Maximum Entropy Source Waveform Estimation

by Jon F. Claerbout

Examples of the geometric inequality (see Geometric Programming, Duffin, Peterson and Zenor, John Wiley & Sons, New York, 1967) are:

$$\frac{1}{2} v_1 + \frac{1}{2} v_2 \geq (v_1 v_2)^{1/2} \tag{1a}$$

$$\frac{3}{4} v_1 + \frac{1}{4} v_2 \geq v_1^{3/4} \; v_2^{1/4} \tag{1b}$$

$$\frac{1}{3} v_1 + \frac{1}{3} v_2 + \frac{1}{3} v_3 \geq (v_1 v_2 v_3)^{1/3} \tag{1c}$$

The general case is

$$\sum_{i=1}^{N} w_i v_i \geq \prod_{i=1}^{N} v_i^{w_i} \tag{2}$$

with the restrictions

$$v_i, w_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{N} w_i = 1 \tag{3}$$

The geometric inequality becomes an equality if and only if all the v_i are identical.

Let us define

$$p_i = w_i v_i \geq 0$$

and specialize to the case

$$w_i = \frac{1}{N} \qquad \text{(all } i \text{)} \tag{5}$$

Then (2) becomes

$$\Sigma p_i \geq \Pi \left(\frac{p_i}{w_i} \right)^{w_i} = \Pi (N p_i)^{1/N} \tag{6}$$

Take the exponential of the logarithm of the right hand side

$$\Sigma\, p_i \;\geq\; \exp\, \ln\, \Pi\, (N\, p_i)^{1/N} \;=\; \exp\, \Sigma\, \frac{1}{N}\, \ln\, N\, p_i \tag{7}$$

Because of the inequality we can define a coefficient

$$c \;=\; \frac{\exp\, \Sigma\, \frac{1}{N}\, \ln\, N\, p_i}{\Sigma\, p_i} \;\leq\; 1 \tag{8}$$

which, like a coherency, is upper bounded by unity. The numerical value of c is independent of the scaling of p_i. The important property of c is that the larger it is, the more similar are the p_i. Alternatively, we can discuss I, the negative logarithm of c, namely

$$I \;=\; -\ln c \;=\; \ln\, \Sigma\, p_i - \frac{1}{N}\, \Sigma\, \ln\, N\, p_i \;\geq\; 0 \tag{9}$$

The smaller I, the more nearly equal are the p_i. What other mathematical functions have this property? One is the sample variance, namely

$$\hat{\sigma} \;=\; \Sigma\, (\, v_i - \frac{1}{N}\, \Sigma\, v_i\,)^2 \;\geq\; 0 \tag{10}$$

Of course (10) is not independent of the scale of the v_i, nor is it particularly constrained to have the v_i positive. A less familiar function σ' which is more similar in behavior to (9) would be

$$\sigma' \;=\; \Sigma\, \left|\, \ln\, v_i - \frac{1}{N}\, \Sigma\, \ln\, v_i\, \right|^n \;\geq\; 0 \tag{11}$$

It turns out (accidentally?) that deconvolution, the industry standard technique for elimination of shallow water multiples, results from minimization of (9). (So does Burg's maximum entropy spectral analysis.) The basic model of such a seismogram X is that it is made of random reflection coefficients convolved with a highly resonant shallow water filter. The idea is that a filter A should be determined so that the output power spectrum

$$P = A^* X^*X A = A^* R A \tag{12}$$

tends toward a constant. The filter A then tends toward the inverse of the shallow water resonance function. Instead of the summation in (9) one has an integration over frequency so (9) becomes

$$I = \ln \int_0^{2\pi} P(\omega)\, d\omega - \frac{1}{2\pi} \int_0^{2\pi} \ln [\, 2\pi\, P(\omega)\,]\, d\omega \geq 0 \tag{13}$$

Let us compute the derivative of I with respect to A*

$$\frac{\partial I}{\partial A^*} = \frac{\int \frac{dP}{dA^*}\, d\omega}{\int P\, d\omega} - \frac{1}{2\pi} \int \frac{1}{P} \frac{\partial P}{\partial A^*}\, d\omega \tag{14}$$

Setting this derivative equal zero cannot be expected to determine a scaling factor in A because I was defined to be independent of such a scaling factor. Let us choose the scaling factor by the auxiliary condition

$$2\pi = \int P\, d\omega \tag{15}$$

Then (14) reduces to

$$\frac{\partial I}{\partial A^*} = \int [\, \frac{\partial P}{\partial A^*} - \frac{1}{P} \frac{\partial P}{\partial A^*}\,]\, d\omega \tag{16}$$

Inserting the derivative of (12) into (16) we get

$$\frac{\partial I}{\partial A^*} = \int [\, R A - \frac{1}{A^*}\,]\, d\omega \tag{17}$$

Now we presume that A can be chosen minimum phase so that we have convergence on the unit circle for the expansion

$$\frac{1}{A^*(Z)} = b_0 + b_1/Z + b_2/Z^2 + \dots = B^*(1/Z) \tag{18}$$

Since the integral around the unit circle in (17) selects the coefficient of Z^0 we see, as in SEP-1, p. 258 that choosing zero for the derivative

$$\frac{\partial I}{\partial a_k^*} = \frac{\partial I}{\partial A^*} \frac{A}{\partial a_k^*} = \frac{\partial I}{\partial A^*} Z^{-k}$$

leads to the familiar Toeplitz equations.

This method has been tried also on deep water multiple reflections but it doesn't work. Perhaps the reason is that the model is so far from reality. A better model for the elimination of deep water multiple reflections is the Noah model (see SEP-3). In this model most practical multiple reflection problems arise from the free surface reflection. One would like to observe the upcoming wave U if the downgoing wave D were an isolated impulse, but the free surface returns the upcoming wave back down as an augmented downgoing wave. Thus, ideally,

$$U = \frac{U}{D} = \frac{R'}{1+R'} \tag{19}$$

where R' is the seismogram with multiples. A practical problem is that R' is not observed directly. It is seen through some source waveform window $B(Z)$, that is, we observe

$$Q(Z) = B(Z)\, R'(Z) = R'(Z) \,/\, A(Z) \tag{20}$$

and we imagine that the spectrum of random reflection coefficients belongs to

$$U = \frac{Q\,A}{1 + Q\,A} \tag{21}$$

Then

$$P = U^*U = \frac{A^* Q^*}{1+A^*Q^*} \frac{Q A}{1+Q A} \qquad (22)$$

and

$$\frac{\partial P}{\partial A^*} = \left[\frac{1}{A^*} - \frac{Q^*}{1 + A^* Q^*} \right] P \qquad (23)$$

This can easily be inserted into the gradient expression (16), but the result is not a linear expression in the unknown A. Thus, we won't have an analytic solution.

Nevertheless, some sort of numerical approach to the problem might be quite reasonable. The frequency axis could be divided into zones within which $p(\omega_i)$ fluctuates but A is considered a complex constant. First we could easily compute I at many locations in the complex A plane to see if there is any real danger of multiple minima. If there is, the formulation may be in serious trouble. If there isn't, the solution should be easy to find by any kind of gradient descent procedure. In this the descent is

$$d\,I = \frac{\partial I}{\partial A^*} dA^* + \frac{\partial I}{\partial A} dA$$

Alternatively we could abandon the geometric inequality and seek another such that we would get linear equations.