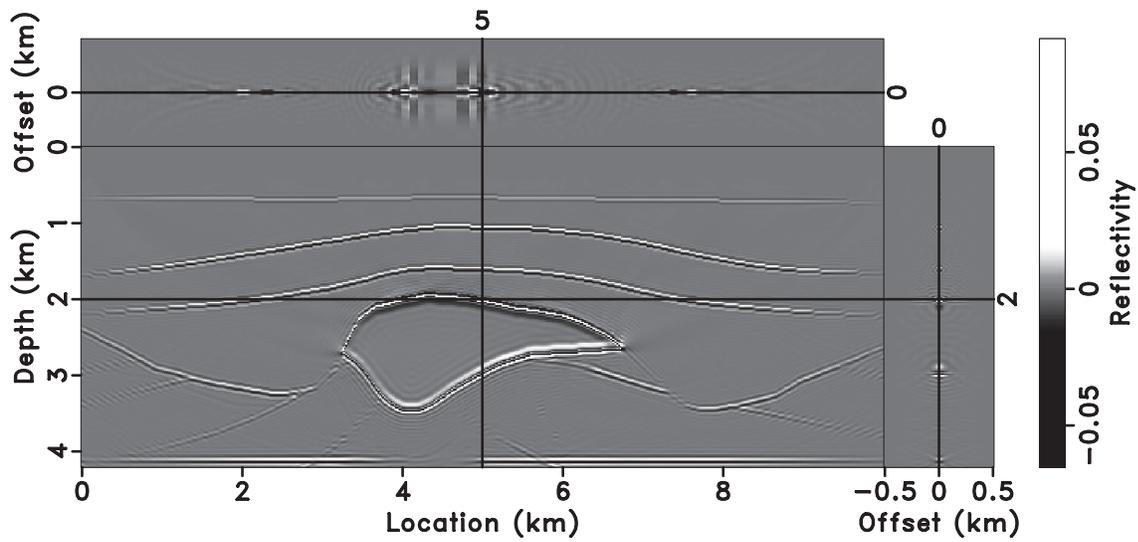# THE RICE INVERSION PROJECT

*Mario Bencomo, Raanan Dafni, Lei Fu, Jie Hou, Guanghui Huang, Yin Huang,
Rami Nammour, William Symes, and Muhong Zhou*

Annual Report 2015

TRIP15 — TABLE OF CONTENTS

ii

INTRODUCTION TO THE 2015 ANNUAL REPORT

Welcome to the 2015 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 finished Annual Report web page will also include the program of the Annual Review Meeting and links to slides sets presented there, as in prior years.

At this writing, two of these papers ("Accelerating Extended Least Squares Migration with Weighted Conjugate Gradient Iteration", Hou and Symes, and "Scattering and dip angle decomposition in relation with subsurface offset extended wave equation migration", Dafni and Symes) have been returned to *Geophysics* after minor revision, and "Flexibly Preconditioned Extended Least Squares Migration in Shot Record Domain", Huang and Symes, has been submitted to *Geophysics* and is in review. Expanded abstracts for various meetings have been submitted, accepted, or presented, as noted in the text.

TRIP members received several awards during 2015. Jie Hou received the SEG's 2014 Best Student Paper award for his presentation "An Approximate Inverse to the Extended Born Modeling Operator". A paper of the same title is part of the 2014 Report, and appeared in the November-December 2015 issue of *Geophysics*. Jie's 2015 SEG presentation "Accelerating Extended Least Squares Migration with Weighted Conjugate Gradient Iteration" has been designated as one of the 31 best papers presented at that meeting. Finally, I received the Desiderius Erasmus award from the EAGE at its 2015 meeting in Madrid.

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, 15 January 2016

Accelerating Extended Least Squares Migration with Weighted Conjugate Gradient Iteration

Jie Hou and William W. Symes

ABSTRACT

Least Squares Migration (LSM) iteratively achieves a mean square best fit to seismic reflection data, provided that a kinematically accurate velocity model is supplied. The subsurface offset extension adds extra degrees of freedom to the model, thereby allowing LSM to fit the data even in the event of significant velocity error. This type of extension also implies additional expense per iteration from cross-correlating source and receiver wavefields over the subsurface offset, and therefore places a premium on rapid convergence. We accelerate the convergence of Extended Least Squares Migration, by combining the Conjugate Gradient algorithm with weighted norms in range (data) and domain (model) spaces that render the extended Born modeling operator approximately unitary. Numerical examples demonstrate that the proposed algorithm dramatically reduces the number of iterations required to achieve a given level of fit or gradient reduction, compared to Conjugate Gradient iteration with Euclidean (unweighted) norms.

INTRODUCTION

Least Squares Migration (LSM) iteratively seeks a short-scale reflectivity model so as to achieve a best fit to the seismic reflection data in least squares sense, via repeated migrations and demigrations (Nemeth et al., 1999; Kuehl and Sacchi, 2003). LSM is in fact least squares inversion based on Born (linearized) modeling (Bourgeois et al., 1989): since it is an inversion, the amplitudes are likely to be physically reasonable, and many authors have noted the effective spatial deconvolution and increase in apparent resolution resulting from data fitting (Dutta et al., 2014).

In the cited references and many others, the LSM optimization problem is solved iteratively. Since each iteration involves both (prestack) migration and demigra-

tion, the expense of iterative LSM can be considerable. This paper presents a technique for accelerating the convergence of *extended* LSM (ELSM), which constructs an extended or non-physical model depending on extra parameters besides spatial coordinates. When the velocity and data are compatible, ELSM outputs a volume focused in physical space, from which one may extract a physical inverted model equivalent to LSM's. Unlike LSM, however, ELSM may be approximated by a modified extended reverse time migration (RTM) operator, involving only wave equation (as opposed to ray-theoretic) computations (Hou and Symes, 2014, 2015). Also unlike LSM, this modified extended RTM accurately inverts the extended Born modeling operator even when the velocity is not kinematically correct. In this paper, we use the components of the modified extended RTM to modify a conjugate gradient algorithm for ELSM, and demonstrate dramatic convergence speedup of the modified algorithm.

To explain how convergence acceleration is accomplished, begin with the observation that LSM data fitting depends critically on the accuracy of the background velocity model. Velocity error leads to mispositioned and defocused structures in the image domain, and related data misfit. That is as one would expect: the model depends on fewer parameters than the data, so only in special case (correct velocity) can the data be fit well throughout.

One might guess that data-fitting might be decoupled from choice of velocity model, at least to some extent, by extending the model with extra dimensions, to equalize the model and data dimension. Claerbout's survey-sinking migration concept (Claerbout, 1985) inspires one possibility: Stolk and De Hoop (2005), Symes (2008b), Stolk et al. (2009) describe how to add an internal *subsurface* offset variable to the model, so that incident wavefield at one point can interact with (cause) a reflected wavefield at a nonzero offset. Extended least squares migration (ELSM) with this subsurface offset extension is the algorithm explored in this paper. It is able to fit data equally well with correct or incorrect background velocity model, within broad limits (Liu et al., 2013).

On the other hand, LSM (extended or not) via conjugate gradient or another relatively fast iterative algorithm generally requires tens of iterations for an acceptable result, each costing as much as two migrations. The additional parameters in ELSM add extra cost: for subsurface offset extension, computational loops over the offset axes are implicit in the definition of extended modeling or migration. This heavy cost could be alleviated in two ways : either lower the cost for each iteration or accelerate the convergence rate.

Several acceleration techniques of the first kind have been suggested. For example, one might reduce the cost of modeling/migration operator through simultaneous source (Beasley et al., 1998), blended data (Berkhout, 2008) or phase encoding method (Romero et al., 2000; Ikelle, 2007), together with various technologies to suppress the crosstalk artifacts (Krebs et al., 2009; Schuster et al., 2011; Xue et al., 2014). Another approach to cost reduction couples downsampling on all space-time axes (including subsurface offset) with low-pass filtering and velocity updating. As the velocity improves, the necessary range of offsets required to fit data is also reduced, so that upsampling and increased frequency content can be accommodated with a constant (and relatively small) number of gridpoints in offset (Fu and Symes, 2015).

In this paper, we confine ourselves to the second kind of acceleration, often accomplished by conditioning the linear system to be solved, that is, creating a related system by linear change of variable, for which convergence is faster. For example, an approximate inverse least squares Hessian can be computed by inverting the diagonal matrix elements, perhaps complemented by a limited number of off-diagonals, and used to compensate for poor illumination and thus accelerate convergence (Chavent and Plessix, 1999; Pratt, 1999; Shin et al., 2001; Rickett, 2003; Symes, 2008a; Tang, 2009; Aoki and Schuster, 2009).

Recently ten Kroode (2012) explained the construction of a computable approximate inverse to the subsurface offset extended Kirchhoff modeling operator. Hou and Symes (2014, 2015) modified ten Kroode's construction for the Born operator, and observed that the geometric optics based derivation of the computable approximate inverse owes its feasibility to a remarkable identity, due to Zhang et al. (2003), linking the so-called Beylkin determinants of Kirchhoff migration-inversion and the geometric amplitudes. For the subsurface offset extension (and only for this and similar extended modeling operators) this identity implies that a simple modification of the least-squares Hessian is to leading order computable without resort to ray-theoretic constructs, thus purely via numerical solution of wave equations and a few explicit filtering operations. Inverting these filtering operations, one obtains an approximate inverse, asymptotically correct in the same sense as Generalized Radon Transform inversion (Beylkin, 1985; De Hoop and Bleistein, 1997; Operto et al., 2000) but without ray-tracing.

Hou and Symes (2015) also demonstrate that the approximate inverse $\bar{F}^{\dagger}$ of the subsurface offset extended Born modeling operator $\bar{F}$ takes the form

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

in which $\bar{F}^T$ is the transpose of $\bar{F}$ (extended RTM), W_{model} and W_{data} are positive-definite symmetric operators on model and data spaces respectively, explicitly computable directly from knowledge of the velocity field. Since $\bar{F}^\dagger \bar{F} \approx I$, it follows that $\bar{F}$ is approximately unitary with respect to the norms defined by W_{model} and W_{data}. Therefore a conjugate gradient or similar algorithm (for example, LSQR) formulated with weighted norms in data and model spaces, with weight operators W_{data} and W_{model} respectively, will converge much more rapidly than the same algorithm formulated with the Euclidean norm.

In this paper, we explain exactly how to compute W_{model} and W_{data}, and numerically verify the dramatic improvement in convergence rate resulting from using these operators to define the norms appearing in the conjugate gradient algorithm applied to ELSM.

The work we report here differs from that cited above, amongst others, in several ways. As noted above, Hessian approximations of various sorts, mostly based on extraction of diagonal or near-diagonal elements, have long been used as convergence accelerators in various forms of full waveform inversion. These approximations, however, are of uncertain accuracy, whereas approximate inversion as we use it here is an asymptotic inverse, increasingly precise as frequency increases, Asymptotic inversion has also been used to accelerate iterative waveform inversion - for a recent example, see **?**. However so far as we know all prior work of this type has relied explicitly on ray-trace constructions (Generalized Radon Transform inversion), as opposed the purely "wave equation" construction explained here. Wave equation based asymptotic inversion is often termed "true amplitude migration", and has been even been developed for various types of extended models, for instance by Xu et al. (2011); Tang et al. (2013), not however used to accelerate iterative extended inversion. Finally, the remarkable form of the approximate inverse (5), its relation to the mapping properties of Born modeling, and the implications for iterative inversion do not seem to have been exploited before.

The rest of the paper is organized as follows. We first review the theory of ELSM and the approximate inverse operator. We then explain how to compute the weight operators, and how to write conjugate gradient iteration to accommodate them. We call the reformulated conjugate gradient (CG) algorithm with weighted norms the weighted conjugate gradient (WCG) algorithm. We present three synthetic examples, illustrating the considerably faster convergence of WCG over (unweighted) CG. In these examples, the number of WCG iterations necessary to reduce the least squares gradient to any reasonable proportion of its original

size (5%, 1%,...) is an order of magnitude or more smaller than the corresponding number for CG. We end with a discussion of consequences of our work and open questions.

THEORY

The wave physics underlying the theory to be explained in this paper is the Born (linear, single scattering) approximation to constant density acoustics. Write $v = v_0 + \delta v$ for the acoustic wave velocity v resulting from a perturbation δv of a reference velocity v_0. To first (linear) order in δv, the corresponding change in the pressure field p is $p \approx p_0 + \delta p$, in which p_0 corresponds to v_0. The reference and perturbation pressure fields, p_0 and δp, satisfy a coupled system of wave equations

$$\frac{\partial^2 p_0}{\partial t^2} - v_0^2 \nabla^2 p_0 \;\; = \;\; f \tag{2}$$

$$\frac{\partial^2 \delta p}{\partial t^2} - v_0^2 \nabla^2 \delta p \;\; = \;\; 2 v_0 \delta v \nabla^2 p_0 \tag{3}$$

in which f is a representation of the acoustic energy source, localized near source position $\mathbf{x}_s$. The approximation $p \approx p_0 + \delta p$ is most accurate when v_0 is smooth (transparent) on the wavelength scale, and all model oscillations (reflectivity) are confined to δv (Symes, 2009).

The Born modeling operator $F[v_0]$ is parametrized by the reference velocity v_0 and relates the perturbations δv and δp:

$$F[v_0] \delta v = \delta p \tag{4}$$

Implicit in this definition is sampling at receiver positions $\mathbf{x}_r$, possibly dependent on the source position $\mathbf{x}_s$, also the sampling time interval.

With this notation, the constant density acoustic linearized, or Born, inverse problem is: given a reference velocity $v_0(\mathbf{x})$, source field $f(\mathbf{x}, t; \mathbf{x}_s)$, and data perturbation field $\delta d(\mathbf{x}_r, t; \mathbf{x}_s)$, find a velocity perturbation δv so that

$$F[v_0] \delta v = \delta d. \tag{5}$$

The migration operator approximately solves this linearized inverse problem by computing the adjoint of the Born modeling operator. However, the migrated images typically suffers from quality degradation, such as amplitude imbalance and

wavelet sidelobes (Bednar and Bednar, 2006; Mulder and Plessix, 2004). Least squares migration (LSM) (Nemeth et al., 1999; Kuehl and Sacchi, 2003) can significantly reduce all of these defects. LSM is actually synonymous with solution of the least squares problem

$$J_{LS} = \frac{1}{2} \|F[v_0]\delta v - \delta d\|^2,$$ (6)

and equivalent to solving the normal equation:

$$F[v_0]^\dagger F[v_0]\delta v = F[v_0]^\dagger \delta d.$$ (7)

where $F^\dagger$ is the adjoint relative to the choice of norms in data and model spaces. LSM will generate a model fitting the data as well as possible, but that may not be very well for even relatively small velocity errors.

Figure 1 shows LSM images of a flat reflector in correct and incorrect constant background velocity respectively. The wrong velocity image is mispositioned and has incorrect amplitude. The velocity error deprives LSM of the inversion property, that is, the image does not fit the data well (misfit plot is shown in Figure 4a). The reason for this misfit is overdetermination: the dimension of the model is less than the dimension of the data, so that the data can only be fit well in the exceptional case that the background velocity v_0 is kinematically correct.
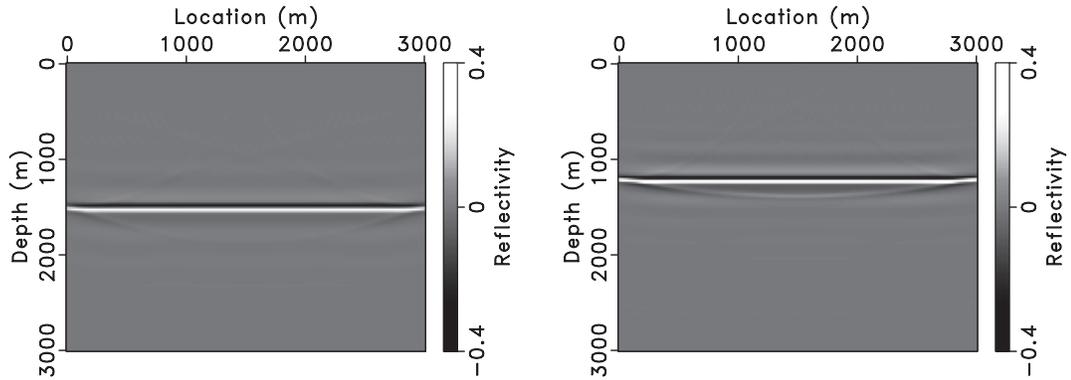

Figure 1: Least Squares Migration result after 30 CG iterations using (a) correct background velocity model, 2.5km/s (b) incorrect background velocity model, 2km/s.

Extended Least Squares Migration

In order to equalize the model dimension with data dimension, it is natural to introduce extra degrees of freedom by extending the physical model. Common choices for additional dimensions include reflection angle and surface offset. Another possible extension is the subsurface offset, which is essentially the (horizontal) offset h between sunken source and sunken receiver in Claerbout's survey-sinking imaging condition (Claerbout, 1985; Symes, 2008b; Stolk et al., 2009).
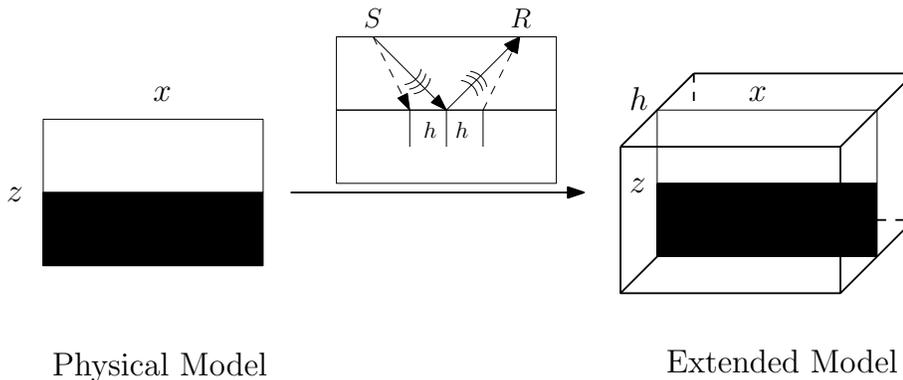

Physical Model   Extended Model

Figure 2: Sketch of the subsurface offset extension. The subsurface offset is half the distance between subsurface scattering points. This extension allows stress to produce strain at a distance.

Physical (non-extended) models δv give rise to subsurface offset extended models $\delta \bar{v}$ via multiplication by $\delta(h)$: in 2D, $\delta \bar{v}(x, z, h) = \delta v(x, z)\delta(h)$. That is, as extended models, physical models are *focused*. See Figure 2 for an illustration of the 2D variant, which we will discuss for the remainder of this paper. The operator extension $\bar{F}$ integrates $\delta \bar{v}$, over h, hence coincides with F when $\delta \bar{v}$ is physical. Note that the data space is the same in both cases. We call least squares migration with extended operators (solution of (4) or (5) with F replaced by $\bar{F}$) Extended Least Squares Migration (ELSM).

Figure 1 and Figure 3 show LSM and ELSM model estimates computed with Conjugate Gradient (CG) iteration (Nocedal and Wright, 1999), in a precise sense the best iterative method for this type of problem. With the extra dimension, all the data information can be preserved in the model space. Correct velocity model will force the energy focus at $h = 0$ section. Incorrect background velocity model will spread event energy to nonzero h. The LSM data residual is large for incor-

rect velocity, but the ELSM data residual remains small with correct or incorrect velocity.
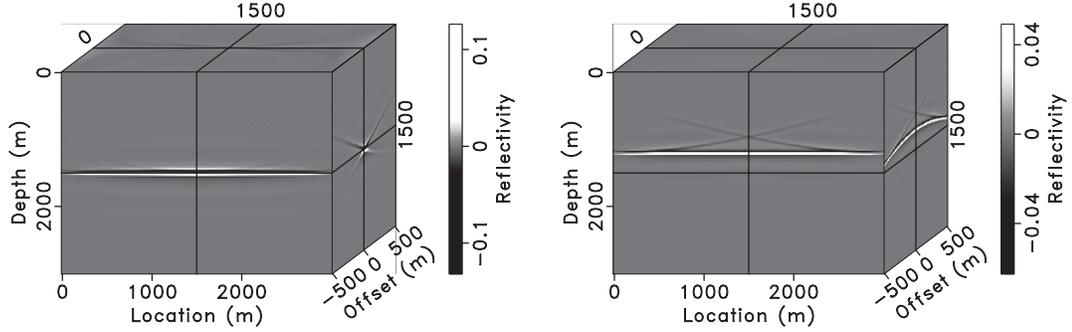

Figure 3: Extended Least Squares Migration result after 30 CG iterations using (a) correct background velocity model, 2.5km/s (b) incorrect background velocity model, 2km/s. The subsurface offset range is from -500m to 500m.

Figure 4 compares the relative misfit plot between LSM and ELSM as a function of conjugate gradient iteration, confirming ELSM's tolerance of velocity error.

The information in the offset plane can be used to update the background velocity model, if energy is not focused. However, the integration over h makes $\bar{F}$ more expensive than F, therefore fast convergence is very desirable. The convergence rate of CG depends on the operator spectrum: error components associated with clustered eigenvalues (for example, near 1) of the normal equation (5) are reduced by a large factor in a single iteration. Since the definition of the operator adjoint $\bar{F}^\dagger$ depends on the norms chosen for domain and range spaces, accelerating convergence can be accomplished by choosing these norms to move many eigenvalues of $\bar{F}^\dagger \bar{F}$ close to 1.

Unitary Property of Extended Modeling

ten Kroode (2012); Hou and Symes (2014, 2015) show how to compute an approximate inverse to the extended Born modeling operator, using only solutions of wave equations and local algebraic computations in physical and Fourier space, without the use of ray-tracing. This approximate inverse has exactly the form (5), with

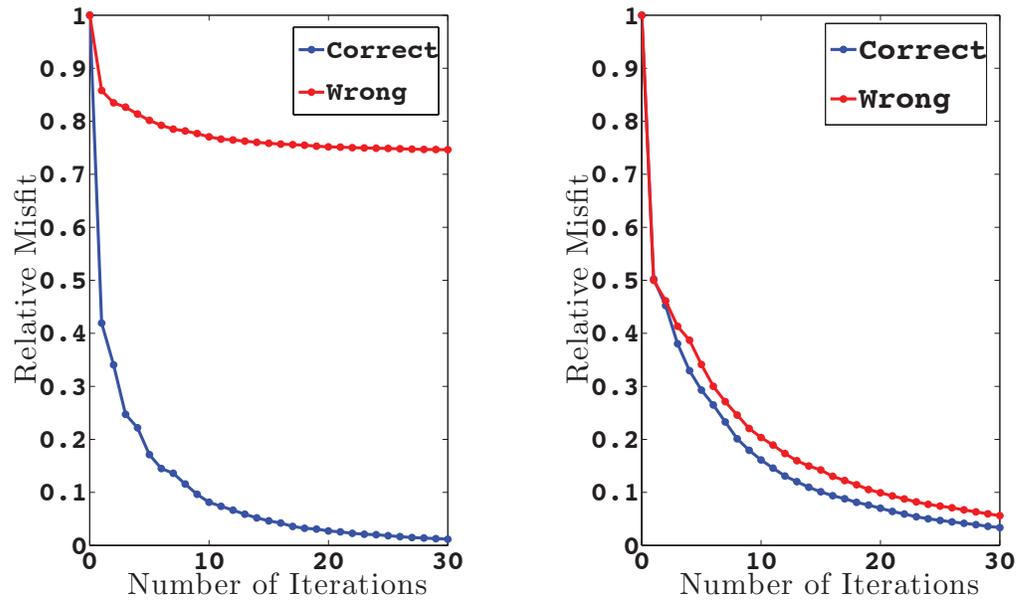$$W_{\text{model}}^{-1} = 4v_0^5 LP, \ W_{\text{data}} = I_t^4 D_{z_s} D_{z_r} \tag{8}$$

Figure 4: Relative data misfit plot for (a) LSM and (b) ELSM with correct and wrong velocity model. The blue line represents correct background velocity model and the red line shows the incorrect background velocity model. Notice ELSM will converge to zero no matter with correct or incorrect velocity model.

where $L = \sqrt{\nabla^2_{(x,z)} \nabla^2_{(h,z)}}$, I_t is time integration, $\bar{F}^T$ is the Euclidean adjoint of extend Born modeling operator (extended RTM) and D_{z_s}, D_{z_r} are the source and receiver depth derivatives. P is a Fourier-like operator whose amplitude is a known algebraic function of $v_0(x,z), v_0(x \pm h, z)$, and (k_x, k_z, k_h), given explicitly in (Hou and Symes, 2015), Appendix A.

The expression above for W_{model} is not symmetric, but can be symmetrized with negligible error. Both weight operators are positive definite.

The operator L is easily realized in Fourier space, using 2D discrete Fourier transforms. I_t is approximated by cumulative sum. Approximation of D_{z_s}, D_{z_r} is more subtle, if only pressure data is recorded, as in conventional streamer aquisition. For streamer simulation with shallow tow depth, Hou and Symes (2015) point out that free surface ghosts effectively supply the vertical derivatives, and therefore the data weight operator in the approximate inverse formula (5) requires only the use of the absorbing boundary adjoint (extended RTM). Conversely, for data without free surface ghosts, the free surface extended RTM operator should be used. This is the approach taken in this paper. Of course this approximation is good only up to the first ghost notch. ten Kroode (2012) points out that the vertical source and receiver derivatives may be expressed as square root operators in t, x_s, x_r, due to the field at the receivers being upcoming (for absorbing surface data). Besides supplying the grounds for treating $D_{z_s} D_{z_r}$ as positive-definite symmetric (see (Hou and Symes, 2015) for more on this), this observation suggests a computation of the data-side weight operator using one-way operators, which could be valid in settings other than shallow towed streamer acquisition, for example ocean-bottom recording. Finally, we note that the required data is actually recorded by multicomponent ocean bottom and land instruments, and more recently by specialized streamers.

It remains to address the operator P. $P = I$ in two special cases: for input data focused at (that is, only nonzero near) $h = 0$, or for any data if v_0 is independent of x. Numerical experiments, including those reported here, tend to indicate that $P \approx I$ for slowly varying v_0, so we have adopted the approximation $P = I$ throughout this paper. While the effect of this approximation on the accuracy of the approximate inverse appears to be small, we have not yet assessed it explicitly by implementing P and including it in the model-side weight operator: that remains a topic for further research.

See Hou and Symes (2015) for further discussion of the numerical implementatio of formula (5).

Weighted Conjugate Gradient Algorithm

As pointed out in the introduction, the form of the approximate inverse (5) implies that the extended Born modeling operator is unitary, provided that we define inner products in model and data space use the weight operators discussed in the last section:

$$\langle \delta \bar{v}_1, \delta \bar{v}_2 \rangle_{\text{model}}^2 = \sum_{x,z,h} \delta \bar{v}_1(x,z,h)(W_{\text{model}} \delta \bar{v}_2)(x,z,h), \tag{9}$$

and similarly for data space. Therefore, a Krylov subspace algorithm formulated using these inner products and the corresponding adjoint would be expected to converge much more rapidly than the same algorithm using Euclidean inner products and the corresponding adjoint (standard RTM).

Conjugate gradient iteration is a standard choice of Krylov space iteration for symmetric positive definite linear systems such as (5) (Björk, 1997). This and similar iterations take the same form regardless of choice of inner product, so long as the adjoint operator is defined relative to the inner product used. We call the following algorithm *Weighted CG* (WCG), remarking that it is really just the CG algorithm using the weighted inner products (9). We formulate it in the notation of the target problem, the extended version of the normal equation (5):

$$\bar{F}^{\dagger}[v_0]\bar{F}[v_0]\delta \bar{v} = \bar{F}^{\dagger}[v_0]\delta d. \tag{10}$$

Inputs to the algorithm are the operators $\bar{F} = \bar{F}[v_0], \bar{F}^{\dagger}[v_0]$, the data δd, and an initial estimate of extended velocity perturbation $\delta \bar{v}_0$ (which may be the zero field).

Algorithm 1 Weighted Conjugate Gradient Algorithm

1: $z_0 \leftarrow \bar{F}^\dagger(\delta d - \bar{F}\delta\bar{v}_0)$

2: $p_0 \leftarrow z_0$

3: $k \leftarrow 0$

4: **repeat**

5:     $\alpha_k \leftarrow \dfrac{\langle z_k, z_k \rangle_{\text{model}}}{\langle \bar{F}p_k, \bar{F}p_k \rangle_{\text{data}}}$

6:     $\delta\bar{v}_{k+1} \leftarrow \delta\bar{v}_k + \alpha_k p_k$

7:     $z_{k+1} \leftarrow z_k - \alpha_k \bar{F}^\dagger \bar{F} p_k$

8:     $\beta_{k+1} \leftarrow \dfrac{\langle z_{k+1}, z_{k+1} \rangle_{\text{model}}}{\langle z_k, z_k \rangle_{\text{model}}}$

9:     $p_{k+1} \leftarrow z_{k+1} + \beta_{k+1} p_k$

10:    $k \leftarrow k + 1$

11: **until** Error is sufficiently small, or max iteration count exceeded

---

In the termination statement, "Error" refers to the error $\sqrt{\langle z_k, z_k \rangle}$ in the solution of the normal equation (10). In all of the experiments reported below, termination occurred when the maximum iteration count was reached, as our intent was to explore the asymptotic behaviour of the iteration.

This algorithm may be transformed by introducing temporary vectors such as $q_k = \bar{F}p_k, y_k = \bar{F}^T W_{\text{data}} \bar{F} p_k$, and so on, and the algebra re-arranged to minimize the applications of the weight operators. As these are relatively cheap, most of the computational work resides in the extended modeling ($\bar{F}$) and RTM ($\bar{F}^T$) steps, so the cost per step is negligibly greater than that of standard CG with Euclidean inner products.

NUMERICAL EXAMPLES

In this section, we will compare the performance of ELSM with CG and WCG iterations using several examples. Besides the performance comparison, we will also illustrate the robustness of both data-fitting via ELSM and accelerated convergence of WCG against velocity error.

In all examples, we use a 2nd order in time, 8th order in space centered finite difference algorithm to approximate wave propagation, and apply absorbing boundary conditions on all four sides of the rectangular computational domain. We use

fixed spread geometry. The Born modeling operator uses a point source and the migration operator adopts a dipole source to simulate the source/receiver depth derivative operators required by the approximate inverse, as discussed in the preceding section.

Simple Example

The first example involves a simple layered model, shown in Figures 5. The computational domain is a 3×3 km rectangular grid with 10 m grid interval in all directions. Subsurface offset ranges over the interval [-250 m, 250 m], with the same 10 m interval as the spatial coordinates. We generate 76 shots using a 2.5-5-30-35 Hz zero phase trapezoidal bandpass wavelet and a discrete isotropic point source representation. The time step is 1 ms. All 76 shots are recorded in the 301 receivers spaced 10 m apart. Shot spacing is 40 m, and the first source and receiver are located at $x = 0$ m. Depth of both sources and receivers is 10 m.
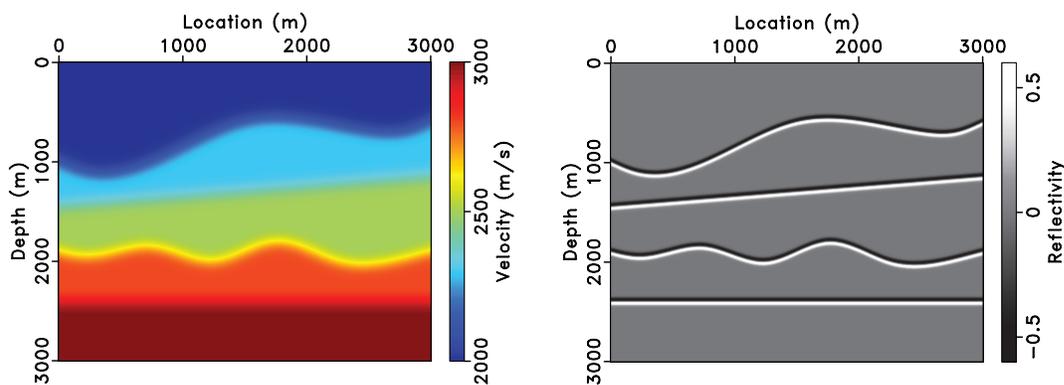
Figure 5: An example layered model with simple structures. (a) Background velocity model (b) Reflectivity model.

The inverted ELSM models using 20 iterations of CG and WCG appear in Figure 6. The zero offset section for both cases depicts the main structures of the model with focused energy. However, one can observe obvious lower resolution for the CG result in comparison to the WCG result: in fact, CG is much further from convergence.

As discussed in Hou and Symes (2014, 2015), stacking the output extended image over offset produces a physical image. The stacked image from the WCG inversion
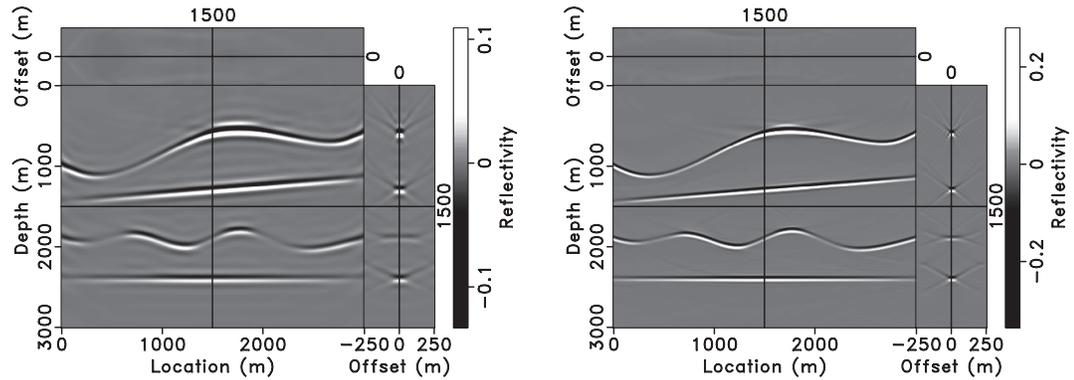
Figure 6: 20 iterations result of ELSM via (a) CG iteration (b) WCG iteration.

volume shown in Figure 7a reconstructs the original reflectivity model with very small error. The difference with the original model is mainly composed of edge artifacts (Figure 7b).
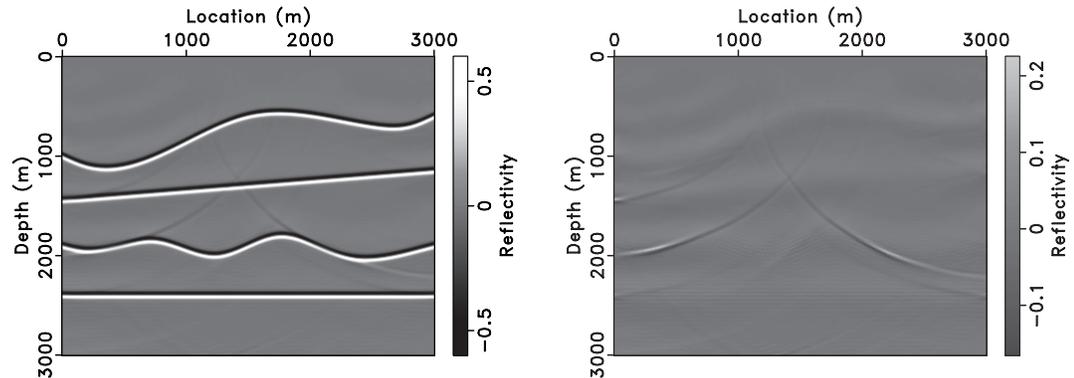
Figure 7: (a) Physical image obtained by stacking 20 WCG iterations ELSM result along subsurface offset axis (b) Difference between the obtained physical image and the original reflectivity model. All figures are plotted on the same grey scale.

The misfit plot, displayed in Figure 8a, exhibits the remarkable acceleration of WCG over CG. The required number of iterations for a given error level in WCG is substantially smaller than that required for CG. In fact, the first iteration of WCG, which is the approximate inverse operator itself, is comparable to the 20 iteration result of CG. Figure 8b plots the norm of the normal residual (least squares

gradient), again confirming the accelerated convergence of WCG compared to CG.

Figure 8: (a) Relative data misfit plot against the number of iterations (b) Relative normal residual plot against the number of iterations. The blue line represents the WCG iteration and the red line shows the CG iteartion.

Marmousi Example

Our second example is based on the Marmousi model (Versteeg and Grau, 1991). We smooth the original model as the background model (Figure 9a) for simulation and migration and take the difference as the reflectivity model (Figure 9b). The synthetic data correspond to a fixed spread acquisition with a source spacing of 40 m and receiver spacing of 20 m. Both the first source and receiver are at $x = 0$ m. 231 sources and 461 receivers are all placed at 20 m depth. Spatial grid interval is 20 m in x, z, and (subsurface offset) h, and the subsurface offset range is [-500 m, 500 m]. The source wavelet is a 2.5-5-20-25 Hz trapezoidal bandpass wavelet with 2 ms time interval, also the time step in the finite difference computations.
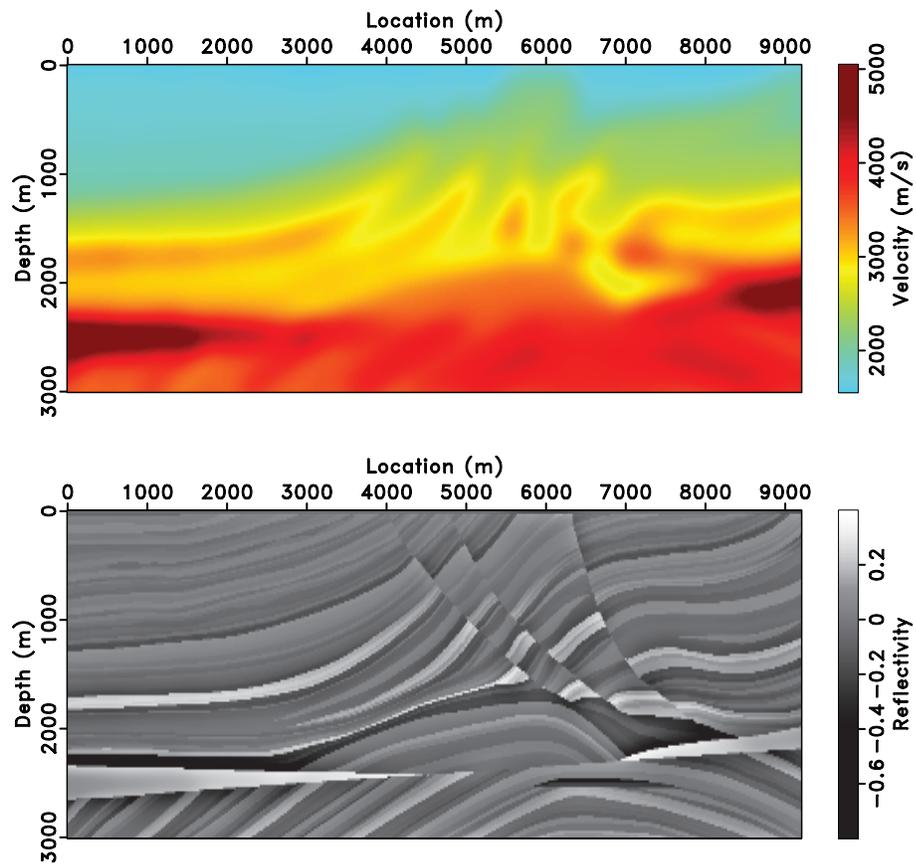
Figure 9: Marmousi Model. (a) Background velocity model obtained by smoothing the original Marmousi model (b) Reflectivity model obtained by taking the difference between orginal velocity model and background velocity model.

Figure 10 shows the results of 20 iterations. The stacked images from both CG iteration and WCG iteration are shown in Figure 11. The WCG algorithm yields an inversion with noticeable higher resolution than does CG with the same number of iterations.
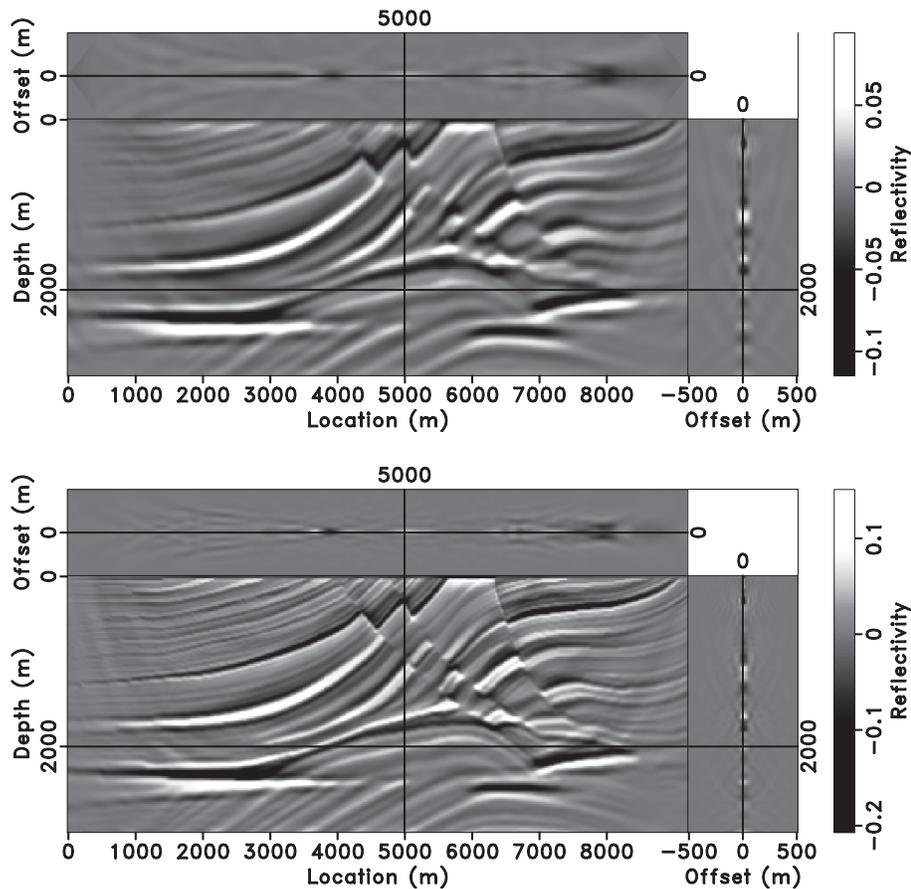
Figure 10: 20 iterations result of ELSM via (a) CG iteration (b) WCG iteration.

Figure 12 plots the normalized data misfit and the normal residual as a function of the iteration number for the Marmousi example. As in the first example, WCG exhibits dramatically faster convergence than CG.

As indicated in the theory section, ELSM should permit good fit to data regardless of the accuracy of the background velocity, within broad limits. Also, as established by Hou and Symes (2015), the approximate inverse described earlier is accurate even with erroneous velocity, that is, approximates the solution of the
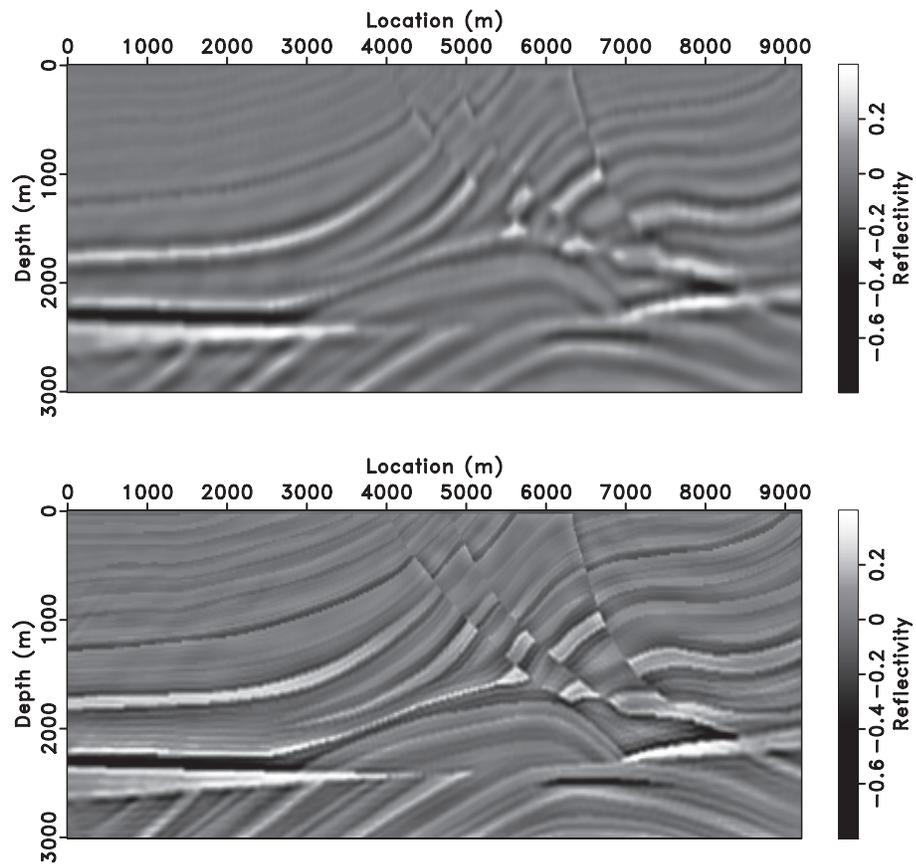
Figure 11: (a) Physical image obtained by stacking 20 CG iterations ELSM result along subsurface offset axis (b) Physical image obtained by stacking 20 WCG iterations ELSM result along subsurface offset axis.

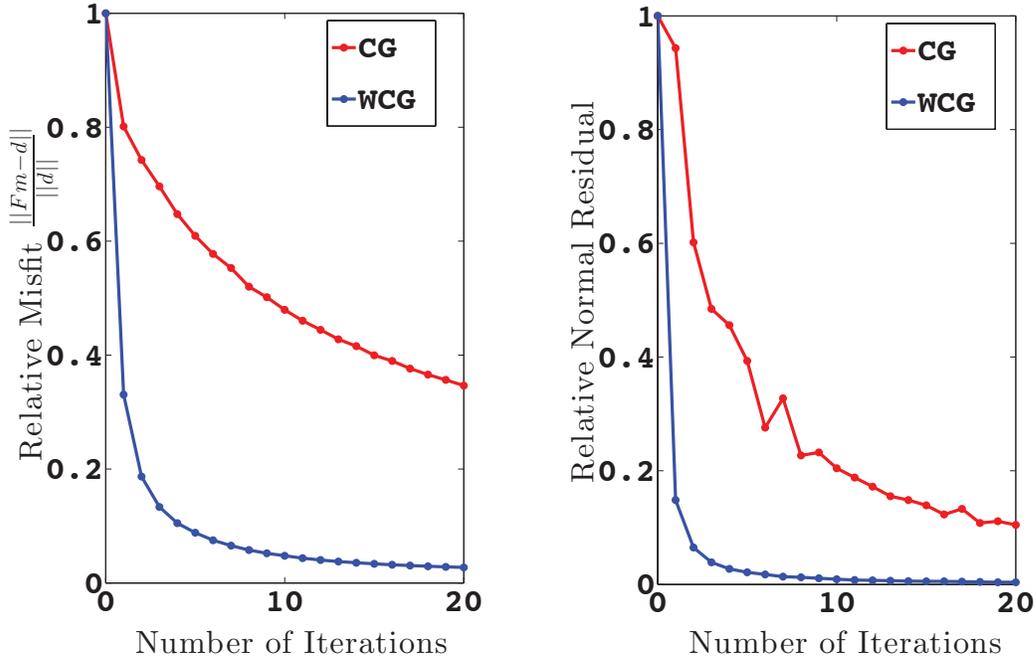Figure 12: Marmousi example: (a) Relative data misfit plot against the number of iteartions (b) Relative normal residual plot against the number of iterations. The blue line represents the WCG iteration and the red line shows the CG iteartion.

ELSM problem. Therefore we expect the same acceleration of WCG over CG even in the presence of obvious velocity error. We conduct the same experiment with an incorrect velocity model, by reducing the background velocity everywhere by a factor of 0.9. After 20 iterations CG and WCG, we obtain the results shown in Figure 13. The subsurface offset ranges over [-1000 m, 1000 m]. Due to the incompatibility of the background velocity used in this ELSM with the data, the inversion is not focused at zero offset: energy spreads significantly to nonzero offset.
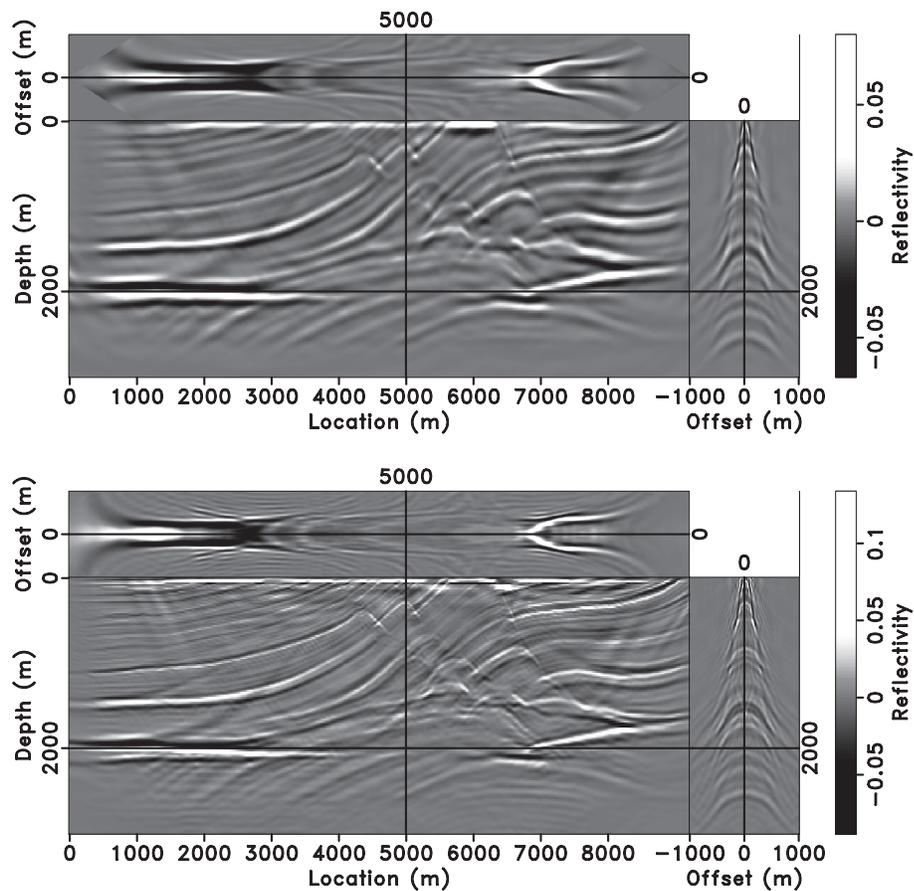
Figure 13: 20 iterations result of ELSM via (a) CG iteration (b) WCG iteration.

Of course it is the presence of nonzero offset reflectivity that permits ELSM to fit data even with erroneous velocity, although in order to make this happen, the offset range needed to be twice that used in the preceding experiment. Both CG and WCG versions of ELSM yield increasing data fit with iteration, though WCG

reduces misfit much faster than does CG. The convergence behaviour of the two algorithms is very similar to that shown earlier for the correct background velocity (Figures 14).

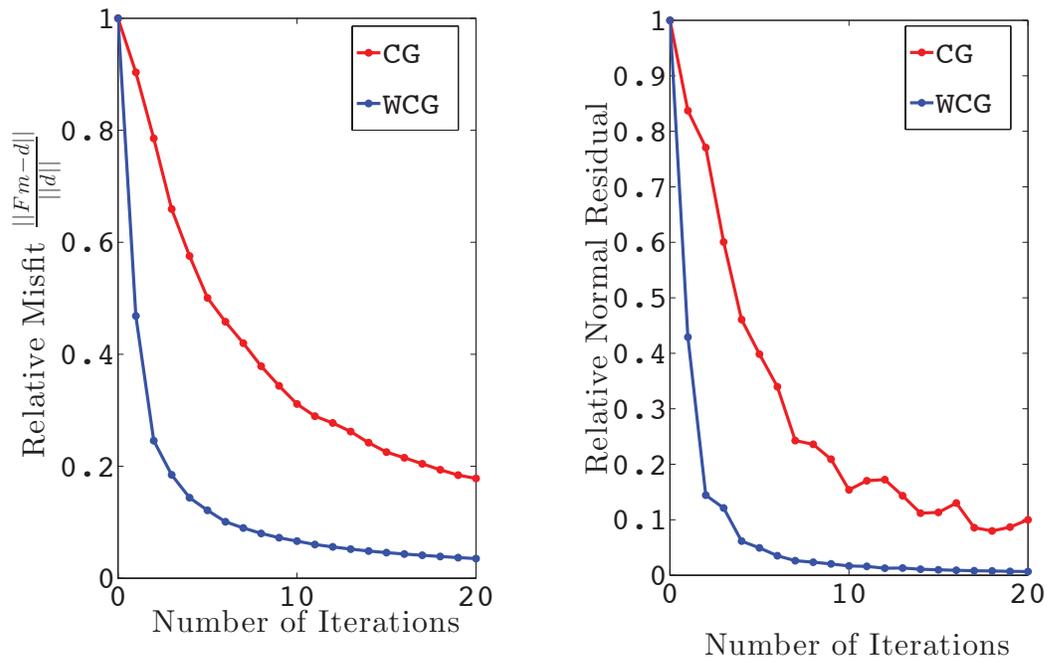Figure 14: (a) Relative data misfit plot against the number of iterations (b) Relative normal residual plot against the number of iterations. The blue line represents the WCG iteration and the red line shows the CG iteration.

Salt Example

Our third example is based on a 2D section of the SEG/EAGE salt model (Aminzadeh et al., 1997). The background velocity model shown in Figure 15a is achieved by (a) removing the salt from the original salt model, (b) smoothing the salt-free result, and (c) inserting the salt body into the smoothed background. The difference between the background model and original velocity model is used as the reflectivity model, displayed as Figure 15b. Note that the reflectivity vanishes inside the salt. The model is discretized with 20 m grid spacing in all directions. The experiment simulates 251 evenly spaced shots and records at 501 evenly spaced receivers, both at 20 m depth. The shots begin at $x = 0$ m with a shot

interval of 40 m and the receiver array begins at the same position with a receiver interval of 20 m. The subsurface offset range is [-500 m, 500 m], and the offset interval is also 20 m. The source waveform is a 2.5-5-20-25 Hz bandpass wavelet with 2 ms time interval. The velocity discontinuity at the salt boundary violates the scale-separation hypotheses underlying the theory of the approximate inverse (Hou and Symes, 2015): that is, the velocity model has significant spectral content at spatial frequencies excited by the data. We will use this example to show that the WCG algorithm has an improved convergent rate even in the presence of discontinuities in the background model, at least in some cases.
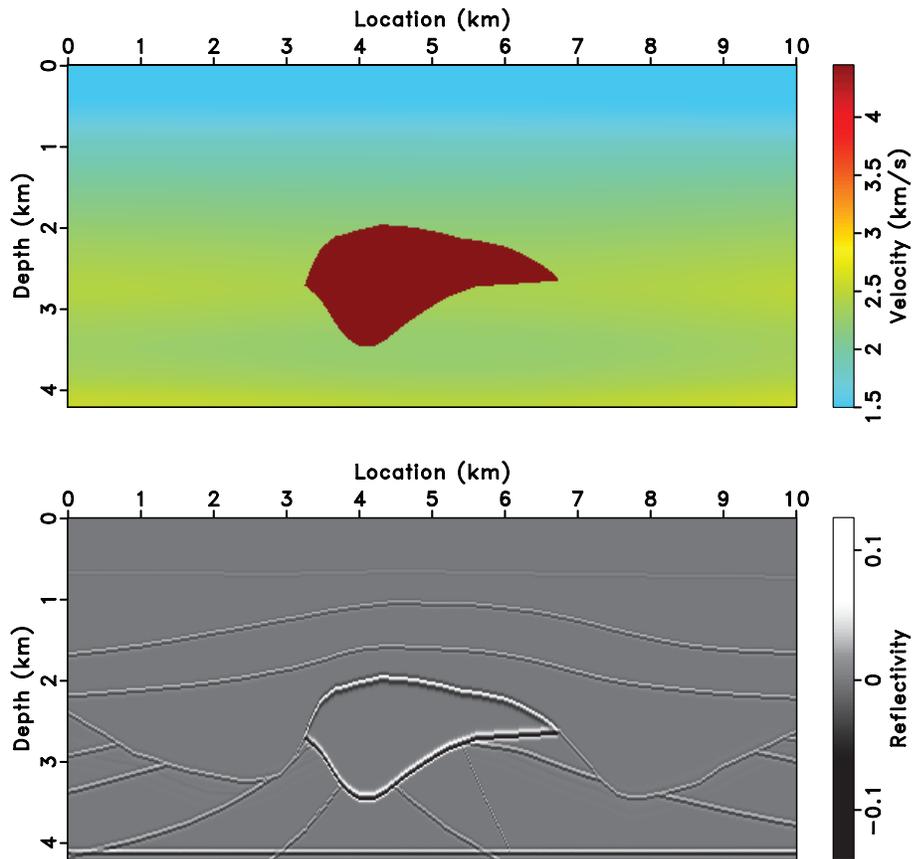
Figure 15: SEG/EAGE salt model. (a) Smooth background velocity model with salt overlaid (b) Reflectivity model.

Both CG and WCG are used to generate the extended inversions, displayed in Figure 16.
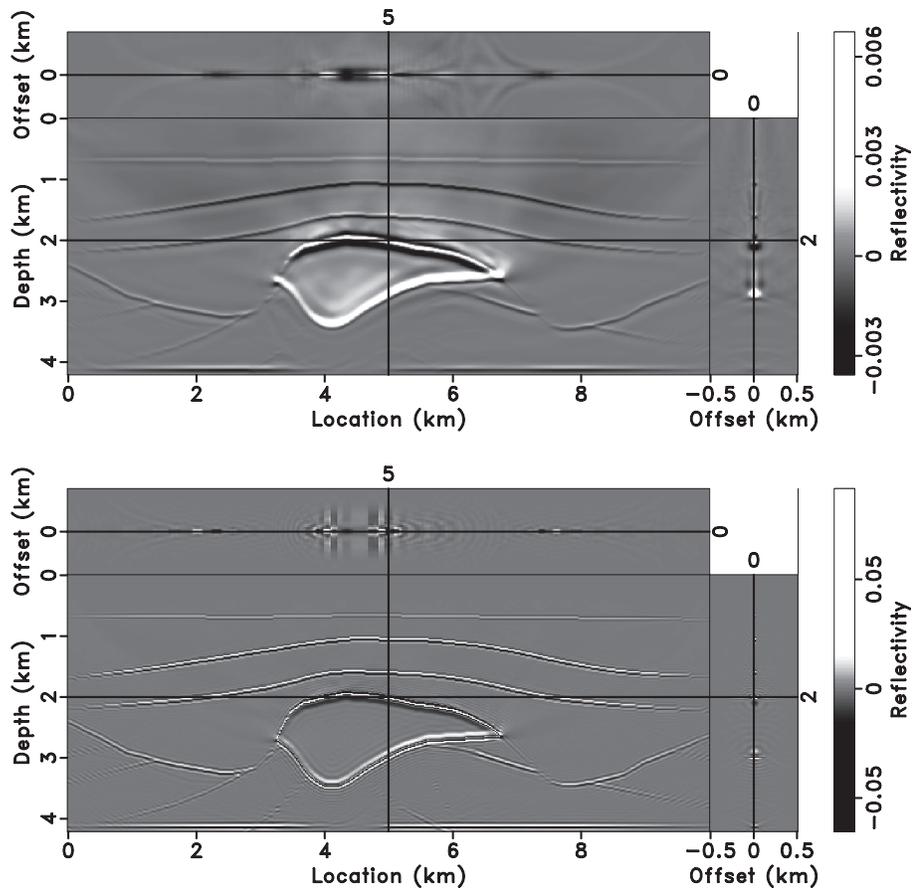
Figure 16: 20 iterations result of ELSM via (a) CG iteration (b) WCG iteration.

Both algorithms correctly recover the main structures and salt boundary of the salt. Classical low frequency noise above the salt is still present in the CG result while WCG removes most of the top salt noise, probably a side effect of the Laplacian filtering implicit in the model space weight operator W_{model}. WCG also better recovers subsalt reflectivity. Stacking over offset yields the image shown in Figure 17.
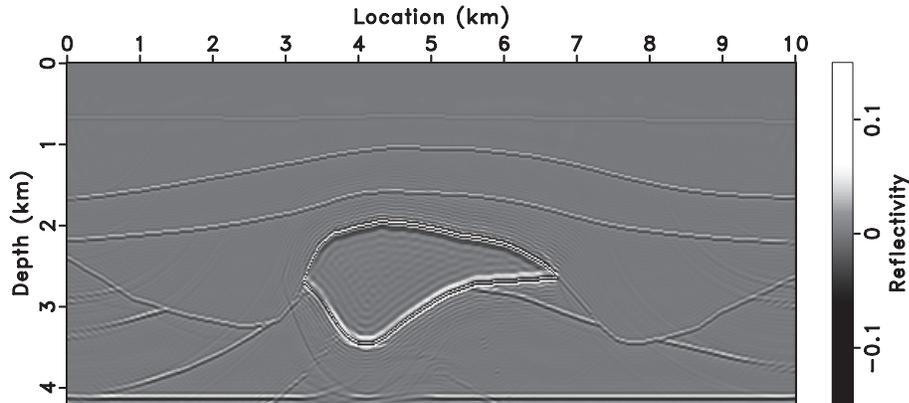
Figure 17: Physical image obtained by stacking 20 WCG iterations ELSM result along subsurface offset axis. Plotted on the same grey scale with the reflectivity model.

Figure 18 shows the misfit plot and normal residual plot for CG and WCG. From the figure, it is clear that the proposed WCG has substantially faster convergence for this example also. The salt in the background appears to considerably slow the convergence of (unweighted) for CG iteration, while barely influencing the progress of WCG iteration.

DISCUSSION

We have described a method to accelerate the convergence of conjugate gradient iteration for extended linearized inversion (ELSM), and illustrated the acceleration so obtained with several numerical examples. In this section we note several interesting aspects of this topic that merit further study.

The careful reader will have noticed that in Figure 4, the misfit for ELSM is considerably larger at 30 iterations than is that for LSM (with correct velocity model), and one could wonder whether the convergence curve might be flattening out and
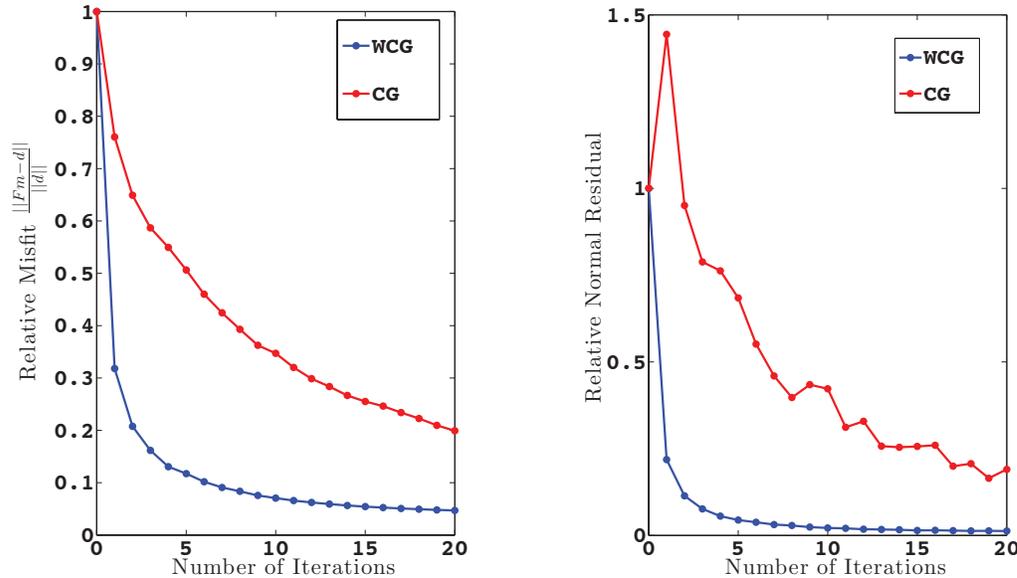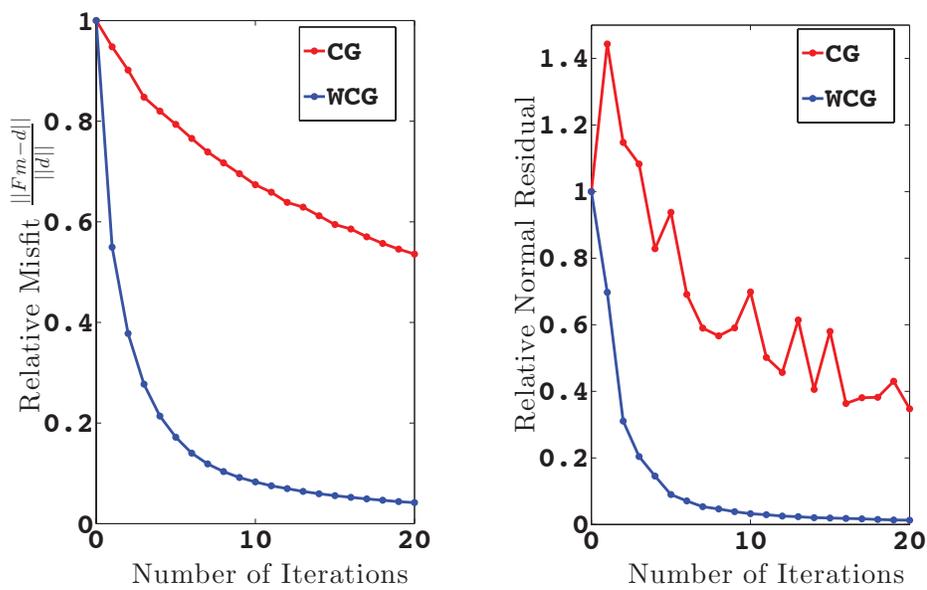
Figure 18: (a) Relative data misfit plot against the number of iteartions (b) Relative normal residual plot against the number of iterations. The blue line represents the WCG iteration and the red line shows the CG iteration.

possibly even asymptotic to a positive value. Similar behavious is visible in other convergence plots, for both CG and WCG (though the latter reduces the residual norm much faster than the former). The residual norm would be expected to converge to a positive value if the data were not in the range of the modeling operator - in principle impossible, as our experiments are all inverse crimes, with data computed using the same algorithms as used in the inversion. Instead, we suggest that the addition of many degrees of freedom via the subsurface offset extension has expanded the spectrum of the linearized modeling operator, and thus enlarged its condition number and in particular the number of singular values near zero. This would be true even for the weighted norms: as remarked in the theory section, the weight operators are actually only semidefinite in the continuum limit. Therefore it would not be so surprising that the error reduction curves for ELSM would have relatively long tails. This conclusion would also be consistent with the rapid convergence to zero of the gradient norms. The spectrum of the extended modeling operator may have some relation to the ray theory of extended modeling, and is surely worth some attention.

The extra offset axis is an unattracitve aspect of extended modeling and inversion. It adds to both memory and computation intensity of subsurface offset ELSM: as noted by Mulder and van Leeuwen (2008), the cycles used in summations over offset can easily overwhelm the other finite difference operations. Against this considerable cost one must count the robustness of data fit against velocity error, noted and illustrated above. However LSM assuming a sufficiently correct velocity is of considerable interest, and one must wonder whether carrying the (in principle) useless subsurface offsets is really necessary for an efficient LSM algorithm. In fact, we have found that in the special case of ELSM for data-consistent velocity, it is possible to enforce the focusing at $h = 0$ that should eventually result by using only one sample in offset for the bulk of the computation. By resorting to multiple offsets only for the computation of the model weighting operator, an inexpensive step as it's outside of the time loop, it is still possible to obtain the sort of acceleration over CG described here. We will discuss the resulting effcient LSM algorithm elsewhere.

This paper has adopted the simplest useful wave physics, constant density acoustics, and it is natural to wonder whether similar reasoning could be applied to generate efficient extended inversion algorithms for more physically complete descriptions of seismic wave motions. ten Kroode (2012) based his "wave equation" approach to asymptotic Kirchhoff inversion on variable density acoustics, and a similar treatment is certainly possible for Born inversion. The most important fur-

ther step is surely to anisotropic elasticity. It remains to be seen, to what extend asymptotic approximate inversion via wave equation computations is possible in that setting, in particular whether multicomponent data and/or polarization is required. Finally, wave attenuation and dispersion are ubiquitous, and it is not even clear what "asymptotic approximation" should mean when those phenomena are significant.

CONCLUSION

We have proposed a model-dependent choice of norms in conjugate gradient algorithm, based on the construction of a high-frequency asymptotic approximate inverse to the extended Born modeling operator of constant-density acoustics, to accelerate the convergence of iterative extended least squares migration. The weighted norm algorithm has roughly the same computational cost as the unweighted algorithm commonly used for least squares migration, and in particular involves no ray tracing. Numerical examples suggest that weighted conjugate gradient iteration converges considerably more rapidly, and to a better extended linearized inversion estimate of reflectivity, than does unweighted conjugate gradient iteration. The improved convergence survives velocity error, as theory predicts, and even model features that contradict the theoretical underpinnings of the wave-equation based asymptotic approximate inverse construction.

ACKNOWLEDGMENTS

REFERENCES

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

Aoki, N., and G. Schuster, 2009, Fast least-squares migration with a deblurring filter: Geophysics, **74**, no. 6, WCA83–WCA93.

Beasley, C. J., R. E. Chambers, and Z. Jiang, 1998, A new look at simultaneous sources: 69th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 133–135.

Bednar, J. B., and C. J. Bednar, 2006, Two-way vs. one-way: a case study style comparison: 76th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2343–2347.

Berkhout, G., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, 924–938.

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.

Björk, A., 1997, Numerical methods for least squares problems: Society for Industrial and Applied Mathematics.

Bourgeois, A., B. Jiang, and P. Lailly, 1989, Linearized inversion: A significant step beyond pre-stack migration: Geophysical Journal International, **99**, 435–445.

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

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

De Hoop, M. V., and N. Bleistein, 1997, Generalized Radon transform inversions for reflectivity in anisotropic elastic media: Inverse Problems, **16**, 669–690.

Dutta, G., Y. Huang, W. Dai, X. Wang, and G. T. Schuster, 2014, Making the most out of the least (squares migration): 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4405–4410.

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.

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.

Ikelle, L., 2007, Coding and decoding: Seismic data modeling, acquisition and processing: 77th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 66–70.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-waveform seismic inversion using encoded

sources: Geophysics, **74**, no. 6, WCC177–WCC188.

Kuehl, H., and M. Sacchi, 2003, Least-squares wave equation migration for AVP/AVA inversion: Geophysics, **68**, 262–271.

Liu, Y., W. Symes, and Z. Li, 2013, Multisource least-squares extended reverse time migration with preconditioning guided gradient method: 83rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 3709–3715.

Mulder, W. A., and R.-E. Plessix, 2004, A comparison between one-way and two-way wave equation migration: Geophysics, **69**, 1491–1504.

Mulder, W. A., and T. van Leeuwen, 2008, Automatic velocity analysis and multiples: 78th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 1184–1191.

Nemeth, T., C. Wu, and G. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

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.

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

Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth migration: Geophysics, **68**, 1371–1379.

Romero, A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

Schuster, G. T., X. Wang, Y. Huang, W. Dai, and C. Boonyasiriwat, 2011, Theory of multisource crosstalk reduction by phase-encoded statics: Geophysical Journal International, **184**, 1289–1303.

Shin, C., S. Jang, and D.-J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

Stolk, C. C., and M. V. De Hoop, 2005, Modeling of seismic data in the downward continuation approach: SIAM Journal on Applied Mathematics, **65**, 1388–1406.

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

Symes, W., 2008a, Approximate linearized inversion by optimal scaling of prestack depth migration: Geophysics, **73**, no. 2, R23–35.

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

——, 2009, The seismic reflection inverse problem: Inverse Problems, **25**,

123008:1–24.

Tang, B., S. Xu, and Y. Zhang, 2013, 3D angle gathers with plane-wave reverse time migration: Geophysics, **78**, no. 2, S117–S123.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian: Geophysics, **74**, no. 6, WCA95–WCA107.

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.

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

Xue, Z., Y. Chen, S. Fomel, and J. Sun, 2014, Imaging incomplete data and simultaneous-source data using least-squares reverse-time migration with shaping regularization: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 3991–3996.

Zhang, Y., G. Zhang, and N. Bleistein, 2003, True amplitude wave equation migration arising from true amplitude one-way wave equations: Inverse Problems, **19**, 1113–1138.

Flexibly Preconditioned Extended Least Squares Migration in Shot Record Domain (*Geophysics*, submitted)

Yin Huang[1] and Rami Nammour[2] and William Symes[1]

ABSTRACT

This paper demonstrates a method for accelerating the convergence of iterative least squares migration. The algorithm uses a pseudodifferential scaling (dip- and spatially-varying filter) preconditioner together with a variant of conjugate gradient iteration with iterate-dependent ("flexible") preconditioning. The migration is formulated without the image stack, thus producing a shot dependent image volume that retains offset information useful for velocity updating and amplitude-versus-offset analysis. Numerical experiments show that flexible preconditioning with pseudodifferential scaling not only attains considerably smaller data misfit and gradient error for a given computational effort, but also produces higher resolution image volumes with more balanced amplitude and fewer artifacts than is achieved with a non-preconditioned conjugate gradient method.

INTRODUCTION

While prestack reverse time migration (RTM) (Baysal et al., 1983; Loewenthal and Mufti, 1983; Mulder and Plessix, 2004) can produce accurate images of complex subsurface structure, it can also suffer from unbalanced amplitude and illumination artifacts, acquisition footprint, and imperfect focusing of the seismic wavelet. Least squares RTM, or LSM, an alternate name for iterative least-squares linearized inversion (Bourgeois et al., 1989; Chavent and Plessix, 1999; Nemeth et al., 1999; Tang, 2009; Aoki and Schuster, 2009; Dutta et al., 2014; Luo and Hale, 2014; Valenciano et al., 2015), can resolve the aforementioned problems of

RTM (its first iteration) to some extent. Since each iteration requires a migration-modeling pair, the expense of LSM can be considerable, yet achieving its benefits requires that sufficient iterations be performed.

In this paper, we show how to accelerate iterative LSM by combining optimal pseudodifferential scaling (Symes, 2008a; Nammour and Symes, 2009; Nammour, 2009) with the Flexibly Preconditioned Conjugate Gradient (FPCG) algorithm (Notay, 2000; Knyazev and Lashuk, 2006). Our algorithm is designed to invert each shot record individually, and thus produce image gathers that can be used for velocity and amplitude-versus-offset analysis. We argue that the theoretical properties of this algorithm suggest that it should substantially reduce the number of iterations required for good data fit and small LSM gradient, and illustrate this conclusion with a number of examples showing the benefits of accelerated LSM in better image amplitude and phase for a given computational effort.

Many previous works have addressed computational efficiency of LSM, or linearized inversion. Our work belongs to the *scaling* genre, in which the (computable) action of the normal operator or Hessian $F^T F$ of the scattering operator F on an image or set of images is used to constrain a matrix representation, which is then inverted to approximately solve the normal linear system equivalent to least squares minimization. The type of algorithm explored in this paper originated in the work of Claerbout and Nichols (1994), who computed an approximate inverse Hessian scale factor by point-wise division of migrated image $F^T d$ by the output of the Hessian applied migrated image, that is, $(F^T F)F^T d$. Rickett (2003) polished this idea and compared it with other alternatives for replacing the migrated image in the denominator. This type of approximation implicitly presumes that the Hessian acts mostly by space-dependent scaling of its input, that is, as a spatially diagonal operator: in effect, that the Hessian acts mostly to modify amplitudes. However, Chavent and Plessix (1999) showed that the Hessian is not spatially diagonal, or even diagonally dominant: that is, the Hessian is also a filter of non-zero width. They suggested taking this non-diagonality into account via a partial mass-lumping method, that is, adding the off-diagonal terms to the diagonal, and then scaling the migrated image by the inverse of resulting diagonal matrix. Shin et al. (2001) suggested another narrow-band approximation to the normal operator, which they call the pseudo-Hessian matrix. Guitton and Kaelin (2006) replaced the diagonal multiplier of Claerbout and Nichols (1994) and Rickett (2003) with a localized filter, and demonstrated a more accurate approximate inversion in some cases. Symes (2008a) explained the non-diagonality observed

by Chavent and Plessix (1999) and others: under well-understood conditions, the normal operator is the product of a known power of the Laplace operator, which is not a spatially diagonal operator, and another operator that really does act as a spatially diagonal operator, at least so long as the image has a well-defined dip field in most places, as seismic images of sedimentary structures tend to do. This paper also introduced an inverse problem for finding the optimal scaling operator, to replace pointwise division. In essence, optimal scaling is possible because the normal operator (LSM Hessian) is *pseudodifferential*, that is, a space-varying filter behaving like a polynomial in spatial frequency for large frequencies (Beylkin, 1985; Rakesh, 1988; ten Kroode et al., 1998; Stolk, 2000). Nammour and Symes (2009); Nammour (2009) used this observation to extend optimal scaling to imaging problems with conflicting dips, using an algorithm due to Bao and Symes (1996) for approximating pseudodifferential operators efficiently. Herrmann et al. (2008) implemented the same idea using a different approach to computing pseudodifferential operators, based on their approximate diagonalization in curvelet frames. Demanet et al. (2012) used the low-rank property implicit in the pseudodifferential nature of the normal operator to approximately constrain a matrix representation by its action on a randomly chosen set of vectors, then efficiently inverting the matrix.

The works cited so far approximate the inverse normal operator, or approximately solve the normal equation. It is natural to think that an approximate inverse could be used to precondition iterative methods for faster convergence, and many of the ideas described in the last paragraph have been used in this fashion, beginning with Chavent and Plessix (1999). We mention in particular Herrmann et al. (2009), who based a preconditioning construction on the use of curvelet representation as in Herrmann et al. (2008). See Pan et al. (2014) for a comparison of several scaling methods used as preconditioners.

Note that under some circumstances approximate least squares solutions can be constructed via asymptotic analysis. The seminal paper Beylkin (1985) led to much work on so-called Generalized Radon Transform inversion, which uses ray tracing quantities to build a Kirchhoff-type integral approximating a pseudo inverse to the scattering operator. See **?**Lameloise et al. (2015) for recent examples of the use of asymptotic inversion based on ray-tracing to accelerate iterative LSM or full waveform inversion. Recently it has been recognized that asymptotic approximate inverses may be constructed without any ray-trace computations whatsoever: see (Zhang et al., 2003; Stolk et al., 2009a; Hou and Symes, 2015b), and

(Hou and Symes, 2015a) for the use of ray-free asymptotics for preconditioning.

The scaling operator concept (and our use of it) does not rely explicitly on high frequency asymptotics or ray tracing, or involve any ray computations. However its justification rests on the pseudodifferential property of the normal operator or Hessian of the scattering operator, and that in turn is a consequence of the ray asymptotics of the wave equation.

Our work differs from that described so far in two main respects. First, we accelerate the solution of the *shot record extension* of linearized (or "Born") forward modeling, in which each shot is modeled by a reflectivity or model perturbation proper to that shot, while sharing a background or reference model with all other shots. The result of a least-squares Born inversion with this extended modeling principle is an inverted *extended* reflectivity volume, analogous to the prestack shot-record migration image volume - which is, in fact, the result of the first iteration of most iterative algorithms for solving the extended problem. The fixed horizontal position slices of this migrated image volume are known as *shot record common image gathers* or CIGs, and are crucial ingredients in migration-based velocity analysis and amplitude-versus-offset analysis. All of the works cited so far have concerned the non-extended version of the Born inverse problem. Our algorithm iteratively inverts the extended Born modeling operator. The extended reflectivity volume so constructed may be similarly sliced into higher-resolution and more artifact-free version of migration CIGs, as we shall illustrate with several examples.

The other major innovation in our work stems from the nature of scaling algorithms, which do not in general produce an approximate inverse of the modeling operator, but rather an operator that approximately solves the least squares problem with specific data. The scaling operator so produced likely does not solve the least squares problem with any data other than that used to generate it. Thus use of a scaling operator as preconditioner requires an algorithm that admits a different preconditioner each iteration. In contrast, all of the work cited so far uses conventional (fixed) preconditioned iterations. Fortunately, algorithms using iteration dependent preconditioning, called Flexibly (or Variably) Preconditioned, have been studied in the computational mathematics literature (Notay, 2000; Knyazev and Lashuk, 2006). We combine a flexibly preconditioned conjugate gradient iteration with the pseudodifferential approximation of Bao and Symes (1996) and the optimal coefficient selection of Symes (2008a) and Nammour and Symes (2009); Nammour (2009) to substantially accelerate the conju-

gate gradient algorithm for shot-record Born inversion, and improve the quality of its output in the bargain.

We note that the preconditioning strategy presented in Hou and Symes (2015a) and related algorithms are based on the use of *subsurface offset extension*, which involves more expensive modeling and migration in comparison with ordinary modeling and migration, but functions well with highly refractive velocity models ("complex structure", see Stolk et al. (2009b)). In contrast, while the shot-record extension leads to accurate imaging only for mild lateral heterogeneity (no multi-pathing, see Stolk and Symes (2004)), it is no more expensive than standard Born modeling and RTM. Optimal pseudodifferential scaling doubles the cost of each iteration of LSM, but considerably reduces the required number of iterations, for significant overall reduction in computational cost, as we shall show.

The rest of the paper is organized as follows: In the theory section, we will briefly discuss the LSM in shot record domain, illustrate its ill-condition property, then review the pseudodifferential scaling optimization and describe a flexibly pre-conditioned conjugate gradient (FPCG) method to solve the extended LSM problem in shot record domain. In the numerical results section, we apply the FPCG method to two simple synthetic problems and to a Born inversion problem derived from the Marmousi model, and compare its efficiency with ordinary conjugate gradient method. We end with a discussion of some possible future developments and applications of FPCG-accelerated extended LSM.

THEORY

We choose constant density acoustics for modeling wave propagation in the derivation of the following theory and algorithm. The constant density acoustic wave equation with squared velocity $m(\mathbf{x}) = c^2(\mathbf{x})$ and isotropic radiator source with wavelet $w(t)$ at source position $\mathbf{x}_s$ is

$$\left(\frac{\partial^2}{\partial t^2} - m(\mathbf{x})\nabla^2\right)u(\mathbf{x}, t; \mathbf{x}_s) = w(t)\delta(\mathbf{x} - \mathbf{x}_s). \tag{1}$$

The shot-record extended Born approximation (Kern and Symes, 1994; Symes, 2008b) allows the model perturbation δm to depend on the shot-record parameter as well as spatial location $\mathbf{x}$. Since we use shot position $\mathbf{x}_s$ to parametrize the isotropic point radiator in equation 1, the extended model perturbation takes the

same functional form $\delta m(\mathbf{x}; \mathbf{x}_s)$ as the image volume of prestack migration sorted by shot record. The corresponding perturbational (or Born) wave equation is

$$\left(\frac{\partial^2}{\partial t^2} - m(\mathbf{x})\nabla^2 \right) \delta u(\mathbf{x}, t; \mathbf{x}_s) = \delta m(\mathbf{x}; \mathbf{x}_s)\nabla^2 u(\mathbf{x}, t; \mathbf{x}_s) \tag{2}$$

The *extended Born modeling operator* is defined by sampling the perturbation field δu at the receiver positions $\mathbf{x}_r$ corresponding to each source:

$$F[m]\delta m(\mathbf{x}_r, t; \mathbf{x}_s) = \delta u(\mathbf{x}_r, t; \mathbf{x}_s). \tag{3}$$

The adjoint operator F^T of the extended Born modeling operator F is the reverse time migration (RTM) with the standard cross-correlation imaging condition (Tarantola, 1984b), but without the final stack over shot position: thus the output of F^T is a shot-dependent image volume field of the same type as δm. The corresponding adjoint operator of the non-extended Born modeling operator stacks the output of the extended adjoint over shot position to produce an image (function of spatial position).

With these notational conventions, the linearized least squares, or LSM, problem takes the following form: given a data perturbation d, find a model perturbation δm to minimize

$$\frac{1}{2}\|F[m]\delta m - d\|^2. \tag{4}$$

As (4) is quadratic in δm, the global minimizer is obtained at any stationary point. Thus the minimizer of the problem (4) is equivalent to the solution of the normal equations

$$F^T F \delta m = F^T d, \tag{5}$$

Since we will fix the background model m in almost all the cases, we write the extended Born modeling operator as F instead of $F[m]$ for simplicity. Due to the size of the problem, the computation of $(F^T F)^{-1}$, or solution of equation 5, by any variant of Gaussian elimination is infeasible even for 2D field-scale examples. Thus iterative methods are used to solve equation 5. Most of these converge at rates that depend on the spectrum of the normal operator $F^T F$, and are faster when the spectrum has small extent. However $F^T F$ tends to be ill-conditioned (have a spectrum spread over a large interval) for several reasons. The limited aperture of field survey geometry implies that some structural perturbations have no appreciable

impact on the Born seismogram, and imply existence of very small eigenvalues. The dynamic range of geometric spreading also tends to spread the spectrum. Most dramatic in effect is the frequency-dependent scaling between model and data perturbation: in 2D, an oscillatory localized plane wave component of δm, of spatial frequency k, results in an output of $F^T F$ amplified by $O(k)$. The analogous amplification in 3D is $O(k^2)$. This frequency-dependent scaling has been understood since Beylkin (1985). In effect, the 2D version of $F^T F$ has the square root of the Laplace operator as a factor, whereas the 3D version has the Laplace operator itself. In the continuum limit, the spectrum is without bound; for discretization, the largest eigenvalue grows in proportion to the Nyquist frequency.

Ill-conditioning due primarily to the frequency-dependent scaling of $F^T F$ has a disastrous effect on convergence of iterative methods for solution of equation 5. While the analysis simply suggests that the condition of equation 5 grows worse with increasing frequency, numerical experiment suggests that it is quite bad already for length and frequency scales characteristic of exploration seismology. We illustrate this effect with a simple example, in which we attempt to invert for a model perturbation δm representing a single reflector in the presence of a slow Gaussian anomaly in the background model, as shown in Figures 1a and 1b.
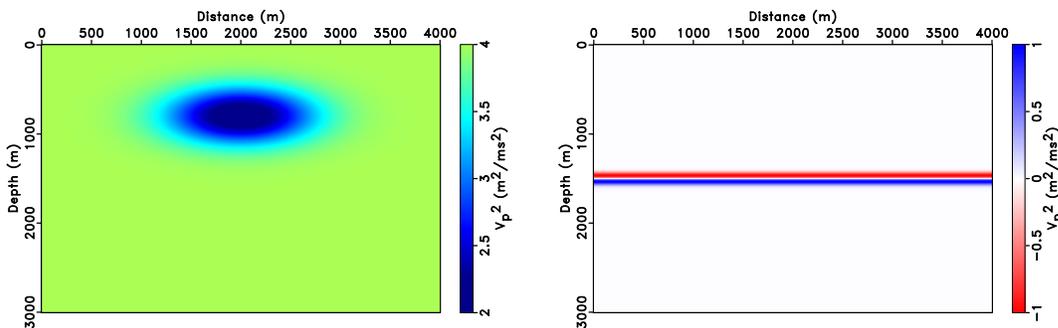
Figure 1: Slow Gaussian anomaly model: (a) background m, (b) perturbation δm.

We use point sources with impulse wavelets bandpass filtered with corner frequencies 1 Hz, 7 Hz, 28 Hz, 35 Hz. A total of 81 point sources are placed at depth 20 m with a spacing of 50 m. 401 fixed receivers with a spacing 10 m are placed at depth 10 m. Data is recorded until 2.4 s.

We use a very simple iterative algorithm, the steepest descent method with fixed
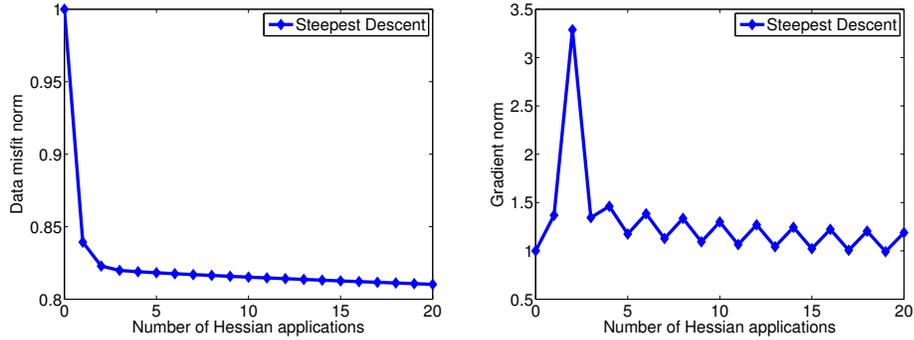
Figure 2: Convergence curve of steepest descent method for the slow Gaussian anomaly problem: (a) normalized data misfit (b) normalized gradient length.

step length chosen as a crude estimate of the reciprocal maximum eigenvalue of $F^T F$ (Boyd and Vandenberghe, 2004). Although this choice of step length guarantees decrease of the objective function value at each iteration, very slow convergence rate is expected if the condition number of $F^T F$ is large. The data misfit history (Figure 2a) shows that there is no big decrease in the error after 3 steps of iteration. The zigzag decrease of the normal residual (length of the difference between the left hand side and right hand side in equation 5, Figure 2b) is typical of the steepest descent method and indicates the ill-conditioning property of the normal operator $F^T F$.

This simple example strongly indicates that a reformulation of equation 5 is required for reasonably fast convergence of iterative solvers.

Pseudodifferential Scaling

The second author introduced an approximate inversion by scaling (Nammour and Symes, 2009; Nammour, 2009), based on a fundamental observation about the normal operator $F^T F$ established by Beylkin (1985); Rakesh (1988); ten Kroode et al. (1998); Stolk (2000), namely that $F^T F$ is a *pseudodifferential operator* (ΨDO) under some conditions, a synopsis of which appears at the end of this section. This fact underlies the effectiveness of migration as an imaging algorithm.

We point out several important properties of ΨDOs:

- their action preserves the location and orientation of localized oscillatory plane waves;

- to each is associated an *order*, which captures the power of frequency with which localized plane waves are scaled; and

- they are expressed via an oscillatory integral, examples below, and take the form of a spatially variable filter.

The first of these facts explains the connection between the ΨDO nature of the normal operator and the imaging property of the RTM operator F^T: its effect on the data created by $F\delta m$ is to recover localized oscillatory plane waves components of δm in location and orientation.

The oscillatory integral form of the ΨDO $F^T F$ is

$$F^T F \delta m(\mathbf{x}) \approx \int a(\mathbf{x}, \mathbf{k}) \hat{\delta m}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \tag{6}$$

The function of position and wavenumber $a(\mathbf{x}, \mathbf{k})$ is the *principal symbol* of $F^T F$; a is homogeneous of order s in wavevector $\mathbf{k}$ (identical to the order of the operator, mentioned above) and depends on the background model m. Note that if a were independent of $\mathbf{x}$, then equation 6 would define a filter, and thus could be computed inexpensively via FFT. However, the dependence on $\mathbf{x}$ is essential: includes geometric amplitudes and other factors. The theory of ΨDOs (Taylor, 1981) requires that a is *smooth*, that is, with arbitrarily many well-defined derivatives, in both $\mathbf{x}$ and $\mathbf{k}$. Smoothness of m implies this property for the principal symbol of $F^T F$, under some assumptions that will be detailed below.

The order s is dimension-dependent: for dimension d, the order is $s = d - 1$ (Rakesh, 1988). The sense of approximation suggested in equation 6 is asymptotic, in the sense that the difference between the two sides is $O(|\mathbf{k}|^{s-1})$, in terms of spatial frequency.

From here on, we will consider the 2D case only. In particular, $s = 1$.

The pseudodifferential scaling problem (Nammour and Symes, 2009; Nammour, 2009) constructs an approximate solution of equation 5: for a given image, for

example $b = F^T d$ with d the observed data, find a pseudodifferential operator Q with principal symbol $q(\mathbf{x}, \mathbf{k})$ of order -1,

$$Qu(\mathbf{x}) = \int q(\mathbf{x}, \mathbf{k})\hat{u}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}d\mathbf{k}, \tag{7}$$

such that

$$QF^T Fb \approx b. \tag{8}$$

That the order of Q should be -1 is a consequence of one of the rules for combining ΨDOs: the product of two ΨDOs is another ΨDO, whose order is the sum of the orders of the factors (Taylor, 1981). Since the identity operator is a ΨDO of order 0 (principal symbol $\equiv 1$!), and the 2D version of $F^T F$ has order 1, the approximate inverse ΨDO should have order -1.

Whatever approach one might use to solve the pseudodifferential scaling problem, it is necessary to compute the actions of the operators Q and $F^T F$, if only to gauge success. We will use a finite difference time-stepping method to approximate both Born modeling (F) and RTM (F^T). To approximate the action of Q, we use the algorithm of Bao and Symes (1996). This algorithm is based on the observation that since the principal symbol $q_{-1}(\mathbf{x}, \mathbf{k})$ of Q is homogeneous of order -1 in $\mathbf{k} = k\hat{\mathbf{k}} = (k\cos\theta, k\sin\theta)$, $q_{-1}(\mathbf{x}, \mathbf{k}) = k^{-1}q_{-1}(\mathbf{x}, \hat{\mathbf{k}})$ is a 2π-periodic smooth function of θ, hence approximated using its first $N + 1$ Fourier components with exponentially small error in N:

$$q_{-1}(\mathbf{x}, \hat{\mathbf{k}}) \approx \frac{1}{2}c_0(\mathbf{x}) + \sum_{n=1}^{N} c_n(\mathbf{x})C_n(\hat{\mathbf{k}}) + s_n(\mathbf{x})S_n(\hat{\mathbf{k}}) \tag{9}$$

The spatial functions $\{c_n\}$ and $\{s_n\}$ are Fourier coefficients, and $\{C_n\}$ and $\{S_n\}$ are polynomials in the unit vector $\hat{\mathbf{k}}$ defined by the conditions:

$$\begin{aligned} C_n(\hat{\mathbf{k}}) &= \cos n\theta \\ S_n(\hat{\mathbf{k}}) &= \sin n\theta \end{aligned} \tag{10}$$

Recursion rules derived from the sum-of-angle formulas are the easiest way to compute $\{C_n\}$ and $\{S_n\}$:

$$\begin{aligned} C_{n+1}(\hat{\mathbf{k}}) &= 2\hat{k}_1 C_n(\hat{\mathbf{k}}) - C_{n-1}(\hat{\mathbf{k}}); C_0 = 1, C_1(\hat{\mathbf{k}}) = \hat{k}_1 \\ S_{n+1}(\hat{\mathbf{k}}) &= 2\hat{k}_2 C_n(\hat{\mathbf{k}}) + S_{n-1}(\hat{\mathbf{k}}); S_0 = 0, S_1(\hat{\mathbf{k}}) = \hat{k}_2 \end{aligned} \tag{11}$$

The reader will note that the first of these rules defines the Chebyshev polynomials in $\hat{k}_1$.

It remains to discretize the equation 7 in $\mathbf{x}$. Bao and Symes (1996) use cubic spline approximation on a coarse nodal grid, taking advantage of the smoothness of $q(\mathbf{x}, \mathbf{k})$ and the implied smoothness of $\{c_n(\mathbf{x})\}$ and $\{s_n(\mathbf{x})\}$. In terms of a B-spline basis $\{\psi_j(\mathbf{x}) : 1 \le j \le J\}$,

$$c_n(\mathbf{x}) = \sum_{j=1}^{J} c_n^j \psi_j(\mathbf{x})$$

$$s_n(\mathbf{x}) = \sum_{j=1}^{J} s_n^j \psi_j(\mathbf{x}) \tag{12}$$

Combine equations 9 through 12 with the definition 7 and rearrange to obtain the approximation of the action of Q in terms of the coefficient sequence $\mathbf{c} = \{c_n^j, s_n^j : 0 \le n \le N, 1 \le j \le J\}$:

$$Q[\mathbf{c}]u(\mathbf{x}) =$$

$$\sum_{j=1}^{J} \left(\sum_{n=0}^{N} \left[c_n^j \int k^{-1} C_n(\hat{\mathbf{k}}) \hat{u}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} + s_n^j \int k^{-1} S_n(\hat{\mathbf{k}}) \hat{u}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \right] \right) \psi_j(\mathbf{x}). \tag{13}$$

Each integral inside the square bracket is actually a filter hence can be computed with the aid of the FFT at low cost. There are only $O(N)$ such filters to compute, for a much lower total cost per application than a straightforward gridded quadrature method applied to the definition 7. This reduction in cost is possible because the approximation 9 has replaced a joint function of $\mathbf{x}$ and $\mathbf{k}$ with a sum of products of $\mathbf{x}$ and $\mathbf{k}$ separately, and each of these can be computed using coarse grid computations by virtue of the smoothness of q. See Bao and Symes (1996); Nammour and Symes (2009) for a careful discussion of the computational cost of this algorithm.

If the goal were to calculate the action of Q given the principal symbol q, then the coefficient array $\mathbf{c}$ would have to be extracted from q by FFT and solution of a large linear system. Instead, Nammour (2009) simply reformulates the optimal pseudodifferential scaling problem to extract $\mathbf{c}$ directly from its data, via solution of

an optimization problem: given an image b, for example $b = F^T d$, find coefficients $\mathbf{c}$ to minimize

$$\|b - Q[\mathbf{c}](F^T F)b\|^2, \tag{14}$$

Problem 14 is a quadratic least squares problem, and could be solved using an appropriate numerical optimization method, for example conjugate gradient method. Note that the number of unknowns here is $(2N+1)J$, which could be much smaller than the unknowns of extended LSM: based on our experience, $N \approx O(10)$ is adequate, and J can be taken to be 2-3 orders of magnitude smaller than the number of grid points of the model b (or m). Given the work estimate mentioned above for each application of Q, the optimal pseudodifferential scaling problem is cheap to solve, compared with the iterative extended LSM.

Nammour (2009) shows that the solution of problem 14 satisfies

$$Q[\mathbf{c}]b \approx (F^T F)^\dagger b = (F^T F)^\dagger F^T d \approx \delta m, \tag{15}$$

for $b = F^T d$ with $d = F \delta m$. $(F^T F)^\dagger$ denotes the Moore-Penrose pseudoinverse (Moore, 1920).

To summarize, to compute the optimal pseudodifferential scaling,

- Compute extended migration $b = F^T d$;

- Apply normal operator to get re-migrated image $(F^T F)b$;

- Solve problem 14 to get $\mathbf{c}$;

- Compute $Q[\mathbf{c}]b \approx (F^T F)^{-1} b$.

We end this discussion by mentioning the restrictions that must hold in order that the underlying assumption of this construction, namely that $F^T F$ be a ΨDO, is valid. Smoothness of the velocity model m is essential: presence of interfaces typically implies the well-known tendency for reflectors to be imaged in multiple locations, whereas ΨDOs preserve reflector locations. For the physical (non-extended) Born operator, the necessary additional conditions are rather mild: essentially, no rays can arrive at the receiver or source arrays tangent to the surfaces on which they lie, and scattering over π (diving waves) must also be ruled out (or removed from the output of F) (Stolk, 2000). For the shot record extension, an additional and rather strict assumption must be satisfied, namely that the ray

field associated to m have no caustics in the region where $\delta m \neq 0$. This condition effectively restricts preconditioning by pseudodifferential scaling to models with mild lateral heterogeneity. Nolan and Symes (1996); Stolk and Symes (2004) thoroughly discuss the reasons for this restriction, which is related to failure of shot-record extended prestack migration to act as an imaging operator in the presence of caustics or multipathing.

Flexibly Preconditioned Conjugate Gradient Algorithm

Preconditioned conjugate gradient (CG) iteration (Nocedal and Wright, 1999; Golub and van Loan, 1996) is simply the CG algorithm formulated with an inner product different from the usual Euclidean dot product. Applied to least-squares problems such as equation 5, the effect is to alter the application of the adjoint operator (F^T) by appending another operator M. If $MF^TF \approx I$, convergence is much accelerated.

It is tempting to use the approximate inverse Q, computed by solving equation 14 as in the previous section, for M in preconditioned CG, however that does not result in accelerated convergence. The reason is that while Q satisfies equation 8 approximately, it is not generally an approximation to $(F^TF)^\dagger$ - it only approximates the action of the latter operator on a particular vector.

There is however a variant of preconditioned CG, called *flexibly preconditioned CG* (FPCG), that makes use of this type of approximate solution of the normal equation as opposed to an approximate inverse of the normal operator. The algorithm is listed explicitly in **algorithm** 2. The action of a conventional preconditioner is replaced by an approximate solution of the normal equation at each step - thus each iteration requires that Q, hence **c**, be constructed anew, and an additional application of Hessian per iteration is required. In the algorithm listing, the construction of the iteration-dependent preconditioning step is highlighted in red.

One thing to note is the formula for the direction update scalar β in algorithm 2 should be the Polak-Ribiere variant (Nocedal and Wright, 1999), as recommended by Notay (2000), who proved that as long as the variation in the preconditioner is sufficiently small, the FPCG method will have a convergence rate that is comparable with CG method with an approximate inverse of F^TF as the iteration-independent preconditioner. While the conditions for convergence developed in

---

Algorithm 2 A flexibly preconditioned conjugate gradient method

---

Given data d and $\delta m_0 = 0$, Born modeling operator F, RTM operator F^T
Given maximum iterations MaxIter and $i = 0$
$r_0 = d$, $G_0 = F^T r_0$
compute $\mathbf{c}_0$ by solving equation 14 with $b = G_0$
$g_0 = Q[\mathbf{c}_0]G_0$
$p_0 = g_0$
while (not converge && i < MaxIter) **do**
    $q_i = Fp_i$
    $\alpha_i = \dfrac{\langle g_i, G_i \rangle}{\langle q_i, q_i \rangle}$
    $\delta m_{i+1} = \delta m_i + \alpha_i p_i$
    $r_{i+1} = r_i - \alpha_i q_i$
    $G_{i+1} = F^T r_{i+1}$
    compute $\mathbf{c}_{i+1}$ by solving equation 14 with $b = G_{i+1}$
    $g_{i+1} = Q[\mathbf{c}_{i+1}]G_{i+1}$
    $\beta = \dfrac{\langle g_{i+1}, G_{i+1} - G_i \rangle}{\langle g_i, G_i \rangle}$
    $p_{i+1} = g_{i+1} + \beta p_i$
    $i = i + 1$
end while

(Notay, 2000; Knyazev and Lashuk, 2006) do not appear to be verifiable in examples, the algorithm nontheless appears to perform reasonably well, as the examples to come will show.

We note that the behavior of the algorithm may be improved by modifying the definition of F so that $F^T F$ is better approximated by a ΨDO. For example, appropriate mutes and window functions to remove diving wave energy and shallow acquisition footprint will improve the approximate solution by pseudodifferential scaling. High pass filtering may also make this approximation more accurate, since we keep only the leading order terms in frequency throughout our constructions.

NUMERICAL EXAMPLES

In this section, we show several numerical results for the FPCG method and comparisons with conjugate gradient and steepest descent methods.

Wave equations are discretized using regular grid finite-difference scheme: 2nd order accurate in time and 4th order accurate in space. We use free surface top boundary condition and place sources and receivers below the surface. The choice of free surface boundary condition and this acquisition geometry introduce two derivatives to the data: vertical derivative over the source position and vertical derivative over the receiver position (Hou and Symes, 2015b), which are approximately equivalent to $\dfrac{\partial^2}{\partial t^2}$. Thus composing the original Born modeling operator with the square of integral operator over time I_t^2 roughly undo the effect of using this free surface boundary condition and acquisition geometry, where $I_t = \displaystyle\int_t dt$.

The FPCG method used in following examples contains two steps of preconditioning: first, apply I_t^2 to both the input data and the source wavelet; second, apply Algorithm 2 to the integrated data and the integrated source wavelet.

The first two examples concern recovering a single reflector under the presence of a slow and a fast Gaussian anomaly in background model respectively. These two simple synthetic examples are used to demonstrate the behavior of the proposed algorithm. The last example is the Marmousi model with 450 m extended water

layer, which is used to illustrate the capability and highlight the superiority of the proposed method when a model has complex structures.

For each example, we show the convergence curves of solving extended LSM using FPCG method and other methods, display the inverted extended image $\delta m(\mathbf{x}, \mathbf{x}_s)$. We also show the stacked image $\sum_{\mathbf{x}_s} \delta m(\mathbf{x}, \mathbf{x}_s)$, the target data, and the data residual (the difference of the predicted and target data).

Slow Gaussian anomaly model

The slow Gaussian anomaly model is illustrated in Figures 1a and 1b. The acquisition geometry is as described in the last section. Source and 3 shots of the data are shown in Figures 3a and 3b.
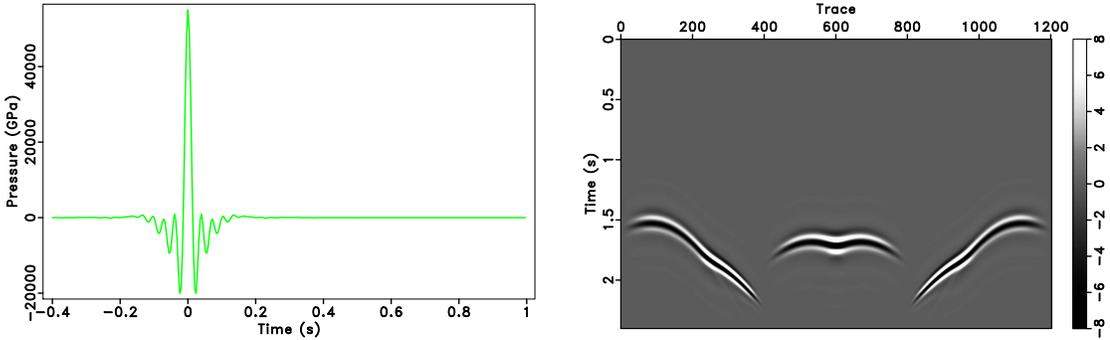

Figure 3: (a) source wavelet with corner frequency 1 Hz, 7 Hz, 28 Hz, 35 Hz, used in Gaussian anomaly examples (b) 3 of 81 Born shot records for slow Gaussian anomaly example (Figures 1a and 1b.)

Data shown in Figure 3b is used as the observed data in this example (81 shots - only 3 are shown here). Figure 4a shows the normalized data misfit (the value of the objective function in equation 4) as a function of number of Hessian applications. Figure 4b plots the normalized normal residual (the L2 norm of the difference of the right hand side and the left hand side with or without preconditioning of equation 5). CG with windowing is the conjugate gradient method with the forward map F replaced by FW, the composition of F and a windowing operator W which multiplies a smooth cutoff function that sets $\delta m(\mathbf{x}, \mathbf{x}_s)$ to

zero from the surface to depth = 200 m: this is the depth range in which imaging is dominated by acquisition footprint and large amplitude, making the Hessian more ill-conditioned. The improvement of conditioning resulting from windowing out these effects is evident in the convergence plots, but is not sufficient to bring the residual below 40% of its initial value in 20 iterations. Unwindowed CG converges even more slowly.
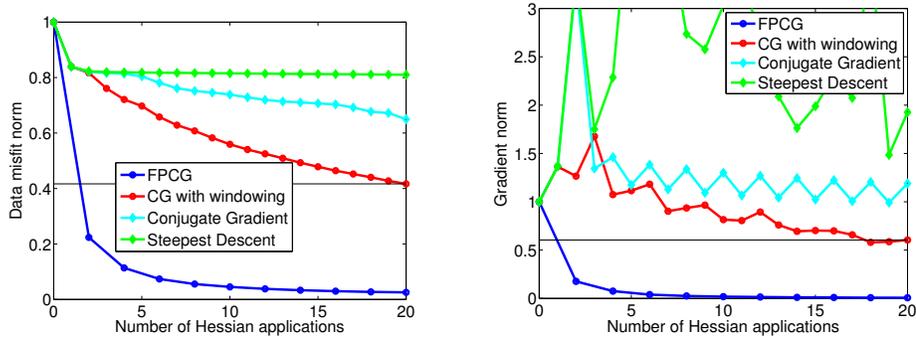

Figure 4: Slow Gaussian anomaly example: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.

In contrast, FPCG (Algorithm 2, without windowing or muting) achieves 4% of the initial residual RMS in 10 iterations. Note that each iteration of FPCG requires a second application of the Hessian, hence is twice as expensive as each iteration of CG. The convergence plots are arranged to take this expense differential into account: the horizontal axis unit is Hessian applications. In particular, note that FPCG with 2 Hessian applications fits the data as well as CG with 20 Hessian applications (and depth windowing as described above, Figure 4a). In this example, FPCG is the only method to approximate a stationary point of the least squares objective, reducing the gradient length to a small fraction (0.8%) of its original value in 20 iterations (Figure 4b).

Figures 5a and 5b show the inverted extended images or extended model perturbations, the solutions of equation 5 using FPCG and CG method with windowing. They are 2D views of 3D volumes. The physical model axes are labeled as "Distance" and "Depth". The third axis without label is the shot record. Figures 6a, 6b, 7a and 7b compare the stacked images and data residuals of FPCG method and CG method with windowing. Figures 5a and 5b and stacked images in Figures 6a and 6b show that the frequency contents in the final images are quite different with and without preconditioning. Fewer artifacts are observed in the image

Figure 5: Slow Gaussian anomaly example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Figure 6: Slow Gaussian anomaly example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Figure 7: Slow Gaussian anomaly example: data residual, same 3 shots as in Figure 3b after 20 Hessian applications using FPCG (a), and using CG with windowing (b).

from FPCG method (Figures 5a and 6a). It is obvious from Figures 7a and 7b that preconditioning improves the data fitting greatly as well as helps to attenuate unwanted noise.

Fast Gaussian anomaly model

The fast Gaussian anomaly model is shown in Figure 8a. The acquisition geometry, source wavelet, and the true reflectivity model are the same as the slow Gaussian anomaly model example. 3 shots of the data are shown in Figures 8b. Very similar convergence behavior (Figures 9a and 9b) as in the slow Gaussian anomaly model example is observed in this set of tests. See Figures 10, 11 and 12 for the comparison of inverted images, stacked images and data residuals by FPCG method and CG method.

Marmousi model

To evaluate the potential of the FPCG algorithm to recover reflectivity with a wide range of dips and strengths, we apply it to the a model derived from the Marmousi model (Versteeg and Grau, 1991). We smoothed the model and added a 450 m deep water layer to produce the smooth background model shown in Figure 13a.
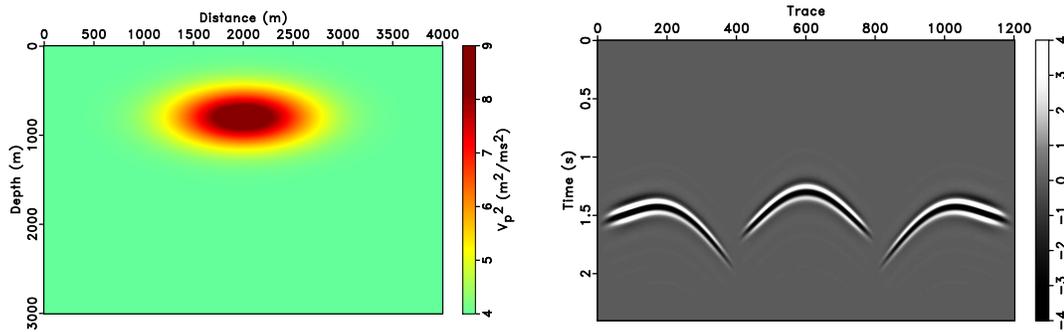
Figure 8: Fast Gaussian anomaly example: (a) background model, and (b) 3 of 81 Born shot records.

Figure 9: Fast Gaussian anomaly example: convergence curves of numerical methods for (a) normalized data misfit and (b) normalized gradient length.
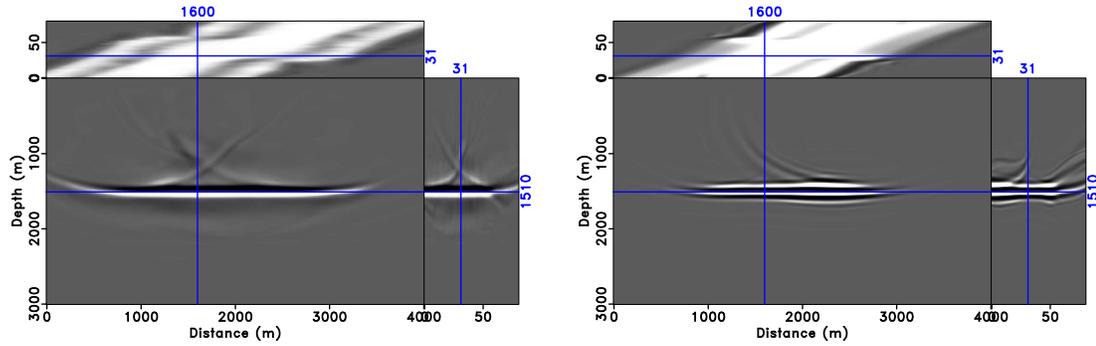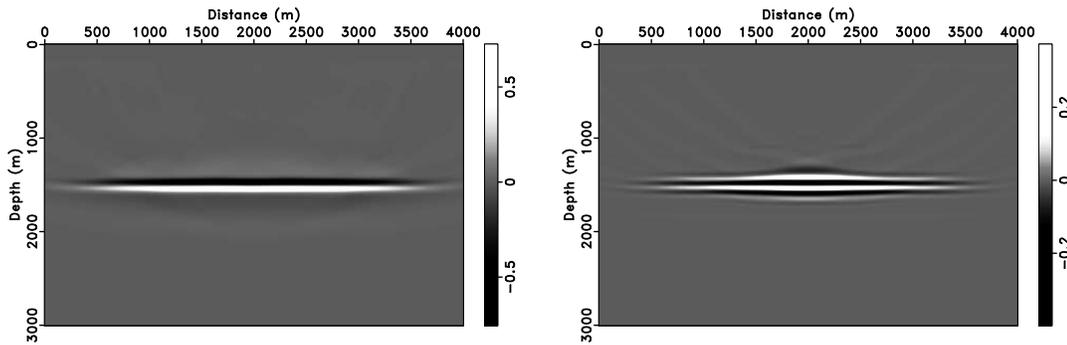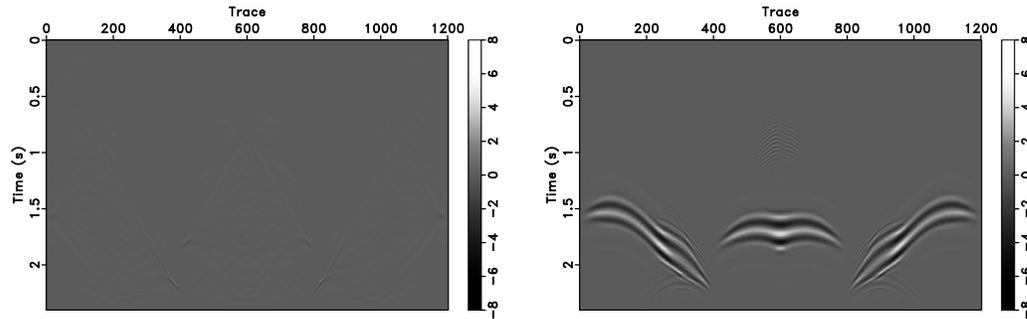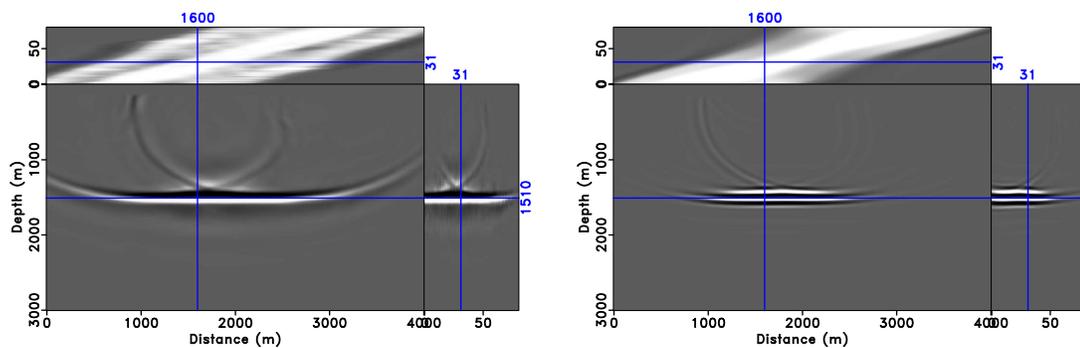

Figure 10: Fast Gaussian anomaly example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).
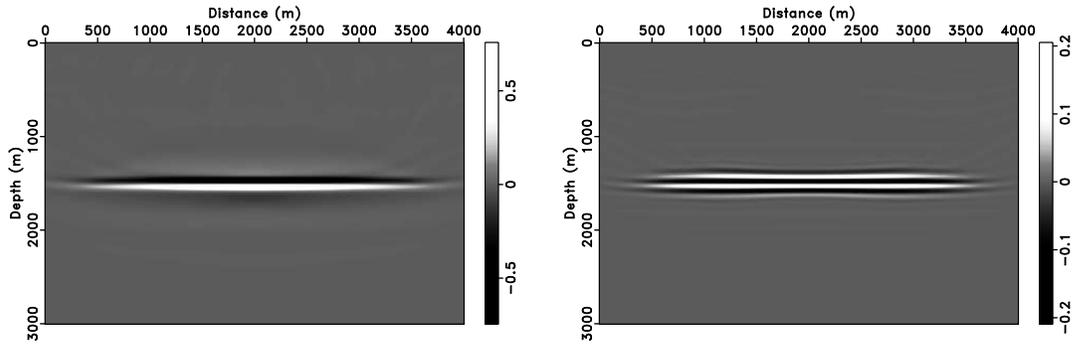
Figure 11: Fast Gaussian anomaly example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).
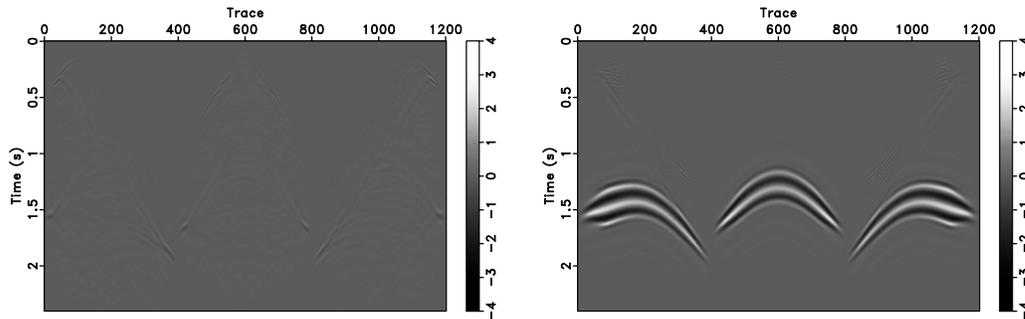
Figure 12: Fast Gaussian anomaly example: Data residual after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

The difference between the original and a slightly less smoothed model serves as the model perturbation or reflectivity field (Figure 13b). The model was gridded at 8 m in both directions, and a finite difference method used to produce Born data (Figures 14a, 14b). The source wavelet was a zero-phase bandpass filter with corner frequencies 1 Hz, 6 Hz, 25 Hz, 30 Hz. 128 point sources were located at depth 8 m with 48 m spacing. 1151 fixed receivers with 8 m spacing were placed at depth 6 m. Final recording time was 3.4 s.
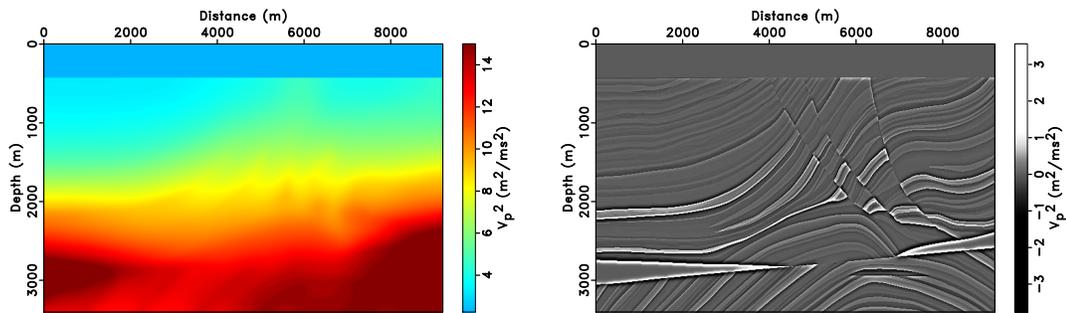
Figure 13: Marmousi example: (a) background model, (b) model perturbation.
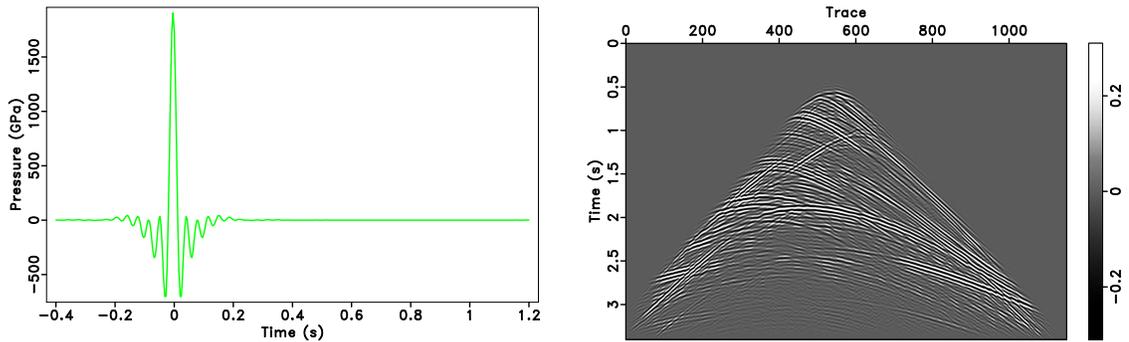
Figure 14: Marmousi example: (a) source wavelet with corner frequencies = 1 Hz, 6 Hz, 25 Hz, and 30 Hz. (b) Born shot record, shot index 65.

Although we use Born modeling to generate the observed data, diving waves kinematics are still present since diving waves generated by the background model are also perturbed. We have muted the diving waves, both in the target data and by composing the modeling operator with a the same mute.

As in the previous examples, we used both CG with windowing and FPCG to invert this data. The windowing operator in this case zeroes the first 450 m in
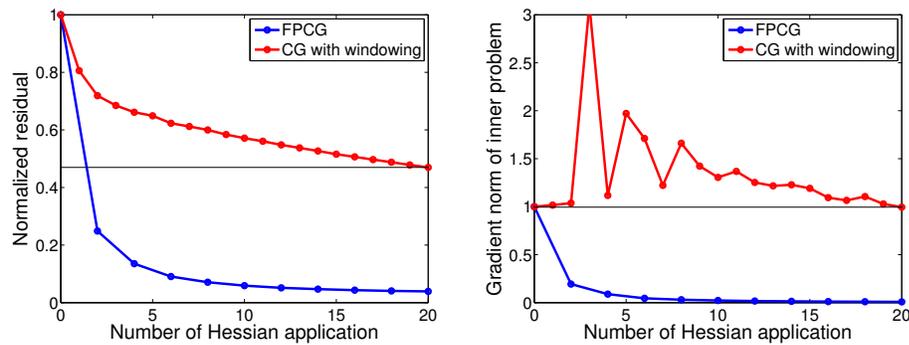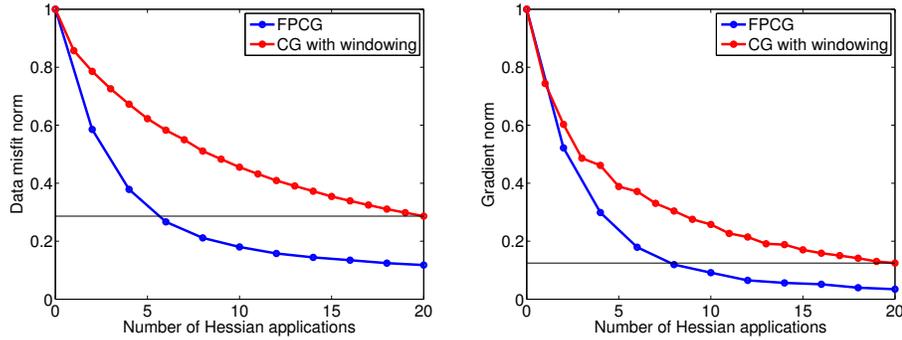
Figure 15: Marmousi example: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.

depth, with a smooth taper to avoid introducing high-frequency artifacts into the inversion. Figure 15a shows the normalized data misfit as a function of Hessian applications. Figure 15b plots the normalized normal residual or the gradient of the CG and FPCG method. The FPCG method with 6 Hessian applications fits the data better than the CG method with 20 Hessian applications (Figure 15a), similarly for the gradient length.
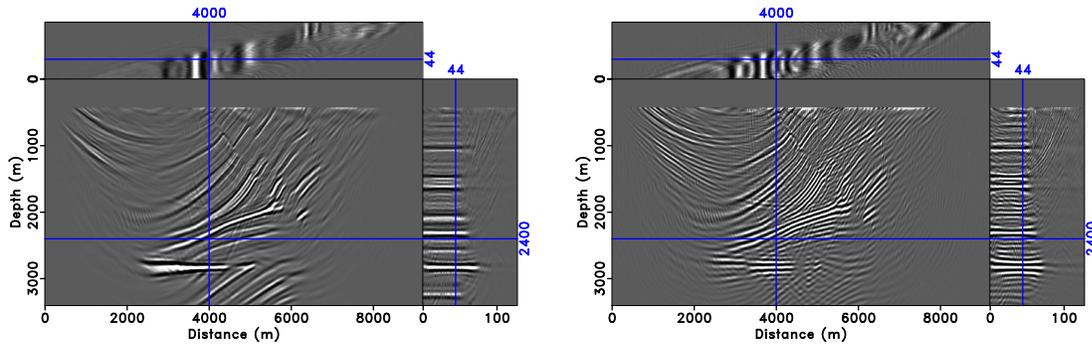
Figure 16: Marmousi example: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Figures 16a and 16b show the inverted extended model perturbations produced by these methods. The FPCG result (Figure 16a) shows both more accurate amplitudes and better forused wavelet than does the non-preconditioned CG result (Figure 16b). The superiority of the FPCG method is especially evident in

the gather panel: the FPCG-produced events are better separated and flatter, with fewer artifacts contaminating the image volume, than are the CG-produced events.
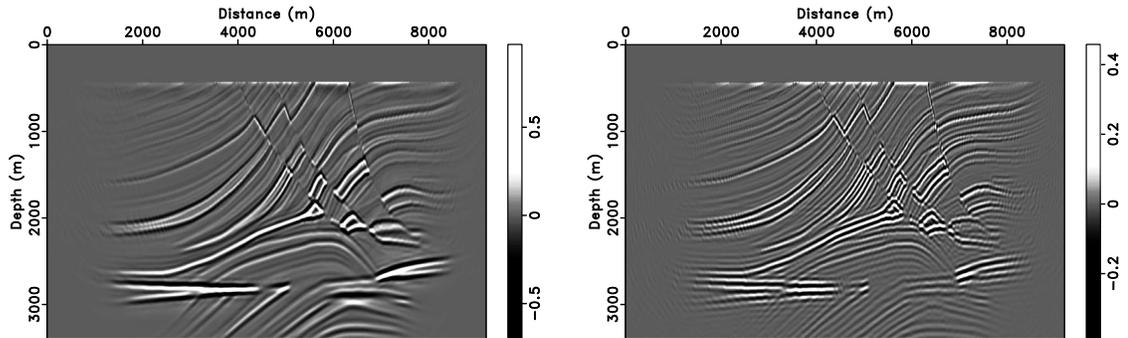
Figure 17: Marmousi example: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

Figures 17a and 17b compare the stacked version of the inverted extended images of FPCG method and CG method with windowing. We see that preconditioning is effective in mitigating artifacts and producing a better balanced amplitude in the image,as well as improving the resolution of the image especially in the deep part (below 2.5km).

Figure 18a compares the well logs at $x = 4000$ m of inverted model perturbations from the 44-th shot data by FPCG and CG with windowing methods with the true reflectivity model. Both FPCG and CG with windowing recover the amplitude of the shallow part of the model, while FPCG method restores relatively better the amplitude of the deep part of the model than the non-preconditioned method. The spectra of traces of inverted images corresponding to shot 44 are shown in Figure 19a. It is clear that for the wavenumber range that could be explained by these methods (between 0.003 1/m and 0.013 1/m), the amplitude of the inverted image by FPCG method resembles that of the target image.

Since the LSM is in the shot record domain, the inverted image volume minimizes the data residual, while the stack of the inverted image does not, although the stacked image contains a lot of useful informations as well. That is why after stacking, the amplitude does not match the true reflectivity as shown in Figure 18b, but the amplitude from FPCG method is still relatively closer to the true one than that from CG with windowing at the deep part of the model (below
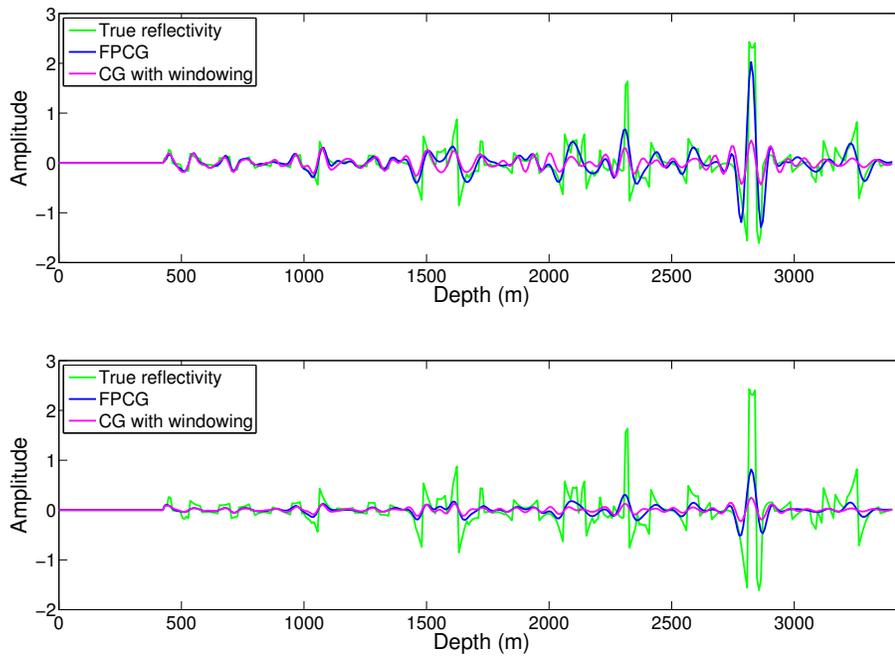
Figure 18: Marmousi example: Vertical trace comparison at $x = 4000$ m of target model perturbation with inverted model perturbation cube from the 44-th shot (a) and with stacked inverted image (b) by FPCG and CG with windowing methods after 20 Hessian applications.
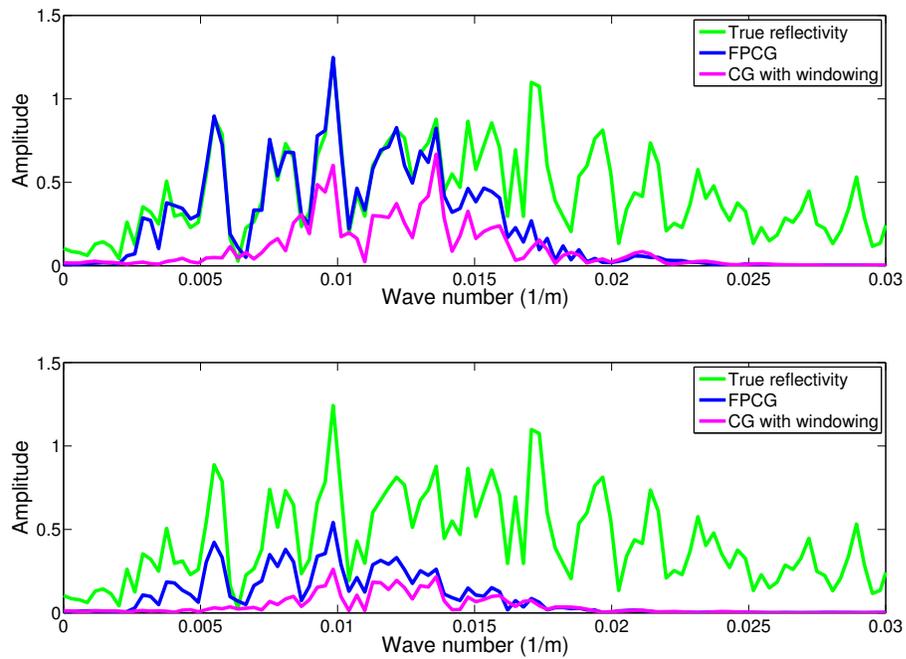
Figure 19: Marmousi example: Spectrum comparison of the vertical trace shown in Figure 18 of target model perturbation with inverted model perturbation cube from the 44-th shot (a) and with stacked inverted image (b) by FPCG and CG with windowing methods after 20 Hessian applications.

2500 m). The corresponding spectra comparison of traces of stacked images are shown in Figure 19b. Note that the traces and spectra of stacked images are for visualization purposes and have no quantitative meaning.
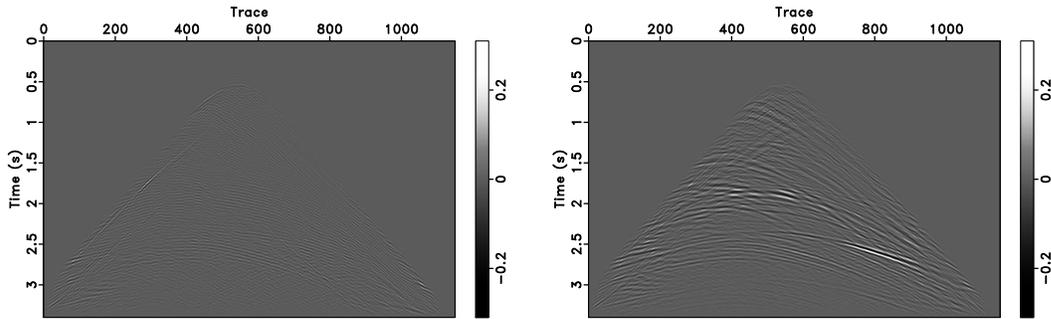
Figure 20: Marmousi example: data residual, same shot record as in Figure 14b, after 20 Hessian applications using FPCG (a), and using CG with windowing (b).

Data residuals from FPCG method and CG method with windowing are shown in Figures 20a and 20b respectively. With preconditioning, data residual is small everywhere, while the non-preconditioned CG is struggling to fit the deep part (around and below 2s) of the data. We also see steeply-dipping events in the residual from FPCG method, which is better modeled by the non-preconditioned CG method. These events are close to diving wave perturbations, and are poorly explained by the pseudodifferential scaling operator.
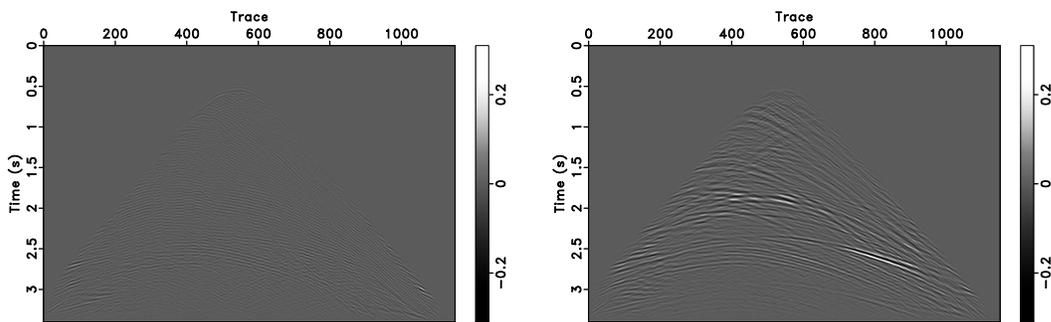
Figure 21: Marmousi example: data residual after dip filtering with 20 Hessian applications using FPCG (a), and using CG with windowing (b).

Since these straight-line events are mainly due to refracted waves, with a particular apparent velocity, they can be suppressed through dip filtering. Figures 21a

and 21b show the same data residual, but after application of a dip filter with cut-off velocity here is 2500 m/s, so that events with slopes larger than 0.0004 s/m are excluded. Figures 21a and 21b show the data residual after application of the dip filter.
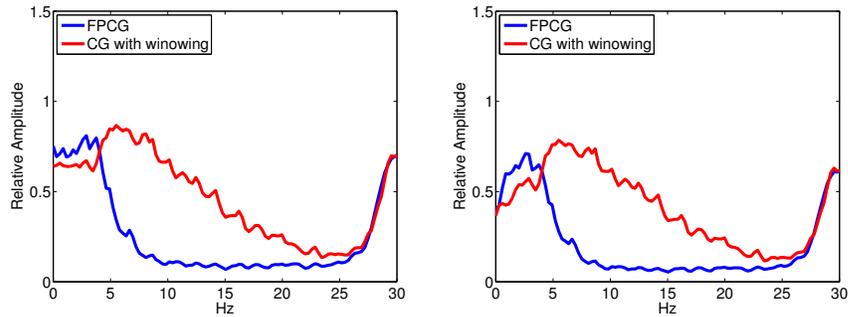
Figure 22: Marmousi example: Relative power spectra of data residual (a) and data residual after dip filtering (b) with respect to observed data.

The relative data residual power spectrum gives another way of appreciating the effect of preconditioning. FPCG method suppresses error regardless of frequency components, while CG method fits different frequency components in the data differently, as can be observed in Figure 22a. The error spectra with dip filter (Figures 22b) are slightly lower in amplitude than those without dip filter.

One advantage cited earlier for extended Born modeling is ability to fit data even with incorrect velocity. The final example of this section illustrates this property. The smooth background model depicted in Figure 23 is obtained by adding 70% of the target model in Figure 13a and 30% of the constant model $m = 2.25$ m^2/s^2. Note that the background model is scaled, thus the reflection coefficients at the ocean bottom is different from the target model and the refracted waves generated by the target model will not be predicted correctly. Perturbations of these refracted waves form a part of the data that could not be fit using shot record domain LSM, preconditioned or not, which is a feature of shot record domain model extension. However, for the part of data that could be fitted, using shot record domain model extension, preconditioning will still significantly improve the convergence rate. We show the convergence curves for inversion of the Marmousi-Born data (Figure 14b) with this incorrect velocity model in Figure 24a. As was the case with correct velocity, a reasonably good fit is obtained, with most of the remaining residual energy located near the diving wave perturbations, and the

FPCG iteration is considerably more efficient than the CG iteration, by about the same factor as for the correct velocity case.
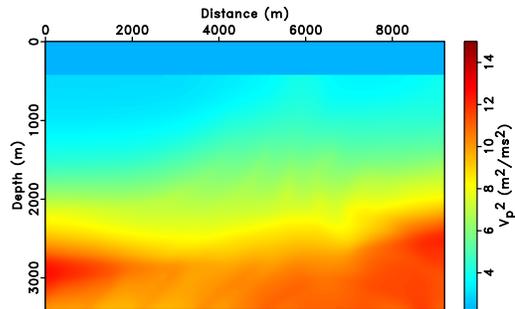
Figure 23: Marmousi example, incorrect background model
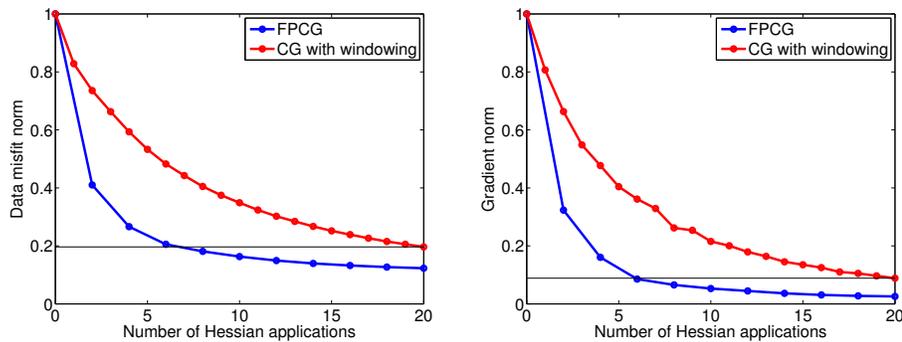
Figure 24: Marmousi example, incorrect background model: convergence curves of numerical methods, (a) normalized data misfit and (b) normalized gradient length.

Figures 25a and 25b show the inverted extended model perturbations. The gather information is more valuable than the single shot image in this example, since the background model is wrong. We see a better focused gather with preconditioning (Figure 25a) compared with non-preconditioning (Figure 25b). The curvatures in image indicate that the background model is smaller than the true model and thus could be used to update the background model.

Figures 26a and 26b compares the stacked images of FPCG method and CG method with windowing at wrong background model (as shown in Figure 23). We see that
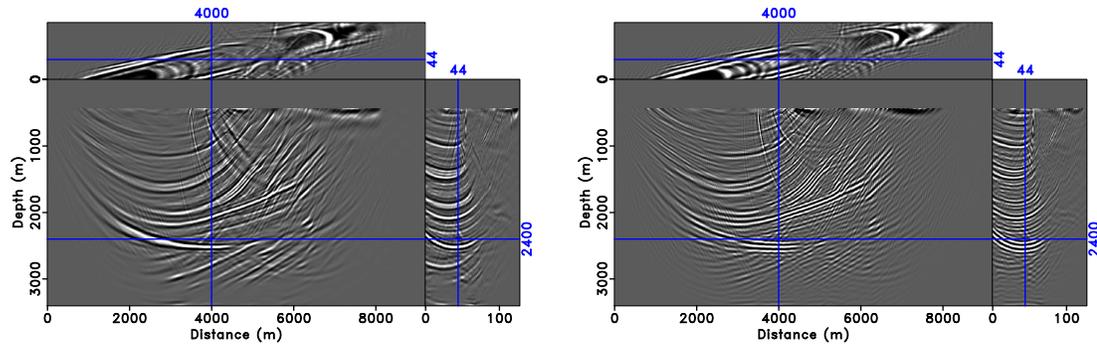
Figure 25: Marmousi example, incorrect background model: inverted model perturbation cube after 20 Hessian applications using FPCG (a) and using CG with windowing (b).
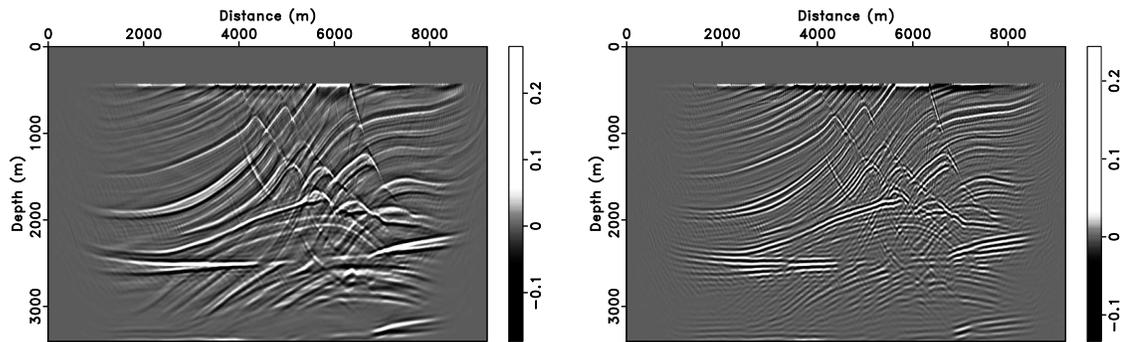
Figure 26: Marmousi example, incorrect background model: Stacked image after 20 Hessian applications using FPCG (a) and using CG with windowing (b).

the preconditioning is still helpful in mitigating artifacts in the image, and produces better balanced amplitude as well as improves the resolution of the image, although the reflector positions are wrong.
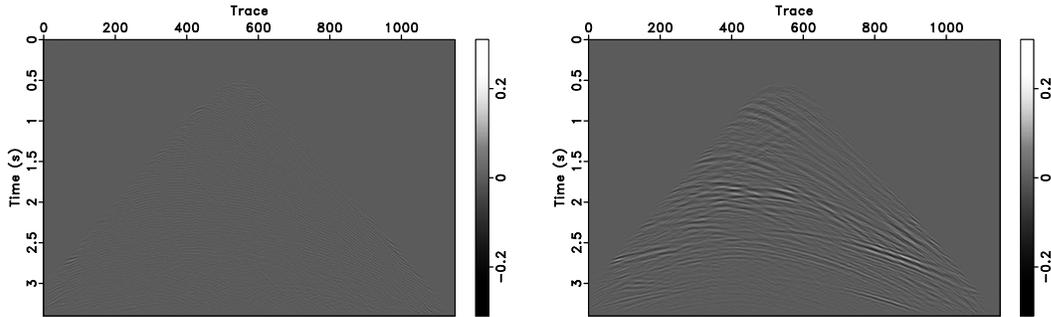
Figure 27: Marmousi example, incorrect background model: data residual, same shot record as in Figure 14b, after 20 Hessian applications using FPCG (a), and using CG with windowing (b).

Finally, the data residuals depicted in Figures 27a and 27b, using the same grey scales as in Figures 20a and 20b, confirm the ability of shot record extended Born modeling to fit data very well even with incorrect velocity. Once again, the FPCG residual is considerably smaller than the CG residual, even though the same computational effort has been expended in producing both.

DISCUSSION

The acceleration technique presented here suffers from some strict limitations on its validity, noted in the section on pseudodifferential scaling: for example, it depends on more or less unique energy paths between sources, receivers, and scattering points. However, within the domain of validity - mild lateral velocity variation and slow changes in spatially averaged index of refraction - it provides an inexpensive and effective way to accelerate iterative LSM.

We have examined only the simplest relevant wave physics in this paper, namely constant density acoustics. However, the foundations have been laid long since for extension to variable density acoustics and indeed to elasticity (Beylkin and Burridge, 1990; Burridge et al., 1998; De Hoop and Bleistein, 1997) This extension

amounts to a rational version of amplitude versus offset (AVO) analysis, with illumination based amplitude artifacts have been removed (Tarantola, 1984a; Lörtzer and Berkhout, 1989; Castagna, 1993; Luh, 1993). Accelerated convergence in these settings can in principle be achieved using similar ideas - for some first steps see Nammour and Symes (2011).

A fast extended LSM available could be used in a form of optimization based migration velocity analysis based on Born inversion (Liu et al., 2014; Symes, 2015; Huang and Symes, 2015). In fact, inversion (as opposed to migration), with its improved amplitude and especially phase, is actually essential for reliable functioning of such objective velocity analysis algorithms (Kern and Symes, 1994; Lameloise et al., 2015).

CONCLUSION

We presented a flexibly preconditioned conjugate gradient algorithm with pseudodifferential scaling preconditioner to improve the convergence rate of shot record extended least squares migration method. Flexible preconditioning differs from the usual preconditioning using a fixed operator, in that the preconditioner is computed anew at each step. Scaling preconditioning, costing one additional application of the Hessian per iteration, requires the use of an iteration-dependent preconditioner. We have shown how such a an algorithm can be based on a fast algorithm for application of pseudodifferential operators.

Although the computational cost associated with one image update via FPCG method is effectively twice that of ordinary CG method per step, the FPCG method has been shown to converge much faster. The faster convergence more than compensates for the increased cost per step: FPCG produces more accurate inverted extended model perturbations, with better amplitudes and more focused events, than does CG, for the same computational effort.

ACKNOWLEDGMENTS

REFERENCES

Aoki, N., and G. Schuster, 2009, Fast least-squares migration with a deblurring filter: Geophysics, **74**, no. 6, WCA83–WCA93.

Bao, G., and W. Symes, 1996, Computation of pseudo-differential operators: SIAM J. Sci. Comput., **17**, 416–429.

Baysal, E., D. D. Kosloff, and J. W. C. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

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.

Bourgeois, A., B. Jiang, and P. Lailly, 1989, Linearized inversion: A significant step beyond pre-stack migration: Geophysical Journal International, **99**, 435–445.

Boyd, S., and L. Vandenberghe, 2004, Convex Optimization: Cambridge University Press.

Burridge, R., M. V. De Hoop, D. Miller, and C. Spencer, 1998, Multiparameter inversion in anisotropic elastic media: Geophysical Journal International, **134**, 757–777.

Castagna, J. P., 1993, AVO analysis - Tutorial and review, *in* Offset Dependent Reflectivity - Theory and Practice of AVO Analysis: SEG, 1, 3–36.

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

Claerbout, J. F., and D. Nichols, 1994, Spectral preconditioning: Technical Report 82, Stanford Exploration Project, Stanford University, Stanford, California,

USA.

De Hoop, M. V., and N. Bleistein, 1997, Generalized Radon transform inversions for reflectivity in anisotropic elastic media: Inverse Problems, **16**, 669–690.

Demanet, L., P. Létourneau, N. Boumal, H. Calandra, J. Chiu, and S. Snelson, 2012, Matrix probing: a randomized preconditioner for the wave-equation Hessian: Applied and Computational Harmonic Analysis, **32**, 55–168.

Dutta, G., Y. Huang, W. Dai, X. Wang, and G. T. Schuster, 2014, Making the most out of the least (squares migration): 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4405–4410.

Golub, G. H., and C. F. van Loan, 1996, Matrix computations, 3rd ed.: Johns Hopkins University Press.

Guitton, A., and B. Kaelin, 2006, Least squares attenuation of reverse time migration artifacts: 76th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SPMI 1.6.

Herrmann, F., P. Moghaddam, and C. Stolk, 2008, Sparsity- and continuity-promoting seismic image recovery with curvelet frames: Applied and Computational Harmonic Analysis, **24**, 150–173.

Herrmann, F. J., C. R. Brown, and Y. A. Erlangga, 2009, Curvelet-based migration preconditioning and scaling: Geophysics, **74**, A41–A46.

Hou, J., and W. Symes, 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.

Huang, Y., and W. 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.

Knyazev, A. V., and I. Lashuk, 2006, Steepest descent and conjugate gradient methods with variable preconditioning: SIAM J. Matrix Anal. & Appl., **29(4)**, 1267–1280.

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 Ab-

stracts, European Association of Geoscientists and Engineers.

Loewenthal, D., and I. Mufti, 1983, Reversed time migration in spatial frequency domain: Geophysics, **48**, 627–635.

Lörtzer, G. J. M., and A. J. Berkhout, 1989, Linear AVO inversion of multicomponent seismic data: 59th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 967–972.

Luh, P. C., 1993, Wavelet attenuation and AVO, *in* Offset Dependent Reflectivity - Theory and Practice of AVO Analysis: SEG, 3, 190–198.

Luo, S., and D. Hale, 2014, Least-squares migration in the presence of velocity errors: Geophysics, **79**, S153–S161.

Moore, E. H., 1920, On the reciprocal of the general algebraic matrix: Bulletin of the American Mathematical Society, **26**, 394–395.

Mulder, W. A., and R.-E. Plessix, 2004, A comparison between one-way and two-way wave equation migration: Geophysics, **69**, 1491–1504.

Nammour, R., 2009, Approximate inverse scattering using pseudodifferential scaling: Technical Report 09-09, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

Nammour, R., and W. Symes, 2009, Approximate constant density acoustic inverse scattering using dip-dependent scaling: 80th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SI 3.6.

Nammour, R., and W. W. Symes, 2011, Multiparameter inversion: Cramer's rule for pseudodifferential operators: International Journal of Geophysics, **2011**, 780291 (12 pp). (DOI:10.1155/2011/780291).

Nemeth, T., C. Wu, and G. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

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.

Notay, Y., 2000, Flexible conjugate gradients: SIAM J. Sci. Comput., **22(4)**, 1444–1460.

Pan, W., K. Innanen, and M. G., 2014, A comparison of different scaling methods for least-squares migration/inversion: 76th Annual International Conference and Exhibition, Expanded Abstract, European Association for Geoscientists and Engineers, WE G103 14.

Rakesh, 1988, A linearized inverse problem for the wave equation: Communications on Partial Differential Equations, **13**, 573–601.

Rickett, J. E., 2003, Illumination-based normalization for wave-equation depth

migration: Geophysics, **68**, 1371–1379.

Shin, C., S. Jang, and D.-J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

Stolk, C. C., 2000, Microlocal analysis of a seismic linearized inverse problem: Wave Motion, **32**, 267–290.

Stolk, C. C., M. V. de Hoop, and T. P. M. Op 't Root, 2009a, Reverse time migration inversion from single-shot data: 79th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2995–2999.

Stolk, C. C., M. V. de Hoop, and W. Symes, 2009b, 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.

Symes, W., 2008a, Approximate linearized inversion by optimal scaling of prestack depth migration: Geophysics, **73**, no. 2, R23–35.

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

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

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian: Geophysics, **74**, no. 6, WCA95–WCA107.

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.

Taylor, M., 1981, Pseudodifferential Operators: Princeton University Press.

ten Kroode, F., D. Smit, and A. Verdel, 1998, A microlocal analysis of migration: Wave Motion, **28**, 149–172.

Valenciano, A., S. Lu, N. Chemingui, and J. Yang, 2015, High resolution imaging by wave equation reflectivity inversion: Presented at the 77th Annual International Conference and Exhibition, Expanded Abstract, European Association for Geoscientists and Engineers.

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

Zhang, Y., G. Zhang, and N. Bleistein, 2003, True amplitude wave equation migration arising from true amplitude one-way wave equations: Inverse Problems, **19**, 1113–1138.

Scattering and dip angle decomposition in relation with subsurface offset extended wave-equation migration

Raanan Dafni and William W. Symes

ABSTRACT

An angle-dependent reflection coefficient is recovered by seismic migration in the angle-domain. We propose a post-migration technique for computing scattering and dip angle common-image gathers (CIGs) from seismic images, extended by the subsurface offset, in relation with 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 post-migration subsurface offset image is employed to decompose scattering and dip angle CIGs individually, or to decompose a multi-angle CIG by showing simultaneously both angles on the gathers axis. A unique dip-angle response of seismic reflections is introduced as 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 examples demonstrate 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.

Note: The full text of this paper part of the 2015 TRIP annual report:

`http://www.trip.caam.rice.edu/reports/2015/trip2015_report.html`

Joint Model and Minimal-Source Full Waveform Inversion for Seismic Imaging Under General Anisotropic Sources (Thesis Proposal)

Mario J. Bencomo

71

ABSTRACT

Accurate representation and estimation of seismic sources is essential to the seismic inversion problem. General sources can be approximated by a truncated series of *multipole point-sources*, depending on the finiteness and anisotropy of the source. Unfortunately the number of terms in the multipole expansion needed to accurately represent a given source is problem dependent and not known a priori in general. Most research in joint determination of source and medium/model parameters assume isotropic point-sources resulting in an inability to fit the anisotropy observed in data, ultimately impacting the recovery of model parameters.

In my work I propose a novel method of joint model-source inversion that handles general sources by incorporating two methodologies: *Occam's inversion*, also referred to as *minimum structure inversion*, and the *variable projection (VP) method*. Occam's inversion regularizes ill-conditioned parameter estimation problems by minimizing a complexity criterion subject to a data fit constraint. In my approach I represent the seismic source as a series of multipole point-sources with minimal number of terms, hence minimal-source, satisfying certain data misfit criteria as model parameters are updated. The VP method solves nonlinear least squares problems with separable structure, i.e., residual function is a linear combination of nonlinear functions. My proposed algorithm incorporates VP for solving the joint model-source *full waveform inversion (FWI)* problem, where source parameters are linear with respect to data residual. The approach I propose here does not claim to solve the difficulties associated with model inversion but instead seeks to preserve the recoverability of model parameters under unknown anisotropic sources. Thus, difficulties associated with model estimation will be minimized by choosing starting models correct enough for FWI to produce decent results; predicted traveltimes from starting models are at most half a wavelength off from the data.

In this proposal I present some numerical examples showing that convergence rates for finite difference (FD) solvers can be preserved in the presence of singular source terms, mainly spatial delta functions and its derivative. I also give some joint model-source inversion results for simple 2-D constant-density acoustic media with assumed isotropic source term.

My proposed work will consist of implementing the model and minimal-source inversion algorithm, and related forward solvers (e.g., elasticity FD solvers), as part of the seismic inversion software package IWAVE. I will investigate the interplay between source and medium anisotropy that may result in potential source-model ambiguity, in particular for the case of *vertical transversally isotropic (VTI)* elastic media. I will also investigate the sensitivity of my proposed inversion algorithm to modeling errors for controlled synthetic and field test cases, primarily inverting data from an elastic medium modeled with acoustics.

INTRODUCTION

Accurate representation and estimation of seismic sources is essential to the seismic inversion problem. I propose a novel method for joint model-source parameter estimation that handles general sources with unknown anisotropy. Joint recoverability of model-source parameters, with an emphasis on the interplay between anisotropic medium and general anisotropic sources, is the primary interest of this work and will be explored under some reasonable assumptions. The approach I propose here does not claim to solve the difficulties associated with model inversion but instead seeks to preserve the recoverability of model parameters under unknown sources.

Motivation

Seismic data in the form of seismograms contain information about the source and receiver responses as well as properties of the Earth's medium, the latter being of priority in exploration seismology. Conventional methodology focused on removing the source dependency of seismograms as a preprocessing step in seismic imaging by estimating the source time signature through statistical methods like predictive deconvolution Robinson (1957) and homomorphic deconvolution Ulrych (1971). Ziolkowski has criticized these statistical methods for imposing unrealistic, and at times theoretically unjustifiable, constraints on both the source and medium, yielding arguable results vulnerable to subjectivity Ziolkowski (1991). Alternatively, source time signatures for the Vibroseis™ and air-gun array sources have been estimated using near field measurements with some success, see Ziolkowski (1991) and Landrø and Sollie (1992) . Source estimation via near field measurements will depend on how the source is modeled (i.e., represented mathematically) and the near-field data's dependency on the medium, which can be unknown in applications where the source-receiver path is partially submerged in unknown medium or the direct arrival cannot be cleanly separated from the data.

Starting in the late 1980's, efforts in decoupling source-medium interactions in the context of seismic imaging shifted from removal of the source in seismic traces to the joint determination of source and medium parameters. Early attempts focused on proving theoretically the co-determinability of source time-

dependecy and model parameters from reflectivity data under simplifying assumptions: acoustic layered models, quasi-impulsive and non-impulsive sources, primaries only data, inverting for only one model parameter (reflectivity); Ramm (1985),Lewis (1989),Bube et al. (1988),Minkoff and Symes (1995). Later works have implemented and tested the feasibility of joint model and source inversion in a variety of synthetics and field data, testing the limits on the assumptions of joint model-source recoverability theory and going beyond in some cases, e.g., Minkoff and Symes (1997), Wang et al. (2009), Zhou et al. (1997).

Despite some successful inversions, the joint determination of source and model parameters under more realistic and general model assumptions can be difficult and in some cases impossible as demonstrated by Delprat-Jannaud and Lailly (2005). A simple 1-D isotropic acoustic medium is considered while inverting for an isotropic point source and a "not so simple" impedance profile, assuming velocity is known. Errors due to incorrect source time-dependency (i.e., wrong phase, amplitude, and time shift) were shown to lead to accumulating errors in depth. Moreover, source-model ambiguity is shown to be significant for both constant and varying velocity cases, in which authors concluded that codetermination of both source and model is infeasible due to high sensitivity and ambiguity of model parameters with respect to the source time-dependency. The authors highlight a fundamental limitation of joint source-model inversion but also, from an optimistic point of view, motivate the further study of model-source ambiguities in order to overcome them. More importantly, their work emphasizes just how detrimental errors in the source can be to model inversion, or conversely how important a "correct" source is model inversion.

Accurate estimation of seismic sources raises the more fundamental question of source representation. All of the works mentioned above, with the exception of Minkoff and Symes (1997), idealize the source contribution to that of an isotropic source with spatial point support. Assuming that sources are essentially of point-support is justified by the fact that the spatial dimension of seismic sources considered in exploration seismology are considerably smaller (typically by a magnitude) than the propagating wavelengths of seismic waves. The isotropy assumption however is questionable, particularly when the source is known or potentially expected to have an anisotropic radiation pattern. A perfect example is that of an array of air-guns in marine seismic surveying where the effective radiation pattern of the source is known to exhibit vertical directivity. As a matter of example, Minkoff and Symes (1997) demonstrate the importance of accounting for

anisotropy in the airgun array source for marine reflection field data where both reflectivity and source terms were inverted for. Their results show that inverting for anisotropic source terms, as oppose to using a given modeled source or even inverting for an isotropic source, allows them to account for 25% more of the data and were able to achieve 90% data fit up to a gas-sand target. Moreover recovered medium parameters matched closely expected lithology, but only with an isotropic source representation. Active work in seismic anisotropic source inversion is primarily focused on the estimation of earthquake mechanisms, and recently microseismic events (small earthquakes) resulting from hydraulic fracturing. Sources are typically assumed to be point double couples (DC) as to approximate the radiation pattern exhibited by a slipping fault. I refer to Aki and Richards (2002), Shearer (2009), Jost and Herrmann (1989), Julian et al. (1998) for an introductory discussion on this topic. Higher order *multipole* representations have also been considered in cases where finiteness of source (i.e., size of fault) leads to anisotropy unaccounted for in a DC model; Li et al. (2006), Stump and Johnson (1982).

General sources of finite volume can be approximated by a truncated series of *multipole point-sources* depending on the finiteness and anisotropy of the source; this is the concept of the *multipole expansion* of sources commonly used in electromagnetism. Authors in Santosa and Symes (2000) analyze and quantify the accuracy of multipole point-source approximations for the context of acoustic sources using a combination of far-field and spherical harmonic expansions. Unfortunately, the number of terms in the multipole expansion needed to accurately represent a given source is problem dependent and not known a priori for the types of applications considered here. In Minkoff and Symes (1997) the number of source terms needed was derived by trial by error during their preliminary studies, where fewer terms would have lead to unexplained amplitude-versus-angle behavior in the data.

In my work I seek to incorporate the importance of an accurate source representation for fitting seismic data in the context of seismic imaging, in hopes that it will lead to better inversion results for model parameters. I propose a joint source-model inversion algorithm that will automate, in a sense, the trial and error procedure from Minkoff and Symes (1997) as a means for handling generally sources, where no structure on the source representation is assumed. My algorithm is based on *Occam's inversion*, also referred to as *minimum structure inversion*: higher order multipole point-source terms are systematically introduced

in the source representation, under the condition that the data can be better explained. In other words, the algorithm is biased towards more simplistic/minimal sources of lower multipole terms and only considers adding complexity if it benefits in driving the data residual down enough.

An added bonus of using a multipole point-source representation is that it yields a linear relationship between source parameters and data, mainly through convolution in time with an appropriate Green's function. The linearity of source parameters qualifies the joint source-model inversion problem for separable nonlinear programming methods, primarily that of the *variable projection* (VP) method developed by Golub and Pereyra (1973). The VP method reduces the nonlinear LS problem, resulting from an inversion formulation, by essentially eliminating the source parameters and redefining the objective function in terms of model parameters only. In practice, VP methods, coupled with Gauss-Newton type algorithms for solving nonlinear LS problems, have been proven to outperform their non-reduced counterparts in some cases, Ruhe and Wedin (1980). My proposed algorithm will also incorporate the VP method as a means to alleviate the difficulty of joint source-model inversion.

Claim

Accurate representation and estimation of seismic sources is essential to the seismic inversion problem where the goal is to recover medium parameters from recorded seismograms. Most work in joint determination of model and source parameters assumes that the structure of the source is given a priori, a practical but unrealistic assumption in some applications. I propose the method of *model and minimal source estimation* for handling the joint source-model inversion problem under general, possibly anisotropic, point-sources. My proposed algorithm will incorporate the *variable projection* (VP) method to mitigate the difficulties of joint source-model inversion, as well as Occam's inversion for recovering the best minimal source that explains the data sufficiently well.

My proposed work will consist of implementing the model and minimal-source inversion algorithm, and related forward solvers (e.g., elasticity FD solvers), as part of the seismic inversion software package IWAVE. Joint recoverability of model-source parameters, in particular for anisotropic medium and general sources, is the primary interest of this work and will be explored under some reasonable

assumptions. The approach I propose here does not claim to solve the difficulties associated with model inversion but instead seeks to preserve the recoverability of model parameters under unknown sources. I consider *full waveform inversion* (FWI) for the model parameter estimation aspect of the joint model-source inversion. Moreover, the difficulties associated with model estimation will be minimized by choosing starting models correct enough for FWI to recover the true model; predicted traveltimes from starting models are at most half a wavelength off from the data. I will investigate the interplay between source and medium anisotropy that may result in potential source-model ambiguity, in particular for the case of *vertical transversally isotropic (VTI)* elastic media. I will also investigate the sensitivity of my proposed inversion algorithm to modeling errors for controlled synthetic and field test cases, primarily inverting data from an elastic medium modeled with acoustics.

Agenda

The following sections in this document cover some of the methodology and preliminary results for this project. In the methods section, I develop the formulation for the proposed model and minimal-source inversion algorithm to better elucidate some of its key ideas. I also discuss discretization techniques used for representing singular source terms, in particular multipole point-sources, on uniform grids points so as to preserve convergence order of finite difference schemes used to solve the forward problem. A numerical convergence study for finite difference schemes with discretized multipole point-sources is covered in the preliminary results section, along with some very preliminary work on joint model-source inversion, mainly 2-D constant-density acoustics with assumed isotropic source term.

METHODS

A clear challenge and point of interest for this project will be the interplay between anisotropy of source and medium, and ultimately its impact on the inversion of medium parameters. Thus in the development of my proposed model and minimal-source inversion algorithm I consider the linear elasticity equations as I formulate the related forward and inverse problems. As mentioned earlier, I plan

on applying my inversion process to recover model parameters from VTI elastic media, which follows as a special case of the linear elasticity equations. The multipole point-source representation is also presented in the development of the forward problem as a parametrization of source parameters for general anisotropic sources. After developing the forward problem, the joint model-source inverse problem is first stated as an output least squares problem, after which is reformulated through by application of the VP method. The last methods subsection derives the proposed model and minimal-source inversion algorithm, highlighting some of the key aspects of the method for handling general anisotropic sources by an Occam's inversion type strategy.

The second half of this section concerns methodology for incorporating multipole point-source terms into forward solvers, particularly finite difference (FD) methods. I cover a technique developed by Waldén (1999) for discretizing constituents elements of multipole point-sources, that is derivatives of the delta function $\delta(\mathbf{x})$. The goal is to approximate the singularity of the source term by a more regular function of compact support such that the overall convergence rate of the FD method is maintained.

Model and Minimal-Source Inversion Formulation

Forward Problem

For the forward problem, I will be considering the 3-D linear elasticity equations in displacement-stress form:

$$\rho \ddot{u}_i = f_i + \frac{\partial}{\partial x_j} \sigma_{ij} \tag{1a}$$

$$\sigma_{ij} = c_{ijmn} \frac{\partial}{\partial x_n} u_m \tag{1b}$$

for $i, j, m, n = 1, 2, 3$, where ρ is density, $\boldsymbol{u} \in \mathbb{R}^3$ is the displacement field vector, $\boldsymbol{f} \in \mathbb{R}^3$ is the body force density, $\boldsymbol{\sigma} \in \mathbb{R}^{3 \times 3}$ is the second-order stress tensor, and c_{ijmn} is the fourth-order Hooke's tensor from linear elasticity. Note that summation over repeated indices is assumed, unless otherwise specified.

The general solution to Eq.(1) over a volume V, assuming homogeneous boundary

and initial conditions, can be shown to be given by

$$u_i(\boldsymbol{x}, t) = \int_V dV(\boldsymbol{\eta}) \, f_j(\boldsymbol{\eta}, t) * G_{ij}(\boldsymbol{x}, t; \boldsymbol{\eta}) \tag{2}$$

where '$*$' denotes convolution in time and $G_{ij}(\boldsymbol{x}, t; \boldsymbol{\eta})$ is the Green's function that satisfies the elasticity equations with an impulsive source, i.e.,

$$\rho \frac{\partial^2}{\partial t^2} G_{ij} = \delta_{ij} \delta(\boldsymbol{x} - \boldsymbol{\eta}) \delta(t) + \frac{\partial}{\partial x_j} \left\{ c_{ijmn} \frac{\partial}{\partial x_n} G_{mn} \right\}, \quad \forall \boldsymbol{x} \in V,$$

with homogeneous boundary and initial conditions. See Aki and Richards (2002) for a derivation of these results. Work by Santosa and Symes (2000) has shown that small acoustic anisotropic sources can be approximated to any degree of accuracy by a multipole-point-source series, where the number of terms in the series depends on the dimensions of the source region, desired accuracy, and anisotropy of source. Thus, it suffices to equate f in Eq.(1) to some truncated multipole point-source series centered at some $\boldsymbol{\eta}^* \in V$,

$$f_j(\boldsymbol{\eta}, t) = \sum_{|\mathbf{s}| \leq N} (-1)^{|\mathbf{s}|} F_{j;\mathbf{s}}(t) D_{\boldsymbol{\eta}}^{\mathbf{s}} \delta(\boldsymbol{\eta} - \boldsymbol{\eta}^*), \tag{3}$$

where I have used multi-index notation with $\mathbf{s} \in \mathbb{N}_0^3$, and hence

$$D_{\boldsymbol{\eta}}^{\mathbf{s}} = \left(\frac{\partial}{\partial \eta_1} \right)^{s_1} \left(\frac{\partial}{\partial \eta_2} \right)^{s_2} \left(\frac{\partial}{\partial \eta_3} \right)^{s_3}, \quad |\mathbf{s}| = \sum_{i=0}^{3} s_i.$$

I will assume that results from Santosa and Symes (2000) can be extended to the elasticity case. In Eq.(3), the 4th-order tensor $F_{j;\mathbf{s}}$ for $j \in \{1, 2, 3\}$ and $|\mathbf{s}| \leq N$ is related to the N-order *force-moment tensor*, reminiscent of the *seismic moment tensor* in earthquake seismology; see Backus and Mulcahy (1976a) and Backus and Mulcahy (1976b). Lastly, inserting Eq.(3) into Eq.(2) yields

$$u_i(\boldsymbol{x}, t) = \sum_{|\mathbf{s}| \leq N} F_{j;\mathbf{s}}(t) * G_{ij,\mathbf{s}}(\boldsymbol{x}, t; \boldsymbol{\eta}^*), \tag{4}$$

where

$$G_{ij,\mathbf{s}}(\boldsymbol{x}, t; \boldsymbol{\eta}^*) := D_{\boldsymbol{\eta}}^{\mathbf{s}} G_{ij}(\boldsymbol{x}, t; \boldsymbol{\eta}) \Big|_{\boldsymbol{\eta} = \boldsymbol{\eta}^*}.$$

Least Squares Formulation

I first formulate the inverse problem as output least squares (OLS) and then introduce a matrix-vector formulation, leading to the application of variable projection.

The OLS problem seeks to minimize the error in a least squares sense between observed and predicted data, i.e., minimize OLS objective function Φ_{OLS};

$$
\begin{aligned}
\Phi_{OLS} \quad &:= \quad \frac{1}{2} \sum_{r=1}^{N_r} \sum_{k=1}^{N_k} \left| u_{i_r}(\boldsymbol{x}_r, \omega_k) - d(\boldsymbol{x}_r, \omega_k) \right|^2 \\
&= \quad \frac{1}{2} \sum_{r=1}^{N_r} \sum_{k=1}^{N_k} \left| \sum_{|\mathbf{s}| \leq N} F_{j;\mathbf{s}}(\omega_k) G_{i_r j, \mathbf{s}}(\boldsymbol{x}_r, \omega_k; \boldsymbol{\eta}^*) - d(\boldsymbol{x}_r, \omega_k) \right|^2 .
\end{aligned}
$$

The indices r and k correspond to a discrete sample of receiver and frequency points $\{\boldsymbol{x}_r\}_{r=1}^{N_r}$ and $\{\omega_k\}_{k=1}^{N_k}$ respectively. Note that indexes i_r allude the possibility of multicomponent data. Also, note that we are minimizing the Fourier transform of the observed and predicted data, as hinted by their dependency on a sample of frequencies ω_k; this is done to decouple the source parameters from the Green's function terms since convolution is multiplication in frequency space. For each frequency ω_k, I assemble the Green's function $G_{i_r j, \mathbf{s}}(\boldsymbol{x}_r, \omega_k; \boldsymbol{\eta}^*)$ into a matrix $\mathbf{G}_k$, source parameters $F_{j;\mathbf{s}}^{(n)}(\omega_k)$ into vector $\mathbf{f}_k$, and data points $d(\boldsymbol{x}_r, \omega_k)$ into vector $\mathbf{d}_k$ such that

$$
\{\mathbf{G}_k \mathbf{f}_k\}_r = \sum_{|\mathbf{s}| \leq N} F_{j;\mathbf{s}}(\omega_k) G_{i_r j, \mathbf{s}}(\boldsymbol{x}_r, \omega_k; \boldsymbol{\eta}^*), \quad \text{and} \quad \{\mathbf{d}_k\}_r = d(\boldsymbol{x}_r, \omega_k).
$$

Thus,

$$
\Phi_{OLS} = \frac{1}{2} \sum_{k=1}^{N_k} (\mathbf{G}_k \mathbf{f}_k - \mathbf{d}_k)^T (\mathbf{G}_k \mathbf{f}_k - \mathbf{d}_k) = \frac{1}{2} \sum_{k=1}^{N_k} \|\mathbf{G}_k \mathbf{f}_k - \mathbf{d}_k\|^2 .
$$

The dimensions of the vectors and matrix given above;

$$
\begin{aligned}
dim(\mathbf{d}_k) &= 2N_r \\
dim(\mathbf{G}_k) &= 2N_r \times 2N_* \\
dim(\mathbf{f}_k) &= 2N_*
\end{aligned}
$$

where N^* is the total number of terms in the multipole expansion, and the factor of 2 accounts for the imaginary and real components of the now complex matrix and and vector elements. Finally, I introduce the following augmented matrix **G** and vectors **f** and **d**,

$$
\mathbf{G} := \begin{bmatrix} \mathbf{G}_1 & & & \\ & \mathbf{G}_2 & & \\ & & \ddots & \\ & & & \mathbf{G}_{N_k} \end{bmatrix}, \quad \mathbf{f} := \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_{N_k} \end{bmatrix}, \quad \mathbf{d} := \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \vdots \\ \mathbf{d}_{N_k} \end{bmatrix}, \tag{5}
$$

with

$$
\begin{aligned}
dim(\mathbf{d}) &= 2N_r N_k \\
dim(\mathbf{G}) &= 2N_r N_k \times 2N_k N_* \\
dim(\mathbf{f}) &= 2N_k N_*
\end{aligned}
$$

I now obtain the final form of the OLS objective function:

$$
\Phi_{OLS}[\mathbf{f}, \mathbf{m}] = \frac{1}{2}\|\mathbf{G}(\mathbf{m})\mathbf{f} - \mathbf{d}\|^2, \tag{6}
$$

where the dependency of the Green's functions matrix **G** on vector of model parameters **m** is implicit. The form of Φ_{OLS} in Eq.(6) conveniently demonstrates explicitly the linearity of the calculated data with respect to source parameters, neatly packed into vector **f**. Note that the matrix-vector form given in Eq.(6) could have also been derived in the time domain, though with a more complicated matrix **G** and convoluted arrangement of source parameters.

Applying Variable Projection Method

The nonlinear LS problem from Eq.(6) exhibits a special structure, where parameters we wish to invert for can be split into two groups: parameters such that the forward map

$$
\mathcal{F}[\mathbf{f}, \mathbf{m}] = \mathbf{G}(\mathbf{m})\mathbf{f} \tag{7}
$$

is a linear and nonlinear function with respect to said parameters, i.e., source parameters **f** and model parameters **m** respectively. The variable projection (VP) method, as developed by Golub and Pereyra (1973), seeks to exploit the separable structure in nonlinear LS problems by eliminating the linear parameters, via an orthogonal projection resulting from a linear LS problem.

I now formulate the VP problem related to the previously discussed OLS problem: minimize VP objective function $\Phi_{VP}[\mathbf{m}]$, where

$$\Phi_{VP}[\mathbf{m}] := \Phi_{OLS}[\mathbf{f}(\mathbf{m}), \mathbf{m}], \quad \text{and} \quad \mathbf{f}(\mathbf{m}) := \underset{\mathbf{f}}{\mathrm{argmin}}\, \Phi_{OLS}[\mathbf{f}, \mathbf{m}]. \tag{8}$$

Note that evaluating $\mathbf{f}(\mathbf{m})$ requires solving a linear LS problem. Furthermore, we can obtain an explicit form of $\mathbf{f}(\mathbf{m})$ by solving the normal equations associated with the respective LS problem;

$$\mathbf{G}(\mathbf{m})^T \mathbf{G}(\mathbf{m}) \mathbf{f}(\mathbf{m}) = \mathbf{G}(\mathbf{m})^T \mathbf{d}. \tag{9}$$

A clear benefit of VP is the reduction of parameters to invert for, though it may come at the price of a "more nonlinear" reduced function. Fortunately, the reduced objective function preserves the global minimizer as well as stationary points

Model and Minimal-Source Inversion Algorithm

Occam's inversion, also referred to a minimum-structure inversion, was first introduced by Constable et al. (1987) for generating 1-D smooths models from electromagnetic sounding data. This inversion formulation is based on Occam's principle, where "... among competing hypotheses that predict equally well, the one with the fewest assumptions should be selected."[1] Constructing models from electromagnetic sounding data is a difficult task due to several difficulties, including non-uniqueness and ill-conditioning issues. Occam's inversion attempts to regularize the problem by seeking the smoothest model that best explains the data, in order to avoid overfitting the data with unnecessary features in the estimated parameters. Ideally the resultant smooth model will be a lower resolution representative of the real model, reflecting the limited information inherent in the data.

In the spirit of Occam's inversion I propose a joint model-source inversion that promotes "simplicity" in the source representation as the model parameters are refined. In other words, higher order multipole terms will not be considered unless deemed necessary to fit the data. A two pronged approach is taken to promote

---

[1]From Wikipedia.

a bias towards minimal-sources of low order multipole terms during the inversion process. First, a regularization term is added to Φ_{OLS} to better-condition the source inversion step implicit in the definition of the VP objective function;

$$\Phi_{OLS}[\mathbf{f}, \mathbf{m}] = \frac{1}{2}\|\mathbf{G}(\mathbf{m})\mathbf{f} - \mathbf{d}\|^2 + \frac{\mu}{2}\|\mathbf{f}\|^2$$

thus modifying Φ_{VP},

$$\Phi_{VP}[\mathbf{m}] = \Phi_{OLS}[\mathbf{f}(\mathbf{m}), \mathbf{m}],$$

where $\mathbf{f}(\mathbf{m})$ now solves the normal equations

$$\left[\mathbf{G}(\mathbf{m})^T\mathbf{G}(\mathbf{m}) + \mu\mathbf{I}\right]\mathbf{f}(\mathbf{m}) = \mathbf{G}(\mathbf{m})^T\mathbf{d}.$$

Note that the regularization term will bias $\mathbf{f}(\mathbf{m})$ towards more minimal-sources, i.e., complex sources with more multipole terms in the expansion will be penalized. The second approach, and the key component of the algorithm, consist of a strategy for gradually introducing complexity in the source representation if a significantly better data fit can be achieved. I know give a sketch of the proposed algorithm to better elucidate the minimal-source strategy.

Let $\mathbf{G}^N(\mathbf{m})$ and $\mathbf{f}^N$ denote the Green's function matrix and source parameter vector as defined in Eq.(5), though now indexed by the highest multipole order N from expansion in Eq.(3). For example, the $N = 0$ case only includes the monopole terms in $\mathbf{G}^0(\mathbf{m})$ and $\mathbf{f}^0$, while, say $N = 1$, contains both monopole and dipole terms in $\mathbf{G}^1(\mathbf{m})$ and $\mathbf{f}^1$. The joint model and minimal-source inversion algorithm is given below.

Algorithm 1. *Joint model and minimal-source inversion:*

    1. Compute initial model $\mathbf{m}^1$

    2. for $k = 1, 2, ...$

        2.1. for $N = 0, 1, ...$

- *compute and assemble* $\mathbf{G}^{N,k} := \mathbf{G}^N(\mathbf{m}^k)$.
- *solve normal equations for* $\mathbf{f}^{N,k} := \mathbf{f}^N(\mathbf{m}^k)$,

$$\left[(\mathbf{G}^{N,k})^T \mathbf{G}^{N,k} + \mu \mathbf{I}\right] \mathbf{f}^{N,k} = (\mathbf{G}^{N,k})^T \mathbf{d}.$$

- *compute current misfit* $X_N^2 := \|\mathbf{G}^{N,k} \mathbf{f}^{N,k} - \mathbf{d}\|^2$, *and* $\Delta X^2 :=$ $X_N^2 - X_{N-1}^2$ *for* $N \geq 1$.
- *Stopping criteria for N-loop:*
    (i) **if** $X_N^2 \leq \epsilon^2$, **then** *exit N-loop with* $N^* = N$.
    (ii) **if** $N \geq 1$ *and* $\Delta X^2 > 0$, **then** *exit N-loop with* $N^* = N - 1$.
    (iii) **if** $N \geq 1$ *and* $-\Delta < \Delta X^2 < 0$, *for some* $\Delta > 0$, **then** *exit N-loop with* $N^* = N - 1$.

end *N-loop*

2.2 Update model parameter $\mathbf{m}^k \to \mathbf{m}^{k+1}$ *from the reduced objective function* Φ_{VP}, *with* $\mathbf{G}(\mathbf{m}) = \mathbf{G}^{N^*}(\mathbf{m})$.

2.3 Check stopping criteria for k-loop.

end *k-loop*

The *for*-loop over N in step 2.1 is the crucial component to minimal-source inversion, it incrementally incorporates higher order multipole terms in the source representation until any of the exiting *if*-statements are satisfied. The first *if*-statement truncates the multipole expansion at the current value of N if the misfit is smaller than some given tolerance ϵ^2. In other words, if the data is fitted well enough with an N-order multipole-point source adding higher order terms would be unnecessary. Note that in the early iterations, as $\mathbf{m}$ is refined, it may be difficult or simply not plausible to satisfy *(i)* for a reasonable N since the current model $\mathbf{m}^k$ may be far enough from the true model. *If*-statements *(ii)* and *(iii)* alternatively give criteria in terms of relative improvement of misfit as higher

order multipole terms are added. Namely, if the misfit increases in the next N-iteration ($\Delta X^2 > 0 \iff X_N^2 > X_{N-1}^2$), then the multipole expansion is truncated at the previous step, i.e., $N - 1$. If however the next N-iteration yields a smaller misfit ($\Delta X^2 < 0$), but the improvement in misfit is too small ($-\Delta < \Delta X^2$), then the multipole expansion is still truncated at the previous step. To summarize, step 2.1 automates the process of gradually incorporating higher order multipole terms in the source representation in such a manner that if the data can be fitted reasonably well, or if adding more terms does not reduce the misfit enough, then a lower order source representation is preferred.

Multipole Point-Source Approximation

Formulating the source term f as a truncated series of multipole point-source terms resulted in a forward map $\mathcal{F}[\mathbf{f}, \mathbf{m}]$, see Eq.(7), that is linear with respect to source parameters $\mathbf{f}$. Consequently, in order to evaluate the forward map, or elements of Green's function matrix $\mathbf{G}$, one must solve a PDE with increasingly singular source terms that involve derivatives of the spatial delta function, $D_{\boldsymbol{\eta}}^{\mathbf{s}} \delta(\boldsymbol{\eta} - \boldsymbol{\eta}^*)$. Errors from the discretization of singularities can have a detrimental effect on the overall simulated waves, in particular it can hamper the forward solver's convergence rate if special care is not taken. Waldén (1999) develops an approximation to singular sources in a finite difference (FD) setting for the 1-D Helmholtz equation:

$$\begin{cases} \dfrac{d^2}{dx^2} u + a^2 u = \delta^{(s)}(x - x^*), \\ \quad a \in \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}, \\ \qquad x \in \mathbb{R}, \end{cases}$$

where $\delta^{(s)}$ denotes the sth derivative of the delta function. He proofs that overall the convergence rate of FD schemes can be preserved if the singular source terms satisfies certain discrete moment conditions. In the following subsection I derive the method for discretizing singular sources as proposed by Waldén (1999), though I will follow some of the notation used by Tornberg and Engquist (2004).

Approximation to 1-D $\delta^{(s)}(x)$

I first consider the 1-D case for approximating $\delta^{(s)}(x)$ for $s = 0$ in order to introduce some notation and key ideas. Let $\delta_\epsilon \in C_0(\mathbb{R})$ denote an approximation to $\delta(x)$ such that its support is contained in $[-\epsilon, \epsilon]$.

Definition 1. *Given an approximation δ_ϵ, some uniform grid $\{x_i | x_i = ih, i \in \mathbb{Z}, h > 0\}$, and some $x^* \in \mathbb{R}$, define $M_r(\delta_\epsilon, x^*, h)$ as the discrete r^{th} moment of δ_ϵ:*

$$M_r(\delta_\epsilon, x^*, h) := h \sum_{i \in \mathbb{Z}} \delta_\epsilon(x_i - x^*)(x_i - x^*)^r. \tag{10}$$

Definition 2. *Given a uniform grid $\{x_i | x_i = ih, i \in \mathbb{Z}, h > 0\}$, and some $q \in \mathbb{N}$, define the following subset of compact continuous functions on $\mathbb{R}$ denoted by $Q_q(h)$:*

$$Q_q(h) := \left\{ \phi \in C_0(\mathbb{R}) \,\middle|\, M_r(\phi, x^*, h) = \delta_{r0}, \forall x^* \in \mathbb{R}, \ 0 \leq r < q \right\}$$

where δ_{r0} is the Kronecker delta,

$$\delta_{r0} = \left\{ \begin{array}{ll} 1, & r = 0 \\ 0, & r \neq 0 \end{array} \right. .$$

The space $Q_q(h)$ thus consists of continuous functions with compact support that have unit mass and vanishing moments in a discrete sense with respect to some uniform grid, mimicking the properties that define the distribution $\delta(x)$;

$$\int \delta(x - x^*) f(x) \, dx = f(x^*),$$

which holds true for $f \in C_0^n(\mathbb{R}), \forall n \in \mathbb{N}_0$. The number of vanishing moments that are satisfied by the approximation δ_ϵ will control the error in some sense as stated in the following proposition.

Proposition 1. *Suppose $\delta_\epsilon \in Q_q(h)$, for some $q > 0$, and that $f \in C^q(\mathbb{R})$. Then*

$$E := \left| h \sum_{i \in \mathbb{Z}} \delta_\epsilon(x_i - x^*) f(x_i) - \int \delta(x - x^*) f(x) \, dx \right| \leq Ch^q,$$

and $E = 0$ if f is constant.

Ideally, one would be inclined to pick approximations δ_ϵ with the smallest ϵ width possible, however the following proposition states that ϵ is in actuality constrained by the desired order accuracy.

Proposition 2. *There exists $\delta_\epsilon \in Q_q(h)$ if and only if $2\epsilon \geq qh$.*

Proofs of the aforementioned propositions are given in Tornberg and Engquist (2004). I will extend and provide proofs for their more general counterparts based on approximations of the distribution $\delta^{(s)}(x)$, that is the sth derivative of the delta function.

Definition 3. *Given a uniform grid $\{x_i | x_i = ih, i \in \mathbb{Z}, h > 0\}$, and some $q \in \mathbb{N}$ and $s \in \{0\} \cup \mathbb{N}$, define the following subset of compact continuous functions on $\mathbb{R}$ denoted by $Q_q^s(h)$:*

$$Q_q^s(h) := \left\{ \phi \in C_0(\mathbb{R}) \,\Big|\, M_r(\phi, x^*, h) = s!(-1)^s \delta_{rs}, \forall x^* \in \mathbb{R}, \ 0 \leq r < q+s \right\}.$$

Theorem 1. *Suppose $\delta_\epsilon^s \in Q_q^s(h)$, for some $q > 0$ and $s \geq 0$, and that $f \in C^{q+s}(\mathbb{R})$. Then*

$$E := \left| h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) f(x_i) - \int \delta^{(s)}(x - x^*) f(x) \, dx \right| \leq Ch^q,$$

and $E = 0$ if f is constant.

Proof. The first thing to do is expand $f(x_i)$ by its Taylor series approximation;

$$f(x_i) = \sum_{n=0}^{q+s-1} \frac{f^{(n)}(x^*)}{n!}(x_i - x^*)^n + R(x_i),$$

where $R(x_i)$ is the remainder term. Using the formula above in the summation term of E yields the following:

$$
\begin{aligned}
h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) f(x_i) &= h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) \left\{ \sum_{n=0}^{q+s-1} \frac{f^{(n)}(x^*)}{n!}(x_i - x^*)^n + R(x_i) \right\}, \\
&= \sum_{n=0}^{q+s-1} \frac{f^{(n)}(x^*)}{n!} h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*)(x_i - x^*)^n + h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) R(x_i), \\
&= \sum_{n=0}^{q+s-1} \frac{f^{(n)}(x^*)}{n!} M_n(\delta_\epsilon^s, x^*, h) + h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) R(x_i).
\end{aligned}
$$

Assuming that $\delta_\epsilon^s \in Q_q^s(h)$, then we have $M_n(\delta_\epsilon^s, x^*, h) = s!(-1)^s \delta_{ns}$, i.e.,

$$
h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) f(x_i) \;=\; (-1)^s f^{(s)}(x^*) + h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) R(x_i),
$$

$$
=\; \int \delta^s(x - x^*) f(x)\, dx + h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) R(x_i).
$$

Thus,

$$
E = \left| h \sum_{i \in \mathbb{Z}} \delta_\epsilon^s(x_i - x^*) R(x_i) \right|.
$$

Recall that δ_ϵ^s is of compact support, primarily supp $\delta_\epsilon^s(x) \subset [-\epsilon, \epsilon]$. Thus, there exist a finite subset of integers $\{i_n\}_{n=0}^N$ such that $\delta_\epsilon^s(x_i - x^*) = 0$ if $i \notin \{i_n\}_{n=0}^N$. Moreover, it is assumed that $f \in C^{q+s}(\mathbb{R})$, thus there exists y such that

$$
R(x_i) = \frac{f^{(s+q)}(y)}{(s+q)!}(x_i - x^*)^{s+q}.
$$

The error E takes the form

$$
\begin{aligned}
E &= \left| h \sum_{n=0}^N \delta_\epsilon^s(x_{i_n} - x^*) \frac{f^{(s+q)}(y)}{(s+q)!}(x_{i_n} - x^*)^{s+q} \right|, \\
&\leq \left| \frac{f^{(s+q)}(y)}{(s+q)!} \right| \max_n |x_{i_n} - x^*|^q \left| h \sum_{n=0}^N \delta_\epsilon^s(x_{i_n} - x^*)(x_{i_n} - x^*)^s \right|, \\
&\leq Ch^q |M_s(\delta_\epsilon^s, x^*, h)|, \\
&\leq Ch^q.
\end{aligned}
$$

$\square$

Theorem 2. *There exists $\delta_\epsilon^s \in Q_q^s(h)$ if and only if $2\epsilon \geq (q+s)h$.*

Proof. Note that $2\epsilon \geq (q+s)h$ implies that the support of δ_ϵ^s must contain at least $q+s$ grid points, in other words $\delta_\epsilon^s(x_i - x^*)$ is at the very least potentially nonzero at $s+q$ grid points. Suppose that the support of δ_ϵ^s contains N grid points indexed

by $\{i_n\}_{n=1}^N$. In order for δ_ϵ^s to belong to $Q_q^s(h)$ it must satisfy $q+s$ conditions related to its discrete moments, i.e.,

$$M_r(\delta_\epsilon^s, x^*, h) = h \sum_{n=1}^N \delta_\epsilon^s(x_{i_n} - x^*)(x_{i_n} - x^*)^r = s!(-1)^s \delta_{rs}.$$

These constraints can be written in matrix-vector form as $\mathbf{A}\mathbf{d} = \mathbf{b}$, with

$$\{\mathbf{A}\}_{mn} = (x_{i_n} - x^*)^{m-1}, \quad \{\mathbf{d}\}_n = \delta_\epsilon^s(x_{i_n} - x^*), \quad \{\mathbf{b}\}_m = s!(-1)^s \delta_{m-1,s},$$

for $m = 1, ..., q+s$ and $n = 1, ..., N$. Thus, in order to have a solution $\mathbf{d}$ to the linear system resulting from satisfying conditions on the moments M_r, the number of unknowns N has to be greater than or equal to the number of constraints $q+s$. $\quad\square$

Proposition **??** provides a measure of the error between the discrete behavior of the approximation δ_ϵ^s and the analytic properties of the distribution $\delta^{(s)}$ when applied to a smooth function f. A number of discrete moments must be constrained in order to yield an approximation with desired rate of convergence. Moreover, Proposition 2 states that δ_ϵ^s must be defined over enough grid points to be properly resolved.

Implementing 1-D Approximation

Given a source location x^* only the values $\{\delta_\epsilon^s(x_i - x^*) \mid i \in \mathbb{Z}\}$ are required for implementation purposes regarding FD methods. As mentioned in the proof of Proposition 2, $\delta_\epsilon^s(x_i - x^*)$ will take on (possibly) nonzero values at a subset of indexes $\{i_n\}_{n=1}^N \subset \mathbb{Z}$. The grid points $\{x_{i_n}\}_{n=1}^N$ are referred to as the stencil of $\delta_\epsilon^s(x-x^*)$. Let $\mathbf{d} \in \mathbb{R}^N$ be a vector containing $\delta_\epsilon^s(x - x^*)$ evaluated at stencil points $\{x_{i_n}\}_{n=1}^N$,

$$d_n := \delta_\epsilon^s(x_{i_n} - x^*).$$

In computing $\mathbf{d}$, I will follow the same methodology given by Petersson and Sjogreen (2010).

For a given x^*, let $i_* \in \mathbb{Z}$ such that $x_{i_*} \le x^* < x_{i_*+1}$, and let $\alpha := (x_* - x_{i^*})/h \in [0, 1)$. Conditions on discreet moments M_r can be reformulated to yield an equivalent subspace of $C_0(\mathbb{R})$ functions.

Definition 4. *Given a uniform grid $\{x_i | x_i = ih, i \in \mathbb{Z}, h > 0\}$, and some $q \in \mathbb{N}$ and $s \in \{0\} \cup \mathbb{N}$, define the following subset of compact continuous functions on $\mathbb{R}$:*

$$\tilde{Q}_q^s(h) := \left\{ \phi \in C_0(\mathbb{R}) \,\middle|\, M_r(\phi, x^*, h) = (-1)^s \frac{r!}{(r-s)!}(x^* - x_{i_*})^{r-s}, \forall x^* \in \mathbb{R}, \ 0 \le r < q+s \right\}.$$

Theorem 3. *The two spaces $Q_r^s(h)$ and $\tilde{Q}_r^s(h)$ are equivalent.*

The constraints on discrete moments $M_r(\delta_\epsilon^s, x^*, h)$, specified by membership of approximation δ_ϵ^s to $Q_q^s(h)$, can be further simplified by taking advantage of the fact that $x_i - x_j = (i-j)h$ for all $i, j \in \mathbb{Z}$:

$$\sum_{n=1}^{N} d_n (i_n - i_*)^q = \begin{cases} \dfrac{(-1)^s}{h^{s+1}} \dfrac{r!}{(r-s)!} \alpha^{r-s}, & r-s \ge 0 \\[2mm] 0, & r-s < 0 \end{cases}, \quad \text{for } r = 0, 1, \dots, q+s-1. \quad (11)$$

Thus we have $q + s$ equations and N unknowns. Approximation functions with narrow widths are preferred in practice, and thus I will choose width ϵ such that $N = q + s$. I know reformulate Eq.(11) in matrix-vector form: $\mathbf{Ad} = \mathbf{b}$ where

$$\begin{aligned}
\{\mathbf{A}\}_{mn} &= (i_n - i_*)^{m-1}, \\
\{\mathbf{d}\}_n &= \delta_\epsilon^s(x_{i_n} - x^*), \\
\{\mathbf{b}\}_m &= \begin{cases} \dfrac{(-1)^s}{h^{s+1}} \dfrac{(m-1)!}{(m-1-s)!} \alpha^{m-1-s}, & m-1-s \ge 0 \\[2mm] 0, & m-1-s < 0 \end{cases}
\end{aligned}$$

for $m, n = 1, \dots, q+s$. Note that the matrix $\mathbf{A}$ given above is independent of h and α, and hence its inverse can be explicitly computed offline and stored to later generate the vector $\mathbf{d}$ for the desired approximation δ_ϵ^s.

2-D and 3-D Approximations

The delta function, and its derivatives, in higher dimensions can be handled by taking tensor products of the 1-D counterparts. For example, the 3-D multi-derivative of the delta function is approximated as follows:

$$D^{\mathbf{s}} \delta(\mathbf{x}) \approx \prod_{i=1}^{3} \delta_{\epsilon_i}^{(s_i)}(x_i)$$

is a q-order approximation if $\delta_{\epsilon_i}^{s_i} \in Q_q^{s_i}(h)$ for each $i = 1, 2, 3$.

PRELIMINARY RESULTS

The proposed model and minimal-source inversion algorithm is currently under development as an IWAVE application. IWAVE is a software framework written in C and C++ for time-dependent PDE solvers, mainly FD methods on uniform grids. The *Rice Vector Library* (RVL) is another software package that provides the underlying framework for inversion schemes based on Newton-like methods in IWAVE. For more information on simulation based optimization via IWAVE and RVL see *the Rice Inversion Project* (TRIP) technical report Symes et al. (2004), or published works Padula et al. (2009), Symes et al. (2011).

Preliminary work shown here considers 2-D acoustic media as a starting point for this project. In particular, the forward problem will consist of solving the 2D acoustic equations, in velocity-pressure form, as given by the following system of first order PDEs:

$$\rho(\boldsymbol{x})\frac{\partial}{\partial t}v_1(\boldsymbol{x},t) + \frac{\partial}{\partial x_1}p(\boldsymbol{x},t) = 0, \tag{12a}$$

$$\rho(\boldsymbol{x})\frac{\partial}{\partial t}v_2(\boldsymbol{x},t) + \frac{\partial}{\partial x_2}p(\boldsymbol{x},t) = 0, \tag{12b}$$

$$\frac{1}{\kappa(\boldsymbol{x})}\frac{\partial p}{\partial t}(\boldsymbol{x},t) + \frac{\partial}{\partial x_1}v_1(\boldsymbol{x},t) + \frac{\partial}{\partial x_2}v_2(\boldsymbol{x},t) = f(\boldsymbol{x},t), \tag{12c}$$

where p is the pressure field, $[v_1, v_2]$ is the particle velocity field, and $[\rho, \kappa]$ are medium parameters (density and bulk modulus respectively). The depth axis will be assumed to be oriented along the x_1-axis, while x_2 corresponds to horizontal displacement. Note that the source term f is a scalar and is injected in the pressure equation. The forward problem will be solved by *staggered grid FD* methods implemented in IWAVE based on the 2-2 and 2-4 order schemes originally developed for elasticity by Virieux (1986) and Levander (1988). The '2-4' is used to indicate the temporal and spatial discretization order of the time-space FD method, i.e., 2-4 refers to 2nd-order in time and 4th-order in space. In this section I present some preliminary results in the context of 2-D acoustic media: a convergence study for FD methods used to solve the forward problem with the singular source approximation and some preliminary joint model-source inversions for an isotropic point-source.

Convergence Rate Analysis for Point-Sources

The method of discretizing point sources as developed by Waldén (1999) has been used in other works related to seismic wave propagation: Petersson and Sjogreen (2010) make use of the method when solving the 1-D scalar wave equation, as well as extend the approximation to 2-D and 3-D, demonstrating through numerical examples that convergence rates of finite difference methods can be preserved. With similar results, Geiser (2009) applies the singular source term approximation to wave propagation in elastic media. As proof of concept I offer a convergence rate analysis for the FD methods considered here when applying the singular source discretization technique. In particular, I will be testing the 2nd and 4th order approximations to $\delta(x_1, x_2)$, and $\dfrac{\partial}{\partial x_2} \delta(x_1, x_2)$.

A homogeneous 2-D acoustic model, spanning $8\,km$ in depth and $8\,km$ horizontally, is used throughout this study. Density and bulk modulus are set to $1\,g/cm^3$ and $16\,GPa$ respectively, resulting in a medium velocity of $4\,km/s$. I have chosen to use the same source time function considered in the numerical examples by Geiser (2009), denoted by $F(t)$;

$$
F(t) = \begin{cases} exp\left(-\dfrac{1}{t(1-t)}\right) & , t \in [0,1] \\[2em] 0 & , t \in (-\infty, 0) \cup (1, \infty) \end{cases}
$$

Note that $F \in C_0^\infty(\mathbb{R})$.

Errors and Convergence Rates

Convergence rates are approximated by comparing seismic traces, that is time series of computed pressure field p at a receiver location $\mathbf{x}_r = (x_{1,r}, x_{2,r})$. Let p_h denote the numerical pressure field computed via a FD method with spatial grid size h. The convergence rate at a particular receiver location $\mathbf{x}_r$ is approximated via the following formula:

$$
R(\mathbf{x}_r) = \log_2\left(\frac{\|p_h(\mathbf{x}_r, \cdot) - p_{h/2}(\mathbf{x}_r, \cdot)\|}{\|p_{h/2}(\mathbf{x}_r, \cdot) - p_{h/4}(\mathbf{x}_r, \cdot)\|}\right), \tag{13}
$$

where the norm $\|\cdot\|$ is chosen to be either $\|\cdot\|_2$ or $\|\cdot\|_\infty$ defined by

$$\|p(\boldsymbol{x}_r,\cdot)\|_2 := \sqrt{\sum_k |p(\boldsymbol{x},t_k)|^2},$$

$$\|p(\boldsymbol{x}_r,\cdot)\|_\infty := \max_k |p(\boldsymbol{x}_r,t_k)|.$$

Results for $\delta(x_1,x_2)$

Convergence rates are first tested for both 2-2 and 2-4 staggered grid FD methods with source term

$$f(\boldsymbol{x},t) = F(t)\delta(\boldsymbol{x}-\boldsymbol{x}^*),$$

with $\boldsymbol{x}^* = [-4,4]\,km$. The 2-D delta function is approximated by the tensor product of 1-D approximations, i.e., $\delta(x_1,x_2) \approx \delta_{\epsilon_1}(x_1)\delta_{\epsilon_2}(x_2)$. Space-time discretization choices for both 2-2 and 2-4 schemes are given in Tab.(1). The spatial grid size are refined by factors of two in order to use Eq.(13) to approximate convergence rates. The time step for each FD scheme, constant under h refinement, was chosen to be several times smaller than the required size Δt_{CFL} from stability estimates for the smallest h. Small enough time step ideally decreases $\mathcal{O}(\Delta t^2)$ errors from the temporal discretizing to help isolate the convergence rate associated with the spatial discretization, essentially producing a semi-discrete solution.

	h-refinement [m]	time step [ms]
2-2	10, 20, 40	0.5
2-4	20, 40, 80	0.5

Table 1: Space-time discretization choices for convergence rate study with δ source term.

Convergence rates are computed for seismic traces along a line of receivers at depth $-4\,km$, using both $\|\cdot\|_2$ and $\|\cdot\|_\infty$ norms, see Fig.(1) and Fig.(2). Indeed convergence rate is preserved when using point source approximations of appropriate order for both 2-2 and 2-4 schemes, even at 2 grid points from the source location.

Results for $\dfrac{\partial}{\partial x_2}\delta(x_1, x_2)$

Convergence rates are are tested for both 2-2 and 2-4 staggered grid FD methods with source term

$$f(\boldsymbol{x}, t) = F(t)\frac{\partial}{\partial x_2}\delta(\boldsymbol{x} - \boldsymbol{x}^*),$$

with $\boldsymbol{x}^* = [-4, 4]\,km$. The 2-D delta function is approximated by the tensor product of 1-D approximations, i.e., $\delta(x_1, x_2) \approx \delta_{\epsilon_1}(x_1)\delta^{(1)}_{\epsilon_2}(x_2)$. The same space-time discretization choices given in Tab.(1) were chosen for this test. Approximate convergence rates are shown in Fig.(3) and Fig.(4), with expected results. One thing to notice is that as the derivative and approximation order increases, so does the size of the point-source approximation stencil, and along with it the region of grid points with affected convergence rate. As proof of concept, Fig.(5) shows convergence rates for the 2-4 scheme when using a second order approximation to point-source term. It is observe that the convergence rate drops from 4 to 2.

Preliminary Joint Model-Source Inversion

In this section I offer some preliminary results concerning joint model-source inversion for a simple 2-D constant-density acoustic medium. Inversion results via VP are compared against model-only inversions assuming known source.

The forward problem in this context consists of: computing time series of pressure field at given receiver locations $\boldsymbol{x}_r$, that is $p(\boldsymbol{x}_r, t_k)$, where $p(\boldsymbol{x}, t)$ is the solution to Eq.(12) for given model parameters $[\rho, \kappa]$ and source term

$$f(\boldsymbol{x}, t) = F(t)\delta(\boldsymbol{x} - \boldsymbol{x}^*).$$

Source and model parameters $[\mathbf{f}, \mathbf{m}]$ are given by discretized counter parts of $F(t)$ and $\kappa(\boldsymbol{x})$ respectively. Let $\mathcal{F}[\mathbf{f}, \mathbf{m}]_{r,k} = p(\boldsymbol{x}_r, t_k)$, thus emphasizing the dependency on source-model parameters $[\mathbf{f}, \mathbf{m}]$.

Model inversion, with known source, is carried out here by solving the OLS problem: minimize OLS objection function Φ_{OLS} with respect to model parameters $\mathbf{m}$,

$$\Phi_{OLS}(\hat{\mathbf{f}}, \mathbf{m}) := \frac{1}{2}\sum_r \sum_k \left|\mathcal{F}[\hat{\mathbf{f}}, \mathbf{m}]_{r,k} - d(\boldsymbol{x}_r, t_k)\right|^2,$$

where $\hat{\mathbf{f}}$ denotes the true source parameters and $d(\boldsymbol{x}_r, t_k)$ is the observed data. Joint model-source inversion is done using the VP method, that is minimize VP objective function Φ_{VP} with respect to model parameters $\mathbf{m}$, where

$$\Phi_{VP}[\mathbf{m}] := \Phi_{OLS}[\mathbf{f}(\mathbf{m}), \mathbf{m}], \quad \text{and} \quad \mathbf{f}(\mathbf{m}) := \underset{\mathbf{f}}{\text{argmin}} \, \Phi_{OLS}[\mathbf{f}, \mathbf{m}].$$

A gradient descent method, with a quadratic interpolation line search, is used for the numerical minimization of the OLS and VP objective functions. More sophisticated numerical optimization will of course be considered in future work.

Transmission Test Case

The first inversion test case considers the recovery of a smooth bulk modulus anomaly using transmission data, as depicted in Fig.(6). The true source-time function was chosen to be a Ricker wavelet with $15\,Hz$ peak frequency, centered at $0.3\,s$, see Fig.(7). A homogeneous bulk modulus model of $15\,MPa$ is used as the starting model for the inversion process.

Recovered models are shown in Fig.(8a) for OLS with known source-time function and Fig.(8b) for joint source-model inversion via VP. Both inversions are able to deduce a high bulk modulus anomaly, correctly centered in the horizontal direction. On the other hand, both inversions effectively smear the anomaly in the depth axis, not too surprising since the recorded data does not provide information to penalize such artifacts. The resulting source time-function, given by $\mathbf{f}(\mathbf{m})$ in the VP method, is shown in Fig.(9a) along with its deviation from the true source, Fig.(9b). The recovered source is correctly centered at $0.3\,s$ with similar waveform to that of the true source, deviating at most by 6% of max amplitude. Lastly, Fig.(10) plots the relative data residual as the model is refined, showing that the VP method produces a model with better fit. Gradient decent iteration terminates when model update is too small.

CONCLUSIONS

Solving the forward problem accurately and efficiently is of upmost importance to the inverse problem, "quality of the results of any inverse problem depends heavily on the realism of the forward modeling," Gauthier et al. (1986). The forward

problem, for the inversion algorithm proposed here, will consist of solving the elasticity (or acoustics) equations with singular source terms in the form of multipole point-sources. Numerical results demonstrate that convergence rates of FD methods, particularly the 2-2 and 2-4 staggered grid FD methods for acoustics, can be preserved away from the source when approximating appropriately source terms of the form $f(\boldsymbol{x}, t) = F(t)\dfrac{\partial^s}{\partial x^s}\delta(\boldsymbol{x} - \boldsymbol{x}^*)$, particularly for $s = 0, 1$. The main idea is to approximate the 1-D distribution $\dfrac{\partial^s}{\partial x^s}\delta(x)s$ by a $C_0(\mathbb{R})$ function $\delta_\epsilon^{(s)}(x)$ that satisfies a number of discrete moment conditions to yield a desired accuracy, Waldén (1999).

Some preliminary joint-source model inversions were presented for simple 2-D constant-density acoustic models with assumed isotropic source term. In particular, the joint model-source inversion via the VP method was compared against model-only inversion with correct source. Results for both transmission test case analyzed here demonstrate similar recovered model parameters, localizing the the presence of anomaly in the horizontal direction, though creating artifacts in the depth axis possibly due inadequacy of data and inversion formulation. Moreover, accurate source-time function from the VP method was also recovered. It was also observed that the VP inversion lead to a smaller objective function value, i.e., better fit to data. These results further validate the use of VP for dealing with the source estimation aspect of joint model-source inversion.

FUTURE WORKS

The proposed joint model and minimal-source inversion algorithm is designed to tackle the problem of model inversion under unknown general anisotropic sources. A key component of the method is the multipole point-source representation of anisotropic sources, resulting in a parametrization yielding an inverse problem that lends itself to reducible nonlinear programming method, in particular the VP method. Preliminary results given here demonstrate the ability for dealing with the singularity of multipole point source terms when solving the forward problem, as well as further validate the use of VP for in joint model-source inversion.

The pseudo code for the proposed joint model and minimal souce inversion served

well as an instructive tool to elucidate the essence of model and minimal-source inversion. I now highlight some of the details that will play a vital role in the success of the overall algorithm and entail future work:

- Solving for $\mathbf{f}^{N,k}$:
  Computing source parameters involves solving the regularized normal equations. A unique $\mathbf{f}^{N,k}$ should be recoverable in theory, however ill-conditioning of $G^{N,k}$ may creep in as N is increased. Thus, solving normal equations should be done is a robust and efficient manner, as well as choosing μ wisely.

- Choosing ϵ^2 and Δ:
  Should ϵ^2 and Δ be dependent on iteration? How should they be chosen? Also, if not enough statistical information is known about the errors in the sampled data can a reasonable ϵ^2 still be chosen?

- Updating $\mathbf{m}^k \to \mathbf{m}^{k+1}$ from $\Phi_{VP}[\mathbf{m}]$:
  This step is perhaps one of the more difficult ones. More sophisticated numerical optimization schemes will need to be incorporated for updating the model when minimizing reduced objective function $\Phi_{VP}[\mathbf{m}]$.

I plan on testing the model and minimal-source inversion algorithm primarily in two scenarios to analyze the influence of source and its impact on recovered medium parameters. The first case will consider an acoustic medium with an effective source modeled after the physics of air-gun arrays in order to test the both the recoverability of medium parameters and the influence of anisotropy from a realistic source. Lastly, the interplay between source and medium anisotropy will be explored in the inversion of VTI elastic medium parameters.

REFERENCES

Aki, K., and P. G. Richards, 2002, Quantitative seismology / keiiti aki, paul g. richards.: Sausalito, Calif. : University Science Books, c2002.

Backus, G., and M. Mulcahy, 1976a, Moment tensors and other phenomenological descriptions of seismic sourcesi. continuous displacements: Geophysical Journal International, **46**, 341–361.

——, 1976b, Moment tensors and other phenomenological descriptions of seismic sourcesii. discontinuous displacements: Geophysical Journal International, **47**, 301–329.

Bube, K., P. Lailly, P. Sacks, F. Santosa, and W. W. Symes, 1988, Simultaneous determination of source wavelet and velocity profile using impulsive point-source reflections from a layered fluid: Geophysical Journal International, **95**, 449–462.

Constable, S. C., R. L. Parker, and C. G. Constable, 1987, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data: Geophysics, **52**, 289–300.

Delprat-Jannaud, F., and P. Lailly, 2005, A fundamental limitation for the reconstruction of impedance profiles from seismic data: Geophysics, **70**, R1–R14.

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

Geiser, J., 2009, Higher-order splitting method for elastic wave propagation: International Journal of Mathematics and Mathematical Sciences, **2008**.

Golub, G. H., and V. Pereyra, 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate: SIAM Journal on numerical analysis, **10**, 413–432.

Jost, M. u., and R. Herrmann, 1989, A students guide to and review of moment tensors: Seismological Research Letters, **60**, 37–57.

Julian, B. R., A. D. Miller, and G. Foulger, 1998, Non-double-couple earthquakes 1. theory: Reviews of Geophysics, **36**, 525–549.

Landrø, M., and R. Sollie, 1992, Source signature determination by inversion: Geophysics, **57**, 1633–1640.

Levander, A., 1988, Fourth-order finite-difference pw seismograms: Geophysics, **53**, 1425.

Lewis, R. M., 1989, Source-velocity identification for a layered model of reflection seismology: PhD thesis, PhD thesis, Department of Mathematical Sciences, Rice University, Houston, Texas, USA.

Li, J., Y.-M. He, and Z.-X. Yao, 2006, Source finiteness and rupture propagation

using higher-degree moment tensors: Bulletin of the Seismological Society of America, **96**, 1241–1256.

Minkoff, S. E., and W. W. Symes, 1995, Estimating the energy source and reflectivity by seismic inversion: Inverse Problems, **11**, 383.

——, 1997, Full waveform inversion of marine reflection data in the plane-wave domain: Geophysics, **62**, 540–553.

Padula, A. D., W. Symes, and S. D. Scott, 2009, A software framework for the abstract expression of coordinate-free linear algebra and optimization algorithms: ACM Transactions on Mathematical Software, **36**, 8:1–8:36.

Petersson, N. A., and B. Sjogreen, 2010, Stable grid refinement and singular source discretization for seismic wave simulations: Commun. Comput. Phys, **8**, 1074–1110.

Ramm, A., 1985, Inversion of the backscattering data and a problem of integral geometry: Physics Letters A, **113**, 172–176.

Robinson, E. A., 1957, Predictive decomposition of seismic traces: Geophysics, **22**, 767–778.

Ruhe, A., and P. Å. Wedin, 1980, Algorithms for separable nonlinear least squares problems: Siam Review, **22**, 318–337.

Santosa, F., and W. W. Symes, 2000, Multipole representation of small acoustic sources: Chinese Journal of Mechanics, **16**, 15–21.

Shearer, P. M., 2009, Introduction to seismology: Cambridge University Press.

Stump, B. W., and L. R. Johnson, 1982, Higher-degree moment tensorsthe importance of source finiteness and rupture propagation on seismograms: Geophysical Journal International, **69**, 721–743.

Symes, W. W., E. Dussaud, H. Dajani, and A. Padula, 2004, A time-stepping library for simulation-driven optimization: Technical report, Technical Report 04-XX, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

Symes, W. 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).

Tornberg, A.-K., and B. Engquist, 2004, Numerical approximations of singular source terms in differential equations: Journal of Computational Physics, **200**, 462–488.

Ulrych, T., 1971, Application of homomorphic deconvolution to seismology: Geophysics, **36**, 650–660.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media; velocity-stress finite-difference method: Geophysics, **51**, 889–901.

Waldén, J., 1999, On the approximation of singular source terms in differential equations: Numerical Methods for Partial Differential Equations, **15**, 503–520.

Wang, K., J. R. Krebs, D. Hinkley, A. Baumstein, et al., 2009, Simultaneous full-waveform inversion for source wavelet and earth model: SEG Int'l. Expo. & Ann. Meeting, Expanded Abstracts, 2537–2541.

Zhou, C., G. T. Schuster, S. Hassanzadeh, and J. M. Harris, 1997, Elastic wave equation traveltime and waveform inversion of crosswell data: Geophysics, **62**, 853–868.

Ziolkowski, A., 1991, Why don't we measure seismic signatures?: Geophysics, **56**, 190–201.
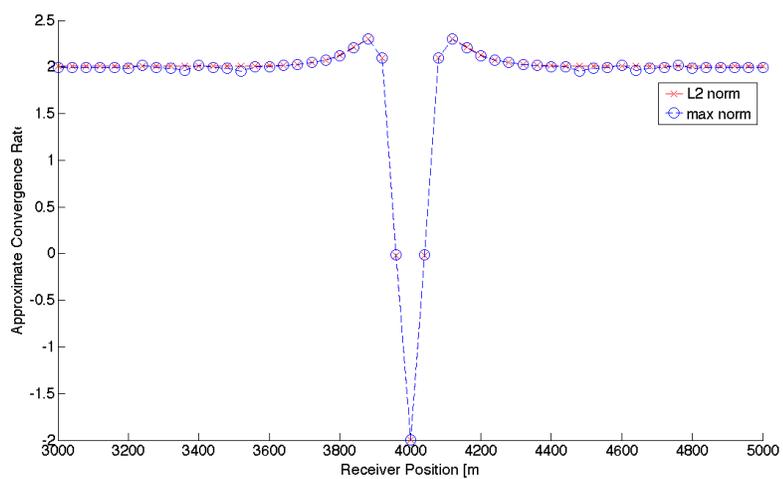
Figure 1: Rates for 2-2 staggered grid scheme with $\delta(\boldsymbol{x})$ source term and approximation $\delta_\epsilon^0(x_1)\delta_\epsilon^0(x_2) \in Q_2^0(h) \times Q_2^0(h)$.
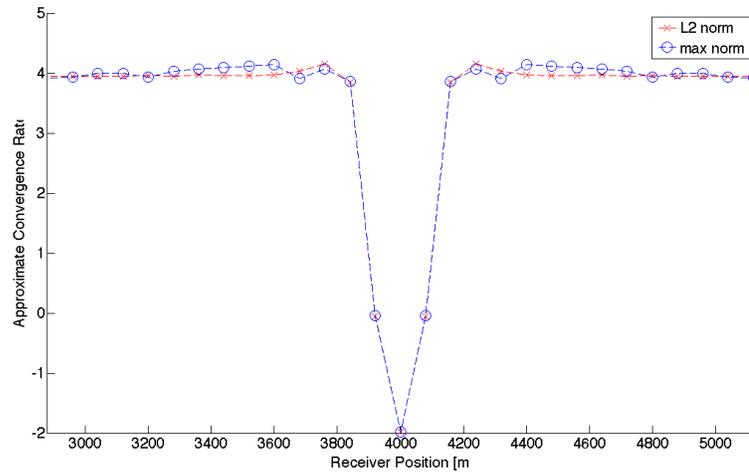
Figure 2: Rates for 2-4 staggered grid scheme with $\delta(x)$ source term and approximation $\delta_\epsilon^0(x_1)\delta_\epsilon^0(x_2) \in Q_4^0(h) \times Q_4^0(h)$.
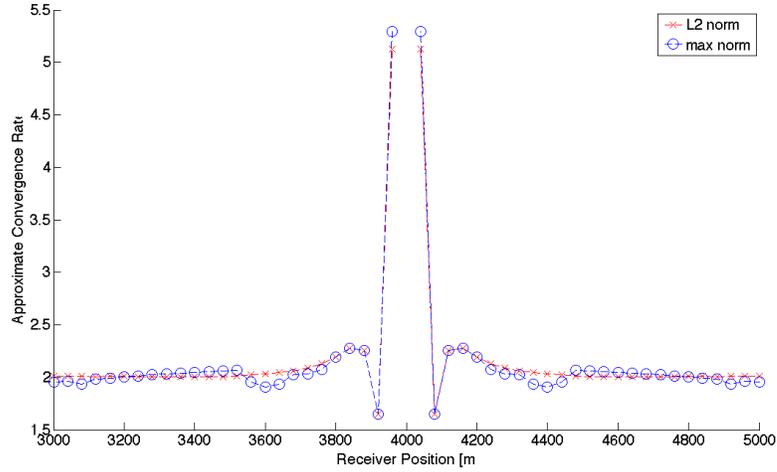

Figure 3: Rates for 2-2 staggered grid scheme with $\dfrac{\partial}{\partial x_2}\delta(x)$ source term and approximation $\delta_{\epsilon_1}^0(x_1)\delta_{\epsilon_2}^1(x_2) \in Q_2^0(h) \times Q_2^1(h)$.
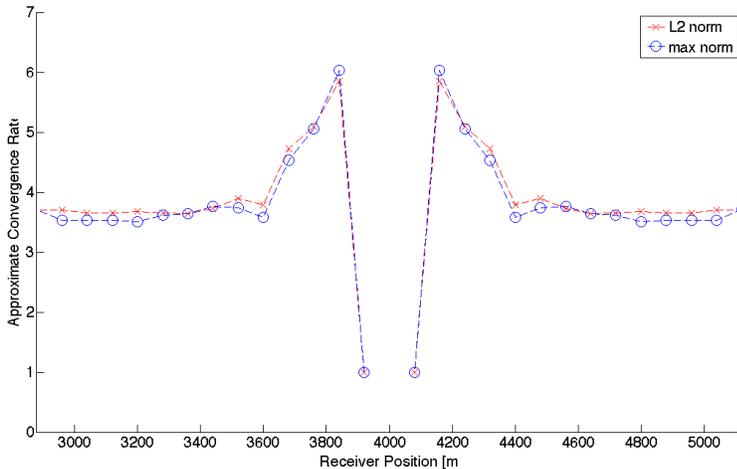
Figure 4: Rates for 2-4 staggered grid scheme with $\dfrac{\partial}{\partial x_2}\delta(\boldsymbol{x})$ source term and approximation $\delta_{\epsilon_1}^0(x_1)\delta_{\epsilon_2}^1(x_2) \in Q_4^0(h) \times Q_4^1(h)$.
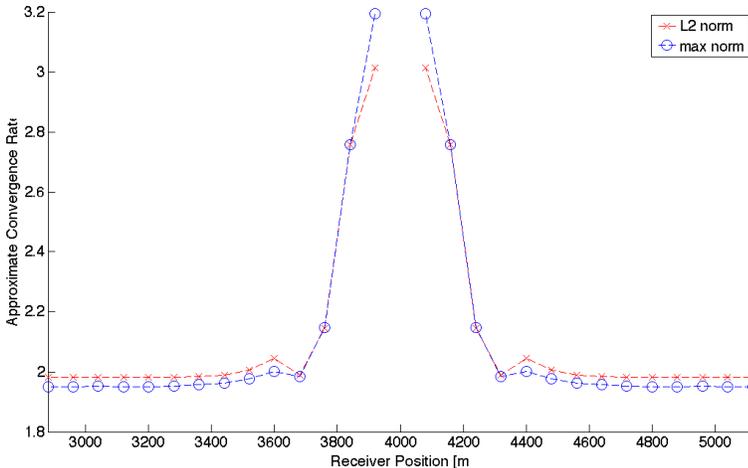

Figure 5: Rates for 2-4 staggered grid scheme with $\dfrac{\partial}{\partial x_2}\delta(\boldsymbol{x})$ source term and approximation $\delta_{\epsilon_1}^0(x_1)\delta_{\epsilon_2}^1(x_2) \in Q_2^0(h) \times Q_2^1(h)$.
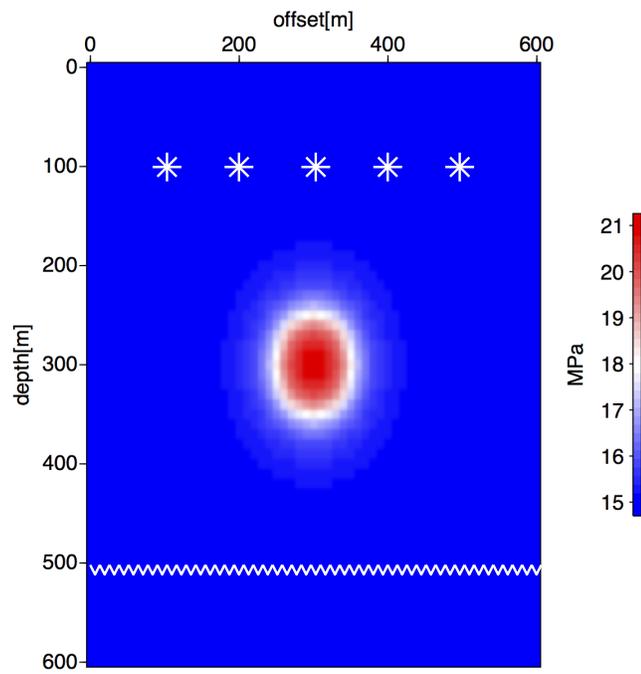
Figure 6: True bulk modulus model and source-receiver configuration; ∗-source, ∨-receiver.
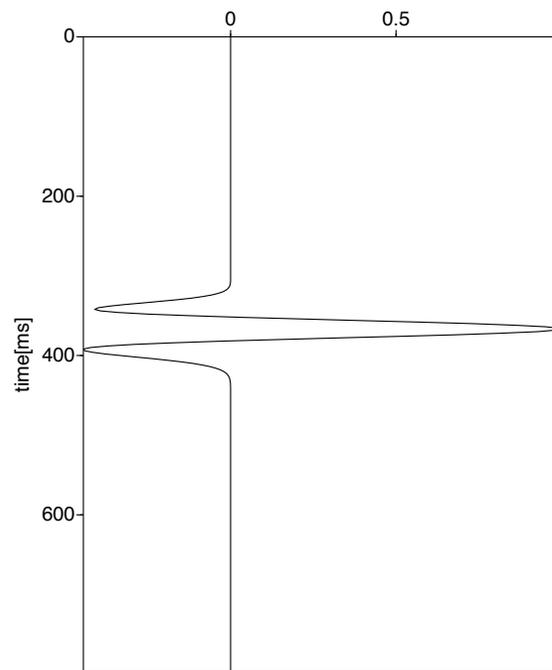
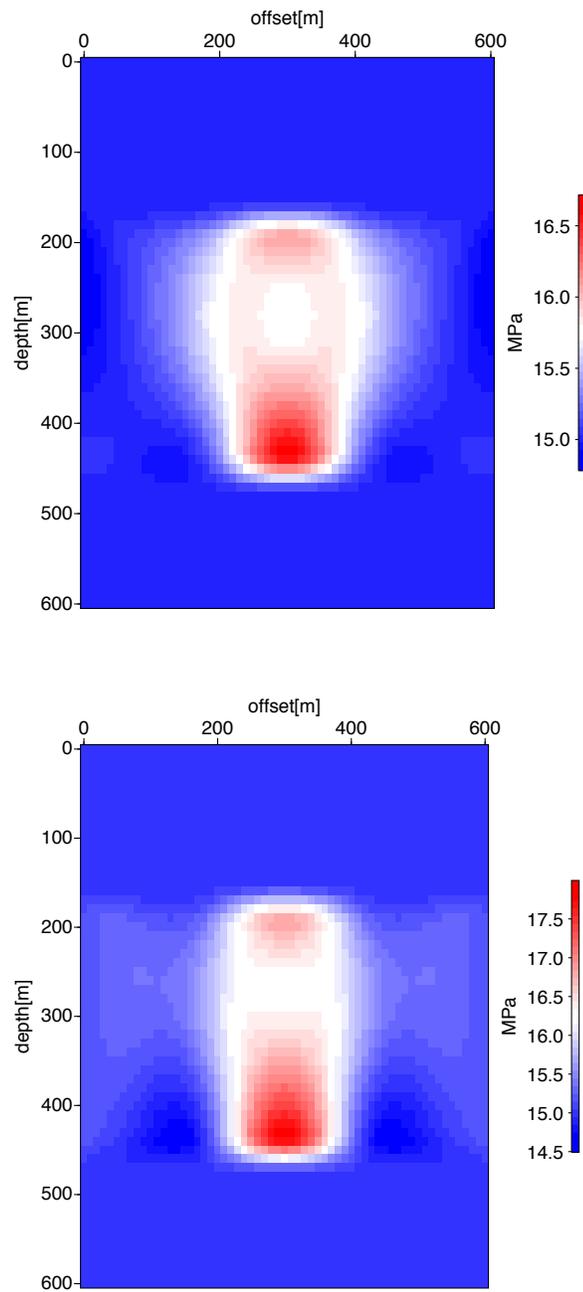Figure 7: True source-time function.

Figure 8: Recovered bulk modulus model for transmission test case via OLS inversion with known source (a) and VP inversion (b).
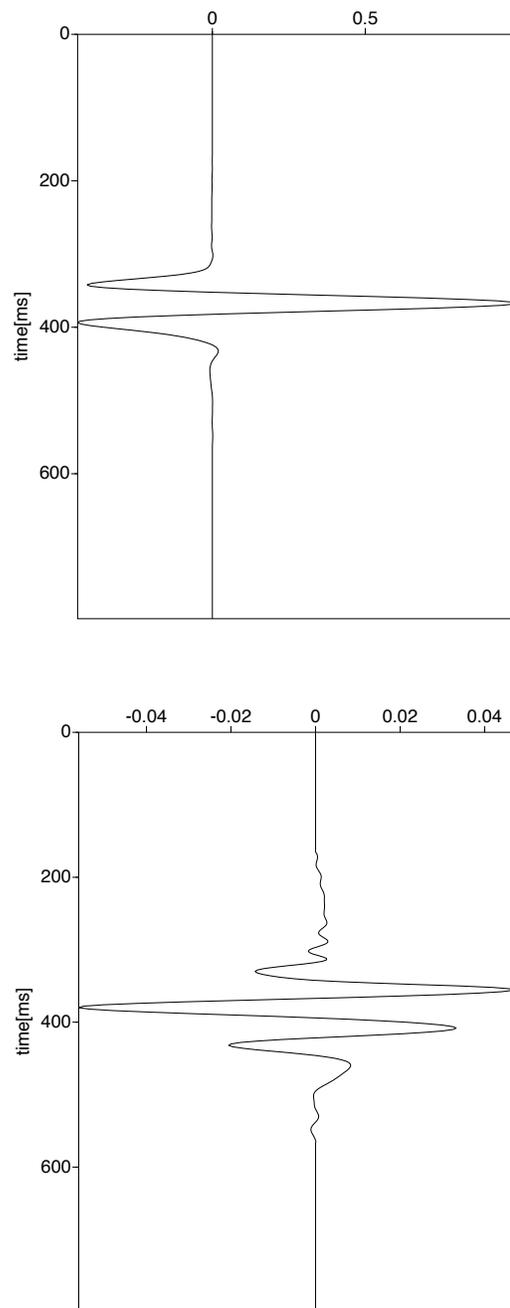
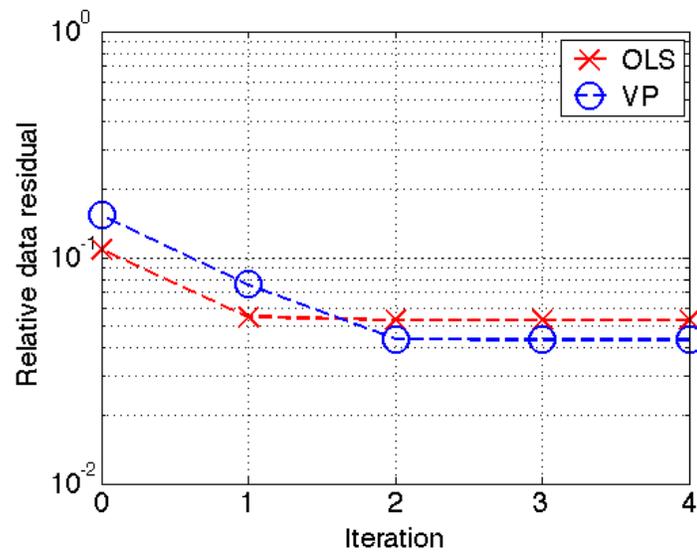Figure 9: Recovered source-time function from VP (a) and its error (b).

Figure 10: Relative data residual per iteration for transmission test case.

Stable Composite Staggered Grid Finite Difference Scheme for Anisotropic Elastic Wave Simulations

Muhong Zhou

ABSTRACT

Based on the energy method, this paper proposes a stable composite staggered grid scheme for anisotropic (up to orthorhombic) elastic wave simulations. To conserve the energy and agree with the transmission condition, the paper derive a rule to update the ghost data at the grid refinement interface. The paper also derive an upper bound for the time step size and show that with the energy-conserving ghost data update rule, the scheme is stable. 1D numerical results demonstrate that both energy-conserving ghost data update rule and the upper bound imposed on the time step size are indispensable to achieve stability.

INTRODUCTION

Many seismic imaging applications used for oil and gas explorations need a fast and stable wave simulation solver. Staggered grid finite difference scheme (Virieux, 1986; Levander, 1988) is often chosen to simulate wave propagations because it is easy to implement and cost effective. Since finite difference step sizes, both spatially and temporally, are restricted by the shortest wavelength in the computing domain, when the simulation domain is composed of subdomains having different characteristic wavelengths, it is desirable to use a composite grid scheme (Gustafsson and Mossberg, 2004; Rodríguez, 2008; Petersson and Sjögreen, 2010; Kristek et al., 2010), i.e., coarser grid for subdomain with longer wavelength (or faster wave propagation speed) and finer grid for subdomain with shorter wavelength (or slower wave speed). A practical example of such domain is in the marine seismic explorations, the P-wave in the water is around 1500m/s, but the S-wave near the ocean bottom can be as low as 100m/s (Fehler, 2012). Another application of

using these composite grid schemes is to resolve local fine-scale diffraction patterns (Rodríguez, 2008). In this paper, I focus on building a composite staggered grid finite difference scheme that is based on first-order hyperbolic wave equation and numerically stable.

My work is mainly based on Petersson and Sjögreen (2010)'s work where they use the energy method to propose a stable composite collocated grid finite difference scheme based on isotropic elastic wave equation in second-order displacement form. They derive a rule to update the interface ghost data by conserving the discrete energy of the scheme over time and by aligning with the transmission condition across the refinement interface. With this rule, they claim that their scheme is stable; however, as I will show later, energy conservation along does not lead to a stable scheme, it needs to meet another stability qualification that I call *boundedness* and this qualification sets an upper bound for the time step size. Their paper (Petersson and Sjögreen, 2010) does not mention how they choose the time step size, but in a previous paper (Petersson and Sjögreen, 2009) they gave an upper bound for the time step size for a uniform collocated grid finite difference scheme; nevertheless how to extend this uniform grid bound to the composite grid case is not clear to me. My work extends their energy-conserving ghost data update rule based on a composite collocated grid finite difference scheme to the composite staggered grid finite difference scheme. Compared with the former scheme, which is based on the isotropic elastic wave equation in second-order displacement form, the latter scheme not only builds upon staggered grids but also is based on the anisotropic (up to orthorhombic) elastic wave equation in first-order hyperbolic form expressed in terms of velocities and stresses; therefore the latter scheme needs a different and more complex discrete energy formulation and a new rule to update the interface ghost data.

To formulate this new discrete energy, I consult Rodríguez (2008)'s work where they build a discrete energy for a composite grid mass-lumping finite element scheme based on the wave equation in first-order hyperbolic form. By conserving this discrete energy they derive an interface condition imposed on traces of velocities and stresses on the interface. They also prove that their scheme will be stable if the time step size satisfies a certain condition. Since they work with a finite element scheme, I re-formulate their discrete energy to suit finite difference scheme via using quadrature rules as they were used in Petersson and Sjögreen (2010) for collocated grid finite difference scheme.

The rest of the paper is organized as follows: in section two I use the energy

method to identify two stability qualifications for the 1D composite staggered grid finite difference scheme based on acoustic wave equation, which yield the energy-conserving interpolation to update the ghost data and pose an upper limit for the time step size respectively; in section three, I obtain a stable 3D composite grid scheme by generalizing the energy-conserving ghost data interpolation and the upper bound for time step size from 1D to 3D case; section four contains 1D numerical results which show that both the energy-conserving ghost data interpolation and restricting the time step size are indispensable to obtain stability; in the last section, I conclude this paper and discuss the future work.

1D STABLE COMPOSITE GRID SCHEME

1D acoustic wave propagation along the x direction is governed by the following equation system:

$$\begin{cases} \rho(x)v_t(x,t) = p_x(x,t) + s(x,t) \\ p_t(x,t) = \kappa(x)v_x(x,t) \end{cases} \tag{1}$$

where v is the particle velocity, p is the pressure, ρ is the density, κ is the bulk modulus and s is the source function.

Consider a hypothetical situation when there exists a material discontinuity along the x direction as shown in Fig.(1): the wave propagation speed is faster in the right subdomain. Accordingly, I construct a composite staggered grid scheme with spacing h for the left half-line and $H(> h)$ for the right: that is

$$\begin{cases} v((i-N_i)h, (n+\frac{1}{2})\Delta t) \mapsto v_i^{n+\frac{1}{2}} & \text{for } i = 0, ..., N_i, \\ v(iH, (n+\frac{1}{2})\Delta t) \mapsto V_i^{n+\frac{1}{2}} & \text{for } i = 0, ..., N_I \end{cases} \tag{2}$$

Pressures live on the dual grid at integer times:

$$\begin{cases} p((i+\frac{1}{2}-N_i)h, n\Delta t) \mapsto p_{i+\frac{1}{2}}^n & \text{for } i = 0, ..., N_i - 1, \\ p((i+\frac{1}{2})H, n\Delta t) \mapsto P_{i+\frac{1}{2}}^n & \text{for } i = 0, ..., N_I - 1. \end{cases} \tag{3}$$

Source function and densities are configured on the same grid locations as veloc-

ities, and moduli are with pressures:

$$
\begin{cases}
\rho((i - N_i)h) \mapsto \rho_i^-, & s((i - N_i)h, (n + \frac{1}{2})\Delta t) \mapsto s_i^{n+\frac{1}{2}} & \text{for } i = 0, ..., N_i, \\
\rho(iH) \mapsto \rho_i^+, & s(iH, (n + \frac{1}{2})\Delta t) \mapsto S_i^{n+\frac{1}{2}} & \text{for } i = 0, ..., N_I
\end{cases}
\tag{4}
$$

$$
\begin{cases}
\kappa((i + \frac{1}{2} - N_i)h) \mapsto \kappa_{i+\frac{1}{2}}^- & \text{for } i = 0, ..., N_i - 1, \\
\kappa((i + \frac{1}{2})H) \mapsto \kappa_{i+\frac{1}{2}}^+ & \text{for } i = 0, ..., N_I - 1.
\end{cases}
\tag{5}
$$

Hence the grid refinement interface coincides with the grid location with velocities $(v_{N_i}^{n+\frac{1}{2}}, V_0^{n+\frac{1}{2}})$, and two ends of the composite grid also coincide with the velocity locations $(v_0^{n+\frac{1}{2}}, V_{N_I}^{n+\frac{1}{2}})$; whereas all the pressure data live on the interior points of each grid. This design eases the implementation of the homogeneous Dirichlet boundary condition which requires zero velocities on the boundary:

$$
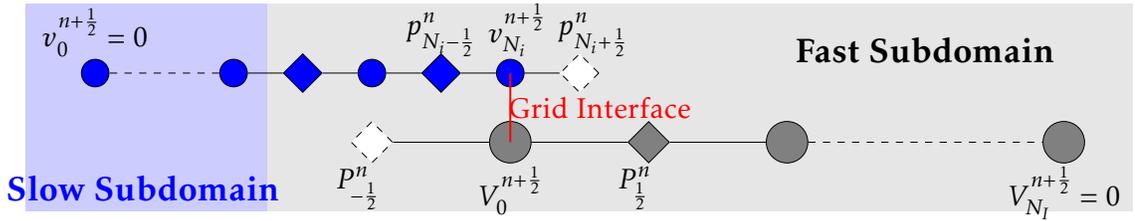v_0^{n+\frac{1}{2}} = V_{N_I}^{n+\frac{1}{2}} = 0.
\tag{6}
$$

Figure 1: A composite staggered grid finite difference scheme (H=2h) in the presence of material discontinuity. **Slow/fast Subdomain** denotes the subdomain in which the wave propagation speed is slow/fast. The region of each subdomain is marked by the blue/gray shade. Blue/grey circles and diamonds represent locations of fine/coarse grid variables with variable names besides them. The red line delineates the grid refinement interface, which is placed inside the fast subdomain. Two white dashed diamonds denote the ghost data.

On the fine grid, the 2-2 (second-order accurate in both time and space) finite difference scheme for this equation system is:

$$
\begin{cases}
\rho_i^- D_{-,t} v_i^{n+\frac{1}{2}} = D_{-,x} p_{i+\frac{1}{2}}^n + s_i^n \\
D_{+,t} p_{i+\frac{1}{2}}^n = \kappa_{i+\frac{1}{2}}^- D_{+,x} v_i^{n+\frac{1}{2}}
\end{cases}
\tag{1D.FDM}
$$

where $D_{+,\cdot}/D_{-,\cdot}$ is the second-order forward/backward finite difference operator along a particular direction, e.g.,

$$
\begin{cases}
D_{+,t} p_{i+\frac{1}{2}}^n = \dfrac{p_{i+\frac{1}{2}}^{n+1} - p_{i+\frac{1}{2}}^n}{\Delta t} \\
D_{-,t} v_i^{n+\frac{1}{2}} = \dfrac{v_i^{n+\frac{1}{2}} - v_i^{n-\frac{1}{2}}}{\Delta t}
\end{cases}
\quad , \quad
\begin{cases}
D_{-,x} p_{i+\frac{1}{2}}^n = \dfrac{p_{i+\frac{1}{2}}^n - p_{i-\frac{1}{2}}^n}{h} \\
D_{+,x} v_i^{n+\frac{1}{2}} = \dfrac{v_{i+1}^{n+\frac{1}{2}} - v_i^{n+\frac{1}{2}}}{h}
\end{cases}
\tag{7}
$$

The 2-2 scheme on the coarse grid is built similarly.

To see how this scheme iterates over time, first suppose all velocities are updated to time level $n + \frac{1}{2}$, pressures to time level $n - 1$. Subsequently, pressures can be updated to time level n uisng Eq.(1D.FDM). However, when using Eq.(1D.FDM) to update velocities from time level $n + \frac{1}{2}$ to time level $n + \frac{3}{2}$, updating interface velocities ($v_{N_i}^{n+\frac{1}{2}}$, $V_0^{n+\frac{1}{2}}$) requires $p(\frac{1}{2}h, n\Delta t)$ and $p(-\frac{1}{2}H, n\Delta t)$, therefore I construct two ghost pressure data near the interface:

$$
p(\tfrac{1}{2}h, n\Delta t) \mapsto p_{N_i+\frac{1}{2}}^n, \quad p(-\tfrac{1}{2}H, n\Delta t) \mapsto P_{-\frac{1}{2}}^n.
\tag{8}
$$

Once we know these ghost data values, interface velocities can be updated using the same finite difference scheme as it is used to update interior velocities and then the iteration can proceed.

An intuitive approach to update these ghost data is by interpolation from existing pressure values in the neighborhood, for example when $H = 2h$,

$$
\begin{cases}
P_{-\frac{1}{2}}^n = 0.5(p_{N_i-\frac{3}{2}}^n + p_{N_i-\frac{1}{2}}^n) \\
p_{N_i+\frac{1}{2}}^n = 0.15625 P_{-\frac{1}{2}}^n + 0.9375 P_{\frac{1}{2}}^n - 0.09375 P_{\frac{3}{2}}^n
\end{cases}
\tag{9}
$$

Above two equations have to be executed in sequence since the input to the second equation relies on the output from the first one. When the solution is smooth

near the interface, according to the Taylor's Theorem these equations are at least second-order accurate to the true solution at these locations. In this paper, I call the composite grid scheme using this intuitive ghost data update rule as *the intuitive composite grid scheme*.

As shown in Fig.(1), the grid refinement interface is placed inside the fast material. If the interface is aligned with the material continuity, then the ghost data will be interpolated from pressure data attributed to a different subdomain. For example, $P_{-\frac{1}{2}}^{n}$ for the fast subdomain will be updated using $p_{N_i-\frac{3}{2}}^{n}, p_{N_i-\frac{1}{2}}^{n}$ from the slow subdomain. Since the fine grid can handle both subdomains whereas the coarse grid can handle the fast subdomain only, it is better to place the interface inside the fast material so that each ghost data comes from the same subdomain as its interpolation source.

However, notice that this intuitive composite grid scheme has two defects:

- First of all, even if the solution is smooth near the interface and the intuitive interpolation approach gives a good approximation to the ghost data, it is hard to analyze the stability of the scheme.

- Second, the intuitive composite grid scheme may not be consistent with the physical transmission condition which requires the velocity and pressure are continuous across the interface.

In the rest of this section, I will first present the energy method in the 1D continuous case and identify two qualifications of the energy that contribute to stability. Then I will use the 1D energy formulated in the continuous case as a guide to formulate a discrete energy on the composite staggered grid, and show that this discrete energy will have those two stability qualifications if (1) updating the ghost data by conserving the energy and meanwhile aligning with the transmission condition, and (2) setting an upper limit for the time step size.

Energy Method in 1D Continuous Case

Consider solving problem Eq.(1) on $v(x,t), p(x,t) \in C^1(\mathbb{R} \times [0, \infty))$ and suppose they vanish as x goes to infinity. The energy $\mathcal{E}$ of this acoustic wave equation system at

time t is often formulated as:

$$\mathcal{E}[v,p](t) = \frac{1}{2}\int_{\mathbb{R}} \rho(x)v^2(x,t)\mathrm{d}x + \frac{1}{2}\int_{\mathbb{R}} \frac{1}{\kappa(x)}p^2(x,t)\mathrm{d}x \qquad (10)$$

The first term of this formula is often called the *kinetic energy*, and the second is called the *strain energy*. Since $\rho(x), \kappa(x) > 0$, there exist constants $c_1(\rho,\kappa), c_2(\rho,\kappa), C_1(\rho,\kappa), C_2(\rho,\kappa) > 0$ so that

$$c_1\|v(x,t)\| + c_2\|p(x,t)\| \leq \mathcal{E}[v,p](t) \leq C_1\|v(x,t)\| + C_2\|p(x,t)\| \qquad \text{(boundedness)}$$

where the L^2-norm $\|\cdot\|$ is evaluated in the x-domain. *Boundedness* refers to grid functions and the energy are mutually controlled by each other.

If Eq.(1) is in its homogeneous form, i.e., when $s(x,t) = 0$,

$$\frac{\mathrm{d}\mathcal{E}[v,p](t)}{\mathrm{d}t} = \int_{\mathbb{R}} \rho(x)v(x,t)v_t(x,t)\mathrm{d}x + \int_{\mathbb{R}} \frac{1}{\kappa(x)}p(x,t)p_t(x,t)\mathrm{d}x$$

$$\overset{Eq.(1)}{=} \int_{\mathbb{R}} v(x,t)p_x(x,t)\mathrm{d}x + \int_{\mathbb{R}} p(x,t)v_x(x,t)\mathrm{d}x \overset{IBP}{=} 0 \qquad \text{(conservation)}$$

where IBP is short for "integration by parts". *Conservation* means that the scheme maintains a constant energy over time.

If perturb the initial solution $(v(x,0), p(x,0))$ by a small amount, and denote the new initial solution by $(\hat{v}(x,0), \hat{p}(x,0))$. If both of them are subject to the same source term and boundary condition, then the error in the later solution $((\hat{v} - v)(x,t), (\hat{p} - p)(x,t))$ satisfies the homogeneous form of Eq.(1), and consequently Eq.(boundedness) and Eq.(conservation) together yield that

$$c_1\|(\hat{v} - v)(x,t)\| + c_2\|(\hat{p} - p)(x,t)\| \leq C_1\|(\hat{v} - v)(x,0)\| + C_2\|(\hat{p} - p)(x,0)\| \qquad (11)$$

i.e., the error in the later solution will be controlled by the initial data perturbation; hence Eq.(1) is stable.

Therefore in the 1D continuous case, *boundedness* and *conservation* together lead to stability. Note that when deriving stability we assume that two solutions are subject to the same source term and boundary condition, so we are looking for these two stability qualifications only for equation or scheme in its homogeneous

form; however, the conditions leading to these qualifications may be imposed on non-homogeneous scheme.

Lastly, above analysis can be applied to 3D case and it is easy to see that stability qualifications in the 3D case are the same as those in the 1D case.

Formulate Discrete Energy on the 1D Composite Grid

Unlike the continuous case, the composite grid scheme works on a truncated domain, and different grid functions are defined on different time levels. To accommodate these facts, I combine features from the energy formulated for composite grid finite element scheme based on wave equation in first-order hyperbolic form (Bécache et al., 2001) and the one for composite collocated grid finite difference scheme based on wave equation in second-order displacement form (Petersson and Sjögreen, 2010), and come up with a new energy formulation for the composite staggered grid finite difference scheme based on wave equation in first-order hyperbolic form.

Define the discrete kinetic energy $\mathcal{K}$ and strain energy $\mathcal{S}$ on the fine grid at time level n as:

$$\mathcal{K}_f^n[v] = \frac{1}{2} \sum_{i=0}^{N_i} \alpha_i h \rho_i^- v_i^{n+\frac{1}{2}} v_i^{n-\frac{1}{2}} \approx \int_{-N_i h}^0 \rho(x) v(x, (n+\frac{1}{2})\Delta t) v(x, (n-\frac{1}{2})\Delta t) \mathrm{d}x \quad (12)$$

$$\mathcal{S}_f^n[p] = \frac{1}{2} \sum_{i=0}^{N_i-1} h \frac{1}{\kappa_{i+\frac{1}{2}}^-} (p_{i+\frac{1}{2}}^n)^2 \approx \int_{-N_i h}^0 \frac{1}{\kappa(x)} p^2(x, n\Delta t) \mathrm{d}x \quad (13)$$

where the subscript "f" by the energies is short for "fine grid", and the quadrature weights are defined as follows:

$$\alpha_i h = \begin{cases} \dfrac{1}{2}h, & i = 0, N_i \\ h, & 0 < i < N_i \end{cases} \quad (14)$$

so Eq.(12) approximates the integral on the right using exactly the Trapezoid Rule, and Eq.(13) uses the Midpoint Rule.

The sum of these two energies defines the discrete total energy $\mathcal{E}$ on the fine grid. Define the kinetic, strain and total energy on the coarse grid in a similar way,

then the sum of the total energies on both grids defines the total energy on the composite grid:

$$\mathcal{E}_c^n[v,p,V,P] = \mathcal{E}_f^n[v,p] + \mathcal{E}_r^n[V,P] \tag{15}$$

where "c" stands for composite grid, "r" for coarse grid.

Qualify for *Conservation*

In the discrete case, *conservation* means

$$\mathcal{E}_c^n[v,p,V,P] = \mathcal{E}_c^{n-1}[v,p,V,P] \tag{16}$$

So next I will show how to update the ghost data in order to achieve this stability qualification.

To simplify derivations afterwards, first I will define some notations. On the fine grid, define the following 1D scalar inner product and norm for grid functions u,v defined on the primary grid:

$$\langle v,u \rangle_{fp} = \frac{1}{2} \sum_{i=0}^{N_i} \alpha_i h v_i u_i \tag{17}$$

$$\|v\|_{fp} = [\frac{1}{2} \sum_{i=0}^{N_i} \alpha_i h (u_i)^2]^{\frac{1}{2}} \tag{18}$$

where the subscript "fp" is the shorthand of "the primary grid of the fine staggered grid".

Similarly, the inner product and norm for grid functions p,q defined on the dual grid are:

$$\langle p,q \rangle_{fd} = \frac{1}{2} \sum_{i=0}^{N_i-1} h p_{i+\frac{1}{2}} q_{i+\frac{1}{2}} \tag{19}$$

$$\|p\|_{fd} = [\frac{1}{2} \sum_{i=0}^{N_i-1} h (p_{i+\frac{1}{2}})^2]^{\frac{1}{2}} \tag{20}$$

Analogous to the IBP property in the continuous case, these discrete inner products satisfy the following summation by parts (SBP) property:

$$\langle v, D_{-,x}p\rangle_{fp} + \langle D_{+,x}v, p\rangle_{fd} = v_{N_i}\overline{p_{N_i}} - v_0\overline{p_0} \qquad \text{(1D.SBP)}$$

for v on primary grid, p on dual grid, $\overline{p_{N_i}}$ being $\dfrac{1}{2}(p_{N_i+\frac{1}{2}} + p_{N_i-\frac{1}{2}})$, $\overline{p_0}$ being $\dfrac{1}{2}(p_{\frac{1}{2}} + p_{-\frac{1}{2}})$.

Consequently, for a homogeneous scheme (i.e., source term is zero) imposed with the homogeneous Dirichlet boundary condition (i.e, $v_0^{n+\frac{1}{2}} = 0$), we have

$$
\begin{aligned}
\mathcal{K}_f^n[v] - \mathcal{K}_f^{n-1}[v] &= \langle \rho^- v^{n+\frac{1}{2}}, v^{n-\frac{1}{2}}\rangle_{fp} - \langle \rho^- v^{n-\frac{1}{2}}, v^{n-\frac{3}{2}}\rangle_{fp}\\
&= \Delta t\langle v^{n-\frac{1}{2}}, \rho^- D_{-,t}(v^{n+\frac{1}{2}} + v^{n-\frac{1}{2}})\rangle_{fp}\\
&\overset{Eq.(1D.FDM)}{=} \Delta t\langle v^{n-\frac{1}{2}}, D_{-,x}(p^n + p^{n-1})\rangle_{fp}\\
&\overset{SBP}{=} -\Delta t\langle D_{+,x}v^{n-\frac{1}{2}}, p^n + p^{n-1}\rangle_{fd} + \Delta t v_{N_i}^{n-\frac{1}{2}}(\overline{p_{N_i}^n} + \overline{p_{N_i}^{n-1}})\\
&\overset{Eq.(1D.FDM)}{=} -\langle \frac{1}{\kappa^-}(p^n - p^{n-1}), p^n + p^{n-1}\rangle_{fd} + \Delta t v_{N_i}^{n-\frac{1}{2}}(\overline{p_{N_i}^n} + \overline{p_{N_i}^{n-1}})\\
&= \mathcal{S}_f^{n-1}[p] - \mathcal{S}_f^n[p] + \Delta t v_{N_i}^{n-\frac{1}{2}}(\overline{p_{N_i}^n} + \overline{p_{N_i}^{n-1}})
\end{aligned}
\tag{21}
$$

i.e.

$$\mathcal{E}_f^n[v,p] - \mathcal{E}_f^{n-1}[v,p] = \Delta t v_{N_i}^{n-\frac{1}{2}}(\overline{p_{N_i}^n} + \overline{p_{N_i}^{n-1}}) \tag{22}$$

On the coarse grid I establish a similar result:

$$\mathcal{E}_r^n[V,P] - \mathcal{E}_r^{n-1}[V,P] = -\Delta t V_0^{n-\frac{1}{2}}(\overline{P_0^n} + \overline{P_0^{n-1}}) \tag{23}$$

Therefore, the total energy difference between adjacent time levels are:

$$\mathcal{E}_c^n[v,pV,P] - \mathcal{E}_c^{n-1}[v,p,V,P] = \Delta t v_{N_i}^{n-\frac{1}{2}}(\overline{p_{N_i}^n} + \overline{p_{N_i}^{n-1}}) - \Delta t V_0^{n-\frac{1}{2}}(\overline{P_0^n} + \overline{P_0^{n-1}}) \tag{24}$$

From the above equation, one may observe that if the interface velocities and the averaged interface pressures meet the following condition for arbitrary nonnegative m, n, then the energy will be conserved over time.

$$\Delta t v_{N_i}^{n-\frac{1}{2}}\overline{p_{N_i}^m} = \Delta t V_0^{n-\frac{1}{2}}\overline{P_0^m} \qquad \text{(1D.ECI)}$$

The transmission condition requires the velocity is continuous across the interface, which leads to the first condition:

$$v_{N_i}^{n+\frac{1}{2}} = V_0^{n+\frac{1}{2}}, \quad n \geq 0 \tag{1D.ECI.a}$$

Substituting this equation back into (1D.ECI) leads to another equation

$$\overline{p_{N_i}^m} = \overline{P_0^m}, \quad n \geq 0 \tag{1D.ECI.b}$$

Interestingly, this second condition is consistent with the other requirement of the transmission condition that the pressure has to be continuous across the refinement interface.

Note that although Eq.(1D.ECI.a) and Eq.(1D.ECI.b) are derived on homogeneous scheme, they can be imposed on non-homogeneous scheme as well. Once a non-homogeneous scheme satisfies these two conditions, then its homogeneous counterpart automatically satisfies these two conditions which in turn leads to the first stability qualification: *conservation*.

Eq.(1D.ECI.a) and (1D.ECI.b) together provide a linear equation system solving for the ghost data at any time level. For example, to solve $P_{-\frac{1}{2}}^n$ and $p_{N_i+\frac{1}{2}}^n$, note that Eq.(1D.ECI.b) explicitly depends on them when $m = n$. On the other hand, according to Eq.(1D.FDM), Eq.(1D.ECI.a) is equivalent to

$$v_{N_i}^{n-\frac{1}{2}} + \frac{\Delta t}{\rho_{N_i}^-}(D_{-,x}p_{N_i+\frac{1}{2}}^n + s_{N_i}^{n-\frac{1}{2}}) = V_0^{n-\frac{1}{2}} + \frac{\Delta t}{\rho_0^+}(D_{-,x}P_{\frac{1}{2}}^n + S_0^{n-\frac{1}{2}}). \tag{25}$$

Thus the coefficient matrix of this linear equation system is

$$\begin{pmatrix} \dfrac{1}{2} & -\dfrac{1}{2} \\ \dfrac{\Delta t}{h\rho_{N_i}^-} & \dfrac{\Delta t}{H\rho_0^+} \end{pmatrix} \tag{26}$$

Its determinant is positive, so the system is solvable.

In this paper, the ghost data update rule provided by Eq.(1D.ECI.a) and Eq.(1D.ECI.b) is called *energy conserving interpolation*, and the composite grid scheme using this ghost data update rule is called *stable composite grid scheme*.

Qualify for *Boundedness*

Now the *stable composite grid scheme* is not necessarily stable: *boundedness* has not been guaranteed yet. To achieve *boundedness*, note that

$$
\begin{aligned}
\mathcal{K}_c^n[v, V] =& \langle \rho^- v^{n-\frac{1}{2}}, v^{n+\frac{1}{2}} \rangle_{fp} + \langle \rho^+ V^{n-\frac{1}{2}}, V^{n+\frac{1}{2}} \rangle_{rp} \\
=& \langle \rho^- v^{n-\frac{1}{2}}, v^{n+\frac{1}{2}} - v^{n-\frac{1}{2}} \rangle_{fp} + \langle \rho^+ V^{n-\frac{1}{2}}, V^{n+\frac{1}{2}} - V^{n-\frac{1}{2}} \rangle_{rp} \\
& + \langle \rho^- v^{n-\frac{1}{2}}, v^{n-\frac{1}{2}} \rangle_{fp} + \langle \rho^+ V^{n-\frac{1}{2}}, V^{n-\frac{1}{2}} \rangle_{rp}
\end{aligned}
\tag{27}
$$

where

$$
\begin{aligned}
& |\langle \rho^- v^{n-\frac{1}{2}}, v^{n+\frac{1}{2}} - v^{n-\frac{1}{2}} \rangle_{fp} + \langle \rho^+ V^{n-\frac{1}{2}}, V^{n+\frac{1}{2}} - V^{n-\frac{1}{2}} \rangle_{rp}| \\
\overset{Eq.(1D.FDM)}{=}& |\Delta t \langle v^{n-\frac{1}{2}}, D_{-,x} p^n \rangle_{fp} + \Delta t \langle V^{n-\frac{1}{2}}, D_{-,x} P^n \rangle_{rp}| \\
\overset{\substack{(1D.SBP), \\ (1D.ECI.a),(1D.ECI.b)}}{=}& |-\Delta t \langle D_{+,x} v^{n-\frac{1}{2}}, p^n \rangle_{fd} - \Delta t \langle D_{+,x} V^{n-\frac{1}{2}}, P^n \rangle_{rd}| \\
\le& \frac{\Delta t}{h} (\gamma_1 \|p^n\|_{fd}^2 + \frac{1}{\gamma_1} \|v^{n-\frac{1}{2}}\|_{fp}^2) + \frac{\Delta t}{H} (\gamma_2 \|P^n\|_{rd}^2 + \frac{1}{\gamma_2} \|V^{n-\frac{1}{2}}\|_{rp}^2)
\end{aligned}
\tag{28}
$$

for any $\gamma_1, \gamma_2 > 0$ and their optimal values are left to be determined afterwards.

As a result,

$$
\begin{aligned}
\mathcal{E}_c^n[v, p, V, P] \le& \left(\langle \frac{1}{\kappa^-} p^n, p^n \rangle_{fd} + \frac{\gamma_1 \Delta t}{h} \|p^n\|_{fd}^2 \right) + \left(\langle \frac{1}{\kappa^+} P^n, P^n \rangle_{rd} + \frac{\gamma_2 \Delta t}{H} \|P^n\|_{rd}^2 \right) \\
& + \left(\langle \rho^- v^{n-\frac{1}{2}}, v^{n-\frac{1}{2}} \rangle_{fp} + \frac{\Delta t}{h \gamma_1} \|v^{n-\frac{1}{2}}\|_{fp}^2 \right) + \left(\langle \rho^+ V^{n-\frac{1}{2}}, V^{n-\frac{1}{2}} \rangle_{rp} + \frac{\Delta t}{H \gamma_2} \|V^{n-\frac{1}{2}}\|_{rp}^2 \right)
\end{aligned}
\tag{29}
$$

and

$$
\begin{aligned}
\mathcal{E}_c^n[v, p, V, P] \ge& \left(\langle \frac{1}{\kappa^-} p^n, p^n \rangle_{fd} - \frac{\gamma_1 \Delta t}{h} \|p^n\|_{fd}^2 \right) + \left(\langle \frac{1}{\kappa^+} P^n, P^n \rangle_{rd} - \frac{\gamma_2 \Delta t}{H} \|P^n\|_{rd}^2 \right) \\
& + \left(\langle \rho^- v^{n-\frac{1}{2}}, v^{n-\frac{1}{2}} \rangle_{fp} - \frac{\Delta t}{h \gamma_1} \|v^{n-\frac{1}{2}}\|_{fp}^2 \right) + \left(\langle \rho^+ V^{n-\frac{1}{2}}, V^{n-\frac{1}{2}} \rangle_{rp} - \frac{\Delta t}{H \gamma_2} \|V^{n-\frac{1}{2}}\|_{rp}^2 \right)
\end{aligned}
\tag{30}
$$

Eq.(29) implies the energy is controlled by the grid functions:

$$
\mathcal{E}_c^n[v, p, V, P] \le C_p \|p^n\|_{fd}^2 + C_P \|P^n\|_{rd}^2 + C_v \|v^{n-\frac{1}{2}}\|_{fp}^2 + C_V \|V^{n-\frac{1}{2}}\|_{rp}^2
\tag{31}
$$

for some constants $C_p, C_P, C_v, C_V > 0$, which are only determined by the densities, moduli, Δt, h, γ_1 and γ_2.

Let $\kappa_{\max}^- = \max\limits_{0 \le i \le N_i} \kappa_i^-$, and $\kappa_{\max}^+, \rho_{\min}^-, \rho_{\min}^+$ are defined similarly. If the time step size satisfies

$$\Delta t < \min\left\{ \frac{h}{\kappa_{\max}^- \gamma_1}, \frac{H}{\kappa_{\max}^+ \gamma_2}, \gamma_1 \rho_{\min}^- h, \gamma_2 \rho_{\min}^+ H \right\} \tag{1D.dt}$$

then the term in every parenthesis of Eq.(30) will be greater than a grid function norm multiplied by a constant factor, i.e.,

$$\mathcal{E}_c^n[v, p, V, P] \ge c_p \|p^n\|_{fd}^2 + c_P \|P^n\|_{rd}^2 + c_v \|v^{n-\frac{1}{2}}\|_{fp}^2 + c_V \|V^{n-\frac{1}{2}}\|_{rp}^2 \tag{32}$$

for some constants $c_p, c_P, c_v, c_V > 0$, which are only determined by the densities, moduli, Δt, h, γ_1 and γ_2; hence the grid functions are controlled by the energy.

With *conservation* and *boundedness* (Eq.(31) and Eq.(32)), for any initial perturbation: $(v^{-\frac{1}{2}}, p^0, V^{-\frac{1}{2}}, P^0) \to (\hat{v}^{-\frac{1}{2}}, \hat{p}^0, \hat{V}^{-\frac{1}{2}}, \hat{P}^0)$, we have

$$\begin{aligned} c_p \|p^n - \hat{p}^n\|_{fd}^2 + c_P \|P^n - \hat{P}^n\|_{rd}^2 + c_v \|v^{n-\frac{1}{2}} - \hat{v}^{n-\frac{1}{2}}\|_{fp}^2 + c_V \|V^{n-\frac{1}{2}} - \hat{V}^{n-\frac{1}{2}}\|_{rp}^2 \le \\ C_p \|p^0 - \hat{p}^0\|_{fd}^2 + C_P \|P^0 - \hat{P}^0\|_{rd}^2 + C_v \|v^{-\frac{1}{2}} - \hat{v}^{-\frac{1}{2}}\|_{fp}^2 + C_V \|V^{-\frac{1}{2}} - \hat{V}^{-\frac{1}{2}}\|_{rp}^2 \end{aligned} \tag{33}$$

Hence the scheme is stable if ghost data are updated according to Eq.(1D.ECI.a) and Eq.(1D.ECI.b), and the time step size is constrained by Eq.(1D.dt).

Finally, to reduce computing time, γ_1, γ_2 should be selected to make Δt as large as possible.

3D STABLE COMPOSITE GRID SCHEME

Elastic wave propagation in a 3D anisotropic material is governed by two equations:

$$\begin{aligned} \text{Conservation of Momentum: } \rho(\mathbf{x})\mathbf{v}_t(\mathbf{x}, t) = \nabla \cdot \sigma(\mathbf{x}, t) + \mathbf{s}(\mathbf{x}, t) \\ \text{Constitutive Law: } \sigma_t(\mathbf{x}, t) = \mathbf{c}(\mathbf{x}) :: \varepsilon(\mathbf{x}, t) \end{aligned} \tag{34}$$

where the simulation variables are particle velocity $\mathbf{v} = (v_i)_{i=1,2,3}$ and stress tensor $\sigma = (\sigma_{ij})_{i,j=1,2,3}$. The material is characterized by density ρ and stiffness tensor

$\mathbf{c} = (c_{ijkl})_{i,j,k,l=1,2,3}$. $\varepsilon = (\varepsilon_{ij})_{i,j=1,2,3}$ is the strain rate tensor defined by $\frac{1}{2}[\nabla\mathbf{v}+(\nabla\mathbf{v})^T]$. $\mathbf{s}$ is the source function; however, as one might observe in 1D case that stability qualifications are obtained on homogeneous form, I will assume the source term is zero in the rest of the section.

Both stress and strain tensors are symmetric, and the stiffness tensor $\mathbf{c}$ satisfies $c_{ijkl} = c_{klij} = c_{jikl}$, so the second equation in Eq.(34) is often expressed in Voigt notation or Kelvin notation (Helbig, 1996). In this paper I adopt Kelvin notation as it will simplify the later derivation for the energy. Eq.(35) shows the Kelvin notation:

$$\frac{\partial}{\partial t}\begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sqrt{2}\sigma_{23} \\ \sqrt{2}\sigma_{13} \\ \sqrt{2}\sigma_{12} \end{pmatrix} = \begin{pmatrix} c_{1111} & c_{1122} & c_{1133} & \sqrt{2}c_{1123} & \sqrt{2}c_{1113} & \sqrt{2}c_{1112} \\ c_{2211} & c_{2222} & c_{2233} & \sqrt{2}c_{2223} & \sqrt{2}c_{2213} & \sqrt{2}c_{2212} \\ c_{3311} & c_{3322} & c_{3333} & \sqrt{2}c_{3323} & \sqrt{2}c_{3313} & \sqrt{2}c_{3312} \\ \sqrt{2}c_{2311} & \sqrt{2}c_{2322} & \sqrt{2}c_{2333} & 2c_{2323} & 2c_{2313} & 2c_{2312} \\ \sqrt{2}c_{1311} & \sqrt{2}c_{1322} & \sqrt{2}c_{1333} & 2c_{1323} & 2c_{1313} & 2c_{1312} \\ \sqrt{2}c_{1211} & \sqrt{2}c_{1222} & \sqrt{2}c_{1233} & 2c_{1223} & 2c_{1213} & 2c_{1212} \end{pmatrix}\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{22} \\ \varepsilon_{33} \\ \sqrt{2}\varepsilon_{23} \\ \sqrt{2}\varepsilon_{13} \\ \sqrt{2}\varepsilon_{12} \end{pmatrix}$$

(35)

Denote stress, strain rate and stiffness tensors in Kelvin notation by $\underline{\sigma}, \underline{\underline{c}}, \underline{\varepsilon}$ respectively, then Eq.(35) can be written as:

$$\underline{\sigma}_t(\mathbf{x}, t) = \underline{\underline{c}}(\mathbf{x})\underline{\varepsilon}(\mathbf{x}, t) \qquad (36)$$

To save subscripts for grid indices, let (u, v, w) denote three components of the velocity, $(txx, tyy, tzz, tyz, txz, txy)$ denote stress tensor components $(\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{23}, \sigma_{13}, \sigma_{12})$, and $(cij)_{i,j=1,...,6}$ denote $\underline{\underline{c}}_{ij}$.

In this work, I use the classical finite difference staggered grid scheme proposed by Virieux (1986) to simulate the wave propagation. Fig.2 shows the configuration of a grid cell of this staggered grid scheme. If the index tuple $[i, j, k, n + \frac{1}{2}]$ is mapped from $[ih, jh, kh, (n + \frac{1}{2})\Delta t]$, then the tuple $[i + \frac{1}{2}, j, k, n]$ is mapped from $[(i + \frac{1}{2})h, jh, kh, n\Delta t]$ and so forth. Notice that in each cell, normal stresses locate on the same grid position; velocity components are diagonal to each other, so are the stress components if viewing the normal stresses (txx, tyy, tzz) as a single component. This configuration can handle schemes with $(cij)_{6\times6}$ confined to the

following pattern:

$$
(cij)_{6\times6} = \begin{pmatrix}
c11 & c12 & c13 & 0 & 0 & 0 \\
c12 & c22 & c23 & 0 & 0 & 0 \\
c13 & c23 & c33 & 0 & 0 & 0 \\
0 & 0 & 0 & c44 & 0 & 0 \\
0 & 0 & 0 & 0 & c55 & 0 \\
0 & 0 & 0 & 0 & 0 & c66
\end{pmatrix}
\tag{37}
$$

which represents a material that is orthorhombic or more symmetric, e.g., transversely isotropic, and these two types of anisotropies are common for fractured shales (Tsvankin, 1997). Handling wave simulations in more general anisotropic material requires a fully staggered grid, or Lebedev grid scheme (Bernth and Chapman, 2011; Rubio et al., 2014), which stores far more variables per grid cell than the classical staggered grid scheme does; therefore if the material is not presenting strong anisotropy, it is better to use the classical scheme which computes faster and requires less memory.
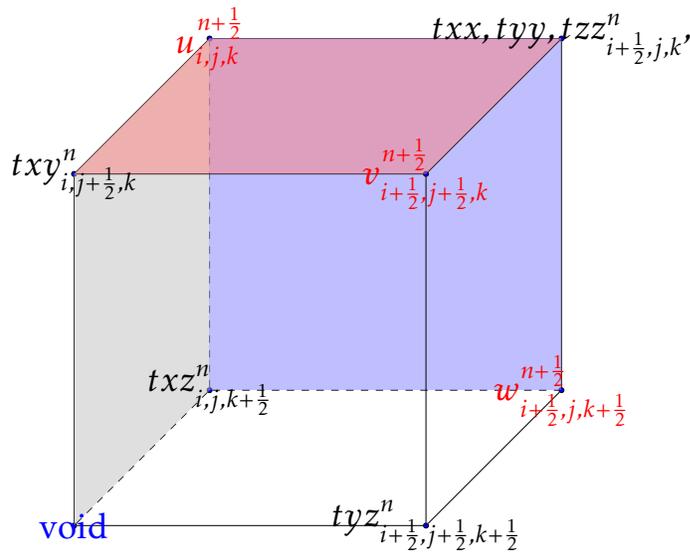
Figure 2: An unit grid cell of the classical staggered grid finite difference scheme.

With the above grid configuration, the classical finite difference scheme for Eq.(34)

is:

$$
\begin{cases}
\rho_{i,j,k} D_{-,t} u_{i,j,k}^{n+\frac{1}{2}} = D_{-,x} txx_{i+\frac{1}{2},j,k}^{n} + D_{-,y} txy_{i,j+\frac{1}{2},k}^{n} + D_{-,z} txz_{i,j,k+\frac{1}{2}}^{n} \\[2mm]
\rho_{i+\frac{1}{2},j+\frac{1}{2},k} D_{-,t} v_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}} = D_{+,x} txy_{i,j+\frac{1}{2},k}^{n} + D_{+,y} tyy_{i+\frac{1}{2},j,k}^{n} + D_{-,z} tyz_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n} \\[2mm]
\rho_{i+\frac{1}{2},j,k+\frac{1}{2}} D_{-,t} w_{i+\frac{1}{2},j,k+\frac{1}{2}}^{n+\frac{1}{2}} = D_{+,x} txz_{i,j,k+\frac{1}{2}}^{n} + D_{-,y} tyz_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n} + D_{+,z} tzz_{i+\frac{1}{2},j,k}^{n}
\end{cases}
$$

$$(\text{3D.FDM.a})$$

$$
\begin{cases}
\begin{pmatrix} D_{+,t} txx_{i+\frac{1}{2},j,k}^{n} \\[1mm] D_{+,t} tyy_{i+\frac{1}{2},j,k}^{n} \\[1mm] D_{+,t} tzz_{i+\frac{1}{2},j,k}^{n} \end{pmatrix} = \begin{pmatrix} c11_{i+\frac{1}{2},j,k} & c12_{i+\frac{1}{2},j,k} & c13_{i+\frac{1}{2},j,k} \\[1mm] c12_{i+\frac{1}{2},j,k} & c22_{i+\frac{1}{2},j,k} & c23_{i+\frac{1}{2},j,k} \\[1mm] c13_{i+\frac{1}{2},j,k} & c23_{i+\frac{1}{2},j,k} & c33_{i+\frac{1}{2},j,k} \end{pmatrix} \begin{pmatrix} D_{+,x} u_{i,j,k}^{n+\frac{1}{2}} \\[1mm] D_{-,y} v_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}} \\[1mm] D_{-,z} w_{i+\frac{1}{2},j,k+\frac{1}{2}}^{n+\frac{1}{2}} \end{pmatrix} \\[5mm]
D_{+,t} tyz_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n} = \frac{1}{2} c44_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}} (D_{+,z} v_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}} + D_{+,y} w_{i+\frac{1}{2},j,k+\frac{1}{2}}^{n+\frac{1}{2}}) \\[3mm]
D_{+,t} txz_{i,j,k+\frac{1}{2}}^{n} = \frac{1}{2} c55_{i,j,k+\frac{1}{2}} (D_{+,z} u_{i,j,k}^{n+\frac{1}{2}} + D_{-,x} w_{i+\frac{1}{2},j,k+\frac{1}{2}}^{n+\frac{1}{2}}) \\[3mm]
D_{+,t} txy_{i,j+\frac{1}{2},k}^{n} = \frac{1}{2} c66_{i,j+\frac{1}{2},k} (D_{+,y} u_{i,j,k}^{n+\frac{1}{2}} + D_{-,x} v_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}})
\end{cases}
$$

$$(\text{3D.FDM.b})$$

To model a material with a flat discontinuity along the *z* direction, I concatenate two uniform staggered grids having grid spacings of *h* and *H*(> *h*) respectively, and place their interface inside the fast material. I will employ same notation rules as those used in the 1D case: lowercase notations are for simulation variables on the fine grid, whereas uppercase notations denote coarse grid simulation variables; fine grid coefficients are appended with "-", and coarse grid coefficients with "+".

Fig.3 demonstrates my 3D composite grid scheme when *H* = 2*h*. The wave propagation speed is slower in the top material than in the bottom material, which is consistent with the fact that the wave propagation speed generally increases with depth in the subsurface of the earth. Hence a fine grid is used for the top area. Both grids are rectangular. Their boundary faces are aligned with colored faces shown in Fig.2, and the grid refinement interface is also aligned with the red face. Table.1 details the grid location and index range of each grid function. To allow

two staggered grids connect seamlessly, I assume

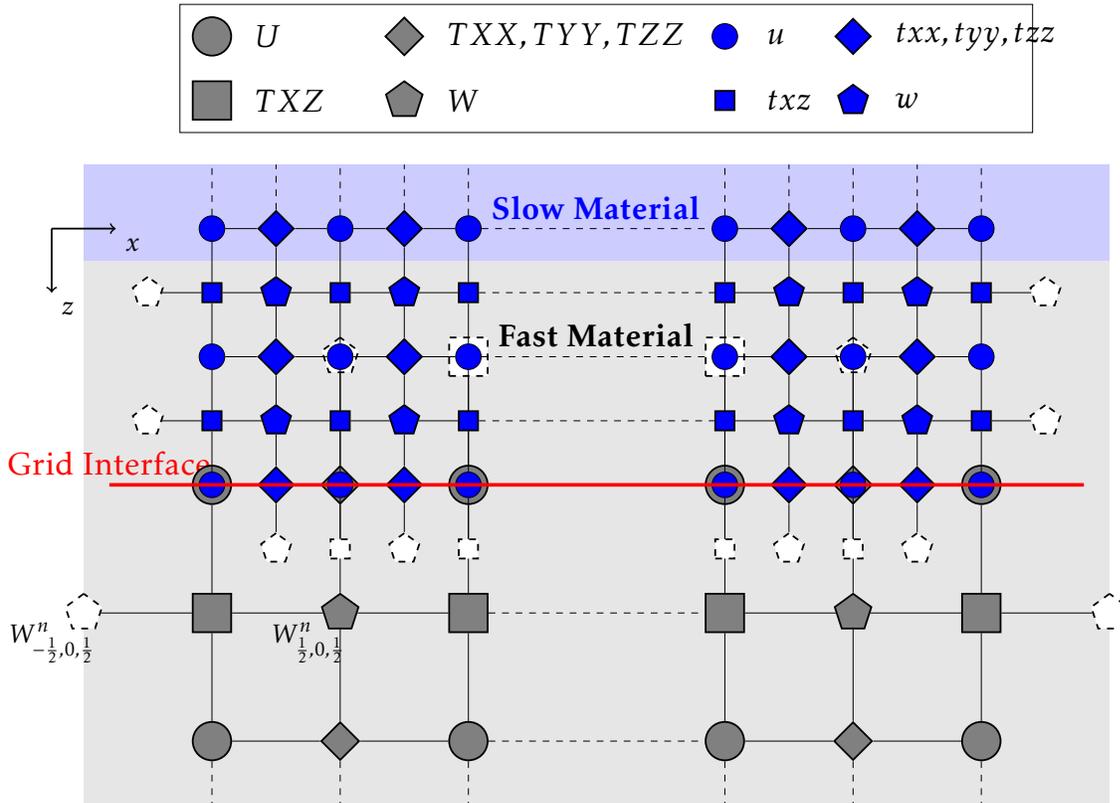$$\begin{cases} N_i h = N_I H \\ N_j h = N_J H. \end{cases} \tag{38}$$

Figure 3: A $x - z$ cross-section of the 3D composite staggered grid (H=2h) implemented with the homogeneous Dirichlet boundary condition. Smaller blue cells represent fine grid variables, bigger gray cells being variables on the coarse grid and white dashed cells being the ghost data. The thick red line denotes the grid refinement interface which is again placed inside the fast material.

To impose homogeneous Dirichlet boundary conditions on all boundary faces, if a velocity locates on some boundary face, e.g. $u^n_{0,0,N_k}$ is on $z = -N_k h$, then it is set to 0; otherwise construct a ghost data at the symmetric grid point of the velocity with respect to the boundary face, and select its value so that their average on the boundary equal to 0. For example, $W^n_{\frac{1}{2},0,\frac{1}{2}}$ (labelled in Fig.3) is only half grid size

Var	Fine Grid Function	Range of (i,j,k)
v_1	$u_{i,j,k}^{n+\frac{1}{2}}$	$[0,N_i]\times[0,N_j]\times[0,N_k]$
v_2	$v_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}}$	$[0,N_i-1]\times[0,N_j-1]\times[0,N_k]$
v_3	$w_{i+\frac{1}{2},j,k+\frac{1}{2}}^{n+\frac{1}{2}}$	$[0,N_i-1]\times[0,N_j]\times[0,N_k-1]$
σ_{11}	$txx_{i+\frac{1}{2},j,k}^{n}$	$[0,N_i-1]\times[0,N_j]\times[0,N_k]$
σ_{23}	$tyz_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}^{n}$	$[0,N_i-1]\times[0,N_j-1]\times[0,N_k-1]$
σ_{13}	$txz_{i,j,k+\frac{1}{2}}^{n}$	$[0,N_i]\times[0,N_j]\times[0,N_k-1]$
σ_{12}	$txy_{i,j+\frac{1}{2},k}^{n}$	$[0,N_i]\times[0,N_j-1]\times[0,N_k]$

Table 1: The table shows locations and ranges of fine grid functions. The fine grid domain is $[0,N_ih]\times[0,N_jh]\times[-N_kh,0]$, where $v_1(ih,jh,(k-N_k)h,(n+\frac{1}{2})\Delta t)\mapsto u_{i,j,k}^{n+\frac{1}{2}}$ and $v_2((i+\frac{1}{2})h,(j+\frac{1}{2})h,(k-N_k)h,(n+\frac{1}{2})\Delta t)\mapsto v_{i+\frac{1}{2},j+\frac{1}{2},k}^{n+\frac{1}{2}}$ and so forth. Notice that along any direction the number of points on the dual grid is always one less than it on the primary grid. The coarse grid is defined on $[0,N_Ih]\times[0,N_Jh]\times[0,N_Kh]$ (not shown) and $v_1(ih,jh,kh,(n+\frac{1}{2})\Delta t)\mapsto U_{i,j,k}^{n+\frac{1}{2}}$ and so forth.

close to the boundary face $x=0$, then $W_{-\frac{1}{2},0,\frac{1}{2}}^{n}$ is built at the symmetric location and select its value so that

$$\overline{W_{0,0,\frac{1}{2}}^{n}}=\frac{1}{2}(W_{\frac{1}{2},0,\frac{1}{2}}^{n}+W_{-\frac{1}{2},0,\frac{1}{2}}^{n})=0. \tag{39}$$

Similar to the 1D case, I built two layers of interface ghost data whose value are determined by the following stability analyses. Since the interface velocities are set to zero on the boundary due to the homogeneous Dirichlet boundary condition, the ghost data is only needed in the interior grid points.

Formulate Discrete Energy on the 3D Composite Grid

Before defining the discrete energy, I will formulate the energy in the continuous case, with variables vanishing in the infinity. I will show that this energy owns two

stability qualifications; therefore, the energy definition in the continuous case can serve as a reference when defining discrete energies on the composite staggered grid.

If $\underline{\underline{c}}$ is invertible, which is true in orthorhombic material, define the total energy as follows:

$$\mathcal{E}[\mathbf{v}, \underline{\sigma}](t) = \frac{1}{2} \iiint\limits_{\mathbb{R}^3} \rho(\mathbf{x}) \mathbf{v}^2(\mathbf{x}, t) \mathrm{d}\mathbf{x} + \frac{1}{2} \iiint\limits_{\mathbb{R}^3} (\underline{\underline{c}}^{-1}(\mathbf{x}) \underline{\sigma}(\mathbf{x}, t)) \cdot \underline{\sigma}(\mathbf{x}, t) \mathrm{d}\mathbf{x} \tag{40}$$

Since density and stiffness tensor are nonnegative, this energy qualifies for *boundedness*. On the other hand, take its derivative over time:

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{E}[\mathbf{v}, \underline{\sigma}](t)}{\mathrm{d}t} &= \iiint\limits_{\mathbb{R}^3} \rho(\mathbf{x}) \mathbf{v}_t(\mathbf{x}, t) \cdot \mathbf{v}(\mathbf{x}, t) \mathrm{d}\mathbf{x} + \iiint\limits_{\mathbb{R}^3} (\underline{\underline{c}}^{-1}(\mathbf{x}) \underline{\sigma}_t(\mathbf{x}, t)) \cdot \underline{\sigma}(\mathbf{x}, t) \mathrm{d}\mathbf{x} \\
&\overset{Eq.(34),(36)}{=} \iiint\limits_{\mathbb{R}^3} (\nabla \cdot \sigma(\mathbf{x}, t)) \cdot \mathbf{v}(\mathbf{x}, t) \mathrm{d}\mathbf{x} + \iiint\limits_{\mathbb{R}^3} \underline{\varepsilon}(\mathbf{x}, t) \cdot \underline{\sigma}(\mathbf{x}, t) \mathrm{d}\mathbf{x} \\
&= \iiint\limits_{\mathbb{R}^3} (\nabla \cdot \sigma(\mathbf{x}, t)) \cdot \mathbf{v}(\mathbf{x}, t) \mathrm{d}\mathbf{x} + \iiint\limits_{\mathbb{R}^3} \varepsilon(\mathbf{x}, t) \circ \sigma(\mathbf{x}, t) \mathrm{d}\mathbf{x} \\
&= \iiint\limits_{\mathbb{R}^3} \sigma_{ij,j}(\mathbf{x}, t)) v_j(\mathbf{x}, t) \mathrm{d}\mathbf{x} + \iiint\limits_{\mathbb{R}^3} \frac{1}{2} (v_{j,i}(\mathbf{x}, t) + v_{i,j}(\mathbf{x}, t)) \sigma_{ij}(\mathbf{x}, t) \mathrm{d}\mathbf{x} \overset{IBP}{=} 0
\end{aligned} \tag{41}$$

where the third equality reveals the advantage of using Kelvin notation that the inner product of $\underline{\epsilon}$ and $\underline{\sigma}$ equals the Hadamard product of ϵ and σ. The last line is written in Einstein summation notation. This equation shows that the energy qualifies for *conservation*.

Towards the discrete case, I will first define some inner products, which can simplify energy definitions and stability analyses afterwards. Table 2 shows a series of inner products defined on the fine grid for different grid functions. Each inner product approximates the product of two grid functions integrated on the fine grid domain $[0, N_i h] \times [0, N_j h] \times [-N_k h, 0]$, and it also naturally induces a norm by the same name. The inner products on the coarse grid can be defined similarly.

With these inner products and their induced norms, define the kinetic energy on

Fine Grid Function	Inner Products
$u_{i,j,k}$	$\langle u,u\rangle_u = \dfrac{1}{2}\displaystyle\sum_{i=0}^{N_i}\sum_{j=0}^{N_j}\sum_{k=0}^{N_k}\alpha_i\alpha_j\alpha_k u_{i,j,k}u_{i,j,k}$
$v_{i+\frac{1}{2},j+\frac{1}{2},k}$	$\langle v,v\rangle_v = \dfrac{1}{2}\displaystyle\sum_{i=0}^{N_i-1}\sum_{j=0}^{N_j-1}\sum_{k=0}^{N_k}\alpha_k v_{i+\frac{1}{2},j+\frac{1}{2},k}v_{i+\frac{1}{2},j+\frac{1}{2},k}$
$w_{i+\frac{1}{2},j,k+\frac{1}{2}}$	$\langle w,w\rangle_w = \dfrac{1}{2}\displaystyle\sum_{i=0}^{N_i-1}\sum_{j=0}^{N_j}\sum_{k=0}^{N_k-1}\alpha_j w_{i+\frac{1}{2},j,k+\frac{1}{2}}w_{i+\frac{1}{2},j,k+\frac{1}{2}}$
$txx_{i+\frac{1}{2},j,k}$	$\langle txx,txx\rangle_{txx} = \dfrac{1}{2}\displaystyle\sum_{i=0}^{N_i-1}\sum_{j=0}^{N_j}\sum_{k=0}^{N_k}\alpha_j\alpha_k (txx_{i+\frac{1}{2},j,k})^2$
$tyz_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}}$	$\langle tyz,tyz\rangle_{tyz} = \dfrac{1}{2}\displaystyle\sum_{i=0}^{N_i-1}\sum_{j=0}^{N_j-1}\sum_{k=0}^{N_k-1}(tyz_{i+\frac{1}{2},j+\frac{1}{2},k+\frac{1}{2}})^2$
$txz_{i,j,k+\frac{1}{2}}$	$\langle txz,txz\rangle_{txz} = \dfrac{1}{2}\displaystyle\sum_{i=0}^{N_i}\sum_{j=0}^{N_j}\sum_{k=0}^{N_k-1}\alpha_i\alpha_j (txz_{i,j,k+\frac{1}{2}})^2$
$txy_{i,j+\frac{1}{2},k}$	$\langle txy,txy\rangle_{txy} = \dfrac{1}{2}\displaystyle\sum_{i=0}^{N_i}\sum_{j=0}^{N_j-1}\sum_{k=0}^{N_k}\alpha_i\alpha_k (txz_{i,j+\frac{1}{2},k})^2$

Table 2: Inner products defined for different grid functions on the fine grid, where α_i is $\dfrac{1}{2}$ at 0 or N_i and is 1 elsewhere; α_j is $\dfrac{1}{2}$ at 0 or N_j and is 1 elsewhere; α_k is $\dfrac{1}{2}$ at 0 or N_k and is 1 elsewhere.

the fine grid as:

$$\begin{aligned}
\mathcal{K}_f^n &= \langle \rho^- u^{n+\frac{1}{2}}, u^{n-\frac{1}{2}}\rangle_u + \langle \rho^- v^{n+\frac{1}{2}}, v^{n-\frac{1}{2}}\rangle_v + \langle \rho^- w^{n+\frac{1}{2}}, w^{n-\frac{1}{2}}\rangle_w \\
&\approx \frac{1}{2}\int_0^{N_i h}\int_0^{N_j h}\int_{-N_k h}^0 \rho(\mathbf{x})\mathbf{v}(\mathbf{x},(n+\frac{1}{2})\Delta t)\mathbf{v}(\mathbf{x},(n-\frac{1}{2})\Delta t)\mathrm{d}\mathbf{x}
\end{aligned} \tag{42}$$

The strain energy on the fine grid is defined as:

$$
\begin{aligned}
\mathcal{S}_f^n =\ & \frac{1}{2} \sum_{i=0}^{N_i-1} \sum_{j=0}^{N_j} \sum_{k=0}^{N_k} \alpha_j \alpha_k h^3 \begin{pmatrix} txx^n \\ tyy^n \\ tzz^n \end{pmatrix}^T \begin{pmatrix} c11^- & c12^- & c13^- \\ c12^- & c22^- & c23^- \\ c13^- & c23^- & c33^- \end{pmatrix}^{-1} \begin{pmatrix} txx^n \\ tyy^n \\ tzz^n \end{pmatrix}\Big|_{i+\frac{1}{2},j,k} \\
& + 2\langle \frac{1}{c44^-} tyz^n, tyz^n \rangle_{tyz} + 2\langle \frac{1}{c55^-} txz^n, txz^n \rangle_{txz} + 2\langle \frac{1}{c66^-} txy^n, txy^n \rangle_{txy} \\
\approx\ & \frac{1}{2} \int_0^{N_i h} \int_0^{N_j h} \int_{-N_k h}^0 (\underline{\underline{\mathbf{c}}}^{-1}(\mathbf{x}, n\Delta t)\underline{\sigma}(\mathbf{x}, n\Delta t)) \cdot \underline{\sigma}(\mathbf{x}, n\Delta t) d\mathbf{x}
\end{aligned}
\tag{43}
$$

where the first line means all grid functions in that term are evaluated at $(i + \frac{1}{2}, j, k)$, and certainly it can be written in $\langle \cdot, \cdot \rangle_{txx}$ notation, but keeping the coefficient matrix eases derivations afterwards.

Then the sum of $\mathcal{S}_f^n$ and $\mathcal{K}_f^n$ is defined as the total energy on the fine grid $\mathcal{E}_f^n$. The energies on the coarse grid can be defined in a similar way, and the total energy on the composite grid $\mathcal{E}_c^n$ is then the sum of $\mathcal{E}_f^n$ and $\mathcal{E}_r^n$.

The rest of the section will present how to update ghost data and restrict time step size in order to qualify $\mathcal{E}_c^n$ for stability.

Qualify for *Conservation*

In 1D stability analyses, (1D.SBP) property is used extensively. This property leads to a list of 3D SBP properties (Eq.(3D.SBP)) and they will be used many

times in the 3D numerical stability analyses.

$$\langle u, D_{-,x}txx \rangle_u + \langle D_{+,x}u, txx \rangle_{txx} = \sum_{j=0}^{N_j} \sum_{k=0}^{N_k} \alpha_j \alpha_k h^2 (u_{N_i,j,k}\overline{txx_{N_i,j,k}} - u_{0,j,k}\overline{txx_{0,j,k}})$$

$$\langle u, D_{-,y}txy \rangle_u + \langle D_{+,y}u, txy \rangle_{txy} = \sum_{i=0}^{N_i} \sum_{k=0}^{N_k} \alpha_i \alpha_k h^2 (u_{i,N_j,k}\overline{txy_{i,N_j,k}} - u_{i,0,k}\overline{txy_{i,0,k}})$$

$$\vdots$$

$$\langle w, D_{+,z}tzz \rangle_w + \langle D_{-,z}w, tzz \rangle_{tzz} = \sum_{i=0}^{N_i-1} \sum_{j=0}^{N_j} \alpha_j h^2 (\overline{w_{i+\frac{1}{2},j,N_k}}tzz_{i+\frac{1}{2},j,N_k} - \overline{w_{i+\frac{1}{2},j,0}}tzz_{i+\frac{1}{2},j,0})$$

$$\text{(3D.SBP)}$$

Proposition 3. *With the Dirichlet homogeneous boundary condition and Eq.(3D.SBP) , the total energy difference over time on the fine grid comes down to:*

$$\mathcal{E}_f^n - \mathcal{E}_f^{n-1} = h^2 \Delta t \Big[\sum_{i=1}^{N_i-1} \sum_{j=1}^{N_j-1} u_{i,j,N_k}^{n-\frac{1}{2}}(\overline{txz_{i,j,N_k}^n} + \overline{txz_{i,j,N_k}^{n-1}})$$

$$+ \sum_{i=0}^{N_i-1} \sum_{j=0}^{N_j-1} v_{i+\frac{1}{2},j+\frac{1}{2},N_k}^{n-\frac{1}{2}}(\overline{tyz_{i+\frac{1}{2},j+\frac{1}{2},N_k}^n} + \overline{tyz_{i+\frac{1}{2},j+\frac{1}{2},N_k}^{n-1}}) \qquad (44)$$

$$+ \sum_{i=0}^{N_i-1} \sum_{j=1}^{N_j-1} \overline{w_{i+\frac{1}{2},j,N_k}^{n-\frac{1}{2}}}(tzz_{i+\frac{1}{2},j,N_k}^n + tzz_{i+\frac{1}{2},j,N_k}^{n-1}) \Big]$$

Proof. The total energy difference over time is the sum of strain energy difference and kinetic energy difference.

Apply the trick: $X^T A X - Y^T A Y = (X+Y)^T A(X-Y)$, and the strain energy difference

over time on the fine grid is:

$$\mathcal{S}_f^n - \mathcal{S}_f^{n-1}$$

$$\overset{(3D.FDM.b)}{=} \frac{1}{2}\Delta t \sum_{i=0}^{N_i-1} \sum_{j=0}^{N_j} \sum_{k=0}^{N_k} \alpha_j \alpha_k h^3 \begin{pmatrix} txx^n_{i+\frac{1}{2},j,k} + txx^{n-1}_{i+\frac{1}{2},j,k} \\ tyy^n_{i+\frac{1}{2},j,k} + tyy^{n-1}_{i+\frac{1}{2},j,k} \\ tzz^n_{i+\frac{1}{2},j,k} + tzz^{n-1}_{i+\frac{1}{2},j,k} \end{pmatrix}^T \begin{pmatrix} D_{+,x}u^{n-\frac{1}{2}}_{i,j,k} \\ D_{-,y}v^{n-\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2},k} \\ D_{-,z}w^{n-\frac{1}{2}}_{i+\frac{1}{2},j,k+\frac{1}{2}} \end{pmatrix} \quad (45)$$

$$+ \Delta t \langle tyz^n + tyz^{n-1}, D_{+,z}v^{n+\frac{1}{2}} + D_{+,y}w^{n+\frac{1}{2}} \rangle_{tyz}$$

$$+ \Delta t \langle txz^n + txz^{n-1}, D_{+,z}u^{n+\frac{1}{2}} + D_{-,x}w^{n+\frac{1}{2}} \rangle_{txz}$$

$$+ \Delta t \langle txy^n + txy^{n-1}, D_{+,y}u^{n+\frac{1}{2}} + D_{-,x}v^{n+\frac{1}{2}} \rangle_{txy}$$

On the other hand, the kinetic energy difference over time on the fine grid is:

$$\mathcal{K}_f^n - \mathcal{K}_f^{n-1} = \langle \rho^- u^{n-\frac{1}{2}}, u^{n+\frac{1}{2}} - u^{n-\frac{3}{2}} \rangle_u + \langle \rho^- v^{n-\frac{1}{2}}, v^{n+\frac{1}{2}} - v^{n-\frac{3}{2}} \rangle_v + \langle \rho^- w^{n-\frac{1}{2}}, w^{n+\frac{1}{2}} - w^{n-\frac{3}{2}} \rangle_w$$

$$(46)$$

Note that

$$\langle \rho^- u^{n-\frac{1}{2}}, u^{n+\frac{1}{2}} - u^{n-\frac{3}{2}} \rangle_u$$

$$\overset{(3D.FDM.a)}{=} \Delta t \langle u^{n-\frac{1}{2}}, D_{-,x}(txx^n + txx^{n-1}) + D_{-,y}(txy^n + txy^{n-1}) + D_{-,z}(txz^n + txz^{n-1}) \rangle_u$$

$$\overset{(3D.SBP)}{=} \Delta t \langle D_{+,x}u^{n-\frac{1}{2}}, txx^n + txx^{n-1} \rangle_{txx} + \Delta t \langle D_{+,y}u^{n-\frac{1}{2}}, txy^n + txy^{n-1} \rangle_{txy}$$

$$+ \Delta t \langle D_{+,z}u^{n-\frac{1}{2}}, txz^n + txz^{n-1} \rangle_{txz} + h^2 \Delta t \sum_{i=1}^{N_i-1} \sum_{j=1}^{N_j-1} u^{n-\frac{1}{2}}_{i,j,N_k} (\overline{txz^n_{i,j,N_k}} + \overline{txz^{n-1}_{i,j,N_k}})$$

$$(47)$$

where the second equation also uses the homogeneous Dirichlet boundary condition. Notice that the result of this equation already cancels some terms in Eq.(45). Apply similar procedures to other terms in the kinetic energy difference, then it yields that the remaining part would be equal to Eq.(44). $\square$

On the coarse grid I obtain a similar result:

$$
\begin{aligned}
\mathcal{E}_r^n - \mathcal{E}_r^{n-1} = &-H^2\Delta t[\sum_{i=1}^{N_I-1}\sum_{j=1}^{N_J-1} U_{i,j,0}^{n-\frac{1}{2}}(\overline{TXZ_{i,j,0}^n} + \overline{TXZ_{i,j,0}^{n-1}}) \\
&+ \sum_{i=0}^{N_I-1}\sum_{j=0}^{N_J-1} V_{i+\frac{1}{2},j+\frac{1}{2},0}^{n-\frac{1}{2}}(\overline{TYZ_{i+\frac{1}{2},j+\frac{1}{2},0}^n} + \overline{TYZ_{i+\frac{1}{2},j+\frac{1}{2},0}^{n-1}}) \\
&+ \sum_{i=0}^{N_I-1}\sum_{j=1}^{N_J-1} \overline{W_{i+\frac{1}{2},j,0}^{n-\frac{1}{2}}}(TZZ_{i+\frac{1}{2},j,0}^n + TZZ_{i+\frac{1}{2},j,0}^{n-1})],
\end{aligned}
\tag{48}
$$

Comparing Eq.(44) with (48) yields that the energy is conserved if velocities and stresses satisfy following equations for any nonnegative n and m.

$$
h^2\sum_{i=1}^{N_i-1}\sum_{j=1}^{N_j-1} u_{i,j,N_k}^{n-\frac{1}{2}}\,\overline{txz_{i,j,N_k}^m} = H^2\sum_{i=1}^{N_I-1}\sum_{j=1}^{N_J-1} u_{i,j,0}^{n-\frac{1}{2}}\overline{TXZ_{i,j,0}^m}
$$

$$
h^2\sum_{i=0}^{N_i-1}\sum_{j=0}^{N_j-1} v_{i+\frac{1}{2},j+\frac{1}{2},N_k}^{n-\frac{1}{2}}\,\overline{tyz_{i+\frac{1}{2},j+\frac{1}{2},N_k}^m} = H^2\sum_{i=0}^{N_I-1}\sum_{j=0}^{N_J-1} v_{i+\frac{1}{2},j+\frac{1}{2},N_K}^{n-\frac{1}{2}}\,\overline{TYZ_{i+\frac{1}{2},j+\frac{1}{2},N_K}^m}
$$

$$
h^2\sum_{i=0}^{N_i-1}\sum_{j=1}^{N_j-1} \overline{w_{i+\frac{1}{2},j,N_k}^{n-\frac{1}{2}}}\,tzz_{i+\frac{1}{2},j,N_k}^m = H^2\sum_{i=0}^{N_I-1}\sum_{j=1}^{N_J-1} \overline{W_{i+\frac{1}{2},j,N_K}^{n-\frac{1}{2}}}\,TZZ_{i+\frac{1}{2},j,N_K}^m
$$

$$
\text{(3D.ECI)}
$$

These equations are analogous to Eq.(1D.ECI) in 1D case. If the ghost data are selected to validate these equations, then the energy is conserved.

However, solving for ghost data certainly requires more equations than the ones provided in Eq.(3D.ECI). Like in the 1D case, I use the transmission condition to create more equations out of Eq.(3D.ECI).

For example, suppose $H = 2h$, then $N_i = 2N_I$, and solving for $TXZ_{i,j,-\frac{1}{2}}^n$ on $(i,j) \in [1, N_I-1]\times[1,N_J-1]$ and $txz_{i,j,N_k+\frac{1}{2}}^n$ on $(i,j) \in [1,N_i-1]\times[1,N_j-1]$ needs $5N_iN_j - 3N_i - 3N_j + 2$ independent equations. Note that the coarse grid point $[i,j,0]$ locates on $(iH, jH, 0)$ on the spatial domain, so does fine grid point $[2i, 2j, N_k]$; fine grid point $[2i+1, 2j, N_k]$ locates in the middle of $[i,j,0]$ and $[i+1,j,0]$ on the coarse grid

and so on. The transmission condition asks for stress continuity which implies the following conditions:

$$
\begin{cases}
\overline{txz^m_{2i,2j,N_k}} = \overline{TXZ^m_{i,j,0}} \text{ on } [1, N_I - 1] \times [1, N_J - 1] \\
\overline{txz^m_{2i+1,2j,N_k}} = \dfrac{1}{2}(\overline{TXZ^m_{i,j,0}} + \overline{TXZ^m_{i+1,j,0}}) \text{ on } [0, N_I - 1] \times [1, N_J - 1] \\
\overline{txz^m_{2i,2j+1,N_k}} = \dfrac{1}{2}(\overline{TXZ^m_{i,j,0}} + \overline{TXZ^m_{i,j+1,0}}) \text{ on } [1, N_I - 1] \times [0, N_J - 1] \\
\overline{txz^m_{2i+1,2j+1,N_k}} = \dfrac{1}{4}(\overline{TXZ^m_{i,j,0}} + \overline{TXZ^m_{i+1,j,0}} + \overline{TXZ^m_{i,j+1,0}} + \overline{TXZ^m_{i+1,j+1,0}}) \text{ on } [0, N_I - 1] \times [0, N_J - 1]
\end{cases}
$$
$$\text{(3D.ECI.txz.a)}$$

Substituting these equations back into the first equation of Eq.(3D.ECI) yields the following condition on $(i, j) \in [1, N_I - 1] \times [1, N_J - 1]$:

$$
\begin{aligned}
16 U^{n+\frac{1}{2}}_{i,j,0} =& 4u^{n+\frac{1}{2}}_{2i,2j,N_k} + 2(u^{n+\frac{1}{2}}_{2i+1,2j,N_k} + u^{n+\frac{1}{2}}_{2i-1,2j,N_k} + u^{n+\frac{1}{2}}_{2i,2j+1,N_k} + u^{n+\frac{1}{2}}_{2i,2j-1,N_k}) \\
& + (u^{n+\frac{1}{2}}_{2i+1,2i+1,N_k} + u^{n+\frac{1}{2}}_{2i-1,2j+1,N_k} + u^{n+\frac{1}{2}}_{2i+1,2j-1,N_k} + u^{n+\frac{1}{2}}_{2i-1,2j-1,N_k})
\end{aligned}
$$
$$\text{(3D.ECI.txz.b)}$$

Note that this condition happens to be consistent with the velocity continuity required in the transmission condition. Let $m = n$, then (3D.ECI.txz.a) and (3D.ECI.txz.b) together provide $5N_i N_j - 3N_i - 3N_j + 2$ equations for $TXZ^n_{i,j,-\frac{1}{2}}$ and $txz^n_{i,j,N_k+\frac{1}{2}}$. The following proposition proves that this equation system is solvable under some mild conditions, and then the problem becomes how to solve it efficiently, which will be part of the future work.

Proposition 4. *If interface densities satisfy Eq.(52) on* $(i, j) \in [1, N_I - 1] \times [1, N_J - 1]$, *then the solution of the equation system provided by* (3D.ECI.txz.a)*(when m = n) and* (3D.ECI.txz.b) *exists and is unique.*

Proof. By (3D.FDM.a), (3D.ECI.txz.b) can be transformed into an equation of $TXZ^n_{i,j,-\frac{1}{2}}$ and $txz^n_{i,j,N_k+\frac{1}{2}}$. Now replace $txz^n_{i,j,N_k+\frac{1}{2}}$ with $TXZ^n_{i,j,-\frac{1}{2}}$ using (3D.ECI.txz.a), then the previous equation contains only $TXZ^n_{i,j,-\frac{1}{2}}$, and is in the following form:

$$
\begin{aligned}
& A_{i-1,j-1} TXZ^n_{i-1,j-1,-\frac{1}{2}} + A_{i-1,j} TXZ^n_{i-1,j,-\frac{1}{2}} + A_{i-1,j+1} TXZ^n_{i-1,j+1,-\frac{1}{2}} \\
& + A_{i,j-1} TXZ^n_{i,j-1,-\frac{1}{2}} + A_{i,j} TXZ^n_{i,j,-\frac{1}{2}} + A_{i,j+1} TXZ^n_{i,j+1,-\frac{1}{2}} \\
& + A_{i+1,j-1} TXZ^n_{i+1,j-1,-\frac{1}{2}} + A_{i+1,j} TXZ^n_{i+1,j,-\frac{1}{2}} + A_{i+1,j+1} TXZ^n_{i+1,j+1,-\frac{1}{2}} = B_{i,j}
\end{aligned}
\tag{49}
$$

where

$$A_{i-1,j-1} = \frac{1}{4\rho^-_{2i-1,2j-1,N_k}} > 0, \quad A_{i-1,j+1} = \frac{1}{4\rho^-_{2i-1,2j+1,N_k}} > 0$$

$$A_{i+1,j-1} = \frac{1}{4\rho^-_{2i+1,2j-1,N_k}} > 0, \quad A_{i+1,j+1} = \frac{1}{4\rho^-_{2i+1,2j+1,N_k}} > 0$$

$$A_{i-1,j} = \frac{1}{\rho^-_{2i-1,2j,N_k}} + \frac{1}{4\rho^-_{2i-1,2j+1,N_k}} + \frac{1}{4\rho^-_{2i-1,2j-1,N_k}} > 0$$

$$A_{i,j-1} = \frac{1}{\rho^-_{2i,2j-1,N_k}} + \frac{1}{4\rho^-_{2i+1,2j-1,N_k}} + \frac{1}{4\rho^-_{2i-1,2j-1,N_k}} > 0 \tag{50}$$

$$A_{i+1,j} = \frac{1}{\rho^-_{2i+1,2j,N_k}} + \frac{1}{4\rho^-_{2i+1,2j+1,N_k}} + \frac{1}{4\rho^-_{2i+1,2j-1,N_k}} > 0$$

$$A_{i,j+1} = \frac{1}{\rho^-_{2i,2j+1,N_k}} + \frac{1}{4\rho^-_{2i+1,2j+1,N_k}} + \frac{1}{4\rho^-_{2i-1,2j+1,N_k}} > 0$$

$$A_{i,j} = A_{i-1,j-1} + A_{i-1,j} + A_{i-1,j+1} + A_{i,j-1} + A_{i,j+1} + A_{i+1,j-1} + A_{i+1,j} + A_{i+1,j+1}$$
$$+ \frac{8}{\rho^+_{i,j,0}} + \frac{4}{\rho^-_{2i,2j,N_k}} - \frac{1}{2\rho^-_{2i+1,2j+1,N_k}} - \frac{1}{2\rho^-_{2i-1,2j+1,N_k}} - \frac{1}{2\rho^-_{2i+1,2j-1,N_k}} - \frac{1}{2\rho^-_{2I-1,2J-1,N_k}} \tag{51}$$

and $B_{i,j}$ is the remaining part that does not involve any ghost data but only known data.

View Eq.(49) as a linear equation system $\hat{A}X = \hat{B}$ solving for the unknown vector $X = (TXZ^n_{1,1,0}, ..., TXZ^n_{1,N_I-1,0}, TXZ^n_{2,1,0}, ..., TXZ^n_{2,N_I-1,0}, ..., TXZ^n_{N_I-1,1,0}, ..., TXZ^n_{N_I-1,N_J-1,0})^T$, and A, B are mapped to $\hat{A}, \hat{B}$ in a similar way.

If density functions satisfy the following inequality on $(i,j) \in [1, N_I - 1] \times [1, N_J - 1]$

$$\frac{8}{\rho^+_{i,j,0}} + \frac{4}{\rho^-_{2i,2j,N_k}} - \frac{1}{2\rho^-_{2i+1,2j+1,N_k}} - \frac{1}{2\rho^-_{2i-1,2j+1,N_k}} - \frac{1}{2\rho^-_{2i+1,2j-1,N_k}} - \frac{1}{2\rho^-_{2I-1,2J-1,N_k}} > 0. \tag{52}$$

then $\hat{A}$ is strictly diagonally dominant, and hence it is non-singular. As a result, X is uniquely determined, and substituting X back to (3D.ECI.txz.a) uniquely determines $txz^n_{i,j,N_k+\frac{1}{2}}$.

The inequality Eq.(52) imposed on density is considered mild because when the

density is locally smooth enough, i.e., densities involved in Eq.(52) are nearly identical, then the inequality is satisfied. □

Here I list two groups of equations to compute the remaining ghost data but omit the derivation process for brevity. Similar to Proposition 4, it is easy to prove that they are solvable because of having strictly diagonally dominant coefficient matrices; but unlike Proposition 4, when solving for these ghost data, achieving strictly diagonal dominance poses no restriction on density or whatever.

Equations below are solved for $TYZ^n_{i+\frac{1}{2},j+\frac{1}{2},0}$ on $[0,N_I-1]\times[0,N_J-1]$, and $tyz^n_{i+\frac{1}{2},j+\frac{1}{2},N_k}$ on $[0,N_i-1]\times[0,N_j-1]$:

$$
\begin{cases}
4V^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2},0} = v^{n+\frac{1}{2}}_{2i+\frac{1}{2},2j+\frac{1}{2},N_k} + v^{n+\frac{1}{2}}_{2I+\frac{3}{2},2j+\frac{1}{2},N_k} + v^{n+\frac{1}{2}}_{2i+\frac{1}{2},2j+\frac{3}{2},N_k} + v^{n+\frac{1}{2}}_{2i+\frac{3}{2},2j+\frac{3}{2},N_k} \\[2mm]
\overline{TYZ^n_{i+\frac{1}{2},j+\frac{1}{2},0}} = \overline{tyz^n_{2i+\frac{1}{2},2j+\frac{1}{2},N_k}} = \overline{tyz^n_{2i+\frac{3}{2},2j+\frac{1}{2},N_k}} = \overline{tyz^n_{2i+\frac{1}{2},2j+\frac{3}{2},N_k}} = \overline{tyz^n_{2i+\frac{3}{2},2j+\frac{3}{2},N_k}}
\end{cases}
$$
$$\text{(3D.ECI.tyz)}$$

The following group of equations are solved for $w^n_{i+\frac{1}{2},j,N_k}$ on $[0,N_I-1]\times[1,N_J-1]$, $W^n_{i+\frac{1}{2},j,0}$ on $[0,N_i-1]\times[1,N_j-1]$:

$$
\begin{cases}
8TZZ^{n+1}_{i+\frac{1}{2},j,0} = 2tzz^{n+1}_{2i+\frac{1}{2},2j,N_k} + 2tzz^{n+1}_{2i+\frac{3}{2},2j,N_k} + tzz^{n+1}_{2i+\frac{1}{2},2j\pm1,N_k} + tzz^{n+1}_{2i+\frac{3}{2},2j\pm1,N_k} \\[2mm]
\overline{W^{n+\frac{1}{2}}_{i+\frac{1}{2},j,0}} = \overline{w^{n+\frac{1}{2}}_{2i+\frac{1}{2},2j,N_k}} = \overline{w^{n+\frac{1}{2}}_{2i+\frac{3}{2},2j,N_k}} \\[2mm]
\overline{w^{n+\frac{1}{2}}_{2i+\frac{1}{2},2j+1,N_k}} = \overline{w^{n+\frac{1}{2}}_{2i+\frac{3}{2},2j+1,N_k}} = \frac{1}{2}\left(\overline{W^{n+\frac{1}{2}}_{i+\frac{1}{2},j,0}} + \overline{W^{n+\frac{1}{2}}_{i+\frac{1}{2},j+1,0}}\right).
\end{cases}
$$
$$\text{(3D.ECI.w)}$$

Qualify for *Boundedness*

Denote the maximum and minimum eigenvalue of the inverse matrix of

$$
\begin{pmatrix}
c11^- & c12^- & c13^- \\
c12^- & c22^- & c23^- \\
c13^- & c23^- & c33^-
\end{pmatrix}\Big|_{i+\frac{1}{2},j,k}
$$

by $e^-_{i+\frac{1}{2},j,k}$ and $E^-_{i+\frac{1}{2},j,k}$. Since all the elements in this matrix are positive, its eigenvalues must be positive, and then $emax_{i+\frac{1}{2},j,k}, emin_{i+\frac{1}{2},j,k} > 0$. Consequently, the

strain energy on the fine grid satisfies:

$$
\mathcal{S}_f^n \begin{cases}
\leq \langle E^- txx^n, txx^n \rangle + \langle E^- tyy^n, tyy^n \rangle + \langle E^- tzz^n, tzz^n \rangle \\
+2\langle \dfrac{1}{c44^-}tyz^n, tyz^n \rangle_{tyz} + 2\langle \dfrac{1}{c55^-}txz^n, txz^n \rangle_{txz} + 2\langle \dfrac{1}{c66^-}txy^n, txy^n \rangle_{txy} \\
\geq \langle e^- txx^n, txx^n \rangle + \langle e^- tyy^n, tyy^n \rangle + \langle e^- tzz^n, tzz^n \rangle \\
+2\langle \dfrac{1}{c44^-}tyz^n, tyz^n \rangle_{tyz} + 2\langle \dfrac{1}{c55^-}txz^n, txz^n \rangle_{txz} + 2\langle \dfrac{1}{c66^-}txy^n, txy^n \rangle_{txy}
\end{cases}
\tag{53}
$$

On the coarse grid a similar result can be obtained.

On the other hand, for the fine grid kinetic energy

$$
\langle \rho^- u^{n+\frac{1}{2}}, u^{n-\frac{1}{2}} \rangle_u + \langle \rho^+ U^{n+\frac{1}{2}}, U^{n-\frac{1}{2}} \rangle_U
$$

$$
\overset{(3D.FDM.a)}{=} \Delta t \langle u^{n-\frac{1}{2}}, D_{-,x}txx^n + D_{-,y}txy^n + D_{-,z}txz^n \rangle_u + \langle \rho^- u^{n-\frac{1}{2}}, u^{n-\frac{1}{2}} \rangle_u
$$

$$
+ \Delta t \langle U^{n-\frac{1}{2}}, D_{-,x}TXX^n + D_{-,y}TXY^n + D_{-,z}TXZ^n \rangle_U + \langle \rho^+ U^{n-\frac{1}{2}}, U^{n-\frac{1}{2}} \rangle_U
$$

$$
\overset{(3D.ECI)}{=} -\Delta t \Big(\langle D_{+,x}u^{n-\frac{1}{2}}, txx^n \rangle_{txx} + \langle D_{+,y}u^{n-\frac{1}{2}}, txy^n \rangle_{txy} + \langle D_{+,z}u^{n-\frac{1}{2}}, txz^n \rangle_u
$$

$$
+ \langle D_{+,x}U^{n-\frac{1}{2}}, TXX^n \rangle_{TXX} + \langle D_{+,y}U^{n-\frac{1}{2}}, TXY^n \rangle_{TXY} + \langle D_{+,z}U^{n-\frac{1}{2}}, TXZ^n \rangle_{TXZ} \Big)
$$

$$
+ \langle \rho^- u^{n-\frac{1}{2}}, u^{n-\frac{1}{2}} \rangle_u + \langle \rho^+ U^{n-\frac{1}{2}}, U^{n-\frac{1}{2}} \rangle_U
\tag{54}
$$

where every inner product is bounded by the norm of vectors, for example,

$$
|\langle D_{+,x}u^{n-\frac{1}{2}}, txx^n \rangle_{txx}| \leq \frac{1}{h}\left(\gamma_{ux}\|u^{n-\frac{1}{2}}\|_u^2 + \frac{1}{\gamma_{ux}}\|txx^n\|_{txx}^2 \right)
\tag{55}
$$

Therefore, let

$$
\begin{aligned}
\beta_1 =& \langle \rho^- u^{n-\frac{1}{2}}, u^{n-\frac{1}{2}} \rangle_u + \langle \rho^+ U^{n-\frac{1}{2}}, U^{n-\frac{1}{2}} \rangle_U + \langle \rho^- v^{n-\frac{1}{2}}, v^{n-\frac{1}{2}} \rangle_v + \langle \rho^+ V^{n-\frac{1}{2}}, V^{n-\frac{1}{2}} \rangle_V \\
& + \langle \rho^- w^{n-\frac{1}{2}}, w^{n-\frac{1}{2}} \rangle_w + \langle \rho^+ W^{n-\frac{1}{2}}, W^{n-\frac{1}{2}} \rangle_W
\end{aligned}
$$

$$
\begin{aligned}
\beta_2 =& \frac{\Delta t(\gamma_{ux} + \gamma_{uy} + \gamma_{uz})}{h} \|u^{n-\frac{1}{2}}\|_u^2 + \frac{\Delta t(\gamma_{Ux} + \gamma_{Uy} + \gamma_{Uz})}{h} \|U^{n-\frac{1}{2}}\|_U^2 \\
& + \frac{\Delta t(\gamma_{vx} + \gamma_{vy} + \gamma_{vz})}{h} \|v^{n-\frac{1}{2}}\|_v^2 + \frac{\Delta t(\gamma_{Vx} + \gamma_{Vy} + \gamma_{Vz})}{h} \|V^{n-\frac{1}{2}}\|_V^2 \\
& + \frac{\Delta t(\gamma_{wx} + \gamma_{wy} + \gamma_{wz})}{h} \|w^{n-\frac{1}{2}}\|_w^2 + \frac{\Delta t(\gamma_{Wx} + \gamma_{Wy} + \gamma_{Wz})}{h} \|W^{n-\frac{1}{2}}\|_W^2 \\
& + \frac{\Delta t}{h\gamma_{ux}} \|txx^n\|_{txx}^2 + \frac{\Delta t}{h\gamma_{vy}} \|tyy^n\|_{tyy}^2 + \frac{\Delta t}{h\gamma_{wz}} \|tzz^n\|_{tzz}^2 \\
& + (\frac{\Delta t}{h\gamma_{uy}} + \frac{\Delta t}{h\gamma_{vx}}) \|txy^n\|_{txy}^2 + (\frac{\Delta t}{h\gamma_{uz}} + \frac{\Delta t}{h\gamma_{wx}}) \|txz^n\|_{txz}^2 + (\frac{\Delta t}{h\gamma_{vz}} + \frac{\Delta t}{h\gamma_{wy}}) \|tyz^n\|_{tyz}^2 \\
& + \frac{\Delta t}{H\gamma_{Ux}} \|TXX^n\|_{TXX}^2 + \frac{\Delta t}{H\gamma_{Vy}} \|TYY^n\|_{TYY}^2 + \frac{\Delta t}{H\gamma_{Wz}} \|TZZ^n\|_{TZZ}^2 \\
& + (\frac{\Delta t}{H\gamma_{Uy}} + \frac{\Delta t}{H\gamma_{Vx}}) \|TXY^n\|_{TXY}^2 + (\frac{\Delta t}{H\gamma_{Uz}} + \frac{\Delta t}{H\gamma_{Wx}}) \|TXZ^n\|_{TXZ}^2 + (\frac{\Delta t}{H\gamma_{Vz}} + \frac{\Delta t}{H\gamma_{Wy}}) \|TYZ^n\|_{TYZ}^2
\end{aligned}
\tag{56}
$$

where $\gamma_{ux}, \gamma_{uy}, ..., \gamma_{Wz} > 0$ are coefficients left to be optimized later on, then

$$
\mathcal{K}_c^n \begin{cases} \leq \beta_1 + \beta_2 \\ \geq \beta_1 - \beta_2 \end{cases}
\tag{57}
$$

Eq.(53) and (57) imply that the total discrete energy on the composite grid is bounded above by a weighted sum of grid function norms, and it is also bounded blow by a weighted sum of grid function norms if the coefficient before the norm square of each grid function in $(\mathcal{S}_c^n + \beta_1 - \beta_2)$ is positive, i.e. the time step size has

to satisfy

$$
\begin{aligned}
\Delta t < \min\Bigg\{ & \frac{h\min\rho^-}{\gamma_{ux}+\gamma_{uy}+\gamma_{uz}}, \frac{H\min\rho^+}{\gamma_{Ux}+\gamma_{Uy}+\gamma_{Uz}}, \\
& \frac{h\min\rho^-}{\gamma_{vx}+\gamma_{vy}+\gamma_{vz}}, \frac{H\min\rho^+}{\gamma_{Vx}+\gamma_{Vy}+\gamma_{Vz}}, \\
& \frac{h\min\rho^-}{\gamma_{wx}+\gamma_{wy}+\gamma_{wz}}, \frac{H\min\rho^+}{\gamma_{Wx}+\gamma_{Wy}+\gamma_{Wz}}, \\
& h\min\{\gamma_{ux},\gamma_{vy},\gamma_{wz}\}\min e^-, H\min\{\gamma_{Ux},\gamma_{Vy},\gamma_{Wz}\}\min e^+, \\
& \frac{2h\gamma_{wy}\gamma_{vz}}{\max c44^-(\gamma_{wy}+\gamma_{vz})}, \frac{2h\gamma_{wx}\gamma_{uz}}{\max c55^-(\gamma_{wx}+\gamma_{uz})}, \frac{2h\gamma_{uy}\gamma_{vx}}{\max c66^-(\gamma_{uy}+\gamma_{vx})}, \\
& \frac{2H\gamma_{Wy}\gamma_{Vz}}{\max c44^+(\gamma_{Wy}+\gamma_{Vz})}, \frac{2H\gamma_{Wx}\gamma_{Uz}}{\max c55^+(\gamma_{Wx}+\gamma_{Uz})}, \frac{2H\gamma_{Uy}\gamma_{Vx}}{\max c66^+(\gamma_{Uy}+\gamma_{Vx})}\Bigg\}
\end{aligned}
\tag{3D.dt}
$$

where $\min\rho^-$ denote the minimum density on the fine grid, and meanings of other similar notations are straightforward. Therefore, once the time step size satisfies Eq.(3D.dt), the scheme qualifies for *boundedness*.

To compute as fast as possible, $\gamma_{ux},\gamma_{uy},...,\gamma_{Wz} > 0$ are then optimized to maximize the time step size.

1D NUMERICAL TESTS

In this section, I perform two series of experiments to numerically verify that in the 1D case both *conservation* and *boundedness* are indispensable to achieve a stable composite staggered grid scheme.

1D Wave Propagation Problem, Exact Solutions and Composite Grid Schemes

I consider the same 1D wave propagation problem as the one used in Petersson and Sjögreen (2010), except that they formulate the problem in terms of displacement and simulate it using the composite collocated grid scheme. The problem is

defined on $x \in [0, 1]$ and formulated as:

$$\begin{cases} v_t(x, t) = c^2 p_x(x, t) + g(t)\delta'(x - x_*) \\ p_t(x, t) = v_x(x, t) \end{cases} \tag{58}$$

where x_* is the source location; $\delta'(x)$ is the first derivative to the Dirac delta function; $c = 1$ represents the propagation speed over $[0, 1]$, and hence there is no material discontinuity in this case; the time function is polynomial:

$$g(t) = \begin{cases} -20t^7 + 70t^6 - 84t^5 + 35t^4, & 0 \le t < 1 \\ 1, & t > 1 \end{cases} \tag{59}$$

which is a smooth function with vanishing derivatives at $t = 0, 1$. Homogeneous Dirichlet boundary condition is imposed on two ends.

When $g(t) = t^n$ and $n \ge 2$, following derivations in Petersson and Sjögreen (2010), I obtain the exact solution for the velocity:

$$\begin{aligned} v(x, t) = & \frac{n(ct + x - x_*)^{n-1}}{2c^{n+1}} \mathcal{H}(x - x_* + ct) + \frac{n(ct - x + x_*)^{n-1}}{2c^{n+1}} \mathcal{H}(x - x_* - ct) \\ & - \frac{n(ct + x - x_*)^{n-1} + n(ct - x + x_*)^{n-1}}{2c^{n+1}} \mathcal{H}(x - x_*) \end{aligned} \tag{60}$$

where $\mathcal{H}$ is the Heaviside step function and t is chosen such that the wave does not hit the boundary because otherwise the solution will be more complex (actually I don't know how to give an analytical expression in this case). Note that $v(x, t)$ is discontinuous only at x_*: the presence of Heaviside step functions indicate that the only possible discontinuous locations are at $x_* - ct$, $x_* + ct$ and x_*; however, the polynomial term before $\mathcal{H}(x - x_* + ct)$ goes to 0 as x goes to $x_* - ct$, therefore the function is continuous at $x_* - ct$, and similarly, the function is also continuous at $x_* + ct$.

When $g(t) = t^n$ and $n \ge 2$, the exact solution for the pressure is

$$\begin{aligned} p(x, t) = & \frac{n(ct + x - x_*)^{n-1}}{2c^{n+2}} \mathcal{H}(x - x_* + ct) - \frac{n(ct - x + x_*)^{n-1}}{2c^{n+2}} \mathcal{H}(x - x_* - ct) \\ & + \frac{n(ct - x + x_*)^{n-1} - n(ct + x - x_*)^{n-1}}{2c^{n+2}} \mathcal{H}(x - x_*) - \frac{t^n}{c^2}\delta(x - x_*) \end{aligned} \tag{61}$$

Similar to the discontinuity arguments developed for the velocity, the pressure is also discontinuous only at x_*, but in this case the discontinuity is caused not

only by $\mathcal{H}(x - x_*)$ but also by $\delta(x - x_*)$, so the numerical solution goes to infinity when x goes to x_*, which flattens out solutions at other positions in the same plot frame. Therefore, to have a clear overall comparison between the exact solution and the numerical solution, I muted the pressure values for both numerical and exact solution near x_* in plots shown later.

The fine grid of the composite grid is on $[0, 0.5]$ and coarse grid on $[0.5, 1]$. The coarse grid size $H = 0.0125$ is twice as large as the fine grid size h. The intuitive scheme uses ghost data interpolation given by Eq.(9), whereas the stable scheme uses ghost data interpolation determined by the linear equation system given by Eq.(1D.ECI.a) and Eq.(1D.ECI.b). According to Eq.(1D.dt), the maximal time step size to maintain stability for the stable composite grid scheme is

$$\Delta t_{\max} = \min\left\{ \frac{h}{\gamma_1}, \frac{H}{\gamma_2}, \gamma_1 h, \gamma_2 H \right\} \tag{62}$$

When $\gamma_1 = \gamma_2 = 1$, $\Delta t_{\max}$ is maximized to attain h. In later experiments, I use $\Delta t = \frac{1}{2}h$.

Dirac Source Discretization

To approximate the source term $\delta'(x)$, I adopt the stencil approach as suggested by Petersson and Sjögreen (2010). Denote the grid function corresponding to $\delta'(x)$ by s_i^n on the fine grid and S_i^n on the coarse grid. The nonzero elements of the grid function (s_i^n, S_i^n) are aggregated only within a neighborhood of x_*, and they constitute a stencil that satisfies

$$\langle s_i^n, P_i^{(q),-} \rangle_{sp} + \langle S_i^n, P_i^{(q),+} \rangle_{rp} = -P'^{(q)}(x_*), \quad q = 0, \ldots, Q \tag{63}$$

where $P^{(q)}(x)$ is any polynomial in x of degree q, $P'^{(q)}(x)$ is its first derivative, and the grid function of $P^{(q)}(x)$ is denoted by $(P_i^{(q),-}, P_i^{(q),+})$.

Since the finite difference scheme is second order accurate in both time and space, the source approximation should be at least second order accurate in both time and space otherwise the accuracy of the finite difference scheme would be wasted. Waldén (1999) showed that Q has to be at least l to obtain overall l-th order convergence. Therefore in this paper, I choose $Q = 3$, and then Eq.(63) yields a

nonzero stencil that is 4-point wide. Fig. 4 shows that the nonzero stencil locations vary with different x_*. When x_* locates on either grid but is not close to the interface, the nonzero stencil points are entirely on one grid, as demonstrated in the first and the last case in Fig. 4; however, when x_* is close to the interface, the nonzero stencil is distributed on both grids as shown in the remaining cases in Fig. 4. Suppose the overlapping point, or the interface point is assigned with a weight of $\mathcal{I}$ as solved by Eq.(63), then as suggested in Petersson and Sjögreen (2010) I assign the weight on the fine grid interface point as $\mathcal{I}$, and $\frac{1}{2}\mathcal{I}$ for the coarse grid interface point.
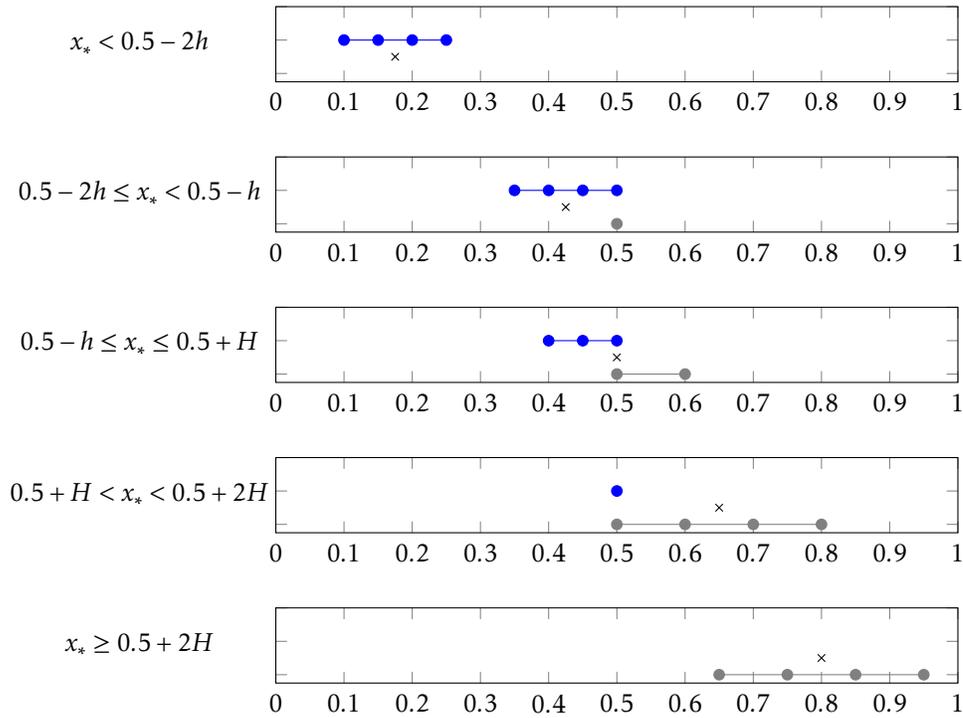
Figure 4: Distribution of nonzero elements in the approximated source term with varying x_*. The cross represents the location of x_*. Blue dots denote grid positions of nonzero s_i^n, gray dots for nonzero S_i^n.

Numerical Results and Discussion

The first series of experiments demonstrate that it is necessary to use the energy conserving ghost data interpolation.

I vary the source location x_* at five different positions on the composite grid as illustrated and categorized in Fig.4 and compare 1D wave propagation numerical results obtained by using the intuitive composite staggered grid scheme and the stable composite staggered grid scheme with its exact solutions. The results are shown in Fig.5 and 6.

When the source is very close to the interface (within two grid sizes), which creates a discontinuity in the solution near the interface, Fig.5 shows that the intuitive composite staggered grid scheme is not stable; whereas Fig.6 shows that the stable composite staggered grid scheme is stable regardless of the source-to-interface distance.

This result is consistent with the previous analyses that the intuitive ghost data interpolation can provide a good approximation only when the exact solution is smooth near the interface; however in this case the exact solution is not smooth near the interface because there is a first order derivative of Dirac delta source which is close to the interface. In contrast, the energy conserving ghost data interpolation successfully solves this issue by incorporating the source influence, see Eq.(25).

One may point out that both schemes fail to provide a good approximation to the velocity near x_* regardless of its location (except using the stable scheme when $x_* = 0.5 - 1.5h$); however, this inconsistency has been observed in Petersson and Sjögreen (2010) as well, and I suspect that this inconsistency is due to the source discretization, which also causes the slight mismatch between exact velocity and numerical velocity computed by the stable scheme when $x_* = 0.5 + 5H$.

The second series of experiments seek to demonstrate the necessity of restricting the time step size to qualify for *boundedness*. As derived earlier, the maximal time step size to maintain stability is equal to h. I examine the numerical solutions computed by the stable composite grid scheme by setting $\Delta t = 0.98h$ and $\Delta t = 1.02h$ respectively. The results shown in Fig. 7 reveal that in this case the restriction ((1D.dt)) on the time step size is not only necessary but also so sharp that any infinitesimal amount of increment would cause instability.
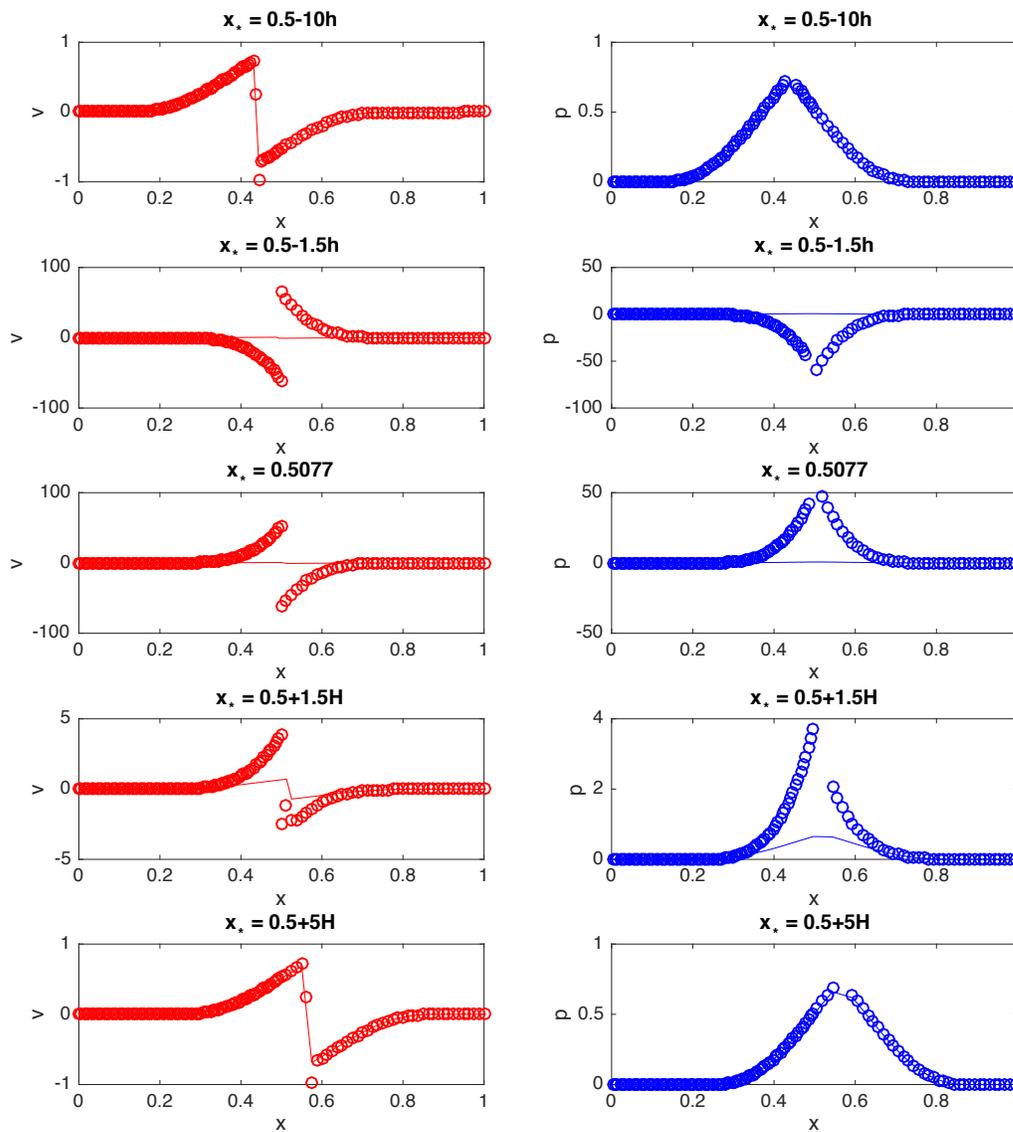
Figure 5: Comparison between exact solutions and numerical solutions computed by the intuitive composite staggered grid scheme with varying x_* location. All the snapshots were taken at $T = 0.33$. The left five panels show exact velocities (red line) and numerical velocities (red circles); the right five panels show exact pressures (blue line) and numerical pressures (blue circles).
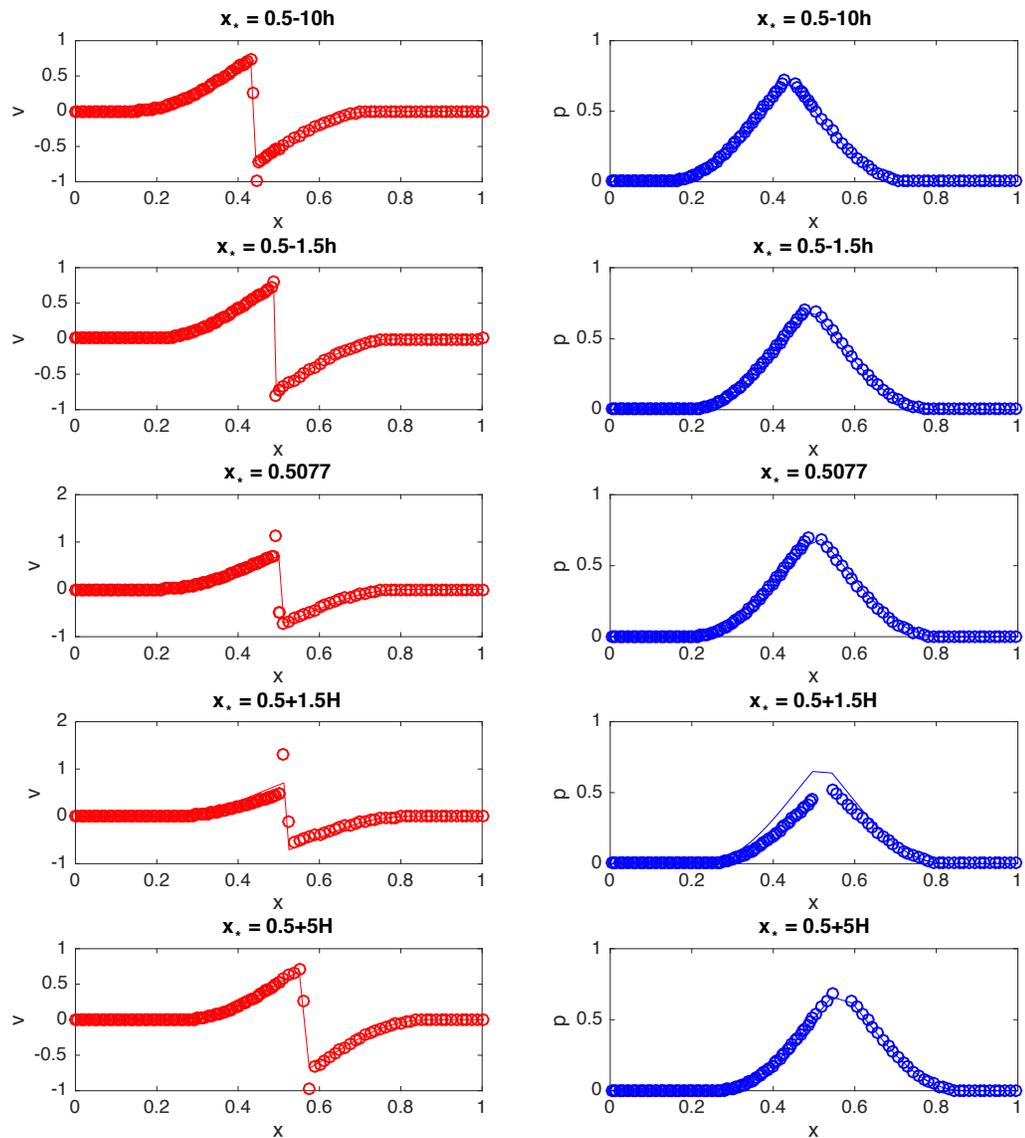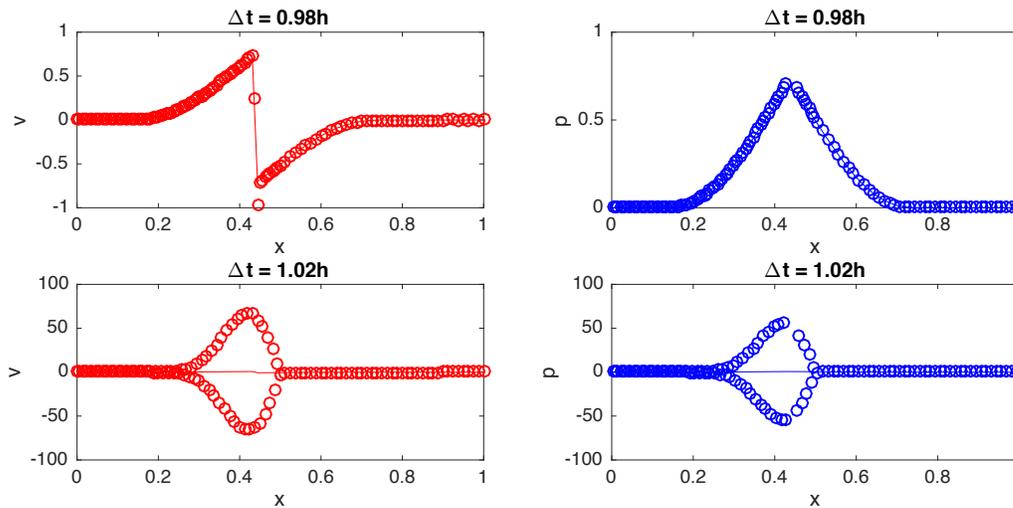
Figure 6: Comparison between exact solutions and numerical solutions computed by the stable composite staggered grid scheme with varying x_*. All the snapshots were taken at $T = 0.33$. The left five panels show exact velocities (red line) and numerical velocities (red circles); the right five panels show exact pressures (blue line) and numerical pressures (blue circles).

Figure 7: Comparison between exact solutions and numerical solutions computed by the stable composite staggered grid scheme with varying Δt. All the snapshots were taken at $T = 0.33$. The left two panels show exact velocities (red line) and numerical velocities (red circles); the right two panels show exact pressures (blue line) and numerical pressures (blue circles).

CONCLUSION AND FUTURE WORK

In this paper, I built a stable composite staggered grid scheme based on the first-order hyperbolic wave equation, on a rectangular domain composed of two subdomains with different characteristic wavelengths. The scheme is equipped with the homogeneous Dirichlet boundary condition. The 3D scheme can handle anisotropic (up to orthorhombic) elastic wave simulations. I use the energy method to show that if the homogeneous form of a scheme satisfies *conservaiton* and *boundedness*, then the scheme is stable. To qualify for *conservation*, the interface ghost data have to be updated in a way so that the energy is conserved. When the refinement ratio is equal to two, I show that this *conservation* qualification, together with the transmission condition, yields a linear equation system to solve for the ghost data, and this system is solvable under mild conditions. When the ghost data are updated to conserve the energy, the scheme can qualify for *boundedness* as long as the time step size is bounded above by the spatial grid size multiplied by some constant which are determined by only coefficients in the wave equation. Finally, The 1D numerical results show that the intuitive composite grid scheme, in which

the interface ghost data are straightforwardly interpolated from pressures nearby and so the energy may not be conserved, is not stable when the solution is not continuous near the interface, whereas the stable composite grid scheme keeps its stability in this case. In addition, the results also show that restricting the time step size is indispensable to achieve stability.

Future work includes:

- **Analyze convergency.** Lax-Richtmyer Equivalence Theorem states that for a linear and consistent numerical scheme, stability is equivalent to convergency. The stable composite grid scheme built in this paper is linear, stable and consistent at all interior grid points, but it is hard to analyze whether or not it is consistent at interface and boundary points. So Lax-Richtmyer Theorem can not be applied in this case unless I can prove consistency at the interface and the boundary.

- **Towards high order scheme.** Currently, the spatial finite difference operators are only of second order accuracy. Using high order spatial finite difference operators can reduce the number of points required per wavelength and hence it saves memory and computational cost (Levander, 1988). Therefore, it is desirable to extend the scheme using high order finite difference operators.

- **Implement the free surface boundary condition.** Free surface boundary condition is used to model the water-air interface, where the traction perpendicular to the interface is **0**, i.e.,

$$\sigma \cdot \mathbf{n} = \mathbf{0}.$$

  Implementing this boundary condition on finite difference scheme based on first-order hyperbolic wave equation is straightforward. If the stress components are aligned on the boundary, then set them to be zero; for those stress components that are half grid size to the boundary, then construct their ghost stress points with respect to the interface so that averaged stress values on the interface are zero. This modeling technique is known as *stress-imaging technique* (Levander, 1988), and this technique can provide an accurate and stable solution when modeling a planar free surface (Robertsson, 1996). In Petersson and Sjögreen's paper (Petersson and Sjögreen, 2010), they obtain a stable collocated grid scheme with free surface boundary condition. So I expect to achieve a stable staggered grid scheme with this boundary condition.

- **Implement the PML boundary condition.** The perfectly matched layer (PML) boundary condition (Bérenger, 1994) is used to to simulate wave propagations in infinite domain by minimizing spurious waves reflected from the artificial boundary, as if the computational domain extends to infinity and the outgoing waves never travel back. It wraps a layer around the computational domain which attenuates the wave energy at an exponential rate regardless of the wave incident angle to the boundary. PML has already been well developed for staggered grid finite difference scheme based on anisotropic (orthotropic or higher symmetric) elastic wave equation (Collino and Tsogka, 1998). In addition, Bécache (Bécache et al., 2001) has derived a necessary and a sufficient stability condition for this PML scheme.

- **Parallelization and performance tuning.** For parallelization, I would like to use MPI+OpenMP guided by HPCToolkit (Mellor-Crummey, 2015). The first task is to efficiently solve and parallelize the energy conserving ghost data interpolation. Petersson and Sjögreen (Petersson and Sjögreen, 2010) found that in practice an iterative block Jacobi relaxation method can efficiently solve these equation systems in parallel for composite collocated grid finite difference scheme.

  For fine tuning, I will explore cache optimizations and ensure full SIMD vectorization. In my master thesis (Zhou, 2014) I built a fully vectorized uniform collocated grid finite difference scheme based on the wave equation in second-order displacement form. I investigated both spatial and temporal cache optimizations, which are designed to be compatible with the OpenMP parallelization. The composite staggered grid scheme simulates more variables than the uniform collocated grid scheme, so it incurs higher memory pressure and hence cache optimizations might be more effective in improving the performance.

REFERENCES

Bécache, E., S. Fauqueux, and P. Joly, 2001, Stability of perfectly matched layers, group velocities and anisotropic waves: Technical report.

Bérenger, J.-P., 1994, A perfectly matched layer for the absorption of electromagnetic waves: Journal of computational physics, **114**, 185–200.

Bernth, H., and C. Chapman, 2011, A comparison of the dispersion relations for

anisotropic elastodynamic finite-difference grids: GEOPHYSICS, **76**, WA43–WA50.

Collino, F., and C. Tsogka, 1998, Application of the PML absorbing layer model to the linear elastodynamic problem in anisotropic heteregeneous media: Technical report.

Fehler, M., 2012, SEAM update: Elastic simulations: The Leading Edge, **31**, no. 1, 24–25.

Gustafsson, B., and E. Mossberg, 2004, Time compact high order dfference methods for wave propagation: SIAM Journal on Scientific Computing, **26**, 259–271.

Helbig, K., 1996, Kelvin and the early history of seismic anisotropy, *in* Seismic Anisotropy: Society of Exploration Geophysicists, 15–36.

Kristek, J., P. Moczo, and M. Galis, 2010, Stable discontinuous staggered grid in the finite-difference modeling of seismic motion: Geophysical Journal International, **183**, no. 3, 1401–1407.

Levander, A., 1988, Fourth-order finite-difference P-SV seismograms: Geophysics, **53**, 1425–1436.

Mellor-Crummey, J., 2015, Performance analysis of MPI+OpenMP programs with HPCToolkit.

Petersson, N. A., and B. Sjögreen, 2009, An energy absorbing far-field boundary condition for the elastic wave equation: Communications in Computational Physics, **6**, 483–508.

———, 2010, Stable grid refinement and singular source discretization for seismic wave simulations: Communications in Computational Physics, **8**, 1074–1110.

Robertsson, J. O., 1996, A numerical free-surface condition for elastic/viscoelastic finite-difference modeling in the presence of topography: Geophysics, **61**, 1921–1934.

Rodríguez, J., 2008, A spurious-free space-time mesh refinement for elastodynamics: International Journal for Multiscale Computational Engineering, **6**, 263–279.

Rubio, F., M. Hanzich, A. Farrés, J. de la Puente, and J. M. Cela, 2014, Finite-difference staggered grids in GPUs for anisotropic elastic wave propagation simulation: Computers & Geosciences, **70**, 181–189.

Tsvankin, I., 1997, Anisotropic parameters and P-wave velocity for orthorhombic media: GEOPHYSICS, **62**, 1292–1309.

Virieux, J., 1986, P-SV wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.

Waldén, J., 1999, On the approximation of singular source terms in differential equations: Numer. Methods Partial Differential Eq., **15**, 503–520.

Zhou, M., 2014, Wave equation based stencil optimizations on a multi-core CPU: Master's thesis, Rice University.

Algorithmic Aspects of Extended Waveform Inversion

William. W. Symes

ABSTRACT

Extended waveform inversion uses inference of an extended model, depending on non-physical parameters, to extract velocity information directly from waveform data. It has been observed a number of times that data fit at all stages of velocity updates, a side-effect of model extension, appears to be important to assure reliability of algorithms of this type. An analysis of the modeling operator derivative provides an explanation for this observation, and reveals the link between extended waveform inversion and traveltime tomography.
A version of this paper was presented at the EAGE 2015 Anual Meeting.

INTRODUCTION

Waveform velocity analysis uses *extended modeling* to detect inconsistency between velocity model and data: extra parameters are introduced into the model to enable data fit even in the initial stages of model updating. The final model must depend only trivially on these additional parameters, and model updates are constructed to reduce the dependence. Such methods may operate primarily in data domain or image domain, or combine elements of both, and may distantly or closely relate to full waveform inversion (FWI). Some recent contributions to this technology include Lameloise et al. (2014); Gao and Williamson (2014); Biondi and Almomin (2014); Weibull and Arntsen (2014) and a number of papers mentioned below. For extensive discussion and older references, see Symes (2008).

This paper concerns algorithmic features of waveform velocity analysis common to both image- and data-domain approaches. Kern and Symes (1994) describe an optimization formulation, and observe that in one particular case: (a) model updates should be nested, with short-scale components updated prior to each update

151

of long-scale velocities; (b) within each inner step, short-scale model components should be *inverted*, that is, made to fit data, rather than merely imaged; (c) computation of the gradient involves the adjoint derivative of the modeling operator, which may be computed with an RTM-like procedure, so Newton-like descent methods are feasible. The evidence for these assertions presented by Kern and Symes (1994) was partly numerical. The purpose of this paper is to identify the mathematical structure underlying these conclusions, thus demonstrating their applicability beyond the particular case considered in Kern and Symes (1994). I will also show how item (b) resolves the "gradient artifact" problem (Fei and Williamson, 2010; Vyas and Tang, 2010) for a natural class of objective. A byproduct of the resolution is positive semidefiniteness of the Hessian at noise-free data, which to some extent explains why these methods work.

THEORY

The variants of waveform modeling considered in this abstract are *separable*: the model consists of two components, a background model m and a linear component r. In many cases this division results from linearization ("Born approximation") of a nonlinear model: m is the background or reference model, r is a perturbation of m (Tarantola, 1984). In others, r is an intrinsically linear parameter, such as a trace-dependent source wavelet or filter (Plessix, 2000; Warner and Guasch, 2014).

Denote by $F[m]$ the linear modeling operator parametrized by m. $F[m]r$ is then predicted data from the separable model (m, r). For the cases discussed in this paper, $F[m]$ can be represented as an integral operator with oscillatory kernel, whose *normal operator* $F[m]^*F[m]$ (the superscript $*$ means adjoint or transpose) is also an oscillatory integral with phase function $\mathbf{k} \cdot (\mathbf{x} - \mathbf{y})$. The amplitudes of these operators are homogenous in wavenumber $\mathbf{k}$, to leading order, and the order can be arranged to be $= 0$ by appropriate choice of other model parameters (for example, source wavelet). See for example ten Kroode et al. (1998); Stolk et al. (2009). Operators of this type have a *calculus*: sums, adjoints, and (operator) products are operators of the same type. The order follows the same rules as for differential operators: in particular, the product AB of operators A of order p and B of order q has order at most $p+q$, but the *commutator* $[A, B] = AB - BA$ has order at most $p+q-1$. Indeed the common name for this operator type is *pseudodifferential* (Taylor, 1981). The importance of order stems from the effect of these operators

on frequency components of their inputs: for an operator of order p, components of wavenumber $\mathbf{k}$ are scaled roughly by k^p in the output (again, similar to the effect of differential operators).

An *annihilator A* is an operator whose null space consists of the non-extended models, that is, those not depending on the extra extension variables. For the separable models considered here, only the linear model parameter r need be extended. Using $\|\cdot\|$ to denote norm (RMS), and $\langle\cdot,\cdot\rangle$ inner product, a canonical formulation of the extended waveform inversion problem is

$$\text{minimize } \frac{1}{2}\|Ar\|^2 \text{ subject to } F[m]r - d = 0 \tag{1}$$

Because of the added degrees of freedom provided by extended modeling, essentially arbitrary data may be fit essentially perfectly: ignoring some technical issues, write $r = F[m]^{-1}d$ to satisfy the constraint. $F[m]^{-1}$ is an operator of the same type as $F[m]$. The objective in equation (1) may be written as

$$J_{\text{DSO}}[m] = \frac{1}{2}\|AF[m]^{-1}d\|^2 = \frac{1}{2}\langle d, (F[m]^*)^{-1}A^*AF[m]^{-1}d\rangle \tag{2}$$

If A^*A is pseudodifferential, then the operator "sandwich" in the middle of the inner product is also pseudodifferential, and its amplitude depends on the model through ray tracing and various algebraic combinations. $J_{\text{DSO}}[m]$ is therefore smooth in m: the sizes of its derivatives are essentially independent of data frequency content. Conversely, in at least some cases smoothness of $J_{\text{DSO}}[m]$ requires that A^*A be pseudodifferential (Stolk and Symes, 2003).

Analysis of the gradient and Hessian of $J_{\text{DSO}}[m]$ requires analysis of the derivative DF of F (Symes, 2014; ten Kroode, 2014):

$$DF[m, \delta m]r = \lim_{h \to 0} \frac{1}{h}(F[m + h\delta m] - F[m])r = F[m]Q[m, \delta m]r, \tag{3}$$

in which $Q[m]$ is an operator of order 1 - that is, it scales the input at wavenuber $\mathbf{k}$ by a factor proportional to k, thus enhancing high frequency components. Q is skew-symmetric to leading order: $Q + Q^*$ is an operator of order 0.

EXAMPLES

A simple setting for waveform velocity analysis is linearized constant density acoustics. The dynamical laws for this model consist of two couple partial dif-

ferential equations:

$$\frac{1}{v^2}\frac{\partial^2 p}{\partial t^2} - \nabla^2 p = f; \; \frac{1}{v^2}\frac{\partial^2 \delta p}{\partial t^2} - \nabla^2 \delta p = 2\frac{\delta v}{v^3}\frac{\partial^2 p}{\partial t^2}; \; p, \delta p = 0, t < 0 \tag{4}$$

plus suitable boundary conditions. Sampling the pressure perturbation δp at receiver locations defines a separable modeling operator: $F[m]r = \{\delta p(x = x_r, z = z_r, t)\}$, with $m = v^{-2}$ and $r = \delta v/v$. The derivative $DF[m, \delta m]$ is related in the same way to the solution of a coupled system of four wave equations, not written here.

To illustrate the effect of Q, I used a linearized model derived from the Marmousi exaample (Versteeg and Grau, 1991), and a normally incident plane wave source with the $-1/2$ derivative of a (2, 3.5, 10, 12.5) Hz bandpass filter as waveform. Even over the small bandwidth of this example, this presence of an additional derivative (factor of frequency) in the output of DF becomes obvious in a plot of average trace spectra - see Figure 1.
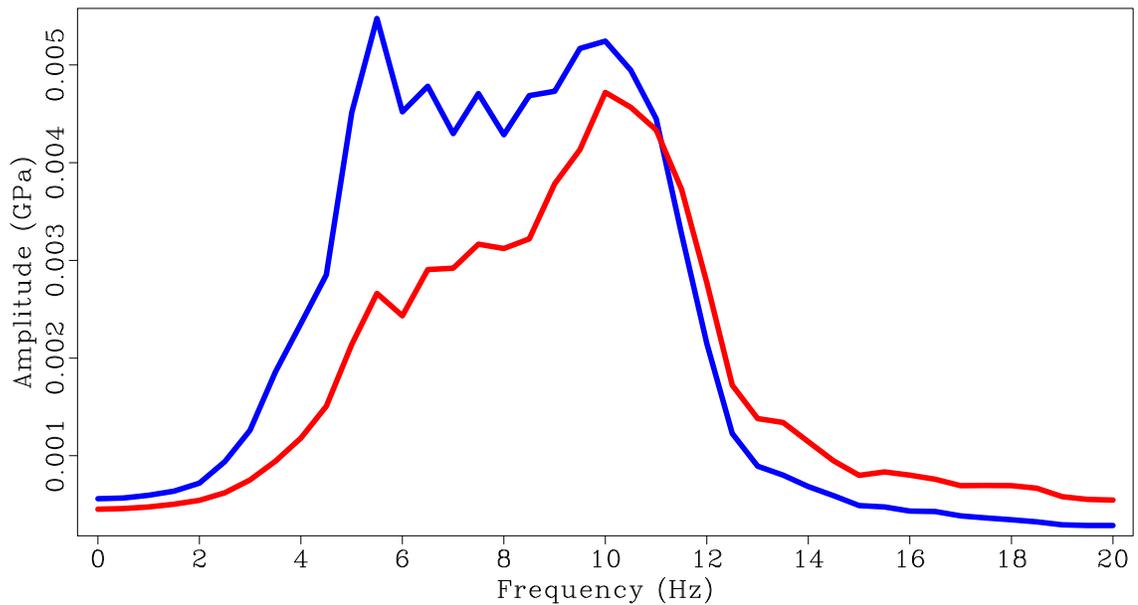
Figure 1: Average spectra of typical linearized modeling output (blue line) and of its model derivative (red line). madrid15WS/project/ specs

RESULTS

Several central facts about waveform inversion in the linearized, separable setting explained above follow from the derivative structure theorem (8).

To begin with, a straightforward version of linearized full waveform inversion could be written in the notation introduced above as: given d, find m, r to

$$\text{minimize } \frac{1}{2}\|F[m]r - d\|^2 \text{ subject to } Ar = 0 \tag{5}$$

The derivative of the least squares objective (5) with respect to the background model m is

$$\langle F[m]r - d, DF[m, \delta m]r \rangle = \langle F[m]^*(F[m]r - d), Q[m, \delta m]r \rangle.$$

Because $F[m]$ is of order 0, the size of the objective is roughly proportional to the norm of d. However, the derivative of the objective, involving as it does an operator of order 1, becomes very large if d or r have high-frequency components. This disparity in size between the objective and its model derivative, arbitrarily large if arbitrarily large wavenumbers are permitted in the data and/or linear model, is another way of looking at the phenomenon commonly known as "cycle skipping", and is responsible for the multimodal nature of the least squares objective.

Remarkably, switching the role of objective and constraint creates the optimization problem (1), with very different properties from those just described. As will be explained below, the derivative structure theorem implies that (1) exhibits local convexity and relations to traveltime tomography that (5) does not.

Some earlier work on (1) used approximation $\tilde{J}_{\text{DSO}}$ of J_{DSO}, in which the migrated image $F[m]^*d$ is substituted for the inversion $F[m]^{-1}d$ (Shen et al., 2003). The attraction of this formulation is clear, as the adjoint or transpose operator may have much lower computational complexity than the inverse. However, several authors have pointed out that the gradient of this modified objective has a tendency to oscillate in space, detracting from its utility as a model update (Fei and Williamson, 2010; Vyas and Tang, 2010). This feature is not a defect in the gradient, but in the modified objective. With $F[m]^*d$ in place of $F[m]^{-1}d$, a quick calculation shows that

$$D^2\tilde{J}_{\text{DSO}}[m](\delta m_1, \delta m_2) = \langle AQ[m, \delta m_1]^*F[m]^*d, AQ[m, \delta m_2]^*F[m]^*d \rangle +$$

$$\langle AF[m]^*d, AQ[m,\delta m_1]^*Q[m,\delta m_2]F[m]^*d\rangle + ... \tag{6}$$

in which the ellipses stand for similar terms. Now presume that the data is actually consistent with the model: $F[m]r = d, Ar = 0$ - note however that there is no guarantee that m is a global minimizer of $\tilde{J}_{\mathrm{DSO}}$. The Hessian form becomes

$$D^2\tilde{J}_{\mathrm{DSO}}[m](\delta m_1,\delta m_2) = \langle [A,Q[m,\delta m_1]^*F[m]^*F[m]]r, [A,Q[m,\delta m_2]^*F[m]^*F[m]]r\rangle +$$

$$\langle [A^*A, F[m]^*F[m]]r, Q[m,\delta m_1]^*Q[m,\delta m_2]F[m]^*F[m]r\rangle + ... \tag{7}$$

The first term on the RHS is a positive semidefinite quadratic form of order 0, stable against (mean-square) small errors in r, and in fact related to traveltime error (Symes, 2014), therefore tomographic. The second term however dominates: for oscillatory (therefore reflective) r, the net order 1 operator in of this form makes it far more important than the first, and $\tilde{J}_{\mathrm{DSO}}$ is not locally convex, hence not globally minimized, at the "truth". This observation accounts for the gradient "artifacts" mentioned earlier, as well as the minimization of $\tilde{J}_{\mathrm{DSO}}$ at data-inconsistent models observed by Khoury et al. (2006). It justifies the assertion of Kern and Symes (1994) that inversion is required.

Returning to the definition proposed in equation (1),

$$DJ_{\mathrm{DSO}}[m]\delta m = -\langle AF[m]^{-1}DF[m,\delta m]F[m]^{-1}, AF[m]^{-1}d\rangle$$

$$= -\langle Q[m,\delta m]F[m]^{-1}d, A^*AF[m]^{-1}d\rangle \tag{8}$$

From equation (11) follows a formula for the gradient of J_{DSO} in terms of the adjoint derivative: for a suitable choice of inner product $\langle \cdot,\cdot\rangle_M$ in background model space, $\langle DF^*[m,d],\delta m\rangle_M = \langle d, DF[m,\delta m]\rangle$. Kern and Symes (1994) showed how to compute the adjoint, also known as the *tomographic operator* (Biondi and Sava, 2004), using an RTM-like procedure.

At model-consistent data, each appearance of $F[m]^*F[m]$ in (7) is replaced by the identity operator, and many terms disappear, leaving the stable semidefinite form

$$D^2J_{\mathrm{DSO}}[m](\delta m_1,\delta m_2) = \langle [A,Q[m,\delta m_1]]r, [A,Q[m,\delta m_2]]r\rangle \tag{9}$$

with tomographic meaning in specific instances (Symes, 2014). Further, m is a global minimizer of J_{DSO}. Thus aspects of m that are controlled by certain forms of tomography are also controlled by J_{DSO}.

CONCLUSIONS

This abstract has briefly reviewed the theoretical foundations of extended wave-form inversion. The basic lessons of Kern and Symes (1994) - that inversion should be nested, with short scale components in the inner loop, and these should be inverted, not merely imaged (an observation also made by Lameloise et al. (2014)) - have been justified using recently developed understanding of the modeling operator and its derivatives.

As mentioned before, certain technical issues have been overlooked in this discussion. Most importantly, $F[m]$ is invertible in only a few special cases. Generally, it is only invertible approximately (modulo lower-frequency errors, and requires regularization) and microlocally, that is, in certain regions of model phase space. These caveats do not invalidate the conclusions stated above in any significant way, but complete arguments are considerably more complex. Recently, computable RTM-like approximate inverses with smoothing errors have been established for several important modeling cases (ten Kroode, 2012; Hou and Symes, 2014). These are essential ingredients in practical implementation.

ACKNOWLEDGEMENTS

REFERENCES

Biondi, B., and A. Almomin, 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.

Fei, W., and P. Williamson, 2010, On the gradient artifacts in migration velocity analysis based on differential semblance optimization: 80th Annual Inter-

national Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4071–4076.

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.

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.

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, 2014, Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis: Geophysical Prospecting, **62**, (in press).

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

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., 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, 2003, Smooth objective functionals for seismic velocity inversion: Inverse Problems, **19**, 73–89.

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

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

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

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: Presented at the Inverse Problems - from Theory to Application, Proceedings, Institute of Physics.

ten Kroode, F., D. Smit, and A. Verdel, 1998, A microlocal analysis of migration: Wave Motion, **28**, 149–172.

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

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, 2014, Anisotropic migration velocity analysis using reverse time migration: Geophysics, **79**, R13–R25.

Full Waveform Inversion via Matched Source Extension

Guanghui Huang and William W. Symes

ABSTRACT

Matched Source Waveform Inversion introduces additional degrees of freedom into waveform modeling in the form of trace-dependent source modification, allowing close data fit at all stages of the inversion process. Penalizing source modification leads to an optimization problem with the same global minimum as Full Waveform Inversion, but with less tendency to develop local minima caused by cycle-skipping. We implement and analyze a simple variant of this technique, using constant density acoustics and a transmission configuration, and explain the very close relation of this method to traveltime tomography for mildly heterogeneous velocity models. In common with other data domain waveform tomography methods, the Matched Source objective function may develop multiple local minima, despite avoiding cycle-skip in data residual, if multiple ray paths connect sources and receivers.

A version of this paper was presented at the SEG 2015 Anuual International Meeting.

INTRODUCTION

Full waveform inversion via nonlinear least squares tends to produce many, mostly noninformative local minima (Gauthier et al., 1986; Virieux and Operto, 2009), a serious difficulty since efficient local optimization methods are mandatory due to problem size. The root cause of local minima is *cycle-skipping*: the tendency of synthetic data to be more than a half-wavelength out of phase with observed data, hence enjoying spurious fit. Many recent advances in FWI technology have reduced this impediment - most obviously, it lessens with low-freqquency, high s/n data, However the difficulty remains, and has inspired many researchers to

suggest alternative approaches. One family of alternatives relies on *extended modeling*, that is, additional degrees of freedom leading to closer data fit, hence less tendency to cycle-skip, throughout the inversion process - see Symes (2008) for an overview of these ideas.

This paper reviews a particularly simple extended modeling variant, Matched Source Inversion, which exhibits several important features of other extended model inversion, but is simpler to analyse. The additional degrees of freedom in Matched Source Inversion are introduced through permitting source waveforms to depend on source and receiver coordinates.

THEORY

The discussion to follow uses a simple model of seismic trace data synthesis based on constant density linear acoustics: the excess pressure field $p(\mathbf{x}, t; \mathbf{x}_s)$ generated by a causal isotropic point radiator at source position $\mathbf{x} = \mathbf{x}_s$ solves the wave equation

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

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

The forward modeling operator $S[v, f]$ relates the velocity field $v(x, z)$ and wavelet function $f(t)$ to the scattered field at the receiver,

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

in which $\mathbf{x}_r$ and $\mathbf{x}_s$ range over survey source and receiver positions.

With these conventions, Full Waveform Inversion may be stated as follows: given recorded traces $d(\mathbf{x}_r, t; \mathbf{x}_s)$, find v and f so that $S[v, f] \simeq d$. A common formulation asks that v, f be chosen to minimize the residual sum of squares (Tarantola, 1984; Virieux and Operto, 2009):

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

The objective function J_{FWI} is quadratic in f, but quite non-quadratic in v:

Extended modeling and matched sources

The tendency of FWI to cycle-skip is fundamentally due to the failure to fit data, given kinematically wrong v. This misbehaviour disappears if the source wavelet *depends on the trace*: that is, replace $f(t)$ on the RHS of equation (1) with $\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s)$, call the resulting source-receiver dependent pressure field $\bar{p}$ and define a new modeling operator by $\bar{S}[v, \bar{f}](\mathbf{x}_r, t; \mathbf{x}_s) = \bar{p}(\mathbf{x}_r, t; \mathbf{x}_s)$. Note that computing $\bar{S}$ does not require solving a wave equation for each source-receiver pair, since

$$\bar{S}[v, \bar{f}] = \bar{f} \star S[v, \delta] \tag{4}$$

(trace-by-trace convolution with the causal Green's function of (1)).

If $\bar{f}(\mathbf{x}_r, t; \mathbf{x}_s) = f(t)$ is independent of source and receiver coordinates: then $\bar{S}[v, \bar{f}] = S[v, f]$. This observation suggests supplementing the data-fit condition

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

with a penalty for source dependence on source-receiver coordinates. Several such penalties have been suggested. For example, assume a transmission configuration: suppose that the sources $\mathbf{x}_s$ lie in a plane, say $x = x_s$, and the receivers lie in another, say $x = x_r$. Song and Symes (1994), Symes (1994) suggest (essentially) minimizing the mean-square of differences of neighboring source traces, that is, the mean square of $\nabla_{y_s, z_s, y_r, z_r} \bar{f}$. Plessix et al. (2000), Plessix (2000) applied this "differential semblance" concept to invert field crosswell data. Warner and Guasch (2014) notes that $\bar{f} = f^*$ implies that the match filter $\bar{w}$ required to make $\bar{w} \star f^* = \bar{f}$ is an impulse at zero lag for every trace, hence in the null space of multiplication by t. Luo and Sava (2011) use a similar approach, and Plessix et al. (2000) used a moment penalty to enhance the stability of the differential semblance approach.

Analysis of the gradient

All of the approaches explained in the last paragraph take the form:

$$J_{\mathrm{MS}}[v] = \frac{1}{2} \sum_{\mathbf{x}_s, \mathbf{x}_r} \int dt \, |A\bar{f}|^2 \tag{6}$$

subject to the data fit condition (5)). This condition makes $\bar{f}$, hence J_{MS} implicitly a function of v. A is an *annihilator*-like operator, such as $\nabla_{y_r,\dots}$ or multiplication of the match filter by t, whose null space consists of trace-independent $\bar{f}$.

The gradient of J_{MS} can be computed in the same way as the gradient of J_{FWI}, via the adjoint state method (that is, RTM). On the other hand, in some cases the structure of the gradient follows from a remarkable relation between the modeling operator $\bar{S}$ and its derivative, or first order perturbation with respect to velocity $D_v\bar{S}$. This analysis reveals a very close link between Matched Source Inversion and traveltime tomography.

Presume also that v is slowly varying (smooth), so that the geometric acoustics approximation to the Green's function is accurate, and that for all source-receiver pairs in the survey, a single-arrival traveltime field τ, with geometric ampiltude a, describes the propagation. That is,

$$S[v,\delta](\mathbf{x}_r,t;\mathbf{x}_s) \approx a(\mathbf{x}_r;\mathbf{x}_s)\delta(t - \tau(\mathbf{x}_r,\mathbf{x}_s)). \tag{7}$$

The remarkable relation mentioned above is this: up to an error of lower order in frequency,

$$D\bar{S}[v,\bar{f}]\delta v \approx \bar{S}[v,Q[v,\delta v,\bar{f}]] \tag{8}$$

In this expression, $Q[v,\delta v,\bar{f}]$ is linear in δv and $\bar{f}$ (separately). Q is essentially a derivative, in its action on $\bar{f}$: it is given by an Fourier-like integral whose amplitude contains a factor of ω. Also, Q is approximately skew adjoint: $Q^T + Q \approx 0$, the error not involving derivatives.

For the very simple matched source extended model discussed here, this relation follows directly from the geometric optics approximation (7), in fact with an explicit formula: $Q[v,\delta v,\bar{f}] = -(D\tau[v]\delta v)(\partial_t\bar{f})$, where $D\tau[v]\delta v$ is the first order perturbation of traveltime τ. Similar relations hold for more complex extended models: see for example Symes (2014) and ten Kroode (2014) for similar relation in space-shift (subsurface offset) extended modeling.

Evidently

$$DJ_{\text{MS}}[v]\delta v \approx \sum_{\mathbf{x}_r,\mathbf{x}_s} \int dt\,(A^T A)\bar{f}(D\tau[v]\delta v)\partial_t\bar{f}$$

so

$$\nabla J_{\text{MS}}[v] \approx \sum_{\mathbf{x}_r, \mathbf{x}_s} D\tau[v]^T \left(\int dt \, (\partial_t \bar{f})(A^T A \bar{f}) \right) \tag{9}$$

Here $D\tau[v]^T$ is the adjoint operator of $D\tau[v]$, which backprojects its argument along rays (Song and Symes, 1994)

To see what information is backprojected, we examine two of the possibilities mentioned above. For simplicity, consider 2D wave propagation and a single receiver position $z = z_r$ in the plane $\{x = x_r\}$, and assume that the data is noise-free and well-approximated by geometric acoustics (7):

$$d(\mathbf{x}_r, t; \mathbf{x}_s) \approx a^*(\mathbf{x}_r, \mathbf{x}_s) f^*(t - \tau(\mathbf{x}_r, \mathbf{x}_s)). \tag{10}$$

Denote by $\Delta\tau(z_r, z_s) = \tau[v^*](z_r, z_s) - \tau[v](z_r, z_s)$ the traveltime residual. Then from equations (5) and (7),

$$\bar{f} \approx (a^*/a) f^*(t - \Delta\tau)$$

1) differential semblance (Song and Symes, 1994; Plessix et al., 2000): $A = \partial_{z_s}$, then

$$\int dt \, \partial_t \bar{f} A^T A \bar{f} \approx -\left(\frac{a^*}{a} \right)^2 \left(\int dt \left| \frac{\partial f^*}{\partial t} \right|^2 \right)$$

$$\times \left(\frac{\partial}{\partial z_s} \right)^T \left(\frac{\partial}{\partial z_s} \Delta\tau(z_r, z_s) \right) + (...). \tag{11}$$

up to an error of lower order in peak frequency of f^*; this is equivalent to backprojecting $\partial_{z_s} \Delta\tau$ by $D_v(\partial_{z_s}\tau)^T$, a form of stereotomographic update (Billette and Lambaré, 1998).

2) moment of $\bar{f}$ (after Luo and Sava (2011), Warner and Guasch (2014) - note that Adaptive Waveform Inversion as in Warner and Guasch (2014) employs the match filter mentioned earlier ($\bar{w}$ above) hence differs somewhat from this construction): $A =$ multiplication by t,

$$\int dt \, \partial_t \bar{f} A^T A \bar{f} \approx (a^*/a)^2 \int dt \, t^2 \frac{1}{2} \partial_t (f^*(t - \Delta\tau))^2$$

$$\approx -\Delta\tau (a^*/a)^2 \int dt \, (f^*)^2. \tag{12}$$

where f^* is assumed to be zero-phase.

That is, the gradient of J_{MS} is closely related to the gradient of a tomographic objective - backprojection of either traveltime residual or traveltime slope residual, scaled by the amplitude ratio.

Analysis of the Hessian

J_{MS} is locally convex near a local minimizer if its Hessian $D^2 J_{MS}$ is positive semidefinite. For consistent (noise-free) data, the global minimum is zero, so $A\bar{f} = 0$, and as a result many terms vanish, leaving

$$D^2 J_{MS}[v^*](\delta v, \delta v) \approx \sum_{\mathbf{x}_s, \mathbf{x}_r} \int dt \, |[A, Q](v, \delta v, \bar{f})|^2$$

The notation $[A, Q](...)$ means $AQ[v, \delta v, \bar{f}] - Q[v, \delta v, A\bar{f}]$. A similar conclusion holds for other variants of extended modeling, see (Symes, 2014; ten Kroode, 2014; Symes, 2015). Substituting the explicit expression for Q derived above, obtain the following approximations to
$D^2 J_{MS}[v^*](\delta v, \delta v)$:

1) for $A = \partial_{z_s}$,

$$\approx \left(\int dt \left| \frac{\partial f^*}{\partial t} \right|^2 \right) \left(\sum_{\mathbf{x}_s, \mathbf{x}_r} \left| \frac{\partial}{\partial z_s} D\tau[v^*](\delta v) \right|^2 \right).$$

2) for $A = $ multiply $\bar{f}$ by t,

$$\approx \left(\int dt \, |f^*|^2 \right) \sum_{\mathbf{x}_s, \mathbf{x}_r} |D\tau[v^*]\delta v|^2.$$

That is, the Hessian of J_{MS} is proportional to the Hessian of a traveltime objective function: at least locally, J_{MS} is as convex as a tomographic objective, and minimizing it determines the same information about the velocity, in the high frequency limit.

NUMERICAL EXAMPLES

We focus on the numerical performance of the annihinator $A = t$ (case (2) above), and assume that $f^* =$ Ricker wavelet with peak frequency 10 Hz. Implementation of $\bar{S}$ is via equation (4), with $S[v, \delta]$ approximated via finite difference modeling (numerical Green's function), and $\bar{f}$ estimated via trace-by-trace deconvolution.

Example 1. *Plot objective function for various values of velocity.*

Figure 1 plots the values of the objective function for data from a homogeneous medium $v = 2$ km/s, with sources and receivers in lines $x = const.$ separated by 1 km. At least along this line segment in velocity model space, the only stationary point occurs at the global minimizer.
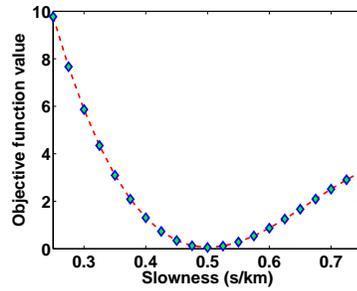
Figure 1: Matched source objective function for homogeneous velocity 0.25 s/km $\leq v^{-1} \leq 0.75$ s/km. Correct velocity is 2 km/s.

Example 2. *In this example, the target velocity consists of two Gaussian velocity anomalies embedded in a $v = 2$km/s background:*

$$v(x, z) = 2 - 0.6e^{-\frac{(x-0.25)^2+(z-0.3)^2}{(0.2)^2}} - 0.6e^{-\frac{(x-0.25)^2+(z-0.7)^2}{(0.1)^2}},$$

where $x \in [0, 0.5]$km, $z \in [0, 1]$km. The initial model is given by the constant velocity $v_0 = 2$ km/s. 50 shots are simulated on $\mathbf{x} : x_s = 0.01$ km, from $z = 0.01$ km to $z = 0.99$ km with $\Delta z_s = 0.02$ km. For each shot, 99 receivers are uniformly distributed from $z = 0.01$ km to $z = 0.99$ km at $x_r = 0.49$ km. See Figure 3a

The inversion procedure computes the gradient of J_{MS} by the adjoint state method - not by use of formula (9), in fact no use whatever is made of geometric acoustics in the computations. These computations are input to the Limited Memory BFGS algorithm (Nocedal and Wright, 1999) globalized with a backtracking line search.

The estimated velocity v_{est} after 50 LBFGS iterations, beginning with $v_0 = 2$ km/s, appears as the right-hand panel in Figure 3b. We compare the target data with the re-synthesized data, using the inverted velocity v_{est} and the known wavelet f^* - see Figure 2a. The data fit error, computed in this way, is about 23%, and the velocity is close enough that most of the remaining residual could be eliminated by FWI. Note that the extended model residual $\bar{S}[v_{est}, \bar{f}_{est}] - d$ is negligible by construction.

DISCUSSION

Since the inversion algorithm does not require ray-tracing computations, one might wonder if the relation between traveltime tomography and the matched source objective function would hold without the assumption of single arrivals (only one raypath connecting source and receiver. The following example suggests otherwise:

Example 3. *This example uses precisely the same velocity field as the previous one, but places the source and receiver lines 1 km apart. The anomalies are centered between the lines (see Figure 7a). The initial model is also given by the constant velocity $v_0 = 2$ km/s. Apart from the distance between the lines being doubled, the source-receiver geometry is the same.*

A typical shot gather in the left of Figure 6a shows that the later arrivals are now distinct and have energy comparable to the first. Resimulation using the estimated velocity from 50 iterations of the procedure previously described, starting from $v_0 = 2$ km/s (right hand panel in Figure 7b), fits the first arrival reasonably well but is otherwise kinematically incorrect, and would not be a useful starting estimate for FWI.

The reason for this failure is the presence of events in the estimated source $\bar{f}$ at time lags equal to differences of traveltime branches between the trial and target velocities - see Figure 8. These difference branches do not collapse to zero

lag, but must cancel, which does not happen until the predicted traveltimes coincide with the observed traveltimes for all branches, not just the first arrivals, to within a half wavelength. Therefore matched source inversion is prone to an effect similar to cycle-skipping once multiple arrivals achieve sufficient energy. See Symes (1994) for a complete discussion and more synthetic examples, Plessix et al. (2000); Plessix (2000) for more discussion of the implications in the crosswell context, Nolan and Symes (1997); Stolk and Symes (2004) for examples of similar pathology afflicting other data-domain model extensions, and Symes (2008); Stolk et al. (2009) for an account of extensions that are immune, at least to some extent.
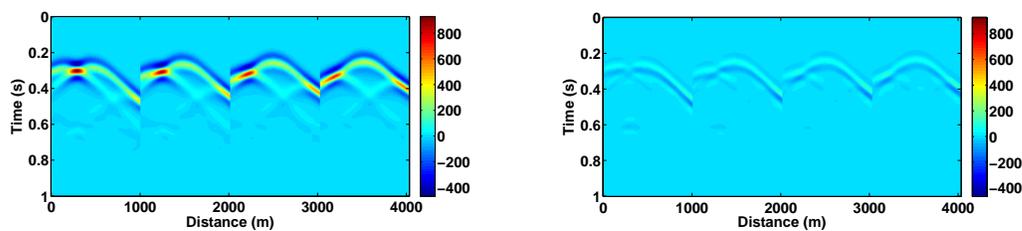
Figure 2: Four typical shots synthesized data (up) and difference between synthesized and inverted data (bottom)

CONCLUSION

We have shown a close relation between Matched Source Waveform Inversion and traveltime tomography, in the case of simple ray geometry, for both moment-type and differential semblance annihilators. Numerical examples show that this relation persists if the energy carried by later arrivals is not dominant. However with sufficient energy propagating along multiple ray paths, an effect akin to cycle-skipping takes place, and MSWI fails in the same fashion as standard FWI. Our examples use an idealized crosswell configuration, but the same conclusions should apply to a matched source approach to diving wave tomography.

ACKNOWLEDGEMENTS
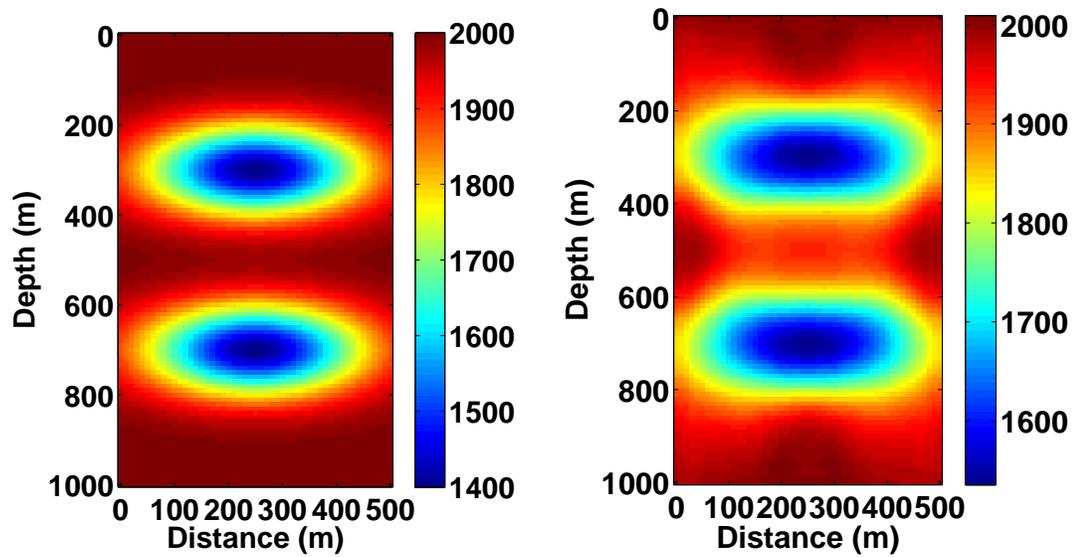
Figure 3: True model (left) and inverted model (right) after 50 iterations, velocity model fitting error is 2.2%.
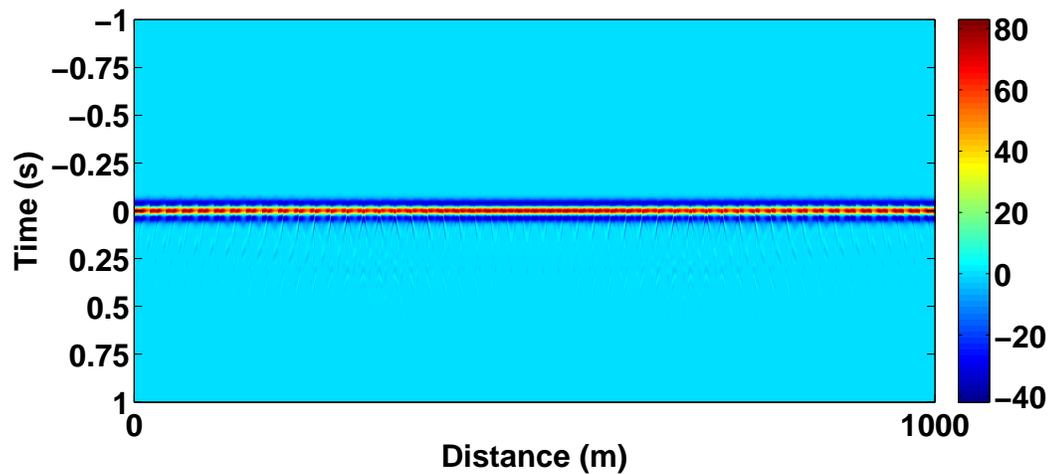
Figure 4: Extended source functions
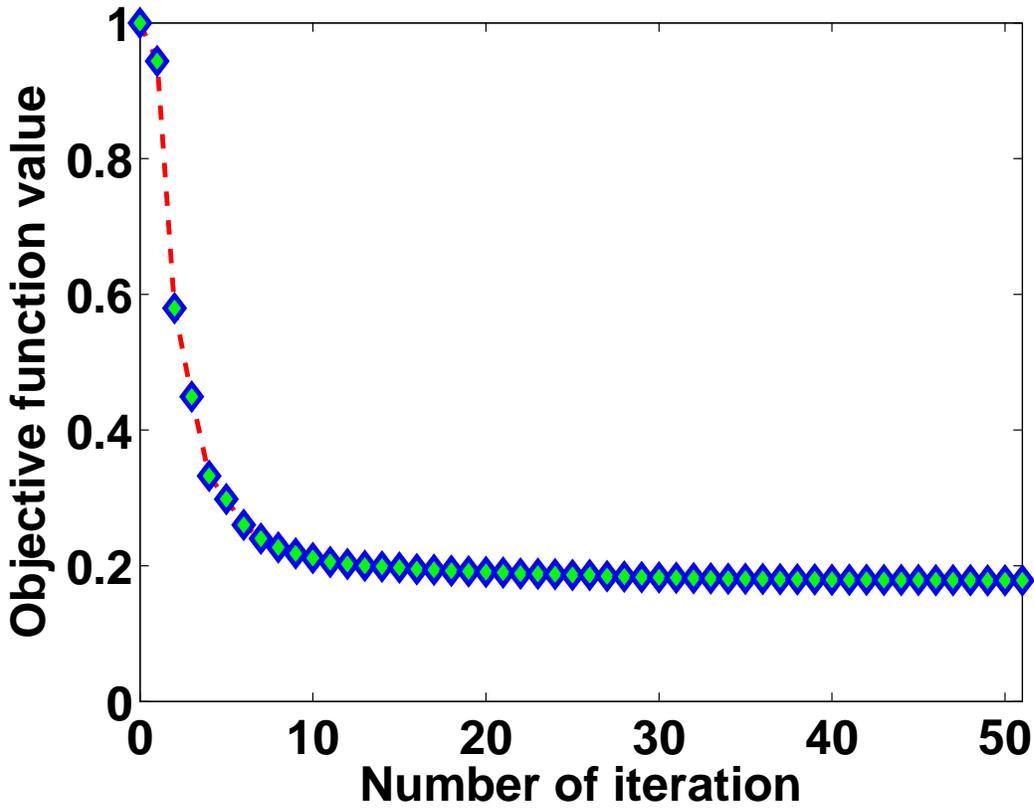
Figure 5: Convergence history of MSWI
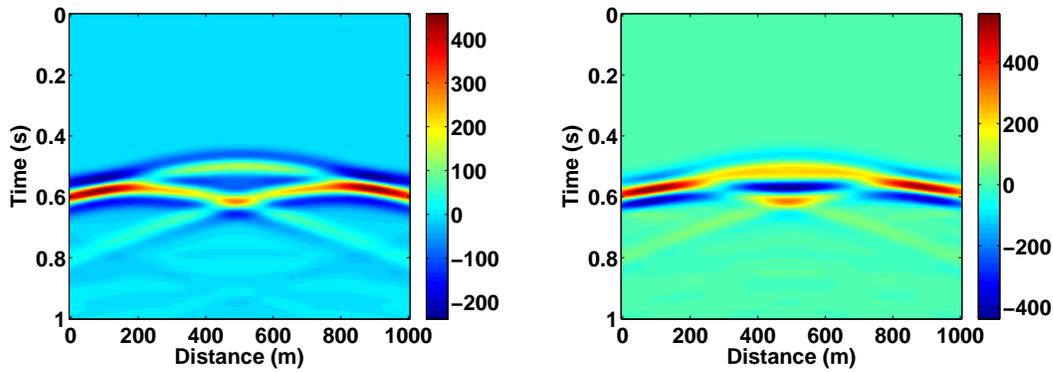
Figure 6: Single shot synthesized data (left) and difference between synthesized and inverted data (right) at shot position $z = 510$m
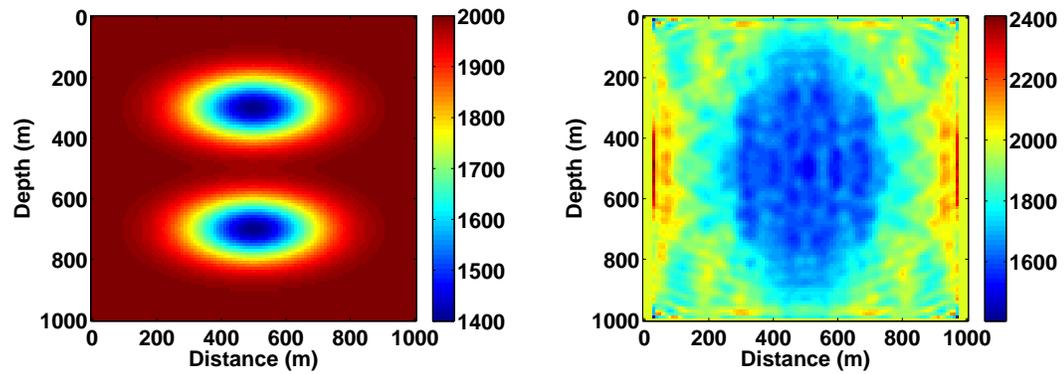
Figure 7: True model (left) and inverted model (right) after 50 iterations.
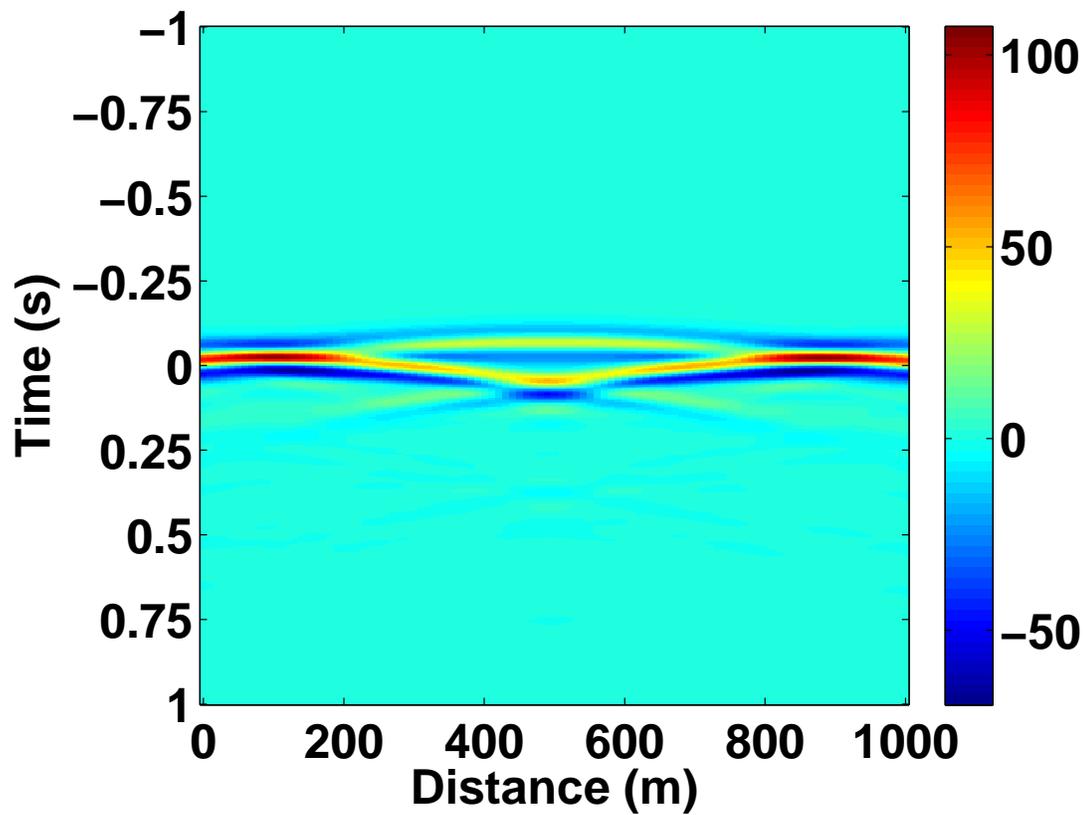
Figure 8: Extended source function at $z_s = 510m$
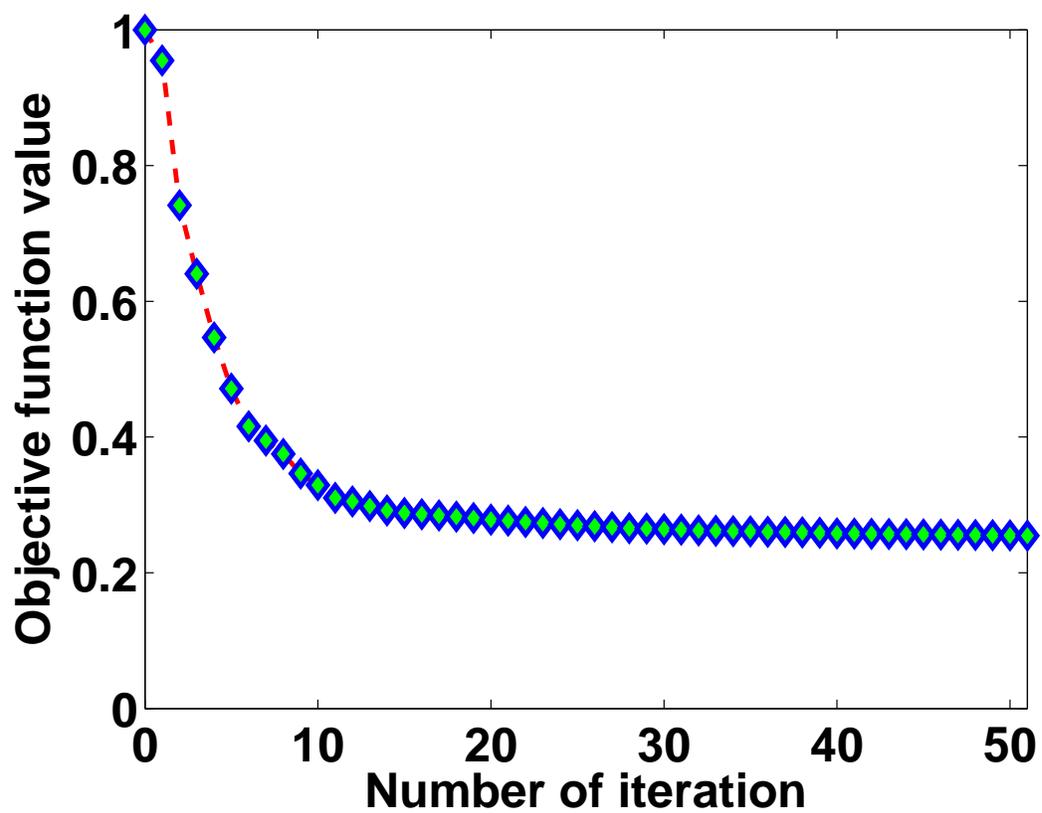
Figure 9: Convergence history of MSWI

REFERENCES

Billette, F., and G. Lambaré, 1998, Velocity macro-model estimation from seismic reflection data by stereotomography: Geophysical Journal International, **135**, 671–680.

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

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.

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

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

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.

Song, H., and W. Symes, 1994, Inverison 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.

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: Presented at the Inverse Problems - from Theory to Application, Proceedings, Institute of Physics.

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

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: Presented at the Inverse Problems - from Theory to Application, Proceedings, Institute of Physics.

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.

Accelerating Extended Least Squares Migration with Weighted Conjugate Gradient Iteration

Jie Hou, William W. Symes

ABSTRACT

Least Squares Migration (LSM) iteratively achieves a mean square best fit to seismic reflection data, provided that a kinematically accurate velocity model is supplied. The subsurface offset extension adds extra degrees of freedom to the model, thereby allowing LSM to fit the data even in the event of significant velocity error. This type of extension also implies additional expense per iteration from cross-correlating source and receiver wavefields over the subsurface offset, and therefore places a premium on rapid convergence. We accelerate the convergence of Extended Least Squares Migration, by combining the Conjugate Gradient algorithm with weighted norms in range (data) and domain (model) spaces that render the extended Born modeling operator approximately unitary. Numerical examples demonstrate that the proposed algorithm dramatically reduces the number of iterations required to achieve a given level of fit or gradient reduction, compared to Conjugate Gradient iteration with Euclidean (unweighted) norms.

A version of this paper was presented at the SEG 2015 Anuual International Meeting.

INTRODUCTION

Least Squares Migration (LSM) iteratively seeks a short-scale reflectivity model so as to achieve a best fit to the seismic reflection data in least squares sense, via repeated migrations and demigrations(Bourgeois et al., 1989; Nemeth et al., 1999; Kuehl and Sacchi, 2003). Since it is an inversion, the amplitudes are likely to be physically reasonable, and many authors have noted the effective spatial deconvolution and increase in apparent resolution resulting from data fitting (Dutta et al., 2014).

However, ability to fit the data depends critically on the accuracy of the background velocity model. Velocity error leads to mispositioned and defocused structures in the image domain, and related data misfit. That is as one would expect: the model depends on fewer parameters than the data, so only in special case (correct velocity) can the data be fit well throughout.

A natural solution is to extend the model with extra dimensions, to equalize the model and data dimension. One possibility is described in Stolk and De Hoop (2005), Symes (2008), Stolk et al. (2009): add an internal offset variable to the model, so that incident wavefield at one point can interact with (cause) a reflected wavefield at a positive distance. Extended Least Squares Migration (ELSM) with this subsurface offset extension is able to fit the data equally well with correct or incorrect background velocity model (Liu et al., 2013).

On the other hand, LSM generally requires tens of iterations for an acceptable result, each costing as much as two migrations. The additional parameters in ELSM add extra cost: for subsurface offset extension, computational loops over the offset axes are implicit in the definition of extended modeling or migration. This additional cost could be alleviated in two ways : either lower the cost for each iteration or accelerate the convergence rate. The first kind of acceleration may involve simultaneous source (Beasley et al., 1998), blended data (Berkhout, 2008) or phase encoding method (Romero et al., 2000; Ikelle, 2007), together with various technologies to suppress the crosstalk artifacts (Krebs et al., 2009; Schuster et al., 2011; Xue et al., 2014). This paper discusses only the second kind of acceleration, which is often accomplished by preconditioner. Numerous preconditioners have been investigated in this scope, such as approximate diagonal of Hessian (Pratt, 1999; Shin et al., 2001; Tang, 2009), debluring filter (Aoki and Schuster, 2009). Recently ten Kroode (2012) explained the construction of a computable approximate inverse to the subsurface offset extended Kirchhoff modeling operator. Hou and Symes (2014) modified ten Kroode's construction for the Born operator. Since an approximate inverse can often be used as a preconditioner, it seems likely that this concept could be the source of convergence acceleration.

This paper points out that the approximate inverse $F^{\dagger}$ of the subsurface offset extended Born modeling operator F takes the form

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

in which F^T is the transpose of F (extended RTM), W_{model} and W_{data} are positive-definite symmetric operators on model and data spaces respectively, explicitly

computable directly from knowledge of the velocity field. Since $F^\dagger F \approx I$, it follows that $F^\dagger$ is approximately unitary with respect to the norms defined by W_{model} and W_{data}. Therefore a properly formulated Conjugate Gradient (or LSQR) algorithm will converge very rapidly, much more rapidly than the same algorithm formulated with the Euclidean norm. We will explain exactly how to compute W_{model} and W_{data}, and numerically verify the dramatic improvement in convergence rate.

We first review the theory of ELSM and the approximate inverse operator. We then explain how to compute the weight operators, and how to write CG iteration to accommodate them. We end with two synthetic examples, demonstrating WCG algorithm possesses faster convergence rate comparing to normal CG algorithm.

THEORY

The solution of the constant density acoustic wave equation can be expressed as

$$\mathcal{F}[v] = u \tag{2}$$

where $\mathcal{F}$ is the full wave equation solution or modeling operator, v is the velocity model, u is the acoustic potential. Most of seismic imaging technologies depend on the (partial) linearization of the nonlinear modeling operator $\mathcal{F}$ with respect to v, which is often referred to as Born (linear, single scattering) approximation. Write $v = v_0 + \delta v$, and treat the influence of δv by first order perturbation (linearization). The corresponding approximation $u \approx u_0 + \delta u$ is most accurate when v_0 is smooth (transparent) on the wavelength scale, and all model oscillations (reflectivity) are confined to δv (Symes, 2009). Then the Born approximation is expressed similarly to (2) as

$$F[v_0]\delta v = \delta u \tag{3}$$

where $F = \dfrac{\partial \mathcal{F}}{\partial v}$ is the linear Born modeling operator.

Least squares migration (LSM) (Nemeth et al., 1999; Kuehl and Sacchi, 2003) is synonymous with solution of the least squares problem

$$J_{LS} = \frac{1}{2}\|F[v_0]\delta v - \delta d\|^2[+\text{Regularizing terms}], \tag{4}$$

or equivalent to solving the normal equation:

$$F[v_0]^\dagger F[v_0]\delta v = F[v_0]^\dagger \delta d. \tag{5}$$

where $F^\dagger$ is the adjoint relative to the choice of norms in data and model spaces. LSM will generate a model fitting the data as well as possible, but that may not be very well for even relatively small velocity errors. Figure 1 shows LSM images of a flat reflector in correct and incorrect constant background velocity respectively. The wrong velocity image is mispositioned and has incorrect amplitude, but it is also not an inversion, i.e. does not fit the data well (misfit plot is shown in Figure 3a).
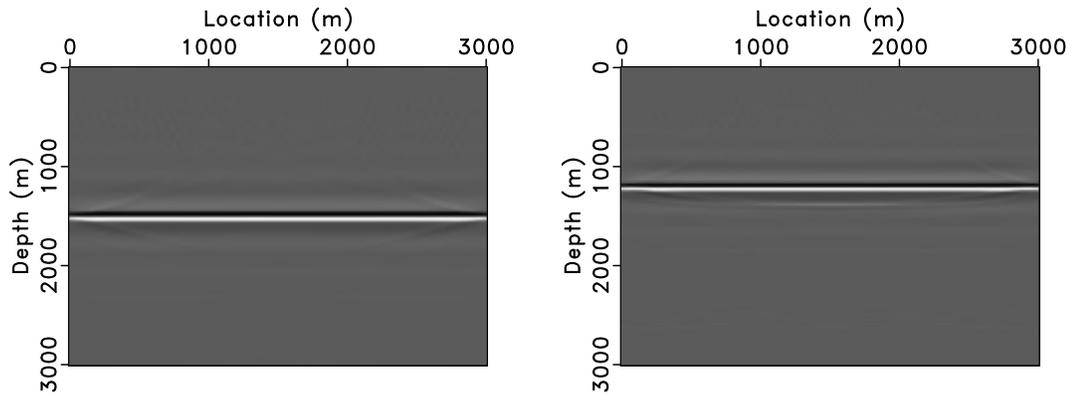
Figure 1: Least Squares Migration using (a) correct background velocity model (b) incorrect background velocity model (90%)

Extended Least Squares Migration

In order to equalize the model dimension with data dimension, it is natural to introduce extra degrees of freedom by extending the physical model. Common choices for additional dimensions include reflection angle and surface offset. Another appropriate extension is the subsurface offset, which is essentially the (horizontal) offset h between sunken source and sunken receiver in Claerbout's survey-sinking imaging condition (Claerbout, 1985; Symes, 2008; Stolk et al., 2009).

Physical (non-extended) models δv give rise to subsurface offset extended models $\delta\bar{v}$ via multiplication by $\delta(h)$: in 2D, $\delta\bar{v}(x,z,h) = \delta v(x,z)\delta(h)$. That is, as extended models, physical models are *focused.* See Figure 2 for an illustration of the 2D variant, which we will discuss for the remainder of this paper. The operator extension $\bar{F}$ integrates $\delta\bar{v}$, over h, hence coincides with F when $\delta\bar{v}$ is physical. Note
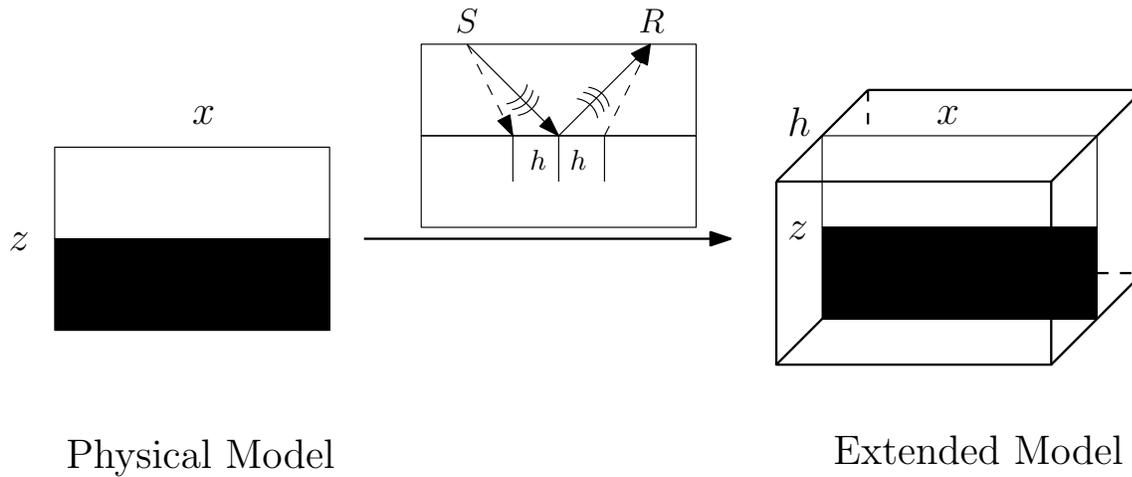
Figure 2: Sketch of subsurface offset extended model

that the data space is the same in both cases. We call least squares migration with extended operators (solution of (4) or (5) with F replaced by $\bar{F}$) Extended Least Squares Migration (ELSM).

Figure 1 shows LSM and ELSM model estimates computed with Conjugate Gradient (CG) iteration (Nocedal and Wright, 1999), in a precise sense the best iterative method for this type of problem. With the extra dimension, all the data information can be preserved in the model space. Correct velocity model will force the energy focus at $h = 0$ section. Incorrect background velocity model will spread event energy to nonzero h. The LSM data residual is large for incorrect velocity, but the ELSM data residual remains small with correct or incorrect velocity. Figure 3 compares the relative misfit plot between LSM and ELSM as a function of conjugate gradient iteration, confirming ELSM's tolerance on velocity error.

However, the integration over h makes $\bar{F}$ more expensive than F, therefore fast convergence is very desirable. The convergence rate of CG depends on the operator spectrum: error components associated with clustered eigenvalues (for example, near 1) of the normal equation (5) are reduced by a large factor in a single iteration. Since the definition of the operator adjoint $\bar{F}^\dagger$ depends on the norms chosen for domain and range spaces, accelerating convergence can be accomplished by choosing these norms to move many eigenvalues of $\bar{F}^\dagger\bar{F}$ close to 1.
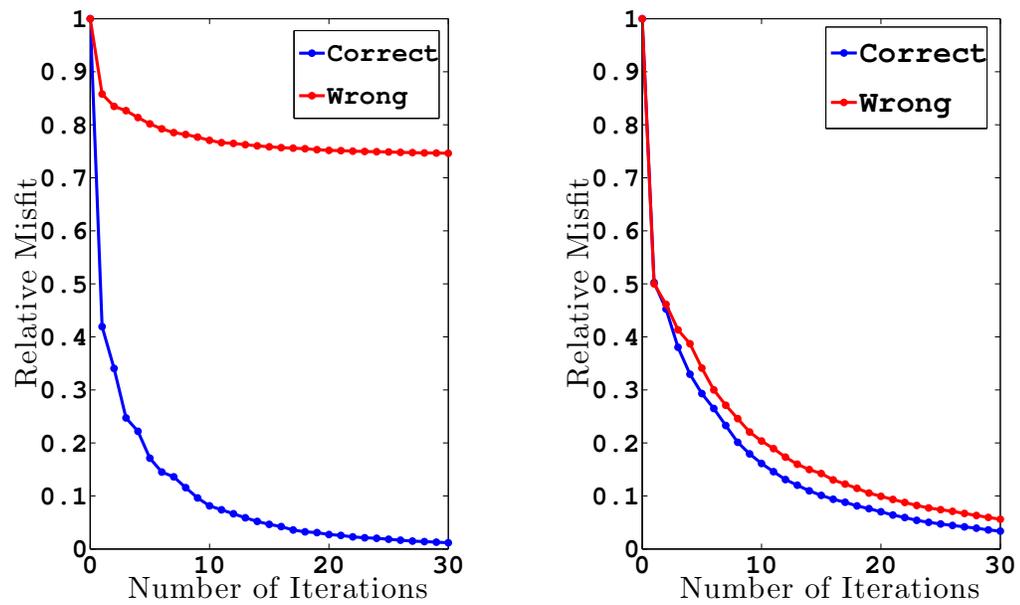
Figure 3: Misfit plot for (a) LSM and (b) ELSM with correct and wrong velocity model. Notice ELSM will converge to zero no matter with correct or incorrect velocity model.

Unitary Property of Extended Modeling

The subsurface offset extension actually makes it possible to establish a computable approximate inverse to the extended Born modeling operator (ten Kroode, 2012; Hou and Symes, 2014). This approximate inverse has exactly the form (5), with

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

where $L = \sqrt{\nabla_{(x,z)}^2 \nabla_{(h,z)}^2}$, I_t is time integration, F^T is the Euclidean adjoint of extend Born modeling operator (extended RTM) and D_{z_s}, D_{z_r} are the source and receiver depth derivatives, which may be expressed as square root operators in t, x_s, x_r, due to the field at the receivers being upcoming (free surface not included in model!). P is a Fourier-like operator whose amplitude is a known algebraic function of $v_0(x,z), v_0(x \pm h, z)$, and (k_x, k_z, k_h). $P = 1$ for $h = 0$. The expression above for W_{model} is not symmetric, but can be symmetrized with negligible error. Both weight operators are positive definite. See ten Kroode (2012); Hou and Symes (2014) for further discussion.

Weighted Conjugate Gradient Algorithm

Use the weight operators of the last section to define norms in model and data space:

$$\|\delta \bar{v}\|_{\text{model}}^2 = \sum_{x,z,h} \delta \bar{v}(x,z,h)(W_{\text{model}} \delta \bar{v})(x,z,h),$$

and similarly for data space. We call the following algorithm *Weighted CG* (WCG), even though it is really just the CG algorithm, because of the weight operators involved in relating the norms and adjoints to their Euclidean counterparts:

Algorithm 3 Weighted Conjugate Gradient Algorithm

1: $z_0 \leftarrow F^\dagger(d - Fx_0)$

2: $p_0 \leftarrow z_0$

3: $k \leftarrow 0$

4: **repeat**

5:     $\alpha_k \leftarrow \dfrac{< z_k, z_k >_{\text{model}}}{< Fp_k, Fp_k >_{\text{data}}}$

6:     $x_{k+1} \leftarrow x_k + \alpha_k p_k$

7:     $z_{k+1} \leftarrow z_k - \alpha_k F^\dagger Fp_k$

8:     $\beta_{k+1} \leftarrow \dfrac{< z_{k+1}, z_{k+1} >_{\text{model}}}{< z_k, z_k >_{\text{model}}}$

9:     $p_{k+1} \leftarrow z_{k+1} + \beta_{k+1} p_k$

10:    $k \leftarrow k + 1$

11: **until** Error is sufficiently small

NUMERICAL EXAMPLES

In this section, we compare ELSM using CG and WCG on a 2D model with simple structures and on a salt model.

Simple Example

The first example is a layered model with simple structures. The background velocity model and reflectivity model are shown in Figure 4. A (2.5-5-30-35) Hz bandpass wavelet with 1ms time interval is used to generate 76 shots data on the surface.

Both CG and WCG algorithm are used to carry out ELSM. The images after 20 iterations appear in Figure 5. The zero offset section for both cases depicts the main structures of the model with focused energy. However, one can observe obvious lower resolution for the CG result comparing to the WCG result.

As discussed in Hou and Symes (2014), stacking the output extended image over offset can produce a physical image. The stacked image from WCG result shown in Figure 6a perfectly reconstruct the original reflectivity model. The difference
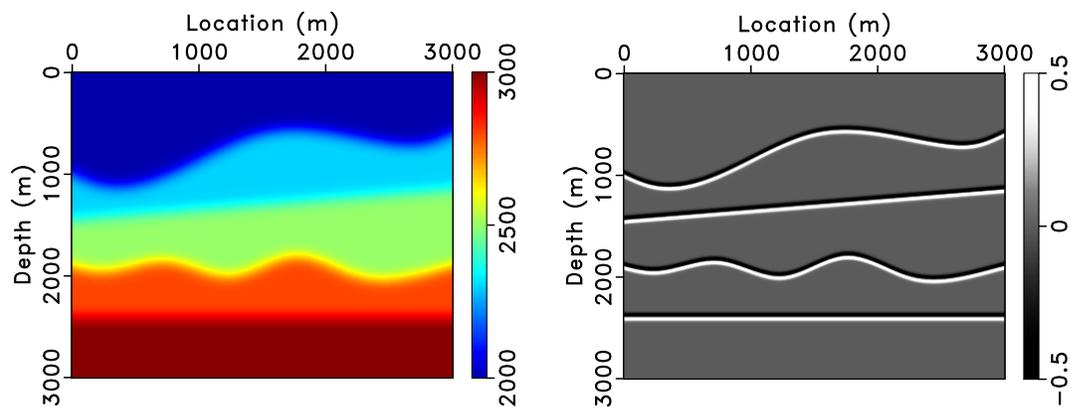
Figure 4: (a) Background velocity model (b) Reflectivity model
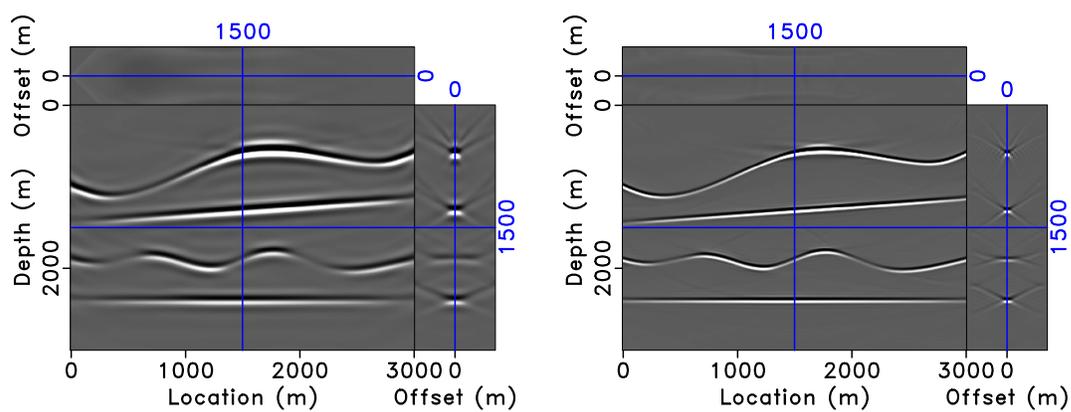
Figure 5: 20 iteration ELSM via (a) CG (b) WCG

with the original model (Figure 6b) confirms this observation.
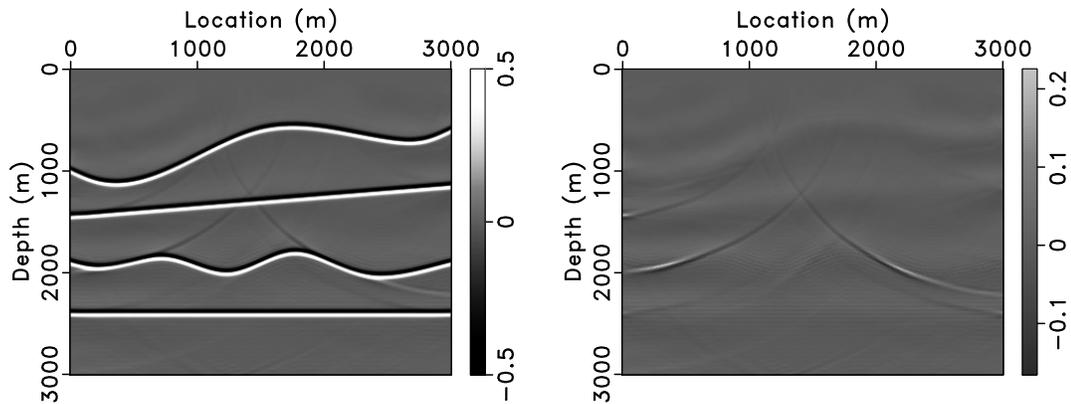
Figure 6: (a) Stacked image from WCG result (b) Difference with original reflectivity model. Plotted on the same grey scale.

The misfit plot, displayed in Figure 7a, exhibits the remarkable acceleration for WCG over CG. The required number of iterations for a given error level in WCG is substantially smaller comparing to CG. As a matter of fact, the first iteration result of WCG, which is the inverse operator itself, is comparable to the 20 iteration result of CG. The Normal Residual $F^T(Fm_k - d)$ plot, from a different perspective, confirms the fast convergence of WCG.

Salt Example

The salt example described here is based on a 2D section of the SEG/EAGE salt model(Aminzadeh et al. (1997)). The background velocity model shown in Figure 8a is achieved by smoothing the original salt model. The difference between the background model and original velocity model is used as the reflectivity model, displayed as Figure 8b. The model is discretized with a vertical and horizontal space of 20m. The velocity discontinuity at the salt boundary violates the hypotheses underlying the theory of the approximate inverse (Hou and Symes, 2014). We will use this example to show that the WCG algorithm has an improved convergent rate even in the presence of salt.

Both CG and WCG are used to generate the extended images, displayed in Figure
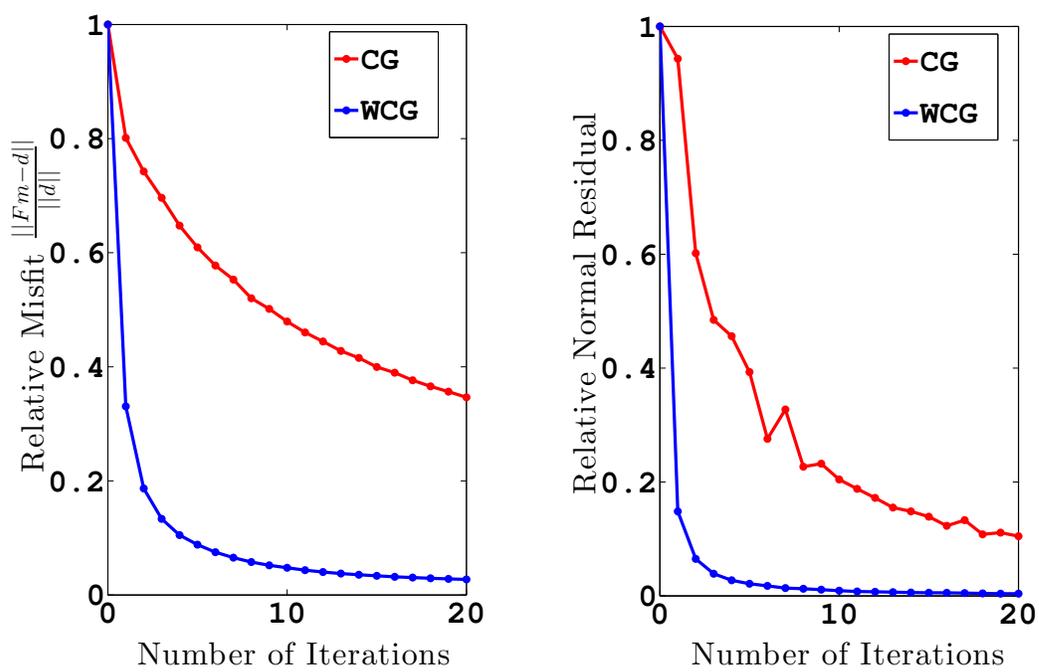
Figure 7: (a) Relative Misfit (b) Relative Normal Residual
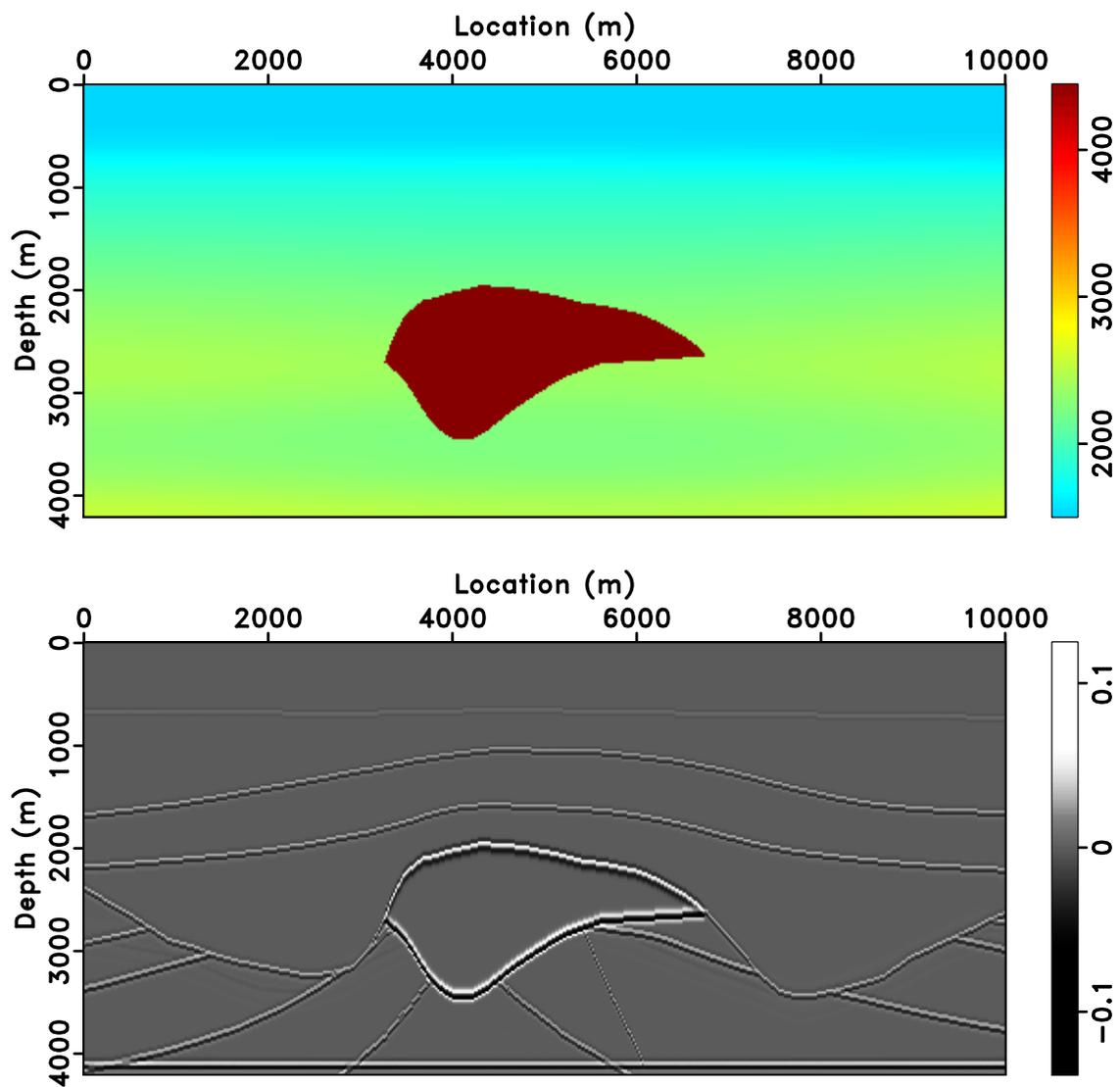
Figure 8: (a) Background velocity model (b) Reflectivity model

9.

Main structures and boundary of the salt are clearly observed for both cases. Classical low frequency noise above the salt is still present in the CG result while WCG remove most of the top salt noise. Stacking over offset yields the image shown in Figure 10.

Figure 11 shows the misfit plot and normal residual plot for CG and WCG. From the figure, it is clear that the proposed WCG has substantially faster convergence for this example also.

CONCLUSION

We propose a WCG algorithm, the CG algorithm with weighted norms chosen from the construction of an approximate inverse operator, to accelerate ELSM. Numerical examples suggest that the WCG iteration converges more rapidly to a better extended LSM estimate of reflectivity than does unweighted CG iteration.

ACKNOWLEDGMENTS

REFERENCES

Aminzadeh, F., J. Brac, and T. Kunz, 1997, 3-D salt and overthrust models: SEG.
Aoki, N., and G. Schuster, 2009, Fast least-squares migration with a deblurring filter: Geophysics, **74**, no. 6, WCA83–WCA93.
Beasley, C. J., R. E. Chambers, and Z. Jiang, 1998, A new look at simultaneous

Figure 9: 20 iteration ELSM via (a) CG (b) WCG
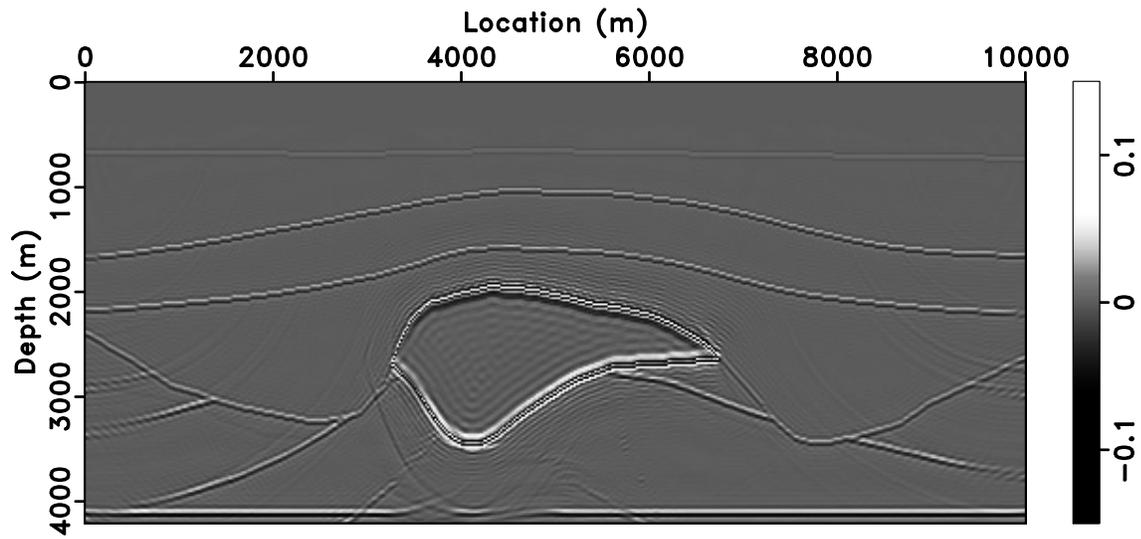
Figure 10: Stacked Image from WCG result. Plotted on the same grey scale with the reflectivity model.
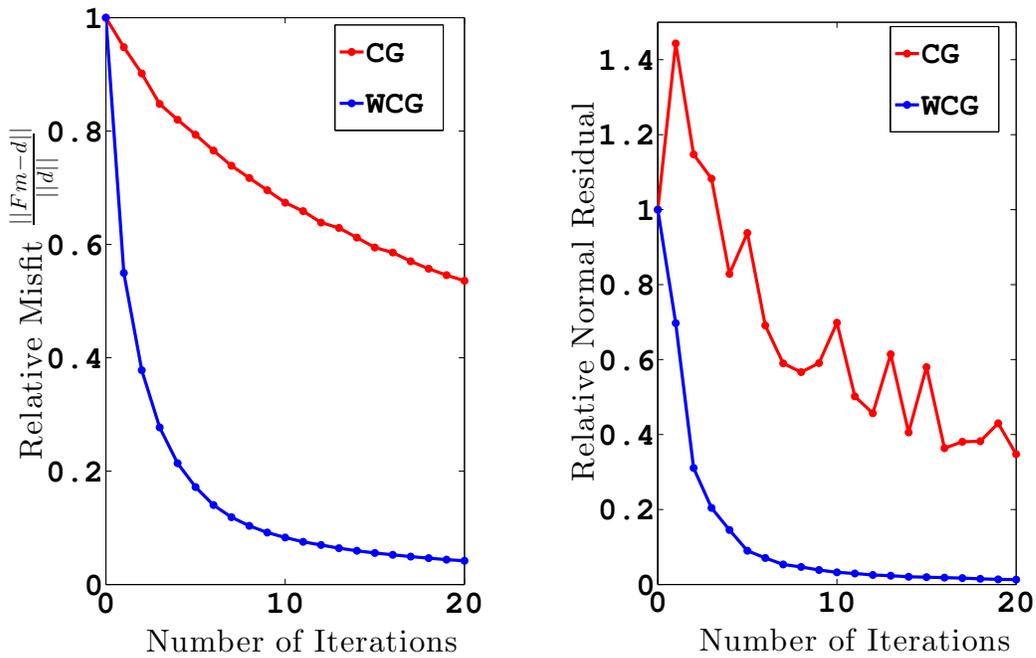
Figure 11: (a) Relative Misfit Plot (b) Normal Residual Plot.

sources: 69th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 133–135.

Berkhout, G., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, 924–938.

Bourgeois, A., B. Jiang, and P. Lailly, 1989, Linearized inversion: A significant step beyond pre-stack migration: Geophysical Journal International, **99**, 435–445.

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

Dutta, G., Y. Huang, W. Dai, X. Wang, and G. T. Schuster, 2014, Making the most out of the least (squares migration): 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 4405–4410.

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.

Ikelle, L., 2007, Coding and decoding: Seismic data modeling, acquisition and processing: 77th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 66–70.

Krebs, J. R., J. E. Anderson, D. Hinkley, R. Neelamani, S. Lee, A. Baumstein, and M.-D. Lacasse, 2009, Fast full-waveform seismic inversion using encoded sources: Geophysics, **74**, no. 6, WCC177–WCC188.

Kuehl, H., and M. Sacchi, 2003, Least-squares wave equation migration for AVP/AVA inversion: Geophysics, **68**, 262–271.

Liu, Y., W. Symes, and Z. Li, 2013, Multisource least-squares extended reverse time migration with preconditioning guided gradient method: 83rd Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 3709–3715.

Nemeth, T., C. Wu, and G. Schuster, 1999, Least-squares migration of incomplete reflection data: Geophysics, **64**, 208–221.

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

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

Romero, A., D. C. Ghiglia, C. C. Ober, and S. A. Morton, 2000, Phase encoding of shot records in prestack migration: Geophysics, **65**, 426–436.

Schuster, G. T., X. Wang, Y. Huang, W. Dai, and C. Boonyasiriwat, 2011, Theory of multisource crosstalk reduction by phase-encoded statics: Geophysical Journal International, **184**, 1289–1303.

Shin, C., S. Jang, and D.-J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.

Stolk, C. C., and M. V. De Hoop, 2005, Modeling of seismic data in the downward continuation approach: SIAM Journal on Applied Mathematics, **65**, 1388–1406.

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.

———, 2009, The seismic reflection inverse problem: Inverse Problems, **25**, 123008:1–24.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian: Geophysics, **74**, no. 6, WCA95–WCA107.

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

Xue, Z., Y. Chen, S. Fomel, and J. Sun, 2014, Imaging incomplete data and simultaneous-source data using least-squares reverse-time migration with shaping regularization: 84th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 3991–3996.

Born Waveform Inversion via Variable Projection and Shot Record Model Extension

Yin Huang and William W. Symes

ABSTRACT

Born waveform inversion is a partially linearized version of full waveform inversion based on Born (linearized) modeling, in which the earth model is separated into a smooth background model and a short scale reflectivity, and both are updated to fit observed trace data. Because kinematic variables (velocity) are updated, the possibility of cycle-skipping and consequent trapping at local minimizers exists for Born waveform inversion, just as it does for full waveform inversion. Extended Born waveform inversion allows reflectivity to depend on additional parameters, potentially minimizing the likelihood of cycle skipping by permitting data fit throughout the inversion process. Extended or not, the Born waveform inversion objective function is quadratic in the reflectivity, so that a nested optimization approach is available: minimize over reflectivity in an inner stage, then minimize the background-dependent result in a second, outer stage. This paper uses a 2D acoustic modeling, reflectivity permitted to depend on shot coordinates (shot record extension), a differential semblance penalty to control this dependence, and the variable projection variant of nested optimization. Our examples suggest that neither extended modeling nor variable projection alone are sufficient to enable convergence to a global best-fitting model, but the two together are quite effective.

A version of this paper was presented at the SEG 2015 Anuual International Meeting.

INTRODUCTION

Seismic full waveform inversion (FWI) is used to infer the interior structure of the earth from observed seismic waves by posing model-based data fitting as a

nonlinear least squares problem. Studied in the 1980's by Tarantola and others (Tarantola, 1984), it has recently become a viable model building strategy (Virieux and Operto, 2009). Because of the band-limited feature of seismic data, the FWI objective function may exhibit many local minima sharing few features with a best-fitting model (Gauthier et al., 1986).

Replacement of full waveform modeling by linearized or Born modeling in the formulation of FWI yields a partly linear least squares problem, in fact underlying much of seismic imaging theory and practice. The Born approximation is most accurate when the background is smooth on the wavelength scale, hence transparent, and the perturbation contains all short scale or oscillatory features of the earth model. Therefore we will adopt the convention that the linearization is based on such a long/short scale decomposition. The linearized forward model is linear, hence the mean square error objective function is quadratic, in the perturbation or reflectivity component of the Born model. The objective is still quite non-convex in the background model, hence in principle as likely to suffer from cycle-skipping and local minima as the FWI objective.

Recently, it has been suggested that a reduced or *variable projection* objective, obtained by minimizing over reflectivity for fixed background model, might be less likely to exhibit local minimizers (van Leeuwen and Mulder, 2009; Xu et al., 2012). Alternatively, since local minimizers appear to arise from cycle-skipping, a particular form of data misfit, extension of the reflectivity to depend on more parameters than just the spatial coordinates, might assist global convergence by permitting better data fit. Of course, any such dependence on extra parameters is non-physical, and must be controlled and ultimately removed from the model to give a satisfactory inversion. See Symes (2008) for an overview of extended modeling and inversion, and Shen and Symes (2008), Sun and Symes (2012), Biondi and Almomin (2014), Weibull and Arntsen (2014), and Lameloise et al. (2014) for more recent examples.

In this paper, we present examples which suggest that both extended modeling and variable projection (or, more generally, nested optimization) appear to be necessary ingredients in an effective Born waveform inversion algorithm. Kern and Symes (1994) showed how to combine variable projection and extended modeling. We use the same framework as Kern and Symes (1994), namely 2D constant density acoustic modeling and the shot-record model extension, to explore the importance of these two concepts. Shot-record extended models (in the Born inversion context) permit the reflectivity to depend on shot coordinates. Since single shot

data is easy to fit by Born modeling, with little constraint on background velocity, simply by appropriate choice of reflectivity, this extension allows good data fit throughout the inversion process and in particular reduces the likelihood of cycle skip. Also like Kern and Symes (1994), we use a *differential semblance* penalty to control the non-physical dependence of reflectivity on shot. This penalty is essentially the only choice for the shot-record extension leading to a smooth objective function, amenable to gradient-based optimization (Stolk and Symes, 2003).

Our examples are based on the Marmousi velocity model (Bourgeois et al., 1991). The target background model is created by smoothing the Marmousi velocity model. The target reflectivity model is the difference of the original model and a less stringent smoothing. We truncate and scale the model and data in two ways, creating two examples. The first example provides a relatively good starting model; nonetheless, its results suggest that without variable projection, neither Born waveform inversion nor its extended variant are likely to be successful, whereas variable projection added to either approach yields accurate inversion. The second example poses a more difficult problem, that of convergence from a constant background velocity. Only the variable projection extended approach appears to produce constructive velocity and reflectivity updates.

THEORY

Born modeling separates the model into a background m and a perturbation δm. For wave propagation, linearized modeling is generally most accurate when the background contains the long scales in the model and the perturbation the short (wavelength) scales. We shall call the perturbation the *reflectivity*, as the background is transparent hence δm is responsible for reflections. Denoting observed data by d and the linearized modeling operator by $F[m]$, the least squares objective of Born waveform inversion (BWI) is:

$$J_{\mathrm{BWI}}[m, \delta m] = \frac{1}{2}\|F[m]\delta m - d\|^2 \tag{1}$$

For each m, $F[m]$ is an operator on perturbational models $M = \{\delta m(\mathbf{x})\}$. We refer to this set as the the physical model space, to distinguish it from the extended model space $\bar{M} = \{\bar{m}(\mathbf{x}, \mathbf{h})\}$; in all cases, the physical model space appears somehow as a subspace of the extended model space. The variable $\mathbf{h}$ is a (scalar or vector)

parameter, such as shot coordinate, subsurface offset, or scattering angle, which characterizes additional degrees of freedom in the extended model space. The extended modeling operator is denoted $\bar{F}[m]$: note that background models are not extended. The extended Born waveform inversion (EBWI) problem is: given data d, find $m \in M$, $\delta\bar{m} \in \bar{M}$ that minimizes

$$J_{\text{EBWI}}[m, \delta\bar{m}] = \frac{1}{2}\|\bar{F}[m]\delta\bar{m} - d\|^2 + \frac{\alpha^2}{2}\|A\delta\bar{m}\|^2. \tag{2}$$

The second term in this sum involves an operator ("annihilator") A whose null space is precisely the physical models M. Minimizing it drives extended models toward physical (non-extended) models. For shot coordinate model extension, $\mathbf{h} = \mathbf{x}_s$ and $A = \nabla_{\mathbf{x}_s}$. Thus physical models are exactly those extended models that are constant in shot coordinate. The weight α in (2) controls emphasis on physicality: as $\alpha \to \infty$, the minimization of J_{EBWI} resembles more and more minimization of J_{BWI} (Gockenbach et al., 1995).

Variable Projection

The variable projection reduced objective for Born waveform inversion, $J_{\text{VP}}[m]$ is the least value attained by $J[m, \delta m]$ over the model space of reflectivity δm:

$$J_{\text{VP}}[m] = \min_{\delta m} J_{\text{BWI}}[m, \delta m]. \tag{3}$$

van Leeuwen and Mulder (2009) and Xu et al. (2012) have proposed closely related objective functions.

Similarly, define a variable projection objective function for extended Born modeling by

$$J_{\text{VPE}}[m] = \min_{\delta\bar{m}} J_{\text{EBWI}}[m, \delta\bar{m}]. \tag{4}$$

The analysis of VP objective function is similar, but simpler than VPE objective function. Thus we use $J_{\text{VPE}}[m]$ as an example. The value of J_{VPE} at a given background velocity m is the minimum value of J_{EBWI} over $\delta\bar{m}$ for fixed m, d. Since J_{EBWI} is quadratic in $\delta\bar{m}$, its minimum value $J_{\text{VPE}}[m]$ is attained at $\delta\bar{m}$ which solves the *normal equation*

$$(D\bar{F}[m]^T D\bar{F}[m] + \alpha^2 A^T A)\delta\bar{m} = D\bar{F}[m]^T \delta d. \tag{5}$$

Assume that the inner product in the background model space takes the form $\langle \bar{m}_1, \bar{m}_2 \rangle_{\bar{M}} = \langle \bar{m}_1, \Lambda \bar{m}_2 \rangle$, where $\langle \cdot, \cdot \rangle$ is the ordinary Euclidean inner product and Λ is a weight or roughening operator, chosen to enforce smoothness (slow variation) of the background models. In the examples below, Λ is a power of the Laplace operator. Assuming also that $\delta \bar{m}$ solves equation (5),

$$\nabla J_{\text{VPE}}[m] = \Lambda^{-1} D^2 \bar{F}[m]^T [\delta \bar{m}, D\bar{F}[m]\delta \bar{m} - d] \tag{6}$$

The transposed second derivative $D^2 \bar{F}[m]^T$ has been called the tomographic or WEMVA operator (Biondi and Sava, 2004; Biondi and Almomin, 2012). Note that both $D^2 \bar{F}[m]$ and $D^2 \bar{F}[m]^T$ depend on the background model parameter. The gradient of $J_{\text{VP}}[m]$ has the same form, but without model extension.

NUMERICAL EXAMPLES

We create a Born model $(m, \delta m)$ by modifying the 2D Marmousi model (Bourgeois et al., 1991). In this case, m is the velocity-squared field, and δm is its perturbation. We extended the water layer to 450 m depth, then smoothed the velocity-squared field with a moving box average to produce the background velocity model in Figure 1. The reflectivity (δm) is the result of subtracting from the Marmousi model a less aggressive smoothing. Both m and δm were resampled to the 16 m × 16 m simulation grid. Then Born data were computed by solving the perturbational wave equations using a centered finite difference scheme of order 2 in time and 4 in space. The source is a (finite difference version of) an isotropic point radiator with 6 Hz peak frequency Ricker wavelet. Trace data were synthesized for 110 shots starting from 1km with spacing 64m, depth 6 m. The receiver spread is symmetric about zero offset, with 481 receivers spaced 16m apart. Receiver depth is 10 m.

Two examples will be presented in this section. The first one compares inversion with and without variable projection. We give only the results for EBWI, as those for BWI are similar. In all cases, the parameter α in the definition (2) of J_{EBWI} and of J_{VPE} is set to 0.01, on the basis of trial-and-error. Our principal quality control display will be the stack of the extended reflectivity

$$\delta m_{\text{stack}}(\mathbf{x}) = \sum_h \delta \bar{m}(\mathbf{x}, h)$$

in comparison to the similar stack computed the target velocity-squared model: if events appear in the same positions with roughly the same amplitude, the inversion is successful.

In the first example, we use 2.6s data and truncate the model at 2km depth. The target (correct) background velocity-squared model for this example appears in Figure 1. The initial velocity-squared chosen for the first example (Figure 1, 2nd panel) is a linear combination: 70% of the target model and 30% homogeneous "water" ($(1.5 \text{ km/s})^2$). Because we have limited the depth range to 2 km and the maximum recording time to 2.6 s, the kinematic deviation from the target model is not great: this is a relatively easy velocity estimation problem. The difference is however great enough that the stack δm_{stack} incorrectly positions reflectors and faults (Figure 2, compare 1st and 2nd panels).

We approximate minimization of J_{VPE} by a very crude optimization algorithm: steepest descent with bisection backtracking line search. The gradient is computed by evaluating formula (6) using finite difference implementation of the adjoint state method, adapted to compute the tomographic operator $D^2\bar{F}[m]^T$ (Symes and Santosa, 1988). After 7 steps of this process, a lot of reflectors which we could not see at the initial model appear, and most reflectors positions are correct (Figure 2, 3rd panel).

We also attempt minimization of J_{EBWI} (equation (2)) simultaneously for m and δm, without reduction by variable projection. We use the Limited Memory BFGS algorithm (Nocedal and Wright, 1999), with the same backtracking line search used in the VPE minimization. We re-start the algorithm every 50 steps: each block of 50 steps has cost similar to that of one variable projection iteration, so 350 steps of LBFGS is roughly equivalent to 7 steps of variable projection optimization. The stacked reflectivity obtained at step 350 is displayed in Figure 1, 4th panel. The update of the velocity-squared shows no hint of has kinematic correction: all reflectors remain essentially in their initial, incorrect positions (Figure 2, 4th panel).

From this example, we conclude that minimization of J_{EBWI} for $m, \delta\bar{m}$ jointly is unlikely to succeed, whereas minimization of the variable projection function J_{VPE} produces a useful velocity update. A similar conclusion holds for J_{BWI} and J_{VP}.

The second example compares extended and non-extended variable projection, that is, minimization of J_{VPE} and J_{VP}. Since both appear successful for "easy"
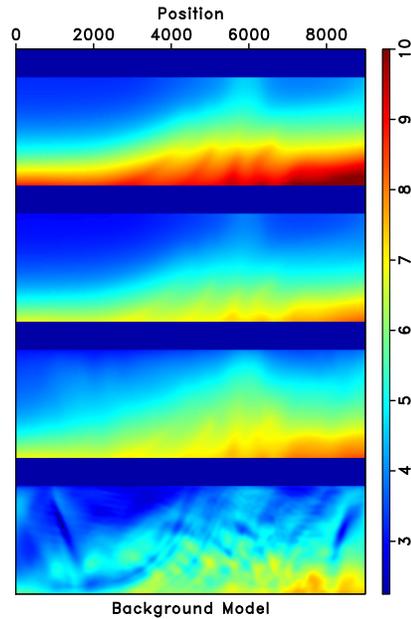
Figure 1: From top to bottom are true velocity-squared background model, initial model, estimate after 7 iterations of VPE, estimate after 350 iterations of EBWI.

velocity estimations, we create a more difficult problem by using more data (4 s), the full depth range (3.5 km) in the model, and choosing a more drastically incorrect initial model, namely a homogenous $v = 1.5$ km/s (Figure 3, top two panels). The stack δm_{stack} of the inverted extended reflectivity at the initial model is weak in amplitude with reflector positions are in error throughout (compare top two panels of Figure 4).

For both optimizations, we use steepest descent with backtracking line search as described above. The solution of equation (5) for both VP and VPE is estimated by the CG method with number of iterations increasing with background model update. This is based on our observation that when the background velocity is far from the true model, an inaccurate gradient estimate is still adequate to produce an acceptable update. In all cases, however, the residual error in normal equation (5) is reduced to less than 5 percent of its initial size.

After 13 steepest descent steps for J_{VPE} each involving 5 - 30 CG iterations per function evaluation and several backtrack steps, positions of events down to about
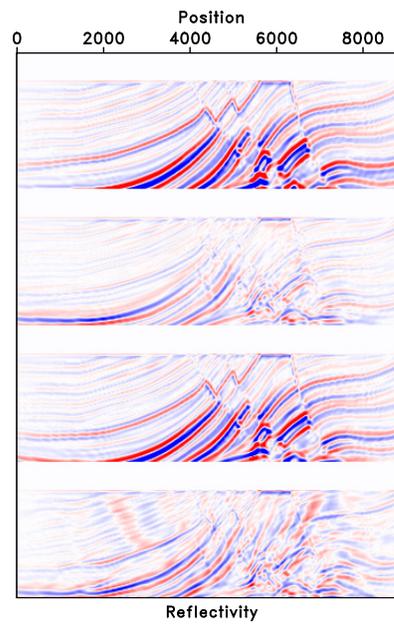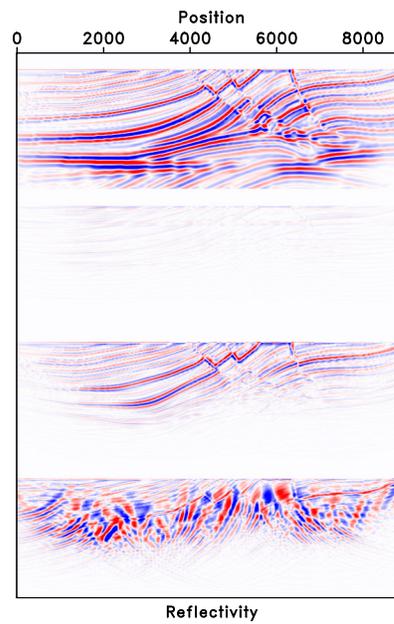
Figure 2: From top to bottom are inverted reflectivity at true background model, initial model, velocity model after 7 iterations of VPE, 350 iterations of EBWI, plotted on the same color scale. Note that weak amplitude indicates destructive interference due to inaccurate velocity.

1.5 km, somewhat deeper on the left, are largely corrected and amplitudes are considerably improved in the stacked inversion (Figure 4, 3rd panel). Although the inverted model at iteration 13 is not close to the true model at below 1.5km, the trend is promising (Figure 3, 3rd panel).

The first few VP updates, on the other hand, have partly repositioned the shallowest reflectors, subsequent steps do not improve the kinematics of the stack and do not correct deeper events that are well-positioned by VPE (Figures 3, 4, 4th panels).
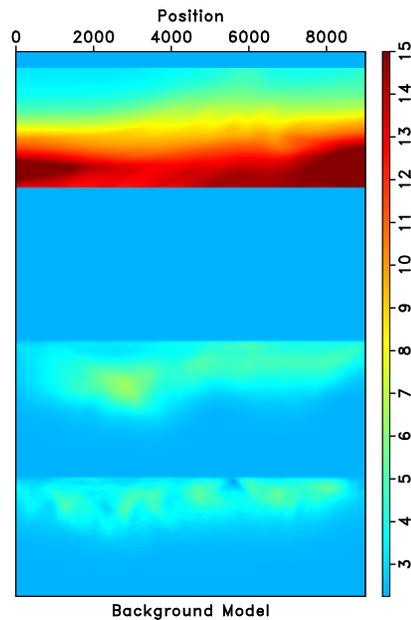
Figure 3: From top to bottom are true background model, initial model, velocity-squared model after 13 iterations of VPE and 13 iterations of VP.

DISCUSSION

The examples presented in the last section suggest that both extended modeling, which permits data fit throughout the inversion process, and variable projection, which enforces it, are critical ingredients in waveform inversion. Without both of these ingredients, gradient-based algorithms fail to constructively update kinematically inaccurate initial guesses. In particular, since the variable projection

Figure 4: From top to bottom are stacks of inverted reflectivity at the true background model, initial model, velocity model after 13 iterations of VPE and 13 iterations of VP, plotted on the same color scale, except the last which is scaled by 0.2 for display.

algorithms for Born and extended Born inversion involve similar numbers of migration/modeling pairs, the extended variant would appear to be the superior choice.

The Achilles' heel of this approach is its overall cost: the number of modeling/migration cycles required for the rather simple 2D examples presented here was in the hundreds, and such computational largesse is likely infeasible for industry-scale problems. Most of the cycles in these exercises go into the iterative solution of the normal equation (5), which must be fairly precise in order that the error in the gradient formula (6) be controlled. Convergence of CG and other iterative methods can be accelerated through the use of preconditioning, that is, inexpensive approximate inverse operators. Several methods for constructing preconditioners for (5) have been proposed (Tang, 2009; Stolk et al., 2009; Nammour and Symes, 2009), and should be evaluated for their potential to accelerate the VPE algorithm.

CONCLUSION

This paper has reviewed the variable projection extended inversion algorithm introduced by Kern and Symes (1994), and evaluated its performance on two 2D models.

We compared this method with iterative Born inversion, both with and without variable projection reduction, and with extended inversion without variable projection reduction. Our examples appear to indicate that both model extension and nested optimization are crucial attributes of a robust Born modeling based waveform inversion.

ACKNOWLEDGMENTS

providing HPC resources that have contributed to the research results reported here.

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.

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

Bourgeois, A., P. Lailly, and R. Versteeg, 1991, The Marmousi model, *in* The Marmousi Experience: IFP/Technip.

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

Gockenbach, M., W. Symes, and R. Tapia, 1995, The dual regularization approach to seismic velocity inversion: Inverse Problems, **11**, 501–531.

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, 2014, Improving the gradient of the image-domain objective function using quantitative migration for a more robust migration velocity analysis: Geophysical Prospecting, **62**, (in press).

Nammour, R., and W. Symes, 2009, Approximate constant density acoustic inverse scattering using dip-dependent scaling: 80th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, SI 3.6.

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

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

Stolk, C. C., M. V. de Hoop, and T. P. M. Op 't Root, 2009, Reverse time migration inversion from single-shot data: 79th Annual International Meeting, Expanded Abstracts, Society of Exploration Geophysicists, 2995–2999.

Stolk, C. C., and W. Symes, 2003, Smooth objective functionals for seismic velocity inversion: Inverse Problems, **19**, 73–89.

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

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.

Tang, Y., 2009, Target-oriented wave-equation least-squares migration/inversion with phase-encoded hessian: Geophysics, **74**, no. 6, WCA95–WCA107.

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

van Leeuwen, T., and W. A. Mulder, 2009, A variable projection method for waveform inversion: Presented at the 71st Conference EAGE, European Association for Geoscientists and Engineers. (Expanded Abstract).

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

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

Xu, S., D. Wang, F. Chen, Y. Zhang, and G. Lambare, 2012, Full waveform inversion for reflected seismic data: Presented at the 74th Conference EAGE, European Association for Geoscientists and Engineers, Expanded Abstracts.

Reducing the cost of extended waveform inversion by multiscale adaptive methods

Lei Fu and William W. Symes

ABSTRACT

Extended waveform inversion overcomes the "local minima" obstacle by adding an additional dimension of freedom to the model. However, one main challenge of this method is the computational intensity. In this abstract, we combine multiscale method with an adaptive approach to reduce the computational cost. In the multiscale strategy, the data and the source function are filtered by low-pass filters with low to high cutoff frequencies. Correspondingly, the space decomposition follows a coarse-to-fine scheme. Instead of using fixed subsurface offset range, in our adaptive approach, the adequate range is dynamically determined by measuring the data fitting at each background velocity step. Results from a synthetic example show a great improvement in computational efficiency while maintaining sufficient offset.

A version of this paper was presented at the SEG 2015 Anuual International Meeting.

INTRODUCTION

Least-squares full waveform inversion provides a way to determine the earth properties based on the comparison of the observed data and predicted data obtained from forwarding modeling (Tarantola, 1984). However, due to its highly nonlinear nature, the objective function of typical least-squares functions appears to possess many stationary points (local minima). Aiming to solve the local minima problem, the extended modeling concept links migration velocity analysis with full waveform inversion (Symes, 2008). The extension of the velocity model to subsurface offsets provides a robust solution to the local minima problem of conventional full waveform inversion (Sun and Symes, 2012; Biondi and

Almomin, 2012; Almomin et al., 2012). However, the extended waveform inversion is still a challenging data-fitting procedure. One main challenge is the computational intensity from the introduced additional dimension.

Based on the linearized model of acoustic scattering (Born approximation), the pressure field can be divided into two parts: the reference (incident) pressure field u and the scattered (perturbation) field δu. The reference pressure field u only contains direct waves and refracted waves, provided that the background velocity v is smooth or slowly varying, on the scale of wave length. The perturbed field depends linearly on the velocity perturbation δv, which is presumed to represent the oscillatory character of Earth structure within wavelength scale. In the constant density acoustic case, these two fields can be expressed as:

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

$$\left(\frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)\delta u(t,\mathbf{x};\mathbf{x_s}) = \frac{2\bar{r}}{v^3}\frac{\partial^2 u}{\partial t^2}(t,\mathbf{x};\mathbf{x_s}) \tag{2}$$

where $\bar{r} = \delta\bar{v}$ is the extended perturbed velocity model, and $v(\mathbf{x})$ denotes the acoustic background velocity. The reference field is represented by $u(t,\mathbf{x};\mathbf{x_s})$, a function of position $\mathbf{x}$ and time t. The source term is composed of wavelet time function $w(t)$ and a delta function $\delta(\mathbf{x}-\mathbf{x_s})$ centered at shot postion $\mathbf{x_s}$.

We can then express the solution of equation (2) as

$$\delta u(t,\mathbf{x_r};\mathbf{x_s}) = \frac{\partial^2}{\partial t^2}\int d\mathbf{x}\int d\mathbf{h}\int d\tau \frac{2r(\mathbf{x},\mathbf{h})}{v(\mathbf{x}+\mathbf{h})v^2(\mathbf{x}-\mathbf{h})} \\ G(\tau,\mathbf{x}+\mathbf{h};\mathbf{x_r})G(t-\tau,\mathbf{x}-\mathbf{h};\mathbf{x_s}) \tag{3}$$

where G is Green's function, and $\mathbf{h}$ is the subsurface offset. Here we write the action of the operator $\bar{r}$ as if it were an integral operator, with kernel also denoted $\bar{r}$. This additional integral in dimension of $\mathbf{h}$ increases the computational cost by a factor of N_h, number of grid points in h. Note that in 3D, one more space shift dimension is needed, making it even more expensive.

In order to make the extended waveform inversion feasible in real world applications, improvement of the computational efficiency is indispensable. The multigrid method (Brandt, 1977; Briggs et al., 2000) is a powerful technique to efficiently solve large-scale problems. One of the main advantages of this method is

reducing the computational cost by decomposing a problem by scale. See Bunks et al. (1995) for an early example of the multiscale method in least-squares inversion. By decomposing the problem by scale, the method greatly reduces the computational burden of the inversion, and improves the performance of iterative inversion.

THEORY

Seismic waveform inversion purports to obtain a velocity model, from which synthetic data is generated via a forward solver to best fit the observed seismic data d. The objective function J is a measure of the difference between synthetic data and observed data. The objective function of the linearized extended waveform inversion depends on the background velocity v and the extended velocity perturbation (extended reflectivity) $\bar{r}$, which can be written as:

$$min_{v,\bar{r}}J[v,\bar{r}] = \frac{1}{2}\|\bar{F}[v]\bar{r} - d\|^2 + \frac{\alpha^2}{2}\|A\bar{r}\|^2 \tag{4}$$

where $\|A\bar{r}\|^2$ acts as a regularizing term, weighted by α^2. Operator A is the regularization operator (annihilator), which is specific to differential semblance operator, multiplication by subsurface offset h in this work. Note the weighting parameter α controls the balance between data fitting and model extension: when $\alpha \to 0$, the model has little constraint on the energy distribution in the extended dimension h in order to achieve good data fitting; when $\alpha \to \infty$, the reflectivity model is forced to be physical ($\bar{r} = r$), i.e. the objective function becomes the ordinary non-extended one.

Instead of updating the extended reflectivity model and background velocity model alternately or simultaneously, this seismic inverse problem can be classified as a problem of separable least-squares and be solved with variable projection method (Golub and Pereyra, 1973). The inverse problem is solved by a nested optimization approach. In the inner loop, the objective function is optimized over extended reflectivity $\bar{r}$. The background velocity v is updated in the outer loop.

First, we calculate the gradient of the objective function $J[v,\bar{r}]$ with respective to extended model perturbation $\bar{r}$:

$$\nabla_{\bar{r}}J[v,\bar{r}] = \bar{F}[v]^*(\bar{F}[v]\bar{r} - d) + \alpha^2 A^*A\bar{r} \tag{5}$$

where $*$ denotes adjoint.

Setting the gradient function to zero results in the first order necessary condition for a stationary point of equation (4):

$$\bar{F}[v]^*\bar{F}[v]\bar{r} + \alpha^2 A^* A \bar{r} = \bar{F}[v]^* d \tag{6}$$

where $D\bar{F}[v]^*$ is a version of prestack depth migration and independent of $\bar{r}$. Defining the normal operator $\bar{N}[v] = \bar{F}[v]^*\bar{F}[v] + \alpha^2 A^* A$, equation (6) can be re-written as:

$$\bar{N}[v]\bar{r} = \bar{F}[v]^* d \tag{7}$$

This normal equation can be solved by a linear iterative method, e.g. conjugate gradient (CG) method. We name this as least-squares extended reverse time migration (LSERTM).

Lastly, the background velocity v is updated in the outer loop. The gradient of the reduced objective function $J[v, \bar{r}[v]]$ respect to v can be written as:

$$\nabla_v J[v, \bar{r}[v]] = \Lambda^{-2s} D\bar{F}[v]^T \left(\bar{r}[v], \bar{F}[v]\bar{r} - d \right) \tag{8}$$

where Λ^{-2s} is a smoothing operator for positive s (Symes and Kern, 1994).

As explained above, determining the subsurface offset range is 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 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. However, in the real world, a rough estimation of the velocity error may exist, in most situations, we don't know this error accurately enough to determine the necessary extend of space shift. We propose an adaptive method for determining the offset range by measuring the data fitting:

1. Apply a lowpass filter ($0 - f\ Hz$) on the data and source;

2. Run the least square extended migration;

3. Evaluate the data residual;

4. If data residual < X% and half offset residual < X%, decrease the offset range $h \rightarrow h/2$, return to step 3;

5. If data residual < X% and half offset residual > X%, update the background velocity. Then decrease the grid size by fact of 2 ($dh \rightarrow dh/2$), and increase the frequency band ($f \rightarrow 2f$), return to step 1;

6. If the data residual is more than X%, increase the offset range ($h \rightarrow 2h$), return to step 2;

Here, half offset residual refers to the data residual computed by using half of the subsurface offset range. Combined with multiscale approach, this adaptive method is able to assess the adequacy of the offset range very efficiently: early on, use low frequency, coarse grid, and coarse sampling in offset to update inaccurate model; as model becomes more accurate, increase frequency content, resample to finer grid, and keep computational cost down by reducing offset range. Steps 1-6 form a loop. The loop terminates when the offset range is adequate to resimulate the data and has been reduced as far as possible given the current velocity model. The value of X is determined by the amount of the data that can not be simulated by the current scale of the model (including the noise in the data).

RESULTS

In this section, we will use a single reflector model with constant background velocity to demonstrate the feasibility of the proposed multigrid approach. Note that we have done far more iterations than necessary but only because this is baseline experiment. Here we would like to emphasize that the data fit would not be able to decrease further even with more iterations, when the subsurface offset is inadequate.

The modeling and inversion are both constant-density acoustics based on the Born approximation. The 26 sources and 201 receivers ($0\ km$ to $2.0\ km$) are placed on the surface. Shown in Figure 1a, the true velocity model is consist of one horizontal high velocity perturbations sitting on constant background velocity v of $3.0\ km/s$. Based on Born approximation, the true velocity is split into the background velocity model $v(x,z)$ and the extended velocity perturbation model

$\delta \bar{v}(x, z, h)$. Note that the background velocity model $v(x, z)$ is nonextended, while the extended perturbation model $\delta \bar{v}(x, z, h)$ has nonzero value only at $h = 0\,m$.

The multigrid strategy is a coarse-to-fine space decomposition scheme. In this experiment, the space decomposition is implemented in three steps with grid size $40\,m$, $20\,m$, and $10\,m$. Correspondingly, the observed data and the source function are filtered by low-pass filters with low to high cutoff frequencies. The original source wavelet is a $3 - 24\,Hz$ bandpass wavelet (shown in Figure 1b). Figure **??**-**??** illustrate the source wavelet filtered by $0 - 8\,Hz$ and $0 - 16\,Hz$ bandpass filters, respectively.
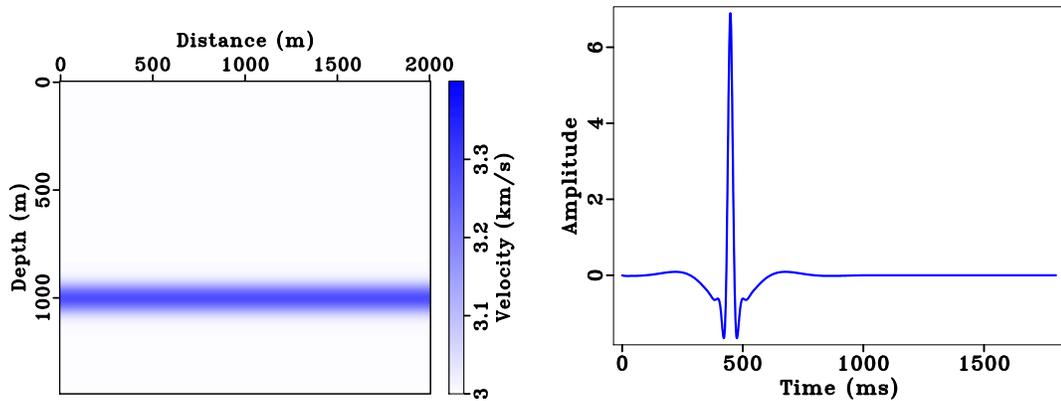
Figure 1: (a) Velocity model, (b) source wavelet (bandpass 3-24 Hz)

Step 1. We start with grid size $40\,m$, frequency band $0 - 8\,Hz$, and the time step interval $6\,ms$. Figure 2a shows the extended reverse time migration (ERTM) image for the $0 - 8\,Hz$ data with slower migration velocity ($2.4\,km/s$). The large data residual after 100 CG iterations indicates that the subsurface offset range is insufficient to resimulate the observed data (see Figure **??**). In the vertical slices at $x = 1000\,m$, the upward curve is truncated at the maximum offset $h = \pm 320\,m$. Figure 2b illustrates that the LSERTM image is distorted by this limited offset range. Because the data residual is greater than $X = 8\%$, we increase the offset range to $640\,m$. The data residual is less than 8%, which indicates the offset is long enough to predict the observed data. The ERTM image and LSERTM are shown in Figure 3a and 3b. Based on equation (8), the gradient of the objective function respect to v is shown in Figure 4b. Note that the theoretical computational cost here is only $(\frac{1}{4})^4 \approx 0.39\%$ of the original problem, in which space grid size $10\,m$ and the time
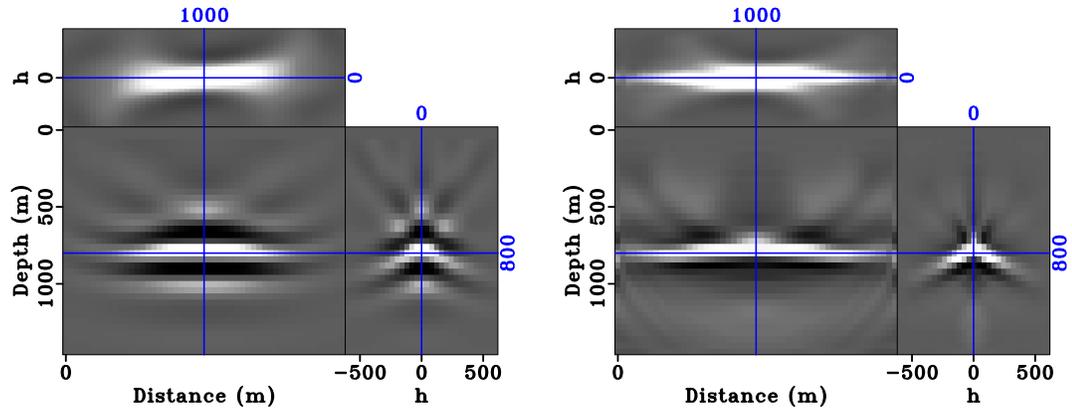
step interval 1.5 *ms* are used.
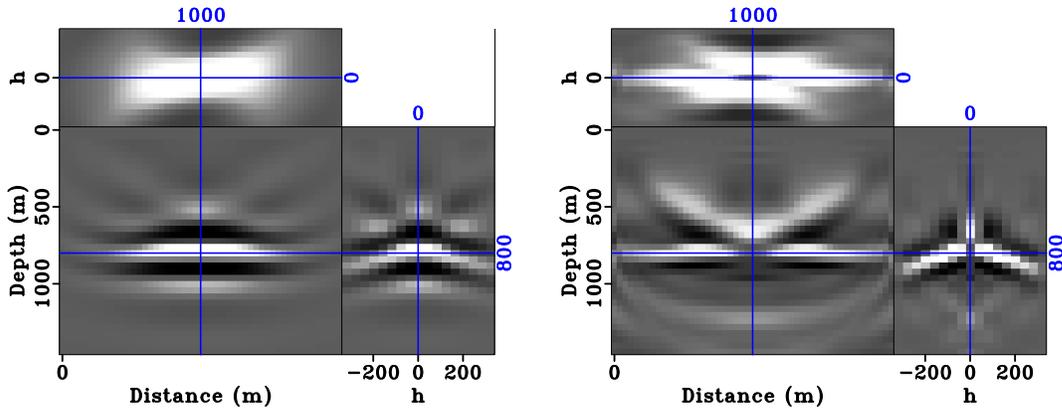
Figure 2: Subsurface offset range -320 m $\leq h \leq 320$ m, velocity $= 2.4$ km/s, 3-8 Hz data: (a) Extended RTM image, (b) Extended LSRTM model, 100 CG iterations

Step 2. Assume that we update the migration velocity correctly to $2.6\,km/s$. Now we filter the source wavelet and observed data by a frequency band $0-16\,Hz$, and decresase the subsurface offset range, the grid size and the time step interval by a factor of 2. The LSERTM image of 100 CG iterations is shown in Figure 5a. As the migration velocity is closer to the correct value, the energy should be more focused toward $h = 0m$, which suggests that shorter subsurface offset range is needed. As a result, an even half of the subsurface offset range $(-320m \leq h \leq 320m)$ is sufficient to predict the observed data, suggested by the data fitting (illustrated in Figure **??**). Note that $1/4$ of the subsurface offset range $(-160\,m \leq h \leq 160\,m)$ is insufficient.

Step 3. Assume that we update the migration velocity to $2.8\,km/s$. Now the original source wavelet and observed data are used. The subsurface offset range, grid size and the time step interval is further decreased by a factor of 2. The subsurface offset range of $-160\,m \leq h \leq 160\,m$ gives good data fitting.

CONCLUSIONS

The objective of this abstract is to address the central problem of extended waveform inversion, the computational cost, by applying the multiscale method and adaptively. In the multiscale approach, the data and the source function are fil-

Figure 3: Subsurface offset range -640 m $\leq h \leq 640$ m, velocity = 2.4 km/s, 3-8 Hz data: (a) Extended RTM image, (b) Extended LSRTM model, 100 CG iterations
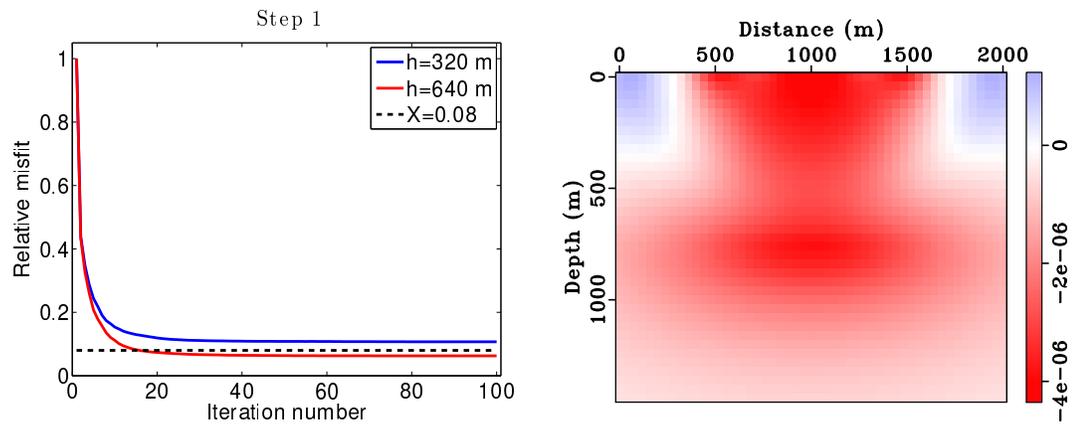
Figure 4: (a) Relative data residual as function if iteration number, various offset ranges; (b) Gradient of variable projection objective function with respect to background velocity
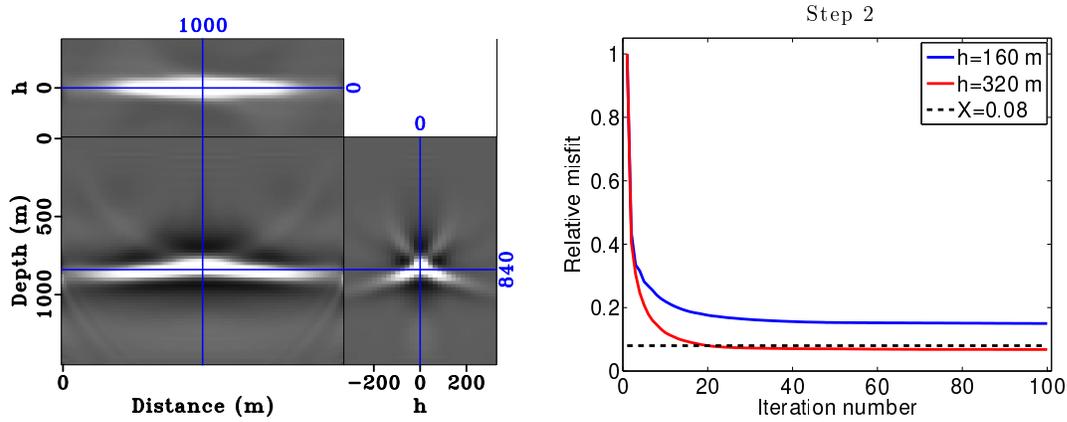
Figure 5: Step 2: velocity updated to 2.6 km/s, space grid size 20 m, subsurface offset range $-320\,m \leq h \leq 320\,m$, and the $3-16\,Hz$ data and source function (a) LSERTM result of 100 CG iterations, (b) the relative data residual as a function of iteration number with different subsurface offset range
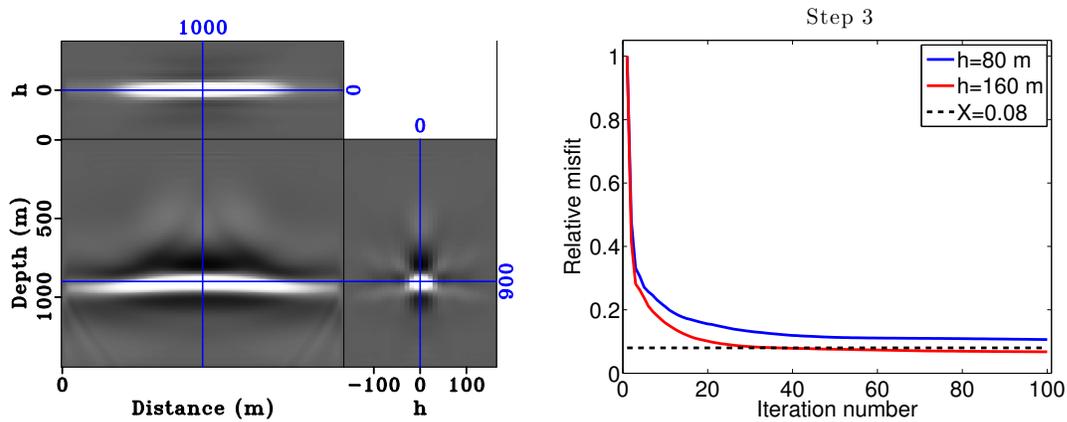
Figure 6: Step 3: velocity updated to 2.8 km/s, space grid size 10 m, subsurface offset range $-160\,m \leq h \leq 160\,m$, and the $3-24\,Hz$ data and source function (a) LSERTM result of 100 CG iterations, (b) the relative data residual as a function of iteration number with different subsurface offset range

tered by low-pass filters with low to high cutoff frequencies. Correspondingly, the space decomposition follows a coarse-to-fine scheme. We also shows an adaptive method for determining the subsurface offset range, in which the data fitting is used to assess the adequacy of the offset range. The numerical example demonstrate that the proposed methods can greatly reduce the computational cost and ensure adequate offset at the same time.

ACKNOWLEDGMENTS

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.

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.

Brandt, A., 1977, Multi-level adaptive solutions to boundary-value problems: Mathematics of computation, **31**, 333–390.

Briggs, W. L., S. F. McCormick, et al., 2000, A multigrid tutorial: Siam.

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

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.

Mulder, W., 2014, Subsurface offset behaviour in velocity analysis with extended reflectivity images: Geophysical Prospecting, **62**, 17–33.

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

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

Symes, W., and M. Kern, 1994, Inversion of reflection seismograms by differential semblance analysis: Algorithm structure and synthetic examples: Geophysical Prospecting, **in press**.

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

Accelerating Least Squares Migration with Weighted Conjugate Gradient Iteration

Jie Hou, William W. Symes

ABSTRACT

Iterative least squares migration can be effectively preconditioned by an adaptation of a preconditioner developed for subsurface offset extended least squares migration. The adaptation avoids interaction with the subsurface offset axis during forward and reverse time propagation, and is thus relatively economical of computer resources.

A version of this paper was submitted to the EAGE 2016 Annual Meeting.

INTRODUCTION

Least Squares Migration (LSM) iteratively achieves a best fit to seismic reflection data in least squares sense. It typically requires a kinematically correct velocity model and involves high computational cost to achieve a reasonable data fit. LSM with subsurface offset extension (ELSM) allows a good fit to the data even with significant velocity error, however, at higher computational expense per iteration. Hou and Symes (2015a) recently showed how to accelerate the convergence of ELSM with Weighted Conjugate Gradient (WCG) iteration, based on a high-frequency asymptotic approximate inverse to the extended Born modeling operator. However, the computationally expensive extension is in principle unnecessary with a correct velocity model, as the data can then be fit well with a conventional (non-extended) reflectivity model.

In this paper we present a modification of the WCG iteration of Hou and Symes (2015a) that accelerates the convergence of conventional (non-extended) iterative LSM without subsurface offset extended wave equation modeling and migration. This paper discusses and illustrates the 2D case; the 3D case is in principle similar.

THEORY

In this section, we will first review the concept of ELSM and WCG algorithm. We will then show how to apply the WCG algorithm to LSM without extended wave modeling.

Extended Born Modeling

The dynamical law of 2D constant density acoustics can be expressed as:

$$\frac{\partial^2 u}{\partial t^2} - m\nabla^2 u = f, \tag{1}$$

where $m = v^2$ is the squared acoustic velocity, $u = u(x, z, t; x_s)$ is the pressure field for source position x_s, and f models the seismic energy source at x_s. Linearization with respect to m yields

$$\frac{\partial^2 \delta u}{\partial t^2} - m\nabla^2 \delta u = \delta m \nabla^2 u \tag{2}$$

Here δm is the model perturbation (or reflectivity model), δu is the pressure perturbation. The Born modeling operator F maps δm to δu sampled at survey source and receiver positions. Given source/receiver data δd, LSM seeks a reflectivity model δm to minimize, $\|F\delta m - \delta d\|^2$, in which the vertical bars denote RMS (Euclidean length) over all axes. The subsurface offset extended model replaces δm with $\delta \bar{m}$ and the right-hand side of equation 2 with

$$\int dh \delta\bar{m}(x - h, z, h)\nabla^2 u(x - 2h, z, t; x_s) \tag{3}$$

The extended Born modeling operator $\bar{F}$ maps $\delta\bar{m}$ to δu sampled at the survey source and receiver positions; here h is essentially the (horizontal) offset between sunken source and receiver in Claerbout's survey-sinking imaging condition (Claerbout, 1985). The extension operator E produces an extended reflectivity from a non-extended or physical reflectivity δm by

$$\delta\bar{m}(x, z, h) = E\delta m(x, z, h) = \delta m(x, z)\delta(h). \tag{4}$$

The relation between extended and non-extended modeling operators is $F = \bar{F}E$.

WCG Algorithm for ELSM

The WCG algorithm is based on the the construction of a computable approximate inverse $\bar{F}^\dagger$ to the extended Born modeling operator (ten Kroode, 2012; Hou and Symes, 2015b):

$$\bar{F}^\dagger = W_{\text{model}}^{-1}\bar{F}^T W_{\text{data}}, \text{ where } W_{\text{model}}^{-1} = 4v_0^5 LP, \ W_{\text{data}} = I_t^4 D_{z_s} D_{z_r}. \tag{5}$$

$L = \sqrt{\nabla_{(x,z)}^2 \nabla_{(h,z)}^2}$ is similar to the Laplacian operator and can be implemented easily in wavenumber domain, I_t is time integration, $\bar{F}^T$ is the Euclidean adjoint of extend Born modeling operator (extended RTM) 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.

The WCG algorithm is the CG algorithm with weighted inner products:

$$\|\delta\bar{m}\|_{\text{model}}^2 = \delta\bar{m}^T W_{\text{model}}\delta\bar{m}, \ \|\delta d\|_{\text{data}}^2 = \delta d^T W_{\text{data}}\delta d \tag{6}$$

Equation 5 can be interpreted as asserting that the extended Born modeling operator is approximately unitary with respect to the weighted norms, hence the WCG algorithm can accelerate the convergence of ELSM dramatically.

WCG Algorithm for LSM

Using the relation between ordinary and extended Born modeling, $F = \bar{F}E$, the basic cost function of LSM with data weight W_{data} can be written as

$$J = \|W_{\text{data}}^{1/2} F\delta m - \delta d\|^2 = \|W_{\text{data}}^{1/2}\bar{F}E\delta m - \delta d\|^2. \tag{7}$$

LSM with WCG algorithm amounts solve the data-weighted normal equation,

$$F^T W_{\text{data}} F\delta m = F^T W_{\text{data}}\delta d. \tag{8}$$

To include the model weight factor, note that the extended image $\bar{F}^T W_{\text{data}}\delta d$ is focused at $h = 0$ when the velocity is kinematically consistent with $\delta d \approx F\delta m$, as is its image under the extended Hessian. For such focused image volumes EE^T acts as the identity, up to scale factor (the adjoint E^T extracts the $h = 0$ section). Thus

$$EF^T W_{\text{data}} F\delta m = EE^T\bar{F}^T W_{\text{data}}\bar{F}E\delta m \approx \bar{F}^T W_{\text{data}}\bar{F}E\delta m. \tag{9}$$

Apply the model weight operator on the left hand side of equation 9 and use equation 5 to obtain

$$E^T W_{\mathrm{model}}^{-1} E F^T W_{\mathrm{data}} F \delta m \approx \delta m \qquad (10)$$

That is, the Born modeling operator F is approximately unitary with respect to the norms defined by $W_{\mathrm{phys}} = E^T W_{\mathrm{model}} E$ and W_{data}, and these weights should play the same role in accelerating LSM as do W_{model}^{-1} and W_{data} in accelerating ELSM. Operationally, application of W_{phys}^{-1} amounts to padding its input with zeroes for $h \neq 0$, applying W_{model}^{-1}, then extracting the $h = 0$ section. Thus model extension is involved, but outside of the time loops that define F and F^T, and therefore at negligible cost.

SYNTHETIC EXAMPLES

Marmousi Model

We first apply LSM with both CG and modified WCG iteration on the Marmousi Model (Bourgeois et al., 1991). The background model, shown in Figure 1a, is computed by smoothing the original model and resampling to 20m grid. The reflectivity model, shown in Figure 1b, is the difference between smoothed background model and original model. 231 evenly spaced shots Born data are simulated with a (2.5-5-20-25) Hz band-pass wavelet and recorded by 461 evenly spaced receivers. Both the source and receiver are placed at 20m depth. Figure 1c,1d plot the reflectivity image after 20 CG and WCG iterations on the same grey scale. The WCG result clearly exhibits higher resolution. The misfit plot, displayed in Figure 2, shows the remarkable acceleration of WCG over CG. Actually, the result of 3 WCG iteration is already better than that of 20 CG iteration.

Overthrust Model

Background and reflectivity models derived from the SEG/EAGE overthrust model (Aminzadeh et al., 1997) are shown in Figure 3a and 3b, respectively. The model is discretized with 25m grid. The experiment simulates 201 evenly spaced shots and records at 401 evenly spaced receivers, both at 25m depth. A (2.5-5-20-25) Hz band-pass wavelet with a 2-ms time interval is used.

Reflectivity images (Figures 3c, 3d) and misfit plot (Figure 4) show clearly that WCG substantially improves convergence.

CONCLUSIONS

We have modified WCG algorithm for LSM. It shows a remarkable acceleration over CG iteration.

ACKNOWLEDGEMENTS

REFERENCES

Aminzadeh, F., Brac, J. and Kunz, T. [1997] *3-D Salt and Overthrust Models*. SEG, Tulsa.

Bourgeois, A., Lailly, P. and Versteeg, R. [1991] The Marmousi model. In: Versteeg, R. and Grau, G. (Eds.) *The Marmousi Experience*, IFP/Technip, The Hague.

Claerbout, J.F. [1985] *Imaging the Earth's Interior*. Blackwell Scientific Publishers, Oxford.

Hou, J. and Symes, W. [2015a] Accelerating extended least squares migration with weighted conjugate gradient iteration. In: *Expanded Abstracts*. Society of Exploration Geophysicists, 4243–4248.

Hou, J. and Symes, W. [2015b] An approximate inverse to the extended Born modeling operator. *Geophysics*, **80**, R331–R349.

ten Kroode, F. [2012] A wave-equation-based Kirchhoff operator. *Inverse Problems*, 115013:1–28.
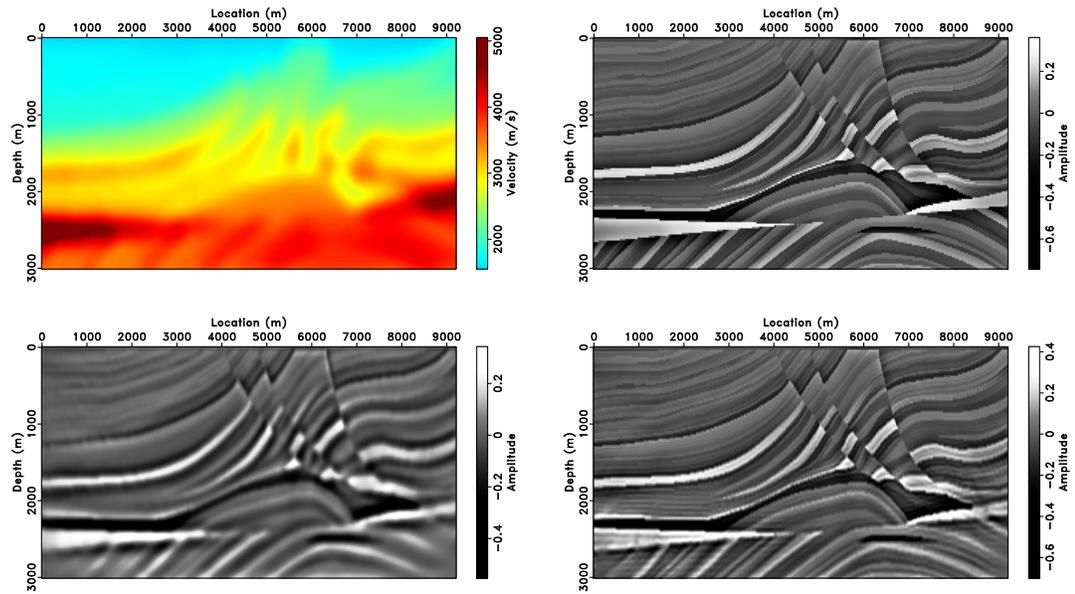
Figure 1: (a) Smoothed background velocity model for Marmousi example (b) Reflectivity model (c) LSM image from 20 iteration CG (d) LSM image from 20 iteration WCG.
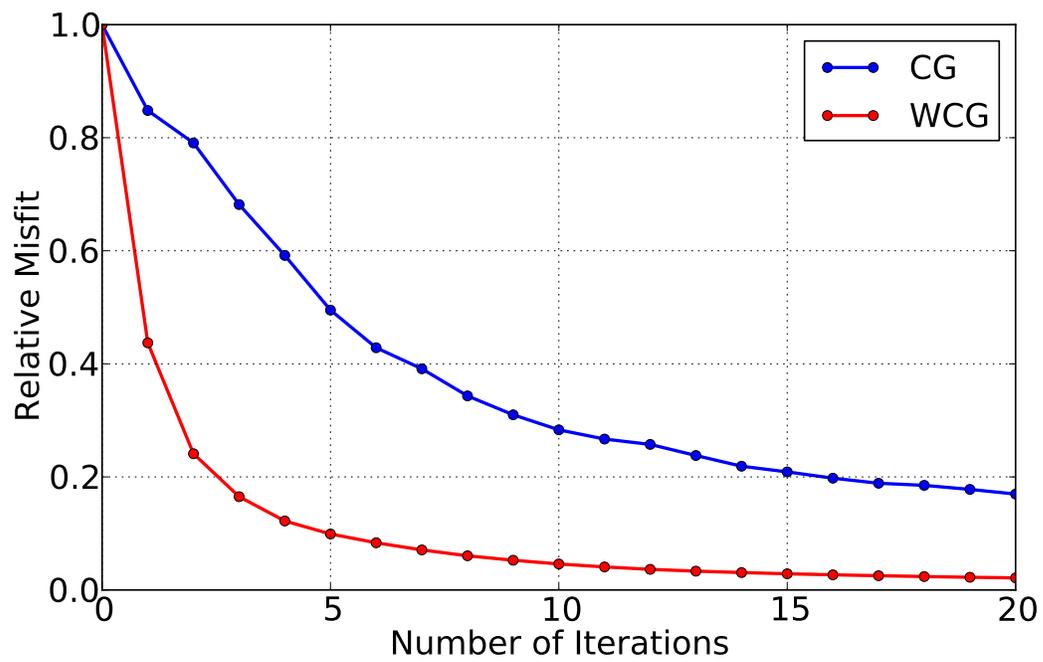
Figure 2: Relative data misfit plot for LSM of Marmousi example data with CG and WCG.
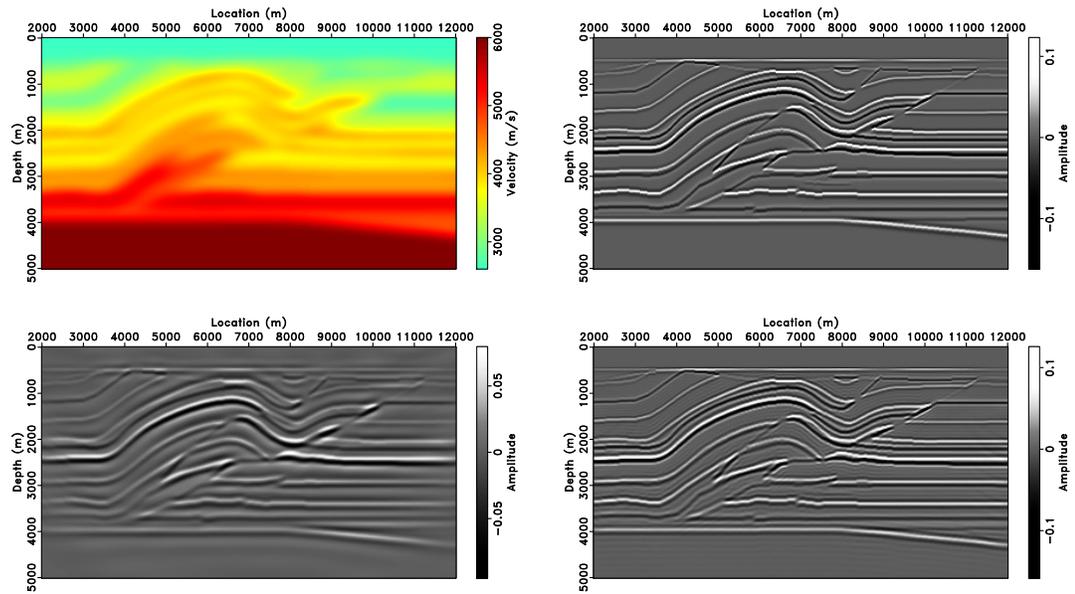
Figure 3: (a) Smoothed background velocity model for Overthrust model (b) Reflectivity model (c) LSM image from 20 iteration CG (d) LSM image from 20 iteration WCG.
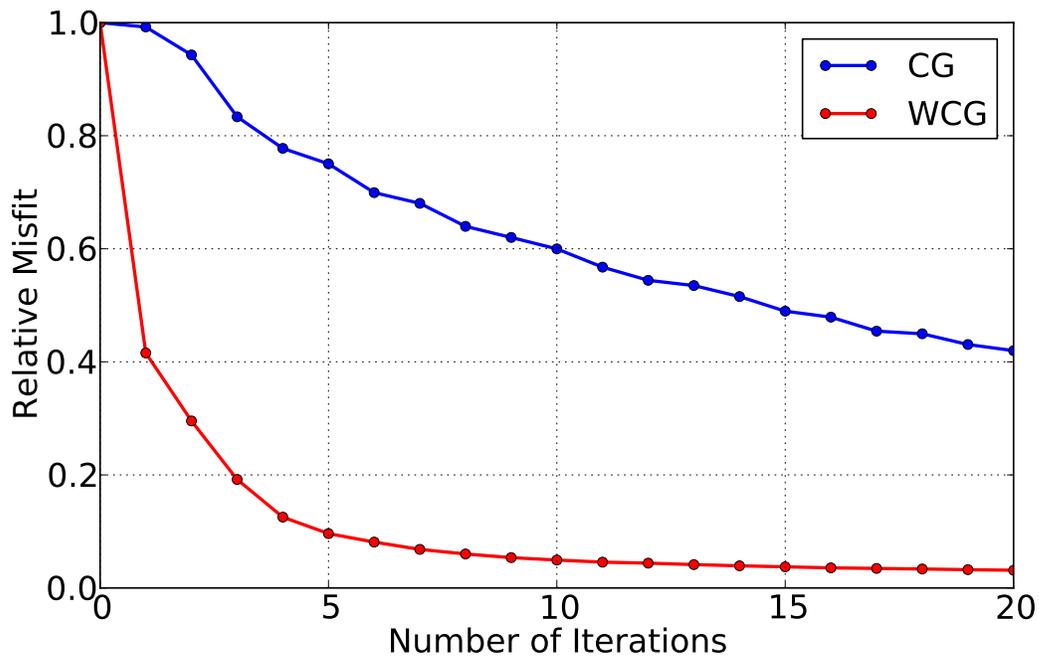
Figure 4: Relative data misfit plot for LSM of Overthrust example data with CG and WCG.

Dip-angle decomposition in relation with subsurface offset extended wave-equation migration

Raanan Dafni and William W. Symes

ABSTRACT

Our proposal provides post-migration techniques for computing angle-domain common-image gathers (CIGs) from seismic images, extended by the subsurface offset, in relation with wave-equation migration methods. In addition to the commonly used decomposition of the scattering-angles, we associate the wave-equation migration with dip-domain image gathers as well. Our methodology suggests a system of Radon transform operators by introducing local transform relations between the subsurface offset image and the angle-domain components. The same subsurface offset extended image is employed to decompose scattering and dip angle CIGs individually, or to decompose a multi-angle CIG by showing simultaneously both angles on the gather axis. 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 expected acceleration of wave-equation inversion methods.

Note: A version of this paper was ubmitted for presentation at the 2016 EAGE Annual Meeting. The full text of this paper is part of the 2015 TRIP annual report:

`http://www.trip.caam.rice.edu/reports/2015/trip2015_report.html`

Kinematic artifacts in the extended subsurface offset domain

Raanan Dafni and William W. Symes

ABSTRACT

We provide a kinematic analysis of the prestack image behaviour in the subsurface offset domain. When the medium properties are perfectly known, the image is expected to focus at the zero subsurface offset, where the incident and the scattering wavefields interact at a common point. However, kinematic artifacts are often observed in subsurface offset common-image gathers (CIGs) away from the zero offset trace, and artificially impairs the expected focusing. These artifacts emerge in relation with the acquisition geometry truncation at the boundaries of the seismic survey extent.

We suggest a formation mechanism for the artifacts emergence by considering seismic migration as a superposition of subsurface offset extended impulse responses, contributed by individual data traces. The accumulation of the image, in a trace-by-trace manner, gives an insight to its fundamental building blocks that better explains the formation of the kinematic artifacts.

We also discuss the defocusing of the subsurface offset image due to an erroneous migration velocity. In such case, the kinematic artifacts are formed by the same mechanism, while interfering with the essential defocusing information of the image away from the zero offset trace.

Geometry of Extended Reflection

William Symes[*]

ABSTRACT

Subsurface offset extended modeling implies relations between reflection properties (reflector dip, scattering angle) and ray geometry. The relation is similar to that of non-extended modeling, but differs in that a simple relation between scattering angle and image wavenumber holds only for zero offset.

INTRODUCTION

Sava and Fomel (2003) introduced the construction of angle domain extended images by Radon transform of subsurface offset extended images. Notably, the "scattering angle" was identified by Sava and Fomel (2003) as linked to the image wavenumber in offset.

Generally, the physical meaning of this "angle domain" has been misunderstood. For example, the "scattering angle" computed via Radon transform is the opening angle between incident and reflected rays only when these rays meet at the scattering point - that is, for zero subsurface offset. All of the energy is focused at zero offset when the velocity if correct (in the infinite frequency limit, of course), but not otherwise.

The ray geometry of the subsurface offset extension has been explored in a number of papers. I use here the account of Hou and Symes (2015), which completely and rigorously develops the geometry of extended reflection. Note that essentially the same relations appear in ten Kroode (2012), and possibly elsewhere; these relations are implicit in the asymptotic analysis of extended modeling and migration.

235

THEORY

In constant-density acoustic Born modeling extended by horizontal subsurface offset, the model perturbation δv is a function of horizontal spatial coordinates **x**, horizontal subsurface offset **h**, and depth z. For 2D modeling, **x** and **h** are 1D vectors, whereas for 3D modeling they are 2D: we will use the same notation in either case.

The dual wavevectors to these coordinates are $\mathbf{k}_x, \mathbf{k}_h,$, and k_z. Two distinguished wavevector subspaces play a key role, namely $\mathbf{k}_h = 0$ with coordinates $\mathbf{k}_x, k_z$ and $\mathbf{k}_x = 0$ with coordinates $\mathbf{k}_h, k_z$. We will write

$$k_{xz} = \sqrt{|\mathbf{k}_x|^2 + k_z^2}, \ k_{hz} = \sqrt{|\mathbf{k}_h|^2 + k_z^2}. \tag{1}$$

For 2D, the angle subtended with the distinguished axis, which for horizontal subsurface offset is of course the k_z axis, are denoted ν for the $\mathbf{k}_x, k_z$ subspace, and γ for the $\mathbf{k}_h, k_z$ subspace:

$$\cos \nu = \frac{k_z}{k_{xz}}, \ \cos \gamma = \frac{k_z}{k_{xz}}. \tag{2}$$

For 3D, these angles are denoted ν_1 and γ_1, and the corresponding azimuths are ν_2 and γ_2 respectively. Equations (2) are still correct, with ν replaced by ν_1 and γ by γ_1.

In the high-frequency asymptotics of reflection, the wavevector $(\mathbf{k}_x, \mathbf{k}_h, k_z)$ is an extended reflector normal. For physical reflectors, the reflectivity Fourier transform is independent of $\mathbf{k}_h$, and the subvector with coordinates $(\mathbf{k}_x, k_z)$ is thus the physical reflector normal. Accordingly, *nu* (2D) or ν_1 (3D) is the dip (polar) angle and ν_2 is the dip azimuth. The physical significance of the angle γ, or γ_1 and γ_2 in the 3D case, is not as obvious, as it is related to the ray geometry of reflection.

For the purposes of this paper, assume that all traveltimes between source or receiver and reflecting point are single-valued, that is, that no caustics occur in the scattering domain. For a particular pair $\mathbf{x}_r, \mathbf{x}_s$ of source and receiver coordinates (assuming that these lie on horizontal planes), the two-way traveltime T is defined in terms of the one-way traveltime τ by

$$T(\mathbf{x}, \mathbf{h}, z; \mathbf{x}_r, \mathbf{x}_s) = \tau(\mathbf{x} + \mathbf{h}, z, \mathbf{x}_r) + \tau(\mathbf{x} - \mathbf{h}, z, \mathbf{x}_s). \tag{3}$$

T is closely related to the phase function in the expression of the scattering normal operator - see Hou and Symes (2015), Appendix A, equation (A-1). Note that in this reference T is denoted ϕ, and the one-way times are denoted as T_r and T_s respectively. The computations leading up to equation (A-6) are essentially the same, with x replaced by $\mathbf{x}$ and k_x, k_h by $\mathbf{k}_x, \mathbf{k}_h$ and so on to accommodate the 3D case. Note that in 3D, two time integrations are required for each copy of $\bar{F}$ on the left-hand side of equation (A-1), in order that the resulting expression be equivalent to integration over a surface. The upshot is essentially the same stationary phase conditions as expressed in (A-6). In particular,

$$(\mathbf{k}_x, k_z) \quad \text{is parallel to} \quad \nabla_{\mathbf{x},z} T, \tag{4}$$

$$(\mathbf{k}_h, k_z) \quad \text{is parallel to} \quad \nabla_{\mathbf{h},z} T, \tag{5}$$

from which it follows that

$$\frac{k_z}{k_{xz}} = \frac{\nabla_z T}{|\nabla_{\mathbf{x},z} T|}, \tag{6}$$

$$\frac{k_z}{k_{hz}} = \frac{\nabla_z T}{|\nabla_{\mathbf{h},z} T|}, \tag{7}$$

and

$$\tan \nu = \frac{|\nabla_{\mathbf{x}} T|}{\nabla_z T}, \tag{8}$$

$$\tan \gamma = \frac{|\nabla_{\mathbf{h}} T|}{\nabla_z T}. \tag{9}$$

The $\mathbf{x}, z$ gradient of T is the sum of the $\mathbf{x}, z$ gradients of τ for the arguments indicated in equation (3). These gradients are in turn the ray velocity vectors from source and receiver. For $\mathbf{h} = 0$, the eikonal equation states that these gradients have the same length, so that their sum is their bisector. Thus in this case equation (4) states Snell's law of reflection, and the bisector of the angle subtended by incident and reflected rays is the dip (reflector normal) vector. For $\mathbf{h} \neq 0$, the values of velocity at the points of evaluation $(\mathbf{x} \pm \mathbf{h}, z)$ are not necessarily the same, so that while the reflector normal is still parallel to the isochron normal, it is not necessarily the bisector of the incident and reflected ray pair.

The scattering (or opening) (half-) angle γ_{scat} is one-half the angle subtended by the incident and reflected rays. Denoting $s = v^{-1}(\mathbf{x}, z), s_\pm = v^{-1}(\mathbf{x} \pm \mathbf{h}, z)$, and using

the eikonal equation several times,

$$|\nabla_{\mathbf{x},z} T(\mathbf{x}, \mathbf{h}, z; \mathbf{x}_r, \mathbf{x}_s)|^2 = s_+^2 + s_-^2 + 2s_+ s_- \cos 2\gamma_{\text{scat}}$$

$$= 4s_+ s_- \left(\frac{s_+}{2s_-} + \frac{s_-}{2s_+} - 1 + \cos^2 \gamma_{\text{scat}} \right). \tag{10}$$

This equation is difficult to interpret in general, however in the special case $\mathbf{h} = 0$ it takes on a familiar meaning:

$$|\nabla_{\mathbf{x},z} T(\mathbf{x}, 0, z; \mathbf{x}_r, \mathbf{x}_s)|^2 = 4s^2 \cos^2 \gamma_{\text{scat}} \tag{11}$$

On the other hand,

$$|\nabla_{\mathbf{x},z} T|^2 = |\nabla_{\mathbf{x}} T|^2 + |\nabla_z T|^2 = \sec^2 v |\nabla_z T|^2. \tag{12}$$

The computation of $|\nabla_z T|^2$ is accomplished in equations (A-21) through (A-26) and accompanying discussion in Appendix A of Hou and Symes (2015), and needs no change other than the lexicographic ones already noted to accommodate 3D: the starting point is the stationary phase (A-6), or its rewrite (4), (5) above. The net result is (for general $\mathbf{h}$, rewritten from equations (A-24), (A-25) and (A-26))

$$|\nabla_z T|^2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \tag{13}$$

in which

$$a = \frac{k_{xz}^2 k_{hz}^2}{k_z^4}, \; b = -2 \left[(s_+^2 = s_-^2) \frac{\mathbf{k}_x \cdot \mathbf{k}_h}{k_z^2} + (s_+^2 + s_-^2) \right], \; c = (s_+^2 - s_-^2)^2. \tag{14}$$

For $\mathbf{h} = 0$, these expressions collapse to

$$|\nabla_z T|^2 = 4s^2 \frac{k_z^4}{k_{xz}^2 k_{hz}^2} \tag{15}$$

so from (11), (12, and (2)

$$4s^2 \cos^2 \gamma_{\text{scat}} = 4s^2 \sec^2 v \frac{k_z^4}{k_{xz}^2 k_{hz}^2}$$

$$= 4s^2 \cos^2 \gamma \tag{16}$$

so the angles γ and γ_{scat} are identical for $\mathbf{h} = 0$. Inspection of (13), (14) suggests that otherwise, when $s_+ \neq s_-$, these two angles are not necessarily equal.

For the 2D case with $\mathbf{h} = 0$, reference to (2), (4), (5) and (15) yield identities previously stated by Sava and Fomel (2003):

$$
\begin{aligned}
\frac{\partial T}{\partial x} &= 2s \sin \nu \cos \gamma, \\
\frac{\partial T}{\partial h} &= 2s \cos \nu \sin \gamma, \\
\frac{\partial T}{\partial z} &= 2s \cos \nu \cos \gamma,
\end{aligned}
\tag{17}
$$

and

$$
\begin{aligned}
-\frac{\partial z}{\partial x} = \tan \nu &= -\frac{k_x}{k_z} \\
--\frac{\partial z}{\partial h} = \tan \gamma &= -\frac{k_h}{k_z}
\end{aligned}
\tag{18}
$$

which are simply the equations (8) and (9) with sign normalization. On the left-hand side, z is regarded implicitly as a function of x, h describing the plane tangent to the extended reflector at the reflection point. Note that the identities (17) are correct in general *only* for $\mathbf{h} = 0$, whereas from the point of view developed in this paper, equations (18) are valid in general.

For completeness, we state the 3D generalizations of these relations, in coordinates, using h_x and h_y for the $x-$ and $y-$ components of subsurface offset:

$$
\begin{aligned}
\frac{\partial T}{\partial x} &= 2s \sin \nu_1 cos\nu_2 \cos \gamma_1, \\
\frac{\partial T}{\partial y} &= 2s \sin \nu_1 sin\nu_2 \cos \gamma_1, \\
\frac{\partial T}{\partial h_x} &= 2s \cos \nu_1 \sin \gamma_2 \cos \gamma_2, \\
\frac{\partial T}{\partial h_y} &= 2s \cos \nu_1 \sin \gamma_1 \sin \gamma_2, \\
\frac{\partial T}{\partial z} &= 2s \cos \nu_1 \cos \gamma_1,
\end{aligned}
\tag{19}
$$

and

$$-\frac{\partial z}{\partial x} = \tan \nu_1 \cos \nu_2 \quad = \quad -\frac{k_x}{k_z},$$

$$-\frac{\partial z}{\partial y} = \tan \nu_1 \sin \nu_2 \quad = \quad -\frac{k_y}{k_z},$$

$$-\frac{\partial z}{\partial h_x} = \tan \gamma_1 \cos \gamma_2 \quad = \quad -\frac{k_{h_x}}{k_z},$$

$$-\frac{\partial z}{\partial h_y} = \tan \gamma_1 \sin \gamma_2 \quad = \quad -\frac{k_{h_y}}{k_z}. \tag{20}$$

Again, equations (19) relate rays to reflectors and hold only for $h = 0$, whereas equations (20) simply identify the angles in relation to the reflector wavenumbers and hold in general.

CONCLUSION

The geometrical reasoning developed by Hou and Symes (2015) encapsulates the relations between ray angles and reflector normals. These relations take the usual form proposed by Sava and Fomel (2003) only for physical reflection, that is, zero subsurface offset.

REFERENCES

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

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

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

Parallel IWAVE

William W. Symes, The Rice Inversion Project

ABSTRACT

IWAVE supports loop-level, block-level, and task-level parallelism. OpenMP parallelization of loops is available, as is domain decomposition via MPI. MPI also enables parallelism over shot records. This brief report desribes the mechanics of IWAVE parallel execution modes, and includes examples of typical use cases.

INTRODUCTION

IWAVE was designed from the outset for parallel execution. The first release (version 1.0, fall 2009) used domain decomposition via MPI, and optionally loop distribution via OpenMP, for simulation (forward map) only. Later releases introduced task-level parallelism, that is, concurrent execution of shots optionally in combination with domain decomposition, and extended all of these modes of parallelism to derivative and adjoint derivative computations.

This paper explains how to invoke IWAVE in various modes of parallel execution. I apply a selection of these modes to a small 2D example, and so illustrate that the same results are obtained for serial execution as for any mix of parallel options. I will discuss parallelization over shots and domain decomposition here: loop distribution via OpenMP will be discussed elsewhere. The SConstruct file in the project subdirectory contains examples of all of the constructions mentioned here, and should be regarded as part of the paper.

THE IWAVE APPROACH TO PARALLELISM

Domain decomposition is *computed* from minimal user input, namely the number of subdomains along each grid axis. The global grid is partitioned into subgrids

of equal, or near-equal, size, and each subgrid is assigned to an MPI process. The subgrids contain ghost points required to execute the scheme stencil. Exchange of data to re-initialize ghost points uses the `MPI_Vectortype` structure to realize virtual subgrids, thus avoiding redundant memory allocation and data motion. OpenMP-based threaded execution co-exists with domain decomposition. See Terentyev (2009) for details on the initial design of IWAVE, Fehler and Keliher (2011) for information on its use in the QC of the SEAM Phase I project. and Symes et al. (2011) for a description of a later release.

IWAVE uses minimal user data, namely the number of shots to execute in parallel, to generate a vector of communicators, one per parallel shot. Each communicator is assigned a subset of shots, the subsets being chosen as close to equal size as possible, and a subset of processes that participate in computing each shot assigned to that communicator - one for pure task-level parallelism, more than one if domain decomposition is also being used.

The same principles apply to the forward map and to its derivatives and adjoint derivatives: all are parallelized in the same ways.

IWAVE currently implements disk-to-disk operations. Common input data (eg. model fields) is read and broadcast, shot-dependent data (such as shot gathers) are read or written as needed or generated. The current i/o design is very MPI 1, that is, all i/o takes place on rank 0 of each communicator (note that there may possibly be multiple communicators active in an IWAVE run). Reduction data (non-extended adjoint output) is presumed to be summed (stacked). In task-parallel domain decomposition, each communicator assembles a partial stack over the shots assigned to it, then these results are summed in a final reduction phase.

BUILDING PARALLEL IWAVE

IWAVE can be built MPI-enabled, but not as part of the overall Madagascar build. To enable MPI execution, run SConstruct in `RSFSRC/trip`, with a `config.py` file specifying MPI compilation and linking. You will need to put the path to the MPI root directory in your environment as `MPIROOT`, or use one of the module loading systems common at supercomputer sites to provide equivalent information to your shell, then follow the model (or use) one of the config files in `RSFSRC/trip/admin`, for example `RSFSRC/trip/admin/linux.mpi.py` - to use, simply

copy to `RSFSRC/trip/config.py`, then `scons`.

This process creates the MPI-enabled executables in the `main` subdirectories, not in the `RSFROOT/bin` directory where other Madagascar commands are found - for example, in this exercise, `RSFSRC/trip/iwave/acd/main/acd.x`. You will need to instrument your `SConstruct` files to find these commands, since the default paths from Madagascar will not work. See `project/SConstruct` for an example.

NOTE: at this writing (2015.09.10) the TRIP directory tree head revision does NOT reside in `RSFSRC` - it is independent of Madagascar, for purely accidental reasons flowing from a catastrophic failure at SourceForge earlier in summer 2015, and lives in the CAAM TRIP repository as `trip2.1`. I anticipate that sometime soon the TRIP tree will migrate back to the new `RSFSRC` GitHub repository. For the time being, where you see `RSFSRC/trip` in this discussion, substitute `trip2.1`. Do NOT use the version of the TRIP tree found in `RSFSRC`, until further notice, as it is increasingly out-of-date.

JOB CONTROL

The TRIP software stack includes a Python module (`RSFSRC/trip/admin/new-batch.py`) defining functions that read standardized job information for serial, command line parallel, or batch execution of IWAVE commands and builds appropriate Madagascar `Flow`s for each case. These data structures specify each essential item of information necessary for job definition with minimal dependencies on other items.

An `SConstruct` file using the IWAVE batch module should include standard boilerplate at the top:

```
from rsf.proj import *
from newbatch import tripExec
```

The `SConstruct` should also define

- an array of *job dictionaries*;

- one or more *parallel environment* dictionaries;

- one or more *batch environment* dictionaries.

and include a call to `tripExec` following these definitions. This section describes these data structures and their relations and uses.

The central data structure is the *job*, a Python dictionary with these fields:

- `job`: job name (string),

- `pre`: preliminary command, to be executed in serial mode. This is mainly provided to include creation of output data files by copy or by another command (from Madagascar or SU, for example) - these need to exist, with correct metadata, before acting as target in an IWAVE command, however they cannot be built in a separate `Flow` since no object can serve as target in two different `Flow`s. Generally these are (low-intensity floating point) commands which one wishes to execute in serial, hence defined separately from the main command below, which may be executed in parallel.

- `src`: list of source files, including all those used in as sources in the IWAVE command and in the preliminary command.

- `tgt`: list of target files.

- `cmd`: main command, to be executed as serial command, under mpirun, or as part of a batch script.

- `exe`: execution environment dictionary - defines type of parallel execution (serial, command-line mpi, or batch) with necessary parameters for each.

Since each project typically specifies several jobs, the project `SConstruct` file should organize its job dictionaries into an array, eg.

```
jobs = [ {job1: {...}}, {job2: {...}},...]
```

Three types of parallel execution environment are recognized, corresponding to three dictionary structures:

- serial: an empty dictionary

- command line mpi: a dictionary with two items,

    - `platf: mpi` - specifies command line mpi
    - `ppn: ...` - number of MPI processes

- batch:

    - `platf:` batch platform name - serves as index into batch environment dictionary, see below
    - `nodes:` number of nodes
    - `ppn:` processes/cores per node
    - `wall:` wallclock time limit, in form xx:xx:xx

Batch execution requires some additional job-independent characteristics of each environment, such as scripting language, launcher name or path, project name and other accounting information, etc. These need to be listed in a *batch environment* dictionary. The information provided in the batch environment dictionary is peculiar to each site and/or machine operated by the site, and may be expected to change as supercomputing sites evolve their software stacks and environments. The dictionary has a standard structure: it uses platform names (from the batch instances in the parallel environment dictionary) as keys; the corresponding values are themselves dictionaries, defining values for the standard keys:

- `batch:` batch system name, eg. pbs, sge, slurm,...

- `queue:` name of execution queue.

- `acct:` account name, to which jobs are to be charged.

- `mail:` email address to which notification of job begin and end should be sent

- `launcher:` name of MPI launcher used, eg. mpiexec, ibrun,...

Having defined the array of job dictionaries, the parallel environment dictionary (or dictionaries, as different jobs can use different parallel environments), and the batch environment dictionary (which may be empty, if your project has no need of batch job submission), include the line

```
tripExec(<name of jobs array>, <name of batch env dict>)
```

in the project SConstruct somewhere below the definitions. The tripExec function parses the information contained in its arguments into Madagascar Flows that execute the various jobs with the chosen modes of parallelism. Each Flow executes in a subdirectory of the project directory, named jobname.work, in which jobname is the job name assigned in the job dictionary (containing the item 'job': jobname). This subdirectory captures the diagnostic output of IWAVE and MPI (for instance, the cout....txt files generated by IWAVE, one for each MPI process). The subdirectory is a Flow target, so scons -c in the project directory gets rid of all of these execution directories and their contents.

Notes:

1. The Flows generated by tripExec do not do standard i/o, so this tool is really suitable only for applications like IWAVE. For SU or Madagascar commands structured as filters (i.e. < inp cmd >outp, write ordinary Madagascar Flows.

2. The implementation of tripExec uses sfbatch, the Madagascar batch utility, enabling use of the Madagascar SOURCES and TARGETS macros in the command definition, and keeping each data file name to a unique location (in the src or tgt list). sfbatch incorporates current standard choices for SLURM and PBS at many sites, but given the variety and continued evolution of supercomputer batch environments it's natural to suspect that it may need updating - so beware.

3. The various modes of execution may be mixed in a Madagascar project SConstruct. In particular, dependence of results on other results produced by batch commands is respected: a Flow dependent on the result of a batch computation executes when the batch computation is complete, rather than when the batch submission command is executed.

4. For all but serial execution models, the number of processes, and for domain decomposition their geometric layout, must be described. This information can be captured in various ways, but probably the easiest is via a set of Python integer variables, as in the example described in the next section. For example, a script fragment describing a 2D 2×3 domain decomposition of 6 shots, computed 3 at a time, would include in the definition of the IWAVE driver command something like

```
NP1=2
NP2=3
NPT=3
...
cmd='... mpi_np1=' + str(NP1) + ' mpi_np2=' + str(NP2) +
    ' partask=' + str(NPT)
```

For command-line MPI execution, one would naturally use exactly the required number of threads, so

```
jobs=[..., {...,'exe': {'platf': 'mpi', 'ppn': str(NP1*NP2*NPT)} },...]
```

whereas queue management algorithms for batch sites argue for an independent definition of the number of threads, like

```
jobs=[..., {...,'exe': {'platf': 'euclid',
                        'nodes': str(NODES),
                        'ppn': str(PPN),
                        'wall': '04:00:00'} },...]
```

The IWAVE MPI environment function checks that sufficient resources are available: thus IWAVE will abort if `NP1*NP2*NPT > NODES*PPN`.

5. **WARNING:** In many cluster environments, partial use of a node (using fewer cores than available on the number of nodes allocated) can result in jobs hanging. Therefore, until we see some portable and transparent way around this limitation, the number of tasks defined by IWAVE (`NP1*NP2*NPT` in the preceding example) should be the same as the number of cores allocated (`NODES*PPN` in the example).

EXAMPLES

I have used the simple 4-layer OBC example from (Symes, 2014) to illustrate the various parallel IWAVE modes for serial and command line mpi execution. The command used in these examples is `iwave/acd/main/acd.x`, which implements

the forward map, its first and second derivatives, and their adjoints. This application is described in some detail in Symes (2014).

The paper (Symes, 2015) describes a similar example. The reader should examing the `project/SConstruct` files for both papers to see "live" examples of the framework. In particular, the examples attached to this paper are configured for command-line MPI, rather than batch. To see an explicit example of the mechanics of batch submission, see the examples attached to the companion paper (Symes, 2015).

The velocity model for the 4-layer OBS example appears as Figure 1. Other details of the simulation are as described in Symes (2014), and in the figure captions. Figures 2a, 2b, 2c, and 2d show the results of modeling an OBC gather (effectively, a shot on the seafloor and near-surface receivers) with no domain decomposition (or 1×1), versus 1×2, 2×1, and 2×2 decompositions, all plotted on the same scale. The differences are at round-off level. The parameters `mpi_np1` and `mpi_np2` describe the number of domains along the first and second axes, and were set at 1 and 2 respectively. Thus, 1, 2, 2, and 4 MPI processes were involved in these simulations respectively. It is an error to provide too few processes to assign each domain a process. However providing more processes is not an error; the unused processes are simply idled.

Figures 5a, 5b, 5c, and 5d describe the analogous results for the Born approximation. Once again, the differences are at round-off level. Finally, Figures 6a, 6b, 6c, and 6d describe the reverse time migration (adjoint linearized modeling) of the corresponding Born data.

All of these results should be compared the analogous plots in Symes (2014).

Finally, to illustrate shot parallelization and its combination with domain decomposition, I simulated 9 shots, with 3 computed concurrently (`partask=3`). I obtained the result displayed in Figure 7a. This runs under mpi command line control with `-np 3` (resulting from the parallel environment dictionary entry `'ppn'`: `'3'`. Initially, shots 0, 3, and 6 are distributed to processes 1, 2, and 3. Then shots 1, 4, and 7 are distributed in the same way, and finally 2,5, and 8.

Note that the communicator design described above causes the algorithm to behave like an asynchronous queue - for example process 2 could be working on shot 4 when process 1 is still on shot 0. In general, IWAVE takes an optimistic

point of view regarding load balancing: it simply relies on similar-sized jobs taking similar lengths of time to complete. This approach has definite limits, but it is perhaps reasonable as a first cut for a solver based on uniform rectangular meshes.

Figure 7b shows the same job processed with 3 concurrent shots and a 1×2 domain decomposition. In this case, processes 1 and 2 start on shot 0, processes 3 and 4 on shot3, and processes 5 and 6 on shot 6, and so on. As was the case in other examples, the differences between results obtained with different levels of parallelism are within round-off of each other.
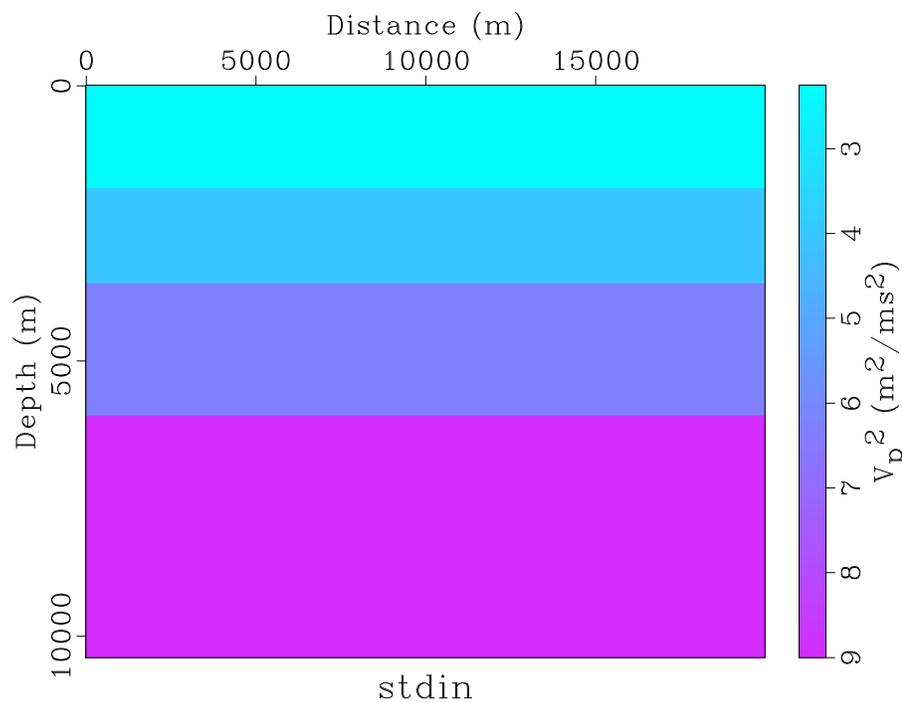
Figure 1: 4-layer velocity-squared model. First layer has H_2O properties: model sea bottom is at depth = 1875 m. iwavepar/project/ csq4layer

REFERENCES

Fehler, M., and J. Keliher, 2011, SEAM Phase 1: Challenges of subsalt imaging in Tertiary basins, with emphasis on deepwater Gulf of Mexico: Society of Explo-
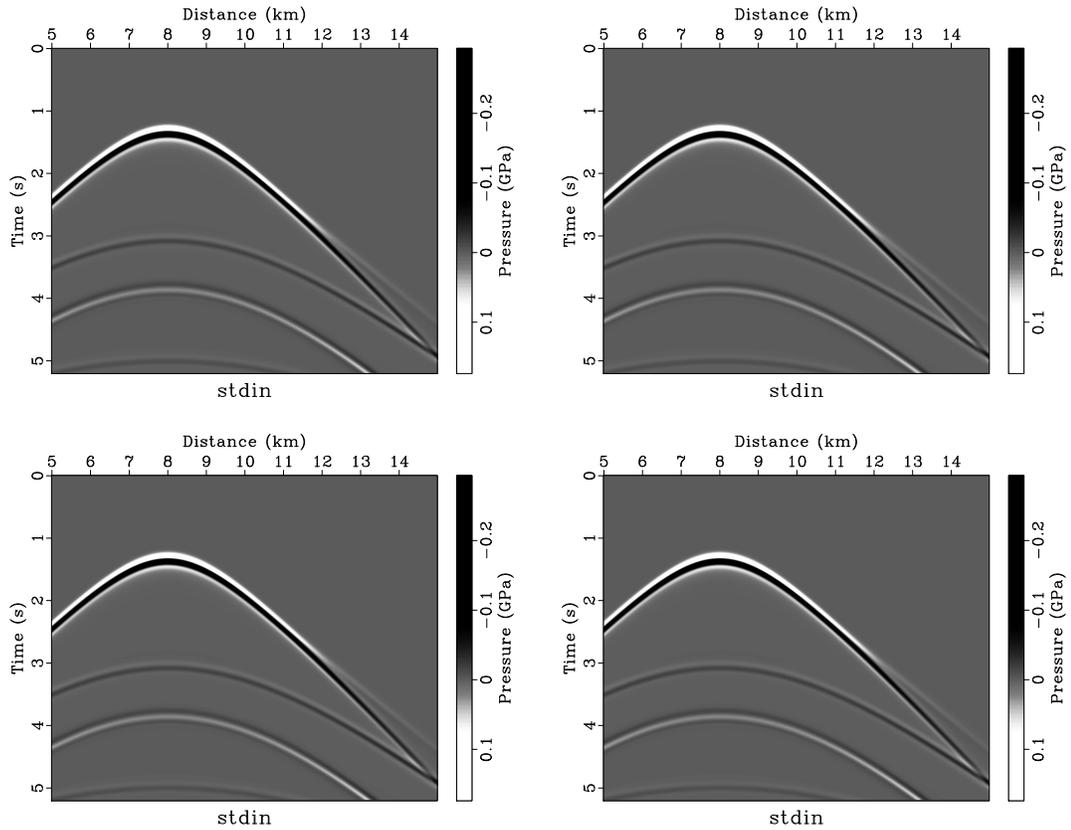
Figure 2: Shot over model of Figure 1 at OBC position $x = 8$ km, $z = 1.875$ km, 400 receivers in 5 km $\leq x \leq$ 15 km, $\Delta x = 25$ m. Isotropic point source as constitutive law anomaly, Gaussian derivative wavelet, 5 Hz peak frequency. Domain decomposition parameters (mpi_np1,mpi_np2) = (1,1) (no domain decomp), (1,2), (2,1) and (2,2). iwavepar/project/ shot11,shot12,shot21,shot22
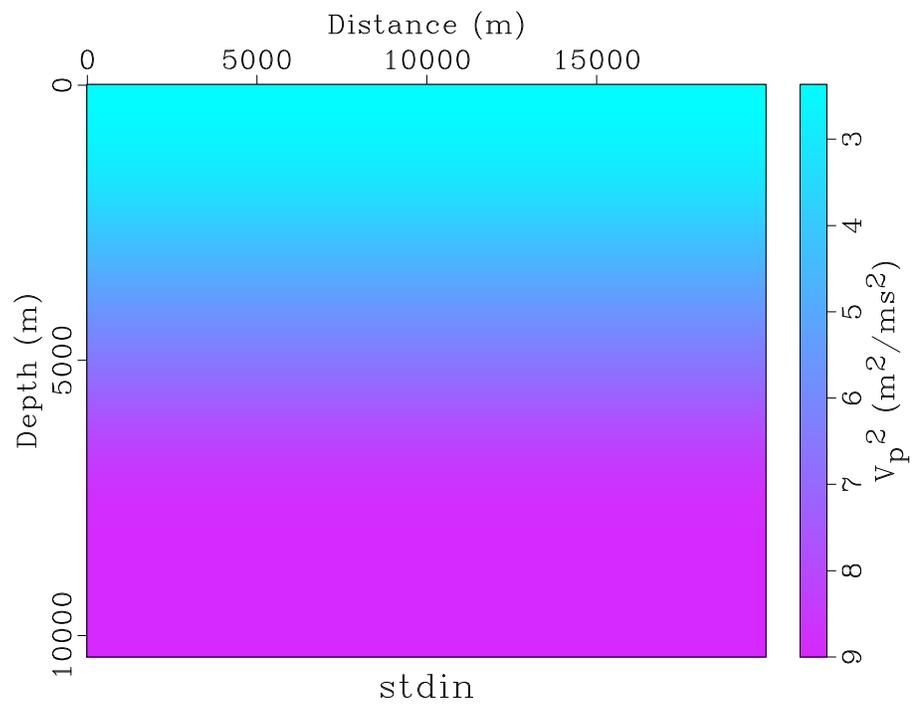
Figure 3: Background model: smoothing of 4-layer velocity-squared. iwavepar/project/ csq4layersm
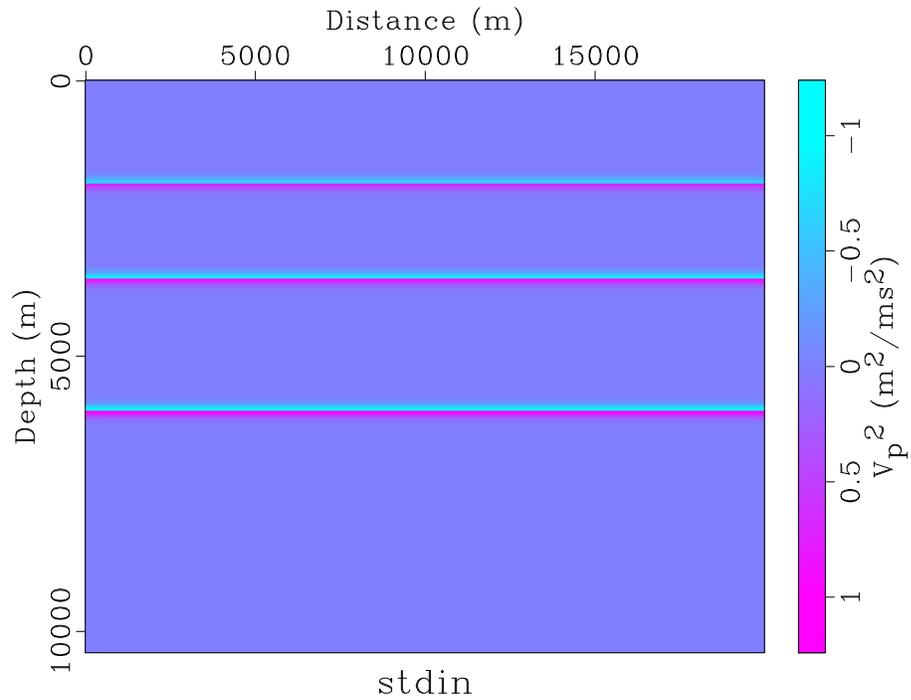
Figure 4: Reflectivity model: difference between velocity-squared (Figure 1) and a less aggressive smoothing than that shown in Figure 3. iwavepar/project/ dcsq4layer
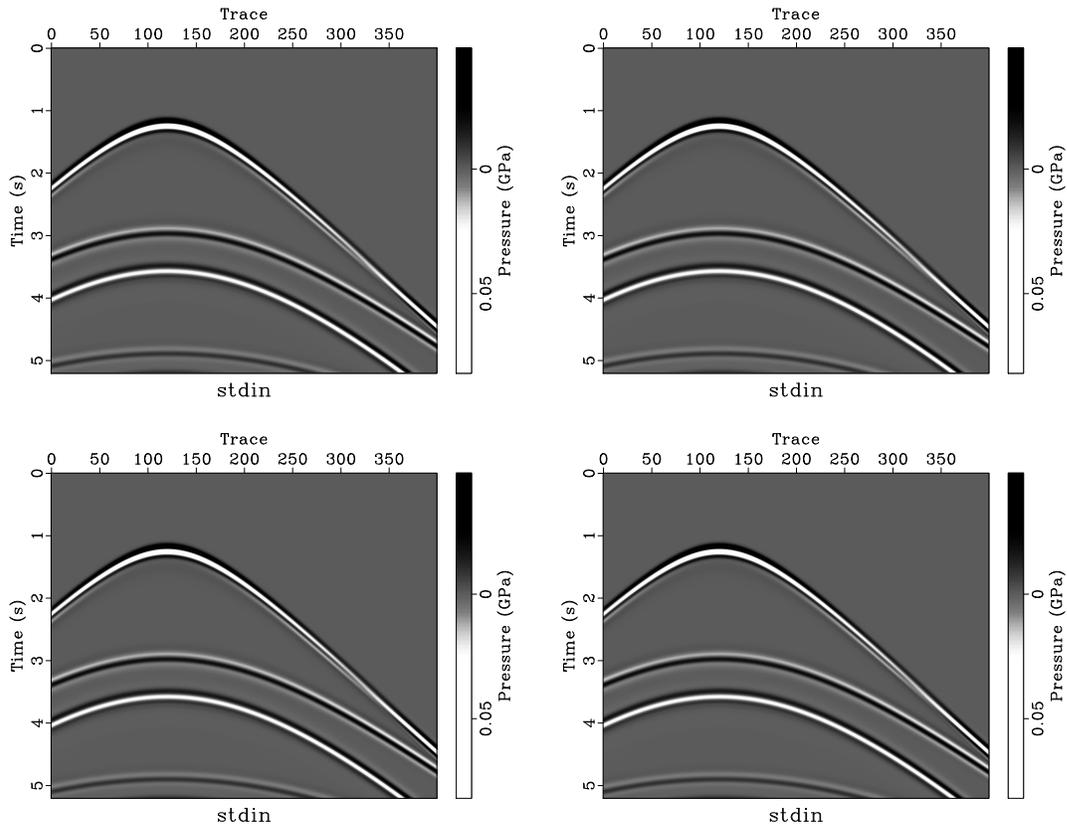
Figure 5: Linearized ("Born") shot over model of Figure 1 at OBC position $x = 8$ km, $z = 1.875$ km, 400 receivers in 5 km $\leq x \leq$ 15 km, $\Delta x = 25$ m. Isotropic point source as constitutive law anomaly, Gaussian derivative wavelet, 5 Hz peak frequency. Velocity-squared as in Figure 3, reflectivity (velocity-squared perturbation) as in Figure 4. Domain decomposition parameters (mpi_np1,mpi_np2) = (1,1) (no domain decomp), (1,2), (2,1) and (2,2). iwavepar/project/ born11,born12,born21,born22
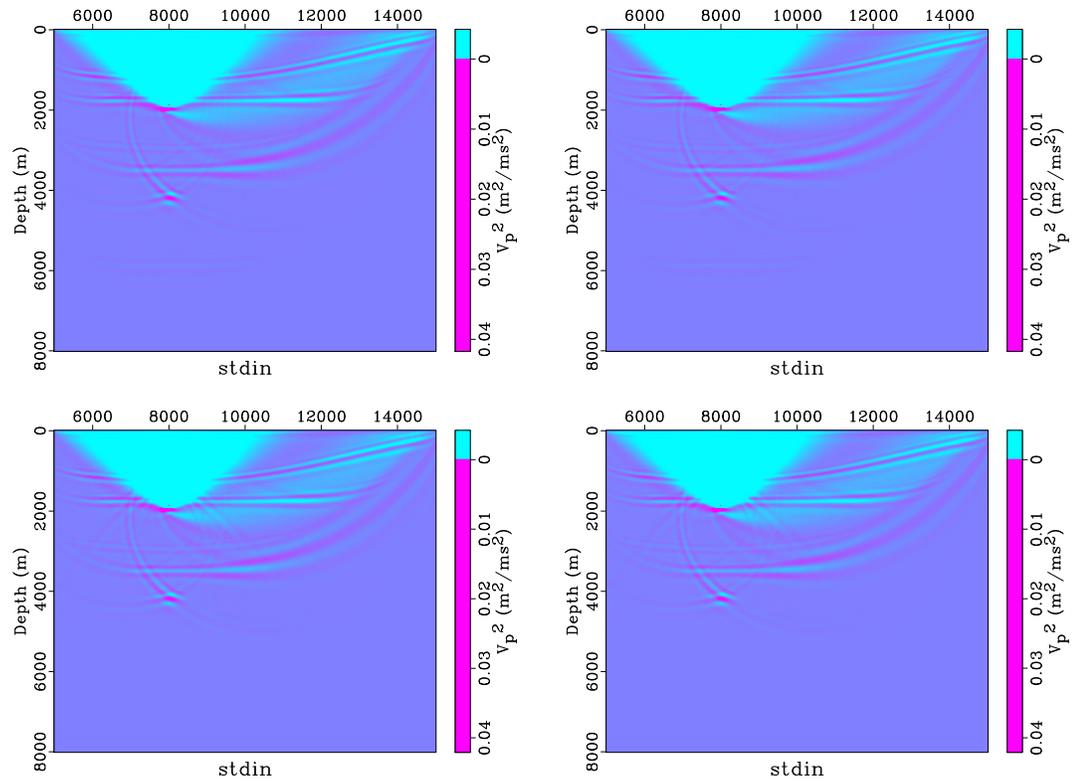
Figure 6: Adjoint linearized map (RTM) applied to data of Figures 5a, 5b, 5c, and 5d, with same domain decomposition parameters. iwavepar/project/ rtm11,rtm12,rtm21,rtm22
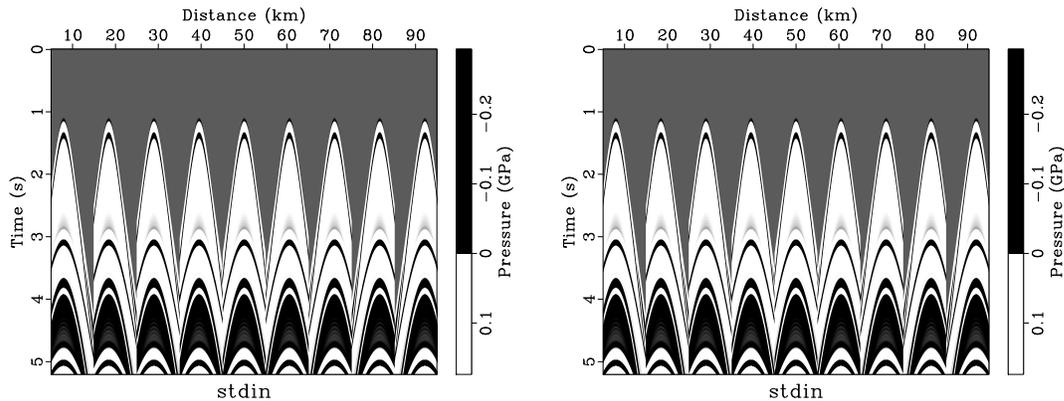
Figure 7: 9 shots from $x_s = 8$ km to $x_s = 12$ km, 3 shots computed at once (partask=3); no domain decomposition on left, 1×2 domain decomposition on right. iwavepar/project/ shot8-12km,shot8-12kmdd

ration Geophysicists.

Symes, W., 2014, IWAVE structure and basic use cases, *in* TRIP 2014 Annual Report: The Rice Inversion Project. (available summer 2015).

———, 2015, Extended IWAVE, *in* TRIP 2015 Annual Report: The Rice Inversion Project. (available summer 2016).

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).

Terentyev, I. S., 2009, A software framework for finite difference simulation: Technical Report 09-07, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, USA.

Extended IWAVE

Raanan Dafni and William W. Symes, *The Rice Inversion Project*

ABSTRACT

IWAVE accommodates several extended modeling modes. This paper explains how to implement shot record extension (independent simulation of shots with one model per shot), subsurface offset extension (mechanical parameters as non-diagonal operators), and source extension (independent source parameters for each trace, or each shot, or both), for constant-density acoustic modeling. The modeling operators and their first and second derivatives inherit all features of IWAVE simulation - dataflow design, shot and multidomain parallelism, computed memory allocation, command line job control, etc.

INTRODUCTION

MATHEMATICAL SETTING

The central design principle of IWAVE/RVL can be stated as follows: *the mathematics is the API*. So the first task is to describe the mathematics of extended modeling.

To make the illustrations in this paper as simple as possible, we choose constant density acoustic wave propagation as the underlying physics. The sole material parameter field is the velocity squared, $m = v^2$, proportional to bulk modulus by the (constant) density. The second order wave equation for causal acoustic dynamics is

$$\frac{\partial^2 p}{\partial t^2} - m \nabla^2 p = f, \; p = 0, t << 0. \tag{1}$$

in which $p = p(\mathbf{x}, t, \mathbf{x}_s)$ is the pressure field, $\mathbf{x}_s$ is the source parameter (which may be location, or slowness, or ...), $f = f(\mathbf{x}, t, \mathbf{x}_s)$ is the source field. The modeled

seismogram is the trace, or sampling, of p at (possibly $\mathbf{x}_s$-dependent) receiver locations. For the moment, we will regard the source as known and fixed, and write $\mathcal{F}(m)$ for the collection of traces, so determined, a point in the data space D, the range of the forward map $\mathcal{F}$ whose domain is the model space M.

An extended model space $\bar{M}$ is simply larger: M "$\subset$" $\bar{M}$ - as we will see, the inclusion need not be literally true. Typically extended modeling means parameters added somehow to the model, beyond those specified by the basic physics, but the fundamental requirement is that there be an extended modeling operator $\bar{\mathcal{F}} : \bar{M} \to D$ and an extension operator $E : M \to \bar{M}$ so that

$$\bar{\mathcal{F}}(E(m)) = \mathcal{F}(m), \; m \in M. \tag{2}$$

Shot Record Extension

Subsurface Offset Extension

Dataflow Control

IWAVE controls the selection of extension (or no extension) by *input and/or output data format*. This idea is described in (Symes, 2014). Briefly,

- if the grid of the input fields (or output fields, in the case of the adjoint linearized map) contains one or more extended axes, then IWAVE computes an extended map. As explained in (Symes, 2014), IWAVE uses an enhanced version of the Madagascar data structure to differentiate extended from non-extended axes: added keywords are

    - `dim` = spatial dimension, should be same as reference grid
    - `gdim` = global dimension, including extended axes - for typical 2D extended modeling, this is `dim` + 1.
    - `idxxx` for `xxx=1,...dim-1` are the id's of the spatial axes, `xxx=dim` signifies the time axis, `xxx=dim + 1,...,99` are available for external extended axes such as shot number, and `xxx=100,...` signify internal extended axes such as subsurface offset. "External" means that only a single

value participates in the simulation of a single shot; "internal" means that all (or many) points on the axis particpate in single shot simulation, as is the case for spatial axes.

– for linearized or adjoint linearized maps,

* if the grid of the input reference fields is the same as the grid of the perturbation inputs or outputs (csq_d1 or csq_b1 for acd first derivative and adjoint first derivative respectively), then IWAVE computes the linearization (or adjoint linearization) of the (extended or non-extended) forward map.

* if the grid of the input reference fields is not extended (has no extended axes), but one or more perturbations is extended, then IWAVE computes the extended linearization (or its adjoint) about the implicitly extended reference field. This case is typical for migration velocity analysis.

In all cases, the identification keywords of extended axes tell IWAVE what sort of extension to compute.

EXAMPLE: COARSE GRID MARMOUSI

This example uses the Marmousi model subsampled to $dx = dz = 24$ m. The Born simulation uses a well-smoothed background model (Figure 1) and a reflectivity derived from the original model with a less agressive smoothing removed (Figure 2). Velocity-squared and its perturbation are plotted, as these are the quantities used in the simulation. The acquisition geometry is the original, subsampled by a factor of 4 to give a 60-shot simulated towed streamer geometry. Time of recording is cut back to 2 s.

All of the computations reported here were performed at the Texas Advanced Computing Center (TACC), University of Texas-Austin, using TACC's Stampede Linux cluster. The source directory for this paper contains the project/SConstruct script file describing every command, and illustrates the configuration of reproducible computational experiments in a batch environment (SLURM). The (small) jobs described below ran on 32 threads (2 nodes) and complete in well under a minute. See the companion paper SymesIWPAR:15 for a description of the software tools used in these examples. Note that the number of tasks (keyword
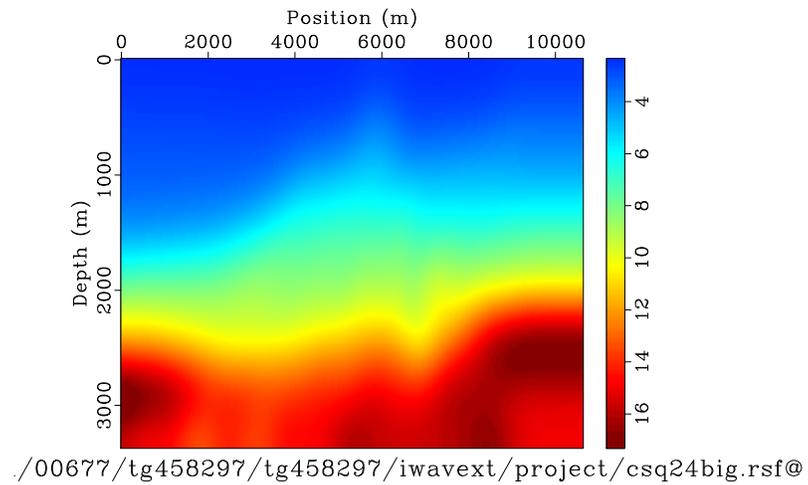
./00677/tg458297/tg458297/iwavext/project/csq24big.rsf@

Figure 1: Background model: square of smoothing of Marmousi velocity.

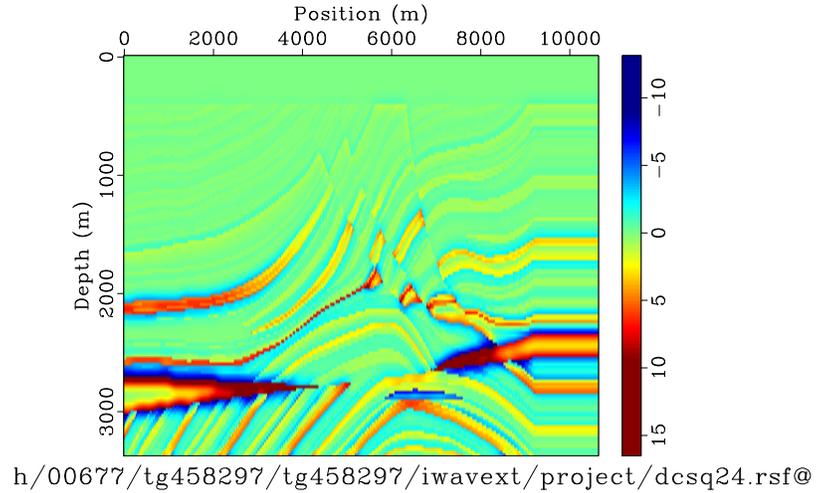h/00677/tg458297/tg458297/iwavext/project/dcsq24.rsf@

Figure 2: Reflectivity model: difference of squares of Marmousi velocity and a smoothing.

partask = number of shots to execute in parallel) was also 32, for reasons explained in (Symes, 2015). Since this example involves 60 shots, IWAVE idles some threads near the end of the run.

Shot Record Extension

All output (and, of course, input) data objects must exist for IWAVE to function properly. To build a prototype shot record extended model that can be used as an output object for shot record extended migration, use `sfspray` to duplicate the perturbational model (or any other RSF data object with the reference grid) 60 times. [For implementations of this and all other computations described here, see the `SConstruct` file in the `project` subdirectory.]

In addition to duplicating the reference grid, your command must decorate the rsf header file as described above: use `sfput` to add `id1=0`, `id2=1`, `id3=3`, `dim=2`, and `gdim=3` to the header file. Note that `id3=3` sets the first axis numbered above the time axis (hence extended) to `dim + 1`, which is the index of the shot record axis in the simulation grid (see (Symes, 2014) for explanation). These choices will cause IWAVE to overwrite a shot record extended image on the output (Figure 3). As expected, the common image gathers reflect the correct velocity in being as flat as possible subject to the limited aperture of shot record migration and presence of edge effects (Figure 3, right panel).

Subsurface Offset Extension

In this case, use `sfpad` to add a subsurface offset axis to the perturbational model to create a prototype for migration output. I chose to use 20 grid cells to the left and right of offset 0, with offset increment the same as the other spatial axis increments (24 m), so a maximum subsurface offset of 480 m. Header corrections for this case are `id1=1`, `id2=0`, `id3=100`, `d3=24`, `o3=-480`, `dim=2`, and `gdim=3`. Note all of the spatial fields are transposed in this case: while the model field fetched from the TRIP web site has *z* as the fast axis, *horizontal* subsurface offset extension is *required* to use *x* as fast axis. There are two reasons for this choice: (1) the organization of loops in the numerical kernals can be maximally vectorized with this choice, and (2) domain decomposition is not permitted on
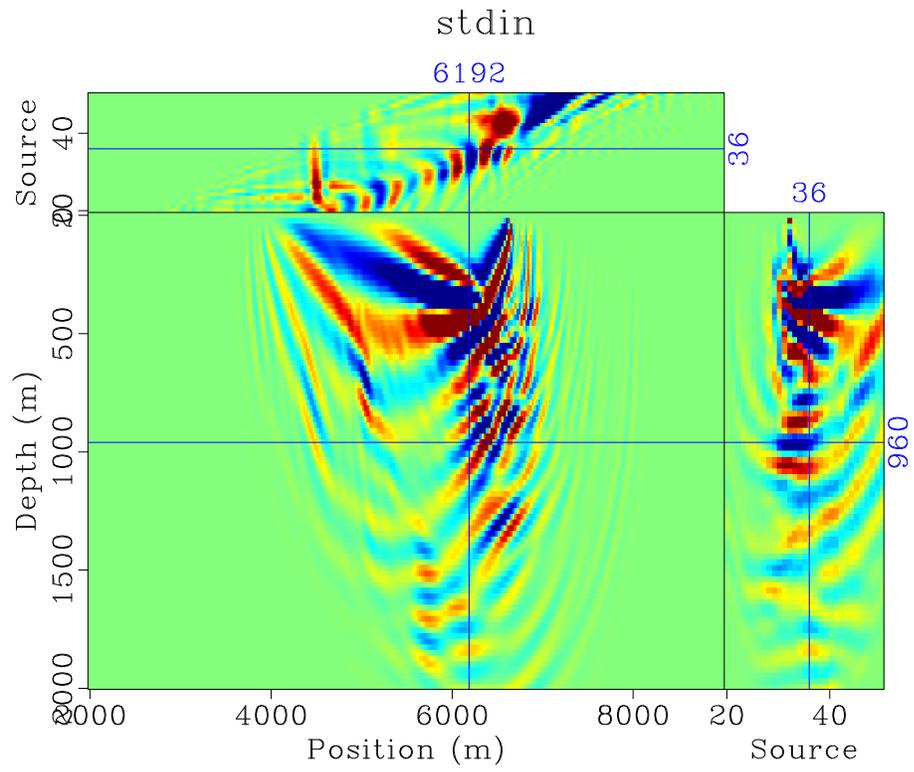
Figure 3: Shot record extended migration cube: front face is image

the fast axis, since the effective stencil width is very large in this dimension, and the restriction to x allows a simple implementation of this restriction. Since the physical significance of the axes is part of the data structure, IWAVE can take axis ordering into account so that it is transparent to the simulation. The data setup must include transposing of the velocity squared field and all perturbations, of course, and to display the output of the migration in the normal way it must be transposed back to fast-z order. The choice id3=100 toggles internal extension: at the level of IWAVE itself, this choice causes appropriate memory allocation to occur. The interpretation of the extended axis as horizontal subsurface offset is an attribute of the numerical kernels that go into acd (not of the core IWAVE code).

The resulting migration is reasonably well-focused at offset zero (Figure 4, right panel), and the zero-offset section is identical to non-extended migration (Figure 4, front panel).
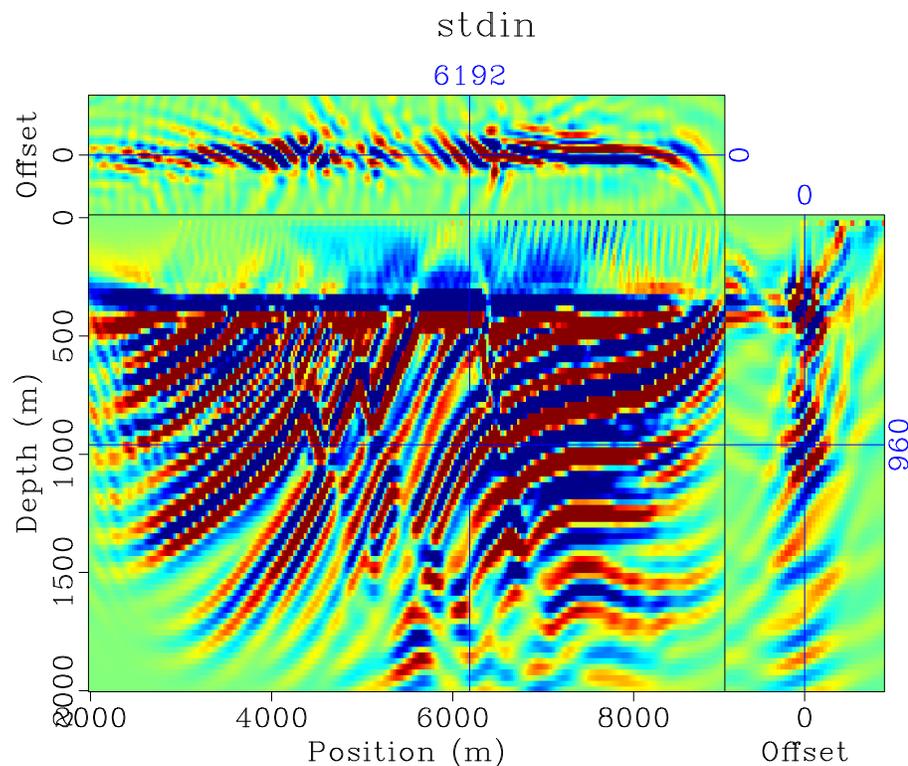
Figure 4: Subsurface offset extended migration cube: front face is image

REFERENCES

Symes, W., 2014, IWAVE structure and basic use cases, *in* TRIP 2014 Annual Report: The Rice Inversion Project. (available summer 2015).

——, 2015, Parallel IWAVE, *in* TRIP 2015 Annual Report: The Rice Inversion Project. (available summer 2016).

Basic Imaging: Viking Graben

William W. Symes, The Rice Inversion Project

ABSTRACT

Presents very simple processing sequence resulting in post-stack and prestack migrated images of Viking Graben data.

INTRODUCTION

The purpose of this paper is to explain how to use simple scripts and public-domain software to create images of the subsurface from reasonably tame 2D seismic reflection data. By "reasonably tame", I mean having few amplitude anomalies and mostly primary reflection energy. These conditions pretty much imply high quality shallow water marine data from regions with low structural relief and well-defined dip trends, shot along dip lines, with successful multiple removal applied.

The Viking Graben (or Mobil AVO) data placed in the public domain by Mobil Research in 1994 appears to be reasonably tame, in the sense just described."Viking Graben" refers to a 25 km long marine line from the North Sea Norwegian sector. The data released by Mobil included both raw field tape (`tape1093`) and parabolic Radon demultiple (`paracdp`) version of the data, along with two logs of v_p, v_s, and ρ, and a far-field wavelet recording. Data are in SEGY format (IBM 4-byte floats, plus text and binary reel headers).

Robert Keys and Douglas Foster were the organizers of the SEG post-convention workshop at which the initial results of various groups' efforts to invert this data were presented. Besides the workshop report volume (Keys and Foster, 1998), Keys and Foster have recently published a web page on this data:

`s3.amazonaws.com/open.source.geoscience/open_data/`

```
Mobil_Avo_Viking_Graben_Line_12/mobil_avo.html
```

This paper discusses so-called standard processing of the Viking Graben data, based on hyperbolic normal moveout (NMO). The connection between inversion and standard processing is covered in detail in my MSRI lecture notes from 2013, available on the TRIP web page, downloadable materials, short course and summer school materials, especially part 2.4. I also show how to transform the byproducts of standard processing into suitable input for prestack imaging via RTM.

My workflow is recorded in the project/SConstruct file. The reader should refer to this file for the detailed form of the commands used. The following sections outline the rationale for the choice of commands and their sequence. To reproduce my results, the reader will need to install SU, Madagascar, and trip2.1 packages. SU should be version 44 or later. The prestack part of the workflow requires parallel installation of trip2.1, and at least 10 threads of execution, preferably more, to finish in reasonable time.

PRELIMINARY STEPS

Extract the data using Fetch, from the TRIP data repository. The result is paracdp.segy, a SEGY-format file. The SU command segyread converts this file to SU format (no reel headers, little-endian 4-byte floats, same number of traces). Besides parabolic radon demultiple, this data was delivered with a rather strict mute applied, with no signal before 1 s. No further mute is applied in the processing sequence described here.

Evidently there are some bad traces in some of the low-number CDPs, with floating point garbage in some of the samples. On average, random binary digits form numbers many orders of magnitude larger than the $O(10^8)$ samples encountered in "good" CDPs. So an adequate strategy is to use sugain to reject samples in the 99%ile, and this command is included in the conversion to SU format.

EXTRACT ANALYSIS CDPS

I chose 4 CDPs (Figure 1a, Figure 1b, Figure 2a, Figure 2b) roughly evenly spaced, including the end zones, separated by about 8 km. This is a preliminary guess as to the necessary lateral resolution for velocity analysis. It turned out to be reasonable.

Also, I truncated the traces at 3 s, as the logged zone indicated that only reflection arriving before that time would likely image the exploration targets. This is of course an ex post facto decision, but in fact it becomes clear quickly that most of the energy after perhaps 4 s must be other than primary reflection, either multiples or sideswipe.

VELOCITY ANALYSIS

Create velan panels, one for each CDP: Figure 3a, Figure 3b, Figure 4a, and Figure 4b. Note that in each case there is a trend of strong bullets or streaks at lower velocities than some nearby higher velocity peaks. The higher velocity peaks are weaker, but pick them anyway. The stronger lower velocity features represent multiple reflections - the suppression of multiple energy notwithstanding. (If you carry out this exercise vvwith the field tapes, you will see nothing but water and near-water velocity features).

To QC your picks, plot the NMO correction of the corresponding CDPs as Figure 5a, Figure 5b, Figure 6a, and Figure 6b.

These CDPs are pretty easy to pick from the velocity spectra; however it is instructive to see what might go wrong. For CDP 2000, suppose you were to pick velocities of 1800 m/s at t=1.2, 1.6. and 1.8 s, and 2200 m/s at t=2.1 s - each of these coordinates with a feature on the velocity panel - instead of 1950, 2050, and 2250 m/s at 1.25, 1.6, and 2.25 s, as was done to create Figure 6b. Then you would generate Figure 7 instead.

STACK AND POST-STACK MIGRATION

As explained in the short course lectures, the velocity-denominated quantity determined by flattening NMO-corrected CDP gathers is the so-called RMS velocity, or v_{rms}, the root mean square of a local wave velocity expressed as a function of vertical travel time t_0:

$$v_{\mathrm{rms}}^2(t_0) = \frac{2}{t_0} \int_0^{t_0} d\tau \tilde{v}^2(\tau).$$ (1)

Vertical traveltime is related to depth by

$$z = \int_0^{t_0} d\tau \tilde{v}(\tau)$$ (2)

in which $\tilde{v}$ is the local (or *interval* wave velocity as a function of t_0. Velocity as a function of depth is then given by composing $\tilde{v}$ with the inverse of the tranformation (2).

This construction can be carried out for every midpoint. If the subsurface structure is sufficiently laterally homogeneous, then to good approximation the actual wave velocity is well-approximated by the interval velocity in the "well" below each midpoint. The interval velocity as function of t_0 or z can then be used to perform migration of zero-offset data, to good approximation. Also, to good approximation zero-offset data is similar to the stack of NMO-corrected CDPs.

This process requires a velocity at every CDP. So far we have only determined four such RMS velocity profiles. However the command sunmo will interpolate between midpoints, and extrapolate towards the ends of the line as necessary, and even output the interpolated v_{rms} as a function of t_0 and midpoint (keyword voutfile). This output is critical, as it can be converted later into other forms of velocity suitable for time and depth migration. Of course, the command also produces a stacked section (Figure 8).

Because the deeper events are much less energetically imaged than the shallow, automatic gain control (amplitude equalization) produces a much more informative Figure 9. This AGC'd stack will be input to all poststack processes.

You will note the precence of many *diffraction tails* especially below 2 s. These may be partially collapsed either by poststack time migration using Gazdag's algorithm (Figure 10) or, more effectively, by poststack depth migration using one

of several methods. Gazdag poststack time migration assumes layered velocity structure, hence uses only the first column of the interval velocity as function of tim. Nonetheless Figure 10 shows considerably less interference from diffraction tails; the graben structures that give this prospect its name are now clearly in evidence. Poststack depth migration (that is, what one would naturally mean - an approximation to the adjoint linearized zero-offset modeling operator) does even better - Figure 11 displays the output of Phase-Shift-Plus-Interpolation poststack depth migration, using the velocity $v(z, x)$ depicted in Figure 14 (obtained via use of the SU utility `velconv`, see `project/SConstruct` for details). The image gives reasonable depths for the various horizons. Also, this velocity model is the beginning of the next, prestack phase of Viking Graben processing.

With an eye to the prestack processing to be discussed in another paper, I include analogous results for downfiltered data. Jie Hou suggested a 5-10-30-40 Hz band-pass filter. The resulting AGC'd stack (Figure 12) shows the expected decrease in resolution over the original data, which has significant energy over 50 Hz.

[A question: the process, as you will see from the SConstruct, consists in filtering the data then applying NMO. Since the stack is a collection of time traces, you might think that you could get the same result by applying NMO first, then filtering. Is this true?]

The PSPI post-stack depth migration (Figure 13) reveals the same structure as does the stack, with the same decreased resolution over Figure 11, but with diffraction tails suppressed and stretched to depth.

CONCLUSION

Use of simplified physics (NMO, stack, poststack migration) based on layered modeling produces plausible subsurface structure images from the Viking Graben data. Public domain software - in this exercise, mostly SU - provides enough functionality to carry out this imaging task, provided that some basic pre-processing is performed (multiple suppression, mute). Mobil provided a suitably preprocessed version of the data in 1994.

The sequence of processing steps takes us far enough from the basic physics of wave propagation to raise questions about whether the images produced here are

actually images of anything real. In a subsequent paper, we turn to prestack processing based on the acoustic wave equation - only a step further, but meaningful nonetheless. To telegraph part of the punch: prestack processing supports the validity of the images obtained in this paper by poststack processes. It seems that in the case of this "tame" data, classic seismic data processing does not steer you wrong.

REFERENCES

Keys, R., and D. Foster, 1998, Comparison of seismic inversion methods on a single real dataset: Society of Exploration Geophysicists.
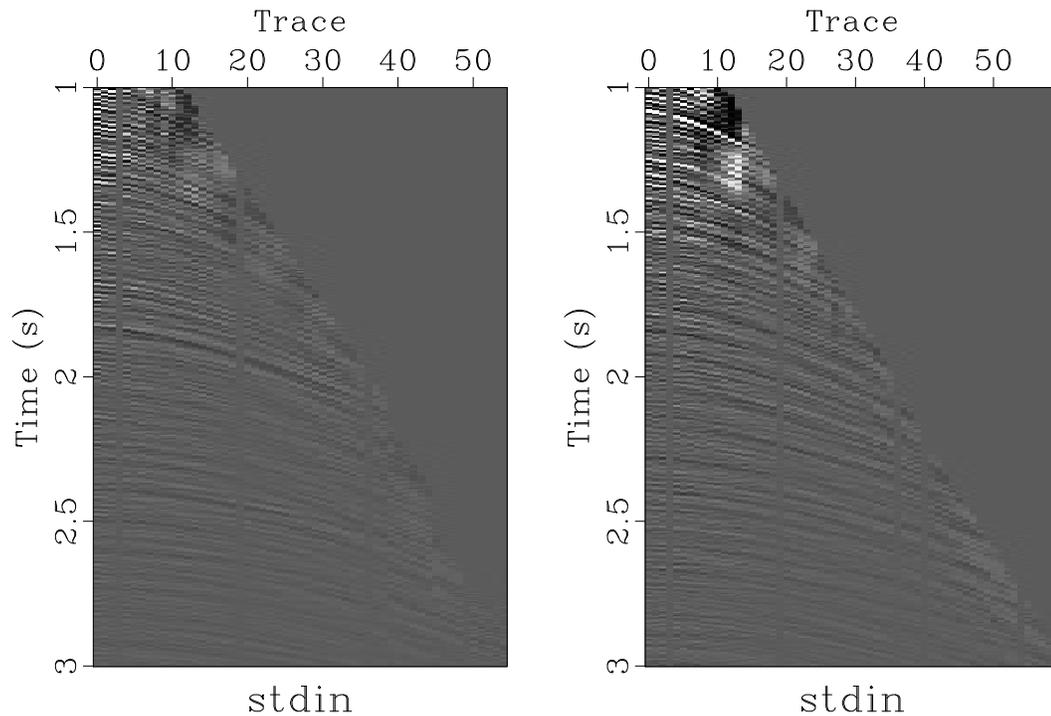
Figure 1: Left: CDP 200; Right: CDP 700
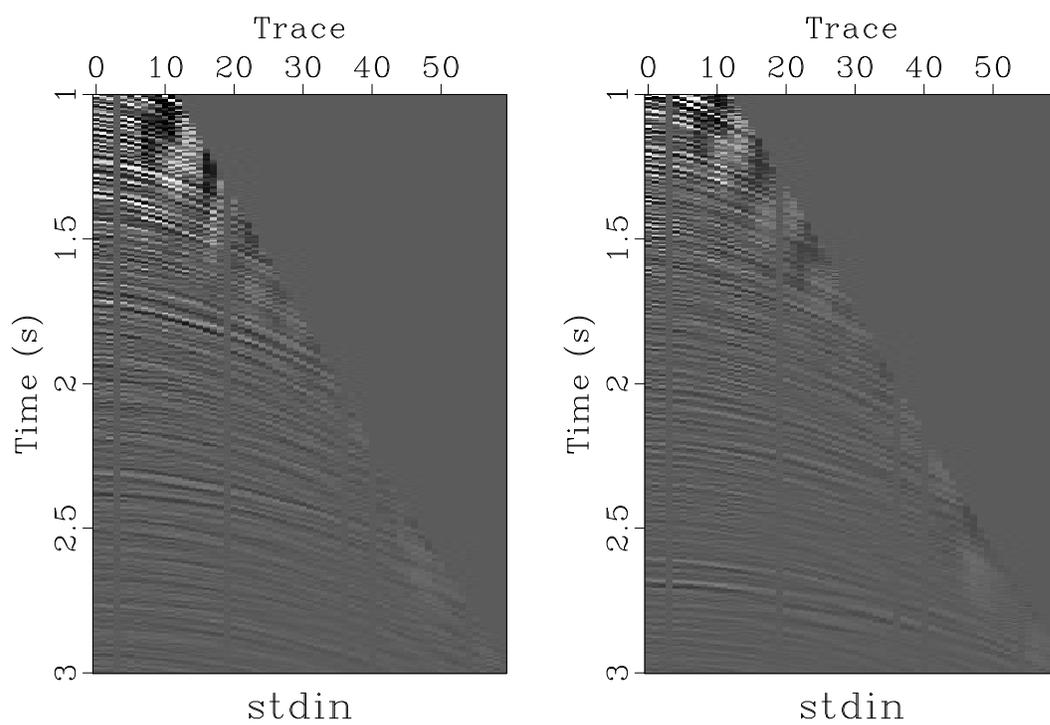
Figure 2: Left: CDP 1300; Right: CDP 2000
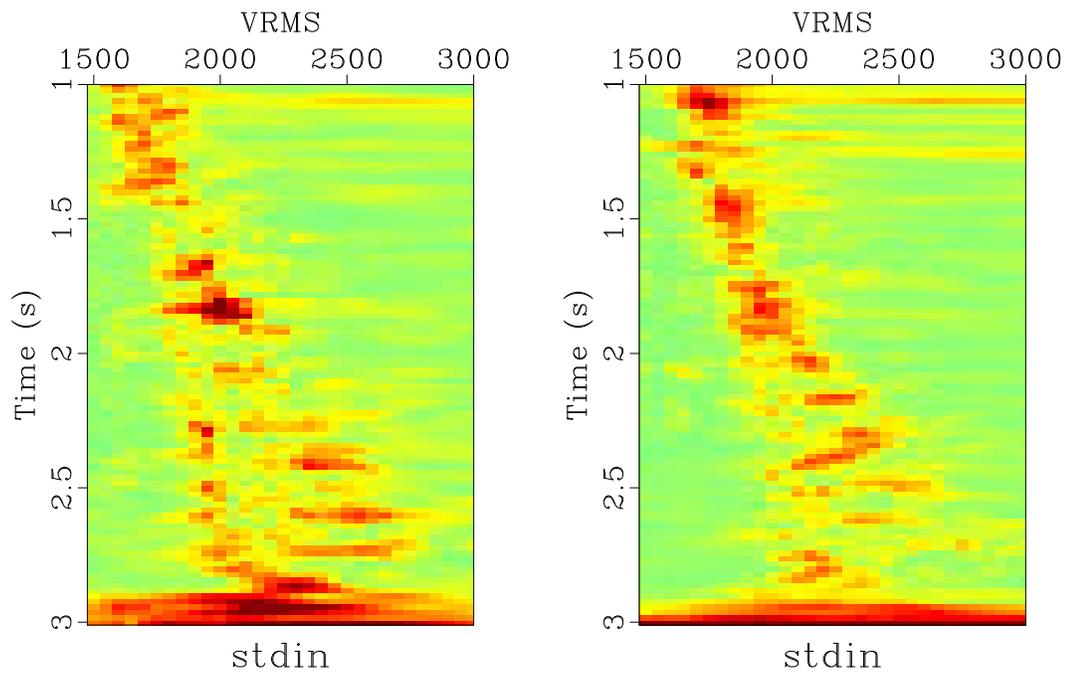
Figure 3: Left: Velocity spectrum for CDP 200; Right: Velocity spectrum for CDP 700
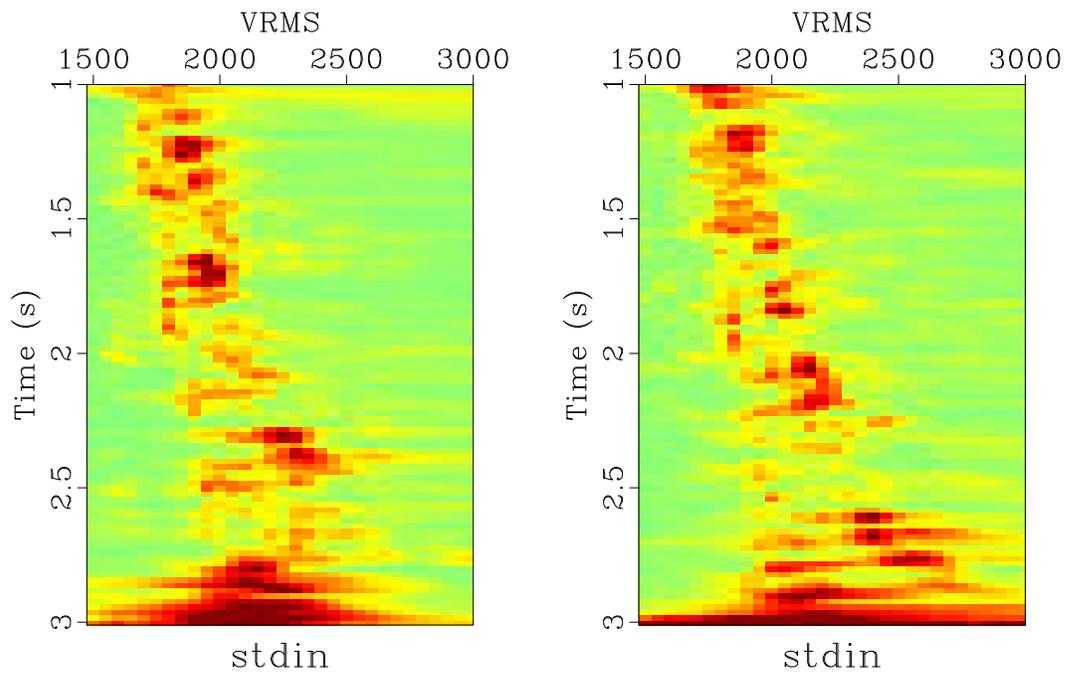
Figure 4: Left: Velocity spectrum for CDP 1300; Right: Velocity spectrum for CDP 2000
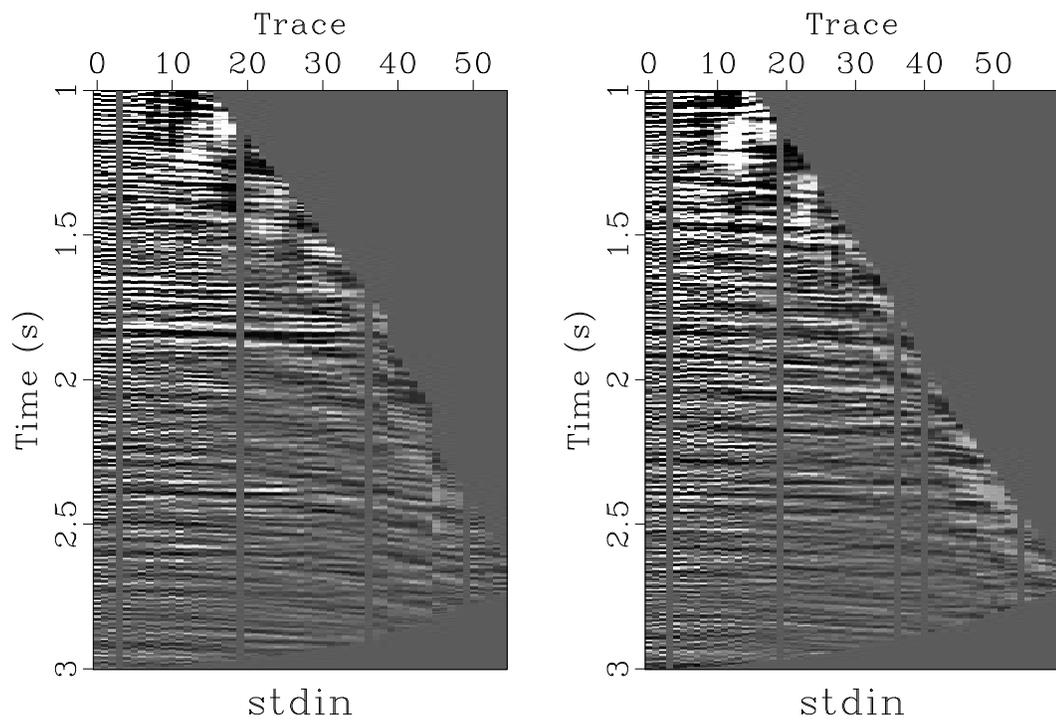
Figure 5: Left: NMO corrected CDP 200; Right: NMO corrected CDP 700
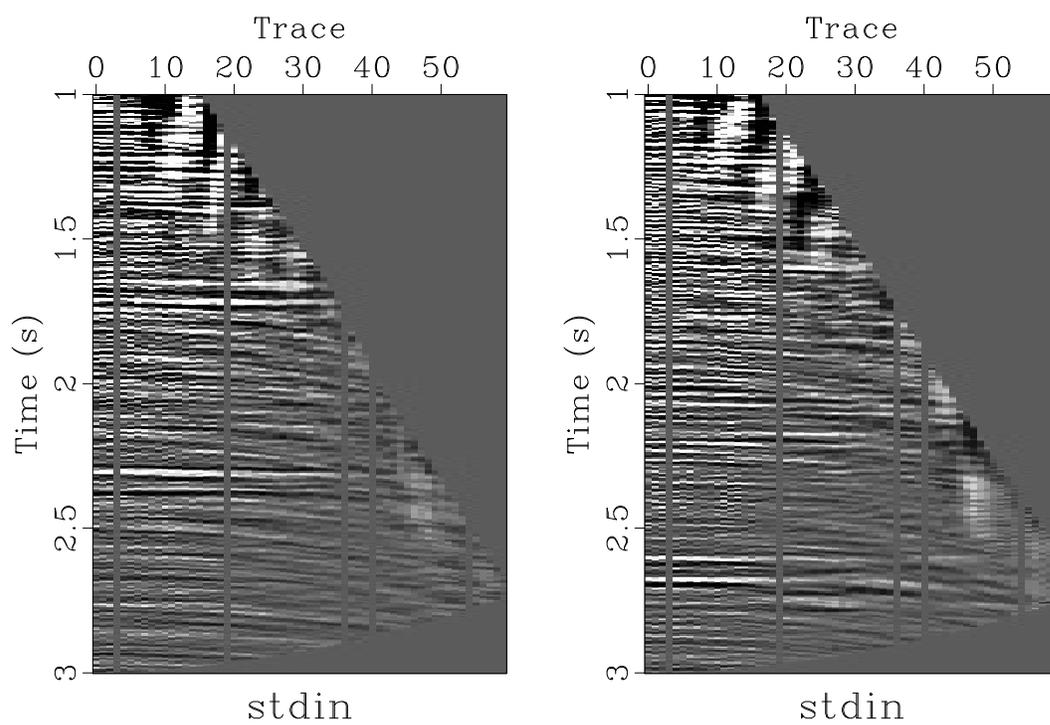
Figure 6: Left: NMO corrected CDP 1300; Right: NMO corrected CDP 2000
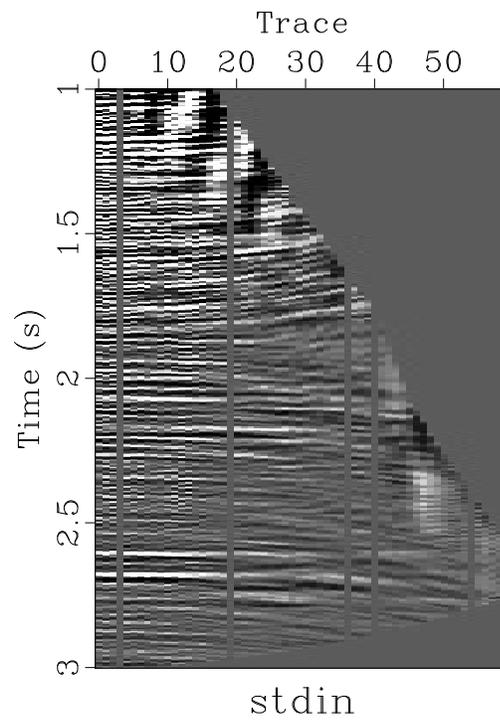
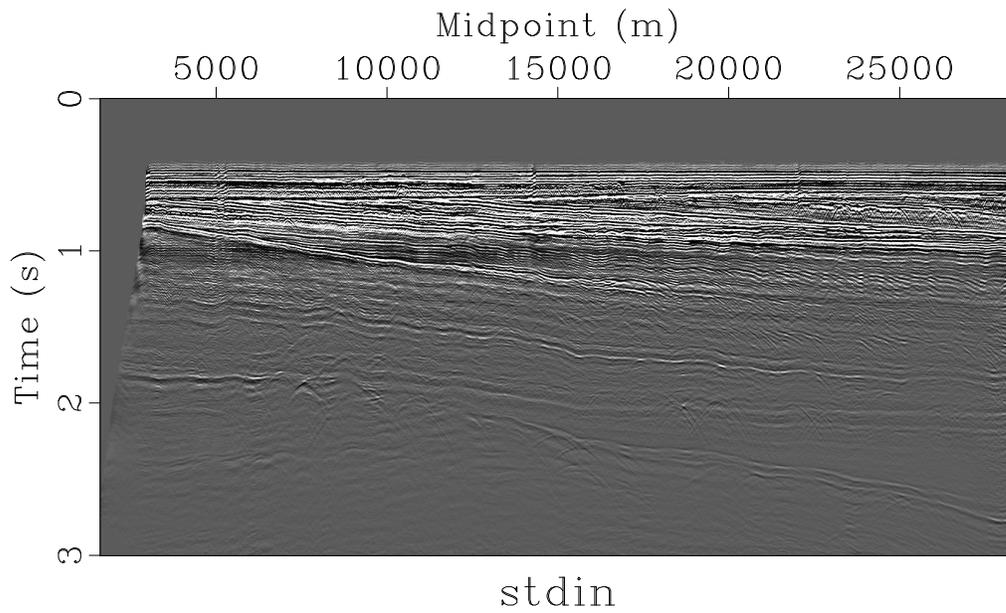Figure 7: NMO corrected CDP 2000 with tnmo=0.0,1.2,1.5,1.8,2.1,2.6,2.75 vnmo=1500,1800,1800,1800,2200,2400,2700

Figure 8: Stack with linearly interpolated, constant extrapolated v_{rms}
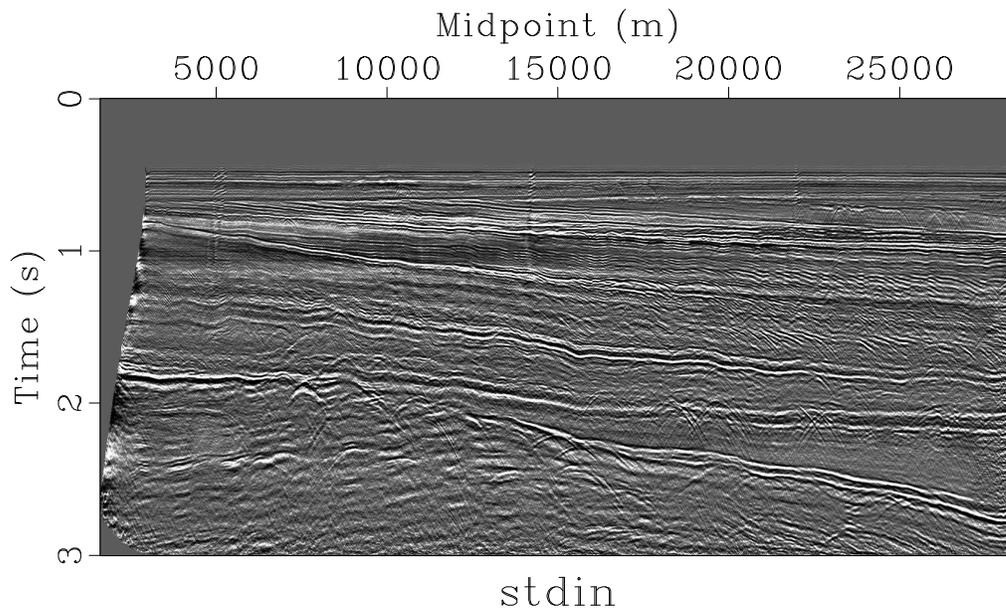
Figure 9: Automatic Gain Control (AGC) applied to data of Figure 8. Note pronouced diffraction hyperbolae in the deeper part of the secion.
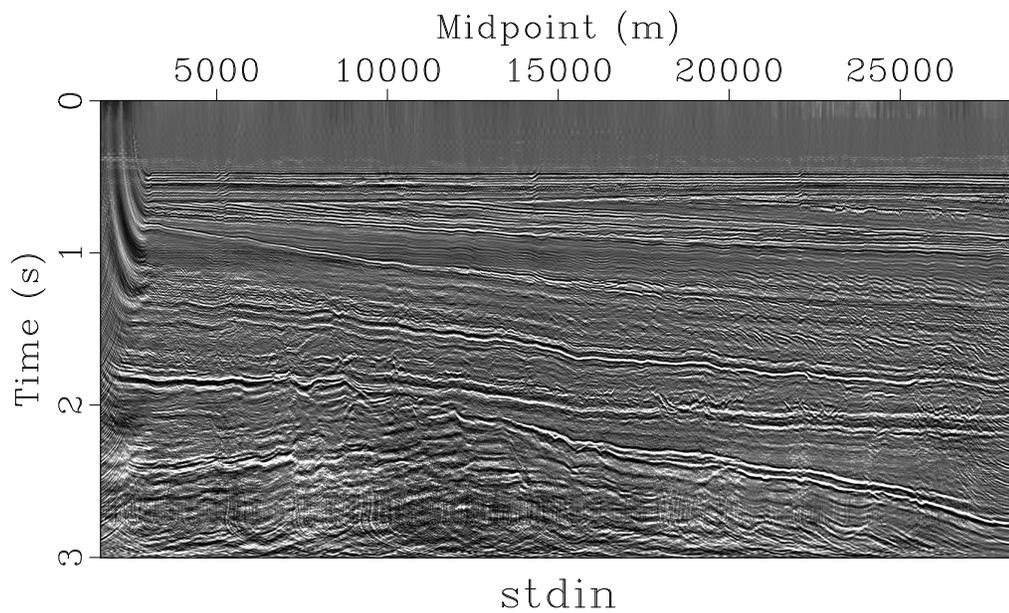
Figure 10: Gazdag's phase-shift time migration applied to the stack (Figure 8), followed by AGC. Diffraction artifacts are largely collapsed to the diffracting points that caused them, and the graben structures that give this prospect its name begin to be clearly visible. This migration assumes horzontal layering, which is not correct on the scale of 25 km.
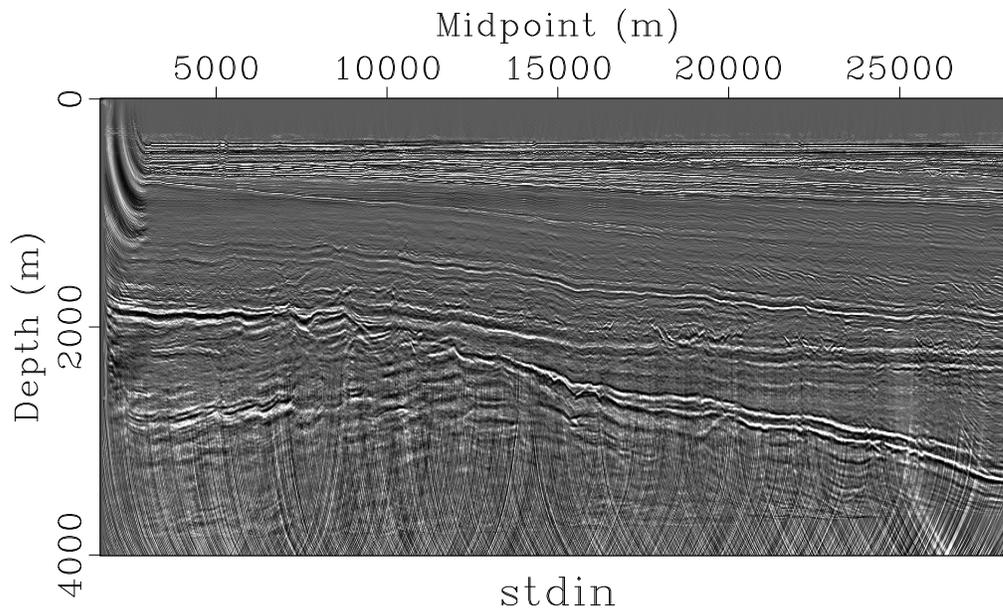
Figure 11: Gazdag's phase-shift-plus-interpolation post-stack depth migration applied to the stack (Figure 8), followed by AGC. Accounts for lateral velocity variation - uses interval velocity as function of midpoint and depth - more accurately than does the time migration of Figure 10. The geology is even more clearly delineated, and depths should approximate those to be obtained by more sophisticated imaging.
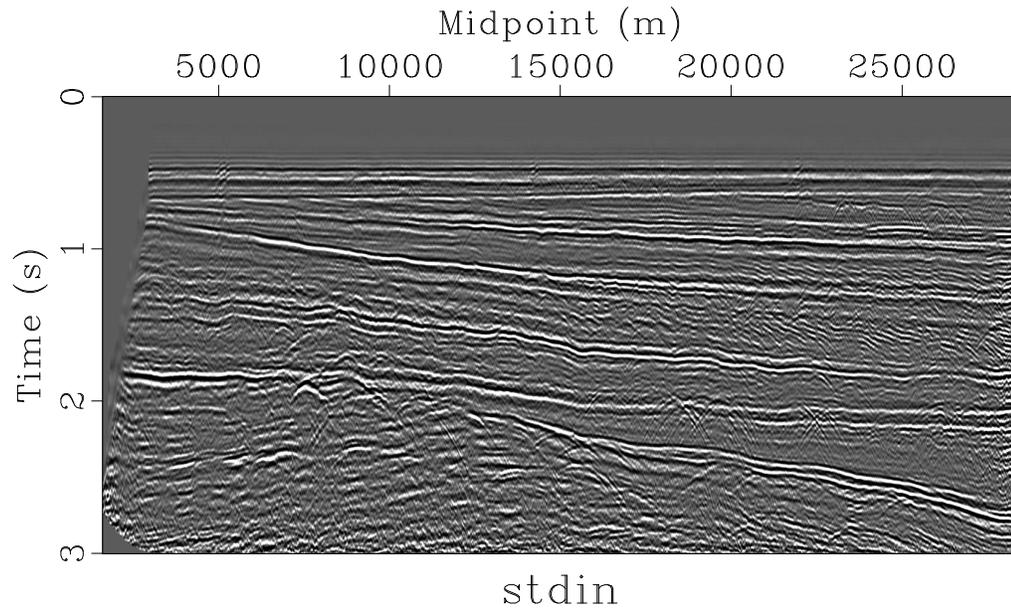
Figure 12: NMO stack of (5,10,30,40) Hz bandpass filtered data, after AGC. Compare to Figure 8.
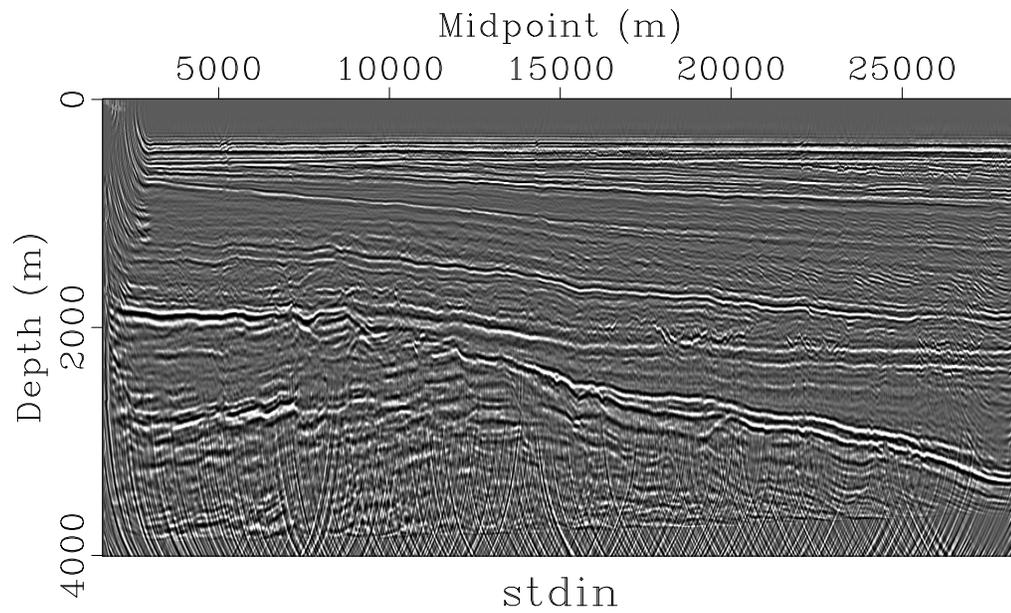

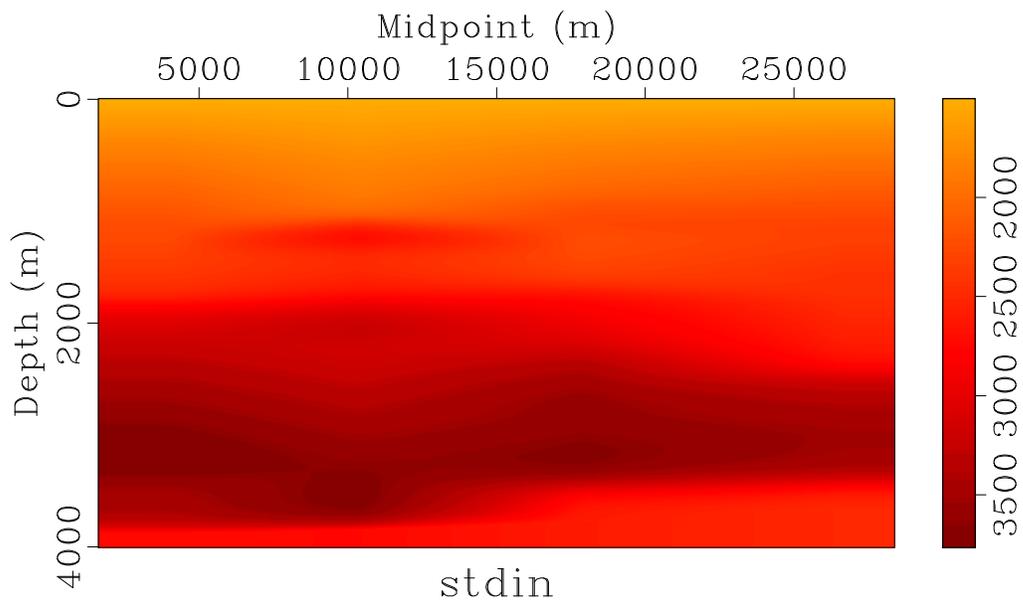Figure 13: PSPI post stack migrated image from stack in Figure 12.

Figure 14: Interval velocity as function of depth, derived from NMO velocity analysis. Probably not to be taken seriously below 3 km - for initial MVA estimate should extend by contant from 3 km and smooth.