

Multicomponent least-squares Kirchhoff depth migration

by

Landon Safron

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Geophysics

Department of Physics

University of Alberta

©Landon Safron, 2018

Abstract

Conventional seismic migration operators produce an image that suffers from low resolution, sampling artifacts, and poorly balanced amplitudes. An improved image can be obtained by casting migration as a least-squares optimization problem in which the goal is to minimize the difference between the predicted synthetic data and the true observed data. This thesis aims to expand the technology surrounding least-squares Kirchhoff depth migration. First, an anti-aliased forward/adjoint Kirchhoff operator is derived for performing seismic modeling and migration. The anti-aliased Kirchhoff operator utilizes a triangular filter that can be efficiently customized on-the-fly. Unlike alternative Kirchhoff operators, the anti-aliased Kirchhoff operator proposed in this thesis is well-suited for least-squares migration because only one copy of the data is required to be produced and stored. Second, the importance of regularizing the least-squares cost function is discussed, and a preconditioning strategy for promoting lateral continuity within common image gathers is reviewed. Since the preconditioner promotes stable features in the model, the eigenvalue distribution of the operator is improved, thereby increasing the rate-of-convergence of conjugate gradients. Lastly, a new cost function is proposed for performing converted wave least-squares depth migration in the presence of S-velocity errors. Due to the low signal-to-noise ratio and strong aliasing of converted waves, S-velocity models are typically less accurate than P-velocity models; consequently, PP seismic images are typically more accurate than PS seismic images. The proposed converted wave migration cost function exploits this observation by registering the PS image volume with the presumably correct PP image volume. The new converted wave cost function is found to be an improvement over the conventional cost function in terms

of the accuracy of the image and the flatness of the common image gathers. Moreover, the registration strategy allows the true depths of the converted wave reflectors to be recovered, despite the presence of errors in the S-velocity model.

Acknowledgements

I would like to thank my supervisor, Dr. Mauricio Sacchi, without whom this thesis would not be possible. His insight into mathematics and inverse theory has forever changed the way I approach problems. I would also like to thank my colleagues in the Signal Analysis and Imaging Group for all of their valuable suggestions and insightful conversations over the past two years. Additionally, I would like thank my committee members for their feedback during the creation of this thesis. Finally, I would like to thank my family for their endless support.

Contents

1	Introduction	1
1.1	Background	1
1.2	Seismic migration	3
1.3	Image registration	5
1.4	Scope of this thesis	7
1.4.1	Main contributions	7
1.4.2	Organization	8
2	The Kirchhoff operator	9
2.1	Introduction	9
2.2	The wave equation	10
2.2.1	The momentum equation	10
2.2.2	The elastic wave equation	11
2.2.3	The acoustic approximation	12
2.3	Linear scattering	13
2.4	The Kirchhoff approximation	16
2.5	Kirchhoff migration	19
2.6	Operator aliasing	23
2.7	Examples	29
2.7.1	Example: The forward/adjoint Kirchhoff operator	30
2.7.2	Example: Kirchhoff operator aliasing	33
2.8	Summary	35

3	The wavefront construction method	36
3.1	Introduction	36
3.2	Ray tracing	37
3.2.1	Background velocity model	37
3.2.2	The ray tracing equations	39
3.3	Wavefront construction	43
3.3.1	Wavefront propagation	43
3.3.2	Parameter estimation	48
3.4	Examples	51
3.4.1	Example: SAIG velocity model	51
3.4.2	Example: Accuracy of the wavefront construction method	55
3.5	Summary	57
4	Least-squares migration	58
4.1	Introduction	58
4.2	Regularized least-squares migration	60
4.3	Preconditioned least-squares migration	62
4.4	Examples	63
4.4.1	Example: Damped vs preconditioned LSM	64
4.4.2	Example: Kirchhoff operator aliasing in LSM	69
4.5	Summary	72
5	Converted wave LSM with S-velocity errors	74
5.1	Introduction	74
5.2	Warped converted wave LSM	75
5.3	Example: LSM with S-velocity errors	78
5.4	Summary	83

6	Conclusions	84
6.1	Summary	84
6.2	Future work	86
	Bibliography	88
	Appendices	
A	The eikonal and transport equations	95
B	The Runge-Kutta method	97
C	Conjugate gradients for least-squares	99
D	Dynamic image warping	101
E	Software	104
E.1	The Julia programming language	104
E.2	Open-source software	104

List of Tables

E.1	Description of the main Julia functions developed during the creation of this thesis.	105
-----	---	-----

List of Figures

2.1	Diagram showing the intersection of a sphere D with a plane S . The subscript “a” denotes features in the upper hemisphere and the subscript “b” denotes features in the lower hemisphere.	15
2.2	Coordinates for Kirchhoff AVA migration. Figure from Feng (2004).	20
2.3	Illustration of the Kirchhoff operator (red curve). The forward Kirchhoff integral sprays each model-domain scatterer over a data-domain diffraction. The adjoint Kirchhoff integral collapses each data-domain diffraction into a model-domain scatterer. In practice, the adjoint operator recovers a blurred version of the true image because it is not orthogonal to the forward operator. 24	
2.4	Kirchhoff migration operator (red curve) cutting through a seismic event. (a) The trace spacing is small, so the Kirchhoff operator intersects several traces. The sum of the samples along the operator is nearly zero, so an aliasing artifact is not produced. (b) The trace spacing is large. The steeper Kirchhoff operator intersects only a single trace, so the sum of the samples along the operator is non-zero. Therefore, the steeper Kirchhoff operator produces an aliasing artifact. The less steep Kirchhoff operator intersects a few traces, so the sum of the samples along the operator is nearly zero. Therefore, the less steep Kirchhoff operator does not produce an aliasing artifact. Figure from Gray (2013).	25
2.5	The amount of operator moveout (grey boxes) is proportional to the local dip of the Kirchhoff operator (red curve). The steepest parts of the Kirchhoff operator experience the greatest amount of moveout and are therefore the most susceptible to aliasing.	26
2.6	Amplitude spectrum for the triangular filter in equation 2.45 with $\ell = 10$ and $\delta t = 4$ milliseconds. Most of the triangular filters energy is contained within the first spectral notch.	28
2.7	(a) Synthetic velocity model (m/s). (b) Reflectivity model for (a). The reflectivity is the same for each offset bin.	31
2.8	Forward modeled shot gather for a source located at $x=2500\text{m}$ for the model in figure 2.7.	32
2.9	Stacked seismic image obtained by migration via the adjoint Kirchhoff operator. 32	

2.10	(a) Synthetic reflectivity model. (b) Zero-offset seismic section produced using the forward Kirchhoff operator \mathbf{L} . (c) Migrated seismic image using the aliased adjoint Kirchhoff operator \mathbf{L}^\dagger based on Algorithm 1. Migrated seismic image using the anti-aliased adjoint Kirchhoff operator \mathbf{L}^\dagger based on Algorithm 2.	34
3.1	Rays with different takeoff angles can follow a different path to reach the same scatterer. The curved ray travels a longer distance at a slower velocity, and thus requires a larger traveltime to reach the scatterer. The red star denotes the source and the grey shape represents a low velocity zone.	42
3.2	A new wavefront (dashed line) is constructed by tracing the existing rays (black circles) forward one time step. Additional rays (white circles) are interpolated when the distance between rays (red line) is greater than the threshold DS_{\max}	44
3.3	The new wavefront crosses itself as it propagates across a low-velocity caustic. Only the rays corresponding to first arrivals are kept when using the first arrival mode. Thus, the part of the wavefield behind the crossing (point number 5 and 6) is removed. An imaginary link (red dashed line) is created to fill in the gap; however, no new rays are interpolated in this gap.	45
3.4	Geometrical relationship between the arc length of three rays and the opening angle.	47
3.5	Traveltimes are interpolated by propagating the wavefront forward within each piece of the mesh.	49
3.6	(a) SAIG velocity model. (b) Smoothed version of (a). Smoothing was applied to the slowness model using a Gaussian blurring kernel with a radius of 100m. The color-bars show the velocity in units of m/s.	52
3.7	The wavefront construction method. (a) Wavefront mesh for a shot located at $x=1500\text{m}$ on the surface. (b) Location of the wavefront at 60ms intervals. The color-bars show the velocity in units of m/s.	53
3.8	(a) Interpolated traveltimes and (b) interpolated amplitudes using the wavefront construction method. The amplitudes are heavily clipped to allow small-scale variations to be seen. The color-bar for (a) shows the traveltime in units of seconds. The colorbar for (b) shows the amplitude (unit-less).	54
3.9	Error of the wavefront construction method for a linear velocity model. (a) Absolute traveltime error. (b) Relative amplitude error. The color-bar for (a) shows the traveltime error in seconds. The color-bar for (b) shows the relative amplitude error (unitless).	56
4.1	SAIG velocity model (m/s).	65
4.2	Forward modeled shot gather at $x=3500\text{m}$ for the velocity model in figure 4.1.	65
4.3	Stacked migrated image for (a) migration via the adjoint, (b) damped least-squares migration, and (c) preconditioned least-squares migration with a 5-point triangular smoothing filter.	66

4.4	Common image gather located at $x=5000\text{m}$ for (a) migration via the adjoint, (b) damped least-squares migration, and (c) preconditioned least-squares migration with a 5-point triangular smoothing filter.	67
4.5	Convergence curves for damped and preconditioned least-squares migration.	68
4.6	(a) Synthetic reflectivity model. (b) Zero-offset seismic section produced using the forward Kirchhoff operator \mathbf{L} . (c) Least-squares migrated image using the aliased Kirchhoff operator based on Algorithm 1. Least-squares migrated image using the anti-aliased Kirchhoff operator based on Algorithm 2.	70
4.7	Convergence curve for aliased and anti-aliased least-squares Kirchhoff depth migration.	71
5.1	(a) True P-wave velocity model. (b) True S-wave velocity model. (c) Errors in S-velocity model. The color-bars show the velocity in units of m/s.	79
5.2	(a) PP data and (b) PS data produced by elastic finite difference modeling for a source located at $x=800\text{m}$. The direct wave was removed using a top mute.	80
5.3	Migration in the presence of S-velocity errors. (a) Adjoint PP image. (b) Adjoint PS image. (c) Conventional PS LSM image. (d) Warped adjoint PS image. (e) Warped PS LSM image.	81
5.4	Common image gathers located at $x=1000\text{m}$ for (a) PP adjoint, (b) PS adjoint, (c) conventional PS LSM, (d) warped PS adjoint, and (e) warped PS LSM. Panel (f) shows the shifts that were used/estimated by the warping operator; the color-bar shows the vertical shifts in units of meters.	82

CHAPTER 1

Introduction

1.1 Background

Seismic exploration is a branch of geophysics that focuses on searching for natural resources in the subsurface by recording, processing, and interpreting seismic data. Of the numerous tools available to exploration seismologists, seismic reflection is by far the most powerful. Seismic reflection is an active-source experiment that uses seismic waves to illuminate features in the subsurface. After processing the seismic reflection data, an image of the subsurface reflectivity can eventually be recovered. These images ultimately reveal the interfaces between different geological units, and are therefore invaluable in oil and gas exploration because they provide detailed information about the subsurface structure and stratigraphy. In addition to this, amplitude variations within the pre-stack image volume can provide direct evidence of hydrocarbons, thereby increasing the probability of success.

Several methods have been developed for generating seismic images. The earliest approaches were based on the assumption that the subsurface consists of laterally continuous, horizontally layered geological units. Under this assumption, each trace in a common midpoint (CMP) gather samples the same region in the subsurface (Yilmaz, 2001). Therefore, a time-domain pre-stack image volume can be obtained by applying a normal moveout (NMO) correction to the traces in each CMP gather. The time-domain image can be directly analyzed, or it can be converted into a depth-domain image by performing a time-to-depth conversion using the subsurface velocity model (Cameron et al., 2008). The image volume can be analyzed for variations in seismic amplitude vs. offset (AVO), or it can be stacked to form a structural image. This process has been widely used in industrial seismic exploration and remains a fundamental step in modern seismic processing workflows.

The normal moveout correction assumes the subsurface consists of flat geological layers. However, hydrocarbons are often located in structurally complex areas where the flat-layer assumption is invalid. Since the velocity model varies laterally in these areas, the NMO correction is incapable of completely flattening the CMP gathers. If the geology is simple enough that it can be represented by constant-dip layers, a dip moveout (DMO) correction can be used to improve the image (Deregowski, 1982; Hale, 1984; Notfors and Godfrey, 1987; Liner, 1990). The DMO correction is an extension of the NMO correction that aims to remove the dip-dependency of the stacking velocities, thus allowing dipping reflectors to be properly flattened.

Although the DMO correction can improve the accuracy of the seismic image, structurally complex geological regions require a more rigorous mathematical treatment. Moreover, even the simplest geological structures such as anticlines and synclines produce artifacts that the DMO correction cannot account for. Therefore, a more robust method for seismic imaging is required. This problem is addressed using a technique called seismic migration. Unlike the NMO/DMO method, which is based on a geometrical interpretation and generates a seismic image by shifting events vertically within the data domain, seismic migration aims to estimate a model that explains the data by *migrating* features from the data domain into the model domain. Since seismic migration is based on modeling, it satisfies the physics of the seismic experiment and is therefore capable of generating an accurate image in even the most structurally complex geological environments. For these reasons, migration has become a crucial part of the seismic processing workflow. However, as we will find throughout this thesis, limitations in the frequency content of the data as well as approximations used in the migration operator ultimately reduce the resolution and accuracy of the migrated seismic image. This thesis aims to advance the technology surrounding seismic migration by proposing modifications to the existing algorithms as well as incorporating some new ideas.

In the remainder of this thesis, the terms “seismic imaging” and “seismic migration” will be used interchangeably. However, a more formal treatment would differentiate between the two by noting that seismic imaging is a general term that describes the process of generating a seismic image, whereas seismic migration is a specific technique used for performing seismic imaging.

1.2 Seismic migration

In the seismic reflection experiment, a seismic source is used to send elastic waves down into the subsurface where they are reflected, refracted, and transmitted before being recorded back at the Earth's surface. Reflection seismology aims to use this data to recover an image of the geological interfaces off of which the recorded seismic waves reflected. Of the many techniques available for performing seismic imaging, seismic migration is the most powerful.

Seismic wave propagation is governed by the wave equation. Therefore, the seismic data that is recorded at the Earth's surface is a subset of the full wavefield. Seismic migration aims to undo the physics of the wave equation by numerically back-propagating the recorded seismic data down into the subsurface until it reaches its reflection/scattering point. This is typically accomplished using a two-step process: first extrapolate the source-side and receiver-side wavefields downward, and then apply an imaging condition to extract an image of the subsurface (Claerbout, 1971; Lumley, 1989; Rickett and Sava, 2002; Sava and Fomel, 2003, 2006). Seismic migration can therefore be viewed as a boundary value problem in which the goal is to estimate the wavefield in the subsurface for a given set of boundary observations.

In order to estimate the wavefield in the subsurface, seismic migration requires a method for simulating the propagation of seismic waves. Two general classes of algorithms have been developed for this purpose: wave equation and Kirchhoff/scattering. The two classes are typically referred to as wave equation migration (WEM) and Kirchhoff migration, respectively. As its name suggests, wave equation migration extrapolates the source-side and receiver-side wavefields downward via the wave equation. This can be accomplished in the time-space domain by solving the wave equation using the finite difference method (Boore, 1972), or in the frequency-wavenumber domain by iteratively extrapolating the wavefield downward using the one-way wave equation (Gazdag, 1978; Stolt, 1978; Stoffa et al., 1990; Mulder and Plessix, 2004). The former approach is a special type of wave equation migration that is given its own name: reverse time migration (RTM). Although WEM and RTM solve the wave equation in different domains, the final objective is the same: find a solution that satisfies the wave equation. Solutions of the full "two-way" wave equation generally produce the best seismic images because they most accurately represent the physics of the seismic experiment. However, the two-way wave equation is generally too expensive to solve in most practical applications. For this reason, one-way WEM has gained popularity as a low-cost alternative (Mulder and Plessix, 2004).

Although WEM and RTM are accurate, they suffer from a number of fundamental issues

that limit their practicality. As previously mentioned, they are computationally expensive; thus, the hardware requirements are generally unrealistic for large-scale 2D and 3D problems (Zhu and Lines, 1998; Mulder and Plessix, 2004). In addition to this, WEM and RTM require a uniform modeling grid; this poses a problem in areas with topography, although some effort has been made to address this issue (Shragge, 2014). Wave equation methods also require regularly spaced data; any sources and receivers that are unaligned with the modeling grid must be binned or interpolated to the nearest grid point. For these reasons, industry has been slow to adopt wave equation-based migration algorithms.

The aforementioned issues can be resolved by utilizing the second class of seismic migration algorithms: Kirchhoff migration. Kirchhoff migration seeks an integral approximation to the wave equation by making two key assumptions: (i) the data satisfies the high-frequency assumption, and (ii) the upward scattered wavefield is linearly related to the incident wavefield by a factor called the reflectivity. Evidently, Kirchhoff migration is based on scattering theory. Kirchhoff migration assumes the subsurface is a grid of scatterers and the seismic data consists of a sum of diffractions. An image of the subsurface reflectivity can therefore be obtained by collapsing the data-domain diffractions back to their model-domain scattering point. In order to accomplish this task, Kirchhoff migration requires a method for simulating the propagation of seismic waves from the sources & receivers at the surface to the scattering points in the subsurface; this is done using Green's functions (Schneider, 1978; Bleistein et al., 2001). For each source/receiver pair in the data and each scattering point in the model, a source-side Green's function is used to simulate the downward propagation of the incident wavefield and a receiver-side Green's function is used to simulate the upward propagation of the scattered wavefield. Since estimating Green's functions is substantially less expensive than solving the wave equation, Kirchhoff migration has a relatively low computational cost (Zhu and Lines, 1998). In addition to this, Kirchhoff migration does not require a uniform modeling grid, nor does it require evenly spaced data. Thus, Kirchhoff migration is naturally capable of handling topography, and does not require binning or interpolating the recorded seismic data. However, since Kirchhoff migration utilizes an integral approximation to the wave equation, it is less accurate than WEM/RTM and may be inadequate for imaging highly complex geological structures (Zhu and Lines, 1998). Despite this shortcoming, Kirchhoff migration remains the workhorse of most industrial seismic imaging workflows.

Regardless of whether WEM or Kirchhoff migration is used, the final image will always contain a number of artifacts due to the presence of noise in the data, errors in the velocity model, and poor subsurface illumination (Herron, 2000; Gray et al., 2001; Gray, 2013). Regularizing the seismic data can help minimize the number of artifacts in the image (Fomel

and Guitton, 2006). Least-squares migration (LSM) has also been shown to help with this problem (Nemeth et al., 1999; Chavent and Plessix, 1999; Duquet et al., 2000; Kuehl and Sacchi, 2003; Wang and Sacchi, 2006; Kaplan et al., 2010; Dong et al., 2012). Unlike traditional seismic migration algorithms, which generate an image by applying a linear migration operator to the data, least-squares migration produces an image by solving an optimization problem. The LSM cost function typically contains two terms that must be minimized simultaneously: (i) a data misfit term, which measures how well the predicted synthetic data fits the observed data, and (ii) a regularization term, which measures the prevalence of undesirable features in the model. Minimizing the misfit term ensures the inverted seismic image satisfies the observed data, while minimizing the regularization term reduces the prevalence of undesirable features in the model. Thus, if the relative weight/tradeoff between the two terms is correctly chosen, the inverted seismic image will fit the observed data and have minimal artifacts.

The main shortcoming of least-squares migration is its computational cost. In order to solve the LSM optimization problem, an iterative solver such as the conjugate gradient (CG) method must be used. The conjugate gradient method is a Krylov subspace method; each iteration of CG requires one application of the forward modeling operator and one application of the adjoint migration operator. Thus, the total cost of each iteration of CG is approximately double that of traditional seismic migration. Since multiple iterations of CG are required to reach convergence, the cost of LSM is substantial. One option for improving the rate of convergence LSM is to precondition the inverse problem (Wang et al., 2004). Preconditioning is generally capable of incorporating the same constraints as regularization, but also improves the rate of convergence because it improves the condition number of the pseudoinverse.

Since least-squares migration is solved iteratively, this thesis will periodically refer to LSM as the “iterative” approach to seismic migration. The alternative approach, which estimates a seismic image through application of a linear migration operator, will be referred to as the “direct approach”.

1.3 Image registration

Image registration is a computer vision technique for aligning the coordinates of two or more similar images. One image is designated as the reference, and the others are transformed (registered) to align their features with the reference. After the images have been registered, their features share common coordinates and can therefore be easily compared. Thus, image registration is a crucial step when comparing data and images that were collected during

different measurements.

Image registration has been used in a variety of applications including automatic targeting systems, audio signal processing, and medical imaging. More recently, image registration has been used in seismic exploration. One example is the alignment of synthetic and real seismograms (Liner and Clapp, 2004). Synthetic seismograms typically differ from real seismograms due to errors in the velocity model and approximations made during forward modeling. Thus, image registration can be used to align the synthetic events with the corresponding real events. This is particularly valuable in least-squares migration, where subtle differences between the synthetic and real seismograms can lead to notable changes in the migrated image. For example, Zeng et al. (2015) and Luo and Hale (2014) found that registering the synthetic events with the real events resulted in a more focused seismic image with better overall continuity. Image registration has also been used for estimating trim statics by registering the far-offset traces with the near-offset traces (Zhang et al., 2014). Another application has been the alignment of different seismic phases (Hale, 2013). Moreover, in order to compare PP¹ and PS² migrated images, the events in the PS image must be registered with the PP image to align their reflectors. By aligning the PP and PS phases, amplitude differences between the two images can be easily compared to make inferences about the subsurface geology. Lastly, image registration has been used for estimating the ratio between the P-wave and S-wave velocities (Compton and Hale, 2013; Gaiser, 1996; Liang and Hale, 2012; Nickel and Sonneland, 2004).

This thesis investigates a specific type of image registration called dynamic time warping (DTW). DTW was first used in the field of audio signal analysis for applications involving voice recognition (Sakoe and Chiba, 1978; Müller, 2007), and has more recently been used in seismic exploration (Hale, 2013). The DTW algorithm aims to estimate the relative shifts between two signals by “forward accumulating” the relative error followed by “backtracking” to estimate the shifts. The method is naturally capable of incorporating smoothness constraints; thus, DTW is robust in the presence of noise and other artifacts.

Conventional dynamic time warping is limited to working with one dimensional signals. However, most seismic data are inherently multidimensional. The most straightforward extension of DTW to multidimensional image registration has been shown to be computationally infeasible (Keysers and Unger, 2003). Moreover, the cost of this approach grows exponentially with the number of dimensions. Several computationally efficient alterna-

¹The abbreviation “PP” is commonly used to denote the compressional-wave to compressional-wave seismic phase.

²The abbreviation “PS” is commonly used to denote the compressional-wave to shear-wave seismic phase.

tives have been proposed to address this issue (Pishchulin, 2010; Mottl et al., 2002). One algorithm that has gained substantial popularity in recent years is the dynamic image warping (DIW) method proposed by Hale (2013). DIW estimates the relative shifts between two images on a trace-by-trace basis, much like DTW. However, DIW also includes lateral smoothness constraints to ensure the relative shifts are continuous across each dimension of the image. Consequently, DIW is stable in the presence of noise and other non-linearities. Hale and Compton (2013) also showed that performing DIW on a subsampled image volume can further improve the smoothness and continuity of the estimated shifts.

In this thesis, a new application of multidimensional image registration is reviewed. Multidimensional image registration has frequently been used for registering PP and PS images. However, it is typically assumed that the P- and S-velocity models are correct. In practice, S-velocity models are typically less accurate than P-velocity models due to the strong aliasing³ and low signal-to-noise ratio⁴ (SNR) associated with converted waves. As a result, the migrated PS image is typically less accurate than the PP image. Dynamic image warping can compensate for the S-velocity errors by warping the PS image to align its reflectors with the corresponding PP reflectors. Moreover, embedding the DIW operator in the converted wave LSM cost function allows for the recovery of a high-resolution image that is structurally accurate, despite the presence of S-velocity errors.

1.4 Scope of this thesis

1.4.1 Main contributions

This thesis aims to expand the technology surrounding least-squares Kirchhoff depth migration. The main contributions are:

1. Formulate an anti-aliased forward and adjoint Kirchhoff operator that are numerically adjoint to each other,
2. Demonstrate the benefits of preconditioned least-squares migration by promoting smoothness within common image gathers, and
3. Provide a new method for performing converted wave least-squares migration in the presence of S-velocity errors by embedding a dynamic image warping operator into the cost function.

³Converted waves propagate as shear-waves as they arrive at the surface. Since shear waves travel slower than compressional waves, they are more susceptible to aliasing than compressional-wave arrivals.

⁴Converted waves are predominantly measured on the horizontal component of seismic receivers. Many sources of noise (e.g. the wind) also travel horizontally; thus, converted waves often suffer from low signal-to-noise ratio.

1.4.2 Organization

This thesis is organized as follows. Chapter 2 derives the forward and adjoint Kirchhoff operators from first principles. The concept of Kirchhoff operator aliasing is discussed and a solution is presented based on a triangular filter that can be customized on-the-fly. Chapter 3 derives the traditional ray tracing equations from the eikonal equation. The wavefront construction method is then introduced as an extension of ray tracing that propagates the full wavefront rather than individual rays. Finally, a robust technique is reviewed for estimating the amplitude component of the seismic Greens functions by comparing the wavefront at consecutive time steps. Chapter 4 reviews least-squares migration. Two different regularization strategies are discussed: (i) damping by minimizing the norm of the model, and (ii) preconditioning by promoting smoothness within common image gathers. Chapter 5 provides a review of the dynamic image warping method. A new converted wave least-squares migration cost function is then formulated that uses dynamic image warping to register the PS image with the PP image, thereby reducing the sensitivity of the cost function to errors in the S-velocity model.

CHAPTER 2

The Kirchhoff operator

2.1 Introduction

Forward modeling is the process in which a known model is used to generate an observable outcome. In seismic exploration, the model consists of the geophysical parameters that characterize the subsurface, and the observations are the seismic data that are collected during a field experiment. The set of equations that relate the geophysical parameters to the observed seismic data is generally referred to as the forward problem. Therefore, the forward problem is a simulation that aims to mimic the physical processes that occur during a seismic experiment.

The complete physical representation of the forward problem is often too complex and computationally expensive to solve for practical applications. Instead, it is common to approximate the forward problem by using a simplified set of governing equations. Since the Earth's subsurface is an elastic medium, the propagation of seismic waves is governed by the elastic wave equation. However, simulating the elastic wave equation is prohibitive and is generally not necessary to achieve satisfactory results. Since it is often possible to isolate or extract P-wave arrivals from the recorded wavefield, it is generally sufficient to replace the elastic wave equation with the acoustic wave equation and treat the Earth as an acoustic medium. The wave equation can be further simplified by assuming the wavefield and model satisfy the high-frequency assumption, under which the contribution of multiply-scattered waves is negligible. As a result, the recorded wavefield can be expressed explicitly via an integral equation known as the Kirchhoff integral. The Kirchhoff integral maps scatterers from the model domain into diffractions in the data domain based on computed travel-times, then scales the amplitude using three factors: (i) the reflectivity of the scatterer,

(ii) the geometrical spreading factor, and (iii) the scattering angle. The resulting data is a primaries-only approximation of the acoustic wave equation.

Forward modeling is an interesting and useful topic on its own. However, industrial applications of seismology are generally more interested in solving the inverse problem. That is, for a given set of a geophysical data, what was the physical model that caused it? Inverse problems are typically more complicated than forward problems because the forward modeling operator is generally not invertible. Geophysicists must therefore utilize alternative methods for solving inverse problems. This thesis focuses on a class of solvers that aims to find a solution by iteratively minimizing the error between the true observed data and the predicted synthetic data. In order to adopt this strategy, a forward modeling operator and its exact adjoint must be derived. In this chapter, we derive a forward & adjoint pair for performing forward modeling and migration of seismic data. The analysis begins by deriving the momentum equation for an arbitrary medium. Next, the elastic wave equation is formulated for an isotropic medium with homogeneous layers. Following this, the acoustic wave equation is presented and linear scattering under the high-frequency assumption is reviewed. Next, the Kirchhoff approximation is applied, which leads to the development of a forward & adjoint operator pair based on the Kirchhoff integral. Finally, the problem of Kirchhoff operator aliasing is discussed, including an inexpensive solution based on a triangular filter.

2.2 The wave equation

Seismic imaging algorithms are based on solutions to the wave equation. Multicomponent seismic imaging requires an algorithm based on the elastic wave equation, while single-component seismic imaging can achieve satisfactory results using the acoustic wave equation. In this section, the elastic wave equation and the acoustic approximation are derived from first principles.

2.2.1 The momentum equation

We begin by considering Newton's second law of motion, which describes the relationship between a force F_i and acceleration \ddot{u}_i in the i^{th} spatial dimension:

$$F_i = m\ddot{u}_i, \quad (2.1)$$

where u_i is the displacement, the two overdots denote the second derivative with respect to time, $m = \rho dx_1 dx_2 dx_3$ is mass, ρ is mass density, and i can take on the value 1, 2, or 3 to

represent the three spatial dimensions. Force is related to stress via

$$F_i = (\partial_j \tau_{ij} + f_i) dx_1 dx_2 dx_3 , \quad (2.2)$$

where $\partial_j = \frac{\partial}{\partial x_j}$ denotes differentiation with respect to the j^{th} spatial dimension, τ_{ij} is the stress tensor that describes the force acting in direction \hat{i} on a plane whose normal faces in direction \hat{j} , and f_i is the body force term. Combining equation 2.1 and 2.2 leads to the momentum equation

$$\rho \ddot{u}_i = \partial_j \tau_{ij} + f_i , \quad (2.3)$$

where ρ is a function of position \mathbf{x} , and u_i , τ_{ij} , and f_i are functions of position \mathbf{x} and time t . The momentum equation is a fundamental equation that lays the foundation for much of seismology. It constrains the physical system by requiring changes in particle momentum to be balanced by the sum of external forces acting on them. In seismology, the body force generally consists of a gravity term and a source term. The gravity term is significant in low-frequency normal mode seismology, but can generally be ignored in seismic exploration because its contribution is negligible for typical seismic wavelengths (Shearer, 2009). The source term, on the other hand, is not negligible and must be considered. Therefore, f_i is often referred to as the source term in seismic exploration.

2.2.2 The elastic wave equation

Equation 2.3 describes the displacement in terms of the stresses involved. The linearized stress-strain equation (Hooke's law) describes the relationship between those stresses and the associated strain:

$$\tau_{ij} = C_{ijkl} e_{kl} , \quad (2.4)$$

where C_{ijkl} is the elastic moduli (stiffness) tensor and e_{kl} is the strain tensor. The subscripts i, j, k , and l can take on the value 1, 2, or 3. Therefore, the stiffness tensor has 81 elements. If the medium is isotropic, the number of unique elements is reduced to 2 and the stiffness tensor is given by

$$C_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}) , \quad (2.5)$$

where λ and μ are called the Lamè parameters. The Lamè parameters are material-dependent and therefore characterize the physical properties of the medium. The first Lamè parameter, λ , has no obvious physical meaning (Laursen, 2013). The second Lamè parameter, μ , is the shear modulus and describes the resistance of the material to shearing.

Next, we assume the material displacements are small compared with each dimension of the body (infinitesimal strain theory). Under this assumption, the strain tensor can be

linearized by eliminating the second-order terms. As a result, the 9 elements of the strain tensor are given by

$$e_{kl} = \frac{1}{2}(\partial_k u_l + \partial_l u_k) . \quad (2.6)$$

Combining equation 2.3, 2.4, and 2.6 results in

$$\partial_j(C_{ijkl}u_{kl}) - \rho\ddot{u}_i = -f_i . \quad (2.7)$$

Finally, subbing equation 2.5 into equation 2.7 results in the isotropic, elastic wave equation for homogeneous media

$$(\lambda + 2\mu)\nabla(\nabla \cdot \bar{u}) + \mu\nabla \times \nabla \times \bar{u} - \rho\frac{\partial^2 \bar{u}}{\partial t^2} = -\bar{f} , \quad (2.8)$$

where \bar{u} is the displacement vector and ∇ denotes the spatial gradient defined by

$$\nabla = \hat{i}\frac{\partial}{\partial x_1} + \hat{j}\frac{\partial}{\partial x_2} + \hat{k}\frac{\partial}{\partial x_3} . \quad (2.9)$$

Helmholtz theorem states that any smoothly varying vector field can be expressed as the sum of an irrotational vector field and a solenoidal vector field. Therefore, defining

$$\bar{u} = \nabla\phi + \nabla \times \bar{\psi} , \quad (2.10)$$

we obtain the equation for a compressional wave

$$\nabla^2\phi - \frac{1}{\alpha^2}\frac{\partial^2\phi}{\partial t^2} = -f_\phi , \quad (2.11)$$

and a shear wave

$$\nabla^2\bar{\psi} - \frac{1}{\beta^2}\frac{\partial^2\bar{\psi}}{\partial t^2} = -\bar{f}_\psi , \quad (2.12)$$

where $\alpha = \sqrt{\frac{\lambda+2\mu}{\rho}}$ is the P-wave velocity, $\beta = \sqrt{\frac{\mu}{\rho}}$ is the S-wave velocity, f_ϕ is the compressional component of the source, and \bar{f}_ψ is the rotational/shear component of the source. Helmholtz theorem can therefore be used to decompose an elastic wavefield into separate P- and S-potentials. As a result, it is possible to process and image P- and S-wave data separately, as will be exploited later in this thesis.

2.2.3 The acoustic approximation

The elastic wave equation is prohibitively complicated and computationally expensive to solve. It is therefore desirable to simplify the seismic modeling equation. As several authors have exploited (e.g. Kosloff and Baysal (1983); Tarantola (1984b); Bleistein (1987); Lambare

et al. (1992)), treating the subsurface as a fluid greatly simplifies the problem while still allowing for the recovery of ample subsurface information. Thus we proceed under the acoustic approximation. Since fluids are incapable of supporting shear deformation, $\mu = 0$ and the elastic wave equation simplifies to

$$\nabla(\lambda \nabla \cdot \bar{u}) - \rho \frac{\partial^2 \bar{u}}{\partial t^2} = -\bar{f}. \quad (2.13)$$

Defining the pressure as

$$p = -\lambda \nabla \cdot \bar{u}, \quad (2.14)$$

we obtain the acoustic wave equation

$$\nabla^2 p - \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2} = -\tilde{f}, \quad (2.15)$$

where $c = c(\mathbf{x})$ is the acoustic velocity and $\tilde{f} = \nabla \cdot \frac{1}{\rho} \bar{f}$. The acoustic wave equation describes the propagation of compressional waves in a fluid medium. Although the Earth's subsurface is strictly elastic, the acoustic wave equation is capable of modeling the dilatational part of the wavefield. Therefore, the acoustic wave equation can be used for seismic modeling if the P-wave data can be extracted from the recorded wavefield with minimal crosstalk. A number of techniques have been proposed to address this challenge. For example, Dellinger and Etgen (1990) proposed the application of Helmholtz decomposition to elastic data in the frequency-wavenumber domain. van der Baan (2006) proposed a technique for separating PP and PS data using independent component analysis. Other techniques based on plane-wave decomposition have also been investigated (Cary, 1998). In some cases, the P-wave data is nearly polarized with the vertical axis and can therefore be extracted without performing additional processing steps. This is often the case for land seismic data due to the presence of a low-velocity weathered layer near the Earth's surface. For these reasons, the acoustic approximation has been used as the foundation of many seismic processing and imaging workflows (Schneider, 1978; Berryhill, 1979; Baysal et al., 1983; Gazdag and Sguazzero, 1984; Tarantola, 1984a; Claerbout, 1985; Symes and Carazzone, 1991).

2.3 Linear scattering

Since seismologists are mainly interested in primary reflections, it is common to simplify the problem based on linear scattering (Bleistein et al., 2001). We begin by considering the temporal Fourier transform of equation 2.15, which leads to the inhomogeneous Helmholtz equation

$$\nabla^2 p(\mathbf{x}, \mathbf{s}, \omega) + \frac{\omega^2}{c(\mathbf{x})^2} p(\mathbf{x}, \mathbf{s}, \omega) = -\tilde{F}(\omega) \delta(\mathbf{x} - \mathbf{s}), \quad (2.16)$$

where $\tilde{F}(\omega)$ is the temporal Fourier transform of \tilde{f} , ω is the angular frequency, \mathbf{x} is an arbitrary position in the subsurface, and \mathbf{s} is the location of the source. Equation 2.16 can be expressed more compactly as

$$\mathcal{L}(\mathbf{x}, \omega)p(\mathbf{x}, \mathbf{s}, \omega) = -\tilde{F}(\omega)\delta(\mathbf{x} - \mathbf{s}) , \quad (2.17)$$

where $\mathcal{L}(\mathbf{x}, \omega) = \nabla^2 + \frac{\omega^2}{c(\mathbf{x})^2}$ is the Helmholtz operator. Imposing the high-frequency assumption diminishes the contribution of multiples. Strictly speaking, the high-frequency assumption is only valid when λ and μ vary on a length scale that is large compared with typical seismic wavelengths. However, it turns out that primary reflections can still be accurately modeled even if this condition is violated, so long as the input model is sufficiently smoothed prior to modeling (Gray, 2000). The total wavefield can then be written as the sum of the incident wavefield $p_I(\mathbf{x}, \mathbf{s}, \omega)$ and the scattered wavefield $p_S(\mathbf{x}, \mathbf{s}, \omega)$:

$$p(\mathbf{x}, \mathbf{s}, \omega) \approx p_I(\mathbf{x}, \mathbf{s}, \omega) + p_S(\mathbf{x}, \mathbf{s}, \omega) , \quad (2.18)$$

where $p_I(\mathbf{x}, \mathbf{s}, \omega)$ satisfies the Green's function for the source:

$$\mathcal{L}(\mathbf{x}, \omega)p_I(\mathbf{x}, \mathbf{s}, \omega) = -\tilde{F}(\omega)\delta(\mathbf{x} - \mathbf{s}) . \quad (2.19)$$

Equation 2.19 is the inhomogeneous Helmholtz equation for the incident wavefield. It describes the propagation of acoustic waves from the source down into the subsurface, ignoring reflections. Thus, equation 2.19 can be used to extrapolate the source-side wavefield downward.

Subbing equation 2.18 into equation 2.17 results in

$$\mathcal{L}(\mathbf{x}, \omega)p_I(\mathbf{x}, \mathbf{s}, \omega) + \mathcal{L}(\mathbf{x}, \omega)p_S(\mathbf{x}, \mathbf{s}, \omega) = -\tilde{F}(\omega)\delta(\mathbf{x} - \mathbf{s}) . \quad (2.20)$$

The terms containing the scattered wavefield can then be isolated by subbing equation 2.19 into equation 2.20, which leads to

$$\mathcal{L}(\mathbf{x}, \omega)p_S(\mathbf{x}, \mathbf{s}, \omega) = 0 . \quad (2.21)$$

Equation 2.21 is the homogeneous Helmholtz equation for the scattered wavefield. It describes the propagation of acoustic waves that have already been reflected once and is therefore responsible for modeling the primary reflections in a seismic record. In order to solve equation 2.21, Green's theorem is used. Following Bleistein et al. (2001), we define a

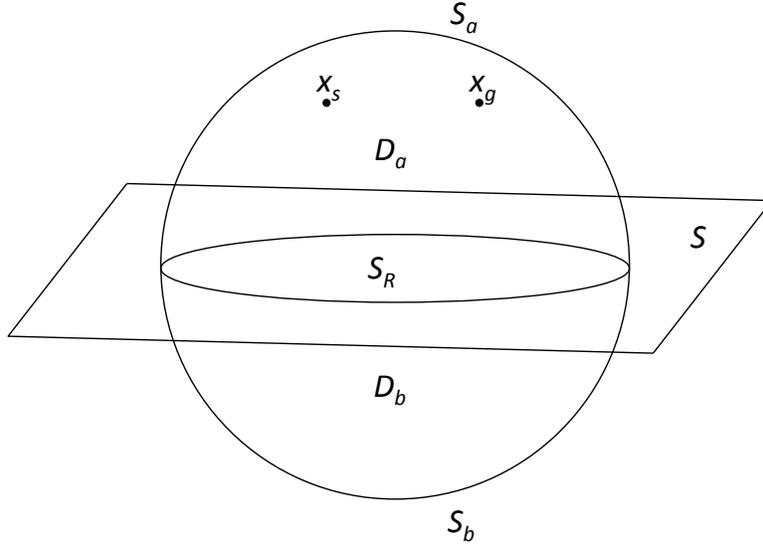


Figure 2.1: Diagram showing the intersection of a sphere D with a plane S . The subscript “a” denotes features in the upper hemisphere and the subscript “b” denotes features in the lower hemisphere.

receiver-side Green’s function $G(\mathbf{x}, \mathbf{r}, \omega)$ subject to

$$\mathcal{L}^*(\mathbf{x}, \omega)G^*(\mathbf{x}, \mathbf{r}, \omega) = -\delta(\mathbf{x} - \mathbf{r}) , \quad (2.22)$$

where \mathbf{r} is the location of the receiver and the superscript “ \star ” is used to denote the adjoint. Applying Green’s theorem to $p(\mathbf{x}, \mathbf{s}, \omega)$ and $G^*(\mathbf{x}, \mathbf{r}, \omega)$ results in

$$\int_{D_a} [G^* \mathcal{L}p - p \mathcal{L}^* G^*] dV = \int_{S_a + S_R} \left[G^* \frac{\partial p}{\partial n} - p \frac{\partial G^*}{\partial n} \right] dS , \quad (2.23)$$

where the dependent variables have been dropped to improve readability. The integration regions D_a , S_a , and S_R are the upper hemisphere volumes and surfaces created when a sphere D intersects a reflection surface S , as shown in figure 2.1. Specifically, D_a is the upper hemisphere volume, S_a is the upper hemisphere surface that bounds D_a on the top, and S_R is the part of S inside of D that bounds D_a on the bottom. Lastly, $\frac{\partial}{\partial n}$ is the outward normal derivative on the boundary surface. Solving the integral on the left side of equation 2.23 yields

$$\begin{aligned}
\int_{D_a} [G^* \mathcal{L}p - p \mathcal{L}^* G^*] dV &= p(\mathbf{r}, \mathbf{s}, \omega) - \tilde{F}(\omega) G^*(\mathbf{s}, \mathbf{r}, \omega) \\
&= p_I(\mathbf{r}, \mathbf{s}, \omega) + p_S(\mathbf{r}, \mathbf{s}, \omega) - \tilde{F}(\omega) G^*(\mathbf{s}, \mathbf{r}, \omega) \\
&= \int_{S_a + S_R} \left[G^* \frac{\partial p}{\partial n} - p \frac{\partial G^*}{\partial n} \right] dS .
\end{aligned} \tag{2.24}$$

Performing a similar analysis with $p_I(\mathbf{x}, \mathbf{s}, \omega)$ and $G^*(\mathbf{x}, \mathbf{r}, \omega)$ leads to

$$\begin{aligned}
\int_{D_a} [G^* \mathcal{L}p_I - p_I \mathcal{L}^* G^*] dV &= p_I(\mathbf{r}, \mathbf{s}, \omega) - \tilde{F}(\omega) G^*(\mathbf{s}, \mathbf{r}, \omega) \\
&= \int_{S_a + S_R} \left[G^* \frac{\partial p_I}{\partial n} - p_I \frac{\partial G^*}{\partial n} \right] dS .
\end{aligned} \tag{2.25}$$

Next, combining equation 2.24 and 2.25 gives the result

$$p_S(\mathbf{r}, \mathbf{s}, \omega) = \int_{S_a + S_R} \left[G^* \frac{\partial p_S}{\partial n} - p_S \frac{\partial G^*}{\partial n} \right] dS . \tag{2.26}$$

If we consider the limit where the radius of the sphere approaches infinity, then S_R approaches S and the integral over S_a approaches zero (Bleistein et al., 2001). Therefore, equation 2.26 becomes

$$p_S(\mathbf{r}, \mathbf{s}, \omega) = \int_S \left[G^*(\mathbf{x}, \mathbf{r}, \omega) \frac{\partial p_S(\mathbf{x}, \mathbf{s}, \omega)}{\partial n} - p_S(\mathbf{x}, \mathbf{s}, \omega) \frac{\partial G^*(\mathbf{x}, \mathbf{r}, \omega)}{\partial n} \right] dS . \tag{2.27}$$

Finally, the theorem of reciprocity states that the adjoint Green's function is related to the Green's function via $G(\mathbf{r}, \mathbf{x}, \omega) = G^*(\mathbf{x}, \mathbf{r}, \omega)$. Therefore, we arrive at the result

$$p_S(\mathbf{r}, \mathbf{s}, \omega) = \int_S \left[G(\mathbf{r}, \mathbf{x}, \omega) \frac{\partial p_S(\mathbf{x}, \mathbf{s}, \omega)}{\partial n} - p_S(\mathbf{x}, \mathbf{s}, \omega) \frac{\partial G(\mathbf{r}, \mathbf{x}, \omega)}{\partial n} \right] dS . \tag{2.28}$$

Equation 2.28 is an expression for the scattered wavefield in terms of an integral of itself. Since p_S is present both inside and outside of the integral, it is impossible to rearrange the equation to isolate it. Therefore, an alternative approach must be adopted to obtain an explicit formulation for the scattered wavefield. This problem is addressed in the next section by applying the Kirchhoff approximation.

2.4 The Kirchhoff approximation

The Kirchhoff approximation is a technique for linearizing the non-linear Kirchhoff integral in equation 2.28. It assumes the following linear relationship between the incident wavefield

and the scattered wavefield at the location of each scatter:

$$\begin{aligned} p_S(\mathbf{x}, \mathbf{s}, \omega) &= R(\mathbf{x}, \theta) p_I(\mathbf{x}, \mathbf{s}, \omega) , \\ \frac{\partial p_S(\mathbf{x}, \mathbf{s}, \omega)}{\partial n} &= -R(\mathbf{x}, \theta) \frac{\partial p_I(\mathbf{x}, \mathbf{s}, \omega)}{\partial n} , \end{aligned} \quad (2.29)$$

where $R(\mathbf{x}, \theta)$ is the angle-dependent reflectivity and θ is the reflection angle (Bleistein et al., 2001). Subbing equation 2.29 into equation 2.28 gives

$$p_S(\mathbf{r}, \mathbf{s}, \omega) = \int_S R(\mathbf{x}, \theta) \left[p_I(\mathbf{x}, \mathbf{s}, \omega) \frac{\partial G(\mathbf{r}, \mathbf{x}, \omega)}{\partial n} + G(\mathbf{r}, \mathbf{x}, \omega) \frac{\partial p_I(\mathbf{x}, \mathbf{s}, \omega)}{\partial n} \right] dS . \quad (2.30)$$

According to the product rule for differentiation, equation 2.30 is equivalent to

$$p_S(\mathbf{r}, \mathbf{s}, \omega) = \int_S R(\mathbf{x}, \theta) \frac{\partial (p_I(\mathbf{x}, \mathbf{s}, \omega) G(\mathbf{r}, \mathbf{x}, \omega))}{\partial n} dS . \quad (2.31)$$

Equation 2.31 is the Kirchhoff integral for the upward scattered wavefield. It shows that the primary reflections for a known reflectivity model can be computed so long as the source and receiver Green's functions are known. Therefore, the accuracy of the Kirchhoff integral is ultimately limited by accuracy of the Green's functions $p_I(\mathbf{x}, \mathbf{s}, \omega)$ and $G(\mathbf{r}, \mathbf{x}, \omega)$.

The Green's function for an arbitrary background field has the form

$$G(\mathbf{x}, \mathbf{y}, \omega) = A(\mathbf{x}, \mathbf{y}) e^{i\omega\tau(\mathbf{x}, \mathbf{y})} , \quad (2.32)$$

where $A(\mathbf{x}, \mathbf{y})$ is the amplitude (geometrical spreading factor) and $\tau(\mathbf{x}, \mathbf{y})$ is the traveltime. By substituting the WKBJ approximation for the wavefield into the Helmholtz equation, it can be shown that $A(\mathbf{x}, \mathbf{y})$ obeys the transport equation

$$2\nabla\tau(\mathbf{x}, \mathbf{y}) \cdot \nabla A(\mathbf{x}, \mathbf{y}) + A(\mathbf{x}, \mathbf{y}) \nabla^2 \tau(\mathbf{x}, \mathbf{y}) = 0 , \quad (2.33)$$

and $\tau(\mathbf{x}, \mathbf{y})$ obeys the eikonal equation

$$\frac{1}{c^2(\mathbf{x})} - (\nabla\tau(\mathbf{x}, \mathbf{y}))^2 = 0 . \quad (2.34)$$

Therefore, the source and receiver Green's functions can be obtained by solving equation 2.33 and 2.34. See Appendix A for a complete derivation of the eikonal and transport equations via the WKBJ approximation. For now, we proceed under the assumption that a solution to the eikonal and transport equations can be found. In chapter 3, a numerical technique for estimating the Green's functions will be reviewed in detail.

Using equation 2.32 as a template for the source and receiver Green's functions, equation 2.31 becomes

$$p_S(\mathbf{r}, \mathbf{s}, \omega) = i\omega \tilde{F}(\omega) \int_S R(\mathbf{x}, \theta) A(\mathbf{r}, \mathbf{x}, \mathbf{s}) [\mathbf{n} \cdot \nabla \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})] e^{i\omega \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})} dS, \quad (2.35)$$

where

$$\begin{aligned} A(\mathbf{r}, \mathbf{x}, \mathbf{s}) &= A(\mathbf{r}, \mathbf{x}) A(\mathbf{x}, \mathbf{s}), \\ \tau(\mathbf{r}, \mathbf{x}, \mathbf{s}) &= \tau(\mathbf{r}, \mathbf{x}) + \tau(\mathbf{x}, \mathbf{s}). \end{aligned}$$

Equation 2.35 can be further simplified to (Bleistein and Gray, 2002)

$$p_S(\mathbf{r}, \mathbf{s}, \omega) = S(\omega) \int_S R(\mathbf{x}, \theta) A(\mathbf{r}, \mathbf{x}, \mathbf{s}) \left| \frac{2\cos(\theta)}{c(\mathbf{x})} \right| e^{i\omega \tau(\mathbf{r}, \mathbf{x}, \mathbf{s})} dS, \quad (2.36)$$

where $S(\omega)$ is the source signature given by

$$S(\omega) = \begin{cases} -i\omega \tilde{F}(\omega) & \text{for 3D} \\ |\omega| \tilde{F}(\omega) & \text{for 2D} \\ \sqrt{|\omega|} \sqrt{\frac{\sigma_s \sigma_r}{\sigma_s + \sigma_r}} e^{3i\pi \cdot \text{sgn}(\omega)/4} \tilde{F}(\omega) & \text{for 2.5D.} \end{cases} \quad (2.37)$$

The parameters σ_s and σ_r are responsible for describing out-of-plane behavior for the 2.5 dimensional case (Bleistein and Gray, 2002). Equation 2.36 can be expressed more compactly using linear algebra notation:

$$\mathbf{d} = \mathbf{Lm}, \quad (2.38)$$

where $\mathbf{d} = \mathbf{d}(\mathbf{r}, \mathbf{s}, \omega)$ is the seismic data in the frequency domain, $\mathbf{m} = \mathbf{m}(\mathbf{x}, \theta)$ is the angle-dependent or offset-dependent reflectivity, and \mathbf{L} is the linear Kirchhoff forward modeling operator that consists of two main parts: (i) integration over each scatterer in the model to accumulate the Earth's impulse response, and (ii) multiplication with a wavelet to honor the seismic source. The advantage of using linear algebra notation will become clear when the concept of least-squares migration is introduced in chapter 4.

The Kirchhoff integral can be used to predict seismic data for a known reflectivity model. In terms of practical implementation, it suggests that seismic data can be computed by summing the contribution from each subsurface scatterer. For a given source-receiver pair and a single subsurface scatterer, a source-side Green's function must be computed to simulate the downward propagation of seismic waves from the source to the scatterer, and a receiver-side Green's function must be computed to simulate the upward propagation of scattered waves from the scatterer to the receiver. Therefore, the core of Kirchhoff modeling consists of computing source and receiver Green's functions for each source-receiver pair in the data

and each scatterer in the model.

Our analysis of Kirchhoff modeling was carried out in the frequency domain. In practice, Kirchhoff modeling is typically performed in the time domain. The change from frequency to time is trivial and can be achieved by taking the inverse Fourier transform of equation 2.36. As a result, multiplication of the source gets replaced with convolution, and integration of the complex exponential gets replaced with integration of an impulse response $\delta(t - \tau)$. Time-domain implementations of Kirchhoff modeling therefore have the same algorithmic structure as frequency-domain implementations: first estimate the source-side and receiver-side Green's functions to form the Earth's impulse response, then perform convolution with a wavelet to honor the source. Importantly, time-domain Kirchhoff modeling can be expressed using the same linear algebra notation that was used in the frequency-domain: $\mathbf{d} = \mathbf{L}\mathbf{m}$, where $\mathbf{d} = \mathbf{d}(\mathbf{r}, \mathbf{s}, t)$ is the seismic data in the time domain.

So far, we have derived a forward modeling integral that can be used to generate seismic data. Forward modeling is a valuable tool that has many applications in seismic exploration. For example, it can be used to design seismic surveys and pick optimal parameters when up-scaling from 2D to 3D. It can also be used to verify geological interpretations or to estimate the seismic response for a specific geological feature. Although forward modeling has many uses, it is incapable of recovering direct, quantitative information about the subsurface. In order to recover a quantitative description of the subsurface, an inverse problem must be solved.

2.5 Kirchhoff migration

An inverse problem is a mathematical framework that allows the unknown physical parameters of a system to be estimated from a set of observations. In the case of seismic imaging, the observations are the recorded seismic data, and the unknown physical parameters consist of an image of the subsurface reflectivity. The mathematical framework that relates the seismic data to the subsurface reflectivity is the Kirchhoff integral. Intuitively, an inverse problem should be solved by undoing the physics of the forward problem. This implies that an image of the subsurface reflectivity can be obtained by multiplying the recorded seismic data with the inverse of the forward modeling operator: $\hat{\mathbf{m}} = \mathbf{L}^{-1}\mathbf{d}$. In practice, \mathbf{L}^{-1} is either unstable or impossible to estimate, so an alternative approach must be used.

Two classes of techniques exist for performing seismic migration: direct and iterative. Direct migration techniques aim to transform the seismic data into an image of the subsurface via an explicit formula. Moreover, they seek an approximation of the inverse that is capable

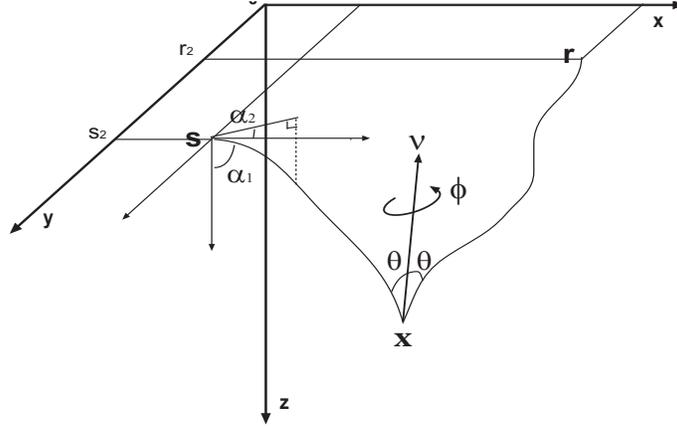


Figure 2.2: Coordinates for Kirchhoff AVA migration. Figure from Feng (2004).

of preserving the dominant features in the migrated image (Beylkin, 1985; Bleistein, 1987; Bleistein and Gray, 2002). For example, Bleistein and Gray (2002) proposed the following algorithm for common-angle migration:

$$R(\mathbf{x}, \theta, \phi) = \frac{1}{8\pi^3} \left[\frac{2\cos(\theta)}{c(\mathbf{x})} \right]^2 \int \int S(\omega) \frac{p_S(\mathbf{r}, \mathbf{s}, \omega)}{A(\mathbf{r}, \mathbf{x}, \mathbf{s})} e^{-i\omega\tau(\mathbf{r}, \mathbf{x}, \mathbf{s})} \left| \bar{\nu} \cdot \frac{\partial \bar{\nu}}{\partial \alpha_1} \times \frac{\partial \bar{\nu}}{\partial \alpha_2} \right| \cdot \delta(\theta' - \theta) \delta(\phi' - \phi) \left| \frac{\partial(\alpha_1, \alpha_2, \theta', \phi')}{\partial(s_1, s_2, r_1, r_2)} \right| ds_1 ds_2 dr_1 dr_2 d\omega, \quad (2.39)$$

where $\bar{\nu}$ is the unit normal vector for the reflector, θ is the reflection angle, ϕ is the azimuth, and α_1 and α_2 are the raypath angles, as shown in figure 2.2. Direct approaches for migration have gained substantial popularity in industrial seismic imaging due to their low computational cost, simplicity, and versatility. However, they also suffer from a number of drawbacks. Most notably, the predicted synthetic data for the migrated image does not match the true observed data. Therefore, the migrated image does not satisfy the physics of the forward problem, which in turn reduces our confidence in the solution and can make it difficult to evaluate the accuracy of the inversion. Additionally, direct seismic migration techniques are less robust in the presence of noise and incomplete seismic data (Nemeth et al., 1999). As a result, the inverted image is often contaminated with noise and artifacts.

The second class of techniques for performing seismic migration takes an iterative approach. Unlike direct seismic migration, which attempts to approximate the inverse using an explicit transformation, iterative seismic migration aims to estimate the true model by approximating the pseudoinverse. Moreover, iterative methods aim to solve an optimization problem by

repeatedly updating the model until it optimally satisfies the recorded seismic data. Since the predicted synthetic data satisfies the observed wavefield, iterative methods are generally considered an improvement over direct transformations. Chapter 4 discusses iterative methods for seismic migration in detail. In order to implement an iterative solver, the forward modeling operator and its adjoint must be derived. We have already developed an expression for the forward problem, as shown in equation 2.36. Therefore, all that remains is to derive its adjoint. Since the Kirchhoff integral is a Fredholm integral of the first kind, its adjoint is given by

$$R(\mathbf{x}, \theta) = \int S^*(\omega) p_S^*(\mathbf{r}, \mathbf{s}, \omega) A(\mathbf{r}, \mathbf{x}, \mathbf{s}) \left| \frac{2\cos(\theta)}{c(\mathbf{x})} \right| e^{i\omega\tau(\mathbf{r}, \mathbf{x}, \mathbf{s})} ds d\mathbf{r} d\omega, \quad (2.40)$$

where ds and $d\mathbf{r}$ denote the sources and receivers, and $d\omega$ is the frequency increment. Much like the forward Kirchhoff integral, the adjoint integral has two main parts: (i) multiplication with the complex conjugate of the source wavelet, followed by (ii) summation over a traveltime isochron to collect scattered energy.

Since the adjoint maps data-domain features into the model domain, it can be viewed as a seismic migration operator. In this thesis, migration via the adjoint will serve as the benchmark for evaluating the performance of our iterative methods. Note that the adjoint is not orthogonal and is therefore a poor approximation of the inverse. As a result, using the adjoint integral for seismic migration will result in a blurred image that is incapable of satisfying the observed seismic data (Youzwishen, 2003).

The adjoint Kirchhoff integral can be expressed using linear algebra notation as

$$\hat{\mathbf{m}} = \mathbf{L}^\dagger \mathbf{d}, \quad (2.41)$$

where the superscript “ \dagger ” denotes the adjoint (Hermitian transpose) and $\mathbf{d} = \mathbf{d}(\mathbf{r}, \mathbf{s}, \omega)$ is the seismic data in the frequency domain. The adjoint Kirchhoff integral is typically implemented in the time domain, which can be achieved by taking the inverse Fourier transform of equation 2.40. This can be accomplished by taking the same steps that were used for the forward integral: replacing multiplication of the source with convolution, and replacing the complex exponential with an impulse response. The resulting time-domain integral can be expressed using the same linear algebra notation that was used in the frequency domain: $\hat{\mathbf{m}} = \mathbf{L}^\dagger \mathbf{d}$, where $\mathbf{d} = \mathbf{d}(\mathbf{r}, \mathbf{s}, t)$ is the seismic data in the time domain.

Equations 2.36 and 2.40 inherently assume that the subsurface is a continuous medium and that the spacing between neighboring sources/receivers is infinitesimally small, thereby justifying an integral formulation. In practice, computers are limited to working with dis-

cretized models, and seismic surveys are performed using a finite number of sources and receivers. As a result, the integrals in equations 2.36 and 2.40 get replaced with finite summations. The result is a discretized forward and adjoint Kirchhoff operator (\mathbf{L} and \mathbf{L}^\dagger) that can be implemented based on the pseudocode in Algorithm 1.

The “rhofilter” in Algorithm 1 applies a phase-shift to the data and scales its amplitude spectrum by a factor of ω to account for the $i\omega$ in $S(\omega)$. Practical implementations of the Kirchhoff forward/adjoint operator (including the one used in this thesis) are more complicated than Algorithm 1 and include several additional features. For example, rounding of indexes generally gets replaced with interpolation to improve the stability of the operator. Additionally, the Kirchhoff operators aperture is usually limited to a specified range of angles to reduce its cost and limit the number of migration artifacts. A number of tapers, including edge tapers and aperture tapers, are typically introduced to reduce the impact of boundaries on the Kirchhoff operator. Lastly, many Kirchhoff operators also include features to reduce the effects of operator aliasing, which is discussed in the next section.

Algorithm 1 Forward/adjoint Kirchhoff operator

```

procedure KIRCHHOFF(d, m, forward/adjoint, wav, sx, rx, dz, dx, dt)
  if adjoint then
    d = rhofilter(d, adjoint)
    d = convolution(d, wav, adjoint)
  end if
  for itrace = 1:Ntrace do
    for ix = 1:Nx do
      for iz = 1:Nz do
        w = weight(sx[itrace], rx[itrace], vel, ix, iz, dx, dz)
        t = travelttime(sx[itrace], rx[itrace], vel, ix, iz, dx, dz)
        it = round(t/dt) + 1
        if forward then
          d[it,itrace] += w*m[iz,ix]
        else if adjoint then
          m[iz,ix] += w*d[it,itrace]
        end if
      end for
    end for
  end for
  if forward then
    d = convolution(d, wav, forward)
    d = rhofilter(d, forward)
  end if
end procedure

```

2.6 Operator aliasing

The Kirchhoff integral associates each scatterer in the model domain with a diffraction in the data domain, as shown in figure 2.3. The forward Kirchhoff integral generates seismic data by spraying the reflectivity of each scatterer over a diffraction hyperbola. Likewise, the adjoint Kirchhoff integral computes the reflectivity of each scatterer by collapsing and summing the amplitudes along the corresponding diffraction. Since the forward and adjoint Kirchhoff integrals map model-domain scatterers to data-domain diffractions (and vice versa), the diffraction trajectory is sometimes referred to as the Kirchhoff operator. Within this section, any usage of the term “Kirchhoff operator” refers to a diffraction trajectory. Outside of this section, usage of the term “Kirchhoff operator” refers to the linear operator \mathbf{L} .

Our derivation of the forward and adjoint Kirchhoff integrals assumed the seismic data was collected using an infinite number of sources and receivers, therefore justifying an integral formulation. Under this assumption, the Kirchhoff operator is perfectly sampled, so aliasing is not a concern. However, real seismic data is collected using a limited number of sources and receivers. Therefore, the forward and adjoint Kirchhoff integrals get replaced with finite summations. Although replacing integration with summation is a relatively inconsequential change, it does introduce a few fundamental problems. Among them is the issue of operator aliasing. Operator aliasing occurs when the Kirchhoff operator cuts through an event at too steep an angle to adequately sample it, as shown in figure 2.4. As a result, the migration operator calculates a non-zero reflectivity for a scatterer that should have otherwise been assigned a value of zero, thus introducing an “alias” artifact into the image.

Aliasing is ultimately a sampling problem. As the sampling rate approaches infinity, discrete signals becomes asymptotically similar to analog signals and aliasing ceases to be a problem. Conversely, as the sampling gets sparser, discrete signals loses high-frequency information and aliasing becomes an increasingly important issue. This suggests that operator aliasing cannot be ignored if the array of receivers is sparsely sampled. In order to remain unaliased, each point on the Kirchhoff operator must satisfy the dip Nyquist criterion (Lumley et al., 1994; Abma et al., 1999):

$$f_k^{max} = \frac{1}{2\Delta t_k} = \frac{1}{2(\partial t_k / \partial \xi)\Delta \xi}, \quad (2.42)$$

where f_k^{max} is the maximum unaliased frequency that can supported by the operator at the k^{th} trace, Δt_k is the local operator moveout, $\partial t_k / \partial \xi$ is the local dip of the operator, and $\Delta \xi$ is the spacing between adjacent traces. Equation 2.42 indicates that the maximum unaliased frequency is inversely proportional to the slope of the operator. Therefore, gently

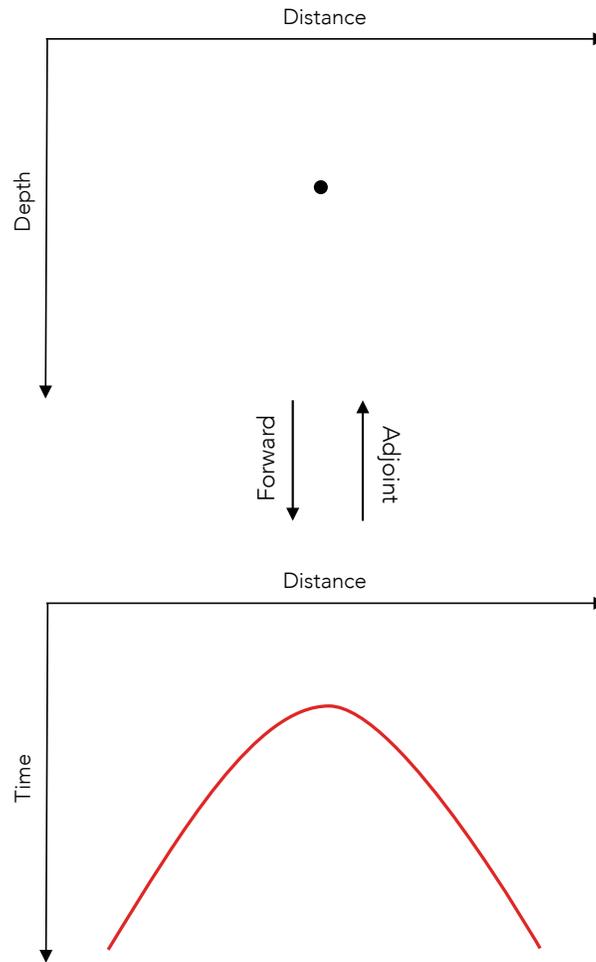


Figure 2.3: Illustration of the Kirchhoff operator (red curve). The forward Kirchhoff integral sprays each model-domain scatterer over a data-domain diffraction. The adjoint Kirchhoff integral collapses each data-domain diffraction into a model-domain scatterer. In practice, the adjoint operator recovers a blurred version of the true image because it is not orthogonal to the forward operator.

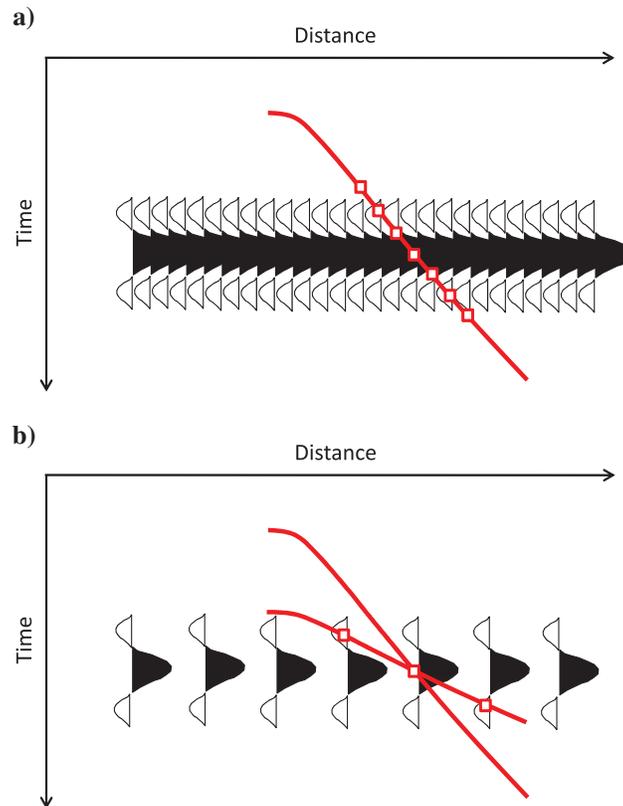


Figure 2.4: Kirchhoff migration operator (red curve) cutting through a seismic event. (a) The trace spacing is small, so the Kirchhoff operator intersects several traces. The sum of the samples along the operator is nearly zero, so an aliasing artifact is not produced. (b) The trace spacing is large. The steeper Kirchhoff operator intersects only a single trace, so the sum of the samples along the operator is non-zero. Therefore, the steeper Kirchhoff operator produces an aliasing artifact. The less steep Kirchhoff operator intersects a few traces, so the sum of the samples along the operator is nearly zero. Therefore, the less steep Kirchhoff operator does not produce an aliasing artifact. Figure from Gray (2013).

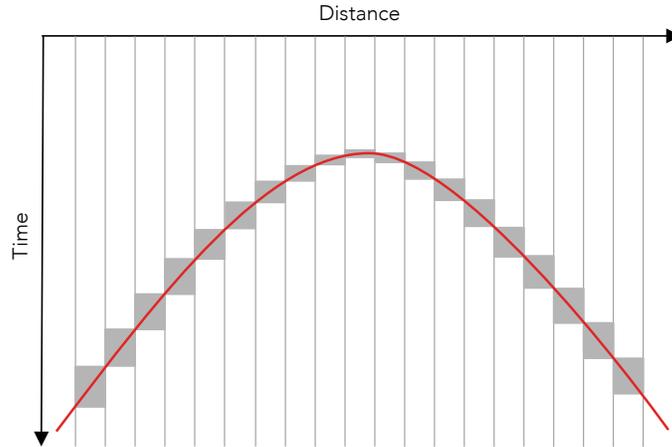


Figure 2.5: The amount of operator moveout (grey boxes) is proportional to the local dip of the Kirchhoff operator (red curve). The steepest parts of the Kirchhoff operator experience the greatest amount of moveout and are therefore the most susceptible to aliasing.

dipping Kirchhoff operators can support higher frequencies than steeply dipping Kirchhoff operators. Additionally, since the slope of the Kirchhoff operator varies from trace to trace, so too does the amount of operator moveout (figure 2.5). This suggests that the maximum unaliased frequency is different for each point on the Kirchhoff operator.

A number of approaches have been proposed for dealing with operator aliasing. The simplest solution is to apply a single lowpass filter to the entire data set, thereby suppressing the aliased energy. However, picking the high-cut frequency for the lowpass filter is problematic. Since the maximum unaliased frequency is different for each point on the Kirchhoff operator (and the Kirchhoff operator varies from scatterer to scatterer), the high-cut frequency must be set to the lowest possible value of f_k^{max} . Therefore, a considerable amount of high-frequency data gets suppressed that theoretically could have been used. A substantial improvement can be achieved by producing several different copies of the data, each with a different high-cut frequency (Gray, 1992). The Kirchhoff operator can then extract amplitudes from the data set whose high-cut frequency most closely matches the local value of f_k^{max} . This approach has been used extensively in industrial applications of seismic imaging. However, it is only feasible for migration methods that are based on a direct transformation. Since iterative seismic migration require an exact forward and adjoint, Gray's approach for anti-aliasing would require each data set to be forward modeled at each iteration. This would increase the cost of the algorithm and partially negate the benefits of using a Kirch-

hoff operator. Thus an alternative anti-aliasing technique is desired for iterative seismic migration.

An ideal anti-aliasing filter should be customized on-the-fly based on the local maximum unaliased frequency. The first step is to find an efficient way to compute f_k^{max} . For Kirchhoff time migration, an analytical expression of f_k^{max} can be derived in terms of the RMS velocity and the source-side & receiver-side traveltimes (Lumley et al., 1994). In contrast, the depth migration anti-aliasing criteria does not have an analytical expression because Kirchhoff operators in $c(\mathbf{x})$ media do not have simple hyperbolic forms. Instead, depth-domain Kirchhoff operators can estimate f_k^{max} by differencing the traveltimes on each side of the k^{th} trace:

$$f_k^{max} = \frac{1}{|t_{k+1} - t_{k-1}|}. \quad (2.43)$$

Given that f_k^{max} can be computed for each point on the Kirchhoff operator, all that remains is to design a lowpass filter than can be applied locally and customized on-the-fly. Lumley et al. (1994) proposed a clever method to efficiently implement a triangular anti-aliasing filter for this purpose. The filter can be expressed in the Z-domain as

$$g(Z) = \frac{-Z^{-\ell-1} + 2Z^0 - Z^{\ell+1}}{\alpha(1 - Z^1)(1 - Z^{-1})}, \quad (2.44)$$

where ℓ is the half-length of the triangular filter and $\alpha = (\ell + 1)^2$. The amplitude spectrum of the triangular filter is given by (Lumley et al., 1994)

$$\Gamma(\omega) = \frac{\sin^2(\omega\delta t(\ell + 1)/2)}{\sin^2(\omega\delta t/2)}, \quad (2.45)$$

and has spectral notches located at

$$f_n = \frac{\omega_n}{2\pi} = \frac{n}{(\ell + 1)\delta t}, \quad (2.46)$$

where δt is the time sampling interval for each receiver. As an example, figure 2.6 shows a plot of the normalized amplitude spectrum with $\ell = 10$ and $\delta t = 4$ milliseconds. Most of the filters energy is contained within the first spectral notch. Therefore, the optimal half-length of the triangular filter can be determined by choosing the local maximum unaliased frequency to correspond with the first spectral notch, i.e., $f_k^{max} = f_1$. Equation 2.46 then becomes

$$f_k^{max} = \frac{1}{(\ell + 1)\delta t}, \quad (2.47)$$

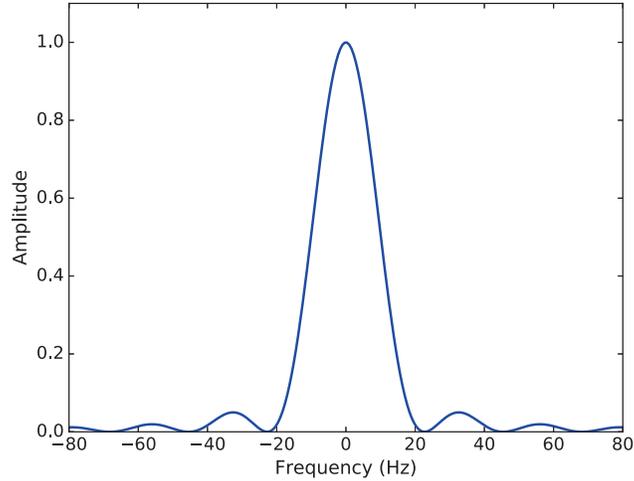


Figure 2.6: Amplitude spectrum for the triangular filter in equation 2.45 with $\ell = 10$ and $\delta t = 4$ milliseconds. Most of the triangular filter's energy is contained within the first spectral notch.

which can be rearranged to solve for the optimal half-length of the triangular anti-aliasing filter:

$$\ell = \frac{1}{f_k^{max} \delta t} - 1. \quad (2.48)$$

Finally, the $1/(1 - Z^1)$ and $1/(1 - Z^{-1})$ in equation 2.44 can be Taylor expanded as

$$\frac{1}{1 - Z^1} = Z^0 + Z^1 + Z^2 + Z^3 + Z^4 + Z^5 + \dots, \quad (2.49)$$

and

$$\frac{1}{1 - Z^{-1}} = -Z^{-1} - Z^{-2} - Z^{-3} - Z^{-4} - Z^{-5} - \dots, \quad (2.50)$$

which represent causal and acausal integration, respectively (Lumley et al., 1994). The numerator of equation 2.44 represents a gapped three-point Laplacian, where the gap length controls the length of the triangular filter and the high-cut frequency. The anti-aliasing filter can therefore be implemented by replacing the Dirac-delta shaped diffraction trajectory with a gapped-Laplacian shaped diffraction trajectory, followed (preceded) by causal and acausal integration for the forward (adjoint) operator, as shown in Algorithm 2. The structure for the aliased Kirchhoff operator (Algorithm 1) and the anti-aliased Kirchhoff operator (Algorithm 2) is nearly identical. Practical implementations of Algorithm 2 include all of the extra features mentioned at the end of section 2.5.

Algorithm 2 Anti-aliased forward/adjoint Kirchhoff operator

```

procedure KIRCHHOFF(d, m, forward/adjoint, wav, sx, rx, dz, dx, dt)
  if adjoint then
    d = rhofilter(d, adjoint)
    d = convolution(d, wav, adjoint)
    d = integrate_causal_acausal(d, adjoint)
  end if
  for itrace = 1:Ntrace do
    for ix = 1:Nx do
      for iz = 1:Nz do
        w = weight(sx[itrace], rx[itrace], vel, ix, iz, dx, dz)
        t = travelttime(sx[itrace], rx[itrace], vel, ix, iz, dx, dz)
        it = round(t/dt) + 1
         $\Delta t = \text{abs}(t_{itrace+1} - t_{itrace-1})/2$ 
         $f^{max} = \frac{1}{2\Delta t}$ 
         $\ell = \text{round}(\frac{1}{f^{max}dt}) - 1$ 
        if forward then
          d[it- $\ell$ ,itrace] -= w*m[iz,ix]
          d[it,itrace] += 2*w*m[iz,ix]
          d[it+ $\ell$ ,itrace] -= w*m[iz,ix]
        else if adjoint then
          m[iz,ix] += -w*d[it- $\ell$ ,itrace] + 2*w*d[it,itrace] - w*d[it+ $\ell$ ,itrace]
        end if
      end for
    end for
  end for
  if forward then
    d = integrate_causal_acausal(d, forward)
    d = convolution(d, wav, forward)
    d = rhofilter(d, forward)
  end if
end procedure

```

2.7 Examples

The following synthetic examples demonstrate the functionality of the forward and adjoint Kirchhoff operator. The examples use simple models with the intent of illustrating the Kirchhoff operators basic behavior, strengths, and limitations, rather than investigating its full capability. More complicated models are investigated when least-squares migration is discussed in chapter 4.

2.7.1 Example: The forward/adjoint Kirchhoff operator

This example demonstrates the Kirchhoff operators ability to map model-domain features into the data domain and vice versa. To begin, consider the simple anticline velocity model in figure 2.7(a). The forward Kirchhoff operator requires a reflectivity model as its input. The reflectivity should generally be computed based on the angle of incidence, density, and P and S velocities (Aki and Richards, 2002). In this example, we simplify the reflectivity formula by using a constant density and assuming the rays reflect at near-vertical incidence

$$R = \frac{v_2 - v_1}{v_2 + v_1},$$

where v_1 and v_2 are the velocities of the layers above and below each hypothetical interface, respectively. Using this approach, we obtain the reflectivity model in figure 2.7(b), which serves as the input for forward modeling via the Kirchhoff operator.

The Kirchhoff operator is highly flexible and can handle common-shot or common-offset data. In this example, the Kirchhoff operator works in the common-shot domain and thus produces a set of common shot gathers as its output. An example for a source located at $x=2500\text{m}$ is shown in figure 2.8. Comparing figure 2.7(b) with figure 2.8, it is clear that the Kirchhoff operator has modeled only the primary reflections. Care must therefore be taken when comparing real data with modeled data because the Kirchhoff operator ignores the direct wave and multiples. A closer inspection of figure 2.8 reveals a number of low-amplitude artifacts that do not correspond to features in the model. These artifacts arise from two main sources: (i) incomplete destructive interference, which is responsible for the low-amplitude diffraction “wings” underneath the first reflection, and (ii) edge effects, which are responsible for the spurious events protruding from the edge at approximately $t = 1.5\text{s}$. The former can typically be overcome by using a denser modeling grid, which provides additional scatterers to help cancel the remaining diffraction “wings”. The latter arises due to the presence of model boundaries, and can be partially mitigated by using a harsher edge-taper. In this example, the artifacts can be ignored because their amplitudes are relatively small compared with the reflections.

The adjoint Kirchhoff operator takes a seismic data set as its input, then outputs an image of the subsurface reflectivity. Figure 2.9 show the migrated image for the forward modeled data. Although the migrated image is structurally accurate, it incorrectly estimates the reflectivity of each interface and suffers from poor resolution. For these reasons, the adjoint is generally considered a poor approximation of the inverse and a more accurate solution is desired.

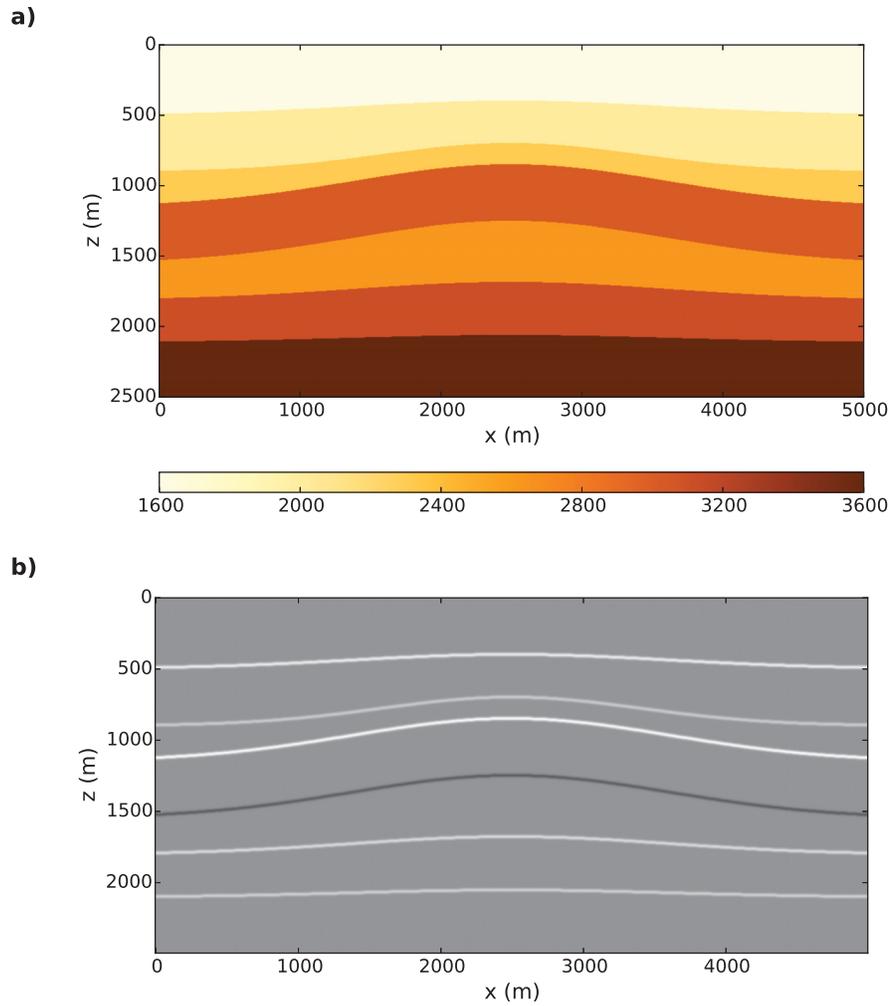


Figure 2.7: (a) Synthetic velocity model (m/s). (b) Reflectivity model for (a). The reflectivity is the same for each offset bin.

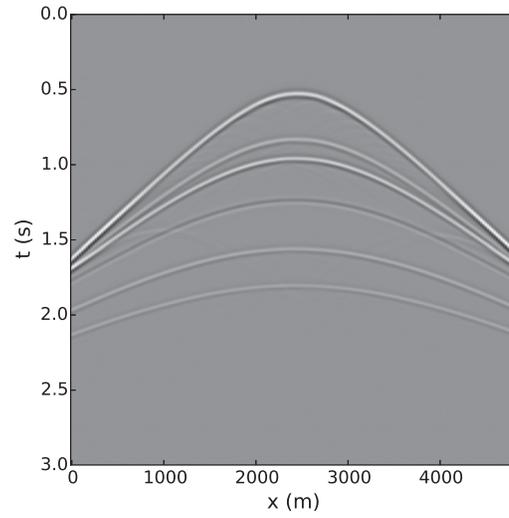


Figure 2.8: Forward modeled shot gather for a source located at $x=2500$ m for the model in figure 2.7.

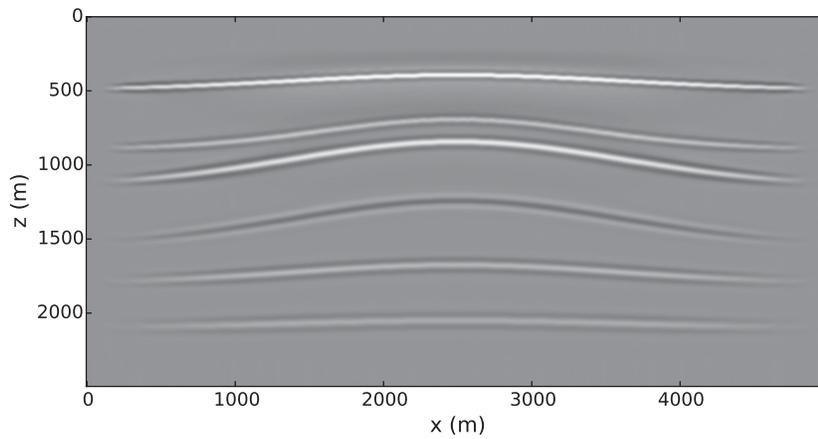


Figure 2.9: Stacked seismic image obtained by migration via the adjoint Kirchhoff operator.

2.7.2 Example: Kirchhoff operator aliasing

To illustrate the impact of aliasing on the Kirchhoff operator, consider the two-interface reflectivity model in figure 2.10(a). Operator aliasing is more prevalent for low-velocity models because traveltime is inversely proportional to velocity. Therefore, to emphasize the effect of operator aliasing, a constant velocity of 2000m/s was used to generate the zero-offset section in figure 2.10(b). The data consists of 50 receivers that were uniformly spaced over a distance of 2500m. Each receiver used a sampling interval of 4ms to ensure the data is adequately sampled in time. Thus, the spatial and temporal sampling of the data is sufficient to avoid data aliasing. Any artifacts appearing in the migrated image must therefore be a consequence of Kirchhoff operator aliasing.

In figure 2.10(c), the input data was migrated using the adjoint Kirchhoff operator without accounting for operator aliasing (Algorithm 1). Since the effect of aliasing was ignored, the Kirchhoff operator incorrectly migrated the high-frequency features in the data. This is responsible for the low-amplitude precursor artifacts that precede the real reflectors. This style of high-frequency artifact is well documented and is known to reduce the quality of the migrated seismic image (Biondi, 2001). Because these artifacts are caused by operator aliasing, it stands to reason that they can be suppressed by locally filtering the frequencies that violate equation 2.42. In figure 2.10(d), the seismic data was migrated using the anti-aliased adjoint Kirchhoff operator (Algorithm 2). Comparing figure 2.10(c) with figure 2.10(d), it is clear that the triangular anti-aliasing filter has suppressed the aliased energy, resulting in an image with fewer artifacts. If enough clipping is applied to figure 2.10(d), it would be possible to see low-amplitude remnants of the aliasing artifacts. This is a consequence of the fact that the triangular anti-aliasing filter does not completely suppress the frequencies above f_k^{max} . Moreover, the triangular filter has non-zero spectral side-lobes that extend beyond f_k^{max} (figure 2.6). However, the presence of the spectral side-lobes clearly has a negligible effect on the migrated image.

As expected, the triangular anti-aliasing filter marginally increased the computational cost of the forward/adjoint Kirchhoff operator, resulting in only a 2% increase in run time. This increase is probably comparable with (or even less than) the cost of alternative anti-aliasing filters, which have additional input/output costs associated with accessing multiple copies of the data. Additionally, since the triangular filter was implemented on-the-fly, the anti-aliased Kirchhoff operator had the same computer memory requirements as the aliased Kirchhoff operator. This is a considerable advantage over alternative anti-aliasing filters, which require extra computer memory for each copy of the data. These observations suggest that the triangular anti-aliasing filter is well suited for iterative seismic migration.

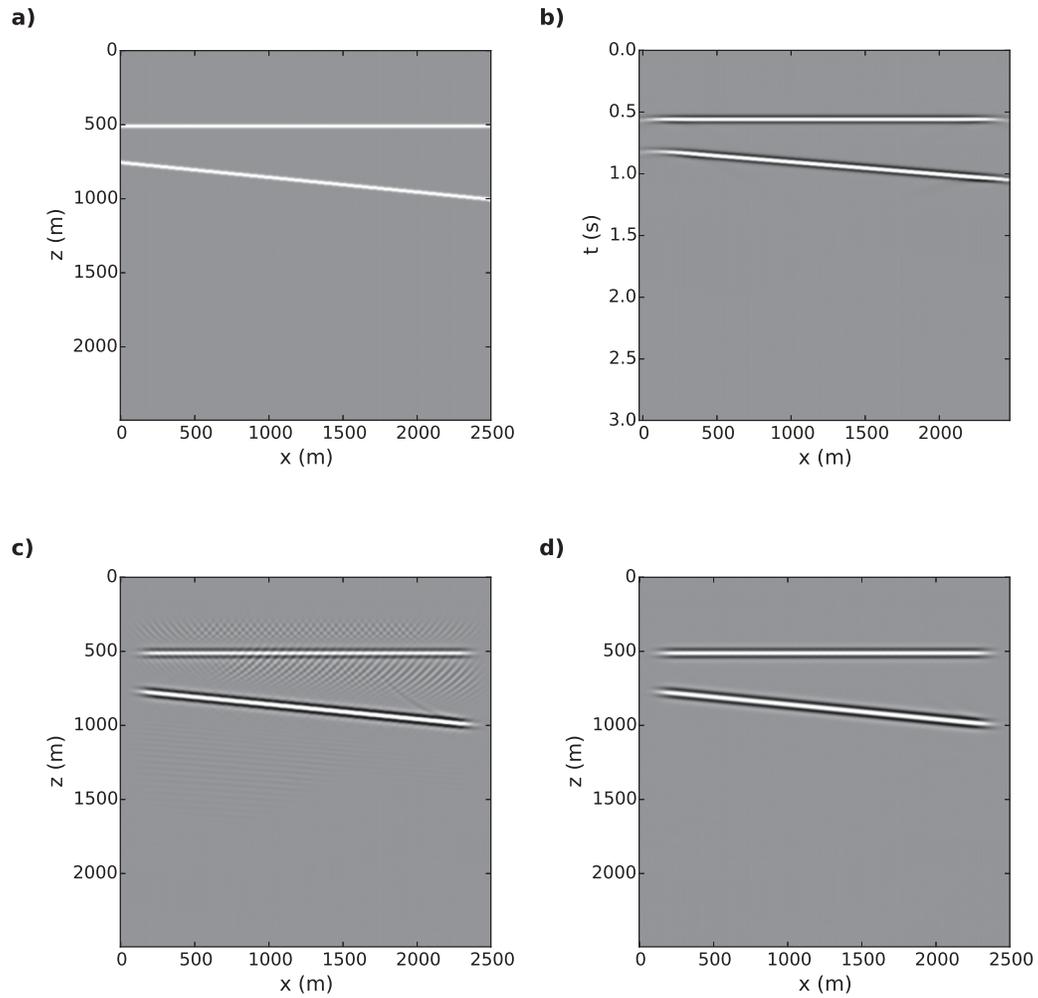


Figure 2.10: (a) Synthetic reflectivity model. (b) Zero-offset seismic section produced using the forward Kirchhoff operator \mathbf{L} . (c) Migrated seismic image using the aliased adjoint Kirchhoff operator \mathbf{L}^\dagger based on Algorithm 1. Migrated seismic image using the anti-aliased adjoint Kirchhoff operator \mathbf{L}^\dagger based on Algorithm 2.

2.8 Summary

Seismic migration algorithms are based on the wave equation. Solving the full elastic wave equation is generally too expensive for practical purposes. A less expensive alternative is to approximate the Earth as a fluid (acoustic approximation), followed by linearization via the high-frequency assumption and the Kirchhoff approximation. The resulting forward modeling operator is capable of calculating the primary reflections for a known reflectivity model. Its adjoint transforms a primaries-only data set into a blurred image of the subsurface reflectivity.

Operator aliasing is an issue that arises due to the use of a finite number of receivers. Aliasing has the greatest effect on the steepest dipping parts of the Kirchhoff operator, where the operator moveout is the largest. Conventional anti-aliasing techniques that involve lowpass filtering the data are unpractical for iterative seismic migration. Instead, an efficient triangular anti-aliasing filter can be implemented and customized on-the-fly by replacing the Dirac-delta shaped migration operator with a gapped-Laplacian shaped operator, followed (preceded) by causal and acausal integration for the forward (adjoint) operator.

CHAPTER 3

The wavefront construction method

3.1 Introduction

The Kirchhoff operator integrates over the model to accumulate the Earth's impulse response. For each scattering point in the model, a source-side Green's function is used to simulate the downward propagation of the source and a receiver-side Green's function is used to simulate the upward propagation of the scattered wavefield. If either of the Green's functions contain errors, the Kirchhoff operator will incorrectly propagate the seismic waves, resulting in erroneous seismic data (for the forward operator) or an incorrect seismic image (for the adjoint operator). Therefore, the forward and adjoint Kirchhoff operators require an accurate method for estimating the Green's functions for seismic waves.

The traveltime and amplitude components of the Green's functions can be estimated by solving the eikonal and transport equations, respectively. There are two classes of solvers for the eikonal and transport equations: grid-based and ray-based. Grid-based methods aim to find a direct solution to the eikonal and transport equations using numerical techniques such as the finite difference method (Vidale, 1988) or the fast marching method (Sethian, 1996). The motivation for using a grid-based method is obvious: they are robust and can handle even the most structurally complex velocity models. However, the pitfalls of grid-based methods generally outweigh its strengths. For example, their accuracy directly depends on the grid-spacing. Therefore, highly accurate grid-based solvers come with a substantial computational cost, especially for three dimensional models. Additionally, most practical implementations of grid-based methods are only capable of estimating first arrivals, which may be insufficient when the subsurface is structurally complex (Operto et al., 2000). Lastly, grid-based methods require separate solvers for the traveltime and amplitude com-

ponents; thus, the two solutions may be inconsistent with each other. Instead, the Green's functions are typically estimated using ray tracing (Dines and Lytle, 1979; Červený, 1987; Lomax, 1994; Vinje et al., 1993). Ray tracing seeks a solution to the eikonal equation by solving a set of coupled ordinary differential equations (ODEs). Unlike grid-based methods, ray-based methods can estimate amplitudes without having to solve the transport equation. In addition to this, most implementations of ray tracing are capable of computing later arrivals. Thus, most practical implementations of Kirchhoff migration rely on ray tracing to estimate Green's functions.

In this chapter, I review the wavefront construction method (WCM) for estimating Green's functions. WCM is similar to conventional ray-based methods in that it solves a set of coupled ODEs to propagate rays forward in time. However, instead of propagating each ray independently, WCM propagates each ray simultaneously by reconstructing the wavefront at each time step. Moreover, WCM utilizes ray tracing principles to propagate the wavefront itself, rather than individual rays. As a result, it resolves the ray multipathing and ray shadow zone problems, thereby improving the accuracy and stability of the interpolated traveltimes. Additionally, the relative divergence/convergence of each point on the wavefront can be compared at successive time steps to obtain an accurate estimate of the geometrical spreading factor. This is an advantage over grid based methods, which require the traveltimes and geometrical spreading factor to be estimated independently. In summary, WCM has the same robustness as grid-based methods, while retaining the benefits of ray tracing, including its accuracy and ability to estimate amplitudes (geometrical spreading factors) and later arrivals. WCM is therefore an ideal choice for estimating the Green's functions used in our Kirchhoff operator.

3.2 Ray tracing

Most implementations of the Kirchhoff forward/adjoint operator utilize ray tracing to estimate the source-side and receiver-side Green's functions. In this section, the method of characteristics is used to derive the ray tracing ODEs from the eikonal equation. Following this, a naive technique is presented for estimating the amplitude of the Green's functions based on the ray tracing solution.

3.2.1 Background velocity model

The forward & adjoint Kirchhoff integrals and the eikonal & transport equations are only valid under the high-frequency assumption. This suggests that the migration velocity model

should contain only low-wavenumber features; moreover, the high-wavenumber perturbations get decoupled from the velocity model and are instead contained within the reflectivity model, where they can eventually be inverted for (Aki and Richards, 2002). Unfortunately, there isn't an absolute rule or guideline for distinguishing between low-wavenumber and high-wavenumber features. Instead, low and high wavenumber features are distinguished by comparing the length-scale of the feature with the frequency of the signal. Therefore, features that are considered "low wavenumber" at one frequency may be considered "high wavenumber" at another frequency.

As Bleistein et al. (2001) noted, one way to distinguish between low and high wavenumber features is to use the Rayleigh criterion, which states that two features can be "just resolved" so long as the maximum (i.e. the peak) of one feature corresponds with the zero crossing of the other (Born and Wolf, 1999). Using this criterion, a feature in the model is considered to be large if its length-scale L satisfies

$$L \geq \frac{\Lambda}{4}, \quad (3.1)$$

where Λ is the dominant wavelength in the wavefield. If the seismic waves travel at a background speed of c_0 , the Rayleigh criterion can be rewritten as

$$1 \leq \frac{4L}{\Lambda} = \frac{4Lf}{c_0}, \quad (3.2)$$

where $c_0 = \Lambda f$, and f is the dominant frequency in the wavefield. If we define the dimensionless parameter ζ as

$$\zeta = \frac{4Lf}{c_0}, \quad (3.3)$$

then the Rayleigh condition for identifying large features becomes (Bleistein et al., 2001)

$$\zeta \geq 1. \quad (3.4)$$

If the frequency f of the seismic data and the length-scale L of the features in the velocity model satisfy equation 3.4, then the data and model are said to satisfy the high-frequency assumption.

When seismic data is collected, the frequency bandwidth of the data is known and the average velocity can be estimated. Therefore, equation 3.4 determines the length-scale of the features that can exist in the velocity model while satisfying the high-frequency assumption. This length-scale can also be interpreted as the size of the blurring kernel that should be used to smooth the velocity model. As an example, consider a typical seismic data set

in which the signal falls primarily between 5 Hz and 100 Hz. For convenience, let's assume the average velocity is $c_0 = 2000$ m/s. When $f = 5$ Hz, features in the velocity model must vary on a length scale of at least 100 m to satisfy the high-frequency assumption. On the other hand, when $f = 100$ Hz, features in the velocity model must vary on a length scale of at least 5 m to satisfy the high-frequency assumption. As expected, the higher frequencies are capable of supporting features that vary on a shorter length-scale. This suggests that the lowest frequency in the signal ultimately determines the length-scale over which features in the velocity model should vary. In other words, the lowest frequency in the signal determines the optimal size L of the blurring kernel that should be used to smooth the velocity model. For our example, the velocity model should be smoothed using a blurring kernel with a diameter of approximately $L = 100$ m. Based on my experience, I recommend using a normalized Gaussian blurring kernel that is truncated at three standard deviations. However, any diameter L blurring kernel should suffice.

Although the high-frequency criterion in equation 3.4 is a nice theoretical development, it represents only one interpretation of the high-frequency assumption and should not be considered an absolute rule that must be followed. In practice, it is generally sufficient to smooth the velocity model just enough so that sharp variations are not obviously visible. In fact, Gray (2000) found that it is generally better to apply less smoothing rather than too much; as the size of the blurring kernel becomes grossly large, the velocity model loses all geological plausibility, and the migrated image suffers as a result.

If less smoothing is better than more smoothing, then the question is raised: what happens if the high-frequency assumption is violated? Bleistein et al. (2001) pointed out that the frequency content of the data primarily impacts the resolution of the image. If the data contains only low frequencies, the migration operator becomes incapable of resolving sharp reflectors. As a result, the migrated image contains blurred reflectors rather than sharp ones. As we will find out in chapter 4, migration via the pseudoinverse (least-squares migration) helps to boost the resolution of the migrated image, and therefore allows us to partially violate the high-frequency assumption.

3.2.2 The ray tracing equations

The traveltimes component of the Green's functions can be estimated by solving the eikonal equation. Although the eikonal equation can be solved directly, it is generally advantageous to recast the problem as a coupled system of ordinary differential equations via the method of characteristics (Kravtsov and Orlov, 1990; Bleistein et al., 2001). The resulting system

of ordinary differential equations is given by

$$\begin{aligned}\frac{d\mathbf{x}}{d\tau} &= c^2(\mathbf{x})\mathbf{p}, \\ \frac{d\mathbf{p}}{d\tau} &= -\frac{\nabla c(\mathbf{x})}{c(\mathbf{x})},\end{aligned}\tag{3.5}$$

where $\tau(\mathbf{x})$ is the travelttime (phase) and $\mathbf{p}(\mathbf{x})$ is the slowness vector defined by

$$\mathbf{p}(\mathbf{x}) = \nabla\tau(\mathbf{x}) = \frac{1}{c(\mathbf{x})} \left[\hat{i} \sin(\theta)\cos(\phi) + \hat{j} \sin(\theta)\sin(\phi) + \hat{k} \cos(\theta) \right],\tag{3.6}$$

where θ is the inclination angle with respect to vertical and ϕ is the azimuthal angle. If the rays are propagated using a small step in time & space, each segment of the raypath is approximately linear. Thus, θ and ϕ are nearly constant for each step, which means equation 3.5 can be further simplified by replacing $\frac{\partial\mathbf{p}}{\partial\tau}$ with $\frac{\partial\theta}{\partial\tau}$ and $\frac{\partial\phi}{\partial\tau}$. The resulting system of equations is given by

$$\begin{aligned}\frac{\partial x_1}{\partial t} &= c(\mathbf{x})\sin(\theta)\cos(\phi), \\ \frac{\partial x_2}{\partial t} &= c(\mathbf{x})\sin(\theta)\sin(\phi), \\ \frac{\partial x_3}{\partial t} &= c(\mathbf{x})\cos(\theta), \\ \frac{\partial\theta}{\partial t} &= -\cos(\theta) \left[\cos(\phi)\frac{\partial c(\mathbf{x})}{\partial x_1} + \sin(\phi)\frac{\partial c(\mathbf{x})}{\partial x_2} \right] + \sin(\theta)\frac{\partial c(\mathbf{x})}{\partial x_3}, \\ \frac{\partial\phi}{\partial t} &= \frac{1}{\sin(\theta)} \left[\sin(\phi)\frac{\partial c(\mathbf{x})}{\partial x_1} - \cos(\phi)\frac{\partial c(\mathbf{x})}{\partial x_2} \right].\end{aligned}\tag{3.7}$$

The ray tracing equations can be further simplified for two dimensional models by choosing $\phi = 0$, which gives

$$\begin{aligned}\frac{\partial x_1}{\partial t} &= c(\mathbf{x})\sin(\theta), \\ \frac{\partial x_3}{\partial t} &= c(\mathbf{x})\cos(\theta), \\ \frac{\partial\theta}{\partial t} &= -\cos(\theta)\frac{\partial c(\mathbf{x})}{\partial x_1} + \sin(\theta)\frac{\partial c(\mathbf{x})}{\partial x_3}.\end{aligned}\tag{3.8}$$

Since the ray tracing ODEs form a coupled system, x_1 , x_2 , x_3 , θ , and ϕ must be updated simultaneously. An analytical solution usually cannot be found for most 2D and 3D velocity models. Therefore, a numerical approach must be used instead. Several numerical methods exist for solving systems of ODEs. The most straightforward approach is the Euler

method, which replaces derivatives with finite differences. The Euler method is intuitive to understand, straightforward to implement, and inexpensive to solve. However, the Euler method's error is proportional to the square root of the step size; thus the Euler method is only accurate to the first order. Seismic migration requires precise traveltimes in order to obtain a high-quality image. Therefore, the Euler method's error is too large for this task. A more accurate solution can be obtained using higher-order solvers such as the Runge-Kutta method (Franklin, 2013). The Runge-Kutta method solves each ODE by evaluating each derivative as a weighted sum of points along the step length. Since more information is utilized at each step, the Runge-Kutta method can achieve much higher accuracy. The classical version of the Runge-Kutta method utilizes information from four points along each step and is accurate to the fourth order. Appendix B provides a review of the fourth order Runge-Kutta method.

Ray tracing provides a means for numerically integrating individual rays forward in time. For a single ray, solving the ray tracing equations yields the location of the ray as a function of time. If several rays - each with a different takeoff angle - are traced outward away from the source, then the traveltimes in the subsurface can be estimated by interpolating between the raypaths. This process can be repeated for each source and receiver in the data to estimate the traveltime component of the seismic Green's functions. An improved estimate of the traveltimes can be obtained by utilizing strategies such as the shooting method (Červený et al., 1977), although this greatly increases the computational cost.

So far, ray tracing has been used to solve for the traveltime component of the Green's functions; next, we solve for the amplitude component. As shown in Appendix A, the amplitude component of the Green's functions is given by the solution to the transport equation. However, solving the transport equation is an added computational cost that ideally should be avoided. Instead, it is common to estimate the amplitude of the Green's functions based on the solution to the ray tracing equations. A naive estimate of the amplitude can be obtained based on the length of each raypath. Moreover, for a constant velocity model, the amplitude (geometrical spreading) factor is given by

$$A = \frac{1}{R} \tag{3.9}$$

for 3D models and

$$A = \sqrt{\frac{1}{R}} \tag{3.10}$$

for 2D models, where R is the length of the raypath. Thus, for relatively simple geological models (e.g. horizontally layered strata), the amplitude factor can be estimated using equation 3.9 or 3.10, where R is computed by summing the length of each segment of the

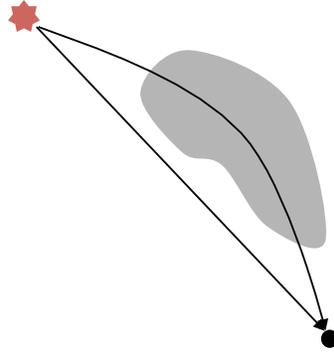


Figure 3.1: Rays with different takeoff angles can follow a different path to reach the same scatterer. The curved ray travels a longer distance at a slower velocity, and thus requires a larger traveltime to reach the scatterer. The red star denotes the source and the grey shape represents a low velocity zone.

ray path. The amplitudes can then be interpolated between the raypaths using the same strategy that was discussed for the traveltimes.

Ray tracing is an inexpensive alternative to solving the eikonal and transport equations. However, it suffers from a number of fundamental issues. First, for complex geological models, ray paths with different takeoff angles may eventually intersect/cross each other, as shown in figure 3.1; this phenomenon is typically referred to as ray multipathing. Because the rays traveled a different path through the model, they will generally intersect at different times. Moreover, one ray will arrive before the other, which causes difficulties when interpolating the traveltimes (Operto et al., 2000). Which traveltime should be used? The traveltime for the first ray? The traveltime for the second ray? Or an average of the two? Conventional ray-tracing has no way to deal with this problem because each ray is propagated independently without any knowledge of the other rays; thus, instabilities and inaccuracies will be introduced into the interpolated traveltimes. Second, high-velocity caustics can cause shadow zones which the raypaths are incapable of penetrating. Therefore, conventional ray tracing methods cannot estimate the traveltimes for these parts of the velocity model. Third, the ray coverage becomes increasingly sparse as the rays propagate farther away from the takeoff location. Therefore, the traveltimes must be interpolated using a decreasing amount of raypath information, which can produce additional interpolation errors and ultimately decreases the accuracy of the interpolated traveltimes. All of these issues can be address by using a modified version of ray tracing called the wavefront construction method.

3.3 Wavefront construction

The conventional method for ray tracing involves shooting a fan of rays outward from the source by propagating each ray independently. The ray fan is then used to interpolate the traveltimes and amplitudes of each scatterer in the subsurface. This approach generally produces poor results when the geology is structurally complex, as discussed in the previous section. A substantial improvement can be achieved by using the ray tracing equations to propagate the wavefront itself, rather than individual rays; this approach is typically referred to as wavefront construction (Vinje et al., 1993). The wavefront construction method is capable of monitoring the raypath density on-the-fly; thus, additional rays can be interpolated as needed. This ensures that the ray density remains approximately the same throughout the simulation, and also eliminates the issues posed by ray shadow zones. In addition to this, the wavefront construction method resolves the multipathing problem by including a criterion that automatically determines which rays should be kept and which ones should be terminated. The most common strategy is to keep the ray that arrives first and terminate all later arrivals; this is typically referred to as “first-arrival mode”. However, an argument can be made that the ray with the greatest amplitude should be kept instead (Vinje et al., 1993). Unless the model is characterized by extreme structural complexity, the first arrival will typically correspond with the highest amplitude. Therefore, choosing between first-arrival mode and highest-amplitude mode is relatively inconsequential.

3.3.1 Wavefront propagation

Consider the wavefront given by the solid line in figure 3.2. The wavefront consists of 8 rays that are approximately evenly spaced; these 8 rays serve as the starting point for the next iteration of ray tracing. If each of the rays is traced forward, the wavefront will eventually reach the dashed line. The distance DS between adjacent rays on the wavefront are then compared with a predefined maximum distance DS_{\max} . If the distance between adjacent rays exceeds DS_{\max} , an additional ray is interpolated on the wavefront. In figure 3.2, the distance between each ray on the new wavefront exceeds DS_{\max} , so new rays are interpolated. For each new ray, the following parameters must be interpolated:

1. The position of the new ray,
2. The inclination angle of the new ray,
3. The azimuthal angle of the new ray, and
4. The amplitude (geometrical spreading factor) of the new ray.

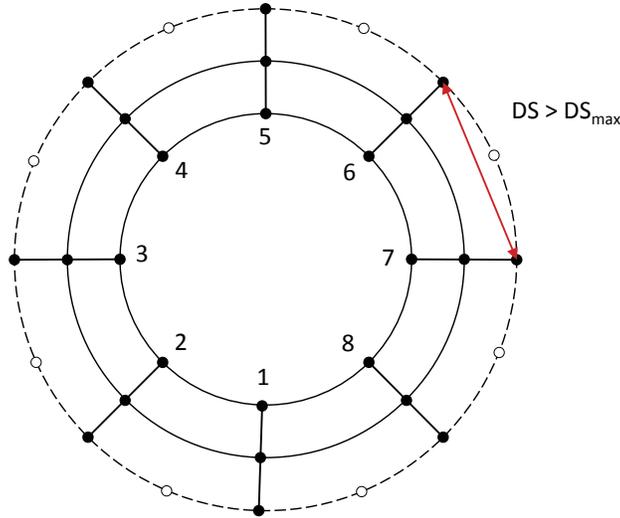


Figure 3.2: A new wavefront (dashed line) is constructed by tracing the existing rays (black circles) forward one time step. Additional rays (white circles) are interpolated when the distance between rays (red line) is greater than the threshold DS_{max} .

Note that the time does not need to be interpolated because the wavefront is defined as a constant time surface. Thus, the interpolated ray should be assigned the same time as its neighbors. Vinje et al. (1993) provided a detailed review of the ray-interpolation procedure. By interpolating the rays using this strategy, the raypath density should remain approximately constant throughout the simulation, regardless of whether the wavefield is diverging or converging. This approach is particularly beneficial when dense ray coverage is needed in the deeper parts of the model.

So far, a condition has been introduced for determining when new rays should be interpolated. Just as importantly, conditions must be put in place to determine when rays should no longer be traced (i.e. terminated). Generally, there are two cases in which a ray should be terminated:

1. The wavefront crosses itself (i.e. rays intersect), or
2. Part of the wavefront crosses the model boundary.

Wavefront self-crossings occur when two or more parts of the wavefront travel different paths through the model to reach the same point. Since each part of the wavefront travels a different distance at a different velocity, one part of the wavefront will arrive before the

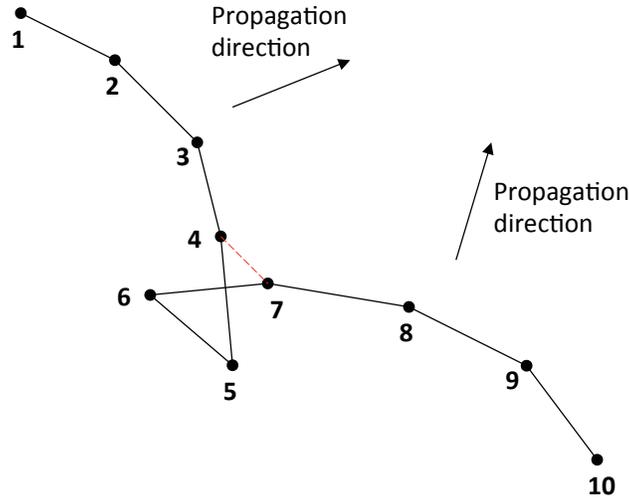


Figure 3.3: The new wavefront crosses itself as it propagates across a low-velocity caustic. Only the rays corresponding to first arrivals are kept when using the first arrival mode. Thus, the part of the wavefield behind the crossing (point number 5 and 6) is removed. An imaginary link (red dashed line) is created to fill in the gap; however, no new rays are interpolated in this gap.

others. Figure 3.1 illustrates this phenomenon from a ray-based perspective; since each ray is propagated without any information about its neighboring rays, it is impossible to properly address this issue. Figure 3.3 illustrates the same phenomenon, but from a wavefront construction perspective; since the full wavefront is propagated simultaneously, it is possible to determine which rays arrived first based on the geometry of the wavefront. In particular, the part of the wavefront behind the crossing (point no. 5 and 6 in figure 3.3) corresponds to a secondary arrival and should therefore be removed. When the self-crossing is removed, the wavefront gets divided into two parts that should each be treated as a separate wavefront. This means the wavefront will eventually be broken up into several smaller wavefronts if the velocity model is structurally complex. These smaller independent wavefronts may eventually interact and cross each other; when this happens, it should be dealt with using the same strategy as before. The ray-selection criterion illustrated in figure 3.3 is based on the first-arrival mode. Alternatively, highest-amplitude mode can be implemented by keeping the part of the wavefront with the greatest amplitude. Most implementations of the wavefront construction method (including the one used in this thesis) use the first-arrival mode due to its simplicity and low computational cost.

The second case in which rays are terminated occurs when part of the wavefront propa-

gates beyond the boundary of the model. Since traveltime and amplitude information is not needed outside of the modeling region, any rays that cross the model boundary should be terminated. However, since the wavefront construction method requires each ray to have information about its neighbors (e.g. for identifying wavefront self-crossings), it may be necessary to propagate each ray one or two steps beyond the model boundary.

Upon completion of wavefront construction, the modeling region will be covered by a series of wavefronts that sample the velocity model in a spider web-like mesh. Each piece of the mesh will have a width of approximately DS and a length of approximately $c(\mathbf{x})\Delta\tau$. In areas where the wavefront is diverging, the size of the mesh will gradually increase until $DS > DS_{\max}$, at which point a new ray will be interpolated and the size of the mesh will be halved. In areas where the wavefront is converging on itself, the size of the mesh will gradually decrease until the wavefront crosses itself, at which point some of the rays will be terminated. Evidently, the size of the mesh is controlled by DS_{\max} and $\Delta\tau$. Smaller mesh are generally preferred because it increases the sensitivity of the wavefront construction method to local variations in the velocity. However, decreasing the size of the mesh results in an increase in computational cost. Therefore, a trade-off must be made between having a small mesh and a low computational cost. Generally speaking, more complicated velocity models require a smaller mesh and thus have a greater computational cost.

Successive iterations of wavefront construction shows how the wavefield propagates as a function of time; thus, the wavefront mesh provides information about the traveltime component of the seismic Green's functions. In the following paragraphs, it will be shown that the wavefront mesh can also be used to obtain information about the amplitude component of the Green's functions.

Consider the three rays in figure 3.4. As previously discussed, the amplitude (geometrical spreading factor) for a homogeneous 2D velocity model can be estimated from the length R of the raypath:

$$A = \sqrt{\frac{1}{R}}. \quad (3.11)$$

Equation 3.11 becomes inaccurate for most heterogeneous media. In order to obtain an accurate estimate of the amplitude, equation 3.11 must be rewritten in terms of the relative divergence/convergence of the wavefront. This can be accomplished by utilizing the following geometrical relationship:

$$L = R\nu, \quad (3.12)$$

where $L = L_{1 \rightarrow 2} + L_{2 \rightarrow 3}$ is the sum of the arc lengths between rays 1 & 2 and rays 2 & 3, and ν is the angle between rays 1 & 3. Subbing equation 3.12 into equation 3.11 yields a

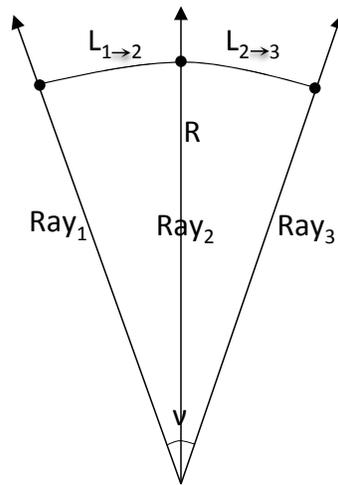


Figure 3.4: Geometrical relationship between the arc length of three rays and the opening angle.

new expression for the amplitude factor:

$$A = \sqrt{\frac{\nu}{L}}. \quad (3.13)$$

For most heterogeneous media, ν is approximately constant from one iteration to the next so long as the time step is small. Thus, for the i^{th} ray on the wavefront, the ratio of amplitudes at iterations $j + 1$ and j is given by

$$\frac{A_{i,j+1}}{A_{i,j}} = \frac{\sqrt{\nu/L_{i,j+1}}}{\sqrt{\nu/L_{i,j}}} = \sqrt{\frac{L_{i,j}}{L_{i,j+1}}}, \quad (3.14)$$

where $A_{i,j}$ is the amplitude of the i^{th} ray at the j^{th} iteration of wavefront construction, and $L_{i,j}$ is the arc length between ray $i - 1$ and ray $i + 1$. Estimating the arc length precisely requires fitting a polynomial to the wavefront. However, Kirchhoff migration is less sensitive to amplitude errors than traveltimes; thus, it is generally sufficient to use a straight-line approximation to estimate the arc length between neighboring rays. Rearranging equation 3.14 yields a recursive expression for updating the amplitude factor:

$$A_{i,j+1} = A_{i,j} \sqrt{\frac{L_{i,j}}{L_{i,j+1}}}. \quad (3.15)$$

The ratio $\frac{L_{i,j}}{L_{i,j+1}}$ is a measurement of the relative divergence/convergence of the i^{th} ray from one iteration to the next. Thus, equation 3.15 measures the amplitude factor based on the local behavior of the wavefront; as a result, it is accurate for both homogeneous and heterogeneous media. Parts of the wavefront that are rapidly diverging will experience the greatest decrease in amplitude, while parts of the wavefront that are rapidly converging will experience the greatest increase in amplitude. Equation 3.15 is also valid for plane waves, which neither converge nor diverge and are thus characterized by a constant amplitude. A similar expression can be derived for 3D velocity models using the same approach.

3.3.2 Parameter estimation

Each time the wavefront is stepped forward, a number of intermediate steps must be completed. Four of those steps have already been reviewed: (i) removing wavefront self-crossings, (ii) terminating rays that cross the model boundary, (iii) interpolating additional rays, and (iv) estimating the amplitude of each ray. Generally, the computed wavefronts will not coincide with the exact location of the scatterers. Thus, the last intermediate step is to interpolate the wavefront properties at the location of the scatterers. In the case of Kirchhoff

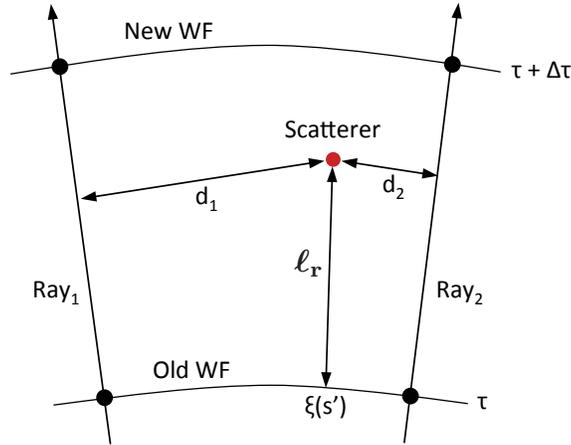


Figure 3.5: Traveltimes are interpolated by propagating the wavefront forward within each piece of the mesh.

migration, the following properties must be interpolated:

1. The traveltime of the wavefront at each scatterer in the model,
2. The amplitude of the wavefront at each scatterer in the model, and
3. The incidence angle of the wavefront at each scatterer in the model.

A lookup table containing this information must be generated for each source and receiver. Altogether, the lookup tables contain all of the information required to compute the source-side and receiver-side Green's functions as well as the angle-dependent obliquity factor in the forward and adjoint Kirchhoff operators.

Traveltime errors have a greater impact on Kirchhoff migration than amplitude or angle errors. Thus, the amplitudes and incidence angles can be interpolated using conventional interpolation schemes such as cubic splines. The traveltimes could also be interpolated using this strategy; however, due to the high sensitivity of Kirchhoff migration to traveltime errors, it is generally better to interpolate the traveltimes using a method that honors the wavefront propagation.

Consider figure 3.5, which shows a typical piece of the wavefront mesh. This particular piece of mesh contains a known scatterer inside of it; thus, the traveltime must be interpolated at the location of the scatterer. Since the traveltimes τ and $\tau + \Delta\tau$ are known, the traveltime at the scatterer can be obtained by propagating the wavefront forward within the

piece of mesh. Vinje et al. (1993) proposed to accomplish this by estimating the following values:

1. The perpendicular distances (d_1 and d_2) from the two rays to the scattering point,
2. The normalized distance s' along the wavefront: $s' = d_1/(d_1 + d_2)$; and
3. The distance ℓ_r from $\xi(s')$ to the scattering point.

The position $\xi(s')$ can be estimated by fitting a low-order polynomial (e.g. a parabola) to the wavefront. Vinje et al. (1993) then reasoned that the traveltimes at the location of the scatterer could be computed by propagating the wavefront forward within the mesh:

$$\tau_{scatterer} = \tau + \frac{\ell_r}{v_{mid}}, \quad (3.16)$$

where τ is the traveltimes of the old wavefront and v_{mid} is the velocity in the middle of the piece of mesh. Evidently, the accuracy of this approach depends on the size of the mesh. As each piece of the mesh gets smaller, v_{mid} becomes a better approximation of the average velocity and thus the traveltimes is interpolated more accurately. In my experience, interpolating the traveltimes using cubic splines results in roughly the same accuracy as the method proposed by Vinje et al. (1993). However, Vinje et al. (1993) method has an intuitive physical basis which may result in more accurate traveltimes when the velocity model is structurally complex. Algorithm 3 summarizes the wavefront construction method.

Algorithm 3 Wavefront Construction

```

procedure WAVEFRONT_CONSTRUCTION(vel, x0, z0,  $\theta_0$ , dx, dz, dt, DSmax)
  for each source and receiver do
    initialize ray parameters
    while (number of points on wavefront) > 3 do
      propagate the wavefront
      compute amplitudes along the wavefront
      remove wavefront self-crossings
      terminate rays that go out of bounds
      interpolate new rays
      interpolate incidence angles, amplitudes, and traveltimes
    end while
  end for
end procedure

```

3.4 Examples

The following examples demonstrate the wavefront construction method. The first example illustrates how the wavefront construction method is used to estimate the traveltime and amplitude components of the seismic Green's functions. The second example investigates the accuracy of the wavefront construction method compared with an analytical solution.

3.4.1 Example: SAIG velocity model

Consider the velocity model in figure 3.6(a). The model contains sharp discontinuities and is therefore unsuitable for ray tracing. Moreover, ray-based methods require a smooth velocity model to satisfy the high-frequency assumption. Thus, a Gaussian blurring kernel was used to generate the smooth velocity model in figure 3.6(b). Importantly, the blurring operator was applied to the slowness model (rather than the velocity model) in order to preserve the kinematics of the ray tracing equations.

Figure 3.7(a) shows the wavefront mesh for a source located at $x=1500\text{m}$ with $DS_{\max} = 50\text{m}$ and 50 initial rays; an additional 443 rays were interpolated throughout the simulation, though only a sparse selection of the mesh is shown because the full mesh is too dense to plot. At the beginning of the simulation, the wavefront propagated through a relatively simple part of the velocity model; thus, the wavefront mesh exhibits straightforward behavior near the source. In particular, the rays diverged uniformly, resulting in a semi-circular wavefront. As the wavefront propagated farther into the model, it encountered complex geological structures. Since these structures are characterized by strong velocity gradients, the rays began to rapidly bend. This phenomenon is particularly noticeable at $(x, z) = (3500\text{m}, 2300\text{m})$, where the wavefront encountered a low-velocity layer. The part of the wavefront that passed through the low-velocity layer lagged behind the rest of the wavefront, forcing the rays to converge toward each other. Since the wavefront construction method keeps only the first arrivals, all later arrivals were terminated, resulting in the *ray stitching* pattern observed from $x=3500\text{m}$ to $x=6000\text{m}$ at approximately $z=2500\text{m}$ depth. A similar pattern can be seen in locations where the rays have turned upward.

Despite the complexity of the raypaths, the wavefront remained relatively smooth throughout most of the simulation, as shown in figure 3.7(b). However, the wavefront became less smooth as it propagated through the aforementioned low-velocity layer. In particular, the part of the wavefront that passed through the low-velocity layer lagged behind the rest of the wavefront, resulting in a sharp “v” feature. However, the wavefront immediately healed itself after exiting the low-velocity layer.

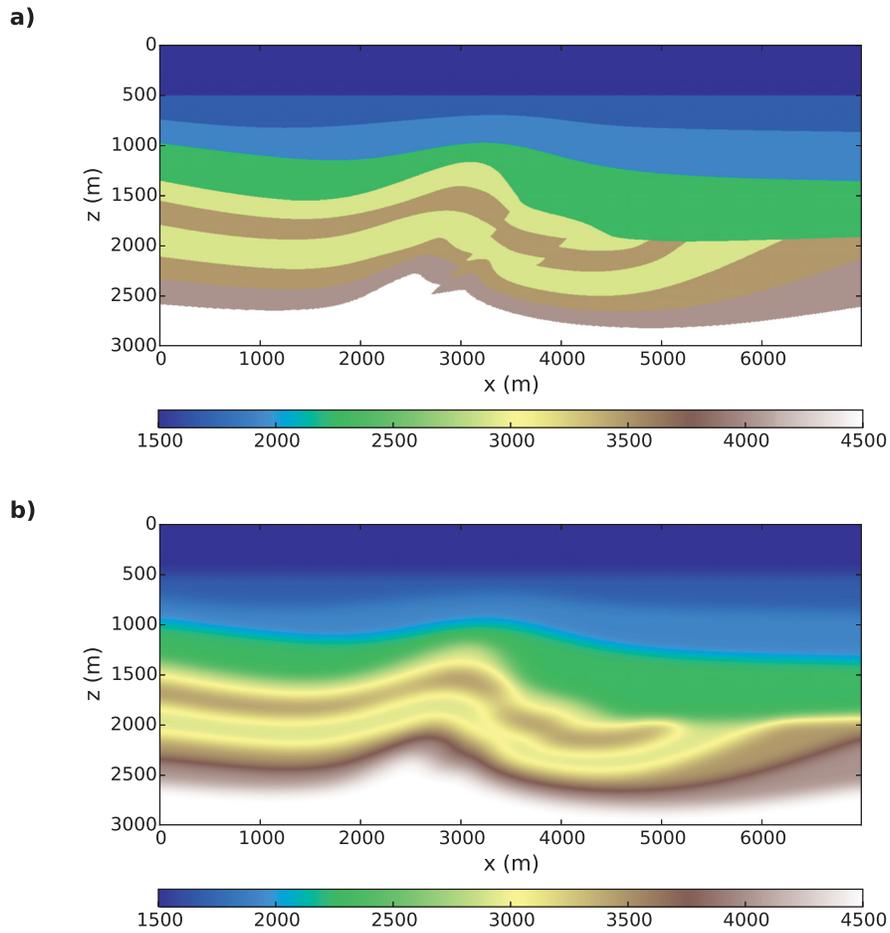


Figure 3.6: (a) SAIG velocity model. (b) Smoothed version of (a). Smoothing was applied to the slowness model using a Gaussian blurring kernel with a radius of 100m. The color-bars show the velocity in units of m/s.

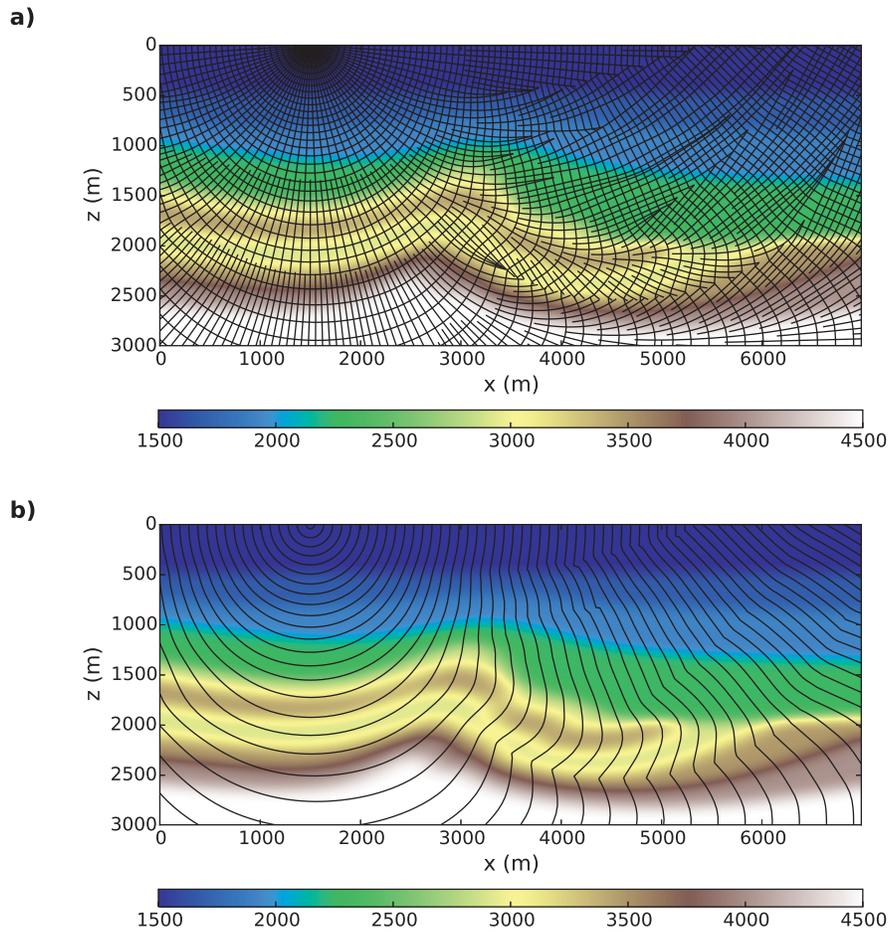


Figure 3.7: The wavefront construction method. (a) Wavefront mesh for a shot located at $x=1500\text{m}$ on the surface. (b) Location of the wavefront at 60ms intervals. The color-bars show the velocity in units of m/s.

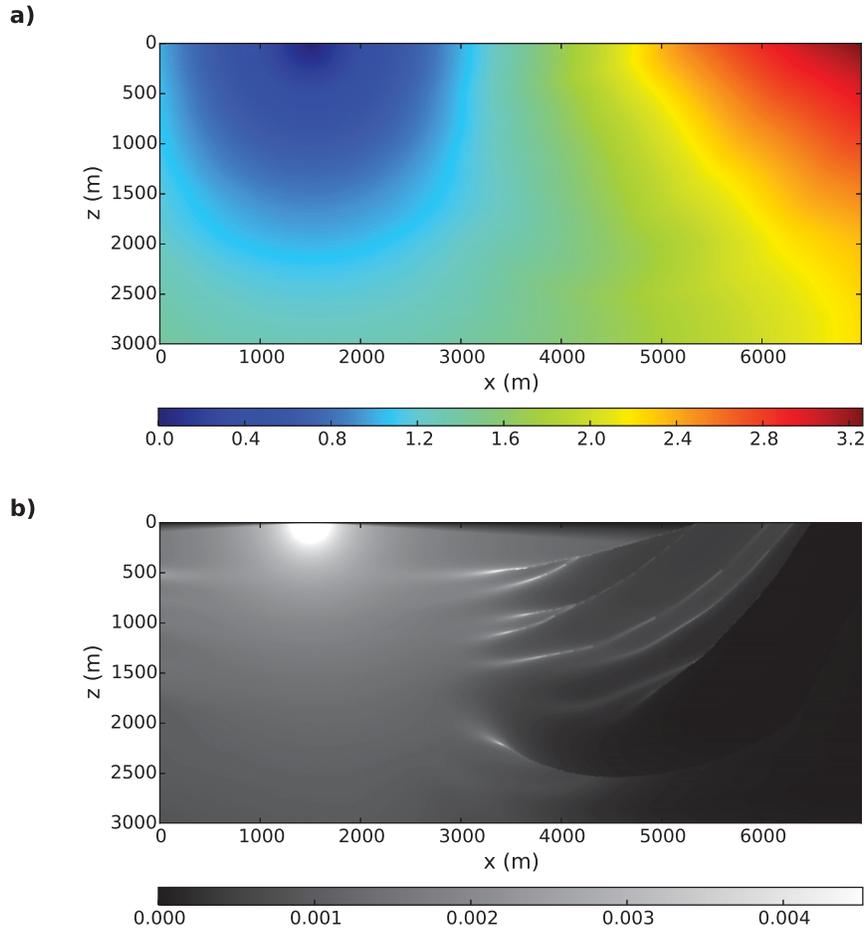


Figure 3.8: (a) Interpolated traveltimes and (b) interpolated amplitudes using the wavefront construction method. The amplitudes are heavily clipped to allow small-scale variations to be seen. The color-bar for (a) shows the traveltime in units of seconds. The colorbar for (b) shows the amplitude (unit-less).

Each piece of the mesh in figure 3.7(a) contains traveltimes and amplitude information about the wavefront. Interpolating this information onto a regular grid using the method discussed in section 3.3.2 produces the traveltimes and amplitude fields shown in figure 3.8(a) and figure 3.8(b), respectively. As expected, the traveltimes increase away from the source, and the traveltimes gradient follows the direction of the raypaths. The interpolated amplitudes exhibit a more complex behavior. For the most part, the amplitudes decrease away from the source, as expected. However, the amplitudes also rapidly increase in a few isolated regions. By comparing figure 3.7(a) with figure 3.8(b), it is clear that these high-amplitude regions correspond with locations in which the wavefront is converging on itself. Evidently, the wavefront construction method is capable of capturing amplitude variations arising from local changes in the behavior of the wavefront.

3.4.2 Example: Accuracy of the wavefront construction method

The wavefront construction method eliminates a number of fundamental issues associated with ray tracing. However, the method is only feasible if it can accurately estimate the traveltimes and amplitude components of the Green's functions. To verify the accuracy of our Runge-Kutta-based wavefront construction scheme, we compare our estimated traveltimes and amplitudes with those computed using an analytical expression. Consider the linear velocity model given by

$$c(\mathbf{z}) = a + b\mathbf{z} ,$$

where \mathbf{z} denotes depth, a is the velocity at the surface, and b is the vertical velocity gradient. For this example, we chose $a=1000\text{m/s}$ and $b=1.2\text{s}^{-1}$. The traveltimes τ and amplitudes A for this model can be computed analytically as (Vanelle, 2002)

$$\tau = \frac{1}{b} \operatorname{arcosh} \left(1 + \frac{b^2 r^2}{2v_s v_g} \right) ,$$

and

$$A = \sqrt{\frac{2}{\sqrt{b^2 r^4 + 4v_s v_g r^2}}} ,$$

where $\operatorname{arcosh}()$ is the hyperbolic cosine function, r is the distance between the source and receiver, and v_s and v_g are the velocities at the source and receiver, respectively.

Figure 3.9(a) shows the absolute error between the estimated and analytical traveltimes; the average error is approximately 0.25ms. Typical sampling rates for seismic data range from 1ms to 4ms. Thus, the traveltimes errors are less than one sample, suggesting they

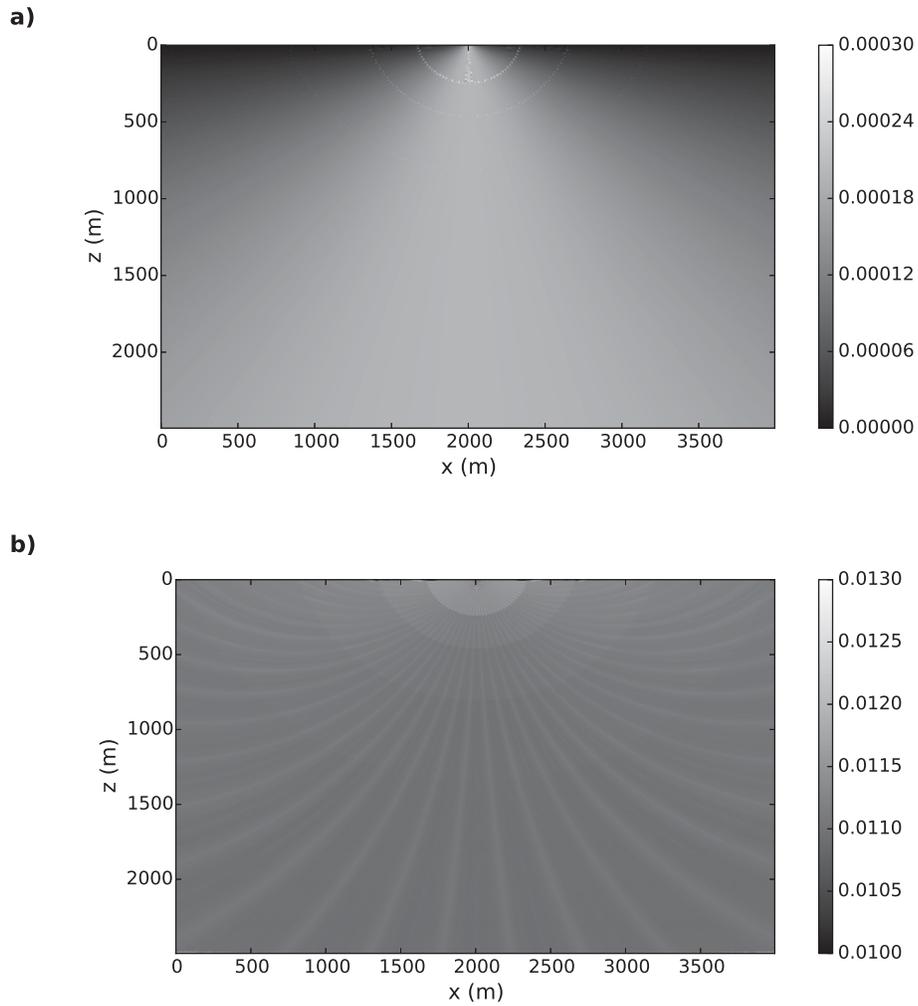


Figure 3.9: Error of the wavefront construction method for a linear velocity model. (a) Absolute travelttime error. (b) Relative amplitude error. The color-bar for (a) shows the travelttime error in seconds. The color-bar for (b) shows the relative amplitude error (unitless).

are negligible. Furthermore, the errors measured in this example are comparable with (or less than) the values reported by other authors (Rickett and Fomel, 1999; Zhang et al., 2011).

Figure 3.9(b) shows the relative error between the estimated and analytical amplitudes; the average error is approximately 1% at each point in the model. This level of error is acceptable because Kirchhoff migration is substantially less sensitive to amplitude errors than traveltimes. Additionally, Buske and Kästner (2004) verified that the amplitudes computed using the wavefront construction method are consistent with those obtained using a finite difference solution of the transport equation.

3.5 Summary

The wavefront construction method addresses a number of shortcomings that conventional ray-based schemes suffer from. In particular, a ray interpolation scheme is introduced that guarantees the ray-density remains approximately constant. As a result, deep parts of the model can be adequately sampled and ray shadow zones are no longer an issue. The wavefront construction method also incorporates a set of rules for determining which rays should be kept and which ones should be terminated if two or more rays cross, thereby improving the stability and accuracy of the interpolated traveltimes. In addition to this, WCM is flexible enough to choose between propagating the first-arrival wavefront, highest-amplitude wavefront, or multiple wavefronts - an obvious advantage compared with grid-based methods. Moreover, the wavefront construction method retains the accuracy and versatility offered by ray-based methods, while incorporating the robustness of grid-based methods. For these reasons, we utilize WCM to estimate the Green's functions in our Kirchhoff operator.

The wavefront construction method can be implemented in a variety of ways. Most notably, the ray tracing portion of the algorithm can be performed using a variety of different techniques. We showed that the Runge-Kutta method is accurate enough to compute the traveltimes and amplitudes with a negligibly small error. However, a lower-order ray tracing scheme could probably be used to reduce the computational cost of the algorithm without substantially impacting its accuracy.

CHAPTER 4

Least-squares migration

4.1 Introduction

Conventional seismic migration algorithms estimate an image of the subsurface by applying a linear migration operator to the seismic data. Since these methods estimate an image via a direct transformation, they are relatively inexpensive and can achieve satisfactory results when the recorded seismic data is densely sampled. However, seismic data is generally incomplete. Moreover, seismic data is often irregularly sampled and contaminated with noise and other artifacts. Consequently, images obtained using conventional migration algorithms typically suffer from poor illumination, low resolution, discontinuous reflectors, and migration artifacts (Nemeth et al., 1999).

Posing seismic migration as an inverse problem yields a higher resolution image with fewer artifacts (Nemeth et al., 1999; Duquet et al., 2000; Kuehl and Sacchi, 2003; Wang et al., 2005; Wang and Sacchi, 2006; Tang, 2009; Dai et al., 2012; Dong et al., 2012; Xue et al., 2015; Yao and Jakubowicz, 2016). This is accomplished by casting migration as a least-squares optimization problem. The least-squares cost function typically contains two terms: (i) a data misfit term, which measures how well the predicted synthetic data matches the observed data, and (ii) a regularization term, which measures the prevalence of undesirable features in the image. Seismic data that is noisy or irregular typically yields a low-quality image and therefore requires a larger amount of regularization. Conversely, high-quality seismic data typically yields an artifact-free image and therefore requires minimal regularization. It is therefore crucial to assign the relative weight of the misfit and regularization terms based on the quality of the seismic data. If the tradeoff between the two terms is chosen correctly, a true-amplitude image can be recovered that satisfies the observed seismic

data and is relatively free of artifacts.

Due to its high computational cost, most practical implementations of LSM have focused on migrating data under the acoustic approximation. The earliest attempts utilized a Kirchhoff operator due to its relatively low computational cost and versatility (Nemeth et al., 1999; Duquet et al., 2000). As computer technology advanced, LSM was made more powerful by utilizing one-way wave equation-based operators (Kuehl and Sacchi, 2003; Wang and Sacchi, 2006; Kaplan et al., 2010). Wave equation-based operators simulate wave propagation more accurately than Kirchhoff operators; consequently, wave equation-based LSM yields a superior image. Modern implementations of LSM go a step farther by using two-way RTM operators (Dong et al., 2012; Dai and Schuster, 2013; Yao and Jakubowicz, 2016). Recently, least-squares migration has been extended to handle elastic data (Stanton and Sacchi, 2017; Chen and Sacchi, 2017). Elastic LSM properly addresses the vector nature of seismic data; thus, it is superior to acoustic LSM when there is considerable crosstalk between different components of the data. Elastic LSM also yields an image of the PS reflectivity, which can provide insight into the location of fluid contacts and help illuminate targets beneath gas-filled zones.

The least-squares cost function is typically solved using the conjugate gradient method. Each iteration of conjugate gradients requires one application of the forward modeling operator and one application of the adjoint migration operator. Thus, least-squares migration has a substantial computational cost. The rate of convergence of least-squares migration can be improved by preconditioning the inverse problem. Moreover, by promoting “good” features in the image at each iteration, the algorithm converges faster and therefore has a lower computational cost (Wang et al., 2004).

This chapter reviews least-squares migration. The benefits of several different regularization strategies are discussed in terms of their computational cost and ability to suppress artifacts. A regularization strategy for promoting smoothness within common image gathers is reviewed, and an equivalent approach based on preconditioning is derived. Damped and preconditioned least-squares migration are then compared using a synthetic example. Lastly, as a continuation of our earlier discussion on Kirchhoff operator aliasing, an example is shown to illustrate the benefits of using an anti-aliased Kirchhoff operator in least-squares Kirchhoff depth migration.

4.2 Regularized least-squares migration

Seismic data is produced when the forward modeling operator acts on the reflectivity model:

$$\mathbf{d} = \mathbf{L}\mathbf{m} + \mathbf{n} , \quad (4.1)$$

where \mathbf{d} is the seismic data, \mathbf{L} is the forward modeling operator, \mathbf{m} is the reflectivity model, and \mathbf{n} is additive noise. In this thesis, \mathbf{L} is the Kirchhoff forward modeling operator described in chapter 2. Thus, from this point onward it is assumed that the seismic data is dominated by primary reflections. An image of the subsurface reflectivity should ideally be obtained by applying the inverse of the forward modeling operator to the seismic data:

$$\hat{\mathbf{m}} = \mathbf{L}^{-1}\mathbf{d} , \quad (4.2)$$

where $\hat{\mathbf{m}}$ is the estimated reflectivity model. In practice, the inverse is numerically unstable or impossible to estimate. Therefore, an alternative approach must be used to perform seismic migration. As discussed in chapter 2, it is a relatively straightforward task to derive the adjoint \mathbf{L}^\dagger of the forward modeling operator. Since the adjoint maps data-domain features into the model domain, it is a type of migration operator. Thus, an image of the subsurface can be obtained by applying the adjoint migration operator to the seismic data:

$$\hat{\mathbf{m}} = \mathbf{L}^\dagger\mathbf{d} . \quad (4.3)$$

The adjoint migration operator produces an image that suffers from low resolution, incorrect amplitudes, and sampling artifacts. Consequently, the estimated image does not satisfy the observed seismic data, which reduces our confidence in the model and makes it difficult to evaluate the accuracy of the inversion. Over the last two decades, least-squares migration has been widely advocated as an alternative approach that can improve the resolution of the image, suppress artifacts, and recover true amplitudes (Nemeth et al., 1999; Duquet et al., 2000; Kuehl and Sacchi, 2003; Kaplan et al., 2010). Least-squares migration estimates an image of the subsurface reflectivity by solving the optimization problem

$$\hat{\mathbf{m}} = \arg \min_{\mathbf{m}} \|\mathbf{W}(\mathbf{L}\mathbf{m} - \mathbf{d})\|_2^2 + \lambda^2 R(\mathbf{m}) , \quad (4.4)$$

where $\|\cdot\|_p$ denotes the p-norm, \mathbf{W} is a diagonal weighting matrix for de-emphasizing low-quality data, $R(\mathbf{m})$ is the regularization term, and λ is the regularization/tradeoff parameter. The first term in equation 4.4 measures the misfit between the predicted synthetic data and the true observed data, and the second term measures the prevalence of undesirable features in the model. As its name suggests, the tradeoff parameter controls the relative weight of the

two terms. Low-quality seismic data typically yields an image that suffers from a number of artifacts and therefore requires a large amount of regularization. Conversely, high-quality seismic data typically yields an artifact-free image and thus requires minimal regularization. If the balance between the misfit and regularization terms is chosen correctly, the inverted image will satisfy the observed seismic data and be relatively free of artifacts.

The regularization term can take a number of different forms, the most popular of which include:

- $R(\mathbf{m}) = \|\mathbf{m}\|_2^2$, the 2-norm, also known as the minimum norm (or damped) solution. Minimizing the 2-norm of the model is analogous to finding the “smallest” solution. Thus, damping via the minimum norm helps to diminish the amplitude of artifacts in the inverted image.
- $R(\mathbf{m}) = \|\mathbf{D}_h \mathbf{m}\|_2^2$, also a 2-norm, where \mathbf{D}_h is a first or second-order derivative operator applied along the angle (or offset) dimension of the model. Minimizing the 2-norm of the derivative along offset is analogous to finding a model with smoothly varying common image gathers. Moreover, since sharp changes in AVO are unrealistic, this regularization strategy ensures the common image gathers are well-behaved.

These regularization strategies have gained popularity because they are stable and robust, and because the 2-norm can be solved efficiently using the conjugate gradient method. The former (i.e. damped LSM) has become particularly popular due to its ability to suppress artifacts arising from noise and incomplete seismic data.

If we define the cost function as

$$F(\mathbf{m}) = \|\mathbf{W}(\mathbf{L}\mathbf{m} - \mathbf{d})\|_2^2 + \lambda^2 R(\mathbf{m}) , \quad (4.5)$$

then a solution to the least-squares optimization problem can be found by finding the global minimum of equation 4.5. Since equation 4.5 is a convex function, its global minimum can be obtained by solving

$$\frac{\partial F(\mathbf{m})}{\partial \mathbf{m}} = 0 . \quad (4.6)$$

When $R(\mathbf{m}) = \|\mathbf{m}\|_2^2$, the closed form solution for the model is given by

$$\hat{\mathbf{m}} = (\mathbf{L}^\dagger \mathbf{L} + \lambda^2 \mathbf{I})^{-1} \mathbf{L}^\dagger \mathbf{d} , \quad (4.7)$$

where \mathbf{I} is the identity matrix and \mathbf{W} has been absorbed into \mathbf{L} . Similarly, when $R(\mathbf{m}) = \|\mathbf{D}_h \mathbf{m}\|_2^2$, the closed form solution is given by

$$\hat{\mathbf{m}} = (\mathbf{L}^\dagger \mathbf{L} + \lambda^2 \mathbf{D}_h^\dagger \mathbf{D}_h)^{-1} \mathbf{L}^\dagger \mathbf{d}. \quad (4.8)$$

Evidently, least-squares migration recovers a model by applying the regularized pseudoinverse to the data rather than the exact inverse. The forward modeling operator and adjoint migration operator are typically implemented as computer programs rather than matrices. Thus, the pseudoinverse cannot be directly computed. Instead, most least-squares problems are solved using efficient iterative solvers such as the conjugate gradient method (see Appendix C).

4.3 Preconditioned least-squares migration

Common image gathers should consist of laterally continuous events if the migration velocity model is correct. It is therefore appropriate to utilize a cost function that penalizes sharp changes in model properties across the offset dimension. As discussed in the previous section, this can be accomplished by using the cost function

$$F(\mathbf{m}) = \|\mathbf{Lm} - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{D}_h \mathbf{m}\|_2^2, \quad (4.9)$$

where the data weighting operator \mathbf{W} has been absorbed into \mathbf{L} and \mathbf{d} . Regularization is used to penalize undesirable features in the model. Since sharp changes in AVO are non-physical, we seek a model in which these features have been suppressed. In equation 4.9, the derivative operator \mathbf{D}_h amplifies high-frequency variations in AVO which the least-squares solver then minimizes, ultimately yielding a model with smooth common image gathers. \mathbf{D}_h can therefore be viewed as a “bad pass” operator that emphasizes undesirable features in the model. By making the change of variables $\mathbf{z} = \mathbf{D}_h \mathbf{m}$, equation 4.9 can be rewritten as

$$F(\mathbf{z}) = \|\mathbf{L} \mathbf{D}_h^{-1} \mathbf{z} - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{z}\|_2^2, \quad (4.10)$$

where the model \mathbf{m} is given by $\mathbf{m} = \mathbf{D}_h^{-1} \mathbf{z}$. Equations 4.9 and 4.10 are equivalent only if \mathbf{D}_h^{-1} exists. In practice, \mathbf{D}_h^{-1} doesn’t need to exist because it can be replaced with a “good pass” operator that behaves as one wishes. Moreover, since \mathbf{D}_h acts as a high-pass filter (“bad pass” operator) that emphasizes undesirable features in the model, its inverse must act as a low-pass filter (“good pass” operator) that promotes desirable features in the model. Thus, \mathbf{D}_h^{-1} can be replaced with an operator that promotes smoothness in the model.

Cost functions that penalize sharp variations in the model typically use a sparse derivative

operator based on a finite difference approximation. For example, \mathbf{D}_h is often implemented as a sparse matrix with 1's on one diagonal and -1's on another diagonal. Finite difference operators are beneficial for a number of reasons: (i) they are intuitive to understand, (ii) any order of derivative can be implemented, and (iii) they are represented by sparse matrices. The last point is particularly important because sparse matrix-vector multiplication has a low computational cost. Thus, it is relatively inexpensive to use a finite difference operator to penalize the derivative of the model. However, the inverse of a sparse matrix is generally non-sparse; thus, if \mathbf{D}_h^{-1} exists, the product $\mathbf{D}_h^{-1}\mathbf{z}$ requires M^2 floating point operations. Choosing \mathbf{D}_h as a sparse finite difference operator is not a requirement. Moreover, the same effect on the model can be achieved by choosing \mathbf{D}_h^{-1} as a sparse “good pass” operator and allowing \mathbf{D}_h to be a dense matrix. Therefore, we choose a sparse matrix \mathbf{P} as our preconditioner, where the behavior of \mathbf{P} is similar to the expected behavior of \mathbf{D}_h^{-1} . Subbing the preconditioner into equation 4.10 yields the formulation for preconditioned least-squares migration

$$\begin{aligned}\hat{\mathbf{z}} &= \arg \min_{\mathbf{z}} \|\mathbf{LPz} - \mathbf{d}\|_2^2 + \lambda^2 \|\mathbf{z}\|_2^2, \\ \hat{\mathbf{m}} &= \mathbf{P}\hat{\mathbf{z}}.\end{aligned}\tag{4.11}$$

Preconditioning the inverse problem improves the eigenvalue distribution of the operator, thereby decreasing its condition number and improving its invertibility (Saad, 1992). Consequently, equation 4.11 requires fewer iterations of CG to reach convergence compared with other cost functions such as damped LSM.

The preconditioner \mathbf{P} is a smoothing operator. A natural choice for promoting lateral continuity within common image gathers is a triangular filter that is applied across the offset dimension. Triangular filters are represented by sparse matrices and are therefore computationally efficient to implement. In fact, they can be implemented using the same methodology that was discussed for the triangular anti-aliasing filter in chapter 2. Additionally, they retain the advantage of placing more weight on the central sample than on neighboring samples. Any choice of smoothing filter should, however, produce similar results.

4.4 Examples

The following synthetic examples demonstrate the pros and cons of damped and preconditioned least-squares migration. The first example illustrates the benefits of damped and preconditioned LSM compared with migration via the adjoint in terms of the resolution and amplitude balance of the image. The second example illustrates the impact of Kirchhoff

operator aliasing on least-squares migration.

4.4.1 Example: Damped vs preconditioned LSM

Least-squares migration reduces the prevalence of sampling artifacts and helps improve the resolution of the image, especially in complex geological settings. To illustrate the benefits of damped and preconditioned least-squares migration, consider the velocity model in figure 4.1. 50 shot gathers were forward modeled using an array of 200 receivers with a sampling rate of 4ms and a 20Hz Ricker wavelet; figure 4.2 shows a shot gather for a source located at $x=3500\text{m}$. Figure 4.3 (a), (b), and (c) show the stacked image obtained by performing migration via the adjoint, damped LSM, and preconditioned LSM with a 5-point triangular smoothing filter, respectively. Similarly, figure 4.4 (a), (b), and (c) show a common image gather located at $x=5000\text{m}$ for migration via the adjoint, damped LSM, and preconditioned LSM with a 5-point triangular smoothing filter, respectively. The adjoint migration operator produced an image volume that suffered from a number of issues. First, reflectors were poorly resolved in the faulted/structurally complex parts of the model as well as the deeper layers. Second, amplitudes were poorly balanced, particularly in the deeper parts of the model. Third, the common image gathers contained incorrect amplitude versus offset signatures at far offsets. Fourth, the far offsets in the common image gathers suffered from migration stretch, which further reduced the resolution of the reflectors in the stacked image. Lastly, the common image gathers were contaminated with low-amplitude artifacts due to incomplete destructive interference of the Kirchhoff operator.

Least-squares migration solves for a regularized model that is capable of predicting the seismic data; thus, LSM is expected to resolve the aforementioned issues. As expected, damped LSM offered a substantial improvement over migration via the adjoint. In particular, damped LSM boosted the resolution of the reflectors throughout the image, balanced the amplitudes of the faults and deeper reflectors, corrected the AVO signatures in the common image gathers, reduced the amount of migration stretch at far offsets, and suppressed the artifacts arising from incomplete destructive interference of the Kirchhoff operator. However, the image volume obtained using damped LSM did suffer from minor artifacts. Specifically, the common image gathers contained low-amplitude noise arising from the non-smooth nature of minimum norm solutions. Since seismic amplitudes are expected to vary smoothly with offset, promoting lateral continuity within common image gathers is expected to yield a more stable solution. In this example, lateral continuity is promoted via preconditioning by applying a 5-point triangular filter across the offset dimension of the common image gathers. As shown in figures 4.3 and 4.4, preconditioned LSM yielded the highest quality image volume with the fewest artifacts. Moreover, the preconditioner suppressed high-frequency

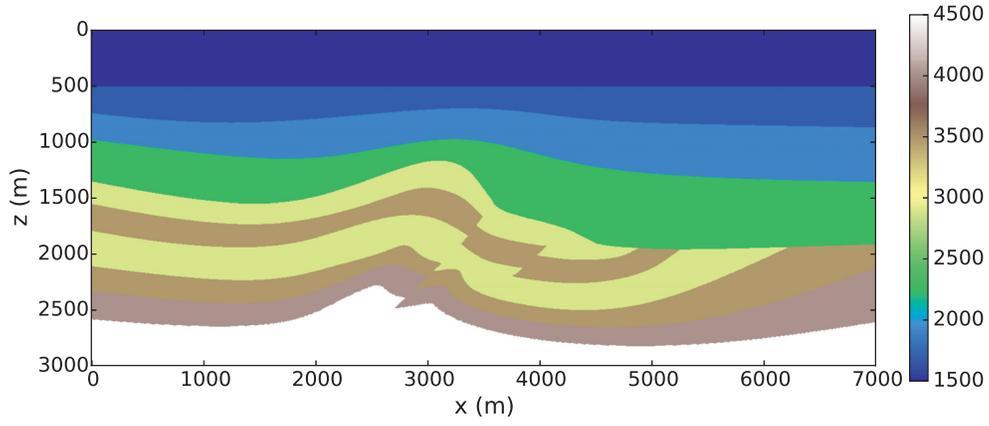
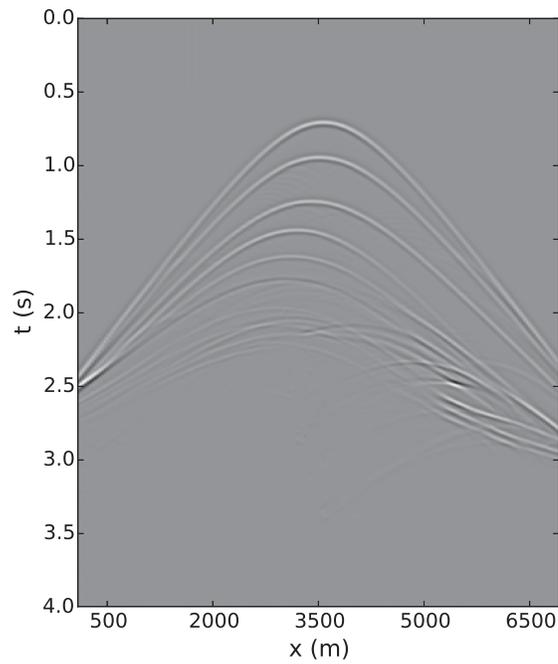


Figure 4.1: SAIG velocity model (m/s).

Figure 4.2: Forward modeled shot gather at $x=3500$ m for the velocity model in figure 4.1.

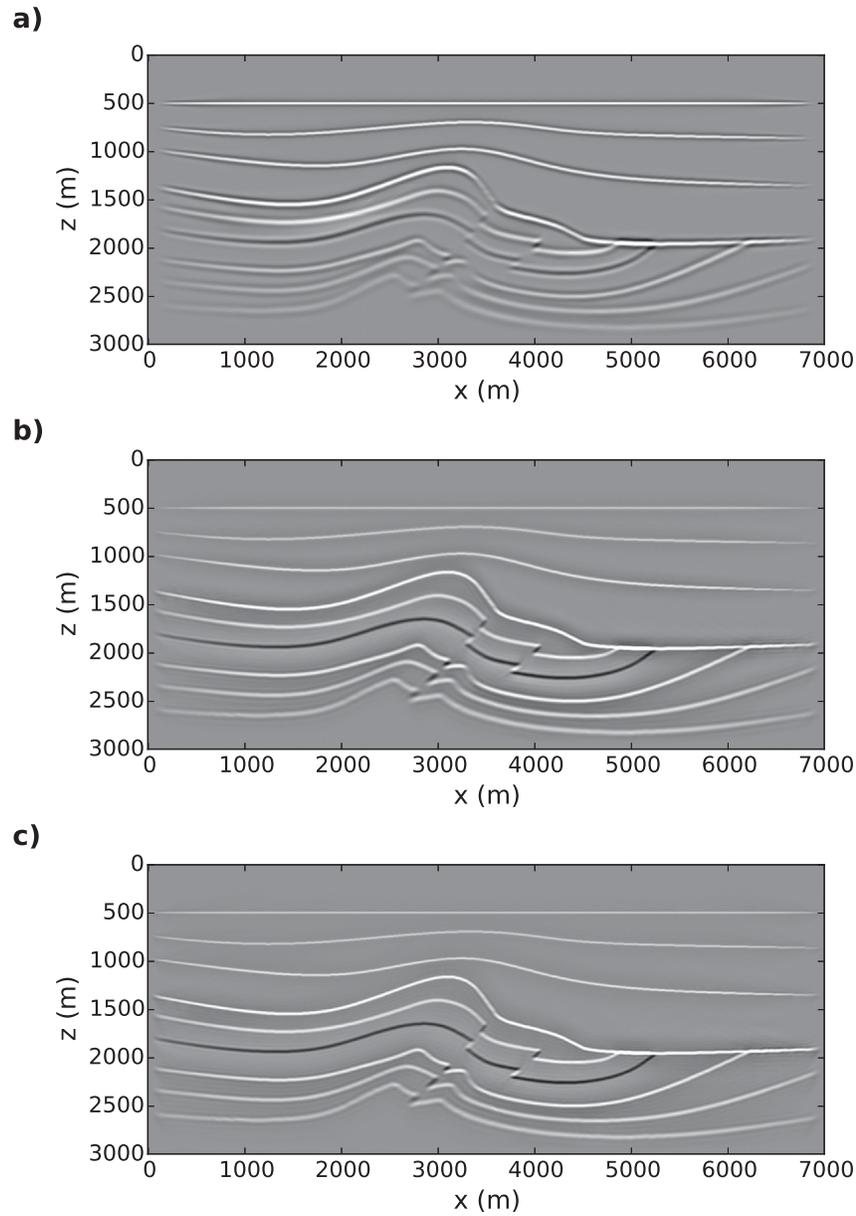


Figure 4.3: Stacked migrated image for (a) migration via the adjoint, (b) damped least-squares migration, and (c) preconditioned least-squares migration with a 5-point triangular smoothing filter.

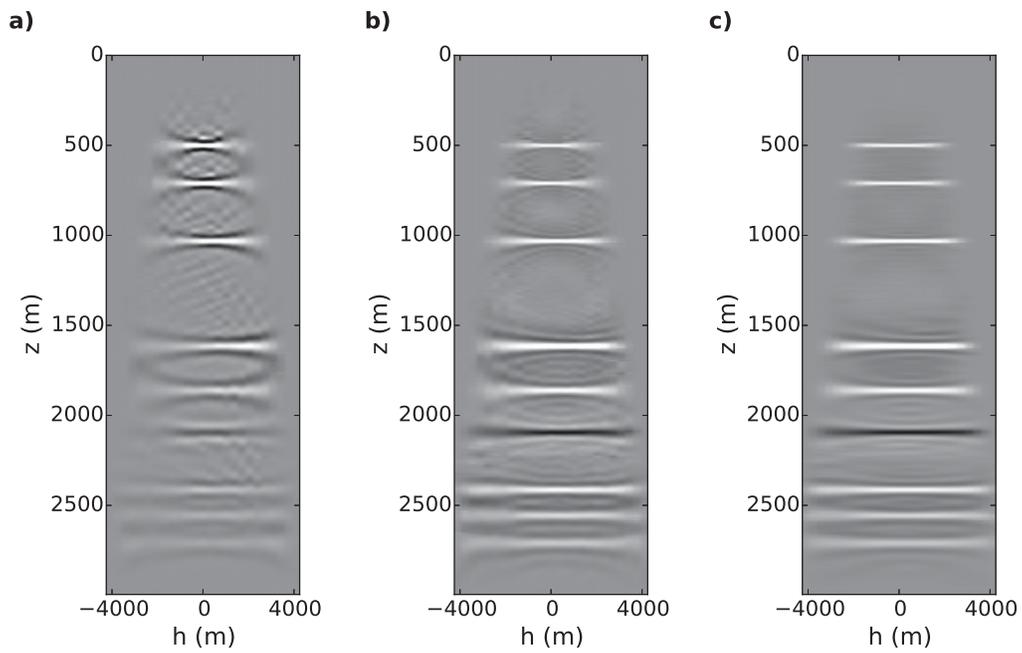


Figure 4.4: Common image gather located at $x=5000\text{m}$ for (a) migration via the adjoint, (b) damped least-squares migration, and (c) preconditioned least-squares migration with a 5-point triangular smoothing filter.

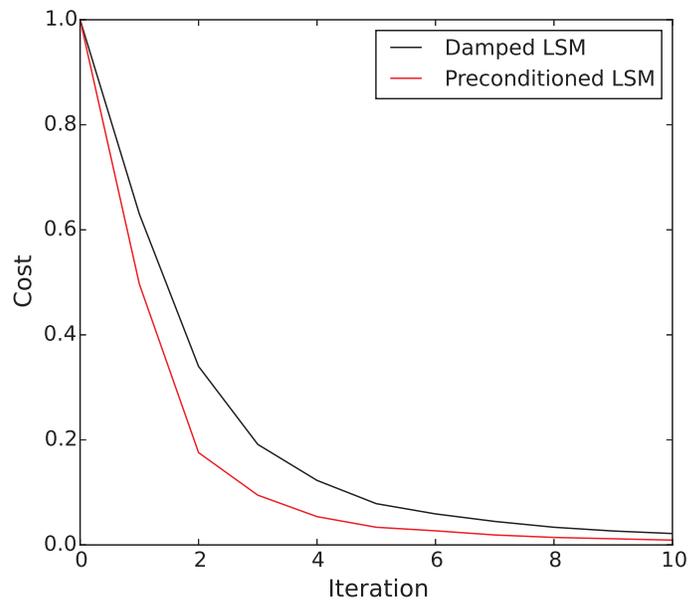


Figure 4.5: Convergence curves for damped and preconditioned least-squares migration.

variations in AVO and promoted lateral continuity of real events, resulting in laterally smooth common image gathers. Additionally, low-amplitude random noise was averaged out, yielding a relatively artifact-free image volume. Preconditioning also helped to reduce the amount of stretch at far offsets, thereby improving the resolution of the stacked image. Lastly, and perhaps most importantly, preconditioned LSM experienced a higher convergence rate than damped LSM, as shown in figure 4.5. In fact, damped LSM required 36 iterations to reach convergence while preconditioned LSM required only 20 iterations. This difference is substantial because the high computational cost of least-squares migration is its largest drawback. Thus, preconditioning increases the feasibility of applying least-squares migration to large 2D or 3D data sets.

4.4.2 Example: Kirchhoff operator aliasing in LSM

Chapter 2 discussed the impact of Kirchhoff operator aliasing when performing migration via the adjoint. An example was also provided to illustrate the effect of operator aliasing with a two-reflector model. The current example is a continuation of the one from chapter two, but differs in that it illustrates the impact of operator aliasing in the framework of least-squares migration. This example uses the same two-reflector model as before, as shown in figure 4.6(a). Additionally, the same constant-velocity model with $v=2000\text{m/s}$ was used to generate the zero-offset section in figure 4.6(b). The data consists of 50 receivers that were uniformly spaced over a distance of 2500m. Each receiver used a sampling interval of 4ms to ensure the data is adequately sampled in time. Thus, the spatial and temporal sampling of the data is sufficient to avoid data aliasing. Any artifacts appearing in the least-squares migrated image must therefore be a consequence of operator aliasing.

In figure 4.6(c), the input data was migrated using damped least-squares migration without accounting for Kirchhoff operator aliasing. Since operator aliasing was ignored, the Kirchhoff operator incorrectly migrated the data at each iteration of conjugate gradients. Consequently, the image is contaminated with low-amplitude artifacts that precede the real reflectors. Compared with figure 2.10(c), it is clear that LSM has partially suppressed the aliased energy. However, an ideal migration operator should eliminate the aliased energy altogether. Therefore, it is appropriate to perform least-squares migration using the anti-aliased Kirchhoff operator presented in chapter 2. In figure 4.6(d), the input data was migrated using damped least-squares migration with an anti-aliased Kirchhoff operator. Comparing figure 4.6(c) with figure 4.6(d), it is clear that the anti-aliased Kirchhoff operator has suppressed the aliased energy in the image, as expected. Evidently, Kirchhoff operator aliasing has consequences in both migration via the adjoint and least-squares migration.

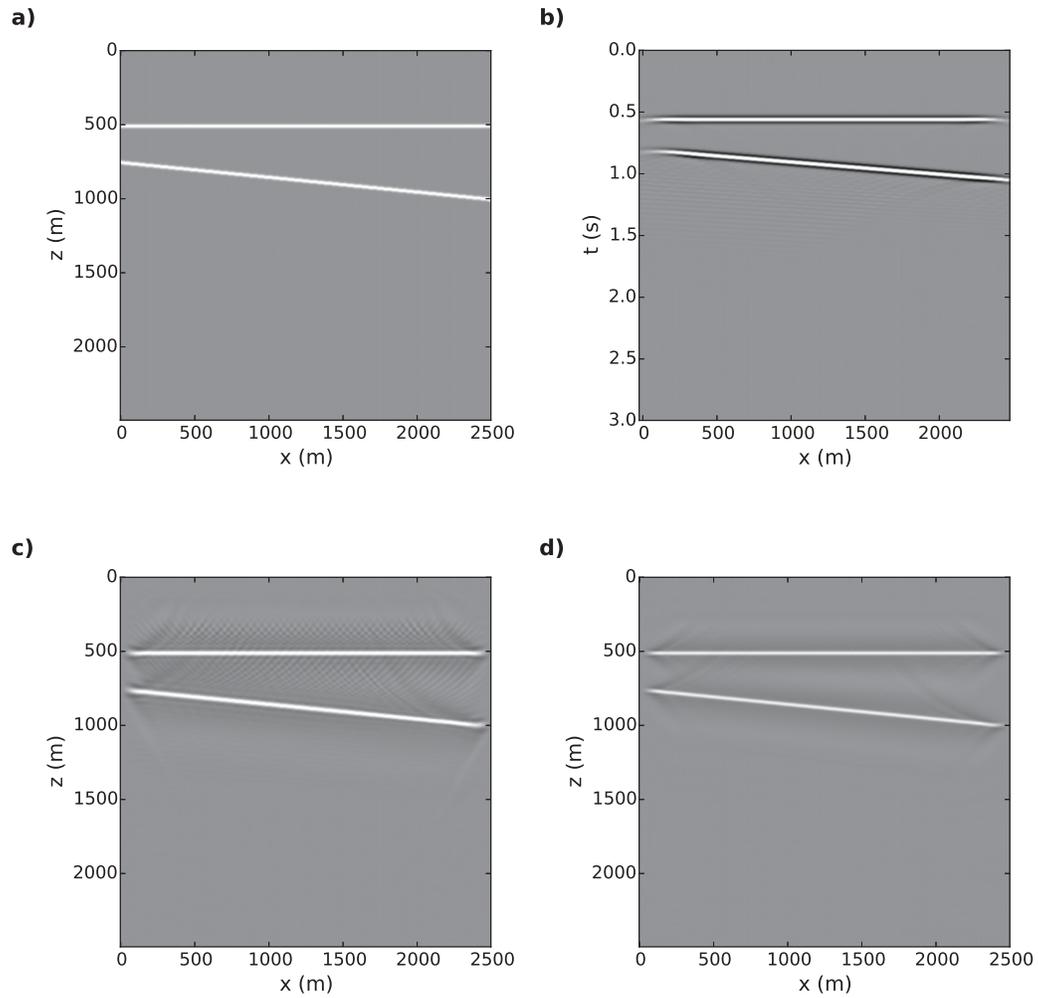


Figure 4.6: (a) Synthetic reflectivity model. (b) Zero-offset seismic section produced using the forward Kirchhoff operator \mathbf{L} . (c) Least-squares migrated image using the aliased Kirchhoff operator based on Algorithm 1. Least-squares migrated image using the anti-aliased Kirchhoff operator based on Algorithm 2.

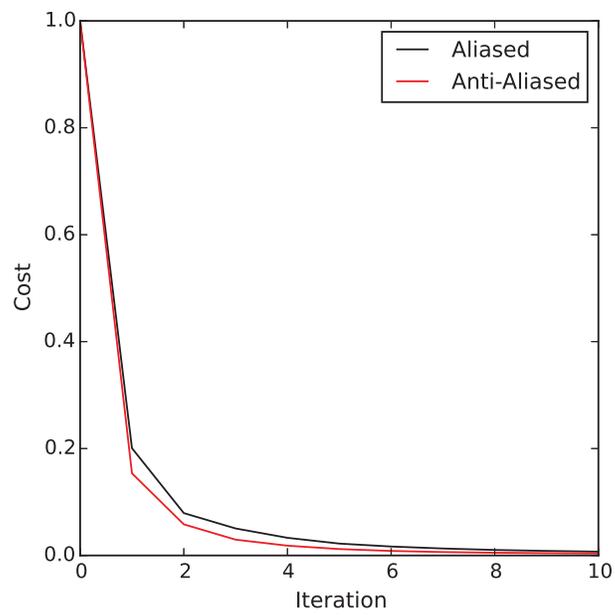


Figure 4.7: Convergence curve for aliased and anti-aliased least-squares Kirchhoff depth migration.

Since least-squares migration is an expensive, iterative process, it is tempting to ignore operator aliasing. However, it turns out that using an anti-aliased Kirchhoff operator can decrease the total cost of least-squares migration. When anti-aliasing is ignored, the adjoint operator introduces aliasing artifacts into the updated image at each iteration. These artifacts reduce the model's ability to predict the seismic data, which in turn increases the misfit between the predicted and observed data. Consequently, aliased Kirchhoff operators are expected to converge slower. Anti-aliased Kirchhoff operators, on the other hand, only migrate features whose frequency content is below the Nyquist frequency of the operator. Thus, artifacts are not introduced into the updated image, which in turn decreases the misfit between the predicted and observed data. These claims can be verified by comparing the convergence curves for aliased and anti-aliased least-squares Kirchhoff migration, as shown in figure 4.7. Specifically, the anti-aliased Kirchhoff operator converges slightly faster than the aliased Kirchhoff operator, thereby decreasing the cost of LSM. However, it should be noted that the difference in convergence rates is relatively small. This is likely due to the fact the forward Kirchhoff operator is a convolutional operator; consequently, the positive and negative aliasing artifacts approximately cancel each other when the forward Kirchhoff operator is applied to the model.

4.5 Summary

Conventional seismic migration algorithms produce an image that suffers from sampling artifacts and low resolution, and is incapable of predicting the observed seismic data. LSM addresses these issues by casting migration as an optimization problem in which the goal is to solve for a regularized model that is capable of predicting the observed data. Most cost functions utilize minimum norm regularization, which seeks to find a model whose norm is small. Minimum norm solutions are generally non-smooth and therefore result in low-amplitude artifacts. An improved image can be obtained by preconditioning the inverse problem by promoting smoothness along the offset dimension of each common image gather. The resulting common image gathers are more stable; thus, conjugate gradients converges in fewer iterations.

Although LSM is a powerful technique, it suffers from a number of shortcomings that limit its practicality. First, choosing the tradeoff parameter is a difficult, expensive process. Moreover, an L-curve or χ^2 test must be generated using multiple different values of the tradeoff parameter in order to identify the optimal choice. In practice, LSM is too expensive to repeat multiple times. Thus, there is considerable uncertainty when choosing the tradeoff parameter. Second, LSM is highly sensitive to errors in the velocity model. If the velocity

model contains errors, least-squares migration will contaminate the migrated image with artifacts in order to minimize the cost function. Thus, LSM must be preceded by a robust velocity inversion technique. Lastly, the computational cost of LSM limits its practical application for large 2D and 3D data sets. Preconditioning has been shown to improve the rate of convergence of LSM. However, more work on this topic is required to improve the feasibility of LSM with large data.

CHAPTER 5

Converted wave LSM with S-velocity errors

5.1 Introduction

Recently, there has been a growing interest in using PS converted waves as a supplementary tool in seismic exploration. Moreover, PS waves provide complimentary information that can be used to infer subsurface properties that cannot be estimated using PP waves alone. A few noteworthy examples include the identification of fluid contacts, imaging beneath gas-filled zones, and constraining the location of weak reflectors.

Migration is a crucial step in the converted wave processing workflow. Ideally, it should be implemented as the inverse of the forward modeling operator. However, because the inverse generally cannot be estimated, the adjoint is often used instead. The adjoint is capable of recovering a structurally correct image, but with low resolution and incorrect amplitudes. Therefore, the adjoint is considered a poor approximation to the inverse. Modern migration algorithms resolve these issues by solving for the pseudoinverse via least-squares migration. For example, Nemeth et al. (1999) found that LSM with a Kirchhoff operator improved the resolution of faults and reduced the acquisition footprint. More recently, Stanton and Sacchi (2014) found that converted wave LSM produced a higher-resolution image and fewer crosstalk artifacts compared with migration using the adjoint.

All seismic migration algorithms involve numerical propagation of seismic waves. Therefore, an accurate velocity model for the target region must be estimated beforehand. A great effort has been made to advance the technology surrounding velocity model estimation (Tarantola, 1984a; Bishop et al., 1985; Luo and Schuster, 1991; Pratt, 1999; Pratt and Shipp, 1999). Despite these advances, no technique is capable of recovering the true model without errors. This is particularly evident for S-velocity inversion, which requires the use

of converted wave data. Converted waves suffer from a low signal-to-noise ratio and are inherently more complex than PP waves. Consequently, S-velocity models are generally less accurate than P-velocity models. This poses a problem for converted wave migration, which requires an accurate estimate of both the P- and S-velocity models. If the errors in the S-velocity model are substantial, the migration operator will incorrectly propagate the converted waves, resulting in a structurally incorrect, unfocused image (Yan and Sava, 2010).

Luo and Hale (2014) recently addressed the problem of performing acoustic LSM in the presence of velocity errors. The authors introduced a dynamic image warping operator into the cost function to align the observed data with the predicted synthetic data prior to computing the misfit. This ultimately reduced the sensitivity of the inversion to traveltimes errors, thereby enabling the cost function to focus on fitting amplitudes. Zhang et al. (2014) proposed a related use for DIW: flattening common image gathers (CIGs). While both of these techniques are capable of improving the focus and resolution of the image, neither of them can ascertain the true location for the reflectors.

In this chapter, we modify the conventional cost function used in converted wave LSM to improve its accuracy in the presence of S-velocity errors. Our modified cost function includes a DIW operator that compensates for S-velocity errors by warping the image prior to computing the predicted synthetic data. Unlike the methods proposed by Luo and Hale (2014) and Zhang et al. (2014), our method is capable of recovering the correct location for the reflectors. However, the method is only suitable for converted wave migration, and is based on the assumption that the P-velocity model and PP image have been accurately estimated.

5.2 Warped converted wave LSM

Converted wave LSM aims to recover an image of the subsurface by minimizing the cost function

$$J = \|\mathbf{L}_{PS}\mathbf{m}_{PS} - \mathbf{d}_{PS}\|_2^2 + \|\mathbf{m}_{PS}\|_2^2, \quad (5.1)$$

where $\mathbf{L}_{PS} = \mathbf{L}_{PS}(v_P, v_S)$ is the converted wave forward modeling operator, v_P and v_S are the P- and S-velocity models, \mathbf{m}_{PS} is the converted wave image volume, and \mathbf{d}_{PS} is the converted wave data after wavefield separation. Equation 5.1 is typically solved using an iterative method such as conjugate gradients (CG). Each iteration of CG involves applying the forward operator to predict synthetic data, followed by migrating the data residual using the adjoint. Therefore, the cost for each iteration of CG is roughly double that of migration using the adjoint.

Equation 5.1 assumes the P- and S-velocity models have been accurately estimated. When this assumption holds true, the migration operator simulates the true propagation of converted waves. Therefore, equation 5.1 should approximately converge to the true image. However, if v_P or v_S contain errors, the migration operator will incorrectly propagate the converted waves, resulting in an incorrect image with non-flat CIGs. In the current study, we focus on improving migration in the latter case. Due to the increased complexity of converted waves compared with PP waves, we assume the S-velocity model contains the majority of the errors. Under this assumption it is implied that the P-velocity model is approximately correct, and therefore an accurate PP image can be obtained. Additionally, it is known *a priori* that PP and PS reflectors are co-located in space. Therefore, structural differences between the PP and PS images can be primarily attributed to errors in the S-velocity model. We propose to compensate for these errors by utilizing image registration to estimate the relative shifts between the PP and PS images.

Several algorithms have been developed for performing image registration. In this work, we utilize the dynamic image warping technique proposed by Hale (2013). DIW aims to recover the sequence of shifts $u[1 : N]$ that aligns two traces, f and g , such that

$$f[i] \approx g[i + u[i]] . \quad (5.2)$$

The shifts are estimated by solving the optimization problem

$$u[i] = \arg \min_l \sum_1^N e[i, l[i]] , \quad (5.3)$$

subject to the constraints

$$u_l \leq u[i] \leq u_u , \quad r_l \leq u[i] - u[i - 1] \leq r_u , \quad (5.4)$$

where $e[i, l]$ is the alignment error array given by

$$e[i, l] = (f[i] - g[i + l])^2 . \quad (5.5)$$

Here, u_l and u_u are lower and upper bounds on shifts $u[i]$, and r_l and r_u are lower and upper bounds on the strain. Therefore, u_l and u_u control the minimum and maximum shifts allowed, while r_l and r_u determine how rapidly the shifts can vary.

Equation 5.3 is capable of registering a single trace. DIW can therefore be extended to multi-dimensional images by solving equation 5.3 for each trace in the image. However, this approach neglects to impose continuity constraints across each dimension, and is therefore prone to introducing artifacts into the registered image. Hale (2013) demonstrated that a

better result can be achieved by smoothing the alignment error array across each dimension prior to solving equation 5.3. See Appendix D for a detailed review of multidimensional dynamic image warping.

Ideally, DIW should estimate the relative shifts between the least-squares PP and PS image volumes. In practice, it is too expensive to perform LSM beforehand. However, because the relative shifts between the PP and PS image volumes are expected to vary smoothly, it should be sufficient to perform DIW using the adjoint image volumes.

After the relative shifts have been estimated, we construct a warping operator, \mathbf{W} , such that its transpose, \mathbf{W}^\dagger , applies the shifts to the converted wave image. An approximation of the converted wave image, $\hat{\mathbf{m}}_{PS}$, can therefore be obtained by solving

$$\hat{\mathbf{m}}_{PS} = \mathbf{W}^\dagger \mathbf{L}_{PS}^\dagger \mathbf{d}_{PS} . \quad (5.6)$$

The image obtained using equation 5.6 is approximately equivalent to what would have been obtained using the adjoint migration operator with the true S-velocity model. Therefore, \mathbf{W}^\dagger reduces the sensitivity of the migration operator to S-velocity errors. Note that \mathbf{W}^\dagger is incapable of compensating for P-velocity errors. Errors in the P-velocity model will introduce inaccuracies into the PP image, which in turn will be transmitted to the converted wave image via the warping operator.

Since \mathbf{W}^\dagger aligns the images, \mathbf{W} must have the opposite effect, i.e., \mathbf{W} effectively applies the shifts in the opposite direction. If the true converted wave image, \mathbf{m}_{PS} , is known, \mathbf{W} can compensate for S-velocity errors by warping the true image prior to demigration:

$$\mathbf{d}_{PS} \approx \mathbf{L}_{PS} \mathbf{W} \mathbf{m}_{PS} . \quad (5.7)$$

\mathbf{W} warps the true image volume by an amount that roughly offsets the error introduced by the erroneous S-velocity model. Therefore, the true observed data is approximately equal to the predicted synthetic data.

Although equation 5.6 is capable of compensating for S-velocity errors, it fails to recover true amplitudes and suffers from low resolution. An improved image can be obtained by solving a least-squares problem that minimizes the difference between the observed data and the predicted synthetic data in equation 5.7. Therefore, we propose a new cost function that is a modification of equation 5.1:

$$J = \|\mathbf{L}_{PS} \mathbf{W} \mathbf{m}_{PS} - \mathbf{d}_{PS}\|_2^2 + \|\mathbf{m}_{PS}\|_2^2 . \quad (5.8)$$

Equation 5.8 can be readily solved using an iterative scheme such as conjugate gradients.

5.3 Example: LSM with S-velocity errors

To demonstrate the efficacy of our modified cost function, we performed a synthetic test using an anticline model with P- and S-velocities shown in figures 5.1(a) and 5.1(b), respectively. Migration velocity models were obtained by smoothing the true models with a 2D Gaussian filter. We then contaminated the S-velocity model with the errors in figure 5.1(c). 50 two-component shot gathers were generated using elastic finite difference modeling with a 40Hz Ricker wavelet. PP and PS data was then extracted by taking the divergence and curl of the data, respectively; figure 5.2 shows the PP and PS data for a source located at 800m. For simplicity, we utilized a 2D Kirchhoff depth migration/demigration operator to migrate the data, where the traveltimes were estimated using the wavefront construction method Vinje et al. (1993).

Figures 5.3(a) and 5.4(a) show the stacked PP image and a CIG obtained using the adjoint PP migration operator. The reflectors are correctly positioned and the CIGs are flat. However, the image suffers from low resolution and poorly balanced amplitudes. Figures 5.3(b) and 5.4(b) show the stacked PS image and a CIG obtained using the adjoint PS migration operator. A comparison of the PP and PS CIGs suggests that the S-velocity model contains errors. Therefore, when we performed converted wave LSM using the conventional cost function, we obtained an incorrect image with non-flat CIGs, as shown in Figures 5.3(c) and 5.4(c).

Because the S-velocity model was known to contain errors, we proceeded to perform converted wave LSM using our modified cost function. We began by performing DIW using the adjoint PP and PS image volumes; figure 5.4(f) shows the estimated shifts for the CIG located at $x=1000\text{m}$. The magnitude of the shifts generally increases with offset. This is due to the fact that the converted wave CIGs were non-flat, with the far offsets requiring a larger correction than the near offsets. Applying the estimated shifts to the adjoint PS image volume resulted in the stacked PS image and CIG in figures 5.3(d) and 5.4(d). Compared with the original PS CIGs, the reflectors in the warped CIGs have been flattened and shifted back to their true location. Therefore, the warping operator successfully compensated for the errors in the S-velocity model. However, the image still suffers from low resolution and incorrect amplitudes. Therefore, an improved image can be obtained by utilizing our modified cost function in equation 5.8. Our modified cost function converged to the stacked PS image and CIG in figures 5.3(e) and 5.4(e). Unlike the conventional cost function, our modified cost function was able to flatten the CIGs and recover the true location of the PS reflectors. Additionally, our modified cost function retained the traditional advantages associated with least-squares migration, including true amplitudes and improved resolution.

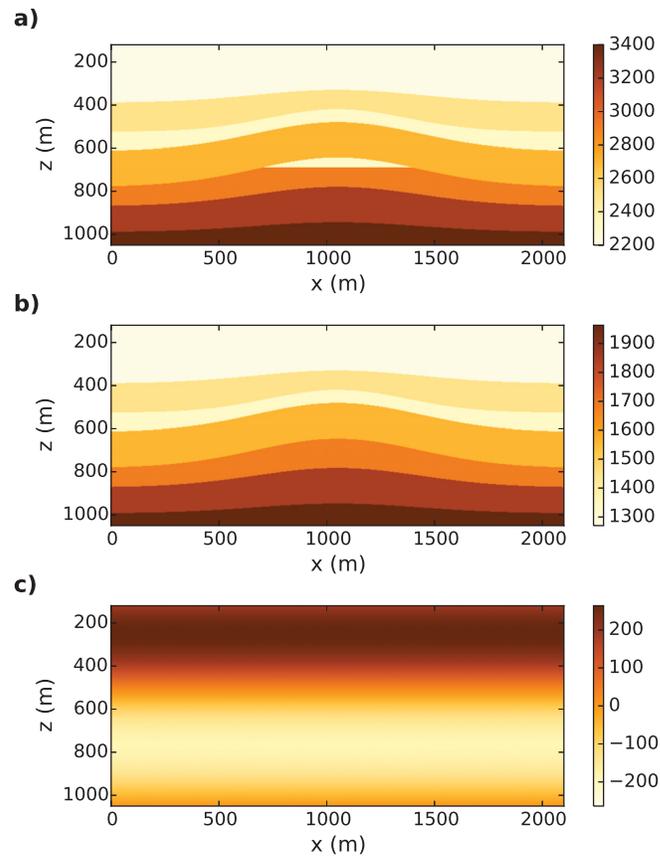


Figure 5.1: (a) True P-wave velocity model. (b) True S-wave velocity model. (c) Errors in S-velocity model. The color-bars show the velocity in units of m/s.

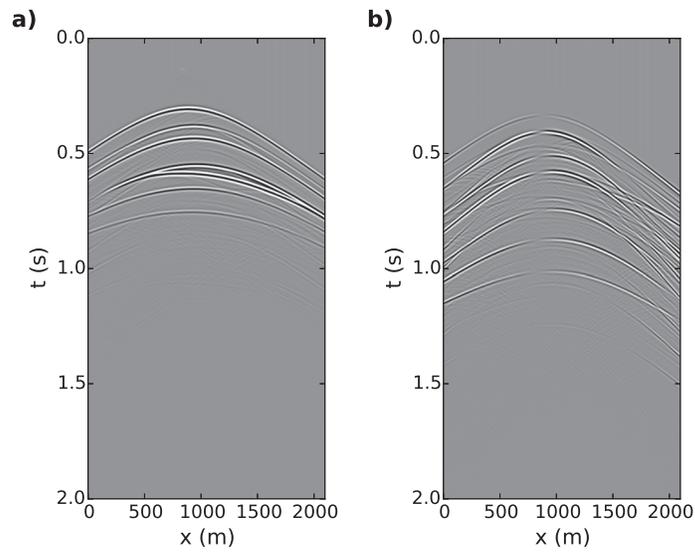


Figure 5.2: (a) PP data and (b) PS data produced by elastic finite difference modeling for a source located at $x=800\text{m}$. The direct wave was removed using a top mute.

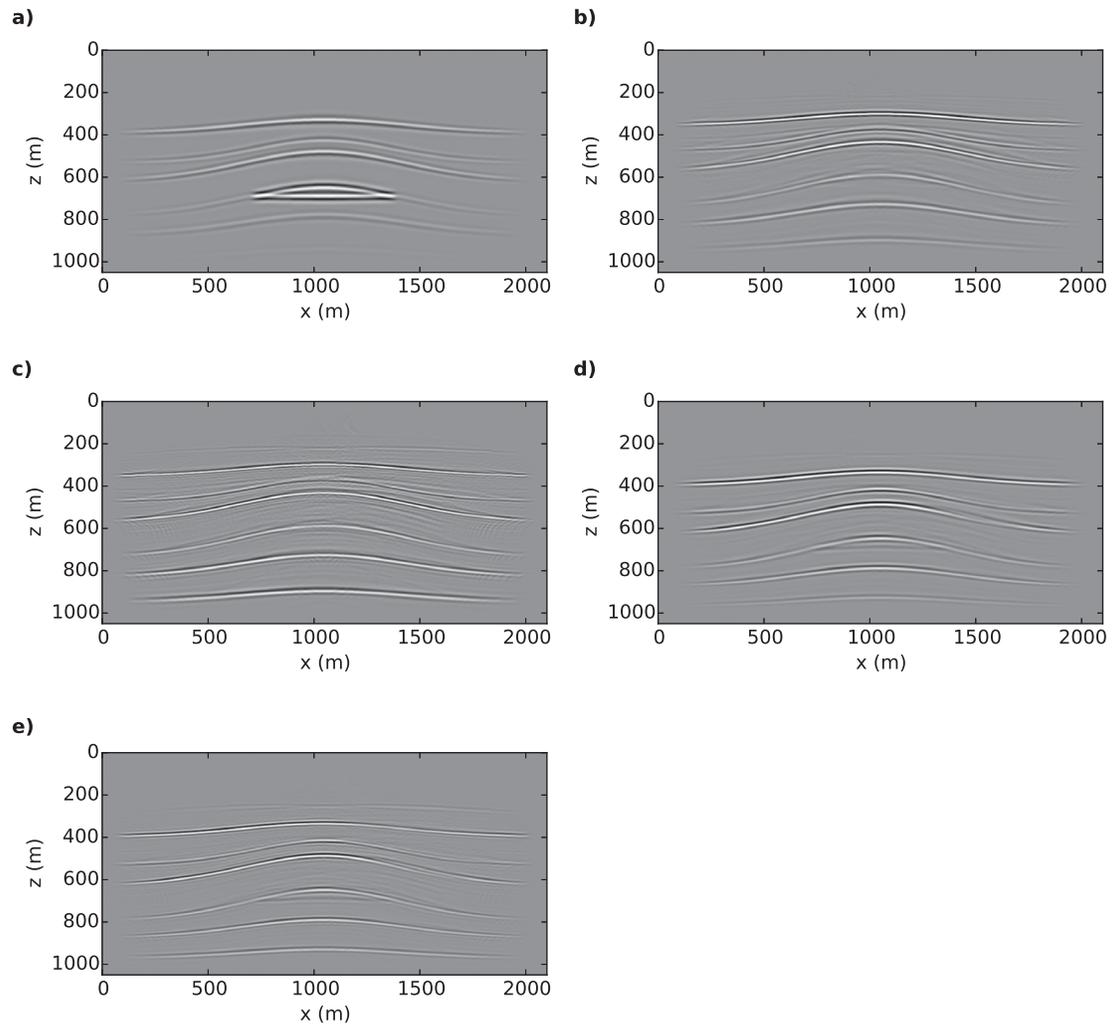


Figure 5.3: Migration in the presence of S-velocity errors. (a) Adjoint PP image. (b) Adjoint PS image. (c) Conventional PS LSM image. (d) Warped adjoint PS image. (e) Warped PS LSM image.

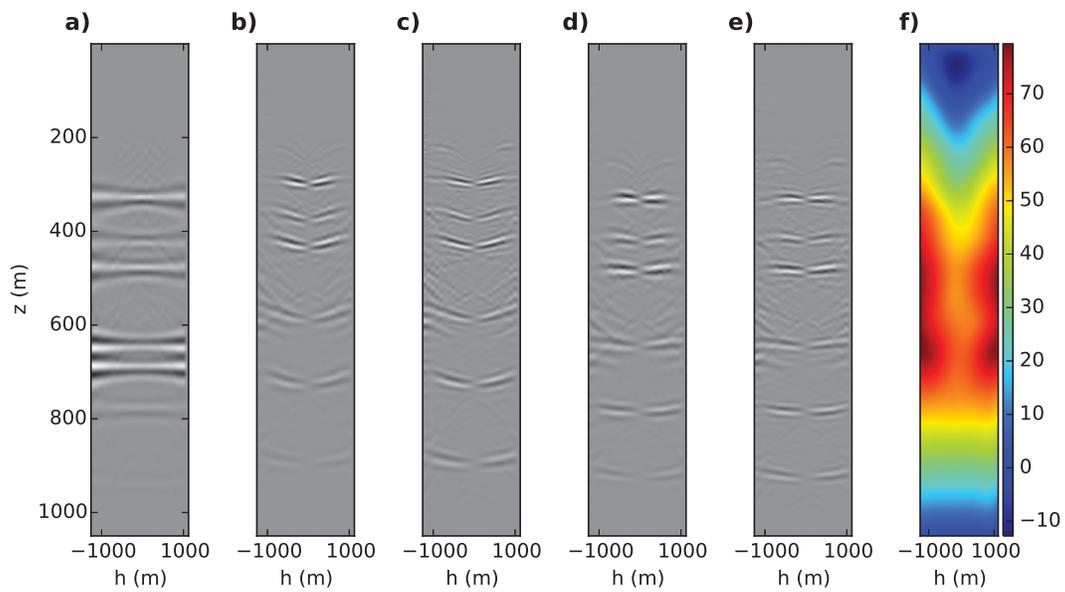


Figure 5.4: Common image gathers located at $x=1000\text{m}$ for (a) PP adjoint, (b) PS adjoint, (c) conventional PS LSM, (d) warped PS adjoint, and (e) warped PS LSM. Panel (f) shows the shifts that were used/estimated by the warping operator; the color-bar shows the vertical shifts in units of meters.

5.4 Summary

Converted wave least-squares migration minimizes the misfit between the observed PS data and the predicted synthetic PS data. If the P- or S-velocity models contain errors, the migration operator will incorrectly propagate the converted waves; consequently, the conventional cost function will converge to an incorrect image that is contaminated with artifacts. In this chapter, we modified the conventional cost function used in converted wave least-squares migration to reduce its sensitivity to S-velocity errors. Our modified cost function includes a dynamic image warping operator that compensates for S-velocity errors by warping the image prior to computing the predicted synthetic data. Moreover, the relative shifts between the adjoint PP and PS images are used to compensate for S-velocity errors by warping the PS image prior to forward modeling. The shifts applied by the warping operator offset the errors introduced by the erroneous S-velocity model; consequently, the predicted synthetic PS data matches the observed PS data, so a correct PS image can be recovered.

We demonstrated with synthetic data that our modified cost function converges to a more accurate image than the conventional cost function when the S-velocity model contains errors. In particular, dynamic image warping was found to be sufficiently accurate for estimating the relative shifts between the adjoint PP and PS images. Applying the estimated shifts to the PS image aligned the PS reflectors with the PP reflectors. Conversely, when the adjoint of those shifts were applied to the PS image prior to forward modeling, the predicted PS data was found to approximately match the observed PS data. As a result, our modified cost function yielded a high-resolution, true-amplitude PS image that was structurally correct. Additionally, since our modified cost function has a reduced sensitivity to S-velocity errors, it was found to produce substantially fewer artifacts than the conventional cost function.

Although the initial results are promising, our modified cost function suffers from a fundamental issue that limits its practicality. Specifically, if the geology is structurally complex and the velocity errors are large, reflectors may intersect within common image gathers. The conventional dynamic image warping method does not include any *a priori* constraints for addressing this issue; consequently, the estimated shifts may contain errors. More work on this topic is therefore required to improve the robustness of the method. For example, it may be beneficial to decouple the warping procedure into three separate steps: (i) apply a residual moveout correction to each common each gather, (ii) stack the common image gathers to form a post-stack image, then (iii) perform image registration using the post-stack images. Decoupling the warping procedure in this manor should improve the stability of the shifts and lower the computational cost.

CHAPTER 6

Conclusions

6.1 Summary

Conventional seismic migration operators yield an image that suffers from low resolution, poorly balanced amplitudes, and sampling artifacts. Least-squares migration addresses these issues by finding a model that minimizes the difference between the true observed data and the predicted synthetic data. This thesis aimed to advance the technology surrounding multicomponent least-squares Kirchhoff depth migration by improving the stability of the forward/adjoint Kirchhoff operator and modifying the conventional PP and PS migration cost functions.

The Kirchhoff operator maps scatterers from the model domain into diffractions in the data domain. Since seismic data is collected using a limited number of traces, the data-domain diffractions are sampled a finite number of times; consequently, the Kirchhoff operator is subject to the issue of aliasing. Kirchhoff operator aliasing can be circumnavigated using one of two strategies: (i) limiting the dip of the Kirchhoff operator, or (ii) suppressing the frequencies that violate the dip Nyquist criterion. Limiting the dip is viable when the geology is flat; however, if the geology is structurally complex, information is needed from the steeply dipping parts of the Kirchhoff operator in order to generate an accurate image. The latter approach for anti-aliasing was investigated in chapter 2 of this thesis. Moreover, an efficient triangular anti-aliasing filter was embedded in the forward/adjoint Kirchhoff operator to suppress the frequencies that violated the dip Nyquist criterion. Unlike alternative anti-aliasing methods, which require multiple copies of the data to be forward modeled and saved, the triangular anti-aliasing filter used in this study required the use of only one data set, thereby decreasing the input/output cost of the algorithm and improving the

memory-efficiency of the operator. Anti-aliasing the Kirchhoff operator was found to yield an improved image for both conventional migration and least-squares migration. Moreover, the triangular anti-aliasing filter suppressed the aliased frequencies in the data, resulting in a relatively artifact-free image. Importantly, the inclusion of the triangular anti-aliasing filter came with a negligible increase in computational cost, making it a natural choice for least-squares migration. Finally, since the anti-aliasing filter suppressed the aliased energy at each iteration of conjugate gradients, least-squares migration with an anti-aliased Kirchhoff operator converged in fewer iterations than least-squares migration with an aliased Kirchhoff operator. Thus, anti-aliasing not only improved the quality of the image, but also increased the rate of convergence of least-squares migration.

The Kirchhoff operator uses Green's functions to simulate the propagation of seismic waves. Specifically, a source-side Green's function is used to simulate the downward propagation of seismic waves from the source into the subsurface, and a receiver-side Green's function is used to simulate the upward propagation of scattered waves to each receiver. Kirchhoff migration therefore requires a robust method for estimating Green's functions. Chapter 3 of this reviewed the wavefront construction method for estimating seismic Green's functions. The wavefront construction method utilizes ray tracing principles to propagate the wavefront itself, rather than individual rays. As a result, the wavefront construction method can compute and distinguish between multiple arrivals based on the geometry of the wavefront. Alternatively, a single arrival, such as the first arrival or highest-amplitude arrival, can be isolated and propagated on its own. This is a substantial advantage compared with other robust techniques for estimating Green's functions, which are typically limited to computing first arrivals. Additionally, the relative divergence of each point on the wavefront can be compared at successive iterations to obtain an estimate of the amplitude of the Green's functions. Thus, the wavefront construction method is capable of accurately estimating all of the quantities required by the Kirchhoff operator.

Least-squares migration is typically solved iteratively using the conjugate gradient method. Each iteration of conjugate gradients requires one application of the forward modeling operator and one application of the adjoint migration operator. Thus, each iteration of CG costs approximately double that of conventional migration. Since several iterations are required to reach convergence, least-squares migration has a substantial computation cost. For this reason, industry has been slow to adopt LSM. Many industrial implementations of LSM circumnavigate this issue by truncating CG after a fixed number of iterations. Although this strategy lowers the computational cost of LSM, it also reduces the quality and resolution of the migrated image. Therefore, there is considerable motivation to improve the rate-of-convergence of LSM. Chapter 4 of this thesis investigated a preconditioning strategy

for reducing this cost. Moreover, a “good-pass” smoothing filter was embedded in the misfit term of the cost function to promote lateral continuity within common image gathers. Since seismic amplitudes are expected to vary smoothly with offset, the preconditioner stabilized the inverse problem. Moreover, preconditioning the inverse problem improved the eigenvalue distribution and decreased the condition number of the operator, thereby improving its invertibility. As a result, fewer iterations of conjugate gradients were needed to reach convergence.

Lastly, this thesis sought to improve the accuracy of converted wave migration in the presence of S-velocity errors. Due to the difficulty in estimating S-velocity models, converted wave images are typically less accurate than PP images. However, it is known *a priori* that PP and PS reflectors are collocated in space. This thesis exploited this observation by utilizing image registration to automatically warp/align the PS reflectors with the presumably correct PP reflectors, thereby reducing the sensitivity of converted wave migration to errors in the S-velocity model. The warping operator was then embedded in the converted wave migration cost function, allowing for the recovery of a high-resolution, true-amplitude PS image that is structurally correct, despite the presence of errors in the S-velocity model. The method was applied to a simple anticline velocity model as a proof-of-concept. The modified cost function was found to flatten the common image gathers and shift the PS reflectors to their correct location in space. Additionally, since the predicted synthetic data better matched the observed data, fewer artifacts were introduced into the final image.

To summarize, the main contributions of this thesis were:

1. The development of an anti-aliased forward/adjoint Kirchhoff operator that satisfies the dot product test,
2. A demonstration of the improved convergence rate resulting from using an AVO-smoothing preconditioner in least-squares migration,
3. The proposal of a new converted wave least-squares migration cost function that utilizes image registration to reduce the impact of S-velocity errors.

6.2 Future work

This thesis proposed a new cost function for performing converted wave least-squares migration in the presence of S-velocity errors. The method relies on using image registration to estimate the relative shifts between corresponding reflectors in the PP and PS image volumes. Those shifts are then embedded in a warping operator that is used to compensate for

S-velocity errors. Evidently, the method is limited by the accuracy and robustness of the image registration technique. This thesis focused on a robust, computationally efficient image registration technique called dynamic image warping. Dynamic image warping proved to be sufficiently stable for the simple velocity model used in chapter 5. However, real models are often substantially more complicated than the one used in this work. Moreover, real PP and PS images often contain closely-spaced reflectors. If the errors in the S-velocity model are substantial, the reflectors may intersect each other in the common image gather domain. It is not yet known if dynamic image warping is robust enough to correctly handle this situation. Therefore, more work is required to determine the validity of our cost function when the model is structurally complex. Moving forward, it may be advantageous to break the warping procedure into three steps: (i) apply a residual moveout correction to each common each gather, (ii) stack the common image gathers to form a post-stack image, then (iii) perform image registration using the post-stack images. Decoupling the method in this manor should improve the stability of the estimated shifts and decrease the computational cost.

This work focused on least-squares migration with a Kirchhoff operator. It would be an interesting and useful experiment to investigate the benefits of using a wave equation-based operator in our proposed converted wave LSM cost function. Since image registration is affected by artifacts, the greater accuracy associated with wave equation-based operators should allow for the recovery of more accurate relative shifts, which in turn should result in an improved converted wave image.

Bibliography

- Abma, R., Sun, J., and Bernitsas, N. (1999). Antialiasing methods in Kirchhoff migration. *Geophysics*, 64(6):1783–1792.
- Aki, K. and Richards, P. G. (2002). *Quantitative seismology*. University Science Books, Sausalito, Calif.
- Baysal, E., Kosloff, D. D., and Sherwood, J. (1983). Reverse time migration. *Geophysics*, 48(11):1514–1524.
- Berryhill, J. R. (1979). Wave-equation datuming. *Geophysics*, 44(8):1329–1344.
- Beylkin, G. (1985). Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized radon transform. *Journal of Mathematical Physics*, 26(1):99–108.
- Biondi, B. (2001). Kirchhoff imaging beyond aliasing. *Geophysics*, 66(2):654–666.
- Bishop, T. N., Bube, K. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, R. T., Spindler, D. A., and Wyld, H. W. (1985). Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, 50(6):903–923.
- Bleistein, N. (1987). On the imaging of reflectors in the earth. *Geophysics*, 52(7):931–942.
- Bleistein, N., Cohen, J. K., Stockwell, J. W., and Berryman, J. G. (2001). *Mathematics of Multidimensional Seismic Imaging, Migration, and Inversion*, volume 13. Springer, New York, New York.
- Bleistein, N. and Gray, S. H. (2002). A proposal for common-opening-angle migration/inversion. *Center for Wave Phenomena Research Report*, 420.
- Boore, D. M. (1972). Finite difference methods for seismic wave propagation in heterogeneous materials. *Methods in computational physics*, 11:1–37.
- Born, M. and Wolf, E. (1999). *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Cambridge University Press, 7th edition.
- Buske, S. and Kästner, U. (2004). Efficient and accurate computation of seismic traveltimes and amplitudes. *Geophysical Prospecting*, 52(4):313–322.
- Cameron, M., Fomel, S., and Sethian, J. (2008). Time-to-depth conversion and seismic velocity estimation using time-migration velocity. *Geophysics*, 73(5):VE205–VE210.

- Cary, P. W. (1998). P/S wavefield separation in the presence of statics. *CREWES Research Report*, 10:30–1.
- Červený, V. (1987). Ray tracing algorithms in three-dimensional laterally varying layered structures. In *Seismic Tomography*, pages 99–133. Springer.
- Červený, V., Molotkov, I. A., and Pšenčík, I. (1977). *Ray method in seismology*. Univerzita Karlova.
- Chavent, G. and Plessix, R. (1999). An optimal true-amplitude least-squares prestack depth-migration operator. *Geophysics*, 64(2):508–515.
- Chen, K. and Sacchi, M. D. (2017). Elastic least-squares reverse time migration via linearized elastic full waveform inversion with pseudo-Hessian preconditioning. *Geophysics*, 82(5):1–89.
- Claerbout, J. F. (1971). Toward a unified theory of reflector mapping. *Geophysics*, 36(3):467–481.
- Claerbout, J. F. (1985). *Imaging the Earth's interior*. Blackwell scientific publications Oxford, Cambridge, Massachusetts.
- Compton, S. and Hale, D. (2013). Estimating Vp/Vs ratios using smooth dynamic image warping. In *SEG Technical Program Expanded Abstracts 2013*, pages 1639–1643. Society of Exploration Geophysicists.
- 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. (2013). Plane-wave least-squares reverse-time migration. *Geophysics*.
- Dellinger, J. and Etgen, J. (1990). Wave-field separation in two-dimensional anisotropic media. *Geophysics*, 55(7):914–919.
- Deregowski, S. M. (1982). Dip-moveout and reflector point dispersal. *Geophysical Prospecting*, 30(3):318–322.
- Dines, K. A. and Lytle, R. J. (1979). Computerized geophysical tomography. *Proceedings of the IEEE*, 67(7):1065–1073.
- Dong, S., Cai, J., Guo, M., Suh, S., Zhang, Z., Wang, B., and Li, Z. (2012). Least-squares reverse time migration: towards true amplitude imaging and improving the resolution. In *SEG Technical Program Expanded Abstracts 2012*, pages 1–5. Society of Exploration Geophysicists.
- Duquet, B., Marfurt, K. J., and Dellinger, J. A. (2000). Kirchhoff modeling, inversion for reflectivity, and subsurface illumination. *Geophysics*, 65(4):1195–1209.
- Feng, J. (2004). *Rock properties inversion with Kirchhoff AVA migration/inversion*. PhD thesis, University of Alberta.
- Fomel, S. and Guitton, A. (2006). Regularizing seismic inverse problems by model reparameterization using plane-wave construction. *Geophysics*, 71(5):A43–A47.

- Franklin, J. (2013). *Computational methods for physics*. Cambridge University Press, New York, New York.
- Gaiser, J. E. (1996). Multicomponent Vp/Vs correlation analysis. *Geophysics*, 61(4):1137–1149.
- Gazdag, J. (1978). Wave equation migration with the phase-shift method. *Geophysics*, 43(7):1342–1351.
- Gazdag, J. and Sguazzero, P. (1984). Migration of seismic data by phase shift plus interpolation. *Geophysics*, 49(2):124–131.
- Gray, S. H. (1992). Frequency-selective design of the Kirchhoff migration operator. *Geophysical prospecting*, 40(5):565–571.
- Gray, S. H. (2000). Velocity smoothing for depth migration: How much is too much? In *SEG Technical Program Expanded Abstracts 2000*, pages 1055–1058. Society of Exploration Geophysicists.
- Gray, S. H. (2013). Spatial sampling, migration aliasing, and migrated amplitudes. *Geophysics*, 78(3):S157–S164.
- Gray, S. H., Etgen, J., Dellinger, J., and Whitmore, D. (2001). Seismic migration problems and solutions. *Geophysics*, 66(5):1622–1640.
- Hale, D. (1984). Dip-moveout by Fourier transform. *Geophysics*, 49(6):741–757.
- Hale, D. (2013). Dynamic warping of seismic images. *Geophysics*, 78(2):S105–S115.
- Hale, D. and Compton, S. (2013). Smooth dynamic warping. *Center for Wave Phenomena Research Report*, 764.
- Herron, D. A. (2000). Pitfalls in seismic interpretation: Depth migration artifacts. *The Leading Edge*, 19(9):1016–1017.
- Hestenes, M. R. and Stiefel, E. (1952). *Methods of conjugate gradients for solving linear systems*, volume 49. NBS.
- Kaplan, S. T., Routh, P. S., and Sacchi, M. D. (2010). Derivation of forward and adjoint operators for least-squares shot-profile split-step migration. *Geophysics*, 75(6):S225–S235.
- Keyzers, D. and Unger, W. (2003). Elastic image matching is np-complete. *Pattern Recognition Letters*, 24(1):445–453.
- Kosloff, D. D. and Baysal, E. (1983). Migration with the full acoustic wave equation. *Geophysics*, 48(6):677–687.
- Kravtsov, Y. A. and Orlov, Y. I. (1990). *Geometrical optics of inhomogeneous media*. Springer-Verlag, Berlin, Germany, 1st edition.
- Kuehl, H. and Sacchi, M. D. (2003). Least-squares wave-equation migration for AVP/AVA inversion. *Geophysics*, 68(1):262–273.
- Kutta, W. (1901). Beitrag zur näherungsweise integration totaler differentialgleichungen. *Zeitschrift für Mathematik und Physik*, 46:435–453.

- Lambare, G., Virieux, J., Madariaga, R., and Jin, S. (1992). Iterative asymptotic inversion in the acoustic approximation. *Geophysics*, 57(9):1138–1154.
- Laursen, T. A. (2013). *Computational contact and impact mechanics: fundamentals of modeling interfacial phenomena in nonlinear finite element analysis*. Springer Science & Business Media.
- Liang, L. and Hale, D. (2012). Automatic registration of PP and PS images using dynamic image warping. *Center for Wave Phenomena Research Report*, 724.
- Liner, C. L. (1990). General theory and comparative anatomy of dip moveout. *Geophysics*, 55(5):595–607.
- Liner, C. L. and Clapp, R. G. (2004). Nonlinear pairwise alignment of seismic traces. *The Leading Edge*, 23(11):1146–1150.
- Lomax, A. (1994). The wavelength-smoothing method for approximating broad-band wave propagation through complicated velocity structures. *Geophysical Journal International*, 117(2):313–334.
- Lumley, D. E. (1989). Kirchhoff prestack depth migration: Imaging conditions and amplitude recovery. In *SEG Technical Program Expanded Abstracts 1989*, pages 1336–1339. Society of Exploration Geophysicists.
- Lumley, D. E., Claerbout, J. F., and Bevc, D. (1994). Anti-aliased Kirchhoff 3-D migration. In *SEG Technical Program Expanded Abstracts 1994*, pages 1282–1285. Society of Exploration Geophysicists.
- Luo, S. and Hale, D. (2014). Least-squares migration in the presence of velocity errors. *Geophysics*, 79(4):S153–S161.
- Luo, Y. and Schuster, G. T. (1991). Wave-equation travelttime inversion. *Geophysics*, 56(5):645–653.
- Mottl, V., Kopylov, A., Kostin, A., Yermakov, A., and Kittler, J. (2002). Elastic transformation of the image pixel grid for similarity based face identification. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 3, pages 549–552. IEEE.
- Mulder, W. A. and Plessix, R. (2004). A comparison between one-way and two-way wave-equation migration. *Geophysics*, 69(6):1491–1504.
- Müller, M. (2007). *Information retrieval for music and motion*, volume 2. Springer-Verlag, Secaucus, New Jersey.
- Nemeth, T., Wu, C., and Schuster, G. T. (1999). Least-squares migration of incomplete reflection data. *Geophysics*, 64(1):208–221.
- Nickel, M. and Sonneland, L. (2004). Automated PS to PP event registration and estimation of a high-resolution V_p - V_s ratio volume. In *SEG Technical Program Expanded Abstracts 2004*, pages 869–872. Society of Exploration Geophysicists.
- Notfors, C. D. and Godfrey, R. J. (1987). Dip moveout in the frequency-wavenumber domain. *Geophysics*, 52(12):1718–1721.

- Operto, M. S., Xu, S., and Lambaré, G. (2000). Can we quantitatively image complex structures with rays? *Geophysics*, 65(4):1223–1238.
- Pishchulin, L. (2010). *Matching algorithms for image recognition*. PhD thesis, RWTH Aachen University Aachen.
- Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model. *Geophysics*, 64(3):888–901.
- Pratt, R. G. and Shipp, R. M. (1999). Seismic waveform inversion in the frequency domain, part 2: Fault delineation in sediments using crosshole data. *Geophysics*, 64(3):902–914.
- Rickett, J. and Fomel, S. (1999). A second-order fast marching eikonal solver. *Stanford Exploration Project Report*, 100:287–293.
- Rickett, J. E. and Sava, P. C. (2002). Offset and angle-domain common image-point gathers for shot-profile migration. *Geophysics*, 67(3):883–889.
- Runge, C. (1895). Über die numerische auflösung von differentialgleichungen. *Mathematische Annalen*, 46(2):167–178.
- Saad, Y. (1992). *Numerical methods for large eigenvalue problems*. St Martins Press.
- Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26(1):43–49.
- Sava, P. and Fomel, S. (2006). Time-shift imaging condition in seismic migration. *Geophysics*, 71(6):S209–S217.
- Sava, P. C. and Fomel, S. (2003). Angle-domain common-image gathers by wavefield continuation methods. *Geophysics*, 68(3):1065–1074.
- Scales, J. A. (1987). Tomographic inversion via the conjugate gradient method. *Geophysics*, 52(2):179–185.
- Schneider, W. A. (1978). Integral formulation for migration in two and three dimensions. *Geophysics*, 43(1):49–76.
- Sethian, J. A. (1996). A fast marching level set method for monotonically advancing fronts. *Proceedings of the National Academy of Sciences*, 93(4):1591–1595.
- Shearer, P. M. (2009). *Introduction to seismology*. Cambridge University Press, New York, New York, 2nd edition.
- Shragge, J. (2014). Reverse time migration from topography. *Geophysics*, 79(4):S141–S152.
- Stanton, A. and Sacchi, M. D. (2014). Least squares migration of converted wave seismic data. *CSEG Recorder*, 39(10):48–52.
- Stanton, A. and Sacchi, M. D. (2017). Elastic least-squares one-way wave-equation migration. *Geophysics*, 82(4):S293–S305.
- Stoffa, P. L., Fokkema, J. T., de Luna Freire, R. M., and Kessinger, W. P. (1990). Split-step Fourier migration. *Geophysics*, 55(4):410–421.
- Stolt, R. H. (1978). Migration by Fourier transform. *Geophysics*, 43(1):23–48.

- Symes, W. W. and Carazzone, J. J. (1991). Velocity inversion by differential semblance optimization. *Geophysics*, 56(5):654–663.
- Tang, Y. (2009). Target-oriented wave-equation least-squares migration/inversion with phase-encoded Hessian. *Geophysics*, 74(6):WCA95–WCA107.
- Tarantola, A. (1984a). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.
- Tarantola, A. (1984b). Linearized inversion of seismic reflection data. *Geophysical prospecting*, 32(6):998–1015.
- van der Baan, M. (2006). PP/PS wavefield separation by independent component analysis. *Geophysical Journal International*, 166(1):339–348.
- Vanelle, C. (2002). *Traveltime-based true-amplitude migration*. PhD thesis, University of Hamburg.
- Vidale, J. (1988). Finite-difference calculation of travel times. *Bulletin of the Seismological Society of America*, 78(6):2062–2076.
- Vinje, V., Iversen, E., and Gjøystdal, H. (1993). Traveltime and amplitude estimation using wavefront construction. *Geophysics*, 58(8):1157–1166.
- Wang, J., Kuehl, H., and Sacchi, M. D. (2004). Preconditioned least-squares wave-equation AVP migration. In *CSEG Annual Convention, CDROM*.
- Wang, J., Kuehl, H., and Sacchi, M. D. (2005). High-resolution wave-equation AVA imaging: Algorithm and tests with a data set from the western canadian sedimentary basin. *Geophysics*, 70(5):S91–S99.
- Wang, J. and Sacchi, M. D. (2006). High-resolution wave-equation amplitude-variation-with-ray-parameter (AVP) imaging with sparseness constraints. *Geophysics*.
- Xue, Z., Chen, Y., Fomel, S., and Sun, J. (2015). Seismic imaging of incomplete data and simultaneous-source data using least-squares reverse time migration with shaping regularization. *Geophysics*.
- Yan, J. and Sava, P. (2010). Analysis of converted-wave extended images for migration velocity analysis. In *SEG Technical Program Expanded Abstracts 2010*, pages 1666–1671. Society of Exploration Geophysicists.
- Yao, G. and Jakubowicz, H. (2016). Least-squares reverse-time migration in a matrix-based formulation. *Geophysical Prospecting*, 64(3):611–621.
- Yilmaz, Ö. (2001). *Seismic data analysis: Processing, inversion, and interpretation of seismic data*. Society of exploration geophysicists.
- Youzwishen, C. F. (2003). *Non-linear sparse and blocky constraints for seismic inverse problems*. MSc thesis, University of Alberta.
- Zeng, C., Dong, S., Wu, Z., Ji, J., Armentrout, D., and Wang, B. (2015). Adaptive least-squares RTM and application to freedom WAZ subsalt imaging. In *SEG Technical Program Expanded Abstracts 2015*, pages 4059–4064. Society of Exploration Geophysicists.

- Zhang, D., Wang, X., Huang, Y., and Schuster, G. (2014). Warping for trim statics. In *SEG Technical Program Expanded Abstracts 2014*, pages 3946–3950. Society of Exploration Geophysicists.
- Zhang, J., Huang, Y., Song, L. P., and Liu, Q. H. (2011). Fast and accurate 3-D ray tracing using bilinear traveltimes interpolation and the wave front group marching. *Geophysical Journal International*, 184(3):1327–1340.
- Zhu, J. and Lines, L. R. (1998). Comparison of Kirchhoff and reverse-time migration methods with applications to prestack depth imaging of complex structures. *Geophysics*, 63(4):1166–1176.

APPENDIX A

The eikonal and transport equations

The Kirchhoff operator uses Green's functions to simulate the propagation of the source-side and receiver-side wavefield. It is therefore crucial to establish a method for accurately and efficiently estimating the Green's functions. For seismic waves, the Green's functions take the following form:

$$G(\mathbf{x}, \omega) = A(\mathbf{x})e^{i\omega\tau(\mathbf{x})}, \quad (\text{A.1})$$

where ω is the angular frequency, \mathbf{x} is an arbitrary position in the subsurface, $A(\mathbf{x})$ is the amplitude (geometrical spreading factor), and $\tau(\mathbf{x})$ is the traveltime. Each Green's function therefore has two unknowns that must be solved for: $A(\mathbf{x})$ and $\tau(\mathbf{x})$. In order to solve for the unknowns, the wavefield must be expanded using a trial solution. The most common trial solution is the high-frequency WKB approximation, which assumes the wavefield $u(\mathbf{x}, \omega)$ can be written as (Bleistein et al., 2001)

$$u(\mathbf{x}, \omega) \approx \omega^\beta e^{i\omega\tau(\mathbf{x})} \sum_{j=0}^{\infty} \frac{A_j(\mathbf{x})}{(i\omega)^j}, \quad (\text{A.2})$$

where β is an integer number and A_j is the j^{th} order wave amplitude. As Bleistein et al. (2001) noted, there are several motivations for using a trial solution of this form. First, it can be rigorously shown that linearly independent solutions of the 1D Helmholtz equation take the form of equation A.2 as $|\omega|$ approaches infinity. Second, there are many known cases where the exact solution of the higher dimensional Helmholtz equation also takes the form of equation A.2. Third, it makes sense from a physical perspective. Division by $(i\omega)^j$ represents successive integrations in the time domain. Since integration is a smoothing operation, the lowest-order terms in equation A.2 represent the sharpest (highest frequency) part of the solution. Therefore, the WKB trial solution is naturally capable of incorporating the high-frequency assumption.

Away from the source, the wavefield satisfies the homogeneous Helmholtz equation

$$\mathcal{L}(\mathbf{x}, \omega)u(\mathbf{x}, \omega) = \left[\nabla^2 + \frac{\omega^2}{c(\mathbf{x})^2} \right] u(\mathbf{x}, \omega) = 0, \quad (\text{A.3})$$

where $c(\mathbf{x})$ is the velocity model, ∇^2 is the Laplace operator, and $\mathcal{L}(\mathbf{x}, \omega) = \nabla^2 + \frac{\omega^2}{c(\mathbf{x})^2}$ is the Helmholtz operator. Subbing the WKBJ trial solution into equation A.3 gives the result

$$\begin{aligned} \mathcal{L}u = \omega^\beta e^{i\omega\tau} \sum_{j=0}^{\infty} \frac{1}{(i\omega)^j} \left[\omega^2 \left\{ \frac{1}{c^2} - (\nabla\tau)^2 \right\} A_j \right. \\ \left. + i\omega \left\{ 2\nabla\tau \cdot \nabla A_j + A_j \nabla^2 \tau \right\} + \nabla^2 A_j \right], \end{aligned} \quad (\text{A.4})$$

where the dependent variables have been dropped to improve readability. In order to satisfy the homogeneous Helmholtz equation, equation A.4 must be equal to zero. The terms with different powers of ω will generally not cancel each other. Therefore, the coefficient of each power of ω must be independently equated to zero, starting with the largest possible exponent, which is $\beta + 2$. Since $\omega^{\beta+2}$ is only present in the first term of the summation, and it only occurs when $j = 0$, setting its coefficient to zero is a simple task that yields the eikonal equation

$$\frac{1}{c^2(\mathbf{x})} - (\nabla\tau(\mathbf{x}))^2 = 0. \quad (\text{A.5})$$

Note that we intentionally avoided the trivial solution in which $A_0 = 0$. This choice also guarantees that the entire first term in the summation is zero for all higher order A_j , which means only the second and third term must be considered in the remainder of this discussion. The second largest power of ω is $\beta + 1$. Of the remaining two terms, $\omega^{\beta+1}$ shows up in only the middle term, and only when $j = 0$. Therefore, equating the coefficient of $\omega^{\beta+1}$ to zero gives the (first) transport equation

$$2\nabla\tau(\mathbf{x}) \cdot \nabla A_0(\mathbf{x}) + A_0(\mathbf{x}) \nabla^2 \tau(\mathbf{x}) = 0. \quad (\text{A.6})$$

Note that when $j = 0$, the order of the third term in the summation is β . Next, we proceed to $j = 1$. When $j = 1$, the order of the second term is β and the order of the third term is $\beta - 1$. Therefore, the solution for higher order A_j can be obtained recursively using the formula

$$2\nabla\tau(\mathbf{x}) \cdot \nabla A_j(\mathbf{x}) + A_j(\mathbf{x}) \nabla^2 \tau(\mathbf{x}) = -\nabla^2 A_{j-1}(\mathbf{x}). \quad (\text{A.7})$$

In conclusion, the traveltime and amplitude components of the Green's functions can be estimated by solving the eikonal and transport equations, respectively.

APPENDIX B

The Runge-Kutta method

The Runge-Kutta method is a numerical method that was proposed by the mathematicians Runge (1895) and Kutta (1901) for estimating the solution of ordinary differential equations of the form

$$\frac{dy}{dt} = f(t, y), \quad y(t_0) = y_0, \quad (\text{B.1})$$

where $y = y(t)$ is the unknown function we want to estimate, t is the time, f is a function of y and t that can be evaluated analytically, and y_0 and t_0 are the known initial conditions (e.g., see Franklin, 2013). The Runge-Kutta method aims to solve equation B.1 by numerically integrating the solution forward in time; thus, the Runge-Kutta method is iterative. The method is initialized using the conditions given by y_0 and t_0 . The solution at each sequential iteration is then computed based on the solution at the current iteration as well as the value of the function f at multiple trial points along the step interval.

Several versions of the Runge-Kutta method have been proposed, each with a different level of accuracy. The most common version of the Runge-Kutta method is accurate to the fourth order and is typically referred to as “RK4” or the “classical Runge-Kutta method”. For a given temporal step size h , the classical Runge-Kutta method integrates from iteration n to $n + 1$ using the following scheme:

$$\begin{aligned} y_{n+1} &= y_n + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4), \\ t_{n+1} &= t_n + h, \end{aligned} \quad (\text{B.2})$$

where y_n and t_n are the solution and time at the n^{th} step, and k_1 , k_2 , k_3 , and k_4 are given by

$$\begin{aligned}
 k_1 &= f(t_n, y_n) , \\
 k_2 &= f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_1\right) , \\
 k_3 &= f\left(t_n + \frac{h}{2}, y_n + \frac{h}{2}k_2\right) , \\
 k_4 &= f(t_n + h, y_n + hk_3) .
 \end{aligned}
 \tag{B.3}$$

Thus, the classical Runge-Kutta method estimates the the value of y at the next step based on the value of y at the current step plus a weighted average four other values: k_1 , k_2 , k_3 , and k_4 . k_1 represents the update computed using the slope at the beginning of the interval. k_2 and k_3 represent the update computed using two different estimates of the slope at the midpoint of the interval. Finally, k_4 represents the update computed using the slope at the end of the interval. By averaging the four increments, low-order errors terms can be eliminated, resulting in a more accurate solution. Not surprisingly, more weight is given to the updates computed at the midpoint, as they provide the best approximation of the average solution. For many numerical problems, the function f is known only at discrete points. Therefore, the Runge-Kutta method requires an interpolation scheme for evaluating f at an arbitrary point. In this thesis, linear interpolation is used because of its low computational cost; when f depends on two or three dimensions, as is the case in seismic ray tracing, bilinear or trilinear interpolation can be used.

APPENDIX C

Conjugate gradients for least-squares

The conjugate gradient method is an algorithm for solving the following system of linear equations (Hestenes and Stiefel, 1952; Scales, 1987):

$$\tilde{\mathbf{A}}\mathbf{x} = \tilde{\mathbf{b}}, \quad (\text{C.1})$$

where $\tilde{\mathbf{A}}$ is a positive-definite matrix or operator, \mathbf{x} is a vector of model parameters, and $\tilde{\mathbf{b}}$ is a vector of data. In damped least-squares migration, the goal is to minimize a cost function of the form

$$F(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda^2\|\mathbf{x}\|_2^2, \quad (\text{C.2})$$

where \mathbf{A} is the forward modeling operator that simulates the seismic experiment, \mathbf{b} is the recorded seismic data, and λ is the regularization parameter. Equation C.2 can be expressed in the form of equation C.1 by rewriting it as

$$\begin{bmatrix} \mathbf{A} \\ \lambda \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}, \quad (\text{C.3})$$

where $\mathbf{0}$ is a vector of zeros. Thus, $\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A} \\ \lambda \end{bmatrix}$ and $\tilde{\mathbf{b}} = \begin{bmatrix} \mathbf{b} \\ \mathbf{0} \end{bmatrix}$. Damped least-squares migration can therefore be solved using the conjugate gradient method (Algorithm 4). Importantly, the conjugate gradient method does not require \mathbf{A} to be explicitly stored as a matrix. Moreover, it only requires information about the action of \mathbf{A} and \mathbf{A}^T on a general vector; consequently, \mathbf{A} and \mathbf{A}^T can be implemented as operators in the form of computer programs. Care must therefore be taken to ensure that \mathbf{A}^T is the exact adjoint of \mathbf{A} .

The same idea can be used to extend the conjugate gradient method to solve more compli-

cated cost functions. For example, preconditioned least-squares migration (e.g. see chapter 4) can be solved by concatenating the preconditioner \mathbf{P} with the forward modeling operator \mathbf{A} . Thus, for preconditioned LSM, $\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{AP} \\ \lambda \end{bmatrix}$. When implementing preconditioned LSM, it is important to make use of the following linear algebra identity: $(\mathbf{AP})^T = \mathbf{P}^T \mathbf{A}^T$. Thus, preconditioned LSM can be solved with Algorithm 4 by replacing \mathbf{A} with \mathbf{AP} and replacing \mathbf{A}^T with $\mathbf{P}^T \mathbf{A}^T$.

Algorithm 4 Conjugate gradients for least-squares

```

b ← vector of observed data
λ ← regularization parameter
Niter ← maximum number of iterations
tol ← convergence tolerance
r = d
s = g = ATr
x = 0
γ = gTg
γ0 = γ
cost0 = rTr
for j = 1 : Niter do
  t = As
  Δ = tTt + λsTs
  if Δ ≤ tol then
    BREAK
  end if
  α = γ/Δ
  x = x + αs
  r = r − αt
  g = ATr
  g = g − λx
  γ0 = γ
  γ = gTg
  cost = rTr + λxTx
  β = γ/γ0
  s = βs + g
  if √γ ≤ √γ0tol then
    BREAK
  end if
end for
return x

```

APPENDIX D

Dynamic image warping

Dynamic image warping (DIW) is a multidimensional extension of 1D dynamic time warping (DTW) (Hale, 2013). Therefore, we begin with a review of dynamic time warping. Consider two seismic traces, f and g , that contain the same reflection events, but are unaligned in time. The two sequences are approximately related by $f[i] \approx g[i + s[i]]$, where $s[i]$ is the relative shift between the i^{th} element of f and g . If the relative shifts are known, they can be applied to g to register it with f .

In order to estimate the relative shifts between the two signals, the alignment error must first be computed. Hale (2013) suggested to measure the alignment error based on the squared difference between the two signals:

$$e[i, l] \equiv (f[i] - g[i + l])^2, \quad (\text{D.1})$$

where l is the integer lag between the two signals. The alignment error $e[i, l]$ approaches zero when the lag l is approximately equal to the true shift $s[i]$. The estimated vector of relative shifts $u[i]$ can therefore be obtained by solving the optimization problem

$$u[i] = \arg \min_l \sum_1^N e[i, l[i]], \quad (\text{D.2})$$

subject to the constraints

$$u_l \leq u[i] \leq u_u, \quad r_l \leq u[i] - u[i - 1] \leq r_u, \quad (\text{D.3})$$

where u_l and u_u are lower and upper bounds on the magnitude of the shifts, and r_l and r_u are lower and upper bounds on the rate of change of the shifts (i.e. the strain). Therefore,

u_l and u_u control the minimum and maximum shifts allowed, while r_l and r_u determine how rapidly the shifts can vary.

Equation D.2 is solved using a two-step process: (i) forward accumulation of alignment errors, followed by (ii) backtracking to estimate the optimal shifts. The first step is a recursive process in which the alignment errors $e[i, l]$ are summed to obtain the “distances” $d[i, l]$. The forward accumulation procedure depends on the strain rates r_l and r_u ; for the simplest case where $r_l = r_u = 1$, the distances are obtained using the framework

$$d[i, l] = e[i, l] + \min \begin{cases} d[i-1, l-1] \\ d[i-1, l] \\ d[i-1, l+1] \end{cases} . \quad (\text{D.4})$$

It is generally desirable to limit the strain to less than one. Moreover, decreasing the strain increases the smoothness of the estimated shifts, thereby improving the stability of the algorithm (Hale, 2013). The most straightforward approach for limiting the strain is to sample the lags at fractions of a time step. However, this is not ideal because it increases the computational cost of the algorithm (Hale, 2013). A more efficient strategy for limiting the strain is to allow the lag to vary by only one time step over a period of several samples. For example, to limit the strain to 50%, any change in lag for one sample should be followed by no change in lag for the next sample. The distances for a 50% strain are therefore computed using the recursive framework

$$d[i, l] = e[i, l] + \min \begin{cases} d[i-2, l-1] + e[i-1, l-1] \\ d[i-1, l] \\ d[i-2, l+1] + e[i-1, l+1] \end{cases} . \quad (\text{D.5})$$

This strategy can be used to limit the strain to any rational number of the form $1/b$, where b is an integer.

After the distances have been forward accumulated, the optimal shifts can be estimated through a process called backtracking. Backtracking is initialized by selecting the optimal shift $u[N]$ for the last sample in the trace, where N is the number of samples in the trace. Since $d[N, l]$ represents the accumulated alignment error for the last sample in the trace, the optimal shift $u[N]$ for the last sample is obtained by choosing the lag l that minimizes $d[N, l]$. The remaining shifts are then estimated using a recursive framework. For the simplest case where $r_l = r_u = 1$, the shifts are estimated by recursively solving

$$u[i-1] = \arg \min_{l \in \{u[i]-1, u[i], u[i]+1\}} d[i-1, l] \quad (\text{D.6})$$

for each sample i in the trace. Equation D.6 states that the optimal shift at sample $i - 1$ is either $u[i] - 1$, $u[i]$, or $u[i] + 1$, depending on which one yields the smallest “distance”. Equation D.6 can be modified for strains less than 100% by shifting sequences in blocks of two or more samples (Hale, 2013).

Dynamic time warping can be used to register multidimensional images by performing DTW on one row or column of the image at a time. However, the resulting shifts are only guaranteed to be smooth and continuous along the dimension in which DTW was performed; moreover, the shifts are generally discontinuous and rapidly varying along the other dimensions. Hale (2013) showed that a substantial improvement can be achieved by smoothing the alignment error array across each dimension of the image prior to performing DTW; this modified approach is referred to as dynamic image warping. Hale (2013) recognized that the forward accumulation algorithm represents integration and is therefore a smoothing operation. The forward accumulation algorithm is therefore a natural choice for smoothing the alignment errors. For the simplest case in which the image has only 1 dimension, the smooth alignment error array \tilde{e} is given by

$$\tilde{e}[i, l] = \tilde{e}_f[i, l] + \tilde{e}_r[i, l] - e[i, l], \quad (\text{D.7})$$

where $\tilde{e}_f[i, l]$ is obtained by forward accumulating the alignment errors and $\tilde{e}_r[i, l]$ is obtained by reverse accumulating the alignment errors. The smoothing operation can be extended to multidimensional images by applying equation D.7 to each dimension of the image. Importantly, the behavior of the multidimensional DIW operator can be controlled by imposing different strain limits (r_l and r_u) across each dimension of the image volume. Thus, by smoothing each dimension of the alignment error array using appropriate strain limits, it is possible to recover a set of shifts that is continuously smooth across each dimension, despite the fact that DTW is applied across only one dimension.

APPENDIX E

Software

E.1 The Julia programming language

The Julia programming language offers high-level syntax with relatively low-level performance. In addition to this, the language has an extensive library of mathematical functions and built-in functionality for distributed memory parallelism, thereby making the language well-suited for solving research-level numerical problems. Moreover, Julia's simple syntax allows for fast-paced development of prototype software that is capable of solving large problems efficiently. For these reasons, this thesis relied on the Julia language for the development of both large-scale programs and small-scale tests.

E.2 Open-source software

The software produced during the creation of this thesis was written in the Julia programming language. Some of this software has been made available as part of the open-source seismic processing package named *SeismicJulia*, which is maintained by the Signal Analysis and Imaging Group at the University of Alberta. The *SeismicJulia* package can be found by visiting <https://github.com/SeismicJulia>. Table E.1 describes the main programs developed and used during the creation of this thesis.

Function	Description
kdmig2d	Kirchhoff depth migration for two-dimensional models. This function represents the forward and adjoint Kirchhoff operator for performing seismic modeling and migration. A triangular anti-aliasing filter is used to suppress artifacts arising from Kirchhoff operator aliasing. Parallelism is implemented by distributing the shot gathers over all of the available worker processes.
WFC	The wavefront construction method. This function uses the wavefront construction method for estimating seismic Greens functions. Parallelism is implemented by distributing the source locations over all of the available worker processes.
SDIW	Smooth dynamic image warping. This function uses smooth dynamic image warping to estimate the sequence of shifts that optimally aligns one image with a reference image. This function can also directly apply those shifts to the image to register it with the reference image.
DistributeArray	Distribute an array over multiple processes. This function distributes a multidimensional array over all of the available worker processes by breaking that array into evenly-sized sub-arrays along one of its dimensions. This function should be used in combination with FetchArray.
FetchArray	Fetch/retrieve the components of an array that have been distributed across multiple processes. This function fetches the sub-arrays of a larger array that have been distributed over all of the available worker processes. This function should be used in combination with DistributeArray.
SendTo	Send arguments from the master process to worker processes. This function takes a list of arguments that are defined on the master process and sends them to user-defined worker processes.
ReadFile	Read a binary file from disk. This function reads a binary file containing an array from disk. It should be used in combination with WriteFile.
WriteFile	Write a binary file to disk. This function saves an array to disk as a binary file. It should be used in combination with ReadFile.
CopyFile	Copy a binary file on disk. This function generates a copy of a binary file that is currently saved on disk.
Integrate!	Perform causal integration or acausal integration. This function performs causal integration (forward operator) and acausal integration (adjoint operator) along one dimension of a multidimensional array.
TriFilter	Triangular filter. This function applies an n-point triangular filter across one dimension of a multidimensional array.
bilinear	Bilinear interpolation. This function uses bilinear interpolation to interpolate the velocity model between grid points.

Table E.1: Description of the main Julia functions developed during the creation of this thesis.