Accelerating Hessian-free Gauss-Newton full-waveform inversion via improved preconditioning strategies
Wenyong Pan, Kristopher A. Innanen, Department of Geoscience, CREWES Project, University of Calgary, Wenyuan Liao, Department of Mathematics and Statistics, University of Calgary

SUMMARY
Gradient-based methods for full-waveform inversion (FWI) have the potential to converge globally but suffer from a slow convergence rate. Newton-type methods provide quadratic convergence, but they are computationally burdensome for large-scale inverse problems. The Hessian-free (HF) optimization method represents an attractive alternative to these above-mentioned optimization methods. In this research, we propose to construct the signed as diagonal Hessian approximations or inverse Hessian approximations. In this research, we propose to construct the HF Gauss-Newton method. Traditionally, the preconditioners are designed as diagonal Hessian approximations or inverse Hessian approximations. In this research, we propose to construct the \( l \)-BFGS inverse Hessian preconditioner with the diagonal Hessian approximations as initial guess. As is shown, the quasi-Newton \( l \)-BFGS preconditioning scheme with the pseudo diagonal Gauss-Newton Hessian as initial guess shows the best performances in accelerating the HF Gauss-Newton FWI.

INTRODUCTION
Seismic full-waveform inversion (FWI) holds the promise of providing high-resolution estimates of subsurface properties. FWI iteratively reconstructs the model parameters by minimizing an L2 norm misfit function (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). Traditional optimization methods for FWI in exploration geophysics are gradient-based methods (i.e., steepest-descent (SD) and non-linear conjugate-gradient (NCG) methods), which are computationally attractive for large-scale inverse problems. However, they suffer from slow convergence rates.

The search direction can be significantly enhanced by multiplying the gradient with the inverse Hessian matrix (Pratt et al., 1998). For multi-parameter FWI, the multi-parameter Hessian is also expected to suppress parameter cross-talk (Operto et al., 2013; Innanen, 2014a,b; Pan et al., 2015c). Furthermore, the second-order term in the Hessian matrix can help to remove the second-order scattering artifacts in the gradient (Pratt et al., 1998; Pan et al., 2015b,c). However, explicit calculation, storage and inversion of the Hessian at each iteration is computationally impractical for large-scale inverse problems. Hence, various approaches have been proposed for approximating the Hessian (Shin et al., 2001a) or inverse Hessian (Nocedal and Wright, 2006; Nammour and Symes, 2009; Demanet et al., 2012). In Gauss-Newton method, an approximate Hessian is introduced by involving only the first-order term (Pratt et al., 1998; Pan et al., 2015a) used phase-encoding technology to construct the diagonal Gauss-Newton Hessian efficiently. Shin et al. (2001a) proposed a pseudo-Hessian approximation by replacing the Fréchet derivative wavefield with the virtual source during the auto-correlation process. Preconditioning the gradient with the diagonal pseudo-Hessian resembles a deconvolution imaging condition (Pan et al., 2014, 2015a).

Instead of constructing the Hessian explicitly, the quasi-Newton \( l \)-BFGS methods approximate the inverse Hessian iteratively by storing the model and gradient changes from a number of \( M \) \((M < 10)\) previous iterations (Nocedal and Wright, 2006). Compared to gradient-based methods, \( l \)-BFGS methods provide faster convergence rates for large-scale inverse problems (Brossier et al., 2010; Ma and Hale, 2012). The convergence performance of \( l \)-BFGS method is closely related to the initial guess of inverse Hessian approximation (Brossier et al., 2009; Guitton and Díaz, 2012). In the numerical section, we give examples to examine the convergence rates of the \( l \)-BFGS methods with different diagonal Hessian approximations as initial guess.

The Hessian-free optimization method represents an attractive alternative to the above-described optimization methods (Nash, 1985; Santosa and Symes, 1988). At each iteration, the search direction is computed by approximately solving the Newton equations using a matrix-free fashion of the conjugate-gradient (CG) algorithm (Nash, 1985; Hu et al., 2009). This linear iterative solver only requires the Hessian-vector products instead of forming the Hessian operator explicitly (Fichtner and Trampert, 2011; Métivier et al., 2014). In this paper, the full Hessian is replaced with the Gauss-Newton Hessian, which is always symmetric and positive semi-definite. One issue of the HF optimization method is that obtaining the search direction approximately requires a large number of CG iterations. Our main goal in this paper is to precondition the CG algorithm, reduce the number of CG iterations, and accelerate the HF Gauss-Newton FWI (Nash, 2000; Sainath et al., 2013).

Preconditioning makes the problem well-conditioned, hence easier to solve, and it reduces the number of CG iterations. The preconditioner for the CG algorithm is designed by approximating the Hessian or its inverse (Nash, 2000). Different preconditioning schemes are developed for comparison in this paper. Diagonal pseudo-Hessian and diagonal Gauss-Newton Hessian are first considered as preconditioners for the CG algorithm. A pseudo diagonal Gauss-Newton Hessian approximation is also introduced as the preconditioner. Quasi-Newton \( l \)-BFGS inverse Hessian approximations also serve as effective preconditioners for CG iterative solver (Nash, 1985). In this paper, we furthermore propose that the \( l \)-BFGS inverse Hessian preconditioners be constructed with diagonal Hessian approximations as initial guesses. We demonstrate with numerical examples that the \( l \)-BFGS inverse Hessian preconditioning strategy can accelerate the HF Gauss-Newton method effectively and furthermore, employing the diagonal Hessian approximations as initial guess can improve its efficiency further.

THE NON-LINEAR LEAST-SQUARES INVERSE PROBLEM
FWI seeks to estimate the subsurface parameters by iteratively minimizing the difference between the synthetic data \( d_{\text{syn}} \) and observed data \( d_{\text{obs}} \) (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). The misfit function \( \Phi \) is formulated in a least-squares form:

\[
\Phi(m) = \frac{1}{2} \sum_{i} \sum_{x_i} \sum_{s} \sum_{o} \| \Delta d(x_i, s, o) \|_2^2,
\]

where \( \Delta d = d_{\text{obs}} - d_{\text{syn}} \) is the data residual vector, and \( \| \cdot \| \) means the L2 norm. Here, the synthetic data \( d_{\text{syn}} \) is related to the seismic wavefield \( u \) by a detection operator \( \mathcal{P} \), which samples the wavefield at the receiver locations: \( d_{\text{syn}} = \mathcal{P} u \). To minimize the quadratic approximation of the misfit function, the updated model at the \( (k+1) \)th iteration can be written as the sum of the model at the \( k \)th iteration and the search direction \( \Delta m_k \):

\[
m_{k+1} = m_k + \mu_k \Delta m_k,
\]

where \( \mu_k \) is the step length, a scalar constant calculated through a line search method satisfying the weak Wolfe condition (Nocedal and Wright, 2006). Within a Newton optimization framework, the search direction \( \Delta m_k \) is the solution of the Newton linear system:

\[
H_k \Delta m_k = -g_k.
\]
where \( g \) and \( H \) indicate gradient and Hessian respectively. Within the adjoint-state formulation (Plessix, 2006), the gradient can be expressed as (Siringue and Pratt, 2004; Pan et al., 2015a):

\[
g(x) = \sum_{s} \sum_{x} \sum_{\omega} \mathcal{R}\left(\omega^2 f(x, \omega) G(x, x, \omega) G(x, x, \omega) \Delta x^\lambda (x, x, \omega)\right),
\]

(4)

where \( G(x, x, \omega) \) and \( G(x, x, \omega) \) indicate source-side and receiverside Green’s functions respectively. Following equation (4), the gradient can be constructed efficiently by cross-correlating the forward modelled wavefield with the back-propagated data residual wavefield (Virieux and Operto, 2009). The gradient is poorly-scaled due to geometrical spreading, and it is also contaminated by spurious correlations because of finite-frequency effects and doubly-scattered energy (Pratt et al., 1998). Multiplying the gradient with the inverse Hessian can greatly enhance the model update, which provides a quadratic convergence rate.

The Gauss-Newton approximate Hessian \( \tilde{H} \) is constructed by correlating two Fréchet derivative wavefields, which only accounts for the first-order scattering effects (Tang, 2009):

\[
\tilde{H}(x, x') = \sum_{s} \sum_{x} \sum_{\omega} \mathcal{R}\left(\omega^4 f(x, \omega) G(x, x, \omega) G(x, x, \omega) \right) \times G^\prime (x', x, \omega) G^\prime (x, x, \omega),
\]

(5)

where the element \( \tilde{H}(x, x') \) in the Gauss-Newton Hessian is formed by correlating the two Fréchet derivative wavefields at the receivers’ locations due to model perturbations at positions \( x \) and \( x' \).

For these Newton-type methods, explicit evaluation and inversion the Hessian matrix \( H \) and Gauss-Newton Hessian \( \tilde{H} \) at each iteration are required. Though Newton-type methods benefit from fast convergence rate, the computation, storage and inversion of Hessian at each iteration are prohibitively expensive for large-scale inverse problems. Gradient-based methods approximate the Hessian matrix \( H \) as an identity matrix \( I \) and they are computationally more attractive but they suffer from slow convergence rates.

**HESSIAN-FREE OPTIMIZATION METHOD**

Instead of constructing Hessian or inverse Hessian approximations, the Hessian-free (HF) optimization method obtains the search direction by solving the Newton linear system approximately using a conjugate-gradient (CG) method (Saad, 2003; Anagaw and Sacchi, 2012). The CG method is an optimal algorithm for solving a symmetric positive definite system \( Wx=b \) and it only requires computing the Hessian-vector products \( \tilde{f}(x) \) (Fichtner and Trampert, 2011):

\[
\tilde{f}(x) = \sum_{s} H(x, x') u(x'),
\]

(6)

where \( u(x') \) is an arbitrary vector. In this paper, the adjoint-state method is employed to calculate the Hessian-vector products. In the context of HF optimization method, the Hessian \( H \) is always replaced with Gauss-Newton Hessian \( \tilde{H} \), which is always symmetric and positive semi-definite:

\[
(\tilde{H} + \varepsilon \tilde{A}) \Delta x_k = -g_k,
\]

(7)

where \( \varepsilon \tilde{A} \) is the damping term ensuring that \( \tilde{H} + \varepsilon \tilde{A} \) is positive definite, \( \varepsilon \) is a small constant value and \( \tilde{A} \) indicates a diagonal matrix consisting of the diagonal elements of the Gauss-Newton Hessian. The resulting algorithm becomes a Levenberg-Marquardt method (Levenberg, 1944; Marquardt, 1963). A Hessian-free optimization method can be made more competitive with further enhancements, such as, an effective preconditioner for the linear system and appropriate stopping criteria for the inner iterative algorithm.

**Preconditioning**

The CG iterative algorithm requires many iterations to obtain the approximate solution of a linear system. The convergence rate of the CG method depends on the spectral properties of the coefficient matrix (Nash, 2000). It is often convenient to transform the equation system into one which has the same solution but more favorable spectral properties. This can be achieved by applying a suitable preconditioner \( \mathcal{M} \) on the linear system: \( \mathcal{M}^{-1}Wx = \mathcal{M}^{-1}b \). Thus, the preconditioned Newton system for the HF Gauss-Newton FWI is given by:

\[
\mathcal{M}^{-1}(\tilde{H} + \varepsilon \tilde{A}) \Delta x_k = -\mathcal{M}^{-1}g_k.
\]

(8)

The solution of equation (8) can be obtained by the preconditioned conjugate-gradient (PCG) method, which is expected to reduce the number of inner iterations, improve the convergence rate and accelerate the HF Gauss-Newton FWI.

**Diagonal Hessian Approximation Preconditioners**

The preconditioner for the CG method is always devised to approximate the Hessian or the inverse Hessian. We first consider the traditional Hessian approximations (e.g., diagonal pseudo-Hessian and diagonal Gauss-Newton Hessian) as the preconditioners. The pseudo-Hessian \( H \) is constructed by replacing the Fréchet derivative wavefield with the virtual source \( f(x, \omega) \) in the correlation process (Shin et al., 2001b):

\[
\tilde{f}(x, \omega) = -\omega^4 f(x, \omega) G(x, x, \omega).
\]

(9)

The diagonal pseudo-Hessian is obtained by auto-correlating two virtual sources:

\[
\mathcal{H}_{\text{diag}}(x) = \sum_{s} \sum_{x} \sum_{\omega} \mathcal{R}\left(\omega^4 |f(x, \omega)|^2 G(x, x, \omega) G^\prime (x, x, \omega)\right),
\]

(10)

Constructing the diagonal pseudo-Hessian at each iteration does not involve any additional cost. While when employing the diagonal Gauss-Newton Hessian \( \mathcal{H}_{\text{diag}} \) as a preconditioner, more computation cost is required for constructing the receiverside Green’s functions.

In this paper, we introduce a pseudo diagonal Gauss-Newton Hessian approximation \( \mathcal{H}_{\text{diag}} \) as the preconditioner for the CG algorithm in the inner loop by assuming that the sources and receivers are co-located. The pseudo diagonal Gauss-Newton Hessian is written as:

\[
\tilde{H}_{\text{diag}}(x) = \sum_{s} \sum_{x} \sum_{\omega} \mathcal{R}\left(\omega^4 |f(x, \omega)|^2 G(x, x, \omega) G^\prime (x, x, \omega)\right).
\]

(11)

This diagonal approximation can be constructed at no additional cost. Summarily, the diagonal pseudo-Hessian, diagonal Gauss-Newton Hessian and pseudo diagonal Gauss-Newton Hessian preconditioners are given by:

\[
\mathcal{M}^{-1}\mathcal{H}_{\text{diag}}^\text{DPH} + \lambda \mu_k,
\]

(12)

\[
\mathcal{M}^{-1}\mathcal{H}_{\text{diag}}^\text{DGH} + \lambda \mu_k,
\]

(13)

\[
\mathcal{M}^{-1}\mathcal{H}_{\text{diag}}^\text{PDGH} + \lambda \mu_k,
\]

(14)

where \( \mu_k \) and \( \mu_k^\text{PDGH} \) are the stabilization terms. These three different preconditioning strategies are referred to as DPH-GN, DGH-GN and PDGH-GN in this paper. When the parameter \( \lambda \) is very large, these preconditioning methods approach the non-preconditioned HF Gauss-Newton (CG-GN) method.

**l-BFGS Inverse Hessian Preconditioners**

Furthermore, we develop an l-BFGS preconditioning scheme for the HF optimization method, namely l-BFGS-GN method (\( \mathcal{M}_L = I \)). In BFGS method, we are given a symmetric and positive definite matrix \( \mathcal{M}_L \) that approximates the inverse of the Hessian, and a pair of vectors \( s_k = m_{k+1} - m_k \) and \( y_k = g_{k+1} - g_k \) that indicates the model and...
Preconditioned HF Gauss-Newton FWI

Gradient changes. The inverse Hessian approximation \( \mathcal{H}_{k+1} \) is given by:

\[
\mathcal{H}_{k+1} = \mathcal{H}_k + \frac{1}{\gamma_k} \mathcal{H}_k \mathcal{A}_k + \frac{1}{\gamma_k} \mathcal{A}_k \mathcal{H}_k, \quad (15)
\]

where \( w_k = 1/\gamma_k s_k \) and \( v_k = I - w_k \mathcal{H}_k s_k \). A limited-memory BFGS (l-BFGS) method is always developed by storing the model and gradient changes from a limited number of previous iterations (typically \( M < 10 \)) (Nocedal, 1980). The approximated inverse Hessian \( \mathcal{H} \) can also be used as a preconditioner for the CG iterative method:

\[
\mathcal{H} (l + e \mathcal{A}) \Delta m_k = - \mathcal{H} g_k, \quad (16)
\]

Traditionally, an identity matrix \( I \) is usually set as the initial guess \( \mathcal{H}_0 \). While, the initial guess \( \mathcal{H}_0 \) is closely related to the performance of \( l \)-BFGS method. In this paper, to improve the \( l \)-BFGS preconditioning scheme, we consider using the stabilized diagonal pseudo-Hessian, diagonal Gauss-Newton Hessian and pseudo diagonal Gauss-Newton Hessian as the initial guess for constructing the \( l \)-BFGS preconditioners. These methods are referred to as \( l \)-BFGS-\( GN \), \( l \)-BFGS-\( GN \)-\( DPH \) and \( l \)-BFGS-\( GN \)-\( PDGH \) methods respectively.

“Over-solving” the Newton equation will not produce a better search direction (Nash, 2000). The CG algorithm should be terminated with an appropriate stopping criteria. We can define the maximum inner iteration number \( k_{\text{max}} \) and relative residual \( \gamma_k \):

\[
\gamma_k = \frac{\| \mathcal{H}_k \Delta m_k + g_k \|}{\| g_k \|}. \quad (17)
\]

where \( k \) indicates the CG inner iteration index. The inner iteration is stopped when \( \gamma_k < \gamma_{\text{min}} \), where \( \gamma_{\text{min}} \) indicates the relative residual tolerance.

NUMERICAL EXAMPLES

A modified Marmousi model is used to examine the efficiency of different preconditioning schemes for the HF Gauss-Newton FWI. The truncated Marmousi model has \( 100 \times 100 \) grid cells with a grid interval of \( 10 \) m in both horizontal and vertical directions. We deploy 49 sources from 20 m to 980 m at depth of 20 m with a regular source spacing of 20 m. Fifty receivers are arranged from 10 m to 1000 m every 20 m at the depth of 20 m. The source function is a Ricker wavelet with a dominant frequency of 30 Hz. Figures 1a and 1b show the true P-wave velocity model and initial P-wave velocity model. The frequencies used for inversion are increased from 5 Hz to 30 Hz with a partial overlap-frequency selection strategy, in which a group of 3 frequencies are used for inversion simultaneously. The frequency group increases from low to high with 2 frequencies overlapped and for each frequency band, a number of 5 outer iterations are performed. The stopping criteria for the inner iteration are \( k_{\text{max}} = 10 \) and/or \( \gamma_{\text{min}} = 2.0e-1 \). The stabilization parameters are \( \epsilon = 1.0e-2 \) and \( \lambda = 1.0e-2 \).

Figure 1: (a) True P-wave velocity model; (b) Initial P-wave velocity model.

Figures 2a and 2b illustrate the inversion results obtained by SD and \( l \)-BFGS (\( \mathcal{H}_0 = I \)) methods. Figures 2c and 2d show the comparison of well log data at 0.1 km and 0.6 km. The SD method is limited in recovering the deep parts of the model. The \( l \)-BFGS method (\( \mathcal{H}_0 = I \)) provides better inversion result compared to SD method but the deep parts of the inversion result are still not satisfactory. Figures 3a, 3b and 3c show the inversion results by \( l \)-BFGS methods with different diagonal Hessian approximations as initial guess. Figures 3d, 3e and 3f show the well log data comparison. With diagonal Hessian approximations as initial guess, the inversion results by \( l \)-BFGS methods have been improved.

Figure 2: (a) SD method (\( \phi = 0.31 \)); (b) \( l \)-BFGS method (\( \phi = 0.07 \)); (c) and (d) are the comparison of the well log data at 0.1 km and 0.6 km.

Figure 3: (a), (b) and (c) show the inverted models by \( l \)-BFGS methods with diagonal pseudo-Hessian (\( \phi = 4.28e-2 \)), diagonal Gauss-Newton Hessian (\( \phi = 2.0e-2 \)) and pseudo diagonal Gauss-Newton Hessian (\( \phi = 2.50e-2 \)) as initial guess respectively. (d), (e) and (f) show the well log data comparison.

Figure 4a is the reconstructed model by the non-preconditioned CG-GN method. Figures 4b, 4c and 4d show the inversion results by \( l \)-BFGS-\( GN \) (\( M = 5 \) and \( \mathcal{H}_0 = I \)), DPH-GN and \( l \)-BFGS-\( GN \)-\( DPH \) methods respectively. The deep parts of the reconstructed models by HF Gauss-Newton methods have been enhanced obviously compared to SD and \( l \)-BFGS methods. Furthermore, the inversion results by preconditioned HF Gauss-Newton methods are further improved compared to that by non-preconditioned CG-GN method.

Figure 4: (a) CG-GN method (\( \phi = 4.1e-3 \)); (b) \( l \)-BFGS-\( GN \) method (\( \phi = 1.1e-3 \)); (c) DPH-GN method (\( \phi = 4.4e-3 \)); (d) \( l \)-BFGS-\( GN \)-\( DPH \) method (\( \phi = 1.1e-3 \)); (e) and (f) show the comparison of well log data at 0.1 km and 0.6 km.

Figures 5a and 5b show the inversion results by DGH-GN and \( l \)-BFGS-
GN-DGH methods with the stabilization parameter $\lambda = 1.0e-2$. The DGH-GN and $l$-BFGS-GN-DGH inverted models are contaminated by artifacts. This is because incorporating the diagonal Gauss-Newton Hessian for preconditioning increases the instability of the inversion process. Figures 5c and 5d are DGH-GN and $l$-BFGS-GN-DGH inversion results with the stabilization parameter $\lambda = 5.0e-2$. It can be seen that the DGH-GN and $l$-BFGS-GN-DGH methods with diagonal Gauss-Newton Hessian approximation become more stable and the model can be reconstructed very well. Figures 6a and 6b show the inversion results by PDHG-GN and $l$-BFGS-GN-PDGH methods with the stabilization parameter $\lambda = 5.0e-2$. The PDHG-GN and $l$-BFGS-GN-PDGH inversion results based on the proposed pseudo diagonal Gauss-Newton Hessian can reconstruct the velocity model stably and efficiently. The $l$-BFGS-GN-DGH and $l$-BFGS-GN-PDGH methods with $\lambda = 5.0e-2$ give the best inversion results.

Figure 5: (a) DGH-GN method ($\lambda = 1.0e-2$ and $\phi = 8.5e-3$); (b) $l$-BFGS-GN-DGH method ($\lambda = 1.0e-2$ and $\phi = 5.2e-3$); (c) DGH-GN method ($\lambda = 5.0e-2$ and $\phi = 2.5e-3$); (d) $l$-BFGS-GN-DGH method ($\lambda = 5.0e-2$ and $\phi = 8.4e-4$); (e) and (f) are the comparison of the well log data at 0.1 km and 0.6 km.

Figure 6: (a) PDHG-GN ($\lambda = 5.0e-2$ and $\phi = 2.1e-3$); (b) $l$-BFGS-GN-PDGH ($\lambda = 5.0e-2$ and $\phi = 8.3e-4$); (c) and (d) are the comparison of the well log data at 0.1 km and 0.6 km.

Figure 7 shows the convergence history of the $l$-BFGS methods and HF Gauss-Newton methods. The HF Gauss-Newton methods provide faster convergence rates than $l$-BFGS methods. Except for DGH-GN and $l$-BFGS-GN-DGH methods ($\lambda = 1.0e-2$), the preconditioned HF Gauss-Newton methods converge faster than the non-preconditioned CG-GN method. Generally, the methods with $l$-BFGS preconditioners converge faster than those methods preconditioned by diagonal Hessian approximations. The $l$-BFGS-GN-DGH and $l$-BFGS-GN-PDGH methods with $\lambda = 5.0e-2$ give the fastest convergence rates.

In Figure 8a we plot the normalized misfit vs. number of forward problems solved for the HF Gauss-Newton methods. Figure 8b shows the normalized misfit vs. computation time (s). The preconditioned HF Gauss-Newton methods are more efficient than the non-preconditioned CG-GN method. Proper stabilization parameter $\lambda$ should be determined to ensure the effectiveness and stability of the DGH-GN and $l$-BFGS-GN-DGH methods. Except for the $l$-BFGS-GN-DGH method, the $l$-BFGS preconditioning schemes with diagonal Hessian approximations as initial guess can reduce the computation burden more effectively and reconstruct the velocity model better than $l$-BFGS-GN method with $\lambda = 1$. Although, $l$-BFGS-GN-DGH method ($\lambda = 5.0e-2$) provides fast convergence rate as shown in Figure 7, more computation cost is required for calculating the receiver-side Green’s functions. The $l$-BFGS-GN-PDGH method ($\lambda = 5.0e-2$) shows the best performance in reducing the computation burden and accelerating the HF Gauss-Newton FWI.

Figure 7: Comparison of the convergence history for the HF Gauss-Newton methods with different preconditioning schemes.

Figure 8: (a) Normalized misfit (log) vs. Number of forward problems solved for different preconditioning strategies; (b) Normalized misfit (log) vs. Computation time (s).

CONCLUSIONS

In this paper, we implement a Hessian-free Gauss-Newton FWI, which obtains search direction by solving the Newton equation with conjugate-gradient algorithm. To accelerate the HF GN FWI, we develop different preconditioning schemes. A pseudo diagonal Gauss-Newton Hessian is introduced as preconditioner. Furthermore, we propose to improve the $l$-BFGS preconditioning by employing the diagonal Hessian approximations as initial guess. We present numerical examples to show that the $l$-BFGS preconditioning method with pseudo diagonal Gauss-Newton Hessian as initial guess can speed up the HF GN FWI most efficiently.

ACKNOWLEDGMENTS

This research was supported by the sponsors of CREWES project and National Science and Engineering Research Council of Canada (NSERC, CRDPJ 461179-13).
Preconditioned HF Gauss-Newton FWI

REFERENCES


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


