The Discrete Multiple Wavelet Transform and Thresholding Methods

T. R. Downie[*] B. W. Silverman[†]

September 18, 1996

SP EDICS: SP 2.4.4 (Wavelets and Filter Banks)

Permission to publish this abstract seperately is granted.

Abstract

Orthogonal wavelet bases have recently been developed using multiple mother wavelet functions. Applying the discrete multiple wavelet transform requires the input data to be preprocessed to obtain a more economical decomposition. We discuss the properties of several preprocessing methods and their effect on the thresholding results. We propose a multivariate thresholding method for multiwavelets where each wavelet coefficient vector is thresholded as opposed to each element. A multiwavelet universal threshold is obtained using the asymptotic behaviour of the χ^2 distribution. A simulation study demonstrates that when using the Geronimo multiwavelet with an appropriate preprocessing method and bivariate thresholding, better results are obtained compared with using Daubechies wavelets.

1 Introduction

The development of wavelet theory has given rise to practical methods of data smoothing and data compression. These methods in general give better results than other non-parametric smoothing methods such as spline smoothing or Fourier analysis and uses a computationally fast algorithm. The idea is to express a function in terms of an orthonormal wavelet basis, where the basis functions are translations and dilations of a *mother wavelet function*. The wavelet transformation is often good at compressing a signal into a few large coefficients. If the original signal contains noise then a thresholding method will throw away the small 'noisy' coefficients and keep the large signal coefficients. A brief introduction to the theory of wavelet methods is given in Nason & Silverman [12]. For a more comprehensive and theoretical account, consult Daubechies [3].

[*]School of Mathematics, University of Bristol, Bristol BS8 1TW, email: *tim.downie@bristol.ac.uk*.

[†]as above

Multiple wavelets [6] have recently been formulated, which use translates and dilates of more than one mother wavelet function. This paper describes a method of thresholding such multiple wavelets, by considering the magnitude of the wavelet coefficient vector rather than its individual elements. A synopsis of single and multiple wavelet methods is given in Section 2.

One major difference between single and multiple wavelets is that the data have to be preprocessed before applying the discrete multiple wavelet transform. Section 3 discusses the properties of some prefilters and how these affect thresholded estimates and other wavelet applications. Section 4 derives the thresholding rules and the multivariate universal threshold. The implementation and performance of this method has been assessed using simulation; the results are documented in Section 5.

2 Single and Multiple Wavelet Theory

2.1 Single Wavelets

A standard wavelet is a univariate function $\psi : \mathbf{R} \to \mathbf{R}$, whose translates and dilations generate a suitable functional basis. A wavelet has short support, integrates to zero and has an associated *scaling function* ϕ. We shall also require that all integer translations and dyadic dilations of the wavelet and integer translates of the scaling function are orthogonal.

Using j to indicate the dilation, and k the translation of ψ and ϕ, define $\phi_{J,k} = 2^{-\frac{J}{2}}\phi(2^{-J}x - k)$ and $\psi_{j,k} = 2^{-\frac{j}{2}}\psi(2^{-j}x - k)$, for $j, J, k \in Z$. We approximate a function f in terms of a wavelet expansion

$$f(x) \approx \sum_{k\in Z} c_{J,k}\phi_{J,k}(x) + \sum_{j=0}^{J-1}\sum_{k\in Z} d_{j,k}\psi_{j,k}(x). \qquad (1)$$

For two functions to be a valid wavelet and scaling function pair they must satisfy several conditions including the dilation equations,

$$\phi(x) = \sum_{k\in Z} h_k\phi(2x-k) \quad \text{for some } h_k \in R,$$

and

$$\psi(x) = \sum_{k\in Z} g_k\phi(2x-k) \quad \text{for some } g_k \in R.$$

Under our assumption of orthogonality on $\psi_{j,k}$ we also require that ψ and ϕ are orthogonal. Under these conditions it can be shown that the filter coefficients, h_k and g_k, also satisfy the following condition,

$$g_k = (-1)^k h_{1-k}. \qquad (2)$$

The discrete wavelet transform DWT (Mallat [10]) takes observations from a function f at 2^J equally spaced points, and recursively computes the wavelet coefficients using the following algorithm.

Assign $c_{0,k} = f(x_k),\ k = 0, \cdots, 2^J - 1$.
Then for $j = 1, \cdots, J$ recursively compute

$$c_{j,k} = \sqrt{2} \sum_n h_n c_{j-1,n+2k} \tag{3}$$

and

$$d_{j,k} = \sqrt{2} \sum_n g_n c_{j-1,n+2k}. \tag{4}$$

To reconstruct fully $f(x_k)$, apply the inverse discrete wavelet transform IDWT.

Given the wavelet coefficients $d_{j,k}, k = 0, \cdots, 2^j - 1, j = 1, \cdots, J$ and $c_{J,0}$ recursively compute $c_{j,n}$ for $j = J-1, \cdots, 0$ using

$$c_{j,n} = \sqrt{2} \sum_k (h_n c_{j+1,n+2k} + g_n d_{j+1,n+2k}).$$

The original data is then $f(x_k) = c_{0,k}$.

We shall asume that the starting coefficients $c_{0,k}$ correspond to sampling at the integers ($c_{0,k} = f(k), k = 0, \ldots, 2^J - 1$), giving the wavelet representation in Equation 1. If the signal is sampled at another rate $x_k = k/T$, for fixed T and $k = 0, \ldots, 2^{J'} - 1$, then the wavelet expression in Equation 1 is simply rescaled by replacing $f(x)$ with $f(x/T)$. Thus, the accuracy of an approximation can be increased by sampling at closer intervals. Note that other authors may use a different indexing convention (eg. [4] [12]), with level $j = 0$ corresponding to the lowest frequencies and level $j = J$ corresponding to the highest frequencies. This is the converse of our notation.

Each wavelet coefficient $d_{j,k}$ indicates the frequency component of f at the jth resolution level within a neighbourhood around $2^j k$. If f is sufficiently regular within an interval, all the coefficients whose corresponding wavelet lies within the interval will be zero. So for many functions of practical importance a wavelet decomposition will contain many zero coefficients with non zero ones appearing only at low frequencies, "near" discontinuities or other "irregularities". It is for this reason that wavelet methods are good at data compression and data smoothing. To smooth a noisy function all coefficients with a magnitude less than a specific value or *threshold* are assumed to represent only the noise component whereas the larger coefficients contain some signal component and should be kept. Full details of thresholding methods are given in Section 4.

How well these methods work depend in part on the properties of the wavelet function itself. In general, better signal compression is possible when the wavelet function is symmetric, has short support, and has a large number of vanishing moments (the latter determines the smoothness of the thresholded estimate). Daubechies [2] shows that the support, symmetry and number of vanishing moments are linked and that for orthogonal wavelets, an improvement in say vanishing moments would be offset by longer support or by worse symmetry.

2.2 Multiple Wavelets

A multiple wavelet basis uses translations and dilations of $L \geq 2$ scaling functions $\phi_1, \cdots, \phi_L$, and L wavelet functions $\psi_1, \cdots, \psi_L$. The expression of a function in terms of a multiple wavelet basis is

$$f(x) \approx \sum_{k \in Z} \sum_{l=1}^{L} c_{J,k,l} 2^{-\frac{J}{2}} \phi_l(2^{-J}x - k) + \sum_{j=1}^{J} \sum_{k \in Z} \sum_{l=1}^{L} d_{j,k,l} 2^{-\frac{j}{2}} \psi_l(2^{-j}x - k).$$

Assume that $L = 2$ unless stated otherwise; generalisations to more than two wavelet functions are straightforward. We can write the above equation in vector notation, as

$$f(x) \approx \sum_{k \in Z} C_{J,k}^{\top} 2^{-\frac{J}{2}} \Phi(2^{-J}x - k) + \sum_{j=1}^{J} \sum_{k \in Z} D_{j,k}^{\top} 2^{-\frac{j}{2}} \Psi(2^{-j}x - k),$$

where

$$\Phi(x) = \begin{bmatrix} \phi_1(x) \\ \phi_2(x) \end{bmatrix}, \ \Psi(x) = \begin{bmatrix} \psi_1(x) \\ \psi_2(x) \end{bmatrix}, \ C_{j,k} = \begin{bmatrix} c_{j,k,1} \\ c_{j,k,2} \end{bmatrix}, \text{ and } D_{j,k} = \begin{bmatrix} d_{j,k,1} \\ d_{j,k,2} \end{bmatrix}.$$

We shall assume that each wavelet is orthogonal to translations and dilations of itself *and* translations and dilations of every other wavelet. Ie.,

$$\int \psi_l(2^{-j}x - k)\psi_{l'}(2^{-j'}x - k')dx = \begin{cases} 1 & \text{for } j = j', k = k', l = l' \\ 0 & \text{otherwise.} \end{cases}$$

The scaling function vector and wavelet vector satisfy dilation equations, similar to the those for single wavelets,

$$\Phi(x) = \sum_{k \in Z} H_k \Phi(2x - k),$$
$$\Psi(x) = \sum_{k \in Z} G_k \Phi(2x - k),$$

where H_k and G_k are 2×2 matrices.

The discrete multiple wavelet transform (DMWT), is analogous to the DWT and is given by,

$$C_{j,k} = \sqrt{2} \sum_{n} H_n C_{j-1,n+2k}$$

and

$$D_{j,k} = \sqrt{2}\sum_n G_n C_{j-1,n+2k}$$

recursively for $j = 0, \cdots, J$.

So for a starting set of scaling function coefficients $C_{0,k}$ the wavelet coefficients can be computed recursively to give $C_{J,k}$ and $D_{j,k}$. The inverse DMWT is given by

$$C_{j-1,n} = \sqrt{2}\sum_k (H_n^\top C_{j,n+2k} + G_n^\top D_{j,n+2k}).$$

Note that the inverse DMWT uses the transpose of the coefficient matrices, because orthogonality gives us that $\sum_k H_k H_k^\top = 1$. For the inverse (single wavelet) DWT the transpose is unnecessary, as the coefficients are scalars.

Another important difference between single and multiple wavelet bases is that given the set of filter coefficients, H_k, there is no simple formula for computing the G_k [14]. In the single wavelet case, using discrete Fourier transforms for g_k and h_k, it is possible to express the orthogonality conditions and the dilation equations in the Fourier domain. Simple arithmetic on these conditions gives the formula shown in Equation 2. However the discrete Fourier transforms for the matrix coefficients are also matrices. An explicit formula for G_k cannot now be formed as commutativity, used in the single wavelet example, no longer holds.

The motivation for developing multiple wavelet bases is that it is possible to have wavelet functions which are mutually orthogonal, symmetric, have short support and have a higher degree of accuracy than a single wavelet of the same support [14]. This would suggest that multiple wavelet bases will give better noise reduction and signal compression ratios than single wavelets, as demonstrated by Strela *et al* [15] and the results in Section 5. The methods developed in the following sections apply to any general orthogonal multiple wavelet bases. Examples given in this paper will use the double wavelet basis derived by Geronimo, Hardin and Massopust [6] using the high pass filter coefficients first published in Strang & Strela [14], shown in Figure 1. The matrix coefficients for the Geronimo wavelets are given by

$$H_0 = \frac{1}{10}\begin{bmatrix} 3\sqrt{2} & 8 \\ \frac{-1}{2} & \frac{-3}{\sqrt{2}} \end{bmatrix} \quad H_1 = \frac{1}{10}\begin{bmatrix} 3\sqrt{2} & 0 \\ \frac{9}{2} & 5\sqrt{2} \end{bmatrix} \quad H_2 = \frac{1}{10}\begin{bmatrix} 0 & 0 \\ \frac{9}{2} & \frac{-3}{\sqrt{2}} \end{bmatrix} \quad H_3 = \frac{1}{10}\begin{bmatrix} 0 & 0 \\ \frac{-1}{2} & 0 \end{bmatrix}$$

$$G_0 = \frac{1}{10}\begin{bmatrix} \frac{-1}{2} & \frac{-3}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 3 \end{bmatrix} \quad G_1 = \frac{1}{10}\begin{bmatrix} \frac{9}{2} & -5\sqrt{2} \\ \frac{-9}{\sqrt{2}} & 0 \end{bmatrix} \quad G_2 = \frac{1}{10}\begin{bmatrix} \frac{9}{2} & \frac{-3}{\sqrt{2}} \\ \frac{9}{\sqrt{2}} & -3 \end{bmatrix} \quad G_3 = \frac{1}{10}\begin{bmatrix} \frac{-1}{2} & 0 \\ \frac{-1}{\sqrt{2}} & 0 \end{bmatrix}.$$

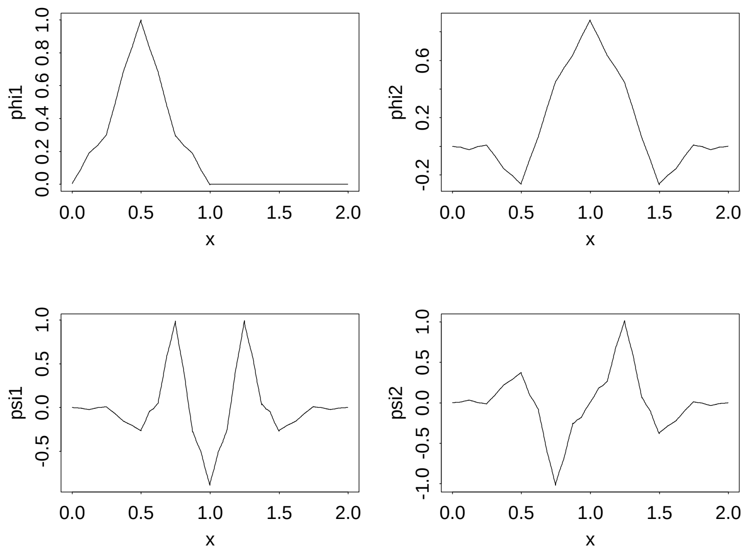

Figure 1: The scaling functions and mother wavelets for the Geronimo multiwavelet.

3 Preprocessing

In the previous section we described the discrete multiple wavelet transform. The algorithm requires that the starting coefficients ($C_{0,k}$) are expressed as vectors but usually the signal is single valued ($f(x_k)$). Clearly a method has to be used that will map this univariate data to a bivariate set of vectors. In a more general setting, a *preprocessor* is a linear operator $\mathcal{Q}$ that maps $f(x_k)$ to the starting coefficients $c_{0,k}$ or $C_{0,k}$. This operator usually takes the form of a filter, termed a *prefilter*. As described below preprocessing is seldom considered for single wavelets but is important for multiple wavelets

3.1 Preprocessing Single Wavelets

Many texts on single wavelets advocate that each coefficient be of the form $c_{j,k} =< f, \phi_{j,k} >$ and $d_{j,k} =< f, \psi_{j,k} >$. This results in the wavelet approximation (Equation 1) giving the best approximation of f (Jawerth and Sweldens [8],Daubechies [3]). If f is only observed at a discrete set of points, these coefficients can only be approximated. However to implement the DWT, it is enough to know the coefficients $c_{0,k}$, as all the other coefficients can be computed directly using the equations 3 and 4. Thus the problem reduces to approximating the "starting coefficients" which are most accurately estimated using,

$$c_{0,k} = \sum_n f(x_n)\phi(x_n - k) \approx < f, \phi(\cdot - k) > .$$

This can be thought of as applying a prefilter $\mathcal{Q}$, whose coefficients are $q_n = \phi(n)$, $n \in Z$, ie.

$$c_{0,k} = \sum_n q_n f(x_{n+k}).$$

We shall refer to a prefilter defined as above as the interpolation prefilter, since the wavelet reconstruction interpolates $f(x_k)$.

In practice, however, the interpolation prefilter is hardly ever used. Instead the starting coefficients are defined to be the same as the observed data, ie $c_{0,k} = f(x_k)$. This can be thought of as applying the identity prefilter, ($q_0 = 1$ and $q_n = 0 \,\text{for}\, n \neq 0$). There are several reasons why the identity prefilter is usually used. Firstly, it is easier to program and compute than the interpolation prefilter and is independent of the wavelet/scaling function. The interpolation prefilter has more coefficients and so more wavelet coefficients will contain (non-zero) signal components. It can be shown that any polynomial that gives all zero wavelet coefficients using the interpolation prefilter also does so when using the identity prefilter. Finally, the frequency response for the identity prefilter is trivial and does not affect the high and low pass properties of the DWT filters. Hence, normally the trivial identity prefilter is used, but the concept of a prefilter is not usually acknowledged. This explains why prefilters have been so poorly studied in the single wavelet literature.

3.2 Prefilters and Repeated Signals for Multiple Wavelets

3.2.1 Example

Suppose we were to use the DMWT without any prefilter, but just assigning the data in pairs to the starting coefficients (which using the prefilter terminology is equivalent to the identity prefilter). If the signal is piecewise constant (Figure 2 top left) then, whatever type of wavelet is used, we would expect only those coefficients whose accompanying wavelets span the discontinuity to be non-zero. But when the identity prefilter is used (top right) all the coefficients in level 1 have non-zero elements. This is catastrophic when trying to compress the signal, and is just bad for thresholding, as all these small non-zero coefficients are reduced to zero introducing a systematic high frequency component after reconstruction (as shown in the bottom left figure). However, by using a suitable prefilter only coefficients whose accompanying wavelets span the discontinuity are non-zero, as shown in the bottom right hand figure.

3.2.2 Definitions

There are two general forms for signal preprocessing, the *prefilter* method partitions the data into a sequence of two vectors and applies the filter $\mathcal{Q}$ which is defined by a sequence of 2×2 matrices

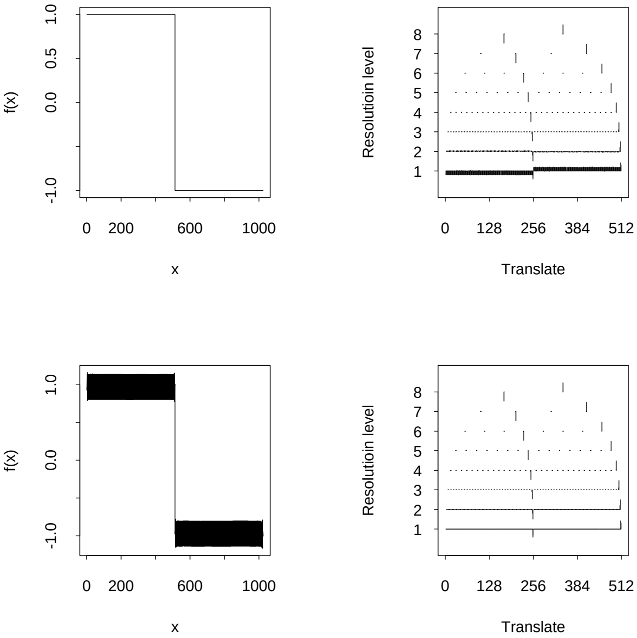

Figure 2: The need for a prefilter: top left) a piecewise constant signal, top right) the resulting wavelet coefficients using identity prefilter, bottom left) reconstruction of thresholded wavelet coefficients using identity prefilter, bottom right) wavelet coefficients using interpolation prefilter

Q_n. Ie.

$$C_{0,k} = \sum_n Q_n \begin{bmatrix} f_{2(n+k)} \\ f_{2(n+k)+1} \end{bmatrix}. \tag{5}$$

The second method convolves a sequence of 2-vectors with the (scalar) observations. This is equivalent to sampling each observation twice, using different weights. Given a filter Γ of vectors γ_n, the starting coefficient are,

$$C_{0,k} = \sum_n \gamma_n \cdot f_{n+k}$$

We shall call this method of preprocessing a *repeated signal* filter, to emphasise the different approach from the matrix prefilters above. Because a repeated signal oversamples there are twice as many vector coefficients in the DMWT decomposition, as compared to using a prefilter.

3.2.3 Recovering the Signal

To recover the signal after reconstruction the output coefficients have to be postprocessed. The postfilter $\mathcal{P}$ that accompanies the prefilter $\mathcal{Q}$ must satisfy $\mathcal{PQ} = \mathcal{I}$, where $\mathcal{I}$ is the identity filter. So if one applies a prefilter, DMWT, Inverse DMWT and postfilter to any sequence the resulting output will be identical to the input. Any valid prefilter must have a well defined and unique postfilter.

The postprocessing that complements a repeated signal takes the form of a filter of row vectors. This restores the output signal to one dimension.

3.3 Criteria for Assessing Preprocessing Methods

In the next subsection we present four prefilters and one repeated signal and compare their suitability as preprocessing methods for the Geronimo multiwavelets. To assess the performance of a prefilter or repeated signal we propose three different criteria: the length, degree of approximation and frequency response of a prefilter or repeated signal.

3.3.1 Length of a prefilter

The length of a prefilter is simply the span of the non-zero coefficients in the filter. It is desirable for a prefilter or repeatedsignal to have as small a length as possible to preserve the time localisation of the wavelet decomposition.

3.3.2 Degree of a prefilter

The degree of approximation of a prefilter or repeated signal is similar to the notion of vanishing moments of the wavelet functions. If a single wavelet ψ has v vanishing moments then the minimum degree polynomial $p(x)$ for which $\int p(x)\psi(x)dx \neq 0$ is v. The DWTs of all polynomials of degree less than v have all their wavelet coefficients equal to zero (except perhaps at boundaries). With multiple wavelets the number of vanishing moments of the vector wavelet $\Psi(x)$ is simply the minimum number of vanishing moments of ψ_1 and ψ_2. In Example 3.2.1 we showed that a piecewise constant function did not give zero wavelet coefficients after applying the DMWT, even though the Geronimo Multiwavelets have 2 vanishing moments [13]. We introduce the *degree* of a prefilter $\mathcal{Q}$ (or repeated signal Γ), as the lowest degree of polynomial which gives non-zero wavelet coefficients, after preprocessing by $\mathcal{Q}$ (or Γ) and transformation by the DMWT. In the above example the identity prefilter has degree zero, (constant signal gives non-zero coefficients).

3.3.3 Frequency Response of Prefilters

We assume that the first level of filter coefficients $C_{0,k}$ are defined on the integers. Thus when a prefilter is used, as in equation 5, we assume that the signal is observed at the half integers. Define the vector of pairs of observed data (observed at the half integers),

$$\mathbf{f}_k = \begin{bmatrix} f(k) \\ f(k+\frac{1}{2}) \end{bmatrix},$$

and the Fourier transforms

$$C(\omega) = \sum_k C_{0,k} e^{-i\omega k},$$

$$Q(\omega) = \sum_k Q_k e^{i\omega k}.$$

Define the Fourier transform of the lth element of the vectors $\{\mathbf{f}_k\}$ as

$$f_l(\omega) = \sum_k f(k + \frac{1}{2}(l-1)) e^{-i\omega(k+\frac{1}{2}(l-1))} \qquad l = 1, 2.$$

We can express $C(\omega)$ in terms of $Q(\omega)$ and $f_l(\omega)$

$$\begin{aligned} C(\omega) &= \sum_k \sum_n Q_n \mathbf{f}_{n+k} e^{-i\omega k} \\ &= Q(\omega) \sum_s \mathbf{f}_s e^{-i\omega s}. \end{aligned}$$

Considering the lth element of the above sum,

$$\begin{aligned} \left[\sum_k \mathbf{f}_k e^{-i\omega k} \right]_l &= \sum_k f(k + \frac{1}{2}(l-1)) e^{-i\omega k} \\ &= f_l(\omega) e^{i\omega \frac{1}{2}(l-1)} \qquad l = 1, 2. \end{aligned}$$

Thus

$$C(\omega) = Q(\omega) \begin{bmatrix} f_1(\omega) \\ f_2(\omega) \cdot e^{i\omega/2} \end{bmatrix}.$$

This phase shift in the last term must be accounted for in any assessment in the frequency response of a prefilter Q, in other words, we analyse $\hat{q}(\omega) = Q(\omega)(1, e^{i\omega/2})^\top$ rather than just $Q(\omega)$.

As the prefilter is specific for each multiple wavelet basis, it is important to consider the frequency response of the lowpass filter in series with the prefilter and similarly the high pass filter in series with the prefilter as described in Xia *et al* [16]. We would like the combined filter $\hat{h}(\omega) = H(\omega)\hat{q}(\omega)$ to be low pass ($H(\omega) = \sum_k H_k e^{-i\omega k}$), that is $|\hat{h}(0)|$ to be non-zero in both elements and $|\hat{h}(\pi)| = (0,0)^\top$. Also, ideally $\hat{g}(\omega) = G(\omega)\hat{q}(\omega)$ should be high pass ie. $|\hat{g}(0)| = (0,0)^\top$ and $|\hat{g}(\pi)|$ be non-zero. Xia *et al* [16] showed that this ideal is unattainable for the Geronimo Multiwavelet but prefilters can be devised so that the deviations from this ideal are small.

Finally we shall consider $|\hat{q}(\omega)|$ for values of $\omega \in [0, \pi]$. If any element of $|\hat{q}(\omega)|$ increases with ω then the prefilter will increase the high frequency component relative to the low frequency component of a signal, effectively decreasing the signal to noise ratio. This is clearly undesirable when the wavelet decomposition is used for thresholding purposes. If $|\hat{q}(\omega)|$ decreases with ω then the SNR will be increased, enhancing the thresholding properties.

3.3.4 Frequency Response of Repeated Signals

The frequency response of a repeated signal is easier to obtain than for a prefilter. This is because $C_{0,k} = \sum_n \gamma_n f_{n+k}$ which corresponds to sampling at the integers. Thus, there is no phase shift in the expression of $C(\omega)$ which is simply,

$$C(\omega) = \hat{\gamma}(\omega) f(\omega), \tag{6}$$

where $\hat{\gamma}(\omega) = \sum_k \gamma_k e^{iwk}$ and $f(\omega) = \sum_k f(k) e^{iwk}$ (a scalar function). The high pass/low pass filters in series with the repeated signal filter are similarly obtained. The properties we desire for $\hat{\gamma}(\omega)$ are exactly the same as specified for $\hat{q}(\omega)$ above.

3.4 Specific Prefilters

3.4.1 Identity Prefilter

The identity prefilter is defined by

$$Q_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad Q_n = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \text{ for } n \neq 0.$$

Giving

$$\hat{q}(\omega) = \begin{bmatrix} 1 \\ e^{i\omega/2} \end{bmatrix}.$$

This prefilter has length 1 but has degree 0. As demonstrated in Example 3.2.1 the poor approximation of the identity prefilter means that it should not be used in practise.

3.4.2 Interpolation Prefilter

The interpolation prefilter is defined

$$Q_0 = \begin{bmatrix} \frac{3}{8\sqrt{6}} & \frac{5}{4\sqrt{6}} \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad Q_1 = \begin{bmatrix} \frac{3}{8\sqrt{6}} & 0 \\ \frac{1}{\sqrt{3}} & 0 \end{bmatrix}.$$

and the postfilter

$$P_0 = \begin{bmatrix} 0 & 0 \\ \frac{4\sqrt{6}}{5} & \frac{-3\sqrt{3}}{10} \end{bmatrix}, \quad \text{and} \quad P_1 = \begin{bmatrix} 0 & \sqrt{3} \\ 0 & \frac{-3\sqrt{3}}{10} \end{bmatrix}.$$

It has length 2 and degree 2. The frequency response is given by

$$\hat{q}(\omega) = \frac{1}{\sqrt{6}} \begin{bmatrix} \frac{1}{8}(3 + 10e^{i\omega/2} + 3e^{i\omega}) \\ \sqrt{2} e^{i\omega} \end{bmatrix}.$$

$|\hat{q}(\omega)|_1$ is slightly decreasing in ω and $|\hat{q}(\omega)|_2$ is constant in ω. The approximation of degree 2 means that the prefilter gives good results when thresholding but the length becomes more critical when compressing a signal.

3.4.3 Xia Prefilter

The Xia prefilter

$$Q_0 = \begin{bmatrix} 2+\frac{\sqrt{2}}{10} & 2-\frac{\sqrt{2}}{10} \\ \sqrt{2}+\frac{3}{20} & \sqrt{2}-\frac{3}{20} \end{bmatrix}$$

was first defined in Xia [16] (fixing the parameter values at $\epsilon_1 = 0$ and $\epsilon_2 = 0.1$). It was devised so that the composite filters $\mathcal{HQ}$ and $\mathcal{GQ}$ have near optimal low and high pass properties respectively. The frequency response of the prefilter is

$$\hat{q}(\omega) = \begin{bmatrix} 2+\frac{\sqrt{2}}{10} + (2-\frac{\sqrt{2}}{10})e^{i\omega/2} \\ \sqrt{2}-\frac{3}{20} + (\sqrt{2}+\frac{3}{20})e^{i\omega/2} \end{bmatrix}.$$

The moduli of the elements of $\hat{q}(\omega)$ are shown in Figure 3, which both decrease with ω.

3.4.4 A Minimal Prefilter

For a multiwavelet system with L wavelet functions and v vanishing moments there exists a prefilter which has degree $m = \min(L, v)$ and length 1, which we shall call a *minimal prefilter*. We shall construct a minimal prefilter for the Geronimo multiwavelet which has degree 2 and only one coefficient. Strang & Strela [14] show that the hat function is a linear combination of the scaling functions ϕ_1 and ϕ_2, specifically

$$H(x) = \frac{1}{\sqrt{2}}[\phi_1(x) + \phi_1(x-1)] + \phi_2(x).$$

This allows us to express constant and linear functions as a linear combination of the scaling functions, since the hat function is a linear spline function. So

$$1 = \sum_k \sqrt{2}\phi_1(x-k) + \phi_2(x-k) \quad \forall x \in R, \tag{7}$$

and

$$x - 1 = \sum_k \frac{2k-1}{\sqrt{2}}\phi_1(x-k) + k\phi_2(x-k). \tag{8}$$

A matrix which maps a constant signal to the coefficients in equation (7) and a linear signal (defined at the half integers) to the coefficients in (8) will have degree 2. Thus the minimal prefilter Q satisfies the equations

$$Q\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \sqrt{2} \\ 1 \end{bmatrix},$$

and

$$Q\begin{bmatrix} k \\ k+\frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{2k-1}{\sqrt{2}} \\ k \end{bmatrix}.$$

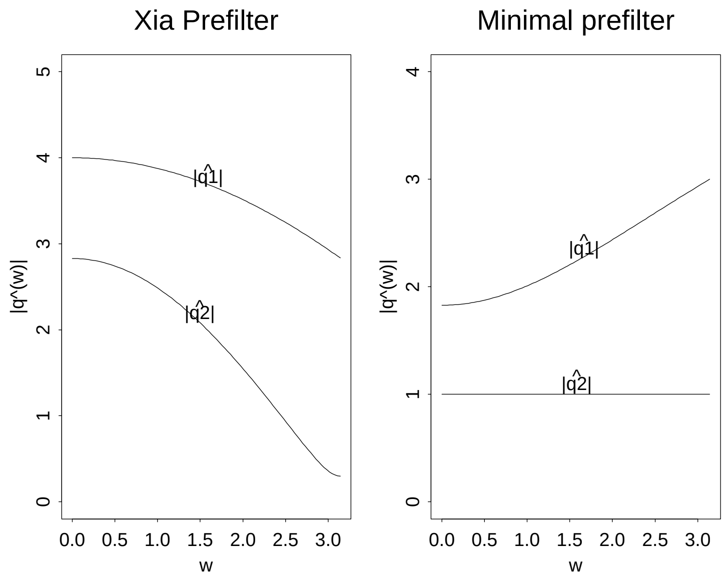

Figure 3: The frequency response $|\hat{q}(\omega)|$ of the Xia and minimal Prefilters.

The matrix Q that solves this is

$$Q = \begin{bmatrix} 2\sqrt{2} & -\sqrt{2} \\ 1 & 0 \end{bmatrix}.$$

The postfilter P is simply Q^{-1}, so

$$P = \begin{bmatrix} 0 & 1 \\ \frac{-1}{\sqrt{2}} & 2 \end{bmatrix}.$$

The frequency response of Q is

$$\hat{q}(\omega) = \begin{bmatrix} 2(\sqrt{2} - e^{i\omega/2}) \\ 1 \end{bmatrix}.$$

The moduli of the elements of $\hat{q}$ are also shown in Figure 3. Notice that $|\hat{q}(\omega)|_1$ increases by a relatively large amount as ω increases. This means that the minimal prefilter amplifies the high frequency component of the signal.

3.4.5 A Minimal Repeated Signal

Strela *et al* [15] proposed the repeated signal with length one $\gamma_0 = (\sqrt{2}, 1)^\top$ which is the first eigenvalue of $H(0)$. This ensures that a constant gives exact representation and so the repeated signal has degree one. It can be shown that there is no repeated signal of length 1 and degree 2. Referring to equation 6, as the minimal repeated signal has length one, its frequency response is constant. The composite filter $G(\omega)\gamma(\omega)$ has good highpass properties but $H(\omega)\gamma(\omega)$ has poor low pass properties.

3.4.6 Discussion

The thresholding and signal compression abilities for the above prefilters and repeated signal are demonstrated in Section 5. The prefilter that is chosen is often critical to the results obtained and this choice should be made depending on the application in hand and the type of signal to be processed. The identity prefilter is very poor and one would expect better results using other preprocessing methods. The minimal prefilter is appropriate for signal compression but it too has poor thresholding properties despite having degree 2, because the signal to noise ratio is amplified using the minimal prefilter. The Xia prefilter has near ideal frequency response properties but only has degree 1. In our simulations other preprocessing methods out performed this prefilter. The minimal repeated signal appears to be the best method for thresholding problems. Although it only has degree one, the oversampling appears to be more of an advantage than the two degrees of approximation in the interpolation and minimal prefilters. The disadvantage is that there are twice as many coefficients in the decomposition giving twice the computational effort, and giving an obvious disadvantage when compressing a signal. The interpolation prefilter give good results when thresholding and for signal compression, but it does not usually give the best results. It would seem that its length of two is not as critical to the decomposition as the frequency response and degree. We suggest that interpolation prefilter should be used if it is not clear which prefilter to use.

4 Thresholding Multiple Wavelets

There have been several different thresholding techniques described for single wavelet decompositions [4] [5] [11] [1]. A thresholding method adapts wavelet coefficients using a rule depending on the magnitude of the coefficients usually with the purpose of obtaining a smoother version of the observed signal. The simplest and most applicable technique is the universal threshold method of Donoho & Johnstone [4]. Strela *et al* [15] have applied the Donoho & Johnstone thresholding method to a multiple wavelet decomposition and concluded that thresholding results were an improvement on thresholding using Daubechies extremal phase wavelets with 2 vanishing moments. We have developed a thresholding technique specifically for multiwavelets, based on the multivariate properties of the decomposition. As demonstrated in the next section this method gives even better thresholding results than the univariate method. We shall first recap the theory behind the single wavelet thresholding method, then develop the multivariate method. Hard and soft thresholding rules and the universal threshold have an exact analogy between univariate and the multivariate methods. Finally in this section, we shall investigate the variance structure of a multiwavelet decomposition of a noisy signal and discuss the implications for thresholding theory.

4.1 Thresholding Single Wavelets

We shall assume throughout that a function f is observed at discrete time points of equal interval, which comprises a signal g and a noise component ϵ,

$$f_i = g_i + \epsilon_i \quad \text{for } i = 0, ..., N-1 \; (N = 2^J).$$

The noise terms are independent identically distributed random variables with $N(0, \sigma^2)$ distribution. Applying an orthogonal discrete single wavelet transform to a function of the above form gives us wavelet coefficients of the form

$$d_{j,k} = d^*_{j,k} + \epsilon_{j,k}$$

for $j = 0, ..., J-1$ and $k = 0, ..., 2^j - 1$. The $d^*_{j,k}$ are the wavelet coefficients of the signal and $\epsilon_{j,k}$ are the wavelet coefficients of the noise component which are also i.i.d. $N(0, \sigma^2)$. The rationale for wavelet thresholding is that for most purposes the signal will be compressed into a few large coefficients whereas the noise component will give rise to small coefficients everywhere.

There are two thresholding rules, *hard* and *soft*. Hard thresholding transforms all the coefficients with absolute value below the threshold to zero and leaves all the others untouched (*keep or kill*). Soft thresholding again kills the small coefficients but will reduce the magnitude of all the larger coefficients by an amount equal to the value of the threshold (*shrink or kill*). The universal thresholding rule by Donoho & Johnstone [4] is given by $\lambda_n = \sqrt{2\sigma^2 \log n}$, where n is the number of wavelet coefficients to be thresholded. As only coefficients at resolution levels $j = 1, \cdots j_0 < J$ are thresholded, n will be less than the number of data points. The universal threshold is the asymptotic limit of the maximum of n $N(0, \sigma^2)$ random variables, thus any coefficients which contain no signal component will be thresholded to zero provided n is large enough. Furthermore Donoho & Johnstone [4] show that the universal threshold is asymptotically optimal.

4.2 Multivariate Thresholding

The trivial extension of the universal threshold to multiple wavelet coefficients is to use the threshold above applied to each coefficient element. This is the method used by Strela *et al.* [15].

Although the DMWT is an orthogonal transformation, using any prefilter other than the identity prefilter will give correlated coefficients, in particular the elements within each coefficient will be highly correlated. As mentioned in the previous section the identity prefilter gives poor results when thresholding. In addition, if there is a signal component present at a particular time frequency location then one would expect other elements in that coefficient to contain some signal component, even if that component is small. Accordingly we develop a method which accounts for the noise and signal components within the whole vector.

Applying the discrete multiple wavelet transform with an appropriate prefilter to a noisy function, we get L-vector coefficients on the form,

$$D_{j,k} = D^*_{j,k} + E_{j,k}$$

where $E_{j,k}$ has a multivariate normal distribution $N(0, V_j)$. The matrix V_j is the covariance matrix for the error term which depends on the resolution level j. Using the standard transformation

$$\theta_{j,k} = D^\top_{j,k} V_j^{-1} D_{j,k} \tag{9}$$

we obtain a positive scalar value which in the absence of any signal component will have a χ^2_L distribution. It is these values that should be thresholded, and the coefficient vectors can then be adapted accordingly.

Suppose we use a threshold of λ. The analogous hard thresholding rule is

$$\hat{D}_{j,k} = D_{j,k} \cdot I(\theta_{j,k} \geq \lambda),$$

which *keeps* the vector if $\theta_{j,k}$ is greater than the threshold and *kills* it otherwise. The Soft thresholding rule can be expressed as

$$\hat{D}_{j,k} = D_{j,k} \cdot \frac{\max\{\theta_{j,k} - \lambda, 0\}}{\theta_{j,k}}.$$

If $\theta_{j,k} > \lambda$ then each element of $D_{j,k}$ is shrunk proportionately, otherwise it is killed (see Figure 4). The shape of the ellipse in Figure 4, which denotes the region in which the vector is killed, depends on the variance matrix V_j.

4.3 Obtaining the Universal Threshold

The universal estimator for single wavelets is the sequence λ_n such that it is asymptotically the minimax estimator for squared error loss. If $X_1, \cdots, X_n$ are i.i.d. $N(0, \sigma^2)$ random variables and $M_n = \max_{i=0,\cdots,n} |X_i|$, we can define the set Λ such that

$$\{\lambda_n\} \in \Lambda \Leftrightarrow P(M_n < \lambda_n) \to 1 \text{ as } n \to \infty$$

Here $\lambda_n = \sqrt{2\sigma^2 \log n}$ is the asymptotic infimum of Λ, and is the asymptotic optimal threshold [4]. This means that for an orthogonal wavelet decomposition of a function with the above noise structure and n is sufficiently large then any coefficient whose absolute value is greater than this threshold is almost certain to contain some signal component.

Turning to multiple wavelet bases with L wavelets, we have already noted that a coefficient vector which has no signal component gives $\theta_{j,k} \sim \chi^2_L$. To find the universal estimator we need to find the asymptotic maximum of n i.i.d. χ^2_L random variables.

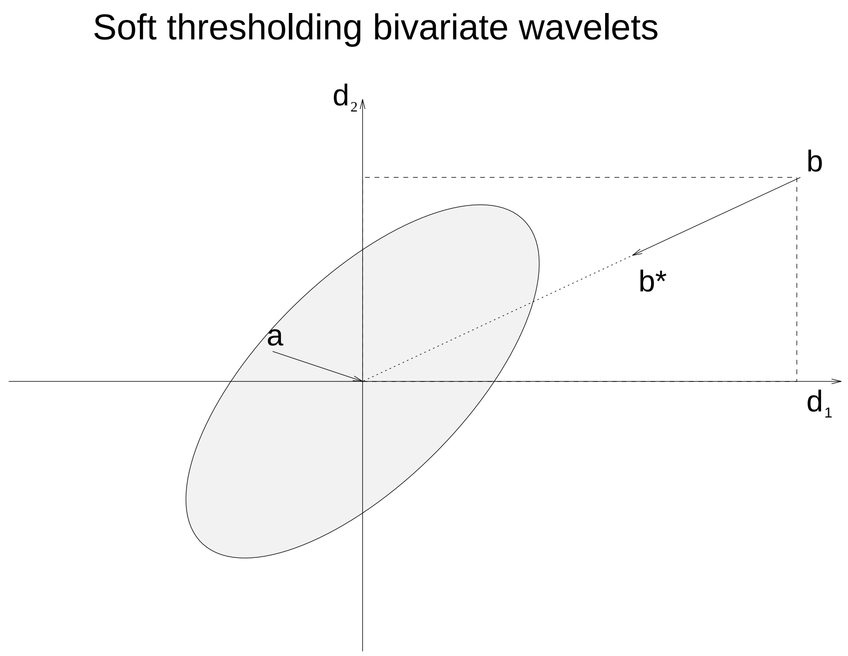

Figure 4: Any point within the ellipse is thresholded to zero ($a \to 0$), any point outside the ellipse is reduced towards the origin by the distance from the edge of the ellipse to the origin ($b \to b'$).

Theorem If M_n is the maximum of n i.i.d χ^2_p random variables ($p > 1$) and let λ_n be the infimum of all sequences a_n that satisfy

$$P(M_n \leq a_n) \to 1 \text{ as } n \to \infty,$$

then

$$\lambda_n = 2 \log n + (p-2) \log \log n.$$

Outline Proof For the result, to hold it is sufficient to show that

$$nP(X > \lambda_n) \to \alpha \text{ as } n \to \infty, \tag{10}$$

where $X \sim \chi^2_p$ and α is a positive real. Using the distribution function for a χ^2_p distribution and repeated integration by parts, it can be shown that the limit in Equation 10 is attained when $\lambda_n = 2 \log n + (p-2) \log \log n$.

When $L = 2$, the universal threshold for multiple wavelets simplifies to $\lambda_n = 2 \log n$. Note that unlike in single wavelet thresholding the variance term does not appear in the universal threshold formula. This is because the theta coefficients are sphered by the inverse variance term in Equation 9. When $L > 2$ then it may still be preferable to use $\lambda_n = 2 \log n$ without the $\log \log n$ term, since the universal threshold is generally an overestimate of the optimal threshold [4].

4.4 Covariance Structure

With the single wavelet DWT, usually no prefilter is used and the transform itself is orthogonal. Hence if the noise component is gaussian white noise with variance σ^2 then the noise component of each coefficient will also be i.i.d. with variance σ^2. The covariance structure of a multiple wavelet decomposition is more complex.

If the identity prefilter is used, then each coefficient vector will be i.i.d. with variance matrix $\sigma^2 I$. However in practise the identity prefilter is not used and so correlation is introduced at every level. For example, if the prefilter has length one then the variance of a starting level coefficient vector is

$$\mathrm{Var}(C_{0,k}) = \sigma^2 Q_0 Q_0^\top \quad \text{for all } k = 0, \dots, 2^J$$

and all the vectors $C_{0,k}$ will be mutually independent. Applying the high and low pass filters will produce vector coefficients which are correlated with their neighbours. The between vector correlations increase with the length of the prefilter. However, for any prefilter, the full covariance matrix for a given level (which contains the covariances between each coefficient element) will be block symmetric and circulant (assuming periodic boundary conditions). By a simple extension of the results in Johnstone and Silverman [9], a vectorwise thresholding routine will be asymptotically optimal.

Using a repeated signal of length one, each $C_{0,k}$ will be i.i.d. with

$$\mathrm{Var}(C_{0,k}) = \sigma^2 \gamma\gamma^\top$$

This is a singular $L \times L$ matrix. Applying the high and low pass filters however spreads the dependence between vector coefficients and the within vector covariances. So $\mathrm{Var}(D_{1,k})$ will be non-singular and invertible, despite the over sampling.

To obtain the $\theta_{j,k}$ in the thresholding routine it is clear that we need to know $V_j = \mathrm{Var}(D_{j,k})$. In theory one can obtain directly the the covariance structure of the wavelet coefficients at each level, in terms of σ^2. We obtained better results when V_j is estimated directly from the observed coefficients using a robust covariance estimation method [7] than by estimating σ^2 and then recursively computing V_j.

5 Simulation

A series of simulations were carried out to assess the efficacy of bivariate thresholding using the Geronimo multiple wavelets, and to compare the results of thresholding using Daubechies wavelets. We investigated the effect of specific prefilters, hard and soft thresholding the magnitude of the

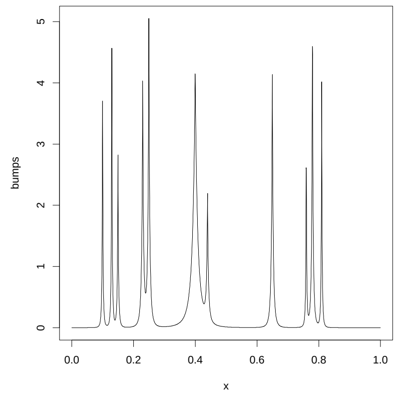

Figure 5: Bumps signal from Donoho & Johnstone

noise component and the original signal. In addition the bivariate method was compared to the univariate method for multiwavelets and the universal threshold to the *ideal* threshold. Finally we consider the signal compression of noise free signals using different prefilters.

Unless specified explicitly, assume that the following details applied. Each simulation used the *bumps* signal evaluated at 2048 points, as defined in Donoho & Johnstone [4] and shown in Figure 5. This function and the others used later in this Section are defined on $[0,1]$, we imply a trivial scaling so that the function is sampled at the integers (or half integers). Gaussian white noise was added to the signal so that the signal-to-noise ratio was 4 (SNR$=\sqrt{\text{var}(f)/\sigma^2}$). The DMWT was applied with the specified prefilter and the coefficients thresholded using the bivariate universal threshold and the hard thresholding method. The inverse wavelet transform and postfilter were applied to obtain the smoothed estimate, the estimate was then compared with the original signal by computing the mean square error MSE $=\frac{1}{n}\sum_i(f_i-\hat{f}_i)^2$. For each table entry (with the exception of table 4) only one simulation was done but the same noise component was used within each table. In table 4 one-hundred different replications were made for each entry.

The total number of levels in the decomposition is 10 for the Geronimo wavelet with the identity, interpolation, minimal,and Xia prefilters, but for the Geronimo wavelet with the (minimal) repeated signal and the Daubechies wavelets there are 11 levels. For each simulation, levels $1,\ldots,j_0$ were thresholded, where j_0 was chosen a) so that all but the five levels corresponding the lowest frequencies are thresholded, labeled as *same frequencies* and b) to give the minimum mean square error, *optimal frequencies*. The number of levels in each simulation that were thresholded are given in each table.

Table 1 compares the thresholding results using different preprocessing methods. The noisy function and the five smoothed estimates thresholded over the same frequencies are shown in

Table 1: Thresholding using different prefilters

	Same frequencies		Optimal frequencies	
Preprocessing	levels	MSE	levels	MSE
Identity	5	0.0123	3	0.0121
Xia	5	0.1088	4	0.1087
Minimal	5	0.0562	4	0.0530
Interpolation	5	0.0078	3	0.0077
Repeated Signal	6	0.0080	4	0.0066

Mean square error of noisy signal=0.0276.

Table 2: Univariate and Bivariate thresholding

		Same frequencies		Optimal frequencies	
Method	Preprocessing	levels	MSE	levels	MSE
Univariate	Interpolation	5	0.0543	2	0.0141
Bivariate	Interpolation	5	0.0061	5	0.0061
Univariate	Repeated signal	6	0.0292	3	0.0137
Bivariate	Repeated signal	6	0.0064	6	0.0064

Mean square error of noisy signal=0.0276.

Figure 6. The best results occurred when using the interpolation prefilter or the repeated signal. The minimal prefilter gives an estimate that is generally close to the true signal but has large unwanted spikes appearing in the estimate. The identity prefilter achieves a better MSE than the minimal prefilter but leaves a very high frequency component which can just be seen in Figure 6 as a thick line. The Xia prefilter has the problems of both the minimal and identity prefilter and gives poor thresholding results. For the remaining simulations only the interpolation prefilter and the repeated signal will be considered.

Table 2 compares univariate and bivariate thresholding of the Geronimo multiwavelet. As can be seen, a large improvement can be obtained using the bivariate method. Figure 7 shows estimates obtained using the two methods.

Table 3 compares hard and soft thresholding. Hard thresholding gives a lower MSE and is less critical to the number of levels that are thresholded.

Table 4 compares thresholding using Daubechies wavelets with the Geronimo multiwavelet under

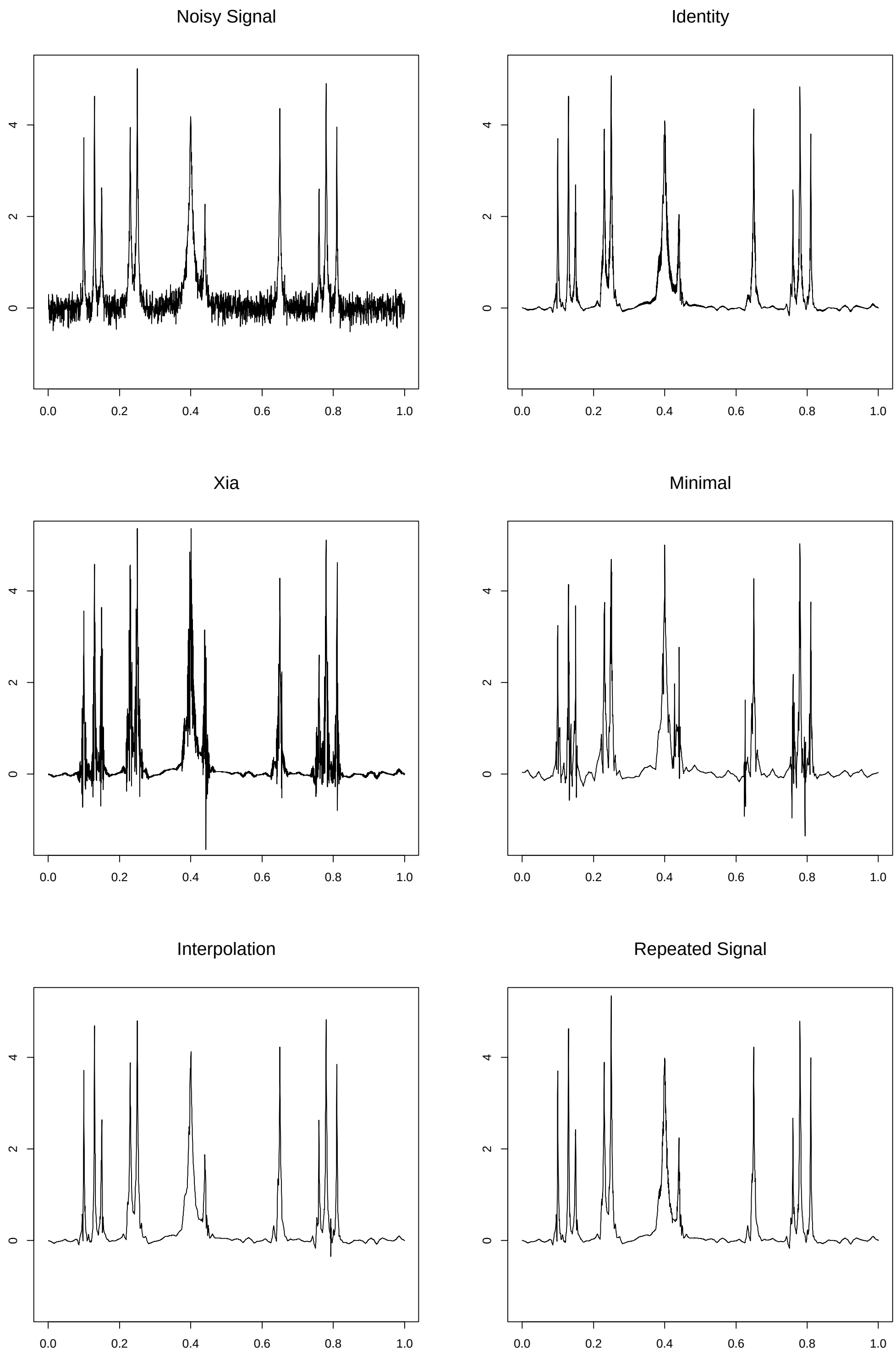

Figure 6: Bumps with noise added, and thresholded using 5 different preprocessing methods.

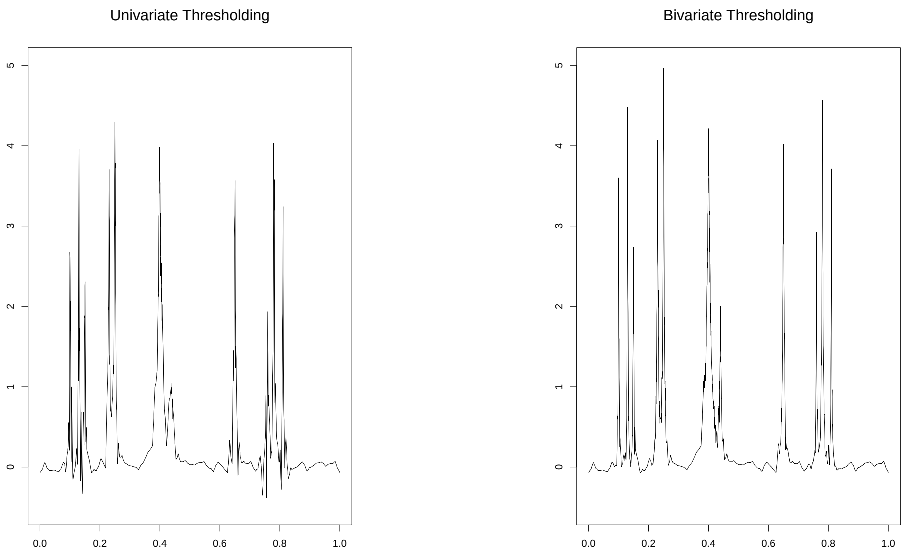

Figure 7: Thresholding of Geronimo multiwavelets using univariate and bivariate thresholding methods, (using repeated signal).

Table 3: Hard and Soft Thresholding

		Same frequencies		Optimal frequencies	
Method	Preprocessing	levels	MSE	levels	MSE
Soft	Interpolation	5	0.0131	3	0.0074
Hard	Interpolation	5	0.0071	3	0.0064
Soft	Repeated Signal	6	0.0134	4	0.0075
Hard	Repeated Signal	6	0.0061	6	0.0061

Mean square error of noisy signal=0.0290.

Table 4: Effect of signal to noise ratio

SNR	Wavelet	Preprocessing	Same frequencies			Optimal frequencies		
			levels	MSE	s.e. $\times 10^{-5}$	levels	MSE	s.e. $\times 10^{-5}$
10	Geronimo	Interpolation	5	0.00121	1.4	5	0.00121	1.4
	Geronimo	Repeated Signal	6	0.00150	1.5	5	0.00145	1.2
	Daubechies 2		6	0.0344	13.	2	0.00250	1.6
	Daubechies 6		6	0.0390	15.	2	0.00213	1.2
7	Geronimo	Interpolation	5	0.00259	2.3	5	0.00259	2.3
	Geronimo	Repeated Signal	6	0.00268	2.4	5	0.00255	1.7
	Daubechies 2		6	0.0347	14.	2	0.00411	2.4
	Daubechies 6		6	0.0400	19.	2	0.00351	2.1
4	Geronimo	Interpolation	5	0.00640	9.1	5	0.00640	9.1
	Geronimo	Repeated Signal	6	0.00652	6.4	5	0.00657	7.6
	Daubechies 2		6	0.0372	19.	2	0.0148	6.2
	Daubechies 6		6	0.0372	19.	2	0.0094	5.8
2	Geronimo	Interpolation	5	0.0257	42.	4	0.0232	2.5
	Geronimo	Repeated Signal	6	0.0271	43.	5	0.0245	22.
	Daubechies 2		6	0.0477	31.	3	0.0261	20.
	Daubechies 6		6	0.0606	43.	3	0.0278	18.

Bivariate universal thresholding used for Geronimo wavelets and Universal thresholding for the Daubechies wavelets, the number of replications per entry was 100.

four different noise levels. The two Daubechies wavelets chosen were *Extremal Phase* with two and six vanishing moments. In general thresholding using the Geronimo multiwavelet gave better results than the Daubechies wavelets; the amount depending on the SNR. The Geronimo multiwavelet performs best for high SNR levels which is a result of it having giving a better signal compression ratio than Daubechies wavelets [15]. The Daubechies estimates gave a lower MSE when only two or three levels were thresholded, but this gave the estimate a much rougher appearance. Figure 8 shows estimates after thresholding with Daubechies and Geronimo wavelets.

Table 5 compares the universal threshold with the *ideal* threshold, the ideal threshold being the value which gives the smallest MSE with the number of levels to be thresholded fixed. The MSE for the universal threshold with the Geronimo multiwavelets is very similar to the MSE for the ideal threshold despite the threshold values being reasonably different. The universal threshold for the Daubechies wavelets however gives a MSE approximately five times greater than that obtained with the ideal threshold.

Table 6 shows the results of thresholding the signals “blocks” and “heavisine” (again as defined in Donoho & Johnstone [4]). The results for the noisy blocks are very similar to those found above when the signal was “bumps”. However the results obtained for the noisy heavisine signal were very similar regardless of which wavelet was used. The Daubechies wavelets slightly outperforming the Geronimo multiwavelet. Donoho & Johnstone noted that heavisine gave better thresholding results (for Daubechies wavelets) than when thresholding blocks and bumps and that this was due to the efficient wavelet decomposition for heavisine. Figure 9 shows the original signals and

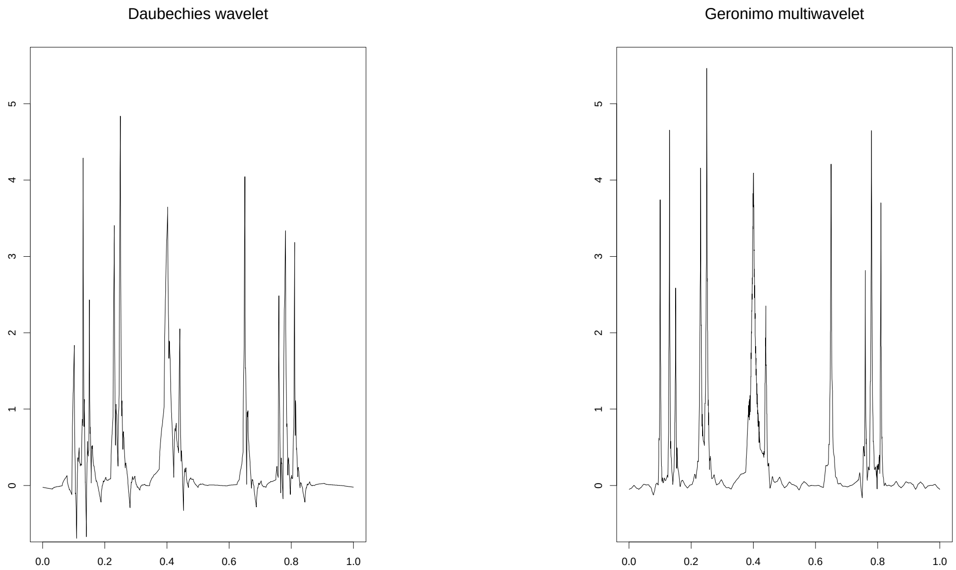

Figure 8: Thresholding using Daubechies 2 wavelet and Geronimo multiwavelets (using repeated signal).

Table 5: Universal and Ideal thresholding

		Universal		Ideal	
Wavelet	Preprocessing	threshold	MSE	threshold	MSE
Geronimo	Interpolation	13.8	0.00689	8.6	0.00596
Geronimo	Repeated Signal	15.2	0.00651	9.6	0.00614
Daubechies 2		1.84	0.0378	0.52	0.00659
Daubechies 6		1.93	0.0455	0.46	0.00766

Mean square error of noisy signal=0.0291

Same frequencies were thresholded for each entry.

Table 6: Thresholding different signals

Method	Preprocessing	Same frequencies		Optimal frequencies	
		levels	MSE	levels	MSE
Blocks, SNR=4, mse=0.2316					
Geronimo	Interpolation	5	0.05808	5	0.05808
Geronimo	Repeated Signal	6	0.05883	5	0.04885
Daubechies 2		6	0.1159	3	0.06427
Daubechies 6		6	0.1842	2	0.07743
Heavisine, SNR=4, mse=0.5424					
Geronimo	Interpolation	5	0.02603	6	0.02330
Geronimo	Repeated Signal	6	0.02600	7	0.02342
Daubechies 2		6	0.02039	6	0.02039
Daubechies 6		6	0.02268	6	0.02268

thresholded estimates.

Table 7 assesses the signal compression of the three noiseless signal functions using the five different preprocessing methods. The measure of compression used was to count the number of non-zero coefficient elements in the wavelet decomposition, arbitrarily regarding any coefficient element less than 10^{-5} as zero. The repeated signal gives many more nonzero coefficients than any of the prefilters since the decomposition has twice as many coefficients. The best overall performance is attained by the Minimal and interpolation prefilters. Because the identity prefilter has degree 0 it gives very poor signal compression.

6 Conclusions

The theoretical development of wavelet bases that have more than one mother wavelet function have been developed in recent years. It has been proposed that such multiwavelet bases should be better at wavelet applications than other wavelet bases because these multiple wavelets can be orthogonal, symmetric, and have shorter support for a comparable number of vanishing moments. There are two major differences between single and multiple wavelets in terms of applying the discrete transform and thresholding methods. The data have to be preprocessed in order to obtain an efficient decomposition, and better thresholding results occur when the thresholding method is based on the coefficient vectors rather than their elements.

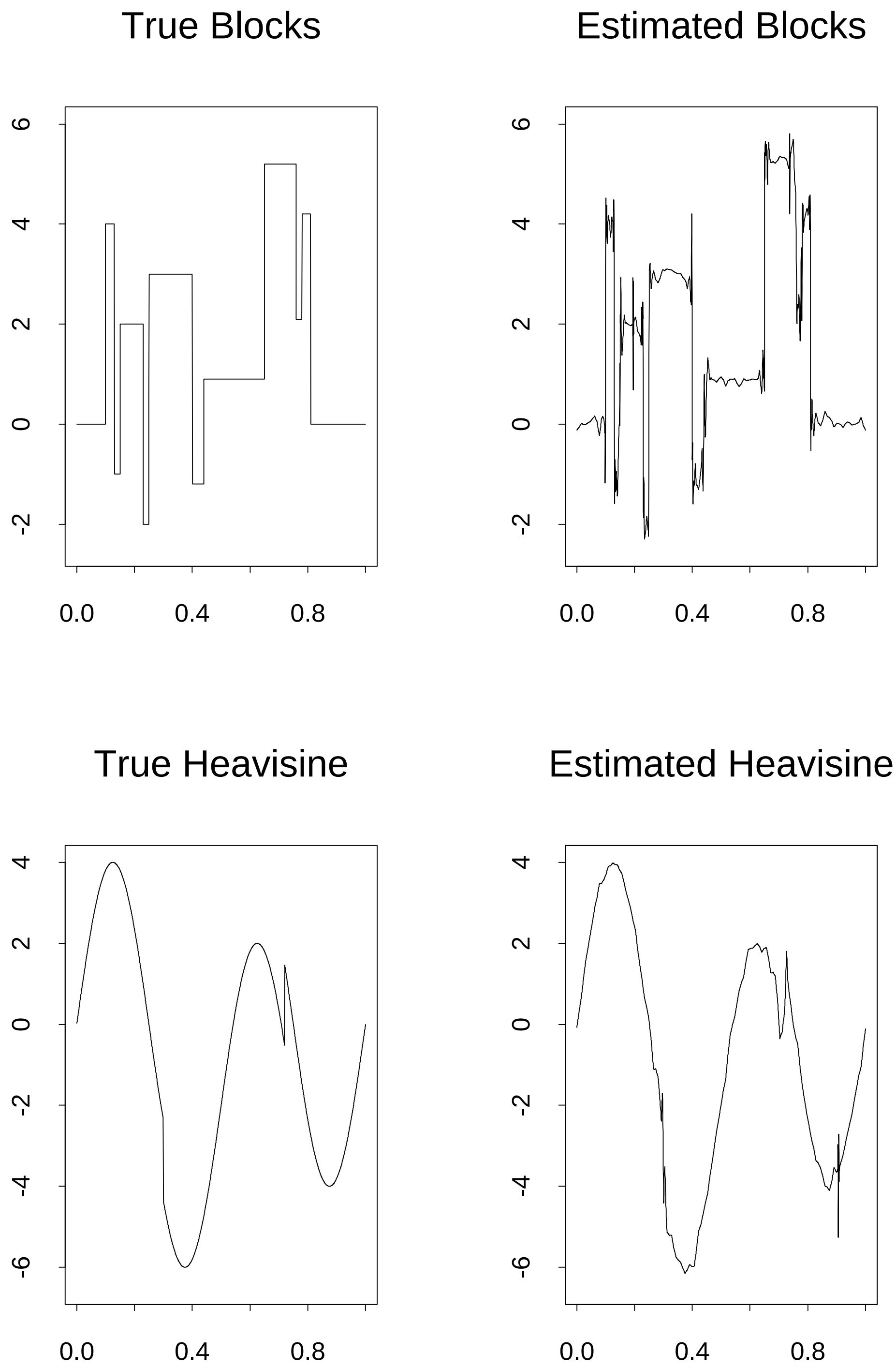

Figure 9: Thresholding blocks and heavisine signals, true signals and thresholded estimates using Geronimo multiwavelets and repeated signal.

Table 7: Signal compression using different preprocessing methods

No. of elements with magnitude $> 10^{-5}$			
Signal	Blocks	Bumps	Heavisine
Interpolation	258	1145	1180
Repeated Signal	1610	3501	4086
Minimal	246	1200	1238
Xia	246	1596	2034
Identity	856	1736	2003

Prefilters can be categorised according to their length and degree. We develop a prefilter which has degree 2 and only one filter coefficient, which has good compression properties, but is inefficient when used for thresholding. We propose that the choice of prefilter should depend on the application in hand and recommend the repeated signal or interpolation prefilter for thresholding and the minimal prefilter or the interpolation prefilter for data compression.

The time/frequency structure of a multiple wavelet decomposition and the dependency within coefficients introduced by a non-orthogonal prefilter are reasons to consider multivariate thresholding. This method computes a value for each coefficient vector which corresponds to its magnitude, such that the noise component of this value has a chi-squared distribution. The thresholding rules are then applied to this value and the coefficient vector is adapted accordingly. A universal threshold based on the chi-squared distribution is developed which is $\lambda_n = 2 \log n$ when there are two mother wavelets.

Simulation shows that this thresholding method gives good results, which are insensitive to the value of threshold. In general the mean square error of estimates using the bivariate thresholding method are much better than using Daubechies wavelets especially when the signal to noise ratio is high.

References

[1] F. Abramovich and Y. Benjamini. Adaptive methods of wavelet coefficients as multiple hypothesis testing proceedure. Submitted for publication, 1994.

[2] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.*, 41:909–996, 1988.

[3] I. Daubechies. *Ten Lectures on Wavelets.* Society for Industrial and Applied Math., Philadelphia, 1992.

[4] D.L. Donoho and I.M. Johnstone. Ideal spatial adaption via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[5] D.L. Donoho, I.M. Johnstone, G Kerkyacharian, and D Picard. Wavelet shrinkage: Asymptopia? *Journal of the Royal Statistical Society, Series B*, 57:301–369, 1995.

[6] J. S. Geronimo D. P. Hardin and P.R. Massopust. Fractal functions and wavelet expansions based on several scaling functions. *Journal of Approximation Theory*, 78:373–401, 1994.

[7] P. J. Huber. *Robust Statistics.* Wiley & Sons, New York, USA, 1981.

[8] B. Jawerth and W. Sweldens. An overview of wavelet based multiresolution analyses. Technical report, University of South Carolina, Feb. 8, 1993.

[9] I. M. Johnstone and Silverman B. W. Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society, Series B*, 1997. to appear.

[10] S. G. Mallat. A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans. on Patt. Anal. and Mach. Intell.*, 11:674–693, 1989.

[11] G. P. Nason. Wavelet shrinkage using cross-validation. *Journal of the Royal Statistical Society, Series B*, 58:463–479, 1996.

[12] G.P. Nason and B.W. Silverman. The discrete wavelet transform in S. *Journal of Computational and Graphical Statistics*, 3:163–191, 1994.

[13] G. Strang and V. Strela. Orthogonal multiwavelets with vanishing moments. *Optical Engineering*, 33:2104–2107, 1994.

[14] G. Strang and V. Strela. Short wavelets and matrix dilation equations. *IEEE Trans. on Signal Proccessing*, 1994.

[15] V. Strela, P.N. Heller, G. Strang, P. Topiwala, and C. Heil. The application of multiwavelet filterbankes to image processing. Technical report, Massachusets Institute of Technology, USA, 1995. Submitted to IEEE trans on Image Processing.

[16] X.-G. Xia, J. Geronimo, D. Hardin, and B. Suter. On discrete multiwavelet transforms. *IEEE Trans. on Signal Processing*, 44:25–35, 1996.