# Wavelet Identification of Transients in Noisy Time Series

René A. Carmona[1]
Department of Mathematics
University of California at Irvine
Irvine, CA 92717, USA

Talk given at the Meeting of the Interface'93

[1]Partially supported by ONR N00014-91-1010

# 1 Introduction

The detection of transients in noisy time series is an important part of modern signal analysis because of the importance of its civil and military applications. The author addressed this problem in [1] where a new approach was proposed. The latter was based on a mixture of statistical and wavelet techniques. Nonparametric tests were implemented in the multiresolution framework of wavelet analysis. This strategy made it possible to assess the presence in the input signals of components significantly different from the noise. The multiresolution analysis gives a wide spectrum of signal to noise (SNR for short) ratios and the test procedures chose to work at the most advantageous of these resolutions. This was the main attraction of the multiresolution tests presented in [1] where it was shown that the presence of specific transients could be detected at unusually low SNR's. The main shortcoming of the method is the actual identification of *the component which is detected.* This talk addresses this weakness. We propose a new denoising procedure. The output gives the best guess for the component of the input signal which was burried in noise.

The idea of the procedure owes much to the reconstruction algorithm of Mallat and Zhong. See [8]. Their goal was to reconstruct a signal from its multiscale edge representation. We showed in [2] that their reconstruction algorithm could be simply understood as a constrained minimization problem: one has to find the unique signal which minimizes a smoothness penalty among all the signals having a prescribed wavelet transform at a certain number of points in the time-scale plane. These points were chosen as the points where the modulus of the wavelet transform of the input signal was maximum. The hope was that these points would contain most of the information contained in the wavelet transform. These reconstructions are only approximate but they do a very nice job at capturing the main features of a signal and the results are very satisfactory.

Our algorithm has two main components. The first one is devoted to the identification of the noise component of the signal and the typical effects of this noise component on the wavelet transform of the input signal. More precisely, this information on the noise is used to identify the points in the time-scale space which cannot be extrema of the wavelet transform, unless something else than noise was present in the input signal. This is done by bootstrap in general. Direct Monte Carlo simulations can be used when parametric knowledge on the distribution of the noise is available. The second part deals with the actual reconstruction of what is believed to be the component of the input which is to be identified. This part of the algorithm uses the reconstruction procedure of Mallat and Zhong (as revised by the author in [2]), the main novelty being the fact that this procedure is fed with the set of points in the time-scale plane which passed the trimming test of the extrema of the wavelet transform.

We illustrate the efficiency of our reconstruction algorithm on the examples of transients used in [1]. The results are surprisingly good. A detailed analysis of the ideas presented in this talk, together with extensions and a Bayesian interpretation of our approach will be given in the forthcoming paper [3].

# 2 Setting of the Problem

A signal (or equivalently a time series) is for us a finite sequence $\{S(t)\}$ of real numbers indexed by a discrete parameter $t$. As the terminology suggests, $t$ will have the interpretation of time. For the sake of simplicity we shall assume that all the time series have the same number of entries. We shall denote by $np$ the common length of these time series. We shall

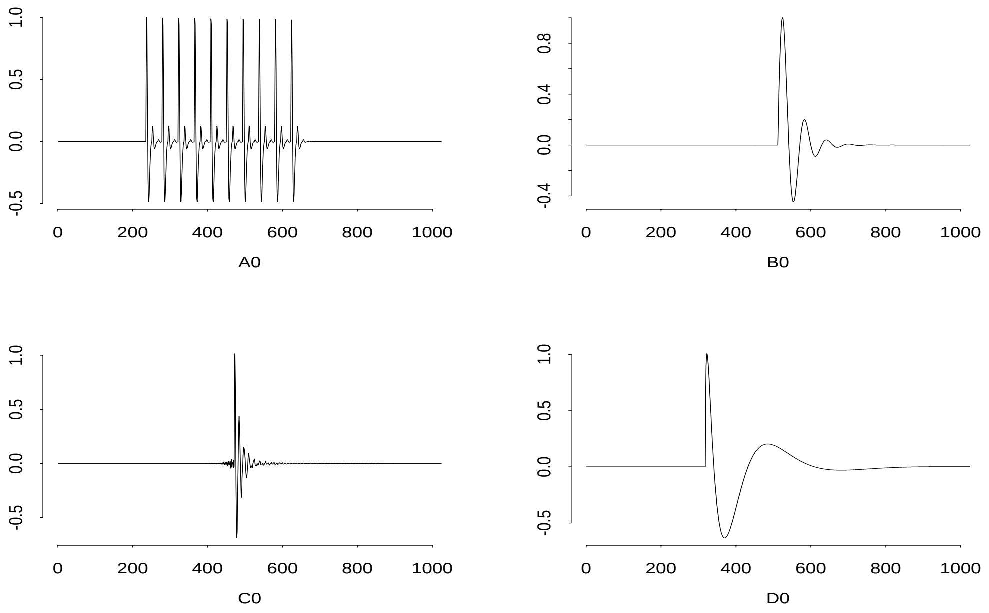


Figure 1: *Plot of the four transients $A0$, $B0$, $C0$ and $D0$ considered in this talk.*

also assume, without any loss of generality, that $np$ is a power of 2. This is achieved, if necessary, by padding the beginning and the end of the input signal with zeroes. Any given time series is supposed to be of the form:

$$S(t) = s(t) + \epsilon(t) \qquad\qquad t = 0, 1, 2, \cdots\cdots, np-1$$

where:

- $\epsilon(t)$ is a noise component which one would like to get rid of, and
- $s(t)$ is the component of the signal which is to be identified.

We assume that $s(t) \neq 0$ only for $t$ in an interval (called the support of $s$) of length significantly smaller than the length $np$ of the time series. For this reason we call $s(t)$ a *transient*. Notice that $s(t)$ can be identically 0 when there is nothing to detect. Since Fourier analysis is obviously inadequate, spline and/or wavelet analysis has been proposed as a natural alternative. See for example [6] for a discussion of the detection of underwater transient signals and [10] for a discussion of the detection of ventricular late potentials in the analysis of abnormalities of clinical electrocardiograms.

Figure 1 gives the plot of the four transients we restrict ourselves to when it comes to illustrating the results of approach. We consider only the cyclic skews, exponentially dampened sine waves, chirps and exponentially skewed nonlinear sine waves studied in [1]. We call them $A$, $B$, $C$ and $D$ respectively. In other words, we have four functions $s_A(t)$, $s_B(t)$, $s_C(t)$ and $s_D(t)$ and Figure 1 gives the plot of these four functions for specific values of the starting point $t_0$ of the transient.

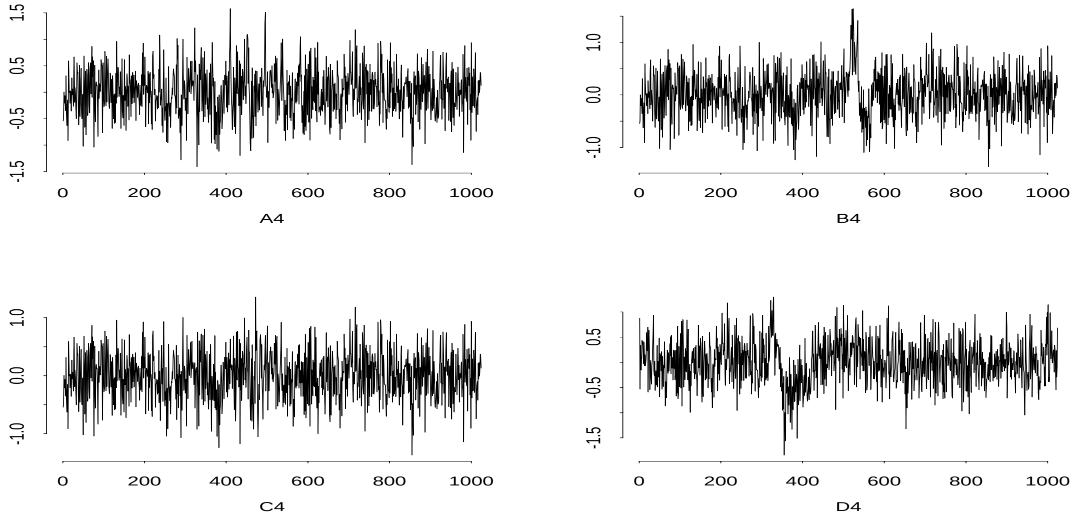


Figure 2: *Plot of the above transients after a variance .4 normal white noise was added.*

For the purpose of this talk, a typical time series will be of the form $S(t) = s(t - t_0) + \epsilon(t)$ where $s(t) = s_A(t)$, $s(t) = s_B(t)$, $s(t) = s_C(t)$, $s(t) = s_D(t)$ or $s(t) \equiv 0$ and where we write $\epsilon(t)$ for the noise component. The noise term $\{\epsilon(t)\}$ is assumed to be a sample from a sequence of independent identically distributed random variables with mean 0. We shall use the notation $\sigma^2$ for the common variance of the random variables $\epsilon(t)$. In the actual case study discussed in this talk, the noise terms are normally distributed. This assumption is not a serious restriction as we shall explain later.

Figure 2 gives the plot of the four transients of Figure 1 in the presence of a normal noise with variance .4. In this case the respective SNR's are $-11$ db, $-14$ db, $-19$ db and $-11$ db.

# 3 Wavelet Transforms

The continuous wavelet transform of a (continuos time) signal $S(t)$ of finite energy is the function $[W_\psi S](a, b)$ of the real variables $a$ and $b$ defined by:

$$[W_\psi S](a, b) = < S, \psi_{a,b} > = \int S(t)\overline{\psi_{a,b}(t)}\, dt = \frac{1}{\sqrt{a}} \int S(t)\overline{\psi(\frac{t-b}{a})}\, dt.$$

where the function $\psi_{a,b}$ is the shifted and scaled version of the original wavelet $\psi$ given by:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi(\frac{t-b}{a})$$

If one thinks of the function $\psi$ as an analyzing bump, then $\psi_{a,b}$ can be viewed as an analyzing bump rescaled by $a$ and centered around $b$. The function $\psi$ is called an analyzing wavelet. The only requirement on $\psi$ is an admissibility condition which is slightly stronger than demanding that the Fourier transform of $\psi$ vanishes at the origin. Since the wavelet coefficient $[W_\psi S](a,b)$ is the $L^2$-inner product of the original signal and $\psi_{a,b}$, it contains information about the values of $S(t)$ for times $t$ near $b$ at the scale $a$.

The main property of the continuous wavelet transform is the conservation of energy formula which becomes, once in its polarized form, a reconstruction formula: in other words, the original signal can be completely recovered from the values of its wavelet transform. Most of the successes of the wavelet transform are based on partial reconstructions.

In most applications, the coefficients $[W_\psi S](a,b)$ are computed on a grid $G = \{(a_m, b_{m,n})\}$ of the form:

$$a_m = a_0^m \qquad \text{and} \qquad b_{m,n} = a_m b_0 n$$

for integers $m$ and $n$. The choices $a_0 = 2$ and $b_0 = 1$ are the most popular. We follow this convention throughout our analysis. For the sake of convenience, we slightly change the notations and we set:

$$\psi_{m,n}(t) = 2^{-m/2}\psi(2^{-m}t - n).$$

A first remarkable fact is that it is still possible to reconstruct the original signal from its wavelet coefficients $< S, \psi_{m,n} >$ under some conditions on the analyzing wavelet $\psi$. Moreover, in order to get rid of the possible redundancy in the wavelet coefficients, it is also possible to construct the analyzing wavelet $\psi$ so that the $\psi_{m,n}$ form an orthonormal basis. Finally, it is even possible to construct the analyzing wavelet $\psi$ in such a way that it has compact support. There are excellent books on the subject. See for example [9], [4] and/or [5].

The numerical examples presented in this talk have been computed in the framework of finite sequences extended by periodicity. If necessary, we pad the signals with zeros in order to make sure that their lengths are powers of 2. The wavelet transforms are computed with the help of finite filters. For the sake of comparison we chose to use the filters given in [8] by Mallat and Zong. The corresponding wavelet is not orthogonal but it has a small support and it has the advantage to be symmetric.

# 4 Extrema Representations and Reconstructions

We follow the idea of [8]. At each resolution $j = 1, \cdots, J$ we consider the sequence of couples $\{(x_{j,n}, X_{j,n})\}_{n\geq 1}$ where $x_{j,n}$ is the location of a local maximum of the absolute value of the wavelet transform at the resolution $j$ and $X_{j,n}$ is the actual value of the wavelet transform at the point $x_{j,n}$ (and resolution $j$). Figure 3 gives the wavelet transform of C0 and the extrema representation of the same signal.

The operation which consists in computing the extrema representation of a signal is essentially reversible. Indeed, as demonstrated in [8], the knowledge of the extrema $(x_{j,n}, X_{j,n})$ for $1 \leq j \leq J$ and $n \geq 1$ is sufficient to reconstruct a very good approximation of the original signal. As explained in [2], the idea of the reconstruction is simple. One looks for the signal which minimizes a given smoothness penalty function among all the signals whose wavelet transform at resolution $j$ is equal to $X_{j,n}$ at $x_{j,n}$. The solution of this elementary constrained optimization problem has a a nice interpretation. The quadratic form defining the smoothness penalty defines a duality notion and the latter can be used to determine a dual wavelet. The solution is obtained by superposing contributions given by the individual extrema. Each extremum, say $(x_{j,n}, X_{j,n})$ gives rise to a bumb at the point $x_{j,n}$ obtained by

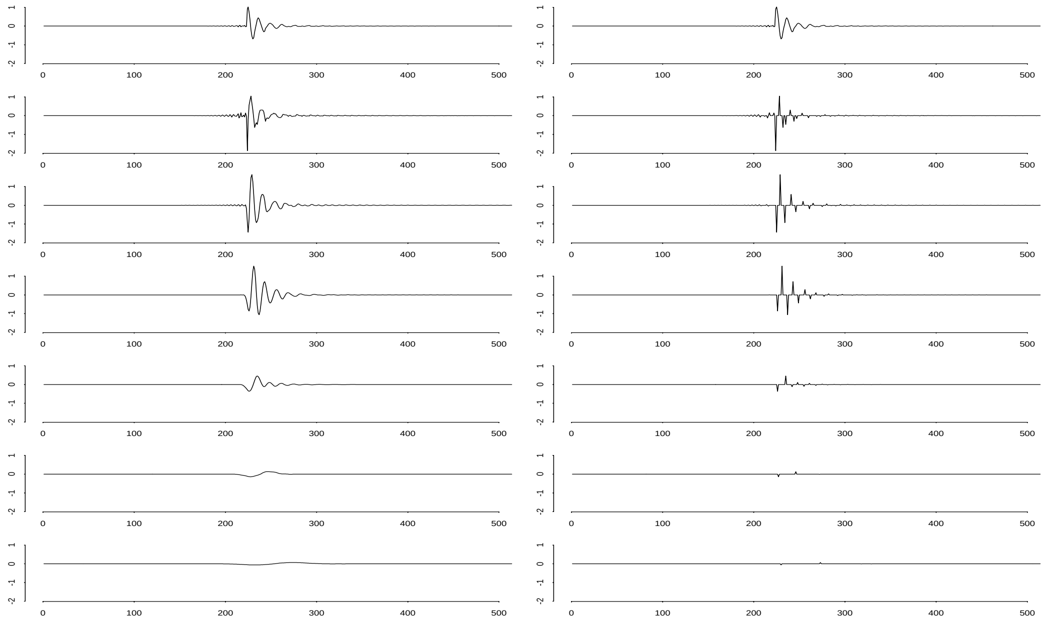


Figure 3: *The transient C0 and its wavelet transform are given in the column on the left. Its extrema representation is given in the rightmost column.*

scaling the dual wavelet according to the resolution $j$ and of amplitude given by the value of the Lagrange multiplier of the problem. The results are surprisingly good.

Figure 4 shows the reconstruction of the signal C0. The original C0 is on the first row, the second row gives the reconstruction from the extrema representation and the third row gives the superposition of the original and the reconstructed signal. The lack of color makes it difficult to notice that this third row actually contains the graphs of two signals!

# 5 Trimming the Extrema Representation of a Noisy Signal

As explained earlier, we assume that the input signal $S(t)$ has a noise component $\epsilon(t)$ which forms a realization of an i.i.d. sequence of mean zero random variables having a common variance $\sigma^2$. We also assume that the other component is a transient in the sense that its support is a small localized part of the signal. For the sake of simplicity we assume that the first part of the signal is free of transient. In other words, we assume the existence of an integer $wl$ such that the *window* $S(0)$, $S(1)$, $\cdots$ , $S(wl-1)$ of length $wl$ is a sample from the noise. We subtract a running mean from this shorter time series and we use it to generate $B$ samples $S^{(b)}(0)$, $S^{(b)}(1)$, $\cdots$ , $S^{(b)}(wl-1)$ indexed by $b = 1, \cdots, B$ of the same length $wl$ by bootstrap. We use direct Monte Carlo simulations when we have some parametric information on the distribution of the noise. In particular, if we know that the sequence $\{\epsilon(t);\ t = 0, \cdots, np-1\}$ is a normal sample then we use the initial window to estimate the variance of this sequence, say $\hat{\sigma}^2$, and we generate $B$ samples of lengths $wl$

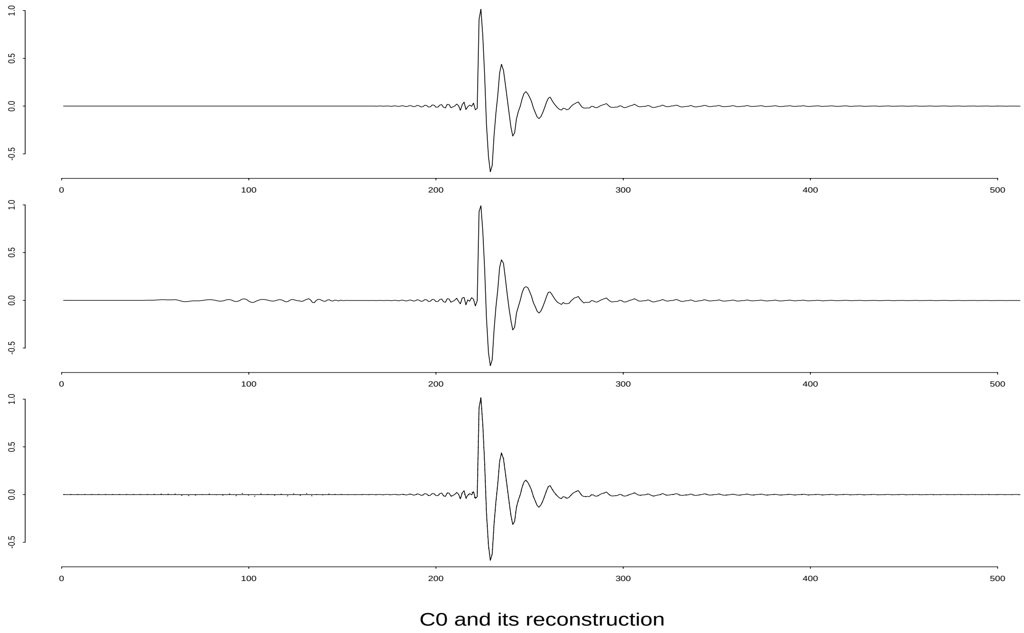


Figure 4: *Top row: plot of the original transient C0. Middle row: plot of the reconstruction from the extrema representation. Bottom row: superposition of C0 and its reconstruction.*

from the normal distribution $N(0, \hat{\sigma}^2)$. For each of the samples produced by simulation we compute the wavelet transform and the extrema representation. Then we construct, for each resolution $j = 1, \cdots, J$, a histogram $\mathcal{H}_j$ for the size of a typical extremum at the resolution $j$. Let us choose a level of significance $\alpha$ in the interval $(0, 1)$ and let us denote by $h_j(\alpha)$ the $100(1-\alpha)$ percentile computed from the histogram $\mathcal{H}_j$. We use $\alpha = .05$ in the analysis reported in this talk.

We trim the list of extrema of the wavelet transform of $S(t)$ in the following way. If an extremum appears at resolution $j$, we compare its absolute value to $h_j(\alpha)$. If its absolute value is smaller than $h_j(\alpha)$, the extremum is suspected to be a result of the noise component of the signal and it is deleted form the list of extrema. If on the other hand, its absolute value is greater than or equal to $h_j(\alpha)$, then the extremum is kept because there is no statistical evidence (at the level $\alpha$) that it is a result of the noise component $\epsilon(t)$. Figures 5 and 6 give the extrema representation of the noisy signals A4 and C4 before and after the trimming.

In other words, the trimmed extrema representation is the list of the extrema (from the extrema representation of the input signal $S(t)$) for which the null hypothesis $H_0: \; s(t) \equiv 0$ is rejected at the level $\alpha$ by a bootstrap test based on the test statistic given by the absolute value of the extremum in question.

**Remark**

Notice that a statistical detection test can be based on the trimmed extrema representation.

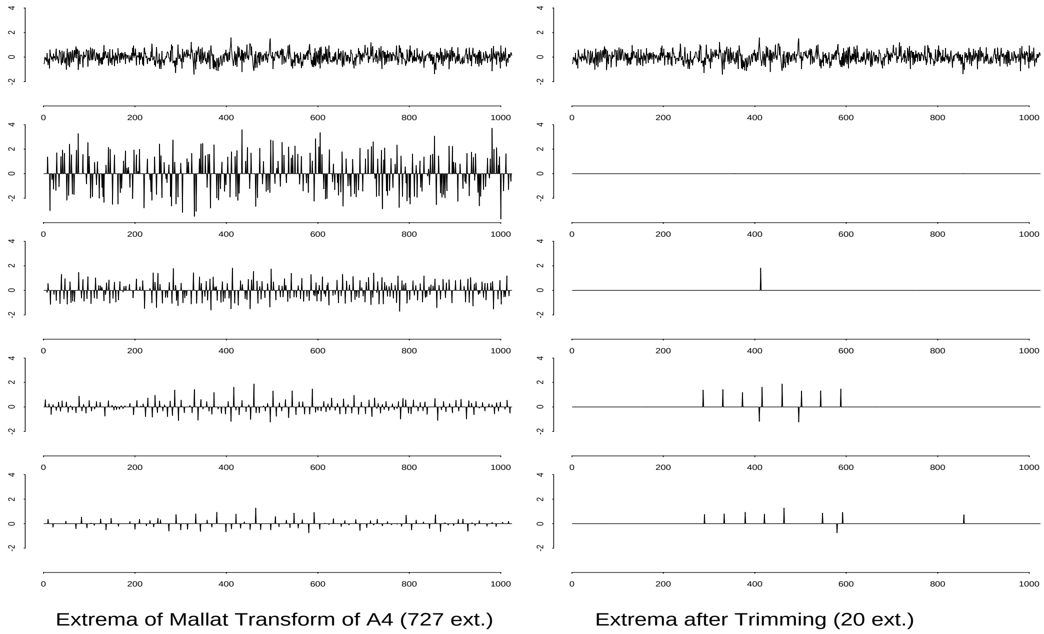


Figure 5: *The extrema representation of the noisy signal A4 is given on the left. The same extrema representation is given on the right after trimming.*

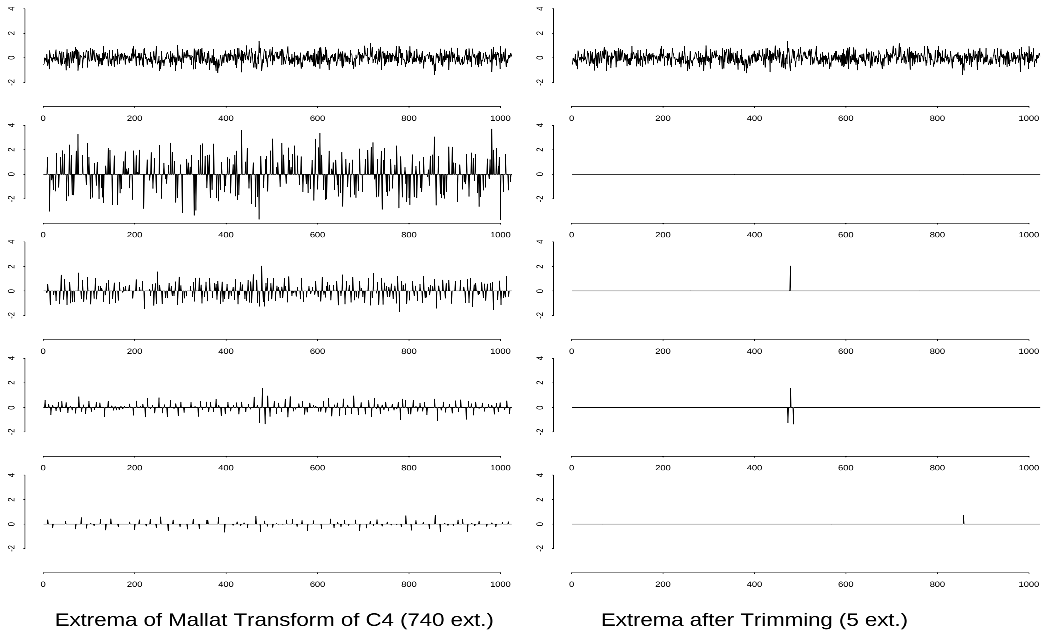


Figure 6: *The extrema representation of the noisy signal C4 is given on the left. The same extrema representation is given on the right after trimming.*

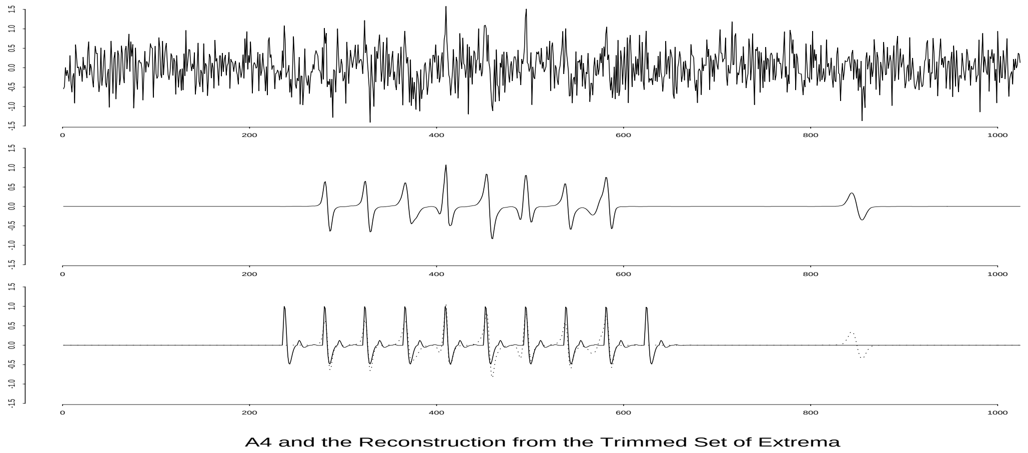


Figure 7: *Top row: plot of the noisy signal A4. Middle row: plot of the reconstruction from the trimmed extrema representation. Bottom row: superposition of the pure transient A0 and its reconstruction from the trimmed extrema representation.*

Indeed, the presence of extrema in the trimmed list is a sign that there is something else than noise in the input signal $S(t)$. This qualitative remark can be given a quantitative version and a specific test can be derived if one so desires. We shall refrain from doing that because the problem of the multiresolution detection was addressed in details in [1].

# 6 Reconstruction after Trimming

The reconstruction procedure works from an extrema representation. If we use it with the extrema representation of the input signal $S(t)$ we reconstruct an approximation of $S(t)$ including its noise component $\epsilon(t)$. If on the other hand we run the reconstruction program on the trimmed extrema representation, it is reasonnable to expect that most of the contribution to $S(t)$ due to the noise component $\epsilon(t)$ will not be reconstructed, giving rise de facto to a denoising procedure of the input signal.

Figures 7 and 8 give the results of this denoising procedure when the input signal $S(t)$ is chosen to be the signals A4 and C4 from the examples given in Figure 2.

As before, the original (noisy) signal is on the first row, the second row gives the reconstruction from the trimmed extrema representation and the third row gives the superposition of the pure transient to identify and the reconstructed signal. These results illustrate the great potential of the denoising method.

# 7 Concluding Remarks

The identification algorithm proposed in this talk can be summarized as follows:

- identification of the points $x_{j,n}$ in the time-scale plane at which the wavelet transform of the input signal $S(t)$ is significantly different from the wavelet transform of pure

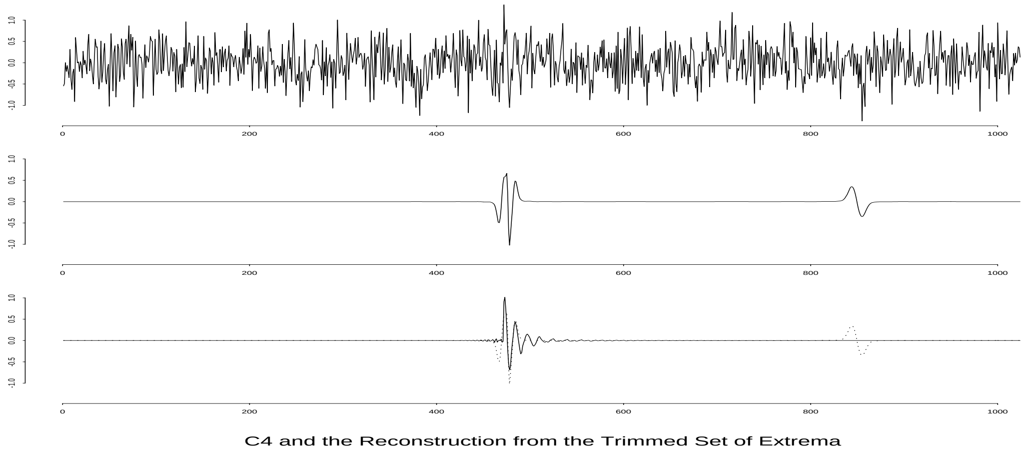


Figure 8: *Top row: plot of the noisy signal C4. Middle row: plot of the reconstruction from the trimmed extrema representation. Bottom row: superposition of the pure transient C0 and its reconstruction.*

noise.

- estimation of the transient $s(t)$ from the values $X_{j,n}$ of the wavelet transform of the $S(t)$ at the time-scale points $x_{j,n}$.

The identification of the significant features of the wavelet transform of $S(t)$ is done by means of a bootstrap test on the size of a given extremum of the absolute value of the wavelet transform. Other test statistics could be used. More importantly, the estimation of $s(t)$ is done by computing the unique signal which minimizes a smoothness penalty among all the signals whose wavelet transform is equal to $X_{j,n}$ at the point $x_{j,n}$. The model $S(t) = s(t) + \epsilon(t)$ says that $X_{j,n}$ is equal to the wavelet transform of $s(t)$ at the point $x_{j,n}$ **PLUS** the value of the wavelet transform of $\epsilon(t)$ at the same time-scale point $x_{j,n}$. In other words, the observations $X_{j,n}$ are corrupted by noise. The noise terms are strongly correlated but this correlation can be controlled since it depends only on the unknown variance $\sigma^2$ and the wavelet filter used in the analysis. If one adopts this point of view, the second part of our identification algorithm can be viewed as a particular case of the so-called *generalized smoothing spline* problem. See [11] for a complete exposition of the solution. This new emphasis will be main theme of the forthcoming paper [3] where the preliminary results presented here will be extended.

*Acknowledgments*

I would like to thank A. Wang for writing the code and producing the graphs included in this paper.

# References

[1] R. Carmona (1992): Wavelet Identification of Transients in Noisy Time Series (preprint)

[2] R. Carmona (1993): Variations on the Reconstruction Algorithm of Mallat & Zhong. (preprint)

[3] R. Carmona (1993): Bayesian Approach to Wavelet Detection of Transients. (in preparation)

[4] C.K. Chui (1992): An Introduction to Wavelets. *Academic Press.*

[5] I. Daubechies (1992): Ten Lectures on Wavelets. *SIAM.*

[6] H.T. Le and E.J. Wegman (1991): Testing the equality of two regression curves using linear smoothers. *J. Acoust. Soc. Amer.* ***89***, 274-279.

[7] S.G. Mallat and W.L. Hwang (1992): Singularity Detection and Processing with Wavelets. *IEEE Trans. Info. Theory* ***38***, 617-643.

[8] S.G. Mallat and S. Zhong (1992): Characterization of Signals from Multiscale Edges. *IEEE Trans. Pattern Anal. Machine Intell.* (to appear).

[9] Y. Meyer (1990): Ondelettes et Opérateurs. I *Hermann.*

[10] F.B. Tuteur (1988): Wavelet Transformations in Signal Detection. (preprint).

[11] G. Wahba (1988): Spline Models for Observational Data. CBMS-NSF Reg. Conf. Ser. in Applied Math. # 59. *SIAM.*