different approaches. The curvelet implementation starts with defining a “mother” curvelet in the Fourier domain. Moreover, its scaled and sheared copies form a partition of the frequency plane. As a result, the curvelet coefficients are obtained by multiplying the Fourier samples of the input signal with curvelet window functions at different scales and directions, followed by spatial downsampling. The downsampling for curvelets is done in an alias-free fashion, which requires the curvelet functions to be strictly localized in the frequency domain. Since lower redundancy requires a higher subsampling ratio, this poses a trade-off between redundancy and spatial localization of curvelets, and subsequently implies a narrower transition band in the frequency supports.

It is important to note that aliasing is allowed in the  $N$ -dimensional filter bank that is implemented in the surfacelet transform. At the end, the aliasing is cancelled out by carefully designed filters. As a consequence,

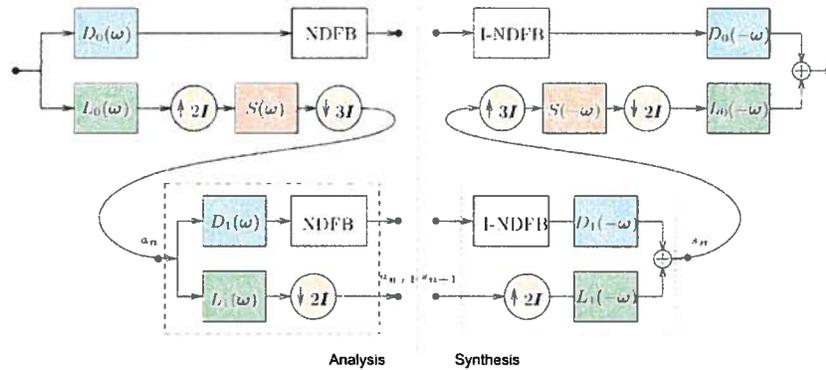


Figure 2.4: (Adapted from Lu and Do (2007)) Left: first two level of the analysis filter bank of the surfacelet transform. Subsequent levels of decomposition are achieved by recursively inserting the point  $a_n$  into  $a_{n+1}$ . Right: synthesis filter bank of the surfacelet transform

surfacelets are significantly less redundant than curvelets (a factor of 5 for surfacelets compared to a factor of 8 for curvelets). The surfacelet transform can use filters with fast spatial decay without losing redundancy efficiency, since the filters do not need to be strictly bandlimited (Lu and Do (2007)). Once again, we list the essential properties of the surfacelet transforms;

1. Just like the curvelet transform, the 2-D surfacelet transform provides a redundant, multiscale and multidirectional decomposition for functions  $f$ . The transform is tight frame (see Appendix A.1).
2. Unlike curvelets, which are strictly localized to one angular wedge, a surfacelet is spread across an entire scale in the frequency domain. In the spatial domain surfacelets also obey a parabolic scaling relation of  $width \propto length^2$ , with most of the oscillations occurring along the minor axis (see Figure 4.16).
3. In the 2-D case the surfacelet transform, as implemented in Lu and Do (2007), is approximately 5 times redundant.
4. Unlike the curvelet directional selectivity, where the number of wedges doubles at every *other* scale, in the surfacelet transform the number of wedges doubles at every scale. The finest and second-finest scales have the same number of wedges.

## Chapter 3

# Sparse seismic signal recovery

### 3.1 Motivation

A seismic survey is an experiment to estimate properties of the Earth's subsurface from reflected waves. The experiment usually consists of a source of energy, such as an air gun, or a vibrator (usually called a Vibroseis) and an array of receiver microphones that measure the reflected waves from the subsurface. By recording these acoustic pressure changes, it is possible to estimate the depth of the singularity that caused the reflection.



Figure 3.1: A typical land seismic survey setup (adapted from [www.iongeo.com](http://www.iongeo.com)). The survey consists of Vibroseis trucks and an array of geophones spread around the survey area. There are physical restrictions on where sensors can be placed: e.g., one cannot put sensors on the roads or the lake.

As illustrated in Figure 3.1, seismic data is often irregularly sampled

along the spatial coordinates due to practical constraints. The irregular sampling is usually due to the fact that at certain locations along the survey area a receiver physically cannot be placed. In marine surveys the irregular sampling occurs as a result of dead/severely contaminated traces or due to cable feathering. A lot of the multi-trace processing algorithms do not handle irregularly sampled data (data irregularities often result in post-processing image artifacts), and thus the data must be interpolated onto a regular grid. There exist relatively simple methods like linear interpolation of neighboring traces to handle this problem. Unfortunately this method does not account for wavefronts in the data, and the results of linear interpolations are not always satisfactory. In practice, the methods that are used to interpolate seismic data are divided into three main groups: offset continuations (Stovas and Fomel (1993)), shot continuation (Schwab (1993)) and transform based methods (Sacchi et al. (1998)). In this work we build on the transform-based approach for signal recovery, as was proposed by Herrmann and Hennenfent (2008) and Hennenfent and Herrmann (2008b).

Frequency domain methods for interpolation also require the data to be mapped onto a regular grid. This is often the case when the temporal frequencies of a signal are sampled in an irregular manner (Lin et al. (2008)). This is an often-occurring issue in seismic wavefield modeling in the frequency domain; when solutions to the time-harmonic Helmholtz equation are available for all frequencies, modeling in the frequency domain can be transformed back to modeling in the physical domain with time-stepping methods. However the solutions of the Helmholtz equation for each temporal frequency are expensive to compute. On the other hand, if we are to invert a seismic wavefield from only a partial set of frequencies, aliasing artifacts will be visible in the resulting modeled wavefield. In this work we extend the ideas presented by Herrmann and Hennenfent (2008) and Hennenfent and Herrmann (2008b) to interpolate the subsampled frequencies of a seismic signal to obtain an alias free result.

## 3.2 Model

One application of transforms described in the previous chapter is recovering seismic wavefields from (i) subset of traces and (ii) a subset of angular frequencies. To describe the problem of signal recovery, we assume the forward model

$$\mathbf{y} = \mathbf{R}\mathbf{M}\mathbf{f} + \mathbf{n}, \tag{3.1}$$

where  $\mathbf{y}$  is a vector, each component of which is a linear measurement of the full data vector  $\mathbf{f}$ . The matrix  $\mathbf{M}$  refers to the domain from which our measurements are acquired. In problem (i)  $\mathbf{M}$  is the identity or the Dirac measurement basis, so  $\mathbf{M} := \mathbf{I}$  and in problem (ii)  $\mathbf{M}$  is the Fourier measurement basis, so  $\mathbf{M} := \mathbf{F}$ . In this text we denote by  $\mathbf{I}_d$  the  $d \times d$  identity matrix, and by  $\mathbf{F}_d$  the  $d \times d$  discrete Fourier transform matrix. We drop the subscript when it is not necessary to specify the dimensions. The matrix  $\mathbf{R}$  is a restriction operator, i.e., it extracts those rows from  $\mathbf{M}$  that represent the samples that are actually acquired. In real life problems the measurement basis and the restriction operator are determined by the acquisition geometry.

In what follows, we wish to use some transform which “sparsifies” our signal. If we can find such a transform, we can then exploit this fact by approximating our signal by a linear combination of a few “x-let” atoms. More precisely, let  $\mathbf{S}$  be the analysis operator of the “x-let” transform (whose atoms form a tight frame), and let  $\mathbf{S}^H$ , the conjugate transpose of  $\mathbf{S}$  be the synthesis operator. Given a set of incomplete measurement  $\mathbf{y}$  of our signal  $\mathbf{f}$ , we wish to obtain an approximation  $\tilde{\mathbf{f}}$  such that (i)  $\tilde{\mathbf{f}} = \mathbf{S}^H \tilde{\mathbf{x}}$  where  $\tilde{\mathbf{x}}$  is sparse, (ii)  $\|\mathbf{f} - \tilde{\mathbf{f}}\|_2$  is small. In certain cases, such an  $\tilde{\mathbf{x}}$  can be computed by solving the optimization problem

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{RMS}^H \mathbf{x}\|_2 \leq \epsilon. \quad (3.2)$$

In what follows, the matrix  $\mathbf{S}^H$  refers to the synthesis matrix of one of the four transforms domains described in the previous chapter. This problem is setup in such a way that the solution has small  $\ell_1$  norm (which, at least under certain assumptions, ensures that it is sparse), while at the same time it fits the data up to a tolerance  $\epsilon$  ( $\epsilon$  is related to the amount of noise present in our data, and in the noise-free case  $\epsilon = 0$ ). Herrmann and Hennenfent (2008) have proposed to solve the problem given in Equation 3.2 when the underlying transform is the curvelet transform to recover a seismic signal from a subset of traces (Problem i). In this case  $\mathbf{S}^H$  in Equations 3.2 is the curvelet matrix, and  $\mathbf{M}$  is the Dirac measurement basis. They use the principle of alignment, i.e., there is a large correlation between curvelets and wavefronts that locally have the same direction and frequency content, to explain heuristically why curvelets achieve good compressibility (one of the necessary conditions for signal recovery) of seismic data.

### 3.3 Compressive sampling

In recent years, the theory of compressive sampling, pioneered by papers like Candès and Donoho (2004), Candès et al. (2006), and Donoho (2006) to name a few, has appeared in a lot of literature. As stated by Candès and Wakin (2007), data acquisition is usually extremely wasteful; large amounts of data are collected, only to be discarded later on. The acquired data is discarded at the compression stage which is usually necessary for transmission and storing purposes. Compressive sampling works under the philosophy of as if “it were possible to directly acquire just the important information about the object of interest”. By taking only a small number of random projections, one has enough information to reconstruct the signal with an accuracy that is at least as good as the best compressed representation of the object. In this type of data acquisition, analog signals are essentially already in a compressed digital form. In principle, one can obtain “super-resolved” signals from just a few measurements and all that is required is to “decompress” the measured data.

Candès and Romberg (2007) have shown that when our signal  $f$  is sufficiently sparse, one can exactly recover the signal exactly via  $\ell_1$  minimization.

**Theorem 1** (Candès and Romberg (2007)): Fix  $f \in \mathbb{R}^n$  and suppose that the coefficient sequence  $x$  of  $f$  in the basis  $\Psi$  is  $S$ -sparse. Select  $m$  measurements in the domain  $\Phi$  uniformly at random. Then if

$$m \geq C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot \log n, \quad (3.3)$$

for some positive constant  $C$  and mutual coherence  $\mu$ , the solution to Equation 3.2 is exact with overwhelming probability.

The quantity  $\mu$  is known as the *coherence* between the sensing matrix  $\Phi$  and the measurement matrix  $\Psi$ , and is defined by (Candès and Romberg (2007))

$$\mu(\Phi, \Psi) = \sqrt{n} \cdot \max_{1 \leq k, j \leq n} |\langle \varphi_k, \psi_j \rangle|, \quad (3.4)$$

where  $\varphi_k$  is the  $k^{\text{th}}$  row of  $\Phi$  and  $\psi_j$  is the  $j^{\text{th}}$  row of  $\Psi$ . This quantity measures the spread of vectors from the measurement basis in the sparsifying domain.

The “overwhelming probability”, mentioned in the above theorem exceeds  $1 - \delta$  if  $m \geq C \cdot \mu^2(\Phi, \Psi) \cdot S \cdot \log(n/\delta)$  (see Candès and Romberg (2007) for details). The theorem suggests a very specific type of acquisition; a non-adaptive sampling in an incoherent domain, followed by one-norms

solution. Essentially, this would acquire the analog signal in a compressed form. After which, the data is “decompressed” via  $\ell_1$  minimization.

The theory of compressive sampling does not really apply to our case as our “sparsity transforms” correspond to redundant expansions which are very structured and do possess bad coherence properties for any of these theorems to apply. Still, empirically  $\ell_1$  minimization seems to perform well to capture sparse representations in these cases. One of the main goals of this thesis is to speculate on possible empirical explanations for this unexpectedly well performance of compressed sensing-like methods, at least in the case of seismic datasets with the four mentioned transforms.

### 3.4 $\ell_1$ minimization algorithm

The solution of the unconstrained optimization problem given in Equation 3.2 has been extensively studied and to this day many different solvers exist to tackle this problem. Without being exhaustive, a few of them are: *iterative reweighted least-squares*, introduced by Gersztenkorn et al. (1986), *SPGL $_1$* , proposed by Berg and Friedlander (2007), a variation of the *projected gradient* method, proposed by Birgin et al. (2000), *iterative hard-thresholding*, recently introduced by Blumensath and Davies (2008); and *iterative soft-thresholding*, introduced by Daubechies et al. (2004). Herrmann and Hennenfent (2008) used the iterative soft thresholding (IST) to solve the problem given in Equation 3.2. The IST is a very attractive solver to use, especially for large problems, since it only consists of matrix-vector multiplications and thresholding operations.

However, before we can apply this solver directly to the constrained problem given in Equation 3.2, we must reformulate it as a series of unconstrained problems. Following the strategy of Elad et al. (2008), Problem 3.2 can be rewritten as an unconstrained problem using Lagrange multipliers  $\lambda$  as

$$\tilde{\mathbf{x}}_\lambda = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{RMS}^H \mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (3.5)$$

where  $\lambda$  is a parameter that controls the tradeoff between the sparsity of the solution (minimization of  $\ell_1$  norm) and the  $\ell_2$  data misfit. Now equation 3.5 is in such a form that we can directly apply the IST solver. To simplify the notation, we define the matrix  $\mathbf{A}$  to be  $\mathbf{A} := \mathbf{RMS}^H$ .

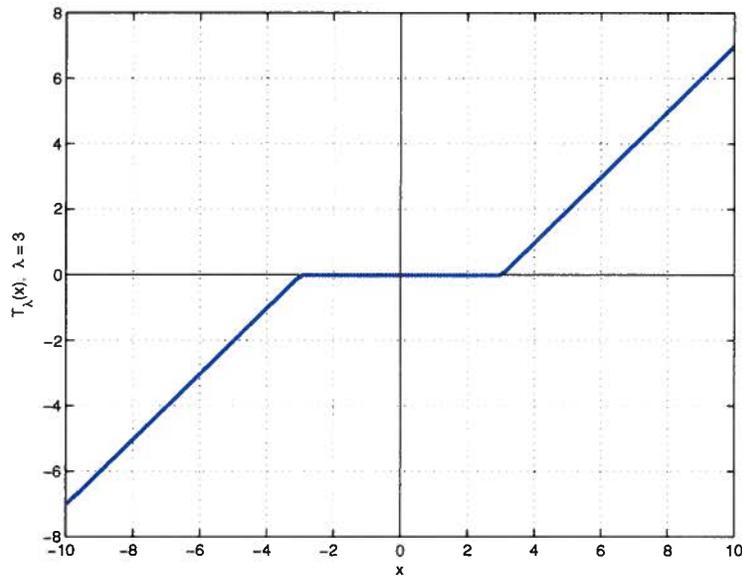
Daubechies et al. (2004) have shown that iterative update of  $\mathbf{x}$  by

$$\mathbf{x} \rightarrow T_\lambda(\mathbf{x} + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x})) \quad (3.6)$$

converges to the solution of the problem given in Equation 3.5, for a particular value of  $\lambda$ , as the number of iterations goes to infinity. In Equation 3.6,  $T_\lambda$  is a soft-thresholding function, defined by

$$T_\lambda(x) = \begin{cases} x - \text{sign}(x)\lambda & \text{if } |x| \geq \lambda \\ 0 & \text{if } |x| < \lambda. \end{cases} \quad (3.7)$$

The operation  $T_\lambda(x)$  thresholds and shrinks the entries of  $x$  by the threshold level  $\lambda$ . An example of this function is shown in Figure 3.2 for  $\lambda = 3$ .



(a)

Figure 3.2: Example of the soft-thresholding function  $T_\lambda(x)$  with a threshold level  $\lambda = 3$ .

To solve the full problem in Equation 3.5, the Lagrange multiplier  $\lambda$  must be cooled, and each subproblem is solved with this particular value of  $\lambda$ . Cooling is a common strategy to solve such problems (see e.g., Daubechies et al. (2004)); during this cooling process the subproblems are solved approximately for a decreasing  $\lambda$  (Hennenfent (2008)). In practice, a limited number of inner iterations, about 5, are required to give an accurate estimate for  $\tilde{x}_\lambda$ . In terms of computational complexity this is practical since

each update of  $\mathbf{x}$  in Equation 3.7 requires only two matrix-vector multiplications.

**Result:** Estimate for  $\mathbf{x}$

```

1 initialization;
2  $\mathbf{x}_0 \leftarrow$  initial guess for  $\mathbf{x}$ ;
3  $\lambda_0 \leftarrow$  initial Lagrange multiplier;
4  $L \leftarrow$  number of inner iterations;
5 while  $\|\mathbf{Ax} - \mathbf{y}\|_2 \geq \epsilon$  do
6   for  $\ell = 1$  to  $L$  do
7      $i \leftarrow i + 1$ ;
8      $\mathbf{x}_{i+1} \leftarrow \mathcal{S}_{\lambda_k}(\mathbf{x}_i + \mathbf{A}^H(\mathbf{y} - \mathbf{Ax}_i))$ ;
9   end
10   $\lambda_{k+1} \leftarrow \alpha_k \lambda_k$  with  $0 < \alpha_k < 1$ ;
11   $k \leftarrow k + 1$ ;
12 end
```

**Algorithm 1:** Iterative soft thresholding, proposed by Figueiredo and Nowak (2003).

In the next chapter we apply the theory of compressive sampling combined with  $\ell_1$  minimization to investigate which transform performs best when recovering seismic signals from different types of restrictions.

# Chapter 4

## Results

In this section, we demonstrate the performance of recovery using Algorithm 1 on a synthetic seismic wavefield. We also compare the compressibility and mutual coherence of the transforms in the physical and frequency domains. We try to determine if there is indeed a correlation between the sparsity/coherence and the signal recovery performance of a particular transform.

### 4.1 Recovering from a set of missing traces

The first problem that we are considering is the sparse recovery of a seismic shot gather from a partial set of seismic traces. The model, consisting from a full set of traces and the data, consisting from a partial set of traces are shown in Figures 4.1(a) and 4.1(b) respectively.

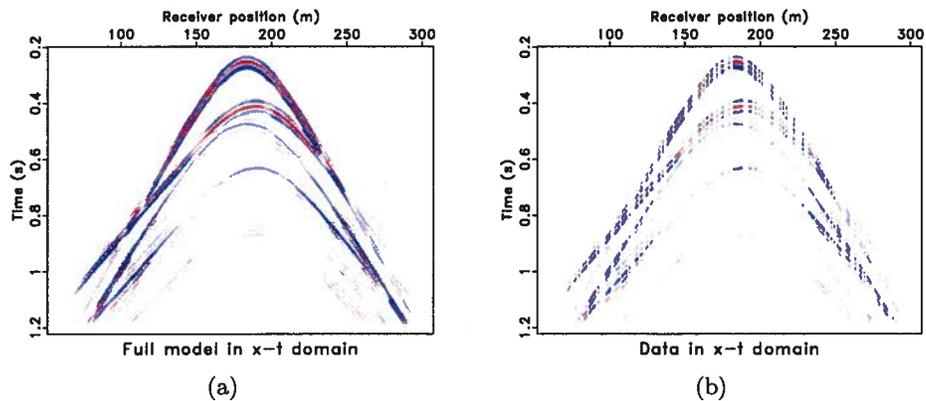


Figure 4.1: Seismic data and missing traces. (a) Model - synthetic seismic wavefield. (b) Data - same seismic wavefield as in (a) but with 50% of the traces removed

In this case the data is generated by randomly removing 50% of the traces. In section 4.2.2, we will demonstrate that the success of the recovery also

depends on in which manner the data is sampled. We apply the iterative soft-thresholding algorithm to the data in Figure 4.1(b) and test the recovery by the four different transforms - curvelets, surfacelets, shift-invariant and complex wavelets. In this problem the measurement basis is the Dirac measurement basis. Mathematically,  $\mathbf{S}$  in Equation 3.5 is one of the four transforms and  $\mathbf{M} := \mathbf{I}$ .

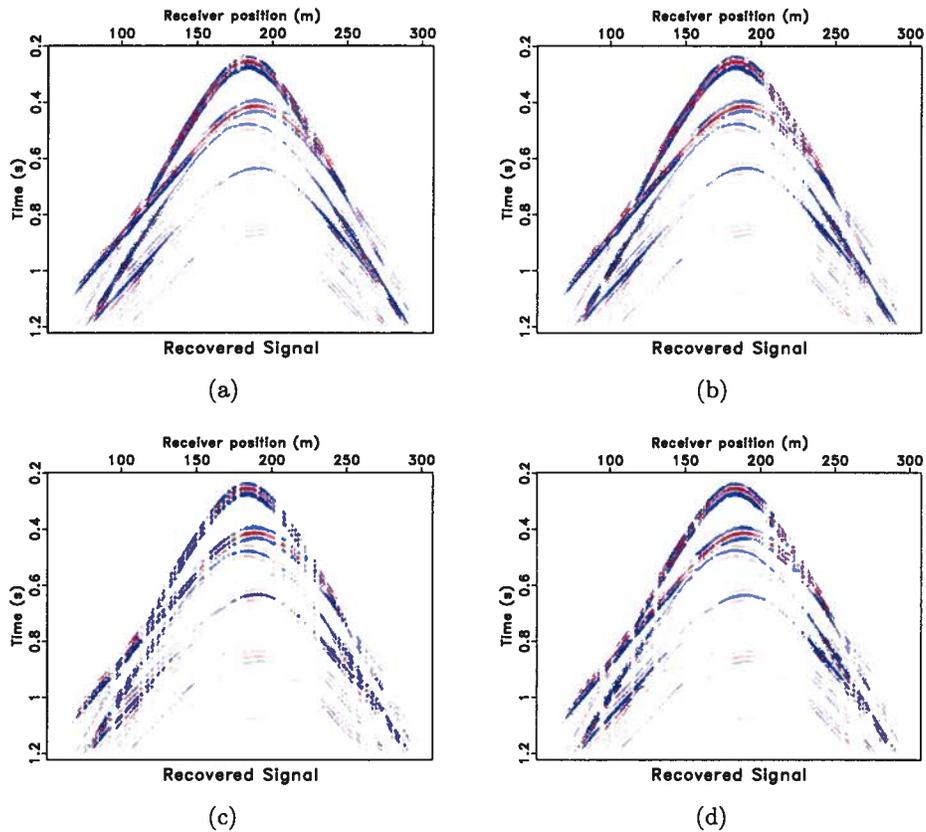


Figure 4.2: Recovered seismic wavefield via (a) curvelets; (b) surfacelets; (c) shift-invariant wavelets; (d) complex wavelets

The recovery results by the iterative soft-thresholding algorithm using curvelets, surfacelets shift-invariant wavelets and complex wavelets are plotted in Figures 4.2(a), (b), (c) and (d) respectively. Using the IST algorithm (Algorithm 1), we have limited ourselves to 5 inner-loop and 30 outer-loop

iterations. The signal-to-noise ratios, computed by

$$SNR = -20 \log \|\mathbf{f}_0 - \tilde{\mathbf{f}}\|_2 / \log \|\mathbf{f}_0\|_2, \quad (4.1)$$

where  $\mathbf{f}_0$  is the full model and  $\tilde{\mathbf{f}}$  is the recovered signal, are: 7.49 dB for curvelets, 5.36 dB for surfacelets, 3.99 dB for shift-invariant wavelets and 4.34 dB for complex wavelets. As expected from the coherency argument and the empirical decay of coefficients, curvelets (the ones that are most incoherent with the Dirac measurement basis and have the fastest coefficient decay rate) are the most successful in the recovery. Compared to curvelets, shift-invariant and complex wavelets had very limited success. The recovery results from missing seismic traces are summarized in Table 4.1.

Transform	Shift-invar. wavelet	complex wavelet	curvelet	surfacelet
SNR (dB)	3.99	4.33	7.49	5.36

Table 4.1: Physical restriction performance summary

## 4.2 Recovering from a set of missing angular frequencies

In the second problem that we are investigating, we are assuming that our data comes from only a subset of temporal frequencies. The model and the data are shown in Figures 4.3(a) and 4.3(b), respectively.

The data is generated by taking the frequency spectrum of the unrestricted synthetic wavefield from Figure 4.1(a) and randomly removing 50% of the temporal frequency content. In this case, just as in the previous case, the matrix  $\mathbf{S}$  in Equation 3.5 is still one of the four transforms, but now the measurement basis is Fourier, so  $\mathbf{M} := \mathbf{F}$ . The problem of recovering a signal from a set of missing seismic traces (section 4.1) is an interpolation problem in the physical domain, while the recovery of a signal from a set of temporal frequencies is in fact an anti-aliasing problem in the physical domain.

This idea is shown in Figure 4.4, where the data from Figure 4.3(b) is shown in the physical domain. The data contain aliasing artifacts that are the comparable to coherent noise. We again apply the iterative soft-thresholding algorithm to test the recovery performance of the four different transforms. Again, just as in the previous problem, we limit ourselves to 5 inner-loop and 30 outer-loop iterations.

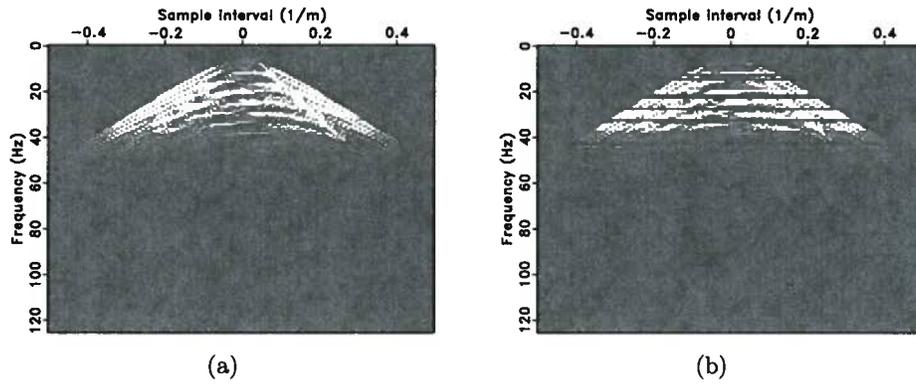


Figure 4.3: Fourier spectrum and missing frequencies. (a) The model - frequency spectrum of the synthetic wavefield in Figure 4.1(a). (b) The data - 50% of the temporal frequencies are missing.

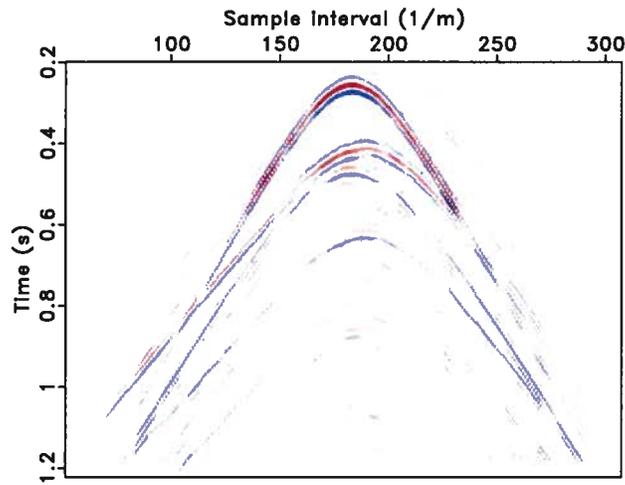


Figure 4.4: The data from Figure 4.3(b) in the physical domain. Aliasing artifacts are present throughout the entire image.

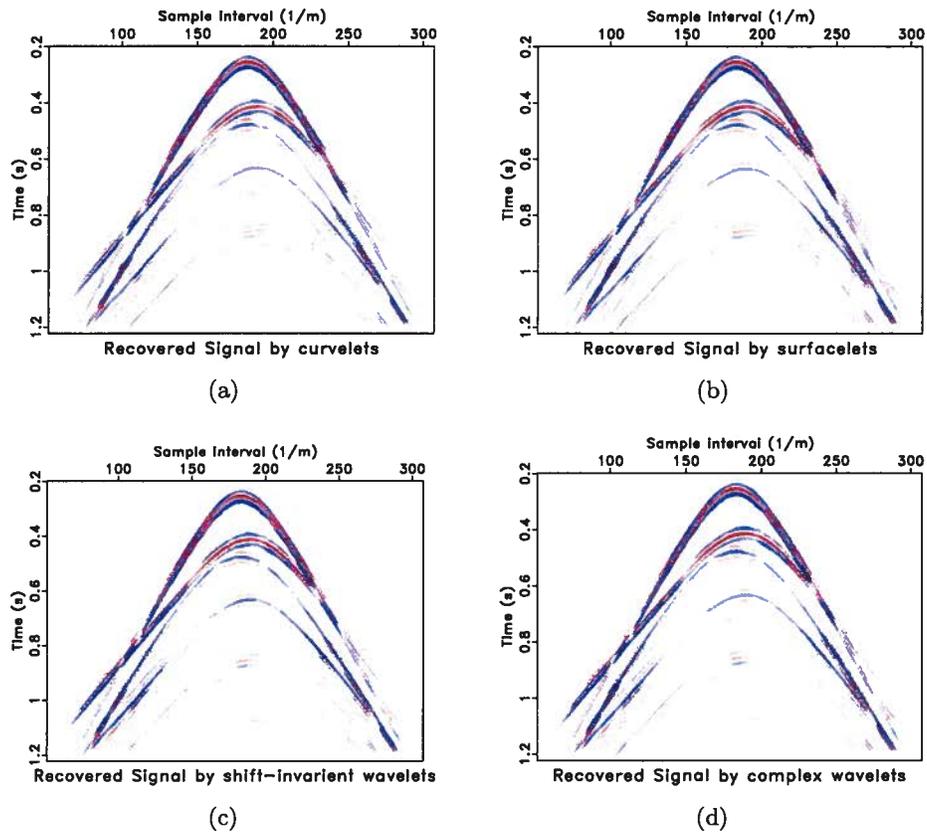


Figure 4.5: Recovered seismic wavefield via (a) curvelets; (b) surfacelets; (c) shift-invariant wavelets; (d) complex wavelets.

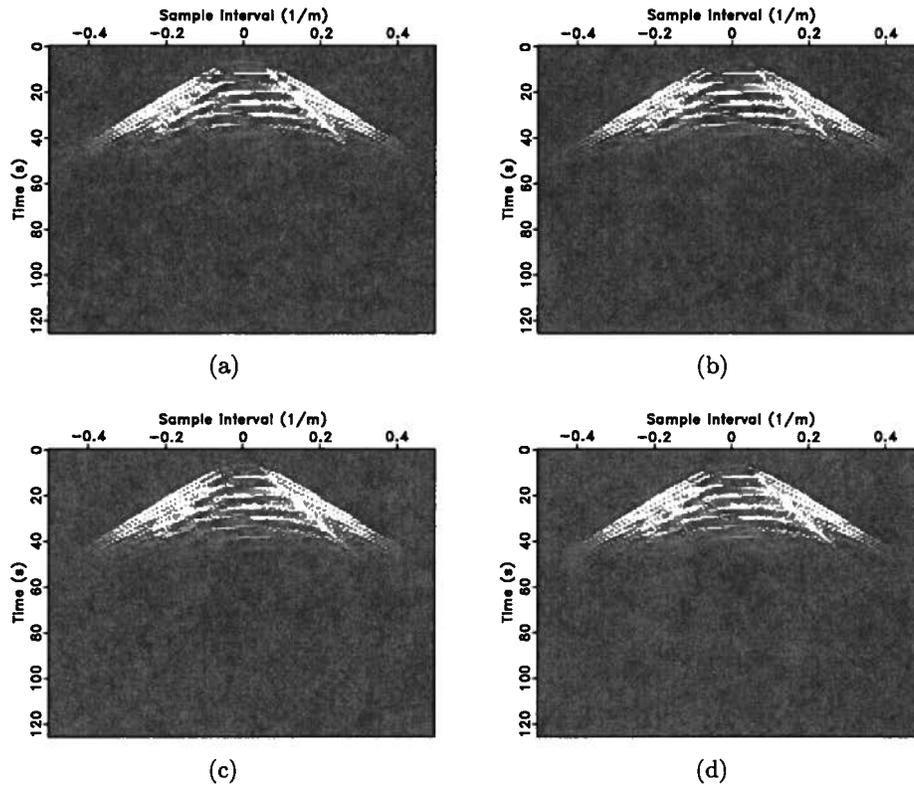


Figure 4.6: Fourier spectra of the recovered signals via (a) curvelets; (b) surfacelets; (c) shift-invariant wavelets; (d) complex wavelets.

The recovered signals by curvelets, surfacelets, shift-invariant wavelets and complex wavelets are shown in Figures 4.5(a), 4.5(c), 4.5(c) and 4.5(d). The corresponding Fourier spectra are shown in Figure 4.6. The signal-to-noise ratios for the four different cases are: 5.77 dB for curvelets, 6.17 dB for surfacelets, 11.0 dB for shift-invariant wavelets and 8.01 dB complex wavelets. Again, just as expected from the coherency argument, out of the four transforms shift-invariant wavelets had the best success in recovering the signal from missing frequencies. The wavelet recovery result has the highest SNR and the fewest aliasing artifacts left over. The recovery results from missing angular frequencies are summarized in Table 4.1.

Transform	Shift-invar. wavelet	complex wavelet	curvelet	surfacelet
SNR (dB)	11.0	8.01	5.77	6.17

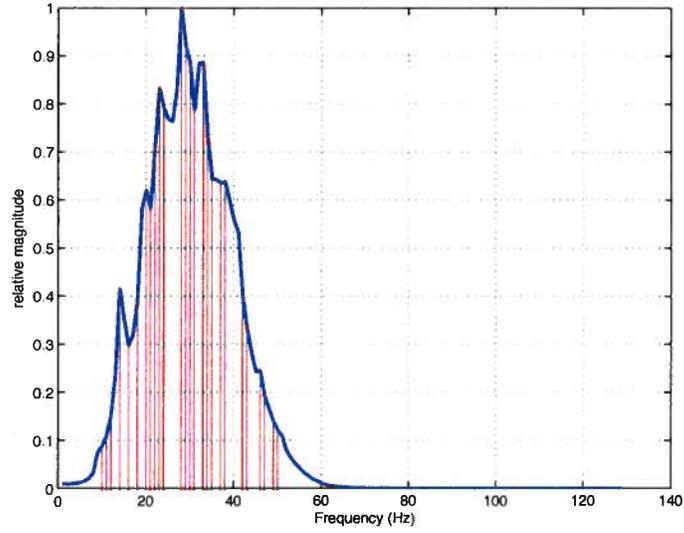
Table 4.2: Frequency restriction performance summary.

#### 4.2.1 A different sampling scheme

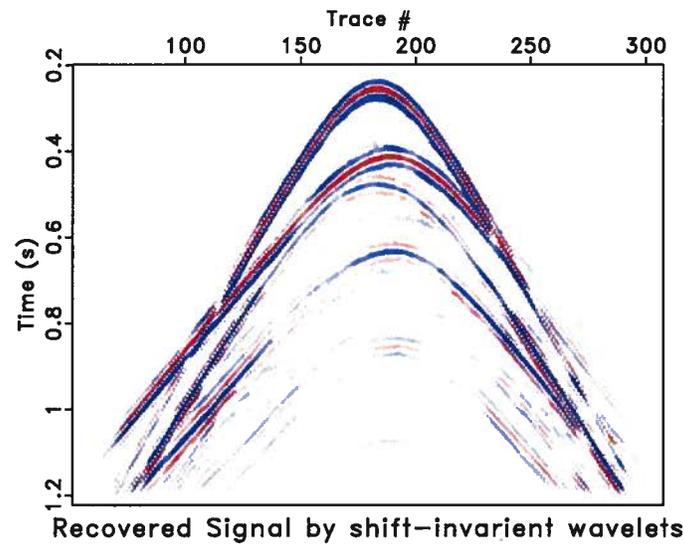
The sampling scheme used in section 4.2 assumes that we have no previous knowledge of the distribution of the frequency content of our data. That is, we assume the the frequencies are uniformly distributed over the full spectrum, and we sample this spectrum in a uniform manner. From Figure 4.3(b) it can be seen that most of the frequency content of the wavefield is concentrated around the 5 Hz to 50 Hz band. If we have this previous knowledge we can use this to sample our data only in the region that is know to have to highest frequency content.

The averaged frequency spectrum, shown in Figure 4.7(a) is now used as a pseudo probability density function to determine where our samples will come from. That is, the averaged spectrum is used as an envelope for the new sampling scheme, i.e., most of the samples are measured from frequency intervals that have the highest relative magnitude.

The signal in Figure 4.7(b) shows the recovered result by taking the samples from Figure 4.7(a) (red lines). As expected, we have a significant improvement over the purely random sampling scheme result in Figure 4.5(c). The signal-to-noise ratio for the band random sampling scheme is 11.86 dB, while in the uniform random sampling case the SNR is 11.0 dB. This is a significant improvement since in this case we were recovering from only 20% of the original temporal frequencies while in the previous case we were recovering from 50% of the original frequencies.



(a)



(b)

Figure 4.7: Frequency-weighted sampling scheme. (a) Averaged frequency distribution with 20% of the frequencies used to sample the signal (red lines); (b) recovered signal via shift-invariant wavelets with the frequency samples taken from (a).

### 4.2.2 The importance of random sampling

Lets return to the scenario where we assume to have no previous knowledge on the location of the frequency content of the data. One might think that if this is the case, then instead of sampling the data randomly, where we could miss locations of high frequency content, we could try to uniformly sample the data. In this part we will show that the success of signal recovery is very much dependent on the sampling scheme used. Lets first consider a 1d example.

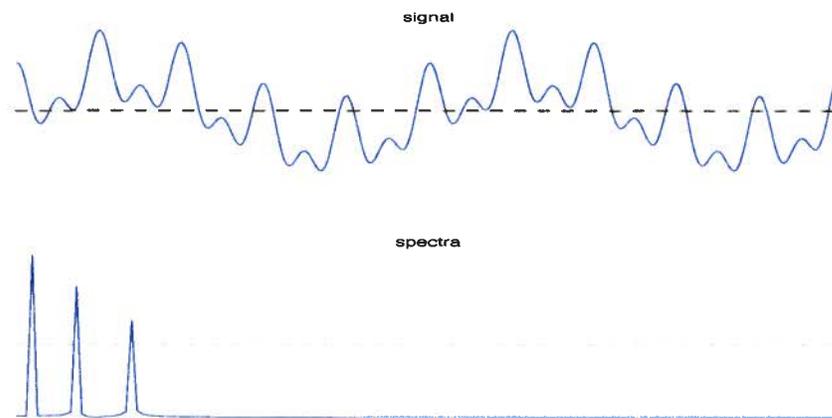


Figure 4.8: A 1D signal and it's Fourier spectrum

The signal in Figure 4.8 is a signal composed of a superposition of three cosine functions with different frequencies. The corresponding Fourier spectrum has three distinct peeks corresponding to the three different frequencies. Lets subsample the signal in a random manner (in the physical domain).

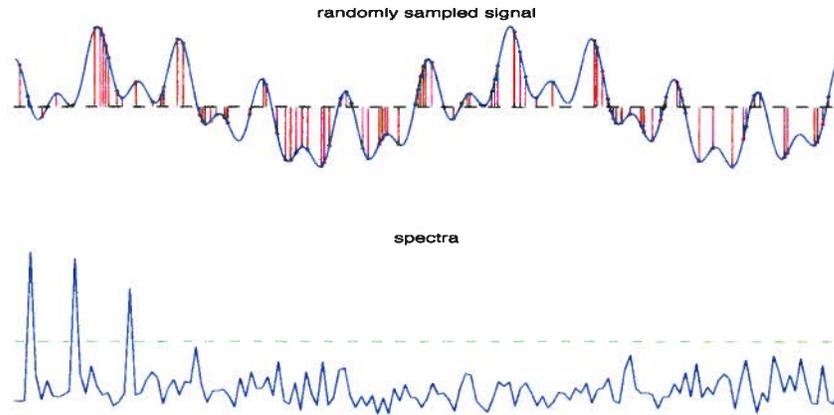


Figure 4.9: Randomly subsampled 1D signal and it's Fourier spectrum

Figure 4.9 show the original signal randomly subsampled by a factor of 3. The red lines represent the samples actually acquired. The Fourier spectra of the subsampled signal is noisy, but never-the-less we can still see the three distinct spikes corresponding to the three different frequencies. If we set to zero all the coefficients that are below the threshold level (the green line), that when we synthesize back to the physical domain, we can exactly recover our signal. Now lets subsample the signal in a uniform manner.

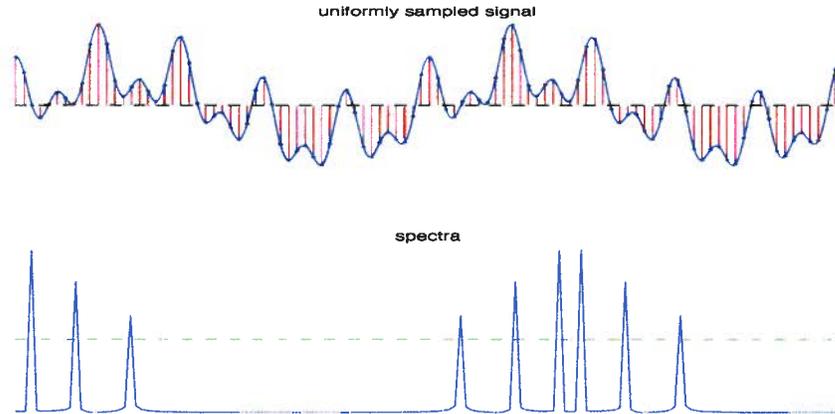


Figure 4.10: Uniformly subsampled 1D signal and its Fourier spectrum

The signal in Figure 4.10 is a three-fold uniformly subsampled signal. The Fourier spectrum is no-longer noisy; it now consists of the three original spikes plus some coherent aliasing artifacts. All of these spikes live above the threshold level (that same threshold level as in the previous Figure). Therefore, after thresholding and synthesizing back to the physical domain we would not get back our original signal, but instead a distorted and aliased copy of it. We now demonstrate this idea on our 2-D synthetic wavefield.

Figure 4.11(c) shows a three-fold uniformly subsampled frequency spectrum of the synthetic wavefield. Figure 4.11(b) shows the attempt to recover the signal with the iterative soft-thresholding algorithm using the shift-invariant wavelet transform. Clearly the algorithm is unsuccessful since we have coherent aliasing artifacts left in the recovered result. In this case the signal to noise ratio is 1.58 dB.

### 4.2.3 Direct Fourier spectra interpolation

Motivated by the success of curvelets to recover a signal from a set of missing seismic traces, we would like to see if we can improve upon the recovery of a signal from a set of missing angular frequencies, using curvelets. This is done by direct interpolation of the subsampled Fourier spectra. In this case

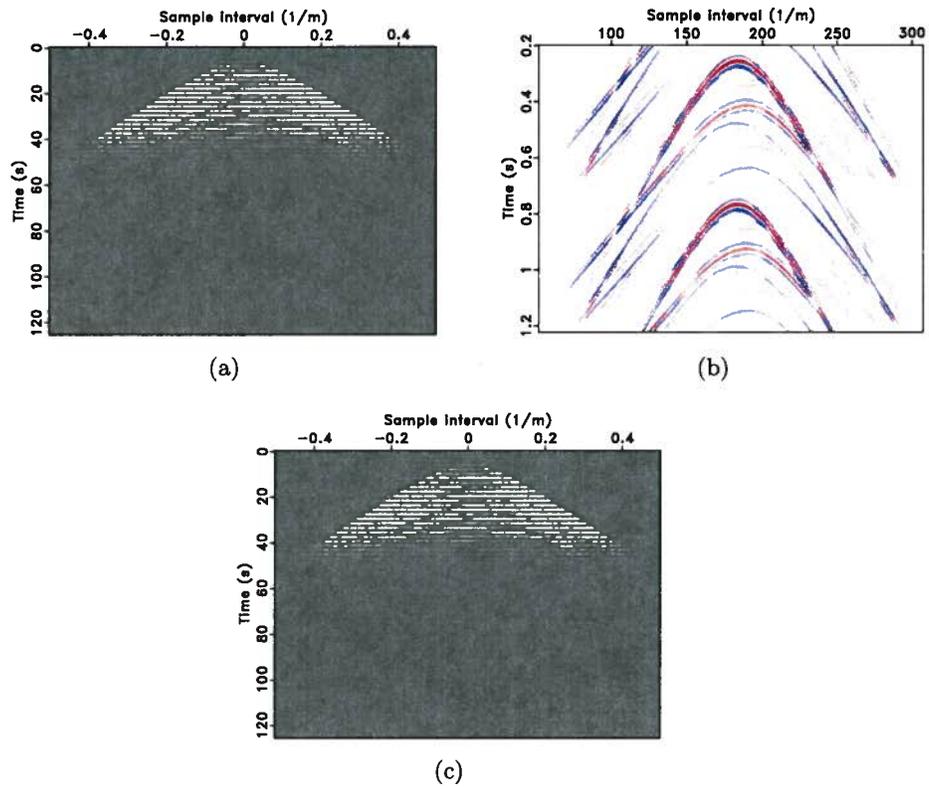


Figure 4.11: Uniform sub-sampling of angular frequencies. (a) Three-fold uniformly subsampled frequency spectra of the synthetic wavefield; (b) Attempted recovery result by shift invariant wavelets with a uniform subsampling scheme from (a); (c) interpolated spectrum.

the optimization problem can be formulated as

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon, \quad (4.2)$$

where  $\mathbf{A} = \mathbf{RIS}^H$ ,  $\mathbf{y} = \mathbf{FRIf}_0$  and  $\mathbf{S}^H$  is the synthesis of the curvelet transform. The approximation to  $\mathbf{f}_0$  is calculated by  $\tilde{\mathbf{f}} = \mathbf{F}^H\mathbf{S}^H\tilde{\mathbf{x}}$ .

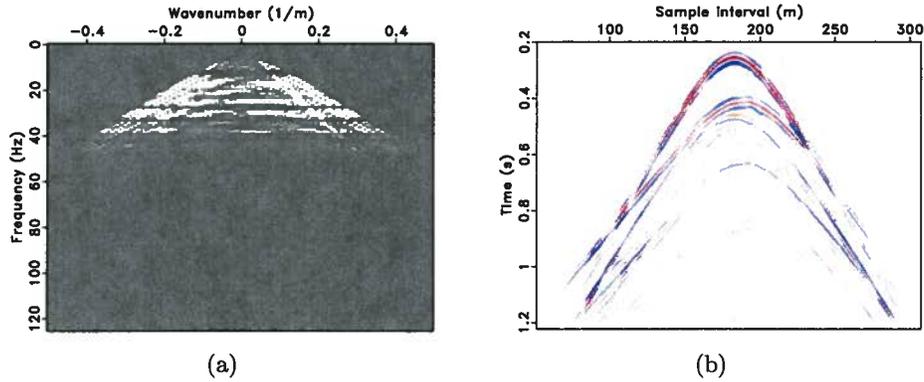


Figure 4.12: Subsampled signal. (a) Subsampled Fourier spectrum of a seismic signal; (b) same signal as in (a), but in the physical domain.

The subsampled Fourier spectrum and the signal in the physical domain are shown in Figure 4.12. The full model is shown in Figure 4.12(b). The spectra in Figure 4.12(a) is two-fold under-sampled.

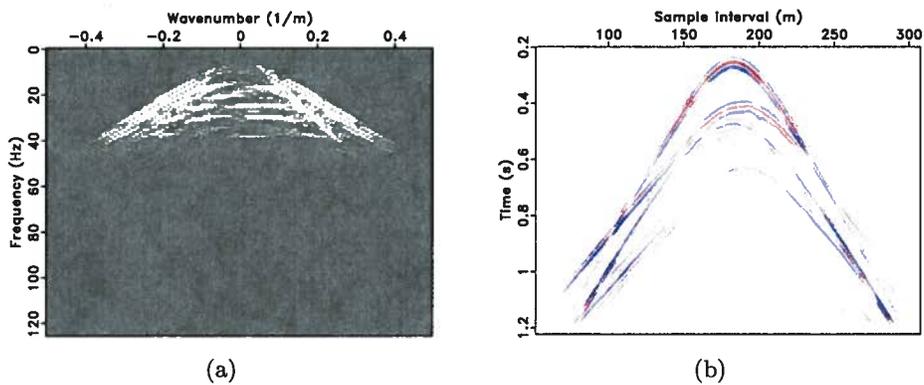


Figure 4.13: Directly interpolating the Fourier spectra. (a) Interpolated spectra from Figure 4.12(a); (b) interpolated signal in the physical domain.

The results of direct Fourier spectra interpolation are shown in Figure 4.13. The interpolated spectrum is shown in Figure 4.13(a) and the recovered signal in the physical domain is shown in Figure 4.13(b). Again the IST solver, with 5 inner and 30 outer loop iterations was applied to Problem 4.2 to obtain these results.

Unfortunately the results from directly interpolation the Fourier spectrum are not as favorable as the ones observed in section 4.2. The recovered signal in Figure 4.13(b) has a significant number of aliasing artifacts left over. The SNR for the recovered signal is 4.64 dB. Using shift-invariant wavelets we observed better results when recovering a signal from missing angular frequencies, as formulated in Problem 3.2 The wavelet recovery for a two-fold subsampled signal in the frequency domain had a SNR of 11.0 dB.

### 4.3 Compressibility

In this section, we compare the compressibility of seismic data with respect to the four transforms. In this context, compressibility means that only a few coefficients are required to capture most of the energy of a signal.

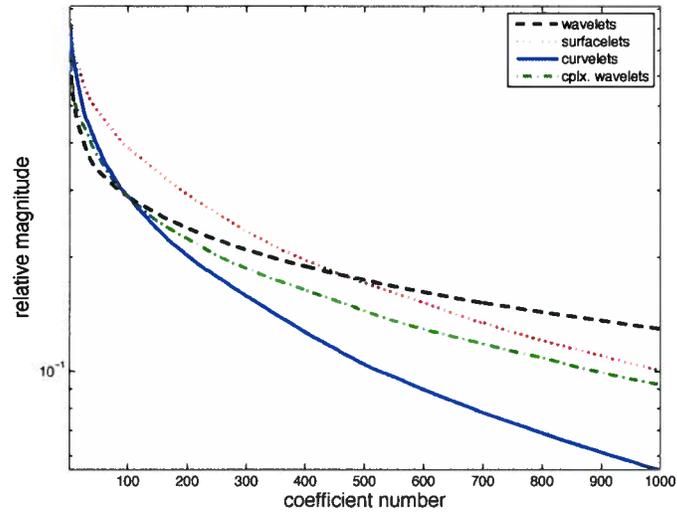
#### 4.3.1 Nonlinear approximation

Since each of the four transforms provides an over-complete signal representation with different redundancy factors, it would not be a fair to just compare the decay rates of the sorted coefficients corresponding to each different transform. Two different approaches have been used in the literature to remedy the problem of different redundancy factors (see e.g., Tatum and Collins (2003)). The first approach is a Monte Carlo type of sampling of coefficients where we account for the different redundancies. For example, the curvelet transforms is approximately 8 times redundant (in 2-D) so we randomly remove 7 out of 8 of the curvelet coefficients. This procedure is repeated  $n$  times ( $n$  should be large, say 100) and the resulting coefficients are averaged and the resulting *averaged* coefficients are plotted in magnitude-descending order. The same procedure is done for the other three transforms. The results are shown in Figure 4.14(a).

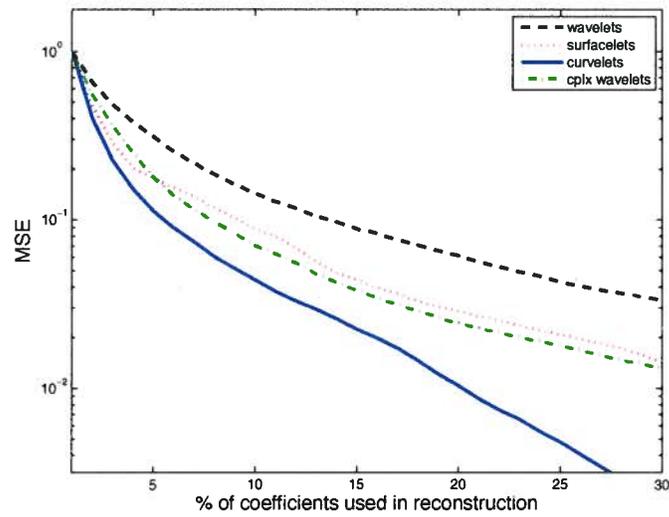
The other approach is, instead of analyzing at the relative decay rate of the coefficients we can investigate the mean squared error (or the  $\ell_2$  error) as a function of the *percentage* of coefficients used in the reconstruction. This type of comparison is shown in Figure 4.14(b).

From the two plots in Figure 4.14 we can see that curvelet coefficients decay in magnitude significantly faster than the other three transforms' coefficients. The results are similar in the compression plot - a significantly smaller fraction of coefficients are required to attain the same  $\ell_2$  error from a partial reconstruction of curvelet coefficients than from any other transforms partial set of coefficients, i.e., 30% of shift-invariant wavelet coefficients are required to get the same MSE compared to when we use only 16% of curvelet coefficients. In both plots (compression and  $\ell_2$  error) the experiments were performed on a seismic dataset, shown in Figure 4.1(a). The same number of scales (5) was used in every transform.

From Figures 4.14(a) and 4.14(b) we get an idea about the compressibility of seismic datasets with respect to different transforms. But never the less this is a rather rudimentary way of trying to determine which transform will perform better in seismic signal recovery. The Pareto curve (discussed in the next section) is perhaps a more suitable tool.



(a)



(b)

Figure 4.14: Nonlinear approximation. (a): Randomly sampled transform coefficients plotted in magnitude-descending order. (b) MSE of the reconstruction from an incomplete set of transform coefficients

### 4.3.2 The Pareto curve

Another tool to measure the compressibility of a signal could be obtained using the Pareto curve (see e.g., Berg and Friedlander (2008)) associated with the optimization problem given in Equation 3.2. The use of this curve has been previously proposed to compare the compressibility of seismic signals by curvelets and Fourier in Hennenfent and Herrmann (2008a). We extend this approach to include surfacelets, shift-invariant and complex wavelets in the compressibility comparison, and test whether one can indeed infer the efficiency of a transform for sparse recovery from the corresponding Pareto curves.

The Pareto curve traces the optimal tradeoff between the data misfit and some prior. We use  $SPGL_1$  (Berg and Friedlander (2007)) to solve the basis pursuit problem ( $BP_0$ ) (Chen et al. (1999)), where  $BP_\sigma$  is

$$BP_\sigma : \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ s.t. } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \sigma. \quad (4.3)$$

In our case we are assuming a noise-free dataset ( $\sigma = 0$ ). The matrix  $\mathbf{A}$  corresponds to the synthesis matrix of one of the above-described transforms and  $\mathbf{y}$  is the data. The columns of each  $\mathbf{A}$  are normalized to have unit-norm. Stated in plain English, we try to find a vector of coefficients that has the smallest  $\ell_1$ -norm that fits our data.

The lines in Figure 4.15 represent the Pareto curves for the four different transforms, plus Fourier. The curves were computed on a synthetic seismic dataset shown in Figure 4.1(a). For each transform, the point at the bottom of the  $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$  axis represents the  $\ell_1$  norm of the solution to the  $BP_0$  problem, i.e., the vector of transform coefficients with the smallest  $\ell_1$  norm that perfectly describes the data.

In Figure 4.15 we can see that Fourier achieves the poorest compressibility of the signal. Curvelets provide the sparsest representation of the data with  $\|\mathbf{x}_{\text{curvelet}}\|_1 = 32.6$ , the smallest  $\ell_1$  norm solution among all transforms. Hennenfent and Herrmann (2008a) suggested that this type of analysis provides a better understanding of signal compressibility than just comparing the empirical decay rates of the magnitude sorted coefficients as was done by Candès et al. (2006). By tracing the solution path to the  $BP$  problem the Pareto curve provides us with a concrete way of comparing transform compressibility. The compressibility comparison of the magnitude-sorted coefficients is more suitable when comparing transforms which have the same redundancy factors. It is important to note that the transforms' recovery performance, observed in the next chapter, does not correspond to either

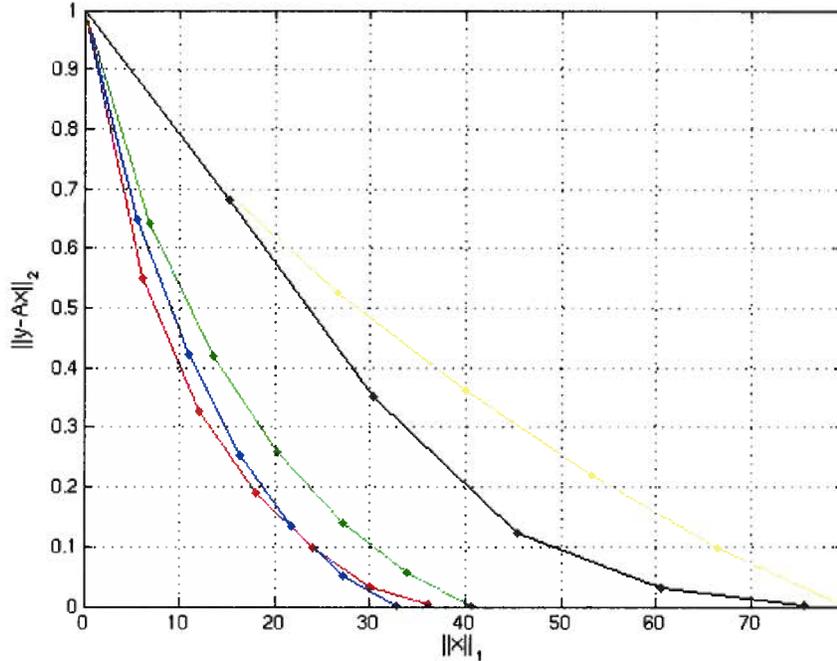


Figure 4.15: Pareto curves for the different transforms. The lines in the plot represent Pareto curves for compressibility of seismic data by: Fourier (yellow), curvelets (blue), surfacelets (red), complex wavelets (green) and shift-invariant wavelets (black)

the order of  $BP$  solutions observed in the Pareto curve, nor the empirical decay rates of the sorted coefficients.

#### 4.4 Mutual coherence

In this section we adapt this theory of compressive sampling to our problems of interest, i.e., we would like to determine the coherence  $\mu$  of the two measurement domains Dirac ( $\mathbf{M} := \mathbf{I}$ ) and Fourier ( $\mathbf{M} := \mathbf{F}$ ), for the four transform domains - curvelets, surfacelets, shift-invariant and complex wavelets. Although it is computationally expensive to calculate this quantity even for small 2-D signals, we can get an idea of its relative values from

one 'x-let' in the physical domain and its corresponding frequency response.

Figure 4.16 demonstrates this coherence idea. It shows one curvelet, surfacelet, shift-invariant wavelet and complex wavelet plotted in the physical and frequency domains, all at the same scale. One can see that a curvelet appears to be most incoherent in the physical domain (i.e.  $\mu(\mathbf{I}, \mathbf{S}_{curvelet})$  is the smallest), that is, there are large oscillations along one direction, and it appears to be the least-localized in the physical domain. Shift-invariant wavelets appear to be the least localized in frequency and thus they are most incoherent in the Fourier domain ( $\mu(\mathbf{F}, \mathbf{S}_{wavelet})$  is the smallest).

The basis functions in Figure 4.16(a) were generated by applying a synthesis transform on a vector with a single non-zero entry. The corresponding frequency response, Figure 4.16(b), is generated by taking an *fft* of the basis functions in Figure 4.16(a), applying an *fftshift* (zero-frequencies are now in the center) and plotting their absolute values.

#### 4.4.1 Coherence approximation

A more deterministic way of looking at the coherence of the four transforms is to use a Monte-Carlo (Nagata and Watanabe (2008)) type of method. Rather than calculate the full set of inner products, as defined by Equation 3.4, we approximate the coherence for only a subset of columns as follows; start by defining the quantity  $\nu_i$  by

$$\nu_i = \max |\mathbf{A}^T \mathbf{A} \mathbf{v}_i| \text{ with } \mathbf{A} = \mathbf{R} \mathbf{M} \mathbf{W} \mathbf{S}^H, \quad (4.4)$$

where  $\mathbf{S}$  is one of the four sparsifying transforms,  $\mathbf{W}$  is a diagonal matrix that ensures that the columns of  $\mathbf{A}$  have unit norm,  $\mathbf{M}$  is an measurement basis (either physical or frequency) and  $\mathbf{R}$  is a restriction operator. In the physical domain  $\mathbf{R}$  restricts the columns that correspond to missing traces and in the frequency domain  $\mathbf{R}$  restricts the rows that correspond to missing angular frequencies. The vector  $\mathbf{v}_i$  is a vector that singles out one column of  $\mathbf{A}$ , i.e.,  $\mathbf{v}_i$  is a vector of zeros with a single 1 at a random location. The coherence of the transform and the measurement basis is then computed by

$$\tilde{\mu} = \max\{\nu_i\}_{i \in [1..N]}, \quad (4.5)$$

where  $\nu_i$  is defined in Equation 4.4.  $N$  is the number of columns for which we wish to calculate  $\tilde{\mu}$ . We perform this computation on 10000 random vectors  $\mathbf{v}_i$  and the results are summarized in Table 4.3.

Chapter 4. Results

	Physical Domain			Frequency Domain		
	$\tilde{\mu}$	mean	var	$\tilde{\mu}$	mean	var
Curvelets	0.876	0.748	0.0056	0.924	0.818	0.0010
Surfacelets	0.871	0.692	0.0042	0.841	0.691	0.0037
Complex Wavelets	0.998	0.738	0.0305	0.876	0.612	0.0499
Shift-invar. wavelets	0.997	0.741	0.0413	0.810	0.690	0.0091

Table 4.3: Coherence approximations

The mean and variance are calculated from the set  $\{\nu_i\}$  for  $\nu_i$  as defined in Equation 4.4. From the coherence approximations in Table 4.3, one can see that complex and shift-invariant wavelets are the most coherent in the physical domain. Shift-invariant wavelets are the most incoherent with the frequency domain. One could use this Table 4.3 as a rough-predictor of which transform will be the most suitable for signal recovery in the two different cases. Curvelets are the most coherent in the frequency domain, and are quite incoherent in the physical domain so we expect curvelets to perform better when recovering a signal from missing traces rather than when we recover a signal from missing frequencies. By the same rational we can expect the best results when recovering a signal from missing frequencies using shift-invariant wavelets. Although we observed the best result when recovering a signal from missing traces (see Chapter 4.1) using curvelets, surfacelets appear to be slightly more incoherent in the physical domain than curvelets.

#### 4.4.2 Data-dependent coherence approximation

We hypothesize that the success of the recovery of a signal in the physical domain using curvelets might be due to the nature of the signal - i.e. curvelets achieve the sparsest representation of seismic signals (see Section 4.3.1). Instead of randomly choosing the  $\mathbf{v}$  vectors in Equation 4.4, we choose the vectors that correspond to 10000 most significant or insignificant transforms' coefficients, and repeat the same calculation for  $\tilde{\mu}$ .

Chapter 4. Results

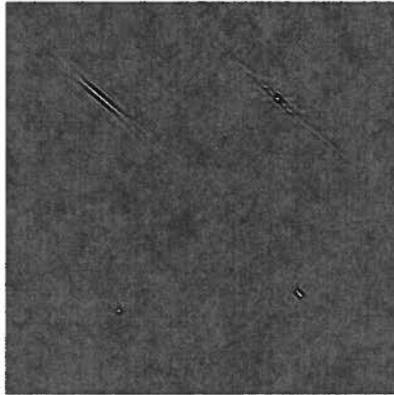
	Physical Domain			Frequency Domain		
	$\tilde{\mu}$	mean	var	$\tilde{\mu}$	mean	var
Curvelets	0.864	0.825	0.0013	0.864	0.819	0.0007
Surfacelets	0.834	0.677	0.0040	0.833	0.681	0.0035
Complex Wavelets	0.875	0.556	0.0298	0.875	0.575	0.0316
Shift-invar. wavelets	0.871	0.599	0.0301	0.872	0.582	0.0211

Table 4.4: Data-dependent coherence approximations using 10000 most *significant* transforms' coefficients

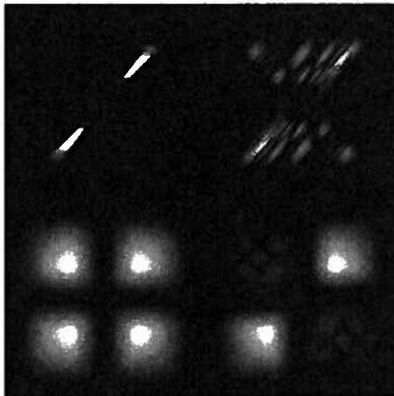
	Physical Domain			Frequency Domain		
	$\tilde{\mu}$	mean	var	$\tilde{\mu}$	mean	var
Curvelets	0.845	0.816	0.0073	0.846	0.815	0.0082
Surfacelets	0.833	0.698	0.0031	0.834	0.671	0.0029
Complex Wavelets	0.875	0.875	1.4e-26	0.875	0.666	0.0490
Shift-invar. wavelets	0.871	0.871	6.4e-19	0.822	0.670	0.0517

Table 4.5: Data-dependent coherence approximations using 10000 most *insignificant* transforms' coefficients

From Tables 4.4 and 4.5 no clear pattern can be observed that would indicate which transform should perform better when it comes to recovering a signal from either missing traces or missing angular frequencies. That is, both tables shows that surfacelets are slightly less coherent than curvelets in the physical domain, but when recovering a signal from missing traces (physical restriction), the best results are observed using curvelets (see Section 5).



(a)



(b)

Figure 4.16: Single atoms of the four transforms in the physical and frequency domains. (a): curvelet (top right), surfacelet (top right), shift-invariant wavelet (bottom left), complex wavelet (bottom right) in the physical domain. (b): Fourier transform of the atoms in the same order.

## Chapter 5

# Conclusions and discussion

Throughout this thesis, we have investigated the recovery of a seismic signal from missing angular frequencies and missing seismic traces using four different transforms: shift-invariant wavelets, complex wavelets, curvelets and surfacelets. The recovery was carried out using the iterative soft-thresholding procedure, and the algorithm was ran for a fixed number of iterations. The signal-to-noise ratios for the different experiments performed are summarized in Table 5.1.

Recovery signal-to-noise ratios		
	<b>Physical restriction</b> 50% of traces missing	<b>Frequency restriction</b> 50% of frequencies missing
<b>Curvelets</b>	<b>7.49 dB</b>	5.77 dB
<b>Surfacelets</b>	5.36 dB	6.17dB
<b>Shift-invar. wavelets</b>	3.99 dB	<b>11.0 dB</b>
<b>Complex wavelets</b>	4.33 dB	8.01 dB

Table 5.1: Recovery performance summary

In the problem of recovering the signal from missing traces the best results were observed using the curvelet transform. Curvelets were able to give the sparsest representation of or seismic signal at the cost of a moderate redundancy factor. Also, out of the four transforms curvelets appear to be most incoherent with the Dirac measurement basis, which explains, from a compressive sampling point of view why the best interpolation results were observed using curvelets.

In the problem of recovering a signal from a subset of angular frequencies the best results were observed using the shift-invariant wavelet transform, although at a cost of high redundancy. These results were found to be significantly better (an improvement by 6.36 dB) than a direct interpolation of the Fourier spectra of the subsampled signal. Out of the four transforms, shift invariant wavelets were observed to be most incoherent with the Fourier measurement basis.

We have also seen that the success of the recovery also depends on the

sampling scheme used. Uniform subsampling results in coherent aliasing artifacts and as a result the recovery algorithm fails to produce meaningful results, as can be seen in Figure 4.11(b). On the other hand, if one has some prior knowledge of where most of the frequency content is concentrated, one can sample in the band. In this case the recovery results had a high SNR (11.86 dB) when recovering from just 20% of the frequencies.

Throughout this thesis we have attempted to find a direct link between a transform's performance in recovering seismic signals from incomplete measurement and the compressibility/mutual coherence of a transform with a particular measurement basis. Unfortunately no such link was observed. This is mainly due to the fact that our sparsifying transforms correspond to redundant expansions with poor coherence properties. A more thorough investigation is required to determine precisely how non-orthonormal systems (like shift-invariant wavelets, complex wavelets, curvelets and surfacelets) tie into the theory of compressive sampling.

# Appendix A

## Wavelet concepts

### A.1 Frames

Frame theory is used to analyze the completeness, stability and redundancy of discrete transforms. It was originally developed by Duffin and Schaefer (1952) to reconstruct band-limited signals  $f$  from irregularly spaced samples  $\{f(t_n)\}_{n \in \mathbb{Z}}$ . Roughly speaking, a frame is a family of vectors  $\{\phi_n\}_{n \in \Gamma}$  in a Hilbert space  $\mathbf{H}$  that characterizes any signal  $f \in \mathbf{H}$  from its inner products  $\{\langle f, \phi_n \rangle\}_{n \in \Gamma}$  in a numerically stable way. More precisely, the sequence  $\{\phi_n\}_{n \in \Gamma} \in \mathbf{H}$  is a frame of  $\mathbf{H}$  if there exist two constants  $A > 0$  and  $B > 0$  such that for any  $f \in \mathbf{H}$

$$A\|f\|^2 \leq \sum_{n \in \Gamma} |\langle f, \phi_n \rangle|^2 \leq B\|f\|^2. \quad (\text{A.1})$$

When  $A = B$  the frame is said to be tight. A frame defines a complete and stable signal representation, which may or may not be redundant. The redundancy of the signal representation is reflected by the frame bounds  $A$  and  $B$  if the frame vectors  $\phi_n$  have unit norm. Indeed, a frame  $\{\phi_n\}_{n \in \Gamma}$ , with  $\|\phi_n\| = 1, \forall n \in \Gamma$ , is an orthonormal basis if and only if the frame bounds satisfy  $A = B = 1$ .

# Bibliography

Bamberger, R. H. and M. J. T. Smith, 1992, A filter bank for the directional decomposition of images: theory and design: *IEEE Transactions on Image processing*, **40**, 882–893.

Berg, E. v. and M. P. Friedlander, 2007, Spgl1: A solver for large-scale sparse reconstruction. (<http://www.cs.ubc.ca/labs/scl/spgl1/?n=HomePage>).

———, 2008, Probing the pareto frontier for basis pursuit solutions: Tech. Rep. TR-2008-01, Department of Computer Science, University of British Columbia, Vancouver. (To appear in *SIAM J. Sci. Comp.*).

Birgin, E. G., J. M. Martínez, and M. Raydan, 2000, Nonmonotone spectral projected gradient methods on convex sets: *SIAM J. on Optimization*, **10**, 1196–1211.

Blumensath, T. and M. E. Davies, 2008, Iterative hard thresholding for compressed sensing (technical report).

Candès, E. and J. Romberg, 2007, Sparsity and incoherence in compressive sampling: *Inverse Problems*, **23**, 969–985.

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

Candès, E. J. and L. Demanet, 2005, The curvelet representation of wave propagators is optimally sparse: *Communications on Pure and Applied Mathematics*, **58**, no. 11, 1472–1528.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: *Multiscale Modeling and Simulation*, **5**, no. 3, 861–899.

Candès, E. J. and D. L. Donoho, 2004, New tight frames of curvelets and optimal representations of objects with piecewise- $C^2$  singularities: *Communications on Pure and Applied Mathematics*, **57**, no. 2, 219–266.

## Bibliography

---

- Candès, E. J. and M. Wakin, 2007, People hearing without listening: an introduction to compressive sampling.: To appear in *IEEE Transactions on Signal Processing*.
- Chen, S. S., D. L. Donoho, and M. A. Saunders, 1999, Atomic decomposition by basis pursuit: *SIAM Journal on Scientific Computing*, **20**, 33–61.
- Dagher, J., A. Bilgin, and M. Marcellin, 2003, Resource-constrained rate control for motion jpeg2000: *Image Processing, IEEE Transactions on*, **12**, 1522–1529.
- Daubechies, I., M. Defrise, and C. De Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: *Communications on Pure and Applied Mathematics*, **57**, no. 11, 1413–1457.
- Donoho, D., 2006, Compressed sensing: *Information Theory, IEEE Transactions on*, **52**, 1289–1306.
- Donoho, D. L. and I. M. Johnstone, 1994, Ideal spatial adaptation by wavelet shrinkage: *Biometrika*, **81**, 425–455.
- Duffin, R. J. and A. C. Schaefer, 1952, A class of nonharmonic fourier series: *Trans. Amer. Math. Soc.*, **72**, 341–366.
- Elad, M., J.-L. Starck, P. Querre, and D. L. Donoho, 2008, Simultaneous cartoon and texture image inpainting using morphological component analysis (mca): **19**, 340–358.
- Figueiredo, M. and R. Nowak, 2003, An em algorithm for wavelet-based image restoration: *Image Processing, IEEE Transactions on*, **12**.
- Gersztenkorn, A., J. B. Bednar, and L. R. Lines, 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: *Geophysics*, **51**, 357–368.
- Hennenfent, G., 2008, Sampling and reconstruction of seismic wavefields in the curvelet domain: PhD thesis, The University of British Columbia, Vancouver, BC Canada.
- Hennenfent, G. and F. J. Herrmann, 2008a, One-norm regularized inversion: learning from the pareto curve: Submitted to the SEG 2008 convention.
- , 2008b, Simply denoise: wavefield reconstruction via jittered under-sampling: *Geophysics*, **73**, no. 3.
- Herrmann, F. J. and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: *Geophysical Journal International*, **173**, 233–248.

## Bibliography

---

- Holschneider, M., R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, 1989, *Wavelets, time-frequency methods and phase space*: Springer-Verlag, Berlin.
- Kingsbury, N., 1999, Image processing with complex wavelets: *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, **357**, 2543–2560.
- Lilly, J. M. and S. C. Olhede, 2008, On the design of optimal analytic wavelets. (technical report).
- Lin, T. Y., E. Lebed, Y. A. Erlangga, and F. J. Herrmann, 2008, Interpolating solutions of the helmholtz equation with compressed sensing: Presented at the SEG.
- Lu, Y. and M. N. Do, 2006, Video processing using the 3-dimensional surfacelet transform: *Signals, Systems and Computers*, 2006. ACSSC '06. Fortieth Asilomar Conference on, 883–887.
- Lu, Y. M. and M. N. Do, 2007, Multidimensional directional filter banks and surfacelets: *Transactions on Image Processing*, **16**, no. 4, 918–931.
- Lustig, M., D. Donoho, and J. M. Pauly, 2007, Sparse mri: The application of compressed sensing for rapid mr imaging: *Magnetic Resonance in Medicine*, **8**, 1176–1194.
- Mallat, S., 1999, *A wavelet tour of signal processing*, second edition: Academic Press.
- Masud, S. and J. McCanny, 1999, Rapid vlsi design of biorthogonal wavelet transform cores: *Signal Processing Systems*, 1999. SiPS 99. 1999 IEEE Workshop, 291–300.
- Nagata, K. and S. Watanabe, 2008, Exchange monte carlo sampling from bayesian posterior for singular learning machines: *Neural Networks, IEEE Transactions on*, **19**, 1253–1266.
- Sacchi, M. D., T. J. Ulrych, and C. J. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: *Transactions on Signal Processing*, **46**, no. 1, 31–38.
- Schwab, M., 1993, Shot gather continuation. (technical report).
- Selesnick, I. W., R. G. Baraniuk, and N. G. Kingsbury, 2005, The dual-tree complex wavelet transform: *IEEE Transactions. Signal processing*, 123–151.

## *Bibliography*

---

- Shensa, M., 1992, The discrete wavelet transform: Wedding the à trous and mallat algorithms: *IEEE Transactions on Information Theory*, **40**, 2464–2482.
- Starck, J.-L., J. Fadili, and F. Murtagh, 2007, The undecimated wavelet decomposition and its reconstruction.: *IEEE Transactions on Image Processing*, **16**, 297–309.
- Stovas, A. M. and S. B. Fomel, 1993, Kinematically equivalent dmo operators: *Russian Geology and Geophysics*, **37**, 102–113.
- Tantum, S. and L. Collins, 2003, Performance bounds and a parameter transformation for decay rate estimation: *Geoscience and Remote Sensing, IEEE Transactions on*, **41**, 2224–2231.
- Tsaig, Y. and D. L. Donoho, 2006, Extensions of compressed sensing: *Signal Process.*, **86**, 549–571.