A THEORY FOR VARIABLE NORM DECONVOLUTION

William Gray

An iterative, multichannel deconvolution technique is developed using statistical methods. It is based on the reflectivity series being characterized by the generalized Gaussian family of probability distributions. The deconvolution filter is estimated by maximizing a measure of the probability that the resulting reflectivity series is a specified member of the family of distributions.

Background

Recently Wiggins (1977, pp. 7-14) presented a powerful deconvolution technique called *minimum entropy deconvolution (MED)*. It was based upon iteratively estimating an inverse filter such that a norm ratio which measures the spikiness of the deconvolved time series is maximized.

Claerbout (SEP-13, pp. 1-9) extended Wiggin's work with the introduction of *parsimonious deconvolution*. Intuitively, Claerbout felt that Wiggin's technique was excessively biased toward the higher amplitude events on a seismogram and a method which "sees" more of the data would result in a better deconvolution.

Variable norm deconvolution (SEP-14, pp. 181-205) was presented in the last SEP report. It was essentially a generalization of Wiggin's work. An algorithm was derived and implemented in the frequency domain. Excellent results using synthetics and field recorded data were presented. In the same report Godfrey (SEP-14, pp. 207-225) extended Claerbout's earlier work.

Since the last SEP report a statistical deconvolution theory has been worked on by Claerbout, Godfrey, Gray, and Muir. The theory has led to an understanding which has spawned new and more powerful deconvolution methods.

Generalized Gaussian distribution

The reflection seismogram can be modeled as the output of a linear filter, given the earth's reflectivity series as input. The linear filter generally represents the source waveform and other undesirable averaging effects. Deconvolution is an attempt to estimate the inputs given the outputs. To solve the problem some property of the filter or inputs must be assumed.

Here it is assumed reflection amplitudes and averaging effects are well described by members of a family of symmetrical probability distributions defined by

$$f(X;\beta,\alpha) = \frac{\alpha}{2\beta\Gamma(\frac{1}{\alpha})} e^{-\left(\frac{|x|}{\beta}\right)^{\alpha}} \qquad (1)$$

where:

$-\infty < x < \infty$ is a random reflection amplitude;

$\Gamma(\cdot)$ is the gamma function;

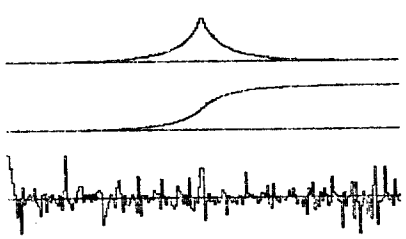
$\beta > 0$ is the scale parameter; and

$\alpha > 0$ is the shape parameter.

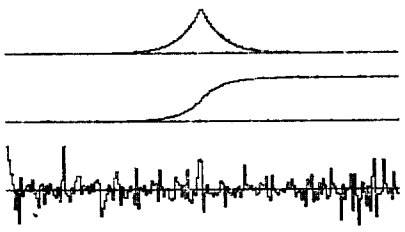
This family covers a wide range of distributions. The certain event, double exponential, Gaussian, and uniform distributions are all members. Figure 1 displays members as a function of the shape parameter.

The family will be called the *generalized Gaussian* after Miller and Thomas (1972, pp. 241-250) who used it in detection theory as a model for non-Gaussian noise. The statisticians initially called it *Subbotin's distribution* (1923, pp. 296-301). It was rarely used until recently when Box and Tiao (1973, pp. 149-241) examined its properties as a prior distribution for Bayesian analysis. Regression models based on the family have been investigated by Goodman and Kotz (1971).

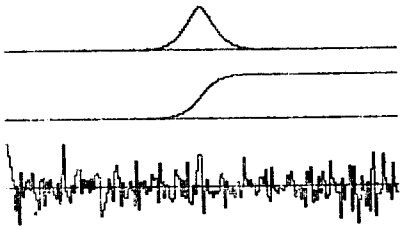
According to Algazi and Lerner (1969) the family for $.1 < \alpha < .6$ is representative of impulsive atmospheric processes. Using the Kolmogorov-Smirnov test Godfrey (SEP-15, p.157) has shown the family to be fairly representative of both reflection seismograms and reflectivity series.



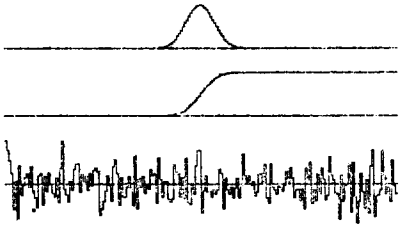
$\alpha = .8$



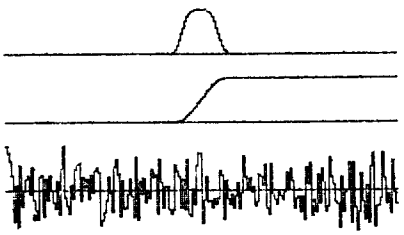
$\alpha = 1.$
(double exponential)



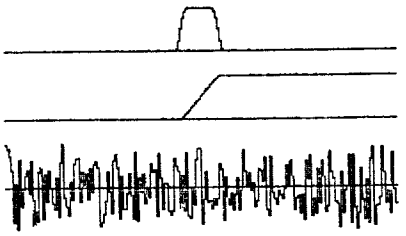
$\alpha = 1.5$



$\alpha = 2.$
(Gaussian)



$\alpha = 4.$



$\alpha = 8.$

FIGURE 1.--Members of the generalized Gaussian family of distributions. For each shape parameter the probability density function is plotted above the cumulative distribution and a representative random sample. The shape parameter determines the rate of exponential decay for the pdf.

Typically gained seismograms will have a shape parameter in the range (1.3,2.5). Reflectivity series, being more spiky, are generally in the range (.4,1.5). The family is believed to fairly represent seismic data because of its "peakness" at the origin for $\alpha \leq 1$.

Properties of the generalized Gaussian

Several properties and estimators for the generalized Gaussian distribution are derived now for later use.

For fixed variance the family of distributions can be parameterized by Shannon's entropy. Let entropy be defined by

$$H(\beta,\alpha) = -\int_{-\infty}^{\infty} f(X;\beta,\alpha) \log f(X;\beta,\alpha)\, dx \tag{2}$$

where f is the distribution given by Equation (1). Performing the integration yields

$$H(\beta,\alpha) = \log\left(\frac{2\beta\Gamma\left(\frac{1}{\alpha}\right)}{\alpha}\right) + \frac{1}{\alpha} \tag{3}$$

This function is shown in Figure 2 for constant scale.

The λ-th absolute moment is derived from

$$E(|x|^{\lambda}) = \int_{-\infty}^{\infty} |x|^{\lambda} f(X;\alpha,\beta)\, dx \tag{4}$$

as

$$E(|x|^{\lambda}) = \frac{\Gamma\left(\frac{\lambda+1}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \beta^{\lambda} \tag{5}$$

The variance, σ^2, is the second absolute moment,

$$\sigma^2 = \frac{\Gamma\left(\frac{3}{\alpha}\right)}{\Gamma\left(\frac{1}{\alpha}\right)} \beta^2 \tag{6}$$

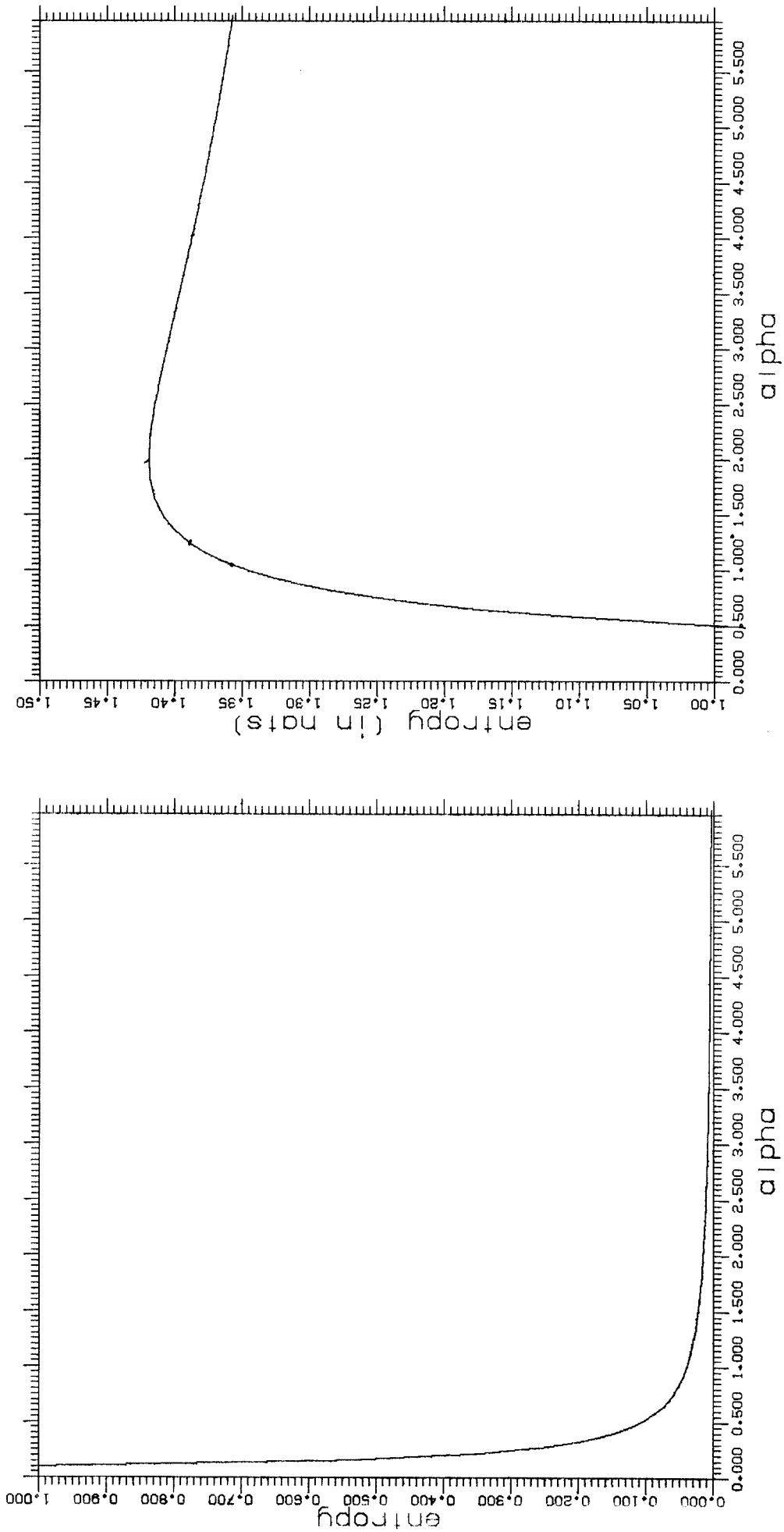


FIGURE 2.--The entropy of the generalized Gaussian plotted for varying shape parameter α. On the left the scale was held fixed, $\beta = .5$. On the right the variance was held fixed, $\sigma^2 = 1$.

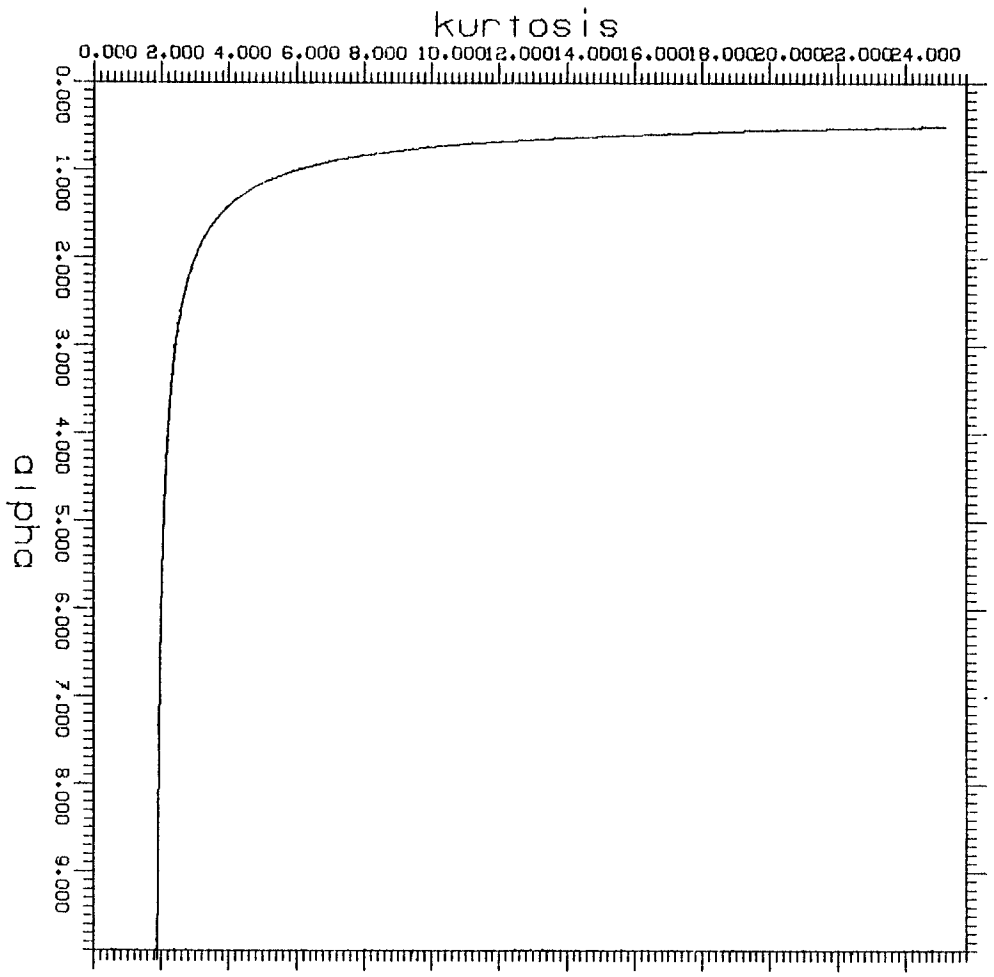


FIGURE 3.--The generalized Gaussian family is ordered by kurtosis. $K(2;X)$ is plotted versus α, the shape parameter.

Substituting this result in Equation (3) gives entropy as a function of variance and the shape parameter,

$$H(\alpha,\sigma^2) = \log\left(\frac{2}{\alpha}\sqrt{\frac{\sigma^2\Gamma\left(\frac{1}{\alpha}\right)^3}{\Gamma\left(\frac{3}{\alpha}\right)^2}}\right) + \frac{1}{\alpha} \tag{7}$$

This function is shown in Figure 2 for different members of the family with constant variance. The Gaussian has maximum entropy as expected. Entropy decreases on both sides. As $\alpha \to 0$ it rapidly goes to $-\infty$ which is the entropy of the certain event. On the other side it falls off slowly to another minimum, the uniform distribution.

Kurtosis is traditionally defined as the ratio of the fourth sample moment and second sample moment squared.

$$K = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^4}{\left(\frac{1}{n}\sum_{i=1}^{n} x_i^2\right)^2} \tag{8}$$

It has long been used as a statistic to describe the tailweight of a random sample and for testing departures from normality.

"Generalized kurtosis," $K(\rho,X)$, is also a measure of tailweight. It is defined as

$$K(\rho;X) = \frac{\frac{1}{n}\sum_{i=1}^{n} |x_i|^{2\rho}}{\left(\frac{1}{n}\sum_{i=1} |x_i|^{\rho}\right)^2} \tag{9}$$

This statistic is useful here because it provides an efficient scale invariant method of estimating a shape parameter best describing a given random sample. Using Equation (5) gives

$$\frac{E(|x|^{2\rho})}{[E(|x|^{\rho})]^2} = \frac{\Gamma\left(\frac{2\rho+1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{\rho+1}{\alpha}\right)^2} \tag{10}$$

Combining Equations (9) and (10) gives

$$\frac{\frac{1}{n}\sum_{i=1}^{n}|x_i|^{2\rho}}{\left(\frac{1}{n}\sum_{i=1}^{n}|x_i|^{\rho}\right)^2} = \frac{\Gamma\left(\frac{2\rho+1}{\alpha}\right)\Gamma\left(\frac{1}{\alpha}\right)}{\Gamma\left(\frac{\rho+1}{\alpha}\right)^2} \tag{11}$$

Solving this equation for α is probably impossible but in a computer the RHS can be evaluated for many values of α and saved in a table. The LHS is computed from the random sample and a shape parameter determined by interpolation on the table. Figure 3 is a plot of $K(2;X)$ versus α.

$K(\rho;X)$ is a family of estimators. Geary (1947, pp. 209-242) has shown that $K(2;X)$ is most efficient when no information on the distribution of the random sample is available. For seismic data we have information on the expected distribution and from Monte Carlo testing we have determined that $K(1;X)$ is a more efficient estimator than $K(2;X)$. Box (1953, pp. 465-468) also concluded that $K(1;X)$ would be a better estimator for a finite sample having greater kurtosis than the Gaussian.

The log likelihood for the generalized Gaussian is defined as

$$L = \log \prod_{i=1}^{n} f(x_i;\beta,\alpha)$$

$$= \log \alpha^n - \log\left[2\beta\Gamma\left(\frac{1}{\alpha}\right)\right]^n - \frac{\sum_{i=1}^{n}|x_i|^{\alpha}}{\beta^{\alpha}} \tag{12}$$

The maximum likelihood estimator for the scale parameter, β, is the solution of

$$\frac{\partial L}{\partial \beta} = 0 \tag{13}$$

Solving the equation gives the estimator

$$\beta = \left(\frac{\alpha}{n} \sum_{i=1}^{n} |x_i|^{\alpha} \right)^{\frac{1}{\alpha}} \tag{14}$$

This yields the mean deviation in the case of the double exponential ($\alpha = 1$). For the normal distribution ($\alpha = 2$), β^2 is estimated by twice the square of the standard deviation and for the rectangular ($\alpha \to \infty$), β is estimated by one-half of the sample range.

Derivation of norm ratio

In statistics the most powerful test of the null hypothesis against another is one which maximizes discrimination and minimizes *Type II error*. Type II error is the probability of accepting the null hypothesis as true when actually the alternative is true. In information theory a statistic that maximizes detectability of a signal while minimizing the probability of a false alarm is most powerful.

Hogg (1972, pp. 422-424) develops a scale-invariant, most powerful test for one member of the generalized Gaussian against another. He considers the family without a scale parameter

$$f(X;\alpha) = \frac{\alpha}{2\Gamma\left(\frac{1}{\alpha}\right)} e^{-|x|^{\alpha}} \tag{15}$$

and uses a procedure described by Hájek and Šidák (1967, pp. 107-108) to derive the test. From the theory of rank statistics the most powerful test for a symmetrical distribution which is scale-invariant is found from the integral

$$I(\alpha) = \int_0^{\infty} \prod_{i=1}^{n} f(\beta x_i;\alpha)\, \beta^{n-1}\, d\beta \tag{16}$$

This is the marginal distribution of the likelihood of α obtained by integrating out the scale. One degree of freedom is omitted because the distribution is assumed to be symmetrical.

Substituting the density into the integral gives,

$$I(\alpha) = \int_0^{\infty} \left[\frac{1}{2\Gamma\left(\frac{\alpha+1}{\alpha}\right)}\right]^n e^{-\left(\beta^{\alpha}\sum_{i=1}^{n}|x_i|^{\alpha}\right)} \beta^{n-1}\, d\beta \tag{17}$$

Using a change of variables

$$z = \beta^{\alpha}\sum_{i=1}^{n}|x_i|^{\alpha} \tag{18}$$

the integral is of the form

$$I(\alpha) = c(\alpha,n)\left(\frac{1}{n}\sum_{i=1}^{n}|x_i|^{\alpha}\right)^{-n/\alpha} \tag{19}$$

where

$$c(n,\alpha) = 2^{-n}\frac{\Gamma\left(\frac{n}{\alpha}\right)}{\left[\Gamma\left(\frac{1}{\alpha}\right)\right]^n}\alpha^{n-1}\, n^{-n/\alpha} \tag{20}$$

To determine if a random sample X is drawn from the distribution $f(X;\alpha_2)$ as opposed to $f(X;\alpha_1)$, a ratio of tests is used

$$V(X;\alpha_1,\alpha_2) = \frac{\left(\frac{1}{n}\sum_{i=1}^{n}|x_i|^{\alpha_2}\right)^{-n/\alpha_2}}{\left(\frac{1}{n}\sum_{i=1}^{n}|x_i|^{\alpha_1}\right)^{-n/\alpha_1}} \tag{21}$$

where the constants $c(n,\alpha_1)$ and $c(n,\alpha_2)$ are omitted because they are not dependent on the sample.

The larger V becomes, the more probable it is that the sample X is drawn from the distribution $f(X;\alpha_2)$ and not $f(X;\alpha_1)$.

If m independent, identically distributed channels are available, the multichannel log likelihood is

$$U(X;\alpha_1,\alpha_2) = \log \prod_{j=1}^{m} \frac{\left(\frac{1}{n} \sum_{i=1}^{n} |x_{ij}|^{\alpha_2} \right)^{-n/\alpha_2}}{\left(\frac{1}{n} \sum_{i=1}^{n} |x_{ij}|^{\alpha_1} \right)^{-n/\alpha_1}} \tag{22}$$

Relation to deconvolution

For the deconvolution problem the recorded data, y, is available. What is needed is an inverse filter, f, which will reverse the undesirable effects of the forward filter, b.

The forward filter drives the distribution of the reflectivity series, X, toward Gaussian and increases entropy. Kurtosis is decreased if the reflectivity series is described by a distribution with $\alpha < 2$, for $\alpha > 2$ it is increased.

The inverse filter would have the opposite effect. It should drive the distribution away from a Gaussian toward a distribution with less entropy and more kurtosis. A method for determining the inverse filter can be derived by iteratively estimating a reflectivity series (and thus an inverse filter) such that the function $U(X;\alpha_1,\alpha_2)$ is maximized where $\alpha_1 > \alpha_2$.

In the earlier version of variable norm deconvolution (SEP-14, pp. 181-205) the function

$$V(X;\alpha) = \log \prod_{j=1}^{m} \frac{\sum_{i=1}^{n} |x_{ij}|^{\alpha}}{\left(\sum_{i=1}^{n} |x_{ij}|^{2} \right)^{\alpha/2}} \tag{23}$$

was maximized. This is equivalent to maximizing $U(X;\alpha,2)$. It is most powerful for detecting reflections in a seismogram with noise slightly more uniform than the Gaussian. The degree of uniformity is determined by the size of α.

It is believed that the distribution α_1 characterizes the undesirable averaging effects in the reflection seismogram and the distribution α_2 represents the expected distribution of the reflectivity series.

The estimate which maximizes the function $U(X;\alpha_1,\alpha_2)$ may not have the distribution α_2. It is an estimate which is "furthest" from α_1 in the "direction" α_2. This causes a problem because the estimated inverse filter sometimes collapses several real events into one in trying to drive the distribution of the reflectivity series past α_2.

Derivation of Newton-type algorithm

Several methods have been investigated for maximizing U. The method described in SEP-14 is valid only when $\alpha > 2$. Here a general Newton-type method is derived.

Newton's method (Avriel, 1976, pp. 288-299) for iteratively extremalizing the multidimensional nonlinear function $U(X;\alpha_1,\alpha_2)$ has the form

$$X^{k+1} = X^k - [\nabla^2 U(X^k;\alpha_1,\alpha_2)]^{-1} \nabla U(X^k;\alpha_1,\alpha_2) \tag{24}$$

where

X^k is the k-th estimate of the reflectivity series,

∇U is the gradient vector,

$$[\nabla U]_i = \frac{\partial U(X^k;\alpha_1,\alpha_2)}{\partial x_i} \tag{25}$$

and

$\nabla^2 U$ is the Hessian matrix

$$[\nabla^2 U]_{ij} = \frac{\partial^2 U(X^k;\alpha_1,\alpha_2)}{\partial x_i \, \partial x_j} \tag{26}$$

For now only a single channel is considered. Later the channels will be averaged over to produce a multichannel estimate of the inverse filter.

The gradient vector is derived from Equation (22) as

$$[\nabla U]_i = \left(\frac{|x_i|^{\alpha_1 - 1}}{\mu_1} - \frac{|x_i|^{\alpha_2 - 1}}{\mu_2} \right) \operatorname{sgn}(x_i) \tag{27}$$

where

$$\mu_1 = \frac{1}{n} \sum_{i=1}^{n} |x_i|^{\alpha_1} \tag{28}$$

and

$$\mu_2 = \frac{1}{n} \sum_{i=1}^{n} |x_i|^{\alpha_2} \tag{29}$$

If the current estimate of the reflectivity series has distribution α_3, μ_1 and μ_2 can be expressed in terms of the scale β as

$$\mu_1 = \frac{\Gamma\left(\frac{\alpha_1+1}{\alpha_3}\right)}{\Gamma\left(\frac{1}{\alpha_3}\right)} \beta^{\alpha_1} \tag{30}$$

$$\mu_2 = \frac{\Gamma\left(\frac{\alpha_2+1}{\alpha_3}\right)}{\Gamma\left(\frac{1}{\alpha_3}\right)} \beta^{\alpha_2} \tag{31}$$

Using these results in Equation (27) gives the gradient

$$[\nabla U]_i = \left[\frac{|x_i|^{\alpha_1-1}}{\beta^{\alpha_1} \Gamma\left(\frac{\alpha_1+1}{\alpha_3}\right)} - \frac{|x_i|^{\alpha_2-1}}{\beta^{\alpha_2} \Gamma\left(\frac{\alpha_2+1}{\alpha_3}\right)} \right] \Gamma\left(\frac{1}{\alpha_3}\right) \operatorname{sgn}(x_i) \tag{32}$$

This can be expressed in terms of the variance by substituting Equation (6) in (32), yielding

$$[\nabla U]_i = \left\{ \frac{|x_i|^{\alpha_1-1}}{\left[\frac{\sigma^2 \Gamma\left(\frac{1}{\alpha_3}\right)}{\Gamma\left(\frac{3}{\alpha_3}\right)}\right]^{\alpha_1/2}} - \frac{|x_i|^{\alpha_2-1}}{\left[\frac{\sigma^2 \Gamma\left(\frac{1}{\alpha_3}\right)}{\Gamma\left(\frac{3}{\alpha_3}\right)}\right]^{\alpha_2/2}} \right\} \operatorname{sgn}(x_i) \tag{33}$$

Two plots of this function are displayed in Figure 4 for various combinations of α_1 and α_2 with x varying. Each plot is for a different current distribution α_3.

The diagonal elements of the Hessian matrix are

$$[\nabla^2 U]_{ii} = \frac{(\alpha_1-1)|x_i|^{\alpha_1-2}}{\mu_1} - \frac{(\alpha_2-1)|x_i|^{\alpha_2-2}}{\mu_2}$$

$$- \frac{\alpha_1}{n}\left(\frac{|x_i|^{\alpha_1-1}}{\mu_1}\right)^2 + \frac{\alpha_2}{n}\left(\frac{|x_i|^{\alpha_2-1}}{\mu_2}\right)^2$$

$$+ 2\delta(x_i)\left(\frac{|x_i|^{\alpha_1-1}}{\mu_1} - \frac{|x_i|^{\alpha_2-1}}{\mu_2}\right) \qquad (34)$$

and the off-diagonal elements are

$$[\nabla^2 U]_{ij} = \left(\frac{\alpha_1}{n}\frac{|x_i x_j|^{\alpha_1-1}}{\mu_1^2} - \frac{\alpha_2}{n}\frac{|x_i x_j|^{\alpha_2-1}}{\mu_2^2}\right) \operatorname{sgn}(x_i)\operatorname{sgn}(x_j) \qquad (35)$$

Inversion of the Hessian matrix is not computationally practical because of its large size. Taking its expected value yields a constant diagonal matrix which is trivial to invert. The expectation of the Hessian is

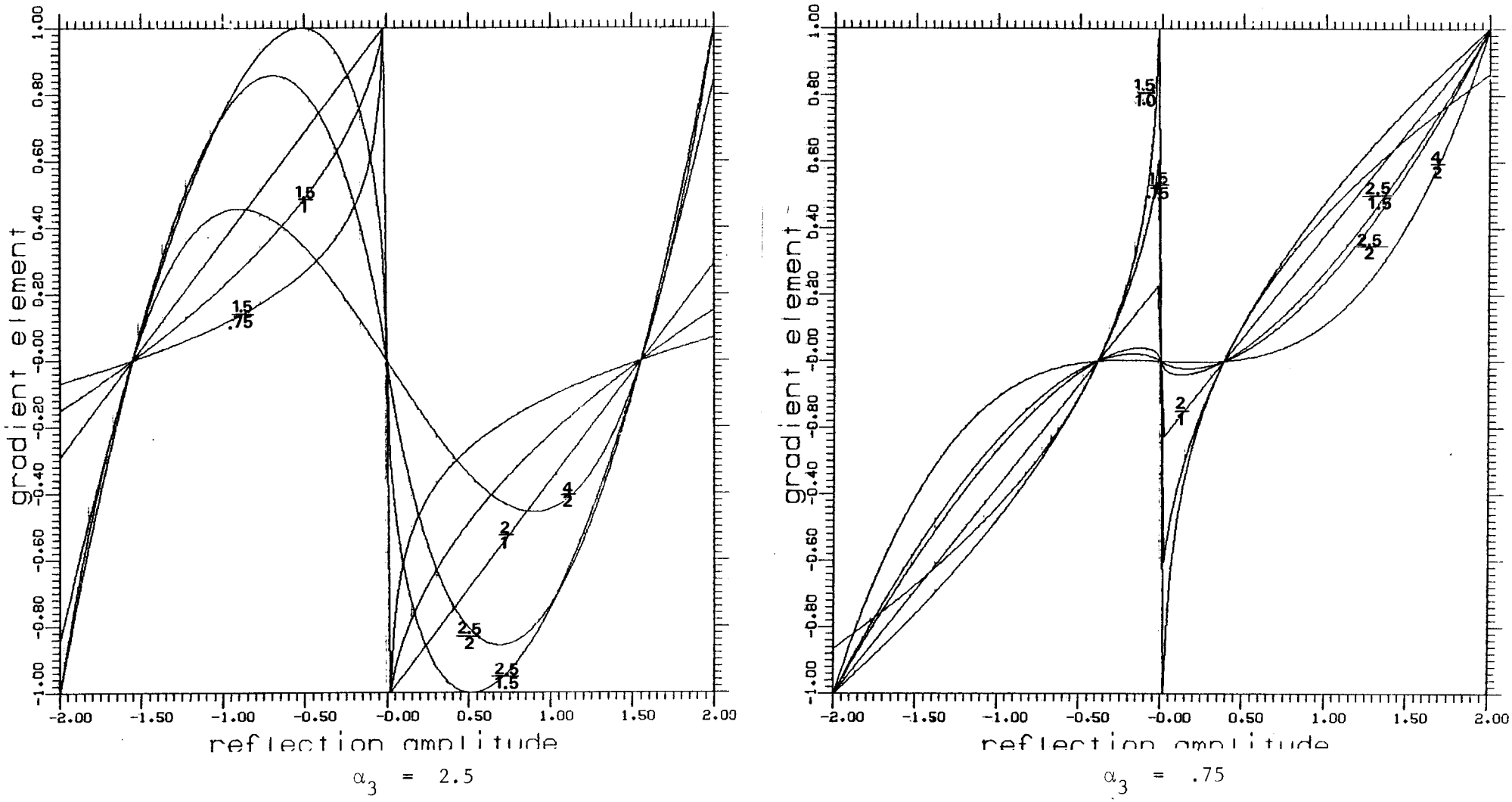
$$E([\nabla^2 U]) = cI \qquad (36)$$

where

$$c = \frac{\alpha_1-1}{\mu_1} E(|x|^{\alpha_1-2}) - \frac{\alpha_2-1}{\mu_2} E(|x|^{\alpha_2-2}) \qquad (37)$$

and I is the identity matrix.

The last term of Equation (34) has a zero expectation because the probability of a random variable being exactly zero is zero so



$\alpha_3 = 2.5$ $\alpha_3 = .75$

FIGURE 4.--Plots of gradient element size versus reflection amplitudes for different norm ratios and different current distributions. The reflection amplitude scale is in standard deviations. Each curve represents a different norm ratio and is labeled α_1/α_2. The gradient element size varied greatly and was scaled to (-1,1) for plotting.

$$E[\delta(x_i)] = 0 \tag{38}$$

Equation (35) and the third and fourth terms of Equation (34) are believed to drop out when taking their expectations because the reflection amplitudes are assumed independent and the terms are scaled by $1/n$.

The constant given by Equation (37) can be simplified using the expression for the absolute moments given by Equation (5). Assume the current estimate of the reflectivity series is described by the shape parameter α_3, then

$$c = \frac{\alpha_3}{\beta^2}\left[\frac{\Gamma\left(\frac{\alpha_1 + \alpha_3 - 1}{\alpha_3}\right)}{\Gamma\left(\frac{\alpha_1 + 1}{\alpha_3}\right)} - \frac{\Gamma\left(\frac{\alpha_2 + \alpha_3 - 1}{\alpha_3}\right)}{\Gamma\left(\frac{\alpha_2 + 1}{\alpha_3}\right)}\right] \tag{39}$$

In deriving Equation (39), the integral

$$(\alpha - 1)\int_0^\infty x^{\alpha-2} e^{-x}\, dx = (\alpha - 1)\,\Gamma(\alpha - 1) \tag{40}$$

had to be evaluated for $\alpha > 0$. The integral

$$\int_0^\infty x^{\alpha-2} e^{-x}\, dx \tag{41}$$

is defined only for $\alpha > 1$. Because

$$(\alpha - 1)\,\Gamma(\alpha - 1) = \Gamma(\alpha) \tag{42}$$

the pole at $\alpha = 1$ is simple and removable. Equation (39) is, therefore, valid for positive α_1, α_2, and α_3.

Using the maximum likelihood estimator for the scale, Equation (14) gives an estimate of the constant as

$$c = \alpha_3\left(\frac{\alpha_3 \sum_{i=1}^{n} |x_i|^{\alpha_3}}{n}\right)^{-2/\alpha_3}\left[\frac{\Gamma\left(\frac{\alpha_1 + \alpha_3 - 1}{\alpha_3}\right)}{\Gamma\left(\frac{\alpha_1 + 1}{\alpha_3}\right)} - \frac{\Gamma\left(\frac{\alpha_2 + \alpha_3 - 1}{\alpha_3}\right)}{\Gamma\left(\frac{\alpha_2 + 1}{\alpha_3}\right)}\right] \tag{43}$$

which is the estimate used in the current algorithm. A less efficient estimate for seismic data written in terms of the variance is

$$c = \frac{\alpha_3 \left(\frac{3}{\alpha_3}\right)}{\sigma^2 \Gamma\left(\frac{1}{\alpha_3}\right)} \left[\frac{\Gamma\left(\frac{\alpha_1 + \alpha_3 - 1}{\alpha_3}\right)}{\Gamma\left(\frac{\alpha_1 + 1}{\alpha_3}\right)} - \frac{\Gamma\left(\frac{\alpha_2 + \alpha_3 - 1}{\alpha_3}\right)}{\Gamma\left(\frac{\alpha_2 + 1}{\alpha_3}\right)} \right] \tag{44}$$

This estimate was derived using Equation (6) to relate the variance and scale. Figure 5 displays the constant for several combinations of α_1 and α_2 versus α_3.

When using Newton's method to maximize a nonlinear function, the Hessian matrix must be negative-definite for a maximum to be obtained. This requires the constant c to be negative. Figure 5 shows it to be negative so long as the estimated reflectivity series is described by a shape parameter less than two.

With seismic data the parameter is sometimes estimated as slightly greater than two. This usually happens when the signal to noise ratio is small. If this occurs in the multichannel case the seismogram can be left out of the estimation. In the single channel case the shape parameter must be arbitrarily set below two.

Define perturbations to the currently estimated reflectivity series as

$$dX^{k+1} = X^{k+1} - X^k \tag{45}$$

From Equation (24) the perturbations are expressible in terms of the variance as

$$dX^{k+1} = [\nabla^2 U]^{-1} \nabla U \tag{46}$$

where $[\nabla^2 U]^{-1}$ is defined by Equation (44) and ∇U by Equation (33). This function is displayed in Figure 6.

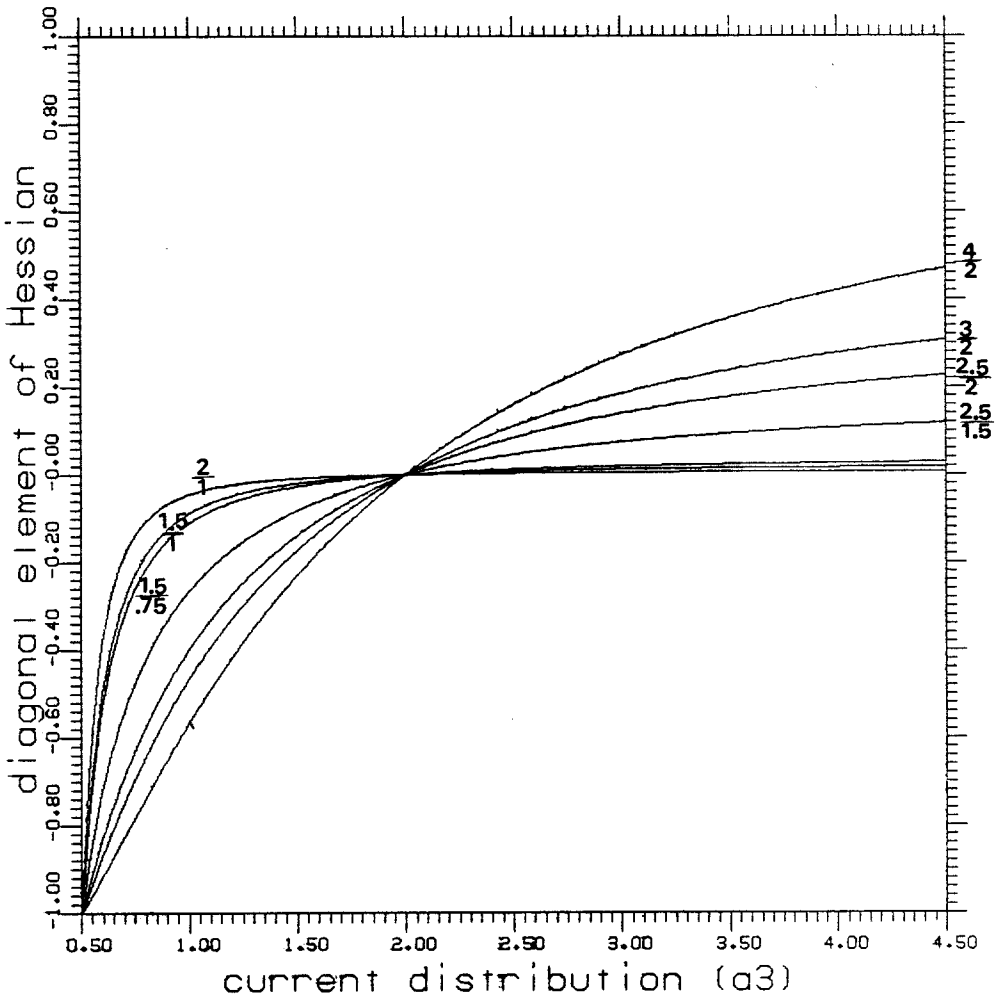


FIGURE 5.--The expected value of the diagonal element of the Hessian is plotted against different current distributions. Each curve is labeled by the norm ratio α_1/α_2. The amplitudes were scaled to the interval (-1,1). The variance of the current distribution is held constant.

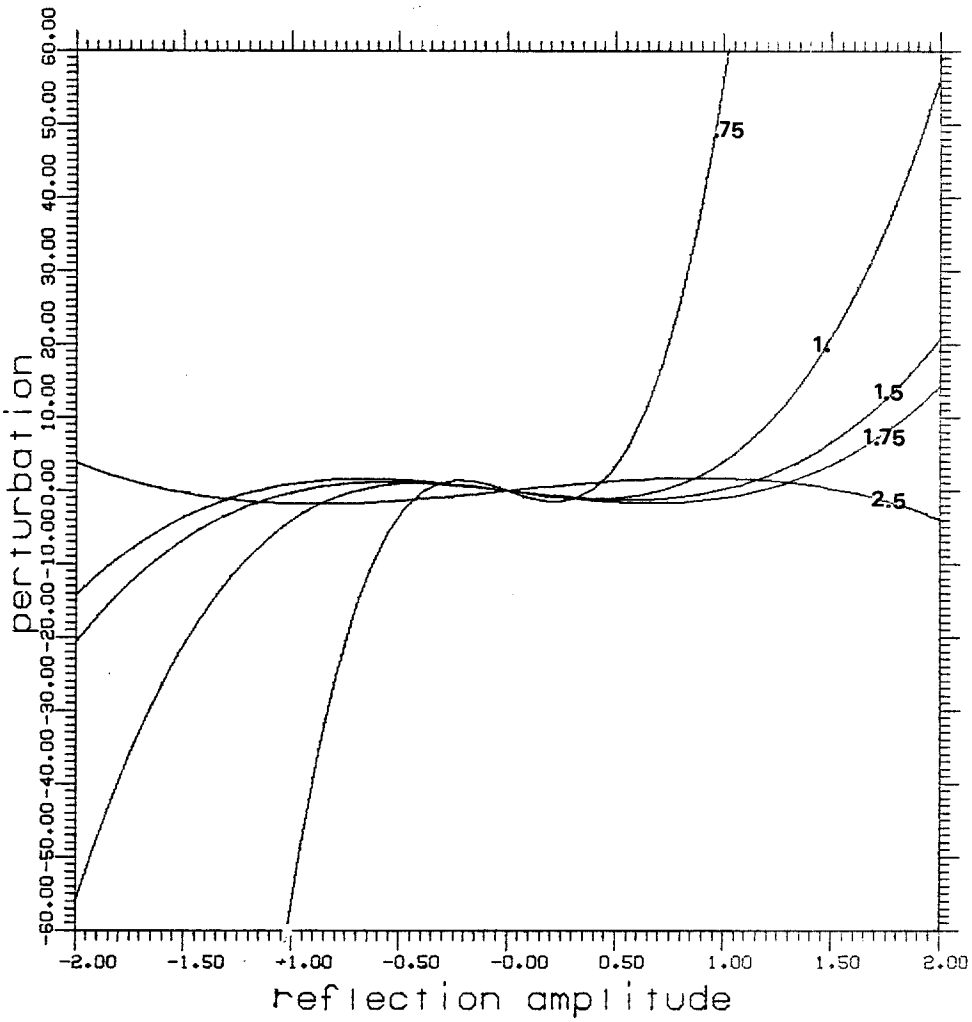
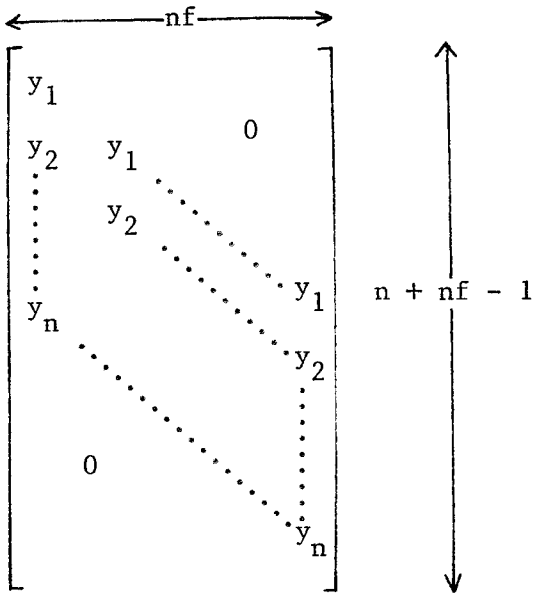


FIGURE 6.--Perturbations to the reflectivity series relative to estimated reflection amplitudes. Both axes are in units of standard deviations. Each curve represents the perturbations for different current distributions. The norm ratio was $\alpha_1 = 4$ and $\alpha_2 = 2$.

Estimating the inverse filter

So far the algorithm has been expressed in terms of the reflectivity series. Now it is related to the inverse filter. Let [Y] represent the convolution matrix



where nf is the inverse filter length.

The k-th estimate of the reflectivity series, X^k, is the convolution of the input and k-th estimate of the inverse filter written (small letters now denote vectors),

$$[Y]f^k = x^k \tag{47}$$

Substituting this into Equation (24), Newton's method, gives

$$[Y]df^{k+1} = -[\nabla^2 U(X^k;\alpha_1,\alpha_2)]^{-1} \nabla U(X^k;\alpha_1,\alpha_2) \tag{48}$$

where

$$df^{k+1} = f^{k+1} - f^k \tag{49}$$

is an unknown perturbation to the current filter estimate.

Because the length of the filter is less than that of the reflectivity series the system is overdetermined. The least squares estimate of the filter

perturbations is found by solving the system

$$[\bar{Y}\ Y]\ df^{k+1} = -[\bar{Y}]\ [\nabla^2 U(X^k;\alpha_1,\alpha_2)]^{-1}\ \nabla U(X^k;\alpha_1,\alpha_2) \tag{50}$$

where $[\bar{Y}]$ is the transpose of $[Y]$. Since the expectation of the Hessian matrix is a constant diagonal matrix the system can be rewritten

$$c[\bar{Y}\ Y]\ df^{k+1} = -[\bar{Y}]\ \nabla U(X^k;\alpha_1,\alpha_2) \tag{51}$$

where c is the constant given by Equation (43). Moving the constant to the LHS is essential to the multichannel algorithm as its inverse blows up when the current distribution is near Gaussian. This is illustrated by Figure 7.

For the multichannel case the filter is assumed common to all channels. Summing Equation (51) over channels yields the system

$$\left[\sum_{j=1}^{m} c_j\ \bar{Y}_j\ Y_j\right]\ df^{k+1} = -\sum_{j=1}^{m} [\bar{Y}_j]\ \nabla U(X_j{}^k;\alpha_1,\alpha_2) \tag{52}$$

which can be written

$$R\ df^{k+1} = -dg^{k+1} \tag{53}$$

where

R is an $nf \times nf$ Toeplitz matrix. It is a weighted sum over channels of the input autocorrelation matrices.

df^{k+1} is a vector of length nf representing unknown perturbations to the k-th filter.

and

dg^{k+1} is a vector of length nf. It is a weighted sum over channels of the crosscorrelations between the inputs and perturbations to the k-th estimated reflectivity series.

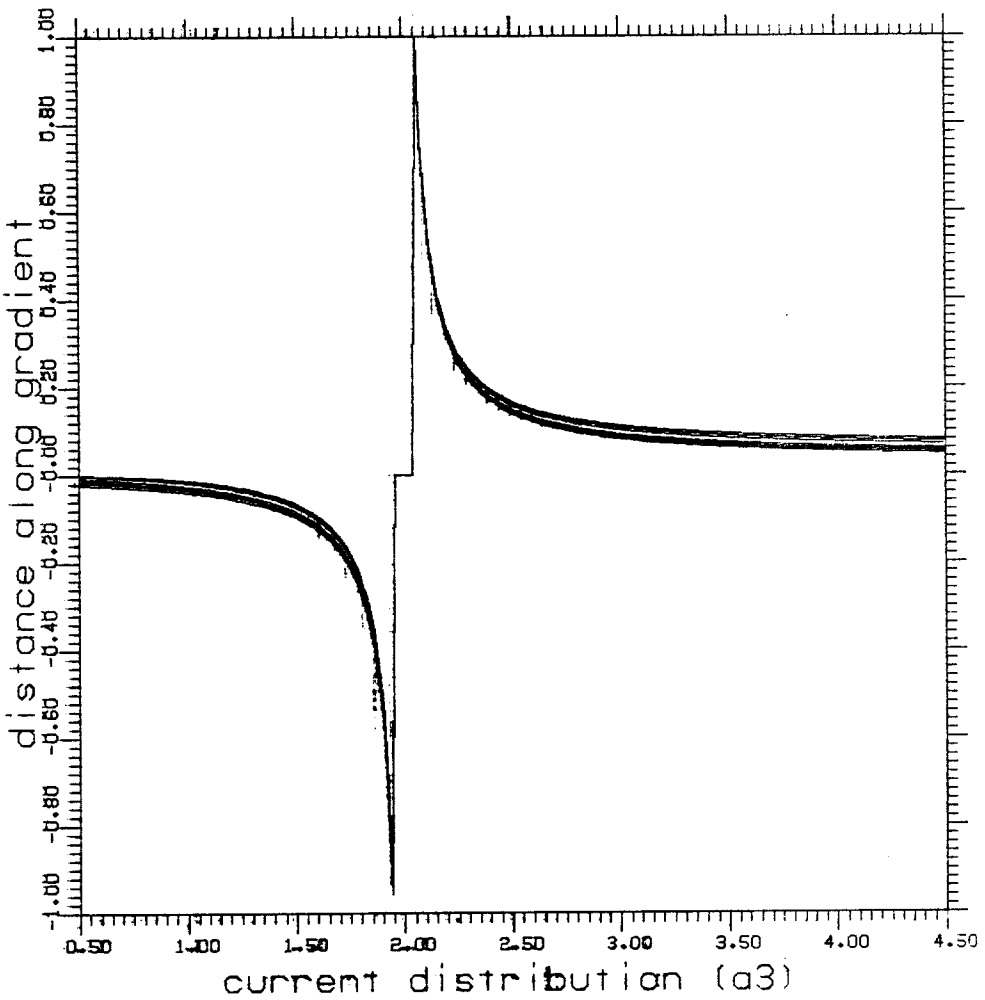


FIGURE 7.--The inverse of the Hessian constant is similar to a distance to move in the direction of the gradient. When the current distribution goes to Gaussian the distance goes to infinity. Distances for distributions near Gaussian are not plotted.

The inverse of the matrix R is efficiently computed using the *Levinson recursion* (Claerbout, 1976, pp. 141-142). To stabilize the inversion a *ridge regression technique* (Marquardt, 1970, pp. 591-612) is easily implemented. The solution is written

$$df^{k+1} = -[R + \varepsilon I]^{-1} \, dg^{k+1}$$

where ε is the constant used for ridge regression. It is dependent on the signal to noise ratio and in the current algorithm it defaults to .001. It might be useful to estimate a different constant for each channel based on the currently estimated distribution of its reflectivity series. This has not yet been tried.

Limiting the step size

After the system [Equation (53)] is solved the filter perturbations are added to the previously estimated filter. Convolving this filter with the inputs gives a new estimate of the reflectivity series. If the initial estimate of the reflectivity series is noisy or far from the "true" solution, the algorithm can become unstable.

A partial remedy is to scale the filter perturbations so the resulting estimate of the reflectivity series maximizes some easily computed objective function. Let

$$f^{k+1} = f^{k} + \lambda df^{k+1} \tag{55}$$

where λ is called the step size or scale. The k+1-th estimate of the reflectivity series is

$$\begin{aligned} x^{k+1} &= [Y]f^{k+1} \\ &= [Y]f^{k} + \lambda[Y] \, df^{k+1} \\ &= x^{k} + \lambda \, dx^{k+1} \end{aligned} \tag{56}$$

where dx^{k+1} is the perturbation to the reflectivity series.

Generalized kurtosis,

$$K(1;x^k + \lambda\, dx^{k+1}) = \frac{\frac{1}{n}\sum_{i=1}^{n} |x_i^k + \lambda\, dx_i^{k+1}|^2}{\left(\frac{1}{n}\sum_{i=1}^{n} |x_i^k + \lambda\, dx_i^{k+1}|\right)^2} \tag{57}$$

is an easily computed objective function. The optimum step size is obtained by varying λ until kurtosis is maximized.

The *Fibonacci method* (Avriel, 1976, pp. 225-240) is theoretically the best search procedure for determining the step size. It requires the objective function to be unimodal within a given search interval. Eleven function evaluations will bracket the optimum step size within one percent of the initial search interval.

Figure 7 shows the estimated distribution (inversely proportional to kurtosis) of the reflectivity series versus step size for successive iterations It is clearly unimodal for a positive step size. That the Newton-type method predicts the step size fairly accurately is evidenced by the fact that the minimum at all iterations is approximately one.

Relation to MED

A robust version of Wiggin's minimum entropy deconvolution maximizes

$$V(x) = \log \prod_{j=1}^{m} \frac{\sum_{i=1}^{n} x_{ij}^4}{\left(\sum_{i=1}^{n} x_{ij}^2\right)^2} \tag{58}$$

by iteratively solving

$$\left[\sum_{j=1}^{m} \frac{\overline{Y}_j\, Y_j}{\sum_{i=1}^{n} (x_{ij}^k)^2}\right] f^{k+1} = \left[\sum_{j=1}^{m} \frac{\overline{Y}_j (X_j^k)^3}{\sum_{i=1}^{n} (x_{ij}^k)^4}\right] \tag{59}$$

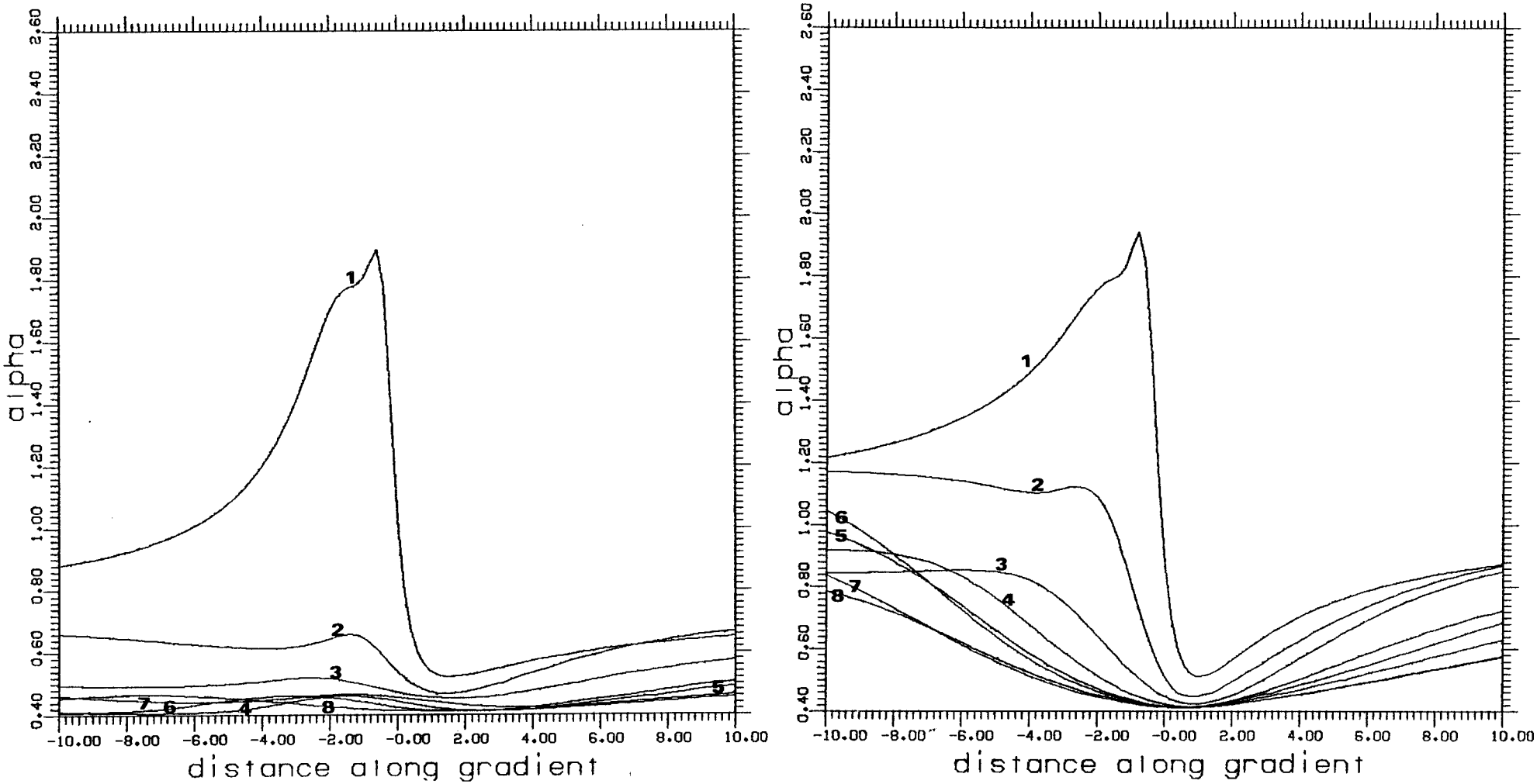


FIGURE 8.--The distribution of the estimated reflectivity series for varying step size is displayed. A step size of 1 is estimated by the Newton-type method. Each curve represents the distribution at different iterations. On the right the norm ratio was $\alpha_1 = 2.5$, $\alpha_2 = 1.5$. On the left the ratio was $\alpha_1 = 4$, $\alpha_2 = 2$. The near offset trace of Figure 9 was being deconvolved.

for an estimate of the inverse filter. $(X_j^k)^3$ represents a vector whose elements are the current reflectivity series cubed. Since

$$f^{k+1} = f^k + df^{k+1} \tag{60}$$

and

$$x_j^k = [Y_j] f_j^k \tag{61}$$

Equation (61) can be rewritten in terms of filter perturbations as

$$\left[\sum_{j=1}^{m} \frac{\bar{Y}_j Y_j}{\sum_{i=1}^{n} (x_{ij}^k)^2}\right] df^{k+1} = \sum_{j=1}^{m} \bar{Y}_j \left[\frac{(X_j^k)^3}{\sum_{i=1}^{n} (x_{ij}^k)^4} - \frac{X_j^k}{\sum_{i=1}^{n} (x_{ij}^k)^2}\right] \tag{62}$$

This is similar to the current algorithm which maximizes $U(X;4,2)$ by solving

$$\left[\sum_{j=1}^{m} c_j \bar{Y}_j Y_j\right] df^{k+1} = -\sum_{j=1}^{m} \bar{Y}_j \left[\frac{(X_j^k)^3}{\sum (x_{ij}^k)^4} - \frac{X_j^k}{\sum_{i=1}^{n} (x_{ij}^k)^2}\right] \tag{63}$$

where c_j is a constant determined by the estimated distribution of X_j^k. The two methods are identical when $\alpha_3 = 1.71$ because from Equation (44) with $\alpha_1 = 4$ and $\alpha_2 = 2$,

$$c_j = \frac{-1}{\frac{1}{n} \sum_{i=1}^{n} (x_{ij}^k)^2} \tag{64}$$

This implies that MED assumes for all iterations that the current distribution is slightly more spiky than a Gaussian.

An alternate algorithm

The Newton-type method can become unstable if the current estimated distributions vary greatly from channel to channel. In practice it is generally stable and converges quickly. The gradient vector defined by Equation (27) is orthogonal to the current estimate of the reflectivity series. Because of this the method is similar to steepest descent (Avriel, 1976, pp. 441-442) where the Hessian constant is thought of as a distance to move in the direction of the gradient determined by each channel's distribution.

An alternate method is to equally weight each channel's contribution to the sums in Equation (52). This has been accomplished by dividing the autocorrelations by their zero lag,

$$e_j = \sum_{i=1}^{n} y_{ij}^2 \tag{65}$$

and scaling the crosscorrelations by

$$d_j = \frac{1}{\sqrt{\left(\sum_{i=1}^{n} dx_{ij}^2\right)\left(\sum_{i=1}^{n} y_{ij}^2\right)}} \tag{66}$$

The system of equations represented by Equation (53) is solved and the step size then determined to maximize the kurtosis.

Application to field-recorded data

The marine seismic gather presented in SEP-14 is used here to show how the current method performs. Figure 9 displays the 48-trace gather used as input to the algorithm. The recorded gather was time-shifted to approximately flatten the reflection from the seafloor. The shifted gather was divided into eight groups of six traces. Each group had a different exponential gain applied to approximately correct for the effects of spherical divergence.

Figure 10 is the deconvolution achieved by maximizing $U(X;2,1)$ over the six near-offset traces. The estimated inverse filter was applied to the

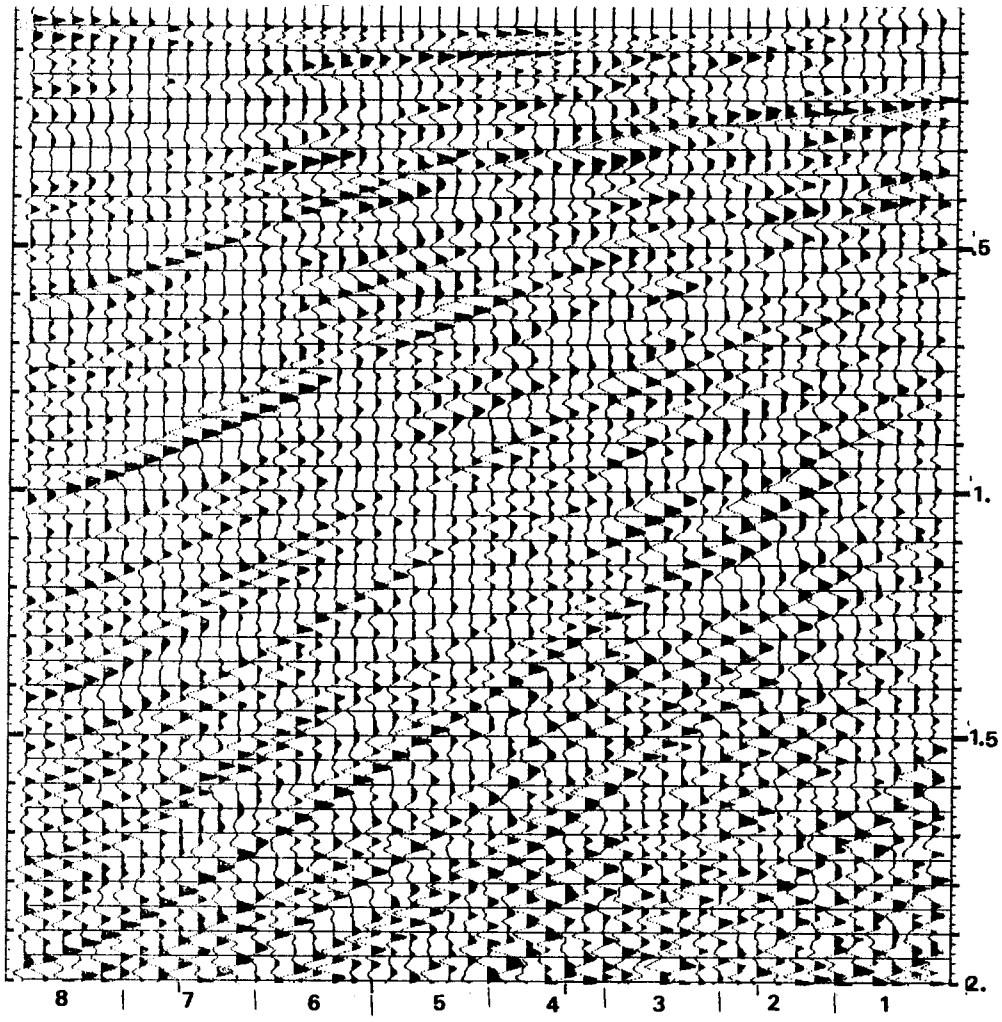


FIGURE 9.--Inputs to deconvolution algorithm. Time shifting and an exponential gain correction were applied.

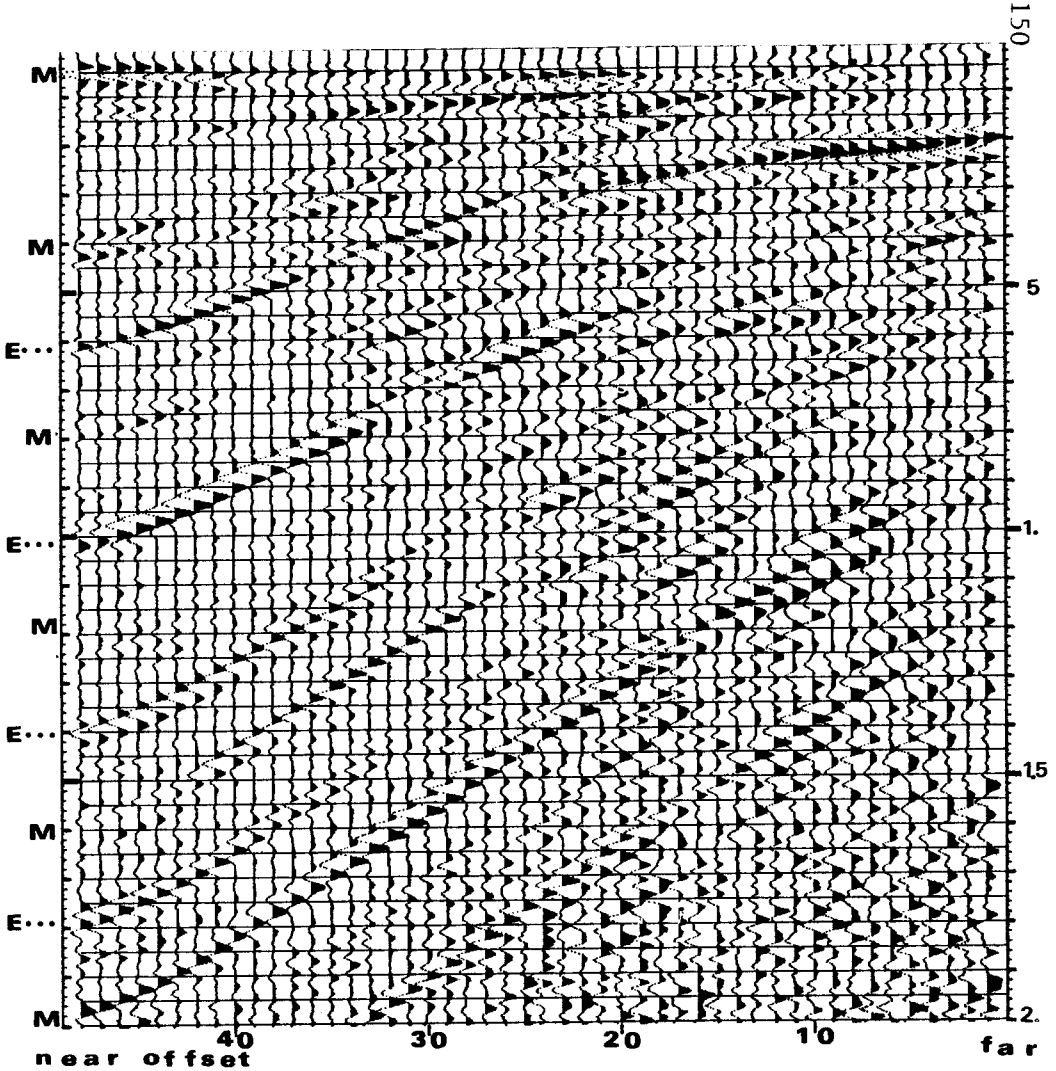


FIGURE 10.--Results obtained using U(X;2,1) to estimate a filter from traces 42-48. The filter was applied to the entire gather. The seafloor and its multiples are marked on the left by M. A strong primary and its peglegs are marked by E.

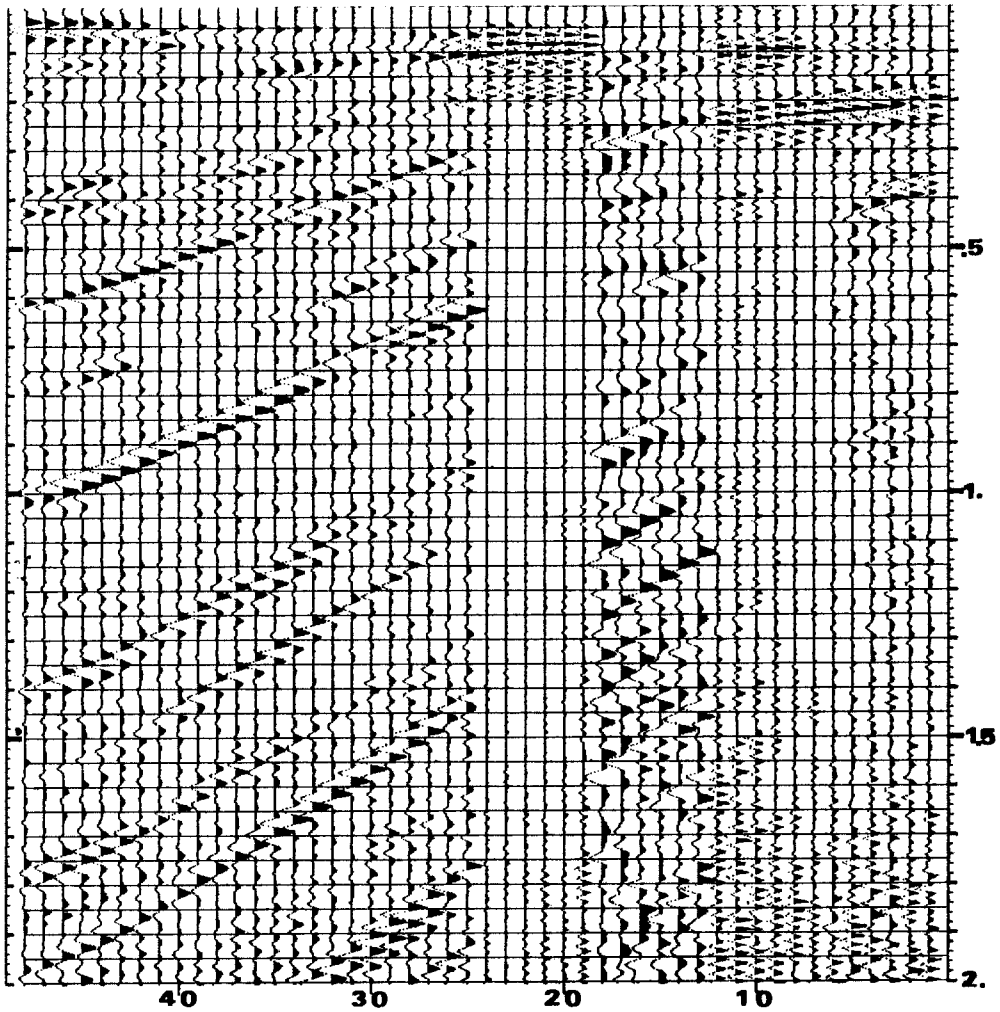


FIGURE 11.--Results obtained using U(X;2,1) to estimate an inverse filter for each six-trace subgroup of the gather.

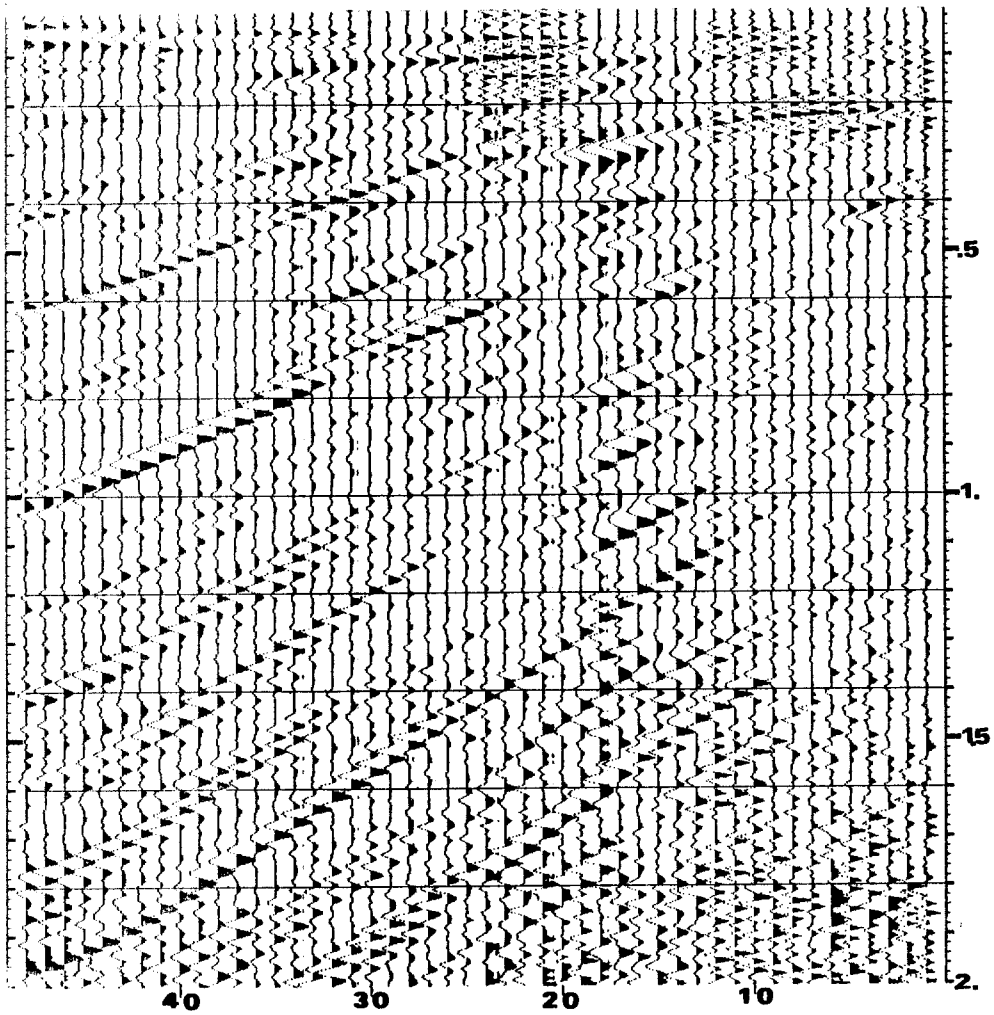


FIGURE 12.--Results obtained using $U\,[X;2\hat{\alpha},\max(\hat{\alpha},1)]$ to estimate a filter for each six-trace group.

entire gather. The deconvolution is noticeably simpler in appearance, especially the near-offset traces. The Newton-type method with no limiting of step size was used. It converged in five iterations.

Figure 11 is the deconvolution obtained by maximizing $U(X;2,1)$ where a different inverse filter was estimated for each group. The Newton method with no limit on the step size became unstable for several of the groups. This is believed to be due to high noise levels, end effects, and the channels not being independent.

Many methods were tried to improve the robustness of the algorithm. Figure 12 is a result obtained after eleven iterations where

$$U\,[X;2\hat{\alpha},\max(\hat{\alpha},1)\,]$$

was maximized using a limited step. In this method the distribution of each channel's estimated reflectivity series, $\hat{\alpha}$, was used to determine the norm ratio. The limiting of step size was determined from only one trace of the gather because of core limitations. The results are considerably better than those of Figure 11. The noise in the lower right-hand corner is believed due to end effects.

Figure 13 is a near-offset section extracted from the same line as the shot gather displayed in Figure 9. Prominent bubble oscillations mask primary events in several areas. The ungained section was deconvolved using the Newton-type method to maximize $U(X;2,1)$. The initial filter was 138 points long. A spike was placed at the 25th point. Figure 14 is the result obtained after three iterations. The bubble oscillations were effectively suppressed. Precursors before the seafloor reflection, a problem caused by earlier algorithms, are absent. This is attributed to the sensitivity of $U(X;2,1)$ to very small reflection amplitudes. $U(X;4,2)$, being sensitive only to the larger amplitudes, yielded noticeable precursors when applied to this dataset.

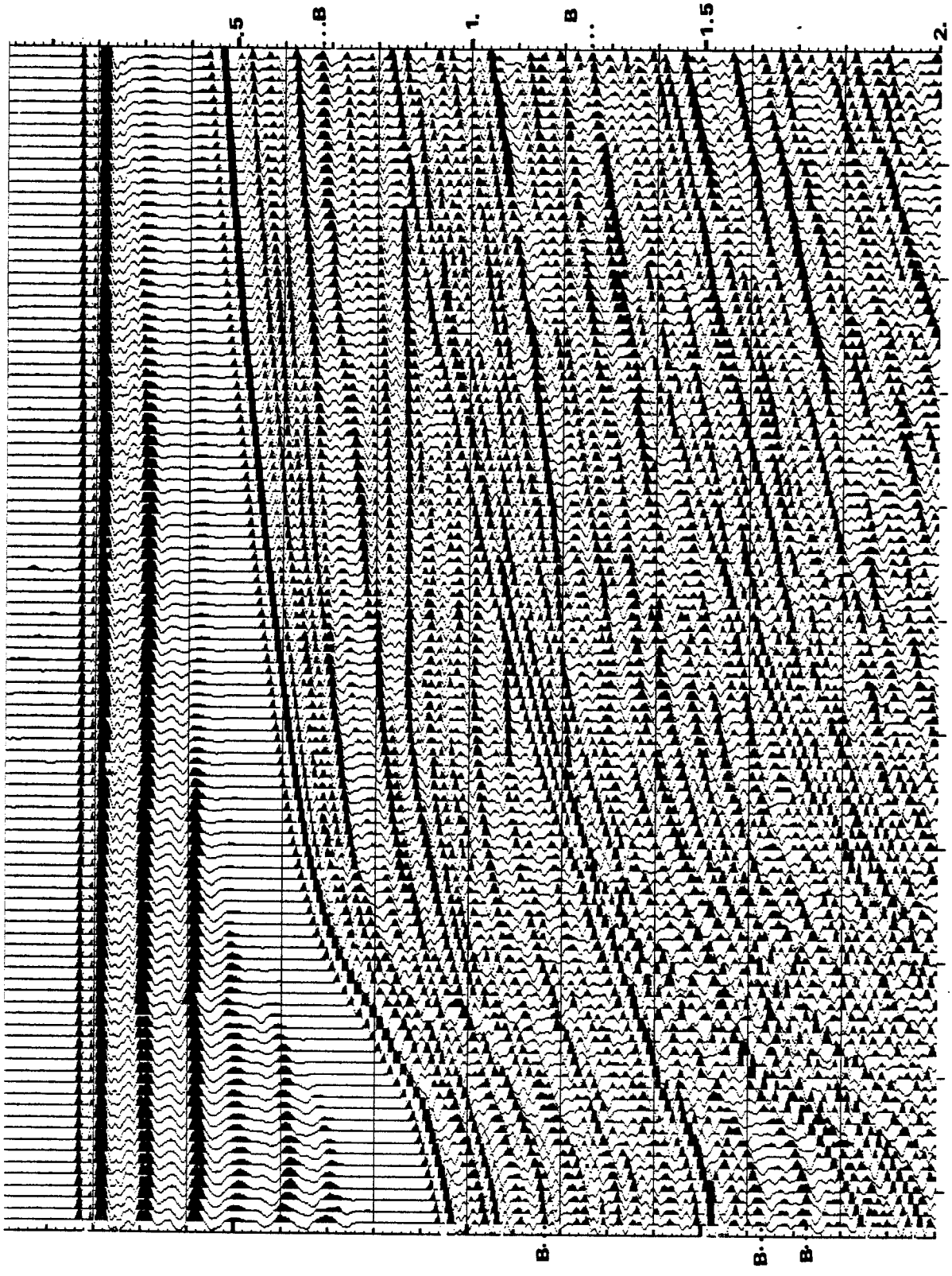


FIGURE 13.--Several areas of this near-offset section display the air gun bubble masking primary reflections. These areas are indicated by B's. The traces were AGC'ed before plotting.

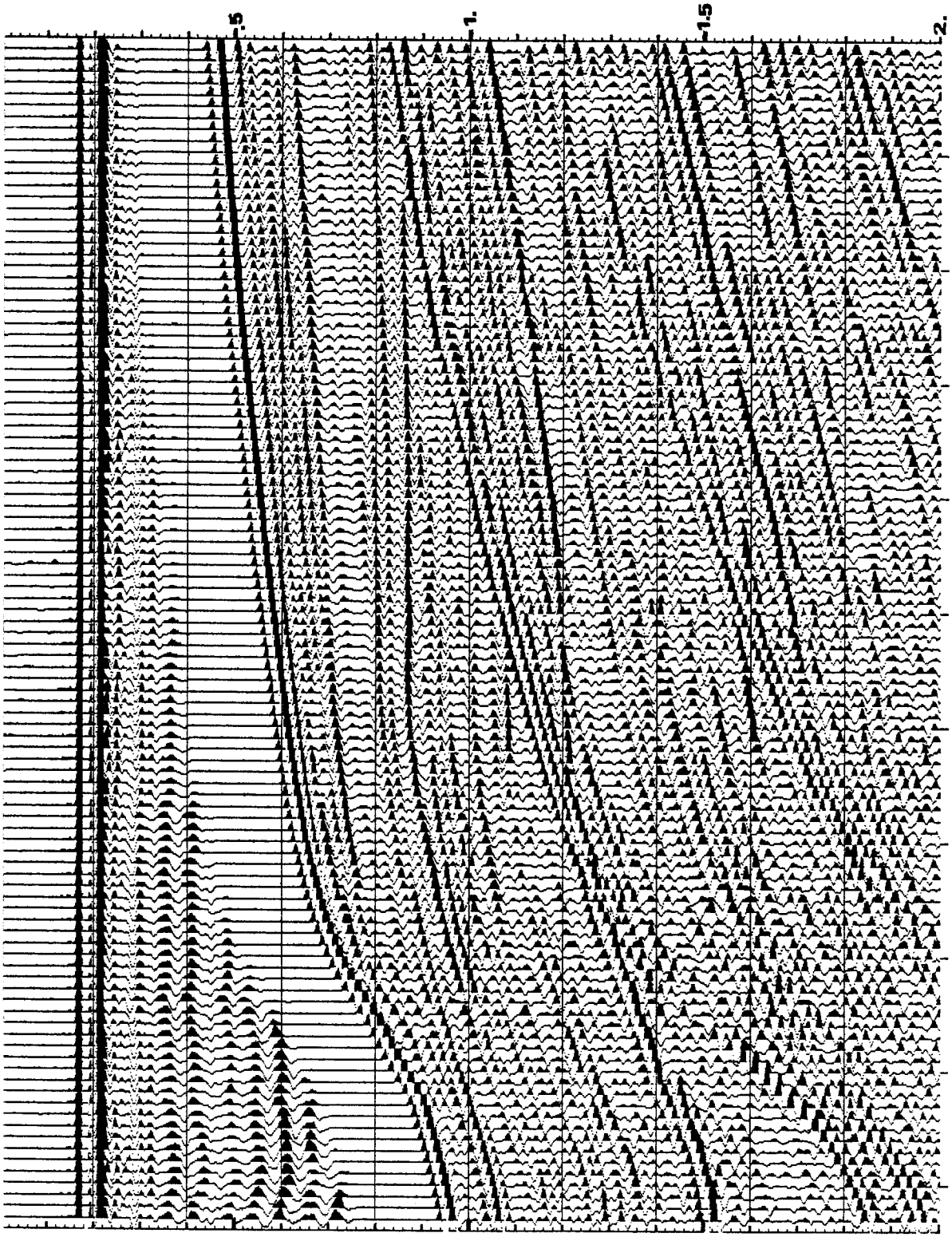


FIGURE 14.--Deconvolved near-offset section obtained by maximizing U(X;2,1). The bubbles oscillating labeled on Figure 13 are suppressed. What is left of the direct arrival indicates the air gun signature varies slightly from shot to shot.

Summary

A deconvolution method has been presented which iteratively determines an inverse filter by maximizing a nonlinear function of the resulting reflectivity series. The function is a measure of the probability that the reflectivity series is one member of a family of distributions and not another. The technique was applied to the generalized Gaussian family. It could be used to derive algorithms for other families, the *generalized Beta* (Miller, 1972, pp. 241-250) being a likely candidate.

The ideas presented here have application to problems other than deconvolution. The paper on exponential gain estimation (this report, p. 183) is a simple example. Mehta (SEP-11, pp. 61-75) tried estimating migration velocity as a function of the non-Gaussianness of the migrated wavefield. The function U(X;2,1) could be used in the same application with tremendous savings in computation.

The results presented show the method is not very robust and can produce deconvolutions which are worse than the inputs. Currently we are working on this problem by trying to reformulate the method in a detection theory framework for greater understanding.

REFERENCES

ALGAZI, V. R. and R. M. LERNER (1969), "Binary detection in white non-Gaussian noise," *M.I.T. Lincoln Lab*, Lexington, Mass., *Tech. Rep. DS-2138*, 1969.

AVRIEL, M. (1976), *Nonlinear Programming: Analysis and Methods* (Englewood Cliffs, New Jersey: Prentice-Hall, Inc.).

BOX, G. E. P. (1953), "A note on regions for tests of kurtosis," *Biometrika*, 40, pp. 465-468.

BOX, G. E. P. and G. C. TIAO (1973), *Bayesian Inference in Statistical Analysis* (Reading, Massachusetts: Addison-Wesley Publishing).

CLAERBOUT, J. F. (1976), *Fundamentals of Geophysical Data Processing* (New York: McGraw-Hill).

GEARY, R. C. (1947), "Testing for normality," *Biometrika*, 34, pp. 209-242.

GOODMAN, I. R. and S. KOTZ (1971), "On structural properties of a multivariate θ-generalized distribution," Eastern Regional Meeting of American Statistical Association, April 21-23, 1971, University Park, Pennsylvania.

HAJEK and SIDAK (1967), *Theory of Rank Tests* (New York: Academic Press), pp. 107-108.

HOGG, R. V. (1972), "More light on the kurtosis and related statistics," *Journal of the American Statistical Association*, 67:33, June 1972, pp. 422-424.

MARQUARDT, D. W., "Generalized inverses, ridge regression, biased linear estimation and nonlinear estimation," *Technometrics*, 12, pp. 591-612.

MILLER, J. H. and J. B. THOMAS (1972), "Detector for discrete-time signals in non-Gaussian noise," *IEEE Transactions on Information Theory*, IT-18:2, March 1972, pp. 241-250.

SUBBOTIN, M. T. (1923), "On the law of frequency of errors," *Matematicheskii Sbornik*, 31:1, January 1923, pp. 296-301.

WIGGINS, R. A. (1977), "Minimum entropy deconvolution," Proceedings of the International Symposium on Computer Aided Seismic Analysis and Discrimination, June 9-10, 1977, Falmouth, Mass., IEEE Computer Society.