Maximum likelihood estimation of residual wavelets

Clement Kostov

ABSTRACT

The method for maximum likelihood estimation of residual wavelets, presented in the SEP-50 report, is further elaborated in two aspects: (1) comparison between two measures of estimation errors, the variance and the bias, provides guidelines for the number of data to be used in the estimation, (2) estimation of residual wavelets from seismic data motivate the the use of models for statistically dependent input series.

INTRODUCTION

A method for blind deconvolution of residual wavelets was presented in the last SEP report (Kostov and Rocca, 1896). The method is an approximate maximum likelihood estimation of a small wavelet.

Blind deconvolution makes no assumptions about the phase of the wavelet. To recover the full wavelet, blind deconvolution methods (Wiggins, 1978, Bellini and Rocca, 1985, Walden, 1985) rely on a model for the probability distribution function (pdf) of the input sequence and track the changes in the pdf as a function of the wavelet.

The estimation errors in blind deconvolution can be broadly classified in two types – errors resulting from the random fluctuations in the input sequence, and errors arising from discrepancies between the data and the model.

The variance of the estimates resulting from random fluctuations of the input sequence is analyzed by deriving explicit expressions for the estimator and for the variance of the estimator as a function of the pdf of the input sequence (reflectivity), of the amount of data available, and of the constraints on the wavelet (Rocca and Kostov, 1987).

The sensitivity of the estimated wavelet to errors in the model – wrong assumption of statistical independence, or error in the parameters of the pdf of the input sequence – are discussed in the present paper.

DERIVATION OF THE ESTIMATOR

Outline

The likelihood function chosen for the estimation of the wavelet is the pdf of the wavelet conditional to the data. First, expressions for this likelihood function and for its gradient are derived. Then an estimator of the wavelet is defined as the product of a gain matrix times the gradient vector. Maximum likelihood estimator are asymptotically unbiased and minimum variance (Rosenblatt,1985). This property is used to characterize estimators of the wavelet and to obtain explicit expressions for the estimator and for its variance in the case of generalized Gaussian input sequences.

Notations

The convolutional model is

$$\underline{y} = \underline{x} + \underline{a}*\underline{x} = \underline{x}*(\underline{\delta} + \underline{a}), \tag{1}$$

where $\underline{x}$ is the input sequence of length N (reflectivity), $\underline{y}$ is the output sequence of length N (data), and $\underline{\delta} + \underline{a}$ is a "deltalike" wavelet . The wavelet $\underline{\delta}$ is the initial guess for the wavelet and the wavelet $\underline{a}$ is the unknown residual wavelet. In the sequel, we always assume $a_0 = 0$.

The likelihood function and its gradient

Applying Bayes theorem, the likelihood function $p_{\underline{a}}(\underline{a} \mid \underline{y})$, which is also the pdf of the wavelet given the data, is related to the pdf of the data given the wavelet

$$p_{\underline{a}}(\underline{a} \mid \underline{y}) = \frac{p_{\underline{y}}(\underline{y} \mid \underline{a})p_{\underline{a}}(\underline{a})}{p_{\underline{y}}(\underline{y})}. \tag{2}$$

An alternative expression of the likelihood function (Kostov and Rocca, 1987), which depends on the pdf model $p_{\underline{x}}$ of the reflectivity sequence $\underline{x}$, rather than on the unknown pdf $p_{\underline{y}}(\underline{y} \mid \underline{a})$ of the data $\underline{y}$, is given by

$$p_{\underline{a}}(\underline{a} \mid \underline{y}) = \frac{p_{\underline{x}}(\underline{y} - \underline{a}*\underline{y} \mid \underline{a})}{p_{\underline{x}}(\underline{y})}. \tag{3}$$

The components γ_k of the gradient $N\underline{\gamma}(\underline{y})$ of the likelihood function defined in (6) are obtained by a Taylor expansion of the likelihood function (Kostov and Rocca, 1987),

$$\gamma_k = \frac{1}{Np_{\underline{x}}(\underline{y})}\sum_{i=1}^{N}\frac{\partial p_{\underline{x}}}{\partial x_i}(\underline{y})y_{i-k} \quad , \text{ for } \mid k \mid \leq \frac{M}{2}. \tag{4}$$

For iid input sequences the components of the gradient simplify further to

$$\gamma_k = \frac{1}{N}\sum_{i=1}^{N}\frac{p'}{p}(y_i)y_{i-k} \quad , \text{ for } \mid k \mid \leq \frac{M}{2}, \tag{5}$$

where $p(x)$ is a one-dimensional pdf function.

The gain matrix for an unbiased estimator

The estimator $\underline{\hat{a}}$ of the convolutional wavelet $\underline{a}$ is written a priori as

$$\underline{\hat{a}} = \underline{G}\ \underline{\gamma}(\underline{y}), \tag{6}$$

where the gain matrix $\underline{G}$ is $(M+1)\times(M+1)$ and the vectors $\underline{\hat{a}}$ and $\underline{\gamma}$ are $(M+1)$.

The gain matrix $\underline{G}$ is obtained by requiring that the estimator $\underline{\hat{a}}$ be unbiased, that is $E(a) = E(\underline{\hat{a}})$. The term $E(\underline{\hat{a}})$ can be related to $E(\underline{a})$ by linearizing equation (6) with respect to $\underline{a}$,

$$\begin{aligned} E(\underline{\hat{a}}) &= \underline{G}E(\underline{\gamma}(\underline{y})) = \underline{G}E(\underline{\gamma}(\underline{x} + \underline{a}*\underline{x})) \\ &= \underline{G}\,[\,E(\underline{\gamma}(\underline{x})) + E(\underline{\gamma}'(\underline{x}))\,]E(\underline{a}). \end{aligned} \tag{7}$$

The expression of the $(M+1)\times(M+1)$ matrix $\underline{\gamma}'(\underline{x})$ is given in Appendix A.

The above equations (7) are compatible with the unbiasedness condition if and only if

$$\underline{G}\ E[\underline{\gamma}(\underline{x})] = 0 \quad \text{and} \tag{8a}$$

$$[\,\underline{I} - \underline{G}\ E[\underline{\gamma}'(\underline{x})]\,]E(\underline{a}) = 0. \tag{8b}$$

When the vector $E(\underline{a})$ is arbitrary, i.e. when no prior information on the coefficients is available, equation (11b) implies

$$\underline{G}\ E[\underline{\gamma}'(\underline{x})] = \underline{I}'\ .$$

Gain matrix and estimator for independent generalized Gaussian variables

For a generalized Gaussian input sequence, it is shown in Appendix A that

$$\begin{aligned} E[\underline{\gamma}(\underline{x})] &= -\underline{\delta}\ , \ \text{and} \\ E[\underline{\gamma}'(\underline{x})] &= t(\alpha)\underline{I} + \underline{I}'\ , \end{aligned} \tag{9}$$

where $\underline{I}'$ is the matrix with entries equal to one along the secondary diagonal and entries equal to zero elsewhere.

The function $t(\alpha)$ in equation (9) depends on the shape parameter α characterizing the pdf of the input sequence; $t(\alpha)$ is plotted in Figure 1 and its analytical expression is given in Appendix B. The function $t(\alpha)$ attains its minimum value of 1 when the shape parameter is equal to 2, i.e. for a Gaussian distribution, and increases monotonically as the distribution moves away from the Gaussian.

The gain matrix $\underline{G}$ is the inverse of the matrix $E[\underline{\gamma}'(\underline{x})]$, given as a function of $t(\alpha)$ in equation (9),

$$\underline{G} = \frac{1}{t(\alpha)^2 - 1}[t(\alpha)\underline{I} - \underline{I}'\ ]. \tag{10}$$

The gain matrix is defined for all distributions except the Gaussian, for which the estimation of the wavelet is in general not possible. The expression for the estimator of the wavelet $\underline{a}$ follows from equation (10)

$$\underline{\hat{a}} = \frac{1}{t(\alpha)^2 - 1}\,[t(\alpha)\underline{I} - \underline{I}'\ ]\,\underline{\gamma}(\underline{y}). \tag{11}$$

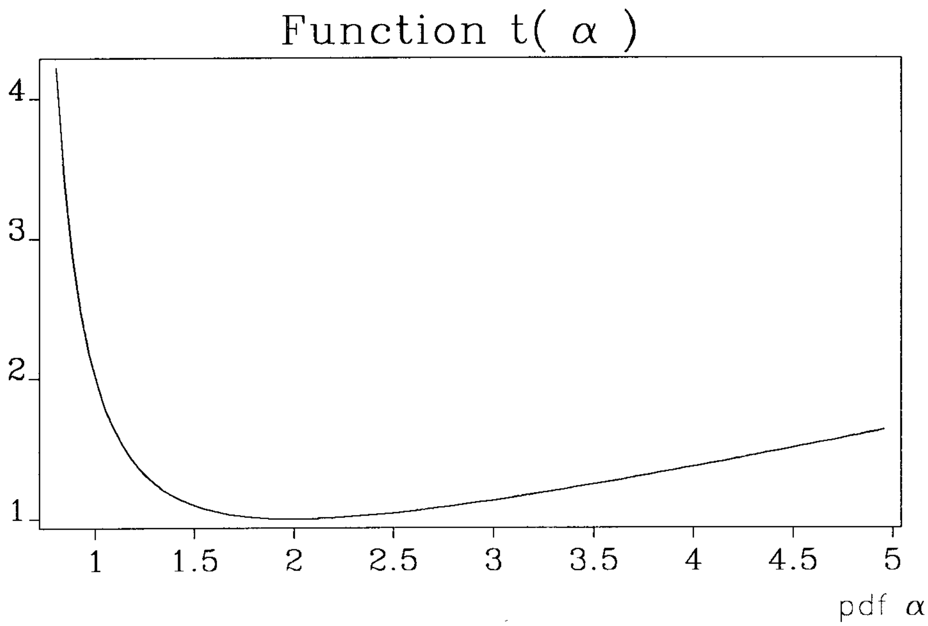

FIG. 1. Function $t(\alpha)$ used in the definitions of the gain factors for different types of wavelets. Notice that the minimum value of the function $t(\alpha)$ is attained for $\alpha = 2$, $t(2) = 1$.

Variance of the estimator

The estimator $\underline{\hat{a}}$ is defined in equation (11) as a function of the data vector $\underline{y}$. To analyze theoretically the variance of the estimator, $\underline{\Upsilon}(\underline{y})$ is approximated to order zero in $\underline{a}$ as $\underline{\Upsilon}(\underline{x})$. The expression of the estimator becomes

$$\underline{\hat{a}} = \frac{1}{t(\alpha)^2 - 1} [t(\alpha)\underline{I} - \underline{I}'] \underline{\Upsilon}(\underline{x}). \tag{12}$$

The $(M+1)\times(M+1)$ covariance matrix of the estimator $\underline{\Sigma}$ is defined as

$$\underline{\Sigma} = E[(\underline{\hat{a}} - \underline{a})(\underline{\hat{a}} - \underline{a})^T]. \tag{13}$$

To first order in the coefficients of the wavelet $\underline{a}$, the covariance matrix of the unbiased estimator $\underline{\hat{a}}$,

$$\underline{\Sigma} = \frac{1}{N} \frac{[t(\alpha)\underline{I} - \underline{I}']}{t(\alpha)^2 - 1}, \tag{14}$$

is obtained by combining equations (13) and (14) and by using the following identity for the covariance matrix of the gradient vector derived in Appendix B,

$$E[\underline{\Upsilon}(\underline{x})\underline{\Upsilon}(\underline{x})^T] = \frac{1}{N}[t(\alpha)\underline{I} + \underline{I}'].$$

The factor $\frac{1}{N}$ implies that the covariance tends to zero as the number of data goes to infinity. The variance of the estimator tends toward infinity as the distribution tends toward the Gaussian (the covariance matrix is proportional to the gain factor of the estimator $\underline{\hat{a}}$, equation (11)).

PRIOR INFORMATION

Non-uniqueness of the unbiased estimator

Prior information about the wavelet, expressed as linear constraints on the wavelet's coefficients, restricts the wavelet $\underline{a}$ and its estimate $\underline{\hat{a}}$ to a subspace F of the space spanned by the gradient vector $\underline{\gamma}$. The conditions defining the gain matrix $\underline{G}$ (equations (7)) become less strict, and there might be several possible choices for gain matrices and correspondingly for the unbiased estimators; causal wavelets for instance have coefficients equal to zero at negative time lags, and therefore the corresponding entries of the gain matrix are arbitrary.

Ambiguities in the choice of an estimator can be resolved by choosing an unbiased minimum variance estimator compatible with the constraints (Rocca and Kostov, 1987). The equations defining unbiased, minimum variance estimators are derived below for the case of linear constraints.

Linear constraints: examples

- Even wavelets
  Even wavelets are defined by $\frac{M}{2}$ independent parameters and $\frac{M}{2}$ independent linear relations, $a_k = a_{-k}$, for $1 \le k \le \frac{M}{2}$.
- Odd wavelets
  Odd wavelets are defined by $\frac{M}{2}$ independent parameters. The $\frac{M}{2}$ relations among the wavelets coefficients are $a_k = -a_{-k}$.
- Causal wavelets
  Causal wavelets are such that $a_k = 0$ for $k \le 0$. The $\frac{M}{2}$ coefficients for positive lags are arbitrary.
- Residual wavelets of known direction
  The direction $\underline{h}$ of the residual wavelet is known; only a scale factor remains to be estimated. For instance, to estimate a small constant phase shift, the residual wavelet could be specified as proportional to the time-domain representation of the Hilbert transform (Appendix C).

Linear constraints: general case

In each of preceding cases the prior constraint on the wavelet's type was expressed as a system of linear equations, $\underline{L}\ \underline{a} = \underline{0}$, satisfied by the wavelets coefficients. More conveniently, the estimator and the variance of that estimator are expressed in terms of the projection onto the solution space of the linear system $\underline{La} = \underline{0}$. Let $\underline{P}$ denote this projection operator. When $\underline{L}$ itself is a projection operator, that is $\underline{L}^2 = \underline{L}$, the relation between $\underline{P}$ and $\underline{L}$ is simply $\underline{P} = \underline{I} - \underline{L}$. To define an even wavelet, an operator $\underline{L}$ equal to $\frac{\underline{I} - \underline{I}'}{2}$, could be used. The equation $\underline{L}\ \underline{a} = \underline{0}$ requires that the odd part of an even wavelet be zero.

Table 1 summarizes the definitions of different wavelet's types in terms of the operators $\underline{L}$ and $\underline{P}$, and lists for each type of wavelet the number of independent coordinates that need to be estimated.

Table 1

Prior information			
type of wavelet	$\underline{L}$	$\underline{P} = \underline{I} - \underline{L}$	number of parameters
arbitrary	$\underline{0}$	$\underline{I}$	M
even	$\frac{\underline{I} - \underline{I}'}{2}$	$\frac{\underline{I} + \underline{I}'}{2}$	$\frac{M}{2}$
odd	$\frac{\underline{I} + \underline{I}'}{2}$	$\frac{\underline{I} - \underline{I}'}{2}$	$\frac{M}{2}$
causal	$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$	$\frac{M}{2}$

The unbiased minimum variance estimator

The a priori expression for the estimator is now $\underline{\hat{a}} = \underline{PG}\,\underline{\gamma}(\underline{y})$, where $\underline{P}$ is the projection onto the space F spanned by the gradient vector. The unbiasedness condition should hold only in the subspace F, and therefore equation (10) should be projected onto F

$$\underline{P}\,[\underline{I} - \underline{G}E(\underline{\gamma}'(\underline{x}))]\underline{P} = 0, \tag{15}$$

We solve for a gain matrix $\underline{G}$ written a priori as $\underline{G} = p\underline{I} + q\underline{I}'$, because the inverse of $E(\underline{\gamma}'(\underline{x}))$ is a linear combination of $\underline{I}$ and $\underline{I}'$ (equation (10)). The coefficients p and q are the unknowns.

After substitution of $\underline{G}$ into equation (15), the unbiasedness condition is obtained in terms of p and q:

$$(pt(\alpha) + q - 1)\underline{P} + (qt(\alpha) + p)\underline{PI}'\,\underline{P} = \underline{0}. \tag{16}$$

When the matrices $\underline{P}$ and $\underline{PI}'\,\underline{P}$ are linearly dependent, equation (16) provides only one relation for the two unknowns p and q. In that case the pair (p, q) which minimizes the variance of the estimator subject to the unbiasedness constraint (16) is chosen. The covariance matrix of the estimator projected onto the the subspace F is given by

$$\begin{aligned}\underline{\Sigma} &= \underline{PGE}(\underline{\gamma\gamma}^T)\underline{GP} = \\ &= [t(\alpha)(p^2 + q^2) + 2pq]\underline{P} + [p^2 + q^2 + 2t(\alpha)pq]\underline{PI}'\,\underline{P}\end{aligned}$$

The covariance matrix is proportional to $\underline{P}$. The proportionality factor $f(p, q, t(\alpha))$ measures the variance of the estimator; it is a quadratic form in the coefficients p and

q , and depends also on the input pdf model through the function $t(\alpha)$. To determine p and q, $f(p, q, t(\alpha))$ is minimized subject to the linear constraint between p and q given by the unbiasedness condition (16).

For linearly independent matrices $\underline{P}$ and $\underline{PI'}\ \underline{P}$ the coefficients of these matrices in equation (16) are set to zero and the same gain factors for the estimation of arbitrary wavelets as in equation (11) are obtained.

Estimates and their variances for particular wavelet types

For even wavelets, odd wavelets, causal wavelets and wavelets of known direction the operators $\underline{P}$ and $\underline{PI'}\ \underline{P}$ are linearly dependent. The method described in the previous section is applied for each wavelet type and the operator $\underline{PI'}\ \underline{P}$, the gain matrix, the estimator and the covariance matrix of the estimator are determined. The results are summarized in Table 2.

Table 2

type of wavelet	$\underline{P}$	$\underline{PI'}\ \underline{P}$	estimator $\underline{\hat{a}}$	covariance $\underline{\Sigma}$
arbitrary	$\underline{I}$	$\underline{I'}$	$\frac{t(\alpha)\underline{I} - \underline{I'}}{t(\alpha)^2 - 1} \underline{\gamma}(\underline{y})$	$\frac{1}{N} \frac{t(\alpha)\underline{I} - \underline{I'}}{t(\alpha)^2 - 1}$
even	$\frac{\underline{I} + \underline{I'}}{2}$	$\frac{\underline{I} + \underline{I'}}{2}$	$\frac{\underline{I} + \underline{I'}}{2(t(\alpha) + 1)} \underline{\gamma}(\underline{y})$	$\frac{1}{N} \frac{\underline{I} + \underline{I'}}{2(t(\alpha) + 1)}$
odd	$\frac{\underline{I} - \underline{I'}}{2}$	$\frac{\underline{I} - \underline{I'}}{2}$	$\frac{\underline{I} - \underline{I'}}{2(t(\alpha) - 1)} \underline{\gamma}(\underline{y})$	$\frac{1}{N} \frac{\underline{I} - \underline{I'}}{2(t(\alpha) - 1)}$
causal	$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$	$\underline{0}$	$\frac{\underline{I}}{t(\alpha)} \underline{\gamma}(\underline{y})$	$\frac{1}{N} \frac{\underline{I}}{t(\alpha)}$

Estimation of a constant phase-shift

The wavelet $\underline{\hat{a}}$, corresponding to a constant phase-shift, could be estimated in one of the two two following methods:

(1) as the projection of $\underline{a}$ along the known odd direction $\underline{h}$ ($\underline{h}$ is the discrete Hilbert transform and $\underline{H}$ is the projection onto $\underline{h}$, defined in Appendix C)

$$\underline{\hat{a}} = \frac{\underline{H}\,\underline{\gamma}(\underline{y})}{(t(\alpha) - 1)},$$

(2) as an odd wavelet (since $\underline{h}$ is odd), given by

$$\underline{\hat{a}} = \frac{\underline{\gamma}(\underline{y}) - \underline{I'}\ \underline{\gamma}(\underline{y})}{2(t(\alpha) - 1)}.$$

In both cases, the estimator of the constant phase shift is $\hat{\theta} = \frac{\underline{h}.\underline{\hat{a}}}{|\underline{h}|^2}$ with variance:

$$var(\hat{\theta}) = E(\hat{\theta}^2) = \frac{\underline{h}^T E(\underline{\hat{a}}\underline{\hat{a}}^T)\underline{h}}{\underline{h}^4} = \frac{1}{2N(t(\alpha)-1)} \times \frac{1}{\sum_{i=1}^{M/2} h_i^2} \qquad (17)$$

Comparison of the variances

Figures 2a and 2b show the variances for four types of wavelets – arbitrary, odd, causal, and even – as a function of the shape parameter α of the generalized Gaussian pdf's. The plots distinguish between super-Gaussian distributions for which $\alpha < 2$ and sub-Gaussian distributions for which $\alpha > 2$, because the variances for arbitrary and odd wavelets tend to infinity as the distribution tends toward the Gaussian.

The variance for the estimate of an arbitrary wavelet, when no prior information is available, is highest. The curves for the estimates of the odd and even parts of a wavelet, are different by several orders of magnitude. The variance on the odd part is practically identical to the total variance.

For any given type of wavelet, the variance is highest (possibly infinite) for the Gaussian distribution. The variances remain finite for Gaussian pdf only when the phase of the full "deltalike" wavelet is known a priori as for even or causal residual wavelets.

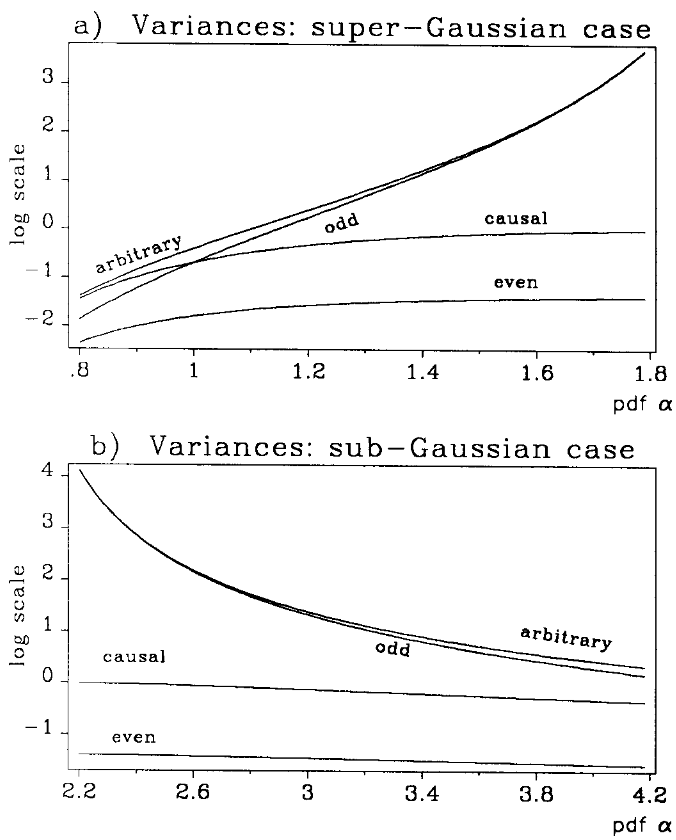

FIG. 2. Theoretical variances of the estimated wavelets represented on a logarithmic scale as function of the input pdf and prior information. The amount of data and the scale of the residual wavelet are fixed. a) super-Gaussian case. b) sub-Gaussian case.

Sensitivity to the pdf model

What if the shape parameter $\tilde{\alpha}$ of the input sequence $\underline{x}$ is different from the parameter α assumed for the estimation?

The gain matrix $\underline{G}$ is no longer the inverse of the matrix $E[\underline{\Upsilon}'(\underline{x})]$ and therefore the unbiasedness condition, equation (8b), is no longer satisfied.

From equation (9), the matrix $E[\underline{\Upsilon}'(\underline{x})]$ is equal to

$$E[\underline{\Upsilon}'(\underline{x})] = t(\tilde{\alpha})\underline{I} + \underline{I}' = t(\alpha)\underline{I} + \underline{I}' + (\Delta t)\underline{I}, \quad (18)$$

where $(\Delta t) = t(\tilde{\alpha}) - t(\alpha)$.

The bias of the estimator can be derived from equation (7),

$$E(\underline{\hat{a}}) - E(\underline{a}) = E(\underline{a})\left\{ \underline{I} + (\Delta t)\underline{G} \right\}.$$

Thus, an error in the shape parameter of the input sequence leads to an error on the gain matrix and to a biased estimator. The bias is proportional to the product of two terms – one related to the pdf model,

$$(\Delta t)\underline{G} = \frac{(\Delta t)t(\alpha)}{t(\alpha)^2 - 1}\underline{I}, \quad (19)$$

and the other, $E(\underline{a})$, related to the magnitude of the residual wavelet. The factor in the bias which depends on the pdf model is plotted in Figure 3.

The comparison between the bias and the variance of the estimator sets some practical guidelines for the estimation of the wavelet. The variance of the estimator, due to the random fluctuations of the input sequence, decreases as $\frac{1}{N}$, where N is the number of data. On the other hand, increasing the number of data beyond the number for which the variance becomes smaller than the bias term is not going to improve the accuracy of the estimation. This limit number is obtained from equations (14) and (19),

$$N_{\max} \leq \frac{1}{(\Delta t)E(\underline{a})}. \quad (20)$$

For instance, when the shape parameter of the input sequence is close to $\alpha = 1$, and the magnitude of the coefficients of the residual wavelet is about 10%, equation (20), provides a relation $N_{\max} \times (\Delta\alpha) \leq 30$.

Numerical examples: estimation of constant phase-shifts

The examples presented in this section illustrate the estimation of small constant phase shifts, both from simulated and from seismic data.

The model and the estimation of small phase shifts are described first using simulated data. Constant phase shifts, from -10^o to $+10^o$ were applied to a time series of 14400 random numbers, drawn from a generalized Gaussian pdf with shape parameter $\alpha = 1.2$. For each constant phase shift, a residual wavelet was estimated, and a constant phase-shift correction was computed from the estimated wavelet according to equation (17). The results are shown in Figure 4; both the sign and the amplitude of the correction factor are well estimated in this range of phase-shifts and make it possible to recover the input sequence by iterating the residual wavelet deconvolution.

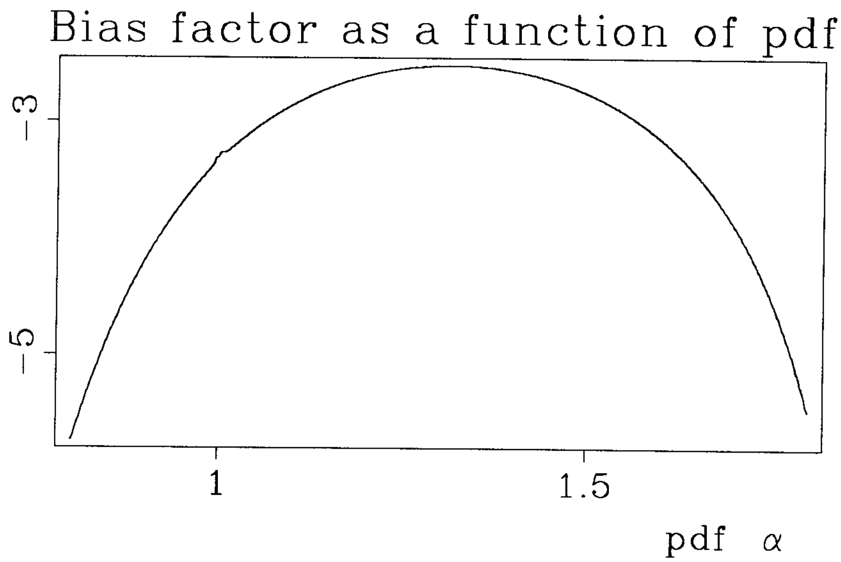

FIG. 3. The bias resulting from an error in the shape parameter of the pdf model is the product of a factor (shown in the figure) which depends on the pdf only, times the magnitude of the residual wavelet.

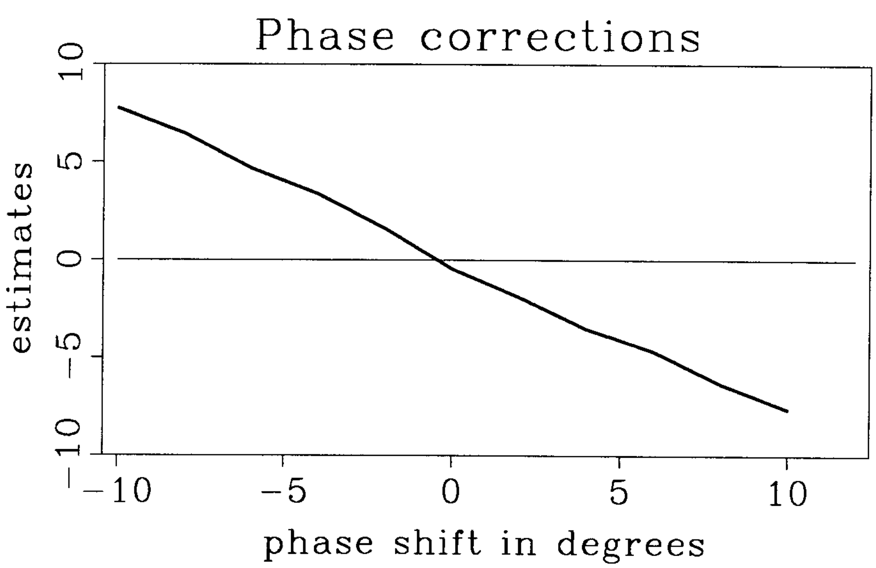

FIG. 4. Phase corrections computed from the estimated residual wavelet for simulated random numbers. The sign and the magnitude of the corrections are well estimated.

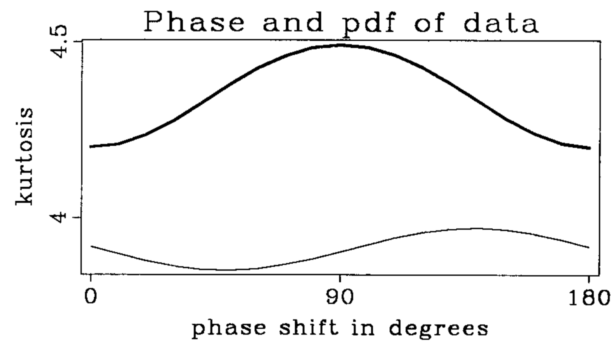

Fig. 5: Light line: change of the kurtosis of the original seismic data as a function of phase-shift. Heavy line: change of the kurtosis of the seismic data as a function of phase-shift after after suppression of long period multiples. Removal of long period multiples increases the non-Gaussianity if the data and the sensitivity of the kurtosis to phase shifts.

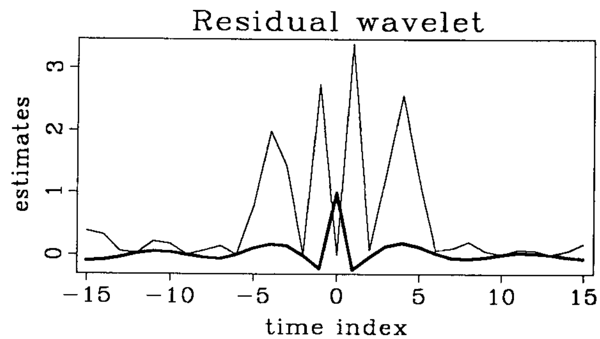

Fig. 6: The residual wavelet (heavy line) computed from seismic data, after removal of long period multiples and 90 degrees phase shift, is nearly symmetric. Thus, the residual wavelet attempts to correct mainly for the non-whiteness of the data.
The variance of the estimates (light line) is high and strongly dependent on the amplitude of the wavelet's coefficients.

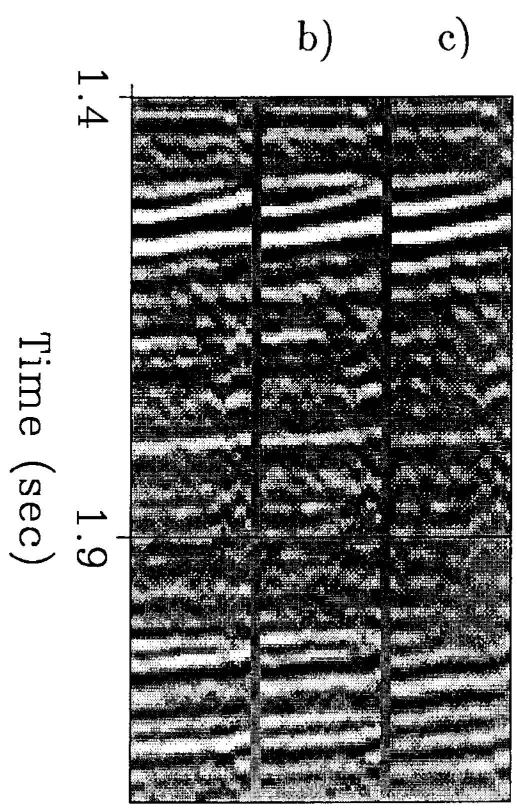

Fig. 7: A window of data after three processes: (a) original data, after normal moveout and removal of long period multiples.
(b) same data as in a), after 90 degrees phase rotation.
(c) same data as in b), after applying the correction computed from the residual wavelet.

A similar experiment – estimation of small constant phase-shifts – was performed with seismic data. The data are a window of 24 traces, each 600 samples long, from a marine shot profile (#27, Yilmaz and Cuomro, 1983). The long period multiples are suppressed in order to balance the spectrum and increase the non-Gaussianity of the data, thus sharpening changes in the pdf with phase-shift (Figure 5). Figure 5 shows that the pdf varies smoothly with phase-shift, reaching an extremum for a phase-shift of 90^o , for which it is the farthest away from the Gaussian pdf.

The coefficients of the residual wavelet estimated from the 90^o phase-shifted data are plotted in Figure 6. The wavelet is nearly symmetric, thus implying that the main correction is for the color of the amplitude spectrum. The variance of the coefficients is also high and depends on the amplitude of the coefficients.

Figure 7 shows the original data (panel 7a), the data after a phase-shift of 90^o (panel 7b), and the data after deconvolution with the residual wavelet (panel 7c). The sharpness of the data increases from panel 7a to panel 7c, as the kurtosis (measure of the pdf) increases from 4.2 for the original data (7a) , to 4.5 for the 90^o phase-shifted data (7b) and reaches 4.6 after three iterations of the residual wavelet deconvolution (7c).

Figure 8 demonstrates that the residual wavelet does not follow correctly changes in the phase of the seismic data. Neither the sign, nor the amplitude of the estimated phase are correct. This effect is expected, because the antisymmetric part of the wavelet in Figure 6 is nearly zero. The residual wavelet attempts to correct for the main discrepancy between the data and the model, which is the colored spectrum of the seismic data. To obtain better results, the model for the estimation of the residual wavelet should account for such a colored spectrum.

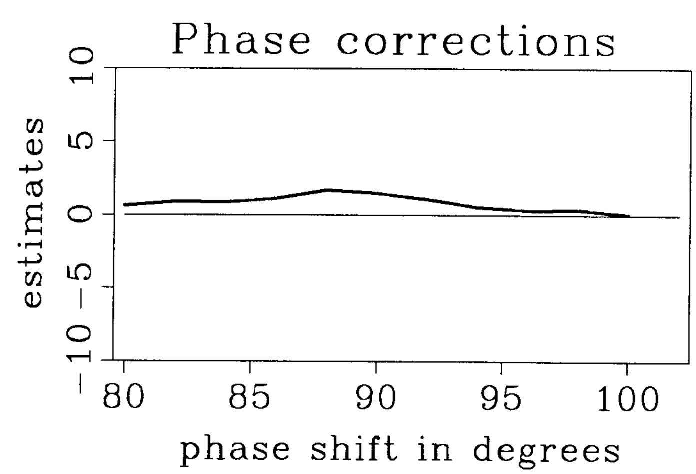

FIG. 8. Phase corrections computed from the estimated residual wavelet for seismic data. The sign and the magnitude of the phase-shifts are incorrectly estimated.

THE GRADIENT FOR NON IID INPUT SEQUENCES

Overview

Models of reflectivity series that do not assume statistical independence fit better the experimental data, according to a recent paper by Walden and Hosken (1985) and other studies referenced in that paper.

The gradient of the likelihood function is derived in this section for an input sequence modeled as a first order Markov process. Higher order Markov chains have been used as models for the logarithm of impedance in the work of Godfrey, Muir and Rocca (1980). For first-order Markov input sequences the gradient remains the cross-correlation between a non-linear transform of the data sequence and the data sequence itself. For iid sequences the non-linear transform has "zero memory" (Godfrey and Rocca, 1981), while for dependent sequences the transform may depend on more than one sample.

First order Markov processes

Given the following pdf model for the input sequence

$$p_{\underline{x}}(\underline{x}) = \prod_{i=1}^{N} p(x_i \mid x_{i-1})$$

the expression for the k^{-th} component of the gradient is of the form

$$\gamma_k = \frac{1}{N}\sum_{i=1}^{N} u(y_{i+1}, y_i, y_{i-1}) y_{i-k} .$$

As in the case of an iid input sequence the gradient is the cross-correlation of the data sequence with a non-linear transform of the data sequence. The non-linear transform is now a function of several variables

$$u(y_{i+1}, y_i, y_{i-1}) = \frac{\frac{\partial}{\partial y_i} p(y_i \mid y_{i-1})}{p(y_i \mid y_{i-1})} + \frac{\frac{\partial}{\partial y_i} p(y_{i+1} \mid y_i)}{p(y_{i+1} \mid y_i)} .$$

We give the particular expressions for the gradient in three cases:

a) independent input samples x_i

Then, $p(x_i \mid x_{i-1}) = f(x_i)$ and the components of the gradient are identical to the components for iid input sequences derived in equation (5):

$$\gamma_k = \frac{1}{N}\sum_{i=1}^{N} \frac{f'(y_i)}{f(y_i)} y_{i-k} \quad \text{for} \quad |k| \leq \frac{M}{2} .$$

In this case the non-linear transform is a function of one variable only ("zero-memory non-linearity"), equal to $f'(y_i)$.

b) independent increments $x_i - x_{i-1}$

In that case, $p(x_i \mid x_{i-1}) = f(x_i - x_{i-1})$ and the components of the gradient are

$$\gamma_k = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{f'(y_i - y_{i-1})}{f(y_i - y_{i-1})} - \frac{f'(y_{i+1}-y_i)}{f(y_i - y_{i-1})}\right] y_{i-k}$$
$$= \frac{1}{N}\sum_{i=1}^{N}\frac{f'(y_i - y_{i-1})}{f(y_i - y_{i-1})}[y_{i-k} - y_{i-1-k}].$$

The gradient in this case is equal to the gradient of the iid sequence of increments. The non linear function depends on three consecutive samples and is equal to:

$$u(y_{i+1}, y_i, y_{i-1}) y_{i-k} = \frac{f'(y_i - y_{i-1})}{f(y_i - y_{i-1})} - \frac{f'(y_{i+1}-y_i)}{f(y_i - y_{i-1})}.$$

c) a combination of the previous cases, $p(x_i \mid x_{i-1}) = \alpha f(x_i) + \beta g(x_i - x_{i-1})$

The gradient and the non linear function for this case are also linear combinations of the gradients and of the non linear functions computed in a) and b). A model of this kind could be used for a process with a few large transition events (occurring with probability $\alpha f(x_i)$) and small variations between transitions (occurring with probability $\beta g(x_i - x_{i-1})$).

CONCLUSION

Two measures of errors in the estimation of residual wavelets, the variance and the bias of the estimator, were discussed. The variance measures the sensitivity of the estimator to random fluctuations in the input sequence, while errors in the model for input the pdf sequence introduce bias. Comparison of the two error measurements provides a bound on the number of data to be used for the estimation of the residual wavelet.

Computations of residual wavelets from seismic data point out the importance of the color of the spectrum for the estimation. Suggestions for modifying the estimator for correlated data modeled as a Markov chain are given.

REFERENCES

Bellini, S., and Rocca, F., Blind deconvolution: Polyspectra or Bussgang techniques?, *in* Biggleri and Prati Eds., Digital Communications, Elsevier Science Publishers, 251-263.

Claerbout, J.F., 1985, Fundamentals of Geophysical Data Processing: Blackwell Scientific Publications.

Godfrey, R., Muir, F. and Rocca, F., 1980, Modeling seismic impedances with Markov chains: Geophysics 45, 1351-1372.

Godfrey, R. and Rocca, F., 1981, Zero memory non-linear deconvolution: Geophys. Prosp., 29,189-228.

Kostov, C. and Rocca, F., 1986, Estimation of residual wavelets: SEP-50,1-20.

Rocca, F. and Kostov, C., 1987, Estimation of residual wavelets: Proceedings of the Conference on Deconvolution and Inversion, Rome, 1986, *inpress* .

Rosenblatt, M., 1985, Stationary Sequences and Random Fields: Birkhauser.

Walden, A.T., 1985, Non-Gaussian reflectivity, entropy, and deconvolution: Geophysics, 50, 2862-2888.

Walden, A.T., and Hosken, J.W.J., 1985, An investigation of the spectral properties of primary reflection coefficients, Geophysical Prospecting 33, 400-435.

Wiggins, R.A., 1978, Minimum entropy deconvolution: Geoexpl., 16,21-35.

Yilmaz, O., and Cuomro, D., 1983, Worldwide Assortment of Field Seismic Records, released by Western Geophysical Company of America, Houston.

APPENDIX A

Matrix $E(\underline{\gamma}'(\underline{x}))$

The $(M+1)\times(M+1)$ matrix $\underline{\gamma}(\underline{x})$ is defined through the relation

$$\underline{\gamma}(\underline{y}) = \underline{\gamma}(\underline{x}) + \underline{\gamma}'(\underline{x})\underline{a}.$$

A Taylor expansion of the k^{-th} component of the vector $\underline{\gamma}(\underline{y})$ leads to:

$$\gamma_k(\underline{y}) = \frac{1}{N}\sum_{i=1}^{N} u(y_i)y_{i-k} =$$

$$= \frac{1}{N}\sum_{i=1}^{N}[u(x_i) + \sum_{l=-M/2}^{l=M/2} a_l x_{i-l} u'(x_i)][x_{i-k} + \sum_{m=-M/2}^{m=M/2} a_m x_{i-k-m}] =$$

$$= \gamma_k(\underline{x}) + \frac{1}{N}\sum_{i=1}^{N}\sum_{l=-M/2}^{l=M/2} a_l [x_{i-l} x_{i-k} u'(x_i) + x_{i-k-l} u(x_i)],$$

where the function $u(x)$ denotes $\frac{p'(x)}{p(x)}$.

From the above equations, we derive the expression for the entry in row k and column l of the matrix $\underline{\gamma}'(\underline{x})$:

$$[\underline{\gamma}'(\underline{x})]_{k,l} = \frac{1}{N}\sum_{i=1}^{N} x_{i-l} x_{i-k} u'(x_i) + \frac{1}{N}\sum_{i=1}^{N} x_{i-k-l} u(x_i).$$

The samples x_i are statistically independent, therefore

$$E[\underline{\gamma}'(\underline{x})]_{k,l} = 0, \text{ for } k \neq l \text{ and } k+l \neq 0,$$

$$E[\underline{\gamma}'(\underline{x})]_{k,l} = E(X^2)E(u'(X)) \text{ for } k = l, \text{ and}$$

$$E[\underline{\gamma}'(\underline{x})]_{k,l} = E(Xu(X)) \text{ for } k+l = 0.$$

For a random variable X with a generalized Gaussian pdf, the expected values $E(X^2)$, $E(Xu(X))$ and $E(u'(X))$ were computed in Kostov and Rocca, 1986, and the expression for the matrix $E(\underline{\gamma}'(\underline{x}))$ is:

$$E(\underline{\gamma}'(\underline{x})) = t(\alpha)\underline{I} + \underline{I}'.$$

Matrix $E(\underline{\gamma}(\underline{x}))$

The k^{-th} component of the gradient vector $\underline{\gamma}(\underline{x})$ is

$$\gamma_k = \frac{1}{N}\sum_{i=1}^{N} u(x_i)x_{i-k}.$$

The expected value $E(\gamma_k(\underline{x}))$ is zero for $k \neq 0$ because the samples x_i are independent. For $k = 0$, and for a random variable X with a generalized Gaussian pdf

$E(\gamma_0(\underline{x})) = E(Xu(X)) = -\delta$, as shown in Kostov and Rocca, 1986.

Matrix $E(\underline{\gamma\gamma}^T(\underline{x}))$

The entry for row k and column l is:

$$E[\underline{\gamma\gamma}^T(\underline{x})_{k,l}] = \frac{1}{N^2}\sum_{i,j=1}^{N} u(x_i)x_{i-k}\,u(x_j)x_{j-l}\,.$$

The samples x_i are statistically independent, therefore only terms such that $i = j$ and $k = l$, or $i-k = j$ and $j-l = i$ contribute to a non-zero mean value,

$$E[\underline{\gamma\gamma}^T(\underline{x})_{k,l}] = 0 \quad \text{for} \quad k \neq l \quad \text{and} \quad k + l \neq 0,$$

$$E[\underline{\gamma\gamma}^T(\underline{x})_{k,l}] = \frac{1}{N}[E(u(X))]^2 E(X^2) \quad \text{for} \quad k = l, \quad \text{and}$$

$$E[\underline{\gamma\gamma}^T(\underline{x})_{k,l}] = \frac{1}{N}[E(Xu(X))]^2 \quad \text{for} \quad k + l = 0.$$

For a generalized Gaussian random variable X,

$$E(\underline{\gamma\gamma}^T(\underline{x})) = \frac{1}{N}(t(\alpha)\underline{I} + \underline{I}'\ ).$$

APPENDIX B

Expression for the gain function of the estimator

The pdf of a generalized Gaussian random variable with shape parameter α and scale parameter β is

$$p(x) = \frac{1}{2\beta\Gamma(\frac{1}{\alpha})} e^{[-|\frac{x}{\beta}|^{\alpha}]}.$$

The logarithmic derivative of $p(x)$ is denoted by $u(x)$,

$$u(x) = \frac{p'(x)}{p(x)} = -\alpha \text{sign}(x)\,|x|^{\alpha-1}.$$

and the gain function of the estimator is,

$$t(\alpha) = E(u'(X))E(X^2) = E(-\alpha\delta(X)\,|X|^{\alpha-1} - \alpha(\alpha-1)\,|X|^{\alpha-2}) =$$

$$= \frac{\Gamma(\frac{3}{\alpha})}{\Gamma^3(\frac{1}{\alpha})}\frac{\alpha(\alpha-1)\pi}{\sin(\frac{\pi}{\alpha})}.$$

APPENDIX C

A small phase shift θ can be applied to an input sequence $\underline{x}$ by convolving $\underline{x}$ with the wavelet $\underline{\delta} + \theta\underline{h}$, where the antisymmetric wavelet $\underline{h}$ is the discrete time domain representation of the Hilbert transform (Claerbout, 1976). A wavelet $\underline{a}$ has the same direction as $\underline{h}$ provided that $[\underline{I} - \underline{H}]\underline{a} = \underline{0}$, where the entries of the matrix $\underline{H}$ are $H_{k,l} = \frac{h_k h_l}{|\underline{h}|^2}$, for $|k|, |l| \leq \frac{M}{2}$ and $|\underline{h}|^2 = 2\sum_{i=1}^{M/2} h_i^2$.