VI. PROPERTIES OF THE VARIABLE NORM RATIO

Seismograms recorded on many channels are available for deconvolution. What is needed is an inverse filter which will reverse the undesirable averaging effects of the forward filter. The forward filter drives the distribution of each channel's reflectivity series towards Gaussian, increases entropy, and decreases kurtosis. If we assume the reflectivity series are spiky members of the generalized Gaussian family, the forward filter can be characterized as increasing their shape parameters towards two.

Design of an inverse filter will be based on reversing these effects. The inverse filter will be estimated such that its outputs are characterized by the smallest possible shape parameter. These outputs are thus less Gaussian, have higher kurtosis and less entropy than the inputs.

In this chapter we assume the recorded seismogram is characterized by a shape parameter α_1 and the underlying reflectivity series by shape parameter α_2. The test statistic developed at the end of the preceding chapter is a measure of the likelihood that a random sample is characterized by shape parameter α_2 and not α_1. Here it is extended to a multichannel measure which is the objective function for variable norm deconvolution. Methods for determining an inverse filter which maximizes this objective function are the subject of the following chapters.

A. Previous Norm Ratio Methods

In 1977 Wiggins (26) presented an iterative, multichannel method called Minimum Entropy Deconvolution (MED). MED attempts to account for the observed seismograms in terms of a small number of large events and with the smallest residual error. An inverse filter is estimated which maximizes a measure of the simplicity of the reflectivity series. The varimax norm ratio,

$$V(X) = \sum_{j=1}^{m} \frac{\sum_{i=1}^{n} x_{ij}^4}{\left(\sum_{i=1}^{n} x_{ij}^2\right)^2} \tag{6-1}$$

where m is the number of channels and n is the number of samples, is used as a measure of simple structure. The varimax norm ratio is analogous to an arithmetic average over channels of the test statistic for determining if a random sample is from a distribution $\alpha_2 = 2$ and not $\alpha_1 = 4$. This norm ratio is believed best for detecting a spiky reflectivity series in convolutional noise which is nearly uniform ($\alpha_1 = 4$). It is not ideally suited to the distributions observed from actual reflectivity series and noise studies.

Claerbout realized this and consequentially developed Parsimonious Deconvolution. Intuitively he felt that MED is excessively biased toward the larger events and a method which "sees" more of the data would result in a better deconvolution. Claerbout proposed minimizing the norm ratio

$$P(X,\alpha) = \lim_{\varepsilon \to 0} \frac{\left[\sum_{j=1}^{m} \sum_{i=1}^{n} |x_{ij}|^{\alpha-\varepsilon}\right]^{1/\alpha-\varepsilon}}{\left[\sum_{j=1}^{m} \sum_{i=1}^{n} |x_{ij}|^{\alpha}\right]^{1/\alpha}} \tag{6-2}$$

by variation of the estimated source wavelet.

Following Claerbout's reasoning, a prototype for Variable Norm Deconvolution was introduced. It was essentially a generalization of Wiggins' work. The proposed ratio was

$$Q(X,\alpha) = \sum_{j=1}^{m} \frac{\sum_{i=1}^{n} |x_{ij}|^{\alpha}}{\left(\sum_{i=1}^{n} |x_{ij}|^{2}\right)^{\alpha/2}} \tag{6-3}$$

where the parameter α governs the relative weightings of the reflectivities. This ratio is believed to be optimum for detecting spiky signals in noise which is more uniform than Gaussian, the degree of uniformity being controlled by the size of α. Deeming (27), about the same time, independently proposed the variable norm and a single algorithm incorporating the above methods.

Ooe and Ulrych (28) proposed maximizing a modified ratio,

$$V'(X) = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij}^{4}}{\left(\sum_{j=1}^{m} \sum_{i=1}^{n} x_{ij}^{2}\right)^{2}} \quad . \tag{6-4}$$

They also proposed applying the transformation

$$z_{ij} = 1 - \exp\left(\frac{-x_{ij}^{2}}{2s^{2}}\right) \tag{6-5}$$

to the estimated reflectivity series where s is a constant in the range

$$\max_{i,j} \frac{|x_{ij}|}{3} \leq s \leq \max_{i,j} \frac{|x_{ij}|}{\sqrt{2}} \quad . \tag{6-6}$$

They reported improved results and faster convergence relative to MED. This is possibly because the transformation changes the distribution of the reflectivity series and noise towards uniformity, the region in which V' is optimal.

B. Multichannel Variable Norm Ratio

Variable norm deconvolution is based on maximizing the multichannel function,

$$U(X,\alpha_1,\alpha_2) = \log \prod_{j=1}^{m} \frac{\left(\frac{1}{n} \sum_{i=1}^{n} |x_{ij}|^{\alpha_1}\right)^{n/\alpha_1}}{\left(\frac{1}{n} \sum_{i=1}^{n} |x_{ij}|^{\alpha_2}\right)^{n/\alpha_2}} \quad . \tag{6-7}$$

This function has several important advantages over the previously proposed measures:

1. It is the geometric mean of the statistic across channels. In testing those ratios which use the arithmetic average it was found that a single spiky channel could dominate the solution as it is given the greatest weight. Often a single channel would be spiked by the inverse filter and the other channels trashed. Using the geometric mean is later shown to weight channels inversely to their spikiness. This does not allow a spiky channel to dominate the solution but may allow noisy channels to slow convergence. Essentially using the geometric mean improves robustness.

A common situation in seismic processing is a data set having a single channel with a huge spike caused by incorrectly reading from magnetic tape. The geometric mean of the statistic over channels would automatically assign the channel zero weight in the estimation of the inverse filter.

2. It is invariant to varying gain levels on different channels. Those proposed by Claerbout and Ooe are not scale invariant. They are essentially single channel measure whereas the varimax norm and $U(X,\alpha_1,\alpha_2)$ are multichannel measures. This property also improves robustness because the scale of typical seismic data can vary greatly across channels and is difficult to remove.

3. It allows the number of samples per channel to vary. In the above expression the number of samples, n, could be subscripted by the channel index, j. This is useful because a spatially varying window may be used to edit out noisy portions of the section being deconvolved. The other norms are not given in terms of the length of the channels. They must assume all channels are of equal length.

4. It allows the norm ratio used on each channel to be determined by that channel's statistics. The shape parameter for each channel can be estimated by generalized kurtosis. This parameter can then be used to determine the corresponding statistic to be maximized. Although the proposed objective function is written in terms of fixed α_1 and α_2, they could be subscripted by the channel index j.

5. It allows the norm ratio to change as the iteration proceeds on each channel. The objective function could be written with α_1 and α_2 subscripted by the channel index and superscripted by the iteration number. This greatly improves robustness as the distribution of the estimated reflectivity series changes both across channels and as the iteration proceeds.

6. The proposed norm ratio is derived from a statistical basis. It is the geometric mean over channels of a scale invariant test statistic for

the hypothesis that a random sample is described by the shape parameter α_2 and not α_1. This enables a greater understanding of what the derived deconvolution method is doing to the data. It also allows one to visualize the inverse filter as one which changes the distribution of the inputs in a known way.

Six advantages of the proposed objective function have been noted above. Now some possible disadvantages are noted.

1. In the next two chapters, methods for determining an inverse filter are derived. They iteratively estimate a reflectivity series (or inverse filter) such that the function $U(X,\alpha_1,\alpha_2)$ is maximized where $\alpha_1 > \alpha_2$. The estimate which maximizes $U(X,\alpha_1,\alpha_2)$ may not have the distribution α_2. It is an estimate which is "furthest" from α_1 in the "direction" α_2. This is a mixed blessing.

It causes problems because the estimated inverse filter sometimes collapses several real events into one by trying to drive the distribution of the reflectivity series past α_2. This problem is partially remedied by requiring the reflectivity series on different channels to be independent. For deconvolving common shot gathers this independence is satisfied by differential moveout. In structurally complex areas the seismograms also have the independence property. For a parallel layered earth, the seismograms are not independent across channels and the algorithms will try to combine all events within a time interval equal to the filter length.

Trying to drive the shape parameter as small as possible is a good property because the shape parameter describing the outputs is rarely known. About all we can say for sure is that it is smaller than that estimated for the inputs to the process.

2. The proposed ratio has a lot of parameters relative to the others mentioned above. This abundance allows flexibility and is believed to improve robustness. On the negative side, it complicates understanding and may prove difficult for others to use. The more parameters a process has, the greater the possibility for incorrectly choosing one of the parameters.

3. A problem, which will become evident when the gradient for the Newton-type algorithm is derived, is the requirement of a threshold parameter to prevent division by zero. This occurs when either α_1 or α_2 is less than 1. We have studied the choice of this parameter and have not yet found a robust criterion for estimating it. This subject is covered again in the chapter on implementation details.

C. Choice of Objective Function

Here several possibilities for choice of the shape parameters are presented. They are broken down into two classifications with three categories each.

1. Variable methods

The coefficients α_1 and α_2 are adjustable to characterize the data being deconvolved. Ideally the setting of these parameters can be made at each iteration by the program, after analyzing the distribution of the currently estimated reflectivity series on each channel. Three philosophies govern this approach. If $\hat{\alpha}$ is the estimated distribution of the current reflectivity series, the first tries to push $\hat{\alpha}$ away from the Gaussian by maximizing $U(X,2,\hat{\alpha})$. The second tries to push $\hat{\alpha}$ towards the certain event by maximizing $\lim_{\varepsilon \to 0} U(X,\hat{\alpha},\varepsilon)$.

The third is similar to Claerbout's Parsimonious Deconvolution. It is most unbiased because it increases U by maximizing $U(X,\hat{\alpha},\hat{\alpha}-\varepsilon)$ where ε is a small constant.

2. Fixed methods

There are three categories of fixed norm methods. The first maximizes $U(X,\alpha_1,2)$ where $\alpha_1 > 2$. This method estimates the reflectivity series by pushing up the larger events. It is believed optimum for a seismogram with convolutional noise described by α_1 and is attractive because the Wiggins-type algorithm can be used. It generally converges in fewer iterations than any of the other methods but has a tendency to combine large events.

The second category maximizes $U(X,2,1)$. This is believed to be closer to the optimum for seismic data because it tries to drive the seismogram's distribution away from the Gaussian toward the double exponential. Generally, it is more robust than the first.

A third category tries to drive the distribution towards the certain event from the Gaussian by maximizing $\lim_{\varepsilon \to 0} U(X,2,\varepsilon)$. Claerbout derived an equivalent function from the geometric inequality which he called Bit Count Deconvolution. The function is a measure of the dissimilarity of the x_is. It is appealing because the convolutional noise is represented by the Gaussian and the spikiest possible deconvolution by ε.

D. Asymptotic Properties

We define a *Bussgang process* as any zero mean random process which has the following property,

$$\frac{E[z(x_t)x_{t+\tau}]}{E[z(x_t)x_t]} = \frac{E[x_t\ x_{t+\tau}]}{E[x_t\ x_t]} \tag{6-8}$$

where z(•) represents the output of an arbitrary memoryless non-linearity such as a cuber.

The property is essentially that the crosscorrelation between the inputs to a non-linearity and its outputs have the same shape as the autocorrelation of the inputs.

A number of random processes belong in this class. Bussgang (29) observed that any correlated Gaussian process has this property. Barrett and Lampard (30) extended Bussgang's result to all processes with exponentially decaying autocorrelation functions. This includes any independent process as its autocorrelation function is an infinitely fast decaying exponential -- a delta function. Other members of this class are Markov chains with arbitrary probability distributions. This type of process is commonly used to model stratigraphic sequences (10) and impedance logs. It is simple to extend the class to include those obtained by differentiating these processes. This is a very significant result because the reflectivity series can be modeled as the differential of an impedance log.

Here we show asymptotically that inputs described by a Bussgang process are not modified by the algorithm. Using expectations the norm ratio for an infinitely long channel is

$$O(X,\alpha_1,\alpha_2) = \ln \frac{\{E[|x|^{\alpha_1}]\}^{1/\alpha_1}}{\{E[|x|^{\alpha_2}]\}^{1/\alpha_2}} \quad . \tag{6-9}$$

Optimization methods presented in the following chapters differentiate the norm ratio with respect to each coefficient of the inverse filter; the resulting gradient is written

$$\frac{\partial O}{\partial f_\tau} = \frac{E[z_1(x_t)x_{t+\tau}]}{E[z_1(x_t)x_t]} - \frac{E[z_2(x_t)x_{t+\tau}]}{E[z_2(x_t)x_t]} \tag{6-10}$$

where $z_\ell(x_t)$ represents the antisymmetric memoryless nonlinearity

$$|x_t|^{\alpha_\ell - 1}\, \mathrm{sgn}(x_t) \ . \tag{6-11}$$

Using the property of the Bussgang process the gradient is identically zero. This proves asymptotically that the algorithms converge when the estimated reflectivity series has this property. It also indicates the parameters α_1 and α_2 are arbitrary positive constants (for infinitely long inputs). This corollary is supported by empirical results which indicate for reasonably varying α_1 and α_2, that the resulting deconvolutions are practically the same when the sample size is large.