Source-independent Hessian-free Gauss-Newton full-waveform inversion

Wenyong Pan, Kris Innanen

ABSTRACT

Full-waveform inversion (FWI) is a powerful and promising technique for estimating the subsurface model parameters by minimizing a l-2 norm misfit function which measures the difference between the modelled data and observed data. While FWI still suffers from a lot of difficulties, one of which being the lack of source information. The estimation of source wavelet is important for successful implementation of full-waveform inversion (FWI). Many FWI algorithms estimate the source signature iteratively in the inversion process. In this paper, a source-independent method is adopted with a data calibration process. Furthermore, the gradient-based methods for FWI suffer from slow local convergence rate. A Hessian-free (HF) Gauss-Newton method is implemented in this research by solving the Newton system with a conjugate-gradient (CG) method. The Hessian-free optimization method only needs Hessian-vector products instead of constructing the Hessian matrix explicitly. In this paper, the Hessian operator in HF Gauss-Newton method is modified by combining with the source-independent strategy. We demonstrate with numerical examples that the proposed strategies can improve the convergence rate and reduce the computational burden.

INTRODUCTION

In recent years, full-waveform inversion (FWI) is becoming increasing popular in estimating the subsurface parameters by iteratively minimizing a l-2 norm misfit function, which measures the difference between the modelled data and observed data (Virieux and Operto, 2009). FWI is exposed to many difficulties, one of which being the lack of information of the source wavelet. Small perturbations of the source wavelet may result in large disturbances in inverted models (Sun et al., 2014). Hence, a robust and efficient FWI algorithm needs the information of source wavelet, which is generally unknown in seismic exploration (Virieux and Operto, 2009; Fichtner, 2010).

In traditional seismic data processing, the source wavelet is estimated from seismic data trace. Many FWI algorithms iteratively update the source wavelet estimation in the inversion process (Canadas and Kolb, 1986; Zhou and Greenhalgh, 2003; Wang et al., 2009; Sun et al., 2014). Zhou and Greenhalgh (2003) consider the source weight as unknown parameter and update it iteratively with unknown velocity. Sun et al. (2014) developed a shallow-response based variable projection type of strategy to estimate the source wavelet alongside the model parameters during FWI. Another class of approach for handling the source effects is to make the inverse problem independent of source wavelet (Lee and Kim, 2003; Li et al., 2013; Kwon et al., 2015). Lee and Kim (2003) developed a source-independent FWI strategy by normalizing the wavefield in the frequency domain with respect to the frequency response of a reference trace, in which the source spectrum is eliminated. In this paper, a variable projection method is used for data calibration in frequency-domain FWI (Golub and Pereyra, 1973; Rickett, 2013; Li et al., 2013).

The traditional optimization methods for FWI in exploration geophysics are gradient-based methods (e.g., non-linear conjugate-gradient (NCG) method). In NCG method, the search direction is just the linear combination of the current gradient and previous search direction (Fletcher and Reeves, 1964; Nocedal and Wright, 2006; Hu et al., 2011). Within the adjoint-state method, the gradient of the misfit function can be calculated efficiently by applying a zero-lag cross-correlation between the forward modelled wavefield and back-propagated data residual wavefield (Pratt et al., 1998; Pan et al., 2014c,a, 2015, 2014b). Thus, the gradient-based methods are computationally attractive for large-scale inverse problem but also suffer from slow convergence rate.

The truncated-Newton method employs second-order adjoint-state formulas to compute the Hessian-vector products instead of calculating the Hessian explicitly (Métivier et al., 2012, 2013). The search direction is obtained by solving the Newton linear system approximately with a conjugate gradient (CG) algorithm. Quasi-Newton methods do not construct the Hessian explicitly, but update the inverse Hessian approximations by storing the information from previous iterations. The BFGS method, due to Broyden (1970), Fletcher (1970), Goldfarb (1970) and Shanno (1970), is a popular quasi-Newton method to iteratively approximate the inverse Hessian. The limited-memory BFGS (l-BFGS) method stores the changes of the gradient and model from a limited number of previous iterations and uses the stored information to implicitly form an inverse of the approximated Hessian (Nocedal, 1980; Byrd et al., 1995; Nocedal and Wright, 2006). In this research, the l-BFGS optimization strategy is implemented with a "two-loop recursion" scheme (Nocedal and Wright, 2006; Anagaw, 2014; Wu et al., 2015).

In this paper, the NCG, l-BFGS and HF Gauss-Newton methods are implemented with the source-independent strategy. In particular, the Gauss-Newton Hessian is corrected with the variable projection strategy, which improves the convergence rate and reduces the computation burden for the Hessian-free Gauss-Newton method.

THEORY AND METHODS

Forward modelling problem in frequency domain

In frequency domain, the forward modelling problem in acoustic medium is governed by the following equation (Helmholtz equation) (Marfurt, 1984):

$$\nabla^2 u(\mathbf{x}, \mathbf{x}_s, \omega) + m(\mathbf{x})\,\omega^2 u(\mathbf{x}, \mathbf{x}_s, \omega) = f_s(\omega)\delta(\mathbf{x} - \mathbf{x}_s), \tag{1}$$

where ω is the angular frequency, $\mathbf{x} = (x, y, z)$ denotes the subsurface location with Cartesian coordinates, ∇^2 is the Laplacian operator, $m(\mathbf{x})$ is the model parameter (square of slowness), $u(\mathbf{x}, \mathbf{x}_s, \omega)$ denotes the pressure wavefield at position $\mathbf{x}$, $\delta(\mathbf{x} - \mathbf{x}_s)$ is the Dirac delta function, and $f_s(\omega)$ means the source signature at position $\mathbf{x}_s$. The solution of equation (1) can be written as the convolution of source $f_s(\omega)$ with Green's function $G(\mathbf{x}, \mathbf{x}_s, \omega)$:

$$u(\mathbf{x}, \mathbf{x}_s, \omega) = f_s(\omega)\, G(\mathbf{x}, \mathbf{x}_s, \omega), \tag{2}$$

where the Green's function is defined as the solution of wave equation due to an impulse source. A 9-point finite difference scheme is employed to discretize the model (Jo et al.,

1996) and a Engquist-Majda boundary condition (first-order) is applied on all of the boundaries of the model (Engquist and Majda, 1977). Equation (1) can be rewritten in a matrix form after discretization:

$$\mathbf{A}\left(\mathbf{m},\omega\right)\mathbf{u}(\mathbf{x},\mathbf{x}_s,\omega)=\mathbf{f}_s(\omega)\delta\left(\mathbf{x}-\mathbf{x}_s\right), \tag{3}$$

where $\mathbf{m}$ is the square of slowness vector, $\mathbf{u}(\mathbf{x},\mathbf{x}_s,\omega)$ and $\mathbf{f}_s(\omega)$ are the discrete pressure wavefield and source vectors. $\mathbf{A}\left(\mathbf{m},\omega\right)$ is the discretized impedance matrix, which is typically sparse and symmetric. In this research, the linear equation (equation (3)) is solved with a direct solver based on multi-frontal Lower Upper (LU) decomposition (Davis and Duff, 1997), which is efficient for a multi-source problem with forward and backward substitutions (Hu et al., 2011).

Non-linear least-squares inversion

FWI allows to reconstruct high-resolution velocity models of the subsurface through the extraction of the full information content of the seismic data. The inversion process is implemented by minimizing a l-2 norm misfit function, which measures the difference between the modelled data and observed data:

$$\Phi\left(\mathbf{m}\right)=\frac{1}{2}\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\sum_{\omega}\left\|\mathbf{d}_{\text{obs}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)-\mathbf{d}_{\text{syn}}\left(\mathbf{m},\mathbf{x}_s,\mathbf{x}_g,\omega\right)\right\|^2, \tag{4}$$

where ω is the angular frequency, $\mathbf{m}$ indicates the model parameter vector, $\mathbf{x}_s$ and $\mathbf{x}_g$ are the source and receiver positions, $\mathbf{d}_{\text{obs}}$ and $\mathbf{d}_{\text{syn}}=\mathcal{P}\mathbf{u}$ indicate the observed data and synthetic data respectively, $\mathcal{P}$ is the sampling operator, and $\|\cdot\|$ means the l-2 norm. Here, we consider to minimize the objective function associated with the source weight:

$$\tilde{\Phi}\left(\mathbf{m},\mathbf{s}\right)=\frac{1}{2}\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\sum_{\omega}\left\|\mathbf{d}_{\text{obs}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)-\mathbf{s}\left(\mathbf{x}_s,\omega\right)\mathbf{d}_{\text{syn}}\left(\mathbf{m},\mathbf{x}_s,\mathbf{x}_g,\omega\right)\right\|^2, \tag{5}$$

where $\mathbf{s}\left(\mathbf{x}_s,\omega\right)$ indicates the source weight vector.

The variable projection method is implemented by minimizing the objective function (equation (5)) with respect to both $\mathbf{m}$ and $\mathbf{s}$. We suppose that $\mathbf{d}_{\text{obs}}$ and $\mathbf{d}_{\text{syn}}$ are known, the source weight $\mathbf{s}$ can be obtained through a least-square approach (Li et al., 2013):

$$\mathbf{s}\left(\mathbf{x}_s,\omega\right)=\frac{\sum_{\mathbf{x}_g}\mathbf{d}_{\text{obs}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)\mathbf{d}^*_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)}{\sum_{\mathbf{x}_g}\mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)\mathbf{d}^*_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)}, \tag{6}$$

where the symbol "*" means complex conjugate transpose. The Jacobian matrix is given as:

$$\mathbf{J}=\frac{\partial\mathbf{d}_{\text{syn}}}{\partial\mathbf{m}}=\mathcal{P}\mathbf{A}\left(\mathbf{m},\omega\right)^{-1}\mathbf{A}_{\mathbf{m}}\left(\mathbf{m},\omega\right)\mathbf{u} \tag{7}$$

where $\mathbf{A}_{\mathbf{m}}\left(\mathbf{m},\omega\right)=\partial\mathbf{A}\left(\mathbf{m},\omega\right)/\partial\mathbf{m}$. The gradient for minimizing objective function in equation (4) is obtained as:

$$\mathbf{g}=\frac{\partial\Phi\left(\mathbf{m}\right)}{\partial\mathbf{m}}=\sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\sum_{\omega}\mathbf{u}^*\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)\mathbf{A}^*_{\mathbf{m}}\left(\mathbf{m},\omega\right)\mathbf{A}^*\left(\mathbf{m},\omega\right)^{-1}\mathcal{P}^*\Delta\mathbf{d}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right), \tag{8}$$

The approximate Hessian used in Gauss-Newton method is expressible as:

$$\mathbf{H}_a = \mathbf{J}^*\mathbf{J} = \mathbf{u}^*\mathbf{A}_{\mathbf{m}}^* \left(\mathbf{m},\omega\right)\mathbf{A}^* \left(\mathbf{m},\omega\right)^{-1}\mathcal{P}^*\mathcal{P}\mathbf{A}\left(\mathbf{m},\omega\right)^{-1}\mathbf{A}_{\mathbf{m}}\left(\mathbf{m},\omega\right)\mathbf{u}. \tag{9}$$

When considering the objective function illustrated in equation (5), the Jacobian matrix becomes (Li et al., 2013):

$$\begin{aligned}
\tilde{\mathbf{J}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right) &= \frac{\partial \mathbf{s}\left(\mathbf{x}_s,\omega\right)\mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)}{\partial \mathbf{m}} \\
&= \mathbf{s}\left(\mathbf{x}_s,\omega\right)\frac{\partial \mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)}{\partial \mathbf{m}} + \frac{\partial \mathbf{s}}{\partial \mathbf{m}}\mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right) \\
&= \mathbf{s}\left(\mathbf{x}_s,\omega\right)\mathbf{J}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right) + \frac{\partial}{\partial \mathbf{m}}\left(\frac{\sum_{\mathbf{x}'_g}\mathbf{d}_{\text{obs}}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)\mathbf{d}^*_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)}{\sum_{\mathbf{x}'_g}\mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)\mathbf{d}^*_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)}\right)\mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right) \\
&= \mathbf{s}\left(\mathbf{x}_s,\omega\right)\mathbf{J}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right) + \frac{\mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)}{\sum_{\mathbf{x}'_g}\left|\mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)\right|^2} \\
&\times \left(\sum_{\mathbf{x}'_g}\mathbf{d}_{\text{obs}}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)\frac{\partial \mathbf{d}^*_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)}{\partial \mathbf{m}} - 2\mathbf{s}\left(\mathbf{x}_s,\omega\right)\sum_{\mathbf{x}'_g}\mathbf{d}_{\text{syn}}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)\mathbf{J}\left(\mathbf{x}_s,\mathbf{x}'_g,\omega\right)\right).
\end{aligned} \tag{10}$$

The corresponding gradient and Gauss-Newton Hessian are expressed as:

$$\tilde{\mathbf{g}} = \frac{\partial \tilde{\Phi}\left(\mathbf{m},\mathbf{s}\right)}{\partial \mathbf{m}} = \sum_{\mathbf{x}_s}\sum_{\mathbf{x}_g}\sum_{\omega}\tilde{\mathbf{J}}^*\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right)\Delta\mathbf{d}\left(\mathbf{x}_s,\mathbf{x}_g,\omega\right). \tag{11}$$

$$\tilde{\mathbf{H}}_a = \tilde{\mathbf{J}}^*\tilde{\mathbf{J}}. \tag{12}$$

Source-independent Hessian-free Gauss-Newton

Instead of constructing Hessian approximations or approximating inverse Hessian, the Hessian-free (HF) optimization method (truncated-Newton or inexact-Newton method), obtains the search direction by solving the Newton linear system approximately using a conjugate-gradient (CG) method with matrix-free fashion (Saad, 2003; Métivier et al., 2014). The CG method is an optimal algorithm for solving a symmetric positive definite system and it only requires computing the Hessian-vector products $\mathbf{H}\upsilon$ instead of forming the Hessian matrix explicitly. The Hessian-vector products can be calculated via finite-difference method (Nocedal and Wright, 2006) or the second-order adjoint-state method (Métivier et al., 2014). In this paper, the second-order adjoint method is employed for Hessian-vector products calculation. In the context of HF optimization method, the Hessian $\mathbf{H}$ is always replaced with Gauss-Newton Hessian $\tilde{\mathbf{H}}$, which is always symmetric and positive definite:

$$\mathbf{H}_a^k\Delta\mathbf{m}^k = -\mathbf{g}^k, \tag{13}$$

where k indicates the iteration index. Thus, the HF Gauss-Newton FWI can be implemented in a double-iterative scheme: the outer loop is to iteratively update the model parameters for the non-linear optimization problem, and the inner loop is to solve the linear system (equation(13)) iteratively with the CG algorithm. The inner iteration is typically stopped or "truncated" before the solution of the Newton equation is obtained. A Hessian-free optimization method will be more competitive if further enhancements are used, for example, an effective preconditioner for the linear system and appropriate stopping criteria for the inner iterative algorithm. With these enhancements, Hessian-free optimization method is a powerful tool for large-scale inverse problem. In this paper, a l-BFGS preconditioning scheme is used for the Hessian-free Gauss-Newton FWI. The source-independent HF Gauss-Newton FWI is implemented by replacing $\mathbf{H}_a^k$ with $\tilde{\mathbf{H}}_a^k$ and replacing $\mathbf{g}^k$ with $\tilde{\mathbf{g}}^k$:

$$\tilde{\mathbf{H}}_a^k\Delta\mathbf{m}^k = -\tilde{\mathbf{g}}^k. \tag{14}$$

NUMERICAL EXPERIMENTS

In this section, we first apply the source-independent Hessian-free Gauss-Newton (SI-HF-GN) FWI on a Gaussian-anomaly model in comparison with non-linear conjugate gradient (NCG), l-BFGS and HF-GN methods. The inversion results verify the effectiveness of the variable projection strategy in improving the convergence rate and reducing the computation cost.

The Gaussian-anomaly model

In this numerical example, the NCG, l-BFGS and HF Gauss-Newton methods are applied on a Gaussian-anomaly model. We examine the quadratic convergence rate of the HF Gauss-Newton FWI in reconstructing the velocity model compared to NCG and l-BFGS methods. We also demonstrate that with the corrected Gauss-Newton Hessian (equation (14)), the convergence rate of the Hessian-free Gauss-Newton method is improved and the computation burden is reduced.

The Gaussian-anomaly model consists of 50×100 grid cells with a grid interval of 10 m in both horizontal and vertical directions. A total of 49 sources are deployed from 20 m to 980 m with a source interval of 20 m and a depth of 20 m. A total of 100 receivers are distributed from 10 m to 1000 m with a receiver interval of 10 m and a depth of 20 m. A Ricker wavelet with a 30 Hz dominant frequency is used as the source function. Figure 1 shows the true Gaussian-anomaly P-wave velocity model. The initial velocity model is a homogeneous model with a constant velocity of 2 km/s.

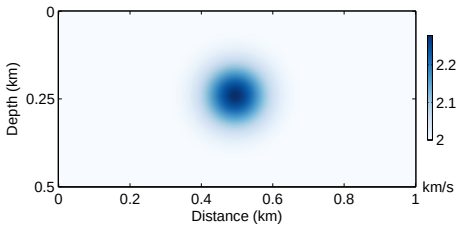

FIG. 1. The true Gaussian-anomaly model.

For comparison of different methods, we set the stopping criteria for inversion as: the maximum outer iteration $k_{max} = 21$ and the minimum normalized misfit $\phi_{min} = 2.0e$-5. To illustrate the performances of different preconditioning strategies, the stopping criteria for the inner CG algorithm is also defined: the maximum inner iteration $\tilde{k}_{max} = 50$ and the minimum relative residual $\gamma_{min} = 1.0e$-2. Figures 2a, 2b, 2c and 2d show the inversion results by SD, l-BFGS, HF-GN and SI-HF-GN methods respectively. Figures 3a and 3b show the normalized misfit vs. iteration and RLSE vs. iteration as the iteration proceeds.

As we can see, compared to SD and l-BFGS methods, the HF Gauss-Newton methods can approach the true model much better. Furthermore, we notice that we the corrected Gauss-Newton Hessian, the source-independent HF Gauss-Newton method can converge faster. We use both of the misfit and RLSE $\epsilon = \frac{\|\mathbf{m}_k - \mathbf{m}_0\|^2}{\|\mathbf{m}_{\text{true}} - \mathbf{m}_0\|^2}$ to evaluate the quality of the inversion result.

We then apply the source-independent Hessian-free Gauss-Newton FWI on a modified Marmousi model. The modified Marmousi model consists of 160×350 grid cells with a grid interval of 10 m in both horizontal and vertical directions. A total of 34 sources are deployed from 10 m to 340 m with a source interval of 10 m and a depth of 20 m. A total of 350 receivers are distributed from 10 m to 3500 m with a receiver interval of 10 m and a depth of 20 m. A Ricker wavelet with a 30 Hz dominant frequency is used as the source function. Figure 4a shows the true P-wave velocity model. The initial velocity model shown in Figure 4a is obtained by smoothing the true P-wave velocity model with a Gaussian function.

The inversion results by NCG, l-BFGS and HF Gauss-Newton and source-independent HF Gauss-Newton methods are shown in Figures 5a, 5b, 5c and 5d. Figure 6 show the model differences between the true model and the inversion results by NCG, l-BFGS and HF Gauss-Newton and source-independent HF Gauss-Newton methods respectively. We observe that the Gauss-Newton Hessian correction, a better inversion result is achieved.

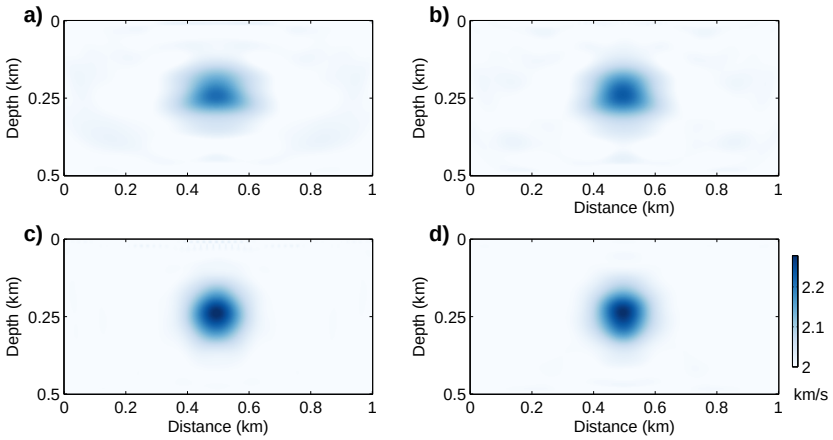

FIG. 2. (a) NCG method; (b) l-BFGS method; (c) HF-GN method; (d) SI-HF-GN method.

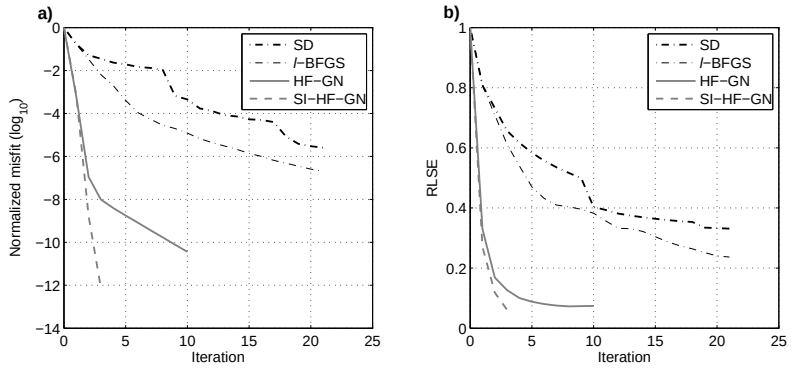

FIG. 3. (a) Normalized misfit ($\log_{10}$) vs. Iteration; (b) RLSE vs. Iteration. The bold-dash-dot, thin-dash-dot, bold-solid and bold-dahs lines indicate SD, l-BFGS, HF-GN, and SI-HF-GN methods respectively.

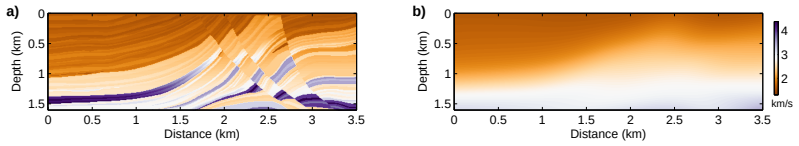

FIG. 4. (a) The true P-wave velocity model; (b) The initial P-wave velocity model.

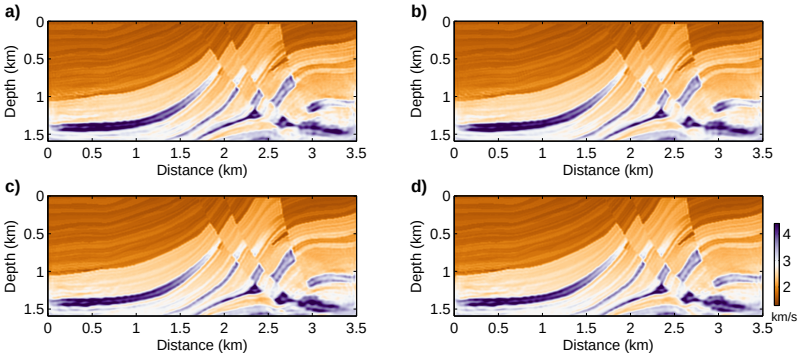

FIG. 5. (a) NCG method; (b) l-BFGS method; (c) Hessian-free GN method ($\epsilon = 0.3793$); (d) Source-independent Hessian-free GN method with corrected Gauss-Newton Hessian ($\epsilon = 0.3528$).

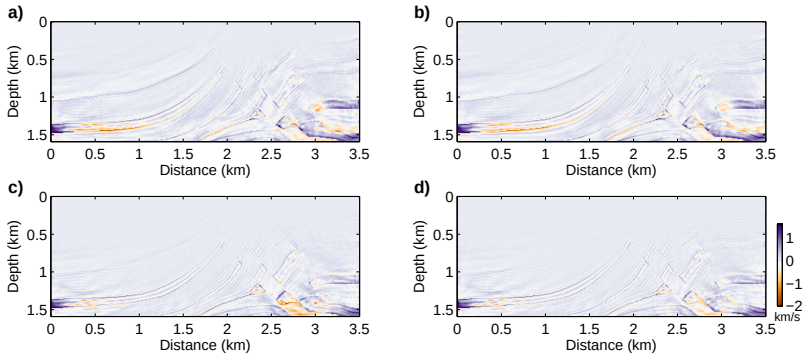

FIG. 6. The model differences between the true model and the inverted results by (a) NCG method; (b) l-BFGS method; (c) Hessian-free GN method; (d) Source-independent Hessian-free GN method with corrected Gauss-Newton Hessian.

CONCLUSIONS

In this paper, we develop a source-independent Hessian-free Gauss-Newton FWI, in which the Gauss-Newton Hessian is corrected with the source-independent strategy. The numerical examples show that the proposed method can provides faster convergence rate and better inversion result with reducing the computational burden.

ACKNOWLEDGEMENTS

The authors thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 461179-13. Author 1 was also supported by a SEG/Chevron scholarship.

REFERENCES

Anagaw, A. Y., 2014, Full waveform inversion using simultaneous encoded sources based on first and second-order optimization methods: Ph.D. thesis, University of Alberta.

Broyden, C. G., 1970, The convergence of a class of double-rank minimization algorithms: IMA Journal of Applied Mathematics, **6**, 222–231.

Byrd, R. H., Lu, P., and Nocedal, J., 1995, A limited memory algorithm for bound constrained optimization: SIAM Journal on Scientific and Statistical Computing, **16**, 1190–1208.

Canadas, G., and Kolb, P., 1986, Least-squares inversion of prestack data: simultaneous identification of density and velocity of stratified data: SEG Expanded Abstracts, 1–6.

Davis, T. A., and Duff, I. S., 1997, An unsymmetric pattern multifrontal method for sparse lu factorization: SIAM Journal on Matrix Analysis and Applications, **18**, 140–158.

Engquist, B., and Majda, A., 1977, Absorbing boundary conditions for the numerical simulations of waves: Mathematical Computation, **31**, 629–651.

Fichtner, A., 2010, Full seismic waveform modelling and inversion: Springer.

Fletcher, R., 1970, A new approach to variable metric algorithms: The Computer Journal, **13**, 317–322.

Fletcher, R., and Reeves, C. M., 1964, Function minimization by conjugate gradients: Comp. J., **7**, 149–154.

Goldfarb, D., 1970, A family of variable-metric methods derived by variational means: Mathematics of Computation, 23–26.

Golub, G., and Pereyra, V., 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variable separate: SIAM Journal on Numerical Analysis, **10**, 413–432.

Hu, W., Abubakar, A., Habashy, T. M., and Liu, J., 2011, Preconditioned non-linear conjugate gradient method for frequency domain full-waveform seismic inversion: Geophysical Prospecting, **59**, 477–491.

Jo, C. H., Shin, C., and Suh, J. H., 1996, An optimal 9-point, finite-difference, frequency-space, 2-d scalar wave extrapolator: Geophysics, **61**, 529–537.

Kwon, T., Seol, S. J., and Byun, J., 2015, Efficient full-waveform inversion with normalized plane-wave data: Geophysical Journal International, **201**, 53–60.

Lee, K. H., and Kim, H. J., 2003, Source-independent full-waveform inversion of seismic data: Geophysics, **68**, 45–59.

Li, M., Rickett, J., and Abubakar, A., 2013, Application of the variable projection scheme for frequency-domain full-waveform inversion: Geophysics, **78**, R249–R257.

Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, **49**, 533–549.

Métivier, L., Bretaudeau, F., Brossier, R., Virieux, J., and Operto, S., 2014, Full waveform inversion and the truncated Newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, 1–23.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2012, The truncated Newton method for full waveform inversion: SEG Technical Program Expanded Abstracts, 1–5.

Métivier, L., Brossier, R., Virieux, J., and Operto, S., 2013, Full waveform inversion and the truncated Newton method: SIAM Journal On Scientific Computing, **35**, B401–B437.

Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.

Nocedal, J., and Wright, S. J., 2006, Numerical Optimization: Springer.

Pan, W., Innanen, K. A., and Margrave, G. F., 2014a, A comparison of different scaling methods for least-squares migration/inversion: EAGE Expanded Abstracts, We G103 14.

Pan, W., Innanen, K. A., and Margrave, G. F., 2014b, Efficient full waveform inversion in the time-ray parameter domain using iteration-dependent sets of ray parameters: CPS/SEG Technical Program Expanded Abstracts, 1–5.

Pan, W., Innanen, K. A., Margrave, G. F., and Cao, D., 2015, Efficient pseudo-Gauss-Newton full-waveform inversion in the τ-p domain: Geophysics, **80**, R225–R14.

Pan, W., Margrave, G. F., and Innanen, K. A., 2014c, Iterative modeling migration and inversion (IMMI): Combining full waveform inversion with standard inversion methodology: SEG Technical Program Expanded Abstracts, 938–943.

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.

Rickett, J., 2013, The variable projection method for waveform inversion with an unknown source function: Geophysical Prospecting, **61**, 874–881.

Saad, Y., 2003, Iterative methods for sparse linear systems: SIAM.

Shanno, D. F., 1970, Conditioning of quasi-Newton methods for function minimization: Mathematics of Computation, **24**, 647–656.

Sun, D., Jiao, K., Vigh, D., and Coates, R., 2014, Source wavelet estimation in full waveform inversion: EAGE Expanded Abstracts, Tu E106 09.

Virieux, A., and Operto, S., 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, WCC1–WCC26.

Wang, K., Krebs, J. R., Hinkley, D., and Baumstein, A., 2009, Simultaneous full-waveform inversion for source wavelet and earth model: SEG Expanded Abstracts, 1–6.

Wu, S., Wang, Y., Zheng, Y., and Chang, X., 2015, Limited-memory BFGS based least-squares pre-stack Kirchhoff depth migration: Geophysical Journal International, **202**, 738–747.

Zhou, B., and Greenhalgh, S. A., 2003, Crosshole seismic inversion with normalized full-waveform amplitude data: Geophysics, **68**, 1320–1330.