Accelerating full-waveform inversion with attenuation compensation

Zhiguang Xue\textsuperscript{1}, Junzhe Sun\textsuperscript{1}, Sergey Fomel\textsuperscript{1}, and Tieyuan Zhu\textsuperscript{2}

\textbf{ABSTRACT}

The calculation of the gradient in full-waveform inversion (FWI) usually involves crosscorrelating the forward-propagated source wavefield and the back-propagated data residual wavefield at each time step. In the real earth, propagating waves are typically attenuated due to the viscoelasticity, which results in an attenuated gradient for FWI. Replacing the attenuated true gradient with a $Q$-compensated gradient can accelerate the convergence rate of the inversion process. We have used a phase-dispersion and an amplitude-loss decoupled constant-$Q$ wave equation to formulate a viscoacoustic FWI. We used this wave equation to generate a $Q$-compensated gradient, which recovers amplitudes while preserving the correct kinematics. We construct an exact adjoint operator in a discretized form using the low-rank wave extrapolation technique, and we implement the gradient compensation by reversing the sign of the amplitude-loss term in the forward and adjoint operators. This leads to a $Q$-dependent gradient preconditioning method. Using numerical tests with synthetic data, we demonstrate that the proposed viscoacoustic FWI using a constant-$Q$ wave equation is capable of producing high-quality velocity models, and our $Q$-compensated gradient accelerates its convergence rate.

\textbf{INTRODUCTION}

Full-waveform inversion (FWI) is a data-fitting procedure used to construct high-resolution quantitative subsurface models by exploiting the full information in the observed data, which has been extensively developed in recent years (Virieux and Operto, 2009). To solve FWI in the framework of the local optimization approach, a model update needs to be computed at each iterative step and each model update involves calculating the gradient of the misfit function with respect to the model parameters (Tarantola, 2005). Due to the non-linearity of FWI in practice, FWI may converge slowly and cause the algorithm to get trapped in a local minimum. To accelerate the convergence rate of FWI, one can calculate a more accurate model update by applying an approximate Hessian inverse to its gradient (Ma and Hale, 2012; Qin et al., 2015; Hou and Symes, 2016). Another more common approach is to precondition the gradient, for instance, using the energy of the source wavefield to normalize the gradient (Gautier et al., 1986), weighting the gradient by depth-dependent functions (Shipp and Singh, 2002; Wang and Rao, 2009), or smoothing/filtering the gradient in the spatial frequency or other transform domains (Guitton et al., 2012; Xue et al., 2017). Because seismic waves propagating in viscoacoustic media are attenuated, a natural gradient preconditioning approach is to compensate for the energy loss caused by attenuation. Here, we limit our focus on intrinsic attenuation, although seismic scattering attenuation caused by heterogeneity also complicates the amplitude analysis (Browaey and Fomel, 2009).

The classic approach of superposition of several standard linear solids (SLSs) has been widely used in the time-domain finite-difference method to approximate the constant-$Q$ seismic wave propagation for a specific frequency band (Robertsson et al., 1994). However, compensating $Q$ effects with SLSs is challenging because reversing the sign of memory variables that account for attenuation does not correctly compensate for the phase (Zhu, 2014, 2016; Guo and McMechan, 2015). To overcome this issue, a better choice is to compensate for amplitude loss by using a modified wave equation, which has separate control over amplitude loss and phase dispersion effects (Zhu and Harris, 2014). This idea has been applied previously in...
reverse time migration (RTM) to get more balanced illumination in seismic images (Zhang et al., 2010; Suh et al., 2012; Bai et al., 2013; Zhu et al., 2014; Sun et al., 2015; Zhu and Sun, 2017), but its application to velocity model building remains to be investigated.

In this paper, the viscoacoustic wave equation proposed by Zhu and Harris (2014) is adopted to formulate a preconditioned approach to $Q$-compensated FWI. This equation involves fractional Laplacians and has separate control over phase dispersion and amplitude loss effects. In this way, compensating for the amplitude loss while preserving the correct kinematics can be done by reversing the sign of the amplitude-loss factor and keeping the sign of the dispersion factor unchanged (Zhu, 2016). This property was used recently by Sun et al. (2016b) to accelerate the convergence rate of least-squares RTM. We use it in this paper to accelerate the convergence rate of FWI. In the following, we first present the theory related to the forward operator, adjoint operator, FWI gradient, and $Q$-compensated gradient. Then, we demonstrate the effectiveness of our preconditioning strategy for FWI using a synthetic example.

**THEORY**

**Forward operator**

To describe the intrinsic attenuation behavior in subsurface media, we choose the constant-$Q$ model (Kjartansso, 1979) that is mathematically simple and it is widely used for applications in exploration seismology. Based on this model, Zhu and Harris (2014) propose a constant-$Q$ viscoacoustic wave equation with decoupled attenuation effects as follows:

$$\frac{1}{c^2} \frac{\partial^2 P}{\partial t^2} = \nabla^2 P + \beta_1 \left( \eta (\nabla^2)^{r+1} - \nabla^2 \right) P + \beta_2 \left\{ \tau \frac{\partial}{\partial t} (\nabla^2)^{r+1/2} \right\} P,$$

where

$$\gamma(x) = \arctan(1/Q(x))/\pi,$$

$$c^2(x) = c^2_0(x) \cos^2(\pi \gamma(x)/2),$$

$$\eta(x) = -c^2_0(x) \omega_0^{-2r(x)} \cos(\pi \gamma(x)),$$

$$\tau(x) = -c^2_0(x) \omega_0^{-2r(x)} \sin(\pi \gamma(x)).$$

The function $P(x, t)$ indicates the pressure wavefield, $Q(x)$ is the quality factor, and $c_0(x)$ stands for the acoustic velocity model defined at a reference frequency $\omega_0$. Note that $\gamma(x)$ defined in equation 2 ranges from 0 to 1/2 and is a dimensionless parameter. The term starting with $\beta_1$ parameter in equation 1 characterizes the velocity dispersion, and the term starting with $\beta_2$ describes the amplitude loss. Thus, $\beta_1$ and $\beta_2$ take values of $\pm 1$ and act like on/off switches to control phase dispersion and amplitude loss effects, respectively.

Equation 1 involves fractional Laplacians and has separate controls over phase dispersion and amplitude loss, which is a big advantage as opposed to other formulations of viscoacoustic wave equation. As shown by Zhu and Harris (2014), this equation is also valid for very low $Q$ values, such as 10–20. In our numerical example, we set the reference frequency $\omega_0$ to be very high and assume that the intrinsic attenuation will delay the propagation of the seismic wave.

Setting $\beta_1$ and $\beta_2$ to one, equation 1 leads to the following viscoacoustic dispersion relation with fractional powers of the wave-number:

$$\frac{\omega^2}{c^2} = -\eta |k|^{2r+2} - i\omega \tau |k|^{2r+1},$$

from which we can solve for the angular frequency $\omega$ as follows (Sun et al., 2015):

$$\omega = \frac{-ip_1 + p_2}{2},$$

where

$$p_1 = \tau c^2 |k|^{2r+1},$$

$$p_2 = \sqrt{-\tau^2 c^4 |k|^{4r+2} - 4\eta c^2 |k|^{2r+2}}.$$

Because the first term under the square root in $p_2$ does not affect the amplitude of wave propagation and only has a relatively small effect on the phase due to the small value of $\tau$ (Sun et al., 2015), equation 9 can be approximated as

$$p_2 \approx \sqrt{-4\eta c |k|^{r+1}}.$$  

In heterogeneous media, the form of $\omega$ in equation 7 can be used to define the phase shift symbol $\phi$ (Zhang and Zhang, 2009). With the use of a one-step extrapolation method, the pressure wavefield $P(x, t)$ becomes complex and satisfies the following first-order partial differential equation:

$$\left( \frac{\partial}{\partial t} + i\phi \right) P(x, t) = 0,$$

which can be solved using the following mixed-domain time marching operator (Sun et al., 2016a):

$$P(x, t + \Delta t) = \int \tilde{P}(k, t) e^{i\phi(x.k, \Delta t)} dk,$$

where $\tilde{P}$ is the spatial Fourier transform of $P$ and the phase function $\phi_1(x, k, \Delta t)$ is defined as

$$\phi_1(x, k, \Delta t) = x \cdot k + q \Delta t = x \cdot k + \frac{-ip_1 + p_2}{2} \Delta t.$$

To implement the wave propagation numerically, we use the low-rank approximation method of Fomel et al. (2013b) to decompose the wave extrapolation matrix $e^{i\phi(x, k, \Delta t)}$ into the following
can be formulated as
\[ e^{i\phi(x,k,D\Delta t-kx)} \approx \sum_{m=1}^{M} W(x,k_m)a_{mn} W(x_n,k). \] (14)

The computation of \( P(x,t+\Delta t) \) then becomes
\[ P(x,t+\Delta t) \approx \sum_{m=1}^{M} W(x,k_m) \left\{ \sum_{n=1}^{N} a_{mn} \left[ \int e^{i\mathbf{k} \cdot \mathbf{x}} W(x_n,k) \tilde{P}(k,t) d\mathbf{k} \right] \right\}. \] (15)

whose computational cost equals that of applying \( N \) inverse fast Fourier transforms per time step, where \( N \) is the approximation rank. We can rewrite equation 15 in an algebraically compact form as
\[ P(x,t+\Delta t) = \mathbf{A} P(x,t). \] (16)

Correspondingly, the full forward-modeling process incorporating the source term \( f(x,t) \) can be expressed as
\[
\begin{bmatrix}
\mathbf{I} & -\mathbf{A} & -\mathbf{A} & \cdots & -\mathbf{A} \\
-\mathbf{A} & \mathbf{I} & -\mathbf{A} & \cdots & -\mathbf{A} \\
& & \ddots & \ddots & \ddots \\
& & & \mathbf{I} & -\mathbf{A} \\
& & & & \mathbf{I}
\end{bmatrix} \begin{bmatrix}
P(x,0) \\
P(x,\Delta t) \\
P(x,2\Delta t) \\
\vdots \\
P(x,n\Delta t)
\end{bmatrix}
= \begin{bmatrix}
f(x,0) \\
f(x,\Delta t) \\
f(x,2\Delta t) \\
\vdots \\
f(x,n\Delta t)
\end{bmatrix}.
\] (17)

The absorbing boundary condition is used to eliminate boundary effects during wave propagation, and it has been intrinsically defined in the wave-propagation matrix (Sun et al., 2016a).

**Adjoint operator**

Based on equation 17, the adjoint operator (conjugate transpose) can be formulated as
\[
\begin{bmatrix}
\mathbf{I} & -\mathbf{A}^* & -\mathbf{A}^* & \cdots & -\mathbf{A}^* \\
-\mathbf{A} & \mathbf{I} & -\mathbf{A}^* & \cdots & -\mathbf{A}^* \\
& & \ddots & \ddots & \ddots \\
& & & \mathbf{I} & -\mathbf{A}^* \\
& & & & \mathbf{I}
\end{bmatrix} \begin{bmatrix}
\tilde{P}(x,0) \\
\tilde{P}(x,\Delta t) \\
\tilde{P}(x,2\Delta t) \\
\vdots \\
\tilde{P}(x,n\Delta t)
\end{bmatrix}
= \begin{bmatrix}
\tilde{f}(x,0) \\
\tilde{f}(x,\Delta t) \\
\tilde{f}(x,2\Delta t) \\
\vdots \\
\tilde{f}(x,n\Delta t)
\end{bmatrix}.
\] (18)

where \( * \) denotes the complex conjugate transpose of a matrix. The adjoint wavefield can be recursively calculated with the following backward time marching scheme (Sun et al., 2016b):
\[ \tilde{P}(x,t) \approx \sum_{n=1}^{N} W(x_n,k) \]
\[ \times \left\{ \sum_{m=1}^{M} a_{mn} \left[ \int e^{-i\mathbf{k} \cdot \mathbf{x}} W(x_n,k) \tilde{P}(x,t+\Delta t) d\mathbf{k} \right] \right\}. \] (20)

Equations 15 and 20 differ from each other in the multiplication order of low-rank decomposed small matrices. In addition, the low-rank matrices used in equation 20 are the complex conjugate transpose of the ones in equation 15. The relative error of the dot-product transpose of the ones in equation 15. The relative error of the dot-product transpose of the ones in equation 15.

**FWI gradient**

The least-squares objective function of standard FWI can be written as

Figure 1. Target and initial models. (a) A portion of a modified Marmousi velocity model that includes a water layer from the surface down to 200 m, (b) a \( Q \) model obtained from the velocity model and including two gas channels, and (c) the initial velocity model.
where $d_{\text{obs}}$ is the observed waveform data and $d_{\text{syn}}$ is the synthetic data.

According to the adjoint-state method (Tarantola, 2005; Plessix, 2006), the FWI gradient can be expressed as

$$
g(x) = \sum_{t} P(x, t) \cdot \frac{\partial F}{\partial c_{0}} \cdot \tilde{P}(x, t),$$

(22)

where the state variable $P(x, t)$ can be obtained by solving equation 17, the adjoint variable $\tilde{P}(x, t)$ is calculated by solving equation 18 with data residual $d_{\text{obs}} - d_{\text{syn}}$ as the source term $\tilde{f}(x, t)$, and $F$ is the forward operator, whose discretized form can be expressed as the square matrix in equation 17. Using equations 7 and 11, $F$ can also be expressed as

$$
F = \left( \frac{\partial}{\partial t} + ip \right) = \left( \frac{\partial}{\partial t} + \frac{p_{1} + ip_{2}}{2} \right).
$$

(23)

With the definitions of $p_{1}$ and $p_{2}$ in equations 8 and 10, respectively, we can derive the gradient of $F$ with respect to velocity $c_{0}$ as follows:

$$
\frac{\partial F}{\partial c_{0}} = -\frac{2 \gamma + 1}{2} c_{0}^{\gamma} \alpha_{0}^{-2\gamma} \sin(\pi \gamma) \cos^{2} \left( \frac{\pi \gamma}{2} \right) |k|^{2\gamma + 1}
$$

$$
+ i(\gamma + 1)c_{0}^{\gamma} \alpha_{0}^{-\gamma} \sqrt{\cos(\pi \gamma)} \cos \left( \frac{\pi \gamma}{2} \right) |k|^{\gamma + 1},
$$

(24)

which is a function of $x$, and can be substituted into equation 22 to calculate the velocity gradient.

**Q-compensated gradient**

The gradient construction in FWI requires the forward-propagated source wavefield and the backward-propagated data residual wavefield. Without compensation, the source wavefield will be attenuated, and the data residual wavefield will be attenuated once again, which means that the crosscorrelation of the source wavefield and data residual wavefield will suffer from the intrinsic attenuation along the entire wave propagation path twice.

To compensate for attenuation, we can simply set $\beta_{2}$ in equation 1 to be $-1$ and keep $\beta_{1}$ unchanged. In this way, the amplitude can be amplified based on the $Q$ model and the dispersion effects will be counteracted (Zhu et al., 2014). Thus, the attenuation-compensated constant-$Q$ wave equation corresponds to the following dispersion relation:

$$
\frac{\omega^{2}}{C^{2}} = -\eta |k|^{2\gamma + 2} + i\alpha |k|^{2\gamma + 1},
$$

(25)

which defines a new phase function (Sun et al., 2015):

$$
\phi_{2}(x, k, \Delta t) = x \cdot k + \frac{i p_{1} + p_{2}}{2} \Delta t,
$$

(26)

where $\phi_{2}(x, k, \Delta t)$ is the complex conjugate of $\phi_{1}(x, k, \Delta t)$. For consistency with the gradient amplitude in acoustic media, we need to use $\phi_{2}(x, k, \Delta t)$ to compute the source wavefield and data residual wavefield. Although the wavefield for calculating the data residual is attenuated along the whole propagation path, the source and data residual wavefields for calculating the velocity gradient are compensated to accumulate the correct compensation factor along the entire propagation path. This resembles the $Q$-compensated RTM. When the crosscorrelation imaging condition is applied, attenuation compensation should be applied to the forward and backward wave propagations; when the deconvolution imaging condition is used, attenuation compensation should be applied just to backward wave propagation (Zhu, 2016).

**NUMERICAL EXAMPLE**

We use a portion of a modified Marmousi velocity model (Figure 1a) and a built $Q$ model (Figure 1b) to test the $Q$-compensated FWI. The models include a water layer from the surface down to 200 m. Figure 1c shows the initial velocity model, which is obtained by using a Gaussian function to convolve the original velocity model with a radius of 450 m, but keeping the water layer unchanged. A fixed-spread acquisition survey is generated, which consists of 26 shots spaced every 200 m. The source wavelet for generating the synthetic data is a Ricker wavelet centered at 10 Hz. The record length of the synthetic data is 2.2 s with a time step of 2 ms. To test the forward operator used in FWI, we compare the acoustic shot record generated by a finite-difference method and the shot record generated by our viscoacoustic low-rank one-step operator with $Q = 10,000$. The comparison in Figure 2

Figure 2. Test low-rank one-step forward operator used in FWI. (a) Low-rank one-step shot record, (b) finite-difference shot record, and (c) single trace comparison.
demonstrates that the low-rank one-step operator can simulate the same shot record even though it is based on a different formulation of the seismic wave equation.

We perform two FWI experiments: one using the conventional velocity gradient and the other one using the proposed $Q$-compensated gradient. The multiscale technique (Bunks et al., 1995) is used to avoid the cycle-skipping problem. The seismic data are filtered into three frequency groups of increasing frequency content: 2–5, 2–9, and 2–15 Hz. The optimization method is nonlinear conjugate gradients. We performed 20 iterations for each of the frequency groups successively. Without $Q$-compensation, source and data residual wavefields get attenuated. Figure 3 shows the (source and data residual) wavefield snapshot and the corresponding gradient comparisons without and with $Q$-compensation. These wavefields have the same traveltime of 1 s and are extracted from the same iteration of FWI. It is apparent that the wavefields with $Q$-compensation have a larger amplitude, especially for the deeper part, in which the accumulated attenuation effects are stronger. With the compensated wavefields, we can construct a $Q$-compensated gradient, which has the same phase as the compensation-free gradient, but a $Q$-dependent amplitude recovery.

Figure 4 shows the FWI results after the inversions with and without $Q$-compensation for the three frequency groups. Comparing Figure 4a and 4b, we observe the obvious improvement in the central and deep parts of the velocity model brought by the $Q$-compensated gradient. This phenomenon is easy to understand. As indicated in Figure 3, $Q$-compensated gradient has a larger amplitude, especially for the deep part. In the subsequent two stages, $Q$-compensated FWI keeps recovering more accurate velocity models at the cost of the same iteration numbers. Figure 4e and 4f shows the final inverted velocities without and with $Q$-compensated gradient. The rock cap (indicated by arrows) in the velocity obtained by $Q$-compensated FWI has a larger value, and it is closer to the true model. Figure 5 shows the data misfit convergence comparison of FWI with and without $Q$-compensation at the three stages. At the initial several iterations of the first stage, $Q$-compensated FWI has a

![Figure 3. Wavefield snapshots at traveltime 1 s, and the corresponding gradients. Source wavefield (a) without and (b) with $Q$-compensation; data residual wavefield (c) without and (d) with $Q$-compensation; and gradient (e) without and (f) with $Q$-compensation.](image)
Figure 4. The multiscale FWI results. Twenty iteration results of 2–5 Hz data (a) without and (b) with $Q$-compensation; 20 iteration results of 2–9 Hz data (c) without and (d) with $Q$-compensation; and 20 iteration results of 2–15 Hz data (e) without and (f) with $Q$-compensation.

Figure 5. Data misfit convergence curves of the multiscale FWI (2–5, 2–9, and 2–15 Hz) without (red dashed line) and with (blue solid line) $Q$-compensation.
faster convergence rate, and then it keeps the smaller misfit as the iterations continue. Comparing the misfit difference through the three stages, we can find that the relative misfit difference becomes bigger as the data frequencies increase. This implies that the attenuation effect is more obvious for high frequencies, and therefore the $Q$-compensated gradient can play a more significant role for high-frequency data inversion.

**DISCUSSION**

We use the proposed method of preconditioning to accelerate the convergence rate of time-domain FWI for velocity reconstruction. The synthetic example shows that the $Q$-compensated gradient can effectively accelerate the rate of convergence, although the low-frequency components of seismic data, which are less sensitive to attenuation, are mainly used in FWI. Previous studies about time-domain FWI in the presence of attenuation have shown that considering attenuation as a smooth background modeling parameter may significantly improve the velocity reconstruction (Kurtzmann et al., 2013; Bai et al., 2014). Compared with those studies, our method not only incorporates the attenuation effects in the wave simulation process but it also aims to construct an amplitude-compensated and phase-preserved gradient based on the attenuation model. Similarly, the attenuation model used in this study is assumed to be known and it is not an inversion parameter. FWI inversion strategies for simultaneous and hierarchical velocity and attenuation inversion have been investigated by Kamei and Pratt (2013). Yang et al. (2016) provide a theoretical review on the formulation of 3D viscoelastic multiparameter FWI. Qu et al. (2017) implement viscoacoustic anisotropic FWI and test their method’s sensitivity to velocity, $Q$, and anisotropic parameters. A background $Q$ model can be estimated using either data-domain or image-domain tomographic techniques (Brzostowski and McMechan, 1992; Quan and Harris, 1997; Hu et al., 2011; Zhou et al., 2011; Venciano and Chemingui, 2012; Shen et al., 2014, 2015; Shen and Zhu, 2015; Dutta and Schuster, 2016).

**CONCLUSION**

We have implemented a viscoacoustic FWI using phase and amplitude decoupled constant-$Q$ wave equation. By reversing the sign of amplitude-loss term in this equation, we have constructed the $Q$-compensated gradient that has the correct phase information and a $Q$-dependent amplitude recovery. The low-rank wave extrapolation technique is used to formulate the exact adjoint operator in the construction of the FWI gradient. Our testing results show that the FWI convergence rate can be notably accelerated by the $Q$-compensated gradient, which helps to rapidly construct the deep structures of the velocity model.

**ACKNOWLEDGMENTS**

We thank the assistant editor D. Draganov, the anonymous associate editor, reviewer J. Hou, and another anonymous reviewer for providing valuable suggestions. We thank the sponsors of the Texas Consortium for Computational Seismology for financial support. Z. Xue and J. Sun were additionally supported by the Statoil Fellows Program at the University of Texas at Austin. We also thank the Texas Advanced Computing Center for providing the computational resources used in this study. All computations presented in this paper are reproducible using the Madagascar software package (Fomel et al., 2013a).

**REFERENCES**


