Geophys. J. Int. (2006) **167,** 1373–1384 doi: 10.1111/j.1365-246X.2006.03162.x

Time domain Gauss–Newton seismic waveform inversion in elastic media

Dong-Hoon Sheen,[1,2] Kagan Tuncay,[1] Chang-Eob Baag[2] and Peter J. Ortoleva[1]

[1] *Laboratory for Computational Geodynamics, Indiana University, Bloomington, Indiana, 47405, USA*
[2] *School of Earth and Environmental Sciences, Seoul National University, San 56–1, Shillim-dong, Gwanak-gu, Seoul, 151–742 Korea.*
E-mail: sheendh@snu.ac.kr

Accepted 2006 August 3. Received 2006 August 3; in original form 2005 March 30

SUMMARY

We present a seismic waveform inversion methodology based on the Gauss–Newton method from pre-stack seismic data. The inversion employs a staggered-grid finite difference solution of the 2-D elastic wave equation in the time domain, allowing accurate simulation of all possible waves in elastic media. The partial derivatives for the Gauss–Newton method are obtained from the differential equation of the wave equation in terms of model parameters. The resulting wave equation and virtual sources from the reciprocity principle allow us to apply the Gauss–Newton method to seismic waveform inversion. The partial derivative wavefields are explicitly computed by convolution of forward wavefields propagated from each source with reciprocal wavefields from each receiver. The Gauss–Newton method for seismic waveform inversion was proposed in the 1980s but has rarely been studied. Extensive computational and memory requirements have been principal difficulties which are addressed in this work. We used different sizes of grids for the inversion, temporal windowing, approximation of virtual sources, and parallelizing computations. With numerical experiments, we show that the Gauss–Newton method has significantly higher resolving power and convergence rate over the gradient method, and demonstrate potential applications to real seismic data.

Key words: finite difference methods, Frechét derivatives, synthetic seismograms, wave equation, waveform inversion.

1 INTRODUCTION

Seismic waveform inversion can be defined as an iterative procedure for obtaining physical properties of the Earth from pre-stack seismic data. It is well known that the inversion of seismic data is a computationally demanding task. Early studies include those of Lailly (1983) and Tarantola (1984) who presented a method of constructing the gradient (or steepest descent) direction for the inversion of the acoustic problem without computing the partial derivatives explicitly. This method constructs the gradient direction by cross-correlating forward propagated wavefields from a seismic source with backward propagated wavefields from the data residuals. As each iterative loop of the inversion requires only several forward simulations for each seismic source, it made seismic waveform inversion feasible in the 1980s. Mora (1987) applied this method to elastic problems in the time domain, whereas Pratt (1990) and Pratt *et al.* (1998) applied it to elastic and acoustic problems, respectively, in the frequency domain.

Tarantola (1984) also presented the Gauss–Newton algorithm called 'total inversion' (Tarantola & Valette 1982), although it was not possible to implement when it was presented because of limited computational resources. In recent years, however, it has become feasible. Pratt *et al.* (1998) used 'virtual source' terms to obtain partial derivative seismic wavefields which had been used for electromagnetic problem (Rodi 1976), and solved seismic waveform inversion with the Newton method. The partial derivative wavefields are obtained from new wave propagation simulations driven by the virtual sources at the grid points where model parameters are determined. In other words, the number of required forward simulations is equal to the number of model parameters.

Shin *et al.* (2001) used an efficient method for calculating partial derivative wavefields using the reciprocity relation between the virtual sources and the receivers. The reciprocity theorem is proven in Aki & Richards (1980) for an elastic anisotropic continuous medium. This theorem allows the source–receiver locations to be interchanged. The recorded seismograms are identical if the sources and receivers are located inside the domain or on its boundary (Eisner & Clayton 2001). Thus, the computation of the partial derivative wavefields doesn't depend on the number of model parameters but depends on the numbers of shots and receivers.

Although developments in computer technology have been impressive, it has been impractical to make use of a Newton type method for high-resolution seismic inversions. To avoid extremely expensive computation of the Jacobian or the Hessian matrix, Hicks & Pratt (2001) proposed a two-step inversion procedure. The backward propagation method is used for finding reflectors consisting

of a large number of parameters, and then the Newton method is applied to background velocities with a much smaller number of parameters. Shin *et al.* (2001) took advantage of the diagonally dominant nature of the 'approximate' Hessian matrix (Pratt *et al.* 1998). Diagonal elements of the approximate Hessian were used as a pre-conditioner for an iterative inversion. In all of these studies, the finite element method was used to solve the forward acoustic problems.

Pratt *et al.* (1998) and Shin *et al.* (2001) used a frequency domain method to carry out the inversion with discrete frequencies. By selecting a few frequencies, the method allows a significant reduction in computational burden. However, success of the method requires a careful selection of frequencies, which highlights the robustness of the time domain approach over the frequency domain approach (Freudenreich & Shipp 2000). Temporal windowing in the time domain approach, like discrete frequencies in the frequency domain approach, is proven to have an important role in reducing computation (Shipp & Singh 2002).

In this work, we apply the Gauss–Newton method to elastic waveform inversion in the time domain. The Gauss–Newton method is known to show local quadratic convergence. Local minimization algorithm, such as a gradient or Newton method, requires us to set a initial model close to a true model. Such inversions, therefore, may converge to a local minimum (Gauthier *et al.* 1986; Mora 1987). Several approaches have been proposed to overcome this difficulty. Traveltime analysis precedes and gives a initial background velocity model to the inversion (Shipp & Singh 2002). Decoupling the high and low wavenumber of velocity variation has also been used to make the solutions to converge to the correct model (Snieder *et al.* 1989; Symes & Carazzone 1991; Clément *et al.* 2001; Hicks & Pratt 2001). They used different types of the misfit function to avoid falling into local minima: the simplex method, the differential semblance optimization, the migration based traveltime, and the adaptive depth stretching. Following Shipp & Singh (2002) and Sirgue & Pratt (2004), we assume that the initial model is close to the global minimum, which can be obtained from the traveltime analysis.

Because seismic waveform inversion is the systematic fitting of synthetic to observed seismograms, it is important to generate accurate seismogram which can account for subsurface elastic features. Thus, we use the velocity-stress staggered-grid finite difference method (Levander 1988) to solve the elastic wave equation in the time domain and implement the perfectly matched layer (PML) method (Berenger 1994) as an absorbing boundary condition. The solution of the elastic wave equation allows us to use multicomponent data. The major obstacle to seismic waveform inversion is the explicit calculation of the Jacobian and the approximate Hessian matrices. To overcome this, we utilize the reciprocity principle and the convolution theorem. A Newton type seismic waveform inversion still requires huge amounts of memory and computation. For elastic problems, much more resources are required than those for acoustics. We surmount the limitation by using different sizes of grids in the spatial and time domains for the inversion stage, temporal windowing, approximation of virtual sources, and parallelizing the method for massively parallel computers using the message passing interface (MPI) approach. Integration of all these features make it possible to implement the classical seismic waveform inversion scheme presented in Tarantola (1984).

This paper is organized as follows: we recap the inverse problem and show how to use the reciprocity principle and the convolution theorem for calculating partial derivatives explicitly. We then present the more detailed inversion scheme used in this work and illustrate how the Gauss–Newton method outperforms the gradient method. We conclude by showing several numerical examples illustrating the application of seismic waveform inversion.

2 INVERSE PROBLEM

Seismic waveform inversion is the problem of finding properties of the Earth from seismic data. In order to infer a set of model parameters which represent the Earth, the inverse problem seeks to minimize the residuals between the model response obtained by forward simulation and the observed seismic data.

In general, the seismic responses $\mathbf{d}$ of the Earth represented by model parameters $\mathbf{m}$ would be recorded at receivers. This relationship can be expressed with the non-linear functional $\mathbf{F}$:

$$\mathbf{d} = \mathbf{F}(\mathbf{m}). \tag{1}$$

The residual error is defined as the difference between the model responses and the observed data:

$$\Delta\mathbf{d} = \mathbf{F}(\mathbf{m}) - \mathbf{d}^{\text{obs}^4}, \tag{2}$$

where $\mathbf{d}^{\text{obs}}$ is the observed data. We now introduce the least-squares problem:

$$S_d(\mathbf{m}) = \frac{1}{2}\|\Delta\mathbf{d}\|^2 = \frac{1}{2}\Delta\mathbf{d}^t\Delta\mathbf{d}, \tag{3}$$

where S_d means the data misfit function, the factor 1/2 allows subsequent simplifications, and the superscript t represents the matrix transpose. The inverse problem becomes the minimization of S_d. Thus, our purpose is to find a model $\mathbf{m}^*$ which minimize $S_d(\mathbf{m}^*)$.

2.1 Gradient method

The gradient method minimizes S_d by updating model parameters in the opposite direction of the gradient of $S_d(\mathbf{m})$ iteratively:

$$\mathbf{m}^{n+1} = \mathbf{m}^n - \alpha\nabla S_d, \tag{4}$$

where the superscripts represent iteration numbers and α is a step length. The gradient direction can be obtained by taking partial derivatives of eq. (3) with respect to model parameters $\mathbf{m}$:

$$\nabla S_d = \left[\frac{\partial \mathbf{F}_i(\mathbf{m})}{\partial m_j}\right]^t \Delta\mathbf{d} = \mathbf{J}^t\Delta\mathbf{d}, \qquad (i = 1, \ldots, NS \times NR;\ j = 1, \ldots, M), \tag{5}$$

where $\mathbf{J}^t$ is a transpose of the Jacobian matrix, the subscripts i and j indicate indices for seismograms and model parameter, respectively, and NS, NR and M are the numbers of shots, receivers and model parameters, respectively.

It is not necessary to compute the Jacobian explicitly so as to find the gradient direction. The backward propagation method requires only several forward computations to construct the gradient direction. Formal derivations for the backward propagation method are given by Tarantola (1984) and Mora (1987).

2.2 Newton method & Gauss–Newton method

It is well recognized that the Newton and Gauss–Newton methods are effective and robust techniques for numerical optimization of non-linear problems and guarantee faster convergence rates than the gradient method when the initial model is close enough to a local or the global minimum. In geophysical inverse problems, however, the Newton method has not been often used because it is difficult to

compute the Hessian matrix and the cost usually surpasses the gain. The Hessian matrix is the analogue of the second-order derivative. Pratt *et al.* (1998) calculated the second-order derivative using the second order virtual source which requires as many additional forward simulations as the number of model parameters. In this work, we use the Gauss–Newton method which neglects the second-order derivative.

The Newton method (Tarantola 1987) is given by

$$\mathbf{m}^{n+1} = \mathbf{m}^n - \mathbf{H}^{-1}\mathbf{J}^t \Delta\mathbf{d}, \tag{6}$$

where $\mathbf{H}$ is the Hessian matrix. Each element of the Hessian matrix can be expressed in differential forms:

$$\mathbf{H} = \nabla^2 S_d(\mathbf{m}) = \frac{\partial}{\partial m_p}\left[\mathbf{J}^t \Delta\mathbf{d}\right] = \mathbf{J}^t\mathbf{J} + \frac{\partial \mathbf{J}^t}{\partial m_p}\Delta\mathbf{d}. \tag{7}$$

Because the second term of eq. (7) is usually small and negligible (Tarantola 1987), we obtain the Gauss–Newton formula

$$\mathbf{m}^{n+1} = \mathbf{m}^n - [\mathbf{J}^t\mathbf{J}]^{-1}\mathbf{J}^t\Delta\mathbf{d} = \mathbf{m}^n - [\mathbf{H}_a]^{-1}\mathbf{J}^t\Delta\mathbf{d}, \tag{8}$$

where $\mathbf{H}_a$ is the approximate Hessian matrix. The elements of the approximate Hessian are obtained from the zero-lag crosscorrelation of the partial derivative wavefields and, specifically, the diagonal elements of the matrix are obtained from the zero-lag auto-correlation of the derivatives. The partial derivative wavefields are mostly uncorrelated with each other in the high-frequency limit. However, the wavefields from adjacent nodes are more or less correlated because the frequency content is finite. Thus, in general, the approximate Hessian matrix will be diagonally dominant due to auto-correlation, and banded due to finite frequencies. It is also known that the inverse approximate Hessian matrix operates like a sharpening or focusing filter (Pratt *et al.* 1998).

In the application of the Gauss–Newton method to geophysical inversion, regularization is particularly useful for stabilizing the system and incorporating *a priori* information to the problem (Tarantola 1987). The regularized misfit function S can be defined as,

$$S(\mathbf{m}) = S_d(\mathbf{m}) + \lambda S_m(\mathbf{m}), \tag{9}$$

where S_m is the model objective function that contains *a priori* information of the model and λ is a scalar value that globally controls the relative importance of the model objective function S_m. The model objective function can be written as combinations of discrete linear operator $\mathbf{L}$:

$$S_m(\mathbf{m}) = \frac{1}{2}\|\mathbf{L}\Delta\mathbf{m}\|^2. \tag{10}$$

Then the regularized Gauss–Newton formula can be written as,

$$\mathbf{m}^{n+1} = \mathbf{m}^n - \alpha^n\left[\mathbf{J}^t\mathbf{J} + \lambda\mathbf{L}^t\mathbf{L}\right]^{-1}\mathbf{J}^t\Delta\mathbf{d}, \tag{11}$$

where α is a step length. We use the conjugate gradient method to invert the regularized approximate Hessian to calculate the model updates $\mathbf{m}^{n+1}$. If $\mathbf{L} = \mathbf{I}$ (the identity matrix), eq. (11) yields the damped least-squares method (Levenberg 1944; Marquardt 1963). If $\mathbf{L}$ is a discrete spatial differential operator, the model objective function controls the roughness of spatial variations among the model parameters. Sasaki (1989) used discrete 2-D Laplacian operator:

$$P_i\Delta\mathbf{m} = (\Delta m_i)^E + (\Delta m_i)^W + (\Delta m_i)^N + (\Delta m_i)^S - 4(\Delta m_i), \tag{12}$$

where the superscripts E, W, N, and S refer to the four neighbours of the ith model parameter and P_i is the ith row of the Laplacian operator matrix whose elements are either 1, −4, or 0. In this work, these model objective functions are used simultaneously:

$$\mathbf{m}^{n+1} = \mathbf{m}^n - \alpha^n\left[\mathbf{J}^t\mathbf{J} + \lambda_1\mathbf{P}^t\mathbf{P} + \lambda_2\mathbf{I}^t\mathbf{I}\right]^{-1}\mathbf{J}^t\Delta\mathbf{d}. \tag{13}$$

A choice of λ value between 0 and infinity produces a compromise result. After several computations, appropriate values of λ are chosen as $\lambda_1 = 0.05$ and $\lambda_2 = 0.0005$. This means that 5 and 0.05 per cent of the maximum value of diagonal elements of $\mathbf{J}^t\mathbf{J}$ are added (Shin *et al.* 2001), respectively.

The simplest way of choosing the step length is to take it as a constant through all the iterations, which can be obtained by trial and error. In this work, an optimal value of the step length is determined by a linearized approach. The optimal value for α^n is given by

$$\alpha^n \cong \frac{[\mathbf{J}^t\mathbf{g}^n]^t\left[\mathbf{F}(\mathbf{m}^n) - \mathbf{d}^{\text{obs}}\right]}{[\mathbf{J}^t\mathbf{g}^n]^t[\mathbf{J}^t\mathbf{g}^n]}, \tag{14}$$

where

$$\mathbf{g}^n = \left[\mathbf{J}^t\mathbf{J} + \lambda\mathbf{L}^t\mathbf{L}\right]^{-1}\mathbf{J}^t\left[\mathbf{d}^n - \mathbf{d}^{\text{obs}}\right]. \tag{15}$$

For more details, see Gauthier *et al.* (1986) and Pica *et al.* (1990).

3 PARTIAL DERIVATIVES

Partial derivatives of seismic waveform inversion represented by the Jacobian matrix $\mathbf{J}$ can be directly obtained from a residual wavefield from two forward simulations with and without perturbations of model parameters. For magnetotellurics data, Rodi (1976) showed that partial derivatives can be obtained from solving forward problems with virtual sources at the perturbation locations. Pratt *et al.* (1998) applied this to acoustic waveform inversion. This approach is appropriate for the case in which there are many stations and a few model parameters. Rodi also proposed a more efficient method when the number of model parameters exceeds the number of the data using the reciprocity principle. Shin *et al.* (2001) adopted this approach for acoustic waveform inversion. In all previous studies, the forward problems were solved in the frequency domain. In this paper, we formulate the wave propagation simulation in the time domain which is based on the finite difference method (Levander 1988), and use the PML boundary condition (Berenger 1994) as an absorbing boundary condition. The convolution theorem is used to utilize the virtual source and the reciprocity principle.

3.1 Partial derivative wavefields from virtual sources

In 2-D Cartesian coordinates, for an isotropic, linearly elastic medium, the equations of motion can be written as a set of first-order hyperbolic equations:

$$\begin{aligned} \rho\,\dot{v}_i &= \tau_{ij,j} + F_i, \\ \dot{\tau}_{ij} &= \mu\,(v_{i,j} + v_{j,i}) + \lambda\,\delta_{ij}\,v_{k,k} + G_{ij}, \end{aligned} \tag{16}$$

where the symbols are given in Table 1.

In this formulation, the model parameters will be density, ρ, or one of Lamé's moduli, λ and μ. In the case of density as the model

Table 1. Definition of symbols.

Symbol	Definition
v	Velocity
τ	Stress tensor
ρ	Density
λ, μ	Lamé's moduli
δ_{ij}	Kronecker delta, $\delta_{ij} = 0$ for $i \neq j$ and $\delta_{ij} = 1$ for $i = j$
F_i, G_{ij}	Body force and traction source
$_{,k}$	Spatial derivative, $\partial/\partial x_k$
$\dot{}$	Temporal derivative, $\partial/\partial t$

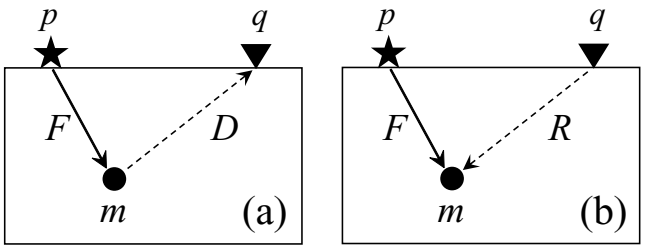

Figure 1. Schematic diagram for general reciprocal relation for partial derivative wavefields. Partial derivative wavefields from a virtual source (a) and a reciprocal source (b). Symbols p, q and m denote the indices for a source, receiver and model parameter, respectively. F, D and R represent the forward wavefield, differential wavefield from the virtual source and reciprocal wavefield from the receiver, respectively.

parameter, eq. (16) is differentiated with respect to a ρ_p:

$$\rho \frac{\partial \dot{v}_i}{\partial \rho_p} = \frac{\partial \tau_{ij,j}}{\partial \rho_p} - \frac{\partial \rho}{\partial \rho_p} \dot{v}_i,$$
$$\frac{\partial \dot{\tau}_{ij}}{\partial \rho_p} = \mu \left(\frac{\partial v_{i,j}}{\partial \rho_p} + \frac{\partial v_{j,i}}{\partial \rho_p} \right) + \lambda \, \delta_{ij} \frac{\partial v_{k,k}}{\partial \rho_p}, \quad (17)$$

where $-(\partial \rho / \partial \rho_p)\dot{v}_i$ is a virtual source for the partial derivative wavefields $\partial v_i / \partial \rho_p$ propagated from the location of a particular model parameter, ρ_p. The virtual source in eq. (17) is obtained from solutions of eq. (16). Partial derivative wavefields with respect to Lamé's moduli, are given in Appendix.

Simple Earth model can be represented by three parameters, such as density and Lamé's moduli or density, and P and S-wave velocities. Tarantola (1986) showed that the choice of model parameters for elastic inversion of seismic data affects the performance of the inversion. From numerical experiments, he chose density and P- and S-wave impedances as adequate parameters for seismic waveform inversion. In this paper, we choose density and P and S-wave velocities as the parameters. Expressions for partial derivatives in terms of density and velocities, and density and impedances are also given in Appendix.

3.2 Seismic reciprocity principle & convolution

Fig. 1 shows how the partial derivative wavefields are generated. The virtual source m samples the wavefield driven by the source p. Then, the partial derivative wavefield propagates from the virtual source to the receiver q. Using the virtual source, thus, explicit computation of the partial derivative wavefields requires as many as the number of model parameters of the forward simulation.

Seismic waveform inversion is usually an underdetermined problem which means the number of parameters exceeds the data. A more efficient way of evaluating partial derivatives is to use the reciprocity principle. The reciprocity principle (Aki & Richards 1980) allows the source–receiver locations to be interchanged. Thus, we can get the same partial derivative wavefield at the virtual source m which is propagated from the receiver q with the forward wavefield F recorded at the virtual source m. This can be calculated by convolution of two wavefields, F and R, at the virtual source m. Then, the number of forward simulations for the derivatives doesn't depend on the number of parameters but depends on the number of shots and receivers.

In the frequency domain, the partial derivative wavefield can be obtained by convolution of the forward wavefield F with the reciprocal wavefield R which is the impulse response of the delta function. This approach, however, is based on the time domain finite difference method and can only use limited frequency band of a source function. In a sense, thus, it is a bandpass filtered partial derivative wavefield that we can obtain from this approach. Because the seismic data for digital processing and numerical analysis are discrete and they can't carry information above a threshold frequency, for example, Nyquist frequency, our limited frequency band of partial derivative wavefields still guarantee the resolution of the problem.

We can decompose the Jacobian matrix **J** into virtual sources and reciprocal wavefields:

$$J^m_{p,q} = F^m_p * D^m_q = F^m_p * R^m_q, \quad (18)$$

where $*$ means convolution.

It should be noted that the reciprocal wavefields are identical only if the orientations of the source and the receiver match (Arntsen & Carcione 2000; Sheen *et al.* 2004). Fig. 2 shows how to generate the partial derivative wavefields with the reciprocity principle. Configuration (i) can be decomposed into (ii) and (iii). By the reciprocity principle, those configurations can be replicated by (iv) and (v). The relations between configurations (iii) and (v) of Fig. 2(a) and between (ii) and (iv) of Fig. 2(b) show that the locations of source and receiver should be reciprocal as well as the orientations. Thus, for example, the reciprocal simulation of the horizontal motion from the source can be generated by summing horizontal and vertical responses from a horizontal source (see Fig. 2a). This suggests that eq. (18) becomes

$$J^m_{p,q} = \left(J^m_{p,q}\big|_x, J^m_{p,q}\big|_z \right), \quad (19)$$

where

$$J^m_{p,q}\big|_x = F^m_p\big|_x * R^m_{q\_x}\big|_x + F^m_p\big|_z * R^m_{q\_x}\big|_z,$$

and

$$J^m_{p,q}\big|_z = F^m_p\big|_x * R^m_{q\_z}\big|_x + F^m_p\big|_z * R^m_{q\_z}\big|_z,$$

where $|_x$ and $|_z$ represent the orientation of recorded wavefields and $\_x$ and $\_z$ mean the orientation of the source component.

3.3 Jacobian & approximate Hessian matrix

The gradient direction in eq. (5) can be obtained by

$$\mathbf{J}^t \Delta \mathbf{d} = J^m_{p,q} \otimes \Delta d_{p,q}, \quad (20)$$

where $\otimes$ means zero-lag crosscorrelation and $\Delta d_{p,q}$ represents the residual at the receiver q of the shot gather with the source p. Before crosscorrelating the Jacobian and the residual, the residual should be convolved with the source function used for the reciprocal wavefields R. The procedure to compute the Jacobian matrix, the gradient direction and the approximate Hessian matrices for multi shots is summarized below:

(i) Simulate all forward wave propagations from the sources and sample forward wavefields at the model parameter m. Compute the residuals;

(ii) Simulate reciprocal wave propagations from the horizontal and vertical component of the receiver q, respectively. Sample reciprocal wavefields at the model parameter m;

(iii) Convolve the source function of step (ii) and the residuals from step (i);

(iv) Construct the Jacobian matrix consisting of partial derivative wavefields obtained from convolving wavefields from steps (i) and (ii);

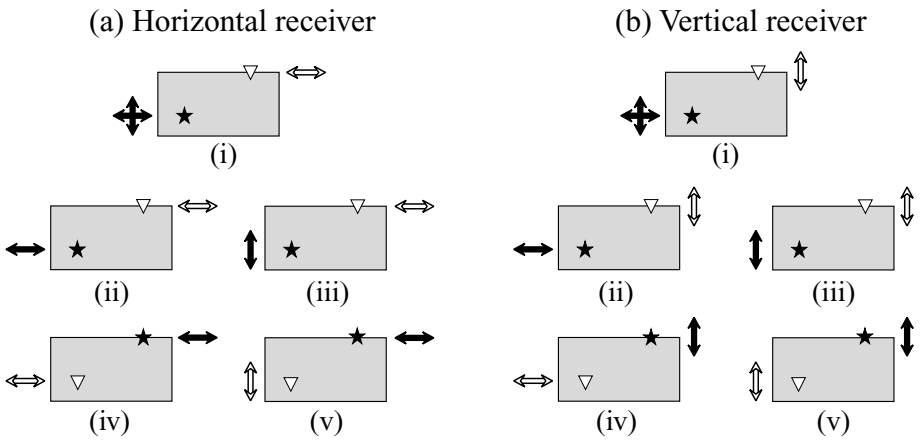

Figure 2. Reciprocal reconstructions of horizontal response (a) and vertical response (b). Star and triangle denote seismic source and receiver, respectively. Black arrows represent source orientations and white arrows receiver orientations, respectively.

(v) Calculate the gradient direction by zero-lag crosscorrelation of the wavefields from step (iv) and (iii).

(vi) Calculate the approximate Hessian matrix by zero-lag crosscorrelation of the Jacobian matrix from step (iv).

The explicit calculation of the Jacobian requires $NS + 2NR$ simulations.

The drawback of the Gauss–Newton method, exclusive of the computational demands, is a huge memory requirement to store the Jacobian and the approximate Hessian matrices. Furthermore, the procedure for constructing the Jacobian already costs huge amounts of the memory [see step (i), above]. However, it turns out that the Jacobian matrix does not have to be computed and stored as a whole. Because we can divide the Jacobian matrix into submatrices, it is possible to compute the approximate Hessian matrix as follows:

$$\mathbf{H}_a = \mathbf{J}^t\mathbf{J} = \left[\begin{array}{ccc} \mathbf{J}_1^t & \cdots & \mathbf{J}_{NR}^t \end{array}\right] \left[\begin{array}{c} \mathbf{J}_1 \\ \vdots \\ \mathbf{J}_{NR} \end{array}\right] = \sum_p^{NR} \mathbf{J}_p^t \mathbf{J}_p. \tag{21}$$

Therefore, the full Jacobian matrix does not have to exist at one time. The approximate Hessian matrix can be computed by summing a series of elements. Each element in $\mathbf{H}_a$ is a zero lag crosscorrelation of two partial derivatives at the receiver locations. At the high-frequency limit, these wavefields would be perfectly uncorrelated with each other. However, the partial derivatives from adjacent nodes are correlated to some extent at finite frequencies. Thus $\mathbf{H}_a$ is diagonally dominant, due to the auto-correlation on the diagonal and banded due to the finite frequency effects (Pratt *et al.* 1998). In addition to this, the approximate Hessian matrix is symmetric so that it's not necessary for all elements to be computed.

Because the full Jacobian matrix is not used, $\mathbf{J}^t\mathbf{g}^n$ in eq. (14) can be obtained by a finite difference approximation:

$$\mathbf{J}^t\mathbf{g}^n \cong \frac{1}{\epsilon}\left[\mathbf{F}(\mathbf{m}^n + \epsilon\,\mathbf{g}^n) - \mathbf{F}(\mathbf{m}^n)\right], \tag{22}$$

where ϵ is a sufficiently small value. Then each iteration step of the inversion requires $2NS + 2NR$ simulations.

3.4 Computational aspects

Although eq. (11) has been proposed several decades ago, it is still a challenging work to solve this problem. The algorithm presented here is implemented with the MPI on an IBM cluster which contains two PowerPC970 2.2 GHz processors and 2 GB of memory in a node and a total of 968 processors. Shipp & Singh (2002) parallelized the backward propagation method so that each simulation runs on different processors simultaneously. In this work, domain decomposition is used for parallelizing the forward and reciprocal simulation. Each processor solves the problem within its small subdomain, communicates with neighbouring processors to update wavefields at each time step and samples forward or reciprocal wavefields within target area in which model parameters are updated.

The resolution required by an accurate forward simulation and the resolution that can be achieved by an inversion are quite different. Numerical stability criteria for wave propagation simulations require smaller grid sizes in the spatial and time domains than those for the inversion. Thus, we use different sizes of grids in the spatial and time domains: a fine grid for forward and reciprocal simulations, and a coarse one for seismic inversion. After both simulations, a virtual source for the inversion is generated by summing wavefields recorded at grid points inside the block for a model parameter.

Because of block parametrization, partial derivatives of each virtual source consist of a large number of partial derivatives inside a block. As the number of grid points describing a virtual source increases, the accuracy of the partial derivative seismogram increases as well. However, there are no significant discrepancies, although derivatives at only 4 grid points are used. Therefore, memory and computation for the Jacobian matrix can be reduced by more than 90 per cent.

In the spatial domain, a model parameter is taken as a block and in the time domain, longer sampling interval and time window of wavefields are used. Required time window for wave propagation simulations is determined to receive all possible reflected waves from the subsurface. The simulation over the entire time window is not necessary for the derivatives but for the residuals because the derivatives are obtained from convolution of waves from a source with those from a receiver. Thus we determine the time window length (NT_p) for the partial derivative by

$$NT_p = NT/2 + (DT + 2/CF)/dt + TP,$$

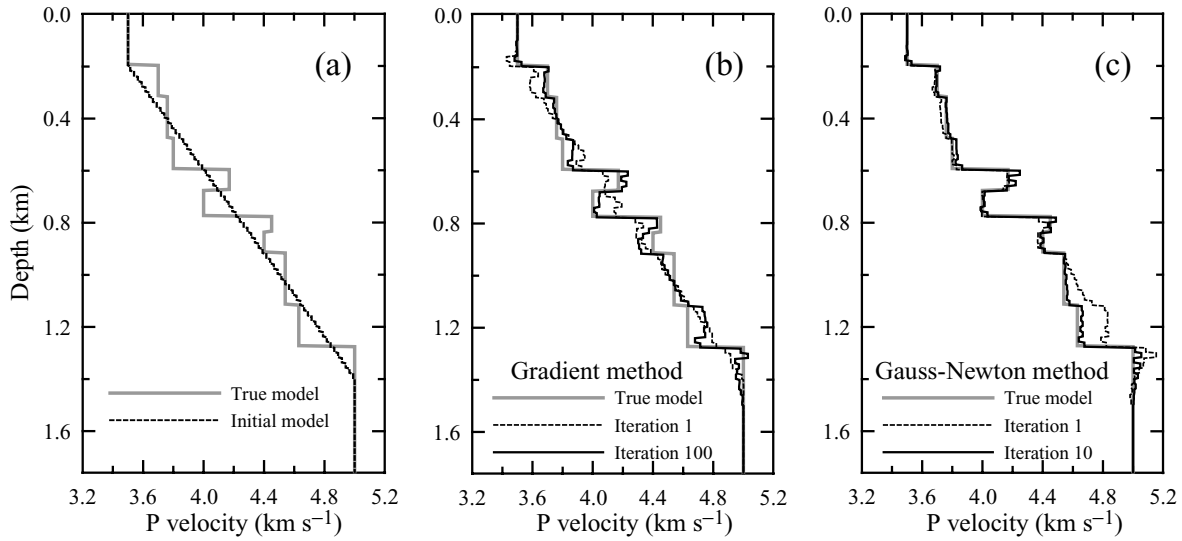

Figure 3. Inversion results from 1-D model: (a) *P*-wave velocity structures of the true model and the initial model and (b and c) updated velocity structures from the gradient method (b) and the Gauss–Newton method (c).

where *NT* and *dt* are the number of time step, the time increment for wave propagation simulations, respectively, *DT* and *CF* represent a delay time and central frequency of a source time function, respectively, and *TP* is a cosine tapering length. Consequently the reciprocal simulations require about 70 per cent of computation of the forward simulations, which reduces the computational and memory burden significantly.

4 NUMERICAL RESULTS

In this section we present some inversion results of synthetic seismic data to examine the resolving power of the Gauss–Newton seismic waveform inversion. In all examples, the source is assumed to be known and the synthetic data have been generated from the same kernel as that used for the inversion.

4.1 Example 1

First, we use a simple layered model in which *P*-wave velocity varies only with depth whereas density and *S*-wave velocity are assumed to be constant and known. *P*-wave velocity of a initial model linearly increases with depth (see Fig. 3). Only the vertical components of seismogram are used as the observed data. Both source and receiver are located at the free surface. The synthetic data has 5 shot gathers which are generated by a point body force of the first derivative of Gaussian function with a dominant frequency of 30 Hz. Each shot gather has 20 receivers with a spacing of 100 m. The minimum and maximum offsets are 100 and 1800 m, respectively. The grid spacing and the time step for wave propagation simulation are 4 m and 0.4 ms, respectively, whereas, for the inversion, vertical and horizontal grid spacings are different which are 20 and 40 m, respectively, and the time step is 4 ms.

Fig. 3 shows the inversion results from the gradient and the Gauss–Newton methods. The result from the gradient method is obtained after 100 iterations and the residual error decreases to 1.0 per cent of its initial value whereas the residual error from the Gauss–Newton method, even only after 10 iterations, achieves 0.4 per cent reduction of the initial. As shown in Kolb *et al.* (1986), the model tends to converge to the true model from the surface downward and the resolution is decreasing with depth. As a rule, deeper part of the model can be resolved after shallower part has been resolved. The low-frequency content of the model is already given by the initial model. In both inversions, the high-frequency content representing the discontinuity converges more rapidly than that of the intermediate frequency representing the layer. As expected, the Gauss–Newton method shows a good convergence rate and the gradient method need much more iterations to resolve the deepest part of the model.

Pre-conditioning of the gradient can improve the convergence rate of the gradient method. However, the choice of pre-conditioner significantly affects its efficiency. Here, a simple pre-conditioner is considered in which we multiply the gradient by an arbitrary power of depth. From several numerical experiments, the pre-conditioner improves the convergence rate as much as about 20 per cent of the gradient method. However, the Gauss–Newton method still shows much faster convergence rate than the gradient method.

Fig. 4 shows the seismograms and the residuals. The scale of the final residuals (Figs 4d and e) is twice as the others. The seismogram from the true model (Fig. 4a) contains significant reflections and multiples, all of which contribute to the inversion, whereas the seismogram from the initial model (Fig. 4b) doesn't show significant events. Thus, the initial residual (Fig. 4c) shows most of events from the true model. In the final residuals from the gradient method (Fig. 4d), there are still intermediate frequency discrepancies generated from the intermediate spatial frequency content at shallow and deep region of the model. The excellent recovery by the Gauss–Newton method is shown in the final residuals (Fig. 4e). All events are well recovered and, consequently, the final model is very close to the true model (see Fig. 3c).

The inverse of the approximate Hessian matrix is known as a focusing filter or a normalizing pre-conditioner. The artefacts in the model update from the gradient method can be effectively removed by the Gauss–Newton method (Pratt *et al.* 1998). Shin *et al.* (2001) showed that a diagonal approximation of the approximate Hessian normalizes the gradient direction so that it improves a poorly scaled image. Fig. 5 shows the model updates after the first iteration of the gradient and the Gauss–Newton methods. The update of the gradient method discriminates the shallow region more clearly than

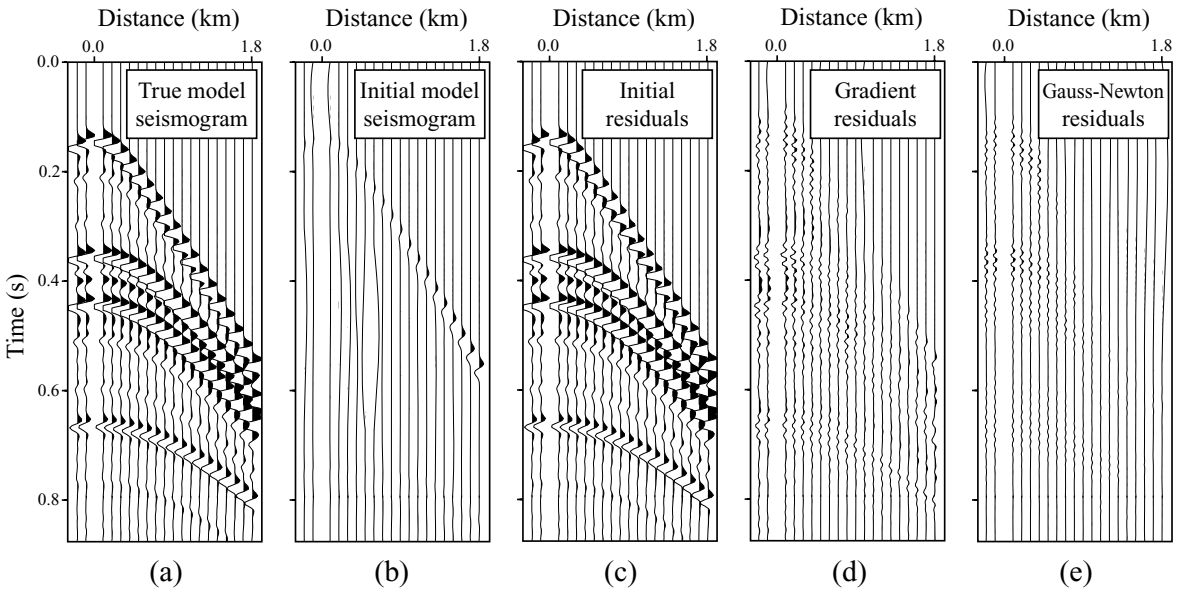

Figure 4. Comparison of vertical component seismograms and residuals from a shot-gather of five shot-gathers: (a) seismogram from the true model, (b) seismogram from an initial model, (c) initial residuals, (d) final residuals from the gradient method (100 iterations) and (e) final residuals from the Gauss–Newton method (10 iterations). For clarity, direct arrivals of (a) and (b) are removed. The scaling factor of the final residuals, (d) and (e), is 2 relative to the others.

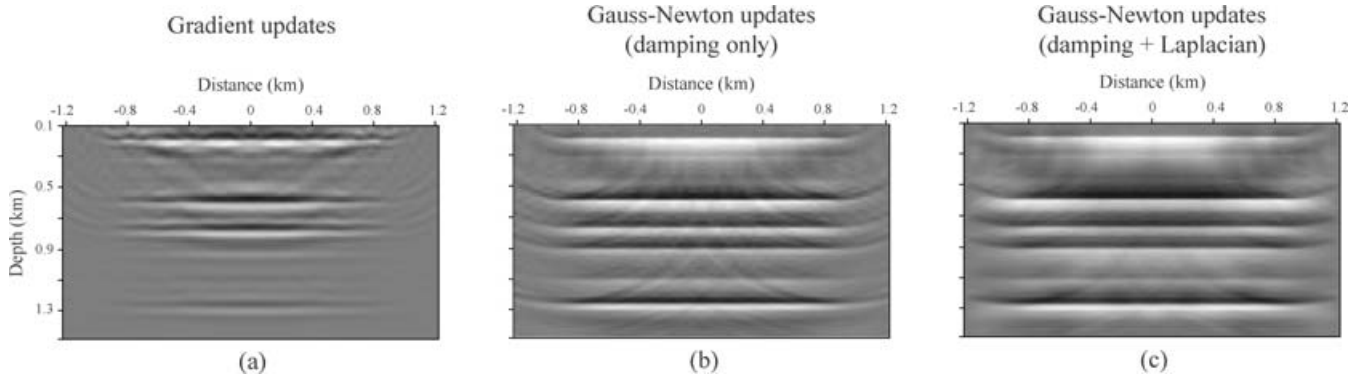

Figure 5. Model updates from the first iterations by the gradient (a), and the Gauss–Newton method (b and c) corresponding to the gradient direction ($\mathbf{J}^t\Delta\mathbf{d}$) and the filtered gradients ($[\mathbf{J}^t\mathbf{J} + \lambda\mathbf{I}^t\mathbf{I}]^{-1}\mathbf{J}^t\Delta\mathbf{d}$ and $[\mathbf{J}^t\mathbf{J} + \lambda_1\mathbf{P}^t\mathbf{P} + \lambda_2\mathbf{I}^t\mathbf{I}]^{-1}\mathbf{J}^t\Delta\mathbf{d}$), respectively. All images are scaled by its own maximum value.

the deeper region whereas that of the Gauss–Newton method discriminates the whole region. It is clearly shown in Fig. 3(c) that, even after the first iteration, the Gauss–Newton method reveals the velocity structure at deep region of the model. Also the former mostly accounts for the high spatial frequency contents of the update information but the latter contains the intermediate frequency contents where is partly because of the Laplacian regularization operator. The damping regularization operator stabilizes the inversion but artefacts from shortage of source–receiver coverage still remain in the update (Fig. 5b). This is reduced by the Laplacian operator (Fig. 5c).

At each iteration, in this example, 5 simulations for the forward simulation, 21 simulations for the reciprocal simulation with a source at the receiver location and 5 simulations for finding an optimal step length were required. The backward propagation method would only need 15 simulations at each iteration because it requires the solution of three times as many forward simulations as the shots (i.e. one to generate synthetic data, another to make the gradient, and the third to compute the step length). Considering the convergence rate and the resolving ability, the Gauss–Newton method is more attractive than the gradient method or the backward propagation method.

4.2 Example 2

The second example has complex geological structures consisting of a fold, discontinuity and dipping layers (Fig. 6a). The model has variations on density and *S*-wave velocity which were assumed to be constant in the previous example. The relationships (Castagna *et al.* 1993) between density, *S*-wave velocity and *P*-wave velocity are assumed to be known as

$$\beta\,(\mathrm{km\,s^{-1}}) = -0.055\,\alpha^2 + 1.017\,\alpha - 1.030,$$
$$\rho\,(\mathrm{g\,cm^{-3}}) = 1.5\,\alpha^{0.225}. \qquad (23)$$

After updating *P*-wave velocities, the other two elastic parameters are updated using these relationships.

The configurations, such as source properties, grid spacing and time step, for the simulation and the inversion are same as before,

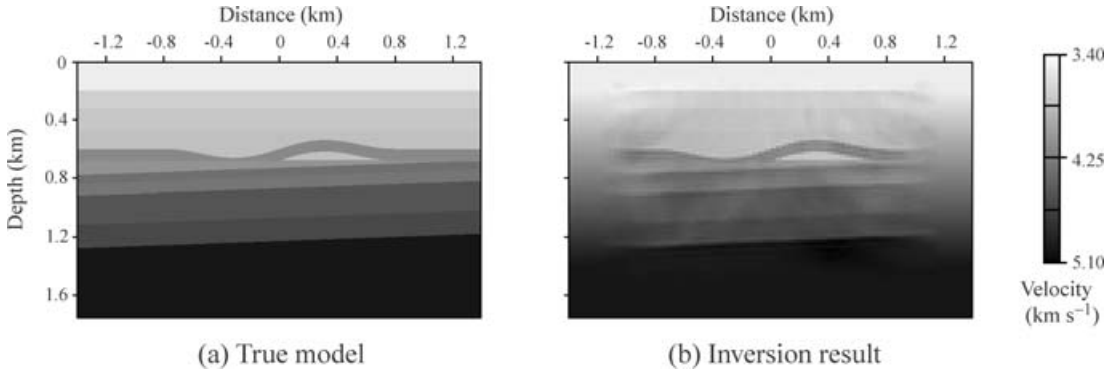

Figure 6. Inversion results of example 2: (a) The true *P*-wave velocity model and (b) inversion result after 5 iterations. Note that the images are at the same scale. Remnant of the initial velocity model can be seen from outsides of updated area of (b) which has vertical velocity variation monotonically increasing with depth.

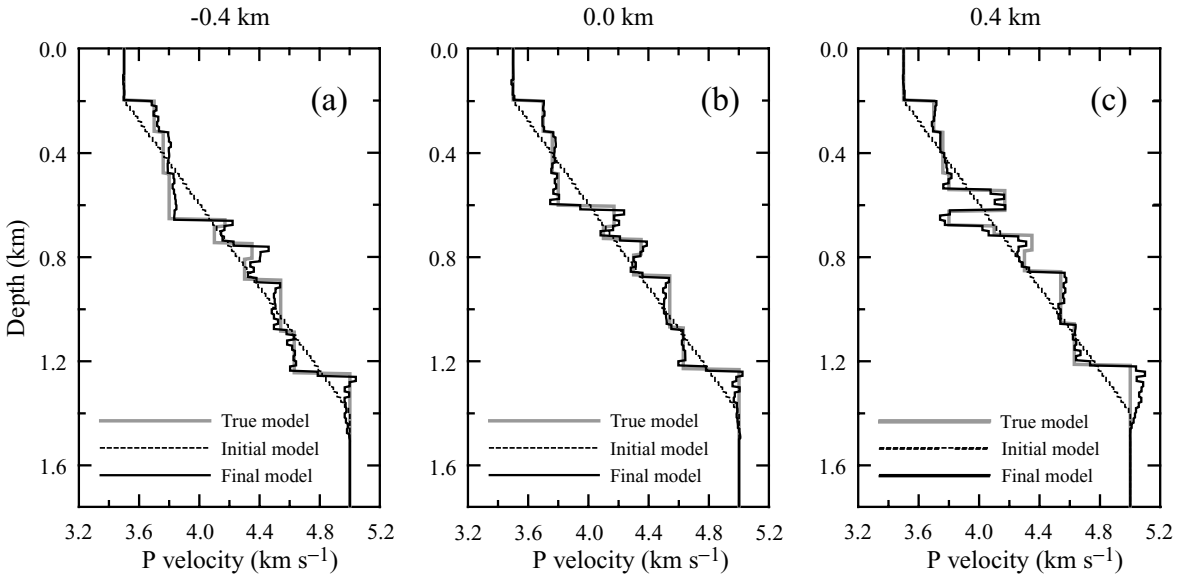

Figure 7. *P*-wave velocity profiles of example 2 at three locations (see Fig. 6). The locations of profiles are at −0.4 km (a), 0.0 km (b) and 0.4 km (c) of the model. The final models are obtained after 5 iterations.

including the initial model. However, the inversion requires more information on the model than that on the previous model because the level of complexity is significantly increased. Therefore, we increase the number of observed data. We use 9 shot gathers and a shot gather has 46 or 47 traces with a spacing of 40 m. The minimum and maximum offsets are 80 and 1920 m, respectively. Both of the horizontal and vertical components of the observed data are used for the inversion.

Fig. 6(b) shows that the structures are correctly resolved within the region of interest. The velocity profiles of the model are plotted in Fig. 7 which compares these to the true model and the initial model, showing that the velocities also agree with the true model. Thus, it may be concluded that the inversion with the Gauss–Newton method can identify a 2-D velocity structure very well and also the properties of the model can be accurately estimated after a few iterations. In this example, the inversion using both of the horizontal and vertical components does not show a notable contribution to the resolving power. Although we do not show the result from the inversion considering only the vertical component of the observed data, the velocities are also recovered accurately.

In this example, 4200 blocks of model parameters are considered. To solve this problem, 9 forward simulations, 98 reciprocal simulations and 9 simulations for the step length are required at each iteration which takes about 14 min on the IBM cluster, described previously, with 48 CPUs. Forward and reciprocal simulations take 3 and 27 per cent of total computing time, respectively, and constructing the Jacobian and approximate Hessian matrix spends 66 per cent. Note that the reciprocal simulation requires about 70 per cent of computation of the forward simulation. The principal memory requirements are about 1.0 GB for saving the forward wavefields, F_p^m in eq. (18), about 75 MB for the Jacobian matrix, $\mathbf{J}_p$ in eq. (21), and about 67 MB for the approximate Hessian matrix, $\mathbf{J}^t\mathbf{J}$ in eq. (21), respectively. By virtue of the approximation of virtual sources, the largest memory requirement, 1.0 GB for the forward wavefields, is reduced to 78 MB without any loss of accuracy.

In order to invert the noise-corrupted data, we added 5 per cent of Gaussian noise to the synthetic data. Fig. 8 shows seismograms without and with Gaussian noises. In example 1, density and *S*-wave velocity are assumed to be constant so that only *P*-wave arrivals are

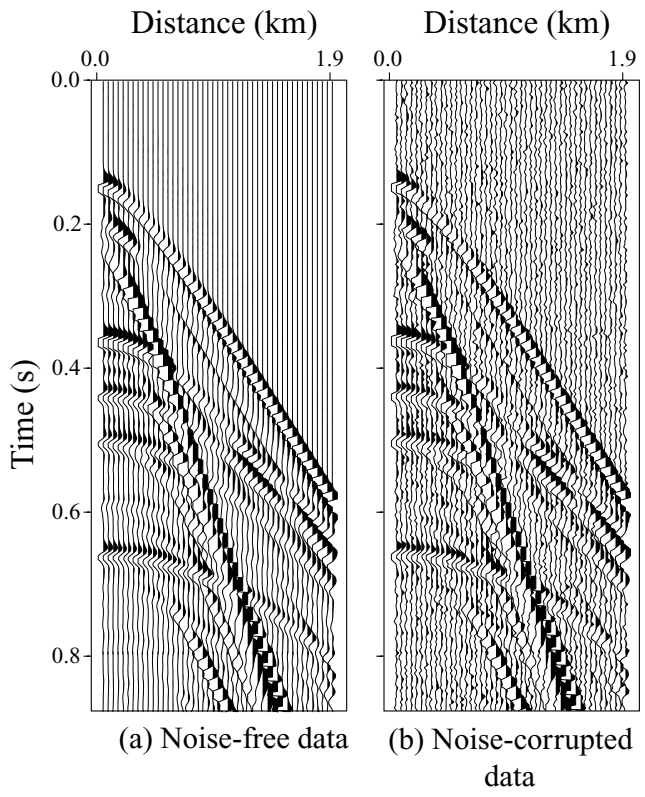

Figure 8. Vertical component seismograms of a shot-gather from example 2: (a) original seismogram and (b) noise-corrupted seismogram. Direct arrivals are removed.

observed. In this example, we consider all parameters have variations and, thus, it shows diverse *P* and *S*-wave arrivals. The inversion of the noise-corrupted data was performed in the same way as before. The result of the inversion, Fig. 9, shows that the noise slightly contaminates the model, but still yields an accurate identification of the model.

As shown by Tarantola (1986), the choice of model parameters is not neutral in the inversion. From numerical experiments, variation in density doesn't affect the inversion much. In order to compare convergent rates of *P*- and *S*-wave velocities, these velocities are updated independently of each other, but density is fixed during the inversion. Noise-corrupted data is used as observed data. Fig. 10 shows *P* and *S*-wave velocity profiles of the model. *P*-wave velocity of the model is recovered well up to the last reflector, whereas the inversion couldn't recover *S*-wave velocity profile at depths larger than 1 km as the data has limited information of *S*-wave velocity beyond this depth.

4.3 Example 3

In this example, we consider a more realistic case instead of applying this method to real data. The data consists of 20 shot gathers collected with a total of 125 receivers and has 50 per cent of Gaussian noise (Fig. 11). The shot and receiver spacings are 600 and 120 m, respectively. The maximum offset of each shot gather is 6 km. The time step for wave propagation simulation is 0.001 s, total record length is 2.25 s, and a dominant frequency of 8 Hz of source is used. The length and depth of region of interest is 14 and 3.5 km, which consists of 9800 blocks of model parameters. The relationships between density, *S*- and *P*-wave velocities are assumed to follow eq. (23).

The inversion results are obtained after 3 iterations and shown in Fig. 12. The result from noise-free data (Fig. 12b) shows excellent recovery of the true model and that from noise-corrupted data (Fig. 12c) also shows good recovery. Higher contamination in deeper region can be explained by relatively low signal to noise ratio of late arrivals. Fig. 13 shows the velocity profiles of the model. The agreement with the true velocity profile is quite satisfactory, considering the crudeness of the noise-corrupted data.

The inversion is performed with 48 CPUs and each iteration takes 3 hr mainly because of construction of the Jacobian and approximate Hessian matrices, taking 83 per cent of computing time. The forward and reciprocal simulations take 3 and 12 per cent, respectively.

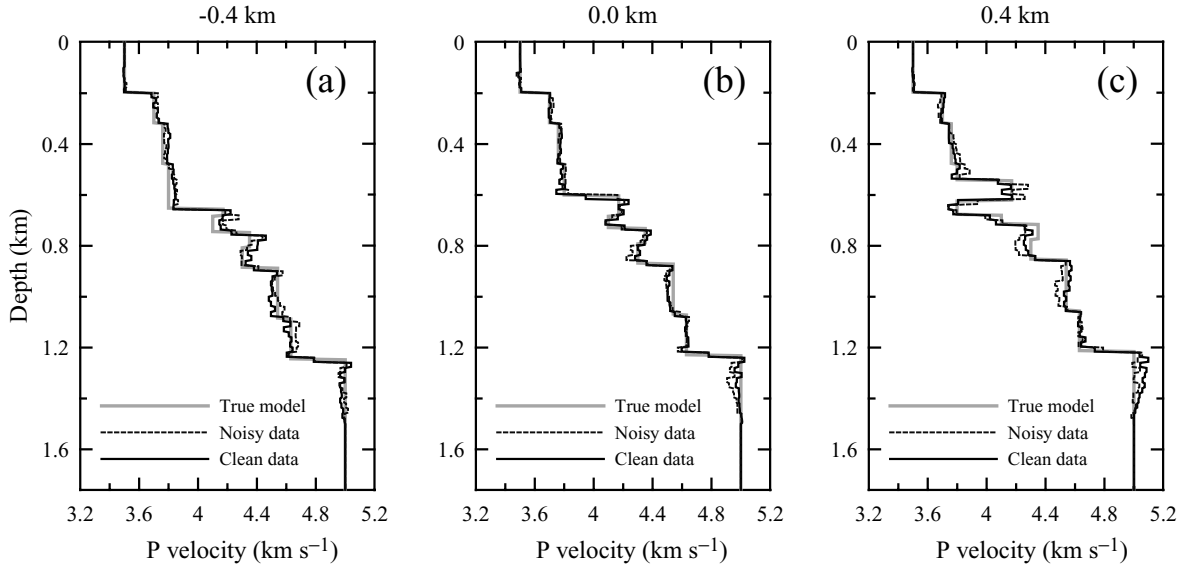

Figure 9. *P*-wave velocity profiles using noise-corrupted data of example 2 at three locations. The locations of profiles are at −0.4 km (a), 0.0 km (b) and 0.4 km (c) of the model. Density and *S*-wave velocity are updated by using eq. (23).

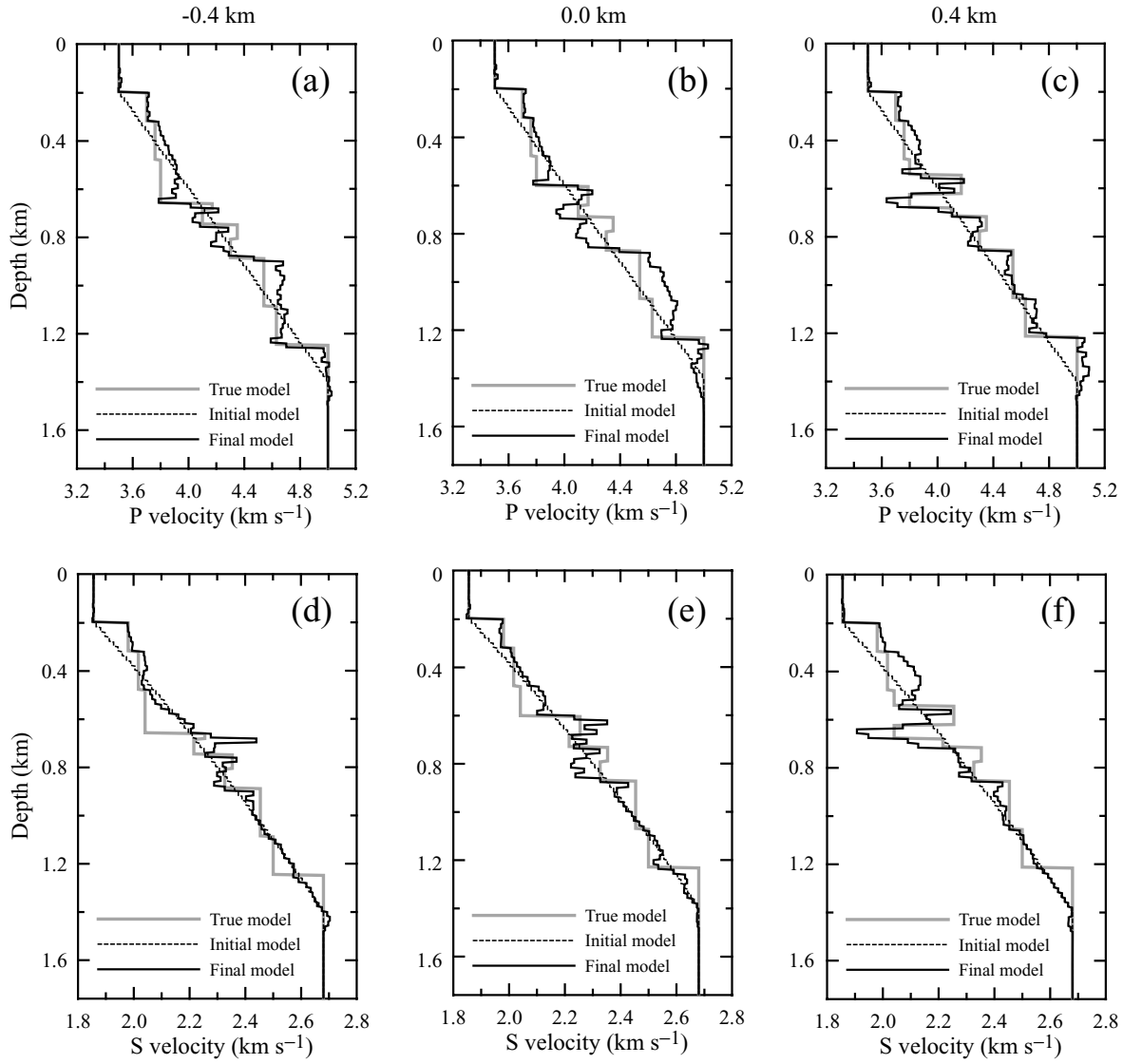

Figure 10. *P*- and *S*-wave velocity profiles using noise-corrupted data of example 2: (a)–(c) *P*-wave profiles, and (d)–(f) *S*-wave profiles. *S*-wave velocity is updated independently of *P*-wave velocity but density is forced to be constant. The locations of profiles are at −0.4 km (a and d), 0.0 km (b and e), and 0.4 km (c and f) of the model.

Comparison with the previous example indicates that the percentage increase of the computation time for the matrices is due to the increase of number of model parameter.

5 CONCLUSIONS

Huge computational and memory requirements have prevented solving the Gauss–Newton seismic waveform inversion (Tarantola 1984). Unfortunately, it is still a difficult problem. In this paper, we have implemented a seismic waveform inversion scheme based on the Gauss–Newton method. We used the velocity-stress staggered-grid finite difference method (Levander 1988) for simulating seismic wave propagation in elastic media and the PML method (Berenger 1994) as an absorbing boundary condition. This approach allows us to accurately simulate seismic wavefields, use multicomponent data, and resolve subsurface elastic features correctly and efficiently. Explicit computations of the Jacobian and the approximate Hessian matrices have been carried out by the reciprocity principle and the convolution theorem, which significantly reduces the number of forward simulations. To overcome the computational and memory limitation, we used different sizes of grids in the spatial and time domains for the inversion, temporal windowing for the partial derivative and approximation of virtual sources, and parallelized all these approaches on massively parallel computers.

In numerical experiments, the Gauss–Newton seismic waveform inversion has been successfully applied to resolve the high and intermediate spatial frequency content of the model. In particular, the results show that the Gauss–Newton method has faster convergence rate than the gradient method for seismic waveform inversion. Furthermore, not only does the method recover the structure correctly, but also estimates velocities accurately. Further work is required to study the effect of the choices of model parameters and, most importantly, to test the method for real seismic data.

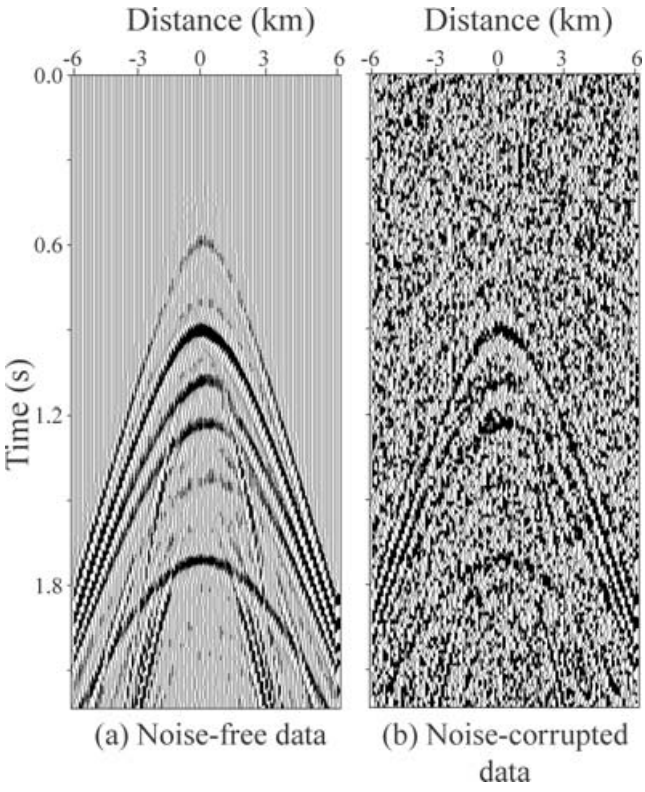

Figure 11. Vertical component seismograms for example 3: (a) original seismogram and (b) noise-corrupted seismogram. Direct arrivals are removed.

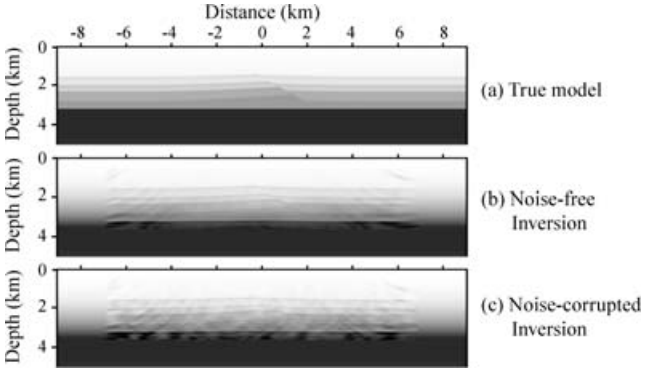

Figure 12. Inversion results of example 3: (a) The true *P*-wave velocity model, (b) inversion result from noise-free data, and (c) from noise-corrupted data. Both inversion results are obtained after three iterations. Note that the images are at the same scale.

ACKNOWLEDGMENTS

This work was partially supported by a grant from the Office of Science (# DE-F602-91ER14175), a contract from the Office of Fossil Energy (# DE-RA26-99FT40160) of the United States Department of Energy, the BK21 program through the School of Earth and Environmental Sciences, Seoul National University, and a project of development of fundamental techniques for evaluating earthquake hazard from the Korea Meteorological Administration. We would like to acknowledge Mary Papakhian (University Information Technology Services, Indiana University) for reserving CPUs on the IBM SP cluster. This work was also supported in part by facilities provided through Shared University Research grants from IBM, Inc.

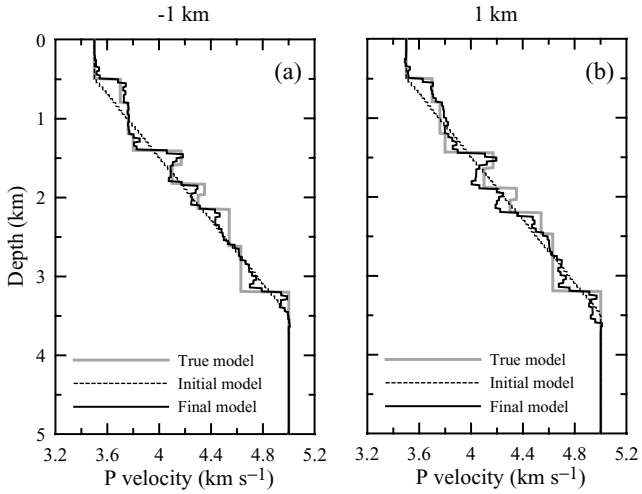

Figure 13. *P*-wave velocity profiles of example 3: The locations of profiles are at −1 km (a) and 1 km (b).

REFERENCES

Aki, K. & Richards, P.G., 1980. *Quantitative Seismology: Theory and Methods*, **Vol. 1,** W.H. Freeman and Company, San Francisco, California.

Arntsen, B. & Carcione, J.M., 2000. A new insight into the reciprocity principle, *Geophysics,* **65,** 1604–1612.

Berenger, J.P., 1994. A perfectly matched layer for the absorption of electromagnetic waves, *J. Comput. Phys.,* **114,** 185–200.

Castagna, J.P., Batzle, M.L. & Kan, T.K., 1993. Rock physics—the link between rock properties and avo response, in *Offset-dependent reflectivity–Theory and Practice of AVO Anomalies*, pp. 135–171, eds Castagna, J.P. & Backus, M.M. no. 8 in Investigations in Geophysics, Soc. Expl. Geophys., Tulsa.

Clément, F., Chavent, G. & Gómez, S., 2001. Migration-based traveltime waveform inversion of 2-d simple structures: A synthetic example, *Geophysics,* **66,** 845–860.

Eisner, L. & Clayton, R.W., 2001. A reciprocity method for multiple-source simulations, *Bull. seism. Soc. Am.,* **91,** 553–560.

Freudenreich, Y. & Shipp, R., 2000. Full waveform inversion of seismic data: frequency versus time domain, *Lithos Science Report,* **2,** 25–30.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics,* **51,** 1387–1403.

Hicks, G.J. & Pratt, R.G., 2001. Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit, *Geophysics,* **66,** 598–612.

Kolb, P., Collino, F. & Lailly, P., 1986. Pre-stack inversion of a 1-d medium, *Proceedings of the IEEE,* **74,** 498–508.

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering: Theory and Application*, pp. 206–220, eds Bednar, J.B., Redner, R., Robinson, E. & Weglein, A., Soc. Ind. Apl. Math., Philadelphia.

Levander, A.R., 1988. Fourth-order finite-difference p-sv seismograms, *Geophysics,* **53,** 1425–1436.

Levenberg, K., 1944. A method for the solutions of certain nonlinear problems in least squares, *Quart. Appl. Math.,* **2,** 164–168.

Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Indust. Appl. Math,* **11,** 431–441.

Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics,* **52,** 1211–1228.

Pica, A., Diet, J.P. & Tarantola, A., 1990. Nonlinear inversion of seismic reflection data in a laterally invariant medium, *Geophysics,* **55,** 284–292.

Pratt, R.G., 1990. Inverse theory applied to multi-source cross-hole tomography. part 2: elastic wave-equation method, *Geophys. Prospect.,* **38,** 311–329.

Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.,* **133,** 341–362.

Rodi, W.L., 1976. A technique for improving the accuracy of finite element solutions for magnetotelluric data, *Geophys. J. R. astr. Soc.,* **44,** 483–506.

Sasaki, Y., 1989. Two-dimensional joint inversion of magnetotelluric and dipole-dipole resistivity data, *Geophysics,* **54,** 254–262.

Sheen, D.-H., Tuncay, K., Baag, C.-E. & Ortoleva, P.J., 2004. Efficient finite difference calculation of partial derivative seismic wavefield using reciprocity and convolution, in *74th Annual meeting*, SEG.

Shin, C., Yoon, K., Marfurt, K.J., Park, K., Yang, D., Lim, H.Y., Chung, S. & Shin, S., 2001. Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion, *Geophysics,* **66,** 1856–1863.

Shipp, R.M. & Singh, S.C., 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophys. J. Int.,* **151,** 325–344.

Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics,* **69,** 231–248.

Snieder, R., Xie, M.Y., Pica, A. & Tarantola, A., 1989. Retrieving both the impedance contrast and background velocity: a global strategy for the seismic reflection problem, *Geophysics,* **54,** 991–1000.

Symes, W.W. & Carazzone, J.J., 1991. Velocity inversion by differential sembrance optimization, *Geophysics,* **56,** 654–663.

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

Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics,* **51,** 1893–1903.

Tarantola, A., 1987. *Inverse Problem Theory: Methods for Data Fitting and Parameter Estimations*, Elsevier Science Publ. Co., New York.

Tarantola, A. & Valette, B., 1982. Generalized non linear inverse problems solved using the least-squares criterion, *Rev. of Geophys. and Space Physics,* **20,** 219–232.

APPENDIX A: PARTIAL DERIVATIVES

In this Appendix we present the differentiations of the wave equation with respect to model parameters for different choices of parameters. In the case of model parameters of density, ρ, and one of Lamé's moduli, λ and μ, the differential equations are given by

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial \rho_p} &= \frac{\partial \tau_{ij,j}}{\partial \rho_p} - \frac{\partial \rho}{\partial \rho_p}\dot{v}_i,\\
\frac{\partial \dot{\tau}_{ij}}{\partial \rho_p} &= \mu\left(\frac{\partial v_{i,j}}{\partial \rho_p} + \frac{\partial v_{j,i}}{\partial \rho_p}\right) + \lambda\,\delta_{ij}\frac{\partial v_{k,k}}{\partial \rho_p},
\end{aligned} \tag{A1}$$

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial \lambda_p} &= \frac{\partial \tau_{ij,j}}{\partial \lambda_p},\\
\frac{\partial \dot{\tau}_{ij}}{\partial \lambda_p} &= \mu\left(\frac{\partial v_{i,j}}{\partial \lambda_p} + \frac{\partial v_{j,i}}{\partial \lambda_p}\right) + \lambda\,\delta_{ij}\frac{\partial v_{k,k}}{\partial \lambda_p} + \frac{\partial \lambda}{\partial \lambda_p}\,\delta_{ij}\,v_{k,k},
\end{aligned} \tag{A2}$$

and

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial \mu_p} &= \frac{\partial \tau_{ij,j}}{\partial \mu_p},\\
\frac{\partial \dot{\tau}_{ij}}{\partial \mu_p} &= \mu\left(\frac{\partial v_{i,j}}{\partial \mu_p} + \frac{\partial v_{j,i}}{\partial \mu_p}\right) + \lambda\,\delta_{ij}\frac{\partial v_{k,k}}{\partial \mu_p}\\
&\quad + \frac{\partial \mu}{\partial \mu_p}(v_{i,j} + v_{j,i}).
\end{aligned} \tag{A3}$$

where eq. (A1) is equvalent to eq. (17). For model parameters of density, ρ, and P- and S-wave velocities, α and β, the equations are given by

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial \rho_p} &= \frac{\partial \tau_{ij,j}}{\partial \rho_p} - \frac{\partial \rho}{\partial \rho_p}\dot{v}_i\,,\\
\frac{\partial \dot{\tau}_{ij}}{\partial \rho_p} &= \mu\left(\frac{\partial v_{i,j}}{\partial \rho_p} + \frac{\partial v_{j,i}}{\partial \rho_p}\right) + \lambda\,\delta_{ij}\frac{\partial v_{k,k}}{\partial \rho_p} + \frac{1}{\rho}\frac{\partial \rho}{\partial \rho_p}\dot{\tau}_{ij},
\end{aligned} \tag{A4}$$

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial \alpha_p} &= \frac{\partial \tau_{ij,j}}{\partial \alpha_p},\\
\frac{\partial \dot{\tau}_{ij}}{\partial \alpha_p} &= \mu\left(\frac{\partial v_{i,j}}{\partial \alpha_p} + \frac{\partial v_{j,i}}{\partial \alpha_p}\right) + \lambda\,\delta_{ij}\frac{\partial v_{k,k}}{\partial \alpha_p}\\
&\quad + 2\,\rho\,\alpha\,\frac{\partial \alpha}{\partial \alpha_p}\,\delta_{ij}\,v_{k,k},
\end{aligned} \tag{A5}$$

and

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial \beta_p} &= \frac{\partial \tau_{ij,j}}{\partial \beta_p},\\
\frac{\partial \dot{\tau}_{xx}}{\partial \beta_p} &= (\lambda + 2\mu)\frac{\partial v_{x,x}}{\partial \beta_p} + \lambda\frac{\partial v_{z,z}}{\partial \beta_p} - 4\,\rho\,\beta\,\frac{\partial \beta}{\partial \beta_p}v_{z,z},\\
\frac{\partial \dot{\tau}_{zz}}{\partial \beta_p} &= \lambda\frac{\partial v_{x,x}}{\partial \beta_p} + (\lambda + 2\mu)\frac{\partial v_{z,z}}{\partial \beta_p} - 4\,\rho\,\beta\,\frac{\partial \beta}{\partial \beta_p}v_{x,x},\\
\frac{\partial \dot{\tau}_{xz}}{\partial \beta_p} &= \mu\left(\frac{\partial v_{x,z}}{\partial \beta_p} + \frac{\partial v_{z,x}}{\partial \beta_p}\right) + \frac{2}{\beta}\frac{\partial \beta}{\partial \beta_p}\dot{\tau}_{xz}.
\end{aligned} \tag{A6}$$

For model parameters of density, ρ, and P- and S-wave impedances, IP and IS, the equations are given by

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial \rho_p} &= \frac{\partial \tau_{ij,j}}{\partial \rho_p} - \frac{\partial \rho}{\partial \rho_p}\dot{v}_i,\\
\frac{\partial \dot{\tau}_{ij}}{\partial \rho_p} &= \mu\left(\frac{\partial v_{i,j}}{\partial \rho_p} + \frac{\partial v_{j,i}}{\partial \rho_p}\right) + \lambda\,\delta_{ij}\frac{\partial v_{k,k}}{\partial \rho_p} - \frac{1}{\rho}\frac{\partial \rho}{\partial \rho_p}\dot{\tau}_{ij},
\end{aligned} \tag{A7}$$

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial IP_p} &= \frac{\partial \tau_{ij,j}}{\partial IP_p},\\
\frac{\partial \dot{\tau}_{ij}}{\partial IP_p} &= \mu\left(\frac{\partial v_{i,j}}{\partial IP_p} + \frac{\partial v_{j,i}}{\partial IP_p}\right) + \lambda\delta_{ij}\frac{\partial v_{k,k}}{\partial IP_p}\\
&\quad + 2\frac{IP}{\rho}\frac{\partial IP}{\partial IP_p}\,\delta_{ij}v_{k,k},
\end{aligned} \tag{A8}$$

and

$$\begin{aligned}
\rho\,\frac{\partial \dot{v}_i}{\partial IS_p} &= \frac{\partial \tau_{ij,j}}{\partial IS_p},\\
\frac{\partial \dot{\tau}_{xx}}{\partial IS_p} &= (\lambda + 2\mu)\frac{\partial v_{x,x}}{\partial IS_p} + \lambda\frac{\partial v_{z,z}}{\partial IS_p} - 4\,\frac{IS}{\rho}\frac{\partial IS}{\partial IS_p}v_{z,z},\\
\frac{\partial \dot{\tau}_{zz}}{\partial IS_p} &= \lambda\frac{\partial v_{x,x}}{\partial IS_p} + (\lambda + 2\mu)\frac{\partial v_{z,z}}{\partial IS_p} - 4\,\frac{IS}{\rho}\frac{\partial IS}{\partial IS_p}v_{x,x},\\
\frac{\partial \dot{\tau}_{xz}}{\partial IS_p} &= \mu\left(\frac{\partial v_{x,z}}{\partial IS_p} + \frac{\partial v_{z,x}}{\partial IS_p}\right) + \frac{2}{IS}\frac{\partial IS}{\partial IS_p}\dot{\tau}_{xz}.
\end{aligned} \tag{A9}$$