Nonlinear multiparameter full waveform inversion based on truncated Newton method

Yu Geng, Kris Innanen and Wenyong Pan

ABSTRACT

Full waveform inversion (FWI) is a powerful tool to reconstruct subsurface parameters. This highly nonlinear inverse problem is normally solved iteratively to minimize a misfit function, which is usually defined as the distance between the observed and predicted data, by gradient-based method or Newton type method. Incorporating more nonlinearity within each update in FWI, especially for multiparameter reconstruction, may have very important consequences for convergence rates and discrimination of different parameter classes. In this study, we focus on acoustic media with variable density, and the goal is to simultaneously update velocity and density, other parameterization is also discussed. We start from the physical interpretation of both the gradient and the Hessian of the misfit function, and derive one approach from the Newton method, to include the additional term of the Hessian, which contains the second-order partial derivative of the wavefield and related to the second-order scattering, into the gradient, to construct a new descent direction. A matrix-free scheme is used to efficiently calculate the product of the Hessian and a vector.

INTRODUCTION

In the last decade, full waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009) has gradually become a mature tool to reconstruct the subsurface parameters. Using the whole information in seismic data, FWI can be seen as a nonlinear least-square minimization problem. A best fit of the subsurface parameters is found iteratively to minimize a misfit function, which measures the distance between the observed and predicted data computed through a forward modelling with these parameters. Several approaches can be used to solve this optimization problem, such as gradient-based method, quasi-Newton method and Gauss-Newton/Newton method.

With the full wavefield accounted in FWI, it is naturally to include more realistic physics in the forward modelling to better match the observed data, such as viscosity, elasticity and anisotropic effects (e.g., Fichtner, 2011; Operto et al., 2013; Plessix et al., 2013; Alkhalifah and Plessix, 2014; Pan et al., 2016). In this case, multiparameter inversion has become feasible to invert parameter classes other than P-wave velocity, such as density, attenuation, shear-wave velocity and so on. However, adding more parameter classes in FWI can increase the ill-posedness of the inverse problem, and inverting multiparameter is much more complicated than the monoparameter inversion due to the potential presence of trade-off/cross-talk between different parameter classes and more degrees of freedom are considered in the parameterization. It is difficult to distinguish the change of the seismic data caused by perturbations of different parameter classes, since they can be more or less coupled. Studies show that hierarchical strategies can be used to invert different parameter classes. However, as many other studies show that in multiparameter inversion, involving the Hessian operator (e.g., Pratt et al., 1998), which is the second-order partial derivative of the misfit function with respect to the parameters, can help to mitigate the cross-talk be-

tween different parameter classes. Gauss-Newton approximation of the Hessian operator is usually used instead of the full Hessian operator, by neglecting the additional terms which contain second-order partial derivatives of the wavefield in the Hessian operator. This additional terms could be essential in FWI by correcting the descent direction with energy from multiple scattering in the residuals with the first-order partial derivative of the wavefield. However, this term needs to be carefully dealt in the Newton method, since adding this term into the Gauss-Newton Hessian operator could destroy the positive definiteness of the Hessian operator, and cause a significant false prediction of the perturbations.

In this study, we present a nonlinear FWI method, based on a discrete frequency-space domain acoustic finite-difference forward modelling. This nonlinear FWI method can be derived from the Newton method, and the second-order scattering term is included in the gradient to form a new descent direction. As in the Gauss-Newton method, Gauss-Newton Hessian operator is used to help mitigating cross-talk between different parameter classes. To avoid explicitly calculation of the Hessian matrix and better account the inverse of the Hessian operator, a matrix-free formalism is then used to compute a product between the Hessian operator and a vector as in the truncated Newton method (Métivier et al., 2013, 2014; Pan et al., 2017). Although we focus on acoustic case, to simultaneously update velocity and density, other parameterizations are also discussed, and similar derivation can be easily expended to elastic case.

FORWARD MODELLING IN ACOUSTIC MEDIA

We will use the frequency domain acoustic wave equation to describe the wave motions

$$\frac{\omega^2}{\kappa(\mathbf{x})}u(\mathbf{x},\mathbf{x}_s,\omega)+\nabla\cdot\left(\frac{1}{\rho(\mathbf{x})}\nabla u(\mathbf{x},\mathbf{x}_s,\omega)\right)=f_s(\omega)\delta(\mathbf{x}-\mathbf{x}_s), \tag{1}$$

where $\kappa(\mathbf{x})$ is the bulk modulus, and $\rho(\mathbf{x})$ is the density, $u(\mathbf{x},\mathbf{x}_s,\omega)$ is the pressure wavefield at position $\mathbf{x}$, generated by a point source located at $\mathbf{x}_s$ with spectrum $f_s(\omega)$. Rewrite the model parameters with different classes into vector form as $\mathbf{m}$, discretized wave equation can be written in matrix form as

$$\begin{aligned}&\mathbf{A}(\mathbf{m},\omega)\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)=\mathbf{f}(\mathbf{x}_s,\omega),\quad\text{and}\\&\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)=\mathbf{A}^{-1}(\mathbf{m},\omega)\mathbf{f}(\mathbf{x}_s,\omega),\end{aligned} \tag{2}$$

Where $\mathbf{A}(\mathbf{m},\omega)$ is the impedance matrix, and it is a sparse banded matrix, as the number of non-zero diagonals are related to the finite-difference scheme, e.g., in this study, we use a five-point finite difference scheme, so the impedance matrix has five non-zero diagonals. Suppose we compute the wavefield in a $nz\times nx=nm$ nodal points on a regular grid, the impedance matrix is a $nm\times nm$ matrix with model parameters $(nm\times npar)\times 1$ column vector (in our case, we have two different parameter classes, so $npar=2$), and both the source and the wavefield vector are $nm\times 1$ column vector. The wavefield can be obtained by the inverse of the impedance matrix, which is usually replaced by direct matrix factorization methods, such as LU decomposition, in most of the applications, to avoid the calculation of inverting the large sparse matrix for each forward modelling. It is also straight forward that when the source spectrum is 1, the columns of $\mathbf{A}^{-1}(\mathbf{m},\omega)$ stands for the discrete approximations of the Green's functions correspond to the related source locations.

NONLINEAR FULL WAVEFORM INVERSION IN FREQUENCY DOMAIN

FWI method is to seek the high resolution estimation of the subsurface model parameters by solving a nonlinear least-squares minimization problem. The misfit function is defined as the L_2 norm of the data residuals

$$\phi(\mathbf{m}) = \frac{1}{2}\sum_{ns}\sum_{n\omega}\|\mathbf{d}_{obs}(\mathbf{x}_s,\omega) - \mathbf{d}_{syn}(\mathbf{m},\mathbf{x}_s,\omega)\|^2 = \frac{1}{2}\sum_{ns}\sum_{n\omega}\delta\mathbf{d}^T\delta\mathbf{d}^*, \tag{3}$$

with $\mathbf{d}_{syn}(\mathbf{m},\mathbf{x}_s,\omega) = \mathbf{R}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ is the synthetic data generated using the current model $\mathbf{m}$, $\mathbf{R}$ is the sampling matrix that sampling the wavefield from the whole space to the receivers' locations, and $\mathbf{d}_{obs}(\mathbf{x}_s,\omega)$ is the observed data, ^{T} is the transpose operator and * is the conjugate operator.

The gradient method

Gradient based methods are usually used in solving the optimization problem, so that the subsurface parameters can be updated iteratively along a descent direction, which is the opposite direction of the gradient, with a step length as

$$\mathbf{m}^{(n+1)} = \mathbf{m}^{(n)} - \alpha^n\mathbf{g}^{(n)}. \tag{4}$$

The gradient of the misfit function is

$$\mathbf{g} = \frac{\partial\phi(\mathbf{m})}{\partial\mathbf{m}} = -\mathrm{Re}\left\{\sum_{ns}\sum_{n\omega}\left(\left(\frac{\partial\left(\mathbf{R}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)}{\partial\mathbf{m}}\right)^T\delta\mathbf{d}^*\right)\right\}, \tag{5}$$

where $\frac{\partial(\mathbf{R}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega))}{\partial\mathbf{m}}$ is the Fréchet derivative matrix $\mathbf{J}$, and real part of the complex-valued vector is taken to ensures the gradient of the misfit function remains real. To calculate the gradient, the Frechet derivative matrix is needed. Taking the partial derivative of the wave equation with respect to the model parameter gives the relation

$$\mathbf{A}(\mathbf{m},\omega)\frac{\partial\left(\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)}{\partial\mathbf{m}} = -\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega), \tag{6}$$

Which shows that the first-order partial derivative of the wavefield $\frac{\partial(\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega))}{\partial\mathbf{m}}$ can be obtained by solving the wave equation with a virtual source

$$\mathbf{f}^g = -\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega), \tag{7}$$

For each model position and model parameter class. The radiation pattern for each parameter class is included in the virtual source, and the calculation of this radiation pattern is depend on the details of the finite approximation method used in the forward modelling (see Appendix A for detail discussion). This also indicates that for each parameter class, the first-order partial derivative of the wavefield with respect to each model position can be interpreted as the wavefield $\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ scattered by a small perturbation of the parameter at this position.

Substituting the virtual source back to the gradient, the gradient becomes

$$\begin{aligned}\mathbf{g} &= -\mathrm{Re}\left\{\sum_{ns}\sum_{n\omega}\left(\mathbf{R}\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\left(-\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)\right)^T\delta\mathbf{d}^*\right)\right\}\\ &= \mathrm{Re}\left\{\sum_{ns}\sum_{n\omega}\left(\mathbf{u}^T(\mathbf{m},\mathbf{x}_s,\omega)\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\right)^T\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\right)^T\mathbf{R}^T\delta\mathbf{d}^*\right)\right\}, \quad (8)\end{aligned}$$

the conjugate of the data residual $\delta\mathbf{d}^*$ in the frequency domain is equivalent to the data time reversed in time domain, and the data residual is back project to the whole space using the operator $\mathbf{R}^T$, before it is propagated back to the subsurface by the operator $\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\right)^T$. Since only the real part of the complexed-valued vector is taken to obtain the gradient, the gradient can be calculated using the adjoint (conjugate transpose) of impedance matrix,

$$\begin{aligned}\mathbf{g} &= \mathrm{Re}\left\{\sum_{ns}\sum_{n\omega}\left(\mathbf{u}^\dagger(\mathbf{m},\mathbf{x}_s,\omega)\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\right)^\dagger\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\right)^\dagger\mathbf{R}^\dagger\delta\mathbf{d}\right)\right\}\\ &= \mathrm{Re}\left\{\sum_{ns}\sum_{n\omega}\left(\mathbf{u}^\dagger(\mathbf{m},\mathbf{x}_s,\omega)\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\right)^\dagger\lambda(\mathbf{m},\omega)\right)\right\}, \quad (9)\end{aligned}$$

where $\lambda(\mathbf{m},\omega)$ is the adjoint variable/adjoint state that is the solution of the adjoint equation (Fichtner et al., 2006a,b; Plessix, 2006)

$$\left(\mathbf{A}(\mathbf{m},\omega)\right)^\dagger\lambda(\mathbf{m},\omega) = \mathbf{R}^\dagger\delta\mathbf{d}, \quad (10)$$

where † stands for the adjoint operator, and $\mathbf{R}^\dagger = \mathbf{R}^T$ since $\mathbf{R}$ is real defined.

The gradient of the misfit function is the multiplication between the first-order partial derivative of the wavefield, sampled at the receiver points, and the data residuals. Therefore, the model perturbation predicted by the gradient is constructed through the contribution of each parameter class at every position in the model space with the data residuals, which is assumed contains only the first-order scattered events. This predicted perturbation could be incorrect when multiscattered events are presented in the data residuals and wrongly accounted with the first-order partial derivative wavefields. In multiparameter FWI, cross-talk between different parameters could be an essential problem as well, since model perturbation can be wrongly predicted, when similar radiation patterns for different parameter classes could provide similar predictions for different parameter classes in the gradient. Involving the Hessian operator (second-order partial derivative of the misfit function) in calculating the model perturbation can help mitigating these issues, which will require the application of the Newton-based methods to solve the optimization problem. In the following section, we will study the application of the Hessian operator to better incorporate both multiscattering and multiparameter in FWI.

Gauss-Newton and Newton method

By expanding the misfit function as a Taylor series up to the second order

$$\phi(\mathbf{m}+\delta\mathbf{m}) = \phi(\mathbf{m}) + \delta\mathbf{m}^T\mathbf{g} + \frac{1}{2}\delta\mathbf{m}^T\mathbf{H}\delta\mathbf{m} + O\left(\|\delta\mathbf{m}\|^3\right), \quad (11)$$

where $\mathbf{H} = \frac{\partial^2 \phi(\mathbf{m})}{\partial \mathbf{m}^2}$ is the second-order partial derivatives of the misfit function, or to say the Hessian operator. Then a perturbation $\delta\mathbf{m}$ is found to minimum of the misfit function under the quadratic approximation

$$\begin{aligned} \mathbf{g} &= -\mathbf{H}\delta\mathbf{m}, \quad \text{or} \\ \delta\mathbf{m} &= -\mathbf{H}^{-1}\mathbf{g}. \end{aligned} \tag{12}$$

The Hessian can be written as (Pratt et al., 1998)

$$\begin{aligned} \mathbf{H} = \frac{\partial^2 \phi(\mathbf{m})}{\partial \mathbf{m}^2} &= -\text{Re}\left\{ \sum_{ns}\sum_{n\omega} \left(\frac{\partial \mathbf{J}^T}{\partial \mathbf{m}^T} (\underbrace{\delta\mathbf{d}^* \dots \delta\mathbf{d}^*}_{nm\times npar}) - \mathbf{J}^T\mathbf{J}^* \right) \right\} \\ &= \underbrace{-\text{Re}\left\{ \sum_{ns}\sum_{n\omega} \left(\frac{\partial \mathbf{J}^\dagger}{\partial \mathbf{m}} (\underbrace{\delta\mathbf{d} \dots \delta\mathbf{d}}_{nm\times npar}) \right)^\dagger \right\}}_{\mathbf{H}_1} + \underbrace{\text{Re}\left\{ \sum_{ns}\sum_{n\omega} \mathbf{J}^\dagger\mathbf{J} \right\}}_{\mathbf{H}_2}, \end{aligned} \tag{13}$$

which contains two terms, the first term $\mathbf{H}_1$ contains the second-order derivatives of the wavefield, and when it is neglected, the second term $\mathbf{H}_2$ becomes the Gauss-Newton approximation of the Hessian operator.

For multiparamter FWI, the Hessian is a large block matrix, with each block the second-order derivative of the misfit function respect to parameter class $\mathbf{m}_i$ and $\mathbf{m}_j$

$$\begin{aligned} \mathbf{H}_{ij} &= \frac{\partial^2 \phi(\mathbf{m})}{\partial \mathbf{m}_i \partial \mathbf{m}_j} \\ &= -\text{Re}\left\{ \sum_{ns}\sum_{n\omega} \left(\left(\left(\frac{\partial^2 \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j} \right)^\dagger \mathbf{R}^\dagger \delta\mathbf{d} \right)^\dagger \right. \right. \\ &\quad \left. \left. - \left(\frac{\partial \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)}{\partial \mathbf{m}_i} \right)^\dagger \mathbf{R}^\dagger \mathbf{R} \left(\frac{\partial \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)}{\partial \mathbf{m}_j} \right) \right) \right\}, \end{aligned} \tag{14}$$

and it is a $nm \times nm$ matrix. Obviously, in the monoparameter case, the Hessian only contains one block, which is second-order derivative of the misfit function respect to the parameter at each position.

The first term in the Hessian $\mathbf{H}_1$ is the multiplication between the second-order partial derivatives of the wavefield recorded at the receiver location and the data residual. Similar to the first-order partial derivatives, which are related to the first-order scattering, the second-order partial derivatives are related to the second-order scattering. To calculate this term, $\frac{\partial^2 \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j}$ is needed, and by taking the second-order partial derivative of the wave equation

$$\begin{aligned} &\mathbf{A}\,(\mathbf{m}, \omega) \frac{\partial^2 \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j} + \frac{\partial \mathbf{A}(\mathbf{m}, \omega)}{\partial \mathbf{m}_j} \frac{\partial \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)}{\partial \mathbf{m}_i} \\ &= -\frac{\partial^2 \mathbf{A}(\mathbf{m}, \omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j} \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega) - \frac{\partial \mathbf{A}(\mathbf{m}, \omega)}{\partial \mathbf{m}_i} \frac{\partial \mathbf{u}(\mathbf{m}, \mathbf{x}_s, \omega)}{\partial \mathbf{m}_j}, \end{aligned} \tag{15}$$

and the second-order partial derivatives can be obtained by solving the wave equation with a second-order virtual source,

$$\frac{\partial^2 \mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j} = \mathbf{A}^{-1}(\mathbf{m},\omega)\mathbf{f}^H,$$

with

$$\begin{aligned}\mathbf{f}^H = &-\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_j}\frac{\partial \mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial \mathbf{m}_i} - \frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i}\frac{\partial \mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial \mathbf{m}_j} \\ &- \frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega).\end{aligned} \tag{16}$$

The first two terms contain the first-order scattered wavefield, and they generates the second-order scattered wavefield. The third term in the virtual source depends on the parameterization and the details of the finite different method (Appendix B). The multiplication between the second-order partial derivatives of the wavefield and the data residual becomes,

$$\begin{aligned}
&-\Big(\frac{\partial^2 \mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j}\Big)^\dagger \mathbf{R}^\dagger \delta\mathbf{d} = \big(\mathbf{A}^{-1}(\mathbf{m},\omega)\mathbf{f}^H\big)^\dagger \mathbf{R}^\dagger \delta\mathbf{d} \\
&= \left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_j}\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\Big(-\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\Big)\right)\right)^\dagger \underbrace{\big(\mathbf{A}^{-1}(\mathbf{m},\omega)\big)^\dagger \mathbf{R}^\dagger \delta\mathbf{d}}_{\text{Adjoint wavefield}\lambda} \\
&\quad + \left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i}\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\Big(-\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\Big)\right)\right)^\dagger \big(\mathbf{A}^{-1}(\mathbf{m},\omega)\big)^\dagger \mathbf{R}^\dagger \delta\mathbf{d} \\
&\quad + \left(\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)^\dagger \big(\mathbf{A}^{-1}(\mathbf{m},\omega)\big)^\dagger \mathbf{R}^\dagger \delta\mathbf{d} \\
&= \Big(-\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\Big)^\dagger \big(\mathbf{A}^{-1}(\mathbf{m},\omega)\big)^\dagger \left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_j}\right)^\dagger \lambda(\mathbf{m},\omega) \\
&\quad + \Big(-\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\Big)^\dagger \big(\mathbf{A}^{-1}(\mathbf{m},\omega)\big)^\dagger \left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i}\right)^\dagger \lambda(\mathbf{m},\omega) \\
&\quad + \left(\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i \partial \mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)^\dagger \lambda(\mathbf{m},\omega).
\end{aligned} \tag{17}$$

With this term $\mathbf{H}_1$ in the Hessian, the inverse of the Hessian can act as a deconvolution operator including terms compensating the artifacts in the gradient generated by the second-order scattered waves.

The second term $\mathbf{H}_2$ in the Hessian is the multiplication of two first-order derivatives of the wavefield at the receiver locations with respect to different parameter class, which measures the similarity of the wavefield recorded at the receiver locations respect to different

parameter classes and different positions,

$$\left(\frac{\partial \mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial \mathbf{m}_i}\right)^{\dagger}\mathbf{R}^{\dagger}\mathbf{R}\left(\frac{\partial \mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)}{\partial \mathbf{m}_j}\right) = \left(-\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_i}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)^{\dagger}\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\right)^{\dagger}\mathbf{R}^{\dagger}\mathbf{R}\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\Big(-\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\Big)\right), \tag{18}$$

It can be seen that, for the same parameter classes $i = j$, this term reaches the maximum at same positions, which are the diagonal terms, and rapidly decreases for the off-diagonal terms. For the different parameter classes $i \neq j$, it measures the coupling between different parameter classes at different positions. Since seismic data is usually band-limited, the second term of the Hessian is diagonally dominant and banded, and it can predict 1) the defocusing caused by the incomplete and uneven illumination for the same parameter classes, 2) the coupling between different parameter classes, 3) the defocusing due to the limited bandwidth of the seismic data. Therefore, applying the inverse of this part of Hessian to filter the gradient, will help focusing the perturbations at each position for each parameter class, and also decoupling different parameter classes to mitigate the cross-talk artifacts.

Nonlinear FWI from Newton method

Explicitly calculating the Hessian and its inverse are normally avoided in large inverse problems due to the large cost, instead, methods, such as quasi-Newton method, truncated Newton method, are usually used to better account the inverse Hessian operator in the inversion. Compared to full Hessian operator, Gauss-Newton Hessian operator takes less computation costs. As pointed out in Pratt et al (1998), including the term $\mathbf{H}_1$ in the Hessian can produce significant false predictions, which may destroy the positive definition of the Hessian, and cause the misfit cannot be further reduced, since in this case, the misfit function is no locally convex anymore. However, it may still be essential to include the term $\mathbf{H}_1$ in the Hessian to better handle the multiscattered energy in the data residuals. It is then naturally to find a better way to incorporate this term, or to say adding more nonlinearity in the descent direction, in the FWI scheme. In our former study, in the monoparameter case of inverting squared-slowness, instead of directly using the inverse of the full Hessian to correct the gradient, we include the double scattered energy in the gradient, and showed the convergence can be improved. In this study, we will extend this method to the multiparameter case, which can be derived from Newton method and seen as an approximate version of Newton method.

Rewrite the full Newton inversion equation

$$\begin{aligned}
&\big(\mathbf{H}_1+\mathbf{H}_2\big)\delta\mathbf{m} = -\mathbf{g},\\
&\big(\mathbf{H}_2^{-1}\mathbf{H}_1+\mathbf{I}\big)\delta\mathbf{m} = -\mathbf{H}_2^{-1}\mathbf{g},\\
&\delta\mathbf{m} = -\big(\mathbf{H}_2^{-1}\mathbf{H}_1+\mathbf{I}\big)^{-1}\mathbf{H}_2^{-1}\mathbf{g},\\
&\delta\mathbf{m} = -\big(\mathbf{I}-\mathbf{H}_2^{-1}\mathbf{H}_1+(\mathbf{H}_2^{-1}\mathbf{H}_1)^2-\dots\big)\mathbf{H}_2^{-1}\mathbf{g}.
\end{aligned} \tag{19}$$

We can find that instead of directly calculating the inverse of the Hessian, a series contains the term $\mathbf{H}_1$, preconditioned by the term $\mathbf{H}_2^{-1}$ can be calculated first to precondition the

perturbation, which is obtained from the Gauss-Newton method. Taking only the first two terms, we can get an approximate version of the full Newton method (Pratt et al., 1998),

$$\delta\mathbf{m} = -\mathbf{H}_2^{-1}\big(\mathbf{g} - \mathbf{H}_1\mathbf{H}_2^{-1}\mathbf{g}\big). \tag{20}$$

Suppose that $\delta\mathbf{m}_1 = -\mathbf{H}_2^{-1}\mathbf{g}$ is the inverted model perturbation from Gauss-Newton method, which can be obtained as the solution of a linearized inverse problem (Gauss-Newton inversion problem). Different implementations can be used to solve the above problem.

Approximate Newton method with subspace method

Two perturbation can be inverted using this approximate version of the full Newton method, which are related to the standard Gauss-Newton method result and second-order terms (Pratt et al., 1998)

$$\begin{aligned}\delta\mathbf{m}_1 &= -\mathbf{H}_2^{-1}\mathbf{g},\\ \delta\mathbf{m}_2 &= \mathbf{H}_2^{-1}\mathbf{H}_1\mathbf{H}_2^{-1}\mathbf{g} = -\mathbf{H}_2^{-1}\mathbf{H}_1\delta\mathbf{m}_1.\end{aligned} \tag{21}$$

Then 2D subspace method can be applied to determine an optimal descent direction with pre-calculated model perturbation $\delta\mathbf{m}_1$ and second-order term $\delta\mathbf{m}_2$

$$\delta\mathbf{m} = \alpha\delta\mathbf{m}_1 + \beta\delta\mathbf{m}_2. \tag{22}$$

In multiparameter case, the model can be divided with different parameter classes, then the 2D subspace method can be extended into $npar \times 2D$ subspace method with more flexible ability to update each parameter class. We will discuss the detail of subspace space method in a companion paper.

Nonlinear inverse method from perturbation analysis

In the perturbative analysis based iterative nonlinear inverse methods (Snieder 1990), linearized perturbation $\delta\mathbf{m}_1$ is first calculated, then using this linearized perturbation, for each shot, for every receiver location i and synthetic data $(\mathbf{Ru}(\mathbf{m},\mathbf{x}_s,\omega))_i$, first-order multiples can be calculated as

$$\delta d_i^{(2)} = \delta\mathbf{m}_1^T \frac{\partial^2\big(\mathbf{Ru}(\mathbf{m},\mathbf{x}_s,\omega)\big)_i}{\partial\mathbf{m}^T\partial\mathbf{m}}\delta\mathbf{m}_1. \tag{23}$$

Then a standard Gauss-Newton method (first-order inverse operator) can be applied to this estimated multiples, and finally obtain the perturbation as

$$\begin{aligned}\delta\mathbf{m} &= \delta\mathbf{m}_1 + \mathbf{H}_2^{-1}\mathrm{Re}\left\{\Big(\frac{\partial\big(\mathbf{Ru}(\mathbf{m},\mathbf{x}_s,\omega)\big)}{\partial\mathbf{m}}\Big)^{\dagger}\mathbf{R}^T\delta\mathbf{d}^{(2)}\right\}\\ &= \mathbf{H}_2^{-1}\mathrm{Re}\left\{\Big(\frac{\partial\big(\mathbf{Ru}(\mathbf{m},\mathbf{x}_s,\omega)\big)}{\partial\mathbf{m}}\Big)^{\dagger}\mathbf{R}^T\big(\delta\mathbf{d} - \delta\mathbf{d}^{(2)}\big)\right\}.\end{aligned} \tag{24}$$

The update is then equivalent to remove the second-order scattering in the data residuals, and use these new data residuals to build up a more linearized inverse problem, which can be solved using the standard Gauss-Newton method.

Approximate Newton method with nonlinear gradient/descent direction

In this study, we are more focused on investigating the method by adding nonlinearity into the descent direction. Rewrite equation (20) as

$$\delta\mathbf{m} = -\mathbf{H}_2^{-1}\big(\mathbf{g} + \mathbf{H}_1\delta\mathbf{m}_1\big), \tag{25}$$

which can be considered as another linearized inverse problem with a new descent direction $\mathbf{g}+\mathbf{H}_1\delta\mathbf{m}_1$, or to say including the second-order scattering caused by the perturbation from a linearized inverse problem in the original gradient.

Substituting $\mathbf{H}_1$ as in equation (17) into the perturbation (20),

$$\begin{aligned}\delta\mathbf{m} = -\mathbf{H}_2^{-1}\mathrm{Re}\Bigg\{&\sum_{ns}\sum_{n\omega}\Bigg(\big(\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\big)^{\dagger}\Big(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\Big)^{\dagger}\lambda(\mathbf{m},\omega)\\
&+\Big(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\Big)^{\dagger}\Bigg(\big(\mathbf{A}^{-1}(\mathbf{m},\omega)\big)^{\dagger}\Bigg(-\sum_{j=1}^{npar}\Big(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}_j}\Big)^{\dagger}\lambda(\mathbf{m},\omega)\delta\mathbf{m}_{1j}\Bigg)\Bigg)\\
&+\Bigg(\Big(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\Big)^{\dagger}\lambda(\mathbf{m},\omega)\Bigg)^{\dagger}\Bigg(\mathbf{A}^{-1}(\mathbf{m},\omega)\Big(-\sum_{j=1}^{npar}\Big(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\delta\mathbf{m}_{1j}\Big)\Big)\Bigg)\\
&+\sum_{j=1}^{npar}\Bigg(\Bigg(\Big(\frac{\partial^2\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}\partial\mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\Big)^{\dagger}\lambda(\mathbf{m},\omega)\Bigg)^{\dagger}\delta\mathbf{m}_{1j}\Bigg)\Bigg)\Bigg\}.\end{aligned} \tag{26}$$

In the monoparameter case, we can find out that it is consistent with the nonlinear descent direction we have studied based on the scattering theory (Geng et al., 2017, see details in Appendix C). Using this approximate version of the full Newton method, by including the double scattered terms into the gradient, we form a new descent direction, which can be used to better correlate to the multi-scattered energy in the data residuals than the Newton method. In this update, data residual generated from the current model (prediction/true perturbation) is used to calculate the adjoint wavefield $\lambda(\mathbf{m},\omega)$, and a perturbation $\delta\mathbf{m}_1$ is obtained first to calculate the new descent direction.

Efficient calculation of approximate Newton method with Hessian-vector product

Explicitly calculating both terms of the Hessian operator in the approximate Newton method is still not possible for the large size of FWI problem, especially in the multiparameter case. In the approximate Newton methods we discussed in the former section, we can see that, the calculation of a product between the Hessian $\mathbf{H}_1$ and a vector is needed to first calculate the second-order scattering related perturbation, and the inverse of $\mathbf{H}_2$ can be added to the perturbation iteratively by a truncated Gauss-Newton method, with the help of a Gauss-Newton Hessian-vector product, while in the nonlinear inverse method derived from the perturbation analysis, the second-order scattered data is predicted before using the linearized inversion to estimate the related perturbation.

As in equation (26), the product of the first term $\mathbf{H}_1$ with a model vector $\mathbf{m}$ is

$$\mathbf{H}_1\mathbf{m} = \mathrm{Re}\left\{\sum_{ns}\sum_{n\omega}\left(\sum_{j=1}^{npar}\left(\left(\left(\frac{\partial^2\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}\partial\mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)^{\dagger}\lambda(\mathbf{m},\omega)\right)^{\dagger}\mathbf{m}_j\right)\right.$$
$$+\left(\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\right)^{\dagger}\lambda(\mathbf{m},\omega)\right)^{\dagger}\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\left(-\sum_{j=1}^{npar}\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\mathbf{m}_j\right)\right)\right)$$
$$\left.+\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)^{\dagger}\left(\left(\mathbf{A}^{-1}(\mathbf{m},\omega)\right)^{\dagger}\left(-\sum_{j=1}^{npar}\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}_j}\right)^{\dagger}\lambda(\mathbf{m},\omega)\mathbf{m}_j\right)\right)\right)\right\}, \tag{27}$$

which can be further written as

$$\mathbf{H}_1\mathbf{m} = \mathrm{Re}\left\{\sum_{ns}\sum_{n\omega}\left(\sum_{j=1}^{npar}\left(\left(\left(\frac{\partial^2\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}\partial\mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)^{\dagger}\lambda(\mathbf{m},\omega)\right)^{\dagger}\left(\mu_1(\mathbf{m},\omega)\right)_j\right)\right.$$
$$\left.+\left(\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\right)^{\dagger}\lambda(\mathbf{m},\omega)\right)^{\dagger}\mu_2(\mathbf{m},\omega)+\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)^{\dagger}\mu_3(\mathbf{m},\omega)\right)\right\}, \tag{28}$$

where $\lambda(\mathbf{m},\omega)$, $\mu_1(\mathbf{m},\omega)$, $\mu_2(\mathbf{m},\omega)$ and $\mu_3(\mathbf{m},\omega)$are the adjoint variables, $\lambda(\mathbf{m},\omega)$ is the solution of the adjoint equation as in the calculation of the gradient as in (10), and $\mu_1(\mathbf{m},\omega)$, $\mu_2(\mathbf{m},\omega)$, $\mu_3(\mathbf{m},\omega)$ are the solutions of the adjoint equations

$$\begin{cases}\mu_1(\mathbf{m},\omega)=\mathbf{m}\\ \mathbf{A}(\mathbf{m},\omega)\mu_2(\mathbf{m},\omega)=-\sum_{j=1}^{npar}\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\mathbf{m}_j\\ \left(\mathbf{A}(\mathbf{m},\omega)\right)^{\dagger}\mu_3(\mathbf{m},\omega)=-\sum_{j=1}^{npar}\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}_j})\right)^{\dagger}\lambda(\mathbf{m},\omega)\mathbf{m}_j.\end{cases} \tag{29}$$

The Hessian-vector product for the second term $\mathbf{H}_2$ of the Hessian operator, which is used in the truncated Gauss-Newton method to solve the linearized inverse problem (12) with Gauss-Newton Hessian by the conjugate gradient method, is reduced into a much more simple form as

$$\mathbf{H}_2\mathbf{m} = \mathrm{Re}\left\{\sum_{ns}\sum_{n\omega}\left(\left(\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\right)^{\dagger}\mu_3(\mathbf{m},\omega)\right)\right\}, \tag{30}$$

with

$$\begin{cases}\mathbf{A}(\mathbf{m},\omega)\mu_2(\mathbf{m},\omega)=-\sum_{j=1}^{npar}\frac{\partial\mathbf{A}(\mathbf{m},\omega)}{\partial\mathbf{m}_j}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)\mathbf{m}_j\\ \left(\mathbf{A}(\mathbf{m},\omega)\right)^{\dagger}\mu(\mathbf{m},\omega)=-\mathbf{R}^{\dagger}\mathbf{R}\mu_2(\mathbf{m},\omega).\end{cases} \tag{31}$$

It can be seen that to obtain the Hessian-vector (28) and (30), besides the forward modelling wavefield $\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ and the adjoint wavefield $\lambda(\mathbf{m},\omega)$, which are also needed

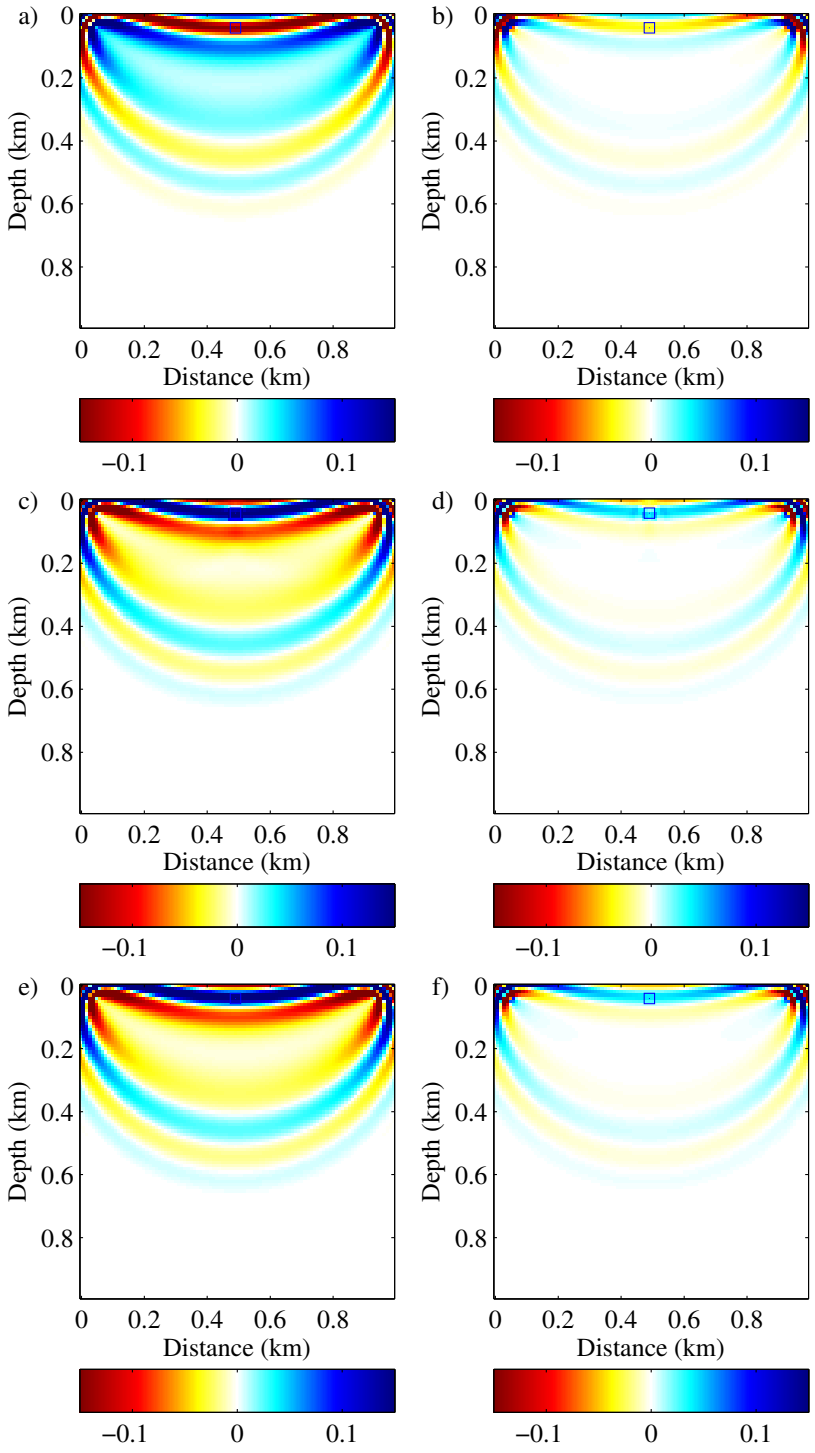
in the calculation of the gradient, two more forward modeling problems are needed to obtain the virtual sources and their related wavefields. Therefore, compared to the truncated Gauss-Newton/Newton method, about twice amounts of the wave equations are needed to be solved in the approximate Newton with nonlinear descent direction.

EXAMPLE

Hessian vector

We will first show the gradient and Hessian-vectors in a constant model with one small scatterer for one source and receiver pair. The model is 1×1 km with grid intervals of 10 m, and the background model is linearly changed along z direction for both velocity and density. The source and receiver are located near to the surface. Ricker wavelet with dominant frequency 15 Hz is used as the source wavelet, and the data is recorded for 1 s with 1 ms sampling. 44 frequencies from 2 Hz to 45 Hz are used to generate the results in the frequency domain. The perturbation is chosen to be 10% of the background model, and is located within a 40×40 m box-shaped area at $x = 0.5$ km. First, we consider velocity perturbation only, as shown in Figure 1-3 the related gradient and Hessian-vectors with perturbations in different depths. It can be observed that, when the data residual is related to the velocity perturbation locating inside the first Fresnel zone, the gradient for velocity has much larger values compared to the gradient for density, which is almost zero, indicating that density perturbations hardly response to first-order scattering at wide scattering angles; on the contrary, when the data residual is related to velocity perturbation locating outside the first Fresnel zone in the deeper region (e.g., corresponds to smaller scattering angles), the gradients for velocity and density are comparable, indicating possible parameter crosstalk. When second-order scattering is considered, the Hessian-vectors calculated using the true perturbation and different part of the Hessian operators are also shown in Figure 1-3. The second-order scattering related Hessian-vectors contain two symmetric lobes for both velocity and density. Compared to the amplitude of Gauss-Newton Hessian-vectors, the amplitude of corresponding second-order scattering related Hessian-vectors is around one order of magnitude smaller. The contribution of the second-order scattering to the full Hessian can hardly be seen in the shallow perturbation case (Figure 1), but are clearly visible in other two cases (Figure 2 and Figure 3). The amplitude of Hessian-vectors for density are small compared to the amplitude of Hessian-vectors for velocity, which indicates a weak second-order scattering from perturbations of different parameter classes.

We then study the density perturbation, Figure 4 shows the gradient and related Hessian-vectors when the density perturbation inside the first Fresnel zone. Compared to the results related to velocity perturbation in the same position (Figure 2), the gradient for both density and velocity are much smaller, and the second-order scattering related Hessian-vectors are nearly zero, which again indicate the difficulty of updating long wavelength component of the density, as well as the very weak second-order scattering related to the density perturbations.


a)
b)
c)
d)
e)
f)
Depth (km)
Distance (km)
0
0.2
0.4
0.6
0.8
−0.1
0
0.1

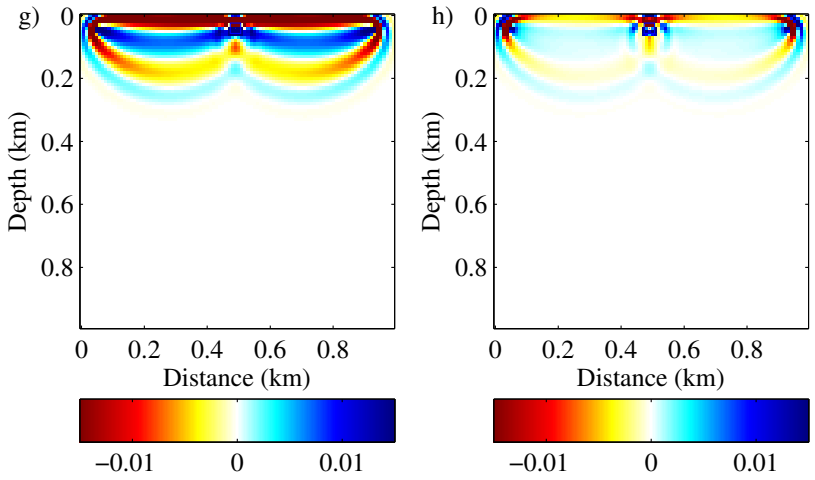
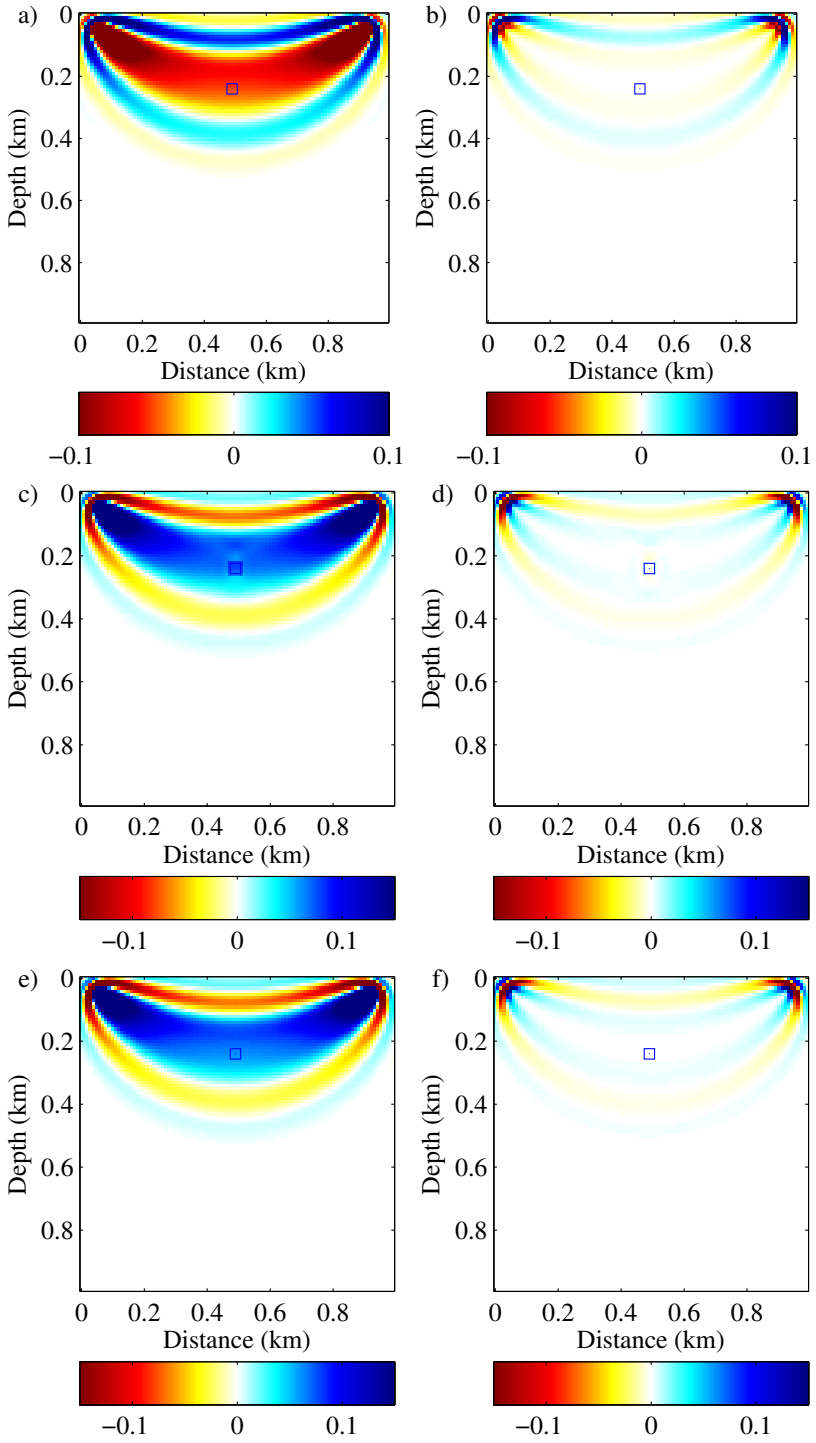


FIG. 1. Gradient and Hessian-vectors for velocity (left column) and density (right column) in a gz media with one small velocity perturbation (indicated with blue box) close to the surface. From top to bottom are gradient, full Hessian-vector, Gauss-Newton Hessian-vector and Hessian-vector with the first term in Hessian (continued).

The Marmousi-II model

The Marmousi-II model is an extension and elastic upgrade of the classic Marmousi model. In this study, we only use part of the original vp and density model and the resize the model into 141×321 grid nodes, with grid intervals of 25 m. The resulting vp and density model is shown as in Figure 5a and 5b, respectively. We use a 64 sources and 321 receivers along the surface, and the water layer is kept constant throughout the iterations. The initial model is obtained by smoothing the exact model using Gaussian smoothing, shown in Figure 5c and 5d. Since the initial model is only a slightly smoothed version of the true model, we simultaneously invert 4 frequencies (3 Hz, 5 Hz, 8 Hz, 12 Hz). A maximum of 20 external iterations and 10 internal iterations are performed for each method. For the approximated Newton method with nonlinear descent direction, 2 internal iterations are used for the truncated Gauss-Newton method to obtain the perturbation before applying the $\mathbf{H}_1$ operator.

The final inverted model using truncated Gauss-Newton, truncated Newton and approximate Newton method with nonlinear descent direction are shown as in Figure 6. Velocity and density profiles at $x = 2.5$ km and $x = 5$ km are shown in Figure 7 and Figure 8. The velocity error, density error, as well as misfit functions with respect to the number of iterations are shown in Figure 9 for all three methods. It can be seen that, truncated Newton method (Figure 6a and 6b) updates both velocity and density; truncated Gauss-Newton method (Figure 6c and 6d) provides the best update for the velocity, but the update of density is overestimated due to the cross-talk. The approximate Newton method with nonlinear descent direction (Figure 6e and 6f) can provide slightly better results for both velocity and density compared to truncated Newton method, and compared to truncated Gauss-Newton method, the update of density has less cross-talk artifacts, e.g., density profile at $x = 2.5$ km between $z = 1.7$ km and 3 km (Figure 7), and density profile at $x = 5$ km between $z = 2$ km and 2.8 km (Figure 8).


a)
b)
c)
d)
e)
f)
Depth (km)
Distance (km)
0
0.2
0.4
0.6
0.8
−0.1
0
0.1

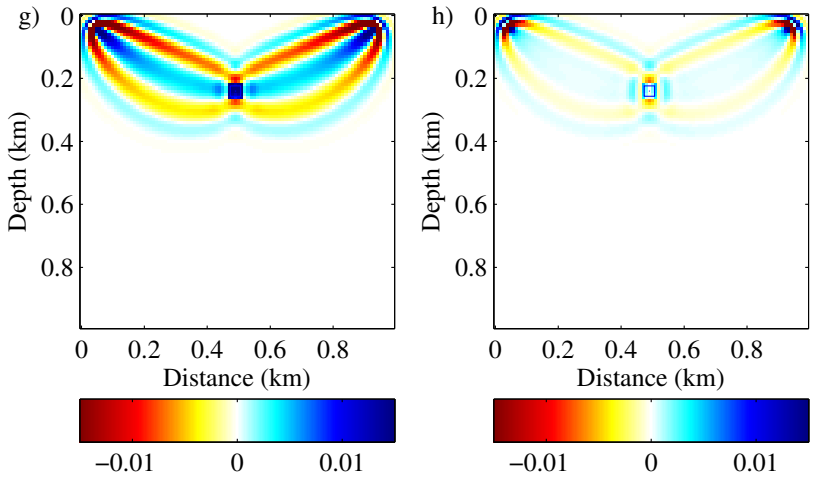


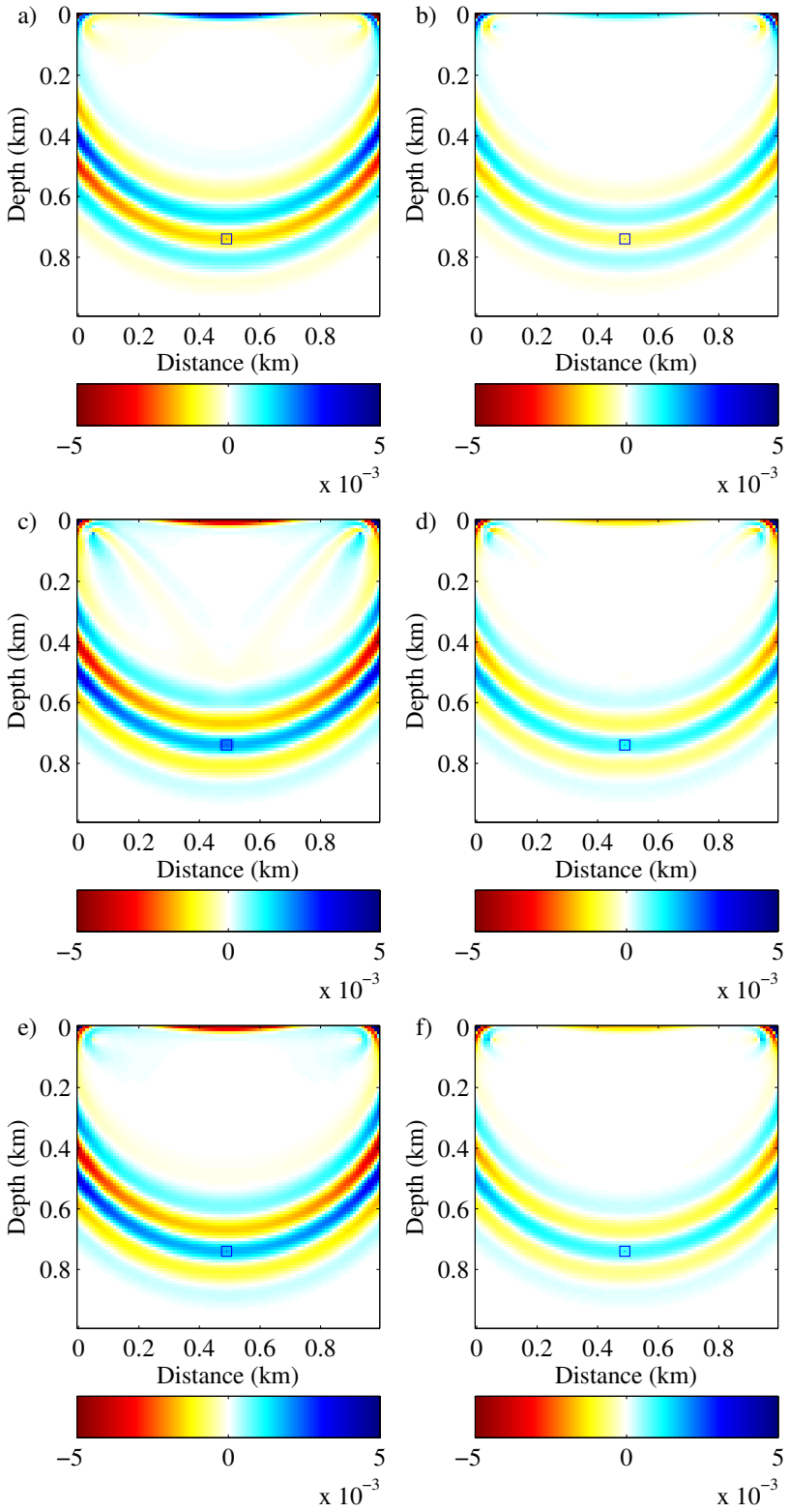
FIG. 2. Gradient and Hessian-vectors for velocity (left column) and density (right column) in a gz media with one small velocity perturbation (indicated with blue box) inside the first Fresnel zone. From top to bottom are gradient, full Hessian-vector, Gauss-Newton Hessian-vector and Hessian-vector with the first term in Hessian.

CONCLUSIONS

The Hessian operator plays a very important role in FWI, especially in the multiparameter case. The inverse of Gauss-Newton Hessian operator helps refocusing the gradient and mitigating the cross-talk artifacts between different parameter classes, and the additional term in the Hessian contains information related to the second-order scattering. However, directly using full Hessian as in Newton method could cause incorrect prediction of the model perturbation, since the second-order scattering energy may be incorrectly to correlate the first-order scattering energy in the gradient, and the positive definiteness of the Gauss-Newton Hessian could be destroyed, which will affect the predict model parameters in a negative way. In this study, we derive a nonlinear descent direction from the Newton method to perform a multiparameter FWI, which alters the gradient with the second-order scattering, and can be used to predict the model perturbation using a truncated Gauss-Newton method.

ACKNOWLEDGMENT

We thank the sponsors of CREWES for support. This work was funded by CREWES and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 379744-08.


a)
b)
c)
d)
e)
f)
Depth (km)
Distance (km)
0
0.2
0.4
0.6
0.8
−5
0
5
x 10^−3

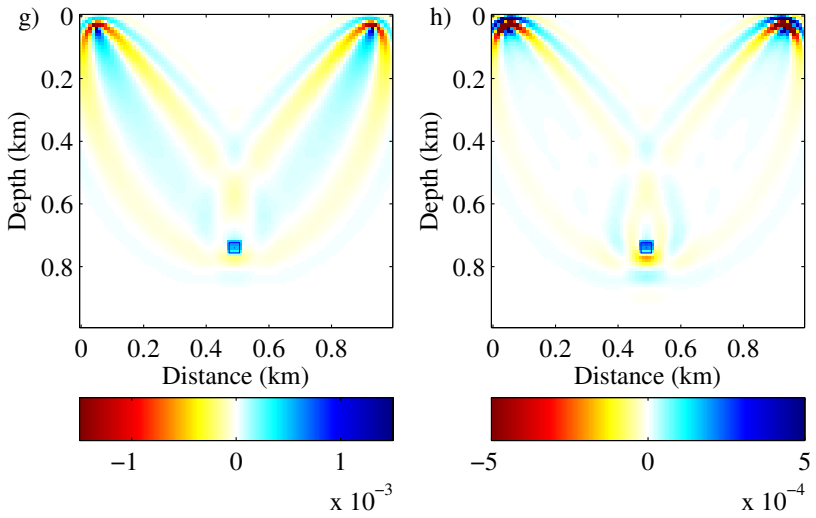
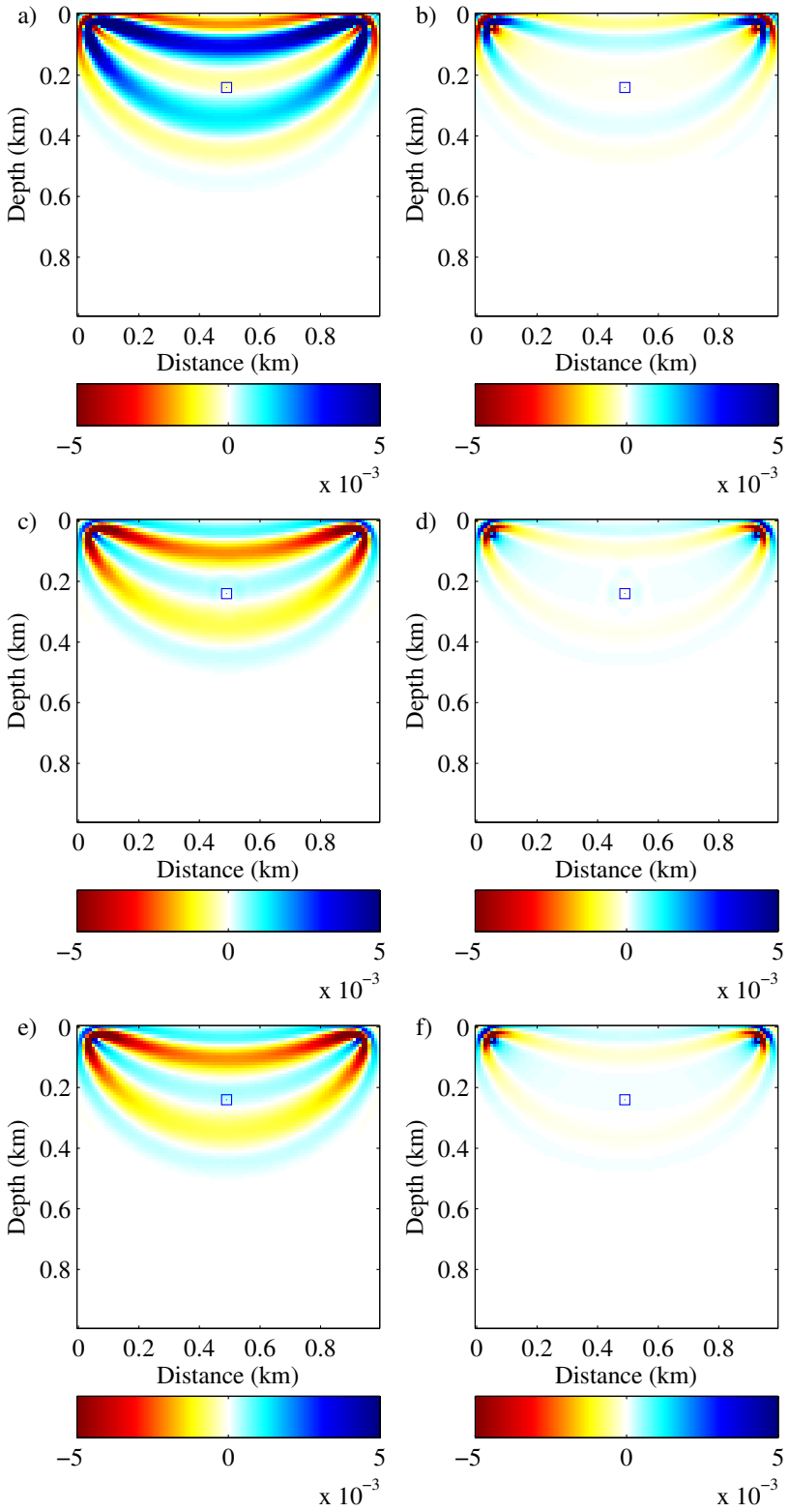


FIG. 3. Gradient and Hessian-vectors for velocity (left column) and density (right column) in a gz media with one small velocity perturbation (indicated with blue box) in deeper part of the model outside the first Fresnel zone. From top to bottom are gradient, full Hessian-vector, Gauss-Newton Hessian-vector and Hessian-vector with the first term in Hessian.


a)
b)
c)
d)
e)
f)
Depth (km)
Distance (km)
0
0.2
0.4
0.6
0.8
−5
0
5
x 10^−3

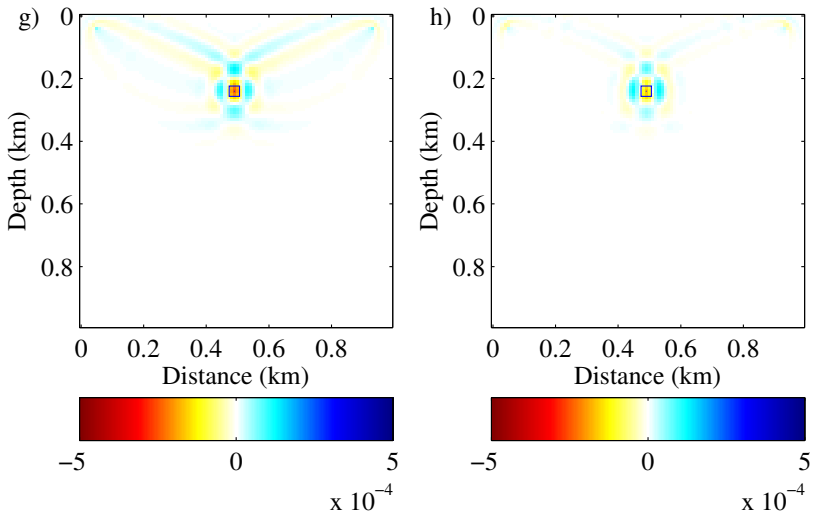


FIG. 4. Gradient and Hessian-vectors for velocity (left column) and density (right column) in a gz media with one small density perturbation (indicated with blue box) inside the first Fresnel zone. From top to bottom are gradient, full Hessian-vector, Gauss-Newton Hessian-vector and Hessian-vector with the first term in Hessian.

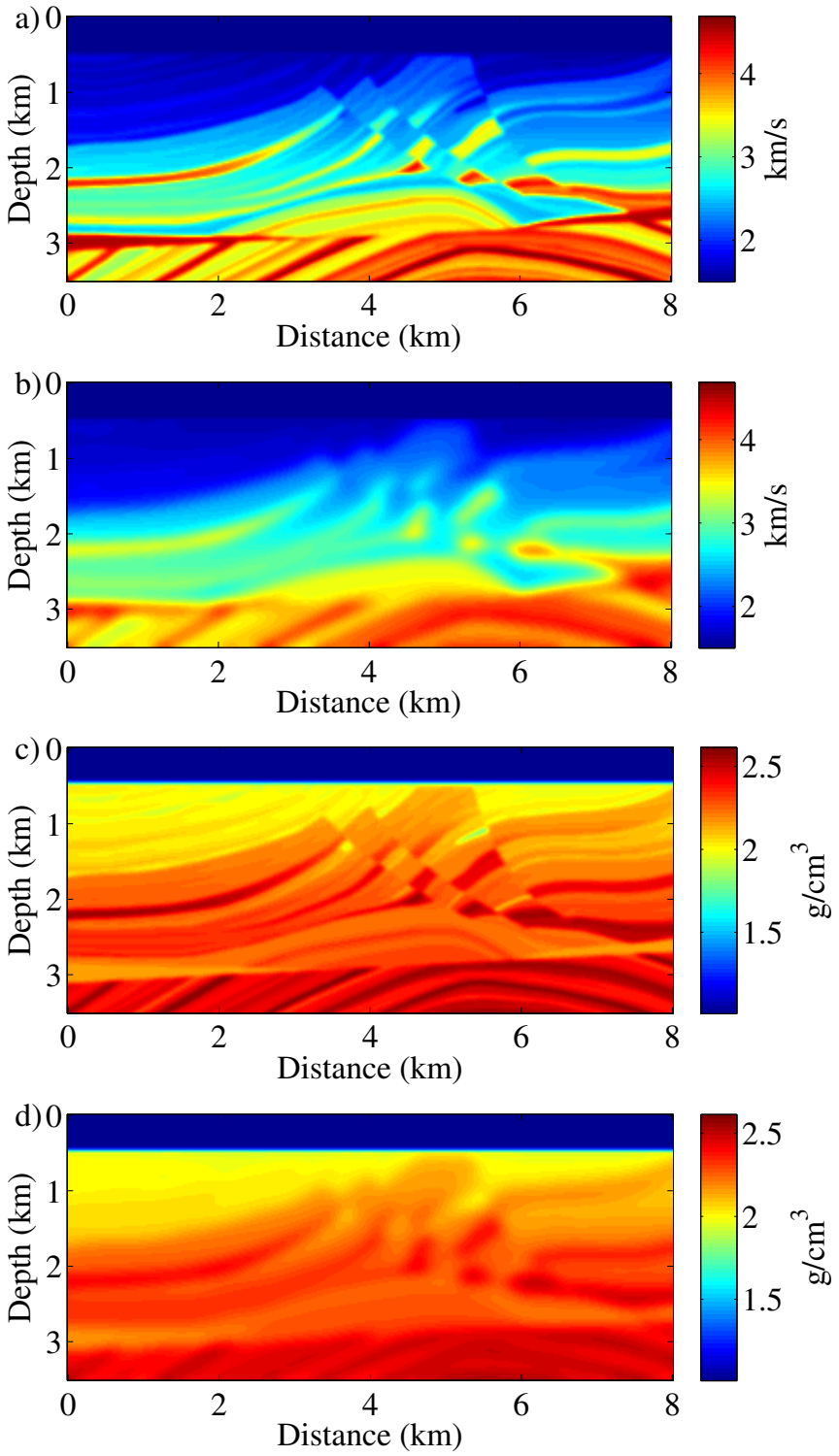
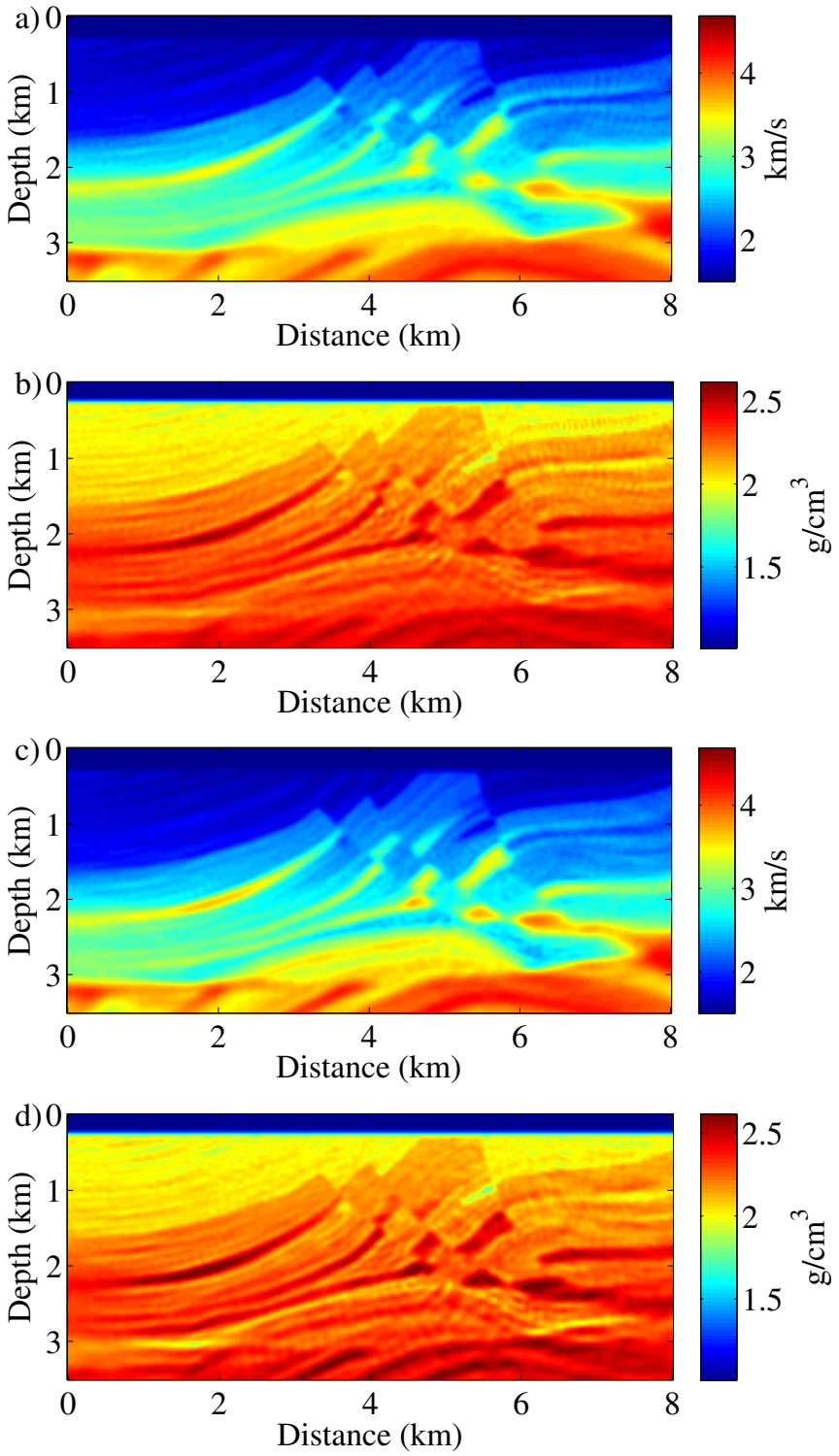


FIG. 5. Marmousi exact velocity model (a), initial velocity model (b), exact density model (c), initial density model (d).


a)
b)
c)
d)
Depth (km)
Distance (km)
0
1
2
3
0
2
4
6
8
2
3
4
km/s
1.5
2
2.5
g/cm3

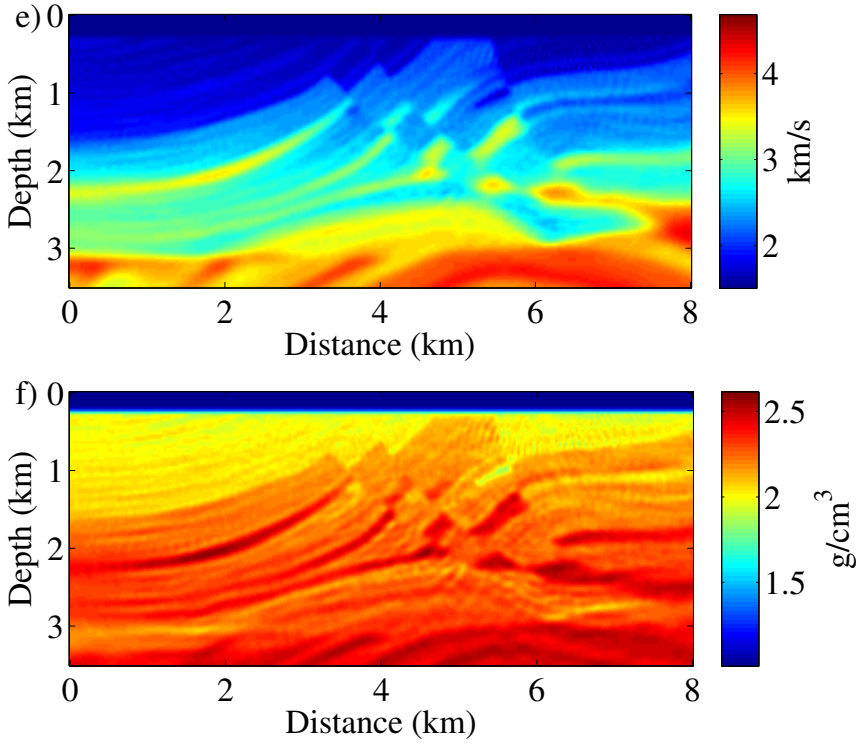


FIG. 6. Inverted velocity and density models for the Marmousi case with truncated Newton method (a) and (b), truncated Gauss-Newton method (c) and (d) and approximate Newton method with nonlinear descent direction (e) and (f).

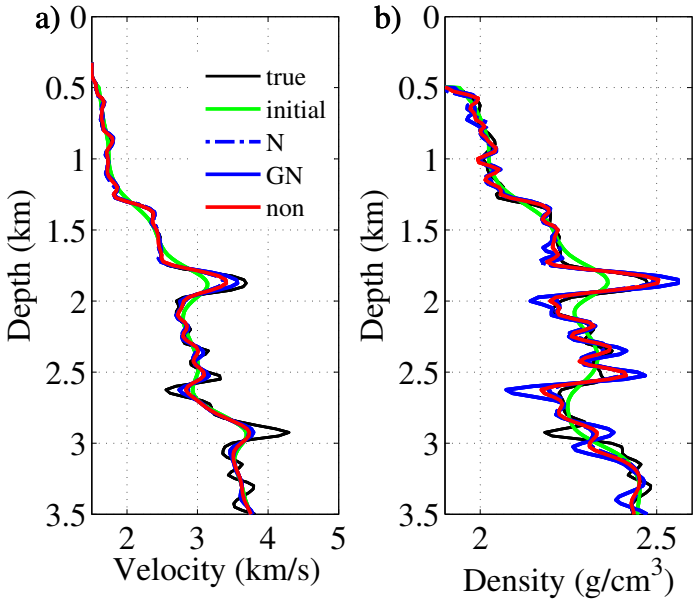


FIG. 7. Velocity profile (a) and density profile (b) at x=2.5km.

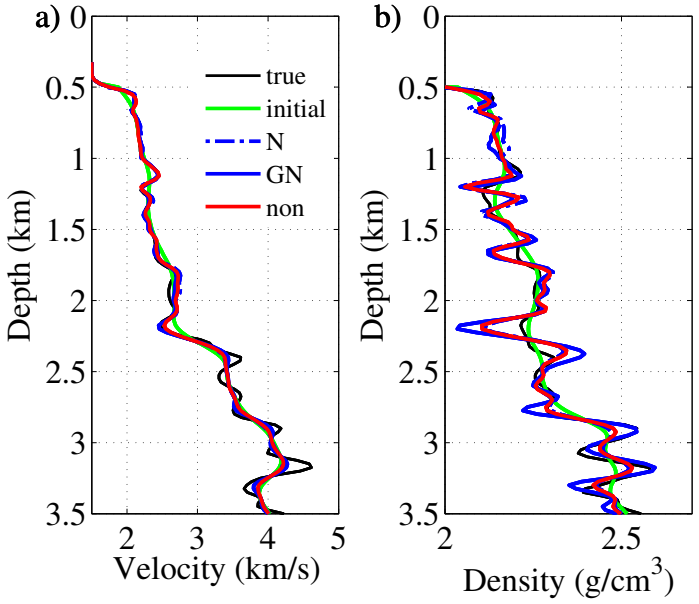


FIG. 8. Velocity profile (a) and density profile (b) at x=5km.

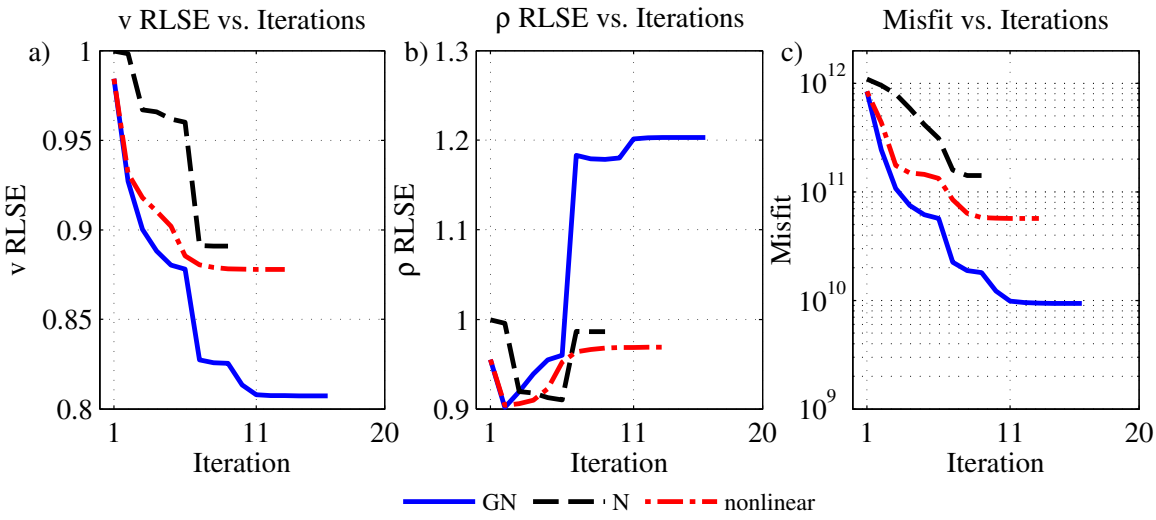


FIG. 9. Velocity profile (a) and density profile (b) at x=5km.

APPENDIX A - CALCULATION OF VIRTUAL SOURCE IN THE GRADIENT

In this appendix, we will review the calculation of the virtual source for different parameterization with a 5-point finite difference scheme in the acoustic case (Jo et al., 1996; Hustedt et al., 2004). We will first show the discretized form of the acoustic wave equation, then we will derive the virtual source of the gradient for different parameter classes from the discretized form of the wave equation. In 2D case, at each point, the wave equation is

$$\frac{\omega^2}{\kappa(x_i,z_j)}u(x_i,z_j,\omega)+\nabla\cdot\left(\frac{1}{\rho(x_i,z_j)}\nabla u(x_i,z_j,\omega)\right)=f(x_i,z_j), \tag{A-1}$$

It can be discretized with five-point finite difference schemes with perfectly matched layer (PML)

$$\begin{aligned}&\frac{\omega^2}{\kappa_{x_i,z_j}}u_{x_i,z_j}+\frac{1}{\Delta_x^2\xi_{x_i}}\left(\frac{1}{\xi_{x_{i+1/2}}}\frac{1}{\rho_{x_{i+1/2},z_j}}\Big(u_{x_{i+1},z_j}-u_{x_i,z_j}\Big)-\frac{1}{\xi_{x_{i-1/2}}}\frac{1}{\rho_{x_{i-1/2},z_j}}\Big(u_{x_i,z_j}-u_{x_{i-1},z_j}\Big)\right)\\&+\frac{1}{\Delta_z^2\xi_{z_i}}\left(\frac{1}{\xi_{z_{j+1/2}}}\frac{1}{\rho_{x_i,z_{j+1/2}}}\Big(u_{x_i,z_{j+1}}-u_{x_i,z_j}\Big)-\frac{1}{\xi_{z_{j-1/2}}}\frac{1}{\rho_{x_i,z_{j-1/2}}}\Big(u_{x_i,z_j}-u_{x_i,z_{j-1}}\Big)\right)=f_{x_i,z_j},\end{aligned} \tag{A-2}$$

where

$$\begin{aligned}\xi_x&=1+i\gamma_x(x)/\omega=1+i\frac{c_{pml}cos\left(\frac{\pi x}{2L}\right)}{\omega},\quad\text{and}\\\xi_z&=1+i\gamma_z(z)/\omega=1+i\frac{c_{pml}cos\left(\frac{\pi z}{2L}\right)}{\omega},\end{aligned} \tag{A-3}$$

are the functions extend the model with a PML layer with width L for both sides along x and z direction. The averaged coefficients on the half grid are

$$\begin{aligned}\frac{1}{\xi_{x_{i+1/2}}}&=\frac{1}{2}\Big(\frac{1}{\xi_{x_i}}+\frac{1}{\xi_{x_{i+1}}}\Big),\quad\text{and}\\\frac{1}{\rho_{x_{i+1/2},z_j}}&=\frac{1}{2}\Big(\frac{1}{\rho_{x_i,z_j}}+\frac{1}{\xi_{x_{i+1},z_j}}\Big).\end{aligned} \tag{A-4}$$

Therefore, the impedance matrix $\mathbf{A}$ in equation (6) is a banded diagonal matrix with five non-zero diagonals, and each coefficient for each point is determined as in equation (A-2).

To calculate the virtual sources for each parameter of different parameterizations, we need to calculate the partial derivatives of the impedance operator with respect to the perturbations first. Here we consider 3 different parameterizations for the acoustic media: a) bulk modulus and density (κ,ρ), b) velocity and density (v,ρ), c) slowness and buoyancy s,b. From the discretized wave equation (A-2), it is obvious that for the parameter on each point (x_i,z_j) (we will use (i,j) for short in this appendix), bulk modulus and also its related parameter velocity and slowness contribute only to the coefficient of this point in

the impedance matrix, however, density ρ or buoyancy contributes to the coefficients of all five neighbour points used in the finite difference. Take the partial derivative of the wave equation respect to the model parameter outside the PML boundaries, for bulk modulus and density, we can get the element for the virtual sources for each point

$$\frac{\partial A_{i,j}}{\partial \kappa_{i,j}} u_{i,j} = -\frac{\omega^2}{\kappa_{i,j}^2} u_{i,j},$$

and

$$\begin{aligned}
\frac{\partial A_{i,j}}{\partial \rho_{i,j}} u_{i,j} &= -\frac{\Big(u_{i+1,j} - u_{i,j}\Big)}{2\Delta_x^2 \rho_{i,j}^2} + \frac{\Big(u_{i,j} - u_{i-1,j}\Big)}{2\Delta_x^2 \rho_{i,j}^2} - \frac{\Big(u_{i,j+1} - u_{i,j}\Big)}{2\Delta_z^2 \rho_{i,j}^2} + \frac{\Big(u_{i,j} - u_{i,j-1}\Big)}{2\Delta_z^2 \rho_{i,j}^2} \\
\frac{\partial A_{i+1,j}}{\partial \rho_{i,j}} u_{i,j} &= \frac{\Big(u_{i+1,j} - u_{i,j}\Big)}{2\Delta_x^2 \rho_{i,j}^2} \\
\frac{\partial A_{i-1,j}}{\partial \rho_{i,j}} u_{i,j} &= -\frac{\Big(u_{i,j} - u_{i-1,j}\Big)}{2\Delta_x^2 \rho_{i,j}^2} \\
\frac{\partial A_{i,j+1}}{\partial \rho_{i,j}} u_{i,j} &= \frac{\Big(u_{i,j+1} - u_{i,j}\Big)}{2\Delta_z^2 \rho_{i,j}^2} \\
\frac{\partial A_{i,j-1}}{\partial \rho_{i,j}} u_{i,j} &= -\frac{\Big(u_{i,j} - u_{i,j-1}\Big)}{2\Delta_z^2 \rho_{i,j}^2}.
\end{aligned} \tag{A-5}$$

For velocity and density

$$\frac{\partial A_{i,j}}{\partial v_{i,j}} u_{i,j} = -\frac{2\omega^2}{\rho_{i,j} v_{i,j}^3} u_{i,j},$$

and

$$\begin{aligned}
\frac{\partial A_{i,j}}{\partial \rho_{i,j}} u_{i,j} &= -\frac{\omega^2}{\rho_{i,j}^2 v_{i,j}^2} u_{i,j} - \frac{\Big(u_{i+1,j} - u_{i,j}\Big)}{2\Delta_x^2 \rho_{i,j}^2} + \frac{\Big(u_{i,j} - u_{i-1,j}\Big)}{2\Delta_x^2 \rho_{i,j}^2} - \frac{\Big(u_{i,j+1} - u_{i,j}\Big)}{2\Delta_z^2 \rho_{i,j}^2} + \frac{\Big(u_{i,j} - u_{i,j-1}\Big)}{2\Delta_z^2 \rho_{i,j}^2} \\
\frac{\partial A_{i+1,j}}{\partial \rho_{i,j}} u_{i,j} &= \frac{\Big(u_{i+1,j} - u_{i,j}\Big)}{2\Delta_x^2 \rho_{i,j}^2} \\
\frac{\partial A_{i-1,j}}{\partial \rho_{i,j}} u_{i,j} &= -\frac{\Big(u_{i,j} - u_{i-1,j}\Big)}{2\Delta_x^2 \rho_{i,j}^2} \\
\frac{\partial A_{i,j+1}}{\partial \rho_{i,j}} u_{i,j} &= \frac{\Big(u_{i,j+1} - u_{i,j}\Big)}{2\Delta_z^2 \rho_{i,j}^2} \\
\frac{\partial A_{i,j-1}}{\partial \rho_{i,j}} u_{i,j} &= -\frac{\Big(u_{i,j} - u_{i,j-1}\Big)}{2\Delta_z^2 \rho_{i,j}^2}.
\end{aligned} \tag{A-6}$$

And for slowness and buoyancy

$$\frac{\partial A_{i,j}}{\partial s_{i,j}}u_{i,j} = \frac{2\omega^2}{\rho_{i,j}v_{i,j}}u_{i,j} = 2\omega^2 b_{i,j}s_{i,j}u_{i,j},$$

and

$$\begin{aligned}
\frac{\partial A_{i,j}}{\partial b_{i,j}}u_{i,j} &= \omega^2 s_{i,j}^2 u_{i,j} + \frac{\left(u_{i+1,j}-u_{i,j}\right)}{2\Delta_x^2} - \frac{\left(u_{i,j}-u_{i-1,j}\right)}{2\Delta_x^2} + \frac{\left(u_{i,j+1}-u_{i,j}\right)}{2\Delta_z^2} - \frac{\left(u_{i,j}-u_{i,j-1}\right)}{2\Delta_z^2} \\
\frac{\partial A_{i+1,j}}{\partial b_{i,j}}u_{i,j} &= -\frac{\left(u_{i+1,j}-u_{i,j}\right)}{2\Delta_x^2} \\
\frac{\partial A_{i-1,j}}{\partial b_{i,j}}u_{i,j} &= \frac{\left(u_{i,j}-u_{i-1,j}\right)}{2\Delta_x^2} \\
\frac{\partial A_{i,j+1}}{\partial b_{i,j}}u_{i,j} &= -\frac{\left(u_{i,j+1}-u_{i,j}\right)}{2\Delta_z^2} \\
\frac{\partial A_{i,j-1}}{\partial b_{i,j}}u_{i,j} &= \frac{\left(u_{i,j}-u_{i,j-1}\right)}{2\Delta_z^2}.
\end{aligned} \tag{A-7}$$

Rewrite the virtual source in the gradient calculation from equation (A-6) and (A-7) into matrix form, it is equivalent to the results obtained under chain rule

$$\begin{aligned}
\mathbf{f}_v^g &= -\left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa}\frac{\partial \kappa}{\partial v} + \frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \rho}\frac{\partial \rho}{\partial v}\right)\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega) = -2\rho v\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega) \\
\mathbf{f}_\rho^g &= -\left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa}\frac{\partial \kappa}{\partial \rho} + \frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \rho}\frac{\partial \rho}{\partial \rho}\right)\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega) \\
&= -\left(v^2\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa} + \frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \rho}\right)\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega) \\
\mathbf{f}_s^g &= -\left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa}\frac{\partial \kappa}{\partial s} + \frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \rho}\frac{\partial \rho}{\partial s}\right)\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega) = 2\rho v^3\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega) \\
\mathbf{f}_b^g &= -\left(\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa}\frac{\partial \kappa}{\partial b} + \frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \rho}\frac{\partial \rho}{\partial b}\right)\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega) \\
&= -\left(-\rho^2 v^2\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa} - \rho^2\frac{\partial \mathbf{A}(\mathbf{m},\omega)}{\partial \rho}\right)\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega).
\end{aligned} \tag{A-8}$$

APPENDIX B - CALCULATION OF SECOND-ORDER DERIVATIVE OF THE IMPEDANCE MATRIX IN THE HESSIAN

To calculate the first term $\mathbf{H}_1$ of the Hessian, the term contains second-order derivative of the forward operator $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \mathbf{m}^2}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ is needed. Same as the calculation of the

virtual source in the gradient in Appendix A, this can be derived by taking second-order partial derivatives of the wave equation with respect to different parameter classes at each point. In the acoustic case, this term contains four components.

First, bulk modulus and density, it will be $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa^2}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$, $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \kappa \partial \rho}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$, $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \rho \partial \kappa}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ and $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \rho^2}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$, and each element is

$$\begin{aligned}
\frac{\partial^2 A_{i,j}}{\partial \kappa_{i,j}^2}u_{i,j} &= -\frac{2}{\kappa_{i,j}}\left(\frac{\partial A_{i,j}}{\partial \kappa_{i,j}}u_{i,j}\right) = \frac{2\omega^2}{\kappa_{i,j}^3}u_{i,j},\\
\frac{\partial^2 A_{i,j}}{\partial \kappa_{i,j}\partial \rho_{i,j}}u_{i,j} &= \frac{\partial^2 A_{i,j}}{\partial \rho_{i,j}\partial \kappa_{i,j}}u_{i,j} = 0,
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 A_{i,j}}{\partial \rho_{i,j}^2}u_{i,j} &= -\frac{2}{\rho_{i,j}}\left(\frac{\partial A_{i,j}}{\partial \rho_{i,j}}u_{i,j}\right)\\
&= \frac{\left(u_{i+1,j}-u_{i,j}\right)}{\Delta_x^2\rho_{i,j}^3} - \frac{\left(u_{i,j}-u_{i-1,j}\right)}{\Delta_x^2\rho_{i,j}^3} + \frac{\left(u_{i,j+1}-u_{i,j}\right)}{\Delta_z^2\rho_{i,j}^3} - \frac{\left(u_{i,j}-u_{i,j-1}\right)}{\Delta_z^2\rho_{i,j}^3}\\
\frac{\partial^2 A_{i+1,j}}{\partial \rho_{i,j}^2}u_{i,j} &= -\frac{2}{\rho_{i,j}}\left(\frac{\partial A_{i+1,j}}{\partial \rho_{i,j}}u_{i,j}\right) = -\frac{\left(u_{i+1,j}-u_{i,j}\right)}{2\Delta_x^2\rho_{i,j}^3}\\
\frac{\partial^2 A_{i-1,j}}{\partial \rho_{i,j}^2}u_{i,j} &= -\frac{2}{\rho_{i,j}}\left(\frac{\partial A_{i-1,j}}{\partial \rho_{i,j}}u_{i,j}\right) = \frac{\left(u_{i,j}-u_{i-1,j}\right)}{2\Delta_x^2\rho_{i,j}^3}\\
\frac{\partial^2 A_{i,j+1}}{\partial \rho_{i,j}^2}u_{i,j} &= -\frac{2}{\rho_{i,j}}\left(\frac{\partial A_{i,j+1}}{\partial \rho_{i,j}}u_{i,j}\right) = -\frac{\left(u_{i,j+1}-u_{i,j}\right)}{2\Delta_z^2\rho_{i,j}^3}\\
\frac{\partial^2 A_{i,j-1}}{\partial \rho_{i,j}^2}u_{i,j} &= -\frac{2}{\rho_{i,j}}\left(\frac{\partial A_{i,j-1}}{\partial \rho_{i,j}}u_{i,j}\right) = \frac{\left(u_{i,j}-u_{i,j-1}\right)}{2\Delta_z^2\rho_{i,j}^3}.
\end{aligned} \tag{B-1}$$

Second, for velocity and density, it will be $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial v^2}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$, $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial v \partial \rho}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$, $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \rho \partial v}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ and $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial \rho^2}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$,

$$\begin{aligned}
\frac{\partial^2 A_{i,j}}{\partial v_{i,j}^2}u_{i,j} &= -\frac{3}{v_{i,j}}\left(\frac{\partial A_{i,j}}{\partial v_{i,j}}u_{i,j}\right) = \frac{6\omega^2}{\rho_{i,j}v_{i,j}^4}u_{i,j},\\
\frac{\partial^2 A_{i,j}}{\partial v_{i,j}\partial \rho_{i,j}}u_{i,j} &= \frac{\partial^2 A_{i,j}}{\partial \rho_{i,j}\partial v_{i,j}}u_{i,j} = -\frac{1}{\rho_{i,j}}\left(\frac{\partial A_{i,j}}{\partial v_{i,j}}u_{i,j}\right) = \frac{2\omega^2}{\rho_{i,j}^2 v_{i,j}^3}u_{i,j},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial^2 A_{i,j}}{\partial \rho_{i,j}^2} u_{i,j} &= -\frac{2}{\rho_{i,j}} \left(\frac{\partial A_{i,j}}{\partial \rho_{i,j}} u_{i,j} \right) \\
&= \frac{2\omega^2}{\rho_{i,j}^3 v_{i,j}^2} u_{i,j} + \frac{\left(u_{i+1,j} - u_{i,j}\right)}{\Delta_x^2 \rho_{i,j}^3} - \frac{\left(u_{i,j} - u_{i-1,j}\right)}{\Delta_x^2 \rho_{i,j}^3} + \frac{\left(u_{i,j+1} - u_{i,j}\right)}{\Delta_z^2 \rho_{i,j}^3} - \frac{\left(u_{i,j} - u_{i,j-1}\right)}{\Delta_z^2 \rho_{i,j}^3} \\
\frac{\partial^2 A_{i+1,j}}{\partial \rho_{i,j}^2} u_{i,j} &= -\frac{2}{\rho_{i,j}} \left(\frac{\partial A_{i+1,j}}{\partial \rho_{i,j}} u_{i,j} \right) = -\frac{\left(u_{i+1,j} - u_{i,j}\right)}{2\Delta_x^2 \rho_{i,j}^3} \\
\frac{\partial^2 A_{i-1,j}}{\partial \rho_{i,j}^2} u_{i,j} &= -\frac{2}{\rho_{i,j}} \left(\frac{\partial A_{i-1,j}}{\partial \rho_{i,j}} u_{i,j} \right) = \frac{\left(u_{i,j} - u_{i-1,j}\right)}{2\Delta_x^2 \rho_{i,j}^3} \\
\frac{\partial^2 A_{i,j+1}}{\partial \rho_{i,j}^2} u_{i,j} &= -\frac{2}{\rho_{i,j}} \left(\frac{\partial A_{i,j+1}}{\partial \rho_{i,j}} u_{i,j} \right) = -\frac{\left(u_{i,j+1} - u_{i,j}\right)}{2\Delta_z^2 \rho_{i,j}^3} \\
\frac{\partial^2 A_{i,j-1}}{\partial \rho_{i,j}^2} u_{i,j} &= -\frac{2}{\rho_{i,j}} \left(\frac{\partial A_{i,j-1}}{\partial \rho_{i,j}} u_{i,j} \right) = \frac{\left(u_{i,j} - u_{i,j-1}\right)}{2\Delta_z^2 \rho_{i,j}^3}.
\end{aligned} \tag{B-2}$$

Third, for slowness and buoyancy, it will be $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial s^2}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$, $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial s \partial b}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$, $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial b \partial s}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$ and $\frac{\partial^2 \mathbf{A}(\mathbf{m},\omega)}{\partial b^2}\mathbf{u}(\mathbf{m},\mathbf{x}_s,\omega)$,

$$\begin{aligned}
\frac{\partial^2 A_{i,j}}{\partial s_{i,j}^2} u_{i,j} &= \frac{1}{s_{i,j}} \left(\frac{\partial A_{i,j}}{\partial s_{i,j}} u_{i,j} \right) = 2\omega^2 b_{i,j} u_{i,j}, \\
\frac{\partial^2 A_{i,j}}{\partial s_{i,j} \partial b_{i,j}} u_{i,j} &= \frac{\partial^2 A_{i,j}}{\partial b_{i,j} \partial s_{i,j}} u_{i,j} = \frac{1}{b_{i,j}} \left(\frac{\partial A_{i,j}}{\partial s_{i,j}} u_{i,j} \right) = 2\omega^2 s_{i,j} u_{i,j},
\end{aligned}$$

and

$$\frac{\partial^2 A_{i,j}}{\partial b_{i,j}^2} u_{i,j} = \frac{\partial^2 A_{i+1,j}}{\partial b_{i,j}^2} u_{i,j} = \frac{\partial^2 A_{i-1,j}}{\partial b_{i,j}^2} u_{i,j} = \frac{\partial^2 A_{i,j+1}}{\partial b_{i,j}^2} u_{i,j} = \frac{\partial^2 A_{i,j-1}}{\partial b_{i,j}^2} u_{i,j} = 0. \tag{B-3}$$

APPENDIX C - CALCULATION OF SECOND-ORDER DERIVATIVE OF THE IMPEDANCE MATRIX IN THE HESSIAN

In the monoparameter case, we use the acoustic wave equation with constant density to describe the wave motion. For simplicity, in this appendix, we assume that the source spectrum $f_s(\omega) = 1$, and invert square slowness $s(\mathbf{x}) = 1/v^2(\mathbf{x})$ for only one receiver, one source and one frequency. In the nonlinear inversion we study before, the standard gradient is

$$\mathbf{g} = \mathrm{Re}\left\{ \mathbf{u}^\dagger(s,\mathbf{x}_s,\omega) \left(\frac{\partial \mathbf{A}(s,\omega)}{\partial s} \right)^\dagger \lambda(s,\omega) \right\} = \mathrm{Re}\left\{ \omega^2 \mathbf{u}^\dagger(s,\mathbf{x}_s,\omega) \lambda(s,\omega) \right\}, \tag{C-1}$$

with

$$\begin{aligned}&\frac{\partial \mathbf{A}(s,\omega)}{\partial s} = \omega^2, \mathbf{u}(s,\mathbf{x}_s,\omega) = \mathbf{A}^{-1}(s,\omega)\mathbf{f}(\mathbf{x}_s,\omega), \quad \text{and} \\ &\lambda(s,\omega) = \left(\mathbf{A}^{-1}(s,\omega)\right)^{\dagger}\mathbf{R}^T\delta\mathbf{d}.\end{aligned} \tag{C-2}$$

It is obvious that source side wavefield is the GreenâĂŹs function for the related source location, and the adjoint wavefield is the receiver side wavefield generated with the conjugate of the data residuals as source. Insert the expression of the adjoint wavefield and the source side wavefield into the gradient, at nth iteration, the gradient at each position $\mathbf{x}$ becomes

$$\mathbf{g} = \text{Re}\left\{\omega^2 G(\mathbf{x},\mathbf{x}_s,\omega|s_n)G(\mathbf{x}_g,\mathbf{x},\omega|s_n)\delta d^*(\mathbf{x}_g,\mathbf{x}_s,\omega|s_n)\right\}, \tag{C-3}$$

and the nonlinear descent direction is

$$g_n(\mathbf{x}) = -\text{Re}\left\{\frac{\partial G(\mathbf{x}_g,\mathbf{x}_s,\omega|s_{n+1})}{\partial s(\mathbf{x})}\delta d^*\left(\mathbf{x}_g,\mathbf{x}_s,\omega|s_n\right)\right\}, \tag{C-4}$$

with the high-order sensitivities for the nth iteration is

$$\frac{\partial G(\mathbf{x}_g,\mathbf{x}_s,\omega|s_{n+1})}{\partial s(\mathbf{x})} = \left(\frac{\partial G(\mathbf{x}_g,\mathbf{x}_s,\omega|s_{n+1})}{\partial s(\mathbf{x})}\right)_0 + \left(\frac{\partial G(\mathbf{x}_g,\mathbf{x}_s,\omega|s_{n+1})}{\partial s(\mathbf{x})}\right)_1 + \ldots, \tag{C-5}$$

where the zero order term

$$\left(\frac{\partial G(\mathbf{x}_g,\mathbf{x}_s,\omega|s_{n+1})}{\partial s(\mathbf{x})}\right)_0 = \frac{\partial G(\mathbf{x}_g,\mathbf{x}_s,\omega|s_n)}{\partial s(\mathbf{x})} = -\omega^2 G(\mathbf{x}_g,\mathbf{x},\omega|s_n)G(\mathbf{x},\mathbf{x}_s,\omega|s_n),$$

and the first order term is

$$\begin{aligned}\left(\frac{\partial G(\mathbf{x}_g,\mathbf{x}_s,\omega|s_{n+1})}{\partial s(\mathbf{x})}\right)_1 = \omega^4 \int \mathrm{d}\mathbf{x}'\delta s_n(\mathbf{x}')\,&[G(\mathbf{x}_g,\mathbf{x}',\omega|s_n)G(\mathbf{x}',\mathbf{x},\omega|s_n)G(\mathbf{x},\mathbf{x}_s,\omega|s_n) \\ &+ G(\mathbf{x}_g,\mathbf{x},\omega|s_n)G(\mathbf{x},\mathbf{x}',\omega|s_n)G(\mathbf{x}',\mathbf{x}_s,\omega|s_n)]\,,\end{aligned}$$

with $\delta s_n(\mathbf{x})$ is the pre-calculated perturbation from the current iteration, and it is a function of the current data residual

$$\begin{aligned}&\delta d(\mathbf{x}_g,\mathbf{x}_s,\omega|s_n) = -\omega^2 \int \mathrm{d}\mathbf{x} G(\mathbf{x}_g,\mathbf{x},\omega|s_n)\delta s(\mathbf{x})G(\mathbf{x},\mathbf{x}_s,\omega|s_n) + \cdots = \mathbf{J}\delta s_n + \ldots, \quad \text{or} \\ &\delta s_n \approx -\mathbf{H}_2\text{Re}\left\{\mathbf{J}^T\delta d^*(\mathbf{x}_g,\mathbf{x}_s,\omega|s_n)\right\}.\end{aligned} \tag{C-6}$$

The first term of the Hessian operator is

$$\begin{aligned}\mathbf{H}_1 = \text{Re}\Bigg\{&\left(\frac{\partial \mathbf{A}(s,\omega)}{\partial s}\lambda(s,\omega)\right)^{\dagger}\mathbf{A}^{-1}(s,\omega)\left(-\frac{\partial \mathbf{A}(s,\omega)}{\partial s}\mathbf{u}(s,\omega)\right) \\ &+\left(\frac{\partial \mathbf{A}(s,\omega)}{\partial s}\mathbf{u}(s,\omega)\right)^{\dagger}\left(\mathbf{A}^{-1}(s,\omega)\right)^{\dagger}\left(-\left(\frac{\partial \mathbf{A}(s,\omega)}{\partial s}\right)^{\dagger}\lambda(s,\omega)\right)\Bigg\}.\end{aligned} \tag{C-7}$$

The related Hessian-vector is

$$\mathbf{H}_1\delta s_n = \mathrm{Re}\left\{\left(\frac{\partial \mathbf{A}(s,\omega)}{\partial s}\lambda(s,\omega)\right)^{\dagger}\mathbf{A}^{-1}(s,\omega)\left(-\frac{\partial \mathbf{A}(s,\omega)}{\partial s}\mathbf{u}(s,\omega)\delta s_n\right)\right.$$
$$\left.+\left(\frac{\partial \mathbf{A}(s,\omega)}{\partial s}\mathbf{u}(s,\omega)\right)^{\dagger}\left(\mathbf{A}^{-1}(s,\omega)\right)^{\dagger}\left(-\left(\frac{\partial \mathbf{A}(s,\omega)}{\partial s}\right)^{\dagger}\lambda(s,\omega)\delta s_n\right)\right\}. \quad \text{(C-8)}$$

Inserting the expression of different terms (C-2) in the Hessian operator and the related Hessian-vector product (C-7) and (C-8), we can write the above Hessian operator and Hessian-vector product into the form of Green's function, For the first term $\mathbf{H}_1$ in the Hessian operator,

$$H_1(\mathbf{x},\mathbf{x}') = -\omega^4\mathrm{Re}\left\{[G(\mathbf{x}_g,\mathbf{x}',\omega|s_n)G(\mathbf{x}',\mathbf{x},\omega|s_n)G(\mathbf{x},\mathbf{x}_s,\omega|s_n)\right.$$
$$\left.+\ G(\mathbf{x}_g,\mathbf{x},\omega|s_n)G(\mathbf{x},\mathbf{x}',\omega|s_n)G(\mathbf{x}',\mathbf{x}_s,\omega|s_n)]\,\delta d^*(\mathbf{x}_g,\mathbf{x}_s,\omega|s_n)\right\}. \quad \text{(C-9)}$$

Similarly, we can get the expression of the second term $\mathbf{H}_2$ in the Hessian operator

$$H_2(\mathbf{x},\mathbf{x}') = \omega^4\mathrm{Re}\left\{G(\mathbf{x}_g,\mathbf{x}',\omega|s_n)G(\mathbf{x}',\mathbf{x}_s,\omega|s_n)G^*(\mathbf{x}_g,\mathbf{x},\omega|s_n)G^*(\mathbf{x},\mathbf{x}_s,\omega|s_n)\right\}. \quad \text{(C-10)}$$

Substituting the above Hessian operators back to the approximate Newton method, we can obtain the descent direction for each position as

$$g(\mathbf{x}) + H_1(\mathbf{x},\mathbf{x}')\delta s_n = \mathrm{Re}\left\{\omega^2 G(\mathbf{x}_g,\mathbf{x},\omega|s_n)G(\mathbf{x},\mathbf{x}_s,\omega|s_n)\delta d^*(\mathbf{x}_g,\mathbf{x}_s,\omega|s_n)\right.$$
$$-\omega^4\int \mathrm{d}\mathbf{x}'\delta s_n(\mathbf{x}')\,[G(\mathbf{x}_g,\mathbf{x}',\omega|s_n)G(\mathbf{x}',\mathbf{x},\omega|s_n)G(\mathbf{x},\mathbf{x}_s,\omega|s_n)$$
$$\left.+\ G(\mathbf{x}_g,\mathbf{x},\omega|s_n)G(\mathbf{x},\mathbf{x}',\omega|s_n)G(\mathbf{x}',\mathbf{x}_s,\omega|s_n)]\,\delta d^*(\mathbf{x}_g,\mathbf{x}_s,\omega|s_n)\right\}. \quad \text{(C-11)}$$

It is equivalent with the nonlinear descent direction we obtained from the scattering theory by perturbing the first-order perturbations, the gradient, with the second-order perturbations, which is obtained from the true data residuals. However, it is different from the nonlinear inverse theory (Snieder, 1990; Innanen, 2014), which can be seen by removing the nonlinear effects (multiscattered) in the data residual to perform a better linearized inversion using the Gauss-Newton method or even preconditioned gradient-based method.

REFERENCES

Alkhalifah, T., and Plessix, R. E., 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, **79**, No. 3, R91–R101.
URL <Go to ISI>://WOS:000338322900029

Fichtner, A., 2011, Full Seismic Waveform Modelling and Inversion: Springer-Verlag Berlin Heidelberg.

Fichtner, A., Bunge, H. P., and Igel, H., 2006a, The adjoint method in seismology: Physics of the Earth and Planetary Interiors, **157**, No. 1-2, 86–104.
URL <Go to ISI>://WOS:000239474200008

Fichtner, A., Bunge, H. P., and Igel, H., 2006b, The adjoint method in seismology - ii. applications: traveltimes and sensitivity functionals: Physics of the Earth and Planetary Interiors, **157**, No. 1-2, 105–123.
URL <Go to ISI>://WOS:000239474200009

Geng, Y., Pan, W. Y., and Innanen, K. A., 2017, Frequency-domain full-waveform inversion with updates based on nonlinear sensitivities, *in* SEG Technical Program Expanded Abstracts, 1330–1335.

Hustedt, B., Operto, S., and Virieux, J., 2004, Mixed-grid and staggered-grid finite-difference methods for frequency-domain acoustic wave modelling: Geophysical Journal International, **157**, No. 3, 1269–1296.
URL <Go to ISI>://WOS:000222667100024

Innanen, K. A., 2014, Seismic full waveform inversion with nonlinear sensitivities: CREWES Annual Report, **26**.

Jo, C. H., Shin, C. S., and Suh, J. H., 1996, An optimal 9-point, finite-difference, frequency-space, 2-d scalar wave extrapolator: Geophysics, **61**, No. 2, 529–537.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migrations, *in* Conference on Inverse Scattering, Theory and Application, Society of Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Métivier, L., Bretaudeau, F., Brossier, R., Operto, S., and Virieux, J., 2014, Full waveform inversion and the truncated newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, No. 6, 1353–1375.
URL <Go to ISI>://WOS:000344222300013

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2013, Full waveform inversion and the truncated newton method: Siam Journal on Scientific Computing, **35**, No. 2, B401–B437.
URL <Go to ISI>://WOS:000318404100035

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Métivier, L., and Virieux, J., 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: The Leading Edge, **32**, No. 9, 1040–1054.

Pan, W. Y., Innanen, K. A., and Liao, W. Y., 2017, Accelerating hessian-free gauss-newton full-waveform inversion via l-bfgs preconditioned conjugate-gradient algorithm: GEOPHYSICS, **82**, No. 2, R49–R64.

Pan, W. Y., Innanen, K. A., Margrave, G. F., Fehler, M. C., Fang, X. D., and Li, J. X., 2016, Estimation of elastic constants for hti media using gauss-newton and full-newton multiparameter full-waveform inversion: Geophysics, **81**, No. 5, R275–R291.
URL <Go to ISI>://WOS:000392752200053

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, No. 2, 495–503.

Plessix, R. E., Milcik, P., Rynja, H., Stopin, A., Matson, K., and Abri, S., 2013, Multiparameter full-waveform inversion: Marine and land examples: The Leading Edge, **32**, No. 9, 1030–1038.

Pratt, R. G., Shin, C., and Hicks, G. J., 1998, Gauss-newton and full newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, No. 2, 341–362.
URL <Go to ISI>://WOS:000073793700010

Snieder, R., 1990, A perturbative analysis of nonlinear inversion: Geophysical Journal International, **101**, No. 3, 545–556.
URL <Go to ISI>://WOS:A1990DG96100004

Tarantola, A., 1984, Inversion of seismic-reflection data in the acoustic approximation: Geophysics, **49**, No. 8, 1259–1266.

Virieux, J., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, No. 6, Wcc1–Wcc26.