American Journal of Computational Mathematics, 2021, 11, 241-266
https://www.scirp.org/journal/ajcm
ISSN Online: 2161-1211
ISSN Print: 2161-1203

Elastic Full Waveform Inversion Based on the Trust Region Strategy

Wensheng Zhang[1,2]*, Yijun Li[1,2]

[1]State Key Laboratory of Scientific and Engineering Computing, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
[2]School of Mathematical Sciences, University of Chinese Academy Sciences, Beijing, China
Email: *zws@lsec.cc.ac.cn

How to cite this paper: Zhang, W.S. and Li, Y.J. (2021) Elastic Full Waveform Inversion Based on the Trust Region Strategy. *American Journal of Computational Mathematics*, **11**, 241-266.
https://doi.org/10.4236/ajcm.2021.114016

Received: August 27, 2021
Accepted: November 12, 2021
Published: November 15, 2021

Abstract

In this paper, we investigate the elastic wave full-waveform inversion (FWI) based on the trust region method. The FWI is an optimization problem of minimizing the misfit between the observed data and simulated data. Usually, the line search method is used to update the model parameters iteratively. The line search method generates a search direction first and then finds a suitable step length along the direction. In the trust region method, it defines a trial step length within a certain neighborhood of the current iterate point and then solves a trust region subproblem. The theoretical methods for the trust region FWI with the Newton type method are described. The algorithms for the truncated Newton method with the line search strategy and for the Gauss-Newton method with the trust region strategy are presented. Numerical computations of FWI for the Marmousi model by the L-BFGS method, the Gauss-Newton method and the truncated Newton method are completed. The comparisons between the line search strategy and the trust region strategy are given and show that the trust region method is more efficient than the line search method and both the Gauss-Newton and truncated Newton methods are more accurate than the L-BFGS method.

Keywords

Elastic Wave Equations, Full-Waveform Inversion, Trust Region Strategy, Line Search Strategy, Newton-Type Method, Time Domain

1. Introduction

The physical properties of the earth can be inverted quantitatively by the reflected seismic waves. In modern seismology, traveltime inversion is the main technique approach in the 19th century. Since the 1980s [1] [2], full waveform inversion (FWI) has become a rapidly developing subject because of its high resolution for reconstructing media parameters.

In FWI, the Fréchet derivative is required to estimate explicitly before proceeding to the inversion of the linearized system [3]. However, the computational cost of calculating Fréchet or gradient is too large. So some other methods have been taken into account. In [1] [2], the authors proposed to compute the gradient by solving an adjoint problem, which is considered as the initial of time-domain FWI. Since then, the FWI has been an important research topic in exploration geophysics. In the 1990s, the time-domain FWI was extended to the frequency domain [4] [5]. The frequency-domain FWI has the advantage of high computational efficiency and flexibility of data selection [6].

The FWI is actually an optimization iterative process to solve a nonlinear optimization problem. Generally, the computational cost of global optimization methods is extensive large. For this reason, local optimization algorithms are adopted. To avoid iteration falling to an unreasonable local minimum point, the multiscale method was proposed [7] in 1995. The multiscale method performs the inversion from the low frequency to high-frequency information step by step. It includes the long-wavelength information of the model and thus it enhances the robustness of inversion effectively. It can overcome the problem of local minima caused by the poor initial model selection. This technique has been successfully applied to the time-domain acoustic FWI, see e.g. [8] [9]. In order to reconstruct the long wavelength of the macroscopic velocity model, the Laplace-domain FWI is proposed [10]. A hybrid domain method combining the Laplace-domain method and the frequency-domain method together is also developed [11]. This method can be regarded as a special case of frequency-domain damping inversion.

In elastic FWI, there are three different parameters, *i.e.*, density and two Lamé parameters. In [12], the authors pointed out that different parameters have coupled effects on the seismic response which is called the trade-off effect or crosstalk. In order to tame the trade-off effect, several hierarchical inversion strategies are developed. In [13], Lamé parameters are inverted with fixed density first and then velocities are updated. In [14], all parameters are inverted simultaneously. However, the overall steplength is estimated by calculating an optimal steplength for every parameter individually.

The Hessian matrix plays an important role in FWI. It contains the information of multiple scattering wavefields and the trade-off effects among different parameters. The off-diagonal blocks of Hessian reflect the relationship between different parameters [12]. The Newton-type method such as the quasi-Newton method, the Gauss-Newton method [15] and the truncated Newton method [16] all contain the Hessian information and so they behave better than the steepest descent method in accuracy. In the calculation of the gradient vector, the adjoint method is usually used [17]. The adjoint method can be extended to the truncated Newton method [18]. For the elastic wave equations, the wave propagator operator is symmetric. The back propagator operator and the forward propagator operator are identical. Besides that, the Gauss-Newton approximation of Hessian matrix is positive and definite. This property can be used in the trust region strategy. Our investigations in this paper show that the trust region method can accelerate the convergence rate and improve the inversion accuracy.

In this paper, we investigate the application of the trust region strategy in FWI. The rest of our paper is organized as follows. In Section 2, the theoretical methods are described in detail. In Section 3, numerical computations and comparisons are presented. Finally, in Section 4, the conclusion is given.

2. Theory

In this section, we will introduce the forward problem first. Then we will describe FWI based on the trust region method. In order to solve the subproblem in the trust region algorithm, the two dimensional subspace method is described.

2.1. Forward Method

The source-free time-domain two dimensional (2D) elastic wave equations can be written as [19]

$$\rho\frac{\partial^2 u}{\partial t^2}=\frac{\partial}{\partial x}\left[\lambda\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+2\mu\frac{\partial u}{\partial x}\right]+\frac{\partial}{\partial y}\left[\mu\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)\right], \tag{2.1}$$

$$\rho\frac{\partial^2 v}{\partial t^2}=\frac{\partial}{\partial y}\left[\lambda\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right)+2\mu\frac{\partial v}{\partial y}\right]+\frac{\partial}{\partial x}\left[\mu\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)\right], \tag{2.2}$$

where ρ is the density, λ and μ are the Lamé parameters, $u(x,y,t)$ and $v(x,y,t)$ are the displacement in the horizontal x and vertical y directions respectively. Furthermore, (2.1) and (2.2) can be rewritten as the following stress-velocity formulation after adding a source

$$\begin{cases}\rho\dfrac{\partial v_x}{\partial t}=\dfrac{\partial\sigma_{xx}}{\partial x}+\dfrac{\partial\sigma_{xy}}{\partial y}+f_x,\\ \rho\dfrac{\partial v_y}{\partial t}=\dfrac{\partial\sigma_{xy}}{\partial x}+\dfrac{\partial\sigma_{yy}}{\partial y}+f_y,\\ \dfrac{\partial\sigma_{xx}}{\partial t}=(\lambda+2\mu)\dfrac{\partial v_x}{\partial x}+\lambda\dfrac{\partial v_y}{\partial y},\\ \dfrac{\partial\sigma_{yy}}{\partial t}=\lambda\dfrac{\partial v_x}{\partial x}+(\lambda+2\mu)\dfrac{\partial v_y}{\partial y},\\ \dfrac{\partial\sigma_{xy}}{\partial t}=\mu\left(\dfrac{\partial v_x}{\partial y}+\dfrac{\partial v_y}{\partial x}\right),\end{cases} \tag{2.3}$$

where v_x and v_y are the particle velocity components in the x and y directions respectively, σ_{xx}, σ_{yy}, σ_{xy} are the stress tensor components. Here we have added the source, *i.e.*, the body force f_x and f_y vector to v_x and v_y components respectively. We rewrite the system (2.3) in the following matrix form

$$\rho\partial_t\underbrace{\begin{bmatrix}v_x\\ v_y\end{bmatrix}}_{\boldsymbol{v}}=\underbrace{\begin{bmatrix}\partial x & \partial y & 0\\ 0 & \partial x & \partial y\end{bmatrix}}_{-D^{\dagger}}\underbrace{\begin{bmatrix}\sigma_{xx}\\ \sigma_{xy}\\ \sigma_{yy}\end{bmatrix}}_{\boldsymbol{\sigma}}+\underbrace{\begin{bmatrix}f_x\\ f_y\end{bmatrix}}_{f_{\boldsymbol{v}}}, \tag{2.4}$$

$$\partial_t \underbrace{\begin{bmatrix} \sigma_{xx} \\ \sigma_{xy} \\ \sigma_{yy} \end{bmatrix}}_{\boldsymbol{\sigma}} = \underbrace{\begin{bmatrix} \lambda+2\mu & 0 & \lambda \\ 0 & \mu & 0 \\ \lambda & 0 & \lambda+2\mu \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} \partial_x & 0 \\ \partial_y & \partial_x \\ 0 & \partial_y \end{bmatrix}}_{-D} \underbrace{\begin{bmatrix} v_x \\ v_y \end{bmatrix}}_{\boldsymbol{v}}, \tag{2.5}$$

where ∂_t, ∂_x, ∂_y denote $\frac{\partial}{\partial t}$, $\frac{\partial}{\partial x}$, $\frac{\partial}{\partial y}$. The symbol $\dagger$ denotes the adjoint operator. Note that the determinant of the matrix C is positive. So from (2.4) and (2.5), we have

$$\underbrace{\begin{bmatrix} \rho\partial_t & D^{\dagger} \\ D & C^{-1}\partial_t \end{bmatrix}}_{A(\boldsymbol{m})} \underbrace{\begin{bmatrix} \boldsymbol{v} \\ \boldsymbol{\sigma} \end{bmatrix}}_{\boldsymbol{w}} = \underbrace{\begin{bmatrix} f_v \\ 0 \end{bmatrix}}_{\boldsymbol{s}}. \tag{2.6}$$

The compact form of the forward problem is

$$A(\boldsymbol{m}(\boldsymbol{x}))\boldsymbol{w}(\boldsymbol{x},t;\boldsymbol{x}_s) = \boldsymbol{s}(\boldsymbol{x},t;\boldsymbol{x}_s), \tag{2.7}$$

where $\boldsymbol{m} = (\rho,\lambda,\mu)$ is the model parameters, $\boldsymbol{w} = (\boldsymbol{v},\boldsymbol{\sigma})$. In (2.7), the wave propagator $A(\boldsymbol{m})$ is symmetric. Given the media parameters ρ, λ and μ, the forward problem is to solve system (2.6) numerically.

We have many numerical methods to solve system (2.6), for example, the finite volume method [20] and the staggered-grid method [21] and so on. In this paper, we use the staggered-grid method for its high computational efficiency and relatively easy code implementation. Since the computational domain is finite, the absorbing boundary conditions [22] are required to absorb the boundary reflections. In this paper, we adopt the perfectly matched layer (PML) method [23] [24]. The forward discrete schemes for solving system (2.3) are given in **Appendix A**.

2.2. Full Waveform Inversion

The FWI is a data-fitting procedure to find the model parameter $\boldsymbol{m}$ by minimizing the data misfit $\delta\boldsymbol{d}$ between the simulated data $\boldsymbol{w}^{cal}$ and the observed data $\boldsymbol{w}^{obs}$. The objective function can be formulated as

$$\chi(\boldsymbol{m}) = \frac{1}{2}\sum_{s=1}^{N_s}\sum_{r=1}^{N_r}\int_0^T \left(\boldsymbol{w}^{obs} - \boldsymbol{w}^{cal}\right)^2 \mathrm{d}t, \tag{2.8}$$

where T is the total recording duration, N_s and N_r are the number of the sources and receivers respectively. Once the media parameters ρ, λ, and μ are inverted, we can yield the P-wave velocity and S-wave velocity by

$$v_p = \sqrt{\frac{\lambda+2\mu}{\rho}}, \quad v_s = \sqrt{\frac{\mu}{\rho}}, \tag{2.9}$$

respectively.

The misfit function $\chi(\boldsymbol{m})$ is minimized iteratively. The update procedure in FWI at k-th iteration can be expressed as

$$\boldsymbol{m}_k = \boldsymbol{m}_{k-1} + \alpha_k \boldsymbol{p}_k, \quad k = 1,2,\cdots, \tag{2.10}$$

where $\boldsymbol{m}_{k+1}$ and $\boldsymbol{m}_k$ are the model parameter at $k+1$ and k iteration respectively, α^k denotes the step length at k iteration, $\boldsymbol{p}_k$ is the descent direction. The step length α_k must be found by using the line search technique for nonlinear inverse problem.

Neglecting the summation in (2.8) and introducing the restriction operator R at receiver points, then the gradient of the objective function with respect to the model parameter $\boldsymbol{m}$ is

$$\nabla_{\boldsymbol{m}}\chi = \frac{\partial \chi}{\partial m_i} = \int_0^T \left(\frac{\partial \boldsymbol{w}}{\partial m_i}\right)^{\dagger} R^{\dagger}\left(R\boldsymbol{w} - \boldsymbol{w}^{obs}\right)\mathrm{d}t, \tag{2.11}$$

where $\dagger$ is the adjoint operator. Furthermore, the Hessian matrix can be obtained by taking derivative of (2.11) with respect to the model parameter

$$H_{ij} = \underbrace{\int_0^T \left(R\frac{\partial \boldsymbol{w}}{\partial m_i}\right)^{\dagger}\left(R\frac{\partial \boldsymbol{w}}{\partial m_j}\right)\mathrm{d}t}_{(H_1)_{ij}} + \underbrace{\int_0^T \left(\frac{\partial^2 \boldsymbol{w}}{\partial m_i \partial m_j}\right)^{\dagger} R^{\dagger}\left(R\boldsymbol{w} - \boldsymbol{d}_{obs}\right)\mathrm{d}t}_{(H_2)_{ij}}. \tag{2.12}$$

The expression of Hessian H is divided into two parts, *i.e.*, H_1 and H_2. The first part H_1 is the Gauss-Newton approximation of Hessian [15], which just contains the 1st-order derivative of the wavefield. The second part H_2 contains the 2nd-order derivative of the wavefield. The full Newton method consists of both the parts H_1 and H_2 [15]. The truncated Newton is based on the computations of descent direction by the conjugated gradient algorithm in the full Newton framework. The BFGS method is an efficient algorithm to compute the inverse of Hessian approximation in the Gauss-Newton method. It is the first quasi-Newton algorithm and named for its discovers Broyden, Fletcher, Goldfarb and Shanno. To save storage memory further, the limited-memory BFGS (L-BFGS) algorithm is usually applied [25] [26] [27] in computations.

2.3. Gradient Calculation

The gradient of the objective function can be obtained by the Lagrangian formulation [17] [18] [28] or the perturbation theory [29]. Following the perturbation theory [29], the gradient of the objective function with respect to model parameters can be rewritten as

$$\frac{\partial \chi}{\partial \rho} = -\sum_{x_s}\int_0^T \left(\frac{\partial u_x}{\partial t}\frac{\partial \tilde{u}_x}{\partial t} + \frac{\partial u_y}{\partial t}\frac{\partial \tilde{u}_y}{\partial t}\right)\mathrm{d}t, \tag{2.13}$$

$$\frac{\partial \chi}{\partial \lambda} = -\sum_{x_s}\int_0^T \left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y}\right)\left(\frac{\partial \tilde{u}_x}{\partial x} + \frac{\partial \tilde{u}_y}{\partial y}\right)\mathrm{d}t, \tag{2.14}$$

$$\frac{\partial \chi}{\partial \mu} = -\sum_{x_s}\int_0^T \left\{\left(\frac{\partial u_y}{\partial x} + \frac{\partial u_x}{\partial y}\right)\left(\frac{\partial \tilde{u}_x}{\partial x} + \frac{\partial \tilde{u}_y}{\partial y}\right) + 2\left(\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_x}{\partial x} + \frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_y}{\partial y}\right)\right\}\mathrm{d}t, \tag{2.15}$$

where $u_x := u$ and $u_y := v$ are the forward displacements in the x and y directions respectively, and $\tilde{u}_x$ and $\tilde{u}_y$ are the corresponding backward displacements in the x and y directions respectively. Equations (2.13)-(2.15) are the gradient expressions in terms of the displacements.

We solve the forward problem based on (2.3) instead of (2.1)-(2.2) by the staggered-grid method. So it is necessary to express the gradient (2.13)-(2.15) with the particle velocity and the stress tensor. The results are the following

$$\frac{\partial \chi}{\partial \rho} = -\sum_{x_s} \int_0^T \left(v_x \tilde{v}_x + v_y \tilde{v}_y \right) \mathrm{d}t, \tag{2.16}$$

$$\frac{\partial \chi}{\partial \lambda} = -\sum_{x_s} \int_0^T \frac{\left(\sigma_{xx} + \sigma_{yy} \right)\left(\tilde{\sigma}_{xx} + \tilde{\sigma}_{yy} \right)}{4\left(\lambda + \mu \right)^2} \mathrm{d}t, \tag{2.17}$$

$$\begin{aligned} \frac{\partial \chi}{\partial \mu} = -\sum_{x_s} \int_0^T \Bigg\{ \frac{\sigma_{xy} \tilde{\sigma}_{xy}}{\mu^2} + \frac{1}{4} \Bigg[\frac{\left(\sigma_{xx} + \sigma_{yy} \right)\left(\tilde{\sigma}_{xx} + \tilde{\sigma}_{yy} \right)}{\left(\lambda + \mu \right)^2} \\ + \frac{\left(\sigma_{xx} - \sigma_{yy} \right)\left(\tilde{\sigma}_{xx} - \tilde{\sigma}_{yy} \right)}{\mu^2} \Bigg] \Bigg\} \mathrm{d}t, \end{aligned} \tag{2.18}$$

where v_x, v_y, σ_{xx}, σ_{yy}, σ_{xy} are the solutions of forward problem (2.3) whereas $\tilde{v}_x$, $\tilde{v}_y$, $\tilde{\sigma}_{xx}$, $\tilde{\sigma}_{yy}$ and $\tilde{\sigma}_{xy}$ are the solutions of corresponding adjoint problem. The derivation is given in **Appendix B**.

2.4. Trust Region Method

There are two strategies to solve a minimization problem. One is the line search method and the other is the trust region method. In order to describe the trust-region method more clearly and directly, we consider minimizing an abstract objection function $f(x)$, *i.e.*, $\min_x f(x)$. Then the line search strategy is

$$x_{k+1} = x_k + \alpha_k p_k. \tag{2.19}$$

The line search chooses a direction p_k first and then finds the step length α_k along this direction. The direction may be the steepest descent direction, the Newton direction and so on. After fixing the search direction, the search length α_k is searched by approximately solving a one-dimensional minimization problem

$$\min_{\alpha > 0} f\left(x_k + \alpha p_k \right). \tag{2.20}$$

Usually, the inexact line search methods such as the Wolfe method and the interpolation method are used [26] since solving (2.20) exactly is too expensive. A popular inexact method is the Wolfe method [30]. In this paper, we use the Wolfe method.

In the trust region method, information gathered from objective function $f(x)$ is used to construct a model function $\tilde{f}_k$ whose behavior near the current point x_k is similar to that of the objective function f_k. Because $\tilde{f}_k$ may not be a good approximation of f when x is far from x_k, we restrict the search for a minimizer of $\tilde{f}_k$ to some region around x_k. Namely, we find the step p by approximately solving the following subproblem

$$\begin{aligned} &\min_p \tilde{f}_k\left(x_k + p \right), \\ &\text{s.t. } x_k + p \in \text{the trust region} \end{aligned} \tag{2.21}$$

If the solution of the problem (2.21) does not produce a sufficient decrease in $f(x)$, it means the trust region is too large, and we shrink it and resolve (2.21). Usually, the trust region is a ball defined by $\|p_k\| \le \Delta$, where $\Delta > 0$ is called the trust region radius. Here the norm is $\|\cdot\|_2$ norm.

The model function $\tilde{f}_k$ in (2.21) is usually defined to be a quadratic function

$$\tilde{f}_k(p) = f_k + \nabla f_k^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} B_k p, \tag{2.22}$$

where f_k and ∇f_k are the objective function and the gradient at x_k respectively, and the matrix B_k is either the Hessian $\nabla^2 f_k$ or some approximation to it like the Gauss-Newton method.

So combining (2.21) and (2.22) together, the form of the trust region method is

$$\tilde{f}_k(p) = f_k + \nabla f_k^{\mathrm{T}} p + \frac{1}{2} p^{\mathrm{T}} B_k p, \quad \text{s.t. } \|p\| \le \Delta_k, \tag{2.23}$$

where Δ_k is the radius. This is the trust region subproblem and we solve it by the two dimensional subspace method [26]. Note that the two dimensional subspace method requires the Hessian is symmetric and positive.

In a sense, the line search method and the trust region method differ in that the line search method starts by fixing the direction p_k and then searches the step length α_k whereas in trust region the direction p_k and the search step α_k are determined together subject to the trust radius Δ_k. Theoretically, the trust region is easier to reach quadric convergence than the line search method. More theoretical details can be found in the references (e.g., [31] [32] [33]).

One key step in the trust region method is the choice of the trust region radius Δ_k at each iteration. The choice is based on the agreement between the objective function f and the model function $\tilde{f}_k$ at previous iteration. Given a step p_k, we define the ratio ρ_k as

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{\tilde{f}_k(0) - \tilde{f}_k(p_k)}, \tag{2.24}$$

where the numerator is the actual reduction, and the denominator is the predicted reduction.

If $\|p_k\|$ reaches the boundary of the trust region or the ratio ρ_k is satisfactory enough, the radius is increased by a constant time:

$$\Delta_{k+1} = \alpha_1 \Delta_k, \quad \alpha_1 > 1. \tag{2.25}$$

If the ratio ρ_k is too small, which means that the model function $\tilde{f}_k$ is not a good approximation to the objective function f under the radius, then we reduce the trust region radius:

$$\Delta_{k+1} = \alpha_2 \Delta_k, \quad 0 < \alpha_2 < 1. \tag{2.26}$$

In this paper, we choose $\alpha_1 = 2$ and $\alpha_2 = \frac{1}{4}$.

2.5. FWI Algorithms

In this section, we present the algorithm of the truncated Newton method with the line search strategy and the algorithm of the Gauss-Newton method with the trust-region strategy.

Algorithm 1 is the FWI algorithm by using the truncated Newton method based on the line search method. There are two loops in **Algorithm 1**. The outer loop is the iteration of the line search strategy, and the inner loop is to compute the iterative direction p_k and the step length α_k. The core of the inner loop is Hessian-vector product $B_k d_j$ and so on. **Algorithm 2** is the FWI algorithm by using the Gauss-Newton method based on the trust region method. In **Algorithm 2**, $\tilde{B}_k$ is the Hessian approximation by the Gauss-Newton method. Comparing **Algorithm 1** and **Algorithm 2**, we know the starting part of both algorithms is the same. The difference is in the rear part of the algorithm. In **Algorithm 1**, the line search strategy consists of the function of computing the step length and the iterative direction. On the contrary, in **Algorithm 2**, the trust region strategy combines the two steps together.

Algorithm 1 Truncated Newton FWI by the line search method

```
Given the initial model $\boldsymbol{m}^0$ and $\epsilon > 0$ and $\delta > 0$.
for $k = 0, 1, 2, \ldots$
    1. Compute forward wavefield $\boldsymbol{w}$. Extract $\boldsymbol{w}^{cal}$ at receiver points.
    2. Compute residuals $\delta\boldsymbol{w} = \boldsymbol{w}^{cal} - \boldsymbol{w}^{obs}$.
    3. Compute backforward wavefield $\widetilde{\boldsymbol{w}}$ generated by $\delta\boldsymbol{w}$ at receiver points.
    4. Compute gradient $\boldsymbol{g}$ based on $\boldsymbol{w}$ and $\widetilde{\boldsymbol{w}}$.
    5. Define tolerance $\epsilon_k = \min(0.5, \sqrt{\|\boldsymbol{g}\|})\|\boldsymbol{g}\|$.
       Set $\boldsymbol{z}_0 = 0$, $\boldsymbol{r}_0 = \boldsymbol{g}$, $\boldsymbol{d}_0 = -r_0 = -\boldsymbol{g}$
    6. for $k = 0, 1, 2, \ldots$
        (a) if $d_j^T B_k d_j \le 0$
                if $j = 0$
                    return $p_k = -\boldsymbol{g}$
                else
                    return $p_k = z_j$
               end (if)
            end (if)
        (b) Set $\alpha_j = r_j^T r_j / d_j^T B_k d_j$
        (c) Set $z_{j+1} = z_j + \alpha_j d_j$
        (d) Set $r_{j+1} = r_j + \alpha_j B_k d_j$
        (e) if $\|r_{j+1}\| < \delta$
                return $p_k = z_{j+1}$
            end (if)
        (f) Set $\beta_{j+1} = r_{j+1}^T r_{j+1} / r_j^T r_j$
        (g) Set $d_{j+1} = -r_{j+1} + \beta_{j+1} d_j$
       end (for).
    7. Compute step length $\alpha_k$ satisfying the Wolfe conditions.
    8. Update model $\boldsymbol{m}^{k+1} = \boldsymbol{m}^k + \alpha_k p_k$.
    9. if $\|f(x_k) - f(x_{k-1})\| / \|f(x_k)\| \le \epsilon$
           break
       end (if)
end (for)
```

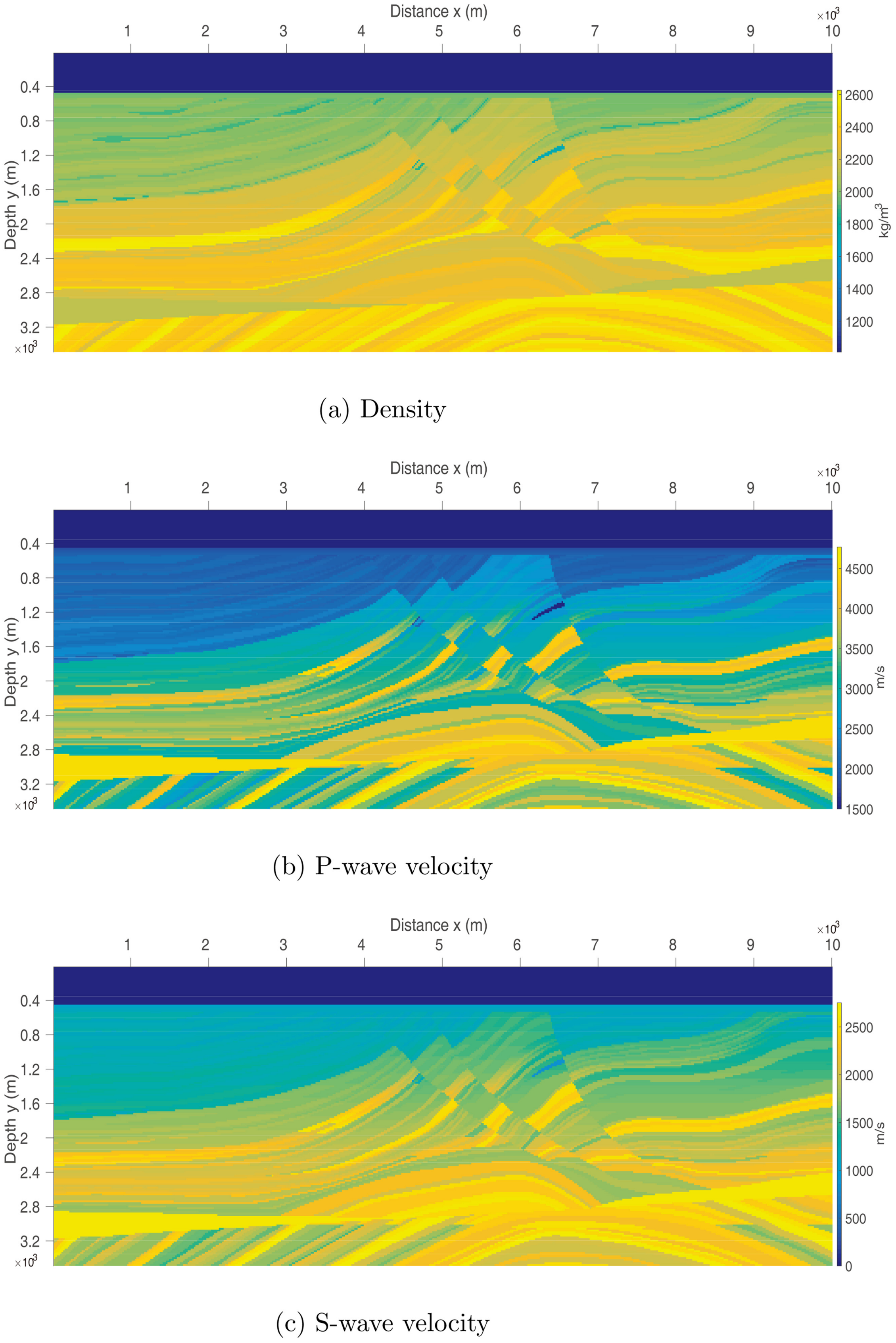
Algorithm 2 Gauss-Newton FWI by the trust region method

```
Given the initial model $\boldsymbol{m}^0$, $\epsilon > 0, \hat{\Delta} > 0, \Delta_0 \in (0, \hat{\Delta})$, and $\eta \in \left[0, \frac{1}{4}\right)$.
for $k = 0, 1, 2, \ldots$
  1. Compute forward wavefield $\boldsymbol{w}$. Extract $\boldsymbol{w}^{cal}$ at receiver points.
  2. Compute residuals $\delta \boldsymbol{w} = \boldsymbol{w}^{cal} - \boldsymbol{w}^{obs}$.
  3. Compute backward wavefield $\widetilde{\boldsymbol{w}}$ generated by $\delta \boldsymbol{w}$ at receiver points.
  4. Compute gradient $\boldsymbol{g}$ based on $\boldsymbol{w}$ and $\widetilde{\boldsymbol{w}}$.
  5. for $k = 0, 1, 2, \ldots$
       (a) if $d_j^T \widetilde{B}_k d_j \leq 0$
               if $j = 0$ return $p_k = -\boldsymbol{g}$
               else return $p_k = z_j$
              end (if)
           end (if)
       (b) Set $\alpha_j = r_j^T r_j / d_j^T \widetilde{B}_k d_j$
       (c) Set $z_{j+1} = z_j + \alpha_j d_j$
       (d) Set $r_{j+1} = r_j + \alpha_j \widetilde{B}_k d_j$
       (e) if $\|r_{j+1}\| < \delta$ return $p_k = z_{j+1}$
           end (if)
       (f) Set $\beta_{j+1} = r_{j+1}^T r_{j+1} / r_j^T r_j$
       (g) Set $d_{j+1} = -r_{j+1} + \beta_{j+1} d_j$
     end (for).
  6. Obtain $\boldsymbol{p}_k$ by solving subproblem (2.23).
  7. Evaluate $\rho_k$ by (2.24).
  8. if $\rho_k < \frac{1}{4}$ then $\Delta_{k+1} = \frac{1}{4}\Delta_k$
     else if $\rho_k > \frac{3}{4}$ and $\|q_k\| = \Delta_k$ then $\Delta_{k+1} = \min\left(2\Delta_k, \hat{\Delta}\right)$
         else $\Delta_{k+1} = \Delta_k$ end (if )
     end (if )
  9. if $\rho_k > \eta$
         $\boldsymbol{m}_{k+1} = \boldsymbol{m}_k + \boldsymbol{p}_k$
     else
         $\boldsymbol{m}_{k+1} = \boldsymbol{m}_k$
     end (if)
 10. if $\|f(x_k) - f(x_{k-1})\| / \|f(x_k)\| \leq \epsilon$ break
     end (if)
end (for)
```

3. Numerical Computations

In this section, numerical computations are presented to illustrate the advantages of the trust region method. We also present the inversion results by the L-BFGS method for comparisons. We choose the L-BFGS method to make comparison as it is the most popular quasi-Newton method.

We consider FWI for the Marmousi model which is usually applied to test the ability of an imaging or inversion algorithm. The exact Marmousi model is shown in Figure 1. Figures 1(a)-(c) are the density, P-wave and S-wave velocities respectively. The model is 10 km in length along horizontal direction and 3.48 km



(a) Density

(b) P-wave velocity

(c) S-wave velocity

Figure 1. The exact Marmousi model. (a) Density; (b) P-wave velocity; (c) S-wave velocity.

in depth. It is grided by the 500 × 174 mesh with 20 m spatial steps in both x and y directions. We set 100 sources and 400 receivers in the data acquisition. The sources are located at 40 m in depth and the first source is set at $x = 800\ \text{m}$ in the horizontal direction. The receivers are located at 460 m in depth with the first receiver is set at $x = 800\ \text{m}$ in the horizontal direction. The source space is 40 m and the receiver space is 20 m. The source is the Ricker wavelet with the time function as

$$s(t) = \left(1 - \frac{1}{2}\omega^2 t^2\right)\exp\left(-\frac{1}{4}\omega^2 t^2\right), \tag{3.1}$$

where $\omega = 2\pi f_0$ and $f_0 = 7\ \text{Hz}$ is the central frequency. The source is loaded on the v_x component. The recording time is 6 s with the time interval 2 ms.

For the elastic wave FWI, the density is more sensitive than the velocities. Some hierarchical inversion strategies developed by [13] and [14] can be applied. In our FWI computations, we don't apply these strategies. We invert the three media parameters simultaneously. However, a multiscale technique by implementing the inversion from low frequency to high frequency is still used. Moreover, the data of the next high frequency band include the data of the previous low frequency band which guarantees the robustness of inversion. In our computations, four groups of frequencies are used to complete inversion step by step, *i.e.*, 2 Hz, 5 Hz, 10 Hz and 20 Hz. For every stage, the stopping criterion is

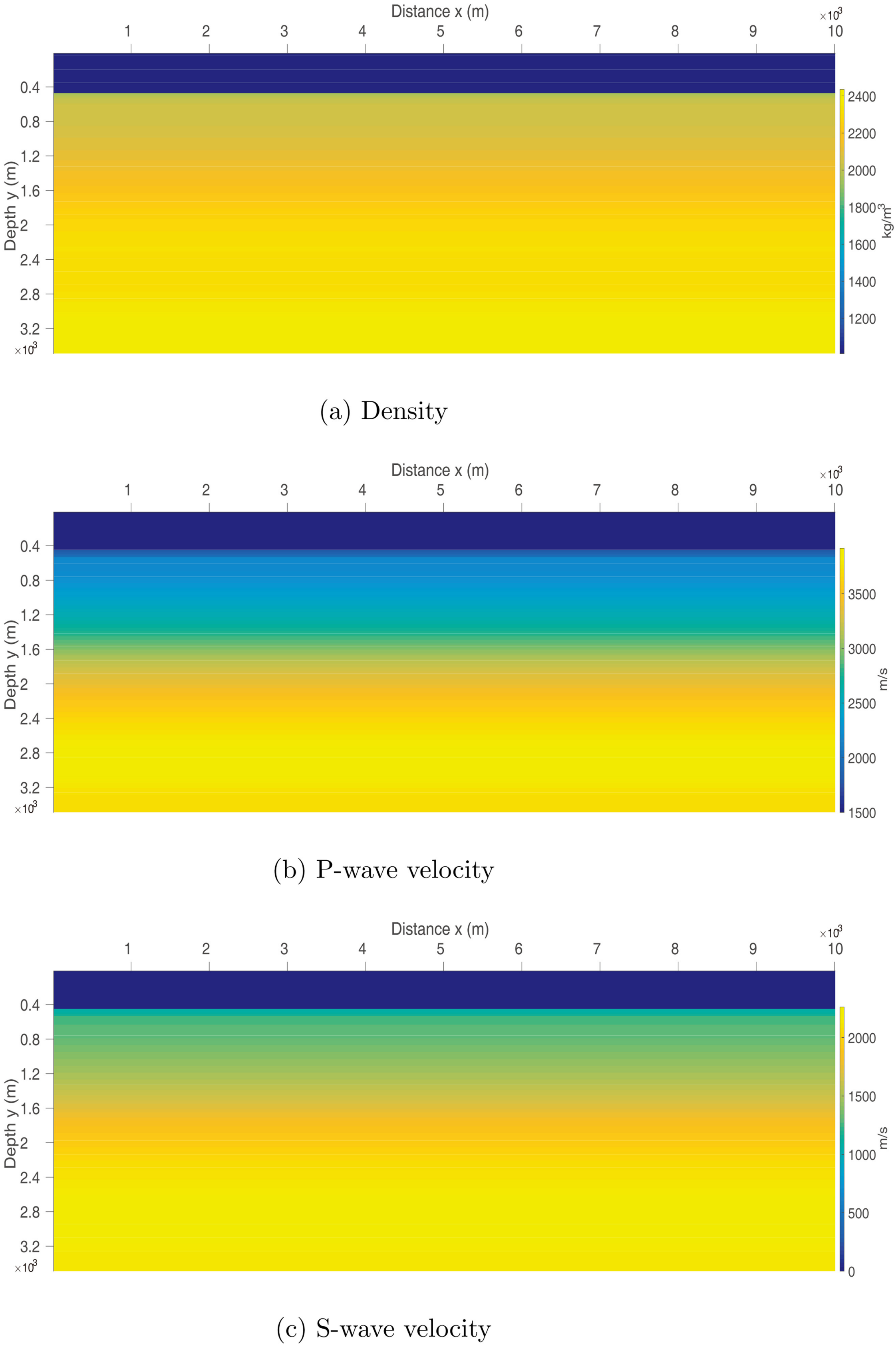
$$\frac{\left\| f(x_k) - f(x_{k-2}) \right\|}{\left\| f(x_k) \right\|} \le \varepsilon, \tag{3.2}$$

where $f(x_k)$ is in fact the objective function computed by (2.8), and ε is set 0.01 in our computation. This stopping criterion is based on the relative decrease of the misfit function.

The initial model for FWI is shown in **Figure 2**. **Figures 2(a)-(c)** are the initial density, P-wave and S-wave velocities respectively. They are generated by arithmetic average smoothing for the exact model along the depth direction only. In **Figure 2**, we can see there are no geological structures such as the faults in the initial model like those in the exact model.

The final inversion result by the Gauss-Newton method with the line search strategy is shown in **Figure 3**. The final inversion result by the Gauss-Newton method with the two dimensional subspace strategy is shown in **Figure 4**. The final inversion result by the truncated Newton method with the line search strategy is shown in **Figure 5**. We also complete the FWI by the L-BFGS. The results are similar and we omit the figures for saving space. We remark that the Hessian in the truncated Newton method is not positive and so the two dimensional subspace strategy can not be applied directly. Comparing **Figures 3-5** with **Figure 1**, we can see that both the Gauss-Newton method and the truncated Newton method with the line search strategy can yield good imaging results. And the case of the Gauss-Newton method with the two dimensional subspace strategy is the same.

Figure 6 shows the profile comparison at the position $x = 5\ \text{km}$ among the

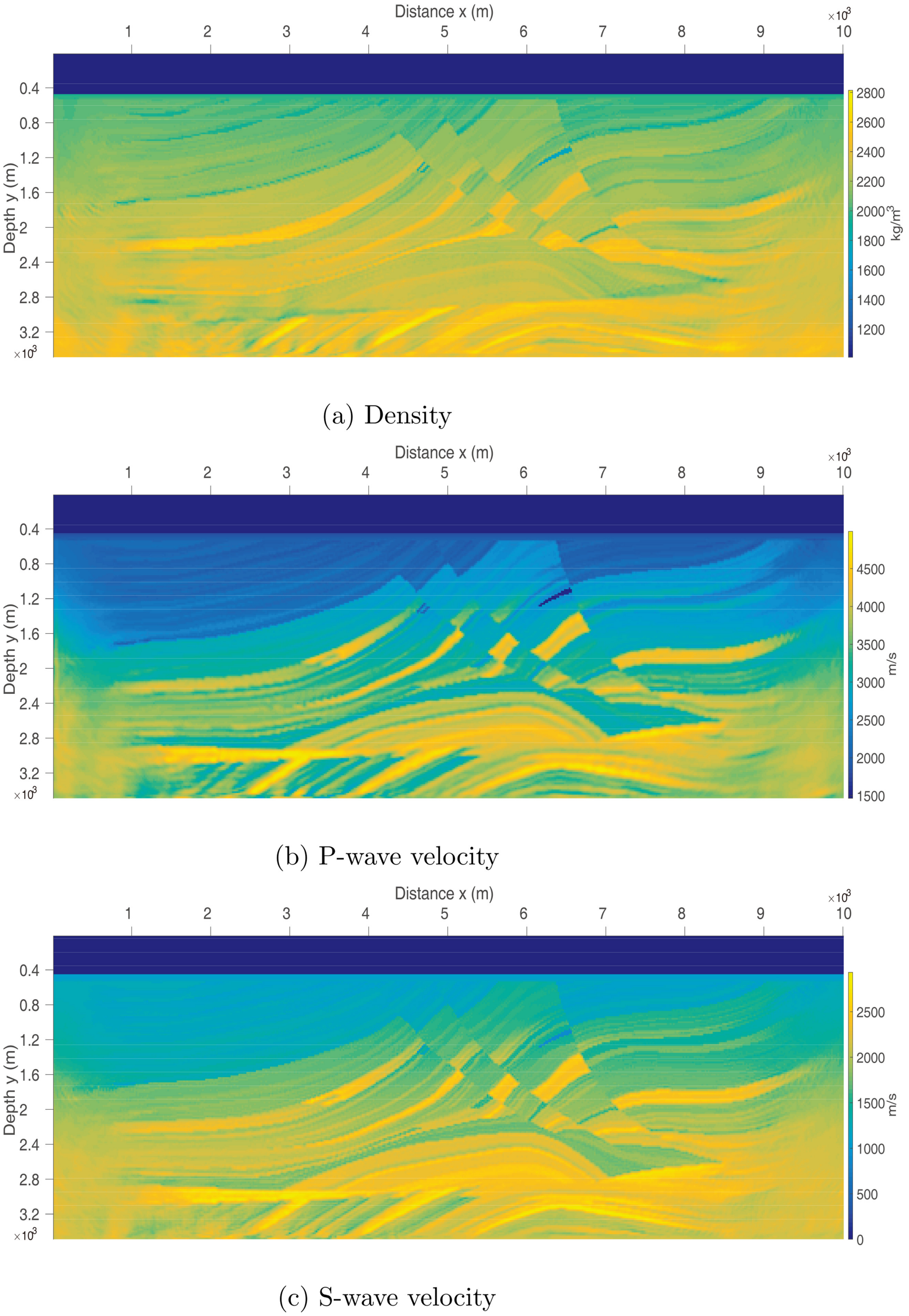


(a) Density

(b) P-wave velocity

(c) S-wave velocity

Figure 2. The initial model for FWI. (a) Density; (b) P-wave velocity; (c) S-wave velocity.

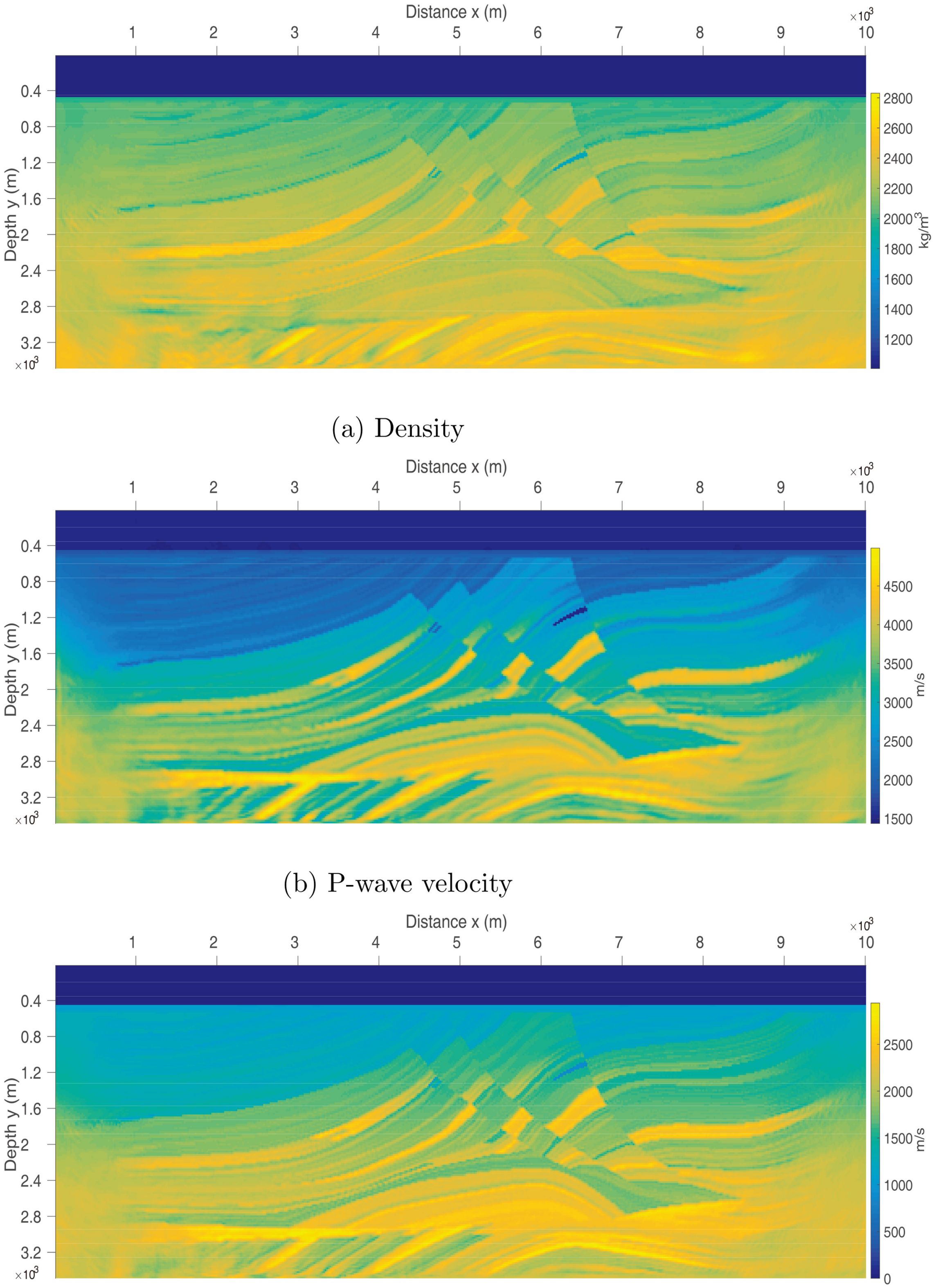


(a) Density

(b) P-wave velocity

(c) S-wave velocity

Figure 3. The finial inversion result of Marmousi model by the Gauss-Newton method with the line search strategy. (a) Density; (b) P-wave velocity; (c) S-wave velocity.

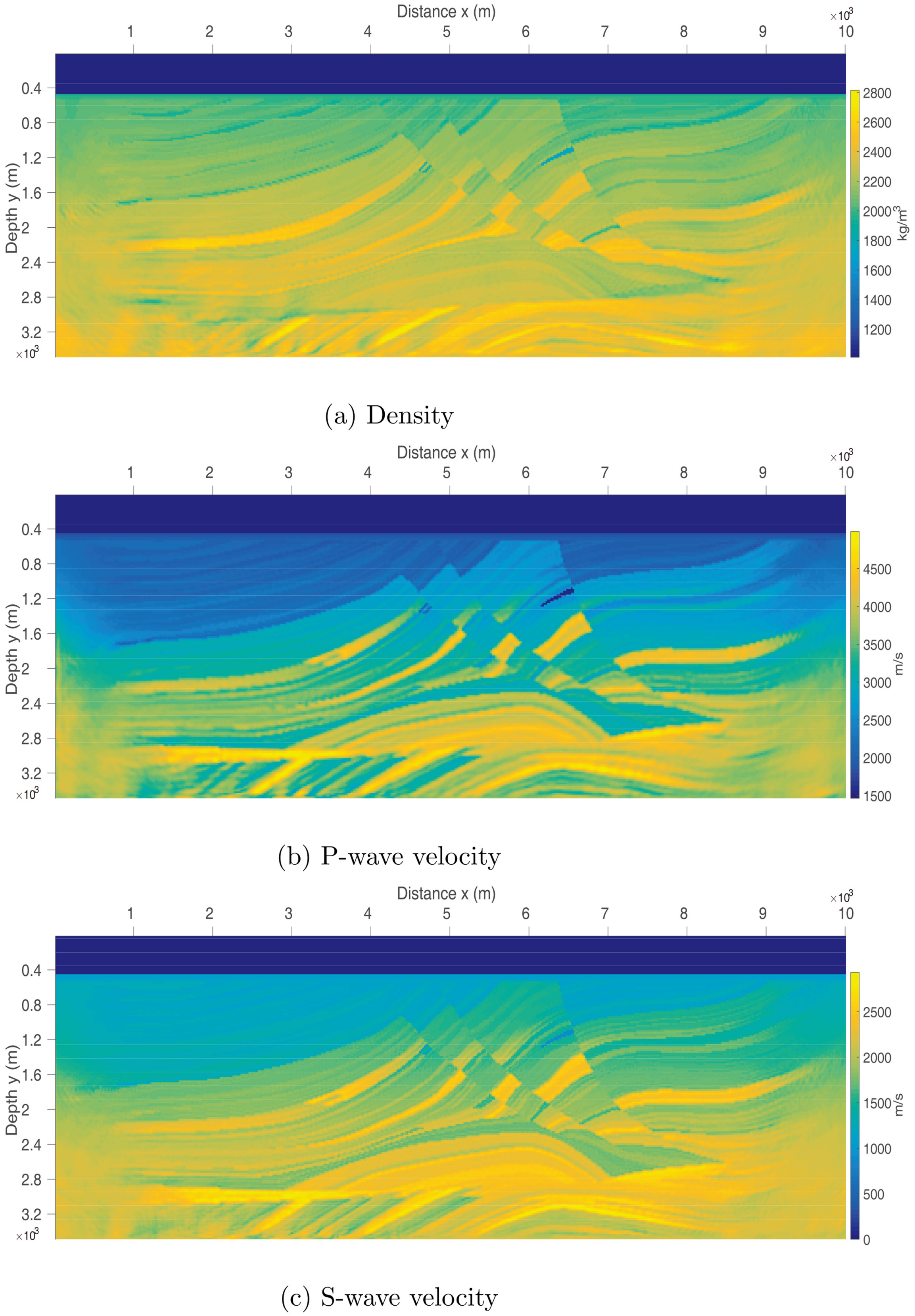


(a) Density

(b) P-wave velocity

(c) S-wave velocity

Figure 4. The finial inversion result of Marmousi model by the Gauss-Newton method with the two dimensional subspace strategy. (a) Density; (b) P-wave velocity; (c) S-wave velocity.

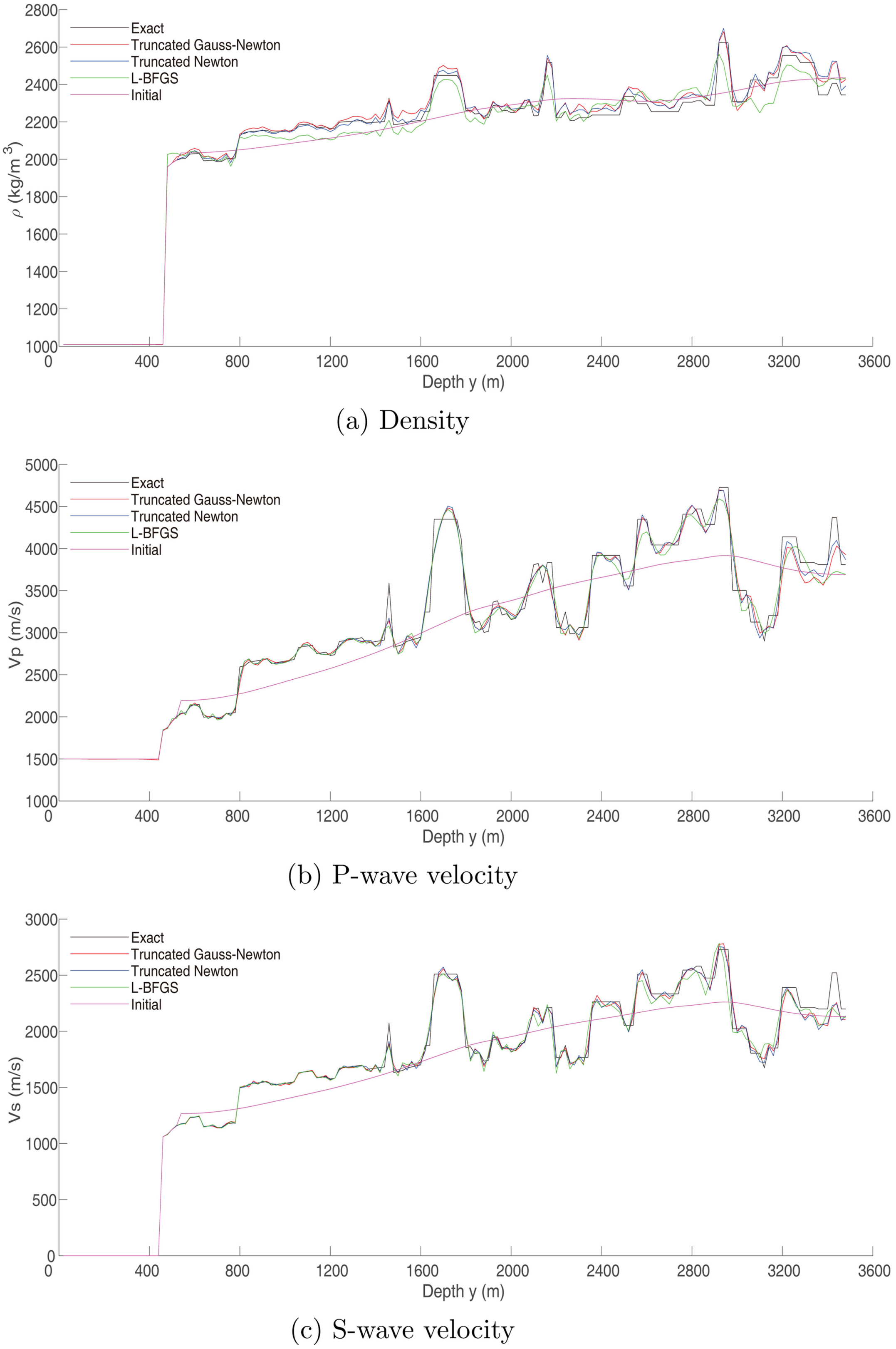


(a) Density

(b) P-wave velocity

(c) S-wave velocity

Figure 5. The inversion result of Marmousi model by the truncated Newton method with the line search strategy. (a) Density; (b) P-wave velocity; (c) S-wave velocity.



(a) Density

(b) P-wave velocity

(c) S-wave velocity

Figure 6. The comparison of inversion results at $x = 5\text{ km}$ inverted by the L-BFGS method, the Gauss-Newton method and the truncated Newton method with the line search strategy. (a) Density; (b) P-wave velocity; (c) S-wave velocity.

inversion results by using the L-BFGS method, the Gauss-Newton method and the truncated Newton method with the line search strategy. In Figure 6, we can see both the inversion results of P-wave velocity and S-wave velocity obtained by the three methods are almost the same. However, the inversion results of density by the Gauss-Newton method and the truncated Newton method are obviously better than the L-BFGS method especially in deep positions.

In Table 1, we present the iteration number at every stage (*i.e.*, 2 Hz, 5 Hz, 10 Hz and 20 Hz) for the L-BFGS method, the Gauss-Newton (GN) method and the truncated Newton (TN) method by the line search (LS) strategy or the two dimensional subspace (Sub) strategy. From Table 1, we can see that the Gauss-Newton method has obvious advantage over other methods from point of total iteration number.

In Table 2, we compare the computational efficiency for different methods. The total computational time and the time of each iteration are listed in Table 2. The total iteration number is also listed for convenience. As we can see, the Gauss-Newton method with the two dimensional subspace method costs least computational time for each iteration.

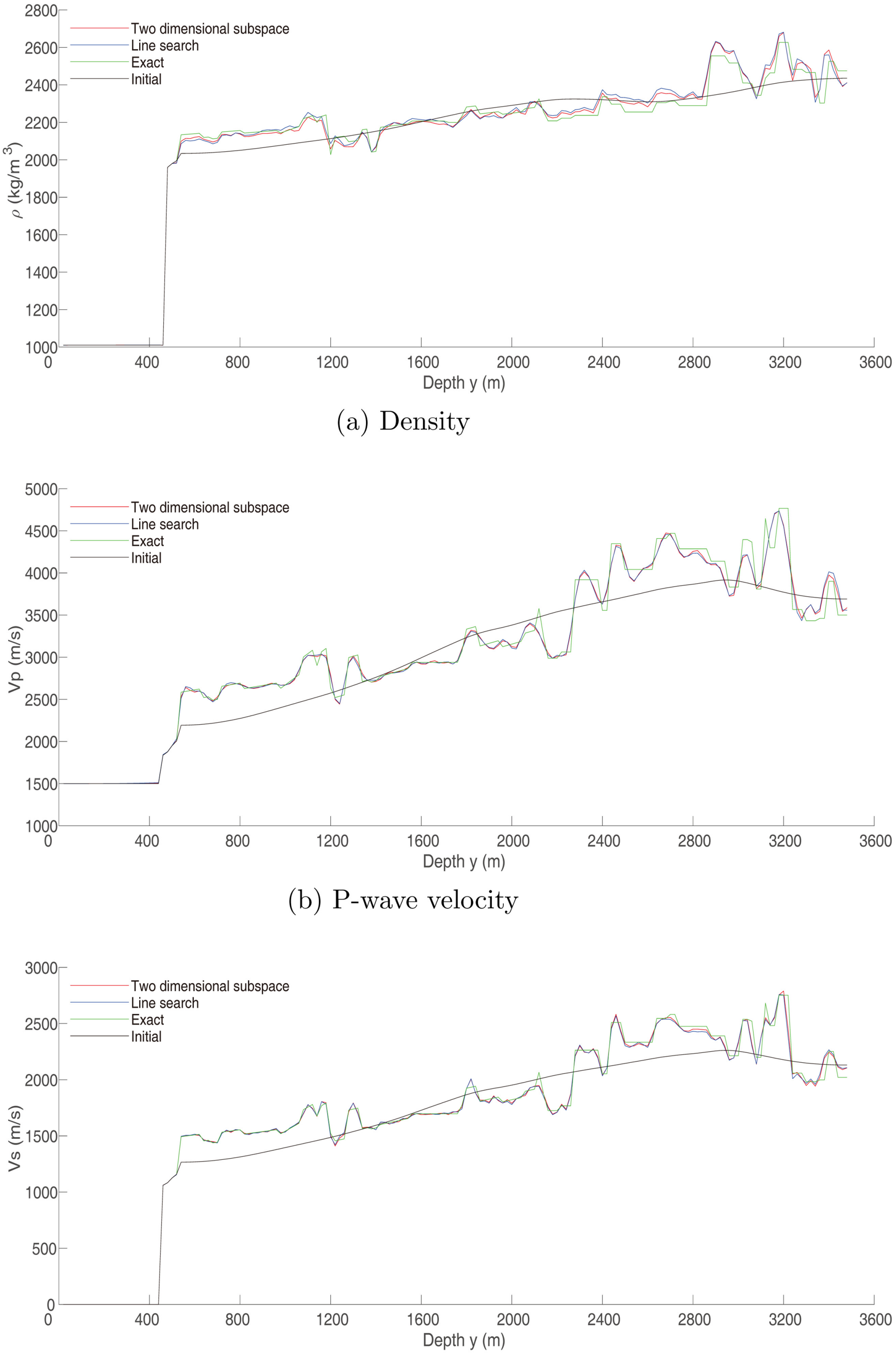
In Figure 7, we present a profile comparison of the inversion results at $x = 6\ \text{km}$ by the Gauss-Newton method with the line search strategy and the trust region strategy. Figures 7(a)-(c) are the inversion results of density, P-wave velocity and S-wave velocity respectively. We can see that both strategies almost have the same inversion accuracy for P-wave velocity and S-wave

Table 1. The number of iterations at each stage for the L-BFGS method, the Gauss-Newton (GN) method and the truncated Newton (TN) method with the line search (LS) strategy or the two dimensional subspace (Sub) strategy.

Stage	L-BFGS (LS)	GN (LS)	GN (Sub)	TN (LS)
One	85	55	51	51
Two	206	55	82	107
Three	150	39	65	69
Four	105	44	62	95
Total	546	193	260	322

Table 2. The total computation time and the time of each iteration for different inversion methods with the line search (LS) strategy or the two dimension subspace (Sub) strategy.

Method	time (s)	iteration	time/iteration (s)
L-BFGS (LS)	243,994	546	447
Gauss-Newton (LS)	53,227	193	276
Gauss-Newton (Sub)	70,055	260	269
Truncated Newton (LS)	95,399	322	296



(a) Density

(b) P-wave velocity

(c) S-wave velocity

Figure 7. The comparison of inversion results for the Marmousi model at $x = 6$ km inverted by using the Gauss-Newton method with the line search strategy and the trust region strategy. (a) Density; (b) P-wave velocity; (b) S-wave velocity.

velocity. For the density, we can see the trust region strategy with the two dimensional subspace strategy has better inversion accuracy than the line search strategy. In Table 3, we present l_2 errors and relative l_2 errors for the different methods. The relative error ε is defined by

$$\varepsilon = \frac{\left\|\boldsymbol{m}^{exa} - \boldsymbol{m}^{cal}\right\|_2}{\left\|\boldsymbol{m}^{exa}\right\|_2}, \tag{3.3}$$

where $\boldsymbol{m}^{exa}$ is the exact model and $\boldsymbol{m}^{cal}$ is the inversion result. The relative l_2 error for the truncated Newton method with the two dimensional subspace strategy is 7.28% and it is the least. This fact is consistent with the theory since the Newton method contains the full information of Hessian. For the Gauss-Newton method, we can see that the relative l_2 error by the two dimensional subspace strategy is 7.38% whereas it is 7.49% for the line search strategy. It shows that the two dimensional subspace strategy is helpful to improve the inversion accuracy. Figure 8 shows the absolute l_2 errors of inversion results versus iteration for the L-BFGS method, the Gauss-Newton method and the truncated Newton method.

Table 3. The l_2 errors and relative l_2 errors of final inversion results for different inversion methods with the line search (LS) strategy or the two dimension subspace (Sub) strategy.

Method	l_2 errors	relative l_2 errors
L-BFGS (LS)	1.050648×10^5	8.57%
GN (LS)	9.194219×10^4	7.49%
GN (Sub)	9.048718×10^4	7.38%
TN (LS)	8.929841×10^4	7.28%

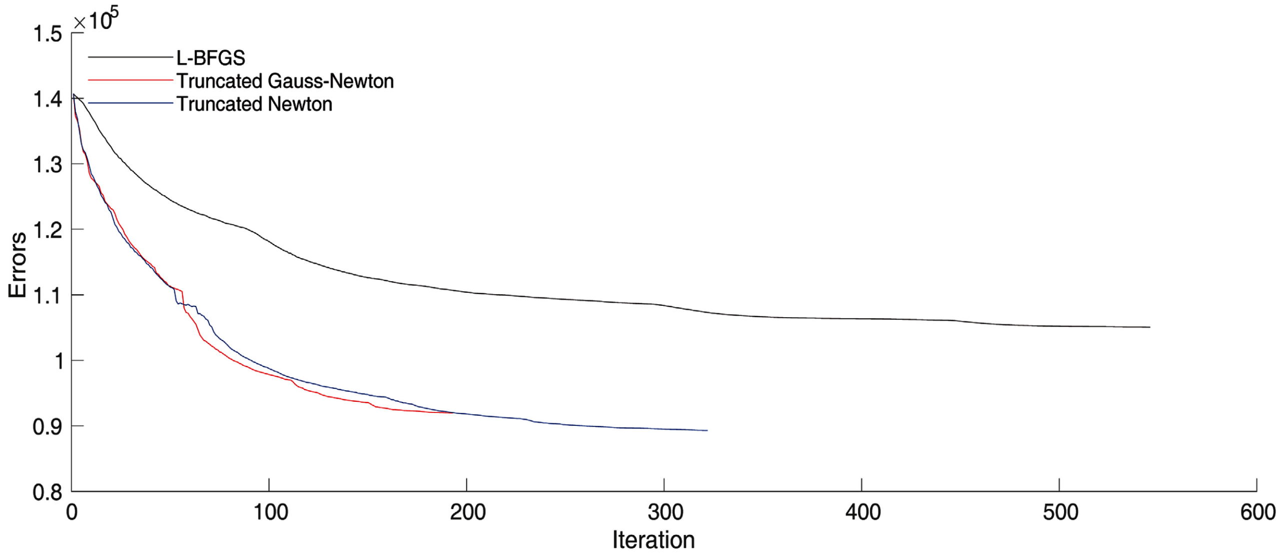


Figure 8. The absolute l_2 errors of inversion results versus iteration for the L-BFGS, the Gauss-Newton method and the truncated Newton method.

4. Conclusions

In FWI, the line search strategy is usually applied. In this paper, we have investigated the application of the trust-region strategy to elastic wave multi-parameter inversion. The theoretical methods and algorithms are described in detail. The trust-region subproblem is solved by the two-dimensional subspace method. Numerical computations for Marmousi model by the Gauss-Newton method and the truncated Newton method in the time domain are completed and compared.

We also have presented the corresponding inversion result by the L-BFGS method for comparison since it is the most popular quasi-Newton method. The comparisons demonstrate that both the Gauss-Newton method and the truncated Newton method are more accurate than the L-BFGS method. The trust region strategy is more efficient than the line search strategy. In the line search strategy, the information of Hessian and gradient is utilized for deciding the search direction whereas the search direction and the search step are determined together in the trust region strategy. In the trust region method, as long as the trust-region radius is well updated, the model updating can be performed for every iteration with the fixed trial step and there are no more extra computations of the forward problem. This is the theoretical reason why the trust-region strategy can save computational time. In the trust-region strategy, the two-dimensional subspace method requires that the Hessian matrix is positive and definite. Unfortunately, this condition is not always satisfied for the truncated Newton method. Future work will be the other measures to update the trust-region radius and the application of the trust-region method to other methods.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under grant number 11471328. It is also supported by the Key Project National Natural Science Foundation under grant number 51739007. The computations in this paper are completed on the parallel cluster of State Key Laboratory of Scientific and Engineering Computing.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Tarantola, A. (1984) Inversion of Seismic Reflection Data in the Acoustic Approximation. *Geophysics*, **49**, 1259-1266. https://doi.org/10.1190/1.1441754

[2] Lailly, P. (1983) The Seismic Inverse Problem as a Sequence of before Stack Migrations. In: Bednar, J.B., Robinson, E. and Weglein, A., Eds., *Conference on Inverse Scattering—Theory and Application*, SIAM, Philadelphia, 206-220.

[3] Woodhouse, J.H. and Dziewonski, A.M. (1984) Mapping the Upper Mantle: Three-

Dimensional Modeling of Earth Structure by Inversion of Seismic Waveforms. *Journal of Geophysical Research*: *Solid Earth*, **89**, 5953-5986. https://doi.org/10.1029/JB089iB07p05953

[4] Pratt, R.G. (1999) Seismic Waveform Inversion in the Frequency Domain, Part 1: Theory and Verification in a Physical Scale Model. *Geophysics*, **64**, 888-901. https://doi.org/10.1190/1.1444597

[5] Pratt, R.G. and Worthington, M.H. (1990) Inverse Theory Applied to Multi-Source Cross-Hole Tomography, Part 1: Acoustic Wave-Equation Method. *Geophysical Prospecting*, **38**, 287-310. https://doi.org/10.1111/j.1365-2478.1990.tb01846.x

[6] Sirgue, L. and Pratt, R.G. (2004) Efficient Waveform Inversion and Imaging: A Strategy for Selecting Temporal Frequencies. *Geophysics*, **69**, 231-248. https://doi.org/10.1190/1.1649391

[7] Bunks, C., Saleck, F.M., Zaleski, S. and Chavent, G. (1995) Multiscale Seismic Waveform Inversion. *Geophysics*, **60**, 1457-1473. https://doi.org/10.1190/1.1443880

[8] Zhang, W. and Luo, J. (2013) Full-Waveform Velocity Inversion Based on the Acoustic Wave Equation. *American Journal of Computational Mathematics*, **3**, 13-20. https://doi.org/10.4236/ajcm.2013.33B003

[9] Zhang, W. and Joardar, A. (2018) Acoustic Based Crosshole Full Waveform Slowness Inversion in the Time Domain. *Journal of Applied Mathematics and Physics*, **6**, 1086-1110. https://doi.org/10.4236/jamp.2018.65094

[10] Shin, C. and Cha, Y.H. (2008) Waveform Inversion in the Laplace Domain. *Geophysical Journal International*, **173**, 922-931. https://doi.org/10.1111/j.1365-246X.2008.03768.x

[11] Kim, Y., Shin, C., Calandra, H. and Min, D.J. (2013) An Algorithm for 3D Acoustic Time-Laplace-Fourier-Domain Hybrid Full Waveform Inversion. *Geophysics*, **78**, R151-R166. https://doi.org/10.1190/geo2012-0155.1

[12] Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L. and Virieux, J. (2013) A Guided Tour of Multiparameter Full-Waveform Inversion with Multicomponent Data: From Theory to Practice. *The Leading Edge*, **32**, 1040-1054. https://doi.org/10.1190/tle32091040.1

[13] Jeong, W., Lee, H.Y. and Min, D.J. (2012) Full Waveform Inversion Strategy for Density in the Frequency Domain. *Geophysical Journal International*, **188**, 1221-1242. https://doi.org/10.1111/j.1365-246X.2011.05314.x

[14] Xu, K. and McMechan, G.A. (2014) 2D Frequency-Domain Elastic Full-Waveform Inversion Using Time-Domain Modeling and a Multistep-Length Gradient Approach. *Geophysics*, **79**, R41-R53. https://doi.org/10.1190/geo2013-0134.1

[15] Pratt, R.G., Shin, C. and Hick, G. (1998) Gauss-Newton and Full Newton Methods in Frequency—Space Seismic Waveform Inversion. *Geophysical Journal International*, **133**, 341-362. https://doi.org/10.1046/j.1365-246X.1998.00498.x

[16] Yang, P., Brossier, R., M'etivier, L., Virieux, J. and Zhou, W. (2018) A Time-Domain Preconditioned Truncated Newton Approach to Visco-Acoustic Multiparameter Full Waveform Inversion. *SIAM Journal on Scientific Computing*, **40**, B1101-B1130. https://doi.org/10.1137/17M1126126

[17] 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. https://doi.org/10.1111/j.1365-246X.2006.02978.x

[18] Métivier, L., Bretaudeau, F., Brossier, R., Operto, S. and Virieux, J. (2014) Full Waveform Inversion and the Truncated Newton Method: Quantitative Imaging of Com-

plex Subsurface Structures. *Geophysical Prospecting*, **62**, 1353-1375. https://doi.org/10.1111/1365-2478.12136

[19] Aki, K. and Richards, P.G. (1980) Quantitative Seismology: Theory and Methods. Freeman, San Francisco.

[20] Zhang, W., Zhuang, Y. and Zhang, L. (2017) A New High-Order Finite Volume Method for 3D Elastic Wave Simulation on Unstructured Meshes. *Journal of Computational Physics*, **340**, 534-555. https://doi.org/10.1016/j.jcp.2017.03.050

[21] Virieux, J. (1986) P-SV-Wave Propagation in Heterogeneous Media: Velocitystress Finite-Difference Method. *Geophysics*, **51**, 889-901. https://doi.org/10.1190/1.1442147

[22] Clayton, R. and Engquist, B. (1977) Absorbing Boundary Conditions for Acoustic and Elastic Wave Equations. *Bulletin of the Seismological Society of America*, **67**, 1529-1540. https://doi.org/10.1785/BSSA0670061529

[23] Berenger, J.P. (1994) A Perfectly Matched Layer for the Absorption of Electromagnetic Waves. *Journal of Computational Physics*, **114**, 185-200. https://doi.org/10.1006/jcph.1994.1159

[24] Komatitsch, D. and Tromp, J. (2003) A Perfectly Matched Layer Absorbing Boundary Condition for the Second-Order Seismic Wave Equation. *Geophysical Journal International*, **154**, 146-153. https://doi.org/10.1046/j.1365-246X.2003.01950.x

[25] Nocedal, J. (1980) Updating Quasi-Newton Matrices with Limited Storage. *Mathematics of Computation*, **35**, 773-782. https://doi.org/10.1090/S0025-5718-1980-0572855-7

[26] Nocedal, J. and Wright, S.J. (2006) Numerical Optimization. Springer Science & Business Media, New York.

[27] Byrd, R.H., Lu, P., Nocedal, J. and Zhu, C. (1995) A Limited Memory Algorithm for Bound Constrained Optimization. *SIAM Journal on Scientific Computing*, **16**, 1190-1208. https://doi.org/10.1137/0916069

[28] Métivier, L., Brossier, R., Operto, S. and Virieux, J. (2017) Full Waveform Inversion and the Truncated Newton Method. *SIAM Review*, **59**, 153-195. https://doi.org/10.1137/16M1093239

[29] Albert, T. (2004) Inverse Problem Theory and Methods for Model Parameter Estimation. SIAM, Philadelphia.

[30] Philip, W. (1969) Convergence Conditions for Ascent Methods. *SIAM Review*, **11**, 226-235. https://doi.org/10.1137/1011036

[31] David, M.G. (1981) Computing Optimal Locally Constrained Steps. *SIAM Journal on Scientific and Statistical Computing*, **2**, 186-197. https://doi.org/10.1137/0902016

[32] Sorensen, D.C. (1982) Newton's Method with a Model Trust Region Modification. *SIAM Journal on Numerical Analysis*, **19**, 409-426. https://doi.org/10.1137/0719026

[33] Moré, J.J. and Sorensen, D.C. (1983) Computing a Trust Region Step. *SIAM Journal on Scientific and Statistical Computing*, **4**, 553-572. https://doi.org/10.1137/0904038

Appendix A. The Staggered-Grid Schemes for Solving (2.3)

Let Δt be the time step in the time direction and h the spatial step in x and y directions. The discrete indexes for x, y and t are denoted by j, i and n respectively. Then the difference schemes for v_x and v_y with the second order accuracy are

$$v_x^{n+1}\left(j,i+\frac{1}{2}\right)=v_x^n\left(j,i+\frac{1}{2}\right)+\Delta t\frac{1}{\rho\left(j,i+\frac{1}{2}\right)}\hat{v}_x^n\left(j,i+\frac{1}{2}\right), \tag{A.1}$$

$$v_y^{n+1}\left(j+\frac{1}{2},i\right)=v_y^n\left(j+\frac{1}{2},i\right)+\Delta t\frac{1}{\rho\left(j+\frac{1}{2},i\right)}\hat{v}_y^n\left(j+\frac{1}{2},i\right), \tag{A.2}$$

where

$$\hat{v}_x^n\left(j,i+\frac{1}{2}\right)=\frac{\sigma_{xx}(j,i+1)-\sigma_{xx}(j,i)}{h}+\frac{\sigma_{xy}\left(j+\frac{1}{2},i+\frac{1}{2}\right)-\sigma_{xy}\left(j-\frac{1}{2},i+\frac{1}{2}\right)}{h},$$

$$\hat{v}_y^n\left(j+\frac{1}{2},i\right)=\frac{\sigma_{xy}\left(j+\frac{1}{2},i+\frac{1}{2}\right)-\sigma_{xy}\left(j+\frac{1}{2},i-\frac{1}{2}\right)}{h}+\frac{\sigma_{yy}(j+1,i)-\sigma_{yy}(j,i)}{h}.$$

The difference schemes for strain σ_{xx}, σ_{xy} and σ_{yy} with the second order accuracy are the following

$$\begin{aligned}\sigma_{xy}^{n+1}\left(j+\frac{1}{2},i+\frac{1}{2}\right)&=\sigma_{xy}^n\left(j+\frac{1}{2},i+\frac{1}{2}\right)+\Delta t\mu\left(j+\frac{1}{2},i+\frac{1}{2}\right)\\&\times\left(\hat{v}_{xy}^n\left(j+\frac{1}{2},i+\frac{1}{2}\right)+\hat{v}_{yx}^n\left(j+\frac{1}{2},i+\frac{1}{2}\right)\right),\end{aligned} \tag{A.3}$$

$$\sigma_{xx}^{n+1}(j,i)=\sigma_{xx}^n(j,i)+\Delta t\lambda(j,i)\left(\hat{v}_{xx}^n(j,i)+\hat{v}_{yy}^n(j,i)\right)+2\Delta t\mu(j,i)\hat{v}_{xx}^n(j,i), \tag{A.4}$$

$$\sigma_{yy}^{n+1}(j,i)=\sigma_{yy}^n(j,i)+\Delta t\lambda(j,i)\left(v_{xx}^n(j,i)+\hat{v}_{yy}^n(j,i)\right)+2\Delta t\mu(j,i)\hat{v}_{yy}^n(j,i), \tag{A.5}$$

where

$$\hat{v}_{xx}^n(j,i)=\frac{v_x\left(j,i+\frac{1}{2}\right)-v_x\left(j,i-\frac{1}{2}\right)}{h},$$

$$\hat{v}_{yy}^n(j,i)=\frac{v_y\left(j+\frac{1}{2},i\right)-v_y\left(j-\frac{1}{2},i\right)}{h},$$

$$\hat{v}_{yx}^n\left(j+\frac{1}{2},i+\frac{1}{2}\right)=\frac{v_y\left(j+\frac{1}{2},i+1\right)-v_y\left(j+\frac{1}{2},i\right)}{h},$$

$$\hat{v}_{xy}^n\left(j+\frac{1}{2},i+\frac{1}{2}\right)=\frac{v_x\left(j+1,i+\frac{1}{2}\right)-v_x\left(j,i+\frac{1}{2}\right)}{h}.$$

Appendix B. The Derivation for Gradient Formulae (2.16)-(2.18)

Note that

$$v_x = \frac{\partial u_x}{\partial t},\ v_y = \frac{\partial u_y}{\partial t},\ \tilde{v}_x = \frac{\partial \tilde{u}_x}{\partial t},\ \tilde{v}_y = \frac{\partial \tilde{u}_y}{\partial t}. \tag{B.1}$$

Substituting (B.1) into (2.13), we have

$$\frac{\partial \chi}{\partial \rho} = -\sum_{x_s} \int_0^T \left(\frac{\partial v_x}{\partial t} \tilde{u}_x + \frac{\partial v_y}{\partial t} \tilde{u}_y \right) \mathrm{d}t, \tag{B.2}$$

which is just (2.16).

From the relation of strain and stress [19], we have

$$\sigma_{xx} = \lambda \left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} \right) + 2\mu \frac{\partial u_x}{\partial x}, \tag{B.3}$$

$$\sigma_{yy} = \lambda \left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} \right) + 2\mu \frac{\partial u_y}{\partial y}, \tag{B.4}$$

where u_x and u_y are the displacement in the x and y directions respectively. Adding (B.3) and (B.4) together, we have

$$\left(\sigma_{xx} + \sigma_{yy} \right) = 2(\lambda + \mu) \left(\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} \right), \tag{B.5}$$

that is

$$\frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} = \frac{\sigma_{xx} + \sigma_{yy}}{2(\lambda + \mu)}. \tag{B.6}$$

Similarly, for the backward wavefield $\tilde{u}_x$ and $\tilde{u}_y$, we have

$$\frac{\partial \tilde{u}_x}{\partial x} + \frac{\partial \tilde{u}_y}{\partial y} = \frac{\tilde{\sigma}_{xx} + \tilde{\sigma}_{yy}}{2(\lambda + \mu)}, \tag{B.7}$$

where $\tilde{u}_x$, $\tilde{u}_y$, $\tilde{\sigma}_{xx}$ and $\tilde{\sigma}_{yy}$ are the backward wavefield of the adjoint problem. Substituting (B.6) and (B.7) into (2.14), we get

$$\frac{\partial \chi}{\partial \lambda} = -\sum_{x_s} \int_0^T \frac{\left(\sigma_{xx} + \sigma_{yy} \right)\left(\tilde{\sigma}_{xx} + \tilde{\sigma}_{yy} \right)}{4(\lambda + \mu)^2} \mathrm{d}t, \tag{B.8}$$

which is just (2.17).

Since

$$\sigma_{xy} = \mu \left(\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} \right), \tag{B.9}$$

we have

$$\frac{\partial u_x}{\partial y} + \frac{\partial u_y}{\partial x} = \frac{\sigma_{xy}}{\mu}, \tag{B.10}$$

and

$$\frac{\partial \tilde{u}_x}{\partial y}+\frac{\partial \tilde{u}_y}{\partial x}=\frac{\tilde{\sigma}_{xy}}{\mu}, \tag{B.11}$$

where $\tilde{\sigma}_{xy}$ is the backward stress field corresponding to σ_{xy}. Multiplying (B.10) and (B.11), we obtain

$$\left(\frac{\partial u_y}{\partial x}+\frac{\partial u_x}{\partial y}\right)\left(\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial \tilde{u}_y}{\partial y}\right)=\frac{\sigma_{xy}\tilde{\sigma}_{xy}}{\mu^2}. \tag{B.12}$$

Subtracting (B.3) and (B.4), we have

$$\sigma_{xx}-\sigma_{yy}=2\mu\left(\frac{\partial u_x}{\partial x}-\frac{\partial u_y}{\partial y}\right), \tag{B.13}$$

that is

$$\frac{\partial u_x}{\partial x}-\frac{\partial u_y}{\partial y}=\frac{\sigma_{xx}-\sigma_{yy}}{2\mu}. \tag{B.14}$$

Similarly, we have

$$\frac{\partial \tilde{u}_x}{\partial x}-\frac{\partial \tilde{u}_y}{\partial y}=\frac{\tilde{\sigma}_{xx}-\tilde{\sigma}_{yy}}{2\mu}. \tag{B.15}$$

Multiplying (B.14) and (B.15), we have

$$\begin{aligned}\left(\frac{\partial u_x}{\partial x}-\frac{\partial u_y}{\partial y}\right)\left(\frac{\partial \tilde{u}_x}{\partial x}-\frac{\partial \tilde{u}_y}{\partial y}\right)&=\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_x}{\partial x}-\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_y}{\partial y}-\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_y}{\partial y}\\&=\frac{\left(\sigma_{xx}-\sigma_{yy}\right)\left(\tilde{\sigma}_{xx}-\tilde{\sigma}_{yy}\right)}{4\mu^2}.\end{aligned} \tag{B.16}$$

Multiplying (B.6) and (B.7), we have

$$\begin{aligned}\left(\frac{\partial u_x}{\partial x}+\frac{\partial u_y}{\partial y}\right)\left(\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial \tilde{u}_y}{\partial y}\right)&=\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_y}{\partial y}+\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_y}{\partial y}\\&=\frac{\left(\sigma_{xx}+\sigma_{yy}\right)\left(\tilde{\sigma}_{xx}+\tilde{\sigma}_{yy}\right)}{4\left(\lambda+\mu\right)^2}.\end{aligned} \tag{B.17}$$

Adding (B.16) and (B.17) together, we have

$$\begin{aligned}&\left(\frac{\partial u_x}{\partial x}-\frac{\partial u_y}{\partial y}\right)\left(\frac{\partial \tilde{u}_x}{\partial x}-\frac{\partial \tilde{u}_y}{\partial y}\right)+\left(\frac{\partial u_x}{\partial x}+\frac{\partial u_y}{\partial y}\right)\left(\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial \tilde{u}_y}{\partial y}\right)\\&=\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_x}{\partial x}-\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_y}{\partial y}-\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_y}{\partial y}\\&\quad+\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_y}{\partial y}+\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_y}{\partial y}\\&=2\left(\frac{\partial u_x}{\partial x}\frac{\partial \tilde{u}_x}{\partial x}+\frac{\partial u_y}{\partial y}\frac{\partial \tilde{u}_y}{\partial y}\right)\\&=\frac{\left(\sigma_{xx}+\sigma_{yy}\right)\left(\tilde{\sigma}_{xx}+\tilde{\sigma}_{yy}\right)}{4\left(\lambda+\mu\right)^2}+\frac{\left(\sigma_{xx}-\sigma_{yy}\right)\left(\tilde{\sigma}_{xx}-\tilde{\sigma}_{yy}\right)}{4\mu^2}.\end{aligned} \tag{B.18}$$

Substituting (B.12) and (B.18) into (2.15), we have

$$\frac{\partial \chi}{\partial \mu} = -\sum_{x_s} \int_0^T \left\{ \frac{\sigma_{xy}\tilde{\sigma}_{xy}}{\mu^2} + \frac{1}{4}\left[\frac{\left(\sigma_{xx}+\sigma_{yy}\right)\left(\tilde{\sigma}_{xx}+\tilde{\sigma}_{yy}\right)}{\left(\lambda+\mu\right)^2} + \frac{\left(\sigma_{xx}-\sigma_{yy}\right)\left(\tilde{\sigma}_{xx}-\tilde{\sigma}_{yy}\right)}{\mu^2} \right] \right\} \mathrm{d}t, \tag{B.19}$$

which is just (2.18).