

Algorithmic strategies for full waveform inversion: 1D experiments

Carsten Burstedde¹ and Omar Ghattas²

ABSTRACT

Full-waveform seismic inversion, i.e., the iterative minimization of the misfit between observed seismic data and synthetic data obtained by a numerical solution of the wave equation provides a systematic, flexible, general mechanism for reconstructing earth models from observed ground motion. However, many difficulties arise for highly resolved models and the associated large-dimensional parameter spaces and high-frequency sources. First, the least-squares data-misfit functional suffers from spurious local minima, which necessitates an accurate initial guess for the smooth background model. Second, total variation regularization methods that are used to resolve sharp interfaces create significant numerical difficulties because of their nonlinearity and near-degeneracy. Third, bound constraints on continuous model parameters present considerable difficulty for commonly used active-set or interior-point methods for inequality constraints because of the infinite-dimensional nature of the parameters. Finally, common gradient-based optimization methods have difficulties scaling to the many model parameters that result when the continuous parameter fields are discretized. We have developed an optimization strategy that incorporates several techniques address these four difficulties, including grid, frequency, and time-window continuation; primal-dual methods for treating bound inequality constraints and total variation regularization; and inexact matrix-free Newton-Krylov optimization. Using this approach, several computations were performed effectively for a 1D setting with synthetic observations.

INTRODUCTION

We consider full-waveform seismic inversion as a strategy for the iterative minimization of the misfit between observed seismic data

and synthetic data, where the synthetic data are obtained by a numerical solution of the wave equation. Arbitrary forward-modeling operators can be used, so full-waveform inversion permits incorporation of the full physics of seismic wave propagation, including effects from anisotropy and attenuation. The resolution in frequency and space can be very high. The inversion algorithm is independent of the dimensionality of the problem; so 1D, 2D, and 3D problems can be accommodated within the same framework. The inversion methodology is also independent of the way the data are selected or preprocessed. Thus, full-waveform seismic inversion provides a systematic, flexible, and general mechanism for reconstructing earth models from observed ground motion.

Full-waveform inversion can be formulated as a nonlinear optimization problem with an objective functional consisting of the difference between observed and synthetic seismic waveforms in a suitable norm, usually L_2 . This choice is also called the output least-squares formulation. A regularization term acting on the earth-model parameters is typically added to the objective functional to penalize variations in the model that are invisible to the data misfit. The optimization is constrained by the governing forward problem (e.g., elastic wave propagation) and by lower and upper bounds on the model parameters.

In practice, however, solving full-waveform seismic inversion problems using nonlinear optimization methods remains a challenge, particularly for high-frequency problems, precluding routine use. This is mostly because of mathematical and numerical difficulties, which include the following.

Multiple minima. — The least-squares data misfit objective functional is oscillatory in directions associated with wavenumber components of the model that are longer than seismic wavelengths, i.e., the smooth components (Santosa and Symes, 1989). The common practice of grid-and-frequency continuation (Kolb et al., 1986; Bunks et al., 1995) can be effective at avoiding local minima, but only when the observations contain sufficient low-frequency content to initialize the continuation process.

Manuscript received by the Editor 31 December 2008; revised manuscript received 3 May 2009; published online 3 December 2009; corrected version published online 17 December 2009.

¹The University of Texas at Austin, Institute for Computational Engineering and Sciences (ICES), Austin, Texas, U.S.A. E-mail: carsten@ices.utexas.edu.

²The University of Texas at Austin, Jackson School of Geosciences, Department of Mechanical Engineering, and Institute for Computational Engineering and Sciences (ICES), Austin, Texas, U.S.A. E-mail: omar@ices.utexas.edu.

© 2009 Society of Exploration Geophysicists. All rights reserved.

Ill-posedness. — The objective functional is relatively insensitive to wavenumber components of the model that are shorter than seismic wavelengths (i.e., the rough components). The eigenvalues of the linearized forward operator corresponding to these components are nearly zero. Tikhonov regularization (Tikhonov and Arsenin, 1977), which penalizes oscillatory components, addresses this ill-posedness but blurs interfaces of layered media. Total variation (TV) regularization (Rudin et al., 1992) is very effective at resolving sharp interfaces but leads to substantially more ill-conditioned inverse operators.

Variational inequalities. — When infinite-dimensional inequalities associated with bound constraints on continuous model parameters are discretized on fine grids, very large sets of inequality constraints result. Active set methods (Nocedal and Wright, 1999) keep track of the subset of parameters that attains the bounds and solves the inverse problem only on the parameters away from the bounds (i.e., the inactive set). These methods can suffer from combinatorial complexity and can exhibit increased computing costs as the grid is refined. Interior-point methods (Conn et al., 2000) often have better scaling properties, but good preconditioners for solving the resulting optimal systems remain elusive.

Scalable optimization algorithms. — Severe nonlinearities associated with waveform data-misfit objective functionals render the least-squares optimization problem difficult to solve. The use of gradient methods such as nonlinear conjugate gradients (Polak and Ribière, 1969), in combination with adjoint computation of the gradient, results in a method that is cheap to apply at each iteration (Chavent, 1974; Tarantola, 1986; Plessix, 2006) yet is only linearly convergent (Nocedal and Wright, 1999). On the other hand, Newton's method is quadratically convergent (Deuffhard, 2004) but requires solving a dense system of equations at each iteration (Pratt et al., 1998), which can be intractable for fine discretizations of continuous model parameters.

We incorporate several techniques that address these difficulties. To help mitigate multiple minima, we use continuation on the grid resolution, on the frequency content, and on the length of the time window (Kolb et al., 1986; Bunks et al., 1995; Sirgue and Pratt, 2004). We appeal to work from image processing on primal-dual TV regularization (Chan et al., 1999) and from optimal control theory on primal-dual active set strategies for bound inequalities (Hintermüller et al., 2003), both of which demonstrate superior numerical performance relative to conventional methods. Finally, we use an exact matrix-free Newton-Krylov method (Santosa and Symes, 1988, 1989) that combines the curvature-exploiting properties of Newton's method with a computational structure resembling a gradient-only method. These methods have demonstrated grid-independent convergence for 2D and 3D highly resolved inverse wave-propagation problems (Akcelik et al., 2002, 2003; Epanomeritakis et al., 2008).

The 1D seismic inverse problem has been studied extensively by Bamberger et al. (1979), Bube and Burridge (1983), Santosa and Symes (1985), and Sacks and Santosa (1987), among others. Unique global solutions exist for noise-free data and Dirac-delta-shaped impulses. In the band-limited case, multiple (at least weak) local minima exist, although the difficulties can be addressed using specialized techniques. On the other hand, multidimensional problems are notoriously difficult to solve numerically because of spurious local

minima.

Although it is certainly possible and more efficient to take advantage of the special properties of the 1D inverse problem, such as reparameterization in traveltime, our objective is to assess inversion algorithms with regularization that apply to a more general class of inverse problems. We choose band-limited 1D experiments because they are simple enough to illustrate the general procedure yet exhibit many of the difficulties prevalent in larger 2D and 3D inversions.

We present numerical experiments based on velocity models derived from borehole measurements. All data are created synthetically and noise free. Our results illustrate that the combination of the algorithms mentioned above yields a scalable, robust method. We conclude by solving a 1D inverse problem that spans a depth of about 100 wavelengths.

FULL-WAVEFORM INVERSION FRAMEWORK

Nonlinear least-squares optimization formulation

The wave equation is a second-order partial-differential equation with a solution in an infinite-dimensional space. Likewise, the density and elasticity parameters can be understood as members of infinite-dimensional spaces. Thus, in this fully continuous setting, a least-squares optimization problem can be posed and the first- and second-order necessary and sufficient conditions for its solution can be derived. Discretizing these continuous optimality conditions and solving them numerically is often referred to as the optimize-then-discretize (OTD) approach. Conversely, discretizing the forward modeling operator and objective functional and afterward deriving all optimality conditions in the discrete setting is often called discretize then optimize (DTO). With appropriate discretization, both approaches converge to the same solution as the grid size approaches zero, yet only DTO ensures consistency of all discrete derivatives (Gunzburger, 2003). We therefore formulate the full-waveform inverse problem as a discrete constrained optimization problem of the following form:

$$\min_{\mathbf{u}, \boldsymbol{\mu}} J(\mathbf{u}, \boldsymbol{\mu}) \text{ where}$$

$$J(\mathbf{u}, \boldsymbol{\mu}) := F(\mathbf{u}) + \alpha R(\boldsymbol{\mu}), \text{ such that } \mathbf{A}_{\boldsymbol{\mu}} \mathbf{u} = \mathbf{f}. \quad (1)$$

Let $u(x, t)$ be the wavefield in space $x \in \Omega$, where $\Omega \subset \mathbb{R}^d$ is the computational domain. The time interval considered is $t \in (0, T)$. Let $f(x, t)$ be the force corresponding to one or more seismic sources, related by the (partial differential) forward operator $\mathcal{A}u = f$. The discretizations of u and f are denoted by $\mathbf{u} \in \mathbb{R}^{KN_w}$ and $\mathbf{f} \in \mathbb{R}^{KN_w}$, where N_w is generally a multiple of the number of space grid points and K the number of discrete time values. The particular (finite-difference or finite-element) discretization used should of course be suited to the forward modeling operator, whose discretized form reads $\mathbf{A}\mathbf{u} = \mathbf{f}$. The nonsymmetric matrix $\mathbf{A} \equiv \mathbf{A}_{\boldsymbol{\mu}}$ is assumed to depend (nonlinearly) on a number N_p of model parameters $\boldsymbol{\mu} \in \mathbb{R}^{N_p}$. These can include density, seismic velocities, elastic moduli, attenuation parameters, and so on.

The observation term $F(\mathbf{u})$ defines the measure of similarity between the synthetics \mathbf{u} and the data \mathbf{u}^* . For output least squares, this is a numerical quadrature formula approximating the integral

$$F_{\text{ls}}(\mathbf{u}) \approx \frac{1}{2} \sum_r \int_0^T \|u^*(x_r, t) - u(x_r, t)\|^2 dt, \quad (2)$$

where x_r are the receiver locations.

The second contribution $\alpha R(\boldsymbol{\mu})$ is the regularization term with a (small) weight factor $\alpha > 0$. In this article, we examine two possibilities — Tikhonov and TV — which are derived from the following equations in the case of a scalar elastic parameter μ :

$$R_2(\mu) = \frac{1}{2} \int_{\Omega} \|\nabla \mu\|^2 dx \text{ for Tikhonov}, \quad (3)$$

$$R_{1,\epsilon}(\mu) = \int_{\Omega} \sqrt{\|\nabla \mu\|^2 + \epsilon} dx \text{ for TV}. \quad (4)$$

Expression 3 generally smooths the model, whereas expression 4 allows jumps yet is more difficult to handle numerically.

Gradient-based optimization

We derive the adjoint equation based on the Lagrangian formulation of the objective functional from equation 1. The general procedure is equivalent to the approach described in Plessix (2006). As mentioned, we consider all variables to be finite-dimensional discretizations of the original physical quantities.

To derive the optimality conditions for the constrained minimization problem (equation 1), we define a Lagrange multiplier \mathbf{p} and form the Lagrangian functional:

$$\mathcal{L}(\mathbf{p}, \mathbf{u}, \boldsymbol{\mu}) := J(\mathbf{u}, \boldsymbol{\mu}) + \mathbf{p}^T (\mathbf{A}_{\boldsymbol{\mu}} \mathbf{u} - \mathbf{f}). \quad (5)$$

Setting the derivatives of the Lagrangian $\nabla_{\mathbf{p}, \mathbf{u}, \boldsymbol{\mu}} \mathcal{L}$ to zero with respect to \mathbf{p} , \mathbf{u} , and $\boldsymbol{\mu}$ yields the state and adjoint equations

$$\mathbf{A} \mathbf{u} = \mathbf{f}, \quad \mathbf{A}^T \mathbf{p} = -\nabla_{\mathbf{u}} F \quad (6)$$

and the reduced gradient in $\boldsymbol{\mu}$,

$$\mathbf{g}^T \nabla_{\boldsymbol{\mu}} \mathcal{L} = \mathbf{p}^T [\mathcal{D}_{\boldsymbol{\mu}} \mathbf{A}(\cdot)] \mathbf{u} + \alpha \nabla_{\boldsymbol{\mu}} R, \quad (7)$$

where we introduce the differential of any function or operator g as

$$\mathcal{D}_{\mathbf{y}} g(\Delta \mathbf{y}) := \lim_{h \rightarrow 0} \frac{g(\mathbf{y} + h \Delta \mathbf{y}) - g(\mathbf{y})}{h}. \quad (8)$$

The differential $\mathcal{D}_{\mathbf{y}} g(\Delta \mathbf{y})$ describes the change in g resulting from an incremental update at point \mathbf{y} in the direction of $\Delta \mathbf{y}$ and is a linear operator acting on $\Delta \mathbf{y}$.

By inspecting the optimality conditions in equation 6, the quantity \mathbf{p} can be identified as the adjoint wavefield. Note that the transpose \mathbf{A}^T naturally leads to a reverse-time propagation of the adjoint \mathbf{p} . The adjoint source emerges as the negative gradient of the observation term and contains the data-misfit information.

Eliminating state and adjoint wavefields (\mathbf{u}, \mathbf{p}) via equation 6 gives an expression for the reduced gradient,

$$\mathbf{g} = -\nabla_{\mathbf{u}} F^T \mathbf{A}^{-1} [\mathcal{D}_{\boldsymbol{\mu}} \mathbf{A}(\cdot)] \mathbf{A}^{-1} \mathbf{f} + \alpha \nabla_{\boldsymbol{\mu}} R, \quad (9)$$

which is the Born approximation Fréchet kernel (Tromp et al., 2005) in DTO formulation. Thus, the gradient can be computed by solving for \mathbf{u} and \mathbf{p} in one forward and one adjoint computation, and assem-

bling it through space-time inner products. Here, these relations are derived exclusively in the discrete setting.

Gradient-based optimization methods such as nonlinear conjugate gradient (NCG) and limited-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) can be implemented at this point (Nocedal and Wright, 1999). A DTO Newton method in the frequency domain is described by Pratt (1999) and by Brenders and Pratt (2007), and an OTD Newton method is described by Chen et al. (2007). Both require explicit factorization of the Hessian matrix. A 2D application of NCG in the OTD setting is demonstrated in Tape et al. (2007).

NEW ALGORITHMIC STRATEGIES

Although it is impossible to eliminate all numerical difficulties in seismic waveform inversion entirely, our goal is to identify and address them one by one with suitable algorithms. In the following, we discuss Newton-Krylov optimization, i.e., the iterative solution of the linear Newton system of equations via a Krylov subspace method (so-called inexact Newton methods; see Dembo et al. [1982] and Saad [2003]). Santosa and Symes (1988) propose this method for seismic inversion. In this context, we also include primal-dual TV and active set strategies.

Inexact Newton-Krylov method

Common gradient-based optimization methods have difficulties scaling to the many inversion variables that result when continuous parameter fields are discretized. This results from two effects: the rate of convergence of gradient-based methods is only linear, and the convergence constant usually deteriorates with increasing grid resolution.

Newton's method for systems of equations converges quadratically, and the convergence rate is grid independent (Allgower et al., 1986; Heinkenschloss et al., 1991; Heinkenschloss, 1993). However, it requires Hessian information, i.e., computing second derivatives. A full derivation of first and second derivatives based on the weak form of the underlying partial-differential equations is detailed in Epanomeritakis et al. (2008). In the discrete setting outlined here, Newton's method takes the form

$$\Delta \boldsymbol{\mu}^{(k)} = \boldsymbol{\mu}^{(k+1)} - \boldsymbol{\mu}^{(k)} = -(\nabla_{\boldsymbol{\mu}}^2 \mathcal{L})^{-1} \nabla_{\boldsymbol{\mu}} \mathcal{L}|_{\boldsymbol{\mu}^{(k)}}, \quad (10)$$

where $\boldsymbol{\mu}^{(k)}$ are subsequent updates of the model parameters. The reduced Hessian matrix $\mathbf{H} := \nabla_{\boldsymbol{\mu}}^2 \mathcal{L}$ is the derivative of the reduced gradient \mathbf{g} (equation 9). It is computed using the chain rule and substituting the state equation 6. The result is

$$\begin{aligned} \mathbf{H} := \nabla_{\boldsymbol{\mu}}^2 \mathcal{L} &= \mathbf{E}_{\mathbf{u}}^T \mathbf{A}^{-T} (\nabla_{\mathbf{u}}^2 F) \mathbf{A}^{-1} \mathbf{E}_{\mathbf{u}} + \alpha \nabla_{\boldsymbol{\mu}}^2 R \\ &+ \mathbf{E}_2 - \mathbf{E}_{\mathbf{u}}^T \mathbf{A}^{-T} \mathbf{E}_{\mathbf{p}} - \mathbf{E}_{\mathbf{p}}^T \mathbf{A}^{-1} \mathbf{E}_{\mathbf{u}}. \end{aligned} \quad (11)$$

The first line contains the positive definite Gauss-Newton approximation, and the second line contains the remaining second-order terms, which can give rise to negative eigenvalues. For ease of notation, we recast the differentials (which are linear operators acting on $\Delta \boldsymbol{\mu}$) as matrices:

$$\mathbf{E}_{\mathbf{u}} \Delta \boldsymbol{\mu} := [\mathcal{D}_{\boldsymbol{\mu}} \mathbf{A}(\Delta \boldsymbol{\mu})] \mathbf{u}, \quad (12)$$

$$\mathbf{E}_{\mathbf{p}} \Delta \boldsymbol{\mu} := [\mathcal{D}_{\boldsymbol{\mu}} \mathbf{A}^T(\Delta \boldsymbol{\mu})] \mathbf{p}, \quad (13)$$

$$\Delta \boldsymbol{\mu}^T \mathbf{E}_2 \Delta \boldsymbol{\mu}' := \mathbf{p}^T [\mathcal{D}_{\boldsymbol{\mu}}^2 \mathbf{A}(\Delta \boldsymbol{\mu}, \Delta \boldsymbol{\mu}')] \mathbf{u}. \quad (14)$$

This form makes it clear that the Hessian \mathbf{H} is a dense matrix that is far too costly to store and factor for fine model discretizations. The remedy is to resort to iterative (and thus inexact) Krylov solvers for the Newton system

$$\mathbf{H}\Delta\boldsymbol{\mu} = -\mathbf{g}, \quad (15)$$

particularly (preconditioned) conjugate gradients (CGs) and the Steihaug variant (Dembo and Steihaug, 1983).

In the inexact Newton-Krylov setting, CG is terminated early to avoid directions of negative curvature, which guarantees a descent direction and thus global convergence with a suitable line search; to prevent oversolving when far from the minimum; and to remain inside the trust region (if this globalization variant is used). (See Eisenstat and Walker [1994, 1996] and Nocedal and Wright [1999] for details.) Moreover, because CG requires only matrix-vector products, the method can be implemented in a matrix-free manner and \mathbf{H} is never stored. Instead, Hessian-vector products are formed on the fly at each CG iteration, each of which requires three forward or adjoint wave-propagation solutions and several projection and extension operations (the matrices \mathbf{E}_u and \mathbf{E}_p and their transposes in equation 11).

An advantage of this approach is its generality: tuning algorithmic parameters and termination criteria alone can realize almost any trade-off between cost per nonlinear iteration and convergence rate, creating a continuum of methods from basic steepest descent (one CG iteration per Newton step) to full Newton (fully converged CG).

Continuation principles

A major drawback of least-squares full-waveform inversion is the oscillatory nature of the least-squares misfit functional in equation 1, especially in 2D and 3D setups. It generally suffers from spurious local minima, which necessitates an accurate initial guess of the low-wavenumber background or significant low-frequency content of the source. This is the most critical point in practice because accurate initial guesses and/or low-frequency information in the data often are not available.

The 1D inverse problem is special in the sense that it admits a complete mathematical analysis. Because of the 1D geometry of the problem, the depth can be reparameterized in terms of traveltimes. Based on this transformation, the 1D inverse problem with a source function of $f(z,t) = \delta(z)\delta(t)$ has been proven to have a unique global solution (Bamberger et al., 1979; Bube and Burridge, 1983; Sacks and Santosa, 1987). However, this does not apply as soon as noisy data or finite-frequency sources (as in this article) are introduced.

Several reformulations of the inverse problem have been developed to mitigate the nonlinearity of the objective functional and to increase the basin of attraction around the global minimum. Examples are differential semblance optimization (DSO; Symes and Carazzone, 1991), traveltimes inversion (Luo and Schuster, 1991), migration-based traveltimes (MBTT; Chavent et al., 1995), and migration velocity analysis (MVA; Symes, 2008). In this article, we focus on grid and frequency continuation approaches (see, e.g., Kolb et al., 1986; Bunks et al., 1995; Sirgue and Pratt, 2004).

The convexity condition for the objective functional can be translated into the following relation involving the characteristic dimension of the inversion domain:

$$L \lesssim \lambda_{\max} = \frac{c_{\max}}{f_{\min}}, \quad (16)$$

where the proportionality constant varies between 1/2 (for very pessimistic estimates) and small integers in our experiments. The value L is the extent (depth) of the domain, f_{\min} is the low-frequency cutoff of the source, and c_{\max} is the maximum occurring wavespeed. The effective longest wavelength λ_{\max} ideally should span the whole domain. This statement gives rise to the following continuation strategies:

- For a fixed-length domain, invert for low frequencies first and increase the frequency in several steps.
- For fixed frequency content, invert for the shallow parts first (small L) and increase the depth in several steps.

Primal-dual TV regularization

The seismic inverse problem without regularization is ill posed, that is, its solution is not unique and/or the solution is unstable to perturbations in the data. A standard approach to make the problem uniquely solvable is to add the Tikhonov regularization term (equation 3) to the objective functional. Although this form of regularization is easy to treat numerically, its smoothing of the parameter model by penalizing the high-wavenumber components precludes recovery of steep jumps in the parameter fields.

TV regularization methods are widely used in image processing to capture sharp interfaces (Rudin et al., 1992; Acar and Vogel, 1994). As defined in equation 4, the TV regularization operator and its gradient read

$$\begin{aligned} R_{1,\epsilon}(\boldsymbol{\mu}) &= \int_{\Omega} \sqrt{\|\nabla \boldsymbol{\mu}\|^2 + \epsilon} dx, \\ \mathcal{D}_{\boldsymbol{\mu}} R_{1,\epsilon}(v) &= \int_{\Omega} \frac{\nabla \boldsymbol{\mu} \cdot \nabla v}{\sqrt{\|\nabla \boldsymbol{\mu}\|^2 + \epsilon}} dx. \end{aligned} \quad (17)$$

The corresponding Hessian $\nabla_{\boldsymbol{\mu}}^2 R_{1,\epsilon}$ is a nonlinear, heterogeneous, highly anisotropic operator, leading to severe numerical difficulties. Convergence of TV in the Newton framework can be quite slow.

Following Chan et al. (1999), we use a primal-dual variant of TV. The main idea is to introduce a slack variable w that simplifies the expression for the gradient (and the Hessian):

$$w := \frac{\nabla \boldsymbol{\mu}}{\sqrt{\|\nabla \boldsymbol{\mu}\|^2 + \epsilon}}, \quad \mathcal{D}_{\boldsymbol{\mu}} R_{1,\epsilon}(v) = \int_{\Omega} w \cdot \nabla v dx. \quad (18)$$

The Newton iterations are modified to include the additional discrete unknown \mathbf{w} . Choosing the initial value $\mathbf{w}^{(0)} = 0$, the enhanced Newton step corresponds to simultaneously solving the system

$$\nabla \mathcal{L}(\mathbf{p}, \mathbf{u}, \boldsymbol{\mu}, \mathbf{w}) = 0, \quad (19)$$

$$\mathbf{w} \sqrt{\|\nabla \boldsymbol{\mu}\|^2 + \epsilon} - \nabla \boldsymbol{\mu} = 0. \quad (20)$$

The condition $|\mathbf{w}^{(k)}| \leq 1$ is enforced by pointwise projection (Hintermüller and Stadler, 2006).

Another approach to speed up TV is the fixed-point or lagged-diffusivity TV iteration (Vogel and Oman, 1996; Dobson and Vogel,

1997). It can be emulated within the primal-dual algorithm by enforcing $\mathbf{w}^{(k)} = 0$ in each Newton iteration.

Primal-dual active set strategy

To satisfy physical constraints and to ensure stability of the wave-propagation solver, it is important to explicitly enforce lower and upper inequality bounds on the model parameters:

$$\mu_-(z) \leq \mu(z) \leq \mu_+(z), \quad z \in \Omega. \quad (21)$$

The inclusion of bound constraints presents considerable difficulties for commonly used active set or interior point methods. We resort to a primal-dual active set strategy following Hintermüller et al. (2003) and Hintermüller and Hinze (2006). The essential idea is to define a dual variable $\lambda(z)$ pointwise by the relation

$$\lambda - \max(0, \lambda + c(\mu - \mu_+)) - \min(0, \lambda + c(\mu - \mu_-)) = 0, \quad (22)$$

with an arbitrary constant $c > 0$. Thus, a point $z \in \Omega$ can be in the upper active set ($\mu(z) = \mu_+(z)$), in the lower active set ($\mu(z) = \mu_-(z)$), or in the inactive set ($\lambda(z) = 0$). From the implementor's point of view, the (inner) Krylov solver for the Newton system (equation 15) is wrapped into a loop that determines the active sets based on λ and solves the Newton system only over the inactive set. This modification is invisible to the (outer) Newton update in equation 10. In conjunction with grid continuation, the convergence is mesh independent and requires only a few additional CG iterations.

1D EXPERIMENTS

In this section, we discuss the results of experiments on seismic inverse problems governed by the 1D wave equation in the time domain:

$$\rho(z) \frac{\partial^2}{\partial t^2} u(z, t) = \frac{\partial}{\partial z} \mu(z) \frac{\partial}{\partial z} u(z, t) + f(z, t) \text{ in } \Omega \times (0, T),$$

$$u(t=0) = 0, \quad \dot{u}(t=0) = 0. \quad (23)$$

The density $\rho(z)$ is assumed known, and we invert for the elastic modulus $\mu(z)$, which is related to the compressional wavespeed $c(z) = \sqrt{\mu(z)/\rho(z)}$. Space is discretized with piecewise linear finite

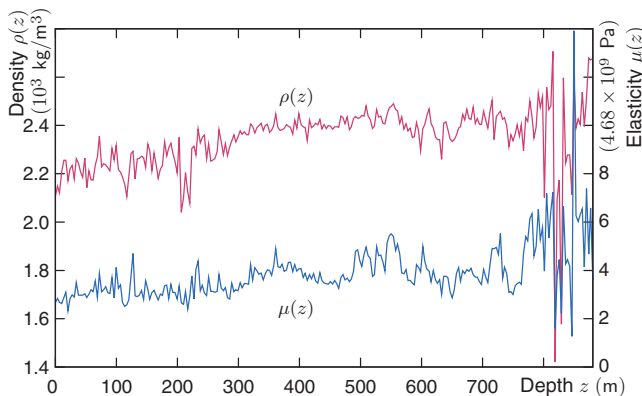


Figure 1. Density $\rho(z)$ and elasticity modulus $\mu(z)$, filtered to $\Delta z = 3.4$ -m intervals for clarity of display. The compressional wavespeeds range from 2160 to 5610 m/s.

elements on the interval $\Omega = (0, L)$. The spatial resolution for u and p is chosen per experiment to ensure at least 10 points per shortest expected wavelength. The resolution for the elasticity modulus $\mu(z)$ can be chosen coarser as appropriate. We invoke the free-surface condition on the top end $z = 0$ and a first-order absorbing boundary condition at $z = L$:

$$\mu(z) \frac{\partial}{\partial z} u(z, t) = 0, \quad z = 0,$$

$$\mu(z) \frac{\partial}{\partial z} u(z, t) = -\sqrt{\rho(z)\mu(z)} \frac{\partial}{\partial t} u(z, t), \quad z = L. \quad (24)$$

The source and receiver are located at the top ($z = 0$), and the source-time function is represented by a Ricker wavelet, taken as the second derivative of a Gaussian. The frequency spectrum of the Ricker wavelet is characterized by its central frequency. It has negligible frequency content above about three times the central frequency.

We use a target earth model derived from borehole measurements and generate synthetic observations by solving the wave equation using this model. The properties of this target model are depicted in Figure 1.

Experiment A — Grid continuation and convergence

In experiment A, we demonstrate grid continuation. The initial guess is an approximation of the true model at a coarse discretization level $j = 4$, where $N = 2^j$ is the number of elements for the discretization of μ . The level j is increased successively from four to seven. The number of outer (Newton) iterations is set to five for all but the last stage, where it is 15. The data $u^*(z = 0, t)$ are created with the same parameter discretization used in the inversion at the respective refinement levels. This is an idealized setting because in reality we do not have a choice on the roughness of the medium. It is, however, related to low-pass filtering of the observed and synthetic data. We use more realistic data in experiment B.

The result from grid continuation is compared with the true model in Figure 2. The relative root-mean-square (rms) error with respect to the observed data is 0.49%. To give an example of the performance of Newton's method, we provide the iteration history for the coarsest level in Table 1. In the beginning of the iteration, progress is

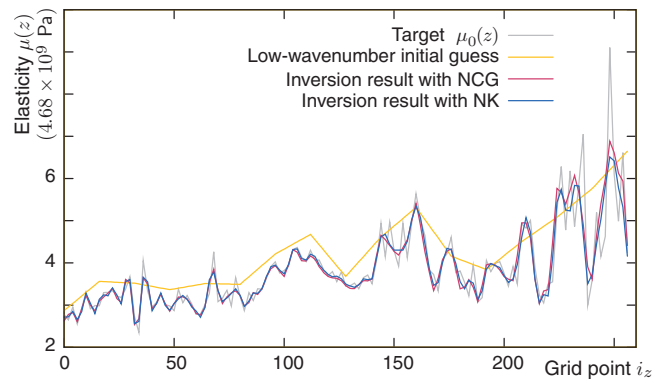


Figure 2. Target model (gray) and results obtained with NCGs (purple) and Newton-Krylov (NK, blue) for level 7: model resolution $\Delta z = 6.8$ m, frequency = 25 Hz. Obtained with Tikhonov regularization and grid continuation. Initial guess (yellow) is 10% too fast. The relative data rms error is 0.88% for NCG and 0.49% for NK.

relatively linear. Once the iterates are sufficiently close to the optimum, we begin to observe quadratic convergence behavior, and the gradient \mathbf{g} and step length are reduced almost to machine precision.

In Figure 3, we compare to inversion without continuation. We examine low- and high-wavenumber initial guesses that are 10% too fast. Although the error in the medium is obviously large, we examine the fit of the data in Figure 4 to find that the rms errors are 6.31% and 18.1%, respectively. This is significantly larger than the rms error obtained with grid continuation (0.49%) but still places the oscillations of the signal at the correct locations. Apparently, a large model error can still provide a moderately good fit of the data, which we attribute to the 1D structure of this experiment.

Table 2 shows accuracy and convergence rates obtained for nonlinear conjugate gradients (NCGs) and the Newton-Krylov method at medium — to high-resolution settings. The setup here is slightly different, using a central frequency of 50 Hz, starting with an initial guess at level 8, and running at most 500 NCG or 20 Newton steps. NCG and Newton’s method converge reasonably well to rms errors

Table 1. Convergence table for Newton-Krylov inversion in very coarse parameter space (level 4, 17 grid points).

Newton iteration	Gradient \mathbf{g}	Relative rms error	Relative error in μ	Step length
Initial	2.01836e-02	2.13699e-01	1.00000e-01	—
1	2.82682e-02	1.97628e-01	9.68066e-02	3.24028e-03
2	1.44228e-02	1.26899e-01	8.27373e-02	1.14733e-02
3	1.30338e-02	9.35648e-02	7.36792e-02	3.67311e-03
4	4.73618e-03	7.09611e-02	6.92431e-02	1.85930e-03
5	4.20749e-03	4.82790e-02	5.54026e-02	1.35189e-03
6	9.07508e-04	4.78811e-02	5.54430e-02	1.72491e-05
7	1.66000e-03	1.47563e-02	2.67686e-02	1.03785e-03
8	5.33378e-04	2.92431e-03	4.06868e-03	1.04789e-04
9	1.83362e-05	1.62347e-04	3.12269e-04	4.28218e-06
10	1.00323e-07	5.88317e-05	1.47565e-04	1.31302e-08
11	4.53587e-11	5.89037e-05	1.47810e-04	1.59807e-13

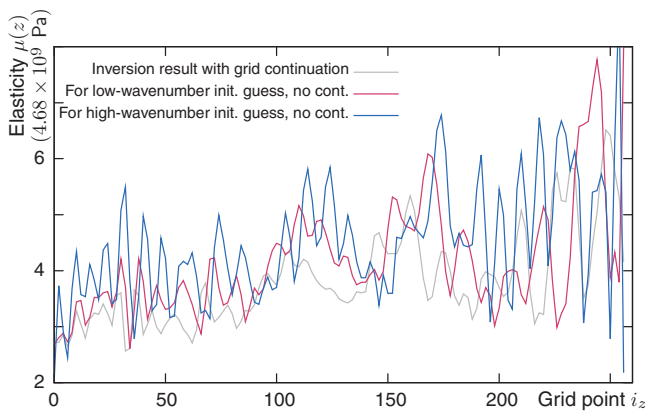


Figure 3. Result obtained with Newton-Krylov and continuation (gray, cf. Figure 2) and two results without continuation for a low-wavenumber initial guess that is 10% too fast (purple) and for a high-wavenumber initial guess that is 10% too fast (blue).

around 1.5% and 2%. NCG makes more progress in the beginning and then slows somewhat, whereas Newton’s method starts slower and increases progress toward the end of the inversion. However, the problem appears sufficiently ill conditioned so that the quadratic convergence of Newton’s method does not show clearly.

We examine the eigenvalues of the Hessian matrix at all final iterates and confirm that these are nonnegative. Considering the small gradient norm, this indicates the process converged to (at least a weak local) minimum rather than a saddle point. Among the different TV implementations, the primal-dual variant is most accurate but also slightly more expensive than standard and fixed-point TV. TV seems to provide slightly lower rms errors than Tikhonov at roughly comparable cost.

Experiment B — Time horizon continuation and TV regularization

In reality, the low-frequency information is often limited, and the continuation used in experiment A may be impractical. In experiment B, we use continuation from short to longer observation times for a fixed frequency that is well represented in the data set. This corresponds to illuminating layers of increasing depth.

First, we choose an initial guess that is coarsened to level 4 and 10% too fast on average, and a Ricker wavelet with a fixed central frequency of 25 Hz. We use a discretization of the parameter domain with 128 elements ($j = 7$). The results are displayed in Figure 5. At the earlier stages when depth information is limited, the Tikhonov regularization leads to smooth models in the unresolved part of the domain. At the final stage with $T = 1$ s, the waves traverse the full model and the structure is recovered at full depth, albeit with less accuracy in the deeper regions. We compare this result (rms 1.31%) with that obtained from a 15% erroneous initial guess (rms 5.91%), which produces a significantly higher error in the velocity model.

The results for primal-dual TV regularization are given in Figure 6. Here, the initial guesses are 5% and 8% off, respectively, resulting in rms errors (1.94% and 6.67%) comparable to Tikhonov regularization. This indicates that the basin of attraction is

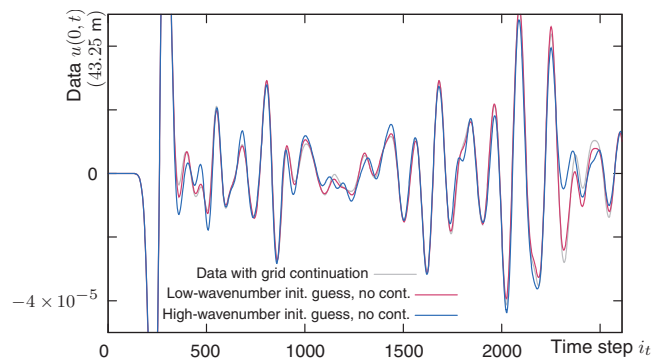


Figure 4. Zoom of the data resulting from inversion with continuation (gray, cf. Figure 2, rms 0.49%) and from inversions without continuation for a low-wavenumber initial guess (purple, rms 6.31%) and for a high-wavenumber initial guess (blue, rms 18.1%).

Table 2. Convergence behavior of nonlinear conjugate gradients (NCGs) with Wolfe line search and Newton-Krylov method. Tikhonov regularization ($\alpha = 5 \times 10^{-6}$), several TV variants ($\alpha = 2 \times 10^{-4}$), model resolution $\Delta z = 1.7$ m, and frequency $f = 50$ Hz. The three sets of rows correspond to progressive stages of the same inversion run.

	NCG		Newton-Krylov			
	Tikhonov	TV	Tikhonov	TV	Fixed-point TV	Primal-dual TV
Initial rms error	35.4%	47.3%	35.4%	47.3%	47.3%	47.3%
Number Newton iterations	—	—	3	3	3	3
Number CG iterations	20	20	29	13	15	15
Gradient	9.96×10^{-3}	9.40×10^{-3}	3.37×10^{-2}	1.29×10^{-1}	1.02×10^{-1}	1.15×10^{-1}
Relative rms error	9.25%	13.9%	9.40%	35.4%	32.4%	33.8%
Relative error in μ	13.4%	13.9%	13.3%	14.3%	14.3%	14.3%
Number Newton iterations	—	—	6	6	6	6
Number CG iterations	100	100	122	63	57	65
Gradient	5.07×10^{-4}	1.37×10^{-3}	1.40×10^{-3}	2.64×10^{-2}	2.78×10^{-2}	3.06×10^{-2}
Relative rms error	3.25%	3.32%	2.25%	8.56%	11.0%	7.75%
Relative error in μ	12.8%	12.9%	12.5%	13.3%	13.4%	13.3%
Number Newton iterations	—	—	20	20	20	20
Number CG iterations	500	500	365	341	375	412
Gradient	8.66×10^{-5}	6.41×10^{-4}	1.75×10^{-4}	7.12×10^{-4}	1.35×10^{-4}	4.09×10^{-4}
Relative rms error	2.16%	1.91%	2.24%	1.75%	1.65%	1.60%
Relative error in μ	12.5%	12.3%	12.5%	12.4%	12.2%	12.3%
Number negative curvature	—	—	3	5	7	5

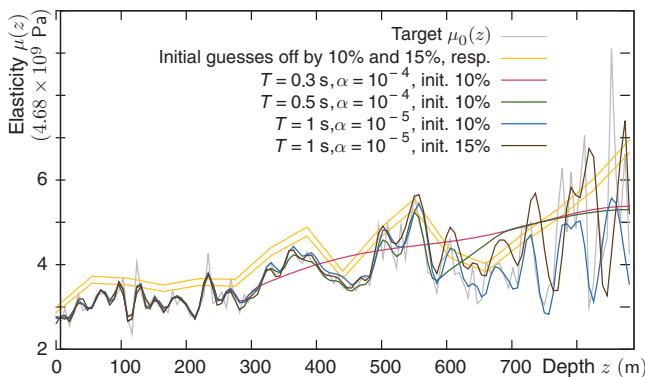


Figure 5. Experiment B: stages of the inversion at 25 Hz, level 7 with Tikhonov regularization. Time horizon T is increased, but regularization weight α is decreased (stage $T = 0.7$ s, $\alpha = 10^{-5}$ not shown). Final rms for the initial guess that is 10% too fast (blue) is 1.31%; for the initial guess that is 15% too fast (brown), it is 5.91%.

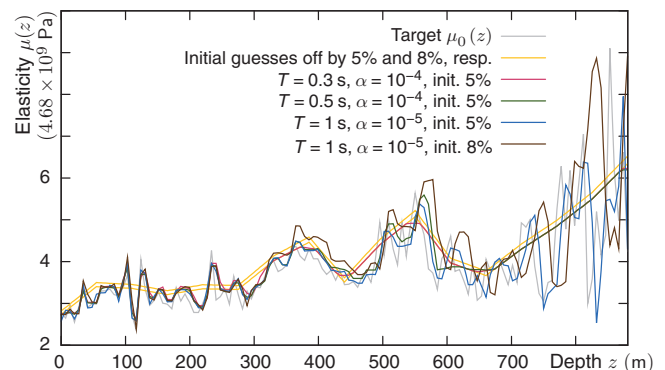


Figure 6. Experiment B with primal-dual TV regularization. Final rms for initial guess that is 5% too fast (blue) is 1.94%; for the initial guess that is 8% too fast (brown), it is 6.67%.

smaller for TV regularization. One can also observe that interface locations are somewhat shifted for depths greater than 500 m. On the other hand, in comparison to Tikhonov regularization, the model is not smoothed but rather is separated into steps of constant velocity, which can be advantageous if the true model is a layered medium.

Experiment C — Frequency continuation and high resolution

In experiment C, we push the resolution limit in a two-stage process. We begin with continuation on the time horizon and then use grid and frequency continuation to reach a dominant signal frequency of 100 Hz. Tikhonov regularization is used throughout this section.

With receiver signals computed at level 7 as in experiment B, 10 Hz is the highest frequency that produces an inversion sufficiently accurate to initiate the second stage. In the second stage, we increase frequency and grid resolution in three steps. The finite-element grid spacing and the source frequency are as follows:

Step 1. $\Delta z = 3.4$ m, $f = 25$ Hz;

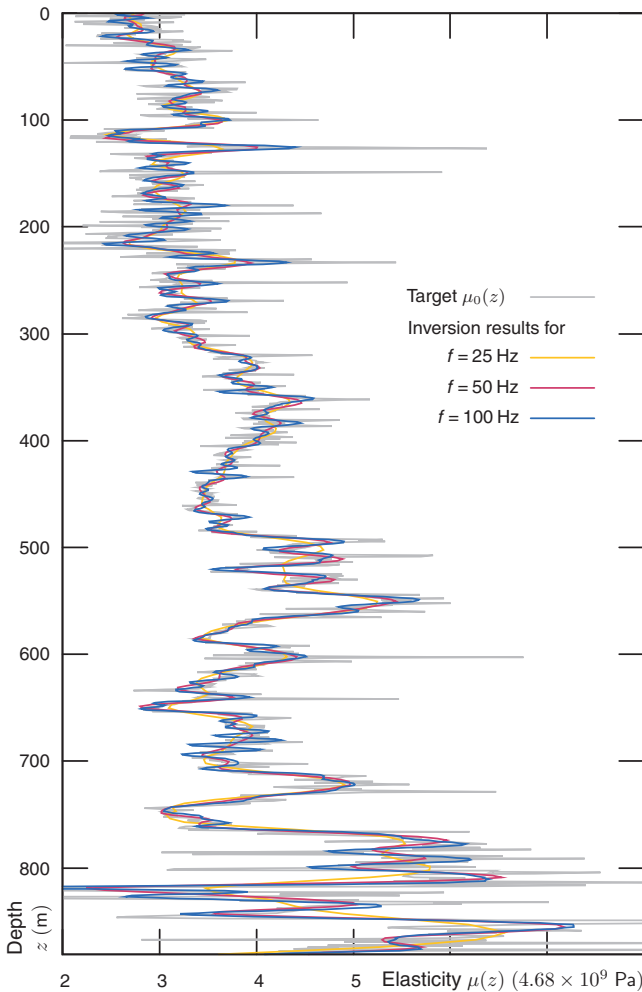


Figure 7. Inversion achieved through three frequency-continuation steps, starting from a first stage result at 10 Hz. Central frequency of the final Ricker wavelet is 100 Hz. Final grid spacing (level 10) is 86 cm.

Step 2. $\Delta z = 1.7$ m, $f = 50$ Hz;

Step 3. $\Delta z = 86$ cm, $f = 100$ Hz.

These settings apply to the generation of the data and the inversion process. The final step at 100 Hz uses more than 1000 elements and 20,000 time steps. This amounts to 40 million total unknowns for the forward and adjoint wavefields. Figure 7 indicates that the inversion achieves subwavelength resolution (< 10 m).

DISCUSSION

The goal of our numerical experiments is to assess several algorithmic strategies for full-waveform seismic inversion for a 1D test bed problem that lends itself to carrying out the necessary experiments. The experiments are conducted with a Newton-Krylov non-linear optimization method that is versatile in the sense that it can handle bound inequalities on the medium and use TV regularization to allow for jumps in the medium, both without significant penalty to convergence or execution time. The inversion algorithm is formulated in a dimensionally independent way, such that it can be applied to 2D or 3D problems by changing out the forward and adjoint solvers and their derivatives.

Because of the band-limited source, we observe convergence to what is likely a local minimum when the initial guess of the medium is outside of the basin of attraction of the global minimum (see Figures 3 and 4). This has been verified by computing all eigenvalues of the Hessian matrix and finding them to be nonnegative, with the gradient being on the order of 10^{-5} . However, as a result of the ill-conditioning of the Hessian matrix, the regime of quadratic convergence of Newton's method is not clearly identifiable.

We compare the Newton-Krylov inversion to the Polak-Ribière NCG algorithm. During the initial steps of the inversion, away from the global optimum, NCG often creates steps that reduce the rms error faster than Newton's method. On the other hand, close to the optimal solution, Newton's method has better convergence properties. In particular, the number of Newton iterations is unaffected by increasing the resolution, and we believe the method to be scalable to much larger problems. We use both Tikhonov and TV regularization. The convergence speeds of the primal-dual TV algorithm and Tikhonov are comparable. TV yields layered instead of smooth models; the basin of attraction of TV seems smaller than that of Tikhonov.

Without continuation of grid, frequency, or the time horizon, we are unable to achieve convergence. When continuation is used, we obtain agreement of the medium down to feature widths that are comparable to the wavelength. We invert successfully using initial guesses with up to 10% relative error. However, our results need to be considered idealized because the data are created synthetically and noise free.

CONCLUSION

We have formulated a flexible, scalable strategy for full-waveform seismic inversion and have assessed its performance on a 1D example problem. Our strategy is based on least-squares inversion and uses techniques to deal with spurious local minima to incorporate regularization that admits discontinuities in the model and to enforce inequality constraints on the model iterates.

To create a globally convergent method in the presence of strong nonconvexity of the objective functional, we confirm that continuation from low to high source frequencies is effective. Iterative elongation of the time horizon provides additional robustness.

In view of the different convergence properties of NCG and Newton-Krylov methods, one attempt at a future algorithm could be to begin with NCG and switch over to Newton's method in the process. However, a similar behavior can be obtained within the inexact Newton-Krylov framework because it supports variable termination criteria that would allow working with low solver tolerances initially, thus performing cheap gradient steps, and tightening the tolerances toward the end for increased accuracy. A further extension could be to consider limited-memory BFGS variants, which provide NCG as a special case, as a preconditioner for the Newton-Krylov method.

A primal-dual implementation of TV regularization proves numerically robust, is comparable in cost and accuracy to standard Tikhonov regularization, and promises a better fit when the expected model contains discontinuities. In response to the apparently smaller basin of attraction of TV, using Tikhonov in the beginning of the inversion process and TV for the final steps could be a valid compromise. Another possibility is to use continuation on the parameter ϵ in the TV operator.

Using the algorithmic strategies described in this article permits solution of inverse problems with inaccurate initial models and a sharply varying background velocity. With central source frequencies of up to 100 Hz, we can recover a 1D model derived from well logs to high detail. The algorithms are essentially dimension independent, so they offer the hope of making full-waveform inversion more practical in large-scale 2D and 3D scenarios.

ACKNOWLEDGMENTS

The authors thank Apache Corporation for cooperation in providing the data and permission to publish this work and thank Sergey Fomel, Georg Stadler, and James Martin for extended discussion. The questions and suggestions of the associate editor and the anonymous referees are greatly appreciated. NSF grants EAR-0326449 and DMS-0724746 are gratefully acknowledged.

REFERENCES

- Acar, R., and C. Vogel, 1994, Analysis of bounded variation penalty methods for ill-posed problems: *Inverse Problems*, **10**, 1217–1229.
- Akcelik, V., J. Bielak, G. Biros, I. Epanomeritakis, A. Fernandez, O. Ghattas, E. J. Kim, J. Lopez, D. O'Hallaron, T. Tu, and J. Urbanic, 2003, High-resolution forward and inverse earthquake modeling on terascale computers: *Proceedings of the IEEE/ACM SC2003 Conference*, 52.
- Akcelik, V., G. Biros, and O. Ghattas, 2002, Parallel multiscale Gauss-Newton-Krylov methods for inverse wave propagation: *Proceedings of the IEEE/ACM SC2002 Conference*, 1–15.
- Allgower, E. L., K. Böhmer, F. A. Potra, and W. C. Rheinboldt, 1986, A mesh-independence principle for operator equations and their discretizations: *SIAM Journal on Numerical Analysis*, **23**, 160–169.
- Bamberger, A., G. Chavent, and P. Lailly, 1979, About the stability of the inverse problem in 1-D wave equations — Application to the interpretation of seismic profiles: *Applied Mathematics and Optimization*, **5**, 1–47.
- Brenders, A. J., and R. G. Pratt, 2007, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: *Geophysical Journal International*, **168**, 133–151.
- Bube, K. P., and R. Burridge, 1983, The one-dimensional inverse problem of reflection seismology: *SIAM Review*, **25**, 497–559.
- Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *Geophysics*, **50**, 1457–1473.
- Chan, T., G. Golub, and P. Mulet, 1999, A nonlinear primal-dual method for total variation-based image restoration: *SIAM Journal on Scientific Computing*, **20**, 1964–1977.
- Chavent, G., 1974, Identification of functional parameters in partial differential equations, in R. E. Goodson and M. Polis, eds., *Identification of parameters in distributed systems*: American Society of Mechanical Engineers, 31–48.
- Chavent, G., F. Clément, and S. Gómez, 1995, Waveform inversion by MBTT formulation: *Proceedings of the 3rd International Conference on Mathematical and Numerical Aspects of Wave Propagation*, 713–722.
- Chen, P., L. Zhao, and T. H. Jordan, 2007, Full 3D tomography for the crustal structure of the Los Angeles region: *Bulletin of the Seismological Society of America*, **97**, 1094–1120.
- Conn, A., N. Gould, and P. Toint, 2000, *Trust-region methods*: SIAM.
- Dembo, R. S., S. C. Eisenstat, and T. Steihaug, 1982, Inexact Newton methods: *SIAM Journal on Numerical Analysis*, **19**, 400–408.
- Dembo, R. S., and T. Steihaug, 1983, Truncated-Newton algorithms for large-scale unconstrained optimization: *Mathematical Programming*, **26**, 190–212.
- Deuffhard, P., 2004, *Newton methods for nonlinear problems: Affine invariance and adaptive algorithms*: Springer.
- Dobson, D., and C. R. Vogel, 1997, Convergence of an iterative method for total variation denoising: *SIAM Journal on Numerical Analysis*, **34**, 1779–1791.
- Eisenstat, S. C., and H. F. Walker, 1994, Globally convergent inexact Newton methods: *SIAM Journal on Optimization*, **4**, 393–422.
- , 1996, Choosing the forcing terms in an inexact Newton method: *SIAM Journal on Scientific Computing*, **17**, 16–32.
- Epanomeritakis, I., V. Akcelik, O. Ghattas, and J. Bielak, 2008, A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion: *Inverse Problems*, **24**, 034015.
- Gunzburger, M. D., 2003, *Perspectives in flow control and optimization*: SIAM.
- Heinkenschloss, M., 1993, Mesh independence for nonlinear least squares problems with norm constraints: *SIAM Journal on Optimization*, **3**, 81–117.
- Heinkenschloss, M., M. Laumen, and E. W. Sachs, 1991, Gauss-Newton methods with grid refinement, in W. Desch, F. Kappel, and K. Kunisch, eds., *Optimal control of partial differential equations*: Birkhäuser, 161–174.
- Hintermüller, M., and M. Hinze, 2006, A SQP-semismooth Newton-type algorithm applied to control of the stationary Navier-Stokes system subject to control constraints: *SIAM Journal on Optimization*, **16**, 1177–1200.
- Hintermüller, M., K. Ito, and K. Kunisch, 2003, The primal-dual active set strategy as a semi-smooth Newton method: *SIAM Journal on Optimization*, **13**, 865–888.
- Hintermüller, M., and G. Stadler, 2006, An infeasible primal-dual algorithm for total variation-based INF-convolution-type image restoration: *SIAM Journal on Scientific Computing*, **28**, 1–23.
- Kolb, P., F. Collino, and P. Lailly, 1986, Pre-stack inversion of a 1-D medium: *Proceedings of the IEEE*, **74**, 498–508.
- Luo, Y., and G. T. Schuster, 1991, Wave-equation traveltime inversion: *Geophysics*, **56**, 645–653.
- Nocedal, J., and S. J. Wright, 1999, *Numerical optimization*: Springer.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503.
- Polak, E., and G. Ribière, 1969, Note sur la convergence de méthodes de directions conjuguées: *Revue Française d'Informatique et de Recherche Opérationnelle*, **16**, 35–43.
- Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: *Geophysics*, **64**, 888–901.
- Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362.
- Rudin, L., S. Osher, and E. Fatemi, 1992, Nonlinear total variation based noise removal algorithms: *Physica D*, **60**, 259–268.
- Saad, Y., 2003, *Iterative methods for sparse linear systems*, 2nd ed.: SIAM.
- Sacks, P., and F. Santosa, 1987, A simple computational scheme for determining the sound speed of an acoustic medium from its surface impulse response: *SIAM Journal on Scientific and Statistical Computing*, **8**, 501–520.
- Santosa, F., and W. W. Symes, 1985, The determination of a layered acoustic medium via multiple impedance profile inversions from plane wave responses: *Geophysical Journal of the Royal Astronomical Society*, **81**, 175–195.
- , 1988, Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology: *Inverse Problems*, **4**, 211–233.
- , 1989, An analysis of least squares velocity inversion: SEG.
- Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: *Geophysics*, **69**, 231–248.
- Symes, W. W., 2008, Migration velocity analysis and waveform inversion: *Geophysical Prospecting*, **56**, 765–790.
- Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential

- semblance optimization: *Geophysics*, **56**, 654–663.
- Tape, C., Q. Liu, and J. Tromp, 2007, Finite-frequency tomography using adjoint methods — Methodology and examples using membrane surface waves: *Geophysical Journal International*, **168**, 1105–1129.
- Tarantola, A., 1986, A strategy for nonlinear elastic inversion of seismic reflection data: *Geophysics*, **51**, 1893–1903.
- Tikhonov, A. N., and V. A. Arsenin, 1977, *Solution of ill-posed problems*: Winston & Sons.
- Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: *Geophysical Journal International*, **160**, 195–216.
- Vogel, C. R., and M. E. Oman, 1996, Iterative methods for total variation denoising: *SIAM Journal on Scientific Computing*, **17**, 227–238.