Imaging of Scattered Wavefields in Passive and Controlled-source Seismology

Thesis by
Abdullah Ahmed AlTheyab

In Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy

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

December, 2015
The thesis of Abdullah Ahmed AlTheyab is approved by the examination committee

Committee Chairperson: G. T. Schuster
Committee Member: Fan-Chi Lin
Committee Member: Daniel B. Peter
Committee Member: Ying Wu
Committee Member: Sherif M. Mahmoud
Seismic waves are used to study the Earth, exploit its hydrocarbon resources, and understand its hazards. Extracting information from seismic waves about the Earth’s subsurface, however, is becoming more challenging as our questions become more complex and our demands for higher resolution increase. This dissertation introduces two new methods that use scattered waves for improving the resolution of subsurface images: natural migration of passive seismic data and convergent full-waveform inversion.

In the first part of this dissertation, I describe a method where the recorded seismic data are used to image subsurface heterogeneities like fault planes. This method, denoted as natural migration of backscattered surface waves, provides higher resolution images for near-surface faults that is complementary to surface-wave tomography images. Our proposed method differ from contemporary methods in that it does not (1) require a velocity model of the earth, (2) assumes weak scattering, or (3) have a high computational cost. This method is applied to ambient noise recorded by the US-Array to map regional faults across the American continent. Natural migration can be formulated as a least-squares inversion to further enhance the resolution and the quality of the fault images. This inversion is applied to ambient noise recorded in Long Beach, California to reveal a matrix of shallow subsurface faults.

The second part of this dissertation describes a convergent full waveform inversion method for controlled source data. A controlled source excites waves that scatter from subsurface reflectors. The scattered waves are recorded by a large array of geophones. These recorded waves can be inverted for a high-resolution image of the subsurface by FWI, which is typically convergent for transmitted arrivals but often does not converge for deep reflected events. I propose a preconditioning approach that extends the ability of FWI to image deep parts of the velocity model, which significantly improves the chances for finding hydrocarbon deposits.
ACKNOWLEDGEMENTS

Nothing that could ever be said describes my gratitude for my wife Ebtihal, and my sons Rawaf, Wafi, and Lafi. I’m also grateful for my advisor Gerard Schuster. I could not have reached this far if it were not for their patience, tolerance, and generosity.

I’m thankful to my committee member Professor Fan-Chi Lin, who exposed me to ambient-noise seismology and how it is used to study regional geology. Much of my interest in velocity estimation started when working with Yi Luo during a summer internship at Saudi Aramco. As an Aramco sponsored student, I had the privilege of receiving the personal support of Abdullah AlNuaim (God bless his soul), Panos Kelamis, Eman AlYami, Abdullah AlQhahtani, and Terry Barr throughout my graduate studies. I’m also thankful for the support of KAUST staff Veronika Pelletier and Luai Lamfon. Computing is a large part of my research. For that, I had the IT support of Mohammed Naseemuddin, Craig Kapfer, Waleed M. Alharbi, and Saber Feki.

I thank my teachers who are too many to list, of whom I especially thank Chris Liner, Jon Claerbout, Bindo Biondi, Albert Tarantola, Tariq AlKhalifah for teaching me much about seismic exploration and seismic inversion.

Graduate studies are difficult without good friends for that I thank my friends at Stanford Exploration Project (SEP) and the Center for Subsurface Imaging and Fluid Modeling (CSIM). Finally, I thank the sponsors of our research for their funding and support.
**TABLE OF CONTENTS**

Examination Committee Approval 2

Copyright 3

Abstract 4

Acknowledgements 5

**I Imaging Near-surface Heterogeneities using Backscattered Ambient Noise** 9

1 Imaging Near-surface Heterogeneities by Natural Migration of Back-scattered Surface Waves 12

1.1 Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 12
1.2 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 13
1.3 Theory . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 14
   1.3.1 Migration of Backscattered Surface-waves . . . . . . . . . . . . . . . . . . 14
   1.3.2 Natural Migration of Backscattering in Passive Data . . . . . . . . . . . . 16
1.4 Numerical Examples . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
   1.4.1 Ambient-Noise Simulation . . . . . . . . . . . . . . . . . . . . . . . . . . . . 18
   1.4.2 Natural Migration Procedure . . . . . . . . . . . . . . . . . . . . . . . . . . 21
1.5 Application to US-Array Passive Data . . . . . . . . . . . . . . . . . . . . . . . . 33
1.6 Conclusions . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 33
1.7 Appendix I: Effect of Elastic Heterogeneities on Migration Images . . . . . . . . . 35
1.8 Appendix II: Backscattered Events in Ambient-noise Cross-correlations . . . . . . 37

2 Natural Inversion of Scattered Surface Waves with Application to Rayleigh Waves in Long-Beach Passive Data 42

2.1 Abstract . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
2.2 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 42
2.3 Method . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 44
2.4 Application to Long-Beach Data . . . . . . . . . . . . . . . . . . . . . . . . . . . 46
II Full-Waveform Inversion of Seismic Reflections

3 Time-domain Incomplete Gauss-Newton Full-waveform Inversion

4 Efficient Wavefront Picking for 3D Tomography and Full-waveform Inversion

5 Full-waveform Inversion of Reflections with Inaccurate Starting Models
Part I

Imaging Near-surface Heterogeneities using Backscattered Ambient Noise
Introduction to Part I

In this part, I present a migration method that does not require a velocity model to migrate backscattered surface waves to their projected locations on the surface. This migration method uses recorded Green’s functions along the surface instead of simulated Green’s functions. Therefore, it is referred to as natural migration. The key assumptions are that the scattering bodies are within the depth interrogated by the surface waves, and the Green’s functions are recorded with dense receiver sampling along the free surface. In Chapter 1, The natural imaging formulas are derived for both active source and ambient-noise data, and computer simulations show that natural migration can effectively image near-surface heterogeneities with typical ambient-noise sources and geophone distributions. We also present the results of applying natural migration to Long-Beach and US-Array passive data. The migration images highlight known discontinuities identified in surface-wave tomograms and correlate well with some of the prominent geological boundaries at two different scales: (1) the tectonic scale such as the edge of the Atlantic Plain Province in southeastern US and (2) the regional scale structure under Long Beach, California. The migration images provide complementary high-wavenumber information to the smoother surface-wave tomograms and can be used to refine the tomographic models.

Significance of the Contributions

With this work, we enable imaging near-surface faults using backscattered surface waves, while avoiding the need for prior elastic models (e.g. velocity and anisotropy parameters). The proposed approaches are robust for imaging near-surface heterogeneities using ambient noise. This is particularly important for imaging faults below populated areas where passive data can be acquired cheaply.

Organization of Chapters

This part of the dissertation has two chapters. The first chapter shows how we can use backscattering in ambient noise to image near-surface faults. We derive the migration equation and test it
on a synthetic model, assuming ideal conditions for noise distribution and survey geometry. With this in mind, we show some preliminary results for application to passive data recorded by the US-Array.

As we apply the method to the Long-Beach dataset, we observed a footprint of the noise distribution and survey geometry. In addition, some known geologic features were not observed in the migration image due to the limited resolution. To resolve these issues, we proposed a least-squares inversion in the second chapter. We estimate impulse responses (point-spread functions) of the natural migration images which demonstrate the effect of the noise and the geometry distribution in addition to the limitation on resolution. Using a least-square inversion scheme, we can minimize the footprints and significantly enhance the resolution of the natural migration images. The final images map known features with remarkable resolution. In addition, they are consistent with previous results from ambient noise tomography.
Chapter 1

Imaging Near-surface Heterogeneities by Natural Migration of Back-scattered Surface Waves

1.1 Abstract

We present a migration method that does not require a velocity model to migrate backscattered surface waves to their projected locations on the surface. This migration method, denoted as natural migration, uses recorded Green’s functions along the surface instead of simulated Green’s functions. The key assumptions are that the scattering bodies are within the depth interrogated by the surface waves, and the Green’s functions are recorded with dense receiver sampling along the free surface. This natural migration takes into account all orders of multiples, mode conversions, and non-linear effects of surface waves in the data. The natural imaging formulas are derived for both active source and ambient-noise data, and computer simulations show that natural migration can effectively image near-surface heterogeneities with typical ambient-noise sources and geophone distributions.
1.2 Introduction

Backscattered surface waves can be imaged for the near-surface heterogeneities (Snieder, 1986). The typical strategy is to 1) linearize the relation between the scattered data $d$ and the model perturbation $m$ (i.e., heterogeneities map) under the Born approximation $d = Lm$, and then 2) find the approximate solution by either an iterative optimization method (Riyanti, 2005; Campman and Riyanti, 2007; Kaslilar, 2007) or by applying the adjoint (Snieder, 1986; Blonk et al., 1995; Campman et al., 2005; Yu et al., 2014) of the modeling operator $L^\dagger$ to the scattered data $d$ to get the migration image $m^{mg} = L^\dagger d$. In all cases, the two key assumptions are that the velocity model (typically, just the smooth component of the surface-wave velocity distribution) is known and the weak-scattering approximation is invoked. For many practical applications, the background velocity model is assumed to be a layered medium. This methodology has found a growing number of uses in earthquake, exploration, and engineering seismology (Snieder, 1986; Blonk et al., 1995; Wijk, 2003; Campman et al., 2005; Riyanti, 2005; Campman and Riyanti, 2007; Kaslilar, 2007).

There are two significant limitations with the above surface-wave inversion methods: the Born approximation is invalid if there are strong velocity contrasts, and the wavefields in complex regions of Earth cannot be accurately modeled without prior knowledge of the elastic parameters of Earth. In either case, the resulting image can contain significant errors. To eliminate these problems, we present a surface-wave imaging method named natural migration (Schuster, 2002; Brandsberg-Dahl et al., 2007; Sinha et al., 2009; Xiao and Schuster, 2009; Hanafy and Schuster, 2014) that does not require the Born approximation or the need to know the velocity model. Instead of computing the Green’s functions with an assumed background velocity model, we estimate the actual Green’s functions $G(x_g|x_s)$ of the earth at the geophone locations $x_g$ for either an active point source at $x_s$, or a virtual point source at $x_s$ computed by cross-correlation and stacking of ambient noise records. These estimated Green’s functions contain all of the effects of scattering, anisotropy, and higher-order modes in the data eliminating the need for compute-intensive elastic modeling operations (Schuster, 2002). The Green’s functions are then used to create the exact modeling operator $L$ that emulates the data, so there is no need to know the velocity model to find
\[ \mathbf{m}^{\text{mig}} = \mathbf{L}^\dagger \mathbf{d} \]. However, the trial image points are restricted to be at the surface, so the migration image provides the scatterer distributions projected from depth to their surface locations. The limitation is that the sampling of the migration image depends on the density of receiver arrays along the free surface. This limitation is mitigated by the recent availability of dense seismic arrays, such as US-Array and the Long-Beach array (Hand, 2014). Our synthetic simulations show that natural migration of both active and passive source data can provide accurate images of the projected distributions of scatterers onto the Earth’s surface as long as the scatterers are within the depth that is sensitive to surface waves.

In this chapter, we derive the natural migration equation, and apply the proposed imaging method to synthetic data. The migration equation starts with the Lippmann-Schwinger equation, but does not assume the Born approximation. Instead of assuming a smooth background model, it uses the empirical Green’s functions recorded in the data. Thus, the migration equation is valid for any type of strong velocity contrast. The next section assesses the effectiveness of natural migration on 3D elastic data generated for a simple fault model. The last section provides a summary of our work.

1.3 Theory

1.3.1 Migration of Backscattered Surface-waves

For an inhomogeneous 3D elastic medium, the scattered wavefield can be represented by (Hudson and Heritage, 1981; Snieder and Nolet, 1987)

\[ u_i(x_s, x_r) = \int_V \gamma_l(\omega) \left\{ \Delta \rho(x) \delta_{pk} \omega^2 G_{lp}(x|x_s) - \Delta c_{kjpq}(x) \frac{\partial}{\partial x_q} G_{lp}(x|x_s) \frac{\partial}{\partial x_j} \right\} G_0^{ki}(x|x_r) \, d^3x, \]

(1.1)

where the particle-displacement vector is given by \( u_i(x_s, x_r) \), \( x_s \) and \( x_r \) are, respectively, the source and the receiver positions. The subscript indices indicate one of the components of the displacement-vector wavefield, where for example \( i \) have the values 1, 2, and 3 for respectively
vertical, horizontal-x, horizontal-y components. Einsteinian summation over dummy indices is assumed. The variable $\omega$ represents the angular frequency, $G_{ij} (x|x_s)$ is the monochromatic Green’s tensor for the $j$-th particle-component point source at the position $x_s$ and the $i$-th component receiver at $x$, and $G_{ij}^0 (x|x_r)$ is the transmitted-wave Green’s tensor (i.e. it only contains the transmitted wavefield without backscattering). The dependence of wavefield variables on the harmonic frequency $\omega$ is silent. Here, $\Delta c_{kjspq} (x)$ represents the arbitrary distribution of elastic perturbations, $\Delta \rho$ is the distribution of the density perturbations, $\delta_{pk}$ is the Kronecker delta function which has the value one when $p = k$ and zero otherwise, and $\gamma_l (\omega)$ is the source-wavelet spectrum. The volume integral in equation [1.1] is over the model volume where perturbations do not coincide with the source or receiver locations. We can derive the migration equation as (see Appendix [1.7] for details)

$$m (x) = \iiint \omega^2 (1 + \delta_{pk}) \gamma_l (\omega) G_{lp} (x|x_s) G_{ki}^0 (x|x_r) u_i (x_s, x_r) d x_s d x_r d \omega.$$

(1.2)

This migration equation can be used to image density and elastic-parameter perturbations.

For surface-waves, the image $m (x)$ in equation [1.2] can be evaluated on the free-surface to produce 2D images which are projections of the scatterer’s locations onto the free surface (Snieder 1986, Blonk and Herman 1994, Campman et al. 2005). These projections are appropriate for scatterer’s at shallow depths that are detectable by surface waves. We shall denote these projections of scatterer’s on the surface as migration shadows. Due to the variable sensitivity with depth for different frequencies, the migration images can be separated according to different frequency bands:

$$m (x, \omega') = \iiint (1 + \delta_{pk}) \beta_{\omega'} (\omega) \gamma_l (\omega) G_{lp} (x|x_s) G_{ki}^0 (x|x_r) u_i (x_s, x_r) d x_s d x_r d \omega,$$

(1.3)

where the band-pass filter $\beta_{\omega'} (\omega)$ is a function designed to smoothly taper the data and Green’s tensors around the central frequency $\omega'$. The $\omega^2$ is considered part of $\beta$ for brevity. A further
decomposition is based on the modes for propagation of incident and scattered wavefields. In addition, equation 3 can be simplified by ignoring the amplitude scaling factor \((1 + \delta_{pk})\). For example, if we consider only Rayleigh-wave scattering due to the Rayleigh-wave incidence wavefield, the migration equation becomes for \(p = k = l = i = 1\)

\[
m_{11}(\mathbf{x}, \omega') = \int \int \int \beta_{\omega'}(\omega) \gamma_{l}(\omega) G_{11}(\mathbf{x}|\mathbf{x}_s) G_{11}^{0}(\mathbf{x}|\mathbf{x}_r) u_1(\mathbf{x}_s, \mathbf{x}_r) \mathbf{x}_s d\mathbf{x}_r d\omega.
\]

(1.4)

This equation is applicable to active-source data. However, special care must be taken because the interpretation of migration shadows is not appropriate for body waves that might not travel along the surface. Therefore, body-wave arrivals must be removed from the data prior to migration. In addition, the source wavelet \(\gamma_l(\omega)\) must be estimated. Fortunately, virtual gathers computed from passive data cross-correlation tend to be exclusively dominated by only surface waves and the phase of the source wavelet \(\gamma_l(\omega)\) is zero after ambient-noise cross-correlation. Therefore, this method is applicable to surface waves in virtual gathers without the need for muting body-wave arrivals. In the following section, we derive the migration equation for passive data.

### 1.3.2 Natural Migration of Backscattering in Passive Data

For surface waves, the time-symmetric ambient noise cross-correlation tensor \(C_{ij}\) is defined as

\[
C_{ij}(\mathbf{x}_A|\mathbf{x}_B, \omega) \overset{\text{def}}{=} \frac{1}{2} \left( \langle d_i(\mathbf{x}_A, \omega) d_j(\mathbf{x}_B, \omega) \rangle + \langle d_i(\mathbf{x}_A, \omega) \overline{d_j(\mathbf{x}_B, \omega)} \rangle \right),
\]

(1.5)

where \(d_i(\mathbf{x}_A, \omega)\) and \(d_i(\mathbf{x}_B, \omega)\) are the \(i\)-th components of the observed particle-displacement at the locations \(\mathbf{x}_A\) and \(\mathbf{x}_B\), respectively. This cross-correlation tensor is related to the Green’s function by the interferometric equation (Weaver and Lobkis, 2004; Snieder, 2004)

\[
\mu \omega C_{ij}(\mathbf{x}_A|\mathbf{x}_B) = \overline{G_{ij}(\mathbf{x}_A|\mathbf{x}_B)} - G_{ij}(\mathbf{x}_A|\mathbf{x}_B),
\]

(1.6)
where i = \sqrt{-1} and \mu is a scalar factor under the far-field approximation, and it depends on the geometrical configuration of the stations, the mode of propagation and the propagation velocity, and the distribution of noise sources. If the scalar factor is ignored, the amplitudes of the empirical Greens function will not be correct. We will disregard this scalar factor in subsequent derivations, keeping in mind that the dynamic information (i.e. amplitudes) in the migration images may be imprecise. Nevertheless, the geometric information of the migration images (i.e. locations of scatterers and fault maps) is still reliable.

If we consider an equation that has a structure similar to that of equation 1.4 but replacing the Green’s functions with ambient noise cross-correlations, we get

\[
\eta_{pk}(x, \omega') \overset{\text{def}}{=} - \iiint \omega^2 \beta_{\omega'}(\omega) \gamma_l(\omega) C_{lp}(x|x_s) C_{0}^{0}_{ki}(x|x_r) u_i(x_s, x_r) \, dx_s dx_r d\omega,
\]

where \(C_{0}^{0}_{ki}\) are the correlations containing only the direct surface waves (i.e. backscattering events are muted). By substituting the right hand side of equation 1.6 into equation 1.7 we get four terms

\[
\eta_{pk}(x, \omega') = \iiint \beta_{\omega'}(\omega) \gamma_l(\omega) \{ G_{lp}(x|x_s) G_{0}^{0}_{ki}(x|x_r) - G_{lp}(x|x_s) G_{0}^{0}_{ki}(x|x_r) - G_{lp}(x|x_s) G_{0}^{0}_{ki}(x|x_r) + G_{lp}(x|x_s) G_{0}^{0}_{ki}(x|x_r) \} u_i(x_s, x_r) \, dx_s dx_r d\omega.
\]

Note that the fourth term is the same as the migration equation 1.4 for active source data, and the first term is equivalent to the fourth term considering the time symmetry in the cross correlations. The contribution of the other two terms to the migration image \(\eta_{pk}(x, \omega')\) can be eliminated
by muting relevant portions of the data $u_i(x_s,x_r)$, as will be demonstrated with the numerical examples in the following sections.

1.4 Numerical Examples

1.4.1 Ambient-Noise Simulation

We will use numerical models and synthetic ambient noise to visualize and analyze how recorded ambient noise can be migrated to produce an image of subsurface heterogeneities, in similar fashion to [Thorbecke and Draganov (2011)]. The 3D model in Figure 1.1 is used to test the effectiveness of natural migration in imaging buried faults near the surface. The model has one shallow layer and one deeper layer, with shear velocities of 577 m/s and 1154 m/s, respectively. The thickness of the shallow layer changes 48 m due to fault displacement, where the shallow layer is 15 m thick at the up-thrown side of the fault and 63 m thick along the down-thrown side. Random scatterers are placed throughout the medium to generate realistic scattering as often observed in field records. The P-wave velocity model is determined by $V_p = \sqrt{3}V_s$, and density is constant with the value $\rho = 2.0$ kg/m$^3$. The grid spacing of the model is 3 m in each direction.

Random noise sources are excited around the model to generate band-limited random noise, as shown in Figure 1.1 (top). Random time functions are generated with a uniform amplitude distribution between -1 and 1, and then we band-pass filter the time functions to the desired range of frequencies. For the numerical examples in this paper, source functions are band-passed between 1 and 20 Hz, and the time interval is 0.3 ms. Synthetic ambient noise is generated by staggered-grid finite-difference simulations of the isotropic elastic wave-equation [Virieux, 1986] with a free-surface boundary condition [Gottschämmmer and Olsen, 2001].

The z-component of the simulated ambient-noise is recorded along the surface using an array of geophone stations. To ensure the validity of the far-field approximation used in the previous derivations, the noise-sources are randomly distributed 30 m away from the recording array during the simulations. The random locations of sources insures that the noise has uniform angular coverage, which will enhance the accuracy of the empirical Green’s function computed by
cross-correlation. In cases of non-uniform angular coverage, scatterers within the random media can enhance the angular coverage and, therefore, improve the accuracy of the empirical Green’s functions (Larose et al. 2006).

About 34 minutes of noise were simulated and recorded, where the total recording time is divided into smaller 1.5-second-long records. Each record is simulated independently with 5 z-component noise sources emitting noise simultaneously; each source has a different signature as described above. The number of noise sources and the length of each record are empirically determined to enhance the quality of the empirical Green’s functions computed by cross-correlating the noise records. A similar quality can be achieved in field passive seismic experiments by recording for a long time (e.g. several days for the Long-Beach data).

See Appendix 1.8 which illustrates the detection of backscattering using cross-correlation for a simple 2D model. Backscattered events are often difficult to observe in ambient noise cross-correlations due to the overwhelming noise and cross-correlation artifacts. However, comparing synthetic data in Figure 1.2A with ambient-noise cross-correlation records in Figure 1.2B, we can observe backscattered events in both figures. Migrating the backscattered events in one virtual gather will partially image the subsurface faults and heterogeneities, and the images will be consistent from one shot to another. This means that when the images from different gathers are stacked they constructively interfere to form a coherent image of the heterogeneities.

Spurious events overlap with the backscattered events in the cross-correlations in Figure 1.2B. Migrating the spurious events in the gather will introduce noise into the migration image. However, such noise is unlikely to be consistent from virtual shot to another. Therefore, when the images are stacked noise will destructively interfere and be attenuated. We assume that this is also the case for deep virtual reflections in the cross-correlations.
Figure 1.1: (Top) A 3D model used to synthesize ambient noise from sources placed around the edges of the model, which are denoted by the star shapes. The point sources (monopoles with vertical velocity component) are randomly distributed on the free surface around 30 m away from the recording array. (Bottom) The deeper half of the model, where the top view shows a buried fault and the change of velocity across the fault.
Figure 1.2: Fault-related backscattering as it appears on synthetic data (A) and ambient-noise cross-correlations (B). The yellow box highlights backscattering from the buried fault. Note that gray-scale images in the paper are variable density plots where extreme positive values are shown as white, negative values are in black, and the gray color indicates zero value.

1.4.2 Natural Migration Procedure

Here, we describe the steps used to compute the natural migration image from ambient-noise records, using the passive-data natural migration equation (i.e. equation 1.7). First, the recorded noise is spectrally normalized \[ \gamma_l(\omega) \approx 1 \] (i.e. amplitudes in the frequency domain are set to one) and then cross-correlated to generate the empirical Green’s functions. The normalization and cross-correlation produce spectrally balanced Green’s functions with zero-phase wavelet, so that it is suitable to assume \[ \gamma_l(\omega) \approx 1 \] in the migration equation 1.7. Next, we normalize the Green’s functions in a virtual source gather by the maximum absolute value of the
amplitudes in the gather. We observe that this normalization partially corrects for the amplification effect that depends on the velocity near the virtual source.

The second step is muting and wavefield separation. The samples near the zero lag are muted to avoid near-field strong artifact\(^1\). We empirically find that muting one period (estimated roughly from the transmitted waves as shown in Figure 1.3A) is sufficient to avoid strong artifacts near sources and receivers.

To compute the natural migration image, we need to separate transmitted (i.e. direct) and scattered wavefields in the empirical Green’s functions so that

\[
C_{ki} (x_A|x_B) = C_{ki}^{\text{trans}} (x_A|x_B) + C_{ki}^{\text{scat}} (x_A|x_B),
\]

(1.9)

where \(C_{ki}^{\text{trans}} (x_A|x_B)\) contains the transmission events, and \(C_{ki}^{\text{scat}} (x_A|x_B)\) contains only the scattered events. This separation can be performed by muting, where an average velocity \(v_{\text{avg}}\) and the period \(T\) of the direct arrivals are estimated and then used to design the muting function,

\[
\tau_{\text{mute}} (x, x_s) = \frac{|x_s - x|}{v_{\text{avg}}} + T
\]

(1.10)

where the transmitted events are above the muting function and the scattered waves are below the function as shown in Figures 1.3B and 1.3C, respectively. Smooth tapering is recommended when applying the mute.

\(^1\)This is related to singularities in the integration domain
Figure 1.3: Muting of cross-correlated traces (A) to separate the direct (B) and scattered (C) events for natural migration. The first period (0.1 second) is muted in the traces prior to migration.
The variables in the migration equation \[1.7\] are defined using the separated wavefields as follows:

\[
\begin{align*}
    u_i(x_s | x_r) &= \sum_k C_{ki}^{\text{scat.}}(x_s | x_r), \\
    C_{ki}^{0}(x | x_r) &= C_{ki}^{\text{trans.}}(x | x_r), \\
    C_{lp}(x | x_s) &= C_{lp}^{\text{trans.}}(x | x_s) + C_{lp}^{\text{scat.}}(x | x_s).
\end{align*}
\]

(1.11) \hspace{1cm} (1.12) \hspace{1cm} (1.13)

Note the summation in equation \[1.11\] over the \(k\) index. In our numerical example, however, we do not record horizontal components and therefore we evaluate \(u_1(x_s | x_r) = C_{11}^{\text{scat.}}(x_s | x_r)\) only.

Now, we compute the natural migration image using equation \[1.7\] and substitute the separated transmission and scattered events respectively into the empirical Green’s functions and backscattered data (i.e. \(u_i\)). Figure 1.4 shows the natural migration result for the simple example above, where the migration image delineates the buried fault. The fault image is a positive and negative doublet where the positive values (white color) identifies the slow part of the model and the negative (black color) identifies the fast side of the fault. The rest of the image has near zero values (gray) or filled with minor noise. Such noise is related to imperfect reconstruction of the Green’s function, and it can be reduced with longer passive data recording, and a more uniform distribution of noise sources.

![Projected Fault-line](image)

Figure 1.4: Natural migration image overlaying the 3D model, where the gray-scale image delineates the fault projected to the surface.

Surface waves travel laterally along the free-surface, where the horizontal boundaries between
the layers do not cause backscattering. Therefore, the proposed method can not directly image changes in the medium along the depth axis, like boundaries between layers. This is useful when the objective of the seismic experiment is to image heterogeneities like faults and scatterers with disregard to the layering details in the subsurface.

Natural migration images are evaluated at geophone stations on the free surface at $z = 0$. Therefore, they do not directly indicate the depth of the anomalies. To develop a sense of depth, we design many bandpass filters $\beta_{\omega'}$ with increasing peak frequency

$$\omega' = \{4.2\text{Hz}, 5.6\text{Hz}, 6.3\text{Hz}, 8.3\text{Hz}, 10.4\text{Hz}, 12.5\text{Hz}, 15.3\text{Hz}, 19.5\text{Hz}, 24.3\text{Hz}, 30\text{Hz}\},$$

where the different frequency bands can be associated with different depth ranges. The filters are chosen to cover the spectrum of the data, using any filter of choice. Frequencies are avoided that violate the Nyquist sampling interval associated with geophone spacing.

Figure 1.5 illustrates a set of bandpass filters in the time and frequency domains. By solving the elastic wave-equation using the average velocity of the first layer, we can estimate the sensitivity of the Rayleigh waves to different depths of velocity anomalies, as shown in Figure 1.6. For filters 1 to 3, most of the energy is concentrated at depths shallower than that of the fault (15 m) and is unable to detect the fault. Filters with a lower range of frequencies, on the other hand, show the significant sensitivity of low-frequency Rayleigh waves to deeply buried velocity heterogeneities.
Figure 1.5: (Top) Band-pass time-domain filters used for migration and (Bottom) their corresponding spectra.
Figure 1.6: Snapshots of laterally traveling Rayleigh waves using the band-pass filters as the source signatures. Note, the change of depth of penetration as a function of frequency contents of bandpass range of frequencies.

We also compute a collection of migration images for different frequency ranges that collectively give an indication of relative depths and sizes of the detected anomalies. Figure 1.7 show the natural migration images as a function of the filter’s range of frequencies. High-frequency filters 1 to 3 do not detect the fault, due to the Rayleigh wave’s shallow depth of penetration. The remaining low-frequency images detect the fault. Figure 1.8 shows a cross-section migration image for $y = 150$ m, where the fault is seen clearly using filters from 4 to 9.

The same numerical experiment was repeated twice for the same 3D model but with fault depths of 21 m and 30 m, and the corresponding natural migration images are shown in Figures 1.9 and 1.10, respectively. With increasing depth, the image of the fault becomes confined to lower frequencies. In the natural migration images, different frequencies detect the fault with different spatial resolution, where the filters with lower-frequency ranges show the fault with lower resolution. This demonstrates the trade-off between depth of penetration and lateral-resolution, which depends on the relationship between frequency and depth of penetration of Rayleigh waves.
In general, the effectiveness of natural migration is limited by the strength of the back-scattered surface waves. This subject is covered by Chai et al. (2012) and Chai et al. (2014). We recommend using a synthetic data test, as the one demonstrated above, for each case where natural migration is applied to assess the abilities and the limits of the method in the given geological settings, noise distribution, and survey geometry. In addition, applications of the method should be in conjunction with other independent methods for studying the subsurface, like surface wave tomography (Lin et al., 2008). This is to validate the interpretation of the natural migration images, and to reject possible false positives generated by uncorrelated noise and imperfect reconstruction of empirical Green’s functions.
Figure 1.7: Natural migration images for different filters, where the decreasing dominant frequency of the filter acts as pseudo-depth. Lower frequency filters are displayed deeper, and the higher frequencies are on top of the 3D natural migration volume. The red arrows indicate the fault image.
Figure 1.8: A migration cross-section demonstrating the detection of the fault as a function of pseudo-depth (filter number), where images of higher frequency filters are displayed at the top and the lower ones are at the bottom of the cross-section. The arrow indicates the top of the fault at $\omega' = 15$ Hz, and the dashed line indicates the trace of the fault.
Figure 1.9: Natural migration images for a fault model where the fault is 21 m deep.
Figure 1.10: Natural migration images for a fault model where the fault is 30 m deep.
1.5 Application to US-Array Passive Data

We demonstrate the potential of passive natural migration by migrating surface waves in the US-Array data set. The US-Array transportable array was deployed over deployment campaigns from the west coast to the east coast of the United States, where Lin et al. (2008) processed the US-Array passive data for ambient noise tomography. We use their cross-correlated data, and we refer the reader to their publications for the prepossessing flow.

1767 stations from the survey were used for migration. Cross-correlations that have less than 30 days of recording is excluded before imaging, and the remaining cross-correlations are normalized by the number of days. Only correlations of the ZZ-traces were used. Prestack images were muted to remove the effect of transmitted waves and singularities near the source and receiver position. The muting area is an ellipse with source and receiver positions as focal points for the ellipse. A smooth Gaussian tapper is used for the muting. The size of the ellipse was determined empirically based on the longest period used in the migrated data.

The US-Array backscattered data were imaged by applying the natural migration to the correlated traces. The high-, mid- and low-frequency migration images are shown in Figures 1.11. There appears to be a good correspondence between many known geological boundaries and the high energy areas in the migration images. This suggests that the boundaries between different geological regions scatter low-frequency surface waves. This conjecture will be examined in more detail with synthetic tests and extensive analysis of the USArray data.

Migration results should be interpreted with great caution near boundaries of any deployment campaign. Limited array coverage could cause false boundaries. To overcome this problem, the scattered data should be inverted for density and velocity perturbations. This is not trivial to accomplish for passive data.

1.6 Conclusions

The migration equations are derived for imaging back-scattered waves using virtual Green’s functions computed by cross-correlating ambient noise. The benefits of this approach are that the
Figure 1.11: Natural migration of US-Array data using three frequency bands (0.01-0.05 Hz, 0.05-0.1 Hz, and 0.1-0.25 Hz).
The back-scattering migration provides complimentary high-wavenumber information to the low-wavenumber transmission tomographic image as is done in exploration seismology. One possibility in the future, is to invert for the perturbation in the least-squares sense, instead of using the migration equation (the adjoint). In this paper, we migrated the backscattered events into pseudo-depths that depends on the frequency ranges of the data. Conversion to absolute depth requires some prior knowledge of the subsurface velocities, and such conversion is the subject of an ongoing research. Another direction of interest is to analyze the coupling between incident Rayleigh-wave and Love scattered waves and vice-versa.

\section*{1.7 Appendix I: Effect of Elastic Heterogeneities on Migration Images}

From equation 1.1, the scattered wavefield due to density perturbations can be quantified using the following equation

\[ u_i (x_s, x_r) = \int \omega^2 \delta_{pk} \gamma_l (\omega) G_{lp} (x|x_s) G^0_{ki} (x|x_r) \times \Delta \rho (x) d^3 x. \]

(1.14)

The corresponding migration equation \cite{Liu2006} is

\[ \Delta \tilde{\rho} (x) = \int \int \int_{x_s \neq x} \omega^2 \delta_{pk} \gamma_l (\omega) G_{lk} (x|x_s) G^0_{ki} (x|x_r) u_i (x_s, x_r) dx_s dx_r d\omega, \]

(1.15)
where the horizontal bar above the integration kernel indicates the complex conjugate of the kernel, and the spatial integration is over the source and receiver planes that exclude the imaging point. This avoids integrating over singular points in the Green’s tensors. For conciseness, we omit the definition of the integration domain over sources and receivers throughout the manuscript.

Similarly, the scattered wavefield due to elastic tensor perturbations is

\[ u_i(x_s, x_r) = -\int \gamma_l(\omega)(x) \frac{\partial}{\partial x_q} G_{lp}(x|x_s) \frac{\partial}{\partial x_j} G_{0}^{\alpha}(x|x_r) \Delta c_{kjpq} d^3x, \quad (1.16) \]

so that the adjoint integral is

\[ \Delta \tilde{c}_{kjpq}(x) = -\int \int \int \gamma_l(\omega) \frac{\partial}{\partial x_q} G_{lp}(x|x_s) \frac{\partial}{\partial x_j} G_{0}^{\alpha}(x|x_r) u_i(x_s, x_r) dx_r dx_s d\omega, \quad (1.17) \]

where \( \Delta \tilde{c}_{kjpq} \) is the image corresponding to the perturbation of the tensor element indicated by the subscripts. Considering the sum of migration images for \( j = q \), the migration equation above can be approximated in the far field by

\[ \sum_j \Delta \tilde{c}_{kjpj}(x) \approx \int \int \int \frac{\omega^2}{v^2(x)} \gamma_l(\omega) G_{lp}(x|x_s) G_{0}^{\alpha}(x|x_r) u_i(x_s, x_r) dx_r dx_s d\omega, \quad (1.18) \]

where \( v(x) \) is the phase velocity for that mode of propagation (e.g. the phase velocity for monochromatic Rayleigh-waves when migrating z-component backscattering using z-component incident wavefield). Here, the spatial derivatives are approximated, under the far-field approximation, for a single mode of propagation using

\[ \sum_j \frac{\partial}{\partial x_j} G_{lp}(x|x_s) \frac{\partial}{\partial x_j} G_{0}^{\alpha}(x|x_r) \approx -\frac{\omega^2}{v^2(x)} G_{lp}(x|x_s) G_{0}^{\alpha}(x|x_r). \quad (1.19) \]

Therefore,

\[ v^2(x) \sum_j \Delta \tilde{c}_{kjpj}(x) = \int \int \int \omega^2 \gamma_l(\omega) G_{lp}(x|x_s) G_{0}^{\alpha}(x|x_r) u_i(x_s, x_r) dx_r dx_s d\omega. \quad (1.20) \]

The right hand side above is the same as the migration equation for density (equation 1.15) when
\( p = k \). This indicates that when we migrate the backscattered data using equation 1.4 we can image both density and velocity perturbations. The left hand-side of equation 1.20 indicates that migrations images of elastic tensor perturbations have an amplification effect that is inversely proportional to the phase velocity.

The sum of the images for density-perturbations and the elastic-tensor perturbations (where \( j = q \)) gives the natural migration image

\[
m(x) = \Delta \tilde{\rho}(x) + v^2(x) \sum_{j,p,k} \Delta \tilde{c}_{pjk}(x) = \int\int\int \omega^2 (1 + \delta_{pk}) \gamma_l(\omega) G_{lp}(x|x_s) G_{ki}^0(x|x_r) u_i(x_s, x_r) \, dx_s, dx_r, d\omega,
\]

(1.21)

In other words, the natural migration image is the sum of several images, each image is a geometric representation of the density or elastic modulus perturbations. As a result, values in the images might not be immediately useful, due to the entangled contributions from different physical variables and the fact that the adjoint integral is not the inverse of the forward scattering equation. Nevertheless, the geometric information in the image is representative of the heterogeneities in the subsurface.

In our derivation above, we deliberately ignored contributions to the images from the elastic-tensor perturbations for \( j \neq p \) to avoid spatial derivatives, which numerically requires a dense sampling of naturally recorded Green’s functions. This does not necessarily mean that the perturbations in those elastic-tensor components can not be imaged using equation 1.4. Further research is needed to understand the effects on those components in the migration image.

1.8 Appendix II: Backscattered Events in Ambient-noise Cross-correlations

In this section, we illustrate how backscattered events are detected in the empirical Green’s function.
Empirical Green’s functions for the \( C_{11} \) are computed by cross-correlating the recorded ambient-noise traces according to equation 1.5 as follows. For a given virtual source located on one of the stations, the recorded trace is referred to as the master trace. For each recording, the master trace is cross-correlated with the trace corresponding to a virtual receiver. Then, the cross-correlations for the given virtual source and receiver are stacked to form the empirical Green’s function for the virtual source-receiver pairs.

Figure 1.12 depicts an empirical Green’s function for the simple 2D case where a homogeneous model has a single scatterer. The prominent features in the gather are the transmitted waves highlighted by the red dotted lines intersecting at the location of the virtual source at the zero-lag time. The rest of the empirical function is dominated by spurious events which are the result of cross-correlating irrelevant events. Such spurious events are weakened with longer recording time. Nevertheless, some key backscattering events can be identified and are highlighted in the yellow dashed lines and labeled A, B, C in Figure 1.12.

The backscattered events are the result of cross-correlating backscattered waves with transmitted events in the ambient noise. The cross-correlation and stacking process of events in the ambient-noise records eliminates common ray paths (Schuster et al., 2004; Schuster, 2010); i.e.

\[
\sum_n e^{i\omega \tau_{ns}} e^{-i\omega \tau_{ns}} \approx e^{i\omega \tau_{sr}},
\]

for ambient-noise sources located at \( n \), where \( \tau_{ns} \) denotes the traveltime from the noise source \( n \) to the virtual source \( s \), and similarly \( \tau_{nr} \) is the traveltime from the noise source to the virtual receiver \( r \). In this simplified analysis of kinematics, we harmlessly ignore the amplitudes and dispersion and highlight the phase of the correlated arrival. Using this simple notion of canceling the phase of common raypaths by cross-correlation, we can analyse the backscattering events in the empirical Green’s functions. Figure 1.13 depicts the three scenarios for redatuming passive events into the A-, B-, and C-labeled backscattered events in the empirical Green’s function in Figure 1.12.

The A-labeled event in Figure 1.12 is characterized by acausal scattering excited by an acausal incident wave. As demonstrated by the corresponding plot labeled A in Figure 1.13, this event is the result of cross-correlating backscattered events from a scatterer at the virtual source position.
with the direct arrival at the virtual receiver position. The common ray path is eliminated and
the results indicate an event that travels from the virtual source to the scatterer to the virtual
receiver, i.e.
\[
\sum_n e^{i\omega \tau_{nr}} e^{-i\omega \tau_n X_s} \approx e^{-i\omega \tau_s X_r},
\]
(1.23)
where \( X \) is the position of the scatterer and the \( \tau \) subscripts denote the points along the raypath.
However, this event has a negative phase (dashed lines Figure I.13 indicate negative phase),
and therefore it appears at a negative time-lag in the empirical Green’s function in Figure I.12.
Similarly, the C-labeled event is the result of cross-correlating the direct arrival associated with the
virtual source position and the backscattered events at the virtual receiver position. Therefore, the
phase delay associated with the common ray path is eliminated, giving rise to a causal scattering
event due to a causal incident wave that appears in the positive time lag in the empirical Green’s
function (\( e^{i\omega \tau_s X_r} \)).

The remaining B-labeled event is related to cases where the scatterer is between the virtual
source and receiver positions. In such cases, backscattered events are cross-correlated with direct
events, or vice versa, giving rise to a mixed phase (\( e^{i\omega \tau X_r - i\omega \tau_s X} \)), where the event could appear at
either positive or negative time lags of the empirical Green’s function. In all cases, however, the
negative phase associated with events traveling from the virtual source to the scatterer while the
positive phase is associated with events traveling from the scatterer to the virtual receiver shown in
Figure I.13B. Therefore, the B-labeled backscattered event is causal backscattering due to acausal
incident waves. If cross-correlation is time symmetric as defined in equation I.5 a mirror of the
B-labeled event will be in the empirical Green’s function, which is acausal backscattering due to
causal incident waves.

The B-labeled backscattering event can be considered non-physical and contradictory to equation I.6. Such non-physical scattering, however, is redundant information and often overlaps
with early arrivals like body waves and strong cross-correlation artifacts. Therefore, we mute such
events between the causal and the acausal transmitted events in the ambient noise cross-correlation
before migration.
Figure 1.12: A shot gather of an empirical Green’s function computed by cross-correlating ambient noise. The red triangle denotes the lateral position of the virtual source (i.e. the master-trace position for the cross-correlation) and the green dashed line denotes the position of a scatterer.
Figure 1.13: Ambient-noise cross-correlation scenarios that give rise to the backscattering events in Figure [1.12]. The start symbol $\star$ denotes convolution between the conjugated phases, shown as dashed lines from the noise source to the virtual source, and phases from noise source to the virtual receiver shown as a solid line.
Chapter 2

Natural Inversion of Scattered Surface Waves with Application to Rayleigh Waves in Long-Beach Passive Data

2.1 Abstract

We use surface-wave scattering in ambient-noise cross-correlations to image near-surface major faults under the populated area of Long Beach, Los Angeles County, California. Images are computed using empirical Green’s functions from ambient-noise cross-correlation, and therefore we eliminate the need for prior velocity models and the costly modeling of surface-waves propagation. The scattered waves are inverted in the least-squares sense using a pseudo-inverse of the Hessian matrix of the ambient-noise scattering. Our results show a number of faults in the area including faults in the Newport-Inglewood fault zone (NIFZ).

2.2 Introduction

Ambient-noise seismology has proved successful in generating virtual Green’s functions from recorded noise by correlating trace pairs to get the virtual Green’s functions (Claerbout, 1968; Rickett and Claerbout, 1999; Weaver and Lobkis, 2004; Snieder, 2004; Campillo and Paul, 2003). The assumption is that there is a uniform glow of incidence energy from all directions (Weaver...
and Lobkis, 2001). In most cases, the records are dominated by the surface wave arrivals that are often inverted for the S-velocity in the subsurface (Shapiro et al., 2005; Sabra et al., 2005; Gerstoft et al., 2006; Bensen et al., 2007; Lin et al., 2008, 2013). However, the S-velocity tomogram is a smooth representation of the subsurface geology and does not easily reveal high-wavenumber details of the subsurface, such as faults.

To detect faults in the subsurface, AlTheyab et al. (2015a,b) proposed natural migration of surface waves to estimate the migration image at the surface, where both the scattered surface waves and the Green’s functions are extracted from correlation of ambient noise. Unlike standard migration, no velocity model is required and there is no weak scattering (i.e., Born approximation) assumption. However, the trial image points are restricted to be along the free surface at the receiver locations, which is appropriate because buried faults cast seismic surface-wave shadows with strong amplitudes in the migration image. This is only true if the fault is shallower than about a third of the dominant wavelength of the surface wave.

AlTheyab et al. (2015a) showed the effectiveness of natural migration of surface waves with 3D synthetic data. Their results demonstrated how buried faults could be detected in the migration image at the surface if the fault was no deeper than about a third of the wavelength. This paper shows the results of applying natural surface-wave migration to ambient-noise data recorded by over 5000 z-component seismometers in Long Beach, California. Results show that surface-wave migration of scattered surface waves can detect the existence of some known faults, and also indicates the existence of unknown buried faults. However, some known faults were not indicated in the migration image, and the cause is most likely due to migration artifacts due to the recording geometry and the non-uniform illumination of the seismic noise. To mitigate the artifacts from the recording geometry, the least-squares migration image is obtained by applying the Hessian inverse to the migration image (Nemeth et al., 1999).

In the following sections, we begin by describing the method used, and the next section describes the acquisition and the processing steps for the Long-Beach ambient-noise data. Then, we analyze the acquisition footprint on the migration images by analyzing the natural migration response to test models. The next section tests the effectiveness of the Hessian inverse for improving
the migration images. Finally, we show results from inverting the Long Beach data and validate them with the known geology of the area. The final section presents the conclusions.

2.3 Method

The scattered surface-wave Greens function $G_1(r_s|\mathbf{r}, \omega)$ for a single mode of surface wave can be written as

$$G_1(r_s|\mathbf{r}, \omega) = -\int \int m(\mathbf{r}) \beta_\omega(\omega) G_0(\mathbf{r}, \omega) G(\mathbf{r}, r_s, \omega) \, d^2\mathbf{r},$$

where the integration is over the recording surface, $\omega$ is the angular frequency, $\mathbf{r}$ indicates a position on the surface, the impulsive source is located at the position $r_s$ on the surface, and $\mathbf{r}$ is the receiver position. $G(\mathbf{r}|r_s, \omega)$ is the incident Green’s function for the reference heterogeneous model, and the scattered Green’s function $G_0(\mathbf{r}, r_s, \omega)$ is defined strictly for a laterally smooth model that does not produce backscattering. Here, superscripts indicating the mode of surface wave are suppressed. $\beta_\omega(\omega)$ is a bandpass filter that has a narrow Gaussian distribution in the frequency domain with it’s peak at $\omega'$, and $m(\mathbf{r})$ is the scattering potential that couples the incident and the scattered modes, for frequencies near $\omega'$. This scattering potential is a 2D representation of the 3D heterogeneities, where heterogeneities above a third of the wavelength below the surface are projected to the surface. Equation 2.1 can be derived from the membrane-wave analog of surface waves (Tanimoto 1990; Tromp and Dahlen 1993) or Snieder’s (1986) formalism for surface-wave holography, as discussed in Appendices A and B, respectively.

Equation 2.1 can be evaluated using Green’s functions estimated from the ambient noise cross-correlations. The surface-wave Green’s function between points $\mathbf{x}_A$ and $\mathbf{x}_B$ for a given mode is kinematically related to the causal part of the time-symmetric cross-correlation $C(\mathbf{x}_A|\mathbf{x}_B, \omega)$ (Lobkis and Weaver 2001; Snieder 2004) in the far field by the approximation

$$G(\mathbf{x}_A|\mathbf{x}_B, \omega) \propto -i C(\mathbf{x}_A|\mathbf{x}_B, \omega),$$

(2.2)
and, therefore, equation 2.1 is approximated as

\[ C_1(r|s,r,\omega) = \int \int m(r) \beta_{\omega'}(\omega) C_0(r|s,\omega) C(r|s,r,\omega) \, d^2r, \quad (2.3) \]

where \( C(r|s,\omega) \) is the cross-correlation function, \( C_0(r|s,\omega) \) is the cross-correlation function containing the transmitted waves (i.e. scattered events are muted), and \( C_1(r|s,\omega) \) is the cross-correlation function containing only the scattered events (i.e. transmitted waves are muted). The corresponding migration equation becomes

\[ \tilde{m}(r) = \sum_{s} \sum_{r_r} \sum_{s} C_1(r|s,\omega) \beta_{\omega'}^*(\omega) C_0^*(r|s,\omega) C^*(r|s,\omega) \cdot (2.4) \]

The system of equations using equation 2.3 can be written as \( d = Lm \), where \( d \) is the vector containing the scattered events in ambient noise cross-correlations, \( L \) is the forward-modeling operator, and \( m \) is the perturbation model (i.e. the scattering potential in equation 2.1). The adjoint operation is \( L^\dagger d \), where the migration operator \( L^\dagger \) is the complex-conjugate transpose to the forward modeling operator. To invert for the model, we employ the least-squares solution \( m = H^{-1} L^\dagger d \), where the Hessian matrix is defined as \( H = L^\dagger L \).

Each column of the Hessian matrix contains the point-scatterer response in the migration image, a.k.a the point-spread-function (PSF), which can be mathematically defined by inserting \( m(r,\omega') = \delta(r - r_{ref}) \) into equation 2.3 and then inserting the result into equation 2.4 to get

\[ H(r, r_{ref}) = \sum_{\omega} \sum_{r_r} \sum_{s} |\beta_{\omega'}(\omega)|^2 C_0(r|r_{ref},\omega) C(r_{ref}|s,\omega) C_0^*(r|r,\omega) C^*(r|r_{ref},\omega). \quad (2.5) \]

Here, \( r_{ref} \) is the point-scatterer position and \( H(r, r_{ref}) \) is its associated PSF.

The Hessian matrix might be rank deficient and affected by the signal-to-noise ratio (SNR) in the ambient-noise cross-correlation. Therefore, we approximate the Hessian inverse by the Singular

\footnote{We are mainly concerned with the phase information in the cross-correlations. Therefore, the amplitudes are always normalized to unity before evaluating any integral.}
Value Decomposition (SVD) procedure (Heath, 1996):

\[ H \overset{\text{def}}{=} U S V^\dagger, \quad (2.6) \]

\[ H^{-1} \approx V R U^\dagger, \quad (2.7) \]

where \( S \) is a diagonal matrix of the singular values, and \( U \) and \( V \) are unitary matrices containing, respectively, the left-singular and and right-singular vectors. The pseudo-inverse is estimated by eliminating the singular values below a certain tolerance value \( \epsilon \), where the diagonals of the matrix \( R \) are defined as

\[ R_{ii} = \begin{cases} \frac{1}{s_{ii}} & s_{ii} > \epsilon \\ 0 & \text{otherwise.} \end{cases} \quad (2.8) \]

The optimal singular value is estimated by computing the images for a range of tolerance values (20\%, 40\%, 60\%, and 80\% of the largest singular value). The most uncertain image features are those that are most sensitive to the small values of singular values.

### 2.4 Application to Long-Beach Data

The Long-Beach passive data were recorded by 5300 stations, where \( z \)-component ambient noise was correlated and stacked to give the empirical Green’s functions. Figure 2.1 depicts the station locations next to the Google map view of the area. The receiver density is about 775 sensors per square kilometer. There are large gaps in the data especially in the north-eastern part of the survey, where the Long-Beach airport is located.

The processing of ambient noise is described by Lin et al. (2013). The cross-correlated records show visible Rayleigh-wave events from 0.5-2.5 Hz. Detecting the scattered surface waves is challenging due to the noise level. Figure 2.2 shows several snapshots of one virtual gather. The transmitted surface waves are reconstructed well in all azimuths, and some spurious events are present in the data.

To prepare the data for migration, the first step is to spectrally normalize the traces by setting the amplitude of every frequency to one, where the bandpass filter \( \beta(\omega) \) in equations 2.3 and 2.4.
shapes the spectrum of the data during the inversion.

The next step is to separate the recorded surface waves into transmitted arrivals $C_0$ and scattered events $C_1$. We apply muting for this purpose by approximating the average Rayleigh wave velocity by $v = 400 \text{ m/s}$ and the period $T = 2 \text{ s}$ in the following muting equations

\[
C_0 (\mathbf{r}_r | \mathbf{r}, t) = \begin{cases} 
    C (\mathbf{r}_r | \mathbf{r}, t) & t < \frac{|\mathbf{r} - \mathbf{r}_r|}{v} + T \\
    0 & \text{otherwise}
\end{cases}, \quad (2.9)
\]

\[
C_1 (\mathbf{r}_r | \mathbf{r}, t) = \begin{cases} 
    C (\mathbf{r}_r | \mathbf{r}, t) & t > \frac{|\mathbf{r} - \mathbf{r}_r|}{v} + T \\
    0 & \text{otherwise}
\end{cases}, \quad (2.10)
\]

where $C_0$ ideally contains the transmitted events, and $C_1$ contains the scattered events in the original cross-correlation function $C$. Note that events between 0 and $T$ are muted to satisfy the far-field approximations used in the derivations. A smooth taper is recommended when applying the mute. Perfect separation might be difficult to attain in practice, so we could shift the muting function with a larger $T$ (or with slower $v$) such that we make sure that $C_1$ does not contain any strong transmitted waves. Figure 2.3 shows an example of data separation by muting.

Finally, the band-pass filters $\beta_{\omega'} (\omega)$ are selected for the modeling and imaging operators during the inversion. We use the modified Morlet wavelet defined in the time domain as

\[
\beta_{\omega'} (t) = \cos (\omega' t) e^{\frac{1}{12} (\omega' t)^2}. \quad (2.11)
\]

The filters span the spectrum from 0.2 Hz to 4.0 Hz as shown in Figure 2.4.

Now, we have all the elements needed for the forward modeling and adjoint operators, which we study in the following sections.

2.4.1 PSFs of Long-Beach Data

To understand the artifacts in the migration image $L^\dagger \mathbf{d}_{\text{obs}}$, we migrate synthetic data generated using a test model $\mathbf{m}_{\text{test}}$ in $\mathbf{d}_{\text{synth}} = L \mathbf{m}_{\text{test}}$. In Figure 2.5, the test model $\mathbf{m}_{\text{test}}$, with several
point-like scatterers, is used to generate synthetic data, which is then migrated using \( \mathbf{m}_{\text{response}} = \mathbf{L}^\dagger \mathbf{d}_{\text{synth}} = \mathbf{H} \mathbf{m}_{\text{test}} \). The response is shown in Figures 2.5b, 2.5c, and 2.5d, for the frequencies 0.67 Hz, 1.1 Hz, and 2.0 Hz, respectively. The smeared features in the migration images are the PSFs for the scatterers in \( \mathbf{m}_{\text{test}} \).

Ideally, the PSFs should be a band-limited version of point scatterers in the test model \( \mathbf{m}_{\text{test}} \), where the PSF of each point scatterer is centered at the correct position and the smearing is invariant with azimuthal angle. This is roughly observed for the PSFs in the middle of the survey. However, toward the edges of the survey, the PSFs have elongated response along the circles centered at the middle of the survey. In addition, the responses tend to have higher amplitudes towards the edges.

The elongated responses of PSFs and the increase in amplitudes are migration artifacts mostly caused by the survey geometry\(^2\). At the center of the survey, virtual sources are uniformly distributed for all azimuths and so provide uniform illumination at the center. Toward the edges of the survey, the majority of the receivers are predominantly in a narrow azimuth range relative to the image point. This uneven distribution of virtual sources and receiver about an image point is manifested by the elongated PSFs.

The PSF patterns have important implications on the geological features in the migration images \( \mathbf{m}_{\text{mig}} = \mathbf{L}^\dagger \mathbf{d}_{\text{obs}} \). Lineaments (e.g. faults) tangential to the circular artifacts will be highlighted in the migration image, while lineaments along a radial direction from the center of the survey will suffer from smoothing. This will minimize the detectability of faults within the NIFZ which are oriented radially with respect to the center of the survey.

The effect on images of faults can be demonstrated with the checkerboard test shown in Figure 2.6. The checkerboard faulting model in Figure 2.6a is oriented roughly along the direction of NIFZ. Synthetic data are generated and migrated as described above. Figures 2.6b, 2.6c, and 2.6d show the migration images for different frequencies. The low-frequency images in Figures 2.6b and 2.6c are adversely affected where many faults are smeared as indicated by the arrows.

In addition to smearing artifacts, the amplitude scaling of PSFs is also affected in part by the

\(^2\) Another source of artifacts is the non-uniform illumination of seismic sources in the ambient noise. Non-uniformity of noise sources leads to significant errors in the virtual Green’s functions.
survey geometry. However, the local velocity at the position of the PSF also affect its amplitude. Figure 2.7a shows the diagonal elements of the 0.67 Hz Hessian matrix. The values of the diagonal elements tend to increase from the center toward the edges of the survey, due to the survey geometry. Nevertheless, structural information is evident in the images where, for example, there is a NW-SE trending high-value zone in the values of the Hessian diagonal elements for 0.67 Hz, possibly due to local velocity variations. Such trends change gradually with increasing frequencies. From Figure 2.7d, we can see how the large Hessian diagonal elements affect the amplitudes of the migration images. The large amplitudes toward the edges of the survey, correspond to large values in the Hessian diagonal elements. In addition, image features which coincide with high-value zones have stronger amplitudes.

Both amplitude and smearing artifacts in the migration images need to be corrected for. In the following section, we correct the images by applying the pseudo inverse of the Hessian matrix to $m_{\text{response}}$ to get the corrected image $m_{\text{inv}}$, i.e. compute $m_{\text{inv}} = H^{-1}L^\dagger Lm_{\text{test}}$.

## 2.5 Natural Inversion

Computing the Hessian matrix with the virtual Green’s functions from the Long-Beach data is computationally expensive. Therefore, we approximate the Hessian by using only 6.25% of the records from the available virtual traces (i.e. using 25% of the sources and 25% of the receivers). The computed Hessian is poorly conditioned, and therefore, we calculate its pseudo inverse using 4000 singular values out of the 5300 singular values. Then this pseudo inverse is applied to the migration images from the checkerboard test in Figure 2.6. Note that the migration images are computed using all of the virtual shot-receiver pairs in the cross-correlated dataset, so our inversion test gives

$$m_{\text{inv}} = H_{\text{partial}}^{-1} H_{\text{full}} m_{\text{test}},$$

(2.12)

where $H_{\text{partial}}^{-1}$ is the pseudo inverse of the Hessian constructed from the 6.25% of the data, where $H_{\text{full}}$ used the full Hessian. This approach of using the partial Hessian is empirically validated by results shown in Figure 2.8, where the checkerboard pattern is geometrically reconstructed by
applying $H^{-1}_{\text{partial}}$ to the migration images in Figure 2.6.

This approach is applied to the cross-correlation scattered data $m_{\text{inv}} = H^{-1}L^\dagger d_{\text{obs}}$ to get the results shown in Figure 2.9. The migration image is shown in a) and the inversion images are shown in 2.9b, 2.9c and 2.9d where different numbers of singular values are used to compute the pseudo inverse. The inversion enhances the spatial resolution and balances the amplitudes, where the resolution increases with an increasing number of singular values. However, we observe that with a higher number of singular values random noise is introduced into the image. Such noise can be removed by post-inversion image processing.

In Figure 2.9 faulting features (indicated by the arrows) start to emerge with an increase in the number of singular values. Such features are smeared and undetectable in the migration image. This is consistent with our observations in the checkerboard test.

Figures 2.10a, 2.10b, 2.10c, and 2.10d depict the migration and inversion results for the peak-frequencies 1.1 Hz and 2.0 Hz. We notice that the inversion images have low-wavenumber components that we filter out by subtracting the local mean of the image values within 250 meters. Subtracting the local mean highlights the faulting discontinuities, as shown in Figures 2.11a and 2.11b. The enhanced images are used in the remaining sections for validation and interpretation.

### 2.6 Cross-validation & Interpretation

In this section, we validate our method by comparing the inversion results to phase-velocity tomograms obtained with the same Long-Beach passive data, then we compare the results to known geological features of the Long-Beach area.

Figure 2.12 compares the 2 Hz inversion results to the ambient-noise phase-velocity tomogram of Lin et al. (2013) in Figure 2.12a. We can see in Figure 2.12b that there is the correspondence between rapid changes of the velocity and the inversion results, as highlighted by the red arrows. Figure 2.13a shows a snapshot of an earthquake wavefield that propagates through the Long-Beach area Clayton (2012). As the wavefield propagates in the complex subsurface, discontinuities develop close to fault boundaries. Several of the lineaments in the 2.0-Hz inversion image coincide with several discontinuities in the observed wavefield. The discontinuities also correlate with
features in inversion images from other frequencies.

Inversion images show some correlation with prominent topographic features in the area. Figure 2.14 illustrates how the lakes are oriented along major near-surface faults. The Bouton lake, for example, is along a fault line trending NE-SW that extend from the lake across the northern edges of the Long-Beach airport and the San Diego freeway to finally intersect the northern part of the Newport-Inglewood fault, and possibly extend further toward the southwest. In the southeastern part of the map, the Colorado lagoon is also aligned along one NW-SE trending fault, which is possibly a part of the NIFZ.

Signal Hill is located in the middle of the survey, and it reaches more than 100 m height above the average ground level of the surrounding area. The hill is visible in the migration images at around 0.65-1.3 Hz peak frequency. Figure 2.15a) shows a roughened topographic map, where typically a crossing from positive to negative (i.e. black to white colors) marks the boundaries of topographic features similar to how migration highlights fault boundaries. The highlighted feature in Figure 2.15b roughly outlines the hill. Further investigation is needed to determine whether the backscattering is related to the hill itself (i.e. perturbation of the free-surface) or the relevant faults in the subsurface (i.e. lateral perturbation in the elastic parameters).

Several other lineaments in the migration images do not correlate to the surface expression of known lineaments or faults. However, the lineaments in the inversion images are very likely to be indicative of near-surface faults, which should be mapped and analyzed to understand their risk. See Figures 2.16 and 2.17 for images that span all the frequencies investigated by this study.

2.7 Conclusions

We demonstrate for the first time that faults can be detected by natural migration of scattered surface waves recorded from ambient seismic noise. No velocity model is required to give the migration image because the data are migrated with the virtual Green’s functions computed from ambient noise. Migrating ambient noise recorded by 5300 z-component stations in Long Beach data suggests the presence of hidden faults that are a continuation of visible fault scarps. The migration image reveals high-wavenumber details of the subsurface geology that are a complement to the
smoothly varying features estimated from Rayleigh-wave velocity tomograms. The implication of this work is that more accurate hazard maps can be created by ambient-noise migration images of the subsurface.

The problem with the migration and inversion images is that they do not reveal all of the portions of the NIFZ that are mapped out by geologists. This suggests that migration images give false negatives in the presence of known faults. One cause is that seismic scattering is sensitive to rapid lateral changes in elastic parameters, and some faults are not associated with large velocity contrasts. This is a limitation that requires further study. In this paper, we applied a disjoint inversion where we ignored secondary contributions from, for example, coupling between different modes. Such coupling can be addressed by a joint inversion or a by applying a post-inversion correction as discussed in Appendix C.

Further research is needed to propose modifications to the inversion for attenuating the noise in the images computed from ambient noise. Image processing techniques could provide a remedy for the noise in the images.

### 2.8 Acknowledgments

This publication is based upon work supported by the KAUST Office of Competitive Research Funds (OCRF) under Award No.OCR-2014-CRG3-62140387/ORS#2300. For computer time, this research used the resources of the Supercomputing Laboratory at King Abdullah University of Science & Technology (KAUST) in Thuwal, Saudi Arabia. We thank Nodal Seismic, and SignalHill for provided access the field data. AlTheyab thanks Saudi Aramco for sponsoring his studies at KAUST.
2.9 Appendix B: Backscattering using the membrane-wave analog of surface wave

The Green’s function $G(r, \omega|r_s)$ for surface-wave potentials satisfy the 2D Helmholtz equation

$$\left[ \nabla^2 + \omega^2 \mu(r, \omega) \right] G(r|r_s, \omega) = \delta(r - r_s), \quad (2.13)$$

where $\omega$ is the angular frequency, $r$ indicates the position on the surface, and the impulsive source is located at the position $r_s$ on the surface, and $\mu(r, \omega)$ is the squared slowness (i.e. the reciprocal to squared velocity).

By perturbing the squared slowness $\mu = \mu_0 + \delta \mu$, where $\delta \mu$ is the squared-slowness perturbation, we can compute the scattered surface-wave Greens function $G_1(r|r_s, \omega)$ using

$$G_1(r|r_s, \omega) = -\int \int \delta \mu(r, \omega) \omega^2 G_0(r,r, \omega) G(r,|r_s, \omega) d^2r. \quad (2.14)$$

Unlike the Green’s function for the reference model $G(r|r_s, \omega)$, the Green’s function for the perturbed model $G_0(r,|r, \omega)$ is defined strictly for a laterally smooth model that does not produce backscattering.

Solving for the perturbation model $\delta \mu(r, \omega)$ by inverting equation 2.14 can be numerically challenging because of (1) the size of the model (i.e. the large number of the unknowns), (2) the sensitivity to noise and artifacts in the cross-correlations, and (3) the bias to high frequencies due to the $\omega^2$ factors in the integration kernels. To simplify the problem, we assume that the model perturbation $\delta \mu(r, \omega)$ is slowly varying with frequency. Therefore, we can explain the data in the vicinity of the central frequency $\omega'$ using a single frequency-independent model $m(r)$ by redefining the perturbation model as

$$\delta \mu(r, \omega) \overset{\text{def}}{=} \frac{1}{\omega^2} \beta_{\omega'}(\omega) m(r). \quad (2.15)$$

where the smooth function $\beta_{\omega'}(\omega)$ is a narrow Gaussian distribution in the frequency domain. By
inserting equation \[2.15\] into equation \[2.14\] we finally reach to equation \[2.1\]

\[ G_1 (r_r | r_s, \omega) = - \int \int \beta_{\omega'} (\omega) m (r) G_0 (r_r | r, \omega) G (r | r_s, \omega) \, d^2 r. \] \[ (2.16) \]

### 2.10 Appendix B: Backscattering using the Sneider’s formalism for surface-wave holography

Sneider (1986) introduced a formalism for multi-mode surface-scattering which is described using

\[ u^{\sigma\nu} (x_r | x_s, \omega) = \int_V p^\sigma (x_r) G_0^\sigma (\tilde{x}_r | \tilde{x}) V^{\sigma\nu} (x, \psi) G^{\nu} (\tilde{x} | \tilde{x}_s) [p^{\nu} (x_s) \cdot F] \, d^3 x, \] \[ (2.17) \]

where

- \( u \) is scattered-wavefield displacement vector,
- the superscripts \( \sigma \) and \( \nu \) denote the mode of surface wave which can be fundamental-Rayleigh, 1\textsuperscript{st}-Rayleigh mode, 2\textsuperscript{nd}-Rayleigh mode, ..., fundamental-Love-waves mode, 1\textsuperscript{st} Love mode, 2\textsuperscript{nd}-Love mode, ..., etc.
- \( x, x_r, x_s \) are the positions of scatter-point, receiver, and source position, respectively.
- The symbol (\( \tilde{\cdot} \)) above a position vector indicates it projected position onto the surface.
- \( G_0^\sigma (\tilde{x}_r | \tilde{x}) \) and \( G^{\nu} (\tilde{x} | \tilde{x}_s) \) are the 2D Green’s functions for wavefield potentials.
- \( F \) is the 3-component source function.
- \( p^\sigma, p^{\nu} \) are the polarization vectors.
- \( V^{\sigma\nu} (x, \psi) \) is the interaction matrix that depends on frequency, position of the scatterer, and scattering angle \( \psi \) on the planner surface.

If we are concerned with the z-component in an isotropic elastic medium due to a z-component impulsive source, we can use the polarization vector \( p^z (x) = \begin{bmatrix} 0 & 0 & i l^{\nu} (x) \end{bmatrix}^T \) at the source,
receiver, and scatterer positions to get the simplified form

\[ u_{z}^{\sigma,\nu}(x_{r}|x_{s},\omega) = -\int_{V} l^{\sigma}(x_{r}) l^{\nu}(x_{s}) V^{\sigma\nu}(x,\psi) G^{\nu}(\tilde{x}_{r}|\tilde{x}) G^{\nu}(\tilde{x}|\tilde{x}_{s}) d^{3}x, \]  

(2.18)

where

\[ V^{\sigma\nu}(x,\psi) = \Delta \rho \omega^{2} [-l^{\sigma}(x) l^{\nu}(x)] \]

\[ + \Delta \lambda [(\partial_{z} l^{\sigma}(x))(\partial_{z} l^{\nu}(x))] \]

\[ + \Delta \mu [2(\partial_{z} l^{\sigma}(x))(\partial_{z} l^{\nu}(x))] \]

\[ + \Delta \mu [k_{\sigma} k_{\nu} l^{\sigma}(x) l^{\nu}(x) \cos(\psi)], \]  

(2.19)

\( l^{\nu}(x) \) and \( l^{\sigma}(x) \) are, respectively, the \( \nu \)–mode and \( \sigma \)–mode eigenfunctions, \( k_{\sigma} \) and \( k_{\nu} \) are the wavenumbers for the scattered mode and the incident wave-mode. \( \Delta \rho, \Delta \lambda, \) and \( \Delta \mu \) are perturbations of the density, Lame first parameter, and Lame second parameter, respectively.

The third term of equation (2.19) is the angle-dependent term of the interaction matrix. If we assume the sheer modulus perturbation is sufficiently small \( \Delta \mu \to 0 \), then the angle dependent term becomes negligible, hence \( V^{\sigma\nu}(x,\psi) \to V^{\sigma\nu}(x) \). Finally, the volume integral can be reduced to a surface integral, and the bandpass filter \( \beta_{\omega}(\omega) \) is introduced in a similar fashion as in Appendix A to get

\[ u_{z}^{\sigma,\nu}(x_{s},x_{r}) = -\int_{S} \beta_{\omega}(\omega) m^{\sigma\nu}(\tilde{x}) G^{\sigma}_{0}(\tilde{x}_{r}|\tilde{x}) G^{\nu}(\tilde{x}|\tilde{x}_{s}) d^{2}\tilde{x}, \]  

(2.20)

where \( m^{\sigma\nu}(\tilde{x}) \approx \{ \int_{0}^{\infty} V^{\sigma\nu}(x) dz \} \). The scalar factor \( l^{\sigma}(x_{r}) l^{\nu}(x_{s}) \) is ignored here due the amplitude normalization in our implementation of equation (2.20).
2.11 Appendix C: From disjoint to joint inversion

In our formulation, we solved the system

\[ \mathbf{d} = \mathbf{Lm} + \mathbf{Jn} \quad (2.21) \]

where we assumed \( \mathbf{Lm} \) is the leading term and \( \mathbf{Jn} \) is the sum of the remaining secondary terms that we ignored, that are related to, for example, anisotropic scattering, mode-polarization, and coupling between different modes. The solution of the joint inverse problem is

\[
\begin{bmatrix}
\mathbf{m}_{\text{inv}} \\
\mathbf{n}_{\text{inv}}
\end{bmatrix} = \mathbf{H}^{-1} \begin{bmatrix}
\mathbf{L}^\dagger \mathbf{d} \\
\mathbf{J}^\dagger \mathbf{d}
\end{bmatrix},
\quad (2.22)
\]

where the Hessian is as follows

\[
\mathbf{H} = \begin{pmatrix}
\mathbf{L}^\dagger \mathbf{L} & \mathbf{L}^\dagger \mathbf{J} \\
\mathbf{J}^\dagger \mathbf{L} & \mathbf{J}^\dagger \mathbf{J}
\end{pmatrix}
= \begin{pmatrix}
\mathbf{H}_{LL} & \mathbf{H}_{LJ} \\
\mathbf{H}_{JL} & \mathbf{H}_{JJ}
\end{pmatrix}
\quad (2.23)
\]

The joint inverse problem can be solve disjointly first and then followed by a correction. The importance of the correction is dependent on the off-diagonal blocks \( \mathbf{H}_{LJ} \) and \( \mathbf{H}_{JL} \) of the Hessian matrix, and how large they are compared to the diagonal ones. In this paper, we solved the disjoint problem

\[
\mathbf{m}_{\text{lsm}} = \mathbf{H}_{LL}^{-1} \mathbf{L}^\dagger \mathbf{d}.
\quad (2.24)
\]

To compute the correction terms, we use the inverse of the Hessian

\[
\mathbf{H}^{-1} = \begin{bmatrix}
\mathbf{H}_{LL}^{-1} + \mathbf{X} & -\mathbf{H}_{LL}^{-1} \mathbf{H}_{LJ} \mathbf{C}^{-1} \\
-\mathbf{C}^{-1} \mathbf{H}_{LJ} \mathbf{H}_{JJ}^{-1} & \mathbf{C}^{-1}
\end{bmatrix},
\quad (2.25)
\]
where

\[
C = H_{JJ} - H_{JL}H_{LL}^{-1}H_{JL}, \quad (2.26)
\]

\[
X = H_{LL}^{-1}H_{LJ}C^{-1}H_{JL}H_{LL}^{-1}, \quad (2.27)
\]

given that \( C \) is invertible. Therefore, the correction terms to the final images are described as

\[
m_{inv} = m_{ls} + \underbrace{Xm_{ls} - H_{LL}^{-1}H_{LJ}C^{-1}J^d}_{\text{Correction terms}} \quad (2.28)
\]

The importance of the correction terms are a subject of further research depending on the case at hand. Yet, we believe that images estimated by disjoint inversion is correct to the first order. We should note that some correction terms can not be computed using natural operators, because they required the wavefield at depth to compute vertical derivatives, for example.
Figure 2.1: The left plot is the Long-Beach ambient-noise survey where each dot denotes a receiver position, and the right plot is the corresponding street-view map. Here, the red line indicate the position of the surface expression of the Newport-Inglewood fault.
Figure 2.2: Snapshots of ambient noise cross-correlations. The star indicates the virtual source position.
Figure 2.3: Separation via muting of cross correlations $C$ (top) into transmitted $C_0$ (middle) and scattered $C_1$ data (bottom). The traces are bandpass filtered for display.
Figure 2.4: Bandpass filters in the time domain (top) and the frequency domain (bottom).
Figure 2.5: a) The point-scatterers used to test the operators and the point-scatterer responses for b) 0.67 Hz, c) 1.1 Hz, and d) 2.0 Hz peak frequencies.
Figure 2.6: a) Checkerboard model used to test the operators and the migration responses for fault-like lineaments at b) 0.67 Hz, c) 1.1 Hz, and d) 2.0 Hz.
Figure 2.7: a) The diagonal elements of the Hessian matrix at 0.67 Hz, and b) the corresponding migration image in Figure 2.6b overlaid with the same diagonal elements. Hot red colors indicate large values and dark blue colors indicate small values of the diagonal elements.
Figure 2.8: a) Checkerboard model used to test the operators and the inversion results for b) 0.67 Hz, c) 1.1 Hz, and d) 2.0 Hz.
Figure 2.9: a) The 0.67 Hz natural migration images $L^\dagger d$ and the inversion images $H^{-1}L^\dagger d$ using b) 1000, c) 2000 d) 4000 singular values. The yellow arrows indicate the Newport-Inglewood fault and the red arrow highlights the southern edge of Signal Hill.
Figure 2.10: The natural migration images for a) 1.1 Hz and b) 2.0 Hz peak frequencies, c) and d) are their respective inversion images with 4000 singular values.
Figure 2.11: a) and b) are enhanced versions of, respectively, Figure 2.10c and 2.10d by subtracting the local mean to highlight discontinuities.

Figure 2.12: a) 2 Hz phase velocity map ([Lin et al., 2013]) from ambient-noise tomography, and b) is an overlay of the corresponding inversion image in Figure 2.11b on top of the velocity map.
Figure 2.13: a) A snapshot of the observed earthquakes waves, and b) is an overlay of the corresponding inversion image in Figure 2.11b on top of the wavefield snapshot.

Figure 2.14: a) The Google Earth map of the Long-Beach area with the arrows highlighting the Bouton lake north and the Colorado lagoon south. b) is the same map with an overlay of the 2.0 Hz inversion image of Figure 2.11b. The position and orientation of the lakes line up with preexisting faults that are observed in the inversion image.
Figure 2.15: a) The Laplacian of the digital elevation model ($\nabla^2 e(x,y)$), where $e(x,y)$ is the elevation model. b) is the 0.67 Hz inversion image. Signal Hill is highlighted by the arrows in both images.
Figure 2.16: Enhanced inversion images for 3000 singular values. Each plot is for a single peak frequency.
Figure 2.17: Enhanced inversion images for 3000 singular values. Each plot is for a single peak frequency.
Part II

Full-Waveform Inversion of Seismic Reflections
Introduction to Part II

The goal of exploration geophysicists is to infer subsurface models below the area of interest by inverting the recorded wavefields on the surface. Inferences about the subsurface can be made by finding the Earth model $m$ which minimizes the misfit between the calculated data and the observed field data. The model $m$ can be numerically estimated by minimizing the functional

$$f(m) = \frac{1}{2} \| d - p(m) \|_2^2,$$

(2.29)

where $d$ is a vector containing all the observed data, and $p$ is a vector containing the predicted data.

The large size of the model vector $m$ and the high computational cost for calculating the prediction vector $p$ mandate the use of iterative local-gradient optimization schemes for minimizing the least-squares functional. In such schemes, the functional is linearized with respect to model updates to give the non-linear Gauss-Newton formula (Pratt et al., 1998)

$$H \delta m = J^\dagger (d - p(m)),$$

(2.30)

where $\delta m$ is the model update, and $H$ and $J$ are respectively the Hessian and the Jacobian matrices (Nocedal and Wright, 2006) of the objective function in equation 2.29. Next, the linearized system of equations is solved for a model update that minimizes the functional. After that, the model is updated and this procedure of linearization, solving for the update, and updating the model is repeated until convergence to a minimum value of the functional. This procedure is known as full-waveform inversion (FWI) (Tarantola, 1984a; Lailly, 1984; Virieux and Operto, 2009).

The functional above is known to be strongly non-linear with many local minima, and iterative local-gradient optimization algorithms often prematurely converge to a local minimum providing implausible subsurface models. Secondary local minima are often caused by cycle-skipping when the observed event and the corresponding predicted event have a phase difference of more than 3

\[3\]

Local minima might be caused by reasons other than cycle-skipping, and sometimes cycle-skipping may not cause a local minima.
half a period. A set of methods have been developed to remedy this issue of cycle-skipping by (1) reducing the complexity of data to be inverted and (2) excluding the cycle-skipped events. The methods are often referred to by many names depending on the formulation such as *multi-scale* (Bunks et al., 1995) or *hierarchical* (Asnaashari et al., 2012) FWI. In general, they use the reweighted objective function

\[
f(m) = \frac{1}{2} \| W(d - p(m)) \|^2_2, \tag{2.31}\]

where \( W \) is a weighting operator that excludes (by multiplication by zero) the cycle-skipped events in the data. As the model is gradually updated during the iterative inversion, previously cycle-skipped events become uncycle-skipped and the weighting operator is updated accordingly to include all the uncycle-skipped events. There are many successful applications using this approach for inverting low-frequencies in the early transmitted arrivals. Inverting scattered or reflection events from surface-seismic data, however, is still a challenge.

If no local minima or cycle-skipping is encountered in reflection FWI, there remain two key problems in inverting reflections and scattered data. The first problem is the slow convergence where a large number of iterations are needed to converge to a subsurface model that can explain the data. I empirically observe that hundreds of iterations are spent to gradually update the low-wavenumber components of the model and adjust the position of the high-wavenumber components accordingly.

The second problem is the non-linearity induced inconsistency which is dependent on both the operator \( H \), the Jacobian matrix \( J \), and the residuals \( (d - p(m)) \); \( H \) and \( J \) are often implemented assuming acoustic single scattering (i.e. under Born approximation), while \( (d - p(m)) \) contains elastic events with multiple scattering. This inconsistency is induced by events that cannot be explained by linearized inversion such as multiple scattering and mode-converted waves. With a large null space of \( H \), the linearized inversion fills this null space with artifacts in an attempt to explain multiple scattering, mode-converted waves, etc. As a result, the final model of reflection FWI (after hundreds of iterations) tends to be implausible and filled with artifacts, even though
it explains the data\(^4\).

The current solution to the three problems (cycle-skipping, the slow convergence with a large number of iterations, and null-space artifacts) is to have an accurate starting model which has the correct low-wavenumber components. This prior-knowledge requirement, however, is non-trivial, requiring a time-consuming velocity analysis, a high-level of expertise, and laborious manual picking. Therefore, this requirement cannot be satisfied for most reflection data, and therefore applications of reflection FWI are very limited to few cases in the literature.

**Proposed Solution**

Here, I build on the previously proposed approaches using the reweighted objective function in equation 5.1 to avoid cycle-skipping, and solve the remaining two problems (i.e. the slow convergence and the null-space artifacts) with two minor pre-conditioning steps to the hierarchical FWI algorithm. The first step is referred to as *offset-rolling*, where a small window of offsets from the uncycle-skipped data are inverted first for few FWI iterations, and then the offset window is shifted gradually toward the uncycle-skipped far offset. It can be shown numerically that this step enhances the low-wavenumber updates allowing for faster convergence when there are low-wavenumber errors in the starting model. The second pre-conditioning step removes the null-space artifacts by smoothing of cumulative updates from offset-rolling step. This is an ad-hoc preconditioning based on Occam’s razor where I favor the least-complex subsurface model, with the minimum high-wavenumber complexity as required to explain the data.

**Significance of the Contributions**

Alleviating the need for an accurate starting model will significantly reduce the time and the financial costs associated with producing velocity models for reflection imaging the subsurface. This will make FWI both practical to implement for many users and provide a much greater depth of inversion than previously possible.

\(^4\)Reflection FWI is an ill-posed problem with many possible solutions that can explain the data.
Organization of Chapters

The first two chapters of this part serve as an introduction to the final chapter which explains in detail the solution proposed above. First, I introduce the incomplete Gauss-Newton optimization scheme for FWI. The novelty in this chapter is that we apply the method in the time domain which has some practical advantages over the frequency domain Gauss-Newton optimization for FWI that has been proposed in previous work. The method is applied to Gulf of Mexico (GOM) streamer data, where we invert the diving waves and observe the challenges in inverting reflections in the field data. In this application, we find that FWI has the elements needed for it to invert for the smooth background velocity and the detailed reflectivity simultaneously. We finally show that FWI with Gauss-Newton optimization can invert for mid-wavenumber components of the velocity model.

In the following chapter, I demonstrate the process of creating a good initial model for 3D early arrival FWI. Traveltime tomography is usually the method for generating such models. However, the picking involved in creating initial model is overwhelmingly time consuming. Therefore, I proposed a strategy of picking the first arrivals where the idea is to start (picking) with the end (FWI) in mind. I developed a method that minimizes significantly the amount of picking needed to mature a smooth velocity model that is good enough as a starting model for FWI. That eventually enabled me to apply 3D FWI to a 3D GOM data in a timely fashion. This approach can help in maturing the shallow depths of the velocity model, but it would not provide a model below the reach of early-arriving waves.

Finally, the third chapter follows naturally after the realization of: (1) the ability of FWI to update the background velocity model and (2) the limitation of early arrivals tomography in maturing accurate deep starting models. In this final chapter, I proposed using FWI with inaccurate starting models, where reflections are used for updating the background model. The FWI method proposed in the chapter is modified for the original FWI, where we dynamically mask out problematic data, and the using Gauss-Seidel-Newton optimization, minimizes a coupling problem that I observed in synthetic and field data. Finally, I show that the proposed procedure is successful in inverting models using reflections, when the starting models are inaccurate.
For the proposed solutions in this part, we assumed a constant-density acoustic wave propagation for simplicity. The concept are, in principal, extendable to a more general elastic media. This, however, will be at a much greater computational expense. As a partial remedy, I propose a method to deal with amplitudes that cannot be explained with acoustic physics. The proposed approach is a prospect for future research with field data tests. Therefore, I include the theory and preliminary synthetic data results in the appendix.
Chapter 3

Time-domain Incomplete Gauss-Newton Full-waveform Inversion

3.1 Abstract

We apply the incomplete Gauss-Newton full-waveform inversion to Gulf of Mexico (GOM) data in the space-time domain. In our application, iterative least-squares reverse-time migration (LSRTM) is used to estimate the model update at each non-linear iteration, and the number of LSRTM iterations is progressively increased after each non-linear iteration. With this method, model updating along deep reflection wavepaths are automatically enhanced, which in turn improves imaging below the reach of diving-waves. The forward and adjoint operators are implemented in the space-time domain to simultaneously invert the data over a range of frequencies. A multiscale approach is used where higher frequencies are down-weighted significantly at early iterations, and gradually included in the inversion.

Synthetic data results demonstrate the effectiveness of reconstructing both the high- and low-wavenumber features in the model without relying on diving waves in the inversion. Results with Gulf of Mexico field data show a significantly improved migration image in both the shallow and deep sections.
3.2 Introduction

For deep subsurface imaging, waveform inversion (Tarantola, 1984b) should invert deeper reflections and later-arrival refractions. Unfortunately, standard FWI has low sensitivity to waveform residuals related to relatively weak deeper reflections compared to the stronger amplitude diving waves. The consequence is slow and often inadequate FWI convergence for reconstructing deep portions of the slowness model. To enhance the effectiveness of FWI for deeper reflections, we use a linear-inversion scheme instead of reverse-time-migration (RTM) (Baysal et al., 1983) for calculating slowness updates. Using this linear inversion, sharp boundaries are incorporated into the slowness model so that they implicitly enhance the model updating along the reflection wavepaths at subsequent iterations. This linear inversion is least-squares reverse-time migration (LSRTM) (Tarantola, 1984b; Plessix and Mulder, 2004; Dai et al., 2012).

Using LSRTM, a slowness-perturbation model is computed based on the Born approximation, where the background slowness is fixed during the linear inversion. The slowness model is then updated with the inverted slowness-perturbation model. After that, the linear inversion is repeated with the updated slowness model as a background slowness. Each linear-inversion and updating of the slowness model constitutes a non-linear iteration. This combined linear and non-linear inversion procedure is cyclically repeated until acceptable convergence.

The above procedure is a variation of Gauss-Newton optimization for FWI (Akcelik, 2002; Akcelik et al., 2002; Erlangga and Herrmann, 2009; Virieux and Operto, 2009). To avoid high computational and memory costs, the linear inversion is computed by an iterative conjugate gradient (CG) solver. The number of CG iterations is increased after each non-linear iteration, and is essential for an accurate model reconstruction. The algorithm is implemented in the time-space domain, and a multiscale approach is used to invert the data for a band of frequencies (Bunks et al., 1995; Boonyasiriwat et al., 2010); starting from a narrow band of low frequencies, and progressively including higher frequencies into the inversion.

In this paper, we review the time-domain incomplete Gauss-Newton FWI (TDIGN-FWI) algorithm, and illustrate its effectiveness on synthetic data that do not contain diving waves. Then, the inversion is applied to Gulf of Mexico (GOM) data. The resulting tomograms show significant
improvements in the deeper section compared to the starting model.

3.3 Theory

Newton’s method \cite{Pratt1998, Nocedal2006} for minimizing the residual difference \( r \) between the calculated and observed data can be written algebraically as

\[
\mathbf{s}_{k+1} = \mathbf{s}_k - \mathbf{H}_f^{-1}(\mathbf{s}_k) \nabla f(\mathbf{s}_k),
\]

where \( \mathbf{s}_k \) is the slowness model, \( \mathbf{H}_f \) is the Hessian matrix and \( \nabla f(\mathbf{s}_k) \) is the gradient of the objective function \( f(\mathbf{s}_k) = \frac{1}{2} \| \mathbf{r}(\mathbf{s}_k) \|^2 \) at the \( k \)-th iteration. By approximating the Hessian as

\[
\mathbf{H} \approx (\mathbf{J}^\top \mathbf{J}),
\]

where \( \mathbf{J} \) is the Jacobian matrix, we get the Gauss-Newton optimization formula

\[
\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_k \left( \mathbf{J}^\top_k \mathbf{J}_k \right)^{-1} \mathbf{J}^\top_k \mathbf{r}_k.
\]

A line search is used to estimate the step length \( \alpha_k \) because the approximation of the Hessian might not be an accurate estimate of the curvature for the non-linear misfit function. Instead of inverting the Hessian matrix, we iteratively solve the overdetermined system of equations

\[
\mathbf{J}_k \mathbf{g}_k = \mathbf{r}_k,
\]

using the same slowness model to get the search direction \( \mathbf{g} \). In other words, LSRTM is used to compute the search direction \( \mathbf{g} \) instead of RTM. Once the search direction \( \mathbf{g} \) and the line-search parameter \( \alpha \) are computed, the slowness model is updated using

\[
\mathbf{s}_{k+1} = \mathbf{s}_k - \alpha_k \mathbf{g}_k.
\]
and the Jacobian operator and the Hessian matrix are also updated according to the new slowness model. In the following section, we review the implementation of the Jacobian operator and its adjoint.

### 3.3.1 Time-domain Implementation of the Jacobian and its Adjoint

We follow a similar procedure to that of [Dai et al. (2012)] in deriving a time-domain implementation of applying the Jacobian matrix $J$ to the slowness perturbation vector $\Delta s$. We will describe our procedure for a single frequency, with the understanding that the method is fully implemented in the space-time domain. Each row of the matrix operation $\Delta p = J \Delta s$ for calculating the wavefield perturbation $\Delta p$ (indexed by receiver position $x_r$ and frequency $\omega$) from the slowness perturbation $\Delta s$ (indexed by the spatial position $x$) for a given source at position $x_s$ with a source wavelet $q(\omega)$ is a discretization of the integral equation

\[
\delta p(x_r, \omega) = 2 \int \omega^2 s_0(x) \delta s(x) p_0(x, \omega) G_0(x | x_r, \omega) \, dx,
\]

(3.5)

where $\omega$ is the frequency, $s_0$ is the background slowness, $\delta s$ is the slowness perturbation, $p_0$ is the background incident wavefield from the source, and $G_0$ is the Green’s function. This equation is the solution to the following system of partial differential equations

\[
(\nabla^2 + \omega^2 s_0^2) p_0(x, \omega) = -\delta(x - x_s) q(\omega),
\]

(3.6)

\[
(\nabla^2 + \omega^2 s_0^2) \delta p(x, \omega) = -2\omega^2 s_0 \delta s(x) p_0(x, \omega),
\]

(3.7)

which indicate that we can evaluate the integral in equation 3.5 by having two wave-propagation simulations.

Similarly, for the adjoint operation $\Delta s = J^\dagger \Delta p$ each row-vector multiplication is the discrete
approximation to the integral

$$\delta s (x) = 2 \int \int \omega^2 s_0 (x) p_0 (x, \omega) \times G_0 (x|x_r, \omega) \delta p^* (x_r, \omega) \, dx_r \, d\omega. \quad (3.8)$$

To evaluate this integral, two wavefields are simultaneously simulated by solving the two wave equations:

$$\left( \nabla^2 + \omega^2 s_0 (x)^2 \right) p_0 (x, \omega) = -\delta (x - x_s) q (\omega), \quad (3.9)$$

$$\left( \nabla^2 + \omega^2 s_0 (x)^2 \right) R^* (x, \omega) = -\delta (x - x_r) 2\omega^2 \delta p^* (x_r, \omega). \quad (3.10)$$

The solution to equation \textbf{3.10} is

$$R^* (x, \omega) = 2 \int \omega^2 G_0 (x|x_r, \omega) \delta p^* (x_r, \omega) \, dx_r. \quad (3.11)$$

By taking the zero-lag correlation and scaling by the background slowness, we get the integral in equation \textbf{3.8} i.e

$$\delta s (x) = \int s_0 (x) R^* (x, \omega) p_0 (x, \omega) \, d\omega = 2 \int \int \omega^2 s_0 (x) p_0 (x, \omega) G_0 (x|x_r, \omega) \delta p^* (x_r, \omega) \, dx_r \, d\omega. \quad (3.12)$$

Similar to the forward modeling operator, the integral in equation \textbf{3.8} can be computed by conducting two wave-propagation simulations and applying the zero-lag cross-correlation to the two wavefields. The above equations are implemented in the space-time domain for the applications in the following sections.
3.3.2 Physical Interpretation of TDIGN-FWI

In this section, we develop by example an intuitive reasoning that explains why TDIGN-FWI provides more accurate tomograms than standard FWI. Figure 3.1a shows a block-velocity model which contains one deeper reflector to constrain the shallow velocity anomalies. Because the deep layer has a slower velocity than the shallow layer, it will not generate refractions. Frequencies below 5 Hz are absent from the data as shown in Figure 3.1b, and the maximum source-receiver offset of 3 km is used for the synthetic data. The initial velocity model is homogeneous with a constant velocity of 2000 m/s. The shallow rectangular anomalies are larger in size than the minimum effective wavelength.

The block model does not generate refractions from the deep interface so that only the reflections from the deeper interface will be employed to reconstruct the square-shaped anomalies. FWI relies on reflections for reconstructing the shallow anomalies, and the velocity updates are attributed to the reflection wavepaths associated with the deep reflector and the boundaries of the anomalies. The construction of reflection wavepaths is dependent on the presence of sharp reflectors in the velocity model. Without the sharp boundaries, standard FWI fails to reconstruct the shallow anomalies as shown in Figure 3.1c.

LSRTM is known to focus reflections and diffractions into a sharp interfaces in the subsurface model. If such sharp interfaces are incorporated into the velocity model as is the case for TDIGN-FWI, these highly resolved reflectors and diffractors generate the wavepaths needed for reconstructing the shallow anomalies. As shown in Figure 3.1d, TDIGN-FWI make use of reflections, diffractions, multiples and prism waves to construct the anomalies and delineate the boundaries with high resolution. The mispositioning of the deeper reflector due to the shallow velocity error is reduced and the reconstructed reflector is nearly flat. In contrast, the reflectors in the standard FWI tomogram in Figure 3.1c are more distorted because the deeper reflections are not fully utilized for the same number of iterations. A prohibitively large number of iterations would be needed to accomplish the same results using a non-linear steepest descent optimization method.
Figure 3.1: a) Test velocity model, b) the source wavelet spectrum for forward-modeling the synthetic data and inversion, c) steepest descent FWI and d) TDIGN-FWI tomograms. The starting velocity model is a constant velocity model with the velocity 2000 m/s, and both of the inversion results have the same computational cost.
3.4 Test on Synthetic Data

We now apply the hybrid FWI to synthetic data computed for the more complicated Marmousi model, where 30 shot gathers are generated from shots evenly spaced along the surface with a fixed-receiver spread on the surface. A 7-Hz Ricker wavelet is used for generating the observed data, and the recording time is 5 seconds, with a sampling interval of 1 ms. The initial model in Figure 3.2a for the inversion is a smoothed version of the correct velocity model. A Gaussian smoothing filter is chosen such that first arrivals of the calculated data for the smoothed model are within half a cycle of the first arrivals in the observed data.

Applying the hybrid FWI algorithm to the synthetic data results in the Figure 3.2b tomogram, which is in almost total agreement with the actual velocity model in Figure 3.2c. The shallow section of the tomogram has relatively high resolution, and the steep dips of the deep section are reconstructed in the tomogram. The smoothness of the tomogram compared to the true model is expected given the low frequencies of the observed data. In this inversion, we start with two sequential linear iterations for the first non-linear iteration, and then increase the number of linear iterations by one for every non-linear iterations.
Figure 3.2: Application of the TDIGN-FWI to the Marmousi model data with a 7-Hz Ricker wavelet, where a) is the initial velocity model, b) is the final tomogram after 30 non-linear iterations, and c) is the true velocity model.
3.5 Application to GOM Streamer Dataset

The TDIGN-FWI is applied streamer data from the Gulf of Mexico. There are 515 shots with a 37.5 meter shooting interval, and the source-receiver offsets are from 198 meters to 6 kilometers, with a 12.5 meter receiver spacing. The trace length is 10 seconds with a 2 ms sampling interval. Prior to inversion, the data spectra are filtered by $\sqrt{i/\omega}$ and gained by $\sqrt{t}$ in the time domain to transform 3D to 2D geometric spreading. The source wavelet is estimated by stacking early arrivals from the near-offset traces.

We start the inversion with the data bandpass-filtered from 0-4 Hz, because there is reliable signal at 4 Hz. At later iterations, we widened the band of data frequencies to 10 Hz, and Figure 3.4 shows the initial velocity model and the final tomogram. The grid size for the tomograms is 301 by 1600 grid points in the vertical and horizontal directions, respectively, and the grid spacing is 12.5 meters. Figure 3.3 shows the convergence curve after the last reset, where the data residual decreases by more than 60 percent.

Figure 3.6 shows the migration images using the initial velocity model and the TDIGN-FWI tomogram. The TDIGN-FWI image is more focused. The spliced common-image-gathers (CIG) are flat in the final image, while the CIG’s are not flat for the initial image. This highlights the significant improvement to the velocity model. Figure 3.9 shows two shot gathers from the observed and calculated data and the match is generally good for early arrivals and most of the deeper reflections.

3.6 Limitations and Future Improvements

A problem with our approach is that the density is assumed to be a constant so that the TDIGN-FWI will introduce sharp velocity boundaries with the wrong velocity values. Those boundaries still help in updating the background velocity, which will improve the migration image. Such sharp boundaries can be removed before applying an FWI algorithm which inverts for more subsurface parameters than the acoustic velocity.

We chose to start the inversion with two LSRTM iterations and increase the number of LSRTM
iterations by one for every non-linear iteration. Our choice is heuristic based on tests with synthetic data. More work is needed for choosing optimal TDIGN-FWI inversion parameters.

3.7 Conclusion

We implemented and applied the TDIGN-FWI to a GOM dataset. The algorithm uses the LSRTM images as the slowness updates instead of the RTM images. The TDIGN-FWI uses the deep reflection data to define sharp boundaries in the velocity model. Those sharp boundaries generate wavepaths that are used by the inversion to build velocity updates for the deeper section. The definition of sharp boundaries and using them in calculating slowness updates are implicit within the algorithm. As a result, the quality of the migration images computed with the TDIGN-FWI tomogram appears to be highly resolved at both the shallow and deeper parts.
Figure 3.4: a) the initial velocity model, and b) final TDIGN-FWI tomogram of the band 0-10 Hz.
Figure 3.5: The wavepaths for diving waves (top), and reflections from different offsets (middle) and (bottom). This demonstrates that reflections wavepaths can update the model below the reach of diving waves.
Figure 3.6: Kirchhoff migration images using the initial velocity model (upper) and the final velocity model (lower). The spliced narrow panels are common image gathers. Flatter events in the common image gathers indicate a better velocity model.
Figure 3.7: Zoom view on the shallow section of the migration images of the initial (left) and the final (right) velocity models.

Figure 3.8: Zoom view on the deep section of the migration images of the initial (left) and the final (right) velocity models.
Figure 3.9: 0-10 Hz shot gathers from different parts of the survey. The observed data are shown in the top panels and the corresponding calculated data at the bottom.
Chapter 4

Efficient Wavefront Picking for 3D Tomography and Full-waveform Inversion

4.1 Abstract

We propose an efficient approach for picking first-break wavefronts on coarsely-sampled time slices of 3D shot gathers. The objective of our work is to compute a smooth initial velocity model for multiscale full-waveform inversion (FWI). Using interactive software, first-break wavefronts are geometrically modeled on time slices with a minimal number of picks. We pick on sparse time slices, perform traveltime tomography, and then compare the predicted traveltimes to the data in-between the picked slices. The picking interval is refined with iterations until the errors in traveltime predictions fall within the limits necessary to avoid cycle-skipping in FWI.

The proposed approach is applied to a 3D OBS dataset. Results indicate about 80% reduction in picking time compared to manual traveltime picking in the shot gathers and 97% reduction in the computer memory requirements. The final tomogram is sufficient as a starting model for early-arrival FWI.
4.2 Introduction

Subsurface velocity inversion methods such as full-waveform inversion (FWI) (Tarantola 1984a, Virieux and Operto 2009) require a good initial velocity model to (1) avoid falling into a local minimum, and (2) minimize the number of computationally-intensive iterations. Inverting traveltimes is the traditional method for generating accurate starting models (Bording et al. 1987, Dessa et al. 2004, Woodward et al. 2008), where first-break picking is the first step toward shallow tomographic inversion (Zhu et al. 1992, Zelt and Barton 1998).

As the seismic industry moves toward denser source and receiver coverage per square area, high density data is becoming a heavy burden for visualization and picking of early arrivals for tomography, especially in the presence of irregularities in the acquisition geometry and complex geology. Many automatic picking algorithms and approaches are proposed (Peraldi and Clement 1972, Gelchinsky and Shtivelman 1983, Coppens 1985, McCormack et al. 1993, Boschetti et al. 1996, Sabbione and Velis 2010, Blias 2012), with varying degrees of success. Some of the picking algorithms require some manual picks as a training set. Therefore, some human picking is still needed to provide reliable picks.

In addition, with many algorithms for auto-picking, the number of erroneous picks increases with lower signal-to-noise ratio (SNR), the presence of spurious events before the first arrivals, and the parameters needed by the auto-picker. Several authors try to analyze the different picks for a given trace to obtain a more reliable pick (Saragiotis et al. 2013, Yalcinoglu 2014). We empirically find that it is invalid to assume that a set automatic picks are representative of uncertainty of traveltimes, as a number of auto-picks might be clustered around the wrong traveltimes giving a false measure of confidence. At the end, someone has to validate the picks before conducting tomography to ensure consistency of the picks and the validity of the final tomogram. Consequently, the data are subjected to increasing rates of picking errors and hence, demand greater efforts in quality control (QC).

We now propose a new iterative approach that minimizes the time of manually picking traveltimes in dense 3D data. In this approach, the coordinates of the first-arrival wavefront are picked on a time slice of a shot gather, where the time-slice interval for picking can be very coarse in
the first iteration. Traveltime tomography is then applied on the picked wavefronts to compute a velocity tomogram. Using the tomogram, new early-arrival traveltimes (wavefronts) are predicted in between the picked time slices and compared to the observed data. When there are significant errors in the prediction, the time-slice interval is reduced in the next picking iteration. This procedure is repeated for the refined picks until the prediction error is less than half the period of the lowest frequency band that will be used in FWI.

A wavefront is parameterized by a number of picks, where the wavefront position between the picks is interpolated by a piecewise smooth function to form a contour. Using contours minimizes the number of picks needed to represent curved wavefronts. This is an alternative to the traditional first-break traveltome picking done on each trace of a common shot gather (CSG). The main advantage of our approach is that it determines the coordinates of the traveltime contours without being overwhelmed by the sheer number of traces within the survey. This is feasible for OBS-like surveys with a relatively small number of gathers with a dense wide-aperture data coverage. However, the approach can be extended to surveys with a large number of gathers if picking is automated on time slices, where the picks are used as seeds for automated picking. The wavefront contours are used as input to traveltime tomography to provide an initial model for FWI.

We apply the approach to ocean bottom seismic (OBS) data, and the results illustrate its effectiveness. The acquisition geometry of the survey is shown in Figure 4.1, where there are 234 OBS stations, spaced roughly 400 meters apart in a 26 by 9 grid with varying water depths with an average of 45 meters, and 124 sail-lines with 50 meters spacing. Each sail-line covers 18 km of distance with a shot spacing of 50 meters. The data contain low-frequencies as low as 2 Hz with a minimum reliable frequency of 4 Hz. To avoid working with a low number of traces per gather, we use common receiver gathers (CRGs) for traveltime picking, traveltime tomography, and FWI.

The paper is organized such that concepts and procedures are demonstrated using the field data example of the OBS survey. First we describe the picking methodology and highlight the merits and the demerits of the proposed approach. Later, we show the results from tomography based on the picks, and finally show the FWI tomogram that uses the starting velocity model from
traveltime tomography. We finally conclude with some remarks and possible future directions.

4.3 Methodology

The key idea is to pick first-arrival traveltimes from time-slices of recorded CRGs instead of the traditional approach of picking traveltimes from each trace of a CSG as depicted in Figure 4.2. For wavefront picking on time-slices, the survey geometry does not have to be regular, but the wave-sampling has to be dense enough to be able to pick events on the time slice. In such dense surveys, conventional picking is difficult to ensure areal consistency of the picks. In addition, generating common-offset or common-azimuth gathers for irregular 3D datasets requires regularization of the data. On the other hand, areal displays of wavefronts are more convenient for picking the first-arrival wavefronts, as shown in Figure 4.3.

First-arrival wavefronts are almost circular in the early time-slices and quickly deform into irregular shapes with complex patterns in the later time slices. Such wavefront shapes can be optimally parameterized by piecewise smooth functions.

The number of picks that can accurately represent a wavefront on a time slice is based on real-time interpolation during the picking process. An interactive application, using a graphical user interface (GUI), is developed that uses a pointer position to compute the wavefront before making the pick. Therefore, the interpreter (i.e., the user) can optimize the number of picks by moving the pointer as far as possible from the last pick as long as the interpolated picks follow the tracked wavefront as shown in Figure 4.4. Next, we discuss pick-interpolation algorithms that can run in real time for faster plotting and picking.

4.3.1 Pick Interpolation

There are several options for interpolating the curve

$$\mathbf{c}(u) = (c_x(u), c_y(u)),$$ (4.1)
Figure 4.1: OBS survey geometry. Red dots represent the location of OBS stations and the green lines represent air-gun sail lines.

Figure 4.2: A schematic plot of the traditional picking approach where first arrivals are picked on every trace in sparse seismic inlines. Red mark represent a source position, the green triangles represent receivers, and the yellow dots are first-arrival traveltime picks.
Figure 4.3: A schematic plot of the proposed picking method where first arrivals are picked on time slices. Colored contours represent early-arrival wavefronts, where the color indicates the traveltime associated with the picked wavefront.
between picks (or the last pick and the pointer position), where the scalar \( u \in [0, 1] \) is the interpolation variable and the points \( c(0) \) and \( c(1) \) are at two consecutive picks. The simplest interpolation is a linear interpolation between consecutive picks at the Cartesian positions \( \mathbf{x}_i = (x_i, y_i) \) and \( \mathbf{x}_{i+1} = (x_{i+1}, y_{i+1}) \), where

\[
\begin{align*}
c_x(u) &= (x_{i+1} - x_i)u + x_i, \\
c_y(u) &= (y_{i+1} - y_i)u + y_i.
\end{align*}
\]

Curved wavefronts, however, require a large number of picks to characterized using linear segments. Splines represent an attractive alternative for interpolating curved functions between picks, which in turn reduce the number of picks needed to track curved wavefronts. Quadratic and cubic Bézier spline curves (Farin, 1997; Gallier, 2000), in particular, are flexible and inexpensive to compute but they require additional picks to characterize the curvature between picks. Appendix A discusses further the details on optimized picking using Bézier splines.

For most time slices, a radial line extending from the source toward any azimuth \( \theta \in [0, 2\pi] \), crosses the first-arrival wavefront only once. For such cases, we can use a simple interpolation scheme in polar coordinates as

\[
\begin{align*}
r(u) &= r_i + (r_{i+1} - r_i)u, \\
\theta(u) &= \theta_i + (\theta_{i+1} - \theta_i)u,
\end{align*}
\]

where the coordinates can be transformed into Cartesian coordinates using

\[
\begin{align*}
c_x(u) &= x_{src} + r(u) \times \sin(\theta(u)), \\
c_y(u) &= y_{src} + r(u) \times \cos(\theta(u)).
\end{align*}
\]

Here, \( x_{src} \) and \( y_{src} \) are the source Cartesian coordinates and \( (r_i, \theta_i) \) are the polar coordinates of the \( i^{th} \) pick; \( r \) is the radial distance from the source to the wavefront contour and \( \theta \) is the azimuth. Even
though this interpolation is not for general cases, it is optimal for early times where wavefronts are nearly circular, which can be represented using few picks.

Figure 4.5 depicts one time slice from the OBS data; the dark-blue areas indicate zones where there is no data coverage. To avoid aliasing, we high-cut filter the data such that the minimum apparent wavelength\(^1\) is at least twice as large as the receiver interval. Plotting time slices of a given shot/receiver gather gives an impression of continuous sampling of the wavefield. Picks could be made interactively on such plots to track the wavefronts.

At the early time slice in Figure 4.5 the wavefront tends to be nearly circular, and therefore, we try to pick it using the interpolation algorithm discussed above. With linear interpolation, a large number of picks is required to accurately represent the curved wavefront. Spline interpolation requires fewer picks, while polar-interpolation requires the minimum number of picks to track the circular wavefront. Later time slices require more picks as in Figure 4.6. Different interpolation methods can be used for different wavefront shapes to minimize the picking time. For the OBS data, we mainly pick with polar interpolation between picks.

### 4.3.2 A Strategy for Selecting Time-Slice Intervals

The time-slice interval for picking is selected according to the goal of early-arrival tomography. As mentioned above, our aim is to generate a starting model for multiscale FWI. The bandwidth for the first FWI run governs the time-slice interval used for picking. To avoid cycle-skipping, data calculated from the initial model should be within half a period of the observed data. To satisfy this condition, the time-slice interval should be within half the period of the dominant frequency used in the first iteration of FWI. For the OBS data, FWI is applied to 2-4 Hz data, and therefore the time-slice interval should be less than 125 milliseconds.

Choosing the picking interval based on the cycle-skipping criteria is rather pessimistic. We can minimize the amount of picking significantly by utilizing redundant information in the first arrivals. For the OBS data, we start picking with 800 ms time-slice interval (i.e. 8 times the required interval), and then we invert the picked wavefront by traveltime tomography to get a

---

\(^1\)estimated for a horizontally propagating wave in the water layer
velocity tomogram. Then, we use the tomogram to predict the waveforms and traveltimes for time-slice interval of 400 ms. If the traveltime difference between the observed and the predicted traveltimes meets the criteria for avoiding cycle-skipping, we stop picking. Otherwise, we refine the picks at 400 ms interval. This procedure is repeated at 200 ms picking time-slice interval, where the prediction errors satisfy the criteria for avoiding cycle-skipping.

With the selected time-slice interval, the memory requirement for storing the data is significantly reduced since intermediate time-slices do not need to be visualized during the picking process. The time-sampling interval is 4 ms for the OBS data, and the associated size of the data containing the first arrivals is more than 80 GB. For the conventional approach, this size requires a high end workstation to hold the volume in memory and an efficient algorithm to minimize delays and cache misses related to frequent jumping between slices and gathers. On the other hand, with time-slice picking, the sparse slices require less than 2 GB of memory for all the gathers. This tremendous reduction in memory cost allows us to perform picking on computers with small memory. In addition, the visualization of the small dataset is fast with minimal plotting delays.

4.3.3 Quality Control

Consistency of traveltime picking is crucial for computing a high-quality tomogram. Therefore, we develop QC tools to monitor the consistency of the picks made throughout the survey. Since wavefronts expand away from the source point with time, the picked wavefront contours from different time slices must not cross. Therefore, for every picked gather, picked wavefronts from different time slices are displayed as in Figure 4.7 to ensure that the picks from different time slices are not crossing one another.

Picks from different gathers must be also be consistent. Figure 4.8 shows the time-slices at 2.4 seconds for different gathers. If the wavefronts of all the gathers are picked, we can generate an apparent velocity plot as a function of OBS position $x_s$, where the apparent velocity is defined as

$$v_a(x_s, \theta, t) = \frac{r(x_s, \theta)}{t}. \quad (4.8)$$
Here, \( r(x_s, \theta) \) is the radial distance from the OBS station to the wavefront as indicated by the arrows in Figure 4.8. Figure 4.9 depicts the apparent velocity as a function of shot position for a given time and azimuth. A color-coded block is plotted on-top of each source position, where the color-coding represents the apparent velocity. If a pick in a given gather does not follow the areal trend, it is reviewed and adjusted to follow the regional trend. This QC exercise is repeated for every picked time-slice for several azimuths before moving to the next step.

### 4.3.4 Inversion using Pseudo- Receivers

The picking and QC procedure described above will result in a set of piecewise-continuous contours of the wavefronts at different times. For ray tomography, the continuous contours are sampled uniformly in the azimuthal direction and each sample is used as a pseudo-receiver with the first-arrival traveltime equal to that of the time slice used in picking the wavefront contour. This sampling produces pseudo-receivers, each with a first-arrival traveltime. For the OBS data, the travel time-contours are resampled uniformly with one degree angular spacing, as shown in Figure 4.10.

We now discuss application of traveltime tomography on these associated traveltimes.

### 4.4 Traveltime Tomography

The early-arrival traveltimes can be modeled using the Eikonal equation (Aki and Richards, 2002; Cerveny, 2005)

\[
|\nabla \tau(x)|^2 = \frac{1}{v^2(x)}.
\]

If multiple first-arrival wavefronts are encountered for a given azimuth, the maximum apparent velocity is plotted. This scenario, however, was not encountered for the OBS dataset.
with the boundary condition

\[ \tau (x_s) = 0, \]  

where \( v(x) \) is the velocity model, and \( \tau (x) \) is the traveltime from the source position \( x_s \) to the \( x \) position. We solve the eikonal equation using the finite-difference method proposed by Jeong et al. (2007), which gives traveltime tables throughout the 3D model. From the tables, we obtain the traveltime values at the pseudo-receivers and compare them with those obtained from our traveltime picks. The traveltimes are inverted for a 3D model, with a 50 m grid spacing in the \( x \)-, \( y \)- and \( z \)- directions.

To compute the tomogram, we find the slowness model \( s \) that minimizes the objective function

\[ f(s) = \frac{1}{2} \sum_{i=1}^{N_t} \left( w_i \delta \tau_i (s) \right)^2, \]  

where \( \delta \tau_i \) is the difference between the observed (from wavefront picking) and the calculated traveltimes, \( w_i \) is a positive weighting factor, and \( N_t \) is the number of data samples used in the inversion. Traveltime picks at the far-offsets are subjected to higher uncertainty due to the very weak amplitudes of the first arrivals. In addition, far-offset traveltime picks often dominate the residual due to their magnitude, compared to the near offsets. To balance the residuals, we use a weighting factor

\[ w_i = \frac{1}{\tau_{\text{obs}}^p}, \]  

where \( \tau_{\text{obs}} \) is the observed traveltimes, and \( p \) is a positive power parameter chosen empirically; the higher the value of \( p \), the more biased the inversion will be toward the near-offset picks. We empirically choose \( p = 0.5 \), which gives balanced updates to the shallow and the deep portions of the model. The objective function is minimized with Gauss-Newton iterations (Heath, 1996) by solving the preconditioned linearized system of equations for the slowness update \( \delta s \):

\[ WJPy = W\delta t, \]  

\[ \delta s = Py, \]
where $\delta t$ is a vector containing the traveltime residuals, $W$ is a diagonal matrix with $W_{ii} = w_i$, $J$ is the Jacobian matrix for the traveltime misfit function, $P$ is a preconditioning operator (a Gaussian smoothing operator), $\delta s$ is the slowness update, and $y$ is a temporary vector.

The adjoint of the Jacobian matrix $J^\dagger$ smears the traveltime residuals along the raypaths computed using the traveltime tables [Dessa et al. 2004]. This back-projection approach is preferred to back projection using adjoint-state method [Sei and Symes 1994; Leung and Qian 2006; Taillandier et al. 2009], due to the low number of rays which is faster to trace.

The role of the preconditioner $P$ is to ensure that the model update is very smooth. At early iterations a Gaussian smoothing filter is chosen to span the whole model and the width of the Gaussian filter is gradually reduced to 800 meters, which is roughly twice the distance between the OBS stations to avoid aliasing. The height of the Gaussian filter is 15\% of its width.

At every nonlinear iteration, the slowness update is obtained using the conjugate gradient (CG) method (three iterations), and the slowness model is updated using $s_{k+1} = s_k + \alpha \delta s$. The step length $\alpha$ is obtained using the golden-section method [Chapra and Canale 2001].

**Tomography Results**

Figure 4.11 shows the final tomogram calculated from the traveltimes picked from the time slices. Figures 4.12 and 4.13 depict, respectively, the depth slices and cross-sections of the final tomogram. The velocity gradually increases with depth, and there is a decrease in velocity at the depth of 1 km. A gentle anticline structure is prominent at the middle of the volume. There is one shallow refractor (depth $\sim$500 m) and a deep one at 1.5 km depth.

The primary objective for traveltime tomography is to compute a starting velocity model, where the calculated waveforms are not cycle-skipped relative to the observed waveforms. Figure 4.14 shows the traveltime error histogram associated with the final tomogram. Since the errors are predominately less than half the period of 4 Hz, the tomogram is used for FWI with a maximum frequency of 4 Hz. Nevertheless, a close look at the nature of traveltime residuals in Figure 4.15 indicates that the residuals are mostly related to short wavelength heterogeneities that cannot not be inverted due to the smoothing constraints used during the inversion. Relaxing the smoothing
(i.e. unregularized/unpreconditioned inversion) will minimize the residuals but will also give geologically unrealistic tomograms. This highlights one of the limitations of traveltime tomography and the proposed method. FWI can however recover such short scale features.

4.5 Full-Waveform Inversion

AlTheyab et al. (2013) describe the theory of Gauss-Newton optimization for time-domain FWI. Gauss-Newton optimization amounts to applying least-squares reverse-time migration (Plessix and Mulder 2004; Dai et al. 2012) to the waveform residuals at each non-linear iteration to estimate the search direction. Here, we apply the same method to the OBS data, using the traveltime tomogram as a starting model. The number of iterations for the linear (LSRTM) inversion is set at five. 2-4 Hz transmission data are inverted, and the results presented here are for the 3D model that excludes the fringe zone (4 km around the OBS nodes). To overcome aliasing caused by the large OBS node spacing, the model updates are preconditioned by Gaussian smoothing similar to the one used in traveltime tomography.

Figures 4.16 shows the initial velocity model and the final tomogram. It is evident that the final tomogram contains some low-wavenumber updates and more geological details. Figure 4.17 compares the predicted data from the traveltime tomogram and the final FWI tomogram. The data fit is better from the FWI tomogram, and the fit is expected to get better after more FWI iterations.

4.6 Conclusion

We present an efficient method for picking first-arrival wavefronts for data with dense wavefield sampling. This method allows for picking geographically on time slices as opposed to the traditional picking of traveltimes in CSGs or common offset gathers. The wavefronts are geometrically modeled with a minimum number of picks using splines and interpolation in polar coordinates.

The picking method is applied to 3D OBS data and the traveltime picks are inverted to get a velocity tomogram. The tomograms predicts first-arrivals that are in-phase with the observed data.
The estimated savings in picking time is about 80% compared to manual picking of traveltimes from each trace of a CSG. In addition, the memory requirement for the proposed approach is less than 5% of that required in the conventional approach. This leads to a robust implementation with modest computer capabilities.

The wavefront picking error grows with increasing trace separation and the wavelengths associated with picked wavefronts. Therefore, this approach is applicable to dense arrays for which high frequencies (higher than 10 Hz) are not aliased spatially.

In our application, we decided to use manual picks directly in traveltime tomography. An alternative approach is to use the picks as seeds for automated picking, that eventually would improve the resolution of traveltime tomography. The picking approach can also be used for picking surface waves and possibly reflection waves.

### 4.7 Appendix A: Picking using Bézier Splines

Bézier splines can be intuitively understood using De Casteljau’s algorithm, which plots a continuous smooth curve in a 2D plane using three control points as follows:

1. Define three control points \( p, q, \) and \( s \) by their x- and y-positions, where the points \( p \) and \( s \) mark the ends of the curve, and the \( q \) controls the curvature of the curve as shown in Figure 4.18.

2. As demonstrated in Figure 4.19, evaluate the continuous curve \( c(u) \) by sampling along the variable \( u \in [0, 1] \), where for a given \( u \) define a new \( p' \) between \( p \) and \( q \) where \( |p' - p| = u|q - p| \), similarly define \( q' \) between \( q \) and \( s \) such that \( |q' - q| = u|s - q| \). Finally, define the interpolated point \( c(u) \) between \( p' \) and \( q' \) such that \( |c - p'| = u|q' - p'| \).

The position of \( q \) controls the curvature and the tangent of the interpolated curves at \( p \) and \( s \).

To use the algorithm above to pick wavefronts, we pick one point along the observed wavefront by pressing the mouse button and then dragging (i.e. continue pressing while moving) to form a line tangential to the wavefront as shown in Figure 4.20. Then, we release the mouse button, and the new mouse position form the end point for the interpolated curve. The interpolation is done
in real-time, such that as we move the mouse pointer, the curve changes accordingly. Therefore, we move as far as we can along the wavefront while still tracking the wavefront. This picking procedure is repeated for the next segments to form a complete pick of the wavefront.
Figure 4.4: Interpolation between picks in the polar coordinates to track curved wavefronts.
Figure 4.5: Early time slice with their picks (yellow dots, and red interpolated polygon in the polar coordinates) using, from top to bottom: linear interpolation, Bézier quadratic splines, and linear interpolation in the polar coordinates. Gray scale images represent the recorded waveform at the active time slice. No data coverage zones are indicated by the dark blue color.
Figure 4.6: Late time slices with their picks (yellow dots, and red interpolated polygon in the polar coordinates). Gray scale images represent the recorded waveform at the active time slice. No data coverage zones are indicated by the dark blue color.

Figure 4.7: Picks within a single gather are checked and adjusted such that the picked time contours must not cross. The red mark indicate the shot position and the colored contour lines indicate wavefront picks, where the color indicates the time at the picked time slice.
Figure 4.8: Estimating apparent velocity from different gathers for a given time-slice and azimuth. The arrows indicate the radial distance from the source to the wavefront for a given azimuth.
Figure 4.9: Panels for QC-ing azimuth-dependent apparent velocity for the 2.4 sec time-slice. The arrow indicates the azimuth for the given apparent velocity. This plot is used to detect anomalous picks that do not follow the trend of surrounding gathers, as the one indicated as mispick. Mispicks are corrected to follow the regional trend.

Figure 4.10: Traveltime picks on pseudo-receivers sampled from the continuous wavefront contours.
Figure 4.11: The final 3D tomogram after 50 iterations.

Figure 4.12: Depth slices from the final 3D tomogram.
Figure 4.13: Cross-sections through the tomogram.

Figure 4.14: Traveltime residuals after 50 iterations of tomography. The traveltime errors are predominantly within half a period ($\pm 0.125$ sec) of the maximum frequency (4 Hz) that is used in the first iteration of FWI.
Figure 4.15: Traveltime residuals for a given shot. Due to strong regularization, tomographic residuals cannot fit short wavelength features below the sea bottom.
Figure 4.16: (Top) Traveltime tomogram and (bottom) early-arrivals FWI tomogram using the traveltime tomogram as a starting model. The survey fringe zones are excluded from the plots.
Figure 4.17: (Left) selected sections from observed data traces, and the corresponding traces from finite-different solutions to the 3D wave equation using (middle) the traveltime tomogram and (right) FWI tomogram (5 iterations). 0.6 seconds automatic gain control (AGC) is applied to the traces for display.

Figure 4.18: Quadratic Bézier curve parameterization by three points in 2D plane.
Figure 4.19: De Casteljau’s algorithm for evaluating Quadratic Bézier curve for $u$ equals, respectively from top to bottom, 0.25, 0.5, and 0.75.

Figure 4.20: Quadratic Bézier curve parameterization by three points in 2D plane.
Chapter 5

Full-waveform Inversion of Reflections with Inaccurate Starting Models

5.1 Abstract

I present a method for inverting seismic reflections using full-waveform inversion (FWI) with inaccurate starting models. For a layered medium, near-offset reflections (with zero angle of incidence) are unlikely to be cycle-skipped regardless of the low-wavenumber velocity error in the initial models. Therefore, we use them as a starting point for FWI, and the subsurface velocity model is then updated during the FWI iterations using reflection wavepaths from varying offsets that are not cycle-skipped.

To enhance low-wavenumber updates and accelerate the convergence, we take several passes through the non-linear Gauss-Seidel iterations are carried out, where traces are inverted from a narrow range of near offsets and finally end at the far offsets. Every pass is followed by applying smoothing to the cumulative slowness update. The smoothing is strong at the early stages and relaxed at later iterations to allow for a gradual reconstruction of the subsurface model in a multiscale manner. Applications to synthetic and field data, starting from inaccurate models, show significant low-wavenumber updates and flattening of common-image gathers after many iterations.
5.2 Introduction

The goal of FWI (Tarantola, 1984a; Virieux and Operto, 2009) is to infer a subsurface model by minimizing the difference between the observed and calculated data. The FWI misfit function, however, is known to be non-linear with many local minima, often caused by cycle-skipping (Gauthier et al., 1986). Therefore, local-gradient optimization algorithms often get stuck at a local minimum providing implausible subsurface models.

A set of methods have been developed to mitigate the cycle-skipping problem, while maintaining the original FWI formulation. Cycle-skipped events are excluded by filtering in different domains (Bunks et al., 1995; Asnaashari et al., 2012), or by evaluating the phase lag between the observed and calculated data for each event and muting cycle-skipped events (Bi and Lin, 2014). In general, such methods apply FWI in multistages. At every stage, part of the data is inverted using the objective function

$$J_i(\delta m_i) = \frac{1}{2} \| W_i \Delta d \left( m_i + \delta m_i \right) \|^2_2,$$

(5.1)

where the subscript $i$ denotes the stage number, $\Delta d$ is the difference between the observed and calculated data, $W_i$ is a weighting operator that excludes (by multiplication by zero) the cycle-skipped events in the data, $m_i$ is the initial model at the $i$–th stage, and $\delta m_i$ is the cumulative model updated to the initial model $m_i$ needed to fit the data. After each stage, a new model is computed

$$m_{i+1} = m_i + \delta m_i,$$

(5.2)

where $m_{i+1}$ is the initial model for the following stage. Note that at each stage the objective function is minimized with many iterations. As the model is gradually updated, previously cycle-skipped events become uncycle-skipped and the weighting operator is updated accordingly to include new uncycle-skipped events in the next stage. Such approaches are effective for inverting transmitted waves (direct arrivals, head waves, and diving waves) which cover the shallow sections of the Earth. However, they are often insufficient for updating the background velocity model using reflections below the reach of transmitted waves.
It is believed that the primary cause of FWI failure in updating the low-wavenumber components of the model is the weakness of the tomographic terms in the update kernels of FWI (Zhou et al., 2012). Therefore, several approaches for enhancing the tomographic components were proposed for reflection FWI while minimizing the difference between the observed and calculated data (Zhou et al., 2012; Xu et al., 2012; Wang et al., 2013; Brossier et al., 2013). Such approaches use scale separation between high-wavenumber and low-wavenumber components of the model where they are inverted at every iteration in separate steps: the imaging step (where the reflectors are mapped in the subsurface) and the tomography step, where the low-wavenumber updates are computed based on the data misfit function.

There are two main disadvantages to the proposed approaches. First, they are based on the single scattering assumption, and therefore higher-order scattering cannot be handled accurately. In addition, the transmitted waves have to be removed from the data before inversion. The second disadvantage is that the implementations bear the additional cost of several wavefield simulations needed in the imaging and the tomography steps. To avoid the single scattering assumption and the additional costs, AlTheyab et al. (2013) proposed using Gauss-Newton optimization when inverting reflections using FWI, where low-wavenumber updates along reflection-wavepaths are naturally enhanced within the FWI iterations.

However, the success of all aforementioned solutions for updating the low-wavenumber components is limited to recovering localized mid-wavenumber anomalies when the reflectors are close to their accurate positions, which requires an accurate starting model. This raises the question whether the weakness of reflection wavepaths is the sole cause for the failure in updating the low-wavenumber components of the model.

We believe that the main problem is the coupling between the low-wavenumber components of the model to all the high-wavenumber components, as well as the contradicting updates along reflection wavepaths from different angles of incidence. To solve this problem, we propose splitting the FWI problem at each stage such that we sequentially invert a narrow-offset range of traces, starting with the short offsets and ending at the far offsets. We will refer to FWI with a narrow-offset range as constant-offset FWI. In each constant-offset FWI, Gauss-Newton optimization is
used, and the final velocity model of a constant-offset FWI is the initial model to the following constant-offset FWI. This approach is closely related to the non-linear block Gauss-Seidel iteration method with overlapping blocks [Tai 1992; Gutiérrez et al. 2011]. Here, each constant-offset FWI is naturally tuned for enhancing low-wavenumber updates along reflection wavepaths.

In the following sections, we will elaborate on the coupling problem and the proposed solution to enhance the low-wavenumber updates. Later, we show the results of applying the proposed method to synthetic and field data, where low-wavenumber errors in the starting model are corrected after many stages. Finally, we comment on the generalization of the method.

5.3 Theory

For a reflection or scattering event, there are a group of receivers along a planar surface where the travel-time gradient of the specified reflection is zero ($\nabla \tau (x, y) \approx 0$), which maps to stationary events in the data where the apparent velocity is zero. This is usually true for near-offset reflections from a layered medium. Such events are unlikely to be cycle-skipped during FWI for any smooth starting model. Therefore, we design the weighting operator $W$ in equation 5.1 to only allow such uncycle-skipped events into FWI, and the windows are gradually widened to include more uncycle-skipped data as we proceed to later stages with an improved initial model. This process can be automated such that cycle-skipping is detected in an adaptive manner as in [Bi and Lin 2014].

This multistage FWI with automatic detection and exclusion of cycle-skipping is extremely slow and impractical if the starting model contains significant low-wavenumber errors. That is because subsurface reflectors will be mispositioned in the early iterations, and the positioning has to be gradually corrected with a large number of iterations. To illustrate the cause of slow convergence, we consider the following scenario.

Consider the two-layer model shown in Figure 5.1 which will give a single reflection from the deep interface between the two layers. For a homogeneous starting velocity model, the angle of incidence, for a horizontally layered model, is related to the model wavenumber via the relation
where $\omega$, $c$, and $\theta$ are, respectively, the angular frequency, the initial velocity, and the angle of incidence. When there are low-wavenumber errors in the initial velocity model, the model phase $\phi(k_z)$ will be a weighted average of the phases from different angles and frequencies that cover the same wavenumber $k_z$ (i.e. the apparent depth of the reflector will be a weighted average of apparent depths from different angles). At the second iteration, phase delays of predicted specular events will vary depending on the angles of incidence $\theta$. It follows, then, that there will be both positive and negative phase delays between the observed and calculated waveforms depending on the angle of incidence.

In the following FWI iterations, the mispositioned reflector will act as a secondary source for updating the low-wavenumber components of the model via reflection wavepaths. Figure 5.2 shows the wavenumber coverage for a model with a single reflector (see Mora (1989) for details) where the tomographic terms are related to reflection wavepaths. Even though there is little overlap between the wavenumber coverage between the diffraction terms from different frequencies and angles of incidence, the corresponding tomographic terms are completely overlapping at the small wavenumbers near the origin. This illustrates the strong coupling between the high wavenumbers (from different angles and frequencies) and the low wavenumbers reconstructed by the tomographic
Figure 5.2: The wavenumber coverage from diffraction terms and tomographic (reflection transmission) terms of different frequencies.
terms in Figure 5.2. Therefore, the positive and negative phase errors result in conflicting updates in the tomographic terms (i.e. the overlapping zone in Figure 5.2), and, consequently, negligible low-wavenumber updates from reflection residuals.

To resolve the strong coupling and the conflicting updates, the objective function in equation 5.1 at the \(i\)-th stage, is regrouped into different terms based on source-receiver offset,

\[
J_i(\delta m_i) = \frac{1}{2} \sum_N^N \| W^h_i \triangle d^h_i (m_i + \delta m_i) \|^2,
\]

where the superscript \(h\) denote the offset index and \(N\) is the number of offsets bins in the data. For the model in Figure 5.1 the source-receiver offset is directly related to angle of incidence \(\theta\) by the relationship \(\text{Sirgue and Pratt, 2004}\)

\[
\cos \theta = \frac{z}{\sqrt{q^2 + z^2}},
\]

where \(z\) is the depth of reflector, and \(q\) is the half the distance between the source and the receiver. Therefore, decomposing the objective function here is equivalent to a decomposition in the wavenumber domain based on the angles of incidence. For general media, the objective function is decomposed such that each term of the decomposed objective function has data that is sensitive to a limited area of the model’s wavenumber spectrum \(\delta m\) using the diffraction terms (see Figure 5.2). With this direct mapping between terms in the decomposed objective function and the high-wavenumber components of the model, we can solve this system using Gauss-Seidel iterations starting from a zero update vector \(\delta m_i^0 = 0\), and the model update at every \(h\)-th iteration is computed using

\[
\delta m_i^h = \delta m_i^{h-1} + \arg \min_x \| W^h_i \triangle d^h_i (m_i + \delta m_i^{h-1} + x) \|^2.
\]

Here, we apply a few iterations of FWI with Gauss-Newton optimization for the model update \(x\) which is added to the initial model of the \(i\)-th stage \(m_i\) and the update from the previous Gauss-Seidel iteration \(\delta m_i^{h-1}\) to minimize the misfit function of the constant-offset data at the \(h\)-th.
offset bin. Note that low-wavenumbers components of the model, coupled by the tomographic terms, are freely updated during Gauss-Seidel iterations.

Due to the coupling problem described above, the low-wavenumber components of the model will oscillate between different passes through the Gauss-Seidel iterations. An under-relaxation scheme is used as a preconditioner to prevent this oscillatory behavior, where the under relaxation operator $S_i$ is used during the update step between stages

$$m_{i+1} = m_i + S_i \delta m_i^N$$

where the $S_i$ is a Gaussian smoothing operator applied to the cumulative update, the Gaussian smoothing filter is wide at early stages to allow updates to the very low-wavenumber components of the model, and the width of the filter is reduced gradually at later stages. We empirically find that smoothing along geological dip further accelerates the convergence process. Now, $m_{i+1}$ is the initial model for the next stage of constant-offset inversions.

Figure 5.3 shows the corresponding FWI workflow. With each pass through the Gauss-Newton FWI block, a few iterations of FWI with the incomplete Gauss-Newton optimization (Akcelik et al., 2002; Erlangga and Herrmann, 2009; AlTheyab et al., 2013) are executed. The block in the workflow labeled select uncycle-skipped data designs the weighting operator $W$ in equation 5.1, which masks the cycle-skipped data. The loop inside the area labeled offset-rolling minimizes the objective function using the Gauss-Seidel iterations.

5.4 Synthetic and Field Data Examples

Now, we demonstrate the effectiveness of the proposed workflow on a synthetic data test. Data with 12 Hz peak-frequency and a 6 km maximum source-receiver offset were generated using the true model shown in Figure 5.4(a). The reweighted Gauss-Newton FWI (excluding cycle-skipped events) without the proposed workflow is applied to the synthetic data using the starting velocity model in Figure 5.4(b), which gives the results in Figure 5.4(c) after 300 iterations. The shallow part above 0.7 km is recovered mostly due to inverting diving waves. However, significantly more
Figure 5.3: The proposed FWI workflow for inverting reflections with Gauss-Seidel iterations (offset-rolling).
(a) True Velocity Model  
(b) Starting Velocity Model  
(c) Conventional FWI Tomogram  
(d) Proposed FWI Tomogram

![Figure 5.4: Conventional and proposed FWI applied to synthetic data.](image)
Figure 5.5: Proposed FWI applied to field data reflections. Reverse-time-migration images are overlaid onto the velocity models, and the three gray-scale panels are angle-domain common-image gathers. Flatter events in the image gathers indicate a more accurate velocity model.

Iterations are needed to recover the low-wavenumber components of the model in the deeper portions of the model. Moreover, the tomogram has many high-wavenumber artifacts that do not relate to any feature in the true model. On the other hand, using the proposed workflow gives the final tomogram in Figure 5.4(d) after 50 stages. The final tomogram is free from the high-wavenumber artifacts while maintaining the high-resolution of the shallow channels. In addition, the deeper fast layer at 1.5 km depth is positioned at the correct depth in the final tomogram.

Figure 5.5 depicts the results for inverting 2D GOM streamer data with a 4 km maximum offset and a 3-10 Hz frequency range. To illustrate that the updates are only inverted by the reflections, transmitted waves are muted. With a homogeneous starting velocity model of 2000 m/s the angle gathers have large moveout. The tomogram seen in Figure 5.5 is after 40 stages and the angle gathers are nearly flat indicating improvements to the velocity model.
5.5 Conclusions

We proposed a method for inverting reflections starting from inaccurate velocity models. At every stage, uncycle-skipped events are grouped according to a narrow-range of on source-receiver offsets, and the data from each offset are inverted sequentially, where the final model of a constant-offset FWI is the initial model for the next inversion at a wider source-receiver offset. Because of the direct mapping from offset to incidence angle in the wavenumber spectrum, this sequential inversion is related to Gauss-Seidel method. The constant offset formulation is applicable to layered media. Complex media might require an alternative formulation of Gauss-Seidel based on angle of incidence and/or frequencies. This will be the subject of a future research.

The Gauss-Seidel iterations have two advantages over the full-domain Newton inversion; it naturally enhances low-wavenumber updates, and it is easier to precondition with under-relaxation schemes. Here, we avoid the single scattering assumption and explicit scale separation, and the reflection data can still be inverted simultaneously with transmitted waves with Gauss-Seidel iteration, as shown in the synthetic data test. Our approach does not require generating common image gathers, which will be an advantage in 3D applications.

We observe that the method fails to invert strong multiples, post-critical reflections, and anisotropic effects in field data. We believe this is due to the physics mismatch, where constant-density acoustic FWI is failing to explain the actual amplitude and waveform variations with offset. To alleviate problems with multiples, we prefer fast starting velocities which minimizes contributions of multiples to FWI updates.

5.6 Appendix A : Convergence of the proposed iterative scheme

To analyze the convergence of the iterative scheme above, we rewrite equation 5.7 as

\[ m_{i+1} = m_i - S_i \Gamma(m_i). \]  

(5.8)
Here, we regard FWI algorithm above as a non-linear process $\Gamma (m_i)$ that takes $m_i$ as an input and give $-\delta m_i^N$ as output, which is the model update needed to fit the observed data. We know that

$$\Gamma (m^*) = 0,$$  \hspace{1cm} (5.9)

where $m^*$ is the true subsurface model, because the data residual for the true velocity is zero. For regularized Gauss-Newton FWI with a single iteration per offset-bin the non-linear update can be expressed as

$$\Gamma_{GN} (m_i) = \sum_{h=1}^{N} \left( (W_i^h J_i^h)^\dagger W_i^h J_i^h + \alpha I \right)^{-1} (J_i^h W_i^h)^\dagger W_i^h \Delta d_i^h (m_i + \delta m_i^{h-1})$$  \hspace{1cm} (5.10)

To characterize the convergence of this optimization scheme, we use the model error $e$ for a given offset at two consecutive iterations

$$m_{i+1} = m^* + e_{i+1} \hspace{1cm} (5.11)$$
$$m_i = m^* + e_i \hspace{1cm} (5.12)$$

which we substitute into the update equation \[5.8\]

$$m^* + e_{i+1} = (m^* + e_i) - S_i \Gamma (m^* + e_i).$$  \hspace{1cm} (5.13)

Using the truncated Taylor expansion,

$$\Gamma (m^* + e_i) = \Gamma (m^*) + \tilde{\Gamma}_i (m^*) e_i,$$  \hspace{1cm} (5.14)

where for regularized Gauss-Newton optimization

$$\tilde{\Gamma}_i (m_i) = \left( (W_i^h J_i^h)^\dagger W_i^h J_i^h + \alpha I \right)^{-1} (W_i^h J_i^h)^\dagger W_i^h J_i^h$$

$$= (H_i^h + \alpha I)^{-1} H_i^h,$$  \hspace{1cm} (5.15)
which has the spectral radius

\[
\rho \left( \left( H^h_i + \alpha I \right)^{-1} H^h_i \right) \leq \frac{\rho \left( H^h_i \right)}{\rho \left( H^h_i + \alpha I \right)} < \frac{\lambda_{\text{max}}}{\lambda_{\text{max}} + \alpha},
\]

where \( \lambda_{\text{max}} \) is the maximum Eigenvalue for the Hessian matrix \( H^h_i \). Therefore, the spectral radius of the linearized operator \( \rho \left( \tilde{\Gamma}_i (m_i) \right) < 1 \). Substituting the truncated Taylor expansion in equation 5.14 into equation 5.13 we get

\[
e_{i+1} \approx e_i - S_i \Gamma (m^*) - S_i \tilde{\Gamma}_i (m^*) e_i \tag{5.17}
\]

and using equation 5.9 we reach the simplified form

\[
e_{i+1} \approx e_i - S_i \tilde{\Gamma}_i (m^*) e_i \tag{5.18}
\]

\[
\approx \left( I - S_i \tilde{\Gamma}_i (m^*) \right) e_i \tag{5.19}
\]

\[
\approx \left( I - S_i \tilde{\Gamma}_i (m^*) \right)^{i+1} e_1, \tag{5.20}
\]

The relaxation operator \( S \) is designed such that \( 0 < \lambda_{1,2,\ldots} (S) < 1 \), which guarantees that the spectral radius of the iteration matrix \( \rho \left( I - S_i \tilde{\Gamma}_i (m^*) \right) \) is less than one at every iteration, or exactly one if it is stuck in a local minima (i.e. model error does not change with stages). Therefore, the system is convergent under the first-order approximation. This procedure, however, does not guarantee monotonic convergence of the original objective function, but it guarantees that the model error is reduced between stages.
REFERENCES


Boonyasiriwat, C., G. Schuster, P. Valasek, and W. Cao, 2010, Applications of multiscale wave-


Cerveny, V., 2005, Seismic ray theory: Cambridge University Press.


Duret, F., and E. Forgues, 2015, 4D surface wave tomography using ambient seismic noise: 77th EAGE Conference and Exhibition - Workshops.


Leung, S., and J. Qian, 2006, An adjoint state method for three-dimensional transmission travel-
——, 2010, Seismic interferometry: Cambridge University Press.


Tai, X.-c., 1992, Parallel function decomposition and space decomposition methods with applications to optimization, splitting and domain decomposition: Presented at the BEIJING MATHEMATICS, Citeseer.


APPENDICES

Appendix A

Kinematic Full-waveform Inversion: Time-domain Implementation

A.1 Abstract

I present a method for inverting signs of the events in a recorded traces. Instead of subtracting the waveforms, the misfit function consists of the normed difference between the arctangent of the predicted and arctangent of the observed waveforms. The proposed inversion yields an algorithm similar to that of conventional FWI, except for a reweighting of the residuals at each iteration. Our numerical results show that this kinematic inversion is less sensitive to strong contrasts in model parameters when we run constant-density acoustic FWI on variable-density elastic data.

A.2 Introduction

Seismic records often have elastic-events with different scales of amplitudes. It is non-trivial to account for these widely varying scales during full-waveform inversion, especially with the assumption of a constant-density acoustic medium. Sigmoid functions have the capability to
threshold the data amplitudes and with the right choice of a sigmoid function, an FWI objective-function can be formulated.

I propose an objective function that uses the arctangent of the waveform. Finding the velocity model that minimizes this objective function in effect inverts for the sign of the waveform and minimizes the influence of the amplitudes. The arctangent function shrinks the values of the argument from the range of \((-\infty, \infty)\) to the range of \((-\frac{\pi}{2}, \frac{\pi}{2})\) (See Figure A.1). In addition, the arctangent function is smooth and continuously differentiable function, with the derivative

\[
\frac{\partial}{\partial x} \arctan(x) = \frac{1}{1 + x^2}. \tag{A.1}
\]

I invert for arctangent of the wavefields in the time-domain for a subsurface model that kinematically reproduces the events, regardless of the amplitude difference. This is useful for many applications that involve difficult data with amplitude issues, as is the case, for example, with traces characterized by a weak receiver-coupling to the ground. In the following section, we derive an FWI objective function that uses the arctangent function.

### A.3 Theory

I define the arctan-data residual as
\[ r(x_s, x_r, t) = \arctan(d_{\text{obs}}(x_s, x_r, t)) - \arctan(d_{\text{clc}}(x_s, x_r, t)) \]

where \(d_{\text{obs}}\) is the observed data, \(d_{\text{clc}}\) is the calculated data which are implicitly dependent on the slowness model \(s(x)\), and \(t, x_s, x_r\) are, respectively, the time, and the source and receiver positions. The observed and calculated data must have amplitudes much larger than \(\pi\), otherwise this condition can be satisfied by straightforward scaling of the observed and calculated data. I do not include such scaling in the following derivations for the sake of conciseness. The objective function is the \(l_2\)-norm of the residuals:

\[ f(s(x)) = \frac{1}{2} \sum_{x_s,x_r,t} r(x_s, x_r, t)^2. \]

The partial derivative of the objective function with respect to slowness field \(s(x)\) at the image point \(x\) is

\[ \frac{\partial}{\partial s(x)} r(x_s, x_r, t) = -\frac{1}{1 + d_{\text{clc}}(x_s, x_r, t)^2} s^2(x) \times \ddot{w}(t) \ast G(x|x_s, t) \ast G(x|x_r, t), \]

\[ (A.4) \]

where \(\ddot{w}(t)\) is the time-second partial derivative of the source signature \(w(t)\), \(G(x|x_s, t)\) and \(G(x|x_r, t)\) are, respectively, the Green’s functions from the source-position \(x_s\) and from the receiver-position \(x_r\) to the image point \(x\). The symbol \(\ast\) denotes convolution in the time-domain. The linearized forward-modeling operator is
\[
\delta r (x_s, x_r, t) = -\frac{1}{1 + d_{clc}(x_s, x_r, t)^2} \sum_x s^2(x) \times \ddot{w}(t) \star G(x|x_s, t) \star G(x|x_r, t) \delta s(x),
\]
(A.5)

and the back-projection of the residual wavefield into the image-space is computed using

\[
\delta \tilde{s}(x) = -\sum_{x_s, x_r, t} \frac{1}{1 + d_{clc}(x_s, x_r, t)^2} s^2(x) \times \ddot{W}(t) \star G(x|x_s, t) \star G(x|x_r, t) \cdot \delta r(x_s, x_r, t).
\]
(A.6)

Equations A.5 and A.6 can be written in algebraic form as

\[
\begin{align*}
\delta r &= WL \delta s, \\
\delta \tilde{s} &= L^\dagger W \delta r,
\end{align*}
\]
(A.7) (A.8)

respectively, where \( L \) is the Born forward-modeling operator and its adjoint \( L^\dagger \) is the reverse-time migration operator, and \( W \) is a diagonal matrix with elements defined as

\[
W_{ij} = \begin{cases} 
0 & i \neq j, \\
\frac{1}{1 + d_{clc}(x_i, x_j, t)^2} & i = j,
\end{cases}
\]
(A.9)

where the superscript indices indicate position within the data vector.

The operator \( L \) and its adjoint \( L^\dagger \) are similar to conventional FWI operators. The only difference in implementation is the diagonal operator \( W \), which has a negligible computational additional cost compared to the cost of the \( L \) operator. The role of the weighting operator \( W \) is to steer the inversion from strong events that are already fitted kinematically in the previous iterations, toward fitting weaker events in the next iteration.
A.4 Physical Interpretation of Objective Function

If we consider one time-sample from a seismic trace, we can evaluate any objective function as a function of amplitude differences between the observed and calculated values. Figure A.2 plots the objective functions with respect to the amplitudes in the observed and calculated time-samples. The values for the conventional FWI objective function (i.e. $f = (d_{\text{obs}} - d_{\text{clc}})^2$) are zero when the residual is zero (i.e. $d_{\text{obs}} = d_{\text{clc}}$), and increase proportionally to the difference-squared for non-zero residuals. For the proposed arctangent misfit function, the function has low values when the amplitudes match in polarity. Therefore, the function is flat in the positive-positive and negative-negative quadrants (See Figure A.2 (bottom)). This flatness is depended on the scale of the data; the stronger the amplitude in the data, the less sensitive the objective function is to amplitude variations and the flatter the objective function. However, overscaling of the data leads the weights in the $W$ operators to be close to zero, which reduces the convergence to unacceptable values. Therefore, I advise a conservative usage of the method by avoiding over-scaling of amplitudes.

For a successful application, the arctangent data should be scaled such that dominating reflectors in the data such as direct arrivals, water-bottom reflections, and top-of-salt reflections transformed to the same values. This is demonstrated in the successful application presented in the following section.

A.5 Synthetic Data Test

Now, we apply the proposed inversion to synthetic pressure data generated using elastic finite-difference modeling. The P-wave and S-wave velocity and the density models used for generating this test data are shown in Figure A.3. In the models, there is a water-bottom at the depth of 400 m and a salt-top reflector at 600 m. Both generate strong and complicated seismic events that dominate the amplitudes. The goal of this test is to try to invert for P-wave velocity by applying a constant-density acoustic FWI to the elastic data. This is a realistic application in seismic exploration considering the higher cost of elastic 3D FWI compared to constant-density
Figure A.2: The objective functions for one time-sample: (top) for conventional objective function and (bottom) for the arctangent objective function.
120 shot gathers were modeled using the elastic model shown in Figure A.3, which has a grid spacing of 10 m. Each shot has receivers with an offset range from 0.5 km to 6 km with a receiver spacing of 10 m. The frequency range is 2-15Hz with time-sampling interval of 0.6 ms, and a total recording time of 3 s.

Figure A.4 depicts the constant-density acoustic and elastic data. The images are scaled such that the amplitudes of the direct arrivals are the same as those at the near-offsets. There is still, however, notable differences in the scale of amplitudes for later events, although they match in kinematics. The stronger amplitudes in the elastic data are due to the contrasts in the elastic parameters in the elastic model used for generating the data, where these amplitudes cannot be acoustically modeled using realistic velocity contrast alone. This is especially true for the water-bottom and salt-top reflections. Figure A.4 also shows the elastic data scaling after taking the arctangent. With the arctangent applied to the data, the amplitude range of the strong events is similar to that of the weaker events, which will give the proposed inversion equal weights for all the events.

Two types of FWI were applied to the elastic data with the initial velocity model shown in Figure A.5. The first one uses the conventional FWI, and the other uses using the proposed arctangent FWI. Figure A.5 shows the final tomograms after 15 iterations for both, where the conventional FWI tomogram has an anomalously strong water-bottom reflector, near source artifacts, and a false image from the water-bottom multiple at the depth of 1200 m, crossing the deep anticline structure. Those issues characterize the failure of conventional FWI in explaining the amplitudes in the elastic data by imposing strong velocity contrasts in the tomogram. The arctangent FWI tomogram, on the other hand, does not suffer from those problems. The water bottom is reconstructed reasonably well, and the inverted water-bottom reflector sufficiently explains the primary event and its free-surface 1st-order multiple. False events still exist in both tomograms, which are likely to be the result of strong mode-converted waves, which cannot be accounted for with the proposed method.
Figure A.3: The P-velocity, S-velocity, and density models used for generating the elastic data.
Figure A.4: Several shot gathers of constant-density acoustic data (top), elastic data (middle), and the arctangent of the elastic data after scaling (bottom).
Figure A.5: Initial velocity model (top), conventional FWI (middle), and arctangent FWI (bottom) tomograms.
A.6 Conclusion

I proposed an objective function that minimizes the effect of strong amplitudes in the observed data. Such amplitudes cannot be explained by a constant-density acoustic model due to many reasons such as incomplete physics, imperfect source and receiver coupling to the ground, to name a few. The proposed method inverts for the arctangent of the data after proper scaling, which scales the strong events down to match the amplitudes of the weaker events. The application to elastic data demonstrates the robustness of the method in inverting strong events in elastic data, when constant-density acoustic medium assumption is used.

Optimal scaling of the data is the subject of future research. Also, other sigmoid functions need to be tested and compared to the arctangent function to choose the function with optimal cost and operator properties.