# Primary estimation with sparsity-promoting bi-convex optimization

by

Tim Tai-Yi Lin

B.Sc., The University of British Columbia, 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)

October 2015

# Abstract

This thesis establishes a novel inversion methodology for the surface-related primaries from a given recorded seismic wavefield, called the Robust Estimation of Primaries by Sparse Inversion (Robust EPSI, or REPSI). Surface-related multiples are a major source of coherent noise in seismic data, and inferring fine geological structures from active-source seismic recordings typically first necessitates its removal or mitigation. For this task, current practice calls for data-driven approaches which produce only approximate multiple models that must be non-linearly subtracted from the data, often distorting weak primary events in the process. A recently proposed method called Estimation of Primaries by Sparse Inversion (EPSI) avoids this adaptive subtraction by directly inverting for a discrete representation of the underlying multiple-free subsurface impulse response as a set of band-limited spikes. However, in its original form, the EPSI algorithm exhibits a few notable shortcomings that impede adoption. Although it was shown that the correct impulse response can be obtained through a sparsest solution criteria, the current EPSI algorithm is not designed to take advantage of this finding, but instead *approximates* a sparse solution in an ad-hoc manner that requires practitioners to decide on a multitude of inversion parameters. The *Robust EPSI* method introduced in this thesis reformulates the original EPSI problem as a formal bi-convex optimization problem that makes obtaining the sparsest solution an explicit goal, while also reliably admit satisfactory solutions using contemporary self-tuning gradient methods commonly seen in large-scale machine learning communities. I show that the Robust EPSI algorithm is able to operate successfully on a variety of datasets with minimal user input, while also producing a more accurate model of the subsurface impulse response when compared to the original algorithm. Furthermore, this thesis makes several contributions that improves

the capability and practicality of EPSI: a novel scattering-based multiple prediction model that allows Robust EPSI to deal with wider near-offset receiver gaps than previously demonstrated for EPSI, as well as a multigrid-inspired continuation strategy that significantly reduces the computation time needed to solve EPSI-type problems. These additions are enabled by and built upon the formalism of the Robust EPSI as developed in this thesis.

# Preface

This thesis consists of my original research, conducted at the Department of Earth, Ocean and Atmospheric Sciences of The University of British Columbia, Vancouver, Canada, under the supervision of Professor Felix Herrmann as part of the Seismic Laboratory of Imaging and modelling (SLIM). I prepared the manuscript of the thesis in its entirety. Some chapters contain previously published materials for which I was the main investigator and author.

A version of Chapter 3 was published in an academic journal (Lin, T.T.Y., and F.J. Herrmann, 2013, Robust estimation of primaries by sparse inversion via one-norm minimization: Geophysics, **78**, R133–R150, doi: 10.1190/geo2012-0097.1).

A version of Chapter 4 has been submitted to an academic journal. Part of this chapter was also published as an extended abstract (Lin, T.T.Y., and F.J. Herrmann, 2014, Mitigating data gaps in the estimation of primaries by sparse inversion without data reconstruction: 84th Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2014-1680.1).

A version of Chapter 5 was published as an expanded abstract (Lin, T.T.Y., and F.J. Herrmann, 2014, Multilevel acceleration strategy for the robust estimation of primaries by sparse inversion: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20140672)

Figures 2.2 and 2.3 are reprinted from Marine Geology, **160**, Saatcilar et al: Character of active faulting in the North Aegean Sea, 339–353, Copyright 1999, with permission from Elsevier (via Copyright Clearance Center license number 3654371168973).

# Table of Contents

# List of Figures

# List of Algorithms

# Glossary

| | |
|---|---|
| **EPSI** | Estimation of Primaries by Sparse Inversion |
| **REPSI** | Robust EPSI |
| **SRME** | Surface-Related multiple Elimination |
| **BP** | Basis Pursuit |
| **BPDN** | Basis Pursuit DeNoising |
| **SPG$\ell_1$** | Spectral Projected Gradient for $\ell_1$ problems |
| **NMO** | Normal MoveOut |

# Acknowledgments

In the spring of 2009, my advisor Prof. Felix Herrmann and I visited his alma mater, Technische Universiteit Delft, in the Netherlands. I had worked on the problem of separating overlapping or simultaneously acquired marine data as an undergraduate, and we had intended to borrow their existing water tank equipment to simulate these (at the time) novel acquisitions patterns, despite knowing that any student who have worked on it have long since graduated. Unimaginably for everyone involved, it proved excruciatingly difficult for current faculty members to get the equipment going again. During the ensuing downtime, I was graciously shown by Dr. Eric Verschuur and his student Gert-Jan van Groenestijn some of their work on what eventually became EPSI. I rapidly gained interest in this new method, and instead of trying to resolve the water tank situation I have decided to work on the EPSI problem, eventually doing enough work to culminate into a PhD program. I sincerely thank Dr. Verschuur and van Groenestijn, as well as the unknown graduate students who made sure that attempts to revive the water tank remains an elusive necromancy, for the beginnings of my involvement in surface-related multiple removal.

I have spent a long time at UBC, and there are many people to whom I am deeply indebted to for teaching me things, or putting up with me in general. Most of my geophysical knowledge came from discussions with colleagues and faculty in the department of Earth, Ocean and Atmospheric Sciences. I would like to thank my predecessors at the Seismic Laboratory for Imaging and Modelling: Gilles Hennenfent, Carson Yarham, and Mohammad Maysami, for taking care of me back while I was an undergraduate co-op student. Thanks to Prof. Michael Bostock for teaching a great course on exploration seismology which bootstrapped me into the field. To my current colleagues at SLIM, you are all amazing people, and I truly enjoyed my time with everyone in the group for the past 6 years.

The staff at SLIM had been overwhelmingly helpful, and I have great memories of harassing these poor folks folks while being reciprocated with nothing but generosity and support. Thanks to Henryk Modzelewski for putting up with me all these years, and to Miranda Joyce for always being so concerned and helpful, especially these last few months concerning my graduation. Cheers to Ian Hanlon for the endless banter exchanges.

I learned a tremendous amount about optimization theory and numerical linear algebra from the privilege of being surrounded with great computer scientists and applied mathematicians. I appreciated discussions with and mentorships from Prof. Michael Friedlander, Ewout van den Berg, Prof. Ozgur Yilmaz, Rayan Saab, Ives Macedo, Yogi Erlangga, Mark Schmidt, Prof. Eldad Haber, Aleksandr Aravkin, Tristan van Leeuwen, Hassan Mansour, and Rongrong Wang. This thesis is written in remembrance of Ernie Esser, with whom I greatly enjoyed conversing and exchanging ideas during his sadly brief yet deeply impactful time among his peers.

The Lasso solver used in Robust EPSI is derived from SPG$\ell_1$, which is the work of Michael Friedlander and Ewout van den Berg. They were also responsible for the Sparse Linear Operator Toolbox (SPOT) upon which the implementation of Robust EPSI was based.

The Pluto 1.5 dataset used in Chapter 3 was provided by the SMAART joint venture.

The synthetic salt dome model used throughout the thesis and the Gulf of Suez dataset used in Chapter 3 were both generously provided by Dr. Eric Verschuur and G.J.A. van Groenestijn. The North Sea dataset used in Chapters 4 and 5 was provided with permission from PGS via Dr. Verschuur.

Last but certainly not least, my eternal appreciation goes to my mentor and advisor Prof. Felix Herrmann, for believing in me nine years ago as wee undergrad, pushing me to work on ever greater things, for providing me with the opportunity to gain exposure with the international community in this field, for his generous support of my meanderings and understand of my shortcomings, and for his infectious enthusiasm in strange and esoteric topics that inevitably turn out to be revolutionary. Thank you for this amazing journey and all the wonderful things that came along the way.

# Chapter 1

# Introduction

The iconic metal-and-glass glazed dome of the Royal Albert Hall in London, as originally envisioned by its architects, was meant to evoke sentiments of the memorable Crystal Palace as the venue of the original Great Exhibition: often considered the most notable legacy of the Hall's namesake. It was one of the first dome structures constructed in this way, and that contributed to oversights that led to what eventually became one of the highest profile acoustic catastrophes for any concert hall in history.

As the future King Edward VII (then still the Prince of Wales) gave the commencement speech on March 29, 1871, reporters from The Times immediately noted a problem (Metkemeijer, 2002), writing: "The address was slowly and distinctly read by his royal Highness, but the reading was somewhat marred by an echo which seemed to be suddenly awoke from the organ or picture gallery, and repeated the words with a mocking emphasis which at another time would have been amusing." An opening concert followed, which only served to further cement the Hall's eventual reputation as "the only place where a British composer could be sure of hearing his work twice." If you would visit today, look up and you can see the latest in a century of engineering effort to defeat the echoes: 135 distinctive-looking acoustic diffusion discs known affectionately as the "Royal Albert Hall mushrooms."

The work in this thesis addresses a similar coherent echo problem in the context of exploration seismology. Exploration seismology utilizes seismic waves that behave very similarly to acoustic waves. In the terminology of the discipline, these coherent echoes are called *multiple reflections* or often simply *multiples*.

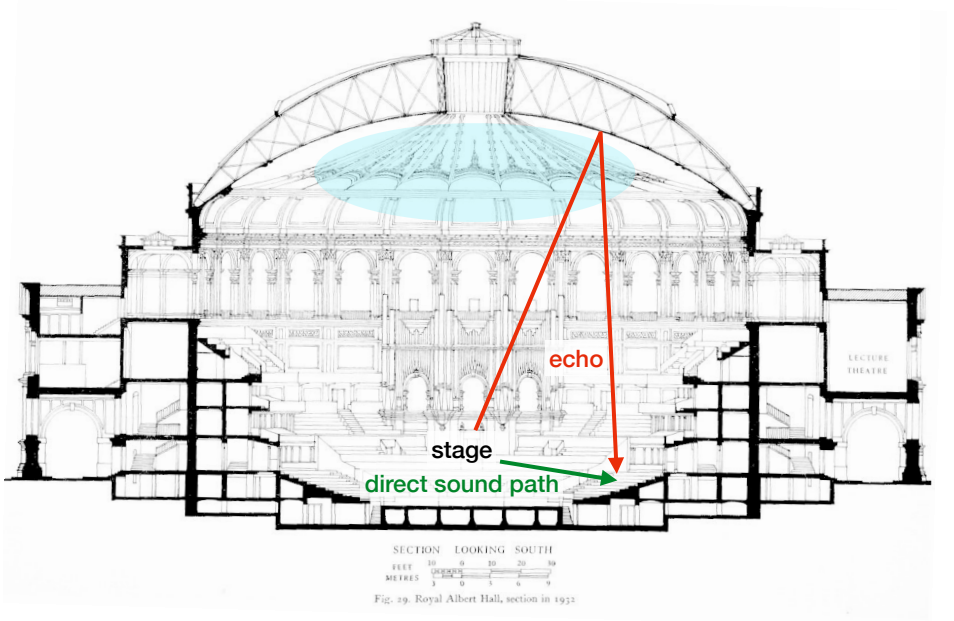



**Figure 1.1** Schematic diagram of the Royal Albert Hall, and the travel paths of direct sound waves (in green) and echoes reflecting off the glazed dome (in red). The canvas velarium (indicated in blue) was later added soon after opening to mitigate some of the echoes.

Much like the Royal Albert Hall, the Earth's subsurface has a problematic "roof": the boundary between the ground and the Earth's atmosphere. Multiples that are entirely caused by this air-ground (or air-water, in the case of surveys over oceans and lakes) interface are called *surface multiples*, and is historically regarded as one of the most severe and perennial sources of unwanted noise in surveying projects. Other types of multiples might also exist in the absence of this interface—usually attributed to reverberations between subterranean interfaces—which are not considered surface multiples. Surface multiples by themselves are of importance because they tend to be systematically the strongest in amplitude for any particular survey. In this thesis, I will develop a comprehensive inversion framework called **Robust Estimation of Primaries by Sparse Inversion (REPSI)**, for reliably identifying and reconstructing the so-called *surface-related primaries*, which is the desired part of recorded seismic datum that are *free* of surface multiples.

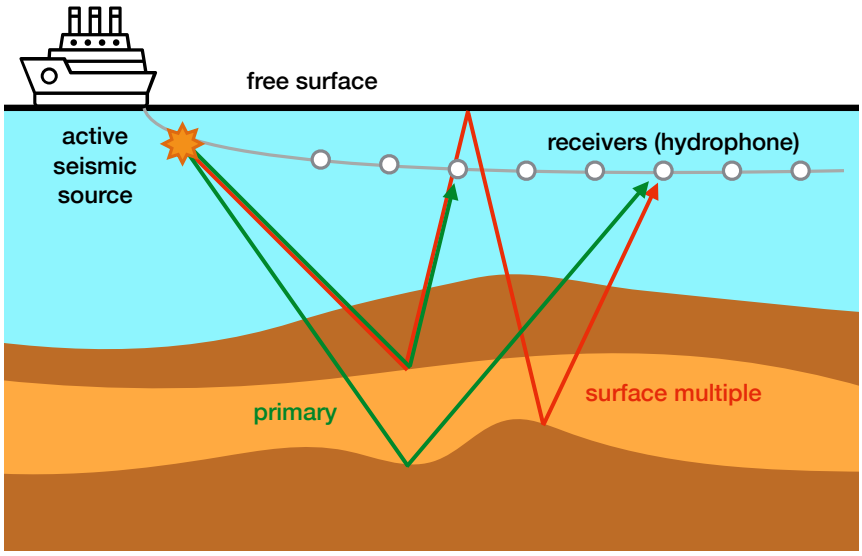



**Figure 1.2** Diagram of a typical marine survey, with primary raypaths indicated in green and a surface multiple raypath indicated in red.

Multiples in general degrade the efficacy and accuracy of seismic surveys by hindering their purpose in several ways. Seismic exploration involves systematically propagating seismic waves through the Earth by exciting point-source energy at its surface. After reflecting or refracting at geological boundaries, some part of these waves return to the surface and are measured over a survey area of several hundred square kilometres for tens of seconds. The arrival time of these responses can be converted into depth values, which allows the spatial distribution of subsurface interfaces to be systematically mapped. This inference process is difficult and delicate, even if one starts with some idealized data without any external noise. The presence of strong surface multiples can further frustrate this goal by either being misconstrued as evidence of a phantom boundary, or, as is more often the case, simply overlap with and thus corrupt responses from a true boundary deeper underground.

The art of removing seismic multiple consists almost entirely of signal-processing techniques applied to the recorded data after the field surveying process. Active or mechanical cancellation during acquisition are non-

existent due to the scale and difficulty of such an approach (unlike with concert halls, survey teams do not usually have the resource to blanket the Earth with elaborately positioned diffusive canvases).

Due to the prevalence and severity of multiples in seismic data, the history of methods for its removal is rich. In the past these methods are generally *filters* that are designed to reject multiples based on its specific characteristics. This is an approach that goes back to the 1950s. Even though multiples look very similar to their primary counterparts, they in general have some particular key characteristics that differentiate them in most seismic data. Surface multiples that reverberate between two very strong reflectors (e.g., the sea surface and the sea bottom) have fixed, predictable delay times between the primary and it subsequent multiples, which can be identified by temporal autoconvolution. Alternatively, it is generally observed that surface multiples will appear to travel at a slower velocity relative to other events that arrive at the same time, leading to a moveout-based filter.

In the past 30 years, significant advancements in surveying equipment and computational power has enabled the full treatment of seismic data as physical wavefields propagating in 3D space, rather than as lower-dimensional approximations. This has enabled seismologists to directly predict the propagation behaviour of recorded wavefronts after reflecting off the surface, and thus identify which one are due to surface multiples.

However, this approach approach of direct modelling raises a paradoxical situation: we are required to know accurately how seismic wavefields propagate, reflect, and refract in the subsurface, but this is the exact information we wish to establish in the first place with a clean, multiple-free seismic survey. To break this paradox of cyclical dependencies, we can either accept some uncertainty in the subsurface model, or in the propagation kernel itself. This choice respectively leads to the two most common approaches in the wave propagation method of multiple prediction:

- **Modelling by wavefield extrapolation:** requires knowing some velocity and interface information about the subsurface which allows surveyors to approximately model the raypaths of surface multiples using either phase space extrapolation or finite-differencing, and subsequently subtract it from the data. In actual usage, this method turned out to be sensitive to slight errors in the subsurface model (es-

pecially the exact reflectivity coefficients), and thus sees limited use in exploration surveys of entirely new areas.

- **Modelling by convolution with data:** exploits the fact that modern acquired seismic wavefields are well-sampled enough to act as approximations to the true kernels (Green's functions) that propagate primaries to their corresponding surface multiples. In general, you trade accuracy of the kernel (i.e., not knowing the source signature, sampling issues, noise, itself containing surface multiples) for not having to know the subsurface.

That latter class of methods became known under the collective term "data-driven methods", and also by the name of its most popular implementation: Surface-Related Multiple Elimination (SRME). However, due to the inherent limitations of using the acquired wavefield as a surrogate for the true Green's function of the subsurface, its formulation necessitates a non-linear adaptive subtraction process to map the inevitably erroneous multiple predictions to the true multiples in the data. This process is often regarded as sensitive to practician tweaking, and can often fluctuate between over-subtraction (wrongly removing primary events of interest) and outright failing to appreciably remove multiples. Although the flexibility and practicality of these data driven methods has led to its widespread adoption, this necessity of the adaptive subtraction paradigm places a ceiling on its theoretical performance. Any cursory literary search today will discover a longstanding body of work attempting to improve the accuracy of adaptive subtraction using a variety of metrics and assumptions on the primary wavefield.

It later was noted that perhaps striving for a much more accurate multiple prediction in the first place, at the cost of needing to solve a full-wavefield inversion problem, can eliminate the need of adaptive subtraction and its many shortcomings altogether. To this end, a proposal to directly solve for the discretized version of the Green's function was proposed by van Groenestijn and Verschuur (2009a), in a method called Estimation of Primaries by Sparse Inversion (EPSI). The key insight of EPSI is that a sparsest solution criteria (in the discretized space-time domain) is enough to uniquely solve for the discretized multiple-free Green's function, without any other input or assumptions other than the acquired data. This Green's function can in turn be directly used as the primary wavefield (after applying an estimate of the source function), or be used to produce a multiple model that is correct

in phase, amplitude, and kinematics, thus obviating the need for adaptive subtraction.

Since its appearance in 2009, EPSI has attracted significant attention from contemporary seismic surveyors seeking results beyond the the limitations to SRME and the prediction-subtraction paradigm. It is still today considered a relatively novel method, and many of its aspects as presented in current literature, especially algorithmically, remain very limited. While the original authors recognized the importance of a minimum sparsity criteria, EPSI still relies on an ad-hoc gradient hard thresholding regularization that depends on parameter tweaking to obtain acceptable solutions. Furthermore, some very interesting possibilities to mitigating incomplete wavefield sampling (such as near-offset gaps) which are enabled by the inversion-based EPSI paradigm remain unexplored. EPSI is also regarded as a computationally expensive method, since it requires many iterations of multiple predictions as function evaluations in an optimization problem, which limits its adaptation as SRME is already regarded as one of the most time-consuming processing methods. My work in the following chapters will aim to address most of these shortcomings.

## 1.1 Main purpose of this work

The primary objective of this thesis is to *propose new formulations and algorithms to apply the nascent EPSI methodology in ways that significantly improve its robustness and practicality* when applied to field data, using insights and frameworks from formal optimization theory and numerical linear algebra. The end result of this work is a method that will provide automated, higher quality surface multiple removal with less practitioner input when compared to current established data-driven methods. In this thesis I will generally refer to my method as Robust EPSI (REPSI) to differentiate it from the original form proposed in van Groenestijn and Verschuur (2009a).

Improvements made in this thesis will especially focus on the issues raised at the end of the previous section. In the extended introduction (chapter 2), I perform a literature survey on the existing field of multiple removal as described here, as well as introduce the mathematical notation necessary for the rest of this thesis. Following that in the subsequent chapters, I will develop and demonstrate my contributions to EPSI, as enumerated and summarized as follows:

### 1.1.1 Summary of contributions

In this thesis, I present the following novel contributions to both EPSI and the study of surface multiple removal made during my PhD studies:

1. Formulated a primary estimation method, based on the forward model used in EPSI, which relies on sparsity induced by formal $\ell_1$-norm minimization, instead of by ad-hoc hard-thresholding of gradient updates (Chapter 3).

2. Demonstrated a way to adapt a convex basis pursuit algorithm based on Pareto root-finding continuations of a series of norm-constraint problems solved by spectral projected gradient methods (SPG$\ell_1$) to the bi-convex EPSI problem above, solving for two unknowns: a wavefield with impulsive and sparse events in time, and a source-side wavelet that is similar to a unknown blur kernel in blind deconvolution (Chapter 3).

3. Using the above two contributions, I proposed a new algorithm for primary estimation called Robust EPSI or REPSI, and demonstrated that it is more effective and easier to apply for surface multiple removal compared to the original EPSI formulation, using both synthetic and real seismic line data (Chapter 3).

4. Augmented the Robust EPSI problem with a new forward modelling operator that accounts for errors in surface multiple prediction caused by an incomplete sampling of the wavefield at the surface (such as the near-offset gap). This is accomplished using a truncated Born-scattering model of the free surface (Chapter 4).

5. Proposed two different variations of the original Robust EPSI algorithm that is shown to be effective at solving the new augmented formulation above, and demonstrated its efficacy on both synthetic and real seismic line data with near-offset gaps for multiple removal (Chapter 4).

6. Presented a strategy to significantly decrease the computation time required to solve the Robust EPSI problem inspired by existing multi-level (and multigrid) methods. This is made possible by exploiting the blind deconvolution aspect of Robust EPSI, as well as the high-order numerical complexity of the evaluation and gradient computation step

of EPSI as a function of the spatial sampling of the wavefield (Chapter 5).

# Chapter 2

# Principles of surface multiple prediction by convolution

This chapter establishes the theoretical background and context necessary for the thesis. The main body of this work builds upon wavefield propagation methods for multiple prediction, which are based on well-known physical behaviour of waves, but in practice facilitated by seismic survey designs that are only becoming feasible and widespread in the past few decades. A brief overview of how subsurface structure information are relayed by active-source seismic responses is included for the benefit of readers without any background in seismic surveying.

I will go into a more detailed literature survey on the early developments of multiple removal techniques alluded to in the previous chapter, followed by a description of more contemporary modelling-based methods that are enabled by recent acquisition designs, expressing them using mathematical integral relations. This discussion culminates in the wave equation-based multiple modelling techniques which make up the physical basis of the EPSI method, and touches on some of its interesting relationships with the blind deconvolution problem. Finally, we look at the shortcomings of the current EPSI methods, from which the subsequent chapters in this thesis are motivated and outlined.

## 2.1 Exploration seismology

The seismic reflectivity time-series is the theoretical response of the subsurface to a perfectly impulsive seismic source of known amplitude. Figure 2.1 depicts the relationship between subsurface structure, reflectivity series, and seismic data.

When seismic waves meet an interface between subsurface materials with differing elastic properties, part of the energy carried by the waves will be reflected and eventually propagated back to the surface. The fraction of energy reflected at each interface depends on the reflection coefficient (or "reflectivity") of the interface, which in turn is determined by the difference in acoustic impedance of the two materials. Both the strength of this reflected signal and its arrival time can be used to further induce information of the subsurface geology.

Typically the subsurface interfaces are assumed to be of zero-order, which results in an impulsive reflection for an impulsive incident event. The arrival timing of these reflections to the surface is a thus function of both the subsurface wave velocity and the location of the interfaces. This relationship is exploited in subsequent imaging steps to produce a structural depiction of the subsurface.

Seismic data processing is the practice of preparing physically acquired seismic data in the field for this type of sensitive inference task. Obtaining a noise-free, amplitude-correct seismic reflectivity time-series is regarded as the goal of seismic processing. Hindering this goal are many practical factors, but the ones most pertinent to this thesis include the presence of both surface-related and internal multiples, as well as band-limited signature of the artificial seismic sources. These are both factors that obfuscate the ideal reflectivity series underlying a seismic recording and makes accurate seismic surveying an inherently difficult task, even in the absence of outside noise.

Due to physical limitations in energy output, artificial seismic sources used in exploration cannot be perfectly impulsive. Instead they typically have a narrow frequency bandwidth determined by the type of equipment used, typically bounded above at 80 Hz and below by 10 Hz, taking on a wavelet-like appearance in time. The recorded reflectivity series from such a source will thus be distorted by the width of the signature wavelet and its amplitude. This lowers the resolving power for thin structures, and obscures the minute differences between reflected and refracted events. In Figure 2.1,

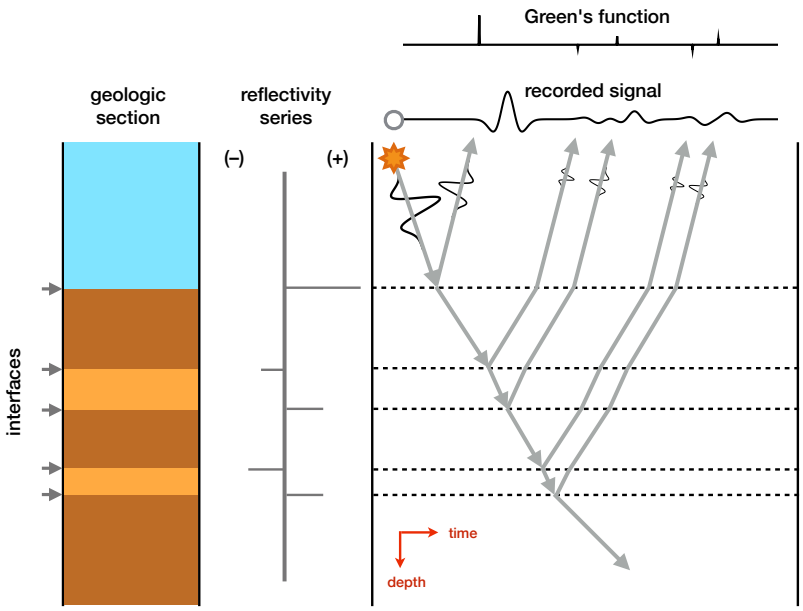



**Figure 2.1** Example of reflectivity series mechanism. Left sections shows six distinct geological layers (including a water column) with five distinct interfaces. Middle section shows the reflectivity coefficient induced by contrast in acoustic impedance between the layers. Right section shows how a single-receiver "trace" of seismogram is produced under an external source with a particular signature. If the source is perfectly impulsive, the trace produced would be the Green's function, with the amplitude of the reflections as a function of the reflection coefficient and the length of the raypath (due to spherical divergence). For a particular finite-energy source signal, the recorded trace is the causal convolution of the down-going source function and the Green's function. This is known as the *convolutional model*.

the thin geological layers between interfaces 2 and 3, as well as 4 and 5, is obscured in the recorded seismogram due to a source signature that has a relatively wide wavelet envelope.

### 2.1.1 The recorded wavefield

Seismic wavefields recorded at the Earth's surface are typically represented as a scalar field (for pressure recording, typically in marine) or vector field

(for particle velocity recording, typically on land) of a space described by two coordinates: a receiver location $\mathbf{x}$ and time $t$ pair. If all receivers lie on a line, in which case the survey geometry is called a *2D* (or *seismic line*) survey, the receiver coordinate may be a single scalar parameterized along the line. In the more general case (typically called a *3D* survey), the receiver coordinate is a 2-vector $\mathbf{x}$ of lateral and vertical coordinates spanning a surface patch. For active seismic surveys, the wavefield is further parameterized by $\mathbf{x}_{\text{src}}$, the location(s) of the active point source responsible for the recorded wavefield, typically in the same coordinate system as the receiver locations.

Thus the recorded seismic wavefield can be written as a function $p(\mathbf{x}, t; \mathbf{x}_{\text{src}})$ of all the dependencies described above. For the purpose of this thesis, we will only deal with scalar pressure fields (thus the use of the letter $p$), although the results are equally valid for vector fields (Fokkema and van den Berg, 1993). Sometimes I will also need to talk about the wavefield under a Fourier transform in the time direction. For brevity, I will not use a separate symbol for the Fourier transform of the physical wavefield, but would simply indicate it by indexing the time coordinate with a suitable symbol (such as $\omega$) denoting frequency, for example writing the above recorded seismic wavefield as $p(\mathbf{x}, \omega; \mathbf{x}_{\text{src}})$.

### 2.1.2 Surface multiples

As discussed in the Introduction, surface multiples are essentially "echoes" of seismic energy off the measurement surface. Much like auditory echoes, they appear in seismic recordings as copies of the primary reflection signal at regular time delays, and are hard to distinguish from wavefronts traveling at a slower velocity than that of the true subsurface. A strong multiple arriving simultaneously with a weak primary reflection introduces difficult ambiguities in velocity tomography and coherent artifacts in structural imaging. Figure 2.3 depicts surface multiples and their impact on seismic recordings.

## 2.2 A brief history of surface-multiple removal

Multiples are by nature coherent wavefronts, which are difficult to remove with rudimentary noise rejection methods utilizing trace averaging and stacking—techniques that otherwise work well with incoherent noise sources.

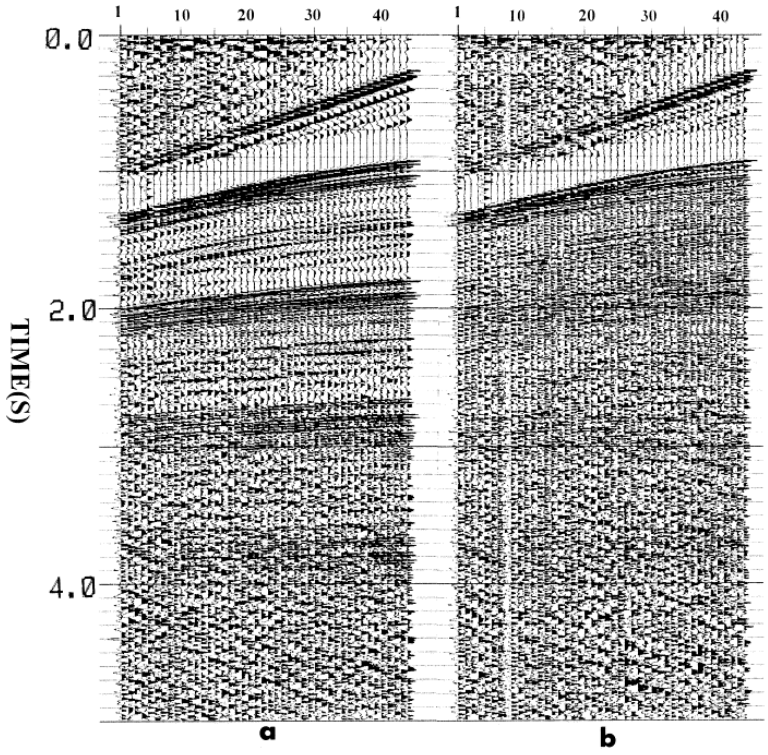



**Figure 2.2** Seismic data **(a)** before and **(b)** after surface multiple suppression, showing the strong, periodic, and coherent appearance of surface multiple events in seismic data. In this case the surface multiple was reverberating within the water layer in a marine acquisition. Adapted from *Character of active faulting in the North Aegean Sea*, Saatcilara et al, 1999.

Seismic surveyors thus typically rely on sophisticated rejection approaches that are developed specifically for this purpose. These methods are generally classified into two broad categories of filtering-based and prediction-based. Furthermore, filtering is separated into predictive deconvolution and move-out filtering. This taxonomy can be found in many survey texts on surface multiple removal methods (Weglein, 1999; Verschuur, 2006).

### 2.2.1 Predictive deconvolution filtering

Predictive deconvolution assumes that certain periodicity in zero-offset (or NMO-stacked) data are due to surface multiples. Filters are designed to re-

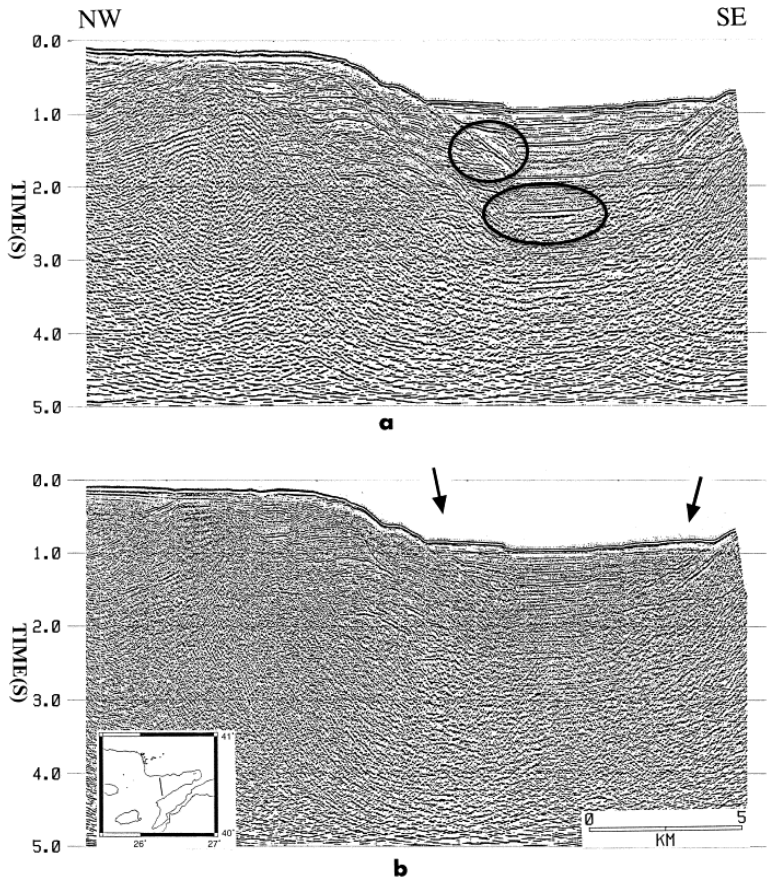



**Figure 2.3** Stacked seismic line **(a)** before and **(b)** after surface multiple suppression. Multiples appear as false, spurious images of non-existent structure. Some examples are encircled in black. Adapted from *Character of active faulting in the North Aegean Sea*, Saatcilara et al, 1999.

move these long-period oscillations in time. The concept for this method can be traced back to Robinson (1954) with additional development by Backus (1959) and Peacock and Treitel (1969).

In its nascent days, active seismic surveys were mostly single-channel in nature. Many processing methods are thus formulated solely on traveltime information contained in the individual trace records, which means that only temporal periodicity is considered here in the simplest case. This leads to one of the biggest advantages of predictive deconvolution, since trace-by-trace processing is all that could be achieved with the equipment

of that time. However, it is also its biggest shortcoming, as the assumptions behind this method is strictly only valid for the zero-offset trace in purely vertically stratified subsurface medium, which means that fundamental approximations to the underlying physics must be made for practically all field surveys.

Due to its simple nature, the predictive deconvolution method can still be found today in practice as a quick way to approximate surface multiple-free data, especially in very shallow water marine acquisition (less than 300 m to the water bottom). It has also increased in sophistication to account for geologically consistent lateral variations in the reverberation period, such as sloping seabeds, and in some instances full lateral heterogeneity in the underlying structure (Taner et al., 1995; Lokshtanov, 1999). Contemporary applications of the predictive deconvolution filter are usually applied in the linear Radon domain (Durrani, 1991) or in the radial direction of the $x$–$t$ domain (Taner, 1980).

### 2.2.2 Moveout filtering

As recording equipments increase in sophistication and surveys began to become increasingly multi-channel in nature (which is around the 1980s), it became possible to discriminate primaries and surface multiple events by its apparent velocity. This is based on an inherently multichannel phenomenon where the subsurface velocity influences the curvature behaviour of arriving wavefronts as a function of surface offset from the shot position.

If a primary and a surface multiple wavefront appears at the same time, the multiple is assumed to have travelled through a slower medium. This assumption is valid as long as wave velocity always increases with depth, which is a commonly held simplification due to increased pressure. The multiple is assumed to have bounced around in a shallower area while a primary that appeared simultaneously ostensibly has to travel through a deeper part of the subsurface to match the arrival time of the multiples.

This was first noticed and exploited as a practical algorithm in the F-K (frequency-wavenumber) domain (Ryu, 1982). It can also be accomplished in the Radon domain (Hampson, 1986; Kelamis et al., 1990; Herrmann et al., 2000), as well as the velocity-stack domain (Thorson and Claerbout, 1985; Yilmaz, 1989; Foster and Mosher, 1992; Sacchi and Ulrych, 1995).

Despite being regarded as more sophisticated than the predictive deconvolution method, there are also many cases where the assumptions behind moveout filtering becomes invalid. In general, this method requires that that subsurface is a strictly increasing function with depth, which is easily invalidated in areas with any geological phenomena where pockets of excessively high or low velocities form, such as salt domes, chalk formations, and gas clouds. Even thrusting faults can easily invalidate this assumption. To get around this limitation, a more accurate physical model is needed accurately identify surface multiples.

### 2.2.3 Beyond filtering: prediction-subtraction paradigms

The above mentioned filtering methods are often satisfactorily adept at picking out the kinematic temporal and spatial locations of the surface multiples (although not entirely error-free). Comparatively, determining the exact amplitude and waveform of multiples is much harder than kinematically determining its location. In general, it requires exact knowledge of the source signature and the bulk properties of the propagation medium, which paradoxically was what we are trying to infer accurately by doing multiple removal. This shortcoming is a real problem because primary and multiple signals often overlap, both in terms of kinematic characteristics and also physically in the seismic record. The only certain way to cleanly suppress multiple event without harming primary events is to also accurately know the amplitude and waveform of the multiples.

This realization came at a time where there is a general trend in the seismic exploration community to utilize more and more wave-equation based methods, driven by survey designs that spatially sample the seismic wavefield at a spacing fine enough, such that there is no severe aliasing in the receiver direction. This resulted in a fundamental shift from simply filtering out multiples to a paradigm of trying to accurately model the multiples using wave-equation principles, and subsequently subtract it from the data.

In the section below I will first introduce some mathematical notion involved to describe these methods, which will be used all throughout the thesis, before doing a survey of the methods specific for multiple removal. The survey will culminate in the method that most famously represent this paradigm: Surface-Related Multiple Elimination (SRME).

## 2.3 Background on wavefield propagation multiple prediction methods

### 2.3.1 Multiple prediction through wavefield convolution

Mathematically, the operation of wavefield convolution is defined such that its physical interpretation is the concatenation of the propagating raypath of wavefields. When two wavefields are convolved correctly, the result is that each point in the first wavefield will act as a local Huygens-type source function, which is subsequently propagated through the path traveled by the second wavefield. The aggregate of this action over all of the physical domain spanned by the first wavefield is the result of the wavefield convolution. Figure 2.4 illustrates this interpretation for a simplified 1D cas.

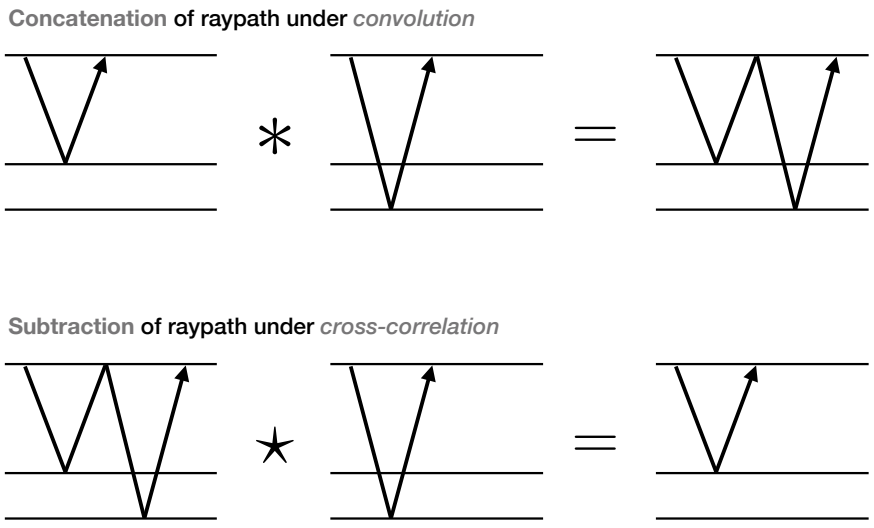



**Figure 2.4** Wavefield convolution concatenates the two propagation raypaths, while its adjoint operation, cross-correlation, subtracts raypaths.

Convolution in the time domain is homomorphic to complex number multiplication in the Fourier domain. For two wavefields that are observed over a coincident plane (or a line in 2D medium) that is large enough, and whose source functions lie on the same plane as described above, the convo-

lution of these two wavefields can be represented as a surface/line integral. To produce a model of the surface multiples through convolution, we need to assume the existence of a wavefield kernel $g(\mathbf{x}, \omega; \mathbf{x}_{\text{src}})$ for all frequencies $\omega$ that, under this surface integral representing wavefield propagation, produces all the multiples $p_m$ generated by that surface from a noiseless pressure wavefield $p(\mathbf{x}, \omega; \mathbf{x}_{\text{src}})$. The physical significance of $g$ will be discussed below, for now we simply require it to satisfy

$$p_m(\mathbf{x}, \omega; \mathbf{x}_{\text{src}}) = \int_S g(\mathbf{x}, \omega; \mathbf{x}') r(\mathbf{x}, \mathbf{x}') p(\mathbf{x}', \omega; \mathbf{x}_{\text{src}}) \, \mathrm{d}\mathbf{x}'. \tag{2.1}$$

Throughout this thesis, our integration surface $S$ is implicitly assumed to be both the multiple-generating free surface and the observation surface on which $\mathbf{x}'$ lies. This assumption can be relaxed by careful treatment of the wavefield directionality of $p$ and $g$ based on where we choose to position $S$ in relation to both surfaces (see Frijlink et al., 2011, for more details). The operator $r$ models the reflection of the free surface, and is typically assumed to be isotropic and well approximated by a Dirac delta function $\delta(\mathbf{x} - \mathbf{x}')$ scaled by a reflection coefficient of $-1$, although this can again possibly be relaxed and estimated (AlMatar, 2010).

The surface integration (2.1) with $g$ effectively sends all wave paths in $p$ *exactly* one more time through the Earth's subsurface. This transforms all detected primary events to the first-order surface multiples, as well as all later $n$-th ordered multiples to the $n{+}1$ order (Anstey and Newman, 1966; Riley and Claerbout, 1976), as illustrated in Figure 2.5. In order for equation 2.1 to only produce multiple events, both $p$ and $g$ should be free of direct arrivals, as they would map to primary events.

The kernel $g$ that satisfies this property can be physically interpreted as the normal derivative (at surface $S$) of a particular wave equation's associated Green's function. The propagation medium of this wave equation has no multiple-generating discontinuity at $S$, but is otherwise identical to the Earth's subsurface which produced $p$. Thus, $g$ can be seen as the primary wavefield $p_o$ component of $p$ that is also deconvolved from the source wavelet $q$, with the relation $p_o(\mathbf{x}, \omega; \mathbf{x}_{\text{src}}) = g(\mathbf{x}, \omega; \mathbf{x}_{\text{src}}) q(\omega)$. To keep this thesis manageable in scope, I assume that all independent seismic experiments at each source location have roughly the same source function in time. I also refer to $g$ as the Green's function instead of its normal derivative, and simply assume that all surface integrals involving it include a well-known obliquity factor.

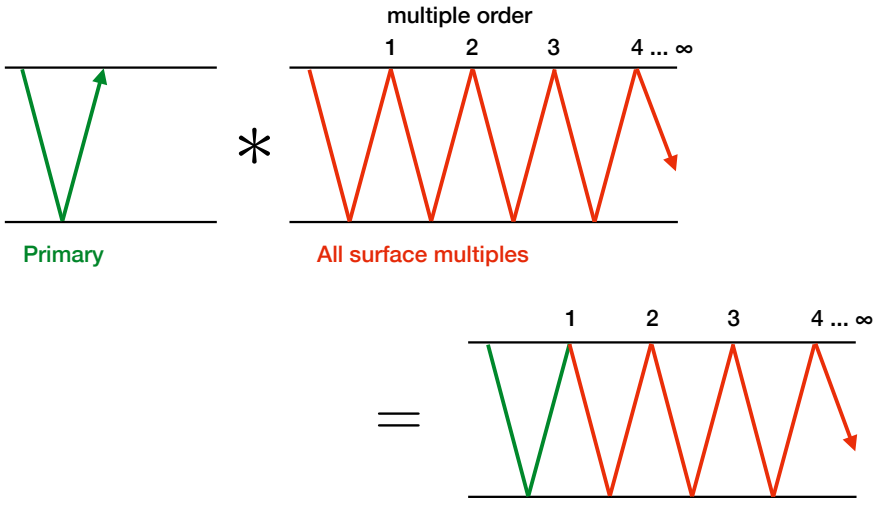



**Figure 2.5** Convolving a wavefield containing surface multiples of different order (up to infinity) with the primary raypath essentially produces "shifts" all orders of surface multiples up by one (i.e., 1st order multiples becomes 2nd order, etc). Primary wavefields can essentially be viewed in this way as a "zeroth order" surface multiple. Convolution of the observed data with the primary raypath therefore produces all surface multiples of order 1 and above, and leaves no primaries. This effect forms the principle behind multiple prediction by wavefield propagation.

### 2.3.2 Propagation based on estimated subsurface models

We see that in order to completely generate all the surface multiples correctly, we need to know the raypath of the primary wavefield. However, this kind of detailed subsurface information is exactly what we have hoped to probe with the seismic survey in the first place! If one only wished to remove the shallow-layer reverberations (such as in the water column), a suitable approximation of the shallow layer can in some cases be estimated well enough to implicitly generate a suitable primary propagation kernel $g$ (typically by solving the wave equation for the estimated medium). This method is typically known as "multiple prediction by wavefield extrapolation" (Bernth and Sonneland, 1983; Berryhill and Kim, 1986; Levin, 1987; Wiggins, 1988; Lokshtanov, 2000).

Anecdotally, the wavefield extrapolation method can be very difficult to attain a high degree of accuracy for many survey fields with complex near-surface geometry; often it is problematically sensitive to even slight errors in the subsurface model. Furthermore, just modelling the shallow raypaths would not be satisfactory for projects that need to see very deep into the Earth. In contemporary usage, unless in very shallow marine acquisitions, the wavefield extrapolation method is largely supplanted in popularity by what are known as "data-driven methods" described below, which are much more straightforward when being applied to arbitrary surveys.

### 2.3.3 Propagation based on recorded data

Instead of explicitly performing wavefield extrapolation by constructing a model and solving the wave equation, we can use the data itself as an approximation of the Green's function of the wave equation. This is valid if the reflecting surface is co-located with the detector surface, which is generally true for surface multiples. The general class of methods that uses the data itself as the multiple-generation wavefield propagator is generally known as "data-driven methods".

This scheme can first be found in Anstey and Newman (1966), which noted that the autoconvolution of a seismic traces containing only primaries transforms it into a trace containing only first order surface multiples. A complete model for surface multiples based on this principle is described by Kennett (1979) for 1D and Riley and Claerbout (1976) for 2D.

It was then reported in Fokkema and Van den Berg (1990; and further established in Fokkema and van den Berg, 1993) that this phenomenon, described by wavefield propagation at a surface, can be formally derived from a reciprocity theorem of the wave equation from its values and normal derivatives at a boundary. This was further generalized for different surface, source, and receiver in van Borselen et al. (1996) and Frijlink et al. (2011).

Another way to look at the data-driven multiple prediction through convolution is by considering scattering from the free surface. Since the ground-air interface can be generally regarded as a perfect scatterer, it is also possible to derive this relationship from scattering theory that relates an incident field to a scattered field. This view was most comprehensively developed in Weglein et al. (1997), and is now commonly known as the Inverse Scattering Series (ISS) method (see Weglein et al., 2003, for a more complete survey). Eventually, this method went on to be popular in internal multiple removal.

## 2.4 Surface-related multiple elimination

Surface-Related Multiple Elimination (SRME) is perhaps of the most commonly used variant out of all the data-driven methods described in the last section. In our formalism, data-driven methods refer to cases where the surface-free Green's function $g$ is approximated with the observed wavefield $p$. Although this absolves us of having to know the subsurface medium, it also leads to an incorrect model of the surface multiples in terms of amplitude and spectral content. This is due to the fact that, unlike the surface-free Green's function, the observed data has in addition been convolved with the source signature, and also includes all orders of surface multiples. This leads to the multiple prediction having the wrong amplitude, wavefront shape (as dictated by the source signature), and over-estimated surface multiples.

It was then realized that the wavelet and amplitude mismatch could be corrected by using a minimum energy assumption on the primary wavefield (Berkhout and Pao, 1982; Verschuur, 1991). Using this criteria one can treat the *subtraction* step as an inverse problem for the primary wavefield. This adaptive subtraction step, combined with the data-driven multiple prediction, became collectively known as the method of Surface-Related Multiple Elimination.

### 2.4.1 Evolution of adaptive subtraction methods for SRME

The aforementioned minimum energy assumption suggests solving a simple least $\ell_2$-norm solution for the primary wavefield to obtain a unique solution from the subtraction. It was realized early on that this is a very aggressive approach that can lead to *over-subtraction* of the multiples. In Berkhout and Verschuur (1997) it was realized that this scheme can be further iterated to improve the accuracy of the wavelet, gradually leading to a more accurate multiple model. However, this still relied on minimum-energy assumption of the primaries. More sophisticated adaptive subtraction method appeared almost immediately.

Some authors advocated simply measuring the acquisition (wavelet) more directly with a careful measurement of the direct arrival, which is typically muted or discarded since only subsurface reflections and refractions are of interest for exploration projects (Ziolkowski et al., 1998, 1999). This enables a deconvolution to be deterministically built into the adaptive subtraction and can improve the reliability of the solution somewhat.

Others advocated an improved selection of priors for the residue wavefield (recorded data minus the multiple model) with what can be viewed as Bayesian-based approaches to adaptive subtraction. One of the earliest example is Spitz (1999) who introduced a scheme based on pattern recognition using spatial predictability of the primary and multiple reflections.

Another successful class of methods uses the sparsity instead of the minimum energy criteria for the primary wavefield (Guitton and Verschuur, 2004), usually under some sparsifying wavefield representations such as Curvelets (Herrmann et al., 2007; Wang et al., 2008; Herrmann et al., 2008; Neelamani et al., 2008). Relatedly, more sophisticated methods were also recently developed using the independent component analysis (ICA) by Kaplan and Innanen (2008). Several authors have built upon this framework (Lu and Liu, 2009; Donno, 2011) including a case that exploits the structure of the wave equation in 3D by subtracting entire seismic volumes at once (Li and Lu, 2013).

### 2.4.2 3D SRME

It should first be noted that the mathematical formalism above applies fully to a 3D acquisition geometry. However, it proved very challenging to sample a surface at a fine enough grid to satisfy aliasing-free requirements for these wavefield propagation methods for multiple modelling. At least one spatial direction will be much more poorly sampled, due to the overwhelming majority of seismic detectors (geophones and hydrophones) being embedded into cables. Despite the sampling challenges involved, many attempts have been made to make SRME practical for 3D data, to address fundamental inaccuracies of the 2D method (Matson and Corrigan, 2000). Many ways are suggested to overcome the sampling deficiencies that is typical of 3D surveys in the cross-line direction. (Nekut, 1998; Biersteker, 2001; van Dedem and Verschuur, 2005; Donno et al., 2010). A good survey of the current situation in 3D is Dragoset et al. (2010).

As a result of the above developments, and due to the general ease of applying the method, SRME today enjoys widespread commercial success, despite its inherent limitation of relying on sophisticated subtraction steps. As of not long ago from the time of writing, it is still recognized in the seismic surveying industry as the most accurate surface multiple removal method for most survey projects. (Verschuur, 1992; Verschuur and Berkhout, 1997; Dragoset and Jeričević, 1998; Verschuur and Prein, 1999; Hadidi et al., 1999).

## 2.5 Estimation of primaries by sparse inversion

In Guitton and Verschuur (2004) and van Groenestijn and Verschuur (2008), it was shown that for adaptive subtraction the correct result is often yielded by the minimizer of a sparsity measure on the primary wavefield. This observation led to the realization that perhaps directly inverting for a deconvolved version of the primary wavefield could be an alternative to improve upon some of the fundamental limitations of SRME. A complete method was eventually introduced in van Groenestijn and Verschuur (2009a) as a procedure called the Estimation of Primaries by Sparse Inversion (EPSI).

The objective of EPSI is to directly solve for a wavefield kernel $g$ that can simultaneously explain both the multiples and the primaries: the multiples through expression 2.1, as well as the primaries using a jointly-estimated wavelet. This fundamental difference was shown to be enough to address deficiencies in the acquisition of $p$ by augmenting the functionals involved in EPSI's variational problem in various ways.

The idea of directly estimating the primary wavefield is not new. Several authors have previously explored ways to estimate either the surface-free Green's function or the primary wavefield directly (van Borselen et al., 1996; Biersteker, 2001; Wang, 2004). The most famous of these is introduced in Amundsen (2001), which instead of having to also estimate the wavelet, instead relies on using the recorded direct wave as proposed by Ziolkowski et al. (1999). However, as of the most recent attempts reported in Majdański et al. (2011), the direct wave still needs to be reconstructed synthetically from measurements.

Following equation 2.1, we have an expression that maps $g$ to both the primary and the surface-related multiple events in the total wavefield $p$. I represent this model as a time-domain operator $M(g, q; p)$ that is a function of $g$ and the source wavelet $q$, parameterized by the data:

$$M(g,q;p) := \mathcal{F}_\omega^{-1}\big[g(\mathbf{x},\omega;\mathbf{x}_{\text{src}})q(\omega) + \int_S g(\mathbf{x},\omega;\mathbf{x}')r(\mathbf{x},\mathbf{x}')p(\mathbf{x}',\omega;\mathbf{x}_{\text{src}})\,\mathrm{d}\mathbf{x}'\big] \tag{2.2}$$

$$\begin{aligned} &= p_o(\mathbf{x},t;\mathbf{x}_{\text{src}}) + p_m(\mathbf{x},t;\mathbf{x}_{\text{src}}) \\ &= p(\mathbf{x},t;\mathbf{x}_{\text{src}}), \end{aligned}$$

where $\mathcal{F}_\omega^{-1}$ represents the continuous inverse Fourier transform. As Figure 2.6 shows in 1D models, knowing the true Green's function greatly improves the accuracy of the modelled multiples.

### 2.5.1 Discretized notation

In practice, EPSI follows the discretize-then-optimize paradigm, which attempts to solve directly for discretized representations $\mathbf{g}$ and $\mathbf{q}$ (discretized entities will be denoted by bold font) using numeric algorithms. We are thus only concerned with the discretized version of expression 2.2. Assuming a regularized sampling of source and receiver coordinates, the surface integral in expression 2.1 can also be written as simple matrix multiplications between monochromatic slices of data volumes, following the detail-hiding notation of Berkhout and Pao (1982). Throughout this thesis I will denote the mono-frequency matrix view of a wavefield by writing the symbol in upper-case letters (dropping the implied frequency dependence for brevity). For example, a mono-frequency data matrix for the wavefield $\mathbf{p}$ (of size $n_\mathrm{t} \times n_\mathrm{rcv} \times n_\mathrm{src}$) is written simply as $\mathbf{P}$ (of size $n_\mathrm{rcv} \times n_\mathrm{src}$). I adhere to the convention of row (column) indices of a data matrix corresponding to the discretized receiver (source) positions.

The discretized forward model for EPSI (equation 2.2) is thus

$$\begin{aligned} M(\mathbf{g}, \mathbf{q}; \mathbf{p}) &= \mathbf{F}_\omega^{-1}\big[M_\omega(\mathbf{G}, \mathbf{Q}; \mathbf{P})\big], \\ M_\omega(\mathbf{G}, \mathbf{Q}; \mathbf{P}) &:= \mathbf{GQ} + \mathbf{GRP}. \end{aligned} \tag{2.3}$$

The operator $\mathrm{F}_\omega^{-1}$ is the (padded) discrete inverse Fourier transform that prevents convolutional wraparound effects in the time domain. The term $\mathbf{GQ}$ produces the primary events, while the $\mathbf{GRP}$ term produces surface multiples. Note that the discretized notation allows for a more arbitrary source function compared to the continuous expression, but I reconcile the two by parameterizing $\mathbf{Q} = q(\omega)\mathbf{I}$. Similarly, the reflectivity operator is assumed isotropic with reflection coefficient $-1$ as above, by setting $\mathbf{R} = -\mathbf{I}$. For clarity, figure 2.7 shows a shot-gather ($\mathbf{x}_\mathrm{src}$ fixed at a particular point) for the various wavefields mentioned and produced in the above relation.

Simultaneously solving for $\mathbf{g}$ as well as the source wavelet for will in general admit non-unique solutions, analogous to nondeterministic deconvolution problems with an unknown blur kernel (I will discuss more on this topic in section 2.5.3). Thus, obtaining a useful solution requires additional

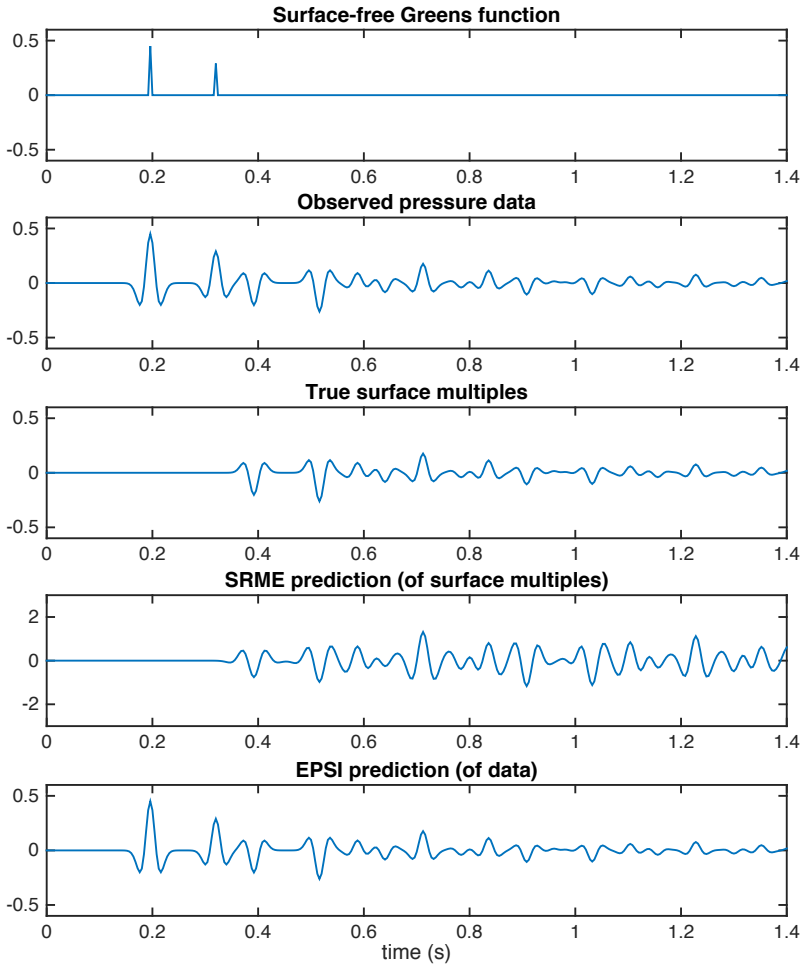



**Figure 2.6** Single trace (1D) surface multiple model of an underlying two-reflector event seismic signal. Convolving the observed pressure data with itself (SRME prediction) produces a wavefield that is similar to the true surface multiple but is erroneous in both the amplitude and the overall shape of the multiple events. The EPSI relation in turn is able to explain the whole data accurately by utilizing the Green's function.

regularization on the unknowns. The insights in van Groenestijn and Verschuur (2008) suggests that this regularization should be minimum sparsity on $\mathbf{g}$. This is also justified by the fact that a discretized representation of the Green's function $\mathbf{g}$ (in 3-D) resembles a wavefield with impulsive or "sparse" wavefronts. Therefore, the EPSI algorithm introduced in this thesis proposes to find the *sparsest* possible $\mathbf{g}$ (in the time domain) through $\ell_0$-norm minimization. Specifically, it solves the following optimization problem:

$$\begin{aligned} \min_{\mathbf{g},\,\mathbf{q}} \|\mathbf{g}\|_0 \quad \text{subject to} \quad f(\mathbf{g},\mathbf{q};\mathbf{p}) \leq \sigma, \\ f(\mathbf{g},\mathbf{q};\mathbf{p}) := \|\mathbf{p} - M(\mathbf{g},\mathbf{q};\mathbf{p})\|_2, \end{aligned} \tag{2.4}$$

where $f(\mathbf{g},\mathbf{q};\mathbf{p})$ is the misfit functional, and the modelling operator $M(\mathbf{g},\mathbf{q};\mathbf{p})$ is as defined in expression 2.3. In the original EPSI formulation (van Groenestijn and Verschuur, 2009a), this minimum sparsity assumption on $\mathbf{g}$ is enforced by an ad-hoc hard threshold operation on gradient updates formed to minimize $f(\mathbf{g},\mathbf{q};\mathbf{p})$. This is a shortcoming that I will address in Chapter 3.

### 2.5.2 Improvements of EPSI over SRME

One of the main motivations behind the proposal of EPSI is that by eliminating the need for adaptive subtraction due to having a much more accurate multiple model, EPSI will produce much more accurate primaries in areas where primaries and multiples severely overlap. There are several reports that show fairly convincing examples on real data that this is indeed the case (van Groenestijn and Verschuur, 2009b; Savels et al., 2010; Baardman et al., 2010).

Another important advantage of EPSI is that it relies on a full iterative inversion/optimization process from a physical model that considers the whole observed data. This allows us to build various kinds of noise into the forward model, and can inherently reject additive noise if the practitioner can avoid overfitting. A related extension is the ability to mitigate missing data in the wavefield which fundamentally leads to erroneous multiple prediction, since the integration surface would effectively be incompletely sampled. This aspect was also successfully demonstrated in the above reports, although I will propose an alternative approach in Chapter 4.

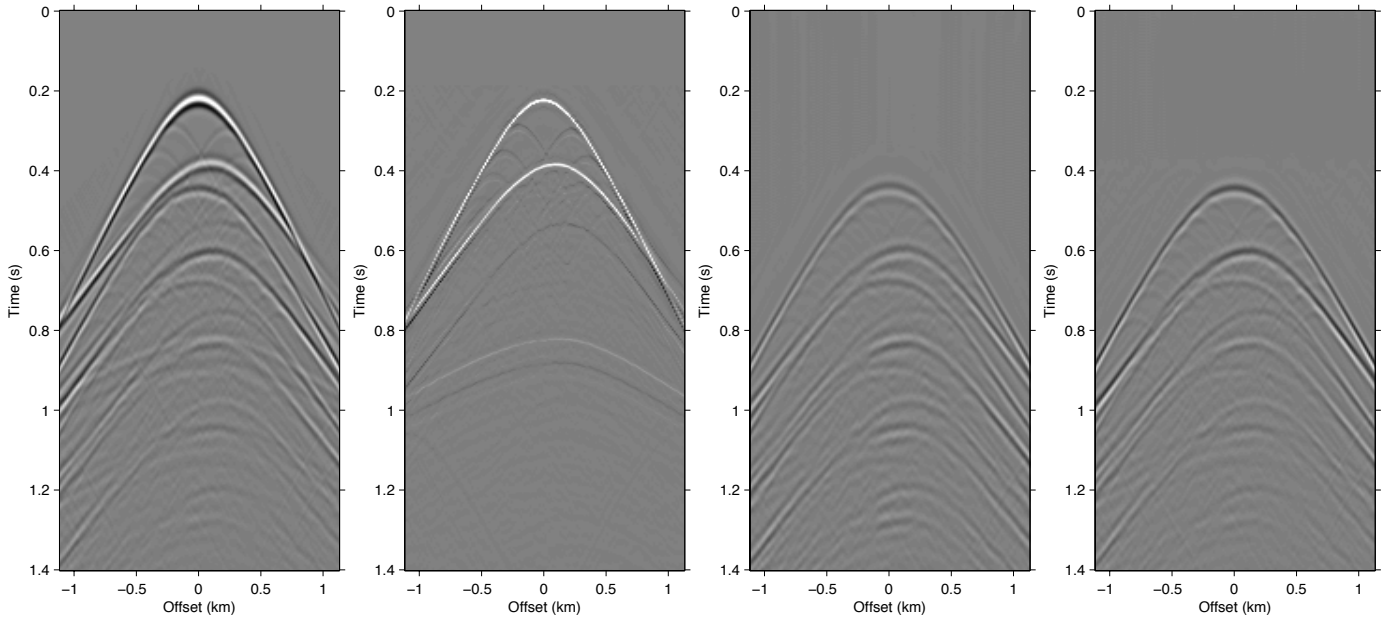



**Figure 2.7** Shot gathers of the various wavefields involved in this paper. **(a)** the complete pressure data **p**, **(b)** the true discrete surface-free Green's function **g**. Producing this is the primary objective of the EPSI problem. **(c)** the SRME multiple model, produced by the equivalent of a **PRP** term, using the observed data as an estimating of **g** the primary raypath propagator. The amplitudes and general shape of the multiples are erroneous when compared to the multiple events in the data. **(d)** the EPSI multiple model produced by the **GRP** term in expression 2.3, which is a much more accurate model of the surface multiple, and can be directly subtracted from the data with satisfactory results.

Finally, EPSI does not return primary wavefield itself, instead the solution is decomposed into the source wavelet and the underlying surface-free Green's function. Essentially, EPSI can be regarded as a method to achieve a physically-consistent (with the surface multiples) blind source signature deconvolution. This aspect has intriguing implications both to the algorithm choice (e.g., it enables a way to accelerate the compute time of EPSI which I introduce in Chapter 6), as well as to potential applications of the solution. In particular, many seismic inference procedures such as velocity tomography and imaging benefit greatly from having increased temporal resolution (wider spectral bandwidth) in the data, and having access to the primary wavefield deconvolved from the source wavelet greatly helps in that regard. In the next section I will briefly segue into discussing why this aspect of EPSI is interesting from a mathematical point of view.

### 2.5.3 The EPSI formulation and its relation to deconvolution

In this section we look at the EPSI problem as a wavefield deconvolution problem. Although we will not further explore this concept for the rest of the thesis (aside from taking advantage of it in chapter 5), it is nevertheless interesting to mention why the EPSI problem can be solved without too much assumption on the source signature $\mathbf{q}$.

Source signature deconvolution is the practice of inverting the convolution model for seismic signals. Specifically, it attempts to solve for $\mathbf{g}$ in the relation

$$\mathbf{p} = \mathbf{g} * \mathbf{q}, \tag{2.5}$$

where $\mathbf{p}$ is a trace of recorded seismogram, $\mathbf{q}$ is the source signature (typically unknown), and the symbol $*$ denotes convolution in the time domain. The deconvolution is performed over all possible source and receiver positions over the surveyed surface, and the aggregate solution we obtain is the Green's function of the subsurface seismic response.

The convolution operator with a typical seismic source signature has a large null space and is difficult to invert. In general, seismic deconvolution is an ill-posed problem, with the difficulty escalating greatly when the source signature is unknown (called *blind deconvolution*). One of the first significant milestones in this field is the Wiener-Levinson deconvolution developed in the 1950s (Robinson, 1957), which is essentially a regularized least-squares method for inverting a known convolution operator. Today this remains

the staple deconvolution method when the source signature is sufficiently well-known.

Most deconvolution problems in seismology are somewhat performed "blind", since it is often practically impossible to precisely determine the exact source function sent into the underground from man-made equipment. Blind deconvolution has a much bigger null space than a typical deconvolution problem in terms of both $\mathbf{g}$ and $\mathbf{q}$. As Figure 2.8 shows, the blind deconvolution model produces the exact same result under complimentary scaling and shifting. Furthermore, a limited bandwidth in $\mathbf{q}$ also means that it will generate the same output even when random high and low frequency noise is added to the model for $\mathbf{g}$.

The late 1970s to the early 1980s saw a big leap of effort in the development of blind source deconvolution using new estimation theory to obtain the source wavelet. The class of "minimum entropy" methods devised estimators that maximizes the non-Gaussianity of the deconvolved reflectivity series (Wiggins, 1978; Donoho, 1981; Ulrych and Walker, 1982; van der Baan and Pham, 2008; Cabrelli, 1985; Oldenburg, 1981). The idea is that convolution always increases the Gaussianity of the signal, and therefore the deconvolved result should be as non-Gaussian as possible. Kurtosis is typically used as the measure of Gaussianity, therefore methods of this type are also called "higher-order statistics" methods. The comparative name is meant to distinguish itself from existing second-order statistics methods, such as spectral whitening.

Even though multiple removal is not usually regarded as a deconvolution process, it comes naturally from the concept of modelling multiples via convolution of two wavefields as introduced in this section. Assuming as before that the surface reflection operator is approximated by $-\mathbf{I}$, the EPSI model (equation 2.3) of the recorded wavefield in the 1D case can be seen as a variation of the simple source signature deconvolution problem of equation 2.5:

$$\begin{aligned} \mathbf{p} &= \mathbf{g} * \mathbf{q} - \mathbf{g} * \mathbf{p} \\ &= \mathbf{g} * (\mathbf{q} - \mathbf{p}). \end{aligned} \tag{2.6}$$

Equation 2.6 essentially contains two convolution relationships: The first term $\mathbf{g} * \mathbf{q}$, with our assumption of a globally stationary source, is simply the source convolution relation stated in equation 2.5. The second term $\mathbf{g} * \mathbf{p}$ produces the multiples caused by the free surface. Intuitively, this means

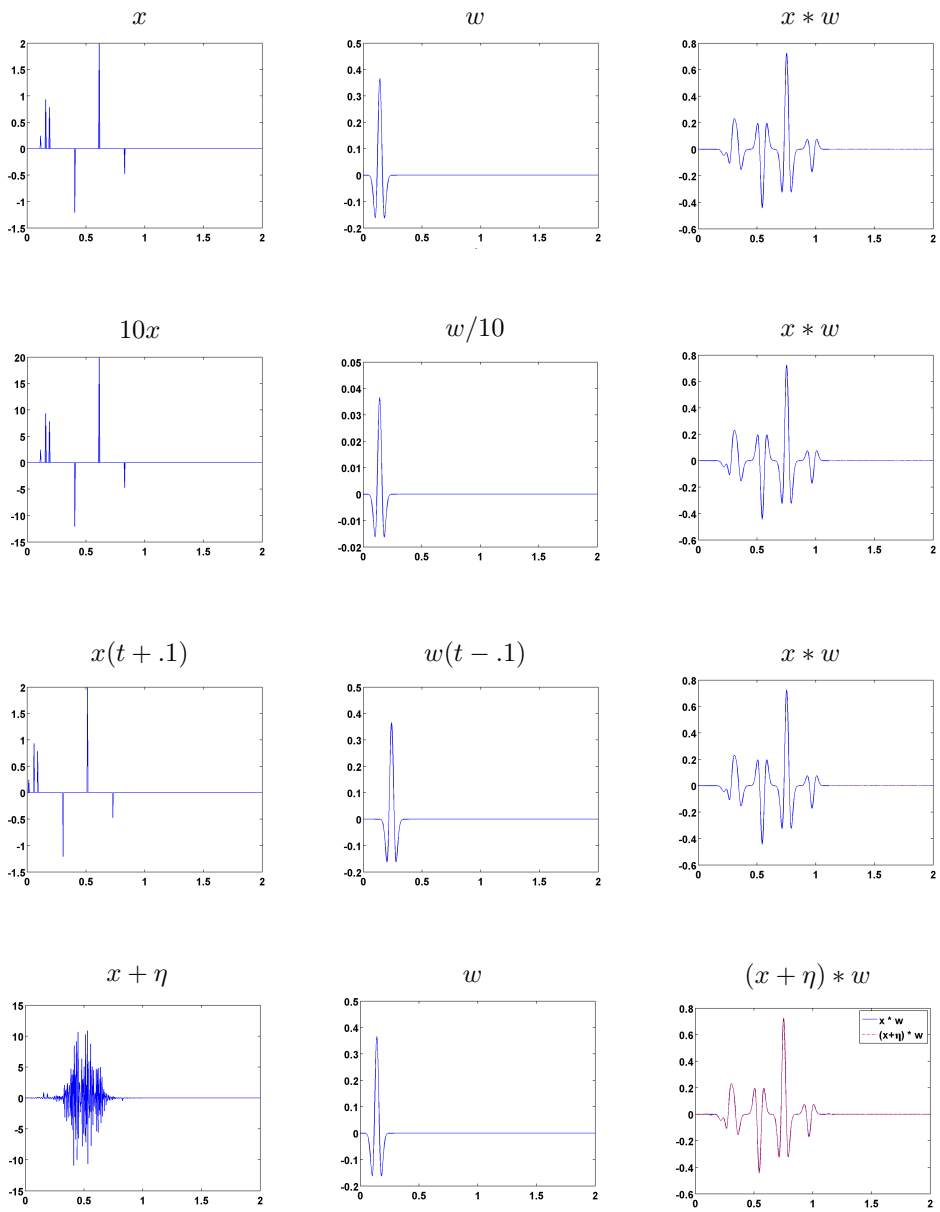



**Figure 2.8** Blind deconvolution is a difficult task. For a given observation (right column), an infinite number of signal $x$ and blur kernel $w$ combinations may reproduce it under a causal convolution relationship. Adapted from *A lifted $\ell_1/\ell_2$ constraint for sparse blind deconvolution*, Esser et. al., 2015

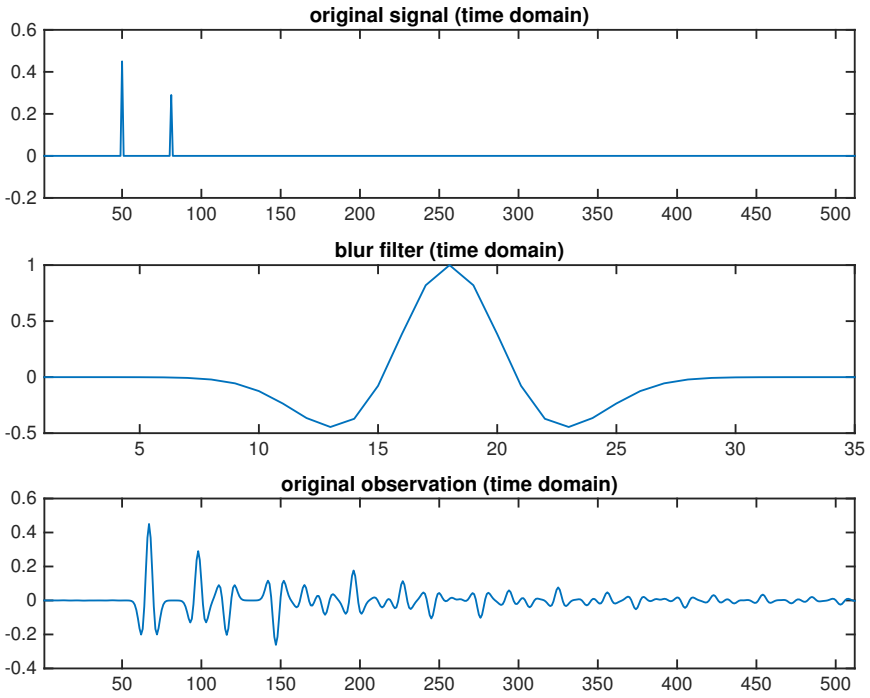



**Figure 2.9** A 1D EPSI model. The Green's function model $\mathbf{g}$ is plotted at the top with two impulsive events. The source signature $\mathbf{q}$ is a standard Ricker wavelet plotted in the middle. The resulting observation $\mathbf{p}$ produced through the EPSI model (equation 2.6) is plotted at the bottom.

that multiples are produced by using the recorded seismic response (with opposite sign) as the source wavefield in a new seismic survey, representing some kind of feedback system. This aligns with the notion that surface multiples are produced by the seismic responses bouncing back underground after hitting the recorders at the Earth's surface. Since convolution distributes over addition, multiple removal can thus also be seen as deconvolving this new "free-surface" source signature $(\mathbf{q} - \mathbf{p})$ from $\mathbf{g}$.

This kind of feedback system introduces some very interesting consequences to the uniqueness of the solutions for 2.6. Specifically, the feedback term prevents any arbitrary scaling of $\mathbf{q}$ (and a complementary scaling of $\mathbf{g}$) to produce the same observation. Figures 2.9 and 2.10 demonstrate this effect, and shows that simply scaling the source signature $\mathbf{q}$ would introduce additional spikes in the model $\mathbf{g}$ in order to produce the same observation $\mathbf{p}$. Thus seeking the sparsest solution for $\mathbf{g}$ also imposes additional amplitude

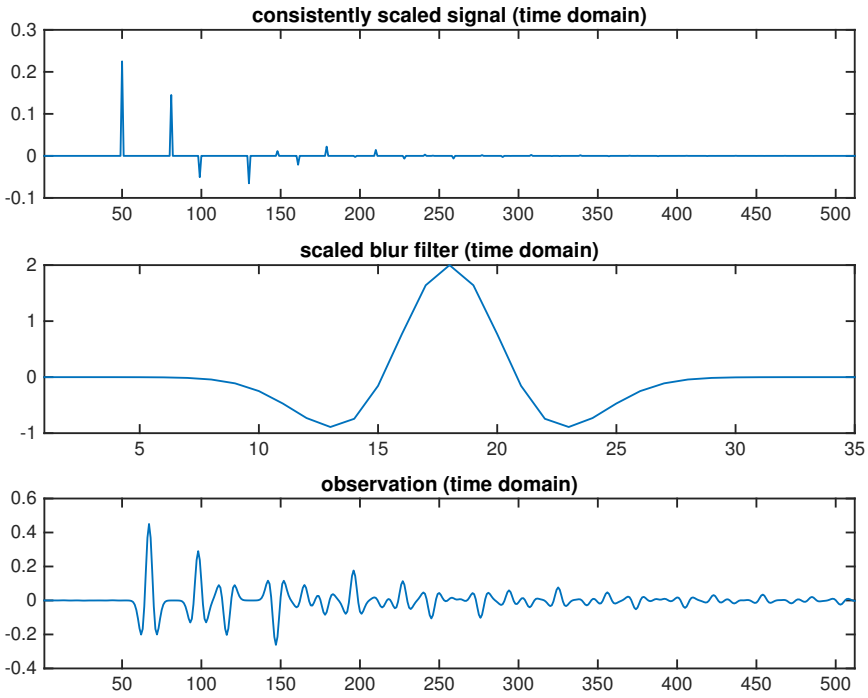



**Figure 2.10** The same 1D EPSI model as shown in Figure 2.9, but with the source signature model **q** (middle plot) scaled by 2. We can still find a Green's function **g** (top plot) that satisfies the original observation **p**, but it is less sparse than the original **g** as shown in the top panel of Figure 2.9.

constraints (I empirically observe a similar situation for phase information) on **q**. This kind of relation is non existent in the typical blind deconvolution, and serves to demonstrate why a minimum sparsity criteria on **g** alone can produce good solutions for the whole model, even though a large part of the EPSI problem essentially consists of blind deconvolution.

## 2.6 Outline of the thesis

Due to the relative novelty of EPSI, there exist many aspects of the method that, despite its inherent advantages, contributes to its current lack of widespread adoption in the industry. In this thesis I will strive to contribute improvements to the EPSI formulation that improve its reliability, capability, and practicality, in ways that will help to further the adoption

of this method. To this end, the following research chapters will focus on some particular aspects of its current shortcomings, as follows:

In **Chapter 3**, I reformulate the EPSI problem into a form that can be more easily solved by convex optimization, by utilizing an $\ell_0$ to $\ell_1$ relaxation and existing machineries in the large-scale optimization community. Despite the authors of EPSI recognizing the importance of a minimum sparsity criteria, the EPSI method only rudimentarily imposes sparsity by only updating the Green's function model with the few largest elements of each gradient update. This does not in general lead to the sparsest solution, and can produce many high-frequency errors in the model update. Instead of relying on hard thresholding of gradient updates, I develop a bi-convex optimization formulation for EPSI called Robust EPSI. I will demonstrate that this method leads to more accurate solutions for the Green's function with much less ad-hoc parameter tweaking.

In **Chapter 4**, I investigate a way to incorporate a free-surface scattering-based model into the convolution-based EPSI forward model of the multiples. This produces individual scattering terms that recreate the surface multiples independently from the convolution model. I show that this can be used to accurately predict the surface multiples inside of areas where there were large missing sample gaps in the acquisition (and integration) surface for the data wavefield, especially in the near-offset area where this kind of error is the strongest. This kind of approach is welcome, since the current methodology is to also estimate the missing pieces of the data, which increase the danger of mutual over-fitting between the different unknown wavefields. I propose modifications to the methods proposed in the previous chapter that are needed in order to successfully invert this augmented forward model, and show that it is more capable than the current methodology in EPSI at mitigating near-offset gaps.

In **Chapter 5**, I introduce a multilevel strategy to significantly shorten the time needed to compute solutions to the Robust EPSI problem. This method exploits the existing high computation complexity of wavefield convolution in reverse, by suggesting to offload early iterations (where accuracy is not important) to coarser spatial grids. Low-pass filtering and post-NMO interpolation is utilized to reliably avoid aliasing issues when moving between different sampling grids. I will analyze the expected computation complexity reduction, and walk through this strategy using a synthetic example dataset, showing how easily it can be implemented without changing the existing Robust EPSI algorithms introduced in the previous chapters.

# Chapter 3

# Robust estimation of primaries by $\ell_1$-norm minimization

The Estimation of Primaries by Sparse Inversion (EPSI) method, introduced in the last chapter, needs to invert for both the surface-free Green's function and the source signature at the same time, as they are both unknowns in the integral equation model for surface multiples. This results in an inversion that is analogous to blind deconvolution, with a high potential for local-minima issues, where spectral leakage can occur between the two unknowns that lead to incorrect solutions for both. It is opposed to the typical SRME procedure where the solution of the two unknowns are implicitly decoupled through the prediction-subtraction process. The main insight of EPSI is that modelling the surface-free Green's function as impulsive spikes poses a prior strong enough to overcome the inherent local-minima issues present in this inversion.

In van Groenestijn and Verschuur (2008; and partly in Guitton and Verschuur, 2004), it was shown that the minimizer of a sparsity measure on the primary wavefield yields the correct demultipled result during adaptive subtraction. Therefore, it would seem natural for EPSI to seek the sparsest possible solution for the surface-free Green's function. However, as we show in this chapter, the EPSI algorithm as proposed in (van Groenestijn and Verschuur, 2009a) does not automatically seek the sparsest solution, and leaves the burden of determining the appropriate level of sparsity to the practitioner. Furthermore, as I will demonstrate, the sparsity control parameters in the original EPSI do not strongly correlate to the final sparsity level of

the solution, so even if the structure of the subsurface is perfectly known (i.e., the number of reflectors) this information cannot be effectively used in the inversion. This shortcoming impacts the effectiveness of the EPSI in terms of its flexibility, reliability, and quality of its solutions, particularly in its ability to recover important late events of low amplitude.

## 3.1 Motivation

The central purpose of this chapter is to propose a new approach to EPSI that does attempt to seek the sparsest surface-free Green's function in a computationally efficient way, and in the process eliminate the free parameters involved in determining the solution sparsity. To achieve this I devise a reformulation of EPSI as a modification of the basis pursuit optimization problem (Chen and Donoho, 1994; Chen et al., 2001), and design a custom algorithm to find its solution. This results in a method that is applicable to a wide variety of datasets without the need to tweak inversion parameters. Because EPSI itself is a much more computationally expensive method compared to even SRME, its success hinges on whether it can reliably give meaningful results with minimal quality checks over a wide range of datasets. I feel that elimination of free parameters and the additional robustness that comes with always asking for the sparsest solution will be a crucial element for EPSI to be accepted as a more standard practice by industry.

I also aim to improve the quality of EPSI's estimated multiple-free Green's function for those that are interested in using it directly, instead of just the primary wavefield. Because I regard EPSI as an optimization problem, our reformulation takes advantage of recent advances in large-scale sparse recovery techniques (specifically that of van den Berg and Friedlander, 2008, 2011) for faster convergence. This alone resulted in a noticeable improvement in the fidelity of the recovered surface-free Green's function compared to the original algorithm under the same number of iterations. Furthermore, our framework is also flexible enough to allow estimating the multiple-free Green's function in arbitrary transform domains without changing the overall algorithm. This allows transform-domain sparsity to be leveraged for further uplifts in the quality of the directly estimated Green's function.

To motivate these goals, I show in Figure 3.1 a brief example of EPSI result obtained before and after our reformulation, focusing on direct recovery of small primary events at later times. Figure 3.1b is the the surface-free

Green's function obtained from running the original EPSI algorithm on the data shown in Figure 3.1a. To obtain this result I had to decide on several things: number of reflection events allowed into the Green's function per update, an estimate of the water-bottom muting time, and the time-window within which the Green's function is allowed to be updated in each iteration. With my approach, I was able to obtain the solution shown in Figure 3.1c without having to set any of those parameters, while taking the same amount of computational time. Figure 3.1d demonstrates the possible uplift in the quality of the direct solution by recovering the solution in a hybrid curvelet-wavelet transform domain. The overall number of updates on the Green's function remains the same in this case.

It is important to note beforehand that in this chapter I do not regard one important aspect of EPSI, which is the simultaneous reconstruction of missing near-offset data throughout the inversion. Rather, our scope is limited to in-depth analysis of the sparse regularization aspects of EPSI. That said, I do feel that enhancing the robustness of the near-offset reconstruction feature is also an important topic, and that the optimization framework I introduce in this chapter will benefit its future investigation.

In the following sections, I first review the central premise of EPSI, and formulate it as an optimization problem to highlight its potential limitations. I will then introduce a reformulation of EPSI by turning it into a basis pursuit problem with two disparate unknowns. I then discuss how to combine a continuation-based approach (Hale et al., 2008) with a block-coordinate descent scheme to efficiently solve this extended objective. A particular feature of our algorithm is that, while exploiting the Pareto curve (van den Berg and Friedlander, 2008; Hennenfent et al., 2008; Daubechies et al., 2008) of basis pursuit problems for the continuation-approach, it also ensures that the source signature is always estimated with sparse approximations to the surface-free Green's function. I illustrate the improvements of this approach over the original EPSI algorithm on synthetic datasets. Finally, I will use field datasets to demonstrate how recovery of primary late-arrivals can further be improved by seeking solutions in alternative wavefield representations.

## 3.2 EPSI in detail

Our goal in this section is to detail the methodology of EPSI as presented in van Groenestijn and Verschuur (2009a) and restate its formulation in a

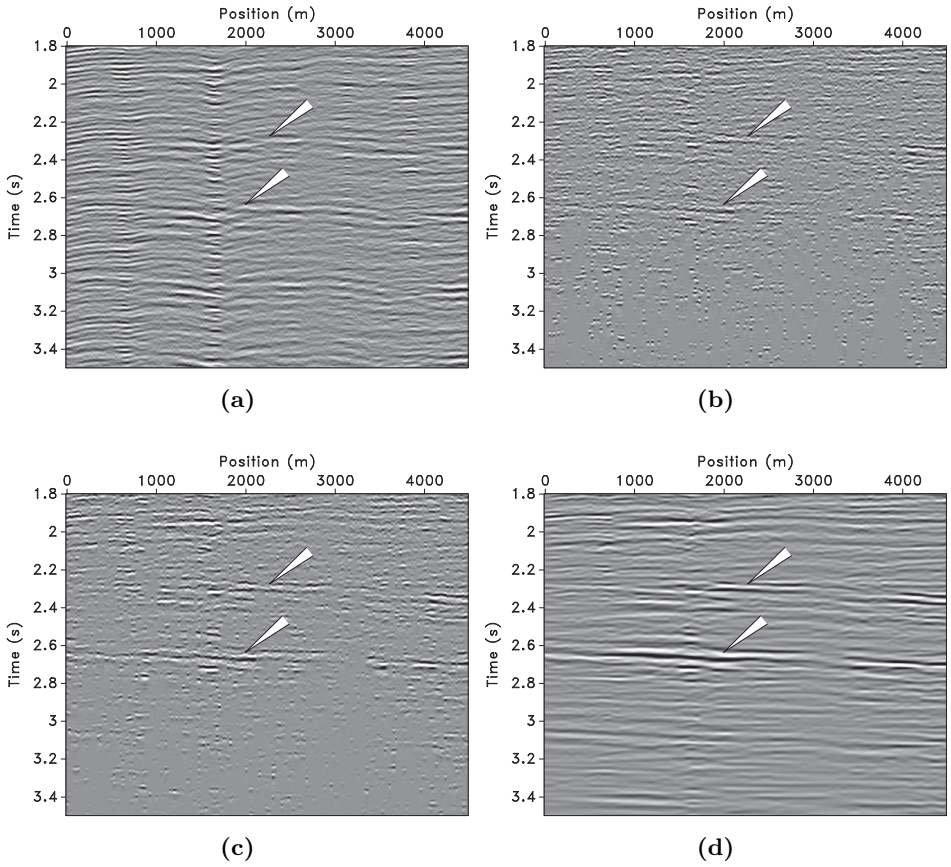



**Figure 3.1** Zoomed common-offset plot (200 m) of **(a)** a prestack field dataset from the North Sea,**(b)** the primary wavefield obtained by the original EPSI algorithm, **(c)** the primary wavefield obtained by our approach in the physical domain, and **(d)** results from our approach obtained under a curvelet-wavelet representation. The results in (b), (c), and (d) all took the same number of gradient updates. Compared to the original EPSI algorithm used in (b), the Robust EPSI algorithm was able to obtain the solution with much fewer input parameters. It also recovered the late primary events under the arrows in a more continuous fashion in the physical domain, but is still not satisfactory without stacking. The result obtained under a hybrid curvelet-wavelet representation shows a significantly improved recovery for this event. What appears to be a fault at 1800 m in the data is in fact localized surface multiple ringing due a syncline structure on the ocean bottom.

manner that more readily (compared to the original works) expresses the underlying optimization problem. This not only enables a more concrete notion of the original algorithm's shortcomings as described in the introduction, but also forms the basis of our extensions.

As a reminder of the notations established in the previous chapter, I present two wavefields pertinent to our discussion: the symbol $g(\mathbf{x}_r, t; \mathbf{x}_{\text{src}})$ which represents the desired surface-free Green's function (also called the primary impulse response) that models the response of the earth at location $\mathbf{x}$ due to an impulsive volume-injection dipole source located at $\mathbf{x}_{\text{src}}$. Aside from the source ghost, it does not include response due to a free surface, so it is free of surface-related multiples but will contain internal multiples. See Figure 3.2 for an illustration of its ray paths. Technically, $g$ is the normal derivative of the surface-free Green's function with respect to the reflecting surface. However, directly inverting for the normal derivative creates instabilities at far offset, so it is customary to include a time-derivative "obliquity factor" in the modelling operator to counteract this. With this change it is, for algorithmic purposes at least, acceptable to think of $g$ as a standard pressure wavefield.

Symbol $p(\mathbf{x}_r, t; \mathbf{x}_{\text{src}})$ represents the total upgoing pressure wavefield data, possibly approximated through receiver deghosting (or wavefield decomposition, if the data allows). Symbols $\mathbf{G}$ and $\mathbf{P}$ represent their discretized monochromatic slices (also called *data matrices*, as indicated by the uppercase). These slices are arranged in a matrix, with the monochromatic shot gathers organized in the columns. The overhead hat symbol is used in place of subscripted frequency indices. I use $q(t)$ to represent the source signature corresponding to our data $p$. Throughout the scope of this chapter I assume $q$ to be the same for all shots and traces, such that $\mathbf{Q}$ can simply be modelled by a frequency-dependent scaling $\mathbf{Q} = q(\omega)\mathbf{I}$ with $\mathbf{I}$ the identity matrix and $q(\omega)$ a scalar function of frequency. The matrix $\mathbf{R}$ models a surface reflection operator that is assumed to be $-\mathbf{I}$ (i.e., an ideal reflector) for the purpose of this chapter. The main physical relationship that underlies EPSI is thus expressed as

$$\mathbf{P} = \mathbf{GQ} + \mathbf{GRP}. \tag{3.1}$$

The main purpose of EPSI is to invert the relationship expressed in equation 3.1 for $g$ and $q$. On the left-hand side of equation 3.1 is $\mathbf{P}$ the observed data, while the right-hand side describes how to model this data using the two unknown quantities $\mathbf{G}$ and $\mathbf{Q}$, using $\mathbf{P}$ as essentially an integration

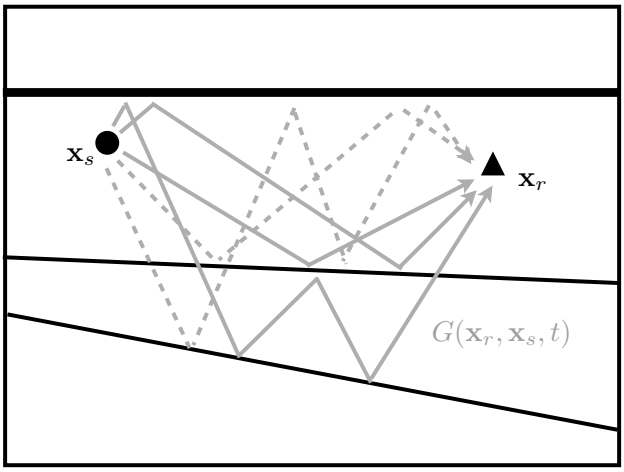



**Figure 3.2** An illustration of the ray paths (solid gray line) travelled by the surface-free Green's function $g(\mathbf{x}_r, t; \mathbf{x}_{\text{src}})$ as described in this chapter. It includes all responses at position $\mathbf{x}$ due to an impulsive pure volume injection source at $\mathbf{x}_{\text{src}}$ that have not been reflected by the free surface, *except* for the source ghost. Rays drawn in dashed grey lines are *not* part of $g$. Our assumed model of the primary wavefield is a source signature function $q$ is injected at $\mathbf{x}_{\text{src}}$ and recorded at $\mathbf{x}$ while following the ray paths of $g$. Think black line indicates the free-surface (e.g., water-air interface) while thin black lines indicate subsurface reflectors.

kernel. This matrix equation models a discretization of the inhomogenous integral equation shown in (2.2). Due to noise, imperfect up-down wavefield decomposition, and inaccuracies in physical assumptions (such as the presence of cross-line dips when processing a 3D line using a 2D algorithm), the data will not lie entirely in the range of the model. Solving equation 3.1 for $\mathbf{G}$ and $\mathbf{Q}$ thus requires an inversion that minimizes the energy of the mismatch between the left-hand and right-hand sides over all frequencies. In addition, the combined unknowns $\mathbf{G}$ and $\mathbf{Q}$ exceed the size of $\mathbf{P}$ and results in an underdetermined system, so the inversion requires additional constraints on the solution. Motivated by the impulsive nature of the Green's function, EPSI uses a time domain sparsity constraint on $g$ as a regularizing factor.

### 3.2.1 EPSI written in optimization form

As was done in the previous chapter, I express EPSI as an optimization problem in the form of minimizing an explicit objective, with the unknowns represented by vector quantities. I again define the EPSI forward-modelling operator $M$ as the right-hand side of equation 3.1 acting over all frequencies on $\mathbf{G}$ and $\mathbf{Q}$ and going under an inverse discrete Fourier transform in the time axis, such that we can restate equation 3.1 as $\mathbf{p} = M(\mathbf{g}, \mathbf{q}; \mathbf{p})$, this time over the whole dataset in physical dimensions instead of monochromatic slices. It is more convenient for $M$ to act on unknowns in the time domain because of the sparsity constraints we will impose in the time domain. For the remainder of this chapter I will drop the explicit parameterization of $M$ on $\mathbf{p}$, thus only writing $\mathbf{p} = M(\mathbf{g}, \mathbf{q})$. I will return to explicitly stating the parameterization in the next chapter, since we will make use of it then.

Note that if we fix either of the arguments of $M(\mathbf{g}, \mathbf{q})$, it becomes a linear operator with respect to the remaining variable; therefore $M$ is a bilinear operator. I denote the linearizations with respect to $\mathbf{g}$ and $\mathbf{q}$ as $\mathbf{M}_{\tilde{g}}$ and $\mathbf{M}_{\tilde{q}}$, and define the partial derivatives $\partial M/\partial \mathbf{q}$ and $\partial M/\partial \mathbf{g}$ evaluated at $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{q}}$ (using the overhead tilde to denote approximations to actual physical entities, or empirically determined quantities)

$$\begin{aligned} \mathbf{M}_{\tilde{q}}\mathbf{g} &:= \left(\frac{\partial M}{\partial \mathbf{g}}\right)_{\tilde{q}} \mathbf{g} = M(\mathbf{g}, \tilde{\mathbf{q}}), \\ \mathbf{M}_{\tilde{g}}\mathbf{q} &:= \left(\frac{\partial M}{\partial \mathbf{q}}\right)_{\tilde{g}} \mathbf{q} = M(\tilde{\mathbf{g}}, \mathbf{q}). \end{aligned} \tag{3.2}$$

Based on this description, we can use our new notation to express the EPSI procedure as equivalent to solving the following standard-form constrained optimization problem

$$\underset{\mathbf{g},\mathbf{q}\in\Lambda}{\text{minimize}} \, \|\mathbf{p} - M(\mathbf{g}, \mathbf{q})\|_2 \quad \text{subject to} \quad \|\mathbf{g}\|_0 \leq \rho, \tag{3.3}$$

where the $\ell_0$-"norm" $\|\mathbf{g}\|_0$ measures the number of non-zero samples in $\mathbf{g}$. Parameter $\rho$ controls the sparsity of the solution for the surface-free Green's function. The final demultipled result can be obtained in one of two ways: as a direct convolution of the final $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{q}}$ (called "direct result") or by subtracting from $\mathbf{p}$ the multiple model $M(\tilde{\mathbf{g}}, 0)$ (called "conservative result"). The latter effectively adds any residue wavefield from the mismatch term

back into the solution, and could be desired if amplitude loss from lack of convergence is an issue.

I use $\Lambda$ to express the set of candidate source signatures $\mathbf{q}$ that satisfy any desired or a priori constraints we wish to impose, e.g., decay rates in time and frequency, as well as phase properties. In this chapter I only impose that $\mathbf{q}$ is constrained in a short time window around $t = 0$, i.e., $q(t) = 0$ for $|t| \geq T$ when $q$ is represented in the time domain. Also, $\mathbf{q}$ is allowed to be anti-causal to account for any possible time shifts in the recorded data. As shown in van Groenestijn and Verschuur (2010), this model is easily extended to accommodate more complex source properties. For example, it is possible to to capture the offset dependence of a dipole source wavelet by parameterizing $\mathbf{Q}$ as diagonal matrix instead of a single scalar. This was not done in our chapter to keep the results directly comparable to that produced by the algorithm described in van Groenestijn and Verschuur (2009a). Furthermore, the time window is really only crucial in the early stages of the inversion to mitigate issues with local minima. In practice I find that it is usually safe to remove this time window entirely in the later stages of the inversion, when $\mathbf{g}$ is sufficiently recovered.

It is important to note that, due to the structure of $M(\mathbf{g}, \mathbf{q})$, our final estimate of $\mathbf{q}$ as the product of an inversion will necessarily capture all preprocessing on $\mathbf{p}$ that can be modelled by a filter of the same structure as $\mathbf{q}$. With the assumptions listed above, this means that an estimate of $\mathbf{q}$ will also capture global scalings, spectrum shaping, and various other preprocessing techniques commonly applied to the data. Conversely, $\mathbf{g}$ is completely determined by the physics of the free-surface, and its amplitude depends only on the relative strength between primary events and their associated multiple reflections in the data. As a consequence, the success of EPSI will depend very much on the internal consistency of the relative amplitudes between primaries and multiples in the input data.

In their original work on EPSI, van Groenestijn and Verschuur (2009a) proposed to approximately solve 3.3 by updating estimates $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{q}}$ (overhead tilde denotes estimated quantities) in an alternating Gauss-Seidel iteration using successive linearizations of $M(\mathbf{g}, \mathbf{q})$ as shown in equation 3.2. The updates to $\tilde{\mathbf{g}}$ and $\tilde{\mathbf{q}}$ are simple gradient updates for linearized problems. This scheme can be described as a cyclic block-coordinate descent algorithm, which I will briefly describe in the next section, before investigating some of the potential shortcomings of their original approach.

### 3.2.2 Optimizing over multiple variables in the EPSI algorithm

The most straightforward strategy for dealing with multiple unknowns is to simply lump them together into a single aggregate variable, and deal with the new problem using the same standard approaches as for single unknowns. However, this approach has immediate difficulties when the different unknowns are strongly coupled and differ in scale (You and Kaveh, 1996), such as in our case of jointly inverting for both a wavefield and its source signature. We also have completely different constraints for $\mathbf{g}$ and $\mathbf{q}$, which exacerbates the problem of working on them in unison. Alternatively, we can decrease the objective by iteratively optimizing over $\mathbf{g}$ with $\mathbf{q}$ fixed, then $\mathbf{q}$ with $\mathbf{g}$ fixed. This is an example of the cyclic block-coordinate descent methods (Bezdek et al., 1987).

Applying this approach to solve equation 3.3 leads to the original EPSI algorithm. We can define two new objective functions that are formed by taking the quadratic objective in equation 3.3 and fixing one of the variables at some value (as discussed above this also linearizes $M(\mathbf{g}, \mathbf{q})$ around that variable), which we write as $f_{\tilde{g}}(\mathbf{q}) := \frac{1}{2}\|\mathbf{p} - \mathbf{M}_{\tilde{g}}\mathbf{q}\|_2^2$ and $f_{\tilde{q}}(\mathbf{g}) := \frac{1}{2}\|\mathbf{p} - \mathbf{M}_{\tilde{q}}\mathbf{g}\|_2^2$. Algorithm 3.1 outlines what a typical cyclic coordinate descent approach would conform to.

**Algorithm 3.1** Cyclic block-coordinate descent for EPSI

1. **repeat**
2. $\quad \tilde{\mathbf{g}}_{k+1} \leftarrow \text{argmin}_{\mathbf{g}} \; f_{\tilde{q}_k}(\mathbf{g}) \quad \text{subject to} \quad \|\mathbf{g}\|_0 \leq \rho$
3. $\quad \tilde{\mathbf{q}}_{k+1} \leftarrow \text{argmin}_{\mathbf{q}} \; f_{\tilde{g}_{k+1}}(\mathbf{q}) \quad \text{subject to} \quad \mathbf{q} \in \Lambda$
4. $\quad k \leftarrow k + 1$
5. **until** $\|\mathbf{p} - M(\tilde{\mathbf{g}}_k, \tilde{\mathbf{q}}_k)\|_2$ converges to some criteria

If we do not individually minimize the two problems and instead simply seek *some* decrease in the objective at each iteration, we arrive at the inexact cyclic block-coordinate descent method. The original EPSI algorithm uses just a single gradient update to form the subsequent $\tilde{\mathbf{g}}_{k+1}$ and $\tilde{\mathbf{q}}_{k+1}$ (where $k$ is the iteration counter), and therefore is an example of such a method.

I now express the original EPSI algorithm in this notation. To begin, a descent direction for $f_{\tilde{q}_0}(\mathbf{g})$ is obtained though its gradient at $\tilde{\mathbf{g}} = 0$. For this first iteration $k = 1$, if we start with $\tilde{\mathbf{q}}_0 = 0$, this gradient is equivalent to the multidimensional autocorrelation of $\mathbf{p}$. To ensure $\|\tilde{\mathbf{g}}\|_0 \leq \rho$ at completion of the algorithm, this gradient is passed to some sparsifying operator $\mathcal{S}$ that zeroes most of its elements. The resulting sparse gradient is then scaled by

a line search and used to form $\tilde{\mathbf{g}}_1$. Next, we form the gradient for $f_{\tilde{g}_1}(\mathbf{q})$ and update $\tilde{\mathbf{q}}_1$ in a straightforward fashion (possibly with an implicit projection to ensure $\tilde{\mathbf{q}} \in \Lambda$). The process then iterates, stopping when a chosen criteria is satisfied, such as an iteration limit or a target misfit. Algorithm 3.2 provides a summary of this procedure, with $\alpha$ and $\beta$ denoting scalings that ensure the objective decreases sufficiently (i.e., line-search scalings).

**Algorithm 3.2** Original EPSI algorithm

1. $\tilde{\mathbf{g}}_0$, $\tilde{\mathbf{q}}_0 \leftarrow$ zero vector
2. $k \leftarrow 0$
3. **repeat**
4. $\quad \tilde{\mathbf{g}}_{k+1} \leftarrow \tilde{\mathbf{g}}_k + \alpha \mathcal{S}(\nabla f_{\tilde{q}_k}(\mathbf{g}_k))$
5. $\quad \tilde{\mathbf{q}}_{k+1} \leftarrow \tilde{\mathbf{q}}_k + \beta \nabla f_{\tilde{g}_{k+1}}(\mathbf{q}_k)$
6. $\quad$ Project $\tilde{\mathbf{q}}_{k+1}$ to satisfy $\tilde{\mathbf{q}}_{k+1} \in \Lambda$
7. $\quad k \leftarrow k + 1$
8. **until** $\|\mathbf{p} - M(\tilde{\mathbf{g}}_k, \tilde{\mathbf{q}}_k)\|_2$ converges to some criteria, or $k = k_{\text{tot}}$

In order to impose the sparsity constraint needed on $\tilde{\mathbf{g}}$, all updates on $\tilde{\mathbf{g}}$ goes through a predetermined sparsifying operator $\mathcal{S}$. However, sparse gradient updates of this form are not actually guaranteed to solve 3.3, so the solution will be heavily influenced by the specific choice of $\mathcal{S}$. As we see in the next section, ambiguity in the choice of $\mathcal{S}$ is also the source of most of the free parameters involved in the original EPSI algorithm.

### 3.2.3 Unpredictable nature of sparse updates

The specifics of the sparsifying operator $\mathcal{S}$ greatly affects the behaviour of the original EPSI algorithm. In order to both satisfy the terminating sparsity constraint $\|\tilde{\mathbf{g}}\|_0 \leq \rho$ and decrease the misfit objective as much as possible, $\mathcal{S}$ selects and keeps the largest amplitude $\rho/k_{\mathsf{tot}}$ elements of the update for $\tilde{\mathbf{g}}$, where $k_{\mathsf{tot}}$ is the number of total gradient iterations used for the problem, while the remaining elements are set to zero. Because $k_{\mathsf{tot}}$ is rarely known beforehand, the practitioner is burdened with the task of estimating an empirical per-update sparsity limit $\tilde{\rho}$ in place of $\rho/k_{\mathsf{tot}}$. To mitigate this somewhat, EPSI makes addition simplifications, such as assuming that seismic reflection events usually appear in all traces. This allows $\mathcal{S}$ to be carried out independently on a per trace basis, and thus $\tilde{\rho}$ can equivalently be specified by the number of largest-amplitude events kept from each trace.

One of the challenges of the original EPSI algorithm is that it is difficult to justifiably choose a $\tilde{\rho}$ for any non-trivial dataset. Intuitively, this parameter seems to correspond to the number of expected reflection events in the Green's function, and would thus correlate to the expected number of subsurface reflectors. However, because the number of gradient updates to recover a particular event with the right amplitude can vary wildly, even in the most favorable case imaginable, where the number of subsurface reflectors are entirely known beforehand, it is still difficult to choose $\tilde{\rho}$ such that the number of events in our final estimate of $\tilde{\mathbf{g}}$ match our expectations.

Furthermore, a poor choice of $\tilde{\rho}$ can have serious consequences for the output of the algorithm. If $\tilde{\rho}$ is allowed to be too large, then $\tilde{\mathbf{g}}$ may no longer resemble a impulsive Green's function after a few iterations. This will ultimately result in poor multiple removal, and I will demonstrate this aspect in the first figure of the Example section. In practice, this outcome is prevented by setting $\tilde{\rho}$ to a very small value (typically 2 to 5 events per trace are kept). Additionally, to avoid any energy being put into $\tilde{\mathbf{g}}$ at the location of multiples, a time-window, also empirically chosen, is applied on the updates. This time-window is allowed to expand with the iteration count, again in an arbitrary fashion, until it covers the entire time axis. The hope is that, as the time window expands to the location of a particular multiple reflection event, its primary event would have already been recovered sufficiently enough that the multiple would not appear in gradient updates.

On the other hand, since a small value of $\tilde{\rho}$ would only allow a small number of events in $\tilde{\mathbf{g}}$ to be updated at each gradient iteration, the practitioner is forced to aggressively ramp the number of iterations in order to recover late low-amplitude events that may be paramount to imaging reservoirs. In the effort to lower our error objective as quickly as possible, large-amplitude early events tend to dominate this $\tilde{\rho}$-event allowance of each update at the early stages of the algorithm, and the weaker events are only updated at higher iteration counts. Because each gradient update for $\tilde{\mathbf{g}}$ is computationally comparable to computing a SRME multiple prediction of the whole dataset (see van Groenestijn and Verschuur, 2009a, for detailed expressions), the additional gradient costs required are substantial. Thus in a typical iteration-constrained application of EPSI, these weaker events may therefore not even be present in the solution. Further exacerbating this issue is the trace-independent nature of $\mathcal{S}$. Because reflection events are not in general recovered at the same rate across all traces, $\mathcal{S}$ does not in turn guarantee that each update will affect the same events across all traces. This tends to affect the late low-amplitude events the most, resulting in

reflections that appear abruptly on some traces but not on adjacent ones. Refer back to Figure 3.1b for an illustration of this issue, and notice the discontinuous appearance of events past 2 s.

For the practitioner, the potential for excessive parameter tweaking, which results from the aforementioned complications due to reliance on $\mathcal{S}$, is highly undesirable. This is avoided if we instead place the sparsity measure of $\tilde{\mathbf{g}}$ in the objective, thus seeking the sparsest solution for the surface-free Green's function. This is closer to the arguments for sparse regularization of the primary Green's function laid out in van Groenestijn and Verschuur (2008), which precedes the formal introduction of EPSI. There, a sparsity measure on the primary wavefield is shown to be a better objective function compared to the energy measure for the purpose of resolving overlapping primary and multiple events in adaptive subtraction for a purely impulsive source.

## 3.3 Robust EPSI

As discussed in the previous section, swapping the data misfit $\|\mathbf{p}-M(\mathbf{g},\mathbf{q})\|_2$ and the sparsity measure of the Green's function $\|\mathbf{g}\|_0$ in equation 3.3 results in a more robust and flexible formulation of the EPSI problem, but also necessitates a different algorithm. Our main contribution here is to demonstrate that this form can be solved in a computationally efficient manner by treating it as an extended $\ell_1$-norm minimization problem (also known as basis pursuit, Chen et al., 2001) for $\mathbf{g}$ that also simultaneously solves for the source signature $\mathbf{q}$. I call this approach the Robust Estimation of Primaries by Sparse Inversion. The prefix *robust* is chosen mainly to reflect a newfound lack of dependence on empirical inversion parameters.

I will first present the Robust EPSI algorithm itself in Algorithm 3.3 before explaining each line in depth. The main part of the algorithm consists of an initialization step to approximate the source signature (lines 2-4), and the main iterations (lines 7-15) of alternating between solving $\ell_1$-norm constrained problems for $\tilde{\mathbf{g}}$ (line 9-10) and matching for the source signature $\tilde{\mathbf{q}}$ (line 12). For now I will leave the technicality of the initialization step to the last part of this section, and focus instead on how the main loop solves the EPSI problem with the goal of obtaining the sparsest solution for $\mathbf{g}$.

**Algorithm 3.3** The proposed Robust EPSI algorithm

1. **Input:** target residual energy $\tilde{\sigma}$, recorded surface seismic data $\mathbf{p}$
2. $\tilde{\mathbf{g}}_0 \leftarrow$ initialize via single event picking as described in text
3. $\tilde{\mathbf{q}}_0 \leftarrow \text{argmin}_{\mathbf{q}} \ \|\mathbf{p} - \mathbf{M}_{\tilde{g}_0}\mathbf{q}\|_2^2 \quad \text{subject to} \quad \mathbf{q} \in \Lambda$
4. $\tilde{\mathbf{g}_0} \leftarrow$ reset to zero vector
5. initialize iteration counter $k \leftarrow 0$, $\ell_1$-norm constraint $\tau_0 \leftarrow 0$
6. if using transforms, set $\mathbf{S}^H$ as synthesis operator, otherwise set $\mathbf{S}^H = \mathbf{I}$
7. **repeat**
8. $\quad \tau_{k+1} \leftarrow$ determine from $\tilde{\sigma}$ and $\tau_k$
9. $\quad \mathbf{x} \leftarrow \text{argmin}_{\mathbf{x}} \|\mathbf{p} - \mathbf{M}_{\tilde{q}_k}\mathbf{S}^H\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau_{k+1}$
10. $\quad \tilde{\mathbf{g}}_{k+1} \leftarrow \mathbf{S}^H\mathbf{x}$
11. $\quad$ scale $\tilde{\mathbf{g}}_{k+1}$ by $s_{k+1}$ according to (3.7)
12. $\quad \tilde{\mathbf{q}}_{k+1} \leftarrow \text{argmin}_{\mathbf{q}} \ \|\mathbf{p} - \mathbf{M}_{\tilde{g}_{k+1}}\mathbf{q}\|_2 \quad \text{subject to} \quad \mathbf{q} \in \Lambda$
13. $\quad$ undo scaling of $\tilde{\mathbf{g}}_{k+1}$ by $s_{k+1}$
14. $\quad k \leftarrow k + 1$
15. **until** $\|\mathbf{p} - M(\tilde{\mathbf{g}}_k, \tilde{\mathbf{q}}_k)\|_2 \leq \tilde{\sigma}$, or a predetermined iteration limit is reached
16. **Output:** estimated Green's function $\tilde{\mathbf{g}}_k$ and source wavelet $\tilde{\mathbf{q}}_k$

### 3.3.1 Sparsity via basis pursuit (lines 7-15)

Directly putting $\|\mathbf{g}\|_0$ in the objective leads to a combinatorial problem, and most heuristics-based algorithms for problems of this type are unstable under noise. Instead, I take the well-known approach of replacing the $\|\mathbf{g}\|_0$ objective with a $\ell_1$-norm objective $\|\mathbf{g}\|_1$ in the problem. This heuristic in the context of inverting linear systems is an established practice (Donoho, 2006) in many fields, including exploration seismology (Claerbout and Muir, 1973), notably in the practice of spiking deconvolution (Taylor et al., 1979; Dossal and Mallat, 2005; Herrmann, 2005), amongst many others.

A naive adoption of the the block-coordinate descent scheme used to solve the original EPSI problem (described in Algorithm 3.1) would then simply iterate over the following two subproblems:

$$\tilde{\mathbf{g}}_{k+1} \leftarrow \underset{\mathbf{g}}{\text{argmin}} \ \|\mathbf{g}\|_1 \quad \text{subject to} \quad \|\mathbf{p} - \mathbf{M}_{\tilde{q}_k}\mathbf{g}\|_2 \leq \sigma \tag{3.4}$$

$$\tilde{\mathbf{q}}_{k+1} \leftarrow \underset{\mathbf{q}}{\text{argmin}} \ \frac{1}{2}\|\mathbf{p} - \mathbf{M}_{\tilde{g}_{k+1}}\mathbf{q}\|_2^2 \quad \text{subject to} \quad \mathbf{q} \in \Lambda, \tag{3.5}$$

where the objective function in the first subproblem (3.4) has been replaced so that the minimization is over $\|\mathbf{g}\|_1$ and an error tolerance constraint $\sigma$

is placed on the misfit function. This is often referred to as a basis pursuit denoise (BPDN, see Fuchs, 2005; Candès et al., 2006; Donoho, 2006; Tropp, 2006) problem, and is a well-studied convex problem, where the only local minimum is the global minimum. Its solution is generally a close approximation to the sparsest (minimum $\ell_0$-norm) solution even under the presence of noise in the observed data. Although the stability of spiking deconvolution using the BPDN approach is currently a well-known issue when significant noise is present, I note that despite similar appearances the EPSI problem is not a pure spiking deconvolution problem, as the Green's function is not only inverted from wavelet convolution but also from the multiple model. The other subproblem (3.5) related to updating the estimate of the source wavelet $\tilde{\mathbf{q}}_{k+1}$ remain untouched compared to Algorithm 3.1.

The size of the optimization problem involved in our application (where the data and the unknowns are both entire prestack wavefields) is very large compared to most applications of BPDN. Fortunately, computationally efficient solvers for BPDN is currently a very active area of research (see Yuan et al., 2010, for a review). However, compared to problems with an well-behaved differentiable objective function such as the $\ell_2$-norm misfit in equation 3.3, the non-differentiability of the $\ell_1$-norm makes BPDN problems inherently more difficult to solve, and in general require more iterations to converge.

On the other hand, the magnitude of our convergence issue is even larger when we consider it as part of a larger overall block-coordinate descent algorithm. The estimate of $\mathbf{q}$ needs to wait until the previous BPDN problem for $\mathbf{g}$ is solved, as in general BPDN problems only produce sparse solutions upon convergence. With these issues in mind, a direct approach of successively solving subproblems 3.4 and 3.5 seems likely to require an unacceptably large number of iterations to produce satisfactory results. Our solution to this problem is to take advantage of the specific structure of continuation methods, as I discuss in the next section.

### 3.3.2 Key insight on the role of continuation techniques (line 8)

The main contribution of our work is recognizing that continuation approaches, which was shown to greatly improve the convergence of BPDN algorithms (Hale et al., 2008), also naturally give us a good way to simultaneously estimate the source signature for the EPSI problem. This allows us to solve our reformulated problem in a feasible way that also retains a con-

vergence guarantee for the BPDN problem. Continuation-based techniques are currently amongst of the fastest methods for solving large-scale BPDN problems. The general idea is to break the BPDN problem down into a series of subproblems, each using the solution of the previous problem as the initial starting point, with their solutions converging to the true BPDN solution. This specific structure is inspired by the general observation that heavily constrained optimization problems generally converge much faster than optimization problems with a more relaxed constraint. This holds true for $\ell_1$-norm constrained problems as well; problems with a tight $\ell_1$-norm constraint on the unknown variable are often much faster to solve than the same problem with a looser $\ell_1$-norm constraint (Donoho and Tsaig, 2008; Malioutov et al., 2005).

To use the continuation scheme in our BPDN problem, we replace the BPDN problem in subproblem 3.4 with a $\ell_1$-constrained problem (typically called a Lasso problem)

$$\underset{\mathbf{g}}{\text{minimize}} \, \|\mathbf{p} - \mathbf{M}_{\tilde{q}}\mathbf{g}\|_2 \quad \text{subject to} \quad \|\mathbf{g}\|_1 \leq \tilde{\tau}_k, \tag{3.6}$$

where $\tilde{\tau}_k$ is a $\ell_1$-norm constraint that varies over the cyclic block-coordinate descent, using $k$ as the iteration counter. If we have a well-defined series of constraints $\tilde{\tau}_0 < \tilde{\tau}_1 < \tilde{\tau}_2 < \ldots < \tilde{\tau}_k$ that converges to the minimum constraint required for the misfit function to reach a chosen error level, we end up converging to the solution to the BPDN problem. It turns out that the series of constraints $\tilde{\tau}_k$ can be chosen in a natural way by applying a Newton root-finding method to the Pareto trade-off curve of our BPDN problem (van den Berg and Friedlander, 2008), as the solution to each of the Lasso problems give the slope of the Pareto curve at $\tilde{\tau}_k$ in closed form. Readers unfamiliar with using Pareto root-finding for $\ell_1$-minimization can refer to Appendix A where it is explained in detail.

The Lasso problems described in equation 3.6 have a differentiable objective function, and are easily solved (most commonly by projected gradient methods, see van den Berg and Friedlander, 2008). Due to the use of continuation, I find that the number of gradient updates required for each Lasso problem is on the order of 10. The overall Newton root-finding method for the correct $\tilde{\tau}$ that solves BPDN, when initialized at $\tau_0 = 0$ and a zero solution vector, also tends to converge quickly, typically within 8 to 10 Newton steps. See Figure 3.3 for a hypothetical solution path of $\tilde{\mathbf{g}}$ for 3 iterations of the Newton root-finding method.

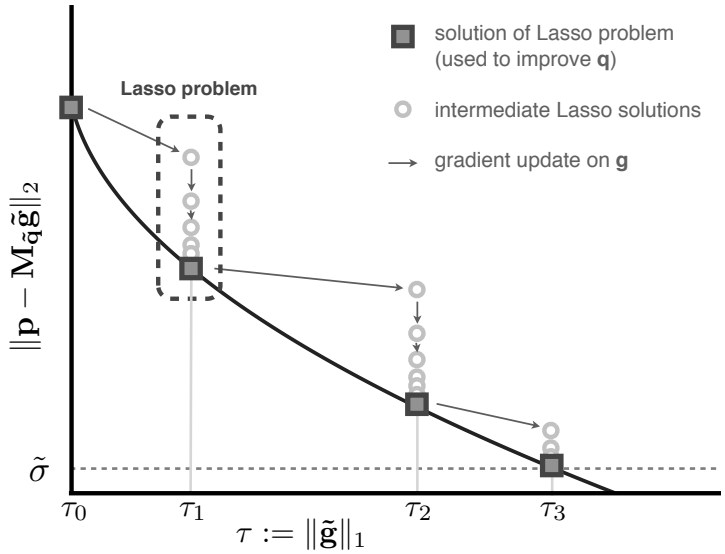



**Figure 3.3** Illustration showing the evolution of $\tilde{\mathbf{g}}$ (grey dots, both circle and squares) in relation to the Pareto curve function as it follows a Pareto root-finding continuation method towards the BPDN solution for a chosen misfit $\tilde{\sigma}$. The optimal (minimum) $\tau$ for the BPDN problem lies where the Pareto curve intersects the dashed line. Solutions to the Lasso problem (square dots) lie on the Pareto curve and is also the solution to an equivalent BPDN problem, and is likely to be sparse. These solutions are used to refine the source signature estimate (line 12 of Algorithm 3.3).

I make the observation that these Lasso problems also serve as the answer to how the source signature should be estimated throughout the course of the algorithm. As each of the Lasso problems are solved, we obtain a series of approximations to the surface-free Green's function $\mathbf{g}$. Due to the nature of constrained convex optimization problems, they are also each solutions to a BPDN problem with a particular data misfit constraint, determined uniquely by $\tilde{\tau}_k$. We therefore can justifiably believe that each $\tilde{\mathbf{g}}_k$ are also sparse. Figure 3.4 illustrates that the Lasso problems lead to sparse estimates of the primary Green's function. These are the best points within the algorithm to refine our estimation of the source signature $\mathbf{q}$, as they allow $\mathbf{q}$ to fully capture the spectral properties of the wavefield. By turning subproblem 3.4 into a Lasso problem, we essentially end up with an analogue of the inexact cyclic block-coordinate descent scheme.

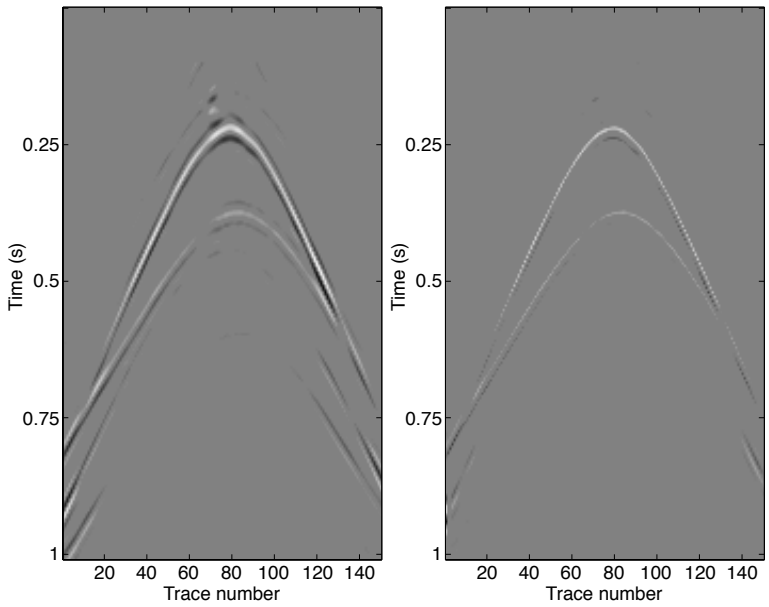



**Figure 3.4** A typical estimate for the Green's function $\tilde{\mathbf{g}}$ at **(a)** the first iteration of a Lasso problem (equation 3.6), and **(b)** at the final iteration. This Lasso problem was solved via a spectral-projected gradient method (van den Berg and Friedlander, 2008) using 8 gradient updates. This demonstrates that Lasso problems tend to result in sparse solutions.

### 3.3.3 Matching for the source wavelet (lines 11-13)

The wavelet matching step is reflected in line 12 of Algorithm 3.3. It just involves solving the subproblem stated in equation 3.5, once the previous Lasso problem is solved and a sparse approximation of the surface-free Green's function $\tilde{\mathbf{g}}_k$ is obtained. I advocate solving the wavelet matching problem to convergence, which happens quickly due to the large amount of observed data compared to the small number of unknowns for the wavelet. If $\mathbf{q}$ is optimal around the current linearization such that its partial derivative of the objective function vanishes, then the next gradient update for $\tilde{\mathbf{g}}$ will be very close to a gradient for the overall EPSI problem, not just the Lasso subproblems (see Aravkin and van Leeuwen, 2012, for an argument of why this is beneficial). Note that subproblem 3.5 essentially involves finding a filter (satisfying any constraint specified in $\mathbf{\Lambda}$) that matches two wavefields:

the current guess of the Green's function $\tilde{\mathbf{g}}$ and the current estimate of the primary wavefield $\mathbf{p} - M(\tilde{\mathbf{g}}, 0)$. It is vastly overdetermined, and if a sophisticated algorithm such as Least-Squares QR (Paige and Saunders, 1982) is used, $\tilde{\mathbf{q}}$ will converge extremely quickly.

The updated wavelet estimate is subsequently used to form a new linearization $\mathbf{M}_{\tilde{q}_k}$ for the next Lasso problem. In theory, this will perturb the underlying Pareto curve, and we can no longer safely expect the original solution $\tilde{\mathbf{g}}_k$ to be approximately optimal after the re-linearization. However, the first-order effect of this perturbation largely contributes to inaccuracies in the computed $\tau_{k+1}$, which mostly governs how quickly the root-finding procedure locate the minimum $\ell_1$-norm $\tau^*$ for the target $\tilde{\sigma}$. The effects of errors in approximating the Pareto curve solution, and its subsequent effect on the convergence rate of the root-finding procedure, is studied in van den Berg and Friedlander (2008), where the sensitivity of the convergence rate of $\tau$ to errors in sampling the Pareto curve is shown to be directly related to the conditioning of each $\mathbf{M}_{\tilde{q}_k}$. Empirically I found that $\tau_{k+1}$ as determined before and after the re-linearization rarely differed by more than 2 to 3 percent, especially towards the end of the Newton iterations, and therefore I regard the perturbation of the Pareto curve to be negligible on the quality of the final solutions.

An important technicality needs to be addressed here: Lasso problems place an $\ell_1$-norm constraint on the Green's functions, which indirectly limits the amplitude of the primary reflection events in $\tilde{\mathbf{g}}_k$. This is a necessary trade-off unique to the BPDN approach that depends on using the $\ell_1$-norm as a heuristic for sparsity. Consequentially, we can expect $\tilde{\mathbf{g}}_k$ to underestimate the amplitudes of the reflection events. To ensure that our estimates of $\mathbf{q}$ have the correct amplitude, we need to appropriately rescale the events in $\tilde{\mathbf{g}}_k$. Under the assumptions of a global source signature for every event in the data, it should suffice to scale the entire $\tilde{\mathbf{g}}_k$ with a single factor $s_k$ that minimizes $\|\mathbf{p} - \mathbf{M}_{\tilde{q}_{k-1}}(s_k\tilde{\mathbf{g}}_k)\|_2$. Note that this scaling does not perform the "debiasing" that is common in the solution of BPDN problems (for example, Figueiredo et al., 2007). It only attempts to mitigate the issue that Lasso solutions early on in the Newton iterations are obtained with a severely underestimated $\ell_1$-norm. This is essentially an exact line-search scaling, which is given by

$$s_k = \frac{\tilde{\mathbf{g}}_k^H \mathbf{M}_{\tilde{q}_{k-1}}^H \mathbf{p}}{\|\mathbf{M}_{\tilde{q}_{k-1}}\tilde{\mathbf{g}}_k\|_2^2}, \tag{3.7}$$

with the symbol $^H$ denoting Hermitian adjoint. This scaling needs to be restored before the next Lasso problem for $\tilde{\mathbf{g}}_{k+1}$ is formulated, to make sure we return to the neighborhood of the Pareto curve. In my experience I find that $s_k$ tend to start at around 2 to 3 and will converge to 1 as $\tau$ converges to $\tau^*$, which suggests that $s_k$ might be useful as a convergence metric for the overall root-finding algorithm. This scaling is reflected in lines 11 and 13 in Algorithm 3.3.

### 3.3.4 Incorporating sparsifying transforms (lines 6, 9-10)

One of the advantages of the Robust EPSI algorithm is that seeking the primary Green's function $\tilde{\mathbf{g}}$ in terms of coefficients of transform domains is relatively straightforward. This is because the sparsity of our solution comes from solving Lasso problems, and the algorithms we utilize are agnostic to the linear forward operator. If the original EPSI approach is used, the sparsifying operator $\mathcal{S}$, which is formulated with physical arguments, might need to be redesigned for the specific transform in a non-trivial way in order to retain convergence. For Robust EPSI, we are free to compound $\mathbf{M}_{\tilde{q}}$ with any linear synthesis transform without changing the underlying algorithm.

Depending on the specific transform chosen, this can be leveraged for several benefits. As discussed in Herrmann (2010), a sparse representation for $\tilde{\mathbf{g}}$, such as curvelets, will enhance the effectiveness of $\ell_1$-norm minimization, which in turn strengthens our BPDN approach. Also, a representation that is composed of atoms that spans over physical space can be called upon to mitigate spatial undersampling issues in $\mathbf{p}$. For example, $\tilde{\mathbf{g}}$ can be sought as coefficients in the hyperbolic Radon domain, analogous to the approach used in van Dedem and Verschuur (2005). I show in Figure 3.1 as well as the Numerical Examples section that this can be leveraged to improve the recovery of weak late arrivals, which can prove challenging even to the Robust EPSI formulation due to comparatively weaker associated multiples from spherical divergence and a naturally lower SNR compared to earlier events).

To modify the Lasso problems for this extension, the synthesis operator represented by the symbol $\mathbf{S}^H$ of the transform domain, which maps the domain coefficients to the physical signal, simply needs to be compounded with $\mathbf{M}_{\tilde{q}}$ in equation 3.6 to form a new Lasso problem

$$\underset{\mathbf{x}}{\text{minimize}} \|\mathbf{p} - \mathbf{M}_{\tilde{q}}\mathbf{S}^H\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau. \tag{3.8}$$

which is solved using exactly the same algorithm as the one used for (3.6). This essentially re-parameterizes the unknown Green's function as coefficients in the transform domain. Once the solution coefficients $\tilde{\mathbf{x}}_k$ are obtained, the physical Green's function can be recovered once again by the synthesis operation $\tilde{\mathbf{g}}_k = \mathbf{S}^H \tilde{\mathbf{x}}_k$. This should be done before subsequently refining $\tilde{\mathbf{q}}$. Both the Lasso algorithm and the Newton iterations do not need to be modified, as long as we redefine $\mathbf{A} = \mathbf{M}_{\tilde{q}} \mathbf{S}^H$ instead of $\mathbf{A} = \mathbf{M}_{\tilde{q}}$ for equations A.1 and A.2. These additions to the algorithm to accommodates the transform domain is reflected in lines 6, 9, and 10 in Algorithm 3.3.

### 3.3.5 Initial calibration of source signature (lines 2-4)

An effective approach when dealing with problems with inherent local minima issues is to use as much a priori information as possible to narrow down the appropriate region of the solution space at the outset of the inversion (see page 136 of Biggs, 1998, for a relevant discussion). For example, in the application of blind photographic deblurring, a common strategy is to first identify edges in the image, which allows the algorithm to quickly establish the correct shape of the unknown point spread function using a contrast-maximizing objective. I find that an analogous technique is similarly effective for our case. It essentially involves careful identification of the largest impulsive event per trace in $\tilde{\mathbf{g}}$, which is then matched with the data to rapidly obtain a reasonable candidate for $\tilde{\mathbf{q}}_0$.

This initialization starts with a multidimensional autocorrelation of $\mathbf{p}$ (equivalent to a gradient update for $\mathbf{g}$ when $\tilde{\mathbf{q}} = 0$). From each trace of this autocorrelation, we pick out the single strongest event that coincides with a primary reflection in $\mathbf{p}$, which will typically be the ocean-bottom reflection. This can be achieved in a variety of ways, such as a combination of muting and automatic first break picking, and can be manually verified if needed. We then identify its peak location as the arrival time of an impulsive reflection, and place a single spike at that time in our otherwise empty initial model of the surface-free Green's function. After this is done for each trace, we ensure that the amplitude of these spikes are in the correct neighborhood, scaling with the exact line-search formula in equation 3.7, setting $\mathbf{M}_{\tilde{q}_{k-1}} \mathbf{g} = \mathbf{M}_0 \mathbf{g} := M(\mathbf{g}, 0)$. A full back-projection can also be done by extending $s$ to a diagonal weighting, allowing each spike to be weighted individually. Physically, this is interpreted as adjusting the amplitudes of every event in the surface-free Green's function so that collectively they model the multiple events as accurately as possible through operator $M(\tilde{\mathbf{g}}, 0)$.

Finally, setting this methodically calibrated single-event approximation of $\mathbf{g}$ as $\tilde{\mathbf{g}}_0$, we can solve equation 3.5 to quickly obtain a good initial source signature estimate $\tilde{\mathbf{q}}_0$ with minimal fear of overfitting. This accomplishes our goal of negating the risk of finding a $\tilde{\mathbf{q}}_0$ so wildly incorrect that it precludes reasonable solutions of $\tilde{\mathbf{g}}$ (even when it satisfies the constraint set $\Lambda$ that we predetermined for $\mathbf{q}$). Once $\tilde{\mathbf{q}}_0$ is obtained, we can optionally choose to minimize the imprint of the initialization by discarding $\tilde{\mathbf{g}}_0$ and reset it to a zero vector $\tilde{\mathbf{g}}_0 = \mathbf{0}$ before starting the cyclic block-coordinate descent iterations. The steps discussed here correspond to lines 2-4 in Algorithm 3.3.

## 3.4 Numerical comparison with original EPSI

I demonstrate the performance of Robust EPSI by applying it to a 2D synthetic seismic dataset that is also used by van Groenestijn and Verschuur (2009a). A shot gather from the center of the model is shown in Figure 3.5a, with arrows indicating the largest surface multiples. The dataset is modelled from a 2D synthetic model that was depicted in figure 1 of the aforementioned paper. It has a water layer approximately 200 m deep and a laterally-varying salt dome layer that is situated roughly between the depths of 400 m and 800 m. The split-spread pressure data (up to zero-offset) is modelled by a second-order in time, fourth-order in space finite-difference acoustic program using a zero-phase Ricker wavelet with peak frequency 30 Hz as source signature, and is deghosted at the receiver-side. We have a well-sampled source line that exactly coincides with the receiver line, so the multidimensional convolution for multiple estimation is free of aliasing issues. A factor of $\sqrt{\omega}$ is included in $\mathbf{M}_{\tilde{q}}$ to account for the technical requirement that $\mathbf{g}$ should be the normal derivative of the Green's function with respect to the free-surface. This matches the "obliquity factor" that is also used in van Groenestijn and Verschuur (2009a), and allows us to compare directly with the original EPSI results.

From this data I obtained two estimates of the surface-free Green's function according to the original EPSI algorithm as described in Algorithm 3.2 after 80 iterations. Figure 3.5b depicts the result obtained without the sparsifying operator $\mathcal{S}$ (i.e., kept each gradient update for $\tilde{\mathbf{g}}$ untouched), and serves to illustrate the importance of sparsity in EPSI. Not only was the primary multiple removal relatively ineffectual (e.g., the water-bottom mul-

tiple at 0.45 s), but also the result clearly does not resemble reflection events due to an impulsive source. Figure 3.5c depicts the result obtained by setting $\mathcal{S}$ to keep the 4 largest amplitude events per trace for every update on $\tilde{\mathbf{g}}$. This greatly improved the multiple rejection and delivered the impulsive appearance we expect from the surface-free Green's function.

Both of these results required a triangle muting mask on $\tilde{\mathbf{g}}$ throughout the inversion procedure (as part of the global time-window implicitly included in $\mathcal{S}$) to prevent placing events before the first water-bottom reflection event around 0.25 s. However, there are still visible artifacts at the boundaries of the triangle mute. Since these anti-causal events are not in the original data, they indicate update errors introduced when $\tilde{\mathbf{q}}$ is severely underestimated in amplitude, most likely at one of the earlier iterations. Because the original EPSI inversion does not attempt to seek the sparsest solution, I find that these errors decay very slowly with the iteration count. To make these artifacts appear more visible in the plots I slightly exaggerated the clipping percentile of the colourmap for plots of the Green's functions only.

Meanwhile, Figure 3.5d shows the result produced by Robust EPSI. I set $\tilde{\sigma}$ to 2% of total data energy in anticipation of edge effects, but the program is terminated early at the equivalent of 82 gradient updates of the Green's function. I find that for Robust EPSI we no longer need to depend on a tight muting mask to reject anti-causal events. Note the absence of the aforementioned anti-causal artifacts, which indicates that we are close to good physical approximations with our $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{q}}$. Furthermore, we can more clearly see here that reflection events show characteristics of the expected dipole response of the Green's function, although it is mitigated somewhat by the inclusion of the obliquity factor. I should note that most of these artifacts appear to be high-frequency and sit in the nullspace of wavelet convolution with $\mathbf{q}$, and will be highly attenuated with a suitable low-pass filter, but not completely. This can be seen in Figures 3.6a and 3.6b which shows the Green's function in Figures 3.5c and 3.5d, respectively, convolved with the estimated source signature $\tilde{\mathbf{q}}$ shown in Figure 3.7. It is interesting to find that Robust EPSI does not produce artifacts of this nature, and serves to demonstrate that $\ell_1$-norm regularization naturally attenuates small random-like events that do not contribute to explaining the data via the forward model. This is one of the reasons why it is often used to heuristically obtain the sparsest solution.

Aside from the absence of artifacts, the fidelity of the Green's function obtained from the Robust EPSI procedure is also qualitatively better, with the weaker reflections around 0.8 s and the diffracted events directly below the water-bottom reflection more cleanly defined compared to the original EPSI results. In both programs the wavelet was allowed to live within a $\pm$100 ms time window. The projected gradient methods used to solve the set of Lasso problems in Robust EPSI converged quickly, typically within 5 to 20 projected gradient updates for each Lasso problem. The root-finding procedure updated $\tau$ a total of 6 times throughout this procedure. At the end of the procedure, the original EPSI achieved a relative data-space $\ell_2$-norm residual of 0.079 while Robust EPSI achieved a lower 0.044; their residue wavefields $\mathbf{p} - M(\tilde{\mathbf{g}}, \tilde{\mathbf{q}})$ are plotted in Figures 3.6c and 3.6d respectively. While the residue of the original EPSI is swamped by the aforementioned artifacts, for Robust EPSI the residue more clearly demonstrates parts of the data that do not fit the forward model, which in this case show a offset-dependent nature, possibly due to either the obliquity factor or imperfect deghosting. As discussed in van Groenestijn and Verschuur (2009b), looking at the residue in this way is useful for identifying potential inconsistencies within the dataset.

## 3.5 Additional numerical examples

### 3.5.1 Synthetic Pluto 1.5 dataset

To demonstrate that the Robust EPSI algorithm is insensitive to optimization parameters, I apply it to different datasets without appreciably changing parameters from the values used in the previous section. Here I use the Pluto 1.5 dataset publicly released by the SMAART JV consortium, which was produced using a fourth-order finite difference P-SV modelling code with a Ricker wavelet of peak frequency 15 Hz. Not only did I keep the exact same parameters used in the previous example, I also directly used the pressure recording at the surface, foregoing any preprocessing such as deghosting or up-down wavefield decomposition. Unlike the example in the previous section, an obliquity factor was not included in the forward model. Nevertheless, the inversion process remained stable and produced a Green's function with a clean impulsive appearance while successfully rejecting multiples. See Figure3.8b for a shot gather plot. Without any change in the

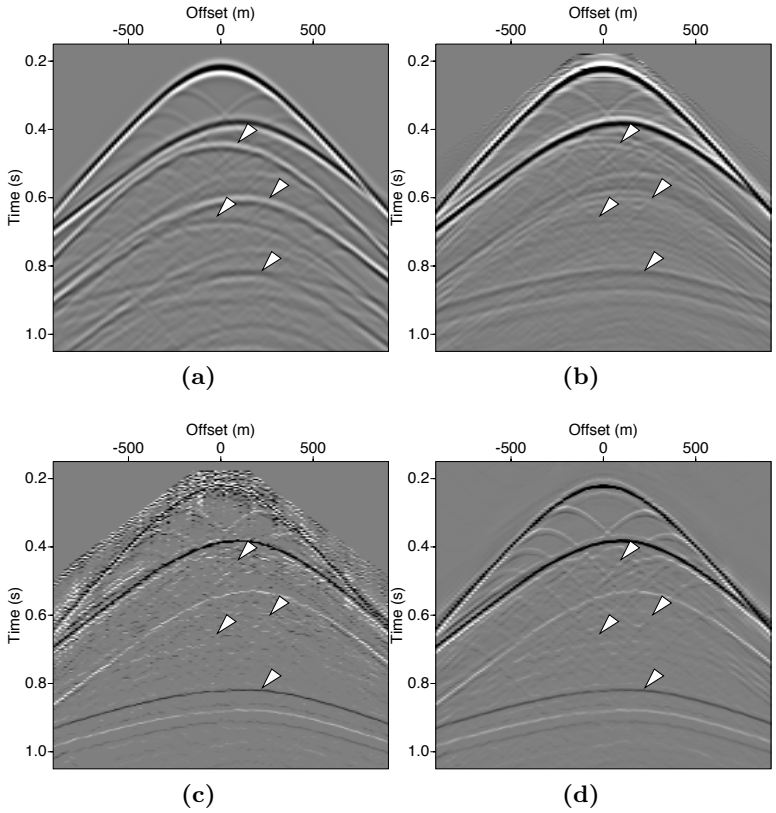



**Figure 3.5** Comparison of the surface-free Green's function obtained via different formulations of EPSI. **(a)** The input data simulated from a 2D marine model. Arrows indicate the largest surface multiples. **(b)** shows the result produced by the original EPSI outlined in Algorithm 3.2 without using the sparsifying operator $\mathcal{S}$ on updates to $\tilde{\mathbf{g}}$. When $\mathcal{S}$ is employed on all updates to $\tilde{\mathbf{g}}$, keeping the 4 largest events per trace per update, the result obtained is shown in **(c)**. The result produced by Robust EPSI outlined in Algorithm 3.3 is shown in **(d)**, using a roughly equivalent number of gradient updates compared to the two other solutions. The reflection events in $\mathbf{g}$ show characteristics of the expected dipole response. A comparison between **(b)** and **(c)** shows that sparsity regularization is important in producing correct estimations of the surface-free Green's function. The solution produced by Robust EPSI is cleaner and more free of artifacts compared to that of the original EPSI with use of the sparsifying operator.

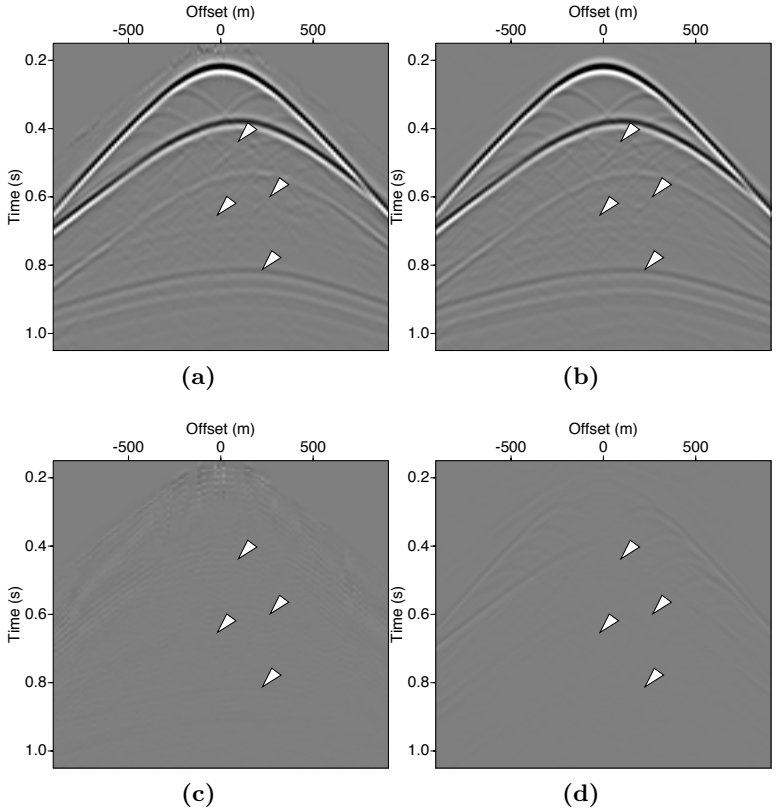
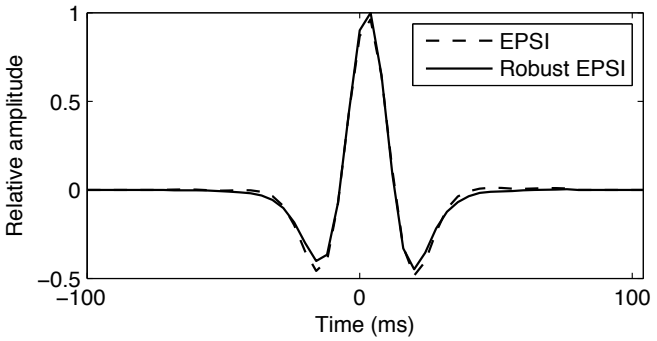


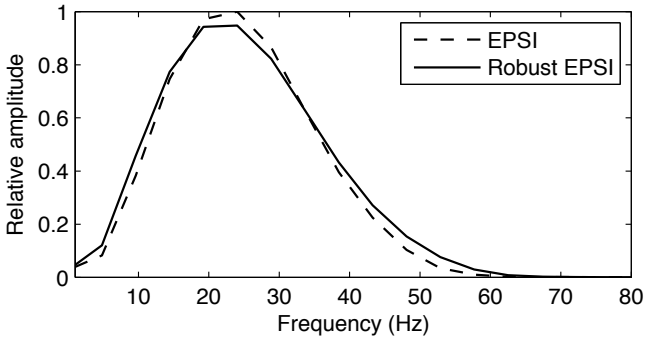

**Figure 3.6** The artifacts introduced by the ad-hoc sparsifying operator in the original EPSI algorithm (shown in Figure 3.5) are diminished but still present after convolution with the estimated signature $\tilde{\mathbf{q}}$, shown in **(a)** for the original EPSI and **(b)** for Robust EPSI. Residue wavefield is shown in **(c)** for the original EPSI and **(d)** for Robust EPSI.

parameters used compared to the previous example, and despite the large difference between the two datasets, Robust EPSI was able to produce a clean estimate of the surface-free Green's function while successfully rejecting the main surface multiples visible at 2 s and 3.8 s, indicated by arrows. The algorithm also showed good convergence with coherent energy barely visible in the total residue $\mathbf{p} - M(\tilde{\mathbf{g}}, \tilde{\mathbf{q}})$ shown in Figure 3.8d.

I also used the Robust EPSI variant that seeks the solution in a transform domain. An ideal candidate is the 3D curvelet frame for its ability to represent prestack wavefields with a high degree of sparsity. However,




(a)




(b)

**Figure 3.7** A comparison between the source-signature estimates produced by the original EPSI and the Robust EPSI approaches as **(a)** time signals, and **(b)** amplitude spectra. The original wavelet used is a Ricker wavelet of peak frequency 30 Hz. Both methods produce comparable models of the source wavelet.

I instead used a separable transform that is more amenable to scaling in a distributed implementation. It consists of 2D curvelets (with 5 scales and 16 angles) on time-slices of the wavefield to promote continuity across source-receiver coordinates. In the time domain I chose the basis associated with the discreet wavelet transform (DWT) constructed from a third-order Battle-Lemarié wavelet (Battle, 1987), due to its compact support in time and ability to sparsely represent seismic time-series. Notationally we can write this hybrid representation as $\mathbf{S}^H := \mathbf{C}^H \otimes \mathbf{W}^H$, where $\mathbf{C}^H$ is the synthesis operator of the 2D curvelet frame and $\mathbf{W}^H$ is the inverse DWT. This hybrid transform had been successfully used before in the sparse recovery of

simultaneously-acquired prestack seismic wavefields (Herrmann et al., 2009). I used this transform for all the examples in the chapter involving transform domains. The total gradient iteration counts were also kept the same as the physical domain experiments. The result can be seen in Figure 3.8c. This method produces a smoother Green's function with fewer continuity issues across traces compared to the examples with sparsity promotion in the physical domain, but some low-frequency artifact in the cross-dip direction can be seen. This may be addressed by a more careful selection of the transform used, or by scale and angle-dependent masking of the transform domain coefficients. This artifact is not present after convolution with the estimated source signature $\mathbf{q}$ (shown in Figure 3.9) as it is outside the dominant seismic signal band.

To assess the quality of the multiple removal achieved by Robust EPSI, I show in Figure 3.10 the full NMO stack of this dataset using the provided velocities. To avoid any potential issue caused by amplitude loss to the residue wavefield, the standard proposed method for EPSI is to produce a amplitude and phase-correct model of the surface multiple wavefield $M(\tilde{\mathbf{g}}, 0)$ from the estimated Green's function, so that it can be directly subtracted from the data. This is called the "conservative primary" result. Figure 3.10b shows the conservative primary produced by subtracting from Figure 3.10a the surface multiple model shown in Figure 3.10c, which is calculated from the Green's function obtained by Robust EPSI in Figure 3.8b. The figures results show good separation of primaries and surface multiples even in areas where the two completely overlap, as indicated by arrows. Note that this demultiple result was again obtained without any parameter tweaking.

### 3.5.2 Gulf of Suez marine data

As a final demonstration of Robust EPSI's capabilities, I look at a marine seismic line from the Gulf of Suez region, with a very shallow ocean-bottom causing strongly ringing surface multiples that permeates the entire dataset. The pre-processing included a time down-sampling from 2 ms to 4 ms, 2x shot spacing interpolation to coincide with receiver spacing, Radon domain near-offset interpolation, and extending via reciprocity to negative offsets when needed. Neither deghosting nor up-down decomposition was performed on the data except for an attempted removal of direct waves by muting. The data contains 361 shot records and 361 hydrophone traces at co-located positions. Robust EPSI was performed on this data using both the physical domain (Algorithm 3.3) and the transform domain variants.

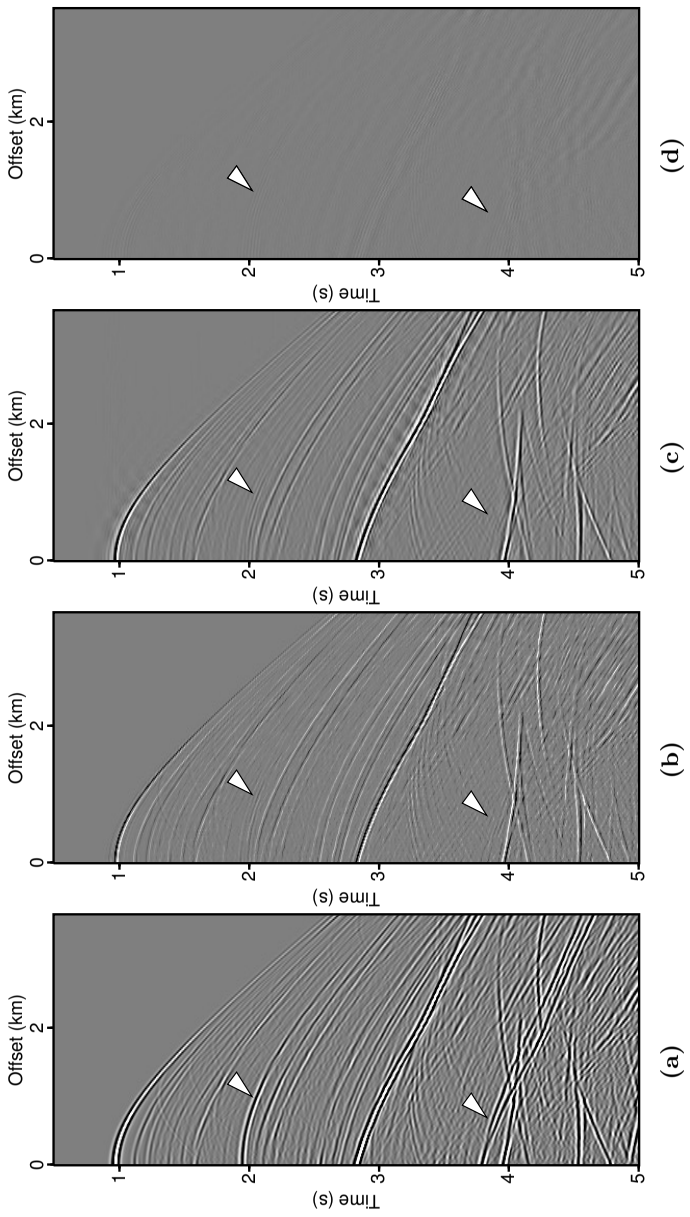
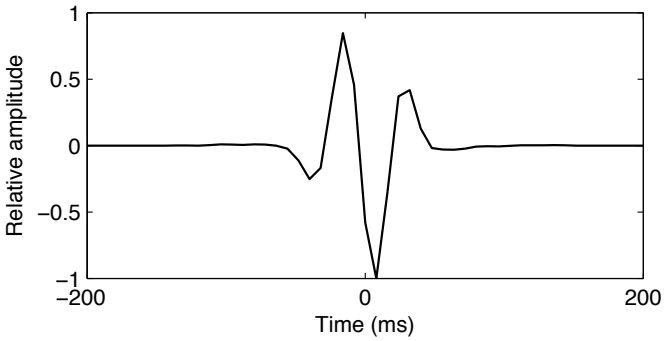
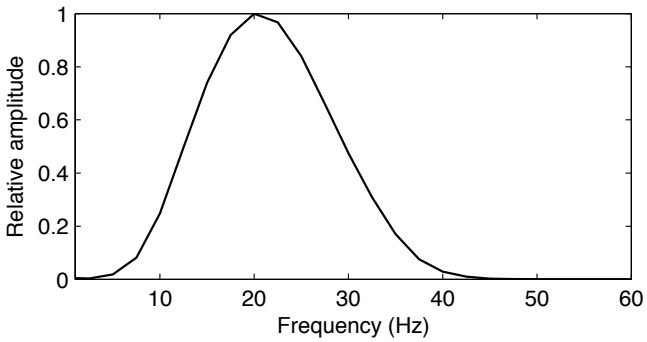



**Figure 3.8** Shot gather of Pluto 1.5 data is shown in **(a)**, while the surface-free Green's function produced by Robust EPSI after 80 gradient updates on the Green's function is shown in **(b)** for the physical domain solution and in **(c)** for the transform domain solution. The total residue from estimating the Green's function in (b) is shown in **(d)**. Parameters for the Robust EPSI algorithm remain unchanged from the ones used to produce Figure 3.5d. Without tweaking, Algorithm 3.3 produced a clean solution for the surface-free Green's function without the surface multiples at 2 s and 3.8 s (indicated by arrows).




**(a)**




**(b)**

**Figure 3.9** Estimate of the source signature $\tilde{\mathbf{q}}$ that was produced in the process of obtaining Figure 3.8b as part of Robust EPSI, plotted as **(a)** time signals, and **(b)** amplitude spectra. The original wavelet used is a Ricker wavelet of peak frequency 15 Hz. The estimated wavelet partially captured the stationary part of the receiver ghost that was not removed from the data prior to Robust EPSI.

I show shot gathers of the Green's function $\tilde{\mathbf{g}}$ produced by Robust EPSI, with and without use of the transform domain, in Figure 3.11. The ringing nature of the surface multiples make it hard to indicate by arrows, but they can be seen throughout the entire shot gather. With the same number of gradient updates we have a very evident improvement in reconstructing the later events past 1 s by using the transform domain. The direct Robust EPSI result only reproduces the strongest sections of the late events, while the transform domain solution is more successful in reconstructing them in their entirety. Our transform domain solution also produced a more

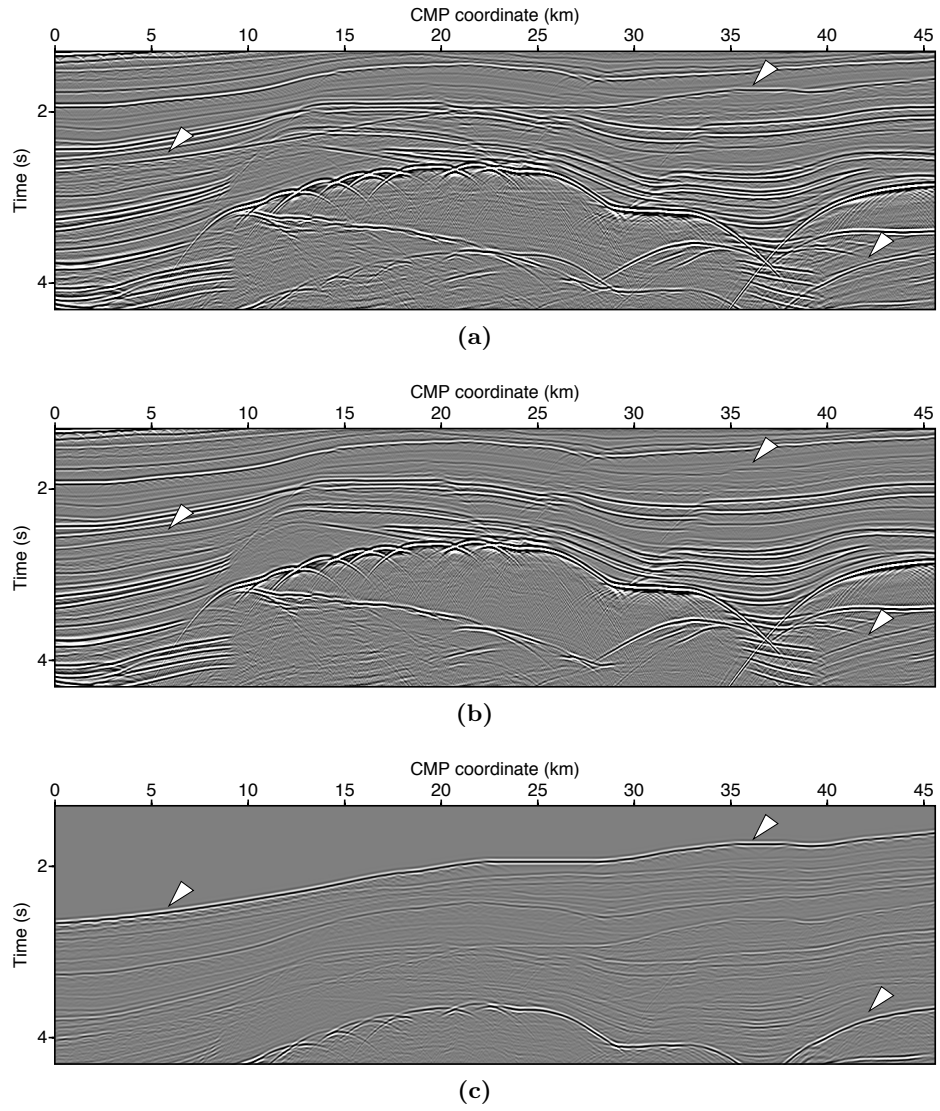



**Figure 3.10** Demonstration of surface multiple removal from Robust EPSI. Stacked section of Pluto 1.5 data is shown in **(a)**. The conservative primary result shown in **(b)** is produced by subtracting the surface multiple model calculated by Robust EPSI in **(c)** from (a). Arrows indicate where surface multiples completely overlap primary events, which are successfully recovered in the conservative primary result.

smoothed appearance, which is consistent with the results in Figures 3.1 and 3.8. The residue wavefield $\mathbf{p} - M(\tilde{\mathbf{g}}, \tilde{\mathbf{q}})$ shown in Figure 3.11d is from the Green's function shown in Figure 3.11b. Figure 3.12 shows the associated estimated source signature, upon which no constraint is imposed aside from a time window that coincides with the boundaries of the plot.

The conservative primary result (see previous section) is shown as a NMO stacked section in Figure 3.13b. The calculated model of surface multiples $M(\tilde{\mathbf{g}}, 0)$ subtracted from the data is shown in Figure 3.13c. The surface multiple model appears physically consistent with no sign of localized amplitude variation that can sometimes result if windowed adaptive subtraction was performed. Arrows indicate where primary events that are obfuscated by strong surface multiples is successfully recovered.

## 3.6 Discussion

### 3.6.1 Robust EPSI in practice

One of the main goals of this chapter is to eliminate most of the implicit free inversion parameters associated with the original EPSI algorithm. Minimizing parameter tweaking is especially important for this expensive algorithm to become successfully adapted in practical use.

Due to the ad-hoc sparse regularization used in EPSI, choosing a suitable sparsifying operator for the gradient updates is crucial. Tweaking of the following parameters are usually conducted the goal of allowing the gradient-sparsifying step in the original EPSI algorithm to build up into an approximation of the the sparsest solution:

- number of events per trace updated in each iteration
- exact arrival time of first reflection event
- size of update window on $\mathbf{g}$ for each iteration

We eliminated the need for all the above parameters in Robust EPSI, because the REPSI problem formulation explicitly makes finding the sparsest solution for a given misfit its goal.

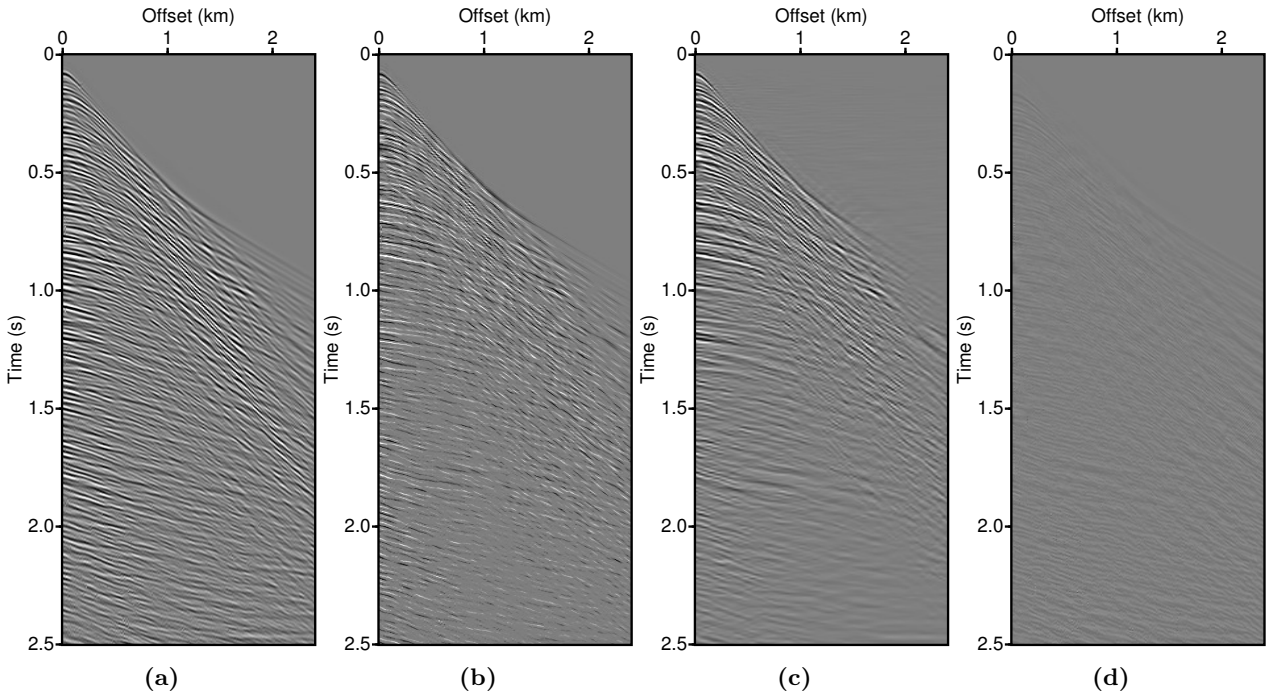
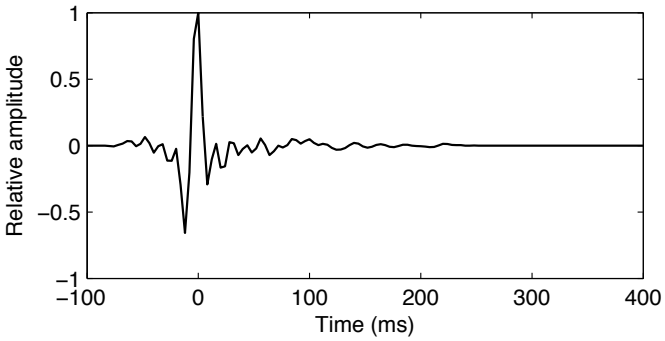
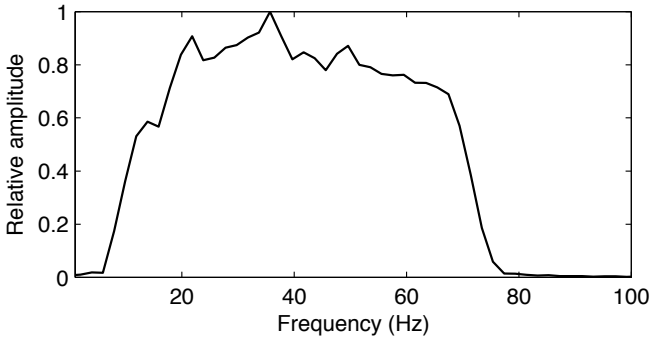



**Figure 3.11** Shot gather of Gulf of Suez data is shown in **(a)**, while the surface-free Green's function produced by Robust EPSI after 80 gradient updates on the Green's function is shown in **(b)** for the physical domain solution and in **(c)** for the transform domain solution. Reflection events in the Green's function show characteristics of the expected dipole response. The total residue from estimating the Green's function in (b) is shown in **(d)**. Parameters for the Robust EPSI algorithm remain unchanged from the ones used to produce Figure 3.8 aside from the length of time window for $\mathbf{q}$.




**(a)**




**(b)**

**Figure 3.12** Estimate of the source signature $\tilde{\mathbf{q}}$ that was produced in the process of obtaining Figure 3.11b as part of Robust EPSI, plotted as **(a)** time signal and **(b)** amplitude spectrum. Note that, as mentioned previously, this wavelet does not necessarily reflect the true physical source signature, and furthermore will capture all preprocessing on the data that be modelled by a global short-time filter, such as the low-cut filter what was applied before downsampling in time.

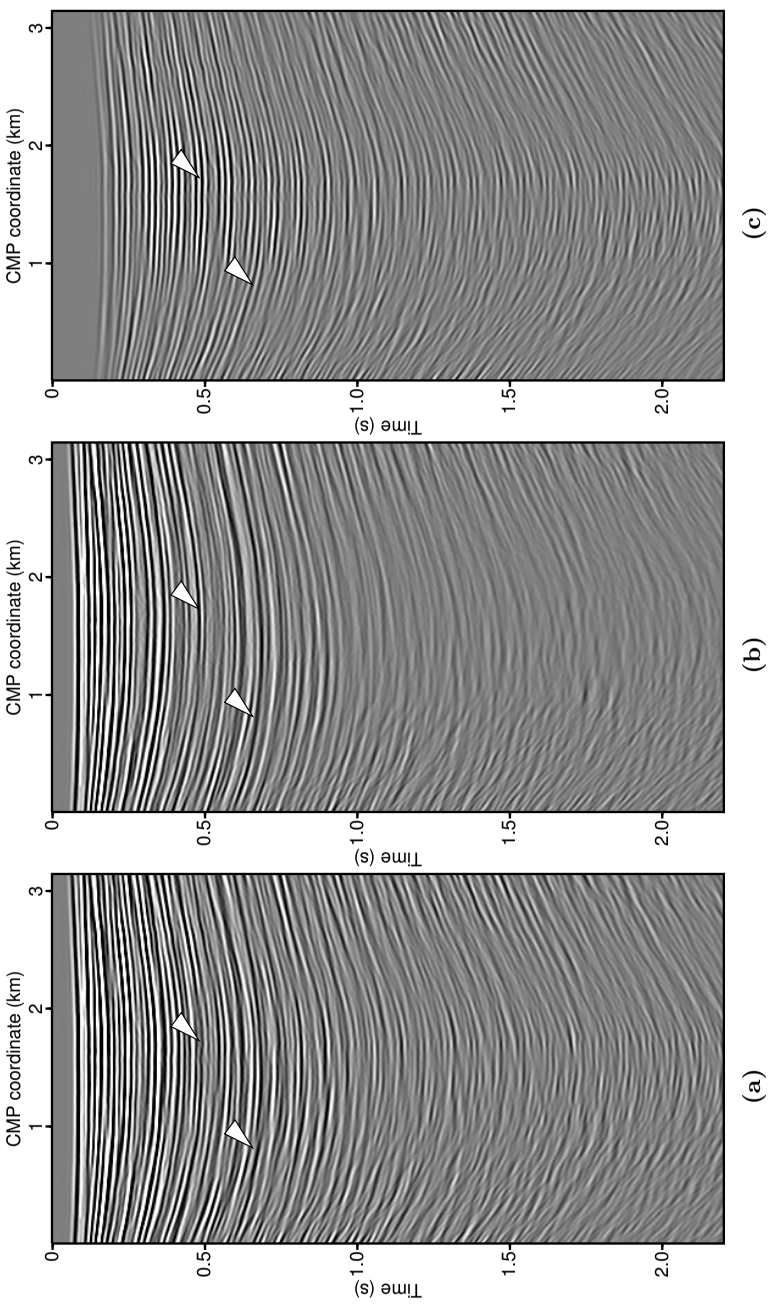



**Figure 3.13** Stacked section of marine data from the Gulf of Suez is shown in **(a)**. The conservative primary result shown in **(b)** is produced by subtracting the surface multiple model calculated by Robust EPSI in **(c)** from (a). Arrows indicate where surface multiples completely overlap primary events, which are successfully recovered in the conservative primary result.

There is still a need to decide how long the global wavelet model is initially allowed to be. But as previously discussed, this length can be relaxed once the danger of local minima (i.e., picking up periodicity with the wavelet when it should have instead been explained by the multiple model) can be ignored. Thus, the only important choice that remains is deciding when to terminate the algorithm.

Depending on the exact needs of the application, the stopping criterion issue may actually matter less compared to most inversion procedures. If multiple removal is the only desired outcome, then the estimated Green's function is actually not used. Instead, it is multidimensionally convolved with the data to produce a amplitude- and phase-correct multiple prediction, which is then simply subtracted from the data. This is called the "conservative solution" in van Groenestijn and Verschuur (2009b). In this case, failure to recover weak events in the $\tilde{\mathbf{g}}$ will not actually remove it from the final product, and therefore the number of iterations used in our examples will typically be sufficient for accurate multiple prediction. It is also worth noting that it is relatively easy to add more iterations for Robust EPSI if it were only terminated at the completion of individual Lasso problems.

If more accuracy in the estimated $\tilde{\mathbf{g}}$ is desired, then transform domain solutions have several inherent advantages over physical domain ones. From a sparse-recovery theory perspective, the ability to fully recover the true Green's function is very much related to its relative sparsity within its representation domain. As an example, curvelet representations of wavefields generally have a higher relative sparsity than physical domain wavefield. This explains why we consistently see improvements in the ability to recover weak signals using curvelets, despite using the same number of optimization iterations.

In addition, the choice of transform domains for Robust EPSI will also result in the rejection of signals that it cannot sparsely represent. Figure 3.1d shows an example of this; the primary events recovered using curvelets appear to be less noisy than any other results, because curvelets do not sparsely represent wide-band noise. In effect, we are getting a curvelet denoising, which operates on the same principle, "for free" as part of Robust EPSI. Perhaps the most important point to emphasize is that the transforms can be chosen to suit the needs of particular datasets and its unique characteristics without needing to change the algorithm itself. The only source of concern is whether sparsity in the transform domain will correspond in

some way to sparsity in the number of reflection events in time, which is still needed to regularize the inversion process itself.

The decision to use transform domains at all in practice will also be situational. Improvements in the direct solution may be a desirable goal, but they need to be assessed with the cost of applying the transforms to the whole dataset in mind. If the estimated Green's function is to simply be used for multiple prediction, and the data is already of high quality, then the additional uplift in Green's function accuracy might not be needed. Conversely, if the Green's function itself is of interest, for example in velocity or attribute analysis where a deconvolved wavefield is advantageous in providing better temporal resolution, then we might favor the use of transform domains at the expense of additional computation time.

### 3.6.2 Computational considerations

The majority of the cost involved in Robust EPSI is the same as that of the original EPSI program; both consist of gradient updates on both the Green's function and on the source signature. Computing the gradient on the Green's function requires a multidimensional cross-correlation of seismic wavefields, while evaluation of the forward modelling operator requires the multidimensional convolution (identical to the SRME prediction step). These are still considered costly operations, even though for most implementations they are carried out in parallel over frequency.

Nonetheless, the engineering task of making these operations computationally economical directly benefit from the widespread efforts in efficient implementations of SRME. In chapter 5, I will introduce a very general solution strategy inspired by multigrid methods from numerical linear algebra, which can somewhat alleviate this cost by exploiting the computational complexity of the multidimensional convolutions of the wavefield in reverse.

Furthermore, due to the nature of EPSI as a wavefield inversion, it is compelling to explore whether the algorithm is tolerant to using low-rank approximations of $\mathbf{p}$ both as part of the modelling operator $M$ and also as the observable. The data-structure of low-rank wavefield approximations can directly be exploited to reduce the cost of the wavefield convolutions and correlations that underlie the evaluation and gradient steps of EPSI. Provided that the low-rank approximation may be obtained at a low overhead cost compared to the overall cost of EPSI, this will greatly increase the computational tractability of EPSI, especially on large 3D datasets where

actual information of the reflection events are overwhelmed by the size of the ambient wavefield dimensions. Exploratory work on this topic had been done in Jumah and Herrmann (2011) with the low-rank approximations built up efficiently via random vector probes into $\mathbf{p}$, using a very limited number of passes through the full data with on-the-fly interpolation.

Another avenue to further reduce the computation cost of wavefield operations come from exploiting simultaneous acquisition. It was shown in van Groenestijn and Verschuur (2011) that the physics of EPSI can work directly on simultaneously acquired data, and that information from the multiple events can even aid in the separation of shot records. Because Robust EPSI uses the same $\ell_1$-norm minimization that had been previously proposed to directly separate simultaneous seismic data (Herrmann et al., 2009; Wason et al., 2011), there is reason to believe that Robust EPSI can recover its solution under a much denser simultaneous acquisition system compared to the method shown in van Groenestijn and Verschuur (2011), using the combined framework proposed in Lin and Herrmann (2009). This opens the possibility of directly using post-acquisition summing of seismic shot records (sometimes referred to as "shot-encoding") to reduce the effective size of $\mathbf{p}$ for the goal of further speeding up the wavefield convolutions and correlations.

The spectral projected gradient method we used (van den Berg and Friedlander, 2008, section 4.1) to solve the Lasso problems for the Green's function incurs negligible computation overhead compared to the large costs of computing the gradients. It mainly introduces a further fixed memory overhead (two to three times) due to having to store one previous solution of the Green's function and its updates, but is responsible for a major acceleration in convergence rate. Overall, I expect the slight increase in computation costs using Robust EPSI to be more than compensated for by the increased convergence rate and the quality of the resulting Green's function, as well as the reduced dependence on parameter search compared to the original EPSI algorithm.

### 3.6.3 Other future extensions

Other important existing extensions of EPSI include the ability to simultaneously estimate missing near-offset reflection data and the capability to account for source-arrays. With the Pareto root-finding approach, I believe there are also straightforward adaptations of these extensions. The near-

offset traces could, for example, be updated along with the source signature upon reaching the Pareto curve (however, in the next chapter I will develop a more reliable alternative to this kind of explicit estimation). The complete decoupling of the source signature estimation from the Green's function in Robust EPSI means we can also easily impose constraints from additional prior knowledge on the source without modification to the algorithm. Due to the optimization framework that it is derived from and the generic solvers it employs, I believe Robust EPSI presents itself as a much more flexible framework compared to the original EPSI formulation for future extensions of the method.

## 3.7 Summary

The stated goal of EPSI is to improve upon existing surface-multiple removal schemes by truly inverting a discrete operator that models the free surface. This method estimates the surface-free Green's function as well as the source-time function associated with the observed data.

My approach follows the sparsity argument of the original EPSI formulation but differs fundamentally in crucial algorithmic details. Instead of alternated enforcing of sparsity on updates of the Green's function and Fourier-domain smoothness on the source-time function, I formulate EPSI in terms of a bi-convex optimization problem. This approach allows us to tap into recent developments in large-scale optimization where sparsity-promoting problems are solved via carefully defined subproblems that allow components to enter into solution in a controllable way.

The main contribution of this work is to combine continuation methods, based on the Pareto trade-off curve, with the estimation of the source function after each subproblem has been solved. By relying on the fact that we solve each of these subproblems uniquely, we arrive at a reliable formulation where all but one of the parameters of the original EPSI have been eliminated. Aside from making the algorithm more robust and hence practical, our formulation has the additional advantage that it offers flexibility to use sparsifying transforms, such as curvelets, with very little effort compared to the original EPSI formulation.

Experimentally, I demonstrated the improved performance of the Robust EPSI algorithm on various synthetic and field datasets. Compared to the original EPSI approach, Robust EPSI produces fewer artifacts for its esti-

mates of the surface-free Green's function. By exploiting transform sparsity promotion, we accomplish these improvements without meticulous control on the sparsity of each update as required by the original EPSI. Because we follow a rigorous formulation and use techniques that are proven to solve $\ell_1$-norm minimization problems, our method produces high quality results on different datasets without parameter changes. While numerous performance related issues will still need to be addressed to extend Robust EPSI to 3D, the resilience of this method to parameter settings makes it highly suitable for large industry-scale multiple-removal problems.

# Chapter 4

# Mitigating acquisition gaps using scattering

One of the best known technical challenges of wave propagation-based multiple prediction is the requirement that wavefields need to be completely sampled over the whole survey domain. This is because primaries in the measured wavefield are used as the source functions for the multiple wavefield. Primaries that are not measured cause gaps in the subsequent ray paths that reflect back down from the free surface, and consequently can lead to serious errors in the final surface multiple prediction. This characteristic behaviour poses serious budget and technical challenges for survey design, and is one of the main motivations behind the the decades-long effort to improve near-offset acquisition, since near-offset primaries are known to be the strongest contributors of surface multiples. While it is now not uncommon for deep-water surveys to use a combination of cables in over/under configurations and off-azimuth recordings to reconstruct the ideal near-offset traces, this is still nearly impossible for shallow-water and on-shore surveys.

Therefore, current practice calls for intricate, on-the-fly trace interpolations, by inserting moveout-corrected versions of kinematically similar traces into non-sampled locations as a preprocessing step of the SRME multiple prediction. However, this paradigm is now being challenged by the increasing ubiquity of simultaneous or blended source acquisition designs, where overlapping wavefronts of conflicting dip and velocity are the norm. Alternative interpolation methods that are more immune to this situation but come at a higher computational cost are generally transform-based, such as

FK-domain reconstruction (sometimes with weighted matching-pursuit), the venerable parabolic Radon-based interpolation (Kabir and Verschuur, 1995), as well as more contemporary methods based on advanced transforms such as curvelets (Herrmann and Hennenfent, 2008; Hennenfent et al., 2010) or exploiting low-rank structures (Ma, 2013; Kreimer et al., 2013; Kumar et al., 2015). Unfortunately, these methods are all at their weakest when reconstructing parts of the wavefront with high curvature and amplitude variation, which is typically characteristic of near-offset traces in seismic data.

Some recent trends have developed where these data are explicitly reconstructed by looking at the multiple information. One such method is EPSI, which involves finding the correct surface-free Green's function that will unambiguously relate all wavefronts in the observed data under a primary-multiple relationship through the action of the free surface. In the original formulation of EPSI, it was proposed to include missing data as yet another piece of unknown that must be found to be consistent with this model, forcing the algorithm to alternate between estimating the surface-free (primary) Green's function, the missing data, and the source wavelet. While the initial results of this explicit reconstruction technique in EPSI have been successful, it fails to take advantage of the tightly-coupled relationship between the missing data and the surface-free Green's function, by ignoring their partial derivatives with respect to each other in the inversion process. My main contribution in this work is coming up with a formulation where the effects of missing data–i.e., a mask acting on the data matrix zeroing entries where data are missing–are explicitly accounted for in the forward model of EPSI through scattering terms at the surface boundary, modelled by repeated autoconvolution with the observed part of the wavefield.

This chapter is organized as follows. I will demonstrate the effect of data gaps in the near-offset on EPSI predictions and contrast it with its impact on SRME. Next, we discuss the method proposed by van Groenestijn and Verschuur (2009a), which extends EPSI to additionally invert for missing parts of the wavefield. I then contrast it with the proposed method in this chapter, where multiple contributions from the missing data are directly accounted for using scattering terms of the estimated model of the Green's function on just the observed part of the wavefield. These terms account for the impact of missing data on the prediction of surface-related multiples. The main difference of my method is that it allows formation of exact gradients for the full Green's function using correlating-type interferometry to explain data that we have, and Born scattering at the surface to emulate data that we don't have. I conclude by demonstrating the efficacy of this

method on both synthetic and field data sets, and make some remarks on the impact of this method on computational costs of REPSI.

# 4.1 Theory

## 4.1.1 Effects of incomplete data coverage

In the previous chapter, we have assumed that the observed wavefield $\mathbf{p}$ has complete coverage over all source-receiver coordinates. Let us now consider the case where $\mathbf{p}$ is spatially well-sampled enough for the multiple model in equation 2.3 to have no aliasing issues, but otherwise has significant gaps in its sampling coverage, such as a near-offset gap. To mathematically describe the effects of coverage issues in field acquisition geometries, I adhere to a notation used by van Groenestijn and Verschuur (2009a) that segregates all entries of $\mathbf{p}$ into actual sampled data $\mathbf{p}'$ and missing data $\mathbf{p}''$, both of equal dimensions to $\mathbf{p}$, such that $\mathbf{p} = \mathbf{p}' + \mathbf{p}''$. Remember that $\mathbf{p}$ is the hypothetical fully sampled (in terms of both source and receiver locations) data at a fixed, regular sampling grid. We can model the available data $\mathbf{p}'$ using a masking operator $\mathbf{K}$ that eliminates (by setting to zero) traces belonging to unrecorded source-receiver coordinates. Similarly, we can map to the missing data $\mathbf{p}''$ by the complement of this mask $\mathbf{K_c} := \mathbf{I} - \mathbf{K}$ (a stencil operator). This gives us

$$\mathbf{p} = \mathbf{p}' + \mathbf{p}'' = \mathbf{K}\mathbf{p} + \mathbf{K_c}\mathbf{p}.$$

The separation can also be done on the monochromatic data-matrices through a binary masking matrix $\mathbf{A}$, with the same dimensions as the data-matrices $\mathbf{G}$ and $\mathbf{P}$. The elements of $\mathbf{A}$ are defined as 1 at sampled positions and 0 otherwise. In this case applying the mask is equal to a Hadamard product $\mathbf{P}' = \mathbf{K}\mathbf{P} := \mathbf{A} \circ \mathbf{P}$. Likewise, the unacquired part of the data-matrix comes from applying the complement stencil $\mathbf{P}'' = \mathbf{K_c}\mathbf{P}$.

When the difference between $\mathbf{p}$ and $\mathbf{p}'$ is large, the incomplete coverage can become a significant source of error in the calculation of the surface multiples, especially at the near-offset locations (Verschuur, 2006). The most well-known example of this situation is missing near-offsets, as illustrated in Figure 4.1. The effects of these gaps on the multiple model are felt

throughout the wavefield volume due to the surface integral that underlies the prediction process, which results in a non-stationary spatial convolution.

The approximation $\mathbf{G} \approx \mathbf{P}$ assumed by SRME doubles the impact of missing data during acquisition. Comparing Figure 2.7d with Figures 4.1b and 4.1c, it is clear that using $\mathbf{P}'$ in place of $\mathbf{G}$ greatly exacerbates kinematic errors in the multiple prediction, resulting in completely unrecognizable wavefronts even when we have a near-offset gap as small as 100 m. Due to this effect, SRME usually requires some type of trace interpolation as a preprocessing step, since adaptive subtraction alone is not able to account for this degree of error.

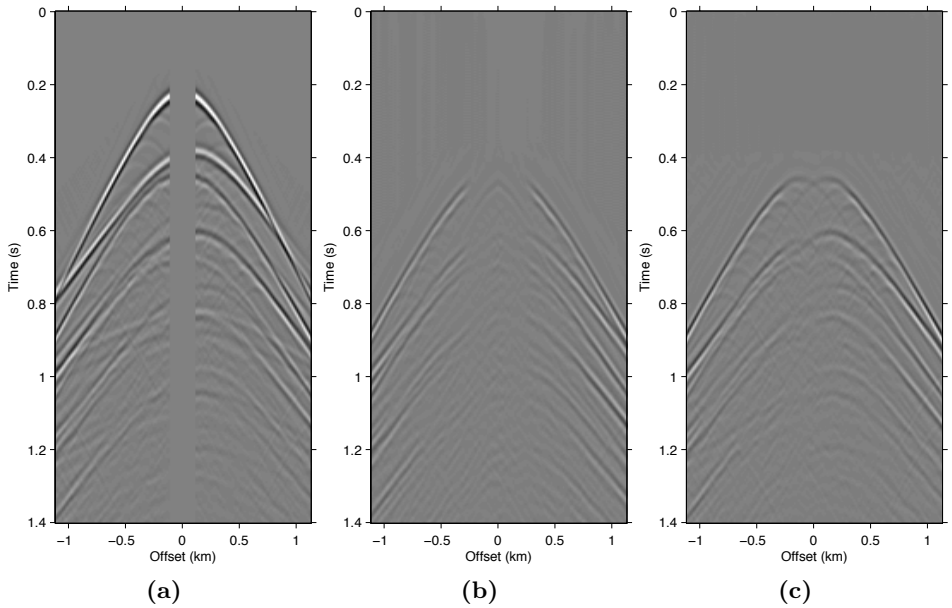
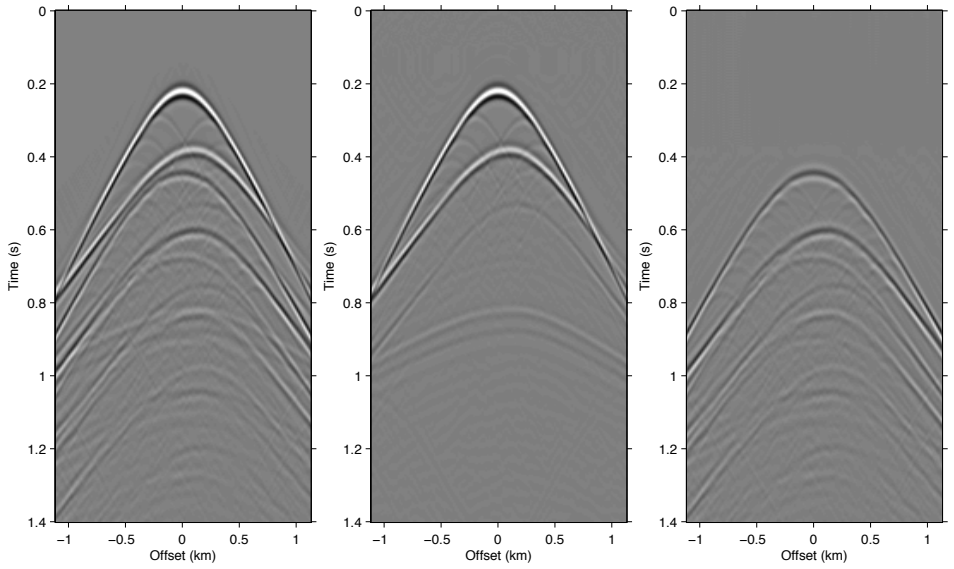



**Figure 4.1** The effects of a 100 m near-offset gap in the observed dataset $\mathbf{p}'$, which is shown in **(a)**, on both SRME and EPSI multiple predictions. **(b)** a shot gather of the SRME multiple prediction produced by auto-convolution of the observed data $\mathbf{P}'\mathbf{P}'$. **(c)** a shot gather of the EPSI multiple prediction produced by convolution with the correct, fully sampled, primary Green's function $\mathbf{G}\mathbf{P}'$. Despite knowing the true primary Green's function, the EPSI multiple prediction still contains errors caused by the missing near-offsets, although it is much more accurate when compared to the SRME prediction which is twice affected by the near-offset gap.




**(a)** Reference data **(b)** True primary **(c)** True multiple

**Figure 4.2** Reference wavefields to the incomplete versions shown in Figure 4.1. A fully recorded wavefield shown in **(a)** is directly comparable to figure 4.1a which shows the same data recorded with a near-offset gap. Once the true primary wavefield shown in **(b)** is taken out, the remaining true surface multiples can be compared to the incorrect models shown in figures 4.1b and 4.1c.

However, on the flip side, the non-local effect of the prediction error ultimately enables us to probe the (actions of the) free surface inside the data gap based solely on measurements outside of it. Since EPSI inherently relies on wavefield inversion, there are enough degrees of freedom in the process to take advantage of this effect. In their original report on EPSI, van Groenestijn and Verschuur (2009a) propose to explicitly reconstruct $\mathbf{p}''$ from intermediate estimates of $\mathbf{g}$, such that the accuracy of $M$ improves as the algorithm converges. Adapting this approach to the REPSI formulation means extending the definition of $M_\omega$ with the unobserved traces as another variable in the forward modelling operator

$$M_\omega(\mathbf{G}, \mathbf{Q}, \mathbf{P}''; \mathbf{P}') := \mathbf{GQ} + \mathbf{GR}(\mathbf{P}' + \mathbf{P}''), \tag{4.1}$$

which leads to a more complicated version of the optimization problem (2.4), with $\mathbf{p}''$ as an additional unknown:

$$\min_{\mathbf{g},\,\mathbf{q},\,\mathbf{p}''} \|\mathbf{g}\|_1 \quad \text{subject to} \quad f(\mathbf{g},\mathbf{q},\mathbf{p}'';\mathbf{p}') \leq \sigma, \tag{4.2}$$
$$f(\mathbf{g},\mathbf{q},\mathbf{p}'';\mathbf{p}') := \|\mathbf{p}' + \mathbf{p}'' - M(\mathbf{g},\mathbf{q},\mathbf{p}'';\mathbf{p}')\|_2.$$

A major issue with this method of independently updating $\mathbf{g}$ and $\mathbf{p}''$ in a cyclically alternating fashion is that we end up needing to include the intermediate and incorrect estimates of $\mathbf{p}''$ into the observation term of the misfit functional. Otherwise, the parts of $\mathbf{G}$ inside the unobserved locations will tend to converge very slowly, if at all, since its gradients would not be driven by the residuals that belong to the primaries, which only appear in $\mathbf{p}''$ inside the gap.

This approach also ignores the fact that $\mathbf{g}$ and $\mathbf{p}''$ are tightly coupled in an obvious way. The strongest parts of these two wavefields (the primaries) overlap and relate to one another through convolution with the wavelet. The full relation is implicitly defined by

$$\mathbf{P}'' = \mathbf{K_c P} = \mathbf{K_c}\big[\mathbf{GQ} + \mathbf{GRP}' + \mathbf{GRP}''\big], \tag{4.3}$$

where it is evident that $\mathbf{P}''$ can be almost completely characterized by $\mathbf{G}$. Therefore, we would expect that their mutual partial derivatives $\partial\mathbf{p}''/\partial\mathbf{g}$ and $\partial\mathbf{g}/\partial\mathbf{p}''$ should not be ignored when independently updating $\mathbf{p}''$ and $\mathbf{g}$. While $\partial\mathbf{p}''/\partial\mathbf{g}$ is straightforward to compute, the expression for $\partial\mathbf{g}/\partial\mathbf{p}''$ is, conversely, quite tricky to compute stably since it necessarily involves $\mathbf{Q}^{-1}$ (deconvolution by the unknown wavelet function $\mathbf{q}$). Motivated by this observation, in this chapter I seek to remove $\mathbf{p}''$ as an inversion variable altogether, and instead account for its multiple contribution with terms that only involve $\mathbf{g}$, $\mathbf{q}$, and the known data $\mathbf{p}'$.

### 4.1.2 Deterministic correction of surface multiple prediction by scattering terms

I start off eliminating $\mathbf{p}''$ from the EPSI modelling function (expression 4.1) by recursively substituting it with the right-hand side of expression 4.3. This results in a new modelling function $\widetilde{M}$ that does not depend on $\mathbf{p}''$, but at the cost of having infinitely many terms in a series expansion. In terms of

mono-frequency data matrices, $\widetilde{M}$ is written as

$$\begin{aligned}\widetilde{M_\omega}(\mathbf{G},\mathbf{Q};\mathbf{P}') = \mathbf{K}\big[&(\mathbf{GQ}+\mathbf{GRP}') \\ &+\ \mathbf{GRK_c}(\mathbf{GQ}+\mathbf{GRP}') && (4.4)\\ &+\ \mathbf{GRK_c}\left(\mathbf{GRK_c}(\mathbf{GQ}+\mathbf{GRP}')\right) && (4.5)\\ &+\ \mathcal{O}(\mathbf{G}^4)\big] \\ &:= \mathbf{K}\sum_{n=0}^{\infty}(\mathbf{GRK_c})^n\left(\mathbf{GQ}+\mathbf{GRP}'\right). && (4.6)\end{aligned}$$

The outermost masking operator $\mathbf{K}$ is needed to match the data $\mathbf{p}''$ only at the observed location, since every term in this expression produces wavefield responses over the entire domain. Note that this expression is no longer linear in $\mathbf{G}$. The $n$=1 term (4.4) is a second-order term since it is effectively a function of $\mathbf{G}^2$, and similarly the $n$=2 term (4.5) is a third-order term.

Since multiples that we observe in nature will decay in amplitude as a function of their order, we can assume from physical arguments that the spectral norm (i.e., largest singular value) of $\mathbf{G}$ is $\|\mathbf{G}\| < 1$, and therefore the infinite-term series expansion 4.6 converges to

$$\mathbf{KP} = \mathbf{K}\left(\mathbf{I}-\mathbf{GRK_c}\right)^{-1}\left(\mathbf{GQ}+\mathbf{GRP}'\right). \tag{4.7}$$

This expression states that we can recover the original EPSI forward model $\mathbf{P} = \mathbf{GQ}+\mathbf{GRP}'+\mathbf{GRP}''$ if it is somehow possible to invert the data mask $\mathbf{K}$ by some hypothetical exact data interpolation.

Although equation 4.7 is not practically useful, it admits a useful interpretation of my approach: that of a Born series solution to a Lippmann-Schwinger relation between the total wavefield $\mathbf{P}$ and an "incident wavefield" composed of the erroneous EPSI prediction $(\mathbf{GQ}+\mathbf{GRP}')$ from incomplete data. The perturbation in this case is the missing contribution coming from within the acquisition gap. This view makes it clear that each term in expression 4.6 is a term in the Born series expansion due to the free-surface reflector within the confines of the acquisition gap (as bound by the stencil $\mathbf{K_c}$). One direct consequence of this interpretation is that we can expect all orders of surface multiples in every term of expression 4.6, because the incident wavefield is not only composed of the primaries $\mathbf{GQ}$ but also parts of all subsequent orders of surface multiples, as produced by the term $\mathbf{GRP}'$.

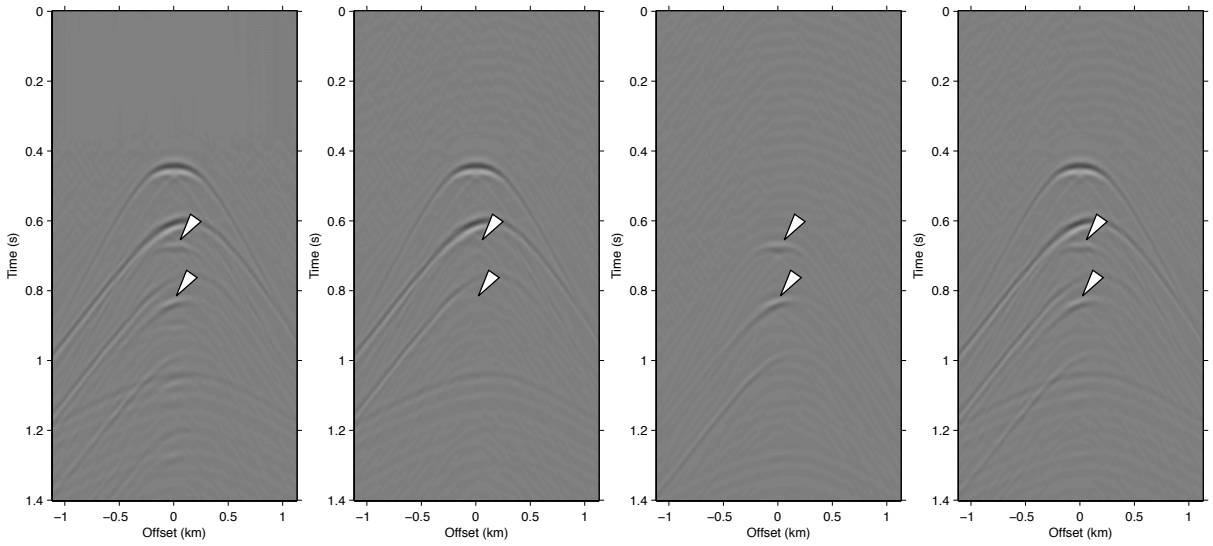
This scattering-based interpretation is perhaps most commonly seen in literature on the inverse scattering series (ISS) approach to multiple removal

(Weglein et al., 1997, 2003). However, unlike typical applications of ISS methodology, where it is used on subsurface reflectors for internal multiple removal, in this chapter we know exactly that the scattering boundary is at the free surface, and have only reflection effects from this scatterer. Perhaps more significantly, we use the derived multiples as part of the modelling function in the context of a full data-fitting inversion problem for the primary estimation, instead of an adaptive subtraction scheme.

As with all series expansion expressions, it is important to investigate how the approximation error behaves as a function of the number of terms used. Since computing these terms involves partial wavefield convolutions that have significant computational costs (although not as much as the full wavefield convolution, as we will discuss later in the chapter), it is practically desirable to limit the number of terms as much as possible. In the next section I claim what is perhaps the main insight of this chapter: that a very small number of these terms, even just two or three, can be enough to account for most of the missing multiple contributions.

### 4.1.3 Effects of term truncation on accuracy

As discussed in the previous section, the whole series up to the $n$-th term of expression 4.6 accounts for all of the missing $n$-th order multiples coming from the part of the free surface confined within the acquisition gap. It also includes some part of all multiples of $n$+1-st order and above, although these contributions may be much smaller. Figure 4.3 shows this with shot-gather representations of the different terms in expression 4.6 for the synthetic dataset shown in Figure 4.1a. We can see that the majority of these terms is composed of the $n$-th order surface multiple by comparing the wavefront shapes and traveltimes of panels 4.3b and 4.3c to the true primaries shown in Figure 4.2b. The arrows in panel 4.3b and 4.3c show two apices of the missing contribution of the second-order surface multiple, which we can see faintly in the $n$=1 term (equation 4.4) and very prominently the $n$=2 term (equation 4.5). The sum of these two terms, shown in panel 4.3d, perfectly accounts for the total missing contribution to the wavefront without relying on any data inside the gap.




**(a)** multiple contribution from $\mathbf{p}''$ (difference of 2.7d and 4.1c)

**(b)** $n$=1 term

**(c)** $n$=2 term

**(d)** sum of b and c

**Figure 4.3** Shot-gathers of various multiple contribution terms in the auto-convolution based forward modelling operator $\widetilde{M}(\mathbf{G},\mathbf{Q};\mathbf{P}')$ (shown in expression 4.6) when applied to the synthetic dataset with 100 m missing near-offsets in $\mathbf{p}''$ shown in Figure 4.1. **(a)** shows the total error in multiple prediction due to missing near-offsets. Panels **(b)** and **(c)** are respectively the first two terms of $\widetilde{M}(\mathbf{G},\mathbf{Q};\mathbf{P}')$ involving auto-convolutions with $\mathbf{g}$ (expressions 4.4 and 4.5). **(d)** shows the sum of panels (b) and (c). Comparing panels (a) and (d), it is evident that just the first two of the higher order terms of equation 4.6 are enough to model most of the significant EPSI multiple prediction errors due to the missing data. The arrows indicate the apices of the first and second order surface multiples of the ocean bottom reflection.

Figure 4.3 also visually demonstrates that the sum of just the first three terms of the series expansion 4.6 is enough to model most of the significant multiple contributions from the missing data $\mathbf{p}''$. Panel 4.3d, which shows the sum of the $n$=1 and $n$=2 terms, explains all the missing contributions from $\mathbf{p}''$ (panel 4.3a) up to the second-order surface multiple, and also contains some kinematic imprints of all higher order ones. Together, these two terms were able to bring the normalized prediction error (defined by $\|\mathbf{p}_m - \widetilde{M}(\mathbf{g},\mathbf{q};\mathbf{p}')\|/\|\mathbf{p}_m\|$, where $\mathbf{p}_m$ is the true surface multiple) down from ≈22% to ≈3%. Figure 4.4 summarizes the effectiveness of these first two terms as a function of the nearest available offset.

From the general trend of the curves in Figure 4.4 we can see that, for a range of typical near-offset gap sizes, just including the $n$=1 term can reduce the normalized multiple modelling error of $\widetilde{M}$ by more than one half (corresponding to a 6 dB improvement), and a further halving of the error is achieved by including up to the $n$=2 term. The model used here is the same one as used by van Groenestijn and Verschuur (2009a), with a gently-varying water column depth around 200 m throughout the model, and a water-bottom reflectivity coefficient of ≈0.41 (seafloor p-wave velocity 2.0 km/s, density 1.8 g/cm$^3$).

## 4.2 Algorithmic considerations

With the new augmented forward-modelling operator $\widetilde{M}(\mathbf{G},\mathbf{Q};\mathbf{P})$ as defined in equation 4.6 using data-matrix notation, we can now solve a new optimization problem by using it in place of the uncorrected modelling operator in the original REPSI problem (c.f. equation 4.2):

$$\min_{\mathbf{g},\,\mathbf{q}} \|\mathbf{g}\|_1 \quad \text{subject to} \quad f(\mathbf{g},\mathbf{q};\mathbf{p}') \le \sigma, \tag{4.8}$$
$$f(\mathbf{g},\mathbf{q};\mathbf{p}') := \|\mathbf{p}' - \widetilde{M}(\mathbf{g},\mathbf{q};\mathbf{p}')\|_2.$$

The main challenge that I will address in this section is that, compared to the original REPSI problem with fully-sampled data, this new one loses the linear relationship of the modelling function with $\mathbf{g}$. In the previous chapter, we relied on this bilinear property of $M(\mathbf{g},\mathbf{q};\mathbf{p}')$ in terms of $\mathbf{g}$ and $\mathbf{q}$ to solve problem (2.4) using an alternating coordinate-descent method. Below I first briefly recount how the original REPSI problem is solved, followed by the proposal of two possible ways to modify our solution strategy to account

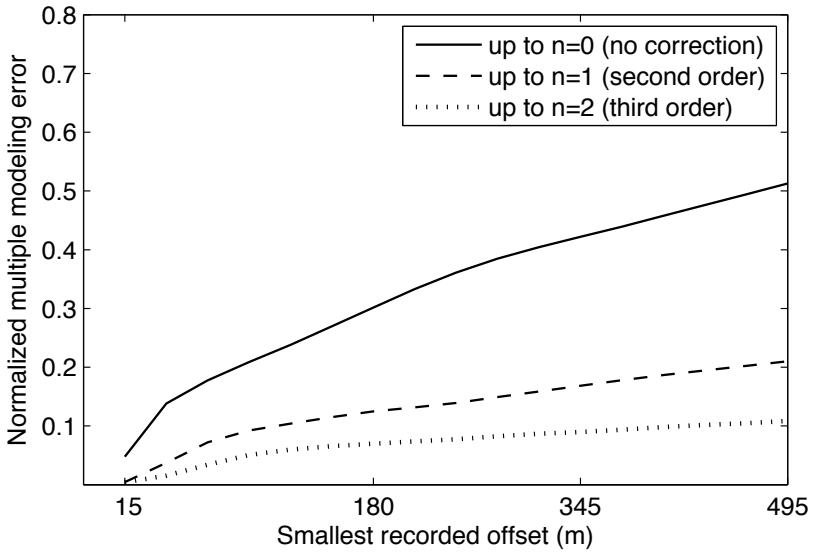



**Figure 4.4** Multiple modelling error for the nonlinear series expansion modelling operator (expression 4.6) as a function of term truncation and the offset gap in the data. The different lines plot this normalized modelling error from truncating all terms from expression 4.6 with $n$ higher than that indicated in the legend.

for the additional nonlinearity due to the existence of higher-order terms in $\widetilde{M}(\mathbf{G}, \mathbf{Q}; \mathbf{P}')$.

The existing approach to solving the original REPSI problem (2.4) is an alternating update of $\mathbf{g}$ and $\mathbf{q}$. Each new iteration of $\mathbf{g}$ is formed by a $\ell_1$-norm constrained problem (called a Lasso problem, Tibshirani, 1996) to give a sparse wavefront, which has a broadband, impulsive appearance in time. We then use it to estimate the wavelet $\mathbf{q}$ by matching the impulsive wavefronts against the data. The number of sparse wavefronts allowed into $\mathbf{g}$ is kept small at the early iterations to capture only the most significant primary events, quickly giving us the best estimate for the wavelet $\mathbf{q}$. Later iterations will then gradually increase the number of events allowed to better fit the remaining data residual.

This relaxation of sparsity constraint happens in a theoretically rigorous way for REPSI. The sparsity of $\mathbf{g}$ is directly influenced by the $\ell_1$-norm constraint value of the Lasso problems. The initial $\ell_1$-norm constraints

start at a small value that is designed to be an underestimate of the true minimum. It is then gradually increased throughout the iterations in a way such that it does not exceed the $\ell_1$-norm of the true solution, yet approaches it quickly. This scheme is enabled by the convex properties of the Pareto trade-off curve for Lasso problems (van den Berg and Friedlander, 2008; Hennenfent et al., 2008; Daubechies et al., 2008).

For the sake of exposition, I now roughly outline and number the main components of the Robust EPSI algorithm as described in the previous chapter (Algorithm 3.3), omitting some initialization details. We first set our solution candidates $\tilde{\mathbf{p}}_0$ and $\tilde{\mathbf{q}}_0$ to zero vectors, and the inversion residue (prediction error) $\mathbf{e}$ to $\mathbf{p}$. Then we iterate over the following steps with iteration counter $k$.

1. Select a new $\ell_1$ norm constraint $\tau_k > \tau_{k-1}$, which is evaluated through a closed-form expression from the current residue $\mathbf{e}_k$, the target misfit $\sigma$, and the $\ell_1$-norm of the current solution $\mathbf{g}_k$.

2. Obtain $\mathbf{g}_{k+1}$ by solving through a spectral projected gradient method (van den Berg and Friedlander, 2008; Birgin et al., 2000), using the previous solution $\mathbf{g}_k$ as the initial guess

$$\min_{\mathbf{g}} \|\mathbf{p} - M(\mathbf{g}, \mathbf{q}_k; \mathbf{p})\|_2 \text{ subject to } \|\mathbf{g}\|_1 < \tau_k. \tag{4.9}$$

3. Do an exact line-search scaling on $\mathbf{g}_{k+1}$ using the multiple term in $M(\mathbf{g}, \mathbf{q}_k; \mathbf{p})$ on $\mathbf{p}$ to minimize the effect of the $\ell_1$-norm constraint on the amplitude of $\mathbf{g}_{k+1}$.

4. Obtain $\mathbf{q}_{k+1}$ by solving a matched-filtering problem

$$\min_{\mathbf{q}} \|\mathbf{p} - M(\mathbf{g}_{k+1}, \mathbf{q}; \mathbf{p})\|_2$$

5. Form a new residue $\mathbf{e}_{k+1} = \mathbf{p} - M(\mathbf{g}_{k+1}, \mathbf{q}_{k+1}; \mathbf{p})$ and check for convergence conditions.

6. (If explicitly reconstructing data as in Problem 4.2) Solve for the missing part of the data $\mathbf{p''}_{k+1}$, using the previous solution $\mathbf{p''}_k$ as the initial guess

$$\min_{\mathbf{p''}} \|\mathbf{p'} + \mathbf{p''} - M(\mathbf{g}, \mathbf{q}_k; \mathbf{p'} + \mathbf{p''})\|_2$$

and form new data by $\mathbf{p} = \mathbf{p'} + \mathbf{p''}_{k+1}$.

For steps 1 and 2 we depend on the notion of a smooth, monotonic Pareto curve (sometimes known as the L-curve) for problem (2.4) to give us sensible $\ell_1$-norm constraint parameters $\tau_k$. The smooth and monotonic property is generally no longer valid when the forward system is nonlinear in $\mathbf{g}$, so below I will discuss two possible approaches to work around this issue, and propose how it can be applied to the above strategy to derive suitable methods for solving 4.8.

### 4.2.1 Modified Gauss-Newton approach

Several existing works on regularized inversion of auto-convolution functions rely on either the Gauss-Newton method or more generally the Levenberg-Marquardt method (Fleischer et al., 1999; Fleischer and Hofmann, 1996). In the same vein, I introduce an approach inspired by Li et al. (2012), which gives heuristically sparse solutions to the new nonlinear REPSI problem (4.8) using a modified Gauss-Newton method.

The crux of this method is to always ensure that the Green's function updates $\Delta\mathbf{g}$ are as sparse as possible for a given amount of decrease in the objective. This is achieved most effectively by taking as updates the solution to a Lasso problem formed around the Jacobian $\mathbf{J}_k$ of $\widetilde{M}$ evaluated at $(\mathbf{g}_k, \mathbf{q}_k)$:

$$\Delta\mathbf{g}_{k+1} = \underset{\mathbf{g}}{\operatorname{argmin}} \|\mathbf{e}_k - \mathbf{J}_k\mathbf{g}\|_2 \quad \text{s.t.} \quad \|\mathbf{g}\|_1 \leq \tau_k, \tag{4.10}$$

where $\mathbf{e}_k$ is the current nonlinear residue $\mathbf{p} - \widetilde{M}(\mathbf{g}_k, \mathbf{q}_k)$. The $\ell_1$-norm constraint $\tau_k$ in this case is again obtained deterministically by the optimization parameters of problem 4.10, using the same expressions used in step 1. However, since we are explicitly solving for updates $\Delta\mathbf{g}$ instead of $\mathbf{g}$ itself, the previous solution $\mathbf{g}_k$ (and hence its $\ell_1$-norm) for this linearized problem is always zero. Each update formed this way will thus tend to have a much smaller $\ell_1$-norm constraint $\tau_k$ compared to the final $\ell_1$-norm of the true solution, but the sum of all the updates may have a final norm that exceeds it. The assumption here is that our series of sparse updates will actually sum into a sparse final solution, which will depend on the Jacobians evolving slowly enough over the iterations, which implies that that all updates $\Delta\mathbf{g}$ should kinematically all consist of very similar wavefront sets. In practice, I find that this assumption is rarely a problem if the acquisition gaps are not

too large, and that we typically indeed usually end up with sparse solutions for $\mathbf{g}$.

The action of the Jacobian $\mathbf{J}_k$ is straightforward to obtain, but actually computing it will pose a challenge. If we explicitly write it out using the data-matrix notation, it becomes evident that the computational effort required escalates very quickly with the order of the series expansion:

$$\begin{aligned}\mathbf{J}_{k\omega}\mathbf{G} = \mathbf{K}\big[&(\mathbf{G}\mathbf{Q}_k + \mathbf{G}\mathbf{R}\mathbf{P}') \\ &+ \mathbf{G}_k\mathbf{R}\mathbf{K}_\mathbf{c}(\mathbf{G}\mathbf{Q}_k + \mathbf{G}\mathbf{R}\mathbf{P}') \\ &+ \mathbf{G}\mathbf{R}\mathbf{K}_\mathbf{c}(\mathbf{G}_k\mathbf{Q}_k + \mathbf{G}_k\mathbf{R}\mathbf{P}') \\ &+ \mathbf{G}_k\mathbf{R}\mathbf{K}_\mathbf{c}\left(\mathbf{G}_k\mathbf{R}\mathbf{K}_\mathbf{c}(\mathbf{G}\mathbf{Q}_k + \mathbf{G}\mathbf{R}\mathbf{P}')\right) \\ &+ \mathbf{G}\mathbf{R}\mathbf{K}_\mathbf{c}\left(\mathbf{G}_k\mathbf{R}\mathbf{K}_\mathbf{c}(\mathbf{G}_k\mathbf{Q}_k + \mathbf{G}_k\mathbf{R}\mathbf{P}')\right) \\ &+ \mathbf{G}_k\mathbf{R}\mathbf{K}_\mathbf{c}\left(\mathbf{G}\mathbf{R}\mathbf{K}_\mathbf{c}(\mathbf{G}_k\mathbf{Q}_k + \mathbf{G}_k\mathbf{R}\mathbf{P}')\right) \\ &+ ...\big].\end{aligned} \tag{4.11}$$

Note that the above expression is only written up to the $n$=2 nonlinear term (expression 4.5) of $\widetilde{M}$. Comparing all the terms here to the original nonlinear expression 4.6, we can see that due to partial derivatives, each of the $\mathcal{O}(\mathbf{G}^n)$ terms expand into $n$ separate terms in the Jacobian. In aggregate, this effectively introduces a quadratic complexity factor in the number of nonlinear terms $n$ over the usual cost of the gradient. Although some reuse is possible in the calculation of these terms, this approach may nonetheless present a prohibitive amount of computation overhead compared to the original problem. In the next subsection I will discuss an alternative scheme that forgoes using the Jacobian and will instead linearize the problem in some other way.

To summarize the modified Gauss-Newton approach to solving problem 4.8, we essentially need to modify steps 2 and 3 of the original algorithm above. In step 2 of each iteration we explicitly compute the updates $\mathbf{g}_{k+1} = \mathbf{g}_k + \Delta\mathbf{g}_{k+1}$ with $\Delta\mathbf{g}_{k+1}$ given by solving problem 4.10. In step 3 we then perform the exact line-search on the update $\Delta\mathbf{g}_{k+1}$ itself against the current residual. Much like in the original algorithm, this mitigates the amplitude loss from the $\ell_1$-norm regularization while keeping its sparsifying effect.

### 4.2.2 Relinearization by substitution

As shown in expression 4.11, having to compute the action of the Jacobian of the operator $\widetilde{M}$ may be unacceptably costly, considering that each wavefield multiplication in these expressions is computationally equivalent to one full SRME multiple prediction step. Since the non-linearity only comes from essentially having quadratic and higher powers of $\mathbf{G}$, a possible alternative to avoid dealing with the Jacobian is to simply substitute the previous iteration $\mathbf{g}_k$ as a constant in the $(\mathbf{GRK_c})^n$ part of equation 4.6. This returns us to the paradigm of inverting a linear operator in $\mathbf{G}$ at step 2, which is written in data-operator notation as

$$\begin{aligned}\widetilde{M}_{k\omega}(\mathbf{G}) = \mathbf{K}\big[&(\mathbf{GQ}_k + \mathbf{GRP}') \\ &+ \mathbf{G}_k\mathbf{RK_c}(\mathbf{GQ}_k + \mathbf{GRP}') \\ &+ \mathbf{G}_k\mathbf{RK_c}\left(\mathbf{G}_k\mathbf{RK_c}(\mathbf{GQ}_k + \mathbf{GRP}')\right) \\ &+ ...\big].\end{aligned} \tag{4.12}$$

Similar to Gauss-Netwon, the approach also relinearizes the problem at step 2, except that we do not have to find each update explicitly. Instead, we simply replicate step 2 of the original REPSI algorithm (solving the Lasso problem in equation 4.9 starting from the previous solution $\mathbf{g}_k$) using the approximate but linear modelling operator of expression 4.12. In effect, we now solve the following in step 2:

$$\min_{\mathbf{g}} \|\mathbf{p} - \widetilde{M}_k(\mathbf{g})\|_2 \text{ subject to } \|\mathbf{g}\|_1 < \tau_k. \tag{4.13}$$

Theoretically speaking, we have to give up some nice properties of Gauss-Newton with this method, such as possible quadratic convergence and guarantees of reaching a stationary point. But in practice I find this method, which I will call "relinearization" for the purpose of this chapter, converges similarly to the Gauss-Newton approach while also having much smaller per-iteration compute costs. In fact, the computational overhead here is marginal compared to the original algorithm, as I will explain in the Discussion section of this chapter. However, to achieve this result I find that it is necessary to manually precondition the first few gradients, where we have zero contribution coming from the missing primary part of $\mathbf{p}''$.

### Improving convergence by gradient trace weighting

If we explicitly write out a (negative) gradient step of the Lasso problem in step 2 (where the residue is at $\mathbf{e}_k$) in the data-matrix notation, using the approximate operator in expression 4.12 up to the $n$=1 term, we obtain, assuming $\mathbf{R} = -\mathbf{I}$ and using $H$ for Hermitian adjoint,

$$(\mathbf{K}\mathbf{E}_k)\mathbf{Q}_k^H - (\mathbf{K}\mathbf{E}_k)\mathbf{P}'^H - \mathbf{K}_\mathbf{c}\mathbf{G}_k^H(\mathbf{K}\mathbf{E}_k)\mathbf{Q}_k^H + \mathbf{K}_\mathbf{c}\mathbf{G}_k^H(\mathbf{K}\mathbf{E}_k)\mathbf{P}'^H.$$

Recall that $\mathbf{E}_k$ is the data-matrix representation of the current residue error $\mathbf{e}_k$. From this we can see that $\mathbf{g}$ is not being updated equally inside the acquisition mask compared to the outside of it. In fact, the first term in this whole expression is actually zero inside the acquisition mask (i.e., $\mathbf{K}_\mathbf{c}\mathbf{K} = 0$ so $\mathbf{K}_\mathbf{c}(\mathbf{K}\mathbf{E}_k)\mathbf{Q}_k^H = \mathbf{0}$), due to the residue being zero inside the mask, and the fact that the source wavelet applies a purely time domain convolution which does not spatially "spread" the wavefield. Since this term contains all the contribution from primary wavefields, which tends to be the strongest part of the wavefield, there is immediately a large loss of amplitude for the updates inside the mask. The job of filling this void then naturally falls to the subsequent terms, which all involve some sort of wavefield convolution and can potentially generate events inside the mask. However, in the first few gradients where $\mathbf{G}_k$ is small in amplitude (due to the $\ell_1$-norm constraints), the third term and above are also unhelpfully close to zero. This creates a convergence "lag" where traces of $\mathbf{g}$ outside of the mask must already be well reconstructed in order for the inside of the mask to begin updating, which is problematic because the outside traces are susceptible to overfit on multiples that should ultimately be explained by contributions coming from the inside of the mask.

We can mitigate the inherent imbalance on the early gradients through a simple weighting applied to all traces inside the acquisition gap. This can be seen as applying a diagonal right preconditioner to the linearized modelling operator. For the Gauss-Newton approach, this kind of scaling is automatically handled by the implicit Hessian, but for the relinearization method we must form and apply this preconditioner manually. An averaged, trace-independent version can be obtained by finding an exact line-search for a scaling $\alpha$ of $\mathbf{K}_\mathbf{c}\mathbf{g}_{k+1}$ which maximally reduces the residual left over from the contributions of $\mathbf{K}\mathbf{g}_{k+1}$ under the action of the next forward operator $\widetilde{M}_{k+1}$ (not that it is linearized around the newly obtained $\mathbf{g}_{k+1}$). In other

words, we seek the scalar $\alpha$, which solves the following problem

$$\min_{\alpha} \|\big(\mathbf{p}' - \mathbf{K}\widetilde{M}_{k+1}(\mathbf{K}\mathbf{g}_{k+1})\big) - \alpha\mathbf{K}\widetilde{M}_{k+1}(\mathbf{K}_\mathbf{c}\mathbf{g}_{k+1})\|_2^2.$$

The computation of this scaling can be effectively free in the process of computing the residual of iteration $k$+1, by producing the new predicted wavefield due to the two parts of $\mathbf{g}_{k+1}$ separately:

$$\begin{aligned}
\mathbf{e}_{\text{outside}} &= \mathbf{p}' - \mathbf{K}\widetilde{M}_{k+1}(\mathbf{K}\mathbf{g}_{k+1}) \\
\tilde{\mathbf{p}}_{\text{inside}} &= \mathbf{K}\widetilde{M}_{k+1}(\mathbf{K}_\mathbf{c}\mathbf{g}_{k+1}) \\
\alpha_{k+1} &= \tilde{\mathbf{p}}_{\text{inside}}^T\mathbf{e}_{\text{outside}}/\|\tilde{\mathbf{p}}_{\text{inside}}\|_2^2.
\end{aligned} \tag{4.14}$$

Note that all the computational cost for this scaling is in the last line, which only involves the vector inner products between $\tilde{\mathbf{p}}_{\text{inside}}$ and $\mathbf{e}_{\text{outside}}$, as well as $\ell_2$ norms. All the terms needed to form these two wavefields are already computed once we have the final residue for this iteration. As an aside, this scaling should be calculated after the line-search over all of $\mathbf{g}$ described in step 3, to minimize the impact of the accuracy of the current wavelet estimate (simply ignoring the primary $\mathbf{GQ}$ term when forming $\tilde{\mathbf{p}}_{\text{inside}}$ and $\mathbf{e}_{\text{outside}}$ will have a similar effect).

After obtaining $\alpha$, we can use it as a right-preconditioner $\alpha\mathbf{K}_\mathbf{c}$ in step 2 of the next iteration, where we now solve

$$\min_{\tilde{\mathbf{g}}} \|\mathbf{p} - \widetilde{M}_k(\alpha\mathbf{K}_\mathbf{c}\tilde{\mathbf{g}})\|_2 \text{ subject to } \|\tilde{\mathbf{g}}\|_1 < \tau_k. \tag{4.15}$$

This will effectively scale all gradient traces inside the stencil by $\alpha$. After the inversion is complete, we obtain the new estimate of the Green's function by $\mathbf{g}_{k+1} = \alpha\mathbf{K}_\mathbf{c}\tilde{\mathbf{g}}$, which removes the imprint of the right preconditioner. As the iteration count increases and $\mathbf{g}$ becomes more accurate, we should see $\alpha$ trend towards 1.

Figure 4.5 demonstrates the effectiveness of this method. Panel 4.5a shows the first gradient update for $\mathbf{g}$ (assuming we use the true wavelet $\mathbf{q}$) we would have obtained from a fully sampled data (shown in Figure 2.7a). Panel 4.5b shows the same gradient step that we would obtain using data with missing near-offsets (figure 4.1a). Evidently, this update contains mostly correct kinematic information about the wavefield, but is much weaker in amplitude for traces inside the near-offset gap. After applying the preconditioner $\alpha\mathbf{K}_\mathbf{c}$, with $\alpha$ computed using the inaccurate gradient

in Panel 4.5b as the approximate solution, we see in Panel 4.5c that this amplitude imbalance is effectively mitigated.

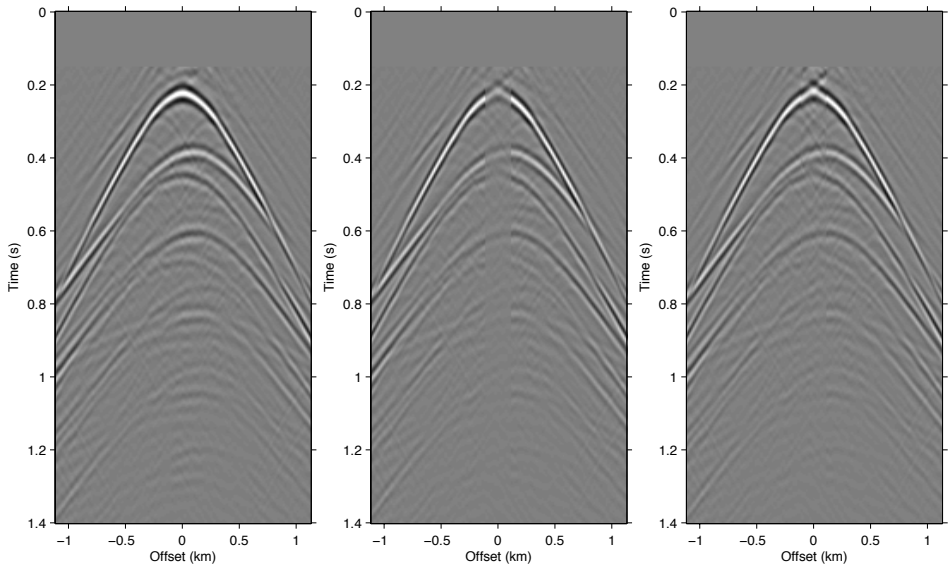



**Figure 4.5** Early gradient updates for $\mathbf{g}$ need to be preconditioned, otherwise convergence will be slow inside the acquisition mask due to amplitude imbalances. **(a)** the exact first gradient for $\mathbf{g}$ from fully sampled data, given exact $\mathbf{q}$. **(b)** the first gradient from data with missing near-offsets (shown in Figure 4.1a) using the relinearized forward model $\tilde{M}_\omega$. Inside the near-offset gap, the events are correct but greatly diminished in amplitude. **(c)** the gradient shown in (b) after applying the preconditioning scaling described. The average trace-independent scaling is effective in balancing the amplitude of the updates between the inside and the outside of the near-offset gap.

To summarize, in the "relinearization" approach we essentially need to replace the steps 2 and 3 of the original algorithm with the following modified steps:

2. Solve the modified Lasso problem (4.15) instead of the original one (4.9). This replaces the wavefield modelling function with the approximately linearized version of $\widetilde{M}$, defined in expression 4.12, and adds a right preconditioner $\alpha_k \mathbf{K_c}$. Once solution is obtained, apply $\alpha_k \mathbf{K_c}$ to invert the preconditioner and get $\mathbf{g}_{k+1}$.

3. Compute the scaling for all of $\mathbf{g}_{k+1}$ as in the original. After applying the scaling, compute the next preconditioner scaling $\alpha_{k+1}$ using expression 4.14.

## 4.3 Numerical examples

In this section I demonstrate the efficacy of my proposed methods, focusing on the missing near-offset problem. I will compare the results given by both the Gauss-Newton and the relinearization approach, using up to the $n = 2$ term (expression 4.5) in the augmented nonlinear modelling operator $\widetilde{M}(\mathbf{g}, \mathbf{q}; \mathbf{p}')$. I also compare my results to the explicit data reconstruction approach (problem 4.2) based on the method suggested by van Groenestijn and Verschuur (2009a), as well as a parabolic Radon-domain interpolation done as a preprocessing step. I will first show results using the synthetic data shown in 2.7a, then move on to a field data example from the North Sea. Although I only show 2D acquisition examples, note that this method makes no assumptions on the coordinate system used for the data matrices, so the derivations are extendable to a 3D survey geometry.

### 4.3.1 Synthetic data example

In this section I focus on how effectively the proposed methods in this chapter handle different sizes of near-offset gaps, and how they fare compared to explicitly inverting for the missing data $\mathbf{p}''$. I use the same 5 km fixed-spread synthetic data used in all the figures so far in the chapter, a shot gather of which is shown in Figure 2.7a. For reference, the corresponding true primary model is shown in Figure 4.2b. This model appears in almost all existing EPSI literature today and serves as an effective benchmark.

Figure 4.6 shows a comparison of methods for a 45 m near-offset gap (in both positive and negative offsets). This is an extremely mild offset gap and typically poses no major problems for SRME workflows. In this case, the data reconstruction (panel 4.6b), the Gauss-Newton (panels 4.6c and 4.6d), and the relinearization (panel 4.6e) strategies all performed very similarly and gave good results without needing to interpolate the near-offsets in preprocessing. All the results plotted here are the direct primary estimation (the final $\mathbf{g}$ from the inversion convolved with the final $\mathbf{q}$ from the same inversion), as the effects of reconstruction inside the offset gap are

clearer this way. In practical applications, usually a multiple model would be formed with the solution $\mathbf{g}$ instead, so it can be subtracted from the data under perhaps more scrutiny.

Comparing the Gauss-Newton results allows us to examine the effects of term truncation in the nonlinear modelling operator $\widetilde{M}$. Panel 4.6c shows the Gauss-Newton solution using up to the $n$=1 term (second-order in $\mathbf{g}$) in $\widetilde{M}$. As discussed in the theoretical section, this term alone cannot fully account for the second-order surface multiples, and thus some part of the first order multiple is put into $\mathbf{g}$ by the algorithm to make up some of the difference. This is especially apparent inside the offset gap, where $\mathbf{g}$ is less constrained by the observed data.

Panel 4.6d shows the solution we can obtain by just using one more term in $\widetilde{M}$, up to the $n$=2 term which is third-order in $\mathbf{g}$. We see that the imprint from the first-order multiple is much less severe in this result. This result, notably, is already very close to the solution from fully sampled data shown back in Figure 4.2b, even though only two additional terms from the series expansion is used. The results of the relinearization method (also using up to the $n$=2 term) shown in Panel 4.6e is almost identical to the Gauss-Newton solution, despite being much faster to compute. In fact it appears to suffer from less overfit of the remaining multiple prediction errors due to the the lack of higher order scattering terms.

Figure 4.7 shows the opposite end of the near-offset gap severity scale with a 225 m offset gap. As we can see in panel 4.7a, the majority of the near-offset wavefront curvature information is missing, which poses a serious challenge for standard workflow tools to accurately reconstruct near-offset traces. In this case the explicit data reconstruction (figure 4.7b) fails to remove some of the later surface multiples, such as those after 0.6 s. The augmented nonlinear modelling operator approaches introduced in this chapter arguably fared better in comparison. The Gauss-Newton approach up to the second-order term (figure 4.7c) is better at recovering the wavefronts inside the offset gap, even building up some of the diffraction events, but also similarly fails to remove some of the later surface multiples. The multiple removal is improved greatly by including up to the third-order term (figure 4.7d), although some multiples still remain, again possibly due to overfit. The relinearization method appears to be the overall winner in this case (panel 4.7e), producing the cleanest primary estimation that is nearly identical to the reference primary events outside of the near-offset gap, while also computing the result much faster than the Gauss-Newton method.

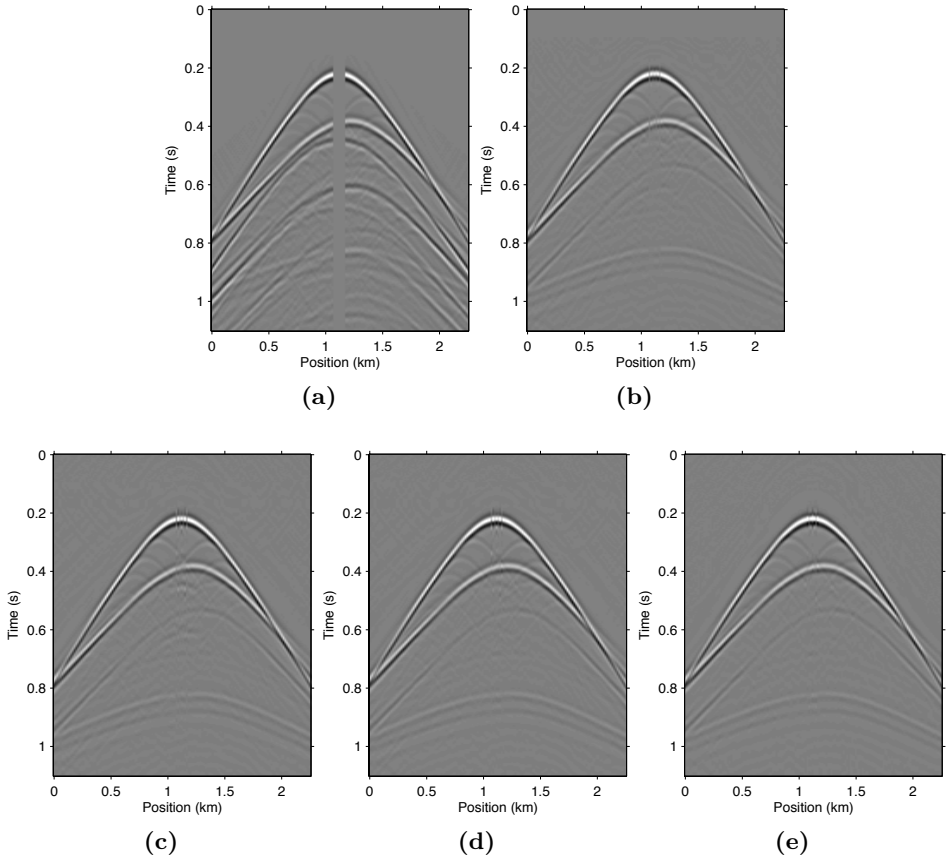



**Figure 4.6** Direct primary estimation results from a 5 km fixed-spread synthetic dataset with a nearest recorded offset of 45 m, for which a shot-gather at 2.5 km is shown in **(a)** up to 1 km offset. **(b)** the estimated primary using explicit data reconstruction (solving problem 4.2). **(c)** estimated primary using the Gauss-Newton method described in this chapter, using up to the $n$=1 term (second-order in $\mathbf{g}$) in $\widetilde{M}$. **(d)** the results obtained with the same method as (c), but including up to the $n$=2 term (third-order in $\mathbf{g}$). **(e)** estimated primary using the relinearization strategy, including up to the $n$=2 term. All the methods produce similar quality results for this small offset gap, although we can clearly see the benefit of using the $n$=2 term in the nonlinear forward model when comparing (c) and (d). The relinearization strategy produced slightly better results over the Gauss-Newton methods, despite also being computationally faster.

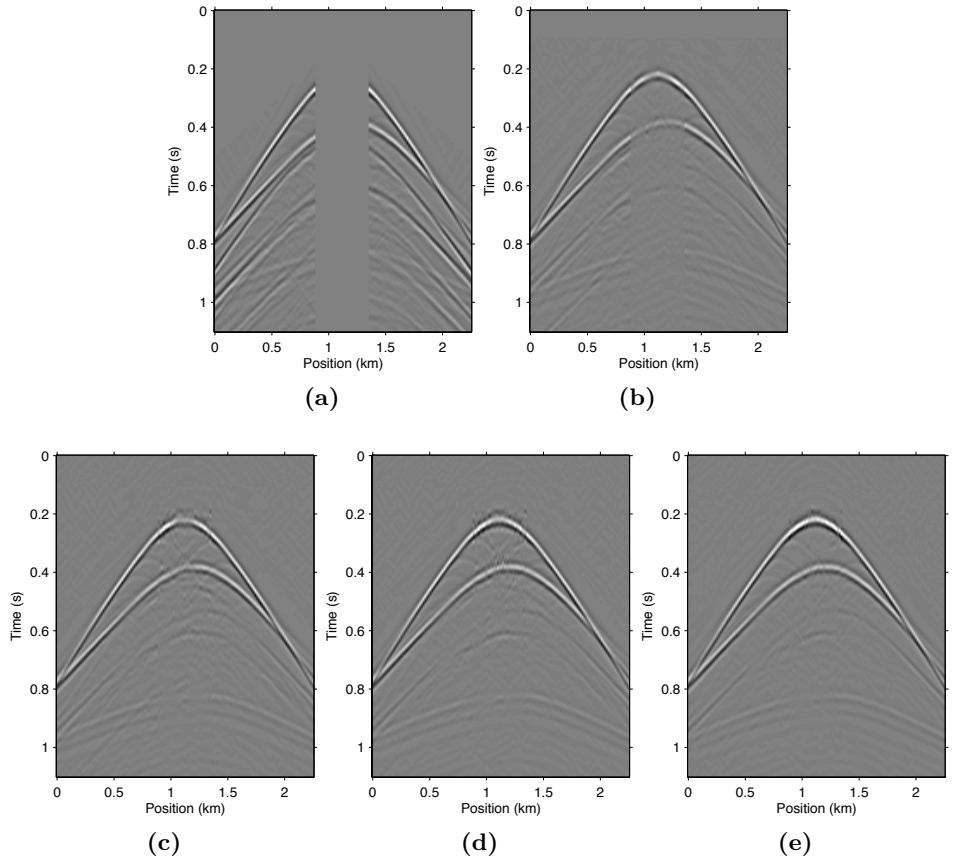



**Figure 4.7** Direct primary estimation results from a 5 km fixed-spread synthetic data with a nearest recorded offset of 225 m, for which a shot-gather is shown in **(a)**. The panels in this figure show the same strategies as the ones shown in Figure 4.6. **(b)** the estimated primary using explicit data reconstruction (solving problem 4.2). **(c)** estimated primary using the Gauss-Newton method described in this chapter, using up to the $n$=1 term (second-order in $\mathbf{g}$) in $\widetilde{M}$. **(d)** the results obtained with the same method as (c), but including up to the $n$=2 term (third-order in $\mathbf{g}$). **(e)** estimated primary using the relinearization strategy, including up to the $n$=2 term. The explicit data reconstruction result in (b) fails to reject some of the later surface multiples, while the methods introduced in this chapter manages to do a better job. As in Figure 4.6, we readily see the benefit of including up to the $n$=2 term, and also that the relinearization strategy manages to produce the cleanest result despite being faster to compute than Gauss-Newton.

### 4.3.2 Field data example

I now demonstrate the effectiveness of the methods proposed here using a North Sea shallow-water marine 2D seismic line with a 100 m near-offset gap. Pre-processing has been applied to convert the collected dual-sensor data to an upgoing pressure wavefield using the method described in Cambois et al. (2009). Data regularization and interpolation were carried out as part of the preprocessing, so in this example I will compare my method with the existing parabolic Radon interpolated near-offsets. An approximate 3D-to-2D correction factor $\sqrt{t}$ has also been applied after data regularization.

Figure 4.8 shows the moveout-corrected stacks of the field data experiments. Panel 4.8a shows the dataset itself. All near-offsets are excluded from the stack to keep comparisons consistent. Panel 4.8b shows the results obtain from using the original REPSI on the regularized data where the near-offset is pre-interpolated independent of REPSI using parabolic Radon-domain methods. Panel 4.8c shows results obtained by inverting the third order (keeping up to $n$=2 terms) nonlinear operator $\widetilde{M}(\mathbf{g},\mathbf{q};\mathbf{p}')$ using the relinearization method, while panel 4.8d shows the same using the Gauss-Newton method.

The methods presented here improve on the results obtained by using unmodified EPSI on pre-interpolation data. With shallow water data most pre-interpolation methods are known to be inaccurate, often under-estimating near-offset amplitudes and incorrectly reconstructing wavefront curvature at the apex. Comparing panel 4.8b with panels 4.8c and 4.8d shows that my methods exhibit a much improved ability to remove large water-bottom multiples that are evident at 0.25 s, 0.55s, and 0.8 s of the stack. This is especially evident by looking at the multiples removed from the pre-interpolation result (panel 4.8e) and comparing it to the multiples removed using the relinearization result (panel 4.8f). By discarding the pre-interpolated near-offset and using the augmented modelling operator $\widetilde{M}$ to account for its effects, we are able to get a much better defined multiple model with more accurate amplitude. Pre-interpolation also led to some ringing kinematic artifacts in the multiple model near the large diffraction events at 3 km and 1.0 s, which is not present in my methods.

Panel 4.8g shows a difference plot between inverting the second order (keeping up to $n$=2 terms) and the third order forward operator using relinearization. We see that the impact of using the additional term is already very small, and mainly consists of the imprint of second-order surface mul-

tiples. The comparison between relinearization and Gauss-Newton is more interesting (panel 4.8h). It is not very clear from the difference plot which method produced a better multiple model, although when carefully examining the primary stacks we see that the amplitude of the first water-bottom multiple (at 0.25 s) is slightly over-estimated for Gauss-Newton. This is possibly due to the over-fitting issues discussed earlier in the synthetic example section. Nevertheless, the relinearization method once again provided a very good result at a fraction of the compute time required for Gauss-Newton.

## 4.4 Discussion

The synthetic results shown in the previous section demonstrated that my proposed scattering-based correction to the multiple modelling with incomplete data can outperform the existing explicit data reconstruction approach used in the literature, especially when larges gaps are encountered in the near-offset. As far as I am aware, this proposed method is also the first practical formulation where a scattering series is rigorously used in a variational problem for primary estimation, with all its numerical approximations explicitly stated and explored. The real data example also demonstrates a significant improvement over using parabolic Radon interpolation for the near-offset. These improvements are tangible regardless of the algorithmic approach used to invert the augmented and non-linear modelling operator.

Perhaps more interestingly, I find that the Gauss-Newton method did not converge appreciably faster than the conceptually simpler and computationally cheaper relinearization method, despite its more solid theoretical convergence guarantees. This might be because the number of total gradient updates typically involved in REPSI is already fairly small (typically around 50), since the projected gradient method that I employ to solve the Lasso problems forms an implicit Hessian of the problem by using spectral step lengths (Birgin et al., 2000). The Gauss-Newton Hession might not have been able to improve much more on this scaling information. On the other hand, I find to my surprise that the relinearization strategy often produced better results. One possible explanation is that in some of the later iterations, the Gauss-Newton method is more prone to over-fitting $\mathbf{g}$, particularly inside the trace mask where it is less sensitive to data misfit, to account for some of the inherent modelling error due to term truncation.

Although I have only shown near-offset acquisition gap effects in this chapter, the derivations here are valid for missing data located anywhere on

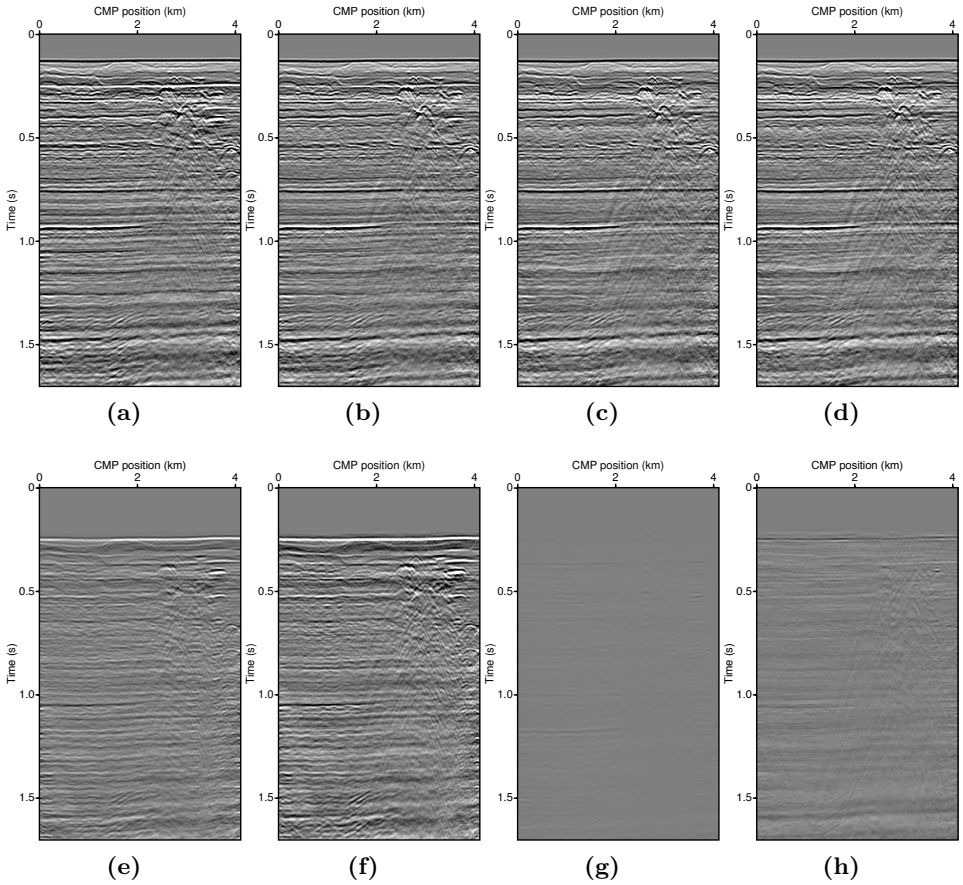



**Figure 4.8** Moveout-corrected stacks (near-offsets excluded) of the North Sea field data with 100 m near-offset gap after conservative multiple removal, with the multiple model generated from the Green's function obtained using the methods introduced in this chapter. **(a)** the recorded field data. **(b)** the results using parabolic Radon near-offset interpolation and the unmodified REPSI algorithm. **(c)** the results obtained by discarding the near-offset traces and using the relinearization method with up to the $n$=2 term in $\widetilde{M}$. **(d)** the same as (c) but using the Gauss-Newton method. **(e)** the final multiple model from the Radon interpolated data result shown in (b). **(f)** the final multiple model from the relinearization method result shown in (c). **(g)** difference plot between solving the relinearization problem using up to the $n$=1 term versus up to the $n$=2 term. **(h)** difference plot between the solutions obtained from the relinearization method and the Gauss-Newton method.

the acquisition geometry. I imposed no special structure on our diagonal masking operator $\mathbf{K}$, so in theory the missing traces can be anywhere. Near-offset traces are well-understood to be responsible for the majority of the constructively interfering part of the multiple contribution gather, so we conveniently use it as a worst-case location for the type of large contiguous detector gaps that are the focus of this chapter.

In theory, my scattering-based method to account for missing data works only when we can trace raypaths that reach the inside of the masked location from the observed data positions, using the estimated Green's function. This would imply that missing traces near the edge of the acquisition grid would be less effectively mitigated, which I do observe in practice. However, I find that the resulting edge effects tend to be overshadowed in magnitude by what would typically be produced in imaging algorithms.

Another important exception to the generality of my method is that it cannot deal with regular undersampling, even if it strictly counts as missing data in the overall framework. This is mainly because the mechanisms employed here also depends on wavefield interferometry, which is fundamentally erroneous when strong, coherent spatial aliasing is present in the data.

Finally, I note that one additional advantage of this approach over explicit data reconstruction is the possibility of exploiting any existing regularization schemes for $\mathbf{g}$, such as the curvelet-based estimation discussed in the previous chapter, although a thorough investigation is outside the scope here. This has the potential to improve the accuracy of the recovery inside the gap using constraints such as wavefront continuity and lateral smoothness. While it is also possible to regularize explicitly reconstructed data, it is much easier in my approach to take advantages of regularizations that are already in place.

### 4.4.1 Computation costs

In each of the scattering terms $\mathbf{GRK_c}$ in expression 4.6, the wavefield convolution is done after stencilling out the traces that lie inside unobserved locations. Therefore the computation cost of applying each of the higher-order terms will depend on the ratio of unobserved to observed locations, and can be much smaller than the convolution cost for the whole wavefield if we are only interested in near-offsets. Since we only have to apply a single $\mathbf{GRK_c}$ to the current term to obtain the subsequent term, the marginal cost of using higher-order terms is small, and the aggregate cost of computing ex-

pression 4.6 will only be a fractional overhead to the uncorrected modelling operator.

The story is different when applying the adjoint of the relinearized operator (expression 4.12) to form the gradient for $\mathbf{g}$ in the relinearization step. In this case each of the scattering terms add a $\mathbf{K_c}\mathbf{R}^H\mathbf{G}^H$ to the previous-order term. Although we still involve the stencil $\mathbf{K_c}$ before computing the next term, there is a special case when computing the $n$=1 term, where we do not have a stencil between $\mathbf{G}^H$ and the residual wavefield, and instead need to compute one full wavefield convolution over the whole grid. The cost of the (linearized) adjoint is thus two times the cost compared to the uncorrected modelling operator, plus the same fractional overhead for higher order terms as mentioned above.

However, note that my method avoids the cost of updating and storing the unknown data $\mathbf{p}''$ as described in step 6 of the original algorithm. The gradients for $\mathbf{p}''$ also require a full wavefield convolution, so the overall cost of updating the unknown wavefield is actually absorbed into the cost of updating $\mathbf{g}$. In fact, if we only use the $n$=1 term, then the overall computational cost of the relinearization strategy is identical to the original REPSI algorithm with explicit data reconstruction (assuming the same number of gradient updates is applied on both $\mathbf{p}''$ and $\mathbf{g}$), while being slightly more memory efficient due to not having to store and compare $\mathbf{p}''$. As discussed in the algorithms section, the cost of using the Gauss-Newton approach is significantly higher due to the partial derivatives in the Jacobian. Depending on the number of terms used, I find that the Gauss-Newton approach adds a two-to-three time computational overhead compared to the relinearization approach.

## 4.5 Summary

In this chapter I presented a modification of the REPSI problem that accounts for large gaps in the acquisition grid (as a function of both source and receiver coordinates), such as the near-offsets, without having to explicitly interpolate or invert for the missing near-offset data. The main idea of this method is to modify the forward modelling operator to fully explain the multiples in the observed data as long as the correct surface-free Green's function is obtained, even though the data operator itself is not completely sampled. This is achieved by augmenting the forward model for the observed wavefield with a truncated scattering series that approximately mimics the

action of the free surface reflector within the acquisition gap. Inverting this modified operator can thus recover the primary Green's function without involving the unobserved data in any way.

Part of the main contribution in this chapter is demonstrating that just two terms in the scattering series is enough to reach a useable accuracy for the REPSI inversion problem. I have demonstrated on both synthetic and real data that this level of approximation already leads to significant improvements over existing methods. The scattering terms involve the same wavefield convolution kernel as the original REPSI problem, so not much effort is required to implement its action. Furthermore, because the effective aperture of the multiple contribution for these scattering terms is limited to the inside of the acquisition gap, each additional term imposes only a small computational overhead compared to the original problem.

Unlike the original REPSI problem, the augmented modelling operator is no longer linear with respect to the primary Green's function; each term in the scattering series involve higher powers of the primary Green's function. I presented two possible modifications to the REPSI algorithm for dealing with this nonlinearity: a Gauss-Newton type approach (which involves a costly computation of the action of the Jacobian), and a straightforward relinearization approach that simply fixes the primary Green's function used in higher-order scattering terms at its previous iteration value. A simple scaling preconditioner allows the relinearization approach to perform similarly to the Gauss-Newton approach while keeping the same computational complexity as the original REPSI algorithm with explicit missing-data inversion, making it a suitable candidate for large industry-scale multiple-removal problems.

# Chapter 5

# Multilevel acceleration strategy for Robust EPSI

The practice of multiple removal on large industry-scale seismic data is a constant struggle between multiple conflicting factors, and can be summarized as a classical trade-off trilemma between accuracy, robustness, and low computational requirements. Current prediction-subtraction methods such as SRME face limits in accuracy and robustness when confronted with undersampled data of limited quality. This shortcoming promoted recent developments in whole-wavefield inversion/deconvolution techniques to decrease dependence on practitioner guesswork and QC, including the Robust EPSI method developed in this thesis. However, these more advanced methods usually require many iterations of multiple prediction to determine an entire unknown discretized wavefield, and thus are far more computationally intensive, posing significant hinderances to practical adaption.

This chapter proposes a strategy for significantly reducing the computational burden of Robust EPSI, without having to fundamentally alter the core implementation of the algorithm. This is achieved by essentially computing a large part of the early model updates on a hierarchy of coarsened spatial sampling grids (the temporal sampling remains untouched), which drastically reduces the computation time required for both the Fourier transform over the time axis as well as the matrix multiplication between the data-matrices. The results computed on the coarser grids are mapped to the full grid in a way that avoids spatial aliasing artifacts, which is the main

contributor of multiple prediction errors due to grid coarsening (Dragoset et al., 2006).

One of the advantages of the method proposed here is that it does not change the underlying physics and optimization problem of REPSI, only the problem size(s), and is therefore essentially "free" to implement. The main ingredients of this method is essentially a way to project and map the observed data, as well as the estimated model of the Green's function $\mathbf{g}$, between different grid sizes. In a way, the thematic structure of this method mimics what is usually known as "multilevel" (or, relatedly, "multigrid") methods in numerical linear algebra for inverting particular classes of linear operators, in that it involves moving intermediate solutions between different samplings in a disciplined way using restriction and prolongation operators.

It is important to note that borrowing nomenclature that is typical of the field of multigrid methods in numerical analysis does not imply that this chapter establishes anything more than a superficial allusion to these classical multigrid methods. Most notably, the operators involved in REPSI does not in general behave like elliptical partial differential operators, and we therefore lose many of the theoretical properties of prolongations in terms of reducing the residue. However, as I will show, the drastic reduction in computation cost when going to a coarser grid is simply enough to reduce the runtime of REPSI to a small fraction of the full problem, and therefore warrants the adoption of this approach.

This chapter will first continue with a discussion of the problems involved in a naive downsampling of the spatial grid, a way to mitigate it in the context of EPSI, and the reduction in computational complexity from the grid coarsening. I will then present a multilevel strategy, which involves an outer loop of downsampling and trace interpolation, and a inner loop of solving REPSI using the solution from the previous outer iteration as a warm start. It will finish with examples on both synthetic and real datasets which demonstrate the speedup obtained from the multilevel strategy, and that we can obtain a close approximation of the solution computed directly on the full grid.

## 5.1 Sampling requirements for EPSI

Wavefield convolution-based multiple prediction methods, such as SRME and EPSI, are essentially operating under an algebra over the set of well-

sampled discrete wavefields. The convolution relationship between different discrete wavefields are described as plain matrix multiplication over mono-frequency data slices. Implicitly, this ties the sampling grid of the wave-propagation operator (the Green's function) to the sampling grid of the data, since the matrices must be of consistent dimensions in order to multiply correctly.

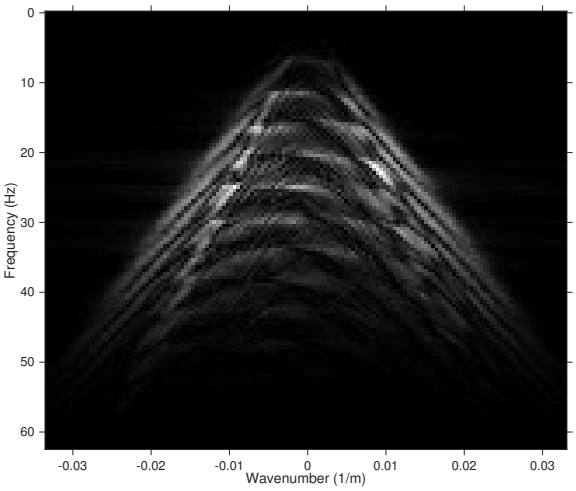
A further implicit assumption is that the sampling interval of this grid must be sufficient to support the frequency spectrum of the wavefield. The largest lateral wavenumber of a seismic recording is an increasing function of temporal frequency, and is typically determined by the slowest medium velocity. The spatial sampling of the whole wavefield is therefore bound by the highest frequency in the effective spectral band of the wavefield. If this relationship is not satisfied, then spatial aliasing occurs which creates further errors in the propagated wavefield, invalidating the integral equations which underlies both SRME and EPSI.

Figure 5.1 shows the F-K (frequency-wavenumber) spectra of a typical seismic common-shot gather. In Figure 5.1a a dataset with a spatial sampling just fine enough (15 m) to support a frequency content of up to 60 Hz. Figure 5.1b shows the spatial aliasing that results when you simply restrict the data to a 30 m grid (2× subsampling). However, since this aliasing only occurs at 30 Hz and above, a low-pass filter is sufficient to remove the aliased part of the spectrum and result in an alias-free wavefield, which can be seen in Figure 5.1c.

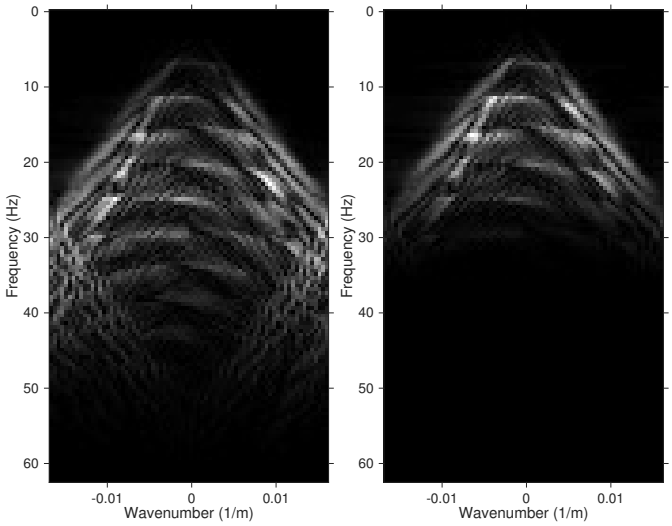
Therefore, in order to satisfy the $\mathbf{P} = \mathbf{GQ} + \mathbf{RGP}$ relation used in previous chapters for EPSI, one of two things can be done to ensure that a particular input data $\mathbf{p}$ is not spatially aliased:

1. Low-pass filter the data to limit the maximum effective frequency, then subsample spatially to a grid that satisfies the Nyquist criteria for the largest surviving wavenumber. We will use this general method, called **decimation**, as a general way to move to coarser grids (corresponding to the **restriction** step of general multigrid methods).

2. Interpolate the data to a finer grid that satisfies the existing Nyquist criteria for largest wavenumber without any low-pass filtering of the existing frequency content. This must be done in a way that does not result in any existing spatial aliases polluting the interpolated signal (i.e., no nearest-neightbour interpolation). Typically this is accomplished by weighted averages of nearby samples after doing a normal-




**(a)** original data




**(b)** 2× subsampled **(c)** low-pass filtered

**Figure 5.1** F-K spectra plot showing that low-pass filtering removes the spatial aliasing caused by subsampling seismic data onto a coarser sampling grid. Black indicates zero amplitude, while white indicates the maximum amplitude for each plot. The original data **(a)** is sampled at 15 m and contains significant frequency content up to 50 Hz. A 2× subsampling to a 30 m grid **(b)** exhibits spatial aliasing behaviour only after 30 Hz due to the minimum wave velocity in the data. After applying low-pass filter at 30 Hz **(c)** the spatial aliased components are removed, at the tradeoff of halving the spectral bandwidth of the data.

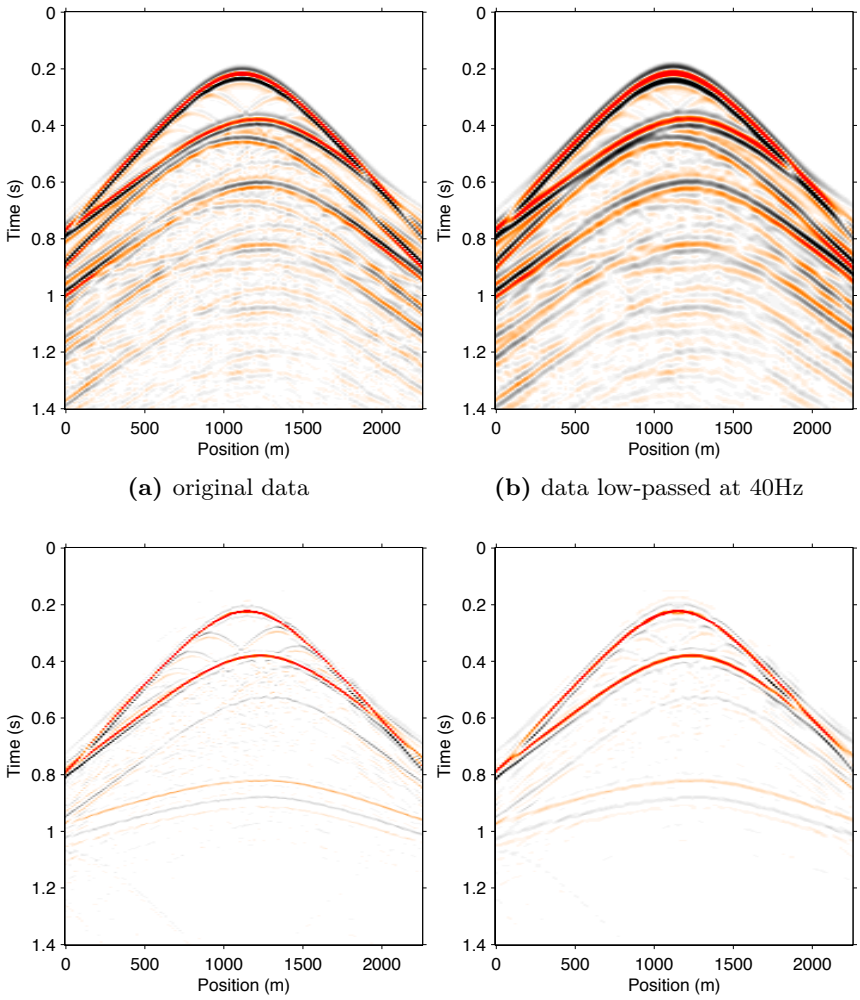
moveout (NMO) correction in the time domain. We will use the term **interpolation** as a general way to map wavefield on coarse grids to finer grids in an alias-free way (corresponding to the **prolongation** step of general multigrid methods).

## 5.2 Low-pass filter, subsampling, and their effects on EPSI

So far we have only involved basic results from Fourier analysis and signal processing, but the effectiveness of the multilevel strategy comes from an interesting characteristic unique to EPSI, and does not apply to SRME. Specifically, I use the fact that EPSI looks for a discretized multiple-free Green's function that is effectively "full-band". This full-band structure is enforced by imposing sparsity constraints in the physical time-space samples of the solution wavefield, similar in concept to spiking wavelet deconvolution. In fact, when we (globally) low-pass filter the input data, we can expect this filter to be captured just by the source signature model $\mathbf{q}$, while the Green's function $\mathbf{g}$ remains full-spectrum.

Essentially, as discussed in Chapter 2, EPSI can also be regarded as a deconvolution process that fundamentally attempts to recover the Green's function on any given spatial grid. Subsampling and low-pass filtering the input data given to EPSI should not, in the most ideal case, fundamentally change the temporal shape of the solution. If we keep the time axis sampling the same as the original data, EPSI should give us the same spikes at the same locations in the Green's function. However, given that the input data is now missing more frequency information due to the low-pass, it necessarily increase the difficulty of the blind deconvolution problem due to a smaller effective measurement. It is important to keep an account of whatever degradation on the estimated Green's function this introduces, in order to make sure that the solution is not totally off-base.

Figure 5.2 demonstrates the effect of this low-pass filtering (without the subsampling) on the Robust EPSI algorithm as described in previous chapters. The top two panels show a synthetic data, whose frequency content can be seen in Figure 5.1a (effectively up to 60 Hz), and the same data with a 40 Hz low-pass filter applied to the time domain using a Hanning window. The bottom two panels show the Green's function obtained respectively from the two dataset above, using the exact same optimization parameters.




**(a)** original data **(b)** data low-passed at 40Hz

**(c)** REPSI solution from (a) **(d)** REPSI solution from (b)

**Figure 5.2** Shot gathers of data (a), its REPSI solution (c), a low-passed 40Hz solution (b), and its REPSI solution (d). For the remainder of the chapter, white indicates zero amplitude, while red and black indicate the maximum and minimum amplitude, respective. Colormaps are normalized for plots within the same figure.

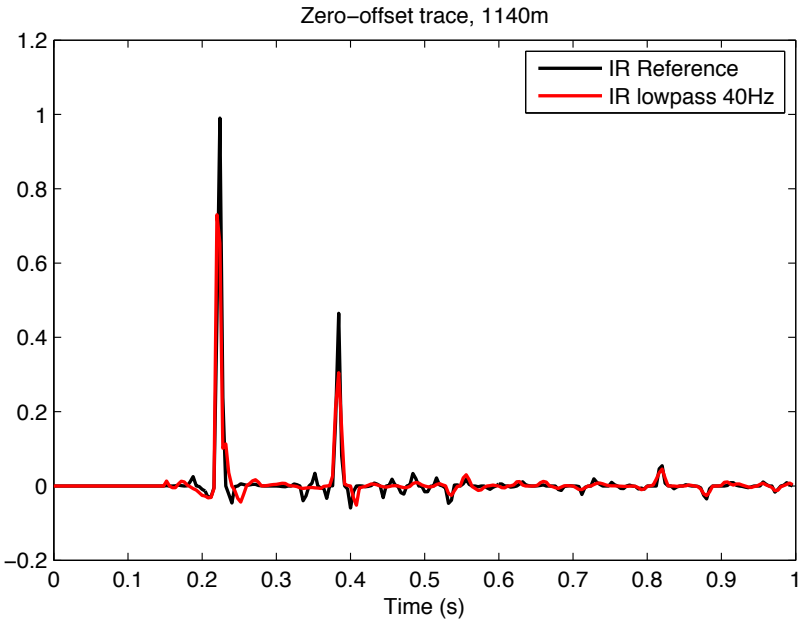



**Figure 5.3** Zero-offset trace of the REPSI solutions shown in Figure 5.2. Solution trace from the low-passed data shares the same general wavefront support as the original data.

As seen in Figure 5.3, the impulsive events in the two solutions both have similar shape and support location. The solutions from low-passed data contains fewer fine-scale features when compared to the solution obtained from the full data, but this is expected behaviour since the source signature was estimated blindly REPSI, and thus will be inherently less effective when resolving data with less spectral content. It is more important to realize that the solution to the low-passed data is already faster to compute, since the matrix-multiplications on the now eliminated higher frequencies do not need to be computed, and that if the 40 Hz data's REPSI solution was used as the initial guess, it will speed up the convergence of the original REPSI problem.

It is not hard to see that, if the reduced computational time of the 40 Hz problem is significant enough, then the *total* runtime of a two-stage optimization as proposed in that last sentence may indeed be shorter than the original problem. The multilevel strategy studied in this chapter is inspired by this idea, and I propose it as a possible generalization of the

multilevel method paradigm for REPSI. Before I introduce the full algorithm, I discuss below the expected reduction in computational complexity that can result when we not only low-pass, as above, but also subsample to a coarser grid which is facilitated by the reduction of spectral content.

### 5.2.1 Expected computational savings due to subsampling

Roughly speaking, REPSI converges through carefully-controlled projected gradient iterations, and by far the most time-consuming part of each iteration is the computation of $M(\mathbf{g}, \mathbf{q}; \mathbf{p})$ and its adjoint operation to form the gradient updates on $\mathbf{g}$. Both of these involve a complete wavefield convolution, modelled in the discretized setting as matrix multiplications between mono-frequency data-slices of the two wavefields.

Assuming that the number of samples in the source and receiver coordinates are the same so $n = n_{\text{rcv}} = n_{\text{src}}$, the data matrices are of size $n \times n$, so each gradient update iteration involves two operations that are of computation complexity

$$\mathsf{Cost}(n) = \mathcal{O}(2n_t n^2 \log n_t) + \mathcal{O}(n_f n^3),$$

where $n_f$ is the number of discrete frequencies, $n_t$ is the number of discrete time samples, to compute the multiple wavefield for a 2D (seismic line) acquisition geometry. The first term is the cost of computing the discrete Fourier transforms using FFT, and the second term is the cost of the matrix multiplications. A simple reduction of grid samples by half (doubling the sample distance), along with a similar halving of the highest frequency from low-passing, results in the computational complexity

$$\mathsf{Cost}\left(\frac{1}{2}n, \frac{1}{2}n_f\right) = \frac{1}{4}\mathcal{O}(2n_t n^2 \log n_t) + \frac{1}{16}\mathcal{O}(n_f n^3),$$

while a further halving (now one-quarter of the original sampling) results in

$$\mathsf{Cost}\left(\frac{1}{4}n, \frac{1}{4}n_f\right) = \frac{1}{16}\mathcal{O}(2n_t n^2 \log n_t) + \frac{1}{128}\mathcal{O}(n_f n^3).$$

Evidently, the $n^2$ and $n^3$ terms in the computational complexity of EPSI gradients which are normally a bane for large-area seismic surveys now helps us to drastically reduce the computational cost for subsampled data. By just halving the sampling, each gradient update is potentially computed an

order of magnitude faster than the full data, and at one-quarter sampling the speedup is now potentially two orders of magnitude. In fact, an entire REPSI problem at the one-quarter sampling can finish before the first gradient step is even finished at the full scale.

For a well-sampled 3D dataset (approximately the same sampling in both spatial directions, so that $n = nx_{\text{rcv}} = ny_{\text{rcv}} = nx_{\text{src}} = ny_{\text{src}}$), this effect is even more drastic, since the $n^2$ and $n^3$ terms for the 2D survey situation now becomes $n^4$ and $n^6$ terms. We end up with the following complexity for 3D surveys:

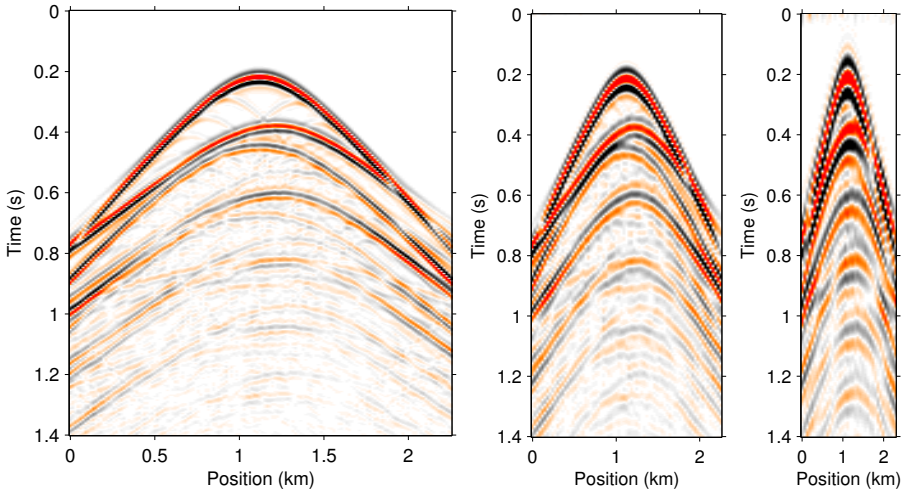
$$\begin{aligned}
\mathsf{Cost}(n) &= \mathcal{O}(2n_t n^4 \log n_t) + \mathcal{O}(n_f n^6) \\
\mathsf{Cost}\left(\frac{1}{2}n, \frac{1}{2}n_f\right) &= \frac{1}{16}\mathcal{O}(2n_t n^4 \log n_t) + \frac{1}{128}\mathcal{O}(n_f n^6) \\
\mathsf{Cost}\left(\frac{1}{4}n, \frac{1}{4}n_f\right) &= \frac{1}{256}\mathcal{O}(2n_t n^4 \log n_t) + \frac{1}{8192}\mathcal{O}(n_f n^6)
\end{aligned}$$

Figures 5.4 and 5.5 demonstrates this reduction in computation as we halve and quarter the original spatial sampling. The coarse-grid solutions were much faster to compute, with the original problem taking 40 minutes, $2\times$ decimation taking 6 minutes, and $4\times$ decimation less than 2 minutes. I will use these examples to illustrate the general multilevel strategy, after first giving a general description and outline of the method in the next section.

## 5.3 A multilevel strategy for REPSI

We observed in the previous section that coarse-scale solutions of REPSI closely resemble decimated versions of the solution at the original sampling interval. It is thus reasonable to postulate that the computation budget would be more efficiently spent on updates at the coarser level first until all coarse-scale information is exhausted, assuming that we also have an effective mapping (i.e., interpolation) from coarser grids to finer grids.

For the sake of simplicity and brevity, we fix a decimation ratio between subsequent grid sizes, such that we can choose our interpolation methods with a known criterion. In this chapter, we simply choose a $2\times$ decimation and interpolation ratio. Today, this degree of interpolation is routinely performed in SRME multiple removal procedures on existing surveys. Typically, field data (at least for marine) will have a shot sampling that is one-half that

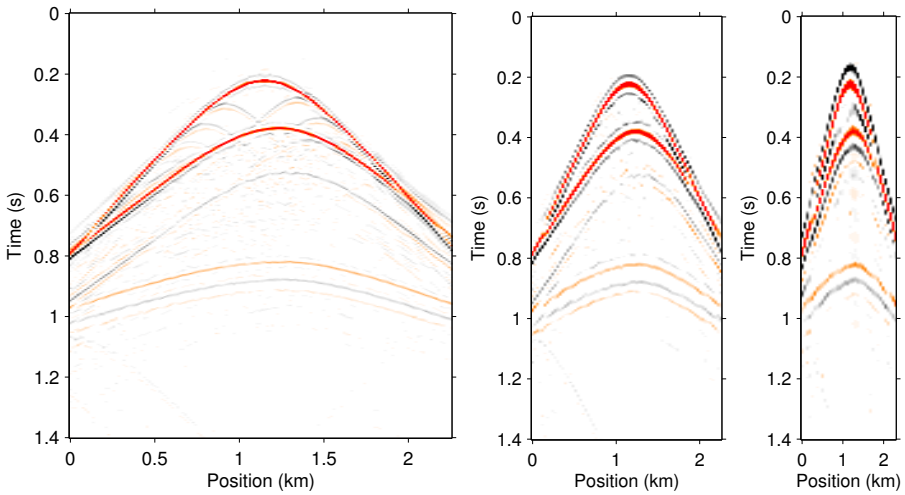



**(a)** Original sampling **(b)** 2× **(c)** 4×

**Figure 5.4** Synthetic data shot gathers at (a) the original spatial sampling and at (b,c) two levels of spatial decimation, with appropriate time-domain low-pass filters (at 30 Hz and 15 Hz respectively) to mitigate artifacts spatial aliasing when computing the surface-related multiple wavefield. The sampling of the time axis remains untouched.

of receiver sampling, so this level of interpolation is considered routine and be achieved reasonably well by, e.g., NMO-corrected trace averaging. Recursion of this approach leads to a multilevel strategy for efficiently solving the REPSI problem, which I outline in Algorithm 1.

As discussed in the introduction, because the REPSI problem inherently involves solving an integral equation of the second kind (Frijlink et al., 2011), our proposed method bears some theoretical connection to the general class of algebraic multigrid algorithms, although we are not able to take advantage of many of the theoretical guarantees of multigrid for parabolic PDE systems due to the hyperbolic nature of the wave equation operator. Nevertheless, this method should outperform unmodified REPSI in computation time provided that the interpolated coarse-scale solution brings $\mathbf{g}$ closer to the true finer-grid solution than the computationally equivalent number of iterations on the finer-grid problem.




**(a)** Original sampling **(b)** 2× **(c)** 4×

**Figure 5.5** Computed multiple-free seismic impulse response shot gathers using a straightforward application of the REPSI algorithm, from the original data and from its decimated, low-pass filtered versions shown in Figure 5.4. Even though its input data was low-pass filtered at 15Hz, the solution shown in (c) retains a wide-band, "deconvolved" appearance with good resolution of the two separate events at t=0.9s. More importantly, the coarse-grid solutions were much faster to compute, with the original problem taking 40 minutes, (b) taking 6 minutes, and (c) less than 2 minutes.

**Algorithm 5.1** Multilevel acceleration strategy for REPSI

1. **Input:** wavefield data $\mathbf{p}$ containing surface multiples
2. Choose maximum data decimation factor $2^S$, set scale variable $s{=}S$
3. Set initial solution $\mathbf{g}_{s=S}$ to zero vector
4. **Repeat**
5. $\quad \mathbf{p}_s \leftarrow$ subsample original data in all spatial coordinates by factor of $2^S$
6. $\quad$ Low-pass filter in the time domain on $\mathbf{p}_s$ to remove spatial aliasing
7. $\quad$ Solve the REPSI problem on subsampled data $\mathbf{p}_s$, initializing from $\mathbf{g}_s$
8. $\quad \mathbf{g}_{s-1} \leftarrow$ interpolate solution $\mathbf{g}_s$ in all spatial coordinates by factor of 2
9. $\quad s \leftarrow s-1$ (go to finer scale)
10. **until** scale variable $s{=}0$ (reached original spatial sampling)
11. Solve the REPSI problem on original data $\mathbf{p}$, initializing from $\mathbf{g}_0$
12. **Output:** approximate solution $\tilde{\mathbf{g}}$ to the REPSI problem (equation 2.4)

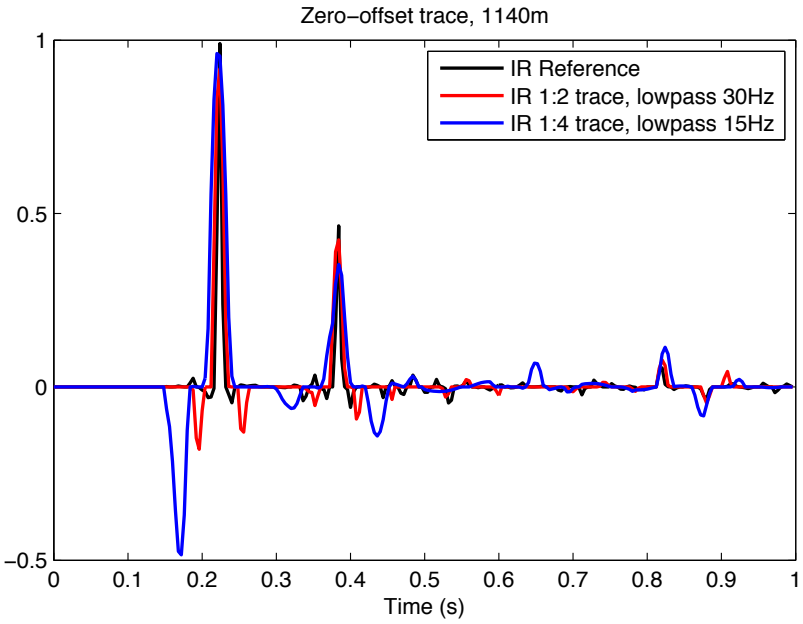



**Figure 5.6** Zero-offset trace for the computed REPSI solutions shown in Figure 5.5. As can be seen, most of the major events coincide, although noticeable degradation can be seen in the 15Hz solution, suggesting that further subsampling might produce results with completely unrecognizable events.

## 5.4 Numerical example

In this section I will illustrate the multilevel strategy using the familiar synthetic dataset shown in the previous sections. I also use the descriptions in this section to expand on some practical details of the multilevel strategy, as well as give tangible figures for computational time savings. I will also use a North Sea dataset similar to the one used in the last section (but with pre-interpolated near-offsets) to demonstrate the viability of this method in practical application, even though the interpolation and decimation tools used here are relatively simple.

### 5.4.1 Synthetic data example

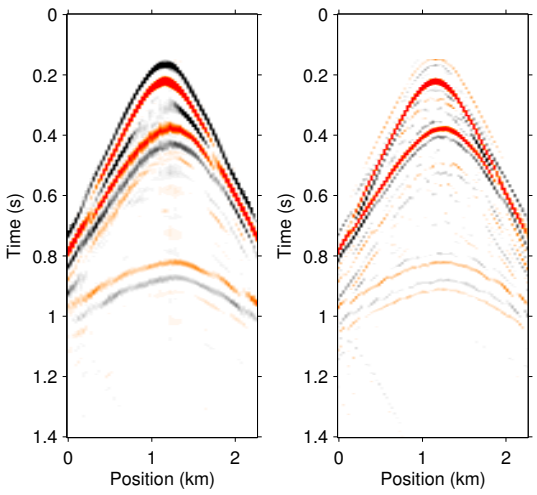
Using the fully-sampled, fixed-spread 2D synthetic seismic line data shown in Figure 5.4a, I illustrate the performance improvements possible with our proposed multilevel strategy. This is the same synthetic dataset that was heavily featured in the previous chapters. The original spatial sampling is 15 m in both the source and receiver coordinates, with 150 source and receiver locations, and 512 time samples. The data was modelled using a source function of a zero-phase Ricker wavelet with peak frequency at 30 Hz, with the highest available frequency at around 60 Hz, near the cutoff for the maximum spatial wavenumber supported by the 15 m sampling.

As shown in Figure 5.5, directly applying the unmodified REPSI algorithm to decimated, smoothed versions of the original data yields decimated, *non-smooth* versions of the original solution, due to the sparse-promoting nature of REPSI. The impulse response shown in Figure 5.5b is computed from the data shown in Figure 5.4b, which is a $2\times$ decimated version of the original data (spatial sampling at 30 m, dropping every other receiver trace and also every other shot record from the original data) followed by low-pass filter with a one-sided Hanning window at 30Hz. Similarly, Figure 5.5c follows from the data shown in Figure 5.4c with 60m spatial sampling and low-pass at 15 Hz. The three different solutions in Figure 5.5 all took roughly 75 iterations of the REPSI algorithm. Due to the cubic scaling of compute operations with the number of grid points, obtaining the coarsest-scale solution in Figure 5.5c is roughly equal in runtime to just 1 iteration of the original undecimated problem.

We now follow the multilevel strategy outlined in Algorithm 5.1, starting with the coarsest decimation factor $4\times$ (corresponding to $s = 2$). As hinted in Figure 5.6, decimation factors coarser than this produced Green's functions that are too erroneous to act as a good starting model. The EPSI problem is solved at this scale (shown back in Figure 5.5c), using 75 iterations as described above. This takes roughly 2 minutes on a contemporary workstation. The Green's function solution at $4\times$ decimation is then mapped to the next finer gird of $2\times$ decimation (corresponding to $s = 1$) via a simple linear interpolation after applying NMO correction in the common-midpoint domain (using a constant velocity of 1600 m/s), which effectively decreases the maximum spatial wavenumber in a deterministic way by reducing the curvature of hyperbolic wavefronts. After interpolation, an inverse NMO correction is then applied.

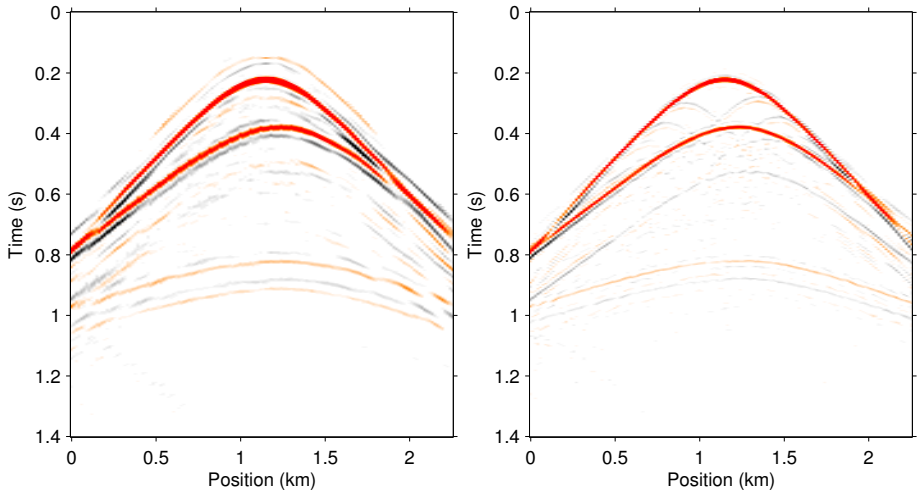
The result of this interpolation and mapping to the 2× grid is shown in 5.7a. I then run the REPSI algorithm using the 2× decimated and low-pass filtered data shown in Figure 5.4b, using the interpolated 4× solution as the initial solution of the iterative algorithm. This would also mean skipping the original spike-picking initialization step outlined in Section 3.3.5. Note that the solution for the wavelet $\mathbf{q}$ at 4× needs to be discarded, since it will have the spectral imprint of the 15 Hz low-pass filter which is applied to the 4× decimated data, and is thus invalid for the 2× decimated data that is low-pass filtered at 30 Hz. A new initial wavelet must be therefore be estimated instead from the interpolated solution.

After 30 iterations at the 2× grid, we obtain a solution that is comparable to the misfit achieved after 75 iterations at 2× grid starting from an empty initial solution. The resulting Green's function is shown in Figure 5.7b, which can be compared to the reference solution for this scale shown in Figure 5.5b.




**(a)** initial at 2× **(b)** final at 2×

**Figure 5.7** The Green's function model at the 2× decimation (30 m) sampling, computed as part of a multilevel continuation strategy for accelerating REPSI. Panel **(a)** shows the NMO-interpolated solution of the REPSI problem at 4× decimation (shown in Figure 5.5c), which is used as initial model for the REPSI problem at 2×. Panel **(b)** shows the final Green's function solution at 2× decimation, using 30 gradient iterations.




**(a)** initial at original sampling **(b)** final at original sampling

**Figure 5.8** The Green's function model at the original (15 m) sampling, computed as part of a multilevel continuation strategy for accelerating REPSI. Panel **(a)** shows the NMO-interpolated solution of the REPSI problem at 2× decimation (shown in Figure 5.7b), which is used as initial model for the REPSI problem at the original sampling. Panel **(b)** shows the final Green's function solution at the original sampling of the data, using just 15 gradient iterations.

Finally, we go to the final $s = 0$ outer loop in Algorithm 5.1, taking the bootstrapped 2× grid solution from Figure 5.8b, and use it (again, after NMO correction and interpolation, shown in Figure [#5.8a]) as the initial solution in a REPSI problem at the original scale. The resulting Green's function for this original scaling is shown in Figure 5.7b, which can be compared to the reference solution for this scale shown in Figure 5.5a. This time, we obtained an acceptable solution after just 15 iterations, comparable in misfit to 75 iterations starting from empty solution.

Figure 5.10 compares the final demultipled wavefields as computed through the multilevel strategy against a reference solution computed completely at the original scale. The results are very similar despite the multilevel strategy requiring only one-third of the runtime of the original method, as described in Figure 5.9.

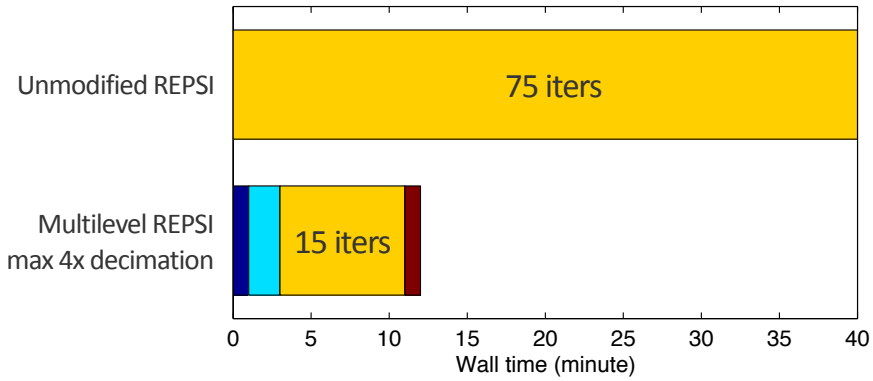
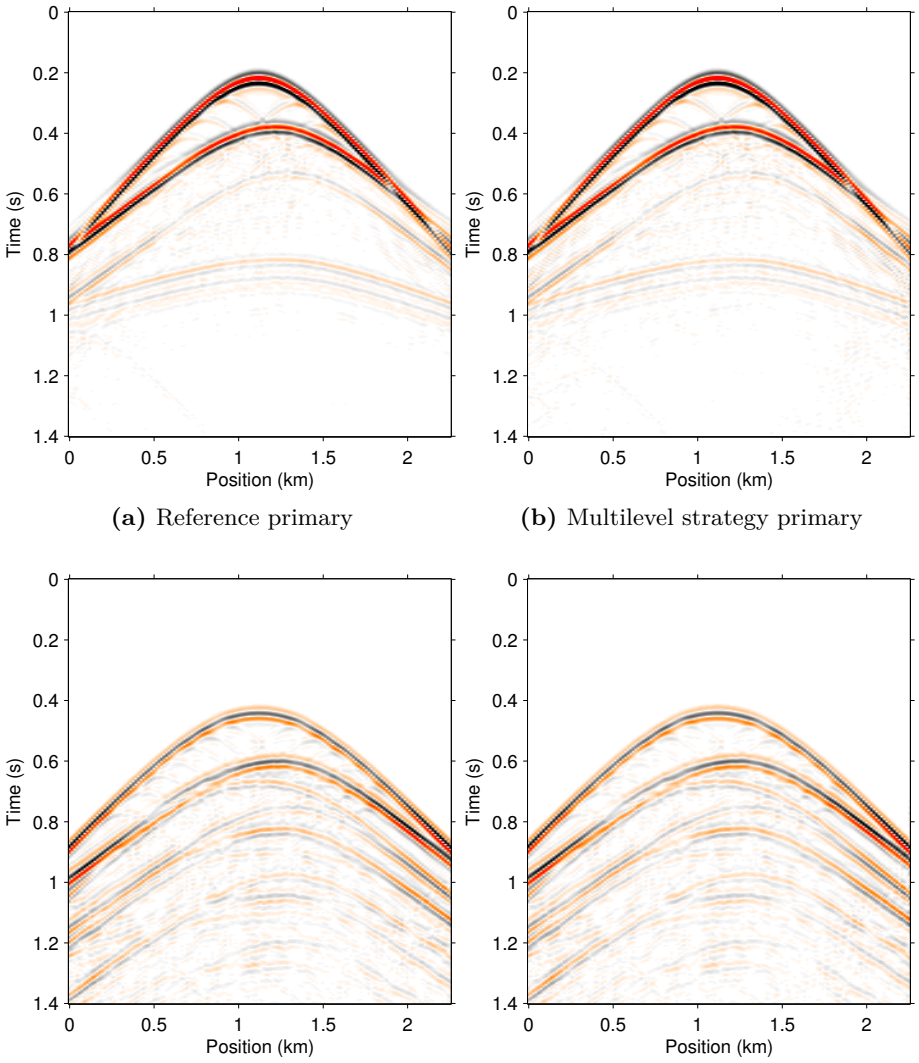



**Figure 5.9** Runtime comparison for the synthetic data example between REPSI run entirely at the original spatial sampling **(40 minutes)**, and the proposed multilevel strategy beginning at 4× decimation of the data **(12 minutes)**. Different colours denote REPSI computation at different decimation levels of the data, with yellow the original sampling, light blue the 2× decimation, and dark blue the 4× decimation. Brown colour denotes computations done for interpolation to a finer grid.

### 5.4.2 Real data example

The dataset used in this section is similar to the North Sea dataset used in Section 4.3.2. However, to isolate the results in this chapter from any algorithmic choices made in the methods in introduced last chapter, I have decided to pre-interpolate any missing near-offset traces using parabolic Radon transform. The original dataset has a 12.5 m spatial sampling, with 361 source and receiver positions and 1024 time samples, making it significantly larger than the synthetic dataset above. Like the synthetic example, I prepare two levels of decimated data at 25 m (2× decimation) and 50 m (4× decimation) sampling, and low-pass filtered at 40 Hz and 20 Hz respectively. These different decimations of the data are shown in Figure 5.11.

For the real data example I will skip plotting the intermediate scale solutions, which are mainly for illustrative purposes, and show directly the final results. A summary of the runtime reduction is shown in Figure 5.12. A three-times speedup was achieved, which is similar to the synthetic case, and largely influenced by the number of iterations at the original scale in the final outer loop.



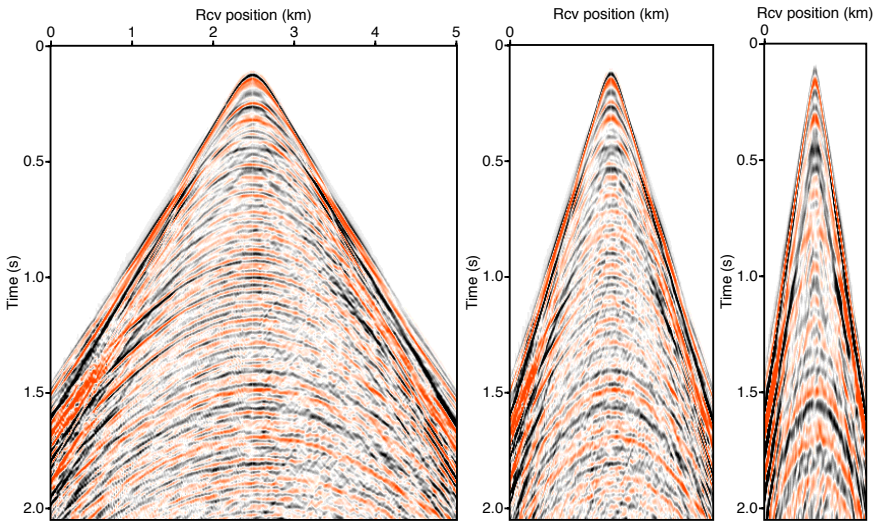

**(a)** Reference primary

**(b)** Multilevel strategy primary

**(c)** Reference multiple

**(d)** Multilevel strategy multiple

**Figure 5.10** Synthetic example comparison shot-gathers of the final demultipled primary wavefield between the REPSI algorithm run at the original spatial sampling, and the proposed multilevel strategy beginning at a 4× decimation of the data.




**(a)** Original sampling **(b)** 2× **(c)** 4×

**Figure 5.11** North Sea field data shot gathers at (a) the original spatial sampling and at (b,c) two levels of spatial decimation, with appropriate time-domain low-pass filters (at 40 Hz and 20 Hz respectively) to mitigate artifacts spatial aliasing when computing the surface-related multiple wavefield. The sampling of the time axis remains untouched.

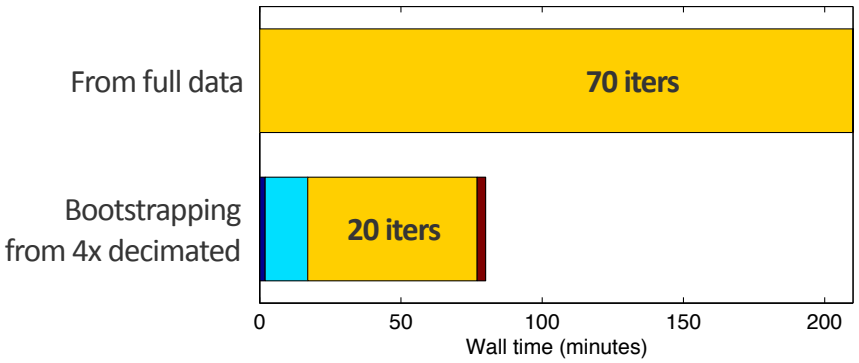



**Figure 5.12** Runtime comparison for the North Sea field data example, between REPSI run entirely at the original spatial sampling **(209 minutes)**, and the proposed multilevel strategy beginning at a 4× decimation of the data **(79 minutes)**. Different colours denote REPSI computation at different decimation levels of the data (c.f. Figure 5.9).

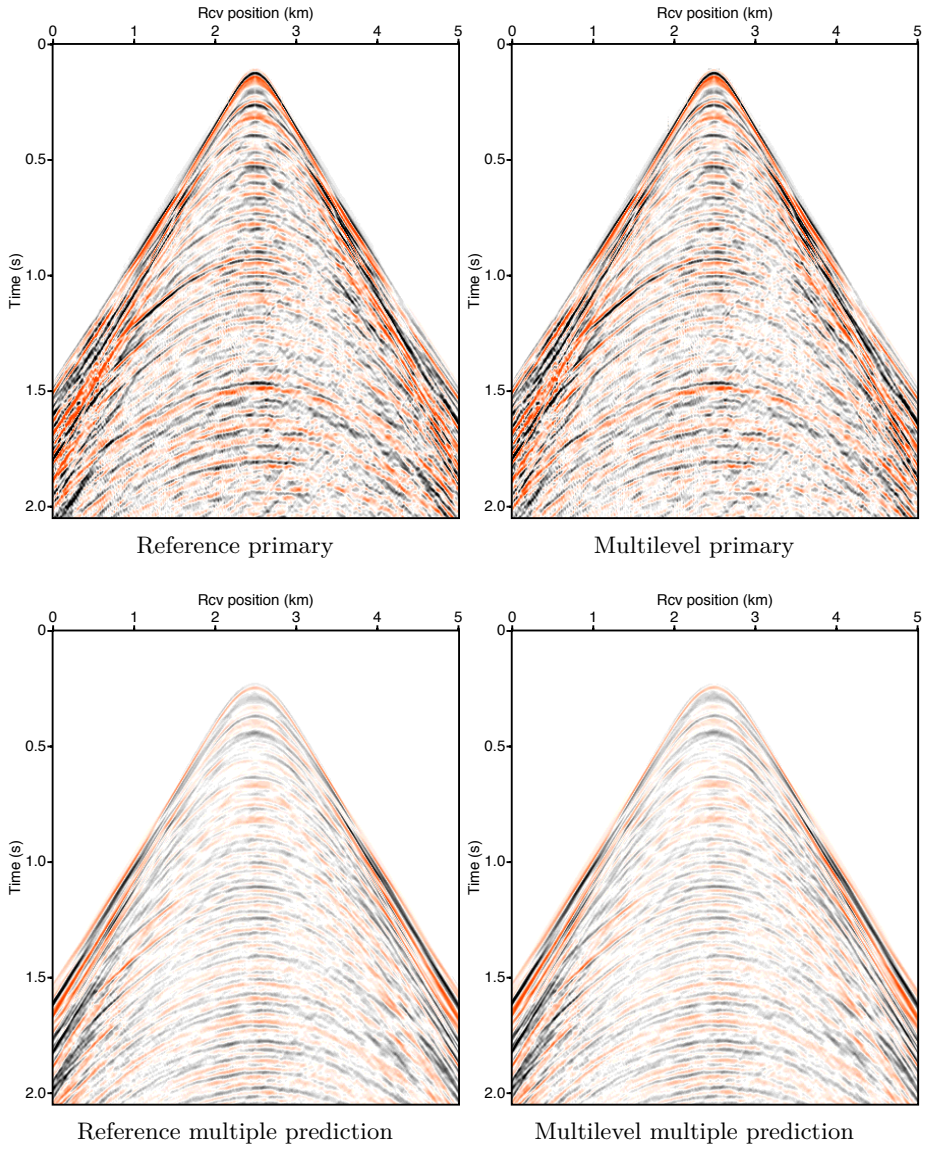
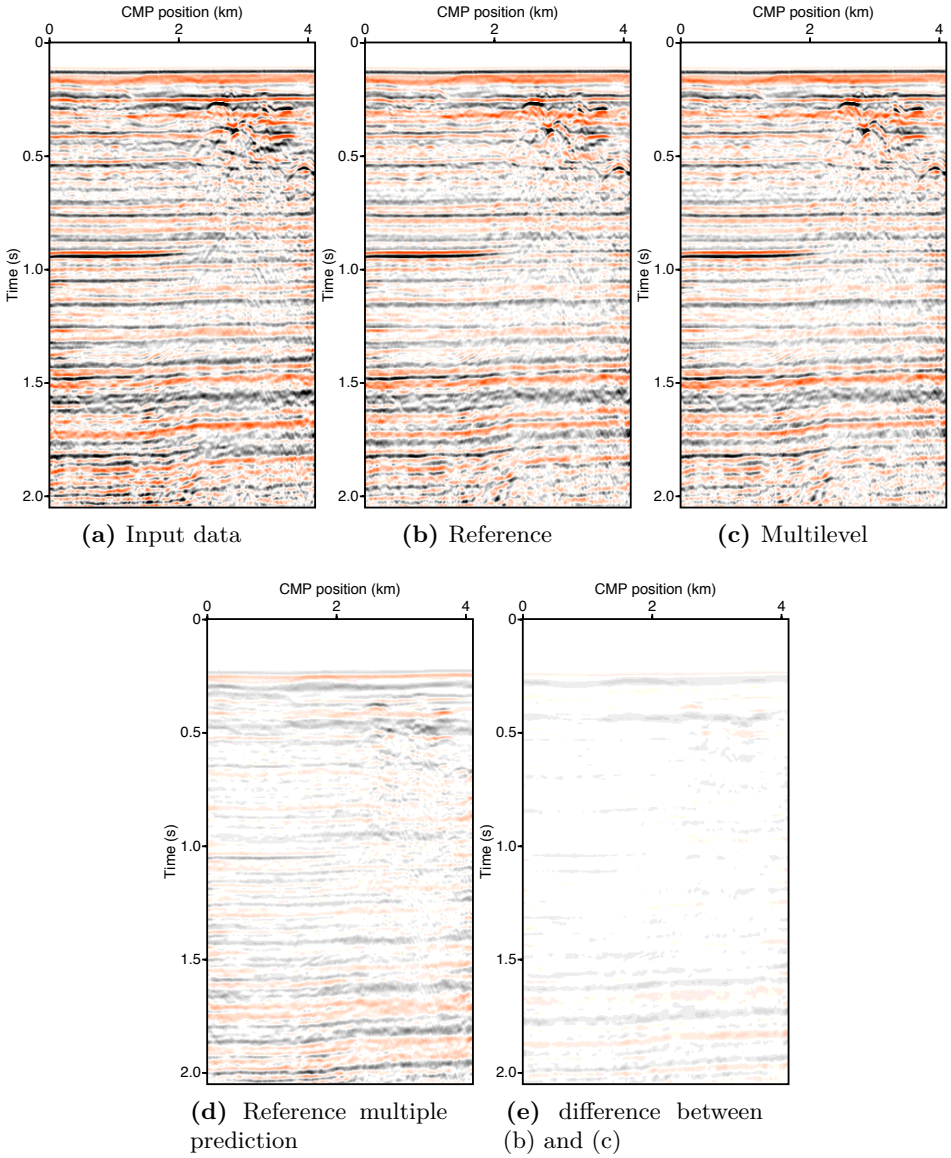



**Figure 5.13** North Sea field data example: shot-gather comparison of the final obtained primary wavefield solution between the REPSI algorithm run at the original spatial sampling, and the proposed multilevel strategy beginning at a 4× decimation of the data.




**(a)** Input data **(b)** Reference **(c)** Multilevel

**(d)** Reference multiple prediction **(e)** difference between (b) and (c)

**Figure 5.14** North Sea field data example: NMO stack comparison of the final obtained primary solution between the REPSI algorithm run at the original spatial sampling, and the proposed multilevel strategy beginning at a 4× decimation of the data.

Figure 5.13 compares the final demultipled wavefields for the real data example as computed through the multilevel strategy, against a reference solution computed completely at the original scale. The two results here are extremely similar, and the differences are hard to discern visually.

To ensure that we do not end up removing erroneous surface multiple events using the multilevel strategy, in Figure 5.14 I show NMO-corrected time domain stacks as a rough approximation of a seismic image computing from this data. Specifically, Figure 5.14d shows an image of the multiple prediction generated from the reference solutions' Green's function when convolved with the original data. Figure 5.14e shows an image of the difference between the primary wavefield generated by both the multilevel strategy solution and the reference solution. Comparing the two, we see that the approximation error is mainly concentrated on small amplitude differences on existing surface multiple events, and no extraneous or spurious artifacts are introduced by following the multilevel strategy.

## 5.5 Summary

The main motivation behind this chapter is the observation that intermediate solutions to the Robust EPSI problem can be computed on a coarser gird and interpolated to a finer grid orders-of-magnitude faster than the time required to reach comparable accuracy when working directly on the finer grid. To exploit this discrepancy, I proposed a method based on a multilevel inversion strategy to substantially reduce the computational requirements of Robust EPSI.

I have also demonstrated on both synthetic and real 2D seismic lines that, without appreciable loss in solution quality, significant speedups for the Robust EPSI algorithm is possible using this approach. Further speedups can be expected for a 3D acquisition geometry due to the higher computation complexity order as a function of grid point density.

# Chapter 6

# Conclusions

Throughout this thesis I have developed a method called Robust EPSI, which is a novel algorithm aimed at reliably solving the EPSI problem for surface-related primary estimation using contemporary methodology inspired by recent trends in large-scale sparsity-promoting optimization. In Chapter 3, I formalized the Robust EPSI problem as a formal bi-convex optimization problem, and proposed an effective solver for this problem that requires less parameter tweaking compared to existing algorithms while delivering improved results. In Chapter 4, I extended the multiple prediction model of Robust EPSI to account for contiguous unobserved gaps in the input wavefield using scattering terms, and as a result was able to mitigate larger gaps than the existing proposed method of explicitly estimating the missing data. In Chapter 5 I proposed a multilevel continuation strategy that significantly speeds up the computation time for the Robust EPSI problem.

In the introduction chapters of this thesis I brought up several shortcomings of the EPSI algorithm as it was originally stated in van Groenestijn and Verschuur (2009a). My work in this thesis results in the following findings which contribute to addressing those shortcomings:

1. The EPSI problem can be effectively described as a dual variable, bi-convex optimization problem. Under a cyclic block coordinate descent scheme, the bi-convex problem decomposes into a basis pursuit denoising (BPDN) problem in the variable describing the surface-free Green's function, and a constraint satisfaction problem in the variable

describing the source signature. I call this formalism the Robust EPSI problem. (Chapter 3)

2. A continuation-based convex solver based on the BPDN Pareto root-finding paradigm can effectively solve the above Robust EPSI problem after a suitable initialization step for the unknown wavelet. The Green's function model is obtained as a typical sparse regularization problem, while the wavelet is updated along the way to minimize the misfit of the forward model by solving a constrained massively overdetermined least-squares fitting before each continuation step. The overall approach is an extension of an existing method called SPG$\ell_1$. The algorithm performs well on test synthetic and field data, and is able to produce more accurate models of the Green's function while also requiring less parameter input when compared to the existing algorithm. I call this solution method the Robust EPSI algorithm. (Chapter 3)

3. The forward wavefield convolution model used in Robust EPSI can be augmented with a truncated scattering series to mitigate multiple errors introduced by propagating an incompletely sampled wavefield. These scattering terms are mainly composed of the corresponding order of surface multiples which decay in amplitude, and therefore just the first few terms of this series is enough to account for the majority of the errors introduced by incomplete wavefield sampling. (Chapter 4)

4. The modified Robust EPSI problem that uses the above augmented forward multiple prediction model no longer takes the form of a simple BPDN problem with respect to the Green's function. However, the Robust EPSI algorithm can be easily adapted to this problem with a straightforward relinearization scheme under a suitable preconditioner. The resulting algorithm is shown to outperform an existing explicit data reconstruction scheme for EPSI, in the sense that it is able to more accurately obtain the whole primary wavefield under a larger near-offset gap in the seismic data. (Chapter 4)

5. A simple multilevel continuation strategy can successfully accelerate the computational runtime needed for Robust EPSI without appreciable changes to the underlying algorithm nor the implementation. This strategy offloads early iterations of the Robust EPSI algorithm to coarser spatial grids run at more iterations. A straightforward choice of restriction and prolongation operations that utilizes low-pass filter-

ing and post-NMO interpolation can be effective at avoiding aliasing issues when moving between different sampling grids. Overall, on 2D real and synthetic data I showed that one can expect the computation time to be 25% of that needed by the unmodified algorithm, without appreciable loss in accuracy. Analysis shows that more drastic computation savings is expected for 3D. Since EPSI is often considered a computationally expensive method, this contribution goes a long way towards improving its practicality on existing computer hardware. (Chapter 5)

## 6.1 Impact to the current field of multiple removal

For the field of surface multiple removal to evolve, it is now considered very important for the inversion-based methods, such as EPSI, to receive more widespread adoption. There are many existing acquisition scenarios for which the current SRME methodologies, even with sophisticated adaptive subtraction methods, fail to acceptably perform multiple removal. An example of this would be shallow-water (less than 200 m) acquisition where surface multiples are exceptionally strong, yet at the same time it is extremely difficult to get acceptable near-offset measurements. Another example would be in cases where surface multiples overlay with very weak but important primary signal, such as sub-salt areas, to the degree that adaptive subtraction can never hope to cleanly preserve the primary.

The contributions made in the thesis thus impact the existing survey industry by removing many of the existing roadblocks leading to practical EPSI adoption, as well improving on its capabilities. The work in Chapter 3 takes most of the guesswork out of using EPSI by removing many free parameters introduced by the ad-hoc sparse regularization, while providing a better solution quality of the Green's function. Since the publication of this chapter in *Geophysics*, some authors have independently implemented the Robust EPSI algorithm and verified that it does indeed provide a more accurate solution for the primary wavefield with less parameter tweaking (Czyczula Rudjord et al., 2015). In Chapter 5, I further improve on the practicality of this method by introducing a simple way to make its computation significantly faster, with tools that practitioner would easily have at their disposal. In Chapter 4, I improved the inherent capabilities of EPSI,

providing a way to deal with even larger near-offset gaps than EPSI was ever previously shown to work in existing literature. Altogether, this collection of work should hopefully make adopting and implementing the general EPSI paradigm a more enticing prospect for working seismic surveyors.

## 6.2 Future research directions

In general, the EPSI technique, and primary estimation by inversion, is in its nascent stages of development. This thesis has made significant progress in terms of making EPSI practical for existing surveying projects, but more could be done in that regard. One obvious direction is towards a full 3D implementation of Robust EPSI. We did not get to it, but many of the contributions in this thesis are done with eventual 3D implementations in mind, and in service of it. For example, the acceleration strategy developed in Chapter 5 was always meant for a full 3D implementation, as that is where we expect to see the most drastic reduction in computing costs compared to a fully sampled 3D EPSI problem that is currently considered impractical on existing hardware.

A major roadblock for a full 3D formulation is one that I did not directly address in the thesis: the issue of *undersampling*. Even in Chapter 4, which concerns mitigating missing data, I have made a careful distinction of sampling patterns that results in a large gap of missing data from one the introduces spatial aliasing issues due to undersampling. The latter is special in that it invalidates the assumptions of the wave equation in such a way that our forward model becomes fundamentally unable to correctly produce a multiple wavefield using the type of convolution discussed in this thesis. So far all, the 2D data we use have only seen a $2\times$ upsampling when interpolating the source coordinate. In 3D, it is very common to require much more difficult interpolations, with cross-line sampling distance to 5 times or even 10 times the in-line receiver sampling. It remains to be seen whether current upsampling techniques introduce prediction errors that severely compromise both EPSI and REPSI. One possible solution to this can come from a hybrid approach, combining some of the the scattering method in Chapter 4 with other kinds of interpolation methods currently available.

In the current literature, many extensions to the EPSI paradigm exist that tweaks the prediction model for slightly acquisition designs. For example, the surface integration does not necessarily need to be coincident with the source and receiver locations (and often are not). As alluded to

the introduction section, for marine streamer acquisitions this can be simply satisfied with proper receiver or source-side deghosting, but one can think of more extreme cases such as when receivers are ocean bottom nodes or cables, in which case it may be separated hundreds of meters from the integration surface. There is existing work that extends EPSI to this kind of acquisition by introducing wave propagation operators into the integral expression for the multiples (van Groenestijn and Ross, 2011). These kinds of extensions can easily be brought to the Robust EPSI scheme. Speaking in broad terms, some more low-hanging contributions can be made by working to relax either the source signature model $\mathbf{q}$ (e.g., van Groenestijn and Verschuur, 2011) or the surface reflection operator $\mathbf{R}$ (e.g., AlMatar, 2010) to capture the physical characteristics of different field surveys.

Finally, a very rich area of investigation lies in developing a deeper theoretical link between blind deconvolution and EPSI. As Figure 3.5 showed, the Robust EPSI approach results in a clean estimate of the Green's function. This opens up the possibility of using Robust EPSI as a physically-consistent way to deconvolve the source signature. This approach will differ from the traditional "spiking" deconvolution in that it additionally exploits information from the multiples to construct the impulse response. Some early work on this topic in terms of spectrum recovery of seismic data can be seen in Lin and Herrmann (2011). It is possible that (Robust) EPSI can play a very important role in preparing seismic data for applications that require wide-band data.

Furthermore, as was discussed in section 2.5.3, the inherent feedback mechanism in EPSI by itself has interesting theoretical implications as a variation on the traditional blind deconvolution models, in in the sense that it introduces a more constrained relationship between the wavelet model and the signal model. There is an existing work with colleagues that exploits this general problem form for a novel deconvolution algorithm (Esser et al., 2015). A more costly but more sophisticated deconvolution algorithm fits naturally into the multilevel strategy in Chapter 5: at the very coarse scales, computation is extremely cheap, but the bandwidth limitation imposed by the low-pass filter becomes more daunting, so being able to get a more accurate solution from a computationally costly algorithm can be welcome. I find this line of research very promising, and definitely plan to pursue this inquiry in the future.

# Bibliography

AlMatar, M. H., 2010, Estimation of Surface-free Data by Curvelet-domain Matched Filtering and Sparse Inversion: Master's thesis, University of British Columbia. → pages 18, 126

Amundsen, L., 2001, Elimination of free-surface related multiples without need of the source wavelet: Geophysics, **66**, 327–341. → pages 23

Anstey, N. A., and P. Newman, 1966, Part I: The sectional auto-correlogram and Part II: The sectional retro-correlogram: Geophysical Prospecting, **14**, 389–426. → pages 18, 20

Aravkin, A. Y., and T. van Leeuwen, 2012, Estimating Nuisance Parameters in Inverse Problems: Inverse Problems, **28**, 16. → pages 50

Baardman, R. H., D. J. Verschuur, R. G. van Borselen, M. O. Frijlink, and R. F. Hegge, 2010, Estimation of primaries by sparse inversion using dual-sensor data: SEG Technical Program Expanded Abstracts, **29**, 3468. → pages 26

Backus, M. M., 1959, Water reverberation - their nature and elimination: Geophysics, **24**, 233–261. → pages 14

Battle, G., 1987, A block spin construction of ondelettes. Part I: Lemarié functions: Communications in Mathematical Physics, **110**, 601–615. → pages 59

Berkhout, A. J., and Y. H. Pao, 1982, Seismic Migration - Imaging of Acoustic Energy by Wave Field Extrapolation: Journal of Applied Mechanics, **49**, 682. → pages 21, 24

Berkhout, A. J., and D. J. Verschuur, 1997, Estimation of multiple scattering by iterative inversion, Part I: Theoretical considerations: Geophysics, **62**, 1586–1595. → pages 21

Bernth, H., and L. Sonneland, 1983, Wave-field extrapolation techniques for prestack attenuation of water reverberations: SEG Technical Program Expanded Abstracts, 264–265. → pages 19

Berryhill, J. R., and Y. C. Kim, 1986, Deep-water peg legs and multiples: Emulation and suppression: Geophysics, **51**, 2177–2184. → pages 19

Bezdek, J. C., R. J. Hathaway, R. E. Howard, C. A. Wilson, and M. P. Windham, 1987, Local convergence analysis of a grouped variable version of coordinate descent: Journal of Optimization Theory and Applications, **54**, 471–477. → pages 42

Biersteker, J., 2001, MAGIC: Shell's surface multiple attenuation technique: SEG Technical Program Expanded Abstracts 2001, Society of Exploration Geophysicists, 1301–1304. → pages 22, 23

Biggs, D. S. C., 1998, Accelerated iterative blind deconvolution: PhD thesis, University of Auckland, Auckland. → pages 53

Birgin, E. G., J. M. Martinez, and M. Raydan, 2000, Nonmonotone spectral projected gradient methods on convex sets: SIAM Journal on Optimization, **10**, 1196–1211. → pages 84, 96

Cabrelli, C. A., 1985, Minimum entropy deconvolution and simplicity: A noniterative algorithm: Geophysics, **50**, 394. → pages 29

Cambois, G., D. Carlson, C. Jones, M. Lesnes, W. Söllner, and H. Tabti, 2009, Dual-sensor streamer data: Calibration, acquisition QC and attenuation of seismic interferences and other noises: SEG Technical Program Expanded Abstracts, 142–146. → pages 95

Candès, E. J., J. Romberg, and T. Tao, 2006, Stable signal recovery from incomplete and inaccurate measurements: Communications on Pure and Applied Mathematics, **59**, 1207. → pages 47

Chen, S. S., and D. L. Donoho, 1994, Basis pursuit: Proceedings of 28th Asilomar Conference on Signals, Systems and Computers, IEEE Comput. Soc. Press, 41–44. → pages 35

Chen, S. S., D. L. Donoho, and M. A. Saunders, 2001, Atomic decomposition by basis pursuit: SIAM Review, **43**, 129. → pages 35, 45

Claerbout, J. F., and F. Muir, 1973, Robust modeling with erratic data: Geophysics, **38**, 826. → pages 46

Czyczula Rudjord, Z., K. de Vos, and P. Hoogerbrugge, 2015, Real Data Comparison of EPSI and Robust EPSI Including the Near Offset Reconstruction: Presented at the 77th EAGE Conference & Exhibition, EAGE. → pages 124

Daubechies, I., M. Fornasier, and I. Loris, 2008, Accelerated projected gradient method for linear inverse problems with sparsity constraints: Journal of Fourier Analysis and Applications, **14**, 764–792. → pages 36, 84, 139, 141

Donno, D., 2011, Improving multiple removal using least-squares dip filters and independent component analysis: Geophysics, **76**, V91–V104. → pages 22

Donno, D., H. Chauris, and M. Noble, 2010, Curvelet-based multiple prediction: Geophysics, **75**, WB255–WB263. → pages 22

Donoho, D. L., 1981, On minimum entropy deconvolution, *in* Applied time series analysis II: Academic Press, 565–608. → pages 29

———, 2006, For most large underdetermined systems of linear equations the minimal L1-norm solution is also the sparsest solution: Communications on Pure and Applied Mathematics, **59**, 797–829. → pages 46, 47

Donoho, D. L., and Y. Tsaig, 2008, Fast Solution of L1-Norm Minimization Problems When the Solution May Be Sparse: IEEE Transactions on Information Theory, **54**, 4789–4812. → pages 48

Dossal, C., and S. Mallat, 2005, Sparse spike deconvolution with minimum scale: Proc. Signal Processing with Adaptive Sparse Structured Representations 2005, 123–126. → pages 46

Dragoset, B., I. Moore, and C. Kostov, 2006, The impact of field-survey characteristics on surface-relatedmultiple attenuation: Geophysical Prospecting, **54**, 781–791. → pages 102

Dragoset, B., D. J. Verschuur, I. Moore, and R. Bisley, 2010, A perspective on 3D surface-related multiple elimination: Geophysics, **75**, 75A245. → pages 22

Dragoset, W. H., and v. Jeričević, 1998, Some remarks on surface multiple attenuation: Geophysics, **63**, 772–789. → pages 22

Durrani, J. A., 1991, Multiple attenuation by predictive deconvolution in the p-t domain: Presented at the EAGE Technical Program Expanded Abstracts. → pages 15

Esser, E., T. Lin, R. Wang, and F. Herrmann, 2015, A Lifted l1/l2 Constraint for Sparse Blind Deconvolution: Presented at the 77th EAGE Conference & Exhibition, EAGE. → pages 126

Figueiredo, M. A. T., R. D. Nowak, and S. J. Wright, 2007, Gradient Projection for Sparse Reconstruction: Application to Compressed Sensing and Other Inverse Problems: IEEE Journal of Selected Topics in Signal Processing, **1**, 586–597. → pages 51

Fleischer, G., R. Gorenflo, and B. Hofmann, 1999, On the Autoconvolution Equation and Total Variation Constraints: ZAMM, **79**, 149–159. → pages 85

Fleischer, G., and B. Hofmann, 1996, On inversion rates for the autoconvolution equation: Inverse Problems, **12**, 419–435. → pages 85

Fokkema, J. T., and P. M. Van den Berg, 1990, Removal of surface-related wave phenomena: The marine case: SEG Technical Program Expanded Abstracts, 1689–1692. → pages 20

———, 1993, Seismic applications of acoustic reciprocity: Elsevier Science. → pages 12, 20

Foster, D. J., and C. C. Mosher, 1992, Suppression of multiple reflections using the Radon transform: Geophysics, **57**, 386–395. → pages 15

Frijlink, M. O., R. G. van Borselen, and W. Söllner, 2011, The free surface assumption for marine data-driven demultiple methods: Geophysical Prospecting, **59**, 269–278. → pages 18, 20, 110

Fuchs, J., 2005, Recovery of Exact Sparse Representations in the Presence of Bounded Noise: IEEE Transactions on Information Theory, **51**, 3601–3608. → pages 47

Guitton, A., and D. J. Verschuur, 2004, Adaptive subtraction of multiples using the L1-norm: Geophysical Prospecting, **52**, 27–38. → pages 22, 23, 34

Hadidi, M. T., M. Sabih, D. E. Johnston, and C. Calderón-Macas, 1999, Mobil's results for the 1997 workshop on multiple attenuation: The Leading Edge, **18**, 100–103. → pages 22

Hale, E. T., W. Yin, and Y. Zhang, 2008, Fixed-Point Continuation for L1-Minimization: Methodology and Convergence: SIAM Journal on Optimization, **19**, 1107–1130. → pages 36, 47

Hampson, D., 1986, Inverse Velocity Stacking for Multiple Elimination: Journal of the Canadian Society of Exploration Geophysicists, **22**, 44–45. → pages 15

Hennenfent, G., L. Fenelon, and F. J. Herrmann, 2010, Nonequispaced curvelet transform for seismic data reconstruction: A sparsity-promoting approach: Geophysics, **75**, WB203–WB210. → pages 74

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**, A23. → pages 36, 84, 139

Herrmann, F. J., 2005, Seismic deconvolution by atomic decomposition: a parametric approach with sparseness constraints: Integrated Computer-Aided Engineering, **12**, 69–90. → pages 46

———, 2010, Randomized sampling and sparsity: Getting more information from fewer samples: Geophysics, **75**, WB173–WB187. → pages 52

Herrmann, F. J., U. Boeniger, and D. J. Verschuur, 2007, Non-linear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, 781–799. → pages 22

Herrmann, F. J., Y. A. Erlangga, and T. T. Y. Lin, 2009, Compressive simultaneous full-waveform simulation: Geophysics, **74**, A35. → pages 60, 70

Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248. → pages 74

Herrmann, F. J., D. Wang, and D. J. Verschuur, 2008, Adaptive curvelet-domain primary-multiple separation: Geophysics, **73**, A17–A21. → pages 22

Herrmann, P., T. Mojesky, M. Magesan, and P. Hugonnet, 2000, De-aliased, high-resolution radon transforms: SEG Technical Program Expanded Abstracts, 1953–1956. → pages 15

Jumah, B., and F. J. Herrmann, 2011, Dimensionality-reduced estimation of primaries by sparse inversion: SEG Technical Program Expanded Abstracts, **30**, 3520. → pages 70

Kabir, M. N., and D. Verschuur, 1995, Restoration of missing offsets by parabolic Radon transform1: Geophysical Prospecting, **43**, 347–368. → pages 74

Kaplan, S. T., and K. A. Innanen, 2008, Adaptive separation of free-surface multiples through independent component analysis: Geophysics, **73**, V29–V36. → pages 22

Kelamis, P. G., E. F. Chiburis, and S. Shahryar, 1990, Radon multiple elimination, a practical methodology for land data. → pages 15

Kennett, B. L. N., 1979, The suppression of surface multiples on seismic records: Geophysical Prospecting, **27**, 584–600. → pages 20

Kreimer, N., A. Stanton, and M. D. Sacchi, 2013, Tensor completion based on nuclear norm minimization for 5D seismic data reconstruction: Geophysics, **78**, V273–V284. → pages 74

Kumar, R., C. D. Silva, O. Akalin, A. Y. Aravkin, H. Mansour, B. Recht, and F. J. Herrmann, 2015, Efficient matrix completion for seismic data reconstruction: Geophysics. → pages 74

Levin, S. A., 1987, Deconvolution with spatial constraints: PhD thesis, Stanford University, Stanford, CA. → pages 19

Li, X., A. Y. Aravkin, T. van Leeuwen, and F. J. Herrmann, 2012, Fast randomized full-waveform inversion with compressive sensing: Geophysics, **77**, A13–A17. → pages 85

Li, Z.-x., and W.-k. Lu, 2013, Adaptive multiple subtraction based on 3D blind separation of convolved mixtures: Geophysics, **78**, V251–V266. → pages 22

Lin, T. T. Y., and F. J. Herrmann, 2009, Unified compressive sensing framework for simultaneous acquisition with primary estimation: SEG Technical Program Expanded Abstracts, SEG, 3113–3117. → pages 70

———, 2011, Robust source signature deconvolution and the estimation of primaries by sparse inversion: SEG Technical Program Expanded Abstracts, 4354. → pages 126

Lokshtanov, D., 1999, Multiple suppression by data-consistent deconvolution: The Leading Edge, **18**, 115–119. → pages 15

———, 2000, Suppression of water-layer multiples—from deconvolution to wave-equation approach: SEG Technical Program Expanded Abstracts, 1981–1984. → pages 19

Lu, W.-k., and L. Liu, 2009, Adaptive multiple subtraction based on constrained independent component analysis: Geophysics, **74**, V1–V7. → pages 22

Ma, J., 2013, Three-dimensional irregular seismic data reconstruction via low-rank matrix completion: Geophysics, **78**, V181–V192. → pages 74

Majdański, M., C. Kostov, E. Kragh, I. Moore, M. Thompson, and J. Mispel, 2011, Attenuation of free-surface multiples by up/down deconvolution for marine towed-streamer data: Geophysics, **76**, V129–V138. → pages 23

Malioutov, D., M. Cetin, and A. Willsky, 2005, Homotopy continuation for sparse signal representation: IEEE International Conference on Acoustics, Speech, and Signal Processing, IEEE, 733–736. → pages 48

Matson, K. H., and D. Corrigan, 2000, A 2.5D method for attenuating free surface multiples based on inverse scattering series: Presented at the 32nd Annual Offshore Technology Conference. → pages 22

Metkemeijer, R. A. A. P. . A. B., 2002, The acoustics of the auditorium of the Royal Albert Hall before and after redevelopment: Presented at the Proceedings of the Institute of Acoustics. → pages 1

Neelamani, R., A. Baumstein, and W. Ross, 2008, Complex curvelet-based adaptive subtraction of several noise templates: SEG Technical Program Expanded Abstracts, 3650–3655. → pages 22

Nekut, A. G., 1998, 3D surface-related multiple prediction: SEG Technical Program Expanded Abstracts, 1511–1514. → pages 22

Oldenburg, D. W., 1981, A comprehensive solution to the linear deconvolution problem: Geophysical Journal International, **65**, 331–357. → pages 29

Osborne, M. R., B. Presnell, and B. A. Turlach, 1999, On the LASSO and Its dual: Journal Of Computational And Graphical Statistics, **9**, 319–337. → pages 141

Paige, C. C., and M. A. Saunders, 1982, LSQR: An algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **8**, 43–71. → pages 51

Peacock, K. L., and S. Treitel, 1969, Predictive deconvolution: theory and practice: Geophysics, **34**, 155–169. → pages 14

Riley, D. C., and J. F. Claerbout, 1976, 2-D multiple reflections: Geophysics, **41**, 592–620. → pages 18, 20

Robinson, E. A., 1954, Predictive decomposition of time series with applications to seismic exploration: PhD thesis, MIT, Cambridge, MA. → pages 14

———, 1957, Predictive decomposition of seismic traces: Geophysics, **22**, 767–778. → pages 28

Ryu, J. V., 1982, Decomposition (DECOM) approach applied to wave field analysis with seismic reflection records: Geophysics, **47**, 869–883. → pages 15

Sacchi, M. D., and T. J. Ulrych, 1995, High-resolution velocity gathers and offset space reconstruction: Geophysics, **60**, 1169–1177. → pages 15

Savels, T., K. de Vos, and J. W. de Maag, 2010, Surface Multiple Attenuation Through Sparse Inversion – Attenuation Results for Complex Synthetics and Real Data: Presented at the 72nd EAGE Conference & Exhibition. → pages 26

Schmidt, M., G. Fung, and R. Rosales, 2007, Fast optimization methods for L1 regularization: A comparative study and two new approaches: ECML '07 Proceedings of the 18th European conference on Machine Learning, Springer Berlin Heidelberg, 286–297. → pages 141

Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, 55–58. → pages 22

Taner, M., 1980, Long period sea-floor multiples and their suppression: Geophysical Prospecting, **28**, 30–48. → pages 15

Taner, M., R. F. O'Doherty, and F. Koehler, 1995, Long period multiple suppression by predictive deconvolution in the x-t domainl1: Geophysical Prospecting, **43**, 433–468. → pages 15

Taylor, H. L., S. Bank, and J. McCoy, 1979, Deconvolution with the L1 norm: Geophysics, **44**, 39. → pages 46

Thorson, J. R., and J. F. Claerbout, 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, **50**, 2727–2741. → pages 15

Tibshirani, R., 1996, Regression shrinkage and selection via the lasso: Journal of the Royal Statistical Society Series B Methodological, **58**, 267–288. → pages 83, 141

Tropp, J. A., 2006, Just relax: convex programming methods for identifying sparse signals in noise: IEEE Transactions on Information Theory, **52**, 1030–1051. → pages 47

Ulrych, T. J., and C. Walker, 1982, Analytic minimum entropy deconvolution: Geophysics, **47**, 1295. → pages 29

van Borselen, R. G., J. T. Fokkema, and P. M. van den Berg, 1996, Removal of surface-related wave phenomena—The marine case: Geophysics, **61**, 202–210. → pages 20, 23

van Dedem, E. J., and D. J. Verschuur, 2005, 3D surface-related multiple prediction: A sparse inversion approach: Geophysics, **70**, V31. → pages 22, 52

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 xi, 35, 36, 48, 50, 51, 70, 84, 139, 141

———, 2011, Sparse optimization with least-squares constraints: SIAM Journal on Optimization, **21**, 1201. → pages 35, 141

van der Baan, M., and D.-T. Pham, 2008, Robust wavelet estimation and blind deconvolution of noisy surface seismics: Geophysics, **73**, V37. → pages 29

van Groenestijn, G. J. A., and W. Ross, 2011, Primary estimation on OBC data by sparse inversion: SEG Technical Program Expanded Abstracts, **30**, 3531. → pages 126

van Groenestijn, G. J. A., and D. Verschuur, 2011, Using surface multiples to estimate primaries by sparse inversion from blended data: Geophysical Prospecting, **59**, 10–23. → pages 70, 126

van Groenestijn, G. J. A., and D. J. Verschuur, 2008, Towards a new approach for primary estimation: SEG Technical Program Expanded Abstracts, **27**, 2487. → pages 23, 26, 34, 45

———, 2009a, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, A23–A28. → pages 5, 6, 23, 26, 34, 36, 41, 44, 54, 74, 75, 77, 82, 91, 122

———, 2009b, Estimation of primaries and near-offset reconstruction by sparse inversion: Marine data applications: Geophysics, **74**, R119. → pages 26, 56, 68

———, 2010, Incorporating the source array into primary estimation: Presented at the 72nd EAGE Conference & Exhibition. → pages 41

Verschuur, D. J., 1991, Surface-related multiple elimination, an inversion approach: PhD thesis, TU Delft. → pages 21

———, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166. → pages 22

———, 2006, Seismic multiple removal techniques: Past, present and future: EAGE Publications. → pages 13, 75

Verschuur, D. J., and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, Part II: Practical aspects and examples: Geophysics, **62**, 1596–1611. → pages 22

Verschuur, D. J., and R. J. Prein, 1999, Multiple removal results from Delft University: The Leading Edge, **18**, 86–91. → pages 22

Wang, D., R. Saab, O. Yilmaz, and F. J. Herrmann, 2008, Bayesian-signal separation by sparsity promotion: application to primary-multiple separation: Technical Report 5, UBC Earth and Ocean Sciences Department. → pages 22

Wang, Y., 2004, Multiple prediction through inversion: A fully data-driven concept for surface-related multiple attenuation: Geophysics, **69**, 547–553. → pages 23

Wason, H., F. J. Herrmann, and T. T. Y. Lin, 2011, Sparsity-promoting recovery from simultaneous data: A compressive sensing approach: SEG Technical Program Expanded Abstracts, **30**, 6. → pages 70

Weglein, A. B., 1999, Multiple attenuation: an overview of recent advances and the road ahead: The Leading Edge, **18**, 40–44. → pages 13

Weglein, A. B., F. V. Araújo, P. M. Carvalho, R. H. Stolt, K. H. Matson, R. T. Coates, D. Corrigan, D. J. Foster, S. A. Shaw, and H. Zhang, 2003, Inverse scattering series and seismic exploration. → pages 20, 80

Weglein, A. B., F. A. Gasparotto, P. M. Carvalho, and R. H. Stolt, 1997, An inverse-scattering series method for attenuating multiples in seismic reflection data: Geophysics, **62**, 1975–1989. → pages 20, 80

Wiggins, J. W., 1988, Attenuation of complex water-bottom multiples by wave-equation-based prediction and subtraction: Geophysics, **53**, 1527–1539. → pages 19

Wiggins, R. A., 1978, Minimum entropy deconvolution: Geoexploration, **16**, 21–35. → pages 29

Yilmaz, O., 1989, Velocity-stack processing: Geophysical Prospecting, **37**, 357–382. → pages 15

You, Y. L., and M. Kaveh, 1996, A regularization approach to joint blur identification and image restoration.: IEEE Transactions on Image Processing, **5**, 416–428. → pages 42

Yuan, G.-X., K.-W. Chang, C.-J. Hsieh, and C.-J. Lin, 2010, A Comparison of Optimization Methods and Software for Large-scale L1-regularized Linear Classification: The Journal of Machine Learning Research, **11**, 3183–3234. → pages 47

Ziolkowski, A. M., D. B. Taylor, and R. G. K. Johnston, 1999, Marine seismic wavefield measurement to remove sea-surface multiples: Geophysical Prospecting, **47**, 841–870. → pages 21, 23

Ziolkowski, A. M., D. J. Taylor, and R. G. K. Johnston, 1998, Multiple wavefields: Separating incident from scattered, up from down, and

primaries from multiples: SEG Technical Program Expanded Abstracts, 1499–1502. → pages 21

# Appendix A

# Pareto root-finding for the minimum $\ell_1$-norm solution

The Pareto curve of a basis pursuit denoising (BPDN) problem characterizes its optimal solutions; it relates any choice of a target misfit $\sigma$ to its corresponding optimal $\ell_1$-norm objective. Because the BPDN problem is convex, this relation is bijective. The most convenient way to express the Pareto curve for our purposes is to write the optimal misfit $\sigma(\tau)$ of a given optimization problem as a function of $\tau$ the $\ell_1$-norm constraint imposed on the unknown. In other words, we define

$$\sigma(\tau) := \underset{\mathbf{x}}{\text{minimize}} \, \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau. \tag{A.1}$$

Note that this form is natural for describing Lasso problems, where $\sigma(\tau)$ is the optimal objective value for a given $\ell_1$-norm ball constraint $\tau$ . See Figure A.1a for an illustration of the Pareto curve in a 2d plane of possible solutions characterized with $\|\mathbf{x}\|_1$ on the horizontal axis and $\|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$ on the vertical axis.

The idea of looking at the Pareto curve for $\ell_1$-norm minimization is not new. It was described in van den Berg and Friedlander (2008) and Daubechies et al. (2008) as the basis of novel solvers, and in Hennenfent et al. (2008) as a way to characterize the various optimization routines for geophysical compressive sensing problems. Specifically, van den Berg and Friedlander (2008) proved that the Pareto curve as defined here (as a function of $\tau$) is convex, decreasing, and continuously differentiable for $\tau \in [0, \tau^*]$, where

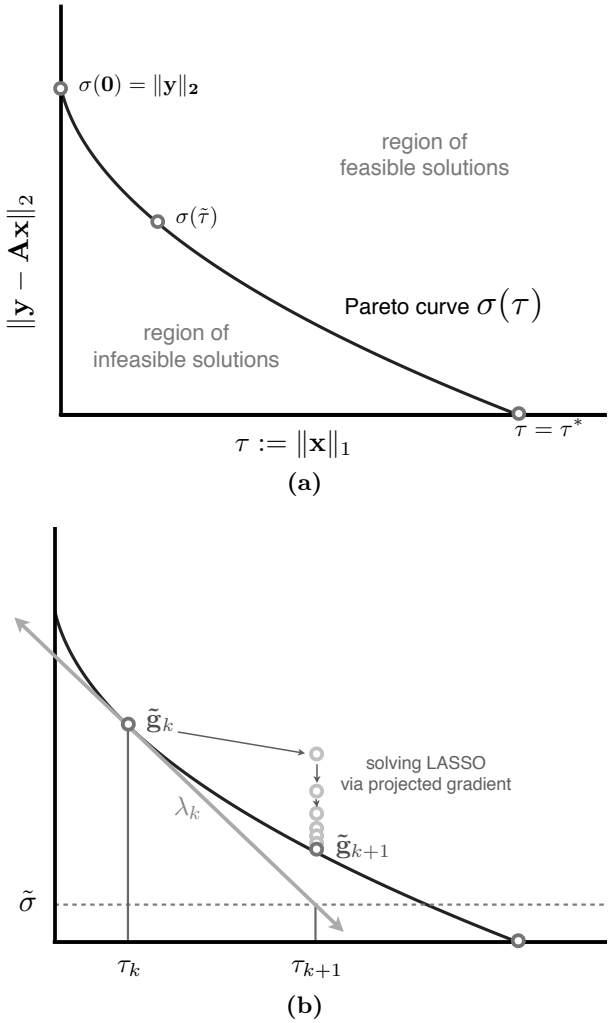



**Figure A.1** Illustrations of the Pareto curve. Panel **(a)** shows the Pareto curve in the parameter space of possible solutions. Panel **(b)** depicts one iteration of the root-finding procedure to find the minimum $\ell_1$ norm that gives the target energy misfit $\tilde{\sigma}$.

$\tau^*$ is the optimal objective of the BP problem $\min_{\mathbf{x}} \|\mathbf{x}\|_1$ s.t. $\mathbf{y} = \mathbf{A}\mathbf{x}$. Furthermore, given an optimal solution $\tilde{\mathbf{x}}$ at any point on the Pareto curve, the slope $\lambda$ of the Pareto curve at that point is given by the closed form expression $\lambda = -\|\mathbf{A}^H\mathbf{r}\|_\infty / \|\mathbf{r}\|_2$, where $\mathbf{r}$ is the residual vector $\mathbf{y} - \mathbf{A}\tilde{\mathbf{x}}$, and

the $\ell_\infty$ norm is equal to the largest absolute value amongst all the elements of the vector.

These properties allow us to find the optimal $\ell_1$-norm $\tilde{\tau}$ for any BPDN problem, by using Newton's method to solve $\sigma(\tilde{\tau}) - \tilde{\sigma} = 0$ (van den Berg and Friedlander, 2008, 2011). Refer to Figure A.1b for a depiction of this procedure. Starting with any solution that lies on the Pareto curve $\tilde{\mathbf{x}}_k$, with $\ell_1$-norm of $\tau_k$ and mismatch $\sigma_k$ (I overload the iteration counter $k$ to keep track of the Newton's method iterations), we iterate over subsequent guesses of the root using the rule $\tau_{k+1} = \tau_k + \Delta\tau_k$ where $\Delta\tau_k = -(\sigma_k - \tilde{\sigma})/\lambda_k$.

In the typical case where noise will cause the observation $\mathbf{y}$ to not be in the range of $\mathbf{A}$, the Pareto curve will not reach $\sigma = 0$, but will instead have an asymptote at the least-squares misfit $\sigma^* = \min \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2$ (van den Berg and Friedlander, 2011). We therefore need to make sure that our target residual $\tilde{\sigma}$ satisfies $\tilde{\sigma} > \sigma^*$. A pragmatic approach to this condition is to monitor $\lambda^{-1}$ through the Newton iterations and stop the inversion when its magnitude becomes unreasonably large, which would indicate the asymptote at $\sigma^*$.

Since obtaining the slope in the Newton root-finding scheme requires knowing an optimal solution $\tilde{\mathbf{x}}$ on the Pareto curve for each of the $\ell_1$-norm restrictions $\tau_k$, its success hinges on the ability to quickly solve problems of the form

$$\underset{\mathbf{x}}{\text{minimize}} \, \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2 \quad \text{subject to} \quad \|\mathbf{x}\|_1 \leq \tilde{\tau}. \tag{A.2}$$

This is typically called a Lasso problem, which differentiates itself from BPDN by an interchanged role between the objective and the regularizing metric, resulting in an objective function is actually differentiable (Tibshirani, 1996; Osborne et al., 1999). Because of this differentiability, a Lasso problem can be solved much faster than BPDN using projected gradient methods (van den Berg and Friedlander, 2008; Schmidt et al., 2007; Daubechies et al., 2008). Furthermore, the algorithm used to solve each Lasso problem can be initialized with the solution of the previous Lasso problem, which distributes the burden of descending the misfit and ensures that each Lasso problem does not require increasingly larger number of iterations. A successive of Lasso problems solved in this way, with the series of $\ell_1$-norm restrictions $\tau_k$ prescribed by the Newton's root-finding method until the optimal $\tau$ is found for the target misfit, forms a continuation-based BPDN solver that was introduced in van den Berg and Friedlander (2008).