Computational analysis of extended full waveform inversion

Ali Almomin

ABSTRACT

I compare the computational cost of conventional full-waveform inversion to the extended full-waveform inversion in both space and time. These model space extensions provide accurate results but increase the cost drastically. I also compare the cost of the nonlinear inversion to linearized inversion by scale-separation. I then propose extending the inversion in data space where there are more underlying assumptions but whose cost competes with the conventional inversion. I test extending the inversion by source ray parameter on the Marmousi model. The results of the synthetic tests show that convergence is possible even with large errors in the initial model which would have prevented convergence of conventional full-waveform inversion.

INTRODUCTION

Seismic velocity analysis methods can be divided into two major groups. First, there are techniques that aim to minimize the misfit in the data domain, e.g., full waveform inversion (FWI) (Tarantola, 1984; Pratt, 1999; Luo and Schuster, 1991). Second, there are other techniques (Symes and Carazzone, 1991; Biondi and Sava, 1999; Shen, 2004; Zhang et al., 2012), that aim to improve the quality in the image domain, such as migration velocity analysis (MVA). These approaches try to measure some quality of the image and then invert for the estimated image perturbation using a linearized wave-equation operator.

There are significant advantages to minimizing the residual in the image space: global convergence, increased signal-to-noise ratio, and decreased complexity of the data (Tang et al., 2008). However, a common drawback to doing velocity analysis in the image domain is that only the transmission effects of the velocity are used. This results in incomplete vertical resolution in the estimated model updates. On the other hand, FWI does not have that problem, since it utilizes the information from both the forward-scattered and back-scattered wavefields. This results in higher resolution model estimates. Moreover, the data misfit is computed in the data space directly without the need to go to another domain or to separate the data into several components. This direct computation results in a relatively simple relationship between the data residuals and the model updates. However, FWI has the disadvantage that its objective function is far from being smooth and convex; it requires the starting model to be very close to the true model to avoid convergence to local minima.

Biondi (2012) presents a generalized framework for full waveform inversion that avoids the cycle-skipping problem while utilizing all the components of seismic data to invert for the medium parameters. This is achieved in two steps: first by extending the velocity model through an additional degree of freedom, and second by imposing a regularization to constrain this added degree of freedom.

In this paper, I compare the cost of conventional full waveform inversion to extended inversion in model space that uses subsurface offset and time lags. I also compare the cost of the extended inversion to linearized inversion by scale separation (Almomin and Biondi, 2012). Next, I propose extending full waveform inversion through a data space axis, such as source location or source ray parameter, instead of model space axes. Finally, I test the source ray parameter extension on the Marmousi model.

COMPUTATIONAL COST

Nonlinear wave equation inversions are commonly performed using gradient-based iterative optimization. Each iteration typically consists of three steps. First, calculate the residuals, which requires one nonlinear forward operation. Second, calculate the gradient, which requires one linearized adjoint operation. Finally, determine the step size by a line search, which requires few nonlinear forward operations. For our calculations, we will assume the line search requires three forward operations results in a total per iteration effort of four forward and one adjoint operations.

Conventional FWI

For conventional full waveform inversion, the modeled data is computed using the nonlinear forward operator as:

$$d(\omega, \mathbf{x}_r, \mathbf{x}_s; v(\mathbf{x})) = \sum_{\mathbf{x}} f(\omega, \mathbf{x}_s) G(\omega, \mathbf{x}, \mathbf{x}_s; v(\mathbf{x})) \delta(\mathbf{x}_r - \mathbf{x}), \tag{1}$$

where $d(\omega, \mathbf{x}_r, \mathbf{x}_s; v(\mathbf{x}))$ is the modeled data, $v(\mathbf{x})$ is the velocity model, $f(\omega, \mathbf{x}_s)$ is the source function, ω is frequency, $\mathbf{x}_s$ and $\mathbf{x}_r$ are the source and receiver coordinates, and $\mathbf{x}$ is the model coordinate. In the acoustic, constant-density case, the Green's function $G(\omega, \mathbf{x}, \mathbf{x}_s; v(\mathbf{x}))$ satisfies:

$$\left(\frac{\omega^2}{v^2(\mathbf{x})} + \nabla^2 \right) G(\omega, \mathbf{x}, \mathbf{x}_s; v(\mathbf{x})) = \delta(\mathbf{x}_s - \mathbf{x}). \tag{2}$$

The propagation can be done in the time domain by convolving each model point with a finite-difference stencil. However, the time marching requires the time axis sampling to satisfy dispersion and stability conditions (Marfurt, 1984), generally much finer than the data sampling. Moreover, each time step requires multiplying the time slice

by the velocity squared. Therefore, the cost of forward modeling can be written as:

$$C_{\text{FWI-F}} = N_x N_y N_z N_{\text{source}}(N_{tp} C_{\text{FDTD}} + N_{tp}), \tag{3}$$

where N_x, N_y and N_z, are the number of points along the three spatial axes, N_{source} is the number of sources, C_{FDTD} is the cost of convolving one model location by the time-domain finite-difference stencil and N_{tp} is the number of time samples for propagation. By linearizing equation 1 over the squared slowness, we can compute the adjoint as:

$$\Delta s^2(\mathbf{x}) = \sum_{\omega,\mathbf{x}_r,\mathbf{x}_s} \omega^2 f(\omega,\mathbf{x}_s) G(\omega,\mathbf{x},\mathbf{x}_s;v(\mathbf{x})) G(\omega,\mathbf{x},\mathbf{x}_r;v(\mathbf{x})) \Delta d^*(\omega,\mathbf{x}_r,\mathbf{x}_s;v(\mathbf{x})), \tag{4}$$

where $\Delta s^2(\mathbf{x})$ is the perturbation in squared slowness and * denotes the complex conjugate. For the adjoint, the imaging time sampling can be much larger than that of propagation since it does not need to satisfy the dispersion and stability conditions. Hence, the cost of computing the adjoint of FWI can be written as:

$$C_{\text{FWI-A}} = N_x N_y N_z N_{\text{source}}(2 \times N_{tp} C_{\text{FDTD}} + 2 \times N_{tp} + N_{ti}), \tag{5}$$

where N_{ti} is the number of time samples for imaging. The total cost of one iteration of FWI becomes

$$C_{\text{FWI}} = N_x N_y N_z N_{\text{source}}(6 \times N_{tp} C_{\text{FDTD}} + 6 \times N_{tp} + N_{ti}). \tag{6}$$

Model-Space Extensions

Biondi and Almomin (2012) introduced an extension to full waveform inversion that can mitigate the cycle-skipping problem and allow for a larger error in the initial model. This is achieved by extending the velocity model along the subsurface offset and then solving the corresponding extended wave equation. The modeled data becomes:

$$d(\omega,\mathbf{x}_r,\mathbf{x}_s;v(\mathbf{x},\mathbf{x}_h)) = \sum_{\mathbf{x}} f(\omega,\mathbf{x}_s) G(\omega,\mathbf{x},\mathbf{x}_s;v(\mathbf{x},\mathbf{x}_h)) \delta(\mathbf{x}_r - \mathbf{x}), \tag{7}$$

and the extended Green's function satisfies:

$$\left(v^2(\mathbf{x},\mathbf{x}_h) *^{-1} \omega^2 + \nabla^2\right) G(\omega,\mathbf{x},\mathbf{x}_s;v(\mathbf{x},\mathbf{x}_h)) = \delta(\mathbf{x}_s - \mathbf{x}), \tag{8}$$

where $\mathbf{x}_h$ is subsurface offset and $*^{-1}$ denotes a deconvolution operator over subsurface offset. This extended wave equation convolves each time slice by all subsurface offsets of velocity. The cost of extended forward modeling becomes:

$$C_{\text{EFWI-F}} = N_x N_y N_z N_{\text{source}}(N_{tp} C_{\text{FDTD}} + N_{tp} N_{hx} N_{hy}), \tag{9}$$

where N_{hx} and N_{hy} are number of subsurface offsets along the x and y axes, respectively. By linearizing equation 7 over the velocity squared, we can compute the adjoint as:

$$\Delta v^2(\mathbf{x}, \mathbf{x}_h) = \sum_{\omega, \mathbf{x}_r, \mathbf{x}_s} \nabla^2 f(\omega, \mathbf{x}_s) G(\omega, \mathbf{x} - \mathbf{h}, \mathbf{x}_s; v(\mathbf{x}, \mathbf{x}_h)) \\ G(\omega, \mathbf{x} + \mathbf{h}, \mathbf{x}_r; v(\mathbf{x}, \mathbf{x}_h)) \Delta d^*(\omega, \mathbf{x}_r, \mathbf{x}_s; v(\mathbf{x}, \mathbf{x}_h)). \quad (10)$$

Therefore, the cost of computing the adjoint of EFWI can be written as

$$C_{\text{EFWI}-\text{A}} = N_x N_y N_z N_{\text{source}} (2 \times N_{tp} C_{\text{FDTD}} + 2 \times N_{tp} N_{hx} N_{hy} + N_{ti} N_{hx} N_{hy}) \quad (11)$$

and the total cost of one iteration of EFWI becomes

$$C_{\text{EFWI}} = N_x N_y N_z N_{\text{source}} (6 \times N_{tp} C_{\text{FDTD}} + 6 \times N_{tp} N_{hx} N_{hy} + N_{ti} N_{hx} N_{hy}). \quad (12)$$

We can see that the computational cost becomes extremely high when we include the subsurface offsets in the velocity model. One way to reduce the cost is presented in Biondi (2012) where the velocity model is extended over time instead of horizontal offset. In that case, the cost becomes a function of one time-lag parameter instead of two horizontal lags in 3D:

$$C_{\text{TimeEFWI}} = N_x N_y N_z N_{\text{source}} (6 \times N_{tp} C_{\text{FDTD}} + 6 \times N_{tp} N_\tau + N_{ti} N_\tau), \quad (13)$$

where N_τ is the number of time lags. The computational disadvantage is that several time slices need to be held in memory for each time instead of the conventional two slices.

Linearized Model-Space Extensions

Any of the previously mentioned inversions can be linearized by scale separation as shown in Almomin and Biondi (2012). To linearize the conventional FWI, the velocity model is separated as:

$$s^2(\mathbf{x}) = b(\mathbf{x}) + r(\mathbf{x}), \quad (14)$$

where $b(\mathbf{x})$ is the background component and $r(\mathbf{x})$ is the perturbation component. The linearized FWI forward operator can be written as follows:

$$d(\omega, \mathbf{x}_r, \mathbf{x}_s; b(\mathbf{x}), r(\mathbf{x})) = \sum_{\mathbf{x}} \omega^2 f(\omega, \mathbf{x}_s) G(\omega, \mathbf{x}, \mathbf{x}_r; b(\mathbf{x})) r(\mathbf{x}) G(\omega, \mathbf{x}, \mathbf{x}_s; b(\mathbf{x})). \quad (15)$$

As shown in equation 15, the cost of the linearized forward operator is equal to the adjoint cost of conventional FWI. Also, the adjoint of the linearized operator with respect to the perturbation is the same as the conventional FWI adjoint and has the same cost as well. The adjoint with respect to the background is:

$$\Delta b(\mathbf{x}) = \sum_{\omega, \mathbf{x}_r, \mathbf{x}_s, \mathbf{y}} \omega^4 f(\omega, \mathbf{x}_s) G(\omega, \mathbf{y}, \mathbf{x}_s; b(\mathbf{x})) r(\mathbf{y}) \\ G(\omega, \mathbf{x}, \mathbf{y}; b(\mathbf{x})) G(\omega, \mathbf{x}, \mathbf{x}_r; b(\mathbf{x})) \Delta d^*(\omega, \mathbf{x}_r, \mathbf{x}_s; b(\mathbf{x}), r(\mathbf{x})) + \\ \omega^4 f(\omega, \mathbf{x}_s) G(\omega, \mathbf{x}, \mathbf{x}_s; b(\mathbf{x})) G(\omega, \mathbf{x}, \mathbf{y}; b(\mathbf{x})) r(\mathbf{y}) \\ G(\omega, \mathbf{y}, \mathbf{x}_r; b(\mathbf{x})) \Delta d^*(\omega, \mathbf{x}_r, \mathbf{x}_s; b(\mathbf{x}), r(\mathbf{x})). \quad (16)$$

Although equation 16 has six Green's functions, only four propagations are required since each background wavefield is the same for two Green's functions. In addition, these background wavefields are the same for the adjoint of perturbation. Hence, the total cost of the linearized FWI per iteration, assuming complete reuse of background wavefields, is

$$C_{\text{LFWI}} = N_x N_y N_z N_{\text{source}}(12 \times N_{tp} C_{\text{FDTD}} + 12 \times N_{tp} + 9 \times N_{ti}). \tag{17}$$

This shows that scale separation by itself increases the cost of the original nonlinear operator since it adds several propagations, imaging and scattering steps. However, a significant cost cutting for extended inversions is possible by extending only the perturbation component without extending background component. By following the same derivation for linearized FWI, I find the cost of linearized extended FWI in subsurface offset when I extend perturbation only to be

$$C_{\text{LEFWI}} = N_x N_y N_z N_{\text{source}}(12 \times N_{tp} C_{\text{FDTD}} + 12 \times N_{tp} + 2 \times N_{ti} + 7 \times N_{ti} N_{hx} N_{hy}). \tag{18}$$

By extending only the perturbation, we remove the subsurface offset multiplication factor N_{tp}. This results in a large reduction in cost because the number of propagation time steps N_{tp} is much larger than the number of imaging time steps N_{ti}. Similarly, the cost of linearized extended FWI in time can be written as

$$C_{\text{TimeLEFWI}} = N_x N_y N_z N_{\text{source}}(12 \times N_{tp} C_{\text{FDTD}} + 12 \times N_{tp} + 2 \times N_{ti} + 7 \times N_{ti} N_{\tau}). \tag{19}$$

For the extended FWI in time, the cost reduction by linearization is less dramatic than the extended FWI in subsurface offset since there is only one additional convolution axis.

Data-Space Extensions

Model space extensions provide an accurate solution to the cycle skipping problem because they decompose the wavefields in the subsurface along the extended axes, either space or time. However, this is also reason for their high cost: each data point will interact with all points of the extended model. To avoid this problem, I propose extending full waveform inversion through a data space axis, such as source location, instead of model space axes. The reason for data space extension is that each extended model component operates on the corresponding component in data space and vice versa. In other words, each experiment can be computed similarly to the conventional way, but the model is changed between different experiments. However, extending the model through a data space axis has the underlying assumption that data components remain separated in the subsurface. This assumption depends on the complexity of the model. For instance, image space angle gathers and data space ray parameter gathers provide similar information in a fairly simple model (Sava and Fomel, 2003), but the latter breaks down in a very complicated model.

The extended inversion, whether in model space or data space, needs to satisfy two conditions. First, the observed data can always be explained by the extended

model regardless of the selection of the initial model (Biondi, 2012). Second, the extended model should allow gradual change by regularization to produce a non-extended model. The data space extensions that can potentially satisfy these conditions are source location and source ray parameter. The source location extension has the advantage of using the exact same propagation engine as the conventional inversion, so its implementation requires minimal adjustment to existing applications. Moreover, the source location satisfies the extension conditions fairly well. Figure 1 shows source location image gathers for a two-layer model when using the correct velocity, a 10 percent lower velocity, and a 10 percent higher velocity. We see that regularizing the additional axis by a derivative can satisfy the second condition. The disadvantage of this extension is having a velocity model for each source location, which can require a very large memory size and burdensome I/O in 3D.

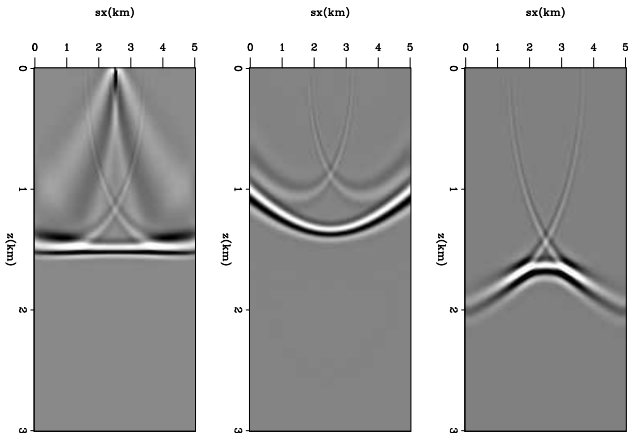


Figure 1: Source location image gather for a two-layer model when using the correct velocity (left), a 10 percent lower velocity (middle), and a 10 percent higher velocity (right). [**CR**]

Extending the model by source ray parameter requires plane-wave encoding of the data (Whitmore, 1995; Zhang et al., 2005; Liu et al., 2006). Figure 2 shows source ray parameter image gathers for a two-layer model when using the correct velocity, a 10 percent lower velocity, and a 10 percent higher velocity. Similar to source location, the source ray parameter seems to also satisfy the second condition of model extension. The first condition is tested in the Synthetic Examples section below. In addition, the number of planes is generally much smaller than the number of source locations, so it both reduces the cost and makes the size of the extended model very manageable.

The cost of source location encoding is the same as conventional FWI, whereas the cost of source ray parameter extension is

$$C_{\text{RayEFWI}} = N_x N_y N_z N_p (6 \times N_{tp} C_{\text{FDTD}} + 6 \times N_{tp} + N_{ti}), \tag{20}$$

where N_p is the number of planes. Figure 3 compares the costs of all mentioned inversions assuming $N_x = N_y = 1000$, $N_z = 100$, $N_{\text{source}} = 10000$, $C_{\text{FDTD}} = 16$, $N_{tp} = 1000$, $N_{ti} = 100$, $N_{hx} = N_{hy} = 100$, $N_\tau = 200$, and $N_p = 1000$, where the costs are normalized by the cost of conventional FWI. The log-scale highlights that the difference in cost between these inversions can be several orders of magnitude.

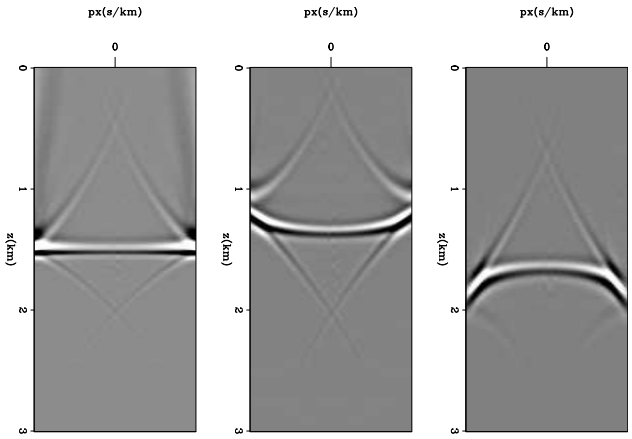


Figure 2: Source ray parameter image gather for a two-layer model when using the correct velocity (left), a 10 percent lower velocity (middle), and a 10 percent higher velocity (right). [**CR**]

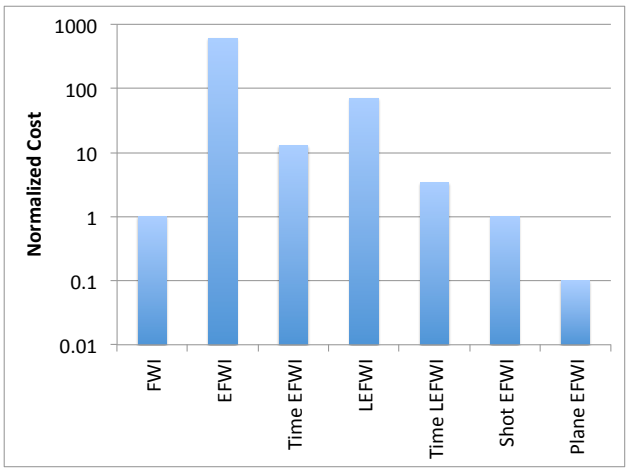


Figure 3: Cost comparison of conventional and extended full wavefrom inversions. [**NR**]

SYNTHETIC EXAMPLES

A modified Marmousi model is used for the synthetic example where a 500m water layer is added to the top. Figure 4 shows the true velocity model. A Ricker wavelet with a fundamental frequency of 15 Hz and temporal sampling of 1.5 ms is used as a source function to model the data. There are 461 fixed receivers with a spacing of 20 m and 21 plane sources with ray parameter ranging from -1 s/km to 1 s/km. The initial model shown in Figure 5 is obtained by strongly smoothing the true model.

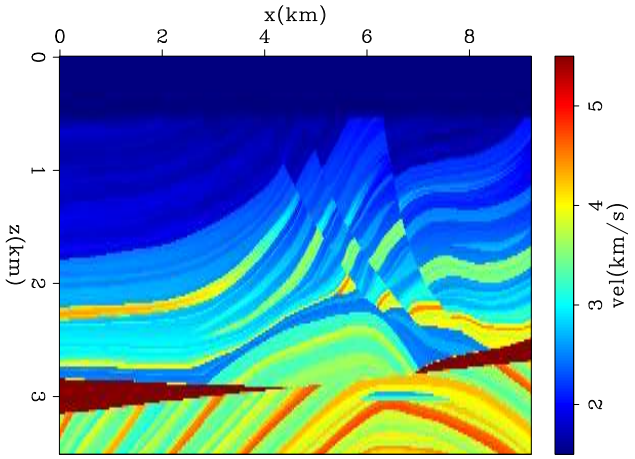


Figure 4: The true velocity of Marmousi model. [**ER**]

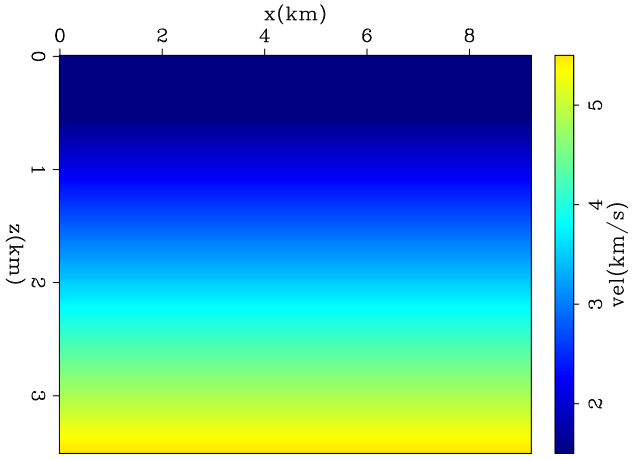


Figure 5: The initial velocity of Marmousi model. [**ER**]

I modeled the observed data with the conventional acoustic wave-equation nonlinear modeling operator. Then, I ran inversion without any tomographic regularization term. This inversion verified the first extension condition, that the modeled data can be explained using the initial model.

I show the results of running 1000 iterations of the unregularized inversion. Figure 6 shows the residual of the data fitting as a function of iterations. The residual decreases monotonically without getting stuck in a local minima and is approaching zero. This means that the first condition is fairly satisfied for this example. Figure 7 shows the difference between the initial and final model at three locations as a function of depth and ray parameter. The three gathers show varying degrees of

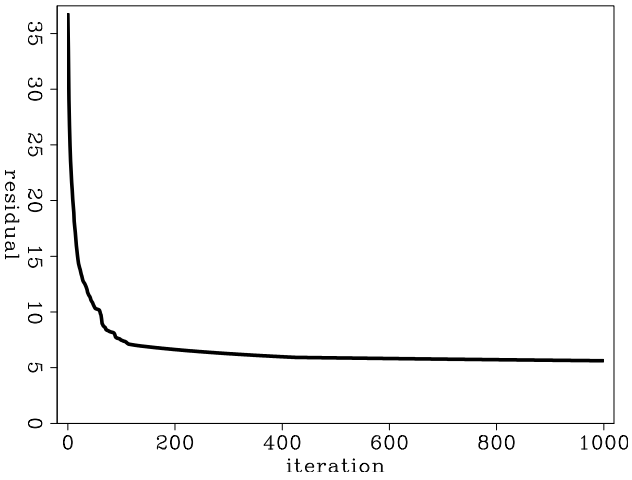


Figure 6: The data residual norm as a function of iterations of Marmousi model unregularized inversion. [**CR**]

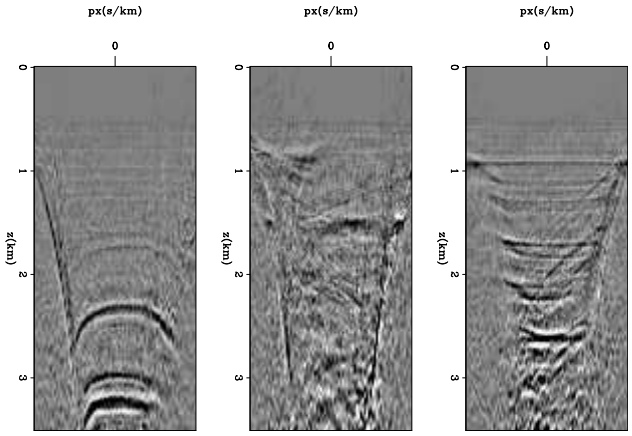


Figure 7: Three ray parameter gathers showing the difference between the initial model and the unregularized inverted model at x=2.5, 5, 7.5 km of Marmousi model. [**CR**]

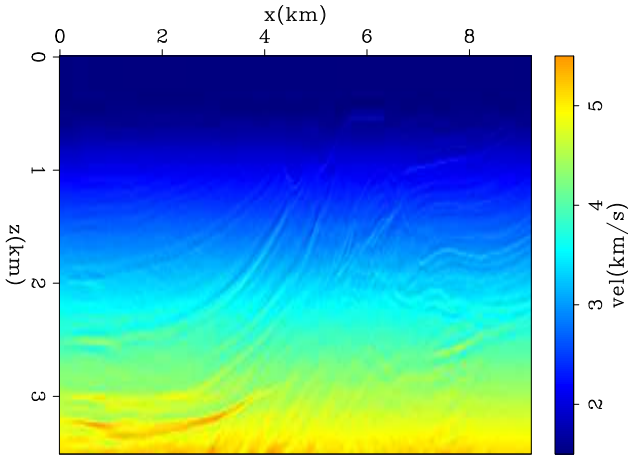


Figure 8: The unregularized inversion results of Marmousi model. [**CR**]

difference across the ray parameter axis, which indicates the inconsistency between these data components given the initial velocity model. Figure 8 shows the average of the final model across the ray parameter axis. Although the final model matches most of the observed data, no low wavenumber components are correctly estimated in this unregularized inversion since the kinematic errors are compensated for by the difference between ray parameter models.

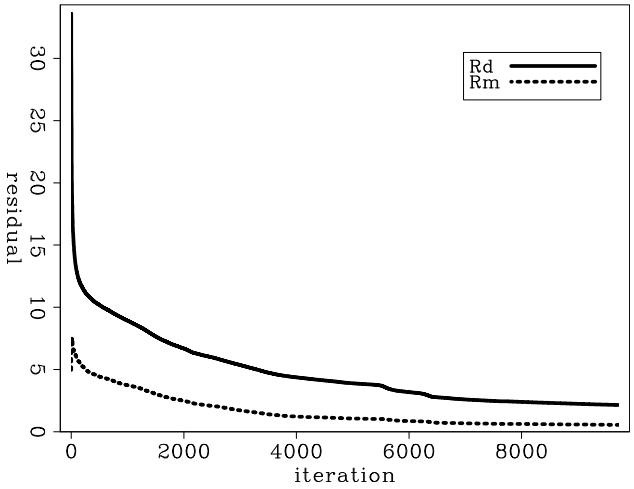


Figure 9: The data residual norm (solid line) and the model residual norm (dashed line) as a function of iterations of Marmousi model regularized inversion. [**CR**]

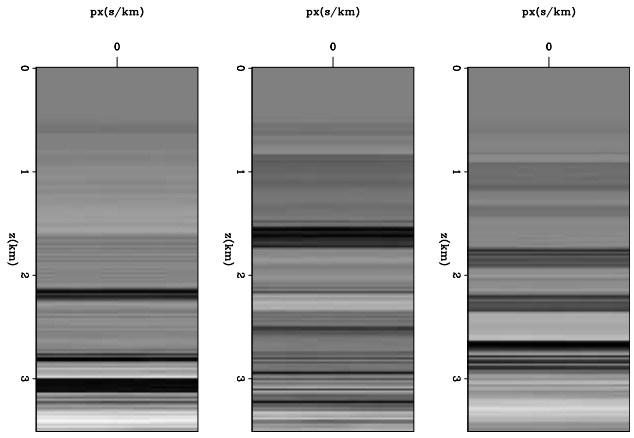


Figure 10: Three ray parameter gathers showing the difference between the initial model and the regularized inverted model at x=2.5, 5, 7.5 km of Marmousi model. [**CR**]

Next, I show the results of running 10000 iterations of the inversion regularized by a derivative across the ray parameter axis. Figure 9 shows the residual of the data fitting and model regularization as a function of iterations. The data fitting residuals decrease slower than the data residuals in the unregularized inversion since two objective functions are competing to be minimized. Nonetheless, it is not getting stuck in a local minima and is approaching zero. Figure 10 shows the difference between the initial and final model at three locations as a function of depth and ray parameter. The three images show no variations across the ray parameter axis, which indicates that the inversion successfully converged the extended model to a physical model. Figure 11 shows the average of the final model across the ray parameter axis. Both low and high wavenumber components are correctly estimated and the inversion converged towards the true answer.

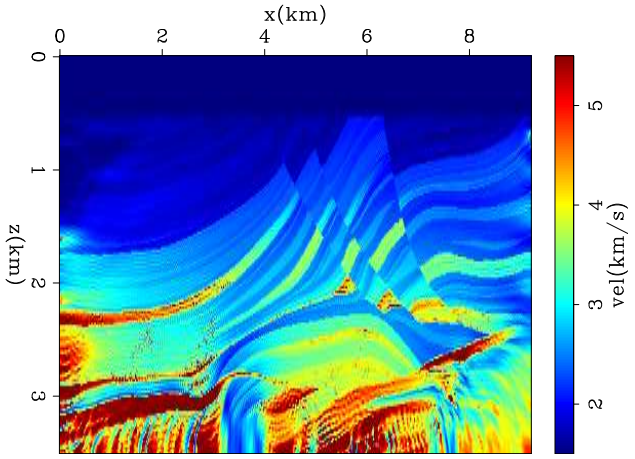


Figure 11: The regularized inversion results of Marmousi model. [**CR**]

CONCLUSIONS

I have shown that extending the velocity in model axes increases the computational cost drastically. Scale separation can reduce the cost but remains much more expensive than conventional FWI. I presented an alternative approach to extended FWI by using data space axes. Although the underlying assumptions might make it less accurate than model space extensions in very complex models, the cost is greatly reduced and becomes similar to the cost of conventional FWI. The synthetic Marmousi example showed remarkable results even when the initial model had large errors. The results of this model are comparable to the results of model space extensions but are significantly cheaper in cost.

ACKNOWLEDGMENTS

I would like to thank the Stanford Exploration Project affiliate companies for financial support. I also thank Saudi Aramco for supporting my Ph.D. studies at Stanford University.

REFERENCES

Almomin, A. and B. Biondi, 2012, Tomographic full waveform inversion: Practical and computationally feasible approach: SEP-Report, **147**, 13–26.

Biondi, B., 2012, Tomographic full waveform inversion and linear modeling of multiple scattering: SEP-Report, **148**, 49–58.

Biondi, B. and A. Almomin, 2012, Tomographic full waveform inversion (TFWI) by combining full waveform inversion with wave-equation migration velocity analysis: SEP-Report, **147**, 1–12.

Biondi, B. and P. Sava, 1999, Wave-equation migration velocity analysis: SEG Technical Program Expanded Abstracts, **18**, 1723–1726.

Liu, F., D. W. Hanson, N. D. Whitmore, R. S. Day, and R. H. Stolt, 2006, Toward a unified analysis for source plane-wave migration: Geophysics, **71**, no. 4, S129–S139.

Luo, Y. and G. T. Schuster, 1991, Wave-equation traveltime inversion: Geophysics, **56**, 645–653.

Marfurt, K. J., 1984, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations: Geophysics, **49**, 533–549.

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

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

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

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

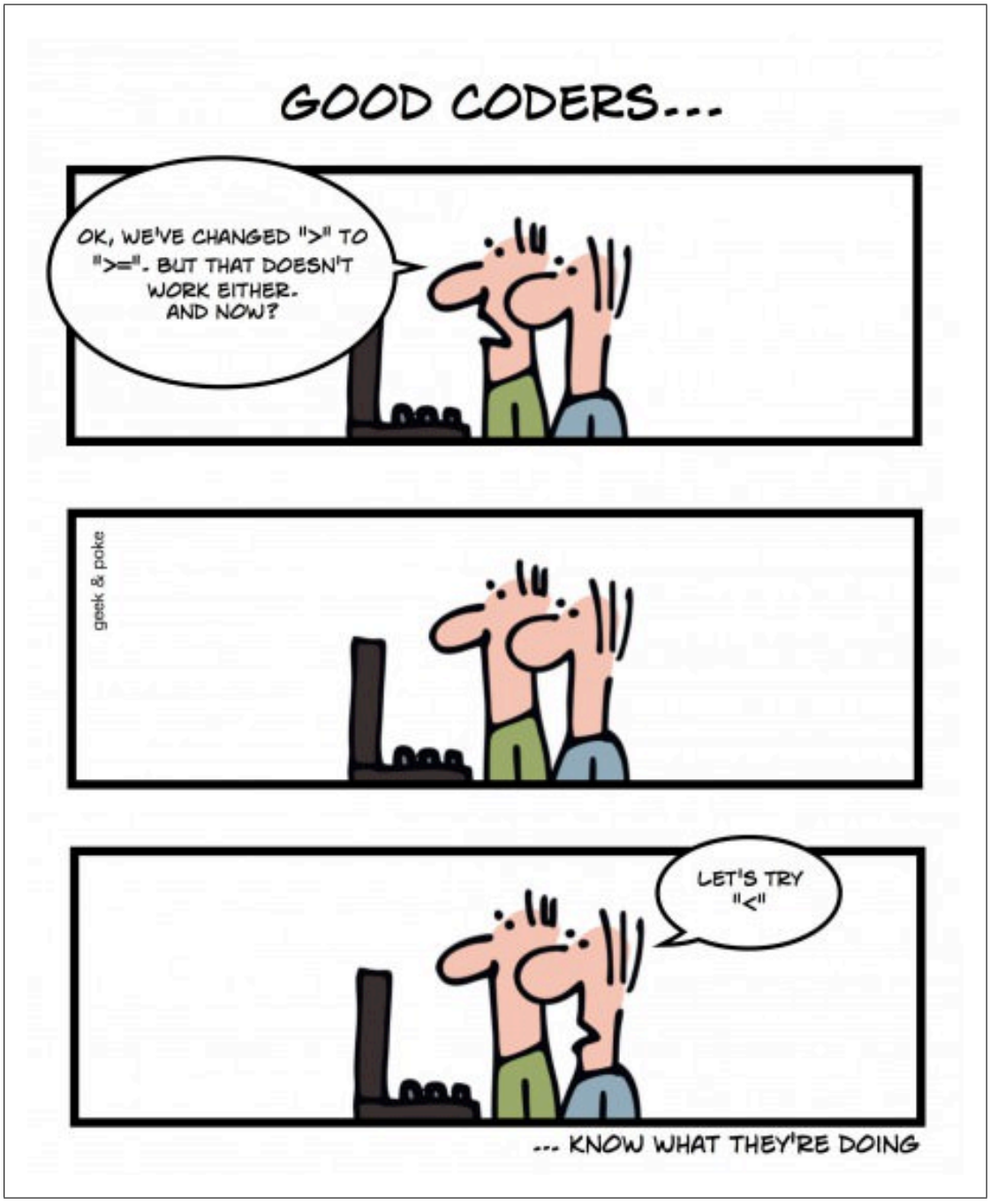
Tang, Y., C. Guerra, and B. Biondi, 2008, Image-space wave-equation tomography in the generalized source domain: SEP-Report, **136**, 1–22.

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

Whitmore, N. D., 1995, An imaging hierarchy for common angle plane wave seismogram: PhD thesis, University of Tulsa.

Zhang, Y., B. Biondi, and Y. Tang, 2012, Residual moveout-based wave-equation migration velocity analysis: SEG Technical Program Expanded Abstracts, **31**, *submitted for publication.*

Zhang, Y., J. Sun, C. Notfors, S. H. Gray, L. Chernis, and J. Young, 2005, Delayed-shot 3D depth migration: Geophysics, **70**, E21–E28.


GOOD CODERS...
OK, WE'VE CHANGED ">" TO ">=". BUT THAT DOESN'T WORK EITHER. AND NOW?
geek & poke
LET'S TRY "<"
... KNOW WHAT THEY'RE DOING