THE RICE INVERSION PROJECT

Raanan Dafni, Lei Fu, Jie Hou, Guanghui Huang, Rami Nammour, and William Symes

Annual Report 2016
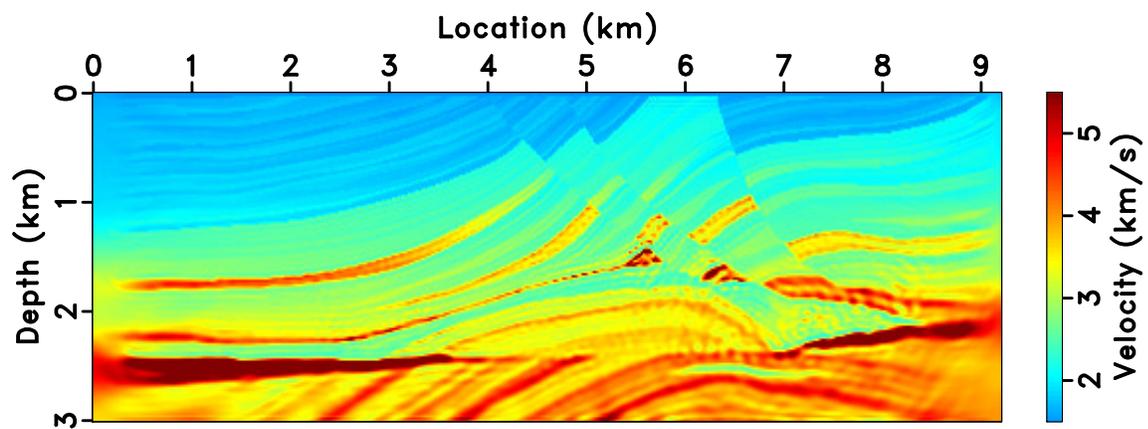
TRIP16 — TABLE OF CONTENTS

ii

INTRODUCTION TO THE 2016 ANNUAL REPORT

Welcome to the 2016 Annual Report volume of The Rice Inversion Project. This volume contains manuscripts of papers, abstracts and reports completed during the course of the project year.

The papers

- "Accelerating Extended Least Squares Migration with Weighted Conjugate Gradient Iteration", Hou and Symes (in the 2015 Annual Report);

- "An alternative formula for approximate extended Born inversion", Hou and Symes;

- "A discrepancy based penalty method for extended waveform inversion", Fu and Symes;

- "An adaptive multiscale algorithm for efficient extended waveform inversion", Fu and Symes;

- "Full Waveform Inversion via Source-Receiver Extension", Huang, Nammour, and Symes

- "Scattering and dip angle decomposition based on subsurface offset extended wave-equation migration", Dafni and Symes; and

- "Kinematic artifacts in the subsurface-offset extended image and their elimination by a dip-domain specularity filter", Dafni and Symes

have been accepted for publication in *Geophysics* and have appeared or will appear in print by mid-2017. "Inversion velocity analysis in subsurface offset domain", Hou and Symes, is in review for the same journal.

I am pleased to acknowledge our debt to Sergey Fomel and other contributors to the Madagascar project, whose reproducible research framework makes our approach to distribution of reports possible.

– WWS, April 2017

Approximate Gauss-Newton iteration for Full Waveform Inversion

Jie Hou*, William W. Symes, The Rice Inversion Project, Rice University

ABSTRACT

Full waveform inversion (FWI) reconstructs the subsurface model from observed seismic data by minimizing the energy of the difference between predicted data and observed data. It is often formulated as an optimization problem and solved by computationally intensive iterative methods. The steepest ("gradient") descent method, with or without line search, uses only first derivative information and is severely slowed by the ill-conditioned behavior of FWI. Newton's method and its relatives take the objective curvature into account, hence converge more quickly, at the cost of more expensive iterates. We describe a preconditioned gradient descent method that approximates the iterates, hence the convergence rate, of the Gauss-Newton method while having essentially the same cost per iterate as the steepest descent method. The preconditioned gradient is based on an approximate inverse to the Born modeling operator. Numerical examples demonstrate that preconditioned gradient descent converges dramatically faster than both conventional gradient descent and the L-BFGS quasi-Newton method.

INTRODUCTION

Full waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009) aims to recover detailed models of the subsurface through a model-based data-fitting procedure, via minimization of the difference between predicted and observed data in the least squares sense. FWI has the potential to recover high-resolution model, given an initial model that predicts traveltimes to within a half-wavelength at frequencies with sufficiently high S/N. It extends in principle to any modeled physics and data geometry, although many questions remain about recovering multiple parameters, robustness, and computational efficiency.

3

Mathematically, FWI poses to a large-scale nonlinear minimization problem. Due to problem size, only local gradient-based methods are feasible (Gauthier et al., 1986; Pratt, 1999; Crase et al., 1990). An obvious candidate, still used in many FWI exercises, is gradient descent, more properly known as the steepest descent algorithm (Nocedal and Wright, 1999). However, steepest descent converges slowly for ill-conditioned problems, for which the objective changes much more rapidly in some directions than in others, and FWI tends to be ill-conditioned. Newton's method, on the other hand, takes second order (curvature) information into account and thus generally converges to a local minimizer in many fewer iterations than required by steepest descent (Pratt et al., 1998; Akcelik et al., 2003; Métivier et al., 2014). The Newton update (and its close relative, the Gauss-Newton update) requires the solution of a linear system, of size equal to the number of degrees of freedom in the model. Such systems must themselves be solved iteratively, due to their size, so that the (Gauss-)Newton steps may be quite expensive. Alternatively, quasi-Newton methods build up an approximation to the inverse Hessian using the gradient at the current step and previous steps. The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method (or its limited memory version, L-BFGS) is arguably the most widely-used quasi-Newton method (Nocedal and Wright, 1999).

The main contribution of this paper is to show how to use an inexpensive modification of RTM to approximate the inverse Gauss-Newton Hessian, or pseudoinverse of the Born modeling operator (Hou and Symes, 2016), thus achieving something close to the Gauss-Newton convergence rate with updates of roughly the same expense as steepest descent. Approximations to the inverse Hessian have a long history in computational work on FWI. Many authors have used a diagonal or small-bandwidth approximation to the inverse (Chavent and Plessix, 1999; Shin et al., 2001; Operto et al., 2004), or a version of the Generalized Radon Transform (Kirchhoff, ray-Born,...) inversion (Jin et al., 1992; Qin et al., 2015; Métivier et al., 2015), to accelerate FWI iterations. Our work is close to the latter in spirit, however we use an approximate inverse or "true amplitude migration" operator that does not require any ray-tracing computations. Similar true amplitude migration operators have been discussed by others (Xu et al., 2011; ten Kroode, 2012), in the context of reconstructing angle-based reflectivity. Hou and Symes (2015b) provide an equivalent construction in the subsurface offset domain, and assess its accuracy as an approximate inverse. Subsequently we showed that this approximate inverse is very effective for accelerating linearized inversion ("least squares migration"), both in the subsurface offset domain (Hou and Symes, 2015a) and for physical domain least squares (Hou and Symes, 2016). Here we illustrate

a similar effect for nonlinear full waveform inversion.

This paper is organized as follow: we first review the theory of FWI and common numerical algorithms used for optimization; we then show how to precondition the gradient and modify the approximate inverse operator; we end with numerical test on 2D Marmousi model, demonstrating FWI with preconditioned gradient exhibits faster convergence than both steepest (gradient) descent method and L-BFGS method, all equipped with the same backtracking line search.

THEORY

The 2D wave equation in acoustic constant density medium can be expressed as

$$\frac{1}{v^2(\mathbf{x})}\frac{\partial^2 u}{\partial t^2}(\mathbf{x},t) - \nabla^2 u(\mathbf{x},t) = f(t,\mathbf{x},\mathbf{x_s});$$
$$u(\mathbf{x},t) \equiv 0, t \ll 0. \tag{1}$$

Here $\mathbf{x}$ denotes the location in the modeling region, $v(\mathbf{x})$ is the acoustic velocity, u is the acoustic potential and f is the source term, also parametrized by source position $\mathbf{x_s}$ and time t. The forward modeling operator $\mathcal{F}$ maps the model $\mathbf{m} = \mathbf{v}^2$ to seismic data $\mathbf{d}$:

$$\mathcal{F}[\mathbf{m}] = \mathbf{d}. \tag{2}$$

FWI is often formulated as an optimization problem by minimizing the following objective function :

$$J[\mathbf{m}] = \frac{1}{2}\|\mathcal{F}[\mathbf{m}] - \mathbf{d}_o\|^2[+\text{regularizing terms}]. \tag{3}$$

Here $\mathbf{m}$ is the velocity model to recover, $\mathbf{d}_o$ is the observed seismic data and $\|\cdot\|^2$ stands for L^2 norm.

Iterative Optimization

Iterative methods for minimization of $J[\mathbf{m}]$ increment the model along a search direction $\mathbf{p}$:

$$\mathbf{m}_{i+1} = \mathbf{m}_i + \alpha_i \mathbf{p}_i, \tag{4}$$

where i is the iteration number and α_i is a step length, chosen by an approximate optimization along the line through $\mathbf{m}_i$ in direction $\mathbf{p}_i$, to reduce the objective function.

Steepest (or gradient) descent uses the negative gradient of J as search direction:

$$\mathbf{p}_i = -\mathbf{g}_i = -F^T(\mathcal{F}[\mathbf{m}_i] - \mathbf{d}_o). \tag{5}$$

Here $\mathbf{g}_i$ represent the gradient, $F = D\mathcal{F}$ is the derivative of $\mathcal{F}$ with respect to model, also known as the Born modeling operator. The superscript ^{T} denotes the adjoint operator, defined with respect to the Euclidean inner product via the "dot product test":

$$F^T \delta \mathbf{d} \cdot \delta \mathbf{m} = \delta \mathbf{d} \cdot F \delta \mathbf{m}, \tag{6}$$

asserted to hold for any data perturbation $\delta \mathbf{d}$ and model perturbation $\delta \mathbf{m}$. F^T is a version of the RTM operator, for which relatively efficient computational methods are available (Plessix, 2006).

The update direction for Newton's method is the solution of a linear system:

$$F^T F \mathbf{p}_i + D^2 \mathcal{F}^T(\mathbf{p}_i, \mathcal{F} - \mathbf{d}_o) = -\mathbf{g}_i. \tag{7}$$

Compared to steepest descent, Newton's method converges extremely fast, at least near the minimum of the objective function. This is because the solution of the Newton system 7 compensates for the ill-conditioning mentioned earlier and to some extent for the nonlinearity of the objective function. The second term on the left-hand side of equation 7 vanishes when the model precisely predicts the data, and may be viewed as a nuisance, as it involves the second derivative of the modeling operator (however see Métivier et al. (2014), who show the importance of this term when nonlinear (multiple) scattering is strong). The Gauss-Newton method neglects the second term, and computes $\mathbf{p}_i$ as the solution of

$$F^T F \mathbf{p}_i = -\mathbf{g}_i. \tag{8}$$

Whether Newton (equation 7) or Gauss-Newton (equation 8 updates are chosen, a very large linear system must be solved, necessarily by an iterative method such as conjugate gradients (Nocedal and Wright, 1999; Akcelik et al., 2003). As a result, each iteration typically requires a considerable number of inner iterations to solve for $\mathbf{p}_i$, each inner iteration involving a Born modeling - RTM cycle.

Quasi-Newton methods build up an approximation to the solution $\mathbf{p}_i$ of either equation 7 or 8 using a generalization of the secant method, combining current

gradient and previous search directions (Nocedal and Wright, 1999) to produce a low rank approximation to the operator on the left hand sides of these equation, allowing easy solution. An example is the BFGS algorithm. The limited memory version BFGS method, L-BFGS, avoids excessive memory use by storing only a limited number of previous search directions on a first-in, first-out basis (Nocedal and Wright, 1999).

Approximating Gauss-Newton

In this section, we will show how to approximately solve the Gauss-Newton system 8 by applying an inexpensive modification to the RTM operator. The update direction (or preconditioned gradient) for this method is

$$\mathbf{p_i} = -F^\dagger(\mathcal{F}(\mathbf{m}_i) - \mathbf{d}) \tag{9}$$

where the modified RTM operator $F^\dagger$ approximates the pseudoinverse of the Born modeling operator $(F^T F)^{-1} F^T$.

The approximate pseudoinverse used in our work originates in the construction of a computable approximate inverse to the subsurface offset extended Born modeling operator $\bar{F}$, which acts on model perturbations depending on a subsurface offset (h-) axis (ten Kroode, 2012; Hou and Symes, 2014, 2015b):

$$\bar{F}^\dagger = W_{\text{model}}^{-1} \bar{F}^T W_{\text{data}}, \tag{10}$$

where $W_{\text{model}}^{-1} = 4v_0^5 LP$, $W_{\text{data}} = I_t^4 D_{z_s} D_{z_r}$. L is similar to the Laplacian: in the wavenumber domain, it is multiplication by $|\mathbf{k}_{xz}||\mathbf{k}_{hz}|$. I_t is time integration. $\bar{F}^T$ is the Euclidean adjoint of $\bar{F}$ (extended RTM) and D_{z_s}, D_{z_r} are the source and receiver depth derivatives. P is an oscillatory integral operator approximately equal to 1 for nearly focused images. We will deal here only with the focused case (kinematically data-consistent velocity model), so we will neglect P. Since the extended image volume is then focused at $h = 0$, we can neglect the energy at nonzero h, so replace $\bar{F}^T$ with ordinary (non-extended) F^T (RTM) in equation 9 to obtain an approximate pseudoinverse for F (Hou and Symes, 2016):

$$F^\dagger = W_{\text{model}}^{-1} F^T W_{\text{data}}. \tag{11}$$

In this case, only W_{model} depends on the subsurface offset h, and can be simply applied by padding with zeroes along the subsurface offset axis. Alternatively, we

can use operators defined only on non-extended ($h = 0$) models and data pertur-
bations: $W_{\text{model}}^{-1} = 8v_0^4|k_z|$, $W_{\text{data}} = |\omega|I_t^4 D_{z_s} D_{z_r}$. This reformulation arises from the
ray geometry of reflection in the focused case, as shown in Figure 1 : at $h = 0$,

$$|k_{xz}| = 2\frac{|\omega|}{v_0}cos\theta$$

$$|k_{hz}| = \frac{|k_z|}{cos\theta}.$$

(12)
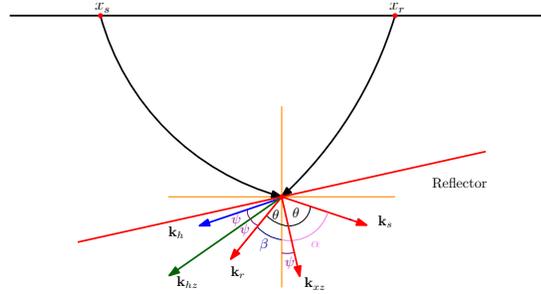
Figure 1: Sketch of the subsurface for h = 0. x_s, x_r represent the location of the source and receiver; $\mathbf{k}_s, \mathbf{k}_r$ are the slowness vectors of the source and receiver rays; $\mathbf{k}_h$ is symmetric to $\mathbf{k}_s$ in terms of vertical axis; θ is half of the angle between $\mathbf{k}_s$ and $\mathbf{k}_r$; α is the incident angle between source ray and vertical direction; β is the incident angle between receiver ray and vertical direction; ψ is half of the angle between $\mathbf{k}_h$ and $\mathbf{k}_r$.

Notice the weight operators W_{model}, W_{data} in either case are filters, and to-
gether are of order zero in frequency. Therefore, they will not change the fre-
quency balance of the RTM image volume, but only alter amplitude and phase.
Notice also that the approximate inverse operator has almost the same computa-
tional cost as the conventional RTM operator.

FWI with Approximate Gauss-Newton Iteration

FWI can be implemented in both time domain and frequency domain. Here we
will test in time domain only. The workflow of FWI with the approximate Gauss-
Newton, or preconditioned gradient descent, method is :

1. compute the data residual and its norm (equation 3);

2. apply the approximate inverse operator on the data residual (equation 9);

3. determine a suitable step length α_i using backtracking line-search method;

4. update the velocity model (equation 4).

NUMERICAL EXAMPLES

We apply the algorithm described above to an example based on the Marmousi model (Versteeg and Grau, 1991). The model is resampled to 20 m grid. A 500 m water layer (not shown in the figures) is added on the top. A fixed spread of 231 sources and 461 receivers are placed at 20 m below the surface. A 2-8 finite difference scheme is used to simulate 4 s seismic data. The source, regarded as known in these experiments and not updated, is a (2.5-5-20-25)-Hz band pass-wavelet with 2ms time sample. The initial model (Fig 7b) used in FWI is the smoothed version of the true model (Fig 5).

Both the gradient and preconditioned gradient at the first iteration are shown in Fig 2. The amplitude distribution of the preconditioned gradient is clearly more balanced: both the shallow part and the deep part are well recovered, while most of the energy in the conventional gradient focuses in the shallow part.
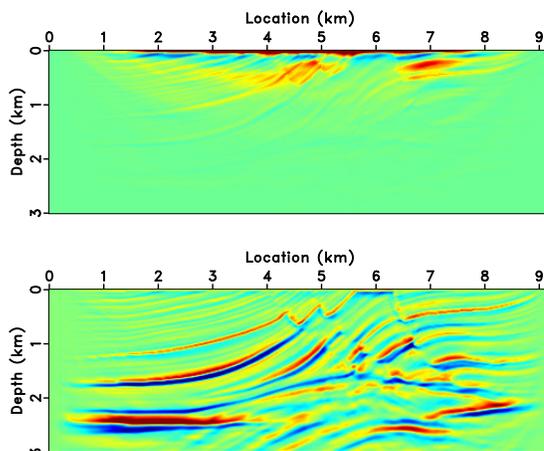
Figure 2: (a) Gradient; (b) Preconditioned gradient for the first iteration.

We then optimize FWI objective function with both gradient descent method and L-BFGS method. The gradient descent method will be applied with both the conventional gradient and preconditioned gradient; as explained above, in the second case, the iteration approximates the Gauss-Newton method. The convergence history is shown in Figure 3. We can see the approximate Gauss-Newton method reduces the objective function to 1 % of its initial value in 10 iterations, whereas steepest descent and L-BFGS, respectively, achieves to reduce the residual only to 30 % and 15 % in 40 iterations.

Figure 6 displays the result of 100 steepest descent iterations; Figure 8, the
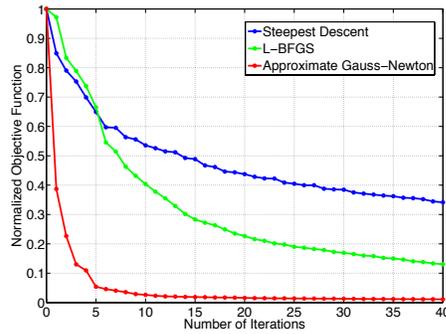
Figure 3: Blue line: residual for the steepest descent method, relative to initial residual. Red line: relative residual for approximate Gauss-Newton method. Cost per step is approximately the same. Green line: residual for the L-BFGS method.

result of 40 L-BFGS; Figure 7, 1 iteration of approximate Gauss-Newton iteration; and Figure 9, 40 iterations of approximate Gauss-Newton iteration. We set the L-BFGS iteration to retain 15 prior search directions. We can see 1 iteration of approximate Gauss-Newton method gives a result comparable to that of 100 iterations of steepest descent or 40 L-BFGS iterations. All three algorithms can recover the shallow part of the model accurately, but only approximate Gauss-Newton recovered the deep part at all well.
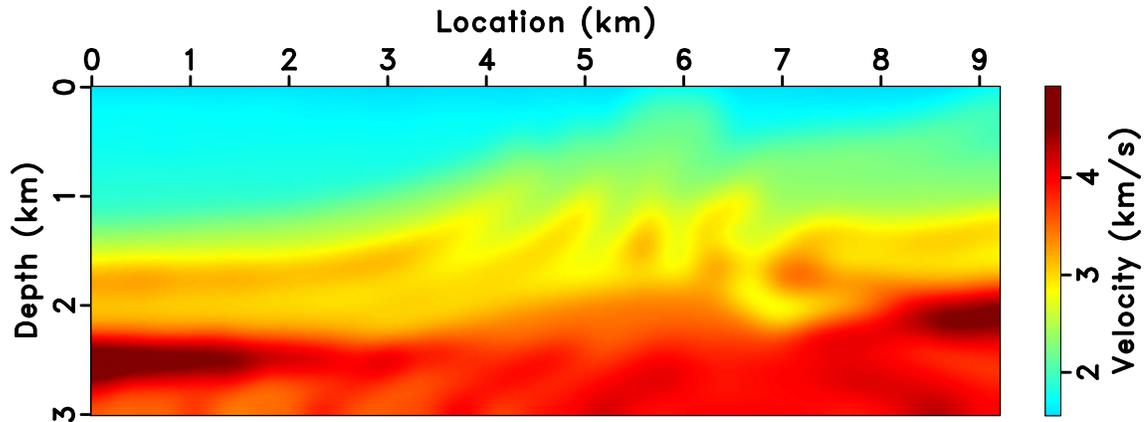
Figure 4: Smoothed version of true model as initial guess
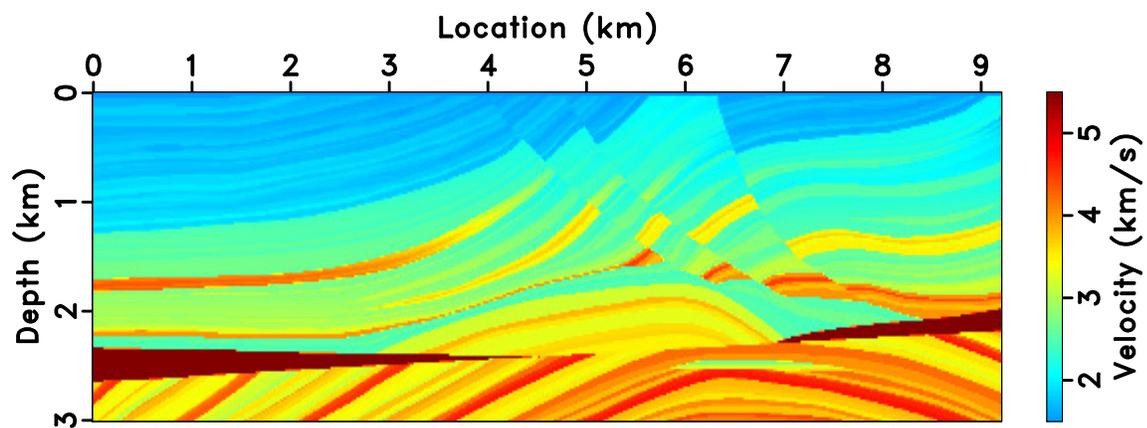
Figure 5: Marmousi model
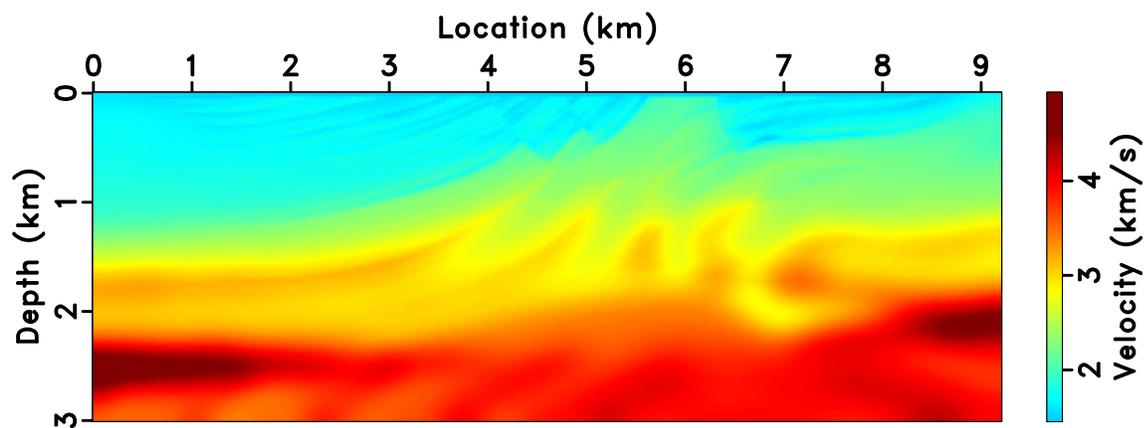
Figure 6: Recovered model after 100 iteration steepest descent method
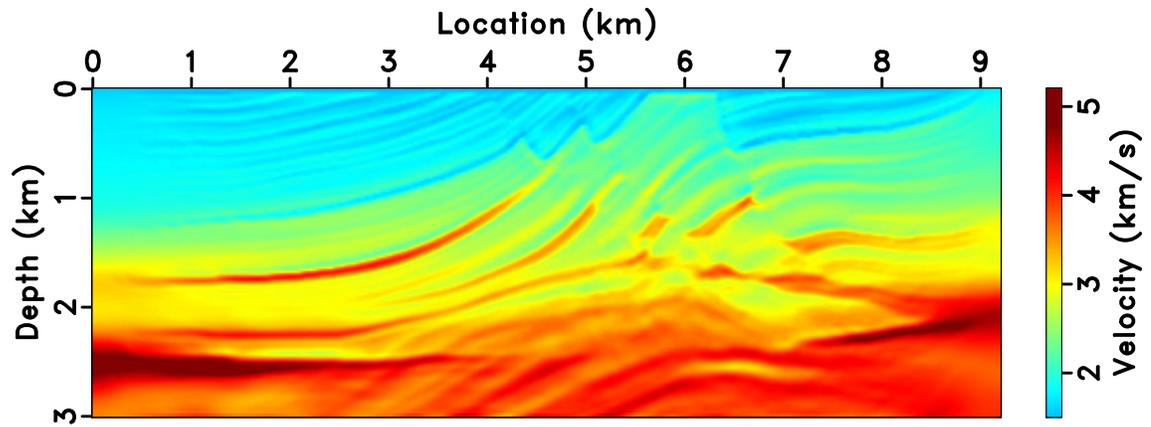
Figure 7: Recovered model after 1 iteration approximate Gauss-Newton method
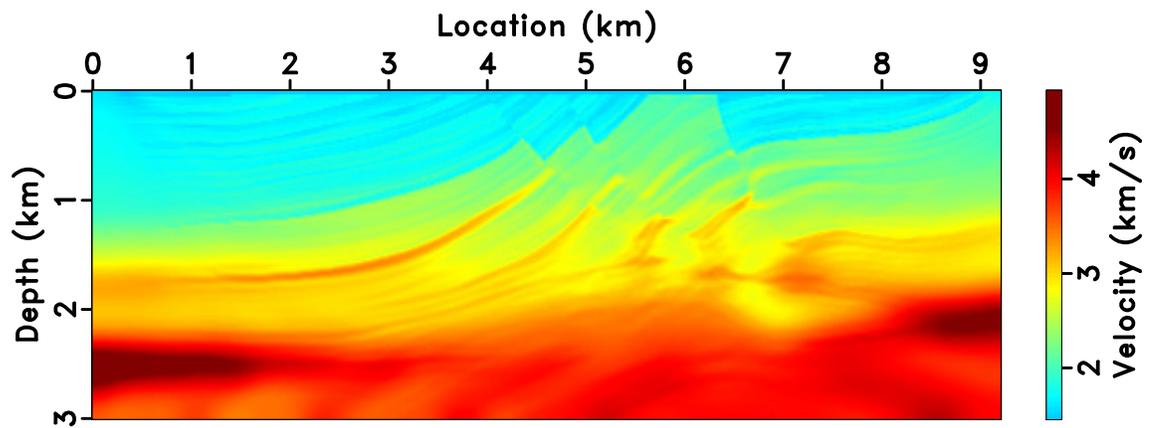
Figure 8: Recovered model after 40 iteration L-BFGS method
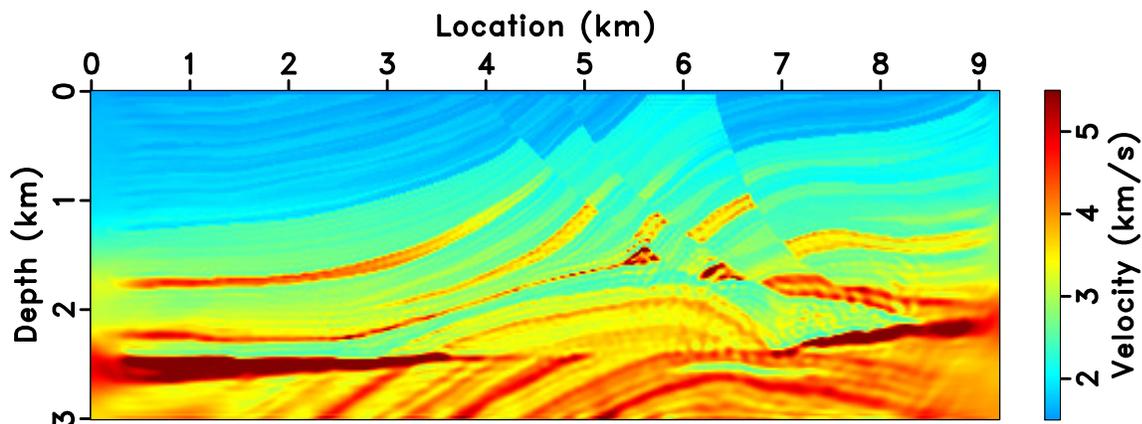
Figure 9: Recovered model after 40 iteration approximate Gauss-Newton method

DISCUSSION AND CONCLUSIONS

We have described a preconditioned gradient descent algorithm that approximates the Gauss-Newton algorithm and its rate of convergence, by replacing the adjoint (RTM) operator in the definition of FWI gradient with an approximate inverse operator. In an initial test using the Marmousi velocity model, the approximate Gauss-Newton algorithm significantly outperformed both steepest (gradient) descent and L-BFGS iterations. Most importantly, the proposed algorithm has computational cost per iteration roughly equal to that of the steepest descent method, that is, computation of the FWI gradient.

Several fundamental puzzles deserved mention. The approximate pseudoinverse to the Born modeling operator, at the heart of the approximate Gauss-Newton method, is based on high-frequency asymptotic analysis of Born scattering, even though the actual computations do not involve any ray tracing or other asymptotic calculations. The asymptotic analysis assumes separation of scales, a model/data relationship not required for the formulation of FWI, and not valid in principle for the model in our example (Marmousi). We do not at present understand the apparent effectiveness of the algorithm in this setting. Multiparameter models (variable density acoustics, elasticity) may admit a similar approach. We do not begin to understand how to formulate high frequency asymptotic approximation in the presence of attenuation.

Finally, we note that the algorithm described here does not solve the cycle-skipping problem. The standing assumption in this work is that the initial ve-

locity model predicts traveltimes within at most a half-wavelength error at significant S/N. Given that assumption, the approximate Gauss-Newton algorithm appears to offer significant efficiency advantages over other common approaches to FWI.

ACKNOWLEDGMENTS

REFERENCES

Akcelik, V., J. Bielak, G. Biros, I. Epanomeritakis, A. Fernandez, O. Ghattas, E. Kim, J. Lopez, D. O'Hallaron, T. Tu, and J. Urbanic, 2003, High resolution forward and inverse earthquake modeling on terascale computers: Presented at the Supercomputing 03, Proceedings, Association for Computing Machinery.

Chavent, G., and R. Plessix, 1999, An optimal true-amplitude least-squares prestack depth-migration operator: Geophysics, **64**, 508–515.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538.

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

Hou, J., and W. Symes, 2014, An approximate inverse to the extended Born modeling operator: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 3784–3789.

———, 2015a, Accelerating extended least squares migration with weighted conjugate gradient iteration: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4243–4248.

———, 2015b, An approximate inverse to the extended Born modeling operator: Geophysics, **80**, no. 6, R331–R349.

———, 2016, Accelerating least squares migration with weighted conjugate gradient iteration: 78th Annual International Conference and Exhibition, Expanded Abstract, European Association for Geoscientists and Engineers, P104.

Jin, S., R. Madariaga, J. Virieux, and G. Lambaré, 1992, Two-dimensional asymptotic iterative elastic inversion: Geophysical Journal International, **108**, 575–588.

Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations, *in* Conference on Inverse Scattering: Theory and Applications: Society for Industrial and Applied Mathematics, 206–220.

Métivier, L., F. Bretaudeau, R. Brossier, S. Operto, and J. Virieux, 2014, Full waveform inversion and the truncated newton method: quantitative imaging of complex subsurface structures: Geophysical Prospecting, **62**, 1353–1375.

Métivier, L., R. Brossier, and J. Virieux, 2015, Combining asymptotic linearized inversion and full waveform inversion: Geophysical Journal International, **201**, 1682–1703.

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

Operto, S., C. Ravaut, L. Improta, J. Virieux, A. Herrero, and P. Dell'Aversana, 2004, Quantitative imaging of complex structures from dense wideaperture seismic data by multiscale traveltime and waveform inversions: a case study: Geophysical Prospecting, **52**, no. 6, 625–651.

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.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory, and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R. G., C. Shin, and G. J. Hick, 1998, Gaussnewton and full newton methods in frequencyspace seismic waveform inversion: Geophysical Journal International, **133**, no. 2, 341–362.

Qin, B., T. Allemand, and G. Lambaré, 2015, Full waveform inversion using preserved amplitude reverse time migration: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1252–1257.

Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin, 2001, Efficient calculation of a partial-derivative wavefield using reciprocity for seismic imaging and inversion: Geophysics, **66**, no. 6, 1856–1863.

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

ten Kroode, F., 2012, A wave-equation-based Kirchhoff operator: Inverse Problems, 115013:1–28.

Versteeg, R., and G. Grau, 1991, Practical aspects of inversion: The Marmousi experience: Proceedings of the EAEG, The Hague.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in explo-

ration geophysics: Geophysics, **74**, WCC127–WCC152.

Xu, S., Y. Zhang, and B. Tang, 2011, 3D angle gathers from reverse time migration: Geophysics, **76**, no. 2, S77–S92.

Inversion velocity analysis via approximate Born inversion

Jie Hou*, *William W. Symes*, *The Rice Inversion Project*, *Rice University*

ABSTRACT

Optimization-based Migration Velocity Analysis (MVA) updates long wavelength velocity information by minimizing an objective function that measures the violation of a semblance condition, applied to an image volume. Differential Semblance Optimization (DSO) forms a smooth objective function both in velocity and data, regardless of the data frequency content. Depending on how the image volume is formed, however, the objective function may not be minimized at a kinematically correct velocity, a phenomenon characterized in the literature (somewhat inaccurately) as "gradient artifacts". In this paper, we will show that the root of this pathology is imperfect image volume formation resulting from various forms of migration, and that the use of linearized inversion (least squares migration) more or less eliminates it. We demonstrate that an approximate inverse operator, little more expensive than RTM, leads to recovery of a kinematically correct velocity.

INTRODUCTION

Full waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009) is capable of recovering detailed models of the subsurface structure through a waveform-based data-fitting procedure. However, it may stagnate at physically meaningless solutions in the absence of a kinematically accurate starting model. Within the limitation to single scattering, migration velocity analysis (MVA) (Yilmaz and Chambers, 1984; Yilmaz, 2001) complements FWI by extracting long scale velocity. Image domain MVA involves construction of an extended image volume, which not only depends on the image point but also on an extra parameter (e.g surface offset, subsurface offset, incidence angle). When velocity and data are kinematically compatible, this volume should have particularly simple structure (flat, focused,...). Deviations from this semblance principle can be

used to drive velocity updates, either by direct measurement (eg. residual move-out picking, (Stork, 1992; Lafond and Levander, 1993; Liu and Bleistein, 1995; Biondi and Sava, 2004)) or via optimization of an objective function.

This paper focuses on a particular choice of optimization approach, namely differential semblance optimization (DSO) in the subsurface offset domain (Symes, 2008). Implementations of this MVA approach have been based on double square root migration (Shen et al., 2003), one-way shot record migration (Shen and Symes, 2008), RTM (Shen, 2012; Weibull and Arntsen, 2013), and various sorts of inversion (Biondi and Almomin, 2012; Liu et al., 2014; Lameloise et al., 2015). Both numerical and theoretical (Symes, 2014; ten Kroode, 2014) evidence suggest that this approach should be effective in recovering velocity macromodels under failure conditions for FWI. However other studies have suggested that the method may produce poor velocity update directions, in particular that the objective gradient may be contaminated with artifacts that prevent rapid convergence to a correct velocity (Fei and Williamson, 2010; Vyas and Tang, 2010).

We show here that use of (linearized) inversion to create image volumes largely eliminates the "gradient artifact" pathology, and describe a computationally efficient method to achieve this goal. In fact, the "artifacts" actually are features of the objective function definition, not of the gradients. This is not a new observation: Khoury et al. (2006) showed that subsurface offset DSO, using common azimuth migration to construct the image volume, could produce erroneous velocities returning lower objective function values than the "true" velocity. Liu et al. (2014); Lameloise et al. (2015) confirm this observation and show that use of inverted (rather than migrated) image volumes tends to improve DSO velocity updates, essentially because the inverted image volume is much better focused at the target velocity. Our innovation is to show that good velocity updates may be achieved with image volumes obtained by an approximation to linearized inversion, costing little more than RTM and involving no ray-theory computations (ten Kroode, 2012; Hou and Symes, 2015).

In this paper, we will compare three different imaging operators : conventional RTM operator, the adjoint of the extended Born modeling operator and an approximate inverse to the extended Born modeling operator. The conventional RTM operator only involves cross-correlating the forward source wavefield and backward receiver wavefield. The adjoint operator differs the RTM operator by time derivatives and velocity scaling. Hou and Symes (2015) modify the adjoint operator into an approximate inverse operator by applying model and data-domain weight operators. To distinguish it from the other possibilities, we call MVA with

the approximate inverse operator Inversion Velocity Analysis (IVA).

In the following sections, we will first review the theory of MVA via DSO in the subsurface offset domain; we then compare three imaging operators and their possible influence on the DSO objective function; we end with numerical test on 2D Marmousi model, demonstrating that better imaging leads to better velocities.

THEORY

The 2D constant density acoustic Born modeling operator $F[v_0]$ can be expressed as

$$(F[v_0]\delta v)(\mathbf{x_s}, \mathbf{x_r}, t) = \frac{\partial^2}{\partial t^2} \int d\mathbf{x} dh d\tau G(\mathbf{x_s}, \mathbf{x} - h, \tau)$$
$$\times \frac{2\delta v(\mathbf{x})}{v_0(\mathbf{x})^3} G(\mathbf{x} + h, \mathbf{x_r}, t - \tau). \tag{1}$$

Here δv is the model perturbation or reflectivity, v_0 is the background velocity model and G is the Green's function.

The Born inverse problem is to fit data $d(\mathbf{x_s}, \mathbf{x_r}, t)$ by proper choice of $v_0(\mathbf{x})$ and $\delta v(\mathbf{x})$. As explained for example in Symes (2008), data fit is impossible unless v_0 is kinematically correct to within a half-wavelength error in traveltime prediction. On the other hand, if δv is extended to depend on subsurface offset h, that is, replaced by $\delta \bar{v}(\mathbf{x}, h)$ and used as input in an extended Born modeling operator $\bar{F}$,

$$(\bar{F}[v_0]\delta \bar{v})(\mathbf{x_s}, \mathbf{x_r}, t) = \frac{\partial^2}{\partial t^2} \int d\mathbf{x} dh d\tau G(\mathbf{x_s}, \mathbf{x} - h, \tau)$$
$$\times \frac{2\delta \bar{v}(\mathbf{x}, h)}{v_0(\mathbf{x})^3} G(\mathbf{x} + h, \mathbf{x_r}, t - \tau). \tag{2}$$

then any model consistent data d can be fit with essentially any v_0 by proper choice of $\delta \bar{v}$. The extended modeled data $\bar{F}[v_0]\delta \bar{v}$ is equal to the (non-extended) Born modeled data $F[v_0]\delta v$ when $\delta \bar{v}(\mathbf{x}, h) = \delta v(\mathbf{x})\delta(h)$, that is, when $\delta \bar{v}$ is focused at $h = 0$, which is the semblance condition for subsurface offset extended modeling.

Imaging Operators

The conventional RTM with space shift imaging condition is often formulated as

$$I(\mathbf{x}, h) = \int d\mathbf{x_s} d\mathbf{x_r} dt d\tau G(\mathbf{x_s}, \mathbf{x} - h, \tau)$$
$$\times G(\mathbf{x} + h, \mathbf{x_r}, t - \tau) d(\mathbf{x_s}, \mathbf{x_r}, t), \tag{3}$$

where d is seismic reflection data. It is, however, not exactly the adjoint to the extended Born modeling operator, which is

$$I(\mathbf{x}, h) = \frac{2}{v_0(\mathbf{x})^3} \int d\mathbf{x_s} d\mathbf{x_r} dt d\tau G(\mathbf{x_s}, \mathbf{x} - h, \tau)$$
$$\times G(\mathbf{x} + h, \mathbf{x_r}, t - \tau) \frac{\partial^2}{\partial t^2} d(\mathbf{x_s}, \mathbf{x_r}, t). \tag{4}$$

The difference is minor: the velocity scaling and second time derivative. Nonetheless, the time derivative is quite important in the sense that the extended normal operator $\bar{F}^T \bar{F}$ is order zero. Without the time derivative, the frequency components will be affected. Hou and Symes (2014, 2015) modify the adjoint operator into an approximate inverse to the extended Born modeling operator, by applying model and data domain weight operators. It has the form,

$$\bar{F}^\dagger = W_{\text{model}}^{-1} \bar{F}^T W_{\text{data}}, \tag{5}$$

with

$$W_{\text{model}}^{-1} = 4v_0^5 LP, \ W_{\text{data}} = I_t^4 D_{z_s} D_{z_r}. \tag{6}$$

Here $L = \sqrt{\nabla_{(x,z)}^2 \nabla_{(h,z)}^2}$, I_t is time integration, $\bar{F}^T$ is the adjoint of extend Born modeling operator and D_{z_s}, D_{z_r} are the source and receiver depth derivatives. P is a Fourier-like operator and approximately equal to 1 near $h = 0$, and will be neglected.

Differential Semblance Optimization

The DSO objective function is

$$J[v] = \frac{1}{2} \|hI(\mathbf{x}, h)\|^2, \tag{7}$$

in which $I(\mathbf{x}, h)$ is produced via the application of either RTM (equation 3), adjoint Born modeling (equation 8), or the approximate inverse $\bar{F}^{\dagger}$ (equation 9) to the data. Since the image volume depends on v_0, so does J.

To examine the effect of imaging operator choice, we first plot the objective function values along a line segment in velocity model space. The model we use combines a constant background velocity model (2.5 km/s) and single flat reflector at 2 km depth. All three imaging operators are applied on the tapered Born data with a range of velocities : from 2 km/s to 3 km/s. The objective function values for different velocities are then plotted in Figure 2. All three objective functions are smooth in velocity and unimodal. However, only the objective function using the approximate inverse operator reaches the minimum at the correct velocity. Both the RTM and adjoint operator version shift toward lower velocity (RTM: 2200 m/s; Adjoint operator : 2400 m/s). Moreover, the objective function using approximate inverse better resolves the minimizer.
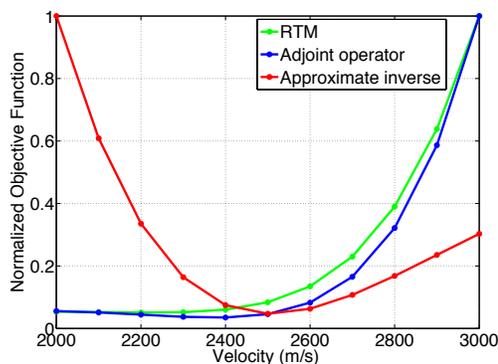
Figure 1: Normalized objective function with three different imaging operators. The velocity ranges from 2 km/s to 3km /s.

NUMERICAL EXAMPLES

The comparison is performed on a truncated Marmousi model (Versteeg and Grau, 1991). The true model, shown in 4a, is a smoothed version of the original model, in order to be wary of low frequency noise in the RTM image. The synthetic data is generated with 2-8 finite difference modeling. The acquisition geometry is a fixed spread of 151 sources and 301 receivers at 20 m depth. A (2.5-5-20-25)-Hz bandpass wavelet with 2ms time sample is used to generate 3s Born data. Optimization is carried out with L-BFGS algorithm (Nocedal and Wright, 1999). We

set the L-BFGS iteration to retain 20 prior search directions. We start the optimization from a highly smoothed initial model (Figure 4b).
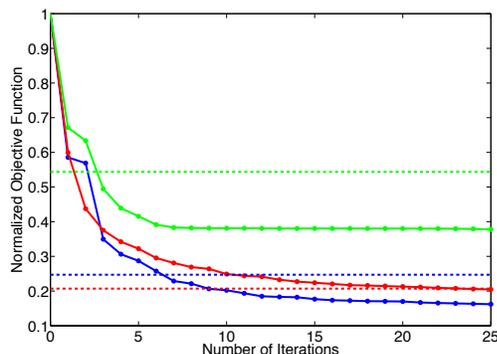
Figure 2: Normalized convergence curves for three different imaging operators. The solid line is the convergence curve and the dashed line is the objective function value at the correct velocity. The green line represents the RTM operator; the blue line represents the adjoint operator; the red line represents the approximate inverse operator.

25 iterations L-BFGS are used to generate the results. Figure 3 displays the normalized convergence curves for the optimization with different imaging operators. All three convergence curves mangage to go beyond the objective function value at the correct velocity, indicating that the objective function doesn't reach the minimum at the correct velocity value. Among them, the approximate inverse version gives the best objective function behavior in the sense that both the objective function value at correct model and the discrepancy between the converged value and objective function value at the correct model are the smallest. Figure 3 shows a comparison of the true, initial, updated models with three different imaging operators. As can be seen in the comparison, same optimization procedure with different imaging operators can produce quite different results. All five models are then used as background model to apply approximate inverse operator on the synthetic data (equation 9). Figure 4 compares the zero-offset images. All three recovered models significantly improve the quality of the images, compared to the clearly distorted initial image. Careful examination reveals that the result using approximate inverse operator is more correct: the location of the reflectors and the amplitude are closer to the image with true velocity. Finally, six offset gathers are pulled out from the middle of the image volume to compare in Figure 5, demonstrating approximate inverse operator produces a more focused image volume.
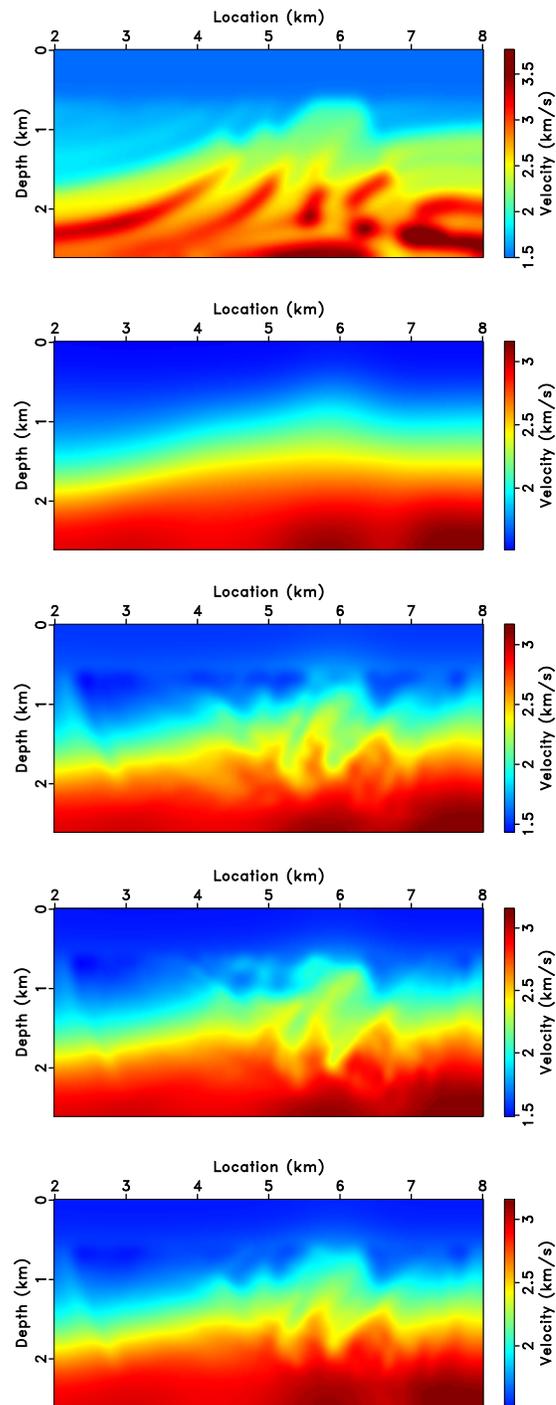
Figure 3: (a) True model; (b) Initial Model; 25 iteration L-BFGS result for (c) RTM operator; (d) the adjoint operator; (e) the approximate inverse operator.
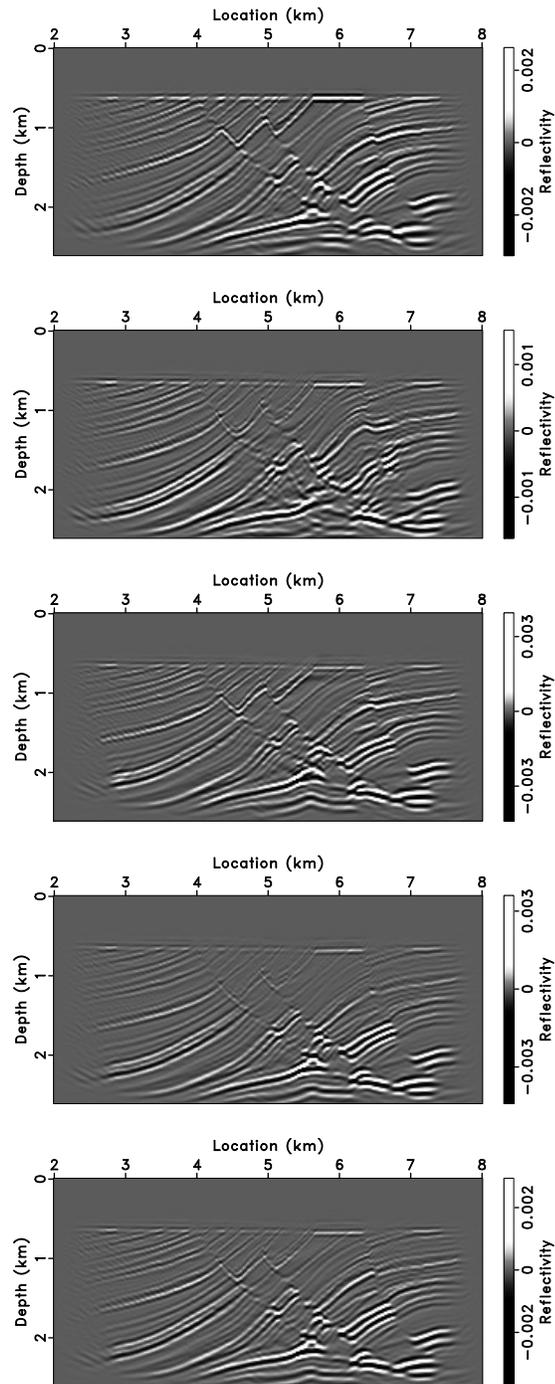
Figure 4: Zero-offset approximate inverse image associated with different velocity models shown in Figure 3.
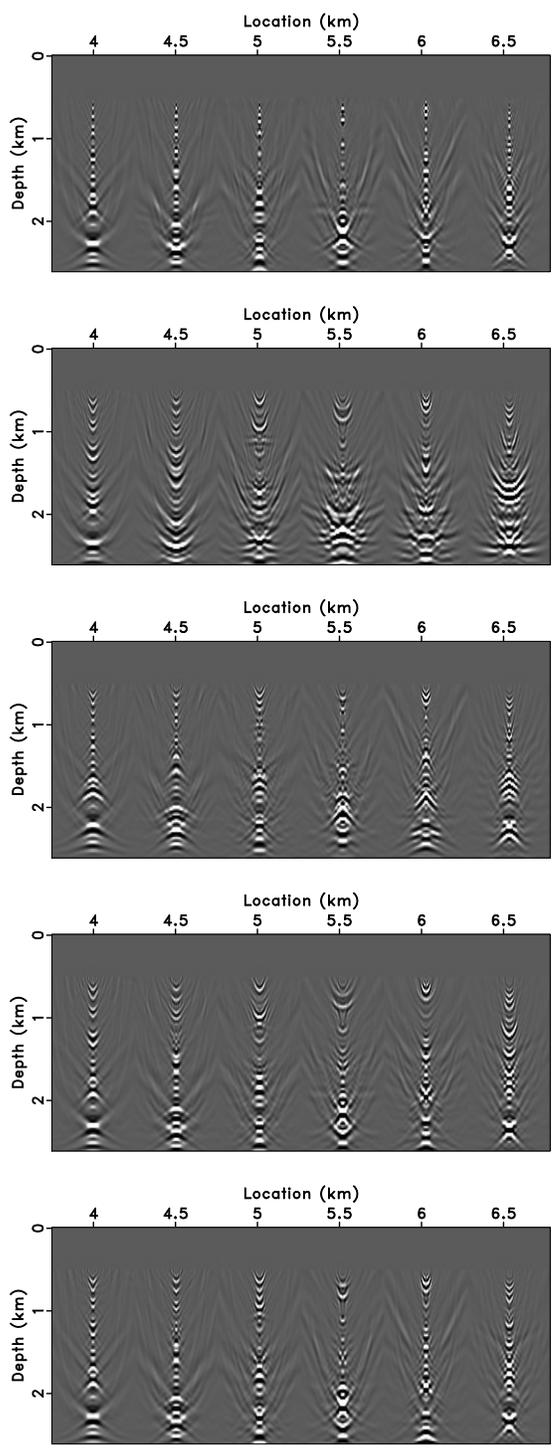
Figure 5: Image gathers in the subsurface offset domain corresponding to the images in Figure 4.

CONCLUSIONS

We have compared velocity analysis via DSO with three different imaging operators and analyzed their corresponding performance. The numerical examples show that the closer that the imaging operator to inversion, the better the DSO velocity estimate. An approximate inverse as in Hou and Symes (2015) adds no additional cost but improves velocity estimation substantially.

ACKNOWLEDGMENTS

REFERENCES

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, Expanded Abstracts, Society of Exploration Geophysicists, SI9.5.

Biondi, B., and P. Sava, 2004, Wave-equation migration velocity analysis - I: Theory, and II: Subsalt imaging examples: Geophysics, **52**, 593–623.

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

Hou, J., and W. Symes, 2014, An approximate inverse to the extended Born modeling operator: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 3784–3789.

———, 2015, An approximate inverse to the extended Born modeling operator: Geophysics, **80**, no. 6, R331–R349.

Khoury, A., W. Symes, P. Williamson, and P. Shen, 2006, DSR migration velocity analysis by differential semblance optimization: 76th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SPMI3.4.

Lafond, C. F., and A. R. Levander, 1993, Migration moveout analysis and depth focusing: Geophysics, **58**, 91–100.

Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations, *in* Conference on Inverse Scattering: Theory and Applications: Society for Industrial and Applied Mathematics, 206–220.

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.

Liu, Y., W. Symes, and Z. Li, 2014, Inversion velocity analysis via differential semblance optimization: Presented at the 76th Annual Meeting, Extended Abstracts, European Association of Geoscientists and Engineers.

Liu, Z., and N. Bleistein, 1995, Migration velocity analysis: theory and an interative algorithm: Geophysics, **60**, 142–153.

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

Shen, P., 2012, An RTM based automatic migration velocity analysis in image domain: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SVE2.2.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

Shen, P., W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2135–2139.

Stork, C., 1992, Reflection tomography in the postmigrated domain: Geophysics, **57**, 680–692.

Symes, W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

———, 2014, Seismic inverse problems: recent developments in theory and practice: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 2–5.

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

ten Kroode, F., 2012, A wave-equation-based Kirchhoff operator: Inverse Problems, 115013:1–28.

———, 2014, A Lie group associated to seismic velocity estimation: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 142–146.

Versteeg, R., and G. Grau, eds., 1991, The Marmousi experience: Proceedings of the EAEG workshop on practical aspects of inversion: IFP/Technip.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in explo-

ration geophysics: Geophysics, **74**, WCC127–WCC152.

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

Weibull, W., and B. Arntsen, 2013, Automatic velocity analysis with reverse time migration: Geophysics, **78**, S179–S192.

Yilmaz, O., 2001, Seismic data processing, *in* Investigations in Geophysics No. 10: Society of Exploration Geophysicists.

Yilmaz, O., and R. Chambers, 1984, Migration velocity analysis by wavefield extrapolation: Geophysics, **49**, 1664–1674.

Matched Source Waveform Inversion: Space-time Extension

Guanghui Huang and William W. Symes, Rice University, and Rami Nammour,
TOTAL E&P R&T USA

ABSTRACT

Matched Source Waveform Inversion (MSWI) uses non-physical (extended) synthetic energy sources to maintain waveform data fit throughout nonlinear inversion. A maximal variant of this approach extends the synthetic source to the entire space-time volume. Velocity is updated via a penalty on the spread of the synthetic source away from the physical source location: reducing this penalty by focusing the source at its correct location forces the velocity to achieve kinematic consistency with the data. Because data fit is maintained during this process, the stagnation typical nonlinear full waveform inversion is avoided, and convergence to an acceptable velocity is obtained without extremely low frequency data or particularly accurate starting models.

INTRODUCTION

Full waveform inversion (FWI) can provide detailed maps of subsurface medium parameters, such as compressional wave velocity and density, by iteratively updating these parameters to optimize the fit between simulated data and recorded seismic data in the least squares sense (Tarantola, 1984a,b). However, both synthetic and field data examples illustrate the tendency of FWI to stagnate at unsatisfactory solutions (often described as local minima) with large data misfit, unless the method is provided with an initial model that predicts arrival times to within a half-period in the lowest usable data frequencies (Bunks et al., 1995; Pratt, 1999; Virieux and Operto, 2009). Initial models and/or low frequency data of suitable accuracy are sometimes acquired, and in that case FWI imaging may yield significant improvements over other methods (Vigh et al., 2010; Plessix et al., 2010), but such information is not always available.

The approach to nonlinear waveform inversion reported in this paper belongs to a variant of FWI, in which additional parameters are added to the model so that data fit can be maintained throughout the update process. Such *model extensions* are non-physical, and must be suppressed during the course of the inversion by means of a penalty term added to the data misfit term of ordinary FWI. Many forms of model extension have been explored. The *space-time source extension* used in the work reported below adds to the model distributed synthetic energy sources, occupying the entire space-time volume used for wave modeling. These extended sources are adjusted, together with the velocity model, so that data is fit and while source energy is focused at the physical source positions. When both goals are achieved, the FWI problem is solved.

Various source extensions have been explored in prior work (Song and Symes, 1994; Symes, 1994; Plessix, 2000; Plessix et al., 2000; Pratt and Symes, 2002; Gao and Williamson, 2014; Warner and Guasch, 2014). Most of these works use the *source-receiver extension*, in which a (possibly) different source wavelet (at the physical source location) is used to model each data trace, and a penalty is imposed for trace dependence of the sources. Huang and Symes (2015a) apply this concept to crosswell tomography, and show that with suitable choice of penalty, provided that no caustics (or multiple energy paths) occur between source and receiver, the objective of this extended version of FWI has the same stationary points as does the traveltime tomography objective, and in particular is convex over a much wider domain in model space than is the FWI objective. As will be explained below, inversion based on the space-time source extension appears to have the same tomographic property but without the requirement that the wavefield be free of caustics.

We mention that another genre of model extension, involving addition of non-physical axes to velocity and other mechanical parameters of the subsurface, has also been explored extensively (Shen and Symes, 2008; Biondi and Almomin, 2012, 2014; Lameloise et al., 2015), also (Symes, 2008) for a survey of older work. These mechanical parameter extensions may achieve roughly the same tomographic goals as do source extensions of the type discussed here, under various conditions.

THEORY

We model seismic wave propagation using the constant density acoustic wave equation with isotropic point source. The pressure field $u(\mathbf{x}, t; \mathbf{x}_s)$ for the source position $\mathbf{x} = \mathbf{x}_s$ satisfies

$$\left(\frac{1}{v^2} \frac{\partial^2 u}{\partial t^2} - \Delta u \right)(\mathbf{x}, t; \mathbf{x}_s) = f(t)\delta(\mathbf{x} - \mathbf{x}_s) \tag{1}$$

$$u|_{t=0} = \left. \frac{\partial u}{\partial t} \right|_{t=0} = 0 \tag{2}$$

The forward operator $S[v]f$ which relates the velocity $v(\mathbf{x})$ and wavelet function $f(t)$ to the scattering field at the receiver $\mathbf{x}_r$, i.e.,

$$S[v]f(\mathbf{x}_r, t; \mathbf{x}_s) = u(\mathbf{x}_r, t; \mathbf{x}_s). \tag{3}$$

Conventional full waveform inversion based on this model consists in finding the pair (v, f) such that the residual wavefield is minimized in the least squares sense:

$$J_{\mathrm{FWI}}[v, f] = \frac{1}{2} \sum_{\mathbf{x}_s, \mathbf{x}_r} \int |S[v]f(\mathbf{x}_r, t; \mathbf{x}_s) - d(\mathbf{x}_r, t; \mathbf{x}_s)|^2 dt.$$

where $d(\mathbf{x}_r, t; \mathbf{x}_s)$ is recorded data.

Extended Model and Annihilator

The extended seismic wavefield $\bar{u}$ corresponding to the extended source model $\bar{f}$ and velocity v satisfies a modification of equation 1:

$$\frac{1}{v^2} \frac{\partial^2 \bar{u}}{\partial t^2} - \Delta \bar{u} = \bar{f}(\mathbf{x}, t, \mathbf{x}_s) \tag{4}$$

$$\bar{u}|_{t=0} = \left. \frac{\partial \bar{u}}{\partial t} \right|_{t=0} = 0 \tag{5}$$

The extended forward modeling operator is defined by

$$\bar{S}[v]\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s) = \bar{u}(\mathbf{x}_r, t; \mathbf{x}_s). \tag{6}$$

Note that the dimensional of data space is smaller than that of the extended source function space, so it is hardly surprising that we can easily fit the data, that is,

solve $d = \bar{S}[v]\bar{f}$ by adjusting $\bar{f}$ for more or less arbitrary v. That is, we now have too many solutions of the inverse problem. To eliminate the nonuniqueness, note that the extended model should be reduced to $\bar{f}(\mathbf{x}, t, x_s) = \delta(\mathbf{x} - \mathbf{x}_s)f(t)$ if given the true velocity, which motivates us to introduce an *annihilator A* to penalize nonphysical source energy away from $\mathbf{x} = \mathbf{x}_s$. Many such annihilators exist. In this work we use

$$A\bar{f}(\mathbf{x}, t; \mathbf{x}_s) = |\mathbf{x} - \mathbf{x}_s|\bar{f}(\mathbf{x}, t; \mathbf{x}_s). \tag{7}$$

Matched Source Waveform Inversion (MSWI)

We combine a data fitting term using the the extended forward map 5 with the mean-square of the annihilator output 2 to produce the matched source waveform inversion (MSWI) objective function,

$$
\begin{aligned}
J_\alpha[v, \bar{f}] \;=\; & \frac{1}{2\alpha} \sum_{\mathbf{x}_r, \mathbf{x}_s} \int |\bar{S}[v]\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s) - d(\mathbf{x}_r, t; \mathbf{x}_s)|^2 dt \\
& + \frac{1}{2} \sum_{\mathbf{x}, \mathbf{x}_s} \int |A\bar{f}(\mathbf{x}, t; \mathbf{x}_s)|^2 dt.
\end{aligned} \tag{8}
$$

Note that the MSWI objective functional is quadratic with respect to extended source function $\bar{f}$, though it is still nonlinear in velocity v. It is therefore natural to adopt a nested approach: solve first for $\bar{f}$ - a linear, hence presumably easier problem - to create a reduced objective depending only on v, then minimize the reduced objective over v. Nested optimization of partially linear objectives was introduced under the name *Variable Projection Method* (VPM) by Golub and Pereyra (1973, 2003) and used to estimate sources and parameter perturbations in waveform inversion (van Leeuwen and Mulder, 2009; Rickett, 2012; Li et al., 2013). In fact, VPM and similar nested algorithms were used in extended waveform inversion from the beginning (Kern and Symes, 1994), and for essential reasons, not merely as a computational convenience (Huang and Symes, 2015b).

Algorithm 1. Vpm for MSWI Method
1. Define the v-dependent extended source $\bar{f}[v]$, by minimizing the objective 8 over $\bar{f}$: this amounts to solving the normal system

$$(\bar{S}[v]^T \bar{S}[v] + \alpha A^T A)\bar{f}[v] = \bar{S}[v]^T d. \tag{9}$$

2. Minimize over v the reduced objective function,

$$J_\alpha^{red}[v] = \frac{1}{2\alpha} \sum_{\mathbf{x}_r,\mathbf{x}_s} \int |\bar{S}[v]\bar{f}[v] - d|^2 dt + \frac{1}{2} \sum_{\mathbf{x},\mathbf{x}_s} \int |A\bar{f}[v]|^2 dt$$
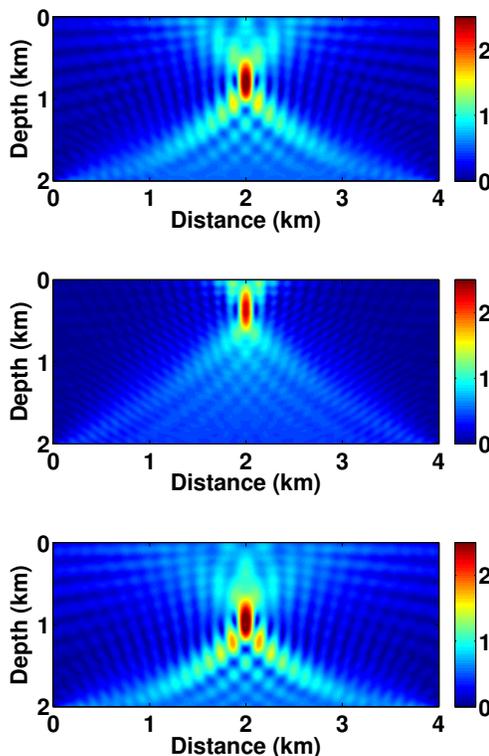
Figure 1: Amplitude of the backpropagation wavefield $\bar{S}^T d$ for (a) true velocity, (b) 10% low velocity, and (c) 10% high velocity

To illustate the mechanics of this algorithm, we use a time-harmonic transmission inversion problem for which the correct velocity is the constant $v = 2$ km/s. Put the single source at $\mathbf{x}_s = (2, 0.6)$ km, and record data at $z_r = 2$ km from $x_r = 0$ km to $x_r = 4$ km. The input source frequency is 10 Hz. We compare the amplitude of backpropagation $\bar{S}[v]^T d$ of the recorded data, i.e. the right hand side of (9), for true velocity and 10% low and high velocity as showed in Figure 1a-1c. In fact, this field is proportional to the first iteration of any gradient-descent method for solving 9, so can serve as a proxy for the solution. Only for the correct velocity is this wavefield focused at the source position, while for other velocities it is not.

Zhang and Gao (2008), Zhang and Wang (2009), and Jin and Plessix (2013) base algorithms for velocity updating on this principle, measuring energy spread via an operator similar to our A. Note that the actual solution $\bar{f}$ of the least-squares problem 9 is much better focused, just as least-squares migration generally produces a better focused image than RTM, and therefore is a more sound basis on which to base a velocity update.

NUMERICAL EXAMPLES

We adopt the frequency continuation approach to solve the MSWI problem optimized by steepest descent gradient method with backtracking line search. For each frequency, a model update is constructed with a fixed number of steepest descent gradient iterations. The updated model is taken as the initial model for inversion at the next frequency. During each iteration, we use a direct matrix solve (Gaussian elimination) to solve the normal equation (9), which requires forming the matrix of the normal operator, to guarantee the accuracy of the gradient.
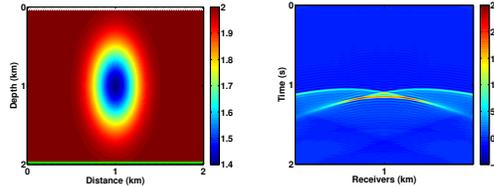
Figure 2: Transmission configuration: (a) target velocity; (b) shot gather at $x_s = 1$km
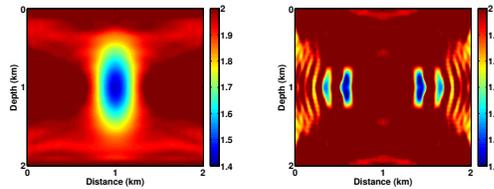
Figure 3: Inverted models after 25 iterations of (a) MSWI and (b) FWI.

Example 1 (Transmission). *The model (Figure 2a) contains a low velocity anomaly embedded in a constant velocity background.* 197 *receivers are placed at depth $z_r = 1.99$ km from* 0.02 *km to* 1.98 *km, spaced* 0.01 *km apart.* 50 *shots are placed at depth $z_s = 0.01$ km from* 0.02 *km to* 1.98 *km, shot interval* 0.04 *km. The frequencies used in inversion are $f = 6, 10, 14, 18$ Hz. The regularization parameter $\alpha = 10^{-6}$.*

Figures 2a-2b show the true velocity and recorded data at the center shot, respectively. The inverted models after 25 iterations for each frequency by MSWI and FWI method are shown in Figure 3a-3b, respectively. Due to the absence of the low velocity anomaly, the initial velocity produces a big arrival time error in the predicted data, which causes FWI to fail, while the main low velocity feature is inverted by MSWI to high precision.
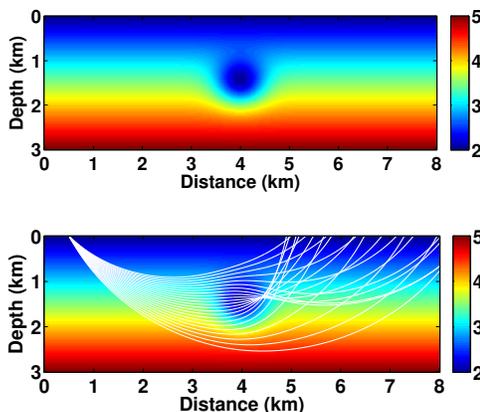
Figure 4: (a) True velocity; (b) ray trace at the first shot position $x_s = 0.05$km

Example 2 (Diving Wave). *The model consists of a low velocity Gaussian anomaly embedded in linear increasing background velocity.* 200 *receivers are placed at depth* $z_r = 0.04$ *km from* $x_r 0.02$ *km to* 7.98 *km with* $dx_r = 0.02$ *km.* 80 *shots are placed at depth* $z_s = 0.02$ *km from* $x_s = 0.05$ *km to* 7.95 *km with* $dx_s = 0.1$ *km. Successively inversion frequencies are* $f = 6, 7, 8, 9, 10$ *Hz. The regularization parameter* $\alpha = 10^{-3}$.

The true velocity is shown in Figure 4a. We perform the ray trace on this model as depicted in Figure 4b. As we can see, there are three types of waves arriving at receivers including direct wave, diving waves, and later arrivals due to the low velocity anomaly. We start the inversion with the correct background (linear) velocity. The initial velocity and inverted velocity are displayed in Figure 5a-5c after 25 iterations for each frequency. The low velocity anomaly is well reconstructed (Figure 4a).

Example 3 (Marmousi Model). *The Marmousi model (Bourgeois et al., 1991) is modified by adding* 0.2 *km water on the top. Thus the model is* 9.2 *km in horizontal extent* 3.2 *km deep. We generate* 114 *shots starting at* $x_s = 0.08$ *km to* 9.12 *km with* $dx_s = 0.08$ *km at depth* $z_s = 0.16$km*. The receivers are placed at depth* $z_r = 0.04$ *km from* $x_r = 0.08$
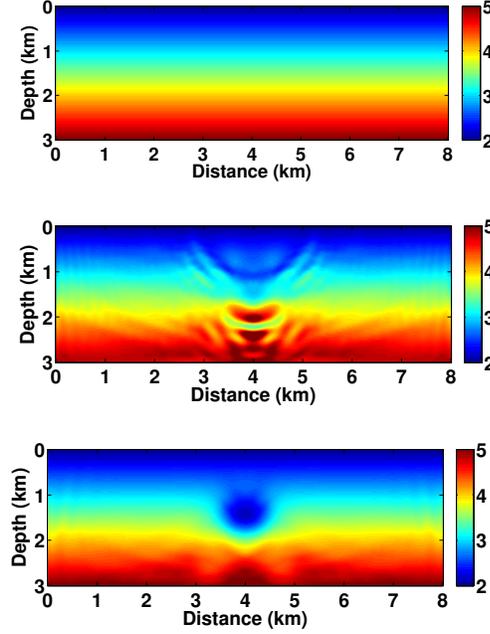
Figure 5: (a) Initial velocity; inverted velocity by (b) FWI method and (c) MSWI using (6-10) Hz data

km to 9.12 *km with* $dx_r = 0.04$ *km. Thus each shot generates* 227 *traces.* 5 *frequencies are used in the inversion:* 4, 5, 6, 7, *and* 8 *Hz. The regularization parameter* $\alpha = 10^{-3}$.

To avoid the obvious "inverse crime", we generate the synthesized data using a finer mesh than is used in the inversion. The target velocity model, 1D linear increasing initial velocity and the inverted velocity after 50 iterations for each frequency are shown in Figure 6a-6c from top to bottom. The inverted and target velocities are quite close, allowing for the limited resolution available to the inversion due to limited frequency range.

DISCUSSION AND CONCLUSION

MSWI is one kind of nonlinear waveform inversion that imposes few constraints on the model or data geometry, hence can succeed with many kinds of waveform data, including transmission, reflection and diving/refraction, without any pre-processing. The source focus annihilator (*A* in equation 2) forces the extended source to focus at survey source positions, and leads to a velocity update not de-
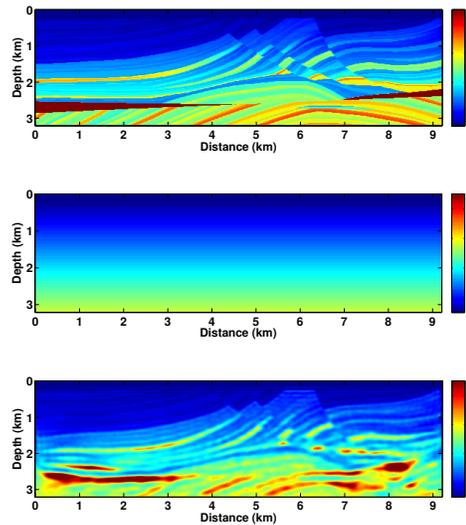
Figure 6: (a) Marmousi model; (b) 1D initial velocity; (c) inverted velocity after 50 MSWI iterations at each frequency.

pending on prior knowledge of the source, an estimate for which is a by-product of the inversion. The data is fit throughout by appropriate choice of the weight α, allowing the method to avoid the stagnation at poorly-fitting models that appears to afflict FWI.

MSWI with space-time extension, as defined here, is closely related to Waveform Reconstruction Inversion (WRI) (van Leeuwen and Herrmann, 2013). WRI updates the velocity model via minimization of the minimum mean square error by which a trial wavefield u fails to solve the wave equation 1, penalized by a multiple of its mean square data misfit. The residual solution error field is exactly the extended source $\bar{f}$ of the present paper, which corresponds 1-1 with the trial wavefield u via the uniqueness theorem for the inhomogeneous wave equation. Besides focusing on the right-hand side rather than the solution, the approach explained here employs a different annihilator than do van Leeuwen and Herrmann (2013).

The implementation used in the examples presented here was formulated in the 2D frequency domain. Direct extension to 3D would require solution of the 3D Helmholtz equation, currently a limiting factor. Alternatively, 3D time domain solution is certainly possible in principle, but requires storage of the full space-time source wavefield $\bar{f}$. Extension to 3D therefore requires either very large floating point or very large storage requirements, a limitation it shares with

several other alternatives to FWI. On the other hand, extension to more general physics, for example various flavors of elastodynamics, is in principle straightforward.

ACKNOWLEDGEMENT

REFERENCES

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, Expanded Abstracts, Society of Exploration Geophysicists, SI9.5.

——, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Bourgeois, A., M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg, 1991, Marmousi, model and data, *in* The Marmousi Experience: Proc. of the 1990 EAEG workshop on practical aspects of seismic data inversion, 5–16.

Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Gao, F., and P. Williamson, 2014, A new objective function for full waveform inversion: differential semblance in the data domain: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1178–1182.

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.

——, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse Problems, **19**, R1–R26.

Huang, G., and W. Symes, 2015a, Full waveform inversion via matched source extension: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1320–1325.

Huang, Y., and W. Symes, 2015b, Born waveform inversion via variable projection and shot record model extension: 85rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1326–1331.

Jin, S., and R.-E. Plessix, 2013, Stack-based full wavefield velocity tomography: 83rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1095–1099.

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.

Li, M., J. Rickett, and A. Abubakar, 2013, Application of the variable projection scheme to frequency-domain full-waveform inversion: Geophysics, **78**, R249–R257.

Plessix, R.-E., 2000, Automatic cross-well tomography: an application of the differential semblance optimization to two real examples: Geophysical Prospecting, **48**, 937–951.

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, Expanded Abstracts, Society of Exploration Geophysicists, 930–934.

Plessix, R.-E., W. Mulder, and F. ten Kroode, 2000, Automatic cross-well tomography by semblance and differential semblance optimization: theory and gradient computation: Geophysical Prospecting, **48**, 913–935.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory, and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R., and W. Symes, 2002, Semblance and differential semblance optimization for waveform tomography: a frequency domain approach: Journal of Conference Abstracts, **7**, 183–4.

Rickett, J., 2012, The variable projection method for waveform inversion with an unknown source function: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1–5.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

Song, H., and W. Symes, 1994, Inversion of crosswell tomography via differential semblance optimization: Technical report, The Rice Inversion Project. (Annual Report).

Symes, W., 1994, Inverting waveforms in the presence of caustics: Technical report, The Rice Inversion Project. (Annual Report).

———, 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

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

———, 1984b, Linearized inversion of seismic reflection data: Geophysical Prospecting, **32**, 998–1015.

van Leeuwen, T., and F. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

van Leeuwen, T., and W. Mulder, 2009, A variable projection method for waveform inversion: 71st Annual Meeting, Extended Abstracts, European Association of Geoscientists and Engineers, U024.

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, Expanded Abstracts, Society of Exploration Geophysicists, 957–961.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1089–1093.

Zhang, Y., and F. Gao, 2008, Full waveform inversion based on reverse time propagation: 78th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1950–1955.

Zhang, Y., and D. Wang, 2009, Traveltime information-based wave-equation inversion: Geophysics, **74**, WCC27–WCC36.

Matched Source Waveform Inversion: Volume Extension

Guanghui Huang and William W. Symes, Rice University, Rami Nammour, TOTAL E&P R&T USA

ABSTRACT

Matched source waveform inversion via volume extension introduces additional, non-physical sources acting at time $t = 0$. The extra sources permit the data to be fit well, even when velocity is kinematically inconsistent with data. Assuming point or near-point sources at known locations, a distance-weighted penalty on source energy is minimized when the data is fit and the sources act only at the known locations, thus solving the FWI problem. Good data fit throughout the inversion process produces a considerably more convex objective function than does the standard data-domain FWI formulation. A good theoretical foundation for this approach exists for pure transmission problems, while numerical examples suggest that in fact it may be a feasible approach to reflection and transmission inversion (separately or together) when the data lack adequate S/N at sufficiently low frequencies for conventional FWI to succeed.

INTRODUCTION

Full waveform inversion (FWI) obtains detailed maps of material parameters interior to the earth by matching simulated to recorded data in the least squares sense (Gauthier et al., 1986; Tarantola, 1984). This highly nonlinear and ill-posed inverse problem is also computationally large, forcing reliance on iterative optimization methods which may stagnate at geologically uninformative solutions (Pratt, 1999; Virieux and Operto, 2009). Many alternative methods have been suggested that may circumvent FWI's requirement of a starting model that predicts traveltimes to within one-half wavelength for frequencies with good S/N. Some recent examples include (Shen and Symes, 2008; Weibull and Arntsen, 2013; Biondi and Almomin, 2014; Warner and Guasch, 2014; van Leeuwen and

Herrmann, 2013). See Symes (2008) for an overview of older literature on this topic.

All of the works just cited expand the convergence domain of FWI by adding parameters to the earth model. The approach presented in this paper belongs to this model extension genre as well: it is a *matched source* waveform inversion method. The physical model consists of coefficients in the wave equation (mechanical paramters) and a source mechanism, represented as a right-hand side in the wave equation. We assume that the physical source is localized, essentially point-like, and adjoin a non-physical auxiliary source field that acts at $t = 0$ (an exploding reflector model, essentially). The role of the auxiliary source field (the additional parameters in this *volume extension* of the acoustic model) is to enable fit to data, for all velocity models, hence the name "matched source". We penalize the auxiliary source field by its mean square, weighted by distance to the source: minimizing this penalty forces the auxiliary source to zero, and recovers the physical source and thus the solution to the inverse problem, since data fit has been maintained throughout the process.

The matched source approach was introduced by Symes (1994): the additional parameters in that work were receiver dependent source wavelets, acting at the physical source positions. This *source-receiver extension* has been used in a variety of contexts (Plessix, 2000; Pratt and Symes, 2002; Luo and Sava, 2011; Warner and Guasch, 2014). For combinations of acquisition geometry and velocity that generate unique ray paths between source and receiver, the objective function is equivalent to the traveltime misfit, which explains the large domain of convergence and essentially tomographic nature of the inversion (Huang and Symes, 2015). However, for more refractive models generating multiple ray paths, the domain of convergence shrinks to resemble that of FWI (Symes, 1994). The advantage of the volume extension presented here is the persistence of the large domain of convergence even for highly refractive models generating multiple ray paths, as will be illustrated below. The volume extension appears to share this advantage with the *space-time extension*, essentially equivalent to Wavefield Reconstruction Inversion (van Leeuwen and Herrmann, 2013). However the volume extension requires only one time level of the field to be stored, as opposed to the full space-time (or space-frequency) volume needed (in principle) in WRI.

We describe the volume extension and the matched source inversion based on it in the next section, along with a sketch of its theoretical justification for transmission inversion (direct or diving waves). Then we present several examples, of both transmission and reflection inversion.

THEORY

We assume that the modeled data $d(\mathbf{x}_r, \mathbf{x}_s, t)$ is the restriction to known survey source and receiver locations $\mathbf{x}_s, \mathbf{x}_r$, of a solution u to the constant density acoustic wave equation,

$$\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} - \Delta u = \delta(\mathbf{x} - \mathbf{x}_s)f(t) \tag{1}$$

$$u|_{t=0} = \left.\frac{\partial u}{\partial t}\right|_{t=0} = 0 \tag{2}$$

where v is the unknown seismic velocity, and $f(t)$ is the input source wavelet.

We introduce forward modeling operator $S[v]f$ to relate the velocity v and source function $f(t)$ to the data as just described. Full Waveform Inversion (FWI) adjusts the velocity to minimize the difference between the modeled data and recorded data in the least squares sense, that is,

$$J_{\text{FWI}} = \frac{1}{2}\sum_{\mathbf{x}_s, \mathbf{x}_r}\int |S[v]f(\mathbf{x}_r, \mathbf{x}_s, t) - d(\mathbf{x}_r, \mathbf{x}_s, t)|^2 dt$$

Gradient-based optimization methods applied to this problem tend to stagnate far from a useful model estimate if the initial model is does not predict travel times of important events in the data to within a half-wavelength at frequencies with good S/N.

Extended modeling and volume extension

The volume extension is defined by solving the extended acoustic wave equation,

$$\frac{1}{v^2}\frac{\partial^2 \bar{u}}{\partial t^2} - \Delta \bar{u} = f(t)\delta(\mathbf{x} - \mathbf{x}_s) + \bar{f}(\mathbf{x} - \mathbf{x}_s)\delta(t) \tag{3}$$

$$u = 0, t < 0 \tag{4}$$

Here $f(t)$ is the physical point source wavelet, the pair $(f(t), \bar{f}(\mathbf{x}, \mathbf{x}_s))$ is the extended source, $\bar{u}(\mathbf{x}, t; \mathbf{x}_s)$ is the acoustic potential field, and $v(\mathbf{x})$ is the velocity field. The extended forward modeling operator is defined by sampling the pressure field,

$$\bar{S}[v](f, \bar{f})(\mathbf{x}_r, t; \mathbf{x}_s) = \frac{\partial}{\partial t}\bar{u}(\mathbf{x}_r, t; \mathbf{x}_s). \tag{5}$$

If the data d is fit, that is, $\bar{S}[v](f,\bar{f}) = d$, and $\bar{f}$ vanishes, so that the total source focuses at the source position $\mathbf{x}_s$, then v solves the FWI problem. We introduce a penalty or annihilator operator $A(f,\bar{f})(\mathbf{x},\mathbf{x}_s) = |\mathbf{x} - \mathbf{x}_s|\bar{f}(\mathbf{x},\mathbf{x}_s)$ for which $A(f,\bar{f}) = 0$ precisely when $\bar{f} = 0$. Then the solution of the FWI problem also solves

$$\text{minimize } \|A(f,\bar{f})\|^2 \text{ subject to } \bar{S}[v](f,\bar{f}) \simeq d$$

As unconstrained optimization problems are easier to attack, we introduce a penalty form of this problem (the variable projection method (Golub and Pereyra, 2003)): we minimize over v

$$J_\alpha[v] = \frac{1}{2} \sum_{\mathbf{x}_r,\mathbf{x}_s} \left(\int |\bar{S}[v](f,\bar{f}) - d|^2 + \alpha \int |A(f,\bar{f})|^2 \right) \tag{6}$$

$$\text{s.t. } N_\alpha[v](f,\bar{f}) \equiv (\bar{S}[v]^T \bar{S}[v] + \alpha A^T A)(f,\bar{f}) = \bar{S}[v]^T d \tag{7}$$

for a penalty parameter $\alpha > 0$ to be determined.

Transmission case

In this section we summarize several important properties of the matched source waveform inversion problem 6 for slowly varying velocity on the wavelength scale. In the asymptotic limit, this simply means that the velocity is smooth as a function of spatial position, so all arrivals are transmitted, i.e. either direct or diving waves, and no reflections exist in the data. This case applies for example to crosswell geometry. We will make a number of assertions about S and related operators, mathematical justifications of which are technical and will be given elsewhere.

We idealize the receiver arrays to continuous subsets of a receiver surface embedded in space (for example, $z = z_r$), and assume that the extended sources $\bar{f}$ are constrained to produce no energy propagation (at least asymptotically) along rays that graze the receiver array, i.e. are tangent to it at some point. Assuming this grazing ray constraint, the normal operator $N_\alpha[v]$ defined in equation 7 is continuous in its argument - that is, a small mean-square change in $(f,\bar{f})$ results in a small mean-square change in $N_\alpha[v](f,\bar{f})$. Also, $N_\alpha[v]$ is regular as a function of v as well.

$N_\alpha[v]$ may be rendered invertible by adding a small multiple of the identity operator (Tihonov regularization). We presume that this has been done, and treat $N_\alpha[v]$ as invertible. Then it is also possible to show that $\bar{S}[v]N_\alpha[v]^{-1}\bar{S}[v]^T$ is smooth in v. Since $J[v]$ may be re-written as

$$J_\alpha[v] = \frac{1}{2\alpha}\langle(I - \bar{S}[v]N_\alpha[v]^{-1}\bar{S}[v]^T)d,d\rangle,$$

$J_\alpha[v]$ is also smooth in v.

Computationally, the gradient of J_α can be obtained via the adjoint state method (Plessix, 2006).

Most importantly, N_α is approximately local, in the sense that a high-frequency pulse input, localized near $\mathbf{x}$ in space and near $\mathbf{k}$ in phase space, will produce a high-frequency pulse localized in the same way. That is, N_α does not move events in the extended source $\bar{f}$. For that reason, it is invertible, up to an error of lower order in spatial frequency: it replaces each localized pulse by a scaled version of itself, to good approximation, with the scaling being positive it the ray generated by the position $\mathbf{x}$ and momentum $\mathbf{k}$ intersects the acquisition geometry. That is, the part of the extended source that generates waves propagating within the experimental aperture is recovered by $N_\alpha[v]^{-1}$, apart from a position- and dip-dependent scaling.

This within-aperture invertibility differentiates the volume extension from the source-receiver extension described by Huang and Symes (2015), which has a similar invertibility property if the velocity model is sufficiently close to homogenous that it does not generate multiple ray paths connecting sources and receivers, but loses invertibility for more complex models.

NUMERICAL EXAMPLES

Example 4 (Crosswell transmission tomography)**.** *The model consists of a low velocity Gaussian anomaly embedded in a constant background velocity $v_0 = 2$ km/s. 99 receivers are placed at depth $z_r = 1.98$ km from 0.02 km to 1.98 km with $dx_r = 0.02$ km for all of 24 shots at depth $z_s = 0.02$ km from 0.08 km to 1.92 km with $dx_s = 0.08$ km to imitate transmission problem. In both volume-based MSWI and FWI we invert the data simultaneously with frequencies from $f = 9$ Hz to $f = 20$ Hz.*
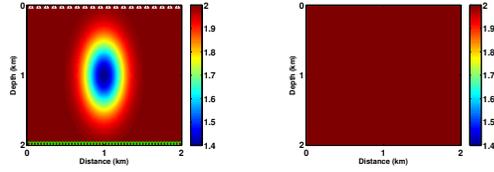
Figure 1: Transmission configuration: (a) true velocity; (b) initial constant velocity
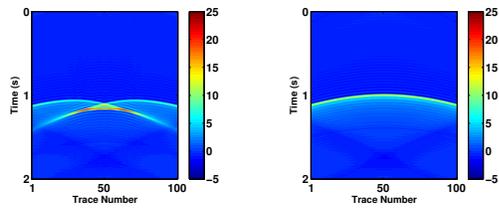
Figure 2: Shot gather at $x_s = 1$km for (a) true velocity and (b) initial constant velocity
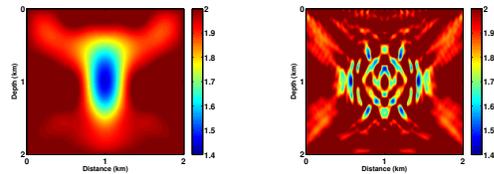
Figure 3: Inverted velocity after 50th iterations by (a) volume-based MSWI method and (b) FWI method

Figure 1a-1b shows the target velocity and the initial velocity, respectively. The recorded data and initial data are ploted in Figure 2a-2b, which show that there is significant multipathing present in the data and a large traveltime difference between recorded data and initial data. The inverted results by MSWI and FWI after 50 LBFGS iterations are shown in Figure 3a-3b, respectively. As expected, FWI fails due to the large traveltime mismatch in the initial model, whereas MSWI recovers a reasonable velocity estimate.

We note that the MSWI based on source-receiver extension (Huang and Symes, 2015) also fails to invert this example, due the multipathing and consequent loss of normal operator invertibility for that extension.

Example 5 (Reflected wave tomography). *The model consists of three layers, with dense sampling of sources and receivers distributed across the surface, Figure 4a. The initial velocity model is $v_0 = 1.5$ km/s.*

MSWI and FWI results are shown in Figure 4b-4c, respectively. FWI method can only reconstuct the first interface position correctly. The volume-based MSWI method produces a rather decent resconstruction with correct velocity between the layers and the right second interface position.
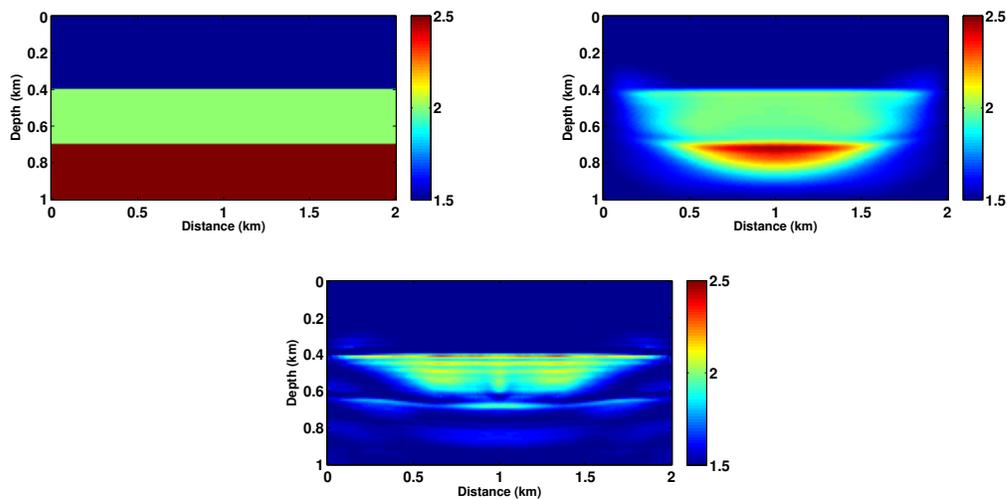
Figure 4: Reflected wave tomography: (a) true velocity , inverted velocity by our method with 5-20Hz data after 200 iterations, and inverted velocity by FWI method with 5-20Hz data after 200 iterations

Example 6 (Salt body reconstruction). *The model (portion of Pluto model (Stoughton et al., 2001)) consists of a high velocity "salt" inclusion in a layered medium. The initial model is a 1D linearly increasing velocity with depth. The frequency band used for inversion is from 4 to 10 Hz.*

The target velocity and initial velocity are showed in Figure 22a-22b, respectively. The inverted velocity by FWI after 200 iterations is ploted in Figure 23. It mischaracterized the structure badly. MSWI method after 100 iterations recovers the structure reasonably accurately (Figure 5d).

DISCUSSION AND CONCLUSION

We have described matched source waveform inversion based on a volume extension, which allows close data fitting throughout the inversion process, and achieve a good reconstruction of the velocity in cases where FWI fails due to lack of low frequency data. We are able to offer a theoretical explanation for the behaviour of the algorithm in the transmission case. Numerical examples suggests that volume extension MSWI is effective in both transmission and reflection modes.

Current implementation of the inner linear subproblem 7 uses direct matrix methods, and forms the the dense matrix explicitly. This approach is feasible in 2D but problem 7 will need to be solved iteratively in 3D. Both preconditioning the inner solve, and computing the gradient accurately with inaccurate inner solves, are important topics for future research. Additionally, for reflected wave tomography, the outer nonlinear iteration converges quite slowly, which may be caused by the problem's extreme ill-posedness.

ACKNOWLEDGEMENT
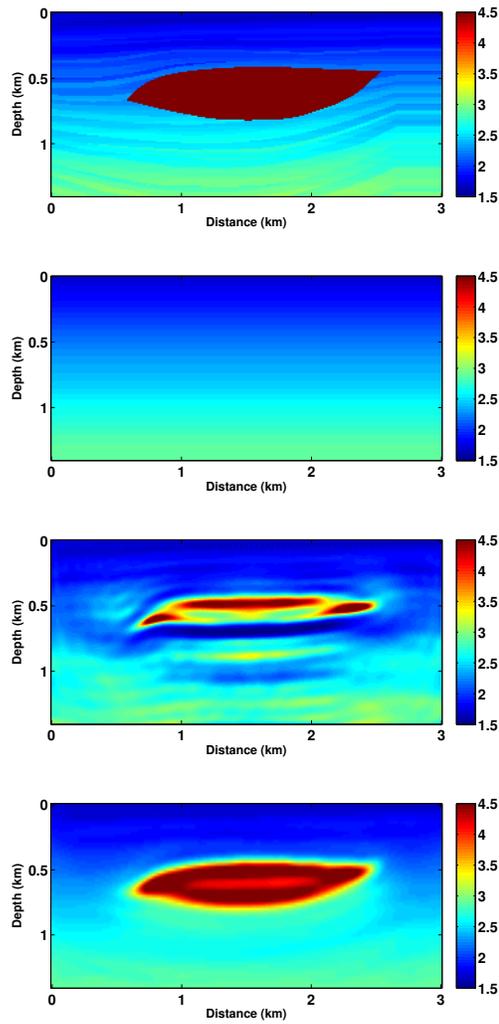
Figure 5: Salt body reconstruction: (a) true velocity; (b) initial 1D velocity; inverted velocity by (c) FWI after 200 iterations, and (b) volume-based MSWI method after 100 iterations

REFERENCES

Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

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

Golub, G., and V. Pereyra, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse Problems, **19**, R1–R26.

Huang, G., and W. Symes, 2015, Full waveform inversion via matched source extension: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1320–1325.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2788–2792.

Plessix, R.-E., 2000, Automatic cross-well tomography: an application of the differential semblance optimization to two real examples: Geophysical Prospecting, **48**, 937–951.

———, 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory, and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R., and W. Symes, 2002, Semblance and differential semblance optimization for waveform tomography: a frequency domain approach: Journal of Conference Abstracts, **7**, 183–4.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

Stoughton, D., J. Stefani, and S. Michell, 2001, 2-D elastic model for wavefield investigations of subsalt objectives, deep water Gulf of Mexico: 71th Annual International Meetting, SEG, Expanded Abstacts, 1269–1272.

Symes, W., 1994, Inverting waveforms in the presence of caustics: Technical report, The Rice Inversion Project. (Annual Report).

———, 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

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

van Leeuwen, T., and F. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1089–1093.

Weibull, W., and B. Arntsen, 2013, Automatic velocity analysis with reverse time migration: Geophysics, **78**, S179–S192.

Mathematical Fidelity and Open Source Libraries for Large Scale Simulation and Optimization

William W. Symes

ABSTRACT

Optimization algorithms used to solve inverse problems in geoscience have abstract mathematical descriptions many of them (Conjugate Gradient iteration, Newtons method,) are socalled matrix free algorithms, that is, they manipulate their mathematical objects (vectors, functions) without reference to their internal details. Similarly, time-stepping algorithms for dynamical simulation may be described in terms of update rules for dynamical states, without reference to the internal structure of these states or the precise action of the rules. Both of these algorithmic settings provide opportunities for creation of re-usable open source code bases, applying to many different tasks. Not only do such libraries save programmer effort and reduce the incidence of errors, but also they could potentially make possible comparison of inversion techniques by providing common implementations for common components. This paper lays out some examples of features that computational types should inherit from their mathematical models, and some solutions to the programming problems that arise in implementing such types. [To be presented in Workshop 8, EAGE, Vienna, June 2016]

INTRODUCTION

One of the more celebrated papers in the theory of turbulence begins a discussion of fluid dynamics with the statement: "The time evolution of a velocity field is given by the Navier-Stokes equations: $\frac{dv}{dt} = X_\mu(v)$" (Ruelle and Takens, 1971). While this description of viscous incompressible fluid flow might seem a bit terse, in fact it was just the right setting for a step change in the understanding of turbulence.

Analogously, the structure of scientific software for dynamical simulation and optimization can also be stripped down to mathematical basics, and the fundamental components reflected accurately in software, with beneficial effect on scope and flexibility. As the concepts involved are mathematical, rather than computational, computer languages unsurprisingly do not provide them, so new types, defined by their behaviour, rather than by their implementation, must be supplied.

The sections to follow will describe some basic types required for expressing algorithms in simulation and optimization, and points out that their computational realizations should possess just those attributes necessary to define their behaviour and relations, and no more. Differing levels of abstraction are appropriate for different parts of a complex application: optimization algorithms deal in vectors and functions, for instance, whereas differential equation solvers are built out of grids and update rules. Open source libraries of optimization algorithms formulated in terms of basic vector calculus types could provide a currently lacking basis for comparison of inversion algorithms in geoscience.

WHAT'S A VECTOR?

A vector is not an array. Consider two length-three arrays, $[1.5, 1.5, 1.5]$ and $[2.0, 2.0, 2.0]$. It is tempting to say that these can be added, but they cannot - did I mention that the units of the first three values are km/s, of the last three, gm/cm^3? The linear algebra concept *vector space* resolves this issue - membership of two vectors in the same vector space asserts that they can sensibly be added. Specification of a vector space of physical data involves specification of units as well, and legitimate manipulation of those units. The operations that vectors must support may be found in the first chapter of a good book on linear algebra, and these operation are attributes of the vector space, not of the vectors that belong to it. Thus a vector is "in" a vector space when the former refers to the latter for its operations, and for compatibility checks. The reward for formulating algorithms in terms of abstract types such as "vector" and "vector space" is that such implementations inherit a guarantee to function in the same way in every specific instance.

Many enormously useful numerical libraries, such as Matlab (Mathworks, 2015) and PETSc (Balay et al., 2015), confound the concept of array with that of vector. One by-product of this confusion is inevitably the demand that dimension be specified - however dimension is not a mandatory attribute of a vector space, even

of a computationally realizable one (consider the vector space of polynomials in a real variable, for instance). Vectors are ultimately implemented with data containers holding floating point samples, of course - usually arrays, but optionally lists, trees, or other suitable data structures. Thus a vector *has* a data container of some sort, but *is not* identical to the container that it owns.

EVALUATING FUNCTIONS

A function f is a triple: a domain V and a range W, both (subsets of) vector spaces, and a rule for assigning to each vector x in V a memeber $f(x)$ of W. Computational realizations of functions may entail many intermediate computations and much storage. Freeing the memory used after each evaluation of the function may be wasteful, since some or all of the intermediate results might be re-used. Not freeing this information when it is no longer needed is also inefficient in a different way.

Consider, for example, a function on a collection of data traces (with samples, headers, etc. making a natural vector space structure) computed by calculating an attribute (for example stack power) of a migrated image. Evaluation requires that the image be computed, but the image is not the result returned by the function. The migrated image must be updated every time the input data vector is changed, so it is not properly an attribute of the function itself. However, if the value corresponding to the *same* data is requested a second time, it is very inefficient to migrate the data all over again. Since the mathematical relation $y = f(x)$ appears to present no opportunity to signal whether intermediate data needs to be re-computed, this appears to be an instance in which the "mathematical API" is inadequate for computational purposes.

Indeed, a number of libraries, such as ROL and NOX in the Trilinos collection (Heroux et al., 2005), create a new "computational function" interface, in effect $y = f(x, isnew)$, giving the programmer the unwelcome responsibility for asserting the new-ness, or otherwise, of an input vector. With a few reasonable conventions, this burden is unnecessary. If Vector data is *encapsulated*, so that it is altered only by basic linear algebra operations and by evaluations of functions, then a *versioning system* is possible, from which in principle it is possible to determine whether vector data has altered since last access. However the reader will see with a moment's reflection that this information cannot reasonably be an attribute of a function. Instead, it is a natural attribute of an *evaluation* type, that

combines a function f and a point x in its domain with the semantics of $f(x)$ *for variable x*. Intermediate data required during evaluation is often re-usable for the computation of gradients, Hessians, and so on. Therefore the evaluation should own an independent copy of the function (the only abstract "handle" on the necessary data) and have access to the vector. The copy of the function should be freed and re-initialized whenver the version count indicates that the vector data has changed.

Figure 1 displays the structure of evaluation, with terminology borrowed from a library that implements it (RVL, described below). Figure 2 flowcharts the evaluation process. The evaluation construction makes the mathematical interface for function evaluation compatible with computational efficiency.
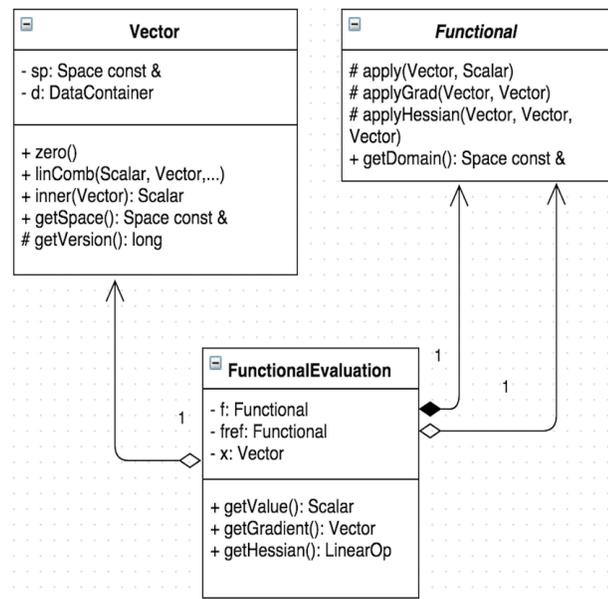

Figure 1: UML-like diagram of a typical function evaluation class. Shows external reference to function and vector, and internal ownership of a copy of the function.
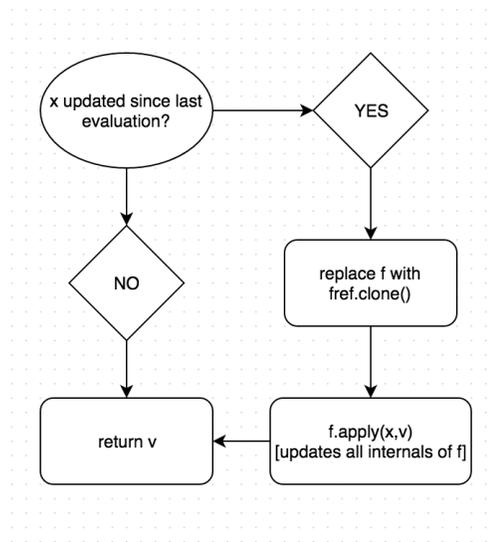
Figure 2: Flow diagram for value computation attribute of an evaluation object.

TIME-DEPENDENT SIMULATION

Any time-domain dynamical law has the form mentioned in the first paragraph. Therefore a natural implementation of any time-stepping method would represent the dynamical state as a vector in the sense described above, and the time step as a function. Spatial gridding then becomes an internal detail of the vector type, and the time step is a vector-valued function (see for example Gockenbach et al. (2002)).

However two compelling reasons argue against this approach. First, the mathematical interface does not conform to computational requirements. Creation and destruction of vectors involves quite a lot of infrastructure not really necessary for time-stepping implementations, and definitely awkward for parallel implementations. Also, assignment is not mathematically legitimate ("x=f(x)"), yet time-stepping is naturally written that way. More practically, excellent libraries such as PETSc (Balay et al., 2015) and Trilinos (Heroux et al., 2005) support implementaton of time stepping but as mentioned above are not consistent with the vector type explained here, at least not without a considerable amount of "wrapper" code.

Therefore array and similar types are the natural domain of time stepping, code, which is naturally used to define functions in the abstract context, and these are available to abstractly formulated algorithms (Symes et al., 2011).

CONCLUSIONS

Several widely-distributed open source libraries incorporate at least some of the observations made here. In particular, the Rapid Optimization Library (ROL) (Kouri et al., 2015), a Trilinos package, uses strong encapsulation of vector data and comes close to a mathematically consistent function interface. It does not however employ a vector space type, nor does it take advantage of the evaluation construction. In a different direction, Madagascar (Holden, 2015) is tied to a specific data representation, but one that is hidden behind a file access interface, and therefore has intrinsically abstract features. So far as the author knows, only the Rice Vector Library (RVL) (Padula et al., 2009) consistently implements the "mathematics as API" concept.

One (of many) obstacles to progress in understanding the effectiveness of Full Waveform Inversion and similar algorithms is the near-impossibility of bench-

mark comparison: details of data representation, simulator structure, and optimization algorithm pervade FWI algorithms and make head-to-head comparison difficult to carry out and even harder to understand. A common approach to organizing the optimization level of FWI applications, based on the intrinsic mathematical structure of these algorithms, would remove one obtstacle to such comparitive research. Use of open-source libraries constructed on this basis is essential to create such a level playing field. Disputes over programming language and other issues have impeded adoption of a common optimization framework, but adherance to a common system of types and practices for their use in algorithms, defined by their mathematical traits rather than by their implementation, could sideline these concerns.

ACKNOWLEDGEMENTS

REFERENCES

Balay, S., S. Abhyankar, M. Adams, J. Brown, K. Brune, P.and Buschelman, L. Dalcin, V. Eijkhout, W. Gropp, D. Kaushik, M. Knepley, L. Curfman McInnes, K. Rupp, B. Smith, S. Zampini, and H. Zhang, 2015, PETSc Web page (http://www.mcs.anl.gov/petsc).

Gockenbach, M., D. Reynolds, P. Shen, and W. Symes, 2002, Efficient and automatic implementation of the adjoint state method: ACM Transactions on Mathematical Software, **28**, 22–44.

Heroux, M., R. Bartlett, V. Howle, R. Hoekstra, J. Hu, T. Kolda, R. Lehoucq, K. Long, R. Pawlowski, E. Phipps, A. Salinger, H. Thornquist, R. Tuminaro, J. Willenbring, A. Williams, and K. Stanley, 2005, An overview of the Trilinos project: ACM Transactions on Mathematical Software, **31**, 397–423.

Holden, J., 2015, The genesis of Madagascar: The Leading Edge, **34**, 1386–1388.

Kouri, D., D. Ridzal, B. van Bloemen Waanders, and G. von Winckel, 2015, ROL Web page (http://trilinos.org/packages/rol).

Mathworks, 2015, Matlab version 8.5.0.197613 (r2015a): The Mathworks, Inc.

Padula, A. D., W. Symes, and S. D. Scott, 2009, A software framework for the ab-

stract expression of coordinate-free linear algebra and optimization algorithms: ACM Transactions on Mathematical Software, **36**, 8:1–8:36.

Ruelle, D., and F. Takens, 1971, On the nature of turbulence: Communications in Mathematical Physics, **20**, 167–192.

Symes, W., D. Sun, and M. Enriquez, 2011, From modelling to inversion: designing a well-adapted simulator: Geophysical Prospecting, **59**, 814–833. (DOI:10.1111/j.1365-2478.2011.00977.x).

Full Waveform Inversion via Source-Receiver Extension

Guanghui Huang and Rami Nammour[†] and William Symes**

ABSTRACT

The source-receiver extension models each seismic trace by its own proper source wavelet. The extra degrees of freedom introduced by this extended source (trace-dependent wavelet) permit good fit to every data trace, regardless of velocity. For kinematically correct velocity, all of the traces in each common source gather are well-fit by the same source wavelet. Thus the variance of the data-fitting source wavelets indicates correctness of velocity. Provided that a common source signature is known, an effective measure of source wavelet variance is the second moment of the squared signature-deconvolved extended source traces. For pure transmission data with single arrival events, this function of velocity approximates weighted mean-square traveltime error, so that its minimization is equivalent to traveltime tomography. Moreover, iterative minimization converges from initial models that would be cycle-skipped for ordinary full waveform inversion. The relation with traveltime misfit breaks down if multiple ray paths connect sources and receivers, due to presence in the extended source of slowly decaying energy at large time lags unrelated to the traveltime error. These slowly decaying signals arise from Green's function spectral notches generated by multiple arrivals of similar energy. Tikhonov regularization (prewhitening) of the least-squares data fitting suppresses these large-lag contributions to the extended source, at the cost of considerably degraded data fit. Numerical examples suggest that this regularized formulation of source-receiver extended inversion is capable of recovering reasonably good velocity models from synthetic transmission and reflection data without stagnation at physically irrelevant models frequently encountered by standard full waveform inversion, but with essentially the same computational cost.

INTRODUCTION

Full waveform inversion (FWI) estimates subsurface structure with high precision by minimizing the differences between the synthesized data and recorded data in the least square sense (Tarantola, 1984; Virieux and Operto, 2009). However, the domain of convexity of the FWI objective function for velocity estimation is generally quite small, on the order of a wavelength in diameter, and iterative optimization methods starting further from the global minimizer may stagnate at physically meaningless apparent optima. The root cause of this behaviour is the tendency of predicted data to be out of phase with, or even orthogonal to, recorded data in large regions of model space ("cycle-skipped"), and therefore very far away in the mean square sense. This problem may be avoided to some extent by a combination of initial model accuracy, high S/N at the lowest recorded frequencies (achieved in some surveys, not in others), and data fitting in expanding frequency bands, from low to high (Bunks et al., 1995; Pratt, 1999; Pratt and Shipp, 1999; Sirgue and Pratt, 2004; Virieux and Operto, 2009). The stagnation problem is somewhat less severe for fitting of refracted energy (diving or transmitted waves), as noted already by Gauthier et al. (1986).

This paper describes an alternative to least-squares data fitting and illustrates its behaviour with several synthetic examples. The basis of this approach is the *source-receiver extension* of constant density acoustic modeling: an independent source trace is provided for each data trace. Appropriate choice of the source trace fits the data trace, for any choice of velocity field. Since the mean-square error is small for all such extended models (velocity plus source traces), some other objective must take over the role of fit error in standard FWI, and drive the extended model towards a physical (non-extended) model that explains the data, hence solves the FWI problem. All traces in a common source gather should share the same source, so it is natural to penalize the deviation from common source. Several penalty functions for measuring this deviation are available. Since only one extended source matches the data for each velocity model, any such penalty is implicitly a function of velocity, and may be used as an objective fuction in an iterative optimization algorithm. If the iteration succeeds in driving the exension penalty to zero, then the process achieves a full waveform inversion: the iterates fit the data well throughout the inversion process, and converge to a physical model.

The main objective of this paper is to demonstrate two facts about source-receiver extended waveform inversion. First, for transmission data with single

arrivals (that is, a unique ray path between each source and receiver), the objective function of source-receiver extended waveform inversion is approximately proportional to the mean-square traveltime error. It follows that under these circumstances, iterative inversion as envisioned in the preceding paragraph succeeds in finding a (physical) model that fits the data, that is, achieves a full waveform inversion. This process produces roughly the same model as does traveltime tomography: in effect, it performs tomography with waveform data. The inversion converges from initial models that are hopelessly cycle-skipped for standard data-domain FWI, as does traveltime tomography.

Second, for transmission data with multiple arrivals (that is, more than one distinct raypath connecting at least some source-receiver pairs), the connection between traveltime tomography and source-receiver waveform inversion is broken, and the source-receiver objective function exhibits the same tendency to stagnate at non-optimal solutions as does FWI. This observation applies in particular to inversion of diving wave data: if the diving waves are triplicated, source-receiver extended inversion is no more likely to converge to a useful velocity model than is FWI. We offer an explanation of this phenomenon, and numerical examples that demonstrate it. In brief, multiple arrivals with certain amplitude relations amongst the branches lead to slowly decaying components in the extended source, contributing energy at large lags having nothing to do with the overall traveltime error between predicted and observed data traces.

We suggest a partial remedy for this misbehaviour: from the spectral point of view, it arises from small eigenvalues of the normal operator or Hessian, equivalent to notches in the Green's function spectrum. As is routine with other ill-conditioned inverse problems, one can suppress the effect of small eigenvalues by Tikhonov regularization (Engl et al., 1996) - in other words, prewhitening deconvolution by the Green's function. Provided with an adequate regularization weight parameter, this regularized source-receiver extended inversion tends to recover the convexity exhibited by the unregularized method in the single-arrival case, and converges to an approximation to the global minimizer. A large regularization weight is required to achieve this goal, causing substantial data misfit (up to 50% in several of our examples). We illustrate all of these claims with numerical examples. The regularization weights used in the examples were chosen by means of a discrepancy principle, that is, setting them so that the initial data misfit using the regularized extended model is within limits proportional to the initial physial model misfit. The necessary proportion of the initial misfit appears to be substantial - we have used 20-50% in our examples.

Although our theoretical developments pertain to transmission data, the source-receiver extension applies also to the reflection configuration. For completeness, we include an example in which a salt lens is recovered from reflection data via source-receiver inversion. Rather heavy regularization seems to be required in this case as well.

The source-receiver extension concept is not new: Song and Symes (1994); Symes (1994); Plessix et al. (2000); Plessix (2000); Pratt and Symes (2002); Luo and Sava (2011) investigate data fitting via source-receiver extension to enhance the convergence of FWI. Warner and Guasch (2014, 2016) uses a very similar approach as part of Adaptive Waveform Inversion, and show its capacity to enlarge the domain of attraction for FWI and its practicality for application to contemporary 3D field surveys. Of several possible choices of penalty, we use the dispersion about zero lag suggested in several of these works (Plessix, 2000; Luo and Sava, 2011; Warner and Guasch, 2014, 2016). The relation between source-receiver extended waveform inversion and traveltime tomography was explained by Song and Symes (1994) in the context of crosswell waveform tomography. Symes (1994) showed that this link is generally broken when multiple arrivals are present in transmission data, using an argument based on causal deconvolution. Neither of these older works actually implemented source-receiver extended inversion, as we do here. Plessix et al. (2000); Plessix (2000) applied the source-receiver extension approach to field crosswell data. They observed the effect described by Symes (1994), and avoided it by using source-receiver pairs with sufficiently different depths to avoid most guided wave energy. Tikhonov regularization is used in many inversion algorithms, including some of source-receiver type mentioned above, but our use of regularization specifically to control the slowly decaying energy produced by multiple arrivals seems to be new. Several authors (Luo and Sava, 2011; Warner and Guasch, 2014, 2016) have applied source-receiver extension methods to reflection configurations.

It should be understood that many factors could lead to failure of source-receiver extension to yield an objective with a large domain of convexity about its global minimizer. To name just a few such factors, reflected waves in transmission data, or multiply reflected waves for reflection data, or out-of-plane reflections for 2D data, or shear or converted waves are all potentially capable of causing failure. However we emphasize that even in the complete absence of these other factors, with data that actually arises in transmission through a slowly varying background, the presence of multiple energetic arrivals is sufficient to derail this approach. As our third example underlines, this conclusion applies particularly

to diving wave fields (the principal data of conventional FWI), which may easily contain multiple arrivals from a localized source.

In the following pages, we first review the theoretical foundation of source-receiver extended waveform inversion, and explain the structure of the algorithm. Then we illustrate its behaviour with four 2D numerical examples. The first two are set in an idealized crosswell geometry, and illustrate the capability of source-receiver extended waveform inversion to provide tomographic-quality solutions, the obstacle to velocity updating posed by multiple arrivals, and a partial remedy through Tikhonov regularization. The other two examples use surface acquisition geometry. The first of these is a pure diving wave problem, with no reflections. The target model generates multiple ray paths in the diving wave field, with the same damaging effect on velocity updating via source-receiver extended wave-form inversion as in the previous crosswell example. Tikhonov regularization, with a regularization weight chosen by trial and error, suffices for these examples to restore convergence to a useful model. Finally, we include an example of surface data inversion via source-receiver extension, in which a model of a salt lens embedded in sediments is recovered from reflection data. All examples begin with homogeneous or simple layered initial guesses. In all cases, attempted full waveform inversion fails, whereas the regularized source-receiver extended inversion succeeds. In the "salt" example, the inclusion emerges from the background without any special effort. All except the first example require regularization, with the regularization weight estimated by trial and error.

We end with a discussion of several obvious or not-so-obvious capabilities and limitations of source-receiver extended waveform inversion.

THEORY

The acoustic model of seismic wave propagation treats the excess pressure field u as the solution of the acoustic wave equation

$$\frac{1}{v^2}\frac{\partial^2 u}{\partial t^2} - \nabla^2 u = \delta(\mathbf{x} - \mathbf{x}_s)f(t), \tag{1}$$

$$u = 0, t << 0 \tag{2}$$

The energy source is modeled here as a point isotropic radiator with source pulse $f(t)$. The forward modeling operator maps the source to the data traces, presumed

to be perfect pressure measurements, and depends on the wave velocity:

$$S[v]f(\mathbf{x}_r, t; \mathbf{x}_s) = u(\mathbf{x}_r, t; \mathbf{x}_s). \tag{3}$$

In equations 1 and 3, the source and receiver locations, $\mathbf{x}_s$ and $\mathbf{x}_r$ respectively, define the acquisition geometry of the survey.

The standard least-squares inversion, or FWI, problem is: given $f(t)$ and $d(\mathbf{x}_r, t; \mathbf{x}_s)$, determine $v(\mathbf{x})$ so as to minimize

$$J_{\text{FWI}}[v] = \frac{1}{2}\|S[v]f - d\|^2. \tag{4}$$

The vertical bars denote the mean square or L^2-norm squared, in other words the sum over all active values of $\mathbf{x}_r, t$, and $\mathbf{x}_s$, possibly scaled by cell volume or other factors.

As mentioned in the introduction, the function defined by equation 4 is difficult to minimize directly, so we will explore an alternative approach to its minimization, through models that violate at least some of the modeling assumptions made above.

Source-Receiver Extended Modeling

The source-receiver extension introduces a trace-dependent source function $\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s)$ to replace $f(t)$. The extended acoustic system is

$$\frac{1}{v^2}\frac{\partial^2 \bar{u}}{\partial t^2} - \nabla^2 \bar{u} = \delta(\mathbf{x} - \mathbf{x}_s)\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s), \tag{5}$$

$$\bar{u} = 0, t << 0 \tag{6}$$

and the extended forward modeling operator is defined by sampling the extended pressure field $\bar{u}$ as before:

$$\bar{S}[v]\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s) = \bar{u}(\mathbf{x}_r, t; \mathbf{x}_s). \tag{7}$$

Since the extended source is trace-dependent, it is straightforward to fit the data, which is not the case in general with the non-extended source unless the data kinematics are well-predicted by the velocity. Denote by $G[v](\mathbf{x}_r, t; \mathbf{x}_s)$ the causal

Green's function of the acoustic wave equation, that is, the solution of the system 1, 2 with $f(t) = \delta(t)$. Then

$$\bar{S}[v]\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s) = G[v](\mathbf{x}_r, t; \mathbf{x}_s) *_t \bar{f}(\mathbf{x}_r, t; \mathbf{x}_s). \tag{8}$$

Equation 8 shows that computing the source-receiver extended forward map involves minimal expense beyond that of the non-extended forward map defined in equation 3: since $G[v] = S[v]\delta(t)$, computing the action of $\bar{S}[v]$ requires computing the action of $S[v]$, followed by one additional convolution per output trace.

If we denote by $(G[v]*)^{-1}$ a convolution inverse to $G[v]$, then

$$\bar{S}[v](G[v]*)^{-1}d(\mathbf{x}_r, t; \mathbf{x}_s) = d(\mathbf{x}_r, t; \mathbf{x}_s). \tag{9}$$

That is, $S[v]^{-1} = (G[v]*)^{-1}$. This operation assumes that the Green's function has a convolution inverse, of course.

In this paper, we will assume that all traces are defined on the same time interval $[0, T]$. To accommodate timing errors associated with erroneous velocities, we extend the time interval for both the (extended) source and the data to $[-T, T]$, padding the data with zeros for $t < 0$. Then we regard all functions as periodic in T of period 2T, and use circulant convolution, which of course in the Fourier domain amounts to multiplication. So the convolution inverse of the Green's function is simply its reciprocal in the Fourier domain, which is *a priori* available only if the Fourier transform has no zeroes. Sometimes this is the case, sometimes not, as will be illustrated below.

Assuming that $S[v]$ is invertible, the extended source that explains the data d is (exactly!)

$$\bar{f}[v] = S[v]^{-1}d. \tag{10}$$

Extended source-receiver waveform inversion

The extended source constructed in equation 10 is mostly likely unphysical, in which the physics defined at the beginning of this section required that sources are uniform across all traces, that is, $\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s) = f(t)$. Since the data can be fit

(assuming deconvolvable Green's function) for any velocity, it is only this requirement that provides velocity updates. A simple way to quantify failure of the extended source to match a physical one uses an *annihilator*, that is, an operator A that produces a zero result when applied to a physicalliy consistent source. Two possible choices are differentiation of the source with respect to $\mathbf{x}_s$ and $\mathbf{x}_r$ (Song and Symes, 1993; Pratt and Symes, 2002), or forcing the convolution quotient of the extended and target (non-extended) sources to resemble the delta function, for example by penalizing the second moment of the squared signal, also known as dispersion about zero lag (Plessix et al., 1999; Luo and Sava, 2011; Warner and Guasch, 2014, 2016). We will use the second option in the work reported below, as it is somewhat simpler to implement. Specifically,

$$A\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s) = t(f^\dagger) * \bar{f} \tag{11}$$

Here $f^\dagger$ is an approximate inverse, or shaping filter, for the known (common to all traces) source wavelet f. It satisfies

$$f^\dagger * f = \chi, \tag{12}$$

in which χ is an approximate (band-limited) delta function, for example a zero-phase bandpass filter.

If we apply A to the v-dependent extended source $\bar{f}[v]$ defined in equation 10, we obtain an index of velocity correctness: if the velocity model and data are kinematically compatible, then all of the inverted sources $\bar{f}[v](\mathbf{x}_r, t; \mathbf{x}_s)$ are approximately the same as the source $f(t)$, hence signature deconvolution should yield an approximate delta function at zero lag, and that is in turn nearly annihilated by multiplication by t. Kinematic disagreement between model and data should lead to larger $A\bar{f}$.

We capture this idea in an objective function:

$$J[v] = \frac{1}{2}\|A\bar{f}[v]\|^2 \tag{13}$$

In effect, the model over which this objective is to be optimized includes both the velocity $v(\mathbf{x})$ and the extended source function $\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s)$, hence is an *extended model*. The optimization is treated as a *nested* problem, with $\bar{f}$ determined as a function of v via deconvolution (equation 10), then v determinined by minimizing $J[v]$ (equation 13).

Remark 1. *The nested design of the optimization problem defined by equations 10, 11, and 13 is essential, not merely a computational convenience. Minimization of alternative objective functions of $v(\mathbf{x}), \bar{f}(\mathbf{x}_r, t; \mathbf{x}_s)$, such as the penalty function*

$$J_\alpha[v, \bar{f}] = \frac{1}{2}\|A\bar{f}[v]\|^2 + \frac{\alpha^2}{2}\|\bar{S}[v]\bar{f} - d\|^2, \tag{14}$$

over v and $\bar{f}$ jointly, turns out to be very inefficient, with (in principle) arbitrarly slow convergence. The cause of this misbehaviour is the very different sensitivities of J_α to v and $\bar{f}$. A change in $\bar{f}$ changes J by an amount proportional to the mean square of $\bar{f}$. In contrast, a change in v results in a change in traveltimes, therefore in general a shift in the events in $\bar{S}[v]\bar{f}$. The rate of change of J_α with respect to v therefore involves the derivative (in t) of $\bar{f}$, which is roughly speaking bigger than $\bar{f}$ in mean square by a factor of the maximum frequency. Therefore relatively high-frequency data, desirable from the point of view of model resolution, yields very different sensitivities to $\bar{f}$ and v, that is, ill-conditioning of the Hessian of J_α. Objectives with ill-conditioned Hessians cause local optimization algorithms to perform poorly - see for example Nocedal and Wright (1999) for information on this point. A problem more amenable to solution via local optimization can be recovered by optimizing J_α first over $\bar{f}$ to create a reduced objective depending only on v, which is then optimized over v - that is, a nested optimization, similar to that defined in equations 13, 10. See Symes (2015) for an explanation, and Huang and Symes (2015b) for an explicit illustration of the performance contrast between nested and non-nested optimization for a different model extension. We will not use the penalty function J_α in the work reported here.

Relation with Traveltime Tomography

Consider for the moment 3D wave propagation through slowly varying velocity fields v for which the associated ray field connects each source-receiver pair with a unique ray. "Slowly varying" means smooth on the wavelength scale, lacking embedded reflectors, dominated by relatively low spatial frequencies. For such smooth single-arrival models, geometric acoustics provides asmptotic Green's function approximation:

$$G[v](\mathbf{x}_r, t; \mathbf{x}_s) \approx a[v](\mathbf{x}_r; \mathbf{x}_s)\delta(t - \tau[v](\mathbf{x}_r; \mathbf{x}_s)). \tag{15}$$

Here a is the geometric amplitude, and τ is the traveltime, for the ray between $\mathbf{x}_s$ and $\mathbf{x}_r$; of course, both depend on v.

Assume that the data d is the image under S of a physical model v_*, f, with v_* producing a single arrival, that is, enjoying the asymptotic Green's function approximation 15. Then

$$d(\mathbf{x}_r, t; \mathbf{x}_s) = S[v_*]f(\mathbf{x}_r, t; \mathbf{x}_s) \approx a[v_*](\mathbf{x}_r, \mathbf{x}_s)f(t - \tau[v_*](\mathbf{x}_r, \mathbf{x}_s)). \tag{16}$$

Assuming that the trial velocity v also produces single arrivals, the extended source $\bar{f}[v]$ defined by equation 10 is

$$\bar{f}[v] \approx \frac{a[v_*](\mathbf{x}_r, \mathbf{x}_s)}{a[v](\mathbf{x}_r, \mathbf{x}_s)} f(t - (\tau[v_*](\mathbf{x}_r, \mathbf{x}_s) - \tau[v](\mathbf{x}_r, \mathbf{x}_s))). \tag{17}$$

From equations 11, 12, and 13,

$$
\begin{aligned}
J[v] &\approx \sum_{\mathbf{x}_r, \mathbf{x}_s} \left(\frac{a[v_*](\mathbf{x}_r, \mathbf{x}_s)}{a[v](\mathbf{x}_r, \mathbf{x}_s)} \right)^2 \int dt\, t^2 (\chi(t - (\tau[v_*](\mathbf{x}_r, \mathbf{x}_s) - \tau[v](\mathbf{x}_r, \mathbf{x}_s))))^2 \\
&= \sum_{\mathbf{x}_r, \mathbf{x}_s} \left(\frac{a[v_*](\mathbf{x}_r, \mathbf{x}_s)}{a[v](\mathbf{x}_r, \mathbf{x}_s)} \right)^2 \int dt\, (t + (\tau[v_*](\mathbf{x}_r, \mathbf{x}_s) - \tau[v](\mathbf{x}_r, \mathbf{x}_s)))^2 \chi(t)^2. \quad (18)
\end{aligned}
$$

Since χ is assumed zero-phase (symmetric about $t = 0$), this

$$
\begin{aligned}
&= \sum_{\mathbf{x}_r, \mathbf{x}_s} \left(\frac{a[v_*](\mathbf{x}_r, \mathbf{x}_s)}{a[v](\mathbf{x}_r, \mathbf{x}_s)} \right)^2 \int dt\, t^2 \chi(t)^2 \\
&\quad + \sum_{\mathbf{x}_r, \mathbf{x}_s} \left(\frac{a[v_*](\mathbf{x}_r, \mathbf{x}_s)}{a[v](\mathbf{x}_r, \mathbf{x}_s)} \right)^2 \int dt\, (\tau[v_*](\mathbf{x}_r, \mathbf{x}_s) - \tau[v](\mathbf{x}_r, \mathbf{x}_s))^2 \chi(t)^2 \quad (19)
\end{aligned}
$$

Since χ is an approximate delta, its square is concentrated near $t = 0$, so the second moment of its square (the integral in the first summand in equation 19) is small. We conclude that

$$J[v] \approx \left(\int dt \chi(t)^2 \right) \sum_{\mathbf{x}_r, \mathbf{x}_s} \left(\frac{a[v_*](\mathbf{x}_r, \mathbf{x}_s)}{a[v](\mathbf{x}_r, \mathbf{x}_s)} \right)^2 (\tau[v_*](\mathbf{x}_r, \mathbf{x}_s) - \tau[v](\mathbf{x}_r, \mathbf{x}_s))^2. \tag{20}$$

That is, $J[v]$ is approximately a weighted mean of the square traveltime error. The weights are the squared amplitude ratios.

Assuming that the trial velocities v are kept safely away from creating caustics, the ratio of amplitudes between trial and target (v_*) velocities lies between two

positive numbers, the exact values of which depend on precisely what is meant by "safely away". If J approaches its minimum value (zero) through a sequence of such velocity models, then the (unweighted) mean square traveltime error must approach zero also, and vice versa. Thus global minimization of J is equivalent to global minimization of the standard mean-square traveltime tomography objective. Similar arguments show that the gradient of J is small only when the gradient of the traveltime tomography objective is small, and that the Hessian of J at the exact solution ($v = v_*$) dominates the Hessian of traveltime tomography objective (see Huang and Symes (2015a) for details). That is, there is a domain in model space, whose size is independent of the bandwidth of the data, in which minimization of J determines the same aspects of the velocity model as does traveltime tomography.

It is possible to make the foregoing statements mathematically precise: that is, the approximations indicated above are all in the same sense (high frequency asymptotics), and sets of models "safely" far from generating caustics can be precisely characterized (Song, 1994). It is also possible to justify precisely the same conclusions for 2D propagation. These refinements are beyond the scope of this paper, but are consistent in every way with the proposition that minimization of the objective J defined here is effectively equivalent to traveltime tomography for models that generate only single arrivals.

Effect of Multiple Arrivals

The story changes dramatically if the data exhibits multiple energetic arrivals, even if in other ways it conforms to the limitations mentioned in the last section, that is, transmission through a slowly varying material model. As pointed out already in Symes (1994), with multiple energetic arrivals, it is possible that the (unregularized) source-receiver extended waveform inversion objective may be as non-convex as the FWI objective. Consequently, the source-receiver extended waveform inversion algorithm can fail to produce kinematically accurate velocity estimates.

We present here a different viewpoint from Symes (1994), who based his discussion on causal deconvolution. As mentioned earlier, our work uses circulant deconvolution (Fourier division) instead. From this point of view, the pathology is a by-product of spectral notches that may develop in the Green's function. Though this argument does not yield the precise conclusion about non-convexity,

it is simpler and indicates another manifiestation of the underlying pathology.

For the generic source-receiver pair (not lying precisely on a caustic), the 3D Green's function takes the asymptotic form

$$G[v](\mathbf{x}_r, t; \mathbf{x}_s) \approx \sum_{i=0}^{N} a_i[v](\mathbf{x}_r; \mathbf{x}_s) H^{v_i} \delta(t - \tau_i[v](\mathbf{x}_r; \mathbf{x}_s)). \tag{21}$$

Here a_i is the geometric amplitude, and τ_i the traveltime, for the ith ray connecting $\mathbf{x}_s$ and $\mathbf{x}_r$. H is the Hilbert transform, and its power v_i is either 0 or 1 and is related to the Maslov index of the ray in the ray field emanating from the source.

$S[v]^{-1}$ amounts to deconvolution by $G[v]$. However the Fourier transform of G is an exponential sum, hence may have zeros or approximate zeros (or, as they are known in this literature, notches). The simplest cartoon example, actually relevant to a synthetic example presented later in this paper, is

$$G(t) \approx a(\delta(t) + \delta(t - \Delta t)) \tag{22}$$

Then $Sf = G * f = 0$ if $f(t) = \cos\dfrac{\pi t}{\Delta t}$: that is, the Fourier transform of G vanishes at odd multiples of $1/\Delta t$.

If G has literal zero Fourier componnents, as in the cartoon example, then equation 9 may not have a solution, that is, G does not have a convolution inverse, and if 9 has a solution, it is not unique. More likely to occur are very small values of the Fourier transform for which the corresonding sinusoid will be vastly overemphasized in the solution of equation 9. If the data d is noise-free data from the same model, that doesn't matter, since the small Fourier component of G is already part of the data. However if the trial model is not very close to the model used to generate the data, then the corresponding data component is likely to be large, resulting in a large sinusoidal contribution at the notch frequency and receiver location to the extended source. Since the sinusoid is non-decaying, $\bar{f}[v]$ defined by equation 10 acquires energy at time lags that have nothing to do with the overall travel time difference between the trial model and the global minimizer. Thus the connection between travel time error and the value of J is broken in this case.

The more refined analysis presented by Symes (1994) actually shows that the domain of convexity of source-receiver extended inversion generally has diameter proportional to a wavelength, similar to FWI: convergence requires that the initial model predict data event times to within a half wavelength.

Regularized Source-Receiver Extended Inversion

To make estimation of the extended source $\bar{f}$ robust against spectral zeroes of the Green's function, we replace equation 9 with the regularized least squares problem: choose $\bar{f}$ to minimize

$$\frac{1}{2}\|\bar{S}[v]\bar{f} - d\|^2 + \frac{\varepsilon}{2}\|\bar{f}\|^2. \tag{23}$$

The minimizer (again denoted $\bar{f}[v]$) solves the normal equation

$$(\bar{S}[v]^T \bar{S}[v] + \varepsilon^2 I)\bar{f}[v](\mathbf{x}_r, t; \mathbf{x}_s) = S[v]^T d(\mathbf{x}_r, t; \mathbf{x}_s) \tag{24}$$

and depends on v, d, and ε. The matrix of the operator on the left-hand side of 24 is the autocorrelation of G, "prewhitened" by addition of ε^2. The effect of the prewhitening is to uniformly increase the entire power spectrum of G, thus making the matrix on the left-hand side of equation 24 better-conditioned. Of course the normal equation 24 is again a convolution equation, and can be solved exactly by means of the Fourier transform.

We define the regularized source-receiver extension objective $J[v]$ again by equation 13, but now with $\bar{f}[v]$ the solution of equation 24.

We expect the modified $J[v]$ to oscillate less than the nonregularized version discussed above: regularization reduces the very large components in the extended source, arising from the spectral notches, proportionally more than other components which presumably contribute energy at time lags more proportional to the overall traveltime error between the trial and target models. It is easy to see from low-dimensional matrix analogues that we cannot expect the regularization parameter ε to be particularly small.

We use a version of the discrepancy principle (control of the data residual) to set ε. We compute the relative initial data residual e_0 and relative extended initial data residual $\bar{e}_0$ using the initial velocity v_0:

$$e_0 = \frac{\|S[v_0]f - d\|}{\|d\|}, \qquad \bar{e}_0 = \frac{\|\bar{S}[v_0]\bar{f}[v_0] - d\|}{\|d\|} \tag{25}$$

in which f is the known source wavelet and $\bar{f}$ is the extended source function, estimated by solving equation 24 with the initial velocity. We adjusted ε by trial-and-error until the ratio $\bar{e}_0/e_0$ lay in the range $0.2 - 0.5$. The (somewhat arbitrary)

bounds are chosen to ensure that ε is large enough that the residual is substantially larger than zero (the expected value for $\varepsilon = 0$), but small enough that the data is substantially better fit than is possible with a physical source ($\bar{f} = f$). Thus ε is large enough to suppress the notch contributions to some exent, but small enough to ensure that the larger events in the data must be fit to some extent. This method for selecting ε has only a heuristic justification, but has functioned well in the examples shown below and in other similar examples not shown here.

Gradient Computation

We use a gradient-based method to minimize J. Appendix A shows how to compute this gradient. Define $\bar{w}$ to be the solution of

$$(\bar{S}[v]^T \bar{S}[v] + \varepsilon^2 I)\bar{w} = A^T A \bar{f} \tag{26}$$

and $\bar{r}$ by

$$\bar{r} = \bar{w} \star (d - \bar{S}[v]\bar{f}) - \bar{f} \star \bar{S}[v]\bar{w} \tag{27}$$

("$\star$" denotes cross-correlation). Then

$$\nabla J[v] = (D\bar{S}[v]\delta)^T \bar{r} \tag{28}$$

in which $(D\bar{S}[v]\delta)^T$ is the well-known impulsive reverse time migration operator. See Appendix A for details.

Computational cost

The computational cost of source-receiver extended waveform inversion is comparable to the cost of standard least-squares FWI. Each function value $J[v]$ requires computation of the Green's function (therefore one modeling step), a deconvolution (solution of equation 24), and some vector algebra. A gradient evaluation $\nabla J[v]$, as defined in the last section, adds another deconvolution (equation 26), and two more convolutions (equation 27) and some more vector algebra, followed by reverse time migration (equation 28). Thus each step of a gradient based optimization is of roughly the same cost as a step of the same algorithm to FWI. Concerning a comparison of total costs over the entire iteration, all that can be said in general is that an algorithm that produces a useful approximation in a reasonable number of steps is in a real sense infinitely cheaper than one that does not.

FWI as post-process

In principle, the principal goal of inversion is production of an explanatory physical model that fits the data. The regularization level required to induce convergence of source-receiver extended inversion from seriously wrong initial models has a negative effect on data fit. However in the examples to follow, regularized source-receiver inversion produces kinematically accurate velocity models, while relaxing the data fit constraint of the unregularized algorithm. Its output should be acceptable input to conventional FWI, and the latter algorithm should reduce the remaining data misfit. Therefore we shall in most cases follow inversion via source-receiver extension with FWI initialized on the final source-receiver model.

NUMERICAL EXAMPLES

We present two sets of numerical examples illustrating the performance of source-receiver extended waveform inversion in comparison to FWI. The first set uses an idealized crosswell geometry, with the source locations on one face of the rectangular scattering domain, the receivers on the opposite face. The second set mimics surface acquisition, with sources and receivers on the same face of the rectangular domain.

We use the limited memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) algorithm with backtracking line search to assure compliance with the weak version of Wolfe's conditions for global convergence to a stationary point (Nocedal and Wright, 1999). For each example, we will apply LBFGS both to solve the full-bandwidth FWI problem, and to minimize the source-receiver extended waveform inversion objective defined in equation 13, and the FWI objective (equation 4). Forward modeling is implemented in frequency domain, using a nine-point (4th order, cross-shaped) stencil to approximate the Helmholtz operator, and a direct matrix solver. The least squares problems defined in equations 24 and 26 are diagonal in the Fourier domain because of our use of circulant deconvolution, so these are solved to machine precision as well. The objective function and gradient are therefore computed to machine precision, at the discrete level.

Crosswell Acquisition Geometry

The first two examples use idealized crosswell acquisition, that is, sources lie on one boundary face of a rectangular scattering region, receivers on the opposite face. The first shows that source-receiver extended waveform inversion without regularization (i.e. $\varepsilon = 0$) may converge to a reasonable solution of inverse problem when FWI fails to do so. In this example, the data exhibits only a single arrival (though partly as a result of wavefront healing). Our previous analysis (Huang and Symes, 2015a) explains the behaviour of source-receiver extended waveform inversion in this case: when raypaths from source to receiver are unique, source-receiver extended waveform inversion is equivalent to least-squares traveltime tomography, and delivers a comparable solution. The second shows that unregularized source-receiver extended waveform inversion may fail in the same way as FWI if multiple arrivals are present in the data with significant energy. We observe that this phenomenon may be understood as ill-posedness of the inner (source-estimation) problem, and that Tikhonov regularization can restore apparent convergence to a tomographic-quality solution.

Weak Low-Velocity Lens

The target velocity model for the first example is a Gaussian low velocity anomaly embedded in a constant background velocity $v_0 = 2$ km/s (Figure 1), i.e.,

$$v(x, z) = 2 - 0.7 e^{-\frac{(x-1)^2}{0.5^2} - \frac{(z-1)^2}{0.25^2}} \text{ km/s} \tag{29}$$

The sources and receivers (indicated with white triangle and green circle in Figure 1, respectively) are placed at $x_s = 0.01$ km and $x_r = 1.99$ km, respectively. Thirty-nine shots are evenly spaced between $z_s = 0.05$ km to $z_s = 1.95$ km, and 199 receivers are located from $z_r = 0.01$ km to $z_r = 1.99$ km with $\Delta z_r = 0.01$ km.

For display purposes, we synthesize a time-domain solution of the target problem from frequency-domain fields, and show the recorded data for the center shot at $z_s = 1$ km in Figure 2a. The effective time-domain source is a boxcar 0.5-30 Hz bandpass filter. We use the constant velocity $v_0 = 2$ km/s as the initial model. The simulated data in Figure 2b using the initial velocity for the centered shot shows a traveltime error larger than a half-wavelength at the median frequency of 12 Hz for the central data traces of the recorded data. For inversion, we use the data frequency band from 3 Hz to 20 Hz. After 21 LBFGS iterations, FWI

is stuck in a physically meaningless solution, whereas source-receiver extended waveform inversion produces a reasonable estimate of v, see Figure 3a-3b for the detailed result, respectively. To further confirm the kinmenatic accuracy of the source-receiver extended waveform inversion solution, we display the extended sources at initial velocity (Figure 4a) and final inverted velocity (Figure 4b).
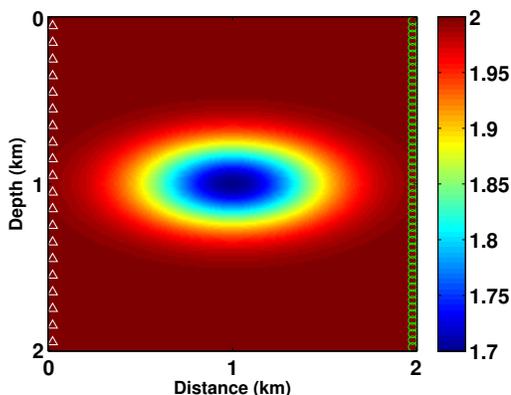

Figure 1: First crosswell example: target velocity model with slow Gaussian anomaly. Lowest velocity is 1.7 km/s

We can understand this example on basis of theory presented in the last section. Examination of Figure 2a reveals that a caustic likely is present in the ray field of this example, but that finite frequency has "healed" it and presented an effective single arrival. Therefore, we would expect the source-receiver extended inversion to converge even from an initial model that is cycle-skipped for FWI to a model of tomographic quality, and indeed it does.

Strong Low-Velocity Lens

The second example (Figure 5) keep the shape velocity anomaly, but makes it stronger:

$$v(x,z) = 2 - 0.6e^{-\frac{(x-1)^2}{0.5^2} - \frac{(z-1)^2}{0.25^2}} \tag{30}$$

Now the lowest velocity of this model is 1.4 km/s. We use the same source and receiver geometry and the data frequency band as the first example. Synthetic data (Figure 6a-6c) for three shot position $z_s = 0.1, 0.5, 1$ km shows energetic later arrivals. For comparison, Figure 7a-7c shows simulated data with initial velocity $v_0 = 2$ km/s at the same shot positions.
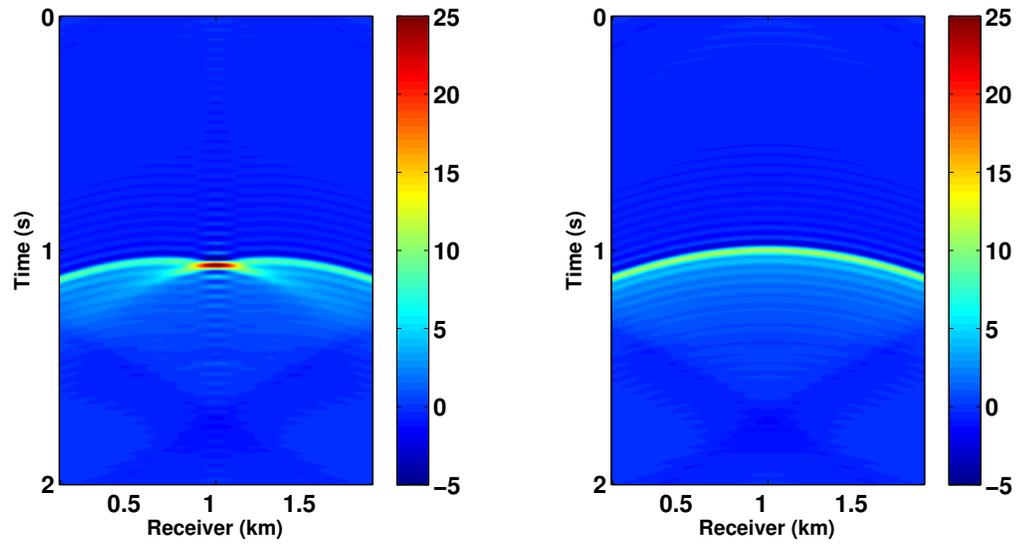
Figure 2: First crosswell example: comparison of shot gathers for the center shot $z_s = 1$ km: (a) target data and (b) simulated data using initial velocity
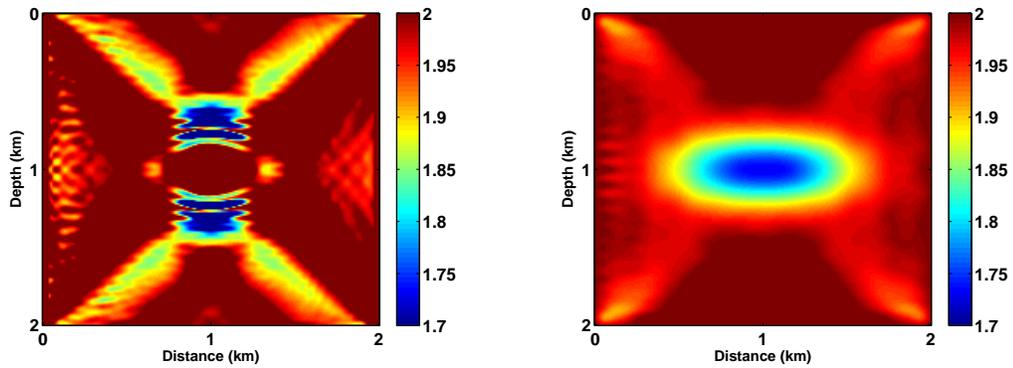

Figure 3: First crosswell example: inverted velocity after 21 LBFGS iterations with 3-20 Hz bandpass data by (a) FWI and (b) source-receiver extended waveform inversion ($\varepsilon = 0$). Initial velocity is $v_0 = 2$ km/s in both cases.
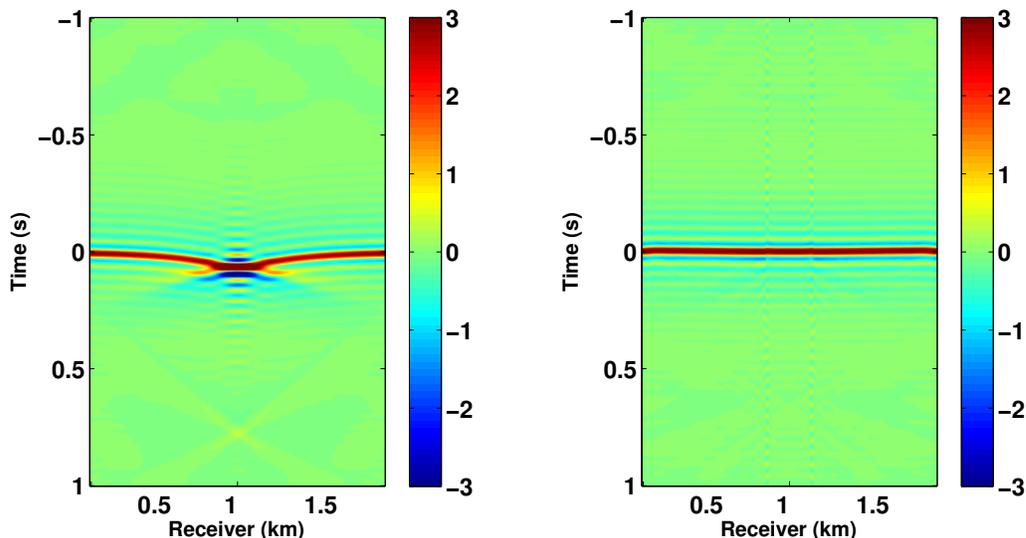
Figure 4: First crosswell example: extended sources for central shot $z_s = 1$ km using (a) initial velocity, and (b) inverted velocity by source-receiver extended waveform inversion

The misfit of traveltimes in the initial data is even more severe than was the case in the first example, and FWI indeed fails as we expect (Figure 8a). However, source-receiver extended waveform inversion without regularization also fails to produce a kinematically accurate velocity after 100 LBFGS iterations (Figure 8b), for the reasons explained in the theory section.

To illustrate the mechanism of failure, we simulated data in Figure 9a-9c with the trial velocity $v_t = 0.8v + 0.2v_0$, which is close to the target velocity model. We can see there are still quite obvious triplications present in the simulated data. Taking a single trace $z_r = 0.55$ km (Figure 10a) for the center shot gather at $z_s = 1$ km for example, we plot the spectrum of the normal operator $\bar{S}^T \bar{S}$, which is the same as the power spectrum of the Green's function, in Figure 10b. For traces with multiple energetic arrivals, the spectrum oscillates and almost vanishes at several frequencies, suggesting the existence of an effective numerical null space of the normal operator.

To illustrate the effect of regularization, we plot the extended source, estimated by solving equation 24 with the trial velocity v_t, for regularization parameters $\varepsilon = 10^{-6}, 0.5, 5$ in Figure 11a-11i. Oscillatory and non-decaying traces are

present in the extended source function $\bar{f}[v_t]$ for $\varepsilon = 10^{-6}$, but are suppressed by increasing ε.

The objective function for different values of $\varepsilon = 10^{-6}, 0.5, 5$ on a line segment in model space, between the initial and target velocities, appears in Figures 12a,12b, and 12c. Parameter $\alpha = 0$ corresponds to the target velocity, $\alpha = 1$ to the initial velocity. These plots show that the region of convexity of source-receiver extended waveform inversion objecitive is quite small for small ε, however expands to include the initial velocity for large enough ε.
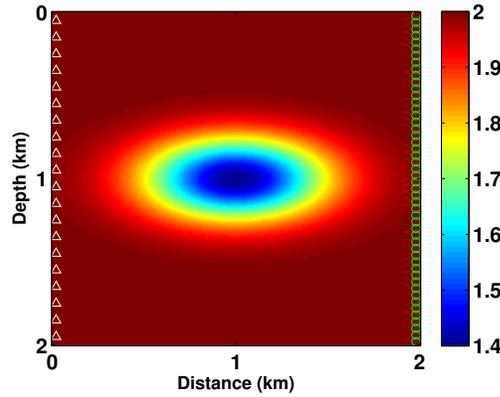

Figure 5: Second crosswell example: target velocity model with slow Gaussian anomaly. Lowest velocity is 1.4 km/s.
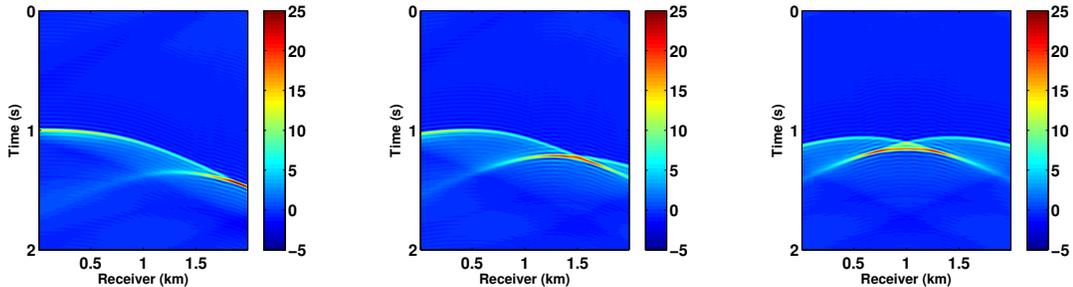

Figure 6: Second crosswell example: shot gathers of recorded data for shot at (a) $z_s = 0.1$ km, (b) $z_s = 0.5$ km and $z_s = 1$ km, respectively

As mentioned before, the arrival time error in the initial 3-20 Hz data (Figure 7a-7c) is too large to permit successful FWI, starting with $v_0 = 2$ km/s (Figure 8a). We find that $\varepsilon = 5$ satisfies the discrepancy criterion articularted in the theory section. With this choice, f ifteen LBFGS iterations produce the velocity estimate in
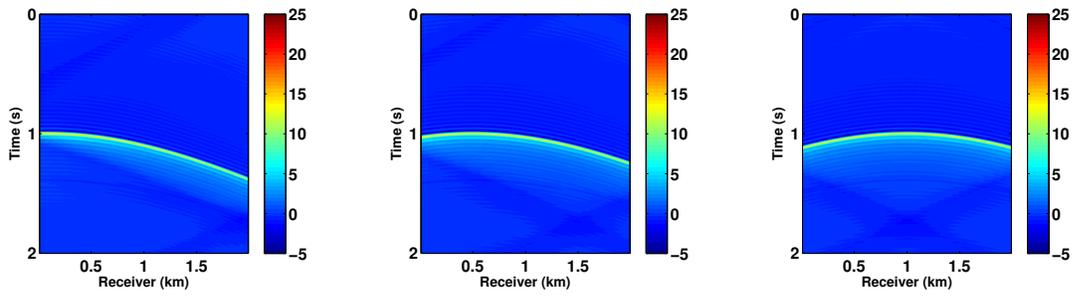
Figure 7: Second crosswell example: shot gathers of simulated data using initial veloctiy v_0 for shot at (a) $z_s = 0.1$ km, (b) $z_s = 0.5$ km and $z_s = 1$ km, respectively
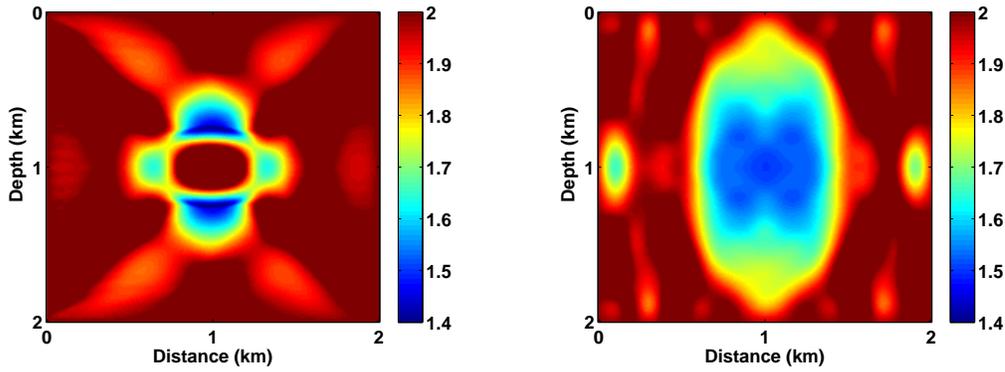

Figure 8: Second crosswell example: inverted velocity after 100 LBFGS iterations with 3-20 Hz data by (a) FWI and (b) unregularized source-receiver extended waveform inversion ($\varepsilon = 0$). Initial velocity is $v_0 = 2$ km/s in both cases.
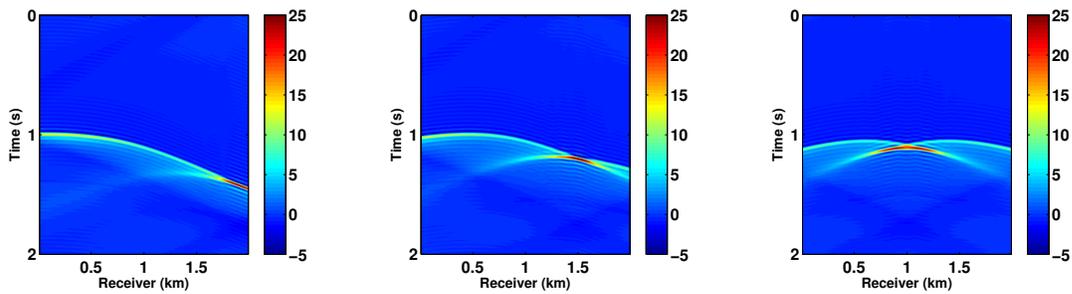

Figure 9: Second crosswell example: shot gathers of simulated data using trial velocity $v_t = 0.8v + 0.2v_0$ for shot at (a) $z_s = 0.1$ km, (b) $z_s = 0.5$ km and $z_s = 1$ km, respectively
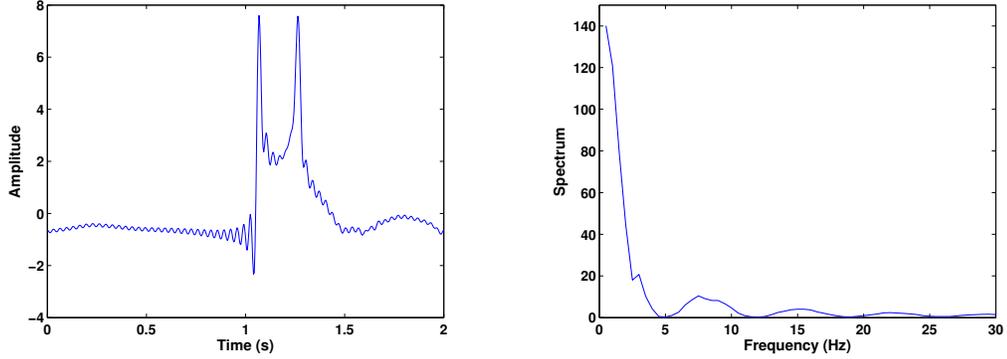
Figure 10: Second crosswell example: (a) data trace at $z_r = 0.55$ km for shot at $z_s = 1$ km and (b) spectrum of normal operator $\bar{S}^T\bar{S}$ of trace (a)

Figure 13a. Both position and shape of Gaussian anormaly are well resolved. Additionally, the extended source functions are almost focused on the zero-lag time after regularized source-receiver extended waveform inversion (Figure 14a-14c), indicating that the final velocity estimate is kinematically accurate. In view of its kinematic accuracy, the final estimate from source-receiver extended waveform inversion should be a usable initial estimate for FWI: indeed, 25 LBFGS iterations of full-bandwidth FWI produces a quite accurate inversion (Figure 13b).

Surface Acquisition Geometry

In the remainder of this section, we apply regularized source-receiver extended waveform inversion to waveform inversion to two examples with surface acquisition geometry, that is, both sources and receivers separated by a hyperplane from the scattering region.

the relative initial data residual e_0 and relative extended initial data residual $\bar{e}_0$:

$$e_0 = \frac{\|S[v_0]f_0 - d\|}{\|d\|}, \qquad \bar{e}_0 = \frac{\|\bar{S}[v_0]\bar{f}_0 - d\|}{\|d\|}$$

in which f_0 is the known source wavelet and $\bar{f}_0$ is the extended source function, estimated by solving equation 24 with the initial velocity. We adjusted ε by trial-and-error until the ratio $\bar{e}_0/e_0$ lay in the range $0.2 - 0.5$. The (somewhat arbitrary)
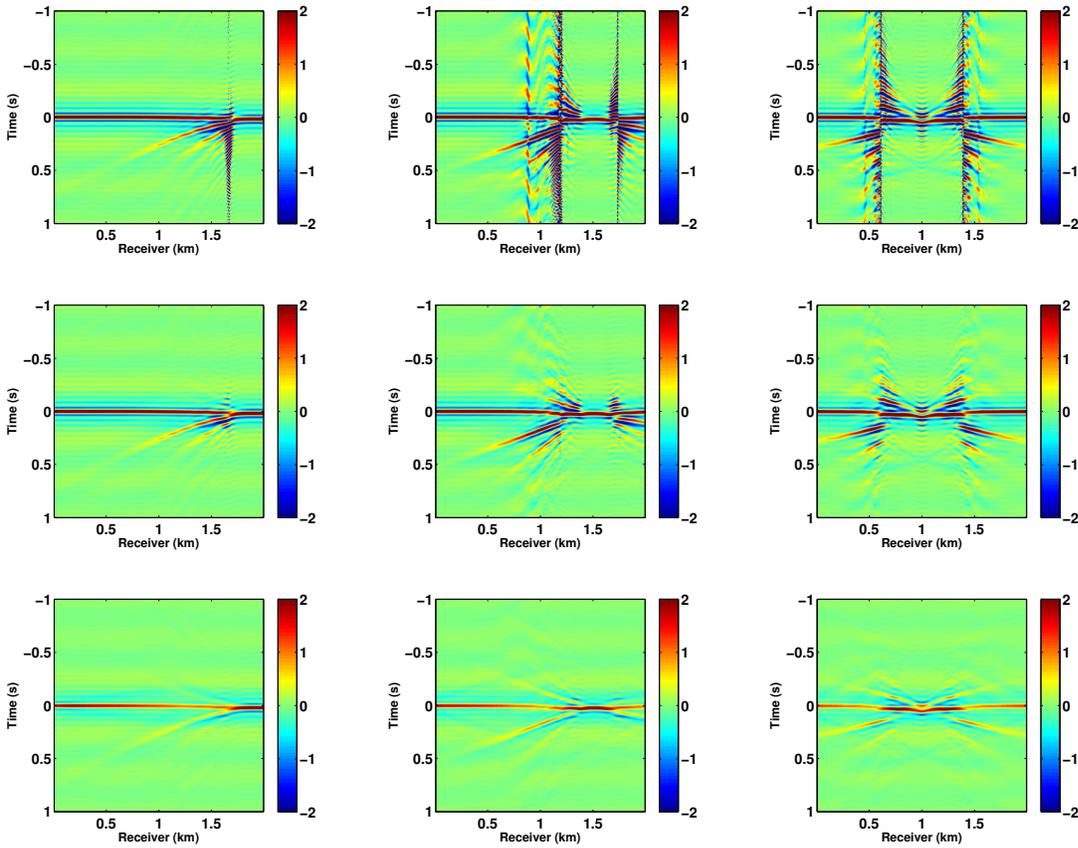
Figure 11: Second crosswell example: plots of regularized extended source functions for three regularization parameters $\varepsilon = 10^{-6}, 0.5, 5$ for shot at $z_s = 0.1, 0.5, 1$ km, respectively. (a) $\varepsilon = 10^{-6}, z_s = 0.1$ km; (b) $\varepsilon = 10^{-6}, z_s = 0.5$ km; (c) $\varepsilon = 10^{-6}, z_s = 1$ km; (d) $\varepsilon = 0.5, z_s = 0.1$ km; (e) $\varepsilon = 0.5, z_s = 0.5$ km; (f) $\varepsilon = 0.5, z_s = 1$ km; (g) $\varepsilon = 5, z_s = 0.1$ km; (h) $\varepsilon = 5, z_s = 0.5$ km; (i) $\varepsilon = 5, z_s = 1$ km;
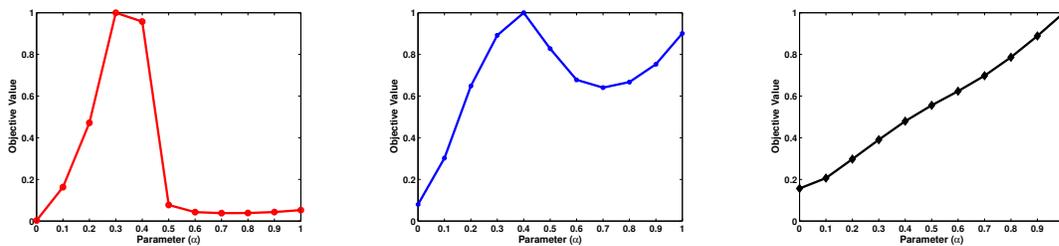

Figure 12: Second crosswell example: objective function evaluated at velocities $(1-\alpha)v + \alpha v_0$, $0 \leq \alpha \leq 1$, for various choices of regularization parameter ε: (a) $\varepsilon = 10^{-6}$, (b) $\varepsilon = 0.5$, (c) $\varepsilon = 5$
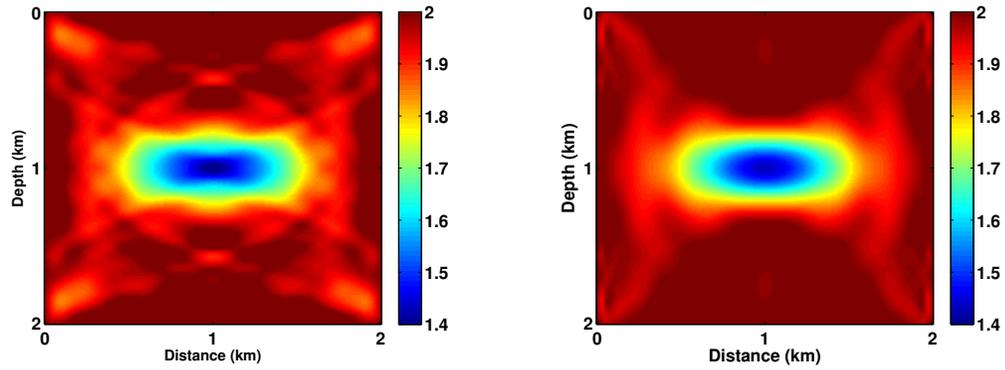
Figure 13: Second crosswell example: (a) inverted velocity by regularized source-receiver extended waveform inversion ($\varepsilon = 5$) after 15 iterations; (b) FWI result after 25 iterations using the regularized source-receiver extended waveform inversion inversion in (a) as initial velocity
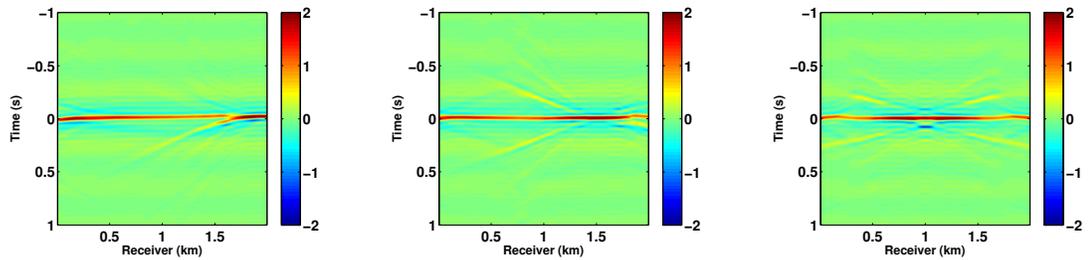

Figure 14: Second crosswell example: extended sources after regularized source-receiver extended waveform inversion inversion for shot at (a) $z_s = 0.1$ km, (b) $z_s = 0.5$ km and $z_s = 1$ km

bounds are chosen to ensure that ε is large enough that the residual is substantially larger than zero (the expected value for $\varepsilon = 0$), but small enough that the data is substantially better fit than is possible with a physical source ($\varepsilon = \infty$). This method is admittedly crude: we will return to the choice of ε in the discussion section.

Diving Wave Inversion

The predominant contemporary use of FWI is to invert diving wave energy (Virieux and Operto, 2009; Vigh et al., 2013). This example examines the use of source-receiver extended waveform inversion, with and without regularization, for a model generating diving waves with triplications. The model is smooth on the wavelength scale, hence transparent: the data consists only of direct and diving waves, with no reflections.

The target model (Figure 15a) consists of a low velocity Gaussian anomaly embedded in linearly increasing background velocity. 100 receivers are placed at depth $z_r = 0.04$ km from $x_r = 0.04$ km to $x_r = 7.96$ km with $\Delta x_r = 0.08$ km. 67 shots are placed at depth $z_s = 0.08$ km from $x_s = 0.04$ km to $x_r = 7.96$ km with $\Delta x_s = 0.12$ km. The frequency band used in inversion is 5-11 Hz. The choice $\epsilon = 10$ gives $\bar{e}_0 = 6.45\%$ and $e_0 = 20.95\%$, satisfying the criterion explained earlier.

The initial model in the iterative inversion is the linearly increasing background (Figure 15b), which produces diving wave arrivals without triplication.

Comparison of the data for target and initial data in Figures 16a-16c and Figure 16d-16f shows that as expected, first arrival times differ by well over a cycle, and of course the triplication structure does not appear in the initial data at all, Therefore one would expect FWI to stagnate far from a useful model estimate, as indeed happens (Figure 17a). Unregularized source-receiver extended waveform inversion ($\varepsilon = 10^{-6}$) also fails: this is a pure transmission problem, and precisely the same phenomenon occurs as in the strong lens crosswell example (Figure 17b). On the other hand, regularized source-receiver extended waveform inversion with the choice of penalty weight $\epsilon = 10$ explained above produces an satisfactory model estimation in the same number of iterations (Figure 18). Examination of extended sources at initial and final regularized source-receiver extended waveform inversion model suggests that the kinematics of the data have been adequately captured in the final model (Figures 19a, 19b, 19c, 19d, 19e, and 19f).

The output of regularized source-receiver extended waveform inversion also performs well as an initial model for FWI, as it has already matched the data arrival times. FWI with 50 iterations of LBFGS, starting at the model shown in Figure 18, produces the slightly more refined model shown in Figure 20. The simulated data (Figure 21a-21c) generated by the final inverted results show the best match with the recorded data in Figure 16a-16c.
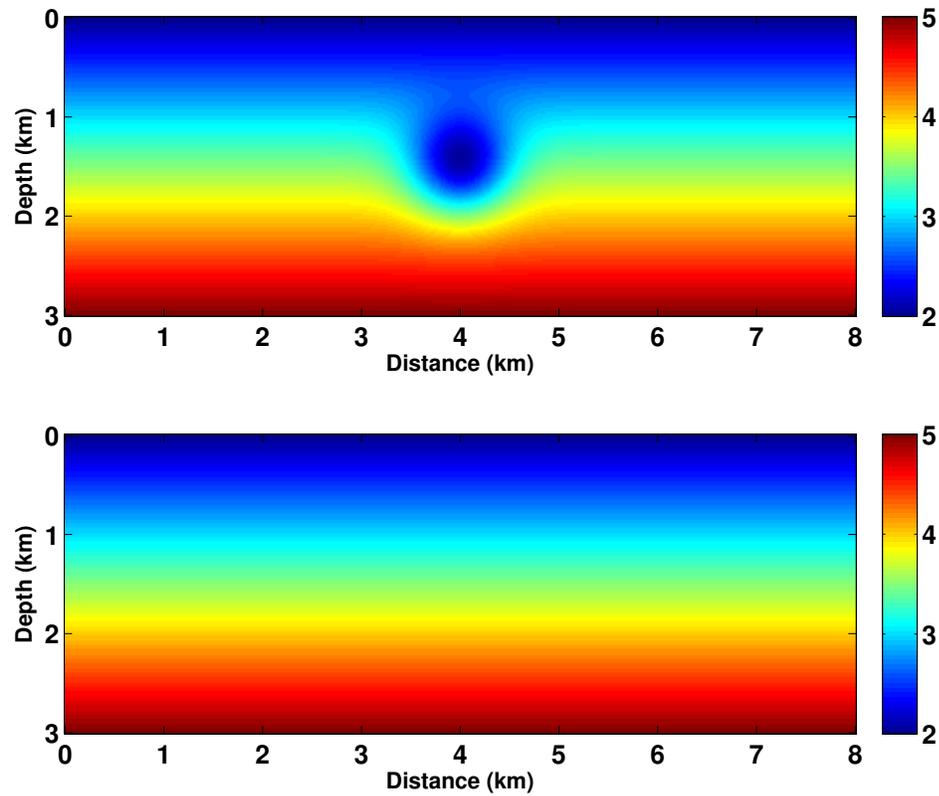

Figure 15: Diving wave example: (a) target velocity model with Gaussian low velocity lens embedded in the linearly increasing background velocity and (b) $v_0(z)$ linearly increasing initial model

Pluto

We modify a portion of the Pluto model (Stoughton et al., 2001) to create the target model in Figure 22a, mimicking an isolated salt pillow, gridded with a cell of 0.01
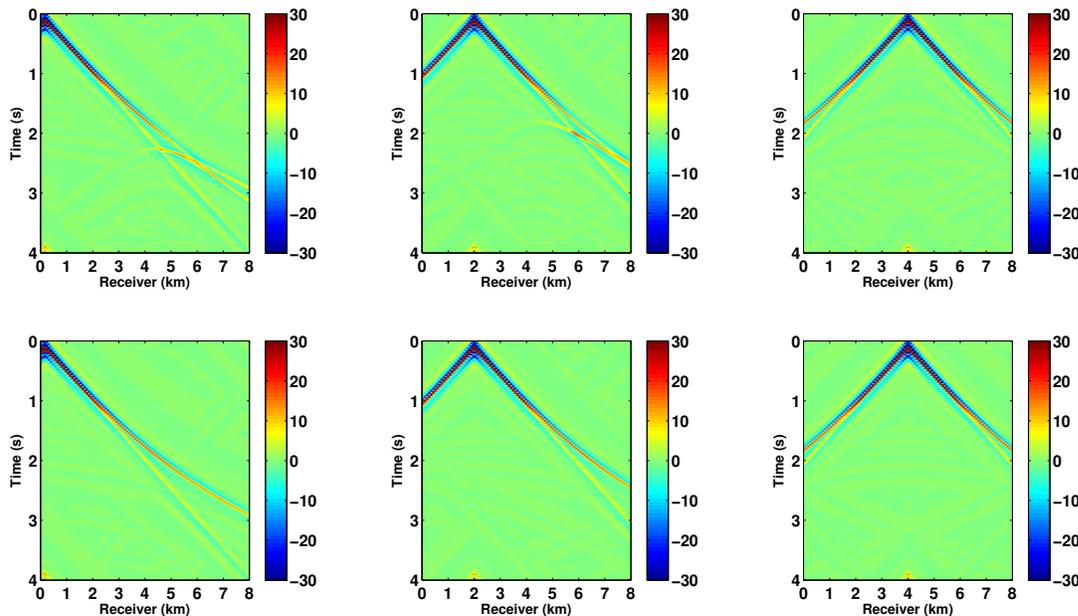
Figure 16: Diving wave example: shot gathers of recorded data for shot at at (a) $x_s = 0.2$ km, (b) $x_s = 2$ km, and (c) $x_s = 4$ km; Shot gathers of simulated data by initial velocity for shot at at (d) $x_s = 0.2$ km, (e) $x_s = 2$ km, and (f) $x_s = 4$ km

km × 0.01 km. The velocity in the "salt" inclusion is 4.5 km/s; the background medium is layer-like with velocity averaging 2.2 km/s. Source depth is $z_s = 0.02$ km. Sources range from $x_s = 0.06$ km to $x_s = 2.94$ km. Fixed spread receivers range from $x_r = 0.02$ km to $x_r = 2.98$ km placed at depth $z_r = 0.04$ km. The maximum time in the recorded data is 4 seconds and the frequency band is 4-10 Hz. We choose the regularization parameter $\epsilon = 1$, for which initial relative errors are $\bar{e}_0 = 3.2\%$ and $e_0 = 6.6\%$, respectively.

Note that in this example, positions of the sources and receivers permit "undershooting", that is, transmitted and reflected ray paths that transit the region under the inclusion. Therefore the data should contain adequate kinematic information to determine the velocity throughout the model, except for the poorly illuminated edges. On the other hand, the large velocity contrast between the "salt" and surrounding "sediments" implies that FWI will likely fail to reconstruct the inclusion from an initial model (Figure 22b) in which it is absent. Indeed, 200 FWI LBFGS iterations method locates the inclusion top, mispositions the bottom, and grossly underestimates the velocity in between (Figure 23).
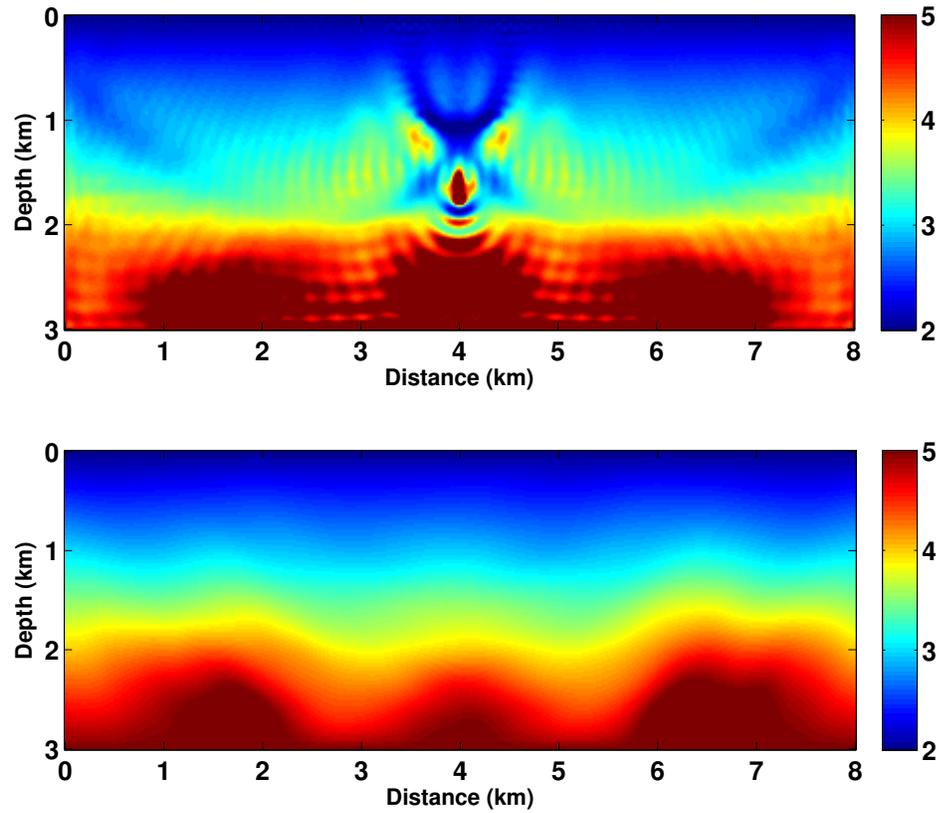
Figure 17: Diving wave example: inverted velocity after 200 iterations using 5-11 Hz data by (a) FWI and (b) unregularized source-receiver extended waveform inversion ($\varepsilon = 10^{-6}$)
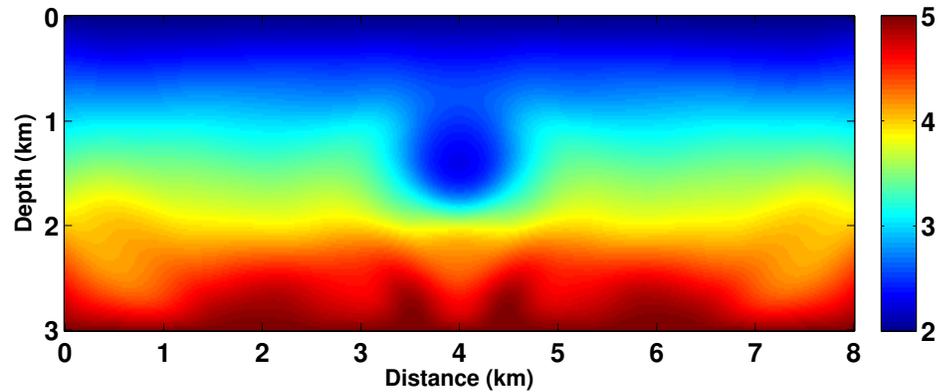
Figure 18: Diving wave example: inverted velocity using 5-11 Hz data after 200 iterations by regularized source-receiver extended waveform inversion ($\varepsilon = 10$)
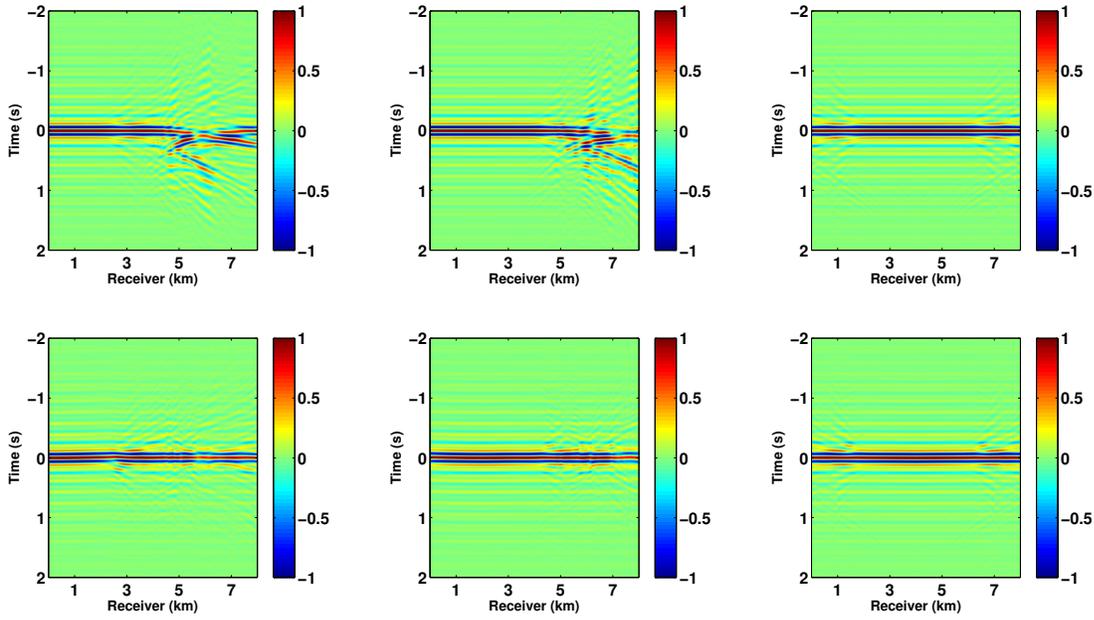
Figure 19: Diving wave example: top row: extended sources by initial model for shot at (a) $x_s = 0.2$ km, (b) $x_s = 2$ km, and (c) $x_s = 4$ km; Bottom row: extended sources by inverted velocity after regularized source-receiver extended waveform inversion inversion for shot at (d) $x_s = 0.2$ km, (e) $x_s = 2$ km, and (f) $x_s = 4$ km
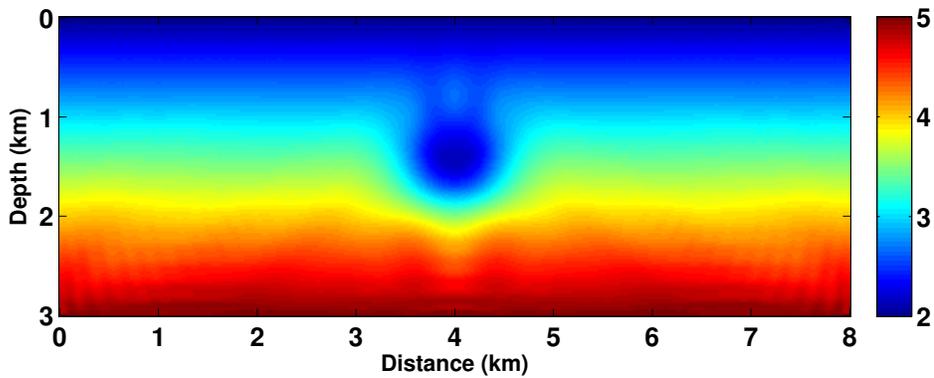
Figure 20: Diving wave example: inverted velocity by 50 FWI iterations, beginning with following regularized source-receiver extended waveform inversion velocity (Figure 18)
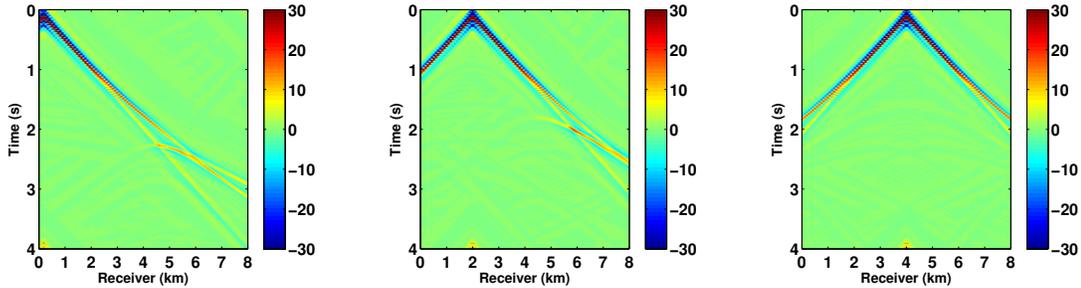
Figure 21: Diving wave example: shot gathers of simulated data with final source-receiver extended waveform inversion+FWI inverted result (Figure 20) for shot at (a) $x_s = 0.2$ km, (b) $x_s = 2$ km, and (c) $x_s = 4$ km, respectively

Regularized source-receiver extended waveform inversion, in contrast, comes much closer to locating both top and bottom of the inclusion, and filling it with approximately correct velocity values, Figure 24a. Focus of the extended source estimates at zero time lag is much improved (Figure 25d-25f) over the initial model (Figure 25a-25c). In fact, the model depicted in Figure 24a appears to be a reasonable initial guess for FWI, providing approximately correct kinematics. A further 50 LBFGS iterations of FWI with frequency band 4-16 Hz data results in an accurate reconstruction of the inclusion (Figure 24b).

The final data residual in Figure 25d-25f implies the inverted model has fit especially the refracted energy in the data quite well, in comparison to the initial data residual, shown in Figure 26a-26c.

DISCUSSION

We have already mentioned in the Introduction that the source-receiver extension used here is the basis for a number of other algorithms, for example Adaptive Waveform Inversion (Warner and Guasch, 2014, 2016). Other source extension concepts have also been productive. Waveform Reconstruction Inversion as described by van Leeuwen and Herrmann (2013) is roughly equivalent to introducing artificial sources "everywhere". See (Huang and Symes, 2016a; Wang and Yingst, 2016) for discussion of other related algorithms. Contrast Source Inversion, described for example by Abubakar et al. (2011), may also be regarded as
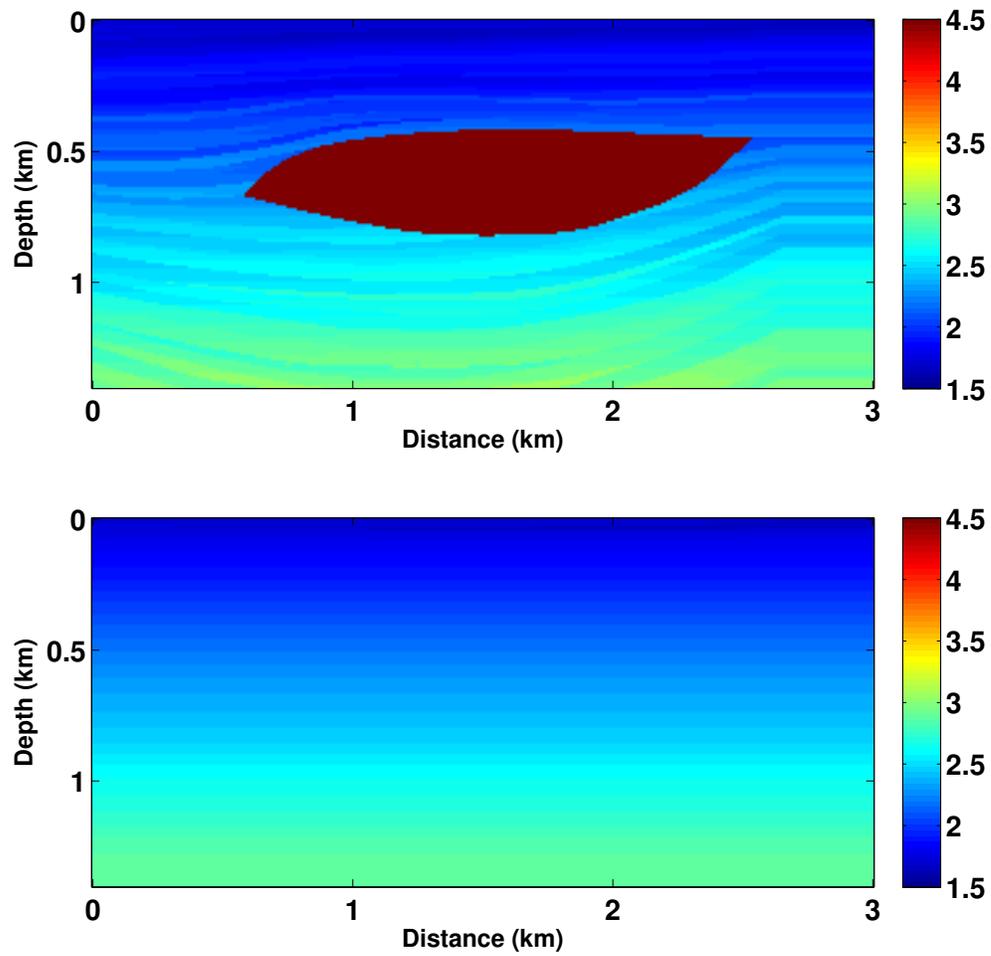
Figure 22: Pluto example: (a) target velocity and (b) 1D initial velocity
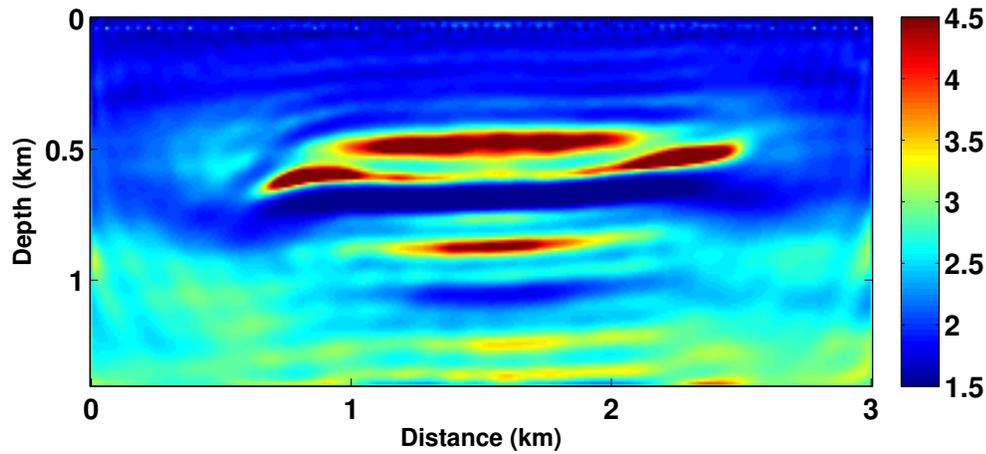
Figure 23: Pluto example: inverted reuslt by FWI method after 200 iterations using 4-10Hz data

a source extension approach. Other extended modeling modifications of FWI have been motivated by wave equation migration velocity analysis ("WEMVA", (Biondi and Sava, 2004)). These WEMVA-like extensions add parameters to the velocity model itself, for example subsurface offset (space shift) (Shen et al., 2003, 2005; Khoury et al., 2006; Shen and Symes, 2008; Shen, 2012; Biondi and Almomin, 2012; Weibull and Arntsen, 2014; Lameloise et al., 2015; Shen and Symes, 2015), time shift (Yang and Sava, 2011; Biondi and Almomin, 2014), scattering angle (De Hoop et al., 2003; Shen and Calandra, 2005), shot coordinates (Symes and Carazzone, 1991; Kern and Symes, 1994; Chauris and Plessix, 2013; Sun and Symes, 2012), and surface offset (Mulder and ten Kroode, 2002; Chauris and Noble, 2001). The common feature in all of these extension based modifications to FWI is their tendency to produce the same sort of long-wavelength velocity updates as does traveltime tomography, that is, to extract kinematic information from the data. We have shown here that source-receiver extended inversion accomplishes this goal in the special case of single arrivals. Similar computations show that subsurface offset extended waveform inversion and reflection slope tomography have proportional Hessians at a global solution (Symes, 2014; ten Kroode, 2014).

We have addressed a major difficulty in source-receiver extended waveform inversion, its tendency to develop apparent multimodality in the event that energy arrives in the data along multiple ray paths. This phenomenon presents a real impediment to using source-receiver extended waveform inversion for cross-
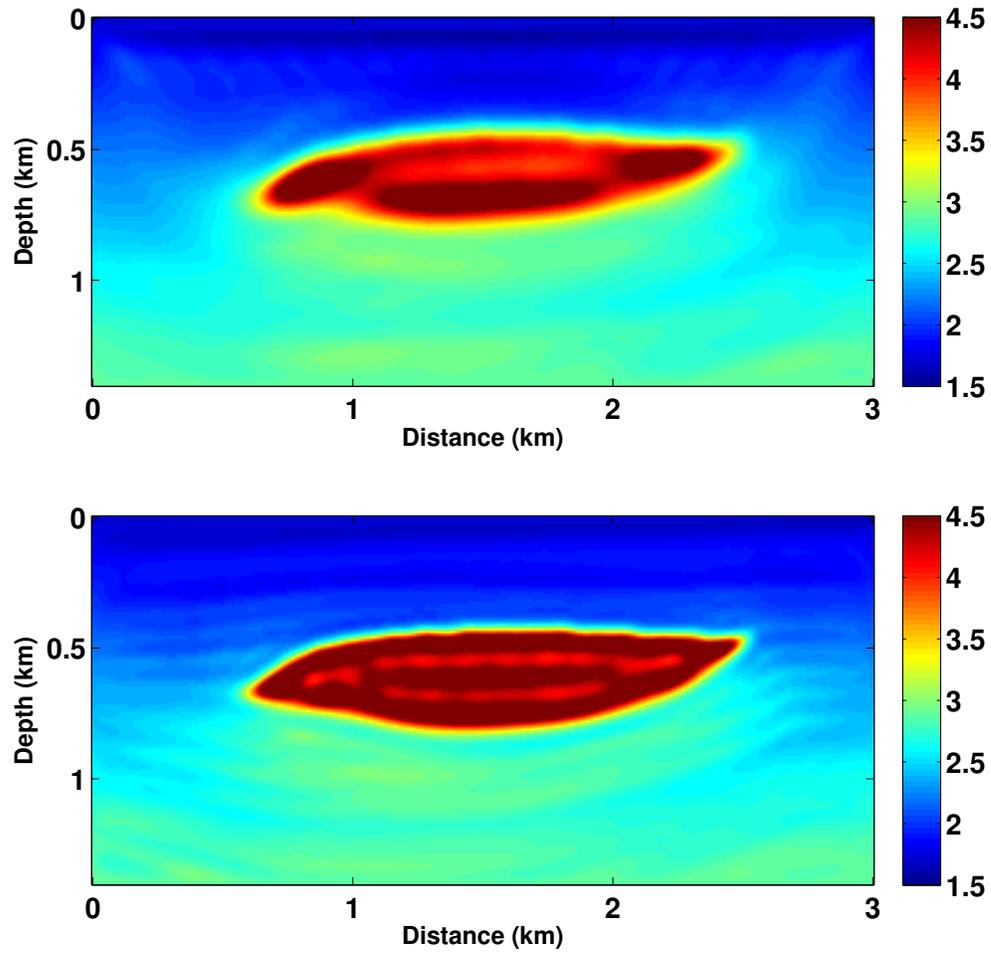
Figure 24: Pluto example: (a) inverted result produced by regularized source-receiver extended waveform inversion after 200 iterations using 4-10 Hz data; (b) FWI result after 50 iterations using 4-16Hz data with regularized source-receiver extended waveform inversion velocity in (a) as initial estimate.
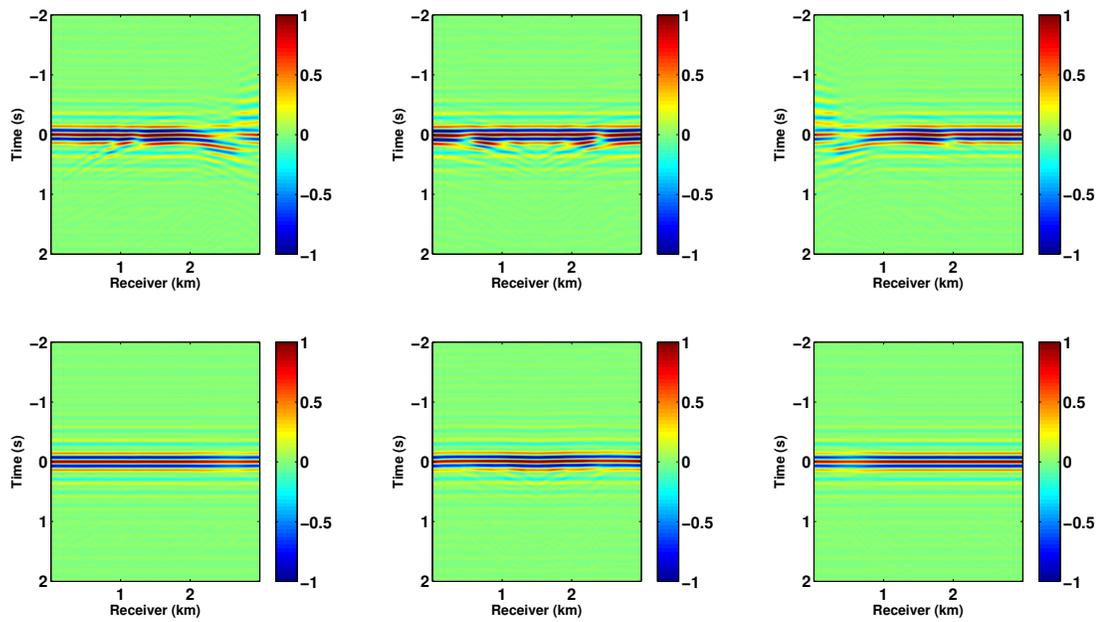
Figure 25: Pluto example: Top row: regularized extended source functions with initial velocity for shot at (a) $x_s = 0.06$ km, (b) $x_s = 1.5$ km, and (c) $x_s = 2.94$ km; Bottom row: regularized extended source functions with regularized source-receiver extended waveform inversion velocity for shot at (d) $x_s = 0.06$ km, (e) $x_s = 1.5$ km, and (f) $x_s = 2.94$ km
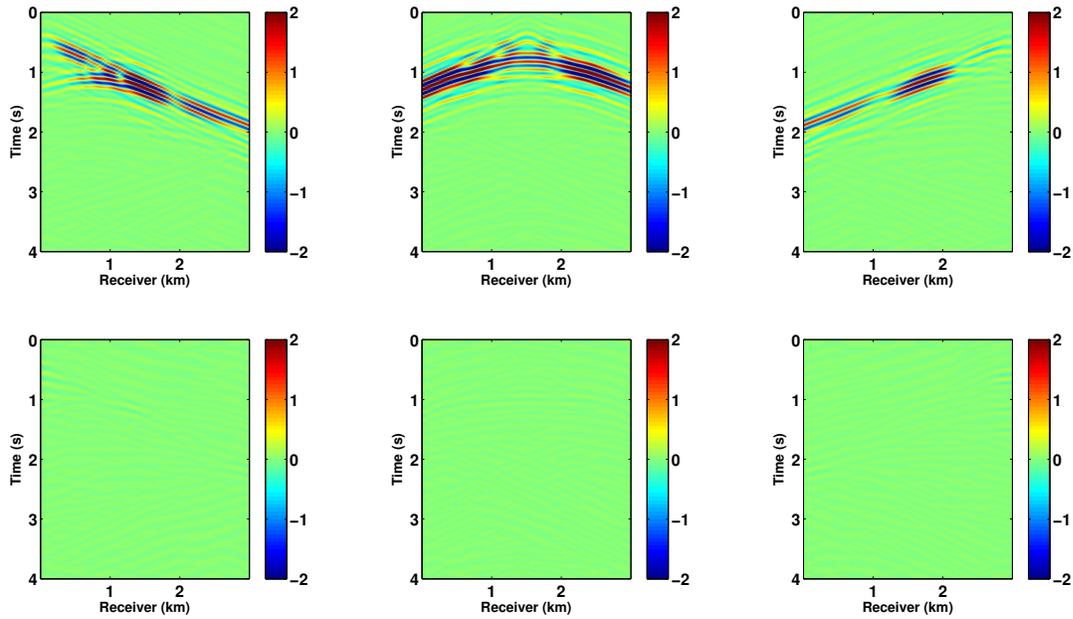
Figure 26: Pluto example: (a) Data residual between recorded data and data simulated with initial velocity for shot at (a) $x_s = 0.06$ km, (b) $x_s = 1.5$ km, and (b) $x_s = 2.94$ km; (b) Data residual between recorded data and simulated data using source-receiver extended waveform inversion + FWI velocity (Figure 24b) for shot at (d) $x_s = 0.06$ km, (e) $x_s = 1.5$ km, and (f) $x_s = 2.94$ km

well tomography, for example, since waveguides are quite commonly encountered in that application. Multiple arrivals are also common in diving wave fields, and can be present for many other reasons. We have advanced an explanation in the context of periodic convolutional modeling, namely the presence of spectral zeros and near-zeros of the Hessian, and partly recovered convexity by Tikhonov regularization, a standard technique for selecting solutions of ill-posed problems. Our justification for rather heavy-handed regularization is only heuristic, but does seem to be effective in many settings, including cartoon examples of crosswell tomography and diving wave FWI, as the examples show.

Other extension methods mentioned above show strong parallels with the observations made here. In particular, the failure of convexity caused by multiple arrivals echoes the similar failure of the surface extension in the presence of source wavefield caustics (Nolan and Symes, 1996; Stolk and Symes, 2004; Symes, 2008). This better-analyzed pathology arises from the impossibility of distinguishing arrivals by midpoint slowness in individual offset gathers, for instance. A consequence is the failure of the inner problem to have a unique solution. Just so, arrival slowness cannot be inferred from a single trace, so the source-receiver extension does not permit separation of arrivals by slowness, and consequently the inverse $(\bar{S}[v])^{-1}$ does not in general exist. For medium based extensions, the remedy is to use so-called subsurface offset instead of surface offset, leading to a well-posed inner problem and consequently smooth and stable reduced objective (Shen et al., 2003, 2005; Khoury et al., 2006; Shen and Symes, 2008; Stolk et al., 2009; Symes, 2014; ten Kroode, 2014). Other source extensions, in contrast to source-receiver, allow similar conclusions to be drawn, and lead to extended inversion algorithms that avoid cycle-skipping, maintain data fit, and do not require strong regularization (Huang and Symes, 2016b).

In the examples presented here, we have used a very large number of LBFGS iterations (hundreds in several examples). The need for so many iterations may be linked to the rather small amount of data used (an almost all cases, 5 - 10 frequencies), which of course also makes the iterations rather inexpensive. We believe that such computational largesse is appropriate for a study designed to explore the source-receiver extended waveform inversion concept. However it certainly begs the question: can reasonable results be obtained in a more reasonable number of RTM applications, say $O(10)$?

Our use of circulant convolution to model the relation between source and data traces is computationally convenient. In particular, it enables inexpensive and machine-precision solution of the normal equation 24 as the matrices in-

volved are diagonal. As our examples are both computed and inverted entirely in the frequency domain, this is an appropriate methodology. However this methodology is commonly termed an "inverse crime" - that is, the same tools used for modeling as for inversion. Field data comes in the form of time domain traces, with finite duration and not necessarily amplitude-decaying. A time cutoff is necessary, either implicitly or explicitly, and that operation does not commute with convolution. Therefore an appropriate version of problem 24 applicable to field data will need a different solution mode, either Gaussian elimination or the Levinson algorithm, or (more likely) an iterative solver such as conjugate gradient iteration.

An inexact solve of the inner problem via an iterative method would bring to the fore another difficulty. The velocity gradient (equation 28) involves reverse-time migration, the transpose operator of the derivative of the modeling operator $S[v]$ with respect to velocity. $S[v]$ amounts to convolution with the Green's function $G[v]$. In the simplest (single arrival) case, ignoring amplitudes, $G[v] = \delta(t - \tau[v]))$. The derivative of $S[v]$ with respect to v is the convolution with the derivative of $G[v]$ with respect to v: from the chain rule, the directional derivative in the direction δv in velocity model space is

$$DG[v]\delta v = -\frac{d\delta}{dt}(t - \tau[v])D\tau[v]\delta v. \tag{31}$$

That is, the derivative of $\bar{S}[v]$ involves convolution with a multiple of $d\delta/dt$ (in this simple case), that is, a shift of the derivative of the input trace, whereas $\bar{S}[v]$ shifts the (undifferentiated) input trace. The Fourier transforms of the two results differ by a factor of frequency. The same is true of the migration operator, that is, the transpose of $D\bar{S}[v]$, and the conclusion remains true in the high-frequency asymptotic sense with the proper amplitudes and without the single-arrival assumption. Iteratively computing the extended source $\bar{f}[v]$ (solution of equation 24) and auxiliary field $\bar{w}$ (solution of equation 26), using (for example) the conjugate gradient algorithm, results in small mean-square errors, as follows from standard theory (Golub and van Loan, 2012; Nocedal and Wright, 1999). However the derivative of an RMS-small error trace is not necessarily RMS-small: it can be as large by a factor of the highest frequency present with significant energy. Therefore the convergence of the gradient obtained by iterative approximation of the expression in equation 28 can be arbitrarily slower (with sufficient bandwidth) than the convergence of the extended source. See (Huang, 2016) for explicit illustration of this effect using a different extension, and (Symes, 2015) for recent theoretical progress in using special accelerators for iterative solution of the inner problem

(equations 24, 26) to accelerate convergence of the gradient.

Note that the use of circulant convolution and consequent machine-precision solution of the inner problem by regularized deconvolution eliminates the source of outer gradient inaccuracy just described.

The sharp-eyed reader familiar with the Warner's AWI algorithm (Warner and Guasch, 2014, 2016) will notice that our definition of the objective function (equation 13) lacks the normalization by the mean square of $\bar{f}$ employed in AWI. There are excellent reasons to normalize such an objective. The principal reason is the amplitude trade-off that occurs in reflection: data fit to a reflection is not affected, to first order, by scaling the reflectivity up and the source down by the same factor. Therefore one would not expect an effective velocity update in the reflection case without normalization of the source. The reader might well ask how we got away with it. The answer is that transmitted waves do not suffer from this scale ambiguity, and all of our examples contained transmitted waves. In particular, the data in the reflection-dominated example (Pluto) contained direct wave energy, as the boundary conditions used here are absorbing in all directions so the dipole effect of the free surface is absent. Clearly this device is not one to rely upon, and a modification of our algorithm to normalize J following the model of Warner and Guasch (2014, 2016) (and many previous works on joint source-model estimation, such as (Minkoff and Symes, 1997)) is indicated. As pointed out by Warner and Guasch (2014, 2016), the additional computational expense from normalization is minimal.

CONCLUSION

We have presented an extended modeling approach to overcome the tendency of FWI to stagnate at uninformative models. The key ingredient in this approach is the addition of parameters to the model, in the form of unphysical source parameters, which allow the model to fit the data at every stage of the inversion. Specifically, we allow the source pulse to depend on the source and receiver coordinates. Other choices of source extension are possible, see below. To eliminate the unphysical additional parameters and recover an FWI solution, we have imposed a penalty on the variance of the extended source across source and receiver. Since the extended source is uniquely determined by the data and the velocity model, the penalty is a function of the velocity.

We have analyzed the use of this penalty as an inversion objective function.

We find that in the simplest case, single-arrival transmission data, this objecive is closely related to mean square traveltime error, so that its minimization yields a tomographic-quality result, even starting from initial models that would grossly cycle-skipped for straightforward FWI. For more complex data with multiple arrivals, the link to traveltime inversion is much more tenuous, and strong regularization is required to produce global convergence. In either case, an iterative local optimization method applied to source-receiver extended waveform inversion produces a sequence of models approaching a kinematically accurate model, even starting with a grossly inaccurate initial velocity estimate. The approach is close enough that convergent FWI iteration can be started from an source-receiver extended waveform inversion solution. Thus in a sense the combination source-receiver extended waveform inversion + FWI appears to globalize the convergence of FWI, at least in some cases.

ACKNOWLEDGEMENT

APPENDIX: COMPUTATION OF GRADIENT

In this section, we show how to compute the gradient of inversion velocity analysis objective function via trace-based extension. The directional derivative of the objective function J with respect to velocity v in the direction (velocity perturbation) δv is given by

$$DJ[v]\delta v = \langle A^T A \bar{f}[v], D\bar{f}[v]\delta v \rangle \tag{A-1}$$

Using the normal equation 24, we have

$$
\begin{aligned}
(\bar{S}[v]^T \bar{S}[v] + \varepsilon^2 I) D\bar{f}[v]\delta v &= (D\bar{S}[v]\delta v)^T d - D(\bar{S}[v]^T \bar{S}[v] + \varepsilon^2 I)\bar{f}[v] \\
&= (D\bar{S}[v]\delta v)^T (d - \bar{S}[v]\bar{f}[v]) - \bar{S}[v]^T (D\bar{S}[v]\delta v)\bar{f}[v] \tag{A-2}
\end{aligned}
$$

Let's introduce an auxiliary extended source $\bar{w}[v](\mathbf{x}_r, t; x_s)$ such that

$$(\bar{S}[v]^T \bar{S}[v] + \varepsilon^2 I)\bar{w}[v] = A^T A \bar{f}[v]. \tag{A-3}$$

Then, we have by combining (A-2)-(A-3)

$$
\begin{aligned}
DJ[v]\delta v &= \langle \bar{w}[v], (D\bar{S}[v]\delta v)^T (d - \bar{S}[v]\bar{f}[v]) - \bar{S}[v]^T (D\bar{S}[v]\delta v)\bar{f}[v] \rangle \tag{A-4} \\
&= \langle D\bar{S}[v]\bar{w}[v], d - \bar{S}[v]\bar{f}[v] \rangle - \langle \bar{S}[v]\bar{w}[v], (D\bar{S}[v]\delta v)\bar{f}[v] \rangle \tag{A-5}
\end{aligned}
$$

Note that $(D\bar{S}[v]\delta v)\bar{f}[v](\mathbf{x}_r, t; \mathbf{x}_s) = \bar{f}[v]*(D\bar{S}[v]\delta v)\delta_t(\mathbf{x}_r, t; \mathbf{x}_s)$, and that $\langle \bar{S}[v]\bar{w}[v], \bar{f}[v]*(D\bar{S}[v]\delta v)\delta_t \rangle = \langle \bar{f}[v] \star \bar{S}[v]\bar{w}[v], (D\bar{S}[v]\delta v)\delta_t \rangle$ where "$\star$" denotes cross-correlation and δ_t is the delta function of t for every trace. A similar transformation of the other term above shows that

$$DJ[v]\delta v = \langle \bar{r}[v], (D\bar{S}[v]\delta v)\delta_t \rangle \tag{A-6}$$

where the "residual" $\bar{r}[v](\mathbf{x}_r, t, \mathbf{x}_s)$ is:

$$\bar{r}[v] = \bar{w}[v] \star (d - \bar{S}[v]\bar{f}[v]) - \bar{f}[v] \star \bar{S}[v]\bar{w}[v] \tag{A-7}$$

(note that this is a trace-by-trace computation). Therefore

$$\nabla J[v] = (D\bar{S}[v](\cdot)\delta_t)^T \bar{r}[v] \tag{A-8}$$

The transpose of the impulsive Born simulation operator $D\bar{S}[v](\cdot)\delta_t$ is the well-known impulsive reverse time migration operator, that is the zero-lag cross-correlation between the incident Green function G and the adjoint field with the residual $\bar{r}[v]$ as the adjoint source.

REFERENCES

Abubakar, A., G. Pan, M. Li, L. Zhang, T. Habashy, and P. van den Berg, 2011, Three-dimensional seismic full-waveform inversion using the finite-difference contrast source inversion method: Geophysical Prospecting, **59**, 874–888.

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, Expanded Abstracts, Society of Exploration Geophysicists, SI9.5.

——, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Biondi, B., and P. Sava, 2004, Wave-equation migration velocity analysis - I: Theory, and II: Subsalt imaging examples: Geophysics, **52**, 593–623.

Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Chauris, H., and M. Noble, 2001, Two-dimensional velocity macro model estimation from seismic reflection data by local differential semblance optimization: applications synthetic and real data sets: Geophysical Journal International, **144**, 14–26.

Chauris, H., and R.-E. Plessix, 2013, Differential waveform inversion - a way to cope with multiples?: 75th Conference EAGE, European Association for Geoscientists and Engineers, WS06–F06. (Expanded Abstract).

De Hoop, M. V., S. Brandsberg-Dahl, and B. Ursin, 2003, Seismic velocity analysis in the scattering-angle/azimuth domain: Geophysical Prospecting, **51**, 295–314.

Engl, H., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.

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

Golub, G. H., and C. F. van Loan, 2012, Matrix computations, 4th ed.: Johns Hopkins University Press.

Huang, G., and W. Symes, 2015a, Full waveform inversion via matched source extension: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1320–1325.

——, 2016a, Matched source waveform inversion: space-time extension: 86th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1426–1431.

——, 2016b, Matched source waveform inversion: volume extension: 86th Annual International Meeting, Expanded Abstracts, Society of Exploration Geo-

physicists, 1364–1368.

Huang, Y., 2016, Born waveform inversion in shot coordinate domain: PhD thesis, Rice University.

Huang, Y., and W. Symes, 2015b, Born waveform inversion via variable projection and shot record model extension: 85rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 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.

Khoury, A., W. Symes, P. Williamson, and P. Shen, 2006, DSR migration velocity analysis by differential semblance optimization: 76th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SPMI3.4.

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.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2788–2792.

Minkoff, S. E., and W. Symes, 1997, Full waveform inversion of marine reflection data in the plane-wave domain: Geophysics, **62**, 540–553.

Mulder, W., and F. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184–1191.

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

Nolan, C. J., and W. Symes, 1996, Imaging and conherency in complex structure: 66th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 359–363.

Plessix, R.-E., 2000, Automatic cross-well tomography: an application of the differential semblance optimization to two real examples: Geophysical Prospecting, **48**, 937–951.

Plessix, R.-E., Y.-H. de Roeck, and G. Chavent, 1999, Waveform inversion of reflection seismic data for kinematic parameters by local optimization: SIAM Journal on Scientific Computation, **20**, 1033–1052.

Plessix, R.-E., W. Mulder, and F. ten Kroode, 2000, Automatic cross-well tomography by semblance and differential semblance optimization: theory and gradient computation: Geophysical Prospecting, **48**, 913–935.

Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory, and verification in a physical scale model: Geophysics, **64**, 888–901.

Pratt, R., and R. Shipp, 1999, Seismic waveform inversion in the frequency do-

main, part 2: Fault delineation in sediments using crosshole data: Geophysics, **64**, 902–914.

Pratt, R., and W. Symes, 2002, Semblance and differential semblance optimization for waveform tomography: a frequency domain approach: Journal of Conference Abstracts, **7**, 183–4.

Shen, P., 2012, An RTM based automatic migration velocity analysis in image domain: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SVE2.2.

Shen, P., and H. Calandra, 2005, One-way waveform inversion within the framework of adjoint state differential migration: 75th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SI3.5.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

———, 2015, Horizontal contraction in image domain for velocity inversion: Geophysics, **80**, R95–R110.

Shen, P., W. Symes, S. Morton, and H. Calandra, 2005, Differential semblance velocity analysis via shot profile migration: 75th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SPVA1.4.

Shen, P., W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2135–2139.

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

Song, H., 1994, On a transmission inverse problem: PhD thesis, Rice University, Houston, Texas, U.S.A.

Song, H., and W. Symes, 1993, A differential semblance formulation of crosswell tomography: Technical Report TR92-22, Department of Computational and Applied Mathematics, Rice University, Houston, TX.

———, 1994, Inversion of crosswell tomography via differential semblance optimization: Technical report, The Rice Inversion Project. (Annual Report).

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

Stolk, C. C., and W. Symes, 2004, Kinematic artifacts in prestack depth migration: Geophysics, **69**, 562–575.

Stoughton, D., J. Stefani, and S. Michell, 2001, 2-D elastic model for wavefield investigations of subsalt objectives, deep water Gulf of Mexico: 71th Annual International Meetting, SEG, Expanded Abstacts, 1269–1272.

Sun, D., and W. Symes, 2012, Waveform inversion via nonlinear differential sem-

blance optimization: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SI3.3.

Symes, W., 1994, Inverting waveforms in the presence of caustics: Technical report, The Rice Inversion Project. (Annual Report).

——, 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

——, 2014, Seismic inverse problems: recent developments in theory and practice: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 2–5.

——, 2015, Algorithmic aspects of extended waveform inversion: 76th Annual International Conference and Exhibition, Expanded Abstract, European Association for Geoscientists and Engineers, WS05–A01.

Symes, W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

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

ten Kroode, F., 2014, A Lie group associated to seismic velocity estimation: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 142–146.

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.

Vigh, D., K. Jiao, W. Huang, N. Moldoveanu, and J. Kapoor, 2013, Long-offset-aided Full-waveform Inversion: 75th Annual Meeting, Expanded Abstracts, European Association of Geoscientists and Engineers, 76th Annual Meeting.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152.

Wang, C., and D. Yingst, 2016, Full-waveform inversion with the reconstructed wavefield method: 86th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, FWI 5.4.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1089–1093.

——, 2016, Adaptive waveform inversion: theory: Geophysics, **81**, R429–R445.

Weibull, W., and B. Arntsen, 2014, Anisotropic migration velocity analysis using reverse time migration: Geophysics, **79**, R13–R25.

Yang, T., and P. Sava, 2011, Wave-equation migration velocity analysis with time-shift imaging: Geophysical prospecting, **59**, 635–650.

Gradients of Reduced Objective Functions for Separable Wave Inverse Problems

William W. Symes, *The Rice Inversion Project*

ABSTRACT

The derivative of the forward (modeling) map for wave inverse problems is unbounded in the natural metric of the model and data spaces. This unboundedness complicates the accurate computation of the reduced gradient in separable wave inverse problems: iterative estimation of the linear variables may not lead to convergent computation of the gradient. If the linear part of the modeling operator possesses a computable inverse up to a smoothing error, however, and if the data is high-frequency in a precise sense, then the gradient may be decomposed into a computable non-iterative part and a convergent remainder. The resulting form of the gradient error control is well-suited to integration with inexact trust-region optimization methods.

INTRODUCTION

Separable (partly linear) modeling operators for wave inverse problems arise either via Born approximation (Kern and Symes, 1994; Chauris and Noble, 2001; Mulder and ten Kroode, 2002; Shen and Symes, 2008; Weibull and Arntsen, 2013; Biondi and Almomin, 2014) or by including source (right-hand side) parameters as model degrees of freedom (Song and Symes, 1994; Plessix, 2000; Pratt and Symes, 2002; Luo and Sava, 2011; Warner and Guasch, 2014; van Leeuwen and Herrmann, 2013). Extended model spaces add yet more degrees of freedom to the linear parameters of a separable modeling operator, to permit data fit for a large set of nonlinear parameter ("velocity model") values. The non-extended ("physical") subspace is the kernel of an "annihilator". Minimizing the norm of the annihilator output over all data-fitting models picks out an approximation to the solution of the physical least squares inverse problem. A natural approach to this

minimization is via elimination of the linear variables to create a reduced function of the nonlinear variables (Kern and Symes, 1994; van Leeuwen and Mulder, 2009; Li et al., 2013). If the nonlinear parameters include velocities, indices of refraction, or equivalent, then the modeling operator is differentiable with loss of regularity: a convergent sequence of approximate solutions of the data-fit problem does not (necessarily) lead to a convergent sequence of derivative values. Consequently, the computable approximate solutions of the data-fit problem do not lead to convergent gradient approximations for the reduced objective, which may reduce the effectiveness of iterative optimization strategies (Huang and Symes, 2015).

In the paragraphs to follow, I show that the gradient computation may be modified to become convergent if three conditions are satisfied:

- the linear part of the forward modeling operator possesses a computable parametrix (inverse modulo compacts, or "asymptotic inverse") in the form of its adjoint with respect to suitable weighted norms, and

- the inverse problem is embedded in a family of problems indexed by a data frequency parameter.

- the (inevitable) Tihonov regularization parameter is coupled to data frequency.

Then the reduced gradient computation may be modified to become convergent as well.

Reliance on parametrices might seem surprising. However, parametrices for linearized inverse wave problems have a long history, mostly as a subtopic of computational geometric optics (Cohen and Bleistein, 1977; Beylkin, 1985; Beylkin and Burridge, 1990; Virieux et al., 1992; Araya and Symes, 1996; Operto et al., 2000). Recently several authors have described direct ("wave equation") parametrix constructions without any ray-geometric computations (Zhang et al., 2005, 2007; Xu et al., 2011; ten Kroode, 2012; Hou and Symes, 2015). For inversion of source parameters, even the usual "exact" inversion by Fourier division is actually a parametrix, due to the inevitable presence of time cut-off.

The existence of the parametrix permits the decomposition of the gradient into two summands: the first does not involve the solution of the inner (linear)

problem at all, while the second does but is manifestly continuous in the linear variables. Hence approximation of the linear part of the solution by a convergent iteration results in a convergent gradient approximation.

While convergence is assured by the three conditions above, properly interpreted, control over the actual error in the gradient is indirect, depending on quantities that are difficult or impossible ot estimate accurately. However I also show that gradient error control of the type established here is compatible with so-called inexact trust-region optimization algorithms (Heinkenschloss and Vicente, 2001), with (in principle) assured convergence to a stationary point. That is, control over gradient error must be linked to convergence metrics of the optimization algorithm.

Some uses of model extension for inverse wave problems have been criticized for so-called "gradient artifacts", that is, computed gradients that appear to give rise to slow convergence or undesired model features (Fei and Williamson, 2010; Vyas and Tang, 2010). However it now appears that nothing was wrong with the gradients - instead, the reduced objective functions of separable least squares problems were defined with insufficiently precise solution of the linear (inner) inverse problem (Kern and Symes, 1994; Liu et al., 2014; Lameloise et al., 2015). The present paper could be viewed as an elaboration upon the latter work.

OPERATORS

This is a Hilbert space story. In fact, the natural objects in this study are *Hilbert space scales*, that is sequences of Hilbert spaces $H = \{H^s, s \in \mathbf{Z}\}$, decreasing in the sense that $H^{s+1} \subset H^s, s \in \mathbf{Z}$. Define the sets

$$H^\infty = \cap_{s \in \mathbf{Z}} H^s, H^{-\infty} = \cup_{s \in \mathbf{Z}} H^s.$$

The inner product in H^s will be denoted $\langle \cdot, \cdot \rangle_s$, and the corresponding norm is $\|\cdot\|_s$. Usually several scales will be involved in most of our assertions; the scale to which the inner product or norm belongs should be clear from context.

An *operator L of order* $k \in \mathbf{Z}$ from a scale H_1 to a scale H_2 is linear function from $H_1^{-\infty}$ to $H_2^{-\infty}$ for which $L|_{H_1^s} \in \mathcal{B}(H_1^s, H_2^{s-k})$ for all $s \in \mathbf{Z}$. Denote by $Op^k(H_1, H_2)$ a set of operators of order k, satisfying the additional order properties: for $L_i \in$

$Op_1^{k_i}(H_1, H_2), i = 1, 2,$

$$L_1 L_2 \in Op^{k_1+k_2}(H_1, H_2), [L_1, L_2] \in Op^{k_1+k_2-1}(H_1, H_2). \tag{1}$$

$Inv^k(H_1, H_2) \subset Op^k(H_1, H_2)$ is the subset of invertible operators of order k.

The motivating example of this structure are scales of L^2 Sobolev spaces, and scalar (pseudo-)differential operators on them.

The domain of a separable extended model consists of two components. The first is a (background) model Hilbert space M_b and an open set of admissible models $U \subset M_b$. In all examples, M_b is a space of (possibly) vector-valued) smooth (C^∞, "low frequency") functions on a suitable physical domain representing the physical parameters of a wave dynamics model, and members of U obey additional constraints (such as bounds) required to make the dynamical laws they parametrize well-posed.

The other component of the model domain is a Hilbert space scale of extended models $\bar{M} = \{\bar{M}_k, k \in \mathbf{Z}\}$. A distinguished subscale $M = \{M_k, k \in \mathbf{Z}\}, M_k \subset \bar{M}_k$, represents the non-extended or physical models. The *annihilator* $A \in Op^0(\bar{M}, N)$ has M as its kernel:

$$\bar{m} \in \bar{M}, \ A\bar{m} = 0 \Leftrightarrow \bar{m} \in M. \tag{2}$$

. The *extension map* $E \in Op^0(M, \bar{M})$ is injective.

The model range or data space is another Hilbert scale D.

The extended modeling operator, or forward map, is a continuous function

$$\bar{F} : U \to Op^0(\bar{M}, D)$$

It is *smooth with loss of regularity*:

$$m_b \mapsto \bar{F}[m_b]|_{\bar{M}^k} \in C^p(U, \mathcal{B}(\bar{M}^k, D^{k-p})), \tag{3}$$

but

$$m_b \mapsto \bar{F}[m_b]|_{\bar{M}^k} \notin C^p(U, \mathcal{B}(\bar{M}^k, D^{k-p+1})), \tag{4}$$

and in particular

$$m_b \mapsto \bar{F}[m_b]|_{\bar{M}^0} \notin C^p(U, \mathcal{B}(\bar{M}^0, D^0)). \tag{5}$$

However, the normal operator is smooth

$$F^T F \in C^\infty(U, Op^0(\bar{M}, \bar{M}), \tag{6}$$

as follows from the *factorization lemma*:

$$D\bar{F}[m_b]\delta m_b = F[m_b](Q_1[m_b]\delta m_b) \tag{7}$$

in which $Q_1 \in C^\infty(U \times M_b, Op^1(\bar{M}, \bar{M}))$ is *essentially skew-symmetric*:

$$Q_1 + Q_1^T = Q_0 \in C^\infty(U \times M_b, Op^0(\bar{M}, \bar{M})). \tag{8}$$

Finally, assume that $\bar{F}$ has the *Egorov property*: if $L_1 \in C^\infty(U, Op^k(\bar{M}, \bar{M}))$, then there exists $L_2 \in C^\infty(U, Op^k(D, D))$ so that for every $m_b \in U$,

$$\bar{F}[m_b]L_1[m_b] = L_2[m_b]\bar{F}[m_b]. \tag{9}$$

and vis-versa.

Remark: In the examples, $\bar{F}[m_b]$ is a Fourier Integral Operator whose canonical relation is a local canonical graph, and whose symbol and canonical relation depends smoothly on m_b. Both the regularity with loss of a derivative (Blazek et al., 2013) and normal regularity of $\bar{F}^T\bar{F}$ follow from these facts: in this case, $\bar{F}^T\bar{F}$ is a pseudodifferential operator. For examples of the factorization lemma, see ten Kroode (2014); Symes (2014). The Egorov property follows from the usual Egorov's theorem (see for example Taylor (1981)).

The physical modeling operator is a similar map

$$F : U \to Op^0(M, D)$$

with the same smoothness with loss of regularity.

The extended and physical modeling operators are related through the extension map: for each $m_b \in U$,

$$F[m_b] = \bar{F}[m_b] \circ E.$$

OBJECTIVE

In the notation just introduced, the separable inverse problem of this paper may be crudely stated as:

$$\text{Given } d \in D^0, \text{ find } m \in M_b, m_l \in M^0 \text{ so that } F[m]m_l \approx d \tag{10}$$

In view of the extension structure, this problem is equivalent to

$$\text{Given } d \in D^0, \text{ find } m \in M_b, \bar{m} \in \bar{M}^0 \text{ so that } \bar{F}[m]\bar{m} \approx d, A\bar{m} \approx 0 \qquad (11)$$

For consistent data (that is, for which "$\approx$" can be replaced with "=" in either of the above statements, a solution is also a global minimizer of both

$$\|\bar{F}[m]\bar{m} - d\|^2 \text{ and } \|A\bar{m}\|^2 \qquad (12)$$

The fundamental difficulty of this type of problem lies in the mapping properties of $m \mapsto F[m]$ or $\bar{F}[m]$: because of these maps are differentiable only with loss of regularity (statements 3, 4, and 5), the first of the two functions in display 12 (the "data residual") is not smooth as a map $M_b \times \bar{M} \to \mathbf{R}$. This lack of smoothness is the ultimate reason that solution of the basic "full waveform inversion" problem 10 is still a topic of active research, rather than a routine tool in exploration and academic seismology, more than thirty years after it was introduced.

Two objectives such as 12 can be combined in various ways to produce constrained and unconstrained optimization problems. Both are quadratic in $\bar{m}$, so minimization of either in this variable seems a tractable problem. Minimizing the second ("semblance") function so that the data residual becomes a function of m only (a so called reduced objective) does not improve matters: the resulting unconstrained optimization is equivalent to 10 and has a non-smooth objective.

Remarkably, eliminating $\bar{m}$ by minimizing the data residual, yields a smooth reduced semblance objective: this conclusion will be justified below. Generally it cannot be assumed that $\bar{F}$ is coercive, even if it is injective. Therefore some form of regularization must be added to the data residual to guarantee existence of a stable minimizer. It is also turns out to be essential to permit more freedom in the metric structure of domain and range spaces also.

Introduce a data weighting operator $W_d \in C^\infty(U, Inv^0(D, D))$. When restricted to D^0, W_d is self-adjoint and positive definite, hence induces a norm on the order 0 data space:

$$\langle d_1, d_2 \rangle_d = \langle d_1, W_d d_2 \rangle_0 \text{ for } d_1, d_2 \in D^0 \qquad (13)$$

Similarly, introduce a model weight operator $W_m \in C^\infty(U, Inv^0(\bar{M}, \bar{M}))$ on the model space, self-adjoint and positive definite when restricted to $\bar{M}^0$, with corresponding norm and inner product

$$\langle \bar{m}_1, \bar{m}_2 \rangle_m = \langle \bar{m}_1, W_d \bar{m}_2 \rangle_0 \text{ for } \bar{m}_1, \bar{m}_2 \in \bar{M}^0 \qquad (14)$$

Reformulate data residual minimization for the linear variables $\bar{m}$ as

$$\min_{\bar{m}} \|\bar{F}[m]\bar{m} - d\|_d^2 + \lambda^2 \|R\bar{m}\|_m^2 \tag{15}$$

in which R is a bounded and coercive regularization operator, whose range is some other Hilbert space. The normal equation is

$$N_\lambda[m]\bar{m} = \bar{F}^\dagger[m]d, \ N_\lambda[m] = \bar{F}^\dagger[m]\bar{F}[m] + \lambda^2 R^\dagger R, \tag{16}$$

in which $\bar{F}^\dagger = W_m^{-1}\bar{F}[m]^T W_d[m]$ is the adjoint of $\bar{F}$ with respect to the weighted norms and $\bar{F}^T$ is the adjoint with respect to the scale 0-norms. Similarly $R^\dagger$ is the adjoint of R with respect to the weighted model-space norm.

Denote by $\bar{m}_\lambda \in C^\infty(U \times D, \bar{M}^0)$ the solution of 7, and define the reduced objective function

$$J_\lambda[m,d] = \frac{1}{2}\|A\bar{m}_\lambda[m,d]\|_0^2 \tag{17}$$

If $J_\lambda[m,d]$ attains its obvious lower bound, then $\bar{m}_\lambda[m,d] \in \bar{M}^0$ and $(m, \bar{m}_\lambda[m,d])$ solves the extended inverse problem 11.

GRADIENT

Minimization of J_λ by any variant of Newton's method requires computation of the gradient of J_λ. Formally,

$$DJ_\lambda[m,d]\delta m = \langle D\bar{m}_\lambda[m,d]\delta m, A^T A\bar{m}_\lambda[m,d]\rangle_0$$

$$= \langle -N_\lambda[m]^{-1}(DN_\lambda[m]\delta m)\bar{m}_\lambda[m,d] + N_\lambda[m]^{-1}(D\bar{F}^\dagger[m]\delta m)d, A^T A\bar{m}_\lambda[m,d]\rangle_0$$

Inserting $W_m W_m^{-1}$, referring to the definition of $\langle \cdot, \cdot \rangle_m$, and noting that $W_m^{-1}A^T = A^\dagger$ and that N_λ is self-adjoint in the model norm, obtain

$$= \langle -(DN_\lambda[m]\delta m)\bar{m}_\lambda[m,d] + (D\bar{F}^\dagger[m]\delta m)d, \bar{q}_\lambda[m,d]\rangle_m \tag{18}$$

where $\bar{q}_\lambda[m,d]$ is the solution of

$$N_\lambda[m]\bar{q}_\lambda[m,d] = A^\dagger A\bar{m}_\lambda[m]. \tag{19}$$

The first term in right-hand side of equation 18 defines a continuous function of $\bar{m}_\lambda[m,d]$. Since $N_\lambda = W_m^{-1}\bar{F}^T W_d\bar{F} + \lambda^2 R^\dagger R$, the rule 9 implies the existence of

$P \in C^\infty(U, Inv^0(\bar{M}, \bar{M}))$ so that $N_\lambda = W_m^{-1} \bar{F}^T \bar{F} P + \lambda^2 R^\dagger R$. This shows in turn that $N_\lambda \in C^\infty(U, Op^0(\bar{M}, \bar{M}))$, thanks to the assumptions on W_d, W_m, 9 once again, and the 6. In particular, $DN_\lambda[m]\delta m \in Op^0(\bar{M}, \bar{M})$, so the bilinear form

$$\bar{q}, \bar{m} \mapsto \langle (DN_\lambda[m]\delta m)\bar{m}, \bar{q}\rangle_m \tag{20}$$

is continuous in $\bar{M}^0 \times \bar{M}^0$. Also,

$$\langle (DN_\lambda[m]\delta m)\bar{m}, \bar{q}\rangle_m = \langle (D(\bar{F}^\dagger[m])\delta m)\bar{F}[m]\bar{m}, \bar{q}\rangle_m$$

$$+\langle (\bar{F}^\dagger[m](D\bar{F}[m]\delta m)\bar{m}, \bar{q}\rangle_m = -\langle W_m[m]^{-1}(DW_m[m]\delta m)\bar{F}^\dagger[m]\bar{F}[m]\bar{m}, \bar{q}\rangle_m$$

$$+\langle \bar{F}^\dagger[m]W_d[m]^{-1}DW_d[m]\bar{F}[m]\bar{m}, \bar{q}\rangle_m + \langle (D\bar{F}[m]\delta m)^\dagger \bar{F}[m]\bar{m}, \bar{q}\rangle_m$$

$$+\langle \bar{F}^\dagger[m](D\bar{F}[m]\delta m)\bar{m}, \bar{q}\rangle_m$$

$$= -\langle (DW_m[m]\delta m)\bar{F}^\dagger[m]\bar{F}[m]\bar{m}, \bar{q}\rangle_0 + \langle (DW_d[m]\delta m)\bar{F}[m]\bar{m}, \bar{F}[m]q\rangle_0$$

$$+\langle \bar{F}[m]\bar{m}, (D\bar{F}[m]\delta m)\bar{q}\rangle_d + \langle (D\bar{F}[m]\delta m)\bar{m}, \bar{F}[m]\bar{q}\rangle_d \tag{21}$$

Concerning the second term in 18, note that

$$(D\bar{F}^\dagger[m]\delta m) = D[(W_m[m])^{-1}\bar{F}[m]^T W_d[m]d](\delta m)$$

$$= -W_m[m]^{-1}(DW_m[m]\delta m)\bar{F}^\dagger[m]d + \bar{F}^\dagger[m]W_d[m]^{-1}DW_d[m]d + (D\bar{F}[m]\delta m)^\dagger d \tag{22}$$

The first two terms in equation 22 involve only bounded operators on $\bar{M}^0$. The contribution to the second term in the right hand side of equation 18 is

$$\langle (D\bar{F}^\dagger[m]\delta m)d, \bar{q}\rangle_m =$$

$$-\langle W_m[m]^{-1}(DW_m[m]\delta m)\bar{F}^\dagger[m]d, \bar{q}\rangle_m + \langle \bar{F}^\dagger[m]W_d[m]^{-1}DW_d[m]d, \bar{q}\rangle_m$$

$$+\langle (D\bar{F}[m]\delta m)^\dagger d, \bar{q}\rangle_m$$

$$= -\langle (DW_m[m]\delta m)\delta m)\bar{F}^\dagger[m]d, \bar{q}\rangle_0 + +\langle (DW_d[m]\delta m)d, \bar{F}[m]\bar{q}\rangle_0$$

$$+\langle d, (D\bar{F}[m]\delta m)\bar{q}\rangle_d \tag{23}$$

Now the difficulty is plain to see: if $\bar{q} = \bar{q}_\lambda[m, d]$ in the last term on the right hand side of equation 23 is replaced with an approximation in the sense of the

0 norm, as would be the result of an iterative process for solving 19 (or equation 7), no bound on the resulting error in the inner product above, hence in the resulting approximation to $DJ_\lambda[m, d]\delta m$ or the gradient of J_λ, can be asserted, since $D\bar{F}[m]\delta m$ is not continuous in the sense of $\bar{M}^0$.

Ignoring for the moment the apparent instability, combine equations 21 and 23 to obtain an expression for the derivative of $J_\lambda[m, d]$:

$$DJ_\lambda[m, d]\delta m = \langle (DW_m[m]\delta m)\bar{F}^\dagger[m](\bar{F}[m]\bar{m}_\lambda[m, d] - d), \bar{q}_\lambda[m, d]\rangle_0$$

$$+\langle (DW_d[m]\delta m)(\bar{F}[m]\bar{m}_\lambda[m, d] - d), \bar{F}[m]\bar{q}_\lambda[m, d]\rangle_0$$

$$-\langle F[m]\bar{m}_\lambda[m, d] - d, (D\bar{F}[m]\delta m)\bar{q}_\lambda[m, d]\rangle_d - \langle (D\bar{F}[m]\delta m)\bar{m}_\lambda[m, d], F[m]\bar{q}_\lambda[m, d]\rangle_d \tag{24}$$

There follows an expression for the gradient, using the partial duals $D\bar{F}[m]^*$, a continuous quadratic form : $D^0 \times \bar{M}^1 \to M_b$ defined by

$$\langle \delta m, D\bar{F}[m]^*(d, \bar{m})\rangle_{M_b} = \langle d, (D\bar{F}[m]\delta m)\bar{m}\rangle_d = \langle (D\bar{F}[m]\delta m)^\dagger d, \bar{m}\rangle_m$$

$$= \langle W_d[m]^{-1}d, (D\bar{F}[m]\delta m)\bar{m}\rangle_0 \tag{25}$$

and $DW_m[m]^t$, a continuous bilinear map $\bar{M}^0 \times \bar{M}^0 \to M_b$ defined by

$$\langle \delta m, DW_m[m]^t(\bar{m}, \bar{q})\rangle_{M_b} = \langle (DW_m[m]\delta m)\bar{m}, \bar{q}\rangle_0$$

$DW_d[m]^t$ is defined similarly. Then the *m*-gradient of J_λ is given by

$$\nabla_m J_\lambda[m, d] = DW_m[m]^t(\bar{F}^\dagger[m](\bar{F}[m]\bar{m}_\lambda[m, d] - d), \bar{q}_\lambda[m, d])$$

$$+DW_d[m]^t(\bar{F}[m]\bar{m}_\lambda[m, d] - d), \bar{F}[m]\bar{q}_\lambda[m, d]$$

$$-D\bar{F}[m]^*(\bar{F}[m]\bar{m}_\lambda[m, d], \bar{q}_\lambda[m, d]) - D\bar{F}[m]^*(\bar{m}_\lambda[m, d], \bar{F}[m]\bar{q}_\lambda[m, d])$$

$$+D\bar{F}[m]^*(d, \bar{q}_\lambda[m, d]) \tag{26}$$

Every part of this expression is manifestly *stable*, that is, substitutions of approximations $\bar{m}_{\lambda,a}, \bar{q}_{\lambda,a}$ in their computation will entail $O(\|\bar{m}_\lambda - \bar{m}_{\lambda,a}\|_0, \|\bar{q}_\lambda - \bar{q}_{\lambda,a}\|_0)$ error, *except* for the last term

STABLILITY

In this section, I will show that the last term in 26 can expressed as the value at $\bar{m} = \bar{m}_\lambda[m, d]$ and $\bar{q} = \bar{q}_\lambda[m, d]$ of of continuous function of $d, \bar{m}, \bar{q}$, thus giving a stable computation of the gradient in the sense explained at the end of the last section.

The construction depends on the availability of a *parametrix* or approximate inverse modulo lower order error, of a particular form. The utility of the construction depends on the computability of the parametrix, that is, on its being a straightforward modification of the transpose. (Hou and Symes, 2015) demonstrated the existence of such special parametrices for a particular separable inverse problem (extended linearized constant density acoustic inversion with horizontal subsurface offset extension) and parametrices with similar properties exist for other extended modeling operators as well.

In fact, most parametrices of the type I've mentioned are only microlocal, and moreover can only be computed approximately. These limitations will be addressed in coming sections.

Assume that the data and model weight operators W_d, W_m can be chosen so that $\bar{F}$ is approximately unitary with respect to the norms $\|\cdot\|_d, \|\cdot\|_m$: satisfies

$$\bar{F}^\dagger \bar{F} - I \equiv S \in C^\infty(U, Op^{-1}(\bar{M}, \bar{M})). \tag{27}$$

Thus $N_\lambda = S + I + \lambda^2 R^\dagger R$). Thus equations 7, 19 are equivalent to

$$\bar{m}_\lambda = (I + \lambda^2 R^\dagger R)^{-1}(\bar{F}^\dagger d - S\bar{m}_\lambda) \tag{28}$$

$$\bar{q}_\lambda = (I + \lambda^2 R^\dagger R)^{-1}(A^\dagger A\bar{m}_\lambda - S\bar{q}_\lambda) \tag{29}$$

$$= (I + \lambda^2 R^\dagger R)^{-1}(A^\dagger A(I + \lambda^2 R^\dagger R)^{-1}(\bar{F}^\dagger d - S\bar{m}_\lambda) - (I + \lambda^2 R^\dagger R)^{-1}S\bar{q}_\lambda \tag{30}$$

Replacing $\bar{q}_\lambda$ in 26 with the right hand side in the last equality of 28, obtain

$$D\bar{F}[m]^*(d, \bar{q}_\lambda[m, d]) = D\bar{F}[m]^*(d, (I + \lambda^2 R^\dagger R)^{-1}A^\dagger A(I + \lambda^2 R^\dagger R)^{-1}\bar{F}^\dagger[m]d$$

$$-(I + \lambda^2 R^\dagger R)^{-1}A^\dagger A(I + \lambda^2 R^\dagger R)^{-1}S[m]\bar{m}_\lambda[m, d]$$

$$-(I + \lambda^2 R^\dagger R)^{-1}S[m]\bar{q}_\lambda[m, d]). \tag{31}$$

The operators appearing in the last two terms of 31 are of order -1, mapping $\bar{M}^0$ continuously to $\bar{M}^1$. So the continuity property of $D\bar{F}[m]^*$ noted above implies

that the last two terms above are stable, that is, substitution of approximations $\bar{m}_{\lambda,a}$ and $\bar{q}_{\lambda,a}$ for $\bar{m}_\lambda$ and $\bar{q}_\lambda$ in the right-hand side of equation 31 will result in an $O(\|\bar{m}_{\lambda,a} - \bar{m}_\lambda\|_0, \|\bar{q}_{\lambda,a} - \bar{q}_\lambda\|_0)$ error.

It remains to be seen that the first term on the right hand side of equation 31 is stable: *a priori*, it only makes sense for $d \in D^1$ (hence $\bar{m}_\lambda \in \bar{M}^1$). A continuous extension to $d \in D^0$ follows however from the factorization property 7. For convenience, define

$$B = (1 + \lambda^2 R^\dagger R)^{-1} A^\dagger A (1 + \lambda^2 R^\dagger R)^{-1}.$$

Then the first term on the right-hand side of 31 is

$$\langle \delta m, D\bar{F}[m]^*(d, B\bar{F}^\dagger[m]d) \rangle_{M_b}$$

$$= \langle d, (D\bar{F}[m]\delta m) B\bar{F}^\dagger[m]d \rangle_d$$

$$= \langle d, \bar{F}[m](Q_1[m]\delta m) B\bar{F}^\dagger[m]d \rangle_d$$

$$= \langle \bar{F}^\dagger[m]d, (Q_1[m]\delta m) B\bar{F}^\dagger[m]d \rangle_m \tag{32}$$

The essentially-skew property of Q (equation 8) holds also for the weighted inner product $\langle \cdot, \cdot \rangle_m$, since the weight operator W_m is invertible and of order 0: specifically,

$$(Q_1[m]\delta m)^\dagger = -(Q_1[m]\delta m) + (Q_{0,m}[m]\delta m), \tag{33}$$

with $Q_{0,m} = Q_0 + W_m^{-1}[Q_1, W_m] \in C^\infty(U, Op^0(\bar{M}, \bar{M}))$. So the right-hand side of equation B-1 is

$$= \langle \bar{F}^\dagger[m]d, ([B, (Q_1[m]\delta m)] + B(Q_1[m]\delta m))\bar{F}^\dagger[m]d \rangle_m$$

$$+ \langle \bar{F}^\dagger[m]d, ([B, (Q_1[m]\delta m)] + ((Q_{0,m}[m] - Q_1[m])\delta m)B)\bar{F}^\dagger[m]d \rangle_m. \tag{34}$$

Add the right hand sides of equations B-1 and 34 and divide by 2 to obtain

$$\langle \delta m, D\bar{F}[m]^*(d, B\bar{F}^\dagger[m]d) \rangle_{M_b}$$

$$= \frac{1}{2} \langle \bar{F}^\dagger[m]d, ([B, (Q_1[m]\delta m)] + (Q_{0,m}[m]\delta m))\bar{F}^\dagger[m]d \rangle_m. \tag{35}$$

The right-hand side of equation 35 is a quadratic form in $\bar{F}^\dagger[m]d$, defined by a self-adjoint operator (the sum of the various commutators and products in the above expression) of order zero. Therefore

$$D\bar{F}[m]^*(d, B\bar{F}^\dagger[m]d)$$

is a $\|\cdot\|_0$-continuous M_b-valued quadratic form in d, hence extends continuously to $d \in D^0$, whence $\nabla_m J_\lambda[m, d]$ is $\|\cdot\|_0$-continuous in d.

For Hilbert spaces H_1, H_2, denote by $\mathcal{Q}(H_1, H_2)$ the Banach space of continuous H_2-valued quadratic forms on H_1. Then the result of the foregoing calculations is summarized in

Theorem 1. *Define $G_\lambda \in C^\infty(U, \mathcal{Q}(D^0 \times \bar{M}^0 \times \bar{M}^0, M_b))$ by*

$$
\begin{aligned}
G_\lambda[m](d, \bar{m}, \bar{q}) =\,& D\bar{F}[m]^*(d, (1 + \lambda^2 R^\dagger R)^{-1} A^\dagger A (1 + \lambda^2 R^\dagger R)^{-1} \bar{F}^\dagger[m]d \\
& - (1 + \lambda^2 R^\dagger R)^{-1} A^\dagger A (1 + \lambda^2 R^\dagger R)^{-1} S[m]\bar{m} \\
& - (1 + \lambda^2 R^\dagger R)^{-1} S[m]\bar{q}]) \\
& + DW_m[m]^t(\bar{F}^\dagger[m](\bar{F}[m]\bar{m} - d), \bar{q}) \\
& + DW_d[m]^t(\bar{F}[m]\bar{m} - d, \bar{F}[m]\bar{q}) \\
& - D\bar{F}[m]^*(\bar{F}[m]\bar{m}, \bar{q}) \\
& - D\bar{F}[m]^*(\bar{m}, \bar{F}[m]\bar{q})
\end{aligned}
\tag{36}
$$

Then

$$\nabla_m J_\lambda[m, d] = G_\lambda[m](d, \bar{m}_\lambda[m, d], \bar{q}_\lambda[m, d]), \tag{37}$$

in which $\bar{m}_\lambda[m, d]$ and $\bar{q}_\lambda[m, d]$ are solutions of equations 7 and 19 respectively.

Proof. The content of the definition 36 of G_λ is that it is a continuous quadratic-form-valued function. This fact has been established for each of the summands: for the first term by equation 35, for the second and third by equation 31 and following discussion, the fourth and fifth by equation 22, and the last two by equations 20 and 21 and surrounding discussion. These calculations also established equation 37.   □

Remark 2. *For computational purposes, it is more convenient to work directly with the defining Hilbert space structure of the background, domain and range spaces, rather*

than with the background, model and data norms, and to formulate the background model space norm also as a weighted norm. For example,

$$\langle \delta m, D\bar{F}[m]^*(d, A^\dagger A \bar{F}^\dagger[m]d)\rangle_{M_b}$$

$$= \langle d, (D\bar{F}[m]\delta m)A^\dagger A \bar{F}^\dagger[m]d\rangle_d$$

$$= \langle W_d d, (D\bar{F}[m]\delta m)W_m^{-1}A^T A \bar{F}^\dagger[m]d\rangle_0.$$

So

$$\nabla_m J_\lambda[m,d] = W_b^{-1}D\bar{F}[m]^t(W_d[m]d, W_m[m]^{-1}A^T A \bar{F}^\dagger[m]d) + K[m](\bar{m}_\lambda, \bar{q}_\lambda), \quad (38)$$

where W_b is the background model space weight operator defining the norm $\|\cdot\|_{M_b}$, and $D\bar{F}[m]^t$ is the transpose with respect to the 0-norms (usually, L^2, or in the discrete case Euclidean length) rather than the M_b and $\|\cdot\|_d$ norms:

$$\langle \delta m, D\bar{F}[m]^t(d, \bar{m})\rangle_0 = \langle d, (D\bar{F}[m]\delta m)\bar{m}\rangle_0. \quad (39)$$

Assuming that $\bar{F}$ is implemented by a time-stepping finite difference or finite element method, this notion of transpose can be computed by a variant of the adjoint state method. For the case that $\bar{F}$ is a linearization (that is, Born approximation), so that $D\bar{F}$ is actually a second derivative, this application of the adjoint state method was introduced by Symes and Santosa (1988), and employed by Kern and Symes (1994) in computations similar to those presented here.

CONTROLLABILITY

The stability result of the last two sections is not in itself sufficient foundation for a convergent optimization algorithm, for two reasons:

1. The solution errors $\bar{m}_{\lambda,a} - \bar{m}_\lambda$ and $\bar{q}_{\lambda,a} - \bar{q}_\lambda$ are not directly observable, since neither $\bar{m}_\lambda$ nor $\bar{q}_\lambda$ are known in practice;

2. typical iterative solution algorithms for equations 7 and 19 do not necessarily reduce their solution errors.

Observable quantities actually reduced by iterative algorithms include the data residual $\|\bar{F}[m]\bar{m}_{\lambda,a} - d\|_d$, the residual for each of the equations 7 and 19, and the norms of the approximate solutions $\bar{m}_{\lambda,a}$ and $\bar{q}_{\lambda,a}$.

It might be objected that the solution errors and residuals for equations 7 and 19 are related, the former being at most λ^{-2} times the latter (λ^2 being a lower bound for the normal operator N_λ). Of course such bounds are not uniform in λ, hence are not useful in formulation of an effective algorithm.

As will be explained in the next section, the necessary form for the gradient error estimate is

$$\|g - \nabla J_\lambda[m,d]\|_{M_b} \le K\epsilon \tag{40}$$

in which $K \ge 0$ and ϵ is a parameter of the estimation process that produces $\bar{m}_{\lambda,a}, \bar{q}_{\lambda,a}$, in effect that $\bar{m}_{\lambda,a}, \bar{q}_{\lambda,a}$ are functions of $m \in U, d \in D^0, e > 0$. Only minimal assumptions will be made about the estimation process - as will be seen below, only requirements on the residuals produced in equations 7 and 19.

However, such estimates appear to require explicit use of scale separation, abstracted in part by introducing a *family* $d = \{d_\lambda : \lambda > 0\} \subset D^0$ of data, and corresponding separable least squares problems. The *high frequency cone condition*

$$\|d_\lambda\|_k \le C_k \lambda^2 \|d_\lambda\|_{k+1}, \, k \ge k_0 \tag{41}$$

gives a quantitative expression of scale separation. The lower index bound k_0 is presumed to be at most -1.

Remark 3. *For the Sobolev scale, this condition identifies the regularization parameter λ^2 with wavelength.*

From here on, $\bar{m}_\lambda[m,d]$ and $\bar{q}_\lambda[m,d]$ denote the solutions of the parametrized systems

$$N_\lambda[m]\bar{m}_\lambda[m,d] = \bar{F}^\dagger[m]d_\lambda \tag{42}$$

and

$$N_\lambda[m]\bar{q}_\lambda[m,d] = A^\dagger A\bar{m}_\lambda[m,d]. \tag{43}$$

Linking the size of the residuals

$$e_{m,\lambda}[m,d] \;=\; N_\lambda[m]\bar{m}_{\lambda,a}[m,d] - \bar{F}^\dagger[m]d_\lambda, \tag{44}$$

$$e_{q,\lambda}[m,d] \;=\; N_\lambda[m]\bar{q}_{\lambda,a}[m,d] - A^\dagger A\bar{m}_{\lambda,a}[m,d,e] \tag{45}$$

to the solution errors $\bar{m}_{\lambda,a} - \bar{m}_\lambda$ and $\bar{q}_{\lambda,a} - \bar{q}_\lambda$ requires some assumption about the approximate solutions $\bar{m}_{\lambda,a}$ and $\bar{q}_{\lambda,a}$. Note that any Krylov method for solution of the systems 42 and 43, preconditioned by $(I + \lambda^2 R^\dagger R)^{-1}$, yields first iterates

$$\bar{m}_{\lambda,1}[m,d] = (I + \lambda^2 R^\dagger R)^{-1} \bar{F}^\dagger[m] d_\lambda \tag{46}$$

$$\bar{q}_{\lambda,1}[m,d] = (I + \lambda^2 R^\dagger R)^{-1} A^\dagger A \bar{m}_{\lambda,1}[m,d] \tag{47}$$

for which the residuals are

$$e_{m,\lambda,1} = S[m](I + \lambda^2 R^\dagger R)^{-1} \bar{F}^\dagger[m] d_\lambda, \tag{48}$$

$$e_{q,\lambda,1} = S[m](I + \lambda^2 R^\dagger R)^{-1} A^\dagger A \bar{m}_{\lambda,1}[m,d], \tag{49}$$

Theorem 2. *Suppose that for $\epsilon > 0$, $\bar{m}_{\lambda,a}[m,d,\epsilon]$ and $\bar{q}_{\lambda,a}[m,d,\epsilon]$ are approximate solutions of 42 and 43 for which the residuals $e_{m,\lambda,\epsilon}$ and $e_{q,\lambda,\epsilon}$ satisfy*

$$\|e_{m,\lambda,\epsilon}\|_m \leq \epsilon \|e_{m,\lambda,1}\|_m \tag{50}$$

$$\|e_{q,\lambda,\epsilon}\|_m \leq \epsilon \|e_{q,\lambda,1}\|_m \tag{51}$$

Then there exists $K > 0$ and $\lambda_0 \geq 0$ so that for $\lambda \leq \lambda_0$,

$$\|\bar{m}_{\lambda,a}[m,d,\epsilon] - \bar{m}_\lambda[m,d]\|_m \leq K\epsilon \|d_\lambda\|_d \tag{52}$$

$$\|\bar{q}_{\lambda,a}[m,d,\epsilon] - \bar{q}_\lambda[m,d]\|_m \leq K\epsilon \|d_\lambda\|_d \tag{53}$$

uniformly in $m \in U$, for fixed choice of $\{C_k\}$ in 41.

Proof. Since

$$N_\lambda[m](\bar{m}_{\lambda,a}[m,d,\epsilon] - \bar{m}_\lambda[m,d]) = e_{m,\lambda,\epsilon},$$

$$\|\bar{m}_{\lambda,a}[m,d,\epsilon] - \bar{m}_\lambda[m,d]\|_0 \leq \frac{K}{\lambda^2}\|e_{m,\lambda,\epsilon}\|_0 \leq \frac{K\epsilon}{\lambda^2}\|e_{m,\lambda,1}\|_0$$

$$\leq \frac{K\epsilon}{\lambda^2}\|d_\lambda\|_{-1} \leq K\epsilon \|d_\lambda\|_d$$

The next-to-the last inequality follows from the definition of $e_{m,\lambda,1}$ and the uniform Op^{-1} bound on $S[m]$, the last from 41 for $k = -1$. Thus 52 is established. Since $e_{q,\lambda,1}$ takes the same form, a similar estimate holds for it. $\qquad\square$

Corollary 1. *Under the assumptions of Theorem 2, there exists $K \geq 0$ for which*

$$\left| \frac{1}{2}\|A\bar{m}_{\lambda,a}[m,d,\epsilon]\|^2 - J_\lambda[m,d] \right| \leq K\epsilon \tag{54}$$

$$\|G_\lambda[m](d, \bar{m}_{\lambda,a}[m,d,\epsilon], \bar{q}_{\lambda,a}[m,d,\epsilon]) - \nabla J_\lambda[m,d]\|_{M_b} \leq K\epsilon \tag{55}$$

for $m \in U$ and sufficiently small λ.

Proof. Follows from continuity of G_λ as described in Theorem 1, and bounds on the solution errors established in Theorem 2.   □

OPTIMIZATION

In view of the error inherent in the calculations explained in the previous sections, standard variants of Newton's method for local minimization cannot be guaranteed to converge to a local minimizer of J_λ. Minimization of the reduced objective requires use of iterations that converge in the presence of inexact gradient and objective function evaluations. Such an algorithm would necessarily need to couple error and step size control.

Since error-contaminated function and gradient computations are hardly rare, it is unsurprising that this topic has a fairly large literature (Dembo et al., 1982; Carter, 1991; Deuflhard, 1991; Carter, 1993; Eisenstat and Walker, 1994). All of these works and many more recent ones share a major drawback: they mandate *absolute* control of function and/or gradient error, implying in particular that the computed gradient be a descent direction. While of course this condition must eventually hold, it is nearly impossible to check in practice: many sources of error are like those considered in the preceding sections, in that they give no direct measure of gradient or function error, but only allow indirect control.

One exception to this pattern is the work of Heinkenschloss and Vicente (2001); Kouri et al. (2013, 2014) on error control in the context of *trust region* methods for sequential qudratic programming, an approach to constrained optimization. Trust region globalization is a general concept that applies also to unconstrained formulations (Conn et al., 2000; Nocedal and Wright, 1999). The condition that Heinkenschloss and Vicente (2001) place on gradient error takes the form: error $\leq K\epsilon$, where ϵ is a control parameter but K is unknown or at least poorly controlled. Thus nothing can be said *a priori* about the size of the gradient error, or even whether the computed gradient is a descent direction, except in an asymptotic sense. Reference to Corollary 1 shows that the theory developed above provides exactly this sort of condition.

The trust region approach to minimization of a C^2 function f defined on a

Hilbert space H bases its kth step on a quadratic model approximating f near x_k.

$$m_k(s) = f_k + \langle s, g_k \rangle + \frac{1}{2}\langle s, H_k s \rangle.$$

Here f_k is the computed value and g_k the computed gradient at x_k. H_k is an approximation, possibly quite crude, to the Hessian of f at x_k. The basic trust region algorithm seeks the step $s = x_{k+1} - x_k$ as the optimum of the constrained problem

$$\text{minimize } m_k(s) \text{ subject to } \|s\| \leq \Delta_k, \tag{56}$$

The *trust radius* Δ_k is also subject to update as the iteration proceeds, so as to satisfy a *sufficient decrease* criterion leading to assured global convergence to a local minimizer, under various conditions (smoothness, Hessian definiteness, etc.) some of which will be mentioned below.

I will describe a simple variant of the trust radius step, depending on two computed quantities: actual reduction,

$$actred = f(x_k) - f(x_k + s)$$

and predicted reduction,

$$predred = m_k(0) - m_k(s) = -(\langle s, g_k \rangle + \frac{1}{2}\langle s, H_k s \rangle)$$

, and on four magic numbers, $0 < \eta_1 < \eta_2 < 1$ and $0 < \gamma_1 < 1 < \gamma_2$.

Choose a *search direction p* by minimizing the model as an unconstrained problem: that is,

$$p = -H_k^{-1} g_k.$$

Note that H_k is only an approximation to the Hessian of f, and that in a rather loose sense: it may be computed by building up a Krylov or quasi-Newton approximation to the Hessian or its Gauss-Newton simplification, for example. Thus p may be only a crude approximation to a Newton or Gauss-Newton step: this algorithm description encompasses so-called truncated Newton-Krylov methods, for example, and the step selection may interact with the computation of H_k.

Next enter the step update loop:

1. set $s = \Delta_k p$

2. compute *actred, predred*.

    (a) if *actred* $\leq \eta_1 predred$, $\Delta_k \leftarrow \gamma_1 \Delta_k$, go to 1.

    (b) if *actred* $\geq \eta_2 predred$, $\Delta_k \leftarrow \min(1, \gamma_2 \Delta_k)$,

3. $x_{k+1} = x_k + s$, $\Delta_{k+1} = \Delta_k$, exit.

To see how this algorithm accommodates function and gradient error, first presume that there is none: that is, $f_k = f(x_k), g_k = \nabla f(x_k)$, and $H_k \geq \beta > 0$ for all k. The theory presumes that the function f is twice continuously differentiable, bounded below, and has uniform bounds on the gradient and condition number of the Hessian over the sub-level sets of the objective. It is easy to see that the condition for a successful step implies that

$$f_k - f_{k+1} \geq C\|g_k\|\Delta_k \tag{57}$$

references, these assumptions imply a natural floor under the trust radius Δ_k. The trust radius decreases only when the actual function reduction is not at least the lower proportion η_1 of the predicted reduction. However as the trust radius, hence the step, gets smaller, the actual reduction and the predicted reduction become close, since the model becomes close to the function to first order. Once the trust radius is small enough, the conditions for further decrease are never met, and the trust radius never decreases beyond a fixed positive threshhold. Since the function values are decreasing from step to step and bounded below, they must converge. Therefore the actual reduction converges to zero, hence the gradient must converge to zero thanks to inequality 57. Note that this reasoning does not establish that $\{x_k\}$ converges - that requires a bit more reasoning. However, if the sequence of iterates does converge, it must converge to a stationary point.

In fact the expected behaviour of this algorithm is that it eventually settles down at $\Delta_k = 1$, takes the full step, and converges at whatever rate the underlying approximation to Newton's method yields.

Convergence with inexact function and gradient evaluations follows from an extension of the same reasoning. Assume for the moment that the function values are exact. Then the primary condition to be imposed on the gradient, according to Heinkenschloss and Vicente (2001), is of the form

$$\|g_k - \nabla f(x_k)\| \leq \xi \min(\|g_k\|, \Delta_k). \tag{58}$$

in which it is merely required that the constant $\xi \geq 0$ be uniform over the sublevel set of f in which the iterates lie. Then once again reduction of Δ_k during a step eventually stops short of a uniform positive lower bound, since the inequality 58 forces g_k, hence the step s, towards the steepest descent step. For short enough steepest descent steps the actual reduction exceeds η_1 times the predicted reduction, which leads both to the step being taken and in the computed gradient g_k being reduced in length.

I embed this mechanism in an algorithm for minimization of J_λ following the pattern of Heinkenschloss and Vicente (2001), section 5. Add $\xi > 0$ to the list of magic numbers required by the trust region algorithm; in theory at least $\xi = 1$ is adequate.

Define $(f, g, H) = JET(m, \epsilon)$ via the algorithm

1. compute initial approximate solutions $\bar{m}_1, \bar{q}_1$

$$
\begin{aligned}
\bar{m}_1 &= (I + \lambda^2 R^\dagger R)^{-1} \bar{F}^\dagger [m] d_\lambda & (59) \\
\bar{q}_1 &= (I + \lambda^2 R^\dagger R)^{-1} A^\dagger A \bar{m}_1 & (60)
\end{aligned}
$$

with residuals $e_{m,1}, e_{q,1}$;

2. compute approximate solutions $\bar{m}, \bar{q}$ of the system 42,43 satisfying with residuals $e = (e_m, e_q)$ satisfying

$$
\begin{aligned}
\|e_m\|_m &\leq \epsilon \|e_{m,1}\|_m & (61) \\
\|e_q\|_m &\leq \epsilon \|e_{q,1}\|_m & (62)
\end{aligned}
$$

3. compute approximate function value, gradient

$$
\begin{aligned}
f &= \frac{1}{2} \|A\bar{m}\|^2 & (63) \\
g &= G_\lambda [m](d, \bar{m}, \bar{q}) & (64)
\end{aligned}
$$

and approximate Hessian H

Formulation of inexact Trust Region algorithm:

1. $(f_k, g_k, H_k) = JET(m_k, \epsilon_k)$

2. while $\epsilon_k > \xi \min(\|g_k\|_m, \Delta_k)$

    (a) $\epsilon_k \leftarrow \xi \min(\|g_k\|_m, \Delta_k)$

    (b) $(f_k, g_k, H_k) = JET(m_k, \epsilon_k)$

3. $s = -\Delta_k H_k^{-1} g_k, m_{k+1} = m_k + s, \epsilon_{k+1} = \epsilon_k, (f_{k+1}, g_{k+1}, H_{k+1}) = JET(m_{k+1}, \epsilon_{k+1})$

$$actred \quad = \quad f_k - f_{k+1} \tag{65}$$

$$predred \quad = \quad \left(1 - \frac{\Delta_k}{2}\right) \Delta_k \langle g_k, H_k^{-1} g_k \rangle_m \tag{66}$$

4. if $actred \leq \eta_1 predred, \Delta_k \leftarrow \gamma_1 \Delta_k$, go to 3.

5. if $actred \geq \eta_2 predred, \Delta_k \leftarrow \min(1, \gamma_2 \Delta_k)$,

6. next k

Corollary 2. *The sequence $\{m_k\}$ produced by the inexact Trust Region algorithm just described converges to a stationary point of J_λ.*

MICROLOCALIZATION

With a few exceptions, no examples of forward maps $\bar{F}$ defined by separable inverse wave problems have actual parametrices, that is, operators $\bar{F}^\dagger$ satisfying the definition 27. Instead, they have *microlocal* parametrices. Abstractly, the microlocal property is captured in an approximate projector $\Pi \in Op^0(\bar{M}, \bar{M})$, self-adjoint with respect to $\|\cdot\|_m$ and close to idempotent in the sense that

$$\|\Pi - \Pi^2\|_m \leq \frac{1}{4}, \ 0 \leq \Pi \leq 1. \tag{67}$$

A microlocal parametrix $\bar{F}^\dagger$ satisfies

$$\bar{F}^\dagger[m]\bar{F}[m] - \Pi = S[m] \in Op^{-1}(\bar{M}, \bar{M}). \tag{68}$$

In this section, the approximate projector Π is locally constant, that is, the relation 68 holds in a neighborhood of $m_0 \in M$.

Augment the linear least squares problem 15 by adding $\langle \bar{m}, (I - \Pi)\bar{m} \rangle_m$ to objective. The corresponding effect on the normal operator is:

$$N_\lambda[m] = \bar{F}^\dagger[m]\bar{F}[m] + (I - \Pi) + \lambda^2 R^\dagger R = I + S[m] + \lambda^2 R^\dagger R.$$

Define as before $\bar{m}_\lambda[m,d]$ to be the solution of $N_\lambda[m]\bar{m}_\lambda[m,d] = \bar{F}^\dagger[m]d$, and the reduced objective J_λ by equation 17.

Computation of J_λ and ∇J_λ goes exactly as before, and Theorems 1 and 2 and Corollary 1 hold verbatim. The only change comes in the relation between the minimizer of J_λ and the solution of the inverse problem $m, \bar{m}$: the data misfit is small only if one assumes that d can be fit with $\bar{m}$ nearly annihilated by $I - \Pi$. That is an *a priori* assumption about the solution, which in examples entails a similar assumption about data.

REFERENCES

Araya, K., and W. Symes, 1996, 2D and 2.5D kirchhoff inversion using upwind finite difference amplitudes: Proc. 66th Annual International Meeting, Society of Exploration Geophysicists, 503–506. (Expanded abstract).

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, **26**, 99–108.

Beylkin, G., and R. Burridge, 1990, Linearized inverse scattering problem of acoustics and elasticity: Wave Motion, **12**, 15–22.

Biondi, B., and A. Almomin, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Blazek, K., C. C. Stolk, and W. Symes, 2013, A mathematical framework for inverse wave problems in heterogeneous media: Inverse Problems, **29**, 065001:1–34.

Carter, R., 1991, On the global convergence of trust region algorithms using inexact gradient information: SIAM J.Numer. Anal., **28**, 251–265.

———, 1993, Numerical experience with a class of algorithms for nonlinear optimization using inexact function and gradient information: SIAM J. Sci. Comp., **14**, 368–388.

Chauris, H., and M. Noble, 2001, Two-dimensional velocity macro model estimation from seismic reflection data by local differential semblance optimization:

applications synthetic and real data sets: Geophysical Journal International, **144**, 14–26.

Cohen, J., and N. Bleistein, 1977, An inverse method for determining small variations in propagation speed: SIAM Journal on Applied Mathematics, **32**, 784–799.

Conn, A., N. Gould, and P. Toint, 2000, Trust region methods: SIAM.

Dembo, R., S. Eisenstat, and T. Steihoug, 1982, Inexact Newton methods: SIAM Journal on Numerical Analysis, **19**, 400–408.

Deuflhard, P., 1991, Global inexact Newton methods for very large scale nonlinear problems: Impact of Computing in Science and Engineering, **3**, 366–393.

Eisenstat, S., and H. Walker, 1994, Globally convergent inexact Newton methods: SIAM Journal on Optimization, **4**, 393–422.

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

Heinkenschloss, M., and L. Vicente, 2001, Analysis of inexact trust-region SQP algorithms: SIAM J. Optimization, **12(2)**, 283–302.

Hou, J., and W. Symes, 2015, An approximate inverse to the extended Born modeling operator: Geophysics, **80**, no. 6, R331–R349.

Huang, Y., and W. Symes, 2015, Born waveform inversion via variable projection and shot record model extension: 85rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 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.

Kouri, D. P., M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders, 2013, A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty: SIAM Journal on Scientific Computing, **35**, A1847–A1879.

———, 2014, Inexact objective function evaluations in a trust-region algorithm for pde-constrained optimization under uncertainty: SIAM Journal on Scientific Computing, **36**, A3011–A3029.

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.

Li, M., J. Rickett, and A. Abubakar, 2013, Application of the variable projection scheme to frequency-domain full-waveform inversion: Geophysics, **78**, R249–

R257.

Liu, Y., W. Symes, and Z. Li, 2014, Inversion velocity analysis via differential semblance optimization: Presented at the 76th Annual Meeting, Extended Abstracts, European Association of Geoscientists and Engineers.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2788–2792.

Mulder, W., and F. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184–1191.

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

Operto, M. S., S. Xu, and G. Lambaré, 2000, Can we quantitativvely image complex structures with rays?: Geophysics, **65**, 1223–1238.

Plessix, R.-E., 2000, Automatic cross-well tomography: an application of the differential semblance optimization to two real examples: Geophysical Prospecting, **48**, 937–951.

Pratt, R., and W. Symes, 2002, Semblance and differential semblance optimization for waveform tomography: a frequency domain approach: Journal of Conference Abstracts, **7**, 183–4.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

Song, H., and W. Symes, 1994, Inversion of crosswell tomography via differential semblance optimization: Technical report, The Rice Inversion Project. (Annual Report).

Symes, W., 2014, Seismic inverse problems: recent developments in theory and practice: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 2–5.

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.

Taylor, M., 1981, Pseudodifferential Operators: Princeton University Press.

ten Kroode, F., 2012, A wave-equation-based Kirchhoff operator: Inverse Problems, 115013:1–28.

———, 2014, A Lie group associated to seismic velocity estimation: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 142–146.

van Leeuwen, T., and F. Herrmann, 2013, Mitigating local minima in full-waveform inversion by expanding the search space: Geophysical Journal International, **195**, 661–667.

van Leeuwen, T., and W. Mulder, 2009, A variable projection method for wave-

form inversion: 71st Annual Meeting, Extended Abstracts, European Association of Geoscientists and Engineers, U024.

Virieux, J., S. Jin, R. Madariaga, and G. Lambaré, 1992, Two dimensional asymptotic iterative elastic inversion: Geophys. J. Int., **108**, 575–588.

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

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1089–1093.

Weibull, W., and B. Arntsen, 2013, Automatic velocity analysis with reverse time migration: Geophysics, **78**, S179–S192.

Xu, S., Y. Zhang, and B. Tang, 2011, 3D angle gathers from reverse time migration: Geophysics, **76**, no. 2, S77–S92.

Zhang, Y., J. Sun, and S. H. Gray, 2007, Reverse time migration: amplitude and implementation issues: 77th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2145–2149.

Zhang, Y., G. Zhang, and N. Bleistein, 2005, Theory of true-amplitude one-way wave equations and true-amplitude common-shot migration: Geophysics, **70**.

An alternative formula for approximate extended Born inversion

Jie Hou, William W. Symes

ABSTRACT

Various modifications of Reverse Time Migration provide asymptotic inverses to the subsurface offset extended Born modeling operator for constant-density acoustics. These approximate inverses have the same quality (asymptotic accuracy) as do Generalized Radon Transform pseudoinverses, but can be computed without any ray tracing whatsoever. In this paper, we describe an approximate inverse of this type whose additional computational cost, above that of subsurface offset extended Reverse Time Migration, is negligible.

INTRODUCTION

Reverse time migration (RTM) with space-shift imaging condition (Sava and Fomel, 2003) is the transpose or adjoint of the space-shift (or subsurface offset) extended Born modeling operator of constant-density acoustics (Stolk et al., 2009). Remarkably, pre- and post-multiplication with relatively low-cost auxiliary operators convert the RTM operator from an adjoint into an asymptotic inverse operator. This observation is implicit in a number of works on "true amplitude migration" (Zhang et al., 2007; Xu et al., 2011), and was made explicit by ten Kroode (2012, 2014) and Hou and Symes (2015b). In one of its forms, the approximate inverse is

$$\bar{F}^{\dagger} = W_{\text{model}}^{-1} \bar{F}^* W_{\text{data}}, \tag{1}$$

in which $\bar{F}$ is the subsurface offset extended Born modeling operator (the dependence on background model is suppressed) and $\bar{F}^*$ is its adjoint (a form of RTM). W_{model} and W_{data} are model- and data-domain weight operators, defined by

$$\begin{aligned} W_{\text{model}}^{-1} &= 4v^5(D_x^2 + D_z^2)^{1/2}(D_h^2 + D_z^2)^{1/2}P; \\ W_{\text{data}} &= I_t^4 D_{z_s} D_{z_r}. \end{aligned} \tag{2}$$

Here, v is the background or migration velocity ($v(x,z)$ in 2D, $v(x,y,z)$ in 3D), D_x, D_h, D_z are x, h, z-direction derivatives (where h is subsurface offset), I_t is the indefinite time integral and D_{z_s}, D_{z_r} are source and receiver depth derivatives. $P(...)$ (the reciprocal is used in Hou and Symes (2015b)) is a so-called pseudodifferential operator (ΨDO) of order zero, an oscillatory integral defined by its symbol or amplitude, also denoted P - see equation 11 below. $\bar{F}^\dagger$ is an asymptotic inverse in the sense that $\bar{F}^\dagger \bar{F} - I$ suppresses high-frequency components, that is, in the same sense as Generalized Radon Transform (GRT) or Kirchhoff inverses (Beylkin, 1985; Bleistein, 1987; Operto et al., 2000). Of course GRT inverses require extensive computation of traveltime and amplitude tables. $\bar{F}^\dagger$ requires no ray-related computations at all.

Note that the operator W_{model} defined by Hou and Symes (2015b) is written here as W^{-1}_{model}, so that $\bar{F}^\dagger$ is expressed as the adjoint of $\bar{F}$ with respect to weighted norms. This observation leads to a dramatic acceleration of iterative extended least squares migration (Hou and Symes, 2016a).

In this paper, we derive and illustrate an alternative approximate inverse

$$\bar{F}^\ddagger = -8v^4 D_z Q F^* D_t W_{\text{data}} \tag{3}$$

in which Q is a ΨDO with properties similar to those of P. We show that $\bar{F}^\ddagger$ is an asymptotic inverse in the same sense as is $\bar{F}^\dagger$, and illustrate its equivalent accuracy with a nontrivial numerical example. Note, however, that the expression 3 for $\bar{F}^\ddagger$ does not take the form of an adjoint of $\bar{F}$ with respect to weighted norms.

The steps necessary to evaluate the right-hand sides of equations 1 and 3 are:

1. Apply the weight operator W_{data} to the data. Our previous paper (Hou and Symes, 2014, 2015b) discusses implementation of this step. For example, for shallow tow-depth streamer data (free surface, small z_s, z_r relative to shortest wavelength to be imaged), the vertical source and receiver derivatives are present in the recorded traces already, as a side effect of ghosting. If the data is free of ghost reflections, shift the vertical derivatives to the migration by using a dipole Green's function in RTM.

2. For equation 3: apply the time derivative, or include one less time integration in the definition of W_{data}.

3. Apply the extended RTM operator $\bar{F}$, formulated with absorbing boundary if source and receiver vertical derivatives are already applied to the data, else formulated with vertical dipole source and receiver.

4. 4.a. For equation 1: apply square roots of Laplacians in x, z and h, z (2D) or x, y, z and h_x, h_y, z (3D). This step requires partial Fourier transforms of 3D data (for 2D inversion) or 5D data (for 3D inversion).

   4.b. For equation 3: apply the z derivative.

5. Apply the operators P (equation 1) or Q (equation 3), or appropriate approximations.

6. Scale by appropriate power of v and numerical factors.

For kinematically accurate v, the output of F^* focuses near $h = 0$, and an appropriate approximation for P or Q is the identity operator, that is, a "no-op". We will say a few words about the general case in the discussion section.

Comparison of steps 4.a. (for equation 1) and 2. and 4.b. (for equation 3) suggests that implementation of equation 3 is likely to be less computationally demanding than implementation of equation 1, as no spatial transforms are involved in the former implementation. At least for kinematically accurate migration velocity, with Q neglected, equation 3 defines a least expensive RTM-based approximation inverse, as computationally feasible as subsurface-offset extended RTM itself.

In the next section, we will outline the theoretical derivation of equations 1 and 3, limiting ourselves to 2D for simplicity. In the following section, we present an example based on the Marmousi model (Bourgeois et al., 1991). The discussion section points out some open questions about the constructions described here.

THEORY

In this section, we sketch an approach for construction of approximate inverse operators, reserving mathematical details for the appendices. We make extensive use of results from Hou and Symes (2015b). To avoid confusion, we prefix all citations of equations from that paper by "HS-". We also use the notation of that

paper, in particular $\phi(x_r, x_s, x, z, h) = T(x_s, z_s, x - h, z) + T(x_r, z_r, x + h, z)$ for offset reflection time as a function of source location (x_s, z_s), receiver location (x_r, z_r), midpoint (x, z), and subsurface offset h. Source and receiver depths are idealized as the same for all traces, so we drop z_s, z_r from the argument list of ϕ. We assume for simplicity that $\partial\phi/\partial z > 0$ in the domain of interest, that is, that reflection time is increasing with depth throughout the region of nonzero reflectivity. This assumption is consistent with the use of horizontal subsurface offset.

The key computation is the asymptotic approximation of a Generalized Radon Transform of the special form

$$M_a u(x, z, h) = \int dx_r dx_s dx' dz' dh' a(x_r, x_s, x, z, h, x', z', h')$$
$$\times \delta(\phi(x_r, x_s, x, z, h) - \phi(x_r, x_s, x', z', h')) u(x', z', h'), \tag{4}$$

defined by the phase ϕ and an amplitude function a. Appendix A shows that

$$M_a u(x, z, h) \approx -\frac{1}{4\pi} \int dk_x dk_z dk_h e^{i(xk_x + zk_z + hk_h)} \frac{\widehat{u}(k_x, k_z, k_h)}{Q(x, z, h, k_x/k_z, k_h/k_z)} \tag{5}$$
$$\times \left[\frac{v^2}{k_z^2} \frac{\partial\phi}{\partial z} \left(\frac{\partial\alpha_s}{\partial x_s} \frac{\partial\alpha_r}{\partial x_r} \right)^{-1} a \right] (x_r(x, z, h, k_x/k_z, k_h/k_z), x_s(x, z, h, k_x/k_z, k_h/k_z), x, z, h).$$

In this representation, $\hat{u}$ denotes the Fourier transform of u, and $Q(x, z, h, k_x/k_z, k_h/k_z)$ is a symbol of order zero, that is, the amplitude (non-oscillatory) factor in the standard representation of a ΨDO. Properties of symbols are described in many standard texts, for example Taylor (1981). Q is defined explicitly in Appendix A, and satisfies

$$Q(x, z, h, k_x/k_z, k_h/k_z) \neq 0 \text{ if } k_z \neq 0, \quad Q(x, z, 0, k_x/k_z, k_h/k_z) \equiv 1. \tag{6}$$

The term in brackets is a function of the source and receiver locations x_r, x_s which are in turn functions of midpoint and offset coordinates and wavenumbers. This functional relationship depends on determining the takeoff slowness vectors $\nabla_{x,z} T(x_{r,s}, z_{r,s}, x \pm h, z)$ of rays connecting sources and receivers to scattering points, from the stationary phase conditions that lead to the asymptotic formula 5. The details are given in ten Kroode (2012) and Appendix A of Hou and Symes (2015b), and also sketched in Appendix A.

A convenient starting point for the derivation of formulae 2 and 3 is equation HS-20, expressing the extended Born operator applied to an extended perturbation $\delta\bar{v}(x, z, h)$ of the (non-extended) velocity model $v(x, z)$, modified by time integrations and vertical source and receiver derivatives, and approximated by ray

theory Green's functions. As in Hou and Symes (2015b), we adopt the notation convention that $\theta_r = \theta(x_r, z_r, x + h, z)$ for the surface takeoff angle of the ray from $(x+h, z)$ to the receiver at (x_r, z_r), and similarly for θ_s and other functions of source and receiver locations and scattering points, with primes for functions of primed variables. The result of equation HS-20 is (we reproduce it here):

$$I_t D_{z_s} I_t D_{z_r} I_t \bar{F} \delta \bar{v}(x_r, x_s, t) \approx \int dx\, dz\, dh\, a_s a_r \frac{\cos \theta_r}{v_r} \frac{\cos \theta_s}{v_s} \delta(t - \phi) \frac{2\pi \delta \bar{v}}{v^3}(x, z, h). \quad (7)$$

Set

$$a^\dagger = -a_s a_r a_s' a_r' \frac{\cos \theta_r}{v_r} \frac{\cos \theta_s}{v_s} \frac{4\pi^2}{v^3 v'^3}, \quad (8)$$

Equation 7 implies that

$$(I_t \bar{F})^* (I_t D_{z_s} I_t D_{z_r} I_t \bar{F}) \approx M_{a^\dagger}. \quad (9)$$

The fact $I_t^* = -I_t$ and application of approximation 5 turns equation 9 into equation HS-21 (derivation details can be found in Appendix A of Hou and Symes (2015b)):

$$M_{a^\dagger} \approx -\frac{1}{32 v(x, z)^5 \pi^3} \int dk_x\, dk_z\, dk_h\, e^{i(k_x x + k_z z + k_h h)} \frac{\widehat{\delta \bar{v}}(k_x, k_z, k_h)}{k_{xz} k_{hz} P(x, z, h, k_x, k_z, k_h)}. \quad (10)$$

Here, $k_{xz} = \sqrt{k_x^2 + k_z^2}, k_{hz} = \sqrt{k_h^2 + k_z^2}$ are the symbols of the operator square roots in the definition of W_{model}^{-1}. Now invoke the multiplicative property of ΨDOs: to leading order in frequency, the symbol of the product (composition) of ΨDOs is the (numerical) product of the symbols. That is, $k_{xz} k_{hz}$ is cancelled by applying W_{model}^{-1}, and $1/P$ by applying the ΨDO

$$(Pu)(x, z, h) = \frac{1}{8\pi^3} \int dk_x\, dk_z\, dk_h\, e^{i(k_x x + k_z z + k_h h)} P(x, z, h, k_x, k_z, k_h) \widehat{u}(k_x, k_z, k_h). \quad (11)$$

Since W_{data} is already implicit in the left-hand side of equation 9, the approximate inverse 1 is established.

To derive the approximate inverse 3, note that

$$\frac{\partial}{\partial t} \delta(t - \phi) = -\left(\frac{\partial \phi}{\partial z}\right)^{-1} D_z \delta(t - \phi). \quad (12)$$

Differentiate both sides of equation 7 with respect to t, use equation 12, and integrate by parts to get

$$D_t I_t D_{z_s} I_t D_{z_r} I_t \bar{F} \delta \bar{v}(x_r, x_s, t)$$

$$\approx - \int dx dz dh\, a_s a_r \frac{\cos \theta_r}{v_r} \frac{\cos \theta_s}{v_s} \left(\frac{\partial \phi}{\partial z}\right)^{-1} \delta(t - \phi) \frac{2\pi D_z \delta \bar{v}}{v^3}(x, z, h). \qquad (13)$$

Accordingly,

$$(I_t \bar{F})^* (D_t I_t D_{z_s} I_t D_{z_r} I_t \bar{F}) \approx M_{a\ddagger} \qquad (14)$$

with

$$a^\ddagger = a_s a_r a'_s a'_r \frac{\cos \theta_r}{v_r} \frac{\cos \theta_s}{v_s} \left(\frac{\partial \phi'}{\partial z}\right)^{-1} \frac{4\pi^2}{v^3 v'^3} D_z. \qquad (15)$$

Another application of the approximation 5, detailed in Appendix B, leads to

$$8v(x, z)^4 (I_t \bar{F})^* (D_t I_t D_{z_s} I_t D_{z_r} I_t \bar{F}) \delta \bar{v}(x, z, h)$$

$$\simeq \frac{1}{8\pi^3} \int dk_x dk_z dk_h e^{i(xk_x + zk_z + hk_h)} \frac{\widehat{\delta \bar{v}}(k_x, k_z, k_h)}{ik_z Q(x, z, h, k_x/k_z, k_h/k_z)}. \qquad (16)$$

Using the fact $I_t^* = -I_t$ and the multiplicative property of ΨDOs as before to move D_z (symbol ik_z) and Q to the left-hand side leads directly to equation 3.

NUMERICAL EXAMPLE

In this section, we use the synthetic Marmousi model (Bourgeois et al., 1991) to compare the accuracy of the two approximate inverse formulae. We use a kinematically accurate velocity, in fact we generate the (Born) data with the same velocity model used to invert it via formulae 1 and 3, so the data to which the operators P and Q are applied is focused near $h = 0$. Thus, these two operators act as approximate identities, and are neglected in our computations.

Figure 1 shows the smoothed background velocity model and reflectivity model. The model is discretized on a 301*921 grid with a spacing of 10 m in both horizontal and vertical directions. The generated Born data have 231 common shot gathers every 40 m, and each shot has 921 receivers every 10 m. The simulation uses a 2.5-5-30-35 Hz bandpass wavelet with 1 ms temporal sampling. The recording length is 4 s.
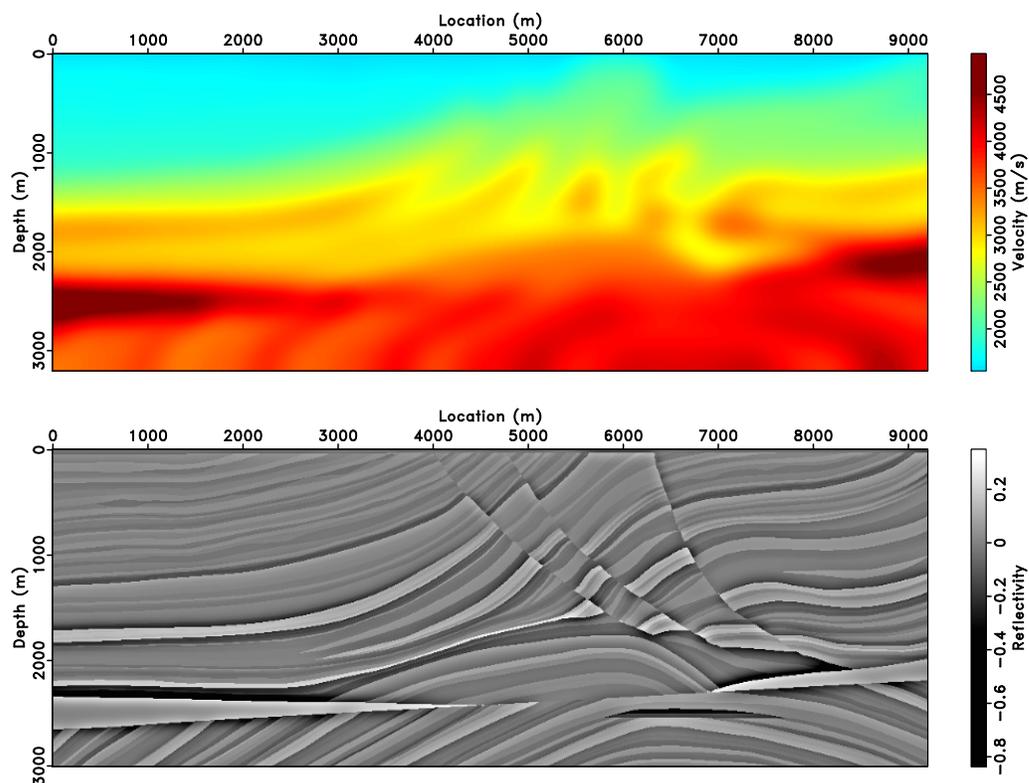
Figure 1: Marmousi model. (a) Background velocity model obtained by smoothing the original Marmousi model; (b) Reflectivity model obtained by taking the difference between the original model and the background model.

Formulae 2 and 3 are applied on the Born data for the Marmousi model. The extended images have an subsurface offset range [-250 m : 250 m] with 10 m spacing. Figure 2 plots the extended inversion produced by the two formulae and their difference. The comparison illustrates that both formulae for the approximate inversion express the same nature and produce very similar results. The difference mainly corresponds to the energy out of the asymptotic framework, e.g., diving wave and refractions. As explained by Hou and Symes (2015b), the physical model can be also recovered by stacking along the subsurface offset axis (nonextended inversion). We plot the stacked images and their difference in Figure 3. These two stacked images are visually almost identical. The artifacts that appearing in the extended images also cancel during the stacking process. The middle trace extracted from the reflectivity model and the two stacked images (Figure 4a) further show the similarity of the two formulae. In order to further verify the effectiveness of the approximate inverse operators, we apply extended Born modeling operator on the inverted extended images to resimulate the data. Figure 5 plots the middle shot of the original data, the resimulated data for the extended images using formulae 1 and 3, and the difference between the two resimulated data. The approximate inversion using both formulae produce almost the same level of relative data misfit, compared to the original data. The middle trace (zero offset) extracted from the middle shot of the original data and two resimulated data confirms this view more clearly (Figure 4b). The substantial difference, especially at large offset, between the two resimulated data (Figure 5d) results from the energy out of the asymptotic framework and shows the effect of the different ΨDOs in two formulae, which is not completely clear yet.

DISCUSSION

We thank Fons ten Kroode for pointing out that the asymptotic inverse 3 may be derived easily from equations 7 and 8 in (ten Kroode, 2014), which summarize an argument analogous to that explained here.

The operator defined by equation 1 is the adjoint of $\bar{F}$ in weighted norms on model and data spaces defined by W_{model} and W_{data}. Since the weighted normal operator $\bar{F}^\dagger \bar{F}$ is approximately the identity, the operator $\bar{F}$ is approximately unitary in the norms defined by the weight operators 2, so that Krylov subspace iterations such as conjugate gradients converge very rapidly to solutions of the extended least squares migration problem (Hou and Symes, 2015a, 2016a,b). It
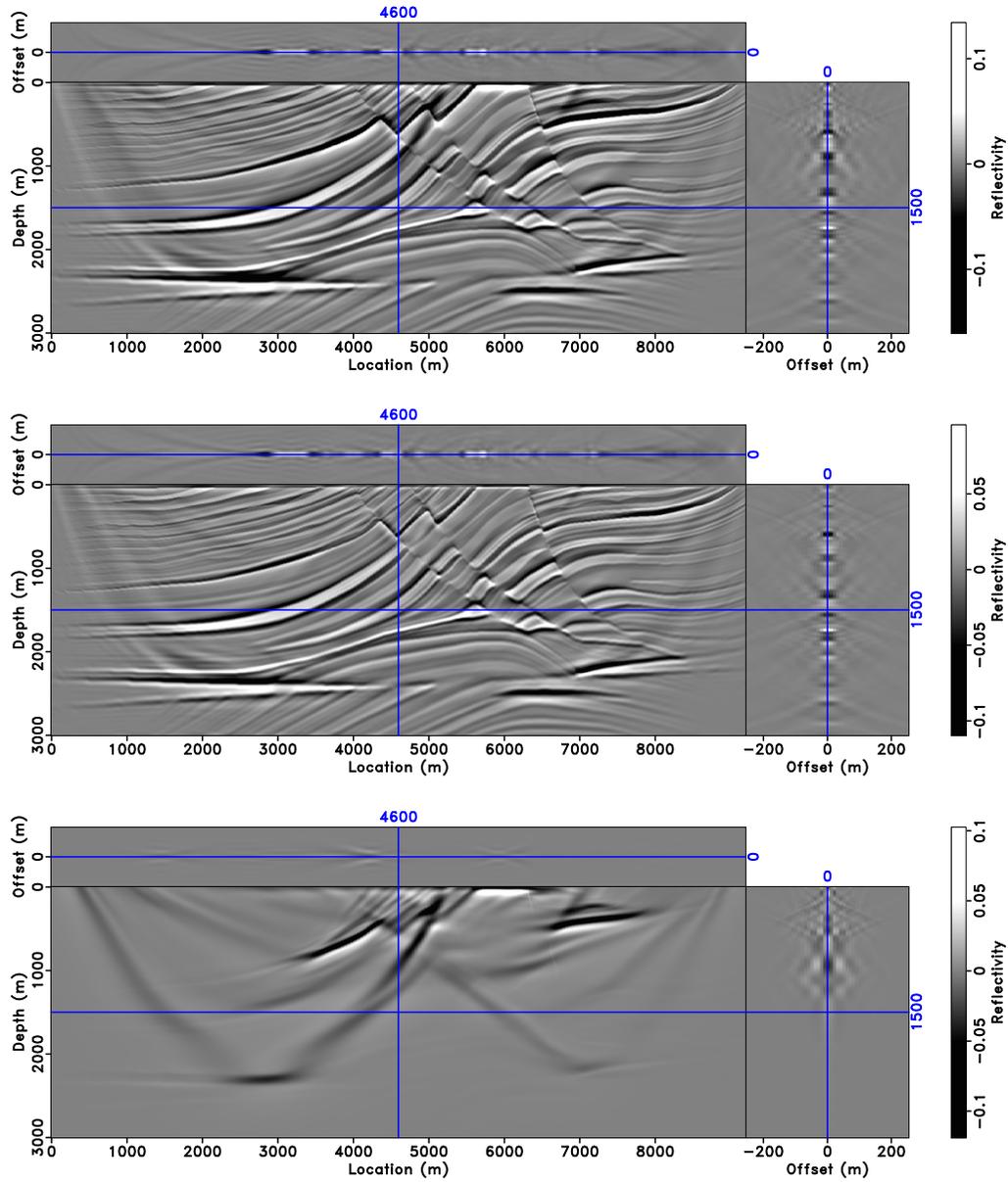
Figure 2: (a) Extended image inverted with formula 1; (b) Extended image inverted with formula 3; (c) The difference between the two extended images.
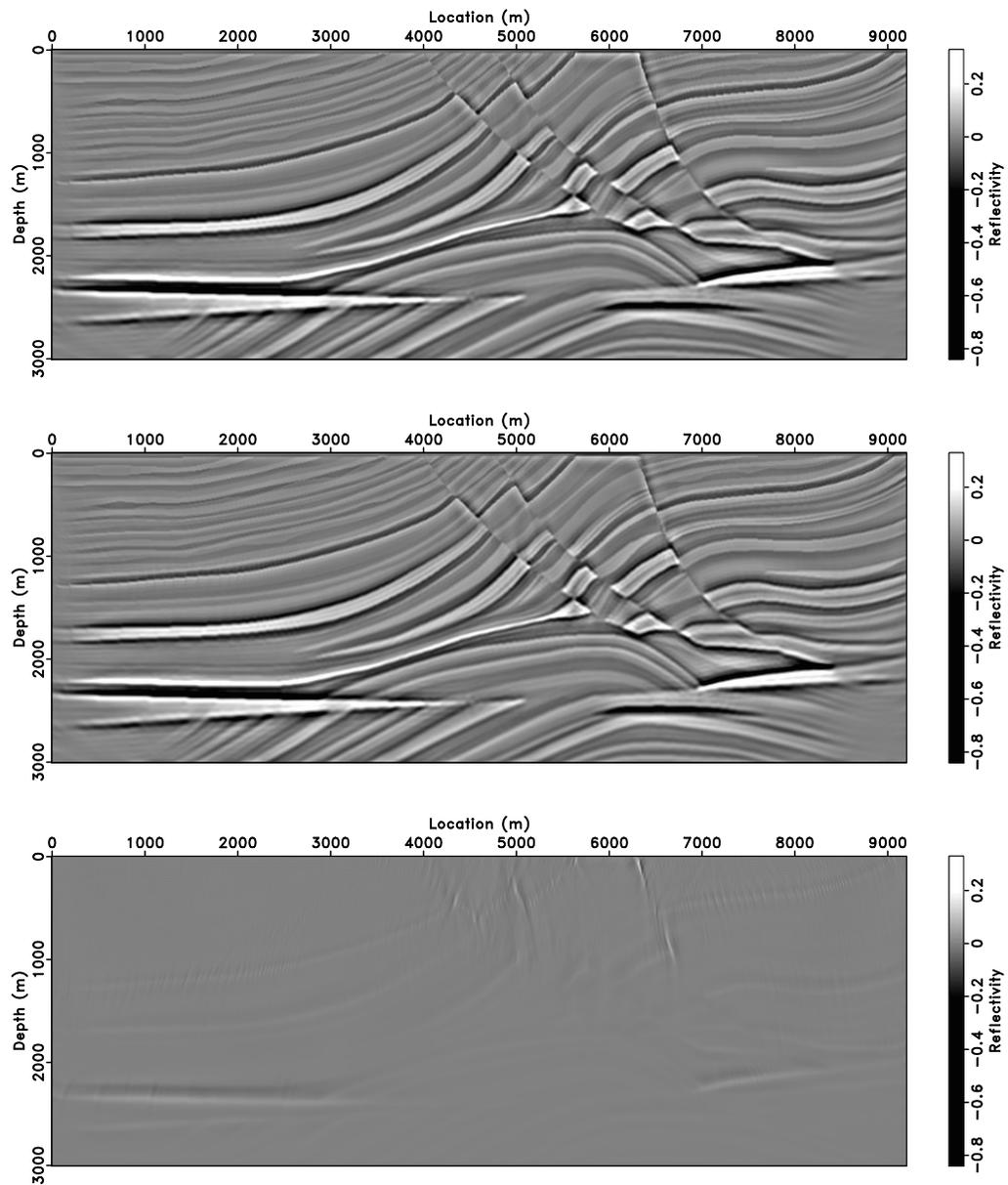
Figure 3: (a) Stacked image corresponding to the extended image in Figure 2a; (b) Stacked image corresponding to the extended image in Figure 2b; (c) The difference between (a) and (b).
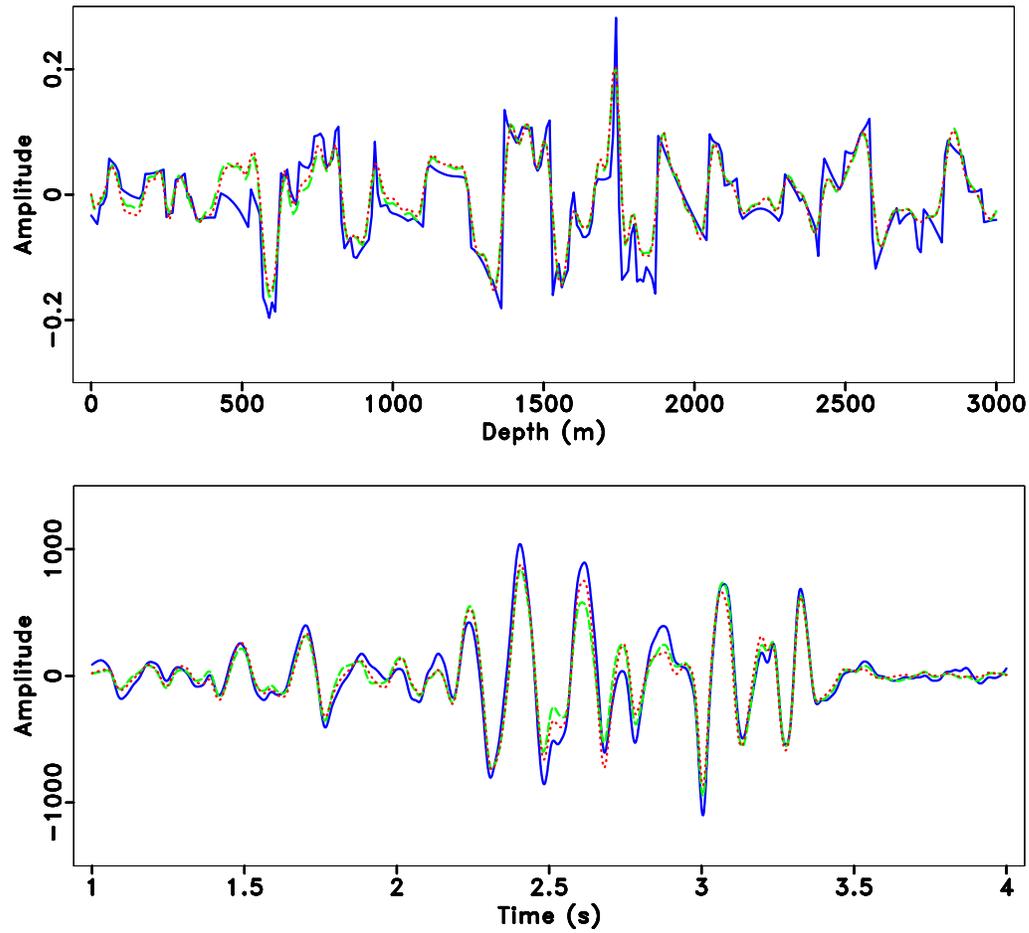
Figure 4: Middle trace comparison. (a) Model comparison: blue line is the reflectivity model; green line is the stacked image using formula 1; red line is the stacked image using formula 3. (b) Data comparison: blue line is the original data, green line is the resimulated data corresponding to the original formula, red line is the resimulated data corresponding to the new formula.
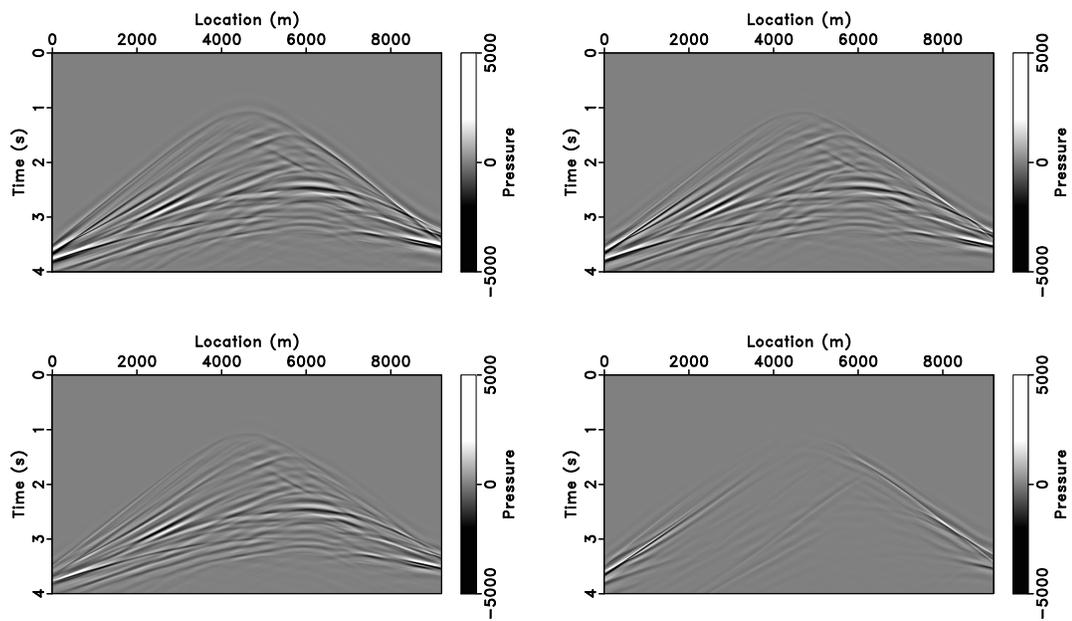
Figure 5: The middle shot of (a) original Born data ; (b) resimulated Born data corresponding to the extended image in Figure 2a; (c) resimulated Born data corresponding to the extended image in Figure 2b;(d) The difference between (b) and (c).

remains to be seen whether the approximate inverse operator shown in equation 3 can play a similar role in accelerating Krylov methods.

The ΨDOs P and Q are in principle necessary for asymptotically accurate recovery of unfocused extended models via 1 and 3. These operators have low approximate rank, and efficient algorithms exist for approximating their action (Bao and Symes, 1996; Demanet and Ying, 2011). However, all of these algorithms have at least the complexity of the Fourier transform, so in fact have substantial cost, especially in 3D. Moreover, only approximations may be computed; at present, the degree of approximation necessary to obtain an accurate approximate inverse is unknown. Experience so far suggests that for convergence acceleration of iterative algorithms, it may be possible to ignore P as we have done here.

The derivations presented here are in principle valid only in the absence of triplications (multiple wave arrivals). However, there is ample evidence that this restriction is not really necessary (Stolk et al., 2009; ten Kroode, 2012, 2014). The absence of turning rays is a more fundamental limitation. To include them would require inclusion of vertical subsurface offsets (Biondi and Symes, 2004).

CONCLUSIONS

In this paper, we have derived a formula for approximate extended Born inversion, whose additional cost beyond that of extended RTM is negligible. By theoretical derivation and numerical experiment, we have demonstrated that the operator derived here, while substantially less expensive to implement (especially in 3D), achieves the same order of asymptotic accuracy as that described by Hou and Symes (2015b).

ACKNOWLEDGEMENTS

provided by the Texas Advanced Computing Center and the Rice University Research Computing Support Group. The Madagascar open source software package has been critically useful in our work. We appreciate the constructive comments from F. ten Kroode and S. Gray.

APPENDIX A

PSEUDODIFFERENTIAL APPROXIMATION OF A GENERALIZED RADON TRANSFORM

In this appendix, we show that equation 5 gives a high-frequency asymptotic approximation to the Generalized Radon Transform 4, by modifying the argument explained in Appendix A of Hou and Symes (2015b). This argument begins with the assumption that $\partial\phi/\partial z > 0$ where the input function $u \neq 0$ (in application, u will be the velocity perturbation $\delta\bar{v}$). Clearly, only those (x',z',h') contribute to the evaluation of $M_a u(x,z,h)$ for which of $\phi(x_r,x_s,x,z,h) = \phi(x_r,x_s,x',z',h')$ for some x_s, x_r. Solve this equation for $z' = Z(x_r,x_s,x,z,h,x',h')$ - given the assumption, the solution is at least locally unique. Note that as a distribution in (x',z',h'),

$$\delta(\phi(x_r,x_s,x,z,h) - \phi(x_r,x_s,x',z',h')) = \left(\frac{\partial\phi}{\partial z'}(x_r,x_s,x',z',h')\right)^{-1} \delta(z' - Z(x_r,x_s,x,h,z,x',h')).$$

$$(A\text{-}1)$$

Substitute the identity A-1 in the definition 4 and introduce the Fourier transform of u to obtain

$$M_a u(x,z,h) = \frac{1}{8\pi^3} \int dk_{x'} dk_{z'} dk_{h'} dx_r dx_s dx' dh' \, a(x_r,x_s,x,z,h,x',Z(x_r,x_s,x,h,z,x',h'),h')$$

$$\times \left(\left|\frac{\partial\phi}{\partial z}\right|^{-1} (x_r,x_s,x',Z(x_r,x_s,x,h,z,x',h'),h') e^{ik_{z'}\left(x'\frac{k_{x'}}{k_{z'}} + h'\frac{k_{h'}}{k_{z'}} + Z(x_r,x_s,x,h,z,x',h')\right)}\right) \widehat{u}(k_{x'},k_{z'},k_{h'}).$$

$$(A\text{-}2)$$

Apply the stationary-phase principle to evaluate the right-hand side of equation A-2 as in equation HS-A4. At the stationary point, $x = x', z = z', h = h'$, and (dropping the primes on the frequency variables)

$$(k_x,k_z)\|\nabla_{x,z}\phi, (k_h,k_z)\|\nabla_{h,z}\phi. \tag{A-3}$$

From relation A-3 it follows that the wavevector (k_x, k_z, k_h) determines the ray connecting the scattering points $(x \pm h, z)$ to the $(x_r, z_r), (x_s, z_s)$ respectively, hence the latter become functions of the covector (x, z, h, k_z, k_z, k_h). In terms of these functions, to leading order in k_z, $M_a u(x, z, h)$

$$
\simeq \int dk_x dk_z dk_h \left[\frac{2\pi}{k_z}\right]^2 \left[\frac{1}{8\pi^3}\right] a(x_r(x, z, h, k_x/k_z, k_h/k_z), x_s(x, z, h, k_x/k_z, k_h/k_z), x, z, h, x, z, h)
$$

$$
\times \left|\frac{\partial\phi}{\partial z}\right|^{-1} (x_r(x, z, h, k_x/k_z, k_h/k_z), x_s(x, z, h, k_x/k_z, k_h/k_z), x, z, h)e^{i(xk_x + zk_z + hk_h)}\widehat{\delta\bar{v}}(k_x, k_z, k_h)
$$

$$
\times |\det \text{Hess}(x_r(x, z, h, k_x/k_z, k_h/k_z), x_s(x, z, h, k_x/k_z, k_h/k_z), x, z, h)|^{-1/2}. \tag{A-4}
$$

The Hessian factor $|\det \text{Hess}|^{-1/2}$ is buried in HS-A27 and HS-A28: we fish it out as

$$
|\det \text{Hess}|^{-1/2} = -\frac{v^2}{2}\left(\frac{\partial\alpha_s}{\partial x_s}\frac{\partial\alpha_r}{\partial x_r}\right)^{-1}\left(\frac{\partial\phi}{\partial z}\right)^2 \frac{1}{Q}, \tag{A-5}
$$

in which α_s, α_r are the ray takeoff angles at source and receiver positions, hence also functions of (x, z, h, k_x, k_z, k_h). Here the arguments are suppressed temporarily. $Q = Q(x, z, h, k_x, k_z, k_x)$ is defined in terms of the z-components of ray slowness

$$
q_r = \frac{\partial T}{\partial z}(x_r, z_r, x + h, z), \quad q_s = \frac{\partial T}{\partial z}(x_s, z_s, x - h, z) \tag{A-6}
$$

by the quantity in square brackets in HS-A28, multiplied by $\left(\dfrac{\partial\phi}{\partial z}\right)^{-2}$:

$$
Q = \frac{v^2}{(q_r + q_s)^2}\left[\frac{q_r^2}{v_-^2} + \frac{q_s^2}{v_+^2} + \left(\frac{1}{v_-^2} + \frac{1}{v_+^2}\right)q_s q_r\right], \tag{A-7}
$$

in which $v_\pm = v(x \pm h, z), v = v(x, z)$. HS-A21 defines q_r, q_s in terms of $v, v_\pm$ and $\partial\phi/\partial z$, and HS-A24, HS-A25, and HS-A26 define the $\partial\phi/\partial z$, as functions of (x, z, h, k_x, k_z, k_h). The reader will note that Q is dimensionless, and homogeneous of degree zero in (k_x, k_z, k_h), hence really a function of $(k_x/k_z, k_h/k_z)$. A bit of algebra using the quoted identities also establishes that

$$
Q(x, z, 0, k_x/k_z, k_h/k_z) = 1. \tag{A-8}
$$

Combining these results with equation A-4, we obtain equation 5.

APPENDIX B

AN ALTERNATIVE FORMULA FOR APPROXIMATE INVERSION

Applying approximation 5 to equation 14, we obtain

$$(I_t\bar{F})^*(D_t I_t D_{z_s} I_t D_{z_r} I_t\bar{F})\delta\bar{v}(x,z,h) \simeq \pi v(x,z)^{-4}\int dk_x dk_z dk_h e^{i(xk_x+zk_z+hk_h)}\widehat{\delta\bar{v}}(k_x,k_z,k_h)$$

$$\times\left[\left(\frac{\partial\alpha_r}{\partial x_r}\frac{\partial\alpha_s}{\partial x_s}\right)^{-1}a_r^2 a_s^2\frac{\cos\theta_r}{v_r}\frac{\cos\theta_s}{v_s}\right](x_r(x,z,h,k_x/k_z,k_h/k_z),x_s(x,z,h,k_x/k_z,k_h/k_z),x,z,h)$$

$$\times\frac{i}{k_z}Q^{-1}(x,z,h,k_x/k_z,k_h/k_z). \tag{B-1}$$

To evaluate the right-hand side of equation B-1, we employ relations HS-B6, HS-B7:

$$a_r^2\frac{\cos\theta_r}{v_r}\left(\frac{\partial\alpha_r}{\partial x_r}\right)^{-1} = \frac{1}{8\pi^2} = a_s^2\frac{\cos\theta_s}{v_s}\left(\frac{\partial\alpha_s}{\partial x_s}\right)^{-1}. \tag{B-2}$$

Substitution of these identities in B-1 leads directly to equation 16, hence to the asymptotic inverse formulat 3 as described in the theory section.

REFERENCES

Bao, G., and W. Symes, 1996, Computation of pseudo-differential operators: SIAM Journal on Scientific Computation, **17**, 416–429.

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, **26**, 99–108.

Biondi, B., and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

Bleistein, N., 1987, On the imaging of reflectors in the earth: Geophysics, **52**, 931–942.

Bourgeois, A., M. Bourget, P. Lailly, M. Poulet, P. Ricarte, and R. Versteeg, 1991, Marmousi, model and data, *in* The Marmousi Experience: Proc. of the 1990 EAEG workshop on practical aspects of seismic data inversion, 5–16.

Demanet, L., and L. Ying, 2011, Discrete symbol calculus: SIAM Review, **53**, 71–104.

Hou, J., and W. Symes, 2014, An approximate inverse to the extended Born modeling operator: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 3784–3789.

——, 2015a, Accelerating extended least squares migration with weighted conjugate gradient iteration: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4243–4248.

——, 2015b, An approximate inverse to the extended Born modeling operator: Geophysics, **80**, no. 6, R331–R349.

——, 2016a, Accelerating extended least-squares migration with weighted conjugate gradient iteration: Geophysics, **81**, no. 4, S165–S179.

——, 2016b, Accelerating least squares migration with weighted conjugate gradient iteration: 78th Annual International Conference and Exhibition, Expanded Abstract, European Association for Geoscientists and Engineers, P104.

Operto, M. S., S. Xu, and G. Lambaré, 2000, Can we quantitativvely image complex structures with rays?: Geophysics, **65**, 1223–1238.

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

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

Taylor, M., 1981, Pseudodifferential Operators: Princeton University Press.

ten Kroode, F., 2012, A wave-equation-based Kirchhoff operator: Inverse Problems, 115013:1–28.

——, 2014, A Lie group associated to seismic velocity estimation: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 142–146.

Xu, S., Y. Zhang, and B. Tang, 2011, 3D angle gathers from reverse time migration: Geophysics, **76**, no. 2, S77–S92.

Zhang, Y., J. Sun, and S. H. Gray, 2007, Reverse time migration: amplitude and implementation issues: 77th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2145–2149.

Subsurface offset extended wave-equation migration velocity analysis with angle filtering

Lei Fu and William W. Symes

ABSTRACT

Migration velocity analysis (MVA) based on subsurface offset extended image provides a way to update the velocity model by measuring the focusing of the extended migrated image. The focusing property is introduced by a horizontal subsurface shift in the imaging condition, which expresses consistency between data and velocity model and can be understood as non-local stress-strain relations in acoustic case. In principle, in a correct velocity, the image should focus at zero subsurface shift; in a wrong velocity, the reflector is imaged as a curve in the extended image, which represents the inconsistency between data and the wrong velocity model. Aiming to recover the correct velocity model by focusing the extended image, differential semblance function penalizes energy in the extended image at non-zero shift. However, under prototypical conditions of acquisition geometry, the existence of artefacts is very likely to deviate the velocity update from its path to the correct velocity. The cause of those artefacts can be studied by means of asymptotic analysis and numerical examples. Here, in order to mitigate those artefacts, a taper in scattering angle domain depending on acquisition geometry and imaging point is introduced. The application of the proposed new method is demonstrated by 2D synthetic examples, which shows migration velocity analysis becomes more robust.

INTRODUCTION

Migration velocity analysis is an image-domain approach to build velocity model by exploiting the redundancy in seismic data. In MVA, the data redundancy is exploited with Common Image Gathers (CIG) by measuring coherency across images obtained for different data offsets. Fundamentally, the physical origin of the

data redundancy is redundant illumination of reflectors by different scattering angles.

One approach to build the CIG is to introduce a subsurface spatial shift. Several application of velocity analysis based on focusing of subsurface extended images have been published, accompanied by the observation of artifacts. Artefacts in extended images are likely to destruct the global convergence property of MVA, leading velocity updates to a wrong direction. Several approaches were proposed to tackle the problem in data domain (Mulder, 2014; Lameloise et al., 2015). By analyzing the origin of those artefacts, this paper proposes a simple but effective approach that directly deals with the artefacts in scattering angle domain.

Proposed by Lailly (1983, 1984), it is an attractive approach to solve the linearized seismic inversion by least-squares minimization of the misfit between observed and modeled data. However, due to its highly nonlinear nature, the objective function of typical least-squares functions appears to possess many 'local minima' far from its global minimum. Here, the global minimum represents the correct physical model that generates the data, while the local minima occurs at the wrong velocity model generating the data similar to the observed one in least-squares sense. After an early example of this obstacle showed by Gauthier et al. (1986), a number of researchers have tried different approaches in an effort to overcome this local minima obstacle. It has been known that FWI succeeds only when wavelength in data and accuracy of initial model combine to predict lowest frequency data to correctly to within half a wavelength. (Cao et al., 1990; Crase et al., 1990; Bunks et al., 1995; Plessix et al., 1999; Shin and Min, 2006). However, neither a" good" initial model nor low-frequency for FWI could be difficult to obtain in the real world (Plessix et al., 2010).

On the other hand, migration velocity analysis possesses global convergence property by exploiting the redundancy in the seismic data. One type of extension makes use of the surface offset or scattering angle as an additional dimension to describe the redundancy in the data, which leads to a surface offset extended modeling (Symes and Kern, 1994; Chauris and Noble, 2001; Mulder and ten Kroode, 2002). Based on the flattening gather principle in conventional stacking velocity analysis or normal moveout (NMO) in seismic processing, the flatness of event in a shot gather or angle gather indicates the correctness of velocity above the event. By adding the surface offset to the model as another degree of freedom, we allow the model to depend on the offset between sources and receivers. However, the surface-oriented extended modeling only works in regions that the lateral hetero-

geneity is not too great. In kinematically complex areas, energy may propagate along multiple paths between source and receiver pairs, so kinematic artifacts strongly violate the consistency or kinematic compatibility between data and velocity model (Nolan and Symes, 1997; Stolk and Symes, 2004).

One alternative is introducing a space shift in the subsurface (Claerbout, 1971; Biondi and Shan, 2002; Biondi and Symes, 2004), rather than on the surface, which leads to the second type of extension: the subsurface offset extension. The principle of MVA is that the kinematic error of the velocity model is measured by an extension parameter (e.g., surface offset, subsurface offset, incidence angle, time shift). For CIG with subsurface offset, the velocity error is compensated by a spatial shift between incident point and reflection point in the subsurface. The subsurface offset extended modeling is originated in Claerbout's "survey sinking" concept. This study implements subsurface offset extension, which is known to be less subject to kinematic artifacts than the surface offset approach. Kinematic artifacts are absent in subsurface image volumes, even in the presence of considerable multipathing, under the sole restriction that rays carrying significant energy do not turn horizontal (Stolk and De Hoop, 2006). Based on theoretical work by Stolk and De Hoop (2001), Shen et al. (2003b) suggested a practical calculation for an update direction designed to focus the subsurface offset image volume at zero offset. Shen and Symes (2008) implemented the differential semblance velocity analysis based on the subsurface offset extending concept. The subsurface offsets extended full waveform inversion is successfully applied by Sun and Symes (2012); Biondi and Almomin (2012); Almomin et al. (2012); Fu and Symes (2015).

Several application of velocity analysis based on focusing of subsurface extended images have been published, accompanied by the observation of artifacts. Artefacts in extended images are likely to lead velocity updates toward a wrong direction. In practice, tapering is the common tool to eliminate migration artefacts due to limitation of acquisition geometry. However, the artefacts in subsurface extended image are different from those observed in conventional migration. Shen (2004) showed some exemplary calculations to identify the relevant ray fields with subsurface space shift. Mulder (2014) gave the formulas to calculate the amplitude in the space-shift extended model for 2D and 3D by stationary phase approximation, provided that the true and trial velocities are constant. Their work shows that the artefacts in subsurface extended image is not only caused by acquisition limitation, more importantly, by peak amplitude from large offset. Almomin et al. (2014) preconditioned the image in angle domain based on amplitude versus an-

gle (AVA) behavior in the acoustic two-way wave-equation. Lameloise et al. (2015) proposed an explicit weight in the CIG derivation. We introduce a taper in scattering angle domain, which depends on acquisition geometry and imaging point.

This paper is organized as follow: we first explain the theory of subsurface offset extended image based on linearized acoustic modelling (Born or single-scattering approximation); we then introduce the new method filter in scattering angle domain; we end with numerical examples, demonstrating that the proposed new approach can significantly improve the quality of velocity updates.

THEORY

Migration Velocity Analysis

Full waveform inversion is known to suffer "cycle skipping" problem, when wavelength in data and accuracy of initial model combine to predict lowest frequency data to correctly to within half a wavelength. On the other hand, even without low frequency data or "good" initial model, migration velocity analysis is able to recover the background velocity by measuring certain property of migrated images (stackpower, flatness, space-shift, time shift).

Proposed by Symes and Carazzone (1991), Differential Semblance Optimization (DSO) approach exploits the model consistency for different subsets of data, providing an automatic tool to update velocity model. In this study, we focus on subsurface spatial shift, which is known to be less subject to kinematic artifacts than other surface offset approaches. When velocity is correct, in ideal case, the energy should focus at $h = 0$; when there is error in velocity model, a reflector is imaged as a curve. The subsurface shift, denoted by h, is restricted to be horizontal, based on the assumption that the rays carrying significant energy do not turn horizontal (Stolk and De Hoop, 2006).

The objective function measures the unfocused energy in the subsurface extended image caused by velocity errors. A simple objective function capturing this concept is

$$J = \frac{1}{2}\|A(h)P(\theta)W(z)I(x,z,h)\|^2 \tag{1}$$

Here, operator $A(h)$ is an annihilator, which penalizes spread-out energy at non-

zero subsurface offset. Many choices of annihilator A have the required relation to the physical model. Amongst the earliest suggested was multiplication by h: $A(h) = h$ (Stolk and De Hoop, 2001; Shen et al., 2003a), which we also use here. $P(\theta)$ is a taper depending on reflection angle θ. The depth weighting $W(z) = (z - z_{min})^p$ is used to compensate the illumination due to the energy decay from depth. The objective function can be minimized only when the velocity is kinematically correct. $I(x, z, h) = F^T[v]d$ is a version of prestack depth migrated image with subsurface offset shift in imaging condition.

The gradient of the objective function J with respect to velocity v may be expressed as:

$$\nabla_v J[v] = SDF[v]^T \left(d, W^T P^T A^2 P W F^T[v]d \right) \tag{2}$$

$DF[v]^T$ is the so-called tomographic or WEMVA operator. As $F[v]$ is the linearization, that is, derivative of the basic acoustic modeling operator, $D\bar{F}[v]$ is actually 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). 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. $F^T[v]d$

Next, the migration wil be reviewed in the setting of the Born approximation of the constant-density acoustics. We will study the problem explicitly in 2D case, but note that much of the discussion carries over without modification to 3D.

In time domain, the linearized ("Born") 2D constant-density acoustics modeling with subsurface shift is governed by wave equation 1 and 4. The reference pressure field $p(x, z, t; x_s)$ solves the acoustic wave equation

$$\left(\frac{\partial^2}{\partial t^2} - v^2 \nabla^2 \right) p(x, z, t; x_s) = w(t)\delta(x - x_s, z - z_s),$$
$$p = 0, t << 0 \tag{3}$$

The reference pressure field is represented by $p(x, z, t; x_s)$, a function of position x, z, x_s and time t. The right-hand side is a simple source representation, which is composed of wavelet time function $w(t)$ and a delta function $\delta(x - x_s, z - z_s)$ centered at source position x_s, z_s.

The perturbational pressure field $\delta p(x, z, t; x_s)$ solves the linearized acoustic wave

equation

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

$$\delta p = 0, t << 0 \tag{4}$$

The reflectivity r here is defined as the perturbation of squared velocity: $r = 2v\delta v$. The extended Born acoustic model used here introduces a horizontal subsurface offset axis, denoted h, and allows the reflectivity to depend on it: $r(x,z,h)$. Since r is (up to a scale factor) the perturbation in the compliance, the extended reflectivity can be understood as a non-local perturbation in the acoustic constitutive relation. Note that only the reflectivity depends on the additional coordinate h - the velocity is non-extended, or *physical*. When the velocity v is correct, $r(x-h,z,h) = r(x,z)\delta(h)$, then equation 4 reduces to ordinary linearized acoustic wave equation.

The extended Born forward modeling operator F is defined in terms of δp by

$$(F[v]r)(x_r,x_s,t) = \delta p(x_r,z_r,t;x_s)$$

$$= \int dx \int dz \int dh \int d\tau\, w(\tau)\nabla^2 G(x-h,z,\tau;x_s)r(x,z,h)G(x+h,z,t-\tau;x_r) \tag{5}$$

where G is the Green's function. 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 $F[v,r]$. The data is the sampled pressure field δp measured at receiver positions $\{(x_r,z_r)\}$ for source positions $\{(x_s,z_s)\}$ at time t. Source and receiver depths are idealized as the same for all traces, so ignored in the notation for the data traces.

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.

The migrated image $I(x,z,h)$ can be expressed as the adjoint of the Born forward modeling operator with subsurface shift

$$I(x,z,h) = (F[v]^*\delta p)(x,z,h)$$

$$= \int x_s \int x_r \int dt \int d\tau\, w(\tau)\nabla^2 G(x-h,z,\tau;x_s)G(x+h,z,t-\tau;x_r)\delta p(x_r,t;x_s) \tag{6}$$

Asymptotic Analysis

In the context of constant velocity, the maximum subsurface offset H can be formulated as a function of the maximum surface offset L and the ratio of the migration velocity and true velocity (Shen, 2004; Mulder, 2014).

In the context of constant velocity, Shen (2004)' showed the curves in the subsurface extended image is an envelope of relevant ray fields of different source and receiver pairs. Let's consider a constant velocity model v_0 with a horizontal reflector located at depth z_0. Sources and receivers are placed on the surface ($z = 0$) and symmetric with respect to the z-axis within range $[-L_{max}, L_{max}]$.

When migration velocity v is correct $v = v_0$, for a source and receiver pair ($x_s = -L, x_r = L$), the travel time t is calculated as

$$t_0 = 2 \frac{\sqrt{L^2 + z_0^2}}{v_0} \tag{7}$$

When migration is wrong $v! = v_0$, according to Claerbout's survey-sinking concept (Claerbout, 1985), the coordinates of the sunken source and receiver are $(x - h, z)$ and $(x + h, z)$ respectively. The travel time is

$$t = 2 \frac{\sqrt{(L - h)^2 + z^2}}{v} \tag{8}$$

The kinematics of the extended image consists of the set of points that the travel time in migration equals to the physical travel time ($t_0 = t$). The image $I(x, z, h)$ is the envelope of a family of curves E parameterized by surface half offset L as

$$E(z, h; l) = (L - h)^2 + z^2 - \rho^2(L^2 + z_0) = 0 \tag{9}$$

where $\rho = v/v_0$ is the ratio between migration velocity and true velocity. The envelop is determined by the contact condition:

$$\frac{\partial E(z, h; L)}{\partial L} = 0 \tag{10}$$

From equation 10, the surface offset L and subsurface offset h has a simple relation:

$$h = L(1 - \rho^2) \tag{11}$$

Eliminate L by solving equation 13 and 9 we get

$$\frac{h^2}{(\rho^2 - 1)z_0^2} + \frac{z^2}{\rho^2 z_0^2} = 1 \tag{12}$$

When the migration velocity is faster than the true velocity ($\rho > 1$), the curve is an eclipse centered at ($h = 0$, $z = 0$); when migration velocity is slower than the true velocity ($\rho < 1$), the curve becomes hyperbola.

Note that the relation $h = L(1 - \rho^2)$ in equation 13 is only valid within the range defined by equation 12. As a result, the maximum subsurface offset h_{max} is defined as

$$when \; \rho > 1, h_{max} = min(z_0\sqrt{\rho^2 - 1}, L_{max}(\rho^2 - 1))$$
$$when \; \rho < 1, h_{max} = L_{max}(1 - \rho^2) \tag{13}$$

In most cases, the maximum subsurface offset h_{max} is determined by the maximum surface offset L_{max}. However, in the faster velocity case ($\rho > 1$), then $h < 0$. When the imaging depth is shallow or the surface offset is large, $h_{max} = z_0\sqrt{\rho^2 - 1}$, so the corresponding maximum surface offset is

$$L_{max} = \frac{h_{max}}{\rho^2 - 1} = \frac{z_0\sqrt{\rho^2 - 1}}{\rho^2 - 1}$$
$$= \frac{z_0}{\sqrt{\rho^2 - 1}} \tag{14}$$

Mulder (2014) demonstrated that in 2D the amplitude increase with subsurface offset h and gave the formulas to calculate the amplitude in the space-shift extended model for 2D and 3D by stationary phase approximation. The amplitude along curve in 2D is

$$A^{2D} = \frac{\rho^5}{128z_0 z}\sqrt{\frac{2\pi v_0^9 r^5}{|\rho^2 - 1|}} \tag{15}$$

where $r = \sqrt{z_0^2 + \frac{h^2}{\rho^2 - 1}}$. Their work showed that the artefacts in subsurface extended image were not only caused by acquisition limitation, more importantly, by peak amplitude from large offset.

The subsurface offset image can also be transform in angle domain by Radon transform (Sava and Fomel, 2003). For correct velocity case, the scatter angle θ can be simply calculated as

$$tan\theta = \frac{L}{z_0} \tag{16}$$

The taper function $P(\theta)$ can be formulated as

$$P(\theta) = cos\left(\frac{\theta}{\theta_{max}}\frac{\pi}{2}\right) \tag{17}$$

where the maximum angle θ_{max} is calculated as $\theta_{max} = \frac{L_{max}}{z_0}$.

For wrong velocity, the angle is

$$tan\theta = \frac{L-h}{z} \tag{18}$$

EXAMPLES

The objective of this section is to demonstrate the feasibility of the proposed approach to eliminated the artefacts by filtering in angle domain. All simulations were performed using a 2D constant-density acoustics, time domain, finite difference method (second order in time, eighth order in space).

We used the method of steepest descent with quadratic backtrack line search to search for the minimum of the objective function. The gradient of the objective function was computed by using equation 9. We performed 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 current model. The optimal step length is estimated by assuming the objective function is quadratic.

Single reflector model

The purpose of this experiment is to investigate the behavior of extended image with different levels of velocity errors thus help us understand the origin of the artefacts.

The background velocity model measures $6.0 \times 2.0 \, km$ with $25 \, m$ cell size uniformly distributed in each dimension. The true background velocity v is constant ($v = 2.5 \, km/s$). In the reflectivity model, there is a horizontal velocity perturbation at depth of $1.0 \, km$. 121 sources ($0 - 6.0 \, km$) and 241 receivers ($0 - 6.0 \, km$) are placed on the surface. Note that the background velocity model $v(x, z)$ is non-extended, while the extended perturbation model $\bar{r}(x, z, h)$ has nonzero value only at $h = 0 \, m$.

Figure 1, 2, and 3 show that the artifacts always exist regardless of migration velocity. Although the shape of the artifacts varies with different migration velocity, one common feature is that artifacts are associated with large angles. This feature is the result of acquisition limitation combined with peak amplitude from large incidence angle. After applying the taper, the artifacts are greatly suppressed.

Overthrust model

The 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 (shown in Figure 21a). The background velocity increases with depth (Figure 2a). The velocity is higher in the center, where the anticline structure sits. The basic information is listed in Table 2.

Parameter	Measurements
Source wavelet	bandpass $5 - 20 \, Hz$
Source position $\mathbf{x_s}$	$x : -2 - 10 \, km$ every $80 \, m$, $z = 0 \, m$
Receiver position $\mathbf{x_r}$	$x : -2 - 10 \, km$ every $20 \, m$, $z = 0 \, m$
Record time	$t = 3 \, s$
Grid size	$dx = dh = dz = 20 \, m$, $dt = 2 \, ms$
Initial velocity	$v = 1.5 \, km/s$

Table 1: Thrust model

Figure 3a shows the initial background velocity is constant ($1.5 : km/s$) and far
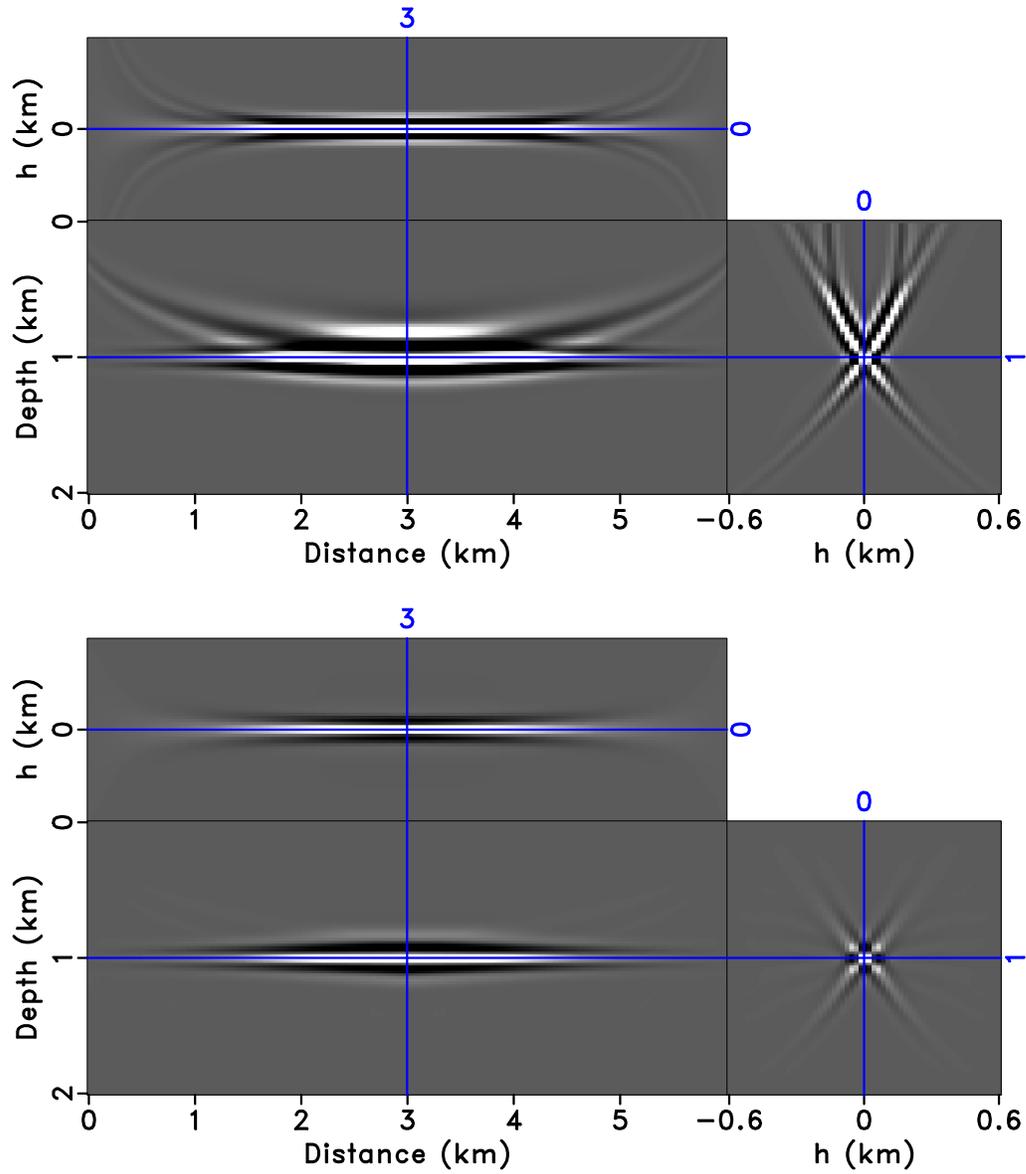
Figure 1: True velocity. (a) Original migrated image; (b) Migrated image with angle taper.
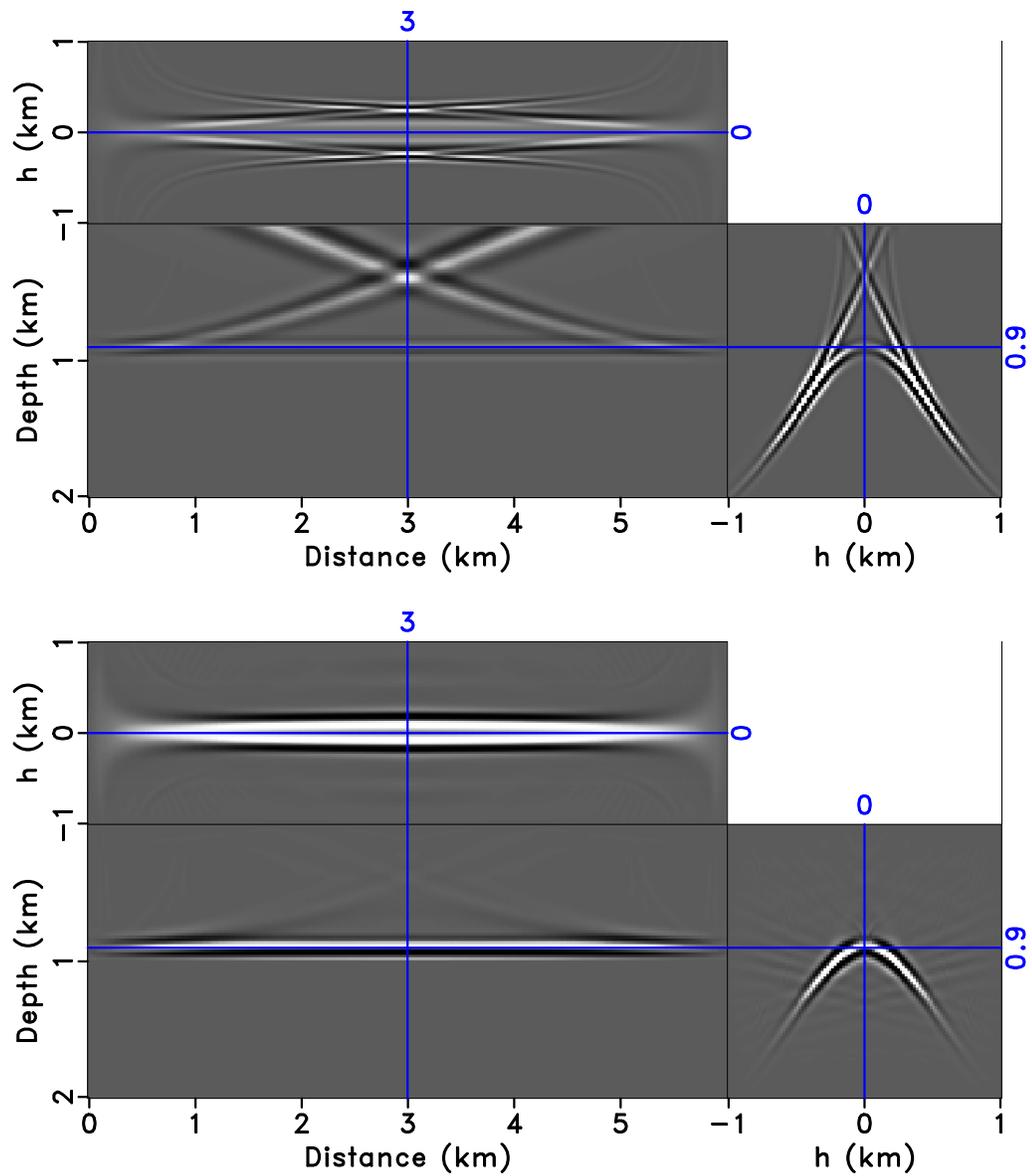
Figure 2: Slower velocity case, migration velocity is 10% slower than true velocity. (a) Original migrated image; (b) Migrated image with angle taper.
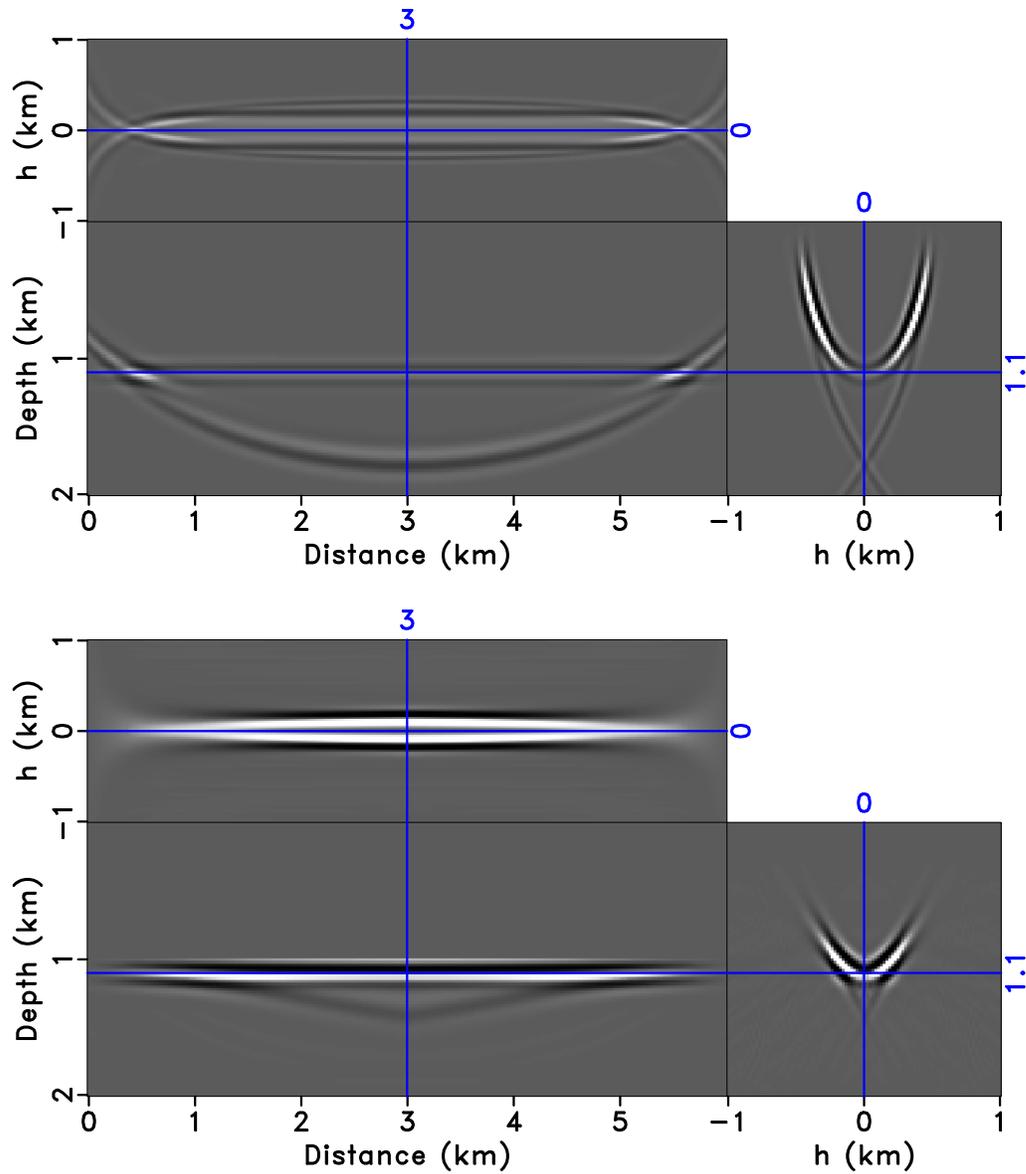
Figure 3: Faster velocity case, migration velocity is 10% fast than true velocity. (a) Original migrated image; (b) Migrated image with angle taper.
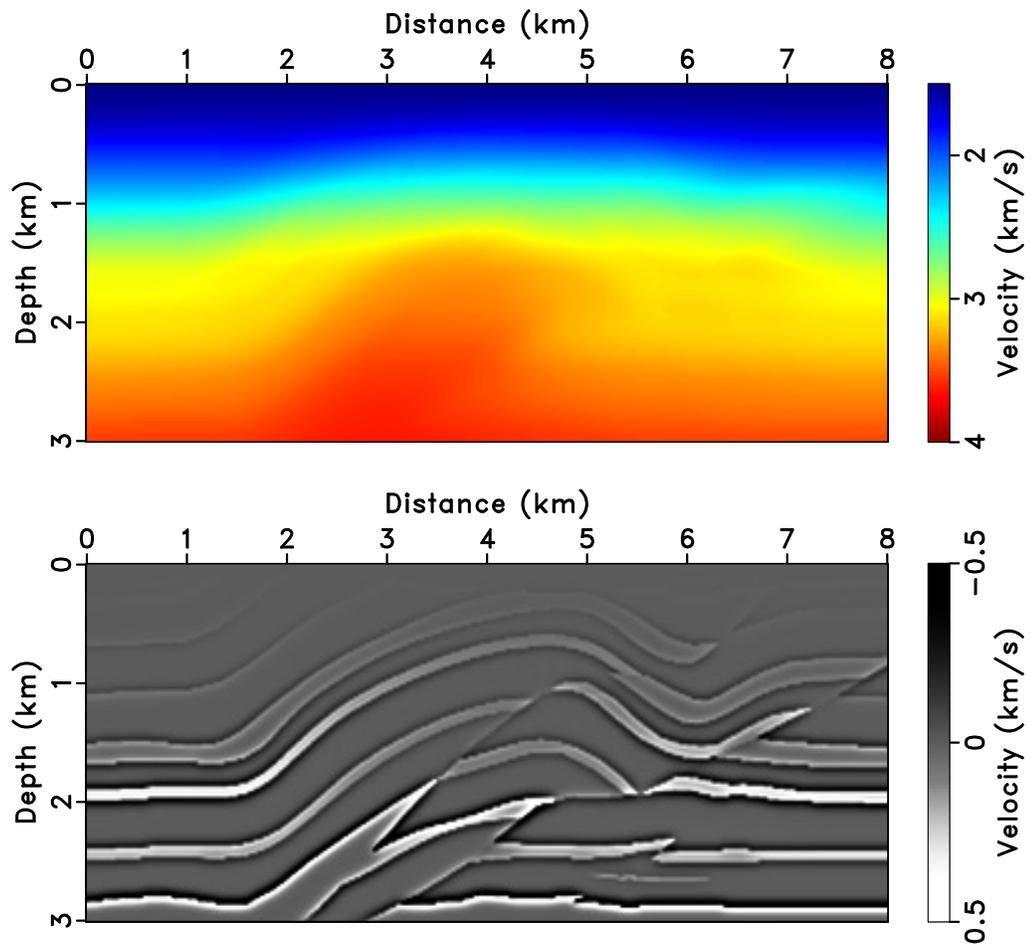
Figure 4: (a) True background velocity model; (b) Reflectivity model.

away from the correct one, so the geological structures can barely be observed from the extended image (Figure 5b). In the subsurface offset gather, the downward curves indict slow velocity.
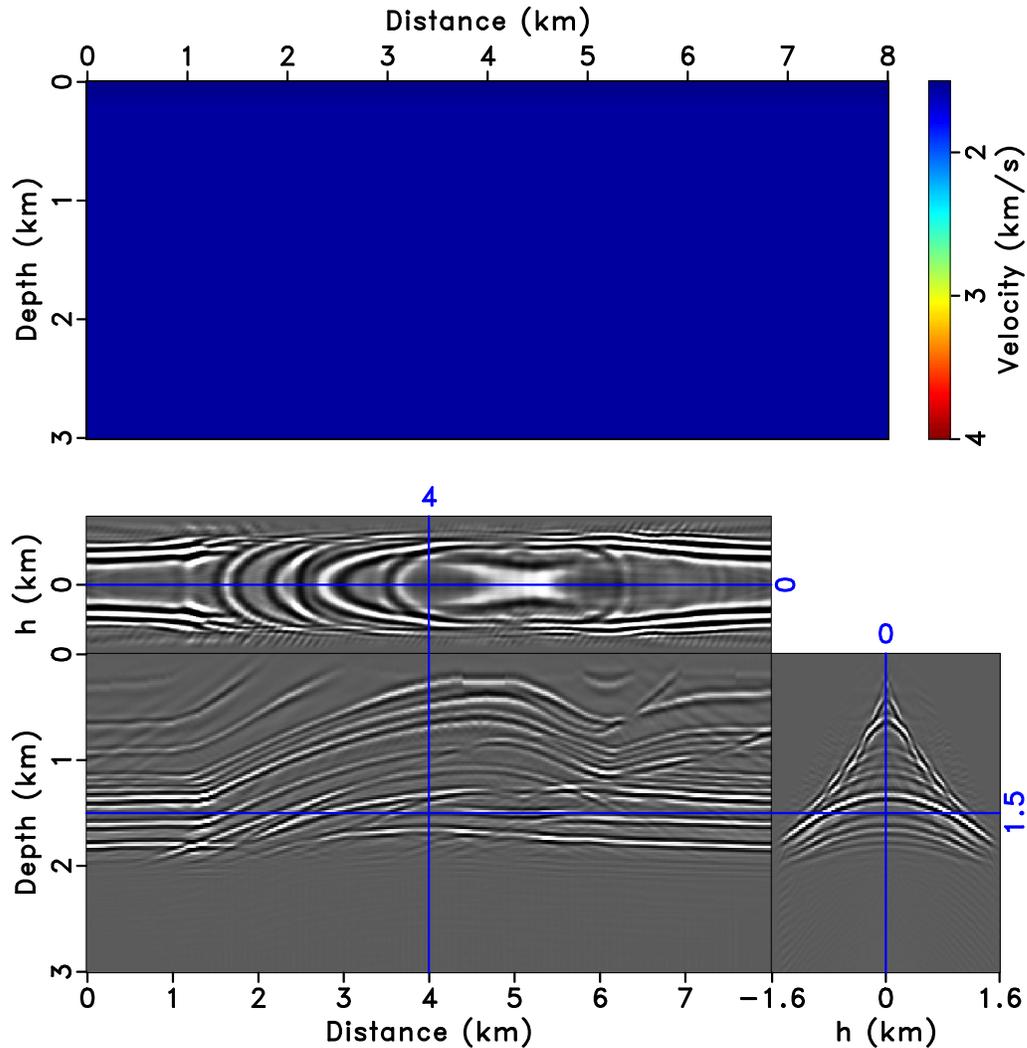
Figure 5: (a) Initial background velocity; (b) Extended reflectivity

The inverted velocity model from 20 iterations and the corresponding extended image are shown in Figure 4a and Figure 6b. The anticline and reverse fault structures can be clearly observed. Furthermore, even the reflector beneath the anticline is imaged correctly. The velocity errors are mostly at the edges and bottoms, which is a result of imperfect illumination. The subsurface gathers are mostly

focused towards $h = 0$ after inversion.

The gathers are much more focused towards $h = 0$ after inversion (see Figure 7a and 7b). For comparison, the CIG without angle taper after inversion is shown in 7c. With the help of our angle filter, these artifacts are eliminated effectively.

Marmousi model

The basic information is listed in Table 2. The true velocity and reflectivity model are shown in Figure 8.

Parameter	Measurements
Source wavelet	bandpass $5 - 20\,Hz$
Source position $\mathbf{x_s}$	$x : -2 - 11.2\,km$ every $80\,m$, $z = 0\,m$
Receiver position $\mathbf{x_r}$	$x : -2 - 11.2\,km$ every $40\,m$, $z = 0\,m$
Record time	$t = 3.6\,s$
Grid size	$dx = dh = dz = 20\,m$, $dt = 2\,ms$
Initial velocity	$v = 1.5\,km/s$

Table 2: Marmousi model

Figure 9a shows the initial background velocity is constant ($1.5 : km/s$) and far away from the correct one, so the geological structures can barely be observed from the extended image

Figure 10a and 10b are the inverted velocity model after 60 iterations and the corresponding migrated extended image. After inversion, the gathers are much more focused towards $h = 0$(Figure 11).

CONCLUSION

As a result of the peak amplitude from the large incidence angles, the subsurface offset images contain many artifacts, which introduce biases in the objective function and its gradient. Aiming to remove those artifacts, we propose an angle
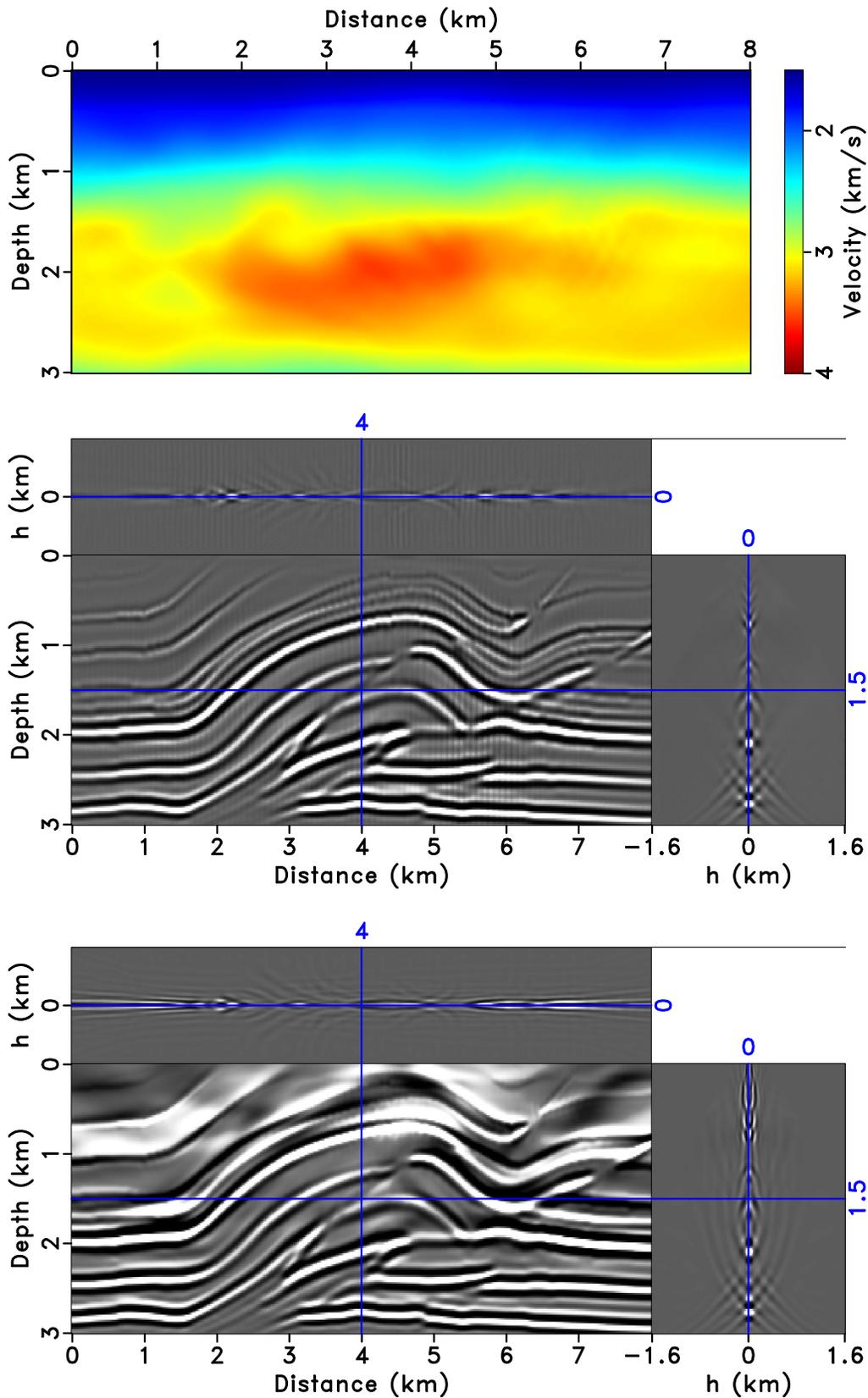
Figure 6: Inversion results after 20 iterations. (a) Background velocity; (b) Extended image; (c) Extended image without angle taper.
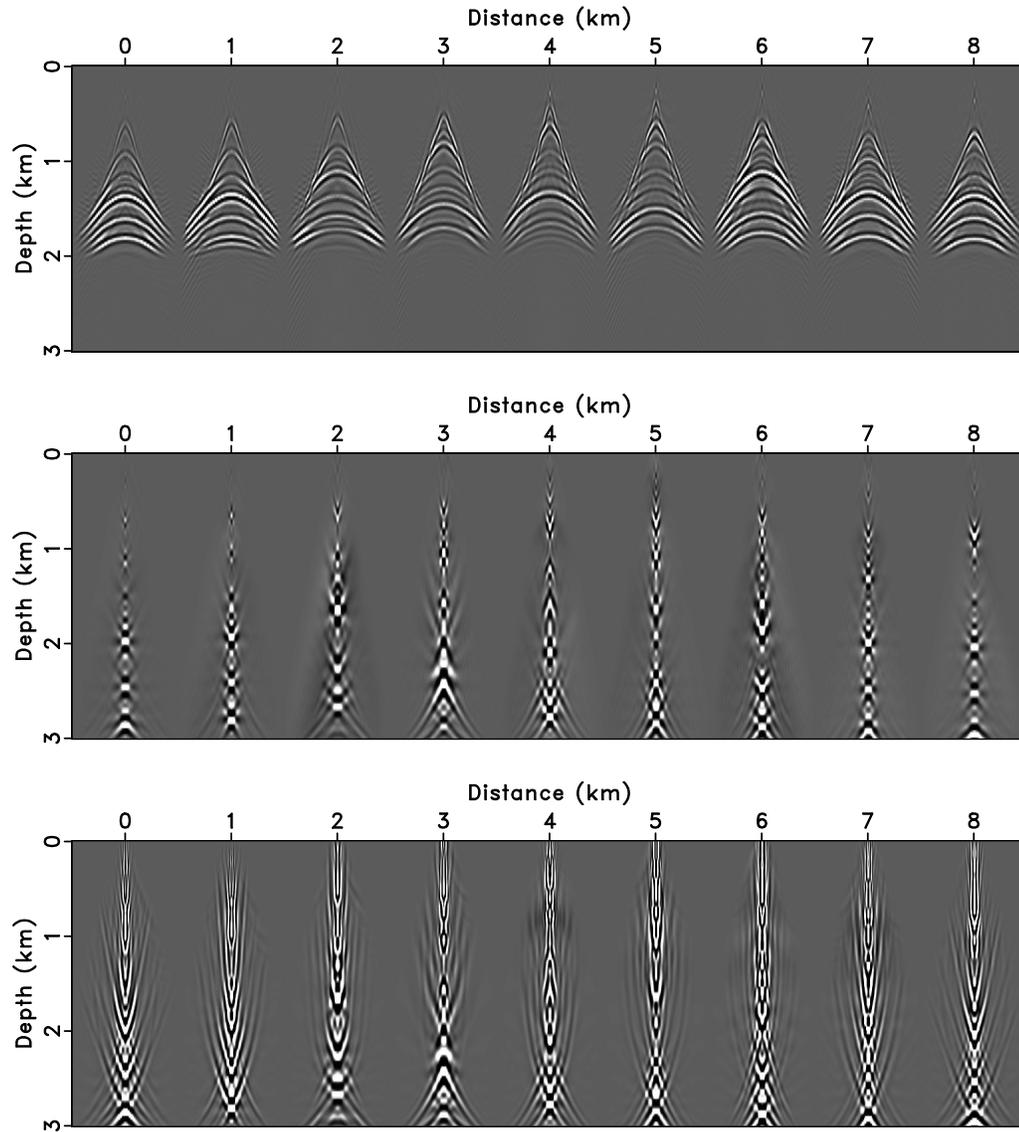
Figure 7: (a) Subsurface offset gathers every 1 *km* using initial background velocity; (b) Subsurface offset gathers using inverted background velocity; (c) Subsurface offset gathers using inverted background velocity without angle taper.
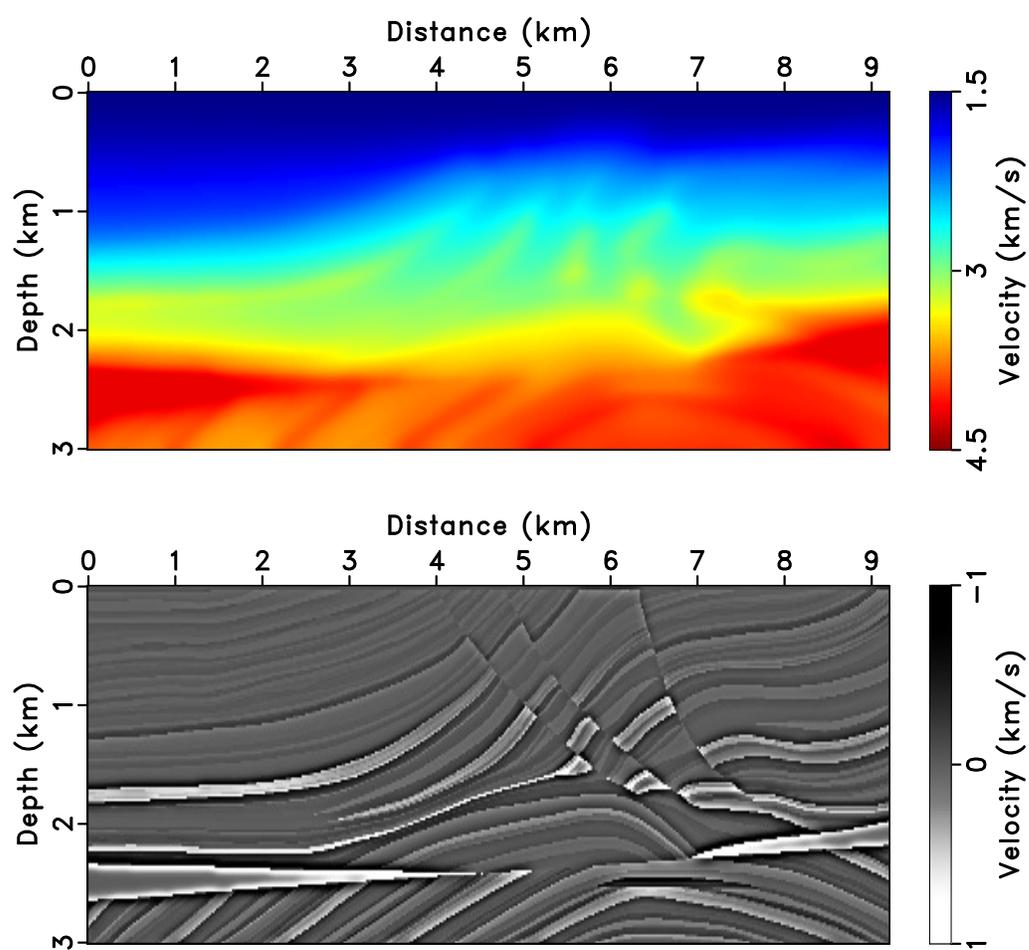
Figure 8: (a) True background velocity model; (b) Reflectivity model.
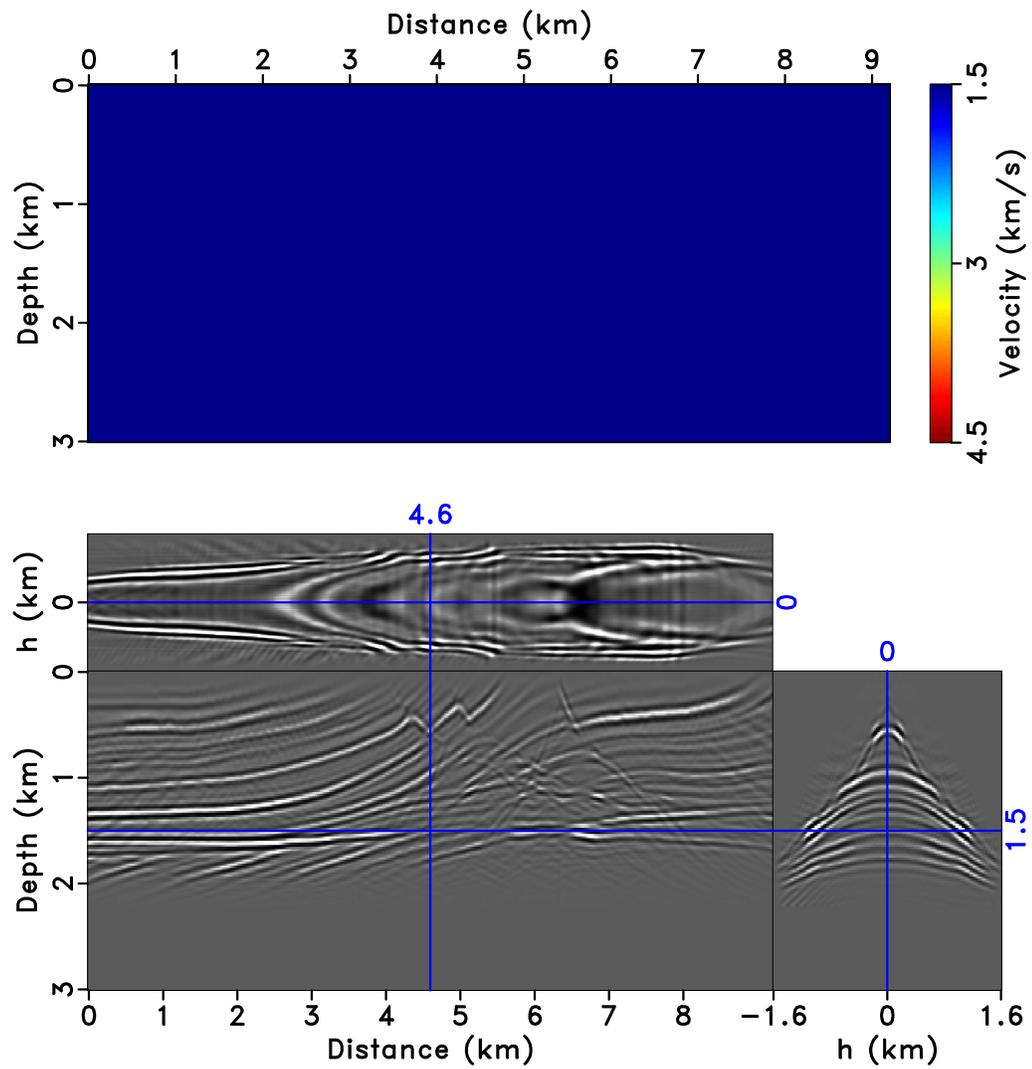
Figure 9: (a) Initial background velocity; (b) Extended reflectivity
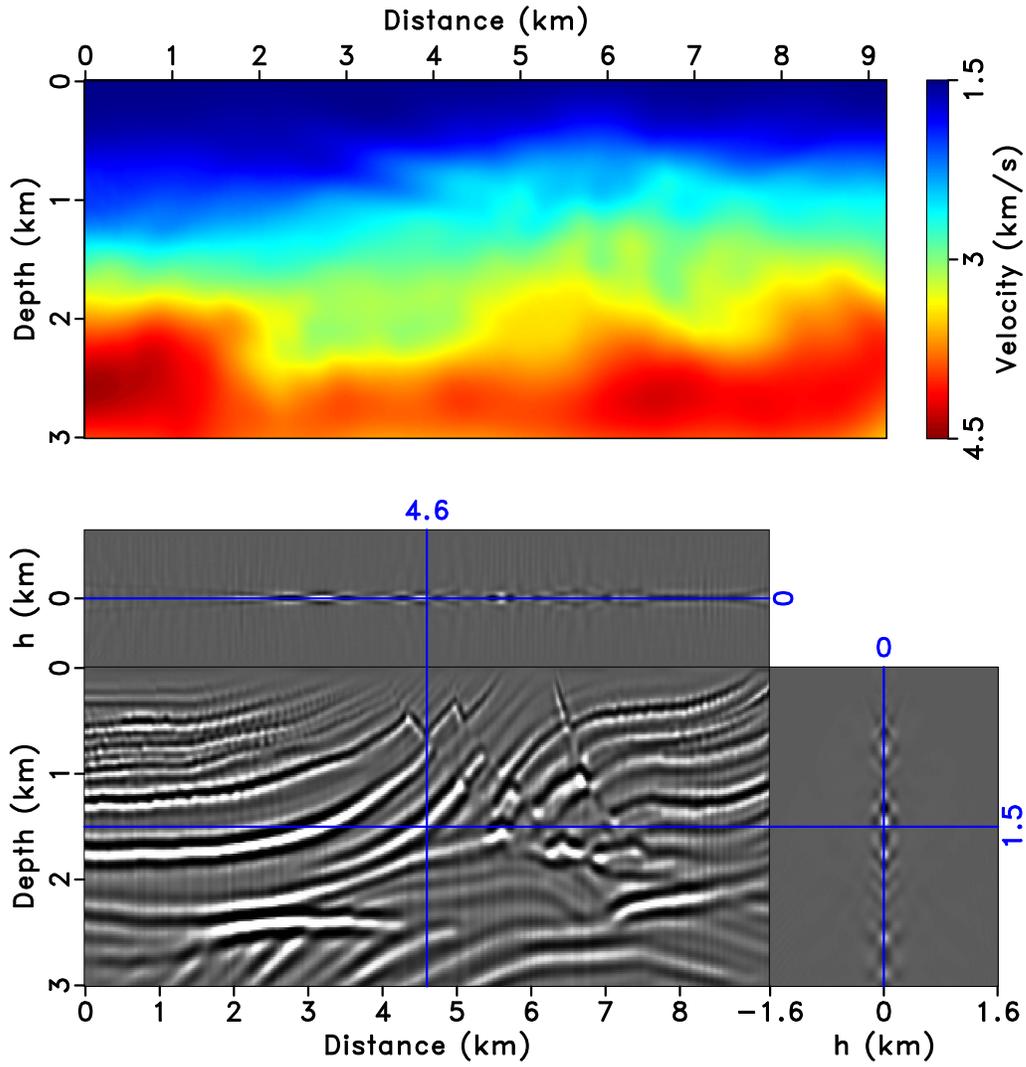
Figure 10: (a) Inverted background velocity after 86 iterations; (b) Extended reflectivity
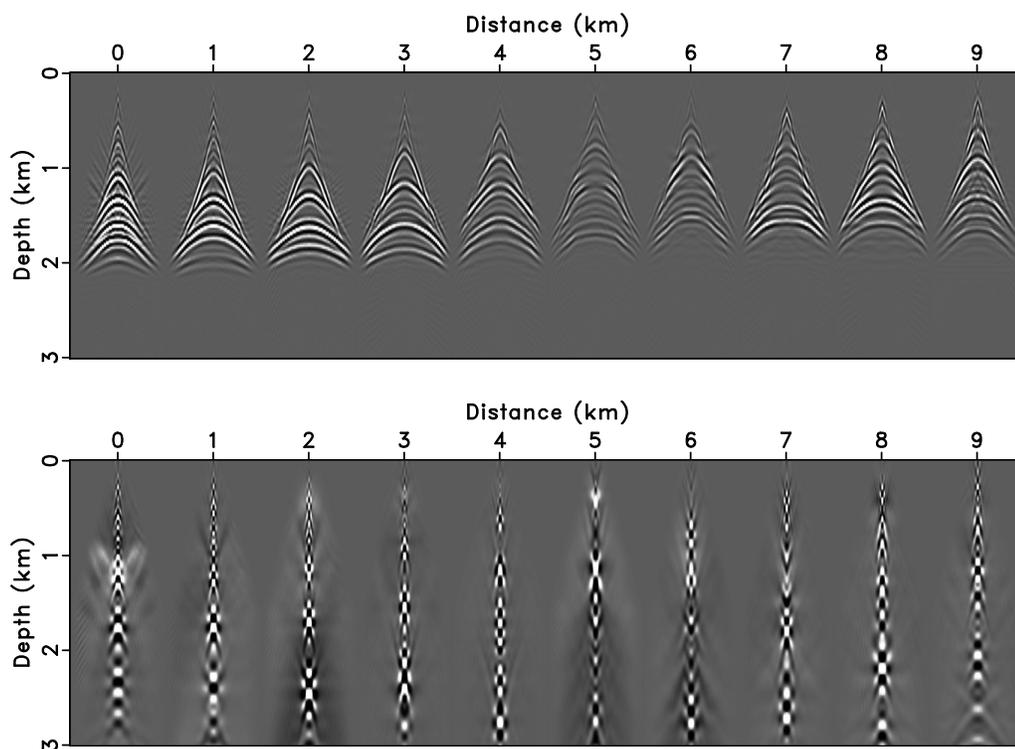
Figure 11: (a) Subsurface offset gathers at $x = 2, 3, 4, 5, 6\,km$ using initial background velocity; (b) Subsurface offset gathers using inverted background velocity

taper, which depends on acquisition geometry and imaging point. 2D synthetic examples show migration velocity analysis becomes more robust with application of the proposed new method.

REFERENCES

Almomin, A., B. Biondi, et al., 2012, Tomographic full waveform inversion: Practical and computationally feasible approach: Presented at the 2012 SEG Annual Meeting, Society of Exploration Geophysicists.

——, 2014, Preconditioned tomographic full waveform inversion by wavelength continuation: Presented at the 2014 SEG Annual Meeting, Society of Exploration Geophysicists.

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, Expanded Abstracts, Society of Exploration Geophysicists, SI9.5.

Biondi, B., and G. Shan, 2002, Prestack imaging of overturned reflections by reverse time migration: 72nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1284–1287.

Biondi, B., and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Cao, D., S. Singh, and A. Tarantola, 1990, Simultaneous inversion for background velocity and impedance maps: Geophysics, **55**, 458–469.

Chauris, H., and M. Noble, 2001, Two-dimensional velocity macro model estimation from seismic reflection data by local differential semblance optimization: applications synthetic and real data sets: Geophysical Journal International, **144**, 14–26.

Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.

——, 1985, Imaging the earth's interior: Blackwell Scientific Publishers.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538.

Fu, L., and W. Symes, 2015, Reducing the cost of extended waveform inversion by multiscale adaptive methods: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1127–1131.

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

Kern, M., and W. Symes, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysical Prospecting, **99**, 565–614.

Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations, *in* Conference on Inverse Scattering: Theory and Applications: Society for Industrial and Applied Mathematics, 206–220.

———, 1984, Migration methods: partial but efficient solutions to the seismic inverse problem, *in* Inverse Problems of Acoustic and Elastic Waves: Society for Industrial and Applied Mathematics.

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.

Mulder, W., 2014, Subsurface offset behaviour in velocity analysis with extended reflectivity images: Geophysical Prospecting, **62**, 17–33.

Mulder, W., and F. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184–1191.

Nolan, C., and W. Symes, 1997, Global solution of a linearized inverse problem for the wave equation: Communications on Partial Differential Equations, **22**, 919–952.

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.

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, Expanded Abstracts, Society of Exploration Geophysicists, 930–934.

Plessix, R.-E., Y.-H. de Roeck, and G. Chavent, 1999, Waveform inversion of reflection seismic data for kinematic parameters by local optimization: SIAM Journal on Scientific Computation, **20**, 1033–1052.

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

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

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

Shen, P., W. Symes, and C. C. Stolk, 2003a, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2135–2139.

Shen, P., W. W. Symes, and C. C. Stolk, 2003b, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, SEG, Expanded Abstracts, 2131–2134.

Shin, C., and D.-J. Min, 2006, Waveform inversion using a logarithmic wavefield: Geophysics, **71**, R31–R42.

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.

——, 2006, Seismic inverse scattering in the downward continuation approach: Wave Motion, **43**, 579–598.

Stolk, C. C., and W. Symes, 2004, Kinematic artifacts in prestack depth migration: Geophysics, **69**, 562–575.

Sun, D., and W. Symes, 2012, Waveform inversion via nonlinear differential semblance optimization: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SI3.3.

Symes, W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

Symes, W., and M. Kern, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysical Prospecting, **in press**.

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.

A discrepancy based penalty method for extended waveform inversion

Lei Fu and William W. Symes

ABSTRACT

Extended waveform inversion globalizes the convergence of seismic wave-form inversion by adding non-physical 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 non-physical (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 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 illustrate 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 is able to yield high-resolution images of sub-surface 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 full waveform inversion (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 S/N. These attributes are available for some surveys but not for others (Plessix et al., 2010). Without them, 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 non-physical degrees of freedom, in such a way that data fit may be achieved (hence cycle-skip avoided) throughout the inversion, even with grossly inaccurate initial model (Symes, 1986; Symes and Carazzone, 1991; Symes, 2008). Since the additional degrees of freedom are non-physical, they must be suppressed if the extended model is to converge to a solution of the full waveform inversion 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 (non-physicality) 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 behaviour 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 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).

Since 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 fundamentally from well established methods such as the L-curve method

(Hansen, 1992), generalized cross-validation (Wahba, 1977), and Morozov's discrepancy principle (Morozov, 1984; 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 should be chosen so that the minimizer is close to an assumed noise-free solution. For example, Morozov's discrepancy principle (Morozov, 1984; 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). Ajo-Franklin et al. (2007), Sen and Roy (2003); Roy (2002, 2005) discuss various methods for choosing regularization parameters in geophysical inverse problems, Morozov's discrepancy principle amongst 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 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 natu-

ral 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. Since 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, see Kern and Symes (1994) and Huang and Symes (2015) for instance.

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 Biondi and Almomin (2014); Weibull and Arntsen (2013); Lameloise et al. (2015a), also Symes (2008) for a review of older work. Plessix (2000), Luo and Sava (2011), Warner and Guasch (2014), and van Leeuwen and Herrmann (2013) 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 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 follow: 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. 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, 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 model and data spaces, distinguishing the (possibly different) norms by context. The data error and model penalty are both norms-squared:

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

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

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 non-physicality of the inner parameter x. Physical inner parameter vectors lie in its

null space. As explained in the introduction, the object of the optimization is to remove the non-physical extended degrees of freedom from the model while maintaining data fit. Reducing p as defined in equation 6 will move the model towards 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 good data fit. When $\alpha \to \infty$, minimization of J_α forces the extended model to be close to physical one, so that the optimization approximates non-extended or physical inversion.

Variable projection method

As mentioned in the introduction, 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

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

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

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

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 simply treat the normal operator as if it were positive definite. The system 6 thus has a unique solution that can be approximated via an iterative method such as conjugate gradient (CG) iteration. Since m and α determine the operator on the LHS 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, LBFGS 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

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

(Golub and Pereyra, 1973). The notation DF signifies the derivative of F. Since 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,

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

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

The Discrepancy Principle

As described in the introduction, 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: since 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 understanding how to reset α when m changes.

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

$$e(\alpha) \;=\; e[m, x[m, \alpha]] \qquad (9)$$
$$p(\alpha) \;=\; p[m, x[m, \alpha]] \qquad (10)$$

Differentiating equation 6 with respect to α leads to the relation

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

whence

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

Note that the inequality in equaion 12 is *strict* if $p > 0$ hence $A^T A x \neq 0$, since 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 a strict inequality if $p > 0$.

Equation 12 together with equation 13 show that increasing α implies increasing e while decreasing p, and

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

In view of equation 12,

$$\frac{de}{d\alpha} \leq 2p. \tag{15}$$

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,

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

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

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

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

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

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

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

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_+. The 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, 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 (i) treat it as hypothetical, with all subsequent results being contingent on it, and choose initial model $(m, x[m, 0])$ for $\alpha = 0$ so that $X = e(0) = e[m, x[m, 0]]$; (ii) if the normal equation 6 is solved approximately by an iterative method (we used conjugate gradient iteration (CG)), choose a number of 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 indirect rather than direct choice of the hypothetical noise level X. Either approach make 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 non-zero 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 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 9; for each update of m, a solution of equation 6 is required to re-compute $x[m, \alpha]$.

   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 to the simple bisection loop (steps 1.3.2.1, 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, as the secant update is usually successful. Since each alpha update requires a solution of the normal equation 6, this is a good thing.
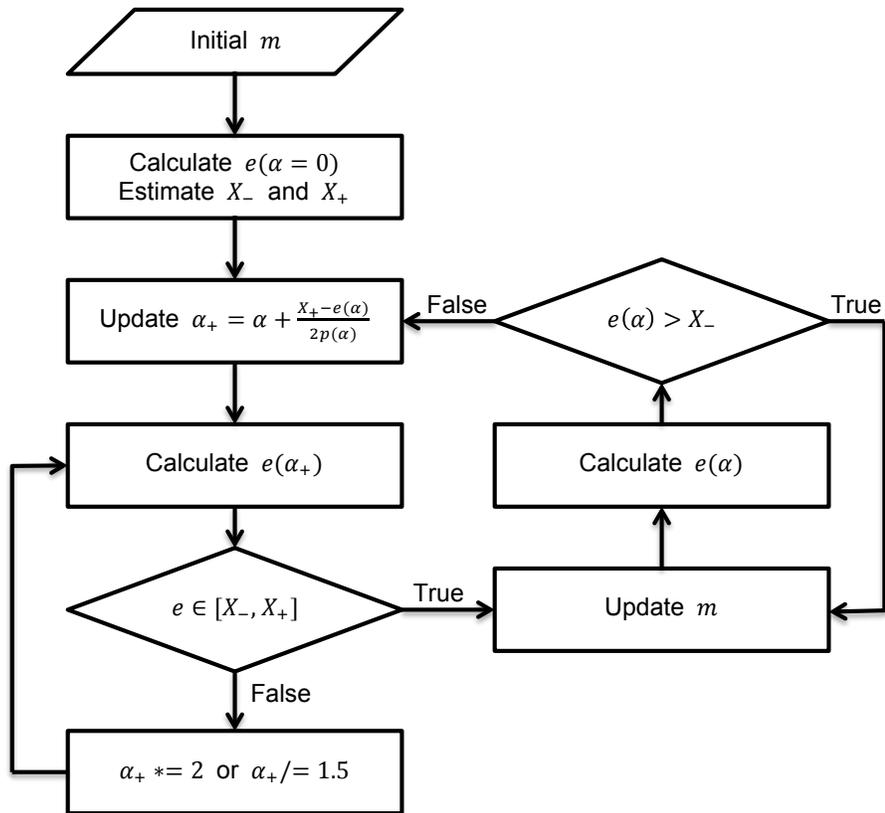
Figure 1: A flowchart for our implementation of our proposed algorithm.

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:

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

$$u = 0, t \ll 0 \tag{21}$$

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

$$\delta u = 0, t \ll 0 \tag{22}$$

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-hand side of equation 1 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

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

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 (a) the background velocity is transparent, that is, contains no reflectors, and (b) 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 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

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

Since $h\delta(h) = 0$, 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 non-physical dimension ($h \neq 0$) and has been used in many prior works on subsurface offset extended waveform inversion (Shen et al., 2003; Symes, 2008; Shen and Symes, 2008; Biondi and Almomin, 2012; Weibull and Arntsen, 2013; Lameloise et al., 2015b).

In numerical implementation of the wave equations 1 and 22, we use a centered finite difference method of order 2 in time and 8 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 the gradient calculation 9. 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 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 macro model, shown in Figure 2a. The difference between the original model (with water layer and padding) and the background model constitutes the reflectivity model, shown in Figure 7a. In the reflectivity model, horizontal layers are distorted by several thrust (reverse) faults (shown in Figure 7a). 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.

Parameter	Measurements
Source wavelet	bandpass $5 - 20\,Hz$
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

Table 1: Parameters for overthrust model example

Figure 3a shows the initial background velocity is constant ($1.5\,km/s$) and far away from the correct one, so the geological 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 relative data residual. Based on the value of $e(0)$, the upper and lower bound of accept range are estimated by formula $X_- = 0.7^2 e(0)$ and $X_+ = 1.2^2 e(0)$. Then the updated weight $\alpha_1 = 1.1e - 6$ is estimated by equation 18. After 1 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 update algorithm, the value of α was doubled three time, which gives data residual in acceptable range. After 3 more velocity updates with weight $\alpha = 3.2e - 5$, data residual became smaller than X_-. Recalculating α and doubling its value made the data residual back to the acceptable range. After 9 more ve-
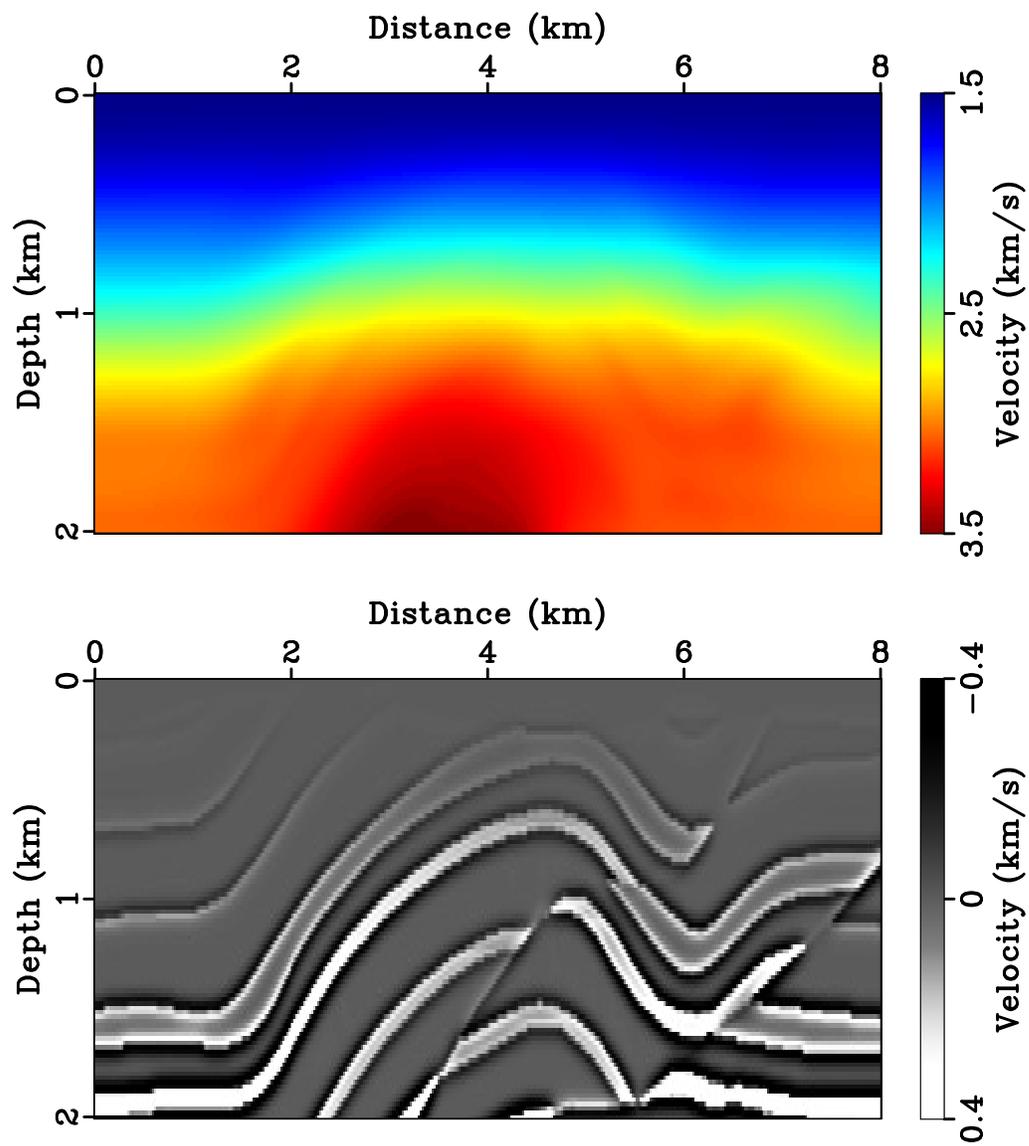
Figure 2: (a) Target background velocity model; (b) Target reflectivity model.
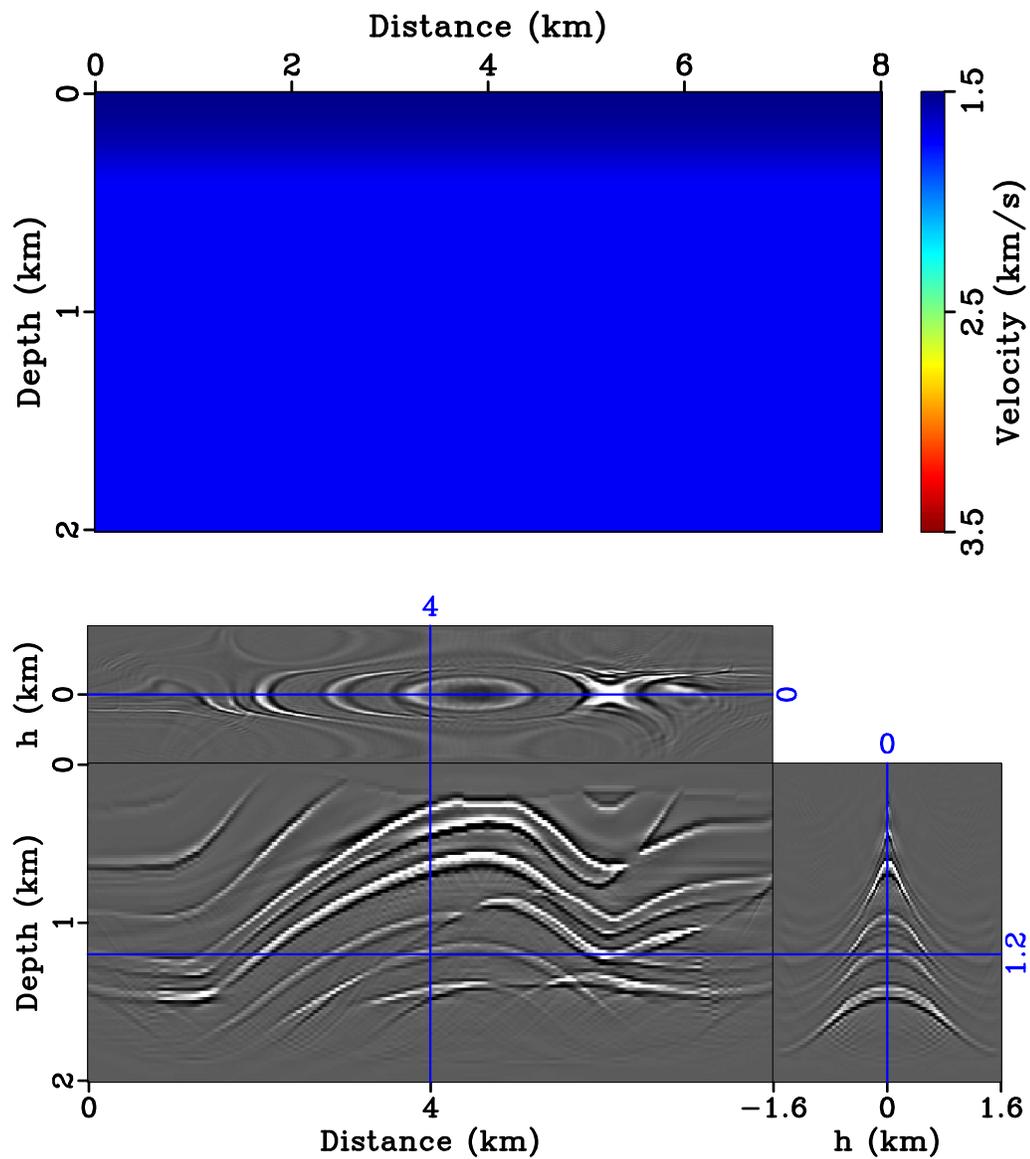
Figure 3: (a) Initial background velocity; (b) Extended reflectivity

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

The subsurface offset gathers are much more focused towards $h = 0$ after inversion (Figure 5). With the inverted background velocity, the geological structures are imaged with much higher resolution at 0-offset section of the reflectivity model (Figure 6). Extending the reflectivity model permits good data fit throughout the inversion process (Figure 7 and 8). Note that the data misfit before and after inversion is measured with different penalty weight α. In other words, good data fit is obtained by models with different amount of extension. Note that in order 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, 2017).

We increase the penalty weight α at iteration number 1, 4 and 13, which significantly accelerates the convergence rate as shown by Figure 9. At the same time, according to 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, computed with the extended reflectivity model. As these are non-physical, 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 both the initial velocity (Figure 3a) and the inverted velocity in Least Squares Migration, which amounts to solving the ptoblem 6 with the constraint that the reflectivity be physical, that is,

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

We used 20 Conjugate Gradient iterations to approximate these minimizations, as we have in other instances of the problem 6. The resulting predicted data appear as Figures 11a, 11b, and the data residuals as Figures 12a, 12b, all plotted on the same grey scale. While the residual is clearly not as small as that attained by the
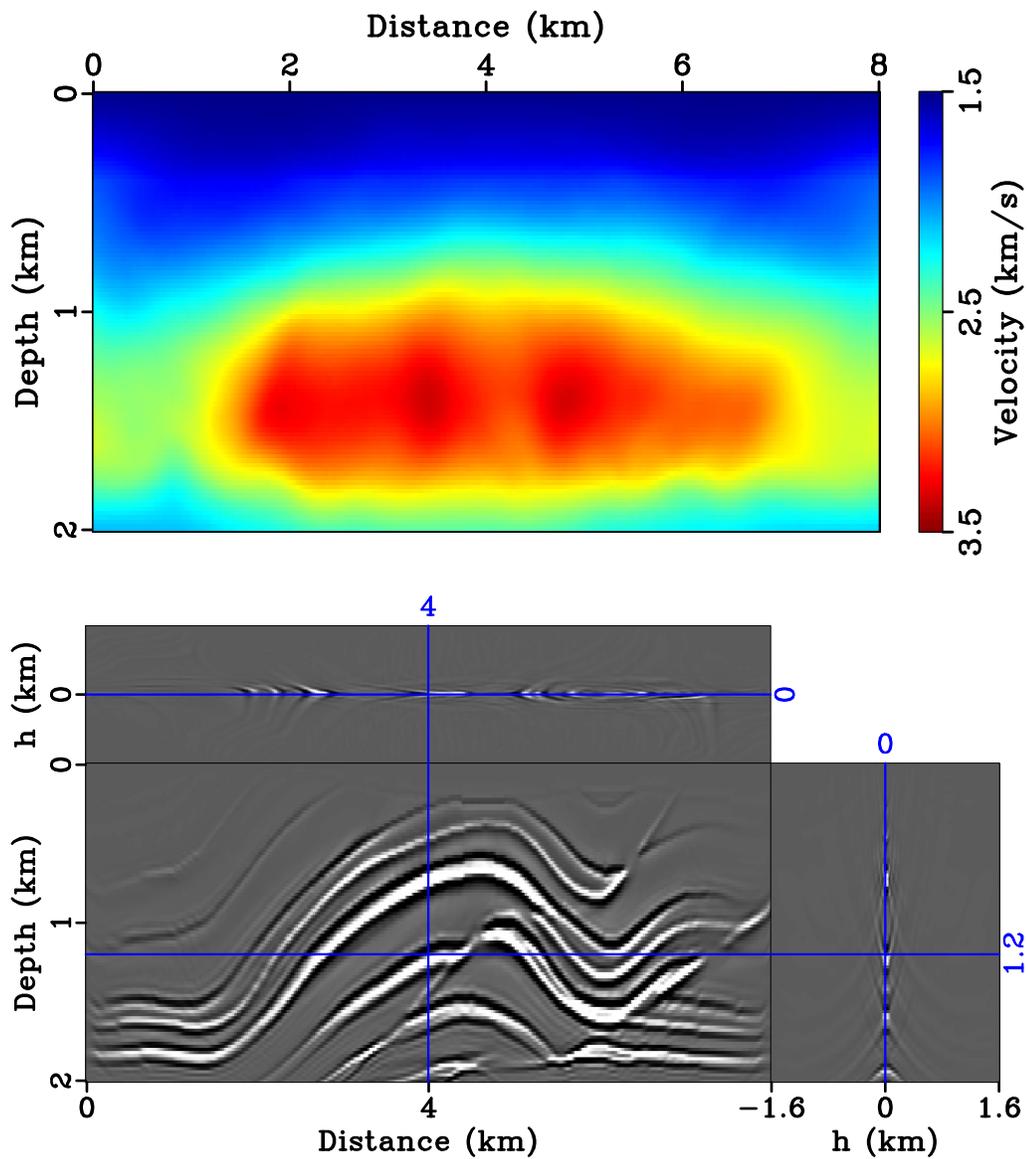
Figure 4: (a) Inverted background velocity; (b) Extended reflectivity
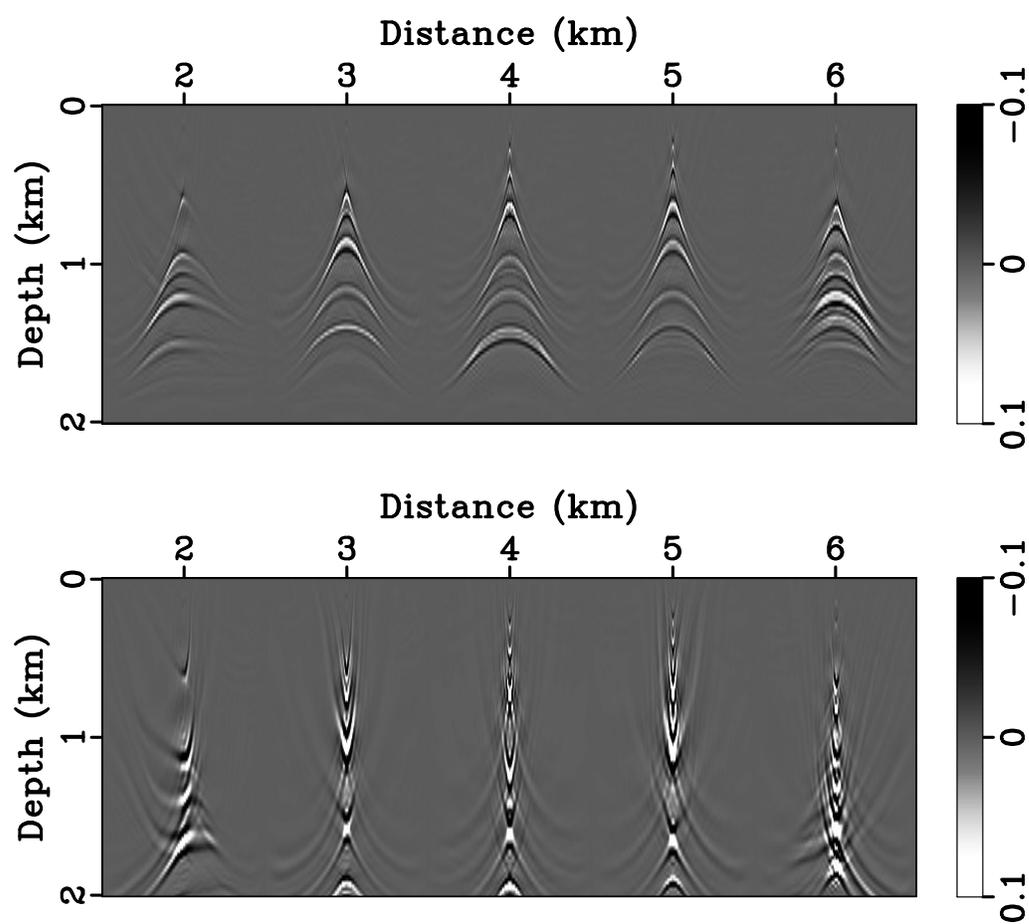
Figure 5: (a) Subsurface offset gathers at $x = 2, 3, 4, 5, 6\ km$ using initial background velocity; (b) Subsurface offset gathers using inverted background velocity
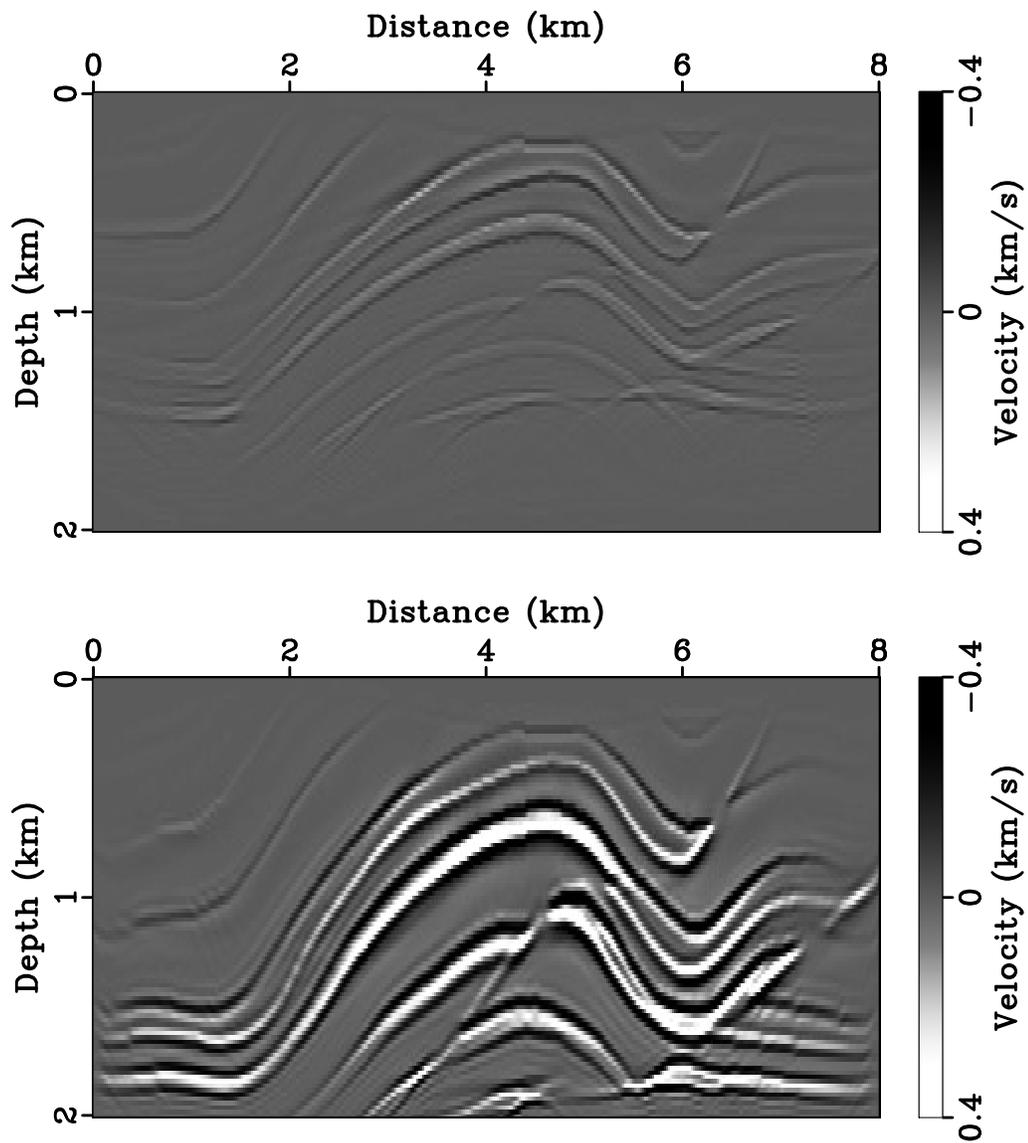
Figure 6: (a) Zero offset section with the reflectivity model using initial background velocity. (b) Zero offset section with the reflectivity model using inverted background velocity.
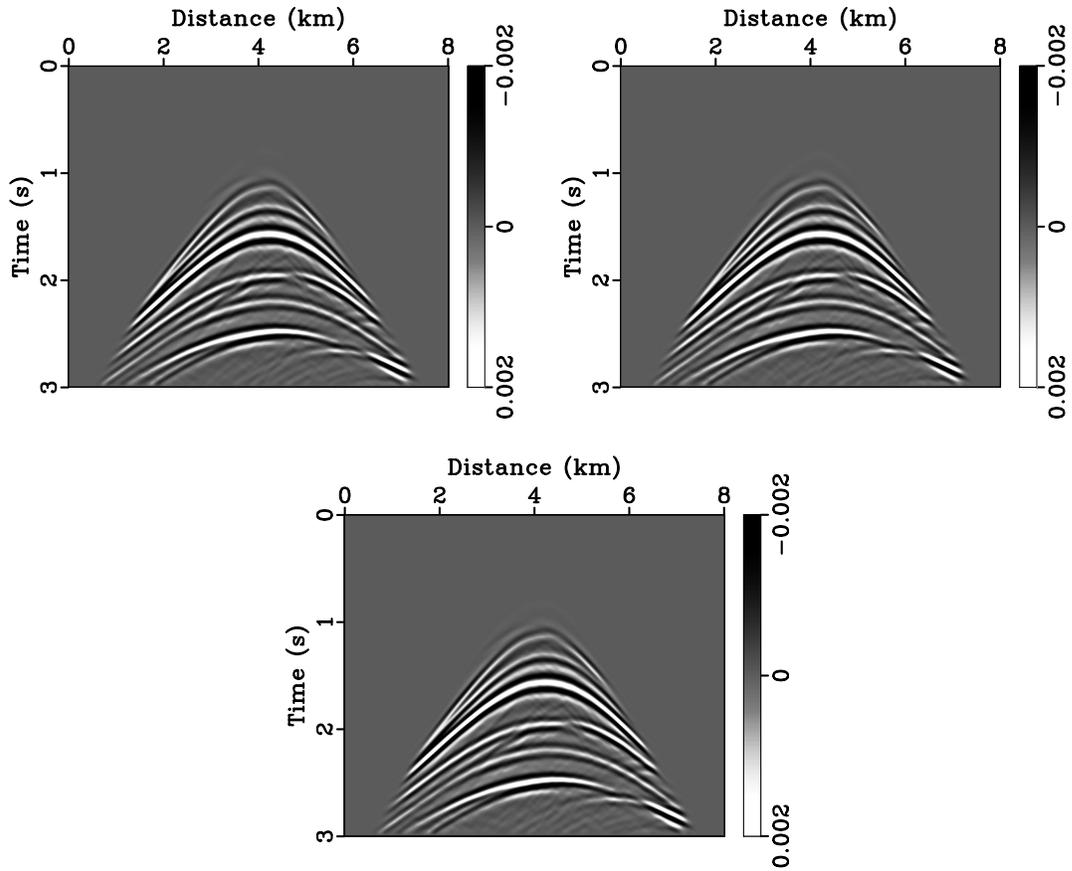
Figure 7: Data gather for a shot at the center of the model (a) Observed data. (b) Predicted data with initial background velocity model. (c) predicted data with inverted background velocity model.
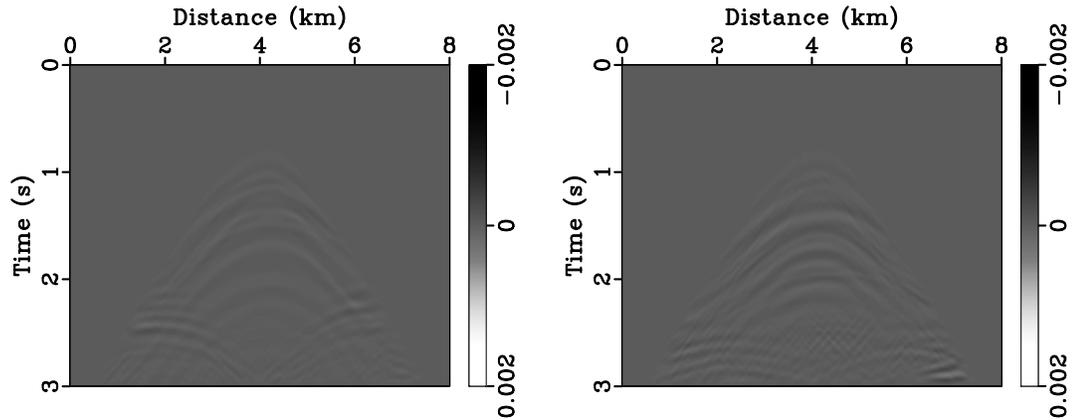
Figure 8: Comparison of data residual (a) with initial background velocity model and (b) with inverted background velocity model

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 Figures 14a 14b, and 14c the traces of the target data (in black) vs. 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 Figures 15a 15b, and 15c the traces of the target data (in black) vs. 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 (for example, the 4 km trace at about 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 Full Waveform Inversion. The potential use of extended waveform
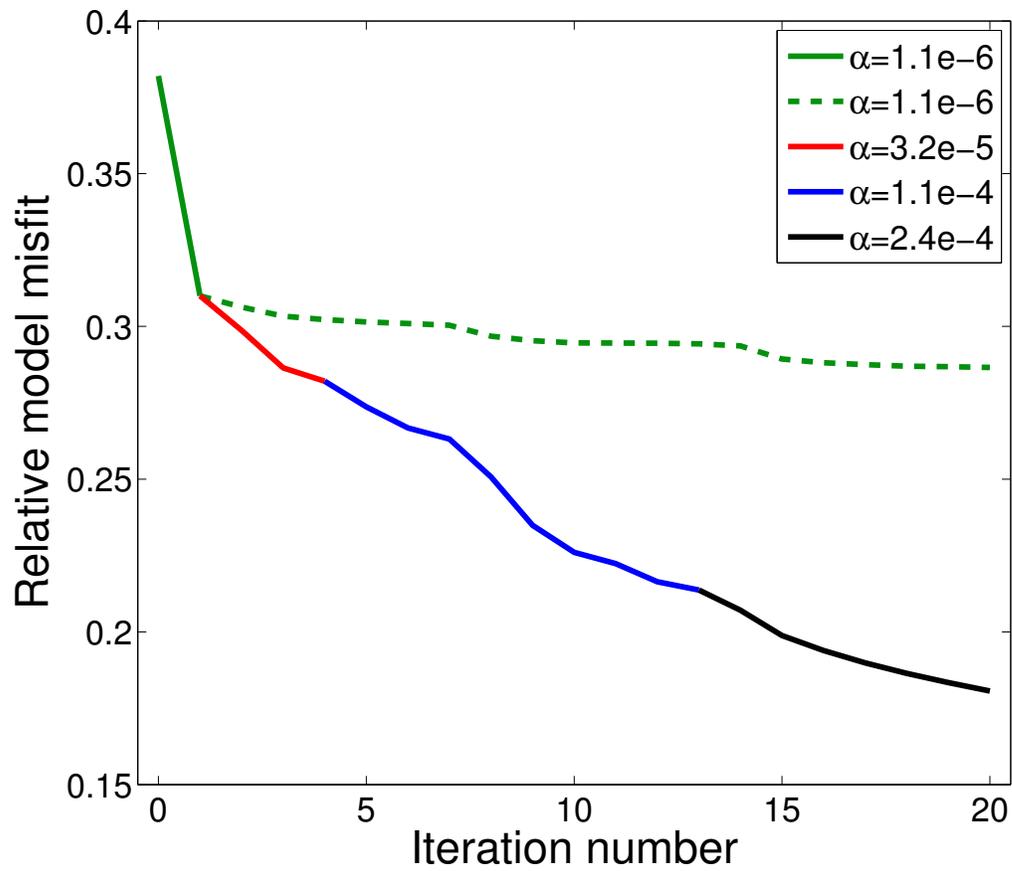
Figure 9: Model convergence rates versus iteration numbers: different colors represent different values of weight α; dashed lines show convergence rate without changing α.
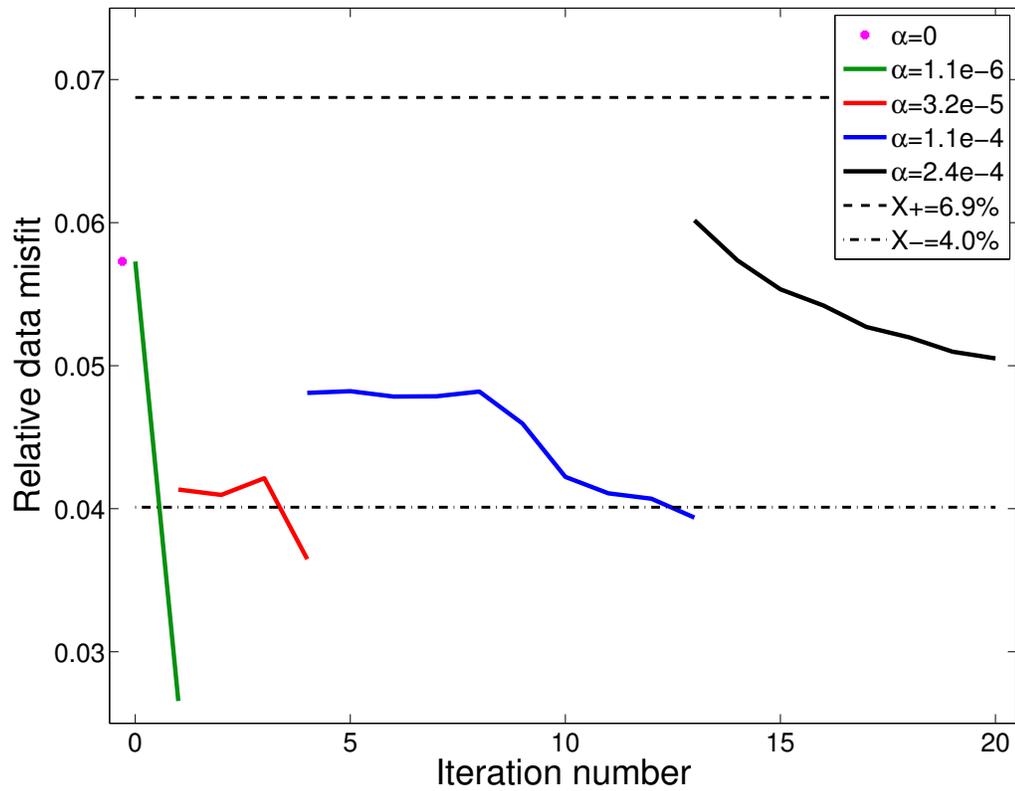
Figure 10: Relative data misfit versus iteration: different colors represent different values of weight α

inversion to produce initial models for Full Waveform Inversion may be a worth-while 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 Figures 13a and 13b. Comparison with Figure 7a suggests that while 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.
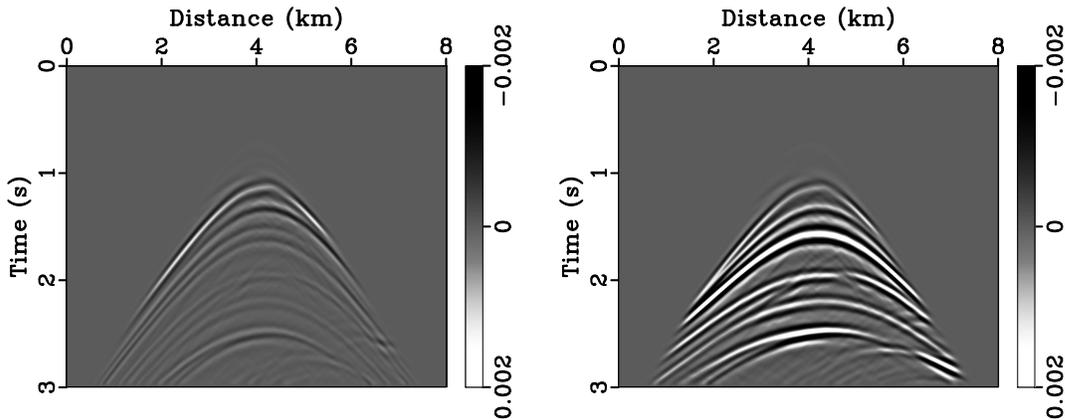
Figure 11: Least Squares Migration predicted data for same shot as shown in Figure 7a. Note that these are non-extended inversions, and not by-products of the inversion algorithm discussed in the text. Produced with 20 Conjugate Gradient iterations. (a) Using initial velocity model (Figure 3a). (b) Using inverted velocity model (Figure 4a).

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 used the variable projection method to estimate *both v* and r_0 (Note that the Least Squares Migration exercise reported in the previous paragraph also constrained r by the condition 25, but did not itself update v, relying instead on the inversion algorithm based on subsurface offset extension to do that). The second term in the definition 4 of J_α is irrelevant, as r satisfying equation 25 also
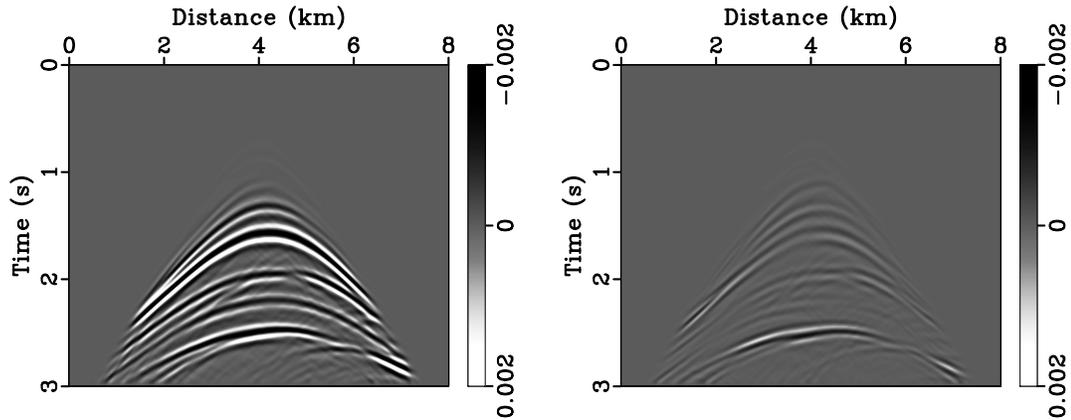
Figure 12: Least squares Migration data residuals for same shot as shown in Figure 7a. Note that these are non-extended inversions, and not by-products of the inversion algorithm discussed in the text. Produced with 20 Conjugate Gradient iterations. (a) Using initial velocity model (Figure 3a). (b) Using inverted velocity model (Figure 4a).

satisfy $Ar(x, z, h) = r_0(x, z)(h\delta(h)) = 0$. 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 "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 Conjugate Gradient 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 non-extended penalty function, since $p = 0$. That is, the same amount of computational effort was expended in this non-extended 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 is shows that Born waveform inversion is just as susceptible to cycle-skipping as is conventional FWI, based on the nonlinear for-
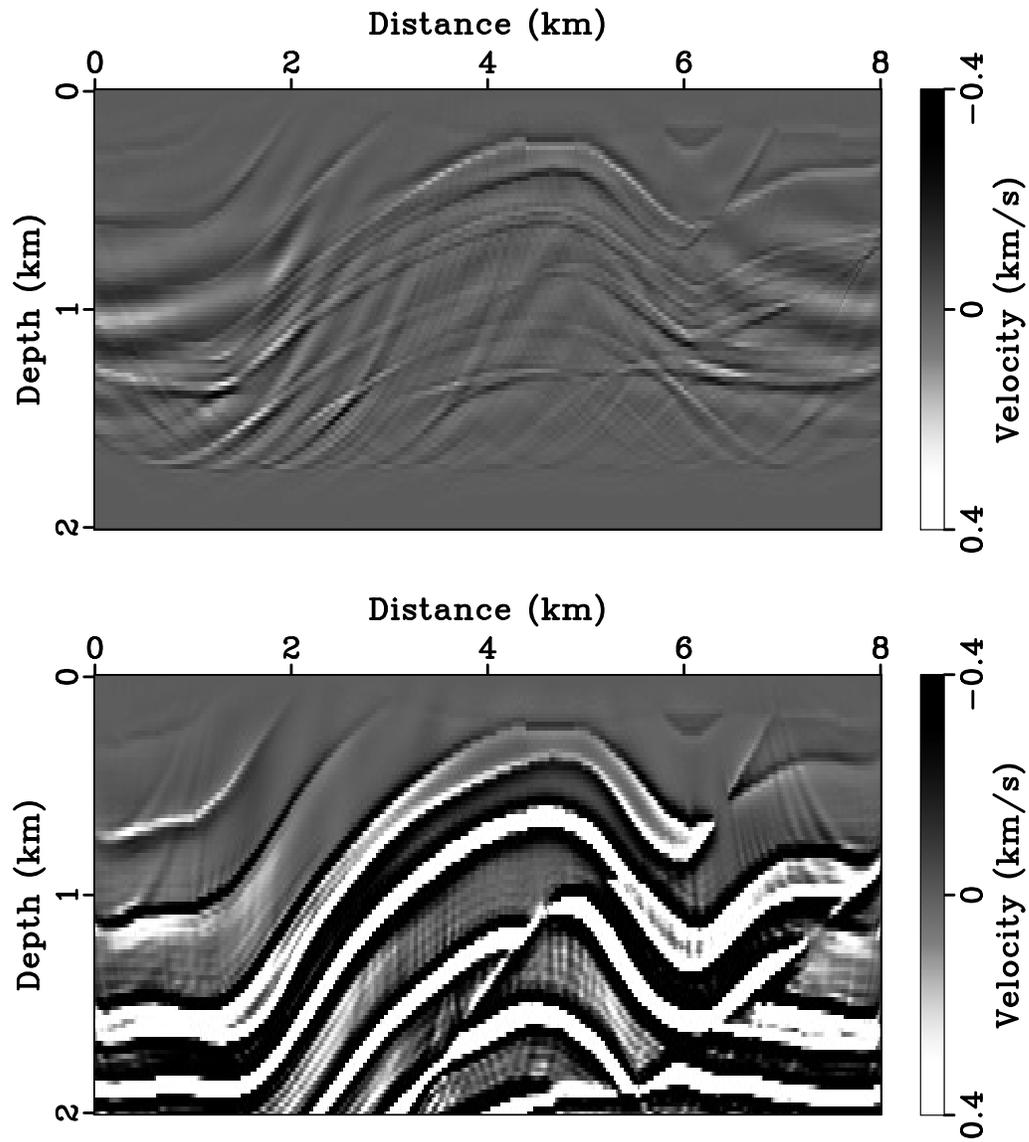
Figure 13: Least Squares Migration images corresponding to Figures 11a, 11b.
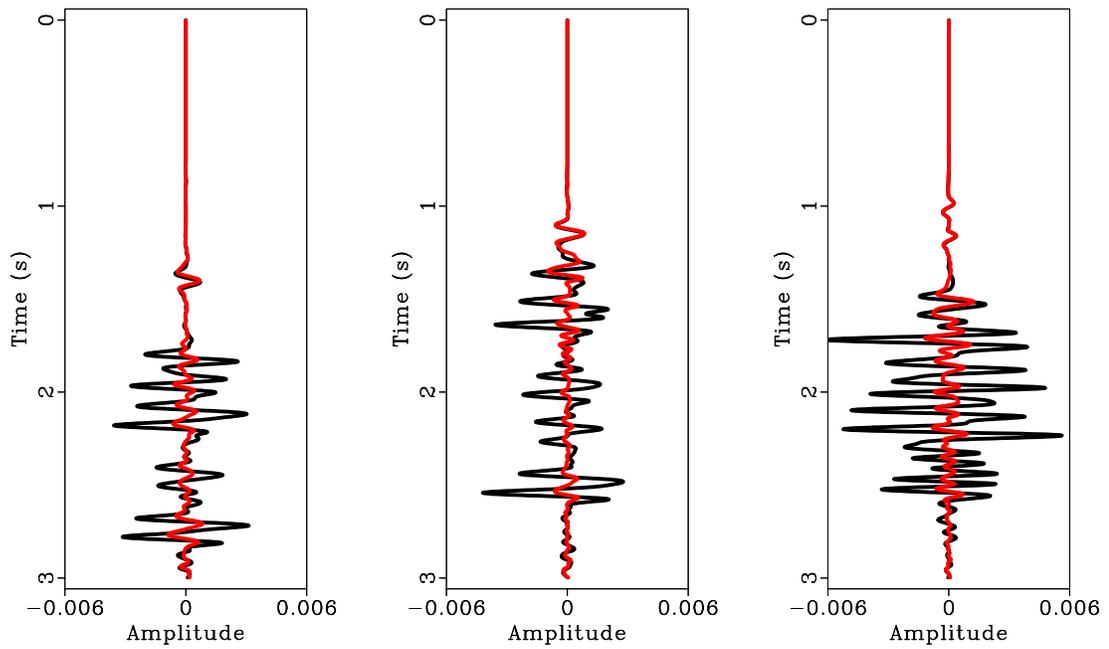
Figure 14: Traces at positions 2 km, 4 km, and 6 km from the shot gather shown in Figure 7a (black), plotted against the same traces from the predicted gather (Figure 11a) using the initial velocity model and the Least Squares Migration reflectivity estimate (Figure 13a) (red).
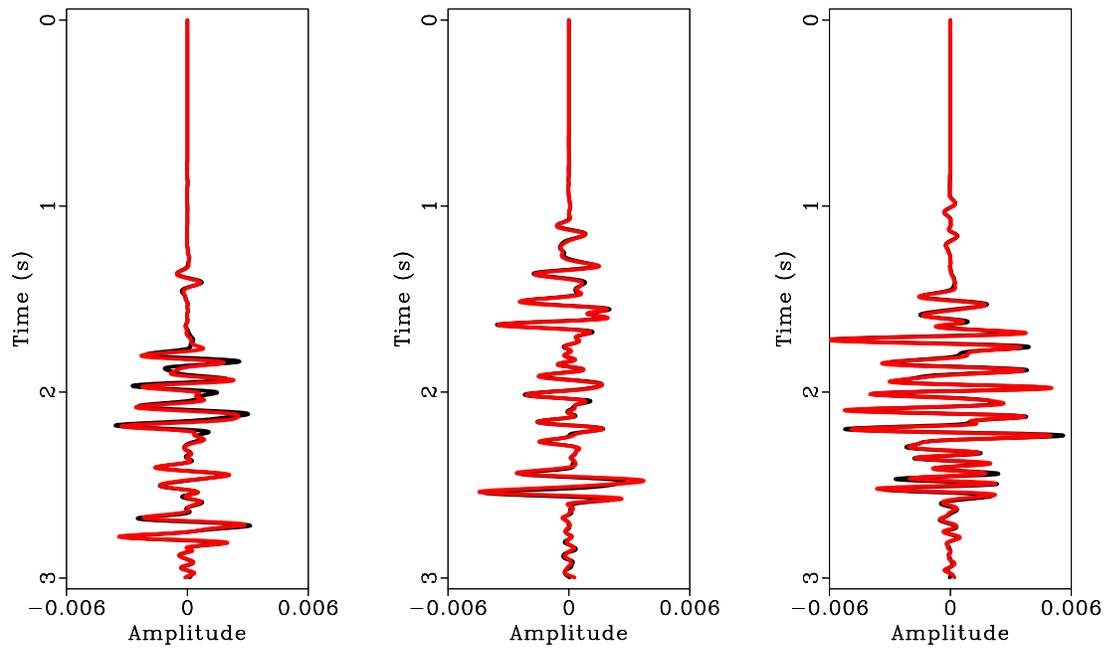
Figure 15: Traces at positions 2 km, 4 km, and 6 km from the shot gather shown in Figure 7a (black), plotted against the same traces from the predicted gather (Figure 11b) using the inverted velocity model (Figure 4a) and the Least Squares Migration reflectivity estimate (Figure 13b) (red).

ward map.

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 behaviour, we present a plot of the reduced objective $J_\alpha[m, x[m, \alpha]]$, for models m on the line segment $m = (1 + \sigma)m_0$, $-0.5 \leq \sigma \leq 0.5$, 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 one-dimensional model space, the behaviour 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 choice of initial model. Our discrepancy based algorithm provides a key ingredient in this procedure, not obvious from plots like Figure 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 (conjugate gradient 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, 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 rootfinding algorithms to ensure convergence. Also, experience suggests that modest accuracy in estimating the inner component is
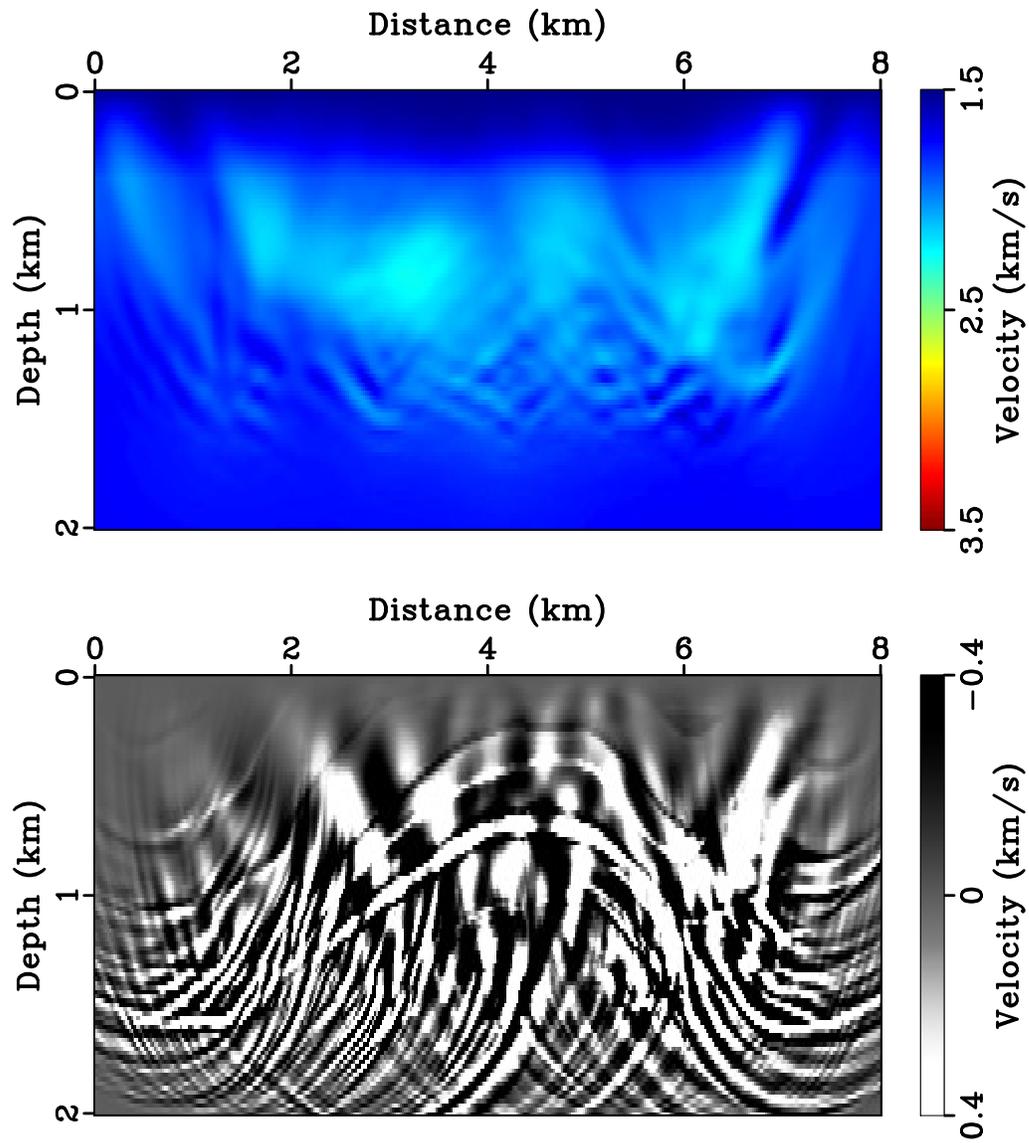
Figure 16: Results of non-extended inversion, as described in the text, using the same algorithms and computational effort as the extended inversion: (a) Inverted background velocity; (b) inverted reflectivity

sufficient to generate a useful α update via equation 18.

We should point out that the variable projection gradient formula (equation 9) 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.

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 $p \to 0$ 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 both outer and inner iteration counts and smaller computational cost per iteration. Many local optimization methods seriously outperform steepest descent, and should lead to more rapid convergence of the outer optimization. The standard text by Nocedal and Wright (1999) explains why, and 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 been explored in the last 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, 2017), 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 towards a practical algorithm.

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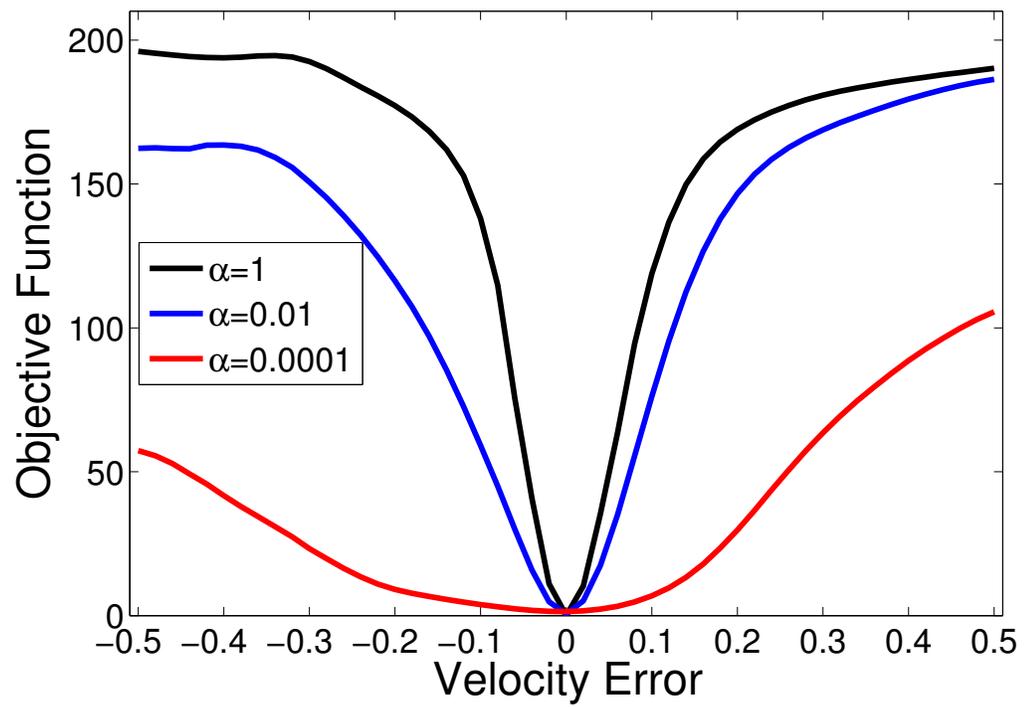
Figure 17: Scan test: objective function with different values of α. Velocity error varies from -50% to $+50\%$.

with fixed penalty parameter.

ACKNOWLEDGMENTS

REFERENCES

Ajo-Franklin, J. B., B. J. Minsley, and T. M. Daley, 2007, Applying compactness constraints to differential traveltime tomography: Geophysics, **72**, R67–R75.

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

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, Expanded Abstracts, Society of Exploration Geophysicists, SI9.5.

——, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Biondi, B., and P. Sava, 2004, Wave-equation migration velocity analysis - I: Theory, and 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, Expanded Abstracts, Society of Exploration Geophysicists, 1127–1131.

——, 2016, Fast extended waveform inversion using Morozovs discrepancy principle, *in* TRIP 2016 Annual Report: The Rice Inversion Project.

Fu, L., and W. W. Symes, 2017, An adaptive multiscale algorithm for efficient extended waveform inversion: Geophysics, **82**, R183–R197.

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.

——, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse Problems, **19**, R1–R26.

Hansen, P., 1992, Analysis of discrete ill-posed problems by means of the l-curve: SIAM Review, **34**, 561–580.

Hou, J., and W. Symes, 2016, Accelerating extended least-squares migration with weighted conjugate gradient iteration: Geophysics, **81**, no. 4, S165–S179.

Huang, Y., and W. Symes, 2015, Born waveform inversion via variable projection and shot record model extension: 85rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1326–1331.

Jackson, D. D., 1972, Interpretation of inaccurate, insufficient, and inconsistent data: Geophysical Journal of the Royal Astronomical Society, **28**, 97–110.

Kelly, K. R., R. W. Ward, S. Treitel, and R. M. Alford, 1976, Synthetic seismograms: A finite-difference approach: Geophysics, **41**, 2–27.

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, 2015a, Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis: Geophysical Prospecting, **63**, 391–404.

——, 2015b, Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis: Geophysical Prospecting, **63**, 391–404.

Luo, S., and P. Sava, 2011, A deconvolution-based objective function for wave-equation inversion: 81th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 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.

——, 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.

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, Expanded Abstracts, Society of Exploration Geophysicists, 930–934.

Roy, I., 2002, A robust descent type algorithm for geophysical inversion through adaptive regularization scheme: Applied Mathematical Modeling, **26**, 619–634.

——, 2005, Iteratively adaptive regularization in inverse modeling with bayesian outlook application on geophysical data: Inverse Problem in Science and Engineering, **13**, 655–670.

Sen, M. K., and I. G. Roy, 2003, Computation of differential seismograms and iteration adaptive regularization in prestack waveform inversion: Geophysics, **68**, 2026–2039.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

Shen, P., W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis

by wave-equation migration: 73rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2135–2139.

Symes, W., 1986, Stability and instability results for inverse problems in several-dimensional wave propagation, *in* Proc. 7th International Conference on Computing Methods in Applied Science and Engineering: North-Holland.

———, 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

Symes, W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.

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.

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

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.

Vigh, D., K. Jiao, W. Huang, N. Moldoveanu, and J. Kapoor, 2013, Long-offset-aided Full-waveform Inversion: 75th Annual Meeting, Expanded Abstracts, European Association of Geoscientists and Engineers, 76th Annual Meeting.

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, Expanded Abstracts, Society of Exploration Geophysicists, 957–961.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152.

Wahba, G., 1977, Practical approximate solutions to linear operator equations when the data are noise: SIAM Journal on Numerical Analysis, **14**, 651–667.

Warner, M., and L. Guasch, 2014, Adaptive waveform inversion: Theory: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1089–1093.

Weibull, W., and B. Arntsen, 2013, Automatic velocity analysis with reverse time migration: Geophysics, **78**, S179–S192.

An adaptive multiscale algorithm for efficient extended waveform inversion

Lei Fu and William W. Symes

ABSTRACT

Subsurface offset extended full waveform inversion may converge to kinematically accurate velocity models without the low-frequency data accuracy required for standard data-domain full waveform inversion. However, this robust alternative approach to waveform inversion suffers from 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 show here 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 remarkable ability to identify subsurface structure with the maximum resolution attainable from seismic data (Vigh et al., 2010, 2013). However 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 model fail to yield substantially better data fit. Amongst 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 non-local 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 data fit attainable by local optimization, at least in some cases (Stolk et al., 2009; Symes, 2014). A number of studies have suggested that the subsurface offset extension may be used as the basis for successful approaches to full waveform inversion, convergent over a much larger region of model space than is standard least-squares FWI (Shen et al., 2003, 2005; Shen and Symes, 2008; Albertin et al., 2006; Symes, 2008; Fei and Williamson, 2010; Vyas and Tang, 2010; Biondi and Almomin, 2012; Shen, 2012; Weibull and Arntsen, 2013; Shan and Wang, 2013; Biondi and Almomin, 2014; Liu et al., 2014; Symes, 2014; ten Kroode, 2014; Lameloise et al., 2015; Fu and Symes, 2015). Much of this work uses linearization (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 non-local 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, moves the extended model closer to a non-extended or physical model concentrated at zero subsurface offset, and (2) low-pass filtered data has enough kinematic content to drive velocity improvement, so long as the filtered data spans 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, as 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 non-increasing. 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 extended model back to physical model by progressively reducing the subsurface offset range throughout 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 modelling (Born or single-scattering approximation). We describe three adaptive parameter adjustments, of maximum subsurface offset, variable projection penalty parameter, and computational grid, and show how to combine them to dramatically enhance 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

While 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; 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: $r = 2v\delta v$. 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 ignored in the notation for the data traces. The pres-

sure field $p(x, z, t; x_s)$ solves the acoustic wave equation

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

The right-hand 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

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

F is defined in terms of δp by

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

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, 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

- an extended model space $\bar{M}$;

- an extension operator $\chi: M \to \bar{M}$;

- an extended modeling operator $\bar{F}: \bar{M} \to D$; and

- an annihilator operator $A : \bar{M} \to \bar{M}$.

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: $\bar{r}(x, z, h)$. Since r is (up to a scale factor) the perturbation in the compliance, one can think of the extended reflectivity as representing a non-local perturbation in the acoustic constitutive relation (we are indebted to Scott 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,

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

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 is 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. Amongst 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-hand 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, $N_h = \dfrac{2H}{dh}$, 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); 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 an 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 non-extended modeling: as noted above, v must predict arrival times within a half-wavelength in order that there exist a non-extended 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 non-physicality, and drives both measures towards zero. A simple objective function capturing this concept is

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

The penalty weight α controls the balance between penalties on data misfit and model extension: when $\alpha \to 0$, the model has little constraint on energy distribu-

tion in the extended dimension h, and can always achieve 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 is 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 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 both 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*

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

The normal equation is equivalent to vanishing of the $\bar{r}$-gradient of $J[v, \bar{r}, \alpha]$. Since the J is positive semidefinite quadratic in $\bar{r}$, that 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:

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

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

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

$(D\bar{F}[v])^T$ is the so-called tomographic or WEMVA operator. As $\bar{F}[v]$ is actually the linearization, that is, derivative of the basic acoustic modeling operator, $D\bar{F}[v]$ is actually 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). 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) showed 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) gave 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 migration velocity to 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 migration (or inversion) velocity accuracy and maximum subsurface offset holds more generally, in a qualitative sense. When the migration velocity model is inaccurate, long offset is needed for good data fit; when the velocity model becomes closer to the correct one after updates, the observed data can be predicted by extended reflectivity model with shorter offset. This relation will be illustrated below, and follows for 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-hand side of the extended linearized wave equation 4. To track this dependence, add it to the notation, and rename the solution $\bar{r}[v, \alpha, H]$. Similarly, 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 $\tilde{J}$:

$$
\begin{aligned}
\tilde{J}[v,\alpha,H] &= e[v,\alpha,H] + \alpha p[v,\alpha,H], \\
e[v,\alpha,H] &= \frac{1}{2}\|\bar{F}[v]\bar{r}[v,\alpha,H] - d\|^2, \\
p[v,\alpha,H] &= \frac{1}{2}\|A\bar{r}[v,\alpha,H]\|^2.
\end{aligned}
\tag{10}
$$

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

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

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 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 re-uses $\bar{r}[v,0,H]$ so 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; 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, since $e_{H/2}$ is not precisely the same as $e[v,\alpha,H/2]$. In the following, we will use $E = \frac{1}{2}X^2\|d\|^2$ as the relative error measure appropriate for comparison with $e(...)$.

Adaptive Penalty Weight

The penalty parameter α is essential in driving the extended waveform inversion towards a solution of the (non-extended) waveform inversion problem. Since p

---

Algorithm 1 H Update

Require: $X > 0, \mu > 1, E = \dfrac{1}{2}X^2\|d\|^2$

  $H_+ = H$
  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 ifreturn H_+

---

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 in (Fu and Symes, 2016). Unlike normal applications of this principle, the 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,

---

Algorithm 2 α Update

Require: $X > 0, E = \dfrac{1}{2}X^2\|d\|^2, 0 < \gamma_- < 1 < \gamma_+, 0 < \beta_- < 1 < 1/\beta_- < \beta_+$

    if initialize **then**

        $\alpha = 0$

    end if

    $\alpha_+ = \alpha$

    if $e[v, \alpha_+, H] < \gamma_- E$ **then**

        $\alpha_+ \leftarrow \alpha_+ + \dfrac{\gamma_+ E - e[v, \alpha_+, H]}{2p[v, \alpha_+, H]}$

        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_+

---

so an algorithm that includes such updates is outside the scope of standard opti-
mization theory.

We do not attempt to create a conceptual framework for such "multi-function
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 previ-
ous iteration as the initial step length for the current one. This carry-over 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 accu-
racy 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 defined 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 display 3.

Algorithm 3 Adaptive Variable Projection Method

Require: choose initial velocity v, α, H, parameters for algorithms 1, 2, continu-
   ous optimzation 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 im-
   provement $\delta = (\tilde{J}[v,\alpha,H] - \tilde{J}[v_+,\alpha_+,H_+])/\tilde{J}[v,\alpha,H]$
       $v \leftarrow v_+, \alpha \leftarrow \alpha_+, H \leftarrow H_+$
   until $\delta < \epsilon$

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 to 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 frequency content from convergence (Symes, 2008, 2014). Instead the point is purely to reduce the computational cost of a majority of the iterations.

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 $k = 0, ..., n$ **do**

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

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

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

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

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

$\quad$ execute algorithm 3 with discretization parameters dt, dx, dz, dh

end for

So 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 non-extended inversion steps stays constant or decreases as the iterations of algorithm 4 proceed.

Cost

It is straightforward to analyze the cost of algorithm 4 occurs 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 a carrying out n steps of Algorithm

4 is

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

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 non-adaptive algorithm, using the finest grid and largest offset axis with $N_h = 2^n N_{h,n}$ gridpoints, is

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

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 $k = 0, ..., n$. Since $N_{t,k} = 2^{k-n} N_{t,0}$, the cost ratio is

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

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

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

For $n = 1$ (2 data octaves), $r \approx 1/2$, for $n = 2$ (3 octaves), $r \approx 1/8$, and for $n = 3$ (4 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 multi-scale approaches by solving a velocity estimation problem modeled on reflection seismology. The simulations are performed using a 2D constant-density acoustics, time domain, finite difference method (second order in time, eighth order in space).

We use the method of steepest descent with quadratic backtrack line search to search for the minimum of the objective function. The gradient of the objective function is computed by 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 current model. The optimal step length is estimated by assuming 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, or 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, as they indicated clearly the degree of focus towards $h = 0$, that is, kinematic correctness of the velocity, attained by our algorithm.

Increased computational cost

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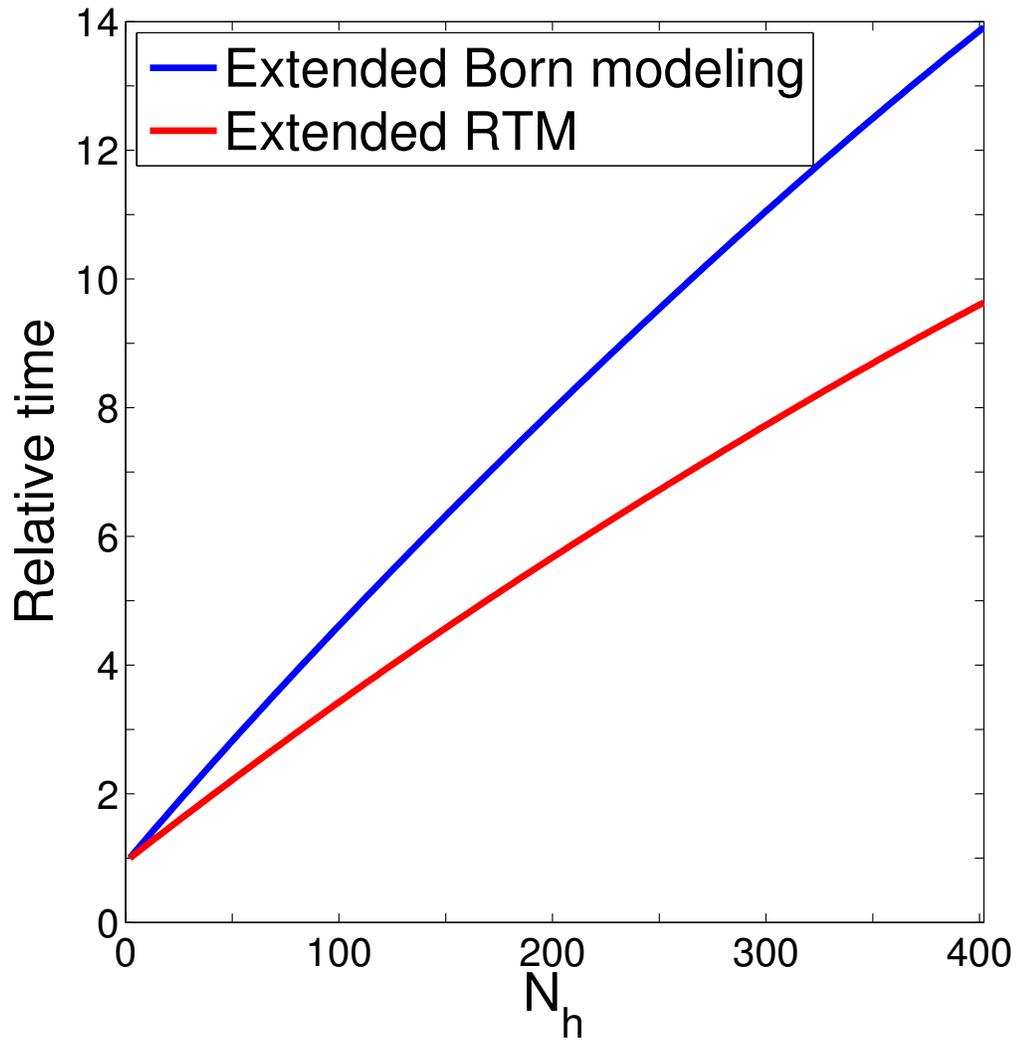
Figure 1: The relative computing time of extended modeling and RTM as a function of number of grid points N_h in subsurface offset axis. The reflectivity grid size is $1000 \times 1000 \times N_h$.

Single reflector model

The purpose of this experiment is to investigate the behavior of extended reflectivity model in 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): that the extent of the subset offset axis necessary for accurate data fit increases with velocity error.

The background velocity model measures $3.0\,km \times 2.4\,km$ with $20\,m$ 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 depth of $1.6\,km$. 61 sources ($0.3\,km$ to $2.7\,km$) and 151 receivers ($0\,km$ to $3.0\,km$) are placed on the surface. Note that the background velocity model $v(x,z)$ is non-extended, while the extended perturbation model $\bar{r}(x,z,h)$ has nonzero value only at $h = 0\,m$. The observed data of shot 31 is depicted in Figure 2b.
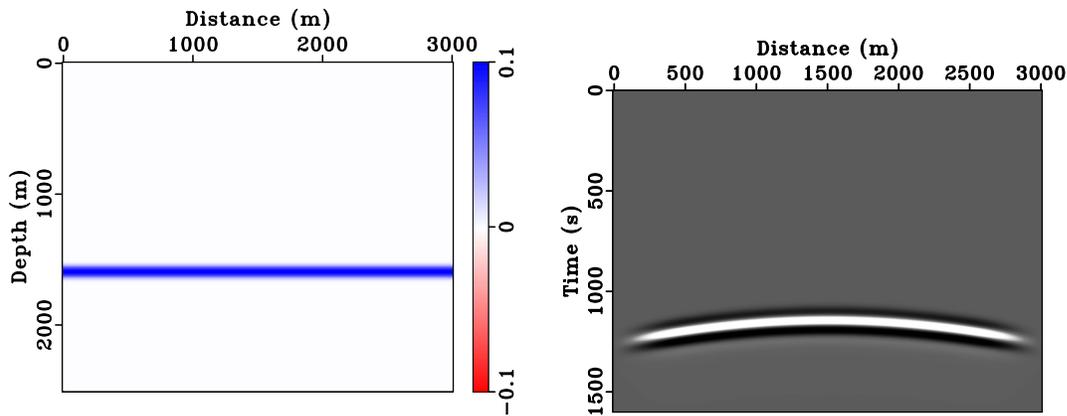
Figure 2: (a) Extended reflectivity $\bar{r}$ at $h = 0\,m$; (b) data of shot 31 at the center

The ELSRTM image with correct velocity after 20 CG iterations is shown in Figure 3a. The energy is focused at zero subsurface offset. The image with 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 h direction. Compared with the correct velocity case, the energy is scattered along h axis. As the velocity error decreases, the energy tends to focus toward the center $h = 0$ (shown in Figure 4a - Figure 4f).

In order to fit the observed data well, the extended reflectivity model must have
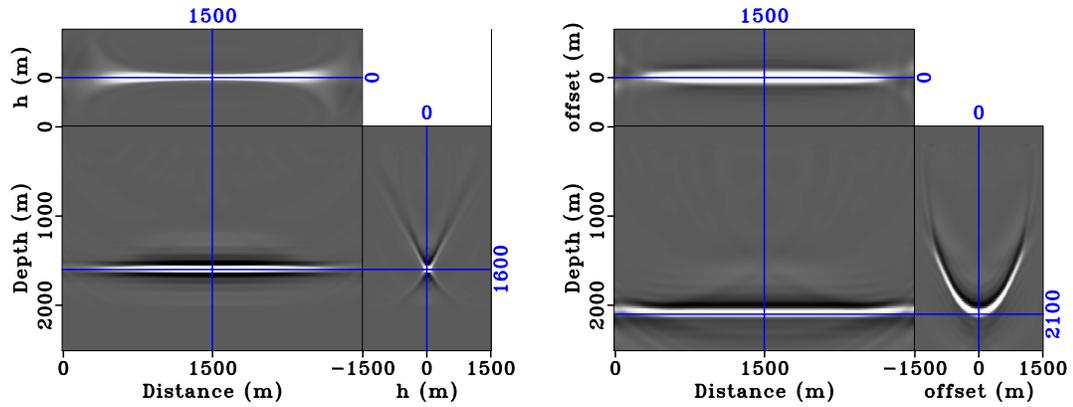
Figure 3: Inverted extended reflectivity $\bar{r}$ after 20 iterations of CG. (a) correct background velocity ($v = v_{true}$) (b) wrong background velocity ($v = 1.3v_{true}$)
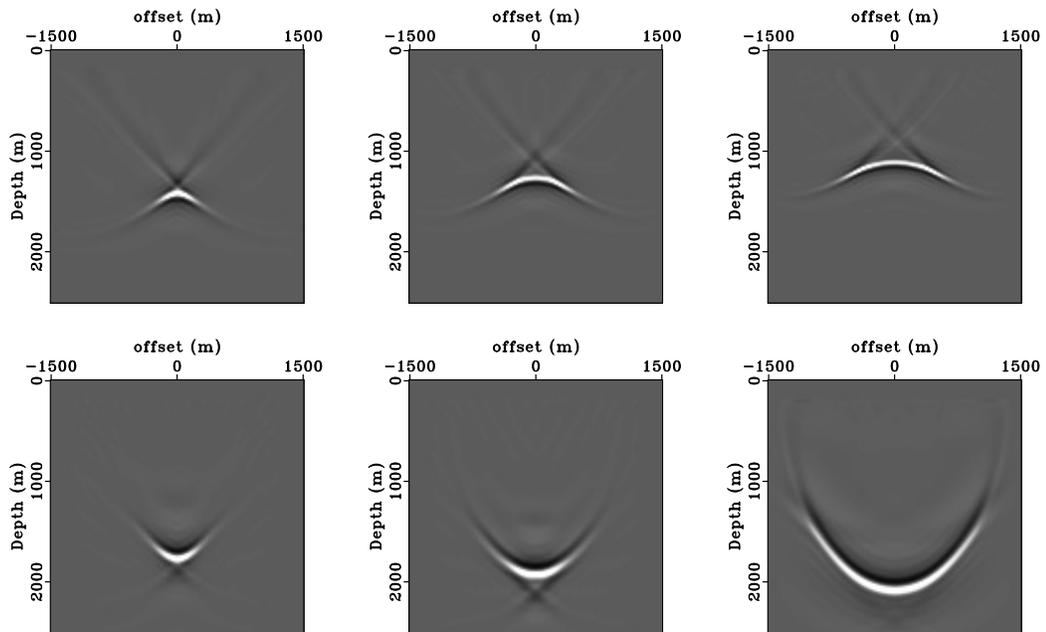
Figure 4: Inverted extended reflectivity $\bar{r}$ in $z - h$ plane after 20 iterations of CG. (a) $v = 0.9v_{true}$, (b) $0.8v_{true}$, (c) $0.7v_{true}$, (d) $1.1v_{true}$, (e) $1.2v_{true}$, (f) $1.3v_{true}$

adequate subsurface extension (see Figure 5a and 5b). On the other hand, the required amount of subsurface extension decreases with the error of the velocity model, which suggests that as the velocity model updates toward the correct one in inversion, shorter subsurface offset is needed (see Figure 6).
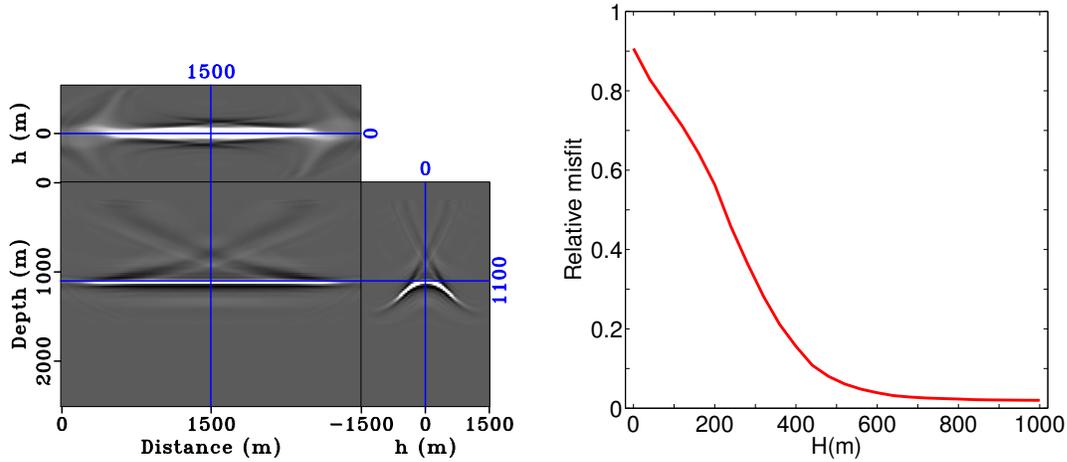
Figure 5: Relationship between the maximum subsurface offset H and the data residual Δd_H. The background velocity is 30 % slower than the true velocity ($v = 0.7v_{true}$). (a) ELSRTM image $\tilde{r}$ after 20 iterations of CG, (b) H vs relative Δd_H

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 is generated with a $3 - 30$ Hz bandpass filter, or approximately three octaves. Accordingly, $n = 2$ and the the observed data and the source function are filtered by three bandpass filters ($3 - 7.5\,Hz$, $3 - 15\,Hz$, and $3 - 30\,Hz$). Correspondingly, the spatial decomposition is implemented in three steps with grid size $50\,m$, $25\,m$ and $12.5\,m$. The grid size in each stage is determined to fulfill the rule $\frac{v_{min}}{f_{max}} > 5dx$ to avoid numerical dispersion. The spatial grid size is same in all dimensions ($dx = dz = dh$). The time step intervals ($dt = 8\,ms, 4\,ms, 2\,ms$) in each stage are chosen to fulfill Courant-Friedrichs-Lewy (CFL) condition ($\dfrac{vdt}{\sqrt{dx^2 + dz^2}} <$
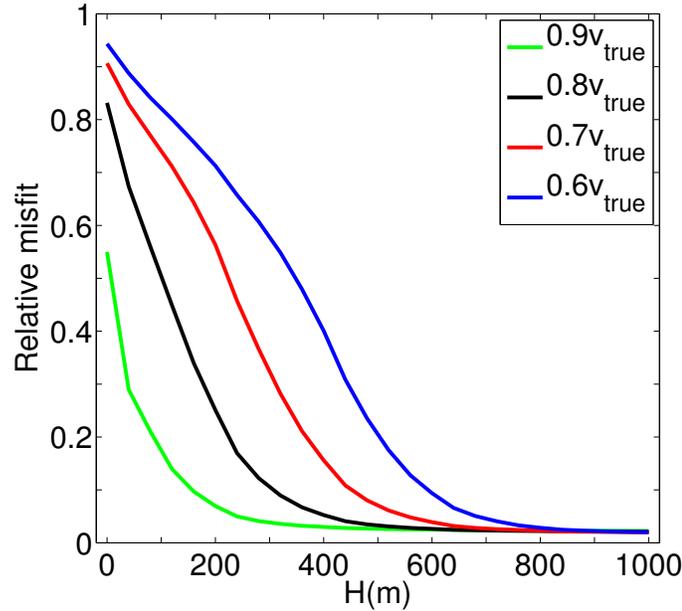
Figure 6: After 20 iterations of CG, the relative data misfit is plotted as a function of maximum subsurface offset H. Different colors represents different levels of velocity error.

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 left α constant thereafter.

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

Stage 1. We start with grid size $50\,m$, frequency band $3 - 7.5\,Hz$, and the time step interval $8\,ms$. The initial background velocity is constant ($3.0\,km/s$). Assume 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 original data after 10 iterations of CG. Meanwhile, reflectivity model with only 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 5% misfit in 20 iterations), so the offset range $H = 800\,m$ is
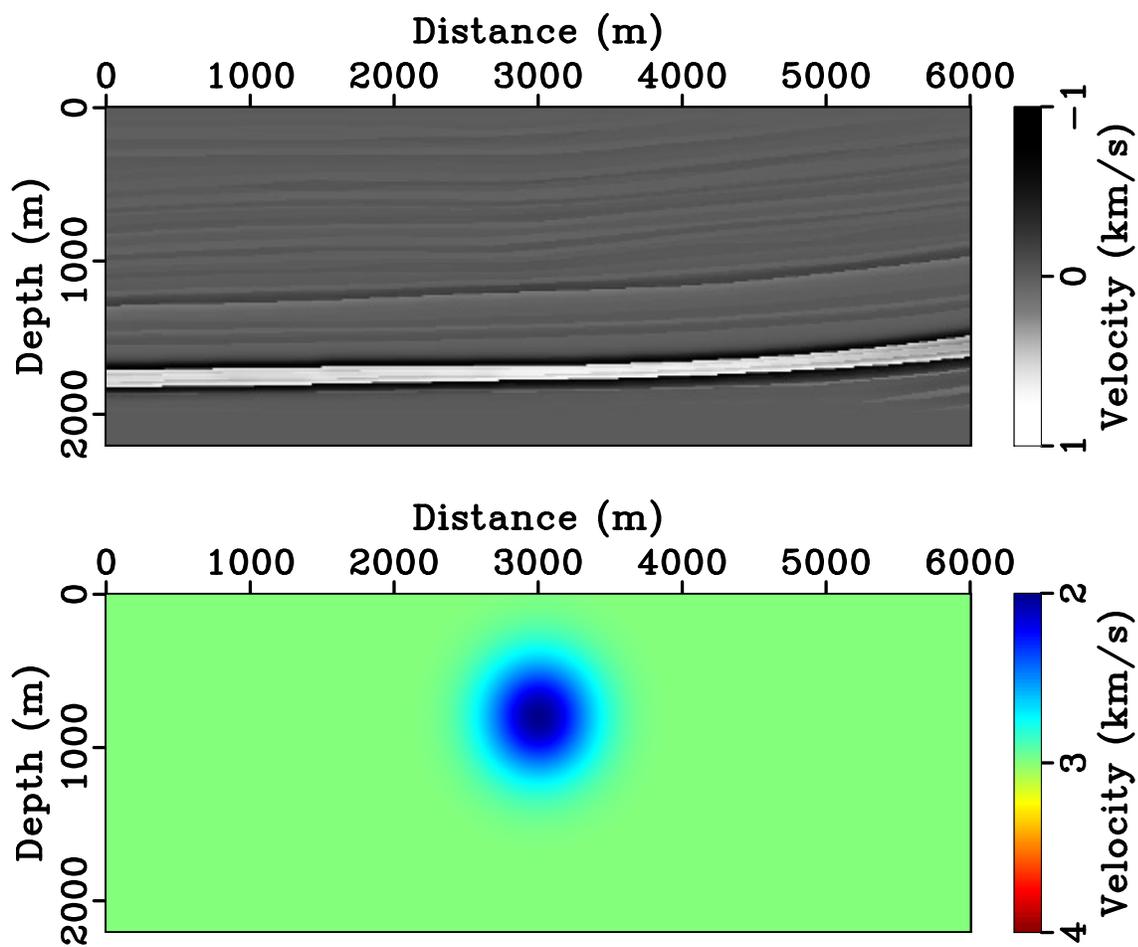
Figure 7: (a) Extended velocity perturbation $\delta\bar{v}$ at $h = 0\ m$ (b) background velocity

Parameter	Measurements
Source wavelet	bandpass $3 - 30\,Hz$
Source position $\mathbf{x_s}$	$x : 0 - 6\,km$ every $50\,m$, $z = 0\,m$
Receiver position $\mathbf{x_r}$	$x : 0 - 6\,km$ 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 inner loop	20

Table 1: Lens model
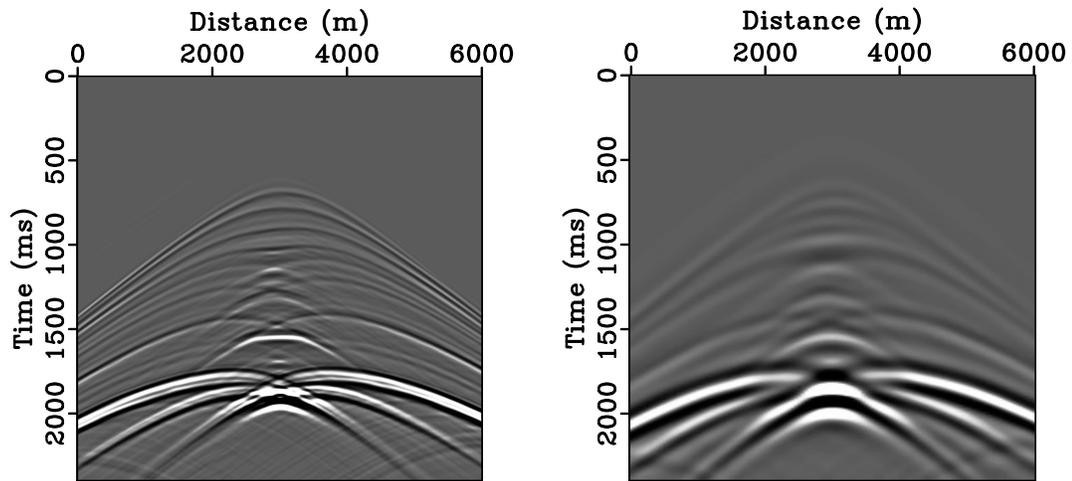
Figure 8: (a) Data of shot 61 at the center $(3 - 30\,Hz)$ (b) bandpass low-frequency data $(3 - 7.5\,Hz)$

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 has been imaged to deeper positions. The gather is barely focused in h direction.

Here, we have run 20 iterations of CG, far more iterations than 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.
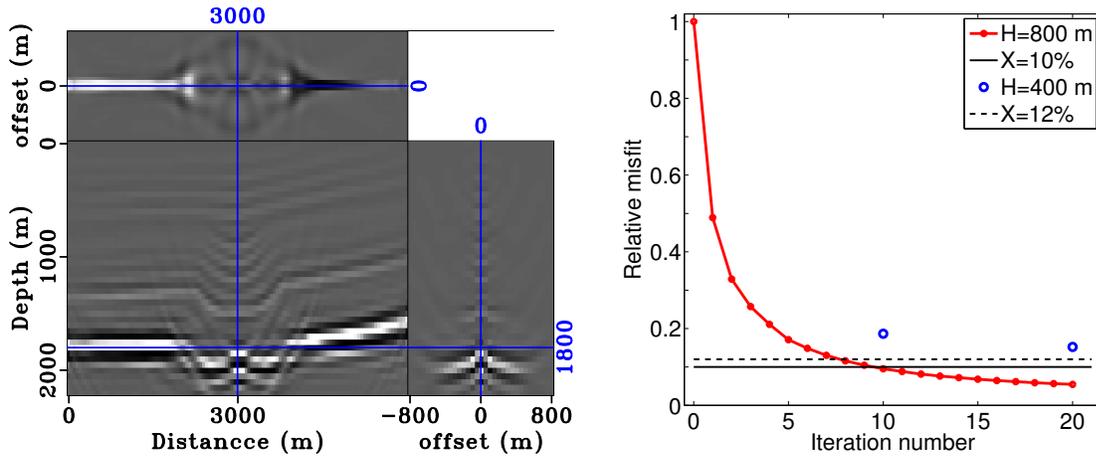
Figure 9: Step1: use initial velocity, spatial grid size $dh = dx = dz = 50\,m$, bandpass filter $f : 3 - 7.5\,Hz$, time step interval $dt = 8\,ms$ (a) ELSRTM result of 20 CG iterations, (b) Relative data residual with different offset ranges.

With the inverted $\bar{r}$ shown in Figure 9a and 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 shallower position and more focused in h direction. As the velocity model becomes closer to the correct one, the energy is more focused toward $h = 0\,m$, which suggests that 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 data misfit less than 12%, shown in Figure 10c. After second velocity update, since half offset range $H = 200\,m$ is not able to provide good data fit, $H = 400\,m$ is still the optimal choice (see Figure 11c). In the next 3 velocity updates (iteration 3, 4, and 5), we are able to reduce the extended offset range by half (see Figure 12, 13, and 14). After the 6th velocity update, reflectivity model even without extension is able to predict the observed data within the tolerance level. And the velocity update satisfies the convergence condition as the 9*th* step shown in previous algorithm, so higher frequency and finer grid can be used, leading the algorithm to the second stage.

Stage 2. The source wavelet and observed data are filtered by lowpass filter ($3 - 15\,Hz$). At the same time, the spatial grid size and the time step interval are reduced by a factor of 2. With the help of the updated velocity model (see Figure
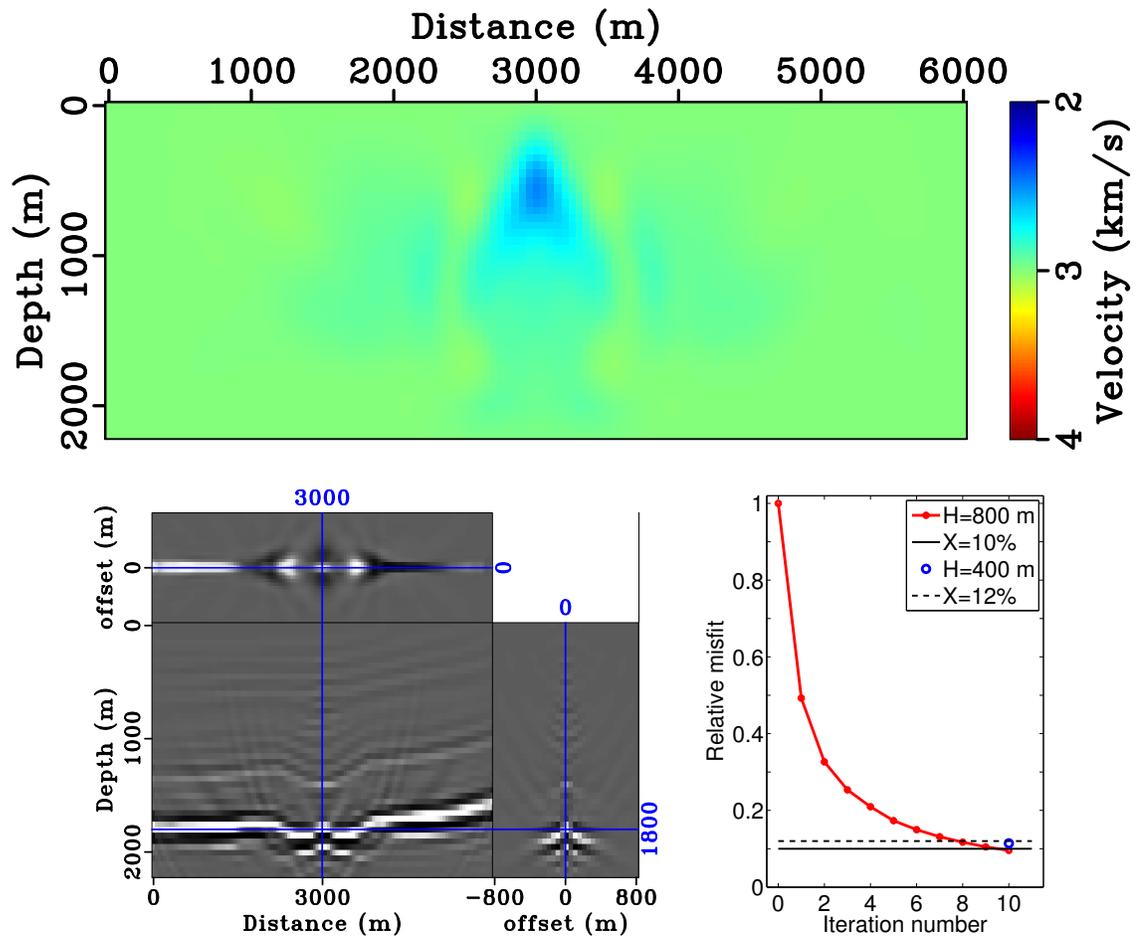
Figure 10: (a) The 1st velocity update; (b) The ELSRTM image by using 1st updated velocity; (c) The relative data residual with different offset ranges.
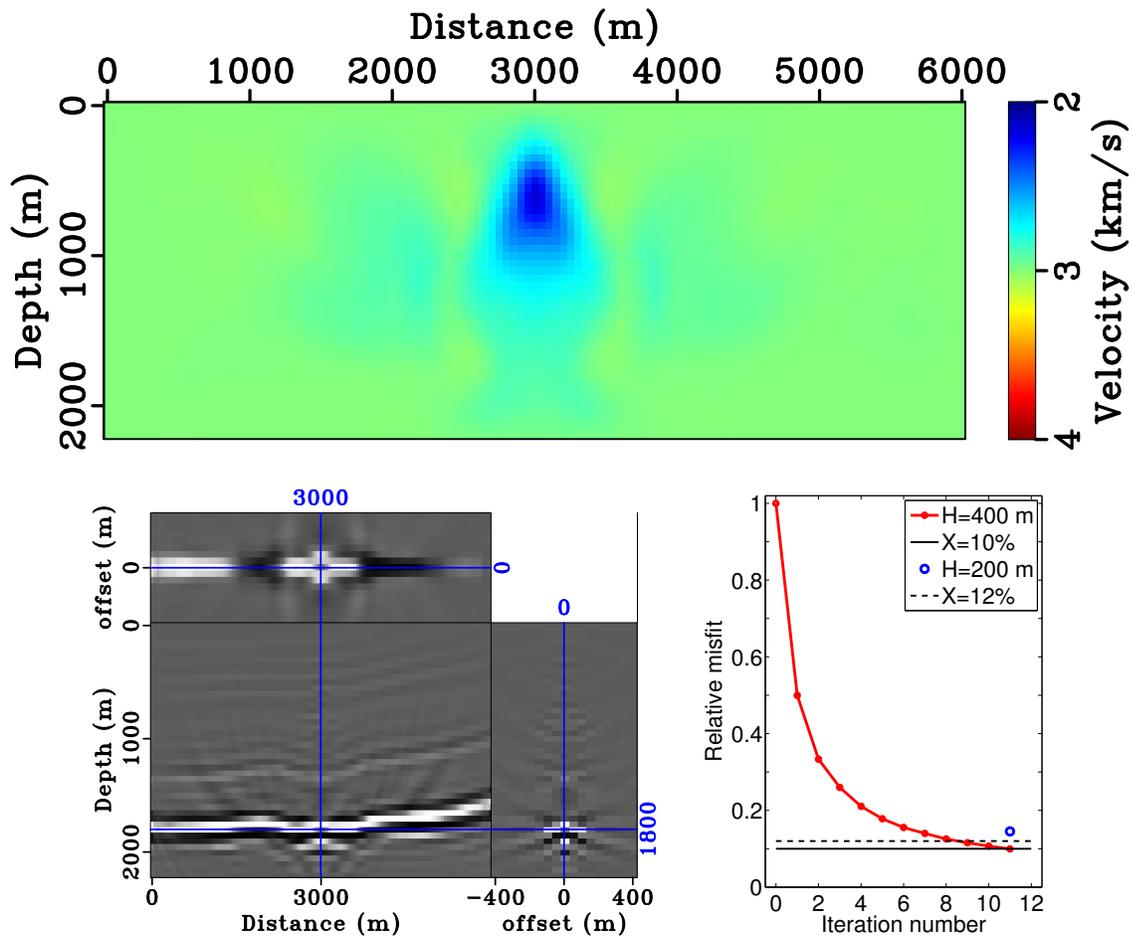
Figure 11: (a) The 2nd velocity update; (b) The ELSRTM image by using 2nd updated velocity; (c) The relative data residual with different offset ranges.
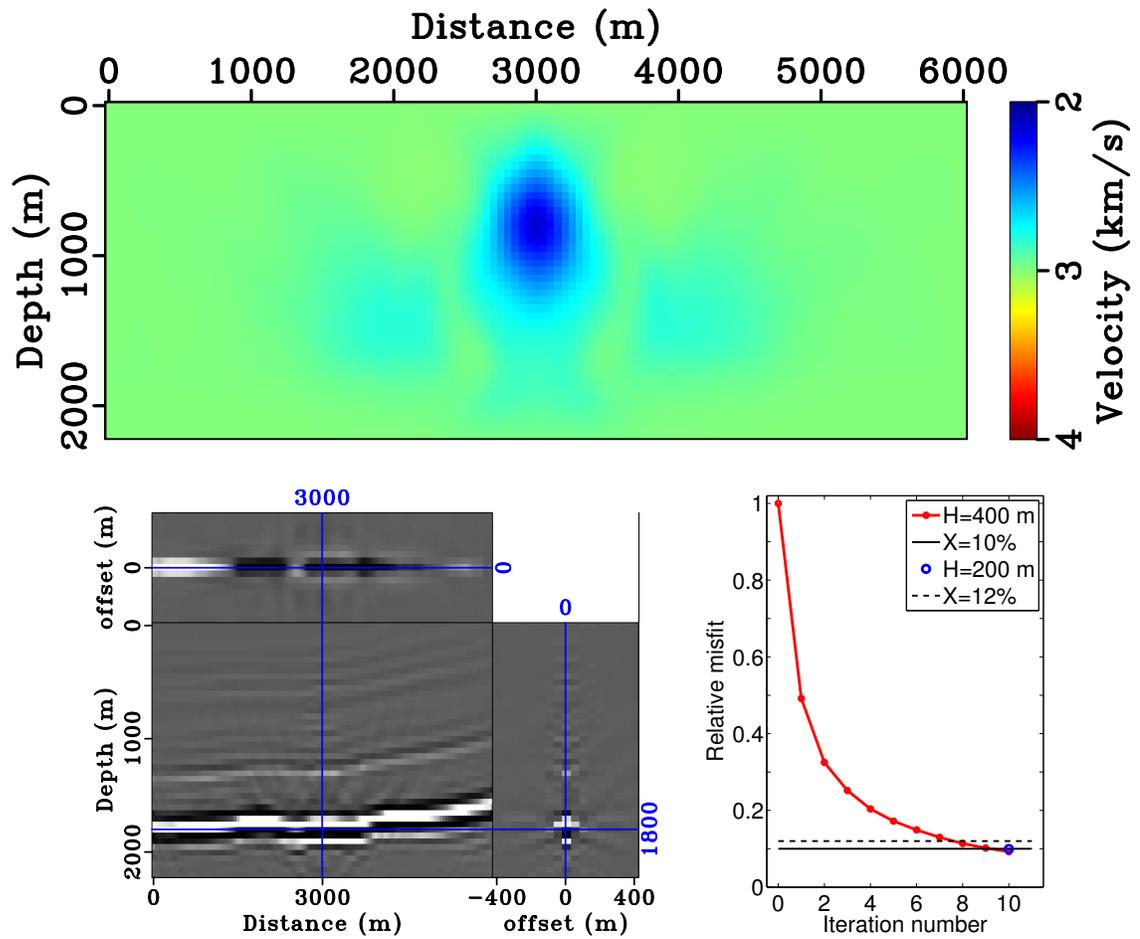
Figure 12: (a) The 3rd velocity update; (b) The ELSRTM image by using 3rd updated velocity; (c) The relative data residual with different offset ranges.
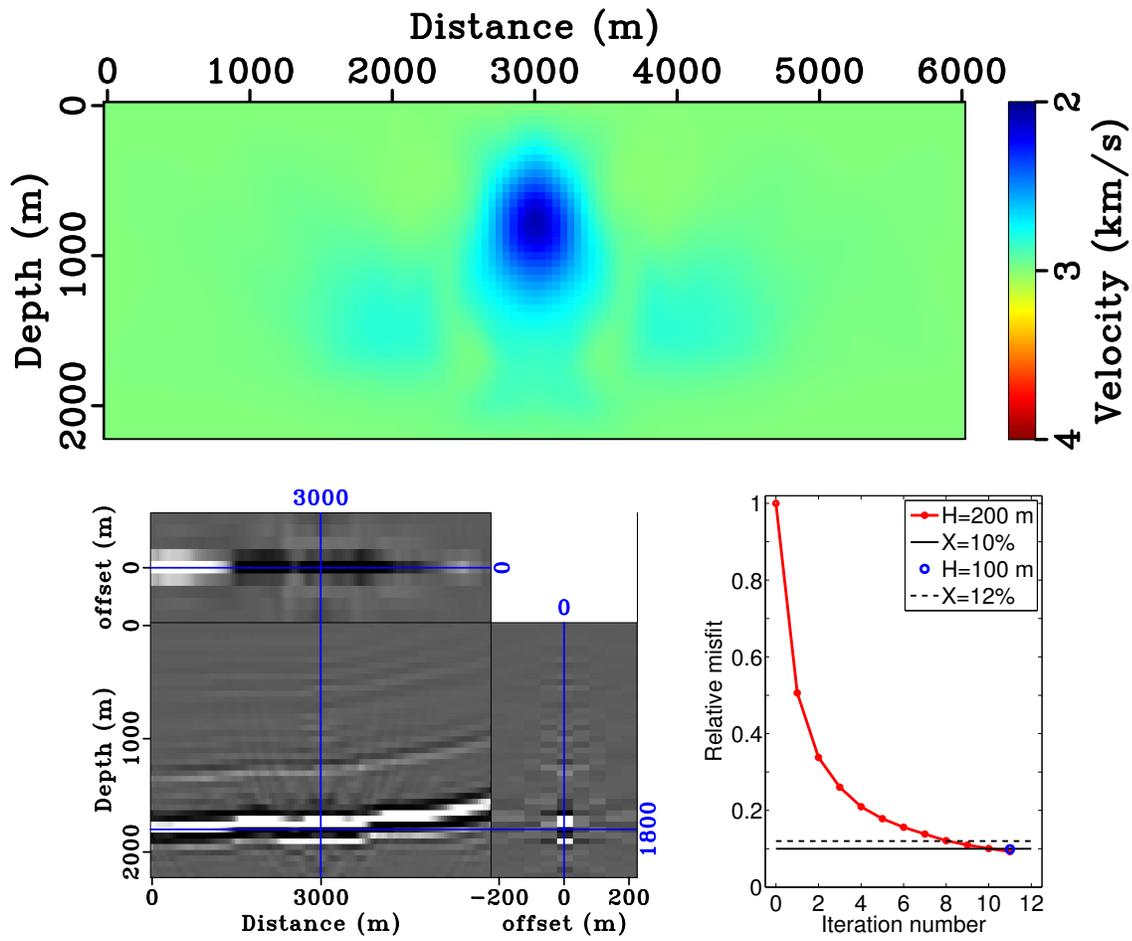
Figure 13: (a) The 4th velocity update; (b) The ELSRTM image by using 4th updated velocity; (c) The relative data residual with different offset ranges.
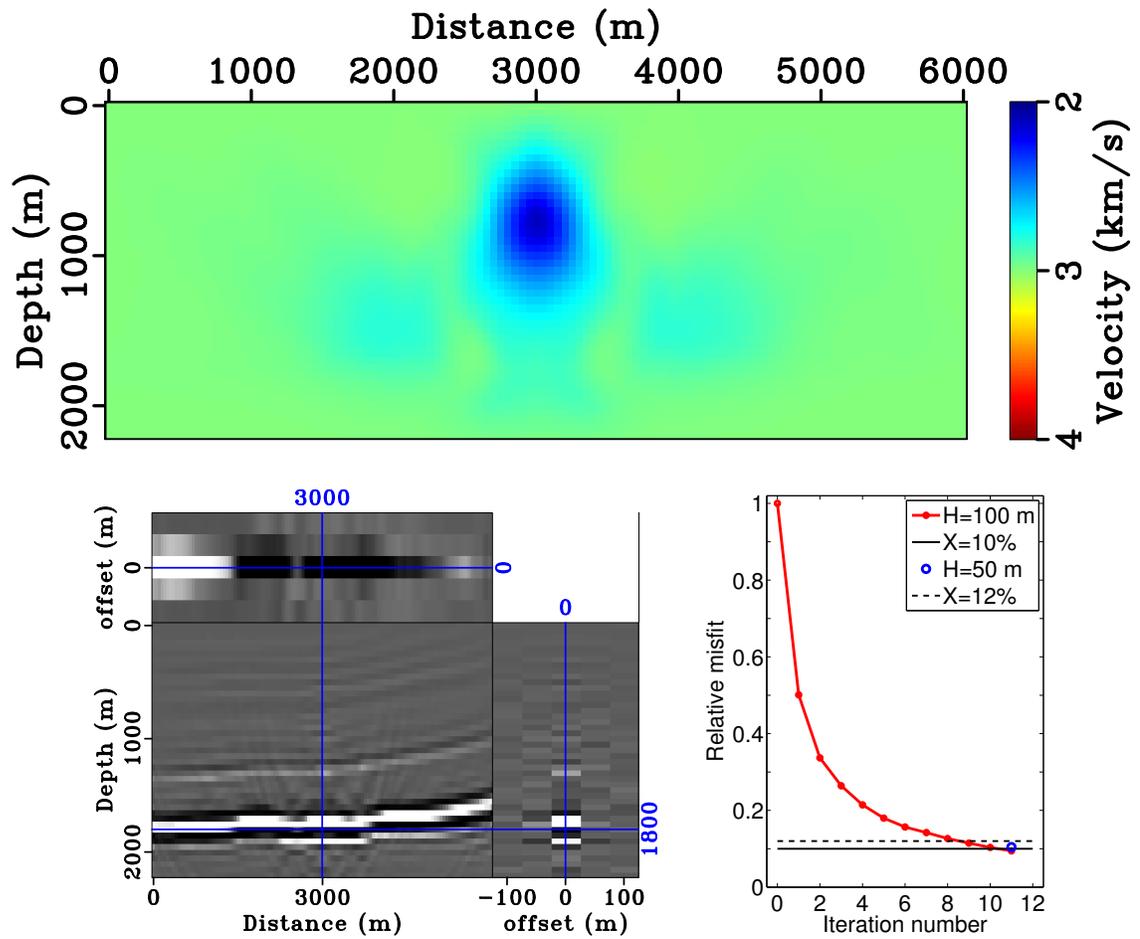
Figure 14: (a) The 5th velocity update; (b) The ELSRTM image by using 5th updated velocity; (c) The relative data residual with different offset ranges.
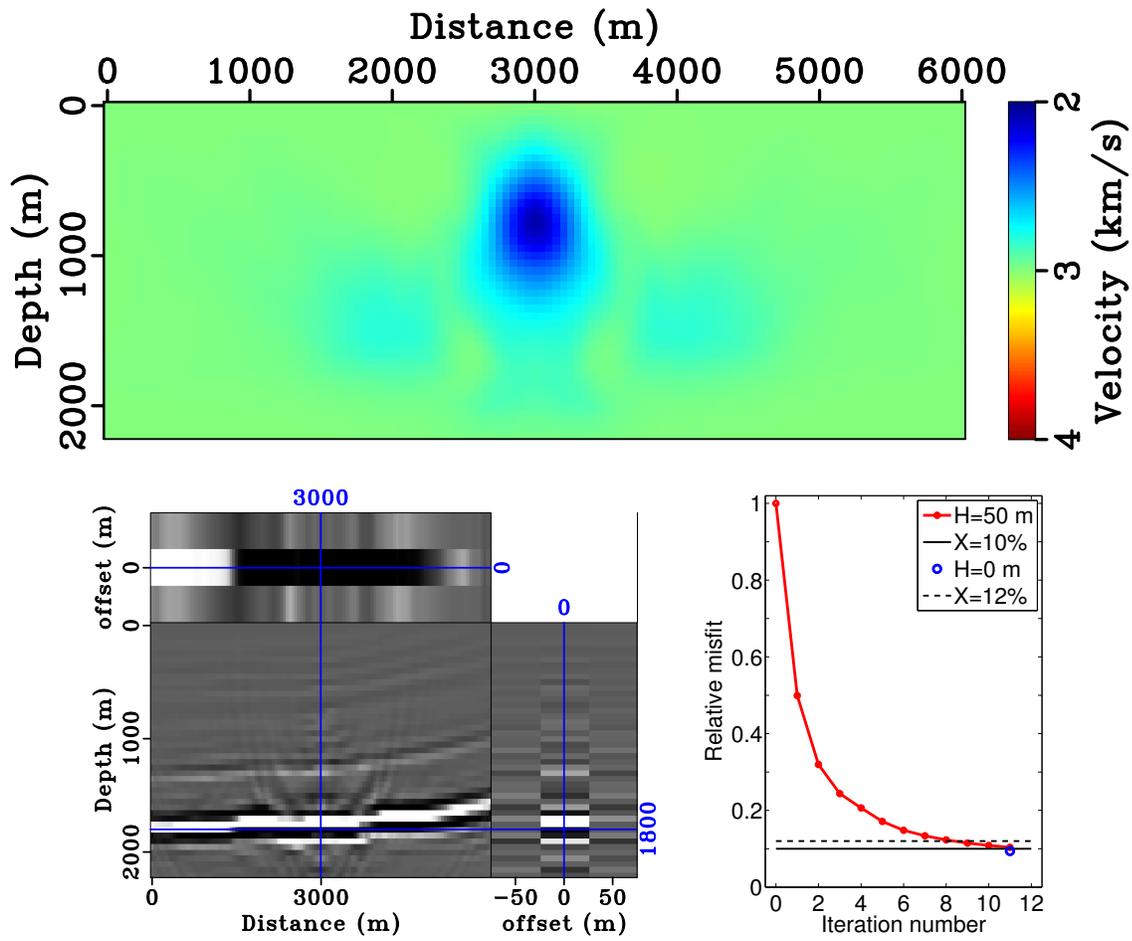
Figure 15: (a) The 6th velocity update; (b) The ELSRTM image by using 6th updated velocity; (c) The relative data residual with different offset ranges.
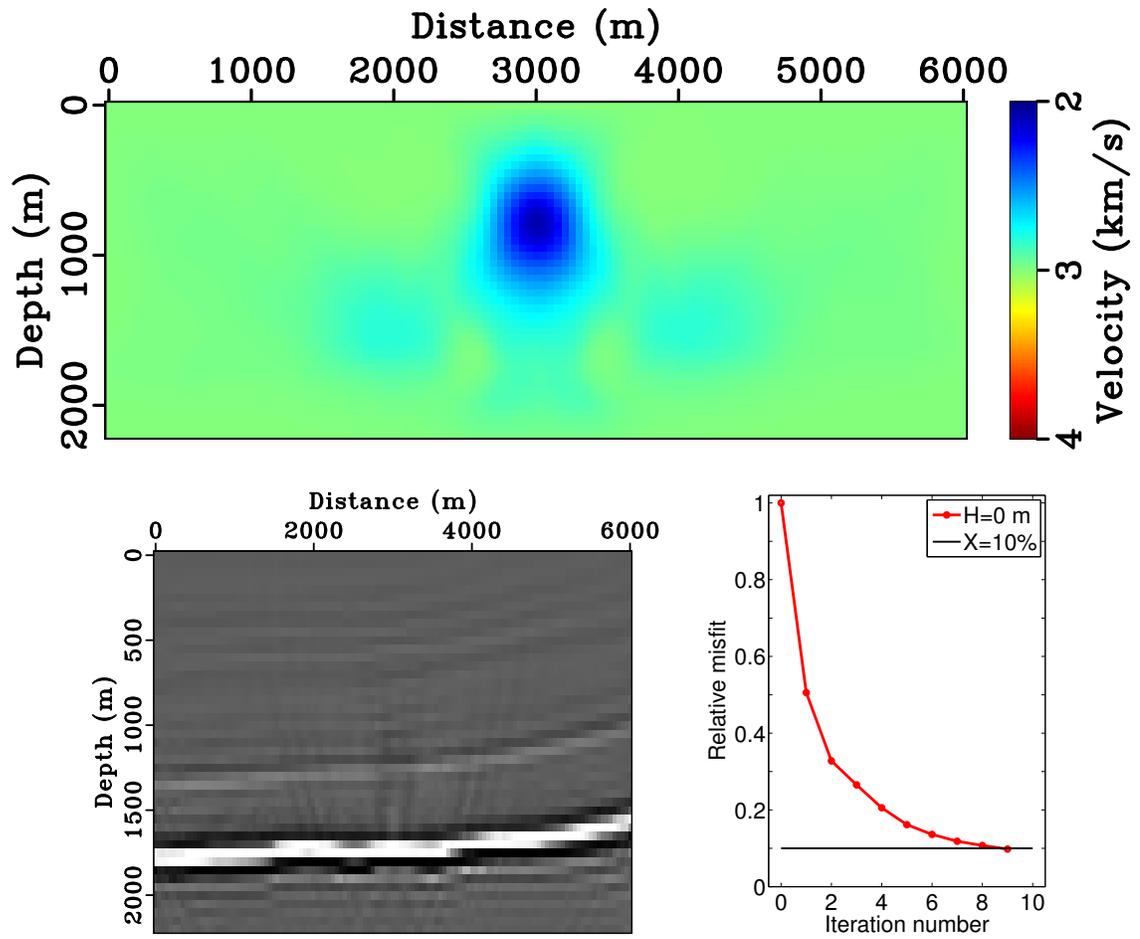
Figure 16: (a) The 9th velocity update; (b) The ELSRTM image by using 9th updated velocity; (c) The relative data residual.

17a) after 2 iterations, higher frequency content, finer spatial grid and time step, the ELSRTM image shows more structural details after 7 CG iterations (shown in Figure 17b). The data residual is shown in Figure 17c.
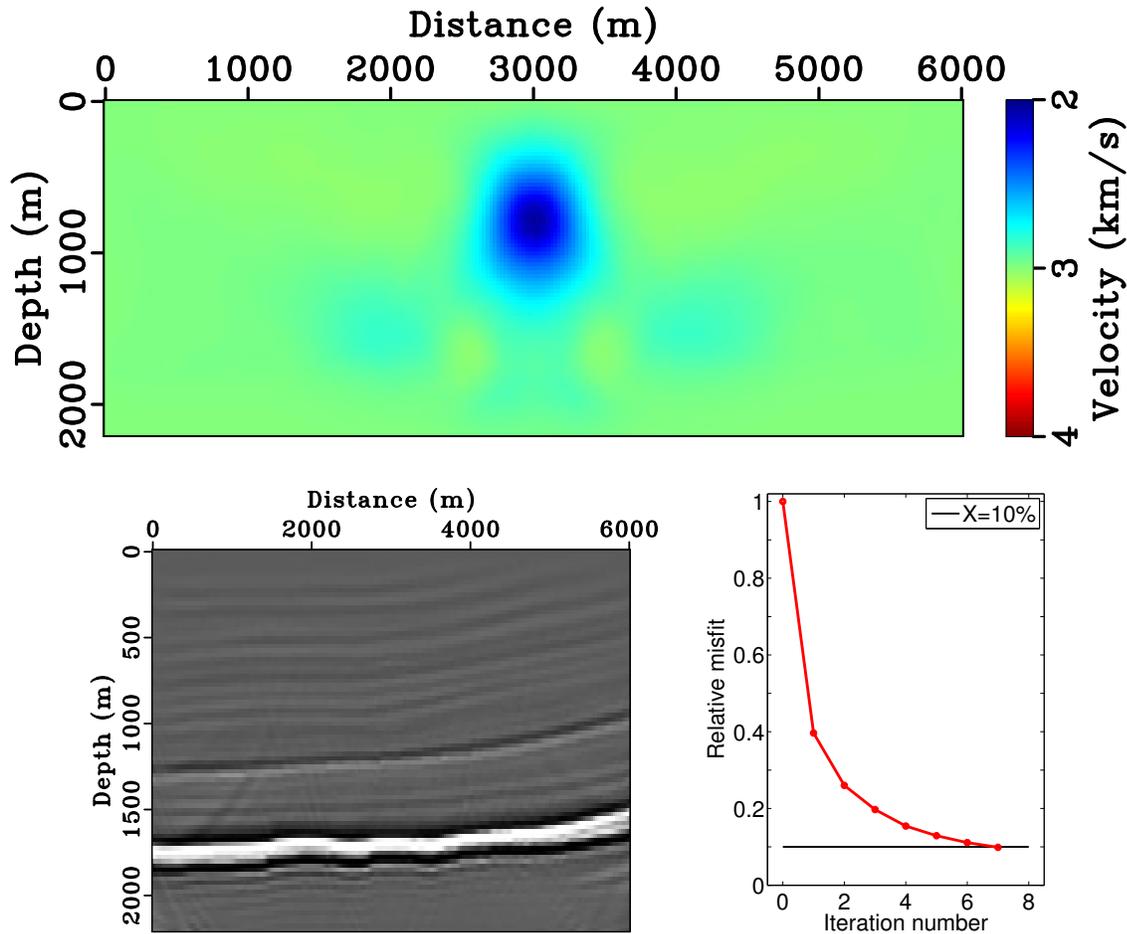
Figure 17: Stage 2 (a) The 2nd velocity update; (b) The ELSRTM image (c) The relative data residual.

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, higher frequency content, finer spatial grid and time step, the background velocity and layered structure are successfully recovered.

For comparison, the inversion results without model extension are shown in Fig-
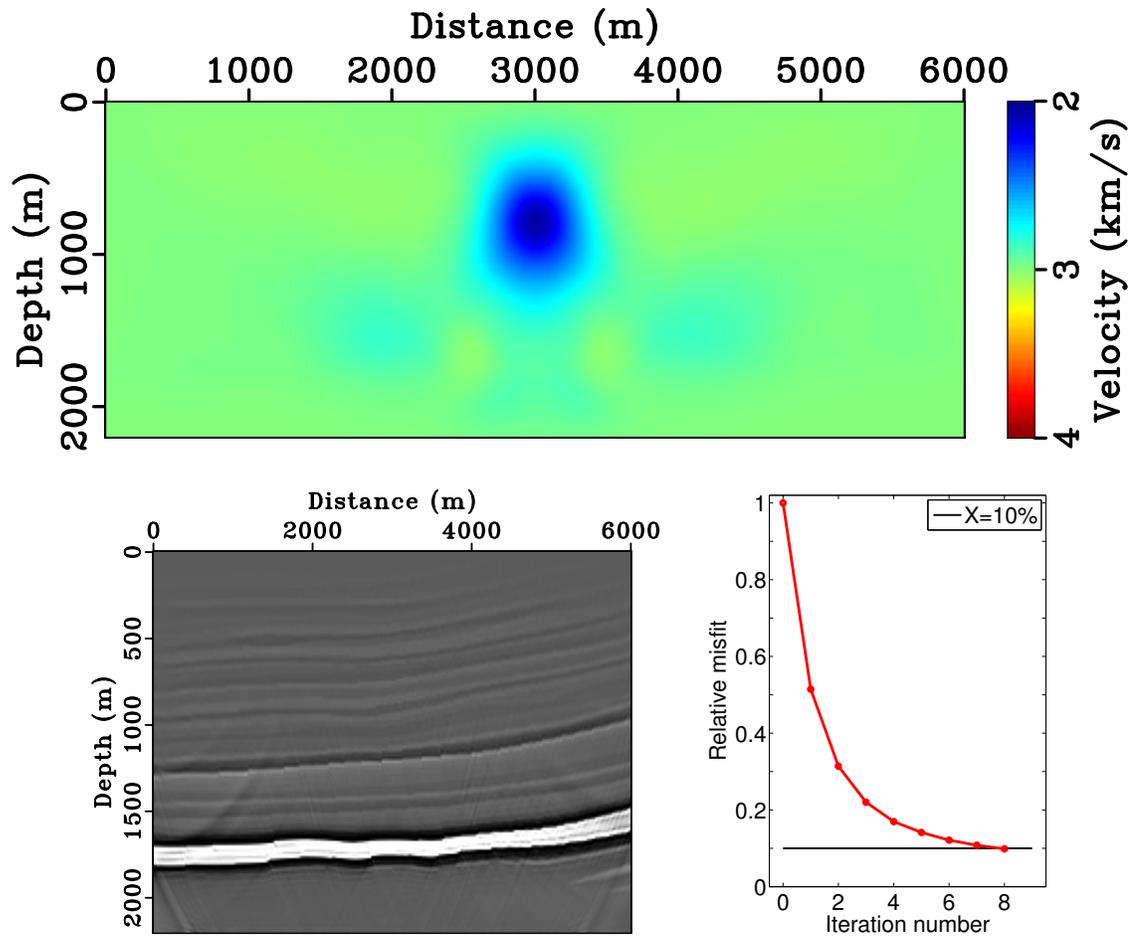
Figure 18: Stage 3 (a) The inverted background velocity model; (b) The ELSRTM image (c) The relative data residual.

ure 19a, 19b without multiscale and Figure 20a, 20b with multiscale approach. Here, without extended dimension h, the objective function is defined with a different model space (M, not $\bar{M}$) and only contains the data misfit term $\frac{1}{2}\|\bar{F}[v]r - d\|^2$. For both cases, we use the same optimization algorithm, steepest descent with quadratic backtrack line search. Both experiments are performed by using same number of velocity iterations at each refinement stages as the extended case. And the reflectivity model is estimated by 20 iterations of CG, which equals 20 modeling/migration pairs. As a result, the total numbers of modeling/migration pairs in these two non-extended experiments are same as 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.

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. The background velocity increases with depth. The velocity is higher in the center, where the anticline structure sits. The basic information is listed in table 2.

Parameter	Measurements
Source wavelet	bandpass $5-20\,Hz$
Source position $\mathbf{x_s}$	$x : 1-7\,km$ every $40\,m$, $z = 0\,m$
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	$v = 1.5\,km/s$
Maximum iterations inner loop	20

Table 2: Thrust model

In this example, bandpass 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 bandpass filters ($5-10\,Hz$ and $5-20\,Hz$. Correspondingly, the space decomposition is implemented with grid size $40\,m$ and $20\,m$, while the time step intervals are $4\,ms$ and $2\,ms$.
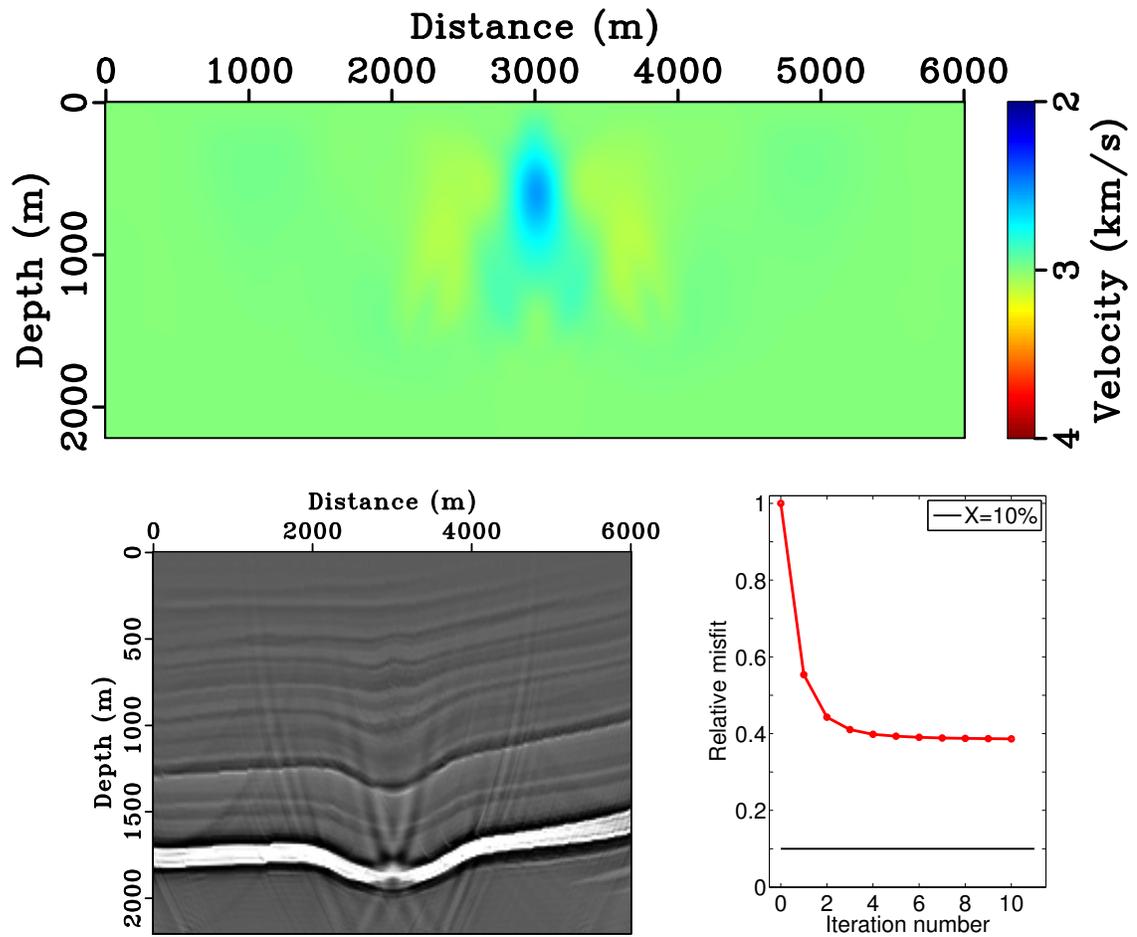
Figure 19: (a) Inverted background velocity model without model extension and multiscale approach (b) Reflectivity model; (c) Relative data residual.
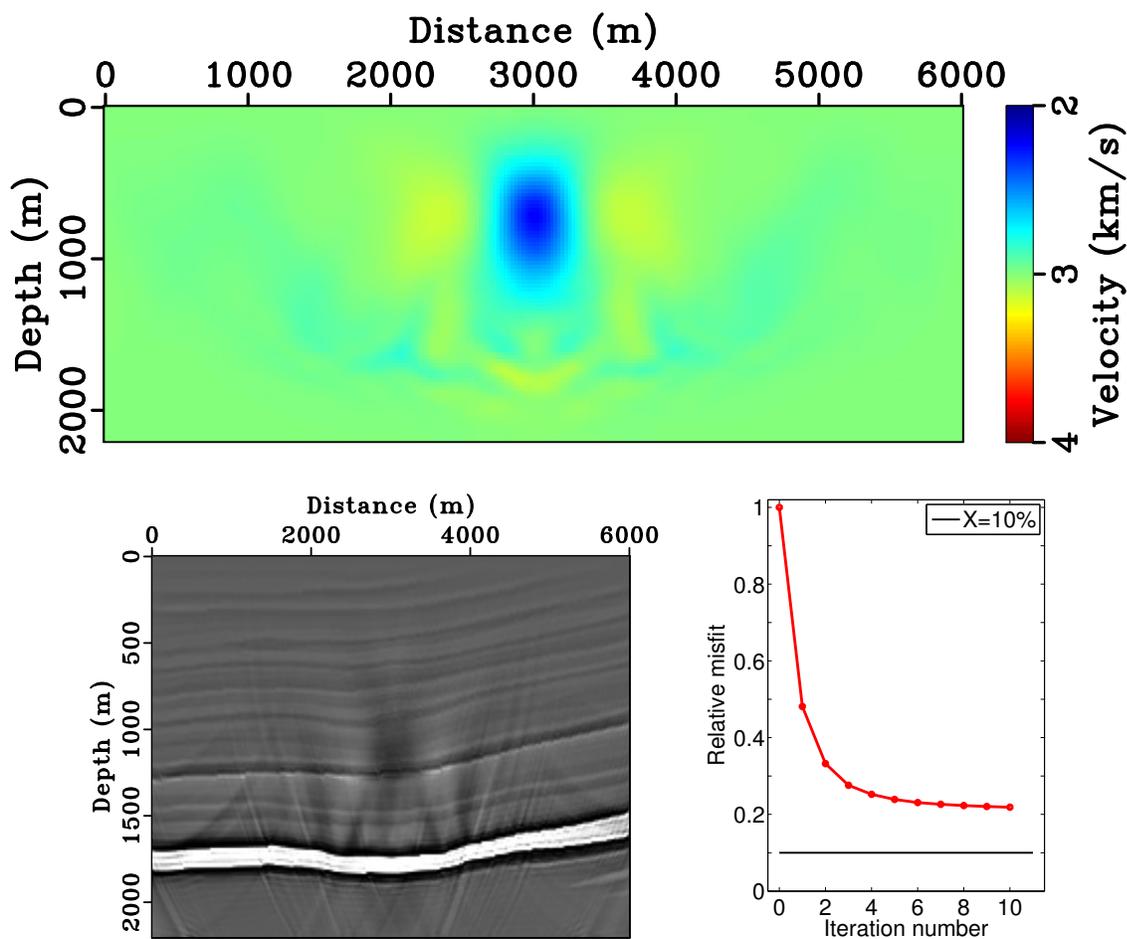
Figure 20: (a) Inverted background velocity model by using multiscale approach but without model extension; (b) Reflectivity model; (c) Relative data residual.

In this example, we used the full Algorithm 4, including the update for α, which resulted in increases of α at four steps ($\alpha = 1.1 \times 10^{-6}, 3.2 \times 10^{-5}, 1.1 \times 10^{-4}, 2.4 \times 10^{-4}$ at iteration 1, 2, 4, 13). More details can be found 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 ($1.5\,km/s$). The optimal subsurface offset range $H = 1.6\,km$ is 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 be barely observed from the ELSRTM image. In the subsurface offset gather, the downward curves indicate slow velocity. After 1st velocity update, the energy becomes more focused towards $h = 0$. Actually, only half of the extension ($0.8\,km$) would provided good data fit. With 2nd velocity update, layer structures are imaged closer to their correct positions. Although the energy is more focused in the gather, half subsurface offset is not adequate to provide good data fit. However, as the background velocity continues to be updated, after 8th iteration, offset $H = 0.4\,km$ is long enough. At iteration 15 and 23, we are able to reduce the offset range by half. As the convergence condition is satisfied at iteration 26, the inversion proceeds to the second stage.

Stage 2. The source wavelet and observed data are filtered by lowpass filter ($5 - 20\,Hz$). At the same time, the spatial grid size and the time step interval are decreased by a factor of 2. With the inverted background velocity model from 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 about $x = 4\,km, z = 2\,km$). 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 h direction.

DISCUSSION

A useful point of view on the range of subsurface offset H, kindly suggested by an anonymous reviewer, relates 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 (non-aliased) angle sampling, and vis-versa. When the velocity is nearly correct, the angle gathers are
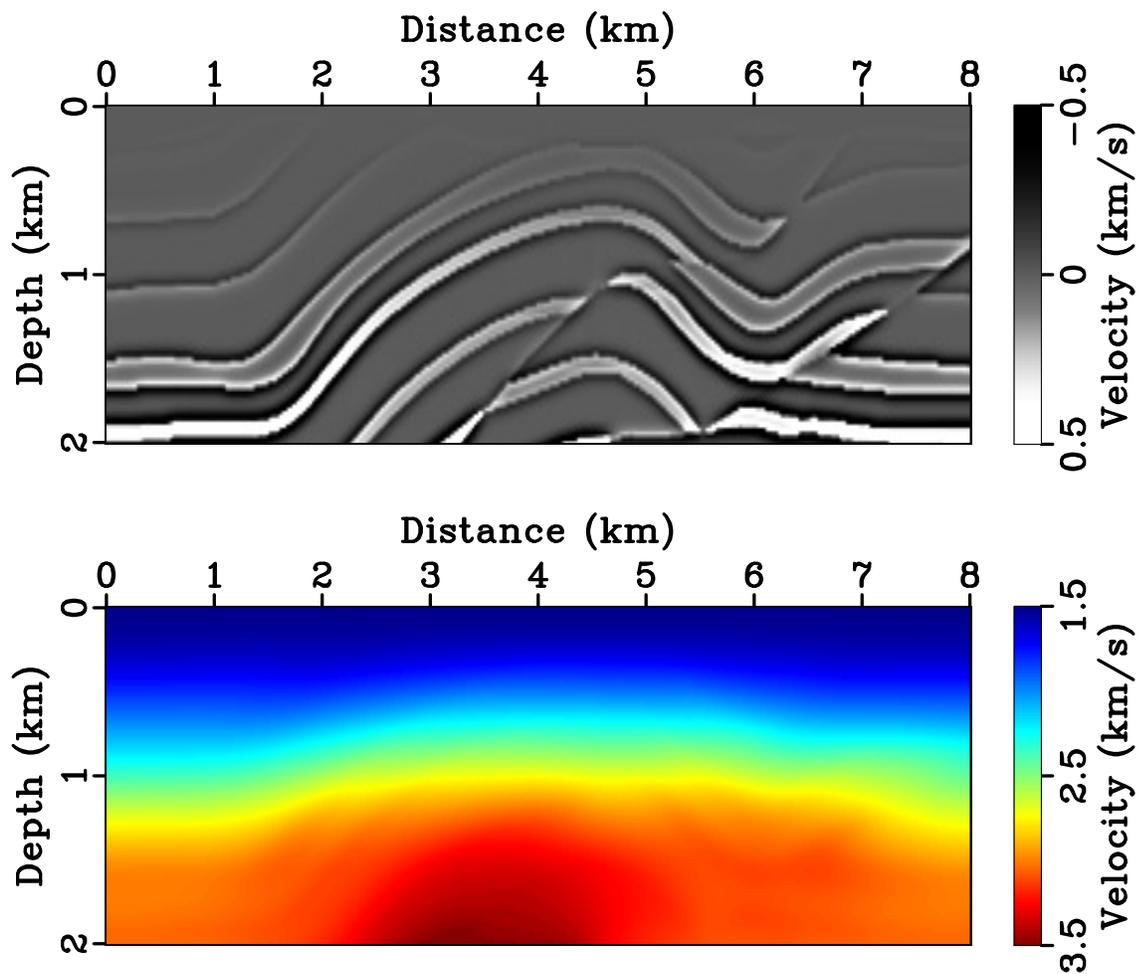
Figure 21: (a) Extended velocity perturbation $\delta\bar{r}$ at $h = 0\,m$ (b) The true background velocity
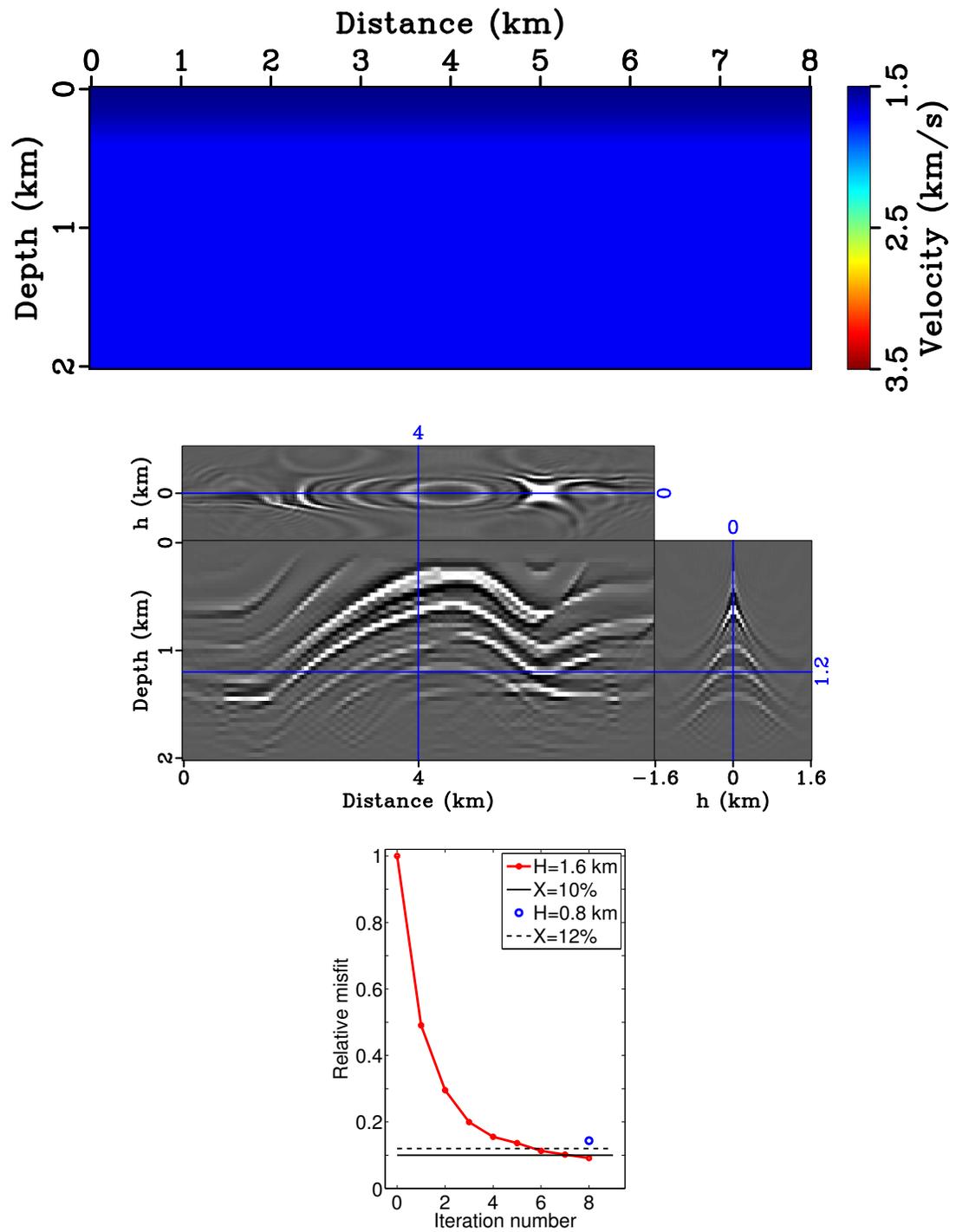
Figure 22: (a)The initial background used in inversion; (b) The ELSRTM image by using initial background velocity; (c) The relative data residual with different offset ranges.
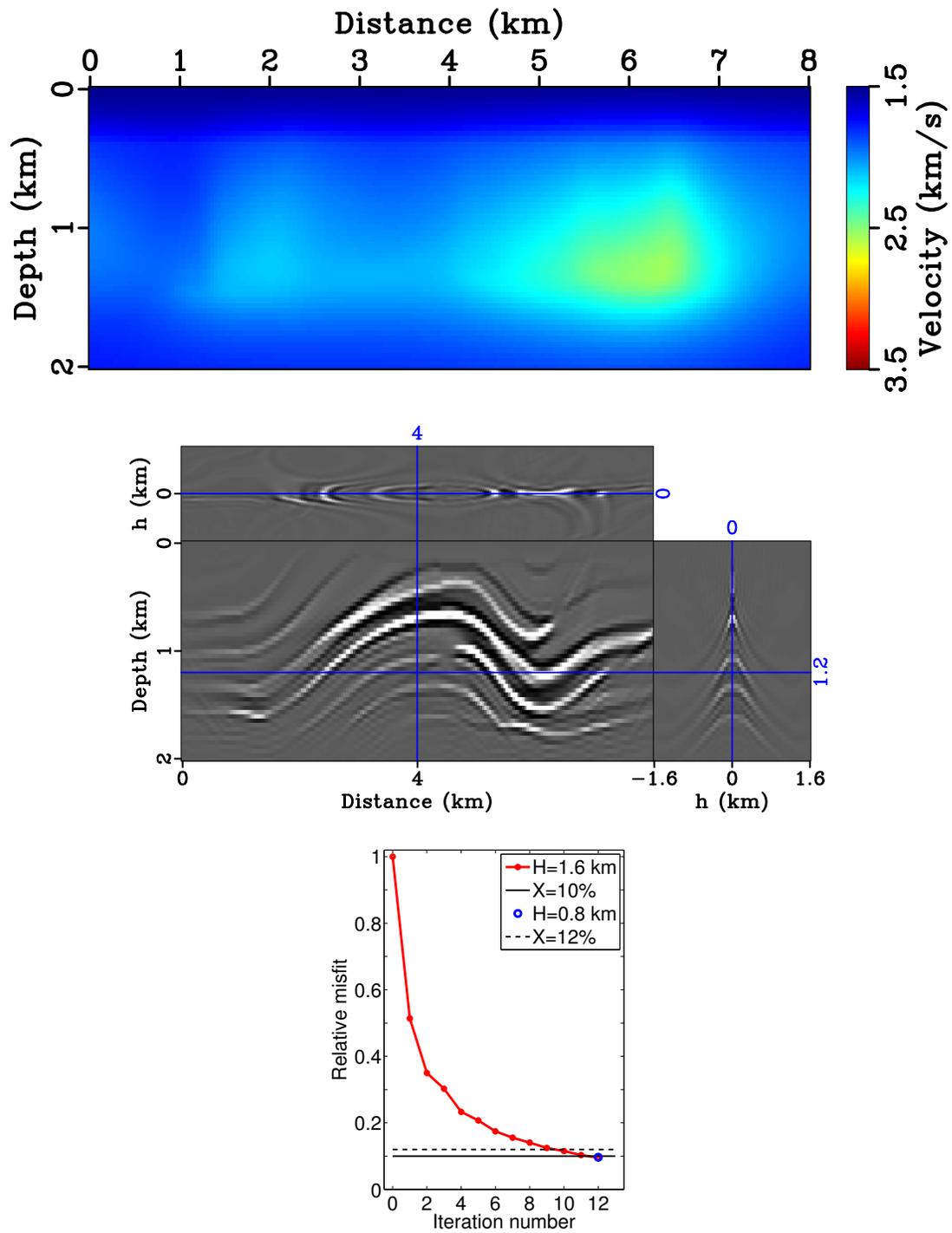
Figure 23: (a)The 1st updated background velocity; (b) The ELSRTM image by using 1st updated background velocity; (c) The relative data residual with different offset ranges.
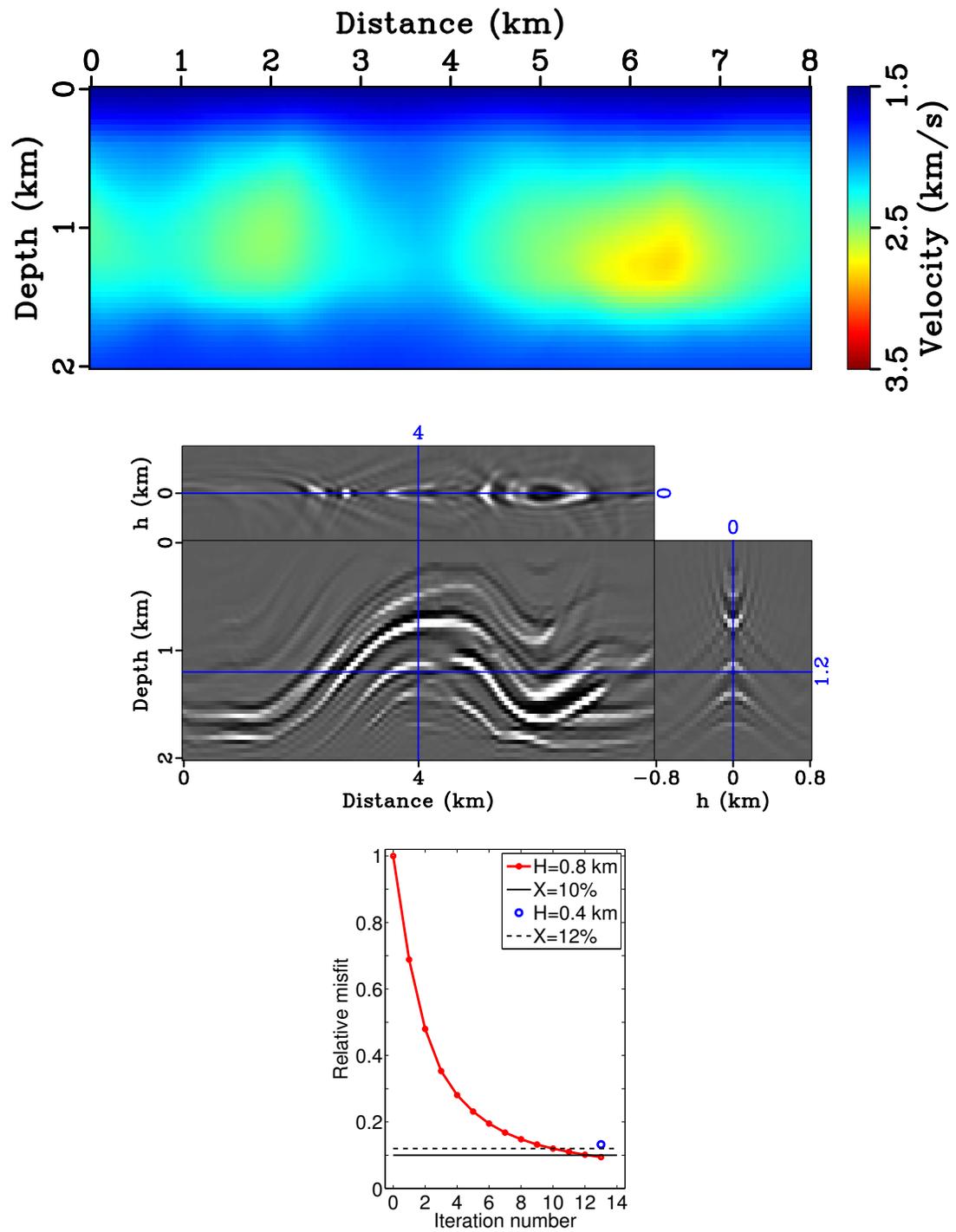
Figure 24: (a) The 2nd updated background velocity; (b) The ELSRTM image by using 2nd updated background velocity; (c) The relative data residual with different offset ranges.
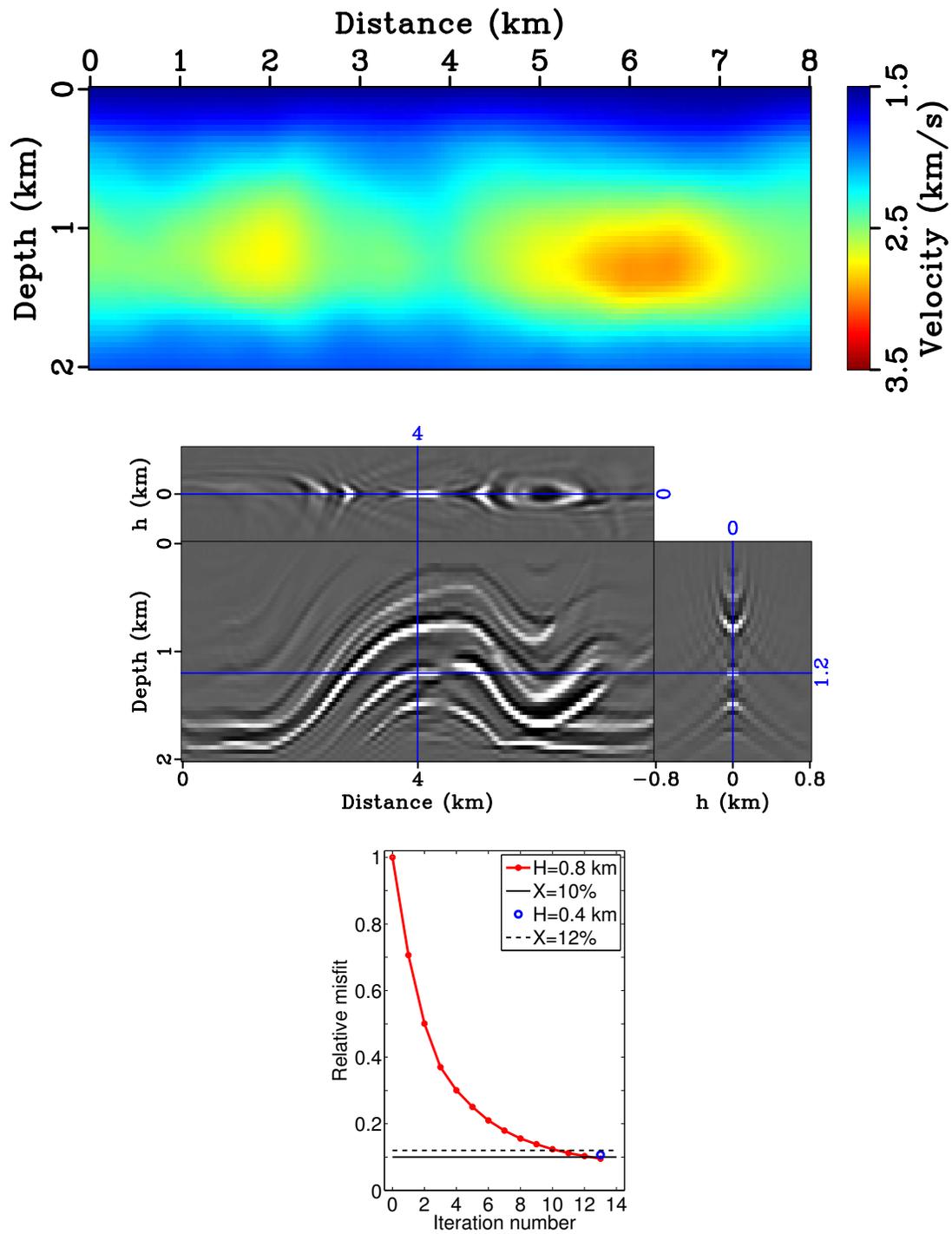
Figure 25: (a) The 8th updated background velocity; (b) The ELSRTM image by using 8th updated background velocity; (c) The relative data residual with different offset ranges.
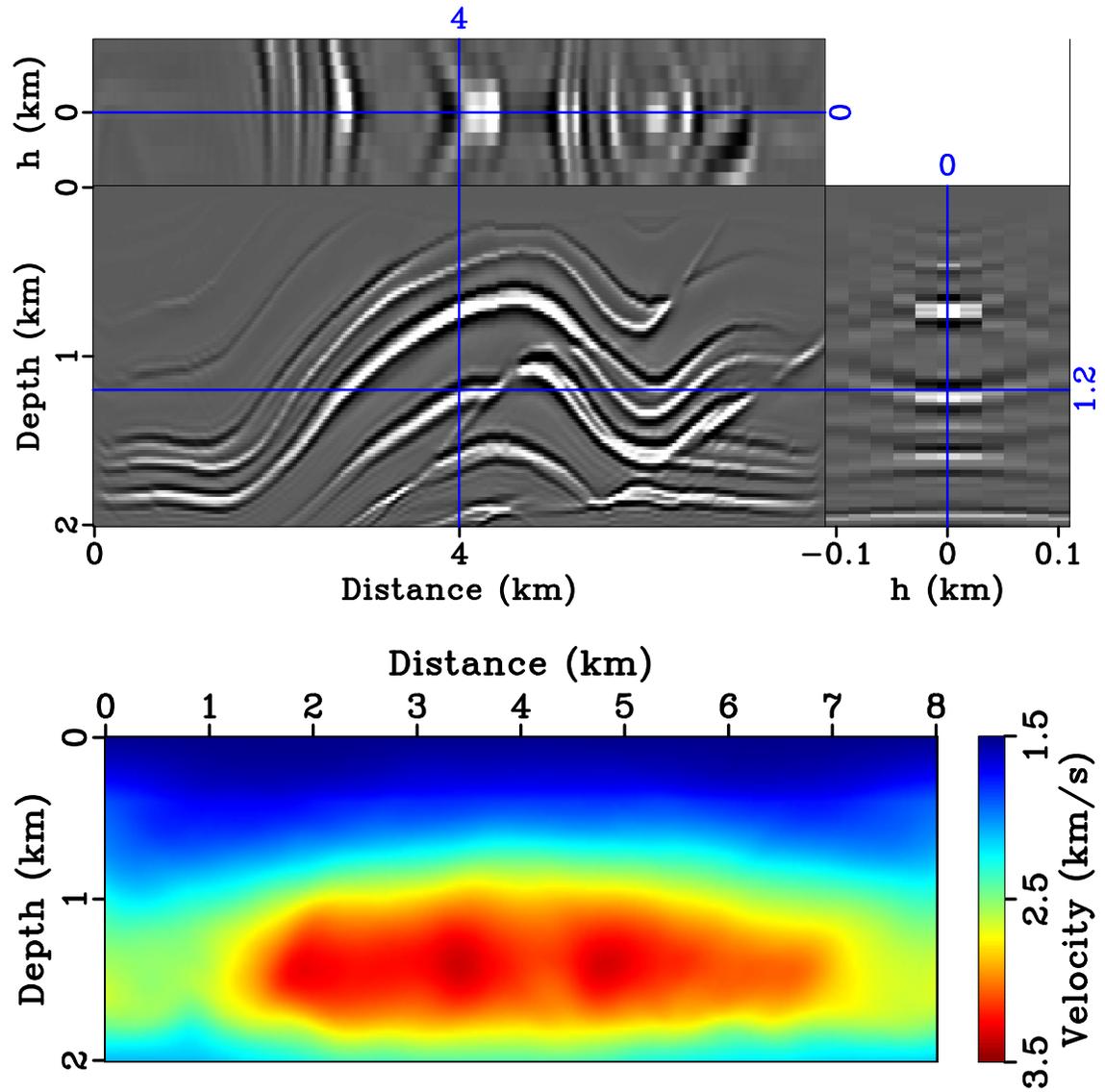
Figure 26: Stage 2 (a) The ELSRTM image; (b) The inverted background velocity.

nearliy flat, and can be reliably sampled with a coarse angle step, corresponding to small H. For substantially incorrect velocity, the angle gathers are far from flat in angle, implying the need for finer sampling for non-aliased representation, corresponding to larger maximum H. The angle-offset correspondence thus provides an alternate view of the relation between 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 about 1% of the original problem by using finest gird and full offset all along, if we assume the convergence rate are the same for both cases. For the overthrust example, although there are only 2 stages, the offset range is dramatically reduced at the first stage with adaptive approach, the cost is also greatly reduced (approximately 2% of the original problem). Although there are only 2 refinement stages, the subsurface offset is greatly reduced at the first few iterations. Note that with more refinements and more iterations for both inner loop and outer loop, the computational performance of the adaptive algorithm will be further improved in comparison with the non-adaptive 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 RMS error attainable with a *physical* (non-extended) model. With one more level of adaptation, our algorithm could attain this goal as well. Since driving the penalty term ($p(v, \alpha, H)$ in the theory section) towards zero is equivalent to approaching the space of physical models, adjusting X so that α can become sufficiently large, or H 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, remains a topic of research.

Note that the extended inversion example converges to a good kinematic solution, while non-extended FWI may not (see lens example) provided with the same number of modeling/migration pairs. There are more ways to reduce the number of iterations. In solving the normal equation 7 by 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, we reminded the reader that subsurface offset modeling and inversion involves (at the discrete level) a full matrix multiply at each time step, the cost of which can easily overwhelm the cost of 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 two-dimensional 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 inline and crossline sampling typical of WAZ streamer surveys, and of the sparse sampling of nodes in OBS surveys? Is even 3 Hz low enough frequency 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 problem of subsurface offset extended waveform inversion, 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 coarsen-

ing, yields a significant improvement in computational efficiency. The adaptive concept progressively reduces the subsurface offset range throughout inversion, which can naturally bring extended model towards physical model. 2D synthetic experiments with a version of this algorithm based on linearized (Born) constant density acoustics suggests that cost reduction of one or two order of magnitude is attainable, with recovery of both 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, Expanded Abstracts, Society of Exploration Geophysicists, TOM2.1.

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, Expanded Abstracts, Society of Exploration Geophysicists, SI9.5.

———, 2014, Simultaneous inversion of full data bandwidth by tomographic full-waveform inversion: Geophysics, **79**, WA129–WA140.

Biondi, B., and G. Shan, 2002, Prestack imaging of overturned reflections by reverse time migration: 72nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 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, Expanded Abstracts, Society of Exploration Geophysicists, 4071–4076.

Fu, L., and W. Symes, 2015, Reducing the cost of extended waveform inversion by multiscale adaptive methods: 85th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1127–1131.

———, 2016, Fast extended waveform inversion using Morozovs discrepancy principle, *in* 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.

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.

———, 2003, Separable nonlinear least squares: the variable projection method and its applications: Inverse Problems, **19**, R1–R26.

Hou, J., and W. Symes, 2016, Accelerating extended least-squares migration with weighted conjugate gradient iteration: Geophysics, **81**, no. 4, S165–S179.

Huang, Y., and W. Symes, 2015, Born waveform inversion via variable projection and shot record model extension: 85rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 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.

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.

Li, M., J. Rickett, and A. Abubakar, 2013, Application of the variable projection scheme to frequency-domain full-waveform inversion: Geophysics, **78**, R249–R257.

Liu, Y., W. Symes, and Z. Li, 2014, Inversion velocity analysis via differential semblance optimization: Presented at the 76th Annual Meeting, Extended Abstracts, European Association of Geoscientists and Engineers.

Mulder, W., 2014, Subsurface offset behaviour in velocity analysis with extended reflectivity images: Geophysical Prospecting, **62**, 17–33.

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.

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, Expanded Abstracts, Society of Exploration Geophysicists, 930–934.

Rickett, J., 2012, The variable projection method for waveform inversion with an unknown source function: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1–5.

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

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

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

———, 2012, An RTM based automatic migration velocity analysis in image domain: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SVE2.2.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

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

Expanded Abstracts, Society of Exploration Geophysicists, SPVA1.4.

Shen, P., W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 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.

Symes, W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

——, 2014, Seismic inverse problems: recent developments in theory and practice: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 2–5.

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.

ten Kroode, F., 2014, A Lie group associated to seismic velocity estimation: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 142–146.

van Leeuwen, T., and W. Mulder, 2009, A variable projection method for waveform inversion: 71st Annual Meeting, Extended Abstracts, European Association of Geoscientists and Engineers, U024.

Vigh, D., K. Jiao, W. Huang, N. Moldoveanu, and J. Kapoor, 2013, Long-offset-aided Full-waveform Inversion: 75th Annual Meeting, Expanded Abstracts, European Association of Geoscientists and Engineers, 76th Annual Meeting.

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, Expanded Abstracts, Society of Exploration Geophysicists, 957–961.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152.

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

Weibull, W., and B. Arntsen, 2013, Automatic velocity analysis with reverse time migration: Geophysics, **78**, S179–S192.

Scattering and dip angle decomposition based on subsurface offset extended wave-equation migration

Raanan Dafni and William W. Symes

ABSTRACT

An angle-dependent reflection coefficient is recovered byseismic migration in the angle domain. We have developed a postmigration technique for computing scattering and dip angle common-image gathers (CIGs) from seismic images, extended by the subsurface offset, based on wave-equation migration methods. Our methodology suggests a system of Radon transform operators by introducing local transform relations between the subsurface offset image and the angle-domain components. In addition to the commonly used decomposition of the scattering angle, the methodology associates the wave-equation migration with dip-domain images as well. The same postmigration subsurface offset image is used to decompose scattering and dip angle CIGs individually or to decompose a multiangle CIG by showing simultaneously both angles on the gathers axis. We show that the dip-angle response of seismic reflections is a spot-like signature, focused at the specular dip of the subsurface reflector. It differs from the well-studied smile-like response usually associated with reflections in the dip domain. The contradiction is clarified by the nature of the subsurface offset extension, and by emphasizing that the angles are decomposed from the subsurface offset image after the imaging condition, without directly involving the propagating incident and scattered wavefields. Several synthetic and field data tests proved the robustness of our decomposition technique, by handling various subsurface models, including seismic diffractions. It is our belief that dip-angle information, decomposed by wave-equation migration, would have a great impact in making the scattering-angle reflection coefficient more reliable and noise free, in addition to the acceleration of wave-equation inversion methods.

259

http://www.trip.caam.rice.edu/reports/2014/trip2016_report.html

Kinematic artifacts in the subsurface-offset extended image and their elimination by a dip-domain specularity filter

Raanan Dafni and William W. Symes

ABSTRACT

Common-image gathers in the dip-angle domain may be computed in relation to wave-equation migration methods, extended by the subsurface offset. They involve the application of a postmigration local Radon transform on the subsurface-offset extended image. In the dip-angle domain, seismic reflections are focused around the specular dip angle of reflection. This focusing distinguishes them from any other event in the image space. We have incorporated the dip-angle information about the presence of specular reflections into the computation of the conventional scattering-angle-dependent reflection coefficient. We have designed a specularity filter in the dip-angle domain based on a local semblance formula that recognizes and passes events associated with specular reflections, while suppressing other sorts of nonspecular signal. The filter is remarkably effective at eliminating either random or coherent noises that contaminates the prestack image. In particular, our dipangle filter provides a method for the suppression of kinematic artifacts, commonly generated by migration in the subsurfaceoffset domain. These artifacts are due to an abrupt truncation of the data acquisition geometry on the recording surface.We have studied their appearance and devised an appropriate formation mechanism in the subsurface-offset and scattering-angle domains. The prominent presence of the kinematic artifacts in image gathers usually impairs the quality of the postmigration analysis and decelerates the convergence of wave-equation inversion techniques. We have determined from testing on synthetic and field data that using the proposed dip-angle-domain specularity filter efficiently eliminates the kinematic artifacts in the delivered gathers. We expect involvement of the specularity filter to increase the reliability and quality of the seismic processing chain and provide a faster convergence of iterative methods for seismic inversion.

Note: This paper appeared in *Geophysics*, **81** (6), S477-S495 (2016). See TRIP 2016 Annual Report web page for doi and link:

`http://www.trip.caam.rice.edu/reports/2014/trip2016_report.html`

Simple AGC-free Preprocessing for Land Data

Raanan Dafni and William W. Symes, *The Rice Inversion Project*

ABSTRACT

A simple workflow using open source software (SU and Madagascar) produces reasonable stacked sections from field data collected at the National Petroleum Reserve, Teapot Dome, Wyoming.

INTRODUCTION

The goal of this paper is to produce stacked sections of quality resembling that of commercial processing without doing anything irreversible to the data, such as ACG. The intent behind this goal is to extract a useful image of the subsurface while maintaining the original amplitude and phase of the data throughout the process, as an initial approximation to waveform inversion.

METHODS

We have used standard SU utiliities to implement our workflow. One exception is the application of statics: rather than `sustatic` , we assigned the total static correction from the trace headers (presumably supplied by the commercial processing firm) to the `delrt` header word, and used `sushift` to apply the corresponding shift. We took this approach because `sustatic` appears to apply a bulk shift to the output, and moreover requires weathering velocity and datum specification which should not be necessary inputs to a simple time shift, for reasons not explained in the self-doc.

Our efforts to reproduce the field tape results using `sftah` functions rather than their SU equivalents have so far been unsuccessful.

The specific results are explained in the captions, below.

To do: surface-consistent amplitude scaling and decon - these are inversion-compatible steps.

RESULTS
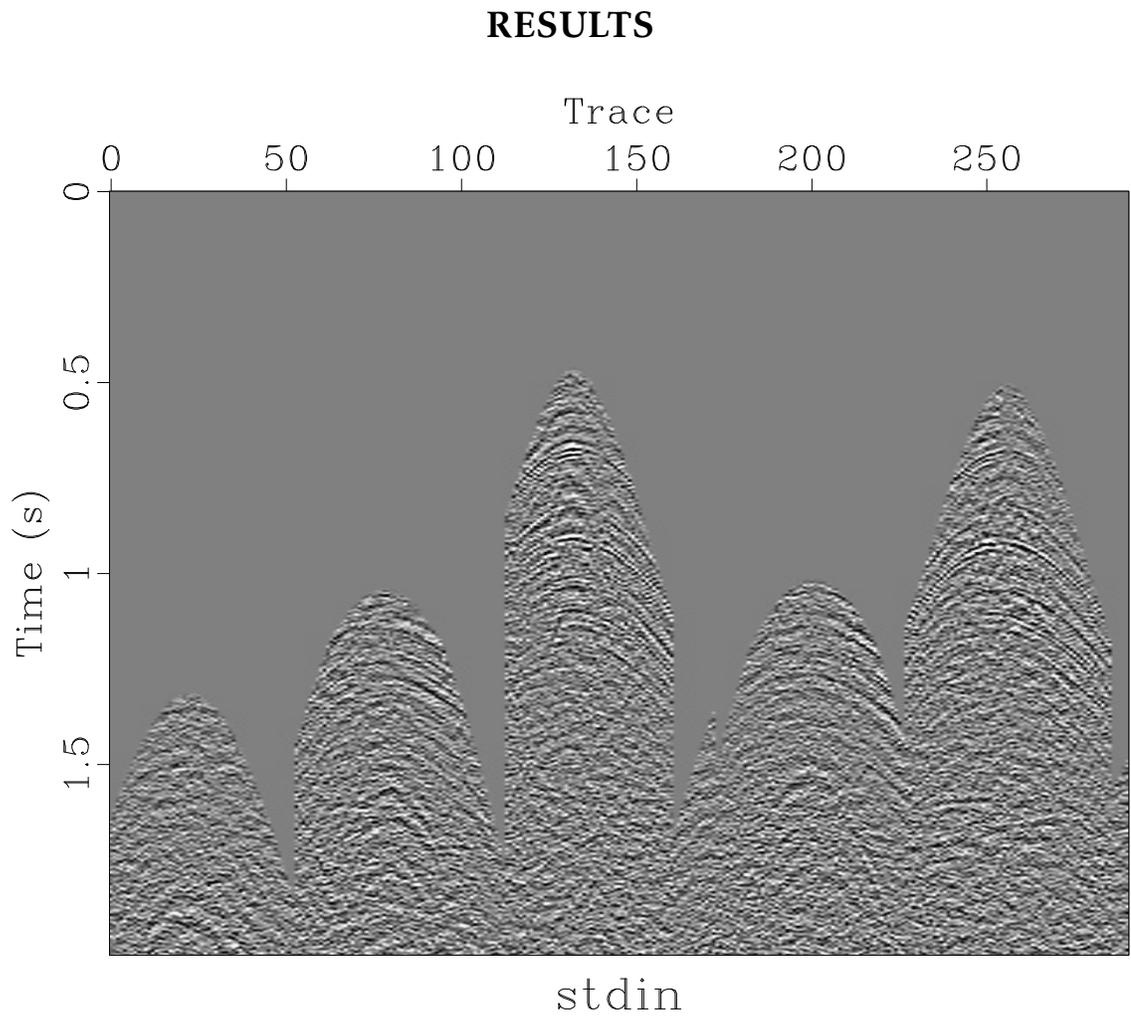
Figure 1: Commercially preprocessed shot gathers from inline 141. teapot/./project/ iline141
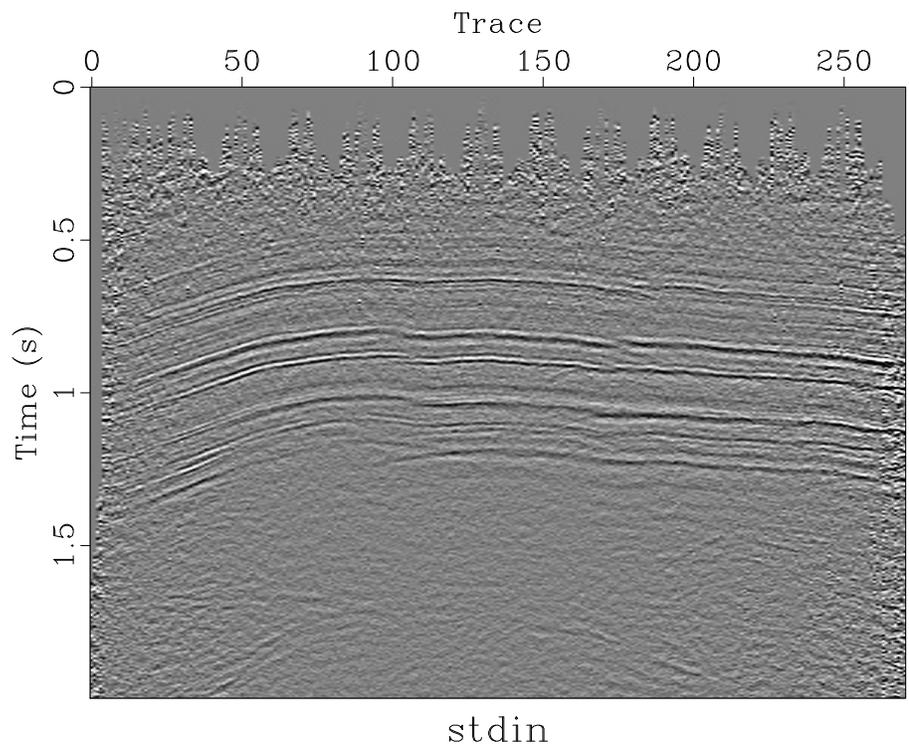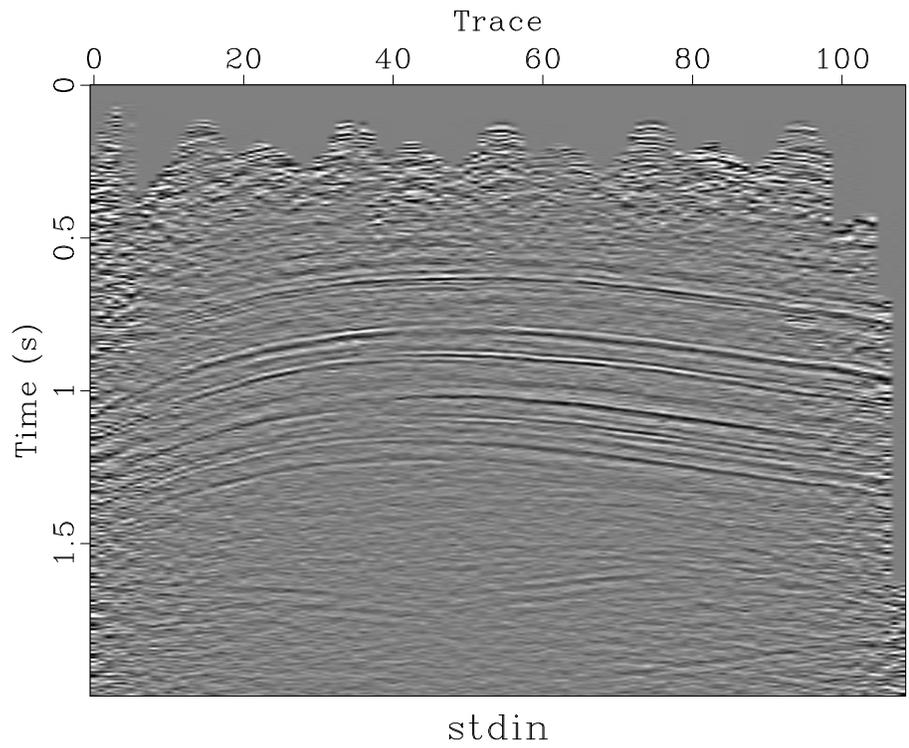
Figure 2: NMO-Stack of commercially preprocessed data using first velocity function in the file: (a) inline, (b) crossline. teapot/./project/ stackiline141,stackxline121
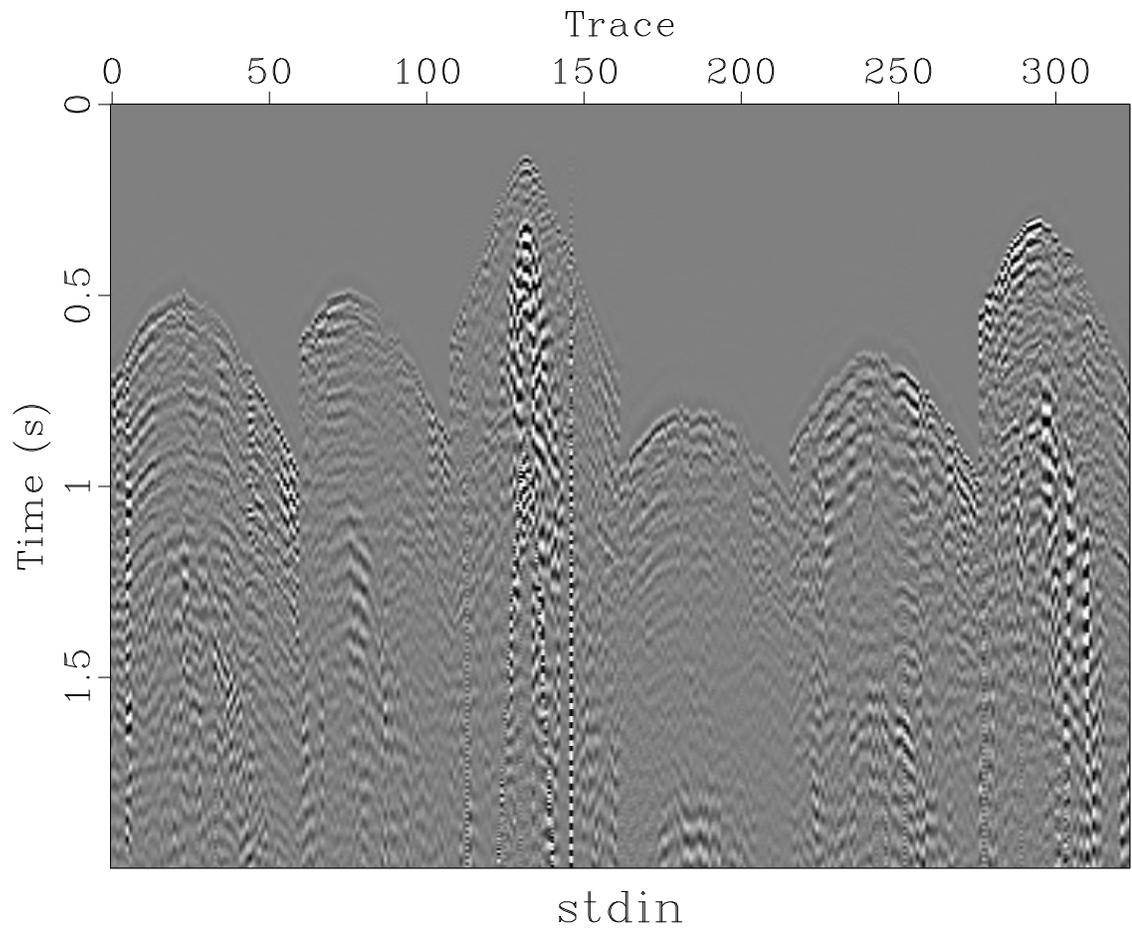
Figure 3: Shot gathers from raw data, inline 141. 80080 ft < sx < 80200 ft. teapot/./project/ filine141
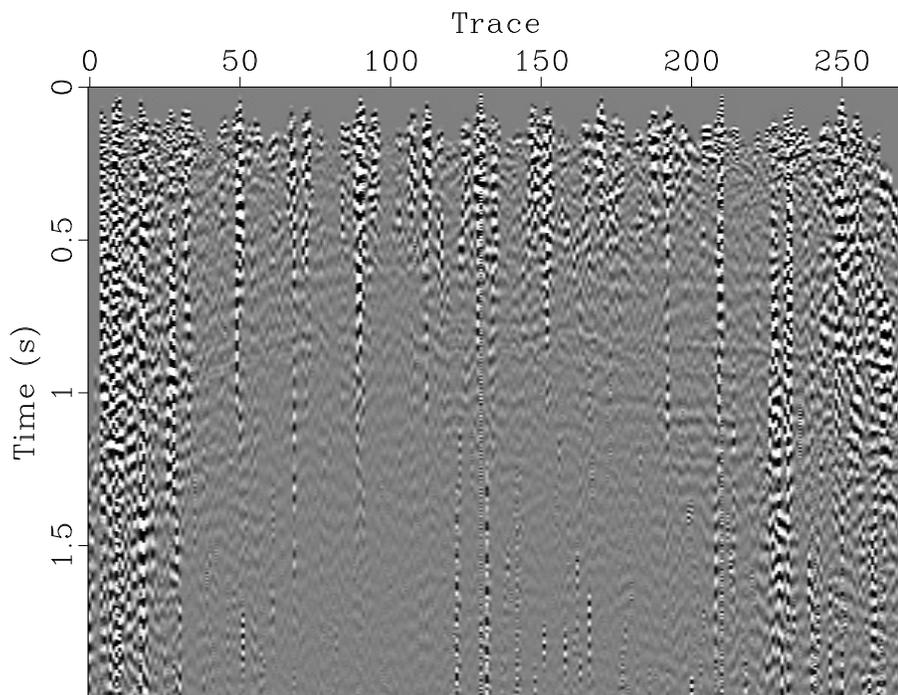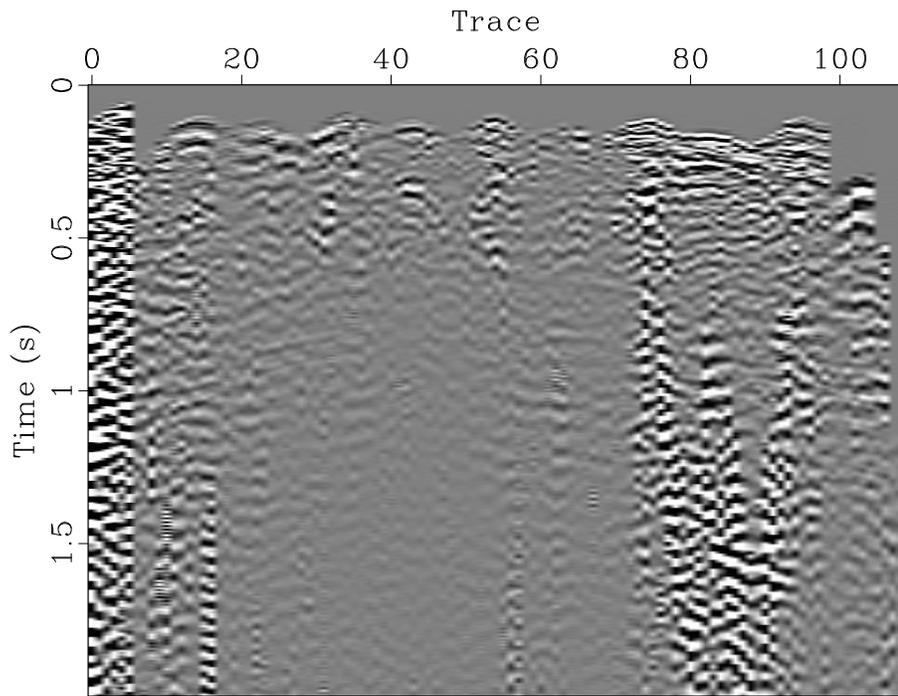
Figure 4: NMO-Stack of raw field data, t^2 scaling, otherwise no preprocessing: (a) inline, (b) crossline. teapot/./project/ fstackiline141,fstackxline121
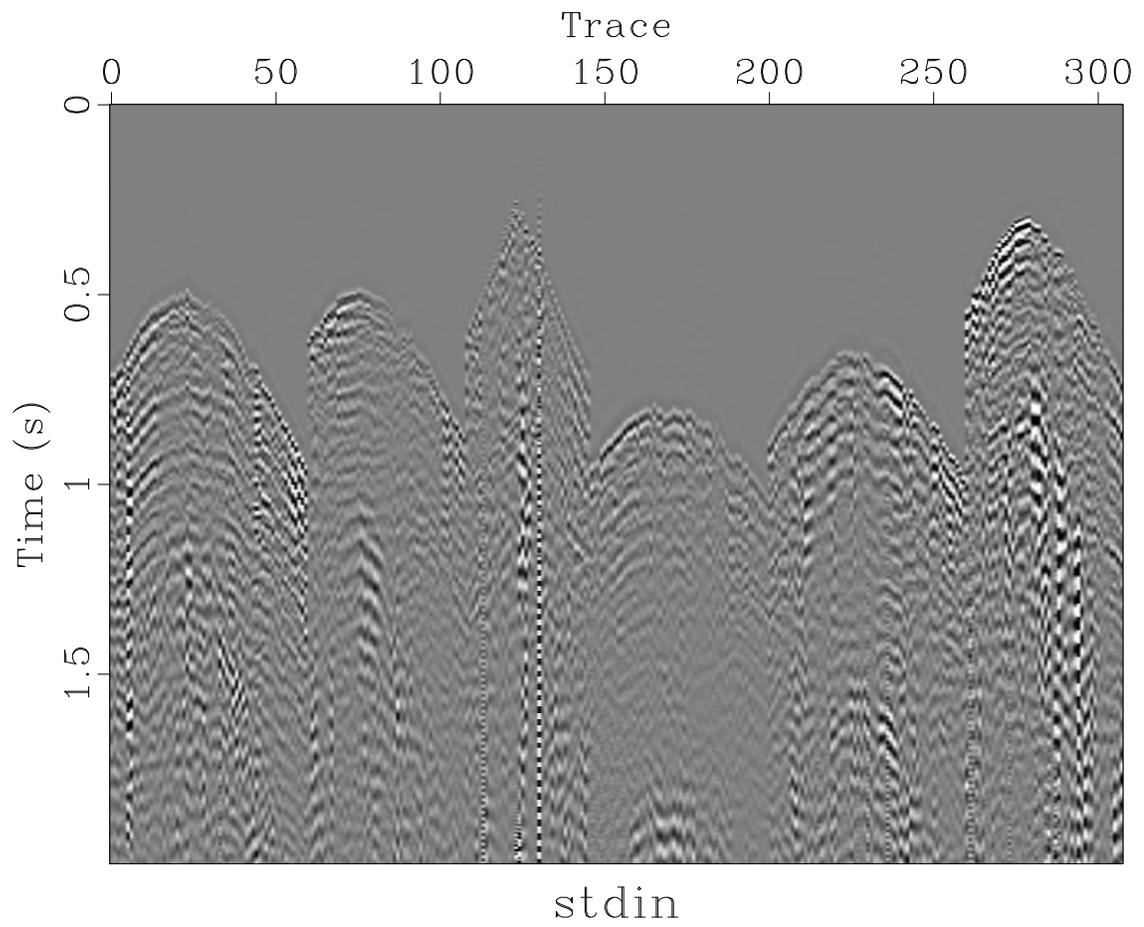
Figure 5: Shot gathers from raw data, inline 141, after deletion of near offset traces. 80080 ft < sx < 80200 ft. teapot/./project/ ffoiline141
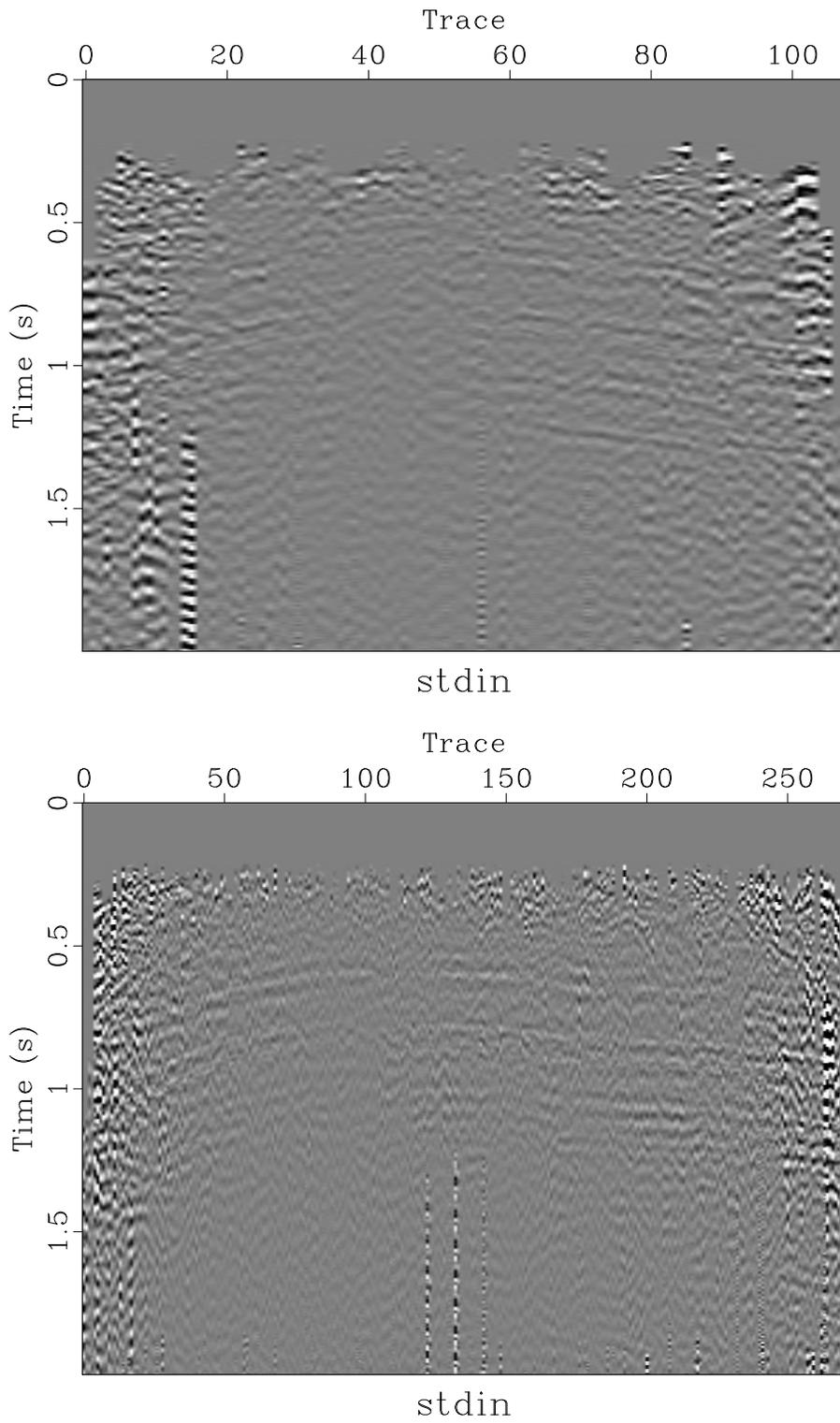
Figure 6: NMO-Stack of raw field data, t^2 scaling, far offsets only (> 2000 ft) : (a) inline, (b) crossline. teapot/./project/ fstackfoiline141,fstackfoxline121
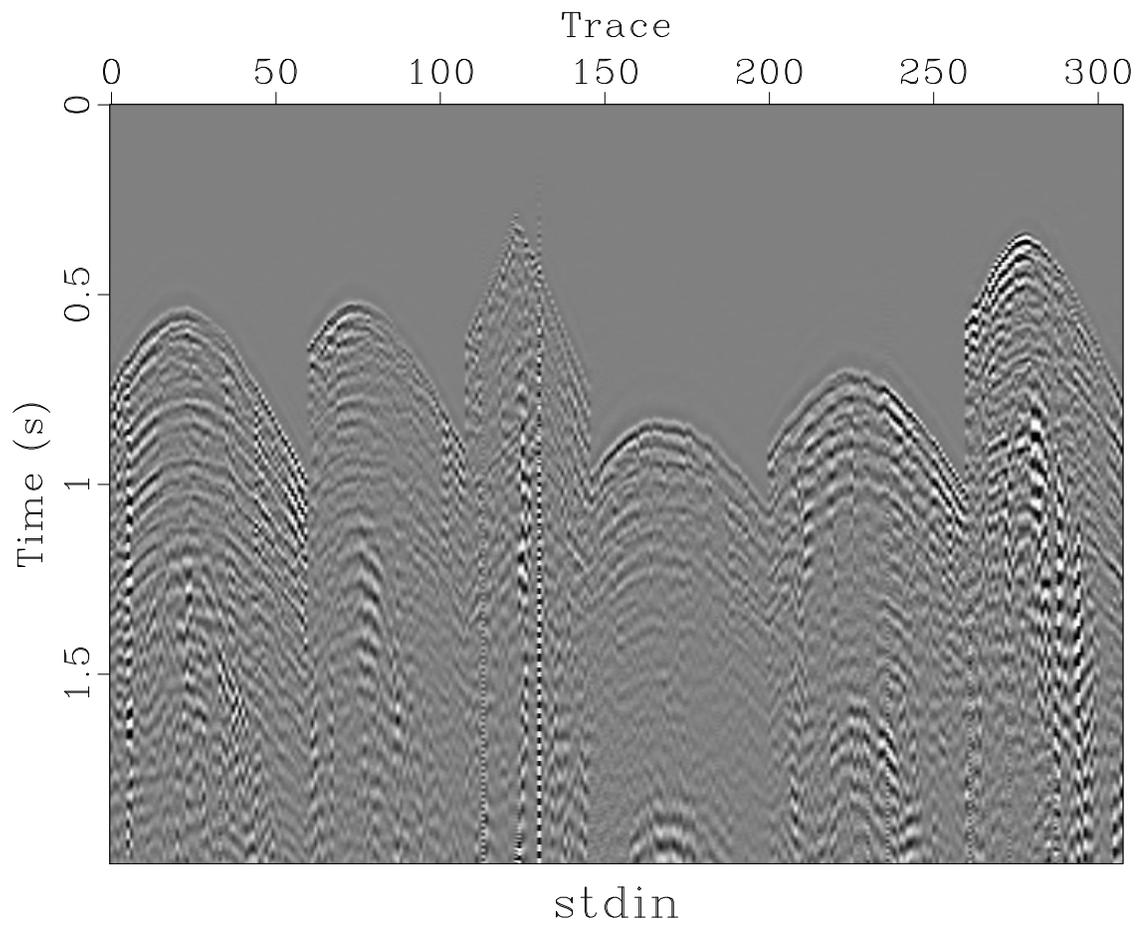
Figure 7: Shot gathers from raw data, inline 141, after deletion of near offset traces, and application of statics. 80080 ft < sx < 80200 ft. teapot/./project/ ffoepiline141
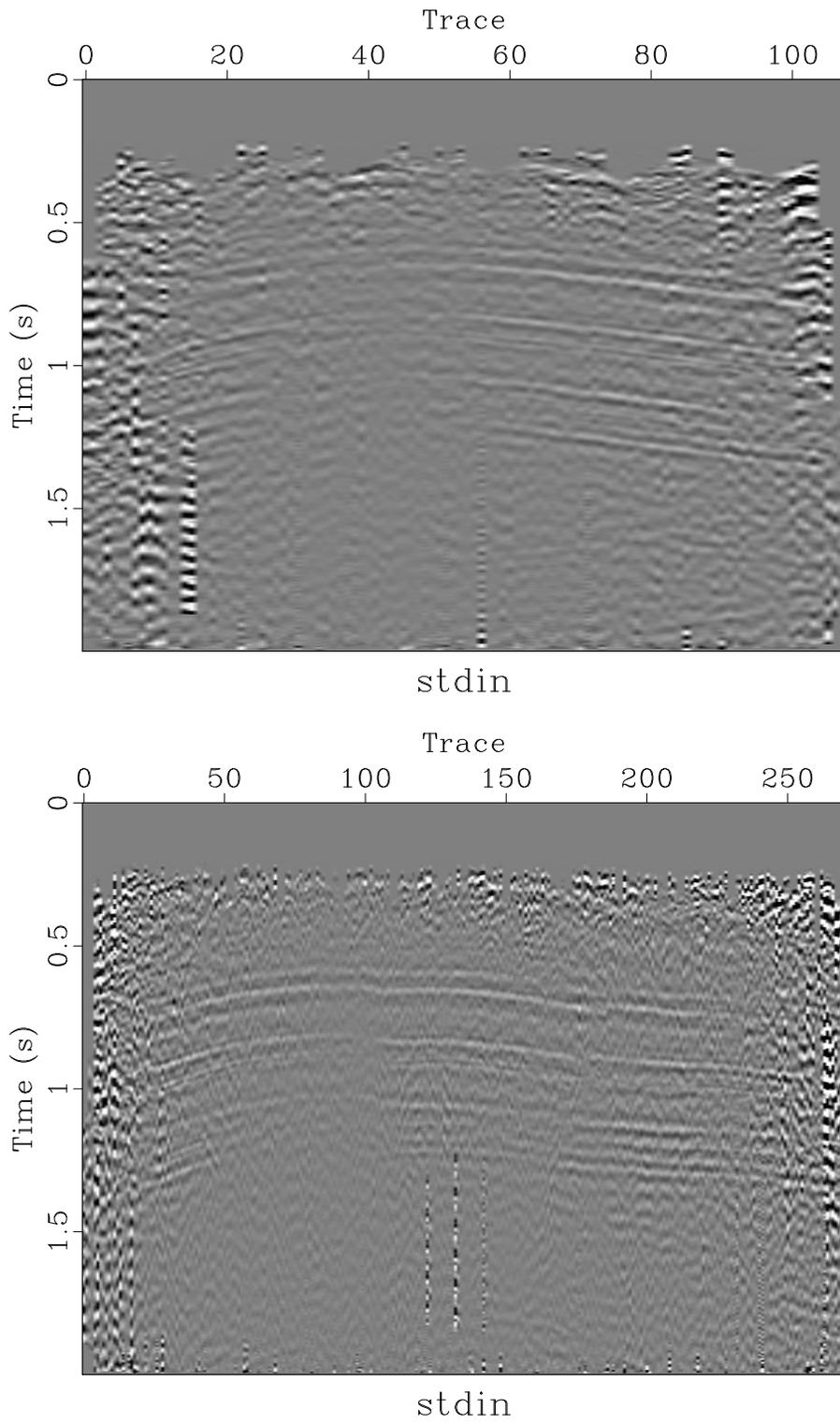
Figure 8: NMO-Stack of raw field data, t^2 scaling, far offsets only (> 2000 ft), static corrections applied per trace headers sstat, gstat: (a) inline, (b) crossline. teapot/./project/ fstackfoepiline141,fstackfoepxline121
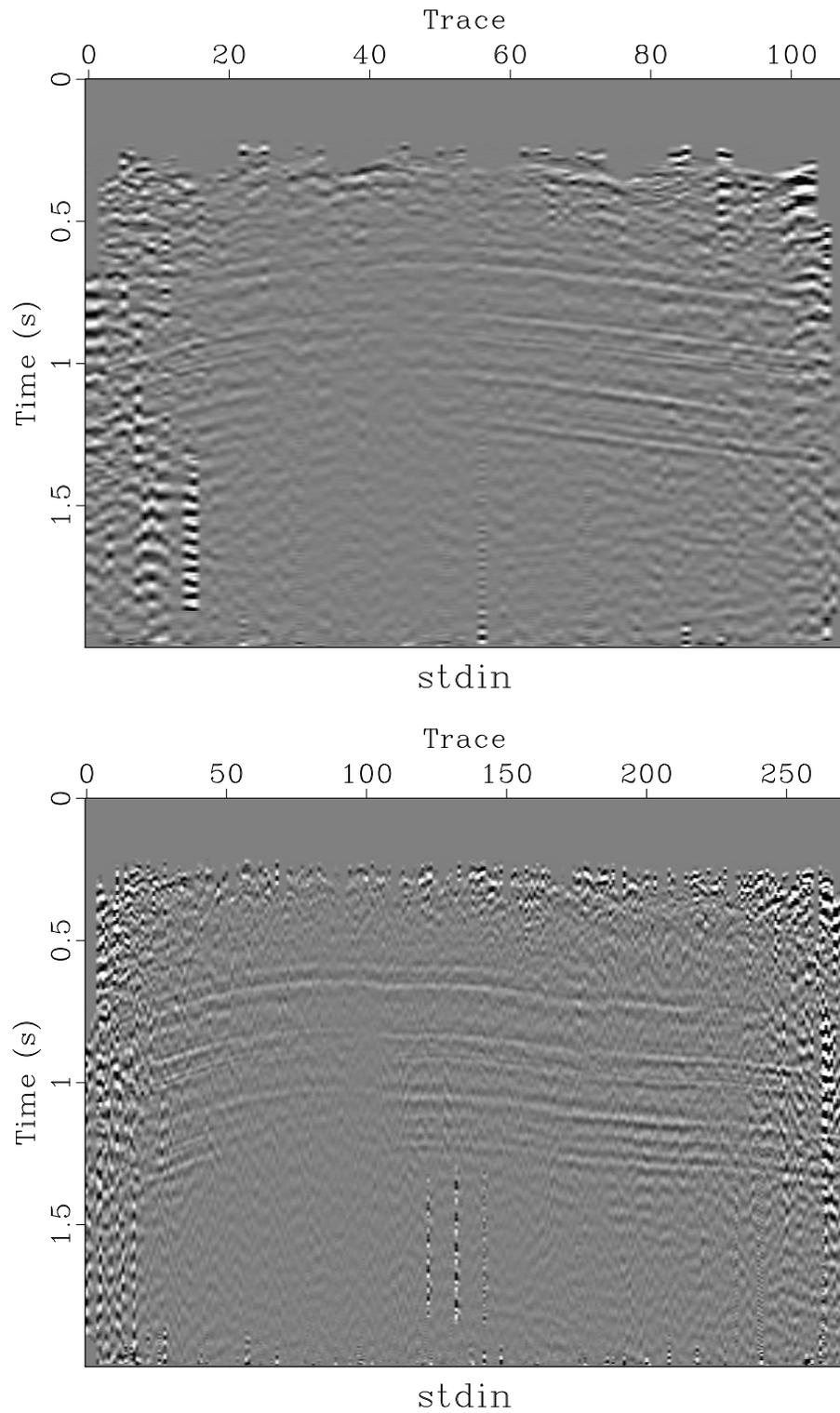
Figure 9: NMO-Stack of raw field data, using NMO velocity function from CDP at junction of inline 141 and crossline 121. t^2 scaling, far offsets only (> 2000 ft), static corrections applied per trace headers `sstat`, `gstat`. Note that washed-out region in previous display near trace 100 in the crossline section is enhanced. (a) inline, (b) crossline.
teapot/./project/ fbstackfoepiline141,fbstackfoepxline121

Inversion velocity analysis in subsurface offset domain

Jie Hou[**], *William W. Symes*[†]

ABSTRACT

Optimization-based Migration Velocity Analysis (MVA) updates long wavelength velocity information by minimizing an objective function that measures the violation of a semblance condition, applied to an image volume. Differential Semblance Optimization (DSO) forms a smooth objective function both in velocity and data, regardless of the data frequency content. Depending on how the image volume is formed, however, the objective function may not be minimized at a kinematically correct velocity, a phenomenon characterized in the literature (somewhat inaccurately) as "gradient artifacts". In this paper, we will show that the root of this pathology is imperfect image volume formation resulting from various forms of migration, and that the use of linearized inversion (least squares migration) more or less eliminates it. We demonstrate that an approximate inverse operator, little more expensive than Reverse Time Migration (RTM), leads to recovery of a kinematically correct velocity.

INTRODUCTION

Full waveform inversion (FWI) (Lailly, 1983; Tarantola, 1984; Virieux and Operto, 2009) is capable of recovering detailed models of the subsurface structure through a waveform-based data-fitting procedure. However, it may stagnate at physically meaningless solutions in the absence of a kinematically accurate starting model. This is the notorious cycle skipping problem. Within the limitation to single scattering, migration velocity analysis (MVA) (Yilmaz and Chambers, 1984; Yilmaz, 2001) complements FWI by extracting long scale velocity. Image domain MVA involves construction of an extended image volume, which not only depends on the

---

Parts of this paper were presented at 2016 SEG annual meeting

image point but also on an extra parameter (e.g surface offset, subsurface offset, incidence angle, source wavelet). When velocity and data are kinematically compatible, this volume should have particularly simple structure (flat, focused,...). Deviations from this semblance principle can be used to drive velocity updates, either by direct measurement (eg. residual moveout picking, (Stork, 1992; Lafond and Levander, 1993; Liu and Bleistein, 1995; Biondi and Sava, 2004)) or via optimization of an objective function.

This paper focuses on a particular choice of optimization approach, namely differential semblance optimization (DSO) in the subsurface offset domain (Symes, 2008). Implementations of this MVA approach have been based on double square root migration (Shen et al., 2003), one-way shot record migration (Shen and Symes, 2008), Reverse Time Migration (RTM) (Shen, 2012; Weibull and Arntsen, 2013), and various sorts of inversion (Biondi and Almomin, 2012; Liu et al., 2014; Lameloise et al., 2015). Both numerical and theoretical evidence (Symes, 2014; ten Kroode, 2014) suggest that this approach should be effective in recovering velocity macromodels under failure conditions for FWI. However other studies have suggested that the method may produce poor velocity update directions, in particular that the objective gradient may be contaminated with artifacts that prevent rapid convergence to a correct velocity (Fei and Williamson, 2010; Vyas and Tang, 2010).

We show here that use of (linearized) inversion to create image volumes largely eliminates the "gradient artifact" pathology, and describe a computationally efficient method to achieve this goal. In fact, the "artifacts" actually are features of the objective function definition, not of the gradients. This is not a new observation: Khoury et al. (2006) showed that subsurface offset DSO, using common azimuth migration to construct the image volume, could produce erroneous velocities returning lower objective function values than the "true" velocity. Liu et al. (2014); Lameloise et al. (2015); Lameloise and Chauris (2016) confirm this observation and show that use of inverted (rather than migrated) image volumes tends to improve DSO velocity updates, essentially because the inverted image volume is much better focused at the target velocity. Our innovation is to show that good velocity updates may be achieved with image volumes obtained by an approximation to linearized inversion, costing little more than RTM and involving no ray-theory computations (ten Kroode, 2012; Hou and Symes, 2015).

In this paper, we will compare two different imaging operators : the adjoint of the extended Born modeling operator (RTM operator) and an approximate inverse to the extended Born modeling operator. The adjoint operator essentially com-

putes the so-called RTM (with space-shift imaging condition) operator. Hou and Symes (2015) modify the adjoint operator into an approximate inverse operator by applying model and data-domain weight operators. An alternative formula, with same quality but substantially less implementation cost, is actually used in this paper (ten Kroode, 2014; Hou and Symes, 2017). To distinguish it from the other possibilities, we call MVA with the approximate inverse operator Inversion Velocity Analysis (IVA).

In the following sections, we will first review the theory of MVA via DSO in the subsurface offset domain; we then compare different imaging operators and their possible influence on the DSO objective function; we end with numerical test on 2D Marmousi model and a field data, demonstrating that better imaging leads to better velocities.

THEORY

In constant density medium, the 2D acoustic wave equation can be written in the following form:

$$\frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = f,\tag{1}$$

where v is acoustic-wave velocity, p is pressure filed and f represents the acoustic energy source. The forward modeling operator $\mathcal{F}$ maps the velocity model into seismic data (sampled pressure field) such that

$$\mathcal{F}[v] = p.\tag{2}$$

The nonlinear nature of the relationship between velocity model v and seismic data p makes it extremely difficult to recover velocity model v from the recorded seismic data.

A standard way to simplify the nonlinear inverse problem above is built on the (partial) linearization of nonlinear modeling operator $\mathcal{F}$, that is, the Born (single scattering) approximation. This linearization starts from the decomposition of the acoustic-wave velocity v into a reference velocity v_0 and a perturbation δv. To first order in δv, we have

$$\mathcal{F}[v] \approx \mathcal{F}[v_0] + F[v_0]\delta v,\tag{3}$$

in which $F = \dfrac{\partial \mathcal{F}}{\partial v}$ is the Born modeling operator. It depends nonlinearly on reference velocity v_0 and linearly on the perturbation velocity δv.

Denoting $p, p_0, \delta p$ as the pressure filed corresponding to $v, v_0, \delta v$, we can define the first order derivative of the original wave equation:

$$\frac{\partial^2 p_0}{\partial t^2} - v_0^2 \nabla^2 p_0 \;=\; f, \tag{4}$$

$$\frac{\partial^2 \delta p}{\partial t^2} - v_0^2 \nabla^2 \delta p \;=\; 2 v_0 \delta v \nabla^2 p_0. \tag{5}$$

As the solution of the coupled system above, the Born modeling operator $F[v_0]$ can be expressed as

$$
\begin{aligned}
(F[v_0]\delta v)(\mathbf{x_s}, \mathbf{x_r}, t) = {} & \frac{\partial^2}{\partial t^2} \int d\mathbf{x}\, dh\, d\tau\, G(\mathbf{x_s}, \mathbf{x}, \tau) \\
& \times \frac{2\delta v(\mathbf{x})}{v_0(\mathbf{x})^3} G(\mathbf{x}, \mathbf{x_r}, t - \tau).
\end{aligned}
\tag{6}
$$

Here δv is the model perturbation or reflectivity, v_0 is the background velocity model and G is the Green's function. With Born approximation, the original nonlinear inverse problem can be recasted into a two-parameter Born inverse problem.

Born Inverse Problem

The so-called Born inverse problem is to fit reflection data $d(\mathbf{x_s}, \mathbf{x_r}, t)$ by proper choice of $v_0(\mathbf{x})$ and $\delta v(\mathbf{x})$. As explained for example in Symes (2008), data fit is impossible unless v_0 is kinematically correct to within a half-wavelength error in traveltime prediction. This is essentially due to the fact that we are trying to solve an overdetermined problem: the dimension of the data is larger than the dimension of the model (e.g. 3D model corresponds to 5D seismic data). This way, data fit is only possible in the special case of correct velocity mentioned above. On the other hand, if δv is extended to depend on extra parameter (subsurface offset, reflection angle ...), that is, replaced by $\delta \bar{v}$, then any model consistent data d can be fit with essentially any v_0 by proper choice of $\delta \bar{v}$.

With the extra parameter, it is possible to preserve all the data-domain information in the extended model domain, thereby providing the ability to formulate the Born inverse problem in extended image domain. As a matter of fact, the inconsistency or consistency between velocity model and seismic data are implicitly expressed in the extended image volume. When the velocity model is kinematically correct, the extended reflectivity model $\delta\bar{v}$ will display certain character (flat, focused...). This is often referred to as the semblance condition. In contrast, the energy in the extended image volume will deviate from the semblance condition with an incompatible velocity model. The Born inverse problem is thus essentially a velocity analysis problem : find a velocity model v_0 so that the corresponding $\delta\bar{v}$ can satisfy the semblance condition. It is often solved with a two-step process:

- Construct the extended image volume via various imaging algorithms

- Update the velocity model by reducing the deviation of the image volume from semblance condition

In this paper, we focus on subsurface offset domain and investigate a certain type of optimization - differential semblance optimization (DSO). Specifically, the reflectivity model $\delta v(\mathbf{x})$ is replaced by $\delta\bar{v}(\mathbf{x}, h)$ and used as input in an extended Born modeling operator $\bar{F}$,

$$(\bar{F}[v_0]\delta\bar{v})(\mathbf{x_s}, \mathbf{x_r}, t) = \frac{\partial^2}{\partial t^2} \int d\mathbf{x} dh d\tau G(\mathbf{x_s}, \mathbf{x} - h, \tau)$$
$$\times \frac{2\delta\bar{v}(\mathbf{x}, h)}{v_0(\mathbf{x})^3} G(\mathbf{x} + h, \mathbf{x_r}, t - \tau). \tag{7}$$

Here, h is essentially half the distance between sunken source and sunken receiver in Claerbout (1985)'s survey sinking imaging condition. The physical meaning of this extension is that we now allow reflection at a distance. This is clearly non-physical extension, which matches the model dimension with data dimension. Data fit is thus achievable by violating the physical rule of wave propagation. Correspondingly, we produce extended image volume $I(\mathbf{x}, h)$ depending not only on the spatial coordinate but also the subsurface offset. As noted above, the subsurface offset gather implies the velocity information : focusness of the gather corresponds to correctness of the velocity.

A simple synthetic example is shown in Figure 1. A 2.5 km/s constant velocity model and a single reflector at 1.5 km depth are used to generate Born data with

a bandpass wavelet. The corresponding imaging gathers with different velocity models will exhibit different behaviors. The energy will focus at zero offset with correct velocity as shown in Figure 1b and spread out when the velocity is incorrect as shown in Figure 1a and 1c.
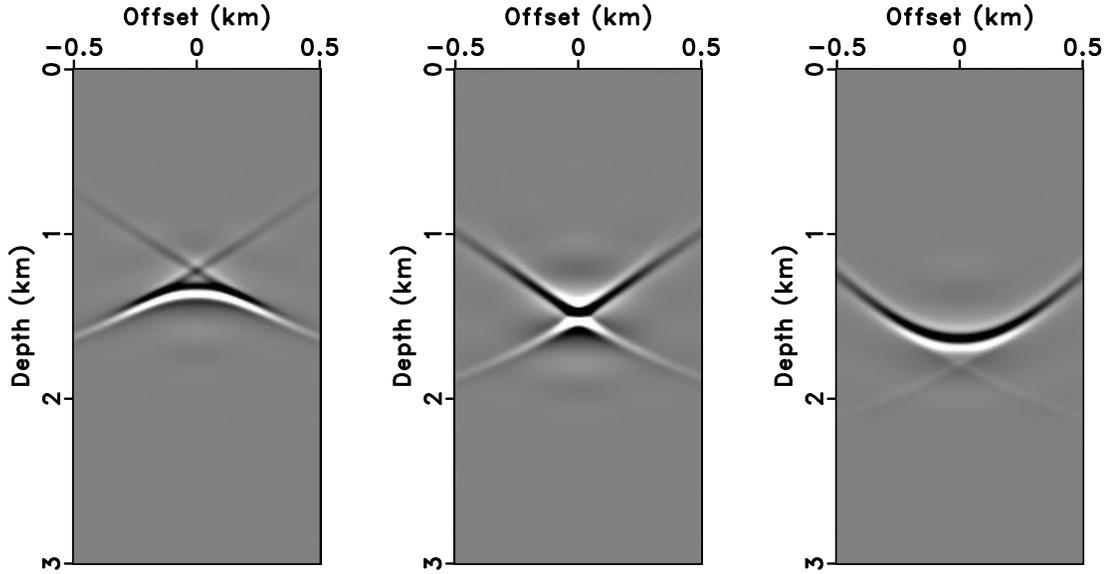

Figure 1: Offset gather with (a) 10% slower velocity (2.25 km/s); (b) correct velocity (2.5 km/s); (c) 10% faster velocity (2.75 km/s).

Imaging Operators

In the past, a variety of imaging operators have been investigated to construct the extended image volume used in the process of velocity analysis. Among them, RTM is known to be the superior imaging operator. Essentially, it computes the adjoint of the extended Born modeling operator :

$$I(\mathbf{x}, h) = \frac{2}{v_0(\mathbf{x})^3} \int d\mathbf{x_s} d\mathbf{x_r} dt d\tau G(\mathbf{x_s}, \mathbf{x} - h, \tau)$$
$$\times G(\mathbf{x} + h, \mathbf{x_r}, t - \tau) \frac{\partial^2}{\partial t^2} d(\mathbf{x_s}, \mathbf{x_r}, t),$$

(8)

where d is seismic reflection data.

Hou and Symes (2014, 2015) modify the adjoint operator into an approximate inverse to the extended Born modeling operator, by applying model and data domain weight operators. It has the form,

$$\bar{F}^{\dagger} = W_{\text{model}}^{-1} \bar{F}^{T} W_{\text{data}}. \tag{9}$$

As a matter of fact, more than one formula (Hou and Symes, 2017) have been proposed for the weight operators:

$$W_{\text{model}}^{-1} = 4v_0^5 LP, \ W_{\text{data}} = I_t^4 D_{z_s} D_{z_r}. \tag{10}$$

$$W_{\text{model}}^{-1} = -8v^4 D_z Q, \ W_{\text{data}} = I_t^3 D_{z_s} D_{z_r}. \tag{11}$$

Here $L = \sqrt{\nabla_{(x,z)}^2 \nabla_{(h,z)}^2}$, I_t is time integration, $\bar{F}^T$ is the adjoint of extend Born modeling operator and D_{z_s}, D_{z_r} are the source and receiver depth derivatives, D_z is depth derivative. Both P and Q are a Fourier-like operator and approximately equal to 1 near $h = 0$, and will be neglected throughout this paper.

These weight operators, although in different form, can achieve similar accuracy. In comparison, Formula 11 is less computationally demanding to implement than formula 10, as no spatial transforms are involved in the former implementation. As a result, formula 11 will be used to implement IVA throughout this paper.

DSO in subsurface offset domain

In the subsurface offset domain, the semblance condition is presented as focused subsurface offset gather. As a result, the simplest choice for the DSO type objective function is

$$J[v] = \frac{1}{2} \|hI(\mathbf{x}, h)\|^2, \tag{12}$$

in which $I(\mathbf{x}, h)$ is produced via the application of either adjoint Born modeling (equation 8), or the approximate inverse $\bar{F}^{\dagger}$ (equation 9) to the data. Since the image volume depends on v, so does J. Minimization of the objective function above penalizes the energy outside zero offset panel and thus focuses the offset gather, which implies correct velocity. This is the basic idea of DSO based velocity analysis.

Behaviour of Objective Function

It has been widely believed, either implicitly or explicitly, that this type of method does not depend on the data fitting. However, numerical experiment suggests the opposite: the better we can fit the data, the closer imaging operator approximates to the real inverse operator, the better this type of this method works.

To examine the effect of imaging operator choice on the objective function, we first calculate and plot the objective function values along a line segment in velocity model space. The model we use combines a constant background velocity model (2.5 km/s) and single flat reflector at 2 km depth. The two different imaging operators mentioned above are applied on the tapered Born data with a range of velocities (from 2 km/s to 3 km/s) to compute the objective function (equation 12). A common variant of MVA objective function (combined with stacking power term) is also investigated (Chauris and Noble, 2001; Shen and Symes, 2008). The objective function values for different velocities are then plotted in Figure 2. All three objective functions are smooth in velocity and unimodal. However, the objective function of MVA shifts toward the lower velocity (2400 m/s). This means the minimization of the DSO objective function will lead to an erroneous velocity. Comparing to DSO term, stacking power term will definitely reaches global minimum at correct velocity. It thus can be regarded as an ad-hoc solution to fix the shift of global minimum. The combination of these two terms, however, is not intuitive. It is notoriously difficult to select the appropriate parameter to balance these two terms. That is to say, we seek to shift the global minimum to the correct velocity without introducing local minimum. In comparison, the objective function using the approximate inverse operator reaches the overall minimum at the correct velocity and better resolves the minimizer. This is because the objective function with an approximate inverse operator, rather than the adjoint operator, is inherently more correct.

The comparison above clearly indicates the so-called "gradient artifacts" problem is not the problem of gradient but the problem of objective function. The objective function based on common imaging algorithm is not behaving as we expect. The gradient is indeed the correct gradient of the wrong objective function. As the imaging operator approaches a true inverse operator, the corresponding DSO objective function should present better property.
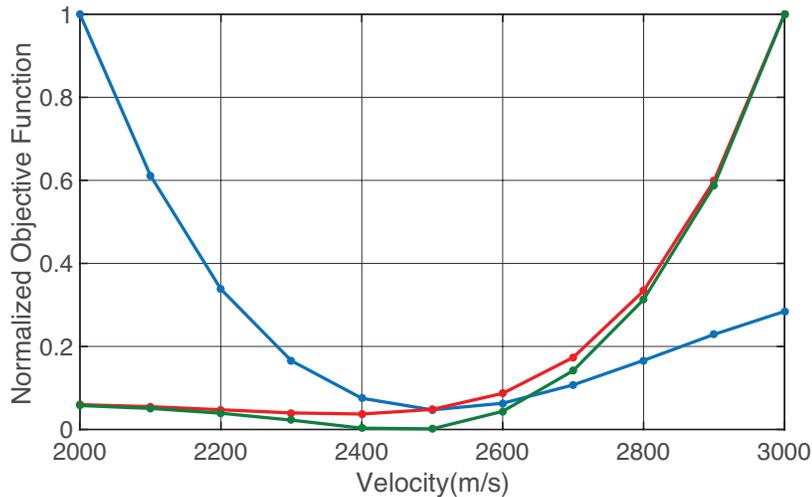
Figure 2: Normalized objective function : the red line represents MVA; the green line represents MVA + Stacking Power; the blue line represents IVA. The velocity ranges from 2 km/s to 3km /s.

NUMERICAL EXAMPLES

In this section, two numerical examples are presented. The first is a synthetic example with truncated Marmousi model (Versteeg and Grau, 1991). Both MVA (the adjoint operator) and IVA (approximate inverse operator) are used to perform the velocity analysis. The second is a field data example. We perform IVA only for this example.

Synthetic Example

The comparison is first performed on a truncated Marmousi model (Versteeg and Grau, 1991). The true model, shown in Figure 4a, is a smoothed version of the original model, in order to be wary of low frequency noise in the RTM image. As a matter of fact, the low frequency noise doesn't affect the approximate inverse operator thanks to the derivative operator in the model weight operators. The synthetic data is generated with 2-8 finite difference modeling. The acquisition geometry is a fixed spread of 151 sources and 301 receivers at 20 m depth. A (2.5-5-20-25)-Hz bandpass wavelet with 2 ms time sample is used to generate 3 s Born

data. Optimization is carried out with L-BFGS algorithm (Nocedal and Wright, 1999). We set the L-BFGS iteration to retain 15 prior search directions. We start the optimization from a linearly increased 1D model (Figure 4b).
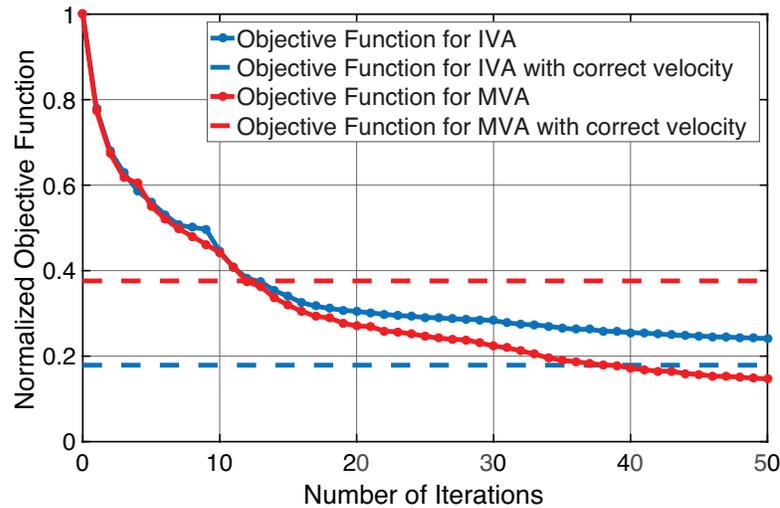

Figure 3: Normalized convergence curves for three different imaging operators. The solid line is the convergence curve and the dashed line is the objective function value at the correct velocity. The blue line represents MVA; the red line represents IVA.

25 iterations L-BFGS are used to generate the results. Figure 3 displays the normalized convergence curves for the optimization with different imaging operators. The convergence curve of MVA mangages to go beyond the objective function value at the correct velocity, indicating that the objective function doesn't reach the minimum at the correct velocity value.The approximate inverse version gives the better objective function behavior in the sense that the objective function at correct model is relatively smaller and more likely to be the global minimum. Figure 4 shows a comparison of the true, initial, updated models with two different imaging operators. As can be seen in the comparison, same optimization procedure with different imaging operators can produce quite different results. All four models are then used as background model to apply approximate inverse operator on the synthetic data (equation 9). Figure 5 compares the zero-offset images produced by the approximate inverse operator. The image associated with MVA velocity is clearly going wrong direction. The result using approximate inverse operator is more correct: the location of the reflectors and the amplitude are closer to the image with true velocity. Finally, six offset gathers are pulled out from the middle of the image volume to compare in Figure 6, demonstrating

approximate inverse operator produces a more focused image volume.
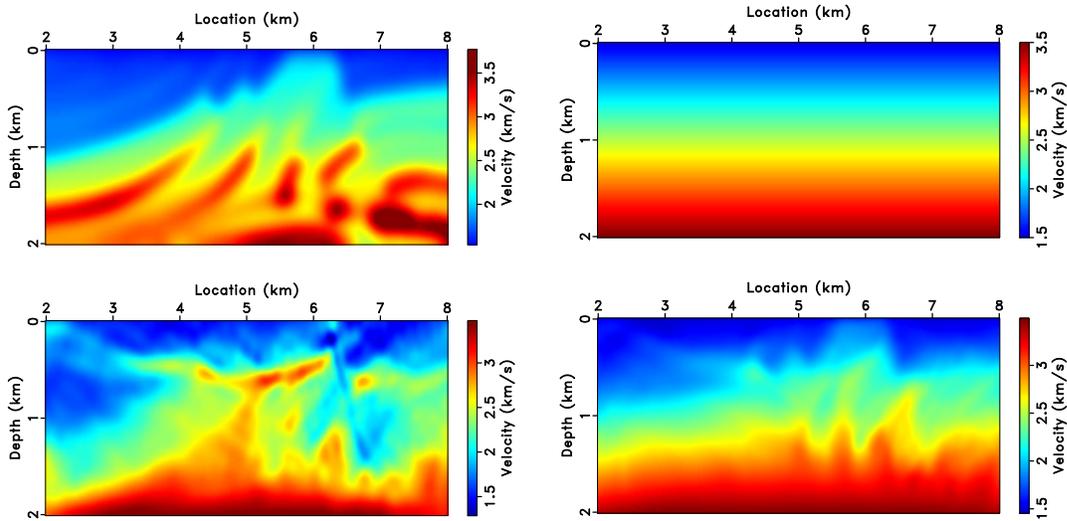
Figure 4: (a) True model; (b) Initial Model; 25 iteration L-BFGS result for (c) the adjoint operator; (d) the approximate inverse operator.

Field Data Example

We further test our method on a field example. The data, known as Mobil AVO Viking Graben Line 12, was initially released for a workshop on seismic inversion methods held on the 64th SEG Internation Exposition and Annual meeting. The data was first placed in publish domain in 1994 and then further published on the internet in 2012.

The detailed information about the dataset can be found in Keys and Foster (1998). The original dataset contain 1011 shots with a 25-m shot interval (11 shots are missing). Each shot is recorded by 120 receivers with 25 m spacing and 262 m minimum offset. The data was recorded for 6 s at a sampling rate of 4 ms. Here, we truncate the data at 3 s and take 330 shots from 10 km to 15 km as our study area. The study area also contains one well log.

The major noise present in this data set is multiple energy. The data used in this paper has been applied a parabolic Radon transform demultiple process by the distributor. However, there is still substantial multiple energy remaining.
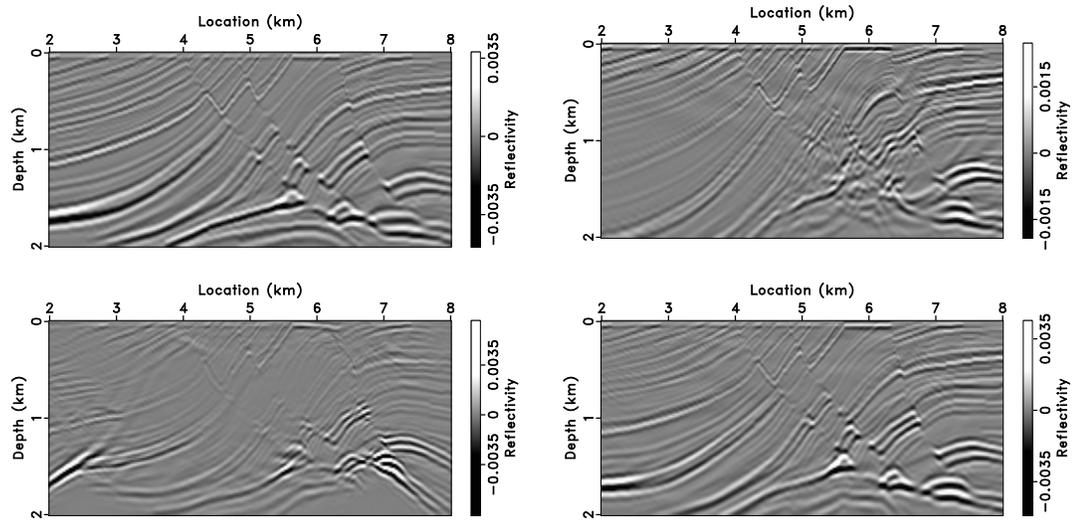
Figure 5: Zero-offset approximate inverse image associated with different velocity models shown in Figure 4.
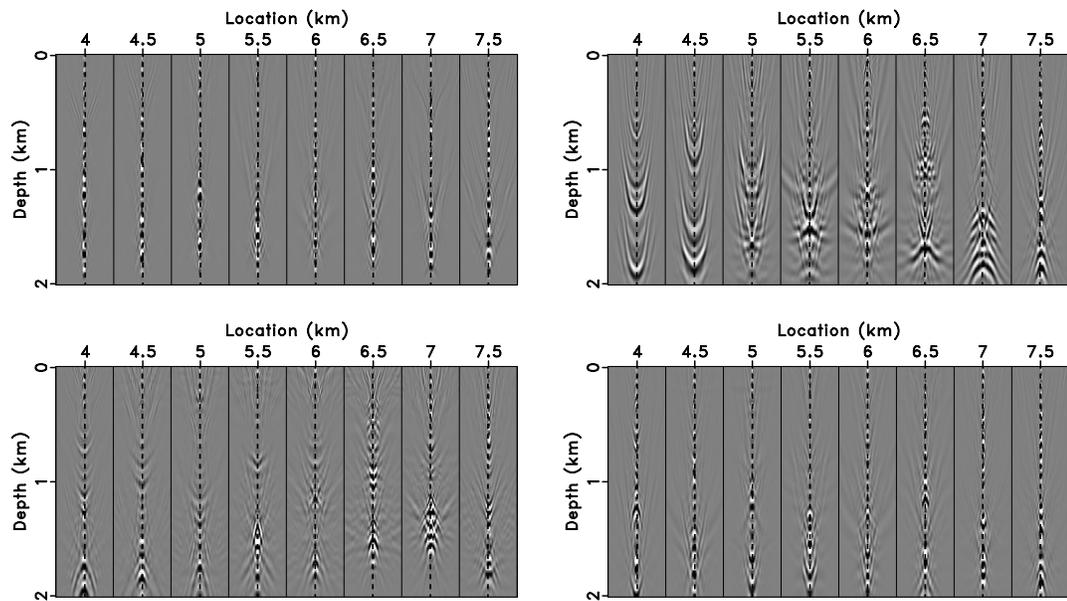

Figure 6: Image gathers in the subsurface offset domain corresponding to the images in Figure 5.

Following Mulder and ten Kroode (2002) and Li and Symes (2007), we apply a dip filter on the NMO-corrected gathers to reject the event with downward moveout. The key assumption here is that multiple reflections exhibit different moveouts from primaries. Three CMP gathers before and after the dip filtering are shown in Figure 7. Other preprocessing includes muting and a 5-10-30-40 bandpass filtering.

We manually pick some CMP gathers to construct the 1D velocity estimates independently, which are subsequently interpolated and smoothed. We then convert this 2D velocity model into interval velocity in depth domain, as shown in Figure 8a. This process is done with the help of open source software Seismic Unix (Cohen and Stockwell, 2015). This model serves as the initial model for subsequent IVA. The updated velocity, after 25 iterations IVA, is shown in Figure 8b. The velocity update is not dramatic because NMO has already done a good job focusing the gathers.

The zero-offset image, produced by the approximate inverse operator, is shown in Figure 9a with well log synthetic overlaid. The generation of the well log synthetic will be explained shortly. The image is clearly in good agreement with well log synthetic. Figure 9b verifies the subsurface offset gather also concentrates at zero offset.

In order to better validate the result, we compare it with the well log data. The well is located around CMP 808 (11700 m). We first use sonic and density log to calculate impedance and reflectivity. The reflectivity is then converted to time domain based on the Vp log data. The image at the well location is also converted into time domain using the IVA updated velocity model. A constant-phase wavelet, with a constant time delay, is generated to match the amplitude spectrum of the seismic image. The wavelet extraction uses CREWES Matlab toolbox (Margrave, 2001). The time delay here is essential because there may be misalignment between the well log and seismic image due to various reason. No squeeze and stretch of the well log are applied in this comparison. We then convolve the reflectivity with the zero-phase wavelet to generate the synthetic seismogram. The comparison in Figure 11 displays a good correlation between the synthetic seismogram and seismic image. The synthetic seismogram is also converted back into depth domain using the Vp log data and overlaid on the image (Figure 9a).
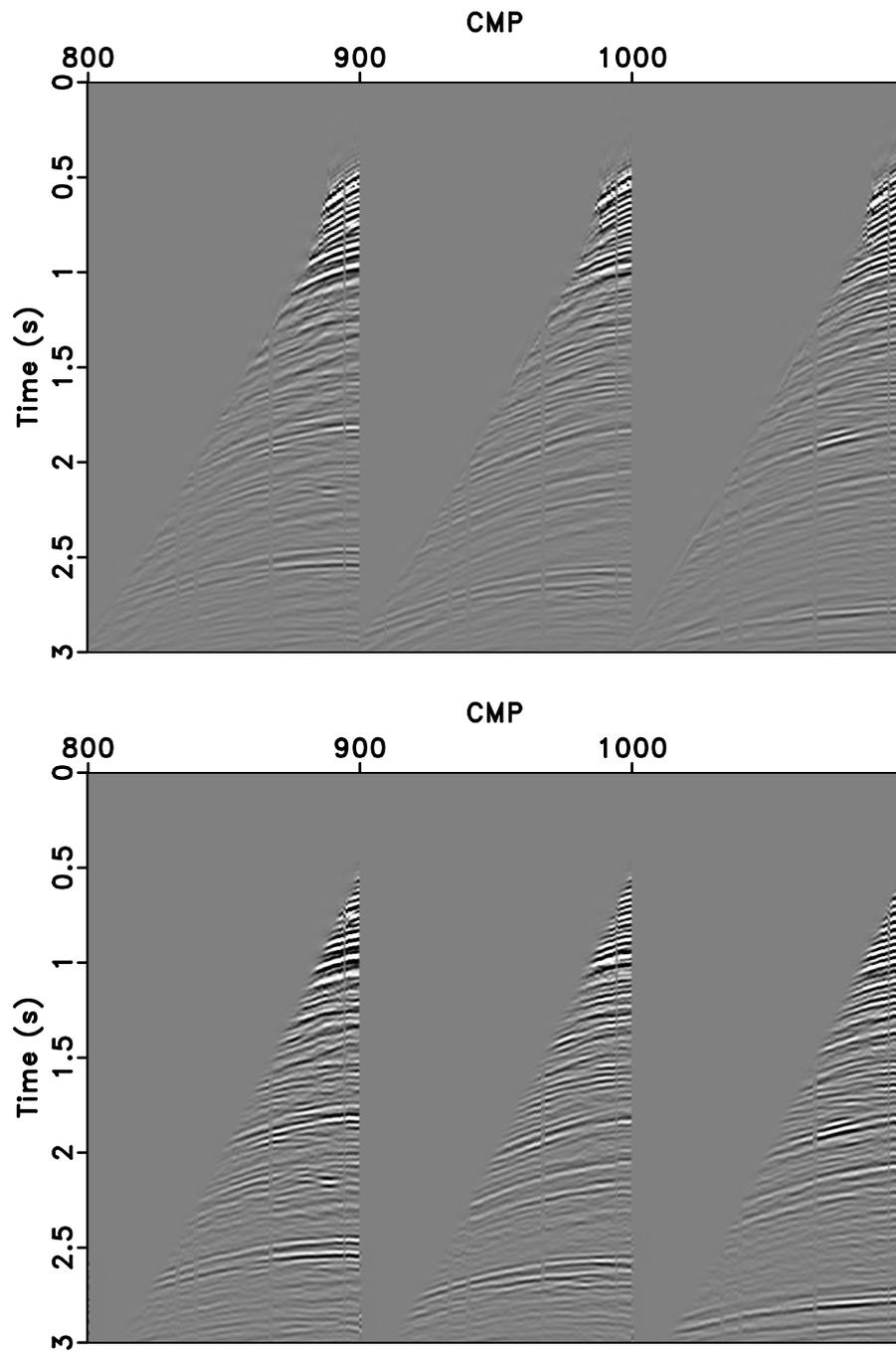
Figure 7: (a) Original CMP gathers from Mobil AVO data; (b) CMP gathers after dip filtering.
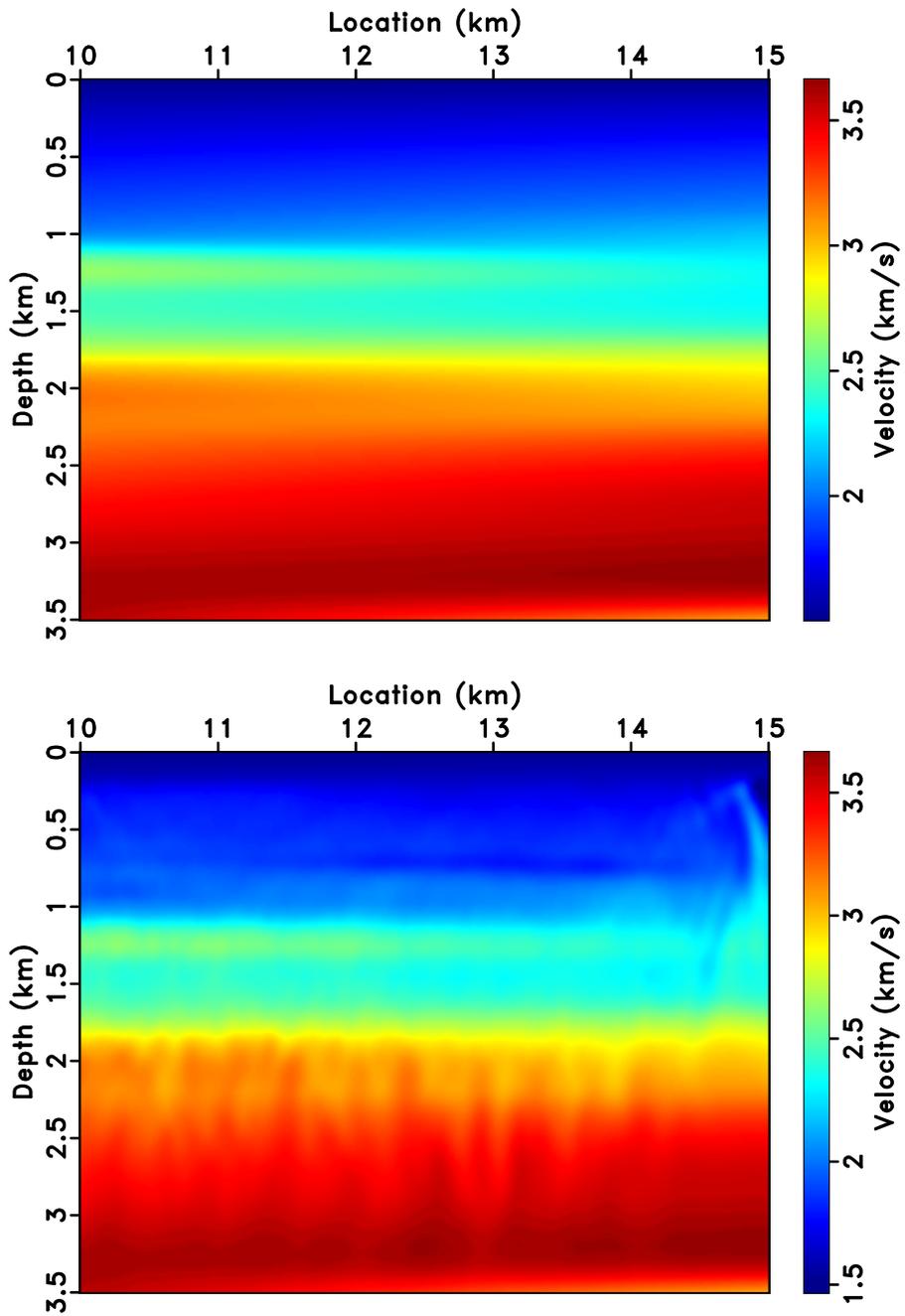
Figure 8: (a) NMO velocity; (b) IVA velocity after 25 iterations.
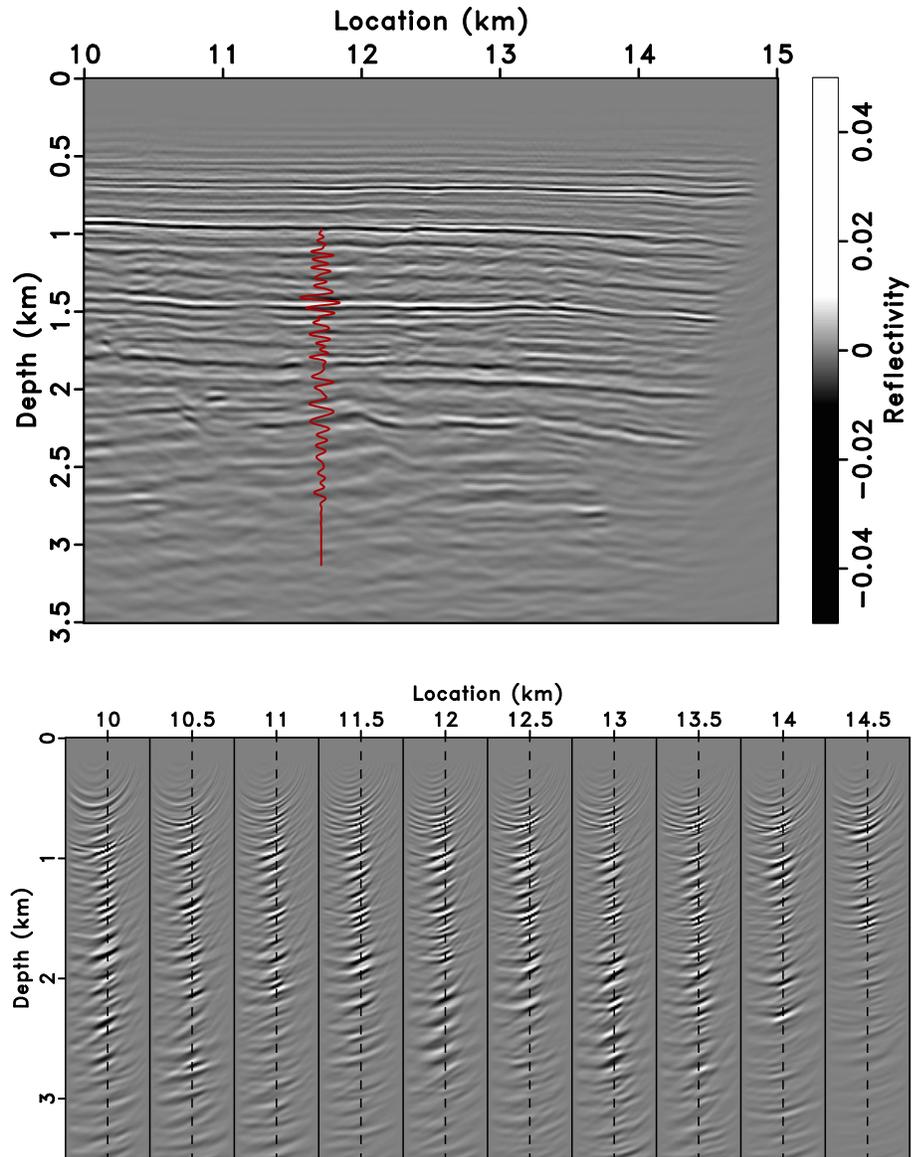
Figure 9: (a) Zero-offset image; (b) Subsurface offset gather produced by the approximate inverse operator.
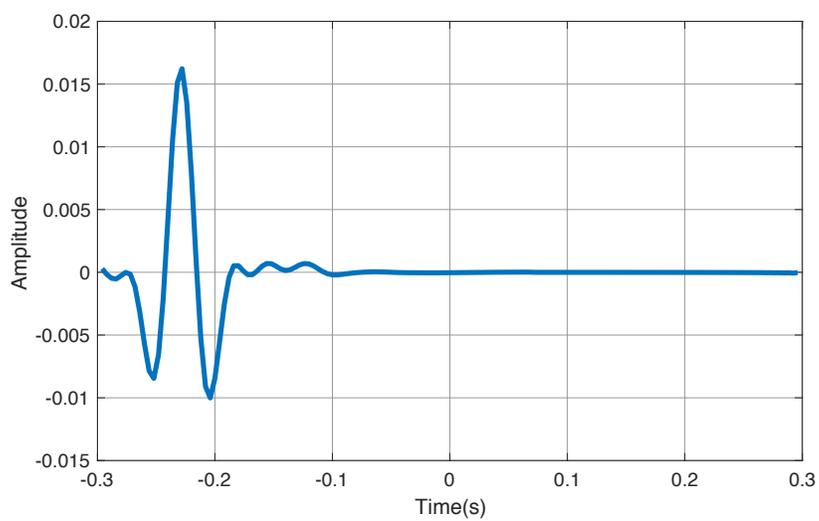
Figure 10: Generated wavelet using CREWS Matlab Toolbox.
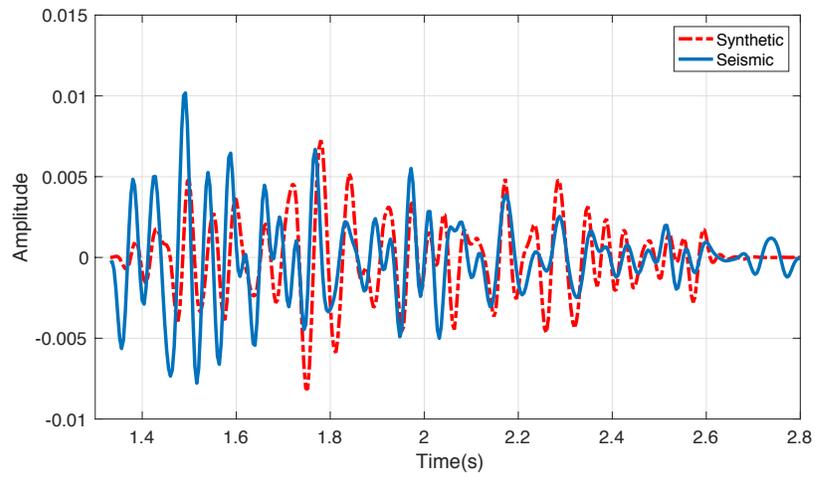
Figure 11: Comparison between seismic and synthetic.

CONCLUSIONS

We have compared velocity analysis via DSO with two different imaging operators and analyzed their corresponding performance. The numerical examples show that the closer that the imaging operator to inversion, the better the DSO velocity estimate. An approximate inverse as in Hou and Symes (2015, 2017) adds no additional cost but improves velocity estimation substantially.

ACKNOWLEDGMENTS

REFERENCES

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, Expanded Abstracts, Society of Exploration Geophysicists, SI9.5.

Biondi, B., and P. Sava, 2004, Wave-equation migration velocity analysis - I: Theory, and II: Subsalt imaging examples: Geophysics, **52**, 593–623.

Chauris, H., and M. Noble, 2001, Two-dimensional velocity macro model estimation from seismic reflection data by local differential semblance optimization: applications synthetic and real data sets: Geophysical Journal International, **144**, 14–26.

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

Cohen, J. K., and J. J. W. Stockwell, 2015, CWP/SU: Seismic Unix release no. 44: a free package for seismic research and processing: Center for Wave Phenomena, Colorado School of Mines.

Fei, W., and P. Williamson, 2010, On the gradient artifacts in migration veloc-

ity analysis based on differential semblance optimization: 80th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4071–4076.

Hou, J., and W. Symes, 2014, An approximate inverse to the extended Born modeling operator: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 3784–3789.

——, 2015, An approximate inverse to the extended Born modeling operator: Geophysics, **80**, no. 6, R331–R349.

——, 2017, An alternative formula for approximate extended born inversion: Geophysics, **82**, no. 1, S1–S8.

Keys, R., and D. Foster, 1998, Comparison of seismic inversion methods on a single real dataset: Society of Exploration Geophysicists.

Khoury, A., W. Symes, P. Williamson, and P. Shen, 2006, DSR migration velocity analysis by differential semblance optimization: 76th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SPMI3.4.

Lafond, C. F., and A. R. Levander, 1993, Migration moveout analysis and depth focusing: Geophysics, **58**, 91–100.

Lailly, P., 1983, The seismic inverse problem as a sequence of before-stack migrations, *in* Conference on Inverse Scattering: Theory and Applications: Society for Industrial and Applied Mathematics, 206–220.

Lameloise, C.-A., and H. Chauris, 2016, Extension of migration velocity analysis to transmitted wavefields: Geophysical Journal International, **207**, 343–356.

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.

Li, J., and W. Symes, 2007, Interval velocity estimation via nmo-based differential semblance: Geophysics, **72**, U75–U88.

Liu, Y., W. Symes, and Z. Li, 2014, Inversion velocity analysis via differential semblance optimization: Presented at the 76th Annual Meeting, Extended Abstracts, European Association of Geoscientists and Engineers.

Liu, Z., and N. Bleistein, 1995, Migration velocity analysis: theory and an interative algorithm: Geophysics, **60**, 142–153.

Margrave, G., 2001, Numerical methods of exploration seismology with algorithms in matlab. (https://www.crewes.org/ResearchLinks/FreeSoftware/).

Mulder, W., and F. ten Kroode, 2002, Automatic velocity analysis by differential semblance optimization: Geophysics, **67**, 1184–1191.

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

Shen, P., 2012, An RTM based automatic migration velocity analysis in image

domain: 82nd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SVE2.2.

Shen, P., and W. Symes, 2008, Automatic velocity analysis via shot profile migration: Geophysics, **73**, VE49–60.

Shen, P., W. Symes, and C. C. Stolk, 2003, Differential semblance velocity analysis by wave-equation migration: 73rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2135–2139.

Stork, C., 1992, Reflection tomography in the postmigrated domain: Geophysics, **57**, 680–692.

Symes, W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790.

——, 2014, Seismic inverse problems: recent developments in theory and practice: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 2–5.

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

ten Kroode, F., 2012, A wave-equation-based Kirchhoff operator: Inverse Problems, 115013:1–28.

——, 2014, A Lie group associated to seismic velocity estimation: Inverse Problems - from Theory to Application, Proceedings, Institute of Physics, 142–146.

Versteeg, R., and G. Grau, eds., 1991, The Marmousi experience: Proceedings of the EAEG workshop on practical aspects of inversion: IFP/Technip.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, WCC127–WCC152.

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

Weibull, W., and B. Arntsen, 2013, Automatic velocity analysis with reverse time migration: Geophysics, **78**, S179–S192.

Yilmaz, O., 2001, Seismic data processing, *in* Investigations in Geophysics No. 10: Society of Exploration Geophysicists.

Yilmaz, O., and R. Chambers, 1984, Migration velocity analysis by wavefield extrapolation: Geophysics, **49**, 1664–1674.