C. R. Mecanique 338 (2010) 364–376

Contents lists available at ScienceDirect

Comptes Rendus Mecanique

www.sciencedirect.com

Inverse problems

Parameter sensitivity analysis of a nonlinear least-squares optimization-based anelastic full waveform inversion method

Analyse de sensibilité aux paramètres d'une méthode d'inversion sismique reposant sur les formes d'onde anélastiques et les moindres carrés non linéaires

Aysegul Askan [a,*], Volkan Akcelik [b,1], Jacobo Bielak [b], Omar Ghattas [c]

[a] *Department of Civil Engineering and Earthquake Engineering Research Center, Middle East Technical University, Ankara, 06531, Turkey*
[b] *Computational Seismology Laboratory, Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[c] *Jackson School of Geosciences, Department of Mechanical Engineering, and Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX 78712, USA*

ARTICLE INFO

Article history:
Available online 16 August 2010

Keywords:
Waves
Waveform inversion
Adjoint formulation
Intrinsic attenuation

Mots-clés :
Ondes
Inversion par forme d'onde
Formulation adjointe
Atténuation intrinsèque

ABSTRACT

In a recent article, we described a seismic inversion method for determining the crustal velocity and attenuation of basins in earthquake-prone regions. We formulated the problem as a constrained nonlinear least-squares optimization problem in which the constraints are the equations that describe the forward wave propagation. Here, we conduct a parametric study to investigate the influence of parameters such as the form of the regularization function, receiver density, preconditioning, noise level of the data, and the multilevel continuation technique on the cost and quality of the inversion. We use the same 2D Los Angeles example as in our earlier study.

RÉSUMÉ

Nous avons décrit dans un article récent une méthode d'inversion sismique, formulée comme un problème de moindres carrés non linéaires assujetti à des contraintes correspondant aux équations décrivant le problème direct de propagation d'ondes, permettant de déterminer la célérité et l'atténuation crustales de bassins situés dans des zones à forte seismicité. Dans cet article, une étude paramétrique est menée afin d'évaluer l'influence sur le coût et la qualité de l'inversion de paramètres tels que la forme de la fonctionnelle de régularisation, la densité de capteurs, le préconditionnement, le niveau de bruit entachant les données et la méthode de raffinement progressif de l'espace des inconnues. L'exemple utilisé (modèle 2D du bassin de Los Angeles) est le même que dans l'article précédent.

* Corresponding author.
E-mail addresses: aaskan@metu.edu.tr (A. Askan), volkan.akcelik@gmail.com (V. Akcelik), jbielak@cmu.edu (J. Bielak), omar@ices.utexas.edu (O. Ghattas).
[1] Present address: Exxon Mobil Research and Engineering, 1545 Route 22 East, Annandale, NJ 08801, USA.

1631-0721/$ – see front matter
doi:10.1016/j.crme.2010.07.002

1. Inverse anelastic wave propagation problem

The fundamental problem of seismic inversion is to recover information on physical properties of the Earth using seismic observations. Full-waveform inversion is one of numerous modern methodologies for solving inverse problems in seismology. Since the pioneering work of Tarantola (e.g. [1,2]) and Tarantola and Valette [3] on the full waveform nonlinear seismic inversion using a least-squares objective function and adjoint operators, the adjoint full-waveform inversion method has been extensively used both in the time domain (e.g. [4,5]) and in the frequency domain (e.g. [6–9]). While most of the studies on full waveform inversion are based on elastic earth models, some authors (e.g. [10–12]) used models for which the intrinsic attenuation is taken into consideration.

In a recent article [13], hereinafter referred to as Paper I, we presented an anelastic full-waveform seismic inversion methodology for estimating shear wave velocity and attenuation in large basins from surface observations of ground motion. The methodology was tested through synthetic inversions for the shear wave velocity and attenuation profiles in a cross section of the shear wave velocity model for a sedimentary soil medium within the Los Angeles basin obtained from the Southern California Earthquake Center (SCEC) Community Velocity Model [14]. Paper I presented inversion results obtained for a well-chosen set of parameters and options in the inversion method. However, the choice of parameters and options can have a strong influence on the performance of the inversion method. In this paper, we perform a number of experiments to study the effects of these parameters on the numerical and qualitative performance of the inversion. These experiments include the sensitivity of the inversion to the type of regularization functional, value of the regularization parameter, receiver density, preconditioning, data noise, and the multilevel technique. We present here results of inversion experiments performed for only one inversion field, the shear modulus, but the conclusions extend to inversion with more than one inversion field [15].

In this section, we briefly summarize the inverse anelastic wave propagation algorithm and the solution technique. In the next section, we present results of the parametric study.

1.1. Nonlinear least-squares formulation

We describe the methodology for inverse anelastic scalar antiplane seismic wave propagation problem in two dimensions. We employ a single generalized standard linear solid (SLS) to represent the anelastic behavior of the constitutive material [13]. For simplicity, we use plane wave absorbing boundaries applied on the truncation surface. Thus, the problem is defined over a finite domain Ω with boundary Γ; Γ consists of a horizontal free surface Γ_{FS} and the truncation boundary Γ_{AB} on which a simple absorbing boundary condition is imposed to limit the occurrence of spurious reflections from the outgoing waves. The system is initially at rest.

We pose the inverse wave propagation problem as a nonlinear least-squares minimization problem. The data misfit objective function is constrained by the partial differential equation (PDE) that describes the forward anelastic wave equation and the ordinary differential equation (ODE) that governs the motion of the anelastic mechanism together with the associated boundary and initial conditions. Receivers at specified locations on the free surface record displacement time histories. The objective of the inversion is to minimize the L^2 norm of the difference between the observed and predicted displacements at the receiver locations with respect to the shear modulus field. The inverse problem for anelastic scalar wave propagation is formulated (for a single event) as:

$$\min_{u,v,\mu} \frac{1}{2} \sum_{j=1}^{N_R} \int_0^T \int_\Omega \left[u^* - u\right]^2 \delta(x - x_j)\, \mathrm{d}\Omega\, \mathrm{d}t + \beta R(\mu) \tag{1}$$

subject to:

$$\rho \frac{\partial^2 u}{\partial t^2} - \nabla \cdot \left[\mu \nabla (u + \eta v)\right] = f(t) \quad \text{in } \Omega \times [0, T] \tag{1a}$$

$$\frac{\partial v}{\partial t} + \alpha v = \frac{\partial u}{\partial t} \quad \text{in } \Omega \times [0, T] \tag{1b}$$

$$\mu \nabla (u + \eta v) \cdot n = 0 \quad \text{on } \Gamma_{FS} \times [0, T] \tag{1c}$$

$$\mu \nabla (u + \eta v) \cdot n = \sqrt{\rho \mu} \frac{\partial u}{\partial t} \quad \text{on } \Gamma_{AB} \times [0, T] \tag{1d}$$

$$u = \frac{\partial u}{\partial t} = 0, \qquad v = 0 \quad \text{at } t = 0 \tag{1e}$$

where N_R is the number of receivers, $\delta(x)$ is the Dirac delta function, x_j is the location of jth receiver, and u^* is the recorded displacement. We refer to the displacements $u(x,t)$ and $v(x,t)$ as the state variables and the shear modulus $\mu(x)$ as the inversion variable. The density, ρ, viscoelastic parameters, α and η, and source function $f(t)$ are assumed to be known. In this formulation, we use as an auxiliary variable in the SLS a displacement v, in contrast to the strain, as is done in general (e.g. [16,17]). When generalized to three-dimensional problems, this results in significant computational savings.

A regularization term, $\beta R(\mu)$, has been added to the data misfit least-squares term in Eq. (1) to address the ill-posedness of the inverse problem. The ill-posedness is due to the band-limited information available from the observations coupled with the infinite dimensionality of the inversion field. Below we present the optimality conditions based on total variation regularization (TVR); however, in the next section we discuss inversion results with the Tikhonov regularization (TR) as well as with the TVR.

We first state the optimality conditions, as derived in Paper I. The Lagrangian functional of the optimization problem is given by:

$$\begin{aligned}\mathscr{L}(u,v,\mu,\lambda,\phi) = {} & \frac{1}{2}\sum_{j=1}^{N_R}\int_0^T\int_\Omega \big[u^* - u\big]^2\delta(x-x_j)\,\mathrm{d}\Omega\,\mathrm{d}t + \beta R(\mu) \\ & + \int_0^T\int_\Omega \lambda\left\{\rho\frac{\partial^2 u}{\partial t^2} - \nabla\cdot\big[\mu\nabla(u+\eta v)\big] - f\right\}\mathrm{d}\Omega\,\mathrm{d}t + \int_0^T\int_{\Gamma_{FS}} \lambda\big\{\mu\nabla(u+\eta v)\cdot n\big\}\,\mathrm{d}\Gamma_{FS}\,\mathrm{d}t \\ & + \int_0^T\int_{\Gamma_{AB}} \lambda\left\{\mu\nabla(u+\eta v)\cdot n - \sqrt{\rho\mu}\frac{\partial u}{\partial t}\right\}\mathrm{d}\Gamma_{AB}\,\mathrm{d}t + \int_0^T\int_\Omega \phi\left\{\frac{\partial v}{\partial t} + \alpha v - \frac{\partial u}{\partial t}\right\}\mathrm{d}\Omega\,\mathrm{d}t \end{aligned} \tag{2}$$

where $\lambda(x,t)$ and $\phi(x,t)$ are the Lagrange multipliers (or adjoint variables) for the PDE and ODE, respectively.

The first order optimality conditions, also known as the Karush–Kuhn–Tucker (KKT) conditions, state that the first variation of the Lagrangian with respect to u, v, μ, λ, and ϕ, respectively, should vanish at the optimum. The variations of the Lagrangian with respect to λ and ϕ respectively yield the original state problems (1a), (1b):

$$\begin{aligned} & \rho\frac{\partial^2 u}{\partial t^2} - \nabla\cdot\mu\nabla(u+\eta v) = f \quad \text{in } \Omega\times[0,T] \\ & \mu\nabla(u+\eta v)\cdot n = 0 \quad \text{in } \Gamma_{FS}\times[0,T] \\ & \mu\nabla(u+\eta v)\cdot n = \sqrt{\rho\mu}\,\frac{\partial u}{\partial t} \quad \text{in } \Gamma_{AB}\times[0,T] \\ & u = \frac{\partial u}{\partial t} = 0 \quad \text{in } \Omega \text{ at } t=0 \end{aligned} \tag{3}$$

and

$$\begin{aligned} & \frac{\partial v}{\partial t} + \alpha v - \frac{\partial u}{\partial t} = 0 \quad \text{in } \Omega\times[0,T] \\ & v = 0 \quad \text{in } \Omega \text{ at } t=0 \end{aligned} \tag{4}$$

Taking variations of the Lagrangian with respect to the state variables u and v leads to adjoint problems in terms of λ and ϕ as follows:

$$\begin{aligned} & \rho\frac{\partial^2\lambda}{\partial t^2} - \nabla\cdot\mu\nabla\lambda + \frac{\partial\phi}{\partial t} = \sum_{j=1}^{N_R}\big(u^* - u\big)\delta(x-x_j) \quad \text{in } \Omega\times[0,T] \\ & \mu\nabla\lambda\cdot n = 0 \quad \text{in } \Gamma_{FS}\times[0,T] \\ & \mu\nabla\lambda\cdot n = -\sqrt{\rho\mu}\,\frac{\partial\lambda}{\partial t} \quad \text{in } \Gamma_{AB}\times[0,T] \\ & \lambda = \frac{\partial\lambda}{\partial t} = 0 \quad \text{in } \Omega \text{ at } t=T \end{aligned} \tag{5}$$

and

$$\begin{aligned} & \frac{\partial\phi}{\partial t} - \alpha\phi = -\eta\nabla\cdot\mu\nabla\lambda \quad \text{in } \Omega\times[0,T] \\ & \phi = 0 \quad \text{in } \Omega \text{ at } t=T \end{aligned} \tag{6}$$

The final KKT condition derived by taking the variation of the Lagrangian with respect to the shear modulus μ yields the material equation as follows:

$$-\beta\nabla\cdot\left[\frac{\nabla\mu}{|(\nabla\mu\cdot\nabla\mu+\varepsilon)^{1/2}|}\right]+\int_0^T\nabla\lambda\cdot\nabla\big(u+\eta(\mu)v\big)\,\mathrm{d}t-\int_0^T D_\mu\eta(\mu)v\nabla\cdot(\mu\nabla\lambda)\,\mathrm{d}t+\int_0^T\phi D_\mu\alpha(\mu)v\,\mathrm{d}t$$
$$=0\quad\text{in }\Omega$$
$$\nabla\mu\cdot n=0\quad\text{on }\Gamma_{FS}$$
$$\frac{\nabla\mu}{|(\nabla\mu\cdot\nabla\mu+\varepsilon)^{1/2}|}\cdot n=\frac{1}{2}\int_0^T\sqrt{\frac{\rho}{\mu}}\lambda\frac{\partial u}{\partial t}\,\mathrm{d}t\quad\text{on }\Gamma_{AB}\tag{7}$$

where the operators $D_\mu\alpha(\mu)$ and $D_\mu\eta(\mu)$ are the first order derivatives of the functions $\alpha(\mu)$ and $\eta(\mu)$ with respect to μ, respectively. We note that in these operators, α and η are related to the reciprocal of the quality factor, Q^{-1}, following Paper I, through the following relationships:

$$\eta\left(Q^{-1}\right)=3.447\,Q^{-1}-0.041$$
$$\alpha\left(Q^{-1}\right)=3.529\,Q^{-1}+2.543$$

Here Q is defined in terms of shear wave velocity, V_S, as follows:

$$Q=a\arctan\left(\frac{V_S}{V_{S_{ref}}}\right)$$

Following Paper I, we use $V_{S_{ref}}=1000$ m/sec, and $a=25.5$ and invert for the shear modulus, μ.

1.2. Solution of the inverse problem

As presented in Paper I, to solve the optimality system (3)–(7), we use a reduced space approach in combination with a line search method. The search direction is computed with an inexact Gauss–Newton (GN) method. Before we describe the GN approximation, we write the Newton update for the full set of optimality conditions (3)–(7) (and therefore the full space of optimization variables):

$$\begin{bmatrix}\delta^2_{uu}\mathscr{L} & \delta^2_{uv}\mathscr{L} & \delta^2_{u\mu}\mathscr{L} & \delta^2_{u\lambda}\mathscr{L} & \delta^2_{u\phi}\mathscr{L}\\ \delta^2_{vu}\mathscr{L} & \delta^2_{vv}\mathscr{L} & \delta^2_{v\mu}\mathscr{L} & \delta^2_{v\lambda}\mathscr{L} & \delta^2_{v\phi}\mathscr{L}\\ \delta^2_{\mu u}\mathscr{L} & \delta^2_{\mu v}\mathscr{L} & \delta^2_{\mu\mu}\mathscr{L} & \delta^2_{\mu\lambda}\mathscr{L} & \delta^2_{\mu\phi}\mathscr{L}\\ \delta^2_{\lambda u}\mathscr{L} & \delta^2_{\lambda v}\mathscr{L} & \delta^2_{\lambda\mu}\mathscr{L} & \delta^2_{\lambda\lambda}\mathscr{L} & \delta^2_{\lambda\phi}\mathscr{L}\\ \delta^2_{\phi u}\mathscr{L} & \delta^2_{\phi v}\mathscr{L} & \delta^2_{\phi\mu}\mathscr{L} & \delta^2_{\phi\lambda}\mathscr{L} & \delta^2_{\phi\phi}\mathscr{L}\end{bmatrix}\begin{Bmatrix}\bar{u}\\ \bar{v}\\ \bar{\mu}\\ \bar{\lambda}\\ \bar{\phi}\end{Bmatrix}=-\begin{Bmatrix}\delta_u\mathscr{L}\\ \delta_v\mathscr{L}\\ \delta_\mu\mathscr{L}\\ \delta_\lambda\mathscr{L}\\ \delta_\phi\mathscr{L}\end{Bmatrix}\tag{8}$$

where the overbar indicates the Newton direction, the right-hand side is the negative of the residual of the KKT conditions. The operators in the above KKT system are defined by their actions on Newton directions as presented in detail in [15].

In a full space approach [18–20], one solves the system (8) simultaneously for $\bar{u}$, $\bar{v}$, $\bar{\mu}$, $\bar{\lambda}$, and $\bar{\phi}$. This approach requires storage of the state, adjoint, and inversion variables in space and time. For large scale problems, storing the time histories of the state and adjoint variables is not feasible. On the other hand, the reduced space method [21,22], in which the optimization problem is projected onto the inversion parameter space, does not require storage of the time-dependent variables. In addition, the regularized reduced Hessian (the Schur complement of $\delta^2_{\mu\mu}\mathscr{L}$ in the full Hessian in (8)) is better-conditioned than the full Hessian. Consequently, in this work we prefer to use the reduced space approach. At each Newton iteration, given a shear modulus estimate μ, we first solve for u and v from systems (3) and (4). With u and v known, we solve for λ and ϕ from Eqs. (5) and (6). Having satisfied the first and second KKT conditions, we now iteratively update the third KKT condition shown as Eq. (7). First, we calculate the reduced gradient $g_\mu=\delta_\mu\mathscr{L}$. Then, in order to determine the Newton update $\bar{\mu}$, we initially solve for $\bar{u}$ and $\bar{v}$ from the last two rows of Eq. (8), then from the first two rows, we solve for $\bar{\lambda}$ and $\bar{\phi}$. Now, we can solve for $\bar{\mu}$ from the middle row to update μ. This block elimination is equivalent in symbolic form to:

$$W_\mu\bar{\mu}=-g_\mu$$

where W_μ is the reduced Hessian [15]. To guarantee a descent search direction, the reduced Hessian must be positive-definite. This may not be the case for the reduced Newton Hessian, unless the solution estimate is close to the minimum. On the other hand, the Gauss–Newton (GN) Hessian is at least positive-semidefinite everywhere [23]. The GN method uses only first order information from the Hessian operator. For our problem, the GN approximation is equivalent to ignoring terms which depend on λ and ϕ in the KKT system. We note that as μ approaches its optimum value, the residual u^*-u approaches zero for noise-free data. Consequently, the GN Hessian approaches the Newton Hessian and the GN method asymptotically exhibits superlinear convergence.

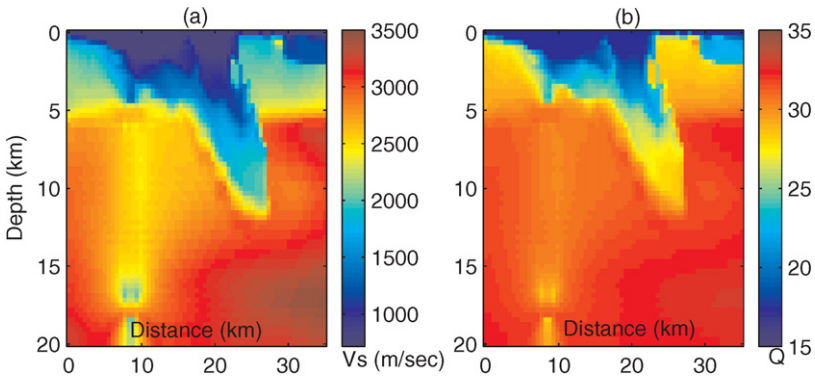

Fig. 1. (a) Target shear wave velocity profile, (b) target damping model.

Table 1
Source and model parameters.

Domain size (km×km)	35×20
Grid dimensions (km×km)	0.54×0.31
Time step Δt (sec)	0.06
Simulation time t_D (sec)	30
Minimum velocity (m/sec)	715
Maximum velocity (m/sec)	3470
Number of receivers	65
Regularization parameter, β	0.001

Due to storage and computational limitations, we cannot solve the reduced GN system with a direct method. In this work we solve the reduced GN system for the search direction with the conjugate gradient (CG) method, which does not require the Hessian operator but only the application of the Hessian operator to vectors. The convergence of the CG algorithm is related to the square root of the condition number of the reduced Hessian. In cases where the eigenvalue spectrum of the Hessian is clustered, it is possible to obtain a faster convergence. In addition, while solving the linear system we terminate CG early resulting in an inexact Newton method [15].

2. Sensitivity of the anelastic waveform inversion algorithm to several parameters

In this section, we perform experiments to study the effects of various parameters on the inversion quality and the algorithmic performance of the inversion technique. We use the same target profile as in Paper I, which is a cross section of the shear wave velocity distribution for a sedimentary soil medium in the LA basin obtained from the Southern California Earthquake Center (SCEC) Community Velocity Model [14]. Fig. 1 shows the target shear wave velocity distribution with the fault location marked and the attenuation profile.

As in Paper I, in our numerical experiments we use waveforms synthesized from the target profile on the free surface as pseudo-observed data. To model the source, we apply a uniform SH kinematic dislocation, with a triangular slip rate. The model parameters, such as the overall dimensions, time step, number of receivers, and regularization parameter, are listed in Table 1. Unless otherwise stated, we use these values in the parametric experiments.

We use a standard Galerkin type finite element scheme for space discretization and a central differences scheme for temporal discretization of the KKT system. The shear modulus μ is approximated as a piecewise bilinear function on each regular grid within the multilevel inversion algorithm, as are the state fields u and v on the 64×64 grid. We note that the wave propagation simulations are performed on a 64×64 grid.

We focus on the performance of the inverse wave propagation method and study the effects of various parameter and algorithmic choices including the type of regularization functional, value of the regularization parameter, receiver density, preconditioning, data noise, and multilevel continuation. We use the term “inversion accuracy” to describe the success of the algorithm in reconstructing the target wave velocity. We use the term “algorithmic performance” to signify such issues as the rate of convergence and algorithmic scalability.

The density of the material is taken as known, since its variability is typically much less pronounced than that of the other system parameters. Thus, even though our inversions are in terms of the shear modulus, μ, we express our results in terms of the shear wave velocity, V_S, which can be readily related to the shear modulus once the density is assumed to be known.

2.1. Effect of regularization type on inversion accuracy

To examine the effect of the form of the regularization on the inversion accuracy, we perform inversions with Tikhonov and total variation regularizations keeping the remaining model parameters fixed. The regularization term in both Tikhonov and total variation regularization penalizes high wavenumber material property components. These two methods differ in the type of penalty function they employ, as briefly discussed below.

2.1.1. Tikhonov regularization (TR)

The form of Tikhonov regularization we use is H^1, i.e. a quadratic penalty on the gradient of the material property function:

$$R_{TR}(\mu) = \frac{1}{2}\int_{\Omega} \nabla\mu \cdot \nabla\mu \,\mathrm{d}\Omega \tag{9}$$

This form of regularization penalizes high wavenumber components of the medium in all directions. For a smooth target medium, Tikhonov H^1 regularization serves to filter out rough components of the medium that are unobservable. Besides, an appropriate choice of β is crucial for the satisfactory performance of the regularization.

The TR functional above becomes unbounded at points of discontinuities in the medium, and thus smoothes material interfaces unnecessarily. Total variation regularization, which we discuss next, performs better than TR in media with interface discontinuities, which is important for a layered medium.

2.1.2. Total variation regularization (TVR)

This is the regularization method used in Paper I. In its continuous form, the TVR functional is as follows:

$$R_{TVR}(\mu) = \frac{1}{2}\int_{\Omega} \sqrt{\nabla\mu \cdot \nabla\mu}\,\mathrm{d}\Omega \tag{10}$$

The above expression is not differentiable when $\nabla\mu = 0$. Our optimization method is a Newton method, which requires twice-differentiable objective functions; thus we make use of a variant of TVR:

$$R_{TVR}(\mu) = \frac{1}{2}\int_{\Omega} \sqrt{\nabla\mu \cdot \nabla\mu + \varepsilon}\,\mathrm{d}\Omega \tag{11}$$

The ε term makes the above functional differentiable everywhere.

Unlike TR, TVR is finite at discontinuities, and thus performs better at recovering interface discontinuities in a target medium [24,25]. Despite the expected improvement, the performance still depends on the value of the regularization parameter. In the next section, we discuss the regularization parameter selection and show inversion results with different parameters.

We show the inverted shear wave velocity profiles for consecutive stages of the multilevel algorithm with two different regularization techniques in Figs. 2 and 3. In the multilevel inversion procedure, where we perform the inversion on successively refined grids, as seen from the figures, the material model is initially taken to be homogeneous, i.e., we assume nothing is known about the target. Thus, the optimization grid starts with a single finite element of constant value, and becomes uniformly finer at each stage of the multilevel process until it matches the 64×64 wave propagation grid. As observed from the figures, the TR smoothes out the inverted profile at the material interfaces. TVR is more successful at localizing interfaces. Therefore, we use TVR in our subsequent examples as the regularization functional.

2.2. Effect of regularization parameter on inversion accuracy and algorithmic performance

The application of regularization necessitates the choice of an appropriate regularization parameter to eliminate oscillatory components of the medium that are unobservable from the band-limited response. If the parameter is too large, the recovered medium is over smoothed. On the other hand, if it is too small, the recovered medium is polluted with highly oscillatory components. Most of the existing work on regularization methods and regularization parameter selection has focused on linear problems. In this work we simply use a trial-and-error approach to tune the regularization parameter. This trial-and-error process is possible only for synthetic problems where the target material property profile is known beforehand. On the other hand, for problems with real data, the selection of the regularization parameter must be automated via alternative techniques. When the noise level of the observations is known, the discrepancy principle may be used to estimate the regularization parameter that yields an observational misfit of the order of the noise. When the noise level is not known, several methods are available, such as Generalized Cross Validation and the L-curve method [25,26]. In general, choosing the optimal regularization parameter requires solution of a sequence of inverse problems. See [25] for a discussion on regularization parameter selection methods.

As a part of this parametric study, we investigate the effects of the regularization parameter on algorithmic performance and inversion accuracy. We solve the example problem using TVR with three different regularization parameters.

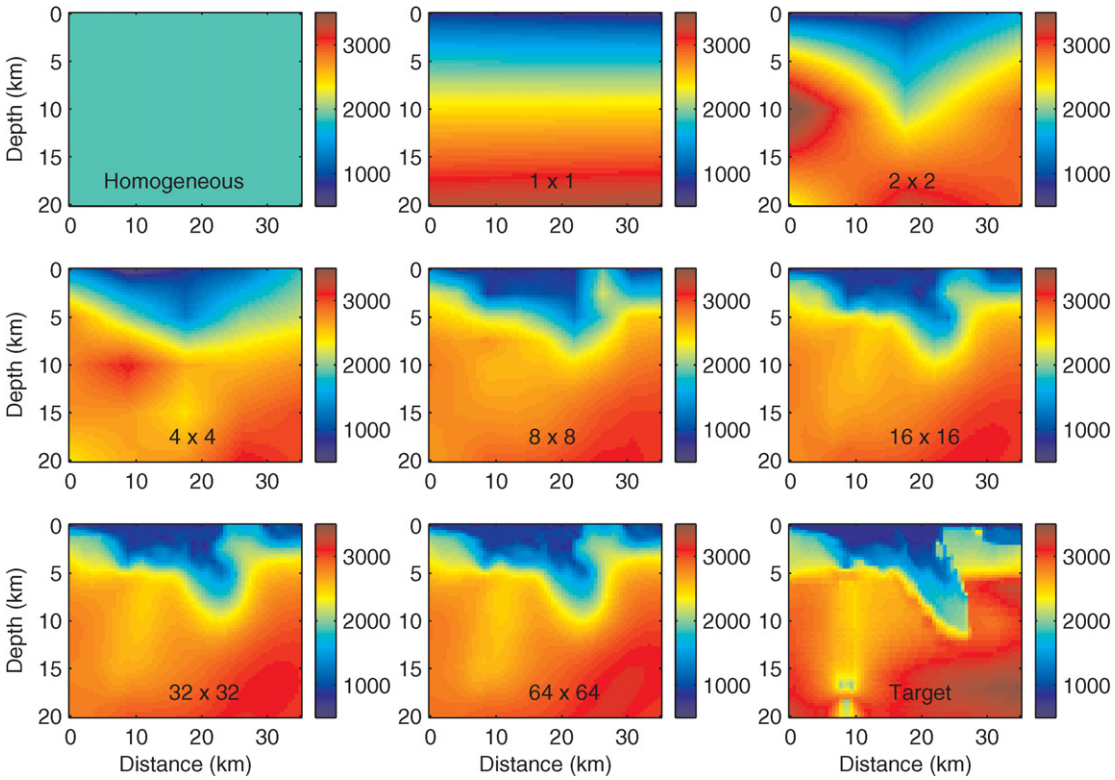

Fig. 2. Inverted shear velocity profiles for the consecutive stages of the multilevel inversion algorithm with Tikhonov regularization ($\beta = 0.5 \times 10^{-3}$) ($V_S$ in m/sec).

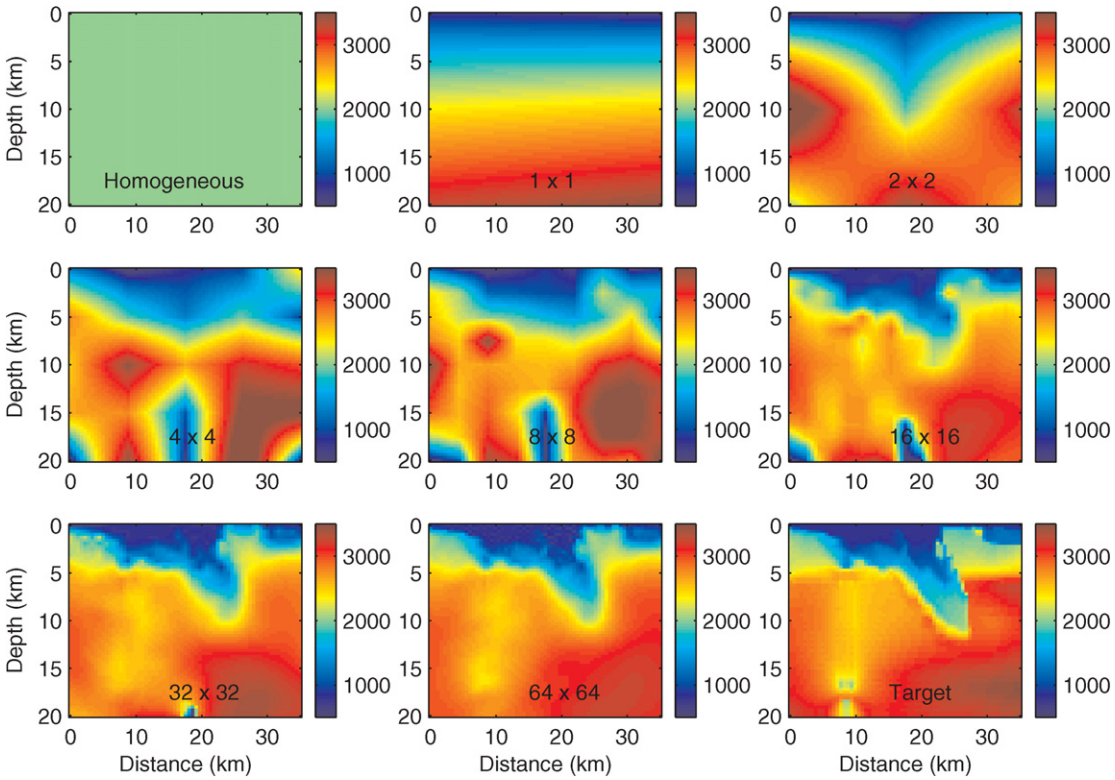

Fig. 3. Inverted shear velocity profiles for the consecutive stages of the multilevel inversion algorithm with total variation regularization ($\beta = 10^{-3}$) (V_S in m/sec).

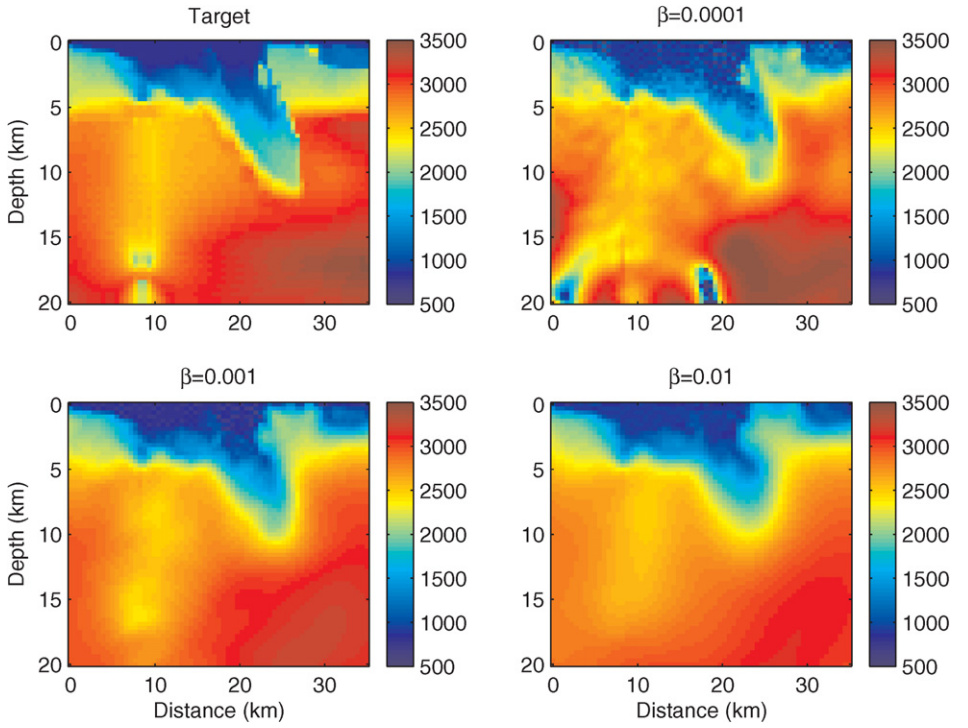

Fig. 4. Effect of different regularization constants on the algorithmic performance (V_S in m/sec, grid size: 64×64).

Table 2
Number of CG and GN iterations for different regularization constants for TVR regularization.

Grid	Total number of CG iterations (number of GN iterations)		
	$\beta = 0.0001$	$\beta = 0.001$	$\beta = 0.01$
Homogeneous	59 (59)	59 (59)	59 (59)
1×1	43 (29)	41 (29)	38 (24)
2×2	63 (36)	58 (33)	61 (36)
4×4	185 (48)	151 (59)	69 (15)
8×8	463 (39)	385 (67)	137 (61)
16×16	501 (56)	262 (53)	99 (13)
32×32	561 (87)	213 (34)	155 (13)
64×64	–	196 (27)	79 (7)

Fig. 4 shows the results with the following regularization parameters: $\beta = 10^{-4}$, $\beta = 10^{-3}$, and $\beta = 10^{-2}$. Table 2 compares the performance of the inversion algorithm in terms of the number of GN and CG iterations as functions of β. For all of the optimization grids, we terminate the algorithm after a maximum of 100 GN iterations. We note that the inversion algorithm within the finest grid with the smallest regularization parameter did not converge after 100 GN iterations.

In Fig. 4, we observe that when β is too large (10^{-2}), regularization over-smoothes the reconstructed profile; whereas when β is too small (10^{-4}), noisy artifacts appear. This observation confirms the need for a problem-specific regularization parameter. As listed in Table 2, the number of CG iterations decreases with increasing β. This is expected due to the decreasing condition number of the Hessian with increasing values of the regularization parameter. On the other hand, as indicated, beyond a certain value, the accuracy decreases with increasing β. After several trials with different values of β, we found that the optimal value is 10^{-3}. We also note that the selected value of the ε term in Eq. (11) is taken to be 10^{-3}; the value of ε does not have as significant effect on the results as the value of β.

2.3. Effect of solution technique on algorithmic performance

To test the effect of the type of optimization method on the speed of convergence, we compute the search direction with the Gauss–Newton (GN) and the steepest descent (SD) methods and compare the number of iterations required for convergence for the two cases. SD is the simplest of the gradient methods, in which the search direction is the negative of the gradient; this is also the initial step of the CG algorithm. Table 3 displays the numerical performance of the inversion

Table 3
Effect of solution technique on numerical performance ($\beta_{TVR} = 10^{-3}$, $\beta_{TR} = 0.5 \times 10^{-3}$).

	Total number of iterations	
	Steepest descent	Gauss–Newton
TR	1000+	124
TVR	1000+	196

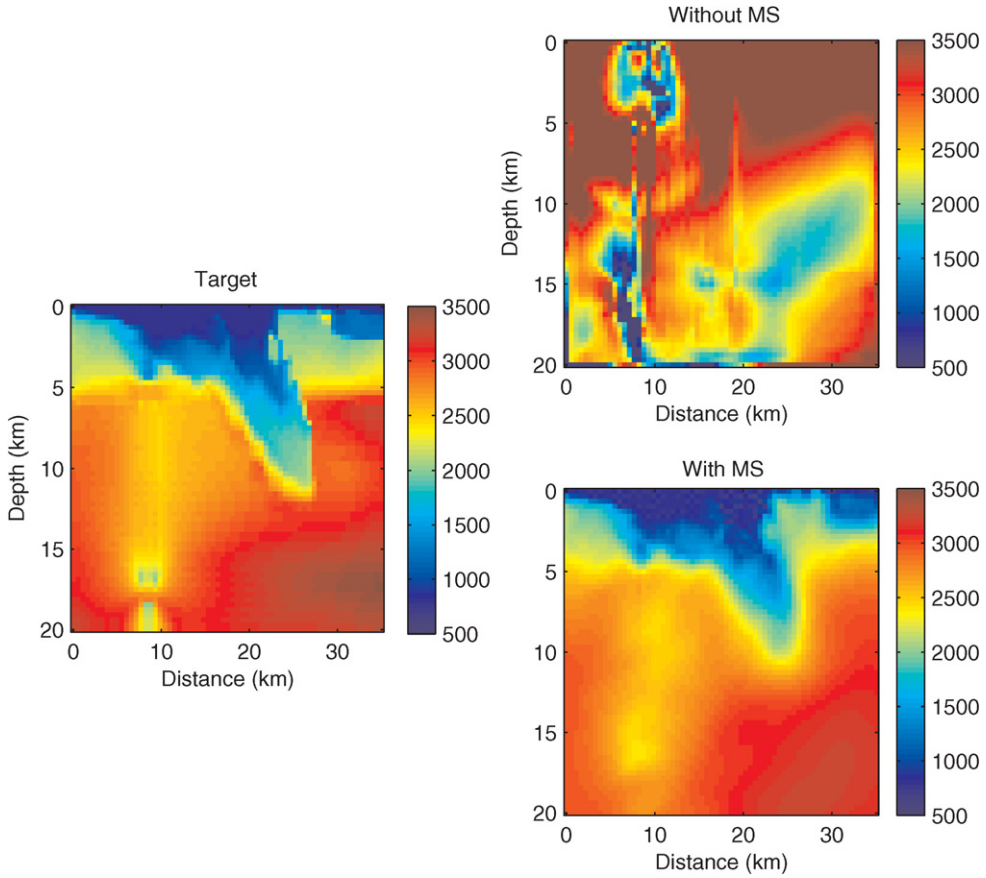

Fig. 5. Effect of multilevel technique on inversion accuracy ($\beta = 0.001$, grid size: 64×64).

algorithm in terms of the total number of iterations for the SD and GN methods for two different types of regularization. The iteration numbers displayed for the GN method are the total number of CG iterations within the final grid of the multilevel algorithm (grid size: 64×64).

We also note that GN here is not exact, i.e. we terminate early the CG solution at each GN iteration to avoid "over-solving", in the style of an inexact Newton-CG method [27].

Regardless of the regularization technique, SD is unable to converge within 1000 iterations whereas GN converges in every case. The slow convergence of SD is apparent particularly for badly scaled systems in which the Hessian matrix is ill-conditioned. Consequently, we prefer to use GN to compute the search direction. When we compare the number of iterations for TR and TVR in conjunction with the GN method, we observe that the TVR converges more slowly than the TR. This is mainly because the same properties of the TVR functional that help to recover the sharp edges of the target profile lead to an anisotropic and more ill-conditioned Hessian than that of TR, resulting in a larger number of iterations. However, since the number of iterations is of the same order for both methods; we, prefer to use the TVR with the GN method for the rest of this work due to its higher accuracy. (A popular alternative to GN method is the Levenberg–Marquardt (LM) method. However, based on numerical experience with similar inverse problems, we note that LM method is more suitable for problems where the GN approximation is not sufficiently accurate and the final misfit is large, which is not the case in the problem presented herein.)

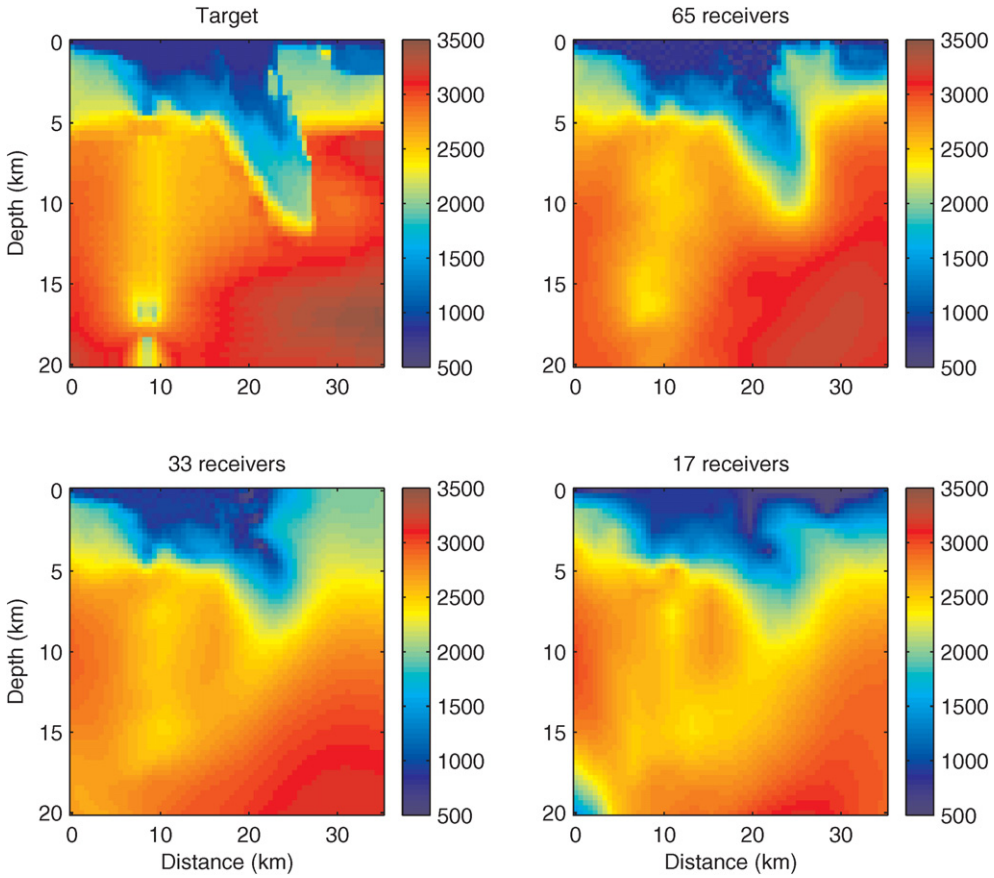

Fig. 6. Effect of receiver density on inversion accuracy ($\beta = 0.001$, grid size: 64×64).

2.4. Effect of local minima and multilevel technique on inversion accuracy

Given an initial guess, an inexact Newton–CG method with appropriate line search is guaranteed to converge at best to a local minimum. When the initial guess is not in the basin of attraction of the global minimum, the existence of multiple local minima is an issue for seismic inversion methods. Since the diameter of the attraction basin for inverse wave propagation problems narrows down with increasing frequency of propagating waves [28], the optimizer should be provided with a good initial guess ito avoid getting trapped into a local minimum. In this work, a multilevel approach is used to generate a sequence of initial guesses that remain within the basin of attraction of the global minimum by means of a strategy similar to that employed by Bunks et al. [28].

Observations on the basin of attraction of the global minimum suggests that inversion be performed initially on a coarse grid with low frequency sources, and the resulting optimal solution used to initialize new inversions on successively refined grids with sources of increasingly higher frequencies. For simplicity, in this work we make use of spatial grid continuation with a single source. We begin by solving the inverse problem on a coarse material grid and prolong the solution to the next finer grid, continuing the refinement until the material grid matches the wave propagation grid. The converged solution on each grid is used as an initial guess for the next scale. In our experience, this multilevel grid continuation strategy keeps the sequence of minimizers within the basin of attraction of the global minimum, provided the seismic source contains sufficient low-frequency energy [21,22].

To examine its effects on the inversion accuracy, we solve the inverse problem with and without the multilevel technique. Fig. 5 compares the target shear wave velocity profile with the inverted profiles obtained from the two cases. As observed in Fig. 5, inversion without the multilevel strategy becomes trapped in a local minimum and fails to recover the target velocity profile. On the other hand, use of the multilevel strategy is observed to be successful in recovering the target velocity profile.

2.5. Effect of receiver density on inversion accuracy

To evaluate the influence of receiver density, we compare shear wave velocity profiles obtained from inversions performed with three different numbers of receivers as shown in Fig. 6. As expected, we recover the target profile more effectively with increasing numbers of receivers, and thus more information regarding the response. Another observation is

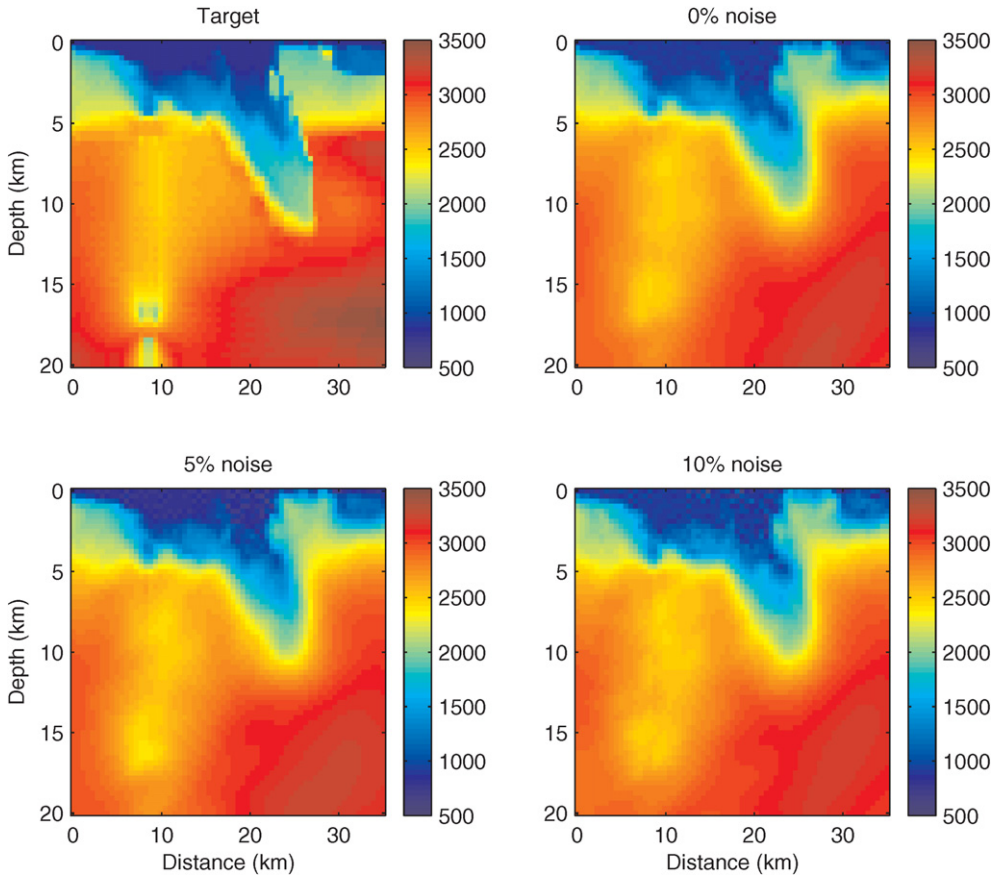

Fig. 7. Effect of noisy data on inversion accuracy ($\beta = 0.001$, grid size: 64×64).

Table 4
Misfit variation for different noise levels.

Noise	$\frac{1}{2}\|u^* - u^*_{\text{noisy}}\|$	Initial misfit	Initial $R_{TVR}(\mu)$	Final misfit	Final $R_{TVR}(\mu)$
0%	0.0	498.64	0.41	0.19	53.78
5%	0.26	562.15	0.41	1.08	53.25
10%	1.04	504.67	0.41	4.02	53.24

that even with 17 receivers we have managed to recover the target profile roughly. This will be beneficial for problems with sparse data.

2.6. Effect of noisy data on inversion accuracy

To account for the uncertainties in recorded response, we add Gaussian noise with zero mean and unit standard deviation to the synthetic waveforms at the receivers u^* (i.e. $u^*_{\text{noisy}} = u^* \times (1 + n \times L)$ where n is the noise generated and L is the percent noise level). To study the effects of noisy data on the quality of the inversion, we solve the example problem using three different levels of noise (0, 5, and 10 percent) added to the original synthetic seismograms. Fig. 7 compares the converged solution on the finest grid within the multilevel algorithm for different noise levels. Table 4 lists the initial and final values of the objective function in terms of both the misfit and size of the regularization term along with the norm of the noise for each case. The initial misfit corresponds to the misfit for the homogeneous initial guess and the final misfit corresponds to the misfit for the final grid in the multilevel algorithm. As can be seen, there is an increase in the final misfit with increasing noise level but this misfit remains below 1 percent. It is noteworthy that noisy data does not increase the number of GN and CG iterations noticeably.

2.7. Effect of preconditioner on algorithmic performance

For an ill-conditioned Hessian, a preconditioner may be applied to the Hessian for faster convergence. The preconditioner approximates the inverse of the Hessian and thus the resulting matrix has a more favorable eigenvalue spectrum, yielding

Table 5
Effect of preconditioning on the number of CG iterations.

Grid	Total number of CG iterations	
	No preconditioner	LBFGS preconditioner
Homogeneous	59	59
1×1	67	41
2×2	386	58
4×4	1080	151
8×8	921	385
16×16	764	262
32×32	576	213
64×64	1563	196

a smaller condition number or clustered eigenvalues. In this work, we examine the effect of a limited memory quasi-Newton (LBFGS) preconditioner [29] by solving the example problem with and without the use of the preconditioner. Table 5 compares the numerical performance of the inversion algorithm for the two cases in terms of the total number of CG iterations. In both cases, we use TVR with a regularization parameter $\beta = 10^{-3}$.

The speed of convergence of CG is greatly accelerated with the use of a preconditioner. For the finest scale in the multilevel algorithm, the number of CG iterations decreases by a factor of 8. We note that in addition to the decrease in the number of CG iterations, the preconditioner provides a reduction in the computing time by a factor of 4 for the entire solution including all of the optimization grids. Epanomeritakis et al. [30] observed a similar algorithmic scalability with the same type of preconditioner employed in a large-scale three-dimensional elastic full-waveform seismic inversion algorithm.

3. Concluding remarks

In this article, we studied the effects of a number of parameters on the quality of inversion and algorithmic performance of the inversion method introduced in Paper I. We applied the method to a two-dimensional example problem, and examined the inversion results for various parameters and algorithmic choices. The highlights are as follows:

- Total variation regularization is more successful at localizing sharp material interfaces than Tikhonov regularization, which over-smoothes the sharp edges.
- The choice of an optimum regularization parameter is important for attaining a successful inversion. When the regularization parameter is too small, the solution is polluted with noisy artifacts; whereas when it is too large, the inverted result is over-smoothed. Every inverse problem requires a problem-specific regularization parameter. In this study, our goal was to illustrate numerically the sensitivity of the solution to the regularization parameter. To this end, we used a trial and error technique to determine the corresponding value of this parameter that leads to the smallest differences between the inverted shear wave velocity profile and the target one.
- By computing search directions with both SD and GN techniques, we found the performance of the GN method to be superior to that of the SD, due to its approximation of the curvature of the objective. SD did not converge even after an unrealistically large number of iterations.
- The use of a multilevel strategy in seismic waveform inversion is crucial. We observed that without the multilevel technique, the optimizer becomes trapped at a local minimum, and consequently the algorithm is not successful in recovering the target medium.
- From tests with different receiver densities, we found that, as expected, the larger the number of receivers, the more effective the inversion is at retrieving the target wave velocity profile. On the other hand, even with sparser receiver arrays, we still obtained reasonable estimates of the shear wave velocity profile. This observation is encouraging for future attempts at waveform inversion using real data where a dense receiver array might not be available.
- Comparing inversions with and without noise, we observed that noise up to 10 percent does not have a significant effect on either the quality of inversion or the algorithmic performance of the inversion method.
- The use of an LBFGS-based preconditioner improved the performance of the algorithm significantly by speeding up the convergence of CG. It is essential to use a preconditioner for waveform inversion problems characterized by ill-conditioned Hessians, and for problems with large numbers of inversion parameters.

In conclusion, in this study, we presented numerical evidence that a careful selection of inversion parameters and application of appropriate remedies for the challenges encountered, yields a full waveform anelastic inversion methodology that performs successfully for a given target shear wave velocity distribution.

Acknowledgements

This research was funded by the National Science Foundation under grant ITR 0326449, Steve Meacham is the cognizant officer. We are grateful for this support. This work is also done in conjunction with Pathway 4 of the Southern California

Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. The SCEC contribution number for this article is 1132.

References

[1] A. Tarantola, Inversion of seismic reflection data in the acoustic approximation, Geophysics 49 (1984) 1259–1266.
[2] A. Tarantola, Inverse Problem Theory and Model Parameter Estimation, SIAM, Philadelphia, PA, 2005.
[3] A. Tarantola, B. Valette, Generalized nonlinear inverse problems solved using the least squares criterion, Rev. Geophys. Space Phys. 20 (1982) 219–232.
[4] J. Tromp, C. Tape, Q. Liu, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, Geophys. J. Int. 160 (2005) 195–216.
[5] P. Chen, L. Zhao, T.H. Jordan, Full 3D tomography for crustal structure of the Los Angeles region, Bull. Seism. Soc. Am. 97 (2007) 1094–1120.
[6] R.G. Pratt, M.H. Worthington, Inverse theory applied to multi-source cross-hole tomography. Part I: Acoustic wave-equation method, Geophys. Prospect. 38 (1990) 287–310.
[7] L. Sirgue, R.G. Pratt, Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, Geophysics 69 (2004) 231–248.
[8] S. Operto, C. Ravaut, L. Improta, J. Virieux, A. Herrero, P. Dell'Aversana, Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study, Geophys. Prospect. 52 (2004) 625–651.
[9] C. Ravaut, S. Operto, L. Improta, J. Virieux, A. Herrero, P. Dell'Aversana, Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt, Geophys. J. Int. 159 (2004) 1032–1056.
[10] A. Tarantola, Theoretical background for the inversion of seismic waveforms including elasticity and attenuation, Pure Appl. Geophys. 128 (1988) 365–399.
[11] G.J. Rix, C.G. Lai, A.W. Spang, In situ measurements of damping ratio using surface waves, J. Geotech. Geoenviron. Engng. 126 (2000) 472–480.
[12] G. Hicks, R.G. Pratt, Reflection waveform inversion using local descent methods: estimating attenuation and velocity over a gas-sand deposit, Geophysics 66 (2001) 598–612.
[13] A. Askan, V. Akcelik, J. Bielak, O. Ghattas, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures, Bull. Seism. Soc. Am. 97 (6) (2007) 1990–2008.
[14] H. Magistrale, S. Day, R.W. Clayton, R. Graves, The SCEC Southern California reference three-dimensional seismic velocity model, Version 2, Bull. Seism. Soc. Am. 90 (2000) S65–S76.
[15] A. Askan, Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures, PhD thesis, Carnegie Mellon University, Pittsburgh, PA, 2006.
[16] S.M. Day, C.R. Bradley, Memory efficient simulation of anelastic wave propagation, Bull. Seism. Soc. Am. 91 (2001) 520–531.
[17] P. Moczo, J.O.A. Robertsson, L. Eisner, The finite-difference time-domain method for modelling of seismic wave propagation, in: Ru-Shan Wu, Valerie Maupin (Eds.), Advances in Wave Propagation in Heterogeneous Earth, in: Advances in Geophysics, vol. 48, Academic Press, New York, MA, 2007, pp. 421–516.
[18] G. Biros, O. Ghattas, Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part I: The Krylov–Schur solver, SIAM J. Sci. Comput. 27 (2) (2005) 687–713.
[19] G. Biros, O. Ghattas, Parallel Lagrange–Newton–Krylov–Schur methods for PDE-constrained optimization. Part II: The Lagrange–Newton solver, and its application to optimal control of steady viscous flows, SIAM J. Sci. Comput. 27 (2) (2005) 714–739.
[20] V. Akcelik, G. Biros, O. Ghattas, J. Hill, D. Keyes, B. van Bloemen Waanders, Parallel algorithms for PDE-constrained optimization, in: M. Heroux, P. Raghaven, H. Simon (Eds.), Frontiers of Parallel Computing, SIAM, 2006.
[21] V. Akcelik, G. Biros, O. Ghattas, Parallel multiscale Gauss–Newton–Krylov methods for inverse wave propagation, in: Proceedings of ACM/IEEE SC2002, 2002.
[22] V. Akcelik, J. Bielak, G. Biros, I. Epanomeritakis, A. Fernandez, O. Ghattas, E.J. Kim, J. Lopez, D. O'Hallaron, T. Tu, J. Urbanic, High-resolution forward and inverse earthquake modeling on terascale computers, in: Proceedings of ACM/IEEE SC2003, 2003.
[23] J. Nocedal, S.J. Wright, Numerical Optimization, Springer, New York, 1999.
[24] R. Acar, R.C. Vogel, Analysis of bounded variation penalty methods for ill-posed problems, Inverse Problems 10 (1994) 1217–1229.
[25] C. Vogel, Computational Methods for Inverse Problems, SIAM, 2002.
[26] P. Craven, G. Wahba, Smoothing noisy data with spline functions, Numer. Math. 31 (1979) 377–403.
[27] S.C. Eisenstat, H.F. Walker, Choosing the forcing terms in an inexact Newton method, SIAM J. Sci. Comput. 17 (1996) 16–32.
[28] C. Bunks, F.M. Saleck, S. Zaleski, G. Chavent, Multiscale seismic waveform inversion, Geophysics 50 (1995) 1457–1473.
[29] J.L. Morales, J. Nocedal, Automatic preconditioning by limited memory quasi-Newton updating, Siam J. Optim. 10 (4) (2000) 1079–1096.
[30] I. Epanomeritakis, V. Akcelik, O. Ghattas, J. Bielak, A Newton–CG method for large-scale three-dimensional elastic full-waveform seismic inversion, Inverse Problems 24 (2008) 034015, 26 pp.