C.1

DEVELOPMENTS IN MINIMUM ENTROPY DECONVOLUTION

by

WILLIAM ALEXANDER NICKERSON

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES
DEPARTMENT OF GEOPHYSICS AND ASTRONOMY

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
JANUARY, 1986

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Geophysics & Astronomy

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date February 12, 1986

DE-6(3/81)

Abstract

Minimum entropy deconvolution (MED) is investigated in light of recent work, and reports of poor performance in the deconvolution of real reflection seismograms. The problems with MED fall into two catagories. The first contains problems related to operator design such as avoiding local extrema in the simplicity criterion, problems with bandwidth, and the selection of filter length and preprocessing parameters. The second is the determination of model validity for the data at hand.

MED is based on the iterative maximization of a data simplicity criterion which exhibits multiple extrema. A geometrical interpretation of this algorithm leads to an intuitive alternative algorithm, the projection method, and to the conclusion that iteration starting filter governs which extremum is reached and hence output wavelet delay. Choice of starting filter is therefore tantamount to a wavelet phase assumption and there is no basis for choosing the standard, centred impulse, starting filter if the global extremum is desired. An optimum lag MED algorithm is proposed to achieve the global maximum. Bandlimiting, data tapering and other aspects of MED operator design are discussed.

No criterion exists for assessing MED model validity. Studies with various synthetic reflectivity statistics reveal that "sparseness" and "simplicity" are poor descriptors of the minimum entropy assumption. Preliminary

attempts to quantify MED applicability are made but fail and indicate that any successful measure must include both underlying reflectivity and wavelet characteristics.

Examples using synthetic data show the advantages of the bandlimited optimum lag MED algorithm over a conventional starting filter method.

Table of Contents

G

Table

List of Figures

Acknowledgements

I am grateful to professor Tad Ulrych, my supervisor, for his support and guidance. Special thanks to Toshi Matsuoka, my surrogate advisor in Tad's absence, for his ideas, encouragement and Japanese-style tutelage. Colin Walker has provided many of the computer programs used in this work and Ken Whittall has been a great help in computer matters. Terry Deeming of Digicon has shown me an industry viewpoint and conversations with him have helped focus this work.

This reseach has been supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

1. INTRODUCTION

1.1 THE DECONVOLUTION PROBLEM

Deconvolution is a technique used in the processing of reflection seismic data aimed at sharpening the images of deep geologic boundaries. Since the mid-1960's, deconvolution in some form has been applied to nearly all digitally recorded data in the presence of a continuing debate over how best to formulate and solve the problem of removing the seismic wavelet. In this section I shall discuss the mathematical formulation of the deconvolution problem, the two general approaches to its solution and the convolution model of the seismogram on which both methods are based.

For the deconvolution approaches considered in this section a discretely sampled seismic trace **x** is represented as a time invariant, one dimensional linear system.

$$\mathbf{x} = \mathbf{w} * \mathbf{r}. \tag{1.1}$$

$$x_i = \sum_{j=1}^{nw} w_j r_{j-i} \qquad i=1,\ldots nx$$

where, using a vector notation for the sampled trace

$$\mathbf{x} = (x_1, x_2, \ldots x_{nx})^{\mathrm{T}}$$

$$\mathbf{w} = (w_1, w_2, \ldots w_{nw})^{\mathrm{T}}$$

and $\mathbf{r} = (r_1, r_2, \ldots r_{nr})^{\mathrm{T}}$

$nx = nw + nr - 1.$

Simply stated, the seismic time sequence is modelled as the convolution of two elements. The first is the wavelet **w**, a transient pulse composed of the seismic source signature, instrument response and some earth filtering effects. The second is an earth impulse response series **r**, the desired signal which has been smoothed by the seismic wavelet. Deconvolution, which means inverse filtering, attempts to solve an equation of the form (1.1) for **r** by designing a linear filter **f** which when convolved with the data gives an estimate of **r**, **y** such that

$$\mathbf{y}=\mathbf{f}*\mathbf{x}. \tag{1.2}$$

It is interesting to note that the convolutional model of the seismogram is just a successful mathematical construction. Based loosely on the wavelet theory of the seismogram developed in a series of papers between 1940 and 1953 by Ricker, the convolution model was adopted by Robinson(1954) to allow the application of predictive deconvolution to seismograms. Physical reasoning based on a flat Earth, plane wave, normal incidence model can be used to justify the convolutional model. However, Zwiolkowski (1984) points out inconsistencies in this and other physical arguments to emphasize that model validity is still open for debate and his refreshing book is recommended to any reader interested in deconvolution.

There are two basic approaches to finding **f** depending on whether or not an estimate of **w** is available. If the source wavelet is indeed known we may take a deterministic approach to solving equation (1.1). This thesis however concerns the problem of deconvolution when the wavelet is unknown. In this case if we allow for an additive noise term, **n**, our seismogram becomes

$$\mathbf{x}=\mathbf{w}*\mathbf{r}+\mathbf{n}. \tag{1.3}$$

Clearly we are seeking to solve one equation in three unknowns. This is done by invoking restrictive statistical assumptions on the nature of the unknown sequences. One particularly successful set of assumptions is that

1. **n** is a gaussian, white stationary process
2. **r** is a gaussian, white stationary process, and
3. **w** is causal, minimum phase and finite in energy.

Under these assumptions the Wold decomposition for stationary time series guarantees a solution, regardless of whether the assumptions are physically justified, and a solution is obtained using the celebrated spiking, or unit gap predictive deconvolution of Robinson (1954).

In recent years a flood of techniques for statistical deconvolution have appeared in the literature, each of which makes its own assumptions in solving the underdetermined problem of equation (1.3) and each raising its own objections. It is evident that no one statistical method is suited to coping with the variety of problems met in seismic data. Thus improvement in present day deconvolution techniques will not likely come from seeking more general statistical assumptions. Instead we must either put more information into the solution or make assumptions which are particularly suited to the task at hand.

The technique treated in this thesis, Minimum Entropy Deconvolution which I shall refer to simply as MED, finds a unique estimate of r from equation (1.3) based on an assumption completely different from those of spiking deconvolution. The MED filter is that one which optimizes the simplicity or sparceness of the output. The term 'minimum entropy' is strictly generic, expressing Wiggins' (1977) goal of increasing the overall apparent order of the data. MED's freedom from the phase and whiteness constraints of traditional deconvolution methods would be a tremendous advantage if a robust and workable MED algorithm were available. This is particularly true in the area of compressing nonminimum-phase marine source signatures in the absence of a useable wavelet estimate.

1.2 OBJECTIVES OF THIS RESEARCH

This thesis has three main objectives. The first is to introduce the reader to Minimum Enropy Deconvolution and recent work in the area by other workers, much of which remains unpublished. The second is to develop a better understanding of the assumptions inherent in MED processing in an attempt at setting guidelines enabling us to avoid applying MED to data for which it is not designed and therefore destined to fail. The third and primary objective of my work is to delineate, and where I can, contribute to, the outstanding problems in the calculation of MED operators with particular emphasis on the problem of multiple extrema.

In Chapter 2 of this thesis I will discuss recent work on MED and the problems which have lead to its apparent fall from grace in the literature. Then in Chapters 3 and 4 I discuss the calculation of MED operators from a heuristic geometrical point of view and address what I consider to be the open problems in MED operator design. Perhaps the most important problem for MED and other statistical techniques, that of when and where to use the method to ensure success is treated in Chapter 5, and that is followed by examples of deconvolution based on results of the previous chapters.

2. MED BACKGROUND

2.1 VARIMAX MED

In 1977 Ralphe Wiggins made what was considered by some to be a major breakthrough in the deconvolution problem. His approach, Minimum Enropy Deconvolution, differed from traditional techniques in that no, often unacceptable, assumptions regarding the reflectivity whiteness or wavelet phase are made. These assumptions, required for spiking deconvolution (Robinson, 1954), are replaced by one simple criterion. That is, that MED designs the linear filter which, when applied to the seismic trace leaves behind the smallest number of large reflectors consistent with the convolutional model and the data. Such a filter operator is designed by maximizing some data simplicity measure with respect to the filter coefficients. The measure chosen by Wiggins(1977) is the varimax norm, defined for a multichannel time series $\underset{\sim}{\mathbf{x}}$ as

$$V(\mathbf{x}) = \sum_{i=1}^{ns} \{ \sum_{j=1}^{nx} x_{ij}^4 / [\sum_{j=1}^{nx} x_{ij}^2]^2 \} \tag{1.4}$$

where x_{ij} is the j'th sample of the i'th trace with nx time samples and ns traces in total. For the single channel case this becomes simply

$$V(\mathbf{x}) = \sum_{j=1}^{nx} x_j^4 / [\sum_{j=1}^{nx} x_j^2]^2 . \tag{1.5}$$

The varimax was borrowed from the field of factor analysis where it is used (Kaiser, 1958) to quantify the simplicity of a matrix of factor loadings. In time series application the behavior of the varimax norm is shown in Figure 1. A single trace consisting of all zeroes and one non-zero spike gives $V=1$. The more non-zero spikes of consistent amplitude there are, the smaller the varimax becomes. The varimax is unaffected by the spacing or polarity of the spikes.

Wiggins coined the term "minimum entropy" since the varimax discriminates against randomness in the deconvolution outcome. Donohoe (1981) verified that for the special case where the output is a random white noise process with time points being identically distributed random variables, the varimax and related norms do indeed reflect the statistical entropy of the trace.

It should be noted that, as defined by equation (1.5) the varimax is closely related to the sample kurtosis K, a statistic representing the peakedness of the probability density function from which the data are sampled. In this case

$$K=\{[\sum_i (x_i-\bar{x})^4/n]/[\sum_i (x_i-\bar{x})^2/n]^2\}$$

$$=nV(\mathbf{x}) \qquad \text{for a zero mean process } \mathbf{x}.$$

where n is the number of points in the sample and we may state that varimax MED designs that filter whose output has the largest kurtosis attainable by filtering.

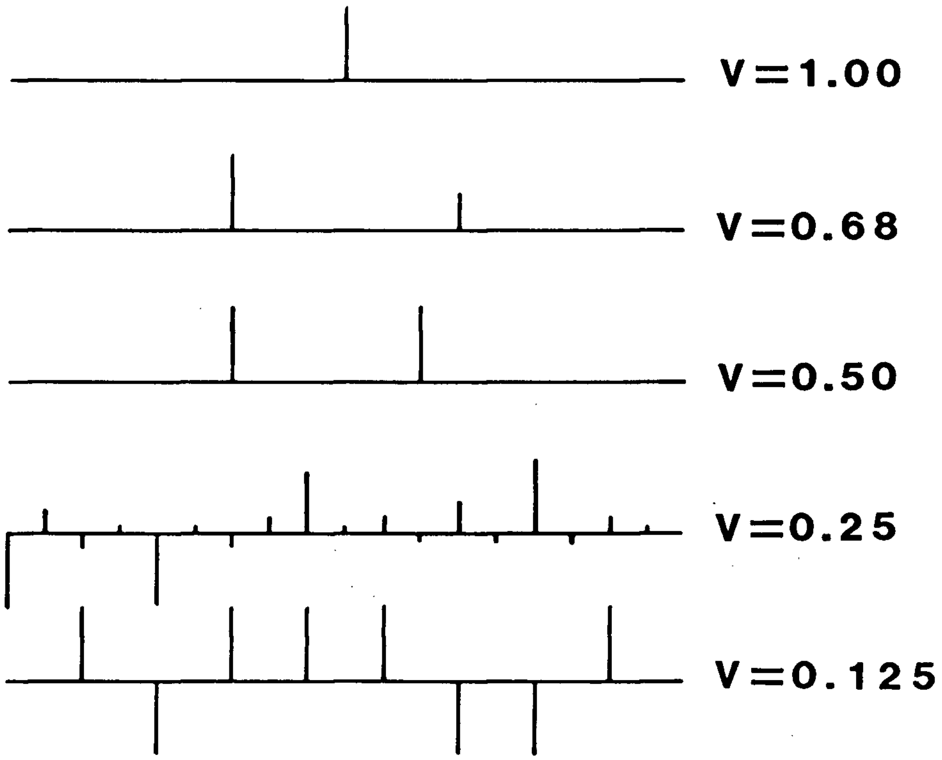


Figure 1.
Varimax norm evaluated for the simple time series shown.

2.2 THE NORMAL EQUATIONS

We have assumed a convolutional model for the seismogram (1.3)

$$\mathbf{x}=\mathbf{w}*\mathbf{r}+\mathbf{n}.$$

It is our aim to obtain some deconvolved seismogram **y** through convolution of the seismogram with a deconvolution filter **f**

$$\mathbf{y}=\mathbf{f}*\mathbf{x}.$$

To find that filter which when convolved with the data as in equation (2.1) yields an output **y** with maximum varimax we maximize $V(\mathbf{y})$ with respect to all filter coefficients f_k, $k=1,\ldots nf$ simultaneously. Setting

$$\frac{\partial V(\mathbf{y})}{\partial f_k} = 0 \qquad k=1,\ldots nf \tag{2.1}$$

we obtain the MED normal equations (Wiggins, 1977) for the single channel case

$$\sum_l f_l \frac{V(\mathbf{y})}{\sum_i y_i^2} \sum_j x_{j-l+1} x_{j-k+1} = \frac{1}{[\sum_i y_i^2]^2} \sum_j y_j^3 x_{j-k+1} \qquad k=1,\ldots nf$$

or,

$$A(\mathbf{y}) \begin{bmatrix} \phi_{xx}(0) & \phi_{xx}(1) . & .\phi_{xx}(nf-1) \\ \phi_{xx}(1) & \phi_{xx}(0) . & . . \\ \cdot & & \\ \cdot & & \\ \phi_{xx}(nf-1) & . . . . & .\phi_{xx}(0) \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ f_{nf} \end{bmatrix} = B(\mathbf{y}) \begin{bmatrix} \phi_{y^3x}(0) \\ \phi_{y^3x}(1) \\ \cdot \\ \cdot \\ \phi_{y^3x}(nf-1) \end{bmatrix} \tag{2.2}$$

where $\phi_{xx}(\tau)$ is the autocorrelation of input **x** at lag τ,

$\phi_{y^3x}(\tau)$ is the crosscorrelation of **x** with $\mathbf{y}^3$ defined as $\mathbf{y}^3 = (y_1^3, y_2^3, \ldots y_{ny}^3)^T$

$$A(\mathbf{y}) = V(\mathbf{y})/\sum_i y_i^2 \text{ , a scalar, and}$$

$$B(\mathbf{y}) = 1/[\sum_i y_i^2]^2 \text{, a scalar.}$$

We must solve equation (2.2) for the f_i given the data x_i but unfortunately the right hand side of (2.2) itself depends non-linearly on **f** since we need it to form **y** and hence $\phi_{y^3x}(\tau)$.

In practice a symmetric starting filter such as

$$_0\mathbf{f} = (0,0,\ldots 1,\ldots 0,0)^T \tag{2.3}$$

is assumed (Wiggins 1977, Ooe & Ulrych 1979, Chacin 1982, Wiggins and Jurkevics 1985), **y**, $\mathbf{y}^3$ and hence $\phi_{y^3x}(\tau)$, A and B are calculated based on this estimate and equation (2.2) solved for **f** using Levinson recursion. This new filter, $_1\mathbf{f}$ is then used to regenerate **y** and the sequence is repeated recursively until it converges on a solution, a process which usually requires 8 to 12 iterations.

The reader familiar with Weiner filtering will note that the MED operator design equation (2.2) is identical in form to the normal equations for design of a least squares waveshaping filter (Robinson and Treitel 1980)

$$\begin{bmatrix} \phi_{xx}(0) & \phi_{xx}(1) \,.\; .\, \phi_{xx}(nf-1) \\ \phi_{xx}(1) & \phi_{xx}(0) \,.\;.\;. \\ \cdot & \\ \cdot & \\ \phi_{xx}(nf-1) \;.\;.\;.\;.\;. & \phi_{xx}(0) \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ f_{nf} \end{bmatrix} = \begin{bmatrix} \phi_{dx}(0) \\ \phi_{dx}(1) \\ \cdot \\ \cdot \\ \phi_{dx}(nf-1) \end{bmatrix} \tag{2.4}$$

Here $\phi_{dx}(\tau)$ is the cross correlation of a desired output **d** with the data **x**.

The solution $\mathbf{f}$ of equation (2.4) is that filter whose output when convolved with $\mathbf{x}$ as in equation (1.2) best approximates $\mathbf{d}$ in a least square sense. We can see therefore, that the single channel MED equations (2.2) differ only by scalar weighting factors A and B, from equations (2.4) above with desired output $\mathbf{y}^3$. This fact allows us to make the simple interpretation of the MED operator design process displayed in Figure 2. Starting with the data $\mathbf{x}$ we form a first estimate of the output ${}_1\mathbf{y}$ by convolution with the starting filter. Then the iteration proceeds by cubing elements of ${}_1\mathbf{y}$ and designing the wave-shaping filter which best estimates ${}_1\mathbf{y}^3$ from the data. Its output ${}_2\mathbf{y}$ is cubed and wave-shaping again is applied to refine the estimate. This process is repeated as shown by the progression of outputs plotted in Figure 2 where the prescripts $i = 1,\ldots 5$ denote the iteration.

Thus far I have considered MED filter design for the single channel case only. That is the case where we can write $\mathbf{x}$ in the form

$$\mathbf{x} = (x_1, x_2, \ldots x_{nx})^{\mathrm{T}}.$$

It is this case which is most adaptable to the geometrical interpretation of part 3. In practice MED is usually applied to find a single deconvolution operator for a multichannel data set $\underset{\sim}{\mathbf{x}}$ consisting of ns segments of data each of length nx and in this case I will write the multichannel data set $\underset{\sim}{\mathbf{x}}$ as:

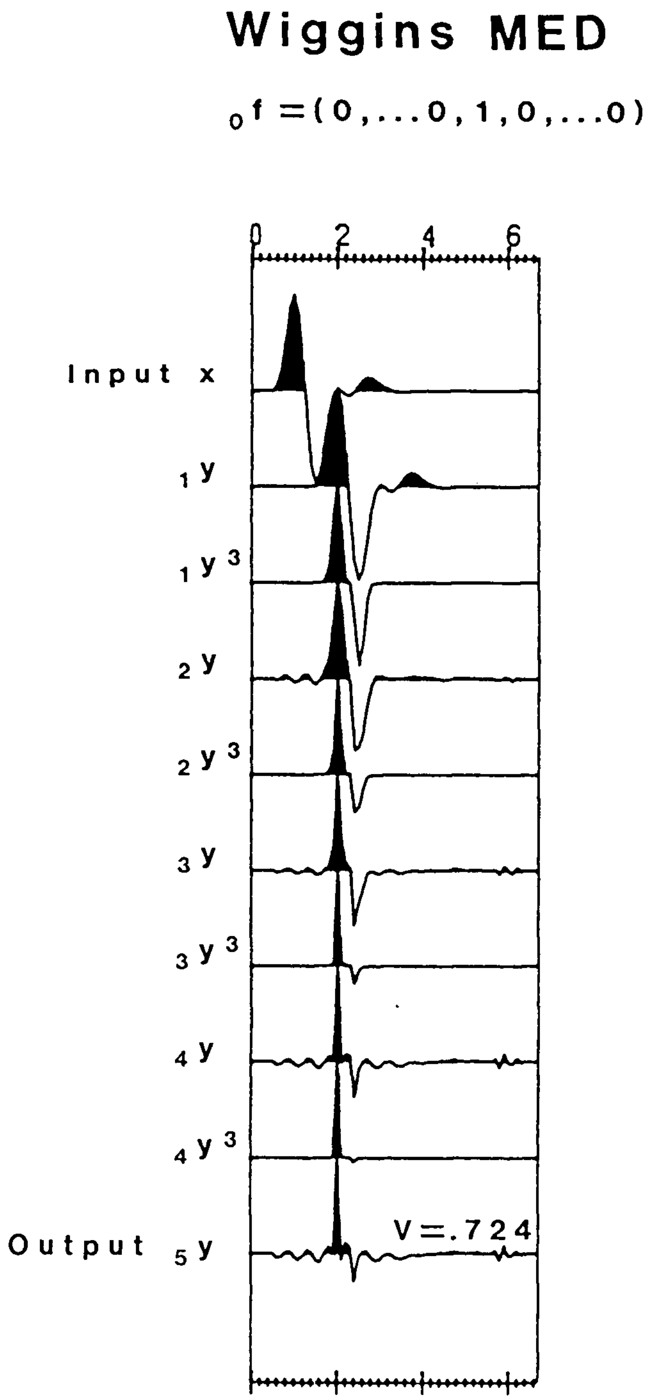


Figure 2.

Successive outputs y and reflectivity estimates y^3 for Wiggins-type MED. Shown from top to bottom are five iterations denoted by prescripts $i=1,\ldots 5$.

$$\underset{\sim}{\mathbf{x}} = \begin{bmatrix} x_{11} & x_{12} & \ldots x_{1,ns} \\ x_{21} & x_{22} & \ldots \\ \cdot & & \\ \cdot & & \\ x_{nx,1} & \ldots & x_{nx,ns} \end{bmatrix} = (\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_{ns}) .$$

Similarly I shall write multichannel filter output $\underset{\sim}{\mathbf{y}}$ as

$$\underset{\sim}{\mathbf{y}} = (\mathbf{y}_1, \mathbf{y}_2, \ldots \mathbf{y}_{ns})$$

and the MED normal equations become:

$$\sum_{i=1}^{ns} A(\mathbf{y}_i) \begin{bmatrix} {}^{i}\phi_{xx}(0) & {}^{i}\phi_{xx}(1) & \ldots & {}^{i}\phi_{xx}(nf-1) \\ {}^{i}\phi_{xx}(1) & {}^{i}\phi_{xx}(2) & \ldots & \\ \cdot & & & \\ {}^{i}\phi_{xx}(nf-1) & \ldots & \ldots & {}^{i}\phi_{xx}(0) \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \cdot \\ \cdot \\ f_{nf} \end{bmatrix}$$

$$= \sum_{i=1}^{ns} B(\mathbf{y}_i) \begin{bmatrix} {}^{i}\phi_{y^3x}(0) \\ {}^{i}\phi_{y^3x}(1) \\ \cdot \\ {}^{i}\phi_{y^3x}(nf-1) \end{bmatrix} \qquad (2.5)$$

where

$$A(\mathbf{y}_i) = V(\mathbf{y}_i)/\sum_l y_{i,l}^2$$

$${}^{i}\phi_{xx}(\tau) = \sum_l x_{i,l}\, x_{i,l-\tau}$$

$$B(\mathbf{y}_i) = 1/[\sum_l y_{i,l}^2]^2 \qquad \text{and,}$$

$${}^{i}\phi_{y^3x}(\tau) = \sum_l y_{i,l}^3\, x_{i,l-\tau} .$$

These equations, which are used in practice in any Wiggins-type algorithm are simply trace averaged versions of equations (2.2), the autocorrelations ${}^{i}\phi_{xx}(\tau)$ and cross correlations ${}^{i}\phi_{y^3x}(\tau)$ for each trace being weighted by trace dependent scalars A and B. Like the single channel case, equation (2.5) is solved recursively by forming $\underset{\sim}{y}$ and hence A, B and $\overset{i}{\phi}_{y^3x}(\tau)$. Then solving for **f**, we renew our estimate of $\underset{\sim}{y}$ and so on. Wiggins (1978) suggests that weights A and B need only be calculated once since they change little in practice from one iteration to the next.

2.3 RELATED APPROACHES

In the period 1977 through 1981 there was a great deal of attention focused on MED. This research dealt mainly with the search for alternative, theoretically justified sparseness criteria. Some of this work undertaken by the Stanford Exploration Project has remained unpublished due to the 4 year delay in public release of SEP reports. Other work, such as that of Deeming (1981), has been presented at SEG meetings and workshops but not in the literature.

MED enjoyed modest industrial application at that time due to its freedom from the cumbersome phase and whiteness constraints of conventional methods. In practice however it failed to live up to the high expectations of its proponents and more recent research such as that of Deeming (1984) and Cabrelli (1985) has focussed on MED shortfalls and how they might be bridged.

In the first paper released after the work of Wiggins (1977), Ooe and Ulrych (1979) provide a clear statement of Wiggins work and its relation to the field of factor anaysis. They propose the use of a modified varimax norm which involves an exponential transform and a slightly different method of summing the varimax over the traces. This norm, the kurtosis norm with exponential transform, K_e, is shown in Table 1. This norm and the varimax with exponential transform contain an adjustable parameter s, which if properly selected improves MED performance in noisy data.

Gray, in a 1979 Ph.D. thesis at Stanford considers the varimax norm as just one in a family of statistics described by Hogg (1972) for testing probability density functions of white random processes. To best understand Gray's reasoning I must first introduce the generalized Gaussian probability density function $f(x,a)$, given by

$$f(x,a) = [a/(2\Gamma(1/a))]exp(-|x|^a) \qquad -\infty<x<\infty \quad (2.6)$$

where $\Gamma(z) = (z-1)!$ is the gamma function.

This distribution and some realizations of independent and identically distributed (i.i.d.) random processes having this probability density function are presented in Figure 3. When $a=1, f(x,a)$ becomes the two-sided exponential distribution. When $a=2, f(x,a)$ becomes the gaussian, and as a tends to infinity $f(x,a)$ tends towards the uniform. Thus the generalized gaussian distribution describes a wide range of random variables from the certain event $a=0$, to the uniformly distributed.

If we consider our sampled time series of length n to be a string of n i.i.d. random variables, then we can model

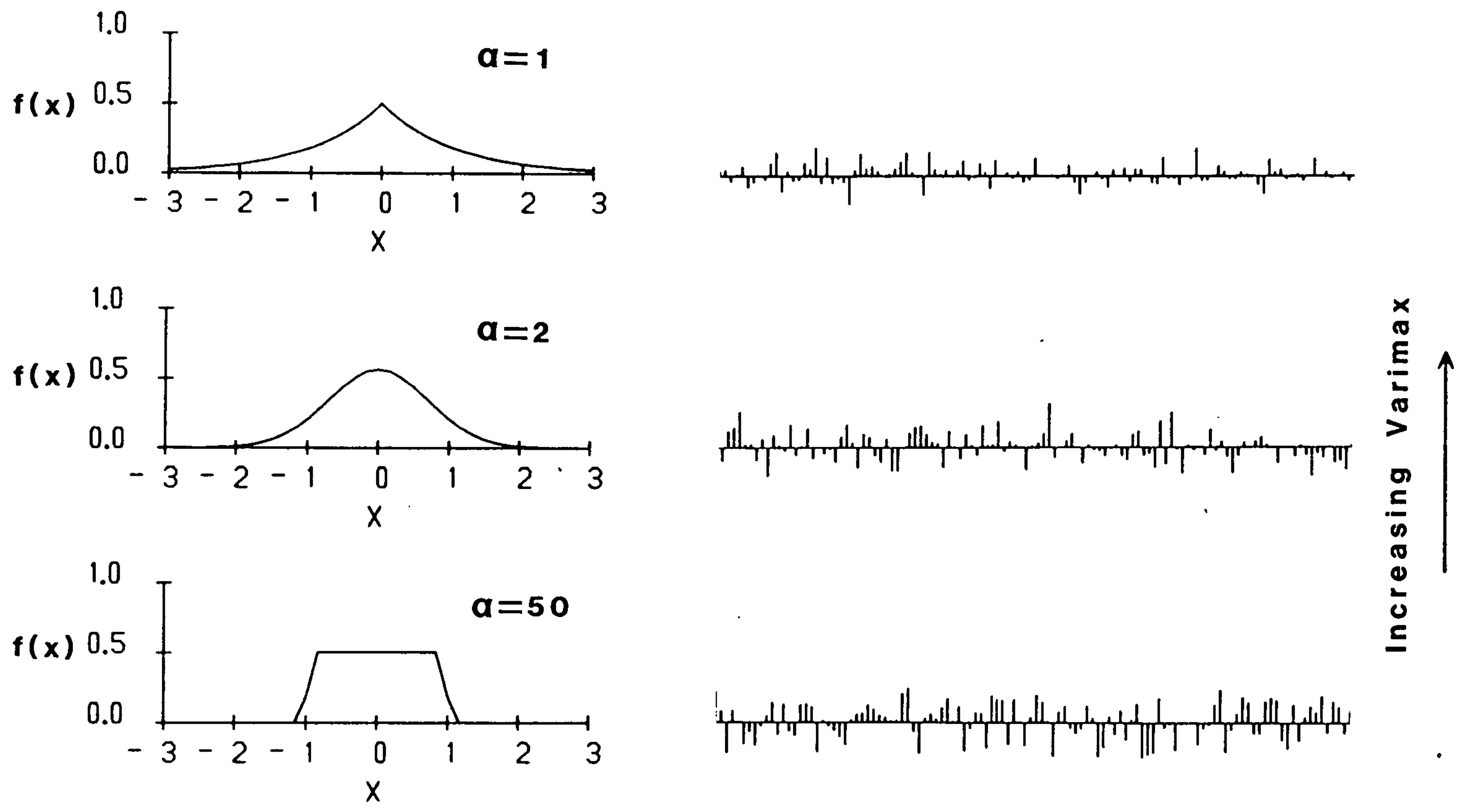


Figure 3.
Three examples of the generalized Gaussian p.d.f. (left) and corresponding realizations of i.i.d. processes (right). Varimax increases with decreasing shape parameter, α, while for a given variance, the Gaussian (α=2) has maximum entropy.

seismograms and reflectivity series as realizations of stationary random processes with amplitude distributions taken from the generalized gaussian family. This is the model for the seismogram and reflectivity series adapted by Gray (1979). Gray notes that the single channel varimax norm is a special case of the statistic $U(\mathbf{x},a_1,a_2)$ given by

$$U(\mathbf{x},a_1,a_2) = \frac{[(1/n)\sum_{i=1}^{n}|x_i|^{a_1}]^{n/a_1}}{[(1/n)\sum_{i=1}^{n}|x_i|^{a_2}]^{n/a_2}} \quad . \tag{2.7}$$

$U(\mathbf{x},a_1,a_2)$ is the "most powerful invariant test statistic" to determine whether a random sample x is drawn from the distribution $f(x,a_1)$ as opposed to $f(x,a_2)$. The larger U becomes, the more probable it is that the sample $\mathbf{x}$ is drawn from the distribution $f(x,a_1)$ and not $f(x,a_2)$. Linear filters designed to maximize the varimax norm are designed therefore to reshape the amplitude distribution of the seismogram away from that described by $a=4$ in the direction of $a=2$. Thus MED is interpreted as an attempt to yield the output with the smallest possible shape parameter a. Under Gray's model the smaller the shape parameter, the more parsimonious or spiky the outcome.

Gray recommends a MED-like scheme "Variable norm ratio deconvolution" which involves adopting the norm $V(\mathbf{y},a_1,a_2)$ shown in Table 1 and choosing $a_1=\hat{a}$, $a_2=2$ where $\hat{a}$ is a shape parameter estimated from the data. Alternatively he presents arguments in favor of $a_1=2$, $a_2=1$. Both schemes allow for adjustment of the a parameters between iterations to control convergence. Furthermore both schemes form the geometric, rather than the arithmetic mean of the norm over several traces.

An interesting outcome of Gray's model is the link between the varimax and Shannon's statistical entropy. Since among probability distributions of fixed variance the Gaussian has maximum entropy, we must maximize the kurtosis (varimax) to minimize entropy when the data are drawn from a distribution which is more peaked than gaussian (leptokurtic). However, we must minimize the varimax when the pdf is less peaked than gaussian (platykurtic). This curious result was observed by Donohoe (1981) and is demonstrated in section 5.1.

Also at Stanford, and at the same time as Gray, Godfrey (1979) completed a Ph.D. thesis entitled "A stochastic model for seismogram analysis" also published as S.E.P. report number 17. Godfrey approaches the problem of unknown wavelet deconvolution by using a stochastic model for the reflectivity explicitly in the problem formulation. The model adapted is a reflectivity sequence in which the reflection coefficients are again i.i.d.. However the reflectivity pdf represents the sum of two populations – isolated large reflectors with a Poisson distribution in time plus Gaussian noise. In this case the pdf describing the reflectivity series is

$$f_{\mathbf{r}}(r) = \lambda\delta(r)+(1-\lambda)G(\sigma_r^2,r^2) \tag{2.8}$$

where $G(\sigma_r^2,r^2)$ is the Gaussian pdf of zero mean and variance

σ_r^2. The Poisson parameter λ sets the mean waiting time between spikes occuring in the reflectivity model at $1/(1-\lambda)$.

Based on this reflectivity model the author builds a model for the deconvolved (or partly deconvolved) seismic trace which is also white but with a pdf known as a 'normal mixture'.

$$f_{\mathbf{x}}(x) = \lambda G(\sigma_n^2, x^2)+(1-\lambda)G(\sigma^2, x^2). \tag{2.9}$$

Here σ^2 is the variance of the seismic trace and σ_n^2 the variance of the "noise" or residual wavelet convolved with the true reflectivity. Based on this probabilistic model of the seismogram a MED-type deconvolution algorithm was proposed which is called "Zero memory non-linear deconvolution" (Godfrey and Rocca 1981). While no norm such as the varimax is explicitly defined in this scheme, deconvolution is achieved by iteratively shaping the data by linear filtering to match a reflection coefficient estimate $\hat{\mathbf{r}}$ for each iteration. That reflectivity estimate is obtained from the data in one of two ways: either as a least squares estimate of the expected value of $\mathbf{r}$ given the seismogram or as the conditional median of $\mathbf{r}$ given $\mathbf{x}$ and the model.

The two reflectivity estimates proposed for non-linear deconvolution are given in Table 1. To calculate them, the probabilistic model parameters σ_r^2, σ_n^2 and λ must first be specified.

Perhaps the best known of all the MED-type deconvolution algorithms is parsimonious deconvolution due to Claerbout (1977a). (See also Postic et al. 1980). After experimenting with a norm of the form $U(\mathbf{x},1,2)$ which was appealing on the basis of the generalized gaussian seismogram model the author abandoned it in favour of the parsimony norm S_2 given by:

$$S_2 = -\sum_i |x_i|^2 \ln |x_i|^2 \tag{2.10}$$

subject to the constraint

$$\sum_i x_i^2 = 1.$$

This norm is identical in form to an expression for parsimony given by Ferguson (1954). Like the varimax, the parsimony was intended for the factor analysis problem of finding that reference frame in which a system of n-dimensional vectors is most simply described. Claerbout derives S_2 from the limit as ϵ tends to zero of $U(\mathbf{x},2,2-\epsilon)$. The method proposed for calculating a MED-type deconvolution operator based on the S_2 norm is an iterative steepest descent algorithm in which convergence can be hastened in the first few iterations by using another norm:

$$S_2 = -\sum_i |x_i|^{2.5} \ln |x_i|^{2.5}$$

subject to

$$\sum_i x_i^{2.5} = 1.$$

Promising examples and a computer program to implement parsimonious deconvolution are included in Claerbout (1977b). Also noteworthy for this thesis are comments by the author that the S_2 norm has many local extrema and that output of the parsimony routine depends on iteration starting point.

Deeming (1981) in a presentation to the S.E.G. annual meeting in Los Angeles drew together all MED work to that date in his "generalized minimum entropy principle" formulation. That is, he demonstrated that all of the norms discussed above can be written in the form.

$$V = (1/nx) \sum_i z_i F(z_i) \tag{2.11}$$

where $z_i = nx \sum_i x_i^2/||\mathbf{x}||^2$

and $||\mathbf{x}|| = [\sum_i x_i^2]^{1/2}$.

In this formulation, Wiggins varimax corresponds to choosing $F(z_i)=nx\cdot z_i$. Grays variable norm deconvolution corresponds roughly to $F(z_i)=z_i^p$. Ooe and Ulrych's exponential transform method corresponds roughly to $F(z_i)=1-e^{-az_i}$ and Claerbout's parsimonious deconvolution corresponds roughly to $F(z_i)=ln(z_i)$. Deeming shows that Wiggins-type normal equations can be developed for any norm of the form (2.11). These equations are similar to equations (2.2) except that the desired output estimate **d** at each iteration is given by the more general

$$d_i = nx\cdot\beta_i/(\sum_i \beta_i z_i) \quad \text{where} \beta_i = \frac{\partial V}{\partial z_i}.$$

Deeming refers to **d** as the "reflection coefficient estimate", formed from the seismogram at each iteration.

A list of the norms discussed above and their respective reflection coefficient estimates is given in

Table 1. It is interesting to plot these reflectivity estimates d_i as a function of the previous iteration output y_i at a given time sample. Figure 4 shows the remarkable similarity among estimates. In all cases the reflectivity estimate surpresses small reflections more than large ones. Only the parsimony is significantly different in character, it surpresses *and negates* small reflections.

Two more recent additions to the MED literature are those of Ulrych and Walker (1982) and Cabrelli (1985). Ulrych and Walker apply MED with exponential transform to the analytic signal. By adopting a complex approach their algorithm is able to cope with a wavelet which undergoes phase shifts within the trace. In his paper, Cabrelli provides a geometrical interpretation of the varimax norm and proposes an alternative, the D-norm, which is supposed to achieve the same goals but using a non-iterative algorithm. The D-norm is also shown in Table 1. It does not however fit into Deeming's generalized minimum entropy formalism and thus requires its own algorithm and has no corresponding reflection coefficient estimate. Further discussion of Cabrelli MED operator design will be saved for section 3 as it is best described in terms of a geometrical interpretation of the varimax norm.

Several other miscellaneous papers on the minimum entropy principle have appeared. In "Seven Essays on Minimum Entropy" (1980) Claerbout proposes that minimum entropy approaches could be pursued not only for the debubbling

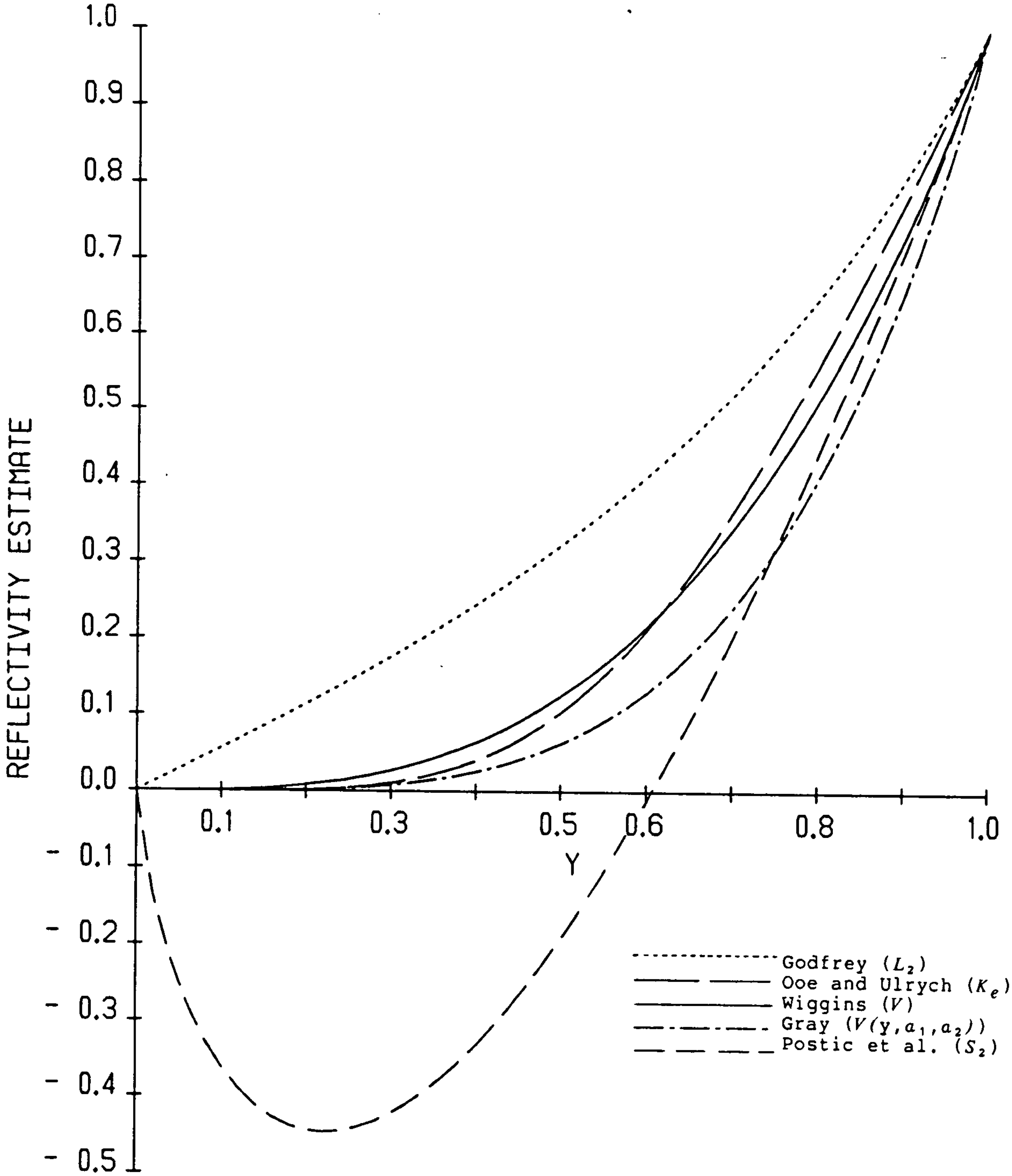


Figure 4.
MED reflectivity estimates d_i plotted as functions of y_i, the seismogram amplitude at the i'th time sample.

problem but also in the selection of absorption parameters in time variant deconvolution, aspects of the migration problem, spectral analysis, and the iterative adjustment of reflection coefficients so as to select and subtract multiples. The author further comments on an alternative formulation of his parsimony norm, the "extrinsic power", comments on the need for spectral balancing and exponential gain recovery before MED processing and speculates on the relationship between parsimony and thermodynamic entropy. The application of minimum entropy methods in decision analysis is considered by Gonzalez-Serano (1980) and in the velocity analysis problem by DeVries and Berkhout (1984).

TABLE I

MED NORMS AND CORRESPONDING REFLECTIVITY ESTIMATES

NORM	REFLECTIVITY ESTIMATE
1. VARIMAX $V(\underset{\sim}{\mathbf{x}})$	
$= \sum_{i=1}^{ns} \left[\dfrac{\sum_{j=1}^{nx} x_{ij}^4}{\left[\sum_{j=1}^{nx} x_{ij}^2\right]^2} \right]$	$d_{ij} = x_{ij}^3$
2. KURTOSIS WITH EXPONENTIAL TRANSFORM $K_e(\underset{\sim}{\mathbf{x}})$	
$= \sum_{i=1}^{ns} \sum_{j=1}^{nx} \dfrac{z_{ij}^4}{\left[\sum_{i=1}^{ns} \sum_{j=1}^{nx} z_{ij}^2\right]^2}$	$d_{ij} = x_{ij} \left[\dfrac{(z_{ij} - z_{ij}^2)}{\sum_{j=1}^{nx} z_{ij}} + K_e z_{ij} \right]$
where $z_{ij} = 1 - exp(-x_{ij}^2/2s^2)$	
3. VARIABLE RATIO NORM $V(\underset{\sim}{\mathbf{x}}, a_1, a_2)$	
$= log \prod_{j=1}^{ns} \dfrac{\left[(1/nx) \sum_{i=1}^{nx} \lvert x_{ij} \rvert^{a_1}\right]^{nx/a_1}}{\left[(1/nx) \sum_{i=1}^{nx} \lvert x_{ij} \rvert^{a_2}\right]^{nx/a_2}}$	$d_{ij} = \lvert x_{ij} \rvert^{a_1 - 1} sgn(x_{ij})$
4. ZERO MEMORY NONLINEAR DECONVOLUTION	
L_2 FORMULATION:	$d_i = \dfrac{\sigma_r^2}{\sigma^2} \left[1 + c \;\; exp\{ \dfrac{-x_i}{2} \left[\dfrac{1}{\sigma_n^2} - \dfrac{1}{\sigma^2} \right] \} \right]^{-1} x_i$
(no explicit norm)	where $c = \lambda\sigma/(1-\lambda)\sigma_n$
L_1 FORMULATION:	$d_i = \begin{cases} 0 & \lvert x_i \rvert < x_c \\ (\sigma_r^2/\sigma^2) x_i & \lvert x_i \rvert \ge x_c \end{cases}$
	where $x_c = \sqrt{2}\sigma_n \{ ln \left[\dfrac{\sigma_r \lambda}{\sigma_n (1-\lambda)} \right] \}^{1/2}$

TABLE I (cont.)

5. PARSIMONY $S_n(\mathbf{x})$

$$= -\sum_{i=1}^{nx} |x_i|^n ln|x_i|^n \qquad d_i = |x_i|^{n-1} sgn(x_i)[1 - ln|x_i|^n]$$

where $\sum_{i=1}^{nx} x_i^2 = 1$

6. D-NORM $D(\underset{\sim}{\mathbf{x}})$

$$= \max_{i,j}(|x_{ij}|/||\mathbf{x}||) \qquad \text{(no equivalent reflectivity)}$$

2.4 PROBLEMS WITH MED

The initial popularity of MED in the geophysical literature lasted only a few years. What at first seemed a promising solution to a perenially unsolved problem in geophysics proved to MED users unworkable. Complaints included poor performance in noisy data compared with conventional methods, (Jurkevicks & Wiggins, 1984) problems whenever high frequency and high amplitude are well correlated in the data (Morley 1980) and poor results when filter length is chosen too long (Morley 1980, Deeming 1985). Cabrelli (1985) shows implicitly that the traditional Wiggins-type algorithm does not succeed in achieving the global maximum of the varimax functional while Chacin (1983) reports gross phase and amplitude errors in simple wavelet estimation problems using a Wiggins-type algorithm with symmetric starting filter given by equation (2.3). It has been further contended by Deeming (1984) that MED's sparse spike train model of the seismogram is neither well understood nor realistic.

The problems with MED, if not the complaints, can be divided into two distinct categories. On the one hand we have unresolved difficulties associated the calculation of MED operators. These include the choice of starting filter and filter length, preprocessing if any, and what type of algorithm to choose. On the other hand there is the problem of model validity, a concern which must be addressed each time MED is applied. If, for example, the reflectivity underlying the seismic trace within our design window does not have the "minimum entropy" property, then MED will design a filter which will give a sparse yet meaningless output. In that case MED has been misapplied to a data set for which the statistical model is invalid and it would be unfair to criticize MED on the grounds that it does not produce sensible results. It is not clear in every case which type of problem is responsible for each of the complaints mentioned at the start of this section. It is however possible that progress can be made on both fronts, that of operator design and selection of applicability criteria. Such progress is the goal of this thesis. Practical and varied real data application will then show whether this work leads to a renewed faith in MED as a workable scheme.

3. A GEOMETRICAL INTERPRETATION

3.1 NOTATION AND INTRODUCTION

There is an intuitive, geometrical interpretation of MED, and least squares filtering in general, which I hope will aid the reader's understanding as it certainly has mine. This heuristic approach is the subject of this section and its development requires the introduction of a few basic concepts.

Consider a segment from a discrete time series **y**

$$\mathbf{y} = (y_1, y_2, \ldots y_{ny})^T.$$

I have written **y** in terms of its sampled values at equally spaced time increments , $i=1,\ldots ny$ in a column vector notation. Indeed as an ordered ny-tuple of real numbers we may consider the time series **y** as a vector in an ny-dimensional Euclidean space, that is, a real vector space over which an inner product is defined. If we consider the set of ny basis vectors for such a vector space:

$$\begin{aligned} {}^1\mathbf{e} &= (1,0,0,\ldots 0)^T \\ {}^2\mathbf{e} &= (0,1,0,\ldots 0)^T \\ &\vdots \\ {}^{ny}\mathbf{e} &= (0,0,0,\ldots 1)^T \end{aligned} \tag{3.1}$$

then we can think of these vectors as forming a coordinate system. Figure 5 shows the time series **y** in the case ny=3, our day to day Euclidean 3-space. Note that the time samples y_1, y_2 and y_3 are simply the coordinates defining the vector and in this section I shall deal as much as possible with

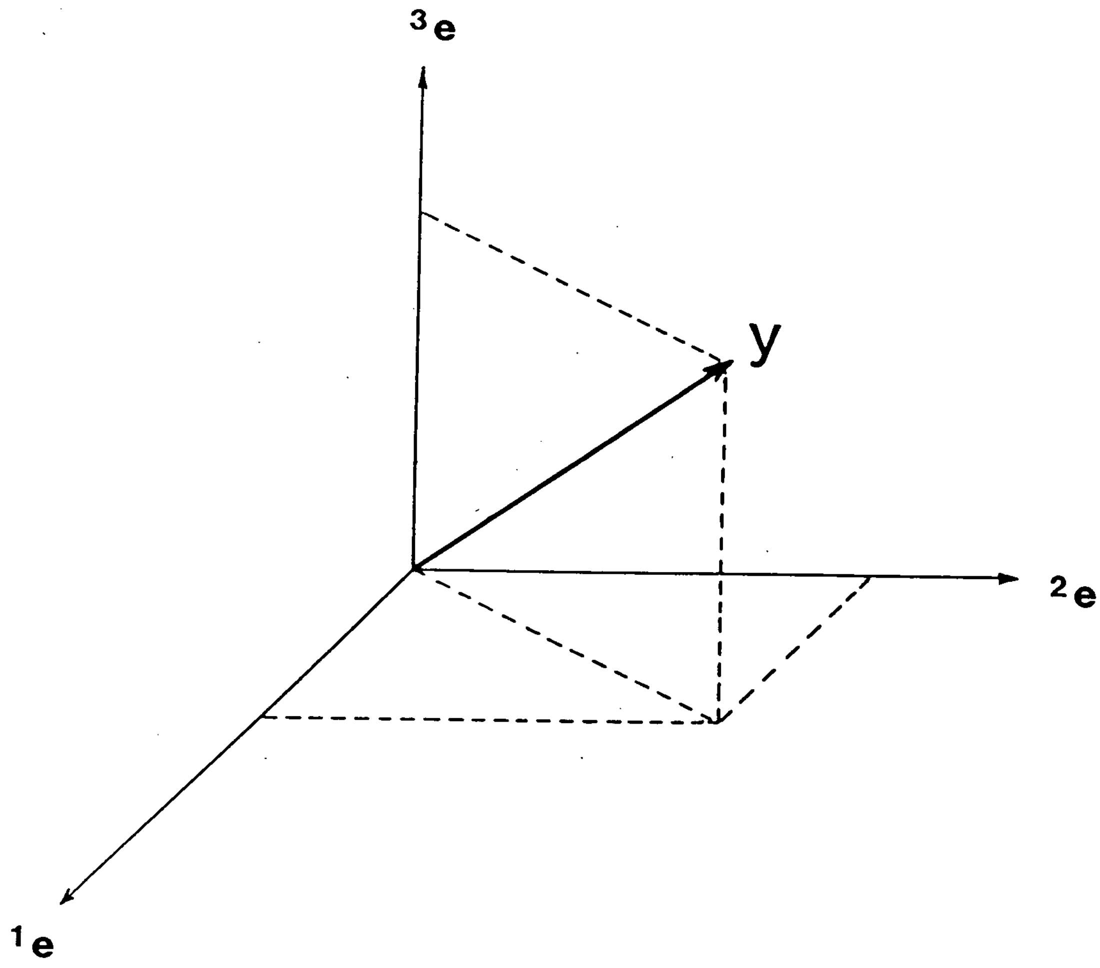


Figure 5.
A time series **y** is considered a vector in an ny-dimensional Euclidean space R^{ny}. If ny=3, time samples y_1, y_2 and y_3 are coordinates describing the vector in R^3.

the easily visualized ny=3 case. An important concept in least squares filtering, the sum of squared errors between two times series, say **y** and **d** is given by

$$\epsilon^2 = \sum_{i=1}^{ny} (y_i - d_i)^2 . \qquad (3.2)$$

This is recognized as simply the squared vector distance between **y** and **d** so

$$\epsilon = ||\mathbf{y}-\mathbf{d}|| . \qquad (3.3)$$

In particular the vector distance given by

$$\epsilon = ||\mathbf{y}-{}^{i}\mathbf{e}|| . \qquad (3.4)$$

that is the L_2 proximity of the **y** vector to the i'th axis in Figure 5 is the squared error between the time series **y** and a unit spike at the i'th lag, the quantity minimized by a Weiner spiking filter.

3.2 GEOMETRICAL INTERPRETATION OF LEAST SQUARES FILTERING

As mentioned in section 2.2 the design of MED operators is a recursive application of Weiner wavelet shaping. Fortunately Weiner shaping or least squares filtering has a simple geometrical analog, one which has been explored extensively by N. de Voogd (1972) and which I discuss here before tackling MED itself.

The Weiner wavelet shaping filter as discussed by Robinson and Treitel (1980) is that filter **f** which when convolved with an input time series **x** best reproduces a predetermined desired output, **d** in a least squares sense.

We may write the convolution as

$$\mathbf{y} = \mathbf{x} * \mathbf{f}.$$

Alternatively we may write output $\mathbf{y}$ as the matrix product of $\mathbf{f}$ with a cyclic matrix $\mathbf{X}$. For the case $nx=2, nf=2$ and $ny=3$ we may write

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} x_1 & 0 \\ x_2 & x_1 \\ 0 & x_2 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \tag{3.5}$$

The columns of $\mathbf{X}$ are themselves vectors given by

$$\begin{aligned} {}^1\mathbf{v} &= (x_1, x_2, 0)^T \\ {}^2\mathbf{v} &= (0, x_1, x_2)^T. \end{aligned} \tag{3.6}$$

Now it can easily be seen that the convolution $\mathbf{y}$ is simply a linear combination of these ${}^i\mathbf{v}$ vectors

$$\mathbf{y} = \sum_{i=1}^{nf} f_i \, {}^i\mathbf{v}. \tag{3.7}$$

What then is the geometrical significance of the convolution? Figure 6 shows that the vectors ${}^i\mathbf{v}$ define a subspace Ω of the output vector space R^3. Since the ${}^i\mathbf{v}$ are linearly independent (see e.g. Cabrelli, 1985) they form a basis for that subspace and any linear combination of these vectors will be confined to Ω, in this case the shaded plane of Figure 6. For a given data set $\mathbf{x}$ then, linear filter outputs are confined to the Ω hyperplane.

Given a desired output $\mathbf{d}$ for a wave shaping filter, the filter output is that vector $\mathbf{y}$ in Ω which is closest to $\mathbf{d}$. As shown in Figure 7 this means $\mathbf{y}$ is simply the projection

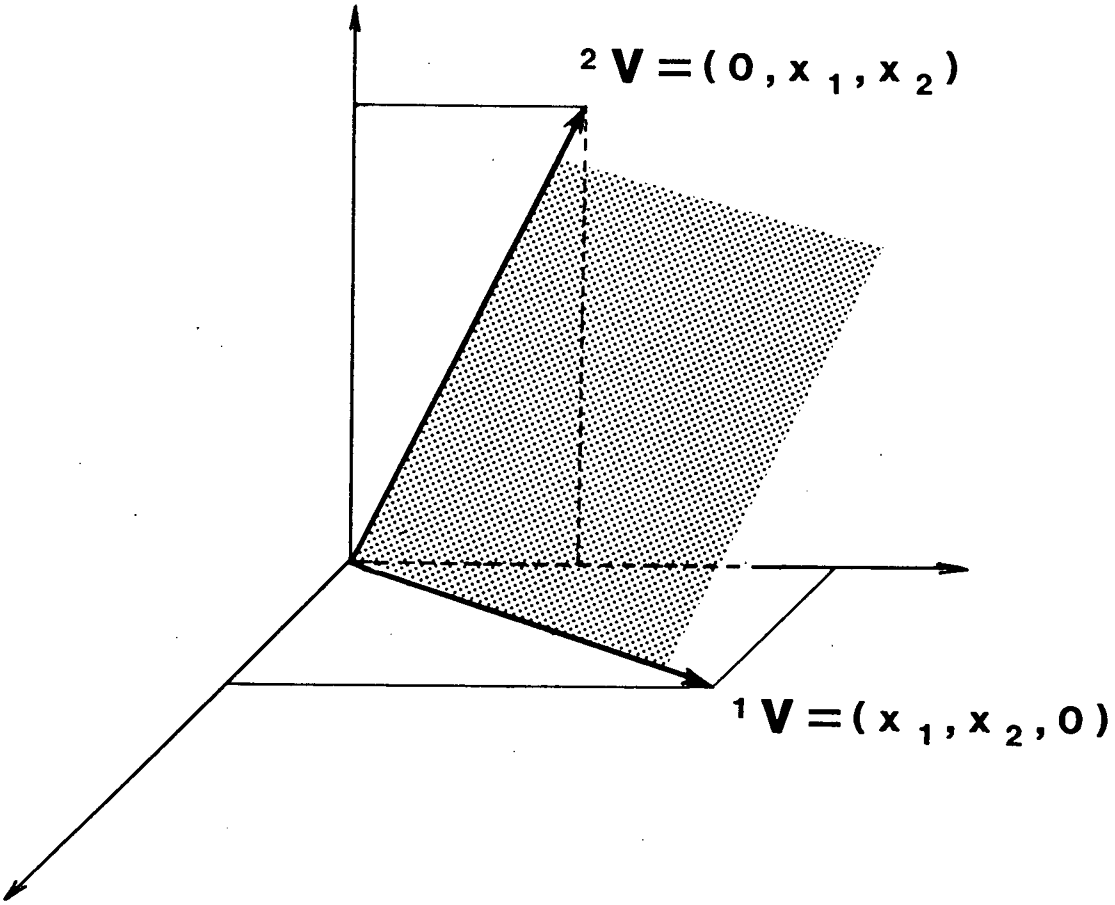


Figure 6.
The vectors $^{i}\mathbf{v}$, $i=1,...nf$ span a subspace Ω of R^{ny} to which all realizable filter outputs, $\mathbf{f} * \mathbf{x}$ are confined. In the case $nx=2$, $nf=2$, $nf=3$, Ω is the shaded plane shown.

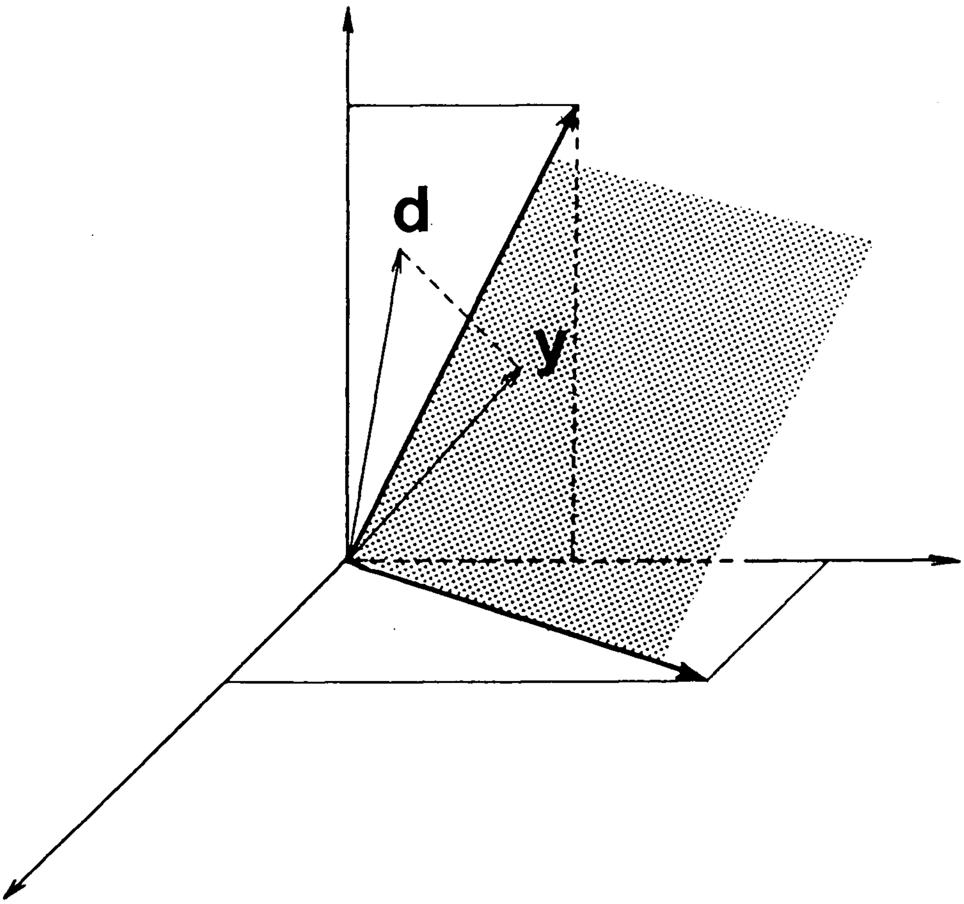


Figure 7.
The least squares filter approximation to **d** is **y**, the orthogonal projection of **d** onto subspace Ω – in this case the shaded plane. For projection MED the vector **y** is used to form a new vector $\mathbf{d}=\mathbf{y}^3$ and hence, by projection, **y** for the next iteration.

of $\mathbf{d}$ onto the plane. This leads us intuitively to the projection method of wavelet shaping. While Robinson and Treitel (1980) describe a design process for wave shaping filters which involves solving the Toeplitz system of equations (2.4) for $\mathbf{f}$ and then convolving $\mathbf{f}$ with the data to get $\mathbf{y}$, the simple alternative of projecting the desired output vector onto the "plane" of possible filter outputs yields the same output. No filter is actually calculated.

The above discussion suggests a quick computational method for performing MED using the projection method. This method, conceived by Toshi Matsuoka, is particularly suited to comparison of the various MED norms of Table 1, since they can each be represented by a reflectivity estimate to be projected onto the common subspace Ω. The projection algorithm proceeds as follows:

1. Compute from the data $\mathbf{x}$, vectors ${}^{i}\mathbf{v}$, $i=1,...nf$ and hence orthonormal basis vectors ${}^{i}a$, $i=1,...nf$ spanning Ω. This is done using the modified Gram Schmidt process described by de Voogd (1974).
2. Compute the desired output $\mathbf{d}$. For instance, to perform Varimax MED we would choose $\mathbf{d}=\mathbf{y}^3$ based on the output $\mathbf{y}$ of the preceding iteration
3. Calculate the projection of $\mathbf{d}$ onto each ${}^{i}a$, a_i given by
$$a_i = \mathbf{d}\cdot{}^{i}a. \tag{3.8}$$
4. Form the filter output $\mathbf{y}$ from a linear combination of the ${}^{i}a$

$$\mathbf{y} = \sum_{i=1}^{nf} a_i \, {}^{i}a. \tag{3.9}$$

The orthogonalization of the ${}^{i}\mathbf{v}$ vectors need only be done once and steps 2 through 4 are repeated until $\mathbf{y}$ converges. Very similar to the well known Gram Schmidt procedure, the method of de Voogd has three advantages:

1. It takes advantage of the cyclic nature of matrix $\mathbf{X}$ to speed calculation.
2. It allows for stabilization of the algorithm through a process equivalent to the addition of a ridge parameter in the autocorrelation matrix of (2.4).
3. It allows calculation of the equivalent filter $\mathbf{f}$ if required.

A simple example of this method is given below. Take the case with data $\mathbf{x}=(1,2)^{\mathsf{T}}$ and starting filter ${}_{0}\mathbf{f}=(0,1)^{\mathsf{T}}$. Our reflectivity estimate $\mathbf{d}$ is given by

$$\begin{aligned} \mathbf{d} &= \mathbf{y}^3 \\ &= (\mathbf{x}*\mathbf{f})^3 \\ &= (1^3, 2^3, 0^3)^{\mathsf{T}} \\ &= (1,8,0)^{\mathsf{T}}. \end{aligned}$$

The next MED iteration therefore seeks to shape the filter output $\mathbf{y}$ into $(1,8,0)$.

The subspace of all possible linear filter outputs is spanned by the vectors ${}^{1}\mathbf{v}$ and ${}^{2}\mathbf{v}$ which are in this case given by

$${}^{1}\mathbf{v} = (1,2,0)^{\mathsf{T}}$$

$$\text{and} \quad {}^{2}\mathbf{v} = (0,1,2)^{\mathsf{T}}.$$

We orthogonalize the ${}^{i}\mathbf{v}$, $i=1,2$ to get an orthogonal basis for Ω such as:

$$^1a = 1/\sqrt{5}(1,2,0)^T$$

$$\text{and} \quad ^2a = 1/\sqrt{105}(-2,1,10)^T.$$

Then the filter output **y** which is closest to **d** in a least squares sense is simply the projection onto the plane of filter outputs.

$$\mathbf{y} = (\mathbf{d}\cdot{}^1a)\,{}^1a+(\mathbf{d}\cdot{}^2a)\,{}^1a$$
$$\approx (3.3, 6.9, .57)^T$$

This result would be used to revise the reflectivity estimate **d**, project it onto the plane and so on.

Armed with an understanding of the projection method one can easily demonstrate its equivalence to the Wiggins MED algorithm. Consider the action of forming a reflectivity estimate $\mathbf{d}=\mathbf{y}^3$, projecting this onto the Ω plane, and cubing the projection to obtain yet another vector to be projected. Such an algorithm will converge when the cube of the vector **y** projected onto Ω yields **y** itself. In this case $\mathbf{y}-\mathbf{y}^3$ is orthogonal to the Ω (hyper)plane and from this simple criterion we may derive the MED normal equations without an explicit norm and without taking any derivatives.

Set

$$(\mathbf{y}-\mathbf{y}^3) \perp \Omega \tag{3.10}$$

$$\text{then} \quad (\mathbf{y}-\mathbf{y}^3) \perp {}^k\mathbf{v} \qquad k=1,\ldots nf.$$

Since the vectors ${}^k\mathbf{v}$ have components ${}^kv_j=x_{j-k+1}$ I may write

$$\sum_{i=1}^{ny} (y_j-y_j^3)x_{j-k+1} = 0. \tag{3.11}$$

Now since **y** is equivalent to a filter output with components

$$y_i = \sum_{l=1}^{nf} f_l x_{i-l+1}$$

(3.11) becomes

$$\sum_{j=1}^{ny} [(\sum_{l=1}^{nf} f_l x_{i-l+1}) - y_j^3] x_{j-k+1} = 0. \tag{3.12}$$

Rearranging I obtain

$$\sum_l f_l (\sum_j x_{j-l+1} x_{j-k+1}) = \sum_j y_j^3 x_{j-k+1} \qquad k=1, \ldots nf. \tag{3.13}$$

Equations (3.13) above are identical to the single channel MED normal equations (2.2) to within a scalar multiplier.

3.3 GEOMETRICAL INTERPRETATION OF MED

Returning now to the field of MED we will look at the varimax norm. The Varimax has been considered by Cabrelli (1985) to be a two stage transformation acting on the vector $\mathbf{y}$ where

$$\mathbf{y} = (y_1, y_2, \ldots y_{ny})^{\mathsf{T}}.$$

The first stage is

$$T_1(\mathbf{y}) = (y_1^2/||\mathbf{y}||^2, y_2^2/||\mathbf{y}||^2, \ldots y_{ny}^2/||\mathbf{y}||^2)$$
$$= \mathbf{y}'$$

and the second stage simply

$$T_2(\mathbf{y}') = ||\mathbf{y}'||^2$$

so that

$$V(\mathbf{y}) = T_2(T_1(\mathbf{y})).$$

The first stage or T_1 transformation, maps the vector from data space R^{ny} to a subset, H defined by

$$H = \{\mathbf{y} \in \mathrm{R}^{ny} \mid \sum_i y_i = 1,\ \ y_i \geq 0,\ \ i = 1, 2, \ldots ny\}.$$

The case for ny=3 is shown in Figure 8. Here H is the hatched area of the plane given by

$$(y_1 - 1/3) + (y_2 - 1/3) + (y_3 - 1/3) = 0$$

and falling in the first quadrant. The varimax norm therefore represents the length of the vector $T_1(\mathbf{y})$ shown in the figure as $\mathbf{y}'$. The varimax then, increases with the distance between $\mathbf{y}'$ or $\mathbf{y}$, and the center of H, Cabrelli then proposed a simpler norm which measures not the distance of $\mathbf{y}'$ from the centre of H but the proximity of $\mathbf{y}$ itself to the axes ${}^i\mathbf{e}$ of Figure 8. To obtain the sparsest output vector $\mathbf{y}$, Cabrelli minimizes the quantity $||\mathbf{y}/||\mathbf{y}|| - {}^k\mathbf{e}||$ over all axes $k=1, \ldots ny$. It can be shown that minimizing the above quantity is equivalent to maximizing $|y_k/||\mathbf{y}||\ |$ so that Cabrelli defines his new norm as follows

$$D(\mathbf{y}) = \max_{1 < k \leq ny} |y_k| / ||\mathbf{y}||.$$

This new norm, where $|y_k/||\mathbf{y}||\ |$ is maximized over all possible vertices is called the "D-norm". Just as the set of realizable filter outputs Ω formed a subset of R^3 so does it form a subset of the H plane. For the ny=3, nx=2 case considered so far, I have found this subset to be an ellipse which changes its character depending on the input data $\mathbf{x}$.

What then occurs when we choose a starting filter ${}_0\mathbf{f}$ and proceed with a Wiggins-type MED algorithm? It is quite easy to visualize in terms of the transformed vector $\mathbf{y}'$. Referring to Figure 9 we see that as the iteration proceeds

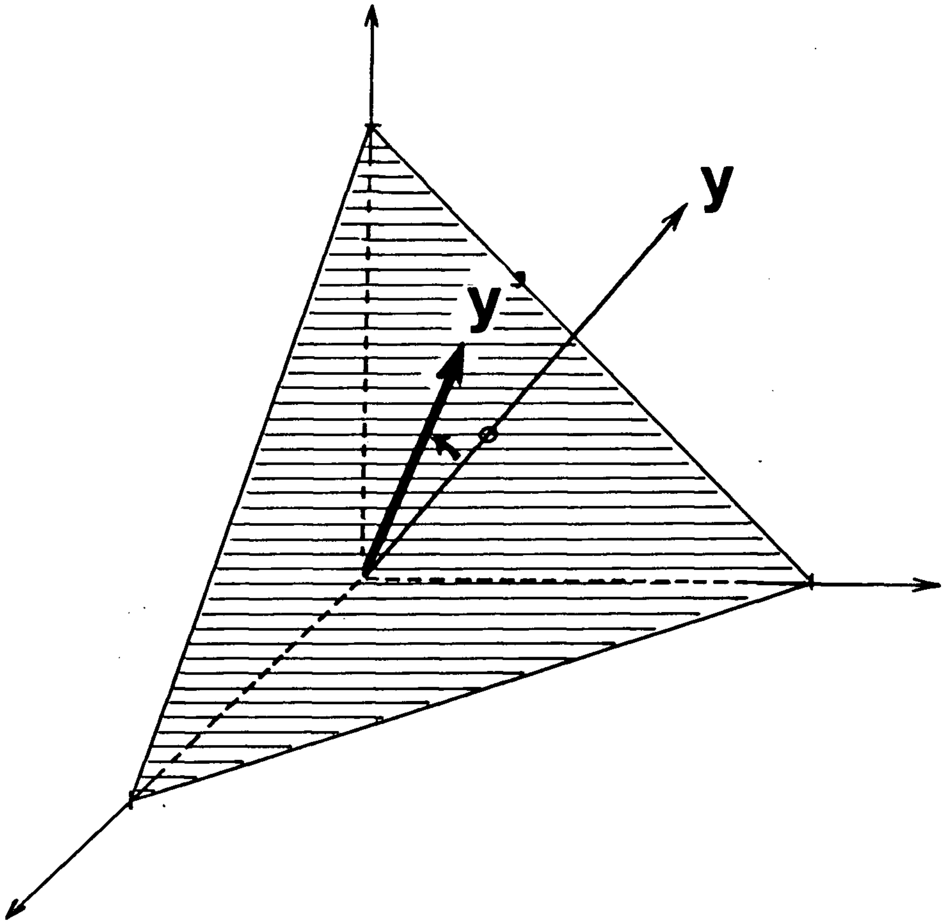
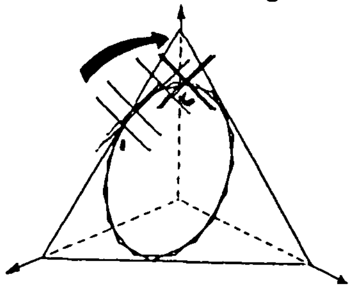


Figure 8.
For the case $ny=3$ the hatched plane is the subset H on which all vectors $\mathbf{y}'=T_1(\mathbf{y})$ lie. Varimax($\mathbf{y}$) is the length of $\mathbf{y}'$ and hence increases toward the axes.

Starting filter=(0.0,1.0)

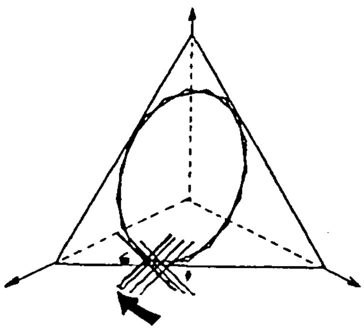


ITERATION	VARIMAX
1	.5148
2	.5428
3	.6043
4	.6250
5	.6257
6	.6257

$\vec{X}$=(1.000,1.190)

Filter output=(-.4599,.3406,1.0567)

Starting filter=(1.0,0.0)



ITERATION	VARIMAX
1	.5148
2	.5251
3	.5301
4	.5307
5	.5308
6	.5308

$\vec{X}$=(1.000,1.190)

Filter output=(.9689,1.4003,.2943)

Figure 9.

Iteration paths and final 3-point output of Wiggins MED after 6 iterations. Crosses indicate the T_1 transformations of y after successive iterations. Starting filters given by (top) ${}_0f=(0,1)^T$, and (bottom) ${}_0f=(1,0)^T$ lead to two distinct extrema.

the vector y' attempts to maximize its length while remaining on the ellipse. Shown in Figure 9 are two examples of the Wiggins MED procedure. At each iteration an 'X' has been plotted representing the endpoint of the corresponding y' vector. After six iterations starting with ${}_0\mathbf{f}=(0,1)^T$ the algorithm has converged to the top of the ellipse shown on the left hand side of Figure 9. However, when we choose ${}_0\mathbf{f}=(1,0)^T$ as shown on the right, quite a different extremum is reached. It is clear from the diagram that there are in this case two possible extrema in the varimax functional and whether the global extremum is reached depends on our choice of starting filter. This and the work in subsequent chapters shows that in general:

1. Minimum entropy norms have multiple extrema, and
2. There is no real basis for choosing a starting filter of the symmetric form

$$ {}_0\mathbf{f} = (0,0,\ldots 1,\ldots 0,0)^T \tag{2.3} $$

if the global extremum is desired.

4. ASPECTS OF MED OPERATOR DESIGN

4.1 MULTIPLE EXTREMA

In the previous section it was shown in simple geometrical terms that the Wiggins MED algorithm may reach local extrema and how these come about. The goal of this section is to ensure that the global extremum is reached.

To envision the workings of the MED algorithm I have been using the unrealistically simple case of $nx=2$, $nf=2$ and $ny=3$. To allow more realistic examples I will now adopt a scheme in which $nf=3$, nx can take on any value, and $ny=nf+nx-1$. Figure 10 shows the three point filter vector

$$\mathbf{f} = (f_1, f_2, f_3)^T.$$

If we specify the constraint

$$||\mathbf{f}|| = 1.$$

That is, let $\mathbf{f}$ lie on the unit sphere, then we can adopt a spherical coordinate representation for $\mathbf{f}$,

$$f_1 = \sin\theta\cos\phi$$

$$f_2 = \sin\theta\sin\phi$$

$$f_3 = \cos\theta.$$

Now any three point filter $\mathbf{f}$ can be represented in terms of two parameters θ and ϕ, for instance

$$\mathbf{f} = (1,0,0)^T$$

becomes $(\theta,\phi) = (\pi/2, 0)$.

More importantly any functional of $\mathbf{f}$ can be contoured in the (θ,ϕ) plane and the particular functional we are interested in is $V(\mathbf{f})$ where

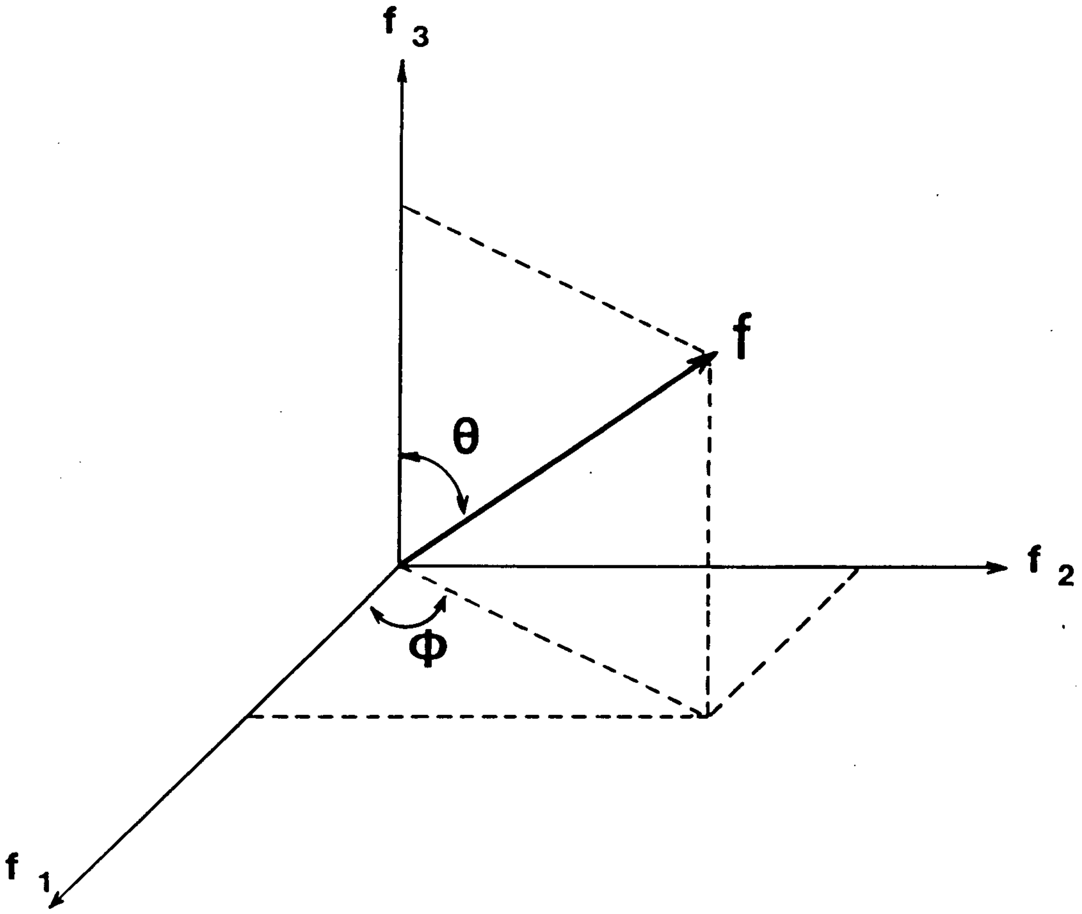


Figure 10.
The three point filter **f** can be parameterized in terms of θ and ϕ.

$$V(\mathbf{f}) = V(\mathbf{f}*\mathbf{x}).$$

Thus for any input data set **x** a contour map representing the shape of the varimax as a functional of the filter applied can be produced. Furthermore, by constructing the data set **x** so as to contain a wavelet with a reasonable three point inverse, it is possible to see which peak in the varimax, if any, coincides with the "right" answer. Figure 11 is an example of such a plot. In this case the input data is a sparse spike series convolved with a simple three point wavelet. To these data I have applied all possible three point filters, at least 1250 of them, sampling the entire range of θ from 0 to π and ϕ from 0 to 2π. The varimax of the resulting filter output has been contoured and it is clearly seen that there are several extrema. One of the first things we notice about Figure 11 is its symmetry. The right hand side of the plot corresponding to $\phi=\pi$ to 2π can be generated by a 180 degree rotation of the left hand side. This symmetry arises from the insensitivity of the varimax to signal polarity. That is, for each maximum in the varimax functional there is a corresponding maximum which will give an output reversed in polarity. MED algorithms may then, induce a polarity reversal in the filtered trace, a fact all too familiar to MED users.

A Wiggins-type MED algorithm was run on the seismogram used to generate Figure 11. The evolution of the filters over the five to six iterations required for convergence is traced with a bold arrow. It is observed that the Wiggins

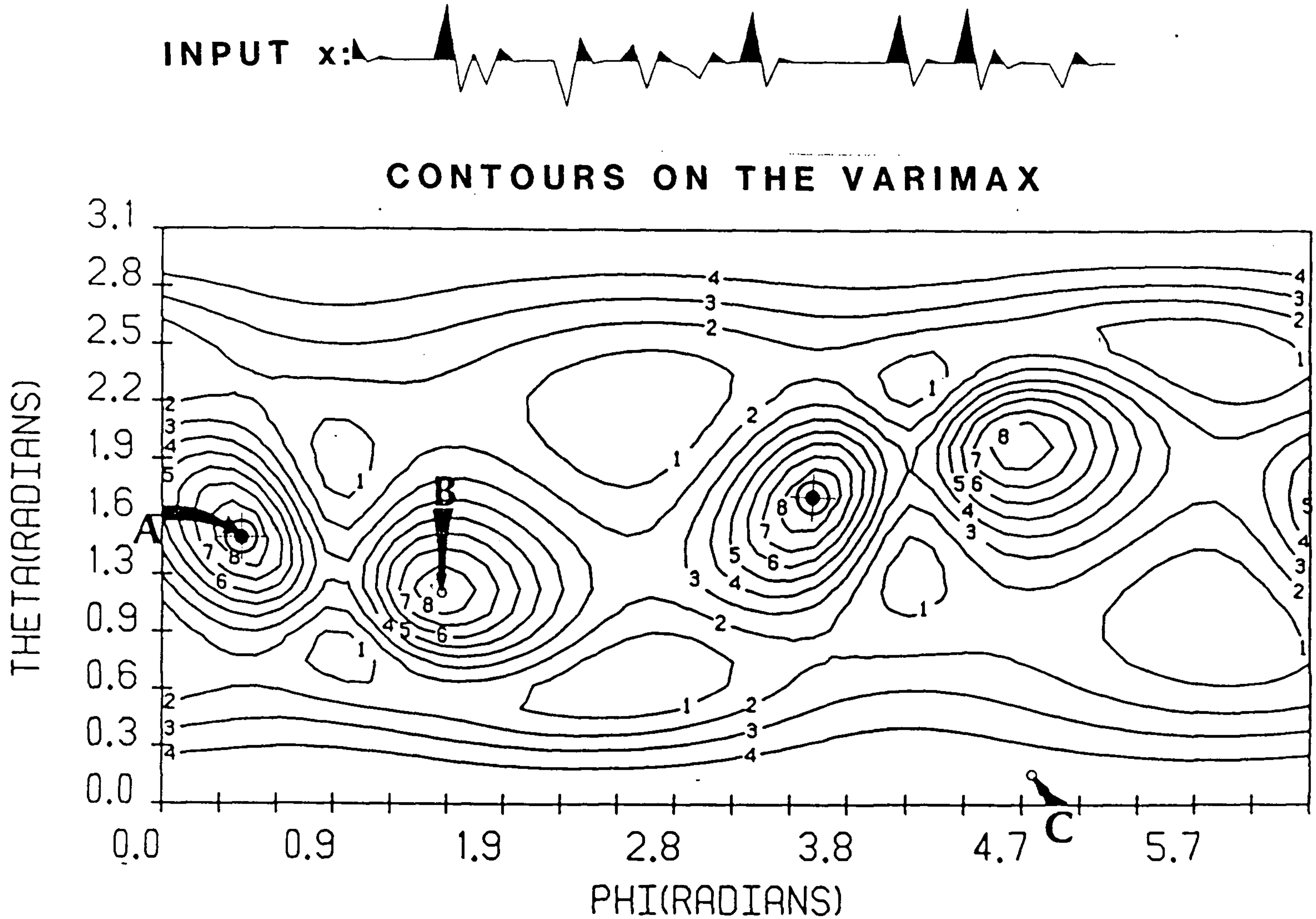


Figure 11.

Relative varimax of the filtered trace contoured for all 3-point filters $f(\theta,\phi)$. Input data set is shown above. Heavy arrows show MED iteration paths for (A): ${}_0f=(1,0,0)^T$, (B): ${}_0f=(0,1,0)^T$, and (C) ${}_0f=(0,0,1)^T$. Circled dots denote global maxima.

MED algorithm does indeed find filters which maximize the varimax, but when the iteration starting filter is varied, different local extrema are reached. Three distinct maxima are sampled by starting filters of the form:

$$
\begin{aligned}
&{}_0\mathbf{f}=(1,0,0)^{T} \\
&{}_0\mathbf{f}=(0,1,0)^{T} \qquad\qquad (3.14)\\
\text{and}\quad &{}_0\mathbf{f}=(0,0,1)^{T}.
\end{aligned}
$$

Those not reached are simply polarity reversals of the first three and are found by negating the starting filter. Still a more interesting example is given in Figure 12. In this case the input data set is the simple minimum phase wavelet given by

$$\mathbf{w} = (.64,.80,.24)^{T}.$$

Although this wavelet has the minimum phase property note that its largest value does not occur at the first time sample. While the contour map presented is for a lone wavelet, the contours for a sufficiently sparse seismogram containing this wavelet have the same character. When the Wiggins-type MED algorithm is run on this example for starting filters of the form (3.14), an interesting and unexpected result surfaces. The global maximum of the varimax is not reached with *any* of the starting filters!

What is the reason for this unexpected finding? The significance of the various extrema lies in the output wavelet lag. In fact the four independent extrema of Figure 12 coincide almost exactly with Weiner filters designed to spike the three point wavelet at lags *1*, *2*, *3*, and *4* of the

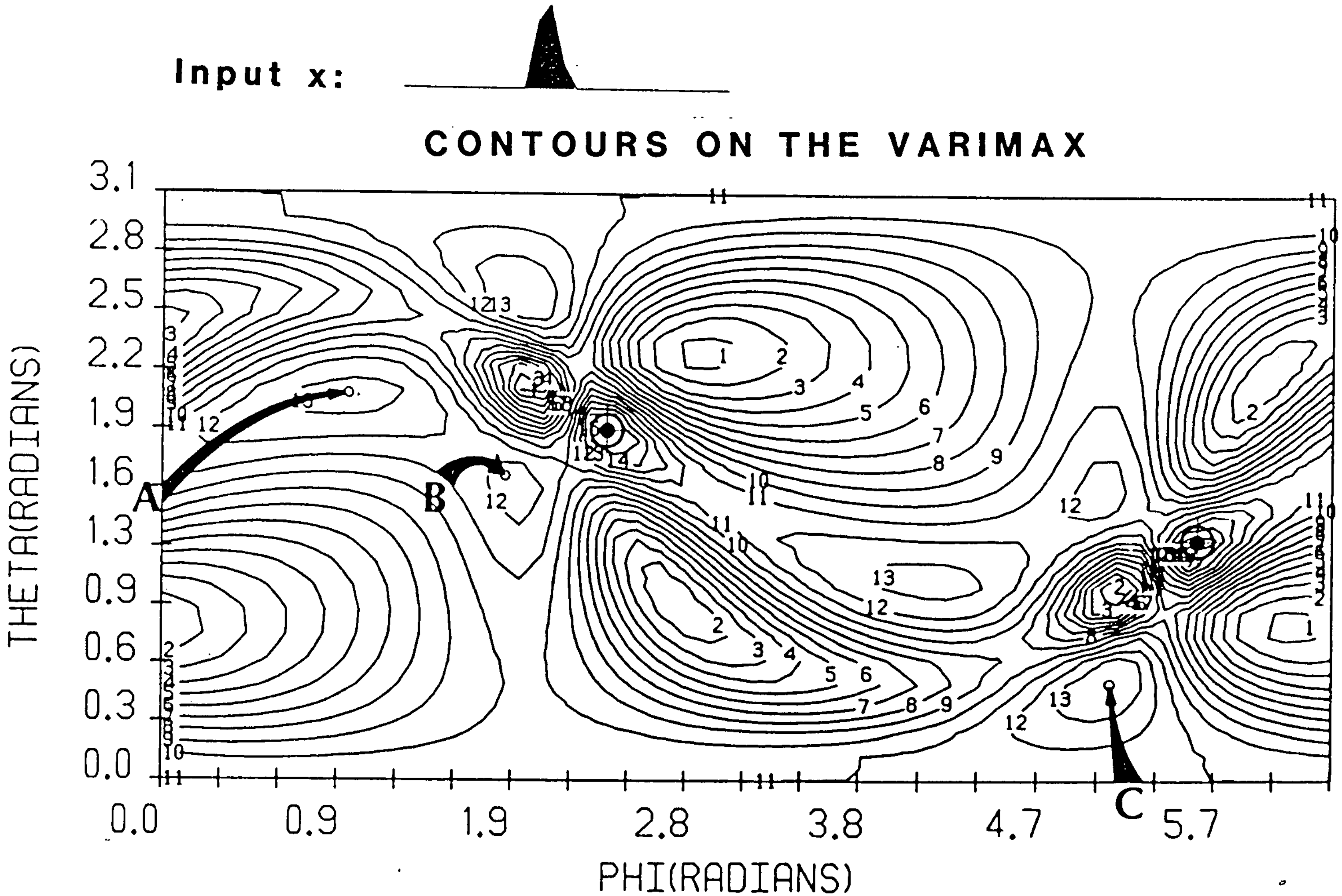


Figure 12.
Relative varimax of the filtered trace contoured for all 3-point filters $f(\theta,\phi)$. Input data set is shown above. Heavy arrows show MED iteration paths for (A): ${}_0f=(1,0,0)^T$, (B): ${}_0f=(0,1,0)^T$, and (C): ${}_0f=(0,0,1)^T$. In this case the global extrema (circled dots) are *not reached*.

filter output $\mathbf{f} * \mathbf{w}$. Recall that the MED algorithm seeks to shape the wavelet into a lagged version of its cube as demonstrated in Figure 2. That lag is determined by the starting filter and for the first iteration of the MED operator design process we have a desired output given by:

$$\mathbf{d} = [(.64,.80,.24,0,0)^3]^T \quad \text{for} \quad {}_0\mathbf{f}=(1,0,0)^T$$
$$\mathbf{d} = [(0,.64,.80,.24,0)^3]^T \quad \text{for} \quad {}_0\mathbf{f}=(0,1,0)^T$$
$$\text{or} \quad \mathbf{d} = [(0,0,.64,.80,.24)^3]^T \quad \text{for} \quad {}_0\mathbf{f}=(0,0,1)^T.$$

The effect of cubing $\mathbf{y}$ is to accentuate the major peaks in the wavelet at the expense of the smaller ones. If, as in this example, the greatest value in the wavelet occurs at w_2, that is $|w_2|>|w_i|$ for all other i, then MED filtering will tend to place a spike at lag two of the output wavelet for a starting filter of the form ${}_0\mathbf{f}=(1,0,0)^T$. Similarly it will place a spike at lag 3 for ${}_0\mathbf{f}=(0,1,0)^T$ or lag 4 for ${}_0\mathbf{f}=(0,0,1)^T$. Weiner filter theory however, (Robinson and Treitel 1980, p.184) suggests the optimal output spike would lie at the first lag for a minimum phase wavelet. This first lag, which also represents the global maximum of varimax, remains to be reached by any of the starting filters (3.14).

It is easy to show that this phenomenon is not unique to the simple example of $nf=3$. Figure 13 shows a minimum phase wavelet to which MED has been applied with three different starting filters. In this case the wavelet is approximately *34* time samples in length and the filter has length $nf=22$. It is observed that for filters of this length, starting filter predetermines output spike lag hence

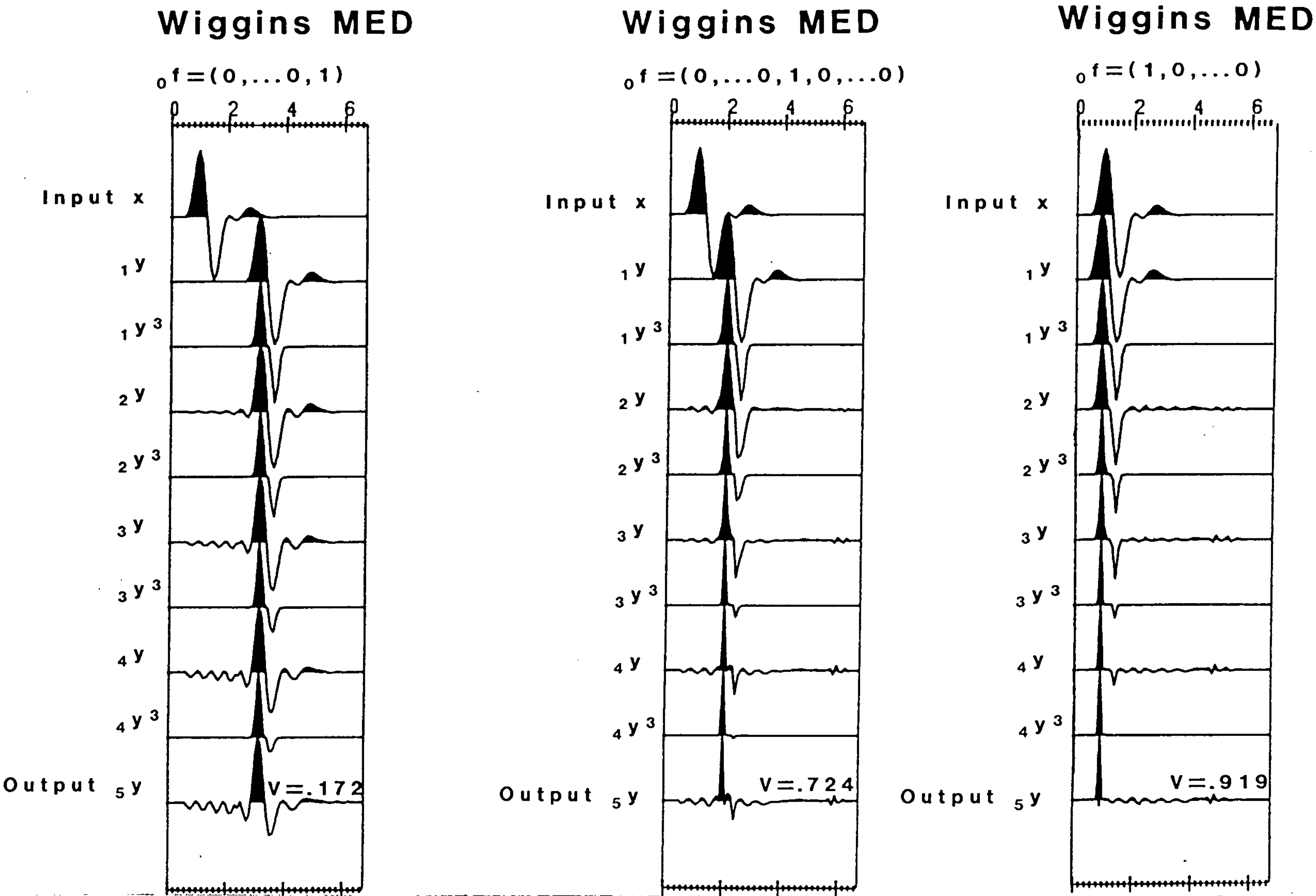


Figure 13.
MED applied to a minimum phase wavelet with three starting filters $_0f$, none of which leads to an output spike at the wavelet onset.

performance of the shaping filter and varimax of the output wavelet. Again none of the filters used are able to spike the wavelet at the first and optimum lag so the global maximum of the varimax remains unreached. It should be noted that further iteration for any of the examples shown in Figure 13 does not result in any significant increase in varimax, and cannot result in a change in output spike placement. Since the increase in varimax with iteration is monotonic (Waldren, 1985), once a local extremum has been reached, there is no possibility for the Wiggins algorithm to skip to another, and hence no possibility the global maximum will be reached.

I propose to overcome this problem by starting the Wiggins MED algorithm not with a starting filter but with a shifted version of the cubed data. By padding the data with zeros and allowing backward as well as forward shifts in the reflection coefficient estimate as a starting point, the algorithm can be designed to spike the wavelet at zero or any other lag. For an input data set **x** containing a wavelet w of length *nw*, the global maximum is reached by testing the varimax for extrema at all $nw+nf-1$ "meaningful" output spike lags of the filtered wavelet, w*f. My algorithm proceeds as follows:

1. Estimate the wavelet length *nw* visually from the data. A generous guess is acceptable.
2. Estimate the wavelet rise time, *L*, that is the number of time samples between the onset of the wavelet and its

peak value. Again, it is okay to overestimate.

3. Pad the data to be filtered, $\mathbf{x}$, with L leading zeros and $(nw-L-1)$ trailing zeros to obtain a new data vector $\mathbf{x}'$ of length $nx'=(nx+nw-1)$.

4. For $i=1$ to $(nw+nf-1)$ start the MED iteration using data $\mathbf{x}'$ and initial reflectivity estimate $\mathbf{d}_i$ with elements d_{ij} where

$$d_{ij} = \begin{cases} 0 & j<i\ ,\ nx+i \le j \le nx'+nf-1 \\ x^3_{j-i+1} & i \le j < nx+i \end{cases} .$$

Stop the iteration upon convergence.

5. Choose as the globally optimum MED operator that one which has achieved the highest varimax over all output lags i.

Example: A seismogram $\mathbf{x}$ is composed of a wavelet.

$$\mathbf{w} = (-.4, 1, .2, -.2)^{\mathsf{T}}$$

convolved with reflectivity

$$\mathbf{r} = (1, 0, 0, 0, .5)^{\mathsf{T}}.$$

That is

$$\mathbf{x} = (-.4, 1, .2, -.2, -.2, .5, .1, -.1)^{\mathsf{T}}.$$

If we take filter length $nf=3$, wavelet length, $nw=4$ and wavelet rise time $L=1$ then, we would form

$$\mathbf{x}' = (0, -.4, 1, .2, -.2, -.2, .5, .1, -.1, 0, 0)^{\mathsf{T}}$$

and start the MED algorithm $(nf+nw-1)$, in this case six, times using desired outputs

$$\mathbf{d}_1 = [(-.4, 1, .2, -.2, -.2, .5, .1, -.1, 0, 0, 0, 0, 0)^3]^T$$
$$\mathbf{d}_2 = [(0, -.4, 1, .2, -.2, -.2, .5, .1, -.1, 0, 0, 0, 0)^3]^T$$
$$\vdots$$
$$\mathbf{d}_6 = [(0, 0, 0, 0, 0, -.4, 1, .2, -.2, -.2, .5, .1, -.1)^3]^T.$$

The "optimum lag" MED output is chosen as that one of the six MED runs which yields the greatest varimax output.

Figure 14 shows the success of this algorithm when applied to an isolated wavelet. The optimum lag MED iteration is shown and achieves both a higher varimax output and visually superior deconvolution than MED using the conventional symmetric starting filter or any of the family of starting filters given by (2.3). It is further noted that this algorithm, when applied to the three point filter example of Figure 12 does indeed select the global extremum as marked.

In closing this section it is noted that this optimum lag algorithm is quite burdensome computationally. However, the optimum lag itself depends on wavelet phase. Provided the phase does not change too drastically from one record to the next we need only perform the full lag scanning on one or two records and then restrict the search to lags in the immediate vicinity of the known optimum. In this way computation cost is held to a few times that of Wiggins MED. Still expensive, but for some practical applications there may well be no satisfactory alternative.

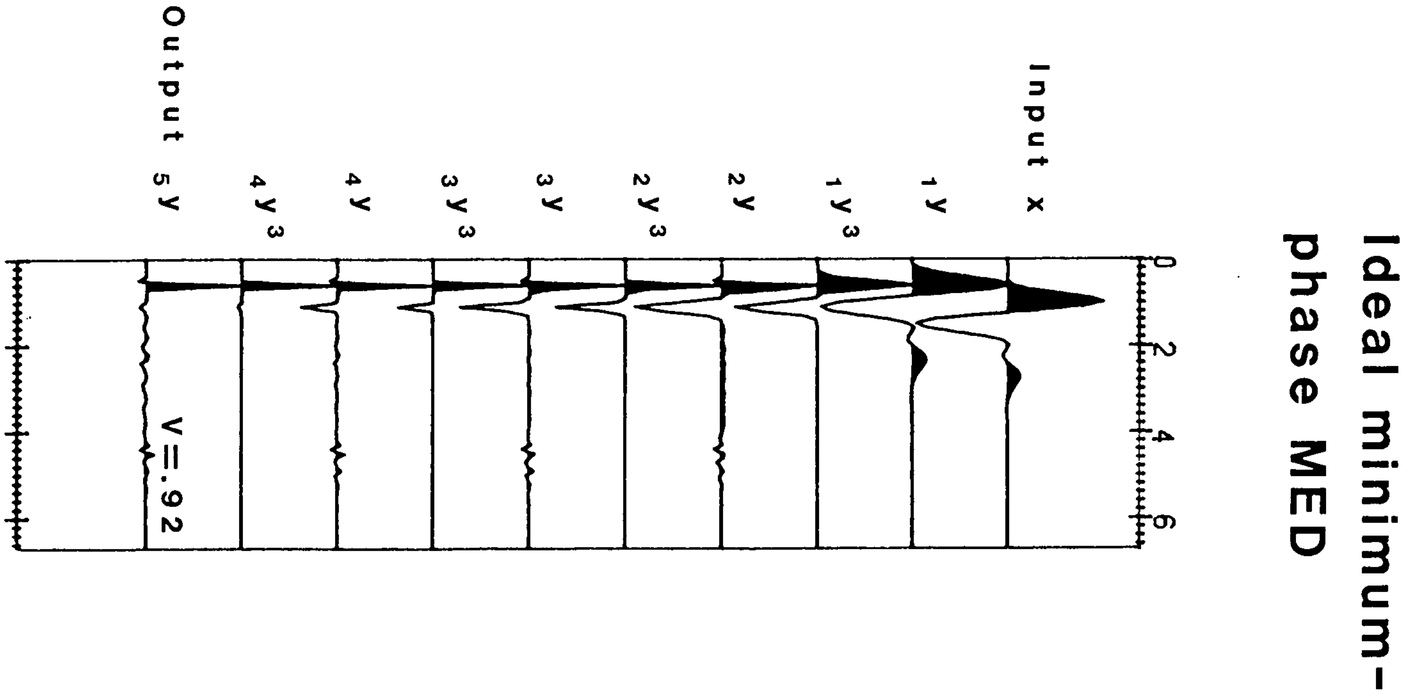


Figure 14.
For a minimum phase wavelet the optimum lag MED algorithm selects that iteration starting point, $_1y$ which leads to a spike output at the output wavelet's zero lag sample.

4.2 BANDLIMITING THE SOLUTION

The digital seismic trace we wish to deconvolve is band-limited and contains noise as well as signal. Is the MED algorithm I have discussed ready to tackle such data? No. It is well understood that least squares filter calculations require the addition of a ridge parameter in the autocorrelation matrix of equation (2.4) to stabilize the solution of the Toeplitz system. This is done by adding a "prewhitening" constant to the diagonal elements – typically less than one percent of the diagonal element itself. We therefore expect the MED algorithm to require a ridge parameter. There is a second problem though, peculiar to MED which must be addressed. DeVries and Berkhout (1984) observe that varimax and related norms increase with bandwidth. This is a problem in the presence of high frequency noise where a MED algorithm called upon to maximize the varimax or similar norm may boost noise at unwanted frequencies to achieve its goal. To overcome this problem I have adopted the band-limiting algorithm described by Deeming (1980) and outlined in Appendix A. By adding to the trace autocorrelation not just a constant at zero lag but a band-limited weighting function, a tradeoff is introduced between maximizing the varimax and surpressing filter energy outside a specified pass band. Examples demonsrating the effectiveness of this technique will be provided in section 6.

4.3 DATA TAPERING AND WINDOW SELECTION

It has been observed in this work, that poor choice of the data segment over which the MED operator is designed can yield disastrous results. Such disasters arise from either time clipped wavelets at the beginning or end of the design window, or the constructive interference of closely spaced reflectors governing filter design.

The first problem arises from the fact that a seismic wavelet, cut short at some point and replaced with zeros is very spiky indeed. Figure 15(b) shows the result of a failed MED run which successfully spiked a clipped wavelet at the edge of the design window shown in Figure 15(a). The unacceptable deconvolution output has a much higher varimax than the correctly deconvolved trace would yield. This problem is not unique to MED, Lazear (1984) reports that such wavelet fragments pose problems for some other deconvolution techniques as well.

Two approaches are recommended to deal with the problem discussed above.

1. Select the time gate carefully without truncating any large events at its edges.
2. Taper the data within the time gate. In this paper I have applied a tapering function due to Colin Walker (personal communication) in which the i'th sample in the design gate is weighted by $B(i)$ defined as

$$B(i) = [4i(m-i)/m^2]^a$$

where $m=ny-1$, and a is selected so that B equals $.5$ at a

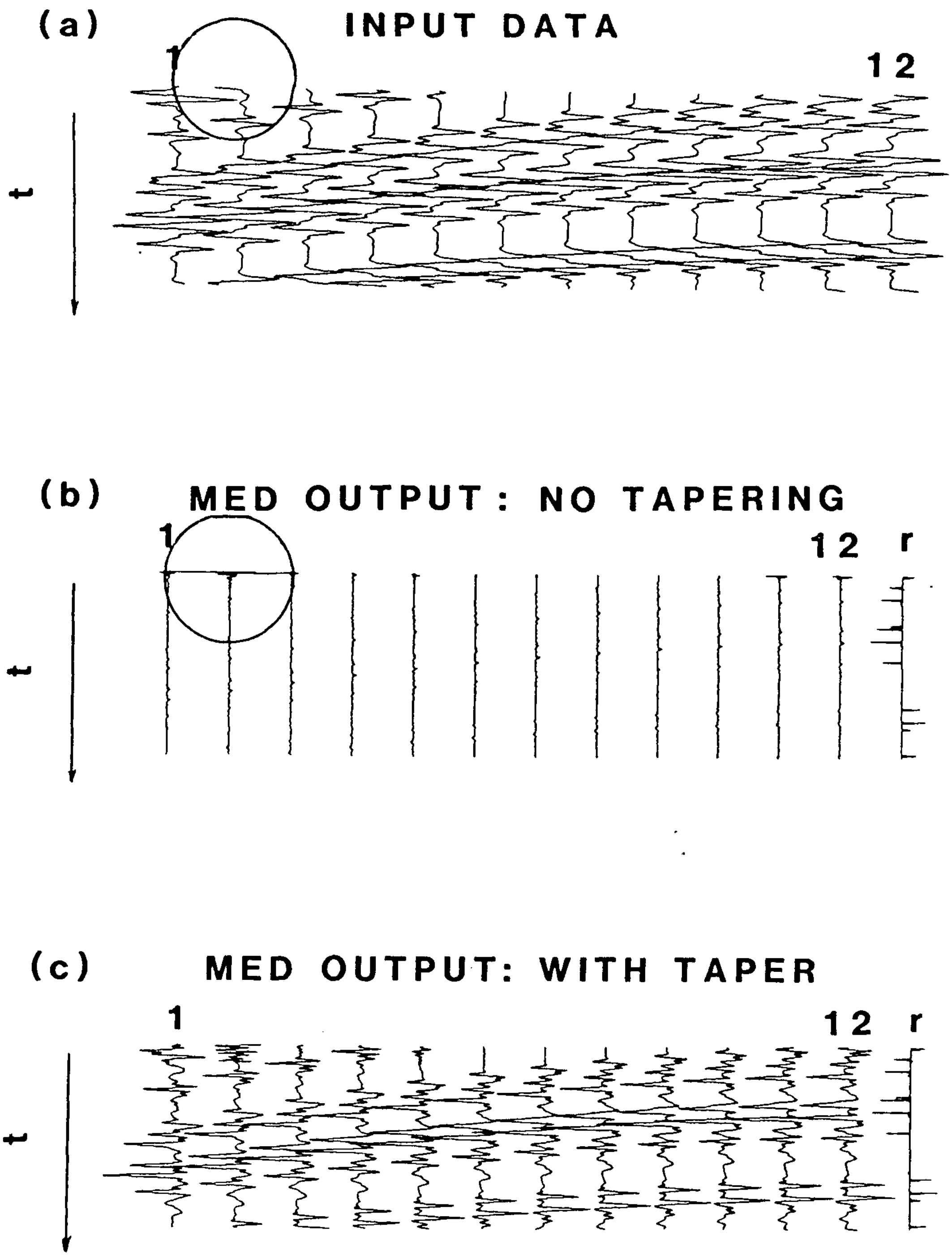


Figure 15.
(a). A multichannel synthetic data set showing normal moveout with truncated reflections at the edge of the time gate (circled). (b). MED output with no data tapering yields a large spike on trace 2 (circled). (c). MED output with data tapered as described. Known reflectivities r are shown at right.

point $nf/2$ samples from the window edge. The effectiveness of this window is shown in Figure 15(c), a repeat run of the data from Figure 15(a) with tapering.

The second problem I discuss in this section is constructive interference. Consider the two reflectivity series shown in Figure 16, convolved with the same wavelet. Series B, yields an excellent MED wavelet inverse while series A does not. Why? Because the closely spaced reflectors in A lead to constructive interference among adjacent wavelets leading MED to spike the combined wavelet. Since MED algorithms base their reflectivity estimates solely on data amplitude, the large amplitude events tend to control the deconvolution. According to Morley (1979) this is particularly true when large amplitude and high frequency are well correlated. To combat this effect we must have a design window containing several large events. Some ways of ensuring this are:

1. Apply MED as a multichannel deconvolution including several traces of a shot record in the filter design. If normal moveout is present the constructive interference for a given event will not remain constant across the record and the algorithm is less likely to nominate a doublet as the true seismic wavelet.
2. Deeming (1984) has suggested applying fast automatic gain control to the data before MED. He attributes the apparent success of this to another argument but it would clearly give smaller, isolated reflectors an equal

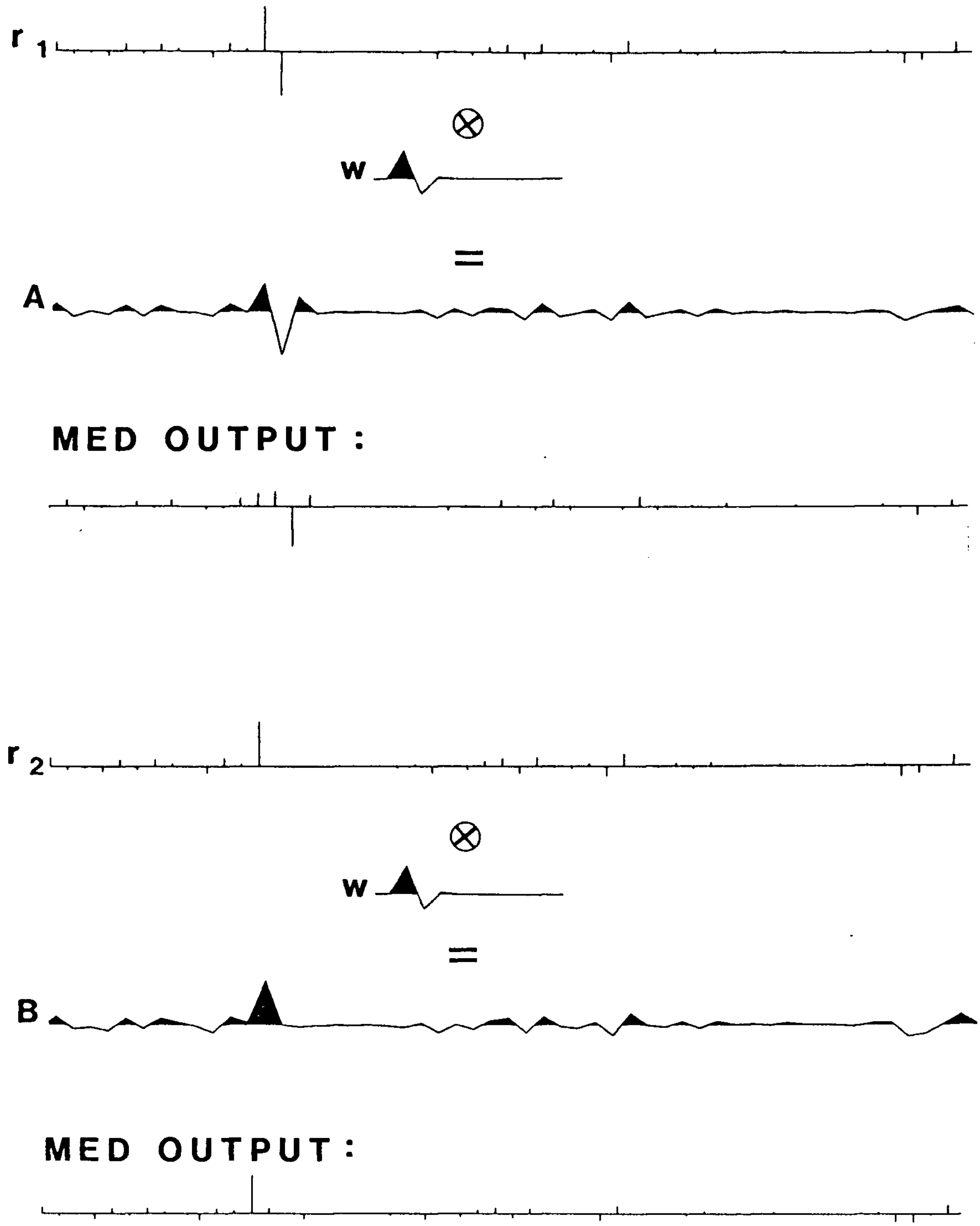


Figure 16.
MED applied to two sparse synthetic traces **A** and **B** derived from reflectivities r_1 and r_2 respectively. **A** yields a poor reflectivity estimate due to constructive interference.

say in controlling the filter design.

3. Deeming (1981) has furthermore suggested applying a 1/varimax trace weighting to the input data. This would give high varimax traces less weight in the MED filter design.

The second point above is undesirable in terms of the convolutional model of the seismogram, and the third in that it discounts what may be valuable information in the high varimax traces. Indeed if the trace varimax is high due to an isolated reflector it is desirable to let it control the filter design. These last two points have however, lead to a more robust algorithm according to Deeming but have not been included in this algorithm. For the present I have applied MED to data which have had spherical divergence correction applied and windows which are carefully selected using the criteria discussed.

4.4 SELECTION OF FILTER LENGTH

Most of the discussion thus far has assumed that the filter length nf is known. In practice, it is up to the user to decide upon a filter length knowing it will likely affect the quality of the deconvolved output. Deeming (1984) in a paper entitled "Why Minimum Entropy Deconvolution Doesn't Work", presents the following argument against the use of very long filters.

Suppose the MED filter length nf were allowed to tend to infinity. In that case the least squares filter output y tends to duplicate exactly its desired output **d**. We would have

$$\mathbf{y} = \mathbf{d}$$

$$= \mathbf{y}^3 \quad \text{for the varimax.}$$

Then $y_i = y_i^3 \qquad i=1,\ldots ny$

and $y_i = 0, 1,$ or $-1 \qquad i=1,\ldots ny.$

A few outputs that might satisfy this criterion are

$$\mathbf{y} = (\ldots,0,0,0,1,0,0,0,\ldots)$$

$$\mathbf{y} = (\ldots,1,0,-1,1,0,0,1,\ldots)$$

$$\text{or } \mathbf{y} = (\ldots,-1,-1,1,-1,1,1,1,\ldots).$$

All of the above deconvolution outputs have two things in common, they are all extrema of the varimax functional, and they are all geologically unreasonable. If we interpret the output **y** as an estimate of the earth reflectivity then all geologic boundaries have the same reflection coefficient, unity. This is not true of any borehole derived reflectivity sequence and is seldom true even of synthetic data! Deeming (1981) suggests that this is a major failing of the MED technique only if very long filters are used. On the other hand, very short filters have larger squared errors ϵ^2 as defined in (3.2). If chosen too short the filter may not adequately represent the wavelet inverse. Ulrych et al. (1978) have suggested choosing a filter length which minimizes a criterion of the form

$$E(L) = \sum_{i=1}^{ns} \{[\mathbf{y}_i(L)-\mathbf{y}_i(L-1)]^T[\mathbf{y}_i(L)-\mathbf{y}_i(L-1)]\}$$

where $\mathbf{y}_i(L)$ is the i'th trace output for a MED filter of length L.

Jurkevics and Wiggins (1984) select filter length by varying nf and choosing that value which sees the varimax of the output trace level off as a function of increasing filter length.

In my examples in section 6 I use filter lengths between 60 and 100 percent of the wavelet length with good results. The cautions of Deeming (1984) whose paper might better be entitled "Why MED Doesn't Work For Long Filters" are valid, and selection of a short filter of a length giving the most subjectively pleasing outcome seems a safe alternative.

As a final comment, the necessity for using the shortest possible filter makes optimum lag scanning essential since, by analogy with optimum lag Weiner filtering, the dependence of filter performance on output lag increases with decreasing filter length.

5. THE APPLICABILITY OF MED

5.1 THE MINIMUM ENTROPY ASSUMPTION

It was noted in chapter 2.4 that besides poor operator design there can be another, fundamentally different, cause for MED failure. This cause is the misapplication of MED to data for which it is inherently unsuited. Recall that a restrictive assumption was adopted to allow us to estimate a reflectivity from the seismic trace, a highly underdetermined problem. We have assumed that the desired reflectivity estimate coincides with the global extremum of the varimax or another entropy measure as a functional of **f**. The reflectivity estimate **y**, our seismic deconvolution output, is valid only if that asssumption is true. Like the minimum phase assumption of spiking deconvolution the minimum entropy assumption is true for some data and not for others. However, unlike the spiking deconvolution case we cannot express, or at least have not yet expressed, the minimum entropy assumption in terms of measurable properties of the seismic trace components.

No satisfactory criterion for MED applicability has yet been proposed. One criterion which I have attempted to apply has failed and I have opted to try MED on seismograms formed from random reflectivity series to gain insight into just what "minimum entropy" means.

The parameter initially tested was $V(\mathbf{y})$, the output trace varimax, which would increase with reflectivity sparseness perhaps yielding a satisfactory deconvolution above a certain threshold value. It was found that this parameter does indeed increase with the sparseness as expected and reflects the quality of deconvolved output in many cases. However, the presence of closely spaced, strong reflectors such as the ones in Figure 16 can yield outputs which are both high in varimax and physically meaningless. This and the frequency dependent nature of the varimax make it unworkable as a threshold criterion.

Our goal in minimum entropy deconvolution is to design a filter from the seismic trace which is itself the MED wavelet inverse. By MED wavelet inverse I mean that filter which would best maximize the varimax of an isolated wavlet. In all cases studied here this has roughly, but not exactly, coincided with the optimum lag Weiner spiking filter for that wavelet. If our data do not satisfy the minimum entropy assumption, then we will be moving further away from the MED wavelet inverse $\mathbf{f}_w$ while searching for a filter $\mathbf{f}_x$ which maximizes the trace varimax. The filters $\mathbf{f}_w$ and $\mathbf{f}_x$, that is, the maxima of the varimax functionals for both seismogram and lone wavelet, will coincide only when the minimum entropy assumption is satisfied by the data. Consider the familiar case with filter length nf=3. The contour diagram mapping the varimax surface for a simple wavelet **w** is shown in Figure 17(a). Below it are varimax contour maps constructed using the same wavelet convolved with (b) some Gaussian i.i.d. noise and (c) a sparse spike series. The

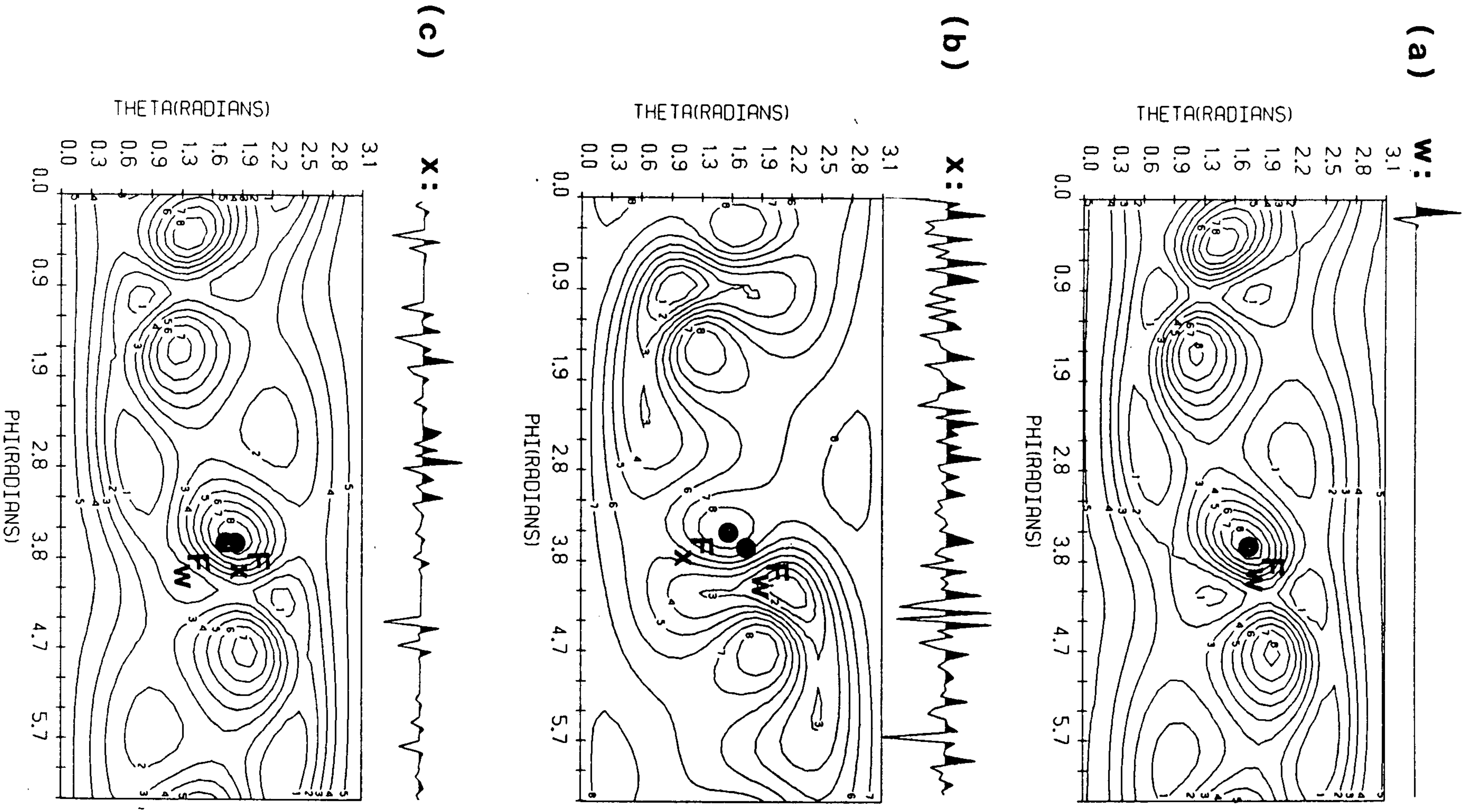


Figure 17.
Varimax of all three point filters contoured for the data sets **x** shown. (a) wavelet alone; (b) gaussian seismogram; (c) gaussian amplitudes with poisson (λ=.3) time distribution. f_w is the MED wavelet inverse from (a), f_x is the MED filter designed from each data set.

similarity between the maps for lone wavelet and sparse seismogram marks MED's applicability in this example. To maximize the varimax of the seismic trace **x** is to find a filter which also maximizes the varimax of the seismic wavelet. The varimax surface portrayed for the Gaussian seismogram (b) however, is markedly different and its global extremum does not coincide with an acceptable wavelet inverse as marked by the peaks on the wavelet contour. Hence the filter output is a poor estimate of the reflectivity. In this latter case MED is therefore not applicable.

For the case $nf=3$ then, one can test MED applicability for a variety of synthetic data sets by comparing the contoured varimax surfaces for wavelet and seismogram. Furthermore we are not restricted to the varimax norm in this case and I have prepared contour diagrams of the varimax with and without exponential transform, parsimony, D-norm and $V(\mathrm{y},2,1)$ for comparison. The data sets used were synthetic seismograms formed by convolving a simple three point wavelet with realizations of the following random processes.

1. Uniform i.i.d.
2. Gaussian i.i.d.
3. Two sided exponential i.i.d.
4. Gaussian amplitudes Poisson distributed in time with a Poisson parameter $\lambda=.3$
5. Gaussian amplitudes Poisson distributed in time with a Poisson parameter $\lambda=.7$.

Realizations of these random processes were 250 points in length and some examples of these are shown in Figures 18 to 20.

Figure 18 shows the D-norm surfaces for known wavelet, Gaussian seismogram and sparse seismogram ($\lambda=.3$).Note that the character of the surface is quite different from that of the varimax norm in Figure 17 indicating that the norms are not equivalent. Furthermore neither of the global maxima $\mathbf{f}_x$ coincide with the wavelet inverse $\mathbf{f}_w$ and D-norm MED is not applicable in either case. Figure 19 shows the behaviour of the parsimony for the same seismograms. Note that the parsimony surface is very similar in shape to the varimax, being inverted with peaks replacing valleys and valleys, peaks. This is to be expected since the parsimony is *minimized* and the rough similarity of parsimony and varimax surfaces was observed with all the data sets tested.

A rather incredible result is shown in Figure 20(b). The varimax surface for a seismogram constructed from a uniform reflectivity sequence is inverted with respect to the surface for wavelet alone, Figure 20(a). This result is predicted by Donohoe (1981), as mentioned in a previous chapter. Indeed if our algorithm were modified to *minimize* the varimax norm then MED would satisfactorily deconvolve this decidedly non-sparse uniform seismogram. Figure 20(c) shows the varimax behaviour for an exponentially distributed seismogram, again filters $\mathbf{f}_x$ and $\mathbf{f}_w$ fail to coincide. The contour diagrams did not suggest that any norm is

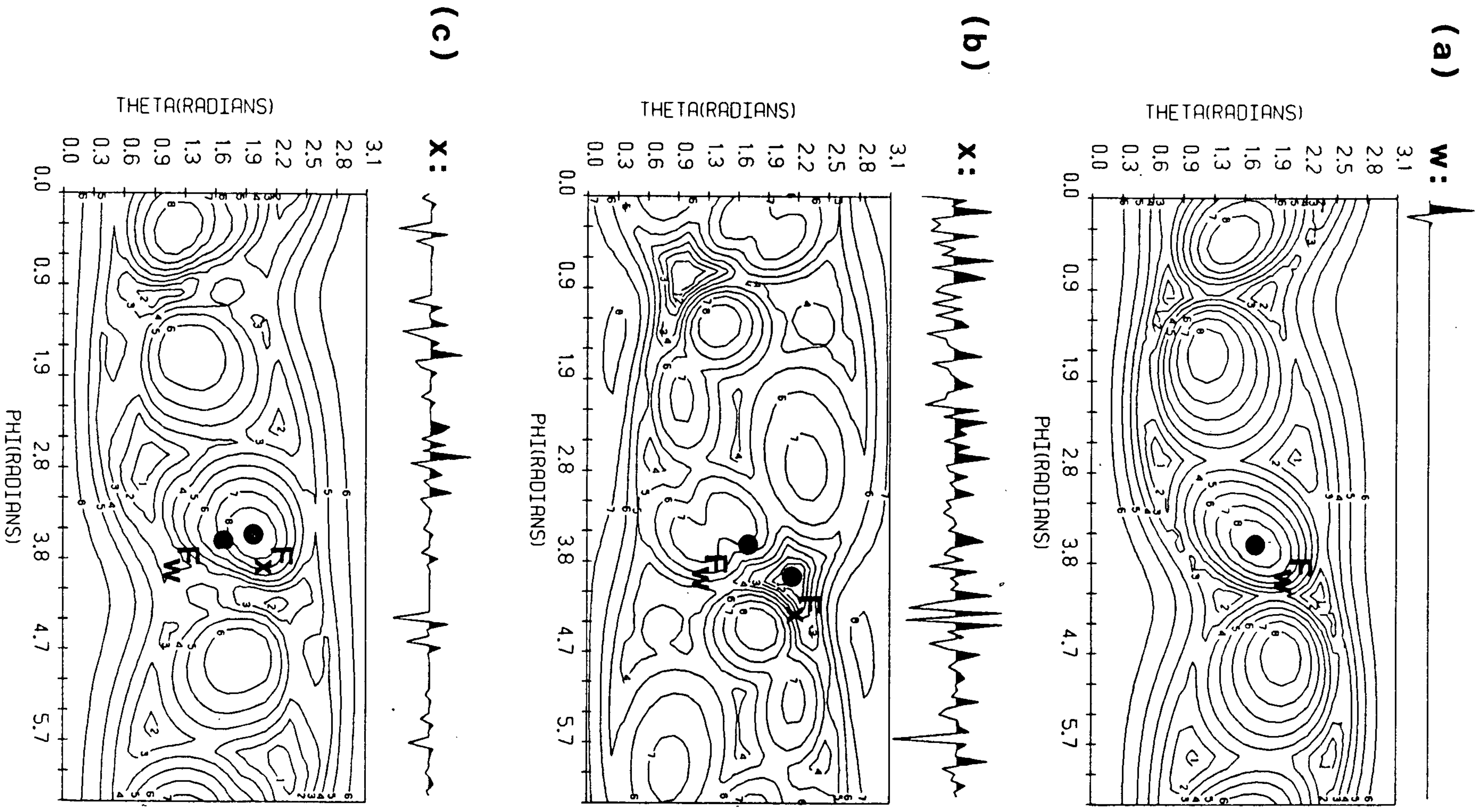


Figure 18.
Contours on the Cabrelli D-norm for the data sets of Figure 17. The distinctly different character of the surface from that of the varimax indicates the non-equivalence of the norms.

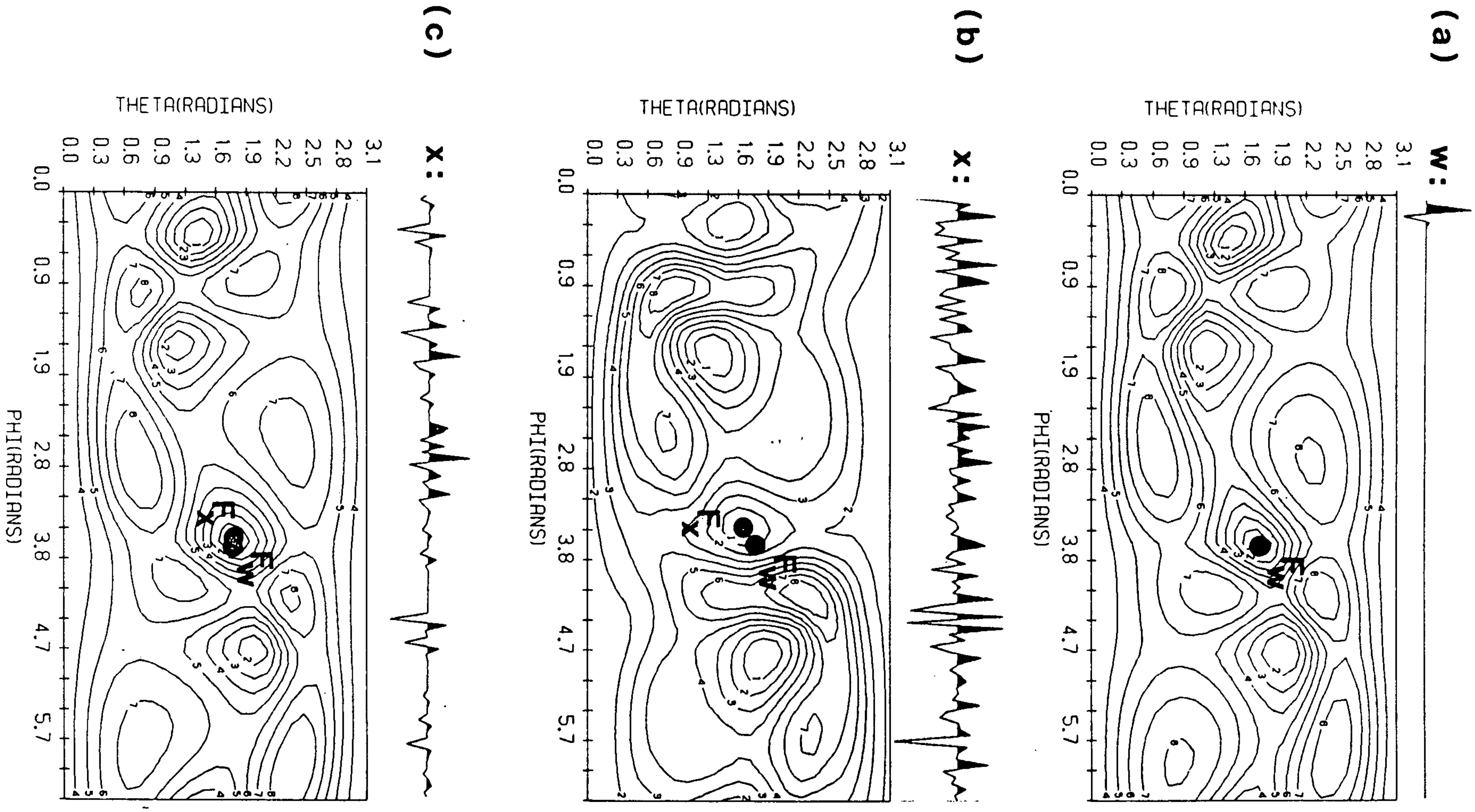


Figure 19.
Contours on the parsimony S_2 for the data sets of Figure 17.

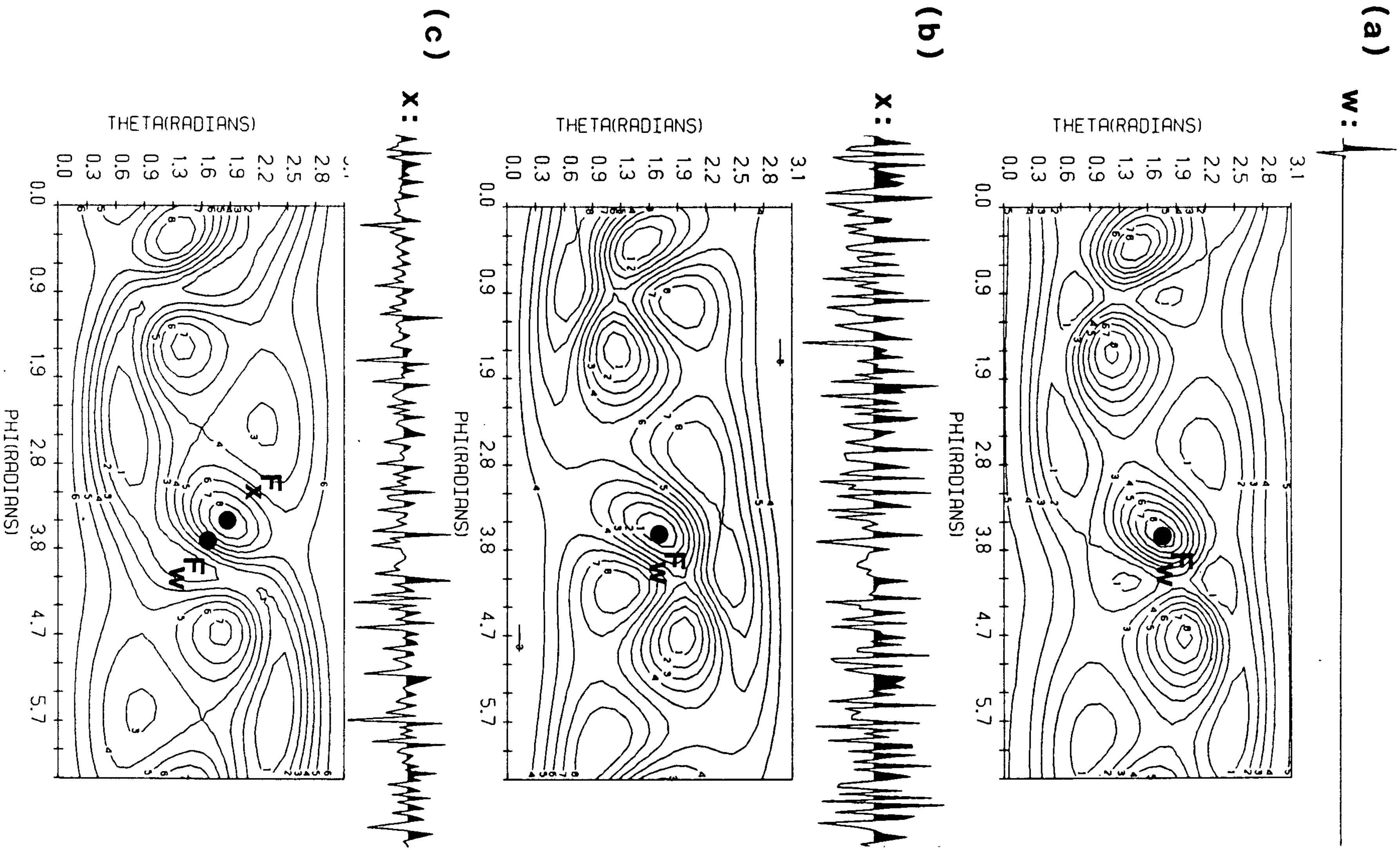


Figure 20.
Contours on the varimax for (a) lone wavelet, (b) uniformly distributed seismogram and (c) exponentially distributed seismogram. Note that (b) is an inverted version of (a).

universally more applicable than any other.

In the course of studing the varimax surfaces for the 3 sample filter case it became apparent that the shape of the surface and therefore the applicability of the method depends on the example chosen. In some cases the contour diagram would change with data set length or the particular realization of random noise used to generate the reflectivity.

It was therefore decided to investigate MED behaviour over many realizations of the random processes and furthermore to do this using more realistic wavelet and filter lengths. Figure 21 shows one realization of each of the reflectivities, which are the same as those used before except Poisson parameters $\lambda=.05$, $.1$ and $.25$ have been used to account for the longer wavelet. The seismograms created by convolving these reflectivities with a 38 sample minimum phase wavelet, and the deconvolved traces for optimum lag MED with $nf=22$ are also shown. After running ten realizations of each reflectivity sequence It became evident that MED applicability bears no simple relation to reflectivity pdf for samples of this size. In this case, since the varimax surface cannot be visualized for $nf=22$ I have judged the applicability by inspection of the residual wavelet $f_x * w$. MED worked most often on the data containing the sparsest reflectivity and least often for the Gaussian. However, for none of the reflectivity series did it either succeed or fail consistently. Furthermore, in an attempt to

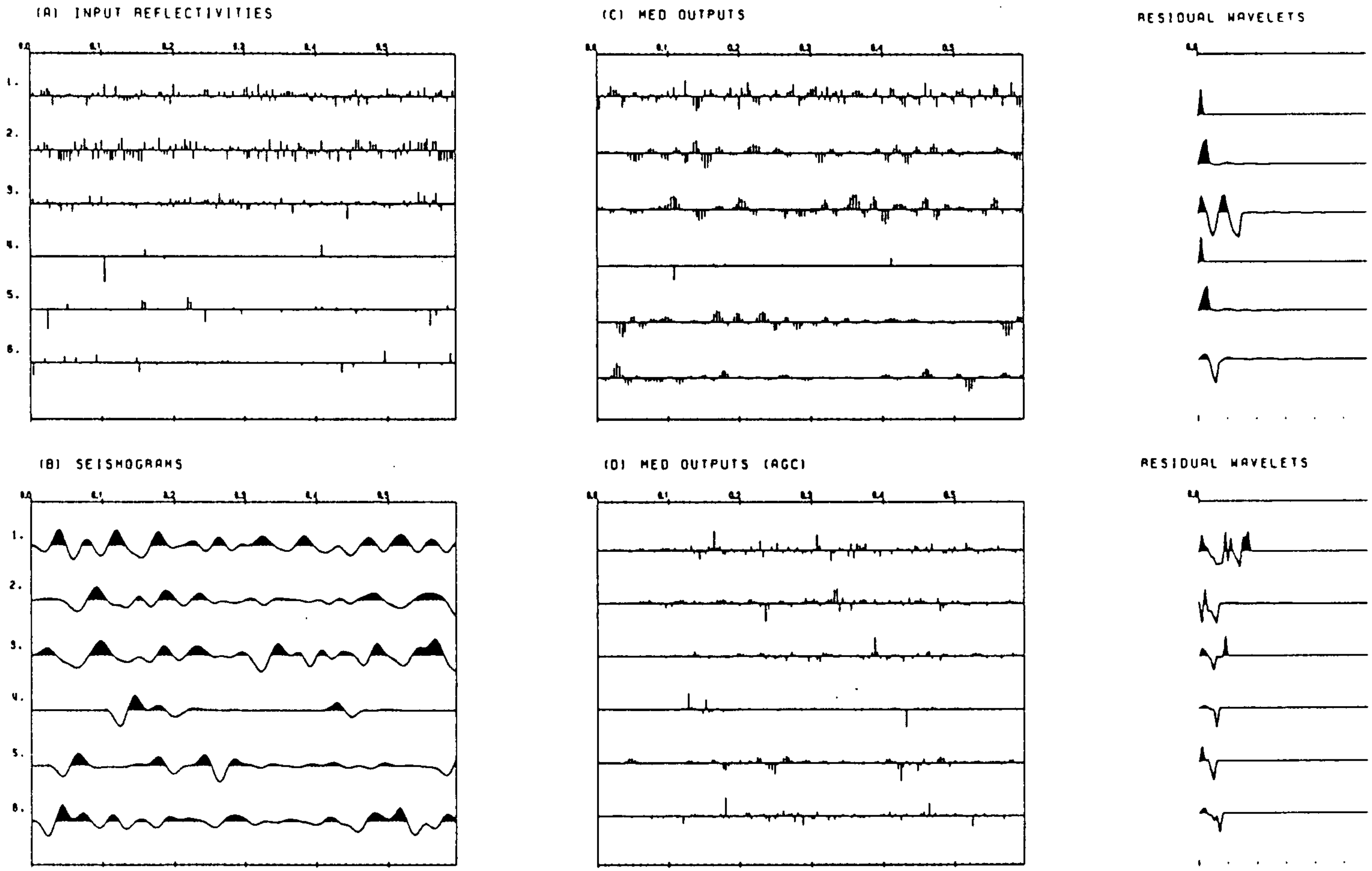


Figure 21.
The effect of reflectivity probability distribution is shown with (a) segments of 1. gaussian, 2. uniform, 3. exponential, 4. gaussian/poisson (λ=.05), 5. gaussian/poisson (λ=.10), and 6. gaussian/poisson (λ=.25) reflectivities. (b) synthetic seismograms and MED outputs with and without AGC are shown in (c) and (d). Residual wavelets $f_x * w$ from (c) and (d) show no improvement with AGC.

quantify the applicability of MED to a certain data set it was found that none of the following parameters proved a dependable guide to MED performance:

1. Input seismogram varimax
2. Deconvolved seismogram varimax
3. Residual wavelet varimax
4. The quantity $\mathbf{f}_x \cdot \mathbf{f}_w$ which is unity when $\mathbf{f}_x = \mathbf{f}_w$ and is a measure of the distance beween $\mathbf{f}_x$ and $\mathbf{f}_w$ on the contour diagrams of the $nf=3$ case.

All of the above were indicators of good MED performance for some data and not for others. AGC was applied to one data set with a variety of gate widths in hopes that it might have a positive effect on the deconvolution as Deeming (1984) suggests. Figure 21(d) shows an example of one trace from each of the various pdfs deconvolved with an AGC gate width of twice the wavelet length. AGC has provided no apparent improvement in this example. Deeming's positive result may have been due to the AGC lending stationarity to his data which my synthetic traces already exhibit.

One conclusion that can be drawn from this work is that MED applicability apparently does *not* depend solely on the statistical properties of the underlying reflectivity function. Given seismograms constructed from two realizations of the same random process one may well find that MED works on one and not the other. This is due to wavelets interfering in such a way as to shift the global varimax extremum away from the true reflectivity regardless of the sparseness, simplicity or non-Gaussian nature of that reflectivity.

The minimum entropy assumption, that the true reflectiviy lies at the extremum of an entropy measure such as the varimax is not just a restriction on the refectivity statistics or even the reflectivity model. It is also a restriction on the seismic wavelet through the way in which the two interact constructively to give very large events. Furthermore, as evidenced by Figure 21, in which MED works on some white noise reflectivities it can be added that the concept of "sparceness" or even "simplicity" does not adequately express the minimum entropy assumption.

6. EXAMPLES USING SYNTHETIC DATA

6.1 INTRODUCTION

In this section I shall test the optimum lag MED routine described in the previous sections and compare results with Wiggins MED using symmetric starting filter. The data used in these tests is a multichannel synthetic data set kindly supplied by Andy Jurkevics and were in fact used in a recent paper for comparison of deconvolution techniques (Jurkevics and Wiggins, 1984).

The data consist of a reflectiviy function whose events have amplitudes drawn from a Gaussian probability distribution and occur Poisson distributed in time with a Poisson parameter $\lambda=.1$. A twelve trace data set has been generated from the reflectivity with simulated normal moveout and is convolved with a minimum phase wavelet. In the noisy examples random Gaussian noise has been generated, bandlimited to less than 60 Hz and added to the traces. In all cases a filter length of nf=22 was used. In filter length tests on noise-free data this was judged the shortest operator yielding satisfactory wavelet estimates.

I had originally planned to include the Cabrelli algorithm in these tests and an algorithm was written for that purpose. As published however, Cabrelli MED requires (ns x nx) times the computational effort of spiking deconvolution. Computing cost was prohibitive for anything but simple single trace examples, and unless significant optimization of that algorithm can be achieved its use as an alternative to iterative MED is in doubt.

6.2 THE EFFECT OF OPTIMUM LAG SCANNING

An example of optimum lag MED versus Wiggins MED with symmetric starting filter is shown in Figure 22. Figure 22(a) shows the data set used for the test, a noise-free multichannel synthetic shot record. The minimum phase seismic wavelet approximately 34 samples long is displayed below the record. MED outputs for nf=22 with and without lag scanning for the global extremum are shown in 22(b) and 22(c) respectively. Wavelet estimates from each MED filter are plotted below with optimum lag MED yeilding a superior result although it does contain a small precursor. No significant improvement in the wavelet estimate is obtained by increasing the filter length, at least up to NF=55 the longest filter tested. In the example shown, the optimum lag MED output is normal in polarity while the polarity of the example on the right is reversed. This is not the case in general however , since even the optimum lag algorithm can come to rest in either of the global extrema corresponding to the two polarities. Final varimax of the optimum lag output is significantly higher than that of the conventional MED algorithm.

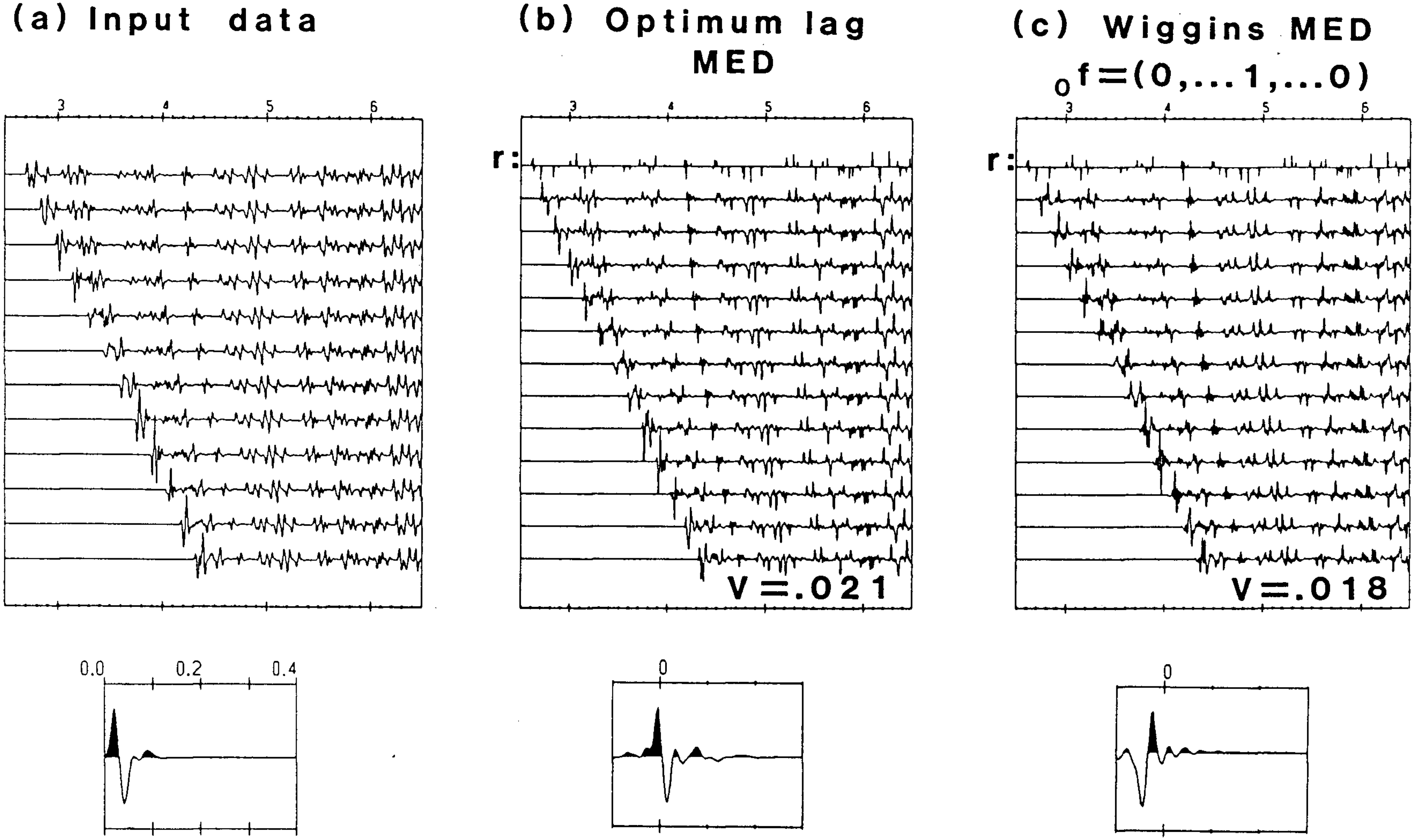


Figure 22.
The effect of optimum lag scanning MED is demonstrated using (left) synthetic multichannel data set containing the wavelet shown, (centre) optimum lag MED output with reflectivity plotted at top and extracted wavelet below and (right) corresponding output for MED with symmetric starting filter.

6.3 THE EFFECT OF ADDITIVE NOISE

The effect of noise on the bandlimited optimum lag MED algorithm is demonstrated below ,again using the data sets from the previous section. To these data I have added 5,10 and 20 percent Gaussian noise bandlimited between 0 and 60 Hz. The noisy data sets and the seismic wavelet they contain are shown in Figure 23. To these data I have applied optimum lag MED with .01 percent prewhitening. The resulting outputs along with wavelets estimated from the MED filters are shown in Figure 24 with the reflectivity plotted at the top of each section. It can be seen that for both the 5 and 10 percent noise cases good results are achieved and a reasonable wavelet estimate, although the precursor is still present and the estimate seems to be degrading with increasing noise. For the 20 percent noise case however, MED has opted to place a spike at the beginning of the data set despite the data tapering applied as discussed in section 4.3. The filter, rich in high frequencies does little to attenuate noise in the data and furthermore yields a poor wavelet estimate.

Next I applied optimum lag MED to the same data set but replaced the prewhitening or ridge parameter with the bandlimiting algorithm described in Appendix A. Output seismograms are displayed in Figure 25. The parameters used were:

1. Passband 0-50 Hz
2. λ=.025 and

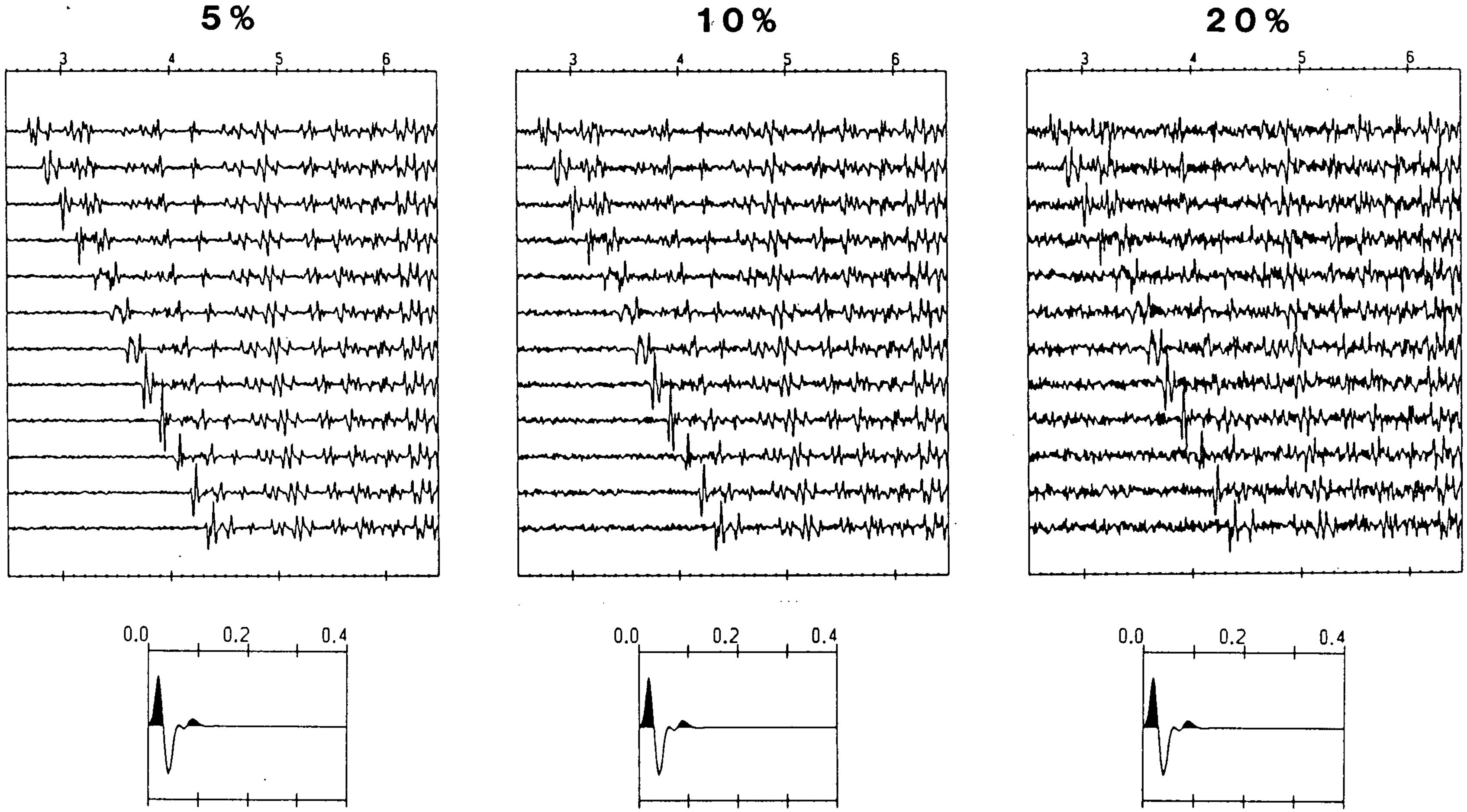


Figure 23.
Synthetic data sets and wavelets used in the noise examples.

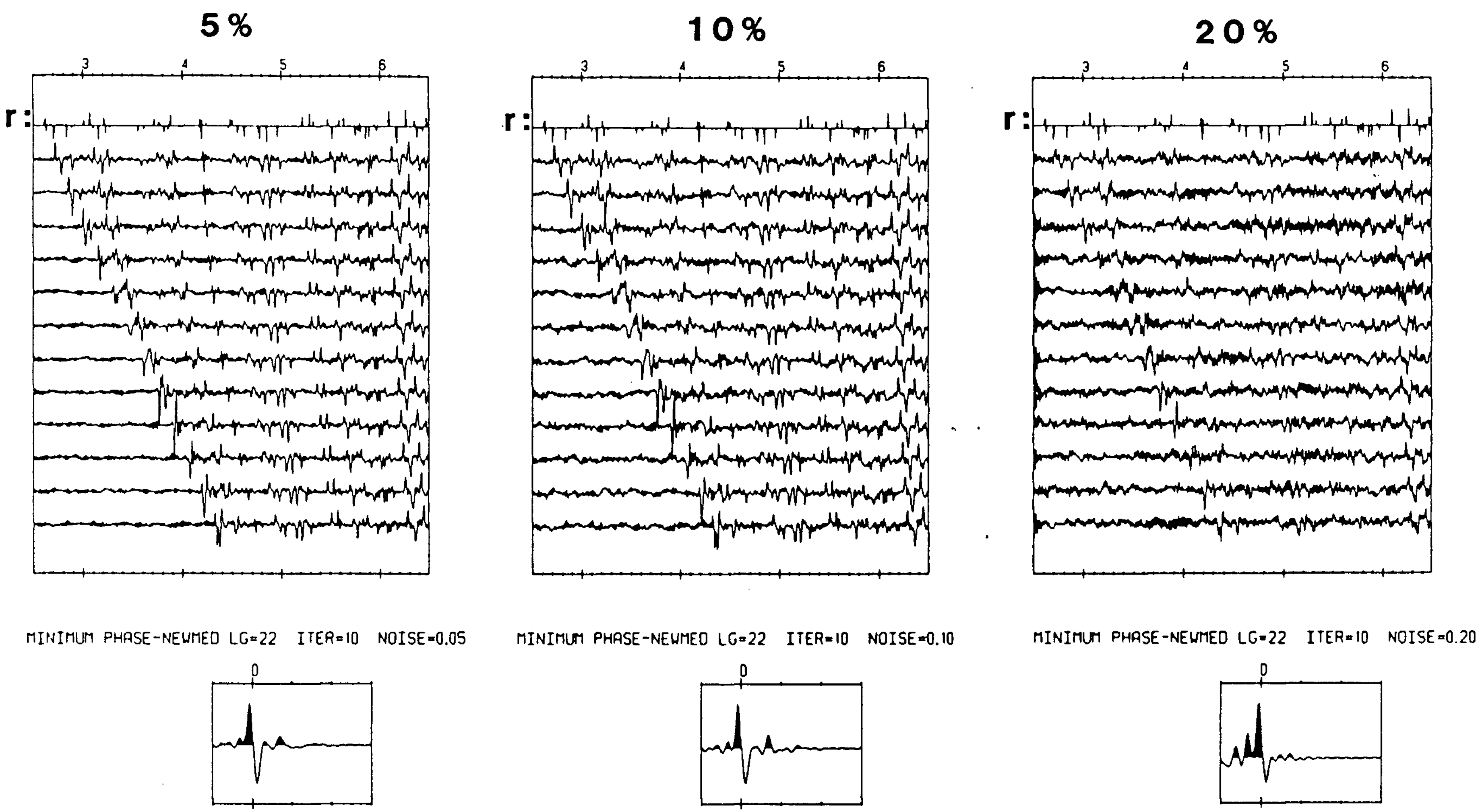


Figure 24.
Optimum lag MED outputs and extracted wavelets for the data sets of Figure 23. with .01% prewhitening.

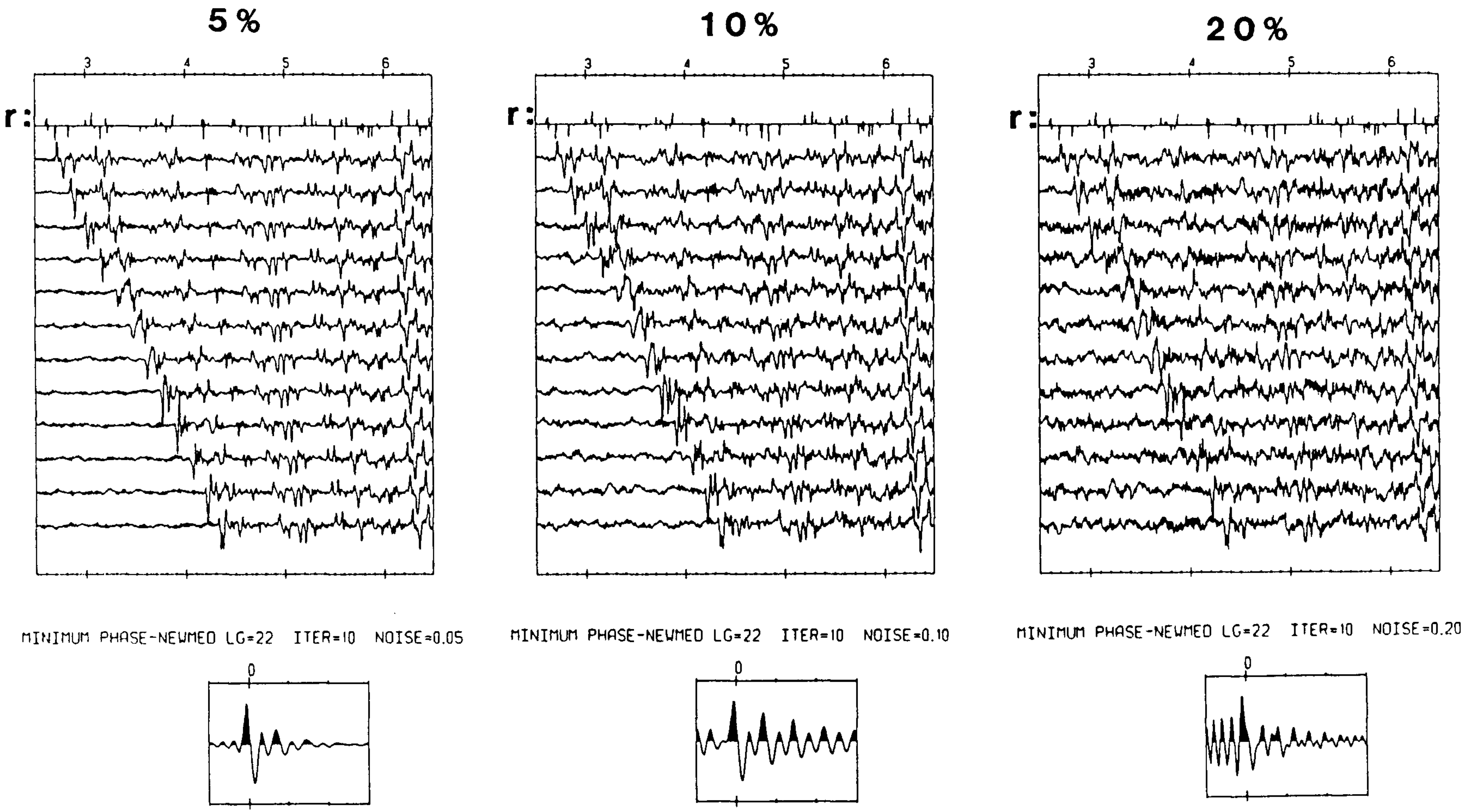


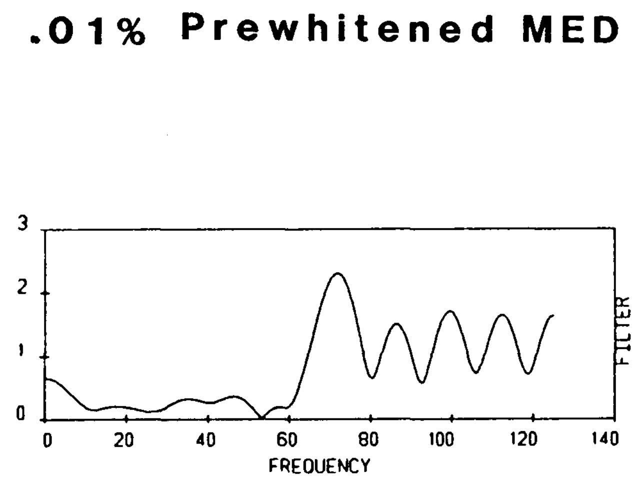
Figure 25.
optimum lag MED outputs and extracted wavelets for the data sets of Figure 23. using the bandlimiting algorithm described in Appendix A.

3. Q=.001.

The deconvolved outputs for the 5 and 10 percent added noise cases are comparable to the previous figure, however the wavelet estimates are not useable. The high frequency noise level in these output traces is lower than in the previous example and the 20% noise example is decidedly superior to that of the previous figure although seriously degraded compared to the 10% case.

Why has the bandlimiting algorithm apparently failed? First it should be noted that the wavelet estimate here is not a good indicator of deconvolution performance. Simply estimated by frequency domain inversion of the filter in the estimated passband of the wavelet, it assumes a delta-type output wavelet. This is inapropriate since I have sought a bandlimited output.

The bandlimiting algorithm has achieved its goal of surpressing filter energy outside the specified frequency band. This is demonstrated in Figure 26. Here power spectra of the MED filters for the 10 percent noise case are plotted, (a): with .01 percent whitening and (b): with bandlimiting as described. As expected, the bandlimiting has done its job of eliminating frequencies above 50 Hz and would therefore reduce the noise of the output trace. However, the shape of the spectra within the passband differs. This is a result of the bandlimiting algorithm minimizing the *sum* of the squared error and the filter energy outside the passband. This in effect creates a



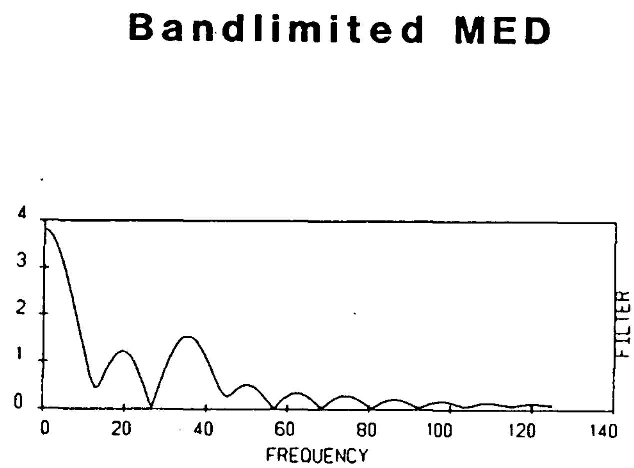


Figure 26.
Power spectra of MED filters designed from the same data set with and without bandlimiting. Shown are the filter spectra from the 10% noise examples of figures 23 and 24.

tradeoff between maximizing the varimax and bandlimiting the filter and in some examples caused the MED algorithm to diverge. If such an algorithm were to simultaneously maximize varimax while minimizing the unwanted filter frequencies we would have a different, and preferable result. As formulated, the bandlimiting scheme of Deeming (1981), when applied to MED is quite dependent on the amount of weighting λ given the high frequency filter energy and leaves the success of any MED algorithm dependent upon λ.

7. DISCUSSION

Minimum entropy deconvolution has the potential to answer some of the perennially unsolved problems in deconvolution such as how to deal effectively with nonminimum phase wavelets and non-white reflectivities. MED is also known to break down in practical application.

MED is not robust and this property is often considered to be inherent in the method. However, in this work I have stressed that there are two distinct types of problem that may be responsible for this failing. The first problem is method of MED operator calculation; the second is judging when to apply MED. These two problems must be kept separate in the minds of the statistical deconvolution user, for no technique, no matter how robust, will yield the right outcome if applied in a situation for which its assumptions do not hold.

Where then do the assumptions of MED hold? For which applications was MED designed? These disarmingly simple questions have yet to be answered, and until they are we can neither set a criterion for when to apply MED nor declare that it is the wrong model for our seismic trace. My examples suggest that a sparse, or even non-Gaussian reflectivity does not adequately describe the seismogram that MED requires. We must further invoke restrictions involving the wavelet and its constructive interaction with the reflectivity function.

Some progress has been made in the area of MED operator calculation. It is now understood that filter lengths should be kept as short as possible, that care must be taken to avoid end effects and that varying the trace weighting and method of summation can improve results. Bandlimiting is clearly a necessary component of a MED filter design but it is not yet clear if the algorithm proposed by Deeming (1981) is the best way to implement this. I have dealt, through a heuristic geometrical approach, with the problem of multiple extrema. This work suggests that the choice of starting filter affects the output spike lag and is therefore tantamount to making a wavelet phase assumption. This can be overcome with the optimum lag MED routine proposed in section 4.1. Terry Deeming (personal communication) makes use of sideways recursion in solution of the MED normal equations in an alternative attempt to achieve the global extremum. Comparison of his technique, which is unpublished, with optimum lag MED would be informative.

The question finally arises whether MED is yet a practical alternative to the deconvolution techniques in use today. The answer seems to be, "Only for the data to which MED is particularly suited." Unfortunately, to this date, we can only determine to which data MED is suited by trying it on that data, hardly an acceptable test given the observation that MED is apparently not suited to most data. The key then is in gaining an understanding of the minimum entropy assumption. Then, with a criterion for application and the advances in operator design that will only come from isolating the computational problems, MED may reach its potential.

REFERENCES

Chacin, O.E., 1985, Wavelet estimation in normal incidence seismograms: Ph.D. Thesis, University of Tulsa.

Claerbout, J., 1980 ,Seven essays on minimum entropy: Stanford Exploration Project Report ,**24**,157-187.

Claerbout, J., 1977a, Parsimonious deconvolution: Stanford Exploration Project Report, **15**,1-9.

Claerbout, J., 1977b ,Examples of parsimonious deconvolution: Stanford Exploration Project Report, **15**,10-32.

Deeming, T.J., 1981, Deconvolution and reflection coefficient estimation, using a generalized minimum entropy principle: 51st Annual International SEG Meeting, October 13, 1981, Los Angeles.

Deeming, T.J., 1984, Why minimum entropy decon doesn't work: SEG Decon Workshop, July, 1984, Vail Colorado.

De Voogd, N., 1974, Wavelet shaping and noise reduction: Geophys. Prosp., **22**,354-369.

DeVries, D. and A.J.Berkhout, 1984, Velocity analysis based on minimum entropy: Geophysics, **49**,2132-2142.

Donohoe, D., 1981, On minimum entropy deconvolution, *in* Findley, D., Ed., Applied Time Series Analysis II, Academic Press.

Ferguson, G.A., 1954, The concept of parsimony in factor analysis: Psychometrika, **19**,281-290.

Godfrey, R.J., 1979, A stochastic model for seismogram analysis: Ph.D. Thesis, Stanford University.

Godfrey, R.J., and F.Rocca, 1981, Zero memory non-linear deconvolution: Geophysical Prospecting, **29**,189-228.

Gonzalez-Serrano, A., 1980, Minimum-entropy decision analysis: Stanford Exploration Project Report, **24**,199-226.

Gray, W.C., 1979, Variable norm deconvolution: Ph.D. Thesis, Stanford University.

Hogg, Robert V., 1972, More light on the kurtosis and related statistics: J. of the Am. Statistical Assoc. **67**,422-424.

Hosken, J.W.J., 1980, A stochastic model of seismic reflections: 50th Annual International SEG Meeting, November 16-20, 1980, Houston.

Jurkevics, A. and R.Wiggins, 1984, A critique of seismic deconvolution methods: Geophysics, **49**,2109-2116.

Kaiser, H.F., 1958, The varimax critereon for analytic rotation in factor analysis: Psychometrika, **23**,187-200.

Lazaer, G.D., 1984, An examination of the exponential decay method of mixed phase wavelet estimation: Geophysics, **49**,2094-2099.

Morley, L., 1980, Minimum entropy deconvolution with the extrinsic power norm: Stanford Exploration Project Report, **24**, 189-198.

Ooe, M. and T.J.Ulrych, 1979, Minimum entropy deconvolution with an exponential transformation: Geophysical Prospecting, **27** ,458-473.

Postic, A., J.Fourmann and J. Claerbout 1980, Parsimonious deconvolution: 50th Annual International SEG Meeting, November 16-20, 1980, Houston.

Robinson, E.A. and S.Treitel, 1980, Geophysical Signal Analysis: Prentice-Hall Inc..

Robinson, E.A. 1954, Predictive decomposition of time series with application to seismic exploration: Ph.D. disertation, M.I.T.: Reprinted in Geophysics, **32**,418-484.

Schoenberger, M., 1985, Special report on the seismic deconvolution workshop, sponsored by the SEG Research Committee, July 18-19, 1984, Vail Colorado: Geophysics, **50**,715-735.

Ulrych, T.J., C.J.Walker and A.J.Jurkevics, 1978, A filter length criterion for minimum entropy deconvolution: CSEG Journal , **14**,21-28.

Ulrych, T.J., 1982, Analytic minimum entropy deconvolution: Geophysics, **47**,1295-1302.

Walden, A.T., 1895, Non-gaussian reflectivity, entropy and deconvolution: Geophysics, **50**,2862-2888.

Wiggins, Ralphe A., 1978, Minimum entropy deconvolution: Geoexploration, **16**,21-35.

Ziolkowski, Anton, 1984, Deconvolution: IHRDC Press.

APPENDIX A: BANDLIMITING THE FILTER OPERATION

In this appendix I discuss the theory and practical application of pass-band constraints on the deconvolution operator. This scheme is based on an as yet unpublished discussion by Deeming (1981) and can be applied instead of a whitening parameter in any weiner filter design process. The commonly used technique of adding a ridge parameter, or white noise to stabilize the filter design process is shown to be a special case of this technique.

The well known Weiner least squares filter design process finds the inverse filter **f** for a wavelet **w** which minimizes the squared error ϵ^2 between the filter output and a desired output **d**. This squared error is defined as

$$\epsilon^2 = \sum_{i=1}^{nd} (d_i - \sum_{j=1}^{nf} f_j w_{i-j})^2 . \tag{A1}$$

Minimizing ϵ^2 with respect to the filter coefficients f_k, $k=1,...nf$ leads to the normal equations

$$\sum_k f_k \sum_i w_{i-j} w_{i-k} = \sum_i w_{i-j} \qquad j=1,...nf. \tag{A2}$$

These equations are solved for the filter coefficients using Levinson recursion. In the practical case however, when **w** may be noisy and band limited the autocorrelation matrix of the wavelet may be illconditioned. In this case we solve the stabilized normal equations:

$$\sum_k f_k (\sum_i w_{i-j} w_{i-k} - \lambda \delta_{jk}) = \sum_{i=1}^{nd} d_i w_{i-j} .$$

Here the addition of a "ridge parameter" ϵ to the zero lag autocorrelation stabilizes the recursion and is equivalent to "prewhitening" the wavelet by the addition of a small amount of white noise to ensure that its spectrum is nowhere zero.

While the normal equations (A2) minimize ϵ^2 it is easy to show that normal equations (A3) minimize a related quantity s where

$$s = \epsilon^2+\lambda \sum_{k=1}^{nf} f_k^2. \tag{A4}$$

This can be shown by differentiating s with respect to the filter coefficients:

$$\begin{aligned}
\frac{\partial s}{\partial f_k} &= \frac{\partial}{\partial f_k}[\epsilon^2+\lambda\sum_l f_l^2] \\
&= \frac{\partial}{\partial f_k}[\sum_{i=1}^{nd}(d_i-\sum_j f_j w_{j-i})^2+\lambda\sum_l f_l^2] \\
&= 2\sum_{i=1}^{nd}(d_i-\sum_j f_j w_{j-i})w_{k-i}+2\lambda f_k \\
&= [\sum_i d_i w_{k-i}-\sum_j f_j \sum_i w_{j-i} w_{k-i}]+2\lambda f_k\delta_{jk}.
\end{aligned}$$

Now by setting the derivative equal to zero to minimize s we obtain the desired result, equations (A3).

$$\sum_j f_j(\sum_i w_{j-i}w_{k-i}+\lambda\delta_{jk}) = \sum_i d_i w_{k-i} \qquad k=1,\ldots nf.$$

An interesting interpretation of this result is possible if we introduce the descrete fourier transform (DFT) $\mathbf{F}$ of $\mathbf{f}$ defined as

$$\mathbf{F}_k = \sum_{j=1}^{nf} f_j e^{-i\,2\pi k\Delta\nu j\Delta t} \tag{A5}$$

with corresponding inverse transform

$$f_j = (1/n) \sum_{k=0}^{n} \mathbf{F}_k e^{-2\pi k\Delta\nu j\Delta t}. \tag{A6}$$

Parseval's theorem then states that the average filter power may be decomposed into contributions from each harmonic.

$$(1/n) \sum_{j=1}^{nf} f_j^2 = \sum_{k=0}^{n} |\mathbf{F}_k|^2$$

where $|\mathbf{F}_k|^2 = \mathbf{F}_k\mathbf{F}_k^*$ the filter energy at frequency $k\Delta\nu$.

This allows us to rewrite (A4) as

$$s = \epsilon^2 + \lambda n \sum_{k=0}^{n} |\mathbf{F}_k|^2$$

and conclude that minimizing this quantity too will lead to the stabilized normal equation (A3). We can now interpret the role of the ridge parameter λ in equations (A3). By designing an operator **f** with these equations we are seeking that operator which minimizes the squared error λ^2 plus some fraction of the filter energy at all frequencies. That fraction is λ, the trade off or ridge parameter. There are two extreme cases

1. $\lambda=0$: In this case the filter output will best reproduce the desired output. However, the filter energy, and hence the deconvolved output variance is unconstrained. If w is bandlimited and noisy the filter may amplify noise outside this frequency band rendering the output noisy if not useless.
2. $\lambda>>1$: In this case the filter power is the dominant term in the error measure s and by minimizing this a filter

with the least energy at all frequencies, will be designed independent of the wavelet.

Between these extreme cases a solution **f** is found which tends to minimize ϵ^2 while limiting the filter power at all frequencies. The question then arises, "Why not weight certain frequencies in the filter power of equation A so that the filter energy is supressed only outside the spectral passband of interest?" Our new error criterion would be

$$s' = \epsilon^2+\lambda n \sum_{k=0}^{n} Q_k |\mathbf{F}_k|^2 \tag{A8}$$

where Q_k, $k=0,\ldots n$ is a weighting function taking on values of unity outside, and near zero inside,the frequency band of interest. In order to implement this criterion we must first rewrite the weighted filter energy in (A8)

$$\begin{aligned} n \sum_k Q_k |\mathbf{F}_k|^2 &= n \sum_k Q_k \mathbf{F}_k{}^* \mathbf{F}_k \\ &= \sum_k Q_k \{ (\sum_j f_j e^{i\,2\pi k \Delta\nu j\,\Delta t}) (\sum_l f_l e^{i\,2\pi k\Delta\nu l\,\Delta t}) \} \\ &= \sum_j \sum_l f_j f_l q_{jl} \end{aligned}$$

where $$q_{jl} = n \sum_k Q_k e^{i\,2\pi k\Delta\nu (j-l)\Delta t} \tag{A9}$$

$\underset{\sim}{q}$ is an nfxnf matrix whose j'th row consists of n^2 times the inverse fourier transform of the window function Q_k centered on the diagonal. Now rewriting the new error criterion as

$$s' = \epsilon^2+\lambda \sum_j \sum_l f_j f_l q_{jl}$$

and differentiating with respect to the filter coefficients:

$$\frac{\partial s}{\partial f_k} = \frac{\partial}{\partial f_k}[\sum_i(d_i-\sum_j f_j w_{i-j})^2+\lambda\sum_{jl}\sum q_{jl}f_j f_l]$$

$$= 2\sum_i (d_i-\sum_j w_{i-j}f_j)w_{i-k}+\lambda 2\sum_j q_{kj}f_k f_j .$$

Now setting $\frac{\partial s}{\partial f_k} = 0 \qquad k=1,...nf \quad gives$

$$0 = \sum_i d_i w_{i-k}-\sum_{ij}\sum w_{i-j}w_{i-k}f_j+\lambda\sum_j q_{kj}f_k f_j$$

or $$\sum_j f_j(\sum_i w_{i-j}w_{i-k}+\lambda q_{kj}) = \sum_i d_i w_{i-k} \qquad k=1,...nf. \qquad \text{(A10)}$$

These are the new band-limited normal equations which again must be solved for filter coefficients f_j. Note the similarity between these and equations (A3) with ridge parameter. Instead of adding a constant to the diagonal of the autocorrelation matrix we add the matrix $\underset{\sim}{\mathbf{q}}$. Indeed if our spectral weights Q_k are chosen to be constant the band-limited normal equations revert to equations (A3). It is well known (Robinson and Treitel, 1980) that the autoccorrelation matrix is positive semidefinite. I will show below that the matrix $\underset{\sim}{\mathbf{q}}$ and hence their sum also has this property enabling the solution of equation (A9) by Levinson recursion. A Matrix $\underset{\sim}{\mathbf{A}}$ is positive, semidefinate, if for any real vector $\mathbf{x}$ we may write

$$\mathbf{x}^T\underset{\sim}{\mathbf{A}}\mathbf{x} \geq 0$$

Evaluating

$$
\begin{aligned}
\mathbf{x}^{\mathsf{T}}\underset{\sim}{\mathbf{q}}\mathbf{x} &= \sum_{r}\sum_{s} x_r q_{rs} x_s \\
&= \sum_{r}\sum_{s} x_r x_s n^2 \sum_k Q_k e^{i\,2\pi k\Delta\nu(r-s)\Delta t} \\
&= n^2 \sum_k Q_k \Big(\sum_r x_r e^{i\,2\pi k\Delta\nu r\Delta t}\Big)\Big(\sum_s x_s e^{i\,2\pi k\Delta\nu s\Delta t}\Big) \\
&= n^2 \sum_k Q_k X_k X_k^* \\
&= n^2 \sum_k Q_k |X_k|^2 \\
&\geq 0 \qquad \text{if } Q_k \geq 0 \text{ for all } k.
\end{aligned}
$$

That is, the matrix $\underset{\sim}{\mathbf{q}}$ is and hence the band-limited autocorrelation matrix is positive semidefinite as long as the weighting function Q_k is nowhere negative.

In this thesis a subroutine was used which implemented a spectral weighting function of the form.

$$Q(\nu_k) = \begin{cases} 1 & \nu_{lo} \le |\nu_k| \le \nu_{hi} \\ c & \text{otherwise } (c \ll 1). \end{cases}$$

The bandlimiting matrix $\underset{\sim}{q}$ can then be calculated using equation (A9) In this case however, it was calculated analytically as if sampled from a continuous spectrum. Then equation (A9) becomes

$$q_{jl} = n^2 \int_{-\nu_n}^{\nu_n} Q(\nu) e^{i 2\pi\nu(j-l)\Delta t} d\nu$$

$$= n^2 [\nu_n \mathrm{sinc}(2\pi\nu_n(j-l)\Delta t) + (c-1)\mathrm{sinc}(2\pi\nu_{hi}(j-l)\Delta t) - (c-1)\mathrm{sinc}(2\pi\nu_{lo}(j-l)\Delta t)]$$

Appropriate values for the nyquist frequency ν_n high and low bandlimits ν_{hi} and ν_{lo} are chosen, then the values q_{jl} are calculated and added to the autocorrelation matrix. Typical values selected might be:

$\nu_{lo} = 0$ Hz

$\nu_{hi} = 50$ Hz

$c = .01$

$\lambda = .05(\phi_{xx}(0))$ where $\phi_{xx}(0)$ is the data autocorrelation at zero lag.

In practice values of q_{jl} are normalized to unity on the diagonal. In its application to MED this bandlimiting algorithm limits the frequency content of the MED filter at each iteration. This is a distinctly different process than bandpassing the final output which in simple weiner filtering might achieve similar results. In MED the frequency content of the MED filter is controlled each time it attempts to match the non-linear reflectivity estimate. A high-variance noisy filter at any stage may disrupt the MED algorithms path to an extremum so that frequency control is essential at each iteration in noisy data application. The regretable practice of bandpassing the data after deconvolution comes too late in the case of MED. for by then a useless Varimax extremum may have already been reached.