



An Algorithm: Seismic Travel Time Tomographic Inversion using Real Coded and Binary Coded Genetic Algorithm.

Richa Rastogi¹, Amit Ray², Abhishek Srivastava^{1*} (abhisheks@cdac.in)

1. Centre for Development of Advanced Computing, Pune

2. Reliance Industries Limited, Mumbai

Summary

An efficient, robust and parallel two-dimensional (2D) seismic traveltimes tomographic inversion scheme using real-coded and binary-coded genetic algorithm has been developed and results of the same is discussed.

Introduction

Seismic imaging, popularly known as tomography, borrowing the term from medical sciences, produces the raster image of the internal structure by combining information from a set of projections obtained at different viewing angles. To engineer high resolution and accurate solutions to mining problems, seismic traveltimes tomography can yield a high resolution image of the subsurface to provide information about lithologic characterization, fracture / void or gallery detection, fluid monitoring etc. The measured characteristics of received wave energy include amplitude and traveltimes.

The resolution of the estimated images in seismic tomography depends on the inversion schemes used, completeness of the data sets and the initially assigned values of the medium parameters in the iterative reconstruction techniques. Since optimal solutions are always aimed at, even from poor initial models, global optimization and search technique such as real-coded and binary-coded genetic algorithm are applied for solving the present imaging problem. The algorithms are tested on a number of complex synthetic subsurface models, of which the results of one of them is presented here. The numerical example is selected to simulate geological conditions for detecting an 'L'-shaped anomalous zone in the subsurface.

Theory and Method

Theories used in the development of the scheme are as follows:

Forward Modelling

John Vidale proposed an algorithm (Vidale, J., 1988) based on finite-difference. The timing process is initiated by assigning point A the travel time of zero. The propagation of 2-D wavefronts is guided by the eikonal equation of ray tracing.

If the times t_0 , t_1 and t_2 at 3 points on a square grid are known, then the time at the fourth point can be estimated as

$$t_3 = t_0 + \sqrt{2(hs)^2 - (t_2 - t_1)^2} \quad \dots(1)$$

where 'h' is the grid spacing and 's' is the slowness.

The finite difference algorithm of Vidale fails when large, sharp velocity contrasts are encountered. Hole and Zelt proposed a modified algorithm (Hole & Zelt, 1995) with minor modification of Vidale's scheme. Two improvements over Vidale's model (i) incorporating head-wave operator, (ii) reverse propagation of travel times.

At the Velocity discontinuities, the travel time has been computed using a head-wave operator

$$t_0 = t_1 + hs \quad \dots (2)$$

This 1-D operator represents a head wave that propagates along the boundary between the two media at the velocity of faster media. Whenever head waves are observed traveling parallel to the face of the expanding square i.e.,



"HYDERABAD 2008"

whenever the head-wave operators are used along the faces of the expanding square, a flag is set in the algorithm. After travel times have been computed throughout the entire model, the travel time field is used as a starting point and times are recomputed starting with a source at the appropriate wall of the model (in this algorithm the bottom face of the model) and expanding perpendicular to that face (upward). Travel times from the new calculations only replace the previously computed times if the new times are smaller.

This algorithm can handle large and sharp velocity contrasts and is faster than several existing algorithms.

Genetic algorithm (GA)

Genetic Algorithm is a population based guided global search technique (Goldberg, 1989) extensively used in geophysical inversion by many researchers. In implementing the Genetic Algorithm for a particular problem, one must first make a choice for parameter representation and genetic operators. This choice is problem dependent. The two main choices available for parameter representation are binary coding or real coding. The binary coded GA has been widely used by the researchers since long time and it is giving good results for the problems having discrete search spaces.

Arguments in support of Binary Coded Genetic Algorithm:

- To choose GA coding, following principles may help:
 1. Principle of Meaningful Building Block: Select coding so that, low-order schemata are relevant to the underlying problem and relatively unrelated to schemata over other fixed position.
 2. Principle of Minimum Alphabet: Select the smallest alphabet that permits a natural expression of the problem..
- In binary coding we can easily find out similarities between individuals due to small cardinality of the alphabet.
- Binary coding offers maximum number of schemata per bit of information of any coding. (Goldberg, D.E., 1989)

Arguments in support of Real Coded Genetic Algorithm:

- When a binary coding is used for continuous search space then Hamming Cliff get associated with certain strings due to which a transition to the neighbouring solution requires alteration in many bits. It causes artificial hindrance to a gradual search in the continuous search space.
- Inability of binary coding to achieve any arbitrary precision in optimal solution. String length in

binary coding should be decided priori to achieve certain precision. To achieve more precision string length should be more and for that population should be increased and increase in population means increase in computation time.

- Recent results have shown that real coded genetic algorithms outperform the binary coded algorithms in most applications (Boschetti, et al., 1996, Qing, et al., 2001).

The first step is to define a search width for each of the model parameters, which can be obtained a priori. Each model is represented by a string of real/binary numbers called *chromosome*. At the start, a finite population of chromosomes is generated at random. Using forward modeling procedure the synthetic data is generated for each chromosome / model of the population and a fitness function is evaluated. Then three steps of GA namely, *selection*, *crossover* and *mutation* are applied to the population in a sequence and the population is modified.

Real coded genetic algorithm individuals are nothing but float values. In case of BCGA individuals used are strings of bits. To work with Binary coded genetic algorithm individual coding and decoding is required. In coding values of genetic algorithm variables is determined and in decoding binary chromosome value is converted double value.

The genetic algorithm involves natural selection and two basic operators corresponding to the biological processes of *crossover* and *mutation* (crossover and mutation are known as genetic operators).

Selection involves the choice of the individuals for the generation of offspring. It produces an intermediate population in such a way that the individuals with higher fitness usually have a greater chance of contributing copies to the intermediate population. Researchers have proposed a number of selection mechanisms in GA literature. Here, we have adopted the tournament selection, which is very simple and efficient.

Crossover is the method of combining (mating) two individuals to produce offsprings. The crossover operator is believed to be the main search operator in the working of a genetic algorithm as an optimization tool. The purpose of a crossover operator is two-fold. Initial random strings representing the problem variables must be searched thoroughly in order to create good strings. Thereafter, good portions of these strings must be combined together to form better strings. A number of crossover operators exist for real coded genetic algorithm in the literature; however the search power to achieve both the above aspects differs from one crossover to another. Here, we have used the blend crossover (BLX- α , $\alpha=0.5$) operator suggested by Eshelman and Schaffer (1993).



For binary coded genetic algorithm a site is chosen randomly and the two parents are crossed to generate two children. Uniform crossover is used here.

Mutation is the random changing of some individuals within the population by altering one or more variables, called *genes*, of a selected chromosome or model so as to increase the diversity in the population. The role of mutation in GA is that of restoring lost or unexplored portions of the search domain into the population to prevent premature convergence of GA to suboptimal solutions. It insures that the probability of reaching any point in the search space is never zero. Among the available mutation operators for RCGA, the non-uniform mutation seems to be very efficient and is implemented here.

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. Here, the whole population is randomly generated in a processor, called *master*, and then the population is distributed among the other processors (called *worker*) for the successive generation.

The algorithm used is Parallel Genetic Algorithm with Hybrid Island Model, flowchart of which is depicted in figure 3.

Worker processors are arranged in ring topology and any communication between worker processors happens in ring fashion only and in clockwise direction.

In the successive generations the sub-population on each worker processor is updated by genetic operators, fitness is evaluated for each individual in the sub-population and some good members of the sub-population are exchanged among the worker processors. This continues till the average fitness of the whole population improves

The code for RCGA and BCGA was executed on PARAM Padma. A small summary of PARAM Padma configuration is given in Table1.

No. of Processors:	248 (Power 4)
Aggregate Memory:	0.5 TeraBytes
Internal Storage:	4.5 TeraBytes



"HYDERABAD 2008"

Operating System:	AIX 5L/LINUX (SuSe)
Aggregate Peak Computing Power	1005 GF (~T F)

Table 1. Summary of PARAM Padma configuration

Example

Initial model estimation

The initial parameter estimation plays a key role in any iterative reconstruction process. The closer the initial guess to the actual solution is, the search space of genetic algorithm will be smaller and convergence will be faster. If the initial assumption is very far off from the actual solution, the search space will be higher and to converge to a reasonable closer model it need to have larger population size and which in turn consumes huge computational resource. So the initial parameterization has to be done very cautiously. In the present investigation we have used back projection method (Wong *et al.*, 1983) for this purpose. The raypaths are assumed to be straight.

If L_{mn} be the length of the straight raypath joining the source m and receiver n and T_{mn} be the recorded first arrival time taken by the ray to travel from the source m to receiver n , the average velocity for the entire region can be written as,

$$\bar{V} = \frac{\sum_{mn} L_{mn}}{\sum_{mn} T_{mn}} \quad (3)$$

With $\bar{V}$ as the initial velocity distribution of the model region, we use RCGA and BCGA for solving the transmission tomographic problem as discussed below.

An 'L' shaped intrusive model

Our first synthetic model is an L-shaped intrusive body (see Fig.1a) with two velocities (3.0km/sec and 3.5km/sec) and with the background velocity of 2.5km/sec. The combined source-receiver configuration of Surface-to-borehole, cross-hole and borehole-to-surface configurations is used for the data generation and inversion purpose. The subsurface area of 10m × 10m in X and Z directions is divided into 11 × 11 square grids. 30 source-receiver pairs have been used for simulating the data acquisition. The generation of synthetic data is done using the forward-difference based algorithm (Hole and Zelt, 1995) for the



"HYDERABAD 2008"

true velocity model. Then the tomographic reconstruction i.e. the estimation of the velocity distribution in the subsurface is carried out using parallel RCGA as well as BCGA. The initial velocity model with a uniform velocity of 2.74 km/sec as shown in Fig.1b, has been estimated by Wong's back projection method. A velocity range of $\pm 28\%$ is used for each cell of this initial guess model for defining the search space for RCGA and BCGA. The reconstructed model after 1250 generations is depicted in Figure 2b for RCGA and in Figure 2a for BCGA.

The GA and the model parameters along with other execution parameters like - rms error in velocity and the execution time etc. for both the coding techniques are given in Table 2.

The reconstructed tomograms demonstrated here, are well resolved, as evident from the low rms errors in velocity, without any a-priori information. The spurious velocities in the lower side of the reconstructed image is due to the less ray coverage in that region, as at the lower end of the model space there is no source-receiver.

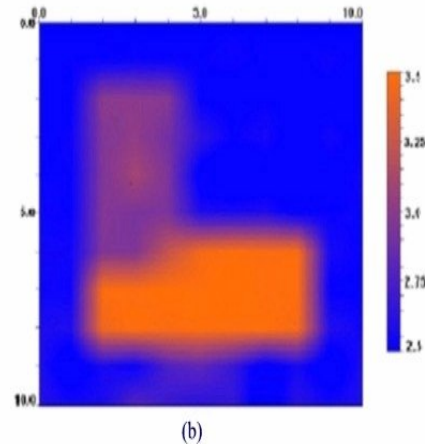
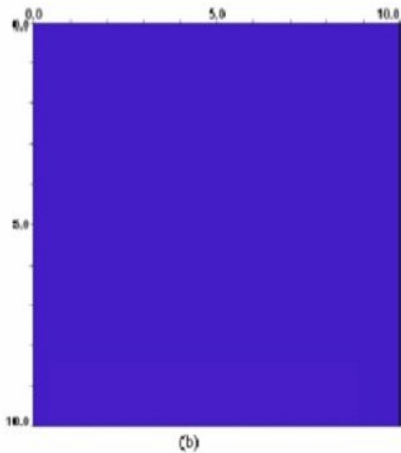
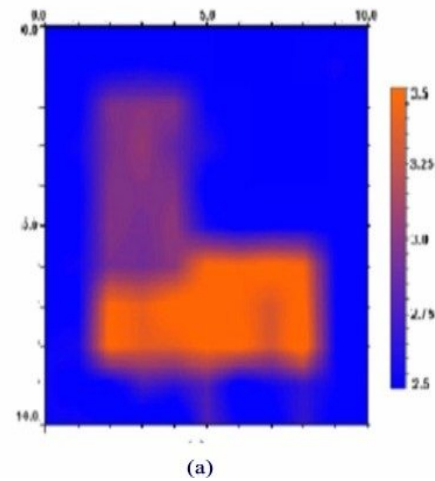
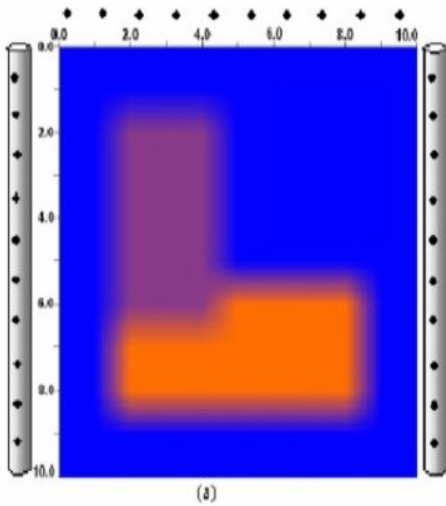


Figure 1: (a) Original Model (b) Initial Model

Figure 2. (a) Reconstructed model using BCGA
(b) Reconstructed model using RCGA



"HYDERABAD 2008"

GA and Model parameters		L-shaped model RCGA	L-shaped model BCGA
	Number of generation	1250	1250
	Number of models per processor	1500	1500
	Number of processor	19	19
	Velocity range used for RCGA for each node (in %)	± 28	± 28
Observations			
	Initial rms error in velocity	12.30	15.51
	Final rms error in velocity	3.69	6.67
	Best individual fitness	0.997497	0.992751
	Time taken for execution on PARAM Padma (1 TF m/c)	16hrs 31mins	19hrs 21mins

Table 2. GA and model parameters along with other observations for both coding technique.

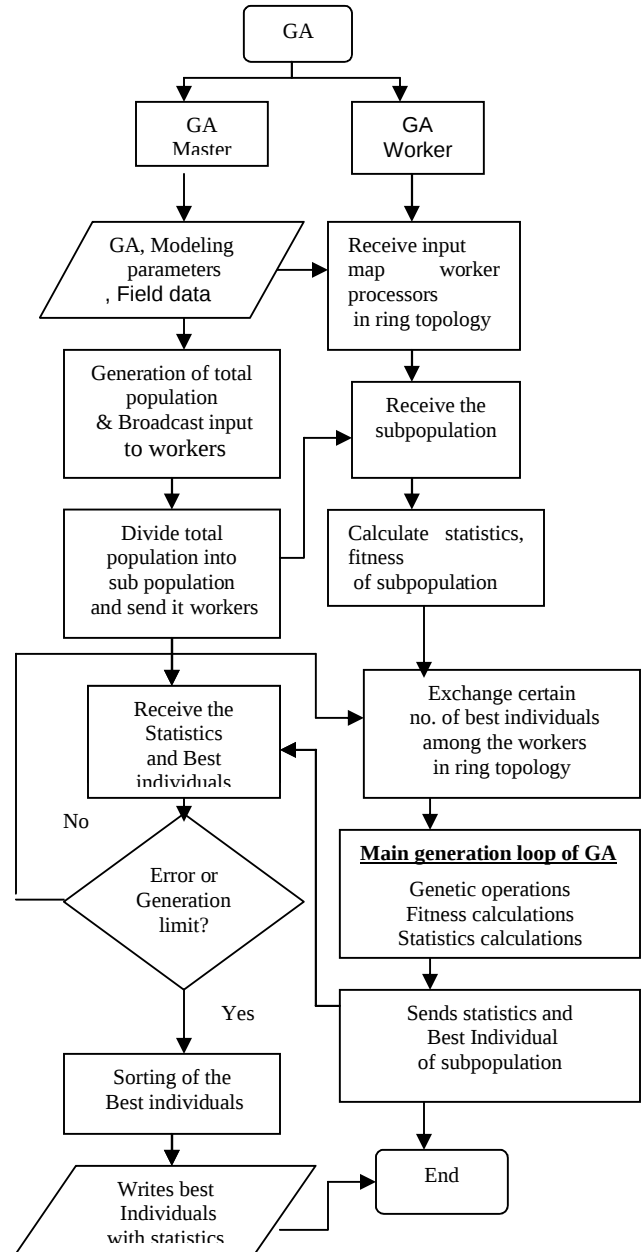


Figure 3: Flow chart of the Parallel Genetic Algorithm – Hybrid Island model



"HYDERABAD 2008"

Conclusions

Seismic Traveltime Tomography Inversion parallel algorithm has been developed using both real coded and binary coded genetic algorithm. GA operators, viz., selection, crossover and mutation for both RCGA and BCGA are different.

Binary coded genetic algorithm took more time for executions because it has an overhead of coding and decoding the strings whereas real coding uses the strings as it is.

Real coding gave better results as the search area for the problem is continuous and not discrete. Real coding can achieve arbitrary precision in optimal solution but this is a drawback for binary coding.

The results for both the coding scheme have been discussed. Results show that both the coding techniques are quite comparable and can be distinguished on the basis of rms error in velocity and execution time on Param Padma.

References

- Boschetti, Fabio, Dentith, M. C., and List, R. D., 1996, Inversion of seismic refraction data using genetic algorithms, *Geophysics*, **61**(6), 1715–1727.
- Bube, K., and Langan, R., 1997, A Hybrid 11/1 2 minimization with applications to tomography: *Geophysics*, **62**, 1183-1195.
- Chakraborty, S., Nath, S. K., and Sengupta, S., 1998, Seismic Transmission Tomography – A configurative approach: *Acta Geophysica Polonica*, XLVI (4), 389-410.
- Chipperfield, A. and Fleming, P., 1996, "Parallel Genetic Algorithms," in *Parallel & Distributed Computing Handbook* by A. Y. H. Zomaya, McGraw-Hill, pp. 1118-1143.
- Eshelman, L. J. and Schaffer, J. D. (1993). Real-coded genetic algorithms and interval schemata. In D. Whitley (Ed.), *Foundations of Genetic Algorithms*, II (pp. 187 - 202).
- Goldberg, D.,E., 1989, *Genetic algorithms in search optimization and machine learning*, Addison-Wesley Publishing Company Inc.
- Hole, J.,A., and Zelt, B.,C., 1995, 3-D finite-difference reflection traveltimes, *Geophys. J. Int.*, **121**, 427-434.
- Iyer, H. M., and Hirahara, K., 1993, *Seismic Tomography - Theory and Practice*: Chapman and Hall, first edition.
- Luo, Y. and Schuster, G.,T., 1991, Wave-equation travelttime inversion, *Geophysics*, **56**, 645-653.
- Mcmechan, G. A., Harris, J. M., and Anderson, L. M., 1987, Crosshole tomography for strongly variable media with applications to scale model data: *Bull., Seis. Soc. Am.*, **77**, 1945-1960.
- Qing, A., Lee, C. K. and Jen, L., 2001, Electromagnetic Inverse Scattering of Two-Dimensional Perfectly Conducting Objects by Real-Coded Genetic Algorithm, *IEEE Transactions on Geoscience and Remote Sensing*, **39**(3), 665-676.
- Vasco, D. W., 1991, Boundary seismic velocities using a tomographic method: *Geophysics*, **56**, 472-482.
- Velis, D.,R., 2001, Travelttime inversion for 2-D anomaly structure, *Geophysics*, **66**, 1481-1487.
- Vidale, J., 1988, Finite-difference calculation of travel times: *Bull., Seis. Soc. Am.*, **78**, 2062-2076.
- Witten, A. J., and King, W. C., 1990, Acoustic imaging of subsurface features: *Journal of Environmental Engineering*, ASCE, **116**, 116-181.
- Wong, J., Hurley, P., and West, G.,F., 1983, Crosshole and seismic imaging in crystalline rocks, *Geophys. Res. Lett.*, **10**, 686-689.
- Yamamoto, T., Nye, T., and Kuru, M., 1994, Porosity, permeability, shear strength: Cross-hole tomography below an iron foundry: *Geophysics*, **59**, 1530-1541.
- Zhang, J., and McMechan, G. A., 1994, 3-D transmission tomography using wide aperture data for velocity estimation of irregular salt bodies: *Geophysics*, **59**, 1620-1630.