Journal of Applied Geophysics 123 (2015) 1–17

Contents lists available at ScienceDirect

Journal of Applied Geophysics

journal homepage: www.elsevier.com/locate/jappgeo

Weighted pseudo-Hessian for frequency-domain elastic full waveform inversion

Hyunggu Jun, Eunjin Park *, Changsoo Shin

Department of Energy Systems Engineering, Seoul National University, Daehak-dong, Gwank-gu 151-744, Republic of Korea

ARTICLE INFO

Article history:
Received 16 March 2015
Received in revised form 4 September 2015
Accepted 18 September 2015
Available online 21 September 2015

Keywords:
Full waveform inversion
Elastic media
Pseudo-Hessian
Frequency domain

ABSTRACT

In full waveform inversion, an appropriate preconditioner for scaling the gradient improves the inversion results and accelerates the convergence speed. The Newton-based method uses the Hessian to scale the gradient and ensures a correctly inverted image with good convergence. However, calculating the full Hessian or approximated Hessian and obtaining their inverse matrices require extensive computation, which has remained an obstacle to the application of Newton-based methods to full waveform inversion. Many attempts have been made to reduce the computational cost of obtaining the Hessian and its inverse; among these alternatives, the pseudo-Hessian method has been widely used. The use of the pseudo-Hessian reduces the computational cost by regarding the zero-lag correlation of the impulse response as a unit matrix, but the pseudo-Hessian cannot properly scale the deep portion of a model. Therefore, we proposed a weighted pseudo-Hessian which can overcome the limitations of the conventional pseudo-Hessian. The weighted pseudo-Hessian was generated by combining a weighting matrix with the conventional pseudo-Hessian. The weighting matrix is an amplitude field and helps the pseudo-Hessian to scale the gradient properly. Therefore, the weighted pseudo-Hessian can effectively scale the gradient from the shallow part to the deep part of the model with great balance. Calculating the weighting function required minimal computation, such that the computational cost of generating the weighted pseudo-Hessian was nearly the same as the computational cost needed to calculate the conventional pseudo-Hessian. To verify the proposed algorithm, Marmousi-2 data were used for the synthetic test. The results indicate that the weighted pseudo-Hessian can effectively scale the gradient from the shallow to deep portions of a model.

1. Introduction

Accurately measuring the properties of subsurface materials is important for imaging subsurface geologic structures. Several methods may be used to invert the properties of subsurface materials from seismic data, such as travel-time tomography, migration velocity analysis and full waveform inversion (FWI). In the past thirty years, many geophysicists have worked to improve FWI (Lailly, 1983; Tarantola, 1984; Mora, 1987; Pratt et al., 1998; Shin and Min, 2006). To evaluate the properties of subsurface materials, FWI updates the model parameters by using an objective function that minimizes the residuals between the observed and modeled data. Therefore, the development of an effective method to minimize the residuals in the objective function has been an important area of research in FWI.

Many optimization methods, such as the steepest-descent method and Newton-based methods, have been used to minimize the misfits between observed and modeled data. In these methods, a certain matrix is used as a preconditioner to accelerate the convergence and improve the FWI results (Pratt et al., 1998; Brossier et al., 2009). Pratt et al. (1998) performed frequency-domain FWI by using the full Newton method and Gauss–Newton method. In the full Newton method, the gradient direction is scaled by using the exact Hessian, which consists of two parts: the first part is an approximated Hessian, and the second part is the cross-correlation between the 2nd-order partial derivatives of the source-side wavefield and the data residuals. The second term of the exact Hessian is difficult to calculate and has relatively small values. Therefore, the last term is generally dropped (Tarantola, 2004), and the resulting formula for scaling the gradient with the approximated Hessian is known as the Gauss–Newton method.

Although both the full Newton and Gauss–Newton optimization methods effectively reduce the misfits in the objective function while updating the model parameters, the Gauss–Newton method is more practical for use in FWI (Habashy and Abubakar, 2004; Sheen et al., 2006) because applying the approximated Hessian reduces the computational cost, although extensive computation is still required. Therefore, some efforts have been made to reduce the computational cost of obtaining the approximated Hessian and its inverse matrix.

Shin et al. (2001b) suggested an efficient method for calculating the Jacobian matrix that involved the use of a reciprocity theorem. The authors effectively calculated the Jacobian matrix by multiplying the

* Corresponding author.
E-mail addresses: h.jun1026@gmail.com (H. Jun), ej0417@snu.ac.kr (E. Park), css@model.snu.ac.kr (C. Shin).

http://dx.doi.org/10.1016/j.jappgeo.2015.09.014
0926-9851/

virtual source with Green's function. Golub and van Loan (1996), Hu et al. (2011) and Bae et al. (2012) applied the conjugate gradient least-square (CGLS) algorithm to the Gauss–Newton method. The CGLS algorithm does not require the explicit calculation of the approximated Hessian. These methods reduce the computational cost that is required to obtain the approximated Hessian, but extensive computation is still required, and adding a stabilizing constant (Marquardt, 1963) to the CGLS algorithm is difficult. The problem of high computational cost is exacerbated when we apply the Gauss–Newton algorithm to elastic FWI because of the considerably larger Hessian compared to the acoustic case. Therefore, the high computational cost remains an obstacle for the practical application of FWI.

Quasi-Newton methods such as L-BFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno) are also widely used (Brossier et al., 2009; Romdhane et al., 2011; Asnaashari et al., 2013) in FWI. L-BFGS estimates the model update, which is scaled by the approximated Hessian at every iteration, by using the model updates and model vectors of the n most-recent iterations. This method can compute the model update properly with less memory cost. However, this method does not always guarantee convergence, and using a proper initial estimation of the approximated Hessian is necessary.

Shin et al. (2001a) suggested a pseudo-Hessian by assuming that the matrix multiplication of the impulse response in the approximated Hessian was a unit matrix, and this approach enhanced the amplitude of later time signals in reverse time migration. Because the pseudo-Hessian is diagonally dominant, the authors scaled the migration image by using the diagonal elements of the pseudo-Hessian. Using this method, the computational cost of calculating the Hessian was greatly reduced. Tarantola (1984) reported that migration is the first step of inversion, and Chavent and Plessix (1999) and Shin et al. (2003) verified that prestack depth-migration is based on the same algorithm as that used in FWI. Therefore, many FWI studies have scaled the gradient direction by using diagonal elements of the pseudo-Hessian (Shin and Min, 2006; Kim et al., 2011). Despite the benefit of the pseudo-Hessian, it is limited in its ability to successfully image the deep region of a model because the pseudo-Hessian neglects the zero-lag correlation of the impulse response of the approximated Hessian, which describes geometric spreading (Choi et al., 2008). This limitation becomes severer when the pseudo-Hessian is applied to elastic FWI because of the complex wave phenomena.

Therefore, Choi et al. (2008) suggested a new pseudo-Hessian that directly addresses geometric spreading. In this method, geometric spreading is effectively simulated by introducing amplitude fields to the diagonal elements of the conventional pseudo-Hessian. The computational cost of constructing a new pseudo-Hessian is similar to that of obtaining the conventional pseudo-Hessian because the amplitude fields are calculated by using a forward modeling procedure. Although the new pseudo-Hessian can invert deep regions better than the conventional pseudo-Hessian, the former tends to neglect the shallow part of the model and focuses too much on the deeper parts of the model.

In this study, we suggest a weighted pseudo-Hessian to overcome the limitation of the conventional and new pseudo-Hessian. To calculate the weighted pseudo-Hessian, we generated an amplitude field which can help the Hessian to scale the gradient properly and combined the amplitude field to the conventional pseudo-Hessian. We used several types of approximation methods to generate the amplitude fields and showed how the approximations were performed. To provide better insight into the advantages of the weighted pseudo-Hessian, we compared the extracted Hessians. Comparing the weighted pseudo-Hessian to the conventional and new pseudo-Hessian, the weighted pseudo-Hessian updates the entire model more properly with a similar computational cost.

We begin by briefly reviewing the theory of frequency-domain FWI, pseudo-Hessian and new pseudo-Hessian. A comparison between the approximated Hessian and pseudo-Hessian follows. Then, we introduce the construction of a weighted pseudo-Hessian. Finally, we explain the strategies for the multi-parameter FWI and demonstrate the effectiveness of the weighted pseudo-Hessian by inverting Marmousi-2 synthetic data.

2. Theory

2.1. Frequency-domain full waveform inversion

An objective function that uses the l_2 norm in the frequency domain is as follows:

$$E = \frac{1}{2}(\mathbf{u}-\mathbf{d})^{\mathrm{T}}(\mathbf{u}-\mathbf{d})^{*}, \tag{1}$$

where $\mathbf{u}$ is the modeled wavefield vector in the frequency domain, $\mathbf{d}$ is the Fourier-transformed observed wavefield vector, T denotes the transpose and * represents the complex conjugate.

The gradient $\boldsymbol{g}$ of the objective function is given by

$$\boldsymbol{g} = Re\left[\boldsymbol{J}^{\mathrm{T}}\mathbf{r}^{*}\right], \tag{2}$$

where $\boldsymbol{J}$ is the partial derivative wavefield matrix and $\mathbf{r} = [\mathbf{u}-\mathbf{d}]$ is the data misfit vector between the modeled and observed wavefield vectors. The adjoint state method (Plessix, 2006) allows the efficient calculation of the gradient instead of the explicit calculation of $\boldsymbol{J}$. The gradient with respect to the kth model parameter m_k is expressed as follows:

$$\boldsymbol{g}_{\boldsymbol{m}_k} = Re\left[\mathbf{v}_k^{\mathrm{T}}\mathbf{S}^{\dagger}\mathbf{r}^{*}\right], \tag{3}$$

where $\mathbf{v}_k = -\frac{\partial\mathbf{S}}{\partial m_k}\mathbf{u}$, $\mathbf{S}$ is the impedance matrix and † indicates the adjoint matrix. The model update $\Delta\mathbf{m}$ is obtained by using a second-order Taylor expansion of the objective function and is expressed as follows:

$$\mathbf{H}\Delta\mathbf{m} = -\boldsymbol{g}, \tag{4}$$

where $\mathbf{H}$ is the Hessian matrix. Therefore, the model parameters at the ith iteration are as follows:

$$\mathbf{m}^{i} = \mathbf{m}^{i-1} - \mathbf{H}^{-1}\boldsymbol{g}^{i-1}. \tag{5}$$

The inverse Hessian $\mathbf{H}^{-1}$ acts as a preconditioner that compensates for geometric spreading (Shin et al., 2001a; Plessix and Mulder, 2004) and addresses the crosstalk between multiple parameters (Operto et al., 2013). When the proper preconditioner is obtained, we can correctly update the model parameters. However, the computation of $\mathbf{H}$ and its inverse matrix requires a huge amount of computation, and many attempts have been made to obtain the proper preconditioner at a lower computational cost, including the use of CGLS (Golub and van Loan, 1996; Hu et al., 2011; Bae et al., 2012), quasi-Newton (Brossier et al., 2009; Romdhane et al., 2011; Asnaashari et al., 2013) and pseudo-Hessian methods (Shin and Min, 2006; Kim et al., 2011). In this study, we used and improved a pseudo-Hessian method. As we

Table 1
Symbols for the Hessians.

Symbol	Name	Description
$\mathbf{H}$	Full Hessian	$Re[\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}^{*}] + Re[(\frac{\partial}{\partial m_1}\boldsymbol{J}^{\mathrm{T}})\mathbf{r}^{*}\ (\frac{\partial}{\partial m_2}\boldsymbol{J}^{\mathrm{T}})\mathbf{r}^{*}\cdots(\frac{\partial}{\partial m_n}\boldsymbol{J}^{\mathrm{T}})\mathbf{r}^{*}]$
$\mathbf{H}_{\mathrm{a}}$	Approximated Hessian	$Re[\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J}^{*}]$
$\mathbf{H}_{\mathrm{p}}$	Pseudo-Hessian	$Re[\mathbf{V}^{\mathrm{T}}\mathbf{V}^{*}]$
$\mathbf{H}_{\mathrm{new-p}}$	New pseudo-Hessian	$Re[\mathbf{V}^{\mathrm{T}}\mathbf{A}\mathbf{V}^{*}]$
$\mathbf{H}_{\mathrm{w-p}}$	Weighted pseudo-Hessian	$Re[\mathbf{V}^{\mathrm{T}}\mathbf{W}\mathbf{V}^{*}]$