Information-theoretic wavelet noise removal for inverse elastic wave scattering theory

Alan J. Van Nevel
U.S. Naval Air Warfare Center, Weapons Division, China Lake, California 93555-6100

Brian DeFacio
Department of Physics and Astronomy, Missouri University, Columbia, Missouri 65211

Steven P. Neal
Department of Mechanical and Aerospace Engineering, Missouri University, Columbia, Missouri 65211
(Received 30 June 1997; revised manuscript received 10 December 1998)

A discussion of noise removal in ultrasound (elastic wave) scattering for nondestructive evaluation is given. The methods used in this paper include a useful suboptimal Wiener filter, information theory and orthonormal wavelets. The multiresolution analysis (MRA), due to Mallat, is the key wavelet feature used here. Whereas Fourier transforms have a translational symmetry, wavelets have a dilation or affine symmetry which consists of the semi-direct product of a translation with a change of scale of the variable. The MRA describes the scale change features of orthonormal wavelet families. First, an empirical method of noise removal from scattered elastic waves using wavelets is shown to markedly improve the l^1 and l^2 error norms. This suggests that the wavelet scale can act as dial to ''tune out'' noise. Maximization of the Kullback-Liebler information is also shown to provide a scale-dependent noise removal technique that supports (but does not prove) the intuition that certain small energy coefficients that are retained contain large information content. The wavelet MRA thereby locates ''islands of information'' in the phase space of the signal. It is conjectured that this method holds more generally. [S1063-651X(99)14403-9]

PACS number(s): 41.20.Jb, 11.80.−m, 62.30.+d, 89.70.+c

I. INTRODUCTION

The problem of understanding and removing noise from measured data [1–7] is important and has been extensively studied. Two areas that could benefit greatly from advances in noise removal are nondestructive evaluation [8–11] and medical diagnostics [12–15]. Wahba [1,2] has published fundamental works on noise removal using spline methods. It is unfortunate that experimental and theoretical physicists have overlooked these references. Coifman and Wickerhauser [3] and Wickerhauser [4] have applied wavelet methods together with Shannon information to the problem of noise removal. Donoho [5] developed a soft thresholding approach also based on wavelets and the smoothness of the signal. The soft-thresholding of n data was implemented by subtracting an amount $\sigma\sqrt{2\ln(n)/n}$ from each wavelet coefficient c_{jk}. The quantity σ, is the variance of the noise and this method treated each scale j identically. The authors [6] treated the wavelet scales empirically and found that two to three scales should retain all coefficients whereas all coefficients $|c_{jk}| < \varepsilon \max(|c_{jk}|)$, where $\varepsilon(0,1)$ is a threshold, should be set equal to zero. These scales which minimized the error norms were called ''exceptional scales'' j_e while all other scales were referred to as ''ordinary scales'' j_0. Later, Van Nevel [7] showed that the Kullback-Liebler information chooses approximately the same scales as the empirical method. This aspect of noise removal will be presented in this paper.

This work combines a diverse group of topics, such as suboptimal Wiener filtering, multiresolution analysis, information theory, and orthonormal wavelets. This synthesis has resulted in a physical basis for our noise removal algorithm, which often is lacking in many other ''one size fits all'' statistical approaches. We conjecture that the method presented here will hold more generally.

The organization of this paper is as follows: In Sec. II the formulation of the elastic wave (ultrasound) nondestructive evaluation applied inverse problem will be presented. In Sec. II A the data and the Wiener filter regularizer are given, with two figures. The Kullback-Liebler cross entropy and other information theoretic considerations are developed in Sec. II B. Section II C contains a short discussion of the multiresolution analysis used by Mallat [16,17], for orthogonal wavelets. An empirical scale-dependent wavelet noise removal method is explained in Sec. II D. In Sec. III, the results of a few calculations are presented. Several results on empirical scale-dependent calculations are given in Sec. III A with four figures and a table, while Sec. III B presents the information-theoretic results, with two figures. Section IV summarizes our conclusions.

II. FORMULATION

The specific inverse problem to be discussed here is scattering amplitude estimation of a flaw in stainless steel, applied to the nondestructive evaluation (NDE) of engineering materials. Given the scattering amplitude estimate, inverse scattering approaches can be used to estimate the size of the flaw, providing a quantitative basis for determining whether, and ultimately when, to replace a flawed part before failure (but not before replacement is necessary). The inverse problem of scattering amplitude estimation is addressed next.

A. Data and Wiener filter regularizer

The signals measured in ultrasonic testing include the effects of the measurement system, and are corrupted by noise.

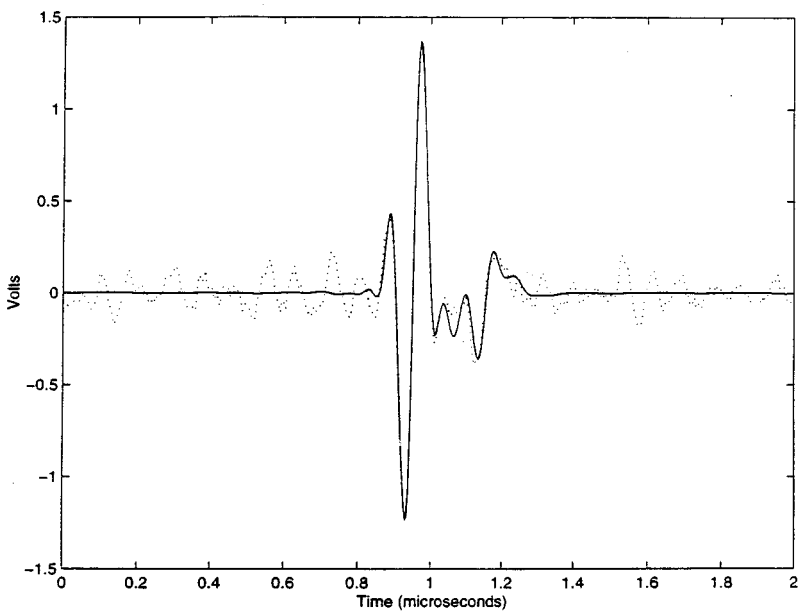

FIG. 1. The calculated time domain signal of a 200-μm sphere shown together with measured elastic wave noise with SNR 4:1 in polycrystalline stainless steel. The solid line is the calculated time domain scattering signal and dotted line has the measured noise added to the signal.

The measurement system response is both band limited and frequency dependent, resulting in measured signals which are blurred and distorted estimates of the actual flaw signatures. The Wiener filter is used to estimate the scattering amplitude of the flaw by attempting to remove the effects of the measurement system in the presence of noise. The optimal form of the Wiener filter requires estimates of the parameters which describe both the noise and scattering amplitude distributions.

The data in this study are taken with a transducer whose square pulses have a central frequency of 12 MHz and drop to -40 dB at 1 and 23 MHz. A measured signal can be modeled as the convolution of the measurement system function with the flaw impulse response function, plus any noise present. In the time domain, this signal is given by

$$f_i = f(t_i) = \int_{-\infty}^{t_i} h(t_i - t')a(t')dt' + n_i, \tag{1}$$

where $n_i = n(t_i)$ is the additive noise and h is the known instrument response of the transducer. Figure 1 shows a time trace of an ultrasonic signal in stainless steel with a signal to noise ratio (SNR) of 4:1. The solid line is a calculated noise-free time domain scattering amplitude from a 200-μm spherical void in steel, and the dotted line is a signal with the grain noise (as measured by one of us, S.P.N.) added. The measurement system used here results in grain scattering noise that is band limited and colored (frequency dependent within the bandwidth) resulting in time domain correlations and autocorrelation behavior in time that do not have Dirac δ falloff. The Fourier transform of Eq. (1) is

$$F(\omega) = H(\omega)A(\omega) + N(\omega) \tag{2}$$

in the continuous frequency case, and

$$F_i = F(\omega_t) = H(\omega_l)A(\omega_l) + N(\omega_l)$$

with $\{I = 1,2,...,M/2\}$ in the discrete case. After the Fourier transformation, the functions $\{F,H,A,N\}$ are all complex valued and there are half as many data as for the real valued functions $\{f,h,a,n\}$ in the time domain. The problem of determining an approximate scattering amplitude $\hat{A}(\omega)$ from Eq. (2) is an ill-posed problem [13–15,1,2,6,7,18–28] because the data are band-limited and corrupted by noise. The regularization of these ill-posed problems can be accomplished by using a Wiener filter [28], W_Q:

$$\hat{A} = W_Q F = \frac{H^* F}{|H|^2 + Q^2(\omega)}, \tag{3}$$

as well as a term from information theory. In Eq. (3), $Q^2(\omega)$ is the regularizing term. The optimal Wiener filter is well known [28] and is given by $Q^2_{\text{opt}} = S_N(\omega)/S_A(\omega)$, where S_N and S_A are the power spectral densities of the noise and the scattering amplitude. Since this is an inverse problem, A (and consequently S_A) are unknown, and the optimal filter cannot be used. The power spectral density of the noise can be estimated from measurements and has been done by one of us (S.P.N.) and others, see, e.g., [11,6,7,29,30]. The frequency independent, suboptimal Wiener filter sets $Q^2 = \varepsilon \max[|H(\omega)|^2]$, where $\varepsilon(0,1)$ is a thresholding parameter (typically is set to 0.01). It corresponds to a limit of -60 dB for the dynamic range of the data. The signal scattered from the flaw and the acoustic noise occupy the same frequency window, so no low-pass filter or smoothing in time can remove the noise without losing information about the flaw itself. The suboptimal Wiener filter acts as a bandpass filter, removing much of the distortion due to the colored frequency response of the measurement system. The suboptimal filter does not achieve noise removal, nor change the SNR within the bandwidth, since information about both the flaw and the noise are passed through the filter. It is well known that Eq. (3) often performs poorly for real data [28], implying that additional information is needed. If one has independent *a priori* information, it may be incorporated by replacing Eq. (3) with the expression

$$\hat{A}(\omega) = W_Q F + C(F), \tag{4}$$

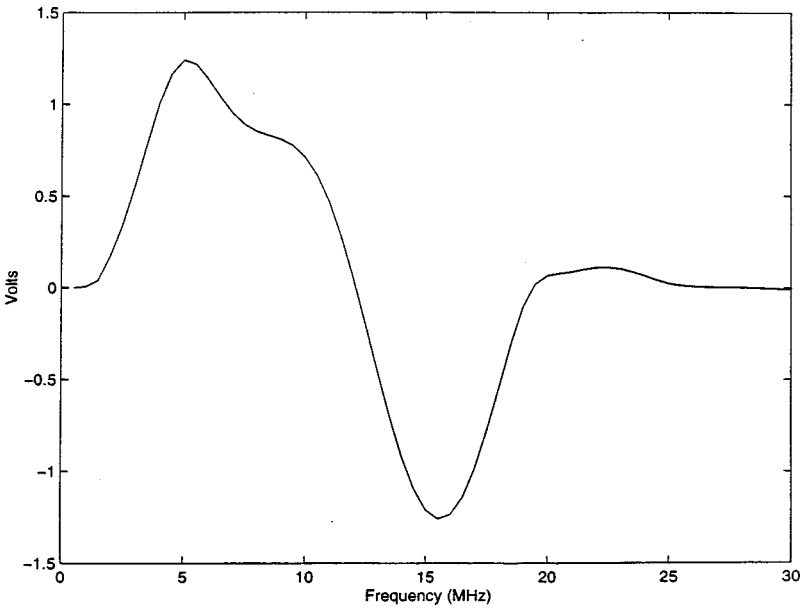

FIG. 2. The real part of the noise free scattering amplitude for a 200-μm sphere as passed through the suboptimal Wiener filter.

where $C(F)$ is a positive-definite cost functional that carries the additional nonredundant information. In Ref. [28] this is done for the case of imaging, where one knows from experience [28,17] that the edge sets are blurred by removing too much of the high-frequency content of the signal. Another common regularizer is a smoothness condition on the approximate solution [5,24]. A different approach using information theory and wavelets that involves part of each of these hypotheses is developed here. It is directly based on recent studies of ultrasonic grain noise in NDE [11,7,29,31,30,32]. The choice of wavelet family will provide the smoothness, and the choice of wavelet scales (to be explained in later sections) gives a nonlinear threshold. By using wavelets and information theory, one can effectively add more information to the problem to improve the noise removal process and further reduce the error norms of the reconstruction. The wavelet multiresolution analysis naturally suggested the present approach and it seems to clarify the usefulness of the scale structure of wavelets for these transient signals.

In order to facilitate the study of scattering amplitude estimation with noise removal, noise corrupted flaw signals were generated using measured acoustic noise from a stainless steel block, the measurement system response for an actual ultrasonic system, and computer generated flaw signatures. The noise corrupted signals for the ith flaw were created as $f(t)=h(t)r(t)+bn(t)$ where b is a scaling factor used to generate the desired SNR [11,29,30,28]. This procedure is described in more detail in Ref. [30]. Dozens of signals were generated for spherical flaws with radii of 75, 100, 150, 200, 250, and 300 μm. The SNR's studied were 10:1, 8:1, 6:1, 4:1, and 2:1. Many examples of signals and reconstructions are available (see Ref. [33]).

The 200-μm sphere is thought to be the critical size which indicates impending failure of a metal. Figure 1 depicts a time trace for a 200-μm spherical void with SNR 4:1 while Fig. 2 shows the Fourier domain representation (real part only) of the signal after filtering with the suboptimal Wiener filter.

B. Information theory concepts

In Ref. [34] the Shannon entropy was used instead of the scattered energy from flaws, to detect flaws in Plexiglas. It was found that the entropy measure outperformed other measures commonly in use. In Refs. [3,4,35], the Shannon entropy was used to select a ''best'' wavelet basis. In the work considered in this paper, the problem is to reconstruct the scattering amplitude which is quite different from these two problems, and it was found that the Shannon entropy performed poorly for our task. This paper shows a new way to combine scattering and information methods.

The problem of defining a measure of information is somewhat recent [28,36,37,35] but much progress has been made. The Shannon information I_S is defined as

$$I_S=k\sum_{i=1}^{N} p_i \log_2(p_i), \tag{5a}$$

where k is a positive constant, p_i is the probability of the ith event, and N events are present in the signal. It is well known that maximizing the Shannon entropy corresponds to the uniform distribution, $p_i=1/N$ for each i.

The Shannon information [36,28] has proven very useful in communications, signal processing, and statistical mechanics, where its applications are called maximum entropy. When there is little or no prior information, the uniform distribution with one or a few physical constraints is often a good choice. However, this choice corresponds to minimum information, which is not a natural goal for many inverse problems. In these problems, maximum information on certain features—such as the boundary of a flaw or a cancer, the size of some foreign object, or the material parameters of the scattering body—is the quality needed. Often, experimentalists (and even theorists) perform a least-squares fit to some known function in order to interpret and characterize an experiment. If there is a firm experimental and theoretical basis for the given function, this can be a powerful technique (see Ref. [23], where a careful inverse analysis showed such a form to occur). In these cases the Fisher information matrix

and the Cramer-Rao theorem [28] are very useful. If there is not a firm basis for the assumed function, however, this approach may not be applicable and can yield nonsensical results.

One method of taking *a priori* information into account is to use the negative of a cross-entropy functional as the measure of information. One cross-entropy functional with a number of desirable features is the Kullback-Liebler information measure [37]. There are a number of other measures of information/entropy, but they are not considered here. The information (entropy) in the Cramer-Rao structure [28,35,38] has been shown to be directly related to shift entropy [39].

Assumption. Of all probability distributions to contain the real and imaginary parts of the scattering amplitude, with the constraints of finite energy, causality, and unitarity, we choose the one which is closest to an *a priori* probability distribution of known scattering amplitudes, $Q=\{q_1,\ldots,q_i,\ldots\}$. When no such probability is given, we take the uniform distribution for N data.

If $\vec{p}=\{p_i\}$ is a measured set of N values of a scattering amplitude, and $\vec{q}$ is a N-vector of values taken from the given probability distribution Q, then

$$I_{\mathrm{KL}}=-\sum_{i=1}^{N} p\log_2\left(\frac{p_i}{q_i}\right) \tag{5b}$$

is the Kullback-Liebler cross information, Equation (5b) can be interpreted as a measure of the discrepancy between $\vec{p}\in P$, the probability space of the reconstructed scattering amplitude, and $\vec{q}\in Q$, the reference scattering amplitude. By minimizing this discrepancy, the maximum information or minimum entropy is obtained. If no such probability distribution Q is known, the uniform distribution is chosen. This reduces the Kullback-Liebler information according to

$$\begin{aligned} I_{\mathrm{KL}} &= -\sum_{i=1}^{N} p_i\log_2\left(\frac{p_i}{1/N}\right) \\ &= \sum_{i=1}^{N} p_i\log_2(1/N)-\sum_{i=1}^{N} p_i\log_2(p_i) \\ &= 1_n(N)+I_S(\vec{p}), \end{aligned} \tag{6}$$

where I_s is the Shannon information as defined in Eq. (5a). Thus, maximum entropy methods are a special case of Kullback-Liebler information when Q is the uniform distribution.

C. Multresolution analysis for orthogonal wavelets

Wavelets [40–44,38,45,46,17] have been shown to have improved localization properties over windowed Fourier transforms, to have good noise removal properties, effective edge detection abilities, and the capability of treating nonstationary stochastic processes. In physics they have recently been used in electronic structure calculations [47], in the formulation and study of correlation functions which arise in particle production [48], and to detect structures in galaxies [49] and two-dimensional (2D) turbulence [50]. Wavelets are basis sets, or frames for finite energy signals in $L^2(\mathbb{R}^N)$. These basis sets satisfy a stability inequality which makes them Riesz bases (a frame is a basis plus many additional redundant elements). Both discrete and continuous wavelets exist, but we will restrict our attention to the discrete case since finite sets of measured data are being studied. This study will use orthonormal bases so that frames are not needed here (continuous wavelets require frames). One family of discrete wavelets is written as $\{\Psi_{jk}(x):(j,k)\in\mathbb{Z}^2\}$ where $\mathbb{Z}$ is the set of integers and $\mathbb{Z}^2$ is the Cartesian product of two (index) sets. The functions Ψ_{jk} are defined by the dilation operator D_{jk}: the product of a translation by k, and a scale change of 2^j, of a single wavelet function $\Psi(x)$ for which

$$\Psi_{jk}(x):=(D_{jk}\Psi)(x)=2^{j/2}\Psi(2^jx-k) \tag{7a}$$

provided Ψ satisfies the three conditions: (a) the linear span (the set of all linear combinations) of all Ψ_{jk}'s form a Riesz basis for $L^2(\mathbb{R}^n)$, (b) the admissibility condition that

$$0<c_\Psi<\infty \tag{7b}$$

with

$$c_\Psi=\int\frac{|\hat{\Psi}(k)|^2}{|k|}dk, \tag{7c}$$

where $\hat{\Psi}(k)$ is the Fourier transform of the function $\Psi(x)$, and (c) $\Psi(x)$ has zero mean

$$\hat{\Psi}(0)=\int_{-\infty}^{\infty}\Psi(x)dx=0. \tag{7d}$$

For signals $f\in L^1(\mathbb{R}^n)\cap L^2(\mathbb{R}^n)$ the Grossmann-Morlet (Calderon) [45] inverse automatically exists. The completeness of the $\{\Psi_{jk}\}$ allows wavelet analysis of any signal $f\in L^2$ by examining and comparing its wavelet coefficients

$$c_{jk}(f)=\langle f|\Psi_{jk}\rangle. \tag{8a}$$

If $f\in L^1(\mathbb{R}^n)\cap L^2(\mathbb{R}^n)$, then

$$f(x)=\sum_{j,k} c_{jk}(f)\Psi_{jk}(x), \tag{8b}$$

and the Grossmann-Morlet inverse of Eq. (8a) exists and is a reconstruction in $L^2(\mathbb{R}^1_+)$.

Strictly speaking, the reconstruction is taken as the inverse solution which minimizes an error norm. Equation (8a) is called the wavelet analysis of a signal f by a fixed wavelet family $\{\Psi_{jk}(x):(j,k)\in\mathbb{Z}^2\}$. Hence Eq. (8b) is called the wavelet reconstruction (synthesis) of $f\in L^1(\mathbb{R}^n)\cap L^2(\mathbb{R}^n)$. The scale changes by 2^j is the new feature of wavelets when compared to Fourier analysis of signals. It corresponds to zooming in to finer details, $x\to 2^{-j}x$, and zooming out to coarser detail, $x\to 2^jx(j>0)$. For this reason, special attention is paid to the scale structure of these inverse reconstructions. The multiresolution structure, which was first introduced by Mallat [16,17], will be used to obtain a new scale-dependent noise removal process from a Wiener filter regularized inverse reconstruction. The reader is reminded that the term ''synthesis'' is general, but that reconstruction should only be used when it is known in what space ''lives.''

Also, a reconstruction in the noise-free case with complete data has a solution that exists, is unique, and depends continuously on the initial data. However, these conditions are not met in measured data.

Since only orthogonal wavelets are used in this study, the MRA of Mallat [16,17] is used; various other types of wavelets (including nonorthogonal) are discussed in Refs. [14,7, 22,40,42–44,38,45].

A multiresolution analysis of $L^2(\mathbb{R}^1)$ is a nested sequence of closed subspaces $\{V_j\}_{j\in\mathbb{Z}}$ of $L^2(\mathbb{R}^1)$ where $\mathbb{Z}$ is the set of all integers, for which conditions (a)–(d) are satisfied: (a) $V_j \subset V_{j+1}$, $\forall_j \in \mathbb{Z}$; (b) $\cup_{j\in\mathbb{Z}} V_j$ is dense in L^2,

$$\cap_{j\in\mathbb{Z}} V_j = \{0\};$$

(c) $f(t)\in V_j$ if and only if $f(2t)\in V_{j-1}$, $\forall_j \in \mathbb{Z}$; (d) a scaling function Φ exists, which $\forall_j \in \mathbb{Z}$ satisfies

$$\{\Phi_{jk}(t) = (D_{jk}\Phi)(t) := 2^{j/2}\Phi(2^j t - k) | k \in \mathbb{Z}\}; \tag{9}$$

and is an orthonormal basis for V_j. At each scale j, the scaling subspace V_j is the low frequency bandpass part of V_{j+1}, and the closed wavelet subspace W_j is the high frequency band pass part of V_{j+1}. The wavelet families written in Eq. (7a) provide an orthonormal basis for the spaces W_j, for each j. The scale functions in Eq. (9), together with the wavelet functions in Eq. (7a), are generated from finite sets of masking coefficients $\{h_k : k = 1,...,N\}$ for which the dilation equations

$$\Phi(t) = \sum_k h_k \Phi(2t-k), \tag{10a}$$

$$\Psi(t) = \sum_k (-1)^k h_{N-k} \Phi(2t-k) \tag{10b}$$

are satisfied for each $k\in\mathbb{Z}$. These masking coefficients are tabulated in a number of the references [40,38,46], and the most recent version of masking coefficients with a number of corrections is available in Ref. [44]. The MRA is a tightly woven mathematical structure since the pair $\{\Phi(t),\Psi(t)\}$ are compactly supported and still generate an orthonormal basis for $L^2(\mathbb{R}^1)$. This amazing mathematical result is due to Daubechies [46]. It is often advantageous to require additional smoothness conditions beyond Eq. (7d) for $k = 1,2,3,...m$ for some $m\in\mathbb{Z}$, $m>2$,

$$\int t^k \Psi(t)\, dt = 0, \tag{11}$$

which increases the ability of the wavelet family generated by Ψ to approximate fine details. These are shown to be useful for ultrasound in the next section. A tradeoff exists between the minimum interval of support of a wavelet family and the order of approximation.

The wavelet analysis and synthesis Eqs. (8a), (8b) will be used as follows in our studies of inverse elastic wave scattering. In the empirical method of Sec. II HD, the analysis is performed by calculating the wavelet coefficients $\{c_{jk}\}$ of a signal and treating different sets of scales as described. The information-theoretic Kullback-Liebler method starts with a useful ''trick'' by Frieden [28] and then uses the maximization of I_{KL} by removing small coefficients on some scales, subject to a signal energy constraint. This constraint requires that the signals have an energy greater than or equal to that implied by the given approximate SNR. The constraint is essential and is assumed in order to avoid removing useful signal energy while eliminating energy due to various noise processes. In this sense, the inverse problem is somewhat analogous to the combined first and second laws of thermodynamics shown in Eq. (4).

Frieden's trick is to take the complex valued estimated scattering amplitudes $\hat{A}(\omega)$ and the noise-free calculated reference scattering amplitude $A_r(\omega)$, and treat them as independent real variables, i.e., $\mathrm{Re}[\hat{A}(\omega)]$, $\mathrm{Im}[\hat{A}(\omega)]$, etc., using the absolute values after an appropriate normalization, to enable their use in the information cost functional. To show the dependence on scales explicitly, expand Eq. (5) as

$$I_{\mathrm{KL}} = \sum_{j,k} (I_{\mathrm{KL}})_{jk} = \sum_j \sum_k p_{jk} \log_2\left(\frac{p_{jk}}{q_{jk}}\right). \tag{12}$$

Using Eq. (8b) for a fixed wavelet family, one has

$$\hat{A}(\omega) = \sum_{j,k} c_{jk} \Psi_{jk}(\omega) \tag{13a}$$

and

$$A_r(\omega) = \sum_{j,k} d_{jk} \Psi_{jk}(\omega). \tag{13b}$$

Using the trick, together with Eqs. (13a), (13b), Eq. (12) becomes

$$I_{\mathrm{KL}} = \sum_j \sum_{k=1}^{N_j} \mathrm{Re}[\hat{a}(\omega)]_{jk} \log_2\left[\frac{\mathrm{Re}[\hat{a}(\omega)]_{jk}}{\mathrm{Re}[a_r(\omega)]_{jk}}\right] + \sum_j \sum_{k=1}^{N_j} \mathrm{Im}[\hat{a}(\omega)]_{jk} \log_2\left[\frac{\mathrm{Im}[\hat{a}(\omega)]_{jk}}{\mathrm{Im}[a_r(\omega)]_{jk}}\right]. \tag{14}$$

The maximum I_{KL} is obtained subject to the energy constraint using the wavelet coefficients of the real and imaginary parts of $\hat{A}(\omega)$. The sums over k run from 1 to N_j (an integer depending on the scale j). The scattering amplitudes were suitably normalized and the absolute values taken to allow for the probabilistic interpretation, the lower case $\hat{a}$ and a_r indicating the estimated and the reference scattering amplitudes divided by $\Sigma\hat{A}(\omega_i)$ and $\Sigma A_r(\omega)$, respectively. Thus $a_r(\omega)$ plays the role of the reference distribution q in the Kullback-Liebler information. A different use of the Kullback-Liebler information was used by Coifman and Saito [35] for classifying features of a signal. The real and imaginary parts of the estimated scattering amplitude were also treated separately here. This approach will also be discussed.

The wavelet analysis of a discrete signal such as f in Eqs. (10) or (2) has a perfect reconstruction by Eq. (8b) if all wavelet coefficients are retained. The signal is said to be compressed if all of the small wavelet coefficients $|c_{jk}| < \varepsilon$ (a threshold) are discarded, provided the imperfect reconstruc-

tion of f is deemed satisfactory. These small coefficients contain a small amount of the signal energy, and compression ratios of 20 or 30:1 are common. We take the opposite viewpoint: a small energy coefficient may be important, especially if it carries a large information content. The locations of the zeros, maxima, and minima of the scattering amplitudes are several key features for this paper. The small coefficients near any of these features increase the accuracy in their determination.

The inverse reconstruction here will begin with the suboptimal Wiener filter regularized estimate of the scattering amplitude. The $l^1(\mathbb{R})$ (or vision norm) is given by

$$\varepsilon_1 := \sum_i^N |\hat{A}(\omega_i) - A_r(\omega_i)|, \tag{15a}$$

and the $l^2(\mathbb{R})$, or energy norm is

$$\varepsilon_2 := \sum_i^N |\hat{A}(\omega_i) - A_r(\omega_i)|^2, \tag{15b}$$

where $\hat{A}$ is the estimated scattering amplitude, A_r is the reference scattering amplitude, and N is the number of data points. The number of data points limits the number of scales j that can be used in the analysis; for $N=256$ or 512 (as in our data), $j_{\max}=7$ or 8, with $j_{\min}=0$. In practice, $j_{\max}$ is chosen to be less than the theoretical maximum. In order to investigate the effects of the order of approximation (smoothness) of the analyzing wavelets, several wavelet families were used: $D4$-$D28$, $DS8$-$DS20$, and $C6$-$C24$, where DN stands for Daubechies minimum phase wavelets, DSN for Daubechies least asymmetric, and CN for Coiflets (as per Daubechies [41]), and the integer N here denotes $2m$ from Eq. (11).

D. Empirical wavelet scales for noise removal

In this subsection we will now discuss the technique of pruning coefficients by scale versus chopping them by comparison with a size threshold ε_t. By chopping we mean the removal of all coefficients $|c_{jk}|<\varepsilon_t$, which is a standard technique. Pruning entails partitioning the scales $\{0,1,\ldots,j_{\max}\}$ into ''ordinary'' and ''exceptional'' scales. The ordinary scales are ones for which the coefficients are chopped, while the exceptional scales do not undergo any chopping. The assumption is that the exceptional scales may contain small coefficients with low signal energies, but that these scales have a high information content with regard to the signal. This suggests the question of whether these exceptional scales $\{j_e\}$ exist and, if so, which values of j are exceptional and what threshold value ε_t should be used. This empirical method, even if successful, does not provide understanding of why this method would improve the reconstructions.

To determine if there is a clear dependence on the scales j, the following numerical experiment was performed: The l^1 and l^2 error norms [Eqs. (15a), (15b)] were calculated for 30 Wiener filtered reconstructions of flaws of known radius (initially 200 μm) and a fixed SNR (initially 8:1) using the suboptimal filter given in Eq. (3). A wavelet decomposition of both the real and imaginary part of the estimated scattering amplitude $\hat{A}(\omega)$ was performed, decomposing the signal into six or seven scales using a fixed wavelet family. In this part of the study we limited the choice of wavelet family to $D4$, $D10$, $D28$, $DS8$, and $C18$. Following the decomposition, a variety of combinations of scales were used as exceptional scales, and the coefficients were pruned. With this new coefficient set, the inverse wavelet transform is performed, which results in a new reconstruction of the scattering amplitude. The error norms were then calculated again using this new estimate. The sets of scales that were used as exceptional scales included several triples (j_1,j_2,j_3) and all doublets (j_1,j_2). The threshold parameter ε_t was found to depend on the SNR, ranging from $\varepsilon_t=0.2$ at SNR 10:1, to $\varepsilon_t=0.35$ at SNR 8:1, to $\varepsilon_t=0.75$ at SNR 2:1. Some results of the empirical pruning and thresholding for sets of exceptional scales and a few wavelet families are shown in Table I. Only one triplet (3,4,5) was found to provide low error norms. A number of other flaw radii were studied in Refs. [6, 7] and many more results from these works are available via the WWW [33].

TABLE I. The l^1 and l^2 error norms of 200-μm spherical voids in stainless steel with SNR 8:1 for $D4$,$D10$,$C18$ wavelets for the ensemble average of 30 signals. The noisy, suboptimal Wiener filter error is given near the top of the list and the error norms for six choices of exceptional scales are presented for comparison.

Wiener filter	l^2 Real 0.86918	l^2 Imag 1.24146	l^1 Real 0.93163	l^1 Imag 1.46722
$D4$ wavelet				
$j_e=1,2$	0.35294	0.38552	0.36349	0.46548
$=2,3$	0.19973	0.30495	0.20180	0.36515
$=3,4$	0.17827	0.26589	0.18186	0.31476
$=4,5$	0.27568	0.25965	0.29410	0.30977
$=5,6$	0.27654	0.25730	0.28591	0.30230
$=3,4,5$	0.17953	0.25774	0.18159	0.30622
$D10$ wavelet				
$j_e=1,2$	0.31748	0.38308	0.36102	0.48856
$=2,3$	0.23943	0.29939	0.25952	0.38376
$=3,4$	0.20858	0.21956	0.21942	0.27978
$=4,5$	0.24385	0.19542	0.27386	0.25554
$=5,6$	0.22349	0.23036	0.25927	0.29983
$=3,4,5$	0.21030	0.21052	0.22525	0.26764
$C18$ wavelet				
$j_e=1,2$	0.33870	0.39060	0.39086	0.46942
$=2,3$	0.23207	0.30046	0.27017	0.35717
$=3,4$	0.17162	0.24619	0.20648	0.28002
$=4,5$	0.22570	0.27710	0.28679	0.31115
$=5,6$	0.23589	0.27680	0.28386	0.31389
$=3,4,5$	0.17225	0.25131	0.20975	0.28050

Next, several existing methods of noise removal were tried in the hopes of explaining that the choice of exceptional scales gave a useful noise removal structure. Methods proposed by Donoho [5], Coifman and Wickerhauser [35], and Wickerhauser [38] were tried. Donoho's method used soft thresholding with different threshold values in different scales. Soft thresholding is different from the chopping described earlier, in that all coefficients are reduced in magnitude by the threshold amount and any coefficient less than the threshold is set to zero. The other methods [35,38] were

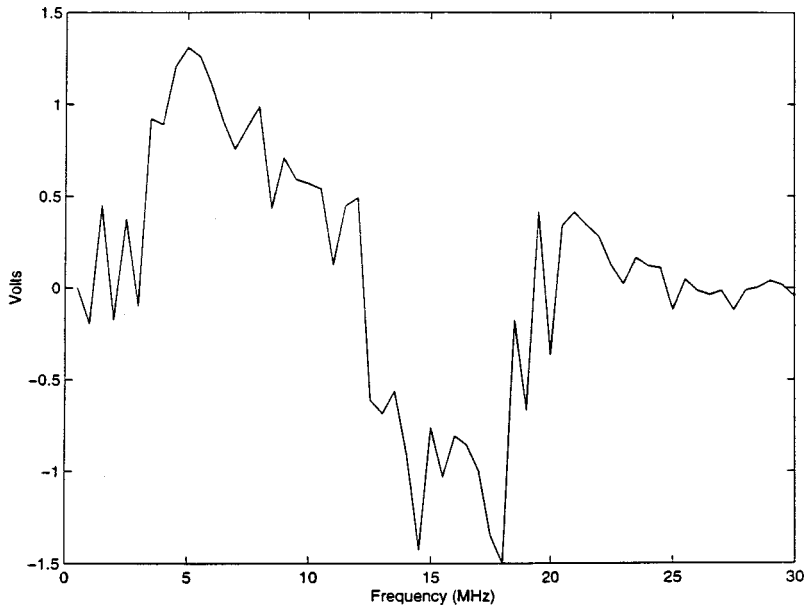

FIG. 3. The real part of the noise-corrupted scattering amplitude, using the suboptimal Wiener filter, with SNR 6:1. Contrast this with Fig. 2.

based on Shannon information (entropy). The methods in [5] produced reconstructions that were comparable to the result in Fig. 3, except that some cases were worse. When computer generated noise was used, it performed well and the published results were reproduced, but the method was completely unable to treat the measured ultrasonic grain noise studied here. The improvements obtained by using [35,38] were smaller and more erratic than the empirical thresholding. The results of the empirical thresholding will be shown in the next section.

E. Kullback-Liebler maximization for noise removal

The information-theoretic technique, which was found to give results very similar to the empirical method just described, was a constrained Kullback-Liebler information [37] maximization approach. This is a maximum information or minimum entropy method, in contrast to the maximum entropy approach. The constraints include the *a priori* information of the SNR, an energy threshold ε_e dependent on the SNR, and the knowledge that a flaw is present in the time window. The reference distribution is taken from the real and imaginary parts of the reference scattering amplitude for a given flaw size. For realistic flaws, the flaw would not be spherical and the effective radius would be different in different incident sonification directions of the backscatter. This situation would require the use of a Radon transform, such as those used in [13,22,27], and is well worth pursuing later. Also, a Lagrange multiplier coefficient λ on the Kullback-Liebler information term in Eq. (16) should be studied.

By maximizing the information measure in Eq. (14), the resulting error norms given by Eqs. (15a) and (15b) are significantly reduced. Given an estimate of the signal to noise ratio one can estimate the expected total energy contained in the signal of interest. The power spectral density of the measured signal (true signal plus noise) is given by $S_T = S_{\hat{A}} + S_N$ where $S_{\hat{A}}$ and S_N are the power spectral densities of the

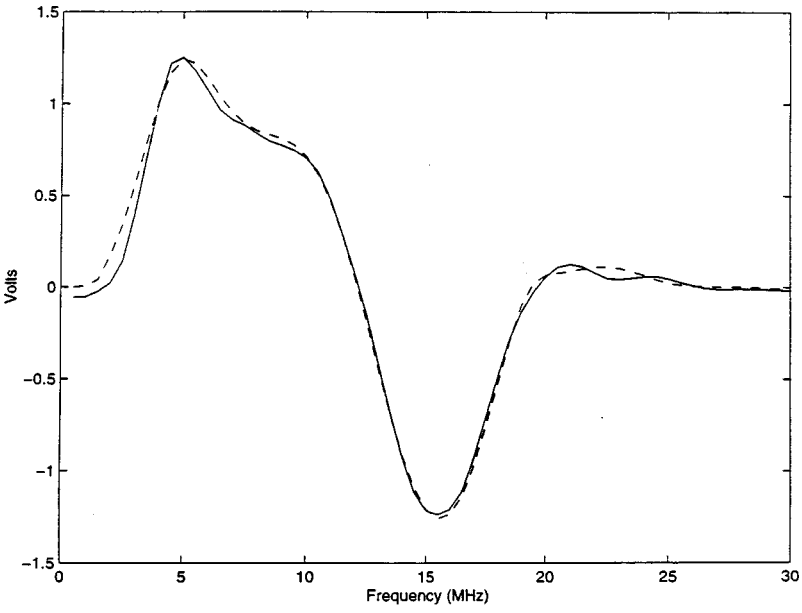

FIG. 4. The result (solid line) of empirical thresholding using $j_e = 3,4,5$, and wavelet family $C18$ compared to the noise free case (dotted line). This is for the real part of the scattering amplitude, with a SNR 10:1.

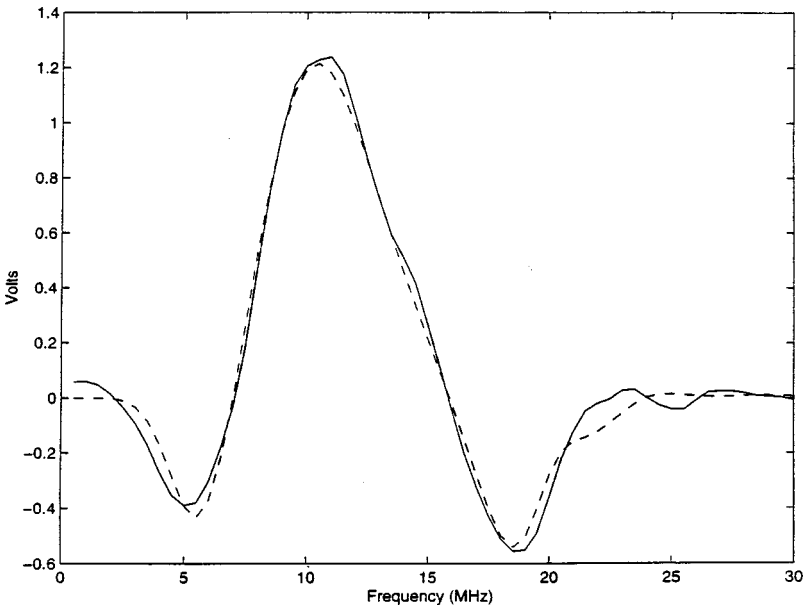

FIG. 5. The result (solid line) of empirical thresholding using $j_c=3,4,5$ and wavelet family $D10$, compared to the noise free case (dotted line). This is for the imaginary part of the scattering amplitude, with SNR 8:1.

estimated signal and the noise, respectively. Since SNR $=S_{\hat{A}}/S_N$ we can write $S_T-S_N=S_N\times(\mathrm{SNR})$. This allows one to approximate the amount of energy that can be discarded safely in the Kullback-Liebler maximization process. No iteration in the algorithm is allowed to violate this constraint.

To maximize Eq. (14), given the needed *a priori* information, one begins by calculating a baseline I_0, using the initial $\hat{A}$. Any changes made to the estimate $\hat{A}$, must increase the information content, i.e., maximize Eq. (14). Following this initialization, a wavelet coefficient c_{jk} [from Eq. (13a)] is chosen at random, from any scale, and reduced by a small percentage. The reduction of a wavelet coefficient is accepted for the new estimate of $\hat{A}$ if the energy constraint is not violated, and if the information content is increased ($I_m > I_{m-1}$). The process is then repeated until several hundred iterations pass without any changes being accepted. Two conditions usually resulted in this convergence: A lower limit on the energy had been reached, or the information given by Eq. (14) had been maximized. In the next section results for both the empirical and the information theoretic approach are presented.

III. RESULTS

A. Empirical wavelet scale-dependent noise removal

A few of the results from the empirical scale-dependent pruning technique will first be presented and discussed. Thirty signals were formed from measured noise in stainless steel combined with calculated scattering amplitudes. Initial noisy estimates of the scattering amplitude were then calculated using the suboptimal Wiener filter. Table I presents the results for a subset of the exceptional scale selections, and a subset of the wavelet families used, for a particular flaw radius, at SNR 8:1. From our calculations and the others available via the WWW [33], two observations follow: (1)

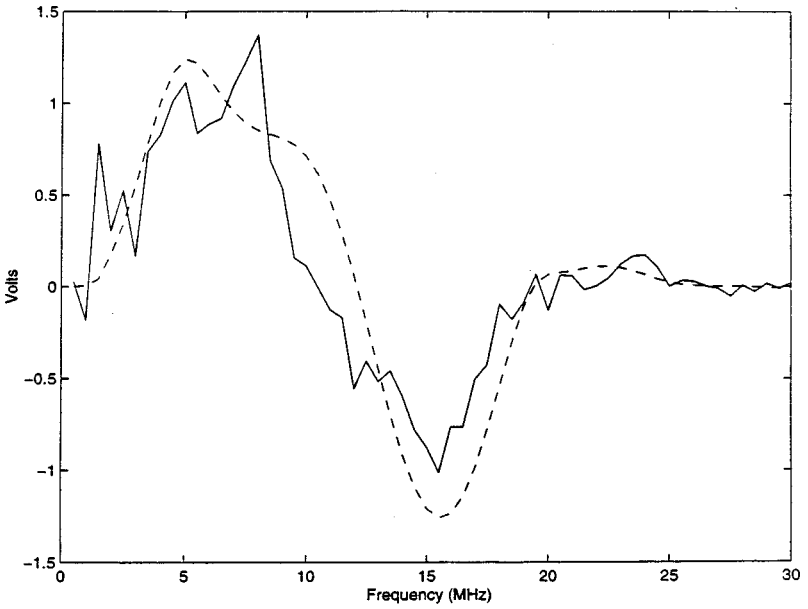

FIG. 6. The result (solid line) of empirical thresholding using $j_c=1,2$ and wavelet family $D4$, compared to the noise free case (dotted line). This is for the real part of the scattering amplitude, with SNR 8:1.

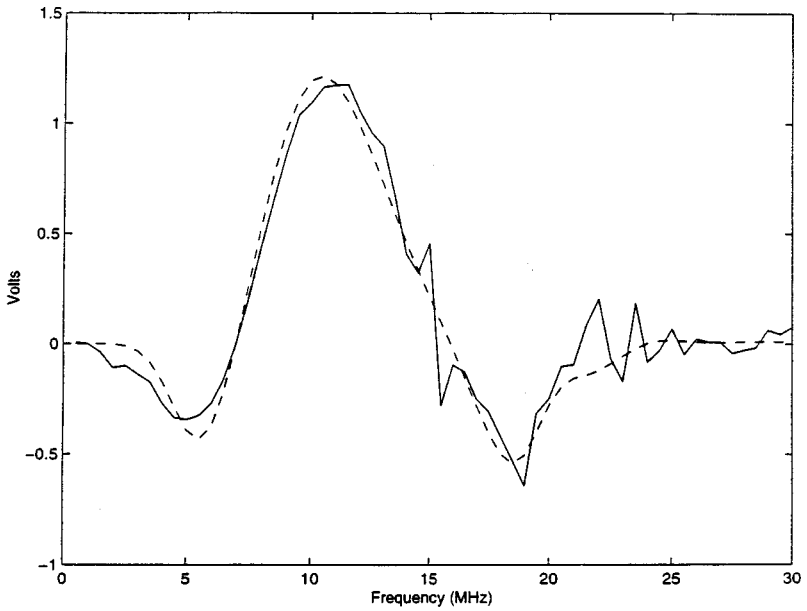

FIG. 7. A Kullback-Liebler maximum information reconstruction of the 200-μm scatterer in stainless steel, using the $D10$ wavelet family, and a 90% threshold for the energy.

There was a clear scale dependence with $j=3,4,5$ or 4,5 (intermediate frequencies) as the best choices, while $j=1,2$ (high frequency) performed poorly. (2) There was no clear ''best'' choice of wavelet. Overall, the $D4$ wavelet performed the worst, while the $D10$ worked the best for $\mathrm{Im}[\hat{A}(\omega)]$, and $C18$ was the best for $\mathrm{Re}[\hat{A}(\omega)]$. Wavelets with more smoothness performed markedly better than those with less differentiability. This agrees with the observations of Tobocman [14,15] with analytic wavelets.

The noisy Wiener filtered scattering amplitude estimates are all similar to Fig. 3. Ideally, the noise removal procedure would result in a plot similar to the one seen in Fig. 2. In order to better understand the 40–50 % reduction in the l^1 and l^2 error norms over the noisy Wiener filtered estimates, some figures will be shown. A few of the cases where the method failed, such as for $j_e=1,2$, or using the $D4$ wavelet, are also presented. Reconstruction of the real part of the scattering amplitude for a 200-μm sphere with SNR 8:1, using empirical scale dependent pruning with wavelet family $C18$, $j_e=3,4,5$, is shown in Fig. 4. The solid line is the theoretical noise-free case for comparison. Figure 5 displays the results for the imaginary part using the $D10$ wavelet, and scales $j_e=4,5$. Note some large amplitude oscillations in the region 20–35 MHz, which would require more smoothing than is provided by the empirical pruning, but the main scattering features up to 20 MHz are accurate. The last example, in Fig. 6, illustrates the dependence of the reconstruction on the choice of exceptional scales and wavelet family. Numerous noise artifacts appear, but the principal features are recognizable (although somewhat inaccurate). The noise removal was performed using a $D4$ wavelet with exceptional scales $j_e=1,2$.

As stated before, a large archive of other calculations, encompassing many other flaw radii, many different wavelet families, and varying SNR's are available via the WWW [51]. These results are a sample from the greater ensemble of work. From the table and the figures it seems that a threshold $\varepsilon_t=1.5/\mathrm{SNR}$, with exceptional scales $j_e=3,4,5$ or $j_e=4,5$ improves the scattering amplitude estimate by removing a good deal of noise. This immediately raises the question of why it works—a question to be addressed in the next section.

B. Kullback-Liebler information theoretic wavelet noise removal

The best answer found so far, to the question of why empirical pruning works is that the relative information (relative to the known reference distribution Q) is maximized.

In Fig. 7, a reconstruction maximizing the Kullback-Liebler information [given in Eq. (14)] is shown. The scattering amplitude shown is for a 200-μm sphere, with SNR 10:1, using a $D28$ wavelet. The result from using the same scattering parameters, but a different wavelet family ($D10$) is shown in Fig. 8, and a plot of the relative error norms for ten signals using $D10$ wavelets is shown in Fig. 9. These reconstructions are of reasonably good quality, and many more of similar quality were also obtained. In fact, the l^1 and l^2 error norms of ten KL information theoretic reconstructions were better than the empirical thresholding method (as can be seen from Fig. 9). One concern was that the algorithm was overtrained, and not robust outside the initial test population of spheres. The concern was that, given a signal with no flaw scattering present, the algorithm would still return a reconstruction that looked like a flaw scattering amplitude. To test the algorithm, the reference distribution Q was kept as the reference scattering amplitude, while the experimentally measured grain noise was input as the signal. The result of the KL noise removal algorithm then gave the dotted line reconstruction in Fig. 10. These results reassured us that the noise removal system was sound and robust.

IV. DISCUSSION

From the table and the ten figures, both scale-dependent empirical thresholding and the information-theoretic Kullback-Liebler noise reduction methods were shown to improve the reconstructions of ultrasound scattering amplitudes over those obtained from a suboptimal Wiener filter. The empirical thresholding has no physical basis, but it does show that the time-frequency representation of wavelets pro-

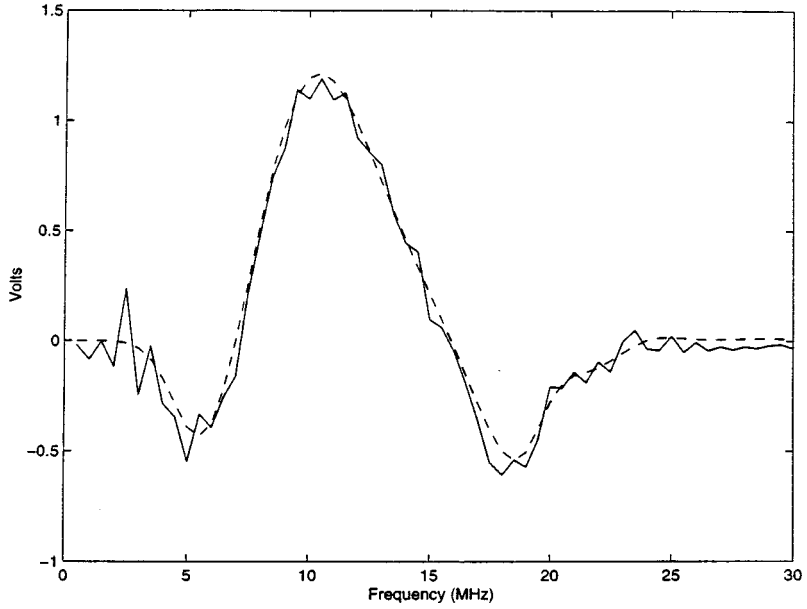

FIG. 8. Same as Fig. 7 with a 50% threshold level for the energy.

vides an additional mathematical tool to improve (or make worse) the inverse reconstructions. The information-theoretic Kullback-Liebler approach minimizes the discrepancy (and thereby error norms) between a noisy Wiener filtered signal and a noise-free reference signal. This corresponds to maximizing the information or minimizing the entropy while preserving the noise free signal energy. Actually this process does not contradict the maximum entropy method because these experiments incorporate additional information. The additional knowledge is encoded in the reference distribution Q which is required for the KL information measure. In addition, there is additional structure which has not yet been used: the unitarity of the scattering amplitude. For forward scattering, $\Theta \cong 0$, this exclusion would not be a good idea, but the problems would be manageable for the backscattering studied here.

The method developed can (and does) fail in some instances. The example in Fig. 6 shows that noise artifacts can persist after applying empirical thresholding. At the lower SNR's all choices of scales and wavelets passed some noise artifacts through the noise removal process. Nevertheless, the noise removal technique shown works much better than the suboptimal Wiener filter, and the Kullback-Liebler cross-information theory gives a physical basis for the method.

Information-theoretic inverse theories have certain advantages over statistical methods, one of which is analogous to the combined first and second laws of thermodynamics in Eq. (4). This combination requires that the information in the cost functional must be nonredundant from that in the L^2 error norm. The fact that the KL method works well suggests that the L^2 energy error and the KL information are not redundant. In thermodynamics, this crucial requirement of nonredundance is taken into account by the ''equation of state'' of the material. In inverse problems, the equation of state is unknown and, in this sense, is the knowledge which is sought here. This study has shown that the KL information is largely independent of the energy error, in contrast to the poor performance of the Shannon information [6]. This line

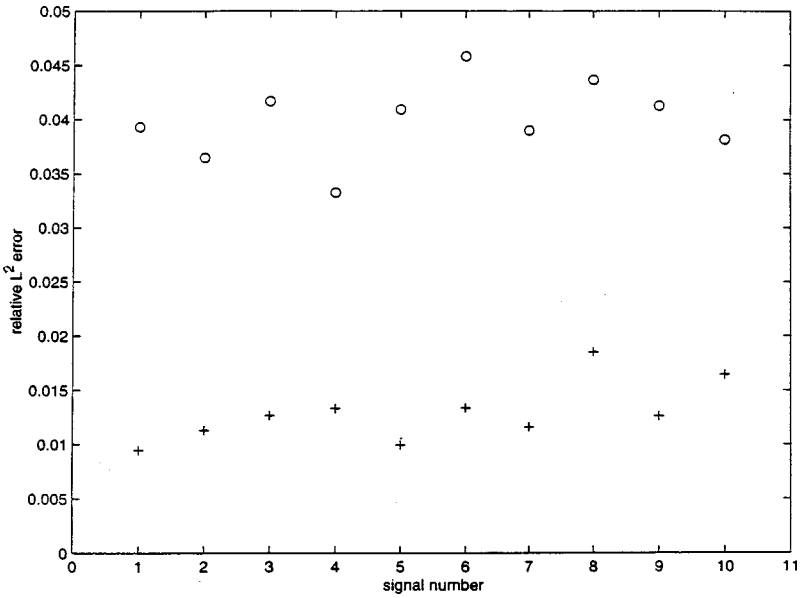

FIG. 9. The relative l^2 error norms for ten signals are shown as circles, while the error norms for signals after noise removal using the KL technique are shown as $+$.

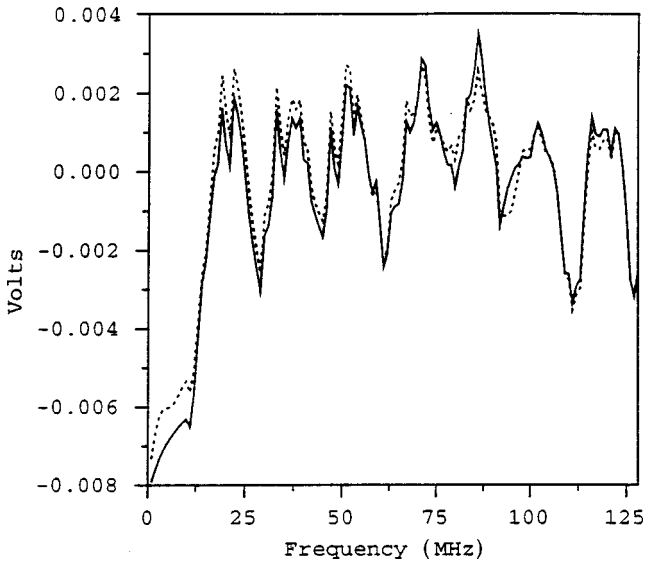

FIG. 10. Grain noise in the time domain is shown as the solid line, while the dotted line represents the signal following noise removal using the KL information maximization, using the noise free scattering amplitude reference distribution *Q*. This figure helps support our claim that the algorithm is not overtrained, and is robust.

of thought is supported by the following interpretation of the results of Hughes [34] who studied the detection problem for large flaws in Plexiglas using Shannon information, but not scattered energy. It was found in Hughes' study that the Shannon information detected large flaws better than the scattered energy. Since the studies that had preceded [34] had good detection performance, though not as good as the Hughes' method, we infer that the Shannon information is largely (but not completely) redundant with the scattered energy.

Many statistical methods suffer from the fact that they are a ''one size fits all'' approach, so that they do not generalize to use additional information easily. Of course, statistics are part of any inverse problem involving real data, but all of the information available about a problem should be used. The cost functional approach [28] allows the use of more of the *a priori* information available, which can appear in two independent parts of the analysis.

Another desirable feature of the KL approach is the fact that minimal input information is required; all that is needed is a reference probability distribution *Q* and an approximate estimate of the SNR. It is also desirable that the Kullback-Liebler information reduces to the maximum entropy method if *Q* is chosen to be a uniform distribution. The fact that, for a few samples, the empirical thresholding produced smaller error norms than the KL method suggests that better choices may exist for a reference probability distribution. The method of Coifman and Saito [35] will be studied in the future. The combination of the performance of the empirical thresholding and its close relation to the maximum information method may say something about orthonormal wavelet families themselves. That is, different scales carry different amounts of energy and information, and they are not the same. The exceptional scales are islands of information in the time frequency plane. We conjecture that this is true for other inverse problems besides the ultrasonic problem studied here [51–55,47–50].

ACKNOWLEDGMENTS

This work was supported by research Grants Nos. AFOSR/NM F49620-96-1-0060 and F49620-96-1-0380, the U.S. Naval Air Warfare Center, Weapons Division, China Lake, CA, and Missouri University, Columbia, MO. This support is gratefully acknowledged. Some of the wavelet code used in this work was written by Eric Veum, a student in our group who received support from the Hughes Medical Foundation. The codes are available by anonymous ftp at theory.physics.missouri.edu. B.D.F. thanks the Department of Mathematics of Texas A&M University for their hospitality during his research leave. Dr. Brett Borden of NAWC, China Lake, CA is thanked for his careful reading of the manuscript and Arjuna Flenner is thanked for his help with the figures.

[1] G. Wahba, SIAM J. Numerical Anal. **14**, 651 (1974).
[2] G. Wahba, *Spline Methods for Observational Data* (SIAM, Philadelphia, 1990).
[3] R. Coifman and M. V. Wickerhauser, in *Wavelets: Mathematics and Applications*, edited by J. J. Benedetto and M. Frazier (CRC Press, Boca Raton, FL, 1994), pp. 399–423.
[4] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software* (A. K. Peters, Wellesley, MA, 1994).
[5] D. Donoho, IEEE Trans. Inf. Theory **41**, 613–627 (1995).
[6] Alan J. Van Nevel, Ph.D. thesis, Missouri University, Columbia, MO (1996).
[7] A. Van Nevel, B. DeFacio, and S. Neal (unpublished).
[8] M. G. Silk, *Fatigue Crack Measurement Techniques and Applications*, edited by K. J. Marsh, R. A. Smith and R. O. Ritchie (Engineering Materials Advisory Services, Boston, MA, 1991), Chap. 5, pp. 95–145.
[9] *Nondestructive Evaluation of Aging Aircraft, Airports, and Aerospace Hardware*, edited by R. D. Rempt and A. L. Brotz (SPIE, Bellingham, WA, 1996).
[10] R. C. Asher, *Ultrasonic Sensors for Chemical and Processing Plants* (Institute of Physics, Bristol, 1997).
[11] S. P. Neal, Ph.D. thesis, Iowa State University, Ames, IA, 1988.
[12] *Medical Imaging 1997 Ultrasonic Transducer Engineering*, edited by K. Kirk Shung (SPIE, Bellingham, WA, 1997).
[13] F. Natterer and F. Wubbeling, Inverse Probl. **11**, 1225 (1995).
[14] W. Tobocman, Inverse Probl. **5**, 1131 (1989).
[15] W. Tobocman, Inverse Probl. **12**, 499 (1996).
[16] S. Mallat, Trans. Am. Math. Soc. **315**, 69 (1989).
[17] S. Mallat, IEEE Trans. Pattern. Anal. Mach. Intell. **77**, 674 (1989).
[18] K. Chadan and P. C. Sabatier, *Inverse Problems in Quantum*

Scattering Theory, 2nd ed. (Springer-Verlag, New York, 1989).
[19] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering* (Springer-Verlag, New York, 1992).
[20] V. G. Romanov, *Integral Geometry and Inverse Problems for Hyperbolic Equations* (Springer-Verlag, New York, 1974).
[21] D. Colton, R. Ewing, and W. Rundell, *Inverse Problems in Partial Differential Equations* (SIAM, Philadelphia, 1990).
[22] A. K. Louis, Inverse Probl. **12**, 175 (1996).
[23] H. Ellmer and D. Semrad, Phys. Rev. E **54**, 3569 (1996).
[24] A. N. Tikhonov and V. Y. Arsenin, *Solutions of Ill-Posed Problems* (Winston, Washington, D.C., 1977).
[25] M. M. Lavrentév, V. G. Romanov and S. P. Shishat-Skii, *Ill-Posed Problems of Mathematical Physics and Analysis* (American Mathematical Society, Providence, RI, 1986).
[26] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Ill-Posed Problems* (Kluwer, Dordrecht, 1996).
[27] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems* (Springer-Verlag, New York, 1996).
[28] B. R. Frieden, *Probability Statistical Optics and Data Testing*, 2nd. ed. (Springer-Verlag, New York, 1991).
[29] S. P. Neal, P. L. Speckman, and M. A. Enright, IEEE Trans. Ultrason. Ferroelectr. Freq. Control **40**, 347 (1993).
[30] S. P. Neal and D. O. Thompson, J. Nondestruct. Eval. **11**, 57 (1992).
[31] C. H. Chen and S. K. Sin, J. Acoust. Soc. Am. **87**, 976 (1990).
[32] A. Abbate, J. Frankel, and P. Das, (unpublished).
[33] URL: http://theory.physics.missouri.edu/~frog/professional.html
[34] M. S. Hughes, J. Acoust. Soc. Am. **91**, 2272 (1992).
[35] R. R. Coifman and M. V. Wickerhauser, IEEE Trans. Inf. Theory **38**, 713 (1992); R. R. Coifman and N. Saito, C. R. Acad. Sci., Ser. I: Math. **319**, 191 (1994).
[36] C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[37] S. Kullback and R. A. Liebler, Ann. Math. Stat. **22**, 79 (1951).
[38] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software* (A.K. Peters, Wellesley MA, 1994).
[39] G. V. Vstovsky, Phys. Rev. E **51**, 975 (1995).
[40] C. K. Chui, *An Introduction to Wavelets* (Academic, San Diego, 1992), Vol. 1; *Wavelets: A Tutorial in Theory and Applications* (Academic, San Diego, 1992), Vol. 2.
[41] I. Daubechies, *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).
[42] G. Kaiser, *A Friendly Guide to Wavelets* (Birkhauser, Boston, 1994).
[43] M. Holschneider, *Wavelets, An Analysis Tool* (Clarendon Press, Oxford, 1995).
[44] C. K. Chui, *Wavelets: A Mathematical Tool for Signal Processing* (SIAM, Philadelphia, 1997).
[45] A. Grossmann and J. Morlet, SIAM J. Math. Anal. **15**, 723 (1984).
[46] I. Daubechies, Commun. Pure Appl. Math. **41**, 909 (1988).
[47] K. Cho, T. A. Arias, J. D. Johnopoulos, and P. K. Lam, Phys. Rev. Lett. **71**, 1808 (1993).
[48] M. Greiner, P. Lipa, and P. Carruthers, Phys. Rev. E **51**, 1948 (1995).
[49] E. A. Slezak, A. Bijoui, and G. Mars, Astron. Astrophys. **227**, 301 (1990).
[50] M. Farge and G. Rabreau, C. R. Acad. Sci., Ser. II: Mec., Phys. Chim., Sci. Terre Univers **307** 1479 (1994).
[51] N.-X. Chen, Phys. Rev. Lett. **64**, 1193 (1990).
[52] Q. Xie and N.-X. Chen, Phys. Rev. E **52**, 5065 (1995).
[53] N.-X. Chen *et al.*, Phys. Rev. E **48**, 1558 (1993).
[54] R. G. Newton, *Scattering Theory of Waves and Particles*, 2nd ed. (Springer-Verlag, New York, 1982).
[55] J. H. Rose and J. A. Krumhansl, J. Appl. Phys. **50**, 2951 (1979).