Estimation of residual wavelets

Clement Kostov and Fabio Rocca

ABSTRACT

A solution is presented for the blind deconvolution problem, where the wavelet is estimated from the convolved output data and from statistical assumptions on the input process and the wavelet. The key steps in this method are the computation of the gradient of a likelihood function, and of the gain operator applied to that gradient.

The case of independent identically distributed (iid) input data samples, drawn from generalized Gaussian distributions, is presented in detail. Expressions for the variance of the estimators are derived as a function of the distribution of the input process, of the number of input data points and of the prior information on the wavelet type. Numerical simulations show also the dependence of the variance on the norm of the wavelet.

Despite a very distinct derivation, the method presented in this paper shows interesting similarities with minimum entropy deconvolution (MED) methods. The form of the estimators and the results on their variances are compared for both methods.

INTRODUCTION

Minimum entropy techniques, first applied to seismic deconvolution by Wiggins (1978), attempt to estimate both the spectrum and the phase of the wavelet. In the original method (Wiggins, 1978), the inverse filter is characterized as the one which maximizes the kurtosis of the deconvolved data. The kurtosis is only one particular measure of the non-Gaussianity of the deconvolved data. Later techniques, developed by Gray (1979), Claerbout (1977), Ooe and Ulrych

(1979) and others, investigated the use of different measures of non-Gaussianity. Donoho (1981) provided a theoretical analysis of these approaches. The related work on "zero memory nonlinear deconvolution" by Godfrey and Rocca (1981) emphasized the importance of a nonlinear function which is used iteratively to estimate the unknown input sequence. As shown in a recent, very clear survey paper by Walden (1985), such nonlinear functions offer an alternative definition of MED algorithms, perhaps more informative than the one based on a objective function.

In this paper, we estimate directly a small residual wavelet. Our derivation starts with the comparison, through a first-order Taylor expansion, of the multivariate probability density function (pdf) of the data and the multivariate pdf model of the reflectivities. The first order term of the Taylor expansion gives a gradient direction. A gain operator, to be applied on the gradient, is derived by requiring that the estimator be unbiased. We obtain thus an explicit formula for the estimator, from which we derive the variance of the estimator as a function of the pdf of the input sequence, and of the number of data. The estimation variances depend also on the type of the wavelet, if known a-priori, and comparisons for odd, even, and causal wavelets are given. These results on the variance provide valuable indications to the analyst for the choice of an estimation method.

Numerical results are given for statistically independent input sequences drawn from generalized Gaussian distributions. These results confirm the theoretical derivations and give an indication on the dependence of the variance on the norm of the residual wavelet.

The estimator obtained in this paper is shown to be similar, although not identical to the MED estimators for residual wavelets. The variances of the estimator are also similar to those derived for MED methods.

The method presented in this paper could be extended in two ways: non convolutional operators, such as migration, can be estimated (van Trier, this report) and more complicated models than iid sequences could be considered for the input process.

DERIVATION OF THE ESTIMATOR

The convolutional model

Let us consider a random sequence $\underline{x}$, convolved with a "delta-like" wavelet $\underline{\delta} + \underline{a}$. The output of the convolution is the vector $\underline{y}$. We introduce the following notations:

$$\underline{y} = [\, y_1 , .. , y_N \,], \qquad (N \times 1)$$
$$\underline{x} = [\, x_1 , .. , x_N \,], \qquad (N \times 1)$$
$$\underline{\delta} = [\, 0 , .. ,0, 1 ,0, .. ,0], \qquad (M \times 1)$$
$$\underline{a} = [\, a_{-M/2} , .. ,a_{-1},0,a_1, .. ,a_{M/2}], \qquad ((M+1) \times 1)$$

and the convolutional model:

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

In the sequel, we always assume $a_0 = 0$. To first order in $\underline{a}$ we have also an expression for $\underline{x}$ as a function of $\underline{a}$:

$$\underline{x} = \underline{y} - \underline{a} * \underline{y} = \underline{y} * (\underline{\delta} - \underline{a}). \tag{2}$$

Gradient of the likelihood function

In order to derive maximum likelihood or conditional mean estimates of the wavelet given the data, we need the probability of the wavelet conditional to the data. We apply first Bayes theorem to relate this probability, $p_{\underline{a}}(\underline{a} \mid \underline{y})$, to the probability of the data given the wavelet, $p_{\underline{y}}(\underline{y} \mid \underline{a})$:

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

Then we use the convolutional model, equation (1), in order to introduce the probability distribution of the input sequence $\underline{x}$ in equation (3). Knowledge of this pdf is one of our hypotheses. The change in the pdf corresponding to a change of variables is given by the following general relation:

$$p_{\underline{y}}(\underline{y} \mid \underline{a}) = p_{\underline{x}}(\underline{x} \mid \underline{a}) \left| \frac{\partial \underline{x}}{\partial \underline{y}} \right| = p_{\underline{x}}(\underline{y} - \underline{a} * \underline{y} \mid \underline{a}) \left| \frac{\partial \underline{x}}{\partial \underline{y}} \right| . \tag{4}$$

We show in appendix A that to first order in the wavelet $\underline{a}$, we have

$$p_{\underline{a}}(\underline{a}) = 1 \quad \text{and} \quad \left| \frac{\partial \underline{x}}{\partial \underline{y}} \right| = 1,$$

while $p_{\underline{y}}(\underline{y} \mid \underline{a}) = p_{\underline{x}}(\underline{y})$, so that equation (4) simplifies to:

$$p_{\underline{a}}(\underline{a} \mid \underline{y}) = \frac{p_{\underline{x}}(\underline{y} - \underline{a} * \underline{y} \mid \underline{a})}{p_{\underline{x}}(\underline{y})}. \tag{5}$$

The pdf of the wavelet, conditional to the data, as given by equation (5) is an important result, which can be used to derive both maximum likelihood estimators and conditional mean estimators.

Since we assume a small residual wavelet $\underline{a}$ we can approximate further the pdf in equation (5), by a linear expression in the coefficients of $\underline{a}$. Expanding, $p_{\underline{x}}(\underline{y} - \underline{a} * \underline{y} \mid \underline{a})$ in a Taylor series around $\underline{y}$ leads to:

$$p_{\underline{a}}(\underline{a} \mid \underline{y}) = \left\{ \frac{1}{N} + \frac{1}{N p_{\underline{x}}(\underline{y})} \sum_{i=1}^{N} \frac{\partial p_{\underline{x}}}{\partial x_i}(\underline{y})[-(\underline{y} * \underline{a})_i] \right\} = \tag{6}$$

$$= [\, \frac{1}{N} - \underline{\gamma}^t(\underline{y})\underline{a} \,],$$

where the components of the gradient vector $\underline{\gamma}$, now only to a zero order approximation in $\underline{a}$, are given by:

$$\gamma_k = \frac{1}{N p_{\underline{x}}(\underline{y})} \sum_{i=1}^{N} \frac{\partial p_{\underline{x}}}{\partial x_i}(\underline{y}) y_{i-k} \quad , \text{ for } |\, k \,| \leq \frac{M}{2} \tag{7}$$

For iid input sequences the gradient components simplify further to:

$$\gamma_k = \frac{1}{N} \sum_{i=1}^{N} \frac{p'}{p}(y_k) y_{i-k} \quad , \text{ for } |\, k \,| \leq \frac{M}{2}, \tag{8}$$

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

Derivation of the gain

A variety of optimization algorithms use the gradient direction to find stationary points of an objective function. Newton type algorithms, for instance, approximate the likelihood function with a quadratic form in the wavelet coefficients and require the matrix of the second derivatives (Hessian). Then, the estimator is the product of the inverse of the Hessian times the gradient vector. However, the derivation of the Hessian is computationally tedious and a common approximation consists in replacing the Hessian by its expected value. For the case of generalized gaussian pdf's we show in Appendix B that this approximation is inadequate.

In this paper, we look for an estimator, which has also the form of a matrix (gain matrix) times the gradient vector. We obtain the gain matrix by requiring the estimator to be unbiased for any expected values of the wavelet coefficients.

The estimator $\underline{\hat{a}}$ is written a-priori as:

$$\underline{\hat{a}} = \underline{G}\ \underline{\gamma}(\underline{y}) = \underline{G}\ \underline{\gamma}(\underline{x} + \underline{a} * \underline{x}).$$

The matrix $\underline{G}$ is $M \times M$ and the vectors $\underline{\hat{a}}$ and $\underline{\gamma}$ are $M \times 1$.

Again, linearizing with respect to $\underline{a}$ leads to:

$$\underline{\hat{a}} = \underline{G}\ [\underline{\gamma}(\underline{x}) + \underline{\gamma}'(\underline{x})\underline{a}], \tag{9}$$

where $\underline{\gamma}'(\underline{x})$ is an $M \times M$ matrix, whose expression is given in Appendix C.

Equation (9) should be satisfied also by the mean vectors of the estimator $E(\underline{\hat{a}})$ and of the wavelet $E(\underline{a})$. A second linear relation between these two mean vectors, comes from the unbiasedness condition, which states that they should be equal. These two linear relations are compatible, provided that:

$$\underline{G}\ E[\underline{\gamma}(\underline{x})] = 0 \quad \text{and} \tag{10}$$
$$\underline{G}\ E[\underline{\gamma}'(\underline{x})] = \underline{I}\ .$$

We show in Appendix C, that for a generalized gaussian input sequence:

$$E[\underline{\gamma}(\underline{x})] = -\underline{\delta}\ , \quad \text{and} \tag{11}$$
$$E[\underline{\gamma}'(\underline{x})] = t(\alpha)\underline{I} - \underline{I}'\ ,$$

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

The gain matrix $\underline{G}$ is then $\underline{G} = \frac{1}{t(\alpha)^2 - 1}[t(\alpha)\underline{I} - \underline{I}']$, and the optimal estimator of the residual wavelet is:

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

A minor technical comment on the derivation of $\underline{G}$ is that zero-th row and column of $\underline{G}$ should be set to zero since $E[\gamma_0(\underline{x})] = -1$.

From equation (12) we can derive estimates of the even part of the wavelet, $\underline{a}^{e}$ and $\underline{a}^{odd}$,

$$\underline{a}^{e} = 1/2[\underline{I} + \underline{I}']\ \underline{\hat{a}} = \frac{1/2}{t(\alpha) + 1}[\underline{I} + \underline{I}']\underline{\gamma} = \frac{1/2}{t(\alpha) + 1}\underline{\gamma}^{e}\ , \tag{13}$$

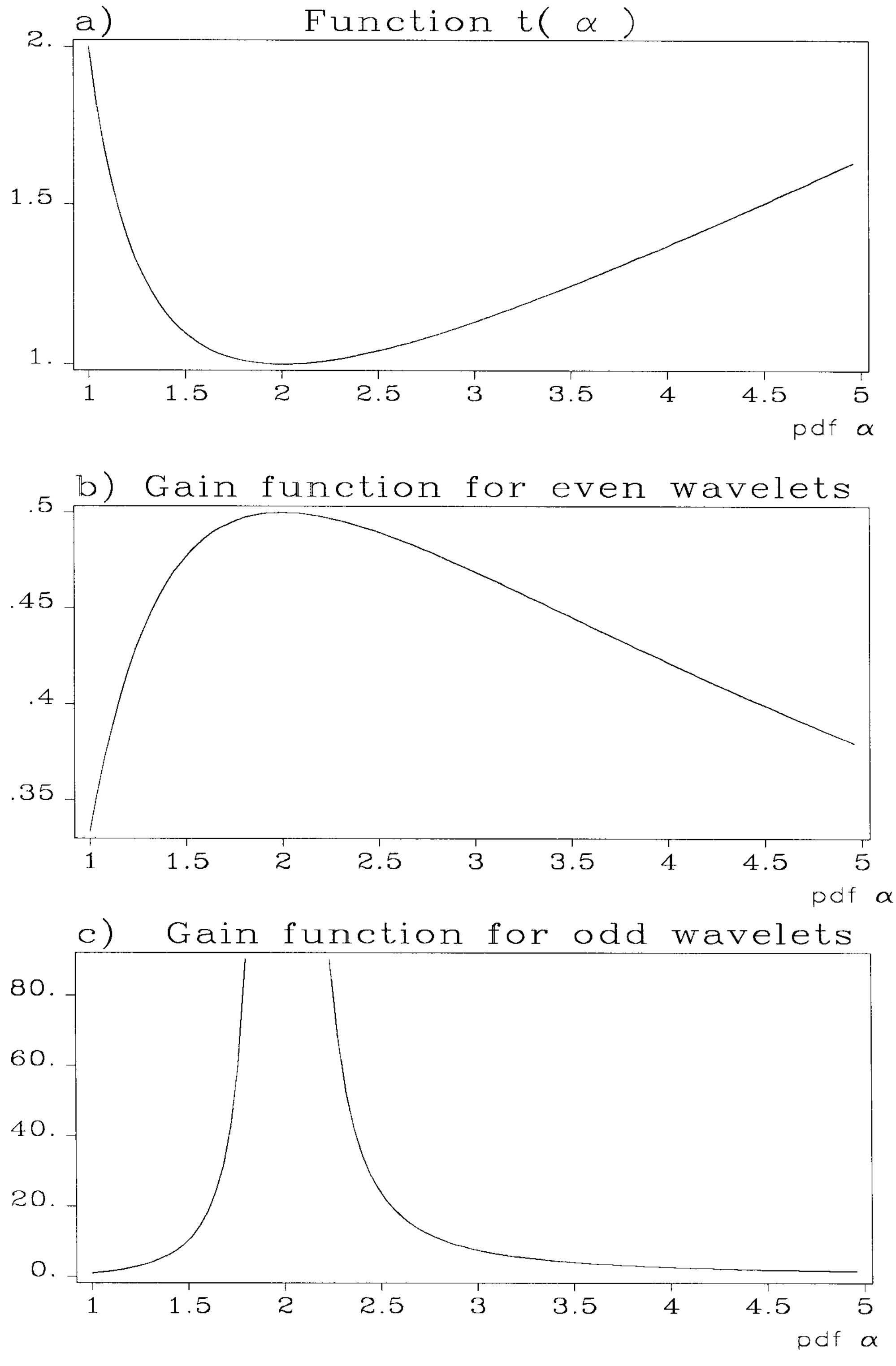

FIG. 1. Functions used to defined the gain factors for different types of wavelets. (a) arbitrary wavelet. (b) even wavelet. (c) odd wavelet.

and

$$\underline{a}^{odd} = 1/2[\underline{I} - \underline{I}'\ ]\underline{\hat{a}} = \frac{1/2}{t(\alpha) - 1}[\underline{I} - \underline{I}'\ ]\underline{\gamma} = \frac{1/2}{t(\alpha) - 1}\underline{\gamma}^{odd}, \tag{14}$$

where $\underline{\gamma}^e$ and $\underline{\gamma}^{odd}$ are the even and odd parts of the gradient.

Notice that the full wavelet $(\underline{\delta} + \underline{a})$ is approximately zero phase for an even residual wavelet and pure phase for an odd residual wavelet.

Figure 1 shows plots of the gain factors of the estimators for arbitrary (equation (12)), even (equation (13)), or odd (equation (14)) wavelets. The function $t(\alpha)$ which appears in the definition of these gain factors, is plotted in Figure 1.a for $1 \leq \alpha \leq 5$, where α is the shape parameter of the generalized gaussian distribution. For gaussian distributions, $\alpha = 2$, and $t(2) = 1$, so that the estimators for arbitrary and for odd wavelets (Figure 1.b) are not defined, since the denominators in equations (12) and (14) are zero. This result was expected, since phase information cannot be inferred from the output data, when the input is gaussian. For even wavelets, there is no phase ambiguity and the estimator (equation (13) and Figure 1.c) is well behaved for all pdf's.

VARIANCE OF THE ESTIMATOR

In equation (12), the estimator is defined as a function of the data vector $\underline{y}$. Such an expression is suitable for computations, but not for theoretical analysis of the estimator, because the pdf of $\underline{y}$ is unknown. Therefore we approximate $\underline{\gamma}(\underline{y})$ to order zero in $\underline{a}$ as $\underline{\gamma}(\underline{x})$. The expression of the estimator becomes:

$$\underline{\hat{a}} = \frac{1}{t(\alpha)^2 - 1}\ [t(\alpha)\underline{I} - \underline{I}'\ ]\ \underline{\gamma}(\underline{x}). \tag{15}$$

The $M \times M$ covariance matrix of the estimator $\underline{\Sigma}$ is defined as:

$$\underline{\Sigma} = E\ [\underline{\hat{a}}\underline{\hat{a}}^T].$$

We substitute $\underline{\hat{a}}$, as given by equation (15) in the above definition of the variance, and use the following expression derived in Appendix C for the covariance of the gradient,

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

in order to obtain the covariance matrix of the estimator:

$$\underline{\Sigma} = \frac{1}{N}\frac{[\ t(\alpha)\underline{I} - \underline{I}'\ ]}{t(\alpha)^2 - 1} \tag{16}$$

This expression for the covariance matrix highlights several important properties of the estimator. Basically, it gives the asymptotic behavior of the estimator as function of the number of data and the pdf of the input sequence. We notice that the factor $\frac{1}{N}$ ensures that the covariance tends to zero as the number of data goes to infinity. This property, a prerequisite for any useful estimator, is called "consistency" in the statistical terminology. Before discussing in detail the influence of the pdf of the input and of the prior information on the wavelet, let us mention what our formula (16) does not tell about the variance.

Equation (16) was derived on the basis of first order Taylor expansions in the vector $\underline{a}$, therefore we also expect also a polynomial approximation for the covariance, depending on terms such as $E(\underline{a})$ and $E[\underline{a}\ \underline{a}^T]$. The first order term in $E(\underline{a})$ vanishes because of the symmetry of the wavelet's pdf. The second order term gives the dependency of the variance on the magnitude of the residual wavelet. However, it requires also a second order approximation of the gradient and therefore considerably more algebra. We will gain insight into the effect of the norm of the residual wavelet from numerical simulations and from the comparison with MED methods.

The covariance matrices for even and odd wavelets can be obtained in a similar way starting from equations (13) and (14). The covariance matrix $\underline{\Sigma}^e$ is equal to:

$$\underline{\Sigma}^e = \frac{1}{2N}\left[\frac{I}{t(\alpha)+1} + \frac{I'}{(t(\alpha)+1)^2} \right], \tag{17}$$

while the covariance matrix for an odd wavelet is:

$$\underline{\Sigma}^{odd} = \frac{1}{2N}\left[\frac{I}{t(\alpha)-1} + \frac{I'}{(t(\alpha)-1)^2} \right]. \tag{18}$$

As a last example, we obtain the variance on the causal part of a wavelet. In that case, we know that $\hat{a}_{-k} = 0$ which yields (after equation (12)) a particular relation between the elements of the gradient vector,

$$\hat{a}_{-k} = 0 \quad => \quad t(\alpha)\gamma_{-k} - \gamma_k = 0 ,$$

so that $\hat{a}_k = \frac{1}{t(\alpha)}\gamma_k$, and:

$$var(\hat{a}_k) = \frac{1}{t(\alpha)^2} E(\gamma_k{}^2) = \frac{1}{Nt(\alpha)}. \tag{19}$$

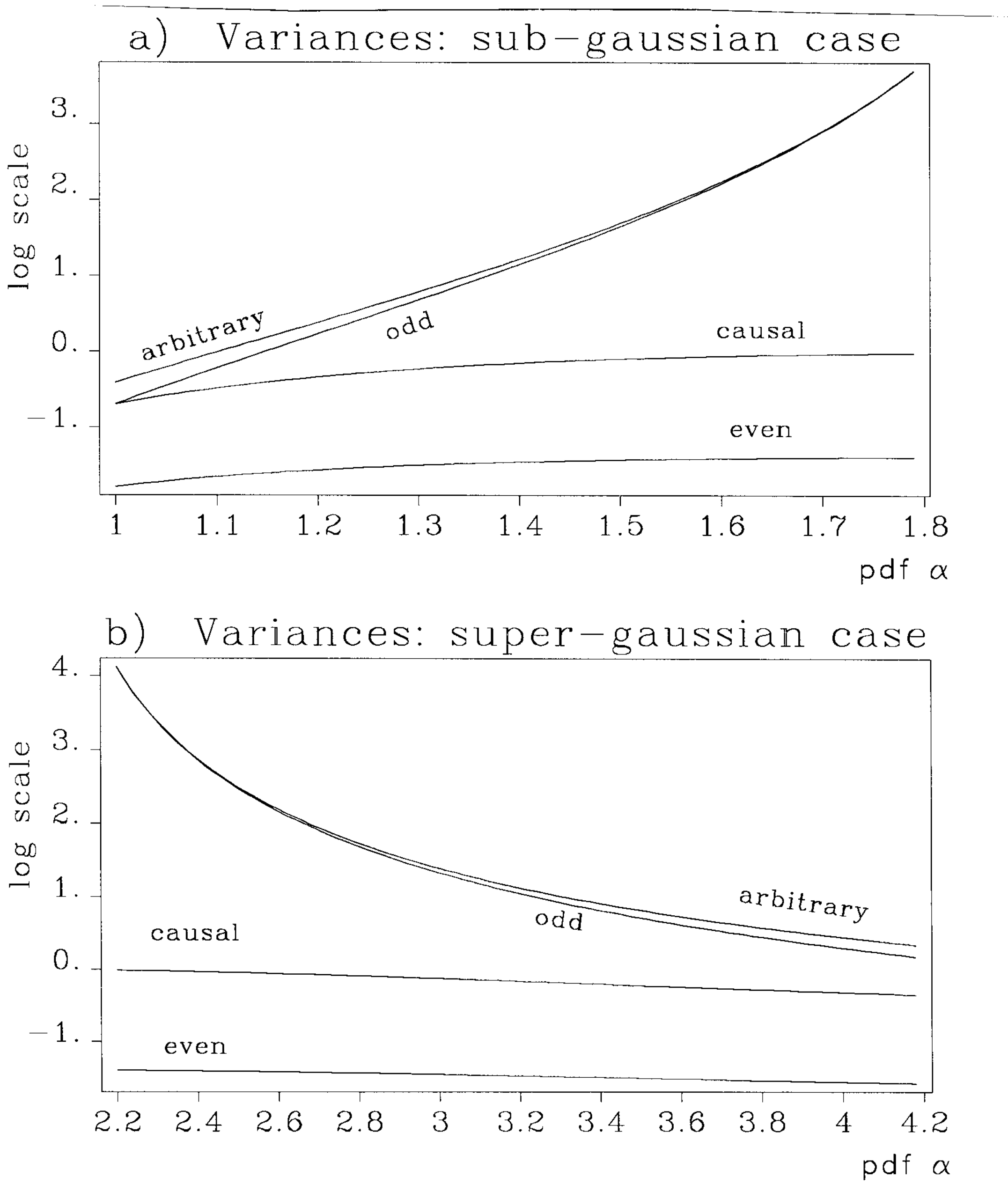

FIG. 2. Variances of the estimated wavelets as function of input pdf and prior information, for fixed numbers of data and scale of the residual wavelet. (a) sub-gaussian case. (b) super-gaussian case.

Figures 2.a and 2.b 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 sub-gaussian distributions for which $\alpha < 2$ and super-gaussian distributions for which $\alpha > 2$, because the variances for arbitrary and odd wavelets tend to infinity as the distribution tends toward the gaussian.

As expected, 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, that is also for the estimates respectively of the phase and of the spectrum of the wavelet, show a difference of several orders of magnitude. The variance on the odd part is practically identical to the total variance.

For small residual wavelets, the full "delta-like" wavelet is invertible, therefore the causal wavelet is also a minimum-phase wavelet, and the estimation variance is well defined for all input pdf's. Notice that the estimation variance for the gaussian distribution is also the highest in the cases of even and causal wavelets, even though at the scale of the plot these variances seem to depend little on the input distribution.

A final remark is that for shape parameters α less than one, the estimation of an odd wavelet becomes more reliable than the estimation of a minimum phase wavelet (compare equations (18) and (19) and notice the trend on Figure 2.a). This property is shared also by MED estimators which outperform least-squares estimation of minimum phase wavelets for "sufficiently heavy tailed" distributions (Walden, 1985).

NUMERICAL EXAMPLES

The numerical tests illustrate in another way how the estimation variance depends on the input pdf and on the type of the wavelet. Numerically we can also simulate the estimation variance as a function of the magnitude of the residual wavelet, a result not given by our theoretical analysis.

Figure 3. shows the wavelets, an even and an odd one, used in the synthetic examples. To obtain residual wavelets, we scale the wavelet coefficients in Figures 3a and 3b by a constant factor. This scale factor is used also as a measure of the magnitude of the residual wavelet. The full wavelet is the sum of the residual wavelet, and a spike, such that its coefficient at time zero is equal to one.

Figures 4a,4b and 4c compare the theoretical variances to the experimental ones for a fixe number of data points, (8 sequences of 8000 points), and fixed scales of the residual wavelets. The scale factor applied to the even wavelet is 10%, and the scale factor applied to the odd one is 20%. Thus for each coefficient of the wavelet there are 8 estimates from which a variance is derived. There are 32 coefficients for the even or odd wavelets, and the plots show the averages of the variances for these coefficients. There was no apparent trend in the variances depending on the coefficient index.

For sub-gaussian sequences the agreement between the experimental and the theoretical curves is excellent both for the even and the odd wavelets (Figures 4a and 4b). As the distribution tends toward gaussian, both the numerator and the denominator of the estimator (equation (12)) tend to zero and therefore the estimation procedure becomes numerically unstable. The function $t(\alpha)$, which appears in the definition of the gain (equation (16)), is asymmetric around $\alpha = 2$, with steeper slopes for $\alpha \leq 2$. This accounts for the lower variances for sub-gaussian estimators observed both on the theoretical and experimental curves, and could at least partially explain the mismatch between the theoretical and experimental curves on Figure 4c. Other reasons for the mismatch could be that the random number generator performs better for sub-gaussian distributions than for super-gaussian, or that the computation of the gradient, for which the data are raised to the power $\alpha - 1$, becomes too sensitive to outliers, when α exceeds 2.

Figures 5a and 5b show how the estimation variance increases as the magnitude of the residual wavelet increases. The scale factors applied to the wavelets shown in Figure 3 vary from 5% to 50%. The numbers of data points, 8×8000 and the pdf of the input data $\alpha = 1.1$ are kept constant. The theoretical estimates shown are the expressions, (independent from the norm of the wavelet), derived in equation (17) for the even wavelet and in equation (18) for the odd wavelet. The agreement between theoretical and experimental data is good up to around 15% for the even wavelet, and up to 30% for the odd wavelet. Afterwards the increase in the curves seems stronger than linear as would be expected from the influence of a quadratic term in the wavelet coefficients.

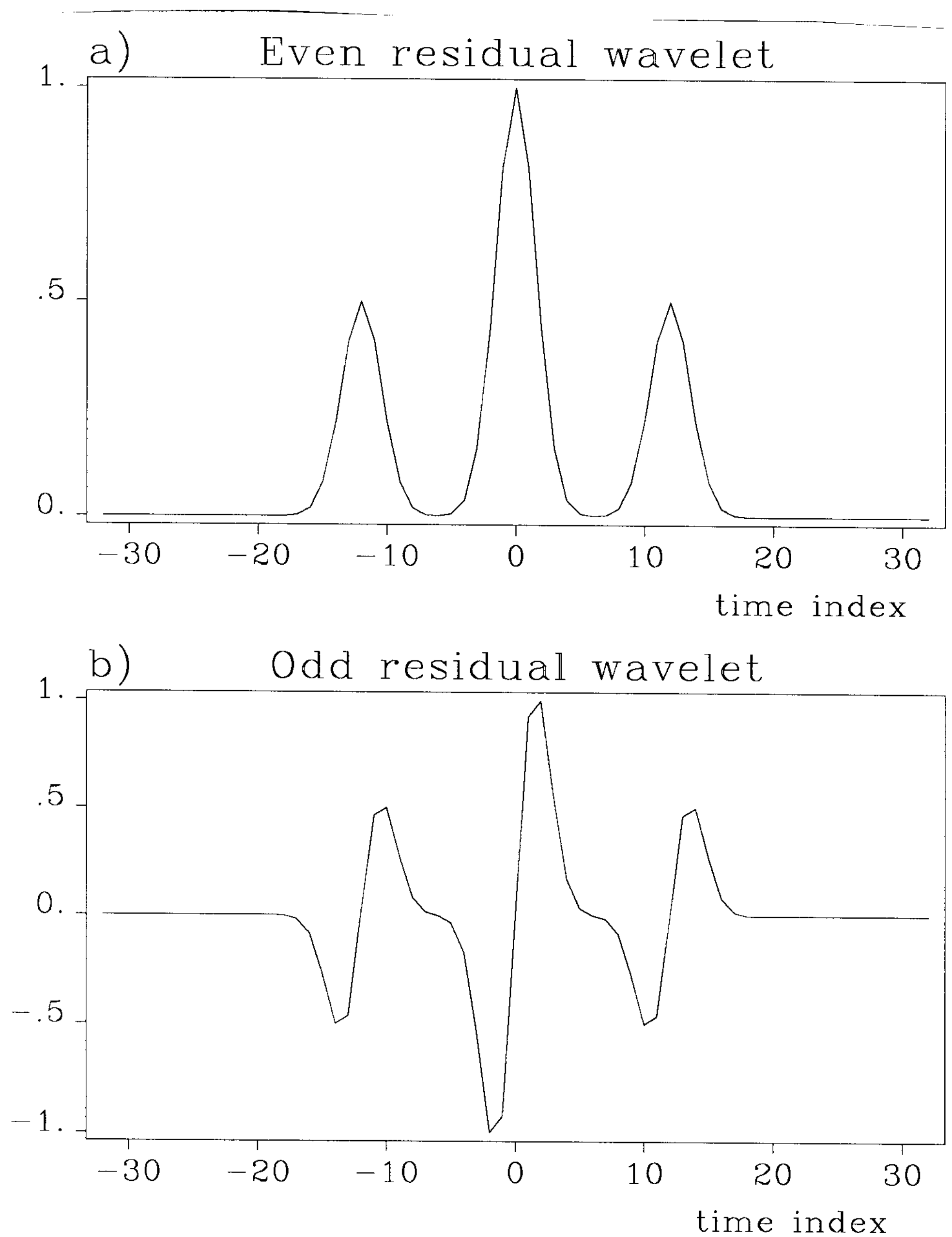

FIG. 3. Residual wavelets, scale 100%. (a) even wavelet. (b) odd wavelet.

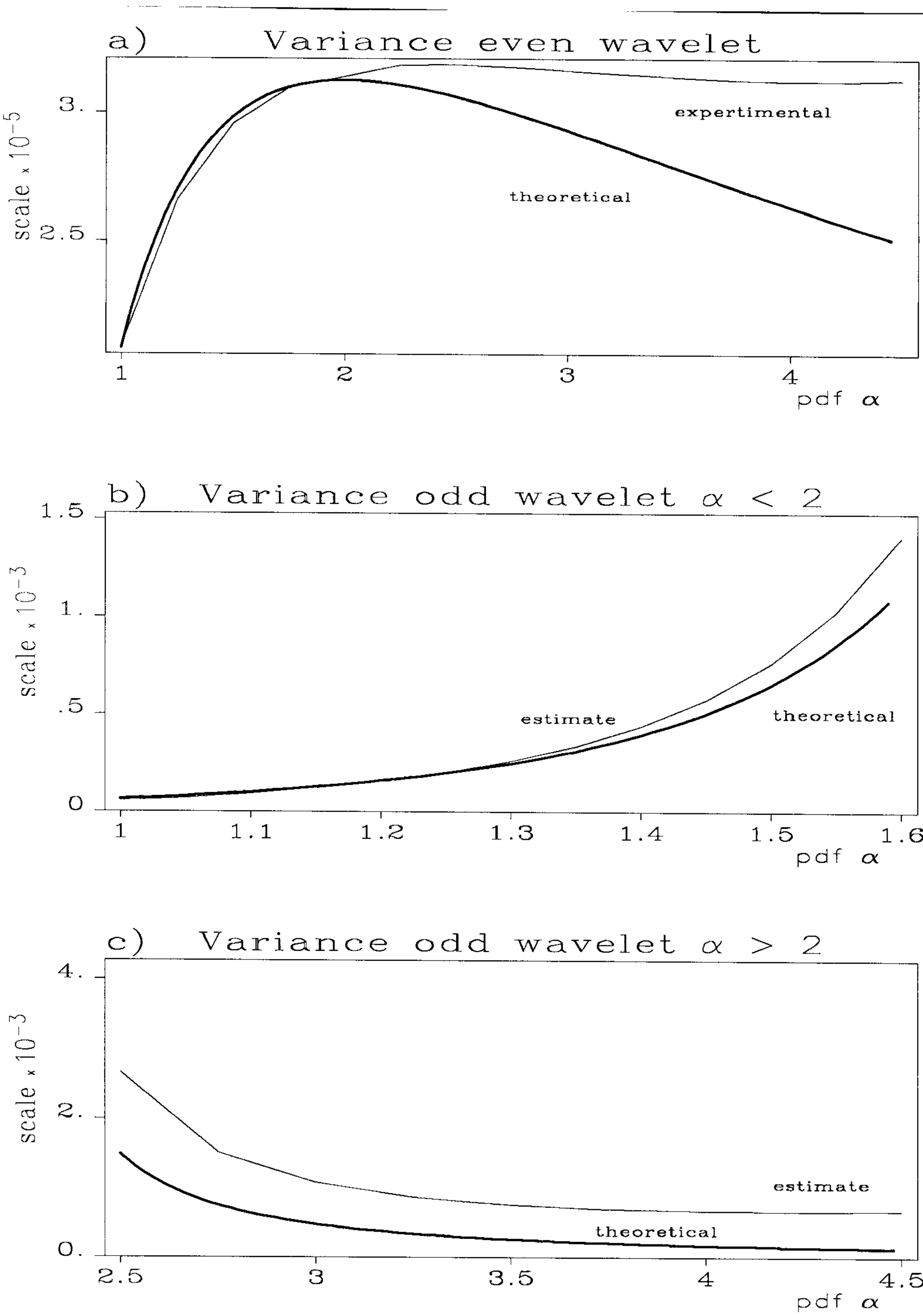

FIG. 4. Comparison of theoretical versus experimental variances. Fixed number of data and scale of the residual wavelet, variable input pdf. (a) even wavelet. (b) odd wavelet, sub-gaussian case. (c) odd wavelet, super-gaussian case.

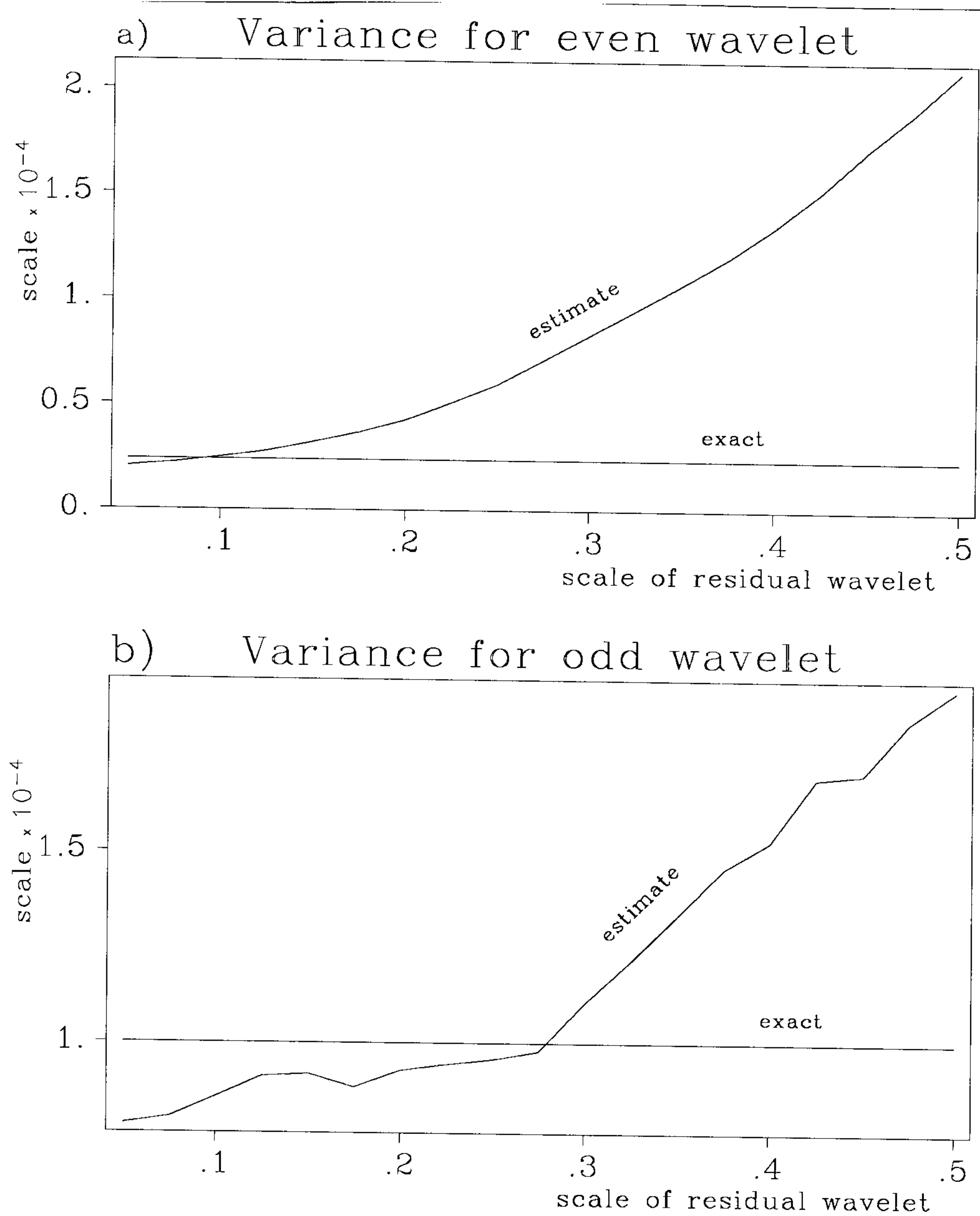

FIG. 5. Variance of estimated wavelets for fixed number of data and input pdf, variable scale of the wavelet. (a) even wavelet. (b) odd wavelet.

COMPARISON WITH MED METHODS

Several features of our method appear familiar from the study of MED algorithms. As in MED algorithms we start by analyzing a change in the probability distribution introduced by a convolution and try afterwards to find the appropriate filter to remove it. We find also, as in MED methods, that the gradient direction for our algorithm is given by the crosscorrelation of the data with a nonlinear function of the input sequence. Our variance estimates, valid to first order in the wavelet, are obtained as the product of three factors (1) the inverse of the number of data, (2) the variance of the input sequence, (3) a term depending only on the pdf of the input. Exactly the same type of result holds for the asymptotic covariance matrix of MED estimators, and more generally for M estimators, as pointed out by Walden (1985).

The major difference between our approach and MED methods is that we look at changes in the joint probability distribution of the data, while MED compares only univariate statistics. Still, how different are the estimators derived by each method?

We attempt to answer this question quantitatively by analyzing a MED type estimator for small residual wavelets. Walden (1985) indicates that the stationary points for several MED objective functions, including those of Wiggins (1978), Claerbout (1977), Ooe and Ulrych (1979), are solutions of a nonlinear Toeplitz systems of the form:

$$\sum_{k,l} y_{k-l}\, y_{k-j}\, b_l = \sum_k (u(x_k) - y_k)\, y_{k-j}, \tag{20}$$

where $-\frac{M}{2} \leq j \leq \frac{M}{2}$, $j \neq 0$, and $k \neq 0$. As previously $\underline{y}$ are the data, $\underline{x}$ the input, and $b_k = -a_k$, is the residual wavelet associated with the inverse filter.

We approximate again the correlation matrix of the data by its expected value, that is for an iid input sequence with unit variance, by the identity matrix. Furthermore we express x_k in equation (20) in terms of $\underline{y}$ and expand $u(x_k)$ in a Taylor series around y_k,

$$b_j = \sum_k u(x_k) y_{k-j} = \sum_k u(y_k) y_{k-j} + \sum_{k,n} y_{k-n}\, y_{k-j}\, u'(y_k) b_n,$$

where b_0 is set to zero.

The term $\sum_{k,n} y_{k-n}\, y_{k-j}\, u'(y_k)$ is a triple correlation product, such that only the expected value of the terms of the form $\sum_k y^2_{k-j}\, u'(y_k)$ is non zero.

The MED estimator appears to be proportional to the gradient γ_j,

$$b_j = \frac{1}{1 - E(u'(x))} \gamma_j \ , \ \text{where} \tag{21}$$

$$\gamma_j = \sum_k y_{k-j} (u(y_k) - y_k)$$

In view of equation (21) we conclude that there are significant differences between the MED estimator and the one proposed in this paper. The gain on the gradient in the case of the MED estimator, at least as applied at each iteration, is proportional to the identity matrix, while the gain applied in equation (12) depends also on the time reverse of the identity matrix ($\underline{I}'$). The effect of the time reversed identity matrix $\underline{I}'$ is important, because it allows to obtain different gain factors and hence different variances on the estimates of on the even and odd parts of the wavelet. On the other hand, the MED filter is obtained iteratively, in which case the gradient direction alone, without the proper gain which might still be sufficient for reaching a stationary point.

Graphical examples of nonlinearities

Figure 6a shows the nonlinearity used in Wiggins' MED method. This nonlinearity is proportional to a cuber $y = u(x) = \lambda x^3$, with a scale factor λ depending on the distribution of the input. Imagine applying this nonlinear function repeatedly on the data. Mathematically, we look at the sequence of points defined by the recursion $x_{n+1} = u(x_n)$, and study the limit points depending on the starting point x_0. Graphically following the arrows on the figures tracks the movement from one iterate to the other, and ultimately leads to a point of convergence. The nonlinear function attenuates small values (small reflection coefficients and noise) and amplifies large ones, its effect on a one dimensional histogram being to increase the spike around zero as well as the tails of the of the histogram. The 1d nonlinearity plotted in Figure 6b refers to the deconvolution method of Ooe and Ulrych. Their exponential transform also attenuates small values, but contrary to the cuber, hardly modifies large ones.

Figures 6c and 6d illustrate the effect of the nonlinearities on pairs of neighboring points. Pairs of neighboring points are also points of a 2d histogram (Ottolini and Rocca, this report), on which bivariate statistical properties such as correlation, independence and phase rotation can be graphically interpreted.

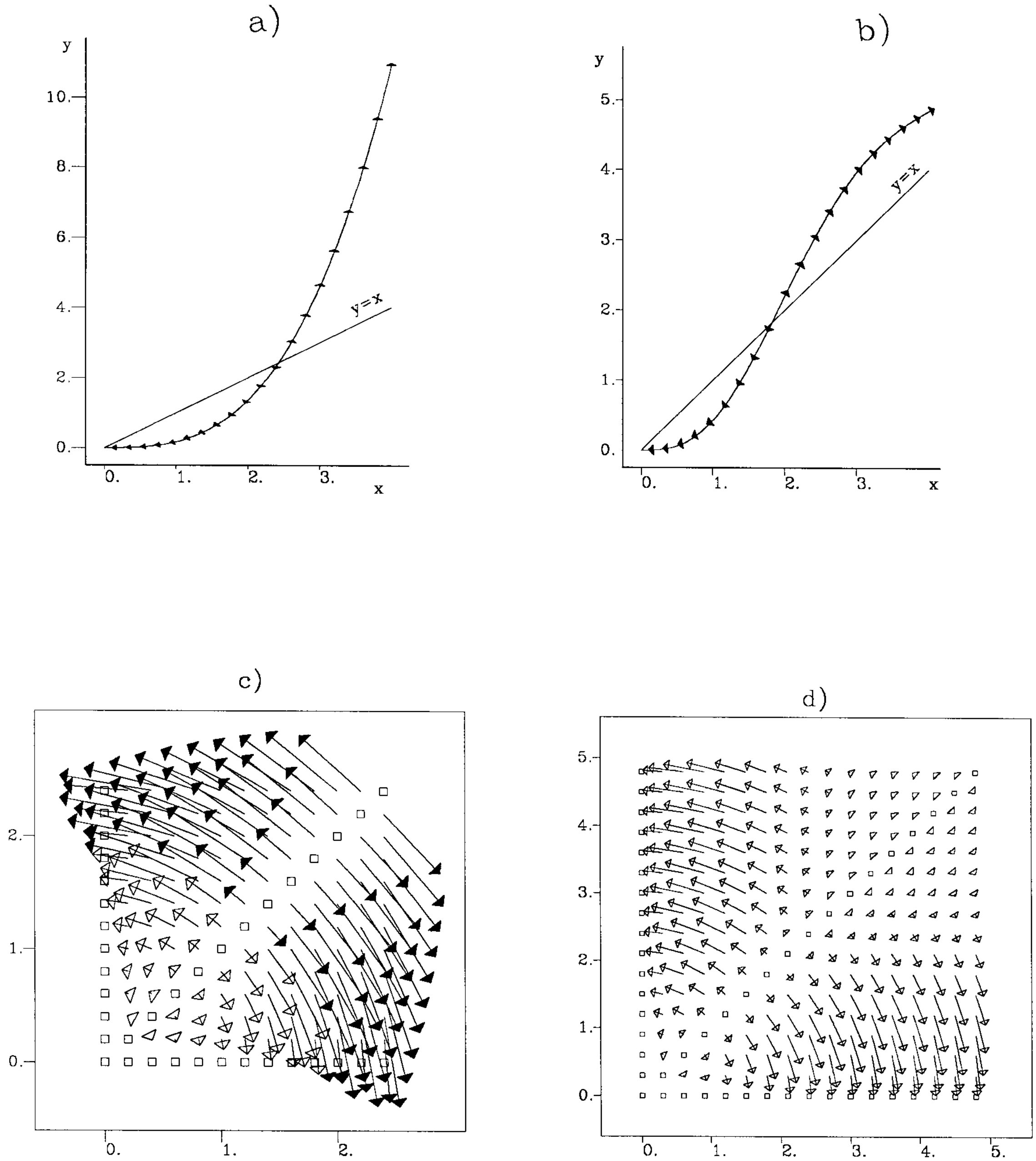

FIG. 6. 1d and 2d nonlinear functions and their effect on histograms. Arrows point toward the image of a point by the nonlinear function. Black arrowheads indicate divergence after a few iterations. Diamonds indicate no displacement. (a) 1d nonlinearity for Wiggins' decon. (b) 1d nonlinearity for Ooe's decon. (c) 2d nonlinearity for Wiggins' decon. (d) 2d nonlinearity for Ooe's decon.

We consider a simple convolutional model, where we try to solve for the rotation angle of a small pure phase shift operator. The 1d nonlinearity maps a pair (X, Y) into a pair $(u(X), u(Y))$. The sine of the angle between these two vectors, which are of equal length, since the transformation is approximately a rotation, is given by:

$$\theta \approx \sin(\theta) = \frac{[Xu(Y) - Yu(X)]}{X^2 + Y^2} \tag{22}$$

At each point of the plots 6c and 6d an arrow is drawn in the radial direction. The magnitude of the arrow is proportional (up to a clip) to the rotation angle as given by equation (22). Points which converge after a few iterations of these rotation are labeled with white arrowheads, points which do not move at all are labeled as diamonds, while black arrowheads indicate divergence. Regions of divergence for Wiggins' nonlinearity are clearly delineated, while Ooe and Ulrych's method is stable for all points. Phase rotations appear to modify the shape of the 2d histogram by decreasing the density of points along the line of unit slope. This is achieved essentially by driving the smaller element of a pair to zero. On the one dimensional histogram, we noticed also such a clipping effect. The difference between the 1d and 2d cases is mainly that the clip is the same for all points in the 1d case, while in the 2d case it is determined from the neighboring points.

The previous discussion gives only some preliminary examples of possible approaches to a simple graphical analysis of the properties of MED estimators. The concept of a nonlinearity characteristic of a MED method seems particularly interesting in this context.

CONCLUSION

We have obtained estimates of the wavelet phases under rather general assumptions about the statistics of the input data. The method of analysis is also an original approach which does not require sophisticated computations. The results could be applied immediately in cases like SAR (Synthetic Aperture Radar), where there are many data. The applications to seismic problems are less straightforward, even if they do not appear as a hopeless case.

REFERENCES

Claerbout, J.F, 1977, Parsimonious deconvolution: Stanford Exploration Project, 13, 1-9.

Donoho, D., 1981, On minimum entropy deconvolution, *in* Findley, D. Ed., Applied time series analysis II: Academic Press Inc.

Hosken, J.W.J., 1986, Maximum kurtosis phase and the phase of the seismic wavelet: Presented at SEG/EAEG Workshop on Deconvolution and Inversion, Rome.

Gray, W., 1979, Variable norm deconvolution: Ph.D. dissertation, Stanford University, also SEP-19.

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

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

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

Ottolini, R., 1986, Movies of joint density histograms: Stanford Exploration Project, 50.

van Trier, J., 1986, Autofocusing of migrated data: Stanford Exploration Project, 50.

Wiggins, R.A., 1977, Minimum entropy deconvolution: Presented at the 39th Ann. Mtg Europ. Ass'n. Explor. Geophys.

APPENDIX A

In this appendix we derive the three formulas needed for the derivation of equation (4) in the text. Although the formulas in the text require only first order approximations in $\underline{a}$, we give the corresponding formulas up to second order, since they could be used to derive other types of estimators.

Approximation of $p_{\underline{a}}(\underline{a})$

For a wavelet with zero mean vector $E(\underline{a})$ the linear terms in the Taylor expansion around zero vanish. The zero-th order term, which is the value of the pdf at zero, is equal to 1. Specifically, for a Gaussian distribution on the wavelet, we have:

$$p_{\underline{a}}(\underline{a})(\det \underline{W})^{\frac{1}{2}}(2\pi)^{\frac{M}{2}} = e^{[\frac{-1}{2}\underline{a}^t \underline{W}^{-1}\underline{a}]} = 1 + \frac{-1}{2}\underline{a}^t \underline{W}^{-1}\underline{a}$$

Approximation of the jacobian

Notice that $|\frac{\partial \underline{x}}{\partial \underline{y}}|^{-1} = |\frac{\partial \underline{y}}{\partial \underline{x}}|$ and that the convolutional equation $\underline{y} = \underline{x} * (\underline{\delta} + \underline{a})$ can be written as a matrix product in the following way:

$$\underline{y} = [\underline{I}(N) + \underline{A}]\underline{x},$$

where $\underline{I}(N)$ is the $(N \times N)$ identity matrix and $\underline{A}$ is a $(N \times N)$ matrix with the wavelet $\underline{a}$ along its diagonal.

Let λ_i, for $1 \leq i \leq N$, be the eigenvalues of $\underline{A}$, then:

$$\det(\underline{I}(N) + \underline{A}) = \prod_{i=1}^{N}(1+\lambda_i) = 1 + \sum_{i=1}^{N}\lambda_i + \sum_{i=1, j>i}^{N}\lambda_i\lambda_j + ..$$

Since the sum of the eigenvalues of a matrix is equal to the trace of the matrix, the following relations hold:

$$\sum_{i=1}^{N}\lambda_i = Tr[\underline{A}]$$

$$(\sum_{i=1}^{N}\lambda_i)^2 = Tr^2[\underline{A}] \quad \text{,and} \quad \sum_{i=1}^{N}\lambda_i{}^2 = Tr[\underline{A}]^2,$$

and are used to derive the identity:

$$\sum_{i=1, j>i}^{N}\lambda_i\lambda_j = \frac{1}{2}((\sum_{i=1}^{N}\lambda_i)^2 - \sum_{i=1}^{N}\lambda_i{}^2) = \frac{Tr^2[\underline{A}] + Tr[\underline{A}]^2}{2}$$

Then, in terms of the components of the wavelet:

$$Tr\,[\underline{A}\,] = Na_0 = 0, \quad \text{and} \quad Tr\,[\underline{A}\,]^2 = Na_{2,0} = N \sum_{k=-\frac{M}{2}}^{\frac{M}{2}} a_k\, a_{-k} = N\underline{a}^t\, \underline{I}'\ \underline{a}\,,$$

where $\underline{I}'$ is the $[(M+1)\times(M+1)]$ matrix with ones along the secundary diagonal and zeroes elsewhere. Finally,

$$\det(\underline{I}(N) + \underline{A}\,)^{-1} \approx (1 - \frac{1}{2} Tr\,[\underline{A}\,]^2) \approx 1 + Tr\,[\underline{A}\,]^2 = 1 + \frac{N}{2}\underline{a}^t\, \underline{I}'\ \underline{a}. \quad (A.1)$$

Approximation of $p_{\underline{y}}\,(\underline{y}\,|\,\underline{a}\,)$

Expanding $(\underline{\delta} + \underline{a}\,)^{-1}$ in a Taylor series up to second order in $\underline{a}$ we obtain:

$$\underline{x}\,(\underline{y}\,,\underline{a}\,) = \underline{y} - \underline{a}*\underline{y} + \underline{a}*\underline{a}*\underline{y} = \underline{y} - \underline{a}*\underline{y} + \underline{a}_2*\underline{y}\ , \quad (A2)$$

where $\underline{a}_2 = \underline{a}*\underline{a}$. Substituting $\underline{x}$ in terms of $\underline{y}$ and expanding $p_{\underline{x}}\,(\underline{x}\,|\,\underline{a}\,)$ in a Taylor series around $\underline{y}$ leads to:

$$p_{\underline{y}}\,(\underline{y}\,|\,\underline{a}\,) = \Bigg\{ p_{\underline{x}}\,(\underline{y}\,) + \sum_{i=1}^{N} \frac{\partial p_{\underline{x}}}{\partial x_i}(\underline{y}\,)[-(\underline{y}*\underline{a}\,)_i + (\underline{y}*\underline{a}_2)_i\,] + \quad (A.3)$$

$$\frac{1}{2} \sum_{i,j=1}^{N} \frac{\partial^2 p_{\underline{x}}}{\partial x_i\, \partial x_j}(\underline{y}\,)[(\underline{y}*\underline{a}\,)_i\, (\underline{y}*\underline{a}\,)_j\,] \Bigg\} [1 + \frac{N}{2}\underline{a}^t\, \underline{I}'\ \underline{a}\,]$$

The previous result gives the zero-th order approximation required in the text.

APPENDIX B

After equation (A.3), we write the expected value of the Hessian T :

$$E\,[T\,] = Np_{\underline{x}}\,(\underline{y}\,)\sum_{k,h} a_k\, a_h\ \ [\underline{I}'\ E\,[\frac{p'\,(x\,)}{p\,(x\,)}x\,) + \quad (A.4)$$

$$1/2E\,(\frac{p''}{p}(x\,))E\,(x^2) + 1/2E\,[\,(\frac{p'}{p}(x\,))^2]]$$

The expressions for the mean values in equation (A.4) are computed in Appendix C. The sum of these three mean values is found to be $-1/2\underline{I}'$, but another second order term in $\underline{a}$, multiplied by $1/2\underline{I}'$ comes from the jacobian (equation (A.1)). So the only second order term in $\underline{a}$ obtained by this approach comes from $p_{\underline{a}}(\underline{a})$ and depends only on the prior information of the wavelet, which is not very interesting.

APPENDIX C

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

$$p_x(x) = \frac{1}{2\beta\Gamma(\frac{1}{\alpha})} e^{[-|\frac{x}{\beta}|^{\alpha}]} . \tag{A.5}$$

The following identities expressed in terms of the gamma function Γ are helpful when working with generalized Gaussian pdf's:

$$E(x^{\lambda}) = \frac{\Gamma[\frac{(\lambda+1)}{\alpha}]}{\Gamma(\frac{1}{\alpha})}, \tag{A.6}$$

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin(\pi z)} \quad \text{and} \quad \Gamma(z+1) = z\Gamma(z). \tag{A.7}$$

From the above relations we derive,

$$E(\frac{p''}{p}$$

$$E(\frac{p'(x)}{p(x)} x) = \gamma_0 = -1$$

and finally

$$E[(\frac{p'}{p})^2 x^2] = t(\alpha)$$

so that the covariance of the gradient components can be derived:

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