Entropies of Shannon and Burg in the Inversion of Seismic Data

J. H. Quezada, UNEB, Brazil
A. Bassrei, Instituto de Física/UFBA and CPGG/UFBA, Brazil

ABSTRACT

In this paper we apply two maximum entropy (ME) approaches to the inversion of geophysical data with discrete data and discrete model parameters (Shannon, 1948; Burg, 1975). This formulation is based on a probabilistic philosophy and on the concept of entropy, where we do not make use of prior information. It is defined an objective function which contains the entropy of the model parameters, which is then minimized under adequate constraints in order to give the output estimate of the model parameters. The tests with synthetic data corrupted with noise in traveltime tomography show the promissing application of the ME approach for the solution of ill-posed inverse problems in geophysics. The results with ME are better, when compared to non-truncated singular value decomposition SVD. The Shannon's algorithm convergence is generally faster when compared to the Burg's ME method.

INTRODUCTION

The Newtonian physics, which dominated from the 17th to the end 19th century, described an universe in which everything should happen precisely in accordance with a law, and this universe was compact, organized, and all future should depend strictly on all past. However, we cannot check by mean of our imperfect experiments, if this or that set of physical laws is passible of verification up to the last decimal figure. The Newtonian conception nevertheless, had to present and express the physical process, as they had in fact, be subjected to those laws. As a consequence, from the middle of the last century a revolution with no foregoing begun in the history of physics. This revolution, based on the idea of a contingent universe, changed the physical thought, and now instead to state that some physical event will happen in any case, whatever the conditions, one states that the event will happen with an overwhelming probability.

The main idea of this process is the concept of entropy, which can be defined in several forms. With the increase of entropy, the universe, and all the closed systems, tend naturally to deteriorate and loose the clearness, to change from a state of minimum probability to another of maximum probability, from a state of organization and differentiation, in which exist forms and distintions, to a state of chaos. In the contingent universe, the order is less probable. The role of entropy is in such a way that Jaynes (1957), states that entropy is a primitive physical concept, even more fundamental than the concept of energy. The concept of entropy was developed by Rudolf Clausius from Germany, in the context of classical thermodynamics and later the Austrian physicist Ludwig Boltzmann gave the statistical interpretation of entropy. The concept of entropy was adopted from statistical mechanics when the Information Theory (IT) was founded by Shannon and Wiener in 1948 (Shannon, 1948; Wiener, 1961). Consider a source S emiting messages $m_1, m_2, \cdots, m_N$ with probabilities $p_1, p_2, \cdots, p_N$ respectively (where $p_1 + p_2 + \cdots + p_N = 1$). The information I_i carried by each message is given by

$$I_i = \log(\frac{1}{p_i}), \tag{1}$$

and entropy (H) is the average information of the source:

$$H(S) = \sum_{i=1}^{N} p_i I_i = -\sum_{i=1}^{N} p_i \log(p_i). \tag{2}$$

Burg (1975) introduced the following definition of entropy within the framework of Spectral Analysis:

$$H = \sum_{i=1}^{N} p_i \log(p_i). \tag{3}$$

It is interesting to find what is the distribution which maximizes the entropy. Since entropy is a measure of uncertainty, the probability distribution which generates maximum uncertainty will have ME. In the absence of prior information, Jaynes (1957), stated that the ME is the less biased estimate from a given information. In the context of prediction theory, the maximization of entropy is not the application of a physical law, but merely a reasoning method which guarantees that no arbitrary inconsistent assumption was used. The maximization of entropy has been very successful in a variety of applications including applied geophysics. Just to mention a few references, Rietsch (1988) applied ME to the inversion of 1-D seismograms, and Bassrei (1993) used a ME algorithm with continuous model parameters in traveltime tomography. A proof of the consistency of the principle of ME is given by Tikochinsky et al. (1984) where it is demonstrated that the ME distribution constrained to average values is the unique consistent induction from the data, for any reproducible experiment. In the present work no prior information is used. For underdetermined inverse problems where the prior information has a key role, an extension of ME, called minimum relative entropy can be used (Bassrei, 1990, 1991, 1994; Ulrych et al., 1990).

MAXIMUM ENTROPY ALGORITHM 1 (SHANNON)

The ME method allows different aproaches and implementations. For instante, it is possible to work with continuous model parameters (Bassrei, 1993) or discrete model parameters, which is the approach used in this work. When continuous model parameters are used, one deals with the probability density function of the model parameters, which are not always available in practice. One the other hand, using discrete model parameters with normalization there is a need to know the total sum of this vector (model parameters), in order to get the proportions. What we do here is to work directly with parameters instead of their probabilities, although this does not follow rigorously the entropy definition.

In the discrete case, we consider that we have M informations and N unknowns, so that the relation between data and model parameters is given by

$$d_j = \sum_{i=1}^{N} g_{ji} m_i, \qquad j = 1, \ldots, M. \tag{4}$$

The objective function is given by

$$\Phi(m_i) = -\sum_{i=1}^{N} m_i \log(m_i) + \sum_{j=1}^{M} \lambda_j \left(\sum_{i=1}^{N} g_{ji} m_i - d_j \right), \tag{5}$$

where the λ_j's are the Lagrange multipliers. Optimizing the above equation we have that

$$\frac{\partial \Phi}{\partial m_i} = -1 - \log(m_i) + \sum_{j=1}^{M} \lambda_j g_{ji} = 0, \tag{6}$$

which gives a solution for m_i:

$$m_i = e^{-1+\sum_{j=1}^{M} \lambda_j g_{ji}}. \tag{7}$$

Substituting the solution m_i in the equation (4) we have that

$$\frac{1}{e} \sum_{i=1}^{N} g_{ji} e^{-1+\sum_{j=1}^{M} \lambda_j g_{ji}} = d_j. \tag{8}$$

This is a non-linear set of equations that can be solved for instance by using the Newton-Raphson method. If we develop d_j through a Taylor's series, we obtain after the truncation,

$$d_j = d_j^0 + \left. \frac{\partial d_j}{\partial \lambda_l} \right|_{\lambda_l = \lambda^0} (\lambda_l - \lambda^0). \tag{9}$$

Defining

$$\Delta d_j^k = d_j^{obs} - d_j^k, \tag{10}$$

$$\Delta \lambda_l^k = \lambda_l^k - \lambda_l^{k-1}, \tag{11}$$

and the matrix

$$R_{jl}^k = \left. \frac{\partial d_j}{\partial \lambda_l} \right|_{\lambda_l = \lambda_l^k}, \tag{12}$$

we get the linearized expression

$$\Delta d_j^k = R_{jl}^k \Delta \lambda_l^k, \tag{13}$$

and finally

$$\Delta \lambda_l^k = \left(R_{jl}^k \right)^{-1} \Delta d_j^k. \tag{14}$$

The λ_j's are updated and are introduced in equation (8) in order to obtain the estimated model parameters from the ME method.

MAXIMUM ENTROPY ALGORITHM 2 (BURG)

The objective function is now given by

$$\Phi(m_i) = \sum_{i=1}^{N} \log(m_i) + \sum_{j=1}^{M} \lambda_j \left(\sum_{i=1}^{N} g_{ji} m_i - d_j \right), \tag{15}$$

where the λ_j's are the Lagrange multipliers. Optimizing the above equation we have that the solution m_i is given by

$$m_i = -\frac{1}{\sum_{j=1}^{M} \lambda_j g_{ji}}. \tag{16}$$

Following the same steps of the last section we can calculate the Lagrange multiplers and then have the model parameters from the equation (16).

NUMERICAL SIMULATIONS IN TOMOGRAPHY

The traveltime along a ray path from a point P (source) to a point Q (receiver) is given by the line integral

$$t = \int_P^Q \frac{dl}{v(\mathbf{x})} = \int_P^Q s(\mathbf{x}) dl, \tag{17}$$

where t is the traveltime, $v(\mathbf{x})$ is the velocity of wave propagation in **x**, dl is the differential length along the ray, and $s(\mathbf{x})$ is the slowness $[s(\mathbf{x}) = 1/v(\mathbf{x})]$. Since the problem is non-linear, it is necessary to linearize it, expanding the function t by a Taylor's series, and considering only the linear term. This yields

$$\Delta \mathbf{t} = \mathbf{G} \Delta \mathbf{s}, \tag{18}$$

where $\Delta \mathbf{t}$ is the traveltime residual resulting from the perturbation, $\Delta \mathbf{s}$ is the slowness distribution residual,

Sixth International Congress of the Brazilian Geophysical Society

and **G** is a matrix which represents the distances between sources and receivers.

In our synthetic model (Figure 1) the medium is discretized in 100 blocks, where the velocity of the homogeneous model is 2000 m/s. There are two heterogeneous features: a low velocity layer of 1700 m/s which represents a negative constrast of velocity of 15%, and a limited body of 2300 m/s which by its turn is a positive constrast of velocity, also of 15%. The velocity contrast is rather low in order to avoid many ray tracing loops during the inverse procedure. The ray tracing procedure is based on Schots (1990). We considered a well-to-well acquisition geometry, with 10 sources in the left hand side well and 10 receivers in the right hand side well. This implies in a determined problem, that is, there are 100 equations and 100 unknowns. However due to the ill-posedness of the problem, the rank of the tomographic matrix is far from 100! In fact, Figure 2 shows the SVD inversion, using all singular values. We still can see the general structure of the model, but the image also generated negative values for velocity! Truncating some values one obtains a good result, but the decision of how many singular values are worthy to be used is rather subjective. The inversion with ME (Burg) is showed in Figure 3, where we can see that the image is much better and is very close to the true model. For the ME inversion we also need to invert a matrix in each iteration for the update of Lagrange multipliers. In order to make a better comparison we also used SVD for this purpose, but in this case with all singular values, or at least with the maximum number of singular values that still allowed the convergence of the ME algorithm. Thus, for the SVD application inside the ME inversion, there is no a demand for a subjective criterion concerning the selection of singular values. Figure 4 shows the convergence speed for the two entropies, where $\lambda = 0.01$ in both algorithms. Note that the convergence is reached in 10 iterations for Shannon's entropy but 100 iterations are not enough for the Burg's one. It is important to mention that the initial value for the Lagrange multipliers is not crucial, that is, they do not have a role like prior information. In Figures 5 and 6 it is possible to see the algorithm convergence, which is allways obtained, for different initial values of the Lagrange multipliers.

CONCLUSIONS

It was shown a stochastic technique based on a ME principle to invert geophysical data. The estimate is the solution of the minimization problem, and is consistent with the input data. The ME solutions showed to be generally superior than SVD. The Shannon's algorithm converges generally faster than the Burg's one. The present technique can be applied to underdetermined systems (the case of the geophysical inverse problems in general) as well as to determined and overdetermined systems.

ACKNOWLEDGMENTS

J. Quezada acknowledges PETROBRÁS for a M.Sc. scholarship. A. Bassrei acknowledges the Brazilian agency CNPq for a research fellowship (300.209/90-8).

REFERENCES

Bassrei, A., 1990. Inversion of seismic data by a relative entropy algorithm. 60th Annual International SEG Meeting, San Francisco, USA. **II**, 1111–1114.

Bassrei, A., 1991. Inversion of 2-D gravity data by a relative entropy algorithm. 53rd EAEG Meeting, Florence, Italy, 410–411.

Bassrei, A., 1993. Inversion of seismic data by a maximum entropy algorithm. 3rd International Congress of the Brazilian Geophysical Society, Rio de Janeiro, Brazil. **I**, 107–110.

Bassrei, A., 1994. Inversion of the Earth radial density distribution by an information theory approach, Brazilian Journal of Geophysics, **12**, 121–134.

Burg, J. P., 1975. Maximum Entropy Spectral Analysis. Ph.D. Thesis, Stanford University.

Jaynes, E. T., 1957. Information theory and statistical mechanics. Physical Review, **106**, 620–630.

Rietsch,E., 1988. Maximum entropy approach to the inversion of one-dimensional seismograms. Geophysical Prospecting, **36**, 365–382.

Shannon, C. E., 1948. A mathematical theory of communication. Bell System Technical Journal, **23**, 379–423, 623–656.

Schots, H. A., 1990. Well-to-well and Well-to-surface Seismic Tomography using Direct Waves (in Portuguese). M.Sc. Dissertation, CPGG/UFBA.

Tikochinsky. Y., Tishby, N. Z., and Levine,R. D., 1984. Consistent inference of probabilities for reproductible experiments. Physical Review Letters, **52**, 1357–1360.

Ulrych, T. J., Bassrei, A., and Lane, M., 1990. Minimum relative entropy inversion of one-dimensional data with applications, Geophysical Prospecting, **38**, 465–487.

Wiener, N., 1961. Cybernetics, 2nd ed. MIT Press, Cambridge, MA.

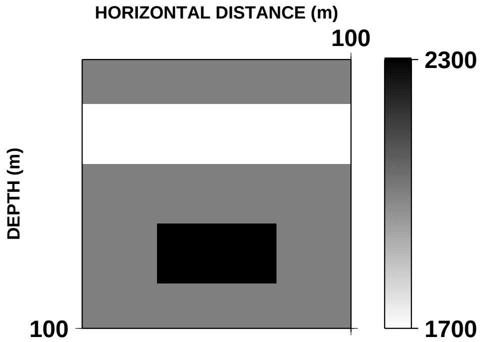

Figure 1: True model, velocity in m/s.

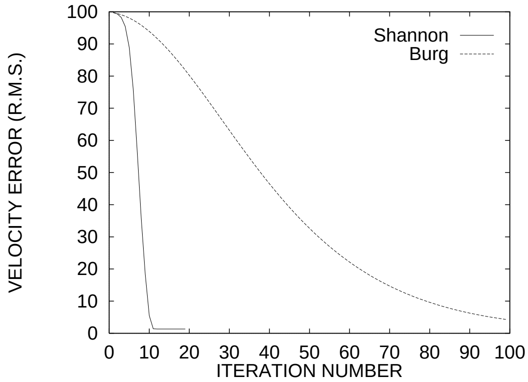

Figure 4: Convergence curves for Shannon and Burg entropies ($\lambda = 0.01$).

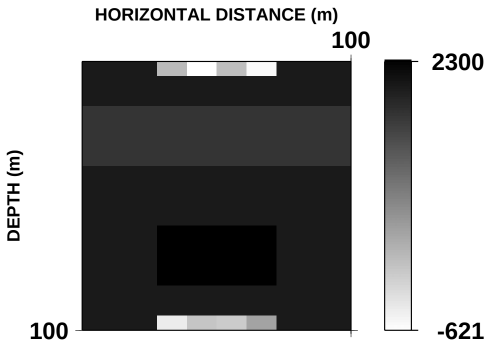

Figure 2: SVD reconstruction using all singular values.

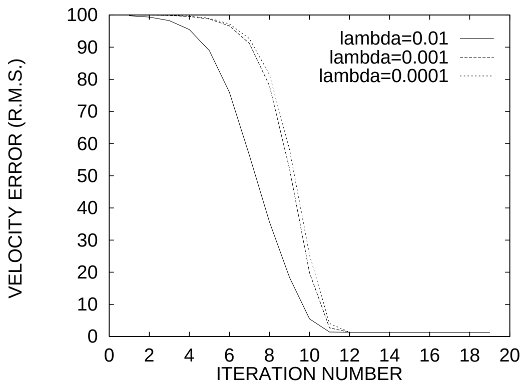

Figure 5: Convergence curves for different initial values of the Lagrange multipliers (Shannon).

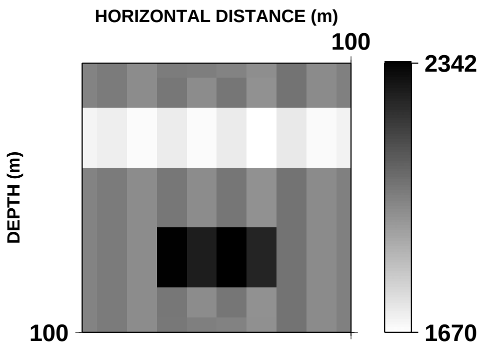

Figure 3: Maximum entropy reconstruction.

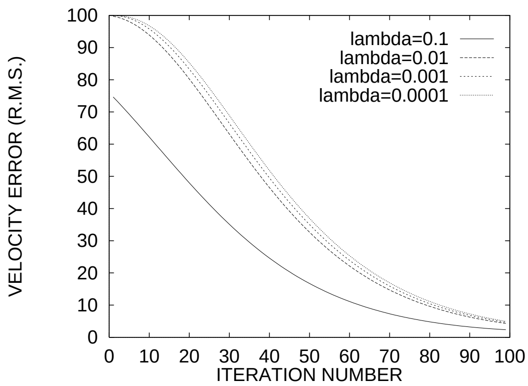

Figure 6: Convergence curves for different initial values of the Lagrange multipliers (Burg).