King Abdullah University of Science and Technology

Earth Science and Engineering

Multi-scale of Full Waveform Inversion of Acoustic Approximation

Author:
Fuqiang Chen

Advisor:
Prof. Gerard Schuster

December 12, 2015

Multi-scale Full Waveform Inversion of Acoustic Approximation

Fuqiang Chen, KAUST*

SUMMARY

In this report, the theoretical aspects and practical implementation of full waveform inversion (FWI) are derived and considered explicitly. Which makes FWI can be easily understood also easily understood how to implement. Many references treat FWI as an integral or the procedure transform into frequency domain. Which is more or less obscure than the derivation in this paper. Our derivation is easy to understand and explicitly describe the adjoint operator of Green's function, reverse propagation and some other terminologies for FWI. Then we implement the FWI in 1D medium. Because the scheme of finite difference (FD) is 2-order in time and 2-order in space, so the large time interval should be chosen, thus results in the narrow of frequency range. For seismic record, which may be dominated by low frequency, therefore, directly applying multi-scale do not improve the result too much. After many iterations, the residual data becomes much more enrich of being high frequency. Then multi-scale scheme can be employed to smooth the residual data. Numerical results show that the residual converges to small value while the velocity model still change subtle.

INTRODUCTION

FWI and FWI with the combination with multi-scale scheme can be found in many publications. To save space, these references do not contain in this part.

THEORY

In this part, the explicit matrix and vector multiplications are used to derive FWI. Forward problems can be stated using

$$p(\mathbf{x}_r,t) = \int_V G(\mathbf{x}_r,t;\mathbf{x}_s) * s(\mathbf{x}_s,t)\, d\mathbf{x}_s \qquad (1)$$

Equation (1) is an elegant expression. Which can be re-written using matrix notation

$$\underbrace{\begin{pmatrix} g_1 & & & & & \\ g_2 & g_1 & & & & \\ g_3 & g_2 & g_1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ g_{n-1} & g_{n-2} & g_{n-3} & \cdots & g_1 & \\ g_n & g_{n-1} & g_{n-2} & \cdots & g_2 & g_1 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_{n-1} \\ s_n \end{pmatrix}}_{\mathbf{Gs}} \qquad (2)$$

By re-writing the forward problem in equation 1, the continuous solution can be represented by discrete form thus for every fixed space point, equation 2 implies the solution to wave equation. By combining the data residual, equation 2 can be used to express the misfit function.

$$\underbrace{\begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \vdots \\ \delta_{n-1} \\ \delta_n \end{pmatrix}^T \begin{pmatrix} g_1 & & & & & \\ g_2 & g_1 & & & & \\ g_3 & g_2 & g_1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ g_{n-1} & g_{n-2} & g_{n-3} & \cdots & g_1 & \\ g_n & g_{n-1} & g_{n-2} & \cdots & g_2 & g_1 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_{n-1} \\ s_n \end{pmatrix}}_{\mathbf{r}^T\mathbf{Gs}} \qquad (3)$$

Equation 3 is the discrete expression of misfit function using inner product. Another equivalent version is

$$\underbrace{\begin{pmatrix} s_1 \\ s_2 \\ s_3 \\ \vdots \\ s_{n-1} \\ s_n \end{pmatrix}^T \begin{pmatrix} g_1 & g_2 & g_3 & \cdots & g_{n-1} & g_n \\ & g_1 & g_2 & \cdots & g_{n-2} & g_{n-1} \\ & & g_1 & \cdots & g_{n-3} & g_{n-2} \\ & & & \ddots & \cdots & \cdots \\ & & & & g_1 & g_2 \\ & & & & & g_1 \end{pmatrix} \begin{pmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \vdots \\ \delta_{n-1} \\ \delta_n \end{pmatrix}}_{\mathbf{s}^T\mathbf{G}^T\mathbf{r}} \qquad (4)$$

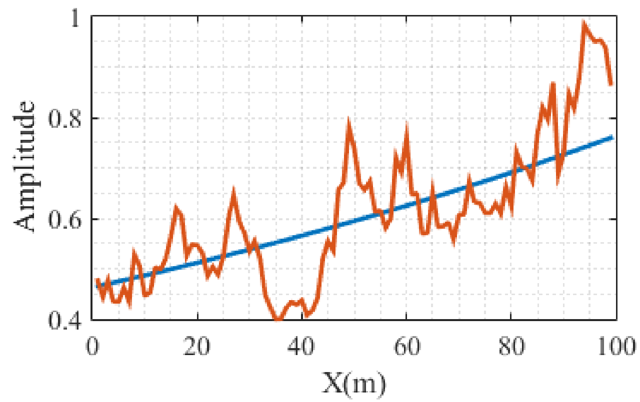
In equation 4, transposed Green's function matrix behaves like a cross correlation process instead of being a convolution. To calculate two latter terms, we can flip the residual vector and transpose the matrix induced by Green's function, thus convolution process is obtained with a reversed time sequence of residual waveform.

$$\underbrace{\begin{pmatrix} g_1 & & & & & \\ g_2 & g_1 & & & & \\ g_3 & g_2 & g_1 & & & \\ \vdots & \vdots & \vdots & \ddots & & \\ g_{n-1} & g_{n-2} & g_{n-3} & \cdots & g_1 & \\ g_n & g_{n-1} & g_{n-2} & \cdots & g_2 & g_1 \end{pmatrix} \begin{pmatrix} \delta_n \\ \delta_{n-1} \\ \delta_{n-2} \\ \vdots \\ \delta_2 \\ \delta_1 \end{pmatrix}}_{\mathbf{Gs}'} \qquad (5)$$

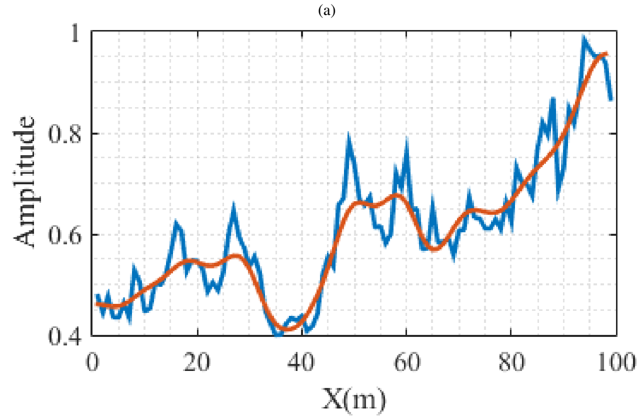
where prime denotes the reversed residual data. Another thing is even though we re-write a part of equation 4 by equation 5, the gap is that misfit function in equation 4 is the inner product time by time from the first time to the end. But the wave-field obtained from equation 5 by forward modeling is reversed, in other words, at the first time, we get the value of wave-field at time nt, thus it cannot be used to calculate inner product (Because now we do not know the wave-field value of another forward modeling using virtual source as excitation unless wave-field of all time is saved). A smart scheme has been used to overcome this disagreement. First we forward propagates the wave-field excited by virtual source, then save the boundary and the wave-field of last two time steps (how many time steps are saved depends on which the scheme of time difference is used, for one-step extrapolation we just need to wave one step.) Then we start to backward propagate this saved wave-field at the same time do the ordinary forward modeling using flipped residual data as excitation then sum them over all of time steps.

NUMERICAL TESTS

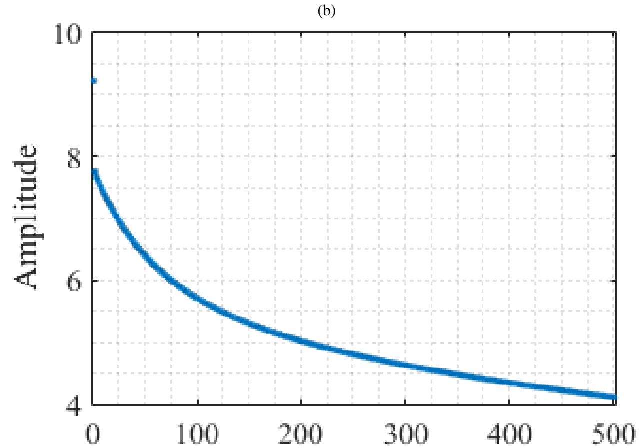
In this part, first we implement the FWI to understand it in detail. the results are as follows. Sorry about errata of ticks and unit. In this project, the space interval is 18m, and label along x axis should be sampling points not the real distance to some reference point. In figure 1, the initial model is chosen so that it closes to true model, which means initial model includes all of the component of low wavenumber, so that we can use gradient-based algorithm to find the minimum. Because the misfit function of FWI is highly non-linear, if the initial model far from the true model in the sense of some norm, then the process of iteration would stuck in some local minimum very fast. Figure 2 is the inverted model (orange line) overlaid on true model (blue line), which shows inverted model approximates the true model much better than the initial model does after 501 iterations. Because observed data is dominated by low frequency and narrow frequency range as illustrated by figure 4, if directly applying multi scale scheme, it seems that does not work very well. However, after 501 iterations, figure 5 shows that the residual data contains high frequency components much more than observed data. So we can smooth the residual data to continue the inversion process. The numerical examples show that misfit function becomes less if the multi scale scheme is employed even though the velocity model hardly changes. As stated before, in this project, if direct applying multi scale scheme, the results does not improve so much. Our explanation to this is because the frequency range is too narrow.



(a)



(b)



(c)

Figure 1: a, True model (orange) overlaid on initial model (blue). b, True model (blue) overlaid by inverted model (orange). c, convergence curve.

CONCLUSION

In this report, we present a much easier way to understand FWI and show how to implement it in explicit way. Then we implement the FWI in 1D medium. FD with 2-order in time axis and 2-order in x axis which requires large time sampling interval thus results in the data dominated by low frequency. However, After iteration of some steps, the residual data contains high frequency components much more than observed data. So we can smooth the residual data to continue the inversion process. In addition, in this project, we also check the reliability of component of high and low frequency of data respectively. Inverted results from data dominated by high frequency shows higher apparent resolution then low frequency does. But which resolution causes ambiguous and, maybe need to be smoothed out for model use at the further processing flows. In other words, results from high frequency is less reliable than the results form low frequency. In addition, data dominated by low frequency converges much faster than data dominated by high frequency does.

ACKNOWLEDGEMENT

First of all, I express my thank to Professor Gerard Schuster for his knowledge, patience, enthusiasm and kind. In addition, special thanks to WWW, which makes learning become more convenient, more effective and more equitable.

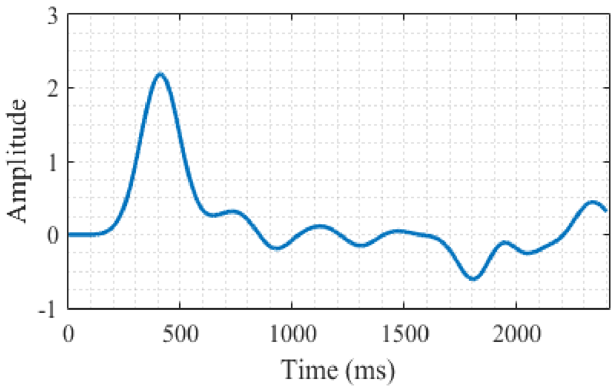


Figure 2: Synthetic data as observed data to retrieve the velocity model. Which is dominated by low frequency because of sampling interval along time axis is large.

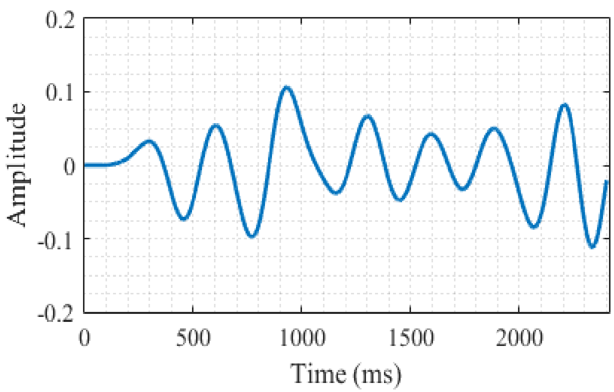


Figure 3: Residual after 501 iterations. Which is input to multi scale scheme of FWI.

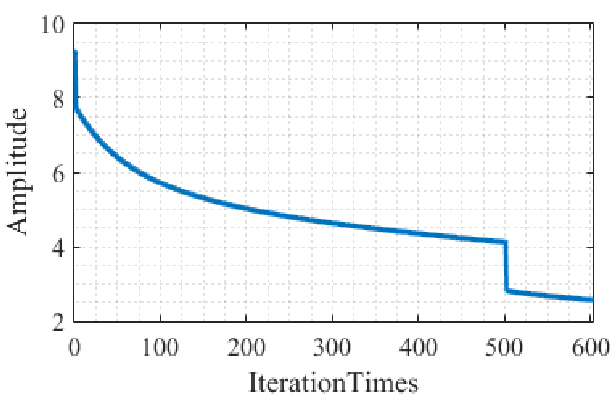


Figure 4: Convergence curve by smoothing residual data at 501th iteration.

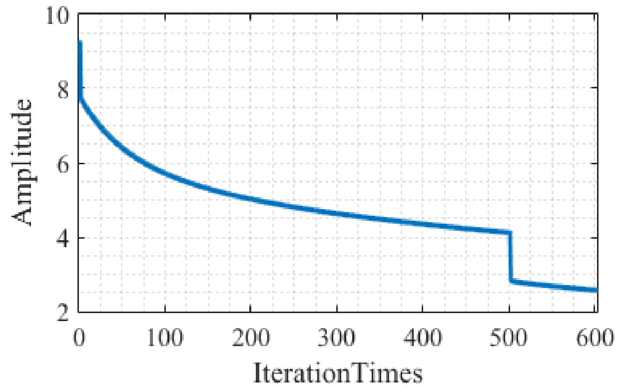


Figure 5: Convergence curve by smoothing residual data at 301th, 401th, 501th iterations.

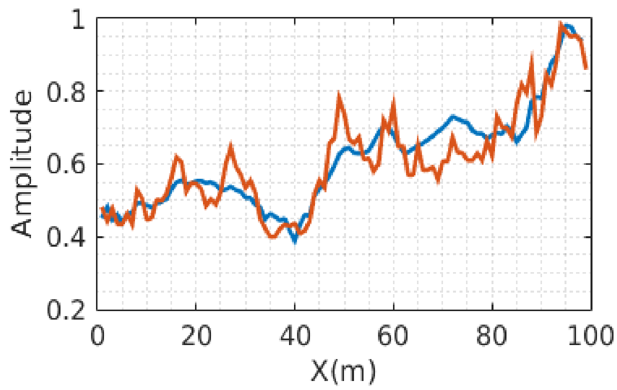


Figure 6: Inverted result from data dominated by high frequency

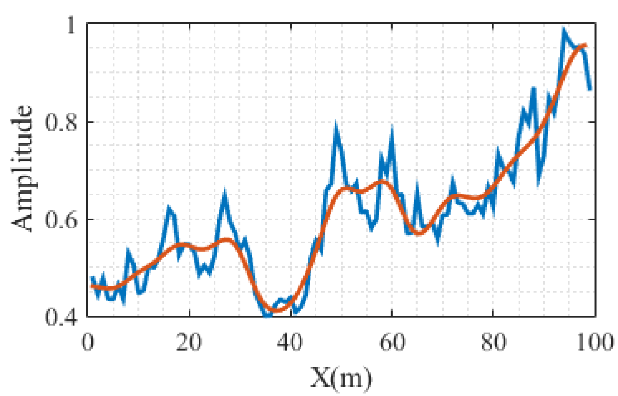


Figure 7: Inverted result from data dominated by high frequency