

# Gauss-Newton and L-BFGS Methods in Full Waveform Inversion (FWI)



Samsul Ariffin Abdul Karim , Mudassar Iqbal , Afza Shafie, and Muhammad Izzatullah 

**Abstract** Full waveform inversion (FWI) is a recent powerful method in the area of seismic imaging where it used for reconstructing high-resolution images of the subsurface structure from local measurements of the seismic wavefield. This method consists in minimizing the distance between the predicted and the recorded data. The predicted data are computed as the solution of a wave-propagation problem. In this study, we investigate two algorithms Gauss-Newton and L-BFGS for solving FWI problems. We compare these algorithms in terms of its robustness and speed of convergence. Also, we implement the Tikhonov regularization for assisting convergence. Numerical results show that Gauss-Newton method performs better than L-BFGS method in terms of convergence of  $l_2$ -norm of misfit function gradient since it provides better convergence as well as the quality of high resolution constructed images. Yet, L-BFGS outperforms Gauss-Newton in terms of computationally efficiency and feasibility for FWI.

---

S. A. Abdul Karim (✉)

Fundamental and Applied Sciences Department and Centre for Systems Engineering (CSE),  
Institute of Autonomous System, Universiti Teknologi PETRONAS, Seri Iskandar,  
Perak Darul Ridzuan, Malaysia  
e-mail: [samsul\\_ariffin@utp.edu.my](mailto:samsul_ariffin@utp.edu.my)

M. Iqbal · A. Shafie

Department of Fundamental and Applied Sciences, Universiti Teknologi PETRONAS,  
Seri Iskandar, Perak Darul Ridzuan 32610, Malaysia  
e-mail: [mudassar.iqbal@buitms.edu.pk](mailto:mudassar.iqbal@buitms.edu.pk)

A. Shafie

e-mail: [afza@utp.edu.my](mailto:afza@utp.edu.my)

M. Iqbal

Department of Mathematical Sciences, Balochistan University of Information Technology,  
Engineering and Management Sciences (BUIITEMS), Quetta, Pakistan

M. Izzatullah

Seismic Modeling and Inversion Group (SMI), King Abdullah University of Science and  
Technology, Thuwal, Saudi Arabia  
e-mail: [muhammad.izzatullah@kaust.edu.sa](mailto:muhammad.izzatullah@kaust.edu.sa)

**Keywords** Full waveform inversion · Gauss-Newton · L-BFGS · Seismic wavefield · Tikhonov regularization

## 1 Introduction

Full-waveform inversion (FWI) is a recent powerful method based on based on nonlinear optimization technique in the area of seismic imaging. FWI was proposed by [1–3] back in the early of 1980s for reconstructing high-resolution images of the subsurface structure from local measurements of the seismic wavefield by minimizing the distance between the predicted and the recorded data [4–6]. Since then there are many numerical studies and new implementation of algorithms have been done [7, 8].

In this study, we investigate two algorithms Gauss-Newton and L-BFGS for solving frequency domain FWI as proposed in [7]. We compare these algorithms in terms of its robustness and speed of convergence via realistic synthetic model with marine exploration seismic setting. Also, we implement the Tikhonov regularization for assisting convergence.

## 2 Problem Formulation

We will formulate the FWI problem in the frequency domain as proposed by **PRatt**. Consider the slowness-squared as model parameters  $\mathbf{m} \in \mathbb{R}^{n_{grid}}$  and the measurement vector  $\mathbf{d} \in \mathbb{C}^{n_{data}}$  are related through a known but nonlinear relationship denoted as

$$\mathbf{d} = F(\mathbf{m}) + \epsilon, \quad (1)$$

where  $\epsilon \sim \mathcal{N}(0, \mathbf{C}_D)$  is additive, normally distributed noise with zero mean and covariance  $\mathbf{C}_D \in \mathbb{C}^{n_{data} \times n_{data}}$ .

The nonlinear forward modeling map  $F(\mathbf{m})$  can be described as

$$F(\mathbf{m}) = \mathbf{P}\mathbf{A}(\mathbf{m})^{-1}\mathbf{q}, \quad (2)$$

where  $\mathbf{q} \in \mathbb{C}^{n_{grid}}$  is the discretized source term which considered known. The operator  $\mathbf{A}(\mathbf{m}) \in \mathbb{C}^{n_{grid} \times n_{grid}}$  represents the discretized Helmholtz operator  $(\nabla^2 + \omega^2 \mathbf{m})$  where  $\omega = 2\pi f$  is the angular frequency. The operator  $\mathbf{P} \in \mathbb{R}^{n_{data} \times n_{grid}}$  denotes the sampling operator which samples the data  $\mathbf{d}$  from the field vector variables  $\mathbf{u}$ , which is the solution of the Helmholtz equation  $\mathbf{u} = \mathbf{A}(\mathbf{m})^{-1}\mathbf{q}$ .

By choosing the matrix that  $\mathbf{L}$  as the first order finite difference operator which commonly referred to as roughening matrix, we can define the least-square misfit function with Tikhonov regularization as

$$V(\mathbf{m}) = \frac{1}{2} \left\| F(\mathbf{m}) - \mathbf{d} \right\|_2^2 + \frac{\alpha}{2} \left\| \mathbf{L}\mathbf{m} \right\|_2^2, \quad (3)$$

where  $\alpha$  is the regularization coefficient. The optimal model  $\mathbf{m}$  can be sought by minimizing the misfit function  $V(\mathbf{m})$  in 3. The resulting optimization problem is typically solved using a gradient-based method which generates iterates of the form

$$\mathbf{m}_{k+1} = \mathbf{m}_k - B_k \nabla V(\mathbf{m}_k), \quad (4)$$

where  $B_k$  includes appropriate scaling/smoothing of the gradient. In this study the matrix  $B_k$  could be represented either as the inverse of the Gauss-Newton approximation or the L-BFGS approximation of Hessian which will be explained in details in the following sections. For the gradient of the misfit function, it can be computed through adjoint-state method [9] and the explicit formula can be described as

$$\nabla V(\mathbf{m}) = \mathbf{J}^T (F(\mathbf{m}) - \mathbf{d}) + \alpha (\mathbf{L}^T \mathbf{L}) \mathbf{m}, \quad (5)$$

with  $\mathbf{J}$  the Jacobian of  $\mathbf{F}(\mathbf{m})$ .

### 3 Gauss-Newton Method

Gauss-Newton method is a method derived from Newton method for solving the nonlinear optimization problem. The issue with Newton method in solving the nonlinear optimization problem especially FWI is the computation of full Hessian. In Eq. 4, the matrix  $B_k$  has two terms based on Newton method which can be presented as

$$\mathbf{H} = \mathbf{J}^T \mathbf{J} + \frac{\partial \mathbf{J}}{\partial \mathbf{m}} (F(\mathbf{m}) - \mathbf{d}). \quad (6)$$

Commonly, the computation of the second term is avoided due to its tedious calculation and which in any case should be small by assuming the problem is approximately linear, which, in practice, implies that the starting model is sufficiently close to the true model. This is where the Gauss-Newton method is being derived from. The difference between Newton and Gauss-Newton method is the negligence of the second term in the Hessian computation. Based on [7, 10], we can safely dropped off the second term in the Eq. 6 because of its value is too small and it is only important if changes in the parameters cause a change in the partial derivative of the Helmholtz equation's solution.

The Gauss-Newton method and its approximation of Hessian can be presented as

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \mathbf{H}_{GN} \nabla V(\mathbf{m}_k), \quad (7)$$

$$\mathbf{H}_{GN} = \mathbf{J}^T \mathbf{J}, \quad (8)$$

where the matrix  $\mathbf{H}_{GN}$  is assumed to have full column rank, and is thus invertible. See [11] for more details regarding to this algorithm.

## 4 L-BFGS Method

The limited- memory BFGS method (L-BFGS) is a quite successful modification of the quasi-Newton methods [11, 12]. In this method, no Hessian approximation is ever actually formed, but rather a collection of the last several  $(s_k, y_k)$  pairs is stored and used to compute the step. Let  $m$ , the memory size, be the number of  $(s, y)$  pairs stored. Then, given an initial matrix  $H_0$ , the matrix  $H_k$  can be defined as follows:

```

 $H \leftarrow H_0$ 
for  $m, \dots, 1$  do
   $H \leftarrow V_{k-i}^T H V_{k-i} + \rho_{k-i} s_{k-i} s_{k-i}^T$ 
end for
 $H_k \leftarrow H$ .

```

The notation is simplified by eliminating the iteration counter  $k$  and choosing to store the most recent value of  $s$ , that is,  $s_k - 1$ , in  $s_m - 1$  and the oldest value,  $s_k - m$ , in  $s_0$ . The vectors  $y_i, i = 0, \dots, m - 1$ , are stored similarly. With these values, it can be shown that the search direction in Eq. 4 can be represented as

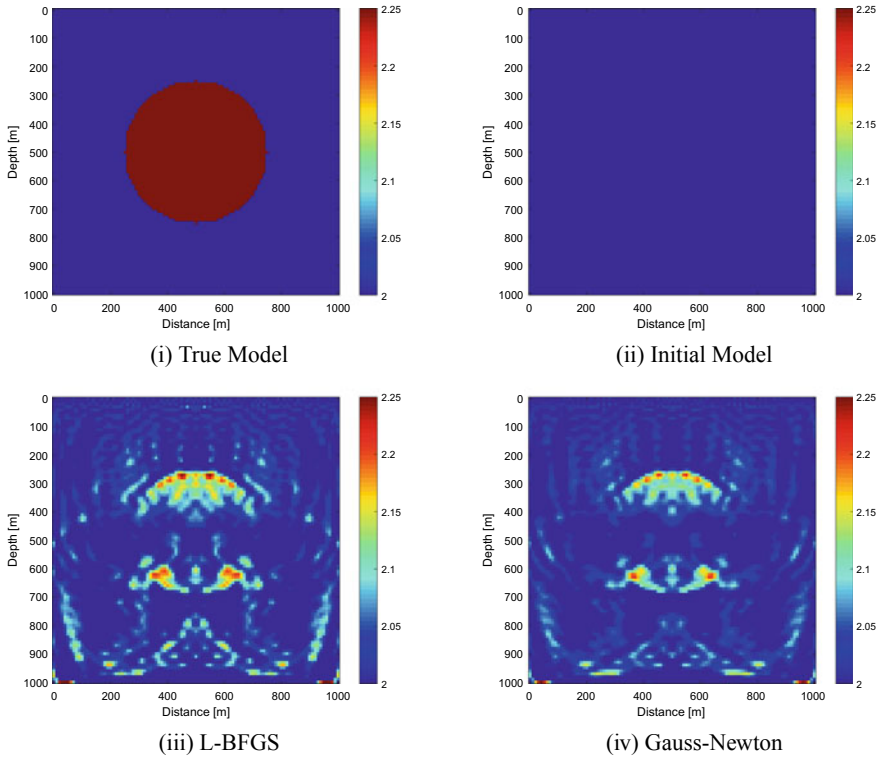
$$B_k \nabla V(\mathbf{m}_k) = H_k \nabla V(\mathbf{m}_k), \quad (9)$$

where the matrix  $H_k$  is the L-BFGS approximation to the inverse Hessian and can be computed through the algorithm presented above.

## 5 Numerical Examples

In these numerical examples, we illustrate the performance of Gauss-Newton and L-BFGS algorithms through solving the frequency domain FWI problem. We solve two FWI problems with two different velocity models with an objective to compare these two algorithms in reconstructing the velocity models from the recorded data.

For first numerical example, we use a homogeneous velocity model with an inclusion in the centre which acts as an reflector, depicted in the Fig. 1a. A standard



**Fig. 1** Numerical example 1: reconstruction of homogeneous velocity model with an inclusion in the centre

finite-difference method is used to solve the Helmholtz equation. The grid size is  $100 \times 100$ , and grid spacing is  $10 \times 10$  m. In this numerical example we consider collocated sources-receivers setting with sources-receivers are located at every 20m. We use frequency content 5 to 25 Hz with frequency sampling of 3.33 Hz.

In the second numerical example, we use the Marmousi model as depicted in the Fig. 3a to perform the numerical studies. A standard finite-difference method is used to solve the Helmholtz equation. The grid size is  $61 \times 220$ , and grid spacing is  $50 \times 50$  m. 50 shots at every 100 and 100 receivers at every 50m are used in this numerical example. This sources-receivers setting is resembling the marine exploration seismic setting. We use frequency content from 0.5 Hz to 3.95 Hz with frequency sampling of 0.5 Hz.

For both numerical examples, we performed 100 Gauss-Newton and L-BFGS iterations each, starting from the initial model depicted in the Figs. 1b and 3b respectively to obtain the optimal model  $\mathbf{m}$  as shown in the bottom row of Figs. 1 and 3. As regularization, we use the Tikhonov regularization method with regularization operator  $\mathbf{L}$  as first order derivative operator and regularization parameter  $\alpha$  equals to 0.01.

In practice, the Hessian is not store explicitly in memory and only its matrix-vectors product are being computed. Thus, for the Gauss-Newton iterations, we are solving a system of linear equations at each iteration using the preconditioned conjugate gradient (PCG) to estimate the descent direction.

## 6 Discussions

Based on two numerical results, both algorithms are performing well and both showing a good convergence of misfit values and the values of  $l_2$ -norm of misfit function gradient as illustrated in Figs. 2 and 4, respectively. As we can observe, the misfit values of L-BFGS is better than Gauss-Newton algorithm, yet the values of  $l_2$ -norm of misfit function gradient for Gauss-Newton algorithm is lower compared to L-BFGS algorithm. In practice, we should consider the values of  $l_2$ -norm of misfit function gradient as it represents the optimal distance of the solution to the truth. This is because the true solution could be obtained when the misfit function gradient is equal to zero or in the vicinity of  $l_2$ -norm of misfit function gradient closes to zero. Thus, based on this practice, Gauss-Newton algorithm is perform better compared to L-BFGS because of its lower value in  $l_2$ -norm of misfit function gradient.

Here we also should discuss the feasibility of each algorithm. Gauss-Newton algorithm needs the matrix-vector product between the inverse of its approximated Hessian and the gradient at each iteration in order to obtain the descent direction. This computation is computationally intensive thus it takes longer time per iteration to solve the optimization problem. Meanwhile, in L-BFGS algorithm no Hessian approximation is ever actually formed, but rather a collection of the last several  $(s_k, y_k)$  pairs is stored and used to compute the step. This makes L-BFGS algorithm is computationally efficient compared to the Gauss-Newton algorithm.

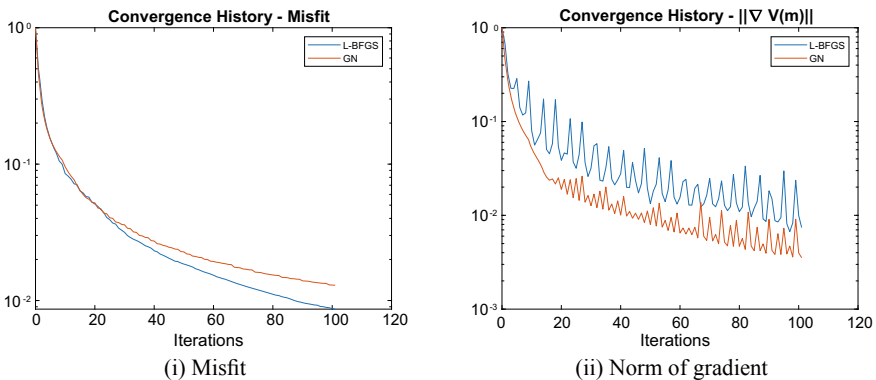
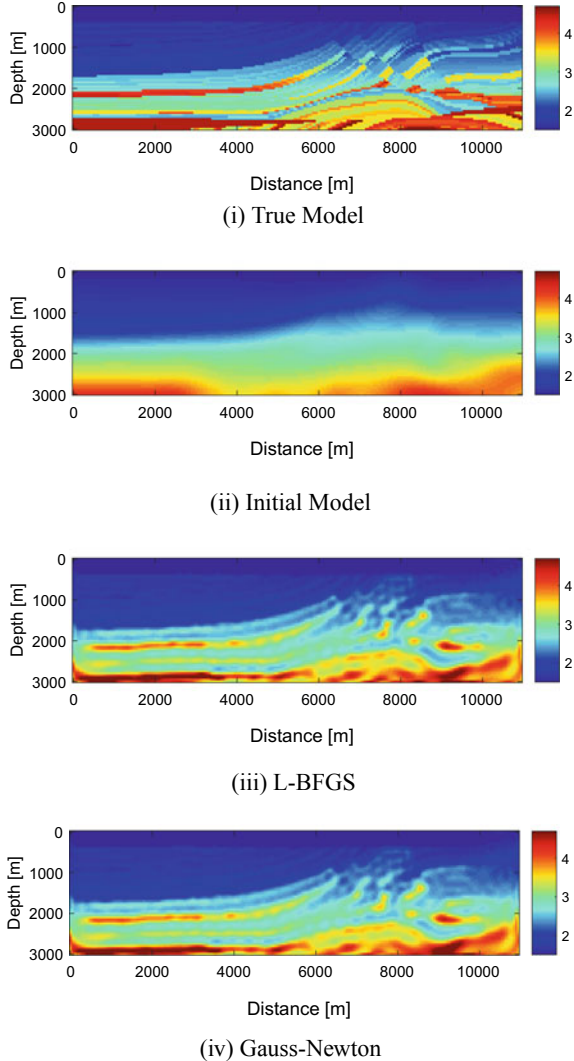


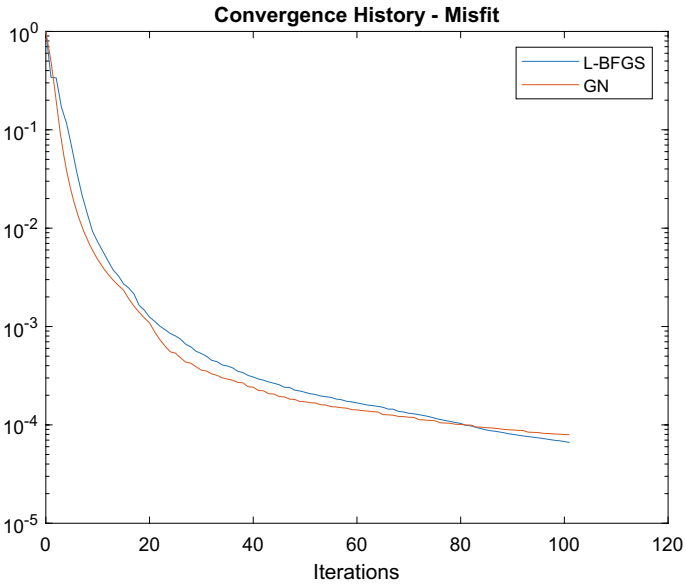
Fig. 2 Numerical example 1: misfit and norm of gradient values at each iteration

**Fig. 3** Numerical example 2: reconstruction of marmousi model

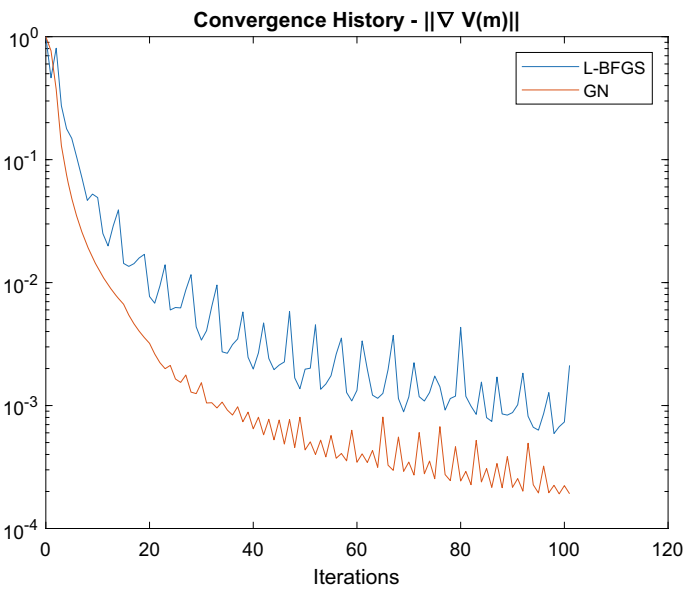


## 7 Conclusion

In conclusion, both algorithms, L-BFGS and Gauss-Newton are comparable to each other in terms of performance. Gauss-Newton algorithm gives a better result in the convergence of  $l_2$ -norm of misfit function gradient sense, yet it is computationally intensive. Meanwhile, L-BFGS performance is comparable to the Gauss-Newton and in terms of computationally efficiency and feasibility, L-BFGS is outperformed the Gauss-Newton for the large scale optimization problems especially in FWI.



(i) Misfit



(ii) Norm of gradient

**Fig. 4** Numerical example 2: misfit and norm of gradient values at each iteration



**Acknowledgements** This research was fully supported by Universiti Teknologi PETRONAS (UTP) through a research grant YUTP: 015LC0-315 (Uncertainty estimation based on Quasi-Newton methods for Full Waveform Inversion (FWI)).

## References

1. Tarantola, A.: Inversion of seismic reflection data in the acoustic approximation. *Geophysics* **49**(8), 1259–1266 (1984)
2. Tarantola, A.: Linearized inversion of seismic reflection data. *Geophys. Prospect.* **32**(6), 998–1015 (1984)
3. Lailly, P.: The seismic inverse problem as a sequence of before stack migrations. In: *Conference on Inverse Scattering, Theory and Applications*, Society for Industrial and Applied Mathematics (1983)
4. Valette, B., Tarantola, A.: Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys.* **20**(2), 219 (1982)
5. Virieux, J., Operto, S.: An overview of full-waveform inversion in exploration geophysics. *Geophysics* **74**(6), WCC1–WCC26 (2009)
6. Virieux, J., Asnaashari, A., Brossier, R., Métivier, L., Ribodetti, A., Zhou, W.: An introduction to full waveform inversion (2017)
7. Gerhard Pratt, R., Shin, C., Hick, G.J.: Gauss–Newton and full Newton methods in frequency–space seismic waveform inversion. *Geophys. J. Int.* **133**(2), 341–362 (1998)
8. Métivier, L., Brossier, R., Operto, S., Virieux, J.: Full waveform inversion and the truncated Newton method. *SIAM Rev.* **59**(1), 153–195 (2017)
9. Plessix, R.E.: A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophys. J. Int.* **167**(2), 495–503 (2006)
10. Tarantola, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics (2005)
11. Nocedal, J., Wright, S.J.: *Numerical Optimization*, 2nd edn. New York, Springer (2006)
12. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. *Math. Program.* **45**(1), 503–528 (1989)