An adaptive multiscale algorithm for efficient extended waveform inversion

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

ABSTRACT

Subsurface-offset extended full-waveform inversion (FWI) may converge to kinematically accurate velocity models without the low-frequency data accuracy required for standard data-domain FWI. However, this robust alternative approach to waveform inversion suffers from a very high computational cost resulting from its use of nonlocal wave physics: The computation of strain from stress involves an integral over the subsurface offset axis, which must be performed at every space-time grid point. We found that a combination of data-fit driven offset limits, grid coarsening, and low-pass data filtering can reduce the cost of extended inversion by one to two orders of magnitude.

INTRODUCTION

Full-waveform inversion (FWI), that is, model-driven least-squares data fitting, has shown a remarkable ability to identify subsurface structure with the maximum resolution attainable from seismic data (Vigh et al., 2010, 2013). However, the lack of data energy at low frequencies relative to other scales may cause iterative gradient-based algorithms to stagnate at uninformative model estimates (Gauthier et al., 1986; Plessix et al., 2010). At such estimates, small changes in the model fail to yield a substantially better data fit. Among the remedies suggested for this malady are various model extensions, which add parameters to the model to provide an avenue for improved data fit and suppress these additional parameters as the inversion progresses via a penalty term incorporated into the extended FWI objective function (Symes, 2008). One of these model extensions allows nonlocal stress-strain relations — in the acoustic case, by adding dependence on a fictitious subsurface-offset axis to the bulk modulus. It has been shown that this subsurface-offset extension makes the data fit attainable by local optimization, at least in some cases (Stolk et al., 2009; Symes, 2014). Several studies have suggested that the subsurface-offset extension may

be used as the basis for successful approaches to FWI, convergent over a much larger region of model space than is standard in least-squares FWI (Shen et al., 2003, 2005; Albertin et al., 2006; Shen and Symes, 2008; Symes, 2008, 2014; Fei and Williamson, 2010; Vyas and Tang, 2010; Biondi and Almomin, 2012; Shen, 2012; Shan and Wang, 2013; Weibull and Arntsen, 2013; Biondi and Almomin, 2014; Liu et al., 2014; ten Kroode, 2014; Fu and Symes, 2015; Lameloise et al., 2015). Much of this work uses linearization (the Born approximation) to simplify the formulation of extended inversion, as do we in the work reported below. Linearized extended waveform inversion is closely related to wave-equation-based migration velocity analysis (Symes, 2008).

A major drawback of the subsurface-offset extension is the computational burden of the nonlocal constitutive law: In terms of time-stepping algorithms, it calls for a full matrix multiplication over at least one spatial axis at every time step (Mulder, 2014). The purpose of this paper is to propose a straightforward strategy to reduce the cost of this class of algorithm, combining frequency continuation and grid coarsening, and reduction of the subsurface-offset axis with control of the penalty parameter in a variable projection formulation (Golub and Pereyra, 1973, 2003; van Leeuwen and Mulder, 2009; Rickett, 2012; Li et al., 2013). Our algorithm relies on two simple observations: (1) improved kinematic accuracy of data-fitting extended models results in improved focus, that is, it moves the extended model closer to a nonextended or physical model concentrated at zero-subsurface offset, and (2) low-pass filtered data have enough kinematic content to drive velocity improvement, so long as the filtered data span an octave or more. We show that the extent of the active interval on the subsurface-offset axis may be shortened, subject to a data-fit criterion because the inversion improves the kinematic accuracy of the model, in concert with refinement of the grid, in such a way that the number of offset grid points is nonincreasing. That is, the cost premium of subsurface-offset inversion iterates over ordinary FWI iterates remains constant or decreases, even as the resolution of the model increases. From another point of view, the adaptive concept introduced in this paper provides a way that naturally brings an extended model back to a physical model by progressively reducing the subsurface-offset

---

range throughout the inversion. In our 2D examples, the cost of this adaptive multiscale extended inversion is a few percent of the cost of the same sequence of iterations applied to a globally adequate fixed offset range and grid. The cost reduction in 3D would be even more dramatic.

This paper is organized as follows: We first explain the theory of subsurface-offset extended waveform inversion based on linearized acoustic modeling (Born or single-scattering approximation). We describe three adaptative parameter adjustments — of maximum subsurface offset, variable projection penalty parameter, and computational grid — and we show how to combine them to dramatically enhance the computational efficiency. We end with two 2D numerical examples, demonstrating that this adaptive multiscale modification can reduce the computational cost of subsurface offset extended waveform inversion by an order of magnitude or more.

THEORY

Although the algorithm to be explained here applies to many models and can be posed abstractly, we choose to explain it in the context of a particular model of wave propagation, linearized constant density acoustics, and inverse problems posed in terms of this model.

Acoustic Born modeling

An abstract setting for seismic waveform inversion problem consists of the model space M, which is a set of physical model of the subsurface structure; the data space D, which denotes a set of the seismic data; and the forward map F, which connects the two spaces M and D $(F:M \rightarrow D)$.

We base our study on linearized ("Born") 2D constant-density acoustics: M consists of pairs (v, r) of (background) velocity field $v(x, z)$ and reflectivity field $r(x, z)$. The reflectivity is the perturbation of squared velocity: $r = 2v\delta v$. Parameter D consists of primaries-only (single-scattering) seismic traces $d(x_r, x_s, t)$ for source positions $\{(x_s, z_s)\}$ and receiver positions $\{(x_r, z_r)\}$. Source and receiver depths are idealized as the same for all traces, so they are ignored in the notation for the data traces. The pressure field $p(x, z, t; x_s)$ solves the acoustic-wave equation:

$$\left(\frac{\partial^2}{\partial t^2} - v^2\nabla^2\right)p(x, z, t; x_s) = w(t)\delta(x - x_s, z - z_s),$$
$$p = 0, t \ll 0. \tag{1}$$

The right side is a simple source representation, an isotropic point radiator with time dependence (pulse) $w(t)$ located at $(x, z) = (x_s, z_s)$. The perturbational pressure field $\delta p(x, z, t; x_s)$ solves the linearized acoustic-wave equation:

$$\left(\frac{\partial^2}{\partial t^2} - v^2\nabla^2\right)\delta p(x, z, t; x_s) = r(x, z)\nabla^2 p(x, z, t; x_s),$$
$$\delta p = 0, t \ll 0, \tag{2}$$

where F is defined in terms of δp by

$$F[v]r(x, z) = \delta p(x_r, z_r, t; x_s), \tag{3}$$

where F produces predicted primary (single-scattering) data traces for the model (v, r). Note that we have used a notational convention suggesting that the action of F on r is linear, rather than writing $F[v, r]$: When v is fixed, the action of F on r (after discretization) could be represented by a matrix multiplication.

Extended acoustic Born modeling

Waveform inversion asks that the model be adjusted, so that the predicted data traces approximate observed data traces, in the mean square sense. As mentioned in the "Introduction" section, this is a very hard optimization problem. For the acoustic Born problem just described, a satisfactory solution is obtained by gradient-descent methods only if the velocity v predicts the times of significant arrivals to within a half-wavelength (Gauthier et al., 1986; Virieux and Operto, 2009). Extended modeling seeks to create an easier optimization problem by enlarging the model space. Extended waveform inversion involves the additional ingredients:

1) an extended model space $\bar{M}$
2) an extension operator $\chi: M \rightarrow \bar{M}$
3) an extended modeling operator $\bar{F}: \bar{M} \rightarrow D$
4) an annihilator operator $A: \bar{M} \rightarrow \bar{M}$.

The operator $\bar{F}$ is an extension of F in the sense that $\bar{F}[E[m]] = F[m]$. The annihilator identifies the "physical" space χM as its null space: $A\bar{m} = 0$ if and only if $\bar{m} = \chi m$ for a physical model $m \in M$.

The extended Born acoustic model used here introduces a horizontal subsurface-offset axis, denoted h, and allows the reflectivity to depend on it: $\bar{r}(x, z, h)$. Because r is (up to a scale factor) the perturbation in the compliance, one can think of the extended reflectivity as representing a nonlocal perturbation in the acoustic constitutive relation (we are indebted to S. Morton for this observation). The extended pressure perturbation $\delta\bar{p}(x, z, t; x_s)$ solves a modification of the linearized wave equation 2,

$$\left(\frac{\partial^2}{\partial t^2} - v^2\nabla^2\right)\delta\bar{p}(x, z, t; x_s)$$
$$= \int_{-H}^{H} dh\bar{r}(x - h, z, h)\nabla^2 p(x - 2h, z, t; x_s),$$
$$\delta p = 0, t \ll 0. \tag{4}$$

The extended model space for horizontal subsurface offset consists of pairs $\bar{M} = \{(v(x, z), \bar{r}(x, z, h))\}$. Note that only the reflectivity depends on the additional coordinate h — the velocity is non-extended or physical. The extension operator is $\chi r(x, z, h) = r(x, z)\delta(h)$. Note that if $\bar{r} = \chi r$, then equation 4 reduces to equation 2.

The extended Born forward modeling operator is defined by

$$\bar{F}[v]\bar{r}(x, z, h) = \delta\bar{p}(x_r, z_r, t; x_s). \tag{5}$$

Because of the previous remark, this operator has the extension property $\bar{F}[v]\chi r = F[v]r$. Note also that we have continued to use the convention that the predicted data are the value of a v-dependent linear operator acting on a reflectivity field.

Many choices of annihilator A have the required relation to the physical model subspace χM. Among the earliest suggested was

multiplication by h: $Ar(x, z, h) = hr(x, z, h)$ (Stolk and De Hoop, 2001; Shen et al., 2003), which we also use here.

The dynamics expressed in equation 4 are closely related to Claerbout's survey-sinking image construction (Claerbout, 1985): (x, z) are the coordinates of the sunken receiver, $(x - 2h, z)$ those of the sunken source (where the source wavefield p is evaluated), so the sunken midpoint is $(x - h, z)$ and the space shift h plays the role of half-offset, as one would expect.

Note that the integration over h on the right side of equation 4 translates into a full-matrix multiply in a finite-difference discretization, and must be performed at every time step. The cost of this integration can easily overwhelm cost of ordinary time-stepping (Mulder, 2014). In 2D, this additional integral in dimension of h increases the computational cost by a factor of N_h, number of grid points in h, $N_h = 2H/dh$, in which dh is the grid size in h. Note that in 3D, another space-shift dimension is needed, making the subsurface-offset extension even more expensive.

Extended waveform inversion

In the extended model, the data $d(x_r, x_s, t)$ and the model $(v(x, z), r(x, z, h))$ depend on the same number of parameters, so you might guess that there would be a 1-1 relation between the two, at least to some extent. Stolk and De Hoop (2001) and Stolk et al. (2009) offer theoretical verification for this guess. We will also show by numerical example that for any "reasonable" data d and velocity model v, there exists a reflectivity $\bar{r}$ for which $\bar{F}[v]\bar{r} \approx d$, provided that the subsurface-offset limit H is large enough. This observation is in sharp contrast to the case for nonextended modeling: As noted above, v must predict arrival times within a half-wavelength in order that there exist a nonextended reflectivity for which $F[v]r \approx d$.

Extended waveform inversion couples a measure of data misfit (usually mean square) to a measure of model nonphysicality, and it drives both measures toward zero. A simple objective function capturing this concept is

$$J[v, \bar{r}, \alpha] = \frac{1}{2}\|\bar{F}[v]\bar{r} - d\|^2 + \frac{\alpha}{2}\|A\bar{r}\|^2. \quad (6)$$

The penalty weight α controls the balance between penalties on the data misfit and the model extension: When $\alpha \to 0$, the model has little constraint on energy distribution in the extended dimension h, and it can always achieve a good data fit. As noted above; when $\alpha \to \infty$, $A\bar{r}$ must $\to 0$, so $\bar{r}$ must approximate a physical reflectivity. For model-consistent (noise-free) data, the global minimum is $J = 0$. Close to the global minimizer, so both summands are close to zero, data are fit, and $\bar{r}$ is approximately physical.

Variable projection method

The objective $J[v, \bar{r}, \alpha]$ is very ill conditioned; that is, its gradient tends to change very rapidly in response to small model changes, making minimization by the gradient-based method very difficult (Kern and Symes, 1994; Huang and Symes, 2015). However, the subproblem of estimating $\bar{r}$ by minimizing $J[v, \bar{r}, \alpha]$, given d and v, is quadratic and relatively well-conditioned. Define $\bar{r}[v, \alpha]$ to be an approximate minimizer of $J[v, \bar{r}, \alpha]$, that is, a solution of the normal equation:

$$(\bar{F}[v]^T\bar{F}[v] + \alpha A^T A)\bar{r} = \bar{F}[v]^T d. \quad (7)$$

The normal equation is equivalent to vanishing of the $\bar{r}$ gradient of $J[v, \bar{r}, \alpha]$. Because the J is positive semidefinite quadratic in $\bar{r}$; that is, it is sufficient to guarantee that $\bar{r}[v, \alpha]$ is an approximate minimizer.

Having chosen the optimal $\bar{r}[v, \alpha]$, substitute it into J to obtain an objective function in v alone:

$$\tilde{J}[v, \alpha] = J[v, \bar{r}[v, \alpha], \alpha]. \quad (8)$$

A minimizer of $\tilde{J}$ is also the v component of a minimizer of J. This is the variable projection principle introduced by Golub and Pereyra (1973, 2003). Moreover, $\tilde{J}$ has a reasonably well-conditioned Hessian and can be minimized effectively with Newton-related techniques, unlike J (Kern and Symes, 1994; Huang and Symes, 2015). Its gradient may be expressed as

$$\nabla_v\tilde{J}[v, \alpha] = S(D\bar{F}[v])^T(\bar{r}[v, \alpha], \bar{F}[v]\bar{r}[v, \alpha] - d), \quad (9)$$

where $(D\bar{F}[v])^T$ is the so-called tomographic or WEMVA operator. Because $\bar{F}[v]$ is actually the linearization, that is, the derivative of the basic acoustic modeling operator, $D\bar{F}[v]$ is its second derivative. The adjoint $D\bar{F}[v]^T$ is computable by a variant of the adjoint-state method (Gauthier et al., 1986; Plessix, 2006) used to compute $F[v]^T$ (Symes and Santosa, 1988; Kern and Symes, 1994). Parameter S is a smoothing operator, or low-pass filter, designed to keep the scales of v and $\bar{r}$ separated after updates using the Hessian. In our work, we used for S a negative power of the spatial Laplace operator.

Adaptive subsurface offset

As explained above, the additional dimension h adds a significant computational cost, so determining the maximum subsurface offset H becomes a crucial problem. Shen (2004) shows some exemplary calculations to identify the relevant ray fields with subsurface space shift, but that does not address the distribution of energy in the space-shift extended model. Mulder (2014) gives formulas to calculate the amplitude in the space-shift extended model for 2D and 3D by stationary phase approximation, provided that the model and migration velocities are constant. In that case, a good estimate for the maximum subsurface offset may be expressed in terms of the maximum surface offset L and the ratio $\rho = v/v_{\text{true}}$ of the migration velocity to the model velocity, as $H = L(1 - \rho^2)$. However, in general, there is no simple direct rule to calculate an appropriate value of H.

Nonetheless, the generic relation between the migration (or inversion) velocity accuracy and the maximum subsurface offset holds more generally, in a qualitative sense. When the migration velocity model is inaccurate, a long offset is needed for a good data fit; when the velocity model becomes closer to the correct one after updates, then the observed data can be predicted by an extended reflectivity model with a shorter offset. This relation will be illustrated below, and it follows for the example from the theoretical analysis by Stolk et al. (2009).

A simple method to estimate appropriate values of H takes advantage of this observation. To explain this method, note that the solution $\bar{r}[v, \alpha]$ of the normal equation 7 depends on H, through the right side of the extended linearized wave equation 4. To track this dependence, we add it to the notation and rename the solution $\bar{r}[v, \alpha, H]$. Similarly, we add H to the argument lists of J, $\tilde{J}$ and introduce the abbreviations e and p for the data misfit and penalty summands in the definition of $\tilde{J}$:

$$\tilde{J}[v, \alpha, H] = e[v, \alpha, H] + \alpha p[v, \alpha, H],$$

$$e[v, \alpha, H] = \frac{1}{2}\|\bar{F}[v]\bar{r}[v, \alpha, H] - d\|^2,$$

$$p[v, \alpha, H] = \frac{1}{2}\|A\bar{r}[v, \alpha, H]\|^2. \qquad (10)$$

Our approach uses the zero-weight residual $e(v, 0, H)$, and it estimates the corresponding residual for offset range $[-H/2, H/2]$ by computing

$$e_{H/2} = \frac{1}{2}\|\bar{F}[v]\Pi_{H,H/2}r[v, 0, H] - d\|^2, \qquad (11)$$

in which $\Pi_{H,H/2}\bar{r}$ denotes the restriction operator that sets $\bar{r} = 0$ for $|h| > H/2$.

Note that $e[v, 0, H]$ depends on $\bar{r}[v, 0, H]$, and therefore it requires the (approximate) solution of the normal equation 7 with zero penalty weight by iteration. However, once $e[v, 0, H]$ is computed, $e_{H/2}$ simply reuses $\bar{r}[v, 0, H]$, so it requires just a forward-modeling step. Therefore, $e_{H/2}$ represents minimal added expense over $e[v, 0, H]$.

Algorithm 1 adapts H to keep the data-fit error e, with zero penalty below a prescribed bound. According to the data-fit property of extended modeling, mentioned above, for any (reasonable) v and tolerance E, a corresponding $\bar{r}$ exists for which $e < E$, provided that H is large enough. At the beginning of the v update cycle, nothing is known about the proper value of H, so increasing it may be necessary. However, as the velocity improves, H should decrease monotonically; an increase of H is necessary only at the beginning of the velocity update cycle. The data fit must be calibrated: We choose an arbitrary relative value X, intended to be eventually a nominal upper bound on actual data noise. We also choose a fudge factor $\mu > 1$, the detection level for reducing H because $e_{H/2}$ is not precisely the same as $e[v, \alpha, H/2]$. In the following, we will use $E = (1/2)X^2\|d\|^2$ as the relative error measure appropriate for comparison with $e(\dots)$.

Adaptive penalty weight

The penalty parameter α is essential in driving the extended waveform inversion toward a solution of the (nonextended) waveform inversion problem. Because p should vanish at a physical solution, one would expect that α should increase as the inversion proceeds. Fu and Symes (2016) show how to systematically increase α to accelerate the convergence of iterative solution for problems with the properties of extended Born waveform inversion. For completeness, we repeat the algorithm of Fu and Symes (2016) as Algorithm 2.

Typical values for the parameters appearing in Algorithm 2 might be $\gamma_- = (0.7)^2, \gamma_+ = (1.2)^2, \beta_- = 0.667, \beta_+ = 2.0$ (we use these values in the experiments reported below).

Note that the update formula used above, and the corrections that follow are guaranteed to yield α, for which $\gamma_-E \leq e[v, \alpha, H] \leq \gamma_+E$. This range condition is a version of the discrepancy principle as explained by Fu and Symes (2016). Unlike normal applications of this principle, Algorithm 2 is designed to increase α systematically.

Adaptive variable projection algorithm

Optimization of $\tilde{J}[v, \alpha, H]$ may be accomplished by any of the commonly used continuous optimization methods, described, for instance, by Nocedal and Wright (1999). However, updating α and/or H actually changes the objective function, so an algorithm that includes such updates is outside the scope of standard optimization theory.

We do not attempt to create a conceptual framework for such "multifunction optimization" here. We merely note that many algorithms include provision for a so-called "warm start"; that is, various parameter values, including solution estimates, may be carried over from one iteration to the next, and a change of objective function may be simply ignored in some cases. For example, a descent algorithm globalized by line search may use the final step length from the previous iteration as the initial step length for the current one. This carryover can be retained even if the objective function changes between iterates. Other auxiliary information may be better discarded; for example, the low-rank inverse Hessian approximation built up in some quasi-Newton methods may lose so much

---

Algorithm 1. The H update.

Require: $X > 0, \mu > 1, E = \frac{1}{2}X^2\|d\|^2$
  $H_+ = H$
  if initialize **then**
    while $e[v, 0, H_+] > E$ **do**
      $H_+ \leftarrow 2H_+$
    end while
  else
    while $e_{H_+/2} < \mu E$ **do**
      $H_+ \leftarrow H_+/2$
    end while
  end if return H_+

Algorithm 2. The α update.

Require:
$X > 0, E = \frac{1}{2}X^2\|d\|^2, 0 < \gamma_- < 1 < \gamma_+, 0 < \beta_- < 1 < 1/\beta_- < \beta_+$
  if initialize **then**
    $\alpha = 0$
  end if
  $\alpha_+ = \alpha$
  if $e[v, \alpha_+, H] < \gamma_-E$ **then**
    $\alpha_+ \leftarrow \alpha_+ + \frac{\gamma_+E - e[v, \alpha_+, H]}{2p[v, \alpha_+, H]}$
    while $e[v, \alpha_+, H] \notin [\gamma_-E, \gamma_+E]$ **do**
      if $e[v, \alpha_+, H] < \gamma_-E$ **then**
        $\alpha_+ \leftarrow \beta_+\alpha_+$
      end if
      if $e[v, \alpha_+, H] > \gamma_+E$ **then**
        $\alpha_+ \leftarrow \beta_-\alpha_+$
      end if
    end while
  end if return α_+

accuracy in a change of objectives that it should be recomputed ab initio. In view of such possibilities, α and H updates must be visible to the optimization algorithm.

A general approach to this issue remains to be worked out. In our experiments, we have used a very simple continuous optimization algorithm, steepest descent with line-search globalization, and we defined the term warm start in the obvious way. Assuming that the notion of warm start is properly chosen, a suitable algorithm structure incorporating the α and H adaptations is given in Algorithm 3.

Adaptive grid

The previous two subsections explained adaptations that pertain, in principle, to the continuum extended waveform inversion problem. This subsection addresses the discrete level, after gridding. The idea is simple and familiar from the FWI literature: Use lower frequency (less resolved) information when the model is further from kinematic data fidelity, and increase the frequency content as the inversion progresses. In the context of extended waveform inversion, the reason is not to enhance the tendency to converge — use of the model extension and appropriate penalty largely decouples the frequency content from convergence (Symes, 2008, 2014). Instead, the point is purely to reduce the computational cost of a majority of the iterations.

As long as Algorithm 3 improves kinematic fidelity enough that the initial H is at least halved by each execution, the number of h grid points N_h will never exceed its initial value. Under this condition, which we have observed in examples, the cost ratio of extended to nonextended inversion steps stays constant or decreases as the iterations of Algorithm 4 proceed.

Cost

It is straightforward to analyze the cost of Algorithm 4 when the numbers of iterations involved in Algorithm 3 are the same for all refinement steps. The cost of each iteration is approximately proportional to number of grid points in space, subsurface offset, and time. The cost of carrying out n steps of Algorithm 4 is

$$a = \sum_{k=0}^{n} s_k * C * N_{h,k} * (N_{t,k})^3, \qquad (12)$$

in which s_k is the number of velocity updates at grid level k; $N_{t,k}$ is the number of time steps for modeling at grid level k; $N_{h,k}$ is the

number of active grid points on the offset axis; and C is a garbage collection factor that contains the ratio of time to spatial grid sizes, the number of floating point operations per velocity update, and the average cost (in time or cycles) per floating point operation. Similarly, the cost of a nonadaptive algorithm, using the finest grid and largest offset axis with $N_h = 2^n N_{h,n}$ grid points, is

$$c = \left(\sum_{k=0}^{n} s_k\right) * C * N_h * N_{t,0}^3, \qquad (13)$$

with the same fudge factor C.

Assume that at each grid-refinement step of the loop in Algorithm 4, the offset range has been halved at least once, then as noted above $N_{h,k} \leq N_{h,0}$ for $k = 0, \ldots, n$. Because $N_{t,k} = 2^{k-n} N_{t,0}$, the cost ratio is

$$r = a/c = 2^{-n} \left(\sum_{k=0}^{n} s_k * 8^{k-n}\right) / \left(\sum_{k=0}^{n} s_k\right). \qquad (14)$$

This ratio is easy to evaluate under the assumption that the same number of velocity updates occurs at each scale. Then, the ratio becomes

$$= 2^{-n}/n(1 - 8^{-n})/(1 - 8^{-1}) \approx 2^{-n}/n. \qquad (15)$$

For $n = 1$ (two data octaves), $r \approx 1/2$, for $n = 2$ (three octaves), $r \approx 1/8$, and for $n = 3$ (four octaves), $r \approx 1/24$.

It should be noted however that, in our experience, the assumption of uniform iterations over grid refinements is unrealistic: In both examples to be presented in the next section, the bulk of the iterations occurred at coarser grid scales, which substantially decreases the cost of the adaptive algorithm, as will be noted below.

EXAMPLES

In this section, we illustrate the performance of the proposed adaptive multiscale approaches by solving a velocity estimation problem modeled on reflection seismology. The simulations are performed

---

Algorithm 3. Adaptive variable projection method.

Require: choose initial velocity v, α, H, parameters for Algorithms 1 and 2, continuous optimization step OPTSTEP, improvement tolerance ϵ.

  repeat

    execute Algorithm 1 to compute H_+

    execute Algorithm 2 to compute α_+

    execute OPTSTEP with warm start to compute v_+; compute relative improvement

    $\delta = (\tilde{J}[v, \alpha, H] - \tilde{J}[v_+, \alpha_+, H_+])/\tilde{J}[v, \alpha, H]$

    $v \leftarrow v_+, \alpha \leftarrow \alpha_+, H \leftarrow H_+$

  until $\delta < \epsilon$

Algorithm 4. Adaptive multiscale variable projection method.

Require: Determine data passband $[f_{\min}, f_{\max}]$, initial discretization parameters dt_0, dx_0, dz_0, dh_0 suitable for accurate simulation at $f_{\max}$

  $n = \text{floor} \log_2(f_{\max}/f_{\min})$

  for $k = 0, \ldots, n$ **do**

    $dt \leftarrow 2^{n-k} dt_0$

    $dx \leftarrow 2^{n-k} dx_0$

    $dz \leftarrow 2^{n-k} dz_0$

    $dh \leftarrow 2^{n-k} dh_0$

    apply band-pass filter $[f_{\min}, f_{\min} + 2^{-n+k}(f_{\max} - f_{\min})]$ to data, source pulse, resample

    execute Algorithm 3 with discretization parameters dt, dx, dz, dh

  end for

---

using a 2D constant-density acoustics, time domain, and finite-difference method (second order in time and eighth order in space).

We use the method of steepest descent with a quadratic backtrack line search to search for the minimum of the objective function. The gradient of the objective function is computed using equation 9. We perform a line-search method to determine the optimal step length. The line search evaluated the objective function for different background velocity models, which were generated by adding multiplication of different step length and search direction to the current model. The optimal step length is estimated by assuming that the objective function is quadratic.

Throughout this section, we refer to the solution of the normal equation 7 with $\alpha = 0$ as extended least-squares reverse time migration (ELSRTM). This calculation is carried out at every velocity update, as part of the H update substep (Algorithm 1). We will display the reflectivity (image) volumes that result from ELSRTM because they clearly indicated the degree of focus toward $h = 0$, that is, the kinematic correctness of the velocity, attained by our algorithm.

Increased computational cost

To illustrate the relation between computational cost and subsurface-offset extension, we perform a simple numerical experiment. The grid dimension for velocity is 1000×1000, and the extended reflectivity grid is $1000 \times 1000 \times N_h$. Note that the computing time is normalized by the time of the nonextended case. The relative computing time is a linear function of the number of grid points N_h in the subsurface-offset axis (see Figure 1). Because the solution of the normal equation consists of iteration over extended Born modeling and RTM, longer subsurface offset directly increases the cost of each iteration.

Single reflector model

The purpose of this experiment is to investigate the behavior of the extended reflectivity model in the subsurface-offset axis with different levels of velocity errors. This experiment illustrates one of the main observations underlying our adaptive algorithm (and extended waveform inversion): The extent of the subset offset axis necessary for accurate data fit increases with the velocity error.

The background velocity model measures 3.0×2.4 km with the 20 m cell size uniformly distributed in each dimension. The true background velocity v is constant ($v = 3.0$ km/s). Shown in Figure 2a, in the extended reflectivity model, there is a horizontal velocity perturbation at the depth of 1.6 km. The 61 sources (0.3–2.7 km) and 151 receivers (0–3.0 km) are placed on the surface. Note that the background velocity model $v(x, z)$ is nonextended, whereas the extended perturbation model $\bar{r}(x, z, h)$ has a nonzero value only at $h = 0$ m. The observed data of shot 31 are depicted in Figure 2b.

The ELSRTM image with correct velocity after 20 CG iterations is shown in Figure 3a. The energy is focused at a zero-subsurface offset. The image with the wrong velocity is depicted in Figure 3b. In the vertical slices at $x = 1500$ m, the reflector is imaged as an upward curve, symmetrical in the h direction. Compared with the correct velocity case, the energy is scattered along h-axis. As the

Figure 1. The relative computing time of extended modeling and RTM as a function of the number of grid points N_h in the subsurface-offset axis. The reflectivity grid size is $1000 \times 1000 \times N_h$.
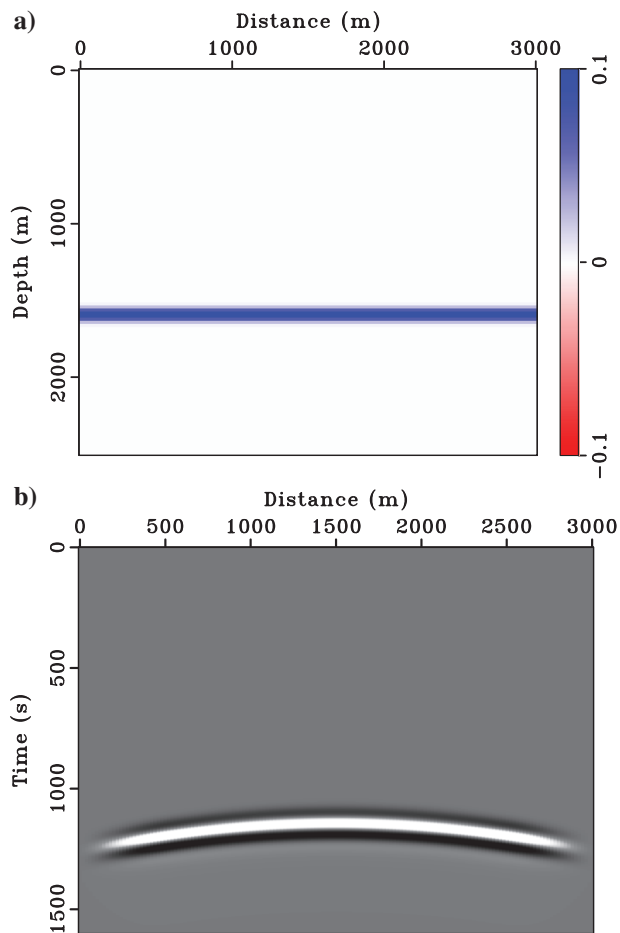

Figure 2. (a) Extended reflectivity $\bar{r}$ at $h = 0$ m and (b) the data of shot 31 at the center.

velocity error decreases, the energy tends to focus toward the center $h = 0$ (shown in Figure 4a–4f).

To fit the observed data well, the extended reflectivity model must have adequate subsurface extension (see Figure 5a and 5b). On the other hand, the required amount of the subsurface extension decreases with the error of the velocity model, which suggests that as the velocity model updates toward the correct one in the inversion, a shorter subsurface offset is needed (see Figure 6).

Lens model

In this example, the reflectivity model contains numbers of horizontal layers with various thickness (shown in Figure 7a). The background velocity model contains a Gaussian low-velocity anomaly sitting on constant velocity ($v = 3.0$ km/s). Acquisition and model geometry parameters are listed in Table 1.

The data are generated with a 3–30 Hz band-pass filter (see Figure 8a), or approximately three octaves. Accordingly, $n = 2$ and the observed data and the source function are filtered by three band-pass filters (3–7.5, 3–15, and 3–30 Hz). Correspondingly, the spatial decomposition is implemented in three steps with grid size 50, 25, and 12.5 m. The grid size in each stage is determined to fulfill the rule $(v_{min}/f_{max}) > 5dx$ to avoid numerical dispersion. The spatial grid size is the same in all dimensions ($dx = dz = dh$). The time-step intervals ($dt = 8, 4, 2$ ms) in each stage are chosen to fulfill the Courant-Friedrichs-Lewy condition ($(vdt/\sqrt{dx^2 + dz^2})$ < 1) in 2D.

In this example, we executed the α update step in Algorithm 3 only once, for the first velocity update at the coarsest grid level, and we left α constant thereafter.

For this example, we choose a target relative misfit level of $X = 0.1$.

Stage 1

We start with a grid size of 50 m, frequency band of 3–7.5 Hz (see Figure 8b), and the time-step interval of 8 ms. The initial background velocity is constant (3.0 km/s). Assuming that the relative velocity error is 10%, then $|1 - \rho^2| \approx 0.2$. Based on equation $H = L(1 - \rho^2)$, the homogeneous medium estimates mentioned in the last section, the initial offset range is estimated as ($H \approx 6 \times 0.2 = 1.2$ km). On the other hand, $dh = 50$ m, according to the relation $H = 2^i * dh$, we choose $i = 4$, so we start with subsurface-offset range $H = 800$ m. The data misfit satisfies the tolerance level $X = 0.1$ (10%) of the original data after 10 iterations of CG. Meanwhile, the reflectivity model with only a half-extension $H = 400$ m fails to fit the data to its satisfactory level $\mu X = 12\%$ even after 20 iterations (whereas with the full offset range, the iteration achieves a 5% misfit in 20 iterations), so the offset range of $H = 800$ m is an optimal choice for the initial velocity model. With low-frequency data, only several thick layers are visible in the inverted image (see Figure 9a). Due to the existence of the low-velocity anomaly, the reflectors beneath it have been imaged to deeper positions. The gather is barely focused in the h-direction.

Here, we have run 20 iterations of CG, far more iterations than are necessary, but only because we would like to emphasize that the data misfit would not decrease further to the tolerance level even with more iterations, when the subsurface offset is inadequate. We will stop iterating when the data misfit satisfies the tolerance level in the following test.

With the inverted $\bar{r}$ shown in Figure 9a and the corresponding data residual $\bar{F}[v]\bar{r} - d$, we update the background velocity model by computing the gradient according to equation 9.

In Figure 10a, the velocity model after first update already reveals the correct location of the top of the anomaly. With the first updated velocity, the linearized inverse problem (equation 7) is solved again. In Figure 10b, the reflector is shifted upward to a shallower position and is more focused in the h-direction. As the velocity model becomes closer to the correct one, the energy is more focused toward $h = 0$ m, which suggests that a shorter subsurface-offset range is needed. As a result, even half of the subsurface-offset range (400 m) is sufficient to predict the observed data well with a data misfit of less than 12%, shown in Figure 10c. After the second velocity update because the half-offset range $H = 200$ m is not able to provide a good data fit, $H = 400$ m is still the optimal choice (see Figure 11c). In the next three velocity updates (iterations 3–5), we are able to reduce the extended offset range by half (see Figures 12, 13, and 14). After the sixth velocity update, the reflectivity model even without extension is able to predict the observed data within the tolerance level (see Figure 15). And, the velocity update satisfies the convergence condition as the ninth step shown in the previous algorithm
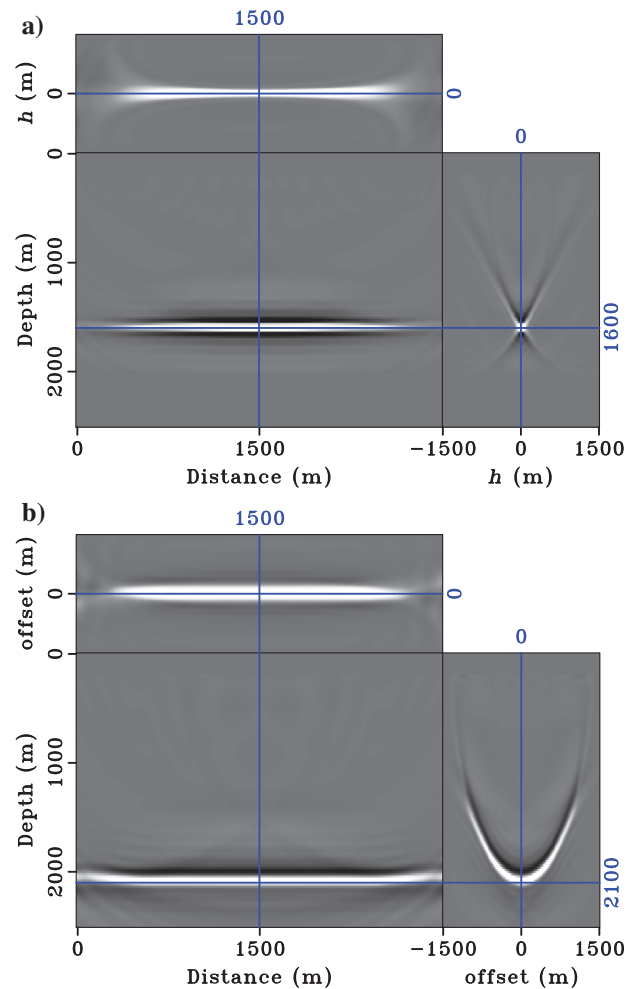

Figure 3. Inverted extended reflectivity $\bar{r}$ after 20 iterations of CG: (a) the correct background velocity ($v = v_{true}$) and (b) the wrong background velocity ($v = 1.3v_{true}$).

(see Figure 16), so a higher frequency and a finer grid can be used, leading the algorithm to the second stage.

Stage 2

The source wavelet and observed data are filtered by a low-pass filter (3–15 Hz). At the same time, the spatial grid size and the time-step interval are reduced by a factor of two. With the help of the updated velocity model (see Figure 17a) after two iterations, a higher frequency content and a finer spatial grid and time step, the ELSRTM image shows more structural details after seven CG iterations (shown in Figure 17b). The data residual is shown in Figure 17c.

Stage 3

The original source wavelet and observed data are used. At the same time, the spatial grid size 12.5 m and the time-step interval 2 ms is used. After one velocity update (Figure 18a), the reflectivity is inverted and illustrated in Figure 18b. With the help of the updated velocity model, the higher frequency content, and the finer spatial grid and time step, the background velocity and layered structure are successfully recovered.

For comparison, the inversion results without a model extension are shown in Figure 19a and

19b without the multiscale approach and Figure 20a and 20b with the multiscale approach. Here, without the extended dimension h, the objective function is defined with a different model space (M, not $\bar{M}$) and it only contains the data misfit term $1/2\|\bar{F}[v]r - d\|^2$. For both cases, we use the same optimization algorithm — the steepest descent with a quadratic backtrack line search. Both experiments are performed by using the same number of velocity iterations at each refinement stage as the extended case. And the reflectivity model is estimated by 20 iterations of CG, which equals
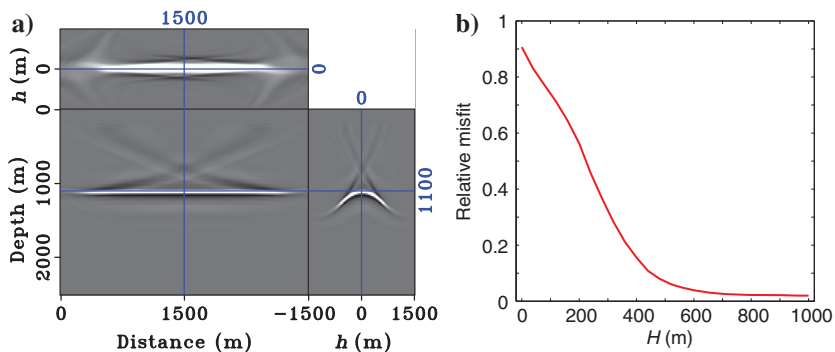
Figure 5. Relationship between the maximum subsurface offset H and the data residual Δd_H. The background velocity is 30% slower than the true velocity ($v = 0.7v_{\text{true}}$). (a) The ELSRTM image $\bar{r}$ after 20 iterations of CG and (b) H versus relative Δd_H.
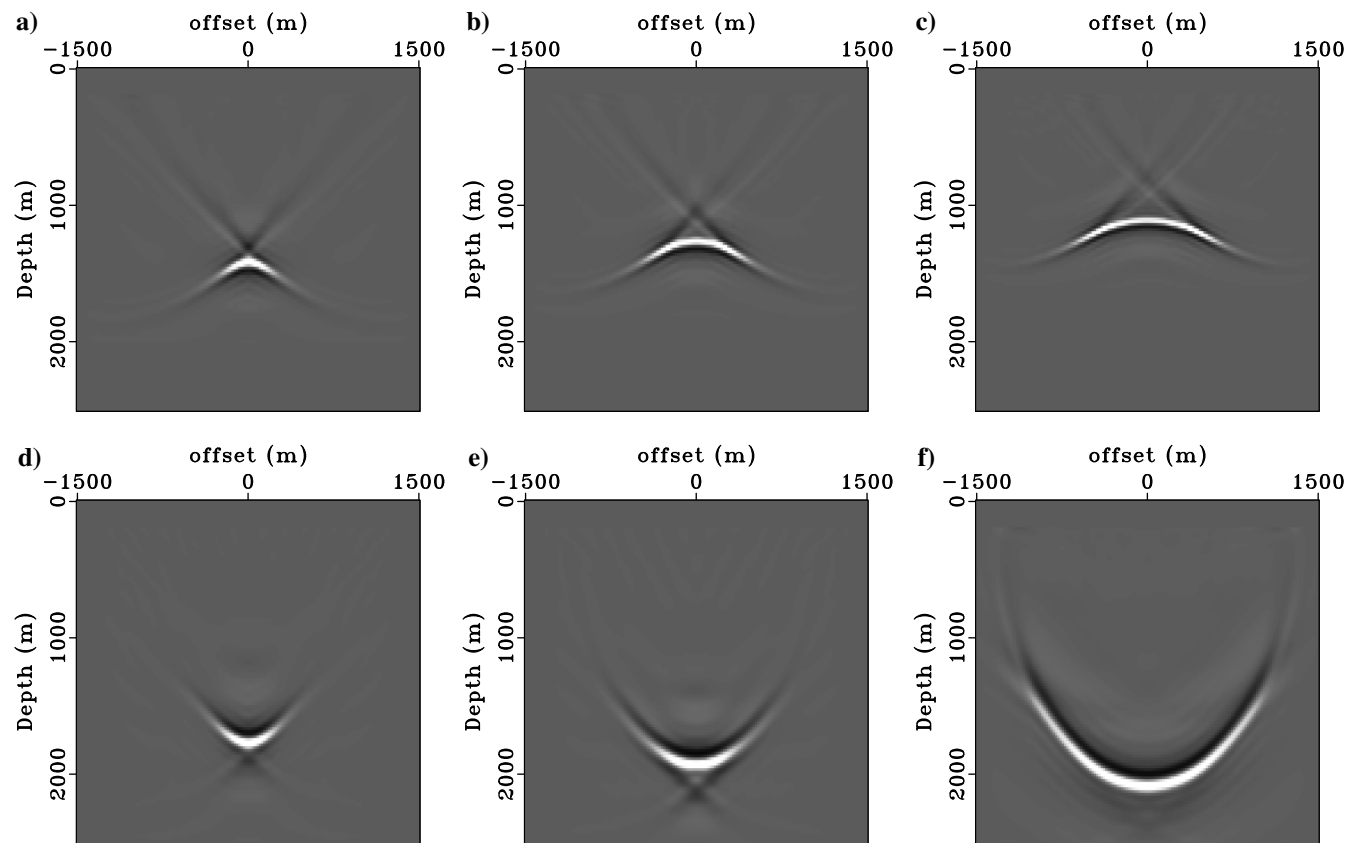
Figure 4. Inverted extended reflectivity $\bar{r}$ in z-h plane after 20 iterations of CG: (a) $v = 0.9v_{\text{true}}$, (b) $0.8v_{\text{true}}$, (c) $0.7v_{\text{true}}$, (d) $1.1v_{\text{true}}$, (e) $1.2v_{\text{true}}$, and (f) $1.3v_{\text{true}}$.
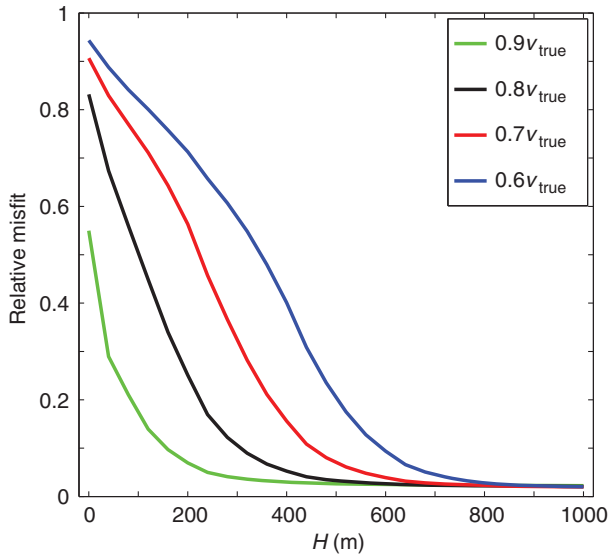
Figure 6. After 20 iterations of CG, the relative data misfit is plotted as a function of maximum subsurface offset H. Different colors represent the different levels of velocity error.

Table 1. Parameters for the lens model example.

Parameters	Measurements
Source wavelet	Band-pass 3–30 Hz
Source position x_s	x: 0–6 km every 50 m, $z = 0$ m
Receiver position x_r	x: 0–6 km every 50 m, $z = 0$ m
Space and time	$x = 6$ km, $z = 2.2$ km, $t = 2.4$ s
Grid size	$dx = dh = dz = 12.5$ m, $dt = 2$ ms
Initial velocity	$v = 3.0$ km/s
Maximum iterations of the inner loop	20
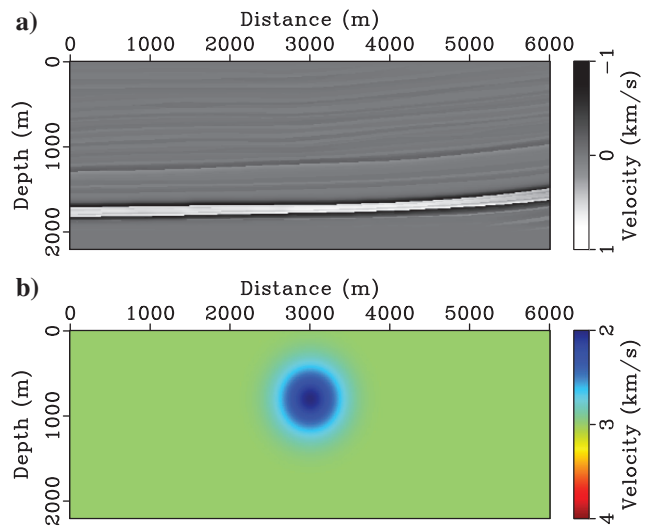
Figure 7. (a) The extended velocity perturbation $\delta\bar{v}$ at $h = 0$ m and (b) the background velocity.

20 modeling/migration pairs. As a result, the total numbers of modeling/migration pairs in these two nonextended experiments are the same as in the previous extended example. In both cases, the inversion fails to correctly recover the accurate background velocity. The layers beneath the low-velocity anomaly are imaged at the wrong depth.
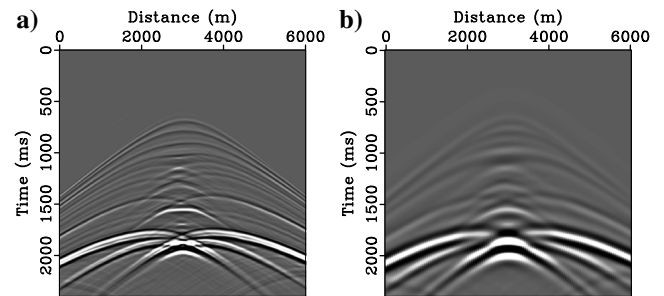
Figure 8. (a) Data of shot 61 at the center (3–30 Hz) and (b) band-pass low-frequency data (3–7.5 Hz).
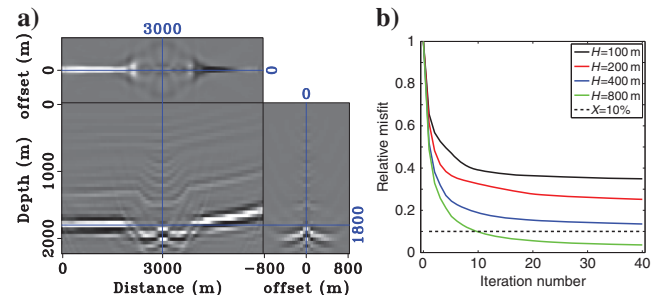
Figure 9. Step 1: Use the initial velocity, spatial grid size $dh = dx = dz = 50$ m, band-pass filter f: 3–7.5 Hz, time-step interval $dt = 8$ ms. (a) The ELSRTM result of 20 CG iterations and (b) the relative data residual with different offset ranges.
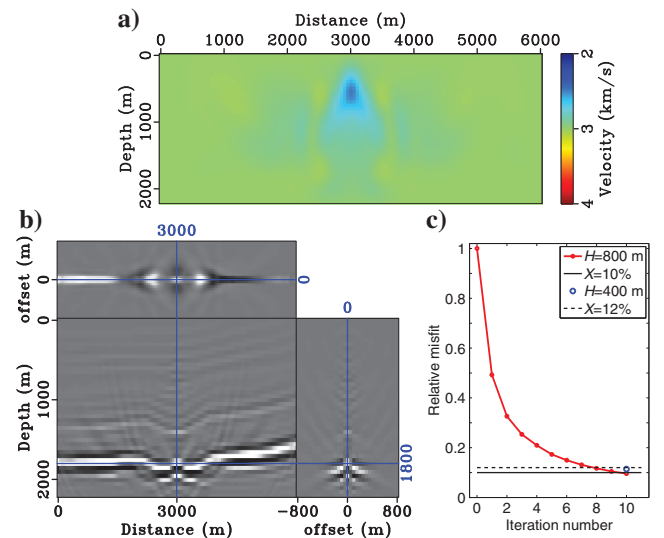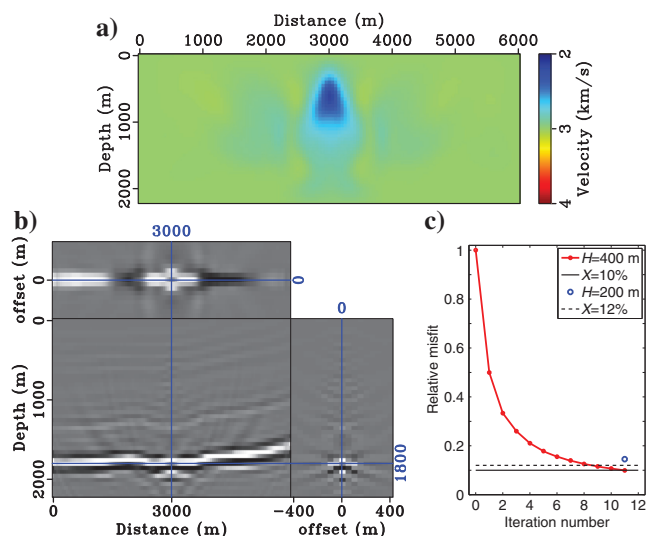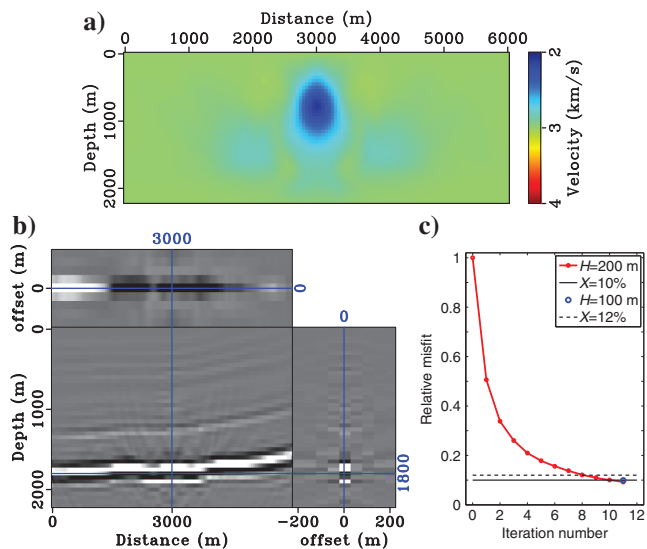

Figure 10. (a) The first velocity update, (b) the ELSRTM image by using the first updated velocity, and (c) the relative data residual with different offset ranges.

Overthrust model

This example is modified from the SEG/EAGE 3D overthrust model (Aminzadeh et al., 1997). In the reflectivity model, horizontal layers are distorted by several thrust (reverse) faults (see Figure 21). The background velocity increases with the depth. The velocity is higher in the center, where the anticline structure sits. The basic information is listed in Table 2.

In this example, a band-pass source wavelet (5–20 Hz) is used, thus $n = 2$ (octaves) of data are available. The inversion is divided into two refinement stages. The observed data and the source function are filtered by two band-pass filters (5–10 and 5–20 Hz). Correspondingly, the space decomposition is implemented with a grid size of 40 and 20 m, whereas the time-step intervals are 4 and 2 ms.

In this example, we used the full Algorithm 4, including the update for α, which resulted in increases of α at four steps ($\alpha = 1.1 \times 10^{-6}$, 3.2×10^{-5}, 1.1×10^{-4}, 2.4×10^{-4} at iterations 1, 2, 4, and 13). More details can be found in Fu and Symes (2016). The target relative misfit level for this experiment is once again 10%.

Stage 1

We use grid size 40 m, frequency band 5–10 Hz, and the time-step interval 4 ms. The initial background velocity is constant (1.5 km/s) (see Figure 22a). The optimal subsurface-offset range $H = 1.6$ km is

Figure 11. (a) The second velocity update, (b) the ELSRTM image by using the second updated velocity, and (c) the relative data residual with different offset ranges.

Figure 13. (a) The fourth velocity update, (b) the ELSRTM image by using the fourth updated velocity, and (c) the relative data residual with different offset ranges.
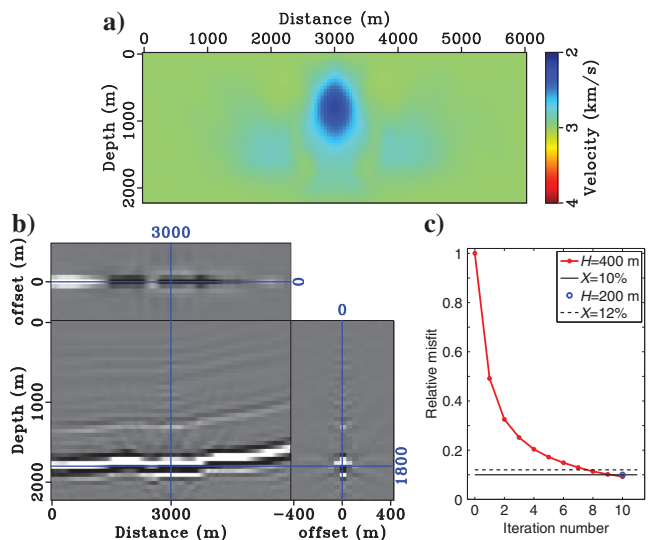
Figure 12. (a) The third velocity update, (b) the ELSRTM image by using the third updated velocity, and (c) the relative data residual with different offset ranges.
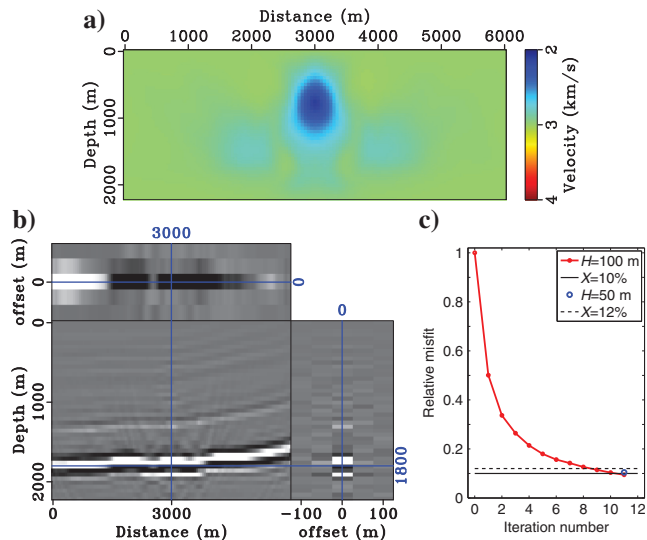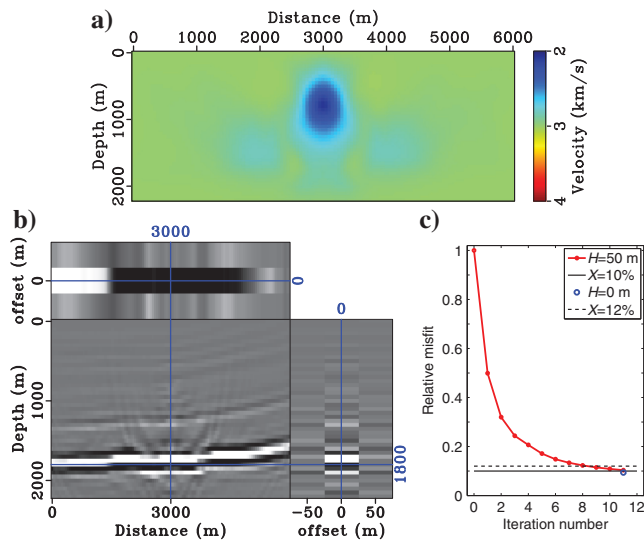

Figure 14. (a) The fifth velocity update, (b) the ELSRTM image by using the fifth updated velocity, and (c) the relative data residual with different offset ranges.

determined by measuring the data misfit during the initialization step in our proposed algorithm. Because the initial velocity is far away from the correct one, the structures can barely be observed from the ELSRTM image (see Figure 22b). In the subsurface-offset gather, the downward curves indicate slow velocity. After the first velocity update (see Figure 23), the energy becomes more focused toward $h = 0$. Actually, only half of the extension (0.8 km) would have provided a good data fit. With the second velocity update (see Figure 24), the layer structures are imaged closer to their correct positions. Although the energy is more focused in the gather, the half-subsurface offset is not adequate to provide a good data fit. However, as the background velocity continues to be updated, after the eighth iteration, offset $H = 0.4$ km is long enough (see Figure 25). At iterations

15 and 23, we are able to reduce the offset range by half. Because the convergence condition is satisfied at iteration 26, the inversion proceeds to the second stage.

Stage 2

The source wavelet and the observed data are filtered by a low-pass filter (5–20 Hz). At the same time, the spatial grid size and the time-step interval are decreased by a factor of two. With the inverted background velocity model from the 28th iteration, the extended reflectivity is shown in Figure 26a. 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,

Figure 15. (a) The sixth velocity update, (b) the ELSRTM image by using the sixth updated velocity, and (c) the relative data residual with different offset ranges.
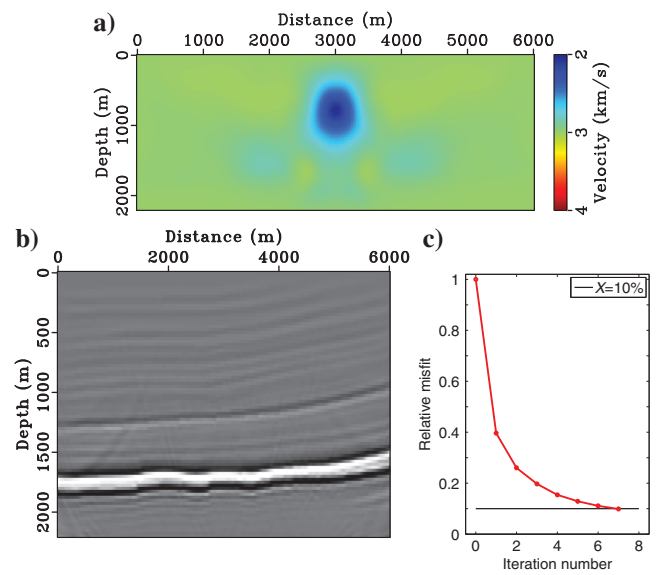

Figure 17. Stage 2: (a) The second velocity update, (b) the ELSRTM image, and (c) the relative data residual.
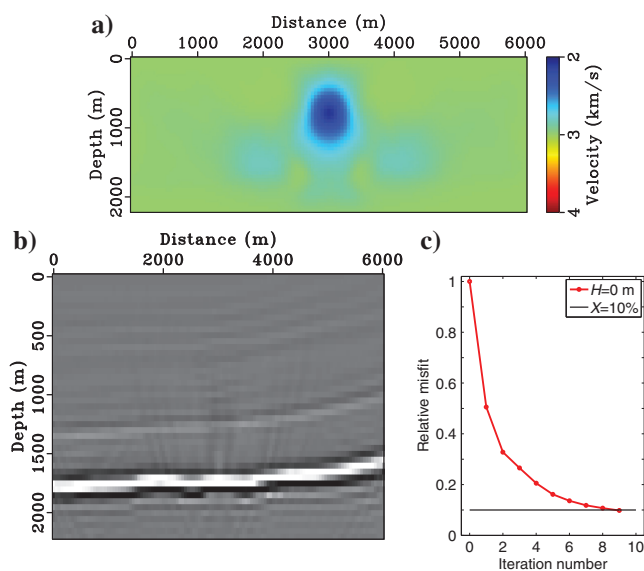

Figure 16. (a) The ninth velocity update, (b) the ELSRTM image by using the ninth updated velocity, and (c) the relative data residual.
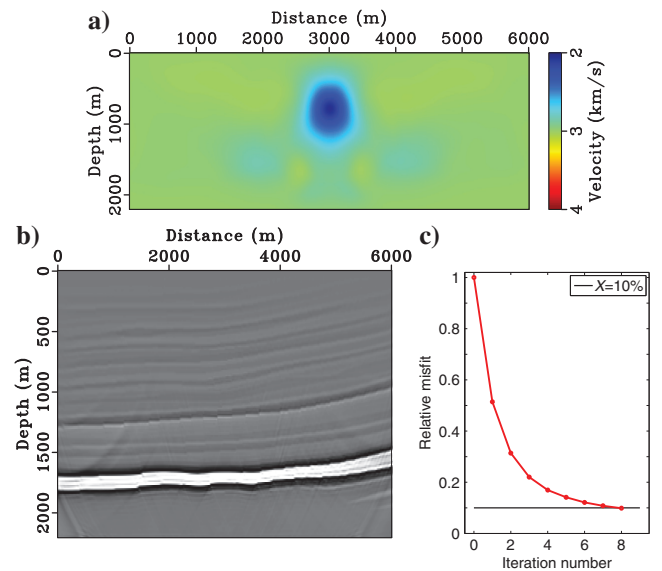

Figure 18. Stage 3: (a) The inverted background velocity model, (b) the ELSRTM image, and (c) the relative data residual.

Fu and Symes

$z = 2$ km). The true velocity model is mostly recovered. The velocity error is mostly at the edges, which is a result of illumination. The energy is well-focused in the h-direction.

DISCUSSION

A useful point of view on the range of subsurface offset H, kindly suggested by reviewer M. Vyas, relates the subsurface offset to the angle domain (Sava and Fomel, 2003): Angle gathers are (essentially) Radon tranforms of subsurface-offset gathers. It follows that the maximum offset H determines (nonaliased) the angle sampling,

Table 2. Parameters for the overthrust model example.

Parameters	Measurements
Source wavelet	Band-pass 5–20 Hz
Source position x_s	x: 1–7 km every 40 m, $z = 0$ m
Receiver position 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, $t = 2$ ms
Initial velocity	$v = 1.5$ km/s
Maximum iterations of the inner loop	20
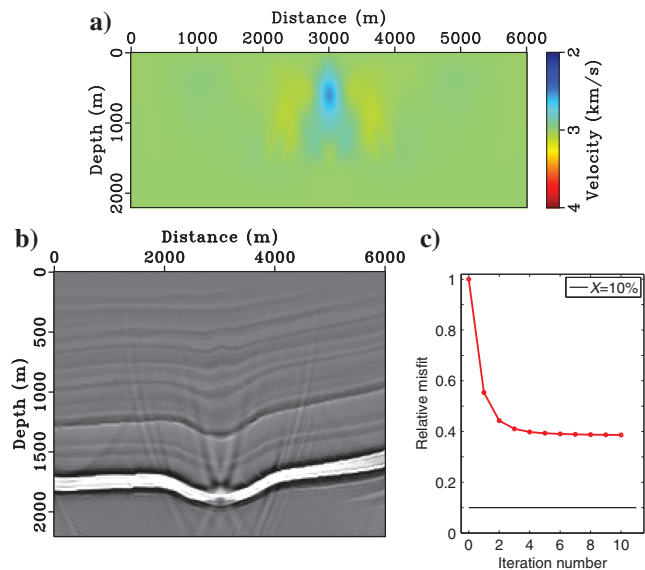
Figure 19. (a) The inverted background velocity model without model extension and multiscale approach, (b) reflectivity model, and (c) relative data residual.
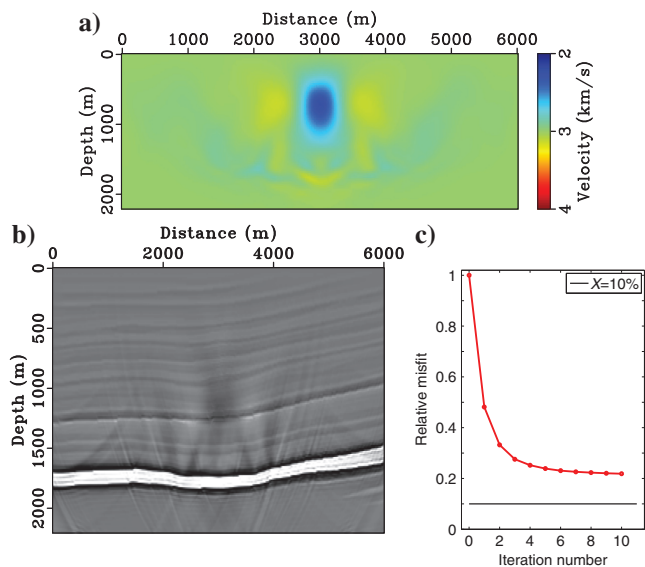
Figure 20. (a) The inverted background velocity model by using multiscale approach but without model extension, (b) the reflectivity model, and (c) the relative data residual.
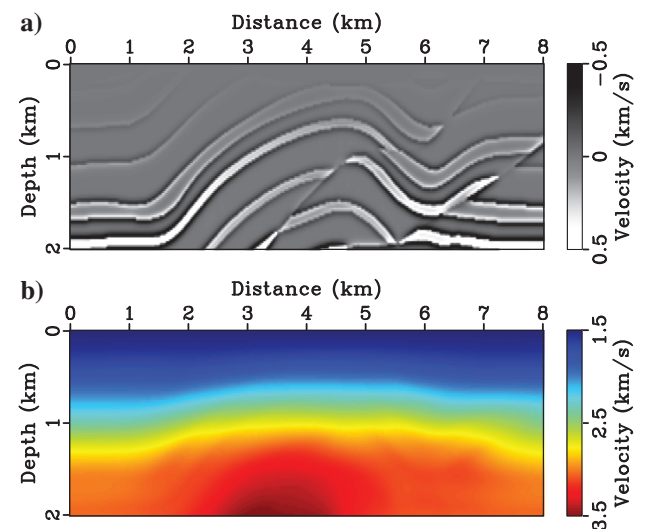

Figure 21. (a) The extended velocity perturbation $\delta\bar{r}$ at $h = 0$ m and (b) the true background velocity.
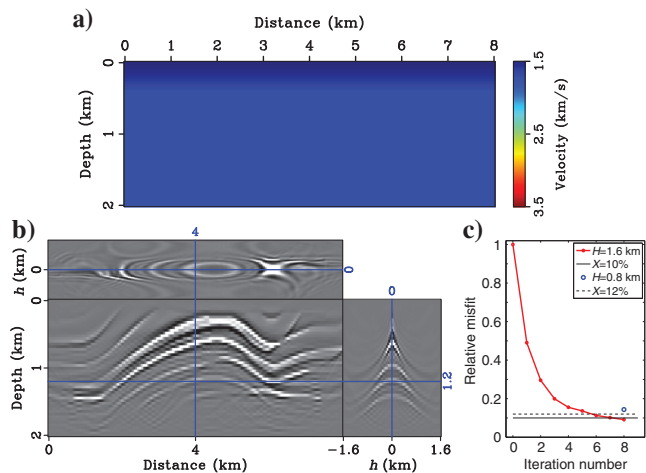

Figure 22. (a) The initial background used in inversion, (b) the ELSRTM image by using the initial background velocity, and (c) the relative data residual with different offset ranges.

and vice versa. When the velocity is nearly correct, the angle gathers are nearly flat, and they can be reliably sampled with a coarse angle step, corresponding to a small H. For a substantially incorrect velocity, the angle gathers are far from flat in angle, implying the need for finer sampling for nonaliased representation, corresponding to larger maximum H. Thus, the angle-offset correspondence provides an alternate view of the relation between the velocity correctness and H.

In all the numerical examples, most of the computational effort is spent on solving the normal equation 7. As a result, the total number of modeling/migration pairs is really the cost in these experiments. The total computational time for the Gaussian example is approximately 1% of the original problem by using the finest gird and a full offset al. along, if we assume the convergence rates are the same for both cases. For the overthrust example, although there are only two stages, the offset range is dramatically reduced at the first stage with the adaptive approach and the cost is also greatly reduced (approximately 2% of the original problem). Although there are only two

refinement stages, the subsurface offset is greatly reduced at the first few iterations. Note that with more refinements and more iterations for the inner and outer loops, the computational performance of the adaptive algorithm will be further improved in comparison with the nonadaptive case. More velocity updates in early stages (coarse grids) will improve the computational efficiency significantly.

As is clear from the discussion in the "Theory" section, from the point of view of our adaptive algorithm, the data misfit tolerance level X is somewhat arbitrary. In our examples, we chose it based on the lowest data misfit achievable with a small number of CG iterations to solve the ELSRTM problem (equation 7 with $\alpha = 0$). However, ideally, this tolerance should approximate the least root-mean-
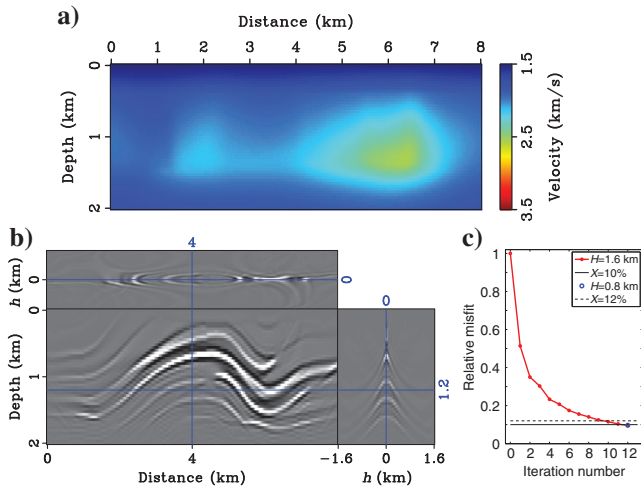

Figure 23. (a) The first updated background velocity, (b) the ELSRTM image by using the first updated background velocity, and (c) the relative data residual with different offset ranges.
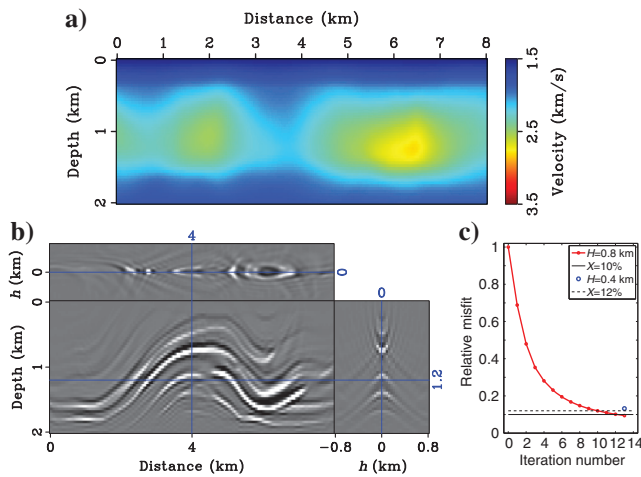

Figure 24. (a) The second updated background velocity, (b) the ELSRTM image by using the second updated background velocity, and (c) the relative data residual with different offset ranges.
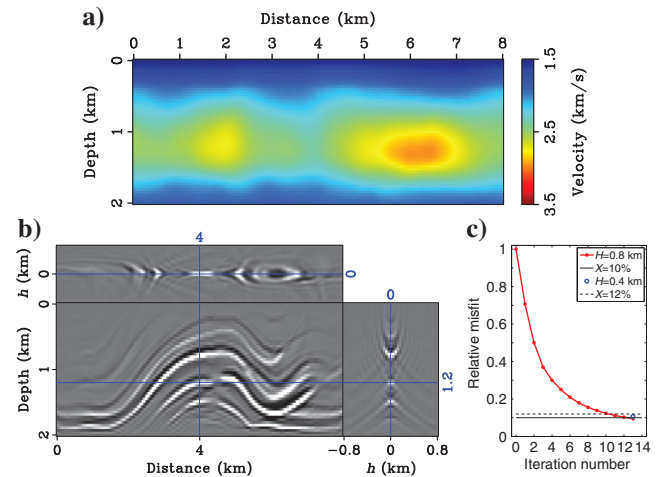

Figure 25. (a) The eighth updated background velocity, (b) the ELSRTM image by using the eighth updated background velocity, and (c) the relative data residual with different offset ranges.
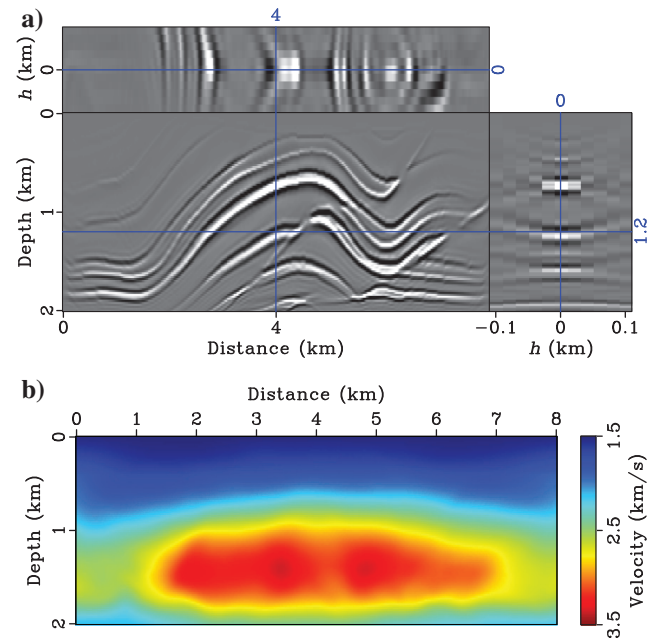

Figure 26. Stage 2: (a) The ELSRTM image and (b) the inverted background velocity.

square (rms) error attainable with a physical (nonextended) model. With one more level of adaptation, our algorithm could attain this goal as well. Because driving the penalty term ($p(v, \alpha, H)$ in the "Theory" section) toward zero is equivalent to approaching the space of physical models, adjusting X so that α can become sufficiently large, or H can become sufficiently small, will assure that X actually approximates the least attainable rms error. How to define "sufficiently" and how to organize an efficient algorithm based on these observations remain a topic of research.

Note that the extended inversion example converges to a good kinematic solution, whereas nonextended FWI may not (see the lens example) provided the same number of modeling/migration pairs. There are more ways to reduce the number of iterations. In solving normal equation 7 by the CG iteration, we have used multiplication by "z^2" as a preconditioner, to accelerate convergence. However, better methods are available (Hou and Symes, 2016) and could potentially further reduce the cost of the inner iteration. Similarly, we used steepest descent with quadratic backtrack line search as the optimization method for updating v. More efficient gradient-based optimization methods can potentially be used to further improve the convergence rate of the outer iteration.

Evidently, the major motivation for the modifications of extended inversion presented in this paper is to bring it closer to practicality for 3D applications. In the "Introduction" section, we reminded the reader that subsurface-offset modeling and inversion involve (at the discrete level) a full-matrix multiply at each time step, the cost of which can easily overwhelm the cost of an ordinary time-domain solution of the wave equation. This additional cost has been a serious impediment to the adoption, or even exploration, of 3D subsurface-offset extended waveform inversion. The techniques introduced in this paper certainly reduce this cost premium, by orders of magnitude for frequency ranges and length scales typical of seismic exploration. However, a definitive assessment of practical feasibility awaits tests with a 3D implementation.

Apart from cost, many other questions about 3D subsurface-offset extended waveform inversion remain to be answered. For example, with a 2D subsurface-offset plane, should the active region be a square of side $2H$, or a rectangle with independently adapted sides, or some more complex region? In dealing with highly refractive subsurface structures, such as salt, gas chimneys, and the like, will vertical offsets be needed to supplement the horizontal (Biondi and Shan, 2002) (the answer is almost certainly "yes," and in 2D as well)? What is the impact of the difference between the inline and crossline sampling typical of wide-azimuth streamer surveys and of the sparse sampling of nodes in ocean bottom seismic surveys? Is 3 Hz even a low enough frequency such that the 5D extended model space is computationally tractable at the necessary sample rates? All of these questions remain to be addressed in future work.

CONCLUSION

The objective of this study is to address one of the central problems of subsurface offset extended waveform inversion, namely, its computational cost, which greatly hinders any practical application of this promising new technique. Adaptive determination of a sufficient offset range to assure data fit, in concert with grid coarsening, yields a significant improvement in computational efficiency. The adaptive concept progressively reduces the subsurface-offset range throughout inversion, which can naturally bring extended model toward physical model. The 2D synthetic experiments with

a version of this algorithm based on linearized (Born) constant density acoustics suggests that cost reduction of one or two orders of magnitude is attainable, with recovery of short- and long-scale features of the velocity model.

ACKNOWLEDGMENTS

REFERENCES

Albertin, U., P. Sava, J. Etgen, and M. Maharramov, 2006, Adjoint wave equation velocity analysis: 76th Annual International Meeting, SEG, Expanded Abstracts, 3345–3349.

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 G. Shan, 2002, Prestack imaging of overturned reflections by reverse time migration: 72nd Annual International Meeting, SEG, Expanded Abstracts, 1284–1287.

Claerbout, J. F., 1985, Imaging the earth's interior: Blackwell Scientific Publishers.

Fei, W., and P. Williamson, 2010, On the gradient artifacts in migration velocity analysis based on differential semblance optimization: 80th Annual International Meeting, SEG, Expanded Abstracts, 4071–4076.

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.

Gauthier, O., A. Tarantola, and J. Virieux, 1986, Two-dimensional nonlinear inversion of seismic waveforms: Geophysics, **51**, 1387–1403, doi: 10.1190/1.1442188.

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.

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.

Kern, M., and W. Symes, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysical Prospecting, **99**, 565–614, doi: 10.1111/j.1365-2478.1994.tb00231.

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.

Li, M., J. Rickett, and A. Abubakar, 2013, Application of the variable projection scheme to frequency-domain full-waveform inversion: Geophysics, **78**, no. 6, R249–R257, doi: 10.1190/geo2012-0351.1.

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.

Mulder, W., 2014, Subsurface offset behaviour in velocity analysis with extended reflectivity images: Geophysical Prospecting, **62**, 17–33, doi: 10.1111/1365-2478.12073.

Nocedal, J., and S. Wright, 1999, Numerical optimization: Springer Verlag.

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.

Rickett, J., 2012, The variable projection method for waveform inversion with an unknown source function: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0663.1.

Sava, P., and S. Fomel, 2003, Angle-domain common-image gathers by wavefield continuation methods: Geophysics, **68**, 1065–1074, doi: 10.1190/1.1581078.

Shan, G., and Y. Wang, 2013, RTM based wave equation migration velocity analysis: 83rd Annual International Meeting, SEG, Expanded Abstracts, 4726–4731.

Shen, P., 2004, Wave equation migration velocity analysis by differential semblance optimization: Ph.D. thesis, Rice University.

Shen, P., 2012, An RTM based automatic migration velocity analysis in image domain: 82nd Annual International Meeting, SEG, Expanded Abstracts, doi: 10.1190/segam2012-0214.1.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, no. 5, VE49–VE59, doi: 10.1190/1.2972021.

Shen, P., W. Symes, S. Morton, and H. Calandra, 2005, Differential semblance velocity analysis via shot profile migration: 75th Annual International Meeting, SEG, Expanded Abstracts, 2249–2252.

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.

Stolk, C. C., and M. V. De Hoop, 2001, Seismic inverse scattering in the 'wave-equation' approach: Technical Report 2001-047, The Mathematical Sciences Research Institute.

Stolk, C. C., M. V. de Hoop, and W. Symes, 2009, Kinematics of shot-geophone migration: Geophysics, **74**, no. 6, WCA18–WCA34, doi: 10.1190/1.3256285.

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., 2014, Seismic inverse problems: Recent developments in theory and practice: Presented at the Inverse Problems — From Theory to Application, Proceedings, Institute of Physics.

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.

ten Kroode, F., 2014, A Lie group associated to seismic velocity estimation: Presented at the Inverse Problems — From Theory to Application, Proceedings, Institute of Physics.

van Leeuwen, T., and W. Mulder, 2009, A variable projection method for waveform inversion: 71st Annual International Conference and Exhibition, EAGE, Extended Abstracts, U024.

Vigh, D., K. Jiao, W. Huang, N. Moldoveanu, and J. Kapoor, 2013, Long-offset-aided full-waveform inversion: 76th 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: 81st 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.

Vyas, M., and Y. Tang, 2010, Gradients for wave-equation migration velocity analysis: 80th Annual International Meeting, SEG, Expanded Abstracts, 4077–4081.

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.