From tomography to full-waveform inversion with a single objective function

Tariq Alkhalifah¹ and Yunseok Choi¹

ABSTRACT

In full-waveform inversion (FWI), a gradient-based update of the velocity model requires an initial velocity that produces synthetic data that are within a half-cycle, everywhere, from the field data. Such initial velocity models are usually extracted from migration velocity analysis or traveltime tomography, among other means, and are not guaranteed to adhere to the FWI requirements for an initial velocity model. As such, we evaluated an objective function based on the misfit in the instantaneous traveltime between the observed and modeled data. This phase-based attribute of the wavefield, along with its phase unwrapping characteristics, provided a frequency-dependent traveltime function that was easy to use and quantify, especially compared to conventional phase representation. With a strong Laplace damping of the modeled, potentially low-frequency, data along the time axis, this attribute admitted a first-arrival traveltime that could be compared with picked ones from the observed data, such as in wave equation tomography (WET). As we relax the damping on the synthetic and observed data, the objective function measures the misfit in the phase, however unwrapped. It, thus, provided a single objective function for a natural transition from WET to FWI. A Marmousi example demonstrated the effectiveness of the approach.

INTRODUCTION

In spite of the many challenges we face in implementing full-waveform inversion (FWI), the high nonlinearity of the objective function, resulting from the sinusoidal nature of seismic wavefields and from the complexity of the earth reflectivity, is our biggest challenge. The quest of finding an initial velocity model that produces synthetic data resulting in a misfit located within the basin of attraction of the global minima of the objective function has rendered the conventional gradient-based waveform inversion convergence conditional. This condition is usually met through unnatural extensions of the model wavelength features from traveltime tomography to FWI (Ellefsen, 2009). Clearly, the size of this basin of attraction and the length of the half-cycle increase at low frequencies and large offsets. However, we rarely acquire data with low enough frequencies or large enough offsets that allow us to start from a crudely calculated initial velocity model.

Phase inversion, that is, considering only the phase of the wavefield in the cost function, has been advocated recently (Kim and Shin, 2005; Shin and Min, 2006; Bednar et al., 2007). Nevertheless, most of the phase inversion implementations are based on extracting the phase without eliminating the wrapping phenomena, and thus, do not contribute much in reducing the nonlinearity (cyclical nature of wavefields) of the inversion. Bednar et al. (2007) attributed this preference to the possibility of losing information in the unwrapping operation. Gelis et al. (2007) cite the importance of unwrapping the phase for their Rytov-based update of an elastic inversion, but they do not implement it in their work. Grzegorczyk et al. (2011), Choi and Alkhalifah (2011), and Alkhalifah and Choi (2012) unwrap such phases so that high-resolution results are maintained without the penalty of introducing unbearable nonlinearity.

Consequently, we are promoting the use of the instantaneous traveltime to measure the misfit between the observed and modeled data as a single objective that allows us to avoid unnatural transition from traveltime tomography to FWI. This objective function measures the misfit in the phase, however, unwrapped, thus eliminating one source of nonlinearity in the objective function that is produced by the cyclical nature of the seismic source and wavefield. In this study in particular, we focus on the instantaneous traveltime (IT) objective function flexibility to take us from traveltime tomography to FWI. We also highlight some of the distinct features that IT brings to the implementation of FWI. We justify our assertions with an implementation of FWI on the ever-versatile and familiar Marmousi model.
THE INSTANTANEOUS TRAVELTIME

An explicit way to unwrap the phase of a signal is suggested by Stoffa et al. (1974), and we write it in terms of traveltime (we refer to it as the IT) instead of phase, as follows:

$$\tau(\omega) = \Im \frac{du(\omega)}{d\omega},$$

(1)

where $u$ is the wavefield in the frequency domain, $\omega$ is the angular frequency, and $\Im$ stands for the imaginary part. We refer to this attribute as the IT because it has a similar formulation to the instantaneous frequency, with for the change of domain and derivative variable (frequency instead of time). Considering that $u(\omega) = A(\omega)e^{i\phi(\omega)}$, then

$$\tau(\omega) = \frac{d\phi}{d\omega} = \frac{\Re u(\omega) \frac{dA(\omega)}{d\omega} - \Im u(\omega) \frac{dA(\omega)}{d\omega}}{(\Re u(\omega))^2 + (\Im u(\omega))^2}. \tag{2}$$

To unwrap the phase, Stoffa et al. (1974) integrate equation 2 over $\omega$, as they use the unwrapped phase for homomorphic deconvolution. A small positive number in the denominator of the right-hand side term in equation 2 could help us avoid a division over zero. As Choi and Alkhalifah (2011) demonstrate, we can easily model $du/d\omega$, needed to evaluate $\tau(\omega)$, using the same extrapolation operator based on the background model used for $u$, but with a virtual source function (see Appendix A). In our case, these calculations are done by solving the wave equation in the frequency domain. Because seismic data are conventionally acquired on the earth surface, their resolving power deteriorates with depth. In addition, reflections, which often arrive at later times compared to refractions, pose a serious nonlinearity problem to FWI. As a result, Shin et al. (2002) introduce Laplace damping in the inversion process to enhance the shallow parts of the data in FWI dominated by reflections. Considering the damping factor $\alpha$, the damped data have the following form: $u_d(t) = u(t)e^{-\alpha t}$, where $u(t)$ is the wavefield in the time domain. Other options for damping later arrivals are available, but the Laplace damping formulation admits a Laplace domain wave equation synonymous to the Helmholtz formulation, and thus, easy to solve (Shin et al., 2002).

THE INFLUENCE OF DAMPING

The attribute of the wavefield given by $\tau$, defined by equation 1, has traveltime units, and for a single event along the trace, it provides the traveltime of the energy for each frequency component included in the wavelet making up the event. However, for multiple events, it provides the average traveltime weighted by the amplitude of each event (Saragiotis et al., 2011). Thus, damping the later arrivals focuses the energy up shallow resulting in smaller IT values, and with strong damping and no noise prior to the first arrival (synthetic data), it provides the traveltime of the first arrival. Figure 1 shows a common shot gather from data 8 of Yilmaz (2001). The arrows point to the two traces from locations 1600 m (solid) and 2800 m (dotted) displayed together in Figure 2a. The spectrum for these two traces are shown in Figure 2b, and the IT (which depends only on the phase component) is shown in Figure 2c. Average traveltimes of 0.5 s for the near-offset trace and 1 s for the far-offset trace (Figure 2c) reflect the average locations of the energy of the two traces. Nevertheless, this traveltime function varies with frequency, and this complex variation holds much of the information necessary to resolve a high-resolution subsurface model using FWI (the phase part). However, finding a model capable of producing reflections giving us synthetic data that has a similar complex phase function is usually an unattainable objective of FWI. Applying a damping factor of 5 ($\alpha = 5$) to the data attenuates the later arrival as shown in Figure 2d. The amplitude spectra (Figure 2e) are much smoother as the complexity in the data has been reduced, and the instantaneous traveltimes (Figure 2f) show, now, distinct average traveltimes for the two traces reflecting more the onset of energy to these geophones. There is still plenty of variation with frequency, but generally less than without damping (Figure 2c). Intuitively, it would be much easier to find a model that produces synthetic data with similar IT responses. Further damping ($\alpha = 30$) enhances the very first arrivals (that may be noise), as shown in Figure 2g. The resultant amplitude spectra are even smoother (Figure 2h). For the near-offset trace, as Figure 2i shows, this damping provides an average traveltime (about 0.12 s) close to first-arrival pick. For the other trace, the strong damping enhanced the near zero-time noise present in this field data, and thus, it provided erroneous traveltimes, corresponding to the energy of the noise. A similar limitation, but with worse consequences, is experienced with the conventional misfit using the traces directly (Figure 2g) because such noise, considering its enhanced amplitude, will result in a large weight for such arrivals in the objective function and the inversion. In both cases, we need to mute such noise as Shin et al. (2002), among others, suggest. For modeled noise-free data this is not an issue, and thus ITs can provide first-arrival picks that can be compared with picked arrivals from field data in a wave equation traveltime tomography (WET) implementation. Thus, the IT representation has the flexibility to serve as a phase inversion (with unwrapped features) as well as, with the right damping, in a WET tomography. However, compared to the classic WET (Luo and Schuster, 1991; Woodward, 1992), with a misfit based on the lag in the crosscorrelation between the observed and modeled data, the IT yields finite frequency traveltime misfit measurements that are more inline with FWI phase characteristics (Dejbbi and Alkhalifah, 2013).

Figure 1. A real seismic shot gather with arrows pointing to the traces that will be used in analyzing the IT. On the vertical axis, the first-arrival traveltime for the two arrowed traces are shown.
WHY DAMP IT?

Laplace damping of the wavefield as a function of time, as Shin et al. (2002) suggest, helps reduce the incredible nonlinearity often accompanying the objective function of FWI, specifically from reflections. Enough damping will transform a trace to an effective $\delta$-function corresponding to the first arrival. This ideal scenario faces many obstacles. As demonstrated in Figure 2e as we compare it with Figure 2b, the amplitude of the near- and far-offset traces can vary considerably due to damping. This will result, unintentionally, in giving the near offset considerably more weight in FWI than the far offsets. A remedy to this amplitude variation problem is suggested by Shin and Min (2006) through using the logarithmic wavefield in the objective function. With logarithmic wavefields, the Laplace damping transforms to a linear function, which is far more bearable. With the IT objective function, because the attribute is traveltime, the damping has little influence on weighting the objective function. Another prevailing issue with damping wavefields for FWI is the early arrival noise (prior to the first arrival) usually present in real data. The damping function will enhance such noise to a point that it negatively effects FWI (Figure 2g). In FWI, it is hard to impose a conditional statement that allows us to remove such enhanced noise from the inversion. In IT-based inversion, as we see in Figure 2i, the traveltime corresponding to the noise will have small values that are easily identifiable.

The calculation of the wavefield and the derivative wavefield needed by equation 1 are performed in the frequency domain. Both wavefields rely on the same Helmholtz solver, but different source functions as demonstrated by Choi and Alkhalifah (2011) and in Appendix A. Despite the frequency-domain implementation, the calculated attribute given by equation 1 is a traveltime variable describing the average location of the energy in the time domain. Thus, damping has a direct effective influence on this attribute by lowering, as we see in Figure 2f and 2i its value for observed and

Figure 2. The two real seismic traces corresponding to the location of arrows in Figure 1 (solid curve = 1600 m, dotted curve = 2800 m) with (a) no damping, (d) $\alpha = 5$, and (g) $\alpha = 30$, and the corresponding amplitude spectrum (b, e, h) and IT (c, f, i), respectively.
modeled data. It also smooths this attribute as a function of frequency. Despite the fact that strong damping can mitigate the nonlinearity, and specifically the wrapping phenomena, for conventional objective functions, for the IT, it actually produces an attribute reflecting the first-arrival traveltime in modeled data that can be compared with picked first arrivals from the observed data.

THE OBJECTIVE FUNCTION

FWI is based on minimizing the difference (the distance based on a norm defined in the Hilbert space) between the observed data (or an attribute of it) and the computationally modeled data (or an attribute of it). We consider the attribute given by equation 2, and thus, we minimize

$$\min_m \| \tau[L(m)] - \tau[d] \|_2,$$

where $m$ corresponds to the model, and $L$ is the modeling operator corresponding to any one of many wave equations and applied to $m$ to generate synthetic data to compare with the observed data $d$. The operator $\tau$ is the implementation of equation 2 (here, $\| \cdot \|_2$ is the $L_2$ norm). The modeling operator that allows us to directly compute $\tau$, and specifically the wavefield and its frequency derivative, is the Helmholtz solver. A strong damping applied to the noise-free, modeled data, by solving the Helmholtz wave equation, can yield an accurate first-arrival traveltime that we can compare with picked ones by replacing $\tau[d]$ with picked first arrivals from the observed data. Thus, with proper Laplace damping, this single objective function can have quasilinear behavior and serve as a WET. It also has the flexibility, as we relax the damping, to provide FWI resolution. In both cases, low or high damping, this objective function has superior qualities to classical WET and FWI. As compared to FWI (or phase inversion), the IT unwraps the phase, and thus, it eliminates the nonlinearity produced by the source. As compared to conventional WET, its sensitivity kernels are free of the “hollow-banana” effect that can hamper proper convergence (Dejbbi and Alkhalifah, 2013).

As Appendix A illustrates, the gradient for this objective function, whether we are using a strong damping and comparing it with picked first arrivals or not, is evaluated using the adjoint state method. In addition to back propagating the instantaneous traveltime residuals, we have to compute the derivative wavefield sensitivity to the model parameters. However, all three state variables are solved using the same operator, which adds little to the cost of the gradient calculation. Most of the cost in solving the Helmholtz equation is in inverting the wave equation operator.

THE OBJECTIVE FUNCTION IN ACTION

Now we test the ability of this single objective function to cope with the spectrum of wavelengths in our model and its influence on the data. As a result, we invert for the infamous Marmousi model (shown in Figure 3) starting from a constant velocity model. Using a finite-difference method applied to the wave equation, we generate data with a frequency band ranging between 2 and 5 Hz. Receivers are considered along the whole surface of the model at a 20-m depth with a spacing interval of 20 m, which coincides with the grid spacing of our model at 20 m in both directions. The shot spacing is 40 m along the whole line. This data set acts as our observed data for the FWI. A shot gather from the resulting modeled data set is shown in Figure 4. The minimum frequency in these observed data is 2 Hz. The Laplace damping will reduce the effective minimum frequency in the observed data with its spectrum whitening implications as is noted by Ha and Shin (2012). The initial velocity model for the inversion is given by a constant velocity of 3 km/s, as shown in Figure 5a. With very strong damping ($\alpha = 40$) and a frequency of 0.125 Hz (in our modeled data only), the instantaneous traveltime admits first-arrival traveltimes that are compared with picked first arrivals from the data, and thus, act like traveltome tomography providing us with the model shown in Figure 5b. As we relax the damping factor to 10 and allow frequencies up to 4 Hz, we obtain, after 100 iterations, the model in Figure 5c. Additional relaxation of the damping provides the model shown in Figure 5d. Finally, with very small damping ($\alpha = 0.5$) and moving up the modeled data frequency range to the data range (2–5 Hz), we end up after 100 iterations with the model shown in Figure 6a. The high-resolution velocity model, considering that the highest frequency in the data is 5 Hz, is reasonably accurate. In all the iterations, we use a single objective function given by equation 3 to take us from WET to FWI.

If we start a conventional FWI (Tarantola, 2004) with the advanced velocity model shown in Figure 5d, after 100 iterations we obtain the velocity model shown in Figure 6b. The difference between Figure 6a and 6b is overall small, though the FWI result seems to have more resolution considering the inversion uses the amplitude information. Some of this higher resolution information do not look accurate, especially up shallow. Depth profiles of the inverted velocity models as they compare with the true Marmousi

Figure 3 for a source located at the surface at location 4 km.

Figure 4. A modeled shot gather for the Marmousi model shown in Figure 3 for a source located at the surface at location 4 km.
model are shown in Figure 7a for the lateral location 4 km and in Figure 7b for the lateral location 8 km. Considering that we limit our inversion to a maximum frequency of 5 Hz, the inverted models are smooth versions of the true one. Results from conventional FWI and the IT show reasonable agreement up shallow, but the IT inversion seems to provide higher velocities and better agreement with the true model at depth. This reflects the phase nature of the IT inversion, which is less sensitive to the weak amplitudes of later arrivals.

DISCUSSION

The IT objective function provides a platform to implement WET tomography with traveltime picks from the observed data as well as phase inversion with the phase of the data. In both implementations, we have the flexibility to use any frequency in the data. Thus, we can use a single objective function to move us from a smooth velocity model to a high-resolution model as we relax the damping and increase the frequencies used in the inversion. However, there are no guarantees that this continuity in the objective function, and in the update kernel, translates to a convergence to a global minima. This objective function, despite its many features over the conventional objective function, does not fill the wavelength gap that we may encounter in the inversion between the initial long wavelength updates and the later short-wavelength updates. The size and nature of this gap are dependent on the frequencies present in the data and the maximum offsets available (Sirgue and Pratt, 2004).

There is, however, another factor to be considered. Damping enhances the low-frequency content by its attempt to whiten the spectrum. Early recorded noise might prevent that, but recently, many options have been suggested to tackle such noise. This potential enhancement of low frequencies allows us to approximately fill the model wavelength gap.

For the Marmousi example the results from the IT inversion showed better agreement with the true model than the results obtained from conventional FWI, especially at depth (Figure 7a and 7b). This is also reflected in the data misfit for the inverted models obtained from the two approaches. Figure 8 shows the normalized root-mean-square fit of the synthetic data generated using the inverted models from conventional FWI (black) and those from the IT inversion (red) at each iteration. Clearly, for the IT inversion, the data fit is overall better. In some cases, the data fit for the IT inversion seemingly deteriorates with iterations because the objective function is not based on the data fit, but on an IT (phase) fit. Nevertheless, such deteriorations are limited.

The full spectrum

Figure 5. (a) The initial constant velocity model, (b) using $\alpha = 40$ and a frequency of 0.125 Hz (equivalent to traveltime tomography), (c) relaxing the damping factor, $\alpha = 10$ and using a frequency range of 0.125–4 Hz and, (d) relaxing the damping even further to 2.5 and using the data frequency range of 0.125–4 Hz.

Figure 6. (a) The final result from the IT based inversion and (b) the final result from conventional FWI starting in the case of FWI with the model given in Figure 5d.
CONCLUSIONS

The IT of the modeled and potentially recorded data provides us with an objective that is flexible enough to provide us with travel-time tomography convergence features to smooth models and FWI resolution features to obtain complex models. The damping of the modeled and recorded data acts as the weighting factor in using these two inversion extremes, and it allows for a natural transition. This has been demonstrated using the Marmousi model.

ACKNOWLEDGMENTS

We thank KAUST for its support. We also thank our colleagues at SWAG for the many useful and fruitful discussions. We thank S. Kaplan and three anonymous reviewers for their valuable suggestions and comments.

APPENDIX A

THE GRADIENT

From equations 1 and 3, the objective function for the IT is written in the following expanded form:

\[ E = \sum_{j=1}^{n} \left[ \Im \left\{ \left( \frac{\partial \hat{u}_j}{\partial \omega} \right) / \hat{u}_j \right\} - \Im \left\{ \left( \frac{\partial \hat{d}_j}{\partial \omega} \right) / \hat{d}_j \right\} \right]^2, \quad (A-1) \]

where \( n \) is the number of recordings and \( u_j \) and \( d_j \) are the modeled and recorded data, respectively. The gradient is obtained by taking the derivative of this equation with respect to the \( i \)th model parameter \( m_i \) and expressed as

\[ \frac{\partial E}{\partial m_i} = 2 \sum_{j=1}^{n} \Im \left[ \frac{\partial^2 \hat{u}_j}{\partial m_i \partial \omega} \left( \frac{\partial \hat{u}_j}{\partial \omega} \right) / \hat{u}_j - \left( \frac{\partial \hat{d}_j}{\partial \omega} \right) / \hat{d}_j \right] \]

\[ + 2 \sum_{j=1}^{n} \Im \left[ - \frac{\partial \hat{u}_j}{\partial m_i} \frac{\partial \hat{u}_j}{\partial \omega} \left( \frac{\partial \hat{u}_j}{\partial \omega} \right) / \hat{u}_j - \left( \frac{\partial \hat{d}_j}{\partial \omega} \right) / \hat{d}_j \right]. \quad (A-2) \]

We can then express the gradient of the objective function in matrix form based on the adjoint-state technique of the wave equation as follows:

\[ \frac{\partial E}{\partial m_i} = 2 \Im \left[ \left( - \frac{\partial \hat{u}^T \hat{u}}{\partial m_i \partial \omega} - \frac{\partial^2 \hat{u}^T \hat{u}}{\partial m_i \partial \omega} \right) \S^{-1} \hat{r}_1 \right] \]

\[ + \left( - \frac{\partial \hat{u}^T \hat{u}}{\partial m_i} \right) \S^{-1} \left\{ \left( - \frac{\partial \hat{d}^T \hat{d}}{\partial \omega} \right) \S^{-1} \hat{r}_1 \right\} \]

\[ - 2 \Im \left[ - \frac{\partial \hat{u}^T \hat{u}}{\partial m_i} \S^{-1} \hat{r}_2 \right], \quad (A-3) \]

where the elements of \( \hat{r}_1 \) are given by

\[ r_{1i} = \frac{1}{u_j} \Im \left[ \left( \frac{\partial \hat{u}_j}{\partial \omega} \right) / \hat{u}_j - \left( \frac{\partial \hat{d}_j}{\partial \omega} \right) / \hat{d}_j \right] \quad (A-4) \]

and the elements of \( \hat{r}_2 \) are given by

\[ \hat{r}_{2i} = \frac{1}{u_j} \Im \left[ \left( \frac{\partial \hat{u}_j}{\partial \omega} \right) / \hat{u}_j - \left( \frac{\partial \hat{d}_j}{\partial \omega} \right) / \hat{d}_j \right] \]
\[ r_{2i} = \left( \frac{\partial \hat{u}_i}{\partial \omega} \right) \cdot \text{Im} \left[ \left( \frac{\partial \hat{u}_j}{\partial \omega} \right) \hat{u}_j - \left( \frac{\partial \hat{d}_j}{\partial \omega} \right) \hat{d}_j \right]. \quad (A-5) \]

To calculate the gradient, we back propagate \( r_1, r_2 \), and \( \{(-\partial S/\partial \omega)^T S^{-1} r_1 \} \), and multiply the back propagated wavefield with virtual wavefields \([-\partial S/\partial m_i) (\partial \hat{u}/\partial \omega) - (\partial S/\partial m_i) (\partial \hat{d}/\partial \omega)]^T \) and \([-\partial S/\partial m_i) (\partial \hat{u}/\partial \omega) - (\partial S/\partial m_i) (\partial \hat{d}/\partial \omega)]^T \), respectively.

We update along the steepest descent direction, which is evaluated using the conjugate-gradient approach (Gill et al., 1981). To obtain a proper step length, we can use a line-search method, but, for simplicity, we fix the value of the step length to be small in a finite difference (stepwise)-like implementation. At low frequencies, this might not be optimal, but at high frequencies, where the Hessian influence and accuracy are reduced, this small fixed step length might provide an efficiency factor. However, the update approach characteristics are beyond the scope of this paper.

REFERENCES


