GEOPHYSICS, VOL. 82, NO. 5 (SEPTEMBER-OCTOBER 2017); P. R287–R298, 17 FIGS., 2 TABLES.
10.1190/GEO2016-0326.1

Check for updates

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

A discrepancy-based penalty method for extended waveform inversion

Lei Fu[1] and William W. Symes[1]

ABSTRACT

Extended waveform inversion globalizes the convergence of seismic waveform inversion by adding nonphysical degrees of freedom to the model, thus permitting it to fit the data well throughout the inversion process. These extra degrees of freedom must be curtailed at the solution, for example, by penalizing them as part of an optimization formulation. For separable (partly linear) models, a natural objective function combines a mean square data residual and a quadratic regularization term penalizing the nonphysical (linear) degrees of freedom. The linear variables are eliminated in an inner optimization step, leaving a function of the outer (nonlinear) variables to be optimized. This variable projection method is convenient for computation, but it requires that the penalty weight be increased as the estimated model tends to the (physical) solution. We describe an algorithm based on discrepancy, that is, maintaining the data residual at the inner optimum within a prescribed range, to control the penalty weight during the outer optimization. We evaluate this algorithm in the context of constant density acoustic waveform inversion, by recovering background model and perturbation fitting bandlimited waveform data in the Born approximation.

INTRODUCTION

Seismic full-waveform inversion (FWI) is able to yield high-resolution images of subsurface structure by iteratively minimizing the difference between predicted data and observed data (Virieux and Operto, 2009; Vigh et al., 2010, 2013). However, the success of FWI depends on an initial model of the earth sufficiently accurate to predict times of data event arrivals to within a half-wavelength for frequencies with adequate signal-to-noise ratio. These attributes are available for some surveys, but not for others (Plessix et al., 2010). Without them, the data fit is poor from the outset, with event times in error by more than a wavelength ("cycle-skipped"), and the model estimate tends to stagnate far from kinematic accuracy.

Extended waveform inversion enlarges the model with nonphysical degrees of freedom, in such a way that data fit may be achieved (hence cycle-skip is avoided) throughout the inversion, even with a grossly inaccurate initial model (Symes, 1986, 2008; Symes and Carazzone, 1991). Because the additional degrees of freedom are nonphysical, they must be suppressed if the extended model is to converge to a solution of the FWI problem, which must necessarily be described only by the parameters of the chosen wave physics. Thus, an optimization formulation of extended waveform inversion must aim to drive a measure of model extension (nonphysicality) p to zero as the iteration progresses, while simultaneously minimizing a measure of data misfit e. A common approach, followed here, is to combine e and p into a single objective function of the (extended) model, $J_\alpha = e + \alpha p$, and to minimize J_α. Choice of the penalty weight α is critical in determining the behavior of such an algorithm: If α is too large, extended models are forced to be essentially physical (not extended), and J_α takes on the characteristics of a typical FWI objective. If on the other hand α is too small, then convergence can be very slow.

The main contribution of this paper is to describe an approach to dynamic control of the penalty parameter α based on discrepancy, that is, on the size of the data misfit e. Maintaining the size of the data residual in a prescribed range throughout the model updating process bounds the error in the final predicted data, and increases α so that p decreases and the final predicted model is close to physical, thus to a solution of the FWI problem with data residual in the prescribed range. Use of a target range, rather than a target value, allows several iterations of a standard nonlinear optimization algorithm to be taken with constant α. The iteration continues through α updates by warm starting the next sequence of iterates (Fu and Symes, 2016).

Because many extended models minimize the data misfit e, the form of the objective $J_\alpha = e + \alpha p$ can be viewed as additive regularization of the data fit e by the penalty term αp, a very well-known approach to the solution of inverse and ill-posed problems (Engl et al., 1996). However, our approach to selection of α differs

Manuscript received by the Editor 20 June 2016; revised manuscript received 17 April 2017; published ahead of production 25 May 2017; published online 05 July 2017.
[1]Rice University, The Rice Inversion Project, Houston, Texas, USA. E-mail: lei.fu.rice@gmail.com; symes@caam.rice.edu.
© 2017 Society of Exploration Geophysicists. All rights reserved.

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

fundamentally from well-established methods, such as the L-curve method (Hansen, 1992), generalized cross validation (Wahba, 1977), and Morozov's (1984) discrepancy principle (Engl et al., 1996). The role of p in these works is to control the otherwise overwhelming influence of data noise on the minimizer of J_α: α and should be chosen such that the minimizer is close to an assumed noise-free solution. For example, Morozov's (1984) discrepancy principle (Engl et al., 1996) is a realization of this idea: Roughly speaking, it chooses an optimal value of α so that the value of p at the minimizer of J_α should be as small as possible subject to a bound on e expressing data variance. This concept appears many times in the geophysical literature in various guises (Jackson, 1972). Roy (2002, 2005), Sen and Roy (2003), and Ajo-Franklin et al. (2007) discuss various methods for choosing regularization parameters in geophysical inverse problems, Morozov's discrepancy principle among them, and note that data variance is often difficult to estimate a priori in geophysical inverse problems.

In contrast, extended inversion should achieve $p = 0$ (or close to it), signifying a physical minimizer (solution of the FWI problem). This value of p is not an indirect expression of unknown data variance, but rather it is a requirement of the physics chosen to represent the wave phenomena under study. The algorithm explained here uses J_α for many values of α, rather than a single, optimal value of α, to attain this objective. Data noise level or variance plays a role in our algorithm: An assumed value is used to set the admissible range for the data discrepancy e, thereby eliminating any danger of cycle skipping. It is actually possible to use the requirement that $p \to 0$ to extend this algorithm to estimate data noise level, as we shall argue in the "Discussion" section. For the purposes of this paper, we suggest a simple and practical estimate of data noise level, from which we derive the target range for e, as the misfit level attained by a "reasonable" computational effort for the problem with $\alpha = 0$ and a more or less arbitrary initial model estimate. Here, reasonable means roughly the effort which we intend to devote to each iteration (objective or gradient evaluation). The net result is that the final, near-physical solution estimate fits the data as well as the initial estimate did, at roughly the same cost.

We note that dynamic control of penalty parameters is not a new idea (see, for instance, Roy, 2002, 2005). However, the underlying principle of our algorithm, based on driving the penalty term p to zero at the solution, does not seem to have been used before.

We use the constant density acoustic extended Born model as the framework for our examples of discrepancy-based inversion. This extended model shares a natural Cartesian product structure with many other extended models (Symes, 2008): Its model vector has two components, inner and outer. The inner or dynamic component is responsible for initiating propagating waves: For the acoustic Born model, this is the velocity perturbation or reflectivity. The outer or kinematic component governs the propagation of waves; in our example, it is the velocity macromodel. The data-prediction operator is linear in the inner or dynamic component. Because the objective summands e and p are quadratic in our examples, minimization over the inner variables is a quadratic optimization problem. The variable projection method (Golub and Pereyra, 2003) takes advantage of this feature to create a reduced objective, the optimum value of J_α over the inner variables, which is a function of the outer variables only, and whose global optimum occurs precisely at the outer component of the global optimum of J_α. Beyond convenience, use of the reduced objective is actually essential for computational efficiency (for instance, see Kern and Symes, 1994; Huang and Symes, 2015).

Other examples of separable extended scattering models are either Born approximations to more complicated scattering physics or modifications of the energy source mechanism that violate the modeled data acquisition scheme. For recent examples of the first type, see Symes (2008), Weibull and Arntsen (2013), Biondi and Almomin (2014), Liu et al. (2014), and Lameloise et al. (2015), also for a review of older work. Plessix (2000), Luo and Sava (2011), van Leeuwen and Herrmann (2013), and Warner and Guasch (2014) describe various examples of the second type of "source" extension. Note that the concept of "model extension," as we describe it, is really a very old idea, implicit in the practice of seismic velocity analysis from its inception.

The remainder of the paper begins with a "Theory" section describing the separable model structure, the variable projection algorithm, the discrepancy principle, and our variable-α algorithm, in abstract form. The following "Examples" section first explains the concrete form taken by the algorithm components for the constant density acoustic Born model and its subsurface offset extension. We then describe the application of this version of the algorithm to an example based on the SEG-EAGE overthrust model. We mention various unresolved issues and possible extensions in the penultimate section, and end by reiterating the conclusions of this study.

THEORY

This section is organized as follows: First, we introduce the separable inverse problem and variable projection method; then, we investigate the role that weight α plays in objective function; we show how to interpret discrepancy principle as a parameter choice rule to keep residual in an acceptable range.

Separable inverse problems

In this section, we present an abstract formulation of the key ideas mentioned in the "Introduction" section. We will give the various components concrete form appropriate for acoustic seismic modeling at the beginning of the next section.

In this formulation, a model is a pair consisting of an outer parameter m and an inner parameter x. The forward (data simulation and modeling) operator F is linear in x and (possibly) nonlinear in m, as reflected in our notation for its evaluation:

$$m, x \mapsto F[m]x. \tag{1}$$

The value $F[m]x$ is a vector in the space of data.

The objective has the form described in the introduction, that is, a linear combination of a data error or misfit term e and a penalty term p, the latter applying only to the inner variables. We will assume that the model and data spaces are Hilbert spaces, with inner products $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We will use the same notation for inner and outer models and data spaces, distinguishing the (possibly different) norms by context. The data error and model penalty are norms squared:

$$e[m, x] = \frac{1}{2}\|F[m]x - d\|^2, \tag{2}$$

$$p[m, x] = \frac{1}{2}\|Ax\|^2, \tag{3}$$

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

and the objective is their weighted sum (the factor of $1/2$ is for later computational convenience):

$$J_\alpha[m,x] = e[m,x] + \alpha p[m,x]. \tag{4}$$

In concrete instances, the regularization operator A measures the extent of nonphysicality of the inner parameter x. Physical inner parameter vectors lie in its null space. As explained in the "Introduction" section, the object of the optimization is to remove the nonphysical extended degrees of freedom from the model while maintaining the data fit. Reducing p as defined in equation 3 will move the model toward the null space of A, which is exactly the physical subspace of the inner parameter vector.

The weight α controls the amount of penalty applied for model extension. When $\alpha \to 0$, the objective function expresses little constraint on model extension and allows a good data fit. When $\alpha \to \infty$, minimization of J_α forces the extended model to be close to the physical one, so that the optimization approximates nonextended or physical inversion.

Variable projection method

As mentioned in the "Introduction" section, the separable nature of this least-squares inverse problems invites use of the variable projection method, a nested optimization approach. First, in the inner loop, the objective function is optimized over linear parameter x with the nonlinear parameter m fixed. The gradient of the objective function $J_\alpha[m,x]$ with respect to x is

$$\nabla_x J_\alpha[m,x] = F[m]^T(F[m]x - d) + \alpha A^T A x, \tag{5}$$

where T denotes the transpose. A stationary point of equation 4 satisfies the normal equation

$$(F[m]^T F[m] + \alpha A^T A)x = F[m]^T d. \tag{6}$$

We presume that this system is positive definite for all values of $\alpha \geq 0$, and in particular that $F[m]^T F[m]$ is also positive definite. In principle, this condition is almost never satisfied, and it is necessary that the forward map be regularized. In practice, we ignore this ill-conditioning of aperture- and bandwidth-limited forward maps, and we treat the normal operator as if it were positive definite. System 6 thus has a unique solution that can be approximated via an iterative method such as conjugate gradient (CG) iteration. Because m and α determine the operator on the left side of equation 6, its solution becomes a function $x[m,\alpha]$ of these quantities.

We minimize the reduced objective $J_\alpha[m,x[m,\alpha]]$ over m via a gradient-based method, such as steepest descent, Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS), or Gauss-Newton iteration (Nocedal and Wright, 1999). The gradient of the reduced objective function $J_\alpha[m,x[m,\alpha]]$ with respect to m is (Golub and Pereyra, 1973)

$$\nabla_m J_\alpha[m,x[m,\alpha]] = (DF[m]^T(F[m]x[m,\alpha] - d))x[m,\alpha]. \tag{7}$$

The notation DF signifies the derivative of F. Because F is an operator-valued function of m, its directional derivative at m in the direction δm, denoted $DF[m]\delta m$, is an operator of the same type. We have used the notation $DF[m]^T$ to denote one of the possible meanings of "transpose" for this operator-valued function: In terms of the inner products $\langle\cdot,\cdot\rangle$ in the various model and data spaces

$$\langle (DF[m]^T d)x, \delta m\rangle = \langle d, (DF[m]\delta m)x\rangle \tag{8}$$

for any data vector d, outer parameter perturbation δm, and inner parameter vector x.

The discrepancy principle

As described in the "Introduction" section, the discrepancy principle (in one of its guises) involves setting an acceptable range of data misfit $[X_-, X_+]$ and adjusting the penalty weight α so that $e[m,x[m,\alpha]]$ lies in this range. The principle so stated applies to the inner optimization over x: Because updating m changes the inner problem, the appropriate condition for such separable problems is that e stays in the range $[X_-, X_+]$ as m is updated. In this subsection, we examine the dependence of e on α with a view to understand how to reset α when m changes.

Accordingly, we regard m as fixed and suppress it from the notation for the remainder of this subsection, and we introduce the abbreviations

$$e(\alpha) = e[m,x[m,\alpha]], \tag{9}$$

$$p(\alpha) = p[m,x[m,\alpha]]. \tag{10}$$

Differentiating equation 6 with respect to α leads to the relation

$$(F^T F + \alpha A^T A)\frac{dx}{d\alpha} = -A^T A x, \tag{11}$$

whence

$$\begin{aligned}\frac{de}{d\alpha} &= \left\langle \frac{dx}{d\alpha}, F^T(Fx-d)\right\rangle \\ &= -\alpha\left\langle \frac{dx}{d\alpha}, A^T A x\right\rangle \\ &= \alpha\langle A^T A x, (F^T F + \alpha A^T A)^{-1} A^T A x\rangle \\ &\geq 0.\end{aligned} \tag{12}$$

Note that the inequality in equation 12 is strict if $p > 0$, hence $A^T A x \neq 0$ because the normal operator is assumed to be positive definite. The derivative of $p(\alpha)$ with respect to α is

$$\begin{aligned}\frac{dp}{d\alpha} &= -\langle A^T A x, (F^T F + \alpha A^T A)^{-1} A^T A x\rangle \\ &\leq 0.\end{aligned} \tag{13}$$

Similarly, there is a strict inequality if $p > 0$.

Equation 12 together with equation 13 shows that increasing α implies increasing e, whereas decreasing p, and

$$\begin{aligned}&\langle A^T A x, (F^T F + \alpha A^T A)^{-1} A^T A x\rangle \\ &= \langle (A^T A)^{1/2} x, [(A^T A)^{-1/2} F^T F (A^T A)^{-1/2} + \alpha I]^{-1} (A^T A)^{1/2} x\rangle \\ &\leq \frac{1}{\alpha}\langle A^T A x, x\rangle = \frac{2}{\alpha} p.\end{aligned} \tag{14}$$

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

In view of equation 12

$$\frac{de}{d\alpha} \le 2p, \tag{15}$$

with this inequality also being strict if $p > 0$.

Suppose the current weight is α_c and denote a candidate for an updated weight by α_+. Then, from inequality 15,

$$e(\alpha_+) - e(\alpha_c) \le \int_{\alpha_c}^{\alpha_+} 2p d\alpha. \tag{16}$$

If $\alpha_+ \ge \alpha_c$, then in view of inequality 13, the above is

$$\le 2p(\alpha_c)(\alpha_+ - \alpha_c). \tag{17}$$

Let us suppose that $e(\alpha_c) < X_+$. Then, setting

$$\alpha_+ = \alpha_c + \frac{X_+ - e(\alpha_c)}{2p(\alpha_c)} \tag{18}$$

implies via inequality 17 that

$$e(\alpha_+) - e(\alpha_c) \le X_+ - e(\alpha_c). \tag{19}$$

Assuming that $p(\alpha_+) > 0$, hence $p(\alpha) > 0$ for $\alpha_c \le \alpha \le \alpha_+$, we conclude that if α_+ is given by the rule 18, then

$$e(\alpha_c) < e(\alpha_+) \le X_+. \tag{20}$$

That is, unless $p(\alpha_+) = 0$ (in which case a physical solution of the inverse problem has been reached), $e(\alpha_+)$ is larger than $e(\alpha_c)$ but in any case does not exceed X_+. Rule 18 therefore provides a feasible updated α consistent with the upper bound in the discrepancy principle.

Practical application of the discrepancy principle

As mentioned in the "Introduction" section, the discrepancy principle requires that a range of data noise (in our notation, one-half data noise squared) $[X_-, X_+]$ be given. We base our algorithm on a data error estimate X, and set $X_- = \gamma_- X, X_+ = \gamma_+ X$, where $\gamma_- < 1 < \gamma_+$ are positive constants at the disposal of the user: Typical values might be $\gamma_- = (0.7)^2, \gamma_+ = (1.2)^2$ (we use these values in the experiment reported below).

The choice of data error estimate X remains. Two approaches to this choice are (1) treat it as a hypothetical, with all subsequent results being contingent on it, and choose initial model $(m, x[m, 0])$ for $\alpha = 0$ so that $X = e(0) = e[m, x[m, 0]]$; (2) if the normal equation 6 is solved approximately by an iterative method (we used CG iteration), choose several iterations to be used throughout and choose $x[m, 0]$ as an approximate solution of equation 6 computed by the chosen number of iterations, and $X = e(0) = e[m, x[m, 0]]$. We used the second method in the experiments described in the next section. Of course, the second approach is really a variant of the first, with an indirect rather than a direct choice of the hypothetical noise level X. Either approach makes sense only for inverse problems of the character described in this paper, in which the unconstrained ($\alpha = 0$) data misfit may be made arbitrarily small by choice of the linear parameter x, for any choice of the nonlinear parameter m. Note the implication for X: In principle, $e(0) = 0$! Therefore, a nonzero X must be selected, which embodies the actual (and initially unknown) data error e for $p = 0$ and a corresponding linear parameter x for which $e(0) = e[m, x]$. The second approach outlined above does this in a "natural" way. However, the arbitrariness of the choice cannot be avoided. The progress of the algorithm and its end result clearly depend on X. We will address this dependence and its implications in the "Discussion" section.

With either approach to the choice of X, the algorithm proceeds as follows (a flowchart is shown in Figure 1):

0) Choose initial m, set $\alpha = 0$, compute $x[m, 0]$ by (approximate) solution of the linear least-squares problem 6, $X = e[m, x[m, 0]]$, $X_\pm = \gamma_\pm X$.
1) While (not done),
   1.1) While $e[m, x[m, \alpha]] \in [X_-, X_+]$, update m by means of a continuous optimization algorithm, using the gradient as given in equation 7; for each update of m, a solution of equation 6 is required to recompute $x[m, \alpha]$.
   1.2) If $e[m, x[m, \alpha]] > X_+$, exit
   1.3) if $e[m, x[m, \alpha]] < X_-$,
      1.3.1. Compute α_+ by equation 18, solve equation 6 to compute $x[m, \alpha_+]$.
      1.3.2. While $e_+ = e[m, x[m, \alpha_+]] \notin [X_-, X_+]$,
         1.3.2.1. If $e_+ < X_-$, set $\alpha_+ \leftarrow \alpha_+ * 2$
         1.3.2.2. If $e_+ > X_+$, set $\alpha_+ \leftarrow \alpha_+ / 1.5$
         1.3.2.3. In either case, solve equation 6 to compute $x[m, \alpha_+]$, calculate $e_+ = e[m, x[m, \alpha_+]]$
      1.3.3. Set $\alpha \leftarrow \alpha_+$

The secant update of α (equation 18, step 1.3.1 above) might seem unnecessarily elaborate, in comparison with the simple bisection loop (steps 1.3.2.1 and 1.3.2.2). However, equation 18 gives a sensible update when $\alpha = 0$, and thus makes the algorithm self-starting. Moreover, our experience is that the bisection loop is seldom invoked because the secant update is usually successful. Because each α update requires a solution of the normal equation 6, this is a good thing.

EXAMPLES

In this section, we illustrate the performance of the discrepancy-based algorithm by solving a velocity estimation problem modeled on reflection seismology. We use the subsurface offset extension of the 2D Born (linearized) constant density acoustic model. The forward-modeling operator F in our case is the subsurface extended Born modeling operator. For a detailed description of this model, its origins, and its properties (see Symes, 2008). In this separable model, the acoustic wave velocity field v is the nonlinear parameter (denoted m in the discussion above), and the reflectivity r (proportional to the perturbation of v) is the linear parameter (x in the abstract discussion). The quantities appear as coefficients in the wave equations satisfied by the pressure field u and its perturbation δu:

$$\left(\frac{\partial^2}{\partial t^2} - v(x,z)^2\nabla^2\right) u(t,x,z;x_s) = w(t)\delta(x - x_s)\delta(z - z_s),$$
$$u = 0, \quad t \ll 0, \tag{21}$$

$$\left(\frac{\partial^2}{\partial t^2} - v(x,z)^2\nabla^2\right)\delta u(t,x,z;x_s) = \int_{-H}^{H} dh r(x,z,h)\nabla^2 u(t,x+2h,z;x_s),$$
$$\delta u = 0, \quad t \ll 0. \tag{22}$$

Note that the velocity v depends on the spatial coordinates x, z, whereas the reflectivity r depends on another coordinate h, representing subsurface (half-)offset. The physical meaning of this dependence is that action-at-a-distance is permitted in this model: To first order in perturbation theory, strain at one space-time position $(x-2h, z)$ is allowed to cause stress instantaneously at a different position at the same depth (x, z). The introduction of spatial shift compensates velocity errors, which permits data to be fit well for arbitrary v. Thus, this model has the feature required by our construction of the discrepancy-based algorithm.

The right side of equation 21 represents an isotropic point radiator located at $x = x_s, z = z_s$ with time dependence $w(t)$.

The forward map $F[v]r$ is defined by

$$F[v]r(x_r, z_r, t; x_s) = \delta u(x_r, z_r, t; x_s), \tag{23}$$

in which x_r, z_r range over the receiver positions of the modeled survey.

We note that the Born forward map in equation 23 is the derivative of a nonlinear forward map (Symes, 2008; Biondi and Almomin, 2014). The linearization error, that is, the amount by which the linear prediction based on the derivative differs from the actual perturbation of the nonlinear forward map, is smallest when (1) the background velocity is transparent, that is, contains no reflectors, and (2) the reflectivity is oscillatory on the wavelength scale. Thus, the most accurate Born approximation to a velocity field uses a spatial average to produce a background model, and it takes the difference between the average and the original velocity fields as the reflectivity. This is the procedure we followed in creating the Born model described below. It is in principle possible to formulate an extended modeling approach based directly on the equations of acoustics, without the intervening linearization used here (Symes, 2008; Biondi and Almomin, 2014). Such an algorithm is beyond the scope of this paper.

If the reflectivity is concentrated or focused at $h = 0$, that is, $r(x, z, h) = r_0(x, z)\delta(h)$, then the perturbation wave equation 22 reduces to the ordinary perturbation equation of linearized acoustics, and the model to the ordinary acoustic Born model. The aim of the inverse problem is to fit data with an appropriate velocity and a physical reflectivity. Thus, an appropriate choice of annihilator is

$$Ar(x, z, h) = hr(x, z, h). \tag{24}$$

Because $h\delta(h) = 0$, the null space of the operator A defined by equation 24 is precisely the collection of r having a factor of $\delta(h)$. This choice of A penalizes energy in the nonphysical dimension ($h \neq 0$), and it has been used in many prior works on subsurface offset extended waveform inversion (Shen et al., 2003; Shen and Symes, 2008; Symes, 2008; Biondi and Almomin, 2012; Weibull and Arntsen, 2013; Lameloise et al., 2015).

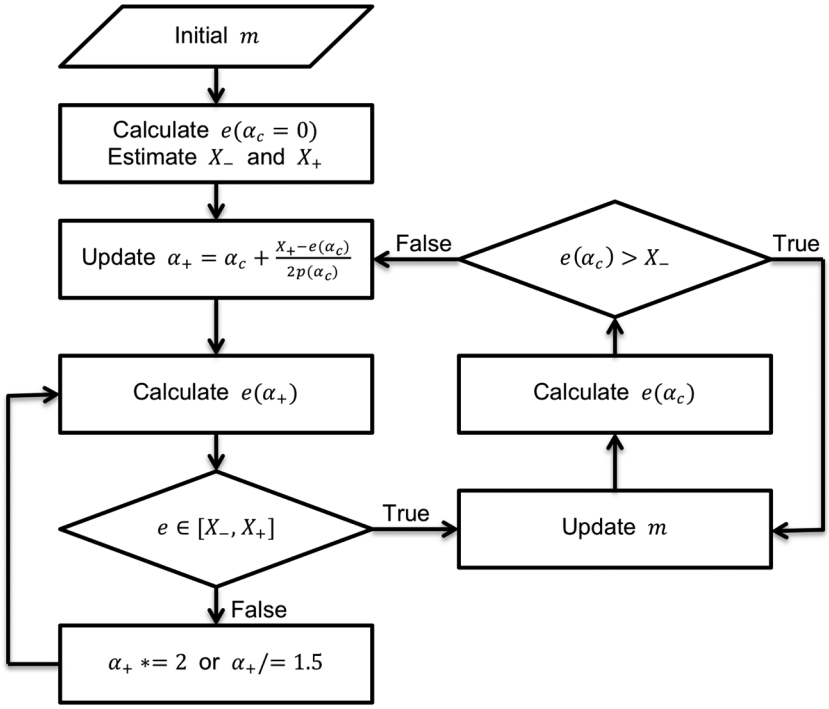

Figure 1. A flowchart for our implementation of our proposed algorithm.

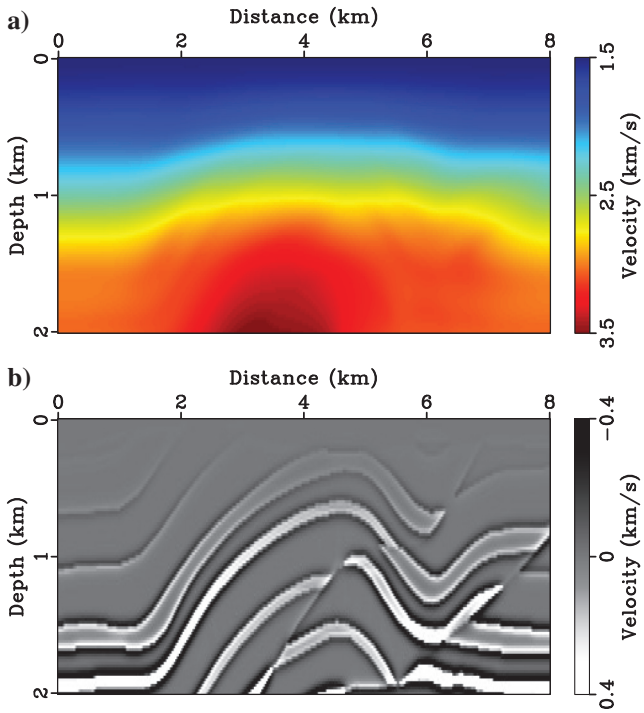

Figure 2. (a) Target background velocity model and (b) target reflectivity model.

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

In numerical implementation of wave equations 21 and 22, we use a centered finite-difference method of order two in time and eight in space (Alford et al., 1974; Kelly et al., 1976). To compute the transpose operator F^T, we use an extended-model version of the adjoint-state method (Plessix, 2006). The derivative transpose DF^T, defined in equation 8, is an essential ingredient in gradient calculation 7. We use a modified version of the adjoint method to compute its value (Symes and Santosa, 1988; Kern and Symes, 1994). We note that DF^T goes under the name "tomographic operator" in the literature on wave-equation migration velocity analysis (Biondi and Sava, 2004).

The example is modified from the SEG/EAGE 3D overthrust model (Aminzadeh et al., 1997). We added a 400 m water layer to the top, and we padded it by a sufficient number of cells to eliminate boundary reflections from the measured data traces (at positions x_r, z_r). We smoothed the model to create a transparent background or macromodel (shown in Figure 2a). The difference between the original model (with the water layer and padding) and the background model constitutes the reflectivity model (shown in Figure 2b). In the reflectivity model, horizontal layers are distorted by several thrust (reverse) faults (shown in Figure 2b). The background velocity increases with depth (Figure 2a). The smoothed velocity is higher in the center, where the anticline structure sits. The basic dimensional information is listed in Table 1.

Table 1. Parameters for the overthrust model example.

Parameter	Measurements
Source wavelet	Band-pass 5–20 Hz
Source position $\mathbf{x_s}$	x:1–7 km every 40 m, $z = 40$ m
Receiver position $\mathbf{x_r}$	x:0–8 km every 40 m, $z = 0$ m
Space and time	$x = 8$ km, $z = 2$ km, $t = 3$ s
Grid size	$dx = dh = dz = 20$ m, $dt = 2$ ms
Initial velocity	$v = 1.5$ km/s
Max iter inner loop	20

Figure 3a shows that the initial background velocity is constant (1.5 km/s) and far away from the correct one, so the geologic structures can barely be discerned from the corresponding extended image (Figure 3b). In the subsurface offset gather, the downward curves indicate slow velocity.

First, given initial velocity, set weight $\alpha = 0$, solving the normal equation by 20 iterations of CG gives the data residual term $e(0) = 2.81e-2$, which approximately equals 5.7% of the relative data residual. Based on the value of $e(0)$, the upper and lower bounds of the acceptable range are estimated by the formulas $X_- = 0.7^2 e(0)$ and $X_+ = 1.2^2 e(0)$. Then, the updated weight $\alpha_1 = 1.1e-6$ is estimated in equation 18. After one velocity update, the data residual drops below the lower bound X_-, so we updated weight $\alpha = 4.1e-6$ again. The residual is still smaller than X_-. According to the update algorithm, the value of α was doubled three times, which gives a data residual in the acceptable range. After three more velocity updates with weight $\alpha = 3.2e-5$, the data residual became smaller than X_-. Recalculating α and doubling its value made the data residual go back to the acceptable range. After nine more velocity updates, the data residual dropped below the lower bound; the updated weight $\alpha = 2.4e-4$ provides data residual in acceptable range.

The inverted background velocity model from 20 iterations and the extended reflectivity are shown in Figure 4a and 4b.

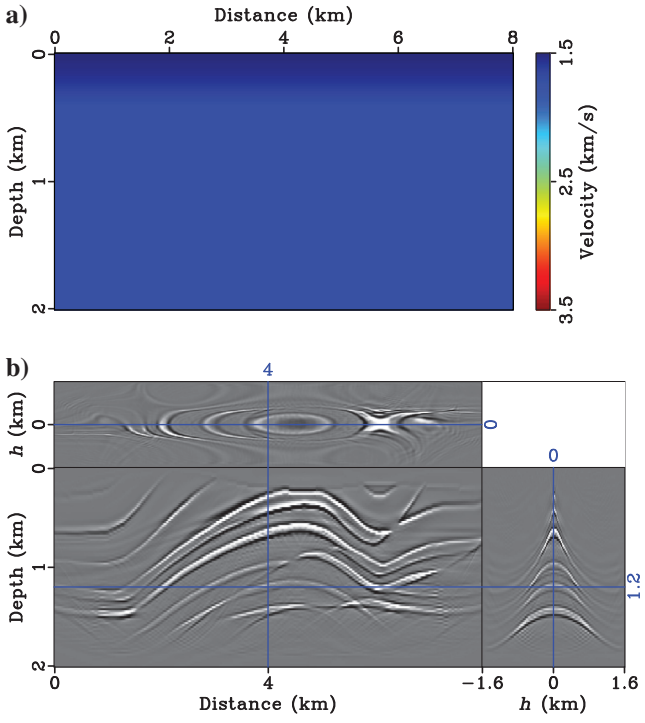

Figure 3. (a) Initial background velocity and (b) extended reflectivity.

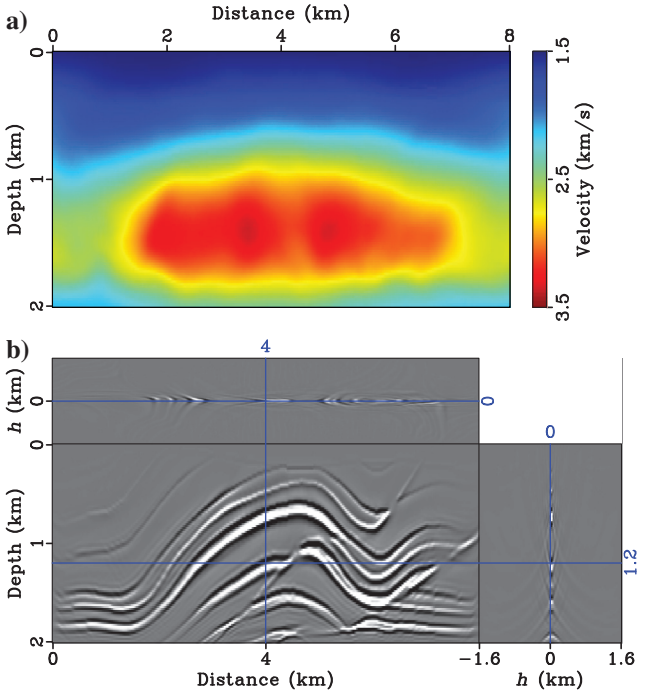

Figure 4. (a) Inverted background velocity and (b) extended reflectivity.

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

The anticline and reverse fault structures can be clearly observed. Furthermore, even the reflector beneath the anticline is positioned correctly (at approximately $x = 4$ km, $z = 2$ km). The velocity error is mostly at the edges, which is a result of imperfect illumination.

The subsurface offset gathers are much more focused toward $h = 0$ after inversion (Figure 5). With the inverted background velocity, the geologic structures are imaged with much higher resolution at the zero-offset section of the reflectivity model (Figure 6). Extending the reflectivity model permits a good data fit throughout the inversion process (Figures 7 and 8). Note that the data misfit before and after inversion is measured with different penalty weight α. In other words, a good data fit is obtained by models with different amounts of extension. Note that to improve the computational efficiency, the adequate subsurface offset range was estimated adaptively by measuring data fit throughout the inversion process (Fu and Symes, 2015, 2017b).

We increase the penalty weight α at iteration numbers 1, 4, and 13, which significantly accelerates the convergence rate, as shown in Figure 9. At the same time, according to the discrepancy principle, the data residual stays in an acceptable range (see Figure 10).

The data misfits plotted in the figures shown so far are extended residuals; that is, they are computed with the extended reflectivity model. Because these are nonphysical and still have some residual energy in $h \neq 0$, as shown in Figure 5b, it is natural to wonder if the plots represent a fair assessment of the quality of the inverted velocity (Figure 4a). To give a more rigorous evaluation, we have used the initial (Figure 3a) and inverted velocities in least-squares migration, which amounts to solving problem 6 with the constraint that the reflectivity be physical; that is

$$r(x, z, h) = r_0(x, z)\delta(h). \tag{25}$$

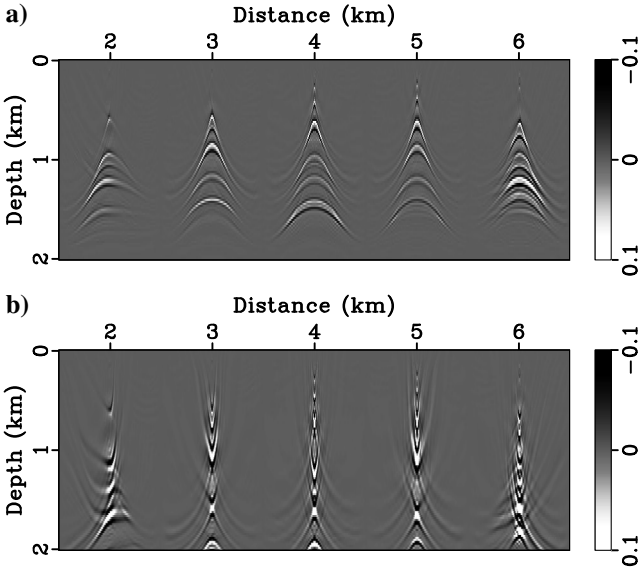

Figure 5. (a) Subsurface offset gathers at $x = 2$, 3, 4, 5, and 6 km using the initial background velocity and (b) subsurface offset gathers using the inverted background velocity.

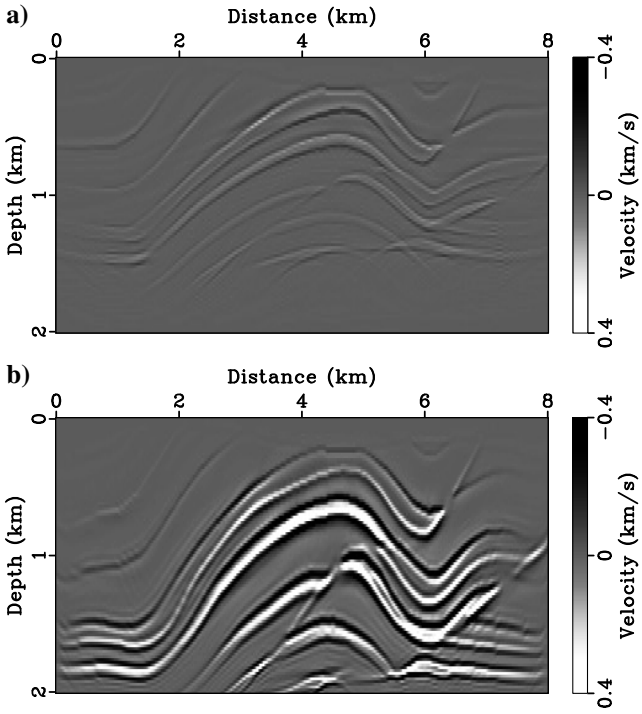

Figure 6. (a) Zero-offset section with the reflectivity model using the initial background velocity. (b) Zero-offset section with the reflectivity model using the inverted background velocity.

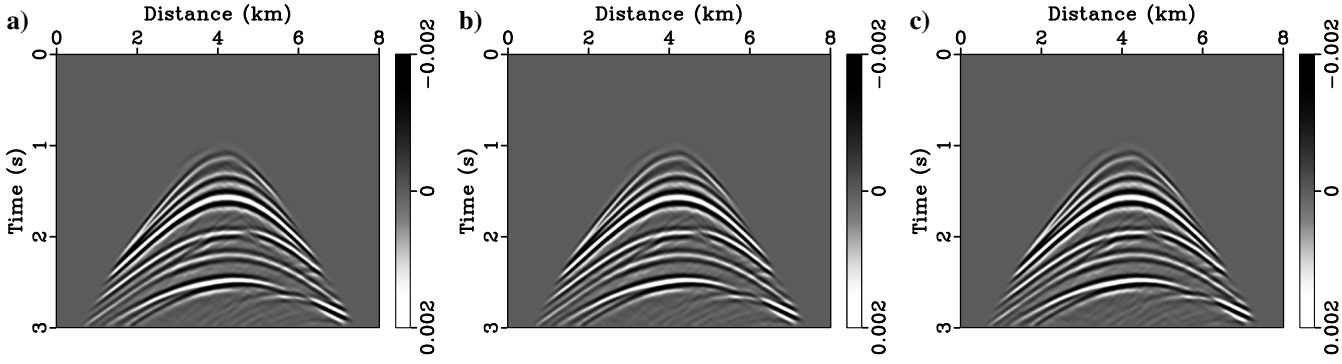

Figure 7. Data gather for a shot at the center of the model. (a) Observed data. (b) Predicted data with the initial background velocity model. (c) Predicted data with the inverted background velocity model.

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

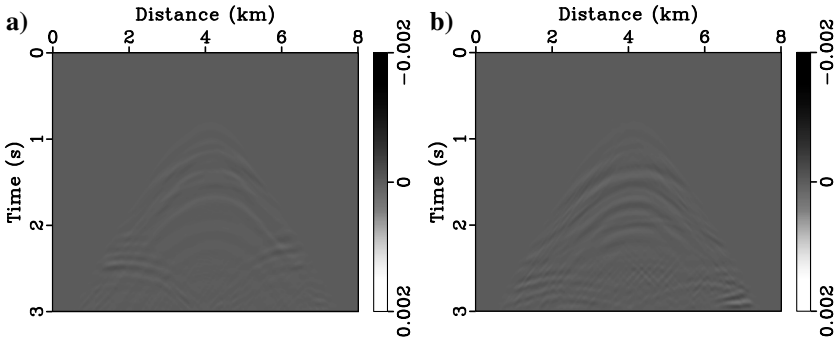

Figure 8. Comparison of data residual (a) with the initial background velocity model and (b) with the inverted background velocity model.

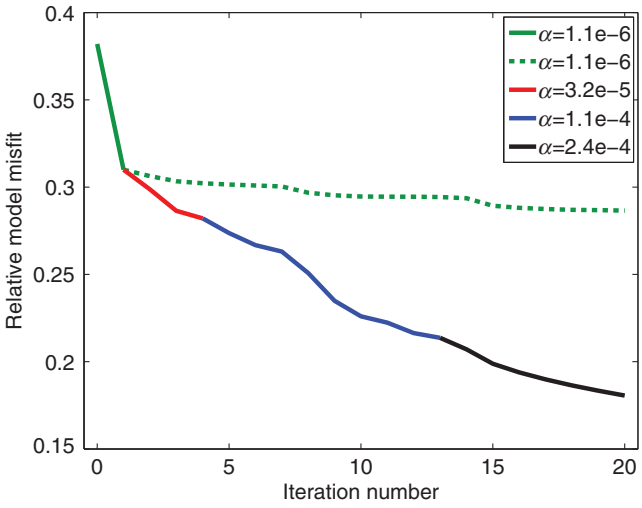

Figure 9. Model convergence rates versus iteration numbers: Different colors represent different values of weight α; dashed lines show convergence rate without changing α.

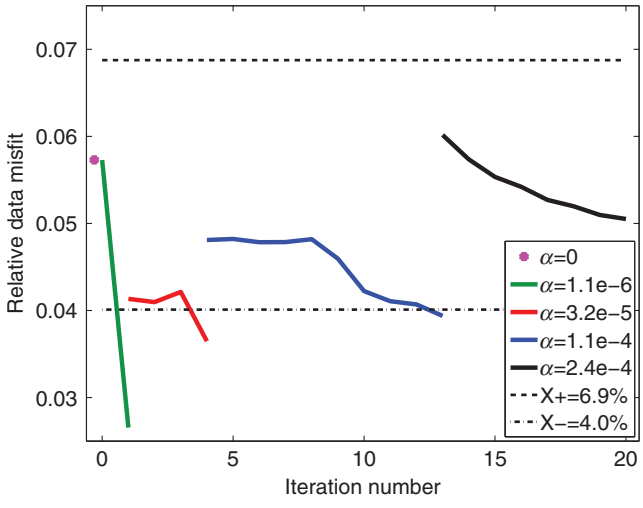

Figure 10. Relative data misfit versus iteration: Different colors represent different values of weight α.

We used 20 CG iterations to approximate these minimizations, as we have in other instances of problem 6. The resulting predicted data appear as Figure 11a and 11b, and the data residuals appear as Figure 12a and 12b, all plotted on the same gray scale. The corresponding physical reflectivities are plotted in Figure 13a and 13b. Although the residual is clearly not as small as that attained by the extended model (Figure 8b), the improvement of Figure 12b over 12a is obvious.

The magnitude of the residual presented in Figure 12b suggests that the model 4a is sufficiently accurate that no cycle-skipping occurs. To better assess the phase accuracy of this model, we present in Figure 14a–14c the traces of the target data (in black) versus the corresponding predicted traces from the reflectivity estimated by least-squares migration with the initial velocity model, in red (Figure 11a). This model is severely cycle skipped. The evidence of cycle skip that turns up in the least-squares migration residual traces is not mispositioning of events — these are time traces, and the migration and modeling are consistent, if wrong. Instead, it is the gross mismatch of amplitudes, due to destructive cancellation in the stack that forms the reflectivity, from which the traces are simulated.

Similarly, we show in Figure 15a–15c the traces of the target data (in black) versus the corresponding predicted traces from the reflectivity estimated by least-squares migration with the inverted velocity model, in red (Figure 11b). The amplitudes are comparable now, and indeed the traces are very close, except for some mild amplitude discrepancies in a few places (e.g., the 4 km trace at approximately 2.5 s). These are likely the result of residual kinematic inaccuracy of the inverted model 4a — as noted above, there is some evidence that the inversion is still progressing. However, it seems likely that the phases are now similar enough that the inverted model could serve as an initial model for (the Born version of) a successful FWI. The potential use of extended waveform inversion to produce initial models for FWI may be a worthwhile subject for further research.

We have also plotted least-squares migration images (physical reflectivities r_0), the input to the modeling that produces the data just described, as shown in Figure 13a and 13b. Comparison with Figure 2b suggests that although not perfect, the physical (least-squares migration) reflector positions and strengths resulting from the inverted velocity are great improvements over those resulting from the initial velocity. The slope of the line plotted in Figure 9 suggests that the iteration is not finished, and that further improvements could be expected with more iterations. Note again that the least-squares migration outputs did not figure at all in the velocity inversion: Least squares migration is used here only as a quality control.

It is also natural to wonder to what extent the model extension used in the algorithm presented here was really necessary to obtain a reasonable velocity inversion for this example. To explore this question, we conducted an analogous experiment, similar in almost all respects, but without the subsurface offset extension. That is, we restricted the reflectivity r to be physical, that is, obey condition 25, and we used the variable projection method to estimate v and r_0 (note that the least-squares migration exercise reported in the previous paragraph also constrained r by condition 25, but did not itself update v, relying instead on the inversion algorithm

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

based on subsurface offset extension to do that). The second term in the definition 4 of J_α is irrelevant because r satisfying equation 25 also satisfies $Ar(x,z,h) = r_0(x,z)(h\delta(h)) = 0$. Thus, the objective function for this experiment is J_0, which can be considered as the Born version of the usual FWI objective (Tarantola, 1984; Virieux and Operto, 2009). Huang and Symes (2015) term this algorithm as "Born waveform inversion."

We used the same optimization algorithm, steepest descent with backtracking line search, applied to the reduced objective function of v produced by the variable projection method. To carry out the reduction, we approximated the solution of the normal equation 6 with 20 iterations of CG iteration, exactly as was the case for the extended inversion. We used the same number (20) of outer iterations (v updates) as well and the same initial model (Figure 3a). Of course, there were no α updates, as α plays no role in the nonextended penalty function because $p = 0$. That is, the same amount of computational effort was expended in this nonextended inversion as in the extended inversion.

The result however is not nearly as satisfying. The inverted velocity (Figure 16a) shows little evidence of the large-scale structure of the target velocity, and the inverted reflectivity (Figure 16b) none of the smaller scale structure. In fact, as noted in the discussion of the extended inversion, the initial model is so far from the target that its predicted traveltimes for deeper reflectors are many wavelengths in error. This example shows that Born waveform inversion is just as susceptible to cycle skipping as is conventional FWI, based on the nonlinear forward map.

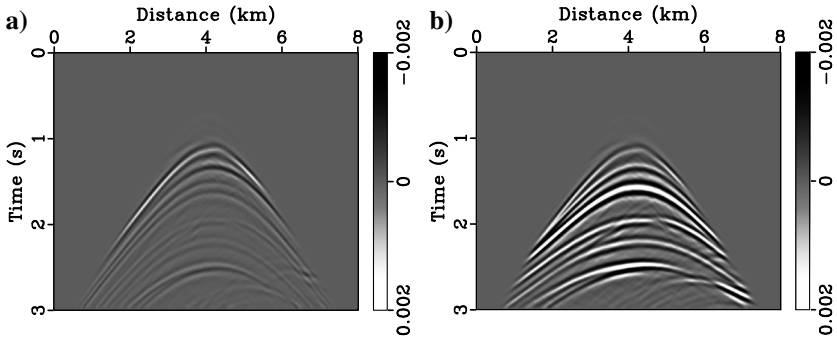

Figure 11. Least-squares migration predicted data for the same shot as shown in Figure 7a. Note that these are nonextended inversions and not byproducts of the inversion algorithm discussed in the text. Produced with 20 CG iterations. (a) Using the initial velocity model (Figure 3a). (b) Using the inverted velocity model (Figure 4a).

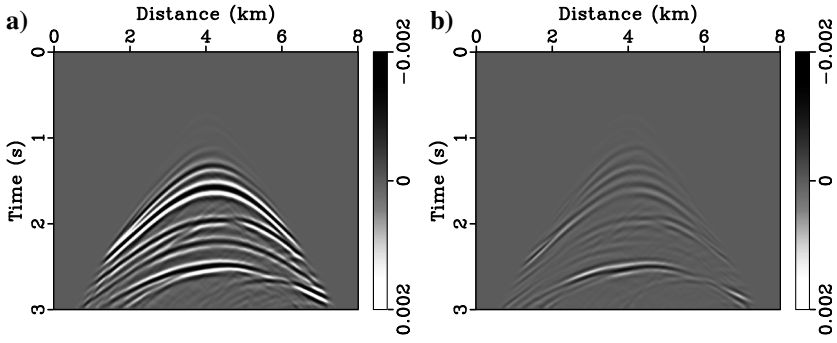

Figure 12. Least-squares migration data residuals for the same shot as shown in Figure 7a. Note that these are nonextended inversions and not byproducts of the inversion algorithm discussed in the text. Produced with 20 CG iterations. (a) Using the initial velocity model

DISCUSSION

The example presented in the last section suggests that adjusting α according to the discrepancy-based algorithm significantly improves the convergence of extended waveform inversion. To shed a little more light on the reasons for this behavior, we present a plot of the reduced objective $J_\alpha[m, x[m,\alpha]]$, for models m on the line segment $m = (1+\sigma)m_0$, $-0.5 \le \sigma \le 0.5$, in which m_0 is the target model of the previous example (Figure 2a). In defining J_α, we use the data of that example as well. The parameter σ is plotted on the horizontal axis of Figure 17, the corresponding value of J_α on the vertical axis, for three choices of α. At least restricted to this 1D model space, the behavior of J_α is as we have suggested: For small α, the objective has only one stationary point, at the global minimum, so convergence would take place from any initial guess, but become rather slow near the minimum. For larger α, the minimum is more highly resolved, but reaching it via descent requires a more accurate initial estimate. Therefore, starting with small α and increasing it as the model is updated should lead to rapid convergence and a highly resolved final model estimate, insensitive to the choice of initial model. Our discrepancy based algorithm provides a key ingredient in this procedure, not obvious from plots, such as Fig-

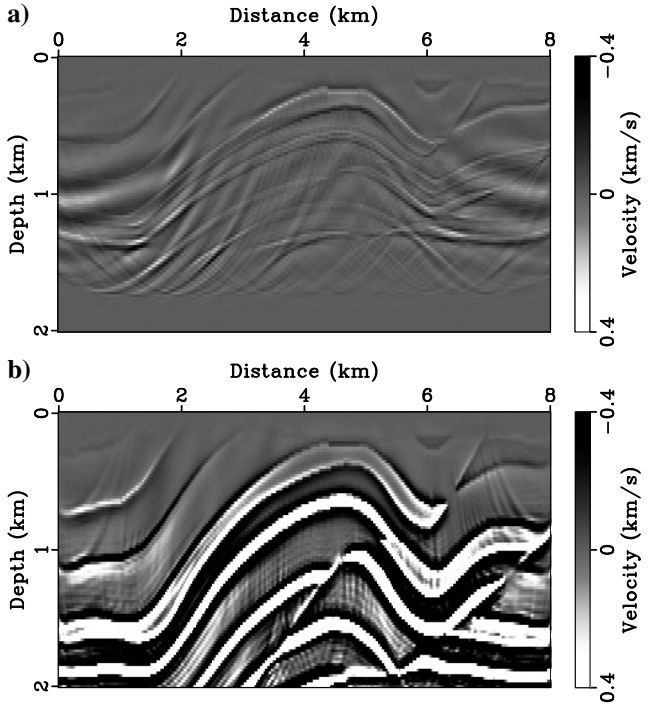

Figure 13. Least-squares migration images corresponding to Figure 11a and 11b.

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

ure 17, namely, a precise recipe for increasing α to keep the model estimate in the domain of attraction of the global minimizer.

As pointed out earlier, the secant update formula 18 makes the discrepancy-based adjustment self-starting. As the derivation shows, it produces a value of α that generates a lower value of the error term e than the target value. However, our experience suggests that it is often accurate enough to use without modification. Note that its derivation is based on the assumption that the inner component ($x[m,\alpha]$ in the notation introduced earlier) is a solution of the normal equation 6. However, we have used an iterative method (CG iteration) and a modest number of iterations to approximate the solution of equation 6. Therefore, the computed α update is contaminated by an error, the magnitude of which we have not estimated. This uncontrolled inaccuracy may not affect the eventual update of α, however. The bisection loop (steps 1.3.2.1 and 1.3.2.2 in the algorithm listing) acts to ensure a proper update satisfying the discrepancy bounds, much as bisection is used in many root-finding algorithms to ensure convergence. Also, experience suggests that modest accuracy in estimating the inner component is sufficient to generate a useful α update via equation 18.

We should point out that the variable projection gradient formula (equation 7) also relies on the solution of the system 6, and therefore is polluted by residual error from iterative approximate solution. The algorithm presented here does not control this error, potentially a serious impediment to convergence. In the example discussed in the last section, the gradient error appeared to be small enough to permit a useful approximation solution of the variable projection problem.

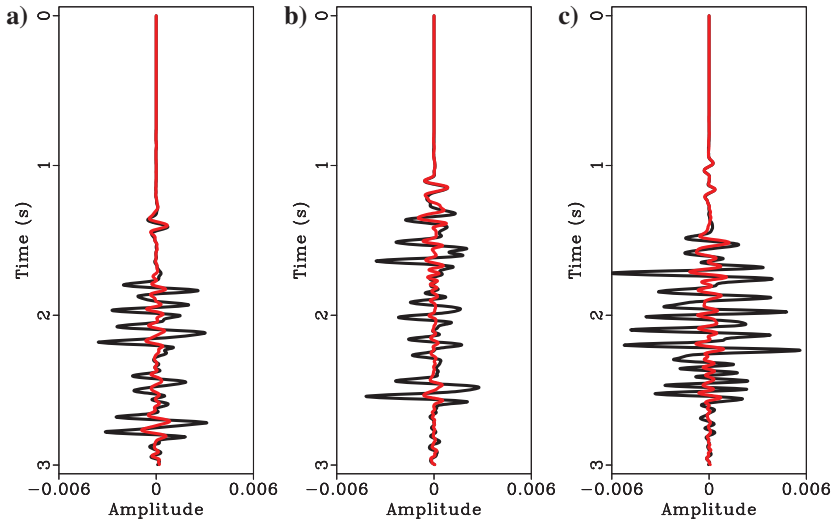

Figure 14. Traces at positions 2, 4, and 6 km from the shot gather shown in Figure 7a (black), plotted against the same traces from the predicted gather (Figure 11a) using the initial velocity model and the least-squares migration reflectivity estimate (Figure 13a) (red).

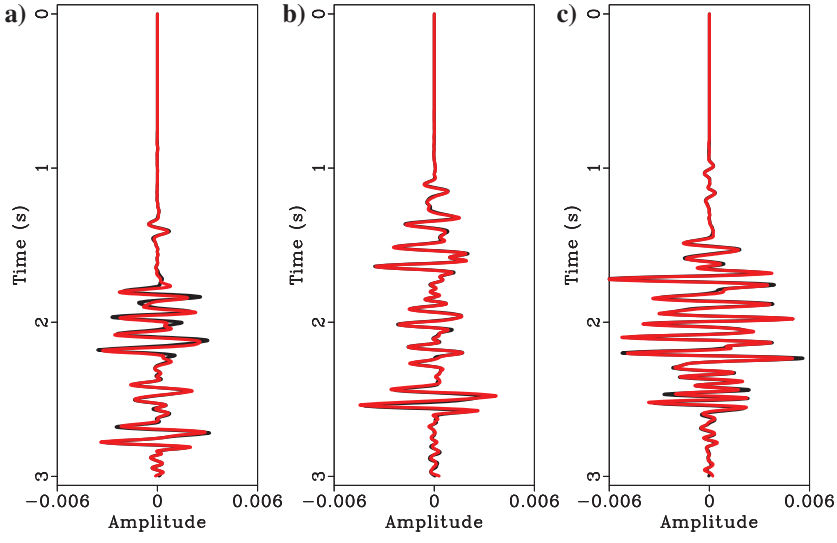

Figure 15. Traces at positions 2, 4, and 6 km from the shot gather shown in Figure 7a (black), plotted against the same traces from the predicted gather (Figure 11b) using the inverted velocity model (Figure 4a) and the least-squares migration reflectivity estimate (Figure 13b) (red).

The actual error level in the data, an estimate for which is denoted by X in the description of our algorithm, plays a central role: It must be small enough that the iterates avoid cycle skipping, but not so small as to impose impossible fit demands given the method used to solve the inner problem in the variable projection method. We have used an ad hoc procedure to pick a value of X, namely, the data-fit error e attained by the iterative solution of the inner problem for the initial outer variable (velocity, in the example), with $\alpha = 0$ and the same parameters (maximum number of iterations, tolerance for the error in satisfying equation 6) to be used in subsequent iterations. Note that our example was an "inverse crime": The same computational method was used to compute the data as was used to fit it, so in fact, the data error level was zero, and the X obtained by our procedure in this instance was an overestimate.

This observation raises the obvious question: Is it possible to correct an initially incorrect estimate for the data error level X? A modification of the algorithm presented here may answer this question in the affirmative. In principle, it is either possible to drive $p \to 0$ and converge to a limit model, with the data misfit e remaining in the range $[X_-, X_+]$ defined in the description of the algorithm above, or it is not. In the former case, by definition, the limit model is physical and solves the FWI problem, with error somewhere in the given range. Otherwise, either eventually $e > X_+$, or $e < X_-$. In the former case, the error estimate X is too small, in the latter, too large. The enhanced algorithm would then increase, respectively decrease, X by a factor >1, and restart.

This description leaves out many details that would be necessary in the formulation of a practical algorithm. For example, the condition

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/

$p \to 0$ must be replaced by $p < \epsilon$ for a tolerance ϵ, which must be chosen somehow. The completion of our sketch into a working algorithm is a subject for further research.

The computational cost of the discrepancy-based algorithm, as described here, is certainly an impediment to its use in practice: We used 20 outer and 20 inner iterations still requiring 400 modeling/migration pairs to solve a relatively tame synthetic 2D problem. Several modifications promise considerable acceleration of this algorithm, through reduction of the outer and inner iteration counts and smaller computational cost per iteration. Many local optimization methods seriously outperform steepest descent, and they should lead to more rapid convergence of the outer optimization. The standard text by Nocedal and Wright (1999) explains why, and it describes several of these methods in detail. To accelerate the inner iteration (solution of the normal equation 6), we have used a crude preconditioner, namely, multiplication of the reflectivity by z^2. Much more effective preconditioners have been explored in the past few years (see, for example, Hou and Symes, 2016). Finally, we have used a simple trick to reduce the length of the active portion of the subsurface offset (h) axis, and thereby a major contributor to the computational cost. This trick is also based on discrepancy: We reduce the maximum $|h|$ whenever it is possible without affecting the discrepancy (e) seriously. In another work (Fu and Symes, 2015, 2017a, 2017b), we have shown how to combine control of maximum $|h|$ with low-to-high frequency continuation to reduce the cost of extended waveform inversion by one to two orders of magnitude. All of these cost-reduction methods, and more, will be essential in moving toward a practical algorithm.

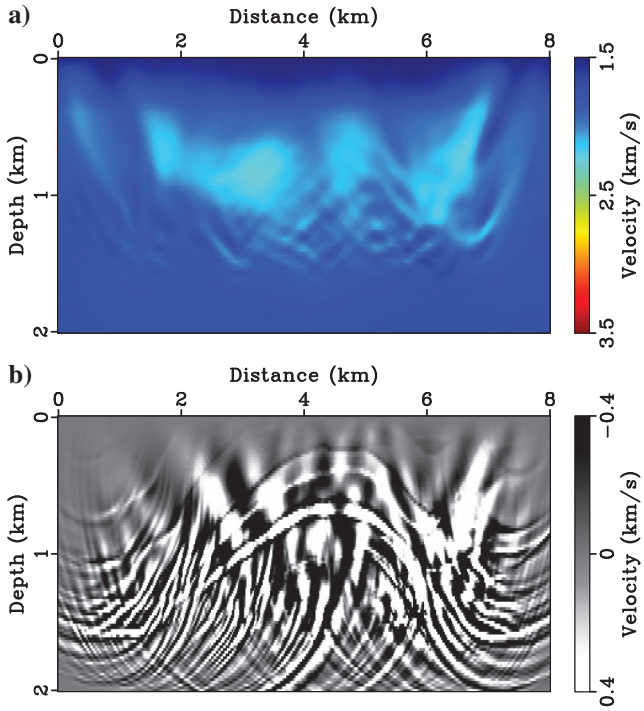

Figure 16. Results of the nonextended inversion, as described in the text, using the same algorithms and computational effort as the extended inversion: (a) inverted background velocity and (b) inverted reflectivity.

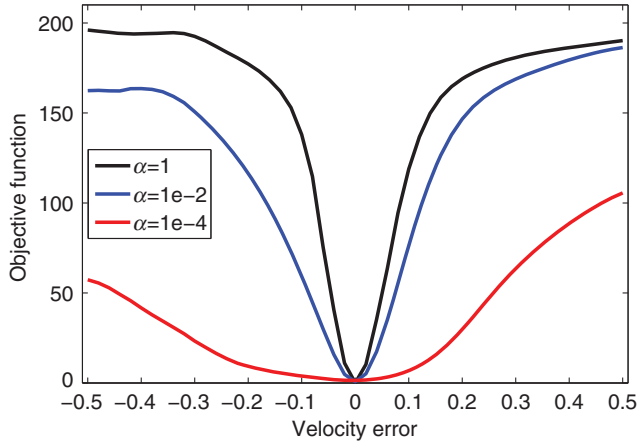

Figure 17. Scan test: objective function with different values of α. Velocity error varies from -50% to $+50\%$.

CONCLUSION

We have introduced a discrepancy-based method for control of a penalty parameter in regularized inverse problems, for which the regularization term vanishes at physically correct solutions. We applied this method to extended waveform inversion based on the Born approximation of constant density acoustic modeling. The discrepancy-based method systematically increases the penalty weight throughout the inversion process, driving the model toward physical consistency while maintaining the data fit within a specified range. Our example suggests that proper choice of the data-fit range allows our algorithm to converge to a kinematically accurate model yielding reasonable image fidelity. Model extension is essential to maintain good data fit and thereby avoid cycle skipping: The initial model in this example produces severely cycle-skipped data, so that an analogous inversion algorithm without extension stagnates at a grossly incorrect model estimate. Dynamic penalty parameter control based on discrepancy dramatically enhances the efficiency of extended inversion: Convergence is much faster than is the case with fixed penalty parameter.

ACKNOWLEDGMENTS

We are grateful to associate editor A. Baumstein, referees I. G. Roy and M. Maharramov, and an anonymous referee for their very helpful comments and suggestions. We would like to thank the sponsors of The Rice Inversion Project (TRIP) for their generous long-term support, which has made our research possible. We acknowledge the Texas Advanced Computing Center and the Rice University Research Computing Support Group for providing essential HPC resources.

REFERENCES

Ajo-Franklin, J. B., B. J. Minsley, and T. M. Daley, 2007, Applying compactness constraints to differential traveltime tomography: Geophysics, **72**, no. 4, R67–R75, doi: 10.1190/1.2742496.

Alford, R., R. Kelly, and D. M. Boore, 1974, Accuracy of the finite difference modelling of the acoustic wave equation: Geophysics, **39**, 834–842, doi: 10.1190/1.1440470.

Aminzadeh, F., J. Brac, and T. Kunz, 1997, 3-D salt and overthrust models: SEG.

Biondi, B., and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation

migration velocity analysis: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0275.1.
Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, no. 3, WA129–WA140, doi: 10.1190/geo2013-0340.1.
Biondi, B., and P. Sava, 2004, Wave-equation migration velocity analysis. Part I: Theory and Part II: Subsalt imaging examples: Geophysics, **52**, 593–623.
Engl, H., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.
Fu, L., and W. Symes, 2015, Reducing the cost of extended waveform inversion by multiscale adaptive methods: 85th Annual International Meeting, SEG, Expanded Abstracts, 1127–1131.
Fu, L., and W. Symes, 2016, Fast extended waveform inversion using Morozovs discrepancy principle: TRIP 2016 Annual Report, The Rice Inversion Project.
Fu, L., and W. Symes, 2017a, Accelerate subsurface-offset extended waveform inversion by adaptive methods: Presented at the CGS/SEG International Geophysical Conference, SEG.
Fu, L., and W. Symes, 2017b, An adaptive multiscale algorithm for efficient extended waveform inversion: Geophysics, **82**, no. 3, R183–R197, doi: 10.1190/geo2016-0426.1.
Golub, G., and V. Pereyra, 1973, The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate: SIAM Journal on Numerical Analysis, **10**, 413–432, doi: 10.1137/0710036.
Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: The variable projection method and its applications: Inverse Problems, **19**, R1–R26, doi: 10.1088/0266-5611/19/2/201.
Hansen, P., 1992, Analysis of discrete ill-posed problems by means of the l-curve: SIAM Review, **34**, 561–580, doi: 10.1137/1034115.
Hou, J., and W. Symes, 2016, Accelerating extended least-squares migration with weighted conjugate gradient iteration: Geophysics, **81**, no. 4, S165–S179, doi: 10.1190/geo2015-0499.1.
Huang, Y., and W. Symes, 2015, Born waveform inversion via variable projection and shot record model extension: 85th Annual International Meeting, SEG, Expanded Abstracts, 1326–1331.
Jackson, D. D., 1972, Interpretation of inaccurate, insufficient, and inconsistent data: Geophysical Journal of the Royal Astronomical Society, **28**, 97–109, doi: 10.1111/j.1365-246X.1972.tb06115.x.
Kelly, K. R., R. W. Ward, S. Treitel, and R. M. Alford, 1976, Synthetic seismograms: A finite-difference approach: Geophysics, **41**, 2–27, doi: 10.1190/1.1440605.
Kern, M., and W. Symes, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysical Prospecting, **99**, 565–614.
Lameloise, C.-A., H. Chauris, and M. Noble, 2015, Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis: Geophysical Prospecting, **63**, 391–404, doi: 10.1111/1365-2478.12195.
Liu, Y., W. Symes, and Z. Li, 2014, Inversion velocity analysis via differential semblance optimization: 76th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20141578.
Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81st Annual International Meeting, SEG, Expanded Abstracts, 2788–2792.
Morozov, V., 1984, Methods for solving incorrectly posed problems: Springer Verlag.
Nocedal, J., and S. Wright, 1999, Numerical optimization: Springer Verlag.
Plessix, R.-E., 2000, Automatic cross-well tomography: An application of the differential semblance optimization to two real examples: Geophysical Prospecting, **48**, 937–951, doi: 10.1046/j.1365-2478.2000.00218.x.
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, doi: 10.1111/j.1365-246X.2006.02978.x.
Plessix, R.-E., G. Baeten, J. W. de Maag, M. Klaassen, R. Zhang, and Z. Tao, 2010, Application of acoustic full waveform inversion to a low-frequency large-offset land data set: 81st Annual International Meeting, SEG, Expanded Abstracts, 930–934.
Roy, I., 2002, A robust descent type algorithm for geophysical inversion through adaptive regularization scheme: Applied Mathematical Modeling, **26**, 619–634, doi: 10.1016/S0307-904X(01)00072-5.
Roy, I., 2005, Iteratively adaptive regularization in inverse modeling with Bayesian outlook application on geophysical data: Inverse Problem in Science and Engineering, **13**, 655–670, doi: 10.1080/17415970500171076.
Sen, M. K., and I. G. Roy, 2003, Computation of differential seismograms and iteration adaptive regularization in prestack waveform inversion: Geophysics, **68**, 2026–2039, doi: 10.1190/1.1635056.
Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, no. 5, VE49–VE60, doi: 10.1190/1.2972021.
Shen, P., W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, SEG, Expanded Abstracts, 2135–2139.
Symes, W., 1986, Stability and instability results for inverse problems in several-dimensional wave propagation: Proceedings of the 7th International Conference on Computing Methods in Applied Science and Engineering.
Symes, W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790, doi: 10.1111/j.1365-2478.2008.00698.x.
Symes, W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663, doi: 10.1190/1.1443082.
Symes, W., and F. Santosa, 1988, Computation of the Newton Hessian for least-squares solution of inverse problems in reflection seismology: Inverse Problems, **4**, 211–233, doi: 10.1088/0266-5611/4/1/017.
Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.
van Leeuwen, T., and F. J. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667, doi: 10.1093/gji/ggt258.
Vigh, D., K. Jiao, W. Huang, N. Moldoveanu, and J. Kapoor, 2013, Long-offset-aided full-waveform inversion: 75th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.20130825.
Vigh, D., W. Starr, J. Kapoor, and H. Li, 2010, 3D full waveform inversion on a Gulf of Mexico WAZ data set: 80th Annual International Meeting, SEG, Expanded Abstracts, 957–961.
Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC127–WCC152, doi: 10.1190/1.3238367.
Wahba, G., 1977, Practical approximate solutions to linear operator equations when the data are noise: SIAM Journal on Numerical Analysis, **14**, 651–667, doi: 10.1137/0714044.
Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, SEG, Expanded Abstracts, 1089–1093.
Weibull, W., and B. Arntsen, 2013, Automatic velocity analysis with reverse time migration: Geophysics, **78**, no. 4, S179–S192, doi: 10.1190/geo2012-0064.1.

Downloaded 01/31/18 to 160.91.216.167. Redistribution subject to SEG license or copyright; see Terms of Use at http://library.seg.org/