THE ANNALS OF "DUNAREA DE JOS" UNIVERSITY OF GALATI
FASCICLE III, 2008, Vol.31, No.1, ISSN 1221-454X
ELECTROTECHNICS, ELECTRONICS, AUTOMATIC CONTROL, INFORMATICS

AN ALGORITHM FOR THRESHOLDING IN THE CONTEXT OF SIGNAL DENOISING WITH WAVELETS

Anisia Gogu, Dorel Aiordachioaie

Dunarea de Jos Galati University, Electrical and Electronics Engineering Faculty
Electronics and Telecommunications Department
Domneasca-47, Galati 800008, Romania
Anisia.Gogu@ugal.ro, Dorel.Aiordachioaie@ugal.ro

Abstract: The problem of estimating a signal corrupted by additive noise has been of interest to many researchers for practical as well as theoretical reasons. Wavelet based methods have become increasingly popular, due to a number of advantages over the linear methods. A simulation-based analysis of some thresholding functions in the context of denoising application of wavelet transform was investigated. A probability based function to compute the threshold parameter is described and implemented in the Matlab-Simulink simulation environment, by using an estimation of the probability density function of the wavelet coefficients. The obtained results are at the same quality level with other procedures currently used in denoising.

Keywords: signal processing, analysis, denoising, filtering, thresholding, simulation, wavelets.

1. INTRODUCTION

The main objective of the paper is to present a simulation based analysis of the signal denoising application of wavelets. Some classical - well-known - functions were considered as well as a probability-based function to estimate the parameters of the thresholding algorithms. Firstly, in section 2, we make a short review of the wavelets theory in order to understand better the background, the state of the art and to prepare the field for new denoising algorithms. Secondly, in section 3, we make an analysis of some well known and intensive used thresholding algorithms. Section 4 describes the proposed algorithm for the computation of the threshold. Section 5 presents and comments the obtained results on various input signals by using the proposed algorithm and referring to other common and intensively used denoising algorithms.

2. BACKGROUND

There are at least two ways to introduce wavelets: one is through the *continuous wavelet transform* (CWT) and another one is through *multiresolution analysis* (MRA), (*Jawerth and Sweldens*, 1993). The MRA could be explained also by starting from signals bases, (Phillips, 2004). Let $\{\psi_{j,k}, \forall j,k\}$ be an orthogonal basis. Let W_j be the set of all signals, $s(t)$, which can be synthesized from the baby wavelets $\{\psi_{j,k}, \forall k\}$, so

$$(1)\quad s_j(t) = \sum_{k=-\infty}^{\infty} c_{j,k} \cdot \psi_{j,k}(t)$$

and

This paper was recommended for publication by Laurentiu Frangu

$$(1.a)\quad s(t) = \sum_{j=-\infty}^{\infty} s_j(t)$$

Another way to express the above idea is to define V_j to be the set of all signals, *s(t)*, which can be synthesized from the baby wavelets $\psi_{i,k}(t)$, where *i<j* and $-\infty < k < \infty$:

$$(2)\quad s(t) = \sum_{i=-\infty}^{j-1} \sum_{k=-\infty}^{\infty} c_{i,k} \cdot \psi_{i,k}(t)$$

Every signal in V_{j+1} is a sum of a signal in V_j and W_j, which means that the spaces W_j are the differences (in the subspace sense) between adjacent spaces V_j and V_{j+1}:

$$(3)\quad V_{j+1} = W_j + V_j$$

The useful wavelets, $\psi(t)$, have a scaling function $\phi(t)$ which can produce the multiresolution spaces V_j. Defining "baby scaling functions"

$$(4)\quad \phi_{j,k}(t) = 2^{j/2} \cdot \phi(2^j t - k)$$

where $-\infty < j,k < \infty$, just as for the wavelet, the scale of $\phi_{j,k}(t)$ is $1/2^j$ and the "position" is $k/2^j$. There it is possible to find scaling functions $\phi(t)$ so that the signals in the space V_j can be synthesized from the baby scale functions $\phi_{j,k}(t)$. Thus we can decompose a signal from any space by using the sub-spaces *V* and *W* with their bases $\phi_{j,k}(t)$ and $\psi_{i,k}(t)$, so we may write:

$$(5)\quad \begin{aligned} s(t) &= \sum_q cA_0(q) \cdot \phi_{J,q}(t) = \sum_k cA_0(k) \cdot \phi_{J-1,k}(t) \\ &+ \sum_k cD_1(k) \cdot \psi_{J-1,k}(t) = A_1(t) + D_1(t) \end{aligned}$$

The signals $A_1(t)$ and $D_1(t)$ are called the *approximation* and *detail* at level 1, and $cA_1(k)$ and $cD_1(k)$ are the approximation coefficients and the details coefficients at level 1. The wavelets and the scales at each index level are orthogonal, (Phillips, 2004). These spaces are orthogonal to each other and it is possible to design any signal by writing the decomposition process of the signal as (Phillips, 2004):

$$(6)\quad \begin{aligned} s(t) &= A_1(t) + D_1(t) = A_2(t) + D_2(t) + D_1(t) \\ &= A_3(t) + D_3(t) + D_2(t) + D_1(t) = \dots \end{aligned}$$

where $D_i(t)$, in W_{-i}, is called the *detail* at level *i* and $A_i(t)$, in V_{-i}, is called the *approximation* at level *i*.

The decomposition process can be iterated, with successive approximations being decomposed in turn, so that one signal is broken down into many lower resolution components. This is called the *wavelet decomposition tree*. Other details and tutorials are presented in (Strang and Nguyen, 1996; The Wavelet Digest , 2008).

3. DENOISING PRINCIPLE

3.1. The basic model

The underlying model for the noisy signal is basically of the following form:

$$(7)\quad s(n) = f(n) + \sigma \cdot z(n)$$

where time *n* is equally spaced. In the simplest model, it is supposed that *z(n)* is a Gaussian white noise *N*(0,1) and the noise level is supposed to be equal to 1. The de-noising objective is to suppress the noise part of the signal *s* and to recover *f,* i.e. to find the best estimate $\hat{f}$ in order to minimize a quality criterion based on, e.g., the mean square error. The method is efficient for families of functions *f* that have only a few nonzero wavelet coefficients. These functions have a sparse wavelet representation. For example, a smooth function almost everywhere, with only a few abrupt changes, has such a property.

From a statistical viewpoint, the model is a regression model over time and the method can be viewed as a nonparametric estimation of the function *f* using an orthogonal basis.

The pioneering work is of Donoho, (Donoho, 1995). Methods based on multiscale decompositions consist of three main steps: First, the raw data are decomposed by means of the wavelet transform, then the empirical wavelet coefficients are shrunk through a thresholding mechanism, and finally, the denoised signal is synthesized from the processed wavelet coefficients through the inverse wavelet transform. The structure is presented in Fig.1, where **TH** stands for thresholding, **cA** for approximation coefficients and **cD** for detail coefficients.

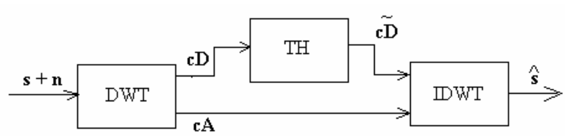

Fig.1 – The structure of the denoising process

3.2. The optimum level of decomposition

From the previous section, we have known that different levels constitute the wavelet transform. The maximum level to apply the wavelet transform depends on how many data points contain in a data set, since there is a down sampling by 2 operations from one level to the next one. A factor that affects the number of level we can reach to achieve the satisfactory noise removal results is the signal-to-noise ratio (SNR) in the original signal. Generally, for high values of the SNR values 4 or 5 for the decompositions levels are enough.

3.3. Some properties of the coefficients

Properties of the coefficients are interesting to study because such set of properties is the generator for thresholding methods and algorithms.

In general, for a one-dimensional discrete-time signal, the high frequencies influence the details of the first levels (the small values of *j*), while the low frequencies influence the deepest levels (the large values of *j*) and the associated approximations. Some known facts, (Misiti, *et al*., 2006), are:(1) If a signal comprising only white noise is analyzed, the details at the various levels decrease in amplitude as the level increases. The variance of the details also decreases as the level increases. The details and approximations are not white noise anymore, as color is introduced by the filters; (2) If the analyzed signal *s* is stationary, zero mean, and contains a white noise, the coefficients are uncorrelated; (3) If furthermore *s* is Gaussian, the coefficients are independent and Gaussian; (4) If *s* is a colored, stationary, zero mean Gaussian sequence, then the coefficients remains Gaussian. For each scale level *j*, the sequence of coefficients is a colored stationary sequence. It could be interesting to know how to choose the wavelet that would de-correlate the coefficients. This problem has not yet been resolved. Furthermore, the wavelet (if indeed it exists) most probably depends on the color of the signal. For the wavelet to be calculated, the color must be known. In most instances, this is beyond our research; (5) If *s* is a zero mean ARMA model stationary for each scale *j*, then there is also a stationary, zero mean ARMA process whose characteristics depend on *j*.

4. THRESHOLDING FUNCTIONS

The parameters of the thresholding functions are presented now but it is important to take into account that the denoising has a composed causes based on both threshold function and thresholding value. Figure 2 presents some thresholding functions.

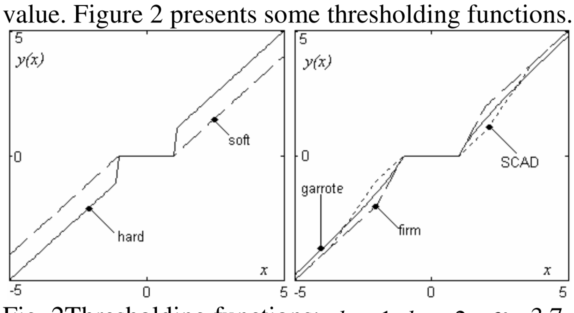

Fig. 2Thresholding functions; $th_1 = 1, th_2 = 2,\ \alpha = 3.7$

Hard thresholding is the simplest method. It can be described as the usual process of setting to zero the elements whose absolute values are lower than the threshold. As it can be seen the hard procedure creates discontinuities at $x = \pm th$. It has been shown that hard thresholding provides an improved signal to noise ratio (Jawerth and Sweldens,1993).

Soft thresholding is an extension of hard thresholding. First setting to zero the elements whose absolute values are lower than the threshold, and then shrinking the nonzero coefficients towards. The soft procedure does not create discontinuities. Soft thresholding has nice mathematical properties and the corresponding theoretical results are available, e.g. see (Donoho,1995).

Simple threshold values with *hard* thresholding results in larger variance in the function estimate, while the same threshold values with *soft* thresholding shift the estimated coefficients by an amount of threshold, creating unnecessary bias when the true coefficients are large. Also, due to its discontinuity, hard thresholding can be unstable, that is, sensitive to small changes in the data (Antoniadis, *et al*., 2001

Thresholding methods can be grouped into two categories: global thresholds and level-dependent thresholds. The former means that we choose a single value *th* to be applied globally to all empirical wavelet coefficients, while the latter means that a possibly different threshold value th is chosen for each resolution level (Antoniadis, *et al*., 2001).

5. THRESHOLD LIMITS

The detail coefficients at the finest scale are essentially noise coefficients with standard deviation equal to σ. The median absolute deviation of the coefficients is a robust estimate of σ.

Many methods for setting the threshold have been proposed. The most time-consuming way is to set the threshold limit on a case-by-case basis. The limit is selected such that satisfactory noise removal is achieved. Commonly these thresholds need an estimate of the noise level, σ. (Donoho, *et al.,* 1994) considered estimating σ in the wavelet domain, more exactly based on the coefficients at the finest resolution level, because the coefficients at this level tend to consist mostly of noise; they proposed a robust estimate of the noise level σ based on median absolute deviation. For example, for a Gaussian noise ; if we apply orthogonal wavelet transform to the noise signal, the transformed signal will preserve the Gaussian nature of the noise, which the histogram of the noise will be a symmetrical bell-shaped curve about its mean value. From theory, four times the standard deviation would cover 99.99% of the noise. Therefore, we could set the threshold be 4.5 times of the standard deviation of the wavelet-transformed signal to remove the Gaussian noise in the signal. How we should do when the noise is not gaussian? We will try to answer in the next section by choosing a criterion based on estimated probability density function.Other commonly used functions to estimate the best value of the threshold are described now.

The *universal threshold* ensures, (Antoniadis, *et al*., 2001), with high probability, that every sample in the

wavelet transform in which the underlying function is exactly zero will be estimated as zero.
It uses a fixed threshold chosen to yield minimax performance for mean square error against an ideal procedure. The minimax principle is used in statistics to design estimators. Since the de-noised signal can be assimilated to the estimator of the unknown regression function, the minimax estimator is the option that realizes the minimum, over a given set of functions, of the maximum mean square error, (Misiti,*et al,*2006;(Donoho, *et al.,* 1994; Donoho and Johnstone, 1998). Since the type of oracle (ideal observer) used has little impact on the minimax thresholds, (Antoniadis, *et al*., 2001) presents a table that can be used as a look-up table in any software.

Wavelet thresholding could suffer from artifacts of various kinds. In other words, in the vicinity of discontinuities, these wavelet thresholding estimators can alternate undershoot and overshoot of a specific target level. (Coifman,*et al*., 1995) proposed the use of the *translation invariant wavelet* thresholding scheme that helps to suppress these artifacts.

The idea of wavelet thresholding can be viewed as a *multiple hypotheses* testing. For each wavelet coefficient $\hat{d}_{jk} \sim N(d_{jk}, \hat{\sigma}^2)$ a hypothesis is tested. If it is rejected the coefficient $\hat{d}_{jk}$ is retained in the model; otherwise it is discarded. Details could be found in (Abramovich and Benjamini,1995).

Other alternative thresholding methods are: (1) methods based on *cross-validation* by minimising the mean integrated squared error (MISE) between a wavelet threshold estimator and the true function, (Nason, 1994; Donoho, *et al.,* 1995; Antoniadis, *et al*., 2001; (2) the *Sureshrink* method, (Donoho, *et al.,* 1994;Antoniadis, *et al*., 2001); (3) methods using *recursive hypothesis testing* problem, in the sense that rather than seeking to include as many wavelet coefficients as possible (subject to constraint) as in (Abramovich and Benjamini,1995), the procedure includes a wavelet coefficient only when there is strong evidence that is needed in the reconstruction, (Ogden,*et al*.,1996).
For the cases where a nonwhite noise *e* is more evident, thresholds must be rescaled by a level-dependent estimation of the level noise (Misiti, *et al*., 2006), so we have an *adaptive thresholding* of wavelet coefficients. The idea is to define level-by-level time-dependent thresholds, and then increase the capability of the de-noising strategies to handle nonstationary variance noise models.

6. DESCRIPTION OF THE METHOD

A method for threshold computation is presented now, which is based on the probability of occurrence of detail coefficients on each level of detail.
The method is applied to each level of detail. The selection criterion is based on the probability of details coefficients. Thus, all details coefficients that have a probability of appearance greater than an imposed value are selected. The method is independent of the probability density function of the noise.

At each detail level the pdf is estimated and used in the computation of the probability limits.

After removing the average we may suppose a symmetric pdf. Then, the threshold is increased until the probability of (| *X(t)*| > *th*) is less then an imposed value , let say P_i.

$$(8)\ P(|N(t)| > th) = P_i$$

The pdf is estimated by using the histogram. The range of data is divided into a number of cells of equal size and the number of data points within each cell is tabulated. The true but unknown probability density function of *X* is *w(x)*, and there are *N* measurements to be placed in *n* cells each of width *W*. Even there it is a loss of information, the approximation is accepted, (Shanmugan, *et al*., 1988).
The pseudocode of the algorithm is presented now.

```
THRESOLD_ALGORITHM
  #1: Data Inputs:
  Pi := Probability of loss ( 0.005)
  w := resolution of histogram (the cell's width)
  #2: Compute histogram
      th := 0;
  #3: LOOP th
      th := th + w;
     UNTIL Eq. (1) is satisfied.
   #4: Data output: th
END
```

The method is quite close to the problem of computation of the threshold by taking into account values which are greater than the noise level, σ, because the last conditions involves also a probability value. The advantages of the method is its generality, because does not matter the type of pdf.

7. SIMULATION RESULTS

A. Case study signals
There are four signals as case studies (1) a signal (randn) whose elements are normally distributed with mean 0, variance σ^2=1, and standard deviation σ=1, N=2000 sample (2) an ECG recorded signal with N = 2000 sample and sampling frequency 1000 Hz, (Popa, 2006); (3) a Doppler signal with noise, (Misiti, *et al*., 2006), with N=1024 samples;(4) a sound signal (mtlb) with noise (noise = cos(2*pi*3*Fs/8*(0:length(mtlb)-1)/Fs)'). The signals are presented in Fig. 3. For simplicity reasons they will be renamed as s_1, s_2,s_3 and s_4, respectively.

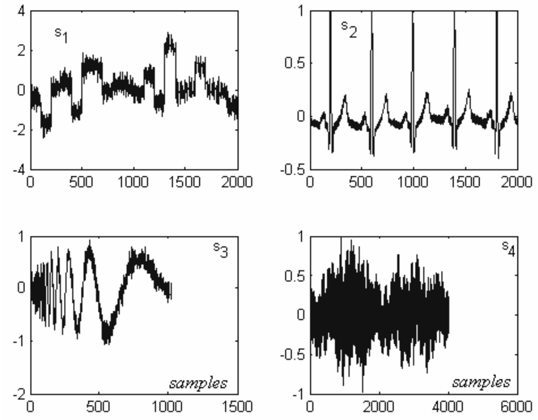

Fig.3: Test case signals

B. Results

We used soft thresholding function for all cases. We were interested in the evaluation of the probability-based function in the computation of the threshold. Two others thresholds were considered, as provided by the Matlab simulation environment, i.e. "*minimax*" and "*sqtwolog*" (universal) functions.

A function to compute the threshold parameter was described and implemented in the Matlab simulation environment. Numerical values of the simulations are presented in Table 1. For each signal – case of study, the thresholds are computed and only three levels of decomposition were considered.

Table 1 - Various thresholding values

minimax	universal	probability based
	Case 1 – s_1	
2.2166	3.7172	0.7861
2.0345	3.5266	0.7916
1.8526	3.3255	0.8388
	Case 2 – s_2	
2.2166	3.7172	0.0213
2.0345	3.5266	0.1134
1.8526	3.3255	0.3671
	Case 3 –s_3	
2.0402	3.5328	0.3291
1.8589	3.3326	0.3366
1.678	3.1201	0.6483
	Case 4 –$s4$	
2.3995	3.8992	0.6222
2.2169	3.7175	0.5366
2.0345	3.5266	0.6215

The approximation criterion is based on the Normalized Mean Squared Error (NMSE). The obtained values are presented in Table.2. The approximation obtained by using a probability based threshold computation is quite close to the results obtained by the other two functions.

Table 2 – The Mean square errors

Minimax	universal	probability based
	Case s1	
0.0246	0.0253	0.0206
	Case s2	
0.0024	0.0024	0.0023
	Case s3	
0.0153	0.0153	0.0151
	Case s4	
0.0502	0.052	0.0494

Numerical values for evolution of the mean square errors with probability of loss for s_1 are presented in Fig.4. To minimize this error, we need to have a small probability of loss. Losing the detail coefficients give a poor estimation of the signal without noise (because details coefficients are delayed) but a more accurate representation for a signal with noise.

The qualitative results presented in Fig. 5, show quite good results of denoising. By the "noise" label the original signal is presented. The "de-noised" label presents the processed signal.

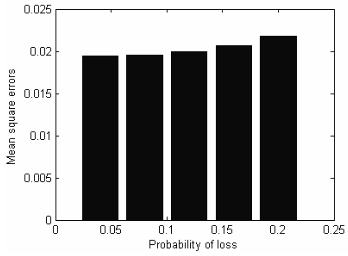

Fig. 4. MSE – Loss Probability dependency

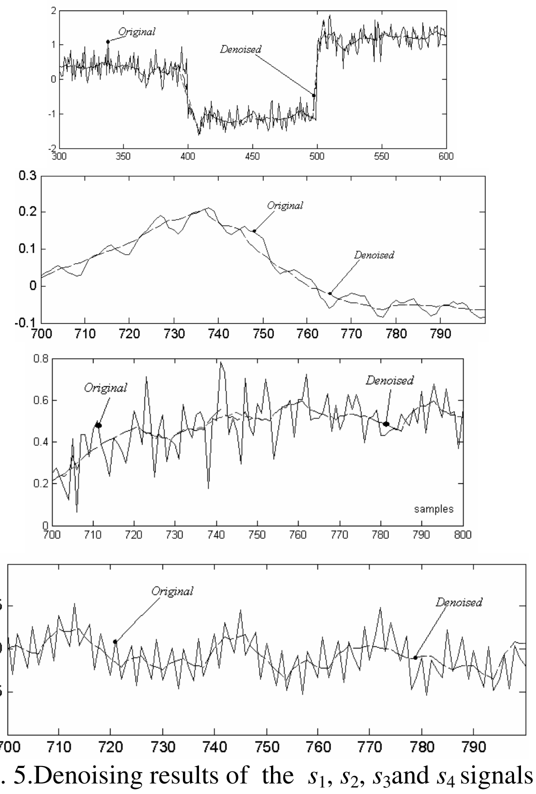

Fig. 5.Denoising results of the s_1, s_2, $s_3$and s_4 signals

8. CONCLUSIONS

In this work we developed and tested a probability based function to compute the threshold parameter by using the estimated probability density of the wavelet coefficients.

The experiments were conducted under simulation environment and the results are at least at the same quality level with other procedures currently used in denoising. In the future we will test our algorithms for 2D signals.

9. REFERENCES

Abramovich F., Trevor C. B., Sapatinas, T., Wavlet Analysis and its statistical appl., *The Statistician*, **49**, P1, pp. 1-29., 2000.

Abramovich, F. & Benjamini, Y., Thresholding of wavelet coefficients as multiple hypotheses testing procedure. In *Wavelets and Statistics*, Antoniadis, A. & Oppenheim, G. (Eds.), Lecture Notes in Statistics, Springer-Verlag, **103**, pp. 5–14, 1995.

Abramovich, F., T. Sapatinas, B.W.Silverman, *J.R. Statist. Soc. B.*, **60**, *Part A*, pp. 725-749, 1998

Achim Alin, Panagiotis Tsakalides, and Anastasios Bezerianos, SAR Image Denoising via Bayesian Wavelet Shrinkage Based on Heavy-Tailed Modeling, *IEEE Transactions On Geoscience And Remote Sensing*, **41** (8), Pag. 1773-1784, August 2003.

Antoniadis A., Bigot J. & Sapatinas, T., Wavelet Estimators in Nonparametric Regression: A Comparative Simulation Study. *Journal of Statistical Software,* **6**, Issue 6, 1-83, 2001.

Antoniadis, A. & Fan, J., Regularization of wavelets approximations . *J. Am. Statist. Ass.*, **96**, 939–967, 2001.

Bjorn Jawerth and Wim Sweldens, *An Overviw of* Wavelet *Based Multiresolution Analysis*, Research Report on DARPA Grant AFOSR 89-0455, USA, 1993.

Coifman, R.R. & Donoho, D.L., Translation-invariant de-noising. In Wavelets and Statistics, Antoniadis, A. & Oppenheim, G. (Eds.), Lect. Notes Statist., **103**, pp. 125–150, New York: Springer-Verlag, 1995.

Donoho D. L., Denoising by soft-thresholding, *IEEE Trans. Inform.Theory*, **41** (3) , pp. 613–627, May 1995.

Donoho, D.L. & Johnstone, I.M. Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Ass.*, **90**, 1200–1224, 1995.

Donoho, D.L. & Johnstone, I.M., Ideal spatial adaptation by wavelet shrinkage, *Biometrika*, **81**, 425–455, 1994.

Gao, H. Y. and Bruce, A. G., WaveShrink With Firm Shrinkage, *Statistica Sinica*, **7**, pp., 855-874, 1997

Gao, H.-Y., Wavelet shrinkage denoising using the non-negative garrote, *J. Comp. Graph. Statist.*, **7**, 469–488, 1998.

Lavielle, M., Detection of multiple changes in a sequence of dependent variables, *Stoch. Proc. and their Applications*, **83**, 2, pp. 79–102, 1999.

Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, *Wavelet Toolbox User's Guide*, 2006.

Nason, G.P., Wavelet regression by cross-validation. *Technical Report* 447, Dept. of Statistics, Standord University, USA., 1994.

Ogden, R.T. & Parzen, E.. Change-point approach to data analytic wavelet thresholding. *Statist. Comput.*, **6**, 1996, 93–99.

Popa,R., *Electronica* Medicala, Editura MATRIX-ROM, Bucuresti, pp. 110-112, 2006.

Shanmugan, K.S. and Breipohl, A.M., *Random Signals, Detection, Est. and Data Analysis*, John Willey & Sons, 1988.

Stephane Mallat and Wen Liang Hwang, Singularity detection and processing with wavelets, *IEEE Transaction Information Theory,* **38** (2), pp. 617-64 3, 1992.

The Wavelet Digest Internet, 2008.