Fast imaging with surface-related multiples

by

Ning Tu

B.Eng., Beijing University of Posts and Telecommunications, 2005
M.Sc., Tsinghua University, 2009

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

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Geophysics)

The University of British Columbia
(Vancouver)

August 2015

Abstract

Surface-related multiples, which are waves that bounce more than once between the water surface and the subsurface reflectors, constitute a significant part of the data acquired in marine seismic surveys. If left untreated, they can lead to misplaced phantom reflectors in the image, and result in erroneous interpretations of the subsurface structure. As a result, these multiples are removed before the imaging procedure in conventional seismic data processing. However, because they interact more with the subsurface medium, they may carry extra information that is not present in the primaries. Therefore instead of removing these multiples, a more desirable alternative is to make active use of them. We derive from the well-established "Surface-Related Multiple Elimination" relation, and arrive at a linearized expression of the wave-equation based modelling that incorporates the surface-related multiples. We then present a computationally efficient approach to iteratively invert this expression to obtain an image of the subsurface from data that contain multiples. We achieve the computational efficiency inside each iteration by *(i)* using the wave-equation solver to implicitly carry out the expensive multiple prediction; and *(ii)* reducing the number of wave-equation solves during data simulation by subsampling the monochromatic source experiments. We show that, compared with directly applying the cross-correlation/deconvolutional imaging conditions, the presented approach can suppress the coherent imaging artifacts from multiples more effectively. We also show that, by curvelet-domain sparsity promoting and occasionally drawing new data samples during the inversion, the proposed inversion method gains improved robustness to velocity errors in the background model, as well as modelling errors incurred during linearization of the wave-equation. To combine the information encoded in both the primaries and the multiples, we then propose a highly-accurate source estimation method to jointly invert the total upgoing wavefield. We show with field data examples that we can reap benefits from both the relative noise-free primaries and the extra illumination coverage of the multiples.

We also demonstrate that the inclusion of multiples help mitigate the amplitude ambiguity during source estimation. We conclude the thesis with an outlook for future research directions, as well as potential extensions of the proposed work.

Preface

All of the presented work is carried out with the supervision of Dr. Felix Herrmann, in the Seismic Laboratory for Imaging and Modelling at the University of British Columbia. Dr. Felix Herrmann was extensively involved in all my research projects as the supervisory author, from conceptual formation to experiment design and manuscript edits.

A version of Chapter 2 has been published [Ning Tu and Felix J. Herrmann, Fast imaging with surface-related multiples by sparse inversion, Geophysical Journal International, 2015, vol. 201, p. 304-317]. I was the lead investigator and the manuscript composer.

A version of Chapter 3 has been published [Ning Tu, Xiang Li, and Felix J. Herrmann, Controlling linearization errors in ℓ_1 regularized inversion by rerandomization, SEG Technical Program Expanded Abstracts, 2013, vol. 32, p. 4640-4644]. I was the lead investigator and the manuscript composer. Xiang Li contributed in the development of the algorithm.

A version of Chapter 4 has been published [Ning Tu, Tristan van Leeuwen, and Felix J. Herrmann, Limitations of the deconvolutional imaging condition for two-way propagators, in SEG Technical Program Expanded Abstracts, 2013, vol. 32, p. 3916-3920]. I was the lead investigator and the manuscript composer. Tristan van Leeuwen was involved in the concept formation.

A version of Chapter 5 has been submitted for publication [Ning Tu, Aleksandr Aravkin, Tristan van Leeuwen, Tim Lin and Felix J. Herrmann, Source estimation with multiples—fast ambiguity-resolved seismic imaging, submitted on May 14 2015]. I was the lead investigator, responsible for concept formation, carrying out numerical experiments, and manuscript composition. Aleksandr Aravkin and Tristan van Leeuwen provided proof of the theory and developed the numerical algorithm that I use the solve the problem. Tim Lin contributed in the concept formation.

A version of Chapter 6 has been published [Ning Tu and Felix J. Herrmann, Fast least-squares imaging with surface-related multiples: application to a North-Sea data set, The Leading Edge, 2015, vol. 34, no. 7, p.

788-794]. I was the lead investigator and the manuscript composer.

A version of Chapter 7 has been published [Felix J. Herrmann, Ning Tu, and Ernie Esser, Fast "online" migration with Compressive Sensing, 77th EAGE Conference & Exhibition, 2015, Madrid, Spain]. Felix motivated and supervised this research project, and composed the manuscript. I contributed to the concept formation, developed the algorithm, and carried out the numerical experiments. Ernie Esser was involved in the concept formation.

A version of Chapter 8 has been published [Rajiv Kumar, Ning Tu, Tristan van Leeuwen, and Felix J. Herrmann, Least-squares extended imaging with surface-related multiples, CSEG GeoConvention 2015, Calgary, Canada]. Rajiv Kumar was the leading investigator and the manuscript composer. I contributed to the concept formation, algorithm development, and experiment design. Tristan van Leeuwen was involved in the early development of the algorithm.

Table of Contents

List of Tables

List of Figures

Glossary

RTM - Reverse-Time Migration
SRME - Surface-Related Multiple Elimination
EPSI - Estimation of Primaries by Sparse Inversion
SPG - Spectral Projected Gradient
BPDN - Basis Persuit De-Noise
LASSO - Least Absolute Shrinkage and Selection Operator
AMP - Approximate Message Passing

Acknowledgments

First and foremost, I would like to thank my research supervisor Dr. Felix Herrmann. The high standards that he abide by in conducting scientific research have profoundly influenced me, and will form a solid foundation for my career as a responsible scientist. The critical thinking skills that I learned through his training will benefit me in all aspects of my life, and I am deeply grateful for this great gift. I am also thankful for the opportunities he provided to get myself exposed to international geophysical communities via all kinds of conferences and meetings.

Thanks to Michael Bostock, Eldad Haber, Michael Friedlander, and Ozgur Yilmaz for sitting on my supervisory committee. Their support and help during the last six years are very important.

My gratitudes also go to all my colleagues at the SLIM lab. I would like to thank Henryk Modzelewski, Manjit Dosanjh, Miranda Joyce, and Ian Hanlon for their professional assistance, and everyone else — especially Xiang Li and Tim Lin — for their valuable scientific insights as research collaborators, and for the past six-year's company and friendship.

Many thanks to Dr. Eric Verschuur and Dr. Joost van der Neut from TU Delft, as well as Dr. Shaoping Lu, Dr. Dan Whitmore, and Dr. Steve Kelly from PGS, for the very insightful technical discussions. I also would like to thank the authors of SPGℓ_1, the SEG/EAGE salt model, and the Sigsbee 2B model, which I used extensively in my research. Many thanks to PGS for providing the field data set that I used in Chapter 6.

I am utterly fortunate to know and marry my wife during my PhD study. Her unconditional support and love encourage me everyday to conquer the obstacles in my research and other aspects of life. I also thank my parents for tolerating me for six years to study and live at such a distant place from their home in China. I am grateful to my kids — they bring new purposes to my life and motivate me to make this world a better place through my efforts.

Thanks to all funding agencies that supported my research — the Chi-

nese Scholarship Council, NSERC, and all SINBAD consortium sponsors. This thesis is not possible without their generous financial support.

To my dear wife Yuan.
You make me who I am today.

To my dear Sarah and Jonathan.
You give my life new purposes.
And a new lifelong job, twice.

To all PhD candidates.
You will make it.

Chapter 1

Introduction

In this chapter, we first briefly explain the reflection seismic method in marine towed-streamer acquisitions, the origin of the surface-related multiples, and the challenges they impose on the following seismic data processing and imaging workflow. We then state the theme of the thesis, and continue with the goals we aim to achieve as a result of the presented work. We conclude the chapter with the outline of this thesis. We will keep this introduction chapter concise, as each of the main chapters also contains its own introduction and literature review.

1.1 Surface-related multiples: the origin and the problem

In marine towed-streamer seismic acquisitions, practitioners collect seismic reflection data by exciting artificial seismic sources, and gather the reflected seismic waves with receivers deployed along one or more cables, known as the *streamers*. Both the source(s) and the streamer(s) are towed by specially-equipped vessels, and are placed at certain depth(s) beneath the surface of the water. Once the data are collected, they undergo a series of processing procedures, such as the removal of both coherent and incoherent noises in the data, before they are *migrated* to produce a seismic image that displays the geological structures of the subsurface. The term *migration*, which we use interchangeably with *imaging* throughout this thesis, as defined in Yilmaz (2001), refers to the process of mapping reflection/diffraction seismic data to their true subsurface positions. Geologists then extract information from the seismic image to locate potential oil/gas reservoirs. The more faithful this image is to the true geology, the more valuable insights it can provide

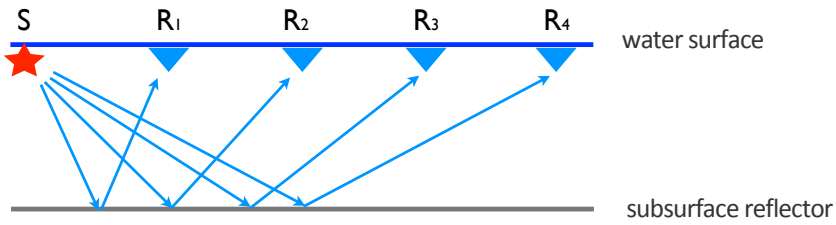
in locating the oil/gas reservoirs.

Cartoon 1.1(a) illustrates a 2D line of the marine towed-streamer seismic survey in a simplified medium. Seismic waves propagate from the source, reflect at boundaries between two different media, and are picked up by the receivers. In this cartoon, $\mathbf{S}$ indicates a seismic source, and $\mathbf{R}_1$ through $\mathbf{R}_4$ indicate several receivers along the streamer. The solid blue lines indicate the ray paths of the *primaries* that are picked up by these receivers. The term *primaries*, as defined in Verschuur (2006), refer to seismic waves that undergo only one upward reflection. However, primaries are not the only wave modes that are picked up by these receivers. Because the water-air contact acts as a perfect reflector, the upgoing waves picked up by the receivers will continue to propagate upward until they hit the water surface, bounce at the surface, and propagate downward into the subsurface again. As a result, the *surface-related multiples* are also picked up by the receivers. As shown in Figure 1.1(b), a primary event is recorded by $\mathbf{R}_1$, and its corresponding first/second/third-order surface-related multiples are recorded by $\mathbf{R}_2$ through $\mathbf{R}_4$ respectively.

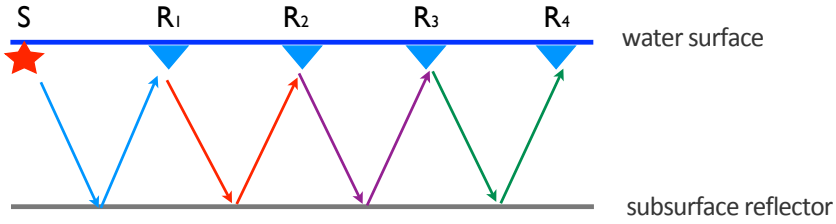
Throughout this thesis, we define *surface-related multiples* as the wavefields that undergo at least one downward reflection at the water surface (Verschuur, 2006). In the data space, these multiples resemble displaced copies of the primaries, and if left untreated in the data, they can lead to phantom reflectors in the seismic image. We use a simple homogenous model with one horizontal reflector to show these multiples in a common-shot gather in Figure 1.2, and the resulting phantom reflectors in the reverse-time migrated (RTM, Baysal et al. (1983)) seismic image in Figure 1.3. We also show the waveforms and the RTM image of the primaries to compare with. The phantom reflectors arising from the multiples can lead to erroneous interpretations, and as a result, conventional data processing entails the removal of these multiple events in the data. A comprehensive literature review of existing de-multiple methods is included in Chapter 2.

1.2 A treasure in disguise

Because multiples undergo extra bounces between the surface and the subsurface reflectors compared with primaries, they are exposed more to the subsurface structures, and may therefore provide extra information that is absent in the primaries. As illustrated in Figure 1.4, for a given seismic source and a certain range of receiver coverage, surface-related multiples can illuminate a wider range of subsurface structures, as all the bouncing points at the water surface now act as secondary sources. In the literature,



(a)



(b)

Figure 1.1: Illustrations of the origin of the surface-related multiples. **(a)** The ray-paths of the primaries. **(b)** The ray-paths of a primary event (in blue) and its corresponding first/second/third-order surface-related multiples (in red/purple/green respectively).

Verschuur (2011) and Lu et al. (2014a) showed that in cases where the illumination from the primaries is not enough, these multiples help to remove the acquisition footprints. Therefore it is desirable to make active use of the surface-related multiples rather than removing them by a de-multiple procedure during seismic data processing. A comprehensive literature review of existing imaging-with-multiple methods — including their respective benefits and limitations — is detailed in Chapter 2.

1.3 Curvelet-domain sparsity promoting

The inversion scheme we use to image surface-related multiples borrows insights from recent developments of compressive sensing. We reduce the per-iteration data-simulation cost by *subsampling* the source experiments, which are proportional to the number of wave-equation solves. To stabilize

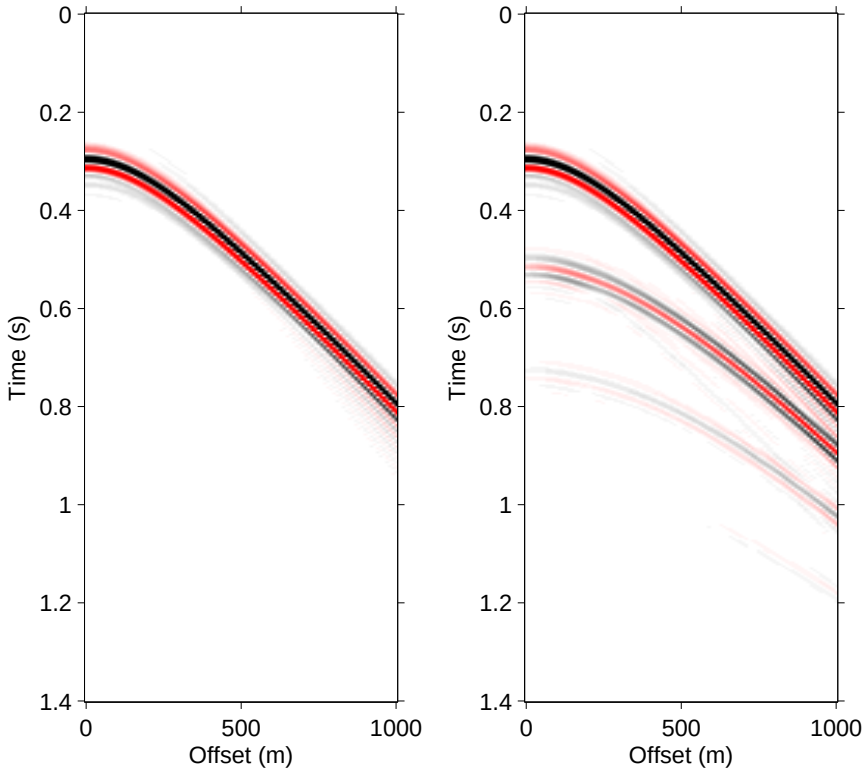
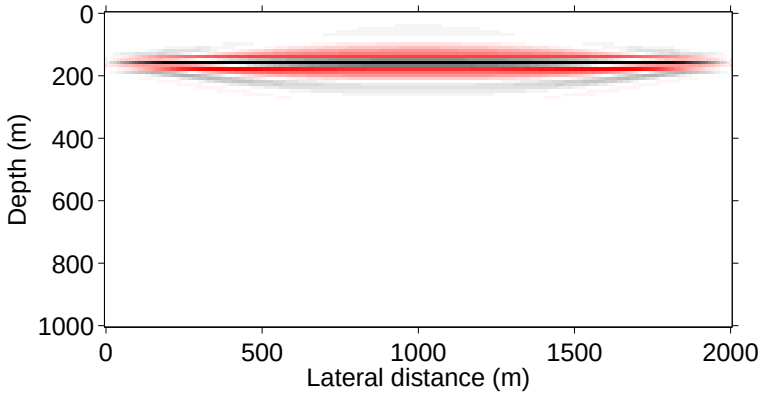


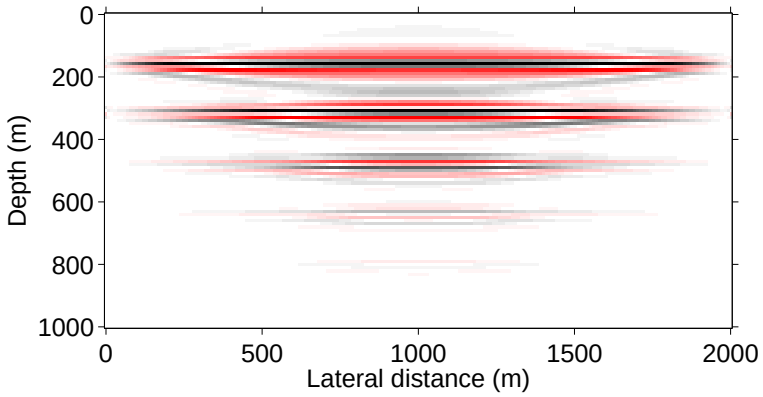
Figure 1.2: **(a)** The waveform of primaries. **(b)** The waveform of data that contain multiples.

the inversion of the *subsampled* system, we promote curvelet-domain sparsity, i.e., we look for the solution that has the most sparse representations in the curvelet domain where the sparsity is measured using the ℓ_1-norm. We use several figures to illustrate the necessity of using the curvelet transform and the ℓ_1-norm.

Curvelets are designed to detect curved singularities (Candès and Donoho, 2004). Figure 1.6 shows several curvelet "atoms" of certain angles and scales (adapted from Herrmann and Hennenfent, 2008). We can see that curvelets are continuous in one direction and oscillatory in the perpendicular direction, plus they are multi-scale and directional. Therefore they can represent curved singularities with only a few coefficients. Figure 1.6 shows the decay of coefficients in the physical domain as well as in the curvelet domain. We



(a)

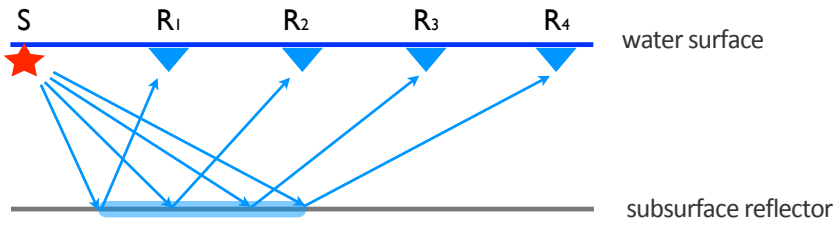


(b)

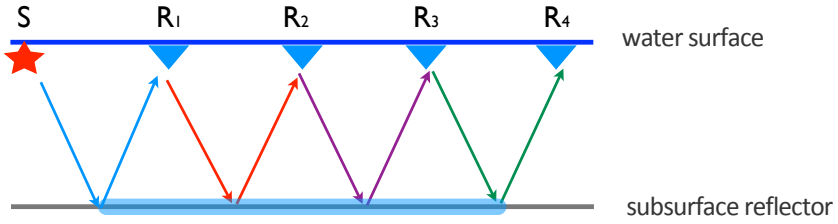
Figure 1.3: **(a)** The RTM image of primaries. **(b)** The RTM image of data that contain multiples.

can see that the curvelet representation is much sparser than in the physical domain.

Figure 1.7 shows that the solution with the minimal ℓ_1-norm is a sparse solution, because the ℓ_1-ball almost always touches the solution space at one of its vertex. Compared with the ℓ_1-norm, using the ℓ_2-norm leads to a non-sparse solution.



(a)



(b)

Figure 1.4: Illustrations showing the wider illumination coverage of multiples compared with primaries. The translucent blue color highlights the illuminated subsurface structure by the primaries (top) and the surface-related multiples (bottom).

1.4 Theme and objectives of the thesis

This thesis focuses on an effective and computationally efficient inversion approach to image the surface-related multiples. We aim to achieve the following objectives through this thesis:

1. To understand the necessity of using an inversion-based approach to image the surface-related multiples, to suppress the coherent artifacts associated with these multiples such as the phantom reflectors shown in Figure 1.3(b). We seek insights by carefully comparing inversion with direct application of either the cross-correlation imaging condition (such as RTM of multiples, as reported by Liu et al. (2011)), or the deconvolutional imaging condition ("deconvolution" in the sense that the source wavefield is deconvolved from the cross-correlated image, as reported by Muijs et al. (2007); Whitmore et al. (2010)).

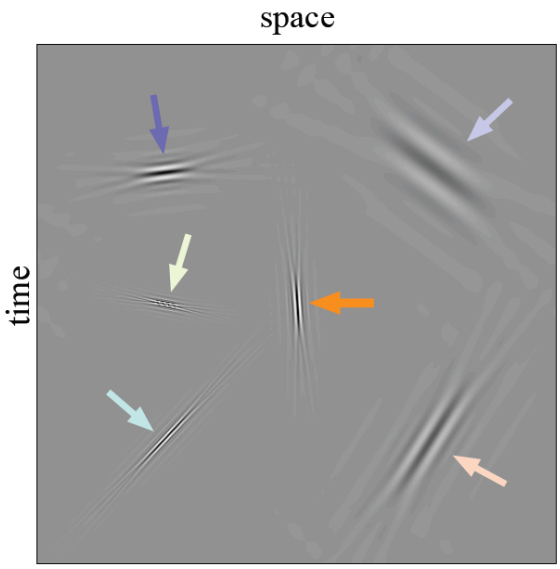


Figure 1.5: Curvelets at different angles and scales.

2. To significantly reduce the high computational costs of the inversion procedure, without compromising the image quality. The computational costs inside each iteration has two main sources: *(i)* carrying out multiple prediction; and *(ii)* solving wave-equations in data simulation. We aim to reduce the computational costs from both factors.

3. To improve the robustness of the inversion procedure w.r.t. modelling errors incurred during the linearization of the wave-equation (i.e., the *linearization* errors), and velocity errors in the background model. As migration is based on the linearization of the wave-equation (migration equates to the adjoint operation of the linearized modelling, as reported by Lailly (1983)), linearization errors can significantly degrade the image quality if not properly dealt with. As the wave-equation is linearized at a given background velocity model (Plessix and Mulder, 2004), and the linear (i.e., first-order) approximation is based on the assumption that the model perturbations (i.e., the difference between the true model and the background model) are small enough for the higher-order terms (of the Taylor expansion of the wave-equation w.r.t. the model parameters) to vanish, a biased estimate of the background velocity model will also introduce errors to the image, for example causing reflectors to shift in depth or even defocus. We aim to develop

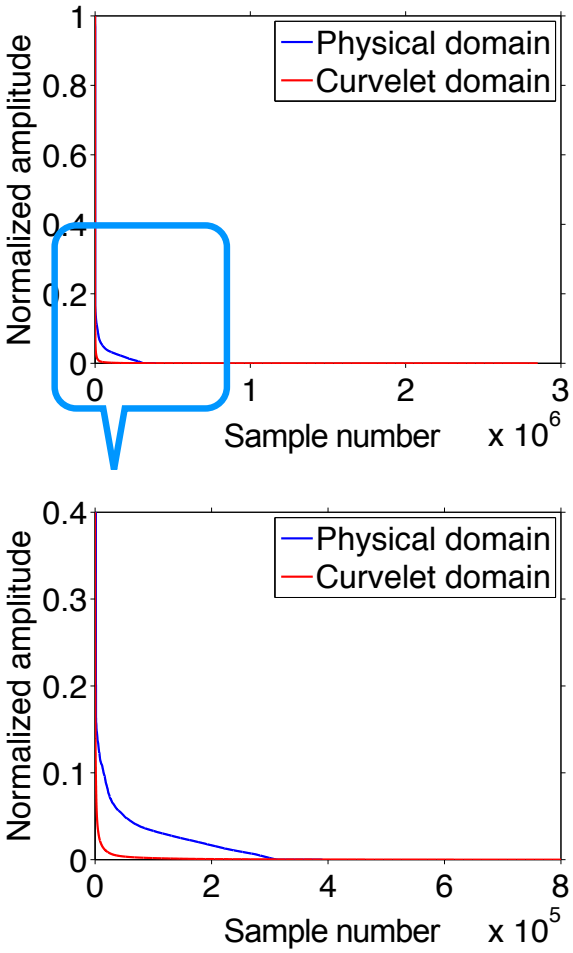


Figure 1.6: Decay of coefficients of a geological model in the physical domain as well as in the curvelet domain. The bottom figure shows a zoomed-in section of the top figure.

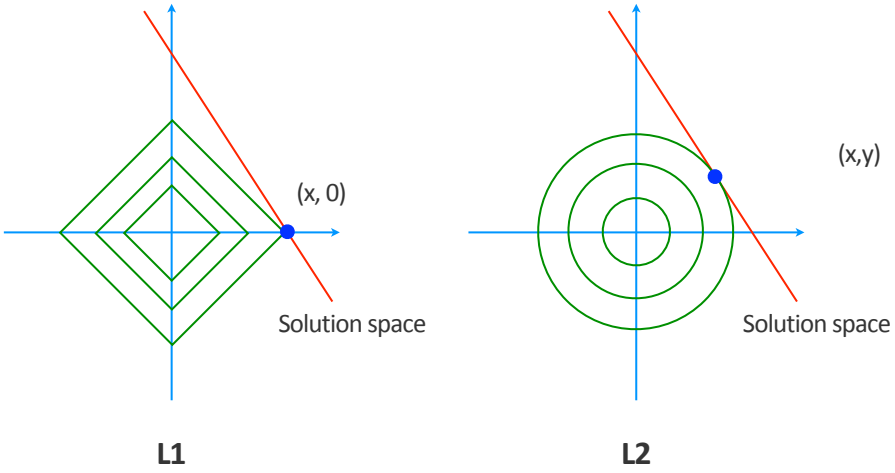


Figure 1.7: Illustration that minimizing the ℓ_1-norm leads to a sparse solution while using ℓ_2-norm does not.

an inversion approach that is relatively robust to both types of errors.

4. To jointly use the information encoded in both the primaries, which are of relatively high signal-to-noise-ratio (SNR), and multiples that provide extra illumination coverage. Compared with imaging the two wavefields separately, joint imaging produces one image, with properly scaled contributions from primaries and multiples without the need to separate them. Given the true source wavelet, we aim to perform the joint inversion by matching the observed total upgoing wavefield including primaries and multiples. For field seismic data, as the source wavelet is typically unknown, we aim to combine source estimation into the joint imaging procedure. As this problem falls into the category of the blind-deconvolutional type of problems (Stockham et al., 1975), it can be plagued by amplitude and phase ambiguities. We also aim to investigate whether and how we can resolve the possible ambiguities by exploiting the extra information from multiples.

1.5 Outline

In Chapter 2, we explain the proposed iterative inversion method and the measures we take to reduce its computational cost, given the knowledge of the background velocity model and the source wavelet. We first pro-

vide a comprehensive review of the literature on existing methods that deal with surface-related multiples in the data. We then follow the well-established Surface-Related Multiple Elimination (SRME) relation, and derive a linearization of the wave-equation-based modelling that incorporates the surface-related multiples. After showing the limitation of the conventional cross-correlation imaging condition in imaging the multiples, we present the proposed sparsity-promoting inversion approach and explain its key ingredients using a stylized synthetic model. Next we discuss the influence of the inversion parameters on the performance of the algorithm, and how we deal with certain practical issues in imaging field seismic data — such as the unknown source wavelet, and possible velocity errors in the background model. We verify the effectiveness of the proposed method using realistic synthetic examples.

In Chapter 3, we focus on a single topic we hardly touch in Chapter 2, namely how the sparsity-promoting inversion approach presented in Chapter 2 gains robustness w.r.t. linearization errors (i.e., modelling errors incurred during the linearization of the wave equation) by a procedure we refer to as "rerandomization". The linearization errors can be interpreted as the coherent mismatch between the linearized modelling and the input data that contain non-linear wave phenomena. We demonstrate the effectiveness of the proposed approach using realistic synthetic examples.

In Chapter 4, the main question to answer is whether we can obviate the iterative inversion procedure — and alternatively use the *deconvolutional* imaging condition with two-way propagators. We provide first a theoretical analysis, and then synthetic examples to reveal the limitation of the deconvolutional imaging condition. We reach a conclusion that a true iterative inversion approach is necessary to properly image the surface-related multiples.

In Chapter 5, we deal with the problem of the unknown source wavelet in seismic imaging, especially in the joint imaging of primaries and multiples (in Chapter 2 we assume the source wavelet is known). We first provide the motivation for this research, review the related literature, and explain the contribution of our work. We then detail the method, and explain the challenge in obtaining deterministic — i.e., ambiguity-resolved — source estimates as well as how we address this issue by actively making use of self-consistency between primaries and the corresponding surface-related multiples. In the end, we verify the efficacy of the proposed approach using realistic synthetic examples, quantitatively analyze the performance of the method, and study its robustness to the initial guess of the wavelet.

In Chapter 6, we apply the proposed inversion approach — together with

the source estimation method — to a North-sea field data set. We study its effectiveness in suppressing the coherent artifacts associated with multiples, and the benefits of the joint imaging of primaries and multiples compared with imaging them separately.

In Chapter 7, we present a prototype of a simplified sparse-inversion algorithm via so-called "linearized Bregman projection". It will be an important topic for our future research, as it leads to fast convergence, yields high-fidelity images, and affords easy algorithmic implementation. We explain the algorithm, compare it with a representative existing approach, and show promising preliminary imaging results — including imaging with multiples — using realistic synthetic models and noise-free data.

In Chapter 8, we present an extension of the proposed work to extended imaging — least-squares extended imaging with surface-related multiples. Extended imaging is widely used in migration velocity analysis. Through this work, we aim for a data processing workflow in the future that obviates the primary-multiple separation procedure altogether, which is currently necessary because velocity analysis relies on multiple-free data. We present the theory and show promising preliminary results using stylized synthetic examples.

Chapter 2

Fast imaging with surface-related multiples by sparse inversion

2.1 Summary

In marine exploration seismology, surface-related multiples are usually treated as noise mainly because subsequent processing steps, such as migration velocity analysis and imaging, require multiple-free data. Failure to remove these wavefield components from the data may lead to erroneous estimates for migration velocity or result in strong coherent artifacts that interfere with the imaged reflectors. However, multiples can carry complementary information compared to primaries, as they interact with the free surface and are therefore exposed more to the subsurface. Recent work has shown that when processed correctly multiples can improve seismic illumination. Given a sufficiently accurate background velocity model and an estimate for the source signature, we propose a new and computationally efficient two-way wave-equation based linearized inversion procedure that produces accurate images of the subsurface from the total upgoing wavefield including surface-related multiples. Modelling of the surface-related multiples in the proposed method derives from the well-known surface-related multiple elimination method. We incur a minimal overhead from incorporating the multiples by having the wave-equation solver carry out the multiple

A version of this chapter has been published in Geophysical Journal Internation, 2015, vol 201, pp. 304-317

predictions via the inclusion of an areal source instead of expensive dense matrix-matrix multiplications. By using subsampling techniques, we obtain high-quality true-amplitude least-squares migrated images at computational costs of roughly a single reverse-time migration with all the data. These images are virtually free of coherent artifacts from multiples. Proper inversion of the multiples would be computationally infeasible without using these techniques that significantly brings down the cost. By promoting sparsity in the curvelet domain and using rerandomization, out method gains improved robustness to errors in the background velocity model, and errors incurred in the linearization of the wave-equation with respect to the model. We demonstrate the superior performance of the proposed method compared to the conventional reverse-time migration using realistic synthetic examples.

2.2 Introduction

While reverse-time migration (RTM) has been very successful, its application to data sets with strong surface-related multiples is challenging because this imaging technology generally hinges on the single-scattering approximation. Ignoring these multiples can lead to erroneous velocity models and images that are prone to coherent artifacts that can lead to misguided interpretation. Throughout this paper, we discuss surface-related multiples unless specified otherwise.

To overcome this problem, industry has been adopting workflows that are designed to remove surface-related multiples prior to velocity analysis and migration. Early methods employed periodicity of multiples, which lead to trace-based predictive deconvolution (Peacock and Treitel, 1969), while later multi-trace techniques, such as NMO and the Radon transform, were used to separate primaries and multiples based on kinematic differences in the data space (Hampson, 1986; Foster and Mosher, 1992).

Unfortunately, these above techniques fail in cases where primaries and multiples overlap and where discrimination based on periodicity and kinematics alone is insufficient. In these cases, wave-equation based methods offer a viable, although computationally expensive alternative. Amongst these wave-equation based methods, we count Surface-Related Multiple Elimination (SRME) (Verschuur et al., 1992; Guitton and Verschuur, 2004; Herrmann et al., 2008c; Wang et al., 2008), model-based multiple prediction methods as reported by Wiggins (1988) and Lu et al. (1999), the inverse scattering method by Weglein et al. (1997), and wave-equation based move-out discrimination by Nemeth et al. (2000) and Sava and Guitton (2005).

An alternative to multiple removal is to actively use them to benefit from the extra information they carry, and to altogether avoid the prediction-subtraction de-multiple paradigm that may hurt the primaries (e.g., SRME). In data space processing, we count Estimation of Primaries by Sparse Inversion (EPSI), which inverts the SRME relation by identifying the surface-free dipole Green's function and the source wavelet as unknowns (van Groenestijn and Verschuur, 2009b; Lin and Herrmann, 2013). Compared to the prediction-subtraction paradigm of SRME where multiples and therefore the information they carry are removed, EPSI maps the surface-related multiples back to the primary impulse response via multi-dimensional deconvolutions. As a result, primaries are better preserved in EPSI (see van Groenestijn and Verschuur, 2009b, Figure (4) for example). However, this progress comes at a price. Its iterative inversion procedure involves significantly more computational work (e.g., multiplications of dense matrices, the computational complexity of which is $\mathcal{O}(n^3)$ for $n \times n$ matrices) than SRME. This may partially explain the relatively slow uptake of EPSI by industry for 3D seismic, where the computation can be prohibitively expensive. In model space processing such as imaging, many works have shown that surface-related multiples are useful signals and can increase illumination coverage when used properly (see examples in Berkhout, 1993; Guitton, 2002; Verschuur, 2011; Long et al., 2013). Among these works, the first school of methods treat the downgoing receiver wavefield as the generalized areal source for the surface-related multiples (Berkhout, 1993; Guitton, 2002; Shan, 2003; Schuster et al., 2004; Berkhout and Verschuur, 2006; Muijs et al., 2007; He et al., 2007; Whitmore et al., 2010; Liu et al., 2011; Verschuur, 2011; Lu et al., 2011; Long et al., 2013; Zuberi and Alkhalifah, 2013, and counting). These methods can be derived from the same multiple prediction principle as in SRME, by identifying the surface-related multiples to be the multi-dimensional convolution of the Green's function and the downgoing receiver wavefield. Most of these methods apply the cross-correlation imaging condition to image these multiples, which results in coherent artifacts in the image (a more detailed explanation can be found in Muijs et al., 2007; Liu et al., 2011). Others propose to use the deconvolutional imaging condition to reduce these artifacts (Shan, 2003; Muijs et al., 2007; Whitmore et al., 2010), but with limited success when complex geology is present (see Poole et al., 2010, and Chapter 4 for a detailed explanation). Methods exploiting least-squares migration to achieve this goal have also emerged in the literature (He and Schuster, 2003; Lin et al., 2010a; Verschuur, 2011; Tu and Herrmann, 2012b; Wong et al., 2012, 2014). However, wide adoption of these iterative methods is largely hindered by the expensive simulation

cost (in terms of the number of wave-equation solves). The second school of imaging-with-multiple methods relies on including a free surface (and / or other reflecting boundaries) in the background model (e.g., Reiter et al., 1991; Brown and Guitton, 2005; Wong et al., 2012). These methods do not fit into current single-scattering based migration paradigm, and are less discussed in the literature. Nevertheless, such a method that accounts for different orders of surface-related multiples still relies on a proper iterative inversion procedure (for example Brown and Guitton, 2005; Wong et al., 2012) and therefore also faces the challenge of expensive computation.

2.2.1 Contributions of this work

Inspired by the computationally expensive but successful SRME-based iterative inversion approaches (e.g., EPSI and the least-squares type imaging-with-multiples methods), we present a new and computationally efficient inversion method to image data with surface-related multiples. As far as we know, our approach is the first instance where primaries and all orders of surface-related multiples are imaged jointly using a two-way wave-equation based linearized inversion method in a computationally efficient manner. As we will show, the method eliminates dense, and therefore prohibitively expensive, matrix-matrix multiplications by solving wave-equations w.r.t. an areal-source term that contains the total downgoing wavefield. By incorporating this areal source into recently proposed compressive imaging (Herrmann and Li, 2012), we arrive at a formulation that creates high fidelity images at low computational costs. We control the coherent imaging artifacts from multiples and the computational cost by promoting curvelet-domain sparsity in the model space.

Without loss of generality, we limit our discussion to a single inversion where the velocity model is not updated and where obtaining high-fidelity artifact-free images is our main emphasis. However, our proposed method can be applied recursively, which would lead to a non-linear migration type of approach as presented by Métivier (2011); Métivier et al. (2011). When the velocity model is sufficiently accurate, this will also take care of internal multiples.

2.2.2 Paper outline

We organize the paper as follows: first, we introduce the matrix equation that establishes the physical relationship between the up- and downgoing wavefields at the free surface. Next, we combine this relationship with lin-

earized forward modelling using the solution operator of the time-harmonic Helmholtz system. This combination enables the use of all the techniques discussed in this paper to bring down the computation. With the linearized forward model, we introduce migration with multiples and sparsity-promoting imaging with multiples, using compressive imaging techniques to bring down the cost of our method to roughly equal that of a conventional RTM. We conclude by evaluating our algorithm using a stylized example and a realistic synthetic case study that demonstrate the superior performance of the proposed method compared to conventional RTM.

2.3 Imaging with multiples

Because we want to image the surface-related multiples, we start our treatise by first introducing the time-harmonic relationship between the up- and downgoing wavefields described by the SRME relation, and then propose to combine this relation with wave-equation based linearized inversion. This will allow us to directly relate the total data, including surface-related multiples, to the perturbations of the subsurface model.

2.3.1 Physics of the free surface

The SRME relation explicitly reveals the role of the surface-free Green's function that relates the total up- and downgoing pressure wavefields at the free surface. Using monochromatic matrices, it can be expressed as (Verschuur et al., 1992):

$$\mathbf{P}_i = \mathbf{X}_i(\mathbf{S}_i + \mathbf{R}_i\mathbf{P}_i), \tag{2.1}$$

where subscript $i = 1, \cdots, n_f$ indexes the frequencies where n_f is the total number of discretized frequencies, $\mathbf{P}_i$ is the discretized total upgoing receiver wavefield at the free surface, $\mathbf{X}_i$ is the surface-free dipole Green's function, and $\mathbf{S}_i$ is the downgoing point-source wavefield. The operator $\mathbf{R}_i$ models the reflectivity at the free surface, and as a result, $\mathbf{R}_i\mathbf{P}_i$ is the *down*going receiver wavefield at the surface that acts as a generalized "areal-source" wavefield for the surface-related multiples. Throughout this paper, we approximate the surface reflectivity to be -1, i.e., the action of $\mathbf{R}_i$ simply flips the sign of $\mathbf{P}_i$. The above relation reveals that the total upgoing wavefield can be represented by a frequency-domain (dense) matrix-matrix product of the surface-free dipole Green's function and the total downgoing wavefield (including the downgoing point-source wavefield and the downgoing receiver

wavefield). Note that because direct waves do not reflect at the surface, they are excluded from both the data $\mathbf{P}_i$ and the dipole Green's function $\mathbf{X}_i$ (Verschuur, 1991; Dragoset and Jeričević, 1998). Internal multiples are part of the primary response (i.e., $\mathbf{X}_i\mathbf{S}_i$) in the SRME relation. To obtain the upgoing wavefield at the surface, receiver ghosts (i.e., the downgoing data) need to be removed (source ghosts should be kept untouched) (Verschuur, 1991; Verschuur et al., 1992; Dragoset and Jeričević, 1998), and then the upgoing data need to be extrapolated from the receiver level to the free surface.

In Equation (2.1), the monochromatic upgoing wavefield is represented as a complex-valued matrix, $\mathbf{P}_i \in \mathbb{C}^{n_r \times n_s}$, with n_s common-shot gathers in its columns and n_r common-receiver gathers in its rows. The dipole Green's function $\mathbf{X}_i$ is organized in the same way. Equation (2.1) assumes fixed-spread acquisition with co-located sources and receivers (i.e., $n_s = n_r$), and therefore the point-source wavefield $\mathbf{S}_i \in \mathbb{C}^{n_r \times n_s}$ can also be organized in this way so that the addition of $\mathbf{S}_i$ and $\mathbf{R}_i\mathbf{P}_i$ makes physical sense. With these assumptions, the matrix product in the above expression corresponds to multi-dimensional convolutions consisting of a temporal convolution along the time dimension and a spatial convolution along the source / receiver dimension. For marine streamer acquisitions, we use reciprocity on-the-fly to mimic fixed-spread acquisition, which leads to minor computational overhead. Equation (2.1) describes 2D seismic lines but it can be extended to 3D seismic data where frequency slices can also be organized into matrices (Verschuur et al., 1992; Dragoset et al., 2010).

Here are more details about $\mathbf{X}_i$ and $\mathbf{S}_i$. In the continuous case, the dipole Green's function is the vertical derivative of the monopole Green's function, which can be alternatively represented using the particle-velocity wavefield (again see Fokkema and van den Berg, 1993, Equation (12.25)). In the discretized case, we approximate $\mathbf{X}_i$ by the response of the subsurface due to the injection of a (closely placed) pair of monopole impulsive sources (with the opposite sign) above and below the free surface, properly scaled to include the subsurface response due to source ghost. The point-source wavefield $\mathbf{S}_i$ is a diagonal matrix, with the diagonal entries being the Fourier spectrum of the source wavelet at a given frequency, i.e., $\mathbf{S}_i = w_i\mathbf{I}$ with w_i the spectrum of the source wavelet at the i^{th} frequency, and $\mathbf{I}$ the identity matrix (Verschuur et al., 1992).

While the Equation (2.1) has been responsible for major breakthroughs in the mitigation of surface-related multiples, including SRME (Verschuur et al., 1992) and EPSI (van Groenestijn and Verschuur, 2009a; Lin and Herrmann, 2013), it poses the following challenges: *(i)* it relies on dense matrix-matrix multiplications that are computationally expensive in terms

of CPU time, memory requirement, and computer input / output, and *(ii)* these dense matrix products require full data sampling, rendering expensive on-the-fly data interpolations necessary to compute these matrix products. While low-rank approximation techniques can be used to alleviate the cost of performing matrix products (Jumah and Herrmann, 2014), we propose to altogether avoid these dense matrix products by using the wave-equation solvers to implicitly carry them out.

2.3.2 Modelling the free surface via areal sources

To arrive at a formulation that is conducive to imaging data with surface-related multiples, we combine the physics of the free surface, described in the previous section, with the wave physics of the Earth interior, i.e., we incorporate the free surface as modelled in Equation (2.1) into the solution of the wave equation. As imaging is based on a linearization of the wave-equation at a smooth background model $\mathbf{m}_0$, we first approximate the direct-wave-free upgoing dipole Green's function $\mathbf{X}_i$ via Born modelling, assuming that the high-order reflections are weak. This approximation can be expressed as

$$\begin{aligned}\mathbf{X}_i &\approx \omega^2 \mathbf{D}_r \mathbf{H}_i[\mathbf{m}_0]^{-1} \mathbf{H}_i[\mathbf{m}_0]^{-1} \mathrm{diag}(\delta\mathbf{m}) \mathbf{D}_s^* \mathbf{I} \\ &\doteq \nabla\widehat{\mathcal{F}_i}[\mathbf{m}_0, \delta\mathbf{m}](\mathbf{D}_s^* \mathbf{I}) \\ &\doteq \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}; \mathbf{I}],\end{aligned} \tag{2.2}$$

where $\nabla\mathcal{F}_i$ denotes linearized forward modelling at the background model $\mathbf{m}_0$, which is linear in both the model perturbations $\delta\mathbf{m}$ and the source $\mathbf{I}$ (denoting the impulsive source array for the Green's function). We use the symbol $\approx$ to indicate that by linearized modelling, the higher-order scatterings are ignored. The symbol "$\doteq$" means "defined as", and the notation $\nabla\widehat{\mathcal{F}_i}$ in the second line denotes the same forward modelling, but is used in a way that explicitly shows the role of the source wavefield. The matrix $\mathbf{D}_r$ is a detection operator that collects data at receiver locations, and $\mathbf{D}_s^*$ injects the source wavefield. Again for the dipole-source effect, a pair of monopole sources is injected at each physical source location (with a positive sign) and the mirrored location w.r.t the free surface (with a negative sign). The matrix $\mathbf{H}_i$ represents the discretized monochromatic Helmholtz operator with absorbing boundaries (including at the surface, as the Green's function is surface free, i.e., it does not include a reflecting surface). The Helmholtz operator is discretized using the finite difference method with a 9-point stencil (Jo et al., 1996). Assuming constant density, the operator

can be parameterized by the square of the slowness (s^2/m^2) discretized in $\mathbf{m}$.

To incorporate surface-related multiples, a linearized modelling of the total upgoing wavefield can be derived by replacing $\mathbf{X}_i$ in Equations (2.1) by Equation (2.2), yielding:

$$\mathbf{P}_i \approx \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}; \mathbf{I}](\mathbf{S}_i - \mathbf{P}_i). \tag{2.3}$$

Compared to Equation (2.1), this linearized modelling involves the same dense matrix-matrix multiplications to predict the multiples, plus we need to simulate the full Green's function $\mathbf{X}_i$ by solving wave-equations for all (sequential) source experiments. To avoid the expensive dense matrix products, we simplify the forward modelling using the associativity of the matrix product, i.e., $\mathbf{ABC} = \mathbf{A}(\mathbf{BC})$, and arrive at the following identity:

$$\begin{aligned}
\mathbf{P}_i &\approx \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}; \mathbf{I}](\mathbf{S}_i - \mathbf{P}_i) \\
&= \nabla\widehat{\mathcal{F}_i}[\mathbf{m}_0, \delta\mathbf{m}](\mathbf{D}_s^*\mathbf{I})(\mathbf{S}_i - \mathbf{P}_i) \\
&= \nabla\widehat{\mathcal{F}_i}[\mathbf{m}_0, \delta\mathbf{m}](\mathbf{D}_s^*(\mathbf{S}_i - \mathbf{P}_i)) \\
&= \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}; \mathbf{S}_i - \mathbf{P}_i].
\end{aligned} \tag{2.4}$$

This simplified expression forms the basis of the proposed fast linearized inversion procedure to image data with surface-related multiples. The seemingly trivial algebraic manipulations in Equation (2.4) have profound implications on the computations. First, the expensive matrix-matrix products between the Green's function and the total downgoing wavefield (the second line) are turned into wave-equations solves with the same downgoing wavefield injected as the source (the third line). This major improvement goes at the expense of working with a more complicated areal source, which may affect the performance of certain modelling engines. We expect this effect to be minor compared with the computational gain. Second, the dense source sampling required by the SRME relation can now be relaxed in seismic imaging: compared with the data-space inversion procedure, namely EPSI (van Groenestijn and Verschuur, 2009b; Lin and Herrmann, 2013), which identifies the entire Green's function $\mathbf{X}_i$ in Equation (2.1) as the unknown, we optimize in the model space and therefore have much less unknowns in the model parameters $\mathbf{m}$. This significant reduction in the number of unknowns leaves room for us to use less source experiments to recover the model parameters. Note that in this case the subsampled sources should still co-locate with certain receivers, i.e., the underlying source and receiver grid should

co-locate but there is not necessarily a source deployed at each receiver location. Third, this simplified expression enables us to reduce the number of wave-equation solves by subsampling the source experiments (note that one source experiment corresponds to one column of the data matrices $\mathbf{P}_i$ and $\mathbf{S}_i - \mathbf{P}_i$, and by source subsampling, we reduce the number of columns of these data matrices) (Herrmann and Li, 2012). According to the second line of Equation (2.4), we still need to solve wave-equations for the fully-sampled sequential sources (i.e., the full source array $\mathbf{I}$ with the number of columns equal to the number of rows) to obtain the full Green's function $\mathbf{X}_i$ to perform matrix-matrix product, i.e., the subsampling does not bring down the simulation cost. According to the third line, we only need to solve wave-equations for the subsampled source wavefield, which has less number of columns than the full source array $\mathbf{I}$ in the second line. Without the simplified expression by Equation (2.4), significant reduction in simulation cost would be impossible.

To facilitate further discussion and problem formulation, we cast the imaging problem in the canonical form of solving a linear system of equations, i.e., $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is a linear operator and $\mathbf{x}$ and $\mathbf{b}$ are vectors. Throughout this paper, lower case symbols refer to vectors and upper-case symbols to matrices or operators. By vectorizing the wavefield $\mathbf{P}_i$ and rewriting the forward modelling in Equation (2.4) in a linear operator form that explicitly reveals the model perturbations as the unknown, we have the following monochromatic expression

$$\begin{aligned} \mathbf{p}_i &= \text{vec}(\mathbf{P}_i) \\ &\approx \text{vec}(\nabla \mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}; \mathbf{S}_i - \mathbf{P}_i]) \\ &\doteq \nabla \mathbf{F}_i[\mathbf{m}_0; \mathbf{S}_i - \mathbf{P}_i]\delta\mathbf{m}, \end{aligned} \tag{2.5}$$

where "vec" means to vectorize a matrix, and $\nabla\mathbf{F}_i$ is the linearized demigration (i.e., Born modelling) operator. The above expression for multiple frequencies can be obtained by stacking over the rows (i.e., vertical concatenation along the columns). This expression relates the medium perturbations $\delta\mathbf{m}$ to the total upgoing wavefield in a linear manner. Remember that these multiples are modelled by injecting the downgoing receiver wavefield as an areal source, rather than by placing a free surface in the background model. With this linearized expression in place, we will now show how conventional RTM (no explicit free-surface in the background model) fails to correctly image the surface-related multiples, which motivates us to exploit the least-squares inversion approach with sparsity promotion.

2.3.3 Migration with multiples by cross-correlation

Imaging with multiples by applying the cross-correlation imaging condition can be mathematically interpreted as applying the adjoint of the linearized demigration operator to the total upgoing data. In our case, we use two-way propagators and this amounts to the reverse-time migration (Baysal et al., 1983; Lailly, 1983). Using Equation (2.5), we obtain the RTM image after summing over all frequencies:

$$\widetilde{\delta\mathbf{m}} = \sum_{i=1}^{n_f} \nabla\mathbf{F}_i^*[\mathbf{m}_0; \mathbf{S}_i - \mathbf{P}_i]\mathbf{p}_i \tag{2.6}$$

where $\widetilde{\delta\mathbf{m}}$ denotes the RTM image, and superscript $*$ denotes the adjoint. As there are primaries and different orders of surface-related multiples in the total up- and downgoing wavefields, cross-correlation between acausal pairs of up- and downgoing wavefields will lead to artifacts in the image. A causal pair of surface reflections comprises the interaction between the k^{th}-order downgoing and the $k+1^{th}$ order upgoing multiples for $k = 0, 1, 2, \cdots$ and with primaries regarded as the 0^{th}-order multiples. By acausal, we mean any pair other than the causal pairs (see detailed explanations by Muijs et al., 2007; Liu et al., 2011).

We use a relatively simple velocity Earth model to illustrate these artifacts. The background model and the model perturbations are shown in Figure 2.1(a) and 2.1(b). The model dimensions are 1250 m deep and 2235 m wide with a five-meter grid distance in both dimensions. We make linearized data according to Equation (2.5) assuming a known source wavelet spectrum $\mathbf{w}$ for which we use a Ricker wavelet, and known areal-source wavefields $-\mathbf{P}_i$ that we simulate using finite-difference modelling with a free-surface boundary condition. The frequency band of the source wavelet goes up to 60 Hz. The total recording time is 2.044 s with 4 ms sampling interval, which corresponds to a total number of 122 non-zero frequencies in the specified frequency range. There are 150 co-located sources and receivers with 15 m spacing.

We show two RTM images here. In both cases, the input data contain surface-related multiples. The first image is based on the conventional single-scattering forward model (i.e., the demigration operator only models the primaries). The result is shown in Figure 2.1(c). We can clearly see the phantom reflectors imaged from multiples (indicated by the arrows in the figure). This is understandable since the multiples are not taken into account by the modelling operator. The second image is by Equation (2.6), which

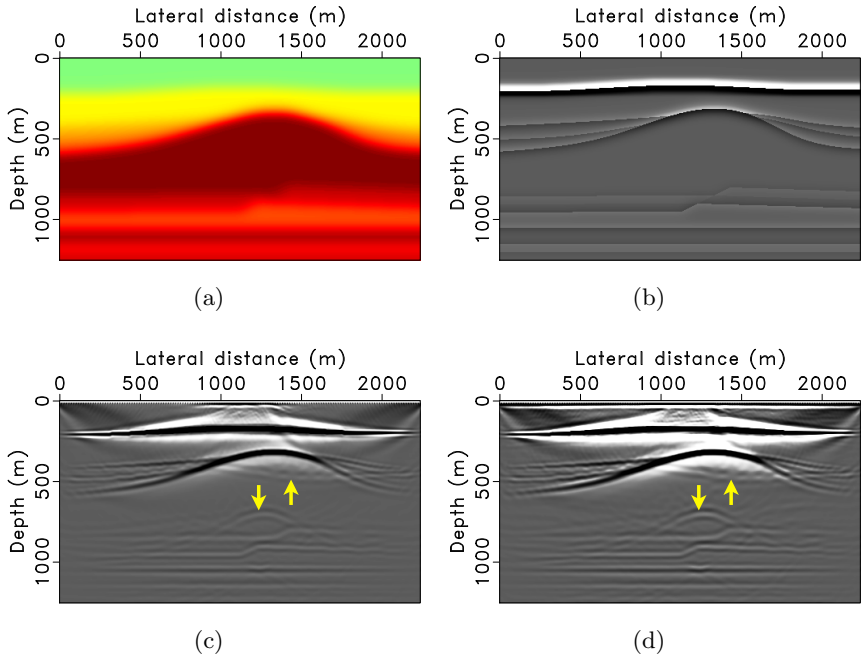


Figure 2.1: The synthetic salt dome model and two RTM images. **(a)** The background model. **(b)** Model perturbation. **(c)** RTM of total data but only with the primary imaging operator. **(d)** RTM of total data by including the total downgoing wavefield in the source wavefield.

models the surface-related multiples by using the total downgoing wavefield as the areal source wavefield. This image is shown in Figure 2.1(d). Although the multiples are modelled by the modelling operator, we still observe severe coherent artifacts (again indicated by the arrows in the figure) caused by acausal cross-correlations between acausal pairs of up- and downgoing wavefields that contain multiples. To overcome these harmful artifacts, we propose to properly invert the linearized modelling operator instead of applying the adjoint.

2.4 Fast sparsity-promoting imaging with multiples

The computational cost of inverting Equation (2.5) for all the frequencies by the iterative least-squares migration is excessive, because each iteration involves one reverse-time migration and demigration (Nemeth et al., 1999). Without storing the wavefield, this amounts to solving four wave equations for each monochromatic source experiment. Therefore the total number of wave-equation solves roughly equals to $4 \times n_f \times n_s \times n_{iter}$, with n_{iter} the number of iterations of the inversion. We reduce this excessive simulation cost by subsampling the monochromatic source experiments and promoting curvelet-domain sparsity of the model parameters (Herrmann et al., 2009; Herrmann and Li, 2012).

We use synthetic examples to help readers understand the different components of the method. In these examples, we use the same model as in Figure 2.1. To obtain a baseline image, we first run an inversion of all the data for 60 iterations. For a fair comparison with the following compressive imaging examples, we cast the inversion as the same Basis Pursuit De-Noise (BPDN) problem (Chen et al., 2001; van den Berg and Friedlander, 2008), and use the same SPGℓ_1 solver (van den Berg and Friedlander, 2008). Details about the BPDN problem and the SPGℓ_1 solver will follow. The result is shown in Figure 2.2(a), which is as expected more or less artifact free as all surface-related multiples are correctly mapped to locations of the true reflectors. The coherent artifacts in Figure 2.1(d) are largely removed by adopting the inversion approach. However, the computation for Figure 2.2(a) is expensive, involving roughly 4.4 million wave-equation solves.

2.4.1 Subsampling monochromatic source experiments

To reduce the number of wave-equation solves, which is proportional to the number of monochromatic source experiments, we follow Herrmann et al. (2009) and subsample over both frequencies and source experiments. First, we only use a randomized subset of frequencies, denoted by Ω. Second, we replace the monochromatic forward modelling in Equation (2.5) by

$$\begin{aligned} \underline{\mathbf{p}}_i &= \mathrm{vec}(\mathbf{P}_i\mathbf{E}) \\ &\approx \nabla\mathbf{F}_i[\mathbf{m}_0;(\mathbf{S}_i-\mathbf{P}_i)\mathbf{E}]\delta\mathbf{m} \\ &= \nabla\mathbf{F}_i[\mathbf{m}_0;\underline{\mathbf{S}}_i-\underline{\mathbf{P}}_i]\delta\mathbf{m}. \end{aligned} \tag{2.7}$$

In this expression, the underlined quantities refer to subsampled source and receiver wavefields by randomized source superimpositions (i.e., forming simultaneous sources) via the action of a tall (i.e., more number of rows than columns) source-encoding matrix $\mathbf{E}$ on the right, which has Gaussian-distributed randomized entries. The number of simultaneous sources is defined as the number of columns of the encoding matrix $\mathbf{E}$. Since there are now fewer source experiments, we have fewer wave-equations to solve. However, this speed-up goes at the expense of introducing source crosstalks into an RTM image, caused by the randomized source superimposition (see Romero et al. (2000), Figure (1), and Herrmann and Li (2012), Figure (1b)). We also show such an example here. Instead of using all 150 sequential sources, we use 10 simultaneous sources. For this example we use all the frequencies. We then compute the RTM image of the 10 simultaneous shot-gathers and show it in Figure 2.2(b). In this image we can identify the noisy source crosstalks as well as the coherent acausal artifacts caused by the surface-related multiples. We propose to remove these artifacts using ideas from sparse recovery.

2.4.2 Removal of source crosstalks by promoting sparsity

Following sparse-recovery and randomized sampling ideas from Compressive Sensing (CS) (Candès et al., 2006b; Donoho, 2006), we remove these artifacts by promoting sparsity of the model perturbations using the ℓ_1-norm. To maximally exploit the sparsity, we incorporate the curvelet transform that affords a sparse representation of the model perturbations (Candès and Donoho, 2004; Herrmann et al., 2008a,b; Herrmann and Hennenfent, 2008; Neelamani et al., 2010). As a result, we aim to solve the following Basis Pursuit De-Noise (BPDN) problem:

$$\begin{aligned} &\underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1 \\ &\text{subject to } \sum_{i\in\Omega} \|\underline{\mathbf{p}}_i - \nabla\mathbf{F}_i[\mathbf{m}_0; \underline{\mathbf{S}}_i - \underline{\mathbf{P}}_i]\mathbf{C}^*\mathbf{x}\|_2^2 \leq \sigma^2, \end{aligned} \tag{2.8}$$

where $\mathbf{C}^*$ represent the curvelet synthesis operator, i.e., its action is to apply the inverse curvelet transform (Candès et al., 2006a), and $\mathbf{x}$ is the curvelet coefficient vector for the image (i.e., $\delta\mathbf{m} = \mathbf{C}^*\mathbf{x}$). This optimization program finds, among all curvelet coefficient vectors, the one with the smallest ℓ_1-norm subject to the fact that its inverse curvelet transform yields demigrated (simultaneous) data that explain the input (simultaneous) data within a user-specified σ that allows for noise in the data and modelling

errors (e.g., from ignoring the higher-order reflections during linearization). We solve the above BPDN problem using the SPGℓ_1 solver, which, instead of directly dealing with the BPDN formulation, solves a series of underlying Least Absolute Shrinkage and Selection Operator (LASSO) subproblems (Tibshirani, 1996). We refer to Appendix A.1 for a more detailed explanation of how SPGℓ_1 solves a general BPDN problem.

The CS theory states that solving such a convex sparsity program (cf. Equation (2.8)) can be used to invert underdetermined systems of equations as long as these systems behave approximately as Gaussian matrices, and $\delta\mathbf{m}$ permits a sufficiently sparse curvelet representation. The latter argument is made relatively easy by referring to the body of literature that has appeared on the compressibility of seismic images w.r.t. curvelets (Herrmann et al., 2008a,b). The first argument is a bit more difficult to make although we can argue that the combination of curvelets, randomized source superimposition, demigration, and free-surface-multiple generation leads to matrices that are well-mixed and not too far off from Gaussian matrices. However, the demigration operator has a null space caused by finite acquisition aperture and complex geologies, which may result in unresolvable steep dips and shadow zones (see e.g., Prucha and Biondi, 2002; Herrmann et al., 2008a). Therefore we cannot expect perfect recoveries.

Using the same setup as for Figure 2.2(b), i.e., 10 simultaneous sources, we obtain an inversion result after 60 SPGℓ_1 iterations. The result is shown in Figure 2.2(c). Compared to Figure 2.2(b), the source crosstalks, as well as the acausal artifacts associated with the multiples, are both removed in Figure 2.2(c) by adopting the above optimization formulation that promotes sparsity in the curvelet domain (cf. Equation (2.8)). Compare to Figure 2.2(a), we reduce the number of wave-equation solves by 15 folds.

While a 15$\times$ speed-up is encouraging, Figure 2.2(c) still involves roughly 8$\times$ the simulation cost of a conventional RTM of all the data. We need to bring down the cost even further to make the method computationally feasible for large data sets. However, the SPGℓ_1 solver converges too slowly if we subsample too much. We show an example where we only use two simultaneous sources and 15 frequencies. The image after 305 iterations is shown in Figure 2.2(d), where we still see lots of remnant noisy artifacts. As explained in recent work on compressive imaging and Approximate Message Passing (AMP), these artifacts are related to the correlation built up between the model iterate and the randomized subsets of monochromatic source experiments (Herrmann and Li, 2012; Herrmann, 2012; Montanari, 2010). We refer to Appendix A.2 for more details about the cause of these artifacts. We will now explain how we remove these artifacts.

2.4.3 Further acceleration by rerandomization

Motivated by recent findings in AMP (Donoho et al., 2009; Montanari, 2010), we proposed a simple rerandomization of the BPDN problem (cf. Equation (2.8)) to remove the remnant incoherent artifacts related to correlation build up shown in Figure 2.2(d). For our problem, this corresponds to selecting a new set of randomized source experiments and a new randomized frequency subset for each LASSO subproblem (Herrmann and Li, 2012; Tu and Herrmann, 2012a). This practice misses a strict theoretical justification but is close in spirit to other randomized algorithms that have let to fast solutions of overdetermined problems, e.g., the randomized (block) Kaczmarz algorithm (Strohmer and Vershynin, 2009; Needell and Tropp, 2014).

To verify the efficacy of rerandomization by comparing with Figure 2.2(d) where no rerandomization is used, we rerun the above example using the same setup (i.e., two simultaneous sources, 15 frequencies, 305 SPGℓ_1 iterations) with rerandomization. The result is shown in Figure 2.2(e), and we can see that the image quality has drastically improved compared to Figure 2.2(d). Compared with Figure 2.2(a), we achieve a 120$\times$ speed-up without notable compromise on the image quality. In fact, the simulation cost to obtain Figure 2.2(e) roughly equals to that of a single RTM of all the data, i.e., the rerandomization technique enables us to obtain a high-fidelity true-amplitude least-squares migrated image when the computational budget only allows for one RTM with all the data.

2.4.4 Sparsity promoting by ℓ_1 or ℓ_2

While the success of the speed-up really hinges on the combination of the two contributing factors—sparsity promoting and rerandomization—the role that the ℓ_1-norm plays in promoting curvelet-domain sparsity should not be understated.

To demonstrate the importance of using the ℓ_1-norm as a measurement of sparsity (Mallat, 2009), we solve a problem that is otherwise the same as Equation (2.8) except that we replace the ℓ_1-norm by the ℓ_2-norm in the objective function. Again we use two simultaneous sources, 15 frequencies, 305 iterations and rerandomization. The result is shown in Figure 2.2(f). We can see that it has more noise compared with Figure 2.2(e) where we use the ℓ_1-norm objective. Therefore the ℓ_1-norm should be used.

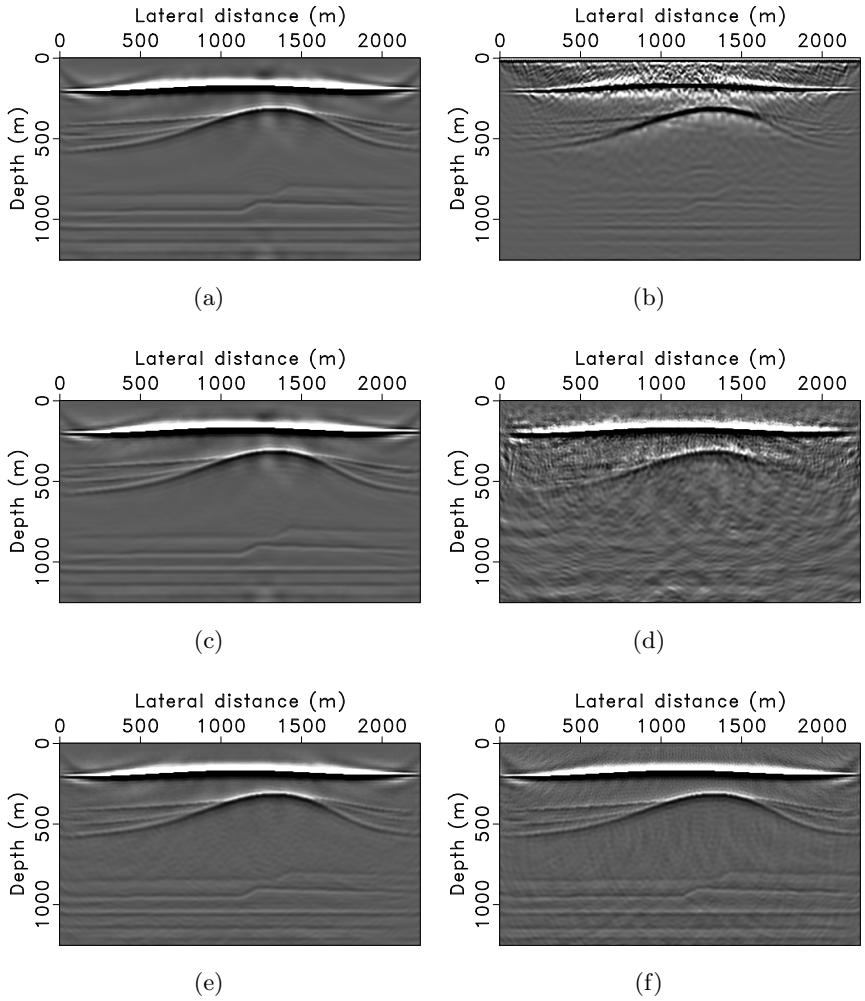


Figure 2.2: Examples using the salt dome model. **(a)** Inversion with all the data. **(b)** Migration with 10 simultaneous sources. **(c)** Inversion with 10 simultaneous sources. **(d)** Inversion with 2 simultaneous sources and 15 frequencies without rerandomization. **(e)** Same as **(d)** but with rerandomization. **(f)** Same as **(e)** but by ℓ_2 norm minimization.

2.4.5 Putting it all together

Now we have explained each component of the proposed method, we summarize and explain the algorithm in Algorithm (1), which is an adapted version of the compressive imaging algorithm by Herrmann and Li (2012).

Input and initialization [lines 1–4]

As we stated before, our inversion algorithm requires estimates of the source wavelet and the background velocity model $\mathbf{m}_0$. The solution vector $\mathbf{x}_0$ is initialized to be a zero vector. We select the tolerance parameter σ simply to be zero (i.e., we only stop when we reach the maximal number of iteration), and choose the number of simultaneous source experiments $n_s' \ll n_s$, the number of randomized frequencies $n_f' \ll n_f$ and the maximal number of iterations k_{max} based on our computational budget (Chapter 3). We use l to index the LASSO subproblems.

Main loop [lines 5–10]

Given the input and after the initialization, we enter into the main loop in which we solve the LASSO subproblems. For each subproblem, we draw new independent randomized subsets of frequencies Ω and source experiments by drawing new source encoding matrix $\mathbf{E}$ (line 6). Without this "redraw" step, the algorithm would be identical to the original SPGℓ_1. Following van den Berg and Friedlander (2008), we compute the sparsity constraint for the l-th LASSO subproblem τ_l using τ_{l-1} and σ using Newton's method on the Pareto curve (line 7), and solve the $\mathbf{LS}_\tau$ subproblem (line 8).

2.5 Performance analysis

While the stylized example clearly showed qualitative improvements in the sparse recovery with rerandomization, we would like to study the behavior of the solver more quantitatively. In particular, we are interested in studying the solution path and the model error as a function of the number of wave-equation solves, as it can give us insights into the way that rerandomization works to speed up the convergence. We are also interested in the recovery error as a function of the number of simultaneous sources, frequencies, and iterations, which will help us to choose the optimal inversion parameters for the proposed algorithm given a fixed computational budget.

Algorithm 1 Fast imaging with multiples

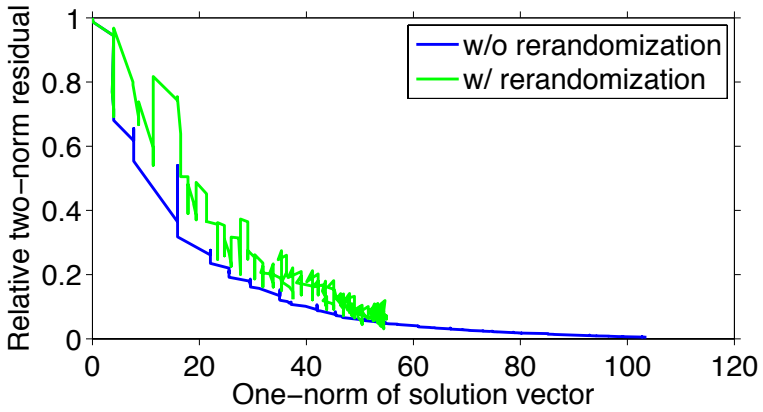
Input:
total upgoing wavefield $\mathbf{P}_i$, point-source wavefield $\mathbf{S}_i$,
background velocity model $\mathbf{m}_0$, tolerance $\sigma = 0$, iteration limit k_{max}
Initialization:
$k \leftarrow 0, l \leftarrow 0, \mathbf{x}_l \leftarrow \mathbf{0}$
while $k < k_{max}$ **do**
  $\Omega_l, \mathbf{E}_l \leftarrow$ new independent draw, $\underline{\mathbf{P}}_i = \mathbf{P}_i\mathbf{E}_l$, $\underline{\mathbf{S}}_i = \mathbf{S}_i\mathbf{E}_l$, $\underline{\mathbf{p}}_i = \mathrm{vec}(\underline{\mathbf{P}}_i)$
  $\tau_l \leftarrow$ determine from τ_{l-1} and σ by root finding on the Pareto curve
  $\mathbf{x}_l \leftarrow \begin{cases} \mathrm{argmin}_{\mathbf{x}} \sum_{i\in\Omega_l} \|\underline{\mathbf{p}}_i - \nabla\mathbf{F}_i[\mathbf{m}_0; \underline{\mathbf{S}}_i - \underline{\mathbf{P}}_i]\mathbf{C}^*\mathbf{x}\|_2^2 \\ \text{subject to } \|\mathbf{x}\|_1 \le \tau_l \end{cases}$ //warm start with $\mathbf{x}_{l-1}$, solved in k_l iterations
  $k \leftarrow k + k_l, l \leftarrow l + 1$
end while
Output: Model perturbation estimate $\delta\mathbf{m} = \mathbf{C}^*\mathbf{x}$

2.5.1 Solution path and model error

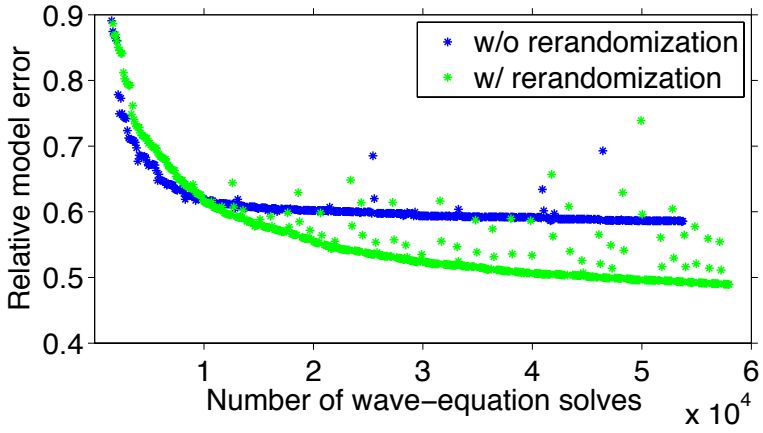
To understand the significant difference between optimizing with and without rerandomization (results in Figure 2.2(d) and 2.2(e)), we plot their respective solution path, i.e., the (relative) ℓ_2-norm of the data misfit as a function of the ℓ_1-norm of the solution vector. We also plot the decreases of the model errors (i.e., the relative ℓ_2-norm of the misfit between the model iterate and the true model perturbations) in the two cases as a function of the number of wave-equation solves.

Shown in Figure (2.3(a)), the solution path for the optimization with rerandomization behaves differently (expected, as we are solving a different problem with rerandomization) from that of the standard SPGℓ_1, where the ℓ_2-norm of the data residual decreased close to zero albeit slowly. Because the residual is larger for the case with rerandomization, one would normally expect a deterioration rather than an improvement in the image quality by using rerandomization. However, we can see in Figure (2.3(b)) that the model errors decay continuously with rerandomization that breaks down the correlation built up between the model iterate and the randomized subsets of monochromatic source experiments, but almost stall at an early stage of the optimization in the case of no rerandomization (i.e., no significant improvement by using more iterations).

Because the artifacts in Figure 2.2(d) with wrongly identified support lie close to the nullspace of the compressive modelling operator (as the blue



(a)



(b)

Figure 2.3: Decreases of the data misfit and model errors, with (in green) and without (in blue) rerandomization. **(a)** Decrease of the relative data misfit. **(b)** Decrease of the relative model errors.

curve decreased close to zero in Figure 2.3(a)), they are difficult to remove even after extended number of iterations. However, by occasionally drawing new compressive modelling operators with independent subsets of frequencies and source experiments, these artifacts can be gradually removed. This observation was also made by Herrmann and Li (2012), where using rerandomization greatly promotes the convergence in terms of model error.

2.5.2 Computational considerations

As we use wave-equation solvers to carry out the SRME type multi-dimensional convolutions implicitly, which also enables us to reduce the simulation cost by subsampling the monochromatic source experiments, the costs of our imaging with multiple algorithm are dominated by the number of wave-equation solves. This raises the question of identifying the best inversion strategy given a certain computational budget in terms of the number of wave-equation solves, which is proportional to the product of the number of monochromatic sources and the number of iterations, i.e., $n_f \times n_s \times n_{iter}$. In conventional inversion approaches, people usually work with all sources and frequencies, but can only afford a few iterations. With the proposed approach, we can work with significantly smaller data subsets and can therefore afford many more iterations because each iteration is now much cheaper. While both approaches are parallelizable over independent monochromatic source experiments in the data space or via domain decomposition in the model space, our approach has the advantage of partially removing the computer input / output bottleneck as the subsampling also reduces the data size.

Using the same synthetic model / data as in the previous section, we compare the performance of the proposed method with different combinations of randomized monochromatic source experiments and iterations. The signal-to-noise ratios (SNR) of the recoveries compared to the true model perturbation are shown in Table (2.1). From the table we can see that more iterations with cheaper per-iteration cost lead to better recoveries, which also offers flexibility in the design of parallel algorithms.

Another aspect is to understand the interplay between the number of simultaneous sources and frequencies once we fix the number of monochromatic source experiments. We compare five different setups and show the results in Table (2.2). We can see that these combinations lead to almost equally good recoveries.

2.6 Synthetic case study

To evaluate the performance of the proposed method for a more realistic setup, we verify our method with a more complex Earth model. The model we use here is cropped from the Sigsbee 2B model that contains abundant sedimentary layers. The Sigsbee 2B model is designed to contain a thin high-velocity ocean bottom layer to generate strong surface-related multiples (The SMAART JV, 2014). We reduced the thickness of the water column

No. of mono. sources.	15	30	60	120	240	480
Aspect ratio of $\nabla\mathbf{F}$	0.02	0.04	0.08	0.16	0.32	0.64
No. of iterations	610	305	152	76	39	20
SNR (dB)	6.5	6.2	5.8	5.1	4.2	3.4

Table 2.1: SNRs as a function of the number of monochromatic source experiments and the number of iterations.

No. of sim. sources	2	3	6	10	15
No. of frequencies	15	10	5	3	2
SNR (dB)	6.2	6.2	6.2	6.3	6.7

Table 2.2: SNRs as a function of the number of simultaneous sources and frequencies. The number of monochromatic sources as well as the number of iterations are fixed.

to allow for more orders of surface-related multiples. The modified true and background models are shown in Figure 2.4. The background model is obtained by smoothing the true model and is kinematically close to the true model. The grid spacing is 7.62 m. There are 261 co-located sources and receivers with 22.86 m spacing. We model the data using iWave with a free-surface boundary condition (Terentyev et al., 2014). We use a Ricker wavelet with 15 Hz peak frequency, and record data for 8.184 seconds. We simulate both the pressure and particle velocity wavefields to separate the upgoing and downgoing wavefields (Wapenaar, 1998). We then remove the direct arrivals by subtracting data modelled with the background model. Next, we extrapolate the upgoing wavefields to the surface level (see Verschuur et al., 1992, for more details). In the imaging procedure, we use our own frequency-domain modelling engine to mimic the discrepancy between the observed data and the modelled data in field seismic data processing (van Leeuwen and Herrmann, 2012a). For the same reason, we use forward modelling instead of linearized Born modelling to simulate the data. However, our proposed method is robust to this linearization error by virtue of using rerandomization (see Chapter 3 for more details).

For comparison, we first computed the RTM image of data with multiples with all the data according to Equation (2.6). The result is shown in Figure 2.5. Again we see the coherent acausal artifacts caused by surface-related

multiples, which are indicated by the arrow.

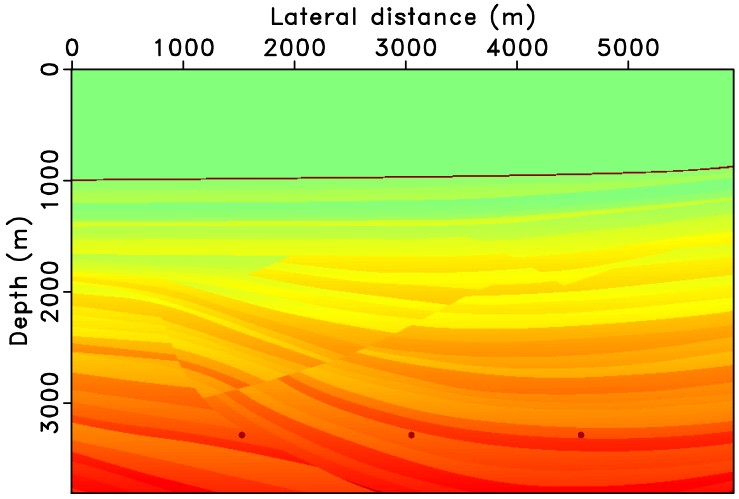
For the proposed method, we use 26 simultaneous sources from all 261 sequential sources (roughly 10× subsampling), 31 frequencies randomly chosen from all 311 discretized frequencies (roughly 10× subsampling), and run the inversion for 50 iterations. In this way the simulation cost remains comparable to a single RTM of all the data (e.g., Figure 2.5). After the inversion is finished, we perform a simple additional curvelet thresholding to remove the remaining incoherent noise in the image (see Herrmann et al., 2008a, for more details). We choose such a threshold level that the filtered-out noise does not contain noticeable coherent energy. The result is included in Figure 2.6(a). This thresholding procedure is also applied to the imaging results in the Discussion section. Again we obtain an image with minimal coherent acausal artifacts arising from the surface-related multiples by using the proposed inversion method. These weak remnant artifacts have the following sources: *(i)* we use different modelling engine for data simulation and inversion; *(ii)* the use of a pair of monopole sources to approximate the source-ghost effect is inexact; and *(iii)* the prediction of multiples is also inexact because of finite aperture. Despite these remnant artifacts, the resulting image is of higher spatial resolution compared to Figure 2.5, as we properly invert the demigration operator that includes the source wavelet.

To highlight the efficacy of the proposed method in reducing the coherent artifacts caused by the surface-related multiples, we run another inversion with the same setup, but ignore the presence of the surface-related multiples (i.e., the modelling operator only models primaries). The result is shown in Figure 2.6(b), where we can see strong coherent artifacts that mimic those in Figure 2.5.

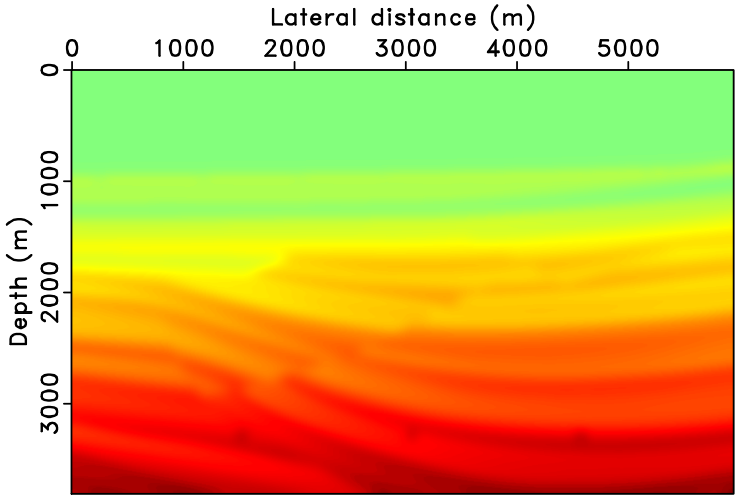
2.7 Discussion

2.7.1 The source wavelet

We assumed that we have access to the true source wavelet. In practice, however, source estimation is never a trivial task because of the amplitude / phase ambiguity in blind deconvolution (Ulrych and Sacchi, 2005, and Chapter 5). The EPSI algorithm successfully eliminates these ambiguities by using information from surface-related multiples (van Groenestijn and Verschuur, 2009b; Lin and Herrmann, 2013). However, it is computationally expensive because it uses an iterative algorithm with dense matrix-matrix multiplies.



(a)



(b)

Figure 2.4: The cropped Sigsbee 2B model. **(a)** The true model. **(b)** The background model.

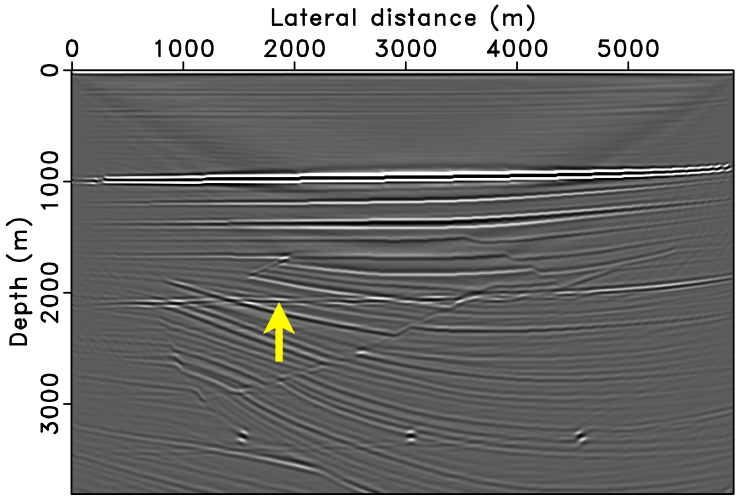
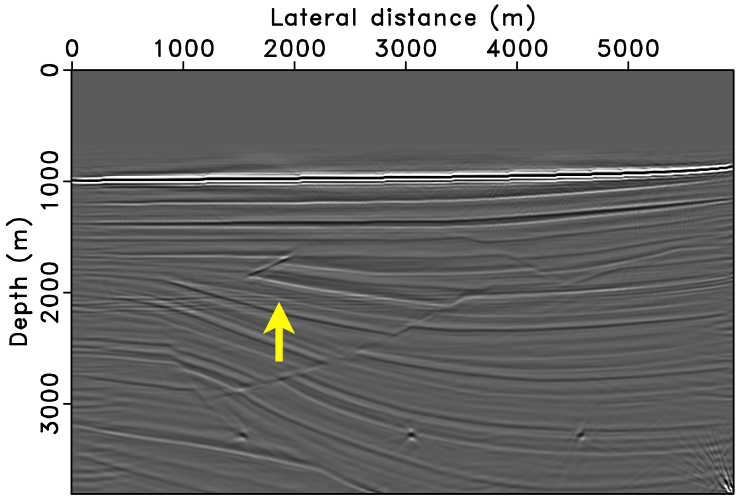


Figure 2.5: RTM image of the Sigsbee 2B model, using data containing multiples.

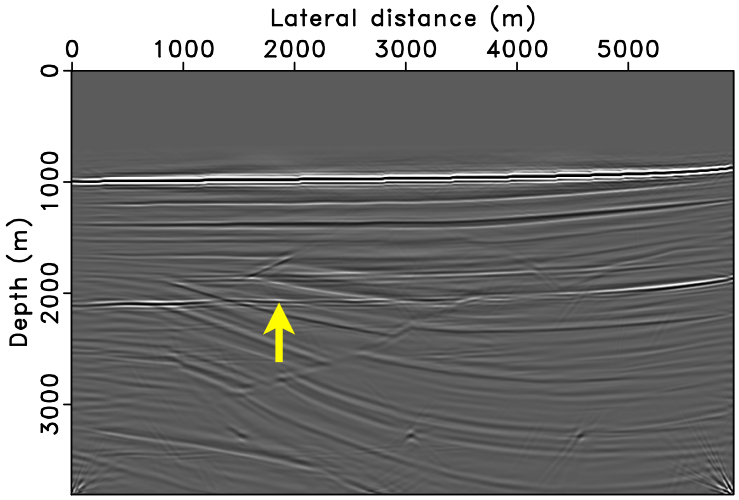
Alternatively, source signatures can be estimated during the imaging itself using an approach known as variable projection (Golub and Pereyra, 2003; Aravkin and van Leeuwen, 2012). Recent work by Aravkin et al. (2013b) and Tu et al. (2013) successfully applied this alternating optimization method to recover the image as well as an estimate for the source function. Figure 2.7(a) contains an example of this method obtained using the same inversion parameters as in the previous section, i.e., we use 26 simultaneous sources, 31 randomized frequencies, and run the inversion for 50 iterations. Compared to Figure 2.6(a), a 50% computational overhead is incurred by the variable-projection step within each iteration, as the primary and multiple wavefields have to be modelled separately to update the source estimate (see Chapter 5 for more details). This extra computation, which is about half of a single RTM of all the data, is still insignificant compared with a conventional least-squares migration of all the data. We report this source-estimation method in more details in Chapter 5.

2.7.2 Imaging with multiples only

Our proposed method can also be used in cases where a reliable primary / multiple separation is available, e.g. after a conventional multiple prediction step. In that situation, we can image multiples separately and ex-



(a)



(b)

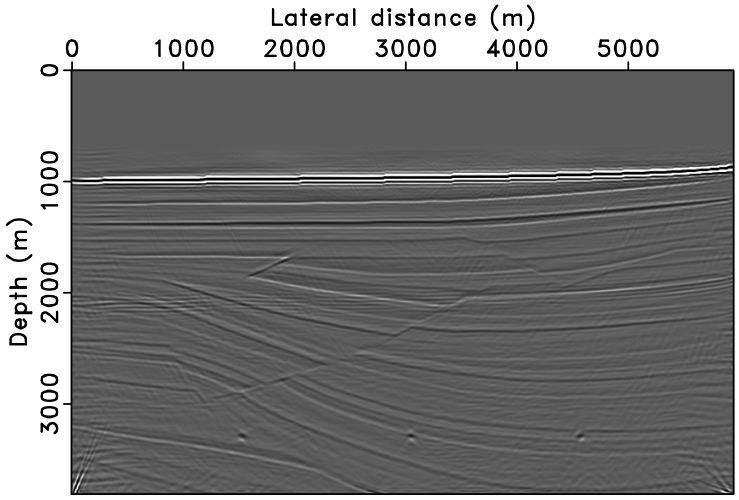
Figure 2.6: Inversion results of the Sigsbee 2B model. Both images are obtained after curvelet thresholding to remove remnant incoherent noise in the image. The arrows in **a**, **b** and in Figure 2.5 point to the same location of the model. **(a)** Fast inversion result using the proposed method. **(b)** Fast inversion result without accounting for the multiples.

ploit their extra illumination properties without explicit knowledge of the source wavelet. (However, multiple predictions by the SRME-type multi-dimensional convolutions typically need to apply the inverse of the wavelet.) As before, we obtain these images by including an areal source, which now consists of the downgoing receiver wavefield only. Compared to earlier work (e.g. Guitton, 2002; Muijs et al., 2007; Whitmore et al., 2010; Liu et al., 2011), our approach leads to a true inversion result while others merely apply the deconvolutional imaging condition or the cross-correlation imaging condition, which is known to lead to undesirable acausal artifacts. We use the same inversion parameters as for Figure 2.7(a), and the result is included in Figure 2.7(b).

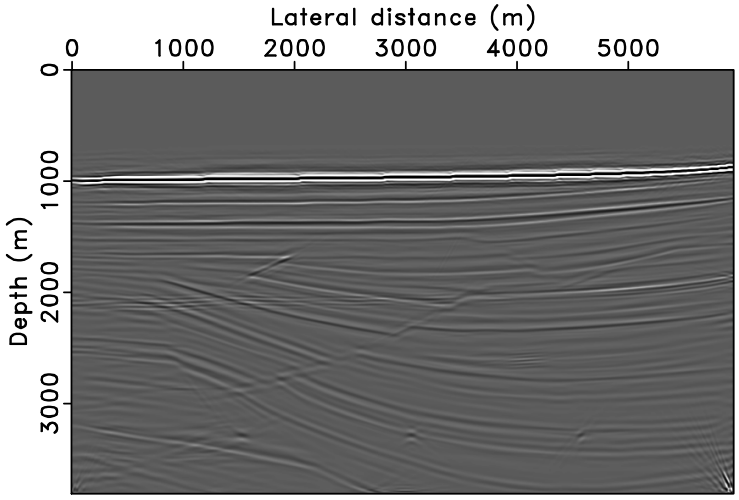
2.7.3 Sensitivity to the background velocity model

Our method hinges, as all other wave-equation based imaging procedures, on the availability of a smooth background velocity model that accurately captures the kinematics of reflected waves. We have also seen that incorporating surface-related multiples in the formulation calls for an iterative sparsity-promoting inversion procedure and the question arises how sensitive this more elaborate imaging procedure is to errors in the background velocity model. This question is particularly pertinent because migration velocity analysis tools also often suffer from surface-related multiples. Moreover, iterative inversion procedures such as (sparsity-promoting) least-squares migration may be more sensitive to velocity-model errors than conventional RTMs and it is not clear how the inclusion of surface-related multiples plays into this.

To address these concerns we conducted an experiment where we deliberately introduce errors in the velocity model that cumulate with depth (see Figure 2.8(a) and 2.8(b)). The results of this exercise are included in Figure 2.9(a) to 2.10(b) and show that for this example least-squares imaging with multiples is relatively well behaved with imaging results that degrade gracefully. As expected, imaging results with curvelet-domain sparsity promotion are more coherent by virtue of that fact that curvelets allow for small kinematic errors as reported in the literature (Herrmann et al., 2008c). The results without curvelet-domain sparsity were obtained by putting the sparsity-constraint parameter τ in each $\mathbf{LS}_\tau$ subproblem to infinity (it is essentially the block Kaczmarz method (see e.g. Needell and Tropp, 2014)), and eliminating the use of the curvelet transform. While more work is needed to fully understand the influence of the background velocity model on our iterative inversion algorithm, these results show that our method



(a)



(b)

Figure 2.7: Dealing with the unknown source signature. Both images are after curvelet thresholding. **(a)** Fast imaging of the total upgoing wavefield with source estimation. **(b)** Fast imaging of multiples only.

is relatively well-behaved w.r.t. possible errors in the background velocity model.

2.7.4 Internal multiples

The proposed work here addresses dominant surface-related multiples in a linearized inversion framework. Aside from surface-related multiples, there are also internal multiples in the acquired seismic data. When used correctly, internal multiples also increase illumination (Cavalca and Lailly, 2005; Malcolm et al., 2009). In cases where fine-scale heterogeneity exists, including internal multiples can be crucial to produce an accurate image (Delprat-Jannaud and Lailly, 2008).

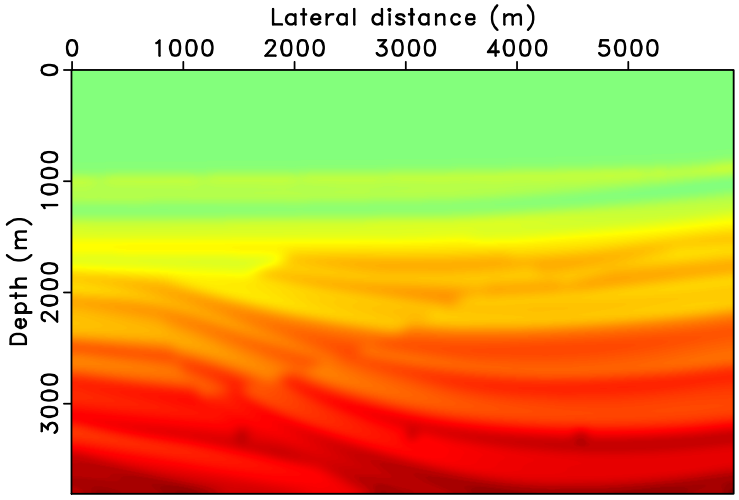
Our proposed approach is based on the same linearization as full-waveform inversion. In addition, our sparsity-promoting inversion can be viewed as a sparsity-constrained Gauss-Newton update (Herrmann and Li, 2012). Therefore, recursive application of the proposed method with updates on the velocity model can be construed as an instance of non-linear migration. As outlined by Métivier (2011); Métivier et al. (2011), this type of recursive inversion will account for internal multiples for a sufficiently accurate starting velocity model.

2.7.5 Other applications

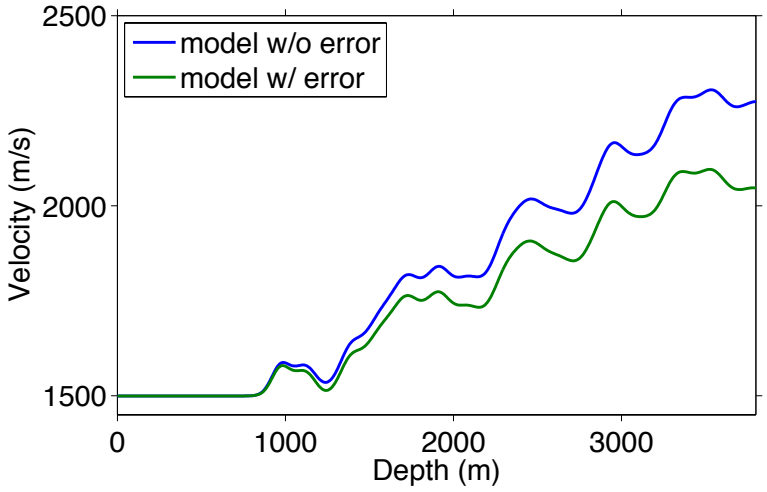
In essence, the proposed method entails multi-dimensional deconvolutions of wavefields combined with a linearized imaging procedure. While we specialized this approach to image with surface-related multiples, there is nothing that prevents the application of our proposed method to other areas as long as the receivers are sampled adequately. For instance, the method could relatively easily be adapted to imaging with receiver functions. In that case, the observed converted shear waves are deconvolved with the pressure waves to image the pressure-to-shear transmission coefficients (Ryberg and Weber, 2000; Shang et al., 2012). Another example would be the deconvolutional interferometric imaging as recently proposed by van der Neut and Herrmann (2013).

2.8 Conclusion

We proposed a new and computationally efficient method to image seismic data with surface-related multiples. Our contributions are threefold. First, our formulation combines the wave equation with surface-related-multiple

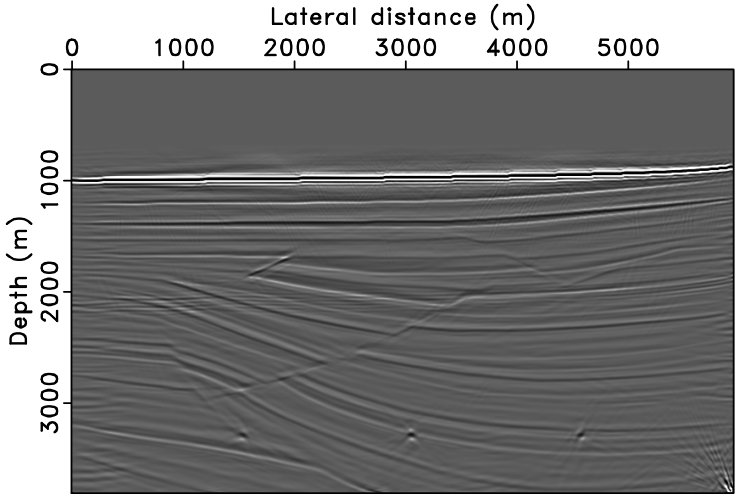


(a)

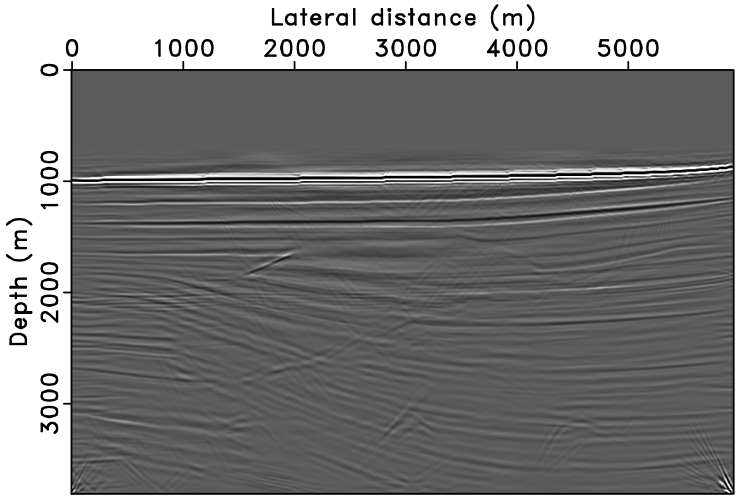


(b)

Figure 2.8: The inaccurate background model. **(a)** The wrong background velocity model. **(b)** One vertical trace of **(a)**.

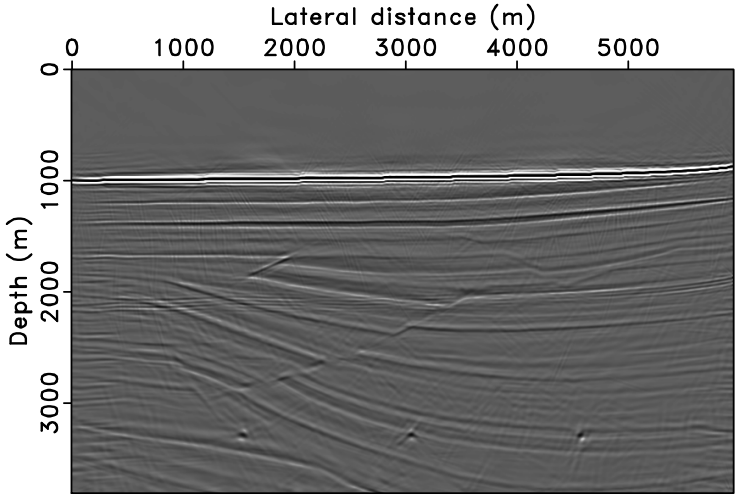


(a)

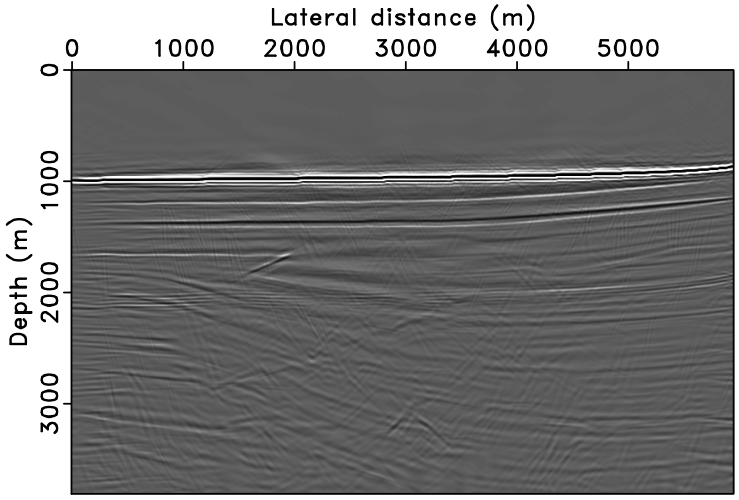


(b)

Figure 2.9: Images using the proposed method, without **(a)** and with **(b)** the velocity error.



(a)



(b)

Figure 2.10: Images without curvelet-domain sparsity promotion, without **(a)** and with **(b)** the velocity error.

prediction by using a generalized areal-source term that contains the downgoing receiver wavefield. By incorporating this source term, we arrive at a linearized formulation that predicts, and therefore correctly images, data with surface-related multiples. Instead of predicting the multiples via explicit dense matrix-matrix multiplications as performed in surface-related multiple elimination, which is prohibitively expensive for large seismic data sets, our method predicts multiples implicitly by solving wave equations for the areal-source wavefields, and therefore eliminate the cost of the matrix product altogether. Second, our combination of multiple prediction and linearized forward modelling with the Born approximation allows us to come up with a proper curvelet-domain sparsity-promoting inversion framework where primaries and surface-related multiples are jointly inverted to produce a high-resolution true-amplitude image of the subsurface medium perturbations. Compared to non-iterative correlation-based imaging methods with multiples, this approach leads to high-resolution images with minimal artifacts. Third, we overcome prohibitive computational cost of sparsity-promoting inversion by choosing to work with small independent randomized subsets of data. This subsampling and rerandomization strategy leads to a simulation cost that is comparable to the cost of a single reverse-time migration with all the data, which makes our method feasible in practice. More importantly, our method produces by virtue of the inversion high-fidelity true-amplitude images, an accomplishment that is difficult, if at all possible, to achieve with conventional methods. Because of its superior performance on complex synthetics, in conjunction with a reduced sensitivity to modelling errors and errors in the velocity model, we expect our method to be suitable to field data.

Chapter 3

Controlling linearization errors in ℓ_1 regularized inversion by rerandomization

3.1 Summary

Linearized inversion is a data fitting procedure that tries to match the observed seismic data with data predicted by linearized modelling. In practice, the observed data is not necessarily in the column space of the linearized modelling operator. This can be caused by lack of an accurate background velocity model or by coherent noises not explained by linearized modelling. Through carefully designed experiments, we observe that a moderate data mismatch does not pose an issue if we can use *all* the data in the inversion. However, artifacts do arise from the mismatch when randomized dimensionality reduction techniques are adopted to speed up the inversion. To stabilize the inversion for dimensionality reduction with randomized source aggregates, we propose to *rerandomize* by drawing independent simultaneous sources occasionally during the inversion. The effect of this rerandomization is remarkable because it results in virtually artifact-free images at a cost comparable to a single reverse-time migration. Implications of our method are profound because we are now able to resolve fine-scale steep subsalt features in a computationally feasible manner.

A version of this chapter has been published in SEG Technical Program Expanded Abstracts, 2013, vol. 32, p. 4640-4644

3.2 Introduction

Linearized inversion aims to obtain the perturbation of the subsurface medium parameterized by a model vector $\mathbf{m}$, by fitting the predicted data to the observed seismic data. The predicted data is identified as a linear function of the model perturbation $\delta\mathbf{m}$ over a background model $\mathbf{m}_0$ where $\mathbf{m} = \mathbf{m}_0 + \delta\mathbf{m}$. The underlying optimization problem can be written as:

$$\mathbf{LS}: \quad \underset{\delta\mathbf{m}}{\text{minimize}} \quad \|\nabla\mathbf{F}[\mathbf{m}_0, \mathbf{q}]\delta\mathbf{m} - \mathbf{d}\|_2, \tag{3.1}$$

where $\nabla\mathbf{F}$ is the linearized Born scattering operator, $\mathbf{q}$ is the vectorized source wavefield for all sources, which is either assumed to be known, or can be estimated by variable projection (Aravkin et al., 2013b; Tu et al., 2013), and $\mathbf{d}$ is the observed data, usually after removal of surface-related multiples.

Despite all its advantages, such as true-amplitude preservation and robustness to missing data (Tu et al., 2013; Nemeth et al., 1999), linearized inversion is not widely adopted because of its high computational cost. Solving problem **LS** usually requires iterative evaluations of the Born scattering operator and its adjoint, while each evaluation is already expensive as it entails several wave-equation solves (e.g., four wave-equation solves without saving the background wavefield) for each monochromatic source experiment. Since the simulation cost is mainly determined by the number of sources and frequencies, dimensionality reduction techniques—such as source encoding and using randomized frequency subsets—are proposed to reduce these costs (Dai et al., 2011; Herrmann and Li, 2012). To overcome the negative effects these subsampling techniques have (e.g., source cross-talks) on the seismic image, Herrmann and Li (2012) proposed to promote Curvelet sparsity in the model space by solving the following ℓ_1 minimization problem (known as the Basis Pursuit De-Noise (BPDN) problem, Chen et al., 2001; van den Berg and Friedlander, 2008):

$$\begin{aligned} \mathbf{BP}_\sigma: \quad & \text{minimize} \quad \|\mathbf{x}\|_1 \\ & \text{subject to} \quad \|\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{q}}]\mathbf{C}^*\mathbf{x} - \underline{\mathbf{d}}\|_2 \leq \sigma, \end{aligned} \tag{3.2}$$

where underlined variables denote subsampled quantities, i.e., $\underline{\mathbf{d}} = \mathbf{RMd}$ where $\mathbf{RM}$ is the subsampling matrix (see Herrmann and Li, 2012, for more details). Matrix $\mathbf{C}^*$ is the Curvelet synthesis operator (Candès et al., 2006a), and σ is adjusted to allow for data mismatch due to noise and modelling errors. To solve the above BPDN problem, we use the SPGℓ_1 solver that

implicitly solves a series of Least Absolute Shrinkage and Selection Operator (LASSO) subproblems with gradually relaxed τ's (Tibshirani, 1996; van den Berg and Friedlander, 2008):

$$\begin{aligned} \mathbf{LS}_\tau: \quad & \text{minimize} \quad \|\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{q}}]\mathbf{C}^*\mathbf{x} - \underline{\mathbf{d}}\|_2 \\ & \text{subject to} \quad \|\mathbf{x}\|_1 \le \tau. \end{aligned} \tag{3.3}$$

In practice, the input data $\mathbf{d}$ does not necessarily lie in the column space of the linearized modelling operator $\nabla\mathbf{F}$. This can be caused by lack of an accurate background model, or by coherent events (e.g., internal multiples) or incoherent noises in the data that the linearized modelling fails to explain. As a result, there will be a mismatch between the input data and the predicted data. However, it is rarely discussed in the literature how to choose the "right" σ for the above $\mathbf{BP}_\sigma$ problem in this "noisy" case. Moreover, how the mismatch is going to affect the image quality, especially when we subsample the data, is not well understood. In this article, we carefully design several experiments to find out the answers to the above questions. We also propose possible solutions if the image quality is compromised due to the data mismatch. We start with the case where hypothetically we can use all the data (i.e., $\mathbf{RM}$ is simply a Dirac operator in this case), and then delve into the case where we randomly subsample the data to speed up the inversion. In the end, we verify our observations with a case study using a realistic synthetic model.

3.3 The full-data case

We use a simple two-layer model to examine how the choice of σ in problem $\mathbf{BP}_\sigma$ affects the inversion. The model is 400 m wide and 400 m deep, with a 5 m grid spacing. The true model $\mathbf{m}$ has two layers. The top layer ranges from the surface to 145 m in depth, and has a velocity of 1500 m/s. The bottom layer ranges from 145 m to 400 m in depth, and has a velocity of 2500 m/s. We smooth the true model to get the background model $\mathbf{m}_0$, and obtain the perturbation $\delta\mathbf{m}$ by subtracting the background model from the true model. There are 21 sources put at a depth of five meters with 20 m lateral spacing, and 81 receivers put at the same depth as the sources with 5 m lateral spacing.

We generate two sets of data here. We make the first data set by linearized modelling (i.e., $\nabla\mathbf{F}[\mathbf{m}_0, \mathbf{q}]\delta\mathbf{m}$), which we refer to as the *linearized* data. We make the second data set by first modelling the entire wavefield with the true model (i.e., $\mathbf{F}[\mathbf{m}, \mathbf{q}]$ where $\mathbf{F}$ is the forward modelling op-

erator), and then subtracting the direct waves modelled with the smooth background model (i.e., $\mathbf{F}[\mathbf{m}_0, \mathbf{q}]$). We refer to this data set as the *forward modelling* data. For both data sets, we use a Ricker wavelet of 25 Hz peak frequency as the source. From an inversion point of view, the major difference between the two data sets is that the linearized data is perfectly in the column space of the Born scattering operator, while the forward modelling data is not. Traces from the two data sets are plotted in Figure 3.1(a). The visual difference between the two data sets is mainly a phase shift, most probably caused by the slightly different kinematics between the true and the background model.

We compare three scenarios here. In the first scenario, we obtain a baseline image by inverting the linearized data. As the data is completely noise-free, we choose $\sigma = 0$ and run for 100 iterations using the SPGℓ_1 solver (van den Berg and Friedlander, 2008). In the second scenario, we use the forward modelling data as the input, and use the *true* σ defined as the ℓ_2-norm of the difference (by subtraction) between the linearized data and the forward modelling data. The third scenario is otherwise the same as the second scenario, except that we simply choose $\sigma = 0$ and let the solver run for 100 iterations.

The inversion results are shown in Figure 3.1(b) to 3.1(d). We can see that the results of the first and the third scenarios are comparable. The failure of the second scenario indicates that the choice of σ is inappropriate. As we can see from Figure 3.1(a), although the phase shift between the two data sets is minimal, the difference (in red) is almost as large as the input data itself. Therefore using the "true" σ, which is the ℓ_2-norm of the difference, will terminate the optimization prematurely. As a result, the second scenario stopped after only four iterations in our experiment. In fact, the linearized modelling in the third scenario well fits the forward modelling data (Figure 3.1(e)). Figure 3.1(f) shows that the phase shift between the two data sets (Figure 3.1(a)) is reasonably explained mainly by a slight spatial shift of the model perturbation. Although this spatial shift is not desirable, at least it leads to a coherent image in Figure 3.1(d) compared with 3.1(c).

3.4 The subsampling case

With the same setup as the previous section, we examine how the data mismatch affects the image quality when we subsample the data to speed up the inversion. Instead of using all 21 sequential sources and 61 frequencies, we use 5 simultaneous sources by randomized source superposition, and 15

randomly selected frequencies (see Herrmann and Li, 2012, for details). Based on our observations in the previous section, we choose $\sigma = 0$ and run for 100 iterations.

We first compare two scenarios. Again, we use linearized data as the input in the first scenario and the forward modelling data in the second scenario. The results are shown in Figure 3.1(g) and 3.1(h). We can see that the mismatch does result in significant interfering artifacts in 3.1(h) when we subsample the data. In other words, the *subsampled* BPDN problem fails to converge to the same solution as the *full* problem. This is closely related to the calibration problem in compressive sensing (Herman and Strohmer, 2010; Gribonval et al., 2012).

In our previous work, we noticed that using a technique that we refer to as *rerandomization* can greatly improve the convergence in terms of model errors decrease (Tu and Herrmann, 2012a), by suppressing the imaging artifacts that arise from source cross-talks due to significant *subsampling*. By rerandomization, we mean to draw an independent random subsampling matrix **RM** for each LASSO subproblem, leveraging SPGℓ_1's warm start feature. To maximize the benefit of rerandomization, we redraw both simultaneous sources and randomized frequency subsets (Tu and Herrmann, 2012a).

As the interfering artifacts from linearization errors are also related to the *subsampling* procedure, we *suppose* the rerandomization technique to be also able to mitigate this type of artifacts. To verify this hypothesis, we setup the third scenario, which is otherwise the same as the second scenario except that we use rerandomization. The result is shown in Figure 3.1(i), which is virtually artifact-free. Although not shown here, we observe that the middle traces of Figure 3.1(d) and Figure 3.1(i) highly resemble each other. Therefore we argue that rerandomization enables the *subsampled* BPDN problem to converge to virtually the same solution as the *full* problem. This also explains the improved convergence observed by Tu and Herrmann (2012a).

3.5 Synthetic examples

In this section, we verify our previous observations using a more challenging model modified from a 2D slice of the SEG /EAGE salt model. We pad 10 grid points to the water layer of the model for a more accurate removal of direct waves (i.e., we can better retain the velocity in the shallow water layer after spatially smoothing this model to obtain the background model, which we use to simulate the direct waves). The model is 3.9 km

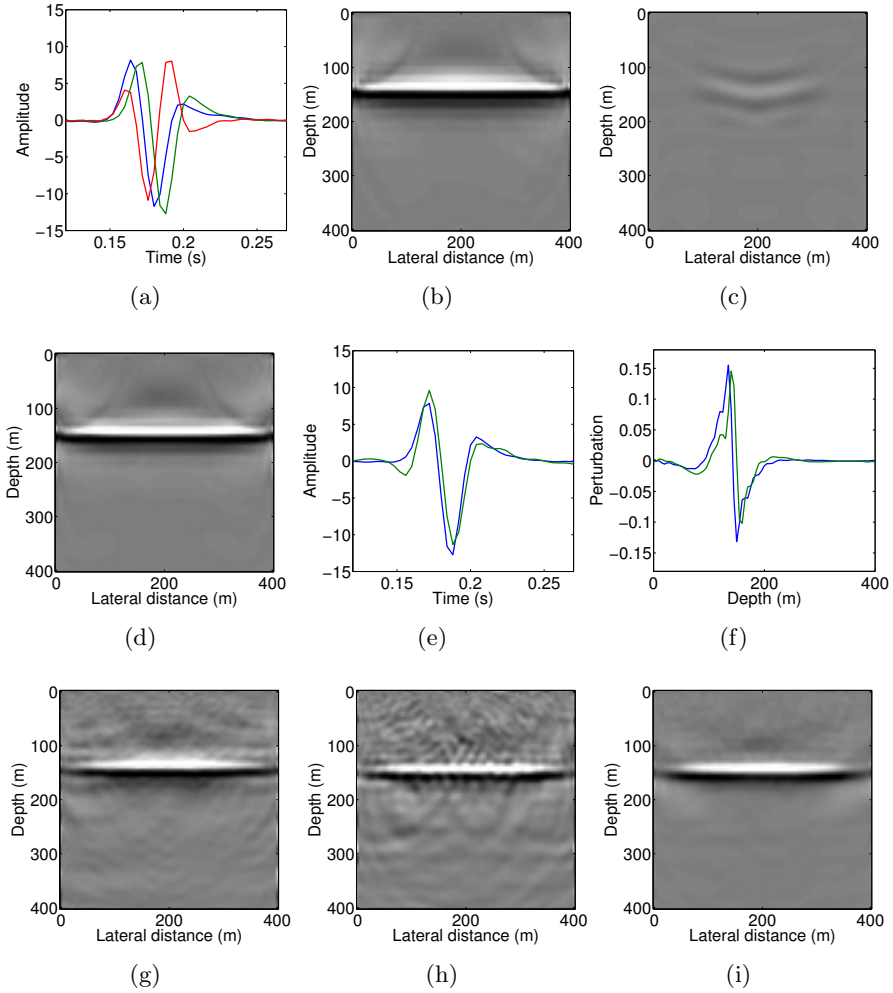


Figure 3.1: Examples with the two-layer model. The images are plotted with the same color scale. **(a)**. Traces of the linearized data (blue), the forward modelling data (green), and their difference (red). **(b)**. Inversion of the full linearized data. **(c)**. Inversion of the full forward modelling data with the true σ. **(d)**. Inversion of the full forward modelling data with $\sigma = 0$. **(e)**. A comparison of the input forward modelling data (blue) and the linearized modelling data with the inversion results of **(d)** (green). **(f)**. The middle traces of **(b)** and **(d)**. **(g)**. Inversion of the subsampled linearized data, $\sigma = 0$. **(h)**. Inversion of the subsampled forward modelling data, $\sigma = 0$. **(i)**. Otherwise the same as **(h)** but with rerandomization, $\sigma = 0$.

deep and 15.7 km wide with 24.384 m (80 feet) grid spacing. We smooth the true model to obtain the background model. As salt creates high velocity contrast with the surrounding medium, smoothing a salt boundary can result in significant linearization errors (remember the linearization of the wave-equation is based on the assumption that the model perturbations are small). We obtain the true model perturbations by subtracting the true and the background models, displayed in Figure 3.2(b). We put 323 co-located sources and receivers every two grid points laterally at the second grid point in depth. We again generate two sets of data: the *linearized* data and the *forward modelling* data. For both data sets, we use a Ricker wavelet of 5 Hz peak frequency as the source and record for 8 seconds. Traces from the two data sets, as well as their difference, are plotted in Figure 3.2(a).

We compare four scenarios here. In the first scenario, we run a reverse-time migration (RTM) with all the data to obtain a baseline image. In the second scenario, we invert from the subsampled linearized data. In the third scenario, we invert from the subsampled forward modelling data. In the fourth scenario, we again invert from the subsampled forward modelling data, but with rerandomization. In all the inversion scenarios, we use 15 frequencies and 15 simultaneous sources. We set the misfit allowance σ to be 0 and run for 100 iterations. With such a setup, the simulation cost is about 1.45X of a single RTM with all the data. We also mute the water layer during each iteration of the inversion, as the artifacts imaged inside the water layer can adversely affect data fitting.

The results are shown in Figure 3.2(c) to 3.2(f). The target zone that we are most interested in lies below the salt structure. Figure 3.2(d) and 3.2(e) show that the linearization errors does significantly degrade the image quality when the input data is subsampled, especially in the target zone. However, by rerandomization, the image quality is greatly improved (Figure 3.2(f)). To understand this significant improvement in the image quality from the perspective of data-fitting, we demigrate the inverted model perturbations from the (subsampled) forward modelling data, both without (Figure 3.2(e)) and with (Figure 3.2(f)) rerandomization, and compare them with the input forward modelling data. Traces from these data sets are plotted in Figure 3.2(g). We can see that improvements in the image quality indeed leads to improved data fitting.

We have yet another important observation from this set of examples. Compared with inversion, RTM (the adjoint operation of the linearized modelling) is usually considered to be more robust to coherent mismatches in the data. However, using the proposed fast linearized inversion approach with rerandomization, we can better resolve the subsalt structures despite

the linearization errors. To provide a detailed comparison, we zoom into the target zones in Figure 3.2(c) and 3.2(f), and compare them in Figure 3.2(h). We can see that the image in Figure 3.2(f) is more faithful to the true model perturbations shown in 3.2(b).

3.6 Conclusion

We analyzed the effects of data mismatches on ℓ_1-norm regularized inversion related to errors in the linearization. This can lead to major artifacts especially when randomized sources are used to make the ℓ_1-norm regularized inversion computationally feasible. Allowing for a tolerance depending on the linearization errors makes the situation only worse because the inversion stops prematurely in that case. We overcome this detrimental weakness of ℓ_1-norm regularized inversion by choosing independent subsets of randomized sources during the inversion. Not only does this rerandomization lead to faster convergence but it also removes many of the artifacts related to errors in the linearization. We tested this algorithm on stylized and realistic synthetic examples. Our method produced virtually artifact-free high-resolution images at low computational cost despite errors in the linearized forward model. Contrary to relying on the fold to stack out imaging related errors during reverse-time migration, our method uses redundancy in the data to stabilize the inversion. This allows us to create high-resolution images at a computational cost comparable to the cost of a single reverse-time migration with all data.

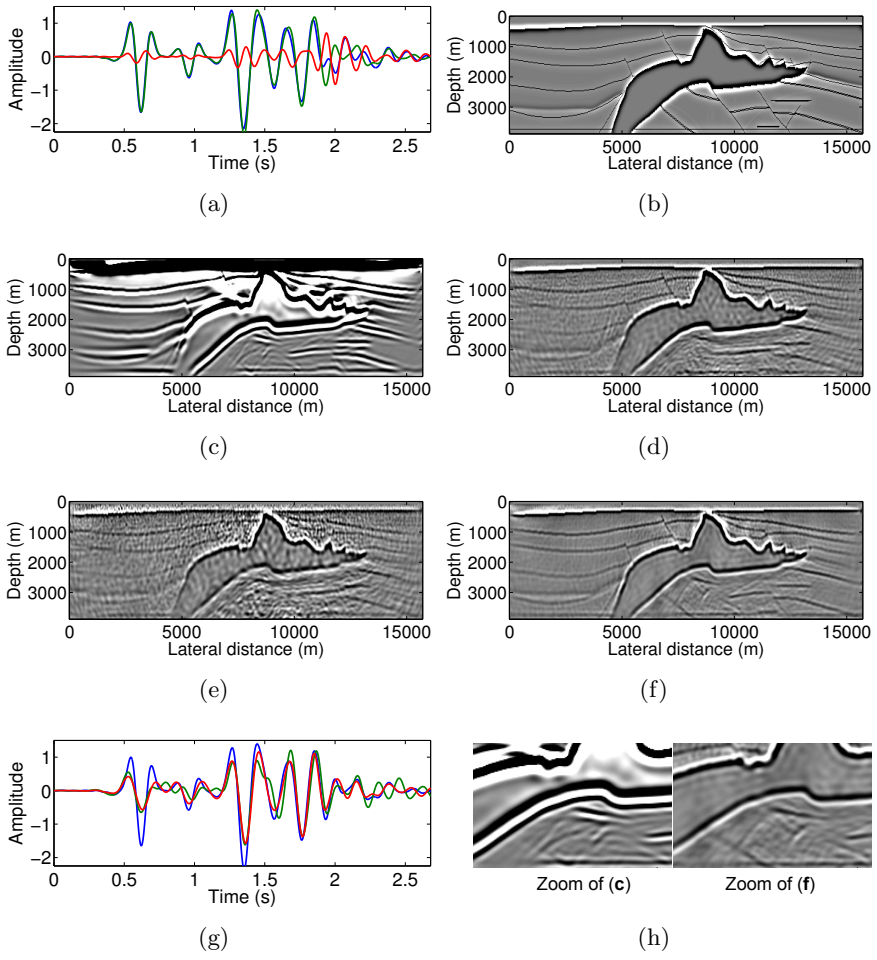


Figure 3.2: Case study with the synthetic SEG/EAGE 2D salt model. All images are plotted with the same color scale *except* the RTM image. **(a)**. Traces of the linearized data (blue), the forward modelling data (green), and their difference (red). **(b)**. True model perturbation. **(c)**. The RTM image of the forward modelling data. **(d)**. Inversion of the subsampled linearized data. **(e)**. Inversion of the subsampled forward modelling data. **(f)**. Otherwise the same as **(e)** but with rerandomization. **(g)**. Traces of the input forward modelling data (blue) and the linearized modelling data with the inversion results of **(e)** (green) and **(f)** (red). **(h)**. Zoomed subsalt areas of **(c)** (left) and **(f)** (right).

Chapter 4

Limitations of the deconvolutional imaging condition for two-way propagators

4.1 Summary

The deconvolutional imaging condition has gained wide attention in recent years, as it is often used to image surface-related multiples. However, we noticed on close inspection that this condition was derived from one-way propagation principles. Now that two-way wave-equation based simulations have become more affordable, we revisit the deconvolutional imaging condition and reveal its limitations for two-way propagators. First, it can distort the image due to receiver-side propagation effects. Second, when used to image surface-related multiples, it is not capable of removing all interfering phantom reflectors.

4.2 Introduction

Surface-related multiples in marine acquisition can be strong enough to interfere with the primary reflections. If not handled correctly, they can result in phantom reflectors that hinder correct interpretation. On the other

A version of this chapter has been published in SEG Technical Program Expanded Abstracts, 2013, vol. 32, p. 3916-3920

hand, if correctly used, they can provide additional information on the subsurface. Based on the SRME formulation (Verschuur et al., 1992), surface-related multiples can be identified as the seismic response of the subsurface medium when the downgoing receiver wavefield at the surface is injected as the source wavefield (Berkhout, 1993). However, even with the source wavefield correctly identified, imaging the multiples still remains a challenge as the cross-correlation imaging condition fails to yield an artifacts-free seismic image due to acausal cross-correlations (Muijs et al., 2007; Liu et al., 2011). To address the issue, some researchers proposed to use the deconvolutional imaging condition (or the smoothing imaging condition derived from it) (Claerbout, 1971; Guitton et al., 2007; Muijs et al., 2007; Whitmore et al., 2010; Lu et al., 2011), while others suggested that an inversion approach (e.g., least-squares migration) should be used (Verschuur, 2011; Tu and Herrmann, 2012a).

The deconvolutional imaging condition was derived from one-way propagation principles (Claerbout, 1971). Now that the two-way wave-equation based simulations have become more affordable, we scrutinize the applicability of the deconvolutional imaging condition for two-way propagators. Our contribution is organized as follows. First, we do a theoretical analysis of the deconvolutional imaging condition in the context of one-way and two-way wave propagations. As it is extensively used to image surface-related multiples, we continue to examine its efficacy for imaging surface-related multiples, and compare it with our linearized inversion approach, namely the sparsity promoting migration with surface-related multiples (Herrmann and Li, 2012; Tu and Herrmann, 2012a).

4.3 Theoretical analysis

For simplicity, we use the detail-hiding matrix notation to describe the one-way wave propagation (Berkhout and Wapenaar, 1993). Tensor $\mathbf{U}^+$ is the down-going wavefield of the subsurface, obtained by downward propagating the down-going source wavefield at the surface $\mathbf{S}^+$, and $\mathbf{V}^-$ is the up-going wavefield of the subsurface, obtained by downward propagating the upgoing receiver wavefield at the surface $\mathbf{D}^-$. In frequency domain modelling, both $\mathbf{U}^+$ and $\mathbf{V}^-$ are representations of the physical position (x, z), the source is indexed by i, and the frequency is indexed by j. Denoting the monochromatic downward continuation operator as $\mathbf{W}_j^+$, and the upward continuation operator as $\mathbf{W}_j^-$, and the reflectivity operator as $\mathbf{R}_{(i,j)}$, we have $\mathbf{U}_{(i,j)}^+ = \mathbf{W}_j^+ \mathbf{S}_{(i,j)}^+$ and the receiver wavefield $\mathbf{D}_{(i,j)}^- = \mathbf{W}_j^- \mathbf{R}_{(i,j)} \mathbf{W}_j^+ \mathbf{S}_{(i,j)}^+$.

Note that with fixed indices (i,j), all wavefields here are vectors. In cases where the exact inverse of the upward continuation operator $\mathbf{W}_j^-$ can be derived so that $\mathbf{V}_{(i,j)}^- = (\mathbf{W}_j^-)^{-1}\mathbf{D}_{(i,j)}^-$, the zero-offset deconvolutional imaging condition can be put as (Claerbout, 1971)

$$\begin{aligned}\widetilde{\mathbf{R}} &= \sum_{i,j} \mathbf{V}_{(i,j)}^- / \mathbf{U}_{(i,j)}^+ \\ &= \sum_{i,j} \frac{(\mathbf{W}_j^-)^{-1}\mathbf{W}_j^-\mathbf{R}_{(i,j)}\mathbf{W}_j^+\mathbf{S}_{(i,j)}^+}{\mathbf{W}_j^+\mathbf{S}_{(i,j)}^+} \\ &= \sum_{i,j} \mathrm{diag}(\mathbf{R}_{(i,j)}),\end{aligned} \tag{4.1}$$

where $\widetilde{\mathbf{R}}$ is the reconstructed image, the division is performed in a element-wise manner, $\mathbf{R}$ is a diagonal operator with the reflectivity coefficients on the diagonal entries (Verschuur and Berkhout, 2009), and notation "diag" means to obtain the diagonal entries. In this case, the image is a summation of the reflectivities for all sources and frequencies. However, as pointed out by Gray (1997), $(\mathbf{W}_j^-)^*$ (star means the adjoint) is often used in practice due to the instability of inverting $\mathbf{W}_j^-$. In this case, the image can be distorted due to receiver side propagation effects, i.e., $(\mathbf{W}_j^-)^*\mathbf{W}_j^-$.

Now we turn to two-way wave-equation migrations. A fundamental difference from its one-way counterpart is that an image is *not* formed from the reflectivity, but from the perturbations in medium parameters $\delta\mathbf{m}$ over a background medium $\mathbf{m}_0$. Instead of having the down-going wavefield $\mathbf{U}^+$ and up-going wavefield $\mathbf{V}^-$ by downward continuation, we now have the temporally forward-propagated source wavefield $\mathbf{U}$ and backward-propagated receiver wavefield $\mathbf{V}$. Denoting the time-harmonic Helmholtz operator as $\mathbf{H}_j = \mathbf{H}[\mathbf{m}_0, \omega_j]$, we can write $\mathbf{U}_{(i,j)} = \mathbf{H}_j^{-1}\mathbf{P}^*\mathbf{S}_{(i,j)}$ where operator $\mathbf{P}$ restricts data at the surface and $\mathbf{P}^*$ maps surface data to the entire domain. By writing $\mathbf{W}_j^+ \doteq \mathbf{H}_j^{-1}\mathbf{P}^*$ and $\mathbf{W}_j^- \doteq \mathbf{P}\mathbf{H}_j^{-1}$ here, we have $\mathbf{U}_{(i,j)} = \mathbf{W}_j^+\mathbf{S}_{(i,j)}$, and the linearized receiver wavefield

$$\mathbf{D}_{(i,j)} = \omega_j^2\mathbf{W}_j^-\mathrm{diag}(\delta\mathbf{m})\mathbf{W}_j^+\mathbf{S}_{(i,j)}.$$

Here we overload the notation "diag" to mean inserting a vector into the diagonal entries of a matrix. Now we clearly see the analogue between the one-way and two-way propagators. By a time-reversal propagation of the

receiver wavefield, we have

$$\mathbf{V}_{(i,j)} = (\mathbf{H}_j^*)^{-1}\mathbf{P}^*\mathbf{D}_{(i,j)} = (\mathbf{W}_j^-)^*\mathbf{D}_{(i,j)}.$$

Since the adjoint of $\mathbf{W}_j^-$ is used, as in the case of using one-way propagators, the propagation effects on the receiver side can again distort the image if the deconvolutional imaging condition is applied.

On the other hand, apply the cross-correlation imaging condition for two-way propagators (i.e., reverse-time migration, RTM) remains the adjoint of the linearized forward modelling (Lailly, 1983). As the deconvolutional imaging condition is often regarded as an "inversion" scheme (see Guitton et al., 2007, equation (3)), we also compare the approach with our iterative linearized inversion approach, where we invert the linearized Born scattering operator by promoting Curvelet sparsity in the model space (Herrmann and Li, 2012). With the notations written above, the monochromatic linearized Born scattering operator $\nabla\mathbf{F}_j$ is expressed as: $\mathbf{W}_j^-\mathrm{diag}(\delta\mathbf{m})\mathbf{W}_j^+\mathbf{S}_j \doteq \nabla\mathbf{F}_j[\mathbf{m}_0;\mathbf{S}_j]\delta\mathbf{m}$, and is a function of the monochromatic source wavefield $\mathbf{S}_j$ and the background model $\mathbf{m}_0$.

Since the deconvolutional imaging condition is plagued by zeros in the denominator (see Equation (4.1)), two stabilized versions have been reported in the literature. First, using a damping term, as reported by Claerbout (1971):

$$\widetilde{\mathbf{R}} = \sum_{i,j} \frac{\mathbf{V}_{(i,j)}\mathbf{U}^*_{(i,j)}}{\|\mathbf{U}_{(i,j)}\|_2^2 + \epsilon}. \tag{4.2}$$

In our experiments, we choose the damping parameter ϵ by $\epsilon = \lambda * \mathrm{mean}(\|\mathbf{U}_{(i,j)}\|_2^2)$. Second, the smoothing imaging condition, as reported by Guitton et al. (2007):

$$\widetilde{\mathbf{R}} = \sum_{i,j} \frac{\mathbf{V}_{(i,j)}\mathbf{U}^*_{(i,j)}}{< \|\mathbf{U}_{(i,j)}\|_2^2 >_{(x,z)}}, \tag{4.3}$$

where the symbol $<>_{x,z}$ means smoothing vertically and horizontally over a window.

4.4 Stylized examples

We design the following experiment to verify our theoretical analysis. The background model we use is shown in Figure (4.1(a)). The left and the right halves are symmetric. There are 3000 m/s velocity anomalies (boundaries are smoothed) over a background velocity of 1500 m/s. The model pertur-

bation is shown in Figure 4.1(b). The model is 594 m deep and 1194 m wide with 6 m grid distance. There are 50 sources put on the left side and 50 receivers symmetrically put on the right side at 6 m in depth with 12 m lateral spacing. By designing such a model and acquisition geometry, the waves to image the left half of the perturbation are more affected by the source-side propagation effects, while the right half is more affected by the receiver-side effects. We generate data by applying the linearized Born scattering operator $\nabla\mathbf{F}$ to the true model perturbation $\delta\mathbf{m}$. A Ricker wavelet of 20 Hz peak frequency is used.

We compare three scenarios here. In the first scenario, we obtain a baseline RTM image. In the second scenario, we use two-way propagators, but use the damped deconvolutional imaging condition (Equation 4.2), with the damping parameter $\lambda = 0.05$, and the smooth imaging condition (Equation 4.3), with a 100 m smoothing window. In the third scenario, we run our sparsity promoting migration with all the data for 100 iterations. The results are shown in Figure 4.1(c) to 4.1(f).

Comparing the images, we can see that the deconvolutional imaging condition, as well as its smoothing variant, lead to wrong positions of the model perturbation on the right half of the model (indicated by arrows). Combining this observation with our theoretical analysis, we conclude that for two-way propagators, the deconvolutional imaging condition can distort the image due to receiver side propagation effects. On the other hand, both RTM and the linearized inversion give us correct images.

4.5 Imaging of multiples

Most applications of the deconvolutional imaging condition that we are interested in are to image surface-related multiples, where it is used somewhat as an "invesion" approach. Now we have shown that it is not really an inversion approach, we are going to investigate the plausibility to use it to image the surface-related multiples by comparing it with the iterative linearized inversion approach.

However, linearized inversion for a model of realistic dimensions is usually considered to be computationally unaffordable as it requires repeated evaluations of the Born scattering operator and its adjoint, while each evaluation itself is already expensive by having to solve four wave-equations (without saving the background wavefield) for each monochromatic source experiment. In the case of imaging multiples, the computational cost is significantly increased by performing multi-dimensional convolution for multiple prediction. However, we have shown in Chapter 2 that the computational

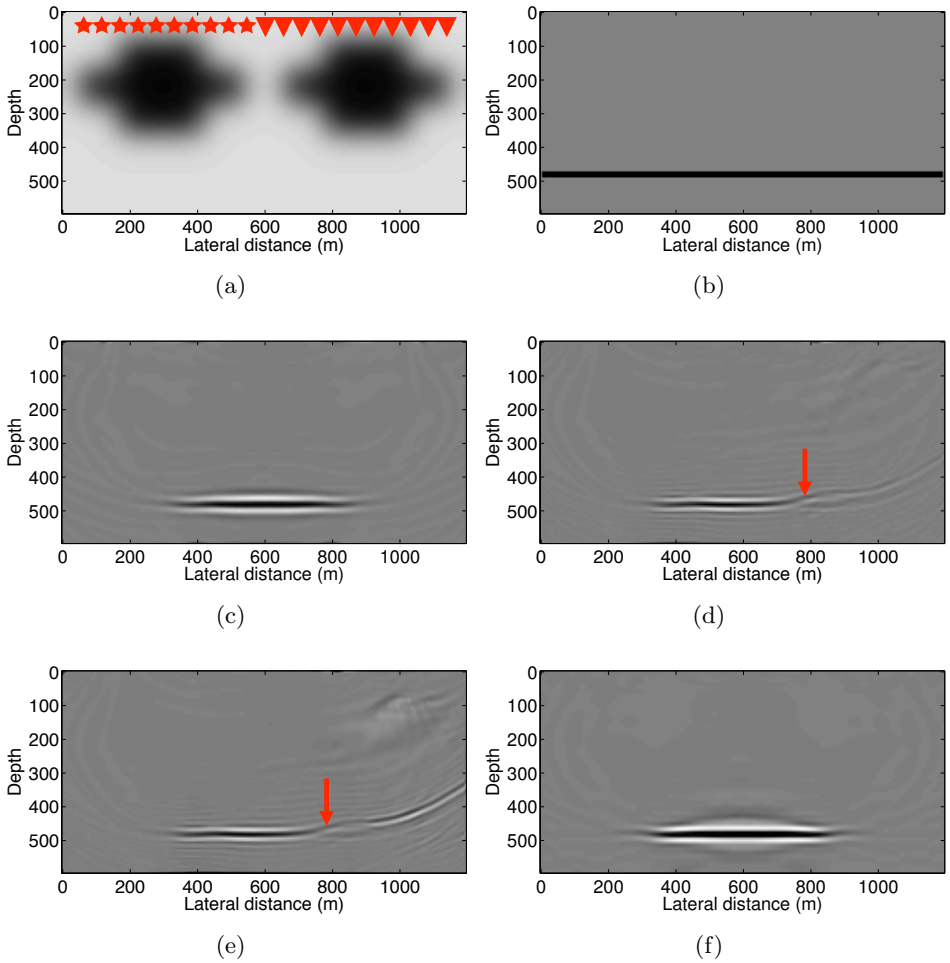


Figure 4.1: The model and the imaging results of the stylized example. The arrow in **(d)** or **(e)** indicates the image distortion. **(a)**. The background model, stars indicate source positions and triangles indicate receivers. **(b)**. True model perturbation. **(c)**. The RTM image. **(d)**. Image by the deconvolutional imaging condition. **(e)**. Image by the smoothing imaging condition. **(f)**. Image by sparsity promoting migration.

cost can be greatly reduced by: *(i)* identifying the total down-going wavefield at the surface (i.e., $\mathbf{D}$ with our notation) as the (generalized) areal sources, and letting the wave-equation solver to carry out the multi-dimensional convolution implicitly; and *(ii)* reducing per-iteration simulation cost by using simultaneous sources and randomized subsets of frequencies. By introducing *rerandomization* (Tu and Herrmann, 2012a), the computational cost can be further reduced without compromising the image quality. Here we use these techniques to achieve a fast inversion.

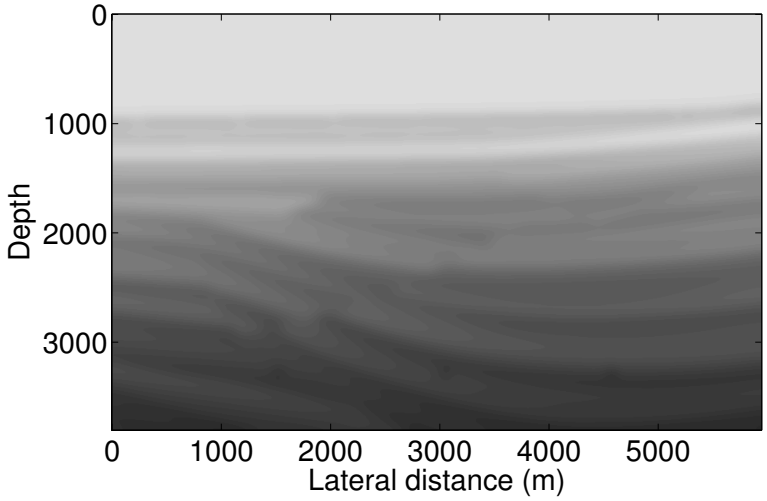
We use a more realistic model cropped from the sedimentary part of the Sigsbee 2B model (courtesy of the SMAART JV). The background model is shown in Figure 4.2(a), and the true model perturbation is shown in Figure 4.2(b). The model is 3.8 km deep and 5.9 km wide, with 7.62 m grid spacing. There are 261 co-located sources and receivers at a depth of 7.62 m with 22.86 m lateral spacing. We generate the surface-related multiples by applying the linearized Born scattering operator with generalized source $\nabla\mathbf{F}[\mathbf{m}_0, \mathbf{D}]$ to the true model perturbation $\delta\mathbf{m}$. We model the total data $\mathbf{D}$ using a time-domain finite difference code with a free-surface boundary condition.

We do the same comparisons as we did for the stylized example. We skip the smoothing imaging condition as it yield similar results as the damped deconvolutional imaging condition (e.g., compare Figure 4.1(d) and 4.1(e)). The damping parameter λ for the deconvolutional imaging is tweaked to be 0.4 (a smaller λ yields to instability in division). In our sparsity promoting migration, we use 15 frequencies, 8 simultaneous sources, and run 338 iterations. This leads to a simulation cost roughly the *same* as a single RTM of all the data. The results are shown in Figure 4.3(a) to 4.4.

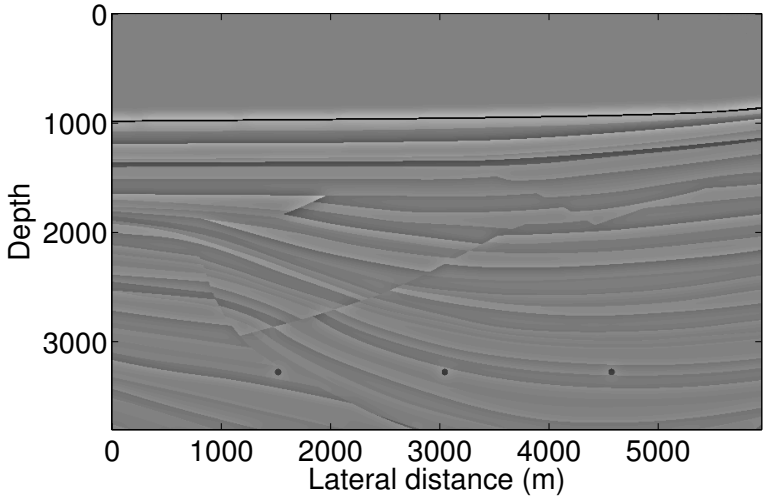
Many authors have concluded that for multiples, the cross-correlation imaging condition (i.e., RTM) introduces artifacts caused by acausal cross-correlations (a replicate of the ocean bottom reflection is indicated by the arrow in Figure 4.3(a), see Muijs et al., 2007; Liu et al., 2011, for details). Now from Figure 4.3(b), we can see that the deconvolutional imaging condition is neither the right choice (the ocean bottom replicate is again indicated by an arrow). The same artifact is also observed by Muijs et al. (2007) (Figure 11b) using one-way propagators. On the other hand, we do get an virtually artifact-free image using our fast iterative inversion method.

4.6 Conclusion

We analyzed the limitations of the deconvolutional imaging condition for two-way propagators. We show that it can distort the image due to receiver-

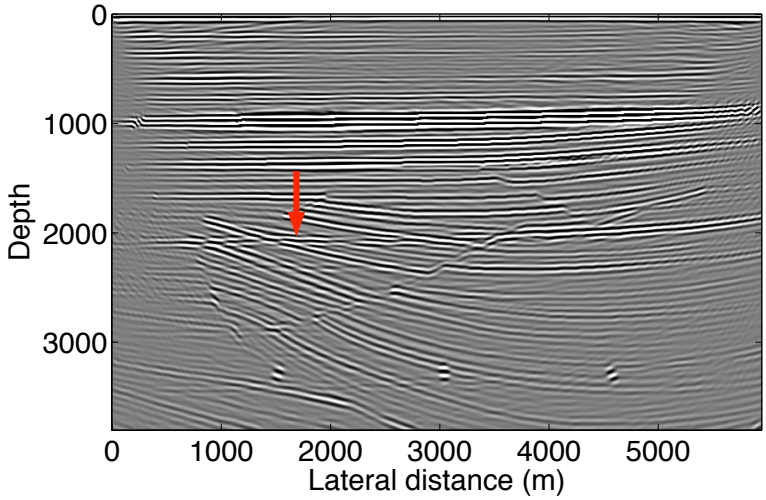


(a)

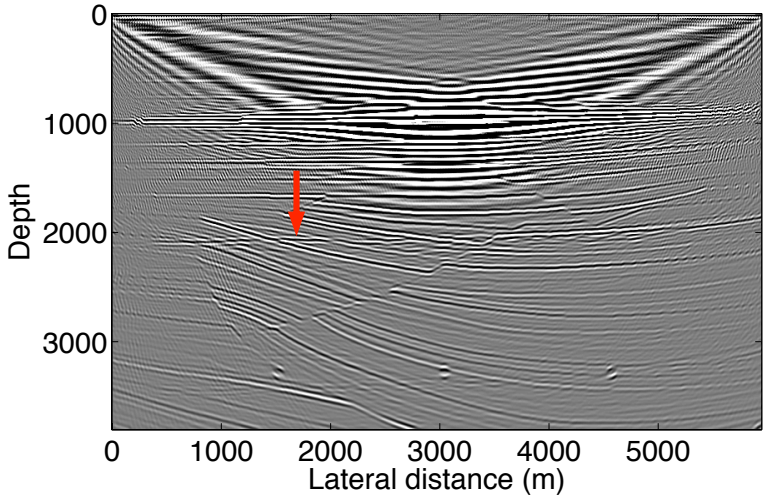


(b)

Figure 4.2: The cropped Sigsbee 2B model. **(a)** The background model. **(b)** True model perturbation.



(a)



(b)

Figure 4.3: The RTM **(a)** and the deconvolutional **(b)** images of multiples.

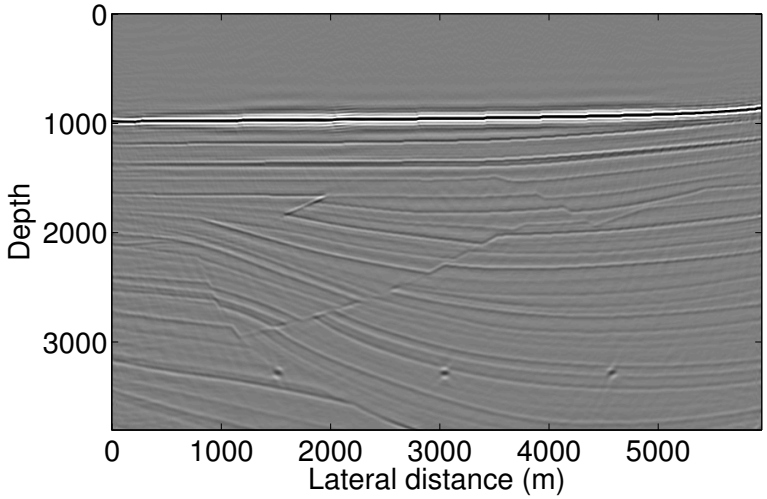


Figure 4.4: Image of multiples by our fast sparsity promoting migration, where the simulation cost is roughly the same as Figure 4.3.

side propagation effects. When used to image the surface-related multiples, it fails to yield an image free of coherent artifacts as it is not really an inversion approach. In light of this, a true inversion scheme is necessary to image the surface-related multiples, and this can be done efficiently using our fast sparsity promoting migration approach with rerandomization.

Chapter 5

Source estimation with multiples

5.1 Summary

We propose a fast "wavelet-free" least-squares imaging procedure that produces high-accuracy seismic images without the knowledge of the source wavelet. Conventional reverse-time migration requires the knowledge of the source wavelet, which is either unavailable or very difficult to accurately determine; inaccurate estimates of the source wavelet can result in seriously degraded reverse-time migrated images, and therefore wrong geological interpretations. To solve this problem, we present a "wavelet-free" imaging procedure that simultaneously inverts for the source wavelet and the seismic image, by tightly integrating source estimation into a fast least-squares imaging framework, namely *compressive imaging*, given a reasonably accurate background velocity model. However, this joint inversion problem is difficult to solve as it is plagued with local minima and the ambiguity with respect to amplitude scalings, because of the multiplicative, and therefore nonlinear, appearance of the source wavelet in the otherwise linear formalism. We have found a way to solve this nonlinear joint-inversion problem using a technique called variable projection, and a way to overcome the scaling ambiguity by including surface-related multiples in our imaging procedure following recent developments in surface-related multiple prediction by sparse inversion. As a result, we obtain highly accurate estimates of the source wavelet and high-resolution seismic images, comparable in quality to

A version of this chapter has been submitted for publication

images obtained assuming the true source wavelet is known. By leveraging the computationally efficient compressive-imaging methodology, these results are obtained at affordable computational costs compared with conventional processing work flows that include surface-related multiple removal and reverse-time migration.

5.2 Introduction

Conventional reverse-time migration (RTM, Baysal et al., 1983) requires the knowledge of the source wavelet as prior information, which is used during the forward propagation of the source wavefield. Unfortunately, this knowledge is usually unavailable, or very difficult to accurately determine for field seismic data. Inaccurate estimates of the source wavelet will introduce errors to the forward propagated source wavefield, and later to the seismic image during the cross-correlation of the source and receiver wavefields. For example, wrong estimates in the phase of the source wavelet can result in misplaced structures in the seismic image, and therefore lead to wrong geological interpretations.

To eliminate the dependence of seismic imaging on the knowledge of the source wavelet, we are motivated to incorporate source estimation into seismic imaging via joint inversion of the seismic image and the source wavelet, by extending the least-squares-migration formalism (Nemeth et al., 1999). Because the forward modelling of the seismic data is linear with respect to the source wavelet, this type of joint inversion problem is known as the separate least-squares problem (Golub and Pereyra, 1973; Kaufman, 1975; Golub and Pereyra, 2003; Aravkin and van Leeuwen, 2012). To reduce the excessive simulation cost of conventional least-squares migrations, we leverage curvelet-domain sparsity of the seismic image (Herrmann et al., 2008a), and tightly integrate source estimation into a computationally efficient least-squares imaging formalism (Herrmann and Li, 2012), also known as the *compressive imaging* method.

Although a separable least-squares problem can usually be effectively solved using variable projection (Golub and Pereyra, 2003), applying variable projection to the proposed joint inversion procedure is complicated by the following two issues. First, the compressive imaging formalism adopts a sparsity-promoting objective function, which is different from typical separable least-squares problems (Golub and Pereyra, 2003). We follow van den Berg and Friedlander (2008) and Aravkin et al. (2013a), and derive an alternative problem formulation that turns the sparsity-promoting objective

into a sparse constraint. Second, there is an ambiguity in the amplitude scaling between the model perturbations and the estimated source wavelet, which cannot be resolved using variable projection alone, as in any blind-deconvolution type of problem (Stockham Jr et al., 1975). Inspired by recent developments in Estimation of Primaries by Sparse Inversion (EPSI, van Groenestijn and Verschuur, 2009a; Lin and Herrmann, 2013), we propose to resolve the ambiguity by incorporating surface-related multiples in the inversion. In this way we leverage the self-consistency between the primary events and their corresponding surface-related multiples, and obtain definite estimates of the source wavelet as well as the seismic image. Assuming negligible modelling errors, our method can recover the original amplitudes of the seismic image and the source wavelet with high fidelity. On a side note, as any other imaging algorithm, our proposed method relies on the knowledge of a reasonably accurate background model.

5.2.1 Related work and our contribution

The variable projection method has found its applications in a variety of geophysical problems in recent years, for example, in characterization of P-S wave conversion (Fomel et al., 2003), in velocity model building (van Leeuwen and Mulder, 2009), and notably in source estimation during full-waveform inversion (Aravkin et al., 2012; Rickett, 2013; Li et al., 2013). In all these applications, an objective function that measures the data misfit is used for the variable projection technique to be applicable. The misfit is usually measured using the ℓ_2-norm (i.e., a least-squares formulation), but other differentiable misfit functions can also be used (Aravkin et al., 2012). In this paper, however, we minimize the ℓ_1-norm of the solution vector to promote sparsity, with the least-squares data-fitting term acting as a constraint (Equation (5.2)). To the authors' knowledge, this work is the first instance of applying the variable projection technique to an optimization problem with a sparsity-promoting objective.

Regarding the use of multiples in source estimation, successful applications have been demonstrated in the literature of Surface-Related Multiple Elimination (SRME, Verschuur et al., 1992), and EPSI (van Groenestijn and Verschuur, 2009a; Lin and Herrmann, 2013; Esser et al., 2015). However, these approaches did not *(i)* recognize and tightly integrate the source estimation with variable projection; neither did they *(ii)* clearly demonstrate why incorporation of the surface-related multiples into sparsity-promoting imaging helps resolve the ambiguity in amplitude scalings that plague source estimations during imaging. This paper is the first, within the context of

seismic imaging, to properly address the above issues.

5.2.2 Paper outline

The paper is organized as follows. First, we formulate the joint sparsity-promoting source- and image-estimation problem, followed by a technical discussion on how to solve ℓ_1-norm minimization problems with source estimation via variable projection. Next, we explain why including surface-related multiples into the formulation helps us resolve the scaling ambiguity during imaging and source estimation. We conclude by demonstrating the advantages of the proposed method over conventional RTM, using carefully elaborated synthetic examples.

5.3 Problem formulation and method

As we propose the joint inversion approach in the compressive-imaging framework, we will first briefly review the compressive-imaging formalism and formulate the source estimation problem. Afterwards, we will explain how to solve the problem using variable projection.

5.3.1 Compressive imaging

Given the knowledge of the source wavelet, compressive imaging aims to obtain least-squares migrated images in a computationally efficient manner, by solving the following Basis Pursuit De-Noise (BPDN) problem (Herrmann and Li, 2012):

$$\begin{aligned} \text{BPDN}(\mathbf{w},\sigma): \quad & \min_{\mathbf{x}} \|\mathbf{x}\|_1 \\ & \text{subject to} \quad \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{d}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2 \le \sigma^2. \end{aligned} \tag{5.1}$$

In this formulation, vector $\mathbf{w}$ represents the Fourier spectrum of the source wavelet, and the tolerance parameter σ allows for data mismatch due to noise in the data and modelling errors. Vector $\mathbf{x}$ is the curvelet coefficients of the image $\delta\mathbf{m}$, i.e., $\delta\mathbf{m} = \mathbf{C}^*\mathbf{x}$ with $\mathbf{C}$ the curvelet synthesis operator (Candès and Donoho, 2004), and $\delta\mathbf{m}$ itself being the perturbations over a given gridded background velocity model $\mathbf{m}_0$ parameterized by the square of the slowness. In the data-fitting constraint of the above optimization program, i indexes the discretized frequencies, and j indexes the different source experiments. Instead of using all n_f discretized frequencies and n_s source

experiments, Herrmann and Li (2012) proposed to subsample the monochromatic source experiments to reduce the simulation cost in forward modelling: notations Ω and Σ denote the randomly selected $n'_f \ll n_f$ frequencies and $n'_s \ll n_s$ source experiments. By subsampling, the number of wave-equation solves in each iteration is reduced by a factor of $(n_f * n_s)/(n'_f * n'_s)$. We refer to Herrmann and Li (2012) and Chapter 2 on the details of the subsampling procedure. The underlined variables $\underline{\mathbf{d}}_{i,j}$ and $w_i\underline{\mathbf{s}}_j$ represent the jth column of the subsampled *primary-only* receiver wavefields (i.e., *simultaneous* data) and *point*-source wavefields (i.e., *simultaneous* sources) respectively. The operator $\nabla\mathbf{F}$ represents the linearized Born scattering operator. Note that the source weight w_i is separable from the source vector $\underline{\mathbf{s}}_j$, i.e., $\nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j]\delta\mathbf{m} = w_i\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\delta\mathbf{m}$. Throughout this paper, we form the operator $\nabla\mathbf{F}$ based on the two-way wave equation, therefore the evaluation of the adjoint of $\nabla\mathbf{F}$ is equivalent to the reverse-time migration (RTM, Baysal et al., 1983; Lailly, 1983). The above formulation means that among all possible solutions, the optimization program looks for the sparsest curvelet coefficients that, after curvelet synthesis and linearized modelling with the simultaneous sources, predict the simultaneous observed data up to a noise threshold specified by σ. With the above compressive-imaging formalism in place, we now continue to formulate the source estimation problem.

5.3.2 Source estimation

Now we identify the source wavelet $w_i, i \in \Omega$ (collected in the vector $\mathbf{w}$) as an additional unknown in the compressive imaging formulation, i.e., we have

$$\begin{aligned} \text{BPDN}(\sigma): \quad & \min_{\mathbf{x},\mathbf{w}} \|\mathbf{x}\|_1 \\ & \text{subject to} \quad \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{d}}_{i,j} - w_i\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2 \le \sigma^2. \end{aligned} \tag{5.2}$$

While conceptually attractive, BPDN(σ) does not lend itself to develop tractable algorithms. Following early work by van den Berg and Friedlander (2008), we interchange the objective and constraint to arrive at an extension—remember the formulation now includes the source as an unknown, which appears nonlinearly in the BPDN(σ) program, of the Least Absolute Shrinkage and Selection Operator (LASSO, Tibshirani (1996)) for-

mulation:

$$\text{LASSO}(\tau): \quad \min_{\mathbf{x},\mathbf{w}} f(\mathbf{x},\mathbf{w}) \doteq \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{d}}_{i,j} - w_i \nabla \mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau. \tag{5.3}$$

As shown by Aravkin et al. (2013a), the set of minimizers of problem BPDN(σ) and LASSO(τ) coincide, given the condition that each minimizer of problem BPDN(σ) satisfies its constraint to equality (i.e., $f(\mathbf{x},\mathbf{w}) = \sigma^2$). Not coincidently, when this condition is met, each minimizer also satisfies the constraint of the LASSO(τ) problem to equality, meaning that we can look for such a τ that the infimum of $f(\mathbf{x},\mathbf{w})$ with $|\mathbf{x}|_1 \leq \tau$ has a value that equals σ^2. Because the objective function of problem LASSO(τ) adopts the canonical form of a separable least-squares problem (Golub and Pereyra, 2003), we can now solve the problem using variable projection combined with projection onto a convex set. In the next few sections, we will first discuss how we solve problem LASSO(τ) with a given τ, and then discuss how we compute the right τ so that problem LASSO(τ) converges to the same solution as problem BPDN(σ).

5.3.3 Variable projection

To incorporate variable projections for the source wavelet into ℓ_1-norm sparsity promoting imaging, let us first consider iterations that involve soft thresholding that undergirds this type of ℓ_1-norm optimization. Model iterates at the k^{th} iteration are in this case projected onto the ℓ_1-norm ball $\|\mathbf{x}\|_1 \leq \tau$, and the involved projected-gradient step is given by

$$\mathbf{x}^{k+1} = \mathrm{P}_{\mathcal{X}}\left[\mathbf{x}^k + \lambda \nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{w})\Big|_{\mathbf{x}=\mathbf{x}^k,\mathbf{w}=\mathbf{w}^k}\right]. \tag{5.4}$$

In this expression, λ is a line-search parameter, $\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{w})\Big|_{\mathbf{x}=\mathbf{x}^k,\mathbf{w}=\mathbf{w}^k}$ is the gradient of the objective function in Equation (5.3) with respect to $\mathbf{x}$ at the k^{th} iteration, and $\mathrm{P}_{\mathcal{X}}$ denotes the projection onto the feasible set $\mathcal{X} = \{\mathbf{x} : \|\mathbf{x}\|_1 \leq \tau\}$.

Contrary to conventional ℓ_1-norm promoting imaging where $\mathbf{w}$ is assumed to be known, we need to evaluate this gradient $\nabla_{\mathbf{x}} f(\mathbf{x},\mathbf{w})\Big|_{\mathbf{x}=\mathbf{x}^k,\mathbf{w}=\mathbf{w}^k}$ with variable projection at each iteration for unknown $\mathbf{w}^k$'s. During each

iteration, we accomplish this by solving the following optimization problem for each frequency:

$$\min_{w_i} \quad \sum_{j\in\Sigma} \|\underline{\mathbf{d}}_{i,j} - w_i \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2, \tag{5.5}$$

which permits the following closed-form solution for the source wavelet (Pratt, 1999):

$$\widetilde{w}_i(\mathbf{x}) = \frac{\sum_{j\in\Sigma}\langle \underline{\mathbf{d}}_{i,j}, \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}{\sum_{j\in\Sigma}\langle \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}, \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}. \tag{5.6}$$

Here, the angular brackets $\langle\cdot,\cdot\rangle$ denote the inner product of two vectors, and $\widetilde{w}_i(\mathbf{x}), i \in \Omega$ denote complex-valued estimates of the source wavelet at different angular frequencies. The above solution enables us to eliminate the unknown source wavelet term in problem LASSO(τ), rendering the problem to solely and nonlinearly depend on $\mathbf{x}$:

$$\begin{aligned} \min_{\mathbf{x}} \overline{f}(\mathbf{x}) &\doteq \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{d}}_{i,j} - \widetilde{w}_i(\mathbf{x})\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2 \\ \text{subject to} \quad & \|\mathbf{x}\|_1 \le \tau. \end{aligned} \tag{5.7}$$

At first inspection, the above problem becomes more complicated than LASSO(τ) as we need to evaluate the derivative of $\widetilde{w}_i(\mathbf{x}), i \in \Omega$ with respect to $\mathbf{x}$. However, Aravkin and van Leeuwen (2012) (Corollary 2.3) have proved that a stationary point of the above problem remains a stationary point of problem LASSO(τ). Furthermore, we have (Aravkin and van Leeuwen, 2012, Theorem 2.1)

$$\nabla_{\mathbf{x}}\overline{f}(\mathbf{x}) = \nabla_{\mathbf{x}} f(\mathbf{x}, \widetilde{\mathbf{w}}(\mathbf{x})). \tag{5.8}$$

The above equation means that in each iteration, after vector $\mathbf{x}$ is modified according to Equation (5.4), we can compute $\nabla_{\mathbf{x}}\overline{f}(\mathbf{x})$ by evaluating $\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{w})$ of problem LASSO(τ) at $\mathbf{w} = \widetilde{\mathbf{w}}(\mathbf{x})$ via Equation (5.6).

While there is no guarantee the variable projection approach will yield the globally optimal solution as the problem remains non-convex, numerical experiments show that in most cases it enables the optimization program to converge within less iterations (Golub and Pereyra, 2003). Rickett (2013) also showed its superior performance over the simultaneous descent method in source estimation for full-waveform inversion applications. Furthermore, the projections themselves (cf. Equation (5.6)) are computationally afford-

able as they do not involve additional expensive operations such as wave-equation solves.

5.3.4 Relaxing the ℓ_1-norm constraint

In this section, we describe our strategy to select the right τ so we actually solve problem BPDN(σ) by solving problem LASSO(τ). We follow the methodology proposed by van den Berg and Friedlander (2008) in SPGℓ_1, where a series of τ's are found by solving a root-finding problem on the Pareto trade-off curve. With the unknown source wavelet, we find the τ's by solving the value function $\nu(\tau) \doteq \inf f(\mathbf{x}, \mathbf{w})|_{||\mathbf{x}|_1 \leq \tau} = \sigma^2$, which yields the following ℓ_1-norm constraint for the $(l+1)^{th}$ LASSO subproblem using the Newton's method with an initial guess of $\tau^0 = 0$:

$$\tau^{l+1} = \tau^l - \frac{\nu(\tau^l) - \sigma^2}{\nu'(\tau^l)}. \tag{5.9}$$

Using this approach, we arrive at the solution of BPDN(σ) by solving a series of LASSO(τ) subproblems for gradually increasing τ's. While $\nu(\tau^k)$ in the above equation can be evaluated straightforwardly with the previous τ estimate, the evaluation of $\nu'(\tau^k)$ is more difficult, because the nonlinearity introduced by the unknown source wavelet violates the linearity assumption on the forward model, which is needed to compute this value (Aravkin et al., 2013a). However, by treating $\mathbf{w}$ as fixed, we approximate $\nu'(\tau^k)$ by (van den Berg and Friedlander, 2008; Aravkin et al., 2013a):

$$\nu'(\tau^k) \approx -\| \sum_{i \in \Omega} \sum_{j \in \Sigma} \mathbf{C} \nabla \mathbf{F}^*[\mathbf{m}_0, w_i \underline{\mathbf{s}}_j] \left(\underline{\mathbf{d}}_{i,j} - \nabla \mathbf{F}[\mathbf{m}_0, w_i \underline{\mathbf{s}}_j] \mathbf{C}^* \mathbf{x} \right) \|_\infty. \tag{5.10}$$

With numerical examples, we will show that this approximation works reasonably well.

5.3.5 Acceleration by redrawing random subsets

Except for the extra variable projection step, solving LASSO(τ) is similar to compressive imaging (Herrmann and Li, 2012), during which computational costs can be reduced by working with randomized subsets of data. As shown by Herrmann and Li (2012) and Chapter 2, solutions of BPDN($\mathbf{w}; \sigma$) for a given source wavelet can be accelerated significantly by selecting new independent randomized subsets of frequencies and simultaneous sources, after each corresponding LASSO subproblem is solved. We found empiri-

cally that working with these new independent subsets leads to faster decay of the model error and to improved robustness with respect to possible linearization errors (Chapter 3)—i.e., $\underline{\mathbf{d}}_{i,j} - \mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j] \approx \nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j]\delta\mathbf{m}$. Both findings can be explained because independent subsets tend to break correlations between errors in the solution and the randomly selected subsets of data (Herrmann, 2012).

While these empirical findings lead to a computationally feasible and robust compressive imaging scheme, we need to justify that working with different randomized subsets does not interact adversely with the proposed source estimation approach. Aside from additional empirical evidence from various examples included below, which suggests that there is no measurable adverse effect, we argue that *(i)* numerical of the value function, $\nu'(\tau)$, and its derivative $\nu'(\tau)$ remain similar as long as we keep the number of frequencies and simultaneous sources the same; and *(ii)* we only draw new subsets after each LASSO(τ^l) is solved. The latter argument assumes that estimates of the source at the l^{th} subproblem have converged, which is a reasonable to make because the number of knowns in Equation (5.6) for each frequency far exceeds the dimensionality of the single unknown complex-value for the source.

5.4 Resolving scaling ambiguities with multiples

Despite the fact that there are indications that we arrived at a practical algorithm to estimate source wavelet as well as high-resolution least-squares images, fundamental issues related to the ambiguity with respect to amplitude scalings remain, which are intrinsic to any blind-deconvolution type of problems where both the source wavelet and the seismic image are unknown (Stockham Jr et al., 1975). However, compared to source estimation in data space using the convolution model (Ulrych et al., 1995), source estimation in the image space does not suffer from ambiguities with respect to global phase shifts, which can be explained by the multiplicity of data and move-out characteristics that are specific to the background-velocity model being used. As stated in the introduction, we assume this background model to be given and to be relatively accurate.

Unfortunately, the same observation does not apply to the ambiguity with respect to amplitude scalings, which correspond mathematically to an invariance of our objective functions with respect to amplitude scalings by

$\alpha \in \mathbb{R}^+$—i.e., we have

$$f(\mathbf{x}, \mathbf{w}) = f(\frac{1}{\alpha}\mathbf{x}, \alpha\mathbf{w}). \tag{5.11}$$

This invariance results in amplitude ambiguity in the source estimation, which originates from the fact that the objective of LASSO(τ) is bi-linear, namely linear in both the source wavelet $\mathbf{w}$ and the solution vector $\mathbf{x}$.

Unfortunately, including sparsity constraints via the ℓ_1-norm or even via the non-convex ℓ_1/ℓ_2-norm (Esser et al., 2015), are inadequate to resolve the above amplitude ambiguity. We therefore alternatively resort to the physics of the free surface. By incorporating the free surface, our problem becomes nonlinear because the forward model now includes a "feedback loop", which generates surface-related multiples from primaries scaled by the (unknown) source wavelet. Extending the physics in this way allows us to resolve the scaling ambiguity by using observed surface-related multiples explicitly to estimate the source wavelet as proposed during EPSI (van Groenestijn and Verschuur, 2009a; Lin and Herrmann, 2013). Following recent work shown in Chapter 2, we propose to do this by incorporating predictions of surface-related multiples into our formulation for compressive imaging.

5.4.1 Compressive imaging with total upgoing wavefields

If we ignore internal multiples, the total upgoing wavefield, including primaries and surface-related multiples, can be modelled by including the observed upgoing wavefields $\mathbf{u}_{i,j}$ at the water surface as areal sources in the linearized Born scattering operator (Chapter 2), yielding

$$\mathbf{u}_{i,j} \approx \nabla\mathbf{F}[\mathbf{m}_0, w_i\mathbf{s}_j - \mathbf{u}_{i,j}]. \tag{5.12}$$

For a perfect free surface with a reflection coefficient of -1, the downgoing spatially impulsive source wavefields $w_i\mathbf{s}_j$ are replaced by "areal" sources $w_i\mathbf{s}_j - \mathbf{u}_{i,j}$, i.e., we include the downgoing receiver wavefields at the water surface $-\mathbf{u}_{i,j}$ into the source wavefields. We obtain $\mathbf{u}_{i,j}$ after careful preprocessing to the observed seismic data, such as applying source-receiver reciprocity (to fill in missing traces), up-down decompositions (for receiver-side deghosting), and extrapolation of the upgoing wavefield from the receiver level to the free surface (see e.g., Verschuur et al., 1992). As a result, we arrive at a formulation that remains conducive to sparsity-promoting imaging with source estimation via variable projection. Because prediction for the multiples are carried out by the wave-equation solver, this forward model is

also computationally viable compared to processing work flows that involve separate multiple-prediction and RTM-imaging procedures.

To arrive at our final formulation, we first replace our objective by

$$f(\mathbf{x},\mathbf{w}) = \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j - \underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}\|_2^2, \tag{5.13}$$

which we obtain by substituting the primary wavefield with the total upgoing wavefield and the spatially impulsive sources with areal sources containing the total downgoing wavefield at the surface. Given this new definition for the objective, we proceed by solving the LASSO(τ) subproblems with variable projections. For this end, we solve for all (simultaneous) sources and for each frequency the following problem:

$$\min_{w_i} \quad \sum_{j\in\Sigma} \|\underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, -\underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x} - w_i\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2, \quad i\in\Omega,$$

which in turn permits the following analytic solution:

$$\tilde{w}_i(\mathbf{x}) = \frac{\sum_{j\in\Sigma}\langle \underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, -\underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}, \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}{\sum_{j\in\Sigma}\langle \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}, \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}. \tag{5.14}$$

To arrive at the modified solution for the source wavelet above, we substitute the primary-only wavefield by an estimate for the primaries given by the total total upgoing wavefield minus the current prediction for the surface-related multiples—i.e., $\underline{\mathbf{d}}_{i,j} \mapsto \underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, -\underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}$. The multiple predictions do not require information on the source wavelet; rather, they are carried out by solving wave-equations with the areal sources given by $-\underline{\mathbf{u}}_{i,j}$. With these substitutions, we are able to estimate the source wavelet by variable projection at additional costs of a single linearized forward modelling operation to predict the multiples.

In return of the increased computational costs, including surface-related multiples during imaging with source estimation has several key advantages, namely *(i)* including the multiples resolves ambiguities in the source estimation, which leads to a correct scaling of the image and therefore of the predicted primaries and multiples; *(ii)* sparsity-promoting imaging with multiples removes coherent artifacts related to surface-related multiples by mapping multiple energy onto primaries and therefore onto the true image; and *(iii)* our inversion procedure properly accounts for the source wavelet resulting in true-amplitude source-deconvolved high-resolution images (Chapter 2). This latter claim of course hinges on the assumption that amplitudes

are correctly modelled by the linearization.

5.4.2 Putting it all together

With the building blocks of our sparsity-promoting imaging with source estimation and multiples in place, we summarize our proposed imaging scheme in Algorithm (2).

Input and initialization [Line 1-6]

As discussed above, our proposed inversion approach requires the knowledge of a reasonably accurate background velocity model $\mathbf{m}_0$. As discussed in Chapter 3, we choose a zero σ to prevent the optimization algorithm from being terminated prematurely (Line 3). The initial guess of the source wavelet is simply an impulse at zero time, which corresponds to a flat Fourier spectrum with unit amplitude and zero phase (Line 6). Both the solution vector $\mathbf{x}$ and the sparsity level τ are simply initialized as zeros.

Main loop [Line 7-16]

In the main loop we solve a series of LASSO(τ^l) subproblems. For each subproblem, we draw new independent subsets of randomized frequencies and simultaneous sources (Line 8). We update the sparsity parameter τ^l using Newton's method (Line 9), and solve for the model parameters collected in vector $\mathbf{x}$ (Line 10) as well as the source wavelet collected in vector $\mathbf{w}$ (Line 13) by variable projection. Note that we do not impose any assumption on the phase of the source wavelet (e.g., minimum phase assumption as used in Robinson, 1967) during source estimation. We terminate the optimization program when the pre-specified maximal number of iterations $k_{\max}$ is reached.

5.5 Examples

To validate and test the proposed imaging scheme with on-the-fly source estimation, we conduct a series of synthetic examples designed to illustrate key aspects of our algorithm including sensitivity to errors in the source wavelet, robustness with respect to modelling errors, and resolving ambiguity with multiples.

Algorithm 2 Fast imaging with source estimation and multiples

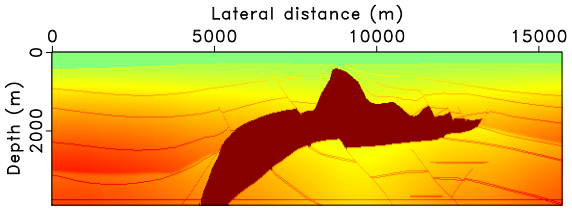
1. **Input:**
2. total upgoing wavefield $\mathbf{u}$, background velocity model $\mathbf{m}_0$,
3. tolerance $\sigma = 0$, iteration limit $k_{\max}$
4. **Initialization:**
5. iteration index $k \leftarrow 0$, LASSO subproblem index $l \leftarrow 0$, $\tau^0 \leftarrow 0$, $\mathbf{x}^0 \leftarrow \mathbf{0}$
6. $w_i = 1$ for all $i \in 1, \cdots, n_f$
7. **while** $k < k_{\max}$ **do**
8. $\quad \Omega_l$, Σ_l, $\underline{\mathbf{u}}_{i,j}$, $\underline{\mathbf{s}}_j \leftarrow$ new independent draw
9. $\quad \tau^l \leftarrow$ determined from τ^{l-1} and σ using Newton's method
10. $\quad \mathbf{x}^l \leftarrow \begin{cases} \text{argmin}_{\mathbf{x}} \sum_{i \in \Omega_l, j \in \Sigma_l} \|\underline{\mathbf{u}}_{i,j} - \nabla \mathbf{F}_i[\mathbf{m}_0, w_i(\mathbf{x})\underline{\mathbf{s}}_j - \underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}\|_2^2 \\ \text{subject to } \|\mathbf{x}\|_1 \le \tau^l \end{cases}$
11. $\quad\quad$ //warm start with $\mathbf{x}^{l-1}$, solved in k_l iterations
12. $\quad\quad$ //in each iteration, update the source by
13. $\quad\quad w_i(\mathbf{x}) = \frac{\sum_{j \in \Sigma} \langle \underline{\mathbf{u}}_{i,j} - \nabla \mathbf{F}[\mathbf{m}_0, -\underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}, \nabla \mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x} \rangle}{\sum_{j \in \Sigma} \langle \nabla \mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}, \nabla \mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x} \rangle}$
14. $\quad k \leftarrow k + k_l, l \leftarrow l + 1$
15. **end while**
16. **Output:** model perturbations $\delta\mathbf{x} = \mathbf{C}^*\mathbf{x}$

5.5.1 Imaging with primaries only

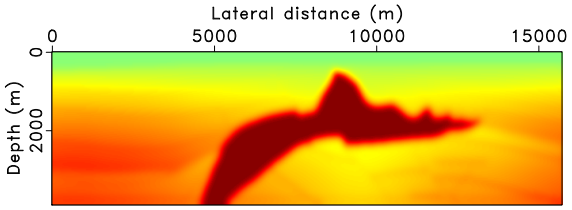
Most conventional seismic data processing work flows use the primary wavefield as the input for migration, i.e., after a de-multiple procedure to the observed data, for example using SRME (Verschuur et al., 1992). We therefore first apply our method to primary-only data to demonstrate its advantages over RTM when used in conventional data work flows.

Experiment setup

We use a 2D slice of the SEG/EAGE salt model. We pad 10 grid points at the top of the model so that the water layer becomes slightly thicker, to better retain the water velocity near the surface when we smooth the true model to get the background model. In this way we can model and therefore remove the direct waves more accurately. The true and the background velocity models are shown in Figure 5.1. The model is 3.9 km deep and 15.7 km long, with a grid spacing of 24.38 m (80 ft). We use a fixed-spread configuration, and put 323 co-located sources and receivers with a 48.77 m lateral grid spacing (i.e., every other grid point) at a depth of 24.38 m (i.e., the second vertical grid point). We use a Ricker wavelet with a peak



(a)



(b)

Figure 5.1: A slightly modified 2D slice of the SEG/EAGE salt model. **(a)** True velocity model. **(b)** Smoothed background velocity model.

frequency of 5 Hz, and the peak of the wavelet is at 0.25 s. We record for 8 seconds, which in the frequency domain yields 96 discretized frequencies up to 12 Hz.

Ideal case

To rule out the influence of any imperfection in the data, we first test our method with an ideal setup, i.e., we synthesize and invert the primary-only data using the same linearized modelling engine. To obtain a baseline image, we run a conventional RTM by fictitiously assuming that we know the true source wavelet (Figure 5.2(a)). To remove the low-wavenumber artifacts in the RTM image, we apply a high-pass filtering along the depth dimension (Mulder and Plessix, 2003). To verify the necessity of the proposed on-the-fly source estimation procedure, we first show the detrimental effect of inverting with a wrong wavelet (Figure 5.2(b)). Next, we compare it with using the true source wavelet (Figure 5.2(c)), and then we show how to avoid the adverse effect wavelet using the proposed method (Figure 5.2(d)). The wrong source wavelet we use is obtained by introducing a time shift of -0.1 s (the minus sign indicates an advance in time) to the true source wavelet.

We can see that compared with conventional RTM (Figure 5.2(a)), the proposed approach produces an image of higher spatial resolution given the true source wavelet (Figure 5.2(b)). However, the image quality is significantly compromised when the wavelet is wrong. In this case the shifted phase of the wavelet not only produces an image contaminated with subsampling-related noises (Figure 5.2(c)), but also leads to a shift of the subsurface structures in depth, which can result in erroneous interpretations (although not shown here, the structures in the RTM image using the wrong wavelet are also shifted in depth). With the proposed source-estimation approach, we obtain a faithful image (Figure 5.2(d)), as if we use the true source wavelet (Figure 5.2(b)).

Remarks on the experiment: In terms of computational cost, we use 16 randomly selected frequencies, 16 simultaneous sources, and 60 iterations for all inversion results; as a result, the four images in Figure 5.2 involve roughly the same number of wave-equation solves. For all inversion results in this and later sections, after the inversion is finished, we apply a curvelet thresholding to remove residual incoherent noises in the image (Herrmann et al., 2008b). The threshold is chosen in such a way that the thresholded noise does not contain noticeable coherent energy. As our solution vector $\mathbf{x}$ is already in the curvelet domain, extra computation incurred in this thresholding step is minimal.

As a quality-control procedure, we also plot in Figure 5.3 the estimates of the source wavelet outputted by the last LASSO(τ) subproblem (in green), and compare it against the true source wavelet (in blue). Both amplitude and phase spectra are plotted. Because of the amplitude ambiguity we discussed above, we normalize the amplitude spectra of both the true and the estimated source wavelet to compare them. The source estimates are highly accurate except for the absolute amplitude scaling.

More realistic case

To test the robustness of our method to imperfections in the data — such as linearization errors (as explained earlier), internal multiples, and discrepancies between the modelling engines used for data simulation and inversion — we now apply the method to a more realistic primary-only data set modelled with iWave (Terentyev et al., 2014). We compare the following examples. For comparison, we again get the baseline image using conventional RTM with the true wavelet (Figure 5.4(a)). Next we obtain the image using the proposed method with the true source wavelet (Figure 5.4(b)) and with source estimation (Figure 5.4(c)). In Figure 5.5 we plot the amplitude (with

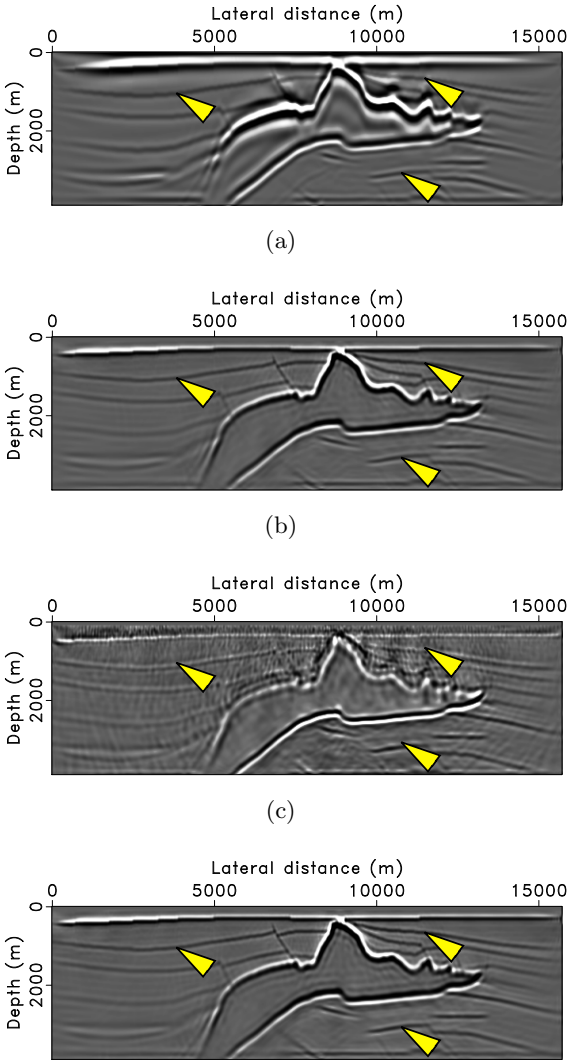
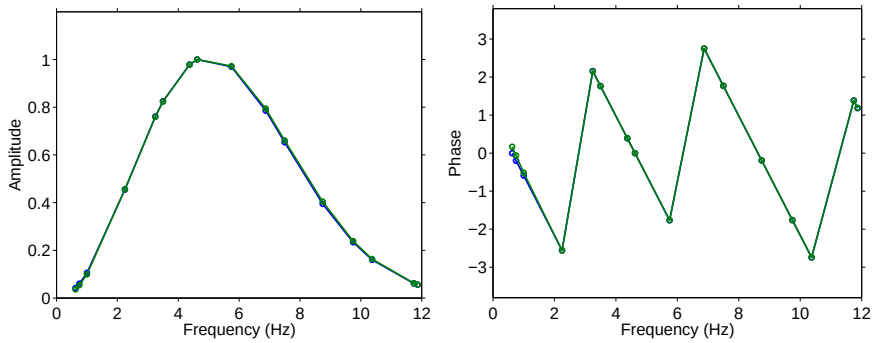


Figure 5.2: Imaging results of the ideal primary-only data set. The arrows point to the same subsurface locations. **(a)** Conventional RTM with the true source wavelet. **(b)** to **(d)** Compressive imaging results with the true source wavelet **(b)**, with the wrong source wavelet that has a -0.1s time shift **(c)**, and with the proposed source-estimation method **(d)**. It is evident that the images in **(b)** and **(d)** are highly comparable, and are of much higher spatial resolution than the conventional RTM image in **(a)**.



(a) (b)

Figure 5.3: Source estimates from the ideal primary-only data set. **(a)** and **(b)**: Amplitude **(a)** and phase **(b)** spectra of the true (in blue) and the estimated (in green) source wavelet during the last LASSO(τ) subproblem, after amplitude normalization.

normalization) and phase spectra of the true wavelet and the estimated wavelet during the last LASSO(τ) subproblem.

We observe that, in the presence of the imperfections in the data, *(i)* the proposed source-estimation method still yields an image that highly resembles the one obtained with the true source wavelet; *(ii)* both images are still of higher spatial resolution than the conventional RTM image; and *(iii)* the source estimates degrade gracefully, compared with the case where we invert the ideal data (see Figure 5.3). The imperfections do result in some imaging artifacts beneath the salt structure, but they do not particularly cause problems for the source-estimation approach. The above observations lead to the conclusion that the proposed source-estimation approach is relatively robust with respect to coherent noises in the primary-only data.

Remarks on the experiment: In generating the data with iWave, we use an absorbing boundary condition at the surface to avoid generating surface-related multiples. The observed data is the difference between the data modelled with the true model and the background model. We use the same inversion parameters as the above set of examples, and as a result, the computational costs of the three images in Figure 5.4 remain roughly the same.

As demonstrated by the two sets of examples above, our proposed source estimation method, when applied to conventional primary-only data, produces seismic images that are comparable to images obtained using true

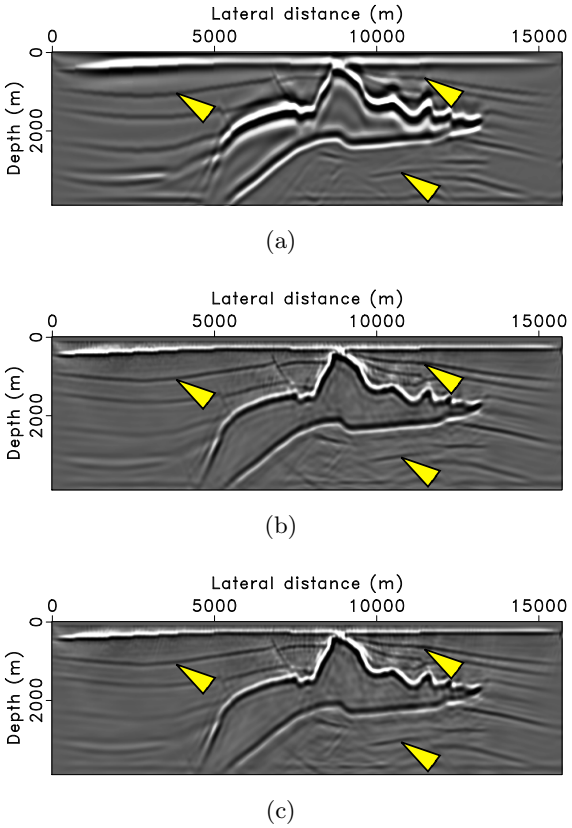
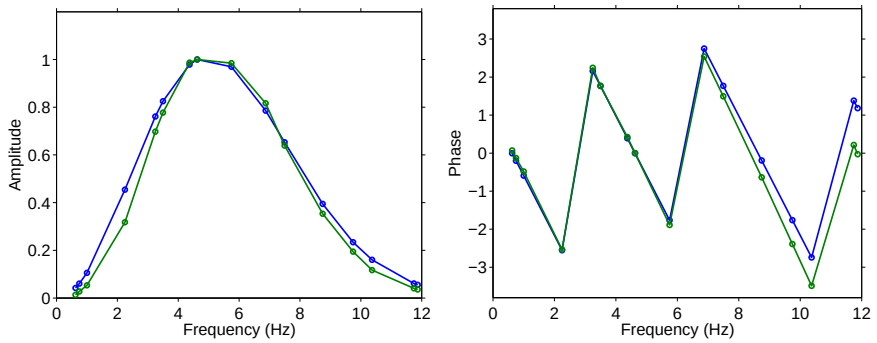


Figure 5.4: Imaging results of the primary-only data set modelled using iWave. The arrows point to the same subsurface locations. **(a)** Conventional RTM image. **(b)** and **(c)** Compressive imaging results using the true source wavelet **(b)** and with the proposed source-estimation method **(c)**.

source wavelets. The method is also relatively robust to imperfections in the input data. However, as we discussed above, the scaling ambiguity cannot be solved by imaging with primaries alone because of the blind-deconvolutional nature of the problem. Next we will demonstrate how we resolve the ambiguity using surface-related multiples.



(a) (b)

Figure 5.5: Source estimates from the primary-only data set modelled using iWave. **(a)** and **(b)**: Amplitude **(a)** and phase **(b)** spectra of the true (in blue) and the estimated (in green) source wavelet during the last LASSO(τ) problem, after amplitude normalization.

5.5.2 Scaling images with multiples

To validate our proposal to incorporate surface-related multiples in the imaging procedure to resolve the amplitude ambiguity, we apply the proposed method to data that contain surface-related multiples. As above, we study the performance of the proposed approach both with an ideal data set and with a more realistic data set. The initial guess for the source wavelet remains an impulse at $t = 0$.

Experiment setup

We use a sedimentary part of the Sigsbee 2B model (The SMAART JV, 2014), which by design has a strong ocean-bottom reflector, to carry out examples using surface-related multiples. We reduce the water depth of the original model, to allow for more orders of multiples to be recorded. The true and the background velocity models are shown in Figure 5.6. The model is 3.8 km deep and 6 km long, with 7.62 m (25 ft) grid spacing. We again use a fixed-spread configuration, and put 261 co-located sources and receivers with 22.86m lateral grid spacing (i.e., every three grid point) at a depth of 15.24 meters (i.e., the third vertical grid point). We use a zero-phase Ricker wavelet with a peak frequency of 15 Hz, and record for 8.184 seconds (i.e., 1024 time samples with 8 ms sampling interval). This setup yields 311

discretized frequencies up to 38 Hz.

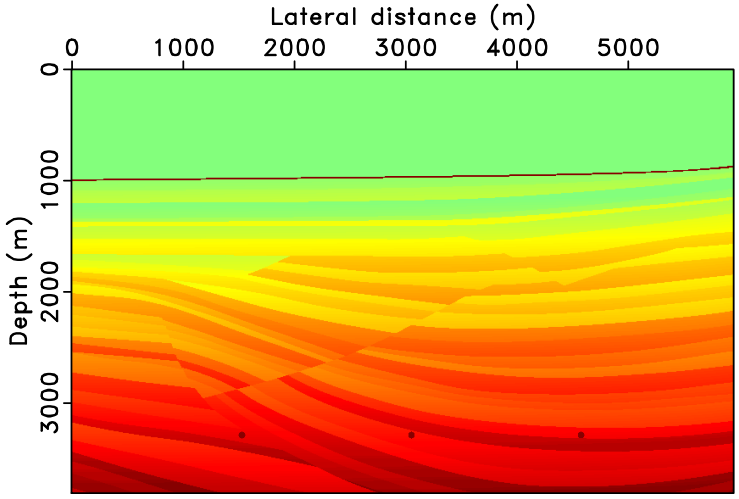
Ideal case

We synthesize the ideal data set including surface-related multiples using linearized modelling following Equation (5.12). Note that in this case, the multiples are not generated by including an explicit free surface in the model, but by injecting the downgoing receiver wavefield as areal sources. To show the removal of amplitude ambiguity, we compare the imaging results obtained with source estimation, including both the image and estimates of the source wavelet during the last LASSO(τ) problem, against the true model and source wavelet. The inverted model perturbations are shown in Figure 5.7(a). To show that it is indeed of the true amplitude and phase, we add it back to the background model to estimate the true model. The result is shown in Figure 5.7(b), and is comparable to the true velocity model shown in Figure 5.6(a). We juxtapose the amplitude and phase spectra of the true and estimated source wavelets in Figure 5.8, without applying any normalization. This comparison shows that in this ideal case, the proposed imaging-with-multiple approach can retrieve the original amplitude (and phase) of both the model perturbations and the source wavelet with high accuracy.

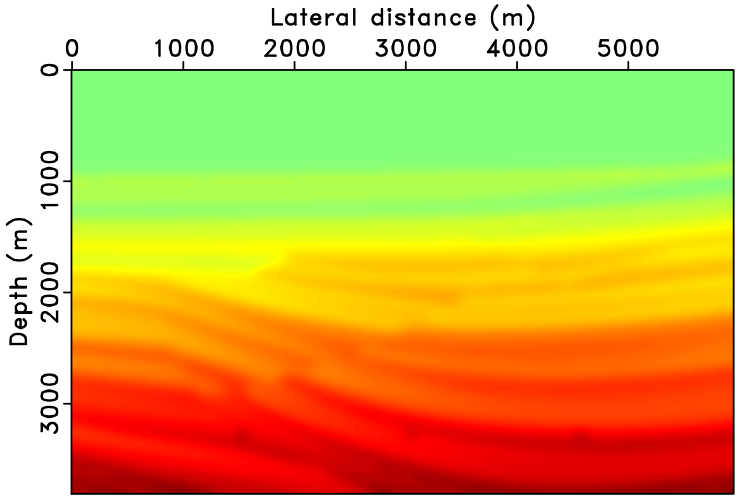
Remarks on the experiment: In the inversion, we use 31 randomly selected frequencies, 26 simultaneous sources, and 50 iterations. As a result, the number of wave-equation solves is roughly 1.5$\times$ of one RTM with all the data (as discussed above, source estimation with multiples introduces a 50% increase in simulation cost for each iteration).

More realistic case

We model a more realistic data set that contains multiples using iWave with a free-surface boundary condition (Terentyev et al., 2014). The observed data is the difference between the data modelled with the true model and the background model. Note that the imperfections in the data now also originate from approximating the observed surface-related multiples by the SRME-type multiple prediction, although this prediction is carried out implicitly using wave-equation solvers in our proposed approach (Chapter 2). Because the modelling engines used in data simulation and inversion have different scalings, we can no longer recover the original amplitude. Therefore we compare the images obtained with the true source wavelet (Figure 5.9(a)) and with source estimation (Figure 5.9(b)). The inversion parameters and

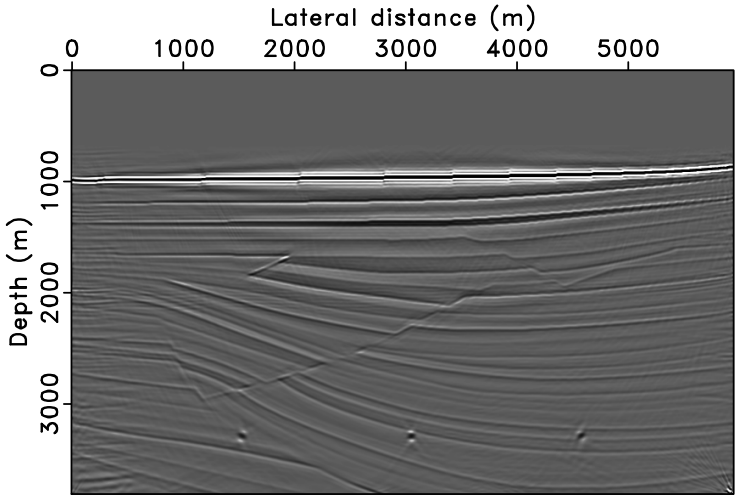


(a)

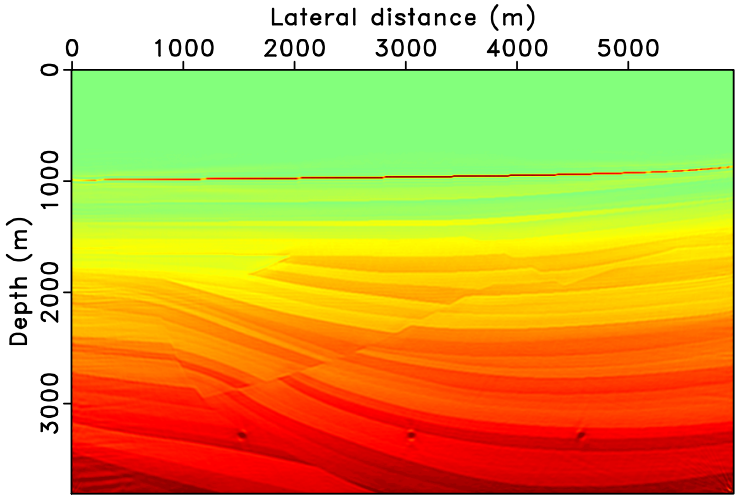


(b)

Figure 5.6: A sedimentary part of the Sigsbee 2B model. **(a)** True velocity. **(b)** Smooth background velocity.

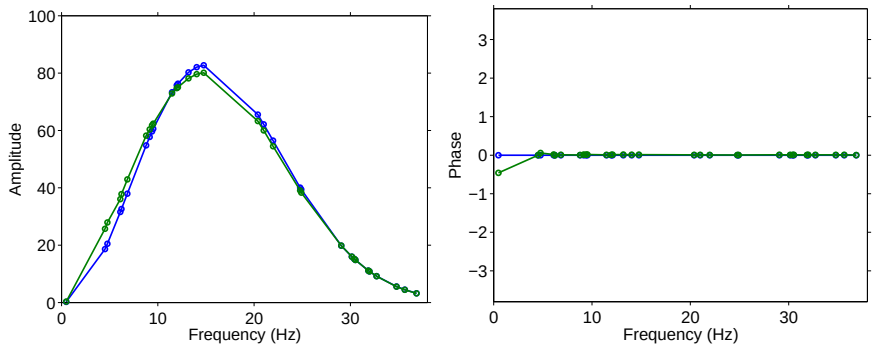


(a)



(b)

Figure 5.7: Compressive imaging-with-multiple result of the ideal data set containing multiples. **(a)** The inverted model perturbations. **(b)** Estimate of the true velocity model by adding the inverted model perturbations back to the background model.



(a) (b)

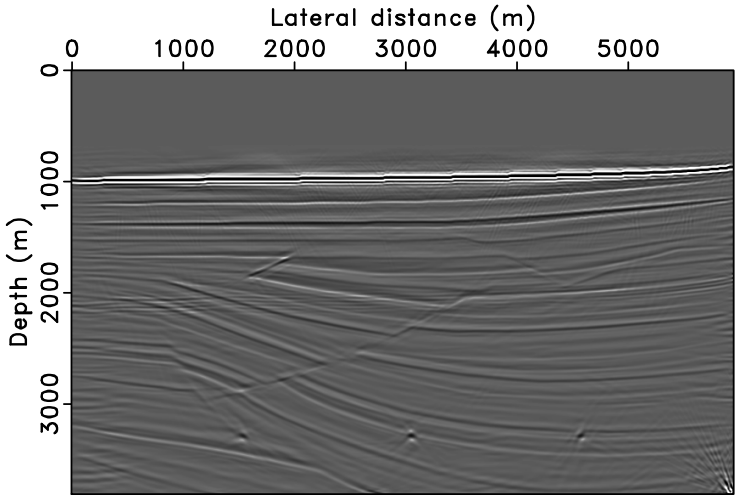
Figure 5.8: Source estimates from the ideal data set containing multiples. **(a)** and **(b)**: Amplitude **(a)** and phase **(b)** spectra of the true (in blue) and the estimated (in green) source wavelet during the last LASSO(τ) subproblem. No normalization is applied.

therefore the computational costs remain the same as the above set of examples. The two images are comparable, both containing minimal coherent artifacts from surface-related multiples (e.g., periodic copies of the ocean bottom at deeper section of the image).

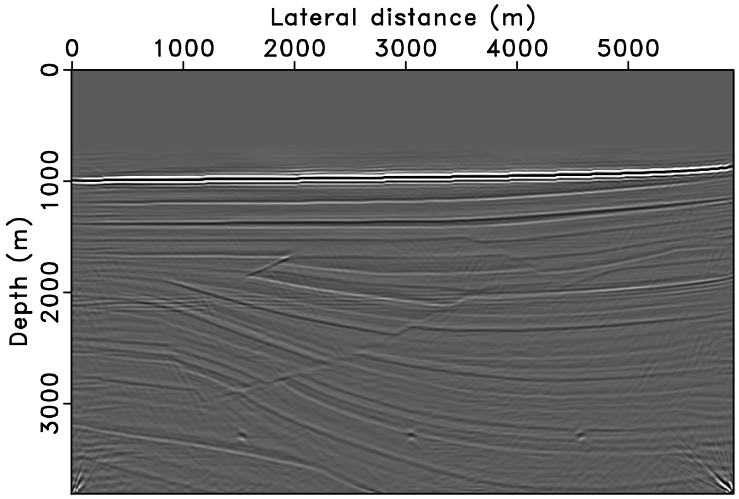
We plot the amplitude and phase spectra of the true and the estimated source wavelets, in Figure 5.10(a) and 5.10(b). Because we do not recover the original amplitude, we need to normalize the amplitude spectra. We observe that in the presence of imperfections in the data, the estimated source wavelet again degrades gracefully, as in the primary-only case. The imperfections in the data do cause the the estimated source wavelet to be biased towards low frequencies, which is presumably caused by the "obliquity" factor (Dragoset and Jeričević, 1998), and can be mitigated by applying the absolute value of the angluar frequencies, i.e., $|\omega|$. We will discuss about the biased source estimates in the Disucssion section with a more detailed analysis.

5.5.3 Convergence analysis

To obtain a more quantitative understanding of the proposed method, we study and compare the convergence behaviors of the recovery algorithm with the true wavelet and with source estimation. We use the normalized cross-correlation (NCC) to measure how well the inverted results approximate true

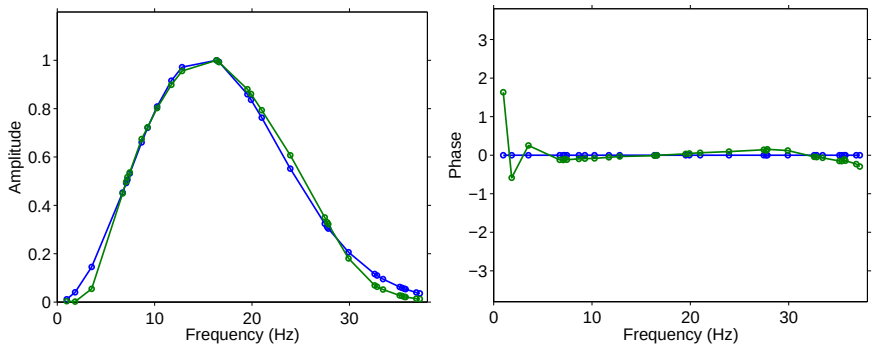


(a)



(b)

Figure 5.9: Compressive imaging-with-multiples results of the data set modelled using iWave. **(a)** and **(b)**: Image obtained with the true source wavelet **(a)** and with source estimation **(b)**.



(a) (b)

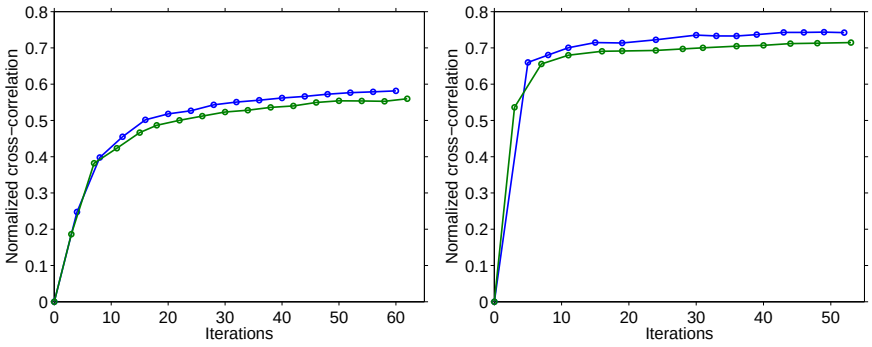
Figure 5.10: Source estimates using multiples from the data set modelled using iWave. **(a)** and **(b)**: Amplitude **(a)** and phase **(b)** spectra of the true (in blue) and the estimated (in green) source wavelet during the last LASSO(τ) subproblem, after amplitude normalization and correction with $|\omega|$.

model perturbations. For two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ (in our case, we vectorize the true and inverted model perturbations first), the NCC is the inner product of two vectors divided by the product of the ℓ_2 norms of the two vectors (i.e., normalized inner product). Geometrically the NCC is the cosine of the angle between two vectors and is not sensitive to the amplitude of either vector. The smaller the angle, the closer the two vectors are, and the closer the NCC is to one. We plot the NCC between the model iterates (we save a model iterate whenever a LASSO subproblem is finished) and the true model perturbation.

We plot the NCC curves in Figure 5.11 for both the imaging-with-primary and the imaging-with-multiple results in the presence of data imperfections. In both cases, we can see that the convergence behaviors of using the true source wavelet and source estimation are comparable, demonstrating the efficacy of the proposed source-estimation method.

5.5.4 Robustness to wavelet initial guess

We use an impulsive source as the initial guess of the source wavelet in all examples shown in this paper and get reasonable recoveries of the seismic images as well as the source wavelets. In this section, we would like to study the performance of the proposed method when the initial guess of the



(a) (b)

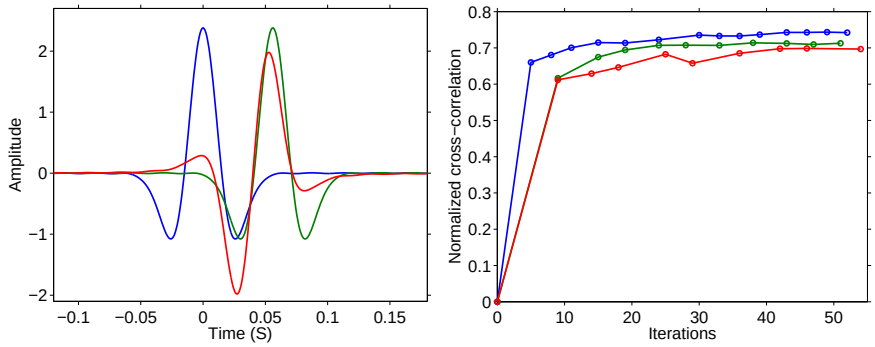
Figure 5.11: Model misfit represented by the normalized cross-correlation (NCC) as a function of the number of iterations. Each dot indicates the finish of one LASSO subproblem. (a) NCC curves for the imaging-with-primary examples in Figure 5.4, using the true wavelet (in blue), and with source estimation (in green). (b) NCC curves for the imaging-with-multiple examples in Figure 5.9, using the true wavelet (in blue) and with source estimation (in green).

source wavelet contains notable phase errors. Robustness of the proposed method to these phase errors is desirable. We use the same experiment setup as the imaging-with-multiples examples. We show the two erroneous initial guesses of the source wavelet in Figure 5.12(a), and show their corresponding solution paths in terms of model errors in Figure 5.12(b). We can see despite that both initial guesses of the source wavelet can lead to local minima, the solution paths and the final model errors are comparable to the case where we use an impulsive initial guess, demonstrating the robustness of the proposed method with respect to errors in the the initial guess of the wavelet.

5.6 Discussion

5.6.1 A note on the "true-amplitude" seismic imaging

From the above discussion, we can see that unless the true source wavelet is known, obtaining the true-amplitude seismic image is out of the question when primary-only data are used due to the amplitude ambiguity. Incorpo-



(a) (b)

Figure 5.12: (a) The true source wavelet and the two erroneous initial guesses of the wavelet. The initial guess in green has a significant time shift, and the one in red has a 90 degree phase rotation plus a time shift. **(b)** The corresponding solution paths in terms of model errors measured using NCC.

rating surface-related multiples in the inversion formulation can resolve the ambiguity, i.e., we get both the source estimates and the image with definite amplitudes. However, these definite amplitudes are the "true" amplitudes only if the linearized modelling engine can accurately reproduce the input wavefield given the true source wavelet, i.e., there is no modelling error. In practice, many factors hinder this requirement from being satisfied, for example, scaling of the modelling engine, and modelling errors including linearization errors (Chapter 3), and oversimplifications of physics such as elasticity and attenuation etc. Nevertheless, the presented method provides a viable approach for the least-squares type of seismic imaging algorithm with an unknown source wavelet to be one step closer towards the retrieval of true amplitudes.

5.6.2 Source estimation as a "garbage collector"

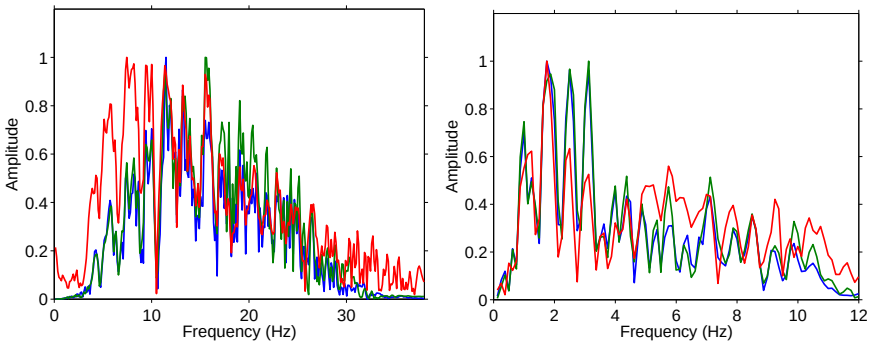
To understand the biased source estimates in Figure 5.10(a), we perform a more careful analysis of the source estimation scheme in the proposed method. According to the objective function of the source estimation step (Equation (5.5)), we estimate the source wavelet in such a way that the linearized modelling produces predicted data that best match the observed data. Because the spectra of the data is the product of the spectrum of the

source wavelet and that of the Green's function, the estimated source wavelet aims to collect whatever in the data that is not adequately explained by the predicted Green's function. In cases where the predicted Green's function is biased because of modelling errors, the source estimates can also depart from the true sources to accommodate this bias.

We therefore plot the frequency spectra of the predicted Green's function (by linearized modelling in the frequency domain using the inverted model perturbations with an impulsive source wavelet), and compare them with the Green's function corresponding to the input data (i.e., the true source wavelet is deconvolved from the data). For easier visualization, we sum over all data traces to obtain one trace that approximately shows the spectra envelope. The result is shown in Figure 5.13(a). We can see that because the predicted Green's function has underestimated low frequencies, the source estimates in return are biased towards the low frequencies to compensate for this underestimation. The underestimation of the predicted Green's function at low frequencies is caused by modelling errors—such as linearization errors—as we do not see the biased source estimates in Figure 5.8(a) where we use the ideal input data. Nevertheless, what is more important is that despite the bias in the source estimates, the predicted Green's function and the inverted model perturbations are comparable to the ones obtained using the true source wavelet (Figure 5.13(a), blue and green lines, and Figure 5.9(a) and 5.9(b)). In contrast, for the imaging-with-primary examples, there is not such a significant underestimate or overestimate of the predicted Green's function (Figure 5.13(b)), and we therefore obtain roughly unbiased source estimates (Figure 5.5(a)).

5.7 Conclusion

We have proposed a fast least-squares imaging procedure that, unlike conventional reverse-time migration, doesn't require prior knowledge of the source wavelet which is difficult to accurately determine. We formulated this wavelet-free imaging procedure as a nonlinear sparse-inversion problem, by tightly integrating source estimation into the compressive imaging framework, and proposed to solve the problem using variable projection. Because of the blind-deconvolutional nature of this problem, applying variable projection alone could not address the scaling ambiguity in estimating the source wavelet and the seismic image. We therefore continued by resolving the issue through the incorporation of surface-related multiples in the inversion process (although at the expense of an increased computational cost). We subsequently provided a series of examples demonstrating that



(a) (b)

Figure 5.13: The spectra of the predicted Green's functions using the inverted model perturbations with the true source wavelet (in blue), with source estimation (in green), versus that of the input data (wavelet is deconvolved from the data, in red). The spectra is obtained by summing over all traces and is amplitude-normalized. **(a)** Imaging-with-multiple examples. **(b)** Imaging-with-primary examples.

our method can produce highly accurate estimates of the source wavelet and high-resolution seismic images that are comparable to the ones obtained with the true source wavelet. We also provided examples where we overcome the scaling ambiguity by incorporating surface-related multiples, yielding definite amplitudes for the estimates of the source wavelet as well as for the seismic image. In all these examples, we controlled the simulation cost of the inversion to be more or less comparable to a single reverse-time migration with all the data, without compromising the quality of the images.

Chapter 6

Application to a field data set

6.1 Summary

In marine seismic acquisition, surface-related multiples constitute a significant portion of the acquired data. Typically, multiples are removed during early-stage data-processing as they can lead to phantom reflectors during migration that may result in erroneous geological interpretations. However, if properly dealt with, multiples can provide valuable extra information and complement primaries in illuminating the subsurface. In this article, we demonstrate the limitation of the reverse-time migration in imaging these multiples, and present an alternative inversion procedure that is computationally efficient, that jointly maps both primaries and multiples to the true reflectors, and where the source function is estimated on the fly. As a result, we obtain high-quality, mostly artifact-free, broad-band images where the imprint of the source-function are partly removed at a computationally affordable expense compared to the combined costs of the wave-equation based surface-related multiple elimination and the reverse-time migration. We achieve all this by including the total downgoing wavefields as areal sources in the least-squares migration in combination with curvelet-domain sparsity promotion. We apply the proposed method to a shallow-water marine data set from the North sea, which contains abundant short-period surface-related multiples, and show its efficacy in eliminating coherent imag-

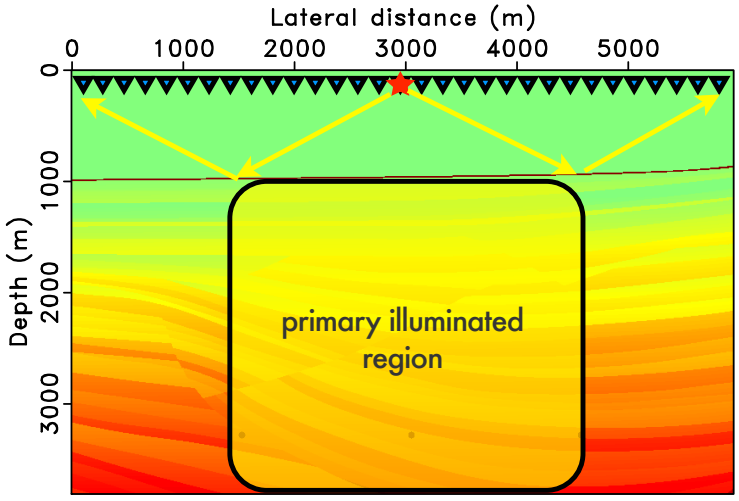
A version of this chapter has been published in The Leading Edge, 2015, vol. 34, no. 7, p. 788-794

ing artifacts associated with these multiples. We also demonstrate the benefits of joint imaging of primaries and multiples compared to imaging these signal components separately.

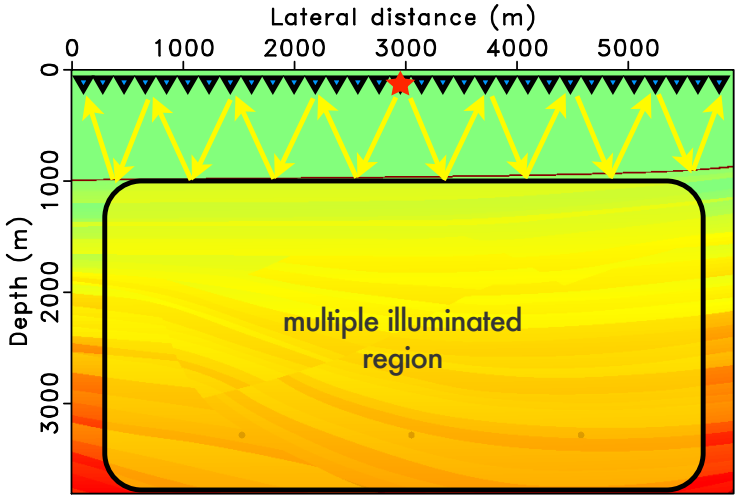
6.2 Introduction

Marine seismic data usually contain strong surface-related multiples, which bounce more than once between the free surface (i.e., the air-water contact) and subsurface reflectors. Amongst these multiples, waves that bounce in the water column are often the most pronounced, which can lead to challenging short-period multiples. Acting on first impulse, surface-related multiples are typically removed during conventional seismic data processing. While these approaches, such as Surface-Related Multiple Elimination (SRME, Verschuur et al., 1992), have been successful in removing the imprint of multiples on the final image, they discard valuable information contained in these multiples whose wave-number contents and apparent aperture are enriched by the fact that multiples undergo multiple bounces. We illustrate this effect in Figure 6.2 . For this reason, we propose to embrace surface-related multiples in our imaging formulation so we can maximally leverage their complementary information.

The idea of using surface-related multiples by itself is not new. Berkhout (1993) presented with his $\mathbf{W}^{-}\mathbf{R}^{+}\mathbf{W}^{+}$ model the theoretical foundation for later attempts to image multiples by modifying the source wavefield, either using the conventional cross-correlation imaging condition (i.e., RTM, Guitton, 2002; Schuster et al., 2004; Liu et al., 2011; Zuberi and Alkhalifah, 2013), or the deconvolutional imaging condition to control the coherent imaging artifacts associated with multiples using RTM (Muijs et al., 2007; Whitmore et al., 2010; Lu et al., 2014b). As pointed out by Poole et al. (2010), as well as in Chapter 4, even the latter approach may fail in the presence of complex geologies, as shown in Muijs et al. (2007), Figure (11b). Recent work by Lin et al. (2010b), Verschuur (2011), and Wong et al. (2012) showed that these artifacts can be largely removed by casting the imaging-with-multiples problem as a proper iterative least-squares migration problem at the expense of expensive, and perhaps prohibitive in 3D, simulation costs. The main contributions of this work is fourfold, namely *(i)* we incorporate surface-related multiples in RTM via wave-equation solves w.r.t. areal sources, and thereby avoid expensive multiple predictions that plague SRME and Estimation of Primaries by Sparse Inversion (EPSI, van Groenestijn and Verschuur, 2009a; Lin and Herrmann, 2013); *(ii)* we estimate the source function on-the-fly during the iterations making use of multiples

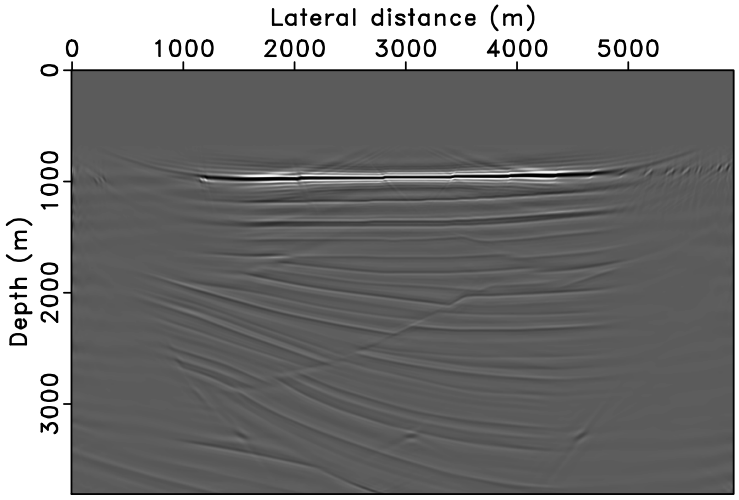


(a)

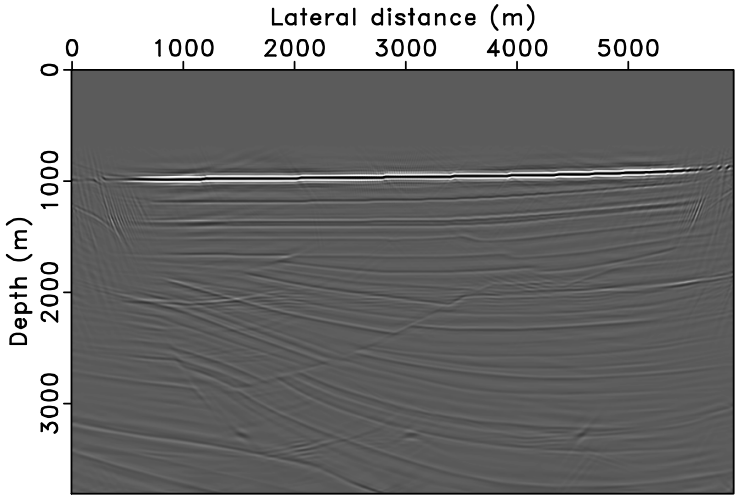


(b)

Figure 6.1: Broader illumination coverage of multiples compared with primaries: an illustration. The model is cropped from the sedimentary part of the Sigsbee 2B FS (with free-surface) model. **(a)** and **(b)**: Illustrations of the illuminated sections from the primaries and the multiples, respectively.



(a)



(b)

Figure 6.2: Broader illumination coverage of multiples compared with primaries: numerical examples. **(a)** and **(b)**: The iterative inversion results (after 60 iterations) of the primaries and the multiples, respectively.

(Chapter 5), following ideas from variable projection and using the unique relation between primaries and multiples and the source (Verschuur et al., 1992); *(iii)* we remove cross-talk between multiples by promoting curvelet-domain sparsity in the model space; and finally *(iv)* we reduce the number of wave-equation solves to roughly the same number of a single RTM using all shots (Chapter 2).

To demonstrate the possible advantages of our joint imaging (i.e., using both primaries and multiples, details in Chapter 2), we apply this method to a North-sea field data set. In particular, we are interested in specific benefits of imaging primaries and multiples together instead separately (e.g., in Guitton, 2002). After demonstrating the performance of the joint inversion approach on a stylized example, we apply it to a real North-sea data set.

6.3 Imaging with multiples & source estimation

For a given background-velocity model $\mathbf{m}_0$, reverse-time migration (RTM) derives from the following linear relationship

$$\delta\mathbf{d} = \nabla\mathbf{F}[\mathbf{m}_0;\mathbf{s}]\delta\mathbf{m} \tag{6.1}$$

between the observed upgoing primary wavefields collected in the vector $\delta\mathbf{d}$, the unknown source wavefields collected in the vector $\mathbf{s}$, and the unknown medium perturbation collected in $\delta\mathbf{m}$. This linear relationship is based on the single-scattering approximation, embodied in the linearized Born scattering operator $\nabla\mathbf{F}$.

During conventional imaging, surface-related multiples are removed from the observed data; the source is estimated, and a background-velocity model is built using (traveltime) tomography or migration-velocity analyses. In this paper, we assume the background-velocity model $\mathbf{m}_0$ to be known while we estimate both the source signature $\mathbf{s}$ and the image $\delta\mathbf{m}$ from the total upgoing wavefield, which contains imprints of both the source signature and the surface-related multiples. Without loss of generality, we assume spatially impulsive sources with the same time-harmonic source signature $\mathbf{s}_i = q(\omega)$ for all n_s sources $i = 1 \cdots n_s$ with ω the angular frequency.

Our formulation for imaging with multiples derives from three important modifications to conventional RTM where the source signature $\mathbf{s}$ is assumed to be known—i.e.,

$$\delta\mathbf{m}_{\text{rtm}} = \nabla\mathbf{F}^*[\mathbf{m}_0;\mathbf{s}(q)]\delta\mathbf{d},$$

where the image is obtained by applying the adjoint of $\nabla\mathbf{F}$ to the *de-*

multipled data—namely,

1. we replace the above spatially impulsive sources by areal sources that contains the downgoing data—i.e., we have

$$\tilde{\mathbf{s}}(q) \mapsto \left[\mathbf{s}(q) - \delta\tilde{\mathbf{d}}\right]. \tag{6.2}$$

   Note that now the observed data $\delta\tilde{\mathbf{d}}$ also contains all upgoing surface-related multiples. Contrary to conventional imaging, the source wavefield is now made up of the downgoing point-source wavefield plus the downgoing data (i.e., the sign-reversed upgoing data as we approximate the surface reflectivity by -1);

2. instead of carrying out migration with the areal source described earlier, we solve the curvelet-domain (denoted by the matrix $\mathbf{C}$) sparsity-promoting optimization program

$$\begin{aligned} &\text{minimize}_{\delta\mathbf{x}} \quad \|\delta\mathbf{x}\|_1 \\ &\text{subject to} \quad \|\delta\tilde{\mathbf{d}} - \nabla\mathbf{F}[\mathbf{m}_0;\tilde{\mathbf{s}}(q)]\mathbf{C}^*\delta\mathbf{x}\|_2^2 \le \sigma^2, \end{aligned} \tag{6.3}$$

   within some user-specified tolerance σ. Similarly to Robust EPSI (REPSI, Lin and Herrmann, 2013), this sparsity-promoting program inverts the surface-related multiples but, instead of estimating the primaries, it produces a sparse curvelet representation for the multi-free least-squares image $\delta\mathbf{x}_{\text{LS}}$ with a give source estimate q. The image is then readily derived by taking the inverse curvelet transform via $\delta\mathbf{m}_{\text{LS}} = \mathbf{C}^*\delta\mathbf{x}_{\text{LS}}$;

3. we estimate the source function on-the-fly after the k^{th} iteration of the above optimization program—i.e,

$$\text{minimize}_q \quad \|\delta\tilde{\mathbf{d}} - \nabla\mathbf{F}[\mathbf{m}_0;\tilde{\mathbf{s}}(q)]\mathbf{C}^*\delta\mathbf{x}^k\|_2^2. \tag{6.4}$$

   As we will demonstrate, the proposed formulation produces high-resolution, high-fidelity, artifact-free images as well as accurate estimates for the source signature. The imprint of the source and surface-related multiples are removed by combining the linearized forward model with SRME. The latter relates the total upgoing wavefield to the surface-free Green's function acting on the total downgoing wavefield. Compared to SRME our approach has the advantage of inverting, instead of removing, the surface-related multiples and thereby preserving multiple-scattering information. In addition, the proposed for-

malism uses wave-equation solvers to carry out the multi-dimensional convolutions implicitly (Chapter 2), which can lead to considerable computational savings. However, as REPSI, our approach is based on an iterative inversion procedure that requires multiple linearized Born scatterings and migrations, which can be prohibitively expensive for large models. Before providing details how to reduce the computational cost, let us first describe common difficulties one encounters when imaging with surface-related multiples.

6.4 Challenges during imaging with multiples

6.4.1 Data preprocessing

As we mentioned earlier, our formulation derives from the SRME relation, which requires dense fixed-spread sampling with co-located sources and receivers. For marine towed-streamer acquisition, we fill in the missing half of data using reciprocity, and interpolate the missing (cross-line/near-offset) traces as in standard SRME. To obtain a reasonable estimate for this total upgoing wavefield, the direct waves and the downgoing wavefield at the surface (i.e., receiver ghosts) need to be removed, and the upgoing wavefield needs to be extrapolated from the receiver level to the free surface (Verschuur et al., 1992). The source-side ghosts, on the other hand, should be left intact so that we mimic a dipole source as required by the theory (Verschuur et al., 1992).

6.4.2 The need to invert

Contrary to previous attempts to mitigate the imprint of surface-related multiples, the proposed formulation calls, as in REPSI, for a sparsity-promoting inversion of the linear forward modelling operator that ties perturbations in the velocity model to data that contain surface-related multiples. Failure to invert this operator leads to images that suffer from the imprint of the source and, perhaps more importantly, from artifacts related to coherent interferences between the areal sources and the observed data due to the presence of multiples.

To explain the physical origin of these artifacts, we included in Figure 6.3 a cartoon of the ray paths of primaries associated with conventional RTM (Figure 6.3(a)) and RTM with areal sources (Figure 6.3(b)). In this plots, the star indicates the seismic source and the triangles indicate the receivers.

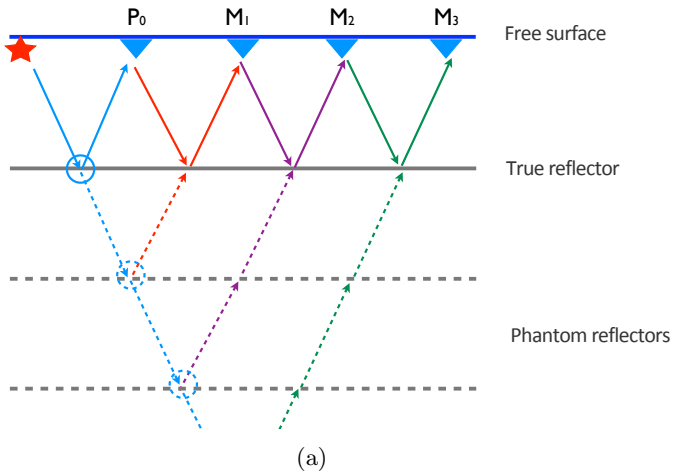
Moreover, the solid arrows depict the ray paths of the primaries $\mathbf{P}_0$, and the 1^{st} to 3^{rd}-order surface-related multiples $\mathbf{M}_1, \cdots, \mathbf{M}_3$. We use solid circles to delineate imaging points that correspond to true reflectors, and dashed circles to delineate imaging points that produce phantom reflectors. As we can see from the illustration in Figure 6.3(a), conventional RTM produces erroneous events that do not correspond to physical reflectors when the input data contains surface-related multiples. Only primaries contribute to imaging the true reflectors. (We ignore internal multiples for now.) Conversely, when we include the total down-going wavefield as an areal source, the RTM image contains a mix of proper and erroneous contributions from interactions between multiples present in the data and the areal sources. Neighboring — e.g., $\mathbf{P}_0$ and $\mathbf{M}_1$, and $\mathbf{M}_1$ and $\mathbf{M}_2$, etc. — pairs of multiples image to physical reflectors while the others create coherent interferences, which are similar to artifacts when carrying out conventional migration on data that contain surface-related multiples.

To illustrate interferences between surface-related multiples on the source and receiver sides, we include a stylized synthetic example consisting of a homogeneous water body, which also serves as the background model, and a single shallow perturbation as depicted in Figure 6.4(a) . In this model, we generate linearized primary-only data, by applying equation 6.1 with spatially impulsive sources, and data with surface-related multiples by injecting the areal sources given by equation 6.2 . The output of conventional RTM with the primary-only data is of reasonable quality, as shown in Figure 6.4(b), demonstrating RTM's ability to reproduce subsurface structures from multiple-free data. Unfortunately, the same observation does not apply when we image data including surface-related multiples (juxtapose Figures 6.4(b) and 6.4(c)). In that case, the image contains erroneous multiple-related artifacts. Unfortunately, including the total downgoing wavefield as areal sources during RTM is insufficient as we can see the result still contains interferences that can lead to wrong interpretations (juxtapose Figures 6.4(b) and 6.4(d)).

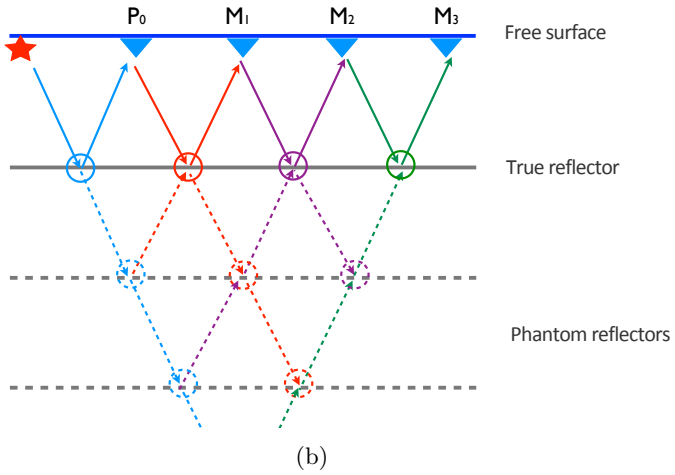
As we will demonstrate below, these artifacts can be removed by properly inverting the multiple-prediction operator, now disguised as areal sources in the Born scattering operator, with curvelet-domain sparsity promotion as outlined above.

6.4.3 Imaging & source estimation with multiples

As the stylized example in the previous section clearly illustrates, using areal sources (in equation 6.2) alone is obviously inadequate to image the



(a)



(b)

Figure 6.3: Illustration of the imaging artifacts stemming from surface-related multiples. In both figures, the input data contain primaries and surface-related multiples. **(a)** Conventional RTM with spatially impulsive sources. **(b)** RTM with areal sources.

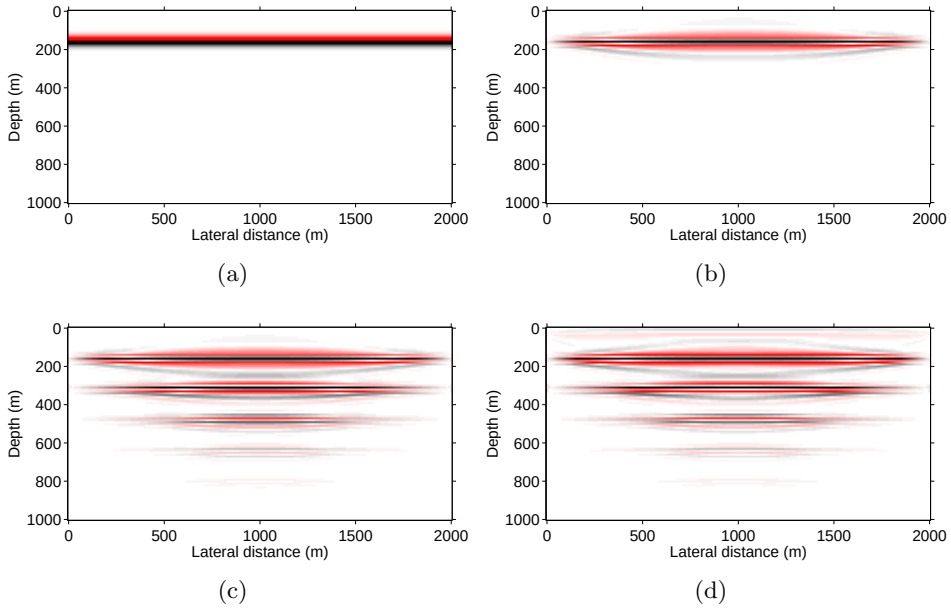


Figure 6.4: True model and the migrated images. **(a)** The true model perturbation. **(b)** Conventional RTM with spatially impulsive sources and multiple-free data. **(c)** The same as **(b)** but now for input data with surface-related multiples. **(d)** The same as **(c)** but now with areal sources given in equation 6.2 .

surface-related multiples. Furthermore, RTM also needs prior knowledge of the temporal source signature of the spatially impulsive sources, which is typically unknown unfortunately. However, when we invert the linearized relationship with sparsity promotion, we obtain far superior results for the above example as shown in Figure 6.5 where we either know the source signature beforehand, or we estimate this source signature on the fly.

We can make the following observations: *(i)* sparsity-promoting inversion of data with multiples, but without using the areal sources, leads to a high resolution image that contains artifacts related to the multiples (see Figure 6.5(a)); *(ii)* when provided with the correct source wavelet, carrying out the inversion with areal sources leads to a high-resolution artifact-free image (see Figure 6.5(b)); *(iii)* the same applies when we invert for the wavelet on the fly (see Figure 6.5(c)); *(iv)* source estimation produces accurate estimates for the source signature (see Figure 6.5(d)).

The above observations are truly remarkable because we have, follow-

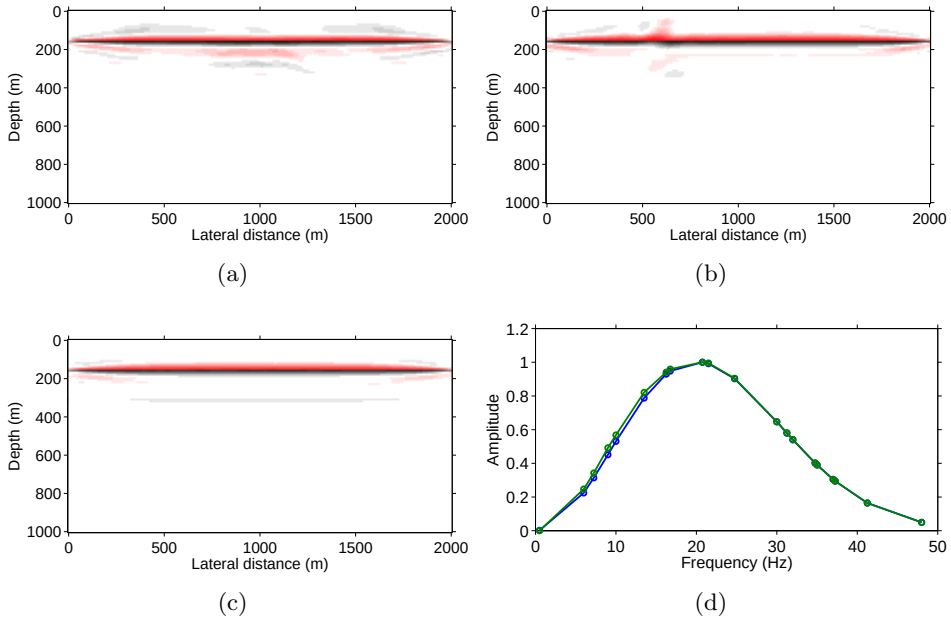


Figure 6.5: Migrated images using the proposed method. **(a)** Imaging of total data but only with primary-imaging operator. **(b)** and **(c)** Joint imaging of primaries and multiples, with the true source **(b)** and with source estimation **(c)**. **(d)** Source estimates (in green) versus the true source function (in blue) at the frequencies selected in the last iteration.

ing the procedure outlined above, been able to solve three problems in one go, namely *(i)* we successfully jointly map primaries and surface-related multiples to true reflectors so we maximally benefit from the higher Signal-to-Noise Ratio (SNR) of the primaries and the extra illumination of the multiples; *(ii)* we estimate the unknown source function accurately while removing its imprint; and *(iii)* we properly scale the contributions from primaries and multiples that both contribute to the image. While the results for this stylized example are certainly encouraging, we will need to first remedy the computational costs associated with this procedure, which relies on many iterations involving computationally expensive migrations and demigrations.

6.4.4 Fast randomized iterative inversion

Least-squares migrations, especially with sparsity-promoting, are in most cases computationally prohibitively expensive because they rely on expensive iterations, the simulation cost of which scales linearly with the number of (monochromatic) source experiments. To reduce these costs, we borrow ideas from Herrmann and Li (2012) and limit the total number of wave-equation solves by working with randomized subsets of frequencies and source experiments during each iteration. As described in Chapter 2, this type of randomized procedure—where we regularly draw new randomized experiments for each subproblem to expedite the convergence—leads to significant cost savings bringing the required number of wave-equation solves of our iterative procedure down to roughly equal a single RTM with all sources. Remember that by virtue of using areal sources, our imaging procedure also avoids expensive iterative multi-dimensional convolutional operations required by SRME and EPSI. Instead, these expensive operations are carried out by the wave-equation solver (Chapter 2), leading to what we believe a computationally viable approach that can be applied to 2D field data sets as shown below, and in the future, to 3D field data.

6.5 North-sea case study

To evaluate the performance of the proposed method, we consider a 2D relatively shallow (90 m water depth) marine data set from the North Sea, acquired by PGS with dual-sensor streamers. Because this type of acquisition gives us access to both the pressure and the vertical particle-velocity wavefields, this data set is ideal for us to accurately separate the observed wavefield into its up- and down-going constituents. To get the best results, we apply the following preprocessing steps: near-offset data interpolation using the Radon transform; up-down wavefield decomposition to remove the receiver ghost and removal of the direct waves. These processing steps, in combination with application of source-receiver reciprocity to mimic a fixed spread acquisition, gives us an estimate for the downgoing wavefield with 401 co-located sources and receivers with 12.5 m grid spacing that serves as input to our imaging algorithm. Only the first 4092 ms of the data record are imaged up to 48 Hz. We derive the 3×5 km background-velocity model with 6.25 m grid spacing from the stacking velocity. To assess the quality of our imaging results, we also ran REPSI to get independent estimates for the primaries and surface-related multiples. The least-squares (LS) image (we will use this term for images obtained by solving the above sparsity-promoting

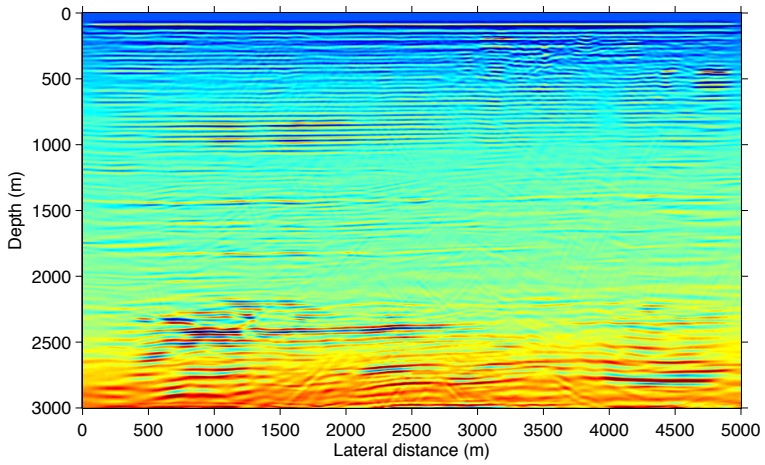
program) from the primary-only data is included in Figure 6.6(a) for reference. The image of the total upgoing data using the proposed method is shown in Figure 6.6(b) . For both images, we estimate the source wavelet on the fly. On the computational side, both images are obtained with roughly the same wave-equation solves as a single RTM with all the data (in the joint-imaging case, the source estimation step necessitates some extra data modelling that results in 50% more simulation cost, we in return reduce the number of iterations by one-third for Figure 6.6(b)).

6.5.1 Joint versus separate inversions

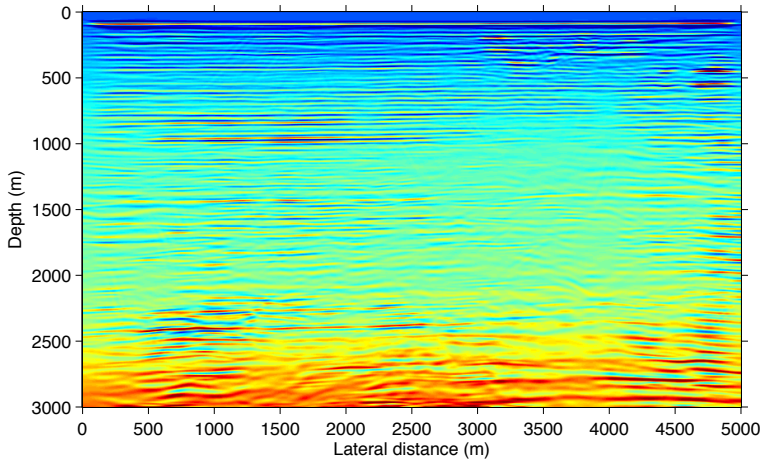
With the proposed method, we can either image primaries and multiples separately, or image the total upgoing wavefields in one step via joint inversion with on-the-fly source estimation. Compared to imaging with primaries and multiples separately, this joint approach has the advantage that the images from these two components are properly scaled as we optimize over the source wavelet to fit both primaries and multiples. As a result, we avoid the source-scaling ambiguity, which hampers imaging with primaries only (Chapter 5). In addition, we produce one combined imaging result that maximally uses the information residing in both primaries and multiples.

To substantiate this claim, we conduct a series of experiments. First, we compare images from primaries only (Figure 6.7(a)); from multiples only (Figure 6.7(b)), and from primaries and multiples jointly (Figure 6.8). All results are obtained with our sparsity-promoting inversion method using their respective source wavefields (see Discussions in Chapter 2). When we compare the images obtained from primaries and multiples only, we find that in general the image of the primaries alone is, as expected, cleaner. For example, the structures indicated by the red arrows are better resolved with the primaries, and so are the two sections marked inside the blue rectangles. However, while scaled differently, the image from the multiples alone reveals most of the major reflectors (cf. Figures 6.7(a) and 6.7(b)) and even leads to an improved image for the reflector indicated by the grey arrows in Figure 6.7(b) .

Comparing our results obtained by inverting primaries- and multiples-only data with the joint inversion suggests that inverting the total data may lead to superior results that benefit from the relatively noise-free primaries and extra illumination from the multiples. As the results in Figure 6.8 indicate, our joint inversion strikes a good compromise and produces a single image that combines the benefits of both primaries and multiples — compared with Figure 6.7(a), the structures indicated by the grey arrows are

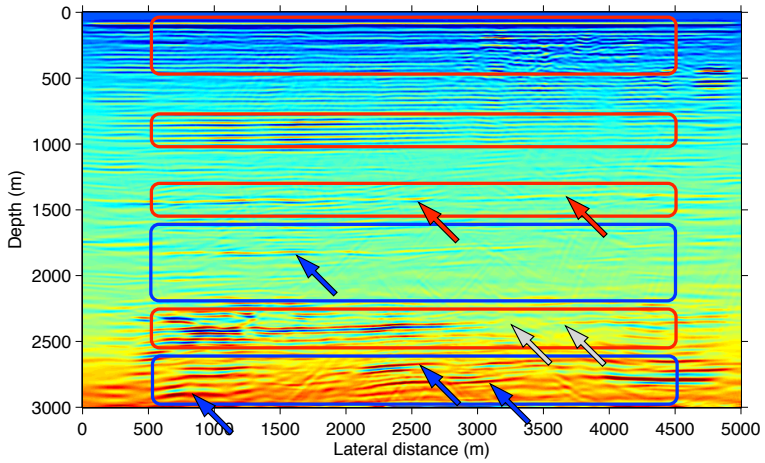


(a)

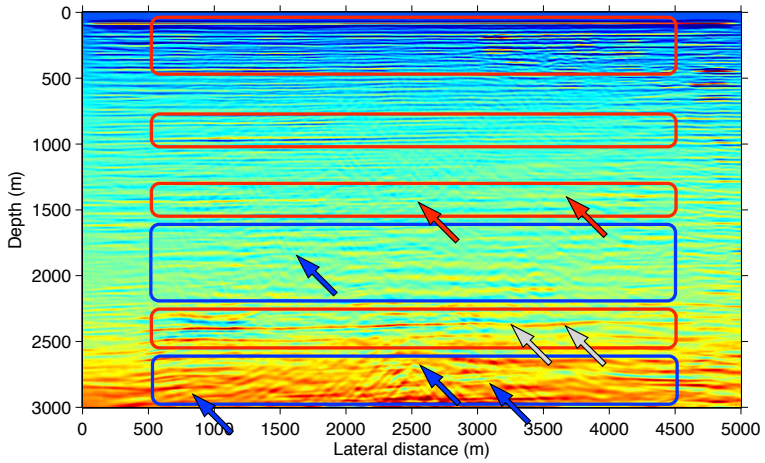


(b)

Figure 6.6: LS images of **(a)** primaries only, and **(b)** total upgoing data.



(a)



(b)

Figure 6.7: Separate images of primary-only data **(a)** and multiple-only data **(b)**.

better imaged in Figure 6.8; and compared with Figure 6.7(b), the structures indicated by the red and blue arrows are better imaged in Figure 6.8.

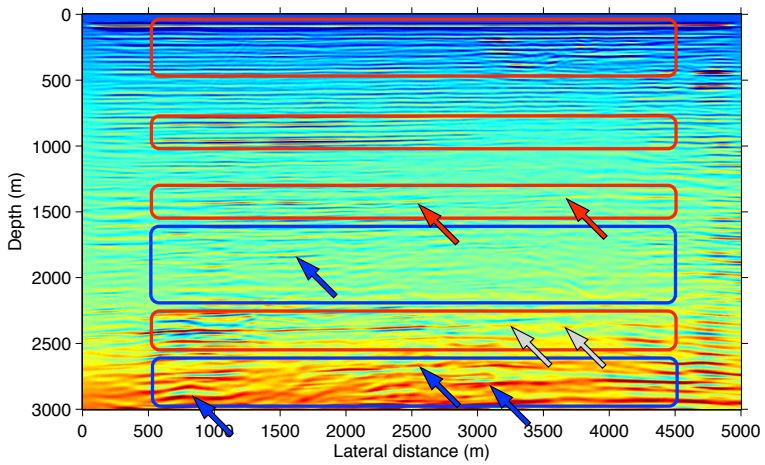
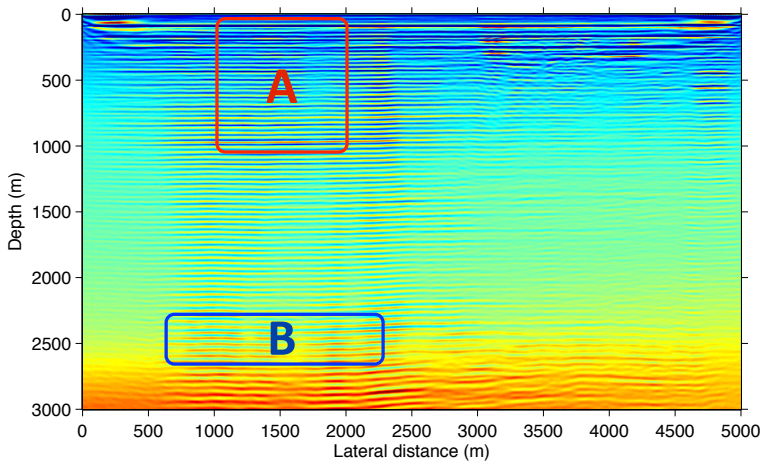


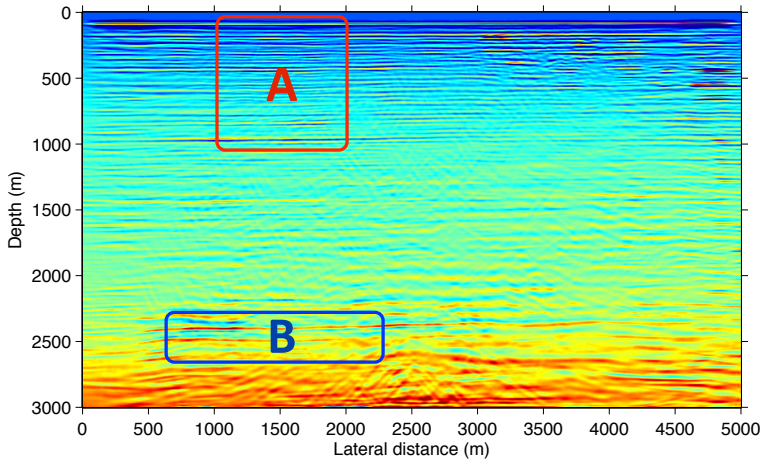
Figure 6.8: Joint imaging of total data consisting of primaries and multiples.

6.5.2 Suppressing imaging artifacts from multiples

To further substantiate our claim that the proposed inversion scheme can remove the coherent artifacts associated with multiples, we study the detrimental effects that surface-related multiples can have when handled incorrectly—i.e., when included but not inverted. For this purpose, we again draw your attention to imaging with multiples only by comparing the result of a single RTM (Figure 6.9(a)) with the result obtained with sparsity-promoting inversion (Figure 6.9(b)). In both cases, we use areal sources defined in equation 6.2 . Even though using areal sources leads to true reflectors such as the sea floor (c.f. the solid circles in Figure 6.3(b)), it also result in interferences (c.f. the dashed circles in Figure 6.3(b)), which rings across the section (Figure 6.9(a)). The presence of these artifacts exposes a fundamental shortcoming of RTM's correlation-based imaging condition, which we overcome largely by sparsity-promoting inversion. As we can be readily observe from Figure 6.9(b) and the zoomed in sections A and B included in Figure 6.10(a) and 6.10(b), the strong artifacts associated with the multiples in Figure 6.9(a) are mostly suppressed in Figure 6.9(b) . These results demonstrate that the energy of surface-related multiples can be mapped to true reflectors via a sparsity-promoting inversion procedure using the downgoing receiver data as areal sources. The results also support our claim that inverting primaries and multiples jointly allows us to benefit from the multiples without the need to separate primaries and multiples upfront, which



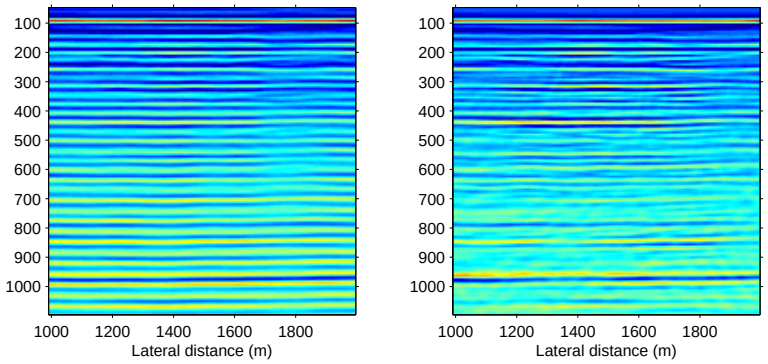
(a)



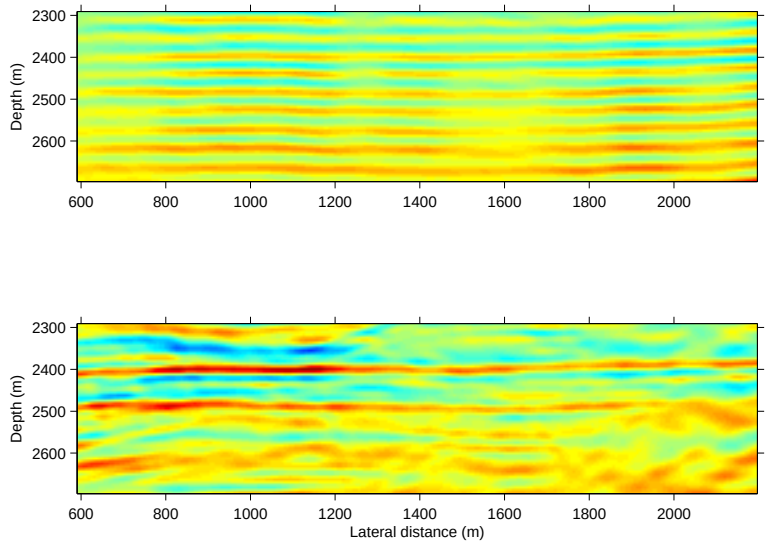
(b)

Figure 6.9: **(a)** RTM of multiples only. **(b)** LS image of multiples only. Both with the downgoing receiver data acting as areal sources.

is an computationally expensive proposition.



(a)



(b)

Figure 6.10: Zoomed-in comparisons of sections A and B between Figure 6.9(a) and Figure 6.9(b).

6.6 Conclusion

Because surface-related multiples bounce more than once between the free surface and the subsurface reflectors, they generate coherent artifacts when not treated properly during imaging—even in situations where conventional sources are replaced by areal sources that include the downgoing data. However, if we combine these areal sources with a proper sparsity-promoting inversion procedure—during which the deconvolved image and source signature are estimated—we arrive at a formulation that maximally leverages primary and multiple energy via a joint-inversion procedure. During this inversion, the adverse coherent artifacts arising from multiples are turned into coherent images, and we can benefit from the multiples that are better sampled than primaries as their bounce points act as secondary sources. Aside from mapping multiple-related artifacts to the correct reflectors, our method also avoids the expensive combination of wave-equation based surface-related multiple elimination and subsequent reverse-time migration of primaries and/or multiples. Instead, our method proposes to image the primaries and multiples jointly during an iterative sparsity-promoting inversion procedure. By adaptively estimating the source wavelet, we achieve an optimal balance between the contributions from primaries and multiples. By using randomized subsampling techniques, we limit the number of required wave-equation solves to roughly equal a single reverse-time migration with all data. Since the modelling of the multiple is done by the wave-equation solver, no explicit multiple prediction and removal are needed. Application of our inversion framework on a real marine field dataset clearly demonstrates the advantage of inverting primaries and multiples jointly compared to estimating and imaging these wave constituents separately.

Chapter 7

Future work: sparse seismic imaging simplified

7.1 Summary

We present a novel adaptation of a recently developed relatively simple iterative algorithm to solve large-scale sparsity-promoting optimization problems. Our algorithm is particularly suitable to large-scale geophysical inversion problems, such as sparse least-squares reverse-time migration or Kirchoff migration since it allows for a tradeoff between parallel computations, memory allocation, and turnaround times, by working on subsets of the data with different sizes. Comparison of the proposed method for sparse least-squares imaging shows a performance that rivals and even exceeds the performance of state-of-the-art one-norm solvers that are able to carry out least-squares migration at the cost of a single migration with all data.

7.2 Motivation

Wave-equation based migration corresponds to computing

$$\delta\mathbf{m} = \sum_{i=1}^{K} \nabla\mathbf{F}_i^*(\mathbf{m}_0, \mathbf{q}_i)\delta\mathbf{d}_i, \tag{7.1}$$

A version of this chapter has been published in the proceedings of the 77th EAGE Conference & Exhibition 2015, Madrid, Spain

where the sum runs over the number of experiments, indexed by $i = 1 \cdots K$. For seismic applications, this number K is typically very large. These experiments—with sources $\mathbf{q}_i$ and corresponding data $\delta\mathbf{d}_i$—involve either individual traces for different source/receiver pairs as in Kirchoff migration or (monochromatic) shot records as in (time-harmonic) reverse-time migration. In either case, the computational costs, including data-storage and movement, grows with K and the size of the survey area. To make matters worse, migration often relies on increasing the fold, to stack out unwanted artifacts in the image space, or on carrying out least-squares migration. In either case, the aforementioned increased costs strain even the largest computational resources stifling the adaptation of sparsity-promoting inversion techniques.

To overcome these costs, which are exacerbated by rapidly increasing costs of RTM as the frequency increases, we proposed randomized algorithms based on certain subsampling techniques that work on randomized subsets of the sources and corresponding data. These techniques have either lead to so-called batching techniques (van Leeuwen and Herrmann, 2012), where the above sum is approximated by carrying out the sum over subsets of randomly selected sources (or simultaneous sources), or to techniques based on Compressive Sensing (Herrmann and Li, 2012). Contrary to batching techniques, where the error is controlled by increasing the size of the subsets, sampling related errors during Compressive Imaging are controlled by promoting curvelet-domain sparsity. By incorporating rerandomization techniques from stochastic optimization into sparsity-promoting imaging, we arrived at an imaging technology that is able to carry out sparsity-promoting least-squares migration working on small subsets at the cumulative costs of roughly one RTM carried out over all data (Herrmann and Li, 2012, and Chapter 2).

While our Compressive Imaging technology is certainly promising, it relies on the combination of a sophisticated ℓ_1-norm solver (SPGℓ_1, van den Berg and Friedlander, 2008), rerandomizations, and warm starts. This makes the resulting algorithm complicated, and therefore difficult to implement on industry-scale problems. In addition, there is, aside from empirical observations, no formal proof what optimization problem this heuristic randomized approach actually solves. Below, we present a new highly flexible sparse inversion methodology that derives from a recently proposed sparse Kaczmarz solver for Compressive Sensing problems using linearized Bregman iterations (Lorenz et al., 2014). The resulting algorithm is relatively simple and suitable for seimic imaging problems where measurements—read application of rows/row-blocks of the linearized modelling operator and

therefore terms in the sum of Equation 7.1—take time and parallel resources to evaluate.

We organize our work as follows. First, we briefly state a simplified version of the algorithm undergirding our Compressive Imaging technology and juxtapose it with the proposed algorithm. Next, we compare the performance of the proposed algorithm on seismic imaging problems, followed by conclusions.

7.3 Fast sparse solvers

We are interested in solving the following (partly) separable—i.e., we have access to individual rows as in trace-based migration or blocks of rows as in shot migration—linear inverse problems, where we aim to estimate the image $\mathbf{x}$ from measurements $\mathbf{b} \in \mathbb{R}^M$ given by $\mathbf{b} = \mathbf{A}\mathbf{x}$, with $\mathbf{x}$ sparse or compressible, and $\mathbf{A} \in \mathbb{R}^{M \times N}$ tall—i.e., $M \gg N$. This overdetermined problem is different from Compressive Sensing where $M \ll N$ and could in principle be solved easily by inverting $\mathbf{A}$ with least-squares or by simply applying the adjoint. Unfortunately, since we are dealing with problems where there are many rows of $\mathbf{A}$ that are expensive to evaluate, we can not follow this approach and we have to resort to working on randomized pairs $\{\mathbf{A}_i, \mathbf{b}_i\}$, made of randomly selected subsets of rows only—i.e., we only have access to $\mathbf{A}_i \in \mathbb{R}^{m \times N}$ and $\mathbf{b}_i \in \mathbb{R}^m$ with $m \ll N \ll M$. Our aim is now to come up with an algorithm where the total amount of work, read the total number of rows participating in the inversion, is roughly the same as the total number of rows of the total system , or terms in the sum of Equation (7.1)—i.e., $L = m \times n_k \sim M$ with n_k the number of iterations of the solver.

7.3.1 Solution by Randomized Iterative Shrinkage-Thresholding Algorithm (RISTA)

While many algorithms have been developed to solve the convex Basis Pursuit problem for underdetermined $\mathbf{A}$'s with $m \ll N$

$$\mathbf{BP}: \qquad \underset{\mathbf{x}}{\text{minimize}} \, \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b},$$

its solution often requires too many iterations, and therefore evaluations of $\mathbf{A}$ and its adjoint $\mathbf{A}^*$, to make this a viable option for industrial problems, where the matrices are either very large or expensive to evaluate or both. To overcome this problem, the authors proposed a method to accelerate this

type of algorithms using ideas from approximate message passing. The key idea of that approach was to "break" buildup of correlations between the model iterates, $\mathbf{x}_k$, and the matrix $\mathbf{A}$ by drawing randomized pairs $\{\mathbf{A}_k, \mathbf{b}_k\}$ for each iteration k. This results in Algorithm 3, a randomized version of the Iterative Shrinkage-Thresholding Algorithm (ISTA,Yin et al. (2008);Beck and Teboulle (2009)). For a fixed pair $\{\mathbf{A}_k, \mathbf{b}_k\}$ and appropriately chosen step length t_k and $\lambda \to 0^+$ (read as λ goes to 0 from above), Algorithm 3 provably solves BP. While RISTA type algorithms—where techniques from stochastic gradients are combined with sparsity promotion—lead to fast and therefore practical algorithms, their performance hinges on choices for the threshold. Not only do small λ's lead to slow convergence, there is also no general proof that Algorithm 3 solves BP when choosing different pairs $\{\mathbf{A}_k, \mathbf{b}_k\}$ for each iteration. So the question is whether there exists an alternative iterative solver that makes few assumptions on $\mathbf{A}$ and that is capable of working with small randomized subsets of rows and that is relatively simple to implement.

Algorithm 3 Randomized Iterative Shrinkage-Thresholding Algorithm (RISTA) for appropriately chosen step lengths t_k's and threshold λ for the soft thresholding with $S_\lambda(x) = \max(|x| - \lambda, 0)$.

Result: Estimate for the model x^k

1. **for** $k = 0, 1, \cdots$ **do**
2. Draw $\mathbf{A}_k$ and $\mathbf{b}_k$
3. $\mathbf{z}_{k+1} = \mathbf{x}_k - t_k \mathbf{A}_k^* (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$
4. $\mathbf{x}_{k+1} = S_\lambda(\mathbf{z}_{k+1})$
5. **end for**

7.3.2 Solution by Randomized Iterative Shrinkage-thresholding Kaczmarz Algorithm with linearized Bregman (RISKA)

Consider the following slight modification—i.e. we replace line 3 of Algorithm 3 by $\mathbf{z}_{k+1} = \mathbf{z}_k - t_k \mathbf{A}_k^* (\mathbf{A}_k \mathbf{x}_k - \mathbf{b}_k)$. We call this Solution by Randomized Shrinkage-Thresholding Kaczmarz Algorithm with linearized Bregman (RISKA,Yin et al. (2008); Maharramov and Biondi (2014)). According to recent work by Lorenz et al. (2014), this modified algorithm solves for step

lengths, $t_k = \frac{\|\mathbf{A}_k\mathbf{x}_k - \mathbf{b}_k\|^2}{\|\mathbf{A}_k^*(\mathbf{A}_k\mathbf{x}_k - \mathbf{b}_k\|^2}$,

$$\text{KB}: \qquad \underset{\mathbf{x}}{\text{minimize}}\, \lambda\|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{x}\|_2^2 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}.$$

For large enough λ (van Leeuwen and Herrmann, 2013), RISKA provably converges to the solution of BP. While the modification w.r.t. RISTA is at first sight relatively minor, its implications are profound and manifold. First, compared to RISTA the inversion involves two iterates, namely $\mathbf{z}_{k+1}$, which seeks to minimize the ℓ_2-norm of the residual, and sparse model iterates $\mathbf{x}_k$ obtained by soft thresholding $\mathbf{z}_{k+1}$ with the threshold λ. Contrary to RISTA, updates for the $\mathbf{z}_k$'s are calculated by subtracting gradients computed with respect to the sparse model iterate $\mathbf{x}_k$. Third, the role of the threshold λ fundamentally changes. Within RISTA, the λ determines the projection of the model iterates on a λ dependent ℓ_1 ball. For a fixed λ the iterations without drawing new pairs $\{\mathbf{A}_k, \mathbf{b}_k\}$ solve the unconstrained problem $\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\| + \lambda\|\mathbf{x}\|_1$. By taking the limit $\lambda \to 0^+$ these unconstrained problems solve BP. Unfortunately, small λ lead to slow convergence (Hennenfent et al., 2008). For large-scale problems this makes it necessary to devise ingenious cooling techniques where the ℓ_1-norm constraint is gradually relaxed, e.g., by decreasing the λ's as the algorithm progresses. This process allows components to enter into the solution slowly while bringing down the residual. Thresholding within RISKA, on the other hand, plays a completely different role since it balances ℓ_1- and ℓ_2-norms on the model. This means in a way that RISKA is "auto tuning"—i.e., when we choose λ too high, the $\mathbf{x}_k$ will be zeros but not the $\mathbf{z}_k$'s. As k increases, the large entries—these are the first candidates to enter the solution—in $\mathbf{z}_k$ will grow while the noisy crosstalk will become smaller if we chose to work with randomized $\{\mathbf{A}_k, \mathbf{b}_k\}$'s. After sufficient iterations, the number of which can be reduced by using intelligent step lengths, components will enter the solution. A too large λ for RISTA, on the other hand, would lead to zeros and would require a lowering of λ. Finally, RISKA solves a constrained problem, i.e., the data is fitted exactly (this can be relaxed to different norms error balls (Lorenz et al., 2014)). This has the advantage that step lengths can be larger leading to a very fast but above all extremely simple algorithm that lends itself well to *(i)* work on (randomized) subsets of the data. Depending on the available computational resources, this gives us the freedom to choose, for a fixed number of passes through the rows L, an optimal batch size m by trading of numbers of iterations and numbers of available nodes partaking in the inversion. For RTM this means that we can control the

number of PDE solves while for trace-based migration this would limit the total number of source-receiver pairs touched by migration; *(ii)* implement RISKA in existing industrial settings as long as one has access to randomized subsets of data $\mathbf{b}_k$, the Born modelling $\mathbf{A}_k$, and its adjoint migration $\mathbf{A}_k^*$. For RTM this means randomly selected shots or simultaneous shots while for Kirchhoff migration one should have access to randomized traces; *(iii)* select data and rows adaptively, e.g., according to the norm of the rows of $\mathbf{A}$ or condition number of block of rows as described in the literature (Strohmer and Vershynin, 2009; Needell and Tropp, 2014).

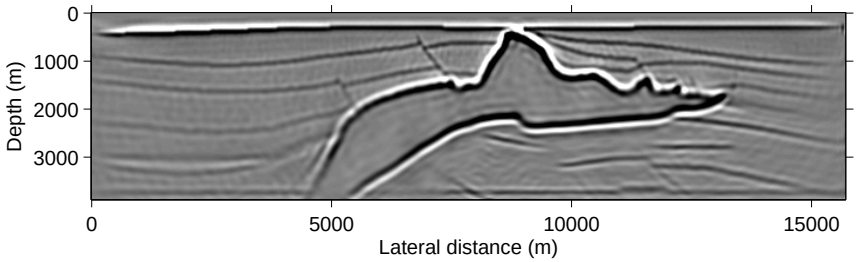
7.4 Examples

To illustrate the performance of RISKA, we compare sparsity-promoting linearized imaging results for primaries only and data including primaries and surface-related multiples obtained with our randomized SPGℓ_1 solver and with the above described algorithm. In all examples, we fix the number of wave-equation solves to be equivalent to a single RTM with all data. The results included in Figure 7.1 show improvements, especially below the Salt (juxtapose Figure 7.1(a) and 7.1(b)). These results are remarkable because we are able to obtain high-fidelity inversion results at the cost of a single RTM. An advantage of RISKA is that the underlying algorithm is simple to implement and relatively robust w.r.t. choices of λ.

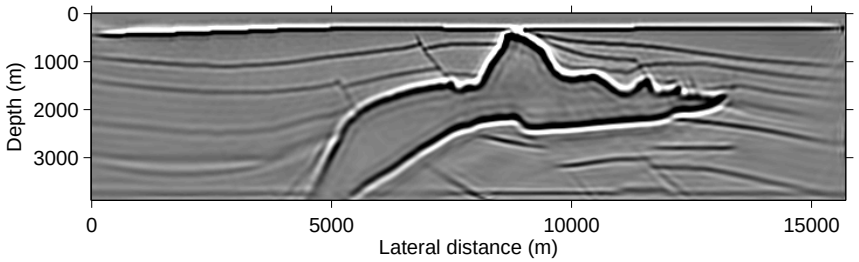
For the same budget (cost of 1 RTM), the proposed algorithm also produces improved results for an imaging example that includes surface-related multiples. In this case, we invert the linearized Born scattering operator with an areal source wavefield that contains the downgoing receiver wavefield (Chapter 2). Again, the results in Figure 7.2 are excellent despite the relative simplicity of the algorithm.

7.5 Conclusion

We presented a simplified randomized sparsity-promoting solver that provably converges to a mixed one- two-norm penalized solution while working on arbitrarily small subsets of data. This latter ability, in combination with its simplicity and "auto-tuning", makes this algorithm particular suitable for large-scale and parallel industrial applications.

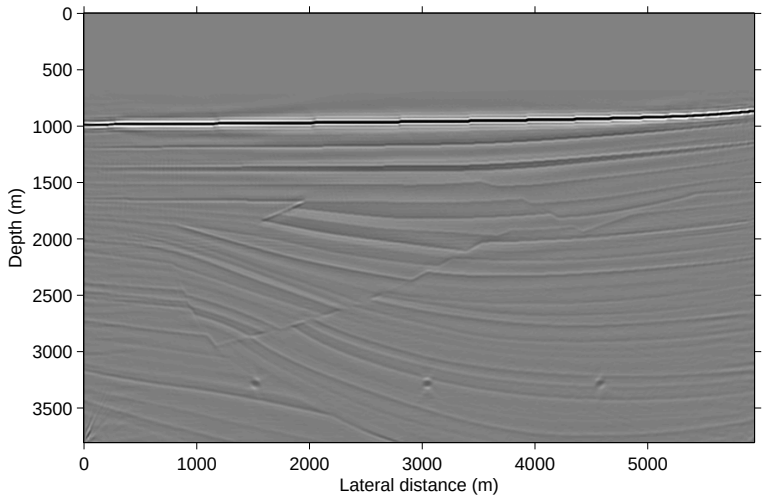


(a)

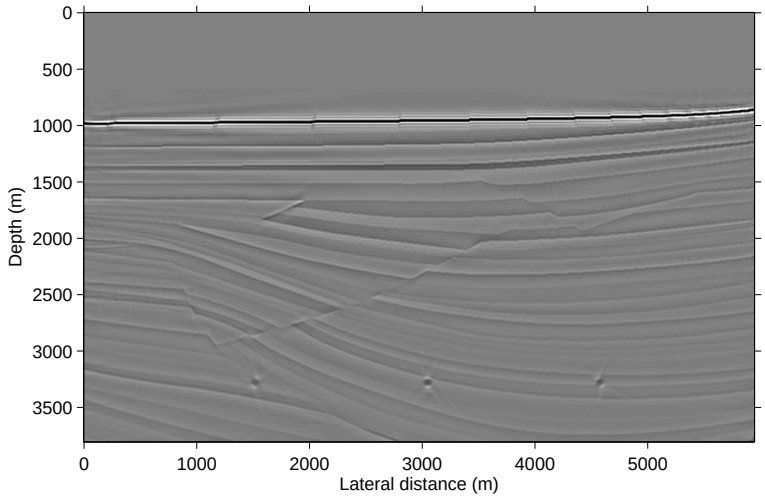


(b)

Figure 7.1: Comparison of randomized sparse inversions on primary only data using SPGℓ_1 **(a)** and RISKA **(b)**. It is clear that for the same computational budget, the inversion result for RISKA is less noisy and recovers the reflectors below the Salt better.



(a)



(b)

Figure 7.2: Comparison of randomized sparse inversions on data with surface-related multiples using SPGℓ_1 **(a)** and RISKA **(b)**. Again, the much simpler RISKA does an excellent job compared to SPGℓ_1.

Chapter 8

Further extension: least-squares extended imaging with surface-related multiples

8.1 Summary

Common image gathers are used in building velocity models, inverting anisotropy parameters, and analyzing reservoir attributes. Often primary reflections are used to form image gathers and multiples are typically attenuated in processing to remove strong coherent artifacts generated by multiples that interfere with the imaged reflectors. However, researchers have shown that, if correctly used, multiples can actually provide extra illumination of the subsurface in seismic imaging, especially for delineating the near-surface features. In this work, we borrow ideas from literatures on imaging with surface-related multiples, and apply these ideas to extended imaging. This way we save the massive computation cost in separating multiples from the data before using them during the formation of image gathers. Also, we mitigate the strong coherent artifacts generated by multiples which can send the migration velocity analysis type algorithms in wrong direction. Synthetic examples on a four-layer model show the efficacy of the proposed formulation.

A version of this chapter has been published in the proceedings of CSEG GeoConvention 2015, Calgary, Canada

8.2 Introduction

An extended image is a multi-dimensional correlation of source and receiver wavefields, as a function of *all* subsurface offsets (see Sava and Vasconcelos, 2011, and references therein for a recent overview). There are many different applications in which extended images are used extensively like construction of angle-domain common-image gathers (ADCIGs) and migration velocity analysis (MVA) (Biondi and Symes, 2004; Shen and Symes, 2008; Symes, 2008; Sava and Vasconcelos, 2011; Kumar et al., 2013, 2014; Wong et al., 2014). Other than that, extended images are used to study the rock properties and fluid indicators by estimating the amplitudes of reflected waves as a function of incident angle at the interface. One of the current limitations is that the extended imaging conditions cannot handle multiple reflections. However, Lu et al. (2014a) showed that primaries may not provide enough illumination to image near-surface targets. To overcome this situation, Lu et al. (2014a) proposed to use multiples where they apply the same imaging condition to multiples by putting in the primaries as the source wavefield and the multiples as the receiver wavefield. However, efforts during seismic data processing to extract multiple wavefields are always challenging, computationally expensive and can risk damage to the underlying signal. Therefore, to mitigate these impediments and to get benefit from the multiples, we follow earlier work on joint imaging of primaries and surface-related multiples (Chapter 2 and Chapter 5) and adapt it to extended imaging which potentially leads to better images. The idea is to combine EPSI (Estimation of Primaries via Sparse Inversion) (van Groenestijn and Verschuur, 2009a) with migration as a way to benefit from surface-related multiples. Following the same ideas, we show that we can exploit the EPSI relationship while forming the image gathers. As a result, we need to move to a least-squares extended imaging rather than a simple correlation-based imaging condition. The proposed way of imaging the surface-related multiples along with primaries do not increase the dominant computational cost, which is the solution of the wave-equation. Finally, we show the efficacy of the proposed formulation on a four-layer synthetic velocity model.

8.3 Extended imaging with free-surface multiples

Given the time-harmonic source and receiver wavefields as matrices $\mathbf{U}$ and $\mathbf{V}$, extended images at a *single* frequency, for *all* subsurface offsets and for *all* subsurface points can be written as the outer product of these two

matrices, i.e., we have

$$\mathbf{E} = \mathbf{V}\mathbf{U}^*, \tag{8.1}$$

where $\mathbf{U}$, $\mathbf{V}$ are calculated using the two-way wave-equation and each column of $\mathbf{U}$, $\mathbf{V}$ corresponds to a source experiment. The explicit expression of $\mathbf{E}$ can be written as

$$\mathbf{E} = \mathbf{H}^{-*}\mathbf{P}_r^T\mathbf{D}\mathbf{Q}^*\mathbf{P}_s\mathbf{H}^{-*}, \tag{8.2}$$

where $\mathbf{H}$ is a discretization of the Helmholtz operator $(\omega^2\mathbf{m} + \nabla^2)$ where $\mathbf{m}$ is the squared slowness, matrix $\mathbf{Q}$ represents the source function, * represents the conjugate-transpose, $\mathbf{D}$ is the primary reflection data matrix, and matrices $\mathbf{P}_s$ and $\mathbf{P}_r$ sample the wavefield at the source and receiver positions (and hence, their transpose injects the sources and receivers into the grid). Note here that $\mathbf{H}^{-1}$ is a PDE solve and constitutes the main computational cost in equation 8.2. In order to incorporate multiples along with primary, we follow the SRME (surface-related multiple estimation) formulation proposed by Verschuur et al. (1992) where the total up- and down-going pressure wavefields, the Green's function and the source are related as follows

$$\mathbf{P} = \mathbf{G}(\mathbf{Q} + \mathbf{R}\mathbf{P}), \tag{8.3}$$

where G represents the Green's function, $\mathbf{P}$ is the total up-going wavefield, $\mathbf{R}\mathbf{P}$ represents the down-going receiver wavefield at the surface that acts as a generalized *areal-source* wavefield for the surface-related multiples, and $\mathbf{Q}$ is the down-going point-source wavefield. In this article we assume that the surface reflectivity can be approximated by -1. Note that each quantity in the above expression represents monochromatic variables. Therefore, we can replace the primary reflections data matrix in equation 8.2 with the total up-going data matrix $\mathbf{P}$ and redefine the extended image $\mathbf{E}$ as

$$\mathbf{E} = \mathbf{H}^{-*}\mathbf{P}_r^T\mathbf{P}(\mathbf{Q} - \mathbf{P})^*\mathbf{P}_s\mathbf{H}^{-*}. \tag{8.4}$$

In 2D, the extended image is a 5-dimensional function of *all* subsurface offsets and temporal shifts. So even in this case, it is prohibitively expensive to compute and store the extended image for all the subsurface points. To overcome this problem, we select l columns from $\mathbf{E}$ implicitly by multiplying this matrix with the tall matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_l]$ yielding:

$$\widetilde{\mathbf{E}} = \mathbf{E}\mathbf{W}, \tag{8.5}$$

where $\mathbf{w}_i = [0, \ldots, 0, 1, 0, \ldots, 0]$ represents a single scattering point with

the location of 1 corresponding to the i^{th} grid location of a point scatterer. Each column of $\widetilde{\mathbf{E}}$ represents a common-image point gather (CIPs) at the locations represented by $\mathbf{w}_i$. We follow van Leeuwen and Herrmann (2012b) to efficiently compute $\widetilde{\mathbf{E}}$ by combining equations (4) and (5) as

$$\widetilde{\mathbf{E}} = \mathbf{E}\mathbf{W} = \mathbf{H}^{-*}\mathbf{P}_r^T\mathbf{P}(\mathbf{Q}-\mathbf{P})^*\mathbf{P}_s\mathbf{H}^{-*}\mathbf{W}. \tag{8.6}$$

As we can see the computational cost of calculating $\widetilde{\mathbf{E}}$ is $2l$ PDE solves plus the cost of correlating the areal-source and the data matrices. Thus, the cost of computing the CIPs does not depend on the number of sources or the number of subsurface offsets, as it does in the conventional methods for computing image gathers (Sava and Vasconcelos, 2011). This is particularly beneficial when we are interested in computing only a few CIPs. Here, Equation 8.6 defines an extended migration operator that maps the total up-going data matrix to an extended image. The corresponding extended demigration operator, $\mathcal{F}$, is defined as

$$\mathcal{F}(\widetilde{\mathbf{E}}) = \mathbf{P}_r\mathbf{H}^{-1}\widetilde{\mathbf{E}}((\mathbf{Q}-\mathbf{P})^*\mathbf{P}_s\mathbf{H}^{-*}\mathbf{W})^*, \tag{8.7}$$

where $\mathcal{F}$ is a linear operator w.r.t the sources. To compute reliable amplitudes for extended image, we estimate the least-squares extended image by solving

$$\underset{\widetilde{\mathbf{E}}}{\text{minimize}} \quad \frac{1}{2}\|\mathbf{P}-\mathcal{F}(\widetilde{\mathbf{E}})\|_F^2, \tag{8.8}$$

where $\|.\|_F^2$ is the Frobenius norm of the matrix (sum of the squared entries). In the above expression we assume that $\mathbf{Q}$ is known, however, in the proposed framework $\mathbf{Q}$ can be estimated as well (Aravkin et al., 2013b).

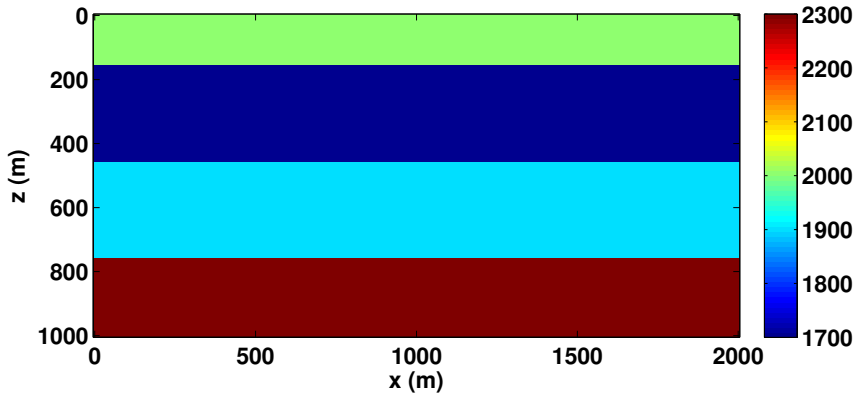
8.4 Results

To test the proposed algorithm for computing common-image gathers, we use a synthetic four-layer velocity model (with a grid sampling of 10 m) as shown in Figure 8.1(a). We place the sources and receivers on the surface with a spatial interval of 10 m. We use a finite-difference frequency-domain code to generate the synthetic data sets. The source signature is a Ricker wavelet with a peak frequency of 10 Hz. Figure 8.1(b) shows the background velocity model used to obtain least-squares reverse-time migration (LSRTM) images and common-image gathers (CIGs). We use 15 LSQR iterations to obtain the LSRTM images and the CIGs. We construct common-image gathers at $x = 1000$ m for all z. Figures 8.2(a) and 8.3(a) show the migrated

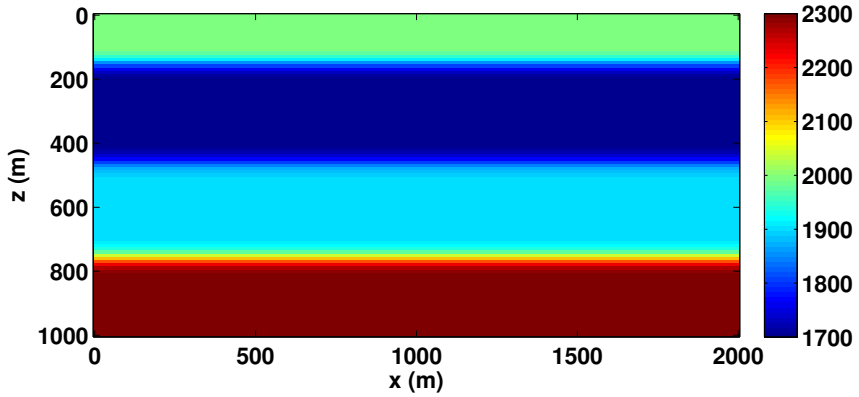
image and CIGs when we only use primary reflection data and $\mathbf{Q}$ as the source function. As expected, the LSRTM image and the image gathers are fully focused. Next, we use the total up-going wavefield $\mathbf{P}$ as input, but still use $\mathbf{Q}$ as the source function. Since the multiple reflections are not properly dealt with, we can see the focused ghost event generated by the multiple wavefields in Figures 8.2(b) and 8.3(b). As mentioned before, the presence of such ghost events can mislead the velocity analysis process. Finally, we obtain the image and the image gathers using the total up-going wavefield $\mathbf{P}$ along with the areal source function $\mathbf{Q}-\mathbf{P}$. We can see in Figures 8.2(c) and 8.3(c) that incorporating the areal sources in the LSRTM and extended images help mitigate the ghost reflector.

8.5 Conclusion

We have proposed a methodology for forming extended image gathers from data containing both primaries and multiples. The approach is based on earlier developments in simultaneous least-squares imaging of primaries and multiples, and matrix-probing methods to compute the extended image in a computationally efficient manner. The result is a method that creates a true-amplitude extended image gather directly from data containing surface related multiples, at a computational cost that is roughly equivalent to a least-squares migration. Numerical results on a simple synthetic model are promising and encourage further research.

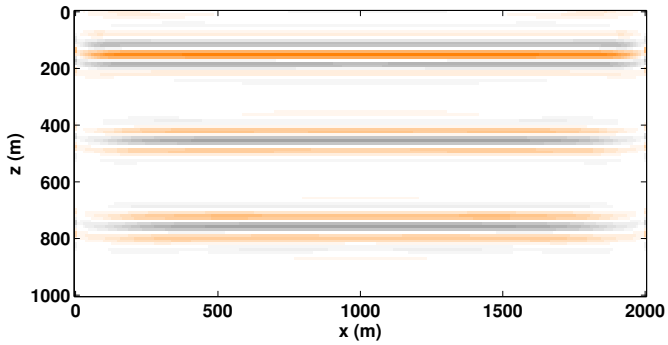


(a)

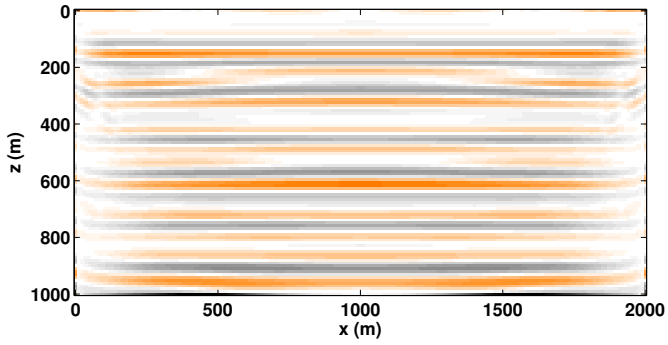


(b)

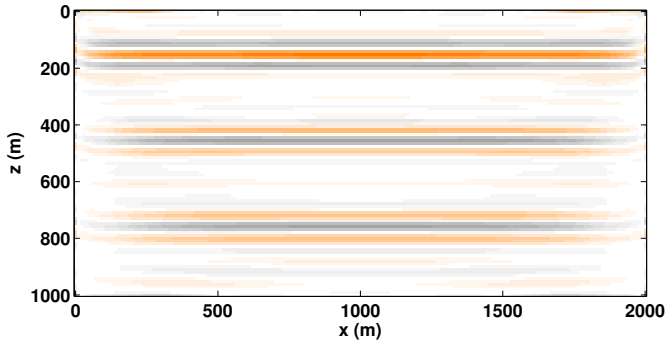
Figure 8.1: **(a)**True velocity model. **(b)** Background velocity model used in extended imaging.



(a)

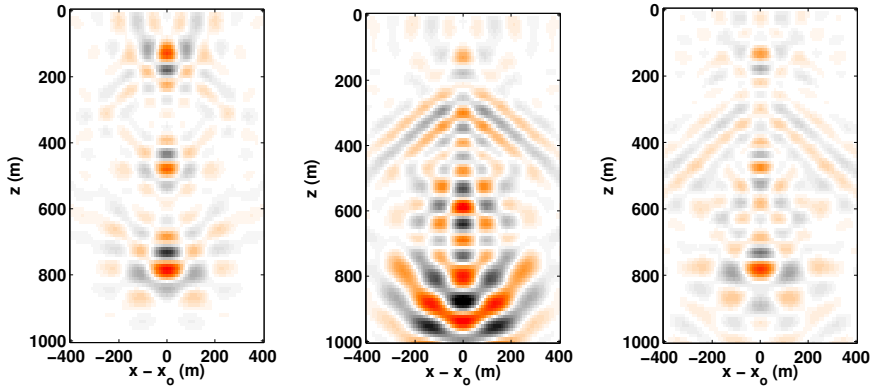


(b)



(c)

Figure 8.2: Least-squares reverse-time migration. **(a)** Primary data only. **(b)** Total up-going data is used but areal source is not used. We can clearly see the ghost reflector when we do not use the areal source. **(c)** Total up-going data is used along with the areal source. Incorporation of areal source removes the ghost reflector.



(a) (b) (c)

Figure 8.3: Least-squares common-image gather extracted along $x =$ 1000m and for all z. **(a)** Primary data only. **(b)** Total up-going data is used but areal source is not used. **(c)** Total up-going data is used along with the areal source. We can clearly see the effect of ghost reflectors (in the middle) which can be removed (on the right) via incorporating the areal source.

Chapter 9

Conclusion

In response of the objectives we aim to achieve through this research, which we detailed in the introduction chapter, we draw the following conclusions:

1. When used to image surface-related multiples, direct application of the cross-correlation imaging condition, i.e., reverse-time migration of these multiples, produces strong coherent artifacts associated with these multiples. Using the deconvolutional imaging condition can suppress these artifacts to some extent, but its performance is compromised in the presence of complex geologies. Through the presented work, we have observed that using an iterative inversion approach, which maps the surface-related multiples to the true "primary-image", can significantly suppress these artifacts.

2. Iterative inversions are prohibitively expensive in computation, as each iteration involves several wave-equation solves for each monochromatic source experiment. The incorporation of surface-related multiples exacerbates the issue as one also need to predict these multiples in each iteration, for example via the surface-related-multiple-elimination type of data convolutions (Verschuur et al., 1992), which are also computationally expensive. In this work, we address the two issues by *(i)* using wave-equation solvers to implicitly carry out the multiple prediction and therefore save most of the cost of performing data convolutions; and *(ii)* reducing the number of wave-equation solves by *subsampling* the monochromatic source experiments. As a result, we can obtain an inverted image with roughly the same number of wave-equation solves as in one reverse-time migration with all the data. We effectively remove the cross-talk artifacts introduced by the subsampling

by adopting curvelet-domain sparsity-promoting and rerandomization (i.e., we draw new data samples every once in a while) techniques, and therefore the image quality is not compromised.

3. We have observed that the proposed method gains robustness *(i)* to velocity errors in the background migration model by curvelet-domain sparsity promoting; and *(ii)* to linearization errors by adopting the rerandomization technique. Robustness to these types of errors are important for the method to be applicable to field data sets.

4. We can obtain true-amplitude images by the joint inversion of primaries and all orders of surface-related multiples. Using field seismic data, we observe that by joint inversion, we can reap benefits from the high signal-to-noise ratio of primaries and the wider illumination coverage of multiples by joint inversion. The joint inversion does not rely on the knowledge of the wavelet — we estimate the source wavelet on the fly — and we show with synthetic examples that the image we produce with source estimation is comparable to that obtained with the true source wavelet. A key contribution of this work is that by incorporating surface-related multiples in the joint imaging process, we resolve the amplitude ambiguity in both the image and the source wavelet, by exploiting the self-consistency between primary events and their corresponding surface-related-multiple events. We show with synthetic examples that, assuming no modelling errors, we can recover the original true amplitude of both the image and the source wavelet.

9.1 Significance and potential application

As we discussed above, an iterative inversion approach, also known as the least-squares migration, is needed to image the surface-related multiples. Although the promising benefits of using surface-related multiples in seismic imaging have been reported in the literature, the wide adoption of the least-squares migration by the oil and gas exploration industry has long been hindered by the expensive computational cost, especially for 3D data sets. We show with synthetic and field data examples that we can produce high-quality least-squares migrated images with roughly the same number of wave-equation solves as one reverse-time migration, compared with the combined cost of one surface-related multiple elimination and one reverse-time migration in conventional seismic data processing. Therefore we believe that the proposed work has great potential to promote the adoption of the

least-squares migration that makes active use of surface-related multiples by the industry, especially for 3D data sets.

Moreover, via the adaptive source estimation method proposed in this work, joint imaging of primaries and all orders of surface-related multiples obviates the necessity of primary/multiple separation during the imaging procedure, given the knowledge of a reasonably accurate velocity model. With potential extension of the proposed methodology to extended imaging, as discussed in Chapter 8, we may be able to avoid this wavefield separation for velocity-model building as well. We are therefore making great progress towards a seismic data processing workflow that does not involve this wavefield separation altogether. This improvement is especially important as seismic acquisition techniques move towards simultaneous/blended acquisitions. As de-multiple methods such as surface-related multiple elimination relies on conventional "sequential" data, geophysical practitioners now first need to recover conventional data from the simultaneous/blended data to separate primaries and multiples. Our contribution will potentially enable geophysical practitioners to directly process seismic data from simultaneous/blended acquisitions.

On a side note, the proposed method, apart from being computationally efficient, also facilitates parallel computing and can therefore benefit from modern parallel computing hardware. The algorithm can be parallelized over frequencies or source experiments. As in each iteration, we only use a small subset of data, we greatly decrease the memory requirement for each computational thread, and therefore the computational time spent on data transmission that greatly stains computer input/output. The above benefit facilitates the use of massively parallel accelerators such as Graphical Processing Units (GPUs).

9.2 Limitation and future work

In this presented work, we implemented and tested the proposed method for 2D seismic data. A future research topic is to extend the work for 3D seismic data, to account for the 3D wave-propagation effects in field seismic data.

We mainly discussed the relatively strong surface-related multiples, and largely ignored the usually much weaker internal multiples. As we discussed in Chapter 2, when properly used, internal multiples can also improve subsurface illumination. By extending our proposed method, via a non-linear-migration type of approach that gradually builds up sharp contrasts in the velocity model, we will investigate how to properly image the internal mul-

tiples.

The proposed method is designed for marine surface data, i.e., both sources and receivers are placed close to the surface, for example, the towed-streamer acquisition. We will investigate how we can adapt the method to work with ocean-bottom data, and how the change of the receiver horizon will change the way that primaries and multiples interact.

The proposed method requires the input data to be free of the downgoing wavefield at the surface (i.e., receiver ghosts). Therefore, a receiver-side deghost process should be applied to a data set before we apply the proposed method to it. We will investigate how we can incorporate the receiver ghosts, instead of removing them, in the proposed inversion approach.

In our implementation of the proposed method, we used the highly sophisticated SPGℓ_1 solver for the sparsity-promoting optimization program. The complexity of the algorithm can hinder the uptake of the proposed method by the industry. We have already started experimenting with alternative approaches, notably the linearized Bregman projection approach that we discussed in Chapter 7. Preliminary results are promising, and we will investigate its performance in more details in our future research.

To avoid the primary/multiple separation altogether in the seismic data processing workflow, we also attempt to extend the proposed methodology to extended imaging used in migration velocity analysis, as we discussed in Chapter 8. We also see promising results and will continue our research on this topic.

Bibliography

Aravkin, A. Y., J. Burke, and M. Friedlander, 2013a, Variational properties of value functions: SIAM Journal on Optimization, **23**, 1689–1717. → pages 64, 68, 70

Aravkin, A. Y., and T. van Leeuwen, 2012, Estimating nuisance parameters in inverse problems: Inverse Problems, **28**. → pages 35, 64, 69

Aravkin, A. Y., T. van Leeuwen, H. Calandra, and F. J. Herrmann, 2012, Source estimation for frequency-domain FWI with robust penalties: Presented at the EAGE technical program, EAGE. → pages 65

Aravkin, A. Y., T. van Leeuwen, and N. Tu, 2013b, Sparse seismic imaging using variable projection: Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2065–2069. → pages 35, 45, 122

Baysal, E., D. Kosloff, and J. Sherwood, 1983, Reverse time migration: GEOPHYSICS, **48**, 1514–1524. → pages 2, 21, 64, 67

Beck, A., and M. Teboulle, 2009, A fast iterative shrinkage-thresholding algorithm for linear inverse problems: SIAM J. Img. Sci., **2**, 183–202. → pages 114

Berkhout, A. J., 1993, Migration of multiple reflections: SEG Technical Program Expanded Abstracts, SEG, 1022–1025. → pages 14, 54, 93

Berkhout, A. J., and D. J. Verschuur, 2006, Imaging of multiple reflections: Geophysics, **71**, SI209–SI220. → pages 14

Berkhout, A. J., and C. P. A. Wapenaar, 1993, A unified approach to acoustical reflection imaging. ii: The inverse problem: The Journal of the Acoustical Society of America, **93**, 2017–2023. → pages 54

Biondi, B., and W. W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283. → pages 120

Brown, M., and A. Guitton, 2005, Least-squares joint imaging of multiples and primaries: GEOPHYSICS, **70**, S79–S89. → pages 15

Candès, E. J., L. Demanet, D. L. Donoho, and L. Ying, 2006a, Fast

discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, no. 3, 861–899. → pages 24, 45

Candès, E. J., and D. L. Donoho, 2004, New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities: Communications on Pure and Applied Mathematics, **57**, 219–266. → pages 4, 24, 66

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. → pages 24

Cavalca, M., and P. Lailly, 2005, 652, *in* Prismatic reflections for the delineation of salt bodies: SEG, 2550–2553. → pages 39

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129. → pages 23, 45

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, no. 3, 467–481. → pages 54, 55, 56

Dai, W., X. Wang, and G. T. Schuster, 2011, Least-squares migration of multisource data with a deblurring filter: Geophysics, **76**, R135–R146. → pages 45

Delprat-Jannaud, F., and P. Lailly, 2008, Wave propagation in heterogeneous media: Effects of fine-scale heterogeneity: Geophysics, **73**, T37–T49. → pages 39

Donoho, D., 2006, Compressed sensing: Information Theory, IEEE Transactions on, **52**, 1289 –1306. → pages 24

Donoho, D., A. Maleki, and A. Montanari, 2009, Message passing algorithms for compressed sensing: Proceedings of the National Academy of Sciences, **106**, 18914–18919. → pages 26, 142

Dragoset, B., E. Verschuur, I. Moore, and R. Bisley, 2010, A perspective on 3d surface-related multiple elimination: Geophysics, **75**, 75A245–75A261. → pages 17

Dragoset, W. H., and Ž. Jeričević, 1998, Some remarks on surface multiple attenuation: Geophysics, **63**, 772–789. → pages 17, 85

Esser, E., T. T. Lin, R. Wang, and F. J. Herrmann, 2015, A lifted ℓ_1/ℓ_2 constraint for sparse blind deconvolution: Presented at the 77th EAGE conference and exhibition 2015, EAGE. → pages 65, 72

Fokkema, J. T., and P. M. van den Berg, 1993, Seismic applications of acoustic reciprocity: Elsevier. → pages 17

Fomel, S., M. Backus, M. DeAngelo, P. Murray, and B. A. Hardage, 2003, Multicomponent seismic data registration for subsurface characterization in the shallow gulf of mexico: Presented at the Offshore Technology Conference. → pages 65

Foster, D., and C. Mosher, 1992, Suppression of multiple reflections using the radon transform: Geophysics, **57**, 386–395. → pages 13

Golub, G., and V. Pereyra, 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate: SIAM Journal on Numerical Analysis, **10**, 413–432. → pages 64

———, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse problems, **19**, R1–R26. → pages 35, 64, 68, 69

Gray, S. H., 1997, True-amplitude seismic migration: A comparison of three approaches: Geophysics, **62**, 929–936. → pages 55

Gribonval, R., G. Chardon, and L. Daudet, 2012, Blind calibration for compressed sensing by convex optimization: Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, 2713–2716. → pages 48

Guitton, A., 2002, Shot-profile migration of multiple reflections: SEG Technical Program Expanded Abstracts, 1296–1299. → pages 14, 37, 93, 96

Guitton, A., A. Valenciano, D. Bevc, and J. Claerbout, 2007, Smoothing imaging condition for shot-profile migration: Geophysics, **72**, S149–S154. → pages 54, 56

Guitton, A., and D. J. Verschuur, 2004, Adaptive subtraction of multiples using the ℓ_1-norm: Geophysical prospecting, **52**, 27–38. → pages 13

Hampson, D., 1986, Inverse velocity stacking for multiple elimination: SEG Technical Program Expanded Abstracts, 422–424. → pages 13

He, R., B. Hornby, and G. T. Schuster, 2007, 3d wave-equation interferometric migration of vsp free-surface multiples: Geophysics, **72**, S195–S203. → pages 14

He, R., and G. Schuster, 2003, Least-squares migration of both primaries and multiples: SEG Technical Program Expanded Abstracts, SEG, 1035–1038. → pages 14

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**. → pages 115

Herman, M., and T. Strohmer, 2010, General deviants: An analysis of perturbations in compressed sensing: Selected Topics in Signal Processing, IEEE Journal of, **4**, 342–349. → pages 48

Herrmann, F. J., 2012, Pass on the message: recent insights in large-scale sparse recovery: Presented at the 74th EAGE Conference & Exhibition, EAGE. → pages 25, 71, 142

Herrmann, F. J., Y. A. Erlangga, and T. T. Lin, 2009, Compressive

simultaneous full-waveform simulation: Geophysics, **74**, A35. → pages 23

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248. → pages 4, 24

Herrmann, F. J., and X. Li, 2012, Efficient least-squares imaging with sparsity promotion and compressive sensing: Geophysical Prospecting, **60**, 696–712. → pages 15, 20, 23, 24, 25, 26, 28, 30, 39, 45, 48, 54, 56, 64, 66, 67, 70, 103, 112

Herrmann, F. J., P. P. Moghaddam, and C. Stolk, 2008a, Sparsity- and continuity-promoting seismic image recovery with curvelet frames: Applied and Computational Harmonic Analysis, **24**, 150–173. → pages 24, 25, 33, 64

Herrmann, F. J., D. Wang, G. Hennenfent, and P. P. Moghaddam, 2008b, Curvelet-based seismic data processing: a multiscale and nonlinear approach: Geophysics, **73**, A1–A5. → pages 24, 25, 77

Herrmann, F. J., D. Wang, and D. J. Verschuur, 2008c, Adaptive curvelet-domain primary-multiple separation: Geophysics, **73**, A17–A21. → pages 13, 37

Jo, C., C. Shin, and J. H. Suh, 1996, An optimal 9-point, finite-difference, frequency-space, 2-d scalar wave extrapolator: Geophysics, **61**, 529–537. → pages 18

Jumah, B., and F. J. Herrmann, 2014, Dimensionality-reduced estimation of primaries by sparse inversion: Geophysical Prospecting, **62**, 972–993. → pages 18

Kaufman, L., 1975, A variable projection method for solving separable nonlinear least squares problems: BIT Numerical Mathematics, **15**, 49–57. → pages 64

Kumar, R., T. van Leeuwen, and F. J. Herrmann, 2013, Ava analysis and geological dip estimation via two-way wave-equation based extended images: Presented at the SEG. → pages 120

———, 2014, Extended images in action: efficient WEMVA via randomized probing: Presented at the EAGE. → pages 120

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on inverse scattering: theory and application, Society for Industrial and Applied Mathematics, Philadelphia, PA, 206–220. → pages 7, 21, 56, 67

Li, M., J. Rickett, and A. Abubakar, 2013, Application of the variable projection scheme for frequency-domain full-waveform inversion: Geophysics, **78**, R249–R257. → pages 65

Lin, T. T., and F. J. Herrmann, 2013, Robust estimation of primaries by sparse inversion via one-norm minimization: Geophysics, **78**, R133–R150. → pages 14, 17, 19, 33, 65, 72, 93, 97

Lin, T. T., N. Tu, and F. J. Herrmann, 2010a, Sparsity-promoting migration from surface-related multiples: SEG Technical Program Expanded Abstracts, SEG, 3333–3337. → pages 14

———, 2010b, Sparsity-promoting migration from surface-related multiples: SEG Technical Program Expanded Abstracts, SEG, 3333–3337. → pages 93

Liu, Y., X. Chang, D. Jin, R. He, H. Sun, and Y. Zheng, 2011, Reverse time migration of multiples for subsalt imaging: Geophysics, **76**, WB209–WB216. → pages 6, 14, 21, 37, 54, 59, 93

Long, A. S., S. Lu, D. Whitmore, H. LeGleut, R. Jones, N. Chemingui, and M. Farouki, 2013, Mitigation of the 3d cross-line acquisition footprint using separated wavefield imaging of dual-sensor streamer seismic: Presented at the 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013, EAGE. → pages 14

Lorenz, D., F. Schpfer, and S. Wenger, 2014, The linearized Bregman method via split feasibility problems: Analysis and generalizations: SIAM Journal on Imaging Sciences, **7**, 1237–1262. → pages 115

Lorenz, D. A., S. Wenger, F. Schöpfer, and M. Magnor, 2014, A sparse Kaczmarz solver and a linearized Bregman method for online compressed sensing: Presented at the Image Processing (ICIP), 2014 21th IEEE International Conference on. → pages 112, 114

Lu, G., B. Ursin, and J. Lutro, 1999, Model-based removal of water-layer multiple reflections: Geophysics, **64**, 1816–1827. → pages 13

Lu, S., D. Whitmore, A. Valenciano, and N. Chemingui, 2014a, Enhanced subsurface illumination from separated wavefield imaging: First Break, **32**, 87–92. → pages 3, 120

———, 2014b, Enhanced subsurface illumination from separated wavefield imaging: First Break, **32**, 87–92. → pages 93

Lu, S., N. D. Whitmore, and A. A. Valenciano, 2011, Imaging of primaries and multiples with 3d seam synthetic: SEG Technical Program Expanded Abstracts, 3217–3221. → pages 14, 54

Maharramov, M., and B. Biondi, 2014, Improved depth imaging by constrained full-waveform inversion: SEP report, **155**, 19–24. → pages 114

Malcolm, A. E., B. Ursin, and V. Maarten, 2009, Seismic imaging and illumination with internal multiples: Geophysical Journal International, **176**, 847–864. → pages 39

Mallat, S. G., 2009, A wavelet tour of signal processing: Academic Press. → pages 26

Métivier, L., 2011, Interlocked optimization and fast gradient algorithm for a seismic inverse problem: Journal of Computational Physics, **230**, 7502–7518. → pages 15, 39

Métivier, L., P. Lailly, F. Delprat-Jannaud, and L. Halpern, 2011, A 2d nonlinear inversion of well-seismic data: Inverse Problems, **27**, 055005. → pages 15, 39

Montanari, A., 2010, Graphical models concepts in compressed sensing: arXiv:1011.4328v3 [cs.IT]. → pages 25, 26, 142

Muijs, R., J. O. A. Robertsson, and K. Holliger, 2007, Prestack depth migration of primary and surface-related multiple reflections: Part i—imaging: Geophysics, **72**, S59–S69. → pages 6, 14, 21, 37, 54, 59, 93

Mulder, W. A., and R. E. Plessix, 2003, 225, *in* One-way and two-way wave-equation migration: SEG, 881–884. → pages 76

Needell, D., and J. A. Tropp, 2014, Paved with good intentions: Analysis of a randomized block kaczmarz method: Linear Algebra and its Applications, **441**, 199 – 221. (Special Issue on Sparse Approximate Solution of Linear Systems). → pages 26, 37, 116

Neelamani, R., A. Baumstein, and W. Ross, 2010, Adaptive subtraction using complex-valued curvelet transforms: Geophysics, **75**, V51–V60. → pages 24

Nemeth, T., H. Sun, and G. T. Schuster, 2000, Separation of signal and coherent noise by migration filtering: Geophysics, **65**, 574–583. → pages 13

Nemeth, T., C. Wu, and G. T. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221. → pages 23, 45, 64

Peacock, K. L., and S. Treitel, 1969, Predictive deconvolution: theory and practice: Geophysics, **34**, 155–169. → pages 13

Plessix, R.-E., and W. A. Mulder, 2004, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, 975–987. → pages 7

Poole, T., A. Curtis, J. Robertsson, and D.-J. van Manen, 2010, Deconvolution imaging conditions and cross-talk suppression: Geophysics, **75**, W1–W12. → pages 14, 93

Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901. → pages 69

Prucha, M., and B. Biondi, 2002, Subsalt event regularization with steering filters: SEG Technical Program Expanded Abstracts,

1176–1179. → pages 25
Reiter, E. C., M. N. Toksoz, T. H. Keho, and G. M. Purdy, 1991, Imaging with deep-water multiples: Geophysics, **56**, 1081–1086. → pages 15
Rickett, J., 2013, The variable projection method for waveform inversion with an unknown source function: Geophysical Prospecting, **61**, 874–881. → pages 65, 69
Robinson, E. A., 1967, Predictive decomposition of time series with application to seismic exploration: GEOPHYSICS, **32**, 418–484. → pages 74
Romero, L., D. Ghiglia, C. Ober, and S. Morton, 2000, Phase encoding of shot records in prestack migration: GEOPHYSICS, **65**, 426–436. → pages 24
Ryberg, T., and M. Weber, 2000, Receiver function arrays: a reflection seismic approach: Geophysical Journal International, **141**, 1–11. → pages 39
Sava, P., and A. Guitton, 2005, Multiple attenuation in the image space: Geophysics, **70**, V10–V20. → pages 13
Sava, P., and I. Vasconcelos, 2011, Extended imaging conditions for wave-equation migration: Geophysical Prospecting, **59**, 35–55. → pages 120, 122
Schuster, G. T., J. Yu, J. Sheng, and J. Rickett, 2004, Interferometric/daylight seismic imaging: Geophysical Journal International, **157**, 838–852. → pages 14, 93
Shan, G., 2003, Source-receiver migration of multiple reflections: SEG Technical Program Expanded Abstracts, SEG, 1008–1011. → pages 14
Shang, X., M. V. de Hoop, and R. D. van der Hilst, 2012, Beyond receiver functions: Passive source reverse time migration and inverse scattering of converted waves: Geophysical Research Letters, **39**, L15308. → pages 39
Shen, P., and W. W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–VE59. → pages 120
Stockham, T.G., J., T. Cannon, and R. Ingebretsen, 1975, Blind deconvolution through digital signal processing: Proceedings of the IEEE, **63**, 678–692. → pages 9
Stockham Jr, T. G., T. M. Cannon, and R. B. Ingebretsen, 1975, Blind deconvolution through digital signal processing: Proceedings of the IEEE, **63**, 678–692. → pages 65, 71
Strohmer, T., and R. Vershynin, 2009, A randomized kaczmarz algorithm with exponential convergence: Journal of Fourier Analysis and Applications, **15**, 262–278. → pages 26, 116
Symes, W. W., 2008, Migration velocity analysis and waveform inversion:

Geophysical Prospecting, **56**, 765–790. → pages 120

Terentyev, I. S., T. Vdovina, W. W. Symes, X. Wang, and D. Sun., 2014, iWave: a framework for wave simulation: http://www.trip.caam.rice.edu/software/iwave/doc/html/index.html. (Online; accessed 16-January-2014). → pages 32, 77, 82

The SMAART JV, 2014, Sigsbee2B FS & NFS 2D synthetic datasets: http://www.delphi.tudelft.nl/SMAART/sigsbee2b.htm. (Online; accessed 24-February-2014). → pages 31, 81

Tibshirani, R., 1996, Regression shrinkage and selection via the lasso: Journal of the Royal Statistical Society Series B Methodological, **58**, 267–288. → pages 25, 46, 67

Tu, N., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2013, Fast least-squares migration with multiples and source estimation: Presented at the 75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013, EAGE. → pages 35, 45

Tu, N., and F. J. Herrmann, 2012a, Imaging with multiples accelerated by message passing: SEG Technical Program Expanded Abstracts, SEG, 1–6, Chapter 682. → pages 26, 48, 54, 59

———, 2012b, Least-squares migration of full wavefield with source encoding: Presented at the EAGE. → pages 14

Ulrych, T. J., and M. D. Sacchi, 2005, Information-based inversion and processing with applications: Elsevier Science. → pages 33

Ulrych, T. J., D. R. Velis, and M. D. Sacchi, 1995, Wavelet estimation revisited: The Leading Edge, **14**, 1139–1143. → pages 71

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. → pages 23, 28, 45, 46, 47, 64, 67, 70, 112, 141

van der Neut, J., and F. J. Herrmann, 2013, Interferometric redatuming by sparse inversion: Geophysical Journal International, **192**, 666–670. → pages 39

van Groenestijn, G. J. A., and D. J. Verschuur, 2009a, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, A23–A28. → pages 17, 65, 72, 93, 120

———, 2009b, Estimation of primaries and near-offset reconstruction by sparse inversion: Marine data applications: Geophysics, **74**, R119–R128. → pages 14, 19, 33

van Leeuwen, T., and F. J. Herrmann, 2012, Fast waveform inversion without source encoding: Geophysical Prospecting. → pages 112

———, 2012a, A parallel, object-oriented framework for frequency-domain wavefield imaging and inversion.: Technical Report TR-2012-03,

Department of Earth and Ocean Sciences, University of British Columbia, Vancouver. → pages 32

———, 2012b, Wave-equation extended images: computation and velocity continuation: Presented at the EAGE technical program, EAGE. → pages 122

———, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667. → pages 115

van Leeuwen, T., and W. A. Mulder, 2009, A variable projection method for waveform inversion: Presented at the 71st EAGE Conference & Exhibition. → pages 65

Verschuur, D. J., 1991, Surface-related multiple elimination, an inverse approach: PhD thesis, Delft University of Technology. → pages 17

———, 2006, Seismic multiple removal techniques: Past, present and future: EAGE Publications BV. → pages 2

———, 2011, Seismic migration of blended shot records with surface-related multiple scattering: Geophysics, **76**, A7–A13. → pages 3, 14, 54, 93

Verschuur, D. J., and A. J. Berkhout, 2009, Target-oriented, least-squares imaging of blended data: SEG Technical Program Expanded Abstracts, 2889–2893. → pages 55

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177. → pages 13, 16, 17, 32, 54, 65, 72, 75, 93, 96, 98, 121, 127

Wang, D., R. Saab, O. Yilmaz, and F. J. Herrmann, 2008, Bayesian wavefield separation by transform-domain sparsity promotion: Geophysics, **73**, 1–6. → pages 13

Wapenaar, K., 1998, Reciprocity properties of one-way propagators: Geophysics, **63**, 1795–1798. → pages 32

Weglein, A., F. Gasparotto, P. Carvalho, and R. Stolt, 1997, An inversescattering series method for attenuating multiples in seismic reflection data: GEOPHYSICS, **62**, 1975–1989. → pages 13

Whitmore, N. D., A. A. Valenciano, and W. Sollner, 2010, Imaging of primaries and multiples using a dual-sensor towed streamer: SEG Technical Program Expanded Abstracts, 3187–3192. → pages 6, 14, 37, 54, 93

Wiggins, J. W., 1988, Attenuation of complex water-bottom multiples by wave-equation-based prediction and subtraction: Geophysics, **53**, 1527–1539. → pages 13

Wong, M., B. Biondi, and S. Ronen, 2012, Imaging with multiples using linearized full-wave inversion: SEG Technical Program Expanded

Abstracts, SEG, 1–5, Chapter 706. → pages 14, 15, 93
———, 2014, Imaging with multiples using least-squares reverse time migration: The Leading Edge, **33**, 970–976. → pages 14, 120
Yilmaz, ., 2001, Seismic data analysis: Society of Exploration Geophysicists. → pages 1
Yin, W., S. Osher, D. Goldfarb, and J. Darbon, 2008, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing: SIAM Journal on Imaging Sciences, **1**, 143–168. → pages 114
Zuberi, A., and T. Alkhalifah, 2013, Imaging by forward propagating the data: theory and application: Geophysical Prospecting, **61**, 248–267. → pages 14, 93

Appendix A

Additional details of methodology

A.1 Solving the BPDN problem using SPGℓ_1

Using canonical linear algebra notations, the BPDN formulation to invert a linear system of equations $\mathbf{Ax} = \mathbf{b}$ with sparsity promoting using the ℓ_1-norm follows

$$\mathbf{BP}_\sigma : \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{b} - \mathbf{Ax}\|_2 \leq \sigma, \tag{A.1.1}$$

where τ is an estimate for the noise level in the data.

To efficiently solve the above $\mathbf{BP}_\sigma$ problem, SPGℓ_1 leverages the Spectrum Projected Gradient (SPG) method that efficiently solves the ℓ_1-constrained least-squares problem, aka the LASSO problem (van den Berg and Friedlander, 2008):

$$\mathbf{LS}_\tau : \quad \min_{\mathbf{x}} \|\mathbf{b} - \mathbf{Ax}\|_2 \quad \text{subject to.} \quad \|\mathbf{x}\|_1 \leq \tau, \tag{A.1.2}$$

where τ is a sparsity constraint on the solution vector $\mathbf{x}$.

van den Berg and Friedlander (2008) found that problem $\mathbf{BP}_\sigma$ and $\mathbf{LS}_\tau$ share the same solution for certain choices of σ and τ, which are related by the Pareto trade-off curve between the ℓ_2-norm of the data residual (i.e., the objective function of problem $\mathbf{LS}_\tau$) and the ℓ_1-norm of the solution vector (i.e., the objective function of problem $\mathbf{BP}_\sigma$). As the Pareto curve is continuously differentiable, SPGℓ_1 uses the Newton's method on the Pareto curve to determine the optimal value of τ given an estimate of σ. As a result,

SPGℓ_1 effectively solves a series of warm-started LASSO subproblems for gradually increasing τ's. By "warm-started", we mean that the solution for the l^{th} subproblem with $\tau^l < \tau^{l+1}$ serves as the initial guess for the next subproblem. Solving each subproblem $\mathbf{LS}_\tau$ itself using the SPG method involves several gradient updates. In each update, we need to evaluate the (expensive) action of $\mathbf{A}$ and $\mathbf{A}^*$, perform a line search, followed by a (cheap) ℓ_1 projection that promotes sparsity.

A.2 Subsampling noise and approximate message passing

In simple terms, the key idea of compressive sensing is to work with a flat sampling matrix $\mathbf{A}$ that creates noisy interferences, i.e., the action of $\mathbf{A}^*\mathbf{A}$ on a sparse vector yield a "noisy" sparse vector where the "noise" actually is incoherent crosstalks with a noise level that depends on degree of subsampling (aspect ratio of $\mathbf{A}$, defined as the ratio of the number of rows to the number of columns) and the sparsity of $\mathbf{x}$. In principle, this interference noise should be easier to remove than coherent noise. As explained by Donoho et al. (2009) and Herrmann (2012), this statement holds true for the first iteration where some noises are removed by projection onto the ℓ_1 ball (which corresponds to a single soft thresholding with the appropriate threshold level). However, later iterations suffer from the correlation built up between the model iterate and the sampling matrix $\mathbf{A}$, which makes the denoising step less efficient and may result in coherent artifacts.

In the context of our compressive imaging problem, this means that when we reduce the number of simultaneous sources and frequencies, which corresponds to a reduced aspect ratio of $\mathbf{A}$, we may get more coherent artifacts in the image that cannot be removed by simply using more iterations.

Recent insights into approximate message passing find that with a correction term in the iterative soft thresholding, this denoising step can continuously bring down the model errors in later iterations (Donoho et al., 2009; Montanari, 2010). While this approach relies on careful analysis, we found that rerandomization of source experiments has statistically the same effect in breaking down this correlation buildup, and fits well in the optimization framework of the proposed method (Herrmann, 2012).