GeoConvention
Calgary • Canada • March 7-11 2016
OPTIMIZING RESOURCES

Accelerating the Hessian-free Gauss-Newton Full-waveform Inversion via Preconditioned Conjugate Gradient Method

Wenyong Pan[1], Kris Innanen[1] and Wenyuan Liao[2]

1. CREWES Project, Department of Geoscience, University of Calgary

2. Department of Mathematics and Statistics, University of Calgary

Summary

Full-waveform inversion (FWI) has emerged as a powerful strategy for estimating the subsurface model parameter. The gradient-based methods promise to converge globally but suffer from slow convergence rate. The Newton-type methods provide a quadratic convergence, but the computation, storage and inversion of the Hessian are beyond the current computation ability for large-scale inverse problem. The Hessian-free (HF) optimization method represents an attractive alternative to these above-mentioned optimization methods. At each iteration, it obtains the search direction by approximately solving the Newton linear system using a conjugate-gradient (CG) algorithm. The main goal of this paper is to accelerate the HF FWI by preconditioning the CG algorithm. In this research, different preconditioning schemes for the HF Gauss-Newton optimization method are developed. It is concluded that the quasi-Newton L-BFGS preconditioning scheme with the pseudo diagonal Gauss-Newton Hessian as initial guess can speed up the HF Gauss-Newton FWI most efficiently.

Introduction

Full-waveform inversion (FWI) promises to provide high-resolution estimates of the subsurface parameter. FWI iteratively reconstructs the model parameters by minimizing a L-2 norm misfit function, which measures the difference between the observed data and synthetic data (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009). The traditional optimization methods for FWI in exploration geophysics are gradient-based methods. The gradient-based methods are computationally attractive for large-scale inverse problem but also suffer from slow convergence rate.

The search direction can be significantly enhanced by multiplying the gradient with the inverse Hessian matrix, which serves as a deconvolution operator for compensating the geometrical spreading effects and de-blurring the gradient. However, explicit calculation, storage and inversion of the Hessian at each iteration are beyond the current computation capability for large-scale inverse problem. In Gauss-Newton method, an approximate Hessian is introduced by only considering the first-order term (Pratt et al., 1998; Pan et al., 2015b). Pan et al. (2015a) used phase-encoding technology to construct the diagonal Gauss-Newton Hessian efficiently. Preconditioning the gradient with the diagonal pseudo-Hessian resembles a deconvolution imaging condition (Pan et al., 2014b, 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 previous iteration.

The Hessian-free optimization method represents an attractive alternative to the above-described optimization methods (Nash, 1985). At each iteration, the search direction is computed by approximately solving the Newton equations through a matrix-free fashion of the conjugate-gradient (CG) algorithm, which is an optimal method for solving a positive definite system. In this paper, the full Hessian is replaced with the Gauss-Newton Hessian, which is always symmetric positive definite. Our main goal in this paper is to precondition the CG algorithm for reducing the CG iterations and accelerating the HF Gauss-Newton full-waveform inversion. Different preconditioning schemes are developed for comparison in this paper.

The Hessian-free Optimization Methods

Instead of constructing Hessian approximations or approximating inverse Hessian, the Hessian-free (HF) optimization method obtains the search direction by solving the Newton linear system approximately using a conjugate-gradient (CG) method. The CG method is an optimal algorithm for solving a symmetric positive definite system and it only requires computing the Hessian-vector products. In the context of HF optimization method, the Hessian is always replaced with Gauss-Newton Hessian, which is always symetric and positive definite:

$$(\tilde{\mathbf{H}}_k + \varepsilon \mathbf{A}) \Delta \mathbf{m}_k = -\mathbf{g}_k, \tag{1}$$

where $\Delta \mathbf{m}_k$ is the search direction in k th iteration, $\tilde{\mathbf{H}}_k$ is the Gauss-Newton Hessian, $\mathbf{g}_k$ is the gradient, $\varepsilon \mathbf{A}$ is the stabilization term, ε is a small constant value and $\mathbf{A=max(\tilde{H}}_k\mathbf{)}$.

Preconditioning

One problem of the CG iterative algorithm is that it requires many iterations when obtaining the approximate solution of a linear system. The convergence rate of the CG method depends on the spectral properties (e.g., its eigenvalues) of the coefficient matrix. It is often convenient to transform the equation system into a system that has the same solution but has more favorable spectral properties. This can be achieved by applying a suitable preconditioner $\mathbf{M}^{-1}$ on the linear system. Thus, the preconditioned Newton system for the HF Gauss-Newton FWI is given by:

$$\mathbf{M}_k^{-1} (\tilde{\mathbf{H}}_k + \varepsilon \mathbf{A}) \Delta \mathbf{m}_k = -\mathbf{M}_k^{-1} \mathbf{g}_k. \tag{2}$$

The solution of equation (2) can be obtained by the preconditioned conjugate-gradient (PCG) method. In this paper, we developed efficient preconditioning schemes for the inner iterative algorithm.

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 for the CG inner iteration. The pseudo-Hessian is constructed by replacing the partial derivative wavefield with the virtual source in the correlation process. We also propose a pseudo diagonal Gauss-Newton Hessian approximation as the preconditioner for the CG algorithm in the inner loop, which can be expressed as:

$$\tilde{\tilde{\mathbf{H}}}_{diag}(\mathbf{x}) = \sum_{\mathbf{x}_g} \sum_{\mathbf{x}_s} \sum_{\omega} \Re(\omega^4 \mid f_s(\omega) \mid^2 \mid G(\mathbf{x}, \mathbf{x}_s, \omega) \mid^4). \tag{3}$$

Furthermore, we develop an L-BFGS preconditioning scheme for the HF optimization method, namely L-BFGS-GN method for $\mathscr{H}_0 = \mathbf{I}$. The approximated inverse Hessian $\mathscr{H}$ can also be used as a preconditioner for the CG iterative method:

$$\mathscr{H}_k^{-1} (\tilde{\mathbf{H}}_k + \varepsilon \mathbf{A}) \Delta \mathbf{m}_k = -\mathscr{H}_k^{-1} \mathbf{g}_k.$$

An identity matrix is usually set as the initial guess for the L-BFGS preconditioner, which is actually important for the L-BFGS method. To improve the performance of 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 preconditioner, which are named as L-BFGS-GN-DPH, L-BFGS-GN-DGH and L-BFGS-GN-PDGH methods respectively.

Numerical Examples

A more complex modified Marmousi model is used to examine the efficiency of different preconditioning schemes for the HF Gauss-Newton full-waveform inversion. The truncated Marmousi model has 100 X 100 grid cells with a grid interval of 10 m in both horizontal and vertical directions. Figures 1a and 1b show the true P-wave velocity model and initial P-wave velocity model. The initial velocity model is obtained by smoothing the true model with a Gaussian function. The inversion process is carried out with a multi-scale approach for mitigating the cycle-skipping problem. The frequency is 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 is $k_{\max} = 10$ or $\gamma_{\min} = 2e-1$. The stabilization parameters are $\sigma = 1.0e-2$ and $\lambda = 1.0e-2$.

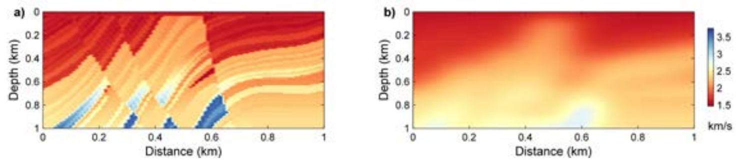

Figure 1. (a) True P-wave velocity model; (b) Initial P-wave velocity model.

Figures 2a and 2b show the inversion results by steepest-descent (SD) and L-BFGS methods. Figures 2c and 2d show the well log data comparison at 0.1 km and 0.6 km respectively. As we can see, the deep parts of the velocity model can not be recovered very well. Figures 3a, 3b, 3c and 3d show the inverted models by CG-GN, L- BFGS-GN, DPH-GN and L-BFGS-GN-DPH methods respectively. We notice that compared to SD and L-BFGS methods, the HF GN methods can recover the deep parts of the velocity model much better. Furthermore, we also see that the inversion results by preconditioned HF GN methods are enhanced compared to non-preconditioned HF GN method (GN-CG).

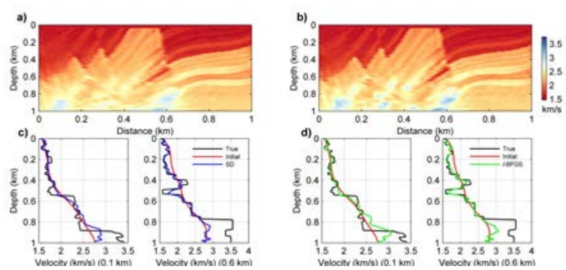

Figure 2. (a) SD method; (b) L-BFGS method. (c) and (d) show the well log data comparison at 0.1 km and 0.6 km respectively.

Figures 4a and 4b show the inversion results by DGH-GN and L-BFGS-GN-DGH methods with $\lambda = 1.0e-2$. The inversion results are contaminated by artifacts. Figures 4c and 4d are the inversion results by DGH-GN and L-BFGS-GN-DGH methods with $\lambda = 5.0e-2$. As we can see, the inversion results become much better. Figure 5a and 5b are the inversion results by PDGH-GN and L-BFGS-GN-PDGH methods with $\lambda = 5.0e-2$. Figure 5c and 5d show the well log data comparison at 0.1 km and 0.6 km respectively. It can be seen that the velocity model has been reconstructed well.

Figure 6 shows the convergence rate of differen methods. Figures 7a and 7b show the normalized misfit vs. number of forward problems solved and normalized misfit vs. computation time. The preconditioned HF Gauss-Newton methods can reconstruct the model more efficiently than the non-preconditioned one (CG-GN method). The L-BFGS preconditioning schemes can reduce the computation burden better than the preconditioning methods based on diagonal Hessian approximations. Proper stabilization parameter should be determined to ensure the effectiveness and stability of the DGH-GN and l-BFGS-GN-DGH methods. Although, L-BFGS-GN-DGH method ($\lambda = 5.0e-2$) provides fast convergence rate as shown in Figure 6. More computation cost is required for calculating the receiver-side Green's functions. The L-BFGS-GN-PDGH method ($\lambda = 5.0e-2$) based on the proposed pseudo diagonal Gauss-Newton does not need additional computational cost compared CG-GN method and it shows the best performance in reducing the computation burden and accelerating the HF Gauss-Newton FWI.

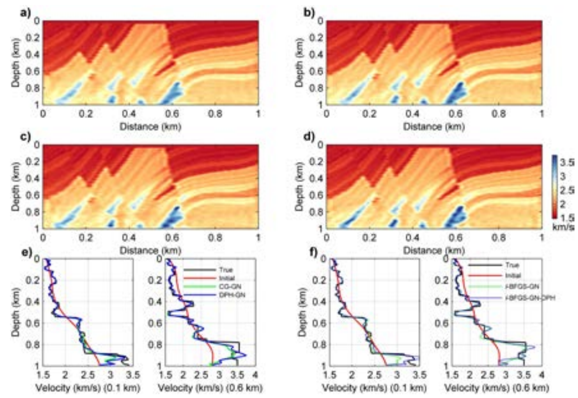

Figure 3. (a) CG-GN method; (b) L-BFGS-GN method; (c) DPH-GN method; (d) L-BFGS-GN-DPH method. (e) and (f) show the well log data comparison at 0.1 km and 0.6 km respectively.

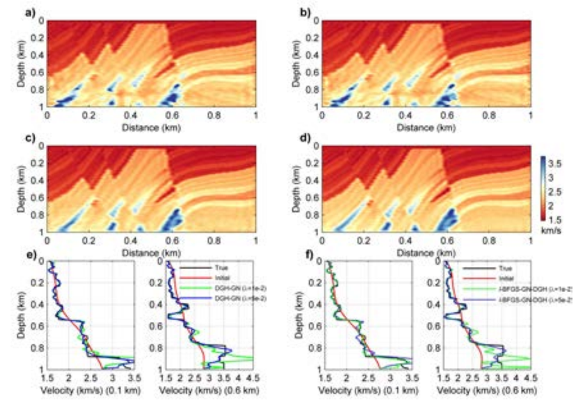

Figure 4. (a) and (b) are the inversion results by DGH-GN and L-BFGS-GN-DGH methods with $\lambda = 1.0e-2$. (c) and (d) are the inversion results by DGH-GN and L-BFGS-GN-DGH methods with $\lambda = 5.0e-2$. (e) and (f) show the well log data comparison at 0.1 km and 0.6 km respectively.

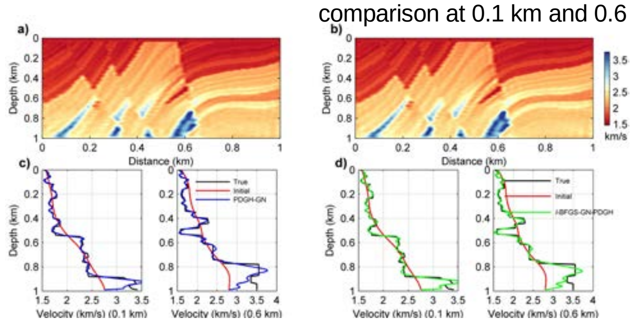

Figure 5. (a) and (b) are the inversion results by PDGH-GN and L-BFGS-GN-PDGH methods with $\lambda = 5.0e-2$. (c) and (d) show the well log data comparison at 0.1 km and 0.6 km respectively.

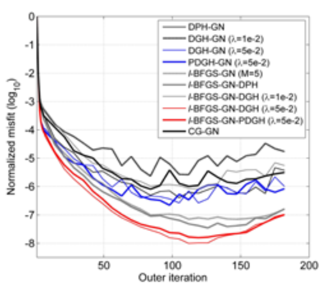

Figure 6. The convergence history of different methods.

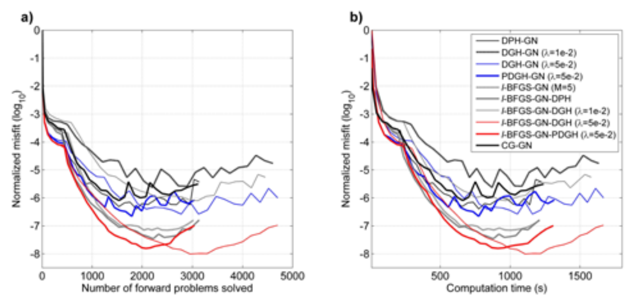

Figure 7. (a) Normalized misfit vs. Number of forward problems solved; (b) Normalized misfit vs. Computation time (s).

Conclusions

In this paper, a Hessian-free Gauss-Newton method is carried out for full-waveform inversion. To accelerate the Hessian-free Gauss-Newton method, we develop different preconditioning schemes for the inner CG algorithm at each outer iteration. A pseudo diagonal Gauss-Newton Hessian is also proposed as preconditioner based on the reciprocal property of the Green's function. We present numerical examples to show that the preconditioning schemes can improve the convergence rate of Hessian-free Gauss-Newton FWI and reduce the computation cost.

Acknowledgements

This research was supported by the Consortium for Research in Elastic Wave Exploration Seismology (CREWES) and National Science and Engineering Research Council of Canada (NSERC, CRDPJ 379744-08).

References

Lailly, P., 1983, The seismic inverse problem as a sequence of before stack migration: Conference on Inverse Scattering, Theory and Applications, SIAM, Expanded Abstracts, 206-220.

Nash, S. G., 1985, Preconditioning of truncated-newton methods: SIAM J. Sci. Statist. Comput, 6, 599-616.

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, 133, 341-362.

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

Pan, W., K. A. Innanen, G. F. Margrave, and D. Cao, 2015a, Efficient pseudo-Gauss-Newton full-waveform inversion in the t-p domain: Geophysics, 80, no. 5, R225-R14.

Pan, W., K. A. Innanen, G. F. Margrave, M. C. Fehler, X. Fang, and J. Li, 2015b, Estimation of elastic constants in HTI media using Gauss-Newton and Full-Newton multi-parameter full waveform inversion: SEG Technical Program Expanded Abstracts, 1177-1182.

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

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

Virieux, A. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, 74, no. 6, WCC1-WCC26.