



Time-domain multiscale full waveform inversion using the rapid expansion method

Adriano W. G. dos Santos, Reynam C. Pestana, INCT-GP/CPGG-UFBA

Copyright 2013, SBGf - Sociedade Brasileira de Geofísica.

This paper was prepared for presentation at the 13th International Congress of the Brazilian Geophysical Society, held in Rio de Janeiro, Brazil, August 26-29, 2013.

Contents of this paper were reviewed by the Technical Committee of the 13th International Congress of The Brazilian Geophysical Society and do not necessarily represent any position of the SBGf, its officers or members. Electronic reproduction or storage of any part of this paper for commercial purposes without the written consent of The Brazilian Geophysical Society is prohibited.

Abstract

This paper presents a new time-domain multiscale full waveform inversion (FWI) using the rapid expansion method (REM) to propagate the source and residual wavefields. Through numerical results, we show the importance of a multiscale approach to the time-domain FWI in order to avoid cycle-skipped local minima. In addition, we show how the REM enables the propagation of the wavefields with large grid sampling intervals, free of numerical dispersion, thus reducing the computational cost when compared to the conventional finite difference (FD) method.

Introduction

The results obtained by full waveform inversion (FWI) in the last years confirmed the potential of this method to yield high accuracy models, even in areas with complex structures, such as sub-salt layers and near-vertical salt boundaries (Vigh et al., 2011).

FWI is defined as an optimization problem, in which the misfit between observed and calculated data should be minimized (Tarantola, 1984). In this process, a velocity or density model is updated iteratively, so the predicted data from the model can match the observed data. Therefore, FWI depends on two major steps: numerical simulation of wave propagation and solution of an inverse problem.

Due to high nonlinearity and nonuniqueness of solution, the main challenge of FWI is to avoid convergence to local minima, while keeping computational cost at acceptable levels. However, more accurate inversion algorithms, like Newton methods, demand more computational resources. One alternative for this dilemma is to reduce the cost of the modeling step, using, for example, GPUs (Graphic Processing Units) (Mao et al., 2012; Wang et al., 2011), pseudo-analytic methods (Ramos-Martinez et al., 2011) and super-gathers, like plane-waves (Vigh and Starr, 2008) or simultaneous sources (Anagaw and Sacchi, 2012).

A different alternative is to take FWI to another domain, like the Laplace-Fourier domain (Cha and Shin, 2009) and, principally, the frequency domain (Pratt, 1999). In these domains is easier to take advantage of an important property of FWI, that high-resolution imaging is expected at half the propagated wavelength (Virieux and Operto, 2009). So it is possible to naturally start the inversion from bigger wavelengths, i.e., lower frequencies and then use

the result as initial model for the higher frequencies, thus reducing the chance of convergence to local minima. One can even select a smaller number of optimal wavelengths (Boonyasirawat et al., 2009), reducing the computational cost.

On the other hand, wave propagation in the frequency domain usually involves inversion of the discretized Helmholtz operator, making it costly for 3D models. In this case, time-domain FWI is more feasible than frequency-domain (Vigh et al., 2009). However, full-bandwidth FWI in time domain is very affected by cycle-skipping artifacts (Virieux and Operto, 2009), which leads the convergence to spurious results.

In this paper, we show how it is possible to implement a time-domain FWI using a multiscale approach similar to the frequency-domain, i.e., from lower to higher frequencies. In addition, we show how the rapid expansion method (REM) (Pestana and Stoffa, 2010) allows the use of large grid sampling intervals, free of numerical dispersion, in contrast with the conventional finite difference (FD) method, thus, reducing computational cost of wave propagation at higher frequencies. Through numerical results using the Overthrust model and the Marmousi model we demonstrate the applicability and efficiency of our implementation.

Time-domain FWI

FWI is defined as an optimization problem that seeks to minimize the error E between an observed data \mathbf{d} and a simulated data, predicted from the model \mathbf{m} :

$$E(\mathbf{m}) = \frac{1}{2} \|\mathbf{d} - F(\mathbf{m})\|^2, \quad (1)$$

where $F(\cdot)$ is the modeling operator and $\|\cdot\|$ the L2-norm. In time domain, the modeled seismic data is generated by numerical solution of a wave equation, such as the acoustic wave equation

$$\frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} = v^2(\mathbf{x}) \nabla^2 p(\mathbf{x}, t) + s(\mathbf{x}, t), \quad (2)$$

with $p(\mathbf{x}, t)$ being the pressure wavefield, $v(\mathbf{x})$ the acoustic velocity (representing the model \mathbf{m}), ∇^2 the Laplacian operator and $s(\mathbf{x}, t)$ a source term. Other equations can be used, e.g., the anisotropic (Warner et al., 2013) and elastic (Podgornova and Charara, 2011) wave equations.

The misfit function (equation 1) is highly nonlinear (Boonyasirawat et al., 2009) and can be minimized using a Newton-like method. With this approach, the model \mathbf{m} is updated iteratively:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \delta \mathbf{m}_k. \quad (3)$$

The model update $\delta \mathbf{m}_k$ is given by (Ma, 2010)

$$\delta \mathbf{m}_k = -\mathbf{H}_k^{-1} \mathbf{g}_k, \quad (4)$$

where \mathbf{H} is the Hessian matrix and \mathbf{g} is the gradient $\frac{\partial E}{\partial \mathbf{m}}$, that can be obtained by the adjoint state method (Bunks et al., 1995). For the acoustic case:

$$\frac{\partial E}{\partial \mathbf{m}} = \frac{2}{v^3} \int_0^T \lambda \frac{\partial^2 p(\mathbf{x}, t)}{\partial t^2} dt, \quad (5)$$

where T is the maximum time recorded on dataset, p is the forward-propagated source wavefield and λ represents the back-propagated residual data, i.e., the residual given by: $F(\mathbf{m}) - d$, is back-propagated using equation 2.

The Hessian matrix, on the other hand, is usually neglected due to its large size and high computational demand. As an alternative, a new iterative method is defined as:

$$\mathbf{m}_{k+1} = \mathbf{m}_k - \alpha_k \mathbf{h}_k, \quad (6)$$

with α_k being a step size, computed by a line search algorithm (Nocedal and Wright, 2006), and \mathbf{h}_k the search direction, determined by gradient methods such as the steepest-descent and the conjugate gradient (Ma, 2010), or quasi-Newton methods, such as the L-BGFS (Deng et al., 2012).

Time-domain FWI using equation 6 and the adjoint state method can then be summarized in five steps:

1. $F(\mathbf{m}_k)$ estimation using forward propagation (equation 2);
2. Gradient (\mathbf{g}_k) computation using equation 5;
3. Determination of search direction \mathbf{h}_k ;
4. Step size (α_k) determination using a line search algorithm;
5. Model update using equation 6.

In this work, we used the steepest-descent method to determine the search direction, i.e., $\mathbf{h}_k = -\mathbf{g}_k$. The step size α was first computed using the adaptive Barzilai-Borwein method, proposed by Zhou et al. (2006) and then a backtracking line search (Nocedal and Wright, 2006) was used to find a step size capable of reduce the misfit.

To test our implementation, we used the Overthrust velocity model (Figure 1). It has 801x187 samples with an interval of 25m in both directions, i.e., $\Delta x = \Delta z = 25m$. We modeled 115 shots using the finite difference scheme with 2nd and 8th order operators in time and space, respectively, with maximum offset of 5km and minimum of -5km. A Ricker wavelet with maximum frequency of 30Hz was used as source.

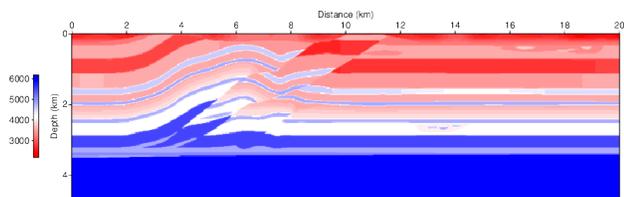


Figure 1: Overthrust true velocity model.

The initial model used for FWI is showed in Figure 2. Usually, lower-resolution velocity analysis methods, like

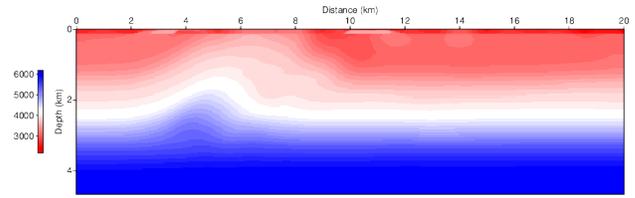


Figure 2: Initial model for FWI.

travel-time tomography and residual migration, are used to estimate a good starting model for FWI (Ma, 2010).

Figure 3 shows the result of time-domain FWI using the full-bandwidth of 30Hz after 30 iterations. The main reflectors were outlined, but the velocities are incorrect. As was said before, this kind of inversion is highly affected by cycle-skipping artifacts, that leads to a local minimum. On the other hand, the lower frequencies are less affected by those artifacts (Virieux and Operto, 2009). Therefore, a multiscale approach is really important to increase the chance of success of FWI, especially when the initial model is far from the true model.

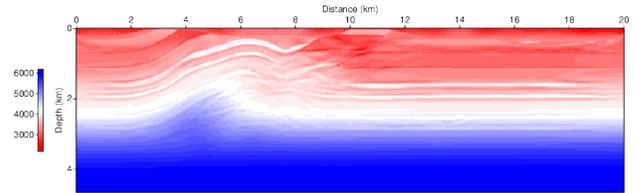


Figure 3: Inverted model with full-bandwidth.

Time-domain multiscale FWI

The time-domain multiscale FWI uses the same principle of the frequency-domain FWI: that large scales, i.e., lower frequencies, have a better chance to converge to the global minimum at that scale (Bunks et al., 1995). The inversion of large scales can then be used as initial model to shorter scales, reducing the risk of convergence to a local minimum.

In the frequency domain, this can be achieved more naturally, because the data is already decomposed in different wavelengths. However, in time domain, one must limit the frequency bandwidth of the observed data, as well as the source wavelet used in the modeling, so that the scale of FWI is consistent with the current model.

The limitation of the frequency bandwidth is done by filtering the observed data and the source wavelet. There are a wide range of filters to choose from. Boonyasiriwat et al. (2009) show how the Wiener low-pass filter can obtain an excellent result when the source wavelet is well estimated. This filter is applied in the frequency domain. It uses the original wavelet and a target wavelet, with the same maximum frequency as the desired filtered data:

$$f_{Wiener}(\omega) = \frac{W_{target}(\omega)W_{original}^{\dagger}(\omega)}{|W_{original}(\omega)|^2 + \varepsilon^2}, \quad (7)$$

where W represents the wavelet, \dagger denotes the complex conjugate and ε^2 is a small number to avoid division by zero.

We tested the time-domain multiscale FWI on the Overthrust model, using the same starting model (Figure 2). The maximum frequency of each multiscale iteration was 6Hz, 10Hz, 14Hz, 20Hz and 30Hz. We ran 30 FWI iterations for each frequency and the inverted velocity model was used as initial model for the next frequency. To filter the data bandwidth, we used the Wiener filter on equation 7.

Figure 4(a-e) show the result of the inversion for the frequencies 6Hz, 10Hz, 14Hz, 20Hz and 30Hz, respectively. It is clear that the multiscale approach achieved a much better inverted model than the one shown in Figure 3. One should notice that the lower frequencies were responsible to recover the most significant structures, while the higher frequencies yield the high-resolution details, characteristic of FWI.

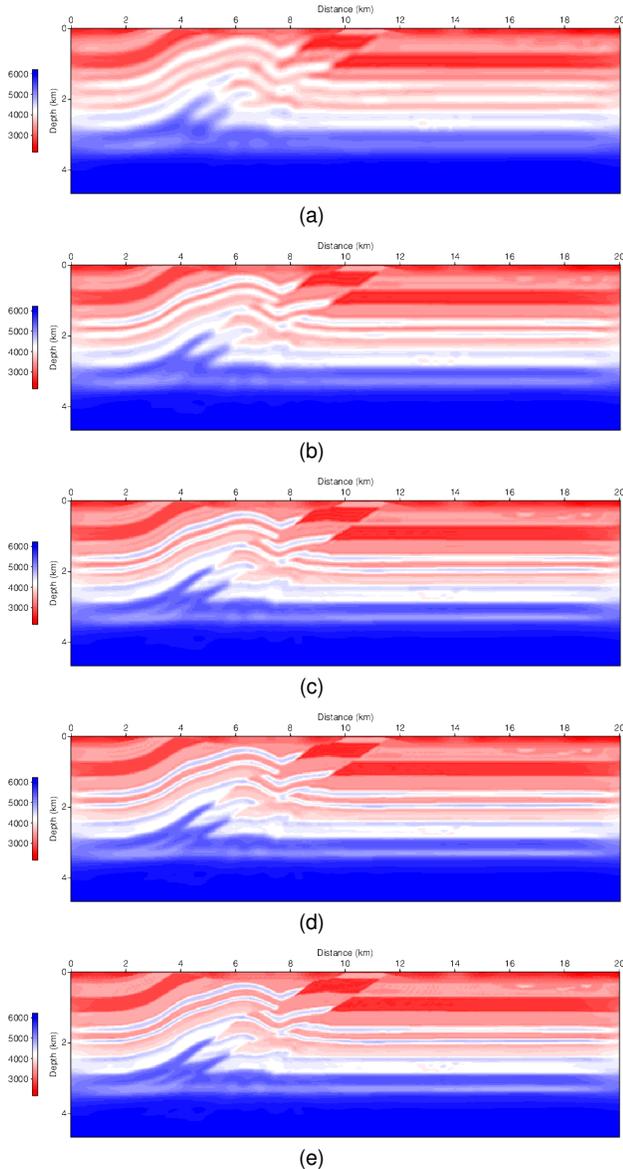


Figure 4: Results of FD time-domain multiscale FWI using: (a) 6Hz, (b) 10Hz, (c) 14Hz, (d) 20Hz and (e) 30Hz.

Figure 5 shows the vertical velocity profiles in the horizontal position of 11.25km of the models in Figures 1, 2, 3 and

4(e). The multiscale inversion was able to match the true velocity model with great precision.

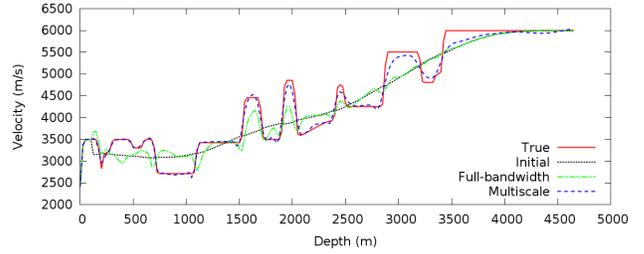


Figure 5: Comparison of a velocity profile at the horizontal position equal to 11.25km: true model (red solid line), initial model (black dotted line), full-bandwidth FWI (green dashed-dotted line) and multiscale FWI (blue dashed line).

In Figure 6, we show the angle-domain common image gathers (ADCIG), obtained during a reverse time migration (RTM) using the following velocity models: (a) initial, (b) inverted with full-bandwidth FWI, (c) inverted with multiscale FWI and (d) true. The initial and full-bandwidth models were not able to completely flatten the events, while the multiscale FWI succeeded.

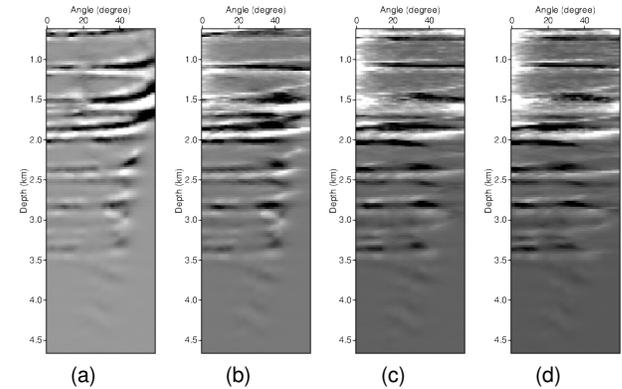


Figure 6: Angle-domain common image gathers at the position 16.25 km obtained with RTM using: (a) initial model, (b) full-bandwidth FWI model (c) multiscale FWI model and (d) true model.

It is important to add that for real datasets, low frequencies and long offsets are not always available. This can cause poor resolution in deeper regions. Recently, Fei et al. (2012) proposed a method to create pseudo low-frequencies to mitigate this limitation.

FWI using the rapid expansion method

As previously stated, simulation of the wave propagation is one of the main challenges of FWI, and the most time consuming. Although the conventional 2nd order in time FD scheme is capable of solving the acoustic wave equation with good precision, its use is constrained by stability and dispersion conditions, given (for the 2D case) respectively by (Bunks et al., 1995):

$$\max(\Delta x, \Delta z) > \sqrt{2} \Delta t \max(v), \quad (8)$$

where Δt is the time interval and Δx and Δz are the space interval; and

$$\max(\Delta x, \Delta z) > \frac{\min(v)}{G f_{\max}}, \quad (9)$$

with f_{\max} being the maximum frequency and G a factor that depends on the order of the FD space operator.

On the other hand, the rapid expansion method (REM) (Pestana and Stoffa, 2010), when combined with the Fourier transform to compute the space derivatives, can march the wavefield with any time step and free of dispersion. This is specially interesting when higher frequencies are used in time-domain FWI. In this case, the FD method needs to use a small grid sampling, to avoid numerical dispersion, thus, increasing the computational cost.

The REM uses the exact solution of wave equation:

$$p(\mathbf{x}, t + \Delta t) = -p(\mathbf{x}, t - \Delta t) + 2 \cos(L\Delta t) p(\mathbf{x}, t), \quad (10)$$

with $-L^2 = v^2(\mathbf{x}) \nabla^2$, for the acoustic case. The cosine is expanded using the method proposed by Tal-Ezer et al. (1987):

$$\cos(L\Delta t) = \sum_{k=0}^{\infty} C_{2k} J_{2k}(\Delta t R) Q_{2k} \left(\frac{iL}{R} \right), \quad (11)$$

where $C_{2k} = 1$ for $k = 0$ and $C_{2k} = 2$ for $k > 0$. For 2D case, $R \approx v_{\max} \pi \sqrt{(1/\Delta x^2) + (1/\Delta z^2)}$. J_{2k} is the Bessel function of order $2k$, and Q_{2k} are the Chebyshev polynomials.

The summation on equation (11) converges exponentially when $k > \Delta t R$. Therefore, the cosine approximation is obtained with good precision for k values greater than $\Delta t R$, enabling the use of any time step without stability issues. In addition, R is inversely proportional to the grid sampling, which means that with bigger grid intervals, a smaller number of terms is required, reducing even further the computational cost.

Another benefit of the REM is that it can reproduce an analytic signal with much better precision than the FD, as showed by (Tessmer, 2011). This can be important in inversion of real datasets, when FD approximations may fail to reproduce the true waveform.

Numerical results

To test the effectiveness of REM to reduce the computational cost of FWI, we selected the Marmousi velocity model, resized to 767x250 samples, with interval of 12m in each direction. The true velocity model is showed in Figure 7.

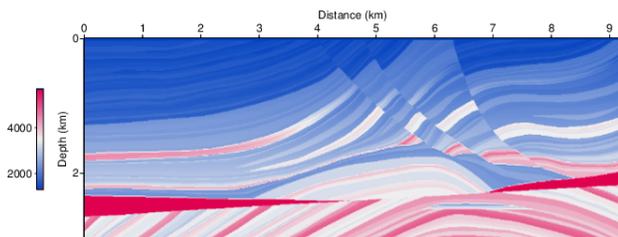


Figure 7: Marmousi true velocity model.

From this model we created two datasets. One using the REM, with a time sampling of 4ms; and another using 8th order FD with time sampling of 1ms, as required by the stability condition 8. The two datasets have the same geometry, with 119 shots and at maximum 250 receivers.

Both were generated using a Ricker wavelet with maximum frequency of 40Hz.

For the inversion, we used a reduced model with 383x125 samples ($\Delta x = \Delta z = 24m$). The starting model, showed in Figure 8, is the strongly smoothed true model.

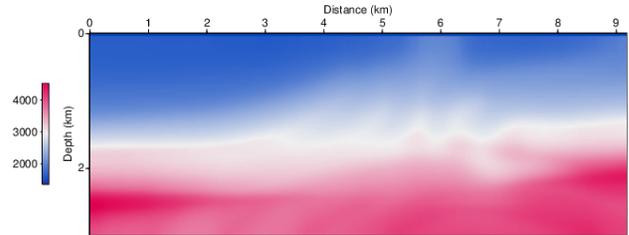


Figure 8: Marmousi initial velocity model.

The selected frequencies for the multiscale FWI were 6Hz, 12Hz, 24Hz and 40Hz. The FWI results were obtained with 30 iterations at each frequency. The simulated and observed data were always computed with the same method, to avoid errors from the incompatibility between FD and REM.

Figure 9 shows the FWI result using REM for maximum frequency equal to 40Hz. In general, the improvement is remarkable, though the deeper layers could have higher resolution if a better selection of frequencies or a more suitable acquisition geometry were made. However, the main purpose of this test was to measure the computational cost of the methods. The rapid expansion method took on average 29.56s to propagate the wavefield during 6s with $\Delta t = 4ms$, which results in 1500 time steps.

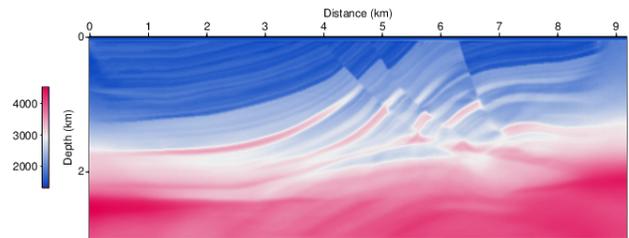


Figure 9: Marmousi inverted model using REM FWI.

On the FWI using finite differences, the numerical dispersion problem arises. The lower frequencies 6Hz and 12Hz are free of this issue because they satisfy the dispersion condition (equation 9). However, using a 4th order operator ($G = 5$), the 24Hz and 40Hz frequencies require grid sampling of 12.5m and 7.5m, respectively. This means an increase of up to ten times in the computational cost, when compared to the 24m grid.

To prove the necessity of avoiding numerical dispersion, we ran the 4th order FD FWI using the 24m grid, ignoring the sampling requirements. Figure 10 shows the result of inversion, damaged by numerical dispersion, as can be seen more clearly in Figure 11.

With the 4th order FD operator, the forward wavefield propagation took on average 46.37s on a 24m sampled grid. The time sampling is $\Delta t = 1ms$, which results in 6000 time steps. In this case, the REM was 1.57 times faster and, even better, free of dispersion.

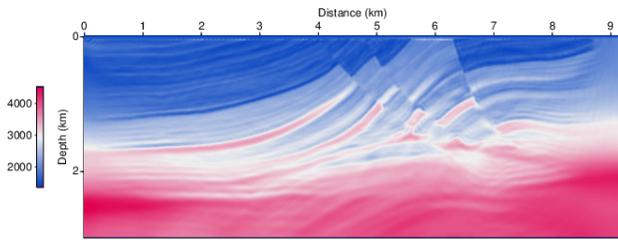


Figure 10: Marmousi inverted model using 4th order FD FWI, ignoring the dispersion condition.

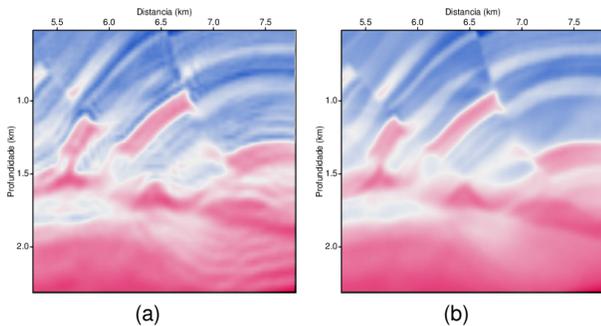


Figure 11: Velocity model highlight showing the effects of dispersion on 4th order FD FWI (a), compared with REM FWI (b).

Additionally, we simulated the forward wave propagation using 4th order FD on $12m$ and $7m$ sampled grids, to satisfy the dispersion condition for frequencies of $24Hz$ and $40Hz$, respectively. With $\Delta x = \Delta z = 7m$, the time sampling used was $\Delta t = 0.8ms$, to satisfy the stability condition. The modeling using the $12m$ sampled grid took, on average, $134.69s$, while on the $7m$ sampled grid, it took $400.57s$, i.e., for the full-bandwidth data ($40Hz$), the REM was 13.55 times faster, using a more sparse grid ($\Delta x = \Delta z = 24m$).

We pushed the limit of REM even further, using a grid spacing of $48m$. The result of this FWI is shown in Figure 12. Despite the natural loss of resolution due to resampling, the inverted model recovered the main features of the true model. The modeling took on average $8.94s$, 44.81 times faster than the worst case scenario FD ($7m$ sampled grid). With this result, we prove that the REM can make time-domain multiscale FWI more flexible, i.e., we can use small grid spacing for higher resolution and large grid spacing to speed up inversion.

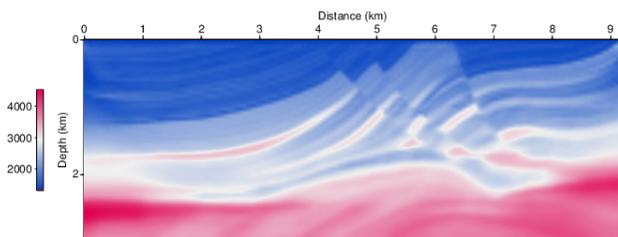


Figure 12: Marmousi inverted model using REM FWI with grid spacing of $48m$.

Although a more detailed study should be done to allow a fairer comparison with the FD method, it is clear that the ability of REM to propagate the wavefield free

of numerical dispersion and stability issues, can greatly reduce the computational cost of time-domain multiscale FWI, specially when high frequencies are used.

Conclusions

In this paper, we showed the importance of a multiscale approach to the time-domain full waveform inversion, in order to avoid convergence to local minima.

In addition, we presented a new FWI using the rapid expansion method to propagate the wavefields in a stable manner and free of numerical dispersion, avoiding time and space resampling, as done in the conventional finite difference method. This can greatly reduce the computational cost of time-domain FWI, specially at higher frequencies, making it more attractive over the frequency-domain FWI.

Acknowledgements

This research was supported by CNPq and INCT-GP/CNPq. The facility support from CPGG/UFBA is also acknowledged.

References

- Anagaw, A. Y. and M. D. Sacchi, 2012, Full waveform inversion with simultaneous sources using the full newton method, chapter 531, 1–5.
- Boonyasiriwat, C., P. Valasek, P. Routh, W. Cao, G. Schuster, and B. Macy, 2009, An efficient multiscale method for time-domain waveform tomography: *GEOPHYSICS*, **74**, WCC59–WCC68.
- Bunks, C., F. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: *GEOPHYSICS*, **60**, 1457–1473.
- Cha, Y. H. and C. Shin, 2009, 2d laplace-fourier-domain full waveform inversion for both velocity and density models: An experience of the 2004 bp velocity-analysis benchmark dataset, chapter 454, 2258–2262.
- Deng, W., L. Han, B. Zhang, F. Huang, and M. Han, 2012, Study of full waveform inversion based on l-bfgs algorithm: *Global Geology*, **15**, 161–165.
- Fei, T. W., Y. Luo, F. Qin, and P. G. Kelamis, 2012, Full waveform inversion without low frequencies: A synthetic study, chapter 484, 1–5.
- Ma, Y., 2010, Full waveform inversion with image-guided gradient: Master's thesis, Colorado School of Mines.
- Mao, J., R.-S. Wu, and B. Wang, 2012, Multiscale full waveform inversion using gpu, chapter 496, 1–7.
- Nocedal, J. and S. Wright, 2006, Numerical optimization, 2. ed. ed. Springer series in operations research and financial engineering: Springer.
- Pestana, R. C. and P. L. Stoffa, 2010, Time evolution of the wave equation using rapid expansion method: *Geophysics*, **75**, 121–131.
- Podgornova, O. and M. Charara, 2011, Multiscale time-domain full-waveform inversion for anisotropic elastic media, chapter 483, 2459–2464.
- Pratt, R., 1999, Seismic waveform inversion in the frequency domain, part 1: Theory and verification in a physical scale model: *GEOPHYSICS*, **64**, 888–901.
- Ramos-Martinez, J., S. Crawley, S. Kelly, and B. Tsimelzon, 2011, Full-waveform inversion by pseudo-analytic extrapolation, chapter 526, 2684–2688.
- Tal-Ezer, H., D. Kosloff, and Z. Koren, 1987, An accurate

- scheme for forward seismic modeling: *Geophysical Prospecting*, **35**, 479–490.
- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *GEOPHYSICS*, **49**, 1259–1266.
- Tessmer, E., 2011, Using the rapid expansion method for accurate time-stepping in modeling and reverse-time migration: *Geophysics*, **76**, S177–S185.
- Vigh, D., J. Kapoor, and H. Li, 2011, Full-waveform inversion application in different geological settings, chapter 467, 2374–2378.
- Vigh, D. and E. Starr, 2008, 3d prestack plane-wave, full-waveform inversion: *GEOPHYSICS*, **73**, VE135–VE144.
- Vigh, D., W. Starr, and K. K. Dingwall, 2009, 3d prestack time domain full waveform inversion.
- Virieux, J. and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *GEOPHYSICS*, **74**, WCC1–WCC26.
- Wang, B., J. Gao, H. Zhang, and W. Zhao, 2011, Cuda-based acceleration of full waveform inversion on gpu, chapter 496, 2528–2533.
- Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Åtekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3d full-waveform inversion: *GEOPHYSICS*, **78**, R59–R80.
- Zhou, B., L. Gao, and Y.-H. Dai, 2006, Gradient methods with adaptive step-sizes: *Comput. Optim. Appl.*, **35**, 69–86.