Sampling and reconstruction of seismic wavefields in the curvelet domain

by

Gilles Hennenfent

Diplôme d'Ingénieur, École Nationale Supérieure de Physique de Strasbourg, 2003
Diplôme d'Études Approfondies, Université Louis Pasteur Strasbourg, 2003

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

Doctor of Philosophy

in

The Faculty of Graduate Studies

(Geophysics)

THE UNIVERSITY OF BRITISH COLUMBIA

April, 2008

Abstract

Wavefield reconstruction is a crucial step in the seismic processing flow. For instance, unsuccessful interpolation leads to erroneous multiple predictions that adversely affect the performance of multiple elimination, and to imaging artifacts. We present a new non-parametric transform-based reconstruction method that exploits the compression of seismic data by the recently developed curvelet transform. The elements of this transform, called curvelets, are multi-dimensional, multi-scale, and multi-directional. They locally resemble wavefronts present in the data, which leads to a compressible representation for seismic data. This compression enables us to formulate a new curvelet-based seismic data recovery algorithm through sparsity-promoting inversion (CRSI). The concept of sparsity-promoting inversion is in itself not new to geophysics. However, the recent insights from the field of "compressed sensing" are new since they clearly identify the three main ingredients that go into a successful formulation of a reconstruction problem, namely a sparsifying transform, a sub-Nyquist sampling strategy that subdues coherent aliases in the sparsifying domain, and a data-consistent sparsity-promoting program.

After a brief overview of the curvelet transform and our seismic-oriented extension to the fast discrete curvelet transform, we detail the CRSI formulation and illustrate its performance on synthetic and real datasets. Then, we introduce a sub-Nyquist sampling scheme, termed jittered undersampling, and show that, for the same amount of data acquired, jittered data are best interpolated using CRSI compared to regular or random undersampled data. We also discuss the large-scale one-norm solver involved in CRSI. Finally, we extend CRSI formulation to other geophysical applications and present results on multiple removal and migration-amplitude recovery.

Table of Contents

Appendices

List of Tables

List of Figures

Preface

This thesis was prepared with Madagascar, a reproducible research software package available at `rsf.sf.net`, in such a way that most of the reproducible results are linked to the code that generated them. Reproducibility facilitates the dissemination of knowledge not only within the Seismic Laboratory for Imaging and Modeling (SLIM) but also between SLIM and its sponsors, and more generally, the entire research community.

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 Python as part of SLIMpy (`slim.eos.ubc.ca/SLIMpy`) with a few exceptions written in Matlab® or Python.

Acknowledgments

I am greatly indebted to my advisor Felix Herrmann. A few words of acknowledgment are just not enough to express my gratitude and appreciation for all he has done for me. Felix' scientific vision and ENTHUSIASM made these last four and a half years of work very exciting. I also wish to thank him for his kind and sometimes (very) energetic encouragements.

I would like to express my gratitude to Henryk Modzelewski, Darren Thompson, Sean Ross Ross, Cody Brown, and the remainder of the SLIM team for their help and friendships. I was very fortunate to be part of this group.

The next persons who have influenced my scientific life are Michael Friedlander, Ozgur Yilmaz, Ewout van der Berg, and Rayan Saab. They helped my research incursions in the areas of applied mathematics and optimization. I would like to thank them for their time and patience.

My appreciation goes to ExxonMobil and Chevron for the hospitality during my internships. In particular, I wish to thank Ramesh Neelamani (ExxonMobil), Warren Ross (ExxonMobil), and Tamas Nemeth (Chevron) for great technical discussions. I also wish to thank Joseph Reilly (ExxonMobil) and Debra Bones (Chevron) for making these internships possible despite the amount of paperwork involved in hiring an international student.

I was very fortunate to meet and interact on a regular basis with Sergey Fomel. I am grateful to him for interesting technical discussions and for sharing with me the Madagascar software since its early days.

I would like to express my gratitude to Uri Ascher, Michael Bostock, Kenneth Bube, and Ozgur Yilmaz for serving on my supervisory committee.

Many thanks to my parents and my sister for their love, support and encouragement throughout my studies.

Finally, I have been very fortunate to meet Daiana five and a half years ago and a few words cannot do justice to the way she has changed forever my life.

To my parents and
my late grandfather, Ferdinand Hennenfent.

Statement of Co-Authorship

Chapter 2 was published with Felix J. Herrmann. The C++ code of the transform described in this chapter is the combination of pieces of existing codes. Colin Russell helped me coding. The manuscript was jointly written with Felix.

Chapter 3 was published with Felix J. Herrmann. The manuscript was jointly written with Felix. He led the more theoretical sections, and I led the more applied ones.

Chapter 4 was published with Felix J. Herrmann. I wrote the manuscript with numerous inputs from Felix.

Chapter 5 was published with Ewout van den Berg, Michael P. Friedlander and Felix J. Herrmann. I wrote the manuscript with inputs from Michael, Felix and Ewout.

Chapter 6 was published with Felix J. Herrmann, Deli Wang, and Peyman P. Moghaddam. The results on focused recovery and primary-multiple separation are due to Deli, and the ones on migration-amplitude recovery to Peyman. The manuscript was written by Felix with inputs from co-authors.

Chapter 1

Introduction

Reflection seismology is a technique widely used by the oil industry to explore and identify potential oil-rich areas into the earth. This technique involves the collection of seismic data that are indirect measurements of the earth's structure. These data are then processed to generate an image of the subsurface that is finally interpreted by geo-scientists.

Seismic data acquisition is a complex and costly operation. On land, dynamite or Vibroseis sources can be used to send energy into the subsurface. This energy propagates and partially reflects at interfaces due to a change in rock properties. The reflected wavefield is recorded at the earth's surface by an array of geophones. At sea, the principle remains the same but the seismic source is typically an air gun and the receivers are hydrophones on streamer lines towed by a seismic vessel. Fig. 1.1 schematically illustrates these two different types of seismic surveys.

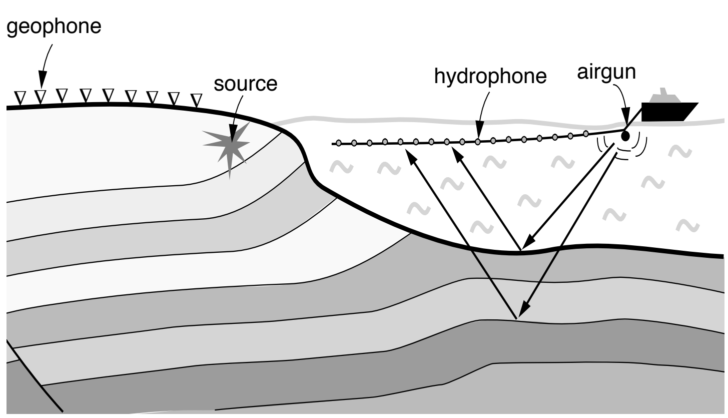


Figure 1.1: Schematic diagram of seismic acquisition.

Processing difficulties generally arise from assumptions made by algorithms, that are not met by acquired data. In particular, most of the commonly-used multi-trace processing algorithms, e.g., Surface-Related Multiple Elimination (SRME - Verschuur et al., 1992) and wave-equation migration (WEM - Claerbout, 1971), assume a dense and regular coverage of the survey area. Field datasets, however, are typically irregularly and/or coarsely sampled along one or more spatial coordinates and need to be interpolated before being processed to avoid artifacts in the final subsurface image.

For regularly-undersampled data along one or more spatial coordinates, i.e., data spatially sampled below Nyquist rate, there exists a wide variety of wavefield reconstruction techniques:

- **Filter-based methods** convolve the incomplete data with an interpolating filter—e.g., the sinc function—in the spatial domain. The most common of these filters are the prediction error filters (PEF's) that can handle aliased events (Spitz, 1991; Fomel, 2000).

- **Wavefield-operator-based methods** represent another type of interpolation approaches that explicitly include wave propagation (Canning and Gardner, 1996; Biondi et al., 1998; Stolt, 2002). They require specific knowledge of a velocity model and they are also typically fairly computationally intensive.

- **Transform-based methods** use *a priori* information about the wavefield in a transform domain—e.g., shape of the temporal and/or spatial spectrum—to solve the reconstruction problem (Sacchi et al., 1998; Trad et al., 2003; Zwartjes and Sacchi, 2007). These methods are generally the fastest approaches and their link with the physics of wave propagation depends on the transform used. For example, Fourier modes correspond to eigenfunctions of a wave equation with constant velocity and the hyperbolic Radon transform relates to the kinematics of the reflection and, hence, to ray theory.

However, for irregularly-sampled data, e.g., binned data with some of the bins that are empty, or data that are continuous random undersampled, the performance of most of the aforementioned interpolation methods deteriorates.

1.1 Theme

The main theme of this thesis is a practical, robust, and geometrical—i.e., transform-based—approach to the seismic wavefield reconstruction problem. The motivation of this approach is two key features of seismic data that are, in our opinion, not used to their full extent in existing approaches, namely

- **High dimensionality** Seismic data is typically 5D—time, two spatial coordinates for the source, and two spatial coordinates for the receiver—for a 3D survey.
- **Strong geometrical structure** Seismic data are a spatio-temporal sampling of the reflected wavefield that contains different arrivals—i.e. wavefronts—that correspond to different interactions of the incident wavefield with inhomogeneities in the Earth's subsurface.

To make the most of these features, our approach uses the curvelet transform (Candès and Donoho, 2004) which is *data-independent*, *multiscale*, and *multidirectional*. The elements of this transform, the curvelets, are localized in the frequency domain and of rapid decay in the physical domain. Because of these properties, curvelets behave as localized eigenfunctions of wave equations with varying velocity (Candès and Demanet, 2005). They are very efficient at representing curve-like singularities—e.g., wavefronts. In other words, only few curvelets are needed to represent the complexity of real seismic data. We use this piece of information, called *sparsity*, to help solve the interpolation problem.

The idea of sparsity-promoting inversion is in itself not new to geophysics. However, we adapt and use new insights from the emerging field of compressive sampling (CS - Candès et al., 2006; Donoho, 2006). These insights clearly identify the three main ingredients that go into a successful formulation of a reconstruction problem, namely a sparsifying transform, a sub-Nyquist sampling strategy that subdues coherent aliases in the sparsifying domain, and a data-consistent sparsity-promoting program.

For interest, curvelets set themselves apart from wavelets by their truly 2D and higher-dimensional nature—i.e., the curvelet transform is non-separable unlike the wavelet transform that is extended to higher dimension by tensor products.

1.2 Objectives

The objectives of this thesis are twofold:

- develop an in-depth understanding of successful sparsity-promoting inversions and their key ingredients,

- formulate a practical sparsity-promoting seismic wavefield reconstruction algorithm whose performance and limitations are well understood.

1.3 Outline

In chapter 2, we first give an overview of the curvelet transform (Candès and Donoho, 2004) and one of its discrete implementation, the fast discrete curvelet transform (FDCT) via wrapping (Candès et al., 2006). We then propose an extension of this implementation that can handle typical seismic data, i.e., data that is irregularly sampled along spatial coordinates and regularly sampled along the time coordinate. This new implementation is coined nonequally sampled fast discrete curvelet transform (NFDCT). Finally, we illustrate the performance of the NFDCT on removing incoherent and coherent noise from nonequally sampled seismic data and on binning.

Chapter 3 deals with the reconstruction of severely spatially-undersampled seismic data. We start by a brief review of CS (Candès et al., 2006; Donoho, 2006) and the key ingredients of its success. We continue by discussing the extension of CS to seismic data recovery and propose a practical algorithm, termed curvelet reconstruction with sparsity-promoting inversion (CRSI). We conclude by showing some reconstruction examples on synthetic and real data sets. For interest, further readings by the author include Herrmann and Hennenfent (2005); Hennenfent and Herrmann (2005); Thomson et al. (2006) and Hennenfent and Herrmann (2006, 2007a).

Chapter 4 focuses on coarse spatial sampling schemes that are favorable for CRSI, a topic touched upon in the previous chapter. First, we propose and analyze a coarse sampling scheme, termed *jittered undersampling* (Leneman, 1966; Dippe and Wold, 1992), which creates, under specific conditions, a favorable recovery situation for seismic wavefield reconstruction methods that impose sparsity in Fourier or Fourier-related domains (see e.g. Sacchi et al., 1998; Xu et al., 2005; Zwartjes and Sacchi, 2007; Herrmann and Hennenfent, 2008). Then, we compare the performance of CRSI on jittered data to its performance on data acquired according to other coarse sampling schemes. For interest, other references on the topic by the author are Hennenfent and Herrmann (2007b,c).

Chapter 5 deals with another topic touched upon in chapter 3, namely one-norm solvers. We draw on the work of van den Berg and Friedlander (2007) and introduce the Pareto curve as a means to understand the compromises implicitly accepted when an algorithm is given limited number of iterations. This situation virtually always occurs in geophysical processing due to the (extremely) large-scale of the problems.

In chapter 6, we show that other geophysical problems—e.g., focused recovery, seismic signal separation, and migration amplitude recovery—can be re-cast in the formulation used for CRSI. This puts in a broader perspective the insights gained during the development of CRSI. For interest, the author also co-authored Herrmann et al. (2007b,a) on this topic.

In Chapter 7, we summarize the work done in this thesis, and discuss some of its aspects in a broader context. Conclusions and recommendations for future research follow.

Appendices A,B, and C contain further details about the curvelet transform and pair with chapter 3. In appendix D, we re-derive a result used in chapter 4 and originally introduced by Leneman (1966).

Bibliography

Biondi, B., S. Fomel, and N. Chemingui, 1998, Azimuth moveout for 3D prestack imaging: Geophysics, **63**, no. 2, 1177 – 1183.

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 (FDCT): 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 C^2 singularities: Communications on Pure and Applied Mathematics, **57**, no. 2, 219 – 266.

Candès, E. J., J. Romberg, and T. Tao, 2006, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: Transactions on Information Theory, **52**, no. 2, 489 – 509.

Canning, A. and G. H. Gardner, 1996, Regularizing 3D data-sets with DMO: Geophysics, **61**, no. 4, 1103 – 1114.

Claerbout, J. F., 1971, Towards a unified theory of reflector mapping: Geophysics, **36**, no. 3, 467 – 481.

Dippe, M. and E. Wold, 1992, Stochastic sampling: theory and application: Progress in Computer Graphics, **1**, 1 – 54.

Donoho, D. L., 2006, Compressed sensing: Transactions on Information Theory, **52**, no. 4, 1289 – 1306.

Fomel, S., 2000, Three-dimensional seismic data regularization: PhD thesis, Stanford University.

Hennenfent, G. and F. J. Herrmann, 2005, Sparseness-constrained data continuation with frames: Applications to missing traces and aliased signals in 2/3-D: Presented at the SEG International Exposition and 75^{th} Annual Meeting.

———, 2006, Application of stable signal recovery to seismic interpolation: Presented at the SEG International Exposition and 76^{th} Annual Meeting.

———, 2007a, Curvelet reconstruction with sparsity-promoting inversion: successes and challenges: Presented at the Curvelet workshop – EAGE 69^{th} Conference & Exhibition.

———, 2007b, Irregular sampling: from aliasing to noise: Presented at the EAGE 69^{th} Conference & Exhibition.

———, 2007c, Random sampling: new insights into the reconstruction of coarsely-sampled wavefields: Presented at the SEG International Exposition and 77^{th} Annual Meeting.

Herrmann, F. J., D. W. G. Hennenfent, and P. P. Moghaddam, 2007a, Seismic data processing with curvelets: a multiscale and nonlinear approach: Presented at the SEG International Exposition and 77^{th} Annual Meeting.

Herrmann, F. J. and G. Hennenfent, 2005, Non-linear data continuation with redundant frames: Presented at the CSEG National Convention.

———, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International. (In press).

Herrmann, F. J., D. Wang, and G. Hennenfent, 2007b, Multiple prediction from incomplete data with the focused curvelet transform: Presented at the SEG International Exposition and 77^{th} Annual Meeting.

Leneman, O., 1966, Random sampling of random processes: Impulse response: Information and Control, **9**, no. 4, 347 – 363.

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.

Spitz, S., 1991, Seismic trace interpolation in the F-X domain: Geophysics, **56**, no. 6, 785 – 794.

Stolt, R. H., 2002, Seismic data mapping and reconstruction: Geophysics, **67**, no. 3, 890 – 908.

Thomson, D., G. Hennenfent, H. Modzelewski, and F. J. Herrmann, 2006, A parallel windowed fast discrete curvelet transform applied to seismic processing: Presented at the SEG International Exposition and 76^{th} Annual Meeting.

Trad, D. O., T. J. Ulrych, and M. D. Sacchi, 2003, Latest view of sparse Radon transforms: Geophysics, **68**, no. 1, 386–399.

van den Berg, E. and M. P. Friedlander, 2007, In pursuit of a root: Technical report, UBC Computer Science Department. (TR-2007-16. http://www.optimization-online.org/DB_FILE/2007/06/1708.pdf).

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, no. 9, 1166 – 1177.

Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, no. 4, V87 – V95.

Zwartjes, P. M. and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased data: Geophysics, **72**, no. 1, V21–V32.

Chapter 2

Seismic denoising with non-uniformly sampled curvelets

2.1 Introduction

Recently introduced curvelets are amongst the latest members of a growing family of multiscale, and now also multidirectional, data expansions (Candès and Donoho, 2004; Candès et al., 2006). The primary aim of these expansions, with respect to a collection of prototype features, is to find a sparse representation for the data. A signal representation is sparse when it is capable of capturing the signal as a superposition of a small number of components. The sparser and the more generic the transformation, the more successful the signal separation.

So what makes the curvelet decomposition an appropriate transform for seismic data processing, and why generalize this transform to non-uniformly sampled data? To answer these questions, let us first describe what seismic data is. Seismic data volumes are recordings of the amplitudes of transient waves at the Earth's surface. These waves are either caused by man-made sources or by naturally occuring earthquakes. Each source and receiver pair generates a trace which is a function of time. A seismic dataset is the collection of these traces. All these traces together provide a spatio-temporal sampling of the wavefield which contains different arrivals that correspond to different interactions of the incident wave field with inhomogeneities in the Earth's subsurface. A common denominator amongst these arrivals is that they represent wavefronts. The main characteristic of a wavefront is its relative smoothness in the direction along the wavefront and its relative

A version of this chapter has been published. G. Hennenfent and F.J. Herrmannn. Seismic denoising with non-uniformly sampled curvelets. *Computing in Science and Engineering*, 8(3), May-June 2006.

oscillatory behavior in the normal direction.

By virtue of their anisotropic shape, curvelets are well adapted to detect wavefronts because aligned curvelets locally correlate well with the wavefront. In that sense, curvelets act as multiscale surf boards riding the incoming wavefronts. However, limitations on data acquisition regarding the positions of the sources and receivers put restrictions on the spatial sampling of seismic wave fields. For instance, in land acquistion for seismic exploration, there are obstacles such as buildings and lakes while in passive seismology there is no control over the source position. Earthquakes occur irregularly along major plate boundaries.

The current implementation of the FDCT assumes a regular sampling along all axes. If we ignore the non-uniformity of spatial sampling, we can no longer expect to detect wavefronts because of lack of continuity. We address this issue by extending the FDCT to non-uniformly sampled data. Through this extension, we are able to not only detect wavefronts in noise but also bring the data to a regular grid in case each grid point contains at least one datum. The example given in Fig. 2.1 clearly illustrates how continuity along wavefronts is destroyed when casting non-uniformly sampled data to a regular grid but is restored when dealing appropriately with the data. Our denoising and binning algorithm is based on this extension and exploits the sparsity of seismic data in the curvelet domain through a nonlinear thresholding on the curvelet coefficients. The term binning refers to interpolation towards a regular grid for the case where the number of irregular samples exceeds the size of the regular grid.

The paper is organized as follows. First, we give a brief overview of curvelets. We demonstrate their sparseness on seismic data, which determines the performance of our denoising. Second, we describe our non-uniform extension to the curvelet transform by the Non-equally sampled Fast Fourier Transform (NFFT, Kunis and Potts, 2003). We show that this extension restores the performance of the transform for non-uniformly sampled data. Third, we introduce a denoising and binning algorithm by nonlinear shrinkage on the curvelet coefficients. We conclude by showing applications to synthetic seismic data.

2.2 The curvelet transform

2.2.1 Main properties

Since their introduction, curvelet transforms (see e.g. Candès et al., 2006, and the references therein) have received increasing interest in the seismic

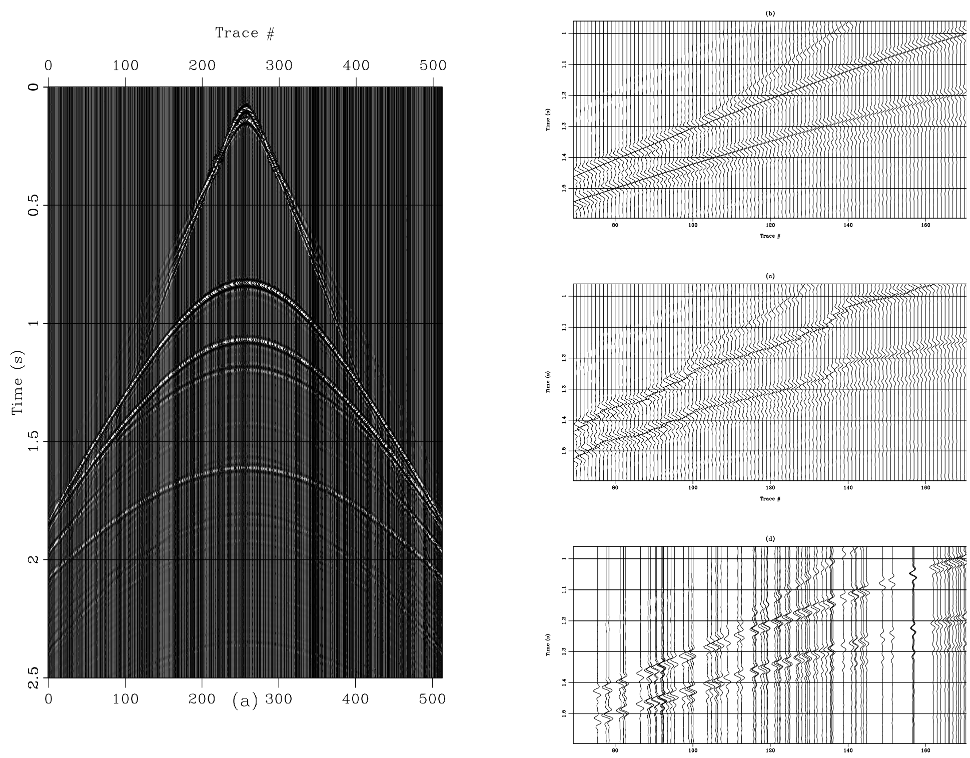


Figure 2.1: Example of synthetic seismic data. **(a)** uniformly (grey-scale plot) and non-uniformly sampled (wiggle trace plot); **(b)** windowed regular sampled data; **(c)** windowed irregular sampled data cast to a regular grid and **(d)** windowed data on the non-uniformly sampled grid. Notice the continuity along the arriving wavefront in **(b)** and **(d)**. Recasting irregular data onto a regular grid destroys the continuity. In this example, the irregularity of the non-uniformly sampled grid has been exaggerated. In this paper, we will only deal with non-uniformly sampled grids with at least one sample for each grid point of the regular binning grid.

research community. The capability of curvelets to detect wavefronts is mainly responsible, and it comes as no surprise that their original construction, through the so-called second dyadic partitioning, came from the field of Harmonic Analysis (Smith, 1998), where curvelets were introduced as expansions for asymptotic solutions of wave equations. This connection has well been recognized by the developers of the FDCTs (Candès et al., 2006) and has resulted in important contributions not only to the compression of Green's functions (Candès and Demanet, 2003), but also to nonlinear approximations of functions with intermittent regularity (Candès and Donoho, 2004). These functions are assumed to be piece-wise smooth with singularities, regions where the derivative diverges, on piece-wise smooth curves. In the Earth, these singularities correspond to geologic unconformities at which waves reflect. In seismic data, these singularities correspond to wavefronts. Geologic boundaries as well as wavefronts contain points of intermittent regularity such as faults or pinch outs along sedimentary layers, or caustics in wavefronts.

The purpose of this paper is not to compress operators. Instead, we are interested in separating different seismic data components which, except for possible incoherent measurement noise, consist of components that are the solution of a wave equation. For this purpose, we employ the curvelet transform as a vehicle that

- is rich enough to account for the multiscale and multidirectional properties of seismic data with intermittent regularity;
- is local in phase space, the space spanned by space and spatial frequency;
- exploits smoothness along, and oscillatory behavior across, the arriving wavefronts;
- differentiates between different signal components on the basis of location, angle and frequency content;
- obtains fast decay of nonlinear approximation error for seismic data;
- permits a fast ($\mathcal{O}(K \log K)$ with K the data size) multi-dimensional (2-/3-D) implementation.

As can be seen in Fig.'s 2.2 and 2.3, curvelets are local in both space and spatial frequency and correspond to a partitioning of the $2-D$ Fourier plane by highly anisotropic elements (for the high frequencies) that obey the paramount parabolic scaling principle (Smith, 1998) width $\propto$ length^2. As opposed to discrete wavelets, designed to provide sparse representations of functions with point singularities, curvelets provide sparse representations for functions with singularities on curves. Moreover, whereas multiscale wavelets consist of a collection of location- and scale-indexed basis functions, curvelets represent a family of functions made out of translations, rotations and parabolic scalings. As such, a frame with moderate redundancy is created. The elements in this transform, which we will call prototype waveforms, are

- *multiscale* with frequency support on dyadic coronae in the 2-D Fourier plane;
- *multidirectional* with angles that correspond to the centers of the wedges (for every other resolution doubling, the number of angles doubles);
- *anisotropic*, obeying the scaling law width $\propto$ length^2;
- *local* allowing for thresholding which locally adapts to the non-stationary signal.

Frames differ from orthonormal bases. Orthonormal transforms (orthonormal matrices) compose an arbitrary finite-energy discretized signal vector $\mathbf{f} \in \mathbb{R}^K$ of length K ($\mathbf{f}$ is a discretization of the multivariate function $f(s,t) : \mathbb{R}^2 \mapsto \mathbb{R}$) according to

$$\mathbf{f} = \mathbf{B}^{-1}\mathbf{B}\mathbf{f} = \mathbf{B}^H\mathbf{B}\mathbf{f} := \sum_{m\in\mathcal{M}} \langle \mathbf{f},\, \varphi_m \rangle \varphi_m, \tag{2.1}$$

with $\mathbf{B}^H$ the matrix adjoint of the decomposition matrix $\mathbf{B}$, and the brackets $\langle\rangle$ denoting the standard discrete inner product $\langle \mathbf{f},\, \varphi_m \rangle = \mathbf{f}^H \varphi_m$ of $\mathbf{f}$ with the m^{th} column vector of $\mathbf{B}^H$. Because $\mathbf{B}$ is an orthonormal basis, its adjoint matrix corresponds to its inverse (inverse transform). The summation in Eq. 2.1 runs over the index set $\mathcal{M}$ of size $M = K$. As opposed to orthonormal transforms, redundant frame expansions decompose a length K signal into a frame expansion with $M > K$ elements. Consequently, the composition matrix is rectangular with the number of columns exceeding the number of rows.

The regularly-sampled FDCT is a frame represented by the matrix $\mathbf{C}$. Applying this matrix to a vector $\mathbf{f}$ creates a multi-index coefficient vector $\mathbf{x} = \mathbf{C}\mathbf{f}$ with $\mathbf{x} := \{x_m\}_{m\in\mathcal{M}}$ with the multi-index m running over the locations, orientations and scales (see Candès et al., 2006, for detail in the discrete constructions of the FDCT). We choose the numerically tight FDCT via wrapping as our curvelet transform. For this transform, the pseudo-inverse (denoted by the symbol †) equals the adjoint and we have $\mathbf{f} = \mathbf{C}^{\dagger}\mathbf{x} = \mathbf{C}^{H}\mathbf{x}$ implying $\mathbf{C}^{H}\mathbf{C} = \mathbf{I}$.

2.2.2 Nonlinear approximation rates

The nonlinear approximation rate expresses the asymptotic decay of the ℓ_2-difference between the original data and the partial reconstruction from the largest M coefficients. In dimension two, Fourier only attains an asymptotic decay rate of $\mathcal{O}(M^{-1/2})$ for data consisting of twice-differentiable functions with singularities on piece-wise twice differentiable curves while curvelets asymptotically obtain the optimal rate $\mathcal{O}(M^{-2})$ ignoring log-like factors. Even though wavelets improve upon Fourier, their approximation rate of $\mathcal{O}(M^{-1})$ is sub optimal.

By virtue of their multiscale and multidirectional construction, curvelets sparsely represent seismic data. Not only do individual curvelets capture the main characteristics of wavefronts locally – they look like little waves – but they also jointly capture the seismic energy effectively. This performance can be observed in Fig. 2.4 where the nonlinear approximation rates are shown for representative seismic synthetic data. The rates are computed for each of the following cases: curvelets and wavelets on regularly-sampled data; curvelets on non-uniformly sampled data (treated as uniformly sampled data); and our extension of the curvelet transform on non-uniformly sampled data. For uniformly sampled data, the nonlinear approximation rate of curvelets outperforms the Daubechies 6 wavelet by a wide margin. This figure also shows the importance of treating non-uniformly sampled data correctly in the curvelet transform. For instance, treating non-uniformly sampled data as uniformly sampled seriously deteriorates the performance.

To address the non-uniformly sampled data issue, binning is used to bring non-uniformly sampled data to the regular grid. To compare the reconstructions, we use space-domain linear interpolation for wavelets and we include NFFT binning as our extension to the FDCT. Fig. 2.5 shows reconstructions for non-uniformly sampled data with binning for only 1 % of the coefficients. The partial reconstruction with the non-uniformly sampled curvelet transform performs nearly as well as the uniformly sampled transform and outperforms the wavelets. Detailed measures on the performance are listed in Table 2.1.

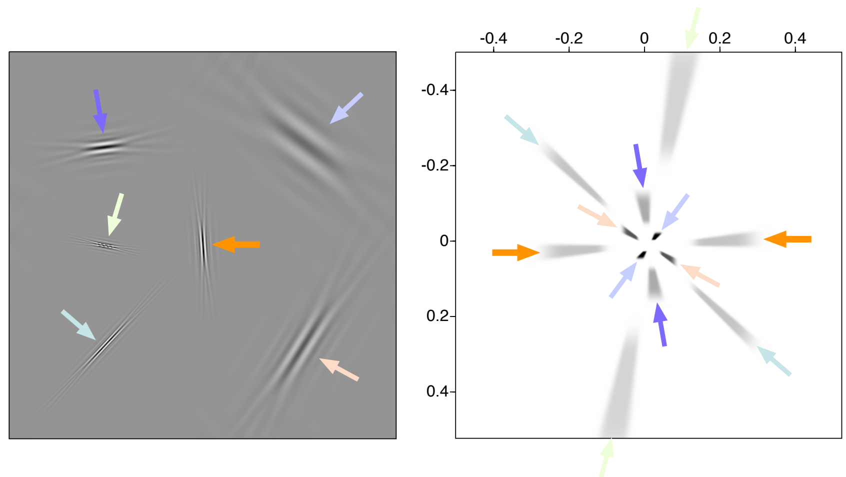


Figure 2.2: Spatial (left) and frequency (right) viewpoints of six real curvelets at different scales and angles. As opposed to complex curvelets, real curvelets live in two angular wedges symmetric about the origin. Comparison of the curvelets in the two domains also shows their micro-local correspondence (Candès and Donoho, 2002), relating the orientation of curvelets in both domains. Because of their rapid decay in the physical space and compact support in the Fourier space, curvelets localize in phase space.

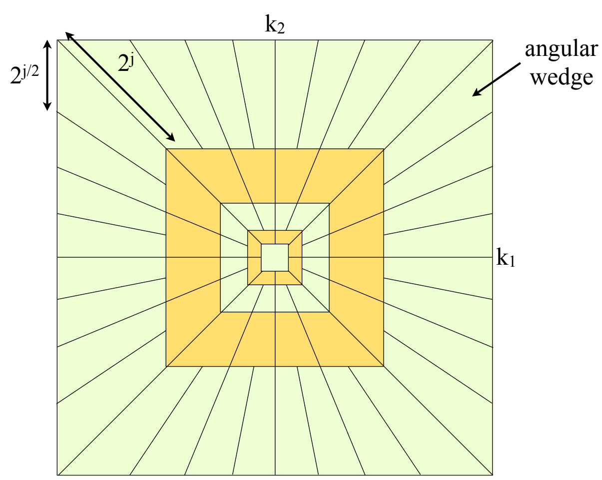


Figure 2.3: Discrete curvelet partitioning of the 2-D Fourier plane into second dyadic coronae and sub-partitioning of the coronae into angular wedges.

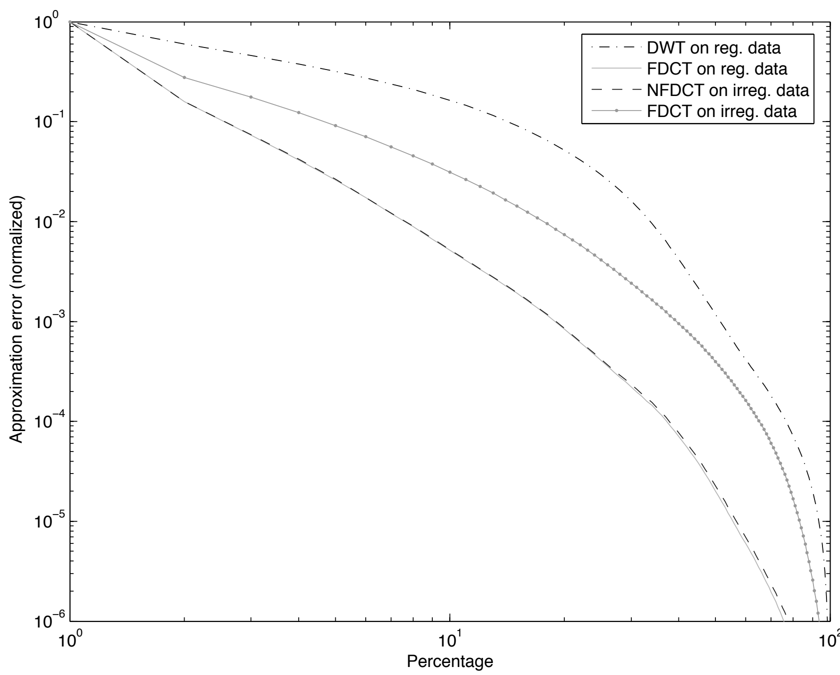


Figure 2.4: Decays of the nonlinear approximation error for (non-uniformly sampled) curvelet transform (N)FDCT and discrete wavelet transform (DWT) using Daubechies 6 on (ir)regularly sampled synthetic seismic data. Curvelets on the regular grid (plain line) clearly outperform discrete wavelets (alternated dash-dot line). Our extension of the curvelet transform for non-uniformly sampled data (dashed line) retains the performance of the regularly-sampled curvelet transform on uniformly-sampled data, as opposed to the inferior performance obtained when irregular data is treated as regular (line with dots).

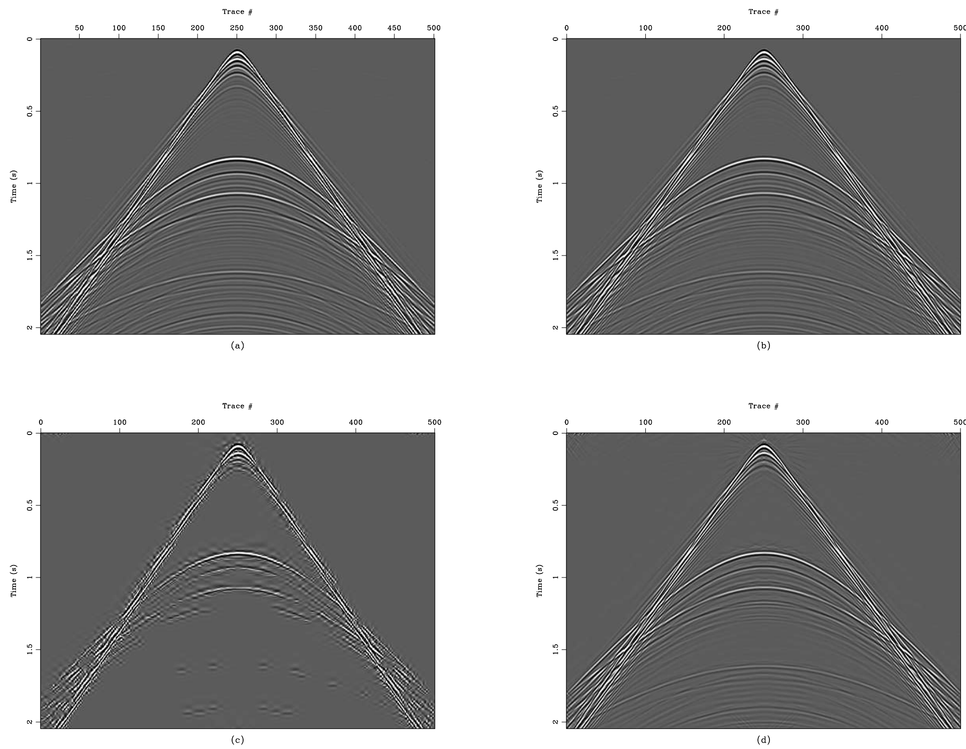


Figure 2.5: Partial reconstructions using 1 % of the wavelet and curvelet coefficients for non-uniformly sampled data. **(a)** linear binning; **(b)** curvelet binning; **(c)** reconstruction of (a) with 1 % of the wavelet coefficients; **(d)** reconstruction of (b) with 1 % of the curvelet coefficients. Visual comparison between the wavelet and curvelet partial reconstructions shows a drastic improvement with the curvelets. This improvement on wavelets is consistent with the nonlinear approximation rates. The numbers listed in Table 2.1 also show improvement for the binning with the NFFT'ed curvelets defined below even though (a) and (b) are visually similar.

	SNR (dB)
Linear binning	-1.96
NFFT binning	9.04
Denoising	13.35
NFFT binning and denoising	8

Table 2.1: Binning (Fig.'s 2.5a and 2.5b) and denoising errors measured by 'signal-to-noise ratio' (SNR) defined as $10 \log_{10} \frac{\|\mathbf{f}\|_2^2}{\|\mathbf{f}-\tilde{\mathbf{f}}\|_2^2}$, with $\mathbf{f}$ the original function and $\tilde{\mathbf{f}}$ its estimate after binning and/or denoising. The SNR is 0 dB for the initial (non-uniformly) noisy data. Notice that, even for this bad SNR, we only lose 1 dB between noise-free NFFT binning and noisy NFFT binning combined with denoising.

2.3 The NFDCT: a curvelet frame for seismic processing

As shown in Fig. 2.4, the performance of curvelet approximations and hence signal separation may seriously deteriorate when non-uniformly sampled data is treated as regular. Because seismic data is more often than not acquired irregularly, failure to account for non-uniformly sampled data may have adverse effects on seismic imaging. The main purpose of this paper is to extend the FDCT towards non-uniformly sampled grids. The FDCT $\mathbf{C}$ on an arbitrary uniformly sampled vector $\mathbf{f}$ factors as $\mathbf{T}$ times $\mathbf{F}$, with $\mathbf{F}$ the orthonormal Fourier transform and $\mathbf{T}$ the curvelet tiling matrix (i.e. $\mathbf{Cf} := \mathbf{TFf}$). Below we replace the ordinary Fourier transform with its non-uniformly sampled counterpart, which is a natural choice since the curvelet construction is defined in the Fourier domain.

From this point on, non-uniformly sampled N-vectors $\underline{\mathbf{f}} \in \mathbb{R}^N$ are denoted by the underbar, and $\underline{\mathbf{f}} := \{f(x_p)\}_{p=1,\cdots,N}$ at the nodes $x_p \in \mathcal{X}$ where $\mathcal{X} := \{x_p = (s_p, t_p) \in \mathbb{R} \times \mathbb{N} : \ -1/2 \leq s_p < 1/2 \quad \text{and} \quad 0 \leq t_p < N_t\}_{p=1,\cdots,N}$, with N the total number of nodes and N_t the number of regular time samples. We consider the number of source/receiver positions larger than the size of the corresponding regular spatial grid.

At the heart of non-equally sampled Fourier transforms of bandwidth limited functions lies the fast evaluation of the following sum (see e.g. Beylkin, 1995; Kunis and Potts, 2003)

$$\underline{\mathbf{f}} := f(x_p) = \sum_{k \in \mathcal{K}} \hat{f}_k e^{-2\pi i k x_p} \quad \text{for} \quad p = 1, \cdots, N. \tag{2.2}$$

This expression corresponds to the discrete inverse Fourier transform from a uniformly-sampled grid $\mathcal{K} := \{k_j = (k_j^s, k_j^t) \in \mathbb{Z}^2 : -K_{s,t}/2 \leq k_j^{s,t} < K_{s,t}/2\}_{j=1,\cdots,K}$ in the Fourier domain (denoted by the symbol), consisting of $K = K_s \times K_t$ samples with $K_t = N_t$, towards the non-uniformly sampled grid $\mathcal{X}$. In matrix-vector notation the above expression becomes

$$\underline{\mathbf{f}} = \mathbf{A}\hat{\mathbf{f}}. \tag{2.3}$$

The NFFT is an implementation that approximately evaluates the above sum with a fast algorithm based on ideas from (Beylkin, 1995). By replacing the regular FFT in the implementation for the FDCT by the pseudo-inverse of the NFFT, we arrive at a transform that takes irregularly sampled data to the regularly sampled Fourier domain.

By limiting the maximum distance between the nodes to K_s^{-1} and having more irregular than regular samples ($N > K$), the pseudo inverse of $\mathbf{A}$ is well conditioned when including an additional diagonal weighting $\mathbf{W}$, proportional to the number of source/receivers per unit on the interval (Kunis and Potts, 2003). The forward non-uniformly sampled fast discrete curvelet transform, NFDCT, is now defined as

$$\mathbf{x} = \mathbf{C}\underline{\mathbf{f}} := \mathbf{T}\mathbf{A}^{\dagger}\underline{\mathbf{f}} \tag{2.4}$$

with $\mathbf{A}^{\dagger} := (\mathbf{A}^H\mathbf{W}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{W}$. Under the above irregular sampling conditions, the non-uniformly sampled forward curvelet transform produces curvelet coefficients that pertain to a regular Fourier grid. Hence, by applying the regular inverse curvelet transform to these curvelet coefficients yields data on the regular grid. This process corresponds to a NFFT-based binning.

2.4 Signal estimation and separation by thresholding

The success of denoising and signal separation depends largely on the ability of a transform to sparsely represent a particular type of image. Discrete wavelet transforms and more recently curvelets accomplish (near) optimal nonlinear approximation rates for certain classes of images (see e.g. Donoho and Johnstone, 1998). As argued before, curvelets appear to be the appropriate choice for seismic data. We discuss estimation techniques both for orthonormal wavelets and overcomplete curvelets.

Denoising by shrinkage Thresholding on the coefficients of an expansion with respect to a collection of prototype waveforms is a key component in the solution of denoising problems with the signal model given by

$$\mathbf{d} = \mathbf{m} + \mathbf{n} \tag{2.5}$$

with $\mathbf{m}$ the unknown deterministic signal component and $\mathbf{n}$ zero-centered white Gaussian noise with standard deviation σ. The Gaussian assumption is fundamental in this work. Whiteness, however, is not a prerequisite.

Soft thresholding on each element of the noisy data coefficient vector solves for the model $\mathbf{m}$ through

$$\tilde{\mathbf{m}} = \mathbf{S}^{\dagger} S_{\mathbf{w}} \left(\mathbf{S}\mathbf{d}\right). \tag{2.6}$$

In this expression, $\mathbf{S}$ stands for an arbitrary sparse signal expansion and S_w for soft thresholding defined element-wise as

$$S_w(x) := \begin{cases} x - \operatorname{sign}(x) w & |x| \geq w \\ 0 & |x| < w \end{cases} \tag{2.7}$$

with $w \geq 0$ a real-valued threshold. The vector $\mathbf{w}$ contains the thresholds for each coefficient. This shrinkage operation by thresholding forms the basis for our denoising and signal separation. In Fig. 2.6 we illustrate the estimation by shrinkage as described in (2.6).

Denoising with orthonormal bases: For arbitrary orthonormal transforms $\mathbf{S} := \mathbf{B}$, we have $\mathbf{S}^{\dagger} = \mathbf{B}^{-1}$ and Eq. (2.6) solves the following minimization problem

$$\tilde{\mathbf{x}} = \arg\min_{x} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \|\mathbf{x}\|_{1,\mathbf{w}} \tag{2.8}$$

with $\{\mathbf{y}, \mathbf{x}\} := \{\mathbf{Bd}, \mathbf{Bm}\}$ the transformed coefficients and $\|\mathbf{x}\|_{1,\mathbf{w}}$ a weighted ℓ_1-penalty functional given by

$$\|\mathbf{x}\|_{1,\mathbf{w}} = \sum_{m \in \mathcal{M}} w_m |x_m|. \tag{2.9}$$

By setting each weight $w_m = 3 \cdot \sigma$, Eq. 2.6 yields an estimate for $\mathbf{m}$. This threshold corresponds to the typical rule for thresholding (see e.g. Mallat, 1998). During this estimation, the quadratic mismatch between the data and model is minimized jointly with the weighted ℓ_1-penalty functional.

The quadratic term is known as the log-likelihood. The model is assumed to be a superposition of prototype waveforms with coefficients drawn independently from a probability function $Pr\{x_m\} \propto \exp\left(-\mathrm{Const}\cdot\lambda|x_m|\right)$ that corresponds to the Laplace distribution which enhances sparsity Starck et al. (2004); Elad (2006).

Denoising with tight frames: The FDCT with wrapping is a tight frame with a synthesis matrix $\mathbf{C}^{\dagger} = \mathbf{C}^{H}$ that has more columns than rows. The coefficient vector exceeds the data size by a factor of roughly 8. In this case, $\mathbf{C}\mathbf{C}^{H} \neq \mathbf{I}$ and Eq. 2.6 is no longer equivalent to the minimization problem in Eq. 2.8. However, for a tight frame with a ℓ_2-norm for the columns of the synthesis matrix close to unity, shrinkage still provides a good approximation to the solution of the above minimization problem (Elad, 2006).

Denoising and binning with the NFDCT By combining the non-uniformly sampled curvelet transform with shrinkage (cf. Eq. 3.7), we arrive at our main result

$$\tilde{\mathbf{m}} = \mathbf{C}^{\dagger} S_{\mathbf{w}}\left(\mathbf{C}\underline{\mathbf{d}}\right), \tag{2.10}$$

accomplishing the joint task of (in)-coherent signal separation on non-uniformly sampled data $\underline{\mathbf{d}}$ and binning. In this expression, the non-uniform data vector $\underline{\mathbf{d}}$ is curvelet transformed with the NFDCT, followed by a thresholding and the regular inverse curvelet transform (FDCT). Under the assumptions we stated before on the bandwidth-limitation of the signal and the unequal sampling, the pseudo-inverse used to Fourier transform the unequally-sampled points can be computed stably. As such, we can safely assume that the regular sampled Fourier data is still close to the Fourier transform of the corresponding uniformly-sampled data. We proceed as if we were dealing with the uniformly sampled case by thresholding and applying the uniformly sampled inverse curvelet transform (IFDCT). The result of this operation is a combined denoising and binning, where irregular bandwidth-limited noisy data is denoised and mapped to a regular grid. This technique is demonstrated in Fig's. 2.7 and 2.8 discussed below.

Coherent signal separation Even though thresholding estimators are primarily used to separate incoherent random from deterministic signal components, extending the thresholding estimations to cases where there are two coherent signal components has been quite successful for cases where there exists a prediction for one of the signal components (this is the case for e.g. primary-multiple separation in seismic exploration, Herrmann et al., 2007).

In this case the signal model becomes slightly more complicated

$$\mathbf{s} = \mathbf{s}_1 + \mathbf{s}_2 + \mathbf{n}, \tag{2.11}$$

with $\mathbf{s}_1$, $\mathbf{s}_2$ the two coherent signal components. Given a prediction $\breve{\mathbf{s}}_2$ for the second component, the first component can be estimated through Eq. 2.6 where the weighting is defined as

$$\mathbf{w} := \max\left(3\sigma,\, \delta|\breve{\mathbf{x}}_2|\right) \tag{2.12}$$

with $\breve{\mathbf{x}}_2 := \mathbf{C}\breve{\mathbf{s}}_2$. This weighting corresponds to a varying threshold defined in terms of the curvelet transform for the predicted signal component. The δ expresses the confidence in the prediction. The above estimator again corresponds to a maximum *a-posteriori* (MAP) estimator minimizing the log-likelihood function with coefficients that are selected from a cross-correlation weighted probability function $Pr\{x_m\} \propto \exp\left(-\mathrm{Const} \cdot w_m|x_m|\right)$ for $m \in \mathcal{M}$. This probability function is weighted by the prediction for the second signal component.

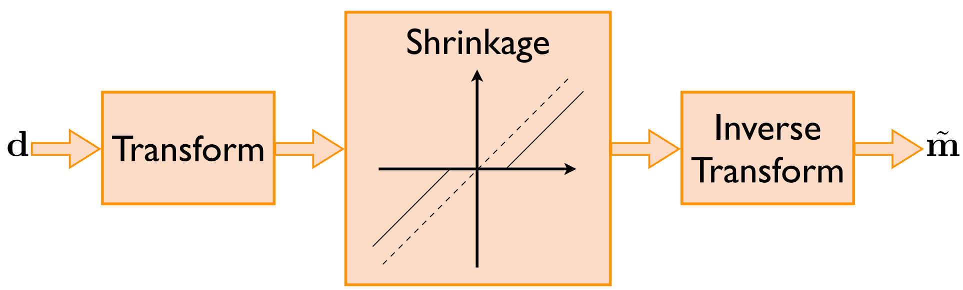


Figure 2.6: 3-step estimation by shrinkage on transformed domain coefficients. Noisy data $\mathbf{d}$ is brought to a transformed domain. Soft thresholding is applied on the coefficients. Finally a denoised estimate $\tilde{\mathbf{m}}$ is obtained by applying the corresponding inverse transform to the thresholded coefficients.

2.4.1 Applications to seismic data

Amongst the striking features of seismic data is that it contains wavefronts possibly contaminated with bandwidth limited Gaussian noise. As shown above, removal of this random component can be accomplished by

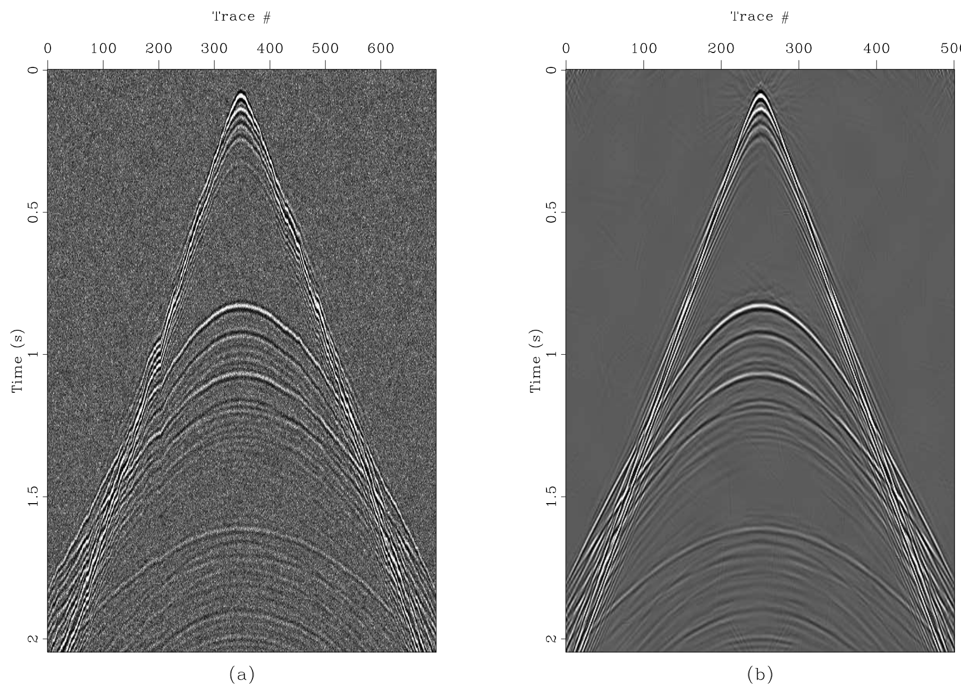


Figure 2.7: Incoherent noise removal through shrinkage (cf. Eq.'s 2.6 and 2.10). **(a)** noisy non-uniformly sampled data plotted in a regular grid and with SNR of 0 dB; **(b)** denoised data including binning (see Eq. 2.10). Notice the significant improvement reflected into the SNR listed in Table 2.1.

forward transforming the (ir)regular data with the (N)FDCT, followed by a simple shrinkage and reconstruction with the IFDCT. Fig. 2.7 illustrates the performance of curvelet denoising by shrinkage. Because the performance of denoising is, besides the binning error, as good as regular denoising, we only show the results for non-uniformly sampled data where denoising is combined with binning. These methods can be extended to the case of coherent signal removal according to the threshold defined in Eq. 2.12. To emphasize the added value of the NFDCT, we include an example where the signal separation is carried out on irregular data cast into a regular grid and on the irregular data itself with the NFDCT.

The removal of ghost events related to multiple interactions of the wavefield with the surface is paramount to the success of seismic imaging based on linearized inverse scattering. These ghosts, also known as multiples, violate the linearization and cause artifacts in the image. Removing these artifacts has proven to be difficult due to the multiple prediction error. Adaptive subtraction techniques based on matched filtering (see e.g. Verschuur et al., 1992) have been developed to counter the inaccuracies and robustly separate the two signal components. Unfortunately, matched filtering suffers from inadvertent removal of primary energy and an unwanted remainder of multiple energy. By formulating this signal separation problem as a weighted shrinkage in the curvelet domain, good results have been obtained as illustrated in Fig. 2.8. These results were obtained using $\tilde{\mathbf{s}}_1 = \mathbf{C}^{\dagger} S_{\mathbf{w}}\left(\mathbf{C}\underline{\mathbf{s}}\right)$ where the weights $\mathbf{w}$ are defined as in Eq. 2.12 with $\breve{\mathbf{s}}_2$ the modeled/predicted multiples. The constants were set to $\delta = 1.6$ and σ according to the noise level. The predicted multiples are left as is. By virtue of the NFDCT, the result for the non-uniformly sampled case is almost as good as the result for the uniformly sampled case.

2.5 Conclusions

In this paper, we demonstrated that curvelet transforms sparsely represent uniformly-sampled seismic data. This property was used to perform denoising and coherent signal separation, including the elimination of multiple reflection events. We also demonstrated that the performance of the curvelet transform is restored by our curvelet transform for non-uniformly sampled data: the NFDCT. Application of this transform to noise removal and signal separation problems on irregular data shows that we recover the performance of the curvelet transform on regular data up to the binning error. The binning error can be controlled at the expense of more computa-

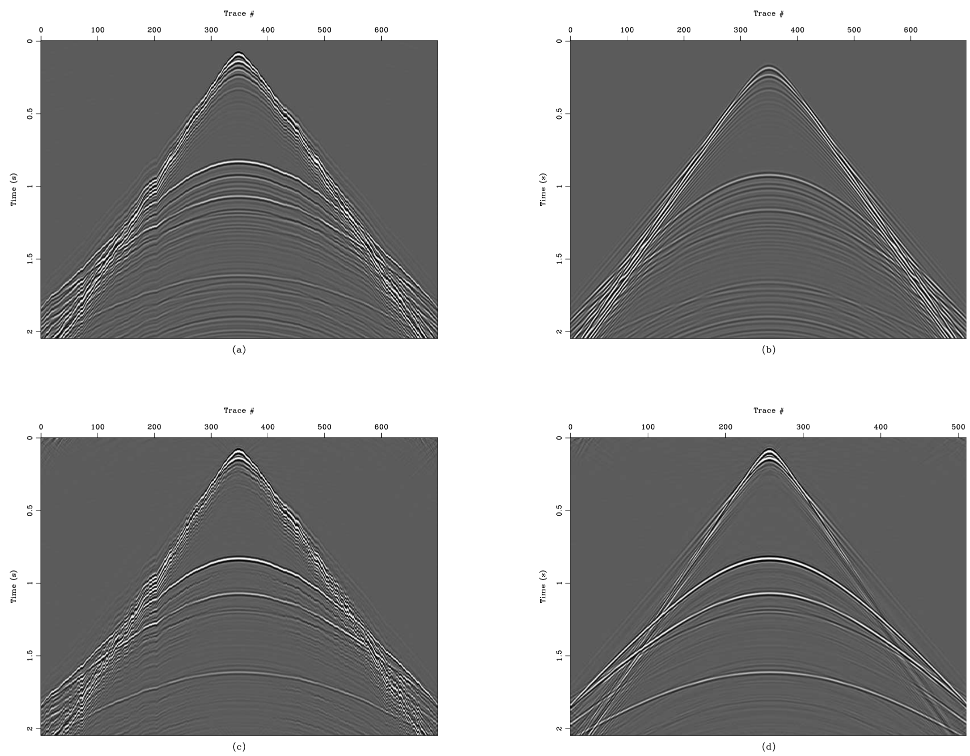


Figure 2.8: Removal of ghost events related to multiple interactions of the wavefield with the surface. **(a)** synthetic non-uniformly sampled data containing primary and multiple reflections treated as regular data; **(b)** predicted multiples; **(c)** estimated primaries using the FDCT on **(a)** and weights as defined in Eq. 2.12; **(d)** estimated primaries using the NFDCT on **(a)** and weights as defined in Eq. 2.12. By virtue of the NFDCT, the result for the non-uniformly sampled case rivals the result for the uniformly sampled case.

tions. In a future paper, we hope to report on an extension of our method to the case where the size of the interpolation grid exceeds the number of unequally sampled data points.

2.6 Acknowledgements

We would like to thank the authors of the FDCT (Candès et al., 2006) and the NFFT (Kunis and Potts, 2003). We also would like to thank Colin Russell for his coding. This work was carried out as part of the SINBAD project with financial support, secured through ITF (the Industry Technology Facilitator), from the following organizations: BG Group, BP, Chevron, ExxonMobil and SHELL. Additional funding came from the NSERC Discovery Grant 22R81254 and from POTSI funded through MITACS.

Bibliography

Beylkin, G., 1995, On the fast Fourier transforms of functions with singularities: Applied and Computational Harmonic Analysis, **2**, no. 4, 363–381.

Candès, E. J. and L. Demanet, 2003, Curvelets and Fourier Integral Operators: Compte Rendus de l'Académie des Sciences, Paris, **336**, no. 1, 395 – 398.

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

Candès, E. J. and D. L. Donoho, 2002, Recovering edges in ill-posed problems: optimality of curvelet frames: Annals of Statistics, **30**, no. 3, 784 – 842.

———, 2004, New tight frames of curvelets and optimal representations of objects with C^2 singularities: Communications on Pure and Applied Mathematics, **57**, no. 2, 219 – 266.

Donoho, D. L. and I. M. Johnstone, 1998, Minimax estimation via wavelet shrinkage: Annals of Statistics, **26**, no. 3, 879 – 921.

Elad, M., 2006, Why simple shrinkage is still relevant for redundant representations?: Transactions on Information Theory, **52**, no. 12, 5559 – 5569.

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007, Nonlinear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, no. 2, 781–799.

Kunis, S. and D. Potts, 2003, Nonequispaced discrete Fourier transform (NFFT): software. (Available at `http://www-user.tu-chemnitz.de/~potts/nfft/`).

Mallat, S. G., 1998, A wavelet tour of signal processing: Academic Press.

Smith, H. F., 1998, A Hardy space for Fourier integral operators: Journal of Geometric Analysis, **7**, no. 4, 629 – 653.

Starck, J.-L., M. Elad, and D. L. Donoho, 2004, Redundant multiscale transforms and their application for morphological component analysis: Journal of Advances in Imaging and Electron Physics, **132**.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, no. 9, 1166 – 1177.

Chapter 3

Non-parametric seismic data recovery with curvelet frames

3.1 Introduction

The methodology presented in this paper addresses two important issues in seismic data acquisition, namely the mediation of imaging artifacts caused by physical constraints encountered during acquisition, and the design of a more economic acquisition, limiting the number of source and receiver positions within the survey. In either case, the data is incomplete and it is our task to recover a fully-sampled seismic data volume as required by wave-equation based multiple elimination (SRME, Verschuur and Berkhout, 1997) and imaging (Symes, 2006). This paper deals with the specific case of seismic data recovery from a regularly-sampled grid with traces missing. As a consequence, the data is undersampled and the Nyquist sampling criterion is violated, giving rise to a Fourier spectrum that may contain harmful aliases.

A multitude of solutions have been proposed to mitigate the impact of coherent aliases on seismic imaging. Our approach derives from three key ingredients, namely a sparsifying transform, a sampling strategy that limits the occurrence of harmful aliases and a nonlinear recovery scheme that promotes transform-domain sparsity and consistency with the acquired data. These three key ingredients form the basis of the emerging field of "compressive sampling" (Candès et al., 2006b; Donoho et al., 2006b) with several applications that include MRI-imaging (Lustig et al., 2007) and A/D conversion (Tropp et al., 2006). Compressive sampling can be seen as a

A version of this chapter has been accepted for publication. F.J. Herrmann and G. Hennenfent. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173:233-248, 2008.

theoretically rigorous justification of empirical ideas on sparsity-promoting inversion that existed in the geophysical literature with applications that include "spiky deconvolution" (Taylor et al., 1979; Oldenburg et al., 1981; Ulrych and Walker, 1982; Levy et al., 1988; Sacchi et al., 1994) analyzed by mathematicians (Santosa and Symes, 1986; Donoho and Logan, 1992) to Fourier and Radon transform-based seismic data recovery, an approach initially proposed by Sacchi et al. (1998) and extended by numerous authors (Trad et al., 2003; Xu et al., 2005; Abma and Kabir, 2006; Zwartjes and Sacchi, 2007). Amongst all these methods, it was observed that a successful solution of these problems depends critically on the number of measurements (or the frequency passband for deconvolution) and the signal's sparsity in some transformed domain, e.g. spikes for deconvolution and Fourier for sparse recovery.

Compressive sampling provides insights into the conditions that determine successful recovery from incomplete data. We leverage these new insights towards a formulation of the large-scale seismic data regularization problem, where a sparsifying transform, anti-alias sampling and a sparsity-promoting solver are used to solve this problem for acquisitions with large percentages of traces missing. These theoretical developments are important since they provide a better intuition of the overriding principles that go into the design of a recovery method and into explicit construction of a sparsifying transform, the sampling strategy and the sparsity-promoting solver.

In this paper, we consider a recovery method that derives from this intuition by using a generic sparsifying transform that requires minimal prior information (although our method benefits like Fourier-based interpolation (Zwartjes and Sacchi, 2007) from dip discrimination by means of specifying a minimum apparent velocity). In that respect our method differs from interpolation methods based on pattern recognition (Spitz, 1999), plane-wave destruction (Fomel et al., 2002) and data mapping (Stolt, 2002), including parabolic, apex-shifted Radon and DMO-NMO/AMO (Trad, 2003; Trad et al., 2003; Harlan et al., 1984; Hale, 1995; Canning and Gardner, 1996; Bleistein et al., 2001; Fomel, 2003; Malcolm et al., 2005), which require, respectively, the omission of surface waves, specific knowledge on the dominant dips and a velocity model.

3.1.1 Our main contribution

The success of our recovery method for seismic data, named curvelet-based recovery by sparsity-promoting inversion (CRSI), derives from a sparsifying transform in conjunction with a sampling scheme that favors recovery. With their well-documented sparsity for seismic data with wavefronts and Fourier-domain localization property (Candès et al., 2006a; Hennenfent and Herrmann, 2006; Herrmann et al., 2007a), curvelets render sparsity-promoting inversion into a powerful constraint for the recovery of seismic data. Our contribution, first reported in Herrmann (2005), lies in the application of this transform (see e.g. Candès et al., 2006a; Ying et al., 2005, for details on the definition and implementation of the discrete curvelet transform) to the seismic recovery problem. Our work includes the adaptation towards a geophysically feasible sampling scheme that eliminates harmful aliases and allows for a dip discrimination by means of a minimum apparent velocity. This combination of sparsity-promotion and sampling permits a solution of a very large-scale ℓ_1-minimization problem at a computational cost comparable to iterative-re-weighted least-squares (IRLS Gersztenkorn et al., 1986).

Our formulation for the solution of the seismic data recovery problem reads

$$\mathbf{P}_\epsilon : \qquad \begin{cases} \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \\ \widetilde{\mathbf{f}} = \mathbf{S}^T \widetilde{\mathbf{x}} \end{cases} \tag{3.1}$$

and is reminiscent of the solution of the "inpainting problem", the problem of infilling missing data, reported by Elad et al. (2005). In this expression, $\mathbf{y}$ is the vector with the incomplete data and $\mathbf{x}$ the unknown coefficient vector that generates the decimated data through the modeling matrix, $\mathbf{A}$. The solution of the recovery problem corresponds to finding the sparsity vector, $\mathbf{x}$ with minimal ℓ_1 norm subject to fitting the data to within a noise-dependent ℓ_2 error ϵ. The estimate for the recovered data vector, $\widetilde{\mathbf{f}}$, is obtained by applying the inverse transform, $\mathbf{S}^T$, to the recovered sparsity vector, $\widetilde{\mathbf{x}}$, that solves $\mathbf{P}_\epsilon$. Above formulation for the recovery problem is known to be stable and extends to (seismic) signals that are not strictly sparse but compressible (Candès et al., 2006b). In that case, the recovery error becomes smaller for transforms that concentrate the signal's energy amongst a smaller fraction of the coefficients.

At this point, the well established ability of curvelets (Candès et al., 2006a; Hennenfent and Herrmann, 2006; Herrmann et al., 2007a) enters into the equation. Compared to discrete wavelets, used for digital storage of multidimensional seismic data volumes (Donoho et al., 1999), curvelets truly honor the behavior of seismic wavefields. They correspond to localized 'little plane waves' that are oscillatory in one direction and smooth in the other direction(s) (Candès and Donoho, 2000, 2004). Like directional isotropic wavelets, they are multiscale and multi-directional, but unlike wavelets, they have an anisotropic shape – they obey the so-called parabolic scaling relationship, yielding a width $\propto$ length^2 for the support of curvelets in the physical domain. Curvelets are also strictly localized in the Fourier domain and quasi localized in the space domain, i.e., they decay rapidly away from the crest where they are maximal. The anisotropic scaling is necessary to detect wavefronts (Candès and Donoho, 2005b,a) and explains their high compression rates on seismic data (Candès et al., 2006a; Herrmann et al., 2007a,b).

3.1.2 Outline

To maximally leverage the recent insights gained from compressive sampling, we tie the important aspects of this theory into the formulation of the seismic recovery problem. After presenting a brief overview of this theory, including an intuitive explanation, we emphasize the importance of compression rates on the quality of the recovery by means of a series of stylized experiments. Based on this experience, the appropriate sparsifying transform, sampling strategy and minimal velocity constraint that controls the mutual coherence are reviewed, followed by the formulation of our sparsity-promoting inversion method. We conclude by applying this method to various datasets with a focus on improvements of curvelet-based recovery over recovery with plane-wave destruction and the additional benefits from shot-receiver interpolation with 3-D curvelets over recovery from shot records with 2-D curvelets.

3.2 Compressive sampling

3.2.1 The basics

Compressive sampling states that a signal with a sparse Fourier spectrum can be recovered exactly from sub-Nyquist sampling by solving a sparsity-promoting program that seeks, amongst all possible solutions, a spectrum with the smallest ℓ_1 norm whose inverse Fourier transform equals the sampled data. During the recovery, the rectangular modeling matrix, $\mathbf{A}$, linking the unknown sparsity N-vector, $\mathbf{x}$, to the incomplete n-data vector, $\mathbf{y}$, is inverted. The recovered data is calculated by taking the inverse Fourier transform of the recovered sparsity vector that solves (denoted by the tilde symbol $\tilde{}$) the sparsity promoting program. Compressive sampling provides the conditions under which this underdetermined system of equations

($n \ll N$) can be inverted. This theory also applies to more general situations, including the presence of noise, compressible instead of strictly sparse signals and more general measurement and sparsity bases, replacing the Fourier basis.

To be specific, compressive sampling theory states that $\mathbf{P}_\epsilon$ (cf. Eq. 3.1) recovers in the noise-free case (for $\epsilon \to 0$) the k non-zero entries of the Fourier N-vector exactly from $n \sim k \times \log N$ samples in the vector, $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ (Candès et al., 2006b). For random sampling, this condition was recently improved to $n = k \times 2 \log(N/k)$ by Donoho and Tanner (2007) in the regime $N \gg k$.

So, what is the rational behind these sampling criteria for k-sparse Fourier vectors? Intuitively, one may argue that taking a single time sample corresponds to probing the data by an inner product with a complex exponential in the Fourier domain. This sinusoidal function intercepts with any non-zero entry of the unknown Fourier spectrum. One can argue that two intersections from two arbitrary samples should suffice to determine the amplitude and phase for each non-zero entry of the spectrum. Extending this argument to a k-sparse spectrum turns this into a combinatorial problem, seeking the smallest number of nonzero entries in the sparsity vector with an inverse Fourier transform that fits the data. The theory of compressive sampling provides conditions under which the above combinatorial problem can be replaced by $\mathbf{P}_\epsilon$ for which practical solvers exist. This theory also provides guidelines for sampling strategies that limit the imprint of interference that leads to coherent aliases. After illustrating the importance of compression for the recovery on a series of stylized experiments, we discuss the design of a compressive sampling procedure that is favorable for the recovery of seismic data with traces missing.

3.2.2 A stylized experiment

Sparsifying transforms form the key component of compressive sampling. As we will show below, the accuracy of the recovery depends on the degree of compression achieved by the sparsifying transform. For signals that are not strictly sparse but compressible, their sparsity properties can be measured by the compression rate, r, defined by the exponent for the powerlaw decay of the magnitude-sorted coefficients. The larger r, the faster the decay of the reconstruction error, measuring the energy difference between the original signal and its approximation from the k largest coefficients. Because $\mathbf{P}_\epsilon$ (cf. Eq. 3.1) recovers the largest k coefficients, the recovery of compressible signals improves in a transformed domain with a large compression rate. The challenge is to find a sparsifying transform that also permits a favorable sampling condition.

A series of experiments is conducted that measures the performance of the recovery as a function of the compression rate and the aspect ratio of the modeling matrix, $\delta = n/N$. This aspect ratio is related to the undersampling rate. As before, a modeling matrix defined in terms of the decimated Fourier matrix is used. The experiments are carried out for varying numbers of measurements, n, and for increasing compression rates, i.e., $(\delta, r) \in (0,1] \times (1/2, 2]$. For each parameter combination, twenty different pseudo-random realizations are generated defining the random sampling and the entries in the sparsity vector, $\mathbf{x}_0$. For each r, this vector is calculated by applying random permutations and signs flips to a sequence that decays with i^{-r} for $i = 1 \cdots N$ with $N = 800$. The incomplete data is generated for each realization with $\mathbf{y} = \mathbf{A}\mathbf{x}_0$ and is used as input to StOMP (Donoho et al., 2006a), a solver that solves $\mathbf{P}_\epsilon$ approximately, for $\epsilon = 0$. As a performance metric, the squared relative ℓ_2 error, $\mathrm{err}_2 = \|\widetilde{\mathbf{x}} - \mathbf{x}_0\|^2 / \|\mathbf{x}_0\|^2$, is calculated and averaged amongst the realizations for fixed $(\delta, r) \in (0,1] \times (1/2, 2]$. This error is encoded in the greyscale of the recovery diagram, which is included in Fig. 3.1. Bright regions correspond to parameter combinations that favor accurate recovery. For r fixed, the relative error decays as the number of measurements increases. For each undersampling ratio, $\delta = n/N$, the error decays rapidly as a function of the compression rate, r. This example underlines the importance of finding a representation that has a high compression rate.

The recovery diagram contains another piece of important information. For a user-defined recovery error and empirical decay rate, the degree of undersampling can be calculated from the intercept of the appropriate contour with a line of constant approximation rate. Conversely, for a given degree of undersampling, the relative recovery error can be determined by looking at the grey value at the specified parameter combination for $(\delta,\, r)$.

Approximately a decade ago Sacchi et al. (1998) showed that a sparse Fourier spectrum can be recovered from sub-Nyquist sampling by a Bayesian argument that amounted to the solution of an optimization problem close in spirit to $\mathbf{P}_\epsilon$. While this work has recently been expanded to large-scale problems in higher dimensions by Trad et al. (2006) and Zwartjes and Sacchi (2007), compressive sampling and the presented recovery diagram provide new insights regarding the abruptness of the recovery as a function of the undersampling and the sparsity, and the importance of the compression rate on the quality of the recovery. Unfortunately, the large number of experiments required to compute the recovery diagram preclude a straightforward extension of these experiments to the seismic situation, where problem sizes

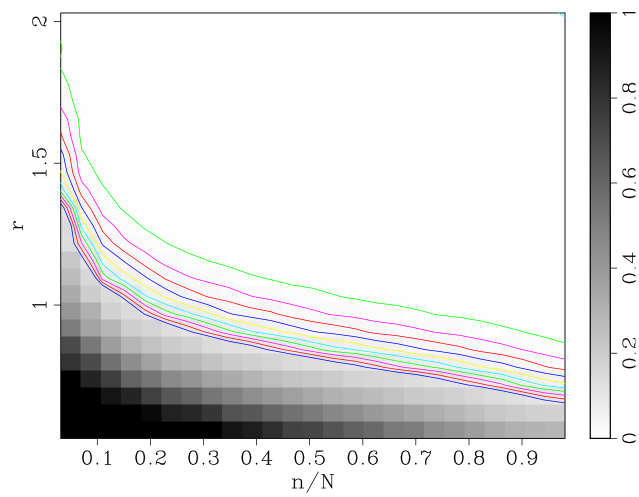


Figure 3.1: Example of a recovery diagram for parameter combinations $(\delta, r) \in (0,1) \times (1/2, 2)$ on a regular grid of 25×25. Notice that the relative ℓ_2 error decays the most rapidly with r. The contour lines represent 1 % decrements in the recovery error starting at 10 % on the lower-left corner and decaying to 1 % in the direction of the upper-right corner.

exceed ($N = \mathcal{O}(2^{30})$). However, this does not mean that abstract concepts of compressive sampling are not useful in the design of a compressive sampling scheme for seismic data.

3.3 Compressive sampling of seismic data

Application of the seismic recovery problem according to the principles of compressive sampling requires a number of generalizations. To make these extensions explicit, the modeling matrix is factored into $\mathbf{A} := \mathbf{RMS}^T$, where $\mathbf{S}^T$ (cf. Eq.3.1) represents the synthesis matrix of the sparsifying transform, $\mathbf{M}$ the measurement matrix and $\mathbf{R}$ the restriction or sampling matrix. The measurement matrix represents the basis in which the measurements are taken and corresponds to the Dirac (identity) basis in seismology and to the Fourier basis in MRI imaging (Lustig et al., 2007). The sampling matrix models missing data by removing zero traces at locations (rows) where data is missing, passing the remaining rows unchanged. The above definition for the modeling matrix is commensurate with the formulation of compressive sampling. As predicted by compressive-sampling theory, the recovery depends quadratically on a new quantity that measures the mutual coherence,

$\mu \geq 1$, between the vectors of the measurement and sparsity bases. This mutual coherence is defined as

$$\mu(\mathbf{M}, \mathbf{S}) = \sqrt{M} \max_{(i,j)\in[1\cdots M]\times[1\cdots N]} |\langle \mathbf{m}_i, \mathbf{s}_j \rangle| \tag{3.2}$$

with $\mathbf{m}_i$ and $\mathbf{s}_j$ the rows of $\mathbf{M}$ and $\mathbf{S}$, respectively. For the Dirac-Fourier pair, where measurements are taken in Euclidean space of a signal that is sparse in Fourier space, this quantity attains its minimum at $\mu = 1$. Because this property quantifies the spread of the vectors from the measurement basis in the sparsifying domain, it explains successful recovery of signals that are sparse in the Fourier domain from a limited number of Euclidean samples. Compressive-sampling theory extends this idea to different measurement and sparsity matrix pairs and this incoherence quantity proves, aside from the compressibility of the to-be-recoverd signal, to be one of the important factors that determines the recovery performance.

3.3.1 Choice for the sparsifying transform

Despite the presence of curved wavefronts with conflicting dips, caustics and a frequency content that spans at least three decades, the curvelet transform attains high compression on synthetic as well as on real seismic data. An intuitive explanation for this behavior lies in the 'principle of alignment', predicting large correlations between curvelets and wavefronts that locally have the same direction and frequency content. This principle is illustrated in Fig. 3.2 and explains that only a limited number of curvelet coefficients interact with the wavefront while the other coefficients decay rapidly away from a wavefront. Remark that curvelets require no knowledge on the location of the wavefronts and do not rely on a NMO correction to reduce the spatial bandwidth. However, additional steps such as focusing (see Herrmann et al., 2008) or spatial-frequency content reduction by NMO will improve the recovery but these require extra prior information.

This compression property of curvelets leads, as shown in Fig. 3.3, to a reconstruction from the largest 1 % coefficients that is far superior compared to Fourier- or wavelet-based reconstructions from the same percentage of coefficients. The curvelet result in Fig. 3.3(d) is artifact free while the Fourier (Fig. 3.3(b)) and wavelet (Fig. 3.3(c)) reconstructions both suffer from unacceptable artifacts. Both for synthetic and real data the observed decays of the magnitude-sorted coefficients, as plotted in Fig. C.1 of Appendix C, support the superior performance of curvelets. By virtue of this property, the curvelet transform is the appropriate choice for our sparsifying transform and we set, $\mathbf{S} := \mathbf{C}$ with $\mathbf{C} \in \mathbb{R}^{N\times M}$ the discrete curvelet transform (Candès et al., 2006a; Ying et al., 2005) with $N > M$ the number of curvelet coefficients and M the size of the fully-sampled data volume, $\mathbf{f}_0 \in \mathbb{R}^M$. See the appendices for further detail on the curvelet transform and its performance on seismic data.

Unlike the Fourier and wavelet bases, curvelets form a frame with a moderate redundancy. Frames share many properties with bases but their redundancy requires care in computing the curvelet coefficients, which are no longer unique. Despite the loss of orthogonality, a technical condition required by compressive sampling, curvelets lead to excellent recovery results, which can be understood intuitively.

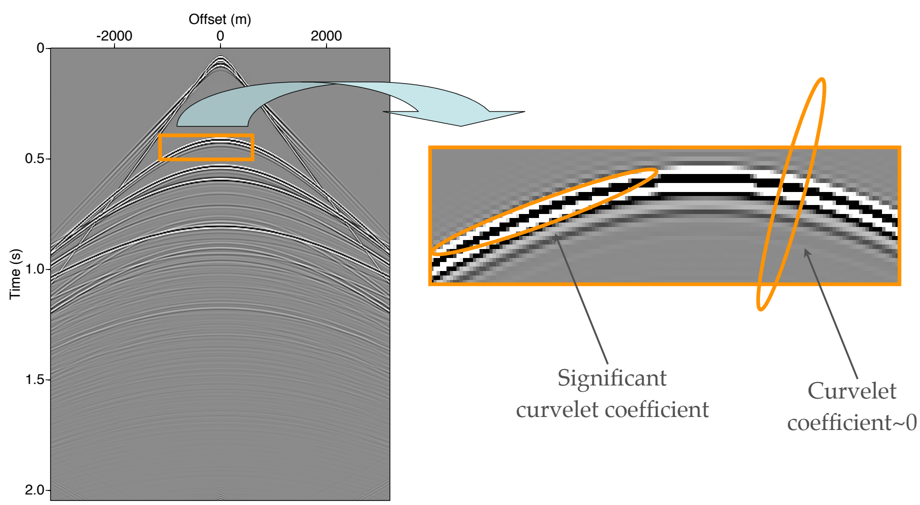


Figure 3.2: Example of the alignment of curvelets with curved events.

3.3.2 The measurement matrix

Sampling of seismic wavefields during acquisition can be considered as taking measurements in the Dirac basis, i.e., $\mathbf{M} := \mathbf{I}$ with $\mathbf{I}$ the identity matrix. This is a good approximation for omnidirectional point sources that are impulsive and for receivers with no directivity and a flat frequency response. For this "choice" of measurement basis – the physics of seismic wavefield acquisition limits this choice to this specific type of measurement basis – the recovery conditions are reasonably favorable according to compressive sampling because the Dirac basis is arguably reasonably incoher-

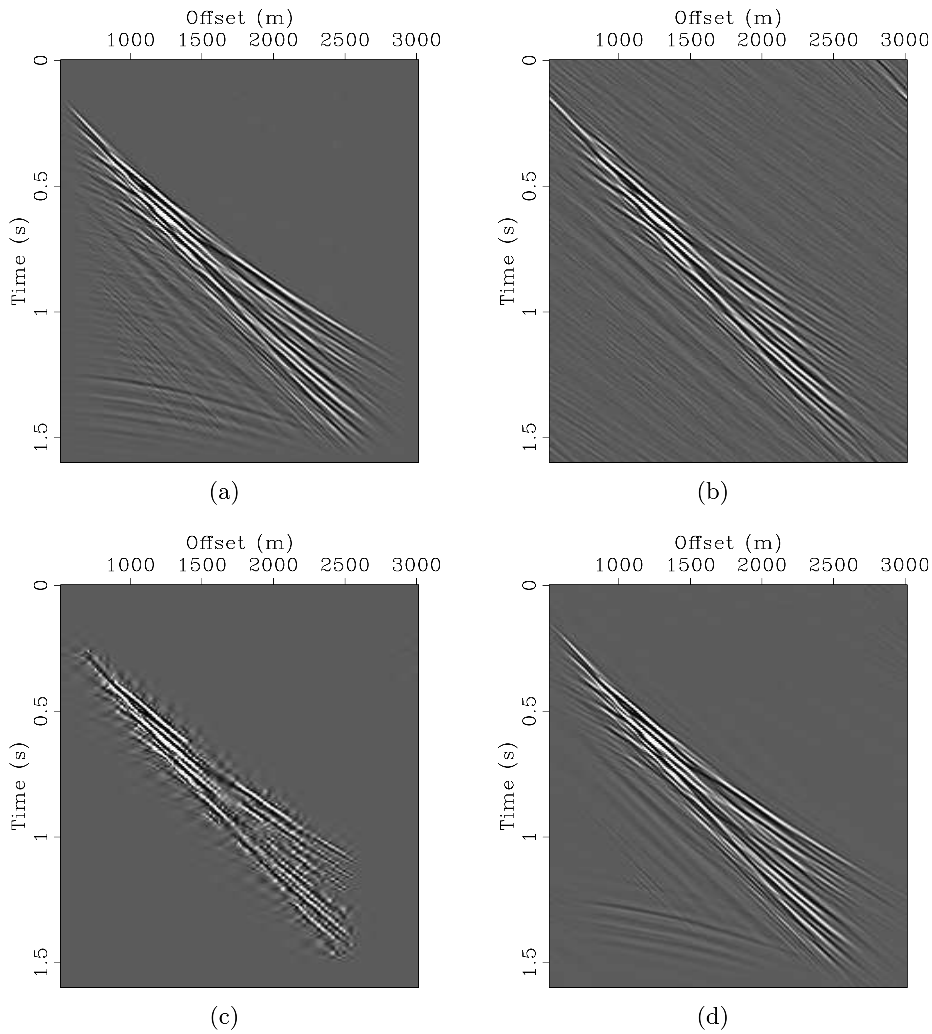


Figure 3.3: Partial reconstruction in different transform domains. **(a)** Original shot record reconstructed from its 1 % amplitude-largest **(b)** Fourier, **(c)** wavelet and **(d)** curvelet coefficients. The curvelet reconstruction is clearly the most accurate approximation.

ent with curvelets, whose Fourier spectrum is confined to localized angular wedges (see Fig. 3.4). We argue that this loss of mutual coherence with respect to the Dirac-Fourier pair is offset by the improved sparsity attained by curvelets (see also our discussion on the role of compression in the stylized examples section). In 3-D this argument gains more strength by virtue of improved sparsity and reduced mutual coherence, i.e., fewer 3-D curvelets are required to capture sheet-like wavefronts while more 3-D curvelets are necessary to cancel each other to approximate a discrete delta Dirac.

Aside from this argument, most if not all practical compressive sampling schemes use sparsifying transforms that are not ideal. For instance, in MRI people use Fourier measurement bases and wavelets as the sparsity basis (Lustig et al., 2007; Candès et al., 2007). At the coarse scales, wavelets become more Fourier-like and hence would adversely affect the recovery. In practice, these less-than-ideal circumstances do not necessarily translate into unfavorable recovery.

Another complication is related to the fact that seismic data is sampled regularly in time and at a subset of source/receiver positions that belong to the acquisition grid. This means that data is fully sampled in time and irregularly along the source/receiver coordinates. This asymmetric trace-by-trace sampling is unfavorable for the recovery because it introduces correlations between vertically-oriented curvelets and vertically-oriented traces along which the data is collected. Fig. 3.4 illustrates this problem schematically.

To incorporate this additional complication in our formalism, we extend the formal definition of mutual coherence (cf. Eq. 6.1) by studying the pseudo mutual coherence between the rows of the acquisition matrix, $\mathbf{RM}$, and the columns of the curvelet synthesis matrix. From this perspective, enforcing a dip discrimination by means of specifying a minimum apparent velocity (see e.g. Zwartjes and Sacchi, 2007), has a natural interpretation in the context of compressive sampling because this discrimination removes steeply dipping curvelets and hence reduces the "mutual coherence" (see Fig. 3.4). This dip discrimination corresponds to Fourier-domain dip filtering and is equivalent to replacing the Dirac measurement basis with a Toeplitz matrix derived from a dip-filtered discrete delta Dirac. In this case, the mutual coherence will also be reduced, yielding a more favorable recovery condition. This observation is consistent with reports in the geophysical literature, where maximal dip limitation for the recovered wavefields are known to improve recovery (Zwartjes and Sacchi, 2007).

Because curvelets are angular selective, it is straightforward to implement the dip discrimination as a diagonal weighting matrix in the curvelet domain. This choice not only avoids having to put infinities in a weighting for the ℓ_1-norm in Eq. 3.1 but it also allows us to redefine the synthesis matrix as

$$\mathbf{S}^T := \mathbf{C}^T\mathbf{W} \quad \text{with} \quad \mathbf{W} = \text{diag}\{\mathbf{w}\} \tag{3.3}$$

with $\mathbf{C}^T \in \mathbb{R}^{M\times N}$ the inverse discrete curvelet transform. The weighting vector, $\mathbf{w}$, contains zeros at positions that correspond to wedges that contain near vertical curvelets and ones otherwise (see Fig. 3.4). However, this redefinition does not impact the actual wavefield because near vertical events can not occur and leads to a reduced mutual coherence between the rows of the acquisition matrix and the columns of the now restricted curvelet synthesis matrix. This restriction removes the curvelets that correlate with traces in the acquisition and therefore leads to a reduction of the mutual coherence, i.e., the sum in Eq. 6.1 no longer runs over the vertically oriented curvelets. The observation that reduced coherence leads to favorable recovery conditions is consistent with the theory of compressive sampling.

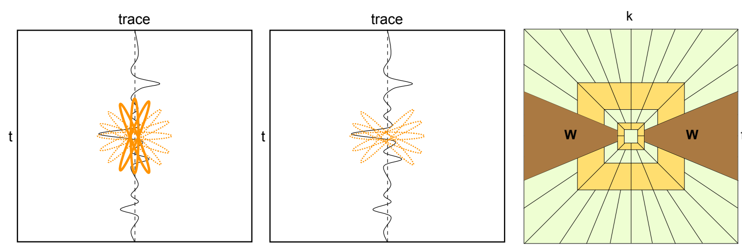


Figure 3.4: Illustration of the angular weighting designed to reduce the adverse effects of seismic sampling. On the left, the increased mutual coherence between near vertical-oriented curvelets and a missing trace. In the middle, a schematic of the curvelets that survive the angular weighting illustrated on the right.

3.3.3 The restriction/sampling matrix

Curvelet-based recovery performs less well in the presence of strong coherent aliases caused by regular undersampling. These coherent aliases are harmful because they lead to artifacts that have large inner products with curvelets, which may lead to falsely recovered curvelets. The performance of transform-based recovery methods depends on a reduction of these aliases that are caused by constructive interference induced by a regular decimation of the data.

Random subsampling according to a discrete uniform distribution – each discrete grid point is equally probable to be sampled – is known to break aliases. For the restricted Fourier matrix, which consists of the Fast Fourier transform (FFT) applied to a vector with zeros inserted at locations where samples are missing, this random sampling turns aliases into a relatively harmless random noise (according to the slogan "*noiseless underdetermined problems behave like noisy well-determined problems*" by Donoho et al., 2006b), allowing for a separation of signal from incoherent interference by a denoising procedure that exploits the sparsifying property of curvelets on seismic data (Hennenfent and Herrmann, 2007a,c). Roughly speaking, this can be understood by arguing that random subsampling according to a discrete uniform distribution corresponds to some sort of a perturbation of the regularly decimated grid that is known to create coherent aliases. As shown in Hennenfent and Herrmann (2007c), this type of sampling, and our extension to jitter sampling, creates a noisy spectrum, where for all wave numbers aliased energy is distributed over the seismic temporal frequency band.

The observation that irregular sampling favors recovery is well known amongst scientists and engineers (Sun et al., 1997; Wisecup, 1998; Malcolm, 2000). Albeit not strictly necessary, we will, for the remainder of this paper, assume that the data is sampled according to a discrete uniform distribution. In practice, there is no need to insist on this condition as long as there is some control on the clustering of the measurements and the size of the largest gaps in the acquisition. Details on this important topic are beyond the scope of this paper and the reader is referred to Donoho and Logan (1992) and to recent applied work by the authors Hennenfent and Herrmann (2007b,c) on jitter sampling.

3.3.4 The modeling matrix

With the sampling and sparsifying matrices in place, the representation for noisy seismic data can now be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n} \quad \text{with} \quad \mathbf{A} := \mathbf{R}\mathbf{I}\mathbf{S}^T, \tag{3.4}$$

$\mathbf{y} \in \mathbb{R}^n$ the noisy measurements and $\mathbf{n} \in \mathbb{R}^n$ a zero-centered pseudo-white Gaussian noise. According to this model, the measurements are related to the sparsity vector $\mathbf{x}_0$ through the modeling matrix $\mathbf{A} \in \mathbb{R}^{n\times N}$. This modeling matrix is defined by compounding the restriction, $\mathbf{R} \in \mathbb{R}^{n\times M}$; measurement, $\mathbf{I} \in \mathbb{R}^{M\times M}$; and inverse transform, $\mathbf{S}^T \in \mathbb{R}^{M\times N}$ matrices. The noisy measurements themselves are given by $\mathbf{y} = \mathbf{R}\mathbf{f}_0 + \mathbf{n}$ with $\mathbf{R} \in$

$\mathbb{R}^{n\times M}$ the restriction matrix taking $n \ll M$ random samples from the full data vector, $\mathbf{f}_0 \in \mathbb{R}^M$. Because the curvelets transform is redundant, the length of the curvelet vector exceeds the length of the full data vector ($N > M > n$). Therefore, our main task is to invert the modeling matrix $\mathbf{A}$ for situations where $\delta = n/N \approx 0.04$ in 2-D and $\delta \approx 0.01$ in 3-D.

3.4 Curvelet Recovery by Sparsity-promoting Inversion (CRSI)

The seismic data regularization problem is solved with matrix-free implementations for the fast discrete curvelet transform (defined by the fast discrete curvelet transform, FDCT, with wrapping, a type of periodic extension, see Candès et al., 2006a; Ying et al., 2005) and the restriction operator. The solution of $\mathbf{P}_\epsilon$ (cf. Eq. 3.1) is cast into a series of simpler unconstrained subproblems. Each subproblem is solved with an iterative soft-thresholding method with thresholds that are carefully lowered. For (extremely) large problems, this cooling leads to the solution of $\mathbf{P}_\epsilon$ with a relatively small number ($\mathcal{O}(100)$) of matrix-vector multiplications.

3.4.1 The unconstrained subproblems

The inversion of the underdetermined system of equations in Eq. 3.4 lies at the heart of compressive sampling. The large system size of seismic data and the redundancy of the curvelet transform exacerbate this problem. Our main thesis is that the matrix, $\mathbf{A}$, can be successfuly inverted with an iterative solution of the sparsity-promoting program $\mathbf{P}_\epsilon$ (cf. Eq. 3.1) by means of a descent method supplemented by thresholding.

Following Elad et al. (2005), the constrained optimization problem, $\mathbf{P}_\epsilon$, is replaced by a series of simpler unconstrained optimization problems

$$\mathbf{P}^\lambda : \quad \begin{cases} \widetilde{\mathbf{x}}_\lambda = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1 \\ \widetilde{\mathbf{f}}_\lambda = \mathbf{S}^T\widetilde{\mathbf{x}}_\lambda. \end{cases} \tag{3.5}$$

These subproblems depend on the Lagrange multiplier λ, determining the emphasis of the ℓ_1-norm over the ℓ_2 data misfit. The solution of $\mathbf{P}_\epsilon$ is reached by solving $\mathbf{P}^\lambda$ for $\lambda \downarrow \lambda_\epsilon$ with $\lambda_\epsilon = \sup_\lambda \{\lambda : \|\mathbf{y} - \mathbf{A}\widetilde{\mathbf{x}}_\lambda\|_2 \leq \epsilon\}$. During the solution of the nonlinear optimization problem $\mathbf{P}^\lambda$, the rectangular matrix $\mathbf{A}$ is inverted by first emphasizing the sparsity-promoting ℓ_1-norm, yielding sparse approximate solutions, followed by a relaxation as λ decreases, increasing the energy captured from the data.

3.4.2 Solution by iterative thresholding

Following Daubechies et al. (2004), Elad et al. (2005); Candés and Romberg (2004) and ideas dating back to Figueiredo and Nowak (2003), the subproblems $\mathbf{P}^\lambda$ are solved by an iterative thresholding technique that derives from the Landweber descent method (Vogel, 2002). According to Daubechies et al. (2004) looping over

$$\mathbf{x} \leftarrow T_\lambda \left(\mathbf{x} + \mathbf{A}^T \left(\mathbf{y} - \mathbf{A}\mathbf{x} \right) \right), \tag{3.6}$$

converges to the solution of $\mathbf{P}^\lambda$ with

$$T_\lambda(x) := \mathrm{sgn}(x) \cdot \max(0, |x| - |\lambda|) \tag{3.7}$$

the soft-thresholding operator. This convergence requires a large enough number of iterations and a largest singular value of $\mathbf{A}$ that is smaller than 1, i.e. $\|\mathbf{A}\| < 1$. Each iteration requires two matrix-vector multiplications.

The descent update, $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{A}^T \big(\mathbf{y} - \mathbf{A}\mathbf{x} \big)$, minimizes the quadratic part of Eq. 3.5. This update is subsequently projected onto the ℓ_1 ball by the soft thresholding. Even though this procedure provably converges to the solution of $\mathbf{P}^\lambda$, the large scale of the seismic regularization problem precludes running these iterations to convergence within a reasonable number of matrix-vector multiplications.

3.4.3 Final solution by cooling

Cooling is a common strategy to solve large to extremely large-scale problems. During this cooling process, the subproblems $\mathbf{P}^\lambda$ are solved approximately for λ decreasing. Because of its simplicity, the iterative-thresholding technique, presented in Eq. 3.6, lends itself particularly well for this approach since it offers a warm start, typically given by the previous outer loop, and control over the accuracy. This accuracy is related to the number of iterations, L, of the inner loop. The higher L the more accurate the solutions of the subproblems become.

The convergence of the overall problem is improved by using the approximate solution of the previous subproblem, the warm start, as input to the next problem for which λ is slightly decreased (Starck et al., 2004; Elad et al., 2005). Sparsity is imposed from the beginning by setting λ_1 close to the largest curvelet coefficient, i.e. $\lambda_1 < \|\mathbf{A}^T \mathbf{y}\|_\infty$. As the Lagrange multiplier is lowered, more coefficients are allowed to enter the solution, leading to a reduction of the data misfit. A similar approach, derived from POCS

Choose: L, K
Initialize: $k = 1$, $\|\mathbf{A}^T\mathbf{y}\|_\infty > \lambda_1 > \cdots > \lambda_K$, $\mathbf{x} = \mathbf{0}$
while $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 > \epsilon$ **and** $k \leq K$ **do**
for $l = 1$ **to** L **do**
$\mathbf{x} = T_{\lambda_k}\left(\mathbf{x} + \mathbf{A}^T\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right)\right)$
end for
$k = k + 1$;
end while
$\widetilde{\mathbf{f}} = \mathbf{S}^T\mathbf{x}$

Table 3.1: The cooling method with iterative thresholding.

(Bregman, 1965), was used by Candés and Romberg (2004) and Abma and Kabir (2006). The details of the cooling method are presented in Table. 3.1.

In practice, five inner loops, i.e., $L = 5$, and 10-50 outer loops, i.e., $10 \leq K \leq 50$, suffice to solve for $\mathbf{x}$ with the series of subproblems $\mathbf{P}^\lambda$. When the cooling is appropriately chosen, the solution of the subproblems converges to the solution of $\mathbf{P}_\epsilon$. The final solution to the seismic data regularization problem, $\widetilde{\mathbf{f}}$, is obtained by applying the weighted-inverse curvelet transform to $\widetilde{\mathbf{x}}$, i.e., $\widetilde{\mathbf{f}} = \mathbf{S}^T\widetilde{\mathbf{x}}$. The total number of matrix-vector multiplications required by this method is similar to those required by iterative-re-weighted least-squares (Gersztenkorn et al., 1986).

3.5 Seismic data recovery with CRSI

The performance of our recovery algorithm is evaluated on synthetic as well as on real data. The first synthetic example is designed to highlight our ability to handle conflicting dips. Next, a synthetic seismic line is used to study the potential uplift for a recovery with 3-D curvelets over a recovery with 2-D curvelets. Finally, our method is tested on real data and compared to a regularization method based on plane-wave destruction (Fomel et al., 2002).

3.5.1 2-D synthetic for a layered earth model

Consider the reflection response of a medium with four plane layers, modeled with a 50-feet (15.24-m) receiver interval, 4-ms sampling interval and a source function given by a Ricker wavelet with a central-frequency of 25-Hz. The dataset contains 256 traces of 500 time samples each. The resulting common-midpoint (CMP) gather after incomplete acquisition is shown in Fig. 3.5(a) together with a close-up in Fig. 3.5(b) of an area with conflicting dips. The incomplete acquisition was simulated by randomly removing 60 % of the traces. This undersampling corresponds to a sub-Nyquist average spatial sampling of 125 feet (38.1 m).

Based on the maximum expected dip of the reflection events in the data, a minimum velocity constraint of 5000 ft/s (1524 m/s) was used. To limit the number of unknowns, the negative dips were excluded. Figs. 3.5(c) and 3.5(d) show the results for the CMP reconstruction with the CRSI algorithm for 100 iterations (5 inner- and 20 outer-loops). The starting Lagrange multiplier was chosen such that 99.5 % of the coefficients do not survive the first threshold. Since the data is noise free, the Lagrange multiplier is lowered such that 99 % of the coefficients survives the final threshold. This corresponds to the situation where $\mathbf{P}_\epsilon$ is solved with a constraint that is close to an equality constraint, i.e., nearly all energy of the incomplete data is captured.

Figs. 3.5(e) and 3.5(f) plot the difference between the recovered and 'ground-truth' complete data. The SNR for the recovery, defined as $SNR = 20 \log \|\widetilde{\mathbf{f}} - \mathbf{f}_0\|/\|\mathbf{f}_0\|$, is about 29.8 dB, which corroborates the observation that there is almost no energy in the difference plots. Curvelet reconstruction clearly benefits from continuity along the wavefronts in the data and has no issue with conflicting dips thanks to the multidirectional property of curvelets.

3.5.2 Common-shot/receiver versus shot-receiver interpolation

Curvelets derive their success in seismology from honoring the multi-dimensional geometry of wavefronts in seismic data. To illustrate the potential benefit from exploiting this high-dimensional geometry, a comparison is made between common-shot interpolation with 2-D curvelets and shot-receiver interpolation with 3-D curvelets. For this purpose, a synthetic seismic line is simulated with a finite-difference code for a subsurface velocity model with two-dimensional inhomogeneities. This velocity model consists of a high-velocity layer that represents salt, surrounded by sedimentary layers and a water bottom that is not completely flat. Using an acoustic finite-difference modeling algorithm, 256 shots with 256 receivers are simulated on a fixed receiver spread with receivers located from 780 to 4620 m with steps of 15 m. The temporal sample interval is 4 ms. The data generated by these simulations can be organized in a 3-D data volume (shot-receiver

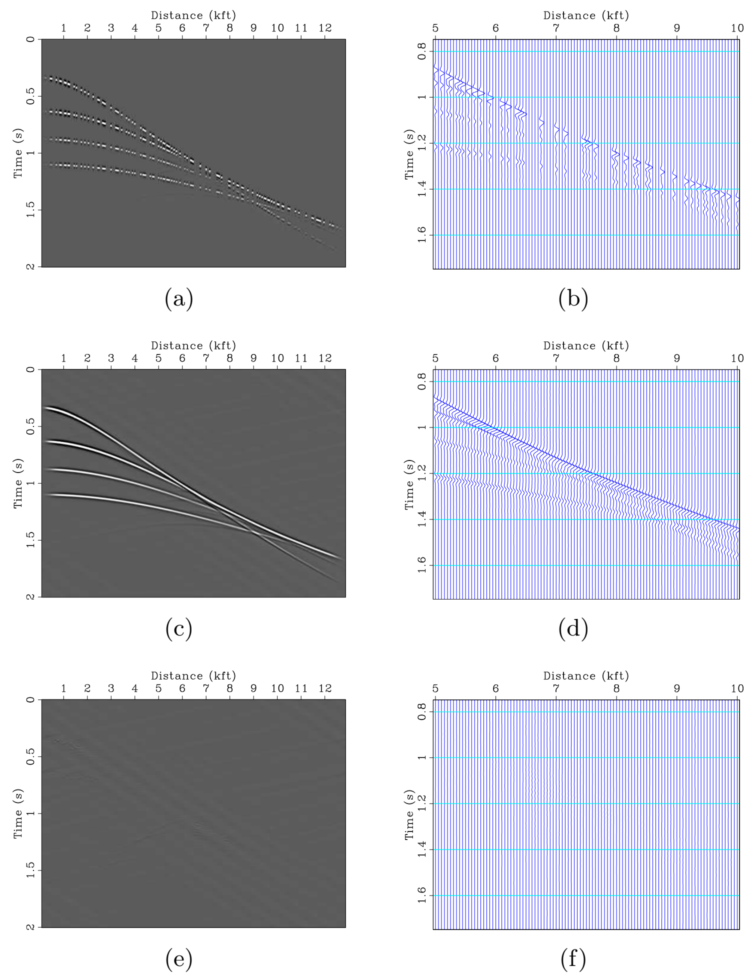
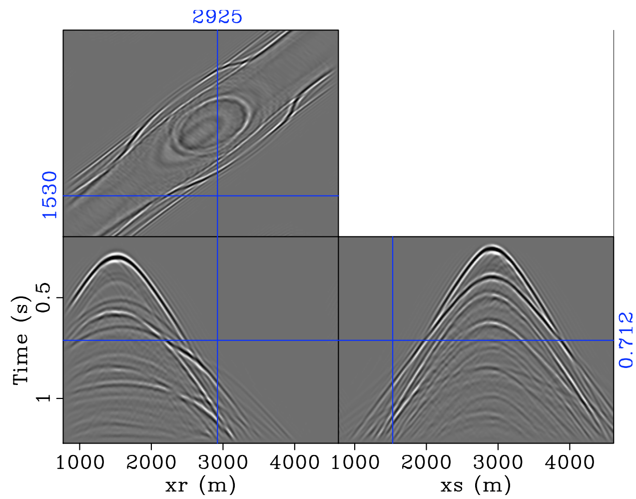


Figure 3.5: Synthetic example of curvelet 2-D reconstruction. **(a)** Simulated acquired data with about 60 % randomly missing traces and **(b)** zoom in a complex area of the CMP gather. **(c)** Curvelet reconstruction and **(d)** same zoom as **(c)**. **(e)** Difference between reconstruction and complete data (not shown here) and **(f)** zoom. Virtually all the initial seismic energy is recovered without error as illustrated by the difference plots (SNR = 29.8 dB).

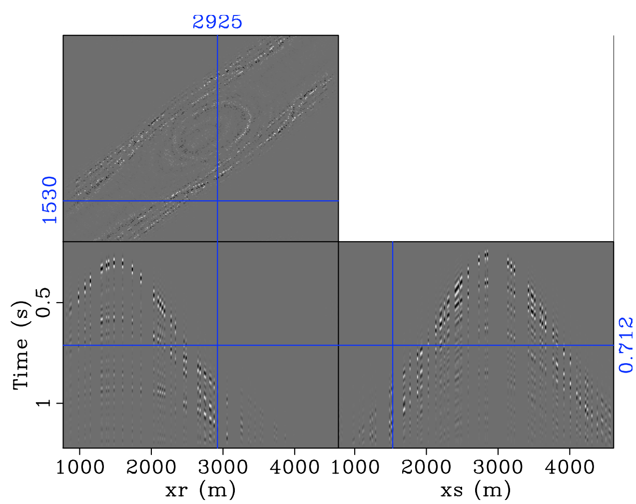
volume) along the shot, x_s, receiver, x_r and time, t coordinates. The full data and the incomplete acquisition are depicted in Figs. 3.6(a) and 3.6(b). The incomplete acquisition is simulated by randomly removing 80 % of the receiver positions for each shot, which corresponds to an average spatial sampling interval of 75 m. Again the full data serves as the ground truth.

To make the comparison, we either solve a series of 2-D dimensional problems on individual shot gathers or we solve the full 3-D interpolation problem. This procedure is outlined in Fig. 3.7 with results for one selected shot record summarized in Fig. 3.8. These results show a clear improvement for the interpolation with the 3-D curvelet transform over the recovery from individual shot records with 2-D curvelets. For both cases results were obtained with 250 iterations and without imposing a minimal velocity constraint. We omitted this constraint because we want to study the uplift without interference from this velocity constraint. Contrasting the results in Figs. 3.8(c) and 3.8(e) confirms the improved recovery by exploiting the 3-D structure, an observation corroborated by the difference plots. The improvement in continuity is particularly visible for the shallow near zero-offset traces where the events have a large curvature. The SNR's for the 2- and 3-D curvelet-based recovery are 3.9 dB and 9.3 dB, respectively, which confirms the visual improvement.

As a possible explanation for the observed performance gain for 3-D curvelets, we argue that 3-D curvelets make up for the increased redundancy (a factor of 24 for 3-D compared to only a factor of 8 in 2-D) by exploiting continuity of wavefronts along an extra tangential direction. This extra direction leads to an improved concentration of the energy amongst relatively fewer curvelet coefficients. The increased dimensionality of 3-D curvelets also makes intersections with areas where data is present more likely. Finally, the theory of compressive sampling tells us that the recovery performance is proportional to the mutual coherence. In 2-D, curvelets are locally line like while 3-D curvelets are locally plate like. Consequently, the mutual coherence between a vertical-oriented 3-D curvelet and a trace is smaller than its 2-D counterpart and this also explains the improved recovery. The result plotted in Fig. 3.9(a) and the difference plot in Fig. 3.9(b) confirm the expected improvement and the recovered data displays a nice continuity along the reconstructed wavefronts. Moreover, there is only minor residual energy in the difference plots for a time slice, common-shot and common-receiver panels. The positions of these slices are indicated by the vertical and horizontal in the different panels. The SNR for the 3-D recovery with the 3-D curvelets is 16.92 dB, which is by all means acceptable.



(a)



(b)

Figure 3.6: Synthetic data volume. **(a)** Complete dataset consisting of $256 \times 256 \times 256$ samples along the source, x_s, receiver, x_r and time coordinates. **(b)** Simulated acquired data with 80 % randomly missing traces.

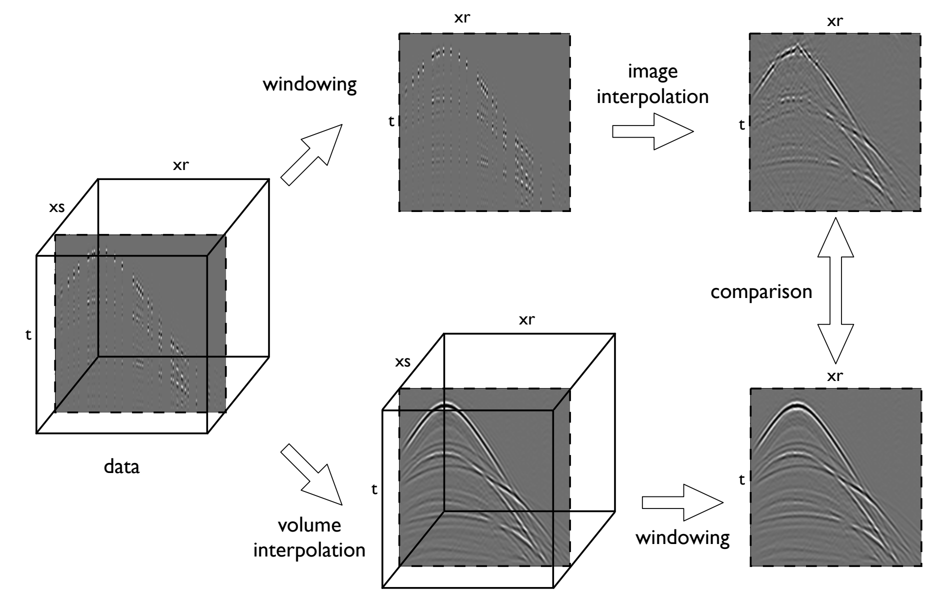


Figure 3.7: Illustration of common shot versus shot-receiver interpolation on the complete data volume.

3.5.3 Comparison between CRSI and plane-wave destruction on 2-D real data

To conclude the discussion, our method is contrasted with an interpolation method based on plane-wave destruction (Fomel et al., 2002). Fig. 4.1(a) displays a real shot record that is used for the comparison. This record is taken from a seismic survey, collected at the offshore Gippsland basin in Australia, and contains traces with the first 1.7 s of data received at 200 hydrophones. The data is sampled at 4 ms with a receiver spacing of 12.5 m. The data is decimated by randomly removing 60 % of the traces, which corresponds to an average spatial sampling interval of 31.25 m. The results obtained with CRSI and the plane-wave destruction method are included in Fig. 3.10. The CRSI result shows a nice recovery with a small residual error. The interpolation result and difference plot for the plane-wave destruction method are included in Figs. 3.10(e) and 3.10(f). These results clearly indicate the challenges imposed by real data, with the recovery performing well for regions with low complexity. However, the plane-wave destruction method does not perform so well for regions where there is more complexity and in particular in regions with conflicting dips. In those areas our

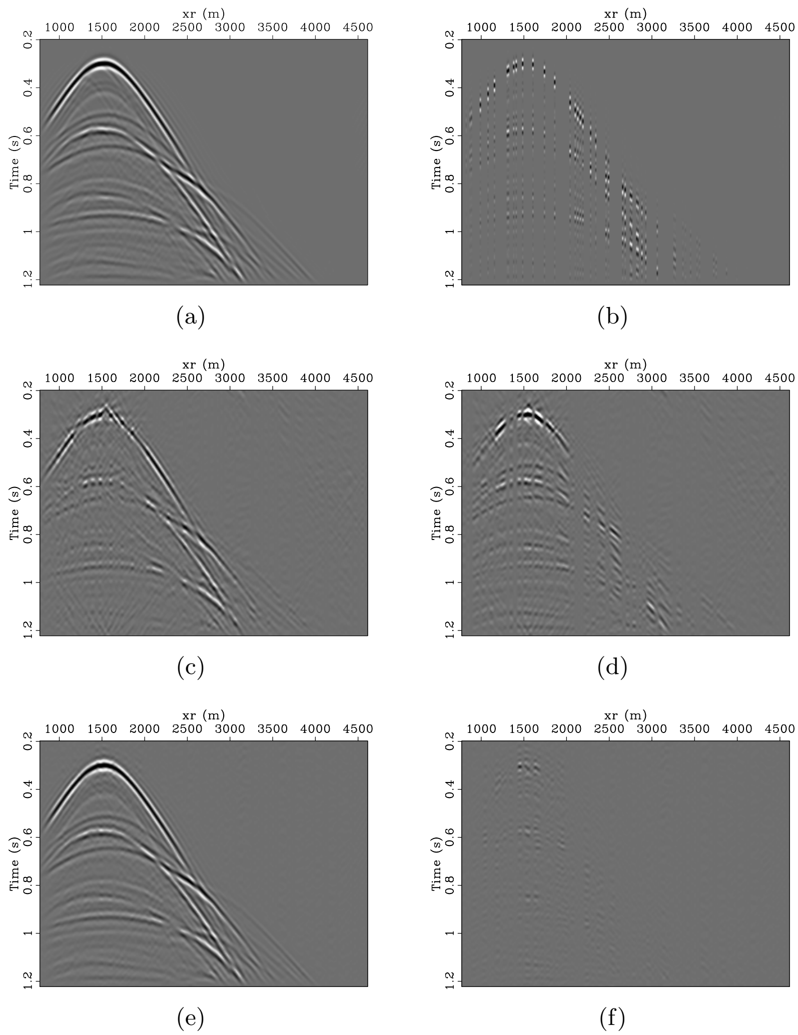
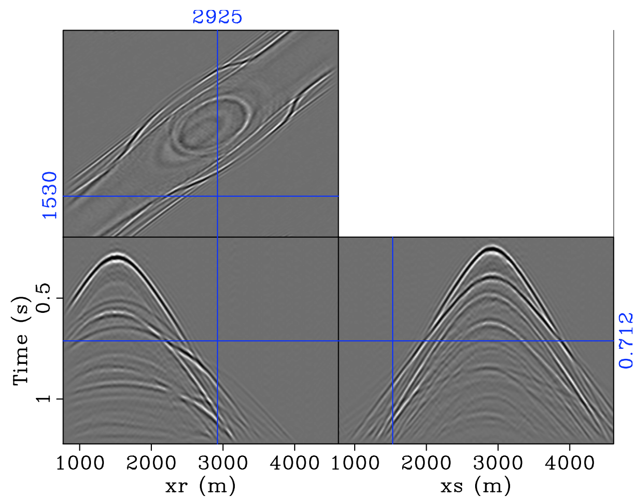
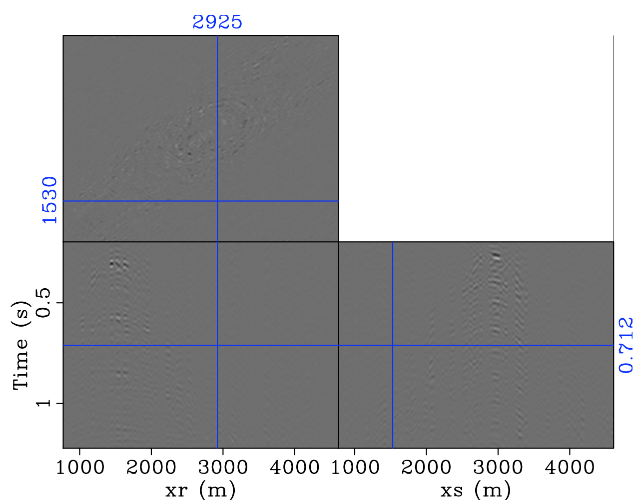


Figure 3.8: Comparison between common-shot (2-D) and shot-receiver (3-D) CRSI. **(a)** Shot from the original data volume. **(b)** Corresponding simulated incomplete data with 80 % randomly missing traces. **(c)** 2-D CRSI result. **(d)** Difference between **(c)** and **(a)**. **(e)** Shot extracted from 3-D CRSI result. **(f)** Difference between **(e)** and **(a)**. 3-D CRSI clearly benefits from 3-D information that greatly improves the reconstruction over 2-D CRSI.



(a)



(b)

Figure 3.9: Synthetic example of curvelet volume interpolation. **(a)** 3-D CRSI result based on the simulated acquired data of Fig. 3.6(b). **(d)** Difference between Fig. 3.6(a) and **(a)**. Notice the continuity and the small difference in the common-shot, common-receiver and time slice. The positions in the cube are indicated by the numbered lines.

curvelet-based method maintains its performance while the plane-wave destruction creates events with erroneous dips. This problem can be related to the inability to assign unique slopes to the reflection events. Curvelets do not experience these difficulties because they can handle multiple dips at the same location. Again, the improved performance is reflected in the SNR's, which is 18.8 dB for 2-D CRSI compared to 5.5 dB for the plane-wave destruction.

3.6 Discussion

3.6.1 Initial findings

Compressive sampling: We showed that the concepts of compressive sampling apply to the seismic recovery problem. Indeed, some of the ideas of compressive sampling are not exactly new to (exploration) seismology, where Fourier, Radon and even migration-based high-resolution approaches have been used to solve the seismic regularization problem. However, compressive sampling offers a clear and concise framework that gives insights into the workings of a successful recovery. These insights offered guidance while making specific choices to exploit the inherent geometry within the seismic wavefield and to eliminate aliases and correlations due to trace-by-trace sampling. Most importantly, compressive sampling tells us that the largest entries of the sparsity vector are recovered thereby underlining the importance of sparsifying transform for seismic data.

Sparsifying transform: An important factor contributing to the performance of our method is the ability of curvelets to parsimoniously capture the essential characteristics of seismic wavefields. This property explains the rapid decay for the magnitude-sorted coefficients and the relative artifact-free reconstruction from a relatively small percentage of largest coefficients. The moderate coherence between the seismic measurement basis and curvelets and the inclusion of the minimal-velocity constraint all contribute to the success of our method. Finally, the results from shot-receiver interpolation showed significant improvement over interpolation on shot records. This behavior is consistent with findings in the literature on Fourier-based recovery (Zwartjes and Sacchi, 2007).

The cooling method: Despite its large scale, the seismic recovery problem lends itself particularly well for a solution by iterative thresholding with

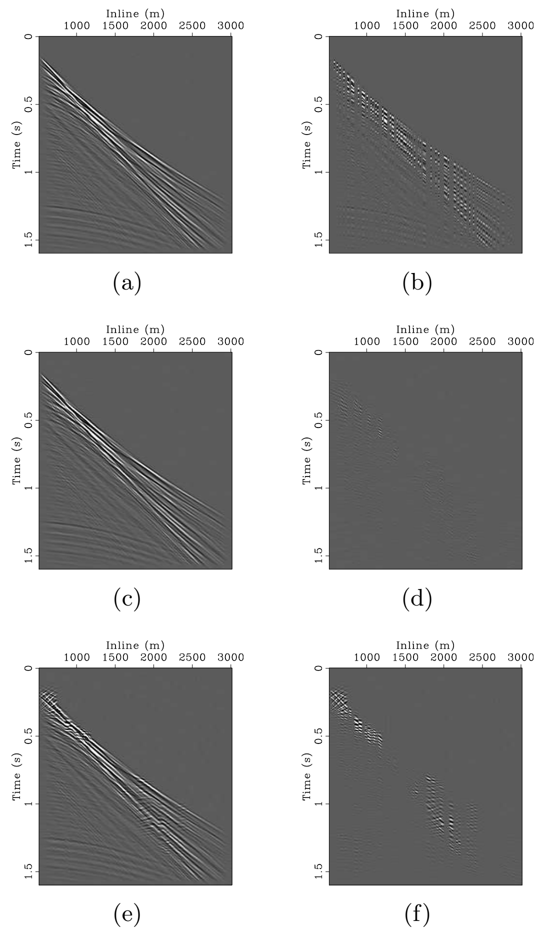


Figure 3.10: Comparison of plane-wave destruction and curvelet-based 2-D recovery on real data. **(a)** Shot-record of a seismic survey from offshore Gippsland basin Australia. Group interval is 12.5 m. **(b)** Incomplete data derived from **(a)** by randomly removing 60 % of the traces (corresponding to average spatial sampling is 31.25 m). **(c)** Result obtained with CRSI. **(d)** Difference between CRSI result and ground truth. **(e)** and **(f)** the same as **(c)** and **(d)** but now obtained with plane-wave destruction. The improvement of the curvelet-based method over the plane-wave destructions is corroborated by the SNR's which are 18.8 dB 5.5 dB, respectively.

cooling. As the threshold is lowered, additional components enter into the solution, which leads to an improved data misfit and controlled loss of sparsity. We find it quite remarkable that this relatively simple threshold-based solver performs so well on the solution of ℓ_1 problems that can be considered as large to extremely large. In a future paper, we plan to report on the properties of this solver compared to other recent developments in solver technology, emerging within the field of compressive sampling (Tibshirani, 1996; Candès and Romberg, 2005; Donoho et al., 2005; Figueiredo et al., 2007; Koh et al., 2007; van den Berg and Friedlander, 2007).

3.6.2 Extensions

Focused CRSI: Our recovery method can be improved when additional information on the wavefield is present. For instance, as part of SRME estimates for the primaries in the data are available. These estimates can be used to focus the energy by compounding the modeling matrix of CRSI with an operator defined by the estimate for the major primaries. As shown by Herrmann et al. (2007c, 2008), this inclusion leads to a better recovery that can be attributed to an improved compression due to focusing with the primaries.

The parallel curvelet transform: Aside from the large number of unknowns within the recovery, seismic datasets typically exceed the memory size of compute nodes in a cluster. The fact that seismic data is acquired in as many as five dimensions adds to this problem. Unfortunately, the redundancy of the curvelet transform makes it difficult to extend this transform to higher dimensions. By applying a domain decomposition in three dimensions, some progress has been made (Thomson et al., 2006). The second problem is still open and may require combination with other transforms.

Jitter sampling: During random sampling there is no precise control over the size of the gaps. This lack of control may lead to an occasional failed recovery. Recently, Hennenfent and Herrmann (2007b) have shown that this problem can be avoided by jitter sampling. During this jitter sampling, the size of the gaps and the occurrence of coherent aliases are both controlled. We report on this recent development elsewhere (Hennenfent and Herrmann, 2007c).

CRSI for unstructured data: The presented interpolation method assumed data to be missing on otherwise regular grids. With the non-uniform fast discrete curvelet transform developed by the authors (Hennenfent and Herrmann, 2006), CRSI no longer requires data to be collected on some underlying grid. This extension makes CRSI applicable in other fields such as global seismology, where irregular sampling and spherical coordinate systems prevail.

Fast (reweighted) ℓ_1 solvers: The success of compressed sensing depends on the ability to solve large-scale ℓ_1 optimization problems. As a result, there has been a surge in research activity addressing this important issue (Tibshirani, 1996; Candès and Romberg, 2005; Donoho et al., 2005; Figueiredo et al., 2007; Koh et al., 2007). One development is particularly relevant and that is the discussion (see Candès et al., 2007, for further details) whether to solve the recovery problem according to Eq. 3.1, known as the synthesis problem or, according to

$$\mathbf{P}_\epsilon^{\mathbf{a}}: \qquad \widetilde{\mathbf{f}} = \arg\min_{\mathbf{f}} \|\mathbf{C}\mathbf{f}\|_1 \quad \text{s.t.} \quad \|\mathbf{R}\mathbf{M}\mathbf{f} - \mathbf{y}\|_2 \leq \epsilon, \tag{3.8}$$

which is known as the analysis problem. Even though there are reports in the literature (Candès et al., 2007) that state that the analysis form (cf. Eq. 3.8) leads to improved recovery results, our experience with (extremely) large problems in CRSI has shown better recovery with the synthesis formulation (cf. Eq. 3.1). Because current hardware affords only $\mathcal{O}(100)$ matrix-vector multiplies, the future challenge will be the inclusion of more sophisticated ℓ_1-norm solvers and the investigation of potential benefits from a possible reweighting and a formulation in the analysis form. The latter corresponds to an approximate solution for the ℓ_0 problem for which encouraging results have been reported (Candès et al., 2007). In a future paper, we plan to report on these issues.

3.7 Conclusions

A new non-parametric seismic data regularization technique was proposed that combines existing ideas from sparsity-promoting inversion with parsimonious transforms that expand seismic data with respect to elements that are multiscale and multidirectional. The compression attained by these elements, which form the redundant curvelet frame, in conjunction with an acquisition that is not too far from random, led to a compressive sampling scheme that recovers seismic wavefields from data with large percentages of traces missing.

Treating the seismic data regularization problem in terms of a compressive sampling problem enabled us to design a scheme that favored recovery. The success of this scheme can be attributed to three main factors, namely the compression of seismic wavefields by curvelets, the control of aliases by (close to) random sampling and the solution of (extremely) large-scale ℓ_1 problems by a cooled iterative thresholding. This combination allowed us to reconstruct seismic wavefields from data with up to 80 % of its traces missing at a cost comparable to other sparsifying transform-based methods. Our method was successfully applied to synthetic and real data. A significant improvement was witnessed for shot-receiver interpolation during which the 3-D geometry of seismic wavefields is fully exploited by 3-D curvelets. Our results also showed a significant improvement on real data with conflicting dips amongst the wave arrivals.

Unfortunately, compressive sampling does not offer explicit sampling criteria for a curvelet-based recovery of seismic wavefields. However, this theory has given us insights that justified the design of our recovery method, where the seismic data regularization problem is solved by sparsity promotion in the curvelet domain.

3.8 Acknowledgments

The authors would like to thank the authors of CurveLab, SparseLab and the Rice Wavelet Toolbox for making their codes available. In particular, we would like to thank Laurent Demanet for providing us with further insights into the curvelet transform. This paper was prepared with Madagascar (`rsf.sourceforge.net`) supplemented by SLIMpy operator overloading, developed by S. Ross Ross. This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and Collaborative Research and Development Grant DNOISE (334810-05) of F.J.H. and was carried out as part of the SINBAD project with support, secured through ITF (the Industry Technology Facilitator), from the following organizations: BG Group, BP, Chevron, ExxonMobil and Shell. The authors would also like to thank ExxonMobil Upstream Research Company for providing us with the real dataset and the Institute of Pure and Applied Mathematics at UCLA supported by the NSF under grant DMS-9810282. Finally, the authors would like to thank the anonymous reviewers whose constructive comments helped improve this paper.

Bibliography

Abma, R. and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, no. 6, E91 – E97.

Bleistein, N., J. Cohen, and J. Stockwell, 2001, Mathematics of multidimensional seismic imaging, migration and inversion: Springer.

Bregman, L., 1965, The method of successive projection for finding a common point of convex sets: Soviet mathematics - Doklady, **6**, 688–692.

Candès, E. and D. L. Donoho, 2005a, Continuous Curvelet Transform II: Discretization and Frames: **19**, 198–222.

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

Candès, E. J. and D. L. Donoho, 2000, Curvelets – a surprisingly effective nonadaptive representation for objects with edges, *in* Rabut, C., A. Cohen, and L. L. Schumaker, eds., Curves and Surfaces, Vanderbilt University Press.

———, 2004, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities: **57**, no. 2, 219–266.

———, 2005b, Continuous Curvelet Transform I: Resolution of the Wavefront Set: **19**, 162–197.

Candés, E. J. and J. Romberg, 2004, Practical signal recovery from random projections: Presented at the Wavelet Applications in Signal and Image Processing XI.

———, 2005, ℓ_1-magic. (Software, `http://www.acm.caltech.edu/l1magic/`).

Candès, E. J., J. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: **59**, no. 8, 1207–1223.

Candès, E. J., M. B. Wakin, and S. P. Boyd, 2007, Enhancing sparsity by reweighted ℓ_1 minimization.

Canning, A. and G. H. F. Gardner, 1996, Regularizing 3-D data sets with DMO: Geophysics, **61**, no. 04, 1103–1114.

Daubechies, I., M. Defrise, and C. de Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constrains: CPAM, 1413–1457.

Donoho, D. L., I. Drori, V. Stodden, and Y. Tsaig, 2005, SparseLab. (Software, `http://sparselab.stanford.edu/`).

———, 2006a, SparseLab.

Donoho, D. L. and B. F. Logan, 1992, Signal Recovery and the Large Sieve: SIAM Journal on Applied Mathematics, **52**, no. 2, 577–591.

Donoho, D. L. and J. Tanner, 2007, How many random projections does one need to recover a k-sparse vector. (submitted for publication).

Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck, 2006b, Sparse solution of underdetermined linear equations by stagewise orthonormal matching pursuit. (Preprint).

Donoho, P., R. Ergas, and R. Polzer, 1999, Development of seismic data compression methods for reliable, low-noise performance: SEG International Exposition and 69th Annual Meeting, 1903–1906.

Elad, M., J.-L. Starck, P. Querre, and D. L. Donoho, 2005, Simulataneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA): **19**, 340–358.

Figueiredo, M. and R. D. Nowak, 2003, An EM algorithm for wavelet-based image restoration.

Figueiredo, M., R. D. Nowak, and S. J. Wright, 2007, Gradient Projection for Sparse Reconstruction. (Software, `http://www.lx.it.pt/~mtf/GPSR/`).

Fomel, S., 2003, Theory of differential offset continuation: Geophysics, **68**, no. 2, 718–732.

Fomel, S., J. G. Berryman, R. G. Clapp, and M. Prucha, 2002, Iterative resolution estimation in least-squares Kirchoff migration: Geophys. Pros., **50**, 577–588.

Gersztenkorn, A., J. B. Bednar, and L. Lines, 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: Geophysics, **51**, 357–369.

Hale, D., 1995, DMO processing: Geophysics Reprint Series.

Harlan, W. S., J. F. Claerbout, and F. Rocca, 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, no. 11, 1869–1880.

Hennenfent, G. and F. J. Herrmann, 2006, Seismic denoising with non-uniformly sampled curvelets: Computing in Science and Engineering, **8**, no. 3, 16 – 25.

———, 2007a, Irregular sampling: from aliasing to noise: Presented at the EAGE 69th Conference & Exhibition.

———, 2007b, Random sampling: new insights into the reconstruction of coarsely-sampled wavefields: Presented at the SEG International Exposition and 77th Annual Meeting.

———, 2007c, Simply denoise: wavefield reconstruction via coarse nonuniform sampling: Geophysics. (To appear).

Herrmann, F. J., 2005, Robust curvelet-domain data continuation with sparseness constraints: Presented at the EAGE 67th Conference & Exhibition Proceedings.

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007a, Nonlinear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, 781–799.

Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2007b, Sparsity- and continuity-promoting seismic imaging with curvelet frames. (To appear, doi:10.1016/j.acha.2007.06.007).

Herrmann, F. J., D. Wang, and G. Hennenfent, 2007c, Multiple prediction from incomplete data with the focused curvelet transform: Presented at the SEG International Exposition and 77th Annual Meeting.

Herrmann, F. J., D. Wang, G. Hennenfent, and P. Moghaddam, 2008, Curvelet-based seismic data processing: a multiscale and nonlinear approach: Geophysics, **73**, no. 1, A1–A5.

Koh, K., S. J. Kim, and S. Boyd, 2007, Simple matlab solver for l1-regularized least squares problems. (Software, `http://www-stat.stanford.edu/~tibs/lasso.html`).

Levy, S., D. Oldenburg, and J. Wang, 1988, Subsurface imaging using magnetotelluric data: Geophysics, **53**, no. 1, 104–117.

Lustig, M., D. L. Donoho, and J. M. Pauly, 2007, Sparse mri: The application of compressed sensing for rapid mr imaging: Magn. Reson. Med. (to appear).

Malcolm, A. E., 2000, Unequally spaced fast Fourier transforms with applications to seismic and sediment core data.: Master's thesis, University of British Columbia.

Malcolm, A. E., M. V. de Hoop, and J. A. Rousseau, 2005, The applicability of DMO/AMO in the presence of caustics: Geophysics, **70**, S1–S17.

Oldenburg, D. W., S. Levy, and K. P. Whittall, 1981, Wavelet estimation and deconvolution: Geophysics, **46**, no. 11, 1528–1542.

Sacchi, M. D., T. J. Ulrych, and C. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: **46**, no. 1, 31–38.

Sacchi, M. D., D. R. Velis, and A. H. Cominguez, 1994, Minimum entropy deconvolution with frequency-domain constraints: Geophysics, **59**, no. 06, 938–945.

Santosa, F. and W. W. Symes, 1986, Linear inversion of band-limited reflection seismogram: SIAM J. of Sci. Comput., **7**, no. 1307-1330.

Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, no. 1, 55–58.

Starck, J.-L., M. Elad, and D. L. Donoho, 2004, Redundant multiscale transforms and their application to morphological component separation: Advances in Imaging and Electron Physics, **132**.

Stolt, R. H., 2002, Seismic data mapping and reconstruction: Geophysics, **67**, no. 3, 890–908.

Sun, Y., G. T. Schuster, and K. Sikorski, 1997, A quasi-Monte Carlo approach to 3-D migration: Theory: Geophysics, **62**, no. 3, 918–928.

Symes, W. W., 2006, Reverse time migration with optimal checkpointing: Technical report, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

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

Thomson, D., G. Hennenfent, H. Modzelewski, and F. J. Herrmann, 2006, A parallel windowed fast discrete curvelet transform applied to seismic processing: Presented at the SEG International Exposition and 76th Annual Meeting.

Tibshirani, R., 1996, Least Absolute Shrinkage and Selection Operator. (Software, `http://www-stat.stanford.edu/~tibs/lasso.html`).

Trad, D. O., 2003, Interpolation and multiple attenuation with migration operators: Geophysics, **68**, no. 6, 2043–2054.

Trad, D. O., J. Deere, and S. Cheadle, 2006, Wide azimuth interpolation: Presented at the 2006 Annual Meeting of the Can. Soc. Expl. Geophys.

Trad, D. O., T. Ulrych, and M. D. Sacchi, 2003, Latest views of the sparse radon transform: Geophysics, **68**, no. 1, 386–399.

Tropp, J., M. Wakin, M. Duarte, D. Baron, and R. Baraniuk, 2006, Random filters for compressive sampling and reconstruction: Presented at the Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP).

Ulrych, T. J. and C. Walker, 1982, Analytic minimum entropy deconvolution: Geophysics, **47**, no. 09, 1295–1302.

van den Berg, E. and M. Friedlander, 2007, In pursuit of a root: Technical report, UBC Computer Science. (TR-2007-16).

Verschuur, D. J. and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, part II: practical aspects and examples: Geophysics, **62**, no. 5, 1596–1611.

Vogel, C., 2002, Computational Methods for Inverse Problems: SIAM.

Wisecup, R., 1998, Unambiguous signal recovery above the nyquist using random-sample-interval imaging: Geophysics, **63**, no. 763-771.

Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, no. 4, V87 – V95.

Ying, L., L. Demanet, and E. J. Candés, 2005, 3-D discrete curvelet transform: , 591413, SPIE.

Zwartjes, P. M. and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased seismic data: Geophysics, **72**, no. 1, V21–V32.

Chapter 4

Wavefield reconstruction via jittered undersampling

4.1 Introduction

While the argument has been made that there is no real theoretical requirement for regular spatial sampling of seismic data (Bednar, 1996), most of the commonly-used multi-trace processing algorithms, e.g., Surface-Related Multiple Elimination (SRME - Verschuur et al., 1992) and wave-equation migration (WEM - Claerbout, 1971), need a dense and regular coverage of the survey area. Field datasets, however, are typically irregularly and/or coarsely sampled along one or more spatial coordinates and need to be interpolated before being processed.

For regularly-undersampled data along one or more spatial coordinates, i.e., data spatially sampled below Nyquist rate, there exists a wide variety of wavefield reconstruction techniques. Filter-based methods interpolate by convolution with a filter designed such that the error is white noise. The most common of these filters are the prediction error filters (PEF's) that can handle aliased events (Spitz, 1991). Wavefield-operator-based methods represent another type of interpolation approaches that explicitly include wave propagation (Canning and Gardner, 1996; Biondi et al., 1998; Stolt, 2002). Finally, transform-based methods also provide efficient algorithms for seismic data regularization (Sacchi et al., 1998; Trad et al., 2003; Zwartjes and Sacchi, 2007; Herrmann and Hennenfent, 2007). However, for irregularly-sampled data, e.g., binned data with some of the bins that are empty, or data that are continuous random undersampled, the performance of the aforementioned interpolation methods may deteriorate.

The objective of this paper is to demonstrate that irregular/random undersampling is not necessarily a drawback for all interpolation methods.

A version of this chapter has been accepted for publication. G. Hennenfent and F.J. Herrmann. Simply denoise: wavefield reconstruction via jittered undersampling. *Geophysics*, 73(3), May-June 2008.

Particular transform-based methods and many other advanced processing algorithms can, indeed, cope with this type of undersampling, as was already observed by other authors (Zhou and Schuster, 1995; Sun et al., 1997; Trad and Ulrych, 1999; Xu et al., 2005; Abma and Kabir, 2006; Zwartjes and Sacchi, 2007). We explain why random undersampling is an advantage for these particular transform-based interpolation methods and how it can be used to our benefit when designing coarse sampling schemes. To keep the discussion as clear and concise as possible, we focus on regular sampling with randomly missing data points, i.e., discrete random (under)sampling. Our conclusions extend to continuous random undersampling though. Unless otherwise specified, the term random is used in the remaining of the text in the discrete sense.

4.1.1 Motivation

Recent results in Information Theory and Approximation Theory established that a signal can be recovered *exactly* from (severely) undersampled data points provided that 1) the signal exhibits sparsity in a known transform domain, 2) the artifacts introduced by undersampling look like incoherent random noise in the sparsifying domain, and 3) a data-consistent sparsity-promoting procedure is used for the recovery. It is possible to build an intuitive understanding of these theoretical results, termed *Compressive Sampling* (CS - Candès et al., 2006; Donoho, 2006; Candès and Romberg, 2006), by considering a simple example.

Figure 4.1(a) shows the superposition of three cosine functions. This signal is sparse in the Fourier domain (condition 1) and is regularly sampled above Nyquist rate. Its amplitude spectrum is plotted in Figure 4.1(b). When the signal is randomly three-fold undersampled according to a discrete uniform distribution as in Figure 4.1(c), its amplitude spectrum, plotted in Figure 4.1(d), is corrupted by artifacts (condition 2) that look like additive incoherent random noise. In this case, the significant coefficients of the to-be-recovered signal remain above the "noise" level. These coefficients can be detected with a denoising technique that promotes sparsity, e.g., nonlinear thresholding (dashed line in Figures 4.1(d) and 4.1(f)), and exactly recovered by an amplitude-matching procedure to fit the acquired data (condition 3). This experiment illustrates a favorable recovery from severely undersampled data points of a signal that is sparse in the Fourier domain.

When the original signal is regularly three-fold undersampled (Figure 4.1(e)), the undersampling artifacts coherently interfere, giving rise to well-known aliases that look like the original signal components (Figure 4.1(f)).

In this case, the above sparsity-promoting recovery scheme may fail because the to-be-recovered signal components and the aliases are both sparse in the Fourier domain. This example suggests that random undersampling according to a discrete uniform distribution is more favorable than regular undersampling for reconstruction algorithms that promote sparsity in the Fourier domain. In general terms, the above observations hint at undersampling schemes that lead to more favorable recovery conditions. Within the field of CS, significant advances have been made regarding the main ingredients that go into the design of an undersampling scheme that favors sparsity-promoting recovery. In this paper, we draw on these results to design a new coarse spatial sampling scheme for seismic data.

4.1.2 Main contributions

We propose and analyze a coarse sampling scheme, termed *jittered undersampling* (Leneman, 1966; Dippe and Wold, 1992), which creates, under specific conditions, a favorable recovery situation for seismic wavefield reconstruction methods that impose sparsity in Fourier or Fourier-related domains (see e.g. Sacchi et al., 1998; Xu et al., 2005; Zwartjes and Sacchi, 2007; Herrmann and Hennenfent, 2007). Jittered undersampling differentiates itself from random undersampling according to a discrete uniform distribution, which also creates favorable recovery conditions (Xu et al., 2005; Abma and Kabir, 2006; Zwartjes and Sacchi, 2007), by controlling the maximum gap in the acquired data. This control makes jittered undersampling very well suited to methods that rely on transforms with localized elements, e.g., windowed Fourier or curvelet transform (Candès et al., 2005a, and references therein). These methods are known to be vulnerable to gaps in the data that are larger than the spatio-temporal extent of the transform elements (Trad et al., 2005).

4.1.3 Outline

After a brief overview of the CS framework and the criteria for a favorable recovery, the effects of different undersampling schemes are studied for signals that are sparse in the Fourier domain. Next, we discuss the advantages of random undersampling and design our jittered undersampling strategy that offers increased control on the acquisition grid. The performance of this new scheme for curvelet-based recovery is illustrated on synthetic and real data.

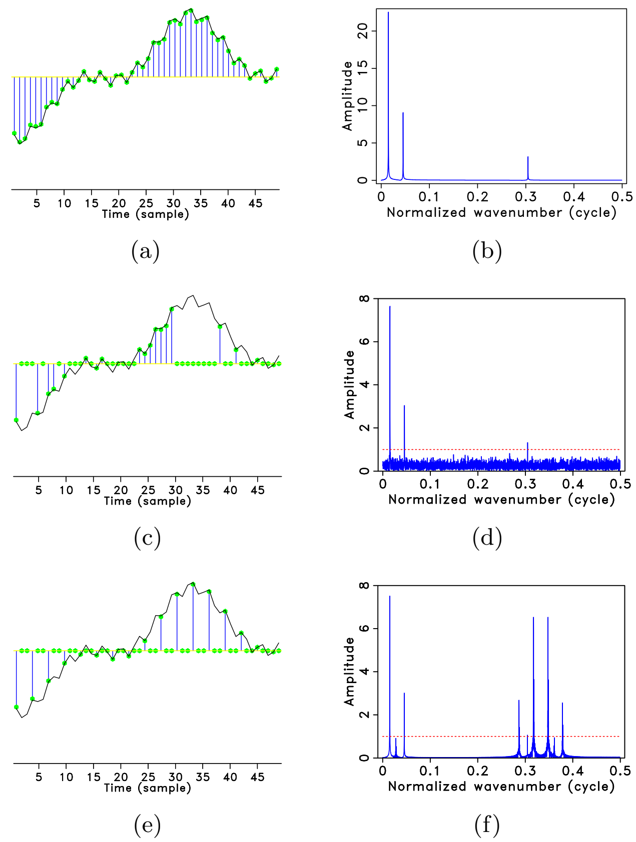


Figure 4.1: Different (under)sampling schemes and their imprint in the Fourier domain for a signal that is the superposition of three cosine functions. Signal (a) regularly sampled above Nyquist rate, (c) randomly three-fold undersampled according to a discrete uniform distribution, and (e) regularly three-fold undersampled. The respective amplitude spectra are plotted in (b), (d) and (f). Unlike aliases, the undersampling artifacts due to random undersampling can easily be removed using a standard denoising technique promoting sparsity, e.g., nonlinear thresholding (dashed line), effectively recovering the original signal.

4.2 Theory

4.2.1 Basics of compressive sampling

An overview of the CS framework and criteria for favorable recovery conditions is given. As mentioned before, CS relies on a sparsifying transform for the to-be-recovered signal and uses this sparsity prior to compensate for the undersampling during the recovery process. For the reconstruction of wavefields in the Fourier (Sacchi et al., 1998; Xu et al., 2005; Zwartjes and Sacchi, 2007), Radon (Trad et al., 2003), and curvelet (Hennenfent and Herrmann, 2005; Herrmann and Hennenfent, 2007) domains, sparsity promotion is a well-established technique documented in the geophysical literature. The main contribution of CS is the new light shed on the favorable recovery conditions.

Recovery by sparsity-promoting inversion

Consider the following linear forward model for the recovery problem

$$\mathbf{y} = \mathbf{R}\mathbf{f}_0, \tag{4.1}$$

where $\mathbf{y} \in \mathbb{R}^n$ represents the acquired data, $\mathbf{f}_0 \in \mathbb{R}^N$ with $N \gg n$ the unaliased signal to be recovered, i.e., the model, and $\mathbf{R} \in \mathbb{R}^{n\times N}$ the restriction operator that collects the acquired samples from the model. Assume that $\mathbf{f}_0$ has a sparse representation $\mathbf{x}_0 \in \mathbb{C}^N$ in some known transform domain $\mathbf{S}$, equation 4.1 can now be reformulated as

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 \quad \text{with} \quad \mathbf{A} \stackrel{\text{def}}{=} \mathbf{R}\mathbf{S}^H, \tag{4.2}$$

where the symbol ^{H} represents the conjugate transpose. As a result, the sparsity of $\mathbf{x}_0$ can be used to overcome the singular nature of $\mathbf{A}$ when estimating $\mathbf{f}_0$ from $\mathbf{y}$. After sparsity-promoting inversion, the recovered signal is given by $\tilde{\mathbf{f}} = \mathbf{S}^H\tilde{\mathbf{x}}$ with

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{4.3}$$

In these expressions, the symbol $\tilde{}$ represents estimated quantities and the ℓ_1 norm is defined as $\|\mathbf{x}\|_1 \stackrel{\text{def}}{=} \sum_{i=1}^N |\mathbf{x}[i]|$, where $\mathbf{x}[i]$ is the i^{th} entry of the vector $\mathbf{x}$.

Minimizing the ℓ_1 norm in equation 4.3 promotes sparsity in $\mathbf{x}$ and the equality constraint ensures that the solution honors the acquired data. Among all possible solutions of the (severely) underdetermined system of linear equations ($n \ll N$) in equation 4.2, the optimization problem in equation 4.3 finds a sparse or, under certain conditions, the sparsest (Donoho and Huo, 2001) possible solution that explains the data.

Favorable recovery conditions

Following Verdu (1998) and Donoho et al. (2006), we define the matrix $\mathbf{L} \stackrel{\text{def}}{=} \mathbf{A}^H\mathbf{A} - \alpha\mathbf{I}$ to study the undersampling artifacts $\mathbf{z} \stackrel{\text{def}}{=} \mathbf{L}\mathbf{x}_0$. The matrix $\mathbf{I}$ is the identity matrix and the parameter α is a scaling factor such that $\text{diag}(\mathbf{L}) = \mathbf{0}$. For more general problems and in particular in the field of digital communications, these undersampling artifacts $\mathbf{z}$ are referred to as *Multiple-Access Interference* (MAI).

According to the CS theory (Candès et al., 2006; Donoho, 2006), the solution $\tilde{\mathbf{x}}$ in equation 4.3 and $\mathbf{x}_0$ coincide when two conditions are met, namely 1) $\mathbf{x}_0$ is sufficiently sparse, i.e., $\mathbf{x}_0$ has few nonzero entries, and 2) the undersampling artifacts are incoherent, i.e., $\mathbf{z}$ does not contain coherent energy. The first condition of sparsity requires that the energy of $\mathbf{f}_0$ is well concentrated in the sparsifying domain. The second condition of incoherent random undersampling artifacts involves the study of the sparsifying transform $\mathbf{S}$ in conjunction with the restriction operator $\mathbf{R}$. Intuitively, it requires that the artifacts $\mathbf{z}$ introduced by undersampling the original signal $\mathbf{f}_0$ are not sparse in the $\mathbf{S}$ domain. When this condition on $\mathbf{z}$ is not met, sparsity alone is no longer an effective prior to solve the recovery problem. Albeit qualitative, the second condition provides a fundamental insight in choosing undersampling schemes that favor recovery by sparsity-promoting inversion.

4.2.2 Fourier-domain undersampling artifacts

Undersampling artifacts in the Fourier domain are studied for two reasons. Firstly, several interpolation methods are based on the Fourier transform (Sacchi et al., 1998; Xu et al., 2005; Zwartjes and Sacchi, 2007). Secondly, the curvelet transform, a dyadic-parabolic partition of the Fourier domain, forms the basis of our recently-introduced recovery scheme (Herrmann and Hennenfent, 2007). Curvelets are in many situations to be preferred over Fourier because of their ability to sparsely represent complex seismic data. For a detailed discussion on this topic, we refer to Candès et al. (2005a) and Hennenfent and Herrmann (2006).

In the coming discussion, the sparsifying transform is defined as the Fourier transform, i.e., $\mathbf{S} \stackrel{\text{def}}{=} \mathbf{F}$. For this definition, the vector generating the Hermitian Toeplitz and circulant matrix $\mathbf{A}^H\mathbf{A}$ is the discrete Fourier transform of the (under)sampling pattern. This pattern has ones where samples are taken, zeros otherwise. Besides, the undersampling artifacts generated by the convolution operator $\mathbf{L}$ are known as *spectral leakage* (Xu et al., 2005).

Regular (under)sampling

When $\mathbf{R}$ keeps all the data points of $\mathbf{f}_0$, i.e., $\mathbf{R} = \mathbf{I}$, the matrix $\mathbf{A}^H\mathbf{A}$ is the identity matrix, as depicted in Figure 4.2(a), $\mathbf{L} = 0$, as plotted in Figure 4.2(d), and there is no spectral leakage. This property holds for any orthonormal sparsifying transform.

When $\mathbf{R}$ corresponds to a regular undersampling scheme, the matrix $\mathbf{A}^H\mathbf{A}$ is no longer diagonal. It now also has a number of nonzero off-diagonals as depicted in Figure 4.2(b). These off-diagonals create aliases, i.e., undersampling artifacts that are the superposition of circular-shifted versions of the original spectrum. Since $\mathbf{x}_0$ is assumed to be sparse, these aliases are sparse as well. Therefore, they are also likely to enter in the solution $\tilde{\mathbf{x}}$ during sparsity-promoting inversion. Because the ℓ_1 norm can not efficiently discriminate the original spectrum from its aliases, regular undersampling is the most challenging case for recovery.

In the seismic community, difficulties with regularly undersampled data are acknowledged when reconstructing by promoting sparsity in the Fourier domain. For example, Xu et al. (2005) write that the anti-leakage Fourier transform for seismic data regularization "may fail to work when the input data has severe aliasing".

Random undersampling according to a discrete uniform distribution

When $\mathbf{R}$ corresponds to a random undersampling according to a discrete uniform distribution, the situation is completely different. The matrix $\mathbf{A}^H\mathbf{A}$ is dense (Figure 4.2(c)) and the convolution matrix $\mathbf{L}$ is a random matrix (Figure 4.2(f)). Consequently, we have

$$\mathbf{A}^H\mathbf{y} = \mathbf{A}^H\mathbf{A}\mathbf{x}_0 \approx \alpha\mathbf{x}_0 + \mathbf{n}, \tag{4.4}$$

where the spectral leakage is approximated by additive white Gaussian noise $\mathbf{n}$. For infinitely large systems (Donoho et al., 2006), this approximation becomes an equality. Because of this property, the recovery problem turns into a much simpler denoising problem, followed by a correction for the amplitudes. Remember that the acquired data **y** are noise-free (cf. equation 4.2) and that the noise **n** in equation 4.4 only comes from the underdeterminedness of the system. In other words, random undersampling according to a discrete uniform distribution spreads the energy of the spectral leakage across the Fourier domain turning the noise-free underdetermined problem (cf. equation 4.2) into a noisy well-determined problem (cf. equation 4.4) whose solution can be recovered by solving equation 4.3. This observation was first reported by Donoho et al. (2006).

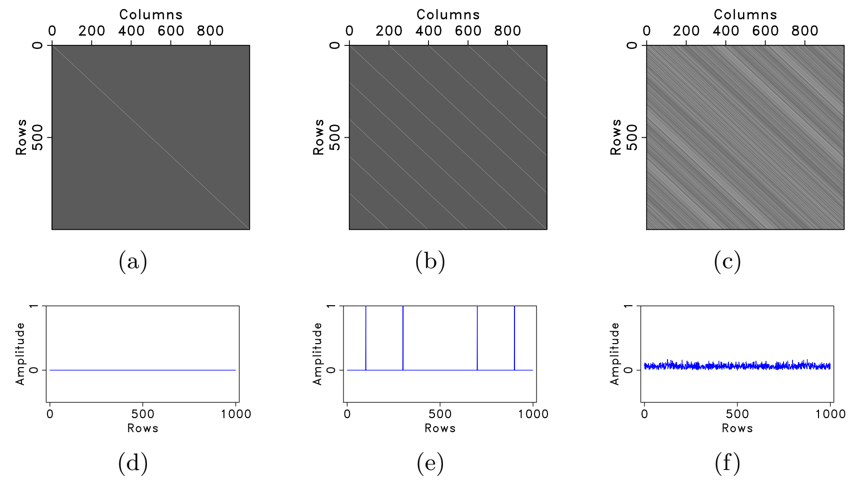


Figure 4.2: Convolution matrix (in amplitude) for (a) regular sampling above Nyquist rate, (b) regular five-fold undersampling, and (c) random five-fold undersampling according to a discrete uniform distribution. The respective convolution kernels (in amplitude) that generate spectral leakage are plotted in (d), (e) and (f). Despite the same undersampling factor, regular and random undersamplings produce very different spectral leakage.

The practical requirement of maximum gap control

As shown in the previous section, random undersampling according to a discrete uniform distribution creates favorable recovery conditions for a reconstruction procedure that promotes sparsity in the Fourier domain. However, a global transform such as the Fourier transform does not typically permit a sparse representation for complex seismic wavefields (Hennenfent and Herrmann, 2006). It requires a more local transform, e.g., windowed

Fourier (Zwartjes and Sacchi, 2007) or curvelet (Herrmann and Hennenfent, 2007) transform. In this case, problems arise with gaps in the data that are larger than the spatio-temporal extent of the transform elements (Trad et al., 2005). Consequently, undersampling schemes with no control on the size of the maximum gap, e.g., random undersampling according to a discrete uniform distribution, become less attractive. The term gap refers here to the interval between two adjacent acquired traces minus the interval associated with the fine interpolation grid, such that adequate sampling has gaps of zero. We present an undersampling scheme that has, under some specific conditions, an anti-aliasing effect, yet offering control on the size of the maximum gap.

4.2.3 Uniform jittered undersampling on a grid

First, the undersampling grid is defined for a discrete uniform jitter. Next, the spectral leakage caused by this scheme is studied.

Definition of the jittered grid

The basic idea of jittered undersampling is to regularly decimate the interpolation grid and subsequently perturb the coarse-grid sample points on the fine grid. As for random undersampling according to a discrete uniform distribution, where each location is equally likely to be sampled, a discrete uniform distribution for the perturbation around the coarse-grid points is considered (see Appendix D and Leneman (1966) for more details).

To keep the derivation of our jittered undersampling scheme succinct, the undersampling factor, γ, is taken to be odd, i.e., $\gamma = 1, 3, 5, \ldots$ We also assume that the size N of the interpolation grid is a multiple of γ so that the number of acquired data points $n = N/\gamma$ is an integer. For these choices, the jittered-sampled data points are given by

$$\mathbf{y}[i] = \mathbf{f}_0[j] \quad \text{for} \quad i = 1, \ldots, n \quad \text{and} \quad j = \underbrace{\frac{1-\gamma}{2} + \gamma \cdot i}_{\text{deterministic}} + \underbrace{\epsilon_i}_{\text{random}}, \tag{4.5}$$

where the discrete random variables ϵ_i are integers independently and identically distributed (iid) according to a uniform distribution on the interval between $-\lfloor(\xi - 1)/2\rfloor$ and $\lfloor(\xi - 1)/2\rfloor$. The jitter parameter $0 \leq \xi \leq \gamma$ relates to the size of the perturbation around the coarse regular grid. The floor function of a real number q, denoted $\lfloor q \rfloor$, is a function that returns the highest integer less than or equal to q. The above sampling can be adapted for the case γ is even.

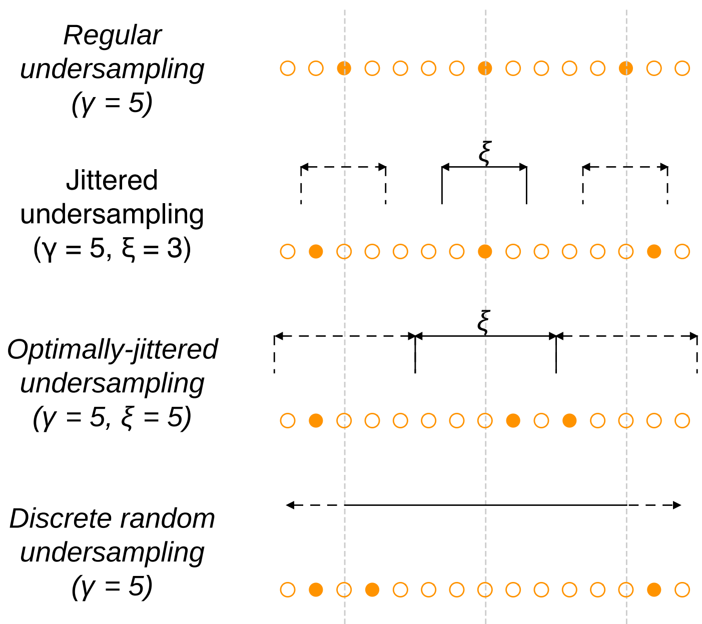


Figure 4.3: Schematic comparison between different undersampling schemes. The circles define the fine grid on which the original signal is alias-free. The solid circles represent the actual sampling points for the different undersampling schemes. The jitter parameter ξ relates to how far the actual jittered sampling point can be from the regular coarse grid, effectively controlling the size of the maximum acquisition gap.

In Figure 4.3, schematic illustrations are included for samplings with increasing randomness. The fine grid of open circles denotes the interpolation grid on which the model $\mathbf{f}_0$ is defined. The solid circles correspond to the coarse sampling locations. These illustrations show that for jittered undersampling, the maximum gap size can not exceed $(\gamma - 1) + 2 \cdot \lfloor (\xi - 1)/2 \rfloor$ data points. For regular undersampling, all the gaps are of size $\gamma - 1$ and for random undersampling according to a discrete uniform distribution, the maximum gap size is $N - n$. Remember that the number of samples is the same for each of these undersampling schemes.

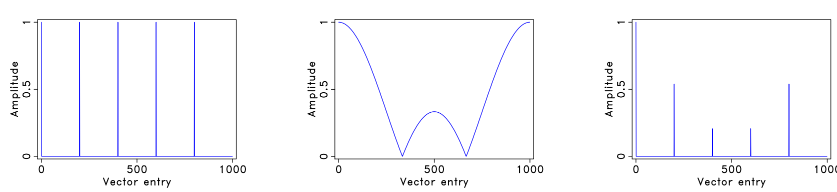
As mentioned earlier, recovery with localized transforms depends on both the maximum gap size and a sufficient sampling randomness to break the coherent aliases. In the next section, we show how the value of the jitter parameter controls these two aspects in our undersampling scheme.

Fourier-domain artifacts of the jittered grid

When $\mathbf{R}$ describes a jittered undersampling scheme according to a discrete uniform distribution, the stochastic expectation $\mathrm{E}\{\cdot\}$ of the first column $\mathbf{a}$ of the circulant matrix $\mathbf{A}^H\mathbf{A}$ is given by

$$\mathrm{E}\left\{\mathbf{a}[k]\right\} \approx \begin{cases} n \cdot \mathrm{sinc}\left(\frac{\xi}{N}(k-1)\right), & \text{if} \quad k = 1 + l \cdot n \quad \text{for} \quad l = 0, \ldots, \frac{\gamma-1}{2} \\ n \cdot \mathrm{sinc}\left(\frac{\xi}{N}(k-1-N)\right), & \text{if} \quad k = 1 + l \cdot n \quad \text{for} \quad l = \frac{\gamma+1}{2}, \ldots, \gamma - 1 \\ 0 & \text{otherwise,} \end{cases} \tag{4.6}$$

where $\mathrm{sinc}(\cdot)$ is the normalized sinc function defined as $\mathrm{sinc}(x) \overset{\text{def}}{=} \sin(\pi x)/\pi x$.



(a) (b) (c)

Figure 4.4: Amplitude spectrum of (a) a five-fold ($\gamma = 5$) regular undersampling vector, (b) a three-sample wide uniform distribution ($\xi = 3$), and (c) the resulting jittered undersampling vector. The first half of the vectors contains the positive frequencies starting with zero, the second half contains the negative frequencies in decreasing order.

The above expression corresponds to an elementwise multiplication of the periodic Fourier spectrum of the discrete regular sampling vector with a sinc function. This sinc function follows from the Fourier transform of the probability density function for the perturbation with respect to a point of the regularly decimated grid.

In Figure 4.4 the amplitudes for this Fourier-domain multiplication are plotted for jittered undersampling with $\gamma = 5$ and $\xi = 3$, i.e., on average four-out-of-five samples are missing for a jitter that includes the decimated grid point, one sample on the right and one sample on the left (cf. Figure 4.3, second row).

Equation 4.6 is a special case of the result for jittered undersampling according to an arbitrary distribution introduced by Leneman (1966) and further detailed in Appendix D. Because these results were originally derived for the continuous case, the above expression is approximate. In practice, however, this formula proves to be accurate, an observation corroborated by numerical results presented below. Consider the following cases for a fixed undersampling factor γ.

Regular undersampling ($\xi = 0$): As observed from the first row of Figure 4.3, there is no jitter in this case and equation 4.6 becomes

$$\mathbf{a}[k] = \begin{cases} n, & \text{for} \quad k = 1 + l \cdot n \quad \text{with} \quad l = 0, \cdots, \gamma - 1 \\ 0, & \text{otherwise.} \end{cases} \tag{4.7}$$

The undersampling artifacts $\mathbf{z}$ consist of aliased energy.

Optimally-jittered undersampling ($\xi = \gamma$): Now the sampling points are perturbed within contiguous windows, as depicted in the third row of Figure 4.3, and equation 4.6 reduces to

$$\mathrm{E}\left\{\mathbf{a}[k]\right\} \approx \begin{cases} n, & \text{for} \quad k = 1 \\ 0, & \text{otherwise.} \end{cases} \tag{4.8}$$

In this special case, the cause of the aliases is removed by the zeros of the sinc function. As with random undersampling according to a discrete uniform distribution, the off-diagonals of the matrix $\mathbf{A}^H\mathbf{A}$ (cf. Figure 4.5(b) and 4.2(c)) are random, turning aliases into noise. Again, the kernel of $\mathbf{L}$ does not contain coherent energy, as observed in Figure 4.5(d), for a five-fold undersampling ($\gamma = 5$) and a jitter parameter of $\xi = 5$. In that sense, this specific relation between the jitter parameter and the undersampling factor is optimal because it creates the most favorable conditions for recovery with a localized transform.

Jittered undersampling ($0 < \xi < \gamma$): In this regime, both coherent aliases and incoherent random undersampling noise are present. Depending on the choice for the jitter parameter, the energy either localizes or randomly spreads across the spectrum. Again, the reduction of the aliases is related to the locations of the zero crossings of the sinc function that move as a function of ξ. As ξ increases, the zeros move closer to the aliases. As expected, the matrix $\mathbf{A}^H\mathbf{A}$, plotted in Figure 4.5(a), still contains the imprint of coherent off-diagonals, resulting in a kernel of $\mathbf{L}$, included in Figure 4.5(c), that is a superposition of coherent aliases and incoherent random noise. Although this regime reduces the aliases, coherent energy remains in the undersampling artifacts. This residue creates a situation that is less favorable for recovery. Depending on the relative strength of the aliases compared to the magnitude n of the diagonal of $\mathbf{A}^H\mathbf{A}$, recovery becomes increasingly more difficult, an observation that can be established experimentally.

In the next section, a series of controlled experiments is conducted to compare the recovery from regularly, randomly according to a discrete uniform distribution and optimally-jittered undersamplings.

4.2.4 Controlled recovery experiments for different sampling schemes

With the favorable sampling schemes identified, it remains to be shown that these samplings lead to an improved recovery compared to the unfavorable regular undersampling. In particular, we want to experimentally confirm that jittered undersampling behaves similarly as random undersampling according to a discrete uniform distribution.

For this purpose, we define the sparsifying transform $\mathbf{S}$ as the Fourier transform $\mathbf{F}$, i.e., $\mathbf{S} \stackrel{\text{def}}{=} \mathbf{F}$, and generate a vector $\mathbf{x}_0$ with k nonzero entries and of length $N = 600$. The nonzero entries of $\mathbf{x}_0$ are distributed at random with random signs and amplitudes. The to-be-recovered signal $\mathbf{f}_0$ is given by $\mathbf{f}_0 = \mathbf{S}^H\mathbf{x}_0$ and the observations $\mathbf{y}$ are obtained by undersampling $\mathbf{f}_0$ regularly, randomly according to a discrete uniform distribution, or optimally-jittered, i.e., $\xi = \gamma$. Finally, the estimated spectrum $\tilde{\mathbf{x}}$ of $\mathbf{f}_0$ is obtained by solving equation 4.3 with the Spectral Projected Gradient for ℓ_1 solver (SPGL1 - van den Berg and Friedlander, 2007). Keep in mind that the number k of nonzero entries of $\mathbf{x}_0$ is not known a priori. Each experiment is repeated 100 times for the different undersampling schemes and for varying undersampling factors γ, ranging from 2 to 6. The reconstruction error is the number of entries at which the estimated representation $\tilde{\mathbf{x}}$ and the true representation $\mathbf{x}_0$ of $\mathbf{f}_0$ in the Fourier domain disagree by more

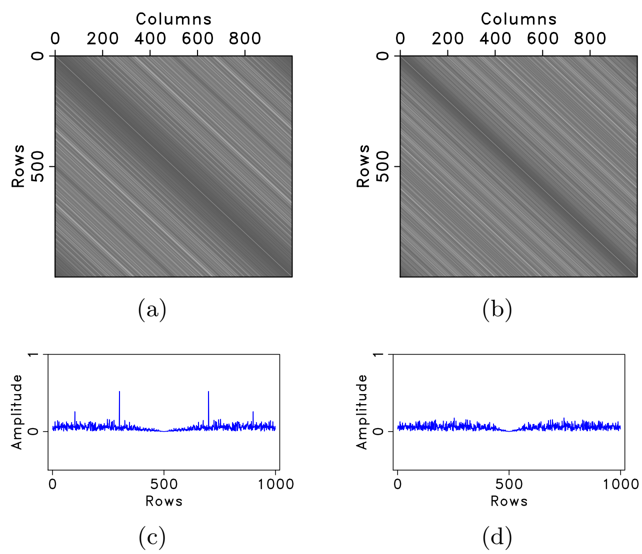
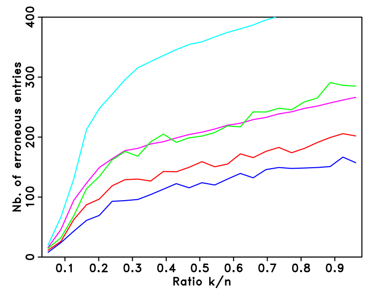


Figure 4.5: Jittered undersampling according to a discrete uniform distribution. (a) Suboptimal and (b) optimal jittered five-fold undersampling convolution matrices (in amplitude). The respective convolution kernels (in amplitude) that generate spectral leakage are plotted in (c) and (d). If the regular undersampling points are not shuffled enough, only part of the undersampling artifacts energy is spread, the rest of the energy remaining in weighted aliases. When there is just enough shuffling, all the undersampling artifacts energy is spread making jittered undersampling like random undersampling, yet controlling the size of the largest gap between two data points.

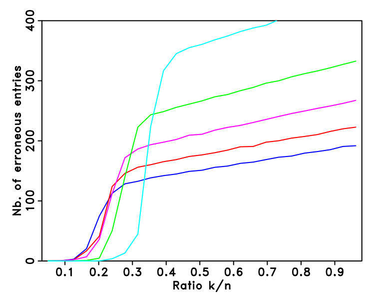
than 10^{-4}. This error accounts for both false positives and false negatives. The averaged results for the different experiments are summarized in Figures 4.6(a), 4.6(b), and 4.6(c), which correspond to regular, random, and optimally-jittered undersampling, respectively. The horizontal axes in these plots represent the relative underdeterminedness of the system, i.e., the ratio of the number k of nonzero entries in $\mathbf{x}_0$ to the number n of acquired data points. The vertical axes represent the average reconstruction error. The different curves represents the different undersampling factors. In each plot, the curves from top to bottom correspond to $\gamma = 2, \ldots, 6$.

Figure 4.6(a) shows that, regardless of the undersampling factor, there is no range of relative underdeterminedness for which $\mathbf{x}_0$ can accurately be recovered from a regular undersampling of $\mathbf{f}_0$. Sparsity is not enough to discriminate the signal components from the spectral leakage. The situation is completely different in Figures 4.6(b) and 4.6(c) for the random and optimally-jittered sampling. In this case, one can observed that exact recovery is possible for $0 < k/n \lesssim 1/4$. The main purpose of these plots is to qualitatively show the transition from successful to failed recovery. The quantitative interpretation for these diagrams to the right of the transition is less well understood but also observed in phase diagrams published in the literature (Donoho et al., 2006). A possible explanation for the observed behavior of the error lies in the nonlinear behavior of the solvers and on an error not measured in the ℓ_2 sense.

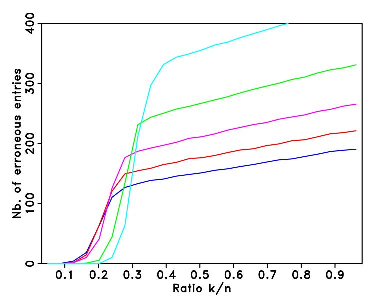
The key observations from these experiments are threefold. First, it is possible, under specific conditions, to *exactly* recover by sparsity-promoting inversion the original spectrum $\mathbf{x}_0$ of $\mathbf{f}_0$ from (very) few data points. Secondly, optimally-jittered undersampling behaves like random undersampling according to a discrete uniform distribution. For practical purposes, the former can thus be seen as equivalent to the latter. Thirdly, not all undersampling schemes for a given undersampling factor are comparable from a CS perspective. Regular undersampling is the most challenging. Random and optimally-jittered undersamplings according to a discrete uniform distribution are among the most favorable. In particular, if the signal is sufficiently sparse, these schemes lead to a reconstruction *as good as* dense regular sampling. Translated to the reconstruction of seismic wavefields, these results hint at a new nonlinear sampling theory based on a sparsifying transform for complex seismic data, e.g., the curvelet transform, and a coarse random sampling scheme that creates favorable recovery conditions for that transform, e.g., optimally-jittered undersampling.



(a)



(b)



(c)

Figure 4.6: Averaged recovery errors for a k-sparse Fourier vector reconstructed from n time samples taken (a) regularly, (b) randomly, and (c) optimally jittered from the model. In each plot, the curves from top to bottom correspond to an undersampling factor ranging from two to six, i.e., $\gamma = 2, \ldots, 6$.

4.3 Application to seismic data

Following recent work on Curvelet Reconstruction with Sparsity-promoting Inversion (CRSI - Herrmann and Hennenfent, 2007), seismic wavefields are reconstructed via $\tilde{\mathbf{f}} = \mathbf{C}^H\tilde{\mathbf{x}}$ where

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad \mathbf{y} = \mathbf{R}\mathbf{C}^H\mathbf{x}. \tag{4.9}$$

In this formulation, $\mathbf{C}$ is the discrete wrapping-based curvelet transform (Candès et al., 2005a). Similarly to any other data-independent transforms, curvelets do not provide a sparse representation of seismic data in the strict sense. Instead, the curvelet transform provides a compressible, arguably the most compressible (Hennenfent and Herrmann, 2006), representation. Compressibility means that most of the wavefield energy is captured by a few significant coefficients in the sparsifying domain. Since CS guarantees, for sparse-enough signal representations, the recovery of a fixed number of largest coefficients for a given undersampling factor (Candès et al., 2005b), a more compressible representation yields a better reconstruction, which explains the success of CRSI.

4.3.1 Synthetic data example

Figure 4.7(a) shows a synthetic dataset sampled above Nyquist rate along both the time and receiver axes. The corresponding amplitude spectrum is plotted in Figure 4.7(b). These two figures serve as references. Comparisons are made between the interpolation results of three-fold spatially undersampled data, collected either regularly or optimally-jittered. As expected, the amplitude spectrum (Figure 4.8(c)) of the regularly undersampled data (Figure 4.8(a)) is severely aliased. Unfortunately, these coherent f-k undersampling artifacts remain coherent in the curvelet domain and hence create a challenge for the reconstruction. To the contrary, there is no observable coherent spectral leakage in the amplitude spectrum (Figure 4.8(d)) for the optimally-jittered undersampled data (Figure 4.8(b)). Instead, the amplitude spectrum looks noisy in the temporal frequency band of the seismic signal.

Figure 4.9 shows the CRSI results for these two experiments. Figures 4.9(a) and 4.9(b) depict the reconstructions given data regularly (Figure 4.8(a)) and optimally-jittered (Figure 4.8(b)) sampled, respectively. Figures 4.9(c) and 4.9(d) represent the corresponding amplitude spectra. Unlike Figure 4.9(d) that is only slightly corrupted by incoherent errors, Figure 4.9(c) still contains substantial energy from the coherent undersampling artifacts.

This observation is corroborated by the respective signal-to-reconstruction-error-ratios of 6.91 dB and 10.42 dB. The signal-to-reconstruction-error-ratio, defined as $20 \cdot \log_{10}(\|\mathbf{f}_0\|_2/\|\mathbf{f}_0 - \tilde{\mathbf{f}}\|_2)$, accounts for the energy of the error but not its type. It is important to keep in mind that the difference in reconstruction quality is solely due to the difference in spatial sampling, the undersampling factor and the recovery procedure were kept the same. This behavior leads us to conclude that, for a given undersampling factor, spatial optimally-jittered undersampling is (much) more favorable for CRSI than regular undersampling.

In addition, Figure 4.10 shows a recovery experiment given randomly three-fold spatially undersampled data. Figure 4.10(a) depicts the simulated acquired data and Figure 4.10(b) the CRSI result. The signal-to-reconstruction-error-ratio is 9.72 dB. Figures 4.10(c) and 4.10(d) contain the corresponding amplitude spectra. As can be observed by comparing Figure 4.8(d) with Figure 4.10(c), both random and optimally-jittered samplings create favorable recovery conditions. However, the larger size of the acquisition gaps in randomly undersampled data deteriorates the overall performance of CRSI. This result corroborates the importance of controlling the size of the maximum gap in optimally-jittered undersampling for reconstruction with curvelets.

4.3.2 Field data example

The far-offsets of a regularly-sampled shot taken from a real marine dataset are considered. Our model consists of 255 traces separated by 6.25 m. The simulated data are obtained by three-fold undersampling this model either regularly (Figure 4.11(a)) or optimally-jittered (Figures 4.11(d)). In both cases, the nominal spatial sampling is 18.75 m. Again, the CRSI algorithm is applied (cf. equation 4.9). No assumption is made regarding the maximum dip in the data. Figures 4.11(b) and 4.11(e) show the CRSI results for the data plotted in Figures 4.11(a) and 4.11(d), respectively. Figures 4.11(c) and 4.11(f) show the differences scaled by a factor of four between the model and the CRSI results. The signal-to-reconstruction-errors are respectively 12.98 dB and 15.22 dB, which corroborates our observations from the synthetic data example. The performance of wavefield reconstruction by CRSI improves when the input data is optimally-jittered sampled.

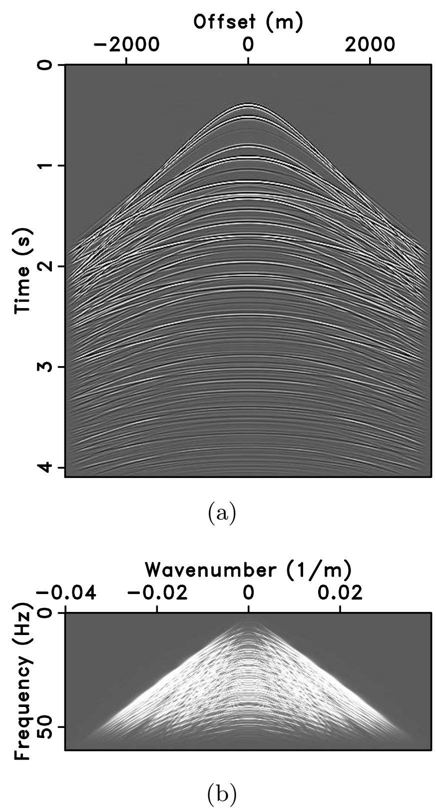
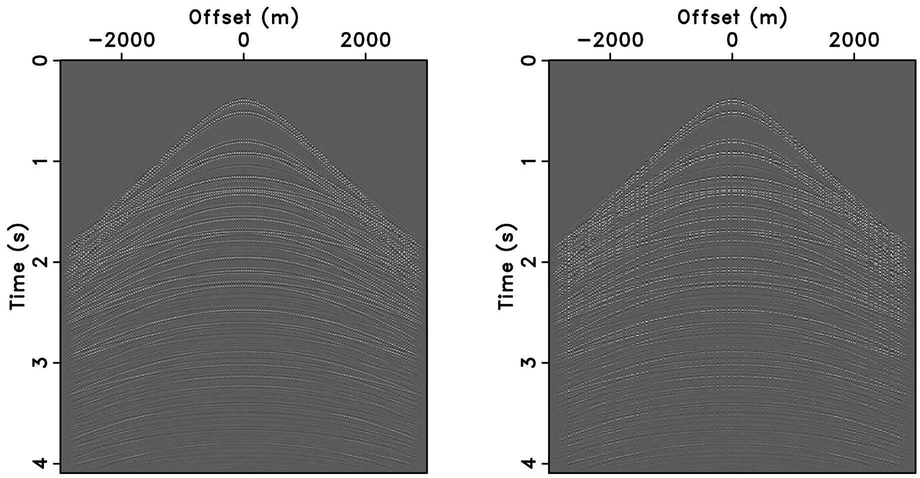
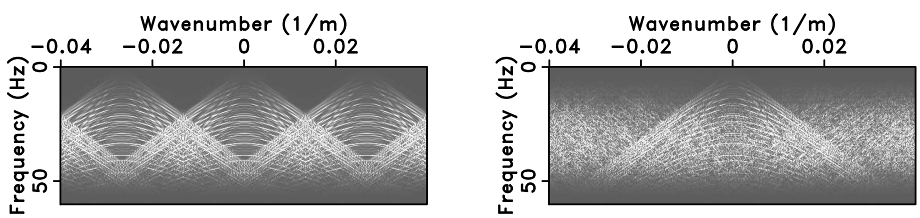


Figure 4.7: Reference model. (a) Synthetic data sampled above Nyquist rate and (b) corresponding amplitude spectrum.



(a)

(b)



(c)

(d)

Figure 4.8: Synthetic data of Figure 4.7 (a) regularly and (b) optimally-jittered three-fold undersampled along the spatial axis. Their respective amplitude spectra are plotted in (c) and (d). For the same amount of acquired data, optimally-jittered undersampling turns the harmful coherent undersampling artifacts of regular undersampling, i.e., aliases, into incoherent random noise.

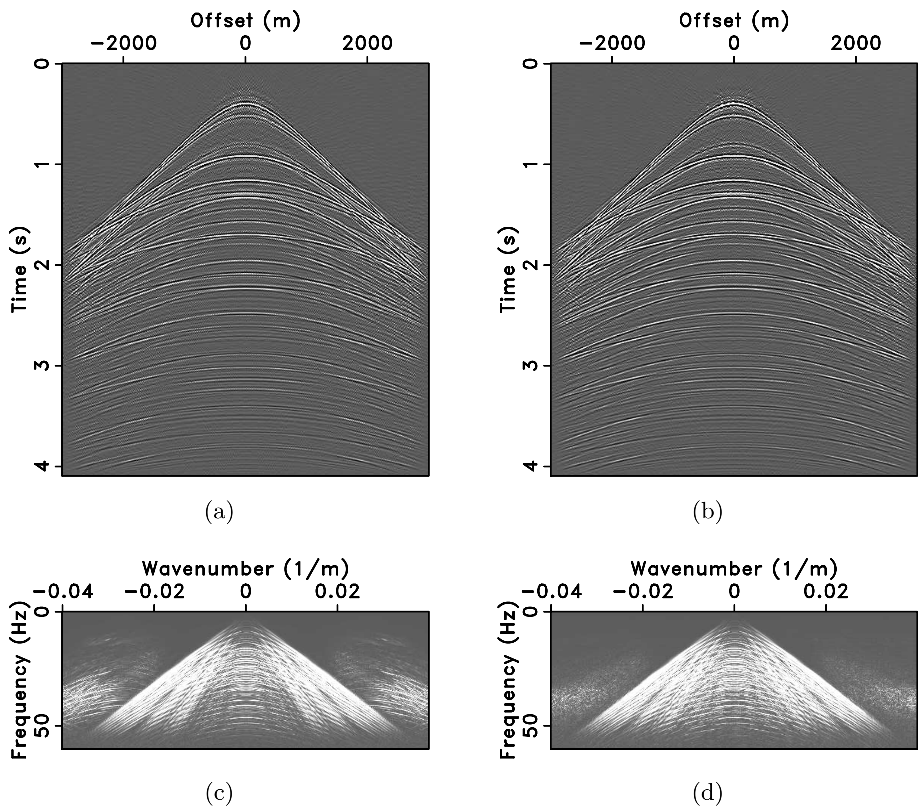
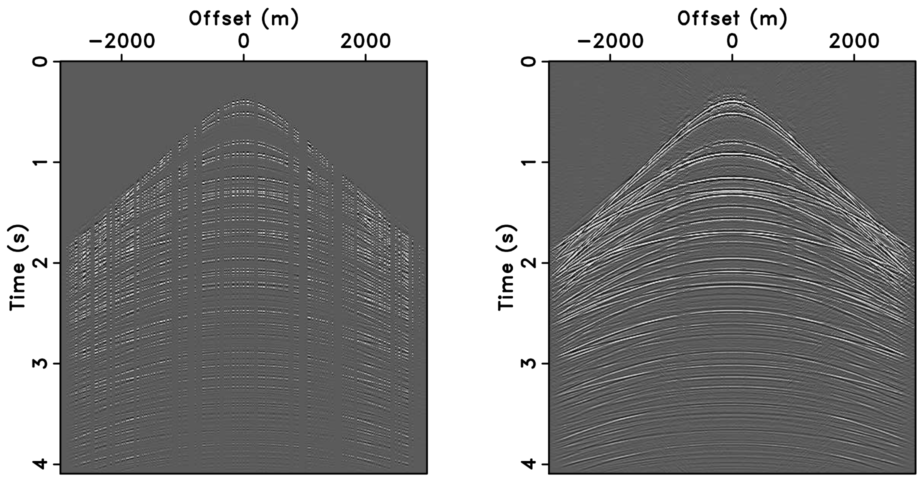
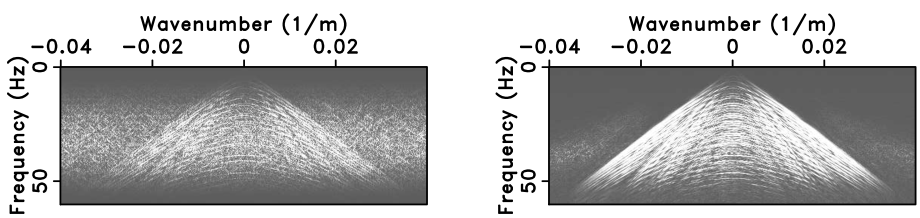


Figure 4.9: Curvelet reconstructions with sparsity-promoting inversion. Results given (a) data of Figure 4.8(a) and (b) data of Figure 4.8(b). The respective signal-to-reconstruction-error-ratios are 6.91 dB and 10.42 dB. For the same amount of data collected in the field, the reconstruction from optimally-jittered undersampled data is much more accurate than the reconstruction from regularly undersampled data.



(a) (b)



(c) (d)

Figure 4.10: Randomly undersampled data and curvelet reconstruction with sparsity-promoting inversion. (a) Synthetic data randomly three-fold undersampled along the spatial axis and (b) curvelet reconstruction with sparsity-promoting inversion. Their respective amplitude spectra are plotted in (c) and (d). The signal-to-reconstruction-error-ratio is 9.72 dB. Although random and optimally-jittered undersamplings create similar favorable recovery conditions (compare (c) with Figure 4.8(d)), the larger size of the acquisition gaps in the randomly undersampled data deteriorates the overall performance of CRSI.

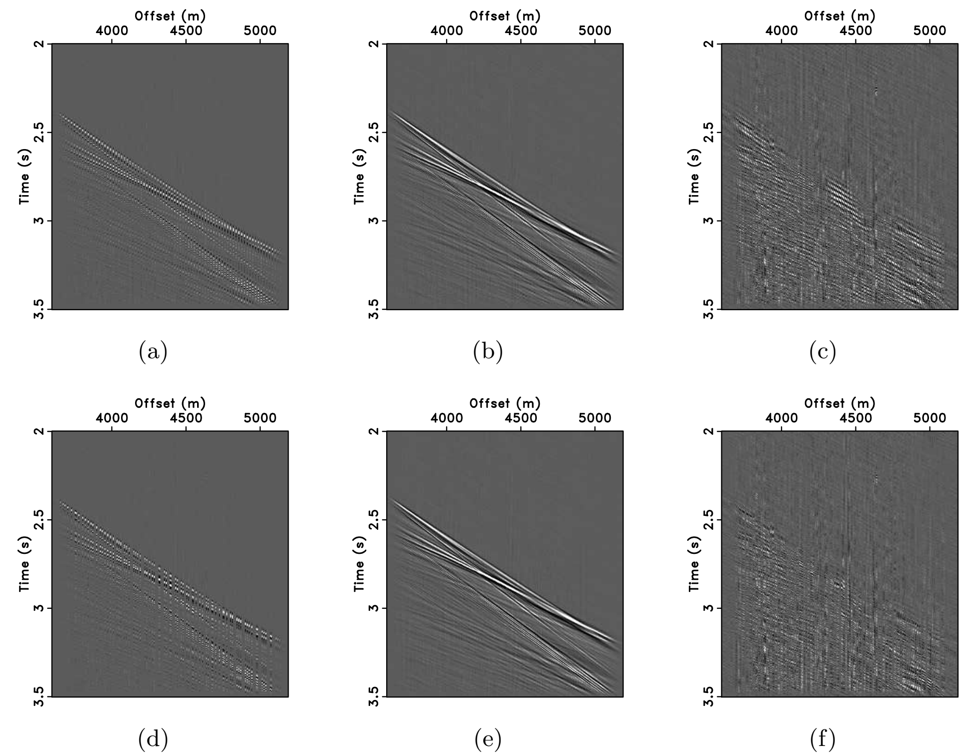


Figure 4.11: Field data example. The original data (not shown) is either (a) regularly or (d) optimally-jittered three-fold undersampled along the spatial coordinate. (b) and (e) are the curvelet reconstructions with sparsity-promoting inversion given data depicted in (a) and (d), respectively. (c) and (f) are differences scaled by a factor of four between the original data and the CRSI results (b) and (e), respectively. The corresponding signal-to-reconstruction-error-ratios are 12.98 dB and 15.22 dB, which corroborates that optimally-jittered undersampling is more favorable than regular undersampling.

4.4 Discussion

4.4.1 Undersampled data contaminated by noise

Although we focused on a noise-free (severely) underdetermined system of linear equations, the CS theory, and hence our work, both extend to the recovery from undersampled data contaminated by noise (Candès et al., 2005b). In this case, the noise $\mathbf{e}$ that corrupts the data adds to the undersampling artifacts in the sparsifying domain. The quantity that relates to the recoverability is now given by $\mathbf{A}^H(\mathbf{A}\mathbf{x}_0 + \mathbf{e}) - \alpha\mathbf{x}_0$ as opposed to $\mathbf{A}^H\mathbf{A}\mathbf{x}_0 - \alpha\mathbf{x}_0$ in the noise-free case. Consequently, the undersampling artifacts $\mathbf{z}$ and the imprint of the contaminating noise in the sparsifying domain, i.e., $\mathbf{A}^H\mathbf{e}$, have to be studied jointly.

4.4.2 From discrete to continuous spatial undersampling

So far, undersampling schemes based on an underlying fine interpolation grid were considered. This situation typically occurs when binning continuous randomly-sampled seismic data into small bins that define the fine grid used for interpolation. Despite the error introduced in the data, binning presents some computational advantages since it allows for the use of fast implementations of Fourier or Fourier-related transforms, e.g., FFTW (Frigo and Johnson, 1998) or FDCT (Candès et al., 2005a). However, binning can lead at the same time to an unfavorable undersampling scheme, e.g., regular or poorly-jittered. In this case, one should consider working on the original data with, e.g., an extension to the curvelet transform for irregular grids (Hennenfent and Herrmann, 2006). Despite the extra computational cost for the interpolation, continuous random sampling typically leads to improved interpolation results because it does not create coherent undersampling artifacts (Xu et al., 2005).

4.4.3 Sparsity-promoting solvers and jittered undersampling

The applicability of CS to the large-scale problems of exploration geophysics heavily relies on the implementation of an efficient ℓ_1 solver. Despite several recent attempts to overcome this bottleneck (Tibshirani, 1996; Figueiredo et al., 2007; van den Berg and Friedlander, 2007), a wide range of large-scale applications still uses approximate ℓ_1 solvers such as iterated re-weighted least-squares (IRLS - Gersztenkorn et al., 1986), stage-wise orthogonal matching pursuit (StOMP - Donoho et al., 2006), and iterative soft-thresholding with cooling (Hennenfent and Herrmann, 2005; Herrmann and Hennenfent, 2007) derived from Daubechies et al. (2004). The success and/or efficiency of these approximate solvers depends upon the implicit-or-explicit assumption that the MAI is incoherent. Because optimally-jittered undersampling creates such a MAI, these solvers can be used for the sparsity-promoting reconstruction with curvelets or other localized Fourier-based transforms. More importantly, jittered undersampling can be useful to evaluate the efficiency/robustness of (approximate) ℓ_1 solvers since the jitter parameter controls the amount of coherent energy that enters the MAI.

4.4.4 Generalization of the concept of undersampling artifacts

Undersampling artifacts are only one particular case of MAI that specifically occurs in the interpolation problem, i.e., $\mathbf{A} \stackrel{\text{def}}{=} \mathbf{R}\mathbf{S}^H$. The study we have done on these artifacts as a function of the restriction operator $\mathbf{R}$ can be extended to more general cases (see e.g. Lustig et al., 2007, in magnetic resonance imaging). For example, when $\mathbf{A}$ is defined as $\mathbf{A} \stackrel{\text{def}}{=} \mathbf{R}\mathbf{M}\mathbf{S}^H$ with $\mathbf{M}$ a modeling/demigration-like operator (Herrmann et al., 2007; Wang and Sacchi, 2007). In this case, $\mathbf{x}_0$ is the sparse representation of the Earth model in the $\mathbf{S}$ domain and $\mathbf{y}$ incomplete seismic data. The study of the MAI now determines which coarse spatial sampling schemes are more favorable than others in the context of sparsity-promoting migration/inversion. Based on observations in Zhou and Schuster (1995) and Sun et al. (1997), we believe that discrete random, optimally-jittered, and continuous random undersamplings will also play a key role.

4.5 Conclusions

Successful wavefield recovery depends on three key factors, namely, the existence of a sparsifying transform, a favorable sampling scheme and a sparsity-promoting recovery method. In this paper, we focused on an undersampling scheme that is designed for localized Fourier-like signal representations such as the curvelet transform. Our scheme builds on the fundamental observation that irregularities in sub-Nyquist sampling are good for nonlinear sparsity-promoting wavefield reconstruction algorithms because they turn harmful coherent aliases into relatively harmless incoherent random noise. The interpolation problem effectively becomes a much simpler denoising problem in this case.

Undersampling with a discrete random uniform distribution lacks control on the maximum gap size in the acquisition, which causes problems for transforms that consist of localized elements. Our jittered undersampling schemes remedy this lack of control, while preserving the beneficial properties of randomness in the acquisition grid. Our numerical findings on a stylized series of experiments confirm these theoretically-predicted benefits.

Curvelet-based wavefield reconstruction results from jittered undersampled synthetic and field datasets are better than results obtained from regularly decimated data. In addition, our findings indicate an improved performance compared to traces taken randomly according to an uniform distribution. This is a major result, with wide ranging applications, since it entails an increased probability for successful recovery with localized transform elements. In practice, this translates into more robust wavefield reconstruction.

4.6 Acknowledgments

G.H. thanks Ken Bube, Ramesh Neelamani, Warren Ross, Beatrice Vedel, and Ozgur Yilmaz for constructive discussions about this research. D.J. Verschuur and Chevron Energy Technology Company are gratefully thanked for the synthetic and real datasets, respectively. The authors thank the authors of CurveLab (`www.curvelet.org`) and the authors of SPGL1 (`www.cs.ubc.ca/labs/scl/spgl1`) for making their codes available. This paper was prepared with Madagascar, a reproducible research package (`rsf.sourceforge.net`). This work was in part financially supported by NSERC Discovery Grant 22R81254 and CRD Grant DNOISE 334810-05 of F.J.H. and was carried out as part of the SINBAD project with support, secured through ITF, from the following organizations: BG Group, BP, Chevron, ExxonMobil, and Shell.

We also appreciate the valuable comments and suggestions from the two reviewers and two associate editors.

Bibliography

Abma, R. and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, E91 – E97.

Bednar, J. B., 1996, Coarse is coarse of course unless...: The Leading Edge, **15**, 763 – 764.

Biondi, B., S. Fomel, and N. Chemingui, 1998, Azimuth moveout for 3D prestack imaging: Geophysics, **63**, 1177 – 1183.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2005a, Fast discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, 861–899.

Candès, E. J. and J. Romberg, 2006, Quantitative robust uncertainty principles and optimally sparse decompositions: Foundations of Computational Mathematics, **6**, 227 – 254.

Candès, E. J., J. Romberg, and T. Tao, 2005b, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Math, **99**, 1207 – 1223.

———, 2006, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: IEEE Transactions on Information Theory, **52**, 489 – 509.

Canning, A. and G. H. Gardner, 1996, Regularizing 3D data-sets with DMO: Geophysics, **61**, 1103 – 1114.

Claerbout, J. F., 1971, Towards a unified theory of reflector mapping: Geophysics, **36**, 467 – 481.

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, **LVII**, 1413 – 1457.

Dippe, M. and E. Wold, 1992, Stochastic sampling: theory and application: Progress in Computer Graphics, **1**, 1 – 54.

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

Donoho, D. L. and X. Huo, 2001, Uncertainty principles and ideal atomic decomposition: IEEE Transactions on Information Theory, **47**, 2845–2862.

Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck, 2006, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit: Technical report, Stanford Statistics Department. (TR-2006-2. `http://stat.stanford.edu/~idrori/StOMP.pdf`).

Figueiredo, M. A. T., R. D. Nowak, and S. J. Wright, 2007, Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems: Technical report, Instituto de Telecomunicacoes. (`http://www.lx.it.pt/~mtf/GPSR/Figueiredo_Nowak_Wright_twocolumn.pdf`).

Frigo, M. and S. G. Johnson, 1998, FFTW: An adaptive software architecture for the FFT: International Conference on Acoustics, Speech and Signal Processing, 1381 – 1384, IEEE.

Gersztenkorn, A., J. B. Bednar, and L. Lines, 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: Geophysics, **51**, 357 – 369.

Hennenfent, G. and F. J. Herrmann, 2005, Sparseness-constrained data continuation with frames: Applications to missing traces and aliased signals in 2/3-D: SEG International Exposition and 75^{th} Annual Meeting, 2162 – 2165.

———, 2006, Seismic denoising with non-uniformly sampled curvelets: Computing in Science and Engineering, **8**, 16 – 25.

Herrmann, F. J. and G. Hennenfent, 2007, Non-parametric seismic data recovery with curvelet frames: Technical report, UBC Earth & Ocean Sciences Department. (TR-2007-1. `http://slim.eos.ubc.ca/Publications/Public/Journals/CRSI.pdf`).

Herrmann, F. J., D. Wang, G. Hennenfent, and P. P. Moghaddam, 2007, Curvelet-based seismic data processing: a multiscale and nonlinear approach. (Accepted for publication in Geophysics. `http://slim.eos.ubc.ca/Publications/Public/Journals/curveletter.pdf`).

Leneman, O., 1966, Random sampling of random processes: Impulse response: Information and Control, **9**, 347 – 363.

Lustig, M., D. L. Donoho, and J. M. Pauly, 2007, Sparse MRI: The application of compressed sensing for rapid MR imaging: Magnetic Resonance in Medicine. (In press. `http://www.stanford.edu/~mlustig/SparseMRI.pdf`).

Sacchi, M. D., T. J. Ulrych, and C. J. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: IEEE Transactions on Signal Processing, **46**, 31 – 38.

Spitz, S., 1991, Seismic trace interpolation in the F-X domain: Geophysics, **67**, 890 – 794.

Stolt, R. H., 2002, Seismic data mapping and reconstruction: Geophysics, **67**, 890 – 908.

Sun, Y., G. T. Schuster, and K. Sikorski, 1997, A Quasi-Monte Carlo approach to 3-D migration: Theory: Geophysics, **62**, 918 – 928.

Tibshirani, R., 1996, Regression shrinkage and selection via the Lasso: Journal of the Royal Statistical Society, **58**, 267 – 288.

Trad, D. O., J. Deere, and S. Cheadle, 2005, Challenges for land data interpolation: Presented at the CSEG National Convention.

Trad, D. O. and T. J. Ulrych, 1999, Radon transform: beyond aliasing with irregular sampling: Presented at the Sixth International Congress of the Brazilian Geophysical Society.

Trad, D. O., T. J. Ulrych, and M. D. Sacchi, 2003, Latest view of sparse Radon transforms: Geophysics, **68**, 386–399.

van den Berg, E. and M. P. Friedlander, 2007, In pursuit of a root: Technical report, UBC Computer Science Department. (TR-2007-16. `http://www.optimization-online.org/DB_FILE/2007/06/1708.pdf`).

Verdu, S., 1998, Multiuser detection: Cambridge University Press.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166 – 1177.

Wang, J. and M. D. Sacchi, 2007, High-resolution wave-equation amplitude-variation-with-ray-parameter (AVP) imaging with sparseness constraints: Geophysics, **72**, S11 – S18.

Xu, S., Y. Zhang, D. Pham, and G. Lambare, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, V87 – V95.

Zhou, C. and G. T. Schuster, 1995, Quasi-random migration of 3-D field data: SEG Technical Program Expanded Abstracts, 1145 – 1148.

Zwartjes, P. M. and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased data: Geophysics, **72**, V21–V32.

Chapter 5

New insights into one-norm solvers from the Pareto curve

5.1 Introduction

Many geophysical inverse problems are ill posed (Parker, 1994)—their solutions are not unique or are acutely sensitive to changes in the data. To solve this kind of problem stably, additional information must be introduced. This technique is called *regularization* (see, e.g., Phillips, 1962; Tikhonov, 1963).

Specifically, when the solution of an ill-posed problem is known to be (almost) sparse, Oldenburg et al. (1983) and others have observed that a good approximation to the solution can be obtained by using one-norm regularization to promote sparsity. More recently, results in information theory have breathed new life into the idea of promoting sparsity to regularize ill-posed inverse problems. These results establish that, under certain conditions, the sparsest solution of a (severely) underdetermined linear system can be *exactly* recovered by seeking the minimum one-norm solution (Candès et al., 2006; Donoho, 2006; Rauhut, 2007). This has led to tremendous activity in the newly established field of *compressed sensing*. Several new one-norm solvers have appeared in response (see, e.g., Daubechies et al., 2004; van den Berg and Friedlander, 2008, and references therein). In the context of geophysical applications, it is a challenge to evaluate and compare these solvers against more standard approaches such as iteratively reweighted least-squares (IRLS - Gersztenkorn et al., 1986), which uses a quadratic approximation to the one-norm regularization function.

In this letter, we propose an approach to understand the behavior of algorithms for solving one-norm regularized problems. The approach consists of tracking on a graph the data misfit versus the one norm of successive

A version of this chapter has been accepted for publication. G. Hennenfent, E. van den Berg, M.P. Friedlander, and F.J. Herrmann. New insights into one-norm solvers from the Pareto curve. *Geophysics*, 2008.

iterates. The *Pareto curve* traces the optimal tradeoff in the space spanned by these two axes and gives a rigorous yardstick for measuring the quality of the solution path generated by an algorithm. In the context of the two-norm—i.e., Tikhonov—regularization, the Pareto curve is often plotted on a log-log scale and is called the L-curve (Lawson and Hanson, 1974). We draw on the work of van den Berg and Friedlander (2008) who examine the theoretical properties of the one-norm Pareto curve. Our goal is to understand the compromises implicitly accepted when an algorithm is given a limited number of iterations.

5.2 Problem statement

Consider the following underdetermined system of linear equations

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n}, \tag{5.1}$$

where the n-vectors $\mathbf{y}$ and $\mathbf{n}$ represent observations and additive noise, respectively. The n-by-N matrix $\mathbf{A}$ is the modeling operator that links the model $\mathbf{x}_0$ to the noise-free data given by $\mathbf{y} - \mathbf{n}$. We assume that $N \gg n$ and that $\mathbf{x}_0$ has few nonzero or significant entries. We use the terms "model" and "observations" in a broad sense, so that many linear geophysical problems can be cast in the form shown in equation 5.1. In the case of wavefield reconstruction, for example, $\mathbf{y}$ is the acquired seismic data with missing traces, $\mathbf{A}$ can be the restriction operator combined with the curvelet synthesis operator so that $\mathbf{x}_0$ is the curvelet representation of the fully-sampled wavefield (Herrmann and Hennenfent, 2008; Hennenfent and Herrmann, 2008).

Because $\mathbf{x}_0$ is assumed to be (almost) sparse, one can promote sparsity as a prior via one-norm regularization to overcome the singular nature of $\mathbf{A}$ when estimating $\mathbf{x}_0$ from $\mathbf{y}$. A common approach is to solve the convex optimization problem

$$\mathrm{QP}_\lambda : \quad \min_{\mathbf{x}} \ \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda\|\mathbf{x}\|_1,$$

which is closely related to quadratic programming (QP); the positive parameter λ is the Lagrange multiplier, which balances the tradeoff between the two norm of the data misfit and the one norm of the solution. Many algorithms are available for solving QP_λ, including IRLS, iterative soft thresholding (IST), introduced by Daubechies et al. (2004), and the IST extension to include cooling (ISTc - Figueiredo and Nowak, 2003), which was tailored to geophysical applications by Herrmann and Hennenfent (2008).

It is generally not clear, however, how to choose the parameter λ such that the solution of QP_λ is, in some sense, optimal. A directly related optimization problem, the basis pursuit (BP) denoise problem (Chen et al., 1998), minimizes the one norm of the solution given a maximum misfit, and is given by

$$\mathrm{BP}_\sigma: \quad \min_{\mathbf{x}} \ \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \sigma.$$

This formulation is often preferred when an estimate of the noise level $\sigma \geq 0$ in the data is available. BP_σ can be solved using ISTc or the spectral projected-gradient algorithm (SPGℓ_1) introduced by van den Berg and Friedlander (2008).

For interest, a third optimization problem, connected to QP_λ and BP_σ, minimizes the misfit given a maximum one norm of the solution, and is given by the LASSO (LS) problem (Tibshirani, 1996)

$$\mathrm{LS}_\tau: \quad \min_{\mathbf{x}} \ \tfrac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tau.$$

Because an estimate of the one norm of the solution $\tau \geq 0$ is typically not available for geophysical problems, this formulation is seldom used directly. It is, however, a key internal problem used by SPGℓ_1 in order to solve BP_σ.

To understand the connection between these approaches and compare their related solvers in different scenarios, we propose to follow Daubechies et al. (2007) and van den Berg and Friedlander (2008) and look at the Pareto curve.

5.3 Pareto curve

Figure 5.1 gives a schematic illustration of a Pareto curve. The curve traces the optimal tradeoff between $\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2$ and $\|\mathbf{x}\|_1$ for a specific pair of $\mathbf{A}$ and $\mathbf{y}$ in equation 5.1. Point 1 clarifies the connection between the three parameters of QP_λ, BP_σ, and LS_τ. The coordinates of a point on the Pareto curve are (τ, σ) and the slope of the tangent at this point is $-\lambda$. The end points of the curve—points 2 and 3—are two special cases. When $\tau = 0$, the solution of LS_τ is $\mathbf{x} = 0$ (point 2). It coincides with the solutions of BP_σ with $\sigma = \|\mathbf{y}\|_2$ and QP_λ with $\lambda = \|\mathbf{A}^H\mathbf{y}\|_\infty / \|\mathbf{y}\|_2$. (The infinity norm $\|\cdot\|_\infty$ is given by $\max(|\cdot|)$.) When $\sigma = 0$, the solution of BP_σ (point 3) coincides with the solutions of LS_τ, where τ is the one norm of the solution, and QP_λ, where $\lambda = 0^+$—i.e., λ infinitely close to zero from above. These relations are formalized as follows in van den Berg and Friedlander (2008):

Result 1 *The Pareto curve i) is convex and decreasing, ii) is continuously differentiable, and iii) has a negative slope* $\lambda = \|\mathbf{A}^H\mathbf{r}\|_\infty/\|\mathbf{r}\|_2$ *with the residual* $\mathbf{r}$ *given by* $\mathbf{y} - \mathbf{Ax}$.

For large-scale geophysical applications, it is not practical (or even feasible) to sample the entire Pareto curve. However, its regularity, as implied by this result, means that it is possible to obtain a good approximation to the curve with very few interpolating points, as illustrated later in this letter.

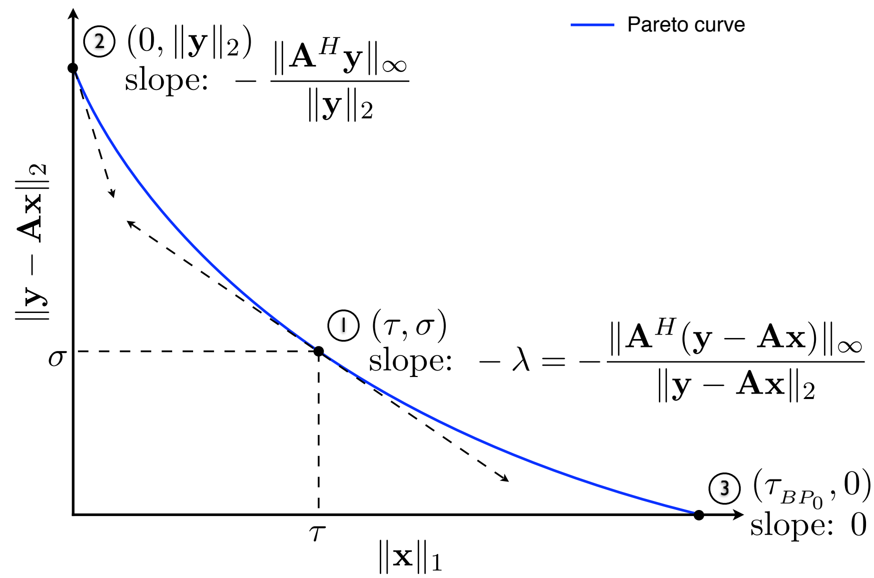


Figure 5.1: Schematic illustration of a Pareto curve. Point ① exposes the connection between the three parameters of QP_λ, BP_σ, and LS_τ. Point ③ corresponds to a solution of BP_σ with $\sigma = 0$.

5.4 Comparison of one-norm solvers

To illustrate the usefulness of the Pareto curve, we compare IST, ISTc, SPGℓ_1, and IRLS on a noise-free problem and compute a solution of BP_σ for $\sigma = 0$, i.e., BP_0. This case is especially challenging for solvers that attack QP_λ—e.g., IST, ISTc and IRLS—because the corresponding solution can only be attained in the limit as $\lambda \to 0$.

We construct a benchmark problem that is typically used in the compressed sensing literature (Donoho et al., 2006). The matrix $\mathbf{A}$ is taken to have Gaussian independent and identically-distributed entries; a sparse solution $\mathbf{x}_0$ is randomly generated, and the "observations" $\mathbf{y}$ are computed according to equation 5.1.

5.4.1 Solution paths

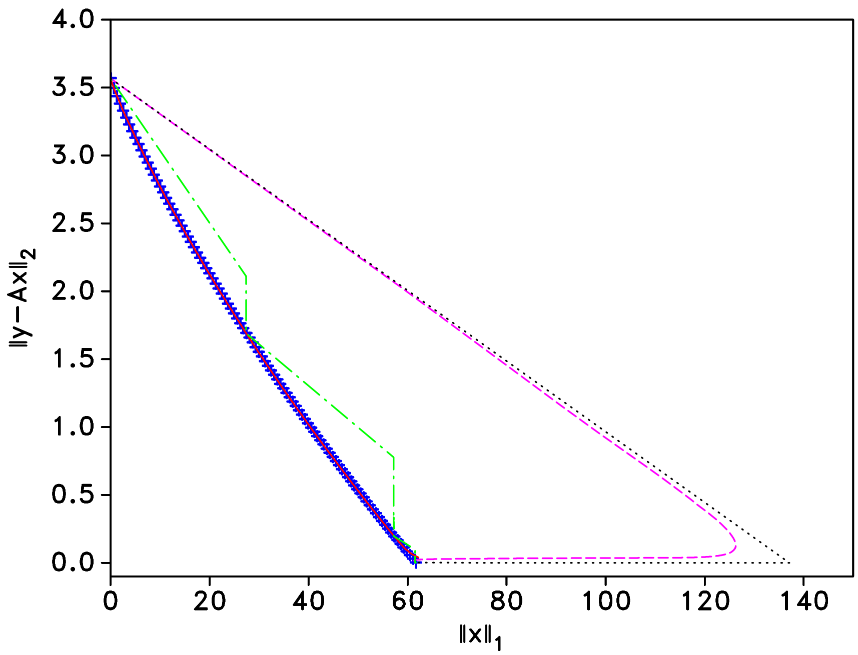


Figure 5.2: Pareto curve and solution paths (large enough number of iterations) of four solvers for a BP_0 problem. The symbols + represent a sampling of the Pareto curve. The solid (—) line, obscured by the Pareto curve, is the solution path of ISTc, the chain ($- \cdot -$) line the path of SPGLℓ_1, the dashed ($- -$) line the path of IST, and the dotted ($\cdots$) line the path of IRLS.

Figure 5.2 shows the solution paths of the four solvers as they converge to the BP_0 solution. The starting vector provided to each solver is the zero vector, and hence the paths start at $(0, \|\mathbf{y}\|_2)$—point 2 in Figure 5.1. The number of iterations is large enough for each solver to converge, and therefore the solution paths end at $(\tau_{BP_0}, 0)$—point 3 in Figure 5.1.

The two solvers SPGℓ_1 and ISTc approach the BP_0 solution from the left and remain close to the Pareto curve. In contrast, IST and IRLS aim at a least-squares solution before turning back towards the BP_0 solution. ISTc solves QP_λ for a decreasing sequence $\lambda_i \to 0$. The starting vector for QP_{λ_i} is the solution of $QP_{\lambda_{i-1}}$, which is by definition on the Pareto curve. This explains why ISTc so closely follows the curve. SPGℓ_1 solves a sequence of LS_τ problems for an increasing sequence of $\tau_i \to \tau_{BP_0}$, hence the vertical segments along the SPGℓ_1 solution path. IST solves QP_{0^+}. Because there is hardly any regularization, IST first works towards minimizing the data misfit. When the data misfit is sufficiently small, the effect of the one-norm penalization starts, yielding a change of direction towards the BP_0 solution. IRLS solves a sequence of weighted, damped, least-squares problems. Because the weights are initialized to ones, IRLS first reaches the standard least-squares solution. The estimates obtained from the subsequent reweightings have a smaller one norm while maintaining the residual (close) to zero. Eventually, IRLS gets to the BP_0 solution.

5.4.2 Practical considerations

In geophysical applications, problem sizes are large and there is a severe computational constraint. We can use the technique outlined above to understand the robustness of a given solver that is limited by a maximum number of iterations or matrix-vector products that can be performed.

Figure 5.3 shows the Pareto curve and the solution paths of the various solvers where the maximum number of iterations is fixed. This roughly equates to using the same number of matrix-vector products for each solver. Whereas SPGℓ_1 continues to provide a fairly accurate approximation to the BP_0 solution, those computed by IST, ISTc, and IRLS suffer from larger errors. IST stops before the effect of the one-norm regularization kicks in; hence the data misfit at the candidate solution is small but the one norm is completely incorrect. ISTc and IRLS accumulate small errors along their paths because there are not enough iterations to solve each subproblem to sufficient accuracy. Note that both solvers accumulate errors along both axes.

5.5 Geophysical example

As a concrete example of the use of the Pareto curve in the geophysical context, we study the problem of wavefield reconstruction with sparsity-promoting inversion in the curvelet domain (CRSI - Herrmann and Hennenfent, 2008). The simulated acquired data, shown in Figure 5.4(a), corresponds to a shot record with 35% of the traces missing. The interpolated

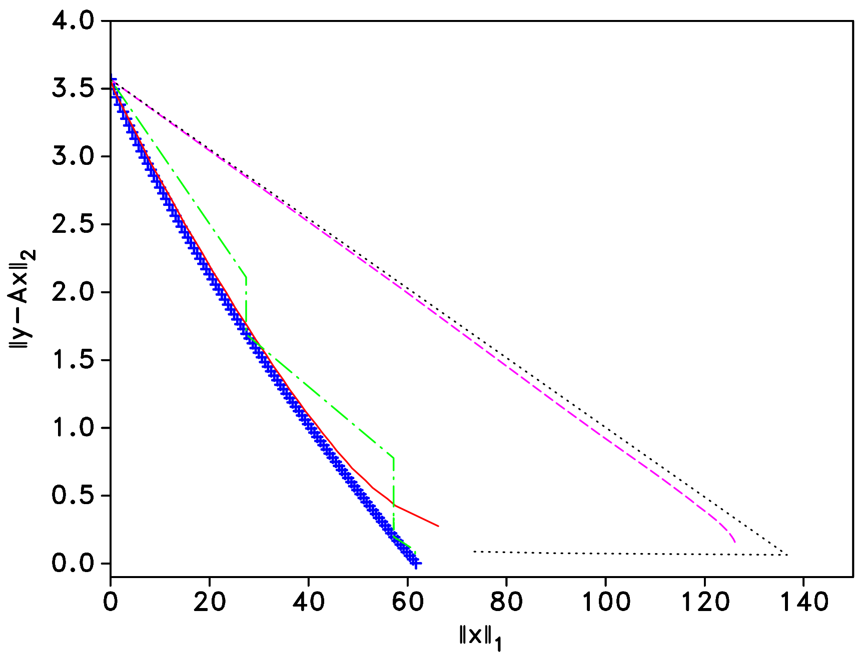


Figure 5.3: Pareto curve and optimization paths (same, limited number of iterations) of four solvers for a BP_0 problem (see Figure 5.2 for legend).

result, shown in Figure 5.4(b), is obtained by solving BP_0 using SPGℓ_1. This problem has more than half a million unknowns and forty-two thousand data points.

The points in Figure 5.5 are samples of the corresponding Pareto curve. The regularity of these points strongly indicates that the underlying curve—which we know to be convex—is smooth and well behaved, and empirically supports our earlier claim. However problems of practical interest are often significantly larger, and it may be prohibitively expensive to compute a similarly fine sampling of the curve.

Because the curve is well behaved, we can leverage its smoothness and use a small set of samples to obtain a good interpolation. The solid line in Figure 5.5 shows an interpolation based only on information from the circled samples. The interpolated curve closely matches the samples that were not included in the interpolation. The figure also plots the iterates taken by SPGℓ_1 in order to obtain the reconstruction shown in Figure 5.4(b). The plot shows that the iterates remain to the Pareto curve and that they convergence towards the BP_0 solution.

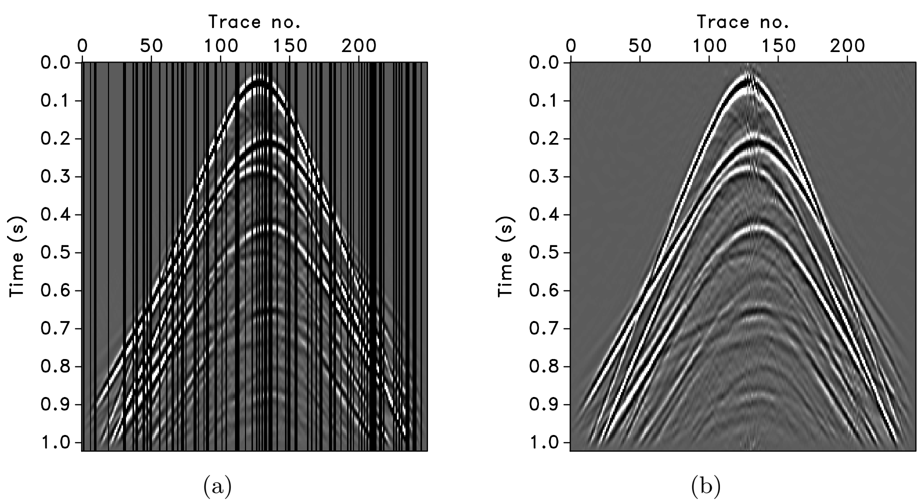


Figure 5.4: CRSI on synthetic data. (a) Input and (b) interpolated data using CRSI with SPGℓ_1.

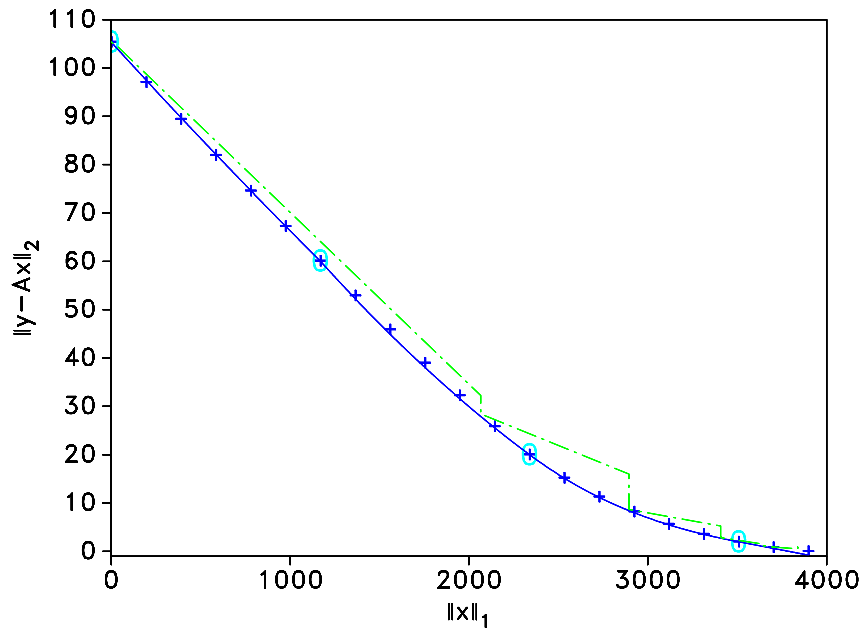


Figure 5.5: Pareto curve and SPGℓ_1 solution path for a CRSI problem. The symbols + represent a fine, accurate sampling of the Pareto curve. The solid (—) line is an approximation to the Pareto curve using the few, circled points, the chain ($- \cdot -$) line the solution path of SPGℓ_1.

5.6 Conclusions

The sheer size of seismic problems makes it a certainty that there will be significant constraints on the amount of computation that can be done when solving an inverse problem. Hence it is especially important to explore the nature of a solver's iterations in order to make an informed decision on how to best truncate the solution process. The Pareto curve serves as the optimal reference, which makes an unbiased comparison between different one-norm solvers possible.

Of course, in practice it is prohibitively expensive to compute the entire Pareto curve exactly. We observe, however, that the Pareto curves for many of the one-norm regularized problems are regular, as confirmed by the theoretical Result 1. This suggests that it is possible to approximate the Pareto curve by fitting a curve to a small set of sample points, taking into account derivative information at these points. As such, the insights from the Pareto curve can be leveraged to large-scale one-norm regularized problems, as we illustrate on a geophysical example. This prospect is particularly exciting given the current resurgence of this type of regularization in many different areas of research.

5.7 Acknowledgments

The authors are grateful to Sergey Fomel and Tamas Nemeth for their valuable input, and to Eric Verschuur for the synthetic data. The authors also thank the anonymous reviewers and associate editor for their comments that certainly helped improve this letter. This publication was prepared using Madagascar (`rsf.sf.net`), a package for reproducible computational experiments, SPGℓ_1 (`cs.ubc.ca/labs/scl/spgl1`), and Sparco (`cs.ubc.ca/labs/scl/sparco`), a suite of linear operators and problems for testing algorithms for sparse signal reconstruction. This research was in part financially supported by NSERC Discovery Grant 22R81254 of F.J.H. and by CRD Grant DNOISE 334810-05 of F.J.H. and M.P.F., and was carried out as part of the SINBAD project with support, secured through ITF, from the following organizations: BG Group, BP, Chevron, ExxonMobil, and Shell.

Bibliography

van den Berg, E. and M. P. Friedlander, 2008, Probing the Pareto frontier for basis pursuit solutions: Technical Report TR-2008-01, UBC Computer Science Department. (`http://www.optimization-online.org/DB_HTML/2008/01/1889.html`).

Candès, E. J., J. Romberg, and T. Tao, 2006, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: IEEE Transactions on Information Theory, **52**, no. 2, 489–509.

Chen, S. S., D. L. Donoho, and M. A. Saunders, 1998, Atomic decomposition by basis pursuit: SIAM Journal on Scientific Computing, **20**, no. 1, 33–61.

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, **LVII**, 1413–1457.

Daubechies, I., M. Fornasier, and I. Loris, 2007, Accelerated projected gradient method for linear inverse problems with sparsity constraints: ArXiv e-prints, **706**, no. 0706.4297. (`http://adsabs.harvard.edu/abs/2007arXiv0706.4297D`).

Donoho, D. L., 2006, Compressed sensing: IEEE Transactions on Information Theory, **52**, no. 4, 1289–1306.

Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck, 2006, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit: Technical Report TR-2006-2, Stanford Statistics Department. (`http://stat.stanford.edu/~idrori/StOMP.pdf`).

Figueiredo, M. and R. Nowak, 2003, An EM algorithm for wavelet-based image restoration: IEEE Transactions on Image Processing, **12**, no. 8, 906–916.

Gersztenkorn, A., J. B. Bednar, and L. Lines, 1986, Robust iterative inversion for the one-dimensional acoustic wave equation: Geophysics, **51**, no. 2, 357–369.

Hennenfent, G. and F. J. Herrmann, 2008, Simply denoise: wavefield reconstruction via jittered undersampling: Geophysics, **73**, no. 3.

Herrmann, F. J. and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International. (doi:10.1111/j.1365-246X.2007.03698.x).

Lawson, C. L. and R. J. Hanson, 1974, Solving least squares problems: Prentice Hall.

Oldenburg, D., T. Scheuer, and S. Levy, 1983, Recovery of the acoustic impedance from reflection seismograms: Geophysics, **48**, no. 10, 1318–1337.

Parker, R. L., 1994, Geophysical inverse theory: Princeton University Press.

Phillips, D. L., 1962, A technique for the numerical solution of certain integral equations of the first kind: Journal of the Association for Computing Machinery, **9**, no. 1, 84–97.

Rauhut, H., 2007, Random sampling of sparse trigonometric polynomials: Applied and Computational Harmonic Analysis, **22**, no. 1, 16–42.

Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and regularization method: Soviet mathematics - Doklady, **4**, 1035–1038.

Tibshirani, R., 1996, Regression shrinkage and selection via the LASSO: Journal Royal Statististics, **58**, no. 1, 267–288.

Chapter 6

Curvelet-based seismic data processing

6.1 Introduction

In this letter, we demonstrate that the discrete curvelet transform (Candès et al., 2006a; Hennenfent and Herrmann, 2006b) can be used to reconstruct seismic data from incomplete measurements, to separate primaries and multiples and to restore migration amplitudes. The crux of the method lies in the combination of the curvelet transform, which attains a fast decay for the magnitude-sorted curvelet coefficients, with a sparsity promoting program. By themselves sparsity-promoting programs are not new to the geosciences (Sacchi et al., 1998). However, sparsity promotion with the curvelet transform is new. The curvelet transform's unparalleled ability to detect wavefront-like events that are locally linear and coherent means it is particularly well suited to seismic data problems. In this paper, we show examples including data regularization (Hennenfent and Herrmann, 2006a, 2007a), primary-multiple separation (Herrmann et al., 2007a) and migration-amplitude recovery (Herrmann et al., 2007b). Application of this formalism to wavefield extrapolation is presented elsewhere (Lin and Herrmann, 2007).

6.2 Curvelets

Curvelets are localized 'little plane-waves' (see Hennenfent and Herrmann, 2006b, and the on-line ancillary material for an introduction on this topic) that are oscillatory in one direction and smooth in the other direction(s). They are multiscale and multi-directional. Curvelets have an

A version of this chapter has been published. F.J. Herrmann, D. Wang, G. Hennenfent, and P.P. Moghaddam. Curvelet-based seismic data processing: a multiscale and nonlinear approach. *Geophysics*, 73(1):A1-A5, January-February 2008.

anisotropic shape – they obey the so-called parabolic scaling relationship, yielding a width $\propto$ length^2 for the support of curvelets in the physical domain. This anisotropic scaling is necessary to detect wavefronts and explains their high compression rates on seismic data and images (Candès et al., 2006a; Herrmann et al., 2007b), as long as these datasets can be represented as functions with events on piece-wise twice differentiable curves. Then, the events become linear at the fine scales justifying an approximation by the linearly shaped curvelets. Even seismic data with caustics, pinch-outs, faults or strong amplitude variations fit this model, which amounts to a preservation of the sparsity attained by curvelets.

Curvelets represent a specific tiling of the 2-D/3-D frequency domain into strictly localized wedges. Because the directional sampling increases every-other scale doubling, curvelets become more anisotropic at finer scales. Curvelets compose multi-D data according to $\mathbf{f} = \mathbf{C}^T\mathbf{C}\mathbf{f}$ with $\mathbf{C}$ and $\mathbf{C}^T$ the forward and inverse discrete curvelet transform matrices (defined by the fast discrete curvelet transform, FDCT, with wrapping, a type of periodic extenstion, see Candès et al., 2006a; Ying et al., 2005). The symbol ^{T} represents the transpose, which is equivalent to the inverse for this choice of curvelet transform. This transform has a moderate redundancy (a factor of roughly 8 in 2-D and 24 in 3-D) and a computational complexity of $\mathcal{O}(n \log n)$ with n the length of $\mathbf{f}$. Even though $\mathbf{C}^T\mathbf{C} = \mathbf{I}$, with $\mathbf{I}$ the identity matrix, the converse is not true, i.e., $\mathbf{C}\mathbf{C}^T \neq \mathbf{I}$. This ambiguity can be removed by adding sparsity promotion as a constraint.

6.3 Common problem formulation by Sparsity-promoting inversion

Our solution strategy is built on the premise that seismic data and images have a sparse representation, $\mathbf{x}_0$, in the curvelet domain. To exploit this property, our forward model reads

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n} \tag{6.1}$$

with $\mathbf{y}$ a vector of noisy and possibly incomplete measurements; $\mathbf{A}$ the modeling matrix that includes $\mathbf{C}^T$; and $\mathbf{n}$, a zero-centered white Gaussian noise. Because of the redundancy of $\mathbf{C}$ and/or the incompleteness of the data, the matrix $\mathbf{A}$ can not readily be inverted. However, as long as the data, $\mathbf{y}$, permits a sparse vector, $\mathbf{x}_0$, the matrix, $\mathbf{A}$, can be inverted by a sparsity-promoting program (Candès et al., 2006b; Donoho, 2006):

$$\mathbf{P}_\epsilon : \qquad \begin{cases} \widetilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \\ \widetilde{\mathbf{f}} = \mathbf{S}^T \widetilde{\mathbf{x}} \end{cases} \tag{6.2}$$

in which ϵ is a noise-dependent tolerance level, $\mathbf{S}^T$ the inverse transform and $\widetilde{\mathbf{f}}$ the solution calculated from the vector $\widetilde{\mathbf{x}}$ (the symbol $\widetilde{}$ denotes a vector obtained by nonlinear optimization) minimizing $\mathbf{P}_\epsilon$. The difference between $\widetilde{\mathbf{x}}$ and $\mathbf{x}_0$ is proportional to the noise level.

Nonlinear programs $\mathbf{P}_\epsilon$ are not new to seismic data processing as in spiky deconvolution (Taylor et al., 1979; Santosa and Symes, 1986) and Fourier transform-based interpolation (Sacchi et al., 1998). The curvelets' high compression rate makes the nonlinear program $\mathbf{P}_\epsilon$ perform well when $\mathbf{C}^T$ is included in the modeling operator. Despite its large-scale and nonlinearity, the solution of the convex problem $\mathbf{P}_\epsilon$ can be approximated with a limited (< 250) number of iterations of a threshold-based cooling method derived from work by Figueiredo and Nowak (2003); Daubechies et al. (2004); Elad et al. (2005). At each iteration the descent update ($\mathbf{x} \leftarrow \mathbf{x} + \mathbf{A}^T\big(\mathbf{y} - \mathbf{A}\mathbf{x}\big)$), minimizing the quadratic part of Equation 6.2, is followed by a soft thresholding ($\mathbf{x} \leftarrow T_\lambda(\mathbf{x})$ with $T_\lambda(x) := \operatorname{sgn}(x) \cdot \max(0, |x| - |\lambda|)$) for decreasing threshold levels λ. This soft thresholding on the entries of the unknown curvelet vector captures the sparsity and the cooling, which speeds up the algorithm, allows additional coefficients to fit the data.

6.4 Seismic data recovery

The reconstruction of seismic wavefields from regularly-sampled data with missing traces is a setting where a curvelet-based method will perform well. As with other transform-based methods, sparsity is used to reconstruct the wavefield by solving $\mathbf{P}_\epsilon$. It is also shown that the recovery performance can be increased when information on the major primary arrivals is included in the modeling operator.

6.4.1 Curvelet-based recovery

The reconstruction of seismic wavefields from incomplete data corresponds to the inversion of the picking operator $\mathbf{R}$. This operator models missing data by inserting zero traces at source-receiver locations where data is missing passing recorded traces unchanged. The task of the recovery is to undo this operation by filling in the zero traces. Since seismic data is sparse in the curvelet domain, the missing data can be recovered by compounding the picking operator with the curvelet modeling operator, i.e., $\mathbf{A} := \mathbf{RC}^T$. With this definition for the modeling operator, solving $\mathbf{P}_\epsilon$ corresponds to seeking the sparsest curvelet vector whose inverse curvelet transform, followed by the picking, matches the data at the nonzero traces. Applying the inverse transform (with $\mathbf{S} := \mathbf{C}$ in $\mathbf{P}_\epsilon$) gives the interpolated data. For details on the conditions that determine successful recovery, refer to Hennenfent and Herrmann (2007a,b) and Herrmann and Hennenfent (2007).

An example of curvelet-based recovery is presented in Figure 6.1 which shows the results of decimating, and then reconstructing, a seismic dataset. The original shot and receiver spacings were 25m, and 80 % of the traces were thrown out at random (see Figure 6.1(b)). Comparing the 'ground truth' in Figure 6.1(a) with the recovered data in Figure 6.1(c) shows a successful recovery in case the high-frequencies are removed. Aside from sparsity in the curvelet domain, no prior information was used during the recovery, which is quite remarkable. Part of the explanation lies in the curvelet's ability to locally exploit the 3-D geometry of the data and this suggests why curvelets are successful for complex datasets where other methods may fail.

6.4.2 Focused recovery

In practice, additional information on the to-be-recovered wavefield is often available. For instance, one may have access to the predominant primary arrivals or to the velocity model. In that case, the recently introduced *focal* transform (Berkhout and Verschuur, 2006), which 'deconvolves' the data with an estimate of the primaries, incorporates this additional information into the recovery process. Application of this primary operator, $\mathbf{\Delta P}$, adds a wavefield interaction with the surface, mapping primaries to first-order surface-related multiples (Verschuur and Berkhout, 1997; Herrmann, 2007). Inversion of this operator, strips the data off one interaction with the surface, focusing primary energy to (directional) sources. This focusing corresponds to a collapse of the 3-D primary events to an approximate line source which has a sparser representation in the curvelet domain.

By compounding the non-adaptive, data-independent, curvelet transform with the data-adaptive *focal* transform, i.e., $\mathbf{A} := \mathbf{R\Delta PC}^T$, the recovery can be improved by solving $\mathbf{P}_\epsilon$. The solution of $\mathbf{P}_\epsilon$ now entails the inversion of $\mathbf{\Delta P}$, yielding the sparsest set of curvelet coefficients that matches the incomplete data when 'convolved' with the primaries. Applying the inverse curvelet transform, followed by 'convolution' with $\mathbf{\Delta P}$ yields the interpolation, i.e. $\mathbf{S}^T := \mathbf{\Delta PC}^T$. Comparing the curvelet recovery with the focused curvelet recovery (Figure 6.1(c) and 6.1(d)) shows an overall

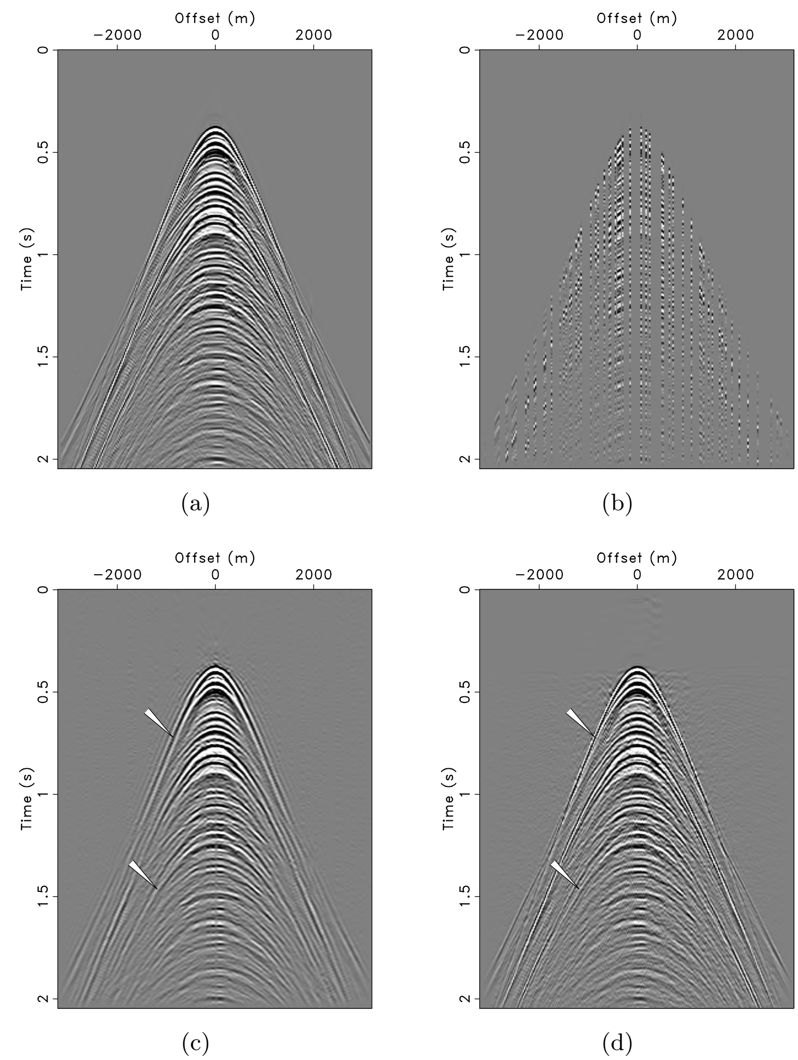


Figure 6.1: Comparison between 3-D curvelet-based recovery by sparsity-promoting inversion with and without focusing. **(a)** Fully sampled real SAGA data shot gather. **(b)** Randomly subsampled shot gather from a 3-D data volume with 80 % of the traces missing in the receiver and shot directions. **(c)** Curvelet-based recovery. **(d)** Curvelet-based recovery with focusing. Notice the improvement (denoted by the arrows) from the focusing with the primary operator.

improvement in the recovered details.

6.5 Seismic signal separation

Predictive multiple suppression involves two steps, namely multiple prediction and primary-multiple separation. In practice, the second step appears difficult and adaptive least-squares ℓ_2-matched-filtering techniques are known to lead to residual multiple energy, high frequency jitter and deterioration of the primaries (Herrmann et al., 2007a). By employing the curvelet's ability to detect wavefronts with conflicting dips (e.g. caustics), a non-adaptive, independent of the total data, separation scheme can be defined that is robust with respect to moderate errors in the multiple prediction. The nonlinear program, $\mathbf{P}_\epsilon$, with $\mathbf{y}$ defined by the total data, can be adapted to separate multiples from primaries by replacing the ℓ_1 norm by a weighted ℓ_1 norm, i.e., $\|\mathbf{x}\|_1 \mapsto \|\mathbf{x}\|_{1,\mathbf{w}} = \sum_\mu |w_\mu x_\mu|$ with μ running over all curvelets and $\mathbf{w}$ a vector with positive weights. By defining these weights proportional to the magnitude of the curvelet coefficients of the 2-D SRME-predicted multiples, the solution of $\mathbf{P}_\epsilon$ with $\mathbf{A} := \mathbf{C}^T$ removes multiples. Primaries and multiples naturally separate in the curvelet domain and the weighting further promotes this separation while solving $\mathbf{P}_\epsilon$. The weights that are fixed during the optimization penalize the entries in the curvelet vector for which the predicted multiples are significant. The emphasis on the weights versus the data misfit (the proportionality constant) is user defined. The estimate for the primaries is obtained by inverse curvelet transforming the curvelet vector that minimizes $\mathbf{P}_\epsilon$ for the weighted ℓ_1 norm ($\mathbf{A} = \mathbf{S}^T := \mathbf{C}^T$).

Figure 6.2 shows an example of 3-D curvelet-based primary-multiple separation of a North Sea dataset with the weights set according to the curvelet-domain magnitudes of the SRME-predicted multiples multiplied by 1.25. Comparison between the estimates for the primaries from adaptive subtraction by ℓ_2-matched filtering (Verschuur and Berkhout, 1997) and from our nonlinear and non-adaptive curvelet-based separation shows an improvement in (i) the elimination of the focused multiple energy below shot location 1000 m, induced by out-of-plane scattering due to small 3-D variations in the multiple-generating reflectors and (ii) an overall improved continuity and noise reduction. This example demonstrates that the multi-scale and multi-angular curvelet domain can be used to separate primaries and multiples given an inaccurate prediction for the multiples. However, the separation goes at the expense of a moderate loss of primary energy which compares favorably compared to the loss associated with ℓ_2-matched filtering (see also Herrmann et al., 2007a).

6.6 Migration-amplitude recovery

Restoring migration amplitudes is another area where curvelets can be shown to play an important role. In this application, the purpose is to replace computationally expensive amplitude recovery methods, such as least-squares migration (Nemeth et al., 1999; Kuhl and Sacchi, 2003), by an amplitude scaling (Guitton, 2004). This scaling can be calculated from a demigrated-migrated reference vector close to the actual reflectivity.

In order to exploit curvelet sparsity, we propose to scale in the curvelet domain. This choice seems natural because migrated images suffer from spatially varying and dip-dependent amplitude deterioration that can be accommodated by curvelets. The advantages of this approach are manifold and include (i) a correct handling of reflectors with conflicting dips and (ii) a stable curvelet sparsity-promoting inversion of the diagonal that restores the amplitudes and removes the clutter by exploiting curvelet sparsity on the model.

The method is based on the approximate identity: $\mathbf{K}^T\mathbf{K}\mathbf{r} \approx \mathbf{C}^T\mathbf{D}_r\mathbf{C}\mathbf{r}$ with $\mathbf{K}$ and $\mathbf{K}^T$ the demigration, migration operators and $\mathbf{D}_r$ a reference-model specific scaling (Herrmann et al., 2007b). By defining the modeling matrix as $\mathbf{A} := \mathbf{C}^T\sqrt{\mathbf{D}_r}$, $\mathbf{P}_\epsilon$ can be used to recover the migration amplitudes from the migrated image. Possible spurious side-band effects and erroneously detected curvelets (Candès and Guo, 2002) are removed by supplementing the ℓ_1 norm in $\mathbf{P}_\epsilon$ with an anisotropic diffusion norm (Fehmers and Höcker, 2003). This norm enhances the continuity along the imaged reflectors and removes spurious artifacts.

Results for the SEG AA' dataset (O'Brien and Gray, 1996; Aminzadeh et al., 1997) are summarized in Figure 6.3. These results are obtained with a reverse-time 'wave-equation' finite-difference migration code. To illustrate the recovery performance, idealized seismic data is generated by demigration, followed by adding white Gaussian noise, yielding a signal-to-noise ratio (SNR) of only 3 dB. This data is subsequently migrated and used as input. Despite the poor SNR, the image in Figure 6.3(a) contains most reflectors, which can be explained by the redundancy of the data, the migration operator's sophistication (diffractions at the bottom of the salt are handled correctly) and the perfect 'match' between the demigration and migration operators. However, the noise gives rise to clutter and there is dimming of

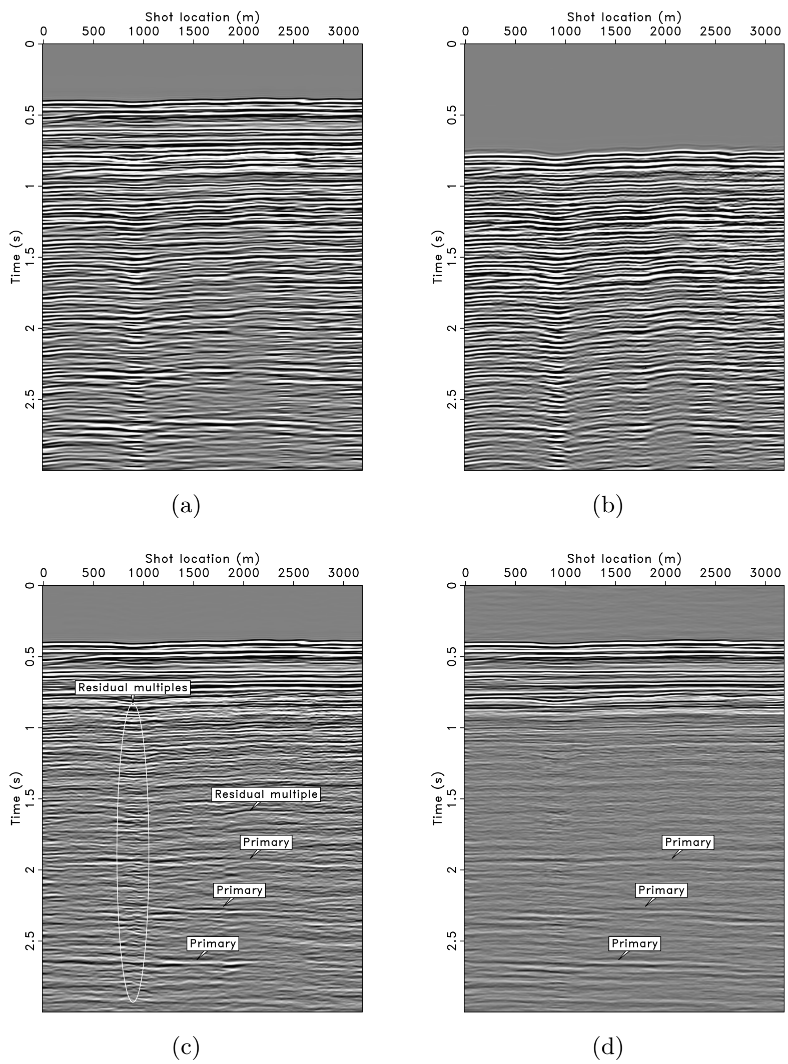


Figure 6.2: 3-D Primary-multiple separation with $\mathbf{P}_\epsilon$ for the SAGA dataset. **(a)** Near-offset section including multiples. **(b)** The SRME-predicted multiples. **(c)** The estimated primaries according to ℓ_2-matched filtering. **(d)** The estimated primaries obtained with $\mathbf{P}_\epsilon$. Notice the improvement, in areas with small 3-D effects (ellipsoid) and residual multiples.

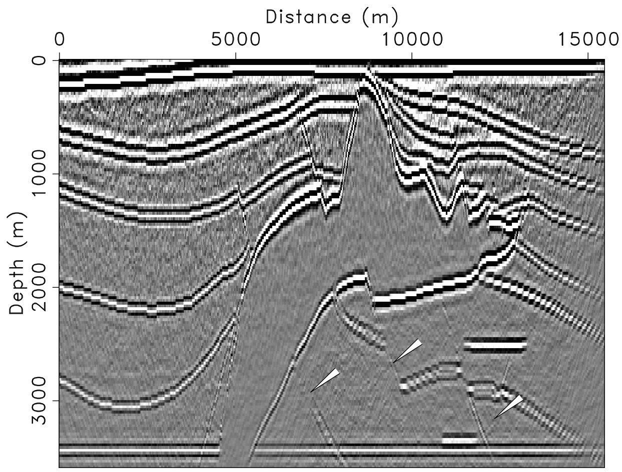
the amplitudes, in particular for steep dips under the salt. Nonlinear recovery removes most of this clutter and more importantly the amplitudes for the sub-salt steep-dipping events are mostly restored. This idealized example shows how curvelets can be used to recover the image amplitudes. As long as the background velocity model is sufficiently smooth and the reflectivity sufficiently sparse, this recovery method can be expected to perform well even for more complex images.

6.7 Discussion and conclusions

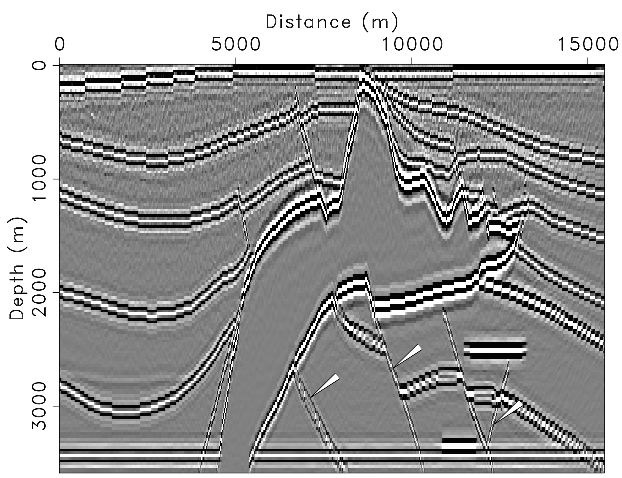
The presented examples show that problems in data acquisition and imaging can be solved with a common problem formulation during which sparsity in the curvelet domain is promoted. For curved wavefront-like features that oscillate in one direction and that are smooth in the other direction(s), curvelets attain high compression rates while other transforms do not necessarily achieve sparsity for these geometries. Seismic images of sedimentary basins and seismic wave arrivals in the data both behave in this fashion, so that curvelets are particularly valuable for compression. It is this compression that underlies the success of our sparsity promoting formulation. First, we showed on real data that missing data can be recovered by solving a nonlinear optimization problem where the data misfit and the ℓ_1-norm on the curvelet coefficients are simultaneously minimized. This recovery is improved further with a combined curvelet-focal transform. Sparsity also proved essential during the primary-multiple separation. In this case, it leads to a form of decorrelation of primaries and multiples, reducing the probability of having large overlapping curvelet entries between these different events. Finally, the sparsity of curvelets on the image itself was exploited to recover the migration amplitudes of the synthetic subsalt imaging example. Through these three examples, the successful application of curvelets, enhanced with sparsity-promoting inversion, opens new perspectives on seismic data processing and imaging. The ability of curvelets to detect wavefront-like features is key to our success and opens an exciting new outlook towards future developments in exploration seismology.

6.8 Acknowledgments

The authors would like to thank D.J. Verschuur and C. Stolk for their input in the primary-multiple separation and migration-amplitude recovery. We also would like to thank the authors of CurveLab (`www.curvelet.org`)



(a)



(b)

Figure 6.3: Image amplitude recovery for a migrated image calculated from noisy data (SNR 3 dB). **(a)** Image with clutter. **(b)** Image after nonlinear recovery. The clearly visible non-stationary noise in **(a)** is mostly removed during the recovery while the amplitudes are also restored. Steeply dipping reflectors (denoted by the arrows) under the salt are also well recovered.

and W. Symes for his reverse-time migration code. The examples were prepared with Madagascar (`rsf.sf.net`), supplemented by SLIMpy operator overloading, developed by S. Ross Ross. Norsk Hydro is thanked for the field dataset. M. O'Brien, S. Gray and J. Dellinger are thanked for the SEG AA' data. This work was in part financially supported by the NSERC Discovery (22R81254) and CRD Grants DNOISE (334810-05) of F.J.H. and was carried out as part of the SINBAD project with support, secured through ITF, from BG Group, BP, Chevron, ExxonMobil and Shell. Finally, the authors would like to thank the anonymous reviewers whose constructive comments helped improve this letter.

Bibliography

Aminzadeh, F., J. Brac, and T. Kunz, 1997, 3-D Salt and Overthrust Model. SEG/EAGE 3-D Modeling Series, No. 1: Society of Exploration Geophysicists, Tulsa.

Berkhout, A. J. and D. J. Verschuur, 2006, Focal transformation, an imaging concept for signal restoration and noise removal: Geophysics, **71**, A55–A59.

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: SIAM Multiscale Modeling and Simulation, **5**, 861–899.

Candès, E. J. and F. Guo, 2002, New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction: Signal Processing, **82**, 1519–1543.

Candès, E. J., J. K. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics, **59**, 1207–1223.

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

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

Elad, M., J. L. Starck, P. Querre, and D. L. Donoho, 2005, Simultaneous Cartoon and Texture Image Inpainting using Morphological Component Analysis (MCA): Journal of Applied and Computational Harmonic Analysis, **19**, 340–358.

Fehmers, G. C. and C. F. W. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, **68**, 1286–1293.

Figueiredo, M. and R. Nowak, 2003, An EM algorithm for wavelet-based image restoration: IEEE Transactions on Image Processing, **12**, 906–916.

Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: Geophysics, **69**, 1017–1024.

Hennenfent, G. and F. J. Herrmann, 2006a, Application of stable signal recovery to seismic interpolation: Presented at the SEG International Exposition and 76th Annual Meeting.

———, 2006b, Seismic denoising with non-uniformly sampled curvelets: IEEE Computing in Science and Engineering, **8**, 16–25.

———, 2007a, Irregular sampling: from aliasing to noise: Presented at the EAGE 69th Conference & Exhibition.

———, 2007b, Random sampling: new insights into the reconstruction of coarsely-sampled wavefields: Presented at the SEG International Exposition and 77th Annual Meeting.

Herrmann, F. J., 2007, Surface related multiple prediction from incomplete data: Presented at the EAGE 69th Conference & Exhibition.

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007a, Nonlinear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, 781–799.

Herrmann, F. J. and G. Hennenfent, 2007, Non-parametric seismic data recovery with curvelet frames: Technical report, UBC Earth and Ocean Sciences Department. (TR-2007-1).

Herrmann, F. J., P. P. Moghaddam, and C. Stolk, 2007b, Sparsity- and continuity-promoting seismic imaging with curvelet frames: Journal of Applied and Computational Harmonic Analysis. (Accepted for publication).

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

Lin, T. and F. J. Herrmann, 2007, Compressed wavefield extrapolation: Geophysics. (Accepted for publication).

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

O'Brien, M. and S. Gray, 1996, Can we image beneath salt?: The Leading Edge, **15**, 17–22.

Sacchi, M. D., T. J. Ulrych, and C. Walker, 1998, Interpolation and extrapolation using a high-resolution discrete Fourier transform: IEEE Transactions on Signal Processing, **46**, 31–38.

Santosa, F. and W. Symes, 1986, Linear inversion of band-limited reflection seismogram: SIAM Journal on Scientific and Statistical Computing, **7**, 1307–1330.

Taylor, H. L., S. C. Banks, and J. F. McCoy, 1979, Deconvolution with the ℓ_1 norm: Geophysics, **44**, 39–52.

Verschuur, D. J. and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, part II: practical aspects and examples: Geophysics, **62**, 1596–1611.

Ying, L., L. Demanet, and E. J. Candés, 2005, 3-D discrete curvelet transform: Wavelets XI, Expanded Abstracts, 591413, SPIE.

Chapter 7

Conclusions

In this chapter we summarize the main contributions of this thesis and discuss some limitations of the work presented. We also suggest follow-up work as well as possible extensions.

7.1 Main contributions

The topic of this thesis is seismic data interpolation. The approach we advocate is to view seismic data from a geometrical perspective. We identify a transform, called the curvelet transform (Candès and Donoho, 2004), to that effect and use it in a new formulation of the wavefield reconstruction problem. This formulation, coined curvelet reconstruction with sparsity-promoting inversion (CRSI), is solved using a large-scale one-norm solver that we introduce and study using the Pareto curve. The reported results on synthetic and real data show that CRSI outperforms other methods but the results also reveal that CRSI's performance depends on the acquisition pattern. We leverage this observation towards the development of a coarse sampling scheme, termed jittered undersampling, that creates, under specific circumstances, favorable recovery conditions for CRSI.

The remainder of this section provides more details about the aforementioned contributions.

7.1.1 Curvelets for seismic data

We use the curvelet transform to exploit the high-dimensional and strong geometrical structure of seismic data. The curvelet transform (Candès and Donoho, 2004), designed to represent curve-like singularities optimally, decomposes seismic data into a superposition of localized plane waves, called curvelets. These curvelets are shaped according to a parabolic scaling law and have different frequency contents and dips to match locally the wavefront at best. These properties guarantee a sparse—arguably the sparsest—data-independent representation of seismic data. In other words, the superposition of only a "few" curvelets captures most of the energy of real seismic data as shown in chapter 2 using our extension of the fast discrete curvelet transform (FDCT - Candès et al., 2006) to irregularly sampled data.

7.1.2 Curvelet reconstruction with sparsity-promoting inversion

Following ideas from compressive sampling (Donoho, 2006; Candès et al., 2006) and existing interpolation algorithms that promote sparsity in a transform domain (Sacchi et al., 1998; Zwartjes and Sacchi, 2007), we formulate a new optimization problem, coined curvelet reconstruction with sparsity-promoting inversion (CRSI), to reconstruct seismic data (chapter 3). In words, CRSI takes as inputs: i) the acquired data, ii) a mask that spatially locates the acquired traces, and iii) an interpolation grid. CRSI returns the sparsest set of curvelet coefficients that explain the acquired data. The interpolated data is reconstructed via the (weighted) inverse curvelet transform of this set.

From a theoretical standpoint, the success of CRSI depends, of course, on the validity of the sparseness assumption but also on the severity of the undersampling, and on the way the data is acquired. The latter point is of particular interest because i) it allows us to give a new interpretation to the minimum velocity constraint that is already successfully used in other interpolation methods, and ii) it motivates the development (chapter 4) of a coarse sampling scheme, termed jittered undersampling, that creates, under specific circumstances, favorable recovery conditions for CRSI. We further discuss this topic in section 7.1.3.

From a practical standpoint, CRSI would not be possible without a robust large-scale one-norm solver. We introduce iterative soft thresholding with cooling (ISTc) to that effect (chapter 3). ISTc is an extension of the iterative soft thresholding algorithm proposed by Daubechies et al. (2004). It reaches an approximation to the desired solution in a (very) limited number of iterations by solving a carefully-chosen sequence of sub-problems. Each of these optimization sub-problems becomes increasingly harder to solve but benefits from an approximate solution of the previous problem as a “warm” start. The solution path is studied in more detail using the Pareto curve (chapter 5). We further discuss this topic in section 7.1.4.

Reported results illustrate that CRSI performs well on synthetic and real data sets and comparatively better than other methods (chapter 3 and Hennenfent and Herrmann, 2006b). We also show on synthetic data that the quality of the reconstruction improves with the dimensionality of the problem.

7.1.3 Wavefield reconstruction via jittered undersampling

The performance of CRSI depends on the acquisition pattern. We explain this phenomenon (chapter 4) by looking at the interpolation problem from a denoising perspective as suggested by Donoho et al. (2006). Indeed, undersampling seismic data in the physical domain translates into adding noise to its curvelet representation. Hence, interpolating consists in separating the undersampling noise from the few significant curvelet coefficients that represent the full data. Because this separation is done by promoting sparsity—i.e., signal's representation is the few large entries—problems arise if an acquisition pattern creates sparse undersampling noise.

We leverage this new insight towards the development of a coarse sampling scheme, termed jittered undersampling, for which CRSI performs at best. At the core of this work is a noise-shaping problem. We show that, under specific circumstances, jittered undersampling creates incoherent random noise in the Fourier and curvelet domains. Furthermore, its construction avoids large acquisition gaps. The combination of these two properties proves to be key in the formulation of a versatile sparsity-promoting wavefield recovery scheme in the curvelet domain as illustrated on a series of examples.

7.1.4 Insights into one-norm solvers from the Pareto curve

We introduce the Pareto curve as a means to understand the behavior and evaluate the performance of one-norm solvers (chapter 5). The technique consists of tracking on a graph the data misfit versus the one norm of successive iterates. By comparing the solution paths to the Pareto curve—the best possible tradeoff between data misfit and sparsity—we are able to assess the performance of the solvers and the quality of the solutions. This prospect is particularly exciting given the current resurgence of one-norm regularization in many different areas of research. In geophysics, such an assessment is relevant, for example, to understand the compromises implicitly accepted when an algorithm is given a limited number of iterations.

Reported results show that ISTc is a robust and reasonably accurate solver under limited number of iterations. These results also reveal that the recently-introduced spectral projected-gradient algorithm (SPGℓ_1 - van den Berg and Friedlander, 2007) could be an interesting alternative to ISTc if its algorithmic complexity scales well with the size of problems.

7.1.5 Curvelet-based seismic data processing

Beside the seismic wavefield reconstruction problem, we recast a few other processing steps—signal separation, migration-amplitude recovery, and deconvolution—in a sparsity-promoting program that exploits the high degree of sparsity attained by curvelets on seismic data and images (chapter 6 and Hennenfent et al., 2005b,a). The promising results obtained shows that the insights gained from the developments of CRSI can be leveraged towards a much broader range of applications. This prospect opens an exciting new outlook towards future developments in exploration seismology.

7.2 Follow-up work

We suggest a few ideas that go beyond the reported experiments.

7.2.1 Interpolation comparisons on complex data

CRSI was tested on different data sets and, in some cases, the results were compared to those of competing algorithms (chapter 3 and Hennenfent and Herrmann, 2006b). We recommend to study further the algorithm on a broader range of complex data. Preliminary experiments on data with strong aliased ground-roll (Hennenfent and Herrmann, 2006a; Yarham et al., 2007) show, for example, that CRSI performs well and may have a competitive advantage over other interpolation methods. Another type of interesting data that comes in mind is data containing diffractions.

7.2.2 Interpolation impact on processing flow

We evaluate the quality of the reported results by comparing the interpolated wavefield to the true wavefield, if available. Although this comparison gives a precise idea of the quality of the reconstruction, it does not measure the impact on processing steps following interpolation—e.g., multiple prediction and elimination—and on what matters most, the final subsurface image. Hence, we recommend to include CRSI in a complete processing flow and compare the final image to the one obtained using a standard flow.

7.3 Current limitations

We examine both the practical and the fundamental weaknesses of the current CRSI, which motivates the extensions we propose in the next section.

7.3.1 Curvelet code

The CRSI results presented in this thesis were obtained using the FDCT based on the wrapping of specially selected Fourier samples (Candès et al., 2005). This implementation breaks down the input image or volume into a number of scales depending on the length of the shortest axis. In other words, if one axis is much shorter than the others, the decomposition along the long axes is unnecessarily limited. Despite an increased implementation complexity, an alternative would be to treat separately the different axes, an idea also proposed by H. Douma (personal communication, 2007). This alteration of the curvelet code would immediately improve, for example, the interpolation of 3D data in the shot domain if the cross-line axis is much shorter than the in-line and time axes.

A more fundamental limitation of the FDCT is related to the redundancy of the transform. Indeed, the FDCT is around 8-redundant in 2D and around 24-redundant in 3D, which precludes, at least for now, tractable higher-dimensional FDCTs. Lu and Do (2007) propose a less redundant N-dimensional ($N \geq 2$) implementation, termed surfacelet transform, by combining a directional filter bank with a multiscale pyramid. However, preliminary results using surfacelets for wavefield reconstruction are not as good as CRSI results (E. Lebed, personal communication, 2007). Another option is to combine the curvelet transform with another transform (Herrmann, 2003; Neelamani et al., 2008) to reduce redundancy and reach higher dimensions. The different treatment of the axes is unsatisfactory in several applications (see, e.g., Neelamani et al., 2008), though. For interest, Kutyniok and Labate (2005) propose yet another N-dimensional ($N \geq 2$) transform, called shearlet transform, but no discrete implementation is available at this point to determine the redundancy and the effectiveness of shearlets for wavefield reconstruction.

7.3.2 CRSI

In chapter 4 we show that CRSI is sensitive to the size of the acquisition gaps. Indeed, CRSI uses localized elements—curvelets—to represent seismic data. If the physical support of these elements is smaller than the acquisition gap (Figure 7.1), these elements will not enter the solution even though they might be useful to interpolate an event obvious to the human eye. We discuss in the next section possible extensions of CRSI to overcome this particular issue.

In chapter 4 we also show that CRSI performs better on irregularly

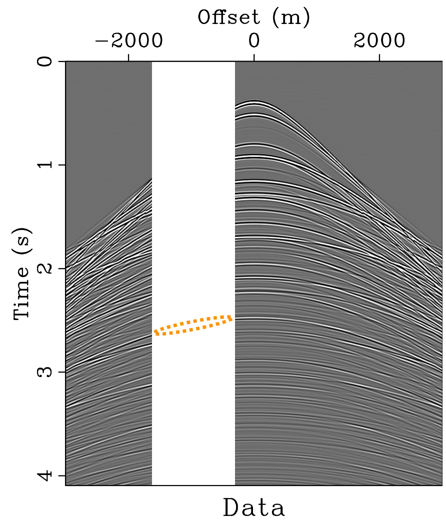


Figure 7.1: Curvelets and large acquisition gap. If the physical support of a curvelet is smaller than the acquisition gap, this curvelet will not participate to the CRSI solution even though this element might be useful to interpolate an event obvious to the human eyes.

undersampled data than on regularly undersampled data. The difference comes from the effectiveness of the sparsity prior to discriminate signal from undersampling noise in either case. Hence, there is an intrinsic difficulty for CRSI as-is to deal with coarse regularly-sampled data. We discuss in the next section the addition of more prior information than sparsity to handle this type of data.

7.4 Extensions

In this last section, we propose some ideas for future work. The common theme of most ideas is the addition of more prior information than sparsity to reconstruct seismic wavefields. In particular, we suggest to incorporate more physics so that CRSI becomes more robust to large acquisition gaps and to regularly-undersampled data.

7.4.1 Curvelet chaining

Seismic data has a sparse curvelet representation but the superposition of a few randomly-selected curvelets is not, in general, a meaningful physical signal. Hence, sparse is only a crude description of seismic wavefields in the curvelet domain. We suggest to use also the relationships between the coefficients to achieve a more accurate description. Indeed, a wavefront is typically represented by a cluster of curvelets that are close one to the other in phase space. La and Do (2005) already use a similar idea with wavelet coefficients of natural images to reach a better solution faster compared to the standard sparsity-promoting program. Their solver, termed tree-based orthogonal matching pursuit (TOMP), searches for a sparse tree representation rather than just a sparse representation.

7.4.2 Physic-based forward model

Rather than adding regularization terms to incorporate more prior information, one can also refine the formulation of the wavefield reconstruction problem—i.e., write a new forward model.

Interpolation with NMO/DMO operators

Zwartjes (2005) uses a normal moveout operator (NMO) or dip moveout operator (DMO) to flatten the input gathers— i.e., to reduce their spatial bandwidth—prior to interpolation. A pseudo-inverse of the NMO/DMO operator is then applied to the reconstructed gather to generate the final result. The advantage of this formulation lies in the reduced spatial bandwidth of the solution that can be enforced during the inversion. We propose to combine this approach with CRSI such that the interpolated data is given by $\tilde{\mathbf{f}} = \mathbf{D}^H\mathbf{C}^H\tilde{\mathbf{x}}$ where

$$\tilde{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{W}\mathbf{x}\|_1 \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{R}\mathbf{D}^H\mathbf{C}^H\mathbf{x}\|_2 \leq \sigma. \tag{7.1}$$

In these expressions, the matrices $\mathbf{R}$, $\mathbf{D}$, and $\mathbf{C}$ represent a restriction operator, an NMO/DMO operator, and a curvelet analysis operator, respectively. The matrix $\mathbf{W}$ is a diagonal weighting in the curvelet domain that enforces a limited spatial bandwidth for the solution. The vectors $\mathbf{y}$ and $\mathbf{x}$ are the acquired data and the curvelet representation of the reconstructed gather flattened, respectively. The symbol ^{H} denotes the conjugate transpose and $\tilde{}$ represents estimated quantities. Finally, σ relates to the noise level in the acquired data.

Interpolation with migration operators

Although computationally more intensive, one can also interpolate with a migration operator (see, e.g., Trad, 2002; Malcolm, 2005; Wang and Sacchi, 2007). In this case, the matrix $\mathbf{D}$ in Equation 7.1 is replaced by the migration operator and the unknown vector becomes the curvelet representation of the subsurface image.

Bibliography

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2005, Curvelab: software. (Available at http://www.curvelet.org).

———, 2006, Fast discrete curvelet transforms (FDCT): 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 C^2 singularities: Communications on Pure and Applied Mathematics, **57**, no. 2, 219 – 266.

Candès, E. J., J. Romberg, and T. Tao, 2006, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information: Transactions on Information Theory, **52**, no. 2, 489 – 509.

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, **LVII**, 1413 – 1457.

Donoho, D. L., 2006, Compressed sensing: Transactions on Information Theory, **52**, no. 4, 1289 – 1306.

Donoho, D. L., Y. Tsaig, I. Drori, and J.-L. Starck, 2006, Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit: Technical report, Stanford Statistics Department. (TR-2006-2. http://stat.stanford.edu/~idrori/StOMP.pdf).

Hennenfent, G. and F. J. Herrmann, 2006a, Application of stable signal recovery to seismic interpolation: Presented at the SEG International Exposition and 76^{th} Annual Meeting. (Slides available at http://slim.eos.ubc.ca/Publications/Public/Conferences/SEG/hennenfent06seg_pres.pdf).

———, 2006b, Recovery of seismic data: practical considerations: Presented at the SINBAD consortium meeting. (Slides available at http://slim.eos.ubc.ca/Publications/Public/Presentations/hennenfent_crsi_sinbad06.pdf).

Hennenfent, G., F. J. Herrmann, and R. Neelamani, 2005a, Seismic deconvolution revisited with curvelet frames: Presented at the EAGE 67^{th} Conference & Exhibition.

———, 2005b, Sparseness-constrained seismic deconvolution with curvelets: Presented at the CSEG National Convention.

Herrmann, F. J., 2003, Multifractional splines: Application to seismic imaging.

Kutyniok, G. and D. Labate, 2005, Shearlets: website. (`www.shearlet.org`).

La, C. and M. N. Do, 2005, Signal reconstruction using sparse tree representations: Presented at the SPIE Conference on Wavelet Applications in Signal and Image Processing XI.

Lu, Y. M. and M. N. Do, 2007, Multidimensional directional filter banks and surfacelets: Transactions on Image Processing, **16**, no. 4.

Malcolm, A., 2005, Data regularization for data continuation and internal multiples: PhD thesis.

Neelamani, R., A. I. Baumstein, D. G. Gillard, M. T. Hadid, and W. L. Soroka, 2008, Coherent and random noise attenuation using curvelet transforms: The Leading Edge. (In press).

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.

Trad, D., 2002, Interpolation with migration operators: Presented at the SEG International Exposition and 72^{th} Annual Meeting.

van den Berg, E. and M. P. Friedlander, 2007, In pursuit of a root: Technical report, UBC Computer Science Department. (TR-2007-16. `http://www.optimization-online.org/DB_FILE/2007/06/1708.pdf`).

Wang, J. and M. D. Sacchi, 2007, High-resolution wave-equation amplitude-variation-with-ray-parameter (AVP) imaging with sparseness constraints: Geophysics, **72**, no. 1, S11 – S18.

Yarham, C., G. Hennenfent, and F. J. Herrmann, 2007, Non-linear surface wave prediction and separation: Presented at the EAGE 69^{th} Conference & Exhibition.

Zwartjes, P. M., 2005, Fourier reconstruction with sparse inversion: PhD thesis.

Zwartjes, P. M. and M. D. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased data: Geophysics, **72**, no. 1, V21–V32.

Appendix A

The discrete curvelet transform

The FDCT by wrapping perfectly reconstructs data after decomposition by applying the transpose of the curvelet transform, i.e., we have $\mathbf{f} = \mathbf{C}^T\mathbf{C}\mathbf{f}$ for an arbitrary finite-energy vector $\mathbf{f}$. In this expression, $\mathbf{C} \in \mathbb{R}^{N\times M}$ represents the curvelet decomposition matrix. The curvelet coefficients are given by $\mathbf{x} = \mathbf{C}\mathbf{f}$ with $\mathbf{x} \in \mathbb{R}^N$. The curvelet transform is an overcomplete signal representation. The number of curvelets, i.e, the number of rows in $\mathbf{C}$, exceeds the number of data ($M \ll N$). The redundancy is moderate (approximately 8 in two dimensions and 24 in three dimensions). This redundancy implies that $\mathbf{C}$ is not a basis but rather a tight frame for our choice of curvelet transform. This transform preserves energy, $\|\mathbf{f}\|^2 = \|\mathbf{C}\mathbf{f}\|^2$. Because $\mathbf{C}\mathbf{C}^T$ is a projection, not every curvelet vector is the forward transform of some function $\mathbf{f}$. Therefore, the vector $\mathbf{x}_0$ can not readily be calculated from $\mathbf{f} = \mathbf{C}^T\mathbf{x}_0$, because there exist infinitely many coefficient vectors whose inverse transform equals $\mathbf{f}$.

A version of this appendix has been accepted for publication. F.J. Herrmann and G. Hennenfent. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173:233-248, 2008.

Appendix B

Curvelet properties

Curvelets are directional frame elements that represent a tiling of the two-/three-dimensional frequency domain into multiscale and multi-angular wedges (see Fig's 2.2 and 2.3). Because the directional sampling increases every-other scale, curvelets become more and more anisotropic for finer and finer scales. They become 'needle-like' as illustrated in Fig. 2.2. Curvelets are strictly localized in the Fourier domain and of rapid decay in the physical domain with oscillations in one direction and smoothness in the other direction(s). Their effective support in the physical domain is given by ellipsoids. These ellipsoids are parameterized by a width $\propto 2^{j/2}$, a length $\propto 2^j$ and an angle $\theta = 2\pi l 2^{\lfloor j/2 \rfloor}$ with j the scale, $j = 1 \cdots J$ and l the angular index with the number of angles doubling every other scale doubling (see Fig. 2.3). Curvelets are indexed by the multi-index $\gamma := (j,\, l,\, \mathbf{k}) \in \mathcal{M}$ with $\mathcal{M}$ the multi-index set running over all scales, j, angles, l, and positions $\mathbf{k}$. Therefore, conflicting angles are possible.

A version of this appendix has been accepted for publication. F.J. Herrmann and G. Hennenfent. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173:233-248, 2008.

Appendix C

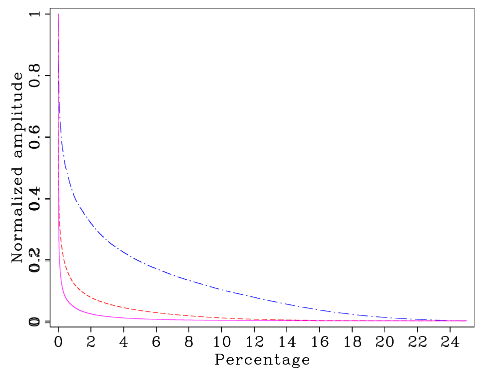
Compression properties of curvelet frames

For 2-D functions that are twice-differentiable and that contain singularities along piece-wise twice differentiable curves, the Fourier transform (ignoring log-like factors in this discussion) only attains an asymptotic decay of the k-term nonlinear approximation error of $\mathcal{O}(k^{-1/2})$. For this class of functions, this decay is far from the optimal decay rate $\mathcal{O}(k^{-2})$. Wavelets improve upon Fourier, but their decay $\mathcal{O}(k^{-1})$ is suboptimal. Curvelets, on the other hand, attain the optimal rate $\mathcal{O}(k^{-2})$. In three dimensions, similar (unpublished) results hold and this is not surprising because curvelets can in that case explore continuity along two directions.

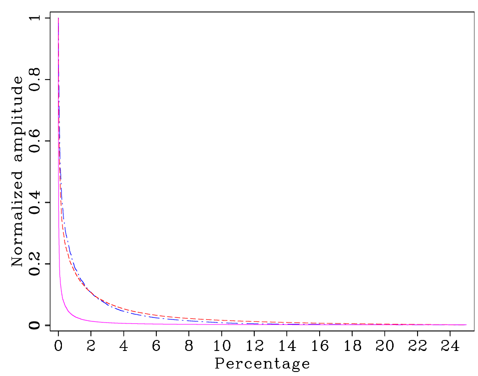
Continuous-limit arguments underly these theoretical estimates, somewhat limiting their practical relevance. Additional facts, such as the computational overhead, the redundancy and the nonlinear approximation performance on real data, need to be taken into consideration. The computational complexity of the curvelet transform is $\mathcal{O}(M \log M)$. The redundancy of the curvelet transform, however, maybe of concern. Strictly speaking wavelets yield the best SNR for the least *absolute* number of coefficients, suggesting wavelets as the appropriate choice. Experience in seismic data recovery, backed by the evaluation of the reconstruction and recovery performance in the 'eye-ball norm', suggest otherwise. Performance measures in terms of the decay rate as a function of the *relative* percentages of coefficients are more informative. For instance, when the reconstruction in Fig. 3.3 of a typical seismic shot record from only 1 % of the coefficients is considered, it is clear that curvelets give the best result. The corresponding reconstructions from Fourier and wavelets coefficients clearly suffer from major artifacts. These artifacts are related to the fact that seismic data does not lent itself

A version of this appendix has been accepted for publication. F.J. Herrmann and G. Hennenfent. Non-parametric seismic data recovery with curvelet frames. *Geophysical Journal International*, 173:233-248, 2008.

to be effectively approximated by superpositions of monochromatic plane waves or 'fat' wavelet 'point scatterers'. This superior performance of the curvelet reconstruction in Fig. 3.3 is also supported by comparisons for the decay of the normalized amplitude-sorted Fourier, wavelet and curvelet coefficients, included in Fig. C.1. In three dimensions, we expect a similar perhaps even more favorable behavior by virtue of the higher dimensional smoothness along the wavefronts. These observations suggest that curvelets are the appropriate choice for the sparsity representation so we set $\mathbf{S} := \mathbf{C}$.



(a)



(b)

Figure C.1: Decay of the transform coefficients for a typical synthetic (the fully sampled data set that corresponds to Fig. 3.2) and real data set (Fig. 3.3(a)). Comparison is made between the Fourier, wavelet and curvelet coefficients. **(a)** The normalized coefficients for a typical 2-D synthetic seismic shot record. **(b)** The same for a real shot record. Coefficients in the Fourier domain are plotted with the blue – dashed and dotted line, the wavelet coefficients with the red – dashed line, and the curvelet coefficients with the pink – solid line. The seismic energy is proportionally much better concentrated in the curvelet domain thus providing a sparser representation of seismic data than Fourier and wavelets.

Appendix D

Jittered undersampling

Jittered sampling locations r_n are given by

$$r_n = n\gamma + \varepsilon_n \quad \text{for} \quad n = -\infty, \ldots, \infty \tag{D.1}$$

The continuous random variables ε_n are independent and identically distributed (iid) according to a probability density function (pdf) p on $[-\zeta/2, \zeta/2]$. The corresponding sampling operator s is given by

$$s(r) = \sum_{n=-\infty}^{\infty} \delta(r - r_n). \tag{D.2}$$

Computing the Fourier transform of the previous expression yields

$$\hat{s}(f) = \frac{1}{\gamma} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{\gamma}\right) \mathrm{e}^{-i2\pi f \varepsilon_n} \tag{D.3}$$

which implies that

$$\mathrm{E}\left\{\hat{s}(f)\right\} = \mathrm{E}\left\{\mathrm{e}^{-i2\pi f \varepsilon_0}\right\} \cdot \frac{1}{\gamma} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{\gamma}\right) \tag{D.4}$$

since the variables ε_n are iid. By definition, the expected value of $\mathrm{e}^{-i2\pi f \varepsilon_0}$ is given by

$$\mathrm{E}\left\{\mathrm{e}^{-i2\pi f \varepsilon_0}\right\} = \int_{-\zeta/2}^{\zeta/2} p(t) \cdot \mathrm{e}^{-i2\pi f t} \mathrm{d}t \tag{D.5}$$

which is the Fourier transform of the pdf of ε_0. Hence,

$$\mathrm{E}\left\{\hat{s}(f)\right\} = \hat{p}(f) \cdot \frac{1}{\gamma} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{\gamma}\right). \tag{D.6}$$

A version of this appendix has been accepted for publication. G. Hennenfent and F.J. Herrmann. Simply denoise: wavefield reconstruction via jittered undersampling. *Geophysics*, 73(3), May-June 2008.

Finally, for a pdf that is continuous uniform on $[-\zeta/2, \zeta/2]$, the expected spectrum of the sampling operator is

$$\mathrm{E}\left\{\hat{s}(f)\right\} = \mathrm{sinc}\left(f\zeta\right) \cdot \frac{\zeta}{\gamma} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{\gamma}\right). \tag{D.7}$$

This result leads us to equation 4.6 since the columns of $\mathbf{A}^H\mathbf{A}$ are circular-shifted versions of the Fourier transform of the discrete jittered sampling vector, i.e., $\mathrm{diag}(\mathbf{R}^H\mathbf{R})$.