A comparison of different scaling methods for least-squares inversion/migration

Wenyong Pan, Kris Innanen, Gary Margrave

ABSTRACT

The least-squares inverse problem, such as full waveform inversion and least-squares migration, can be performed iteratively using a gradient based method. While the convergence rate of this method is very slow for that it assumes the Hessian matrix as an identity matrix. The poorly scaled gradient can be enhanced considerably by multiplying the inverse Hessian matrix. Hessian matrix works as a nonstationary deconvolution operator to compensate the geometrical spreading effects and recover the deep reflectors amplitudes. It can also sharp or focus the gradient by suppressing the multiple scattering effects and improve the resolution of the gradient. While direct calculation of the Hessian matrix is considered to be unfeasible in practical application for its expensively computational burden. Many Hessian approximations have been proposed to scale the gradient and improve the convergence rate of the gradient method. In this research, we compared different scaling methods in the least-squares inverse problem based on different Hessian approximations. The pseudo-Hessian, constructed by two virtual sources, can compensate the geometrical spreading effects obviously. While it is still not enough to balance the amplitude for that it ignores the receiver-side Green's functions. The Hessian approximation based on double illumination method, the linear and chirp phase encoded Hessian can balance the amplitude better for taking the receiver-side Green's functions into consideration. The chirp phase encoding method introduced in this research can approach the exact approximate Hessian better with the same number of simulations compared to the linear phase encoding method.

INTRODUCTION

The least-squares inverse strategy has been considered as an important method in estimating the subsurface parameters (e.g., full waveform inversion) and building the reflectivity profile (e.g., least-squares reverse time migration) through an iterative process. Whether full waveform inversion (FWI) or least-squares reverse time migration (LSRTM) can be implemented using a gradient based method. The gradient can be constructed by applying a zero-lag convolution between the forward modeling wavefields and backpropagated wavefields based on an adjoint state method (Lailly, 1983; Tarantola, 1984). While the gradient based method is regarded as a crude scaling strategy and the gradient is a poorly scaled image which results in a slow convergence rate. The gradient can be considerably enhanced by multiplying the inverse Hessian, the second order partial derivative of the misfit function with respect to the model parameters (Pratt et al., 1998). The Hessian matrix serves as a nonstationary deconvolution operator to compensate the geometrical spreading effects, suppress the multiple scatterings and improve the resolution of the image. While calculating the Hessian matrix directly is prohibitively expensive.

The approximate Hessian in Gauss-Newton method, which ignores the nonlinear term in full Hessian matrix (Pratt et al., 1998; Virieux and Operto, 2009), is equivalent to the Hessian matrix in LSRTM. The approximate Hessian, which can be constructed by a scalar product between two partial derivative wavefields in frequency domain (Pratt et al., 1998), is still too expensive to be calculated directly. Under the assumption of high frequency limit, the Hessian matrix is diagonally dominant. In this case, the inverse Hessian matrix can be approximated by the reciprocals of the Hessian matrix. The diagonal part of the Hessian matrix can serve as a good approximation to recover and balance the amplitudes (Shin et al., 2001a).

Shin et al. (2001b) proposed to use the pseudo-Hessian matrix as a substitution for the approximate Hessian in Gauss-Newton method. He showed how to improve the faint migration image caused by geometrical spreading and transmission loss using the pseudo-Hessian matrix. Based on the assumption of infinite receiver coverage, the pseudo-Hessian can be constructed by the forward modeling wavefields, used as the virtual sources in reverse time migration. And we found that the diagonal part of the pseudo-Hessian is actually equivalent to the source illumination and preconditioning the gradient using the pseudo-inverse of the diagonal pseudo-Hessian matrix is identical to the standard deconvolution imaging condition in reverse time migration. While the diagonal part of the pseudo-Hessian is not effective enough to balance the amplitudes, for that the pseudo-Hessian matrix was derived by ignoring the effects of the zero-lag auto-correlation of receiver-side Green's functions which describe the receiver-side geometrical spreading (Choi and Shin, 2007).

When considering the reciprocity theory, the receiver-side Green's functions can be replaced by the source-side Green's functions under the assumption that the sources and receivers are collocated (Plessix and Mulder, 2006). This strategy forms the double illumination methods when preconditioning the gradient using the diagonal part of the Hessian. The receiver-side Green's functions can also be constructed using the phase encoding method (Tang, 2009) which can reduce the computational burden significantly compared to the traditional shot-by-shot method. Tao and Sen (2013)employed a linear phase encoded Hessian in plane wave FWI. In this research, we introduced a chirp phase encoded method which actually is a combination of the linear (Morton and Ober, 1998; Romero et al., 2000) and random phase encoding methods (Whitmore, 1995; Zhang et al., 2005), to construct the diagonal part of the Hessian.

Firstly, we illustrated a numerical example based on a 50×50 homogeneous model to compare the Hessian approximations. The we presented a multi-layer model and more complex Marmousi model to compare the gradient scaled by different Hessian approximations. The scaling methods proposed in this research can be applied to full waveform inversion or least-squares migration for improving the convergence rate.

THEORY AND METHOD

Gradient Calculation in the least-squares inverse problem

As a least-squares local optimization, full waveform inversion seeks to minimize the difference between the synthetic data and observed data (Lailly, 1983; Tarantola, 1984) and update the model iteratively. The misfit function ϕ is given in a least-squares norm:

$$\phi\left(s_0^{(n)}(\mathbf{r})\right) = \frac{1}{2}\int d\omega \left(\sum_{\mathbf{r}_s}\sum_{\mathbf{r}_g} \|\delta P\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}(\mathbf{r})\right)\|_2\right), \tag{1}$$

where $s_0^{(n)}(\mathbf{r}) = \frac{1}{\left(c_0^{(n)}(\mathbf{r})\right)}$ are the model parameters, the square of the slowness in the nth iteration, and $c_0^{(n)}(\mathbf{r})$ is the velocity. δP mean the data residuals, the difference between the observed data $P\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}\right)$ and synthetic data $G\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}\right)$, $\|\cdot\|_2$ indicates the $\ell - 2$ norm.

The minimum value of the misfit function is sought in the vicinity of the starting model $s_0(\mathbf{r})$ and the updated model can be written as the sum of the starting model and a model perturbation $\delta s_0^{(n)}(\mathbf{r})$ (Virieux and Operto, 2009).

$$s^{(n)}(\mathbf{r}) = s_0^{(n)}(\mathbf{r}) + \mu^{(n)} \delta s_0^{(n)}(\mathbf{r}), \tag{2}$$

where $\mu^{(n)}$ is the step length in nth iteration, which is a scalar constant used to scale the model perturbation and can be obtained through a line search method (Gauthier et al., 1986; Pica et al., 1990).

Applying a second order Taylor-Lagrange development of the misfit function and then taking partial derivative with respect to the model parameters give the model perturbation as:

$$\delta s_0^{(n)} = -\int d\mathbf{r}' H^{(n)-}(\mathbf{r}, \mathbf{r}') g^{(n)}(\mathbf{r}), \tag{3}$$

where $g^{(n)}(\mathbf{r}')$ is the gradient and $H^{(n)}(\mathbf{r}', \mathbf{r})$ is the Hessian matrix.

For gradient, the first order derivative of the misfit function ϕ with respect to the model parameters can be obtained by a zero-lag correlation between the data residuals and the first order partial derivative wavefields.Then apply a perturbation derivation based on the Born approximation, the sensitive matrix can be written as:

$$\frac{\delta G\left(\mathbf{r}_g, \mathbf{r}_s, \omega | s_0^{(n)}\right)}{\delta s_0^{(n)}(\mathbf{r})} = -\omega^2 G\left(\mathbf{r}_g, \mathbf{r}, \omega | s_0^{(n)}\right) G\left(\mathbf{r}, \mathbf{r}_s, \omega | s_0^{(n)}\right), \tag{4}$$

Then we can get the gradient:

$$g^{(n)}(\mathbf{r}) = \sum_{\mathbf{r}_s, \mathbf{r}_g} \int d\omega \Re\left(\omega^2 \mathcal{F}_s(\omega) G(\mathbf{r}, \mathbf{r}_s, \omega) G(\mathbf{r}_g, \mathbf{r}, \omega) \delta P^*\right), \tag{5}$$

where $G(\mathbf{r}, \mathbf{r}_s, \omega)$ and $G(\mathbf{r}_g, \mathbf{r}, \omega)$ are the source-side and receiver-side Green's functions, respectively. And $\mathcal{F}_s(\omega)$ is the source signature. Then the gradient can be calculated using the adjoint state method by applying a zero-lag convolution between the forward modeling wavefields and back-propagated data residuals, which avoids the direct computation of the partial derivative wavefields.

Full Hessian and Approximate Hessian

The image or gradient formed by crosscorrelation imaging condition suffers from geometrical spreading effects, wave attenuation and transmission loss, which result in poor amplitudes for deep reflectors. Multiplying the inverse Hessian matrix can remove the geometrical amplitude decay of the Green's functions (Tao and Sen, 2013).

The full Hessian matrix is the second order partial derivative of the misfit function with respect to the model parameters. And after a series of derivations, the Hessian matrix can be written as the summation of two terms:

$$H^{(n)}(\mathbf{r}', \mathbf{r}) = H_1^{(n)} + H_2^{(n)}, \tag{6}$$

And the first term of full Hessian matrix $H_1^{(n)}$ can be computed by multiplying the data residuals using the second order partial derivative wavefields, which is defined as the variation of the first order partial derivative wavefields corresponding to perturbation in the model parameters. It indicates the nonlinearity of the least-squares inverse problem. The crosscorrelation method produces false anomalies caused by correlating the multiple scattering in data residuals with the partial derivative wavefields (Pratt et al., 1998). The first term $H_1^{(n)}$ predicts these high-order multiple scatterings and works as a de-multiple operator to suppress these artifacts in the gradient.

It is always expensive to compute the first term of the Hessian matrix directly. When the initial model is close to the real model, the first term $H_1^{(n)}$ in the full Hessian matrix is always neglected for computation convenience, which makes the inverse problem approximately linear. Thus the full Hessian matrix can be substituted by the approximate Hessian (the second term in equation (6)), which is used as a preconditioner for the gradient in the Gauss-Newton method:

$$H_a^{(n)} = H_2^{(n)} = \sum_{\mathbf{r}_s, \mathbf{r}_g} \int d\omega \Re \left\{ \omega^4 G^s \left(\mathbf{r}'', \mathbf{r}', \omega | s_0^{(n)} \right) G^g \left(\mathbf{r}'', \mathbf{r}', \omega | s_0^{(n)} \right) \right\} \tag{7}$$

where $G^s = G\left(\mathbf{r}', \mathbf{r}_s, \omega | s_0^{(n)}\right) G^*\left(\mathbf{r}'', \mathbf{r}_s, \omega | s_0^{(n)}\right)$ and $G^g = G\left(\mathbf{r}', \mathbf{r}_g, \omega | s_0^{(n)}\right) G^*\left(\mathbf{r}'', \mathbf{r}_g, \omega | s_0^{(n)}\right)$ indicate the source-side and receiver-side Green's functions respectively.

Each element in the approximate Hessian can be interpreted as the scalar product of two partial derivative wavefields (Pratt et al., 1998), which is equivalent to the zero-lag convolution computation in time domain. The diagonal elements (when $\mathbf{r}' = \mathbf{r}''$) and off-diagonal elements (when $\mathbf{r}' \neq \mathbf{r}''$) of the approximate Hessian can be interpreted as the zero-lag auto-correlation and cross-correlation of the partial derivative wavefields respectively (Shin et al., 2001a), as indicated by the first term $\bar{H}_a^{(n)}$ and second term $\tilde{H}_a^{(n)}$ in equation (8).

$$H_a^{(n)} = \bar{H}_a^{(n)} + \tilde{H}_a^{(n)}, \tag{8}$$

In the high-frequency asymptotics and for the reference model with relatively smooth impedance variations, the two partial derivative wavefields are largely uncorrelated with each other but perfectly self-correlated (Pratt et al., 1998; Tang, 2009), which means that the approximate Hessian is diagonally dominated. In this case, the off-diagonal elements of the approximate Hessian, the second term $\tilde{H}_a^{(n)}$ in equation (8), can be neglected. And the diagonal elements of the approximate Hessian, the auto-correlation between the source-side Green's functions and receiver-side Green's functions, can serve as a good preconditioner for the gradient to deblur the image. What's more, the inverse approximate Hessian can be approximated by the reciprocals of the approximate Hessian $\frac{1}{\bar{H}_a^{(n)}}$ (Ben-Hadj-Ali et al., 1989).

Pseudo-Hessian

Under the assumption of the infinite receiver coverage, the receiver-side Green's functions can be approximated as $d^2(\mathbf{r}_g, \mathbf{r}')$, where $d(\mathbf{r}_g, \mathbf{r}')$ means the distance from position $\mathbf{r}'$ to the receiver positionr $\mathbf{r}_g$. When the depth in vertical is quite smaller than the distance in horizontal, $d^2(\mathbf{r}_g, \mathbf{r}')$ can be approximated as a constant scalar L. This approximation to Hessian matrix is equivalent to the pseudo-Hessian proposed by Shin et al.(2001b). The pseudo-Hessian is constructed using the forward modeling wavefields, used as the virtual sources in reverse rime migration.

$$H_{p\_a}^{(n)} = L \sum_{\mathbf{r}_s} \int d\omega \omega^4 \Re \left\{ G\left(\mathbf{r}'', \mathbf{r}_s, \omega\right) G^*\left(\mathbf{r}', \mathbf{r}_s, \omega\right) \right\}, \tag{9}$$

The diagonal part of the pseudo-Hessian can be formed when $\mathbf{r}' = \mathbf{r}''$.

Double Illumination Method

For finite receiver coverage, the pseudo-Hessian is limited to compensate the geometrical spreading effects and balance the biased amplitudes, for missing the receiver-side Green's functions. By introducing the reciprocity theory, we can get a better approximation. The reciprocity principle proved by Aki and Richards (2002) in elastic media, states that the recorded wavefields are identical when interchanging the locations of sources and receivers. Hence, the receiver-side Green's functions can be constructed by the reciprocal wavefields from the receiver. In this research, we assume that if the lateral velocity variation is small, we can select some of the receivers regularly in the whole acquisition geometry to calculate the reciprocal wavefields, which can reduce the computation cost greatly. A further assumption is that the selected receivers and the sources are collocated. In this condition, the reciprocal wavefields can be substituted by the forward modeling wavefields, which gives the approximation of the diagonal Hessian as (Plessix and Mulder, 2006):

$$\bar{H}_{d\_a} = \sum_{\mathbf{r}_s} \int d\omega \Re \left\{ \omega^4 \parallel G(\mathbf{r}, \mathbf{r}_s, \omega) \parallel^2 \parallel G^*(\mathbf{r}, \mathbf{r}_s, \omega) \parallel^2 \right\}, \tag{10}$$

So,we can notice that preconditioning the gradient using $\bar{H}_{d\_a}$ is equivalent to applying a double illumination compensation to the gradient. That is why we call this method double illumination method. While this method is limited when the sources' distribution is regular corresponding to the receivers' locations.

Linear and Chirp Phase Encoded Hessian

Another strategy to construct the receiver-side Green's functions is using phase encoding technique, which can reduce the computational cost considerably but unfortunately involves strong crosstalk artifacts. Tang (2009) calculated the phase encoded Hessian and compared effects of linear phase encoding and random phase encoding approaches in attenuating the crosstalk artifacts. The phase encoding technique was firstly introduced for imaging (Romero et al., 2000; Liu et al., 2006; Perrone and Sava, 2012; Dai and Schuster, 2013) and then applied for calculating the gradient in FWI (Krebs et al., 2009; Vigh and Starr, 2008; Ben-Hadj-Ali et al., 2011). Tao and Sen (2013) used the plane wave linear phase encoding approach to calculate the gradient for efficient full waveform inversion in frequency-ray parameter domain. The crosstalk artifacts can be reduced effectively with sufficient source and receiver ray parameters. In this research, we used a chirp phase encoding strategy, which is a combination of the linear phase encoding and random phase encoding methods, to calculate the receiver-side Green's functions.

The linear time delay in time domain corresponds to linear phase shift in frequency domain, the linear phase encoded Hessian can be written as:

$$\bar{H}_{l\_a} = \sum_{\mathbf{r}_s} \sum_{\mathbf{p}_g} \int d\omega \Re \left\{ \omega^4 \bar{G}^s(\mathbf{r}', \mathbf{r}', \omega) \bar{G}^g(\mathbf{r}'', \mathbf{r}'', \omega) e^{i\omega p_g (x'_g - x_g)} \right\} \tag{11}$$

where $\mathbf{p}_g$ are the receiver-side ray parameters vector. In equation (11), when $x'_g = x_g$, the phase encoded Hessian becomes the exact approximate Hessian without crosstalk noise. And when $x'_g \neq x_g$, only crosstalk artifacts term is left. So, equation (11) can be written as the summation of the exact Hessian and crosstalk artifacts (Tang 2009).

The chirp phase encoding strategy used in this research is a combination of linear phase encoding and random phase encoding methods. And a random factor is added into the phase shift term of equation (11), which gives:

$$\bar{H}_{c\_a} = \sum_{\mathbf{r}_s} \int d\omega \Re \left\{ \omega^4 \bar{G}^s(\mathbf{r}', \mathbf{r}', \omega) \bar{G}^g(\mathbf{r}'', \mathbf{r}'', \omega) e^{i\omega (p_g + \varepsilon \triangle p)(x'_g - x_g)} \right\} \tag{12}$$

where $\triangle p = rand * p_g, rand \in (0, 1)$ is the random coefficient, ε is the coefficient used to control the amount of dithering.

NUMERICAL EXAMPLES

In this section, three numerical examples are illustrated for analysis and comparison. The first numerical example is a homogeneous model with a constant background velocity. The exact Hessian, Hessian contaminated by the crosstalk artifacts, pseudo-Hessian, Hessian based on double illumination method, linear and chirp phase encoded Hessian are presented for comparison. Then we practice the scaling methods on a multi-layer model and 2D Marmousi model to examine the effects of the preconditioners.

Fig.1 shows the homogeneous model with a constant velocity of 2500m/s used to calculate the Hessian matrix . The model consists of 50×50 grid cells with $5m$ horizontal and

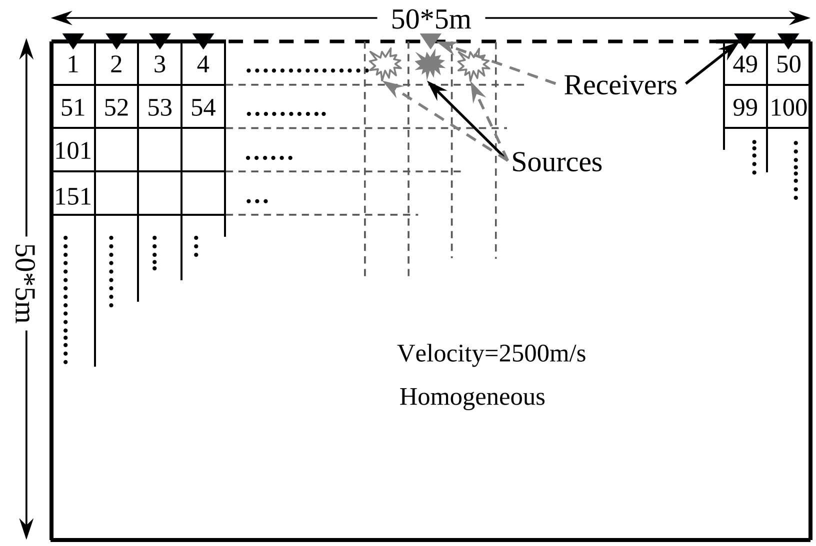

FIG. 1. The homogeneous model used to calculate the Hessian matrix. This model consists of $50 \times 50 = 2500$ grid cells with a grid interval of $5m$. The sources and receivers are located at the top surface

vertical grid intervals, which means that the total number of the parameters is 2500. The sources and receivers are located at the top surface of the model. The source function is a Ricker wavelet with a $25Hz$ dominant frequency bandlimited between 0 and $30Hz$. And the acquisition geometry is designed with 9 sources from $25m$ to $225m$ with a spacing of $125m$ and 50 receivers from $0m$ to $250m$ with a spacing of $25m$.

Fig. 2 shows the exact approximate Hessian and its approximations for this homogeneous model. Fig. 2a is the exact approximate Hessian. As we can see that the Hessian is banded and the energy varies along the diagonal line. So, the Hessian matrix works as a nonstationary deconvolution operator to recover the amplitudes for deep reflectors. Fig. 2b is the pseudo-Hessian, the energy of which is less concentrated on the diagonal band parts. The pseudo-Hessian is a crude approximation to the exact diagonal of the Hessian because it assumes the constant receiver-side Greens functions and ignores the effects of the limited receiver aperture (Tang, 2009). Fig. 2c is formed based on the double illumination method. Fig. 2e and f are the linear and chirp phase encoded Hessian respectively and the ray parameters range from $-0.3s/km$ to $0.3s/km$ with a step of $0.1s/km$. We can see that these three approximations are very close to the exact approximate Hessian. Fig. 2d is Hessian approximation contaminated by crosstalk artifacts when ray parameter $p = 0$ and we can see that the energy is not distributed regularly caused by the crosstalk artifacts. Fig. 3 shows the inverse Hessian corresponding to the Hessian approximations in Fig. 2. For Fig. 3a, we can see that the inverse of the exact approximate Hessian is diagonally dominated. While in the pseudo-Hessian, as shown by Fig. 3b, more energy is distributed in the off-diagonal elements. Fig. 3d shows the inverse Hessian with crosstalk artifacts, which is contaminated by some noise. While the Hessian based on double illumination method(Fig. 3c), the linear phase encoded Hessian(Fig. 3e) and chirp phase encoded Hessian (Fig. 3f) are very close to the exact one.

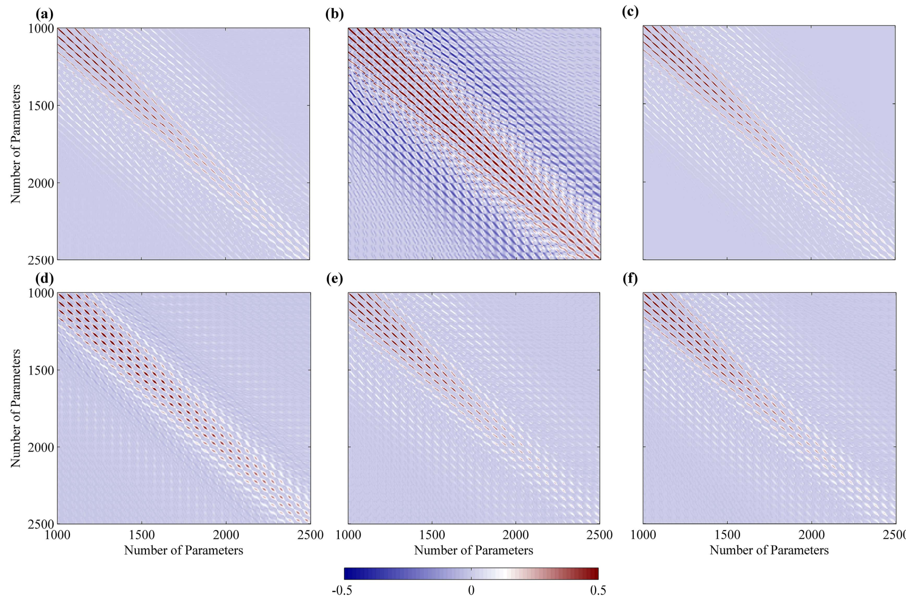

FIG. 2. The approximate Hessian and its approximations constructed with 9 sources and 50 receivers. (a) is the exact approximate Hessian; (b) is the pseudo Hessian; (c) is the Hessian approximation based on double illumination method; (d) is the phase encoded Hessian contaminated by crosstalk artifacts when ray parameter $p = 0$. (e) and (f) are the linear phase encoded Hessian and chirp phase encoded Hessian and the ray parameters range from $-0.3s/km$ to $0.3s/km$ with a step of $0.1s/km$.

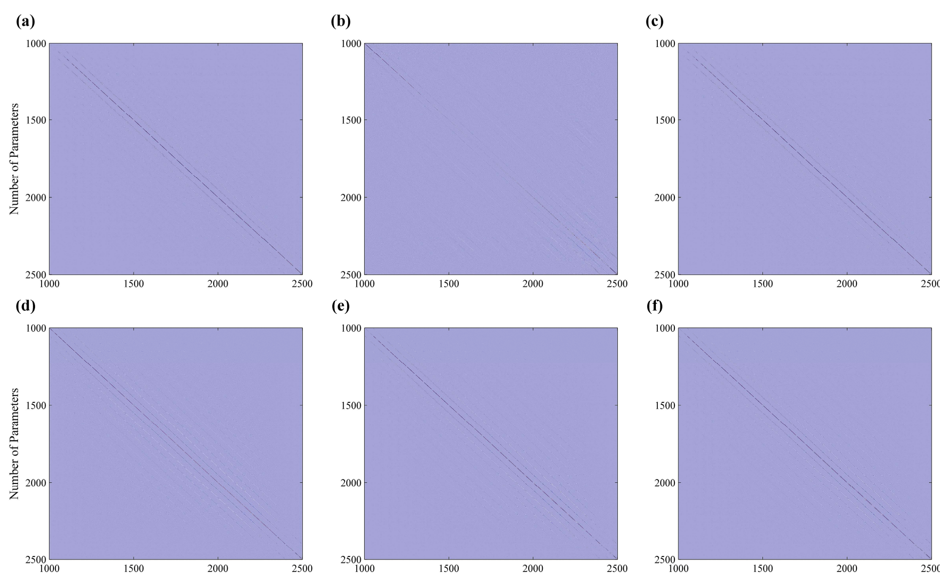

FIG. 3. Inverse Hessian corresponding to the Hessian approximations in Fig. 2.

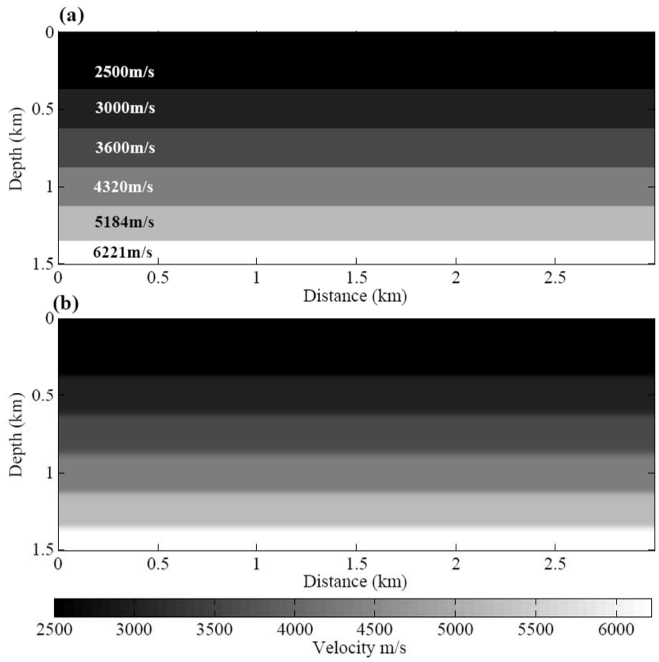

FIG. 4. The six-layer velocity model. (a) shows the exact six-layer model used in this section with the same reflection coefficient at each interface. The five interfaces are located at $0.375km$, $0.625km$, $0.875km$, $1.125km$ and $1.35km$ in depth respectively. And the velocities of the six layers are $2500m/s$, $3000m/s$, $3600m/s$, $4320m/s$, $5184m/s$ and $6221m/s$ respectively. (b) is the smoothed reference velocity model.

One six-layer velocity model with the same reflection coefficient at each interface (as shown in Fig. 4) is designed to verify the effectiveness of the proposed strategies. One source is excited at the location of $(1.5km, 0km)$ with a $30Hz$ dominant frequency Ricker wavelet source function. And 200 receivers with a spacing of $15m$ are arranged at the top surface from $0km$ to $3km$. The velocities of each layer are illustrated in Fig. 4a.

Fig. 5a shows the exact diagonal Hessian which is constructed by 201 simulations. Fig. 5b is the diagonal pseudo-Hessian. As what we have discussed above, the pseudo-Hessian ignores the effects of receiver-side GreenâĂŹs functions and it overestimates the total energy that enters the earth and returns to be recorded by the receivers (Tang, 2009). So, Fig. 5b looks stronger than Fig. 5a but with less accuracy. Fig. 5c is the diagonal part of the Hessian constructed using equation (10). Comparing it with Fig. 5b, it underestimates the total energy for that only one source is used in this numerical example. Fig. 5d is the diagonal part of the phase encoded Hessian obtained by one additional simulation when the ray parameter $p = 0$. It is contaminated by the crosstalk artifacts seriously which results from the interference among the wavefields of different sources.

Both of Fig. 5e and f are the diagonal parts of receiver-side linear phase encoded Hessian with ray parameters ranging from $-0.3s/km$ to $0.3s/km$. Fig. 5e is formed by 7 simulations, while Fig. 5f is formed by 31 simulations. It can be observed that Fig. 5e is still contaminated by crosstalk artifacts obviously. Fig. 5f gives a better result with more simulations and it is very close to the exact approximate Hessian, as shown by Fig. 5a. Fig. 5g and h are obtained using the chirp phase encoding method with the same ray parameter

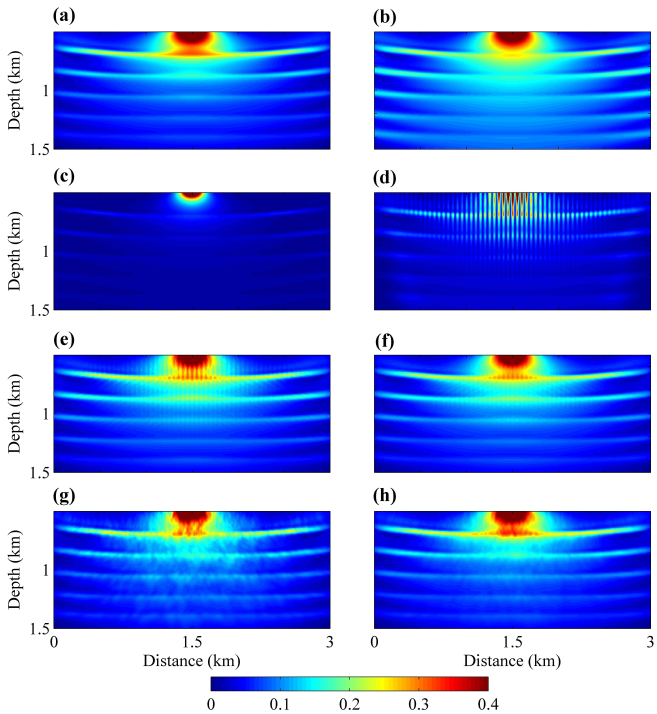

FIG. 5. Diagonal part of the Hessian for the six-layer model with one source and 200 receivers. (a) shows the diagonal part of the exact approximate Hessian; (b) is the diagonal part of the pseudo-Hessian; (c) is the diagonal part of the Hessian constructed by the double source illumination method. (d) is diagonal part of the linear phase encoded diagonal Hessian with the ray parameter $p = 0$; Both of (e) and (f) are the diagonal parts of receiver-side linear phase encoded Hessian with ray parameters ranging from $-0.3s/km$ to $0.3s/km$. (e) is formed by 7 simulations with a ray parameter step of $0.05s/km$. While (f) is formed by 31 simulations with a ray parameter step of $0.02s/km$. Both of (g) and (h) are the receiver-side hybrid phase encoded diagonal Hessian with ray parameters ranging from $-0.3s/km$ to $0.3s/km$. (g) and (h) are formed by 7 and 31 simulations with the ray parameter step of $0.02s/km$ and $0.05s/km$ respectively.

range but different number of simulations. The chirp phase encoding method can reduce the crosstalk artifacts effectively but introduce random noise. Fig. 5h is better than Fig. 5g, because more simulations are practiced.

The diagonal parts of the Hessian shown in Fig. 5 serve as the deconvolution operator to precondition the image. Fig. 6 gives the normalized amplitudes at interface 2, 3, 4 and 5 preconditioned by different Hessian approximations. The bold black lines are the gradients based on crosscorrelation imaging condition without precondition. The thin black lines indicate the gradients preconditioned by the diagonal pseudo-Hessian. The red lines indicate the gradients preconditioned by the exact diagonal Hessian. The blue lines indicate the the gradients preconditioned by the diagonal linear phase encoded Hessian. The red lines indicate the gradients preconditioned by the exact diagonal Hessian. The blue lines indicate the the gradients preconditioned by the diagonal linear phase encoded Hessian.The green lines indicate the the gradients preconditioned by the diagonal chirp phase encoded Hessian. It is easy for us to recognize that the image scaled by the exact diagonal Hessian (the red lines) gives the best amplitudes preservation at different interfaces. The image preconditioned by the diagonal pseudo-Hessian can recover the amplitudes to some extent

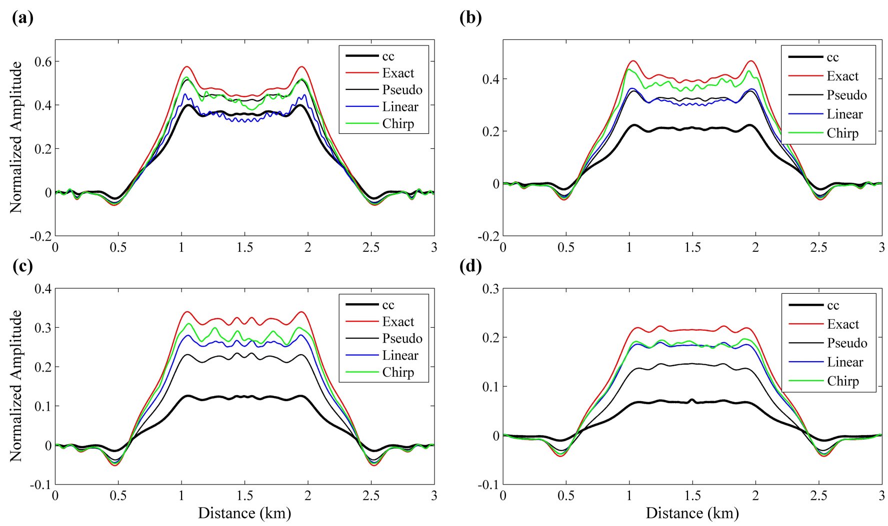

FIG. 6. Amplitude preservation by different conditions. (a), (b), (c) and (d) give the normalized amplitudes by different imaging conditions at layer 2, 3, 4 and 5 respectively. The bold-solid lines (cc) indicate the crosscorrelation imaging condition without precondition. The thin black lines indicate the gradients preconditioned by the diagonal pseudo-Hessian. The red lines indicate the gradients preconditioned by the exact diagonal Hessian. The blue lines indicate the the gradients preconditioned by the diagonal linear phase encoded Hessian. The red lines indicate the gradients preconditioned by the exact diagonal Hessian. The blue lines indicate the the gradients preconditioned by the diagonal linear phase encoded Hessian.The green lines indicate the the gradients preconditioned by the diagonal chirp phase encoded Hessian.

comparing with the crosscorrelation imaging condition. And the images preconditioned by diagonal linear phase encoded Hessian and chirp phase encoded Hessian are both better than that preconditioned by diagonal pseudo-Hessian.

Then the scaling methods mentioned in this research are practiced on the Marmousi Model. The Marmousi model is modified by introducing one water layer with a thickness of $125m$ and P-wave velocity of $1500m/s$. And the modified Marmousi model has 8012301 grid cells with the grid interval of $2.5m$ in horizontal and vertical. 30 point sources are distributed on the surface with a source interval of $125m$ from $875m$ to $4375m$ and 1150 receivers are deployed on the surface with a receiver interval of $5m$ from $5m$ to $5750m$. The source function is a Ricker wavelet with a dominant frequency of $15Hz$. The ray parameter range used for linear phase encoding and chirp phase encoding methods is $[-0.3s/km, 0.3s/km]$ and two ray parameter steps of $0.02s/km$ and $0.05s/km$ are tested for comparison. Fig.7a shows the exact P-wave velocity model and Fig.7b shows the smoothed P-wave velocity model used as the reference model in this research. Fig.8a is the true reflectivity profile obtained following Plessix and Mulder's method (2006)which states that the true reflectivity corresponds to the difference between the true slowness and reference slowness divided by the reference one. Comparing with the true reflectivity, the gradient is poorly scaled. And the amplitudes for the deep reflectors are very weak results from geometrical spreading effects. However, too much energy is concentrated on the shallow part of the image, which results in the overestimated reflectivity for the shallow reflectors.

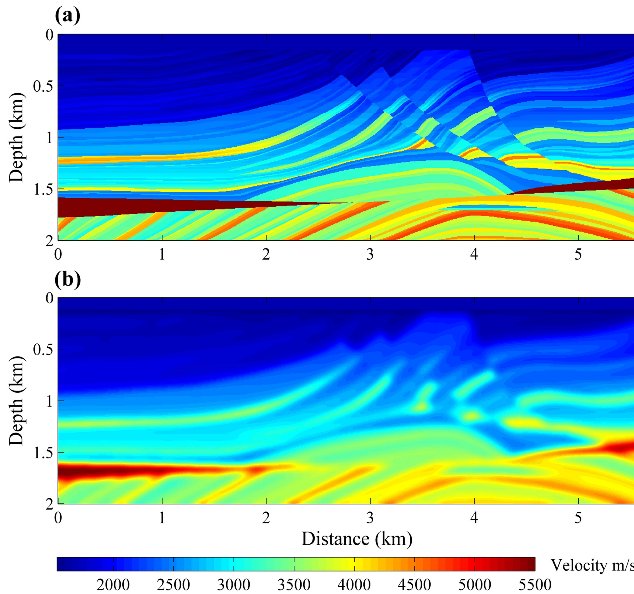

FIG. 7. The modified 2D Marmousi model. (a) P-wave velocity of the true model; (b) The smoothed P-wave velocity model used as the reference model in this research.

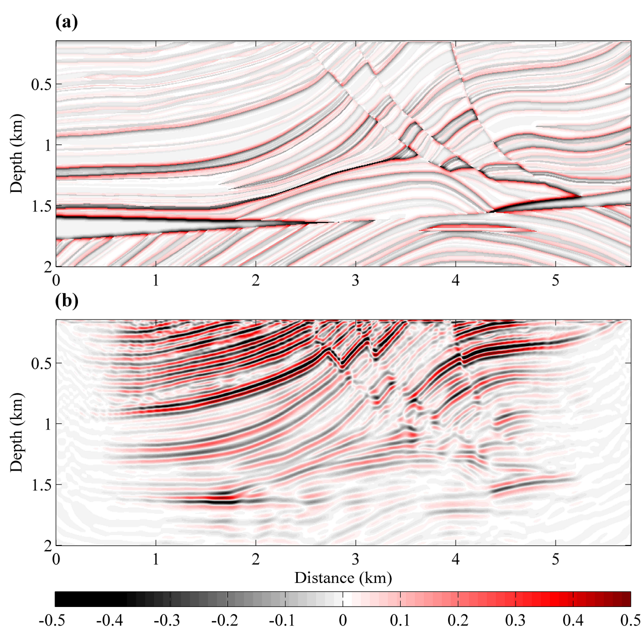

FIG. 8. A comparison between the true reflectivity and the image based on crosscorrelation imaging condition. (a) is the true reflectivity which corresponds to $I = (s_t - s_0)/s_0$, where s_t and s_0 denote the true slowness and the reference slowness respectively. (b) is the gradient based on a crosscorrelation imaging condition.

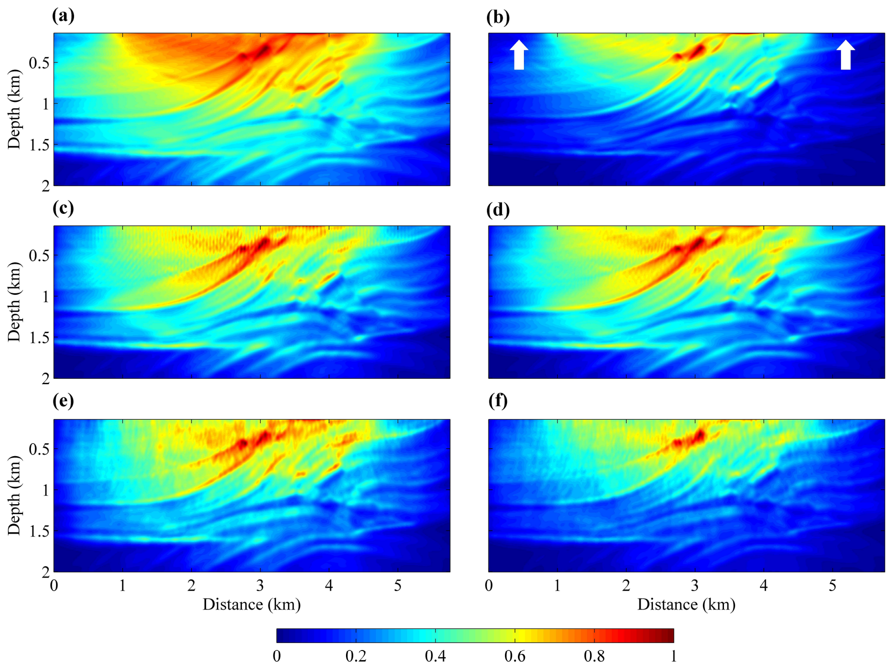

FIG. 9. The diagonal of the Hessian approximations obtained by different strategies. (a) is the diagonal part of the pseudo-Hessian. (b) is the diagonal part of the Hessian based on the double illumination method. (c) and (d) are diagonal parts of the linear phase encoded Hessian with different ray parameter settings; (e) and (f) are diagonal parts of the chirp phase encoded Hessian with different ray parameter settings.

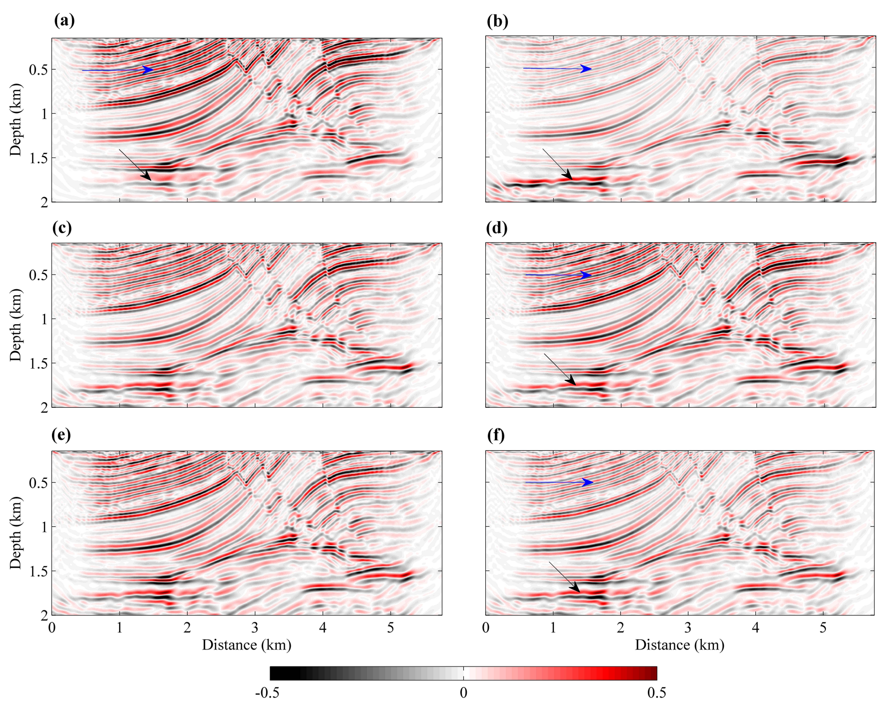

FIG. 10. The gradients obtained through different scaling methods corresponding to the different preconditioners in Fig.5.

Fig.9a shows the diagonal part of the pseudo Hessian which overestimates the illumination energy. Fig.9b is the diagonal Hessian approximation by the double illumination method, which is a very good approximation to the exact diagonal Hessian. While for the areas that the sources don't cover, the energy is underestimated, as indicated by the white arrows. Fig9.c and d are obtained using the linear phase encoding method with 7 and 31 simulations respectively. Fig.9e and f are obtained using the chirp phase encoding method with 7 and 31 simulations respectively. It can be seen that with increasing the number of simulations, the crosstalk artifacts can be suppressed effectively. And for the same number of simulations, the chirp phase encoding method can obtain a better approximation of the diagonal Hessian.

Fig.10 shows the gradients in the first iteration preconditioned by the corresponding Hessian approximations shown in Fig.9. Fig.10a is the gradient scaled by the diagonal pseudo Hessian. We can see that the amplitude of the deep reflectors are recovered obviously. By introducing the receiver-side Green's functions, the multiple scattering effects is suppressed, as denoted by the black arrows and the resolution is improved. The Hessian matrix works like a sharping or focusing filtering to the gradient. What's more, the gradients scaled by the double illumination method based diagonal Hessian and the chirp phase encoded diagonal Hessian have a more balanced amplitudes between the shallow and deep layers. And they are more close to the true reflectivity profile.

CONCLUSION

The Hessian matrix in the least-squares inverse problem can serve as a nonstationary deconvolution operator to compensate the geometrical spreading effects, improve the resolution and suppress the multiple scattering effects. we discussed different Hessian approximations which form the different scaling methods. The diagonal part of the pseudo Hessian, which forms the source-side illumination is limited balance the biased amplitude in the gradient. Three other strategies which also construct the receiver-side Green's functions, can balance the amplitude further, improve the resolution and sharp or focus the gradient. The chirp phase encoding method can approach the exact approximate Hessian better compared to the linear phase encoding method with the same computational cost.

ACKNOWLEDGMENTS

This research was supported by the Consortium for Research in Consortium of Elastic Wave Exploration Seismology (CREWES).

REFERENCES

Aki, K., and Richards, P. G., 2002, Quantitative Seismology: University Science Books, 2nd edn.

Ben-Hadj-Ali, H. S., Operto, S., and Virieux, J., 1989, Elastic ray-born l2-migration/inversion: Geophysical Journal International, **97**, 151–160.

Ben-Hadj-Ali, H. S., Operto, S., and Virieux, J., 2011, An efficient frequency-domain full waveform inversion method using simultaneous encoded sources: Geophysics, **76**, WCC177–WCC188.

Choi, Y., and Shin, C., 2007, Frequency domain elastic full waveform inversion using the new pseudo-hessian matrix: Elastic marmousi-2 synthetic test, 1908–1912.

Dai, W., and Schuster, G. T., 2013, Plane-wave least-squares reverse-time migration: Geophysics, **78**, S165–S177.

Gauthier, O., Virieux, J., and Tarantola, A., 1986, Two-dimensional nonlinear inversion of seismic waveforms: numerical results: Geophysics, **51**, 1387–1403.

Krebs, J. R., Anderson, J., Henkley, D., Neelamani, R., S. Lee, A. B., and Lacasse, M., 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysical Prospecting, **74**, WCC177–WCC188.

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migration, 206–220.

Liu, F., Whitemore, D. W., Day, R. S., and Stolt, R. H., 2006, Toward a unified analysis for source plane-wave migration: Geophysics, **71**, S129–S139.

Morton, S. A., and Ober, C. C., 1998, Faster shot-record depth migrations using phase encodingtest, 1131–1134.

Perrone, F., and Sava, P., 2012, Wave-equation migration with dithered plane waves: Geophysical Prospecting, **60**, 444–456.

Pica, A., Diet, J. P., and Tarantola, A., 1990, Nonlinear inversion of seismic reflection data in a laterally invariant medium: Geophysics, **55**, 284–292.

Plessix, R. E., and Mulder, W. A., 2006, Frequency-domain finite-difference amplitude-preserving migration: Geophysical Journal International, **157**, 975–987.

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**, 341–362.

Romero, L. A., Ghiglia, D. C., Ober, C. C., and Morton, S. A., 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

Shin, C., Jang, S., and Min, M., 2001a, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian: Geophysics, **74**, WCA95–WCA107.

Tao, Y., and Sen, M. K., 2013, Frequency-domain full waveform inversion with plane-wave data: Geophysics, **78**, R13–R23.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266.

Vigh, D., and Starr, E. W., 2008, 3d prestack plane-wave, full-waveform inversion: Geophysics, **73**, VE135–VE144.

Virieux, A., and Operto, S., 2009, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **74**, WCC1–WCC26.

Whitmore, N. D., 1995, An imaging hierarchy for common angle plane wave seismogram: Ph.D. thesis, The University of Tulsa.

Zhang, Y., Sun, J., Notfors, C., Gray, S. H., Chernis, L., and Young, J., 2005, Delayed-shot 3d depth migration: Geophysics, **70**, E21–E28.