Seismic Full Waveform Inversion and Modeling

Richa Rastogi¹, Abhishek Srivastava¹, Kalpita Shirodkar¹* (kalpitas@cdac.in),
S.K Das², Ajith C² PK Choudhury²
1. Centre for Development of Advanced Computing, Pune 2. GEOPIC ONGC, Dehradun

Summary

Use of pre-stack seismic data for quantitative estimation of subsurface properties has gained popularity over the past decades. Most of these quantitative studies rely on the seismic reflection amplitude variation with offset or AVO analysis. But AVO analysis cannot produce the true picture in real scenario especially for a finely layered elastic model, as it takes into account only the simple primary P-wave reflection and does not include the effect of other modes of waves into the amplitude calculation. To properly model such interference, a full waveform pre-stack inversion must be used. Here, in this implementation we have used reflectivity modeling for calculation of forward response of earth models and Real Coded Genetic Algorithm for optimizing or inverting the pre-stack seismic data in 1D.

Genetic Algorithms are well known as guided random search or optimization tools. The search initiates with a random seed value and searches the whole sequence originated from that seed. The parameters to be searched take the values in each generation with in the physically specified limit using the random sequence. Here, in this article we are exploring these two aspects of the GA and trying to judge the convergence.

We have predicted the elastic parameters P, S-wave velocity and density (Rho) assuming the depths of the interfaces are well known, for a synthetically generated pre-stack data which is close to the actual field data and with a known wavelet. The experiment is on two types of scenario, first one is on the physical limit of each parameter, it is fixed or variable, and the second one is using multiple runs (with different seed values) of GA instead of a single run. For both the case the results and convergence has been discussed along with the future directions and improvements. This all has been achieved by using the INVWAV Software developed by C-DAC in collaboration with GEOPIC, ONGC.

Introduction

The objective of seismic inversion is to make inferences about the physical parameters based on a prior knowledge and a set of seismic measurements. In general, these inverse problems are multidimensional, ill posed, and nonlinear in nature. Therefore, local optimization techniques e.g. generalized linear inversion, steepest descent etc. do not provide a satisfactory solution since they normally converge to a local minima, depending upon the choice of the starting model. Global optimization methods, like Genetic Algorithms are therefore suitable for solving these problems.

AVO and AVA analysis is increasingly used in the oil industry for direct detection of hydrocarbons (Sen and Stoffa 1992, Castagna and Smith 1994, Mallick 1995, Cambois 2000). But the inversion of pre-stack seismic data is not commonly used in the industry due to lack of computing resources, convergence issues, and non-uniqueness. Though it is well known that the amplitude information contained in the pre-stack seismic records can be related to lithology. With the advancement in parallel and distributed computing, it is now possible to invert the pre-stack seismic data to obtain acoustic properties, which is critical for increasing our quantitative understanding of lithology, and has direct application in hydrocarbon exploration and exploitation. Seismic waveform inversion
seeks to determine compressional wave (P-wave) velocity, shear wave (S-wave) velocity and density (Rho) by minimizing the differences between observed and synthetic seismic data based on proposed subsurface structures and a particular type of wave propagation.

In the presented research work we have used INVWAV software, which uses Real Coded GA as inversion technique and Reflectivity Modeling as forward modeler. INVWAV is a parallel software and the results presented here are worked out on 4 CPU SGI Prism at CDAC.

Here, the process of inversion in this paper is revised with respect to two main criteria, first is the elasticity of the search space of the inversion parameters and second is the number of GA runs using different random seeds (Multiple realization).

Forward Modeling

A fast and efficient forward modeling algorithm is necessary for any inversion process because more than 90 percent of the computational time is spent in forward modeling calculations. Here we have adopted the reflectivity modeling for this purpose.

The reflectivity method for the computation of synthetic seismogram in layered media is widely used in the industry because of its fastness, compared to the finite difference based elastic wave equation modeling algorithm, and its capacity to compute a total solution of the wave field for a given flat layered-earth model. In the reflectivity method the response of the model in frequency-wave number space is generally computed by a matrix method (Kennett, 1979; Kind, 1976; Kennett, 1983) that automatically includes contributions from all possible generalized rays within the reflecting zone. The seismic response is then converted into the angle gather form for comparing with the observed data.

A simplified model response can also be generated using Elastic Impedance approach, which makes use of the modified zeopritz equation collony(1999).

Real Coded Genetic Algorithm

Genetic Algorithm is a population based guided search technique. It maintains a population of individual solutions that compete amongst themselves based on the Darwin’s Survival of the Fittest principle (Goldberg, 1989). In Genetic Algorithm, two choices for parameter representation are available – binary coded or real coded. Recent researchers have shown that real coded genetic algorithms outperform the binary coded algorithms especially for continuous search space (Boschetti, et al., 1996, Qing, et al., 2001). Therefore, a real coded genetic algorithm was implemented. Unlike binary-coded GA, here an individual is represented by an array of real values, which in the application presented here corresponds to P-wave velocity, S-wave velocity, density and thickness for each layer. In genetic algorithm, the first step is to create a population randomly in a defined search width for each of the model parameters, which can be obtained from the prior information. Using forward modeling procedure the synthetic data is generated for each chromosome / model of the population and a fitness function is evaluated. Then three operators of GA namely, selection, crossover and mutation are applied to the population in a sequence and the population is modified. A number of crossover and mutation operators exist for real coded genetic algorithms in the literature; however the search power to achieve both the above aspects differs from one to another. Here, in this case, we have used the blend crossover (BLX-0.5) operator (Eshelman and Schaffer, 1993) and the non-uniform mutation.

By applying these genetic operators, the population is updated so that the average fitness of the population improves from generation to generation and this finally leads to the convergence. A Hybrid Island Model (Chipperfield and Fleming, 1996) is adopted for the parallel implementation of the genetic algorithms.

Inversion using RCGA

AVA (Amplitude Variation with Angle) inversion is increasingly used in oil and gas industry for estimation of elastic parameters from pre-stack seismic data. Here, RCGA has been used for estimating compressional and shear wave velocities and densities from angle gather. The forward modeling uses the concept of reflectivity for calculation of synthetic angle gathers. The starting model for inversion is derived from the velocity analysis on CMP gathers or via some prior information from near well locations. The CMP gather is converted to angle gathers for comparison of observed and synthetic data. A one-dimensional variation of elastic parameters is assumed at the given CMP location.

Our first criterion of this research is the physical range of inversion parameters Vp, Vs and Rho. Earlier a single range was specified for Vp, Vs and Rho respectively. The values for these parameters then would be randomly selected from within that range. In our modification we are now selecting Vp, Vs, and Rho based on a central guided value given to each of these parameters. The required number of Vp, Vs, and Rho values are then selected with some percentage (+/-) of this central value. In this regards the GA also proceed with a guided search instead of randomly selecting parameters. Observed results show good improvement over the earlier range selection procedure. Results with range and guided central parameters are shown in numerical examples.

Secondly, GA starts with a random seed value. Model is generated randomly with respect to this selected seed value.
The search will therefore proceed in that particular direction of seed area instead of searching the entire available area. In this paper we have taken results with multiple GA runs and the seed value is different for each run. Selecting seed values in this way gives better search possibilities. Finally, we either take the average of the parameters, in different runs or the best of the run. Results for best and average cases compared to the single GA run case are found better in this scenario.

<table>
<thead>
<tr>
<th>Case</th>
<th>GA Runs</th>
<th>Average Fitness (Fitness average whole population) scale: 0.0-1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Hard Coded Range</td>
<td>1</td>
<td>0.745741 (Run 1)</td>
</tr>
<tr>
<td>2. Varying Range</td>
<td>1</td>
<td>0.674840 (Run 1)</td>
</tr>
<tr>
<td>3. Varying Range, multiple runs</td>
<td>3</td>
<td>0.674840 (Run 1), 0.652162 (Run 2), 0.652346 (Run 3)</td>
</tr>
</tbody>
</table>

Table 1: Average Fitness of Population for Single and Multiple Runs. Actual Graphs are given in Figures 3, 6, and 10 respectively

In this way we are also trying to reduce the redundancy in selecting of same values for different runs. Below we present the results of single and multiple GA runs for best and average cases.

Actual average fitness values of the whole population can be seen in Table 1. Discussion pertaining to this is followed in the Discussion and Conclusions section. The colors to the multiple run values in the table depict the respective colors for the runs as in Figure 10.

**Numerical Examples**

The developed inversion algorithm is tested on synthetic data to simulate the real field situations. Figure 1 shows the synthetic offset gather used as observed data for the inversion. This offset gather has been generated through Hampson Russell (HR) software at ONGC using P-sonic, S-sonic and density logs from a recorded offshore well. The high frequency components were blocked to produce a 96-layer earth model. EI method was used to generate the synthetic gather, however, it can be generated with other options in the HR software. The wavelet used is 25 Hz, zero phase ricker wavelet. The inversion interest zone is from 1.6s to 1.9s. The input offset gather is NMO corrected (Figure 1).

This data is inverted using the proposed reflectivity modeling based parallel RCGA scheme in angle gather domain. The red lines show the inverted model after 100 generations with a population size of 900 on 3 processors i.e. 300 individuals or models per processor. The crossover and mutation probabilities used, in these cases, are 0.9 and 0.0001 respectively. Figure 3 illustrates the convergence of the inversion algorithm with iteration (generation/run). The fitness is calculated using the normalized cross-correlation function between the input and output datasets (Sen and Stoffa, 1992). The single GA job took about 1 day 20 hrs and 37 mins whereas 3 GA runs job took 2 days 5 hrs and 24 mins on 4 processors of SGI Prism. Figures 2, 5 and 8 shows the predicted elastic parameters (red lines) in comparison to exact parameters (green lines). Figures 4, 7 and 9 shows the comparison of the input data with the synthetic data for the inverted model in the form of angle gathers. Figures 3, 6 and 10 shows the average fitness graphs for different test cases. One can see a good match between the two data sets from Figures 7 and 10. Following are the test cases performed on the above-mentioned data set:

**Test Case 1: Fixed physical range of parameters with single GA Run**

Ranges for Rho is 2100 – 2500kg/m3, Vp is 2000 – 3400 m/s, and Vs is 700 – 1900m/s.
Test Case 2: Variable physical range of parameters with single GA Run

Varying percentage ranges for variables for a given central value are as follows, ranges for 1st layer, Rho is 0.1, Vp is 0.1, and Vs is 0.1. Ranges for 2nd to n-th layer, Rho is 5.0, Vp is 10.0, and Vs is 10.0.
Test Case 3: Variable physical range with multiple GA runs

Varying percentage ranges for variables for given central value, 1st layer are Rho is 0.1, Vp is 0.1, and Vs is 0.1. Ranges for 2nd to nth layer are Rho is 5.0, Vp is 10.0, and Vs is 10.0. The graphs show the average and best cases for three GA runs.

Discussion and Conclusion

In this paper we have focused throughout on two major concepts, first is the comparison of fixed and variable range selection for the parameter values and second is the comparison of single GA run and average of multiple GA runs for the final convergence. We have achieved improved results in the modified approach, during the experimentation.

In case of the fixed physical range of the parameters the final parameter values and the result is not of the quality as of varying percentage of the parameters around the central guiding value. This shows that in the second case the search space is well defined for GA. Hence, GA has searched better parameter values. See Figures 2 and 5 for comparison of final parameter values and the corresponding AVA comparison in Figures 4 and 7.

The next test case is a comparison between the output of single GA and average/best of multiple GA runs. As can be seen in graphs of figures 5 and 8, the parameter values are better resolved in the multiple GA then in single GA run. Also the output AVA is better in the multiple GA then in single GA run, see Figures 9 and 7. It is expected that average of multiple runs (3) realization to be closer to the real parameter values. The difference in the observed and predicted section could be accounted to the difference in the forward modelers.
Figure 9: A comparison of the observed angle gather with the synthetic angle gather for the inverted model for variable physical range of parameters and multiple GA run for inversion.

Figure 10: The graph showing the average fitness for three GA runs (Red (best), green and Blue) with variable physical range of parameters as a function of iterations (generation). Red one is with the best average fitness.

In stochastic methods like GA diversity of population is always considered as a positive indication of better search. In the current experiment, this has emerged clearly through comparison between hard-coded versus variable range parameter settings. In the variable range case, though the average fitness is away from the best, the diversity is higher and fittest individual is better. In the fixed parameter range case the population is restricted and hence poor diversity and poor best fitness.

The overall conclusion of this modeling and inversion approach is that the multiple GA runs along with guided parameter computation yields better results than single GA run and parameter computation with fixed range.

Future Direction of Work

In future, we will make use of these concepts of multiple GA runs and variable range for inversion in time domain instead of depth domain. Similar attempt will be made with a synthetic data generation using RM and inversion through EI, with addition of random noise to the data and finally real field/processed data inversion with an extracted or unknown wavelet. Typical, inversion input would be a PSTM processed CMP gather, free from multiples to suit the selected forward modeler. A hybrid scheme will also be tried, which will combine the global optimization with local optimization, in order to improve the convergence.

Acknowledgment

Authors wish to thank Director (Exoln), ONGC for funding this research work; ONGC and C-DAC for providing computational facility and permission to publish this work.

Also would like to thank Deepak Sareen, TR Muralimohan from ONGC, Dr. V. Sundararajan and Dr. S. C. Purohit from CDAC for technical discussions.

The views expressed in this work are those of authors and may not necessarily be those of their respective organisations.

References


Kennett, B. L. N., 1983. Seismic wave propagation in stratified media: Cambridge Univ. Press.

