Simulated annealing performance using error measures with different norms

Jorge Magalhães de Mendonça*, Paulo Léo Osorio**

*PETROBRAS SA / BRASIL

**DEE PUC-Rio

Abstract

The seismic inverse problem involves the determination of the subsurface physical properties from data recorded at the earth's surface. A subsurface response mathematical model can be constructed from a physical modeling of the subsurface when excited by a seismic source. This mathematical model has the subsurface physical properties as parameters and provides an estimate of the sample data. This makes possible to compare the sample with the estimated data. A function is defined over the model parameter space that measures the error between sample data and its estimates. The inverse problem can then be stated as a minimization problem of error function over the parameter space. Most geophysical inverse problems are highly nonlinear and are rife with local minima. Classical approaches to this optimization problem are very sensitive to the choice of the initial model, and good starting solutions may not be available. If there is no basis for an initial guess, the theory of Bayesian Inference provides an alternative way to this question taking into account the prior information about the parameter space. The inverse problem can then be stated as an optimization problem whose goal is to maximize the a posteriori probability that the set of parameters has a certain value once given the result of the sample. This problem can be solved by the Simulated Annealing (SA) method, which is a global optimization method that executes an oriented random search in the solution space. The SA method is applied to the solution of 1D seismic inverse problems. The SA performance using error measures with different norms is evaluated in the inversion of an elastic model with additive random noise. Low order norm equalizes the components of the error norm, since it gives more equal weight to different size errors. Moreover that attenuates the difference between sample and synthetic data, reducing the influence of each model parameter individually, preventing the undue bias of parameters. The effect of using a low order norm is the increasing of the SA convergence time and the augmented sensitivity to all parameters in the solution reached.

INTRODUCTION

The earth's subsurface, when excited by a seismic source, produces an elastic wave propagating through itself. In its propagation, this wave is subjected to reflections, refractions and diffractions which result in a distorted image of the subsurface. The seismic processing with techniques as deconvolution and migration tries to remove the effects of the wave propagation, transforming the set of recorded seismograms in sections in depth, that is, spatial images of the some properties such as the compressional wave velocity.

The Seismic Inversion approaches this problem in a different manner. The geophysics have been participating in the development of the Inversion Theory, since they need to infer the true values of a certain physical parameterization of the earth's subsurface, but they are limited to use only data registered at the earth's surface. From a physical modeling of the propagation of elastic waves through the subsurface, a mathematical model can be constructed having as variables the physical modeling parameters. For each specification of feasible values of these parameters, this mathematical model produces an estimate of the data to be observed. The Inversion uses the data recorded and the estimates made by the model to infer the true values of the model parameters. A function is defined over the model parameter space that measures the error between the survey results and the estimates of the model. The inversion problem becomes an optimization problem of this function over the model parameter space.

The earth's subsurface parameterization can be represented by vector $\mathbf{X}=(X_1,...,X_n)$, $\mathbf{X} \in R^n$. If s data samples are recorded in a seismic survey, one might consider vector $\mathbf{D}=(D_1, ... , D_s)$, $\mathbf{D} \in R^s$, a generic element of the data space. Then $\mathbf{X}$ and $\mathbf{D}$ are random variables that assume the specific values of $\mathbf{x}=(x_1,...,x_n)$ and $\mathbf{d}=(d_1,...,d_s)$. Let $\mathbf{G}=(G_1,...,G_s)$ be the subsurface mathematical model, where $\mathbf{G}$ is a vector function composed by s scalar functions G_i. Let $\mathbf{d_e} = \mathbf{G}(\mathbf{x})$ be the estimate of data $\mathbf{d}$ provided by model $\mathbf{G}$ as a function of parameters $\mathbf{x}$. If measurement of data $\mathbf{d}$ is corrupted by an additive random noise $\mathbf{n}$, one can write

$$\mathbf{d} = \mathbf{d_e} + \mathbf{n} = \mathbf{G}(\mathbf{x}) + \mathbf{n} \qquad (1)$$

where $\mathbf{n} = (n_1,..., n_s)$ is a realization of the random noise $\mathbf{N}=(N_1,..N_s)$ in which component N_i is supposedly independent and identically distributed.

The resolution of equation (1) is one of the most ambitious geophysics inverse problems. In a more realistic manner, $\mathbf{x}$ is a model parameter vector in which some of its physical quantities are known. The inversion of equation (1) is usually realized by the use of an optimization method . One searches for $\mathbf{x}$ in order to solve the minimization problem

$$\min_x \; f\,[\,\mathbf{d}\,,\mathbf{G}(\mathbf{x})\,] \qquad (2)$$

where f is the function that measures the error between data $\mathbf{d}$ and estimate $\mathbf{G}(\mathbf{x})$. The solution of problem stated by equation (2) provides the point $\mathbf{x}$ that minimizes the difference between data $\mathbf{d}$ and the model $\mathbf{G}(\mathbf{x})$ measures by function f. In most of geophysical problems, the equation (2) expresses a nonlinear inverse problem where the localization of the global minimum is made in the presence of a great number of local minima. In spite of its complexity, this equation does not introduce an intractable problem, but demands the use of a priori information to be solved.

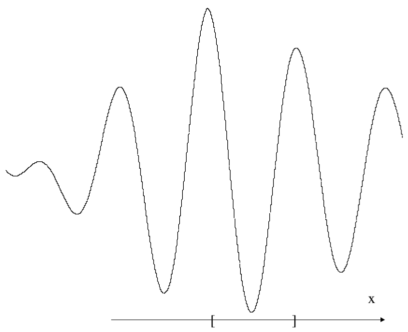

Figure 1 : The interval to a good initial solution

The classical approaches to the solution of nonlinear inverse problems makes use of a starting solution $\mathbf{x}^0$. The solution is improved by an iterative process, perturbing function **G** in vicinity of $\mathbf{x}^0$. This perturbation about $\mathbf{\Delta x} = \mathbf{x} - \mathbf{x}^0$ satisfies approximately a linear relation driving to the resolution of a linear inverse problem at each iteration. From initial solution $\mathbf{x}^0$, the process evolves decreasing the objective function, assuring the condition of a local minimum solution to the final solution obtained. The localization of the global minimum solution depends entirely on having a good initial solution which begins the process and leads it to the global minimum solution. Observing the figure 1, one can note that only guesses in the interior of the indicated interval, lead the process to the global minimum solution, while any other choice drives it to a local minimum solution.

The method SA (Kirpatrick, 1983), which is a global optimization method, makes a oriented random search over the solution space permitting to overcome the absence of a good initial solution. The inverse problem is resolved by the method of SA. The information a priori over the solution space is used to delimitate the searching space. The SA method aims to optimize the a posteriori probability of the parameters assume a certain value, given the sample data as proposed by the Bayesian Inference.

This study evaluates the SA performance by using error measures with different norms, decreasing the SA convergence velocity and making the solution more sensitive to the parameter set (Mendonça, 1997).

ERROR MEASURES

The adequate choice of an error measure between the sample data **d** and its estimate $\mathbf{d}_e$ provided by the model function has a very important role in the SA performance. If vectors **d** and $\mathbf{d}_e$ are elements of a real Euclidian space with an internal product <.,.> and using the norm L_2, the expression h can be defined by equation (3) (Porsani, 1993)

$$h = \frac{2 < \mathbf{d}, \mathbf{d_e} >}{< \mathbf{d}, \mathbf{d} > + < \mathbf{d_e}, \mathbf{d}_e >} \quad (3).$$

The use of the Schwarz inequality can demonstrate that the values of h are contained in the closed interval [-1,1]. The expression h assumes value 1 when $\mathbf{d} = \mathbf{d}_e$ and –1 when $\mathbf{d} = -\mathbf{d}_e$. Another property of h is that it is sensitive to the differences of amplitude and phase between these vectors. As in elastic seismic inversion problems, amplitude and phase of **d** and $\mathbf{d}_e$ must be equal, the expression h is nearly always used.

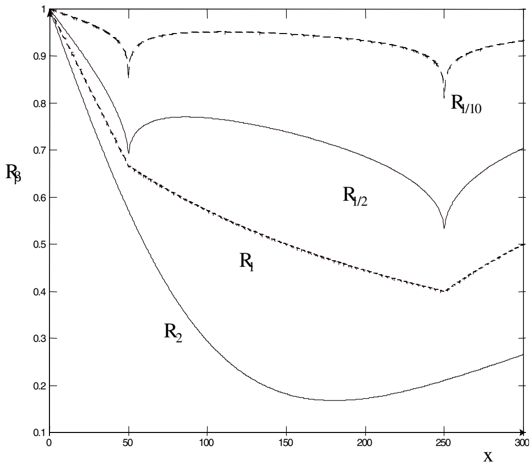

Figure 2 : The smoothness of a R_β-like function

In a real Euclidian space with an internal product and using the norm L_2, the expression h can be rewritten as in equation (4)

$$h = 1 - \frac{2\|\mathbf{d} - \mathbf{d}_e\|^2}{\|\mathbf{d} - \mathbf{d_e}\|^2 + \|\mathbf{d} - \mathbf{d_e}\|^2} = 1 - \frac{2\sum_i |d_i - d_{ei}|^2}{\sum_i |d_i - d_{ei}|^2 + \sum_i |d_i + d_{ei}|^2} \quad (4).$$

Equation (4) suggests the possibility of a generalized expression h_β by the use of a generalized norm L_β. Let the norm L_β of a vector as **d** be defined by $\|\mathbf{d}\| = \sqrt[\beta]{\sum_i |d_i|^\beta}$, then h_β can be expressed by equation (5) :

$$h_\beta = 1 - \frac{2\sum_i |d_i - d_{ei}|^\beta}{\sum_i |d_i - d_{ei}|^\beta + \sum_i |d_i + d_{ei}|^\beta} \quad (5).$$

Let $R_\beta = 1\text{-}h_\beta$, then maximize h_β is the same as minimizing R_β.

There are some considerations than can explain the use of a generalized norm L_β. A first one is the effect of β-values in norm L_β. Small β-values equalize the components of the error norm $\|\mathbf{d} - \mathbf{d}_\mathbf{e}\|$, since it gives more equal weight to different size errors. Moreover it attenuates the difference between **d** and $\mathbf{d}_e$, reducing the influence of each model parameter individually. Another consideration can be made by examining figure 2. This figure shows the behavior of R_β-like functions for some values of β. Function R_β shows a process of smoothness as β-value is decreased, reducing the R_β

-value variation for neighboring points. As at each iteration the SA solution moves among neighboring points, the effect of the smoothness is forcing the SA to visit an increasing number of solutions until it converges to a solution. These two considerations motivate the expectation that the effect of using a small β would be the increasing of the SA convergence time and the augmented sensitivity to all parameters in the solution reached.

THE EXPERIMENT AND THE ANALYSIS OF THE RESULTS

The SA is applied to the inversion of the synthetic seismic data generated by a 1D elastic model. The model parameters are the compressional wave velocity $\mathbf{V_P}$, the shear wave velocity $\mathbf{V_S}$, the density $\boldsymbol{\rho}$ and the vertical travel time $\boldsymbol{\Delta\tau}$. These parameters vary only with depth. The acoustic impedance **I** and the Poisson's ratio $\boldsymbol{\sigma}$ are derived as functions of the other parameters by equations (6) and (7) : $\mathbf{I} = \boldsymbol{\rho}\, \mathbf{V_P}$ (6) and $\boldsymbol{\sigma} = ½\, (\mathbf{V_P^2} - 2\mathbf{V_S^2}) / (\mathbf{V_P^2} - \mathbf{V_S^2})$ (7) . The **I** and $\boldsymbol{\sigma}$ parameters are employed only to the experiment result analysis. For a subsurface model with 7 horizontal layers, 64 plane wave τ-p seismograms with a maximum ray parameter p_{max} of 0.36 sec/km and with a bandwidth of 15 to 40 Hz were used in the inversion. The model included only primary compressional wave reflection with losses due to shear conversion, but did not take into account internal multiples or converted phases. A random noise with uniform distribution was added to these seismograms. This noise was limited in value to 20% of the greatest amplitude in absolute value presented in the original seismograms. The algorithm Very Fast SA (VFSA) as proposed by Ingber (1989) was chosen to implement the SA method.

For the SA performance evaluation using error measures h_β, 10 runs of the VFSA taking β the values 2, 1, ½ and 1/10, in a total of 40 runs were realized. During all 40 runs, the VFSA kept fixed its parameters with the cooling processing lasting for 250 iterations visiting 100 solutions by temperature. The VFSA examined 25000 solutions at each run.

The convergence time measures the time (or iteration) after that h_β-value approximates to vicinity of its final value during the SA runs. For each h_β, the average curve of h_β-value in the 10 runs was computed and the number of the iteration in which the h_β-value reached the 99% of its final value was considered to determine the convergence. The convergence then occurred after 25 iterations for h_2, 53 for h_1, 70 for $h_{1/2}$ and 75 for $h_{1/10}$. These results confirmed that convergence time increased as β-value was decreased.

Table 1 : Percentage of absolute deviations of the best solution parameters from the original parameters

h_2							h_1						
Layer	V_P	ρ	V_S	I	σ	Total	Layer	V_P	ρ	V_S	I	σ	Total
1	.00	3.80		3.80			1	.00	12.43		12.42		
2	.42	8.07	4.65	7.61	.07		2	1.08	3.42	18.48	4.47	.28	
3	.90	8.52	6.54	9.50	1.57		3	1.89	6.51	2.06	4.74	.04	
4	.35	11.60	10.15	11.98	2.65		4	.65	2.41	8.53	1.77	2.26	
5	.01	15.36	9.30	15.35	2.78		5	.11	1.99	8.70	1.87	3.02	
6	.10	7.77	8.92	7.66	3.39		6	.04	1.73	14.96	1.77	7.00	
7	1.13	15.99	.75	17.30	.12		7	.71	1.91	10.58	2.64	3.50	
Σ%	2.91	71.09	40.34	73.20	10.57	198.1	Σ%	4.50	30.39	63.33	29.68	16.10	144.0
Max	1.13	15.99	10.15	17.30	3.39	17.30	Max	1.89	12.43	18.48	12.42	7.00	18.48
$h_{1/2}$							$h_{1/10}$						
Layer	V_P	ρ	V_S	I	σ	Total	Layer	V_P	ρ	V_S	I	σ	Total
1	.00	12.64		12.64			1	.00	.58		.58		
2	.38	.43	16.81	.81	.19		2	.97	9.02	2.53	7.96	.02	
3	.61	.69	7.94	.08	1.57		3	1.20	5.49	14.88	6.75	3.19	
4	1.11	.49	4.70	.61	1.51		4	2.10	5.85	.14	8.07	.59	
5	.01	4.64	5.32	4.65	1.64		5	.25	11.15	2.27	10.88	.82	
6	.08	1.01	.75	1.09	.35		6	.41	10.12	9.63	9.67	4.51	
7	4.16	.93	4.71	5.13	.18		7	5.98	18.61	9.24	11.51	1.12	
Σ%	6.35	20.83	40.26	25.00	5.43	97.87	Σ%	10.91	60.82	38.70	55.42	10.25	176.1
Max	4.16	12.64	16.81	12.64	1.64	16.81	Max	5.98	18.61	14.88	11.51	4.51	18.61

The SA performance using error measures h_β was evaluated by the analysis of the best solution found among the 10 solutions realized for each β-value. Table 1 shows for each h_β the absolute deviation percentage of the best solution model parameters from the original ones. The statistics of the absolute deviation sum (Σ%)[1] and the maximum deviation (Max) were computed for the model parameters $\mathbf{V_P}$, $\boldsymbol{\rho}$, $\mathbf{V_S}$, **I** and $\boldsymbol{\sigma}$ (the vertical travel time values are considered known). These statistics are measures of the length of these deviations made with the use of the L_1 and L_∞ norms respectively. Both statistics pointed out to the predominance of the compressional velocity value $\mathbf{V_P}$ in the best solution using the error measures h_2 and h_1, while the Poisson's ratio $\boldsymbol{\sigma}$ and $\mathbf{V_P}$ values using $h_{1/2}$ and $h_{1/10}$. As $\boldsymbol{\sigma}$ is a function of $\mathbf{V_P}$ and $\mathbf{V_S}$, the predominance of these parameters can be interpreted that the solution has also become more sensitive to the shear velocity values $\mathbf{V_S}$. The measure $h_{1/2}$ has also influenced to make the solution more sensitive to the density values $\boldsymbol{\rho}$, but the measure $h_{1/10}$ did not produce the same effect. In fact, the best solution generated by the VFSA using $h_{1/10}$ was

1 In the table 1, the line indicated by Σ% shows the values to the sum of the absolute deviations and by Max the maximum deviation.

not a good solution. The column named “total” summarizes the two statistics for all parameters. The measure $h_{1/2}$ provided the best result and $h_{1/10}$ the worst. Figure 3 compares graphically the best solution parameters with the original ones.

CONCLUSIONS

The SA method has proved to be a good method to provide solutions to the inverse problems that are highly nonlinear and are rife with local minima. In this experiment, the SA was applied to the solution of 1D seismic inverse problems and its performance was evaluated using error measures h_β. The $h_{1/2}$ produced the most sensitive solution to all parameters. The use of a low order norm $L_{1/2}$ attenuates the difference between the sample data and its estimates, increasing the convergence time of the SA and preventing the undue bias of the solution parameters. The $h_{1/10}$ did not show the same behavior. Similar results are obtained by Porsani (1993) using the method of Genetics Algorithms.

This study allows no generalization, since a rigorous statistical analysis was not done. However, it suggests certain questions. The decreasing of β-value to ½ has provided the most sensitive solution to all parameters, but β-value 1/10 has failed to produce a good solution. This fact raises the question whether exists an inferior limit to β-value in which the SA method using h_β would provide a sensitive solution to all parameters.

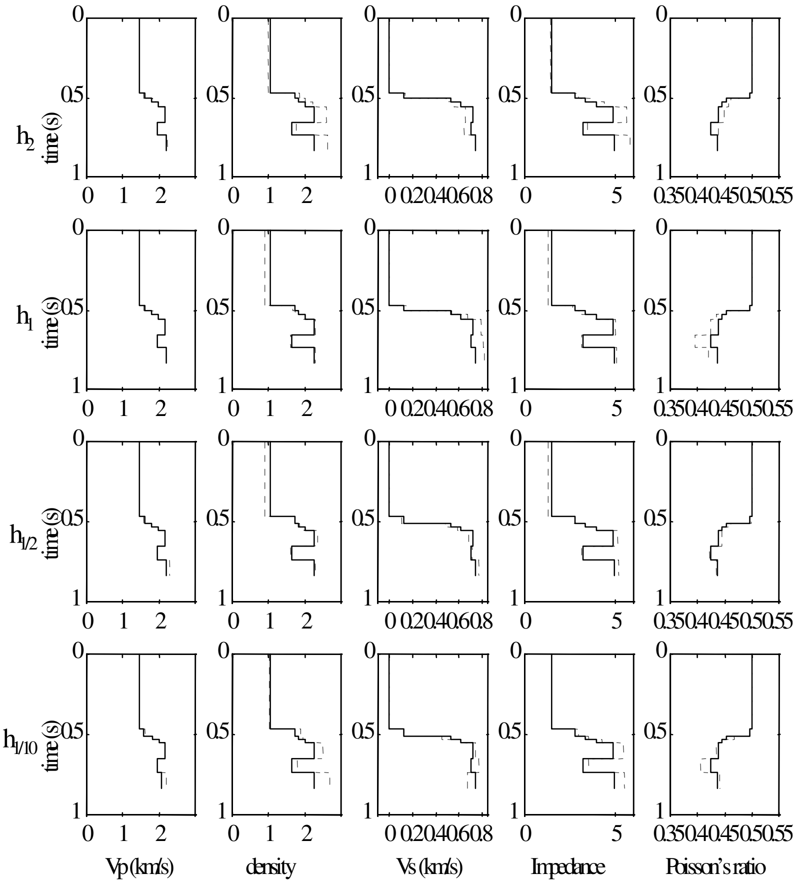

Figure 3 : The best solution parameters and the original parameters

The SA solution using h_2 has shown a high resolution $\mathbf{V_P}$, producing a good approximation to this parameter. However the use of β-value greater than 2 could provide an even better estimate to $\mathbf{V_P}$, closer to its true value.

The value of β has determined a trade-off between $\mathbf{V_P}$ resolution and model resolution. Error measure $h_{1/2}$ has provided a good estimate to $\mathbf{V_P}$, but not so good than the estimate produced by h_2. Nevertheless, $h_{1/2}$ has the advantage of producing an acceptable estimate to the acoustic impedance parameter **I**. A good approximation to the compressional wave velocity $\mathbf{V_P}$ is very important to the seismic processing steps such as NMO or migration and the acoustic impedance **I** is used in the reflectivity studies.

REFERENCES

Ingber, L., 1989, Very fast simulated re-annealing, Mathematical Computer Modeling, 12, 967-973.

Kirkpatrich, S., Gelatt, C.D., Vecchi, M. P., 1983, Optimization by Simulated Annealing, Science, 220, 4598, 671-680.

Mendonça, J.M., 1997, A Inversão Sísmica 1D usando o Simulated Annealing, MSc Thesis, Department of Electric Engineering , PUC-Rio.

Porsani, M.J., Stoffa, P.L., Chunduru, R.K.S., Mrinal, K., 1993, Evaluation of Measures of Error Using Genetic Algorithm, Third International Congress of Brazilian Geophysical Society, 1, 85-90.

ACKNOWLEDGMENTS

Jorge M. Mendonça would like to thank PETROBRAS for permission to publish this work.