Chapter 1

Introduction

Finding the world's remaining petroleum reservoirs will rely on increasingly advanced geophysical techniques. Since the boom of the modern oil industry more than one and a half century ago, petroleum explorationists have found all the oil than can be found with relatively simple methods. Those 'easy oil' are typically buried in shallow depth, located in simple geologic setting, and are easily accessible given the technologies of the time. As a result, the world's remaining oil are located in deep area beneath the earth surface, and are usually within complex geological settings. To pin-point oil reservoir in those areas, advanced geophysical techniques such as 3D seismic imaging, are essential. In fact, 3D seismic imaging is the most important tool in aiding the oil industry to find those 'difficult oil'.

Seismic imaging process can be roughly divided into two steps: first, a velocity model of the earth is obtained by processing surface seismic data; second, the velocity model is used to reposition the surface seismic data to their correct subsurface locations, creating a image of the subsurface. The final products of the process—seismic images depend heavily on the quality of the velocity models. Inaccurate velocity models result in wrong depth of subsurface structures, make true structures disappear or even create spurious structures. These image artifacts are particularly common in geologically complex areas where accurate velocity models are difficult to obtain. To obtain accurate velocity models in those areas, it is important to obtain accurate near-surface velocity models first. In fact, in a lot of the remaining oil rich regions, especially on land, complex geology is synonym to complex near-surface.

The importance of the near-surface has been recognized since the beginning of the seismic imaging era. In the early days, explorers were dealing with relatively simple near-surface, where simple assumptions such as 1D layered earth model is enough. As a result, measuring slopes and arrival time on seismic profiles is enough to derive velocity models (Slitcher, 1932; Dobrin, 1960; Grant and West, 1965). With increasing geological complexity, along with the introduction of computers into seismic imaging, geophysicists start to include more and more lateral variations into the near-surface model assumption, e.g., by introducing dipping layers (Mota, 1954; Musgrave, 1967), or even arbitrary layer configurations (Hagedoorn, 1959; Palmer, 1981; Hampson and Russell, 1984; Schneider and Kuo, 1985). With those more and more realistic assumptions about the near-surface, modern algorithms are much better at estimating near-surface velocity in comparison to their predecessors. The state-of-art methods are based on propagating rays in the near-surface (Olson, 1984; White, 1989) to derive the large-scale structure of near-surface velocity on a gridded model. However, in the few remaining areas representing the 'difficult oil' scenarios, velocity models derived from ray-based methods are not accurate enough for imaging deeper reflectors (Marsden, 1993; Bevc, 1995; Hindriks and Verschuur, 2001).

Applications of ray-based methods in areas with complex geology have two major limitations: first, rays are high-frequency approximation to seismic waves. Such approximation is only accurate when the sizes of velocity structures are at least several times larger than the dominant wavelength of the seismic waves. Yet in geologically complex areas, quite often velocity structures are as large as (if not smaller than) seismic wavelength, breaking the assumption of the ray approximation of seismic waves. Second, ray based methods only use traveltime information from data to update velocity, ignoring waveform information all together, this is insufficient for accurate near-surface velocity model building since waveform carries important information regarding small-scale velocity anomalies.

Wave-equation based velocity estimation methods (Tarantola, 1984; Mora, 1987; Luo and Schuster, 1991; Chavent and Jacewitz, 1995; Pratt et al., 1998; Biondi and Sava, 1999; Zhang and Biondi, 2013) inherently overcome the first disadvantage of ray-based methods by employing the wave equation. However, the use of wave-equation does not necessarily guarantee the use of both traveltime and waveform information. Using wave-equation with traveltime only is still a viable option to perform velocity estimation (Luo and Schuster, 1991). This option is better than ray-based methods, but still yield a smooth model which is unfavorable for accurate near-surface velocity model building. Waveform inversion (Tarantola, 1984; Mora, 1987; Pratt et al., 1998), on the other hand, by using both traveltime and waveform information with wave-equation, is a theoretically promising candidate for accurate near-surface velocity model building. Despite the first waveform inversion theory formulation back in the early 1980s, ray-based methods still were the dominant choice for near-surface velocity model building due to computational power limitations. The rapid growing computational power in recent years has allowed the application of waveform inversion on field datasets (Sheng et al., 2006; Sirgue et al., 2009; Brenders, 2011; Plessix et al., 2010). Such advancement in computational power makes waveform inversion a practically feasible tool for near-surface velocity estimation, in addition to its theoretical advantages.

WAVEFORM INVERSION AND CHALLENGES FOR REAL DATA APPLICATION

Waveform inversion was formulated in the literature for the first time in the early 1980s (Lailly, 1983). It uses wave-equation based forward modeling methods to synthesize seismic data which is then compared against the recorded data, the resulting data difference can be used to update velocity model. The process is usually done iteratively, and the iterative process stops when the data difference is small enough. In this section, first, I will give a summary of how waveform inversion works; second, I will discuss some implementation choices based on geophysical considerations; finally, I will present some major challenges with the conventional waveform inversion schemes that prevent robust field data applications.

Waveform inversion algorithm: how does it work

Waveform inversion is a non-linear problem, in the sense that the modeled waveforms change non-linearly with respect to the changes of the velocity model. Although non-linear problems can be solved by exhaustive search of the solution space, the computational cost of a single waveform-inversion objective-function evaluation prevents such exhaustive search from finishing within any reasonable amount of time. Hence, waveform inversion is usually solved with local descend methods. With these methods, the non-linear problem was linearized around the current solution and solved as a linear problem. Such linearization is carried out successively, at each solution of the previous linear problem. In other words, the waveform inversion problem is solved in an iterative manner, where each iteration represent solving a linearized problem. This is summarized in the algorithm 1:

Algorithm 1 Pseudo code of waveform inversion

```
for iteration =1,n do
    Calculate gradient of objective function with regard to the model.
    Calculate update direction using the current gradient and/or previous gradients.
    Calculate steplength of the update direction.
    Update model.
end for
```

Time domain vs. Frequency domain

As with the other wave-equation based methods in seismic imaging and tomography, waveform inversion can be implemented in both time domain (Tarantola, 1984; Gauthier et al., 1986; Crase et al., 1990; Pica et al., 1990; Sun and McMechan, 1992) and frequency domain (Pratt et al., 1998; Ravaut et al., 2004). The difference is how the wavefields and the gradients are calculated. Time domain implementations use temporal wavefields cross-correlation to obtain gradients, while frequency domain implementations use multiplication of frequency slice(s) of wavefields to obtain gradients. Correspondingly, data differences are measured by comparing (bandpassed) time domain data for time domain implementations, whereas the differences are measured by comparing frequency slice(s) of data for frequency domain implementations. Each type of implementation has its own pros and cons, geophysically and computationally.

First, for large 3D applications, gradients obtained in time domain scheme are superior than that from frequency domain scheme for similar computational cost. Gradient calculation in waveform inversion is similar to reverse time migration. While time domain schemes calculate gradient component from all the data frequencies within a certain bandwidth at once, frequency domain schemes usually calculate gradient from each frequency slice separately. Separate gradient calculation for different frequency slices increases computational cost, and limits the amount of frequency slices that can be utilized.

Second, data residual evaluation for time domain scheme are more readily interpretable by a geophysicist than that from frequency domain scheme. Evaluation of data residuals involves comparing corresponding events in recorded data and observed data. This evaluation makes much more geophysical sense for bandpassed time-domain data than for sinusoid in frequency slices of data. Residual evaluation for time-domain data allows individualized comparison of corresponding events from different interfaces in the subsurface. On the other hand, residual evaluation for frequency domain data only yield a single time-shift for a single frequency slice, at most. The differences in the information obtained regarding data mismatches can make a huge difference in real data applications of waveform inversion, where determining the quality of the results is much more difficult.

However, time domain schemes do have disadvantages. Time domain schemes are usually more computationally intensive than frequency domain schemes. Computational cost of waveform inversion mainly comes from two parts: floating-point operations and I/O cost. For floating-point operations, time domain schemes are dominated by finite difference propagation of wavefields, frequency domain schemes are dominated by matrix factorization for solving the Helmholtz equation. These two operations cost roughly the same in large 3D inversion, while frequency domain schemes has much smaller cost in 2D application and small 3D application (Operto et al., 2007; Vigh and Starr, 2008). However, taking into considerations of the I/O cost, time domain schemes become much more expensive for large 3D applications. As mentioned before, gradient calculation procedure in waveform inversion is similar to that of reverse time migration, hence gradient calculation requires storing at least one wavefield, which is is a huge 4D volume for 3D time domain inversion. Such huge 4D volumes are much bigger than the size of the computer memories, and can only be written out to disk for storage and read back in when needed. But for frequency domain schemes, the wavefield volume is only 3D due to the few number of frequencies slides, which can be stored in memory without incurring the I/O cost. This big difference in wavefield size and resulting I/O cost leads to the computational disadvantage of time domain schemes.

In summary, time domain schemes have the advantage of better gradient quality and easier data residual evaluation, while possess the disadvantage of the I/O and memory cost associated with large 3D applications. In my implementation, I choose time domain scheme over frequency domain scheme because of the first two geophysical advantages, especially the inherent advantage of data residual evaluation. For the computational disadvantage of the time domain scheme, I mitigate it by proper geophysical optimization.

Data choice: Transmission and Reflection

Practically, it is almost impossible to match the entire recorded data in real data applications, given the complex physics phenomena that generate the data and the limited computational resources we have to simulate all these physics phenomena. As a result, it is important to choose the part of data that are suitable for the inversion.

Ideally, data should have good signal to noise ratio, and can be reproduced on synthesized data if the velocity model is correct. In other words, we should only match data that can be correctly modeled by the wave-equation we use in the inversion. For large 3D applications, computational constraints have mandated the use of the acoustic wave-equation in the inversion. On the other hand, in real data, particularly land data, non-acoustic events such as converted waves and surface waves are not uncommon. These non-acoustic events can not be reproduced from the acoustic wave-equation, even when the correct near-surface velocity model is given. As a result, data matching must exclude these non-acoustic data. This criteria leads to different muting strategies for reflection and transmission data.

Reflection data bounce at least once on its path from source to receiver (Figure 1.1). The bounce happens at the interface in the near-surface where impedance changes. In a shot gather, reflections from the near-surface usually arrives at early time, and exist across all the receiver offsets. In complex environments, especially on land, near-offset reflections tend to be contaminated by strong noise such as surface waves. Those noise are usually much stronger than the reflection data itself, making the near-offset reflections not suitable for waveform inversion (Figure 1.3). As a result, only middle to far offset reflections from the near-surface should be used for the inversion.

Transmission data travels from source to receiver without bouncing at any interface in the subsurface (Figure 1.2). Such data includes diving waves and refractions, and is likely to exist in a environment where velocity increase with depth. Since earth velocity generally increase with depth, transmission data are usually present in recorded data, especially with the long offset recording in seismic acquisition these days (Wang et al., 2003; Magesan et al., 2007; Beaudoin, 2010; Moldoveanu et al., 2012). In addition, transmission data mostly goes through near-surface with very good coverage, thus being very useful for near-surface velocity estimation. Most importantly, transmission data usually arrives before other signals and coherent noise (Figure 1.3), making its signal to noise ratio excellent for the purpose of waveform inversion.

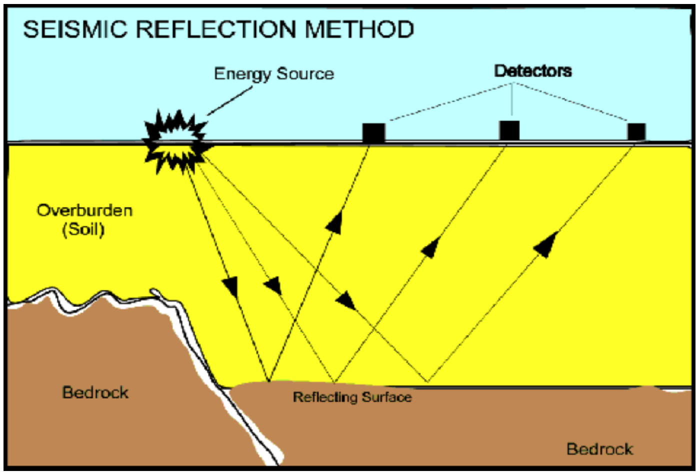

Figure 1.1: Diagram showing reflection in the earth, source:www.geologicresources.com . [**NR**] chap1/. refldiag

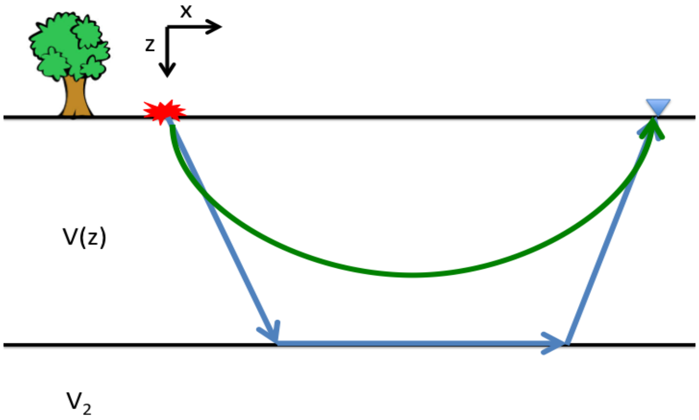

Figure 1.2: Diagram showing transmission in the earth, blue denotes refraction path, green denotes diving wave path. [**NR**] chap1/. refrdiag

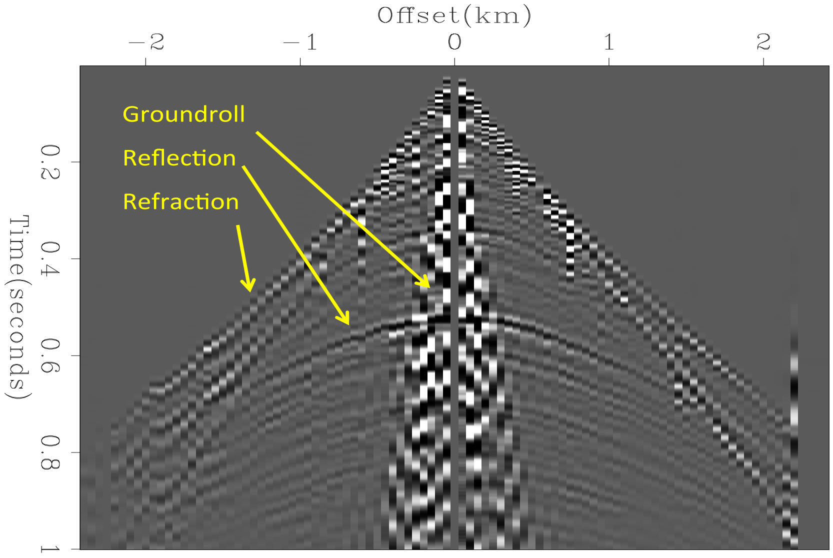

Figure 1.3: Example of typical shot gather acquired onshore. Notice how near-offset reflections are contaminated by strong surface waves while refractions and diving waves are virtually untouched by noise. [**NR**] chap1/. reallandd

Challenges in FWI

In waveform inversion, we can obtain the correct velocity model by using a initial model that is relatively close to the true model and by reproducing the exact recorded data from wave-equation forward modeling. It may sound simple, but those two conditions are very difficult to meet, especially in real data applications. The difficulties are inherent to the theory, which means both time domain schemes and frequency schemes will encounter the difficulties. In addition to that, the choice of time domain schemes brings the aforementioned computational complication that needs to be addressed.

The first condition of close starting model means forward modeled data from the starting model should be within half-cycle traveltime difference of recorded data (Virieux and Operto, 2009). Thus, if the inversion were to work, either we have a close starting model, or recorded data contains such low frequency that the low-frequency quarter-wavelength traveltime difference is big enough even for a bad starting model to converge. Since in oil and gas exploration, such low-frequency data is not commonly available yet, we have to come up with a close starting model.

The second condition of exact data reproduction, especially data amplitude reproduction with the correct velocity model is even more difficult to meet. Current computational power only allows us to use acoustic wave-equation for large scale waveform inversion. In field data applications, even after careful exclusion of non-acoustic events such as converted waves and surface waves, the remaining data are still affected by non-acoustic phenomena such as elastic effects and attenuation. Even with the correct velocity model, acoustic wave-equation can not reproduce the exact phase and amplitude of the recorded data. The aforementioned non-acoustic phenomena mainly affect data amplitude, and data traveltime and phase can be affected by other non-acoustic phnomena such as anisotropy. However, in this thesis, the discussion is only restricted to the isotropic cases.

In summary, the requirement of close starting models, the assumption of exact data reproduction with the correct model, and the I/O challenge for large 3D application of time domain waveform inversion schemes are the main challenges for real data applications of waveform inversion. In my thesis, I will develop various methodologies to address these challenges.

THESIS CONTRIBUTION

The central goal of this thesis is to develop a practical waveform inversion scheme for near-surface velocity estimation that: a), is able to give a high-resolution near-surface velocity structure compared with that from conventional ray-based methods; b) overcomes the geophysical and computational challenges of the conventional waveform inversion in real data applications.

The first major contribution is a new objective function that relies heavily on phase and waveform matching, while ignoring absolute amplitude. Conventional WI objective function relies on matching phase, waveform and absolute amplitude. This is difficult in real data application. Real data amplitude is determined by factors such as source signal strength, earth velocity, earth density, earth attenuation, source/receiver coupling, preprocessing. Most of those factors can not be modeled with the acoustic wave equation used in the inversion. Hence inversion results will be far from the true near-surface models if modeled data matches observed data in terms of the absolute amplitude. In the new objective function, matching absolute amplitude is de-emphasized by proper scaling of the observed data traces and modeled data traces.

The second major contribution is a new wave-equation tomography workflow that relaxes the starting model requirement of waveform inversion, yet still obtain a high-resolution result. Waveform inversion focus on waveform comparison, while assuming minimal traveltime difference between observed data and modeled data. Such assumption is not always guaranteed, especially in geologically complex areas. In those areas, Wave-equation Traveltime Inversion (WTI), through minimizing traveltime differences, can close the gap between a poor starting model and the starting model required by waveform inversion. In the new workflow, I combine the two methods by first running WTI, followed by waveform inversion. In this way, long wavelength errors are fixed before short wavelength errors can be corrected. This workflow combines the advantage of the two methods, overall making the final result high resolution, while only demanding a moderately good starting model.

The last major contribution is a low-frequency Random boundary condition that eliminates the I/O cost associated with storing and accessing large wavefields. Waveform inversion gradient calculation requires the correlation of source wavefield and data residual wavefield. The two wavefields propagate in different time directions. Hence to perform the correlation, at least one of the wavefields has to be saved beforehand. These wavefields are large four dimensional cubes for 3D data, and storing is only possible on disk. As a result, correlation involves a lot of I/O cost associated with writing wavefield to disk and read it in. I overcome this I/O cost by modifying the gradient calculation process, with a modified version of the original random boundary (Clapp, 2009) to accommodate the low-frequency wave propagation in waveform inversion. This alternative completely avoids I/O cost.

THESIS OVERVIEW

Chapter 2: Building initial models by WTI-I propose a two-step workflow to relax the starting model requirement of the waveform inversion. The workflow first uses WTI to update the long-wavelength components of the model, followed by waveform inversion to update the short-wavelength components of the model. I demonstrate the effectiveness of the workflow using synthetic and real land data examples of estimating low-velocity layers in the near-surface. In such cases, waveform inversion alone can not converge to the correct model while the proposed workflow is able to do so. Results of this chapter have been submitted to Saudi Aramco for publication approval in Geophysics (Shen and Luo, 2015)

Chapter 3: Kinematic-Based Inversion Objective Function-I introduce the kinematic objective function for waveform inversion. The new objective function emphasizes phase matching over amplitude matching. It is similar to exponential phase comparison or maximization of zero-time cross-correlation between observed early-arrivals and modeled early-arrivals. Such emphasis of phase comparison does not reduce the inversion resolution for realistically complex models compared with that from using the conventional objective function. More importantly, the new objective function is robust in acoustic inversion of elastic early-arrivals. On the other hand, the amplitude matching in the conventional objective function makes it susceptible to failures in acoustic inversions of elastic early-arrivals.

Chapter 4: Random boundary condition for efficient gradient calculation-I demonstrate a modified random boundary condition that totally eliminates the I/O cost of saving wavefields during gradient calculations. The original random boundary condition was designed for reverse time migration applications. The adaptation for high-frequency data in reverse time migration applications makes the original boundary condition ineffective for the low-frequency wave propagation used in waveform inversion. I proposed to modify the spatial randomness of the random boundary to make it effective for waveform inversion applications, while keep the boundary region the same size. Waveform inversion gradients calculated with the new random boundary condition is virtually the same as the ones calculated with the conventional method. Yet calculating the new gradients does not incur I/O cost of saving wavefields that is associated with the conventional method.

Chapter 5: 3D data examples-I demonstrated the effectiveness of the methodologies developed in the previous chapter using an inversion example of a 3D synthetic dataset and a 3D Brazilian land data example.