Full-waveform inversion based on nonlinear conjugate gradient method, Gauss-Newton method and full Newton method

Yinbin Ma, Musa Maharramov, Biondo Biondi

ABSTRACT

Full-waveform inversion (FWI) generates a high-resolution subsurface model. Robust local minimization algorithms are required for FWI because the objective function is highly nonlinear. In this paper we compare the nonlinear conjugate gradient method, Gauss-Newton method and full Newton method for FWI. These methods use the gradient of objective function and application of Hessian on a model perturbation vector, which can be calculated efficiently with the adjoint-state methods. Numerical results suggest Newton-type methods resolve fine structure faster than the nonlinear conjugate gradient method in terms of number of wave propagation.

INTRODUCTION

Full-waveform inversion (Tarantola, 1984; Virieux and Operto, 2009) is a challenging technique that estimates the high-resolution subsurface model by minimizing the mismatch between observed data and synthetic data. The first order derivative is usually needed for FWI, and the Hessian is used in Newton-type methods. It is known that the adjoint-state method is an efficient method to compute the Frechet derivative (Tromp et al., 2005; Plessix, 2006) and the Hessian (Fichtner, 2011; Fichtner and Trampert, 2011) for FWI.

In this paper, we implement and compare FWI using three methods: nonlinear conjugate gradient (CG) method (Nocedal and Wright, 2006; Maharramov and Biondi, 2013), Gauss-Newton method and full Newton method (Pratt et al., 1998). Nonlinear CG method requires the gradient at each iteration, and the model update is computed based on the current gradient and previous gradient. For Gauss-Newton method and full Newton method, the model update is calculated by applying the approximated inverse of the Hessian to the gradient.

Simple synthetic models are used to test our implementation of FWI algorithms. A model with pinch-out structures is constructed to test the vertical resolution of different methods. A model with wells is used to test the horizontal resolution and the ability to recover vertical structures. Numerical results suggest that Newton-type methods converge faster than nonlinear CG method in terms of number of wave propagation.

We create synthetic model with fine structures based on the Society of Exploration Geophysicists/European Association of Geoscientists and Engineers (SEG/EAGE) model. Preliminary results suggest that Newton-type methods recover deeper structure better than the nonlinear CG method after the same number of wave propagation.

We briefly discuss the different rate of convergence in the last section.

METHOD

Full-waveform inversion for acoustic media

We use the least-squares misfit function for FWI in the time domain, as follows:

$$J(m) = \frac{1}{2}\sum_r \int_0^T ||S_r\mathbf{u} - d_r||_2^2 dt, \tag{1}$$

where S_r is the sampling operator for the receivers, d_r is the observed data at the receiver $\mathbf{r}$, and $\mathbf{u}$ is the synthetic pressure wavefield.

The pressure field $\mathbf{u}$ is computed using the acoustic approximation of wave equation with a non-constant density, as follows:

$$\begin{cases} [\frac{1}{K}\partial_t^2 - \tilde{\nabla}\cdot(\frac{1}{\rho}\nabla)]\mathbf{u} & = \mathbf{f} \\ \mathbf{u}(\mathbf{r}, t=0) & = 0 \\ \partial_t\mathbf{u}(\mathbf{r}, t=0) & = 0, \end{cases} \tag{2}$$

where K is the bulk modulus, ρ is the density, and $\mathbf{f}$ is the source wavefield. Numerically, we solve equation 2 in the time domain using staggered-grid finite difference method, starting from $t = 0$ to maximum recording time $t = \mathrm{T}$.

We formulate FWI as a nonlinear optimization problem and solve it iteratively. The Frechet derivate is estimated at each iteration based on the adjoint-state methods, which requires the correlation of wavefield $\mathbf{u}$ with the receiver wavefield λ, which is defined as follows:

$$\begin{cases} [\frac{1}{K}\partial_t^2 - \nabla\cdot(\frac{1}{\rho}\tilde{\nabla})]\lambda & = \mathbf{d_{res}} \\ \lambda(\mathbf{r}, t=T) & = 0 \\ \partial_t\lambda(\mathbf{r}, t=T) & = 0, \end{cases} \tag{3}$$

where $\mathbf{d_{res}}$ is the difference between synthetic data and observed data. The receiver wavefield is computed backward in time, starting from the maximum recording time $t = \mathbf{T}$ to $t = 0$.

The acoustic wave equation 2 is used to model non-constant density media. We implemented multi-parameter FWI in acoustic media. However, in this paper ρ is assumed to be constant, and we estimate bulk modulus K, which is related to velocity by, as follows:

$$v = \sqrt{\frac{K}{\rho}}. \tag{4}$$

Nonlinear conjugate gradient method

We implement the conjugate gradient method to minimize the objective function in equation 1. A nonlinear CG method generates a sequence of estimated modulus K_i, $i \geq 0$, starting from initial guess K_0.

Assume at iteration i, we have obtained the estimated model K_i. For the next iteration, the gradient is calculated at $\mathbf{g_i} = \mathbf{g}(K_i) \equiv \frac{\partial J}{\partial K_i}$. From the current gradient $\mathbf{g_i}$ and previous gradient $\mathbf{g_{i-1}}$, we get the search direction ΔK_i using the Fletcher Reeves formula and Polak Ribire formula(Nocedal and Wright, 2006; Maharramov and Biondi, 2013).

The local minimum of objective $J(K)$ in the vicinity of K_i is estimated with the line search approach, as follows:

$$\alpha_i = \text{argmin}_\alpha J(K_i + \alpha \Delta K_i), \tag{5}$$

and we update the model with,

$$K_{i+1} = K_i + \alpha_i \Delta K_i. \tag{6}$$

We do not solve the line search problem in equation 5 exactly because the evaluation of $J(K_i + \alpha \Delta K_i)$ is expensive. We want an approximated solution with a few iterations. The line search process is terminated once the Wolfe condition (Nocedal and Wright, 2006) is satisfied, as follows:

$$J(K_i + \alpha \Delta K_i) \leq J(K_i) + c_1 \alpha \mathbf{g_i}^T \Delta K_i \tag{7}$$

$$|\nabla J(K_i + \alpha \Delta K_i)^T \Delta K_i| \leq c_2 |\mathbf{g_i}^T \Delta K_i|, \tag{8}$$

where $0 < c_1 < c_2 < 1$, and ∇J is the gradient of objective function.

We continue this process, until the value of the objective function or the norm of gradient is below a certain threshold.

The gradient used in the nonlinear CG method is computed with the adjoint-state methods (Fichtner, 2011; Fichtner and Trampert, 2011), shown in the following:

$$\mathbf{g}(\mathbf{K}) = \frac{\partial J}{\partial K} = -\int_0^T \lambda \frac{1}{K^2} \partial_t^2 \mathbf{u} dt, \tag{9}$$

where the source wavefield $\mathbf{u}$ satisfies equation 2, and receiver wavefield λ is obtained from equation 2. The details of derivation are shown in the appendix, with multiparameter model:

$$\mathbf{m} = \begin{bmatrix} 1/K \\ 1/\rho \end{bmatrix}. \tag{10}$$

Full Newton method and Gauss-Newton method

Consider the second order expansion of objective function, as follows:

$$J(K + \Delta K) = J(K) + \left(\frac{\partial J}{\partial K}\right)^T \Delta K + \frac{1}{2}\Delta K \mathbf{H}(K) \Delta K + O(\Delta K^3), \tag{11}$$

where the full Hessian $\mathbf{H}$ is the second order derivative of the objective function,

$$\begin{aligned} \mathbf{H} &= \frac{\partial^2 J}{\partial K^2} \\ &= \left(\frac{\partial J}{\partial K}\right)^T \left(\frac{\partial J}{\partial K}\right) + \mathbf{d_{res}}^T \frac{\partial^2 (S_r \mathbf{u})}{\partial K^2}. \end{aligned}$$

The Hessian is reduced to Gauss-Newton Hessian $\mathbf{H}_{\mathrm{GN}}$ by dropping the second term, as follows:

$$\mathbf{H}_{\mathrm{GN}} = \left(\frac{\partial J}{\partial K}\right)^T \left(\frac{\partial J}{\partial K}\right) \tag{12}$$

For the full Newton method and Gauss-Newton method, suppose at iteration $i > 0$, we have obtained modulus K_i. For the next iteration, the gradient $\mathbf{g}(K_i)$ is calculated with equation 9. We scale the gradient by applying the inverse of Hessian, as follows:

$$\Delta K_i = -\mathbf{H}^{-1} \mathbf{g}(K_i). \tag{13}$$

Then we use the line search approach to find the estimate model for the next iteration,

$$\alpha_i = \mathrm{argmin}_\alpha J(K_i + \alpha \Delta K_i) \tag{14}$$

$$K_{i+1} = K_i + \alpha_i \Delta K_i. \tag{15}$$

One advantage of the full Newton method is that when K_i is close to the solution, we will have $\alpha_i \to 1$.

The inverse of Hessian $-\mathbf{H}^{-1}g(K_i)$ is approximated iteratively, which requires the action of Hessian on a model perturbation vector $\mathbf{H}\Delta K$. The action of Hessian can be computed efficiently using the adjoint-state method, as follows:

$$\mathbf{H}\Delta K = -\frac{\Delta K}{K^2}\int_0^T (\mu_a(x,t)+\mu_b(x,t))\partial_t^2\mathbf{u}(x,t)dt - \frac{\Delta K}{K^2}\int_0^T \lambda(x,t)\partial_t^2\delta\mathbf{u}_1(x,t)dt, \tag{16}$$

with source wavefield $\mathbf{u}$, λ from the forward wave propagation, and receiver wavefield μ_a, μ_b and λ from the backward wave propagation, as follows:

$$\begin{cases} [m_K\partial_t^2 - \tilde{\nabla}\cdot(m_\rho\nabla)]\mathbf{u} & = \mathbf{f} \\ [\frac{1}{K}\partial_t^2 - \tilde{\nabla}\cdot(\frac{1}{\rho}\nabla)]^*\lambda & = \sum_r S_r^*(S_r\mathbf{u}-d_r) \\ [m_K\partial_t^2 - \tilde{\nabla}\cdot(m_\rho\nabla)]\delta\mathbf{u_1} & = [\Delta m_K\partial_t^2 - \tilde{\nabla}\cdot(\Delta m_\rho\nabla)]\mathbf{u} \\ [m_K\partial_t^2 - \tilde{\nabla}\cdot(m_\rho\nabla)]^*\mu_\mathbf{a} & = S_r^*S_r\delta\mathbf{u_1} \\ [m_K\partial_t^2 - \tilde{\nabla}\cdot(m_\rho\nabla)]^*\mu_\mathbf{b} & = [\Delta m_K\partial_t^2 - \tilde{\nabla}\cdot(\Delta m_\rho\nabla)]^*\lambda, \end{cases} \tag{17}$$

where we use $m_K \equiv \frac{1}{K}$ and $m_\rho \equiv \frac{1}{\rho}$. The corresponding model perturbations are $\Delta m_K = -\frac{\Delta K}{K^2}$ and $\Delta m_\rho = -\frac{\Delta\rho}{\rho^2}$.

For the Gauss-Newton Hessian, the action can be expressed as the following:

$$\mathbf{H}_{\mathrm{GN}}\Delta K = -\frac{\Delta K}{K^2}\int_0^T \mu_\mathbf{a}(\mathbf{r},t)\partial_t^2\mathbf{u}(\mathbf{r},t)dt. \tag{18}$$

We put the detailed derivation of the Hessian in the appendix with multiparameter model $\mathbf{m} = (1/K, 1/\rho)$. The approximation of the inverse of the Hessian is expensive both in computation time and memory.

RESULTS

Simple synthetic models

We initially test the nonlinear CG method, Gauss-Newton method and full Newton method on simple synthetic models. We use 10 sources with $300\,m$ spacing between neighboring sources. The depth of the sources is $650\,m$, and the receivers are at the same depth as the sources.

Each model is solved with nonlinear CG method for 50 iterations (150 wave propagation for pinch-out model and 166 wave propagation for model contains vertical wells). Four iterations of the full Newton method (160 wave propagation) and Gauss-Newton method (120 wave propagation) are applied, and at each iteration we solve $\Delta K = H^{-1}(-g)$ for 10 CG steps.

The first model Figure 1 contains 5 thin layers with $60-\mathrm{meter}(m)$ thickness, and pinch-out structure. The model has velocity $2400\,m/s$ with source wavelet centered

at approximately 30 Hz. The layers are within the resolution of FWI, while the thin part of the pinch-out is below the resolution. In Figure 3, we plot the objective function versus the number of wave propagation for different methods. Newton-type methods converge faster than the nonlinear CG method in terms of computational cost. We can see the inversion results for the pinch-out model in Figure 2, after 120 wave propagation. Newton-type methods reconstruct sharper boundary comparing with the nonlinear CG method.

The second model shown in Figure 4 contains two vertical wells with $100\,m$ and $30\,m$ width. The objective function versus the number of wave propagation for different methods is shown in Figure 6. We can see the inversion results in Figure 5. For the Newton-type methods, the location of the vertical well is identified properly, while the parameter within the well cannot be properly estimated. On the other hand, the nonlinear CG method has not resolved the deeper part of the $100\,m$ well and has not clearly identified the $30\,m$ well.

SEG/SEAM model

The full Newton method and Gauss-Newton method use the quadratic approximation of the objective function, and ideally they should have better resolution comparing with the nonlinear CG method. They should converge faster when we are close to the true model. In this subsection, we create a synthetic model based on SEG/EAGE model as in Figure 7. We use 10 sources with $300\,m$ spacing between neighboring sources. The depth of the sources is $650\,m$, and the receivers are at the same depth as the sources.

We run the nonlinear CG method for 100 iterations. We apply Full Newton method and Gauss-Newton method for 4 iterations; and at each iteration, we solve $\Delta K = H^{-1}(-g)$ for 15 CG steps. The objective versus the number of wave propagation is shown in Figure 9. We can see that the Newton-type methods converge faster than the nonlinear CG method, at the same computational cost.

We plot the inversion results after 180 wave propagation (56 iterations for the nonlinear CG method, 4 iterations of Gauss-Newton method, and 3 iterations of full Newton method) in Figure 8. We can see significant difference below $1100m$. The nonlinear CG method has not resolved the structure, and the Newton-type methods have reconstructed the fine structure. The Gauss-Newton method has smaller residual comparing with the full Newton method, and we will discuss it in the next section.

DISCUSSION ON RATE OF CONVERGENCE

We show that Newton-type methods converge faster than nonlinear CG method, in terms of number of the wave propagation.

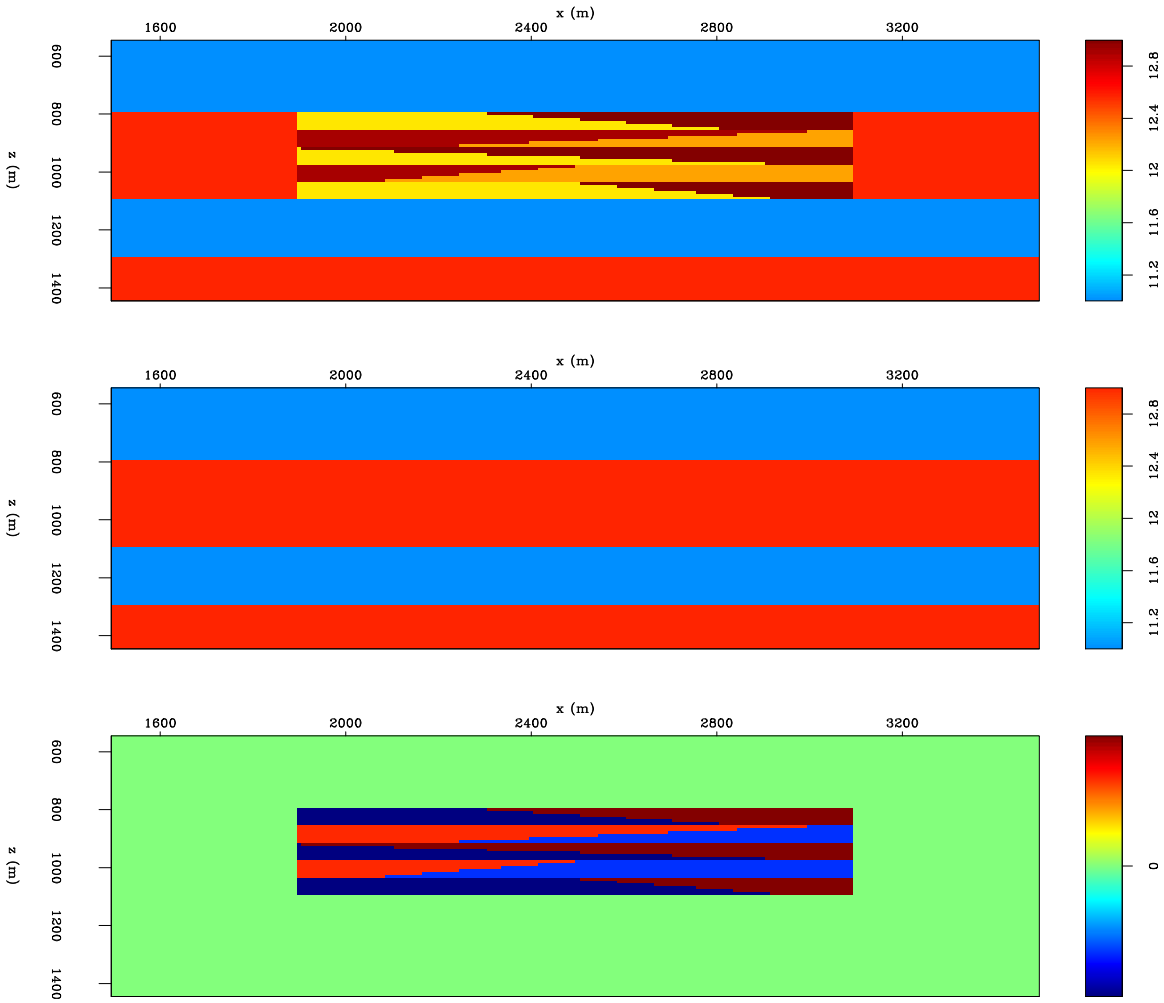

Figure 1: The top panel shows the true modulus model. The middle panel shows the starting modulus model. The bottom panel shows the difference between the true model and starting model.[**ER**]

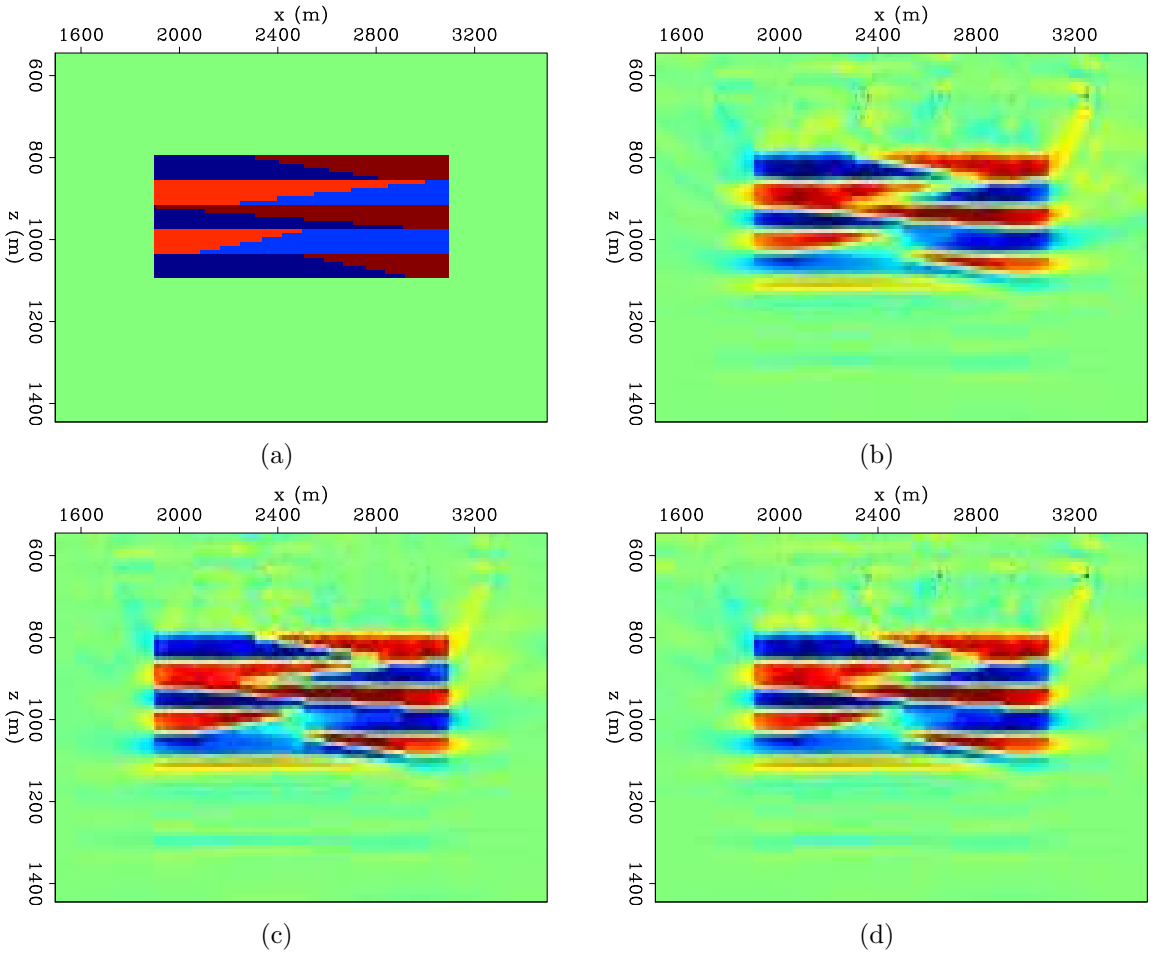

Figure 2: Panel (a) shows the difference between the true model and starting model. Panel (b) is the model update with the nonlinear CG method. Panel (c) is the model update with the Gauss-Newton method. Panel (d) is the model update with the full Newton method.[**CR**]

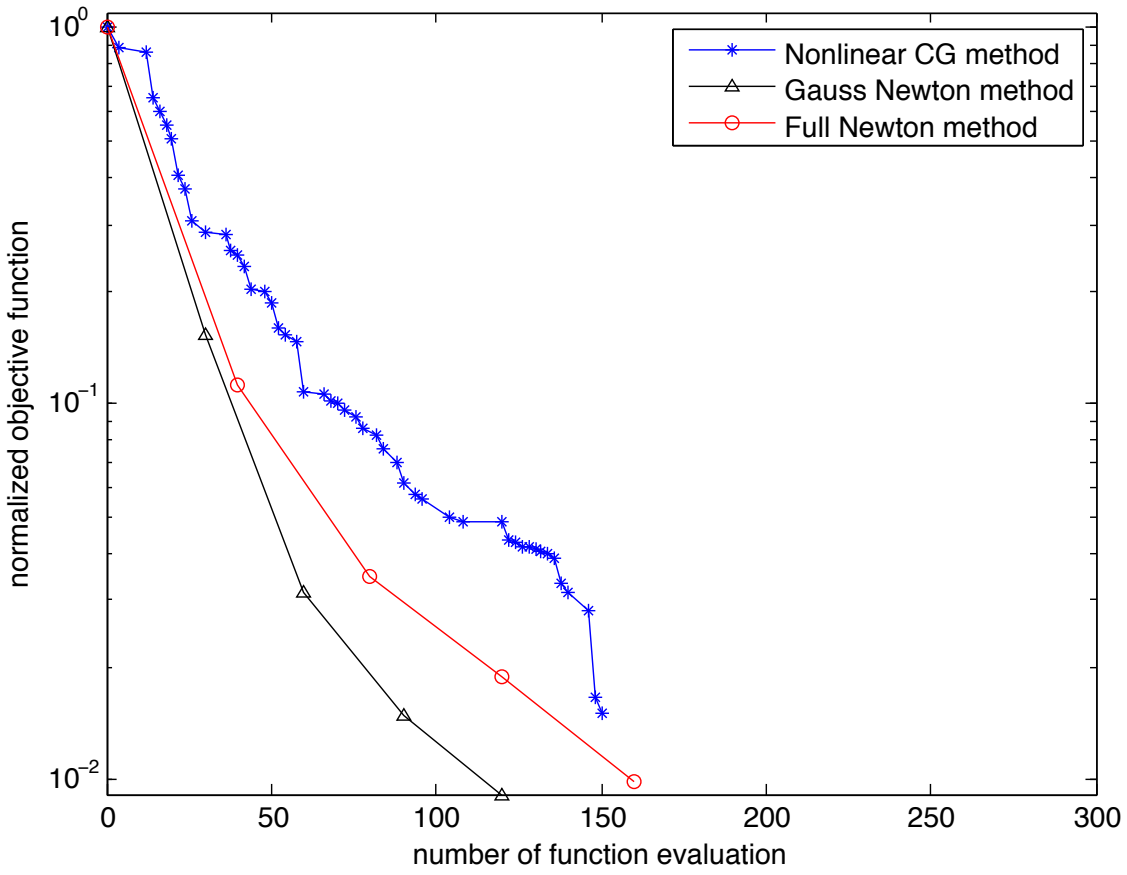

Figure 3: The normalized objective function versus the number of wave propagation. The blue curve represents the nonlinear CG method, black curve represents the Gauss-Newton method and red curve represents the full Newton method.[**CR**]

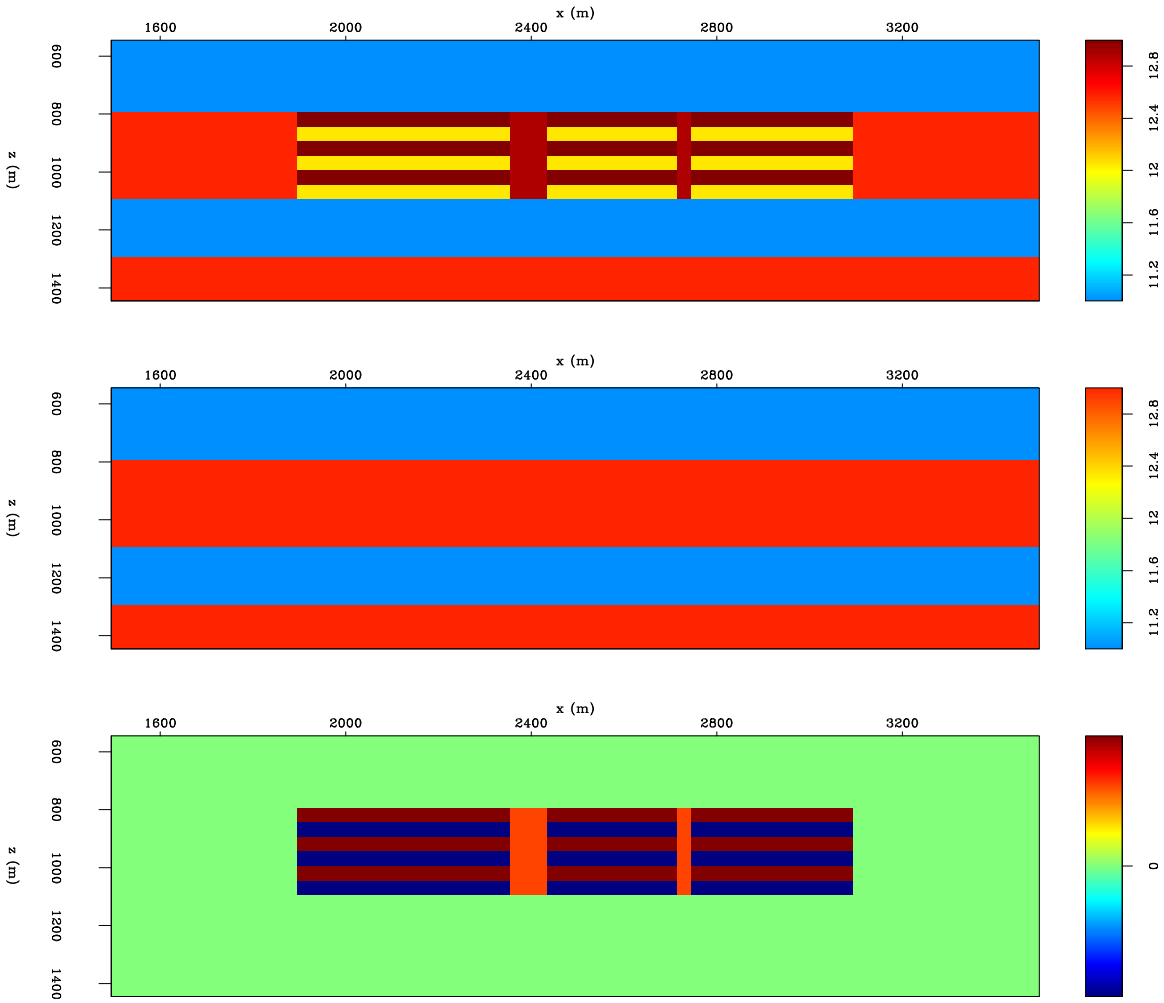

Figure 4: The top panel shows the true modulus model. The middle panel shows the starting modulus model. The bottom panel shows the difference between the true model and starting model.[**CR**]

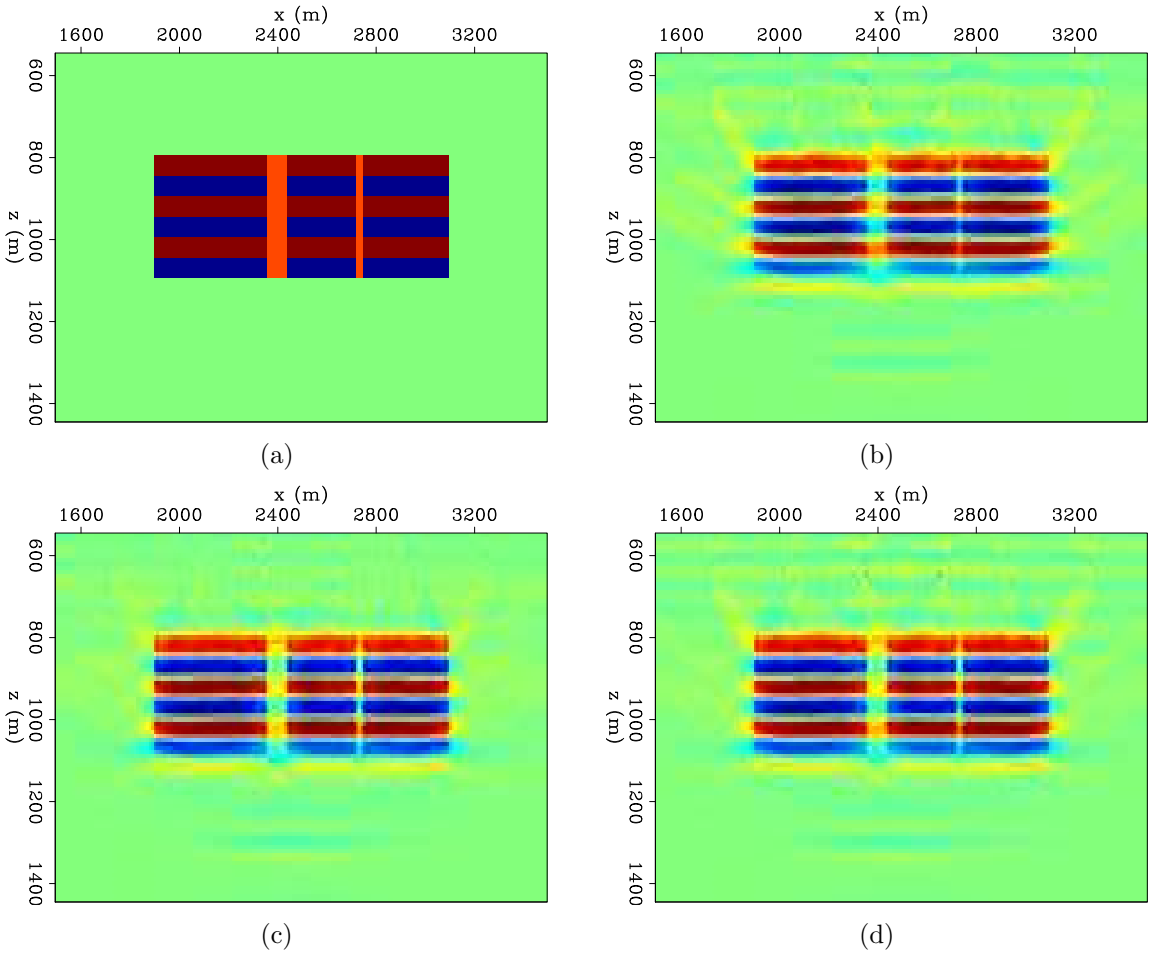

Figure 5: Panel (a) shows the difference between the true model and starting model. Panel (b) is the model update with the nonlinear CG method. Panel (c) is the model update with the Gauss-Newton method. Panel (d) is the model update with the full Newton method.[**CR**]

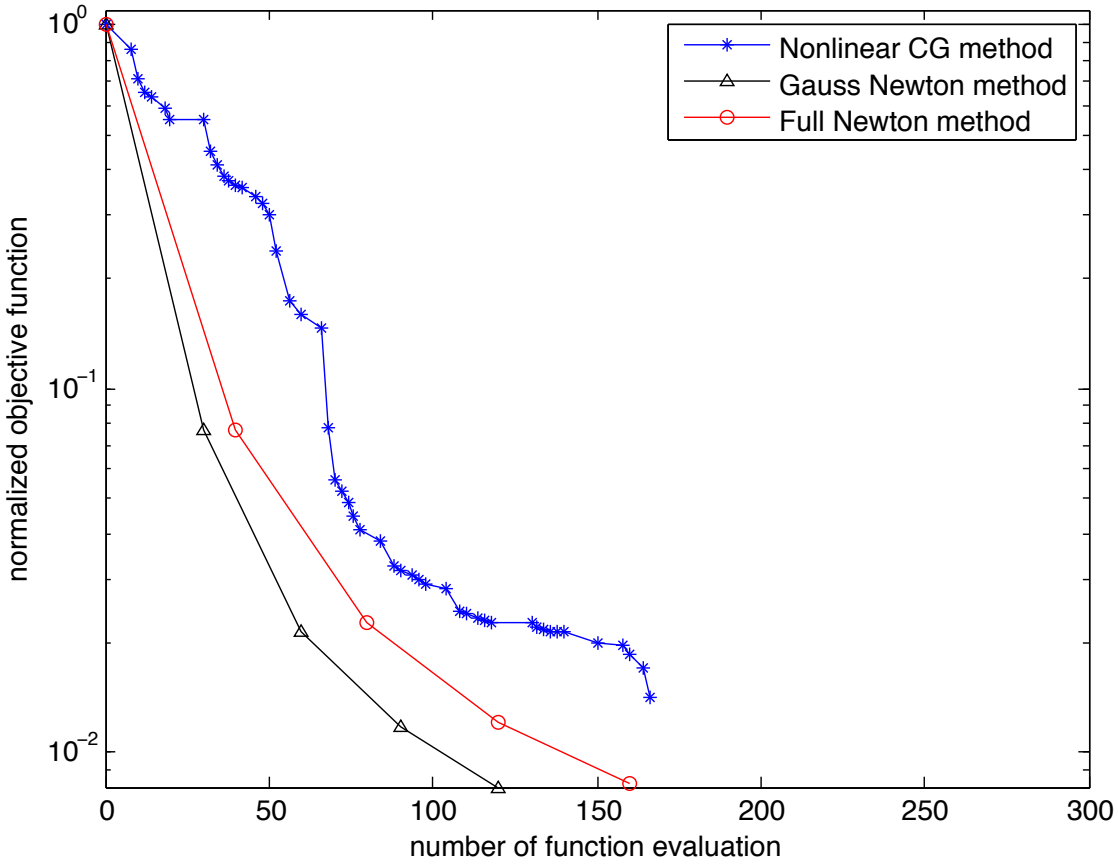

Figure 6: The normalized objective function versus the number of wave propagation. The blue curve represents the nonlinear CG method, black curve represents the Gauss-Newton method and red curve represents the full Newton method.[**CR**]

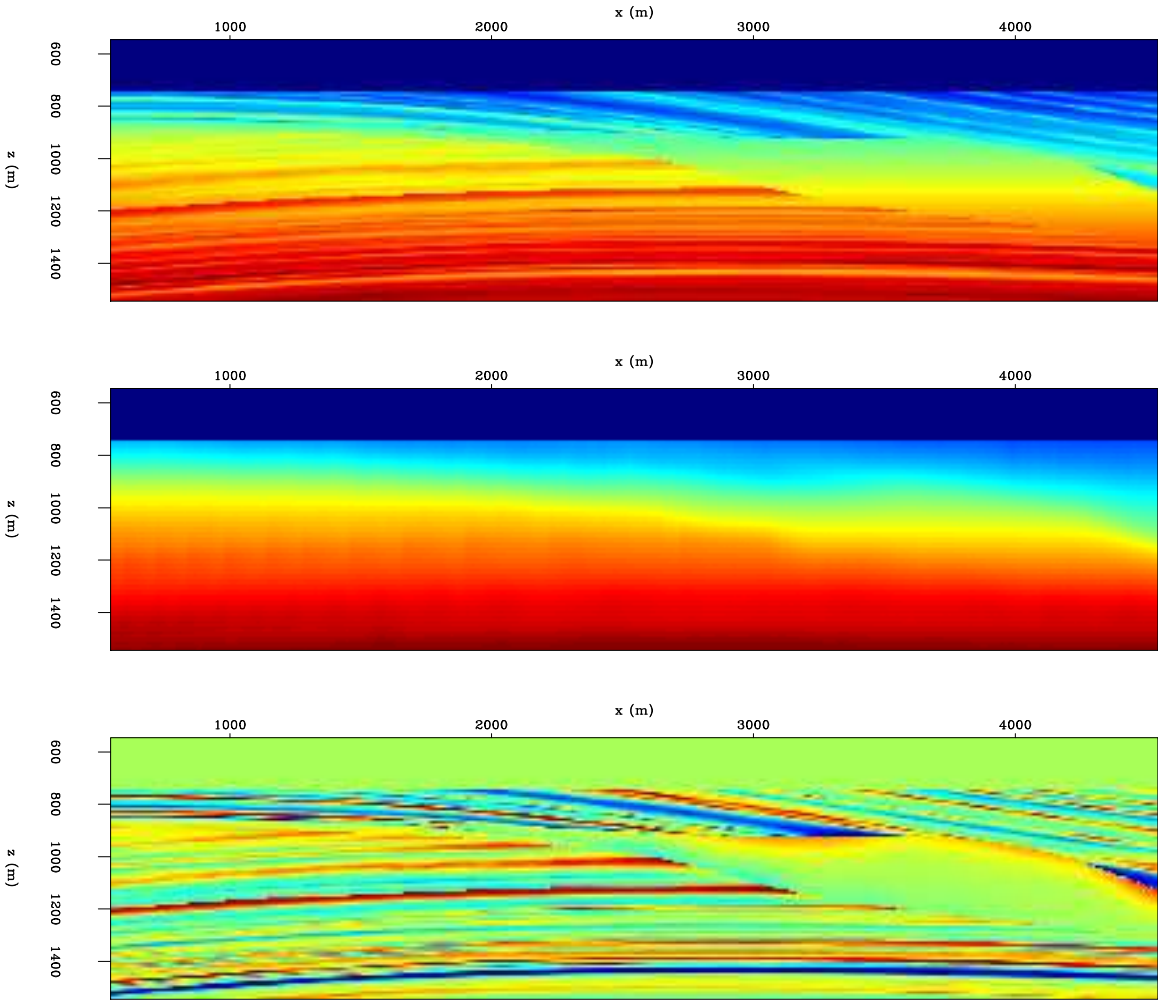

Figure 7: The top panel shows the true modulus model. The middle panel shows the starting modulus model. The bottom panel shows the difference between the true model and starting model.[**ER**]

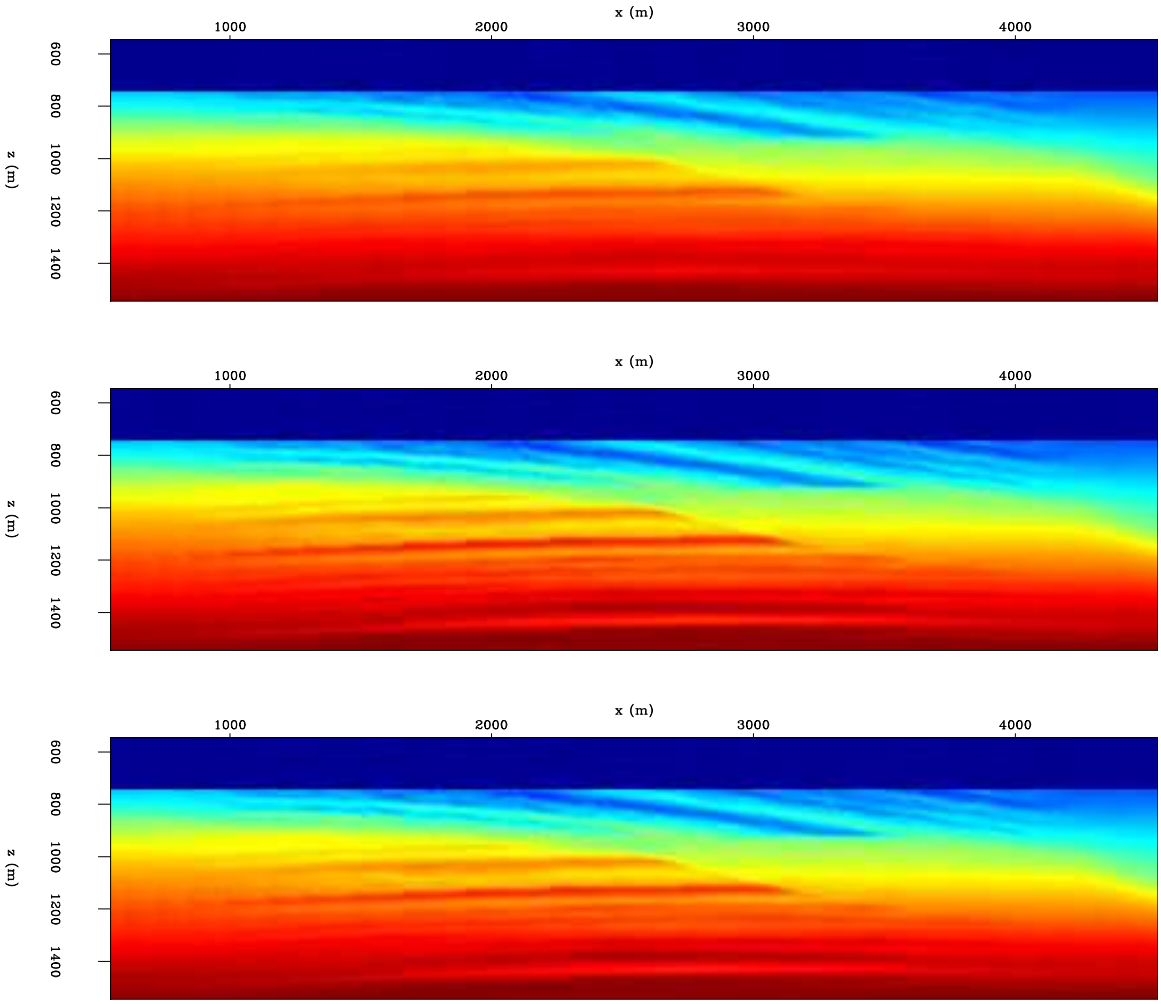

Figure 8: The top panel shows the inversion result with the nonlinear CG method. The middle panel shows the inversion result with the Gauss-Newton method. The bottom panel shows the inversion result with the full Newton method.[**CR**]

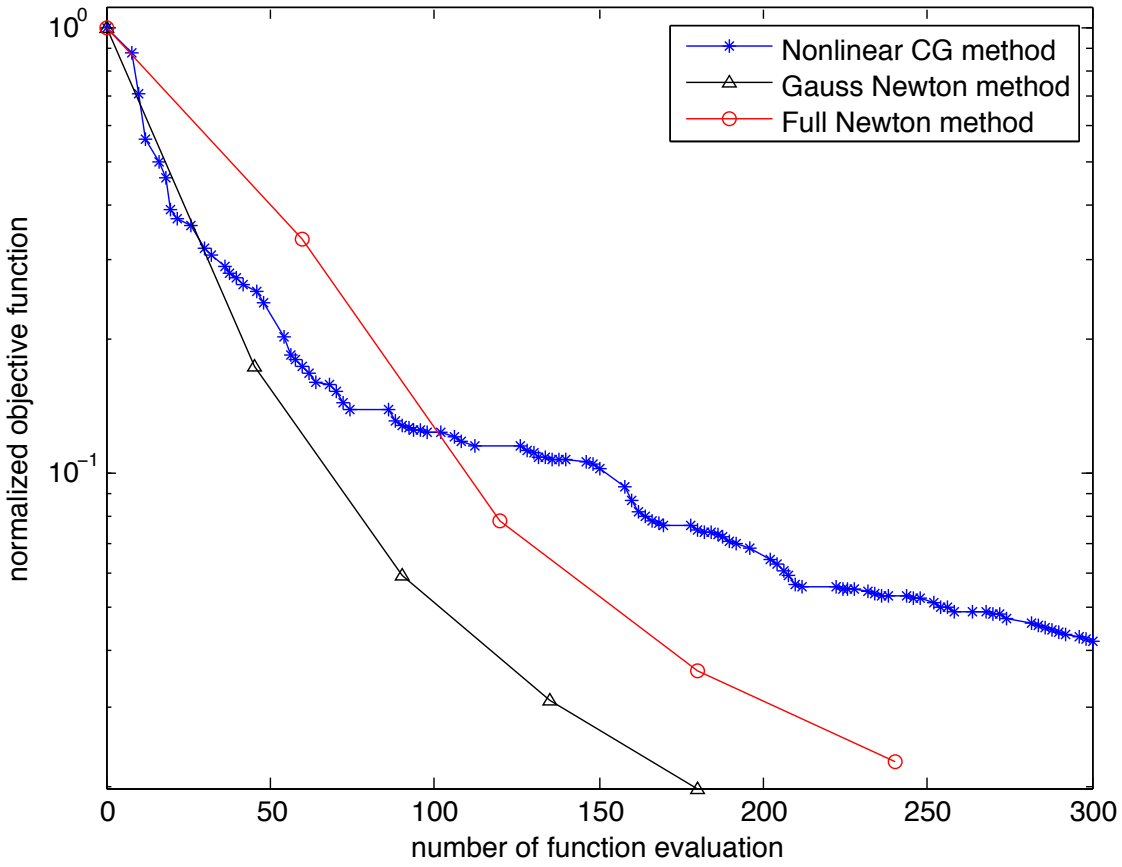

Figure 9: The normalized objective function versus the number of wave propagation. The blue curve represents the nonlinear CG method, black curve represents the Gauss-Newton method and red curve represents the full Newton method.[**CR**]

The nonlinear CG method and the Newton-type methods scale the gradient to get the search direction. It is interesting to examine how the gradient is scaled.

We take the estimated model K_i from the nonlinear CG method at $i = 30$ iteration. The difference between the true model and K_i, which is the ideal search direction, is shown in Figure 10(e). The gradient $\mathbf{g}_i$ is shown in Figure 10(a). The gradient can not be directly used as updating direction because the shallow component is much stronger than the deep component. This property of gradient result in the failure of the steepest descent method in FWI.

In Figure 10(b), we show the correction with nonlinear CG method,

$$\Delta K_i = -\mathbf{g}_i + \beta_i \mathbf{g}_{i-1}, \tag{19}$$

where β_i is obtained from Fletcher Reeves formula. After we scale the gradient in CG method, the search direction has more weight on the deep part. However, the component of search direction in the shallow part are not corrected properly. As the objective function is sensitive to shallow perturbation, we would not expect much decrease in the objective function.

For the Newton-type methods, we apply $\Delta K_{i,Newton} = \mathbf{H}^{-1}(-\mathbf{g}_i)$ for 15 CG steps, and the results are shown in Figure 10(c) and 10(d). The shallow component of the search direction is significantly different for the nonlinear CG method and the Newton-type methods as shown in Figure 10. The search direction for Newton-type methods approximate the ideal search direction reasonably well.

We have pointed out in the previous section, that the Gauss-Newton method converges faster than the Newton method at equivalent number of wave propagation. One reason is that each application full Hessian uses 4 wave propagation, and each application of the Gauss-Newton Hessian uses 3 wave propagation. Solver $\mathbf{H}^{-1}(-\mathbf{g})$ is therefore more expensive for the full Newton method. Another reason might be that because we use single precision float number for wave propagation, the CG algorithm for $\mathbf{H}^{-1}(-\mathbf{g})$ converges slower than the $\mathbf{H_{GN}}^{-1}(-\mathbf{g})$ in terms of CG steps.

CONCLUSION

In this paper, we implemented nonlinear CG method, Gauss-Newton method and full Newton method for solving FWI. We computed the gradient and Hessian with the adjoint-state method. Numerical results suggest Newton-type methods resolve fine structures better than the nonlinear CG method when the computational cost is the same.

ACKNOWLEDGEMENT

We would like to thank Ali Almomin, Robert Clapp for their help on the implementation of full-waveform inversion. We would also like to thank Ettor Biondi, Gustavo

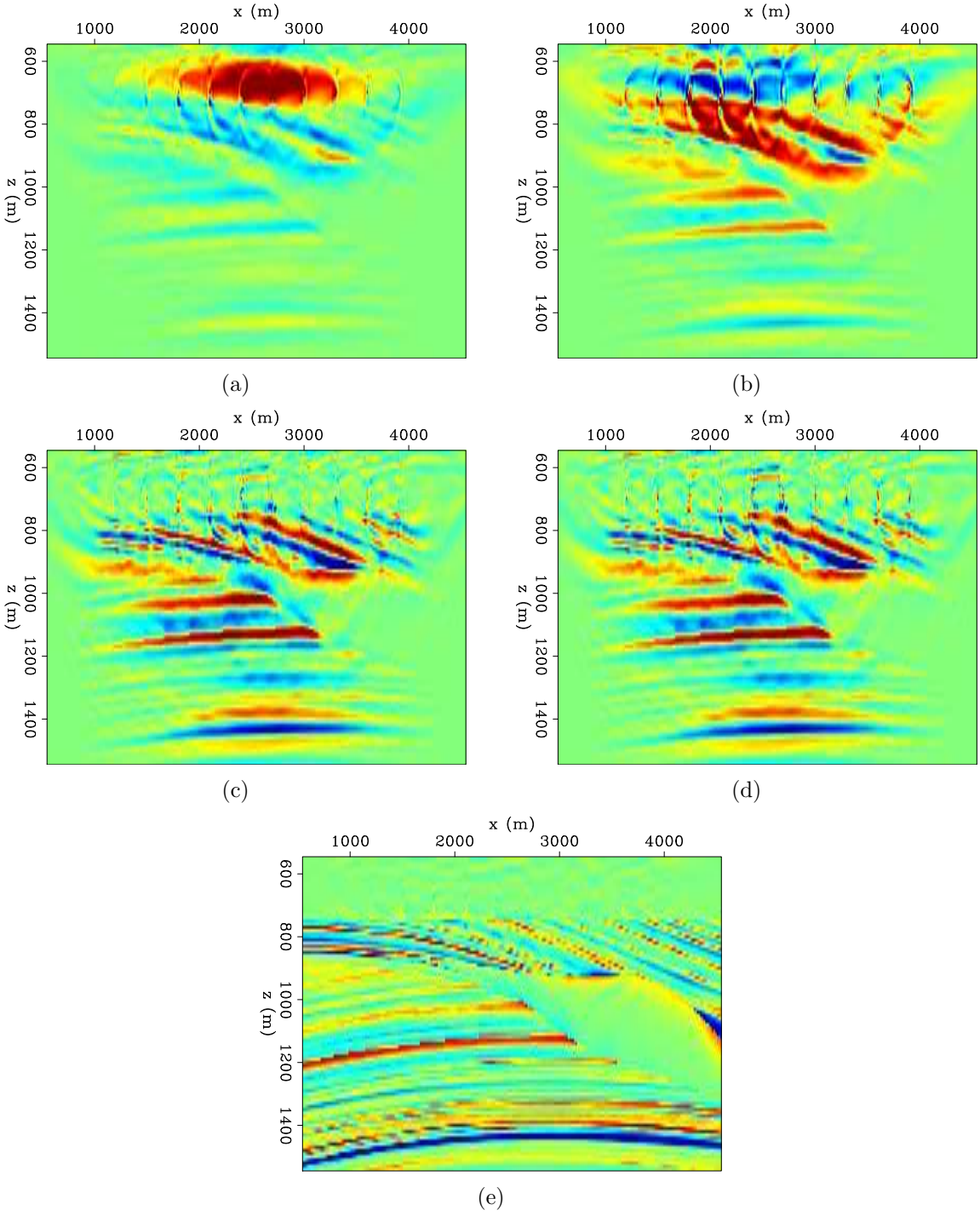

Figure 10: Panel (a) shows the gradient $\mathbf{g}(K_i)$, with K_i estimated from the nonlinear CG method at $i = 30$. Panel (b) is the model update ΔK_i with the nonlinear CG method. Panel (c) is the model update with the Gauss-Newton method after applying $\mathbf{H}_{\mathrm{GN}}^{-1}(-\mathbf{g}(K_i))$. Panel (d) is the model update with the full Newton method after applying $\tilde{\mathbf{H}}^{-1}(-\mathbf{g}(K_i))$. Panel (e) is the ideal model updating direction, which is the difference between true modulus model and current modulus model.[**CR**]

Alves, Ossian O'Reilly for useful suggestions on staggered-grid finite difference.

APPENDIX A: FULL-WAVEFORM INVERSION GRADIENT AND HESSIAN

In this appendix, we derive the Frechet derivative (Tromp et al., 2005; Plessix, 2006) and the Hessian (Fichtner, 2011) for the FWI objective function. We use acoustic approximation of wave equation with non-constant density. For simplicity, we define our model parameter as:

$$m_K \equiv \frac{1}{K}, \tag{20}$$

$$m_\rho \equiv \frac{1}{\rho}, \tag{21}$$

$$m \equiv \begin{bmatrix} m_K \\ m_\rho \end{bmatrix}. \tag{22}$$

The corresponding model perturbation can be expressed as follows:

$$\Delta m_K = -\frac{\Delta K}{K^2}, \tag{23}$$

$$\Delta m_\rho = -\frac{\Delta \rho}{\rho^2}, \tag{24}$$

The wave equation in acoustic media with non-constant density is shown in the following:

$$\begin{cases} [\frac{1}{K}\partial_t^2 - \tilde{\nabla}\cdot(\frac{1}{\rho}\nabla)]\mathbf{u} &= \mathbf{f} \\ \mathbf{u}(\mathbf{r}, t=0) &= 0 \\ \partial_t \mathbf{u}(\mathbf{r}, t=0) &= 0, \end{cases} \tag{25}$$

where K is the bulk modulus, ρ is the density, and $\mathbf{f}$ is the source wavefield. We use operator L to represent the operator:

$$L \equiv [\frac{1}{K}\partial_t^2 - \tilde{\nabla}\cdot(\frac{1}{\rho}\nabla)]. \tag{26}$$

The adjoint of wave equation used to compute receiver wavefield can be written as follows:

$$\begin{cases} [\frac{1}{K}\partial_t^2 - \nabla\cdot(\frac{1}{\rho}\tilde{\nabla})]\lambda &= \mathbf{d_{res}} \\ \lambda(\mathbf{r}, t=T) &= 0 \\ \partial_t \lambda(\mathbf{r}, t=T) &= 0, \end{cases} \tag{27}$$

where $\mathbf{d_{res}}$ is the difference between synthetic data and observed data. The receiver wavefield is computed backward in time, starting from maximum recording time $t = \mathbf{T}$ to $t = 0$.

Frechet derivative

For arbitrary model perturbation, we have

$$\left\langle \Delta m, \frac{\partial J}{\partial m} \right\rangle_{\mathbf{M}} = \left\langle \frac{\partial u}{\partial m}\Delta m, \frac{\partial J}{\partial u} \right\rangle_{\mathbf{U}}, \tag{28}$$

where the subscript **M** indicates the inner product is done in the model space, and and subscript **U** indicates the wavefield space.

Perturb the wave equation, we have

$$\left(\frac{\partial L}{\partial m}\Delta m\right) u + L\frac{\partial u}{\partial m}\Delta, m = 0 \tag{29}$$

Substitute equation (29) into equation (28), we get

$$\begin{aligned} \left\langle \Delta m, \frac{\partial J}{\partial m} \right\rangle_{\mathbf{M}} &= -\left\langle L^{-1}(\frac{\partial L}{\partial m}\Delta m)u, \frac{\partial J}{\partial u} \right\rangle_{\mathbf{U}} & (30)\\ &= -\left\langle (\frac{\partial L}{\partial m}\Delta m)u, L^{-*}\frac{\partial J}{\partial u} \right\rangle_{\mathbf{U}}. & (31)\\ & & (32) \end{aligned}$$

Introducing the receiver wavefield,

$$L^*\lambda = \frac{\partial J}{\partial u}, \tag{33}$$

with the explicit formula shown in equation (27). We have,

$$\left\langle \Delta m, \frac{\partial J}{\partial m} \right\rangle_{\mathbf{M}} = -\left\langle (\frac{\partial L}{\partial m}\Delta m)u, \lambda \right\rangle_{\mathbf{U}}. \tag{34}$$

For acoustic wave equation,

$$\begin{aligned} \left\langle \Delta m, \frac{\partial J}{\partial m} \right\rangle_{\mathbf{M}} &= -\left\langle [\delta m_K \partial_t^2 - \tilde{\nabla}\cdot(\delta m_\rho \nabla)]u, \lambda \right\rangle_{\mathbf{U}} & (35)\\ &= -\left\langle \delta m_K, \int_0^T \lambda \partial_2^t u dt \right\rangle_{\mathbf{M}} + \left\langle \Delta m_\rho, \int_0^T (\tilde{\nabla}^*\lambda)\cdot(\nabla u)dt \right\rangle_{\mathbf{M}} & (36) \end{aligned}$$

Thus,

$$\frac{\partial J}{\partial m} = \begin{bmatrix} \frac{\partial J}{\partial m_K} \\ \\ \frac{\partial J}{\partial m_\rho} \end{bmatrix} = \begin{bmatrix} -\int_0^T \lambda \partial_t^2 u dt \\ \\ \int_0^T (\tilde{\nabla}^*\lambda)\cdot(\nabla u)dt \end{bmatrix}. \tag{37}$$

The Frechet derivative with respect to (K, ρ) is,

$$\frac{\partial J}{\partial (K,\rho)} = \begin{bmatrix} \frac{\partial J}{\partial K} \\ \\ \frac{\partial J}{\partial \rho} \end{bmatrix} = \begin{bmatrix} \frac{1}{K^2}\int_0^T \lambda \partial_t^2 u dt \\ \\ -\frac{1}{\rho^2}\int_0^T (\tilde{\nabla}^*\lambda)\cdot(\nabla u)dt \end{bmatrix} \tag{38}$$

Full Hessian and Gauss-Newton Hessian

We are interested in the action of Hessian:

$$H(\Delta m_1, \Delta m_2) = \langle \Delta m_2, H \Delta m_1 \rangle_{\mathbf{M}} \,, \tag{39}$$

where Δm_1 and Δm_2 are two model perturbation. We use shorthand notation for the bilinear form in this subsection:

$$F(x_1, x_2) \equiv \langle x_1, F x_2 \rangle \,. \tag{40}$$

Our goal is to find an expression,

$$\langle \Delta m_2, H \Delta m_1 \rangle_{\mathbf{M}} = \langle \delta m_2, A(\Delta m) \rangle_{\mathbf{M}} \,, \tag{41}$$

for arbitrary Δm_2, where $A(\Delta m_1)$ does not depend on Δm_2 and can be computed efficiently. Then we claim $A(\Delta m_1)$ is equivalent to the action of Hessian applied to Δm_1.

Expand the bilinear form in equation (39),

$$\begin{aligned} \langle \Delta m_2, H \Delta m_1 \rangle_{\mathbf{M}} &= \left\langle \frac{\partial u}{\partial m} \Delta m_2, \frac{\partial^2 J}{\partial u^2} \frac{\partial u}{\partial m} \Delta m_1 \right\rangle_{\mathbf{U}} & (42) \\ &+ \left\langle \frac{\partial J}{\partial u}, \frac{\partial^2 u}{\partial m^2} (\Delta m_1, \Delta m_2) \right\rangle_{\mathbf{U}} . & (43) \end{aligned}$$

The bilinear form for Gauss-Newton Hessian can be obtained by neglecting line (43),

$$\langle \Delta m_2, H_{GN} \Delta m_1 \rangle_{\mathbf{M}} = \left\langle \frac{\partial u}{\partial m} \Delta m_2, \frac{\partial^2 J}{\partial u^2} \frac{\partial u}{\partial m} \Delta m_1 \right\rangle_{\mathbf{U}} . \tag{44}$$

where H_{GN} represent the Gauss-Newton Hessian.

In order to eliminate $\frac{\partial^2 u}{\partial m^2}(\Delta m_1, \Delta m_2)$ in equation (42 and 43) which depends on Δm_2, we need the first order and second order perturbation of the wave equation:

$$(\frac{\partial L}{\partial m} \Delta m_1) u + L \frac{\partial u}{\partial m} \Delta m_1 = 0, \tag{45}$$

$$(\frac{\partial L}{\partial m} \Delta m_2) u + L \frac{\partial u}{\partial m} \Delta m_2 = 0, \tag{46}$$

and,

$$(\frac{\partial^2 L}{\partial m^2}(\Delta m_1, \Delta m_2)) u + (\frac{\partial L}{\partial m} \Delta m_1)(\frac{\partial u}{\partial m} \Delta m_2) + (\frac{\partial L}{\partial m} \Delta m_2) \frac{\partial u}{\partial m} \Delta m_1 + L \frac{\partial^2 u}{\partial m^2}(\Delta m_1, \Delta m_2) = 0. \tag{47}$$

Substitute equation (47) into equation (42), we get,

$$\langle \Delta m_2, H\Delta m_1 \rangle_{\mathbf{M}} = \left\langle \frac{\partial u}{\partial m}\Delta m_2, \frac{\partial^2 J}{\partial u^2}\frac{\partial u}{\partial m}\Delta m_1 \right\rangle_{\mathbf{U}} \quad (48)$$

$$+ \left\langle \frac{\partial J}{\partial u}, \frac{\partial^2 u}{\partial m^2}(\Delta m_1, \Delta m_2) \right\rangle_{\mathbf{U}}. \quad (49)$$

Define $\delta u_1 \equiv \frac{\partial u}{\partial m}\Delta m_1$ which can be computed from Born approximation, as follows:

$$L(m)\Delta u_1 = -(\frac{\partial L}{\partial m}\delta m_1)u, \quad (50)$$

Use equation (33 and 50), and after some algebra,

$$\begin{aligned}\langle \Delta m_2, H\Delta m_1 \rangle_{\mathbf{M}} &= \left\langle \frac{\partial u}{\partial m}\Delta m_2, -(\frac{\partial L}{\partial m}\Delta m_1)^*\lambda + \frac{\partial^2 J}{\partial u^2}\delta u_1 \right\rangle_{\mathbf{U}} \\ &\quad - \left\langle \lambda, (\frac{\partial^2 L}{\partial m^2}(\Delta m_1, \Delta m_2))u \right\rangle_{\mathbf{U}} \\ &\quad - \left\langle \lambda, (\frac{\partial L}{\partial m}\Delta m_2)\Delta u_1 \right\rangle_{\mathbf{U}}.\end{aligned}$$

We then eliminate $\frac{\partial u}{\partial m}\Delta m_2$ term using equation (46), and get,

$$\langle \Delta m_2, H\Delta m_1 \rangle_{\mathbf{M}} = -\left\langle (\frac{\partial L}{\partial m}\Delta m_2)u, L^{-*}\{-(\frac{\partial L}{\partial m}\Delta m_1)^*\lambda + \frac{\partial^2 J}{\partial u^2}\Delta u_1\} \right\rangle_{\mathbf{U}} \quad (51)$$

$$- \left\langle \lambda, (\frac{\partial^2 L}{\partial m^2}(\Delta m_1, \Delta m_2))u \right\rangle_{\mathbf{U}} \quad (52)$$

$$- \left\langle \lambda, (\frac{\partial L}{\partial m}\Delta m_2)\Delta u_1 \right\rangle_{\mathbf{U}} \quad (53)$$

$$= -\left\langle \mu_a + \mu_b, (\frac{\partial L}{\partial m}\Delta m_2)u \right\rangle_{\mathbf{U}} \quad (54)$$

$$- \left\langle \lambda, (\frac{\partial^2 L}{\partial m^2}(\Delta m_1, \Delta m_2))u \right\rangle_{\mathbf{U}} \quad (55)$$

$$- \left\langle \lambda, (\frac{\partial L}{\partial m}\Delta m_2)\Delta u_1 \right\rangle_{\mathbf{U}}, \quad (56)$$

where we define,

$$L(m)^*\mu_a = -(\frac{\partial L}{\partial m}\Delta m_1)^*\lambda, \quad (57)$$

$$L(m)^*\mu_b = \frac{\partial^2 J}{\partial u^2}\Delta u_1. \quad (58)$$

Thus, our solution for Hessian in the general case:

$$\begin{aligned}
\langle \Delta m_2, H\Delta m_1\rangle_{\mathbf{M}} &= \left\langle \mu, (\frac{\partial L}{\partial m}\Delta m_2)u \right\rangle_{\mathbf{U}} && (59)\\
&+ \left\langle \lambda, (\frac{\partial^2 L}{\partial m^2}(\Delta m_1, \Delta m_2))u \right\rangle_{\mathbf{U}} && (60)\\
&+ \left\langle \lambda, (\frac{\partial L}{\partial m}\Delta m_2)\delta u_1 \right\rangle_{\mathbf{U}}, && (61)
\end{aligned}$$

with each wavefield computed as follows:

$$\begin{aligned}
L(m)u &= f, && (62)\\
L(m)^*\lambda &= \frac{\partial J}{\partial u}, && (63)\\
L(m)\delta u_1 &= -(\frac{\partial L}{\partial m}\delta m_1)u, && (64)\\
L(m)^*\mu_a &= \frac{\partial^2 J}{\partial u^2}\delta u_1, && (65)\\
L(m)^*\mu_b &= -(\frac{\partial L}{\partial m}\delta m_1)^*\lambda, && (66)
\end{aligned}$$

For acoustic wave equation, we get,

$$H\Delta m = \begin{bmatrix} -\int_0^T(\mu_a(x,t)+\mu_b(x,t))\partial_t^2 u(x,t)dt - \int_0^T \lambda(x,t)\partial_t^2\Delta u_1(x,t)dt \\ \int_0^T \tilde{\nabla}^*(\mu_a(x,t)+\mu_b(x,t)\cdot\nabla u(x,t)dt + \int_0^T \tilde{\nabla}^*\lambda(x,t)\nabla\Delta u_1(x,t)dt, \end{bmatrix}. \quad (67)$$

with the 2 forward wave propagation and 2 backward wave propagation,

$$\begin{aligned}
[m_K\partial_t^2 - \tilde{\nabla}\cdot(m_\rho\nabla)]u &= f, && (68)\\
[\frac{1}{K}\partial_t^2 - \tilde{\nabla}\cdot(\frac{1}{\rho}\nabla)]^*\lambda &= \sum_r S_r^*(S_r u - d_r), && (69)\\
[m_K\partial_t^2 - \tilde{\nabla}\cdot(m_\rho\nabla)]\delta u_1 &= -[\Delta m_K\partial_t^2 - \tilde{\nabla}\cdot(\Delta m_\rho\nabla)]u, && (70)\\
[m_K\partial_t^2 - \tilde{\nabla}\cdot(m_\rho\nabla)]^*\mu_a &= S_r^* S_r\delta u_1, && (71)\\
[m_K\partial_t^2 - \tilde{\nabla}\cdot(m_\rho\nabla)]^*\mu_b &= -[\Delta m_K\partial_t^2 - \tilde{\nabla}\cdot(\Delta m_\rho\nabla)]^*\lambda. && (72)
\end{aligned}$$

The Hessian with respect to (K,ρ) therefore can be written as:

$$H\begin{bmatrix}\Delta K\\ \Delta\rho\end{bmatrix} = \begin{bmatrix} \frac{\Delta K}{K^2}\int_0^T(\mu_a(x,t)+\mu_b(x,t))\partial_t^2 u(x,t)dt + \frac{\Delta K}{K^2}\int_0^T \lambda(x,t)\partial_t^2\Delta u_1(x,t)dt \\ -\frac{\Delta\rho}{\rho^2}\int_0^T \tilde{\nabla}^*(\mu_a(x,t)+\mu_b(x,t)\cdot\nabla u(x,t)dt - \frac{\Delta\rho}{\rho^2}\int_0^T \tilde{\nabla}^*\lambda(x,t)\nabla\Delta u_1(x,t)dt \end{bmatrix}. \quad (73)$$

For Gauss-Newton Hessian, we can follow similar derivation, and write:

$$H_{GN}\Delta m = \begin{bmatrix} -\int_0^T \mu_a(x,t)\partial_t^2 u(x,t)dt \\ \\ \int_0^T \tilde{\nabla}^* \mu_a(x,t)\cdot\nabla u(x,t)dt \end{bmatrix}. \tag{74}$$

The difference between the full Hessian and the Gauss-Newton Hessian can be expressed as follows:

$$(H - H_{GN})\Delta m = \begin{bmatrix} -\int_0^T \mu_b(x,t)\partial_t^2 u(x,t)dt - \int_0^T \lambda(x,t)\partial_t^2 \Delta u_1(x,t)dt \\ \\ \int_0^T \tilde{\nabla}^* \mu_a(x,t)dt + \int_0^T \tilde{\nabla}^* \lambda(x,t)\nabla\Delta u_1(x,t)dt \end{bmatrix}. \tag{75}$$

REFERENCES

Fichtner, A., 2011, Full seismic waveform modeling and inversion: Springer.

Fichtner, A. and J. Trampert, 2011, Hessian kernels of seismic data functionals based upon adjoint techniques: Geophysical Journal International, **185**, 775–798.

Maharramov, M. and B. Biondi, 2013, Simultaneous time-lapse full waveform inversion: SEP 150, 1–8.

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

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

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.

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

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216.

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