

**Least-squares Migration and Full Waveform Inversion
with Multisource Frequency Selection**

Dissertation by
Yunsong Huang

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

King Abdullah University of Science and Technology, Thuwal,
Kingdom of Saudi Arabia

September 2013

The dissertation of Yunsong Huang is approved by the examination committee:

Craig Douglas

Committee Member

Signature

Date

Huimin Guan

Committee Member

Signature

Date

Sherif Mahmoud

Committee Member

Signature

Date

Ying Wu

Committee Member

Signature

Date

Gerard T. Schuster

Committee Chair

Signature

Date

ABSTRACT

Least-squares Migration and Full Waveform Inversion with Multisource Frequency Selection

Yunsong Huang

Multisource Least-Squares Migration (LSM) of phase-encoded supergathers has shown great promise in reducing the computational cost of conventional migration. But for the marine acquisition geometry this approach faces the challenge of erroneous misfit due to the mismatch between the limited number of live traces/shot recorded in the field and the pervasive number of traces generated by the finite-difference modeling method. To tackle this mismatch problem, I present a frequency selection strategy with LSM of supergathers. The key idea is, at each LSM iteration, to assign a unique frequency band to each shot gather, so that the spectral overlap among those shots—and therefore their crosstalk—is zero. Consequently, each receiver can unambiguously identify and then discount the superfluous sources—those that are not associated with the receiver in marine acquisition. To compare with standard migration, I apply the proposed method to 2D SEG/EAGE salt model and obtain better resolved images computed at about 1/8 the cost; results for 3D SEG/EAGE salt model, with Ocean Bottom Seismometer (OBS) survey, show a speedup of 40×.

This strategy is next extended to multisource Full Waveform Inversion (FWI) of supergathers for marine streamer data, with the same advantages of computa-

tional efficiency and storage savings. In the Finite-Difference Time-Domain (FDTD) method, to mitigate spectral leakage due to delayed onsets of sine waves detected at receivers, I double the simulation time and retain only the second half of the simulated records. To compare with standard FWI, I apply the proposed method to 2D velocity model of SEG/EAGE salt and to Gulf Of Mexico (GOM) field data, and obtain a speedup of about 4× and 8×.

Formulas are then derived for the resolution limits of various constituent wavepaths pertaining to FWI: diving waves, primary reflections, diffractions, and multiple reflections. They suggest that inverting multiples can provide some low- and intermediate-wavenumber components of the velocity model not available in the primaries. In addition, diffractions can provide twice or better the resolution as specular reflections for comparable depths of the reflector and diffractor. The width of the diffraction-transmission wavepath is on the order of λ at the diffractor location for the diffraction-transmission wavepath.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Gerard T. Schuster, for his guidance, support and encouragement throughout this work. I draw constant inspiration from his superb teaching skills and his unbridled passion for research. I thank the other members of my dissertation committee, Dr. Craig Douglas, Dr. Huimin Guan, Dr. Sherif Mahmoud and Dr. Ying Wu, for their valuable time.

I am grateful for the financial support from KAUST and from the CSIM sponsors. I appreciate the support from the KAUST Supercomputing Laboratory. I thank Dr. Andreas Atle at Total for the mentorship.

Finally, for their help and discussions, I thank my colleagues, Dr. Ola AlHagan, Abdullah Altheyab, Dr. Chaiwoot Boonyasiriwat, Dr. Wei Dai, Xing Wang, Dr. Han Yu, Dr. Ge Zhan, Dr. Dongliang Zhang and Sanzong Zhang.

TABLE OF CONTENTS

Examination Committee Approval	2
Copyright	3
Abstract	4
Acknowledgements	6
List of Abbreviations	10
List of Symbols	12
List of Figures	14
1 Introduction	16
1.1 Motivation and Overview	16
1.1.1 Multisource Crosstalk	17
1.1.2 Seismic Migration with a Stack of Encoded Shot Gathers	19
1.1.3 Resolution Limits of LSM and FWI	22
1.2 Technical Contributions	23
2 Least-Squares Migration with Multisource Frequency Selection	25
2.1 Introduction	25
2.2 Theory	28
2.2.1 Phase Encoding	29
2.2.2 Frequency Selection Encoding	30
2.2.3 Eliminating Marine Acquisition Crosstalk	33
2.3 Method	37
2.3.1 Multisource Objective Function	37
2.3.2 Migration Method	40
2.4 Results	41
2.5 Conclusions	49

3	Full Waveform Inversion with Multisource Frequency Selection on Marine Streamer Data	51
3.1	Introduction	51
3.2	Theory	53
3.3	Numerical Results	57
3.3.1	Synthetic Data	57
3.3.2	GOM Streamer Dataset	59
3.4	Conclusions	67
4	Resolution Limits for Wave Equation Imaging	74
4.1	Introduction	74
4.1.1	Resolution Limits for Traveltime Tomography	75
4.1.2	Resolution Limits for Reflection Imaging	77
4.2	Born Forward and Adjoint Modeling	81
4.2.1	Born Forward Modeling	81
4.2.2	Born Adjoint Modeling	82
4.3	Model Resolution Function and FWI Resolution Limits	83
4.3.1	Model Resolution Equation: $\mathbf{m}^{\text{mig}} = \mathbf{L}^+ \mathbf{L} \mathbf{m}$	83
4.3.2	λ Imaging at the Diffractor	90
4.4	Filling in the Model Spectrum with Multiples	91
4.4.1	Lower Wavenumber Resolution with Prism Waves and Free-surface Multiples	93
4.4.2	Intermediate Wavenumber Resolution with Interbed Multiples	94
4.5	Discussion and Conclusions	95
5	Concluding Remarks	98
5.1	Summary	98
5.2	Future Research Work	99
	Appendices	101
A	The Effect of Frequency Selection on the Hessian	101
B	Prestack Split-Step Migration	102
C	The Relative Computational Cost	105
D	Migration versus Successive Steepest Descent	109
E	Resolution Properties of an Ellipse	113
F	Resolution Limits for Imaging Diving Wave Residuals	115
G	Determinant of a Jacobian Matrix	119

H Publications Based on This Dissertation 121

References **129**

LIST OF ABBREVIATIONS

CG	Conjugate Gradient
CIG	Common Image Gather
COG	Common Offset Gather
CSG	Common Shot Gather
DC	Direct Current
DTFT	Discrete-time Fourier transform
FDM	Frequency-Division Multiplexing
FDTD	Finite-Difference Time-Domain
FLOP	FLoating-point OPeration
FWI	Full Waveform Inversion
FWIMFS	Full Waveform Inversion with Multisource Frequency-Selection
GOM	Gulf Of Mexico
IS	Iterative Stacking
LSM	Least-Squares Migration
LSMMFS	Least-Squares Migration with Multisource Frequency-Selection
MFS	Multisource Frequency-Selection
OBS	Ocean Bottom Seismometer
RTM	Reverse Time Migration

SNR Signal-to-Noise Ratio

LIST OF SYMBOLS

\mathbf{d}^{obs}	Observed data
\mathbf{d}^{pred}	Predicted data
\mathbf{d}_s	Data (gather) associated with source s
j	Angular frequency index
K_{it}	Number of iterative updates
λ	Wavelength
\mathbf{L}_s	Prestack modeling operator associated with source s
\mathbf{m}	A reflectivity or slowness model (gridded), depending on the context
M_{ga}	Number of supergathers
n_{FLOP}	FLOP count
n_{h}	Number of receivers associated with a source ('h' stands for hydrophone)
n_{ω}	Number of discretized frequencies
N_s	Encoding function for source s
n_t	Number of time steps for one run of simulating wavefield propagation
n_x	Number of grid points along the x-axis
n_y	Number of grid points along the y-axis
n_z	Number of grid points along the z-axis, the depth
ω	Angular frequency

S	Number of sources included in a supergather
S_{tot}	Total number of sources
$\tilde{\circ}$	Encoded version of \circ
$W_s(\omega)$	Spectrum of source s

LIST OF FIGURES

1.1	Overview of multisource frequency selection.	21
2.1	Frequency division of marine sources.	34
2.2	SEG/EAGE salt model and a CSG from it.	41
2.3	Normalized model error and objective function over iterations.	43
2.4	Migration images obtained by various methods and settings.	44
2.5	Speedup by the proposed over the conventional method.	46
2.6	3D SEG/EAGE salt velocity model.	47
2.7	Migration images of the 3d model by alternative methods.	48
3.1	The solution to the acquisition misfit problem.	54
3.2	Doubling the simulation time of FDTD to mitigate the transient.	55
3.3	Transient-reduction in forming the gradient of FWI.	56
3.4	FWI of alternative methods on the SEG/EAGE salt model.	58
3.5	Normalized waveform and velocity errors over iterations.	58
3.6	The direct wave and free-surface reflection combine together to give a detected signal.	60
3.7	The estimated source wavelet and spectral division.	60
3.8	Frequency-selection encoding results.	61
3.9	FWI results for the GOM dataset.	65
3.10	Difference of FWI velocity models obtained using the two frequency- selection encoding means.	65
3.11	Reduction of data misfit over FWI iterations.	66
3.12	RTM images of the tomograms.	69
3.13	Zoomed views of the RTM images.	70
3.14	Shot domain Common Image Gatherers (CIGs).	71
3.15	Common Shot Gatherers (CSGs) of various approaches for GOM data of one source at $X = 6$ km.	72
3.16	CSGs of various approaches for GOM data of one source at $X = 12$ km.	73

4.1	First-Fresnel zones for the specular reflection and for the transmission arrival.	76
4.2	Intersection of wavepaths.	77
4.3	Migration amounts to smearing and summation.	78
4.4	Migration-data kernels, wavepaths, and approximate resolution limits.	79
4.5	Model wavenumber for diffraction migration.	80
4.6	Reflection and transmission wavepaths that visit a scatterer.	86
4.7	Model Wavenumber for Transmission Migration.	87
4.8	Wavepaths for migrating diffraction and specular reflection.	89
4.9	Diffraction migration recovers higher wavenumbers than specular transmission migration does.	92
4.10	Prism wavepath for a vertical reflector model.	93
4.11	Wavepaths for free-surface multiple and prism wave.	94
4.12	Interbed multiples with a diffractor.	95
B.1	Flowcharts for prestack split-step modeling and migration.	103
D.1	Scatter plot of the model errors for migration and for SSD.	111

Chapter 1

Introduction

1.1 Motivation and Overview

Oil and gas exploration nowadays typically demands tremendous amount of computation. Consider for example a 3D velocity model with a moderate grid size $n_x \times n_y \times n_z = 1000 \times 1000 \times 500$, with $S_{\text{tot}} = 100 \times 100$ shots and $n_t = 4000$ time steps for one run of wavefield propagation. Suppose one Reverse Time Migration (RTM) (Baysal et al., 1983; Whitmore, 1983; McMechan, 1983) requires three such runs, and the n_{FLOP} per grid point per time step is 25, then the total Floating-point Operation (FLOP) count is $n_{\text{FLOP}} = n_x \times n_y \times n_z \times n_t \times S_{\text{tot}} \times 3 \times 25 = 1.5 \cdot 10^{18}$, or 1.5 exa. Factor in a dozen iterations and more complexity per iteration due to elaborate optimization schemes, as encountered typically in FWI (Lailly, 1984; Tarantola, 1984, 2005), the total n_{FLOP} for this moderately sized model presents a considerable computational cost.

One approach to reduce this cost is known as the multisource technique (Morton and Ober, 1998; Romero et al., 2000; Krebs et al., 2009; Virieux and Operto, 2009; Dai and Schuster, 2009; Tang, 2009), whereby every S sources are active (deemed a supershot) in one finite-difference simulation and every S shot gathers are correspondingly blended into one supergather. This blending of shot gathers is also done for the many shot gathers recorded in the field, which are recorded one shot

gather at a time. Consequently, the number of effective gathers to cope with is cut down by S -fold. The downside of this technique, however, is crosstalk noise that tarnishes the reconstructed model image. Think of a similar task of determining the acoustic property of a lecture hall, while hosting a cocktail party!

1.1.1 Multisource Crosstalk

Crosstalk in the communications industry is defined in the online Merriam-Webster dictionary as “unwanted signals in a communication channel (as in a telephone, radio, or computer) caused by transference of energy from another circuit (as by leakage or coupling).” As we sometimes discover with analog phones, crosstalk can take the form of hearing someone else’s conversation $s(t)_2$ instead of hearing the intended signal $s(t)_1$. Mathematically the noisy signal $s(t)$ can be represented by

$$s(t) = s(t)_1 + s(t)_2, \quad (1.1)$$

where additive random noise can also be considered.

Sometimes, $s(t)_2$ is also considered signal, and the efficient transmission of both signals requires that they both be simultaneously transmitted, but should be separated from one another after recording¹. To achieve this feat, the two signals are encoded and summed together to give $s(t)' = F[s(t)_1] + F[s(t)_2]$ prior to transmission, and then decoded to get the separate signals $F_1^{-1}[s(t)'] = s(t)_1$ and $F_2^{-1}[s(t)'] = s(t)_2$. This efficient means of simultaneously transmitting two signals and their subsequent decoding is a monumental achievement because it greatly reduces both the cost and the number of channels required for today’s

¹This process of combining several input information signals into one output signal is known as multiplexing in the communications industry. Recovering the individual signals from the multiplex signal is known as demultiplexing.

multi-channel communication.

The decoder F_i^{-1} is a device or operation² that does the reverse of the i^{th} channel encoder, undoing the encoding so that the original information can be retrieved. An example in digital electronics is where a decoder can take the form of a multiple-input, multiple-output logic circuit that converts coded inputs into coded outputs. Decoding is also necessary in applications such as data multiplexing, which is seen in the oil industry with transmission of multichannel seismic data.

A more geophysically relevant model of crosstalk is to include correlation terms that must be eliminated, i.e.,

$$s(t) = s(t)_1 + s(t)_2 + \alpha s(t)_1 \star s(t)_2, \quad (1.2)$$

where α is a scalar weighting term, and \star denotes correlation. For our purposes, one might think of two shot gathers summed together and *simultaneously* migrated to give the desirable sum of the individual migration images $s(t)_1$ and $s(t)_2$, and the undesirable crosstalk noise $\alpha s(t)_1 \star s(t)_2$. Similar to the benefits in the communications industry, Morton and Ober (1998) and Romero et al. (2000) tested the possibility of simultaneously migrating a sum of encoded shot gathers (herewith known as a supergather) to tremendously reduce computation time and memory+I/O demands. Unfortunately, their early results did not show significant efficiencies because their method did not easily annihilate the crosstalk term $s(t)_1 \star s(t)_2$.

In fact, decoding is mathematically simpler in the frequency domain where equation 1.2 at ω is given by

$$S(\omega) = S(\omega)_1 + S(\omega)_2 + \alpha S(\omega)_1^* S(\omega)_2, \quad (1.3)$$

where $S(\omega)_i$ is the Fourier transform of $s(t)_i$. This frequency domain representa-

²This definition is adjusted from that in <http://en.wikipedia.org/wiki/Decoder>.

tion presents the opportunity for decoding encoded signals by Frequency-Division Multiplexing (FDM), where interfering signals are transmitted in several non-overlapping frequency ranges (Bates and Gregory, 2007). The distinct frequency band of each signal means that the spectral product in the above equation is zero. It also means that $S(\omega)_1$ and $S(\omega)_2$ can be recovered by appropriate bandpass filters applied to $S(\omega)$. One of FDM's most common applications is cable television, where different TV channels are FDM encoded and sent over the cable simultaneously. The decoding box at home then separates each channel from one another by decoding. There are many other multiplexing schemes such as time-division, statistical, wavelength-division, orthogonal frequency-division, code-division multiplexing (Bates and Gregory, 2007).

1.1.2 Seismic Migration with a Stack of Encoded Shot Gathers

If there are many signals added together and many crosstalk terms then equation 1.3 can be generalized as

$$S(\omega)' = \sum_i S(\omega)_i + \sum_i \sum_j \alpha_{ij} S(\omega)_i^* S(\omega)_j. \quad (1.4)$$

For seismic migration, equation 1.4 represents the migration image formed by adding many shot gathers together and simultaneously migrating them. Unless the input data is encoded, the crosstalk term is too strong and spoils the image.

To reduce this spoilage Morton and Ober (1998) encoded their shot gathers by replacing the wavelet sources with random times series, summing them together

to form a supergather, and migrating to give the migration image

$$S(\omega)' = \sum_i |N(\omega)_i|^2 S(\omega)_i + \sum_i \sum_j \alpha_{ij} N(\omega)_i N(\omega)_j S(\omega)_i^* S(\omega)_j. \quad (1.5)$$

where the encoding function $N(\omega)_i$ is the spectrum of the i^{th} random time series assigned to the i^{th} source. Each random time series was selected to approximate the orthonormality condition

$$\langle N(\omega)_i N(\omega)_j \rangle = \delta_{ij}, \quad (1.6)$$

where $\langle \rangle$ represents ensemble averaging, which is another way of saying that we should sum supergather migration images together, each with a distinctly encoded supergather as the input data. Many other encoders have been proposed. For example, Schuster et al. (2011) (see also references therein) showed that encoding the sources with both random polarities and random phase shifts is better than encoding with either scheme alone at reducing the crosstalk. To make equation 1.6 approximately hold, more (say a factor $F \times$) iterations are required than the conventional migration. As long as this overhead factor F is less than the reduction factor—the number of sources included in a supergather, this multisource endeavor leads to a speedup.

For the marine acquisition geometry, however, such crosstalk is troublesome. The reason is that when S encoded shots are excited at nearly the same time the computer invariably simulates a supergather as in a land geometry with fixed receivers. In contrast, the actual marine shot gather is for a moving receiver string. An example is shown in Fig. 1.1(c), where the marine moving receiver string is denoted by the horizontal green bar at the top, while the fixed receivers in the

simulated land geometry encompass the full width of the color-coded velocity model. This implies that the computer-generated land shot gathers have too many traces per shot compared to the actual marine shot gather. Consequently, there will be a large residual, $\mathbf{d}^{\text{pred}} - \mathbf{d}^{\text{obs}}$, at the uncommon receiver locations, even if the true velocity model is used in the simulation. Note that if there was no crosstalk, i.e., if every receiver could tell apart its contributing sources, then the problem of erroneous residual would vanish, as will be explained in the next paragraph.

A receiver in a multisource simulation would receive signals from all sources in a supershot, although in the marine case only a subset of sources are associated with this receiver. When free of crosstalk, the receiver can determine what signals stem from the superfluous sources in simulation, and thus it can discount those signals before calculating the residual. In this way, no erroneous residual can arise.

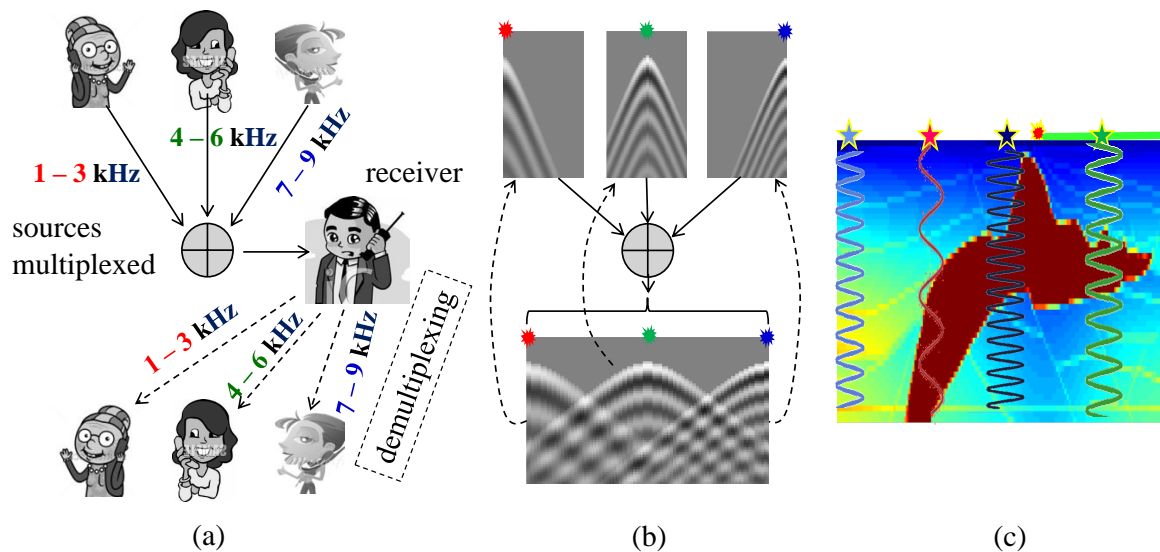


Figure 1.1: Overview of multisource frequency selection. (a) Frequency-division multiplexing/demultiplexing used in telecommunication. (b) Frequency-encoding three shot gathers to form a supergather, which is subsequently disentangled (indicated by the dashed lines) by frequency selection. (c) FWI obtained by multisource frequency selection, with the resulting velocity model color-coded in the background, and sine waves of different frequencies injected at different source locations. Color stars in (b, c) denote sources.

This begs the question whether in multisource seismic inversion there is a

crosstalk-free encoding scheme. Until now, in exploration geophysics no such scheme has been devised. In this dissertation, I have designed a crosstalk-free encoding strategy, similar to the FDM used in the communications industry, as depicted in Fig. 1.1(a). The idea is to assign each shot gather to a unique set of frequencies. Careful assignment ensures no overlap in frequencies from one shot gather to the next, as illustrated in Fig. 1.1(a and b). That is, the source signatures are orthogonal to each other, and so the cross-talk between sources is zero. Using this scheme with marine streamer geometry is illustrated in Fig. 3.1 on page 54, and also in in Fig. 1.1(c). In Chapters 2 and 3, I show how this scheme, called ‘frequency selection’ in this dissertation, can be successfully applied to multisource LSM and multisource FWI, respectively. The benefit is that my numerical results show that for my examples, multisource imaging with frequency selection provides a speedup of more than $8\times$ and $4\times$ compared to conventional migration and FWI, respectively.

1.1.3 Resolution Limits of LSM and FWI

As far as seismic imaging and inversion with the wave equation are concerned, the resolution limits are of fundamental interest, and understanding these limits allows us to economically parameterize the model and design an efficient inversion schedule. Although the resolution limits have been analyzed for traveltimes tomography (Kravtsov and Orlov, 1990; Williamson, 1991; Cerveny and Soares, 1992) and for reflection imaging (Berkhout, 1984; Safar, 1985; Vermeer, 1997; Chen and Schuster, 1999), there lacks a comprehensive treatment of the resolution limits associated with FWI. In this dissertation I conduct a study to fill this gap by carrying out asymptotic analysis to the model resolution function for FWI. The results are summarized in Fig. 4.4 on page 79. Moreover, as noted by Jannane et al. (1989), the absence of intermediate wavenumbers recovered from diffraction and transmission migration poses a serious challenge for waveform inversion. I

address this problem with the aid of diffractor and interbed multiples. The results are summarized in Fig. 4.9 on page 92. This line of work is covered in Chapter 4.

1.2 Technical Contributions

My novel contributions in this dissertation are the following:

- In iterative multisource simulations with marine streamer data, the mismatch between the limited number of live hydrophones in the survey and the pervasive number of live traces generated by modeling is recognized as a form of crosstalk. This crosstalk noise is, for the first time, completely removed by the encoding and decoding scheme of frequency selection.
- This encoding scheme is applied to multisource LSM, regularized by a Cauchy norm in the presence of noise. Relative to conventional migration I show with empirical simulations that multisource LSM with frequency encoding provides a $8\times$ speedup. I also show that if starting from CSG in the time domain, then my iterative method requires preprocessing that incurs $2\times$ the I/O cost of conventional migration. Afterwards my iterative method only requires a minimal fraction, say $1/12$, of the I/O cost of conventional migration.
- While the preceding application uses split-step migration, formulated in the frequency and wavenumber domains, for FWI I adopt the FDTD method, which is preferred by many as the method of choice for 3D FWI. This presents new challenges to the frequency selection scheme, such as spectral leakage, a problem mitigated by doubling the simulation time and retaining only the second half of the simulated records. I show that multisource FWI with frequency encoding achieves a speedup of about $4\times$ and $8\times$ for synthetic and field data compared to the conventional method.
- The resolution limits of various constituent wavepaths pertaining to FWI is

comprehensively studied. For the first time, I show that the gap of missing reconstructed intermediate wavenumbers is partially filled by capitalizing on diffractor and interbed multiples.

- This dissertation has resulted in one journal paper (Huang and Schuster, 2012) published in *Geophysical Prospecting*. A second paper (Huang et al., 2013) is submitted to *Geophysical Journal International*, and a third paper based on this dissertation will soon be submitted for publication. Its associated expanded abstract (Huang and Schuster, 2013) was presented at the EAGE London meeting in June 2013, and was recognized as one of the outstanding papers delivered at the this international meeting. A special letter was sent to me in July 2013 inviting me to publish this work in *Geophysical Prospecting*, the leading applied geophysics journal in Europe. These publications are listed in Appendix H.

Chapter 2

Least-Squares Migration with Multisource Frequency Selection

2.1 Introduction

For large datasets, 3D prestack wave equation migration is a computationally expensive procedure. Its computational workload is proportional to both the number of shots in a survey and the computational complexity of solving the 3D wave equation for a given velocity model. In the case of iterative methods, this workload is proportional to the number of iterations for acceptable convergence.

An attempt to reduce this workload was proposed by Morton and Ober (1998) by migrating one blended supergather, rather than separately migrating individual shot gathers. Here, the supergather is computed by summing a number of shot gathers, each encoded by correlation with a distinct random time series approximately orthogonal with one another. The migration image is then formed by applying a decoded migration operator whose imaging condition is tuned to decode the simultaneous sum of encoded shots. Applying this migration operator to the supergather produces a migration image of good quality only if the number of iterations is sufficiently large. In fact, their results did not show a clear computational cost advantage over the conventional method of wave equation

migration.

To mitigate problems associated with wavelets of long random time series, Jing et al. (2000) and Krebs et al. (2009) proposed a polarity encoder that randomly multiplied shot gathers by either +1 or -1. For phase-encoded multi-source migration, Jing et al. (2000) empirically concluded that the crosstalk term was adequately suppressed when six encoded shot gathers were encoded, summed together, and migrated. On the other hand, Krebs et al. (2009) empirically found that using this strategy with FWI produced acceptable velocity tomograms at a cost saving of at least one order of magnitude. In one of the few exceptions, Gao et al. (2010) used a deterministic encoding to determine a shot's scale factor that gave the most significant update to the velocity model for a specified composite source. Another form of deterministic encoding is plane-wave decomposition (see e.g. Whitmore and Garing, 1993; Duquet et al., 2001; Zhang et al., 2003), which also aims at reducing data volume. Using this method, Vigh and Starr (2008) obtained speedups ranging from three to 10-fold. Other groups, such as Virieux and Operto (2009), Ben-Hadj-Ali et al. (2009, 2011), Dai and Schuster (2009), Boonyasiriwat and Schuster (2010), and Ben-Hadj-Ali et al. (2011) discovered similar cost savings for FWI or least squares migration, except that they used somewhat different encoding recipes such as exclusive use or combinations of random time shifting, frequency selection, source selection, amplitude encoding, and/or spatial randomization of the source locations. A related inversion scheme is by Tang (2009), who used random phase-encoding of simultaneous sources to efficiently compute the Hessian for iterative least squares migration. Almost all of these schemes aimed to efficiently approximate the orthogonality between different encoders in as few iterations as possible.

Is there an encoding scheme that can exactly satisfy this orthogonality condition? The answer is yes. The FDM scheme from the communications industry can

be used to assign each shot gather to a unique set of frequencies. Careful assignment ensures no overlap in frequencies from one shot gather to the next, thereby eliminating the crosstalk. Just as important, FDM also mitigates the acquisition crosstalk noise associated with a marine geometry.

The marine acquisition crosstalk is defined as the migration noise caused by the mismatch in the modeled traces and the recorded traces. In a marine survey the recorded traces are only alive over a moving swath of hydrophones while the generated finite-difference traces are alive everywhere. This induces large residuals in the data misfit functions, leading to large artifacts in the FWI or migration images. As will be discussed later, the FDM strategy eliminates this problem. The downside of this strategy is, however, the reduced resolving power of seismic illumination. To enhance the resolving power, I use the LSM method (Nemeth et al., 1999; Duquet et al., 2000; Tang and Biondi, 2009), varying each shot gather's unique frequency fingerprint at every 3 Conjugate Gradient (CG) updates. The resulting migration algorithm for encoded data can be more than an order of magnitude faster than conventional migration while producing nearly the same image quality.

The rest of this chapter is organized as follows. The theory section presents the theory of frequency-division encoding, how it can be used to remove the crosstalk in migrating supergathers, and the I/O implications for computing systems. The method section, supplemented by appendices, defines the objective function for the frequency-division multisource algorithm, discusses the implications for optimization, and derives the computational complexity. The numerical results for both the 2D and 3D SEG/EAGE salt models are then presented in the numerical results section. Here, the 2D model is used to generate synthetic data emulating a marine survey, and the 3D model is used to test the viability of the proposed technique for 3D data. The final section presents a summary and discussion.

2.2 Theory

I now present the spectral encoding strategy for removing crosstalk artifacts in multisource imaging. First, I identify the source spectrum in the forward modeling equation. Then, I outline a typical phase-encoded multisource procedure, before developing the proposed frequency encoding method.

In the frequency domain a seismic trace with a source at \mathbf{x}_s and a receiver at \mathbf{x} can be expressed (Stolt and Benson, 1986), based on the Born approximation to the Lippman-Schwinger equation, as

$$d(\mathbf{x}|\mathbf{x}_s) = \int G(\mathbf{x}|\mathbf{x}')m_o(\mathbf{x}')G(\mathbf{x}'|\mathbf{x}_s)W_s(\omega) d\mathbf{x}'. \quad (2.1)$$

Here, $G(\mathbf{b}|\mathbf{a})$ denotes the Green's function from \mathbf{a} to \mathbf{b} ; $m_o(\mathbf{x}') \stackrel{\text{def}}{=} s(\mathbf{x}')\delta s(\mathbf{x}')$ is the reflection coefficient-like term at \mathbf{x}' , where $\delta s(\mathbf{x}')$ is the slowness perturbation from an assumed background slowness $s(\mathbf{x}')$; and $W_s(\omega)$ is the spectrum of the s^{th} source weighted by $-2\omega^2$ and can be pulled outside the integral since it is independent of \mathbf{x}' . For conciseness $W_s(\omega)$ is hereafter referred to as 'source spectrum' or simply 'spectrum' for short. As the earth model is discretized into M grid points, equation 2.1 can be recast in matrix-vector form as

$$\mathbf{d}_s = W_s(\omega)\underline{\mathbf{L}}_s \mathbf{m}, \quad \forall s = 1, \dots, S \quad (2.2)$$

which is conventionally expressed as

$$\mathbf{d}_s = \mathbf{L}_s \mathbf{m}, \quad \forall s = 1, \dots, S \quad (2.3)$$

$$\text{where} \quad \mathbf{L}_s = W_s(\omega)\underline{\mathbf{L}}_s. \quad (2.4)$$

Here, $\mathbf{m} \in \mathbb{R}^M$ is the reflectivity model; $\mathbf{d}_s \in \mathbb{C}^{n_h}$ represents the s^{th} shot gather; S is the number of shots; n_h is the number of receivers per shot; $\mathbf{L}_s \in \mathbb{C}^{n_h \times M}$ represents the prestack modeling operator for the s^{th} shot gather, and $\underline{\mathbf{L}}_s$ is \mathbf{L}_s deprived of $W_s(\omega)$.

Equations 2.2 to 2.4 are in the frequency domain and recognize that quantities such as \mathbf{d}_s , \mathbf{L}_s , and \mathbf{L}_s all depend on ω , which is silent to reduce notational clutter; however, ω is explicitly retained in $W_s(\omega)$, because $W_s(\omega)$ represents the proposed frequency encoding function. Note also the subscript in $W_s(\omega)$, implying that different sources may have different spectrums.

2.2.1 Phase Encoding

My frequency encoding scheme will now be developed in the same framework of phase-encoding (Romero et al., 2000), which typically consists of the following three steps. 1) Different shot gathers are uniquely phase encoded. 2) They are summed together to form supergathers, which are then 3) migrated all at once. The first step amounts to multiplying the s^{th} shot gather with a unique phase encoding function N_s , a step expressed as

$$\tilde{\mathbf{d}}_s = N_s \mathbf{d}_s. \quad (2.5)$$

In this dissertation, $\tilde{\circ}$ notes an encoded version of \circ . Then $\tilde{\mathbf{d}}_s$ are summed over all sources to give the encoded supergather $\tilde{\mathbf{d}}$:

$$\tilde{\mathbf{d}} = \sum_{s=1}^S \tilde{\mathbf{d}}_s = \sum_{s=1}^S N_s \mathbf{d}_s, \quad (2.6)$$

$$= \tilde{\mathbf{L}} \mathbf{m}, \quad (2.7)$$

where the multisource phase-encoded prestack modeling operator is defined as

$$\tilde{\mathbf{L}} \stackrel{\text{def}}{=} \sum_{s=1}^S N_s \mathbf{L}_s. \quad (2.8)$$

Finally, the third step involves applying the adjoint operator $\tilde{\mathbf{L}}^\dagger$ to the encoded supergather $\tilde{\mathbf{d}}$ in equation 2.6, before applying the imaging condition, to get the

migrated image $\tilde{\mathbf{m}}$ as

$$\tilde{\mathbf{m}} = \sum_{\omega} \tilde{\mathbf{L}}^{\dagger} \tilde{\mathbf{d}} \quad (2.9)$$

$$= \sum_{\omega} \sum_{s=1}^S \sum_{q=1}^S N_s^* N_q \mathbf{L}_s^{\dagger} \mathbf{L}_q \mathbf{m} \quad (2.10)$$

$$= \hat{\mathbf{m}} + \mathbf{c}, \quad (2.11)$$

where

$$\hat{\mathbf{m}} \stackrel{\text{def}}{=} \sum_{\omega} \sum_{s=1}^S |N_s|^2 \mathbf{L}_s^{\dagger} \mathbf{L}_s \mathbf{m} \quad (2.12)$$

$$= \sum_{\omega} \sum_{s=1}^S \mathbf{L}_s^{\dagger} \mathbf{L}_s \mathbf{m}, \quad (2.13)$$

and

$$\mathbf{c} \stackrel{\text{def}}{=} \sum_{\omega} \sum_{s=1}^S \sum_{q \neq s}^S N_s^* N_q \mathbf{L}_s^{\dagger} \mathbf{L}_q \mathbf{m} \quad (2.14)$$

$$= \sum_{\omega} \sum_{s=1}^S \sum_{q \neq s}^S N_s^* N_q W_s^*(\omega) W_q(\omega) \mathbf{L}_s^{\dagger} \mathbf{L}_q \mathbf{m}. \quad (2.15)$$

Here, $\hat{\mathbf{m}}$ is the sequential shot-gather migration and \mathbf{c} is crosstalk noise. Equation 2.13 follows assuming the phase encoding function N_s is of pure phase so that $N_s^* N_s = 1$, and equation 2.15 follows from equation 2.4.

Note the crosstalk noise, \mathbf{c} , is the only part of $\tilde{\mathbf{m}}$ that depends on the random phase encoding function, over which an ensemble average, denoted by $\langle \cdot \rangle$, is taken to produce

$$\langle \mathbf{c} \rangle = \sum_{\omega} \sum_{s=1}^S \sum_{q \neq s}^S \langle N_s^* N_q \rangle \mathbf{L}_s^{\dagger} \mathbf{L}_q \mathbf{m}. \quad (2.16)$$

2.2.2 Frequency Selection Encoding

While existing approaches such as studied by Schuster et al. (2011) strive to reduce the crosstalk noise by devising phase encoding functions such that $\langle N_s^* N_q \rangle = 0$ for

$s \neq q$, this chapter relies on devising source spectra $W_q(\omega)$'s in order to eliminate the noise term \mathbf{c} defined in equation 2.15. To this end, I single out an arbitrary term in equation 2.15 and investigate how to make it zero. An example of such a term is expressed as

$$\mathbf{c}_{sq} = \sum_{\omega} W_s^*(\omega) \mathbf{v}_{sq}(\omega) W_q(\omega), \quad s \neq q \quad (2.17)$$

where

$$\mathbf{v}_{sq}(\omega) \stackrel{\text{def}}{=} N_s^* N_q \mathbf{L}_s^+ \mathbf{L}_q \mathbf{m}. \quad (2.18)$$

Because the dependence of $\mathbf{v}_{sq}(\omega)$ on ω is typically spatially varying and unknown, it is impossible to construct $W_q(\omega)$'s that can suppress all elements of \mathbf{c}_{sq} , unless the source spectra are non-overlapping. Non-overlapping source spectra ensure that

$$W_s^*(\omega) W_q(\omega) = 0, \quad \text{for } s \neq q, \quad \forall \omega \quad (2.19)$$

and in turn reduce equation 2.15 to zero. I refer to this encoding scheme as frequency selection. The previous analysis contrasts the different roles that the phase encoder N_s and frequency encoder $W_s(\omega)$ play. For notational economy, however, hereafter in the context of frequency selection I recast N_s as a frequency encoder, on which $W_s(\omega)$ is predicated; in addition, ω is discretized, and is identified with a frequency index j running from 1 to n_{ω} . The frequency encoder is given as a binary vector

$$N_s(j) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ frequency belongs to source } s, \\ 0 & \text{otherwise.} \end{cases} \quad (2.20)$$

Note that $N_s(j)$'s are no longer of pure phase; this can be regarded as a form of amplitude encoding (Godwin and Sava, 2010). If no frequency index is shared by

multiple sources, then equation 2.20 leads to

$$N_s(j)N_q(j) = 0, \text{ for } s \neq q, \forall j = 1, \dots, n_\omega. \quad (2.21)$$

Thus equation 2.19 is guaranteed, in my new notation, by the choice

$$W_s(j) = N_s(j)W(j), \forall s = 1, \dots, S, \forall j, \quad (2.22)$$

where $W(j)$ is the intact source spectrum. In shorthand, equation 2.21 can be expressed as

$$\mathbf{N}_s \odot \mathbf{N}_q = \mathbf{0}, \text{ for } s \neq q, \quad (2.23)$$

where \odot represents element-wise multiplication between two vectors. If, moreover, every frequency index is assigned to some source, then equation 2.20 leads to

$$\bigoplus_{s=1}^S \mathbf{N}_s = \mathbf{1}, \quad (2.24)$$

where \oplus represents element-wise addition. Accordingly, following equation 2.22, we have

$$\mathbf{W}_s \odot \mathbf{W}_q = \mathbf{0}, \text{ for } s \neq q, \quad (2.25)$$

$$\bigoplus_{s=1}^S \mathbf{W}_s = \mathbf{W}. \quad (2.26)$$

Given S sources and n_ω frequency indices, frequency selection endeavors to evenly divide the latter among the former. That is, on average each source is assigned n_ω/S frequency indices.

I outline next how n_ω is determined. Suppose the maximal travel time between sources and their associated receivers is T , the peak frequency of the source wavelet is f_0 , and the cutoff high frequency is at $f_{hi} = 2.5f_0$. The Nyquist sampling theorem

dictates $dt < \frac{1}{2f_{hi}} = \frac{1}{5f_0}$, and therefore the total number of time samples is $n_t = T/dt > 5Tf_0$. For real signals devoid of Direct Current (DC), the number of independent angular frequencies is given by

$$n_\omega = 2.5Tf_0. \quad (2.27)$$

For example, the parameters chosen for my 2D and 3D simulations are: $n_\omega = 160$ as $T = 2\text{s}$ and $f_0 = 32\text{Hz}$, and $n_\omega = 360$ as $T = 9\text{s}$ and $f_0 = 16\text{Hz}$, respectively. Note that the effective number of independent frequencies is smaller than n_ω , because the source spectrum $W(j)$ is far from uniform.

2.2.3 Eliminating Marine Acquisition Crosstalk

Once sources have been assigned non-overlapping sets of frequencies, marine acquisition crosstalk can be eliminated. The key idea is, after completing multisource forward modeling by computer simulation, at each receiver h any extraneous frequency component j is pruned; j is considered extraneous if j is assigned to a source, to which, according to the marine geometry, h is not associated. I illustrate the proposed algorithm with an example shown in Figure 2.1. Figure 2.1(a) depicts $S=3$ sources, $n_\omega=5$ frequencies, and a specific way of frequency selection described by the frequency encoders $N_s(j)$'s identified as, from left to right, $N_3 = [1, 0, 0, 0, 1]^T$, $N_2 = [0, 1, 0, 0, 0]^T$, and $N_1 = [0, 0, 1, 1, 0]^T$. Figure 2.1(b) depicts a towed marine geometry, where each source is associated with $n_h = 5$ receivers. For instance, source s_3 is associated with receivers h_5-h_8 , but not with receivers h_1-h_4 . Consider for example $j=5$ at receiver h_4 . Because, according to Figure 2.1(a), $j=5$ is assigned to source s_3 , $j=5$ is considered extraneous at receivers h_4 and should be pruned. The rationale is as follows. When sources are blended, all frequency components are present (see equation 2.26) in forward modeling and consequently at every

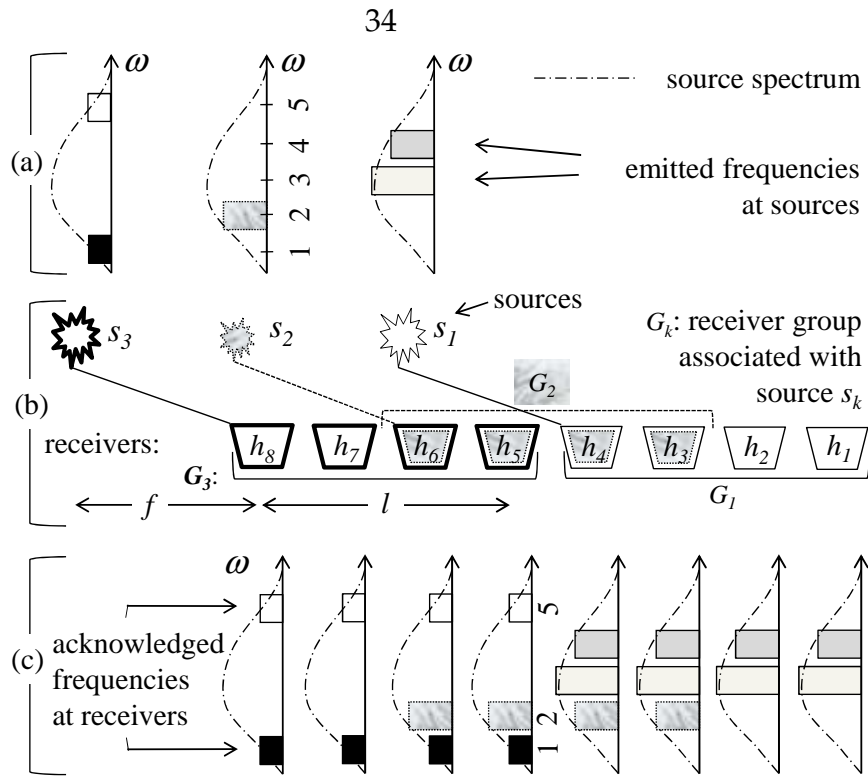


Figure 2.1: Frequency division of sources for one supergather of towed-marine data. Sources and receivers are identified with their indices. (a) Unique spectra assigned to, and hence will be emitted by, the sources. The three spectra patterns are non-overlapping. (b) The association, signified with the same line width and fill style, between sources and their respective receiver groups. f denotes near offset; l denotes line length. (c) Frequencies listened to at each receiver.

receiver. Receiver h_4 would have detected frequency component 5, which comes from source s_3 , but h_4 lies outside the aperture associated to s_3 , and therefore h_4 should not pick up any signal stemming from s_3 . This explains the pruning of the extraneous frequency component 5 at h_4 . This is indicated by the absence of a bar corresponding to $j=5$ at h_4 in Figure 2.1(c). Other unoccupied frequency slots in Figure 2.1(c) are likewise inferred.

The pruning operation is equivalent to selective filling in as follows. Let $F_{\text{blen}}(\text{frequency, receiver})$ of size $n_\omega \times n_{\text{htot}}$ be the outcome³ in the frequency domain detected by receivers generated by forward modeling with blended sources prior to pruning, and let $F_{\text{prun}}(\text{frequency, receiver})$ of the same size be the result of

³in MATLAB notation; likewise for the following arrays in this section

pruning applied to F_{blen} . Here, n_{htot} is the total number of receivers covered by the supergather, and $n_{\text{htot}} = 8$ in this example. F_{prun} is obtained by first initialization with 0 and subsequently filling in with valid entries in F_{blen} ; an entry $F_{\text{blen}}(j, h)$ is valid if frequency component j is not extraneous at receiver h . For instance for $j = 5$, we have

$$F_{\text{prun}}(j = 5, [h_5, h_6, h_7, h_8]) \leftarrow F_{\text{blen}}(j = 5, [h_5, h_6, h_7, h_8]) \quad (2.28)$$

Similarly, the encoded supergather CSG_{enc} , of size $n_\omega \times n_{\text{htot}}$, can be formed as follows, assuming that the observed CSGs have been transformed to the frequency domain and are indexed as $\text{CSG}(\text{frequency}, \text{receiver}, \text{source})$, of size

$$M_{\text{CSG}} = n_\omega \times n_h \times S. \quad (2.29)$$

Here, n_h is the number of receivers associated with each source in acquisition, and $n_h = 4$ in this example. First, $\text{CSG}_{\text{enc}} \leftarrow 0$. Next, fill in CSG_{enc} with the corresponding entries in CSG according to the current frequency encoders. Specifically, loop j over n_ω , and for a given j , find to which source s it belongs, and subsequently find which receivers h 's are associated to this s . Then execute $\text{CSG}_{\text{enc}}(j, h\text{'s}) \leftarrow \text{CSG}(j, :, s)$. An example for $j=3$ is given as

$$\text{CSG}_{\text{enc}}(j = 3, [h_1, h_2, h_3, h_4]) \leftarrow \text{CSG}(j = 3, :, s_1). \quad (2.30)$$

Finally, the misfit function is computed by $F_{\text{prun}} - \text{CSG}_{\text{enc}}$. By pruning or equivalently selective filling in, the mismatch problem between the limited number, n_h , of live traces/shot in observed CSG, and the pervasive number, n_{htot} , of traces in simulation-generated F_{blen} is now resolved.

Note that since there are n_ω equations similar to equation 2.30, each reading n_h

entries, the total number of entries read from CSG by selective filling in is

$$M_{CSG_{enc}} = n_\omega \times n_h, \quad (2.31)$$

$$= \frac{M_{CSG}}{S}. \quad (2.32)$$

In this example $M_{CSG_{enc}} = 5 \times 4 = 20$, coinciding with the number of bars in Figure 2.1(c).

I/O implications

I/O is an important consideration when dealing with industrial-size data sets. Contrary to the naive impression that the I/O cost of the proposed method in K_{it} iterations is K_{it} times that of the standard migration, here I show that the actual I/O cost of the former is only $2 + \epsilon$ times⁴ the latter, assuming the migrations are carried out in the frequency domain while the original data are stored in the time domain. If, however, the data have been transformed into the frequency domain already, then the I/O cost of the former method is only ϵ times the latter.

Let I/O cost be identified with the size⁵ of data passing through I/O, and assume the data size is M_{CSG} . The standard migration entails reading every shot gather, followed by the Fourier transform and then the migration. So the I/O cost is $C_0 = M_{CSG}$. On the other hand, the work flow of the proposed method consists of two stages. (1) Preparation. All input data are read, transformed to the frequency domain and saved to disk. The I/O cost of this stage is $C_1 = 2M_{CSG}$. (2) Migration. The I/O cost per iteration is $M_{CSG_{enc}}$. In K_{it} iterations, the I/O cost is $C_2 = K_{it}M_{CSG_{enc}} = K_{it} \frac{M_{CSG}}{S} = \epsilon M_{CSG}$, where $\epsilon = K_{it}/S \ll 1$ as K_{it} is typically an order of magnitude smaller than S , which is how speedup can be gained by iterative multisource methods. Therefore the I/O cost $C_1 + C_2 = (2 + \epsilon)C_0$ of the proposed method is a

⁴ ϵ is a small fraction, for instance, 1/12.

⁵measured in the number of complex numbers

little more than twice that of the standard approach for any $K_{it} \ll S$. If the data is in the frequency domain already, then the work pertaining to stage (1) is unnecessary. In this case, the I/O cost of the proposed method is only $C_2 = \epsilon C_0$.

If CSG_{enc} can fit in a computer's memory, C_2 can be further reduced as follows. Read the CSGs from disk to form a CSG_{enc} , which is kept in the memory, then make K_{CGit} iterative updates⁶ to the trial model. In this scheme, C_2 is reduced by a factor⁷ of K_{CGit} .

2.3 Method

2.3.1 Multisource Objective Function

Due to frequency division, only a subset of the spectrum will be covered at each source at each iteration, and so ringy migration artifacts are expected. An effective method to reduce migration artifacts (Nemeth et al., 1999; Duquet et al., 2000) is LSM, which works by iteratively updating a trial model in order to minimize a data misfit function. A widely adopted misfit function is the the L_2 norm squared of data error. In addition, regularization with Cauchy norm (Amundsen, 1991; Sacchi, 1997; Wang and Sacchi, 2007) is used in this chapter. In the Bayesian framework (Aster et al., 2005; Debski, 2010), the regularization corresponds to a negative logarithm of the *a priori* distribution of the model. The choice of Cauchy distribution is meant to capture the sparse nature of typical reflectivity models.

⁶Since the Hessian of the objective function is constant given a fixed CSG_{enc} , those K_{CGit} iterations are made using CG.

⁷As K_{CGit} increases, K_{it} may also have to increase in order to produce acceptable result. Therefore this reduction factor is a bit smaller than K_{CGit} .

Following the Bayesian approach, I write the regularization as

$$R(\mathbf{m}) = -\ln p_c(\mathbf{m}) = -\ln \left[\prod_i \frac{c}{\pi(c^2 + m_i^2)} \right] \quad (2.33)$$

$$= \sum_i \ln(c^2 + m_i^2) + \text{constants}, \quad (2.34)$$

where $p_c(\mathbf{m})$ is a 0-median Cauchy distribution with parameter c ; and I write the misfit function as

$$e(\mathbf{m}) = -\ln g_{\sigma^2}(\tilde{\mathbf{d}}|\mathbf{m}) = \frac{1}{2\sigma^2} \|\tilde{\mathbf{d}} - \tilde{\mathcal{Q}}\mathbf{m}\|^2 + \text{constants}, \quad (2.35)$$

where $g_{\sigma^2}(\cdot)$ is 0-mean Gaussian distribution with variance σ^2 . The probabilistic formulations allow us to determine the parameters c and σ^2 by maximum likelihood estimation (MLE). In equations 2.34 and 2.35 the constants are independent of \mathbf{m} . In equation 2.35, $\tilde{\mathbf{d}} \in \mathbb{C}^{n_{\text{tot}}n_{\omega}M_{\text{ga}}}$ and $\tilde{\mathcal{Q}} \in \mathbb{C}^{n_{\text{tot}}n_{\omega}M_{\text{ga}} \times M}$ are respectively formed by concatenating $\tilde{\mathbf{d}}^{(\gamma,j)}$ and $\tilde{\mathbf{L}}^{(\gamma,j)}$ along the column dimension in dictionary order of (γ, j) , where $\gamma = 1, \dots, M_{\text{ga}}$ is the supergather index, with M_{ga} being the number of supergatherers, and $j = 1, \dots, n_{\omega}$ is the frequency index. Here, the descriptor (γ, j) explicates the fact that $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{L}}$ as defined in equations 2.6 and 2.8, respectively, are specific to a particular supergather and frequency. Note that in the case of marine streamer acquisition, the first dimension of $\tilde{\mathbf{d}}$ and $\tilde{\mathbf{L}}$ is extended from n_h to n_{tot} . In contrast, in the standard approach of a single shot gather, the counterparts of $\tilde{\mathbf{d}}$ and $\tilde{\mathcal{Q}}$ would be of sizes $\mathbb{C}^{n_h n_{\omega} S_{\text{tot}}}$ and $\mathbb{C}^{n_h n_{\omega} S_{\text{tot}} \times M}$, respectively, where $S_{\text{tot}} = SM_{\text{ga}}$ is the total number of sources.

The objective function is then constructed as

$$J(\mathbf{m}) = \sigma^2(e(\mathbf{m}) + R(\mathbf{m})) = \frac{1}{2} \|\tilde{\mathbf{d}} - \tilde{\mathcal{Q}}\mathbf{m}\|^2 + \sigma^2 \sum_i \ln(c^2 + m_i^2), \quad (2.36)$$

where additive constants have been dropped. Its negative gradient is given as

$$\mathbf{g} \stackrel{\text{def}}{=} -\nabla_{\mathbf{m}} J(\mathbf{m}) = \widetilde{\mathfrak{L}}^+ (\widetilde{\mathfrak{d}} - \widetilde{\mathfrak{L}} \mathbf{m}) - 2\sigma^2 \sum_i Q(m_i) m_i, \quad (2.37)$$

where

$$Q(m_i) = \frac{1}{c^2 + m_i^2}. \quad (2.38)$$

Note that the shape of the objective function $J(\mathbf{m})$ typically changes over iteration step k because every iteration typically requires a new pass of frequency selection encodings for the M_{ga} supergathers to generate $\widetilde{\mathfrak{d}}$ and to effect $\widetilde{\mathfrak{L}}$. That the objective function depends on k is a topic that is studied in stochastic optimization (Spall, 2003). Our problem (albeit of much larger size) is similar to the ‘stochastic bowl’ studied by Schraudolph and Graepel (2002), because as shown in Appendix A the Hessian of the misfit function pertaining to frequency selection encoded supergathers consists of terms sampled from the standard full Hessian.

As frequency selection encoding could significantly alter the Hessian, the conjugacy condition of CG cannot be maintained if supergathers are formed with a new frequency selection encoding at each iteration, a strategy known as ‘dynamic encoding’. On one hand, in order to accelerate convergence, and on the other, in order to reduce I/O cost, I adopt a strategy of a hybrid CG (termed ‘CG within mimi-batch’ in Schraudolph and Graepel, 2002), whereby supergathers are encoded anew every K_{CGit} iterations. $K_{\text{CGit}} = 3$ is chosen in this study. Given fixed supergathers and a fixed $Q(m_i)$ defined in equation 2.38, K_{CGit} iterations are carried out by a CG scheme (outlined in Algorithm 1 in Appendix C). Then supergathers are randomly encoded again, $Q(m_i)$ ’s are updated, which is known as the ‘Iterative Reweighted Least-Squares’ method (Scales et al., 1988), the parameters c and σ^2 of the probability distributions are re-estimated through MLE, and the search direction of CG is reset to negative gradient.

2.3.2 Migration Method

The migration method is considered next. I choose prestack split-step migration based on the following two reasons: first, the fact that sources are subject to phase and/or frequency encoding demands that prestack migration is the method of choice. Second, aside from computational efficiency and the absence of operator aliasing, the fact that phase shift migration is a spectral technique makes it particularly convenient to perform frequency encoding. To handle smooth lateral variations in the velocity field, I opt for the split-step migration (Stoffa et al., 1990), as did Kuehl and Sacchi (1999). It is a straightforward procedure to adapt this to RTM, with the finite-difference method replacing the spectral method.

The use of LSM requires both the forward modeling and the migration operations. The use of prestack migration requires both a source field and downward continued data field. The details of this migration method are relegated to Appendix B, which is included because of the usefulness in assessing the computational complexities of the algorithms studied in this chapter.

To demonstrate the effectiveness of LSM, its performance will be compared to that of Iterative Stacking (IS). In contrast to the iterative refinement of LSM, IS of encoded migration images (Schuster et al., 2011) at the k^{th} iteration produces a sum of k realizations of migration images. For IS, dynamic encoding is used, so that at each iteration the input supergather, specifically the source wavefield at surface $P(x, z = 0, \omega)$ in Figure B.1(a), is formed using a new frequency assignment.

It is of interest to analyze how the proposed method would fare compared to the standard migration in terms of saving computational cost. This analysis is provided in Appendix C. In addition, the results in Appendix C allow us to compare the convergence performances of LSM and IS on the basis of the same computational cost.

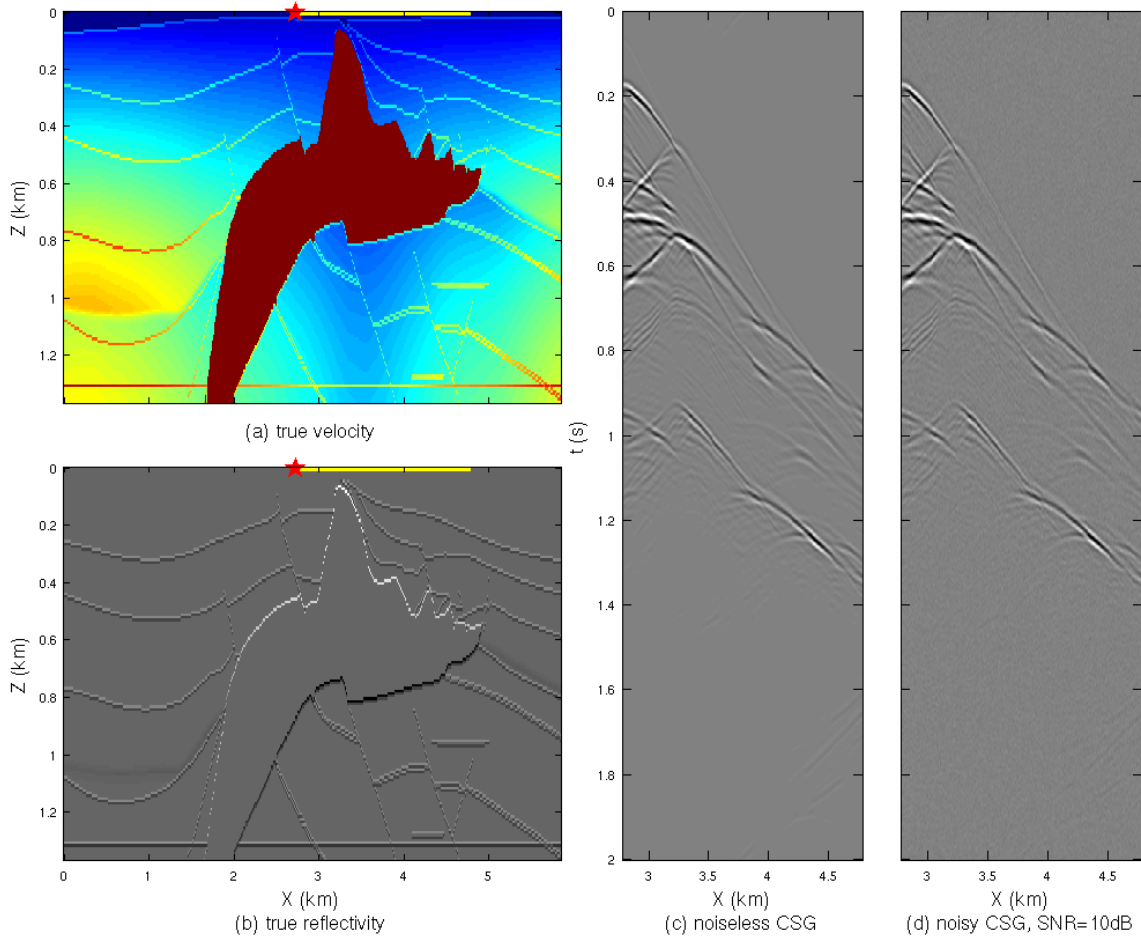


Figure 2.2: (a) The 2D SEG/EAGE salt model and (b) the associated reflectivity model, where the red star denotes a source at $X = 2.725$ (km) and the appending yellow line denotes the receiver aperture of this source. (c) The CSG from this source. (d) This CSG corrupted by bandlimited incoherent noise such that Signal-to-Noise Ratio (SNR)=10 dB.

2.4 Results

The proposed method of Least-Squares Migration with Multisource Frequency-Selection (LSMMFS) is tested on the 2D SEG/EAGE salt model, of size⁸ $n_x \times n_z = 640 \times 150$, with a grid spacing of 9.144 m. The velocity and the reflectivity model are shown in Figure 2.2(a) and (b), respectively.

The following parameters are chosen to emulate a marine acquisition geometry:

⁸ n_x is reduced from the original value of 645 to speed up the FFT.

shot interval = 18.288 m, receiver interval = 9.144 m, near offset = 45.72 m, line length = 2 km. The number, M_{ga} , of supergathers dividing up all $S_{tot} = 304$ sources varies from 1,2,4, up to 8. A Ricker wavelet with a 32 Hz peak frequency is used as the source wavelet, and 160 frequency channels equally divide the frequency range from 0 to 80 Hz, as exemplified alongside equation 2.27. With the true velocity and reflectivity models, a CSG for the source and receivers depicted in Figure 2.2(a, b) is generated for example using split-step forward modeling and is presented in Figure 2.2(c). To probe noise robustness, I contaminate the CSG's with various levels of random noise for a flat spectrum below 80 Hz, to yield SNR=10, 20, 30 dB. Figure 2.2(d) shows a contaminated version of (c). The noisy CSG's are first Wiener filtered, before being migrated. The smoothed velocity model shown in Figure 2.4(a) is used as the migration velocity, and is obtained by applying a 3×3 boxcar filter to the true velocity model shown in Figure 2.2(a).

As the LSM iterations proceed, the trial reflectivity model is updated and surpasses the standard migration image in quality, as demonstrated in Figures 2.3 and 2.4. For comparison, migration with the subsampled CSG's ('Subsmpl Mig') is also considered, which is an alternative means for data reduction and speedup. To yield a speedup of around 8 (see Figure 2.5), comparable to that of my proposed method, the subsampling ratio of 'Subsmpl Mig' is chosen as 1/8. As indicated by the black dashed horizontal lines in Figures 2.3(a–c), the model error of 'Subsmpl Mig' always exceeds that of standard migration, indicated by the black solid horizontal lines. As shown in Figures 2.4(g), the image produced by 'Subsmpl Mig' contains many artifacts that are disruptive because they are of similar spatial frequency and locations to those of reflectors.

Several features in Figure 2.3 are worth commenting. Larger M_{ga} and SNR lead to smaller model error and better convergence. Oscillations in the objective function in panel (d) are the expected behavior of hybrid CG. The objective

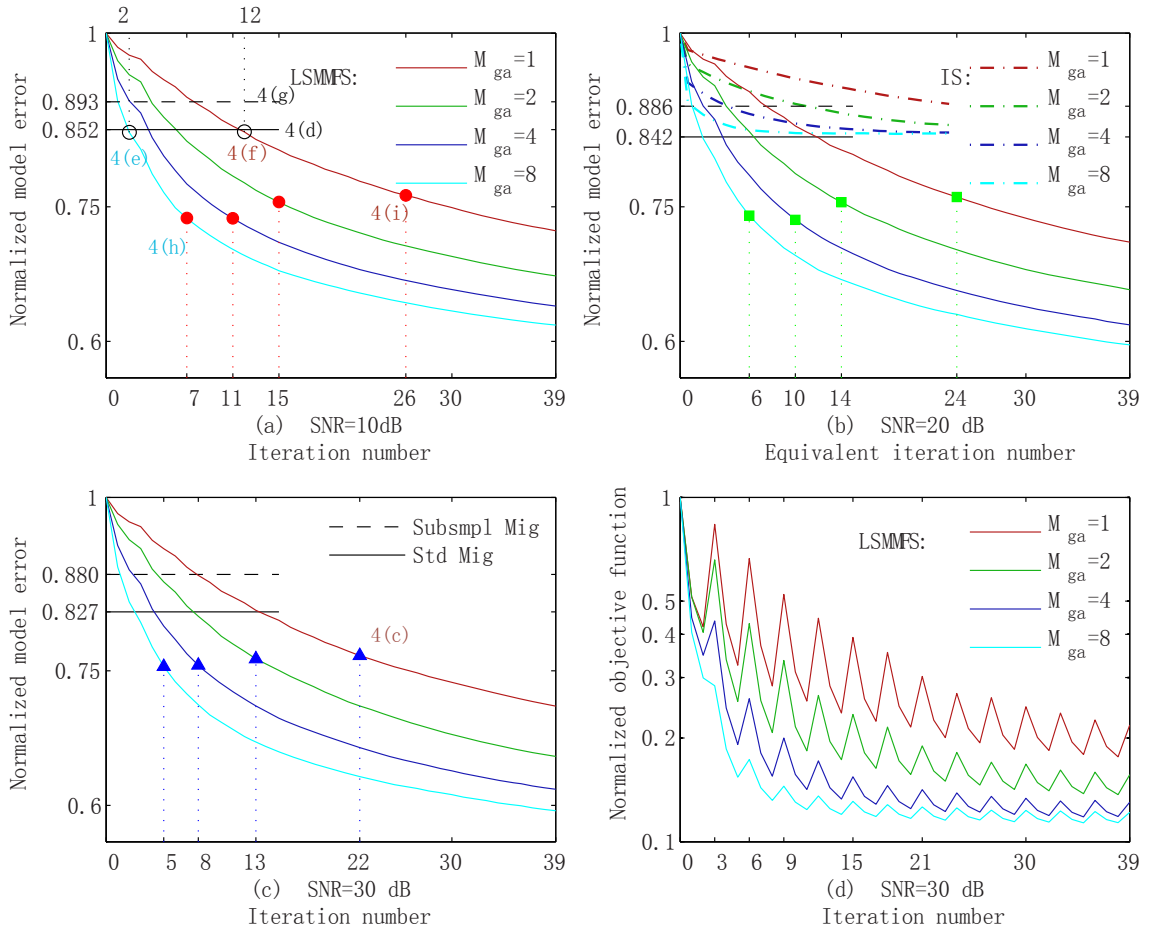


Figure 2.3: Normalized model error (a-c) and normalized objective function (d) for various SNR as a function of iteration number, in solid curves color coded for various M_{ga} , when minimizing by hybrid CG. For space efficiency, the legends in (a,c) are shared among (a-c). Regarding the black horizontal solid and dashed lines in (a,c), the dash-dot curves in (b), and the symbols \bullet , \blacksquare , \blacktriangle , and \circ in (a-c), see text for details. The alphanumeric short labels '4(c)' up to '4(i)' refer to in which figure and panels the corresponding migration images are shown.

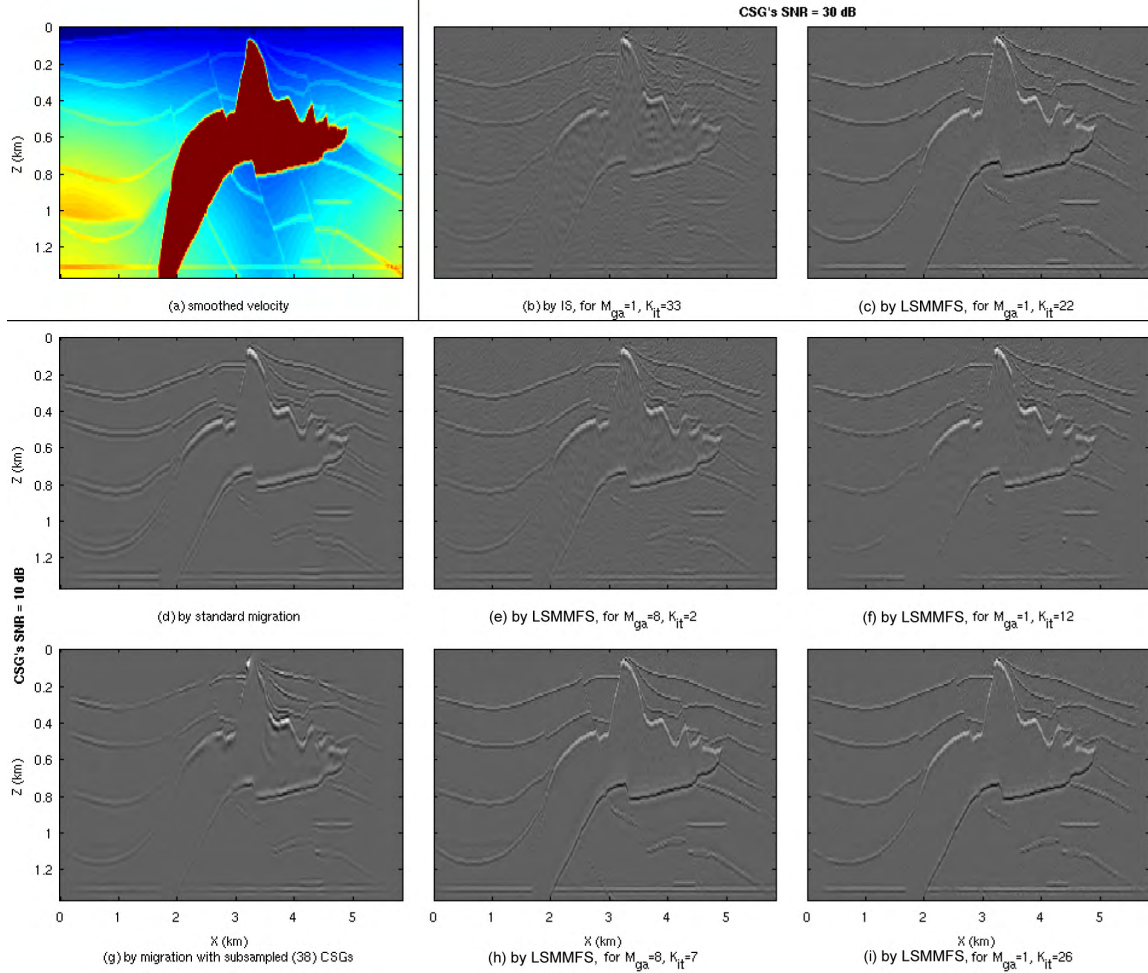


Figure 2.4: Reflectivity distributions obtained by various methods with a smoothed velocity model (a), in various parameter settings of M_{ga} and K_{it} , the iteration number, when applicable. 30 dB of the SNR of CSG applies to (b,c), whereas 10 dB applies to (d-i). (c-i) are respectively referred to in Figure 2.3(a) and (c).

function is consistently reduced by CG within every $K_{CGit} = 3$ updates, but will increase upon the presentation of newly encoded supergathers. This is because the previous optimization efforts are targeted at reducing a differently parameterized objective function. As the iterations proceed, however, the envelope of the oscillatory objective function still decreases, validating the robust performance of hybrid CG.

In terms of model error, the least-squares method can surpass the standard migration in as few as two iterations (see for example the \circ symbol at iteration 2 on the cyan curve in Figure 2.3(a)). This estimate, however, is too optimistic, even though I have made sure to minimize the model error of standard migration image as $\min_{\alpha} \|\alpha \check{\mathbf{m}} - \mathbf{m}\|^2$, where $\check{\mathbf{m}}$ is the migration image and \mathbf{m} is the true model. The reason is that a standard migration image tends to be smooth and the high frequency components are suppressed. Thus, the model error could be large. On the other hand, the image obtained by LSMMFS tends to be sharper, matching the true model better in terms of the L_2 -norm of the model error. The downside, however, is ringy noise, as evident in the corresponding reflectivity images shown in Figure 2.4(e) and (f). That is why it makes sense to involve human subjects in judging the quality of resulting images.

The break-even points where the image quality of LSMMFS is comparable to that of standard migration are indicated by the symbols \bullet , \blacksquare , and \blacktriangle in Figure 2.3(a-c). Three images corresponding to such points are shown in Figure 2.4(c), (h) and (i). To equate the quality of these images with that of the standard migration, shown in Figure 2.4(d), tradeoffs are made. In Figure 2.4(c), (h) and (i), there is some residual high-frequency noise, especially at shallow depths. But this noise is quite distinct from those of reflectors and thus it hardly affects the dominant features. On the other hand, the resolution of Figure 2.4(c), (h) and (i) is better than that of Figure 2.4(d). It is based on these two factors that I choose the break-even

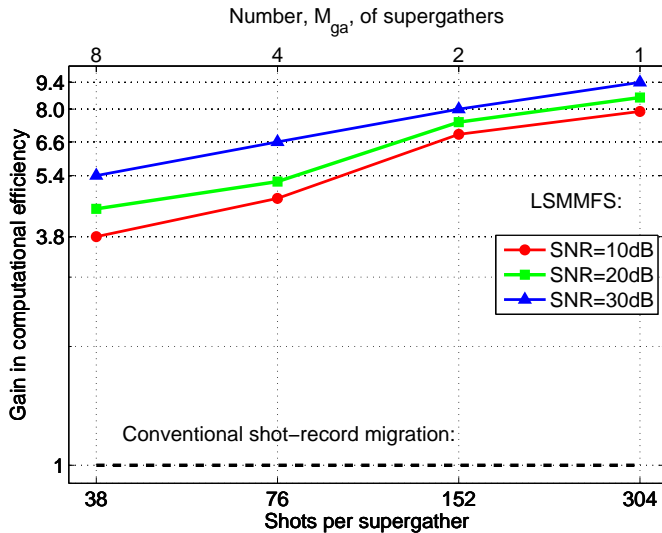


Figure 2.5: The gains in computational efficiency of the proposed LSMMFS as compared to the conventional shot-record split-step migration plotted on a log-log scale for various SNR's, as functions of shots per supergather, or equivalently as functions of M_{ga} (labeled atop).

points in visual quality. Once the abscissae, or K_{it} 's, of these break-even points are known (from equation C.3), I calculate the relative computational cost, or, its reciprocal, termed 'gain in computational efficiency', which is plotted in Figure 2.5. Here we see that, for the parameter settings and the model under study, nearly an order of magnitude of speedup can be achieved.

One may raise the concern that, due to the frequency selection scheme, even with a dozen iterations of dynamic encoding, each source can hardly have the chance to exhaust its spectrum. For example, take $M_{ga} = 2$, $K_{it} = 10$, then $S = S_{tot}/M_{ga} = 304/2 = 152$. So at any one iteration, each source only gets assigned 1/152 of the frequency channels available. With 10 iterations, in the best scenario a source can only cover a mere 10/152 of its spectrum. In light of this analysis, the apparent good performance of the frequency selection scheme seems therefore rather counter intuitive. To address this concern, I maintain that due to least-squares iterations, sources no longer act in straightforward linear superposition as they do in the standard migration. Rather, they act cooperatively and with collaboration between sources the model gets effectively illuminated by a wider range of spectrum than provided by stacking migrations.

To test this idea, I examine the convergence performance of IS, where frequency

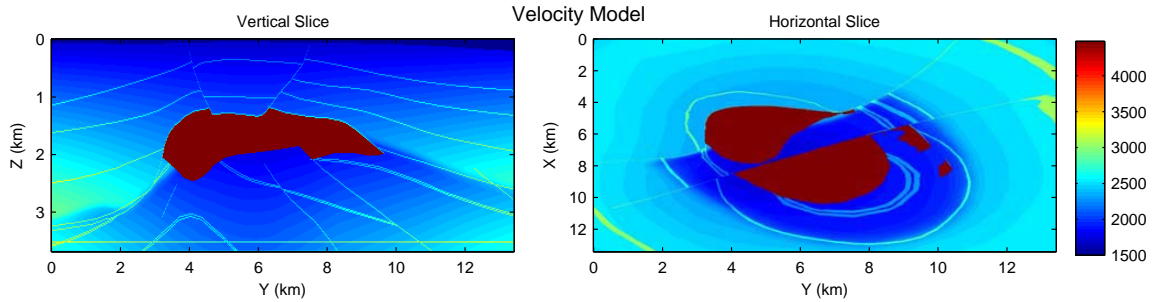


Figure 2.6: The 3D SEG/EAGE salt velocity model, in m/s, sliced at (left panel) $x=6.7$ km and (right panel) $z=1.98$ km.

selection encoding with multisource applies as well. Figure 2.3(b) includes the convergence curves (the dash-dot curves are for IS), plotted according to what is prescribed at the end of Appendix C, and Figure 2.4(b) shows a migration image of IS, which is obtained at the same computational cost as in Figure 2.4(c). Evidently, with this amount of computation, IS does not beat the standard migration in terms of either model error or the quality of migration image. The explanation for this phenomenon is precisely the concern raised earlier, aided by the realization that by random frequency assignments, rarely can a smooth spectrum result, and fluctuations in the spectrum are likely. Non-smoothness in the spectrum corresponds to ringiness in the time domain. Therefore the migration image is always inferior to standard migration image. Contrasting IS and LSMMS, one can see the essential role that least-squares updates play in this frequency selection multisource method. Additional insights are reaped from a comparison study conducted in Appendix D, where I show that iterative refinement likely leads to better solutions than migration does.

To test the viability of the frequency selection multisource method in processing 3D data, I use a 3D SEG/EAGE salt model, of size $n_x \times n_y \times n_z = 672 \times 672 \times 185$ with a grid interval of 20 m. Slices of the velocity model are depicted in Figure 2.6. There is one receiver at each grid point, and $S_{tot} = 64 \times 64 = 4096$ sources are equally distributed on the surface. A Ricker wavelet with a 16 Hz peak frequency

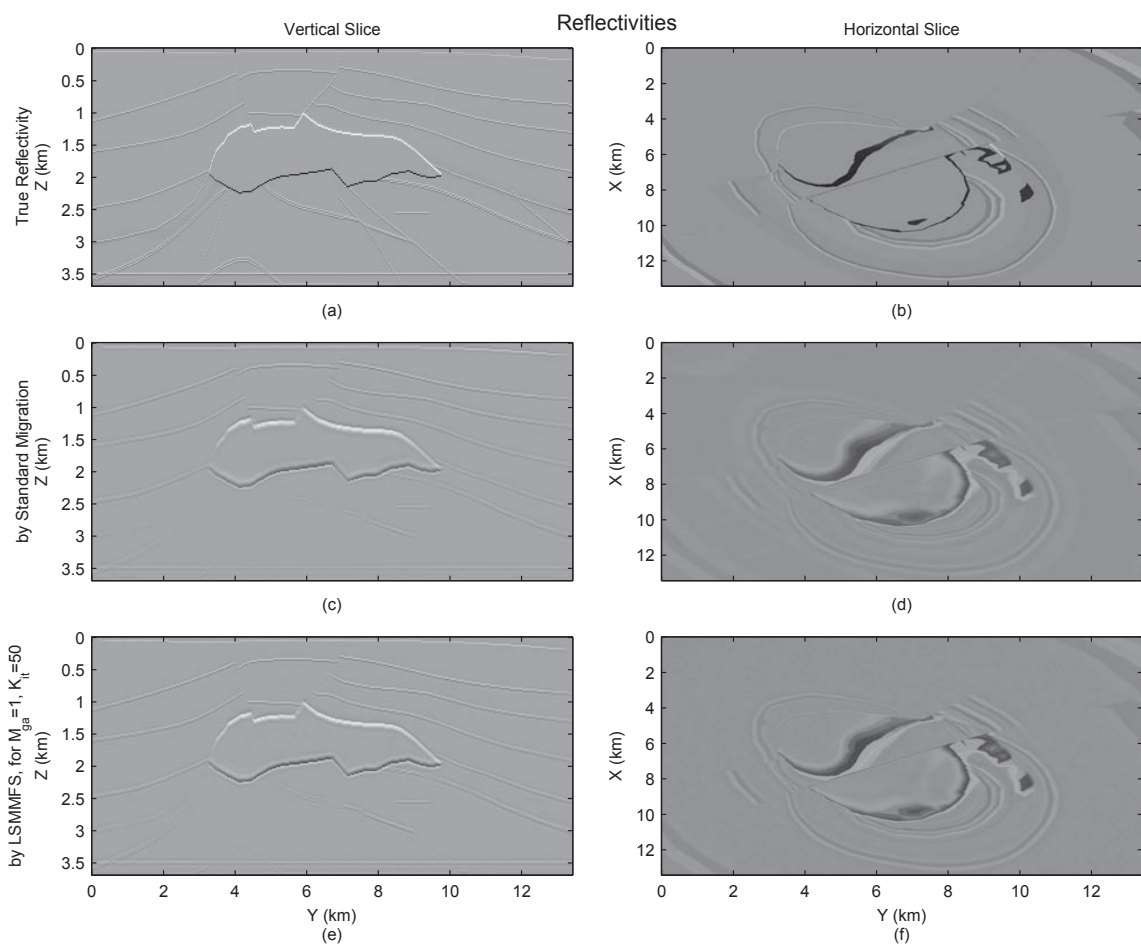


Figure 2.7: Reflectivity model (a,b) and others obtained by (c,d) standard shot-record prestack split-step migration, and (e,f) the proposed LSMMFs for 1 supergather at the 50th iteration of steepest descent.

is used as the source wavelet, and $n_\omega = 360$ frequency channels equally divide the frequency range from 0 to 40 Hz, as exemplified alongside equation 2.27. Here, a fixed acquisition geometry of both the sources and receivers is assumed, as the aim of this study is to test whether the frequency selection multisource method can work on either land or marine 3D data.

Note that in this case the number of sources S_{tot} is far greater than the number of available frequency channels n_ω . If $S = S_{\text{tot}}/M_{\text{ga}} > n_\omega$, then assignment of non-overlapping source spectra is not possible, unless only a small number of sources are turned on at a time, a practice that would discard much useful information. Here I allow overlapping source spectra. If $S \gg n_\omega$, each frequency channel is shared among S/n_ω sources. This assignment can be implemented for example by randomly drawing S/n_ω source indices in turn without replacement to be assigned to each frequency. In addition, a random polarity ± 1 is assigned to each source, in order to reduce the crosstalk among sources sharing a frequency. A comparison of this method with standard migration is given in Figure 2.7, where 50 steepest descent updates of LSMMFS in one supergather yield a result comparable to standard migration. Equation C.5 says, based on the computational cost, that the speedup is $2S/(4K_{it} - 1) = 2 * 4096/199 \cong 41$. On the I/O side, as analyzed at the end of the Theory section, the proposed method requires either $\epsilon \times$ or $(2 + \epsilon) \times$ the I/O cost of the standard approach, depending on whether or not the data have been transformed into the frequency domain.

2.5 Conclusions

I emphasize that the mismatch between the limited number of live hydrophones in a marine-streamer survey and the pervasive number of live traces generated by modeling is essentially a form of crosstalk in multisource migration/inversion

not seen in a fixed spread survey. To completely remove this marine crosstalk I propose a frequency selection encoding scheme similar to the ones used in the communications industry. This scheme demands that any crosstalk-prone sources should only emit signals in non-overlapping frequency bands. This scheme allows any receiver to selectively tune in to the valid source (the source that indeed has made a contribution to this receiver) and to selectively disregard potentially confounding sources. Such sources are grouped with the receiver only at the time of multisource modeling.

I show that my frequency selection technique in a multisource framework is similar in form to the well-studied stochastic optimization problem. In terms of computational cost, blending groups of S sources together to form supergatherers would reduce the subsequent computational cost by a factor of S . Due to the weakened illumination capability, iterations are usually required to produce an image comparable to standard migration. Fast convergence with many fewer steps than S yields an overall speed gain compared to conventional migration.

Numerical results with a 2D salt model and a marine survey show that crosstalk is completely removed with a multisource speedup nearly an order of magnitude faster than standard migration. In the 3D example with a fixed acquisition geometry (a fixed OBS geometry) a speed up of 40 was achieved compared to standard migration. In addition, better resolution was achieved.

Chapter 3

Full Waveform Inversion with Multisource Frequency Selection on Marine Streamer Data

3.1 Introduction

Multisource migration (Morton and Ober, 1998; Romero et al., 2000), LSM, and waveform inversion (Krebs et al., 2009; Virieux and Operto, 2009; Dai and Schuster, 2009; Tang, 2009) of phase-encoded supergathers were developed to significantly reduce the cost of migration and inversion. The key idea is to blend N encoded shot gathers into an N -shot supergather, and iteratively migrate encoded supergathers or, in the case of LSM or FWI, encoded supergather residuals. A representative formula for iteratively estimating the model parameter s_i in the i th cell is given by the steepest descent formula

$$s_i^{k+1} = s_i^k - \alpha \frac{\partial \epsilon}{\partial s_i}, \quad (3.1)$$

where s_i can represent either the reflectivity or the slowness in the i th cell, α is the step length, and ϵ is the misfit function⁹ that is encoded after each iteration with a different encoding function. The benefit of this approach is that wave equation migration of each supergather costs about the same as the migration of a standard shot gather. If the number of iterations is fewer than N , then the computational cost of phase-encoded multisource imaging can be much less than separately migrating each of the N shot gathers (Schuster et al., 2011).

The problem with the above approach is that it is efficiently suited for land data where the receiver spread is fixed for each shot, but not for marine data with a receiver array that moves with each shot. As an illustration, Figure 3.1(a1) shows two shot gathers to be blended, where one shot is at the red source and the other is at the dark blue source; this 2-shot supergather will be denoted as $\mathbf{d}^{obs.}$. Typical of marine surveys, the receiver array is at a different offset for either source so that only certain receivers are *selectively* listening for the red shot but not for the dark blue shot at the uncommon receiver positions. In comparison, the predicted 2-shot supergather $\mathbf{d}^{pred.}$ generated by a finite-difference¹⁰ solution of the wave equation does not *discriminate* and generates traces at every receiver, as shown in Figure 3.1(a2). Hence, there will be discrepancies between the predicted and observed traces at the uncommon receiver positions (indicated by the dashed ovals in Figure 3.1). I denote this problem in multisource FWI as the *aperture mismatch* problem, where the observed supergather is for a blended marine survey while the predicted supergather is for a blended land survey.

The aperture mismatch will lead to a non-zero misfit function $\epsilon = \frac{1}{2} \|\mathbf{d}^{pred.} - \mathbf{d}^{obs.}\|^2$ even if the exact velocity model is used for prediction. The remedy to this mismatch

⁹The misfit function $\epsilon = \frac{1}{2} \|\mathbf{d}^{obs.} - \mathbf{d}^{ref.}\|^2$ relates to the L_2 norm of the encoded difference between the predicted and observed supergathers.

¹⁰A finite-difference simulation of two *simultaneous* sources (a red source and a dark blue source) will compute traces everywhere on the surface that are a superposition of the wavefields from both sources.

is to use an encoding function in the multisource finite-difference modeling that only activates specified receivers for any one shot. This orthogonal encoding function strategy was developed by Huang and Schuster (2012) for wave equation migration, and will now be tested for FWI.

The first part of the chapter provides the theory for multisource FWI with frequency selection, and is followed by results from tests on synthetic and field data. Speedups ranging from 4× to 8× compared to conventional FWI are obtained. The last part presents a summary.

3.2 Theory

The formula for multisource FWI is given in equation 3.1, where the velocity model is updated at each iteration and s_i represents the slowness model in each cell. The frequency selection encoding scheme is described in Huang and Schuster (2012), and summarized by the following steps.

1. Figure 3.1(a) illustrates the problem, and the first step is to assign a non-overlapping frequency spectrum to each of the sources. In Figure 3.1(b), the cyan (red) source is bandlimited around 20 Hz (10 Hz) during the multisource simulation and the supergather is computed.
2. In Figure 3.1(c), a bandpass filter is applied to decode the (b) supergather so that the cyan traces can be separated from the red traces. In each shot gather, only traces recorded by receivers in a hypothetical marine survey are retained while the others are muted.
3. In Figure 3.1(d), the decoded and muted traces are blended together to give the bandlimited marine supergather. This procedure is iterated in equation 3.1, except that a unique non-overlapping frequency is iteratively assigned to

each source. For a sufficient number of iterations, the full bandwidth of the data is employed at each source.

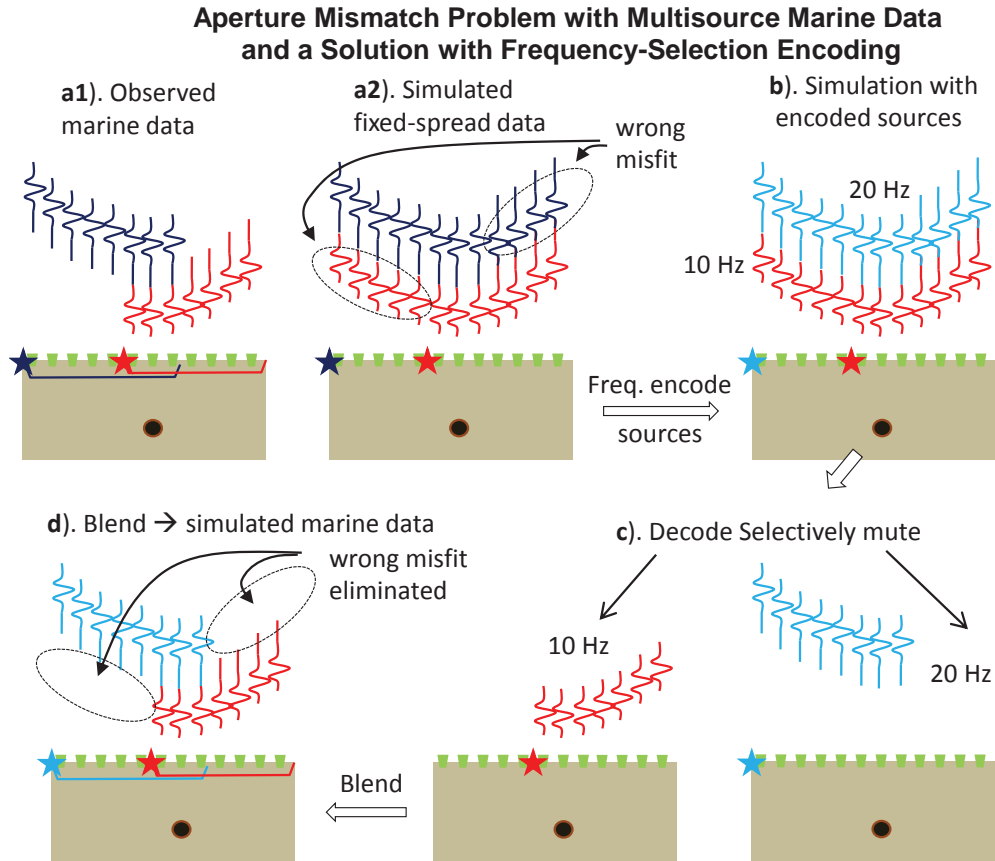


Figure 3.1: The solution to the acquisition misfit problem illustrated in panels (a1) and (a2), is given by the following steps: (b). the cyan (red) source is bandlimited around 20 Hz (10 Hz) during the multisource simulation; (c). a bandpass filter is applied to decode these traces so that the cyan traces can be separated from the red ones. Muting of the specified traces for a marine geometry is applied. (d). The decoded and muted traces are blended together to give the bandlimited marine supergather. Now, there is no aperture mismatch between the simulated and observed supergather, except that the frequency channels of the former are a subset of those of the latter.

To apply the frequency-selection to a FDTD simulation, what are injected at sources (and at receivers when back-propagating the data misfit) are causal sine waves. This causality shows up as transient in received sinusoidal seismograms, causing spectral leakage. To mitigate this problem, assuming the impulse response

of the Earth is active only within time nt , I run the FDTD simulations over $2nt$, and keep the responses only within the second nt , as illustrated in Figure 3.2¹¹. The responses in the second nt are nearly identical to what is obtained by running the simulation over nt with an input of infinitely long sine waves of period nt . Such sine waves are the bases of the Discrete-time Fourier transform (DTFT) of nt , if performed on the data obtained from the FDTD simulation over nt using time-limited source wavelet.

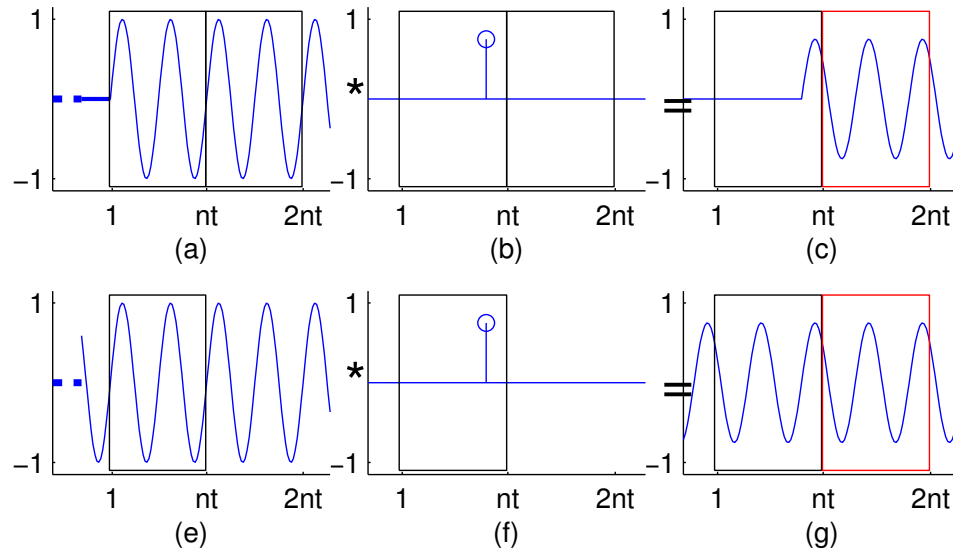


Figure 3.2: Doubling the simulation time of FDTD to mitigate the transients when causal sine waves are injected at the inputs, e.g., sources or receivers. If a (a) causal sine wave $u(t) \sin(2q\pi t/nt)$, where $u(t)$ represents a unit-step function, is convolved with (b) a delayed impulse, the output is a delayed version of the input signal, plotted in (c), where the black box contains the transient, while the red box contains the steady-state. If $\sin(2q\pi t/nt)$, plotted in (e), is injected instead, then the output, plotted in (g), would be a scaled and delayed sine wave. Note that the red box in (g) is identical to that in (c). Due to periodicity in nt of the input sine wave, in (g) the black box is identical to the red.

This transient-reduction scheme also applies when back-propagating the residual wavefield in forming the gradient of FWI, as illustrated in Figure 3.3. Note that only the times slices from $nt + 1$ to $2nt$ of the source wavefield need to be

¹¹Only a single $\delta(t)$ -impulse of the Earth response is shown. Linear superposition generalizes this to an arbitrary impulse response within nt .

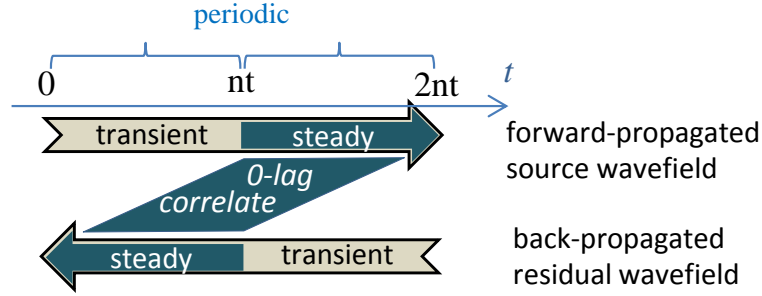


Figure 3.3: Transient-reduction in forming the gradient of FWI. The residual wavefield is back-propagated over $2nt$ steps, starting from step $2nt$. The wavefields for time steps from nt to 1 are steady. The steady wavefields are periodic in time with period nt , inherited from the injected sine waves. Therefore, to form the gradient of FWI, we take the 0-lag correlation between the back-propagated residual wavefield (for time slices from 1 to nt) and forward-propagated source wavefield (for time slices from $nt + 1$ to $2nt$).

reconstructed from the FDTD boundary. The total number of time steps of FDTD propagation thus equals $2nt + nt + 2nt = 5nt$, where the $2nt$ denotes the total time for the forward-propagation and back-propagation of the source and residual wavefields, respectively, while the nt is for reconstruction of the last part of the source wavefield. This compares to $3nt$ time steps in the standard approach in computing the gradient.

In numerical optimization, after the gradient is computed, Brent's method (Press et al., 2007) is used for the line search. This takes on average 5 function evaluations, namely, 5 passes of forward-propagation of the source wavefield. Therefore in the standard approach, per iteration of FWI $8nt$ time steps of FDTD propagation are required, whereas in my transient-reduction scheme, $5nt + 2 \times 5nt = 15nt$ time steps of FDTD propagation are necessary. Roughly, per iteration the latter costs about twice as much as the former does.

3.3 Numerical Results

3.3.1 Synthetic Data

The Full Waveform Inversion with Multisource Frequency-Selection (FWIMFS) method is tested on synthetic data computed for the SEG/EAGE salt model with a marine geometry. The model is decimated by a factor of 3×3 for less overall computational time, as shown in Figure 3.4(b). The source wavelet is a Ricker wavelet peaking at 8 Hz. There are 60 shot gathers evenly distributed across the top of the model with the shot spacing of 82.3 m, the receiver spacing is 27.4 m, and the line length is 2.3 km.

The FWI method uses a preconditioned conjugate gradient method, where the acoustic forward and backward solvers are a finite-difference solution to the 2D space-time wave equation of constant density. The FDTD algorithm is second-order accurate in time and fourth-order accurate in space, denoted as $O(2, 8)$. The source wavelet for the proposed frequency selection method is a pure cosine wave, also employed in Nihei and Li (2007) and Sirgue et al. (2008), at a selected frequency.

The starting model is shown in Figure 3.4(a) and the standard FWI tomogram after 69 iterations is shown in Figure 3.4(d). This result and the associated CPU time will serve as the standard metrics by which the FWIMFS algorithm will be measured.

The FWIMFS strategy produces the tomogram shown in Figure 3.4(c). This result required 439 iterations to achieve the same accuracy as the Figure 3.4(d) result in 69 iterations. This amounts to a factor of $439/69 = 6.362$. The convergence curves shown in Figure 3.5 are plotted in the way that the x-axes for the red dashed curves have been deliberately shrunk by 6.362. With this adjustment, the two convergence curves of velocity error in Figure 3.5(b) almost coincide with each other. Taking this factor into account, among others such as the overhead

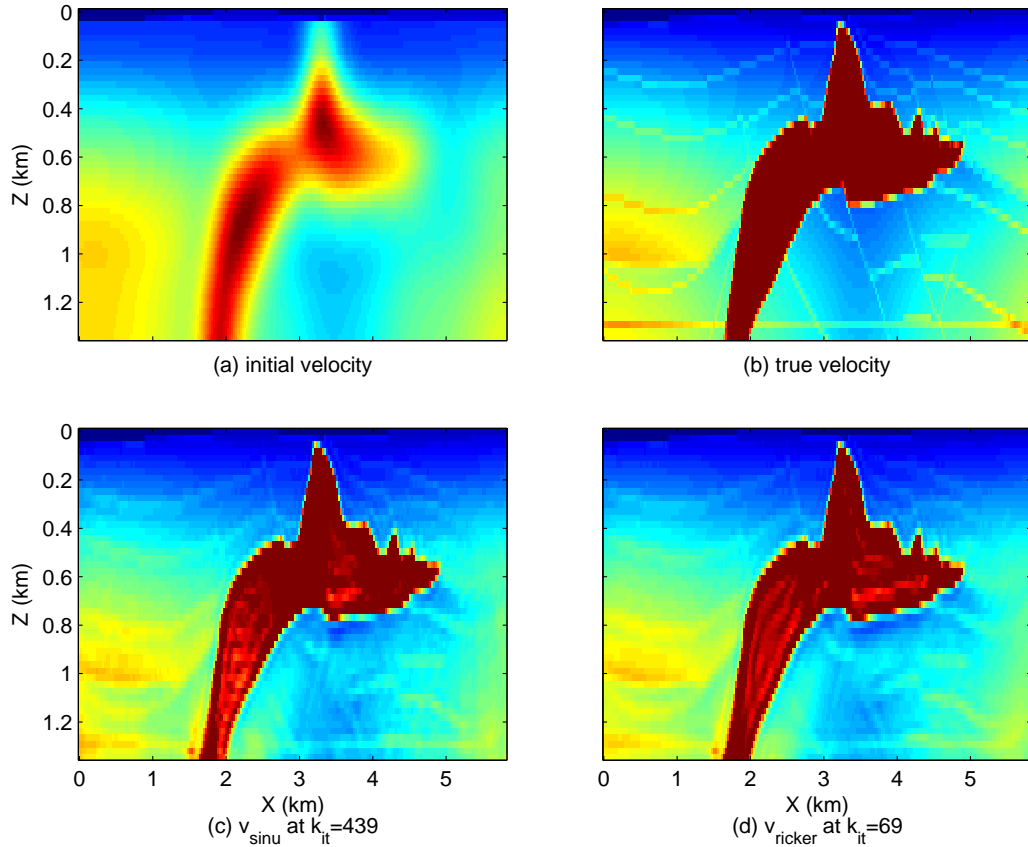


Figure 3.4: (a) The initial velocity, (b) the true velocity, (c) the result of the proposed FWI at the 439th iteration, and (d) the result of standard FWI at the 69th iteration.

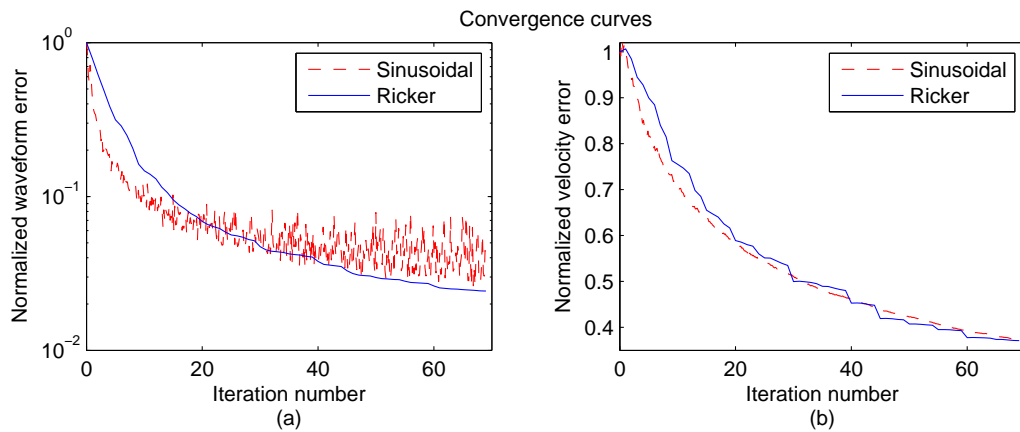


Figure 3.5: Normalized error of waveform (a) and velocity (b) as functions of iteration number, for the proposed method (denoted by ‘Sinusoidal’) and the standard FWI (denoted by ‘Ricker’). Note that the iteration numbers for the red dashed curves are actually 6.362 times those in display. See text for details.

of increased runtime per finite-difference run, the speedup over the conventional FWI scheme is estimated to be about 4. This compares to the $8\times$ speedup reported by Huang and Schuster (2012) for RTM. One reason for this discrepancy is that the implementation in the time domain suffers from a $2\times$ overhead in order to reduce the effect of transient noise in the simulated sinusoidal seismograms.

3.3.2 GOM Streamer Dataset

The FWIMFS is then tested on a GOM streamer dataset. I include 496 shots with a shot interval of 37.5 m. The source-receiver offset ranges from 198 m to 6 km, with a receiver spacing of 12.5 m. Trace length of 7 s is included, with a sampling interval of 2 ms.

This test consists of the following steps:

1. Estimate the source wavelet, by integrating the direct wave.
2. Convert the 3D physical data to 2D, on which grid the simulations are run.
3. Obtain an initial velocity model, shown in Figure 3.9(a), by combining travel-time tomography and the velocity from semblance analysis, referred in short in this chapter simply as ‘traveltime tomography’.
4. Implement a multiscale strategy (Bunks et al., 1995) with FWIMFS.
5. To validate the accuracy of the tomogram, compute the RTM image and CIGs using the tomogram velocity.

The source wavelet is estimated as follows. 1) Stack the direct waves from Common Offset Gatherings (COGs); 2) integrate the stacked trace along time; and 3) set its DC component to 0. The reason for integration is because the received direct wave is $\propto d/dt$ of the source wavelet, as illustrated in Figure 3.6. The estimated source wavelet is plotted in Figure 3.7(a); a bandpass filtered version (filtered

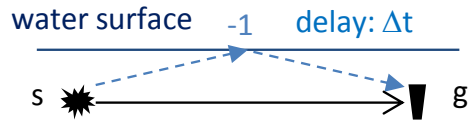


Figure 3.6: The direct wave and free-surface reflection combine together to give a detected signal approximated by $w(t) - w(t + \Delta t)$, proportional to the time derivative of the wavelet $w(t)$.

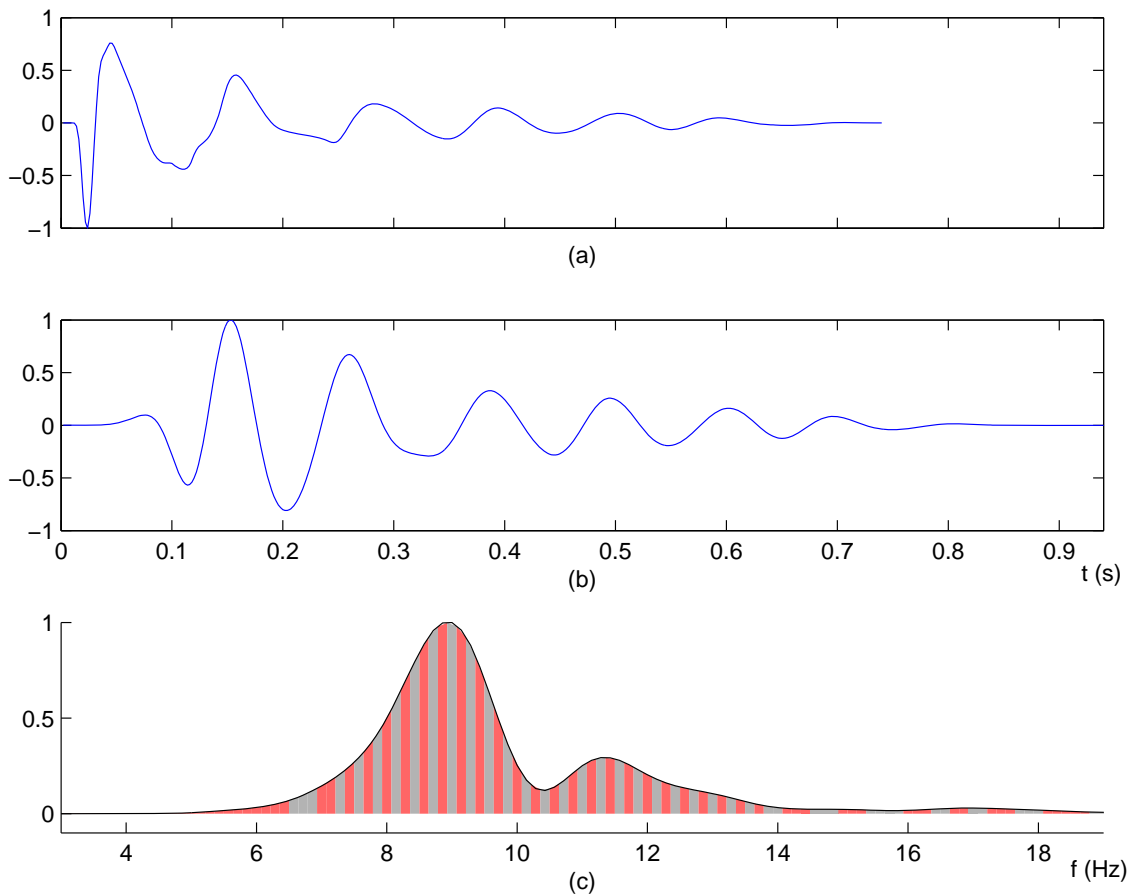


Figure 3.7: (a) The estimated source wavelet, (b) bandpass filtered wavelet with peak frequency 10 Hz. (c) The power spectrum of (b), which is divided into intervals in order to make the energy of each interval comparable. (Due to discreteness of frequencies, the interval energy cannot be made exactly equal.)

with a Ricker wavelet of peak frequency 10 Hz) is plotted in Figure 3.7(b). There are 210 frequency components supporting the power spectrum curve plotted in Figure 3.7(c). As the power is very small at the tails, rather than assigning each frequency component individually, I group them into 62 frequency intervals, filled

in alternating colors in Figure 3.7(c). The intervals are wider near the tails, so that the area under the curve in each interval is equalized as best as one can. Each interval of frequencies is selected as one entity in my frequency-selection scheme. The 3D to 2D conversion is achieved by multiplying the data spectra by $\sqrt{i/\omega}$ (Barton, 1989) and then gaining in the time domain by \sqrt{t} to convert the geometric spreading from 3D to 2D.

Aspects of frequency-selection and stochastic optimization

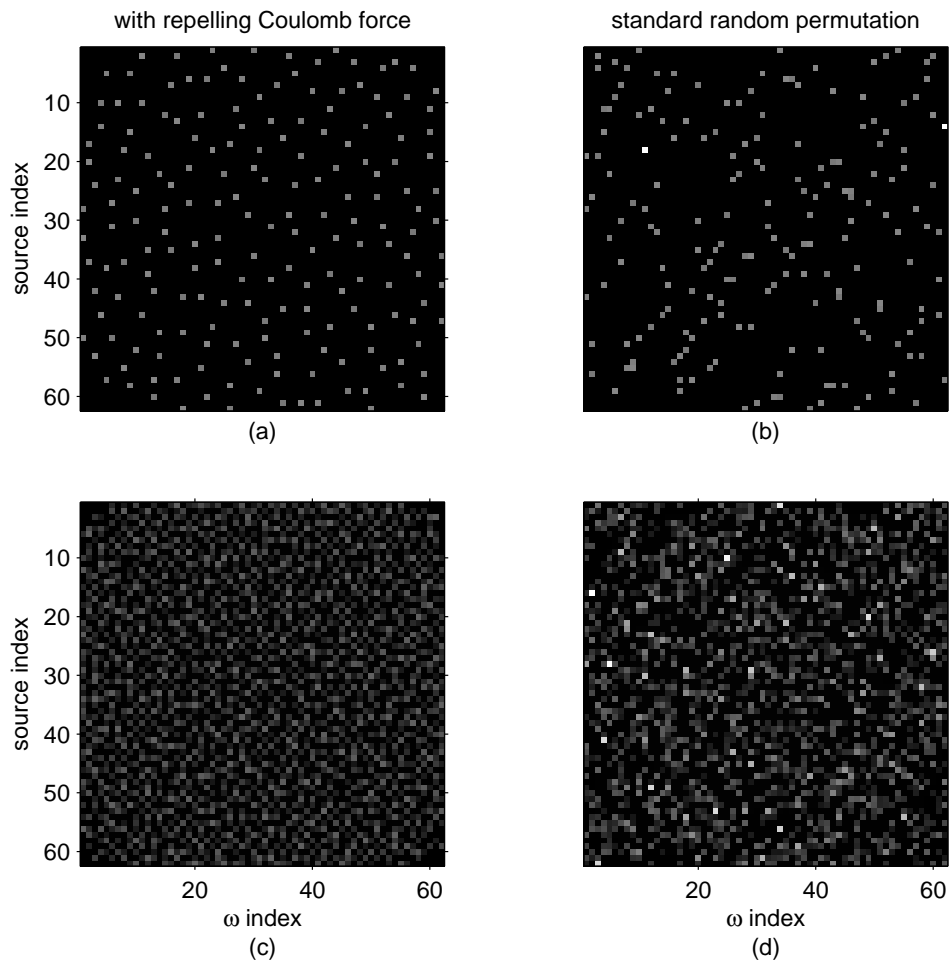


Figure 3.8: Frequency-selection encoding results. Each grey spot denotes a mapping between its two coordinates ω_j and s_i . Upper panels show the mappings accrued over 3 iterations, while lower panels show those over 31 iterations. A bright spot means there are mapping points colliding in history. Earlier mappings are of decayed charges, explaining the varying grey scales of the grey points.

In step 4, for the Multisource Frequency-Selection (MFS), I propose a Quasi-Monte Carlo frequency–source encoding strategy, motivated as follows. It is desirable that every source has a chance of evenly sampling the frequency components assigned to it over the many iterations of FWIMFS. An example of 10 frequency components and 4 iteration steps for a particular source might select the indices like 9,2,5,7, where each number denotes a frequency index; 4 iterations give 4 numbers. An undesirable example is 3,2,2,1, which over-represents the low frequency index 2, omitting the medium and high-frequency components. Moreover, nearby sources tend to illuminate an overlapping region of subsurfaces. Therefore if their frequency contents differ, then they as a whole would cover a wider range of frequency components. Back to the preferable example, its neighbor emitting sine waves of frequency with indices like 2,9,5,7 would be less desirable than emitting a wider band of frequencies with indices 1,10,3,8.

To achieve this end, I introduce repellent Coulomb forces between 2D point charges, each point (ω_j, s_i) denoting a mapping from ω_j to source s_i . This electrostatic system then settles, simulated through greedy optimization, into a low-energy configuration, in which all charged points spread out as much as possible. Examples of this encoding strategy are shown in the left column of Figure 3.8, which appear more uniformly distributed than the counterparts of the standard random permutation shown in the right column.

As the specific frequency-selection code changes over iterations of FWI, this falls in the realm of stochastic optimization (Spall, 2003). While the convergence of a line search in stochastic optimization is still a research problem, I adopt a hybrid approach. Run a gradient descent method with line search for the first K_0 iteration steps of FWI, then switch to a stochastic gradient descent method, where the step size $\propto 1/k$, k being the iteration step index. This is a lightweight and robust algorithm that converges almost surely to a local minimum (Kiwiel, 2001). The

problem remains in choosing the appropriate constant coefficient for this step size formula. The recipe is, first, identify the smallest η , η_{min} , resulting from the first K_0 steps of line search, expressed in the form

$$\|\Delta\mathbf{x}\| = \eta\|\nabla J(\mathbf{x})\|. \quad (3.2)$$

Here, \mathbf{x} is the unknown parameter vector, such as the velocity model; $\Delta\mathbf{x}$ is the update of \mathbf{x} at a step; and ∇J represents the gradient of the objective function. Then, the constant coefficient of the step size formula can be fixed accordingly, as

$$\Delta\mathbf{x}^k = -\eta_{min} \frac{K_0}{k} \nabla J(\mathbf{x}^k), \quad \text{for } k = K_0 + 1, \dots \quad (3.3)$$

A similar recipe of determining the step size of stochastic gradient descent is suggested in Bottou and Bousquet (2011).

As mentioned at the end of Section 3.2, 5 function evaluations are required on average by Brent's method for line search. In contrast, none is required in a stochastic gradient descent method. Consequently, per iteration step of the latter method, the computational cost is only due to the gradient computation, which requires, with the transient-reduction scheme, $5nt$ FDTD propagation steps. This compares to $8nt$ in the standard approach, which, taking advantage of the CG method, needs an accurate line search, and therefore explains the associated overhead. This lightweight stochastic gradient descent approach translates to more iterations and therefore more frequency-selection codes in use.

To further reduce the amount of stochasticity in the gradient, we empirically adopt averaging of two successive gradient calculations. Namely, perform the multisource encoding and gradient computation twice, then stack the gradients. This will double the computational cost. So a tradeoff exists between the size of the number of gradients in the average and the convergence speed of the stochastic

gradient descent. This averaging (also known as the mini-batch) scheme applies to all 8 supergathers, which are formed by dividing up the 496 shot gathers.

I start the inversion with the data bandpass filtered from 0–6 Hz. The initial velocity model is decimated to a grid size of $n_z \times n_x = 51 \times 376$. At later iterations the band is widened to 15 Hz, and the model is upsampled (with interpolation) to $n_z \times n_x = 101 \times 752$. Finally the frequency band covers 0 to 25 Hz, and the velocity model is of grid size $n_z \times n_x = 201 \times 1504$. I only use standard FWI for the first two cases, because the amount of computation is negligible compared to the third case. For example, the second case has a quarter of the model size, half of the time samples (because of doubled time sampling interval), and half the number of shots (due to downsampling). Therefore the computational load per iteration of the second case is only 1/16 of the third.

Results of the GOM dataset

As shown in Figures 3.9(b) and (c), the tomograms computed by standard FWI after 20 iterations and by FWIMFS after 71 iterations appear very similar. The associated computational cost of the latter case is 1/8 of that of the former. In both cases, there appear to be some regions of low-velocity anomalies. One is a horizontal layer at a depth of 0.75 km, where the velocity is slower than in water. This could be due to out-of-plane wavepaths modeled in 2D, resulting in underestimating the velocity, or it indicates unconsolidated sediments. Other low-velocity anomalies seem to lie along the wavepaths of diving waves. This arises due to cycle-skipping of some of the diving waves. Such velocities can be elevated by manual intervention, a trial move to be accepted if it overcomes some cycle-skipping. Rather than going into details of improving the FWI results, this chapter aims at demonstrating that FWIMFS produces essentially the same results as standard FWI does, with the benefit of a reduction in the computational cost. How we massage the velocity

model to address the cycle-skipping is a separate research project.

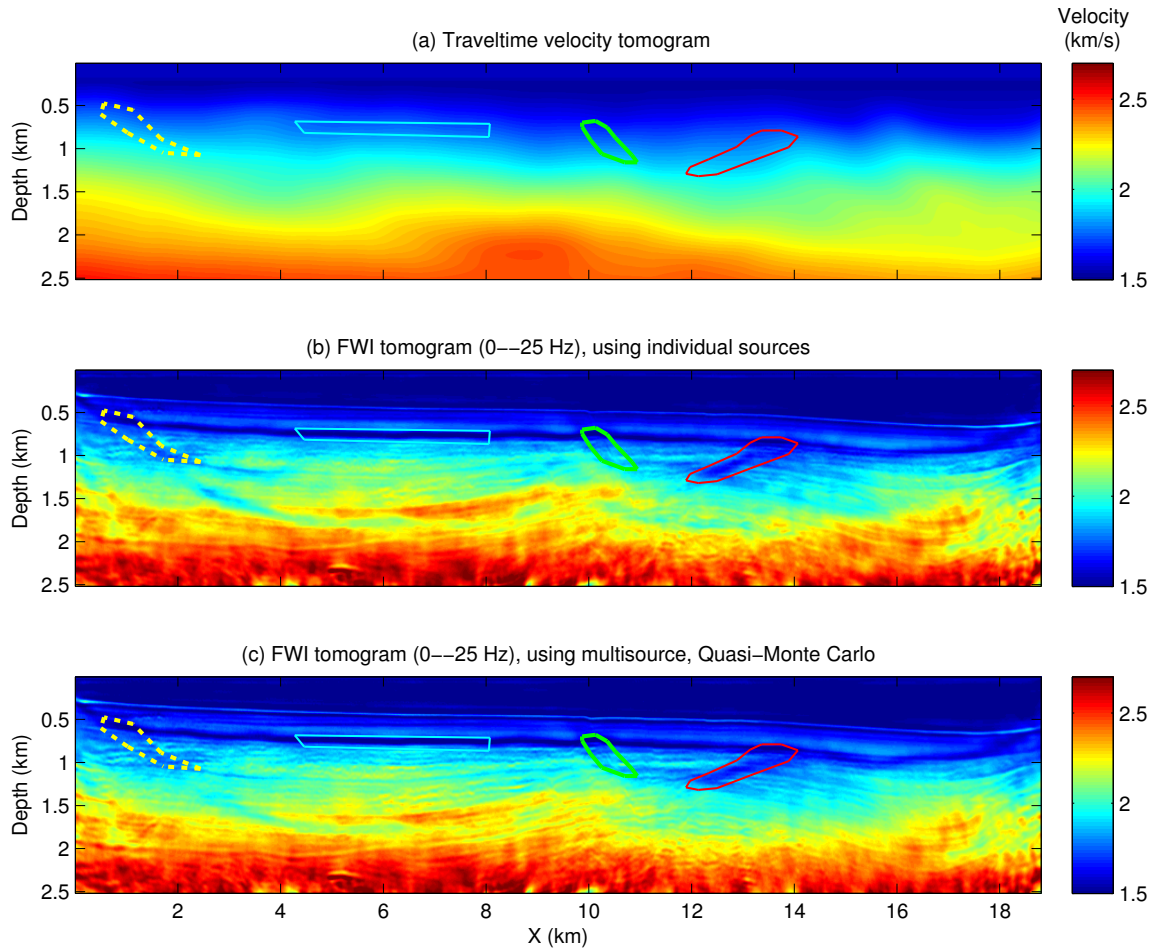


Figure 3.9: FWI results for the GOM dataset. (a) The initial velocity model. The velocity tomograms obtained from (b) standard FWI after 20 iteration steps, and (c) FWIMFS after 71 iteration steps. The encircled regions are low velocity anomalies in the FWI results.

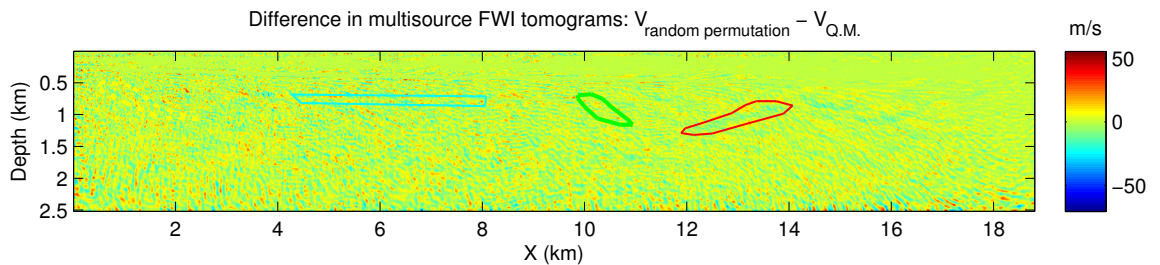


Figure 3.10: Difference of FWI velocity models obtained using Quasi-Monte Carlo versus standard random permutation in frequency-selection encoding. The encirclements are for reference only.

In order to see how my quasi-Monte Carlo encoding scheme accelerates the inversion, Figure 3.10 shows the velocity differences between the tomograms obtained using the two alternative encoding schemes: standard random permutation versus quasi-Monte Carlo. Surprisingly, the two velocity models are almost identical. The other results, such as RTM images and CIGs, appear indistinguishable, and therefore further comparisons between the two are omitted. This implies that the FWIMFS is very robust with regard to specific random codes of frequency-selection. Figure 3.11 plots the convergence curves of the various approaches. The reason why the red curve lies slightly above the green one is explained as follows. By quasi-Monte Carlo encoding, the new multisource data presented to the inversion algorithm is very different from what it has seen in the past, because this data contains entirely new frequency components. In contrast, by standard random permutation, some frequency components may have been used earlier, and thus the new multisource data is not entirely new. So the algorithm using standard random permutation can fare better at fitting the data.

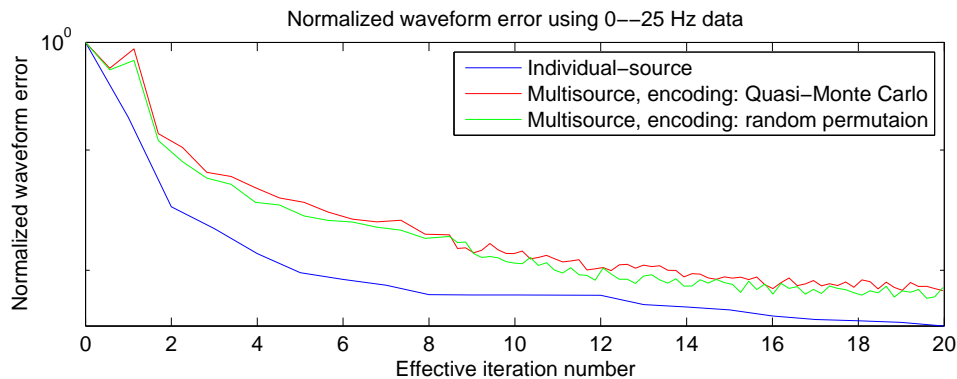


Figure 3.11: Reduction of data misfit (for 0–25 Hz) over FWI iterations, starting from the model obtained from FWI of 0–15 Hz data. The iteration numbers are labeled according to those of the standard FWI (i.e., of individual-source). The curves of the multisource cases are scaled horizontally such that when one draws a vertical line across this figure, the intersected red and green points represent 1/8 of the computational cost than the cost of the intersected blue point.

To probe the qualities of FWI, RTM images are shown in Figure 3.12, with

zoomed views shown in Figure 3.13. Note that the results of standard FWI and of FWIMFS appear identical. In the three zoomed views, the FWI results seem more focused, showing better continuity of reflectors. In the center of the RTM images, around $(X, Depth) = (9.5, 1.5)$ km, however, the traveltome tomogram offers better focusing than FWI tomograms. This is perhaps due to the overhanging low velocity anomalies in the FWI tomograms, as remarked on Figure 3.9.

Comparing the CIGs in Figure 3.14, we see flatter events on the right part, delimited by the red box. This coincides with the observations about the RTM images in that the FWI results fare better (in terms of more focused and continual reflectors) on the right part of the RTM image. In the green box of the CIGs, although the FWI results are flatter than the traveltome counterpart, the FWI velocity is over-corrected to be too small. This is indicated by the events curving up with larger source–midpoint offset (larger offset is on the left side of each CIG strip). This supports the idea that the overhanging low velocity anomalies are not authentic.

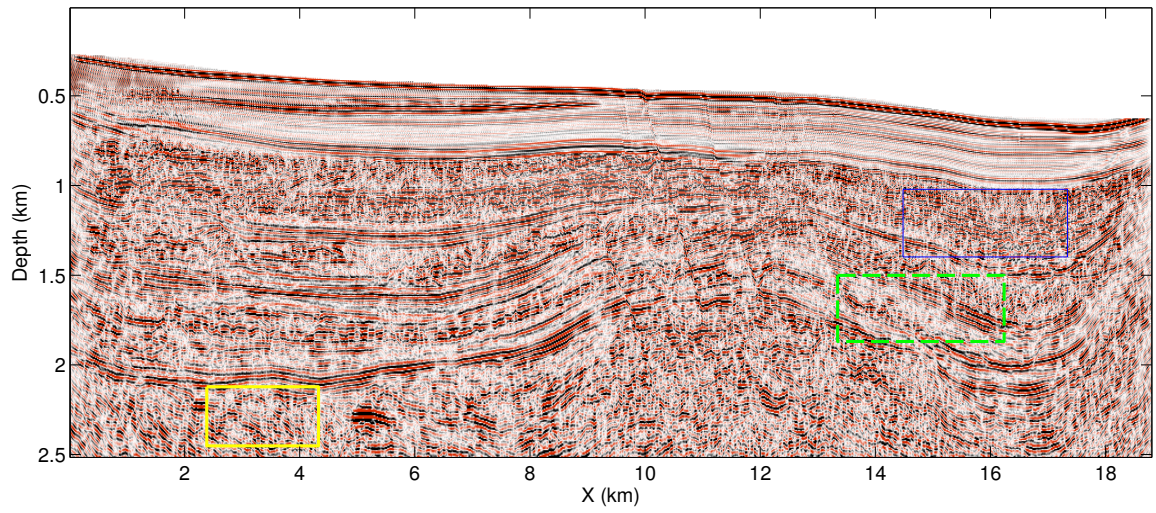
Figures 3.15 and 3.16 show CSGs from two sources, obtained by various methods. The match between the FWI results and the observed ones is generally good, although cycle skipping can be found at offsets = 3.8 km and 6 km, while comparing panels (a) and (c) in Figure 3.15.

3.4 Conclusions

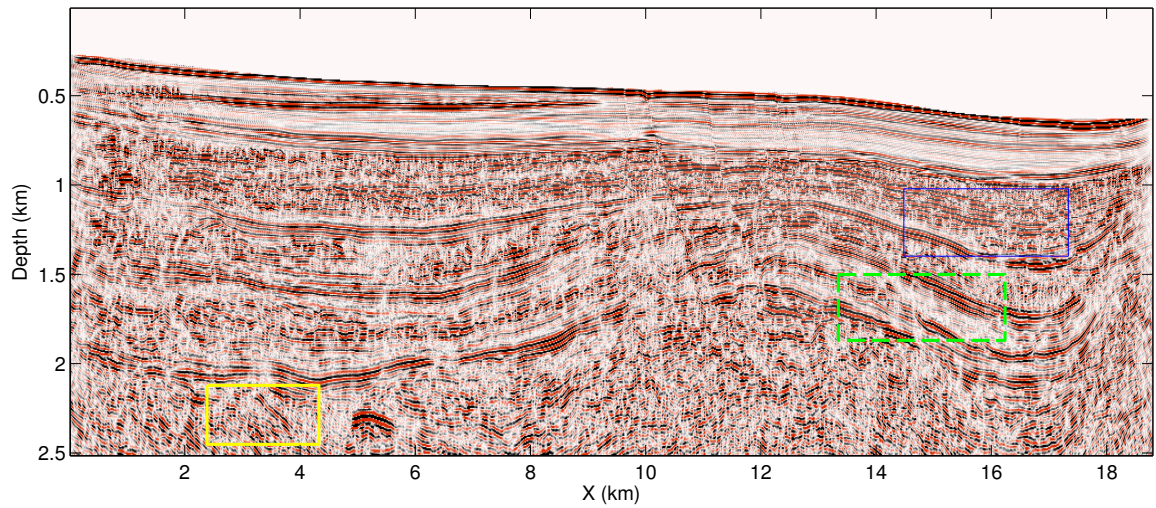
Multisource full waveform inversion of supergatherers for marine data is implemented with a frequency selection strategy. The key enabling property of frequency selection is that it eliminates the crosstalk among sources, thus overcoming the aperture mismatch of marine multisource inversion. This method is now extended to the FDTD from the previous implementation with phase shift migration. Tests on multisource FWI of synthetic marine data set and GOM data set show

speedups of 4× and 8×, respectively, compared to conventional FWI.

RTM using traveltime tomogram



RTM using FWI tomogram (individual-sources)



RTM using FWI tomogram (multisources, Quasi-Monte Carlo)

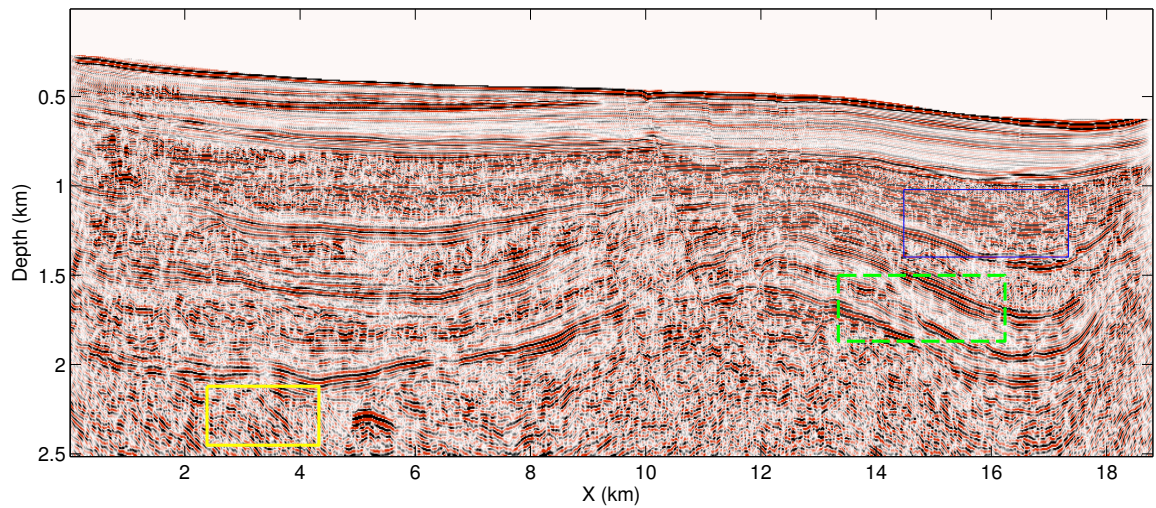


Figure 3.12: The resulting RTM images using various velocity models. The colored boxes are for zoomed views in Figure 3.13.

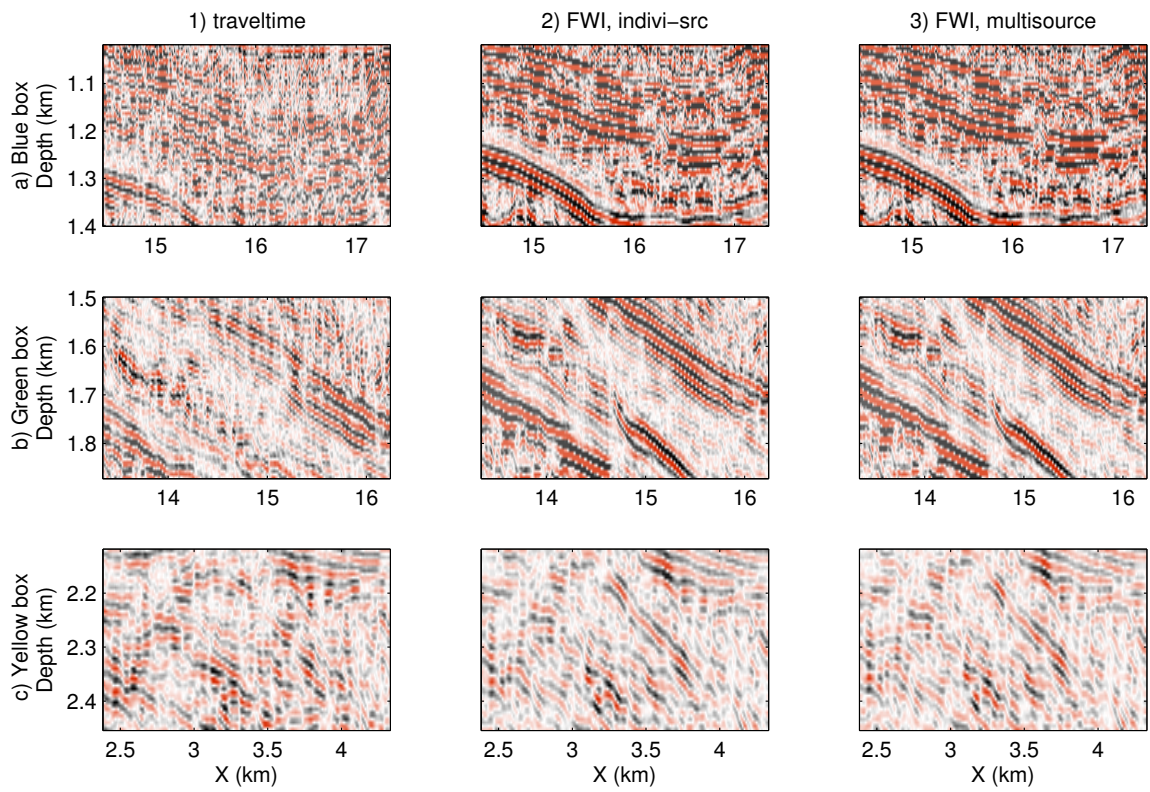


Figure 3.13: Zoomed views of the RTM images in Figure 3.12.

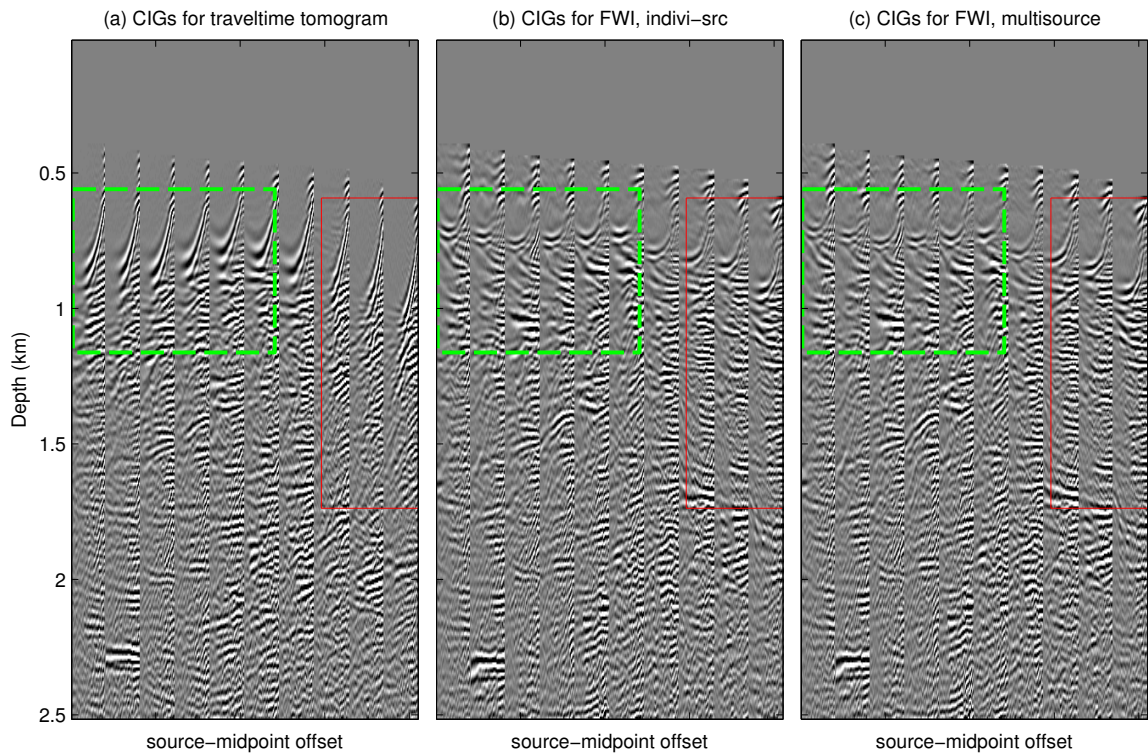


Figure 3.14: Shot domain CIGs. The boxed areas show improved flatness in the FWI results compared to that of the traveltime tomogram.

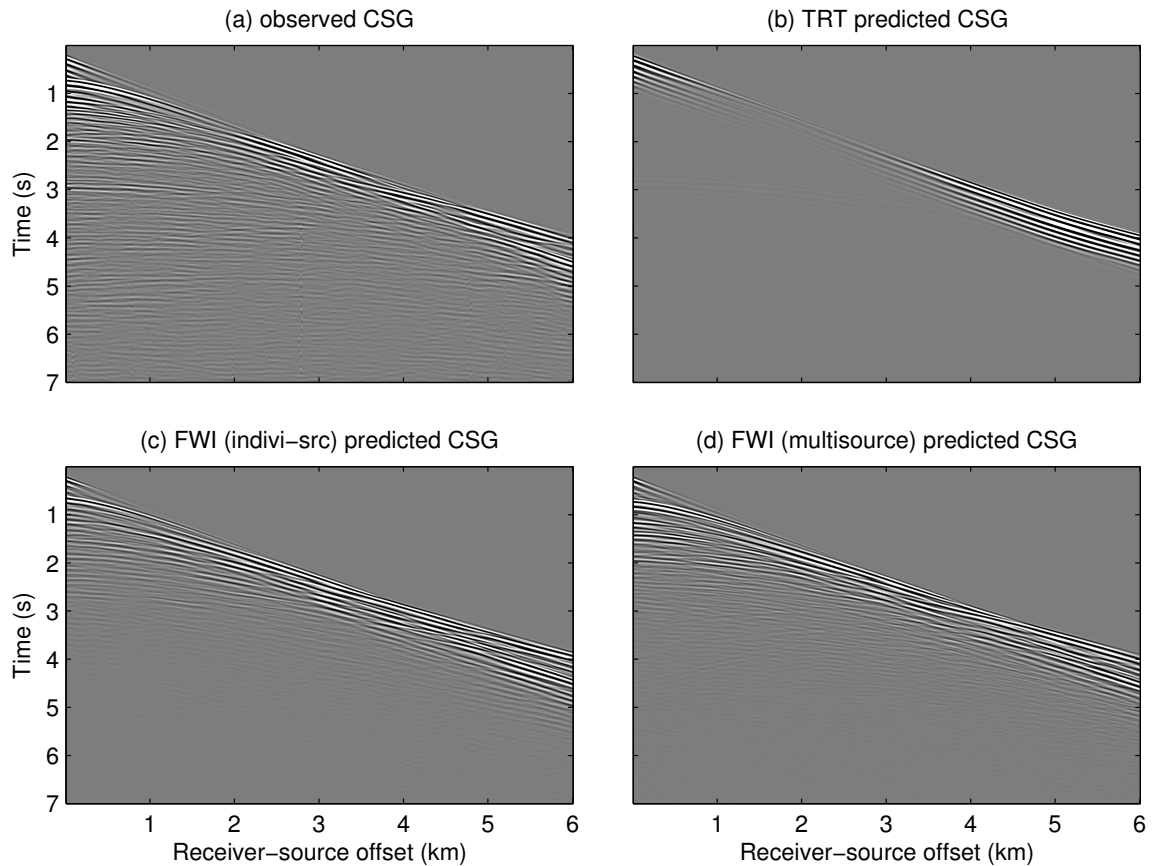


Figure 3.15: CSGs for GOM data of one source at $X = 6$ km. (a) the observed data bandpass filtered to 0–25 Hz. The predicted counterparts obtained with the (b) traveltime tomograms, (c) standard FWI (after 20 iterations), and (d) FWIMFS (after 71 iterations).

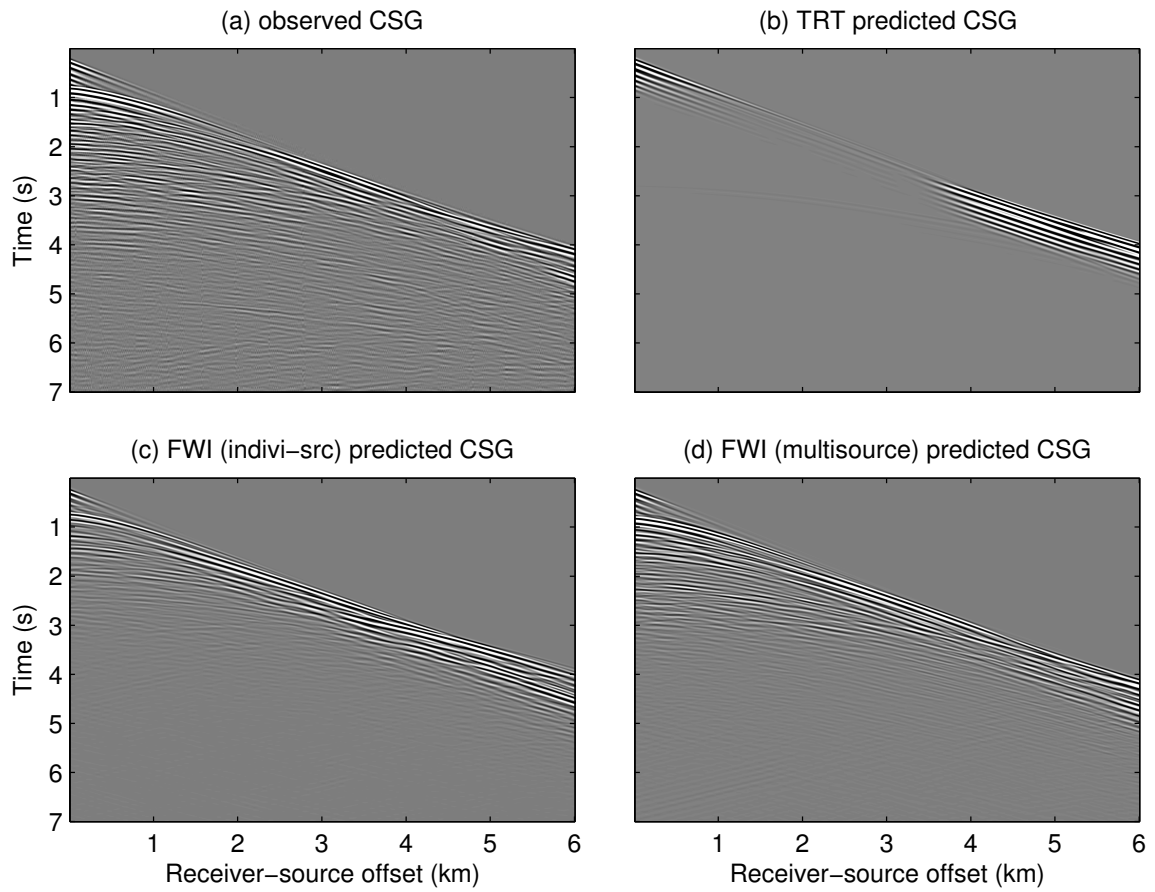


Figure 3.16: CSGs for GOM data of one source at $X = 12$ km. (a) the observed data bandpass filtered to 0–25 Hz. The predicted counterparts obtained with the (b) traveltime tomograms, (c) standard FWI (after 20 iterations), and (d) FWIMFS (after 71 iterations).

Chapter 4

Resolution Limits for Wave Equation

Imaging

4.1 Introduction

FWI (Lailly, 1984; Tarantola, 1984, 2005), RTM (McMechan, 1983; Baysal et al., 1983; Whitmore, 1983), LSM (Nemeth et al., 1999; Duquet et al., 2000; Tang, 2009; Dai et al., 2012) and wave-equation traveltime inversion (Woodward, 1989, 1992; Luo, 1991; Luo and Schuster, 1991; De Hoop and van Der Hilst, 2005) are important tools for imaging seismic data at the engineering (Buddensiek et al., 2008), exploration (Mora, 1988, 1989; Pica et al., 1990; Pratt and Gouly, 1991; Zhou et al., 1995; Shin and Cha, 2008; Krebs et al., 2009; Virieux and Operto, 2009) and earthquake (Marquering et al., 1999; Tong et al., 1998; De Hoop and van Der Hilst, 2005; Van Der Hilst and Maarten, 2005; Tape et al., 2009; Fichtner et al., 2009; Fichtner, 2011; Fichtner and Trampert, 2011b,a) scales. In all of the above methods, the wave equation is solved to estimate the model that minimizes, in some sense, the difference between the predicted and observed data. The main value of these wave equation-based imaging methods is that they overcome the high-frequency assumption of ray-based methods and use many, if not all, of the arrivals to reconstruct a finely detailed earth model. The hope is to find models with spatial

resolution of one-half wavelength, and perhaps even better if evanescent energy can be exploited (de Fornel, 2001; Fink, 2008; Schuster et al., 2012). The main limitations of wave equation imaging are computational cost, extensive preprocessing of the data, and extensive trial-and-error testing compared to the ray-based methods.

To optimize the use of wave equation imaging one must understand its limits of spatial resolution. Without this understanding, models can be over parameterized and lead to solutions that honor the data but violate the wavelength-based resolution limits of wave propagation. Such models should be avoided in our attempts to understand the earth.

In the last 30 years there has been much progress in mathematically defining the resolution limits of seismic images. The two most important categories of seismic imaging and their resolution limits are for traveltimes tomography and reflection imaging.

4.1.1 Resolution Limits for Traveltime Tomography

In raypath traveltimes tomography, the velocity is updated only along the raypath that connects the source at \mathbf{s} with the receiver at \mathbf{g} , whereas in finite-frequency traveltimes tomography, velocity updates can be confined to the first Fresnel zone for the specified source-receiver pair (Harlan, 1990). He states “...band-limited waves can follow paths that are not Fermat raypaths and still cover the distance between two points in almost the same time. All arriving waves that are delayed by less than half a wavelength will add constructively to the first arrival.”

As an example, the raypaths and Fresnel zones for reflection and transmission arrivals are illustrated in Figure 4.1. A point \mathbf{x} is in the Fresnel zone (FZ) if and only if it satisfies the following condition (Kravtsov and Orlov, 1990; Cerveny and Soares, 1992):

$$|\tau_{sx} + \tau_{xg} - \tau_{sg}| \leq T/2, \quad (4.1)$$

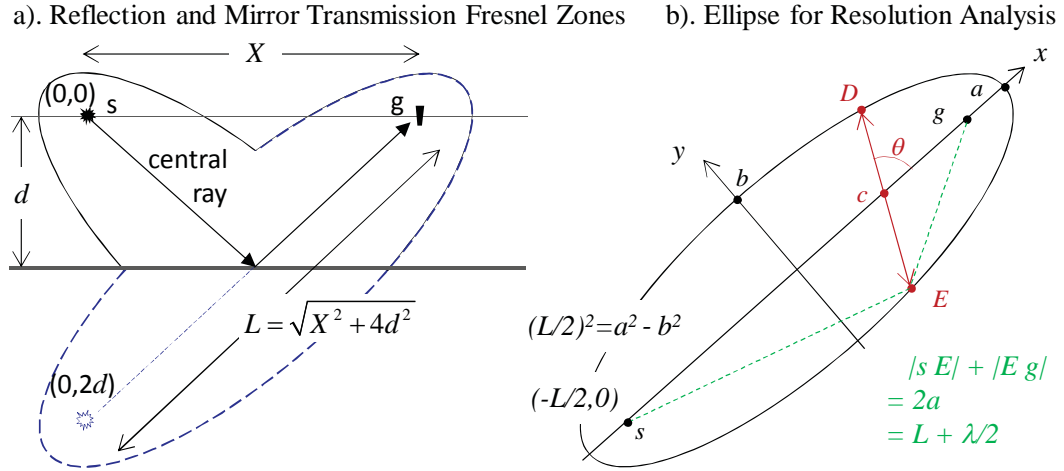


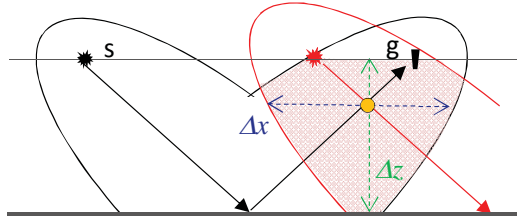
Figure 4.1: a) First-Fresnel zones for the specular reflection and for the transmission arrival excited by the mirror source at $(0,2d)$. In the latter case, the velocity below the reflector has been extended to be the same as the top-layer velocity. b) An ellipse intersected by a line segment DE , where its length $\overline{DE} = \frac{2ab \sqrt{b^2 \cos^2 \theta + (a^2 - c^2) \sin^2 \theta}}{b^2 \cos^2 \theta + a^2 \sin^2 \theta}$ defines the resolution limit (see Appendix E).

where, T is the dominant period of the source wavelet, τ_{sx} is the traveltime for a particular type of wave to propagate from s to the trial image point at x , and τ_{sg} is the traveltime to propagate from s to the specified geophone at g .

In a homogeneous medium, the maximum width of the first Fresnel zone can be shown (Williamson, 1991) to be proportional to $\sqrt{\lambda L}$, where L is the source-receiver distance and λ is the dominant wavelength. Thus, widening the distance between the source and receiver lowers the spatial resolution of the traveltime tomogram. More generally, Appendix E derives the formula for the length between any two points on opposite sides of the ellipse, which provides the horizontal resolution limit for any orientation of the ellipse.

The effective spatial resolution limits Δx and Δz of traveltime tomograms can be estimated (Schuster, 1996) as the minimum width and height of the intersection of first Fresnel zones at the trial image point. As an example, Figure 4.2 shows the intersection zones for both a) reflection and b) transmission rays. At any point on the central ray path, the narrowest width is along the line perpendicular to this ray,

a). Reflection Resolution Limits



b). Transmission Resolution Limits

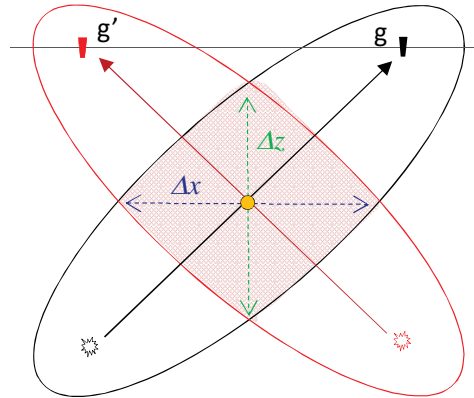


Figure 4.2: Same as Figure 4.1 except there is a new source (red star) along with its red wavepath for each diagram. The minimum width and height of the red-shaded intersection zone defines, respectively, the effective horizontal Δx and vertical Δz resolution limits of the traveltome tomogram at the yellow-filled circle.

which also defines the direction of best resolution. Thus, *a horizontal ray gives the best vertical resolution while a vertically oriented ray provides the best horizontal resolution for transmission tomography*, where the velocity is updated by smearing residuals along the first FZ (also referred to as a wavepath). As will be shown in the next section, this rule of thumb is also true for transmission wavepaths in FWI tomograms, except the *wave form residual* is smeared along the associated wavepath.

4.1.2 Resolution Limits for Reflection Imaging

A seismic migration image is formed by taking the reflection energy arriving at time $\tau_{sx} + \tau_{xg}$ and smearing (Claerbout, 1992) it along the appropriate ellipse in the model-space coordinates x (see Figure 4.3a). For several traces, the migration image in Figure 4.3b is formed by smearing¹² and summing the reflection energy along the appropriate ellipses in the model space. It is obvious that the *narrowest horizontal slice of the fat ellipse is for a trial image point at the far-left and far-right of*

¹²The seismic amplitude is smeared over the thick ellipse shown in Figure 4.3a, where the period T of the trace's source wavelet determines the thickness of the fat ellipse in (x, z) space; Figure 4.3b illustrates that the minimum thickness of the fat ellipse as 0.5λ .

Migration: Smear and Sum Reflection Amplitude Along Ellipses

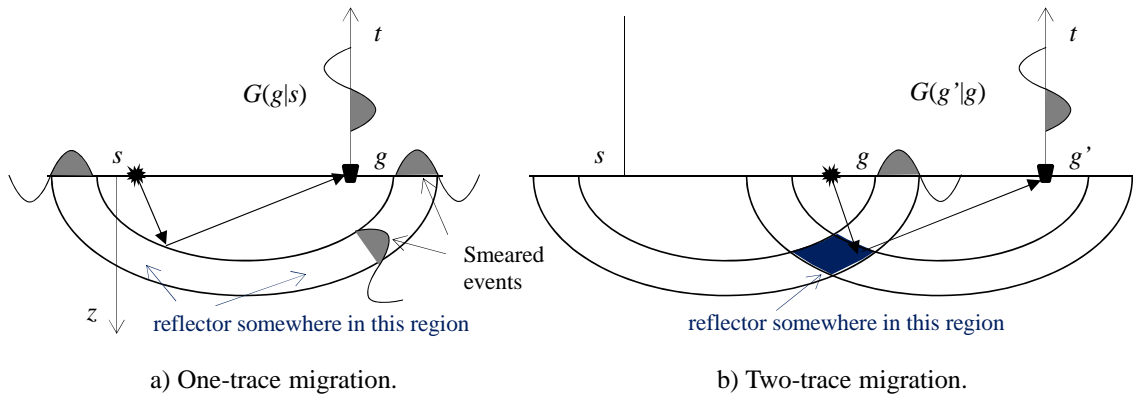


Figure 4.3: Migration is the smearing and summation of trace amplitudes along the appropriate fat ellipses in (x, z) for each source-receiver pair $s - g$ (Claerbout, 1992). Migration of two traces in b) has better spatial resolution than migrating just one trace in a), and the minimum thickness of each fat ellipse is 0.5λ , where T is the dominant period of the source wavelet.

the ellipse to give the best horizontal resolution in the reflection migration image. We also see that the narrowest vertical slice is directly beneath the midpoint of the source-receiver pair to give the best vertical resolution. For poststack data, these resolution limits are given on the rightside of Figure 4.4e, which say that the far-offset (near-offset) trace from a trial image point gives the best horizontal (vertical) resolution.

The resolution limits for migration (Berkhout, 1984; Safar, 1985; Chen and Schuster, 1999; Vermeer, 1997) were later found to be equivalent to those for linearized inversion in a homogeneous (Devaney, 1984; Wu and Toksoz, 1987) and an inhomogeneous medium (Beylkin, 1985) with smooth velocity variations. The key idea is that the model wavenumber vector \mathbf{k} can be equated to the sum of the source-scatterer and geophone-scatterer wavenumbers $\mathbf{k} = \mathbf{k}_{gr_0} + \mathbf{k}_{sr_0}$ shown in Figure 4.5. If \mathcal{D} defines the range of wavenumbers available from the source-receiver positions, then the horizontal Δx and vertical Δz spatial resolution limits of the

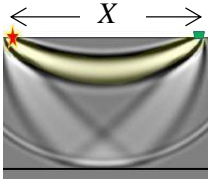
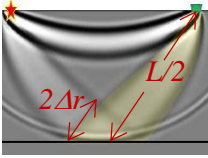
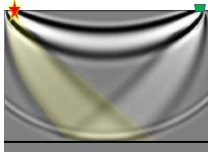
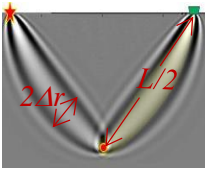
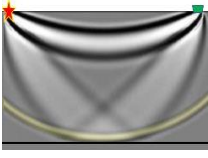
Migration Kernel	Data Kernel	Wavepath	Resolution Limit
a). $[G(\mathbf{x} \mathbf{s})^{\text{dir}} G(\mathbf{g} \mathbf{x})^{\text{dir}*}]$	Diving Wave $G(\mathbf{y} \mathbf{s})^{\text{dir}} G(\mathbf{g} \mathbf{y})^{\text{dir}}$		$2\Delta z = \sqrt{X\lambda}$
b). $[G(\mathbf{x} \mathbf{s})^{\text{rfl}} G(\mathbf{g} \mathbf{x})^{\text{dir}*}]$	Reflection $G(\mathbf{y} \mathbf{s})^{\text{rfl}} G(\mathbf{g} \mathbf{y})^{\text{dir}}$		$2\Delta r = \sqrt{L\lambda}$
c). $[G(\mathbf{x} \mathbf{s})^{\text{dir}} G(\mathbf{g} \mathbf{x})^{\text{rfl}*}]$	Reflection $G(\mathbf{y} \mathbf{s})^{\text{dir}} G(\mathbf{g} \mathbf{y})^{\text{rfl}}$		$2\Delta r = \sqrt{L\lambda}$
d). $[G(\mathbf{x} \mathbf{s})^{\text{dff}} G(\mathbf{g} \mathbf{x})^{\text{dir}*}]$	Diffraction $G(\mathbf{y} \mathbf{s})^{\text{dff}} G(\mathbf{g} \mathbf{y})^{\text{dir}}$		$2\Delta r = \sqrt{L\lambda / 4}$
e). $[G(\mathbf{x} \mathbf{s})^{\text{dir}} G(\mathbf{g} \mathbf{x})^{\text{dir}*}]$	Reflector $G(\mathbf{y} \mathbf{s})^{\text{dir}} G(\mathbf{g} \mathbf{y})^{\text{rfl}}$		$\Delta x = Z\lambda / (4X)$ $\Delta z = \lambda / 4$

Figure 4.4: Migration-data kernels, associated wavepaths, and approximate resolution limits along the middle of the wavepaths for a-d. Here, the dashed and solid lines in cyan represent the raypaths associated with the conjugated kernels and the data kernels, respectively; the trial image points \mathbf{x} and \mathbf{y} are represented by \bullet ; and the resolution limit perpendicular to the wavepath is denoted by $2\Delta r$. The resolution limits for reflection migration in e) are for poststack data, where X corresponds to aperture width, and Δx and Δz correspond to the skinniest width and thickness of the fat migration ellipse.

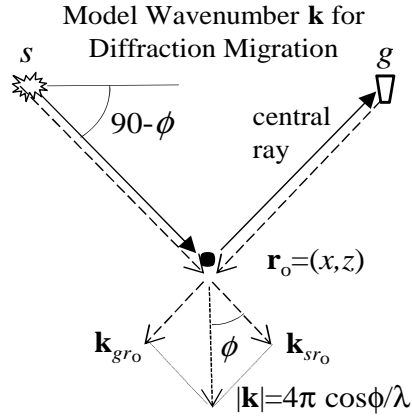


Figure 4.5: Scatterer at $\mathbf{r}_o = (x, z)$ where the sum of the dashed source-scatterer \mathbf{k}_{sr_o} and geophone-scatterer wavenumbers \mathbf{k}_{gr_o} is equal to the recoverable model wavenumber \mathbf{k} . Solid rays define the central raypath.

migration image are defined as

$$\Delta x = \min_{\mathcal{D}} \left[\frac{2\pi}{k_{gx} + k_{sx}} \right],$$

$$\Delta z = \min_{\mathcal{D}} \left[\frac{2\pi}{k_{gz} + k_{sz}} \right]. \quad (4.2)$$

In the farfield approximation these limits are given in Figure 4.4e for poststack migration.

The above resolution analysis have been developed for migration and traveltime tomography, and until now, there has not been a comprehensive treatment of the resolution limits associated with FWI. I now present such an analysis by applying an asymptotic analysis to the model resolution function for FWI. The resulting resolution formulas can be used to better understand and optimize the resolution characteristics of FWI, LSM, and RTM.

This chapter is organized into three sections. The introduction heuristically explains how wavepaths are used to estimate resolution for both traveltime tomography and migration. This leads to an intuitive description of spatial resolution as the minimum width and height of the intersection of Fresnel zones at the trial image point. The next section validates this heuristic definition by rigorously deriving the resolution limits for each type of wavepath, and explains their relationship to the acquisition geometry. Finally, a discussion and summary is given.

4.2 Born Forward and Adjoint Modeling

The model resolution function $\mathbf{L}^+\mathbf{L}$ is a product of forward \mathbf{L} and adjoint \mathbf{L}^+ modeling operators under the Born approximation (Stolt and Benson, 1986). I now define the equations for these modeling operators.

4.2.1 Born Forward Modeling

The trace $d(\mathbf{g}|\mathbf{s})$ excited by a harmonic point source at \mathbf{s} and recorded by a geophone at \mathbf{g} is given by the Born modeling equation:

$$\begin{aligned}\delta d(\mathbf{g}|\mathbf{s}) &= \omega^2 \int_{\Omega} G(\mathbf{g}|\mathbf{x}) \delta m(\mathbf{x}) G(\mathbf{x}|\mathbf{s}) d\mathbf{x}^2 \\ &\rightarrow \delta \mathbf{d} = \mathbf{L} \delta \mathbf{m},\end{aligned}\tag{4.3}$$

where $G(\mathbf{g}|\mathbf{x})$ is the Helmholtz Green's function for the background velocity model, the model function perturbed from the background model is given by $\delta m(\mathbf{x}) = 2\delta s(\mathbf{x})s(\mathbf{x}) \rightarrow \delta \mathbf{m}$, $s(\mathbf{x})$ is the background slowness model, $\delta s(\mathbf{x})$ is the perturbation of the slowness field, and ω is the angular frequency. For notational economy, this equation can be represented in operator notation by $\delta \mathbf{d} = \mathbf{L} \delta \mathbf{m}$, where $\delta \mathbf{d}$ represents the scattered seismic field $\delta d(\mathbf{g}|\mathbf{s})$ under the weak scattering approximation, \mathbf{L} represents the integral operator, and Ω defines the integration points in the model region.

The integration in equation 4.3 is over the entire model space, but if the trace is windowed about a specific event then the integration can be approximated by that over the event's first Fresnel zone associated with the specific source–receiver pair. For example, if the trace only contains the transmitted arrival, then $\Omega = \Omega_{trans}$ defines the points in the yellow colored wavepath in Figure 4.4a of the diving wave's first Fresnel zone; only velocity perturbations in this zone will significantly

affect the character of the diving wave arrival in the trace.

4.2.2 Born Adjoint Modeling

Equation 4.3 can be inverted by the iterative steepest descent formula

$$\delta m(\mathbf{x})^{k+1} = \delta m(\mathbf{x})^k - \alpha \delta m(\mathbf{x})^{\text{mig}}, \quad (4.4)$$

where the misfit gradient $\delta m(\mathbf{x})^{\text{mig}}$ is given by the Born adjoint modeling equation

$$\begin{aligned} \delta m(\mathbf{x})^{\text{mig}} &= \omega^2 \int_D G(\mathbf{g}|\mathbf{x})^* G(\mathbf{x}|\mathbf{s})^* \delta d(\mathbf{g}|\mathbf{s}) dx_g dx_s \\ &\rightarrow \delta \mathbf{m}^{\text{mig}} = \mathbf{L}^\dagger \delta \mathbf{d}, \end{aligned} \quad (4.5)$$

and the integration of points in D is over the range of horizontal source and receiver coordinates along the horizontal recording line at $z = 0$. Here, $\delta d(\mathbf{g}|\mathbf{s}) = d(\mathbf{g}|\mathbf{s}) - d(\mathbf{g}|\mathbf{s})^{\text{obs}}$, \mathbf{L}^\dagger represents the adjoint of the modeling operator \mathbf{L} , the step length is denoted by α , $d(\mathbf{g}|\mathbf{s})$ is the trace predicted from the estimated slowness model, and the observed trace is represented by $d(\mathbf{g}|\mathbf{s})^{\text{obs}}$. The misfit gradient symbol $\delta m(\mathbf{x})^{\text{mig}}$ is superscripted by *mig* because it also represents the migration of the residual. In fact, the first iteration $k = 0$ of equation 4.5 represents the reverse time migration of the scattered data recorded at the surface.

If the windowed event is the reflection, equation 4.5 says that the velocity model is updated by smearing the residual¹³ along the yellow-colored rabbit ears in Figure 4.4b-c and the yellow ellipse in Figure 4.4e. Smearing residuals along the rabbit ears (ellipse) with the b-c (e) migration kernel updates the low-wavenumber (high-wavenumber) portion of the velocity model (Mora, 1989; Zhou et al., 1995; Liu et al., 2011). The spatial resolution limits Δx and Δz associated with any point

¹³The residual can be either the travelt ime residual or the waveform residual.

along the central rays are determined by the, respectively, horizontal and vertical widths of the first-Fresnel zone.

4.3 Model Resolution Function and FWI Resolution Limits

In the Introduction, the model resolution limits were defined for traveltime transmission tomography and reflection migration. Now they will be derived for FWI by applying asymptotic analysis to the model resolution function that relates the model $\delta\mathbf{m}$ to the reconstructed image $\delta\mathbf{m}^{\text{mig}}$.

4.3.1 Model Resolution Equation: $\mathbf{m}^{\text{mig}} = \mathbf{L}^\dagger \mathbf{L} \mathbf{m}$

The forward and adjoint modeling equations can be combined to give the equation for model resolution, i.e., plugging equation 4.3 into equation 4.5 gives

$$\delta m(\mathbf{x})^{\text{mig}} = \omega^4 \int_D \overbrace{[G(\mathbf{g}|\mathbf{x})G(\mathbf{x}|\mathbf{s})]^*}^{\text{migration kernel}} \int_\Omega \overbrace{G(\mathbf{g}|\mathbf{y})G(\mathbf{y}|\mathbf{s})}^{\text{data kernel}} \delta m(\mathbf{y}) dy^2 dx_g dx_s, \quad (4.6)$$

or in more compact notation

$$\delta \mathbf{m}^{\text{mig}} = \mathbf{L}^\dagger \mathbf{L} \delta \mathbf{m}. \quad (4.7)$$

The kernel for the operator $\mathbf{L}^\dagger \mathbf{L}$ is related to the model resolution matrix (Menke, 1989) and is interpreted as the point spread function (Schuster and Hu, 2000) similar to that used in optics, except here, if $\delta m(\mathbf{y}) = \delta(\mathbf{y} - \mathbf{r}_o)$, it is the migration response to a point slowness perturbation in the model at \mathbf{r}_o . The ideal response to a point slowness anomaly is the same point with perfect resolution.

For a two-layer medium, the above Green's function can be decomposed into

its direct and reflection components:

$$\begin{aligned} G(\mathbf{g}|\mathbf{x}) &= G(\mathbf{g}|\mathbf{x})^{\text{dir}} + G(\mathbf{g}|\mathbf{x})^{\text{rfl}}, \\ G(\mathbf{s}|\mathbf{x}) &= G(\mathbf{s}|\mathbf{x})^{\text{dir}} + G(\mathbf{s}|\mathbf{x})^{\text{rfl}}, \end{aligned} \quad (4.8)$$

where $G(\mathbf{g}|\mathbf{x})^{\text{dir}}$ and $G(\mathbf{g}|\mathbf{x})^{\text{rfl}}$ are the, respectively, Green's function for the direct wave and upgoing reflection in the upper layer. Inserting equation 4.8 into the migration kernel in equation 4.6 results in the five migration kernels shown in Figure 4.4, each of which is used to smear residuals along one of the five yellow-colored wavepaths (Liu et al., 2011; Zhan et al., 2013).

Reflection–Migration

Reflection migration smears residuals along the yellow-colored ellipse in Figure 4.4e for a specified source and receiver pair. When two traces are migrated, Figure 4.3b suggests that the minimum width and height of the intersecting fat ellipses defines the resolution limits of reflection migration.

The formulas for migration resolution limits were more rigorously derived (Beylkin, 1985) by applying the migration kernel to traces that only contain the diffraction arrival from a single diffractor. For a localized scatterer¹⁴ in a background medium with smooth velocity variations, Beylkin showed that equation 4.6 asymptotically becomes the Fourier integral over the model wavenumbers k_x and k_z :

$$\delta m(\mathbf{x})^{\text{mig}} = \alpha \int_{\mathcal{D}} e^{-i\mathbf{k}\cdot\mathbf{x}} \delta M(\mathbf{k}) J^{-1} dk_x dk_z, \quad (4.9)$$

where α is related to geometrical spreading, J is the Jacobian, which is derived in Appendix G, and the range of model wavenumbers \mathcal{D} in the integral depends on

¹⁴We will assume a 2D model where the “point” source and scatterer are equivalent to a line source and a line scatterer, with no field variations along the y -axis.

the range of source-receiver pairs. In fact, the model wavenumber vector \mathbf{k} can be equated to the sum of the source-scatterer and geophone-scatterer wavenumbers $\mathbf{k} = \mathbf{k}_{gr_o} + \mathbf{k}_{sr_o}$ shown in Figure 4.5. I will now show how equations 4.6 and 4.9 can be used to estimate the resolution limits of the other wavepaths in Figure 4.4a-d.

Diving-Wave Transmission

Migration of the diving-wave residual along the yellow transmission wavepath in Figure 4.4a provides the low-wavenumber velocity update for waveform inversion (Mora, 1989; Zhou et al., 1995), or wave equation traveltime inversion (Woodward, 1989, 1992; Luo and Schuster, 1991) if the trace residual is replaced by the recorded trace weighted by the traveltime residual. The boundary of the first Fresnel-zone wavepath¹⁵ is defined by values of \mathbf{x} for the delayed diving wave time $\tau_{sg}^{\text{dive}} + T/2 = \tau_{sx} + \tau_{xg}$, where τ_{sg}^{dive} is the diving wave traveltime at the geophone location \mathbf{g} . As illustrated in Figure 4.2b, the minimum width and height of the intersecting fat ellipses defines the effective resolution limits of transmission tomography (Williamson, 1991) or transmission migration (Sheley and Schuster, 2003).

More rigorously, Appendix F shows that the model resolution equation 4.6 for diving waves can be transformed into the Fourier integral

$$\delta m(\mathbf{x})^{\text{mig}} = \alpha \int_{\mathcal{D}_{r_o}} e^{-i\mathbf{k}\cdot\mathbf{x}} \delta M(\mathbf{k}) J^{-1} dk_x dk_z, \quad (4.10)$$

where α is a term related to geometrical spreading and \mathcal{D}_{r_o} defines the range of source-geophone pairs whose first Fresnel zone wavepaths visit the scatterer localized at \mathbf{r}_o . The formulas for resolution limits are the same as in equation 4.2, except \mathcal{D} is replaced by \mathcal{D}_{r_o} .

The range of allowable source-geophone pairs (see Sheng and Schuster (2003))

¹⁵Dahlen (2004) refers to the shape of a diving wavepath as a banana.

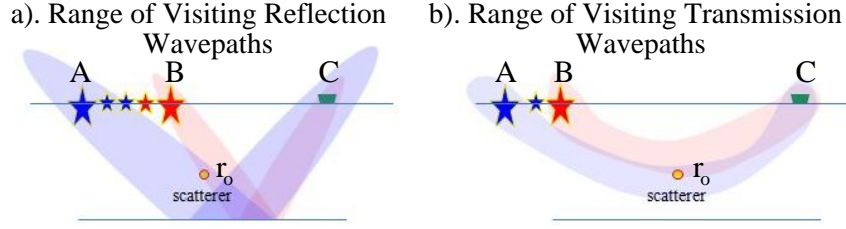


Figure 4.6: Range of sources (blue and red stars) that generate a) reflection and b) transmission wavepaths that visit the scatterer (filled circle). Here, the wavepath is approximated by the first Fresnel zone for the specified source and geophone pair.

in \mathcal{D}_{r_o} is illustrated in Figure 4.6b, where only the sources between the blue and red stars will contribute to the slowness update around the scatterer point at r_o . This differs from the Fourier integral 4.9 for diffraction imaging where *all* source-geophone pairs contribute to the integration domain in \mathcal{D} for a recorded diffraction. Hence, the resolution limits for migrating transmission residuals with the kernel $[G(\mathbf{g}|\mathbf{x})^{\text{dir}}G(\mathbf{x}|\mathbf{s})^{\text{dir}}]^*$ should be worse than migrating diffraction residuals with the same kernel.

The precise connection between intersecting wavepaths in Figure 4.2b, the range of available wavenumbers, and resolution limits in equation 4.2 can be made by assuming a homogeneous medium. In this case, Figure 4.7 shows that the half-width Δz of the first Fresnel zone at the point midway between the source and geophone is equal to

$$\Delta z = \sqrt{L\lambda/4}, \quad (4.11)$$

where L is the distance between the source and geophone, which is equal to that given by equation 4.2. It also shows that Δz is inversely proportional to the sum of the source-scatterer and geophone-scatterer wavenumbers, implying that $\min 1/k_z$ is equivalent to finding the width of the wavepath intersections in Figure 4.2b.

For a single source-geophone pair, the *best direction of transmission spatial reso-*

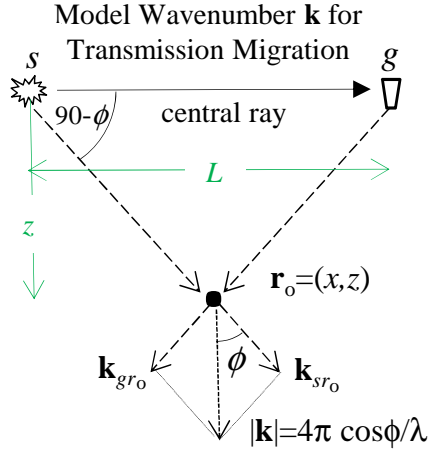


Figure 4.7: Transmission ray and scatterer at $\mathbf{r}_o = (x, z)$ where $x = L/2$ for a homogeneous medium. We assume $L \gg z$ and z is equal to the half-width $z \approx \sqrt{\lambda L/4}$ of the 1st-Fresnel zone in a homogeneous medium. In this case, $\cos\phi = z / \sqrt{L^2/4 + z^2} \approx 2z/L$; inserting the half-width formula gives $\cos\phi = \sqrt{\lambda/L}$ so $k_z = 4\pi \cos\phi / \lambda = 4\pi / \sqrt{L\lambda}$.

lution for a slowness anomaly midway between the source and geophone is perpendicular to the central ray. This means that a slowness anomaly moved perpendicular to the ray from the central ray will lead to the most noticeable change in the transmission arrival. The *worst direction of spatial resolution is along the ray itself* because the slowness anomaly can be slid along it without changing the traveltime; moreover, the model wavenumber $\mathbf{k} = \mathbf{k}_{sx} + \mathbf{k}_{xg}$ is zero all along the transmission central ray.

Reflection–Transmission

Migrating the reflection arrival with any of the kernels in the first column of Figure 4.4b–c leads to the low-wavenumber velocity update along the rabbit-ear wavepaths¹⁶ in Figure 4.4b–c or Figure 4.1a.

The corresponding resolution formula for the rightmost rabbit-ear wavepath is

$$\delta m(\mathbf{x})^{\text{mig}} = \omega^4 \int_{D_{r_o}} [G(\mathbf{g}|\mathbf{x})^{\text{dir}} G(\mathbf{x}|\mathbf{s})^{\text{refl}}]^* \int_{\Omega_{gs}} G(\mathbf{g}|\mathbf{y})^{\text{dir}} G(\mathbf{s}|\mathbf{y})^{\text{refl}} \delta m(\mathbf{y}) d\mathbf{y}^2 dx_g dx_s, \quad (4.12)$$

¹⁶There are two steps for creating an upgoing reflection wavepath: first, generate the migration image and use the reflectors as exploding sources that explode at the traveltime from the source to the reflector. Then, fire off these exploding reflectors to get the upgoing reflection fields $U(\mathbf{x}, t)$. The upgoing rabbit ear wavepath is computed by taking the zero-lag correlation between $U(\mathbf{x}, t)$ and the backpropagated data $B(\mathbf{x}, t)$.

and, as before, can be analyzed for the resolution limits. However, now the asymptotic Green's functions for the transmitted arrival $G(\mathbf{g}|\mathbf{x})^{\text{dir}}$ and the reflection field

$$G(\mathbf{x}|\mathbf{s})^{\text{refl}} = A_{sx}^{\text{refl}} e^{-i\omega\tau_{sx}^{\text{refl}}}, \quad (4.13)$$

are plugged into equation 4.12 to give the resolution limits for updating the velocity by smearing the reflection residual along the rabbit ears. Here, A_{sx}^{refl} accounts for amplitude and phase effects from geometrical spreading and the reflection coefficient, and τ_{sx}^{refl} is the time it takes reflection energy to propagate from the source at \mathbf{s} to the listener at $\mathbf{x} = \mathbf{r}_o$ along the specular dashed raypath in Figure 4.1a.

Estimating the resolution limits for the rabbit-ear wavepaths will result in model resolution formulas similar to that given in Figure 4.4a for transmission imaging. This can be understood without going through the detailed algebra by recognizing that the upgoing reflection wavepath (rightmost rabbit ear in Figure 4.1a) is identical to the transmission wavepath in Figure 4.1b above the interface. This is denoted as a mirror transmission wavepath because it coincides with the first Fresnel zone for a source at the mirror position $(0, 2d)$ in a homogeneous velocity. Thus, the reflection traveltime in a) is identical to the transmission traveltime in b) for any receiver at \mathbf{r}_o . This means that the resolution limits defined by equation 4.2 are applicable to the transmission wavepaths in Figure 4.1b and the reflection wavepaths in Figure 4.1a. However, the range of available wavenumbers for the traces recorded at \mathbf{g} is determined by the limited range of sources in Figure 4.6a that allow for the intersection of their first Fresnel zones with the scatterer. For example, the resolution limit $2\Delta r$ perpendicular to the ray at the midpoint should be equal to the $2\Delta r = \sqrt{\lambda L}$ in Figure 4.1a, except the total length of the reflection ray is $L = \sqrt{X^2 + 4d^2}$.

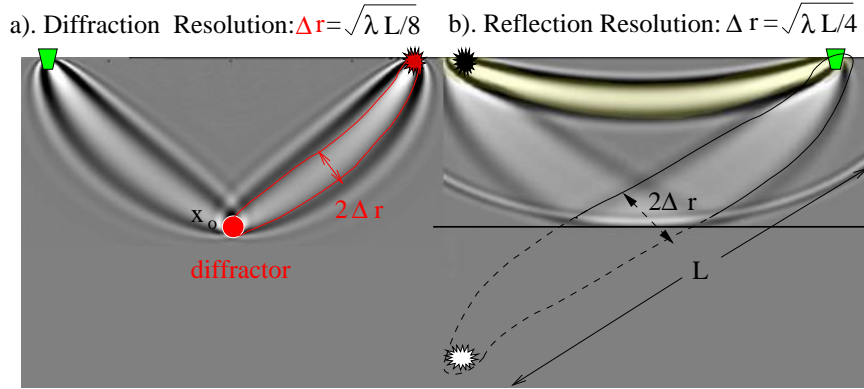


Figure 4.8: Wavepaths for migrating a) diffraction (red ellipse) and b) specular reflection (black ellipse) events along their transmission wavepaths. The diffraction resolution limit $2\Delta r$ perpendicular to the widest part of the diffraction wavepath is about half smaller than that for the specular reflection. Dashed wavepath in b) is the mirror image of the source–side wavepath, with the mirror source denoted by the white star. These wavepaths were obtained by first generating acoustic data for a a) diffractor model and a b) two-layer reflector model. Windowing about the scattered arrivals, the diffraction and reflection traces were then migrated, respectively, with the kernels $[G(\mathbf{g}|\mathbf{x})^{\text{diff}}G(\mathbf{x}|\mathbf{s})^{\text{dir}}]^*$ and $[G(\mathbf{g}|\mathbf{x})^{\text{dir}}G(\mathbf{x}|\mathbf{s})^{\text{dir}}]^*$.

Diffraction–Transmission

How do the resolution characteristics of the diffraction–transmission wavepaths in Figure 4.4d compare to those for the reflection–transmission wavepaths in Figure 4.4b-c? Figure 4.8 suggests that the diffraction resolution limit will be significantly better because the diffraction propagation distance is effectively halved, leading to a narrower wavepath. This means that, if the waveform residuals are used to update the velocity, then the diffraction updates will have significantly better resolution than the reflection updates.

The resolution limits for diffraction-transmission migration can be quantified according to equation 4.11, which says that the maximum resolution limits perpendicular to the diffraction and reflection central rays should be, respectively, $\Delta r^{\text{diff}} \approx \sqrt{\lambda L/4}$ and $\Delta r^{\text{refl}} \approx \sqrt{\lambda L}$. In this case, $L/2$ is the effective length of the central ray between the geophone and the scatterer in Figure 4.8a. These limits can

be rigorously derived by defining the diffraction Green's function $G(\mathbf{x}|\mathbf{s})^{\text{diff}}$ as

$$G(\mathbf{x}|\mathbf{s})^{\text{diff}} = A_{sx_0x}^{\text{diff}} e^{-i\omega(\tau_{sx_0} + \tau_{x_0x})}, \quad (4.14)$$

where the diffractor is located at \mathbf{x}_0 , the trial image point is at \mathbf{x} , and $A_{sx_0x}^{\text{diff}}$ accounts for the effects of geometrical spreading, reflection amplitude, and phase changes due to scattering. Replacing the migration kernel in equation 4.6 by $[G(\mathbf{g}|\mathbf{x})^{\text{diff}}G(\mathbf{x}|\mathbf{s})^{\text{dir}}]^*$ and the data kernel by $[G(\mathbf{g}|\mathbf{y})^{\text{diff}}G(\mathbf{y}|\mathbf{s})^{\text{dir}}]$, and using the explicit expression for the Green's functions yields the model resolution function for diffraction imaging:

$$\delta m(\mathbf{x})^{\text{mig}} = \omega^4 \int_D [A_{sx_0x}^{\text{diff}} A_{gx}]^* \int_{\Omega} A_{sx_0y}^{\text{diff}} A_{gy} e^{i\omega(\tau_{gx} - \tau_{gy} + \tau_{x_0x} - \tau_{x_0y})} dy^2 dx_g dx_s. \quad (4.15)$$

The salient difference between this formula and the one for reflections in equation F.4 is that τ_{x_0x} and τ_{x_0y} replace τ_{sx} and τ_{sy} . This says that the diffraction wavepath is generated by a "virtual" source at the diffractor \mathbf{x}_0 rather than at the actual source location \mathbf{s} . Hence, the diffraction wavepath should be thinner than the specular reflection wavepath in Figure 4.8. In addition, every source–geophone pair has a diffraction wavepath that intersects the diffractor. This means that, similar to diffraction migration, many more diffraction wavenumbers will be available for velocity updates compared to specular reflection–transmission wavepaths.

4.3.2 λ Imaging at the Diffractor

Figures 4.4d and 4.8a illustrate that the width of the diffraction-transmission wavepath is proportional to λ at the diffractor location. This can be mathematically proven by locating the point \mathbf{E} on the Figure 4.1b ellipse so that the line through it and the focus at \mathbf{g} is perpendicular to the elliptical axis. The distance between \mathbf{E} and \mathbf{g} is denoted as \overline{Eg} . In the farfield approximation, $L \gg \overline{Eg} = z_0$ so

we can approximate the ellipse formula for the first Fresnel zone centering about \mathbf{g} as

$$\lambda/2 = \sqrt{z_o^2 + L^2} + z_o - L \approx z_o. \quad (4.16)$$

This suggests that the resolution limit of the updated velocity model is about λ near the scatterer, which is much finer than the resolution limit of $\sqrt{L\lambda/4}$ along the middle of the Figure 4.8a wavepath. This unexpectedly high resolution limit near the reflector boundaries can be observed in wave equation reflection traveltimes (Zhang et al., 2012) and migration velocity analysis (Zhang and Schuster, 2013) tomograms.

To illustrate the range of wavenumbers estimated from diffraction and transmission migration, Figure 4.9 depicts the low wavenumbers (magenta dots) of the model recovered with transmission migration (see Figure 4.7) and the higher wavenumbers (gray dots) recovered by diffraction migration (Figure 4.5). Note the large gap between the recovered low- and high- wavenumber spectra, which will be denoted as the missing intermediate wavenumbers. The absence of such intermediate model wavenumbers is a serious challenge for waveform inversion (Jannane et al., 1989), which will be addressed in the next section.

4.4 Filling in the Model Spectrum with Multiples

The previous sections derived the model resolution equations for diving waves, primaries, and diffractions. What are the resolution benefits for migrating multiples, particularly prism waves or interbed multiples? The short answer is that their associated central rays are longer than those of primaries, so their first-Fresnel zones should be wider. This means that they can reconstruct low-wavenumber and intermediate-wavenumber models that can only be inverted with primaries

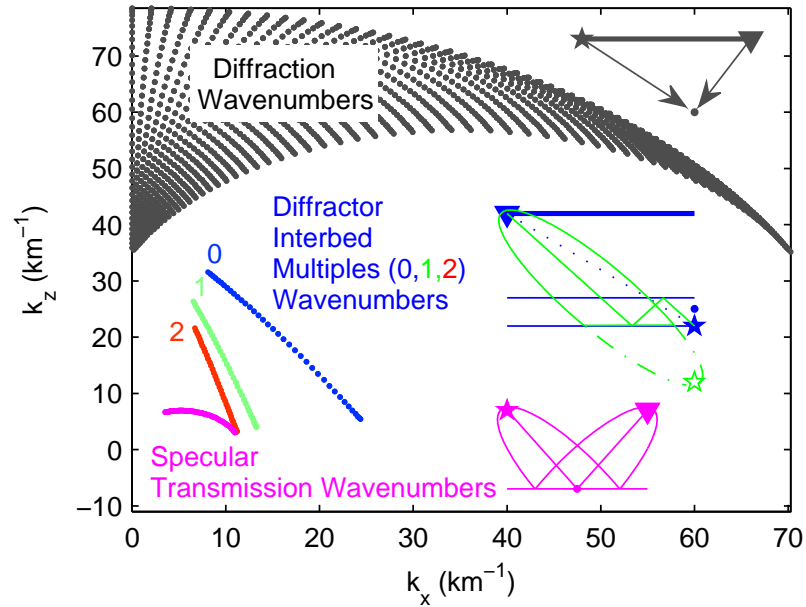


Figure 4.9: Wavenumbers (represented by gray dots) recovered by (i) diffraction migration have much higher values than those (represented by thick magenta curve) recovered by (ii) specular transmission migration. Blue, green and red dots represent the wavenumbers recovered based on (iii) interbed multiples originating from a diffractor, of multiples order 0, 1, and 2, respectively. These wavenumbers somewhat bridge the gap between cases (i) and (ii). The acquisition geometry is a 4 km long line of geophones and sources on the top interface, with the trial image point, denoted by \bullet , at depth 1 km. The diffractor in case (iii) is 80 m below the trial image point.

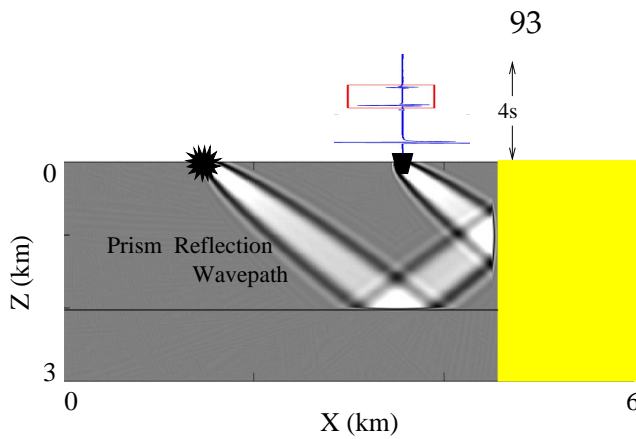


Figure 4.10: Vertical reflector model along with prism wavepath and trace computed by a finite-difference solution to the wave equation. RTM was applied to the windowed events to produce this wavepath.

at impractically wide source-geophone offsets or using sources with unrealistic low-frequencies.

4.4.1 Lower Wavenumber Resolution with Prism Waves and Free-surface Multiples

To demonstrate the enhanced wavenumber coverage of multiples, the point-source response of the yellow vertical reflector in Figure 4.10 is computed by a finite-difference method. The trace is windowed about the reflections and then migrated by RTM to get the prism wavepath image (Dai and Schuster, 2013). As the length L of the prism ray gets longer, the wavepath becomes thicker by $\sqrt{\lambda L}$. In this way, the deep prism-reflection wavepath provides lower wavenumber information about the model compared to primaries. Such low wavenumbers are at the top of the FWI wish list for providing a good starting model for subsalt imaging.

Another example is shown in Figure 4.11. Here, the prism-wave reflection in b) achieves the same low-wavenumber resolution as the 1st-order free-surface multiple in a), but only requires about 1/2 the source-geophone offset of a). The deeper the reflector for the free-surface multiple, the thicker the wavepath and the lower the wavenumber in the estimated model.

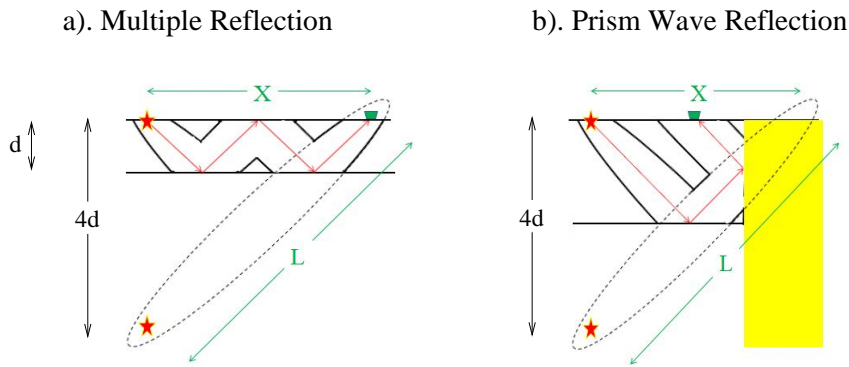


Figure 4.11: Wavepaths for a) the 1st-order free-surface multiple and b) the prism-wave reflection from the yellow block. Note, the source-geophone offset for b) is about 1/2 that for a), and the dashed ellipse is the wavepath for the mirror source at the bottom left.

4.4.2 Intermediate Wavenumber Resolution with Interbed Multiples

Figure 4.9 illustrates that transmission tomography (smearing residuals along rabbit ears) fills in the low wavenumber part of the spectrum, while reflection migration (smearing residuals along ellipse) fills in the high wavenumbers. Now I show that interbed multiples can fill in some of the intermediate wavenumbers denoted by the blue dots.

Figure 4.12a depicts the interbed multiple rays for a thin-bed model with a diffractor at the lower interface. Each order of the multiple will be associated with a different mirror ray, where the depth of the mirror scatterer deepens with the order of the multiple. Therefore the raypath lengthens with order of multiple, and the wavepath thickens as well. I conclude that the mirror wavepath that intersects the thin bed¹⁷ thickens progressively with the order of the multiple, and so should fill in some of the “intermediate wavenumber” gap (Jannane et al., 1989) illustrated by the blue dots in Figure 4.9.

The above analysis can be quantified as in the previous sections by analyzing the

¹⁷The sampling interval between wavenumbers associated with each order of multiple becomes smaller with thinner beds.

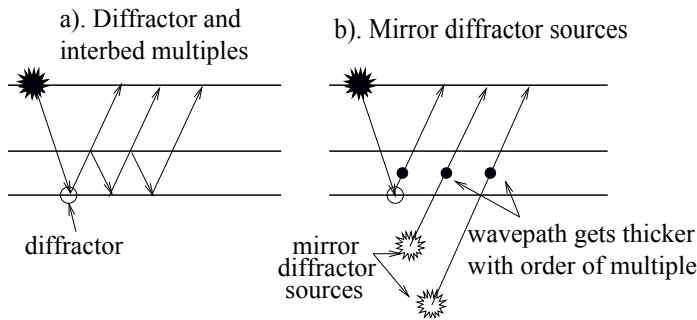


Figure 4.12: a) Ray diagram for interbed multiples generated by a diffractor in a thin layer and b) the associated mirror sources diagram. As the order of the interbed multiple increases, so does the thickness of its transmission wavepaths at the filled circles.

model resolution function. In this case, the forward modeling kernel $G(\mathbf{g}|\mathbf{x})G(\mathbf{x}|\mathbf{s})$ is replaced by one that generates an internal multiple rather than a direct wave or primary reflection¹⁸. The phase term in the Green's function will be replaced by a summation of times corresponding to each leg of the raypaths seen in Figure 4.12a. The migration kernel is also modified by terms that will image the internal multiple to one of its bounce points in the thin layer.

4.5 Discussion and Conclusions

Formulas are derived for the resolution limits of the migration-data kernels in Figure 4.4, as well as those for multiple reflections. They are applicable to images formed by RTM, LSM, and FWI. Their salient implications are the following.

1. Low- and intermediate-wavenumber information about the velocity distribution is estimated primarily by transmission migration of primaries and multiples. The intermediate wavenumbers can be supplied by interbed multiples, while the lower wavenumbers are contained in deep primaries and free-surface related multiples. Inverting multiples can be an opportunity for estimating subsurface velocity information not available in the primary reflections.

¹⁸This kernel corresponds to just one of the terms in the Neumann series expansion of the Lippmann-Schwinger equation (Stolt and Benson, 1986).

2. Inverting diffractions can provide twice or more the resolution compared to imaging primaries. Smearing residuals along the transmission wavepath can achieve a resolution of λ near the diffractor. This high resolution is observed in MVA tomograms. On the other hand, diffraction energy can be more than an order-of-magnitude weaker than primary energy, so the diffraction data will be noisier.
3. Diving waves that bottom out at a certain depth will have a better vertical resolution than horizontal resolution. Therefore it is important to also invert deep reflections to increase both the vertical and horizontal resolution. Since reflections can be an order-of-magnitude weaker than diving waves, it is recommended that diving waves be filtered from the data after a sufficient number of iterations. This might constitute an iterative multi-physics approach to FWI, where inverting a different type of wavefield should be emphasized at different depths and iteration numbers.
4. The transmission migration kernels in Figure 4.4a–d are of the same type as their data kernels. This leads to velocity updates along the transmission wavepaths. In contrast, the traditional migration kernel $[G(\mathbf{g}|\mathbf{x})^{\text{dir}}G(\mathbf{x}|\mathbf{s})^{\text{dir}}]^*$ in Figure 4.4e is a product of two Green's functions for direct waves, while the data kernel is a product of a reflection and a direct-wave Green's function. This mismatch in the type of kernel does not lead to the traditional wavepath where seismic energy propagates, but gives the untraditional wavepath of a fat ellipse. This is the zone where reflection energy could have originated, i.e., the interface.

The limitation of this study is that it does not take into account the non-linear effects of evanescent energy (Fleming, 2008) in determining resolution. Utilizing evanescent energy with FWI could provide, in theory, resolution much better than λ . It

is expected that multiple scattering arrivals between neighboring subwavelength scatterers might provide the extra resolution needed, but not accounted for in this current study.

Chapter 5

Concluding Remarks

5.1 Summary

In Chapter 2 I proposed frequency selection to remove multisource crosstalk, aiming to solve the problem of inconsistent sets of receivers in modeling compared to in marine streamer survey. This approach requires that any crosstalk-prone sources should only emit signals in non-overlapping frequency bands. This approach allows any receiver to selectively tune in to the valid sources: the ones that indeed have contributed to this receiver and to selectively disregard potentially confounding sources, which are not associated with this receiver in the survey. Numerical results with a 2D salt model by marine survey show that the crosstalk is completely removed and compared to the standard migration, the speedup is nearly 10×. In the 3D example with a fixed OBS geometry, a speedup of 40× is achieved.

In Chapter 3 I extended this research avenue to handle multisource FWI for marine streamer data. I adopted FDTD instead of the split-step method employed in Chapter 2. Sine waves of different frequencies are injected at sources that belong to one supergather. Due to delays in arrival time, an injected sine wave shows up at receivers delayed with onsets, which is no longer a pure sine wave. To mitigate this problem of spectral leakage, the simulation time is doubled, keeping only the second half of the duration. Numerical tests on synthetic and GOM data sets

demonstrate speedups around $4\times$.

Of interest to FWI, formulae are derived for the resolution limits of migration-data kernels associated with diving waves, primary reflections, diffractions, and multiple reflections. The formulae suggests that inverting multiples can provide some low- and intermediate-wavenumber components of the velocity model not available in the primaries. Futhermore, diffractions can provide twice or better the resolution as specular reflections for comparable depths of the reflector and diffractor. The width of the diffraction-transmission wavepath is on the order of λ at the diffractor location for the diffraction-transmission wavepath.

5.2 Future Research Work

The work presented in this thesis can be extended in the following directions.

- Multisource frequency selection applied to least-squares redatuming and multiples removal.
- Plane wave and frequency selection combined so that the (reflectivity or velocity) model can be illuminated by several plane waves of different frequencies and of different incident angles together.

APPENDICES

A The Effect of Frequency Selection on the Hessian

The effect of frequency selection on the Hessian matrix is now investigated. For ease of discussion, I restrict the attention to sources $s = 1, \dots, S$ that are used to form one supergather. The Hessian can be identified as

$$\tilde{\mathbf{H}} = \sum_{j=1}^{n_\omega} \tilde{\mathbf{L}}^+ \tilde{\mathbf{L}} \quad (\text{A.1})$$

$$= \sum_{s=1}^S \sum_{j=1}^{n_\omega} N_s(j) |W(j)|^2 \underline{\mathbf{L}}_s^+ \underline{\mathbf{L}}_s. \quad (\text{A.2})$$

Here, equation A.2 follows from equations 2.8, 2.4, 2.22, 2.21, and the fact that $N_s^2(j) = N_s(j)$, as $N_s(j) \in \{0, 1\}$; equation 2.21 ensures that all cross terms in equation A.1 when expanded by plugging in equation 2.8 will vanish.

In contrast, the Hessian in the standard case is

$$\mathbf{H} = \sum_{s=1}^S \sum_{j=1}^{n_\omega} |W(j)|^2 \underline{\mathbf{L}}_s^+ \underline{\mathbf{L}}_s, \quad (\text{A.3})$$

which is equation A.2 lacking the binary encoding function $N_s(j)$.

Comparing equations A.2 and A.3, we see that the encoded Hessian $\tilde{\mathbf{H}}$ consists of a subset of terms in the standard Hessian \mathbf{H} .

B Prestack Split-Step Migration

The steps of prestack split-step migration are described in the flowcharts shown in Figure B.1. This presentation closely follows Kuehl and Sacchi (1999), and is included here for convenience because analysis of computational complexity refers to it.

Consider first the forward propagation of a wavefield. The split-step operator \mathcal{L} per layer can be decomposed into a succession of four linear operators \mathcal{F} , \mathcal{P} , \mathcal{F}^{-1} , and C . First, the seismic wavefield $P(x, z, \omega)$ at z is transformed to the wavenumber k_x domain by the Fourier operator \mathcal{F} . Second, the phase-shift operator \mathcal{P} is applied to the wavefield in the (k_x, ω) domain:

$$P_1(k_x, z, \omega) = P(k_x, z, \omega)e^{-i\Delta z \sqrt{(\omega u_0)^2 - k_x^2}}, \quad (\text{B.1})$$

where u_0 is the mean slowness for the current layer. Third, $P_1(k_x, z, \omega)$ is transformed to the space x domain by the inverse Fourier operator \mathcal{F}^{-1} . Fourth, the phase correction operator C is applied in the (x, ω) domain. This accounts for the lateral slowness variation $\Delta u(x) = u(x) - u_0$:

$$P(x, z + \Delta z, \omega) = P_1(x, z, \omega)e^{-i\omega \Delta z \Delta u(x)}. \quad (\text{B.2})$$

Altogether, it is given that

$$\mathcal{L} = C\mathcal{F}^{-1}\mathcal{P}\mathcal{F}, \quad (\text{B.3})$$

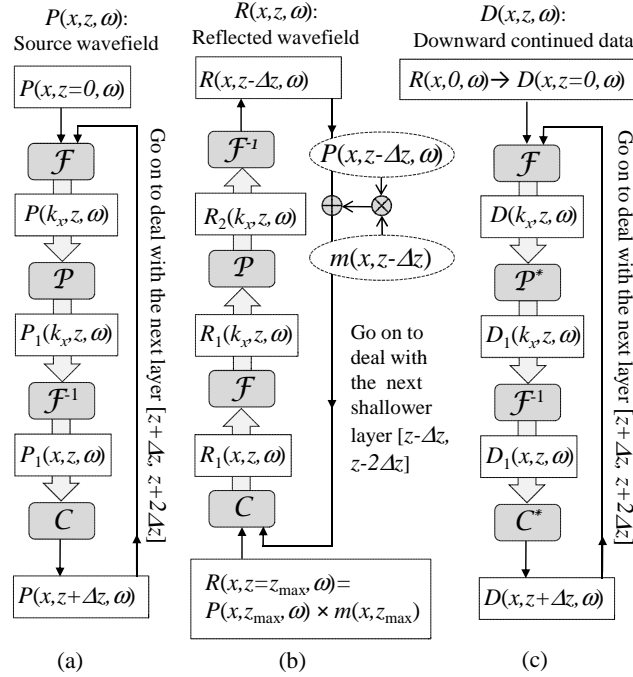


Figure B.1: Flowcharts for prestack split-step modeling and migration. \mathcal{F} , \mathcal{F}^{-1} , \mathcal{P} , and \mathcal{C} denote the Fourier transform, the inverse Fourier transform, the phase shift operator, and the phase correction operator, respectively (see text for details). (a) The source wavefield $P(x, z, \omega)$ is propagated from the surface of the earth to depth z in steps Δz . (b) At each depth z of the earth, a reflected wave is generated by $m(x, z)P(x, z, \omega)$, where $m(x, z)$ is a reflectivity model. The wave is then propagated upward to the surface $z = 0$. The total reflected wavefield $R(x, z, \omega)$ consists of the superposition of the reflected and propagated waves originating from below. The total reflected wavefield collected at the surface is the data, i.e., $R(x, z = 0, \omega) \equiv D(x, z = 0, \omega)$. (c) The data are then downward continued from the surface back to depth z in steps Δz . Finally, the migration image $I(x, z)$ (not shown) is constructed by applying the imaging condition: $I(x, z) = \sum_{\omega} P^*(x, z, \omega)D(x, z, \omega)$, or $I(x, z) = 2 \sum_{\omega > 0} \text{Re}\{P^*(x, z, \omega)D(x, z, \omega)\}$, assuming the DC component is 0.

of which the adjoint is

$$\begin{aligned}\mathcal{L}^\dagger &= \mathcal{F}^\dagger \mathcal{P}^\dagger \mathcal{F}^{-1\dagger} \mathcal{C}^\dagger \\ &= \mathcal{F}^{-1} \mathcal{P}^* \mathcal{F} \mathcal{C}^*,\end{aligned}\tag{B.4}$$

The adjoint operator \mathcal{L}^\dagger applies to the case of ‘backward propagation’, or downward continuation of the data, as illustrated in Figure B.1(c). This ensures that the migration operator is the adjoint of the forward modeling counterpart.

C The Relative Computational Cost

The computational costs of the proposed method and of standard migration (and therefore the relative computational cost) will be derived as follows, assuming the observed CSG's are already in the frequency domain. Since ultimately it is the ratio between the computational costs of different methods that is of interest, I restrict the attention to one frequency and to the S sources encompassed by one supergather.

First, spatial FFT's, repeatedly invoked in Figure B.1(a–c), dominate the computational cost of split-step migration. Element-wise product and dot product between two vectors of length n_x (the grid size along x) incur a cost C_p that is a small multiple of n_x , and the number of times such computation is invoked is comparable to that of FFT_x . On the other hand, the cost of each FFT_x in terms of $n_{\text{FLOP}} \cong \frac{34}{9}n_x \log_2(n_x)$ (Johnson and Frigo, 2006), which $\gtrsim 37n_x$ because n_x is on the order of a thousand or more for typical reflectivity models. This cost far exceeds that of C_p . Since every flowchart in Figure B.1 contains equal number of FFT_x 's, from now on take the computational cost for each of these flowcharts as of unit 1.

Next, the computational costs for forward modeling and for migration, in cases of whether the source field is available, are given by the following *items*:

1. Forward modeling

- (a) The source field, $P(x, z, \omega)$ as depicted in Figure B.1(a), is not available. Therefore $P(x, z, \omega)$ needs to be computed before the reflected field $R(x, z, \omega)$, as depicted in Figure B.1(b), can be obtained. The computa-

tional cost is thus 2.

- (b) The source field is available. Then the only task is to compute $R(x, z, \omega)$, with a computational cost of 1.

2. Migration

- (a) The source field is not available. Then both the source field and the downward continued data field $D(x, z, \omega)$, as depicted in Figure B.1(c), need to be computed. The computational cost is 2. Note: This applies to standard migration.
- (b) The source field is available. Then only the downward continued data field needs to be computed, before forming the final migration image. This computational cost is 1.

With these results in mind, I study the computational cost incurred in a CG algorithm, as listed in Algorithm 1. For standard migration, there are S sources covered by the supergather in question and therefore the computational cost as remarked at the end of item 2(a) needs a factor S . Let κ_0 be this cost, expressed as

$$\kappa_0 = 2S, \quad (\text{C.1})$$

and let κ_{LS} be the cost for LSM, given in the comment besides line 1.25 as

$$\kappa_{LS} = K_{it} \frac{3 + 2K_{CGit}}{K_{CGit}} - 1. \quad (\text{C.2})$$

Therefore the relative computational cost is given by

$$\rho = \frac{\kappa_{LS}}{\kappa_0} = \frac{K_{it}}{S} \frac{3 + 2K_{CGit}}{2K_{CGit}} - \frac{1}{2S} \quad (\text{C.3})$$

$$= \frac{3K_{it} - 1}{2S}, \quad \text{when } K_{CGit} = 3. \quad (\text{C.4})$$

Algorithm 1 Conjugate Gradient algorithm, for solving $\mathbf{Lm} = \mathbf{d}$.

```

1: procedure CONJGRAD( $\mathbf{m}, \mathbf{L}, \mathbf{d}, K_{CGit}$ )
2:   if  $\mathbf{m} = \mathbf{0}$  then
3:      $\mathbf{g} \leftarrow \mathbf{L}^T \mathbf{d}$  ▷ item 2(a)  $\Rightarrow$  cost=2.
4:   else
5:      $\mathbf{g} \leftarrow -\mathbf{L}^T(\mathbf{Lm} - \mathbf{d})$  ▷ item 1(a)
    and 2(b)  $\Rightarrow$  cost=2+1. In forward modeling, the source field is not available,  $\therefore$ 
    item 1(a); in migration, the source field has just become available,  $\therefore$  item 2(b).
6:   end if
7:    $\mathbf{p} \leftarrow \mathbf{g}$ 
8:    $g_{old}^2 \leftarrow \|\mathbf{g}\|^2$ 

9:   if  $K_{CGit} = 1$  then ▷ Steepest Descent
10:     $\alpha \leftarrow \frac{g_{old}^2}{\|\mathbf{Lp}\|^2}$  ▷ Forward modeling. Source fields have been computed in
    line 3 or 5.  $\therefore$  Item 1(b)  $\Rightarrow$  cost=1.
11:     $\mathbf{m} \leftarrow \mathbf{m} + \alpha \mathbf{p}$ 
12:    return  $\mathbf{m}$  ▷ For steepest descent, if this procedure is called  $K_{it}$  times, the
    total cost is  $3 + 4(K_{it} - 1) = 4K_{it} - 1$ .

13:   else ▷ Conjugate Gradient
14:     for  $k \leftarrow 1, K_{CGit}$  do
15:        $\mathbf{q} \leftarrow \mathbf{L}^T(\mathbf{Lp})$  ▷ Forward modeling and migration. Since among CG
        updates, the sources are fixed rather than encoded anew, source fields are thus
        fixed and have been computed in line 3 or 5. As the trial reflectivity model  $\mathbf{m}$  is
        being updated, item 1(b) applies. For migration, item 2(b) applies. So the total
        cost for this line is 2.
16:        $\alpha \leftarrow \frac{g_{old}^2}{\mathbf{p}^T \mathbf{q}}$ 
17:        $\mathbf{m} \leftarrow \mathbf{m} + \alpha \mathbf{p}$ 
18:        $\mathbf{g} \leftarrow \mathbf{g} - \alpha \mathbf{q}$ 
19:        $g_{new}^2 \leftarrow \|\mathbf{g}\|^2$ 
20:        $\mathbf{p} \leftarrow \mathbf{g} + \frac{g_{new}^2}{g_{old}^2} \mathbf{p}$ 
21:        $g_{old}^2 \leftarrow g_{new}^2$ 
22:     end for ▷ Taken together, for this loop, the total cost is  $2K_{CGit}$ .
23:     return  $\mathbf{m}$  ▷ As this procedure is called  $K_{CG}$  times, starting from
    model  $\mathbf{0}$ , the total computational cost is  $\kappa_{LS} = 2 + 2K_{CGit} + (K_{CG} - 1)(3 + 2K_{CGit})$ 
     $= K_{CG}(3 + 2K_{CGit}) - 1$ .
24:   end if
25: end procedure ▷ In an alternative formulation, take  $K_{it} = K_{CG}K_{CGit}$ , and express
     $\kappa_{LS}$  in terms of  $K_{it}$  as  $\kappa_{LS} = K_{it} \frac{3+2K_{CGit}}{K_{CGit}} - 1$ .
  
```

In the case of steepest descent, the relative computational cost is given by

$$\bar{\rho} = \frac{4K_{it} - 1}{2S}, \quad (\text{C.5})$$

which follows from the comment regarding line 1.12 and equation C.1.

In the case of IS with encoded supergathers, the computational cost per iteration is equal to that of standard migration, except without the S factor. For J_{it} stackings, the cost is thus

$$\kappa_{IS} = 2J_{it}. \quad (\text{C.6})$$

Equating equation C.2 and C.6 leads to

$$J_{it} = K_{it} \frac{3 + 2K_{CGit}}{2K_{CGit}} - \frac{1}{2} \quad (\text{C.7})$$

$$= \frac{3K_{it} - 1}{2}, \quad \text{when } K_{CGit} = 3. \quad (\text{C.8})$$

Equation C.7 relates J_{it} , the number of iterations of IS, to K_{it} , the number of iterations of LSM, on the condition that the computational costs incurred by these two methods are equal. This allows us to compare two criterion functions for LSM and for IS, respectively, on the basis of the same computational cost. Let $f_{LSM}(K_{it})$ and $f_{IS}(J_{it})$ be criterion functions of iteration step K_{it} and J_{it} , for LSM and for IS, respectively. Using equation C.7, I write $f_{IS}(J_{it}) = f_{IS}(J_{it}(K_{it}))$. Plotted on the abscissa of K_{it} , the two curves of $f_{LSM}(K_{it})$ and $f_{IS}(J_{it}(K_{it}))$ are thus put on equal footing of computational cost. Examples of such plots are provided in Figure 2.3(b).

D Migration versus Successive Steepest Descent

Given a set of sampled Hessians (or equivalently sampled modeling operators) and the associated data generated according to an underlining model \mathbf{m} , I investigate two strategies, with the knowledge of these sampled Hessians and observed data, to get an estimate of \mathbf{m} : 1) migration and 2) successive steepest descent (SSD). The question I intend to address is, which strategy is better in terms of producing a more accurate estimate?

Let $\mathbf{L}_1, \dots, \mathbf{L}_n$ and $\mathbf{d}_1, \dots, \mathbf{d}_n$ respectively be the n copies of sampled modeling operators and the data generated according to

$$\mathbf{d}_i = \mathbf{L}_i \mathbf{m}, \quad \forall i = 1, \dots, n. \quad (\text{D.1})$$

The migration strategy forms the estimate as

$$\mathbf{m}_{mig} = \sum_{i=1}^n \mathbf{L}_i^\dagger \mathbf{d}_i. \quad (\text{D.2})$$

On the other hand, the SSD strategy consists of the following steps:

1. Start from $\mathbf{m}_{SSD}^{(0)} = \mathbf{0}$.
2. At the k^{th} step, presented with \mathbf{L}_k and \mathbf{d}_k , update the trial model once, accord-

ing to the steepest descent formula, prescribed as :

$$\mathbf{m}_{SSD}^{(k+1)} = \mathbf{m}_{SSD}^{(k)} + \alpha \mathbf{g}, \quad (\text{D.3})$$

where

$$\mathbf{g} = \mathbf{L}_k^\dagger (\mathbf{d}_k - \mathbf{m}_{SSD}^{(k)}), \quad (\text{D.4})$$

$$\alpha = \frac{\|\mathbf{g}\|^2}{\|\mathbf{L}_k \mathbf{g}\|^2}. \quad (\text{D.5})$$

3. Finish at $k = n$.

To fairly evaluate the model error, I introduce

$$\theta^\epsilon(\mathbf{x}) \stackrel{\text{def}}{=} \angle(\mathbf{x}, \mathbf{m}) = \arccos\left(\frac{\mathbf{x}^T \cdot \mathbf{m}}{\|\mathbf{x}\| \|\mathbf{m}\|}\right), \quad (\text{D.6})$$

a criterion that ignores the magnitudes of the vectors in question.

Unable to establish a mathematical bound on $\theta^\epsilon(\mathbf{m}_{mig})$ in comparison to $\theta^\epsilon(\mathbf{m}_{SSD})$, I resort to a Monte Carlo method instead. In this study, $\mathbf{d} \in \mathbb{R}^{10}$ and $\mathbf{m} \in \mathbb{R}^{25}$. The sizes are deliberately chosen to make each individual inverse problem, i.e., to invert equation D.1, which is underdetermined. In addition, the performance of the two strategies in the presence of noise is probed. Specifically, a line of code

$$\mathbf{d}_i \leftarrow \mathbf{d}_i + \mathbf{r} \quad (\text{D.7})$$

follows immediately after equation D.1, and all succeeding operations are based on the noisy \mathbf{d}_i 's. Here, \mathbf{r} is a vector of white Gaussian noise, with its power adjusted to meet a given choice of SNR.

The results are summarized in Figure D.1. Here, the \mathbf{L}_i 's and \mathbf{m} are generated using a Gaussian distribution. I have varied the types of distribution, from Gaussian to a sparse distribution such as binomial, and qualitatively the same trends

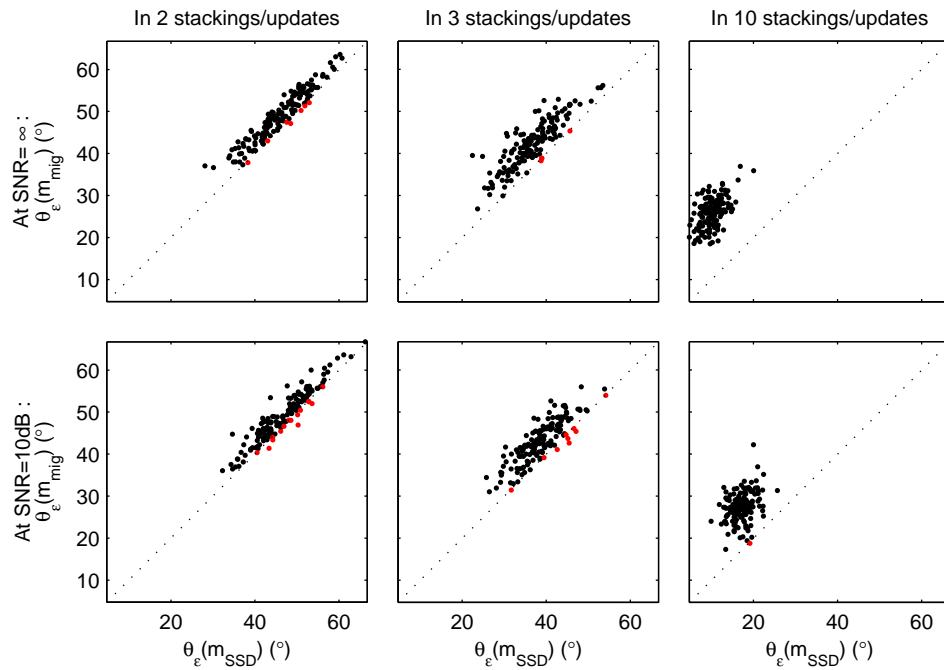


Figure D.1: Scatter plot of the model errors in terms of θ^ϵ , for migration and for SSD, when (upper row) data is not contaminated by noise and (lower row) SNR=10dB, with the set size n varying from (left column) 2, to (middle column) 3, and to (right column) 10. In all the panels, data points corresponding to $\theta^\epsilon(\mathbf{m}_{mig}) \geq \theta^\epsilon(\mathbf{m}_{SSD})$ are plotted in black, while points corresponding to $\theta^\epsilon(\mathbf{m}_{mig}) < \theta^\epsilon(\mathbf{m}_{SSD})$ are plotted in red.

are observed: SSD produces smaller model error than migration does, except for a few rare exceptions. As the size of the random set grows (i.e., more stackings in migration and more updates in SSD), the advantage of SSD becomes even more apparent. As plotted in the upper right panel, $\theta^\epsilon(\mathbf{m}_{mig})/\theta^\epsilon(\mathbf{m}_{SSD}) \approx 24/6$. When \mathbf{d}_i 's are corrupted by noise, however, the performance of SSD deteriorates more than that of migration. But at this level of noise in most cases SSD still outperforms migration.

These observations are intuitively understandable. Migration can be thought of as one step of steepest descent starting from $\mathbf{0}$. Over a set of n samples, migration amounts to averaging n attempts of steepest descent, each starting from $\mathbf{0}$, whereas in SSD, the trial model keeps improving. So it's very probable that the latter outperforms the former. It could happen that the average of these 'first attempts' comes very close to the true model. This explains why exceptions exist. In the presence of white noise, averaging with equal weight over random instances, as migration does, is the most effective means to reduce noise. In SSD, however, the earlier updates influence the iteration trajectory more than the later updates do. In effect, the end result senses a weighted average of the noise contained in each update, with larger weight assigned to early samples and smaller weight to later samples, resulting in less noise reduction than what migration is capable of.

E Resolution Properties of an Ellipse

I first analyze the resolution of the ellipse depicted in Figure 4.1b, and then relate the parameters of this ellipse to those defining a wavepath.

The spatial resolution limit near the point c in Figure 4.1b is related to the reciprocal of the segment length \overline{DE} . This length can be determined by noting that the end points D and E satisfy both the equations of the ellipse and the line, written as

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad (\text{E.1})$$

$$y = \tan \theta (x - c), \quad (\text{E.2})$$

where a and b are the major and minor radii of the ellipse, respectively, and θ is the angle DE makes with the axis of the ellipse. Equations E.1 and E.2 can be reduced to a quadratic equation of one variable y , yielding two roots y_D and y_E . The distance \overline{DE} is then obtained as

$$\begin{aligned} \overline{DE} &= \frac{|y_D - y_E|}{|\sin \theta|}, \\ &= \frac{2ab \sqrt{b^2 \cos^2 \theta + (a^2 - c^2) \sin^2 \theta}}{b^2 \cos^2 \theta + a^2 \sin^2 \theta}. \end{aligned} \quad (\text{E.3})$$

Two special cases of c and θ are immediately verified. We have $\overline{DE} = 2a$ when $\theta = 0$, and $\overline{DE} = 2b$ when $c = 0$ and $\theta = \pi/2$, i.e., the lengths of the major and the minor axes of the ellipse, respectively.

Next, the parameters a and b of the ellipse are related to those defining the first

Fresnel zone, as depicted in Figure 4.1a. Let s and g be the two foci of the ellipse, the distance between s and g be L , and E be an arbitrary point on the ellipse. The first Fresnel zone is delimited by points E on the ellipse that satisfy

$$\overline{sE} + \overline{Eg} = L + \frac{\lambda}{2}. \quad (\text{E.4})$$

Also,

$$\overline{sE} + \overline{Eg} = 2a \quad (\text{E.5})$$

is a property of the ellipse. Another property of the ellipse relates the focal distance to the major and minor radii by

$$L = 2\sqrt{a^2 - b^2}. \quad (\text{E.6})$$

Equations E.4—E.6 give us

$$a = \frac{L}{2} + \frac{\lambda}{4}, \quad (\text{E.7})$$

$$\text{and } b = \sqrt{\frac{\lambda L}{4} + \frac{\lambda^2}{16}}. \quad (\text{E.8})$$

In the limit of $L \gg \lambda$, $b \rightarrow \frac{1}{2}\sqrt{\lambda L}$.

From equations E.7, E.8, and E.3, we see that the resolution can be written as $\overline{DE}(L, \lambda, c, \theta)$, a function of wavepath parameters L, λ , and intersection parameters c and θ .

F Resolution Limits for Imaging Diving Wave Residuals

The resolution limits for imaging diving wave residuals are rigorously derived by multiplying the migration kernel in Figure 4.4a by the expression for the diving-wave arrival

$$\begin{aligned} \delta d(\mathbf{g}|\mathbf{s}) &= \omega^4 \int_{\Omega} G(\mathbf{g}|\mathbf{y})^{\text{dir}} G(\mathbf{y}|\mathbf{s})^{\text{dir}} \delta m(\mathbf{y}) d\mathbf{y}^2, \\ &\approx \omega^4 \int_{\Omega_{gs}} G(\mathbf{g}|\mathbf{y})^{\text{dir}} G(\mathbf{y}|\mathbf{s})^{\text{dir}} \delta m(\mathbf{y}) d\mathbf{y}^2, \end{aligned} \quad (\text{F.1})$$

where the integral over the model-space region Ω is approximated by the one over the region Ω_{gs} . Here, Ω_{gs} coincides with the yellow first Fresnel zone of the diving wave in Figure 4.4a for the source-geophone pair denoted by s and g . This approximation recognizes that only model perturbations within the first Fresnel zone of the diving wave will strongly affect the timing and/or amplitude of the diving-wave arrival at \mathbf{g} .

Plugging equation F.1 into equation 4.6 gives

$$\delta m(\mathbf{x})^{\text{mig}} = \omega^4 \int_{D_{r_0}} \int_{\Omega_{gs}} [G(\mathbf{g}|\mathbf{x})^{\text{dir}} G(\mathbf{x}|\mathbf{s})^{\text{dir}}]^* G(\mathbf{g}|\mathbf{y})^{\text{dir}} G(\mathbf{y}|\mathbf{s})^{\text{dir}} \delta m(\mathbf{y}) d\mathbf{y}^2 dx_g dx_s. \quad (\text{F.2})$$

We now assume a localized subwavelength perturbation $\delta m(\mathbf{y})$ centered at $\mathbf{r}_0 = (x_0, z_0)$ that is non-zero only within a fraction of a wavelength from \mathbf{r}_0 . In this case, the range of source-geophone pairs in D is restricted to the set D_{r_0} of source-

geophone pairs that allow for first Fresnel diving wavepaths to visit the localized perturbation centered at \mathbf{r}_o . These source–geophone pairs are the only ones whose transmitted diving waves¹⁹ will be significantly influenced by the model perturbations centered at \mathbf{r}_o . For example, if the image point is at \mathbf{y} and the geophone is at \mathbf{C} in Figure 4.6b, then D_{r_o} is limited to the sources between A and B . If the wavepaths are those for a specular reflection, then the range of source locations in Ω_{gs} is between A and B in Figure 4.6a.

For a smooth background velocity we assume the following asymptotic Green's function for the migration and data kernels

$$G(\mathbf{x}|\mathbf{y})^{\text{dir}} = A_{xy}e^{-i\omega\tau_{xy}}, \quad (\text{F.3})$$

so that equation F.2 becomes

$$\delta m(\mathbf{x})^{\text{mig}} = \omega^4 \int_{D_{r_o}} \int_{\Omega_{gs}} A_{sx}A_{gx}A_{sy}A_{gy} e^{i\omega(\tau_{gx}+\tau_{sx}-\tau_{gy}-\tau_{sy})} \delta m(\mathbf{y})d\mathbf{y}^2 dx_g dx_s. \quad (\text{F.4})$$

Here, τ_{xy} is the travelttime for the transmitted wave to propagate from \mathbf{y} to \mathbf{x} , and A_{xy} is its attendant geometrical spreading term that satisfies the transport equation.

Assuming that the subwavelength scatterer represented by $\delta m(\mathbf{y})$ is located within a fraction of a wavelength from the trial image point at \mathbf{x} , then τ_{sy} , τ_{gy} , τ_{sx} ,

¹⁹We exclude the case where the scatterer–diving wave interaction produces significant diffractions, so that all source–geophone pairs see significant diffraction energy, not just changes in the diving–wave arrival. This would be the case where the scatterer only has a velocity contrast but no impedance contrast.

and τ_{gx} , can be expanded about its center point \mathbf{r}_o to give

$$\begin{aligned}\tau_{sy} &\approx \tau_{sr_o} + \nabla\tau_{sr_o} \cdot [\mathbf{y} - \mathbf{r}_o], \\ \tau_{gy} &\approx \tau_{gr_o} + \nabla\tau_{gr_o} \cdot [\mathbf{y} - \mathbf{r}_o], \\ \tau_{sx} &\approx \tau_{sr_o} + \nabla\tau_{sr_o} \cdot [\mathbf{x} - \mathbf{r}_o], \\ \tau_{gx} &\approx \tau_{gr_o} + \nabla\tau_{gr_o} \cdot [\mathbf{x} - \mathbf{r}_o].\end{aligned}\tag{F.5}$$

Inserting these approximations into equation F.4 gives

$$\delta m(\mathbf{x})^{\text{mig}} \approx \omega^4 \int_{D_{r_o}} \int_{\Omega_{gs}} A_{sx} A_{gx} A_{sy} A_{gy} e^{-i\omega(\nabla\tau_{gr_o} + \nabla\tau_{sr_o}) \cdot [\mathbf{y} - \mathbf{x}]} \delta m(\mathbf{y}) dy^2 dx_g dx_s.$$

Under the far-field approximation, the geometric spreading terms can be taken outside the integral to give

$$\delta m(\mathbf{x})^{\text{mig}} = \omega^4 \int_{D_{r_o}} A_{sx} A_{gx} A_{sr_o} A_{gr_o} \int_{\Omega_{gs}} e^{-i\omega(\nabla\tau_{gr_o} + \nabla\tau_{sr_o}) \cdot [\mathbf{y} - \mathbf{x}]} \delta m(\mathbf{y}) dy^2 dx_g dx_s.\tag{F.6}$$

Here, the gradient of the travelttime field $\nabla\tau_{sr_o}$ is parallel to the direct wave's incident angle at \mathbf{r}_o , so, according to the dispersion equation, $\omega\nabla\tau_{sr_o} = \mathbf{k}_{sr_o}$ can be identified as the source-to-scatterer point wavenumber vector \mathbf{k}_{sr_o} ; similarly, the geophone-to-scatterer wavenumber is denoted as $\omega\nabla\tau_{gr_o} = \mathbf{k}_{gr_o}$. This means that, by definition of the Fourier transform with a restricted domain of integration $\delta M(\mathbf{k}) = \int_{\Omega_{gs}} e^{-i\mathbf{k} \cdot \mathbf{y}} \delta m(\mathbf{y}) dy^2$, equation F.6 becomes

$$\delta m(\mathbf{x})^{\text{mig}} \approx \omega^4 A_{s_o g_o r_o}^4 \int_{D_{r_o}} e^{i(\mathbf{k}_{gr_o} + \mathbf{k}_{sr_o}) \cdot \mathbf{x}} \delta M(\mathbf{k}_{gr_o} + \mathbf{k}_{sr_o}) dx_g dx_s,\tag{F.7}$$

where $A_{s_o g_o r_o}$ approximates the geometrical spreading for the scatterer at \mathbf{r}_o with the range of allowable source-geophone pairs centered around the pairs denoted by $s_o g_o$, the Fourier spectrum of the model is given by $\delta M(\mathbf{k})$, and the model

wavenumber components $\mathbf{k} = (k_x, k_z)$ are

$$\begin{aligned} k_x &= k_{sx_0} + k_{gx_0} = \omega s(\mathbf{r}_o)(\sin \beta_{sr_0} + \sin \beta_{gr_0}), \\ k_z &= k_{sz_0} + k_{gz_0} = \omega s(\mathbf{r}_o)(\cos \beta_{sr_0} + \cos \beta_{gr_0}), \end{aligned} \quad (\text{F.8})$$

where β_{sr_0} and β_{gr_0} denote the incidence angles of the source and geophone rays, respectively, at the scatterer's location $\mathbf{y} = (x_o, z_o)$. As shown in the appendix, these incidence angles are implicit functions of the source $(x_s, 0)$, geophone $(x_g, 0)$, and scatterer $\mathbf{r}_o = (x_o, z_o)$ coordinates.

The determinant of the Jacobian in equation G.2 (see Appendix G) can be used to map the $dx_g dx_s$ integration in equation F.7 to a $dk_x dk_z$ integration:

$$\delta m(\mathbf{x})^{\text{mig}} = A_{s_0 g_0 r_0} \int_{\mathcal{D}_{r_0}} e^{-i\mathbf{k} \cdot \mathbf{x}} \delta M(\mathbf{k}) J^{-1} dk_x dk_z, \quad (\text{F.9})$$

where \mathcal{D}_{r_0} is the set of wavenumbers that equation F.8 maps from the source–geophone pairs in D_{r_0} for the scatterer at \mathbf{r}_o , and J is the determinant of the Jacobian matrix in equation G.2.

G Determinant of a Jacobian Matrix

The transformation between the data coordinates $(x_g, 0)$, $(x_s, 0)$ and $(k_x, k_z) = (k_{sx_0} + k_{gx_0}, k_{sz_0} + k_{gz_0})$ is given by

$$\begin{pmatrix} dk_x \\ dk_z \end{pmatrix} = \begin{bmatrix} \frac{\partial k_x}{\partial x_g} & \frac{\partial k_x}{\partial x_s} \\ \frac{\partial k_z}{\partial x_g} & \frac{\partial k_z}{\partial x_s} \end{bmatrix} \begin{pmatrix} dx_g \\ dx_s \end{pmatrix}, \quad (\text{G.1})$$

where the 2x2 matrix is the Jacobian matrix. The scaled determinant J of the Jacobian matrix is given by

$$J = \omega^4 \left| \frac{\partial k_x}{\partial x_g} \frac{\partial k_z}{\partial x_s} - \frac{\partial k_x}{\partial x_s} \frac{\partial k_z}{\partial x_g} \right|, \quad (\text{G.2})$$

so that $dk_x dk_z = J dx_g dx_s$. In the case of a homogeneous medium with velocity c and a scatterer at $\mathbf{r}_0 = (x_0, z_0)$, the model wavenumbers are

$$\begin{aligned} k_x &= \frac{\omega(x_0 - x_g)}{c \sqrt{(x_0 - x_g)^2 + z_0^2}} + \frac{\omega(x_0 - x_s)}{c \sqrt{(x_0 - x_s)^2 + z_0^2}}, \\ k_z &= \frac{\omega z_0}{c \sqrt{(x_0 - x_g)^2 + z_0^2}} + \frac{\omega z_0}{c \sqrt{(x_0 - x_s)^2 + z_0^2}}, \end{aligned} \quad (\text{G.3})$$

so that the partial derivatives of the wavenumbers can be easily determined. For a heterogeneous medium, the derivatives can be approximated by finite-difference approximations to the first-order derivatives and the wavenumbers can be computed by a ray tracing method. Under the farfield approximation $z \gg L$, where L

is the aperture width of the source–geophone array, so equation G.3 becomes

$$\begin{aligned}k_x &\approx \frac{\omega(x_o - x_g)}{cz_o} + \frac{\omega(x_o - x_s)}{cz_o}, \\k_z &\approx \frac{2\omega}{c},\end{aligned}\tag{G.4}$$

where the horizontal wavenumbers are now linear functions of the data variables x_g and x_s . This means that equation F.7 represents the inverse Fourier transform of the model spectrum.

H Publications Based on This Dissertation

- Huang, Y. and G.T. Schuster, "Multisource least-squares migration of marine streamer and land data with frequency-division encoding", *Geophysical Prospecting* **60**, 663-680, 2012.
- Huang, Y. and G.T. Schuster, "Multisource Full Waveform Inversion of Marine Streamer Data with Frequency Selection", *EAGE Annual meeting*, London, June, 2013.
- Huang, Y., G.T. Schuster, W. Dai and G. Zhan, "Resolution Limits for Wave Equation Imaging", *Submitted to Geophysical Journal International*, 2013.

REFERENCES

- Amundsen, L. (1991). Comparison of the least-squares criterion and the Cauchy criterion in frequency-wavenumber inversion. *Geophysics*, 56(12):2027–2035.
- Aster, R., Borchers, B., and Thurber, C. (2005). *Parameter Estimation and Inverse Problems*. Elsevier.
- Barton, G. (1989). *Elements of Green's functions and propagation: potentials, diffusion, and waves*. Oxford University Press.
- Bates, R. and Gregory, D. (2007). *Voice & data communications handbook*. McGraw-Hill Osborne Media.
- Baysal, E., Kosloff, D. D., and Sherwood, J. W. C. (1983). Reverse time migration. *Geophysics*, 48(11):1514–1524.
- Ben-Hadj-Ali, H., Operto, S., and Virieux, J. (2009). Three-dimensional frequency-domain full waveform inversion with phase encoding. In *SEG Technical Program Expanded Abstracts*, volume 28, pages 2288–2292.
- Ben-Hadj-Ali, H., Operto, S., and Virieux, J. (2011). An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. *Geophysics*, 76(4):R109–R124.
- Berkhout, A. J. (1984). *Seismic resolution: A quantitative analysis of resolving power of acoustical echo techniques*. Geophysical Press.
- Beylkin, G. (1985). Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized radon transform. *Journal of Mathematical Physics*, 26:99.
- Boonyasiriwat, C. and Schuster, G. (2010). 3D multisource full-waveform inversion using dynamic quasi-monte carlo phase encoding. In *SEG Technical Program Expanded Abstracts*, volume 29, pages 1044–1049.

- Bottou, L. and Bousquet, O. (2011). The tradeoffs of large-scale learning. *Optimization for Machine Learning*, page 351.
- Buddensiek, M.-L., Sheng, J., Crosby, T., Schuster, G., Bruhn, R., and He, R. (2008). Colluvial wedge imaging using traveltimes and waveform tomography along the wasatch fault near mapleton, utah. *Geophysical Journal International*, 172(2):686–697.
- Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(05):1457–1473.
- Cerveny, V. and Soares, J. E. P. (1992). Fresnel volume ray tracing. *Geophysics*, 57(07):902–915.
- Chen, J. and Schuster, G. T. (1999). Resolution limits of migrated images. *Geophysics*, 64(4):1046–1053.
- Claerbout, J. (1992). *Earth soundings analysis: Processing versus inversion*. Blackwell Scientific Publications, Cambridge, Massachusetts.
- Dahlen, F. (2004). Resolution limit of traveltimes tomography. *Geophysical Journal International*, 157(1):315–331.
- Dai, W., Fowler, P., and Schuster, G. T. (2012). Multi-source least-squares reverse time migration. *Geophysical Prospecting*, 60(4):681–695.
- Dai, W. and Schuster, G. T. (2009). Least-squares migration of simultaneous sources data with a deblurring filter. In *SEG Technical Program Expanded Abstracts*, volume 28, pages 2990–2994.
- Dai, W. and Schuster, G. T. (in review 2013). Reverse time migration of prism waves for salt flank delineation. *Geophysical Prospecting*.
- de Fornel, F. (2001). *Evanescent waves: from Newtonian optics to atomic optics*. Springer.
- De Hoop, M. and van Der Hilst, R. D. (2005). On sensitivity kernels for wave-equation transmission tomography. *Geophysical Journal International*, 160(2):621–633.
- Debski, W. (2010). Probabilistic inverse theory. *Advances in geophysics*, 52:1–102.
- Devaney, A. (1984). Geophysical diffraction tomography. *Geoscience and Remote Sensing, IEEE Transactions on*, GE-22(1):3–13.

- Duquet, B., Lailly, P., and Ehinger, A. (2001). 3D plane wave migration of streamer data. In *SEG Technical Program Expanded Abstracts*, volume 20, pages 1033–1036.
- Duquet, B., Marfurt, K. J., and Dellinger, J. (2000). Kirchhoff modeling, inversion for reflectivity, and subsurface illumination. *Geophysics*, 65(4):1195–1209.
- Fichtner, A. (2011). *Full seismic waveform modelling and inversion*. Springer.
- Fichtner, A., Kennett, B. L., Igel, H., and Bunge, H.-P. (2009). Full seismic waveform tomography for upper-mantle structure in the australasian region using adjoint methods. *Geophysical Journal International*, 179(3):1703–1725.
- Fichtner, A. and Trampert, J. (2011a). Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, 185(2):775–798.
- Fichtner, A. and Trampert, J. (2011b). Resolution analysis in full waveform inversion. *Geophysical Journal International*, 187(3):1604–1624.
- Fink, M. (2008). Time-reversal waves and super resolution. In *Journal of Physics: Conference Series*, volume 124, page 012004. IOP Publishing.
- Fleming, M. (2008). *Far-field super resolution*. PhD thesis, Imperial College London.
- Gao, F., Atle, A., and Williamson, P. (2010). Full waveform inversion using deterministic source encoding. In *SEG Technical Program Expanded Abstracts*, volume 29.
- Godwin, J. and Sava, P. (2010). Simultaneous source imaging by amplitude encoding. Technical Report CWP-645, Center for Wave Phenomena, Colorado School of Mines.
- Harlan, W. S. (1990). Tomographic estimation of shear velocities from shallow cross-well seismic data. In *SEG Technical Program Expanded Abstracts*, volume 9, pages 86–89.
- Huang, Y. and Schuster, G. (2012). Multisource least-squares migration of marine streamer and land data with frequency-division encoding. *Geophysical Prospecting*, 60:663–680.
- Jannane, M., Beydoun, W., Crase, E., Cao, D., Koren, Z., Landa, E., Mendes, M., Pica, A., Noble, M., Roeth, G., et al. (1989). Wavelengths of earth structures that can be resolved from seismic reflection data. *Geophysics*, 54(7):906–910.

- Jing, X., Finn, C. J., Dickens, T. A., and Willen, D. E. (2000). Encoding multiple shot gathers in prestack migration. In *SEG Technical Program Expanded Abstracts*, volume 19, pages 786–789.
- Johnson, S. and Frigo, M. (2006). A modified split-radix FFT with fewer arithmetic operations. *Signal Processing, IEEE Transactions on*, 55(1):111–119.
- Kiwiel, K. C. (2001). Convergence and efficiency of subgradient methods for quasiconvex minimization. *Mathematical Programming*, 90(1):1–25.
- Kravtsov, Y. A. and Orlov, Y. I. (1990). *Geometrical optics of inhomogeneous media*. Springer-Verlag.
- Krebs, J. R., Anderson, J. E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., and Lacasse, M.-D. (2009). Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, 74(6):WCC177–WCC188.
- Kuehl, H. and Sacchi, M. D. (1999). Least-squares split-step migration using the hartley transform. In *SEG Technical Program Expanded Abstracts*, volume 18, pages 1548–1551.
- Lailly, P. (1984). Migration methods: partial but efficient solutions to the seismic inverse problem. *Inverse problems of acoustic and elastic waves*, 51:1387–1403.
- Liu, F., Zhang, G., Morton, S. A., and Leveille, J. P. (2011). An effective imaging condition for reverse-time migration using wavefield decomposition. *Geophysics*, 76(1):S29–S39.
- Luo, Y. (1991). Calculation of wavepaths for band-limited seismic waves. In *SEG Technical Program Expanded Abstracts*, volume 10, pages 1509–1512.
- Luo, Y. and Schuster, G. T. (1991). Wave-equation travelttime inversion. *Geophysics*, 56(5):645–653.
- Marquering, H., Dahlen, F., and Nolet, G. (1999). Three-dimensional sensitivity kernels for finite-frequency traveltimes: the banana-doughnut paradox. *Geophysical Journal International*, 137(3):805–815.
- McMechan, G. (1983). Migration by extrapolation of time-dependent boundary values. *Geophysical Prospecting*, 31(3):413–420.
- Menke, W. (1989). *Geophysical data analysis: discrete inverse theory*. Elsevier.

- Mora, P. (1988). Elastic wave-field inversion of reflection and transmission data. *Geophysics*, 53(06):750–759.
- Mora, P. (1989). Inversion= migration+ tomography. *Geophysics*, 54(12):1575–1586.
- Morton, S. A. and Ober, C. C. (1998). Faster shot-record migrations using phase encoding. In *SEG Technical Program Expanded Abstracts*, volume 18, pages 1131–1134.
- Nemeth, T., Wu, C., and Schuster, G. (1999). Least-squares migration of incomplete reflection data. *Geophysics*, 64(1):208–221.
- Nihei, K. and Li, X. (2007). Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (td-psd). *Geophysical Journal International*, 169(3):1069–1078.
- Pica, A., Diet, J., and Tarantola, A. (1990). Nonlinear inversion of seismic reflection data in a laterally invariant medium. *Geophysics*, 55(03):284–292.
- Pratt, R. G. and Goult, N. R. (1991). Combining wave-equation imaging with travelttime tomography to form high-resolution images from crosshole data. *Geophysics*, 56(02):208–224.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (2007). *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition.
- Romero, L. A., Ghiglia, D. C., Ober, C. C., and Morton, S. A. (2000). Phase encoding of shot records in prestack migration. *Geophysics*, 65(2):426–436.
- Sacchi, M. (1997). Reweighting strategies in seismic deconvolution. *Geophysical Journal International*, 129(3):651–656.
- Safar, M. H. (1985). On the lateral resolution achieved by Kirchhoff migration. *Geophysics*, 50(07):1091–1099.
- Scales, J., Gersztenkorn, A., and Treitel, S. (1988). Fast l_p solution of large, sparse, linear systems: Application to seismic travel time tomography. *Journal of Computational Physics*, 75(2):314–333.

- Schraudolph, N. N. and Graepel, T. (2002). Conjugate directions for stochastic gradient descent. In Dorronsoro, J. R., editor, *ICANN*, volume 2415, pages 1351–1356, Madrid, Spain. Springer Verlag.
- Schuster, G. T. (1996). Resolution limits for crosswell migration and travelttime tomography. *Geophysical Journal International*, 127(2):427–440.
- Schuster, G. T., Hanafy, S., and Huang, Y. (2012). Theory and feasibility tests for a seismic scanning tunnelling microscope. *Geophysical Journal International*, 190(3):1593–1606.
- Schuster, G. T. and Hu, J. (2000). Green’s function for migration: Continuous recording geometry. *Geophysics*, 65(1):167–175.
- Schuster, G. T., Wang, X., Huang, Y., Dai, W., and Boonyasiriwat, C. (2011). Theory of multisource crosstalk reduction by phase-encoded statics. *Geophysical Journal International*, 184(3):1289–1303.
- Sheley, D. and Schuster, G. T. (2003). Reduced-time migration of transmitted Ps waves. *Geophysics*, 68(5):1695–1707.
- Sheng, J. and Schuster, G. T. (2003). Finite-frequency resolution limits of wave path travelttime tomography for smoothly varying velocity models. *Geophysical Journal International*, 152(3):669–676.
- Shin, C. and Cha, Y. H. (2008). Waveform inversion in the laplace domain. *Geophysical Journal International*, 173(3):922–931.
- Sirgue, L., Etgen, J., and Albertin, U. (2008). 3d frequency domain waveform inversion using time domain finite difference methods. In *Proceedings of the 70th EAGE Conference and Exhibition, Rome, Italy*, volume 22.
- Spall, J. (2003). *Introduction to stochastic search and optimization: estimation, simulation, and control*. John Wiley and Sons.
- Stoffa, P. L., Fokkema, J. T., de Luna Freire, R. M., and Kessinger, W. P. (1990). Split-step Fourier migration. *Geophysics*, 55(04):410–421.
- Stolt, R. H. and Benson, A. K. (1986). *Seismic Migration: Theory and Practice*. Pergamon Press.

- Tang, Y. (2009). Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian. *Geophysics*, 74(6):WCA95–WCA107.
- Tang, Y. and Biondi, B. (2009). Least-squares migration/inversion of blended data. In *SEG Technical Program Expanded Abstracts*, volume 28, pages 2859–2863.
- Tape, C., Liu, Q., Maggi, A., and Tromp, J. (2009). Adjoint tomography of the southern california crust. *Science*, 325:988–992.
- Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(08):1259–1266.
- Tarantola, A. (2005). *Inverse Problem Theory and methods for model parameter estimation*. SIAM.
- Tong, J., Dahlen, F., Nolet, G., and Marquering, H. (1998). Diffraction effects upon finite-frequency travel times: A simple 2-d example. *Geophysical research letters*, 25(11):1983–1986.
- Van Der Hilst, R. D. and Maarten, V. (2005). Banana-doughnut kernels and mantle tomography. *Geophysical Journal International*, 163(3):956–961.
- Vermeer, G. J. O. (1997). Factors affecting spatial resolution. In *SEG Technical Program Expanded Abstracts*, volume 16, pages 27–30.
- Vigh, D. and Starr, E. W. (2008). 3d prestack plane-wave, full-waveform inversion. *Geophysics*, 73(5):VE135–VE144.
- Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1.
- Wang, J. and Sacchi, M. (2007). High-resolution wave-equation amplitude-variation-with-ray-parameter (avp) imaging with sparseness constraints. *Geophysics*, 72(1):S11–S18.
- Whitmore, N. (1983). Iterative depth migration by backward time propagation. In *SEG Technical Program Expanded Abstracts*, volume 2, pages 382–385.
- Whitmore, N. D. and Garing, J. D. (1993). Interval velocity estimation using iterative prestack depth migration in the constant angle domain. *The Leading Edge*, 12(07):757–762.

- Williamson, P. R. (1991). A guide to the limits of resolution imposed by scattering in ray tomography. *Geophysics*, 56(02):202–207.
- Woodward, M. (1989). *Wave equation tomography*. PhD thesis, Stanford University.
- Woodward, M. (1992). Wave-equation tomography. *Geophysics*, 57(1):15–26.
- Wu, R. S. and Toksoz, M. N. (1987). Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52(01):11–25.
- Zhan, G., Dai, W., Luo, Y., Zhou, M., and Schuster, G. (accepted 2013). Generalized diffraction-stack migration and filtering of coherent noise. *Geophysical Prospecting*.
- Zhang, S. and Schuster, G. (2013). Generalized differential semblance optimization. In *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*.
- Zhang, S., Schuster, G., and Luo, Y. (2012). Angle-domain migration velocity analysis using wave equation reflection traveltime inversion. In *SEG Technical Program Expanded Abstracts*, volume 31, pages 1–6.
- Zhang, Y., Sun, J., Notfors, C., Gray, S., Chernis, L., and Young, J. (2003). Delayed-shot 3D prestack depth migration. In *SEG Technical Program Expanded Abstracts*, volume 22, pages 1027–1030.
- Zhou, C., Cai, W., Luo, Y., Schuster, G. T., and Hassanzadeh, S. (1995). Acoustic wave-equation traveltime and waveform inversion of crosshole seismic data. *Geophysics*, 60(03):765–773.