Frequency-domain waveform inversion using the phase derivative

Yunseok Choi and Tariq Alkhalifah

Physical Science and Engineering Division, King Abdullah University of Science and Technology, 4700 Thuwal, Saudi Arabia.
E-mail: yunseok.choi@kaust.edu.sa; tariq.alkhalifah@kaust.edu.sa

Accepted 2013 September 2. Received 2013 September 2; in original form 2012 September 2

SUMMARY
Phase wrapping in the frequency domain or cycle skipping in the time domain is the major cause of the local minima problem in the waveform inversion when the starting model is far from the true model. Since the phase derivative does not suffer from the wrapping effect, its inversion has the potential of providing a robust and reliable inversion result. We propose a new waveform inversion algorithm using the phase derivative in the frequency domain along with the exponential damping term to attenuate reflections. We estimate the phase derivative, or what we refer to as the instantaneous traveltimes, by taking the derivative of the Fourier-transformed wavefield with respect to the angular frequency, dividing it by the wavefield itself and taking the imaginary part. The objective function is constructed using the phase derivative and the gradient of the objective function is computed using the back-propagation algorithm. Numerical examples show that our inversion algorithm with a strong damping generates a tomographic result even for a high ‘single’ frequency, which can be a good initial model for full waveform inversion and migration.

Key words: Inverse theory; Theoretical seismology; Wave propagation.

INTRODUCTION
Waveform inversion algorithms have been developed to generate a high-resolution and reliable subsurface velocity model. Nevertheless, waveform inversion suffers from an inherent limitation resulting from the oscillatory nature of seismic data. This limitation is somewhat alleviated using low frequencies necessary to reduce the complexity of the objective function for the waveform inversion (Bunks et al. 1995; Zhou et al. 1995; Sirgue & Pratt 2004). However, the available frequency content of seismic data is not usually low enough.

When the starting model is far from the true model, the wrapping around nature of the phase in the frequency domain or the cycle skipping that takes place in the time domain is a major reason for the local minima problem in waveform inversion (Bunks et al. 1995; Virieux & Operto 2009). The phase of the Fourier-transformed wavefield, or what is referred to as the principal value of phase (PVP), has a finite but periodic range \([-\pi, \pi]\), thus the phase of the modelled data could converge to the wrong phase corresponding to a different cycle in the procedure of waveform inversion, which makes the inversion converge to a local minimum. On the other hand, very low frequencies or a close enough starting velocity model to the true one allow us to approach the accurate phase in waveform inversion. In general, however, the frequency band of seismic data is not always low enough and even if it is, the low frequencies are often contaminated with noise. Also, obtaining a good starting velocity model is not a trivial task as it requires using elaborate approaches like traveltime tomography and traveltime picking. Therefore, there is a need to develop an inversion algorithm using a phase attribute free from the wrapping phenomena.

Phase unwrapping for the 1-D case has been tackled in the area of digital signal processing, especially for the computation of the complex cepstrum (Oppenheim & Schafer 1975). The simplest method for the phase unwrapping is to detect the discontinuities (2\pi jumps) in the phase spectrum of the signal, determine the integer multiples of 2\pi according to the discontinuities at each frequency and add it to the principal values of the phases at each frequency (Oppenheim & Schafer 1975). Another method is to use the derivative of phase with respect to the angular frequency (DPAF), the integration of which over frequency gives the unwrapped phase (Oppenheim & Schafer 1975; Tribollet 1977). More evolved methods factorize either the polynomials of the z-transform or the real component of it and use the zeros (roots) of the factorized function to unwrap the phase (McGowan & Kuc 1982; Steiglitz & Dickinson 1982; Treitel et al. 2006). On the other hand, Al-Nashi (1989) developed some measures to detect and count the zeros close to the unit circle (sharp zeros) and unwrapped phases according to the detected sharp zeros. All these methods deal with the 1-D phase of the frequency variable (phases in the frequency coordinate).

On the other hand, the concern of frequency-domain waveform inversion is dealing with the phase values of the offset variable (phases in the offset coordinate) rather than the frequency variable. In the frequency-domain waveform inversion, the modelled wavefield is generated in the spatial coordinate for a single frequency and the gradient of the objective function is calculated selectively for chosen frequencies, thus reducing the wavenumber redundancy.
Another issue with the phase attribute in waveform inversion is its calculation for the objective function. Since waveform inversion minimizes the differences between the observed and modelled data to define the subsurface velocity model, the calculation of phase attribute for both the observed and modelled data should be considered. Since the modelled data set is regenerated every iteration from the updated velocity model, the calculation of the phase attribute from the modelled data also must be repeated in every iteration, whereas a single calculation is enough for the observed data through whole inversion process. Therefore, a simple and fast algorithm to estimate the phase attribute from the modelled data is required in the waveform inversion procedure.

Shah et al. (2010) used the unwrapped phases of the observed and starting modelled data in the source–receiver coordinate to generate an intermediate model, which is used as a new starting model for the next waveform inversion stage. They excluded the part of data in the unwrapping process where PVP cannot be unambiguously unwrapped. However, if the available frequency is not quite low and noise is included in the data, it is very difficult to unambiguously unwrap PVP in major part of data.

For waveform inversion, DPAF can be a good substitute for PVP. PVP has a bound of range, thus suffer from the wrapping phenomena, whereas DPAF has no limitation in range, thus has no wrapping problem. Shin et al. (2003) suggested to calculate the first-arrival traveltime on a given velocity model using DPAF of the modelled wavefield with a high-exponential damping factor. A high-exponential damping factor suppresses the events following first-arrival traveltime on a given velocity model using DPAF of Brossier form inversion (Brenders & Pratt 2007; Shin & Cha 2008, 2009; etc. (Van Leeuwen & Mulder 2010; Djebbi & Alkhalifah 2013). Other difference with the refraction tomography is that we work with the wavefield for several damping factors and frequencies, separately or simultaneously, which results in a natural extension to waveform inversion.

In the following section, we start by analysing PVP and DPAF for different damping factors and frequencies. Next, we present an inversion algorithm using DPAF and show numerical examples for synthetic data.

**THEORY**

In waveform inversion, the difference between the modelled and observed wavefield is minimized to update the subsurface velocity model. Thus, the issue of phase must be considered in both the modelled and observed wavefield. Since the modelled wavefield is regenerated in an iterative fashion, the calculation of the phase attribute of the modelled wavefield is a bigger burden than that for the observed wavefield. Therefore, we focus on the examination of PVP and DPAF estimated from the modelled wavefield. The modelled wavefield using the frequency-domain modelling provides us with the Green’s function, and thus, noise is not an issue for the modelled wavefield. On the other hand, noise must be considered for the observed wavefield, which is discussed in the latter section. Finally, we describe the gradient calculation using the back-propagation algorithm (Pratt et al. 1998) for the inversion of DPAF.

**Exponential-damped wavefield**

The exponential-damped wavefield \( \tilde{u}(t) \) is expressed as a multiplication of the original wavefield \( u(t) \) and exponential damping \( e^{-\alpha t} \):

\[
\tilde{u}(t) = u(t)e^{-\alpha t},
\]

where \( u(t) \) and \( \tilde{u}(t) \) are casual functions and \( \alpha \) is a damping factor. Fourier transform of the damped wavefield is written as

\[
\tilde{u}(\omega) = \int_{0}^{\infty} \tilde{u}(t)e^{i\omega t} \, dt = \int_{0}^{\infty} u(t)e^{i(\omega+i\alpha)t} \, dt,
\]

where \( \omega \) is an angular frequency. In the last term of eq. (2), the damping factor \( \alpha \) is located in the imaginary part of the complex angular frequency. This means that the frequency-domain modelling with a complex angular frequency generates an exponential-damped wavefield, where the imaginary part of the complex angular frequency plays a role of exponential damping.

Fourier transform of the damped wavefield in eq. (2) is represented as a Laplace transform or z-transform of the original wavefield. For a fixed \( \alpha > 0 \), Fourier transform of the damped wavefield lies on a vertical line on the s-plane of Laplace transform (the distance between the vertical line and the vertical axis is \( \alpha \)) and on a circle line on the z-plane of the z-transform (the radius is \( e^{\alpha} \)). The unwrapping procedure for the original wavefield depends on the sharp zeros (zeros near the unit circle) on the z-plane (Al-Nashri 1989). If \( \alpha \) is large enough, the zeros near the circle of \( e^{\alpha} \) radius can be avoided.
If the damping factor $\alpha$ is very high, we can suppress all the events following the first arrival and thus approximate the damped wavefield as

$$\bar{u}(t) \cong \xi e^{-\alpha t} \delta(t - \tau),$$

(3)

where $\xi$ and $\tau$ are the amplitude and traveltime of the first-arrival event, respectively. Fourier transform of eq. (3) is written as

$$\hat{u}(\omega) \cong \xi e^{i(\omega + i\alpha \tau)}.$$

(4)

In eq. (4), the phase is a multiplication of the first-arrival traveltime and the complex angular frequency. Therefore, if we use a high-damping factor, we can estimate the first-arrival traveltime from the phase of the damped wavefield.

Fourier transform with a high-damping factor indicates a circle with a very large radius ($e^{\alpha \omega}$) on the $z$-plane, which means no sharp zeros are near the circle and only a linear component of the phase remains. The linear component of phase, however, still causes the phase-wrapping problem in waveform inversion. On the other hand, as the damping factor (or the radius of the circle) decreases, the density of the sharp zeros near the circle increases and thus the frequency of the phase wrapping increases.

**Wrapped phase estimated from the modelled wavefield**

In waveform inversion, we minimize the difference between the phase of the modelled wavefield and the phase of the observed wavefield. In this section, we describe some of the properties of the wrapped phase (or PVP) of the modelled wavefield.

The frequency-domain modelling is expressed as a matrix form (Marfurt 1984; Pratt et al. 1998):

$$\hat{S}\hat{u} = \hat{f},$$

(5)

where $\hat{S}$ is a symmetric complex matrix incorporating the damping factor (including the complex angular frequency), $\hat{u}$ is a damped wavefield vector. The matrix $\hat{S}$ includes model parameters, such as subsurface velocities. In this paper, we assume the source wavelet is known and $\hat{f}$ is just a shot-positioning vector. The modelled wavefield $\hat{u}$ for a single frequency can be obtained by factorizing the matrix $\hat{S}$ where a direct solver is usually used. The modelled wavefield obtained by a frequency-domain modelling technique has the complex value:

$$\hat{u}_j = |\hat{u}_j| e^{i\arg(\hat{u}_j)},$$

(6)

where $\arg(\hat{u}_j)$ is the phase of $\hat{u}_j$, and the subscript $j$ means the $j$th component (or $j$th offset position) of a vector. The complex logarithm of $\hat{u}_j$ and its inverse are defined (Oppenheim & Schafer 1975) as:

$$\log \hat{u}_j = \log |\hat{u}_j| + i\arg(\hat{u}_j)$$

$$e^{\log \hat{u}_j} = |\hat{u}_j| \cdot e^{i\arg(\hat{u}_j)}.$$  

(7)

The imaginary part of the complex logarithm of $\hat{u}_j$ is the wrapped phase or referred to as the PVP:

$$\arg(\hat{u}_j) = \text{Im} (\log \hat{u}_j),$$

(8)

where ‘Im’ stands for the imaginary part. PVP in eq. (8) always exists between $-\pi$ and $\pi$ and has an ambiguity of a multiple of $2\pi$ associated with $\arg(\hat{u}_j)$. Clearly, any integer multiple of $2\pi$ can be added to PVP without changing the results of the inverse of the complex logarithm in eq. (7).

We generate the modelled wavefield for the Marmousi velocity model shown in Fig. 1 and plot PVP estimated from the modelled wavefield with offset and frequency variables. We place a shot at a distance of 2 km and receivers with an offset range of 0 $\sim$ 6 km to the right side of the shot (distance range of 2 $\sim$ 8 km) with a receiver interval of 20 m. The shot and receiver are placed at a depth of 20 m. We perform the frequency-domain modelling for each frequency ranging from 0.01 to 6 Hz with a frequency interval of 0.01 Hz, thus a total of 600 modelling steps. Fig. 2 shows PVP
maps as a function of offset and frequency for different damping factors. PVP map for a low-damping factor (Fig. 2a) shows the complicated behaviour of phases and the wrapping phenomena. The high-damping factor, on the other hand, yields a simple phase behaviour with wrappings (Fig. 2b). Fig. 3 shows the profiles of each PVP map in Fig. 2 at an offset of 1 km. For the very low damping factor (0.0001 s$^{-1}$), PVP suffers from wrapping effects related to the sharp zeros (Fig. 3a). The wrapping related to the sharp zeros can be recovered using several methods, such as integrating DPAF, factorizing polynomials and so on (Karam 2006). For the high-damping factor (Fig. 3b), PVP shows only simple shapes with wrappings.

On the other hand, our concern in the frequency-domain waveform inversion is dealing with data as a function of offset for a single frequency. In the frequency-domain waveform inversion, a few frequencies are selected to reduce the wavenumber redundancy (Sirgue & Pratt 2004) and the gradient for inversion is calculated independently for each frequency. In Fig. 4, we plot the profiles of each PVP map in Fig. 2 at the frequencies of 0.5–5 Hz. For the very low damping factor, PVP shows a simple behaviour with wrapping effects at the low frequency (Fig. 4a) but complex features with more complicated wrappings at the high frequency (Fig. 4b). PVP for a high-damping factor, however, shows only simple behaviour with different cycle of wrappings according to the frequency (Figs 4c and d). Since the high-damping factor suppresses all events following the first arrival, PVP in Figs 4(c) and (d) represents the first-arrival traveltime multiplied by the angular frequency, as mentioned in eq. (4), with the band including a wrapping effect. Detecting the discontinuities of PVP could un-wrap the phase of the linear component. The detection method, however, cannot be incorporated to the back-propagation algorithm for the calculation of the gradient. In this paper, as an alternative, we propose to use DPAF for the inversion.

Phase derivative estimated from the modelled wavefield

DPAF has been used to estimate the unwrapped phase by integrating it (Oppenheim & Schafer 1975; Tribolet 1977). We, however, propose to invert DPAF itself without the integration to define the subsurface structure.

Figure 3. The profiles of the PVP map in Fig. 2 at an offset of 1 km for damping factors: (a) 0.00001 s$^{-1}$ and (b) 40 s$^{-1}$.

Figure 4. The profiles of PVP map in Fig. 2 at frequencies of (a, c) 0.5 Hz and (b, d) 5 Hz for damping factors: (a, b) 0.00001 s$^{-1}$ and (c, d) 40 s$^{-1}$.
DPAF is defined as the imaginary part of the ratio of the derivative wavefield to the wavefield itself (Oppenheim & Schafer 1975):

$$\text{DPAF}(\omega) = \text{Im} \left( \frac{\partial \hat{u}_j(\omega)}{\partial \omega} \right).$$

For a high-damping factor, DPAF is just the first-arrival traveltime, which is proved by taking the derivative of the damped wavefield in eq. (4) with respect to the angular frequency:

$$\frac{\partial \hat{u}_j(\omega)}{\partial \omega} = i \tau_j \xi_j e^{i(\omega + i\alpha)\tau_j} = i \tau_j \hat{u}_j(\omega).$$

In eqs (9) and (10), DPAF (the traveltime $\tau_j$ in the amplitude part in eq. 10) has no bound of range, thus having no wrappings. To calculate the wavefield derivative ($\partial \hat{u}_j(\omega)/\partial \omega$) for DPAF, Shin et al. (2003) suggested the finite-difference scheme. In this paper, we develop a new algorithm to calculate the wavefield derivative by employing the frequency-domain modelling technique using a virtual source term.

Since the vector $\hat{f}$ is not a function of $\omega$, differentiating eq. (5) with respect to $\omega$ gives

$$\hat{S} \frac{\partial \hat{u}_j}{\partial \omega} = \frac{\partial \hat{S}}{\partial \omega} \hat{u}_j.$$  

In eq. (11), we calculate the wavefield derivative ($\partial \hat{u}_j/\partial \omega$) using the virtual source vector (right-hand side) with the same modelling operator $\hat{S}$. The LU-factorized matrices of $\hat{S}$, used to solve eq. (5), can be reused to calculate the wavefield derivative.

The modelled wavefield $\hat{u}_j$ with a high-damping factor, denominator in eq. (9), has an amplitude of $\xi_j e^{-i\alpha \tau_j}$. If the damping factor ($\alpha$) and the first-arrival traveltime ($\tau_j$) have high values, the modelled wavefield can have a smaller value of amplitude than the precision of a computer. In this case, the modelled wavefield is regarded as a complex zero and the division in eq. (9) is not valid. Therefore, higher level precision might be needed for the high values of damping factor and first-arrival traveltime to avoid the errors caused by the precision of computer. The level of precision can be empirically determined according to the damping factor and maximum value of the first-arrival traveltime related with the model size and velocity.

To examine the DPAF estimated from the modelled wavefield, we again use the same velocity model, selected frequencies and the geometries of source and receivers as set-up in the previous example. We first plot the maps of DPAF with offset and frequency variables for different damping factors in Fig. 5. For the very low damping factor (Fig. 5a), DPAF shows many stream-like fluctuations. Very small amplitude value of $\hat{u}_j(\omega)$, which are related to the sharp zeros, causes singular value (fluctuated behaviour) of division process in eq. (9). On the other hand, DPAF for the high-damping factor (Fig. 5b) shows very simple features without wrapping phenomena. Fig. 6 shows the profiles of each map in Fig. 5 at the offset distance of 1 km. DPAF shows many singular values at low-damping factor (Fig. 6a), but only simple features at high-damping factor (Fig. 6b).

DPAF profiles with offset variable for fixed frequencies and damping factors are plotted in Fig. 7. For the low-damping
Inversion of the phase derivative

Figure 7. The profiles of the DPAF map in Fig. 5 at frequencies of (a, c) 0.5 Hz and (b, d) 5 Hz for damping factors: (a, b) 0.00001 s$^{-1}$ and (c, d) 40 s$^{-1}$.

Phase derivative estimated from the observed wavefield

For waveform inversion, DPAF must be also estimated from the observed wavefield (recorded wavefield). A basic feature of Fourier transform is its capability in simplifying a derivative in the frequency domain by:

$$F[i\tilde{d}(t)] = \frac{\partial \tilde{d}(\omega)}{\partial \omega},$$

(12)

where $i$ is the imaginary unit and $\tilde{d}(t)$ is the damped observed wavefield. Thus, to obtain the derivative of the recorded data with respect to the angular frequency, we multiply the damped signal with ‘$i$’ and take it’s Fourier transform. In this case, DPAF is given by dividing $F[i\tilde{d}(t)]$ by $F[\tilde{d}(t)]$ and taking its imaginary part.

Noise included in the recorded data is a potential source of inaccuracies in estimating DPAF. Therefore, we need to carefully pick the first arrivals and mute the signals prior to the first arrivals before estimating DPAF for noisy data. For a strong damping, estimating DPAF for noisy data is very sensitive to the picked first-arrival traveltine. In this case, the estimated DPAF would be the same as the picked traveltine, and thus the estimated DPAF in the objective function can be replaced with the picked traveltine, which is described again in the next subsection. When strong damping is applied, high-quality picks are required for a successful inversion.

As an alternative, the exponential damping can be shifted to the time nearby the first arrival and then applied to avoid errors caused by the precision of computer in the computation of DPAF of the observed wavefield. Instead of the muting and shifting of damping, we can also implement the dynamic damping function in an iterative fashion and average over frequencies (Choi et al. 2011; Saragiotis et al. 2013).

In Fig. 8, we display an original seismogram and damped seismograms with several damping factors. We generate the original seismogram from the Marmousi model using the time-domain modelling technique. The source wavelet is the first derivative of Gaussian function with a maximum frequency of 35 Hz. The shot is located at the distance of 2.5 km. We apply damping factors of 2, 10 and 40 s$^{-1}$. For the damped seismograms, we normalize each time trace by its maximum absolute value to clearly show the features of the damped seismograms. Even the low-damping factor of 2 s$^{-1}$ suppresses much later arrival events in the seismogram (Fig. 8b). The damping factor of 10 s$^{-1}$ significantly suppresses the seismogram (Fig. 8c) and 40 s$^{-1}$ leaves almost only the first-arrival events (Fig. 8d). The damping factors should be examined before choosing them to include the phase of the wanted events in the seismograms.

Waveform inversion algorithm using DPAF

In waveform inversion, the velocity model is iteratively updated in the direction that minimizes the objective function. The gradient calculation using the Fréchet derivative is very expensive and usually not affordable for waveform inversion. Fortunately, the gradient of the objective function can be obtained by using the backpropagation algorithm similar to the reverse-time migration (Pratt et al. 1998) without calculating the Fréchet derivative. The idea is to formulate equations and gather the factors independent on the
Figure 8. The (a) original seismogram and (b–d) corresponding damped seismograms with damping factors of 2, 10 and 40 s$^{-1}$, respectively. Each time trace of the damped seismograms (b–d) is normalized by its maximum absolute value to clearly show the features of the damped seismograms.

model parameters, thus reducing the number of modelling steps. In this section, we derive the gradient of the new objective function using the back-propagation algorithm.

The objective function using DPAF for one shot and one frequency is expressed as follows:

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

where $\hat{u}_j$ and $\hat{d}_j$ are the modelled and observed wavefields at $j$th receiver. When a high-damping factor is applied for noisy data with muting, DPAF is almost the same to the picked traveltime, thus the second term (DPAF of the observed wavefield) in eq. (13) can be replaced with the picked traveltime. The gradient is obtained by taking the derivative of eq. (13) with respect to the $k$th model parameter $p_k$, where $p$ represents the subsurface velocity, and thus, it is expressed in a vector form:

$$\frac{\partial E}{\partial p_k} = \text{Im} \left[ \left( A \frac{\partial a}{\partial p_k} \right)^T r \right] = \text{Im} \left[ \left( \frac{\partial a}{\partial p_k} \right)^T r \right],$$

where, if the discretized model has total $n$ gridpoints, the vector $a$ has $n$ elements given by

$$a_j = \left( \frac{\partial \hat{u}_j}{\partial \omega} \right) / \hat{u}_j,$$

at all gridpoints of the model, and the vector $r$ also has $n$ elements given by

$$r_j = \text{Im} \left[ \left( \frac{\partial \hat{u}_j}{\partial \omega} \right) / \hat{u}_j \right] - \text{Im} \left[ \left( \frac{\partial \hat{d}_j}{\partial \omega} \right) / \hat{d}_j \right].$$
at the receiver gridpoints and \( r_j = 0 \) at the rest of the gridpoints of the model.

The matrix \( \mathbf{A} \) in eq. (14) is the restriction matrix onto the \( nr \) receivers and can be neglected. To obtain the gradient for all model parameters, we need to calculate the derivative \( \partial \mathbf{a} / \partial p_k \) in eq. (14) for all model parameters, which requires modelling steps of twice the number of model parameters. Instead of calculating the derivative \( \partial \mathbf{a} / \partial p_k \), we can implicitly estimate the gradient using the backpropagation algorithm. Our strategy is to reformulate the derivative \( \partial \mathbf{a} / \partial p_k \) and gather the factors independent of the model parameters as done in conventional frequency-domain waveform inversion (Pratt et al. 1998). The derivative \( \partial \mathbf{a} / \partial p_k \) is reformulated as

\[
\frac{\partial \mathbf{a}}{\partial p_k} = \left( \frac{\partial^2 \mathbf{u}_j}{\partial p_k \partial \omega} \right) / \mathbf{u}_j - \frac{\partial ^2 \mathbf{u}_j}{\partial \omega \partial p_k} .
\]

(17)

Substituting eq. (17) into eq. (14) gives

\[
\frac{\partial E}{\partial p_k} = \text{Im} \left[ \left( \frac{\partial \mathbf{u}_j}{\partial p_k} \right) \mathbf{r}_1 - \left( \frac{\partial \mathbf{u}_j}{\partial \omega} \right) \mathbf{r}_2 \right],
\]

(18)

where the elements of \( \mathbf{r}_1 \) and \( \mathbf{r}_2 \) are given, respectively, by

\[
r_{1j} = \frac{1}{\mathbf{u}_j} \text{Im} \left[ \left( \frac{\partial \mathbf{u}_j}{\partial \omega} \right) / \mathbf{u}_j - \left( \frac{\partial \mathbf{u}_j}{\partial \omega} \right) / \mathbf{u}_j \right] \quad \text{and}
\]

\[
r_{2j} = \left( \frac{\partial \mathbf{u}_j}{\partial \omega} \right) \text{Im} \left[ \left( \frac{\partial \mathbf{u}_j}{\partial \omega} \right) / \mathbf{u}_j - \left( \frac{\partial \mathbf{u}_j}{\partial \omega} \right) / \mathbf{u}_j \right].
\]

(19)

at the receiver gridpoints and \( r_{1j} = r_{2j} = 0 \) at the rest of the gridpoints of the model. The two wavefield derivative vectors in eq. (18) can be rewritten by taking the derivative of the matrix equation in eq. (5) and formulating it. The Fréchet derivative can be rewritten by taking the derivative of the matrix equation in form:

\[
\frac{\partial}{\partial \omega} \hat{\mathbf{S}}_{\mathbf{d} j} \mathbf{r}_j = \hat{\mathbf{S}}^{(-1)} \mathbf{r}_j.
\]

(20)

The second derivative \( \frac{\partial^2}{\partial \omega^2} \hat{\mathbf{S}}_{\mathbf{d} j} \mathbf{r}_j \) obtained by taking the derivative of eq. (5) with respect to \( p_k \) and \( \omega \), expressed as

\[
\frac{\partial^2}{\partial p_k \partial \omega} \hat{\mathbf{S}}_{\mathbf{d} j} \mathbf{r}_j = \hat{\mathbf{S}}^{(-1)} \left[ \frac{\partial \hat{\mathbf{S}}}{\partial p_k} \frac{\partial \mathbf{u}_j}{\partial \omega} - \frac{\partial^2 \hat{\mathbf{S}}}{\partial \omega^2} \hat{\mathbf{S}} - \frac{\partial \hat{\mathbf{S}}}{\partial p_k} \frac{\partial \mathbf{u}_j}{\partial \omega} \right].
\]

(21)

Substituting eqs (20) and (21) into eq. (18) and gathering the factors independent on the model parameters admits the following form:

\[
\frac{\partial E}{\partial p_k} = 2 \text{Im} \left[ \left( \frac{\partial \hat{\mathbf{S}}}{\partial p_k} \frac{\partial \mathbf{u}_j}{\partial \omega} - \frac{\partial^2 \hat{\mathbf{S}}}{\partial \omega^2} \hat{\mathbf{S}} - \frac{\partial \hat{\mathbf{S}}}{\partial p_k} \frac{\partial \mathbf{u}_j}{\partial \omega} \right) \mathbf{r}_1 + \left( \frac{\partial \hat{\mathbf{S}}}{\partial p_k} \frac{\partial \mathbf{u}_j}{\partial \omega} \right) \mathbf{r}_2 \right] \times \hat{\mathbf{S}}^{(-1)} \left[ \left( \frac{\partial \hat{\mathbf{S}}}{\partial p_k} \mathbf{r}_1 \right) + \left( \frac{\partial \hat{\mathbf{S}}}{\partial p_k} \mathbf{r}_2 \right) \right].
\]

(22)

We consider only the regular-grid finite element or finite-difference modelling methods for acoustic or isotropic elastic media. In this case, the inverse matrix \( \hat{\mathbf{S}}^{(-1)} \) is symmetric, as such \( (\hat{\mathbf{S}}^{(-1)})^T = \hat{\mathbf{S}}^{(-1)} \). In eq. (22), we gather the factors that are independent of the model parameters in each term: \( \mathbf{S}^{(-1)} \mathbf{r}_1, \mathbf{S}^{(-1)} (\mathbf{r}_1 \mathbf{r}_1)^T \mathbf{S}^{(-1)} \mathbf{r}_1 \) and \( \mathbf{S}^{(-1)} \mathbf{r}_2 \). These factors, within each term, require only one modelling step for all model parameters. For the other factors, only the matrix \( \hat{\mathbf{S}} \) is dependent on the model parameter \( p_k \). Since we already know the coefficients of the matrix \( \hat{\mathbf{S}} \), calculating the derivative of the matrix \( \hat{\mathbf{S}} \) with respect to the model parameter \( p_k \) is trivial. In eq. (22), we back-propagate \( \mathbf{r}_1, \mathbf{r}_2 \), and \( (\mathbf{r}_1 \mathbf{r}_1)^T \mathbf{S}^{(-1)} \mathbf{r}_1 \), and multiply the back-propagated wavefield with each of the virtual wavefields and sum those to calculate the gradient. In our inversion algorithm, we perform a total of five modelling steps to obtain the gradient for all model parameters: calculating \( \mathbf{u} \) and \( \partial \mathbf{u} / \partial \omega \), in addition to the three back-propagating steps described above. Once we factorize the matrix \( \hat{\mathbf{S}} \) to solve for the background velocity, the factorized matrices can be used for all types of source vectors. In other words, the common feature for all five modelling steps is that they share the same operator, which results in huge savings in the computational time.

We use the conjugate-gradient method to update the model parameters. To obtain a proper step length, we could employ a line-search method but, for simplicity, we use a fixed-step length.

**SYNTHETIC EXAMPLES**

We test both waveform inversions using the phase derivative (DPAF inversion) and the wrapped phase (PVP inversion) for a single-frequency data generated from the Marmousi model (Fig. 1). We use the logarithmic phase objective function for the PVP inversion following Min & Shin (2006) and Bednar et al. (2007). For simplicity, we generate the synthetic-observed wavefield using the frequency-domain modelling technique. We regard the source wavelet is known in the inversion. The initial model for the single-frequency inversion is a linearly increasing velocities from 1.5 to 4 km s\(^{-1}\) (Fig. 9). The grid interval for both horizontal and vertical directions is 20 m. We position 200 shots every 40 m at a 20-m depth and receivers at all gridpoints from 1- to 8.2-km distance at a 20-m depth as well. The frequencies used in the single-frequency inversion are (the low) 0.125 Hz and (the high) 5 Hz. We ran 100 iterations for each inversion using the conjugate-gradient approach. We empirically choose 20 m s\(^{-1}\) as a fixed-step length for updating velocity model.

The damping factors used in our inversion test are 10 and 30 s\(^{-1}\). We empirically determine the high-damping factor of 30 s\(^{-1}\) by considering the precision level of our computer and the maximum value of the first-arrival traveltime related with the model size and velocity to make sure that the value of the synthetic and modelled data are accurate.

Fig. 10 shows the inverted models at the damping factor of 30 s\(^{-1}\) by the PVP inversion and DPAF inversion for single frequencies of 0.125–5 Hz. Both inversion algorithms for the low frequency (0.125 Hz) generate reasonable long-wavelength structures.
Figure 10. The inverted velocity models at a damping factor of 30 s$^{-1}$ using (a, b) the PVP inversion and (c, d) the DPAF inversion for a single frequency of (a, c) 0.125 Hz and (b, d) 5 Hz.

(Figs 10a and c), whereas only the DPAF inversion gives a successful result (Fig. 10d) for the high-frequency case (5 Hz). Since the high-damping factor of 30 s$^{-1}$ eliminates the events, exclusive of the first arrival, the inversion results even for the high single-frequency are similar to the tomographic result. However, because the PVP inversion suffers from the phase-wrapping effect at high frequencies, it generates an incorrect inversion result even with a high-damping factor (Fig. 10b). Fig. 11 compares the residuals of PVP and DPAF at the first and last iterations for a damping factor of 30 s$^{-1}$ and a single frequency of 5 Hz. The residual of PVP is estimated as Im($\log(\hat{u}_j/\hat{d}_j)$) (Min & Shin 2006; Bednar et al. 2007). The geometry of shot and receiver is the same with Figs 4 and 7. The estimated PVP residual at the first iteration shows the wrapping phenomena around 4.2 ∼ 4.5 km (Fig. 11a), in which the modelled and observed phase are in different cycles at the first iteration. Even though the PVP residual decreases at the last iteration (Fig. 11a), the modelled and observed phase are also in different cycles at the last iteration. The wrapping phenomena of the PVP residual mainly results in an incorrect image of inversion at high frequencies (Fig. 10b). On the other hand, the DPAF residual has no wrapping effect and converges well at the last iteration (Fig. 11b), which validates our inversion algorithm.

We perform the inversion tests at the lower damping factor of 10 s$^{-1}$ for the same single frequencies as in the previous inversion example. Fig. 12 shows the inverted models. Both inversion algorithms for low frequency (0.125 Hz) still generate the coincident and successful inversion results (Figs 12a and c). However, the DPAF inversion does not yield successful results for high frequency (Fig. 12d), because the lower damping factor does not considerably eliminate the complexity of wavefield. Fig. 13 shows the comparison of residuals of PVP and DPAF at the first and last iterations for a damping factor of 10 s$^{-1}$ and a single frequency of 5 Hz. The DPAF residual at the last iteration does not converge well (Fig. 13b), which shows an influence of the complexity of high-frequency content at low-damping factor in the inversion. Therefore, we need lower frequency content at lower damping factor to avoid an influence of complexity of seismic data. The PVP residual still shows the cycle-skipping effect at the first iteration (Fig. 13a), resulting in a failure of the inversion. Fig. 14 shows the history of the misfit function of the DPAF inversion for different damping factors of 10 and 30 s$^{-1}$. The rms error for the damping factor of 30 s$^{-1}$...
Inversion of the phase derivative

1913

Figure 12. The inverted models at a damping factor of $10 \text{s}^{-1}$ using (a, b) the PVP inversion and (c, d) the DPAF inversion for a single frequency of (a, c) $0.125 \text{Hz}$ and (b, d) $5 \text{Hz}$.

Figure 13. The residuals of (a) PVP and (b) DPAF at the first (solid line) and last (cross) iterations for a damping factor of $10 \text{s}^{-1}$ and a single frequency of $5 \text{Hz}$. The shot is located at a distance of $2 \text{km}$ (offset distance of $0 \text{km}$).

Figure 14. The history of the misfit function of the DPAF inversion for single frequency of $5 \text{Hz}$ at damping factors of $10 \text{s}^{-1}$ (cross) and $30 \text{s}^{-1}$ (solid line).

decreases below 1 per cent of the initial value, whereas that for the damping factor of $10 \text{s}^{-1}$ hardly converges. We note that the complexity of high-frequency content at low-damping factor interrupt the single-frequency inversion.

For waveform inversion, we can somewhat mitigate the complexity of high frequency at low-damping factor by using a finite frequency band instead of a single frequency. We perform the inversion tests at the lower damping factor of $10 \text{s}^{-1}$ for the frequency band of $2.5 \sim 5 \text{Hz}$. Since the low frequencies of real data are usually not available, we exclude frequencies lower than $2.5 \text{Hz}$. We calculate the gradients for each frequency within the frequency band and sum the estimated gradients. The non-linearity of gradient for each frequency has an inconsistent property with each other, so that the summation of the gradients mitigates the problem of non-linearity. Fig. 15 shows the inverted models for a finite frequency band by the DPAF inversion and PVP inversion. The PVP inversion still does not show the convergence because of the phase-wrapping problem, but the DPAF inversion improves the convergence of inverted velocity model.

We perform the conventional least-squares waveform inversion of low-cut filtered data ($4 \sim 5 \text{Hz}$) starting from the linearly increasing velocities (Fig. 9) and the inverted velocities by our inversion algorithm for high-damping factor and high single frequency (Fig. 10d). Fig. 16 shows the final inverted model after 100 iteration for low-cut filtered data ($4 \sim 5 \text{Hz}$). Fig. 17 shows the depth profiles of the
Figure 15. The inverted models using a damping factor of $10 \text{s}^{-1}$ and a frequency band of $2.5 \sim 5$ Hz by (a) the PVP inversion and (b) the DPAF inversion.

Figure 16. The inverted models by the conventional least-squares inversion starting from (a) the linearly increasing velocities and (b) the inverted velocities obtained from the single-frequency inversion in Fig. 10(d).

Figure 17. Depth profiles of the true model (solid line) and final inverted models in Fig. 16 at the distances of (a) 3 km and (b) 6 km: the final inverted model starting from the linearly increasing velocities (cross) and the final inverted model starting from the single-frequency inversion result (circle).

discussion

Since the exponential damping attenuates later arrival events in return for mitigating the complexity of Green’s function, a proper damping factor must be chosen in the inversion process. In determining a proper damping factor, we must consider the lowest frequency of the available frequency band. If we can use a very low frequency component, we might not need to employ a damping term in the inversion process in the sense of the multiscale approach (Bunks et al. 1995). As the available lowest frequency become higher, a damping factor is needed and depending on the frequency content, we might need a strong damping to suppress the complexity of the high-frequency content. We also must consider the complexity of velocity model (or seismic data) to determine a proper damping factor. If the subsurface velocity model or the acquired seismic data is not complex, we do not need a strong damping term.

It is not easy to find a functional relation between the proper damping factor and the two factors, the available lowest frequency and the complexity of velocity model. Thus, we focus on introducing a new concept of inversion using DPAF and empirically find a proper damping factor. The DPAF component of the new inversion
inherently mitigates the wrapping problem caused by the frequency content of our signal. The damping is required to handle the other source of non-linearity in our inversion caused by the complexity of the velocity model and the resulting complex reflections. As much as the damping is able to attenuate such reflections, thus, relying on turning waves in the inversion, we can converge to a velocity model regardless of the frequency content. A similar assertion can be made if the velocity model is smooth and free of reflections. As soon as the inverted velocity model is able to mildly predict the reflections, which is easier at the lower frequencies, we can slowly loosen the damping screws and allow for reflection information to contribute to the inversion.

With strong damping, our inversion algorithm is similar to the traveltime tomography but more reliable than the conventional ray- and correlation-based traveltime tomography. The ray- and correlation-based tomography algorithms employ the ray path and hallow banana-type sensitivity kernels, respectively, whereas our inversion algorithm with a strong damping has a full sensitivity kernel since full wavefield is used in our inversion algorithm (Van Leeuwen & Mulder 2010; Djebbi & Alkhalifah 2013). It is intuitively understood that the ray path and hallow banana-type sensitivity kernels are less accurate in the traveltime tomography than the full sensitivity kernel.

CONCLUSIONS

We develop a new inversion algorithm using the phase derivative, or what we refer to as the instantaneous traveltime, to avoid the phase-wrapping problem. The phase derivative is defined as the imaginary part of the ratio of the wavefield derivative with respect to the angular frequency to the wavefield itself. In this case, the phase derivative does not have any bounds on range, and therefore, has no wrapping problem. For the inversion, we construct the objective function using the phase derivative. The gradient of the objective function is computed using the back-propagation algorithm.

A damping factor is incorporated in our inversion algorithm. Since a high-damping factor suppresses the events following the first arrival, it eliminates the complexity of seismic data borne of the complex reflection phenomena. Numerical examples show that our inversion algorithm with a high-damping factor generates a tomographic result even for a high single frequency, whereas the wrapped-phase inversion does not because of the phase-wrapping phenomena. Since the real data have insufficient low-frequency content, the inverted model by our inversion algorithm for high single frequency can serve as a good initial model for waveform inversion and migration. On the other hand, our inversion algorithm with a weak damping factor inherits the same non-linearity weakness plaguing the conventional waveform inversion. We can somewhat mitigate the problem of the low-damping factor by including a finite frequency band instead of a single frequency.

ACKNOWLEDGEMENTS

We are grateful to King Abdullah University of Science and Technology for financial supports. We also thank the editor and the reviewers for their fruitful and helpful review of the paper.

REFERENCES


