

Curvelet-Based Migration Amplitude Recovery

by

Peyman P. Moghaddam

B.Sc. in Electrical Engineering, Amirkabir University of Technology, Tehran, Iran,
1995

M.Sc. in Electrical Engineering, Amirkabir University of Technology, Tehran, Iran,
1998

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

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Geophysics)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2010

© Peyman P. Moghaddam 2010

Abstract

Migration can accurately locate reflectors in the earth but in most cases fails to correctly resolve their amplitude. This might lead to mis-interpretation of the nature of reflector.

In this thesis, I introduced a method to accurately recover the amplitude of the seismic reflector. This method relies on a new transform-based recovery that exploits the expression of seismic images by the recently developed curvelet transform. The elements of this transform, called curvelets, are multi-dimensional, multi-scale, and multi-directional. They also remain approximately invariant under the imaging operator.

I exploit these properties of the curvelets to introduce a method called Curvelet Match Filtering (CMF) for recovering the seismic amplitude in presence of noise in both migrated image and data.

I detail the method and illustrate its performance on synthetic dataset. I also extend CMF formulation to other geophysical applications and present results on multiple removal. In addition of that, I investigate preconditioning of the migration which results to rapid convergence rate of the iterative method using migration.

Table of Contents

Abstract	ii
Table of Contents	iii
List of Tables	viii
List of Figures	ix
Preface	xx
Acknowledgments	xxi
Dedication	xxii
Statement of Co-Authorship	xxiii
1 Introduction	1
1.1 Theme	2
1.2 Objectives	4
1.3 Outline	4
Bibliography	7

Table of Contents

2 Sparsity and continuity promoting seismic image recovery	
with curvelet frames	9
2.1 Introduction	9
2.1.1 Existing approaches	11
2.1.2 Our approach	12
2.1.3 Outline	15
2.2 Approximation of the normal operator	18
2.2.1 Zero-order imaging operators	19
2.2.2 Curvelet frames	20
2.2.3 Diagonal approximation of PsDO's	22
2.2.4 Decomposition of the normal operator	25
2.2.5 Estimation of the diagonal	26
2.3 Stable seismic image recovery	31
2.3.1 Curvelet frames for seismic images	31
2.3.2 Stable recovery	35
2.3.3 Solution by iterative thresholding	37
2.3.4 Stable seismic recovery	38
2.4 Sparsity- and continuity-enhanced seismic image recovery	40
2.5 Practical considerations	44
2.6 Numerical experiments	45
2.7 Conclusions	56
2.7.1 Initial findings	56
2.7.2 Extension to prestack imaging	59
2.7.3 Recent related work	60
2.7.4 Choice of the reference vector	60

Table of Contents

Bibliography	62
3 Curvelet-based migration preconditioning and scaling . .	67
3.1 Introduction	67
3.2 Problem formulation	69
3.3 Preconditioning	72
3.3.1 Left preconditioning by fractional differentiation . . .	73
3.3.2 Right preconditioning by scaling in the physical do- main	73
3.3.3 Right preconditioning by scaling in the curvelet do- main	74
3.4 Convergence of least-squares migration	76
3.5 Extensions	78
3.6 Conclusions	80
3.7 Acknowledgments	81
Bibliography	86
4 Curvelet-based seismic data processing	89
4.1 Introduction	89
4.2 Curvelets	90
4.3 Common problem formulation by sparsity-promoting inver- sion	91
4.4 Seismic data recovery	93
4.4.1 Curvelet-based recovery	93
4.4.2 Focused recovery	94

Table of Contents

4.5	Seismic signal separation	96
4.6	Migration-amplitude recovery	99
4.7	Discussion and conclusions	100
4.8	Acknowledgments	102
	Bibliography	104
5	True amplitude depth migration using curvelets	108
5.1	Introduction	108
5.2	Outline of the chapter	110
5.3	Problem Formulation	110
5.4	Diagonal Approximation of the Normal Operator	114
5.5	Curvelets and their invariance under normal operator	115
5.6	Diagonal approximation	117
5.7	Practical Workflow	119
5.8	Examples	121
	5.8.1 Noise-free case	121
	5.8.2 Noisy data	130
5.9	Tolerance analysis	136
5.10	Verification of approximation	138
5.11	Comparison	138
5.12	Conclusion	140
5.13	Acknowledgments	140
	Bibliography	142

Table of Contents

6	Conclusions	146
6.1	Main contributions	146
6.1.1	Sparsity and continuity promoting seismic image recovery with curvelet frames	147
6.1.2	Curvelet-based migration preconditioning and scaling	148
6.1.3	Curvelet-based seismic data processing	149
6.1.4	True amplitude depth migration using curvelets	149
6.2	Follow-up work	150
6.2.1	Full waveform inversion	150
6.2.2	3D true amplitude migration	150
6.3	Adding more priori knowledge to solution	150
6.4	Current limitations	151
6.4.1	Curvelet code	151
	Bibliography	153
 Appendices		
	A Proofs of lemma 1 and theorem 1	154
	B Reverse time wave equation migration	158
B.1	Single scattering	158
B.2	Shot-geophone modeling and migration	159
	Bibliography	161

List of Tables

2.1	Algorithm for the estimation of the diagonal via regularized least-squares inversion. The Lagrange multiplier, η , is increased up to the point that all entries in the vector for the diagonal are positive.	30
2.2	Sparsity-and continuity-enhancing recovery of seismic amplitudes.	44

List of Figures

2.1	Example of the recovery from different percentages of curvelet coefficients. The data set concerns a migrated image derived from the Mobil data set, (a) near perfect recovery from $p = 99\%$ of the coefficients, (b) recovery from only $p = 3\%$ of the coefficients, (c) the difference between near perfect recovery (a) and approximate recovery (b) . This difference does not contain substantial coherent energy.	16
-----	--	----

List of Figures

2.2	Invariance of curvelets under the discretized normal operator Ψ for a smoothly varying background model (a so-called lens model see Fig. 2.4(a)). Three coarse-scale curvelets in the physical domain before (a) and after application of the normal operator (b) in the physical (a-b) and Fourier domain (e-f) . The results for three fine-scale curvelets are plotted in (c-d) for the physical domain and in (g-h) for the Fourier domain. Remark: The curvelets remain close to invariant under the normal operator, a statement which becomes more accurate for finer scale which is consistent with Theorem 1. The example also shows that this statement only holds for curvelets that are in the support of the imaging operator excluding steeply dipping curvelets.	17
2.3	Spatial and frequency representation of curvelets, (a) four different curvelets in the spatial domain at three different scales, (b) dyadic partitioning in the frequency domain, where each wedge corresponds to the frequency support of a curvelet in the spatial domain. This figure illustrates the microlocal correspondence between curvelets in the physical and Fourier domain. Curvelets are characterized by rapid decay in the physical space and of compact support in the Fourier space. Notice the correspondence between the orientation of curvelets in the two domains. The 90° rotation is a property of the Fourier transform.	22

List of Figures

- 2.4 Stylized example of the diagonal estimation for a smooth velocity model and a reflectivity consisting of three reflection events including a fault, **(a)** the smooth background model with the low-velocity lens. This velocity model is used to calculate the linearized scattering and migration operators, **(b)** the bandwidth limited reflectivity, **(c)** the imaged reflectivity from data consisting of 32 shot records with 500 traces each, **(d)** the approximated normal operator for $\text{diag}(\tilde{\mathbf{u}})$ estimated with $\eta = 1$. The bandwidth limited reflectivity in **(b)** and the image in **(c)** served as input for the reference and 'data' vectors for the diagonal estimation procedure outlined in Table 2.1. Remarks: the main dimming of the normal operator is captured by the diagonal approximation. The relative ℓ_2 -error between the actual and the approximate normal operator is 6.1%. 32
- 2.5 Estimates for the diagonal $\tilde{\mathbf{u}}$ are plotted in **(a-d)** for increasing $\eta = \{0.01, 0.1, 1, 10\}$. The diagonal is estimated according the procedure outlined in Table 2.1 with the reference and 'data' vectors, \mathbf{v} and \mathbf{b} , plotted in Fig. 2.4(b) and 2.4(c). As expected the diagonal becomes more positive for increasing η . 33

List of Figures

2.6	Decay rate for the curvelet coefficients compared to the decay for the coefficients of Fourier (dashed line) and discrete wavelet transforms (dot-dashed line), (a) decay rate for the sorted curvelet coefficients (solid line) of the real migrated image included in Fig. 2.1(a), (b) the same as (a) but now for the synthetic reflectivity of the SEG/EAGE AA' salt model (cf. Fig. 2.8(a)). The decay for the curvelet transform clearly compares favorably to these other two transforms.	35
2.7	The SEG/EAGE AA' salt model (Aminzadeh et al., 1997), (a) The original velocity model, (b) the smoothed velocity model with a window size of 720×720 m. Remark: Imaging this model is a challenge because of the presence of the high-velocity (bright) salt.	47
2.8	Imaging of the SEG/EAGE AA' salt model (Aminzadeh et al., 1997), (a) reflectivity defined in terms of the band-pass filtered difference between the smoothed velocity model and a slightly less smoothed velocity model, (b) the imaged reflectivity according to Eq. 2.24. The main reflection events are present but suffer from deteriorated amplitudes, especially under the high-velocity salt and for steep reflectors and faults.	48

List of Figures

2.9	Images that are input to the diagonal approximation scheme outlined in Table 2.1, (a) the reference vector, \mathbf{r} , derived from Fig. 2.8(b), after applying corrections for the spherical-divergence and receiver-array taper, (b) the demigrated-migrated reference vector, $\mathbf{b} = \mathbf{\Psi}\mathbf{r}$, that serves as 'data' for the diagonal estimation, (c) diagonal approximation on the original bandwidth-limited reflectivity plotted in (a) . This diagonal approximation by $\mathbf{A}\mathbf{A}^T\mathbf{m}$ captures the normal operator, $\mathbf{\Psi}\mathbf{m}$, quite well.	50
2.10	Gradient vectors for the reference vector \mathbf{r} plotted in Fig. 2.9(a). The gradients (white \uparrow 's) are used to calculate the tangential directions along which the additional anisotropic smoothing is applied.	51
2.11	Images after nonlinear recovery (\mathbf{P}_ϵ). (a) result with the ℓ_1 -norm recovery only ($\beta = 0$); (b) recovery with the combined ℓ_1 and continuity recovery for $\alpha = \beta = 1/2$. The amplitudes are recovered. The anisotropic diffusion successfully removes the artifacts.	53

List of Figures

2.12	Recovery of the amplitudes after normalization with respect to the first event, (a) seismic traces of the original reflectivity (dot-dashed line), the migrated reflectivity (dashed line) and the recovered reflectivity (solid line), (b) amplitudes along the bottom most horizontal reflector, (c) average amplitude spectra of the depth-dependent reflectivity. The original and recovered spectra are normalized to match. Remarks: The nonlinear recovery corrects for the amplitudes and restores the spectrum.	54
2.13	Image amplitude recovery for noisy data (SNR 3 dB), (a) noise image according to Eq. 2.24, (b) image after nonlinear recovery from noisy data (\mathbf{P}_ϵ). The clearly visible non-stationary noise in (a) is removed during the recovery while the amplitudes are also restored. Steeply dipping reflectors denoted by the arrows are also well recovered.	57
2.14	'Denoising' of a shot record, (a) the noise-free data received by the receiver array, (b) noisy data with a SNR of 3 dB, (c) forward modeled data after amplitude recovery. Observe the significant improvement in the data quality, reflected in an increase for the SNR of 19.2 dB.	58

List of Figures

3.1	The SEG/EAGE AA' salt model, (a) reflectivity defined by the high-pass filtered velocity model, (b) smoothed velocity model, (c) the migrated image according to Equation 3.1. This image suffers from deteriorated amplitudes, especially under the high-velocity salt and for steep reflectors and faults.	82
3.2	Migrated images for different levels of preconditioning, (a) result for left preconditioning (level I, cf. Equation 3.6), (b) result for left-right (including depth-correction) preconditioning (level II, cf. Equation 3.7), (c) the same but now including curvelet-domain scaling (level III, cf. Equation 3.10).	83
3.3	Residual decays for different levels of preconditioning. The dotted blue lines corresponds to least-squares migration without preconditioning, the dash-dotted lines to level I preconditioning, the dashed black lines to level II preconditioning, and the red solid lines to level III preconditioning. This is offset by one iteration to account for the overhead, (a) plot for the decay of the data-space normalized residues μ_k as a function of the number of LSQR iterations, (b) the same but now for the model-space normalized residuals ν_k .	84
3.4	Least-squares migration without and with preconditioning juxtaposed with the original reflectivity (in light blue), (a) least-squares migrated image, (b) least-squares image with level III preconditioning. Notice the improvement in the recovered reflectivity.	85

List of Figures

- 4.1 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. 95

- 4.2 3-D Primary-multiple separation with \mathbf{P}_ϵ 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}_ϵ . Notice the improvement, in areas with small 3-D effects (ellipsoid) and residual multiples. 98

- 4.3 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. . 101

List of Figures

5.1	Spatial and frequency representation of curvelets, (a) four different curvelets in the spatial domain at three different scales, (b) dyadic partitioning in the frequency domain, where each wedge corresponds to the frequency support of a curvelet in the spatial domain. This figure illustrates the microlocal correspondence between curvelets in the physical and Fourier domain. Curvelets are characterized by rapid decay in the physical space and of compact support in the Fourier space. Notice the correspondence between the orientations of curvelets in the two domains. The 90° rotation is a property of the Fourier transform. Courtesy of Herrmann et al. (2008). . . .	115
5.2	Conflicting dips velocity model and reflectivity (a) velocity model, (b) smooth velocity model used to generate our example, (c) reflectivity generated by subtracting smooth velocity model from original one	122
5.3	Diagonal approximation of normal operator, (a) reference image (depth corrected migrated image), (b) normal operator on the reference image (i.e., $\Psi\mathbf{r}$) (c) approximate normal operator on the reference image (i.e., $\mathbf{C}^T\mathbf{W}\mathbf{C}\mathbf{r}$), approximation error is %3.71	124

List of Figures

5.4 Curvelet-domain representation of the curvelet vector obtained from **(a)** reference vector, **(b)** scaling coefficients without smoothing constraints, **(c)**, scaling coefficients with smoothing constraint. The different sub-images represent the curvelet coefficients at different scales (coarsest in the center) and different angles. 126

5.5 Examples of the solution for optimization problem **QP** with soft-thresholding, **(a)** $\lambda = 0$, **(b)** $\lambda = \sigma_{\mathbf{b}}$ **(c)** $\lambda = 2\sigma_{\mathbf{b}}$ 127

5.6 Examples of the solution for optimization problem **BPDN** with SPGL1, **(a)** $\epsilon = \sigma_{\mathbf{b}}$, **(b)** $\epsilon = 50\sigma_{\mathbf{b}}$ **(c)** $\epsilon = 100\sigma_{\mathbf{b}}$ 129

5.7 Reflectivity model, the vertical lines are locations of investigation 130

5.8 Amplitude analysis of the recovered images using proposed methods, **(a)** offset=3480 (m) , **(b)** offset=5160 (m), **(c)** offset=5880 (m) 131

5.9 Amplitude analysis of the recovered images using proposed methods, **(a)** offset= 8640 (m), **(b)** offset=10080 (m) , **(c)** offset=12720 (m), note the salt bottom amplitude mismatch 132

5.10 Synthetic data, **(a)** before , **(b)** after adding the noise, $SNR = 10dB$ 133

5.11 Depth corrected migrated noisy image 134

5.12 Examples of the solution for optimization problem **QP** with soft-thresholding, **(a)** $\lambda = \sigma_{\mathbf{b}}$, **(b)** $\lambda = 2\sigma_{\mathbf{b}}$ **(c)** $\lambda = 3\sigma_{\mathbf{b}}$. . . 135

5.13 Examples of the solution for optimization problem **BPDN** with SPGL1, **(a)** $\epsilon = 10\sigma_{\mathbf{b}}$, **(b)** $\epsilon = 140\sigma_{\mathbf{b}}$ **(c)** $\epsilon = 200\sigma_{\mathbf{b}}$. . 137

List of Figures

- 5.14 Comparison between the approximation terms in Equation 5.21, **(a)** left hand term $E_l = \mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \mathbf{C} \mathbf{C}^T \mathbf{W} \mathbf{x}$, **(b)** right hand term $E_r = \mathbf{C}^T \mathbf{W}^{\frac{1}{2}} \mathbf{x}$, relative error is 4.09 percent. . . . 139

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. I also wish to thank him for his kind and sometimes (very) energetic encouragements.

I would like to express my gratitude to Reza Shahidi, Deli Wang, Cody Brown, Gilles Hennenfent, 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 Chris Stolk, William Symes, Uri Ascher and Chen Grief. 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 Total E&P for the hospitality during my internship. In particular, I wish to thank Eric Dussaud, and Paul Williamson for great technical discussions.

I would like to express my gratitude to Chen Grief, Michael Bostock, Doug Oldenburg for serving on my supervisory committee.

Finally, many thanks to my family for their love, support and encouragement throughout my studies.

To my family

Statement of Co-Authorship

Chapter 2 was published jointly with Felix J. Herrmann and Chris Stolk in the mathematical literature where alphabetical ordering of authors is customary. Aside from providing the numerical examples for this paper, I co-authored the text for the examples and I wrote the description of the part that involved our nonlinear solver.

Chapter 3 was published with Felix J. Herrmann, Cody R. Brown, and Yogi A. Erlangga. My contribution consisted of the development of the curvelet matched filter and the design of the imaging experiment. Cody assisted with generating the results and Felix wrote most of the manuscript with significant inputs from me and the other co-authors.

Chapter 4 is a review paper highlighting three distinctively different applications of the curvelet transform to seismic data processing and imaging. I was responsible for the imaging example and I authored the corresponding text.

Chapter 5 was prepared for submission for publication jointly with Felix J. Herrmann with input from Reza Shahidi. I was responsible for all examples and I authored the text.

Chapter 1

Introduction

Exploration seismology 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. Seismic processing is a technique widely used by the oil industry to explore and identify potential oil-rich areas into the earth using the recorded data.

Seismic migration is vital part of seismic processing. Migration generates an image of the subsurface that is finally interpreted by geo-scientists.

Migration difficulties generally arise from assumptions made in algorithms, that are not met by underlying theory. In particular, modern migration algorithms can accurately position reflecting features in the Earth's subsurface, however they generally do not correctly resolve the amplitudes. This problem has different causes. First, the migration operator is the adjoint but not the inverse of the linearized Born modeling or scattering operator (see e.g. Guitton, 2004; Claerbout, 1985; Gray, 1997; Symes, 2008).

Second, There are also other sources contributing to this inaccuracy including geometrical spreading of wavefront energy, acquisition limitations, velocity heterogeneity and anomalous focusing.

There exists a wide variety of migration amplitude recovery techniques:

- **iterative-based methods** use migration and scattering operators and iteratively recover the amplitude of the reflectors through a Krylov subspace-based method such as LSQR or conjugate-gradient (Kuehl and Sacchi, 2003). They typically fairly computationally intensive
- **match filter-based methods** represent another type of amplitude recovery approach that include the migrated and re-migrated image. They require "only one" time modeling and migration of the image and computationally cheaper than iterative based methods. (Guitton, 2004; Claerbout, 1985; Rickett, 2003; Symes, 2008) are examples of this approach.

1.1 Theme

The method proposed in this thesis follows the match filter-based approach. The main theme of this thesis is a practical, robust, and geometrical—i.e., transform-based—approach to the seismic migration amplitude recovery problem. The motivation of this approach is that two key features of seismic data are, in my opinion, not used to their full extent in existing approaches, namely

- **Representation of the seismic normal operator in transform**

domain The seismic normal operator typically belongs to a class of operators called Pseudo-differential operators. These operators have interesting properties to make them suitable to represent in the transform domain.

- **Strong geometrical structure** Seismic images are a spatio-temporal sampling of the reflectors in the Earth's subsurface. These reflectors are mathematically continuous functions of reduced dimension (i.e., curves in 2D or sheets in 3D)

To make the most of these properties, my approach uses the curvelet transform (Candès and Donoho, 2004) which is *data-independent*, *multi-scale*, and *multidirectional*. The elements of this transform, the curvelets, are localized in the frequency domain and of rapid decay in the physical domain. They are very efficient at representing curve-like singularities—e.g., reflectors. In other words, only a few curvelets are needed to represent the complexity of real seismic image. I use this piece of information, called *sparsity*, to help solve the imaging problem. The idea of sparsity-promoting inversion is in itself not new to geophysics.

The other interesting property of the curvelets is their behavior under the action of acoustic wave propagator (e.g., Candès and Demanet, 2003; Smith, 1998, 1997) . A theorem in chapter 2 proved that the curvelets are approximately invariant under the action of the normal operator . Furthermore I adopt the result of this theorem to introduce match-filtering approach for estimation of this operator.

1.2 Objectives

The objectives of this thesis are two-fold:

- develop an in-depth understanding of successful curvelet match filtering for the approximation of normal operator and its key ingredients.
- formulate a practical sparsity-promoting seismic migration amplitude recovery algorithm.

1.3 Outline

In chapter 2, I observe that curvelets, as a directional frame expansion, lead to sparsity of seismic images and exhibit invariance under the normal operator of the linearized imaging problem. Based on this observation I derive a method for stable recovery of the migration amplitudes from noisy data. The method corrects the amplitudes during a post-processing step after migration, such that the main additional cost is one application of the normal operator, i.e. a modeling followed by a migration. Asymptotically this normal operator belongs to a class of operators known as pseudodifferential operators, for which a diagonal approximation in the curvelet domain is derived, including a bound for its error and a method for the estimation of the diagonal from a compound operator consisting of discrete implementations for the scattering operator and its adjoint the migration operator. The solution is formulated as a nonlinear optimization problem where sparsity in the curvelet domain, as well as continuity along the imaged reflectors, are jointly promoted. To enhance sparsity, the ℓ_1 -norm on the curvelet

1.3. Outline

coefficients is minimized, while continuity is promoted by minimizing an anisotropic diffusion norm on the image. The performance of the recovery scheme is evaluated with a time-reversed 'wave-equation' migration code on synthetic datasets, including the complex SEG/EAGE AA' salt model.

In chapter 3, I introduce a preconditioner for the inversion of the linearized Born scattering operator. This preconditioner approximately corrects for the “square root” of the normal operator. This approach consists of three parts, namely (i) a left preconditioner, defined by a fractional time integration designed to make the migration operator zero order, and two right preconditioners that apply (ii) a scaling in the physical domain accounting for a spherical spreading, and (iii) a curvelet-domain scaling that corrects for spatial and reflector-dip dependent amplitude errors. I show that a combination of these preconditioners lead to a significant improvement of the convergence for iterative least-squares solutions to the seismic imaging problem based on reverse-time migration operators.

In chapter 4, I 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 curvelet sparsity inversion formulation. This puts in a broader perspective the insights gained during the development of curvelet sparsity inversion formulation.

In chapter 5, I focused on the practical approach of amplitude recovery method. I revisit the amplitude recovery method and extend it threefold. First, I replace the linear least-squares formulation for the estimation of the curvelet-domain coefficients (Herrmann et al. (2008)) by a non-linear least-squares formulation that imposes positivity on the scaling coefficients.

1.3. Outline

This comes from the fact that the normal operator is positive semi-definite. Second, I consider the noise term both in the data and that generated by migration due to incomplete data (i.e., few shots). Third, I investigate two methods for recovery and compare them in terms of quality and performance, a soft thresholding-correction technique and a sparsity promotion technique.

In Chapter 6, I summarize the work done in this thesis, and discuss some of its aspects in a broader context. Conclusions and recommendations for future research follow.

Appendix A contains further details about the curvelet transform. It proves the theorem and associated lemma regarding the invariance of the curvelet elements under the action of normal operator. This appendix pairs with chapter 2.

Bibliography

- 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. 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.
- Claerbout, J., 1985, *Earth soundings analysis, processing versus inversion*: Blackwell Scientific Publications.
- Gray, S. H., 1997, True amplitude seismic migration: A comparison of three approaches: *Geophysics*, **62**, 929–936.
- Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: *Geophysics*, **69**, 1017–1024.
- Herrmann, F. J., P. P. Moghaddam, and C. Stolk, 2008, Sparsity- and continuity-promoting seismic image recovery with curvelet frames: *Applied and Computational Harmonic Analysis*, **24**, 150–173.
- Kuehl, H. and M. Sacchi, 2003, Least-squares wave-equation migration for AVP/AVA inversion: *Geophysics*, **68**, 262–273.
- Rickett, J., 2003, Illumination based normalization for wave-equation depth migration: *Geophysics*, **68**, 208–221.

Chapter 1. Bibliography

- Smith, H., 1997, A Hardy space for Fourier integral operators: *Geom. Anal.*, **7**, 784–842.
- , 1998, A parametrix construction for wave equations with $C^{1,1}$ coefficients: *Ann. Inst. Fourier, Grenoble*, **48**, 797–835.
- Symes, W., 2008, Optimal scaling for reverse time migration: submitted to *Geophysics*.

Chapter 2

Sparsity and continuity promoting seismic image recovery with curvelet frames

2.1 Introduction

This paper introduces a general-purpose algorithm for the stable recovery of the amplitudes in seismic images. The method involves the inversion in dimension d of a zero-order pseudodifferential operator (PsDO) of the type

$$(\Psi f)(x) = \int_{\mathbb{R}^d} e^{-ix \cdot \xi} a(x, \xi) \hat{f}(\xi) d\xi \quad (2.1)$$

with

$$\hat{f}(\xi) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} f(x) e^{ix \cdot \xi} dx \quad (2.2)$$

A version of this chapter has been published. Herrman, F.J., Moghaddam, P.P. and Stolk, C. (2008) Sparsity- and continuity-promoting seismic image recovery with curvelet frames. *Applied and Computational Harmonic Analysis*, Vol. 24, No. 2, pp. 150-173, 2008.

2.1. Introduction

the Fourier transform with $x, \xi \in \mathbb{R}^d$ the spatial coordinate and wavenumber vectors and $a(x, \xi)$ the symbol of the pseudodifferential operator. Under certain conditions that include no grazing rays and high-frequency asymptotics, the above linear operator describes seismic data after imaging (ten Kroode et al., 1998; de Hoop and Brandsberg-Dahl, 2000; Stolk and Symes, 2003). A seismic image is derived from noisy data given by

$$d(x_s, x_r, t) = (Km)(x_s, x_r, t) + n(x_s, x_r, t) \quad (2.3)$$

with K the Born scattering operator, $m(x)$ the model with the reflectivity and n zero-mean standard deviation σ white Gaussian noise. The seismic data volume is a function of the source and receiver positions, $x_s \in \mathbb{R}^{d-1}$ and $x_r \in \mathbb{R}^{d-1}$ and of time, $t \in \mathbb{R}_0^+$. After applying the migration operator K^T , (the symbol T is reserved for the transpose), to the data volume an “image” is obtained according to

$$\begin{aligned} (K^T d)(x) &= (K^T K m)(x) + (K^T n)(x) \\ y(x) &= (\Psi m)(x) + e(x). \text{ with } x \in \mathbb{R}^d \end{aligned} \quad (2.4)$$

This equation for the image y corresponds to the normal equation and contains contributions from the normal operator, Ψ , as well as from imaged noise e . The operator K preserves the singularities in y , i.e., $K^T K m \approx \text{Id } m$ with Id the identity matrix and the symbol \approx indicating “approximated by” in the locations of the singularities. With the increased demand for high-quality images, the above approximation of $\Psi \approx \text{Id}$ is no longer justifiable

because it may lead to errors in the relative amplitudes of the imaged reflectors. Our main goal is to derive a method that recovers these imaged amplitudes, i.e. estimate models m from “images” y with clutter.

2.1.1 Existing approaches

With the increasing demand for hydrocarbons and the increasingly hard to image (subsalt), the recovery of correct seismic amplitudes has become more necessary. Approaches in the extensive literature on this topic range from least-squares migration, where the normal or Gramm matrix is inverted with Krylov subspace methods (see e.g. Nemeth. et al., 1999; Chavent and Plessix, 1999; de Hoop and Brandsberg-Dahl, 2000), to high-frequency (microlocal) methods (see e.g. ten Kroode et al., 1998; de Hoop and Brandsberg-Dahl, 2000; Stolk and Symes, 2003), where the normal operator is inverted based on ray-asymptotic arguments, to methods that apply a diagonal scaling (see e.g. (Rickett, 2003; Guitton, 2004; Plessix and Mulder, 2004) and more recently Symes (2006a)) to approximately invert the Gramm operator of ‘wave-equation’ migration. These scaling methods either calculate the weighting analytically (ten Kroode et al., 1998; Plessix and Mulder, 2004), based on certain assumptions regarding the acquisition and velocity model, or estimate the diagonal weighting from the action of the normal operator on some reference vector, an idea first suggested by Symes and reported in (Claerbout and Nichols, 1994). In this paper, a similar approach is followed, where the normal operator is replaced by a diagonal matrix acting on the curvelet coefficients of the image (Candes et al., 2006a; Hennenfent and Herrmann, 2006).

2.1.2 Our approach

The computational cost of evaluating the migration operator is $\mathcal{O}(n^{d+2})$ for a d -dimensional image with n samples in each direction. This large cost makes it a major challenge to conduct least-squares migration for $d > 2$ or large n (Nemeth et al. (1999); Chavent and Plessix (1999); de Hoop and Brandsberg-Dahl (2000); Hu et al. (2001); Kuhl and Sacchi (2003); Plessix and Mulder (2004); Hu et al. (2001)). We address this issue by exploiting recently developed curvelet frames. These frame expansions compress seismic images (see e.g. Candes et al., 2006a; Hennenfent and Herrmann, 2006, for the compression of seismic data and Fig. 2.1 for the compression of a typical migrated seismic image) and consist of a collection of frame elements 'curvelets' that are approximately invariant under PsDO's. These two properties allow for the development of an approach similar to the so-called wavelet-vaguelette method (WVD) (see e.g. Donoho, 1995; Candes and Donoho, 2000b; Lee and Lucier, 2001). In this approach, scale-invariant homogeneous operators are nonlinearly inverted, using the eigenfunction-like behavior of wavelets and curvelets.

Our main contribution to earlier ideas on stable seismic image recovery (see e.g. Herrmann, 2003) is threefold: First, the WVD method is extended to expanding PsDO's with respect to *redundant* curvelet frames. Second, it is observed that order-zero pseudodifferential operators with homogeneous principal symbols act approximately as a multiplication on curvelets. This property implies that such operators can be approximated by a diagonal matrix acting on curvelet coefficients of an image, followed by an inverse

curvelet transform. We refer to this procedure as a curvelet-domain diagonal approximation (or weighting) that becomes more accurate for smaller scales (higher frequencies). Third, a formulation for the amplitude recovery is presented in terms of a nonlinear sparsity-and continuity-enhancing optimization problem.

After discretization, the nonlinear optimization problem for the seismic amplitude recovery (see e.g. Tsaig and Donoho, 2006; Candes et al., 2006b; Daubechies et al., 2005) has the following form

$$\mathbf{P}_\epsilon : \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} J(\mathbf{x}) = J_s(\mathbf{x}) + J_c(\mathbf{x}) & \text{subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \\ \tilde{\mathbf{m}} = (\mathbf{A}^T)^\dagger \tilde{\mathbf{x}}. \end{cases} \quad (2.5)$$

During the optimization, the vector \mathbf{x} is optimized with respect to the penalty functional $J(\mathbf{x})$ and the ℓ_2 -data misfit. The penalty functional $J(\mathbf{x})$ is combination of sparsity penalty function $J_s(\mathbf{x})$ and continuity penalty function $J_c(\mathbf{x})$. The term sparsity vector is used for \mathbf{x} to stress the point that the magnitude-sorted coefficients of this vector are of rapid decay, by virtue of compression by curvelet frames.

The above optimization problem is solved for the model by finding a coefficient vector $\tilde{\mathbf{x}}$ that minimizes the penalty term subject to fitting the data to within a noise-level dependent tolerance level ϵ . The 'tilde' symbol ($\tilde{}$) is reserved for vectors that solve a nonlinear optimization problem. The penalty functional $J(\mathbf{x})$ is designed to promote the *sparsity* of the image in the curvelet domain (i.e., $J_s(\mathbf{x})$) as well as *continuity* along the imaged reflectors (i.e., $J_c(\mathbf{x})$). Bold lowercase symbols refer to discretized vectors

2.1. Introduction

and bold uppercase symbols to discretized operators. Non-boldface symbols refer to continuous infinite dimensional functions and operators.

Estimates for the model $\tilde{\mathbf{m}}$ are calculated by applying the pseudo-inverse (denoted by the symbol \dagger) of the analysis matrix to $\tilde{\mathbf{x}}$. This analysis matrix is given by the adjoint of the synthesis matrix (\mathbf{A}^T). This synthesis is given by the diagonally-weighted curvelet *synthesis* matrix with a weighting designed such that

$$\mathbf{A}\mathbf{A}^T \mathbf{r} \simeq \mathbf{\Psi}\mathbf{r}. \quad (2.6)$$

In this expression, \mathbf{r} represents an appropriately chosen discrete reference vector and $\mathbf{\Psi}$ the discrete normal operator, formed by compounding the discrete scattering and its transpose the migration operator, i.e., $\mathbf{\Psi} := \mathbf{K}^T\mathbf{K}$ with the symbol $:=$ denoting 'defined as'. The symbol \simeq is used to denote that this expression is approximate, a statement we will make more precise with a bound on the error.

After forming the normal operator with one's favorite numerical implementation for the migration operator and its adjoint, our amplitude recovery method consists of the following steps:

1. Calculate an appropriate reference vector, \mathbf{r} , from the data that is close enough to the unknown image. Typically, this reference vector is defined by a migrated image to which a simple amplitude correction is applied;
2. Estimate a diagonal curvelet-domain weighting that approximates the normal operator on the reference vector. This diagonal defines the synthesis matrix \mathbf{A} ;

3. Estimate $\tilde{\mathbf{x}}$ by solving the nonlinear optimization problem \mathbf{P}_ϵ . This program inverts the synthesis matrix;
4. Calculate the model $\tilde{\mathbf{m}}$ from the recovered coefficient vector $\tilde{\mathbf{x}}$ through the pseudo-inverse of \mathbf{A}^T .

The proposed recovery method derives from two essential properties of curvelet frames, namely the ability of this signal representation to compress seismic images and the invariance of curvelets under the normal operator. Fig. 2.1 and 2.2 are included to stress the importance of these properties by confirming that a real migrated image can accurately be recovered from only 3% of the curvelet coefficients, while curvelets remain invariant under the normal operator defined for a smooth background velocity model. During this paper, we use the first property to formulate a sparsity promoting seismic image in term of a nonlinear optimization and the second property for diagonal approximation of the normal operator in curvelet domain.

2.1.3 Outline

First, a diagonal approximation of the normal operator in the curvelet domain is derived. The error of this approximation is bounded, using properties of the curvelet tiling of phase space. In this derivation, use is made of the property that the normal operator can be described as a zero-order pseudodifferential operator. Next, a method is proposed to estimate the diagonal weighting matrix that uses the property that the symbol of the normal operator is smooth in phase space and positive. This diagonal approximation leads to an approximate seismic image representation by a diagonally-

2.1. Introduction

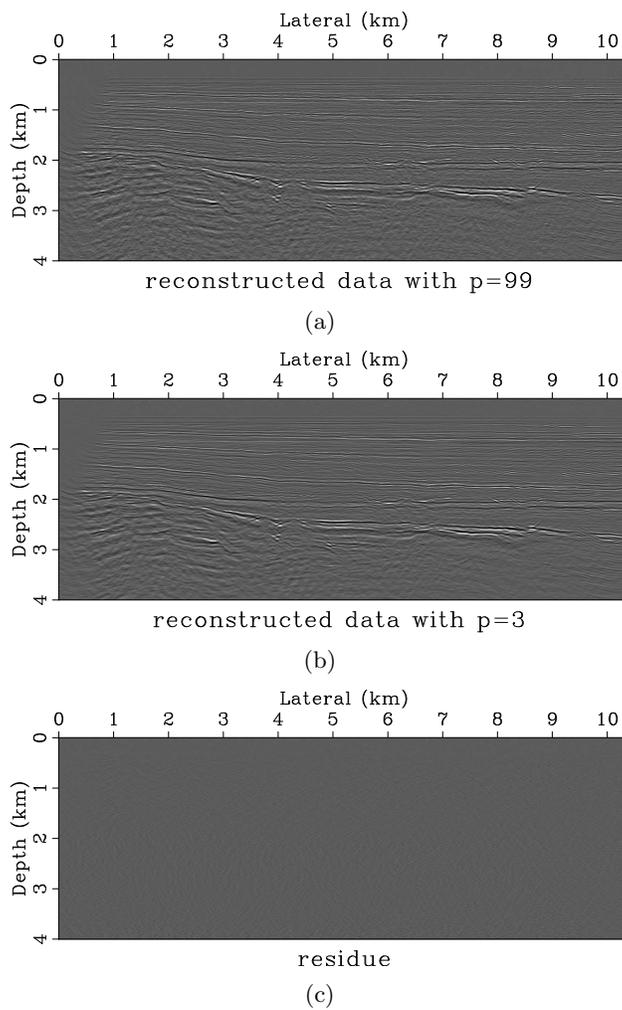


Figure 2.1: Example of the recovery from different percentages of curvelet coefficients. The data set concerns a migrated image derived from the Mobil data set, **(a)** near perfect recovery from $p = 99\%$ of the coefficients, **(b)** recovery from only $p = 3\%$ of the coefficients, **(c)** the difference between near perfect recovery **(a)** and approximate recovery **(b)**. This difference does not contain substantial coherent energy.

2.1. Introduction

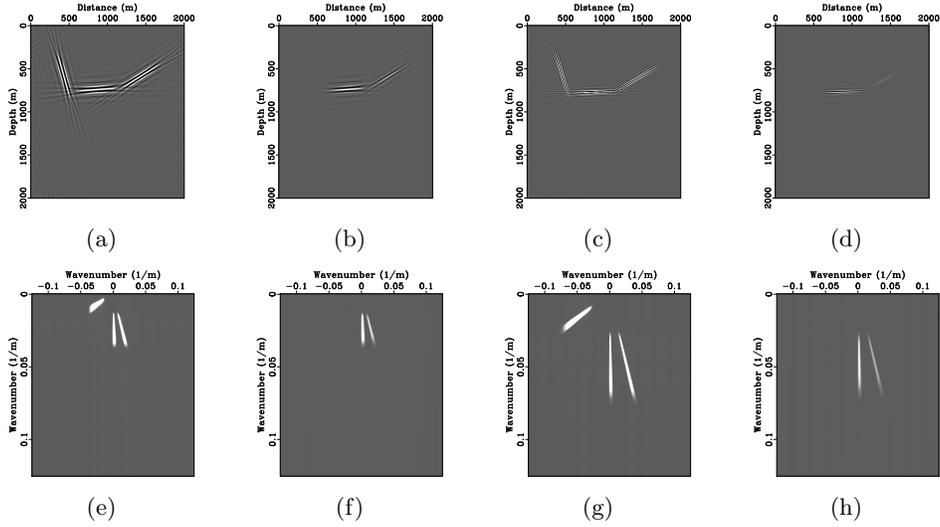


Figure 2.2: Invariance of curvelets under the discretized normal operator Ψ for a smoothly varying background model (a so-called lens model see Fig. 2.4(a)). Three coarse-scale curvelets in the physical domain before (a) and after application of the normal operator (b) in the physical (a-b) and Fourier domain (e-f). The results for three fine-scale curvelets are plotted in (c-d) for the physical domain and in (g-h) for the Fourier domain. Remark: The curvelets remain close to invariant under the normal operator, a statement which becomes more accurate for finer scale which is consistent with Theorem 1. The example also shows that this statement only holds for curvelets that are in the support of the imaging operator excluding steeply dipping curvelets.

weighted curvelet transform. With this representation, the seismic image amplitudes can be recovered by solving a nonlinear optimization problem. During this optimization problem, the image is recovered by minimizing the data mismatch while promoting sparsity in the curvelet domain and continuity along imaged reflectors. This paper is concluded by applying the algorithm to a synthetic data set derived from the SEG AA' salt model.

2.2 Approximation of the normal operator

The primary focus of seismic imaging is to locate singularities in the earth's elastic properties from seismic data recorded at the surface (Beylkin, 1984; de Hoop and Bleistein, 1997; ten Kroode et al., 1998; Stolk, 2000; Bleistein et al., 2001; Brandsberg-Dahl and de Hoop, 2003). A seismic survey consists of multiple seismic experiments in which both the location of the sources and receivers are varied along the surface. The acquired data is subsequently used to create images of the singularities in the subsurface. The main purpose of the recovery method presented in this paper is to recover the relative amplitudes of seismic images from data that is possibly contaminated with noise. For this purpose, a diagonal approximation of the normal operator in the curvelet domain is presented. This approximation derives from the invariance properties of curvelets under the normal operator (see Fig. 2.2). The approximation leads to a SVD-like decomposition for the normal operator and makes the large-scale seismic image recovery problem amenable to a solution by nonlinear optimization.

2.2.1 Zero-order imaging operators

In the high-frequency limit, the scattering operator and the normal operator can, under certain conditions on the medium and ray-geometry, be considered as Fourier Integral Operators (FIO's) (ten Kroode et al., 1998). In dimension two ($d = 2$), the scattering operator, K , and its adjoint the migration operator, K^T , can both be considered as FIO's of order $\frac{1}{4}$, while the leading behavior for their composition, the normal operator Ψ , corresponds to an order-one invertible elliptic PsDO (ten Kroode et al., 1998; de Hoop and Brandsberg-Dahl, 2000; Stolk and Symes, 2003). To make this PsDO amenable to an approximation by curvelets, the following substitutions are made for the scattering operator and the model: $K \mapsto K(-\Delta)^{-1/2}$ and $m \mapsto (-\Delta)^{1/2} m$ with $((-\Delta)^\alpha f)^\wedge(\xi) = |\xi|^{2\alpha} \cdot \hat{f}(\xi)$, with Δ is the Laplacian operator. Alternatively, these operators can be made zero-order by composing the data side with a 1/2-order fractional integration along the time coordinate, i.e., $K \mapsto \partial_t^{-1/2} K$. After these substitutions, the normal operator Ψ becomes zero-order. Similar substitutions are made in the WVD methods where vaguelettes are introduced according to the same mappings.

Our seismic image amplitude problem is different from the recovery of images in ill-posed problems such as inverting the Radon transform (Candes and Donoho, 2000b). In our case, the ill-posedness comes from small values of the symbol at certain positions and wave numbers and is not related to the ill-posedness stemming from the inversion of a normal operator that acts as a (fractional) inverse Laplacian. Our normal operator, without the above substitutions, acts as a (fractional) Laplacian. This behavior makes

the problem ill-posed for the low-frequencies. Since seismic data and the approximations of the operators are both high-frequency, this ill-posedness is negated. The ill-posedness that is a concern are the entries in the symbol of the normal operator that correspond to regions in the model space that are badly insonified.

2.2.2 Curvelet frames

Curvelets are directional frames that represent a specific tiling of the two/three-dimensional frequency plane into multiscale and multi-angular wedges (see Fig. 5.1). Because the directional sampling increases every-other scale doubling, curvelets become more and more anisotropic at finer and finer scales. They become 'needle-like' as illustrated in Fig. 5.1. Curvelets are localized in Fourier space and their smoothness in this domain leads to a rapid decay in the physical domain. Their effective support is given by ellipsoids with a width $\propto 2^{j/2}$ and length $\propto 2^j$ and an angle $\theta_{jl} = 2\pi l 2^{\lfloor j/2 \rfloor}$ with j the scale and l the angular index. Curvelets are indexed by the multi-index $\mu := (j, l, k) \in \mathcal{M}$ with \mathcal{M} the multi-index set running over all scales, j , angles, l , and positions k (see for details Candes et al., 2006a; Ying et al., 2005). Following Candes et al. (2006a), define the continuous curvelet family for $x \in \mathbb{R}^2$ as

$$\varphi_\mu(x) = 2^{3j/4} \varphi(D_j R_{\theta_{jl}} x - k), \quad (2.7)$$

where

- $\varphi(x)$ is a smooth bell-shaped function in the horizontal direction and oscillatory in the vertical direction;

2.2. Approximation of the normal operator

- $D_j = \text{diag}(2^j, 2^{\lfloor j/2 \rfloor})$ is the parabolic scaling matrix;
- $R_{\theta_{jl}}$ is the rotation matrix over angles $\theta_{jl} = 2\pi l 2^{-\lfloor j/2 \rfloor}$ with $0 \leq \theta_{jl} < 2\pi$;
- $k = (k_1, k_2) \in \mathbb{Z}^2$ the translation parameters.

In the frequency domain, curvelets are compactly supported and each element $\widehat{\varphi}_\mu(\xi)$ is localized near the symmetric wedge

$$W_{jl} = \{\pm\xi, 2^j \leq |\xi| \leq 2^{j+1}, |\theta - \theta_{jl}| \leq \pi \cdot 2^{-\lfloor j/2 \rfloor}\},$$

The number of wedges doubles every other scale doubling and hence the directional selectivity increases for finer scales (see Fig. 5.1). For each $\mu \in \mathcal{M}$, the curvelet coefficients are given by the inner product

$$c_\mu = \langle f, \varphi_\mu \rangle := \int_{\mathbb{R}^2} f(x) \overline{\varphi_\mu(x)} dx \quad (2.8)$$

of a function $f \in L^2(\mathbb{R}^2)$ with a curvelet $\varphi_\mu \in L^2(\mathbb{R}^2)$. Curvelets are tight frames, so we have the following reconstruction formula

$$f = \sum_{\mu \in \mathcal{M}} c_\mu \varphi_\mu = C^T C f, \quad (2.9)$$

by virtue of the energy conservation

$$\sum_{\mu \in \mathcal{M}} |c_\mu|^2 = \|f\|_{L^2(\mathbb{R}^2)}^2, \quad \forall f \in L^2(\mathbb{R}^2), \quad (2.10)$$

2.2. Approximation of the normal operator

with C and C^T the curvelet decomposition and composition, respectively. Refer to Candes et al. (2006a), for details on the construction of the fast discrete curvelet transform (FDCT) in dimension two.

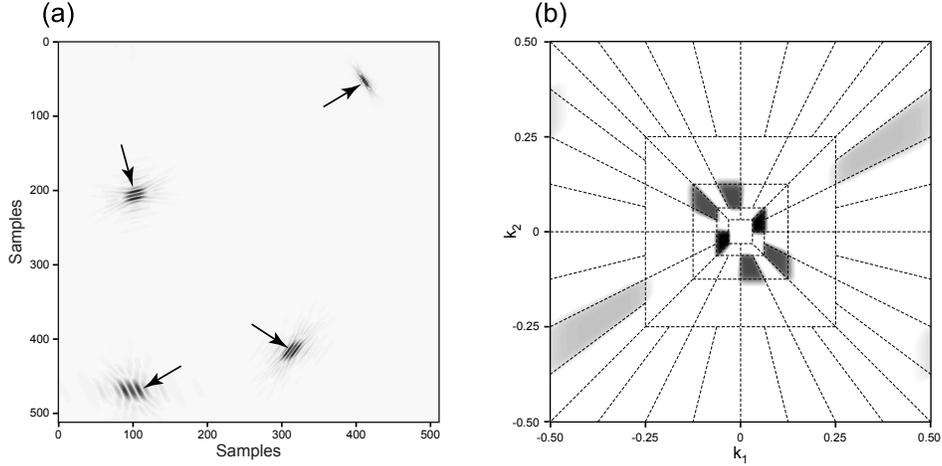


Figure 2.3: Spatial and frequency representation of curvelets, **(a)** four different curvelets in the spatial domain at three different scales, **(b)** dyadic partitioning in the frequency domain, where each wedge corresponds to the frequency support of a curvelet in the spatial domain. This figure illustrates the microlocal correspondence between curvelets in the physical and Fourier domain. Curvelets are characterized by rapid decay in the physical space and of compact support in the Fourier space. Notice the correspondence between the orientation of curvelets in the two domains. The 90° rotation is a property of the Fourier transform.

2.2.3 Diagonal approximation of PsDO's

Data, and the scattering operator K can be so defined that the normal operator is a PsDO of order 0, which is real and self-adjoint, and has a homogeneous principal symbol $a(x, \xi)$. We will show that we can approximate Ψf in the curvelet domain by application of a diagonal matrix, if the

2.2. Approximation of the normal operator

wavenumbers in f are sufficiently high, relative to the smoothness of the symbol of Ψ .

So let $\Psi = \Psi(x, D)$ be a pseudodifferential operator of order 0, with homogeneous principal symbol $a(x, \xi)$. Assuming the operator is self-adjoint and real, the principal symbol a is also real and has an even symmetry under the transformation $\xi \leftrightarrow -\xi$. In addition, we make the technical assumption that Ψ is compact, i.e., if f has compact support, then Ψf also has compact support.

Curvelets in \mathbb{R}^2 are denoted by φ_μ and we define $|\mu| = j$. The scale j ranges from some positive number j_{\min} , to infinity (or j_{\max} in an implementation). The central position and wave vector for the curvelet will be denoted by (x_μ, ξ_μ) , we let $\theta_\mu = \xi_\mu / \|\xi_\mu\|$ be the angle of the curvelet.

Next, we give some consequences of the localization of the curvelet in the space and Fourier domains. The support of a curvelet in the Fourier domain is contained in one of the domains $\pm\xi_2 > \epsilon|\xi_1|$, or $\pm\xi_1 > \epsilon|\xi_2|$, for some small ϵ , say we are in the first case, then ξ_2 can be used as the radial coordinate, with ξ_1/ξ_2 the angular coordinate in the Fourier domain. From the localization properties w.r.t. the angular coordinate, it follows that $(\xi_1/\xi_2 - \xi_{\mu,1}/\xi_{\mu,2})$ is bounded by $C_1 2^{-\lfloor j/2 \rfloor}$ on the support of the curvelet and with C_1 some finite positive constant (from here on we assume that any constant C_i is finite and positive), so that

$$\|(\xi_1/\xi_2 - \xi_{\mu,1}/\xi_{\mu,2})\widehat{\varphi}_\mu\|_{L^2(\mathbb{R}^2)} \leq C_1 2^{-\lfloor j/2 \rfloor}. \quad (2.11)$$

Localization in the space domain follows from the smoothness of the defining

2.2. Approximation of the normal operator

functions in the Fourier domain, in the radial direction we have

$$\|\theta_\mu \cdot (x - x_\mu)\varphi_\mu\|_{L^2(\mathbb{R}^2)} \leq C_2 2^{-j}$$

with C_2 another finite constant. For the other directions we have a weaker estimate, as the curvelet is in the Fourier domain narrower, and therefore in the space domain wider in the directions normal to θ_μ

$$\|(x - x_\mu)\varphi_\mu\|_{L^2(\mathbb{R}^2)} \leq C_3 2^{-\lfloor j/2 \rfloor}.$$

In fact parallel results hold for curvelets in \mathbb{R}^d .

We now compare the application of a PsDO to a curvelet with simply multiplying by the value of the symbol at (x_μ, ξ_μ) . The result will be proved for x in \mathbb{R}^d .

Lemma 1 *Suppose a is in the symbol class $S_{1,0}^0$ (Hormander, 1987), then, with C' some constant, the following holds*

$$\|(\Psi(x, D) - a(x_\nu, \xi_\nu))\varphi_\nu\|_{L^2(\mathbb{R}^n)} \leq C' 2^{-|\nu|/2}. \quad (2.12)$$

To approximate Ψ , we define the sequence $\mathbf{u} := (u_\mu)_{\mu \in \mathcal{M}} = a(x_\mu, \xi_\mu)$ with \mathcal{M} the set of all the curvelet indices. Let \mathbf{D}_Ψ be the diagonal matrix with entries given by \mathbf{u} . Next we state our result on the approximation of Ψ by $C^T \mathbf{D}_\Psi C$.

Theorem 1 *The following estimate for the error holds*

$$\|(\Psi(x, D) - C^T \mathbf{D}_\Psi C)\varphi_\mu\|_{L^2(\mathbb{R}^n)} \leq C'' 2^{-|\mu|/2}, \quad (2.13)$$

where C'' is a constant depending on Ψ .

This main result is proved in Appendix A and it shows that the approximation error for the diagonal approximation decreases for increasingly finer scales, j . The approximation derives from the property that the symbol is slowly varying over the support of a curvelet when the velocity model is sufficiently smooth.

2.2.4 Decomposition of the normal operator

By virtue of Theorem 1, the normal operator can approximately be factorized into

$$\begin{aligned} (\Psi\varphi_\mu)(x) &\simeq (C^T \mathbf{D}_\Psi C \varphi_\mu)(x) \\ &= (AA^T \varphi_\mu)(x) \end{aligned} \tag{2.14}$$

with $A := C^T \sqrt{\mathbf{D}_\Psi}$ and $A^T := \sqrt{\mathbf{D}_\Psi} C$. Because the seismic reflectivity can be written as a superposition of curvelets, the φ_μ can be replaced with the model m . The normal equation (cf. Eq. 2.4) can now be rewritten in the following approximate form

$$\begin{aligned} y(x) &= (\Psi m)(x) + e(x) \\ &\simeq (AA^T m)(x) + e(x) \\ &= Ax_0 + e. \end{aligned} \tag{2.15}$$

The symbol \simeq in these expressions is used to indicate that the approximation is only valid for models that are sufficiently close to a reference vector.

Moreover, the identity only holds to within some constant. This approximation for the forward model forms the point of departure for our seismic amplitude recovery algorithm.

To illustrate the above results, a numerical approximation of a PsDO is applied to a number of curvelets at different location, angles and scales. This PsDO corresponds to the normal operator associated with a seismic experiment with a velocity model containing a low-velocity lens (see Fig. 2.4(a)). The results of these experiments are summarized in Fig. 2.2. Fig. 2.2(b), 2.2(d) display the results of applying the PsDO to three different coarse- and fine-scale curvelets (Fig. 2.2(a), 2.2(c)). The Fourier-domain images are plotted in Fig. 2.2(f), 2.2(h). Comparing the imaged curvelets with their originals, shows that curvelets remain invariant as long as they lie in the support of the normal operator. Steep dipping curvelets are outside the support of the operator and this explains the deterioration of their amplitudes for steep dips. As predicted by Theorem 1, the normal operator becomes more diagonal for increasingly finer scales, an observation reflected in the behavior of the coarse- *versus* the fine-scale curvelets. Not only is this observation consistent with the microlocal correspondence of curvelets reported by Candes and Donoho (2000b) but it is also consistent with the fact that the PsDO's correspond to high-frequency approximations of the normal operator.

2.2.5 Estimation of the diagonal

The discrete normal operator: Before proceeding with the formulation of the seismic image recovery problem, a method is introduced to estimate the diagonal entries of \mathbf{D}_Ψ . These entries are calculated from applying a

2.2. Approximation of the normal operator

discrete implementation of the normal operator to some appropriately chosen reference vector. This discrete normal operator is given by the composition of the discretized scattering operator that relates the discretized reflectivity to the discretized data and its adjoint the migration operator. This compound operator corresponds to a matrix-free implementation for $\Psi := \mathbf{K}^T \mathbf{K} \in \mathbb{R}^{M \times M}$ with $\mathbf{K} \in \mathbb{R}^{L \times M}$ the scattering matrix and L and M the length of the data and image vectors.

The discrete curvelet transform: For the numerical implementation of the curvelet transform, the fast discrete curvelet transform via wrapping (Candes et al., 2006a) is used, which corresponds to a matrix-free implementation of the tall curvelet decomposition matrix $\mathbf{C} \in \mathbb{R}^{N \times M}$ with $N = \#\{\mathcal{M}\} \gg M$ with the symbol $\#$ denoting 'number of'. This transform yields a redundant coefficient vector $\mathbf{c} = \mathbf{C}\mathbf{r}$ with $\mathbf{c} \in \mathbb{R}^N$. For this choice of curvelet transform, the pseudo-inverse equals the transpose, i.e., $\mathbf{C}^T \mathbf{c} = \mathbf{C}^\dagger \mathbf{c}$. The transform is a numerical isometry that preserves energy, i.e., $\|\mathbf{r}\| = \|\mathbf{C}\mathbf{r}\|$, so we have $\mathbf{C}^T \mathbf{C}\mathbf{r} = \mathbf{I}\mathbf{r}$ in the ℓ_2 sense. Since the discrete curvelet representation is overcomplete, with a moderate redundancy (a factor of roughly 8 for $d = 2$), the converse is not the identity, i.e., $\mathbf{C}\mathbf{C}^T \mathbf{r}$ is a projection. Because $\mathbf{C}\mathbf{C}^T$ is a projection, not every curvelet vector is the forward transform of some discrete vector \mathbf{f} .

Non-uniqueness: The estimation of the diagonal \mathbf{D}_Ψ involves the calculation of a diagonal curvelet-domain weighting vector that approximates the action of the normal operator on an appropriately chosen reference vec-

2.2. Approximation of the normal operator

tor \mathbf{r} , i.e., $\mathbf{C}^T \mathbf{D}_\Psi \mathbf{C} \mathbf{r} \simeq \Psi \mathbf{r}$ (cf. Eq. 2.13). This estimation problem can be formulated in terms of a least-squares minimization problem. Define the 'data' vector as $\mathbf{b} = \Psi \mathbf{r}$, i.e., the migrated demigrated reference vector. Let $\mathbf{D}_\Psi = \text{diag}(\mathbf{u})$, with $\mathbf{u} \in \mathbb{R}^N$, be the diagonal curvelet-domain weighting matrix with coefficients $(u_\mu)_{\mu \in \mathcal{M}}$ on its diagonal and calculate the curvelet transform of the reference vector $\mathbf{v} = \mathbf{C} \mathbf{r}$. The diagonal estimation problem can now be formulated as finding the vector \mathbf{u} that minimizes $\|\mathbf{b} - \mathbf{P} \mathbf{u}\|_2$ with $\mathbf{P} := \mathbf{C}^T \text{diag}(\mathbf{v})$ and $\mathbf{b} \in \mathbb{R}^M$ the known demigrated-migrated reference vector. Since the curvelet transform is redundant ($M \ll N$), there are many possible solutions to this minimization problem, which gives rise to non-uniqueness.

Regularization via smoothness in phase space: One of the consequences of the non-uniqueness is the existence of negative entries in the estimated diagonal. These entries are inconsistent with the fact that Ψ is positive definite. Therefore, a regularization is needed that leads to an estimate for the diagonal with positive entries only. We argue that this can be accomplished by imposing smoothness in phase space. This smoothness is a property of the symbol that describes the behavior of the normal operator. In the curvelet domain, this smoothness translates to a smoothness in the amplitudes amongst neighboring points on the curvelet grid. For each entry in the diagonal, this smoothness can be imposed by including an additional penalty term in the least-squares formulation for the estimation of

2.2. Approximation of the normal operator

the diagonal. This damped least-squares problem can be formulated as

$$\tilde{\mathbf{u}} = \arg \min_{\mathbf{u}} \frac{1}{2} \|\mathbf{b} - \mathbf{P}\mathbf{u}\|_2^2 + \eta^2 \|\mathbf{L}\mathbf{u}\|_2^2, \quad (2.16)$$

where $\mathbf{L} = [\mathbf{D}_1^T \quad \mathbf{D}_2^T \quad \mathbf{D}_\theta^T]^T$ is a so-called sharpening operator, penalizing fluctuations amongst neighboring coefficients in \mathbf{u} . $\mathbf{D}_{1,2}$ contain first-order differences at each scale in the $x_{1,2}$ directions, i.e., $\mathbf{D}_{1,2} = [\mathbf{D}_{1,2}^{j_{min}} \cdots \mathbf{D}_{1,2}^{j_{max}}]$ with $\mathbf{D}_{1,2}^j$ the differences at the j^{th} scale in the spatial directions, and \mathbf{D}_θ the first-order differences in the θ direction, i.e., $\mathbf{D}_\theta = [\mathbf{D}_\theta^{j_{min}} \cdots \mathbf{D}_\theta^{j_{max}}]$ with \mathbf{D}_θ^j the difference in the angle direction at the j^{th} scale. These differences are scale dependent because the curvelet grid changes for each scale. The emphasis on the penalty term with respect to the misfit is controlled by η .

Estimation with positivity constraints: To ensure a correct balance between the misfit and the penalty functional, Eq. 2.16 is solved for a series of increasing Lagrange multipliers η . For each η , the following system of equations is inverted

$$\begin{bmatrix} \mathbf{P} \\ \eta \mathbf{L} \end{bmatrix} \mathbf{u} \approx \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix} \quad (2.17)$$

with \approx the approximation in least squares sense and $\mathbf{0}$ a column vector with N zero entries. For a given reference vector, the diagonal estimation procedure consists of the following steps. Apply the normal operator to the reference vector and calculate its curvelet coefficients. Set the Lagrange multiplier to η_{min} and solve Eq. 2.16 by inverting the system of equations in Eq. 2.17. Subsequently, increase η by $\Delta\eta$ and repeat this procedure until all

2.2. Approximation of the normal operator

diagonal elements of $\tilde{\mathbf{u}}$ are nonnegative. Even though we do not have a proof of convergence for this method, increasing the η leads to positive entries for large enough η . The steps of this estimation procedure are summarized in Table. 2.1.

<pre> Calculate: $\mathbf{b} = \Psi\mathbf{r}$ and $\mathbf{v} = \mathbf{C}\mathbf{r}$. Set: $\eta = \eta_{min}$; while $\exists (\tilde{u}_\mu)_{\mu \in \mathcal{M}} < 0$ do Solve $\tilde{\mathbf{u}} = \arg \min_{\mathbf{u}} \frac{1}{2} \ \mathbf{b} - \mathbf{P}\mathbf{u}\ _2^2 + \eta^2 \ \mathbf{L}\mathbf{u}\ _2^2$ Increase the Lagrange multiplier $\lambda = \eta + \Delta\eta$ end while </pre>
--

Table 2.1: Algorithm for the estimation of the diagonal via regularized least-squares inversion. The Lagrange multiplier, η , is increased up to the point that all entries in the vector for the diagonal are positive.

Example: The procedure outlined above is tested on a stylized example with a velocity model given by a low-velocity lens (Fig. 2.4(a)) and a bandwidth-limited reflectivity (Fig. 2.4(b)), consisting of three events. The low-velocity zone is representative of a gas lens in the overburden. Data is generated from this reflectivity with a zero-order linearized Born modeling and consists of 32 shot records with 500 receivers each. For more details on the migration code, refer to Symes (2006b) and to the numerical experiment section at the end of this paper. The data is subsequently imaged (Fig. 2.4(c)). This image and the bandwidth-limited reflectivity (Fig. 2.4(b)) define the 'data' vector \mathbf{b} (cf. Eq. 2.17) and the reference vector \mathbf{r} used to estimate $\tilde{\mathbf{u}}$ according to the algorithm presented in Table 2.1. Fig. 2.5 dis-

plays the curvelet vectors for increasing $\eta = \{0.01, 0.1, 1, 10\}$. These plots clearly illustrate that $\tilde{\mathbf{u}}$ becomes positive for η large enough ($\eta \geq 1$ in this case). Fig. 2.4(d) contains the diagonal approximation of the normal operator for the diagonal estimated with $\eta = 1$. Comparison between the results of applying the normal operator (Fig. 2.4(c)) and its diagonal approximation (Fig. 2.4(d)), shows a nice correspondence. The relative ℓ_2 -error between the actual image its approximation is 6.1%. This value corresponds to a moderate size for the constant C'' in the bound for the approximation error (Theorem 1).

2.3 Stable seismic image recovery

Our formulation for the seismic amplitude recovery problem banks on two fundamental properties of curvelets, namely their ability to **detect wavefronts** and their **invariance** under the normal operator. The combination of these two properties allow for a robust solution of the recovery problem in terms of a nonlinear optimization problem. In this section, the different steps that lead to this formulation are discussed starting with an argumentation why curvelets compress seismic images, followed by a quick review of ℓ_1 -promoted image recovery, its extension to the seismic situation and its solution by a cooling method.

2.3.1 Curvelet frames for seismic images

The sedimentary crust consists of sheet-like layers that correspond to functions with singularities along piece-wise smooth curves. These singularities

2.3. Stable seismic image recovery

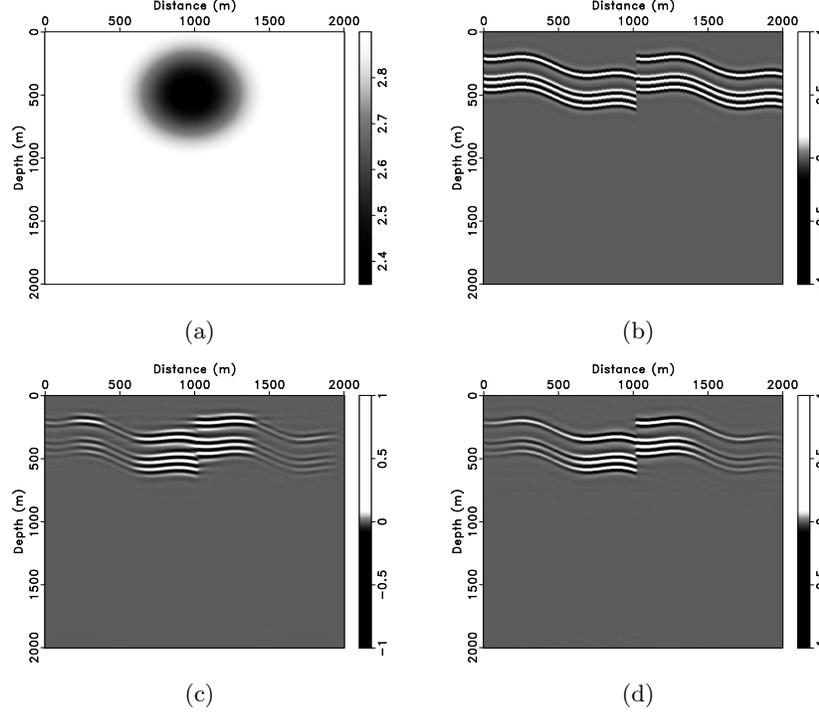


Figure 2.4: Stylized example of the diagonal estimation for a smooth velocity model and a reflectivity consisting of three reflection events including a fault, **(a)** the smooth background model with the low-velocity lens. This velocity model is used to calculate the linearized scattering and migration operators, **(b)** the bandwidth limited reflectivity, **(c)** the imaged reflectivity from data consisting of 32 shot records with 500 traces each, **(d)** the approximated normal operator for $\text{diag}(\tilde{\mathbf{u}})$ estimated with $\eta = 1$. The bandwidth limited reflectivity in **(b)** and the image in **(c)** served as input for the reference and 'data' vectors for the diagonal estimation procedure outlined in Table 2.1. Remarks: the main dimming of the normal operator is captured by the diagonal approximation. The relative ℓ_2 -error between the actual and the approximate normal operator is 6.1%.

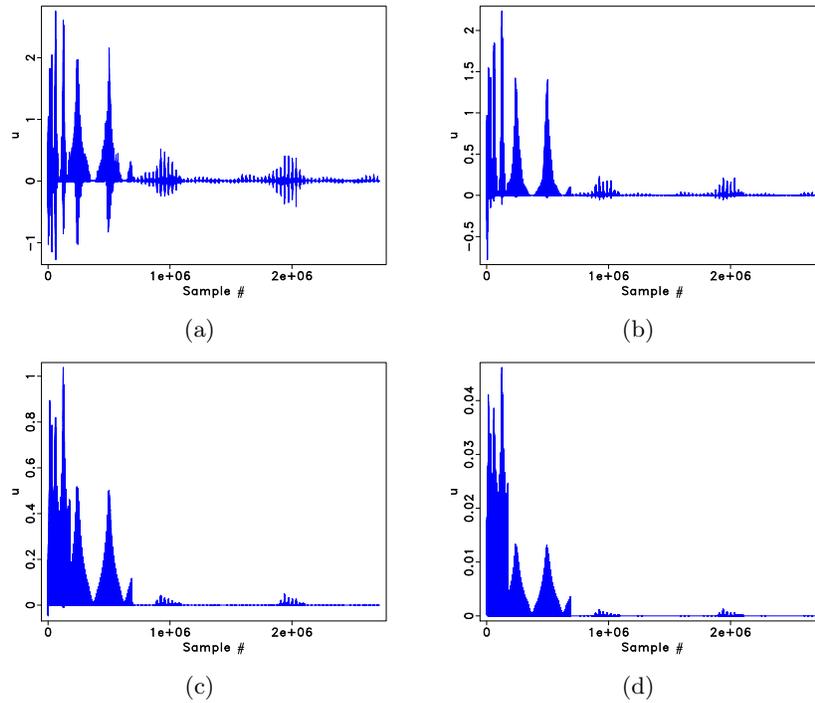


Figure 2.5: Estimates for the diagonal $\tilde{\mathbf{u}}$ are plotted in **(a-d)** for increasing $\eta = \{0.01, 0.1, 1, 10\}$. The diagonal is estimated according the procedure outlined in Table 2.1 with the reference and 'data' vectors, \mathbf{v} and \mathbf{b} , plotted in Fig. 2.4(b) and 2.4(c). As expected the diagonal becomes more positive for increasing η .

are associated with rapid variations in the medium fluctuations at which seismic waves reflect. During seismic imaging, these singularities are recovered. Examples of singularities imaged from real seismic data can be found in Fig. 2.1(a). This image can be considered as a bandwidth limited version of a function with singularities on piece-wise C^2 curves that may include faults and pinch outs (see e.g., Fig. 2.4(b) and 2.8(a) that are examples of synthetic reflectivity models). Fourier (and also wavelet) transforms are known to perform poorly on this type of functions, (Candes and Donoho, 2000a; Candes and Guo, 2002). This poor performance can be explained by the inability of the Fourier transform to localize and the lack of directivity of wavelets. Only when oriented perpendicular to the interfaces, wavelets decay rapidly. Since wavelets lack directionality, they can not resolve these directions, known as the wavefront set. This inability explains why wavelets do not significantly improve the compression rate for seismic images. Curvelet frames, on the other hand, are designed to detect the wavefront set (Candes and Donoho, 2000a, 2004, 2005a,b) and lead, as shown in Fig. 2.6, to better empirical compression rates for real and synthetic seismic images. This improved compression rate explains the accurate reconstruction in Fig. 2.1(a) of a real seismic image from only 3% of the curvelet coefficients and is consistent with the theoretical nonlinear approximations rates reported in the literature (see e.g. Candes and Donoho, 2000a). These rates correspond to a rate of $\mathcal{O}(P^{-2})$ for curvelets, with P the number of largest entries used in the approximation, while Fourier and wavelets only attain $\mathcal{O}(P^{-1/2})$ and $\mathcal{O}(P^{-1})$, respectively. The empirical rates plotted in Fig. 2.6 seem to confirm these theoretical findings.

2.3. Stable seismic image recovery

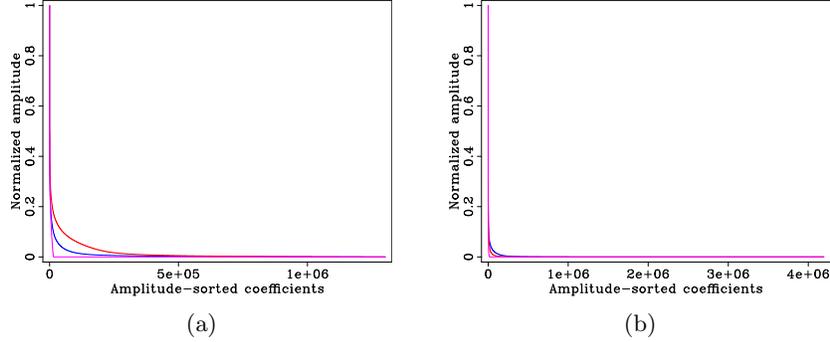


Figure 2.6: Decay rate for the curvelet coefficients compared to the decay for the coefficients of Fourier (dashed line) and discrete wavelet transforms (dot-dashed line), **(a)** decay rate for the sorted curvelet coefficients (solid line) of the real migrated image included in Fig. 2.1(a), **(b)** the same as **(a)** but now for the synthetic reflectivity of the SEG/EAGE AA' salt model (cf. Fig. 2.8(a)). The decay for the curvelet transform clearly compares favorably to these other two transforms.

2.3.2 Stable recovery

Because of the redundancy of the curvelet transform, $\mathbf{C}\mathbf{C}^T$ is a projection and this means that not every curvelet vector is a forward transform of a vector \mathbf{f} . Therefore, the vector \mathbf{x}_0 can't 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} . Recent work on in the field of compressive sensing has shown that rectangular matrices such as the curvelet matrix, $\mathbf{C} \in \mathbb{R}^{N \times M}$ with $N \gg M$, can stably be inverted through a nonlinear sparsity promoting optimization program. These nonlinear recovery methods require fast decay for the magnitude sorted curvelet coefficients (for details refer to the extensive literature on stable recovery also known as compressed sensing. See e.g. Starck et al., 2004; Elad et al., 2005; Candes et al., 2006b).

Following these results, the vector \mathbf{x}_0 can be approximately recovered

from noise-corrupted data

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{n} \quad (2.18)$$

with $\mathbf{A} := \mathbf{C}^T$. As long as the signal is sufficiently compressible \mathbf{x}_0 can successfully be recovered to within the noise level. This recovery involves the solution of a constrained convex optimization problem

$$\mathbf{P}_1 : \quad \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 := \sum_{j=1}^N |x_j| & \text{subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \\ \tilde{\mathbf{m}} = \mathbf{A}\mathbf{x}. \end{cases} \quad (2.19)$$

For certain matrices \mathbf{A} , the solution of this unconstrained optimization problem lies to within the noise level (see e.g. Candes et al., 2006b; Elad et al., 2005). The optimization problem \mathbf{P}_1 is known as the constrained variation of the LASSO (Tibshirani, 1997) and the basis-pursuit denoising (BPDN) (Chen et al., 2001) algorithms.

As part of the optimization, the sparsity vector is fitted within the tolerance ϵ . This tolerance depends on the noise level given by the standard deviation of the noise vector. Since $n_1, n_2, \dots, n_M \in N(0, \sigma^2)$, the probability of $\|\mathbf{n}\|_2^2$ exceeding its mean by plus or minus two standard deviations is small. The $\|\mathbf{n}\|_2^2$ is distributed according the χ^2 -distribution with mean $M \cdot \sigma^2$ and standard deviation $\sqrt{2M} \cdot \sigma^2$. By choosing $\epsilon^2 = \sigma^2(M + \nu\sqrt{2M})$ with $\nu = 2$, we remain within the mean plus or minus two standard deviations. Following (Elad et al., 2005), the above constrained optimization problem (\mathbf{P}_1), is replaced by a series of simpler unconstrained optimization

problems

$$\mathbf{P}_\lambda : \begin{cases} \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \\ \tilde{\mathbf{m}} = \mathbf{A}\tilde{\mathbf{x}}. \end{cases} \quad (2.20)$$

These optimization problems depend on the Lagrange multiplier λ . A cooling method is used, where \mathbf{P}_λ is solved for a Lagrange multiplier λ that is slowly decreased from a large starting value $\lambda_1 < \|\mathbf{A}^T \mathbf{y}\|_\infty$. The optimal $\tilde{\mathbf{x}}$ is found for the largest λ for which $\|\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}\|_2 \leq \epsilon$. During the optimization, the underdetermined frame matrix \mathbf{A} is inverted by imposing the sparsity promoting ℓ_1 -norm. This norm regularizes the inverse problem of finding the unknown coefficient vector (see also Daubechies et al., 2005). We refer to (Donoho et al., 2006; Tropp, 2006) for the recovery conditions for Eq.'s 2.19 and 2.20.

2.3.3 Solution by iterative thresholding

Following (Daubechies et al., 2005; Elad et al., 2005) and ideas dating back to (Figueiredo and Nowak, 2003), Eq. 2.27 is solved by an iterative thresholding technique that derives from the Landweber descent method. The method consist of an outer loop during which the Lagrange multiplier is lowered and an inner loop during which \mathbf{P}_λ is solved. The inner loop is initialized by the solution obtained from the previous outer loop starting with zero vector for the first iteration. After m iterations of the outer cooling loop, the estimated coefficient vector is computed for fixed λ by the following inner loop

$$\mathbf{x}^{m+1} = T_\lambda(\mathbf{x}^m + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x}^m)) \quad (2.21)$$

with $\lambda = \lambda_m$ and

$$T_\lambda(x) := \text{sgn}(x) \cdot \max(0, |x| - |\lambda|). \quad (2.22)$$

As shown by (Daubechies et al., 2005), this iteration for fixed λ converges to the solution of Eq. 2.27 for m large enough and $\|\mathbf{A}\| < 1$. The cost for each iteration is a curvelet synthesis and subsequent analysis.

2.3.4 Stable seismic recovery

The main purpose is to estimate the relative amplitudes of seismic images from data that are possibly contaminated with noise. This data is represented by the discretized linearized forward model

$$\mathbf{d} = \mathbf{K}\mathbf{m} + \mathbf{n} \quad (2.23)$$

with \mathbf{K} the discrete linearized Born modeling operator. Applying the adjoint of this operator to the data vector creates a noisy image (cf. Eq. 2.4)

$$\mathbf{y} = \Psi\mathbf{m} + \mathbf{e}. \quad (2.24)$$

With the curvelet decomposition of the normal operator (cf. Eq. 3.9) in discrete form, i.e.,

$$\Psi\mathbf{m} \simeq \mathbf{A}\mathbf{A}^T\mathbf{m}, \quad (2.25)$$

this expression can be rewritten into the following approximate form

$$\mathbf{y} \simeq \mathbf{A}\mathbf{x}_0 + \mathbf{e} \quad (2.26)$$

with $\mathbf{A} := \mathbf{C}^T\mathbf{\Gamma}$ and $\mathbf{\Gamma} := \sqrt{\mathbf{D}_\Psi}$. To avoid instabilities related to small entries in the estimated diagonal \mathbf{u} , we set $\mathbf{D}_\Psi = \text{diag}(\tilde{\mathbf{u}} + \delta)/\delta$ with δ some small parameter. Comparing this image representation with the one in Eq. 2.18 shows that these expressions are equivalent, aside from the noise term. While the noise in Eq. 2.18 is white and Gaussian, the noise term in Eq. 2.26 is colored by the migration operator. The expectation for the covariance of this colored noise term equals $\mathbf{E}\{\mathbf{e}\mathbf{e}^T\} = \sigma^2\mathbf{\Psi} \simeq \sigma^2\mathbf{A}\mathbf{A}^T$. Because the nonlinear recovery of the vector \mathbf{x}_0 involves the inversion of the matrix \mathbf{A} in which the covariance is factored, the noise term is approximately whitened. To understand this whitening, consider the pseudo-inverse $\mathbf{A}^\dagger = \mathbf{\Gamma}^{-1}\mathbf{C}$ which, when applied to \mathbf{e} , approximately whitens the noise in the curvelet domain. This whitening is a well-known property of WVD techniques.

The nonlinear recovery of seismic images now involves solving

$$\mathbf{P}'_1 : \quad \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 := \sum_{j=1}^N |x_j| & \text{subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \\ \tilde{\mathbf{m}} = (\mathbf{A}^T)^\dagger \tilde{\mathbf{x}}. \end{cases} \quad (2.27)$$

The main assumption underlying this formulation is that the curvelet vector of the model remains (approximately) sparse after application of the normal operator. This assumption is valid when the curvelets remain sufficiently

invariant under this operator. In that case, the nonlinear program (\mathbf{P}'_1) approximately inverts the normal operator in two stages, namely during the ℓ_1 -norm regularized inversion of the synthesis matrix, yielding a sparse estimate for the curvelet coefficient vector, and during the calculation of the model vector *via* the pseudo-inverse of \mathbf{A}^T . During each stage, the ‘square-root’ of the normal operator is approximately inverted.

2.4 Sparsity- and continuity-enhanced seismic image recovery

The above formulation provides a stable recovery for the amplitudes of seismic images, the relative strengths of the fluctuations in \mathbf{m} , without the necessity of multiple evaluations of the normal operator. This recovery is only accurate when the proposed diagonal approximation (cf. Eq. 3.3) provides an adequate approximation to the normal operator. The accuracy of this approximation depends on the available scales in the seismic image (cf. Theorem. 1), on the complexity of the background velocity model, the acquisition geometry and the closeness of the reference \mathbf{r} to the actual unknown model \mathbf{m} . For the remainder of this paper, we assume the acquisition to be ideal and the reference vector to be close to the actual model.

Anisotropic diffusion: Aside from the aforementioned factors that enter into the seismic recovery problem, there is the issue of how to limit spurious curvelet artifacts. These artifacts are either related to the so-called “pseudo Gibb’s” phenomenon (or better side-band effects (see Candes and Guo,

2002)), inherent to curvelet or other harmonic expansions. To reduce these spurious artifacts, the sparsity enhancing penalty functional in Eq. 2.27 is complemented with an additional continuity-enhancing penalty functional. This functional enhances the continuity along the imaged reflectors, which tend to be smooth in the tangential directions. By applying an anisotropic smoothing technique, the 'wavefront' set of the imaged reflectivity is preserved. This technique differs from commonly used edge-preserving penalty functionals such as total variation (TV) (see e.g. Claerbout and Muir, 1973; Schertzer, 2003) that tend to remove the oscillatory behavior of the reflectors in the normal direction.

The anisotropic-diffusion penalty term (see e.g. Fehmers and Hocker, 2003) is given by

$$J_c(\mathbf{m}) = \|\mathbf{\Lambda}^{1/2} \nabla \mathbf{m}\|_2^2, \quad (2.28)$$

with ∇ the discretized gradient matrix defined as $\nabla = [\mathbf{D}_1^T \quad \mathbf{D}_2^T]^T$. The block-diagonal matrix $\mathbf{\Lambda}$ is location dependent (see Fig. 2.10, which plots the gradients) and rotates the gradient towards the tangents of the reflecting surfaces. This rotation matrix is given by

$$\mathbf{\Lambda}[\mathbf{r}] = \frac{1}{\|\nabla \mathbf{r}\|_2^2 + 2v} \left\{ \begin{pmatrix} +\mathbf{D}_2 \mathbf{r} \\ -\mathbf{D}_1 \mathbf{r} \end{pmatrix} \begin{pmatrix} +\mathbf{D}_2 \mathbf{r} & -\mathbf{D}_1 \mathbf{r} \end{pmatrix} + v \mathbf{I} \right\}, \quad (2.29)$$

with \mathbf{D}_i the discretized derivative in the i^{th} coordinate direction and v a parameter that controls the fluctuations for regions where the gradient is small. Following Black et al. (1998), this control parameter is set proportional to the median of $|\nabla \mathbf{r}|$ with $||$ the length of each gradient vector (white

arrows in Fig. 2.10). Similar to the diagonal approximation, a reference vector derived from the migrated image (cf. Eq. 2.24) is used to calculate the tangential directions of the reflecting surfaces.

By combining the two different penalty terms that promote sparsity and continuity, we finally arrive at our formulation for the seismic-amplitude recovery problem

$$\mathbf{P}_\epsilon : \begin{cases} \arg \min_{\mathbf{x}} J(\mathbf{x}) & \text{subject to } \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \leq \epsilon \\ \tilde{\mathbf{m}} = (\mathbf{A}^T)^\dagger \mathbf{x}, \end{cases} \quad (2.30)$$

in which the composite penalty term $J(\mathbf{x})$ is given by

$$J(\mathbf{x}) = \alpha J_s(\mathbf{x}) + \beta J_c(\mathbf{x}), \quad (2.31)$$

with $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$. The $J_s(\mathbf{x}) = \|\mathbf{x}\|_1$ is the ℓ_1 -norm. The second term in the penalty term is given by $J_c(\mathbf{x}) = \|\mathbf{\Lambda}^{1/2} \nabla (\mathbf{A}^T)^\dagger \mathbf{x}\|_2^2$. Because the optimization is carried out over \mathbf{x} and not over the model vector \mathbf{m} , this expression includes a pseudo-inverse that is calculated with a few iterations of the LSQR algorithm (Paige and Saunders, 1982).

The cooling method: The above nonlinear optimization problem (\mathbf{P}_ϵ) is solved with a cooling method (see e.g. Starck et al., 2004). This method consists of a series of thresholded Landweber iterations that solve a series of unconstrained subproblems for decreasing λ . Since this method only requires knowledge of the Jacobians at each iteration, it is relatively straightforward to include the Jacobian of the additional continuity-enhancing penalty func-

tional $J_c(\mathbf{x})$. For data given by Eq. 2.18, the iterations of the cooling method for a particular cooling parameter λ consist of the following three main steps:

Step 1: update of the Jacobian of $\frac{1}{2}\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$:

$$\mathbf{x} \leftarrow \mathbf{x} + \mathbf{A}^T (\mathbf{y} - \mathbf{A}\mathbf{x}); \quad (2.32)$$

Step 2: project onto the ℓ_1 ball $S = \{\|\mathbf{x}\|_1 \leq \|\mathbf{x}_0\|_1\}$ by soft thresholding

$$\mathbf{x} \leftarrow T_\lambda(\mathbf{x}); \quad (2.33)$$

Step 3: project onto the anisotropic diffusion ball $C = \{\mathbf{x} : J(\mathbf{x}) \leq J(\mathbf{x}_0)\}$ by

$$\mathbf{x} \leftarrow \mathbf{x} - \beta \nabla_{\mathbf{x}} J_c(\mathbf{x}) \quad (2.34)$$

with

$$\nabla_{\mathbf{x}} J_c(\mathbf{x}) = 2\mathbf{A}^\dagger \nabla \cdot \left(\Lambda \nabla \left((\mathbf{A}^T)^\dagger \mathbf{x} \right) \right). \quad (2.35)$$

Remember that steps 1 & 2 with $\|\mathbf{A}\| \leq 1$ converge to the solution of Eq. 2.27 for a fixed λ . During step 3, the coefficients are updated according to the gradient of the anisotropic diffusion norm designed to reduce spurious artifacts. The different steps of our algorithm are summarized in Table 2.2. To ensure sparsity, the algorithm is started by setting $\mathbf{x}^0 = \mathbf{0}$.

<pre> Initialize: m = 0; x⁰ = 0; y = K^Td; Choose: M and L A^Ty _∞ > λ₁ > λ₂ > ... while y - Ax̃ ₂ > ε do m = m + 1; x^m = x^{m-1}; for l = 1 to L do x^m = T_{λ_m}(x^m + A^T(y - x^m)) {Iterative thresholding} end for Anisotropic descent update; x^m = x^m - β∇_{x^m} J_c(x^m); end while x̃ = x_m; m̃ = (A^T)[†] x̃. </pre>

Table 2.2: Sparsity-and continuity-enhancing recovery of seismic amplitudes.

2.5 Practical considerations

The presented method depends on the availability of a reference vector with an image that is sufficiently close to the unknown model. As in most Krylov-subspace solvers, we derive the reference vector from the migrated image (matched filter) as an initial guess. Since waves are subject to spherical spreading, a depth-dependent correction is applied to the migrated image (cf. Eq. 2.24). This corrected image serves as input for the diagonal estimation procedure and the calculation of the anisotropic diffusion norm. The image recovery procedure consists of the following steps

1. Obtain the reference vector, \mathbf{r} , by imaging the data with Eq. 2.24,

followed by applying the depth correction.

2. Use the reference vector, \mathbf{r} , to estimate the diagonal matrix, $\text{diag}(\tilde{\mathbf{u}})$, with the procedure outlined in Table 2.1;
3. Define the synthesis and analysis matrices \mathbf{A} and \mathbf{A}^T ;
4. Calculate the rotation matrix, $\mathbf{\Lambda}$, (cf. Eq. 2.29) for the continuity promoting penalty functional;
5. Solve the optimization problem \mathbf{P}_ϵ (Eq. 2.30) with the algorithm outlined in Table 2.2. This yields the estimate for the amplitude-corrected image, $\tilde{\mathbf{m}}$;
6. Optionally goto 1 with $\mathbf{r} := \tilde{\mathbf{m}}$;

2.6 Numerical experiments

The recovery method is applied to a synthetic imaging example for data of the SEG/EAGE AA' salt model (O'Brien and Gray, 1997; Aminzadeh et al., 1997). The original velocity model and its smoothed version that defines the background velocity model, used by the migration operators, are included in Fig. 2.7. To avoid multiple reflections, the background velocity model is averaged over a window of size 720×720 m. The reflectivity (velocity perturbation) is defined as the band-pass filtered difference between the smoothed velocity model and a slightly less smoothed velocity model. The density is assumed to be constant. Fig. 2.8(a) shows the reflectivity model that is 3.6 km by 15.3 km with a grid spacing of 24 m in the vertical and

horizontal coordinate directions. This reflectivity was used to generate data with the linearized Born modeling operator.

The migrated image: A two-way reverse-time migration code with checkpointing was used to conduct the experiments. This two-dimensional ($d = 2$) time-domain code, developed by Symes (2006b), has absorbing boundary conditions and is based on the adjoint-state method. The migration operator is derived from the gradient of a descent algorithm for least-squares inversion (see Symes, 2006b, and the references therein). The data is generated with the adjoint of the migration operator, which corresponds to the linearized Born approximation. This Born modeling operator generates data without nonlinear events such as multiple reflections. The migration operator and its adjoint are made zero-order by applying a half-order fractional integration on the source wavelet. We choose a source wavelet with a relatively broad temporal frequency band $[5 - 60]$ Hz. The data consists of 324 shots with 176 receivers each with a shot at every 48 m, yielding a maximum offset of 4.224 km. Each geophone records 626 time samples with a sample interval of 8 ms, which amounts to 5 s of data.

The 8000 time-step linearized Born modeling takes about 64 min with 68 CPU's on a MPI cluster, while the migration takes about 294 min. These computation times explain why calculating the pseudo-inverse of the normal operator becomes computationally prohibitive in real applications where the images are three dimensional ($d = 3$).

Fig. 2.8 shows the original reflectivity and the migrated image obtained with Eq. 2.24. The image suffers from amplitude deterioration due to spher-

2.6. Numerical experiments

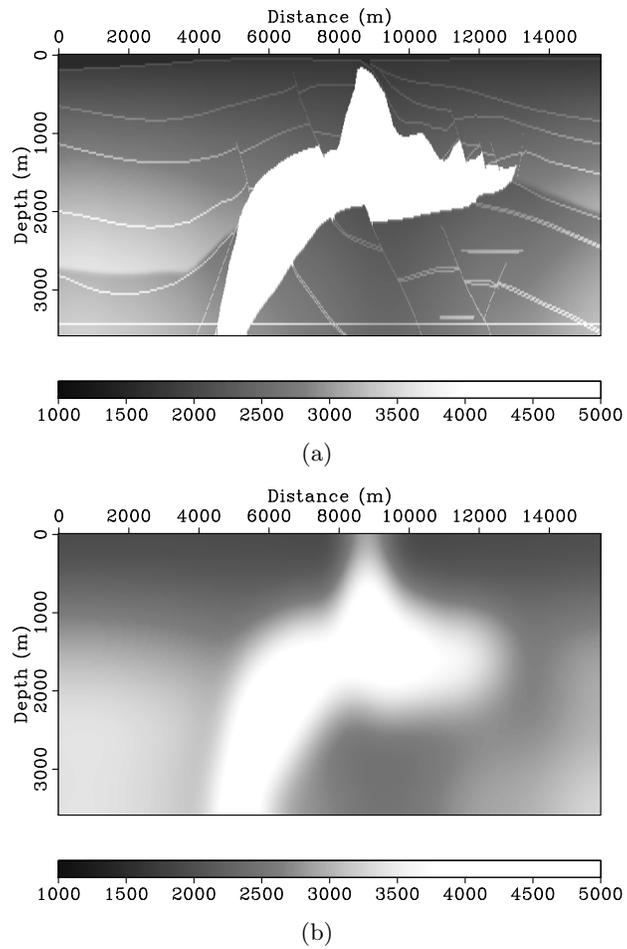


Figure 2.7: The SEG/EAGE AA' salt model (Aminzadeh et al., 1997), **(a)** The original velocity model, **(b)** the smoothed velocity model with a window size of 720×720 m. Remark: Imaging this model is a challenge because of the presence of the high-velocity (bright) salt.

2.6. Numerical experiments

ical spreading and poor illumination. Despite the overall amplitude deterioration for steep reflectors and events at increasing depths, the migration resolves most of the singularities in the image (cf. Fig 2.8(a) and 2.8(b)).

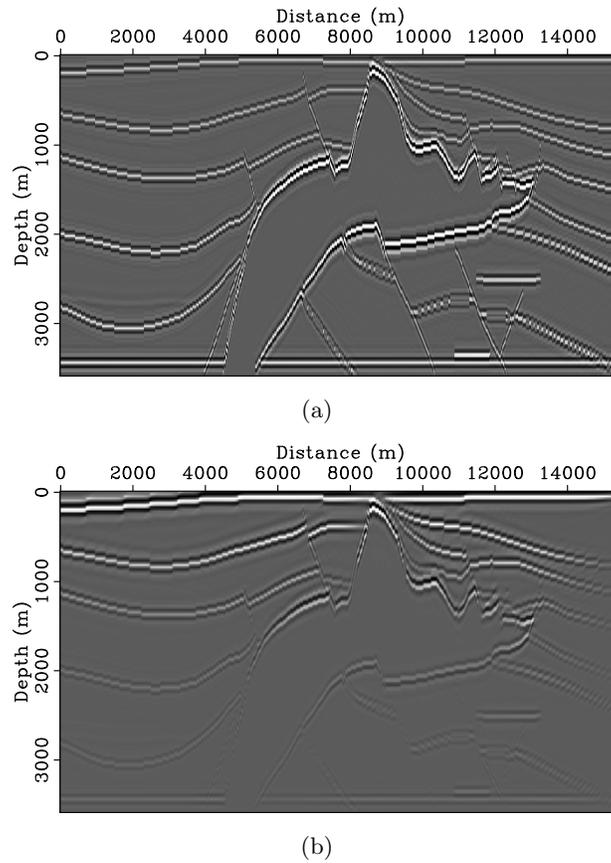
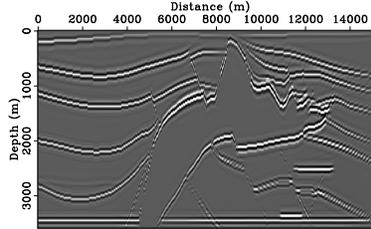


Figure 2.8: Imaging of the SEG/EAGE AA' salt model (Aminzadeh et al., 1997), **(a)** reflectivity defined in terms of the band-pass filtered difference between the smoothed velocity model and a slightly less smoothed velocity model, **(b)** the imaged reflectivity according to Eq. 2.24. The main reflection events are present but suffer from deteriorated amplitudes, especially under the high-velocity salt and for steep reflectors and faults.

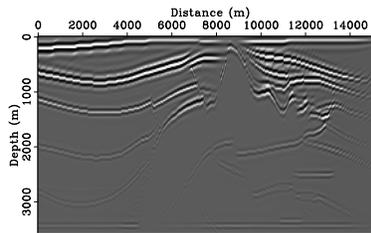
Diagonal estimation: After applying corrections for the spherical spreading, the migrated image is considered as reference vector. This reference vector is used to compute the diagonal approximation for the normal operator and for the calculation of the gradients. The image deterioration between the reference vector (Fig. 2.9(a)) and the demigrated-migrated reference vector (Fig. 2.9(b)) is similar to the amplitude deterioration between the 'unknown' true reflectivity and migrated image (cf. Fig. 2.8), therefore the images in Fig's. 2.9(a) and 2.9(b) should be adequate as input for the diagonal estimation procedure outlined in Table 2.1. To avoid instabilities related to small entries in the estimated diagonal $\text{diag } \tilde{\mathbf{u}}$, we set $\mathbf{\Gamma} = \text{diag}(\tilde{\mathbf{u}} + \delta)/\delta$ with $\delta = 0.2$. The result of applying the estimated diagonal to the bandwidth-limited reflectivity of Fig. 2.8(a) is included in Fig. 2.9(c) and shows that most of the imprint of the normal operator is captured by the diagonal approximation.

Seismic amplitude recovery for the SEG AA' model: With the diagonal approximation in place, the amplitude corrections are applied with the procedure presented in Table 2.2. The anisotropic diffusion penalty term is calculated from the gradients plotted in Fig. 2.10. The results of the amplitude recovery are summarized in Fig. 2.11. Fig. 2.11(a) depicts the result obtained by only applying a ℓ_1 -norm penalty ($\beta = 0$). The image is calculated by lowering the Lagrange multiplier λ in 20 steps from a value that corresponds to a threshold that removes all but 5% of the coefficients to a Lagrange multiplier that only removes 1% of the curvelet coefficients. For each intermediate λ , the inner loop is repeated 5 times ($L=5$). The

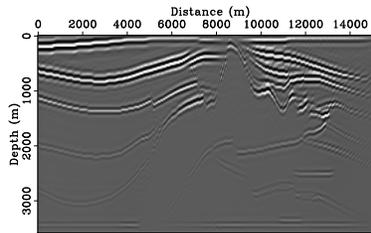
2.6. Numerical experiments



(a)



(b)



(c)

Figure 2.9: Images that are input to the diagonal approximation scheme outlined in Table 2.1, **(a)** the reference vector, \mathbf{r} , derived from Fig. 2.8(b), after applying corrections for the spherical-divergence and receiver-array taper, **(b)** the demigrated-migrated reference vector, $\mathbf{b} = \Psi\mathbf{r}$, that serves as 'data' for the diagonal estimation, **(c)** diagonal approximation on the original bandwidth-limited reflectivity plotted in **(a)**. This diagonal approximation by $\mathbf{A}\mathbf{A}^T\mathbf{m}$ captures the normal operator, $\Psi\mathbf{m}$, quite well.

2.6. Numerical experiments

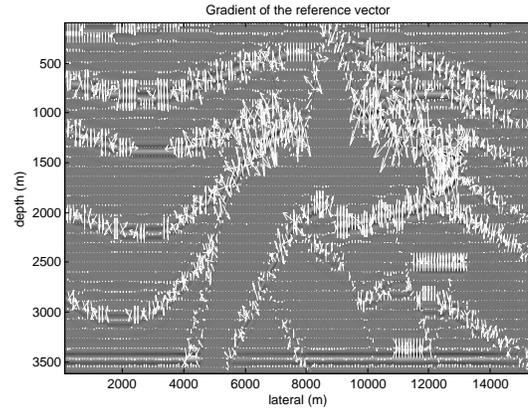


Figure 2.10: Gradient vectors for the reference vector \mathbf{r} plotted in Fig. 2.9(a). The gradients (white \uparrow 's) are used to calculate the tangential directions along which the additional anisotropic smoothing is applied.

result of this norm-one optimization are plotted in Fig. 2.11(a). The image shows a clear improvement over the image plotted in Fig. 2.9(a). Not only is the spherical divergence corrected in a stable manner, but the amplitude of the steep reflections events are recovered as well. There are, however, some small remaining artifacts related to side-band effects (Candes and Guo, 2002). By including the continuity-enhancing penalty functional most of the aforementioned artifacts can be removed as can be seen from Fig. 2.11(b). In this example, the sparsity- and continuity-enhancing norms are weighted equally, i.e. $\alpha = \beta = 1/2$ in Eq. 2.31. To further illustrate the migration amplitude recovery, detailed plots for different traces of the recovered the amplitudes are included in Fig. 2.12. Fig. 2.12(a) compares vertical traces of the original, migrated and amplitude recovered images. The traces are normalized with respect to the first reflector and the plots show that the amplitudes are nicely recovered. Similarly, the amplitudes along the horizontal reflector at $z = 3432$ m, are nicely recovered. The improvement in the recovered amplitudes is also apparent from Fig. 2.12(c) that contains the average vertical-wavenumber amplitude spectra of the original, the imaged and the recovered reflectivity.

Seismic amplitude recovery from noisy data: So far, the examples were noise free. In practice, seismic data volumes contain several sources of clutter, ranging from coherent nonlinear events, such as multiple reflections, to incoherent measurement errors. In this paper, only the recovery from incoherent contamination is considered. Sources of coherent clutter can be removed by using curvelet-based signal separation techniques reported

2.6. Numerical experiments

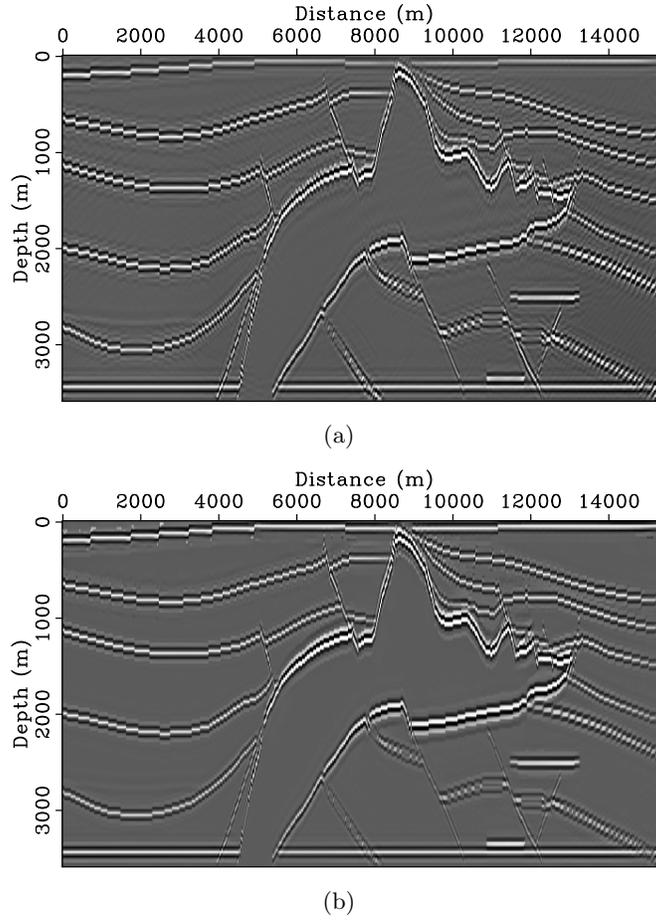


Figure 2.11: Images after nonlinear recovery (\mathbf{P}_e). **(a)** result with the ℓ_1 -norm recovery only ($\beta = 0$); **(b)** recovery with the combined ℓ_1 and continuity recovery for $\alpha = \beta = 1/2$. The amplitudes are recovered. The anisotropic diffusion successfully removes the artifacts.

2.6. Numerical experiments

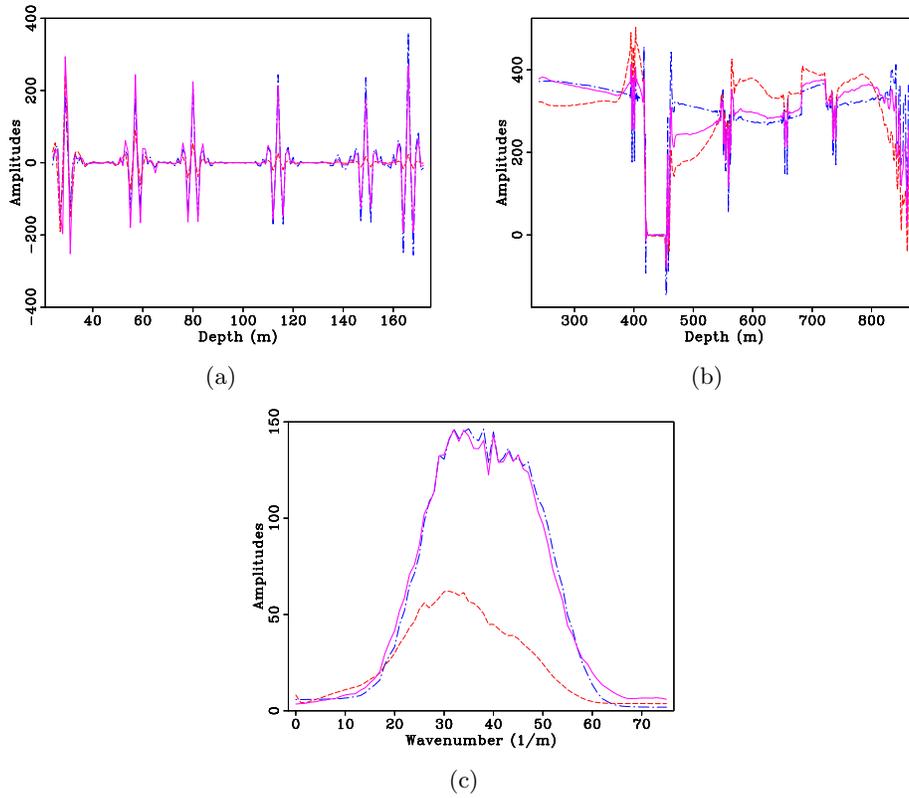


Figure 2.12: Recovery of the amplitudes after normalization with respect to the first event, **(a)** seismic traces of the original reflectivity (dot-dashed line), the migrated reflectivity (dashed line) and the recovered reflectivity (solid line), **(b)** amplitudes along the bottom most horizontal reflector, **(c)** average amplitude spectra of the depth-dependent reflectivity. The original and recovered spectra are normalized to match. Remarks: The nonlinear recovery corrects for the amplitudes and restores the spectrum.

elsewhere (Herrmann et al., 2007). An image computed from data with a signal-to-noise ratio (SNR) of 3 dB is included in Fig. 4.3(a). As shown in Eq. 2.24, the image now contains an additional contribution due to the noise. This noise contribution is clearly visible throughout the image as a non-stationary clutter and leads to a deterioration of the image quality. The clutter in the image space, however, is significantly smaller than in the data space, as can be observed from the noisy shot record plotted in Fig. 2.14(b). The difference in noise levels between the data and image spaces can be explained by the 154-fold redundancy of the data space compared to the image space ($L = 154 \times M$).

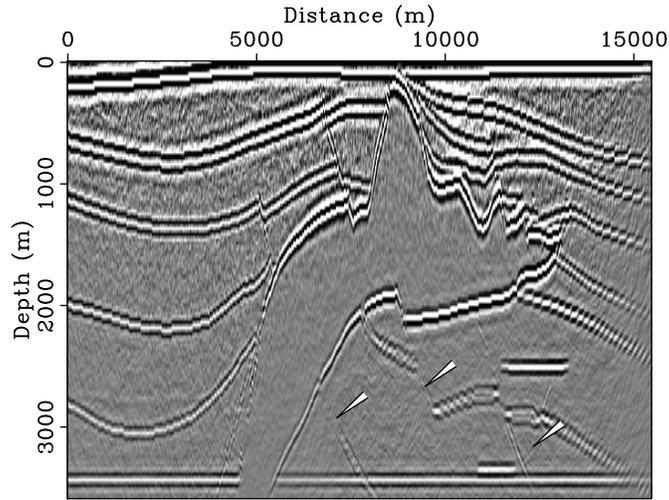
The result for the amplitude recovery of the noisy data are included in Fig. 4.3(b) and this figure shows that the migration amplitudes can stably be recovered, leading to a significant improvement in the SNR for the image (9.2 dB for the image in Fig. 4.3(b) as opposed to (1.5 dB for the noisy image which includes the error due to the imprint of the normal operator). This improved image was obtained using the algorithm of table 2.2, with the same settings as in the noise-free example except for the lowest value of the Lagrange multiplier, which is now set to a larger value depending calculated from the noise level. From Fig. 4.3(b), one can see that the noise contamination has largely been removed. Moreover, the amplitudes have been restored in particular for the steep events at depth. Data generated from the estimated image, $\tilde{\mathbf{d}} = \mathbf{K}\tilde{\mathbf{m}}$ shows a significant removal of the noise (cf. Fig. 2.14(b)-2.14(c)), with reflection events that match the noise-free data plotted in Fig. 2.14(a). This visual improvements leads to an improvement of SNR for the data of 19.2 dB.

2.7 Conclusions

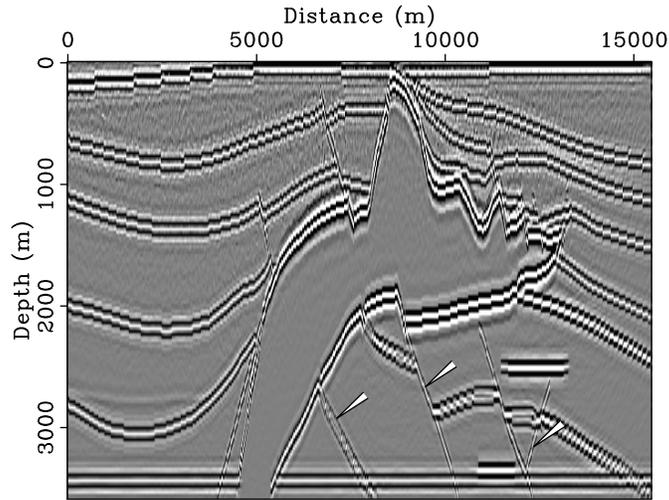
2.7.1 Initial findings

The method presented in this paper combines the compression of images by curvelets with the invariance of this transform under the normal operator. This combination allows for a formulation of a stable recovery method for seismic amplitudes. During the recovery, the normal operator is approximately inverted. Compared to other approaches for migration scaling, the presented method (i) includes a theoretical bound on the L^2 - error for the diagonal approximation in the curvelet domain; (ii) prescribes a procedure for the estimation of the diagonal from numerical implementations of the imaging operators; (iii) formulates the amplitude recovery problem as a nonlinear optimization problem, where the inversion of the diagonalized normal operator is regularized by imposing sparsity in the curvelet domain and continuity along the imaged reflectors.

As long as the background velocity model is sufficiently smooth and there is a reference vector available close enough to the actual reflectivity, the normal operator can be approximated by a diagonal weighting in the curvelet domain. The theoretical approximation error is scale-dependent and decreases for finer scales, an observation consistent with the microlocal properties of the normal operator and curvelets in the high-frequency limit. Estimation of the diagonal exploits smoothness of the symbol by penalizing neighboring entries in the diagonal that fluctuate. These fluctuations are inconsistent with the smoothness of the symbol. For a sufficiently large Lagrange multiplier, the regularized least-squares estimation for the diagonal



(a)



(b)

Figure 2.13: Image amplitude recovery for noisy data (SNR 3 dB), **(a)** noise image according to Eq. 2.24, **(b)** image after nonlinear recovery from noisy data (\mathbf{P}_ϵ). The clearly visible non-stationary noise in **(a)** is removed during the recovery while the amplitudes are also restored. Steeply dipping reflectors denoted by the arrows are also well recovered.

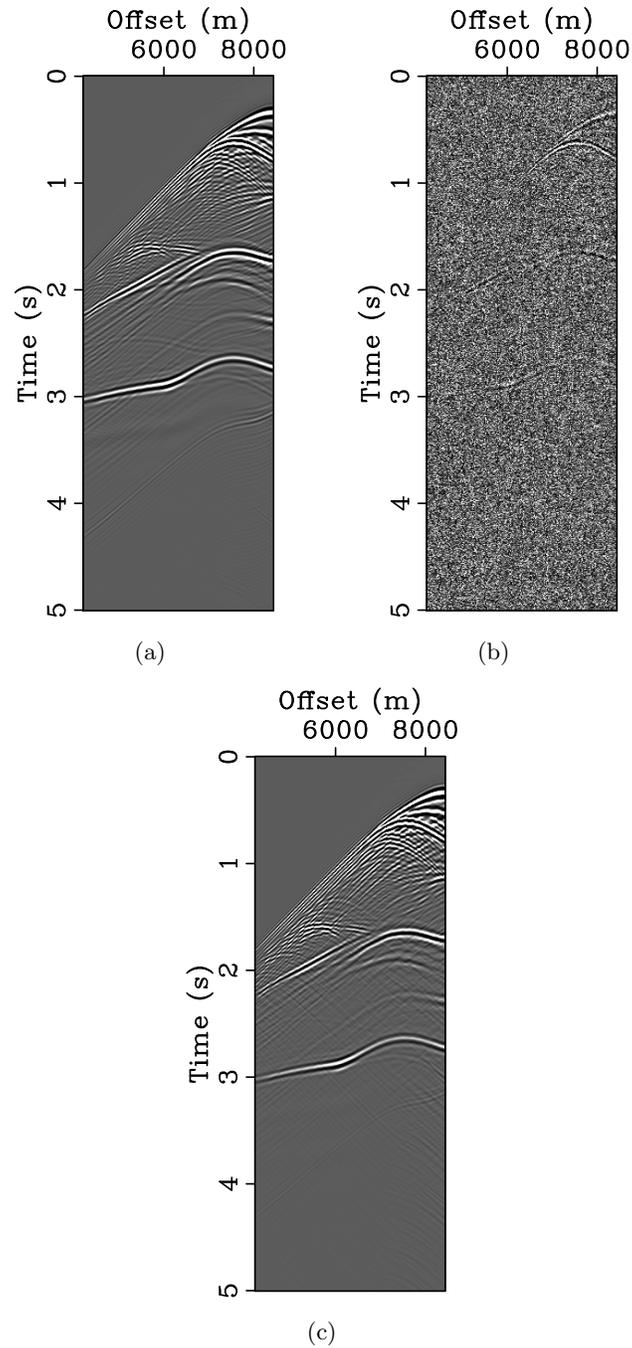


Figure 2.14: 'Denoising' of a shot record, (a) the noise-free data received by the receiver array, (b) noisy data with a SNR of 3 dB, (c) forward modeled data after amplitude recovery. Observe the significant improvement in the data quality, reflected in an increase for the SNR of 19.2 dB.

leads to positive entries. Numerical experiments with the diagonal approximation showed a moderate-sized constant for the bound of the theoretical error.

The diagonal approximation is used to formulate the seismic amplitude recovery in terms of a constrained optimization problem. The amplitude-corrected image is obtained by solving this sparsity- and continuity-promoting optimization problem. The invariance of curvelets under the normal operator preserves the sparsity. The cost of computing the diagonal approximation is one demigration-migration per reference vector, which is much less compared to the cost of Krylov-based least-squares inversion. The recovery results show an overall improvement of the image quality. The joined sparsity- and continuity-enhanced image has diminished artifacts, improved resolution and recovered amplitudes.

2.7.2 Extension to prestack imaging

The discussion has been on 'post-stack' images obtained after applying the zero-offset imaging condition. When the velocity model is correct, images are smooth along the redundant coordinate in pre-stack angle gathers. This smoothness property has been used in velocity analyses methods such as differential semblance (Symes and Carazzone, 1991) and is suitable for the framework laid down in this paper. Besides the ℓ_1 - and continuity norms, the amplitude recovery can also exploit the smoothness along the redundant angle coordinate. The fact that smoothness translates to sparsity is especially exciting because it allows to further exploit the results from the field of compressed sensing.

2.7.3 Recent related work

During the revision, we were informed of recent work by William Symes on optimal scaling of reverse-time migration Symes (2006a). This work and recent work by Guitton (2004) attempt to invert the normal operator by estimating a diagonal approximation from a reference vector, typically given by the migrated image, and the demigrated-migrated reference vector. Smoothness of the diagonal approximation is also imposed by both authors by parameterizing the diagonal as a smooth function in the spatial domain. Our approach goes several steps further by imposing smoothness in phase space and by making use of the curvelet transform that allows for the inversion of the diagonal using the methods of stable image recovery.

2.7.4 Choice of the reference vector

Having access to a proper reference vector is a prerequisite for the success of the method presented in this paper. Selection of the appropriate reference is somewhat reminiscent of finding a good initial guess for Krylov subspace methods. In this paper, we took a data-driven approach by using the migrated image (matched-filter result) as a first guess. Our formulation allows for multiple reference vectors and we hope to report on appropriate selection strategies for reference vectors in the future. This investigation will include an analysis of the sensitivity of our method to the accuracy of the velocity model.

Acknowledgments

The authors would like to thank the reviewers for their constructive comments and suggestions. We also wish to thank the authors of the Fast Discrete Curvelet Transform for making their code available at www.curvelet.org and Dr. William Symes for making his reverse-time migration code available. M. O'Brien, S. Gray and J. Dellinger from BP Amoco are thanked for making the SEG AA' dataset available. The examples presented in this paper were prepared with CWP's Seismic Un*x and with Madagascar (rsf.sourceforge.net). This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (22R81254) and the Collaborative Research and Development Grant DNOISE (334810-05) of F.J.H. This research 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 first two authors would also like to thank the Institute of Pure and Applied Mathematics at UCLA, supported by the NSF under grant DMS-9810282, for their hospitality. The research of the third author was supported by the Netherlands Organization for Scientific Research, through a VIDI grant no. 639.032.509.

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.
- Beylkin, G., 1984, The inverse problem and applications of the generalized Radon transform: *Comm. Pure Appl. Math.*, **37**, 579–599.
- Black, M., G. Sapiro, and D. Marimont, 1998, Robust anisotropic diffusion: *IEEE Trans. Signal Process.*, **7**, 421–432.
- Bleistein, N., J. Cohen, and J. Stockwell, 2001, *Mathematics of multidimensional seismic imaging, migration and inversion*: Springer.
- Brandsberg-Dahl, S. and M. V. de Hoop, 2003, Focusing in dip and AVA compensation on scattering-angle/azimuth common image gathers: *Geophysics*, **68**, 232–254.
- Candes, E. and D. Donoho, 2005a, Continuous curvelet transform i: Resolution of the wavefront set: *Appl. Comput. Harmon. Anal.*, **19**, 162–197.
- Candes, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast discrete curvelet transforms: *SIAM Multiscale Model. Simul.*, **5**, 861–899.
- Candes, E. J. and D. Donoho, 2004, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities: *Comm. Pure Appl. Math.*, **57**, 219–266.

- , 2005b, Continuous curvelet transform ii: Discretization and frames: *Appl. Comput. Harmon. Anal.*, **19**, 198–222.
- Candes, E. J. and D. L. Donoho, 2000a, Curvelets – a surprisingly effective nonadaptive representation for objects with edges: Presented at the Curves and Surfaces, Vanderbilt University Press.
- , 2000b, Recovering edges in ill-posed problems: Optimality of curvelet frames: *Ann. Statist.*, **30**, 784.
- Candes, E. J. and F. Guo, 2002, New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction: *Signal Processing*, 1519–1543.
- Candes, E. J., J. Romberg, and T. Tao, 2006b, Stable signal recovery from incomplete and inaccurate measurements: *Comm. Pure Appl. Math.*, **59**, 1207–1223.
- Chavent, G. and R. Plessix, 1999, An optimal true-amplitude least-squares prestack depth-migration operator: *Geophysics*, **64**, 508–515.
- Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: *SIAM Journal on Scientific Computing*, **43**, 129–159.
- Claerbout, J. and F. Muir, 1973, Robust modeling with erratic data: *Geophysics*, **38**, 826–844.
- Claerbout, J. and D. Nichols, 1994, Spectral preconditioning: Technical report, Stanford Exploration Project.
- Daubechies, I., M. Defrise, and C. de Mol, 2005, An iterative thresholding algorithm for linear inverse problems with a sparsity constraints: *Comm. Pure Appl. Math.*, **57**, 1413–1457.
- de Hoop, M. V. and N. Bleistein, 1997, Generalized Radon transform inver-

- sions for reflectivity in anisotropic elastic media: *Inverse Problems*, **13**, 669–690.
- de Hoop, M. V. and S. Brandsberg-Dahl, 2000, Maslov asymptotic extension of generalized Radon transform inversion in anisotropic elastic media: a least-squares approach.: *Inverse problems*, **16**, 519–562.
- Donoho, D. L., 1995, Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition: *App. and Comp. Harmonic Analysis*, **2**, 101–126.
- Donoho, D. L., M. Elad, and V. Temlyakov, 2006, Stable recovery of sparse overcomplete representations in the presence of noise: *IEEE Trans. Inform. Theory*, **52**, 6–18.
- Elad, M., J. L. Starck, P. Querre, and D. L. Donoho, 2005, Simultaneous cartoon and texture image inpainting using morphological component analysis (mca): *Appl. Comput. Harmon. Anal.*, **19**, 340–358.
- Fehmers, G. C. and C. F. Hocker, 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 Trans. Image Processing*.
- Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: *Geophysics*, **69**, 1017–1024.
- Hennenfent, G. and F. J. Herrmann, 2006, Seismic denoising with non-uniformly sampled curvelets: *Comp. in Sci. and Eng.*, **8**, 16–25.
- Herrmann, F. J., 2003, Multifractional splines: application to seismic imaging: *Proceedings of SPIE Technical Conference on Wavelets: Applications in Signal and Image Processing X*, 240–258.

- Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007, Nonlinear primary-multiple separation with directional curvelet frames: *Geoph. J. Int.*, **170**, 781–799.
- Hormander, L., 1987, The analysis of linear partial differential operators iii: Pseudo-differential operators.
- Hu, J., G. T. Schuster, and P. Valasek, 2001, Poststack migration deconvolution: *Geophysics*, **66**, 939–952.
- Kuhl, H. and M. Sacchi, 2003, Least-squares wave-equation migration for AVP/AVA inversion: *Geophysics*, **68**, 262.
- Lee, N. Y. and B. J. Lucier, 2001, Wavelet methods for inverting the Radon transform with noisy data: *IEEE Trans. Image Processing*, **10**, 79–94.
- 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, 1997, Can we image beneath salt?: *Leading Edge*, **15**, 17–22.
- Paige, C. C. and M. A. Saunders, 1982, Lsqr: An algorithm for sparse linear equations and sparse least squares: *ACM TOMS*, **8**, 43–71.
- Plessix, R. and W. Mulder, 2004, Frequency-domain finite difference amplitude-preserving migration: *Geoph. J. Int.*, **157**, 975–987.
- Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth migration: *Geophysics*, **68**, no. 4.
- Schertzer, O., 2003, Scale-space methods and regularization for denosing and inverse problems: *Advances in Imaging and Electron Physics*, **128**, 445–530.
- 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**.

Stolk, C. C., 2000, Microlocal analysis of a seismic linearized inverse problem: *Wave Motion*, **32**, 267–290.

Stolk, C. C. and W. W. Symes, 2003, Smooth objective functionals for seismic velocity inversion: *Inverse Problems*, **19**, 73–89.

Symes, W. W., 2006a, Optimal scaling for reverse time migration: Technical Report TR 06-19, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

———, 2006b, Reverse time migration with optimal checkpointing: Technical Report 06-18, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

Symes, W. W. and J. Carazzone, 1991, Velocity inversion by differential semblance optimization: *Geophysics*, **56**, 2061–2073.

ten Kroode, A., D.-J. Smit, and A. Verdel, 1998, A microlocal analysis of migration: *Wave Motion*, **28**.

Tibshirani, R., 1997, Regression shrinkage and selection via the LASSO: *J. Royal. Statist. Soc B.*, **58**, 267–288.

Tropp, J. A., 2006, Just relax: convex programming methods for identifying sparse signals in noise: *IEEE Trans. Inform. Theory*, **52**, 1030–1051.

Tsaig, Y. and D. L. Donoho, 2006, Extensions of compressed sensing: *Signal processing*, **86**, 549–571.

Ying, L., L. Demanet, and E. J. Candes, 2005, 3D discrete curvelet transform: Presented at the Wavelets XI, SPIE.

Chapter 3

Curvelet-based migration preconditioning and scaling

3.1 Introduction

Over the years, extensive research has been done to reduce the computational costs of (least-squares) seismic imaging. Improvements in this area are particularly important during iterative least-squares migration, where the linear Born-scattering operator is inverted with iterative Lanczos methods, such as LSQR (Paige and Saunders, 1982; De Roeck, 2002). Examples of these methods can be found in the literature (see e.g. Nemeth et al., 1999; Chavent and Plessix, 1999; Hu et al., 2001; Kuhl and Sacchi, 2003; Yu et al., 2006).

The most successful methods to reduce the cost of migration are the so-called scaling methods where the action of the compound linearized modeling-migration operator—known as the Hessian or normal operator—is replaced by a diagonal scaling in some domain, see e.g. contributions by Claerbout

A version of this chapter has been published. Herrmann, F.J., Brown, C.R., Erlangga, Y.A. and Moghaddam, P.P. (2009) Curvelet-based migration preconditioning and scaling. *Geophysics*, 74, pp. A41, 2009

3.1. Introduction

and Nichols (1994); Rickett (2003); Guitton (2004); Plessix and Mulder (2004), and more recently by Symes (2008) and Herrmann et al. (2008a). These methods vary in degree of sophistication with regard to the estimation of the diagonal, e.g. through migrated-image to remigrated-image matching. They also differ in the way the scaling is applied—i.e., by ‘division’ in the physical domain or via sparsity promotion in the curvelet domain, as reported recently by Herrmann et al. (2008a). During all these methods, imaged amplitudes are restored by applying a scaling as a post-processing step after migration. Even though our approach and the one of Symes (2008) are similar, the curvelet-domain scaling has the advantage that it is able to handle conflicting dips.

In this paper, we take this line of research a step further by using the above scaling argument to apply the proper preconditioning to the system of equations involved in linearized Born scattering. By “proper” we mean a preconditioner with a computational overhead that justifies its improvement by the increase in convergence rate. Because the system is solved iteratively, the preconditioner does not need to be very accurate, avoiding unnecessary extra computational overhead. Note that we use the term preconditioning somewhat loosely compared to its formal definition in numerical linear algebra where the solution is not changed with preconditioning (Calvetti, 2007). Therefore, we also use this term to denote changes in the forward model that favor least-squares inversion. To illustrate the improvements in the images and in the convergence of least-squares migration, we consider three levels of preconditioning. First, we correct the order of the normal operator by introducing a left preconditioning, consisting of a fractional

3.2. Problem formulation

time integration that corresponds to a scaling in the Fourier domain. This first level of preconditioning follows directly from earlier work on migration-amplitude recovery reported by Herrmann et al. (2008a) and Symes (2008). The next level of preconditioning consists of diagonal scaling in the physical domain that compensates for spherical spreading of seismic waves. As a final step, we also include a curvelet-domain scaling as part of the right preconditioning. This step corrects for the remaining amplitude errors that vary spatially as a function of the reflector dip. We conclude by studying the performance of these different levels of preconditioning on the SEG AA' salt model (O'Brien and Gray, 1996; Aminzadeh et al., 1997), using a reverse-time 'wave-equation' migration code with optimal checkpointing (Symes, 2007). Because our main interest is to study the performance of our preconditioner, we work on data generated by the linearized Born-scattering operator instead of data generated by the solution of the full nonlinear forward model.

3.2 Problem formulation

During seismic imaging, the following system of equations needs to be solved

$$\mathbf{Ax} \approx \mathbf{b}, \tag{3.1}$$

where $\mathbf{b} = \mathbf{Ax}_0$ is the multiple-free data with \mathbf{x}_0 the true reflector model, \mathbf{A} the linearized Born-scattering operator, and \mathbf{x} the unknown model vector. Here, the symbol \approx refers to "equality" in the least-squares sense. Con-

3.2. Problem formulation

trary to many inverse problems, the matrix \mathbf{A} , albeit extremely large, is reasonably well behaved and a migrated image can be obtained by applying the adjoint of \mathbf{A} to the data vector—i.e., $\tilde{\mathbf{x}} = \mathbf{A}^*\mathbf{b}$ with \mathbf{A}^* the migration operator. The symbol $*$ denotes the adjoint, and $\tilde{\mathbf{x}}$ is the estimate (denoted by the $\tilde{}$) of the image.

Unfortunately, the output of the above procedure, called migration, produces erroneous results for the amplitudes of the imaged reflectors. To restore these amplitudes, the least-squares solution to Equation 3.1 can be obtained as the solution of the linear system

$$\mathbf{A}^*\mathbf{A}\mathbf{x} \approx \mathbf{A}^*\mathbf{b}, \quad (3.2)$$

with $\mathbf{A}^*\mathbf{A}$ the normal or Hessian operator. Solutions to this system are not unique and correspond to solutions of the least-squares method—i.e.,

$$\tilde{\mathbf{x}}_{LS} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2, \quad (3.3)$$

which finds image vectors, \mathbf{x} , that after modeling fit the data vector, \mathbf{b} . However, the matrix $\mathbf{A}^*\mathbf{A}$ is not invertible—i.e., it has zero or small singular values that correspond to shadow zones. Because the system is large, we employ an iterative matrix-free solution method, such as LSQR (Paige and Saunders, 1982). By limiting the number of iterations for this method, we control the energy of the solution vector $\tilde{\mathbf{x}}$ and we obtain a regularized solution that is equivalent to the solution of a damped least-squares problem (Vogel, 2002). Even though these iterative methods converge relatively

3.2. Problem formulation

quickly (see e.g. Nemeth et al., 1999; Chavent and Plessix, 1999; Hu et al., 2001; Kuhl and Sacchi, 2003; Yu et al., 2006), the sheer size of the imaging problem calls for a further reduction in the number of iterations.

In a perfect world, with infinite computational resources, the ideal preconditioning for the system in Equation 3.1 corresponds to

$$\mathbf{A}\mathbf{M}_R^{-1}\mathbf{u} \approx \mathbf{b}, \quad \mathbf{x} := \mathbf{M}_R^{-1}\mathbf{u}, \quad (3.4)$$

with the right preconditioning matrix, $\mathbf{M}_R := (\mathbf{A}^*\mathbf{A})^{1/2}$, given by the ‘square-root’ of the normal operator. Here, the symbol $:=$ refers to ‘defined as’. In this ideal case, migration recovers the image vector \mathbf{x} , exactly. Unfortunately, in practice (albeit some recent exciting progress has been made by Demanet and Ying, 2008, using discrete symbol calculus for smooth symbols, a development on which we intend to report in the future) the quantity $(\mathbf{A}^*\mathbf{A})^{1/2}$ cannot be computed and we have to resort to appropriate approximations.

In this paper, we propose a combination of left and right preconditioning—i.e., we replace Equation 3.1 by

$$\overbrace{\mathbf{M}_L^{-1}\mathbf{A}\mathbf{M}_R^{-1}}^{\widehat{\mathbf{A}}}\mathbf{u} \approx \overbrace{\mathbf{M}_L^{-1}\mathbf{b}}^{\widehat{\mathbf{b}}}, \quad \mathbf{x} := \mathbf{M}_R^{-1}\mathbf{u}, \quad (3.5)$$

with \mathbf{M}_L^{-1} the left preconditioning matrix. The migrated and least-squares migrated images are given by $\tilde{\mathbf{x}} = \mathbf{M}_R^{-1}\tilde{\mathbf{u}}$, with $\tilde{\mathbf{u}} = \widehat{\mathbf{A}}^*\widehat{\mathbf{b}}$, and by $\tilde{\mathbf{x}}_{LS} = \mathbf{M}_R^{-1}\tilde{\mathbf{u}}_{LS}$, with $\tilde{\mathbf{u}}_{LS} = \arg \min_{\mathbf{u}} \|\widehat{\mathbf{b}} - \widehat{\mathbf{A}}\mathbf{u}\|_2$, respectively. Our preconditioners are derived from the following three observations: (i) under certain

3.3. Preconditioning

conditions—such as the high-frequency limit, smooth background velocity models, and the absence of turning waves (Stolk, 2000)—the normal operator is in d dimensions a $(d - 1)$ -order pseudodifferential operator (Ψ DO, see e.g. recent work by Herrmann et al., 2008a; Symes, 2008, and the references therein), (ii) migration amplitudes decay with depth due to spherical spreading of seismic body waves, and (iii) zero-order Ψ DO’s can be approximated by a diagonal scaling in the curvelet domain (see e.g. Herrmann et al., 2008a). These observations allow us to define a series of increasingly more accurate approximations to the ‘square-root’ of the normal operator, leading to better and better preconditioners. Finally, we also argue that using curvelets will add a certain robustness to Gaussian noise and modeling errors, an observation substantiated by successful applications of this transform in seismic data processing (Herrmann et al., 2008b; Wang et al., 2008).

3.3 Preconditioning

In this section, we introduce different types of preconditioners based on the aforementioned observations. For each preconditioned system, we study the migrated images. Later, we study the convergence of the iterative solver. The examples are computed for the reflectivity and smooth velocity background models plotted in Figure 3.1. To test our preconditioner, data with 324 shots is generated using Equation 3.1. Each shot consists of 176 traces of 6.4s and with a trace interval of 24m. The maximum offset of the data is 4224m.

3.3.1 Left preconditioning by fractional differentiation

As stated before, under aforementioned conditions the normal operator corresponds to a $(d - 1)$ -order Ψ DO. In 2-D image space, this operator corresponds to the leading-order behavior of the square root of Laplacian—i.e., the action of $(\Delta)^{1/2}$ in the physical domain, or to $|\xi| \cdot$ with ξ the wave vector in the spatial Fourier domain. In data space, this action corresponds to a multiplication by $|\omega|$ in the temporal Fourier domain (Herrmann et al., 2008a). We compensate for this action by defining the following left preconditioning:

$$\mathbf{M}_L^{-1} := \partial_{|t|}^{-1/2}, \quad (3.6)$$

where $\partial_{|t|}^{-1/2} \cdot := F^* |\omega|^{-1/2} F \cdot$ with F the Fourier transform and $F^* = F^{-1}$ its inverse. We define the level I preconditioner with \mathbf{M}_L^{-1} as above and $\mathbf{M}_R^{-1} = \mathbf{I}$ (which implies $\mathbf{x} = \mathbf{u}$). Comparison of the migrated images before and after left preconditioning (cf. Figures 3.1(c) and 3.2(a)) shows that the imprint of the Laplacian is removed—i.e., some of the low-frequency content is restored. However, the migrated image still contains dimming of the amplitudes. Note that this left preconditioning corresponds to the solution of the scaled least-squares problem: $\arg \min_{\mathbf{x}} \frac{1}{2} \|\widehat{\mathbf{b}} - \widehat{\mathbf{A}}\mathbf{x}\|_2^2 = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{M}_L^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x})\|_2^2$.

3.3.2 Right preconditioning by scaling in the physical domain

To further correct the amplitudes, we propose to apply a scaling to compensate for the leading-order amplitude decay. This decay is linear because the reflected waves travel from the source down to the reflector, experi-

3.3. Preconditioning

encing an amplitude decay proportional to the square-root of the reflector depth (in 2-D). We correct this linear amplitude decay by defining the right preconditioning matrix:

$$\mathbf{M}_R^{-1} = \mathbf{D}_z := \text{diag}(\mathbf{z})^{\frac{1}{2}}, \quad (3.7)$$

where $z_i = i\Delta z$, $i = 1 \cdots n_z$, with Δz the depth sample interval and n_z the number of samples. Combined with the left preconditioner \mathbf{M}_L^{-1} , we call this the level II preconditioner. The results for the migrated image in Figure 3.2(b) now show further improvement. However, amplitude variations remain, e.g., along the major horizontal reflector just above 3500m.

3.3.3 Right preconditioning by scaling in the curvelet domain

After applying the left preconditioning, the Hessian can be modeled by a zero-order Ψ DO whose action corresponds to that of a nonstationary dip filter—i.e., we have

$$(\Psi f)(x) \simeq \int_{\xi \in \mathbb{R}^d} e^{j\xi \cdot x} a(x, \xi) \hat{f}(\xi) d\xi, \quad (3.8)$$

with Ψ the Hessian of the preconditioned modeling operator and $a(x, \xi)$ a space- and spatial-frequency dependent filter known as the symbol. We use the symbol \simeq to indicate high-frequency approximation and absence of turning waves. Following ideas that go back to Taylor (1981), a pseudodifferential operator can approximately be diagonalized by linear combinations of

3.3. Preconditioning

localized oscillatory functions, such as curvelets (Candès et al., 2006), wave atoms (Demanet and Ying, 2007), and local Fourier bases (Meyer, 1992). In this paper, we use recent results by Herrmann et al. (2008a) who use curvelets because they have the additional advantage of being sparse on the model. The action of the Ψ DO can, after discretization, be approximated by a scaling in the curvelet domain—i.e., we have the following approximate identity

$$\Psi \mathbf{r} \approx \mathbf{C}^* \mathbf{D}_\Psi^2 \mathbf{C} \mathbf{r}, \quad \mathbf{D}_\Psi^2 := \text{diag}(\mathbf{d}^2), \quad (3.9)$$

which is accurate for a reference vector \mathbf{r} close enough to the actual image. In this expression, the matrices \mathbf{C} and \mathbf{C}^* represent the 2-D discrete curvelet transform (see e.g. Candès et al., 2006) for which the adjoint equals the pseudoinverse—i.e., we have $\mathbf{C}^* \mathbf{C} = \mathbf{I}$ with \mathbf{I} the identity matrix. The reciprocal of the curvelet-domain scaling coefficients, \mathbf{d}^{-2} , is found by a remigrated image-to-image matched-filtering procedure that involves the reference vector, typically derived from a conventional migrated image, and the remigrated reference vector. Hence, the cost of forming Equation 3.9 is approximately one modeling and one migration (for further details refer to Herrmann et al., 2008a,c). As we will show below, these additional costs are well offset by the increase in convergence.

By including the above approximation, we define the right preconditioning matrix as

$$\mathbf{M}_R^{-1} = \mathbf{D}_z \mathbf{C}^* \mathbf{D}_\Psi^{-1}, \quad (3.10)$$

which compensates for the remaining amplitude errors. Because there is no need for high accuracy, Equation 3.9 is a good approximation whose ac-

curacy increases with frequency (Herrmann et al., 2008a). Note, however, that this preconditioner by virtue of the underlying assumptions will not help in the removal of nonlocal imaging artifacts. We call the combination of Equations 3.6 and 3.10, the level III preconditioner. The image obtained with this system is plotted in Figure 3.2(c) and shows, as expected, further improvement in the amplitudes of the migrated image. Despite this improvement, image artifacts and amplitude errors are still present and can be attributed to the approximation in Equation 3.9 and to the fact that migration does not correspond to inversion—i.e., $\mathbf{A}^*\mathbf{A} \neq \mathbf{I}$.

3.4 Convergence of least-squares migration

Even though the different preconditioning operators defined so far lead to improvements, problems remain in achieving high-fidelity images. As reported in the literature, the image quality can be further improved by replacing migration with least-squares migration where the (preconditioned) scattering matrix is inverted iteratively. After a limited number of iterations, this method inverts the (preconditioned) system of equations approximately.

The performance of iterative solvers depends on certain properties of the (preconditioned) matrix \mathbf{A} , which include its condition number (ratio of the largest to smallest singular value) and the clustering of the singular values (see e.g. De Roeck, 2002, where these quantities are discussed for a relatively small Kirchhoff-based imaging problem). Because LSQR minimizes the residual during each iteration (cf. Equation 3.3), studying its progress as a function of the number of iterations gives us some way to gauge

3.4. Convergence of least-squares migration

the performance. For this purpose, we introduce the normalized log-based least-squares residual $\mu_k = 20 \log \|\widehat{\mathbf{A}}\mathbf{u}_k - \widehat{\mathbf{b}}\|_2 / \|\widehat{\mathbf{b}}\|_2$, with \mathbf{u}_k the solution of the (preconditioned) system after k iterations and with $\mathbf{u}_0 = 0$. Following De Roeck (2002), we also track the log-based least-squares model-space residual—i.e., $\nu_k = 20 \log \|\widehat{\mathbf{A}}^*(\widehat{\mathbf{A}}\mathbf{u}_k - \widehat{\mathbf{b}})\|_2 / \|\widehat{\mathbf{A}}^*\widehat{\mathbf{b}}\|_2$. Note that in practice, these latter residuals are never computed because they are not a by-product of LSQR, and require an additional matrix-vector multiply per iteration. Contrary to data-space residuals, which possibly contain unmodeled components that may not be in the range of the modeling operator, model-space residuals typically converge to zero. Both quantities are used to empirically establish the performance of the different levels of preconditioning. The results of this exercise are summarized in Figure 3.3, which plots the decay of μ_k and ν_k as a function of the number of iterations k . To account for the overhead, plots for the level III preconditioner are offset by one iteration. This is justified because the cost per iteration of evaluating the curvelet transform, and its inverse, is negligible compared to the cost of migration and demigration.

As we move from a single left preconditioner, towards left and right preconditioners, the data residuals decay faster with a significant improvement obtained by the curvelet-domain scaling. For instance, the residual after 10 iterations for level II preconditioning (fractional integration and depth weighting) is attained by only 5 iterations of level III preconditioning (including curvelet-domain scaling), whereas the result after 10 iterations is approximately 2 dB better. Including the computational overhead, level III preconditioning gains 8 dB in four iterations. Even though the picture for

the model-space residual is less clear, there is a similar trend for the level III preconditioning. For instance, after 10 iterations, we have an improvement of approximately 4.5 dB.

The improvements in convergence for the preconditioned system are also reflected in the least-squares migrated results included in Figure 3.4. Comparing these images shows a clear enhancement for the preconditioned system, plotted in Figure 3.4(b), over the least-squares result obtained without preconditioning. Moreover, juxtaposing the preconditioned least-squares image with the solution for the preconditioned migrated image after one iteration depicted in Figure 3.2(c) shows a significant enhancement of the overall amplitudes and frequency content (cf. Figure 3.1(a) and 3.4(b)).

The bottom line is that each iteration takes 40 and 180 minutes for the modeling and migration on an IBM eServer with 52 processors at 2.2GHz. Aside from one additional modeling and migration, the computation of the diagonal estimation takes approximately 90 minutes on a single CPU. The cost of the 2-D curvelet-transforms part of the preconditioning is less than one minute and is negligible compared to the modeling and migration costs.

3.5 Extensions

The preconditioning methodology presented in this letter constitutes a first step towards a concerted effort to formulate seismic imaging and inversion as an optimization problem. This type of formulation allows us to create high-fidelity images and to make progress towards full-waveform inversion. Having access to appropriate preconditioners is instrumental for this purpose

because it makes the computations tractable. We envisage the following extensions of our work:

- generalization to 3-D, which entails a different power for the fractional integration, a different spherical-spreading correction, and a 3-D formulation of our curvelet-domain matched filter. Moreover, in 3-D the theory of approximating the Hessian by a Ψ DO is less well developed.
- replacement of the level II physical-space preconditioning by more accurate source Green’s function illumination corrections (see e.g., Plessix and Mulder, 2004).
- inclusion of density variations, or in case of elastic wave propagation, the inclusion of variations in the elastic moduli. This extension requires a multi-parameter formulation.
- regularization by curvelet-domain sparsity promotion, replacing Equation 3.3 by

$$\tilde{\mathbf{u}}_{\ell_1} = \arg \min_{\mathbf{u}} \|\mathbf{u}\|_1 \quad \text{subject to} \quad \|\hat{\mathbf{A}}\mathbf{u} - \hat{\mathbf{b}}\|_2 \leq \epsilon \quad (3.11)$$

with ϵ a noise-dependent parameter, and $\mathbf{x}_{\ell_1} := \mathbf{M}_R^{-1}\tilde{\mathbf{u}}_{\ell_1}$. To benefit from our preconditioning, Equation 3.11 requires a solver with Newton-type steps, as opposed to most ℓ_1 -norm solvers that are based on projected gradients. The advantage of this formulation is that it uses curvelet-domain sparsity, which has proven to be a particularly powerful prior (Wang and Sacchi, 2007; Herrmann et al., 2008b,a; Hennenfent et al., 2008). We will report on the solution of Equation 3.11

elsewhere.

- incorporation of our preconditioner in sparsity-promoting full-waveform inversion. This problem requires the solution of the unconstrained optimization problem

$$\min_{\mathbf{z}} \frac{1}{2} \|\widehat{\mathbf{b}} - \widehat{\mathbf{F}}[\mathbf{z}]\|_2^2 + \lambda \|\mathbf{z}\|_1, \quad (3.12)$$

where \mathbf{z} is the curvelet representation of the model, λ a Lagrange multiplier balancing the residual energy and ℓ_1 -norm penalty term, and $\widehat{\mathbf{F}}[\mathbf{z}]$ the nonlinear forward map that links the curvelet-domain model to data. Again, the solution of this optimization problem requires a sophisticated solver that can benefit from our preconditioner.

3.6 Conclusions

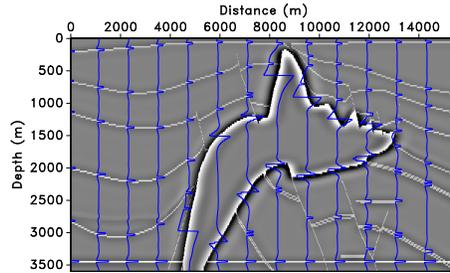
Because of the size of the seismic imaging problem, preconditioning of least-squares migration is an elusive topic, where traditional approaches from numerical linear algebra have not yet found their way. Lack of direct access to the matrices involved and the cost of evaluating matrix-free implementations of the operators are both to blame. Nonetheless, the first few iterations of the LSQR algorithm of least-squares migration are known to make significant progress towards the solution. Unfortunately, even for this limited number of iterations, the computational costs are often still prohibitively large for practical problems. The method presented in this paper partly resolves this issue through a combination of left and right preconditioning

together with a curvelet-domain scaling. Inclusion of the latter proved particularly important because it restores the amplitudes and leads to faster convergence, at a relatively small computational overhead. Aside from this tangible reduction in computational costs of roughly 50%, the use of curvelet frames opens the enticing perspective to use ℓ_1 -norm regularization to improve the quality of images. Preconditioning also plays a pivotal role in making this approach numerically feasible and may extend to a solution of the full-waveform inversion problem with sparsity promotion. Both approaches are justified by ample evidence that curvelets are sparse on the model.

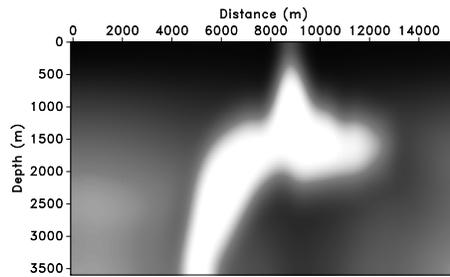
3.7 Acknowledgments

The authors would like to thank C. C. Stolk and W. W. Symes for their input on migration-amplitude recovery and for the use of the reverse-time migration code. We also would like to thank Tamas Nemeth and Michael Friedlander for insightful discussions and the authors of CurveLab (www.curvelet.org). The examples were prepared with Madagascar (rsf.sf.net), supplemented by SLIMpy operator overloading, developed by S. Ross-Ross. This work was in part financially supported by the NSERC Discovery (22R81254) and CRD Grants DNOISE (334810-05) of F.J.H. and was in part carried out within the SINBAD project with support, secured through ITF, from BG Group, BP, Chevron, ExxonMobil and Shell. We finally thank the anonymous reviewer and W. W. Symes for their constructive comments and suggestions that greatly improved our paper.

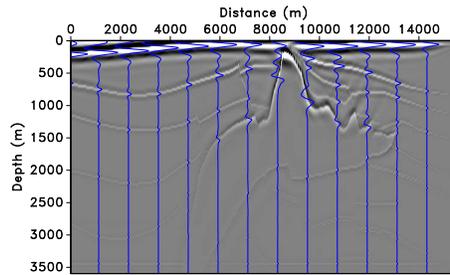
3.7. Acknowledgments



(a)



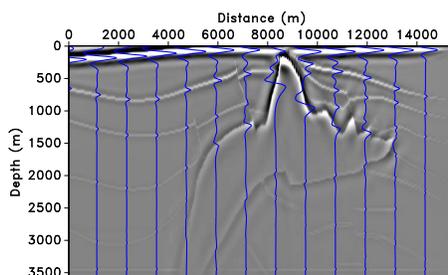
(b)



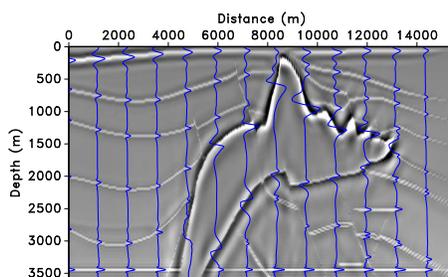
(c)

Figure 3.1: The SEG/EAGE AA' salt model, (a) reflectivity defined by the high-pass filtered velocity model, (b) smoothed velocity model, (c) the migrated image according to Equation 3.1. This image suffers from deteriorated amplitudes, especially under the high-velocity salt and for steep reflectors and faults.

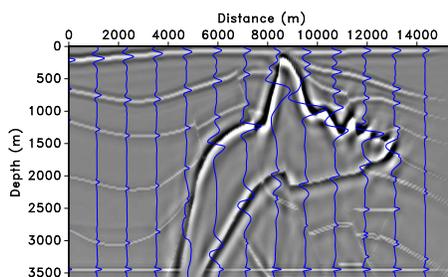
3.7. Acknowledgments



(a)



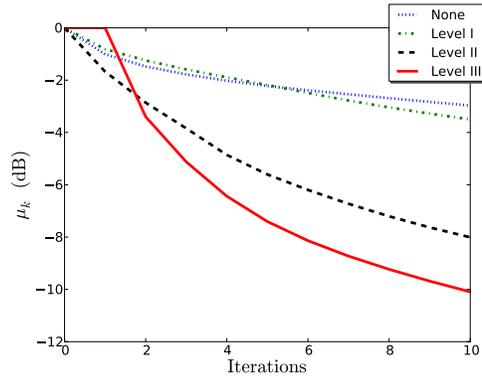
(b)



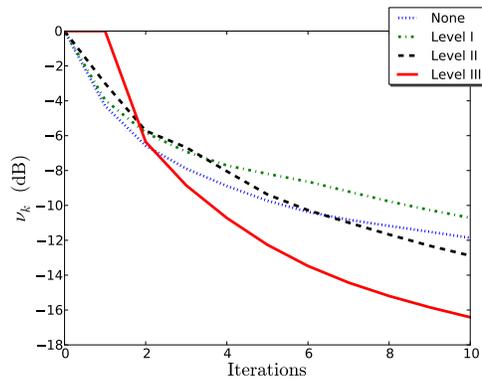
(c)

Figure 3.2: Migrated images for different levels of preconditioning, **(a)** result for left preconditioning (level I, cf. Equation 3.6), **(b)** result for left-right (including depth-correction) preconditioning (level II, cf. Equation 3.7), **(c)** the same but now including curvelet-domain scaling (level III, cf. Equation 3.10).

3.7. Acknowledgments



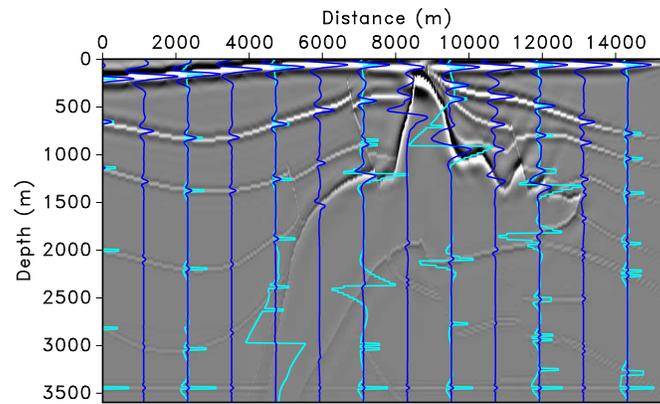
(a)



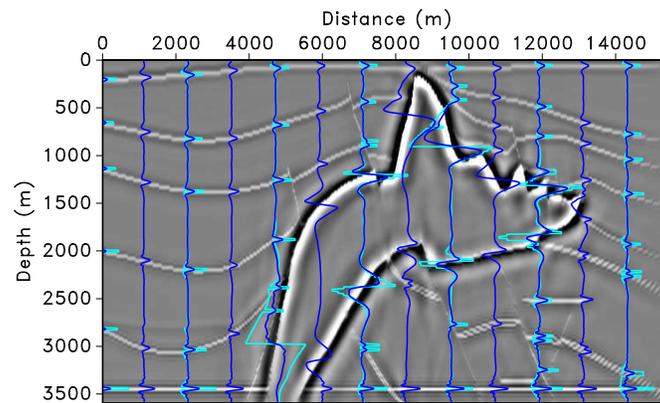
(b)

Figure 3.3: Residual decays for different levels of preconditioning. The dotted blue lines corresponds to least-squares migration without preconditioning, the dash-dotted lines to level I preconditioning, the dashed black lines to level II preconditioning, and the red solid lines to level III preconditioning. This is offset by one iteration to account for the overhead, **(a)** plot for the decay of the data-space normalized residues μ_k as a function of the number of LSQR iterations, **(b)** the same but now for the model-space normalized residuals ν_k .

3.7. Acknowledgments



(a)



(b)

Figure 3.4: Least-squares migration without and with preconditioning juxtaposed with the original reflectivity (in light blue), **(a)** least-squares migrated image, **(b)** least-squares image with level III preconditioning. Notice the improvement in the recovered reflectivity.

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.
- Calvetti, D., 2007, Preconditioned iterative methods for linear discrete ill-posed problems from a Bayesian inversion perspective: *Journal of Computational and Applied Mathematics*, **198**, 378–395.
- Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: *SIAM Multiscale Modeling and Simulation*, **5**, 861–899.
- Chavent, G. and R. E. Plessix, 1999, An optimal true-amplitude least-squares prestack depth-migration operator: *Geophysics*, **64**, 508–515.
- Claerbout, J. and D. Nichols, 1994, Spectral preconditioning: Technical Report SEP-82, Stanford Exploration Project.
- De Roeck, Y., 2002, Sparse linear algebra and geophysical migration: A review of direct and iterative methods: *Numerical Algorithms*, **29**, 283 – 322.
- Demanet, L. and L. Ying, 2007, Wave atoms and sparsity of oscillatory patterns: *Journal of Applied and Computational Harmonic Analysis*, **23**, 368–387.

- , 2008, Discrete symbol calculus. (Submitted for publication. Available at <http://math.stanford.edu/~laurent/papers/DSC.pdf>).
- Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: *Geophysics*, **69**, 1017–1024.
- Hennenfent, G., E. van den Berg, M. P. Friedlander, and F. J. Herrmann, 2008, New insights into one-norm solvers from the Pareto curve: *Geophysics*, **73**, no. 4, A23–A26.
- Herrmann, F. J., P. P. Moghaddam, and C. C. Stolk, 2008a, Sparsity- and continuity-promoting seismic imaging with curvelet frames: *Journal of Applied and Computational Harmonic Analysis*, **24**, 150–173.
- Herrmann, F. J., D. Wang, G. Hennenfent, and P. P. Moghaddam, 2008b, Curvelet-based seismic data processing: a multiscale and nonlinear approach: *Geophysics*, **73**, no. 1, A1–A5.
- Herrmann, F. J., D. Wang, and D. J. E. Verschuur, 2008c, Adaptive curvelet-domain primary-multiple separation: *Geophysics*, **73**, A17–A21.
- Hu, J., G. T. Schuster, and P. A. Valasek, 2001, Poststack migration deconvolution: *Geophysics*, **66**, 939–952.
- Kuhl, H. and M. D. Sacchi, 2003, Least-squares wave-equation migration for AVP/AVA inversion: *Geophysics*, **68**, 262–273.
- Meyer, Y., 1992, *Wavelets and operators*: Cambridge University Press.
- Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: **64**, 208–221.
- O’Brien, M. and S. Gray, 1996, Can we image beneath salt?: *The Leading Edge*, **15**, 17–22.
- Paige, C. C. and M. A. Saunders, 1982, LSQR: An algorithm for sparse

- linear equations and sparse least squares: *ACM TOMS*, **8**, 43–71.
- Plessix, R. and W. Mulder, 2004, Frequency-domain finite difference amplitude-preserving migration: *Geophysical Journal International*, **157**, 975–987.
- Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth migration: *Geophysics*, **68**, 1371–1379.
- Stolk, C. C., 2000, Microlocal analysis of a seismic linearized inverse problem: *Wave Motion*, **32**, 267–290.
- Symes, W. W., 2007, Reverse time migration with optimal checkpointing: *Geophysics*, **72**, SM213–SM221.
- , 2008, Approximate linearized inversion by optimal scaling of prestack depth migration: *Geophysics*, **73**, R23–R35.
- Taylor, M., 1981, *Pseudodifferential operators*: Princeton University Press.
- Vogel, C., 2002, *Computational Methods for Inverse Problems*: SIAM.
- Wang, D., R. Saab, O. Yilmaz, and F. J. Herrmann, 2008, Bayesian wave-field separation by transform-domain sparsity promotion: *Geophysics*, **73**, A33–A38.
- 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.
- Yu, J., J. Hu, G. T. Schuster, and R. Estill, 2006, Prestack migration deconvolution: *Geophysics*, **71**, S53–S62.

Chapter 4

Curvelet-based seismic data processing

4.1 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,

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

2007a), primary-multiple separation (Herrmann et al., 2007) and migration-amplitude recovery (Herrmann et al., 2008). Application of this formalism to wavefield extrapolation is presented elsewhere (Lin and Herrmann, 2007).

4.2 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 anisotropic shape – they obey the so-called parabolic scaling relationship, yielding a "width" proportional to square of the "length" 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., 2008), 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 extension, 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., $\mathbf{C} \mathbf{C}^T \neq \mathbf{I}$.

4.3 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

$$\mathbf{y} = \mathbf{A} \mathbf{x}_0 + \mathbf{n} \tag{4.1}$$

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):

$$\mathbf{P}_\epsilon : \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 & \text{s.t. } \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon \\ \tilde{\mathbf{f}} = \mathbf{S}^T \tilde{\mathbf{x}} \end{cases} \quad (4.2)$$

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

Nonlinear programs \mathbf{P}_ϵ 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}_ϵ 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}_ϵ 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 ($\mathbf{x} \leftarrow \mathbf{x} + \mathbf{A}^T(\mathbf{y} - \mathbf{A}\mathbf{x})$), minimizing the quadratic part of Equation 4.2, is followed by a soft thresholding ($\mathbf{x} \leftarrow T_\lambda(\mathbf{x})$ with $T_\lambda(x) := \text{sgn}(x) \cdot \max(0, |x| - |\lambda|)$) 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.

4.4 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}_ϵ . It is also shown that the recovery performance can be increased when information on the major primary arrivals is included in the modeling operator.

4.4.1 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., $\mathbf{A} := \mathbf{R}\mathbf{C}^T$. With this definition for the modeling operator, solving \mathbf{P}_ϵ 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}_ϵ) 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 4.1 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 4.1(b)). Comparing the 'ground truth' in Figure 4.1(a) with the recovered data in Figure 4.1(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.

4.4.2 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, $\Delta\mathbf{P}$, 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., $\mathbf{A} := \mathbf{R}\Delta\mathbf{P}\mathbf{C}^T$, the recovery can be improved by solving \mathbf{P}_ϵ . The solution of \mathbf{P}_ϵ now entails

4.4. Seismic data recovery

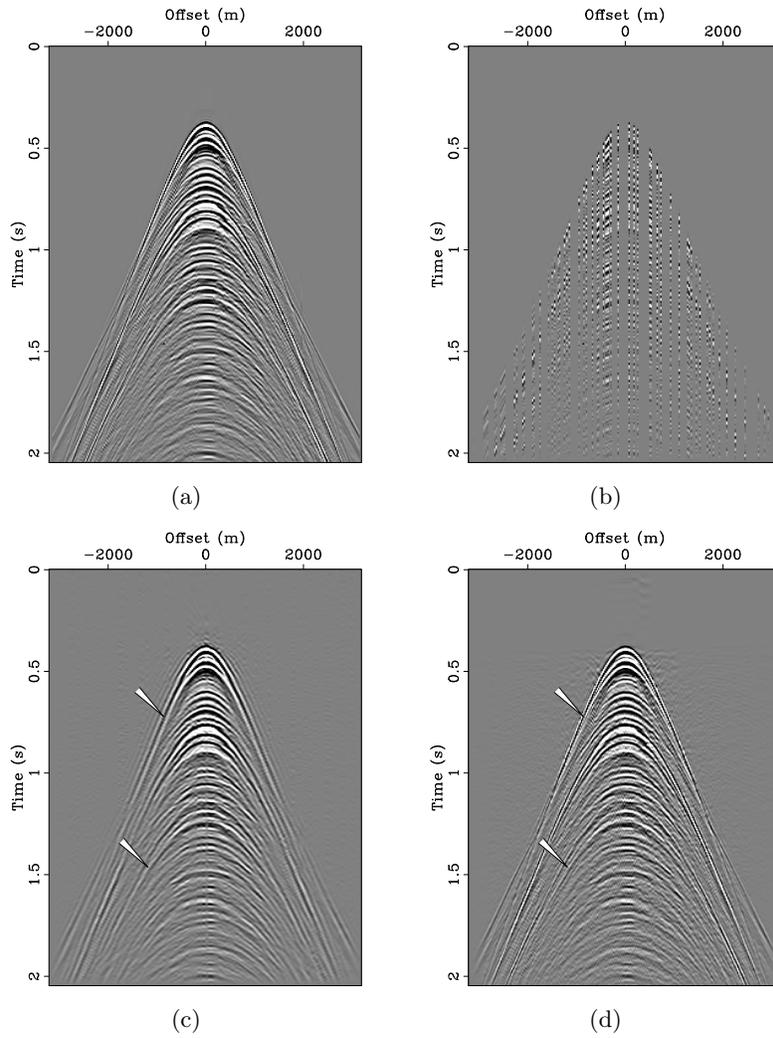


Figure 4.1: Comparison between 3-D curvelet-based recovery by sparsity-promoting inversion with and without focusing, **(a)** fully sampled real SAGA data shot gather, **(b)** randomly subsampled shot gather from a 3-D data volume with 80% of the traces missing in the receiver and shot directions, **(c)** curvelet-based recovery, **(d)** curvelet-based recovery with focusing. Notice the improvement (denoted by the arrows) from the focusing with the primary operator.

the inversion of $\Delta\mathbf{P}$, 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 $\Delta\mathbf{P}$ yields the interpolation, i.e. $\mathbf{S}^T := \Delta\mathbf{P}\mathbf{C}^T$. Comparing the curvelet recovery with the focused curvelet recovery (Figure 4.1(c) and 4.1(d)) shows an overall improvement in the recovered details.

4.5 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., 2007). 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}_ϵ , 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}_ϵ 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}_ϵ . 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}_ϵ for the weighted ℓ_1 norm ($\mathbf{A} = \mathbf{S}^T := \mathbf{C}^T$).

Figure 4.2 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., 2007).

4.5. Seismic signal separation

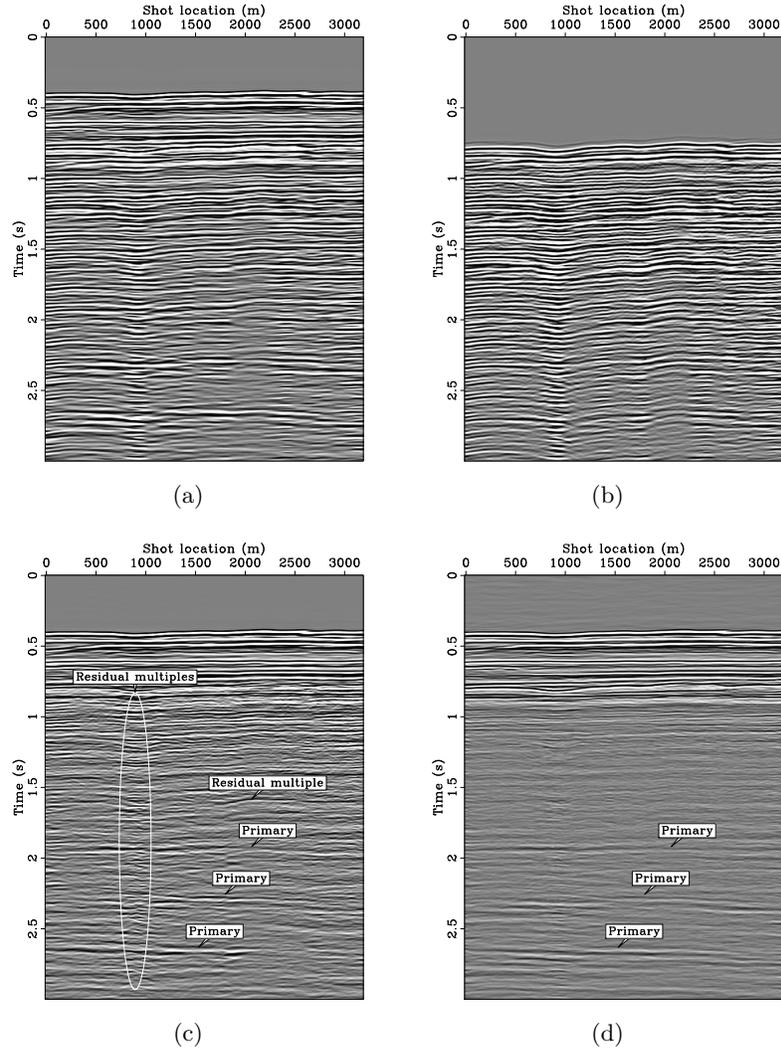


Figure 4.2: 3-D Primary-multiple separation with \mathbf{P}_ϵ for the SAGA dataset, (a) near-offset section including multiples, (b) the SRME-predicted multiples, (c) the estimated primaries according to ℓ_2 -matched filtering, (d) the estimated primaries obtained with \mathbf{P}_ϵ . Notice the improvement, in areas with small 3-D effects (ellipsoid) and residual multiples.

4.6 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.

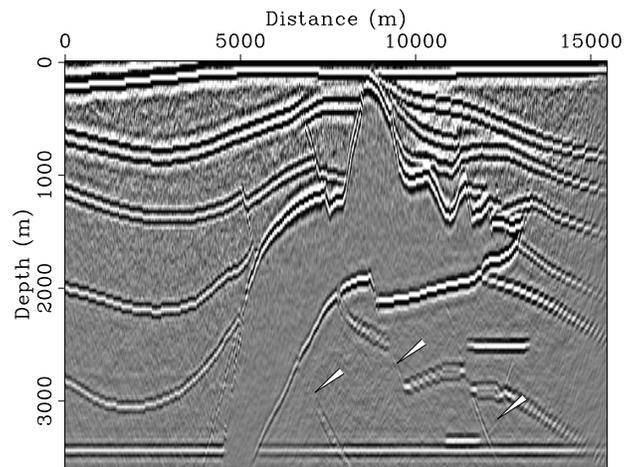
The method is based on the approximate identity: $\mathbf{K}^T \mathbf{K} \mathbf{r} \approx \mathbf{C}^T \mathbf{D}_r \mathbf{C} \mathbf{r}$ with \mathbf{K} and \mathbf{K}^T the demigration, migration operators and \mathbf{D}_r a reference-model specific scaling (Herrmann et al., 2008). By defining the modeling matrix as $\mathbf{A} := \mathbf{C}^T \sqrt{\mathbf{D}_r}$, \mathbf{P}_ϵ 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}_ϵ 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 4.3. 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 4.3(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 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.

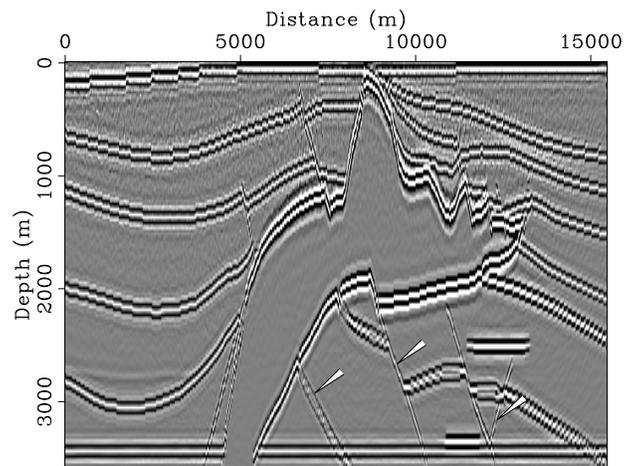
4.7 Discussion and conclusions

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 di-

4.7. Discussion and conclusions



(a)



(b)

Figure 4.3: 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.

rection(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.

4.8 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)

4.8. Acknowledgments

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 Informa-*

tion 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, 2007, 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, 2008, Sparsity- and continuity-promoting seismic imaging with curvelet frames: Applied and Computational Harmonic Analysis.
- 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.
- 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 5

True amplitude depth migration using curvelets

5.1 Introduction

Despite the fact that modern depth migrations can accurately position reflecting surfaces in the Earth subsurface, they often do not correctly recover high-fidelity amplitude information. This is true even when accurate knowledge of the velocity model is available.

Many approaches have been proposed to solve the “true-amplitude” migration problem and they generally fall into three main categories. First, there are approaches that correct for the amplitude “during migration” by applying imaging conditions that take reflector amplitudes into account (see e.g. J. C. Costa and Novais, 2009; Chattopadhyay and McMechan, 2008; J. Schleicher and Novais, 2008; Y. Zhang and Zhang, 2007).

The second category involves linearized inversion where the linearized scattering operator, i.e. the adjoint of migration (see e.g. Guitton, 2004; Claerbout, 1985; Gray, 1997; Symes, 2007), is inverted using iterative Lanc-

A version of this chapter will be submitted for publication. Moghaddam, P.P., Herrmann F. J. and Shahidi, R. (2010) True amplitude depth migration using curvelets.

zos methods (Herrmann et al. (2009)). In this category, a variety of algorithms for true amplitude migration have been introduced which iteratively update the migrated image to minimize the misfit between observed and the simulated data (see e.g. Tarantola, 1987; Nemeth et al., 1999; Kuehl and Sacchi, 2003; Mulder and Plessix, 2004; Herrmann et al., 2009).

Third, there are the so called image-to-image scaling methods, where the normal operator (i.e. migration followed by the scattering operator) is approximated through a transform-domain scaling method. The scaling is done by comparing the migrated image (with some possible pre/post-processings) with a second image obtained by remigration of the image (i.e. after applying the normal operator on the migrated image). Typically, this diagonal scaling is performed in some transform domain (Fourier, curvelets, physical, etc, see e.g. Rickett, 2003; Plessix and Mulder, 2004; Guitton, 2004; Symes, 2008a; Herrmann et al., 2008).

Our method falls in the third category, and we address the amplitude recovery problem by exploiting the recently developed curvelet frames. These frame expansions compress seismic images (see e.g. Hennenfent and Herrmann, 2006; E. J. Candès and Ying, 2006) and consist of a collection of frame elements ‘curvelets’ that are invariant under the normal operator (Herrmann et al. (2008, 2009)). These two properties allow for the development of an approach where the normal operator is nonlinearly inverted, using the eigen-like behavior of curvelets, in combination with the ability of curvelets to handle conflicting dips and sparsely represent the seismic images.

The main contributions of this chapter are three-fold. First, we replace the linear least-squares formulation for the estimation of the curvelet-domain

coefficients (Herrmann et al. (2008)) by a nonlinear least-squares formulation that approximates the symmetric positive definite normal operator by imposing positivity on the scaling coefficients. Second, we apply this method to correct the amplitudes of images that suffer from migration noise caused by small number of shots, as well as noise in the data. Third, we compare two sparsity-promoting methods for recovery, a soft thresholding-correction technique and a large scale sparsity solver.

Our method is tested with reverse-time ‘wave equation’ migration code simulating the acoustic wave equation on different synthetic models.

5.2 Outline of the chapter

In the first section, we derive our formulation and propose two methods for amplitude recovery. Next, we derive the diagonal estimation for the normal operator in the curvelet domain. The chapter ends with an example of each method on the SEG salt model linearized Born dataset and comparison of the proposed methods followed by conclusion.

5.3 Problem Formulation

By assuming our data consist of primaries only (i.e., the surface related and all the internal multiples are assumed to be removed before migration), the data can be expressed as,

$$\mathbf{d} \approx \mathbf{K}\mathbf{m} + \mathbf{n}, \tag{5.1}$$

5.3. Problem Formulation

with \mathbf{d} data, \mathbf{K} the linearized Born scattering operator (also known as the demigration operator), \mathbf{m} the perturbation in the velocity model (i.e., $\frac{\delta \mathbf{v}}{\mathbf{v}^3}$ with \mathbf{v} as velocity, see William W. Symes and Gao (2004); Symes (2008b)) and \mathbf{n} zero-centered Gaussian noise with variance σ_n^2 . Theoretically, migration is the adjoint of the scattering operator. After applying the migration to the data, we obtain the image,

$$\mathbf{y} = \mathbf{K}^T \mathbf{d} \quad (5.2)$$

$$= \mathbf{\Psi} \mathbf{m} + \mathbf{K}^T \mathbf{n} \quad (5.3)$$

where \mathbf{y} is the migrated data, \mathbf{K}^T is the migration operator and the $\mathbf{\Psi}$ is the normal operator defined by $\mathbf{\Psi} = \mathbf{K}^T \mathbf{K}$. In order to find \mathbf{m} in above equation, we need to invert the normal operator ($\mathbf{\Psi}$). To solve the above equation, we follow recent results by Herrmann et al. (2008, 2009) and approximate the normal operator by a curvelet-domain scaling, i.e., we have,

$$\mathbf{\Psi} \mathbf{m} \simeq \mathbf{C}^T \mathbf{W} \mathbf{C} \mathbf{m}, \quad (5.4)$$

with \mathbf{W} a diagonal matrix. This diagonal approximation involves applying the curvelet transform \mathbf{C} , followed by a scaling with the positive diagonal matrix \mathbf{W} and the inverse curvelet transform back to the physical domain by \mathbf{C}^T . [Note that the curvelets are a tight frame (E. J. Candès and Ying, 2006) so the adjoint of curvelet transform (\mathbf{C}^T) is equal to its pseudo-inverse

5.3. Problem Formulation

(i.e., $\mathbf{C}^T \mathbf{C} = \mathbf{I}$]). Inserting this approximation into Equation 5.2 leads to,

$$\mathbf{y} \simeq \mathbf{C}^T \mathbf{W} \mathbf{C} \mathbf{m} + \mathbf{e}, \quad (5.5)$$

with $\mathbf{e} = \mathbf{K}^T \mathbf{n}$ colored noise with covariance,

$$\mathbf{C}_{\mathbf{e}\mathbf{e}} = \mathbf{E}(\mathbf{e}\mathbf{e}^T) = \mathbf{E}(\mathbf{K}^T \mathbf{m} \mathbf{m}^T \mathbf{K}) = \mathbf{K}^T \mathbf{E}(\mathbf{m} \mathbf{m}^T) \mathbf{K} = \sigma_n^2 \mathbf{K}^T \mathbf{K}, \quad (5.6)$$

with $\mathbf{E}(\cdot)$ the statistical expectation and $\mathbf{C}_{\mathbf{m}\mathbf{m}} = \mathbf{E}(\mathbf{m} \mathbf{m}^T) = \sigma_n^2 \mathbf{I}$ the covariance of the white Gaussian noise. To exploit curvelet-domain sparsity on the recovered model, we rewrite Equation 5.5 as,

$$\mathbf{y} = \mathbf{C}^T \mathbf{W} \mathbf{x} + \mathbf{e} \quad (5.7)$$

where $\mathbf{x} = \mathbf{C} \mathbf{m}$. In the context of stable signal recovery (see e.g. Candès et al., 2006; Daubechies et al., 2005), \mathbf{x} can be found by solving the following optimization problem,

$$\mathbf{P}_1 : \quad \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 = \sum_{\mu \in \mathbf{M}} |x_\mu| \text{ subject to } \|\mathbf{C}_{\mathbf{e}\mathbf{e}}^{-\frac{1}{2}} (\mathbf{y} - \mathbf{C}^T \mathbf{W} \mathbf{x})\|_2 \leq v \\ \tilde{\mathbf{m}} = \mathbf{C}^T \tilde{\mathbf{x}}, \end{cases} \quad (5.8)$$

where v is the noise dependent tolerance, $\tilde{\cdot}$ stands for the estimated value and μ is the index set of curvelets. The term $\mathbf{C}_{\mathbf{e}\mathbf{e}}^{-\frac{1}{2}}$ in front of misfit acts as the whitening operator for the data misfit (Tarantola (1987)). During the optimization, the vector \mathbf{x} is found by minimizing the ℓ_1 norm on the curvelet representation for the model and the ℓ_2 norm for data misfit. Equation 5.8

5.3. Problem Formulation

allows us to exploit curvelet-domain sparsity subject to fitting the migrated image within a tolerance by solving the sparsity-promoting program.

Inserting Equation 5.4, the approximation for the normal operator into Equation 5.6 for the noise covariance yields,

$$\mathbf{C}_{ee} = \sigma_n^2 \mathbf{K}^T \mathbf{K} \simeq \sigma_n^2 \mathbf{C}^T \mathbf{W} \mathbf{C}. \quad (5.9)$$

Assuming the curvelet transform is nearly orthonormal, we can approximate the action of $\mathbf{C}_{ee}^{-\frac{1}{2}}$ in Equation 5.8 by,

$$\mathbf{C}_{ee}^{-\frac{1}{2}} \simeq \sigma_n^{-1} \mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \mathbf{C}. \quad (5.10)$$

With this approximation, Equation 5.8 becomes,

$$\mathbf{P} : \begin{cases} \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{b} - \mathbf{C}^T \mathbf{W}^{\frac{1}{2}} \mathbf{x}\|_2 \leq \epsilon \\ \tilde{\mathbf{m}} = \mathbf{C}^T \tilde{\mathbf{x}}, \end{cases} \quad (5.11)$$

with $\mathbf{b} = \mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \mathbf{C} \mathbf{y}$ and $\epsilon = \sigma_n v$. We propose two alternative methods for solving above problem, one is a Basis Pursuit De-Noising problem (BPDN) and other one is a soft-thresholding solution. BPDN proposes a solution to problem \mathbf{P} equivalent to a weighted ℓ_1 problem (by changing the variable $\mathbf{W}^{\frac{1}{2}} \mathbf{x} \mapsto \mathbf{x}$) yielding $\tilde{\mathbf{m}} = \mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \tilde{\mathbf{x}}$,

$$\mathbf{BPDN} : \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{x}\|_{1, \mathbf{W}^{-\frac{1}{2}}} = \sum_i |x_i w_i^{-1/2}| \text{ subject to } \|\mathbf{b} - \mathbf{C}^T \mathbf{x}\|_2 \leq \epsilon \quad (5.12)$$

We solve optimization \mathbf{BPDN} with a projected gradient method (SPGL1),

recently introduced by (van den Berg and Friedlander, 2008). An alternative way can be evaluating \mathbf{P} in unconstrained form. This is equivalent to a quadratic program yielding $\tilde{\mathbf{m}} = \mathbf{C}^T \tilde{\mathbf{x}}$,

$$\mathbf{QP} : \quad \tilde{\mathbf{x}} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{b} - \mathbf{C}^T \mathbf{x}\|_2^2 + \lambda(\epsilon) \|\mathbf{x}\|_{1, \mathbf{W}^{-\frac{1}{2}}} \quad (5.13)$$

where $\lambda(\epsilon)$ is a constant controlled by the noise level in the data. An approximate solution for the optimization \mathbf{QP} can be found with a weighted soft-thresholding method, given by,

$$\tilde{\mathbf{m}} = \mathbf{C}^T \text{sign}(\mathbf{C}\mathbf{b}) \odot \max(0, |(\mathbf{C}\mathbf{b}) \odot \mathbf{w}^{-\frac{1}{2}}| - \lambda(\epsilon)), \quad (5.14)$$

where \odot is the element-wise product, $\mathbf{w} = \text{diag}\mathbf{W}$ is a vectorized form of the diagonal elements for the matrix \mathbf{W} , $\text{sign}(\cdot)$ is the element-wise sign function and $|\cdot|$ the absolute value function. In the examples section, we examine both methods in term of quality of their output.

5.4 Diagonal Approximation of the Normal Operator

In the previous section, we stated that we can diagonally approximate the normal operator in the curvelet domain. This approximation (c.f., Equation 5.4), replaces the construction of the normal operator, which is computationally expensive, with a curvelet domain diagonal scaling, which is computationally cheap. In this section, we provide the theoretical underpinning of such approximation by introducing the curvelets and their invariance

bound under the action of normal operator. We then propose a method to approximate the normal operator with a diagonal scaling in curvelet domain.

5.5 Curvelets and their invariance under normal operator

A 2D curvelet ϕ_μ is defined by its index $\mu = (j, k, \theta)$ with $j = 0, 1, 2, \dots$ consisting of its scale subindex, $\theta = 0, 1, \dots, 2^{\lfloor j/2 \rfloor} - 1$, its orientation subindex ($\lfloor x \rfloor$ is the lower integer part of x), and $k = (k_x, k_z)$ the location in the wavenumber domain subindex which are scale and angle dependent. Figure 5.1(a) shows an example of different curvelets with different scales and orientations and Figure 5.1(b) shows the corresponding Fourier spectrum.

An important feature of curvelets is that the action of the normal operator (i.e., $\Psi = \mathbf{K}^T \mathbf{K}$) on a curvelet ϕ_μ corresponds to its multiplication by a positive scalar. In Herrmann et al. (2008), we proved a theorem bounding the error with the normal operator diagonal approximation in curvelet domain, i.e., we showed,

$$\|\Psi \phi_\mu - \mathbf{C}^T \mathbf{W} \mathbf{C} \phi_\mu\|_2 \leq \beta 2^{-j/2}, \quad (5.15)$$

with \mathbf{W} a positive diagonal matrix as defined before and β a constant. According to this theorem, the approximation error decays with scale (i.e., j), which is consistent with the high-frequency asymptotic behavior that underlies the diagonal approximation. Shahidi and Herrmann (2009a) shows that the constant β is of moderate size.

5.5. Curvelets and their invariance under normal operator

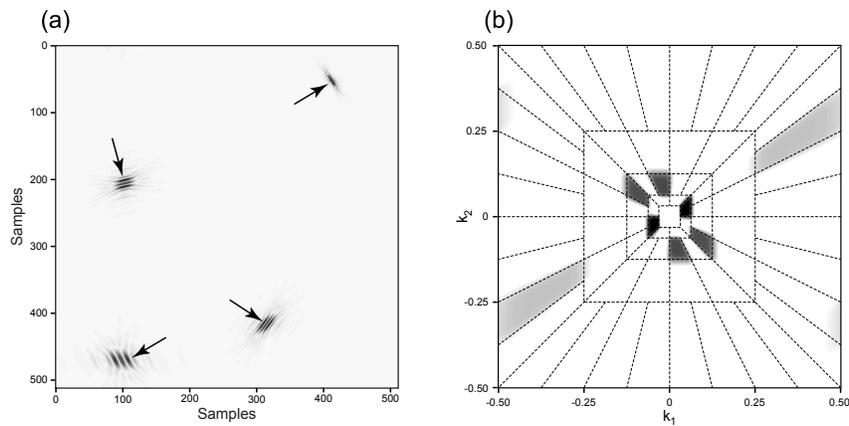


Figure 5.1: Spatial and frequency representation of curvelets, **(a)** four different curvelets in the spatial domain at three different scales, **(b)** dyadic partitioning in the frequency domain, where each wedge corresponds to the frequency support of a curvelet in the spatial domain. This figure illustrates the microlocal correspondence between curvelets in the physical and Fourier domain. Curvelets are characterized by rapid decay in the physical space and of compact support in the Fourier space. Notice the correspondence between the orientations of curvelets in the two domains. The 90° rotation is a property of the Fourier transform. Courtesy of Herrmann et al. (2008).

The normal operator is pseudodifferential when no multipathing occurs (Beylkin, 1985; Rakesh, 1988; Ten-Kroode et al., 1998), and its order in space dimension n is $n - 1$ (Stolk, 2000). The inequality in Equation 5.15 holds when the order of normal operator is zero (Herrmann et al., 2008). Ten-Kroode et al. (1998); Stolk (2000) show that in $2D$ applying an operator of order $-1/2$ to both migration and modeling operators makes the operator of zero order. As it is shown in Herrmann et al. (2009), this is equivalent to fractional integration of the source wavelet in both the modeling and migration operators. Mathematically, this can be written as $\partial_t^{-1/2} \cdot = \mathbf{F}^{-*} |\omega|^{-1/2}$. with \mathbf{F}^{-*} the inverse Fourier transform.

Equation 5.15 states that curvelets are similar to eigenvectors of the normal operator. Therefore, we propose a quasi-eigenvalue decomposition for the normal operator with curvelets as quasi-eigenvectors. This type of approach is similar to the Wavelet-Vaguellette Decomposition (WVD) proposed by (Donoho, 1995; Herrmann et al., 2008).

5.6 Diagonal approximation

In Herrmann et al. (2008), we derive an approximation of the normal operator by remigrating the migrated image again (i.e., $\mathbf{m}_{\text{remig}} = \mathbf{K}^T \mathbf{K} \mathbf{m}_{\text{mig}}$). Subsequently, find a diagonal approximation for the normal operator using these image pairs. Mathematically,

$$\mathbf{m}_{\text{remig}} = (\mathbf{K}^T \mathbf{K}) \mathbf{m}_{\text{mig}} \simeq \mathbf{C}^T \mathbf{W} \mathbf{C} \mathbf{m}_{\text{mig}} \quad (5.16)$$

$$\simeq \mathbf{C}^T \text{diag}(\mathbf{v}) \mathbf{w}, \quad (5.17)$$

with $\mathbf{v} = \mathbf{C}\mathbf{m}_{\text{mig}}$ and \mathbf{w} representing the diagonal elements of \mathbf{W} (i.e., $\mathbf{W} = \text{diag}(\mathbf{w})$).

Equation 5.16 aims to find a curvelet-domain diagonal scaling vector (i.e., \mathbf{w} in Equation 5.16). Because the curvelet transform is redundant (i.e., it has more rows than columns), inverting for \mathbf{w} in Equation 5.16 needs solving an underdetermined system of equations yielding non-unique solutions. To find a unique solution to this estimation problem, we exploit additional properties of the normal operator. First, the normal operator is positive definite, yielding positive entries for the scaling vector \mathbf{w} . Second, the normal operator is elliptic Pseudo-differential (Stolk and De-Hoop, 2006; Shahidi and Herrmann, 2009a) and is invertible when the symbols of this operator are smooth in both physical and Fourier space, when the background velocity model is smooth (Stolk and De-Hoop, 2006; Herrmann et al., 2008). We use these two properties to formulate the diagonal estimation in terms of a nonlinear least-squares problem, which yields positivity of the entries of the diagonal and smoothness amongst neighboring curvelet coefficients in space and angle domain. Compared to our earlier work (Herrmann et al., 2008), the positivity assumption is new in the context of imaging (see related work on application of this method for multiple removal, Herrmann et al. (2007)) and the diagonal estimated by following,

$$\mathbf{P}_d : \begin{cases} \tilde{\mathbf{z}} = \arg \min_{\mathbf{z}} J(\mathbf{z}) = \frac{1}{2} \|\mathbf{m}_{\text{mig}} - \mathbf{C}^T \text{diag}(\mathbf{v}) e^{\mathbf{z}}\|_2 + \kappa \|\mathbf{Lz}\|_2, \\ \tilde{\mathbf{w}} = (\text{diag})(\tilde{\mathbf{w}}) \text{ with } \tilde{\mathbf{w}} = e^{\tilde{\mathbf{z}}}, \end{cases} \quad (5.18)$$

where \mathbf{w} is replaced by $e^{\mathbf{z}}$ to ensure positivity of the estimated diagonal

elements (\mathbf{W}). In above equation, phase-space smoothness is promoted by minimizing ℓ_2 norm of the coefficients after applying the sharpening operator, $\mathbf{L} = [\mathbf{D}_x \ \mathbf{D}_z \ \mathbf{D}_\theta]$. This operator penalizes fluctuations amongst neighboring coefficients in $e^{\mathbf{z}}$, i.e., coefficients that are close in phase space both in position and angle (for more details on this topic see Shahidi and Herrmann (2009b)). The matrices $\mathbf{D}_{x,z}$ contain first-order differences at each scale in the x, z directions, and \mathbf{D}_θ the first-order difference in the θ direction. Because the curvelet grid differs for each scale and angle, this sharpening operator is implemented for each curvelet wedge separately. κ is a smoothness parameter that control the smoothness of the scaling versus data misfit. In this paper, we select this parameter by exhaustive search for automatic parameter selection (Shahidi and Herrmann (2009a)).

Equation 5.18 is convex and can be minimized by an efficient numerical optimization method. We use Limited-memory BFGS (see e.g. Nocedal and Wright, 1999), using following gradient,

$$\nabla_{\mathbf{z}} J(\mathbf{z}) = e^{\mathbf{z}} \odot \text{diag}(\mathbf{v}) \mathbf{C} (\mathbf{C}^T \text{diag}(\mathbf{v}) e^{\mathbf{z}} - \mathbf{m}_{\text{mig}}) + 2\kappa \mathbf{L}^T \mathbf{L} \mathbf{z}. \quad (5.19)$$

5.7 Practical Workflow

Our algorithm consists of two main steps, namely the calculation of the curvelet scaling coefficients via an image-to-remigrated image matched filtering and the subsequent estimation of amplitude-corrected, denoised image by sparsity promotion.

In order to apply the curvelet-domain match filter, we require a curvelet implementation (E. Candes (2005)). Our algorithm also requires access to the Born modeling operator and its adjoint, the migration operator. For this purpose, we use Symes (2007) reverse-time migration with optimal check-pointing and its linearized modeling.

We make the normal operator zero order by applying a zero-phase half-time integration to the source wavelet input to both operators (Ten-Kroode et al., 1998; Symes, 2008a). Our algorithm consists of the following steps:

1. Perform migration and depth correction—i.e., $\mathbf{d} \mapsto \mathbf{D}_z \mathbf{K}^T \mathbf{d} = \mathbf{m}_{\text{mig}}$, with \mathbf{m}_{mig} depth corrected migrated image and $\mathbf{D}_z = \text{diag}(z)$ where $z_i = i\Delta z$, for $i = 1 : n_z$, Δz is the depth grid spacing and n_z is the number of samples in depth (Herrmann et al. (2009)). The reason for depth correction is to amplify the amplitudes of deep events in the migrated image. This ensures that after applying the normal operator, the amplitude of the deep events are partially calibrated so it can be used to extract information from the normal operator via estimation of its diagonal in curvelet domain.
2. Remigrate and resimulate the migrated image—i.e., $\mathbf{K}^T \mathbf{K} \tilde{\mathbf{m}}_{\text{mig}} = \mathbf{m}_{\text{remig}}$.
3. Find the curvelet domain diagonal scaling \mathbf{W} for which $\tilde{\mathbf{m}}_{\text{mig}} \simeq \mathbf{C}^T \mathbf{W} \mathbf{C} \mathbf{m}_{\text{remig}}$ by solving the optimization problem \mathbf{P}_d (cf., Equation 5.18).
4. Construct the original image by solving the minimization problem \mathbf{QP} with weighted soft-thresholding (Equation 5.14) or the optimization problem \mathbf{BPDN} using a solver for large-scale sparse reconstruction.

5.8 Examples

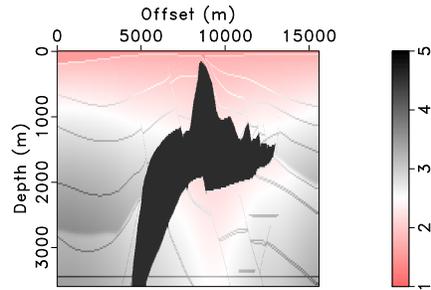
In this section, we apply our amplitude recovery methods on the (noisy) linearized Born dataset with noise using SEG-AA salt model (Aminzadeh et al. (1997); O’Brien and Gray (1996)). The SEGAA model shown in Figure 5.2(a) is 5.28km in depth and 16.8km in offset with grid resolution of 12m. The 5000 time-step linearized Born modeling takes about 24 min with 25 CPUs on a MPI cluster, while the migration takes about 90 min. In this section, we compare the soft-thresholding and the BPDN methods described in this paper in terms of output quality and calculation speed for both no noise in data (migration noise only) and when noise is added to data. We also analyze the accuracy of the estimation by comparing the recovered image traces with the reflectivity.

5.8.1 Noise-free case

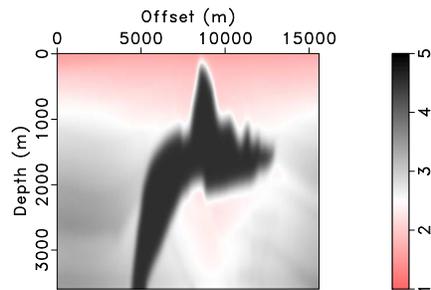
Since the removal of the migration noise is our primary interest, we conduct an imaging experiment with a small number of shots. This dimensionality reduction results in cheaper acquisition and faster computation but goes at the expense of introducing the migration noise. This experiment serves as an illustration on how well our method performs in both the removal of migration noise and restoration of the amplitudes.

We use a land-acquisition scenario with 50 shots between 50m and 16.5km at a 336m spacing and 680 receivers per shot at a 24m interval with offsets between 200m and 16.5km. To generate the linearized Born dataset, we smooth this velocity model as shown in Figure 5.2(b) and use the difference

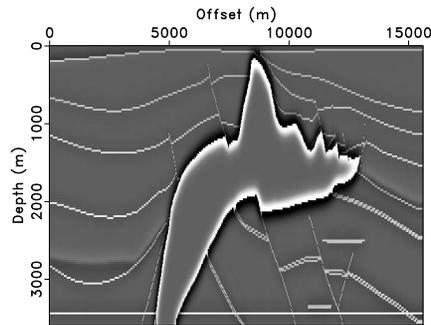
5.8. Examples



(a)



(b)



(c)

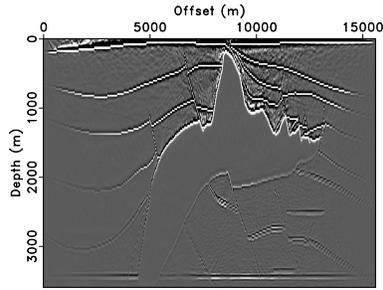
Figure 5.2: Conflicting dips velocity model and reflectivity **(a)** velocity model, **(b)** smooth velocity model used to generate our example, **(c)** reflectivity generated by subtracting smooth velocity model from original one

between the smooth and the original velocity as the reflectivity shown in Figure 5.2(c).

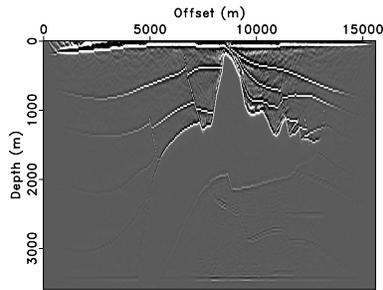
Figure 5.3 summarizes our approximation of the normal operator with the curvelet-domain scaling. We chose the reflectivity in Figure 5.2(c) and applied the scattering operator to generate linearized Born dataset (i.e., $\mathbf{d} = \mathbf{K}\mathbf{m}$). We used the smoothed velocity model shown in Figure 5.2(b) as background velocity for the scattering operator. The depth corrected migrated result is included in Figure 5.3(a). After another modeling and migration of this image, remigrated image is shown in Figure 5.3(b). These two images serve as input to our curvelet scaling coefficient estimation. Applying this curvelet-domain scaling to the reference image yields a good approximation to the action of normal operator as can be seen in Figure 5.3(c). The relative approximation error in this case is 3.71%.

To illustrate the importance of imposing curvelet domain smoothness during the estimation of diagonal scaling, we compare the results of two experiments, one without $\kappa = 0$ in Equation 5.19, and one with a smoothing regularization ($\kappa = 0.01$ found empirically and used to generate the example in Figure 5.3). Figure 5.4 shows the plot of coefficients for each wedge of the reference vector (Figure 5.4(a)), the diagonal estimation without smoothness (Figure 5.4(b)) and with the smoothness (Figure 5.4(c)) constraints. These figures show curvelet decomposition of a shot gather at different frequency band (scale) and angles (dip). Five scales are used for the decomposition. The centre (coarsest scale) shows the DC and low frequency components of the shot gather. The second coarsest scale has 8 angles. The number of angles is doubled to 16 at the third and fourth scale and doubled again to

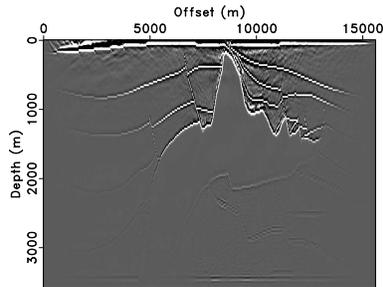
5.8. Examples



(a)



(b)



(c)

Figure 5.3: Diagonal approximation of normal operator, **(a)** reference image (depth corrected migrated image), **(b)** normal operator on the reference image (i.e., $\Psi \mathbf{r}$) **(c)** approximate normal operator on the reference image (i.e., $\mathbf{C}^T \mathbf{W} \mathbf{C} \mathbf{r}$), approximation error is %3.71

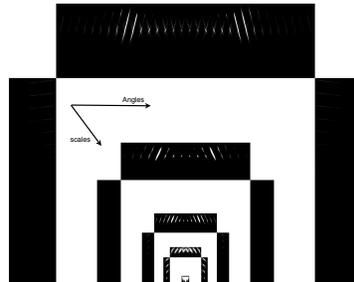
32 for the fifth scale. Note the portions of shot gather captured at various angles and scales. As one can see, the curvelet coefficients for the non zero smoothing regularization $\kappa = .01$ (Figure 5.4(c)) is smooth over different angles and locations.

Soft-thresholding solution

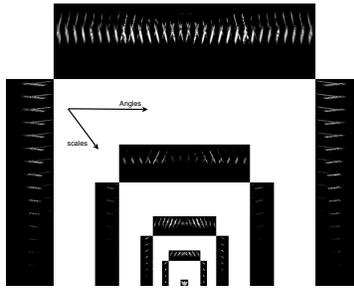
We investigate the effect of the threshold λ on the soft-thresholding solution for the optimization problem **QP**. Figure 5.5 shows the result of solving the optimization problem **QP** with soft thresholding for different values of the threshold λ . The solution of soft thresholding shown in Figure 5.5(a) is for when $\lambda = 0$. This solution is similar to approximately inverting the normal operator in the curvelet domain and applying it to the migrated image (i.e., in Equation 5.11 $\tilde{\mathbf{m}} = \mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \mathbf{C} \mathbf{b}$). Figure 5.5(b) shows the result for $\lambda = \sigma_{\mathbf{b}}$ and Figure 5.5(c) for $\lambda = 2\sigma_{\mathbf{b}}$. We clearly see from the figures that by increasing λ the incoherent noise removed from the image at the expense of dimming the amplitude of the bottom reflectors.

BPDN solution

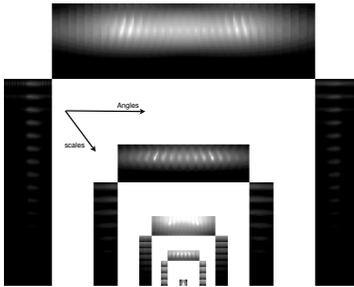
Here we investigate the effect of the tolerance ϵ on the solution for the optimization problem **BPDN** for noise-free data. Figure 5.6 show the results for solving the optimization problem **BPDN** using SPGL1 solver (Berg and Friedlander, 2007; van den Berg and Friedlander, 2008) for different value of the tolerance ϵ . The solution of SPGL1 shown in Figure 5.6(a) is for when ϵ is small ($\epsilon = \sigma_{\mathbf{b}}$ with $\sigma_{\mathbf{b}}$ is the standard deviation of \mathbf{b} in **BPDN**). This solution is similar to Figure 5.5(a) and corresponds to applying the



(a)



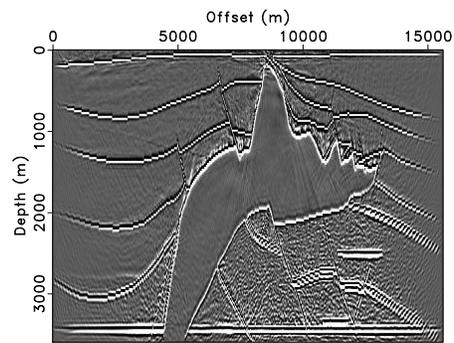
(b)



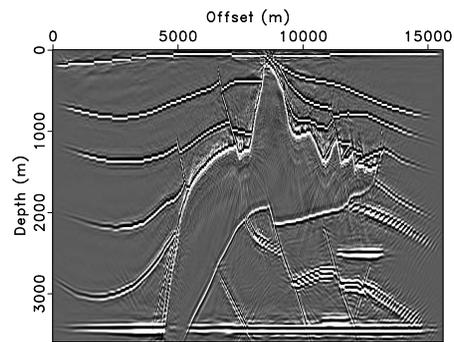
(c)

Figure 5.4: Curvelet-domain representation of the curvelet vector obtained from (a) reference vector, (b) scaling coefficients without smoothing constraints, (c), scaling coefficients with smoothing constraint. The different sub-images represent the curvelet coefficients at different scales (coarsest in the center) and different angles.

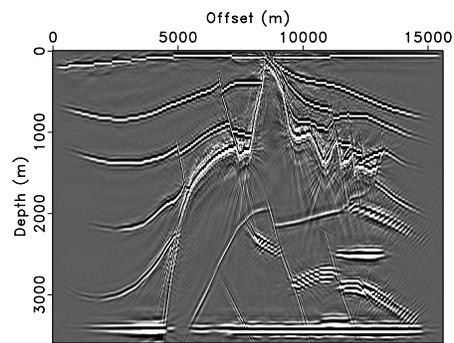
5.8. Examples



(a)



(b)



(c)

Figure 5.5: Examples of the solution for optimization problem **QP** with soft-thresholding, (a) $\lambda = 0$, (b) $\lambda = \sigma_b$ (c) $\lambda = 2\sigma_b$

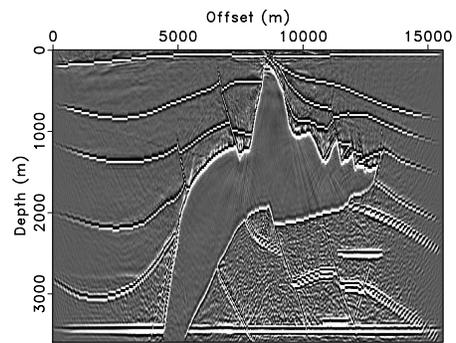
inverse of the diagonal approximation to \mathbf{b} . Figure 5.6(b) shows the result for larger ϵ ($\epsilon = 50\sigma_{\mathbf{b}}$) and Figure 5.6(c) is the solution when the ϵ is very high ($\epsilon = 100\sigma_{\mathbf{b}}$). As can be seen from the figures, when ϵ is small, the image contains many artifacts that increases with depth and when the ϵ is large, these artifacts removed at the expense of dimming the bottom reflectors. The best choice of ϵ is shown in Figure 5.6(b) which is the result of testing a range of different value in the solver. In this image, the deep artifacts are visibly removed and the bottom reflectors are enhanced.

Amplitude Analysis

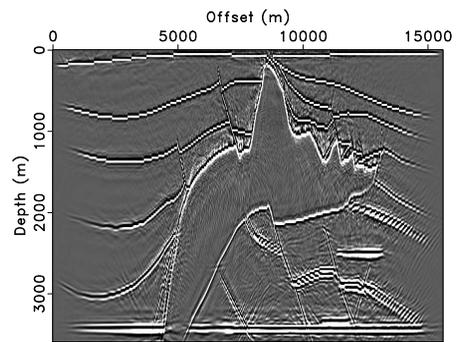
To investigate the performance and accuracy of our amplitude recovery method, we examine a number of traces in the recovered images and compare them to the original reflectivity. Figure 5.7 shows the location of the traces. To examine the performance of our method in the presence of a salt body, we choose a number of traces inside and outside the salt body. The offset location of traces are (3480, 5160, 5880, 8640, 10080, and 12720) m, respectively. Figure 5.8 shows the amplitude of the traces along the offset lines (3480, 5160, and 5880) m, respectively. Except small amount of residual noise in the recovered images, all the strong phases well recovered. Early arrivals are generally well recovered; however, with increasing depth the amount of noise increases.

Figure 5.9 shows the amplitude of the traces along the offset lines of (8640, 10080, and 12720) m, respectively. Note that these traces pass through the salt body. Similar to Figure 5.8, all important phases in the reflectivity are well matched in the recovered image, except the phases that

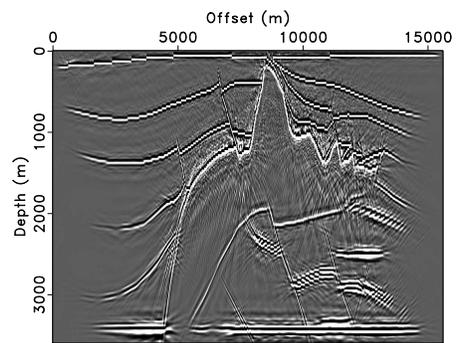
5.8. Examples



(a)



(b)



(c)

Figure 5.6: Examples of the solution for optimization problem **BPDN** with SPGL1, (a) $\epsilon = \sigma_b$, (b) $\epsilon = 50\sigma_b$ (c) $\epsilon = 100\sigma_b$

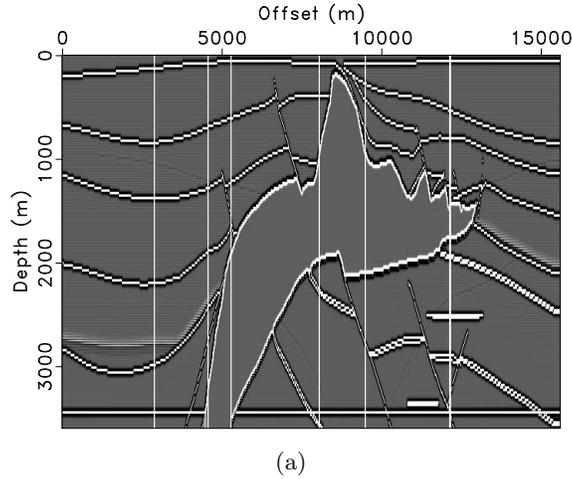
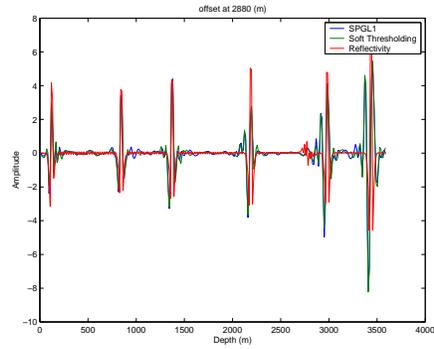


Figure 5.7: Reflectivity model, the vertical lines are locations of investigation come from the bottom salt. Salt bottom imaging is one of the most challenging fields in the seismic imaging. One reason might be the difference in the nature of the salt bottom reflector. This means the amplitude of the reflector should be interpreted as change in the acoustic impedance in that location rather than the simply perturbation in the velocity model.

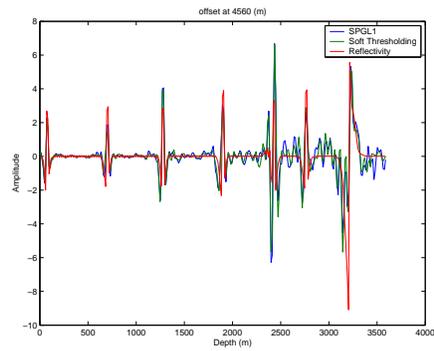
5.8.2 Noisy data

We examine the effectiveness of our recovery method for noisy data. For this purpose, we add white Gaussian noise to the linearized Born data with $SNR = 10 \text{ dB}$ and re-run our method on the data. Figure 5.10 shows data before and after adding the noise. Figure 5.11 shows the depth corrected migrated image for this dataset. Note that the noise is obscuring the deep reflectors in the migrated image.

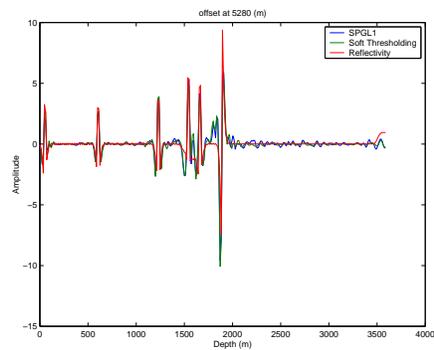
5.8. Examples



(a)



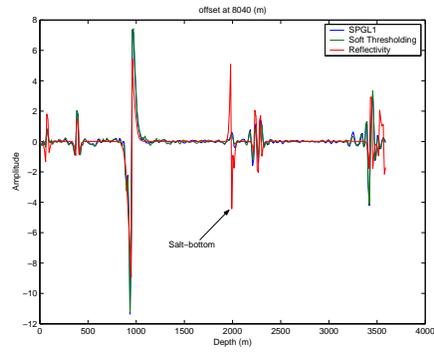
(b)



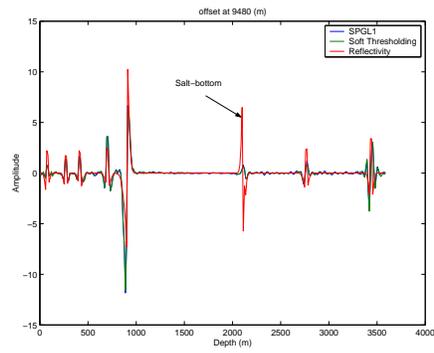
(c)

Figure 5.8: Amplitude analysis of the recovered images using proposed methods, (a) offset=3480 (m) , (b) offset=5160 (m), (c) offset=5880 (m)

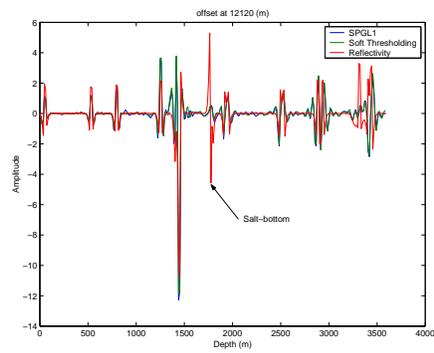
5.8. Examples



(a)



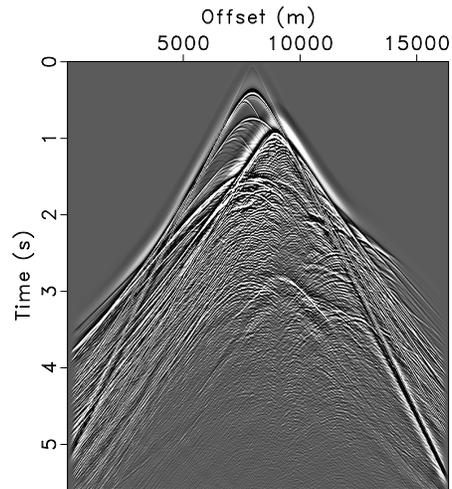
(b)



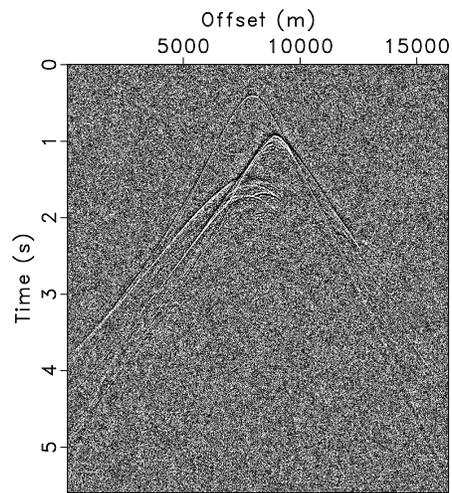
(c)

Figure 5.9: Amplitude analysis of the recovered images using proposed methods, (a) offset= 8640 (m), (b) offset=10080 (m) , (c) offset=12720 (m), note the salt bottom amplitude mismatch

5.8. Examples



(a)



(b)

Figure 5.10: Synthetic data, (a) before , (b) after adding the noise, $SNR = 10dB$

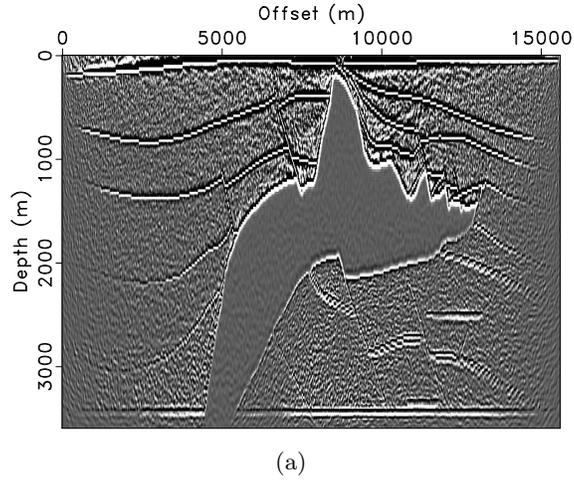


Figure 5.11: Depth corrected migrated noisy image

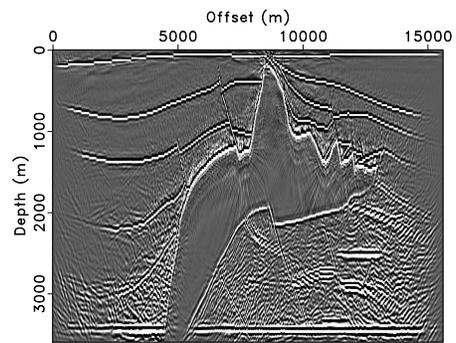
Soft-thresholding solution

We investigate the effect of the threshold λ on the soft-thresholding solution for the optimization problem **QP**. Figure 5.12 shows the result of solving the optimization problem **QP** with soft thresholding for different values of the threshold λ . The solution of soft thresholding shown in Figure 5.12(a) is for when $\lambda = \sigma_{\mathbf{b}}$. Figure 5.12(b) shows the result for $\lambda = 2\sigma_{\mathbf{b}}$ and Figure 5.12(c) is for when $\lambda = 3\sigma_{\mathbf{b}}$. As can be seen from the figures, by increasing λ the incoherent noise is visibly removed from the image at the expense of dimming the amplitude of the bottom reflectors. The best choice of λ is in Figure 5.12(b).

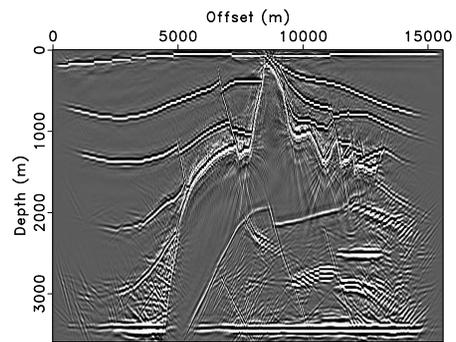
BPDN solution

Figure 5.13 shows the result of solving the optimization problem **BPDN** using SPGL1 solver for different value of the tolerance ϵ . The solution of BPDN

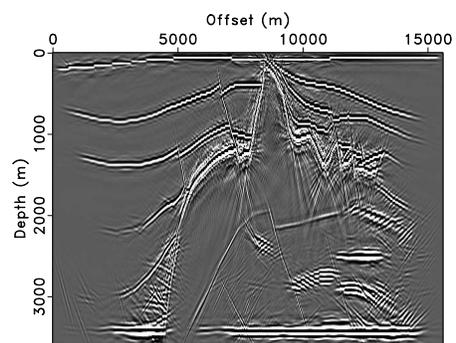
5.8. Examples



(a)



(b)



(c)

Figure 5.12: Examples of the solution for optimization problem **QP** with soft-thresholding, (a) $\lambda = \sigma_b$, (b) $\lambda = 2\sigma_b$ (c) $\lambda = 3\sigma_b$

is shown in Figure 5.13(a) for small ϵ ($\epsilon = 10\sigma_{\mathbf{b}}$ with $\sigma_{\mathbf{b}}$ is the standard deviation of \mathbf{b} in **BPDN**). This solution is similar to approximately inverting the approximation of normal operator in curvelet domain and applying it to migrated image (i.e., in Equation 5.11 $\tilde{\mathbf{m}} = \mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \mathbf{C} \mathbf{b}$). Figure 5.13(b) shows the result for when ϵ is larger ($\epsilon = 140\sigma_{\mathbf{b}}$) and Figure 5.13(c) is the solution when the ϵ is high ($\epsilon = 200\sigma_{\mathbf{b}}$). We can clearly see from the figures that when ϵ is small, the image contains high amount of noise and when the ϵ is large, the noise removed with expense of dimming the bottom reflectors. The best choice of ϵ in Figure 5.13(b) is found through exhaustive process of running the solver on a wide range of ϵ and compare the results.

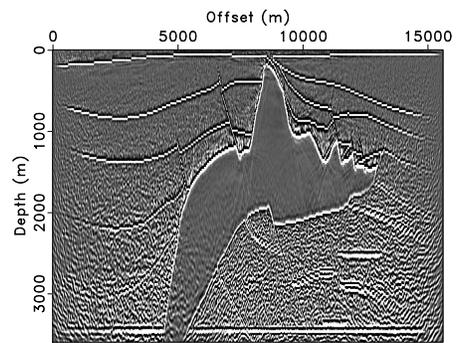
5.9 Tolerance analysis

In the last section, we did not specifically analyze the value of the tolerance ϵ in the optimization problem **BPDN** rather we empirically ran the problem with different values of ϵ and compared the results. A candidate value for the tolerance can be the statistical expectation of the misfit function in **BPDN** in Equation 5.8,

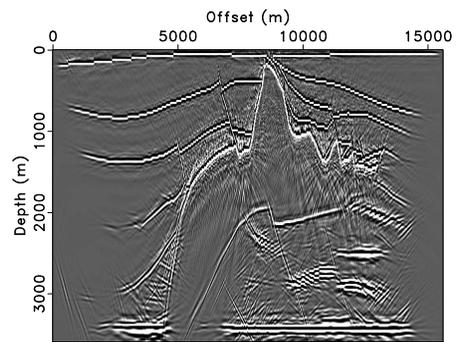
$$\tilde{v} = E[\|\mathbf{C}_{ee}^{-\frac{1}{2}}(\mathbf{y} - \mathbf{C}^T \mathbf{W} \mathbf{x})\|_2] \quad (5.20)$$

with some manipulation $\tilde{v} = N$ with N is the number of elements in \mathbf{y} in lexicographical order. Indeed Figure 5.13(b) is obtained for $v = N$ or $\epsilon = N\sigma_n$.

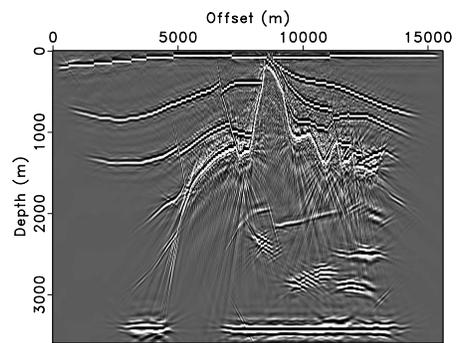
5.9. Tolerance analysis



(a)



(b)



(c)

Figure 5.13: Examples of the solution for optimization problem **BPDN** with SPGL1, (a) $\epsilon = 10\sigma_b$, (b) $\epsilon = 140\sigma_b$ (c) $\epsilon = 200\sigma_b$

5.10 Verification of approximation

Throughout our formulation, we made a subtle approximation from Equation 5.8 to Equation 5.11,

$$\mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \mathbf{C} \mathbf{C}^T \mathbf{W} \mathbf{x} \simeq \mathbf{C}^T \mathbf{W}^{\frac{1}{2}} \mathbf{x}. \quad (5.21)$$

In this approximation, we assume the projection $\mathbf{C} \mathbf{C}^T$ is approximately equal to identity. Here we investigate the accuracy of this approximation.

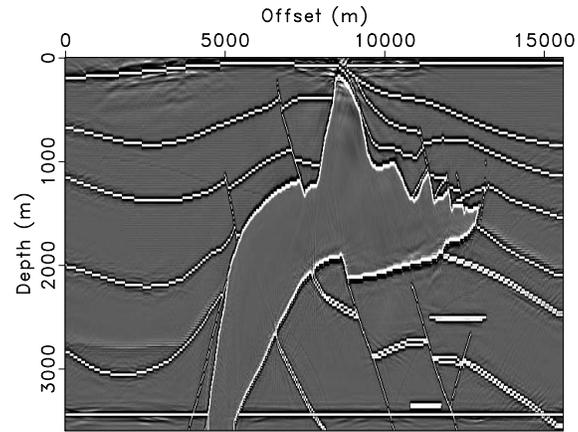
We assume $\mathbf{x} = \mathbf{C} \mathbf{m}$ with \mathbf{m} is the true reflectivity. We call the left hand of above expression as $E_l = \mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \mathbf{C} \mathbf{C}^T \mathbf{W} \mathbf{x}$ and the right hand as $E_r = \mathbf{C}^T \mathbf{W}^{\frac{1}{2}} \mathbf{x}$.

Figure 5.14 show the comparison between the terms in Equation 5.21. Figure 5.14(a) is the left hand E_l expression and Figure 5.14(b) is the right hand E_r expression. The relative error between two images is 4.09 percent.

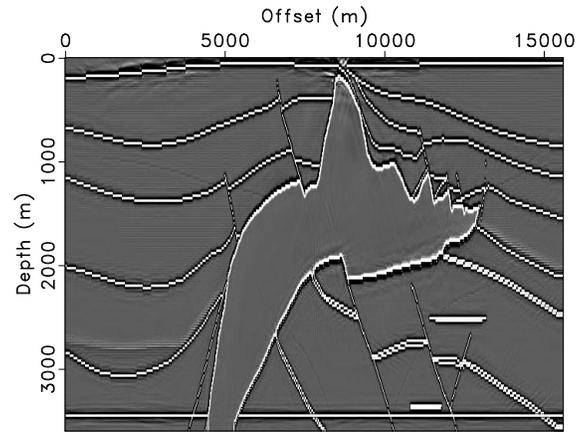
5.11 Comparison

Comparing the output quality and the performance, we suggest using the soft-thresholding technique for amplitude recovery. Soft thresholding in most cases results in better or similar quality of the output with less computational burden comparing to BPDN.

5.11. Comparison



(a)



(b)

Figure 5.14: Comparison between the approximation terms in Equation 5.21, (a) left hand term $E_l = \mathbf{C}^T \mathbf{W}^{-\frac{1}{2}} \mathbf{C} \mathbf{C}^T \mathbf{W} \mathbf{x}$, (b) right hand term $E_r = \mathbf{C}^T \mathbf{W}^{\frac{1}{2}} \mathbf{x}$, relative error is 4.09 percent.

5.12 Conclusion

The method presented in this chapter, is a revisit of the earlier work (Herrmann et al. (2008)), by taking into account the migration noise and the data noise. The method relies on the curvelet transform in both approximation and inversion of the normal operator and has following features,

- speeding up the calculation of the normal operator by diagonalizing it in curvelet domain.
- efficiently using the compression of seismic images by curvelets to solve a curvelet-sparsity optimization problem.
- bringing the amplitude correction problem within the context of stable signal recovery.

The results of applying our method on synthetic data suggest that our migration amplitude recovery can be both efficient and effective in eliminating migration artifacts and recovering the amplitudes in the seismic image. The recovered images showed partial elimination of noise, improved spatial resolution, and enhanced reflectivity amplitude.

5.13 Acknowledgments

The authors would like to thank the authors of CurveLab for making their codes available. The authors would also like to thank W. Symes for providing the migration code and TOTAL E&P for providing the synthetic BP data. This work was in part financially supported by the Natural Sciences

5.13. Acknowledgments

and Engineering Research Council of Canada Discovery Grant (22R81254) and Collaborative Research and Development Grant DNOISE (334810-05) of Felix J. Herrmann 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.

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.
- Berg, E. V. D. and M. Friedlander, 2007, In pursuit of a root: Optimization Online:2007.06.1708, 8.
- Beylkin, G., 1985, Imaging of the discontinuities in the inverse scattering problem by inversion of casual generalized curvelet transform: Journal of Mathematical Physics, **26**, 99–108.
- Candès, E. J., J. Romberg, and T. Tao, 2006, Stable signal recovery from incomplete and inaccurate measurements: **59**, 1207–1223.
- Chattopadhyay, S. and G. A. McMechan, 2008, Imaging conditions for prestack reverse-time migration: Geophysics, **73**, S81–S89.
- Claerbout, J., 1985, Earth soundings analysis, processing versus inversion: Blackwell Scientific Publications.
- Daubechies, I., M. Defrise, and C. de Mol, 2005, An iterative thresholding algorithm for linear inverse problems with a sparsity constraints: **57**, 1413–1457.
- Donoho, D., 1995, Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition: Applied and Computational Harmonic Analy-

sis 2, 101 – 126.

E. Candes, L. Demanet, D. D. L. Y., 2005, Fast discrete curvelet transforms.

E. J. Candès, L. Demanet, D. L. D. and L. Ying, 2006, Fast discrete curvelet transforms: SIAM Multiscale Model. Simul., **5**, 861899.

Gray, S., 1997, True amplitude seismic migration: A comparison of three approaches: Geophysics, **62**, 929–936.

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

Hennenfent, G. and F. J. Herrmann, 2006, Seismic denoising with non-uniformly sampled curvelets: Comp. in Sci. and Eng., **8**, 16–25.

Herrmann, F., C. Brown, Y. Erlangga, and P. Moghaddam, 2009, Curvelet-based migration preconditioning and scaling: Geophysics, **74**, A41–A45.

Herrmann, F., P. Moghaddam, and C. Stolk, 2008, Sparsity- and continuity-promoting seismic image recovery with curvelet frames: Applied and Computational Harmonic Analysis 24, **2**, 150–173.

Herrmann, F. J., D. Wang, and D. J. Verschuur, 2007, Adaptive curvelet-domain primary-multiple separation. ((Submitted to Geophysics)).

J. C. Costa, F. A. Silva Neto, M. R. M. A. J. S. and A. Novais, 2009, Obliquity-correction imaging condition for reverse time migration: Geophysics, **74**, S57–S66.

J. Schleicher, J. C. C. and A. Novais, 2008, A comparison of imaging conditions for wave-equation shot-prole migration: Geophysics, **73**, S219–S227.

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

Mulder, W. and R.-E. Plessix, 2004, A comparison between one-way and

- two-way wave equation migration: *Geophysics*, **69**, 14911504.
- Nemeth, T., C. Wu, and G. Schuster, 1999, Least-squares migration of incomplete reflection data: *Geophysic*, **64**, 208–221.
- Nocedal, J. and W. Wright, 1999, *Numerical optimization*: Springer Verlag.
- O’Brien, M. and S. Gray, 1996, Can we image beneath salt?: *The Leading Edge*, **15**, 17–22.
- Plessix, R. and W. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: *Geophys. J. Int.*, **157**, 975–987.
- Rakesh, A., 1988, A linearized inverse problem for the wave equation: *Comm. on P.D.E.*, **13**, 573–601.
- Rickett, J., 2003, Illumination based normalization for wave-equation depth migration: *Geophysics*, **68**, 208–221.
- Shahidi, R. and F. Herrmann, 2009a, *Curvelet-domain matched filtering*: under preparation.
- , 2009b, *Curvelet-domain matched filtering with frequency-domain regularization*: **28**.
- Stolk, C., 2000, Microlocal analysis of a seismic linearized inverse problem: *Wave Motion*, **32**, 267–290.
- Stolk, C. and M. De-Hoop, 2006, Seismic inverse scattering in the downward continuation approach: *Wave Motion*, **43**, 579–598.
- Symes, W. W., 2007, Reverse time migration with optimal checkpointing: *Geophysics*, **72**, SM213–SM221.
- , 2008a, Approximate linearized inversion by optimal scaling of prestack depth migration: *Geophysics*, **73**, R23–R35.
- , 2008b, Migration velocity analysis and waveform inversion: *Geophys-*

ical Prospecting, **56**, 765 – 790.

Tarantola, A., 1987, Inverse problem theory: Elseweir.

Ten-Kroode, A., D. Smith, and A. Verdel, 1998, A microlocal analysis of migration: *Wave Motion*, **28**.

van den Berg, E. and M. P. Friedlander, 2008, Probing the pareto frontier for basis pursuit solutions: *SIAM Journal on Scientific Computing*, **31**, 890–912.

William W. Symes, Christiaan C. Stolk, B. B. and F. Gao, 2004, Reverse time shot-geophone migration: Technical Report.

Y. Zhang, S. Xu, N. B. and G. Zhang, 2007, True-amplitude, angle-domain, common-image gathers from one-way wave-equation migrations: *Geophysics*, **72**, S49–S58.

Chapter 6

Conclusions

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

6.1 Main contributions

The topic of this thesis is seismic image amplitude correction which is directly applied to seismic imaging. The approach I advocate is to view seismic imaging in a mathematical perspective. I identify a transform, called the curvelet transform (Candès and Donoho, 2004), to that effect and use it in a new formulation of the seismic images amplitude correction problem. The most challenging operator in the seismic imaging problem, called normal operator, is a computationally exhausting. This operator needs to be inverted during the imaging process to give an accurate understanding about the reflectors amplitude in a seismic image. Each element of the curvelet transform, namely curvelet, has an attractive property under the normal operator. In this thesis, I proved that each curvelet remains approximately invariant under the normal operator, which gives us an exceptional ability to approximately invert the normal operator with only one iteration. I call

this approach the curvelet match filtering (CMF).

I leverage this approximation towards the development of a stable amplitude recovery by coining it with sparsity-promoting inversion, which is basically a large-scale one-norm solver.

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

6.1.1 Sparsity and continuity promoting seismic image recovery with curvelet frames

I combine the compression of images by curvelets with the invariance of this transform under the normal operator. This combination allows for a formulation of a stable recovery method for seismic amplitudes. During the recovery, the normal operator is approximately inverted. Compared to other approaches for migration scaling, the presented method (i) includes a theoretical bound on the L^2 - error for the diagonal approximation in the curvelet domain; (ii) prescribes a procedure for the estimation of the diagonal from numerical implementations of the imaging operators; (iii) formulates the amplitude recovery problem as a nonlinear optimization problem, where the inversion of the diagonalized normal operator is regularized by imposing sparsity in the curvelet domain and continuity along the imaged reflectors.

The diagonal approximation is used to formulate the seismic amplitude recovery in terms of a constrained optimization problem. The amplitude-corrected image is obtained by solving this sparsity- and continuity-promoting optimization problem. The invariance of curvelets under the normal opera-

tor preserves the sparsity. The cost of computing the diagonal approximation is one demigration-migration per reference vector, which is much less compared to the cost of Krylov-based least-squares inversion. The recovery results show an overall improvement of the image quality. The joined sparsity- and continuity-enhanced image has diminished artifacts, improved resolution and recovered amplitudes.

6.1.2 **Curvelet-based migration preconditioning and scaling**

I present a method to approximate the normal operator and use this approximation as a preconditioner in the least squares framework. Because of the size of the seismic imaging problem, preconditioning of least-squares migration is an interesting topic. Traditionally, the first few iterations of the LSQR algorithm of least-squares migration are known to make significant progress towards the solution. Unfortunately, even for this limited number of iterations, the computational costs are often still prohibitively large for practical problems. Our method, partly resolves this issue through a combination of left and right preconditioning together with a curvelet-domain scaling. Inclusion of the latter proved particularly important because it restores the amplitudes and leads to faster convergence, at a relatively small computational overhead. Preconditioning also plays a pivotal role in making this approach numerically feasible and may extend to a solution of the full-waveform inversion problem with sparsity promotion. Both approaches are justified by ample evidence that curvelets are sparse on the model.

6.1.3 **Curvelet-based seismic data processing**

Beside the seismic wavefield reconstruction problem, I 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. The promising results obtained show that the insights gained from the developments of CMF can be leveraged towards a much broader range of applications. This prospect opens an exciting new outlook towards future developments in exploration seismology.

6.1.4 **True amplitude depth migration using curvelets**

I presented a fast and robust approach for approximation of the normal operator by revisiting of the earlier work and taking into account the migration noise and the data noise. Compared to other approaches for migration amplitude recovery, some improvements in the design of curvelet amplitude recovery method including, speeding up the calculation of the normal operator by diagonalizing it in curvelet domain, designing an efficient approximation that takes into account a laterally-variant velocity models and steep reflectors. The results of applying method on synthetic data suggest that migration amplitude recovery method can be both efficient and effective in eliminating migration artifacts and recovering the amplitudes in the seismic image. The recovered images showed partial elimination of noise, improved spatial resolution, and enhanced reflectivity amplitude.

6.2 Follow-up work

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

6.2.1 Full waveform inversion

Our estimation for normal operator can be directly applied to the context of full waveform inversion in which the normal operator is inverted and used iteratively. In this context, the velocity model (not just reflectivity in the migration context) is estimated directly from data (see e.g. Symes, 2008).

6.2.2 3D true amplitude migration

Our 2D true amplitude migration can be easily extend to 3D. This can be done in two ways, first, I apply the curvelet smoothing in 2D and for 3rd dimension I can exploit wavelet smoothing. Second, I can use the 3D curvelet transform and extend the method to 3D curvelet diagonal estimation. The latter, needs a sophisticated algorithm to locate the closest curvelets in 3D rather than 2D for smoothing operator.

6.3 Adding more priori knowledge to solution

In this thesis, I propose some ideas for amplitude recovery which presume that the seismic image are sparse in the curvelet domain. The common theme of most ideas is that there is no priori knowledge of the geological structure of the survey area. However, if a priori knowledge about the structure already obtained through the well-log or other geological/geophysical

experiments, it can be used within stable recovery framework in addition of sparsity to reconstruct seismic wavefields.

6.4 Current limitations

I examine both the practical and the fundamental weaknesses of the current curvelet transform.

6.4.1 Curvelet code

The true amplitude migration 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 extend the short axis to be sufficiently large enough. This extension can be done by padding the image along the short axis or wrapping it.

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. Another option is to combine the curvelet transform with another transform (Herrmann, 2003;

6.4. Current limitations

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. In addition, in some cases, defining the smoothing operator in the transform domain is not as straightforward as curvelets since in some cases finding the closest elements in transform domain might require a significant amount of operations.

Bibliography

- Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2005, Curvelab: software. (Available at <http://www.curvelet.org>).
- 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.
- Herrmann, F. J., 2003, Multifractional splines: Application to seismic imaging.
- Kutyniok, G. and D. Labate, 2005, Shearlets: website. (www.shearlet.org).
- Lu, Y. M. and M. N. Do, 2007, Multidimensional directional filter banks and surfacelets: Transactions on Image Processing, **16**, no. 4.
- 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).
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, no. 6, 765–790.

Appendix A

Proofs of lemma 1 and theorem 1

We first prove Theorem 1 from Lemma 1, we then prove the Lemma which appears in chapter 2.

Proof: [Proof of Theorem 1] C^T acting on a vector $\mathbf{c} = \{c_\mu\}_{\mu \in \mathcal{M}}$ in ℓ^2 is simply given by $C^T \mathbf{c} = \sum_\nu c_\nu \varphi_\nu$. Hence,

$$C^T \mathbf{D}_\Psi C \varphi_\mu = \sum_\nu (C \varphi_\mu)_\nu a(x_\nu, \xi_\nu) \varphi_\nu.$$

The difference $(\Psi(x, D) - C^T \mathbf{D}_\Psi C) \varphi_\mu$ can be written as, using that $C^T C = \text{Id}$,

$$\sum_\nu (C \varphi_\mu)_\nu (\Psi(x, D) - a(x_\nu, \xi_\nu)) \varphi_\nu$$

The vector $(C \varphi_\mu)_\nu$ is bounded in $\ell^2(\mathcal{M})$ of vectors with curvelet indices, and its entries can only be nonzero if $|\mu| - 2 \leq |\nu| \leq |\mu| + 2$, since the support of two curvelets in the Fourier domain is disjoint when curvelets are two or

A version of this appendix has been published. Herrman, F.J., Moghaddam, P.P. and Stolk, C. (2008) Sparsity- and continuity-promoting seismic image recovery with curvelet frames. *Applied and Computational Harmonic Analysis*, Vol. 24, No. 2, pp. 150-173, 2008.

more scales apart. It follows that

$$\begin{aligned} \|(\Psi(x, D) - C^T \mathbf{D}_\Psi C) \varphi_\mu\|_{L^2(\mathbb{R}^n)}^2 &\leq C_6 \sum_{\nu} \|(C\varphi_\mu)_\nu(A(x, D) - a(x_\nu, \xi_\nu)) \varphi_\nu\|_{L^2(\mathbb{R}^d)}^2 \\ &\leq C_7 \sum_{\nu} 2^{-|\nu|} |(C\varphi_\mu)_\nu|^2 \leq C_8 2^{-|\mu|/2}. \end{aligned}$$

Proof: [Proof of Lemma 1] Possibly after a coordinate transformation, which does not affect the pseudodifferential nature of Ψ , we may assume that $\xi_\mu = (0, \dots, 0, \lambda)$, we will write $\xi' = (\xi_1, \dots, \xi_{d-1})$.

We may take $|\nu|$ greater than some minimum value, so that the support in the wavenumber domain is away from $\xi = 0$. We can then assume that the principal symbol is homogeneous of order 0 in the support of the curvelet. On the support of the curvelet in the Fourier domain the symbol can be written as $a(x, \frac{\xi'}{\xi_n})$. We apply a preparation theorem to this term. There are C^∞ functions $b_n(x, \xi'/\xi_d)$, $n = 1, \dots, 2d - 1$, such that

$$\begin{aligned} a(x, \xi'/\xi_d) - a(x_\nu, \xi'_\nu/\xi_{\nu,d}) \\ = \sum_{n=1}^d (x - x_\nu)_n b_n(x, \xi'/\xi_d) + \sum_{n=1}^{d-1} \left(\frac{\xi_n}{\xi_d} - \frac{\xi_{\nu,n}}{\xi_{\nu,d}} \right) b_{d+n}(x, \xi'/\xi_d). \end{aligned}$$

Therefore, we can write

$$\begin{aligned} &(\Psi(x, D'/D_d) - a(x_\nu, \xi'_\nu/\xi_{\nu,d})) \varphi_\nu \\ &= \left[\sum_{n=1}^d (x - x_\nu)_n b_n(x, D'/D_d) + \sum_{n=1}^{d-1} b_{d+n}(x, D'/D_d) \frac{D_n}{D_d} \right] \varphi_\nu. \quad (\text{A.1}) \end{aligned}$$

Consider first the contribution for one of the b_n with $1 \leq n \leq d$. The

operator acting on the curvelet reads

$$\begin{aligned} & (x - x_\nu)_n b_n(x, D'/D_d) \varphi_\nu \\ &= b_n(x, D'/D_d) (x - x_\nu)_n \varphi_\nu + [(x - x_\nu)_n, b_n(x, D'/D_d)] \varphi_\nu. \end{aligned}$$

Because of the support and decay properties of curvelets we have $\|(x - x_\nu)_n \varphi_\nu\|_{L^2} \leq C_9^{-|\nu|/2}$, therefore

$$\|b_n(x, D'/D_d) (x - x_\nu)_n \varphi_\nu\|_{L^2(\mathbb{R}^d)} \leq C_{10} 2^{-|\nu|/2}$$

By the calculus of pseudodifferential operators, the operator $[(x - x_\nu)_n, b_n(x, D'/D_d)]$ is a pseudodifferential operator of order -1 , which is continuous $H^{-1}(\mathbb{R}^d)$ to $L^2(\mathbb{R}^d)$ therefore

$$\|[(x - x_\nu)_n, b_n(x, D'/D_d)] \varphi_\nu\|_{L^2(\mathbb{R}^d)} \leq C_{10} \|\varphi_\nu\|_{H^{-1}(\mathbb{R}^d)} \leq C_{11} 2^{-|\nu|}.$$

Adding the previous two estimates, we find that

$$\|(x - x_\nu)_n b_n(x, D'/D_d) \varphi_\nu\|_{L^2(\mathbb{R}^d)} \leq C_{12} 2^{-|\nu|/2}. \quad (\text{A.2})$$

We now consider the b_n , $d+1 \leq n \leq 2d-1$. We must estimate $b_{d+n}(x, D'/D_d) \frac{D_n}{D_d} \varphi_\nu$. By (2.11) we have $\|\frac{D_n}{D_d} \varphi_\nu\|_{L^2(\mathbb{R}^n)} \leq C_{13} 2^{-|\nu|/2}$. Therefore,

$$\|b_{d+n}(x, D'/D_d) \frac{D_n}{D_d} \varphi_\nu\|_{L^2(\mathbb{R}^d)} \leq C_{14} 2^{-|\nu|/2}. \quad (\text{A.3})$$

From (A.1) and the estimates (A.2) and (A.3) for the summands, we find

the result.

Appendix B

Reverse time wave equation migration

This appendix is part of chapter 5. Most migration operators are defined to be the adjoint of what is called the scattering operator. This assumption is also true for reverse time migration.

B.1 Single scattering

The causal acoustic Green's function $G(\mathbf{x}, t; \mathbf{x}_s)$, $\mathbf{x} \in \mathbf{R}^3$ for a point source at $\mathbf{x} = \mathbf{x}_s$ is the solution of

$$\frac{1}{v^2(\mathbf{x})} \frac{\partial^2 G}{\partial t^2}(\mathbf{x}, t; \mathbf{x}_s) - \nabla_x^2 G(\mathbf{x}, t; \mathbf{x}_s) = \delta(\mathbf{x} - \mathbf{x}_s) \delta(t) \quad (\text{B.1})$$

with $G = 0, t < 0$ and v is the acoustic wave velocity field.

Denote by $m(\mathbf{x}) = \delta v(\mathbf{x})/v(\mathbf{x})$ a relative perturbation of the velocity field. Then linearization of the wave equation yields for the corresponding perturbation of the Green's function

A version of this appendix will be submitted for publication. Moghaddam, P.P., Herrmann F. J. and Shahidi, R. (2010) True amplitude depth migration using curvelets.

$$\frac{1}{v^2(\mathbf{x})} \frac{\partial^2 \delta G}{\partial t^2}(\mathbf{x}, t; \mathbf{x}_s) - \nabla_x^2 \delta G(\mathbf{x}, t; \mathbf{x}_s) = \frac{2m(\mathbf{x})}{v^2(\mathbf{x})} \frac{\partial^2}{\partial t^2} G(\mathbf{x}, t; \mathbf{x}_s) \quad (\text{B.2})$$

whose solution has the integral representation at the source and receiver points $\mathbf{x}_r, \mathbf{x}_s$

$$\delta G(\mathbf{x}_r, t; \mathbf{x}_s) = \frac{\partial^2}{\partial t^2} \int dx \int d\tau \frac{2m(\mathbf{x})}{v^2(\mathbf{x})} G(\mathbf{x}, t - \tau; \mathbf{x}_r) G(\mathbf{x}, \tau; \mathbf{x}_s) \quad (\text{B.3})$$

B.2 Shot-geophone modeling and migration

The single-scattered wave field is the time convolution of δG with a source wavelet. The main concern of this paper is the kinematic relationships between data and image, thus we ignore the filtering effect of the source functional and replace it by delta function. This replacement of the source by an impulse does not violate any of our assumptions regarding the adjoint state method, thus the Born modeling operator $K[v]$ is

$$K[v]m(\mathbf{x}) = \delta G(\mathbf{x}_r, t; \mathbf{x}_s) \quad (\text{B.4})$$

The crux of our amplitude recovery method relies on the shot-geophone migration operator to be the adjoint of the shot-geophone modeling operator. The derivation of the adjoint reverse time implementation is a minor variation on the usual implementation of reverse time migration (the 'adjoint state method', ((see e.g. Tarantola, 1987; Whitmore, 1983; Lailly, 1983; Yoon and Hong, 2003; Symes and Stolk, 2004; Symes, 2007)). The

B.2. Shot-geophone modeling and migration

result is

$$K^*[v]d(\mathbf{x}_r, t; \mathbf{x}_s) = \hat{m}(\mathbf{x}) = - \int dx_s \int_0^T dt 2v(\mathbf{x})q(\mathbf{x}, t; \mathbf{x}_s) \frac{\partial^2 G}{\partial t^2}(\mathbf{x}, t; \mathbf{x}_s) \quad (\text{B.5})$$

where the adjoint state or backpropagated field $q(\mathbf{x}, t; \mathbf{x}_s)$ satisfies $q = 0, t > T$ and

$$\frac{1}{v^2(\mathbf{x})} \frac{\partial^2 q}{\partial t^2}(\mathbf{x}, t; \mathbf{x}_s) - \nabla_x^2 q(\mathbf{x}, t; \mathbf{x}_s) = \int dx_r d(\mathbf{x}_r, t; \mathbf{x}_s) \delta(\mathbf{x} - \mathbf{x}_r) \quad (\text{B.6})$$

The migration operator defined by above equations is the reverse time migration operator. Symes and Stolk (2004) showed that the migration operator which is defined in equations B.5 and B.6 is the adjoint of the modeling operator defined in equation B.3. The reverse time migration that is used for this work is based on the Symes and Stolk (2004) implementation, for which the adjointness of the modeling and migration operators is properly tested in the discrete sense. By having the migration and modeling operators properly set, we can proceed with our amplitude recovery method.

Bibliography

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: SIAM, 206–220.

Symes, W., 2007, Reverse time migration with optimal checkpointing: Geophysics, **72**, 213–221.

Symes, W. and C. Stolk, 2004, Reverse time shot-geophone migration: The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

Tarantola, A., 1987, Inverse problem theory: Elseweir.

Whitmore, N. D., 1983, Iterative depth migration by backward time propagation: Presented at 53rd Annual International SEG Meeting, Las Vegas.

Yoon, K. and S. Hong, 2003, 3D reverse-time migration using the acoustic wave equation: An experience with the SEG/EAGE data set.: The Leading Edge, 22–38.