Using non-Gaussianity as an inversion constraint

William Harlan

CHOOSING A FUNCTIONAL

Minimum and maximum entropy methods succeed in simplifying the results of numerous geophysical inversion problems because they either drive random variables toward or away from Gaussianity. This property is not well recognized; as a result, the usual functionals are not optimum. The varimax norm makes some particularly unfortunate assumptions: it can only say whether a given probability density function (pdf) resembles one generalized gaussian more than another. Shannon's entropy does not take account of histogram sampling rates. Least-squares functionals measure only the distance of an amplitude distribution from a Gaussian with *fixed* amplitude and standard deviation. The L1 norm similarly measures the distance of a distribution from an exponential pdf with fixed decay.

Harlan, Claerbout, and Rocca (1984) derived a functional which measures the non-Gaussianity of a pdf. The functional does not compare a specific Gaussian as in least-squares, but rather measures the quality of fit of a pdf with the best fitting Gaussian. This paper will add the ingredients required to use this functional in a non-linear optimization program. I refer readers to the cited paper to understand why one should encourage or discourage non-Gaussianity in inverted parameters.

The development of this paper follows the assumptions that encourage non-Gaussianity. Signal spaces should be chosen so as to be independent, identically distributed (IID) processes. (For Gaussian variables, transformation should diagonalize and balance the covariance matrix.) Thus, information may be contained in a minimum of parameters, with a maximum of non-Gaussianity. The contrary case, with heavy linear summing of random variables, produces Gaussianity. One may encourage Gaussianity in an inversion with the following development, but only by ignoring the accompanying statistical dependence between samples. Such assumptions motivate Burg's maximum entropy spectral analysis, which attempts to maximize Gaussianity in the heavily mixed Fourier transform domain.

First, I shall construct estimates of the functional in terms of values of the random variables. Second, I shall derive the gradient for use in non-linear optimization. The measure will then be in a form usable as a penalty or barrier function. Penalty functions require scaling with respect to other terms in the functional to be optimized. Appropriate scales can only be judged by the interpretability of the inverted result. If *a priori* information allow, a penalty function could be given a fixed barrier value. If not, then several barriers could be tried and the result judged on its interpretability.

CONSTRUCTION OF THE MEASURE

Let us define a histogram $\{q_i\}$ of an array of random variables $\{x_i\}$ with dimension N.

$$q_j \equiv \frac{1}{N}\sum_i^N \Lambda\left(\frac{x_i - \overline{x}}{\Delta} - j\right) \tag{1}$$

where

$$\overline{x} \equiv \frac{1}{N}\sum_i x_i \quad ; \Delta \equiv \frac{\sigma}{M} \; ; \; \sigma^2 \equiv \frac{1}{N}\sum_i (x_i - \overline{x})^2 \tag{2}$$

M is the number of histogram samples per standard deviation σ. Δ is the distance between amplitude samples. Λ is the characteristic function of the histogram. The simplest choice yielding a usable derivative is the triangle function.

$$\Lambda(x) \equiv \begin{cases} 1-|x| & -1<x<1 \\ 0 & elsewhere \end{cases} \tag{3}$$

Harlan, Claerbout, and Rocca (1984) derived the following measure of non-Gaussianity for a continuous pdf (p[x]).

$$F\{p(x)\} \equiv \int p(x)\log p(x) + \frac{1}{2}\log\int (x - \int x'p(x')dx')^2 p(x)dx \tag{4}$$

The simplest continuous estimate $\hat{p}(x)$ derived from q_i is a step function:

$$\hat{p}(x) \equiv \frac{1}{\Delta}\sum_j q_j \,\Pi\left(\frac{x - \overline{x}}{\Delta} - j\right) \tag{5}$$

We choose the boxcar as the characteristic function.

$$\Pi(x) \equiv \begin{cases} 1 & -1/2<x<1/2 \\ 0 & elsewhere \end{cases} \tag{6}$$

Now we can write the equivalent measure of non-Gaussianity for the histogram.

$$F\{\hat{p}(x)\} = \sum_j q_j \log \frac{1}{\Delta} q_j + \log\sigma = \sum_j q_j \log q_j + \log M \tag{7}$$

CALCULATION OF THE GRADIENT

To maximize or minimize the Gaussianity by conventional non-linear descent methods, we must have the Frechet derivative of F with respect to the random variables $\{x_i\}$.

$$\frac{\partial F}{\partial x_l} = \sum_j \frac{\partial q_j}{\partial x_l}(\log q_j + 1) \tag{8}$$

$$\frac{\partial q_j}{\partial x_l} = \frac{1}{N}\sum_i^N \Lambda'(\frac{x_i - \overline{x}}{\Delta} - j)\frac{\partial}{\partial x_l}(\frac{x_i - \overline{x}}{\Delta}) \tag{9}$$

$$\frac{\partial}{\partial x_l}(\frac{x_i - \overline{x}}{\Delta}) = \frac{1}{\Delta}\frac{\partial}{\partial x_l}(x_i - \overline{x}) - \frac{(x_i - \overline{x})}{\Delta^2}\frac{\partial \Delta}{\partial x_l} \tag{10}$$

$$\frac{\partial}{\partial x_l}(x_i - \overline{x}) = (\delta_{i-l} - \frac{1}{N}) \tag{11}$$

$$\frac{\partial \Delta}{\partial x_l} = \frac{1}{M}\frac{\partial \sigma}{\partial x_l} = \frac{1}{M}\frac{\partial}{\partial x_l}(\frac{1}{N}\sum_k^N (x_k - \overline{x})^2)^{\frac{1}{2}} \tag{12}$$

$$= \frac{1}{MN\sigma}\sum_k^N (x_k - \overline{x})(\delta_{lk} - \frac{1}{N}) = \frac{1}{MN\sigma}(x_l - \overline{x})$$

$$\Lambda'(x) = \Pi(x + \frac{1}{2}) - \Pi(x - \frac{1}{2}) \tag{13}$$

Equation (10), with (11) and (12), becomes

$$\frac{\partial}{\partial x_l}(\frac{x_i - \overline{x}}{\Delta}) = \frac{1}{\Delta}[\delta_{il} - \frac{1}{N} - \frac{1}{N\sigma^2}(x_i - \overline{x})(x_l - \overline{x}) \tag{14}$$

Equation (8) with (9), (13), and (14) becomes

$$\frac{\partial F}{\partial x_l} = \frac{1}{N\Delta}\sum_j \sum_i^N \{[\Pi(\frac{x_i - \overline{x}}{\Delta} - j + \frac{1}{2}) - \Pi(\frac{x_i - \overline{x}}{\Delta} - j - \frac{1}{2})] \cdot \tag{15}$$

$$\cdot [\delta_{il} - \frac{1}{N} - \frac{1}{N\sigma^2}(x_i - \overline{x})(x_l - \overline{x})](\log q_j + 1)\}$$

or more simply,

$$\frac{\partial F}{\partial x_l} = \frac{1}{N\Delta}\{[\log q^+(x_l) - \log q^-(x_l)] \tag{16}$$

$$+\frac{(x_l - \overline{x})}{N\sigma^2}\sum_i^N[\log q^+(x_l) - \log q^-(x_l)](x_i - \overline{x})\}$$

where

$$q^+ \equiv \sum_j q_j\,\Pi(\frac{x - \overline{x}}{\Delta} + \frac{1}{2} - j) \qquad (17)$$

$$q^- \equiv \sum_j q_j\,\Pi(\frac{x - \overline{x}}{\Delta} - \frac{1}{2} - j)$$

Let us define the histogram samples to the left and right of an amplitude x as

$$ileft(x) \equiv int(\frac{x - \overline{x}}{\Delta})\ ;\ iright(x) \equiv ileft(x) + 1 \qquad (18)$$

$int(x)$ takes the integer part of x. Equivalently,

$$q^+(x) \equiv q_{iright(x)}\ ;\ q^-(x) = q_{ileft(x)} \qquad (19)$$

During a single iteration, compute Δ and the histogram array $\{q_i\}$, then measure the non-Gaussianity from (7). To find the gradient with respect to one random variable x_l, find the histogram samples above and below its amplitude (determined by equations [18] and [19]) and plug them into equation (16). Note that the logarithm need only be calculated once for each sample.

CONVERGENCE

Unless the histogram $\{q_i\}$ is coarsely sampled, the characteristic function $\Lambda(x)$ will be too narrow in the early iterations for decent convergence. The gradient will more quickly adjust the overall shape of the function pdf if $\Lambda(x)$ is broadened in early iterations and then narrowed. A unique minimum will not change; only the convergence rate.

REFERENCES

Harlan, W.S., Claerbout, J.F., and Rocca, F., 1984, Signal/noise separation and velocity analysis: Geophysics, 49, 1869-1880.