NPL Report CMSC 02/00

A Practical Guide to Wavelets for Metrology

By
G J Lord, E Pardo-Igúzquiza
and I M Smith

June 2000

A Practical Guide to Wavelets for Metrology

G J Lord, E Pardo-Igúzquiza
and I M Smith

June 2000

ABSTRACT

This report aims to give a brief introduction to wavelet analysis and to identify its relevance to areas of metrology. Wavelets provide an economical basis for a wide range of applications as well as offering simultaneously a time and frequency analysis of a signal. Wavelets should not be viewed as a stand-alone tool but as complementing and supplementing well established techniques.

ISSN 1471–0005

Authorised by Dr Dave Rayner,
Head of the Centre for Mathematics and Scientific Computing

National Physical Laboratory,
Queens Road, Teddington, Middlesex, United Kingdom. TW11 OLW

Contents

1 Introduction

Wavelets offer an alternative to traditional Fourier bases for representing functions. Wavelets may be thought of as oscillations that are localized in time (or space) and localized in frequency. A representation of a function or signal based on these localized oscillations is attractive because the coefficients associated with each wavelet convey details in time and frequency. A small coefficient means there is little variation in the function or signal in the vicinity of that particular small oscillation, whereas a large coefficient indicates that there is appreciable change in the signal or function. Analysis of the coefficients allows features in the function or signal to be identified. Furthermore, because coefficients are identified with both a localized frequency and time, small coefficients can be replaced by zero without greatly affecting the reconstruction of the signal. This strategy leads to the efficient representation of functions and algorithms for de-noising of general signals. In summary, the combination of the two localization properties make wavelets attractive for the following applications:

- sparse representation of functions,
- feature detection in signals,
- de-noising of signals.

As an illustration of the de-noising and sparse representation properties, we have plotted in Figure 1(a) a signal arising from the calibration of a hydrophone (see [18]), and in Figure 1(b) the same signal reconstructed from a wavelet decomposition where over 90% of the coefficients have been replaced by zero. The horizontal axis shows the indices of the data points in the signal; the vertical axis is the voltage across the hydrophone (measured in volts).

In this report we discuss some of the fundamental mathematical properties of wavelets and introduce wavelet transforms analogous to the Fourier and fast Fourier transforms. The aim is to introduce the reader to the concepts and terminology associated with wavelets. We cannot hope to give a fully comprehensive review of the mathematics involved and instead refer the reader to references [9, 10, 21]. There is also a great number of on-line resource pages for wavelets, containing information on the whole range of wavelet analyses, introductions to recent research papers and guides to available software. Two good starting points are [37] and [24].

After introducing wavelets and their mathematical properties, we consider the statistical properties associated with wavelets, such as the choice of threshold for signal de-noising and non-parametric regression as well as their use on correlated data sets. We briefly discuss their application to time series analysis before turning to potential applications in metrology. Finally, we comment briefly on our experience with a number of wavelet software packages.

2 Mathematical Properties of Wavelets

2.1 Background

From a signal processing point of view, a wavelet decomposition may be thought of as a combination of high and low pass filters that have particular properties. Let $\mathcal{G}$ be a high pass filter and $\mathcal{H}$ a low pass filter and suppose we are given a function F to examine. Then a wavelet decomposition successively decomposes the function F using the filters into spaces of "approximations" $\{V_j\}$ and "details" $\{W_j\}$. This process is illustrated in Figure 2. The high pass filter $\mathcal{G}$ extracts "detail", whereas the low pass filter $\mathcal{H}$ smoothes the signal. The coefficients in the filters determine the properties of the decomposition.

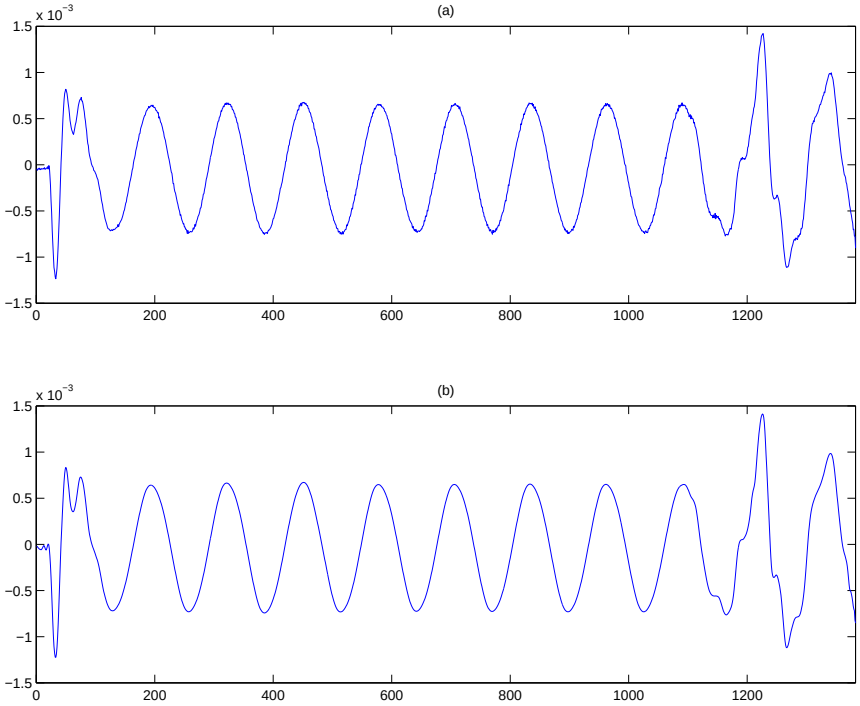

Figure 1: (a) A sample signal; (b) Reconstruction using fewer than 10% of the coefficients from the Coiflet 4 wavelet decomposition.

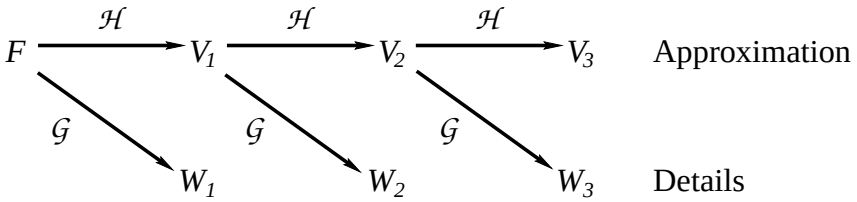

Figure 2: Schematic view of a 3-level decomposition of a function F by wavelets into an approximation space V_3 and detail space W_3 by a combination of low ($\mathcal{H}$) and high ($\mathcal{G}$) pass filters.

An important consequence of this decomposition is that in order to reconstruct the function F, all that are required are the coefficients from the detail spaces $\{W_j | j = 1, \ldots, J\}$ and the final approximation V_J. We can think of decomposition as taking us from a fine to a coarse approximation, while reconstruction takes us in the opposite direction. The reconstruction process is schematically shown in Figure 3, in which we see that to reconstruct F, all that is required are the detail spaces W_1, W_2, W_3 and the final approximation V_3.

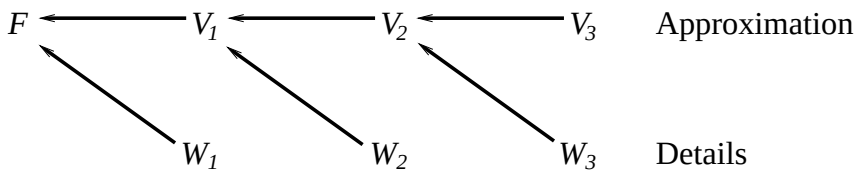

Figure 3: Schematic reconstruction of the function F from the 3-level decomposition. The approximation space V_3 and the detail spaces W_1 ,W_2, W_3 are all that are required.

Another way to introduce a wavelet decomposition is in terms of a wavelet basis and coefficients which convey information both about frequency and spatial/temporal position. To obtain such a basis we use dilations and translations of a **wavelet** function ψ: these wavelet functions form a basis for the detail spaces $\{W_j\}$. There are associated **scaling functions** that form a basis for the approximation spaces $\{V_j\}$. Both wavelet and scaling functions will be discussed in subsequent sections.

Notation: Throughout this report we will use j to denote the scale of the details. Note, however, that there is no standard convention in the literature or computational packages as to whether $j = 0$ represents the finest or coarsest scale. The reader should be aware of this since figures generated using different software will label the scales differently although the identification of each scale is usually obvious. Shifts in position are denoted throughout by k.

There are two important features about the wavelet coefficients that make wavelets so attractive:

1. **Data Compression:** A key to a good basis is being able to represent a function with as few basis functions as possible. For an arbitrary function, a large number of the wavelet coefficients are either zero or very small. Since these coefficients contribute little to the reconstruction of the function, they can be replaced by zero and so good data compression is achieved. This suppression of near-zero coefficients can allow for effective de-noising of a function with noise.

2. **Feature Extraction:** We have seen above that a wavelet decomposition is achieved by a combination of high and low pass filters. This decomposition will essentially identify high, medium and low scales in the function. In addition the non-zero wavelet coefficients can be related to the size of the derivative of the function in a particular region. This relationship is useful for feature extraction, the large coefficients identifying where there are steep gradients or discontinuities in the function.

To help illustrate the use of wavelets, we will consider in the following sections the signal shown in Figure 4. This waveform arises from measurements made in a reverberant tank of the response of an underwater electroacoustic transducer driven by a discrete-frequency tone-burst: see [18] for further details. Two transducers are suspended in a tank of finite size filled with water: one acts as a projector or transmitter and is driven by the tone-burst, the other as a receiver used to record the response of the projector. We can observe in the waveform a number of important features of the measurement. The "turn-on" of the projector is followed immediately by an oscillation representing its resonant behaviour. After three or four cycles of this oscillation, the resonant behaviour is sufficiently damped to observe the steady-state response which takes the form of an undamped oscillation at the frequency of the drive signal. Finally, a change in the waveform

results from the "turn-off" of the projector and the arrival at the receiver of reflections from the tank walls. Throughout, the measured waveform is contaminated by high-frequency noise. In what follows, instead of considering the signal in Figure 4 directly, we modify the waveform to include a discontinuity by *padding* the signal with zeros as in Figure 5. Note that in Figure 4, we plot the voltage (V) against the indices of the data points in the signal, while in Figure 5 the padded waveform is plotted against position in the signal.

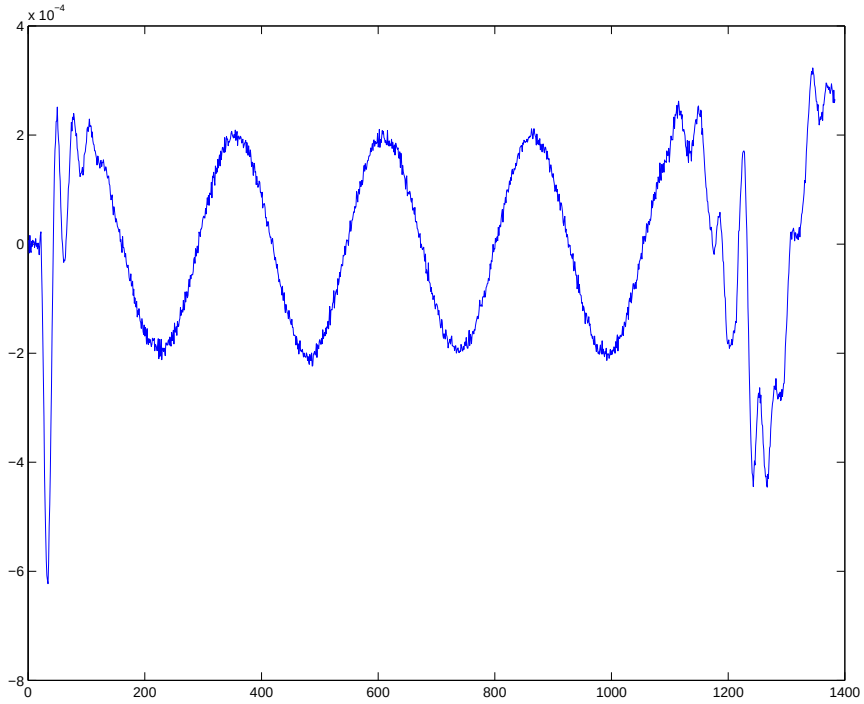

Figure 4: An example signal for wavelet analysis, originating from the calibration of a hydrophone.

In the remainder of this section, we give a general overview of the theory for wavelets. Sections 2.2 and 2.3 describe, respectively, the continuous and discrete wavelet transforms. In section 2.4, we introduce wavelet bases and the idea of multiresolution analyses, before going on to discuss redundant wavelet representations in section 2.6. The properties of wavelets are described and various types of wavelet family introduced in section 2.7, before a discussion on the extensions to higher dimensions is presented in section 2.8.

2.2 Continuous Wavelet Transform (CWT)

Suppose we are given a time signal $f(t)$. A Fourier analysis of this signal extracts information about the frequencies contained in f. The standard Fourier transform

$$(\mathcal{F}f)(w) = \hat{f}(w) = \frac{1}{\sqrt{2\pi}} \int e^{-iwt} f(t)\, dt$$

will give frequency information, but information concerning time-localization is more difficult to obtain from $(\mathcal{F}f)$.

Windowing the signal f is a step towards obtaining such information. Here the signal is first restricted to an interval (with smoothed edges) by multiplying it by a fixed window function, before carrying out a Fourier analysis of the product. Repeating the process with shifted versions of the window function allows localised frequency information throughout the signal to be obtained.

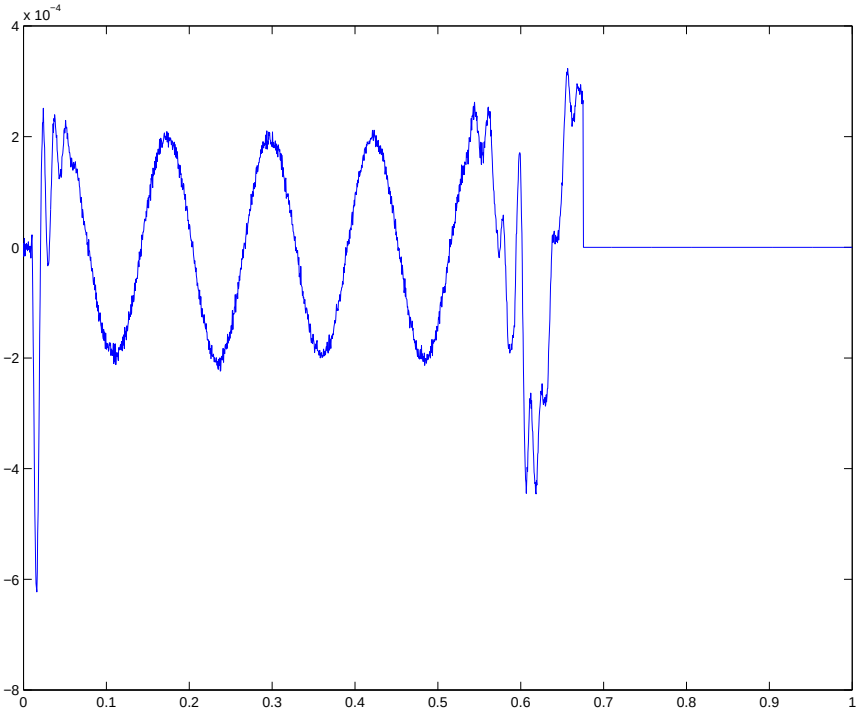

Figure 5: Same example as Figure 4 with zeros added to pad the signal to have a length 2048 and create a discontinuity.

Since the window-width is the same for all frequencies, the amount of localisation remains constant for different frequencies.

Mathematically, the windowed Fourier transform can be written as:

$$\left(T^{\text{win}} f\right)(w,t) = \int f(s)\, g(s-t)\, e^{-iws}\, ds,$$

where g is the windowing function having compact support (see section 2.7).

The **continuous wavelet transform** gives a time-frequency decomposition by taking translations and dilations of a (real or complex) wavelet ψ:

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

where $a(\neq 0), b \in \mathbb{R}$ and ψ may be real- or complex-valued.

The continuous wavelet transform (CWT) of f is then given by

$$\begin{aligned} C(a,b) &= \langle f, \psi_{a,b} \rangle \\ &= \frac{1}{\sqrt{|a|}} \int f(t)\, \overline{\psi\left(\frac{t-b}{a}\right)}\, dt, \end{aligned} \tag{2}$$

where $\overline{\psi}$ denotes the complex conjugate of ψ.

The $\{C(a,b)\}$ are called the **wavelet coefficients**, with a and b referred to as the scale and translation parameters respectively.

It is assumed that the wavelet ψ has zero mean so that

$$\int \psi\, dt = 0.$$

Although this is required for certain mathematical properties to hold, it can be thought of as saying that the wavelet oscillates.

A classical choice for ψ is the "Mexican hat" function:

$$\psi(t) = (1 - t^2)\exp(-t^2/2). \tag{3}$$

This function is localized both in time and frequency and satisfies the zero mean condition. In Figure 6(a) we have plotted the Mexican hat function with $a = 1, b = 0$ and we see it is centred about the origin; in Figure 6(b) we have plotted two scaled and translated versions of the function with $a = 2$, and $b = \pm 5$.

In what follows, we will use the terms "high frequency" and "low frequency" to refer respectively to low and high values of the scale parameter. This "inverse relationship" between scale and frequency is quite intuitive: we can think of a low scale as corresponding to a compressed wavelet, where the details are changing rapidly, similar to the behaviour of a high frequency signal. We also talk about the **resolution** which is simply defined as the reciprocal of the scale parameter.

The Daubechies wavelets are a popular choice as they form an orthogonal basis. These and other types of wavelet are discussed in section 2.7 (see also Figures 13 and 14).

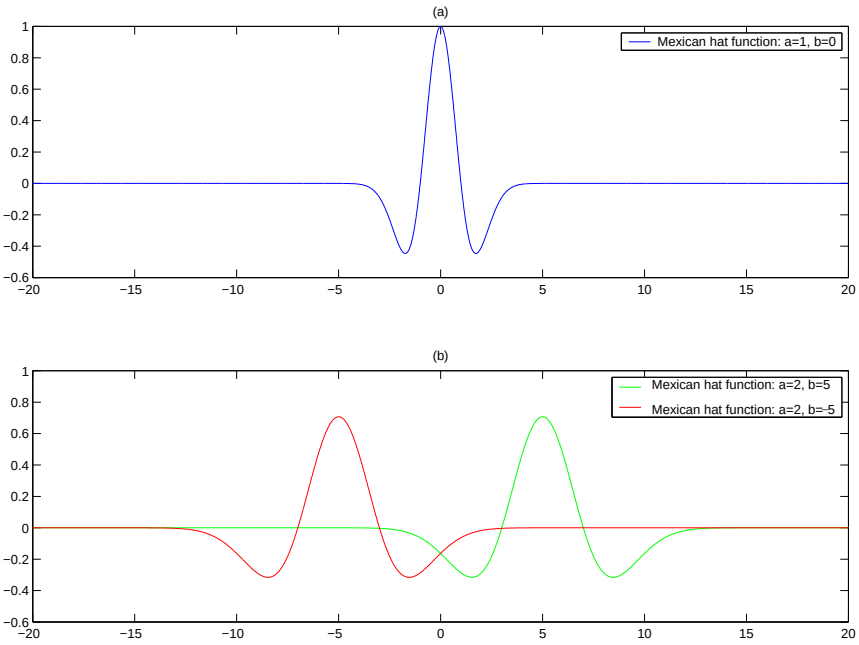

Figure 6: Example of effect of scaling and translation of the Mexican hat function. Note that (a) corresponds to a high frequency whereas in (b) the functions are at a lower frequency and are less localized.

The original function f can be reconstructed from its continuous wavelet transform by evaluating

$$f(t) = C_\psi^{-1} \int\int a^{-2}\, C(a,b)\, \psi_{a,b}(t)\, da\, db, \tag{4}$$

where the constant C_ψ satisfies an **admissibility condition** given by

$$C_\psi = 2\pi \int |\xi|^{-1} \left|\hat{\psi}(\xi)\right|^2 d\xi < \infty.$$

This reconstruction formula (known as the inverse wavelet transform) is proved in [9]; however it is rarely used in practice.

We can visualize the coefficients associated with the CWT, and an example is displayed in Figure 7 for the signal in Figure 5. This particular wavelet transform is based on the Morlet wavelet (see section 2.7). The horizontal axis corresponds to the position in the signal and the vertical axis to the scale (low scale at the bottom). The colour coding indicates the size of the coefficient with black denoting small through red to white as large. There are some clear features in the plot which we discuss below.

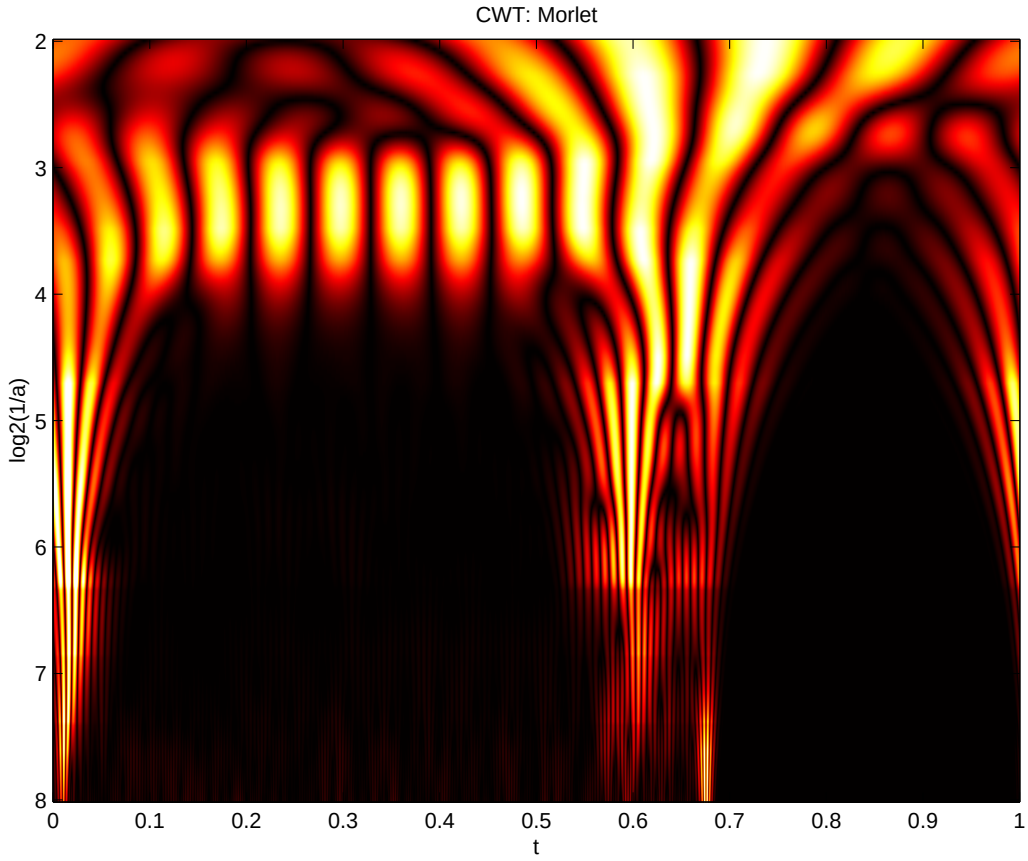

Figure 7: Representation of the wavelet coefficients from the continuous wavelet transform based on the Morlet wavelet. The colour coding indicates the size of the coefficient with black denoting small through red to white denoting large.

Interpretation: The dominant colour in the figure is black - this means that the majority of the wavelet coefficients are zero or close to zero. Consequently these coefficients contribute little to the representation and, for compression purposes, could be replaced by zero. From the bottom of the figure (corresponding to high frequency) there are two areas of non-zero coefficients decreasing in frequency to the main pattern of ovals. The ovals can be interpreted as corresponding to the underlying undamped oscillation that is seen in Figure 5, and the two areas leading up to them in frequency correspond to the resonant behaviour at startup and the first reflection. The lower frequencies show an interference into the region where the signal is zero since the support of the wavelets is quite large at these frequencies.

Note that the signal analysed is not continuous, consisting of a finite number (2^{11}) of data points. This fact puts a maximum limit on the scale parameter a in the decomposition, while for computational purposes the parameters a and b are discretized as well.

2.3 Discrete Wavelet Transform (DWT)

The key difference in a discrete wavelet analysis is that the scale parameter a and translation parameter b in equation (1) are no longer continuous, but instead are integers. Indeed, in the majority of cases, the choice of a and b is limited to the following discrete set:

$$(j,k) \in \mathbb{Z}^2, \qquad a = 2^j, \qquad b = k2^j = ka. \tag{5}$$

The indices a and b in $\psi_{a,b}$ are replaced by j and k respectively, and so equation (1) becomes

$$\psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right), \qquad (j,k) \in \mathbb{Z}^2.$$

Reconstruction of a signal from its DWT is possible provided the wavelet satisfies certain conditions (these are discussed in the following section). The reconstruction formula is analogous to equation (4):

$$f(t) = \sum_{j,k} \langle f, \psi_{j,k} \rangle \, \psi_{j,k}(n), \qquad (j,k) \in \mathbb{Z}^2. \tag{6}$$

In the next section, we introduce an alternative representation for the function f.

Of course, in many practical applications, while functions or signals are continuous, digital-to-analogue conversion yields a discrete form. From a signal processing point of view, the DWT combines the high and low pass filters $\mathcal{G}$ and $\mathcal{H}$ with a **decimation** operator $\mathcal{D}$ that selects alternate members of the sampled signal to get new filters:

$$\mathcal{G}_0 = \mathcal{D}\mathcal{G}, \qquad \mathcal{H}_0 = \mathcal{D}\mathcal{H}.$$

The classic example of $\mathcal{D}$ will keep all the even (or odd) indexed samples. In signal processing terms $\mathcal{D}$ corresponds to a downsampling. In order to reconstruct the signal upsampling is used so that, instead of decimating, zeros are added in before filtering and addition. Because of the compact support of the wavelet filters, the DWT is a fast algorithm: for a signal of length n the decomposition only requires $\mathcal{O}(n)$ operations.

We can visualize the coefficients from the discrete wavelet transform as shown in Figure 8 for the signal in Figure 5. The horizontal axis corresponds to position in the signal and the vertical axis to the scale or frequency with high frequencies at the bottom and low frequencies at the top. At each level we have plotted the signed magnitude of the coefficients.

Interpretation: We see that most of the wavelet coefficients are relatively small and hence for compression purposes we could consider setting them to zero. The large wavelet coefficients correspond to the undamped oscillation and these disappear where the function is set to zero after the discontinuity. The mixed frequency at the start and the first reflection also give rise to non-zero coefficients associated with higher frequencies.

2.4 Construction of Wavelet Bases

Note: While the level of the mathematics contained within this section is quite high, it is included only for completeness. For the reader who is interested in using wavelets, a total understanding of the following concepts is not essential.

The most common and general approach to constructing a wavelet basis is to use a **Multiresolution Analysis**. For an overview see [21] or [9]. This approach leads to the natural construction using high and low pass filters $\mathcal{G}$ and $\mathcal{H}$ discussed in section 2.1. In some parts of the wavelet literature, these filter matrices are called **refinement matrices** .

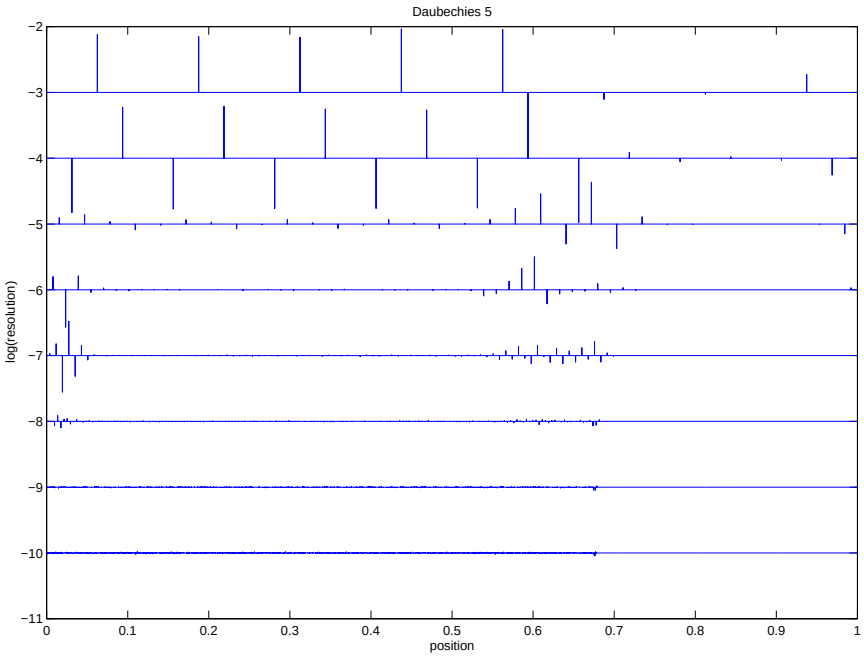

Figure 8: Representation of the wavelet coefficients from the discrete wavelet transform based on the Daubechies 5 wavelet. The horizontal axis corresponds to the position in the signal and the vertical axis to the scale or frequency.

We start by recalling that two functions f_1 and f_2 are orthonormal if

$$\langle f_1, f_1 \rangle = \langle f_2, f_2 \rangle = 1, \tag{7}$$

and

$$\langle f_1, f_2 \rangle = \langle f_2, f_1 \rangle = 0. \tag{8}$$

Also recall that a set of basis vectors $\{e_n\}$ is an **unconditional basis** for a space E if $\{e_n\}$ are linearly independent and the order in which the basis vectors are taken does not matter. For a Hilbert space H an unconditional basis is often called a **Riesz basis**. If A is a bounded operator with a bounded inverse then A maps any orthonormal basis to a Riesz basis.

To generate a (orthonormal) wavelet basis take *two* (orthonormal) wavelets: the **scaling function** (or **father wavelet**) ϕ and **mother wavelet** ψ. Other wavelets in the basis are then generated by translations of the scaling function ϕ and dilations and translations of the mother wavelet ψ using the relationships:

$$\phi_{J_0,k}(t) = 2^{-J_0/2}\phi\left(2^{-J_0}t - k\right), \qquad \psi_{j,k}(t) = 2^{-j/2}\psi\left(2^{-j}t - k\right), \qquad j = J_0, J_0+1, \ldots, \tag{9}$$

where $k \in \mathbb{Z}$.

Figure 14 shows two examples of wavelet and scaling functions.

The dilations and translations of the mother wavelet ψ form a basis for the detail spaces $\{W_j\}$ found by applying the high pass filter $\mathcal{G}$ whereas the scaling functions ϕ form a basis for the approximation spaces $\{V_j\}$.

We can now offer an alternative to equation (6) for the **discrete wavelet representation** of a function f:

$$f(t) = \sum_{k \in Z} c_{J_0,k}\phi_{J_0,k}(t) + \sum_{j=J_0}^{\infty}\sum_{k \in Z} \gamma_{j,k}\psi_{j,k}(t), \tag{10}$$

where

$$c_{J_0,k} = \langle f, \phi_{J_0,k} \rangle \qquad \text{and} \qquad \gamma_{j,k} = \langle f, \psi_{j,k} \rangle .$$

By using the scaling function (from the approximation space), we can convert the double sum over scales in equation (6) to a single-sided sum. Equation (10) with $J_0 = 0$ more accurately describes the wavelet representation used in practice.

Large features are represented in the first term by the scaling functions $\{\phi_{J_0,k}\}$ (the approximation in the space V_k). The wavelets $\{\psi_{j,k}\}$ represent oscillations localized to positions near the point $k2^{-j}$ at frequencies near 2^j and hence give *details* in the signal.

A multiresolution analysis (MRA) (sometimes also known in the literature as a multiscale analysis) formalizes this type of analysis. A MRA of L^2 is defined as a sequence of closed subspaces V_j of L^2, $j \in \mathbb{Z}$, satisfying the following properties [21]:

- $V_{j+1} \subset V_j$,
- $v(t) \in V_j \Leftrightarrow v(2t) \in V_{j+1}$,
- $v(t) \in V_0 \Leftrightarrow v(t+1) \in V_0$,
- $\bigcup_{j=-\infty}^{\infty} V_j$ is dense in L^2 and $\bigcap_{j=-\infty}^{\infty} V_j = 0$,
- A scaling function $\phi \in V_0$ with non-vanishing integral exists such that $\{\phi(t-l) | l \in \mathbb{Z}\}$ is a Riesz basis of V_0.

The first property gives a sequence of nested spaces as shown in the schematic of Figure 2. The second and third statements give the dilation and translation properties. The fourth statement says that the spaces V_j must give an approximation of the space L^2. Finally we have the property that a basis must exist for the space V_0. It follows from the definition of the MRA that there exists a sequence (h_k) such that the scaling function ϕ satisfies

$$\phi(t) = \sum_k h_k \phi(2t-k). \tag{11}$$

This is known as the **refinement equation**, **dilation equation** or **two-scale difference equation**. The wavelet or *detail* spaces W_j are spaces complementing V_j in V_{j-1} i.e. a space that satisfies

$$V_{j-1} = V_j \oplus W_j.$$

This says that every element of V_{j-1} can be written as a sum of an element in V_j and W_j in a unique way, from which we get the decomposition and reconstruction schemes illustrated in Figures 2 and 3.

2.5 Types of Wavelets

Definition: A function ψ is defined as a WAVELET if the collection of functions $\{\psi(t-k) | k \in \mathbb{Z}\}$ is a Riesz basis of W_0. The collection of wavelet functions $\{\psi_{j,k} | j,k \in \mathbb{Z}\}$ is then a Riesz basis of $L^2(\mathbb{R})$.

The most common types of wavelets used are:

Orthogonal Wavelets: $\{\psi(t-k) | k \in \mathbb{Z}\}$ is an orthonormal basis of W_0. Orthogonality is often considered as desirable since:

1. The L^2 norm of the function f is directly related to the norm of the coefficients, i.e. $\|f\| = \sum_{j,k} |\gamma_{j,k}|^2$, and so the replacing of coefficients by zero (e.g. in compression) can be justified in terms of its effect on the fitting measure.

2. The wavelet transform is a unitary transformation (i.e. $A^{-1} = A^T$) and so leads to stable numerical computations. Also, if f is corrupted by independent white noise then so are the wavelet coefficients.

Bi-orthogonal Wavelets: A dual scaling function $\hat{\phi}$ and a dual wavelet function $\hat{\psi}$ which generate a dual MRA can be introduced: one is used for the decomposition and the other for the reconstruction. In terms of the low and high pass filters, bi-orthogonality is expressed as

$$\begin{pmatrix} \mathcal{H}_j \\ \mathcal{G}_j \end{pmatrix}^T \begin{pmatrix} \mathcal{H}_j^* \\ \mathcal{G}_j^* \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}. \tag{12}$$

Real and Complex Wavelets: Wavelets are not restricted to real values. A complex wavelet function will return information about both amplitude and phase. In general, a complex wavelet is suited to oscillatory behaviour, whereas real wavelets are useful for isolating discontinuities and "peak" features.

Boundary Wavelets: We have assumed here that either we are working over the whole of $\mathbb{R}$ or $\mathbb{C}$ or that the interval can be extended periodically. There is a great deal of work in the literature on how to treat boundaries, in particular by deriving boundary filters.

Second Generation Wavelets: These are wavelets which are not necessarily translations and dilations of a single fixed function, and because of the extra freedom they have an adaptability to new applications. One method for constructing these wavelets is the *Lifting Scheme* (see [35]). At the filter level, this adaption is achieved by looking at a bi-orthogonal system and introducing a new set of filters $\{\tilde{\mathcal{H}}_j, \tilde{\mathcal{G}}_j, \tilde{\mathcal{H}}_j^*, \tilde{\mathcal{G}}_j^*\}$ and matrix S such that equation (12) is preserved so

$$\begin{pmatrix} \tilde{\mathcal{H}}_j^* \\ \tilde{\mathcal{G}}_j^* \end{pmatrix} = \begin{pmatrix} I & S \\ 0 & I \end{pmatrix} \begin{pmatrix} \mathcal{H}_j^* \\ \mathcal{G}_j^* \end{pmatrix}, \qquad \begin{pmatrix} \tilde{\mathcal{H}}_j \\ \tilde{\mathcal{G}}_j \end{pmatrix} = \begin{pmatrix} I & 0 \\ -S^* & I \end{pmatrix} \begin{pmatrix} \mathcal{H}_j \\ \mathcal{G}_j \end{pmatrix}.$$

Through the choice of the operator S, particular properties can be introduced.

2.6 Redundancy in the Representation

There is an advantage to introducing redundancy into the wavelet representation of a function as it allows for further basis selection and hence better approximations.

2.6.1 Wavelet Packets

Wavelet packets are an extension to the standard wavelet basis constructed by a multiresolution analysis. This extension is achieved through a property of Riesz basis functions such that given one basis another can be constructed based on dilations and translations (see [21] for an overview and further references to theory). We have essentially already used this property once in the MRA when passing from the approximation space V_j to V_{j+1}.

The key difference in a wavelet packet analysis is that the detail space W_j also undergoes a wavelet decomposition at each level. A level of redundancy is added to the decomposition and this is illustrated in Figure 9 in which the space V_0 is split into subspaces generated by wavelets (this is over 3 levels). The dashed rectangles correspond to the wavelet MRA $V_0 = W_1 \oplus W_2 \oplus W_3 \oplus V_3$. The bold rectangles correspond to a possible wavelet packet splitting.

The main advantage of this representation is that we have more freedom in deciding which basis functions to use to represent a given signal, and there is increased flexibility in adapting the basis to the frequency contents of the signal.

We can choose the set of coefficients in the full binary tree (as illustrated in Figure 9) to represent the function optimally with respect to a certain criterion. This procedure is called **best basis**

selection, where, in general, the criterion depends on the application under consideration, and in the software packages we have used ([26, 11]), is usually entropy-based.

Figure 10 shows a wavelet packet decomposition of the signal in Figure 5 obtained using the Coiflet 3 wavelet, while Figure 11 shows the best basis tree obtained using entropy based on Coifman-Wickerhauser ([6]).

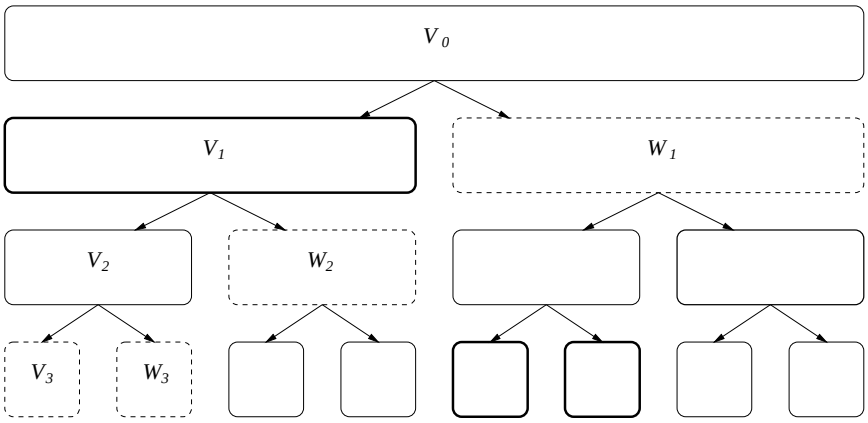

Figure 9: Schematic illustration of a space V_3 split by wavelet packets. The dashed boxes correspond to the MRA whereas the bold boxes indicate a possible wavelet packet splitting.

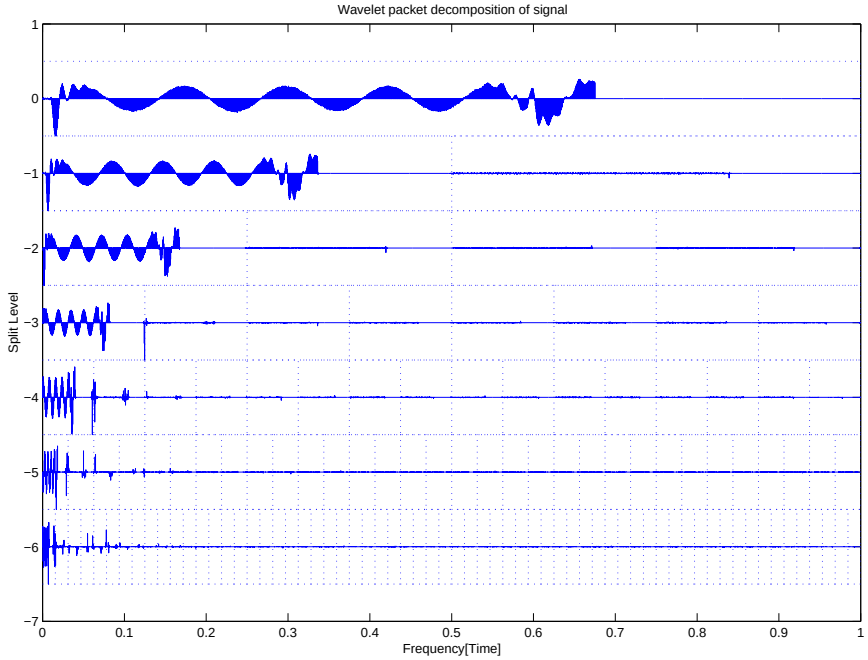

Figure 10: A wavelet packet decomposition by level of the signal given in Figure 5 using the Coiflet 3 wavelet.

See [38] for an acoustic application on best basis selection, while for a summary of basis selection for feature extraction see [19]. The latter reference includes a list of techniques (entropy based algorithms; basis, matching and discriminant pursuit) and references, and very briefly describes an application to flutter and other acoustic signals.

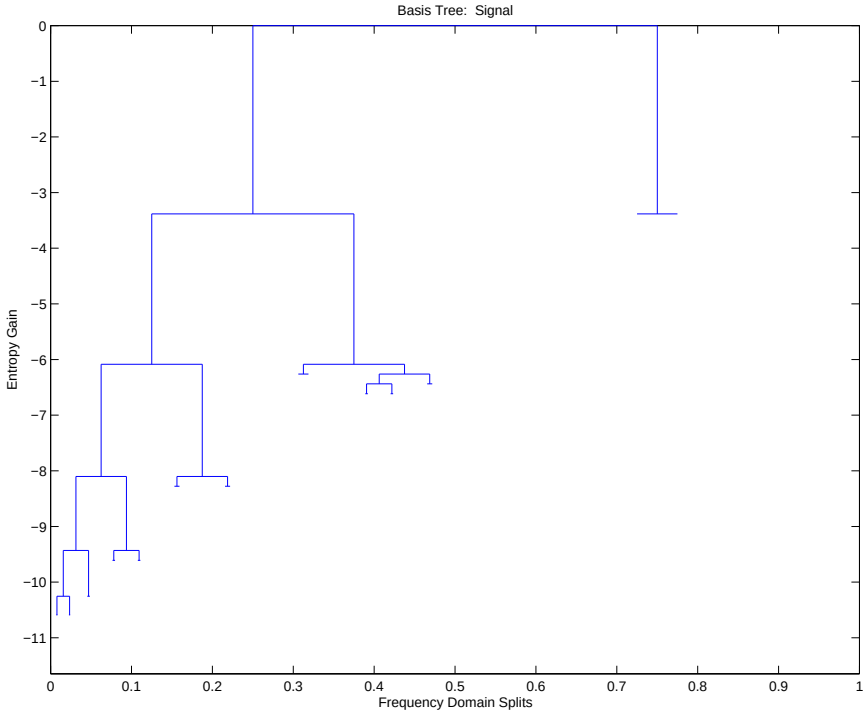

Figure 11: A best-basis wavelet tree for the signal given in Figure 5.

2.6.2 Non-Decimated Wavelet Transform (NDWT)

The non-decimated wavelet transform (also known as a stationary wavelet transform) is a further example of a redundant wavelet representation. From a signal processing view, the key to the NDWT is that both even and odd decimations are used and the results from both are kept at each stage. This decomposition stretches the filters $\mathcal{G}$ and $\mathcal{H}$ by a factor of 2 at each stage. For a given data set, the non-decimated wavelet transform contains the standard discrete wavelet transform for every possible choice of origin in the data. Since the NDWT is over-determined, a choice of basis is required to reconstruct the function from the wavelet coefficients. This choice can either be determined by basis selection or instead (as suggested by [28]) by an average basis approach. The NDWT is similar to a wavelet packet approach where each packet corresponds to either an odd or even decimation and so the coefficients may be viewed either in a time-ordered or packet-ordered fashion.

The NDWT in some sense fills in the gaps in the DWT and so may be thought of as closer to the continuous wavelet transform. Computation of the NDWT takes $\mathcal{O}(n \log n)$ operations for a signal of length n. The advantage of the NDWT can be seen, for example, in signal de-noising.

We can represent the results of a NDWT in a similar way to the Wavelet Packet and DWT by plotting the coefficients on each level. An alternative is a *Multiresolution Analysis Representation* of the wavelet. Here each wavelet space W_j is reconstructed across each scale. Figure 12 shows the NDWT for the signal in Figure 5, with the horizontal axis representing the position and the vertical axis the scale. Note that the vertical scale here has the high frequencies at the top and low frequencies at the bottom.

Interpretation: The frequency mix and sine waves are clearly visible in this representation as is the discontinuity to zero. Note that the lower frequencies are non-zero in the region where the signal is zero due to the larger support (non-zero part) of the low frequency wavelets.

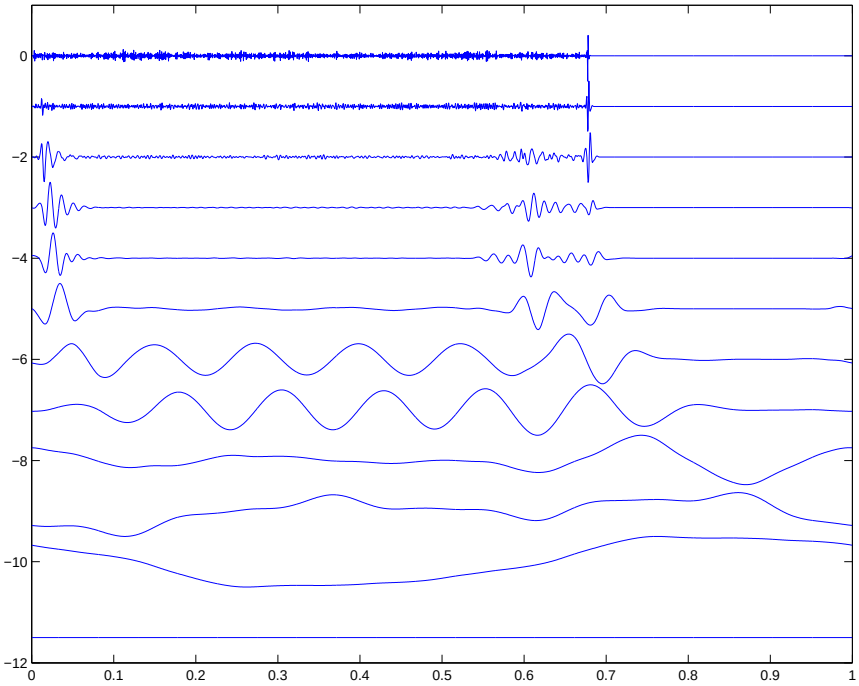

Figure 12: A *Multi Resolution Analysis* representation of the wavelet coefficients from the NDWT based on the Symmlet 6 wavelet. The horizontal axis corresponds to the position in the signal and the vertical axis to the scale or frequency. Note that here the vertical axis is inverted so the high frequencies are at the top of the figure.

2.7 Properties of Wavelets

We can summarize the basic properties of wavelets as follows:

1. Wavelets form an unconditional (Riesz) basis for a wide variety of function spaces (Hilbert spaces).

2. Wavelets are either orthogonal or the dual (bi-orthogonal) wavelets are known.

3. Wavelets (and their duals) are local in space and frequency.

4. Wavelets fit into the framework of a multiresolution analysis which leads to the *fast wavelet transform* to find the wavelet coefficients $\{\gamma_{j,k}\}$.

There are many wavelet families whose qualities vary according to several criteria, and in general the choice of which family to use is motivated by the specific problem under investigation. Desirable properties include:

1. **Compact Support:** A function is said to have compact support when it is non-zero over a finite range. If the scaling function ϕ and wavelet ψ are compactly supported, the filters $\{\mathcal{G}, \mathcal{H}\}$ are finite impulse response filters so that the summations in the discrete wavelet transform are finite. If ϕ and ψ are not compactly supported, then it is desirable to have fast decay to zero so that the filters may be reasonably approximated. For example, the Haar wavelet and Daubechies wavelets (see Figure 13) have compact support, whereas the Mexican hat wavelet (see Figure 6) has infinite support but with fast decay to zero.

2. **Smoothness:** The smoothness of the wavelet is related to the support of the wavelet – the smoother the wavelet, the larger the support. A higher number of **vanishing moments** is desirable for smooth signals.

   If the wavelet has m vanishing moments then

   $$\int t^j \psi(t) dt = 0, \qquad j = 0, \ldots, m. \tag{13}$$

   To detect a discontinuity in the mth derivative of a function, it is necessary to select a wavelet with at least m vanishing moments.

3. **Singularity Detection:** The continuous transform (section 2.2) and NDWT (section 2.6.2) are often used in singularity detection. The typical form of result is that if a function f is Hölder continuous of order α, i.e.

   $$|f(t+h) - f(t)| = \mathcal{O}(h^\alpha),$$

   then the continuous wavelet transform has an asymptotic behaviour:

   $$C(a,b) = \mathcal{O}(a^{\alpha+1/2}) \text{ for } a \longrightarrow 0.$$

   Loosely speaking, the size of a wavelet coefficient indicates the size of the derivative of the function at a particular point.

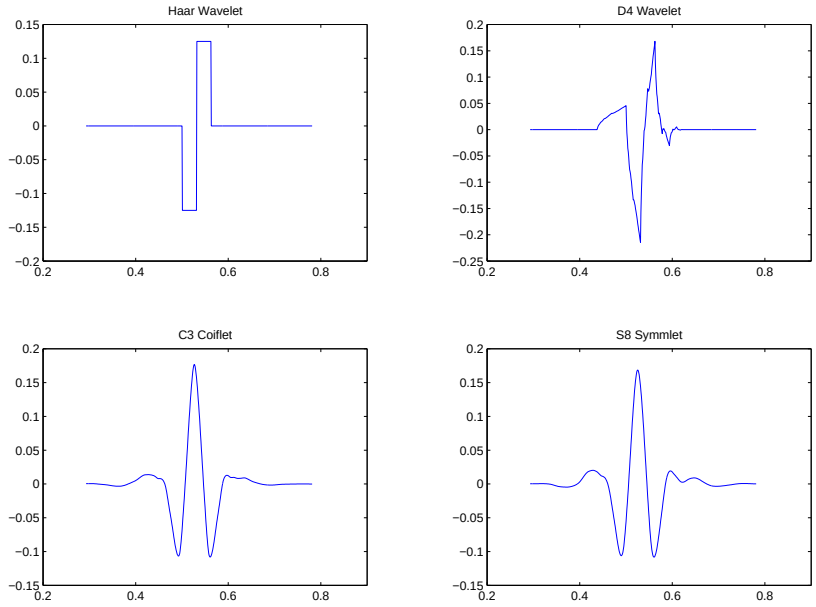

Figure 13: Examples of 4 types of wavelets: the Haar (DB1), the Daubechies 4 wavelet (DB4), Coiflet (C3), and Symmlet (S8).

Common examples of wavelet families (and their properties) are:

Haar: The simplest examples of orthogonal wavelets and the first in the Daubechies family of wavelets.

Meyer wavelet and scaling function. These belong to C^∞ and decay faster than any inverse polynomial, though not exponentially fast. The Fourier transform is compactly supported. The wavelet has an infinite number of vanishing moments.

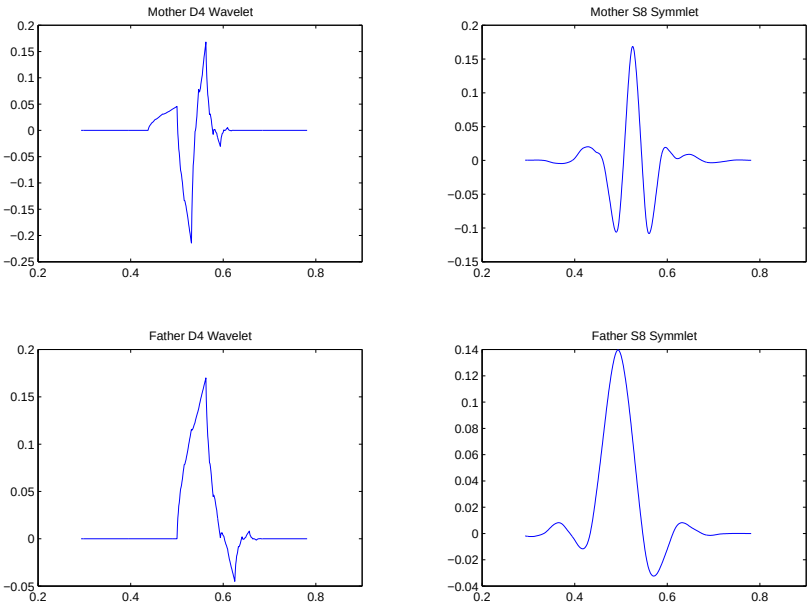

Figure 14: Wavelet and scaling functions for the Daubechies DB4 and Symmlet S8.

Battle–Lemarié wavelets are constructed by orthogonalizing B–spline functions. They have exponential decay. The wavelet with N vanishing moments is a piecewise polynomial of degree $N-1$ that belongs to C^{N-2}.

Morlet: These are complex, symmetrical wavelets that are C^∞ smooth. They do not however form an orthogonal basis. The wavelets are given by a plane wave modulated by a Gaussian

$$\psi(x) = \pi^{-1/4} e^{i\omega x} e^{-x^2/2},$$

where ω is the non-dimensional frequency.

Daubechies: Orthogonal basis indexed by the number of vanishing moments $N(\in \mathbb{N})$ with compact support on an interval $2N-1$ and finite number of continuous derivatives (regularity increases with N). The Haar wavelet corresponds to $N=1$. See Figures 13 and 14 in which we have plotted the wavelet and scaling functions.

Symmlets: These are versions of the Daubechies wavelets with aditional symmetry properties (note though that these are not truly symmetric as can be seen in Figure 13 and 14, in which we have plotted the wavelet and scaling functions.

Coiflets: These are compactly supported and ψ has $2N$ vanishing moments (see Figure 13).

Harmonic: These are complex wavelets discussed in detail in [32, 31] which are useful for vibration analysis since phase information is obtained. The algorithm, like that for the NDWT, yields an over-determined system with K coefficients for all the possible K positions along the time axis.

A more complete table of wavelets and their properties can be found either in the book by Daubechies [9] or in the MATLAB Wavelet Toolbox manual [26] or by using the command *waveinfo* in the MATLAB Wavelet Toolbox.

In a wavelet decomposition, we find at each level a set of wavelet coefficients giving the detail and a set of coefficients for the approximation space V_j. Two principal methods for analysing this information are:

1. Examine the magnitude of wavelet coefficient $\gamma_{j,k}$ on each level j. It may be useful to look at the absolute value $|\gamma_{j,k}|$ and either the relative or absolute magnitudes across levels. In

Figure 5, we plotted the function and in Figure 8 we plotted the wavelet coefficients from the DWT compared relatively between scales. It can be useful to compare the different scales absolutely (plotting each scale with its own particular axis scaling) as doing so shows up large coefficients occurring through the scales. This approach is useful for feature detection.

2. We can construct the representation in the space W_j (i.e. on each level j) sometimes called a multiresolution analysis represention (see section 2.6.2, Figure 12). This construction gives a picture of the detail at each frequency level J, and can be useful for spotting structure in a signal. In Figure 15 we have plotted the constructed detail for the wavelet coefficients in Figure 8.

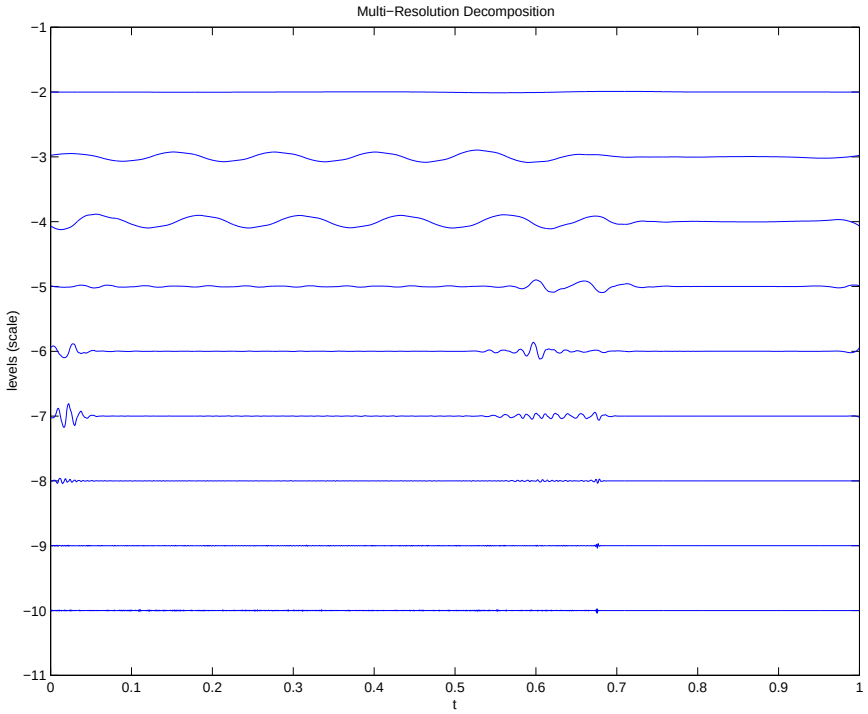

Figure 15: Representation of the wavelets at each level from the discrete wavelet transform based on the Daubechies 10 wavelet. The horizontal axis corresponds to the position in the signal and the vertical axis to the scale or frequency.

In Figures 16 and 17, we consider DWTs of the signal in Figure 5 obtained using different wavelets, looking at the coefficients and the reconstructed wavelet spaces across each level. It is interesting to note the effect that the smoothness has on the decomposition. Similar remarks hold for the NDWT (see Figure 18) and the CWT.

Undesirable Properties/Problem Areas: Wavelets are a relatively new field and as such the theory is still under developement. In general, wavelet methods were conceived for investigating periodic or infinite domains and consequently treating boundaries correctly can be a non-trivial task. Originally the number of data points required was a power of 2, and although theoretically this is no longer the case, it is still assumed in a number of software packages. There are however some standard techniques to deal with general numbers of data points, the simplest of which is to pad the signal with zeros to the desired length (as we have done in Figures 4 and 5).

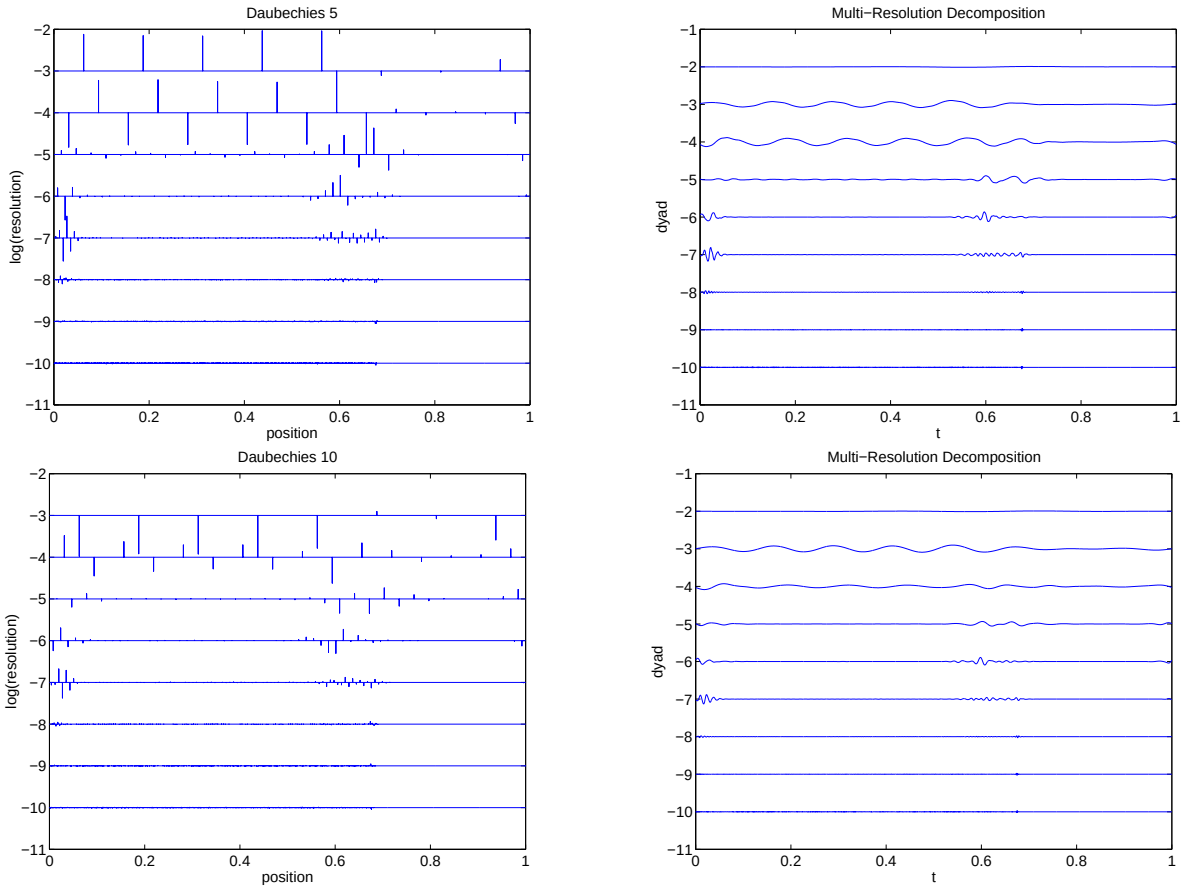

Figure 16: Comparison of the coefficients and reconstructed wavelet spaces in the DWTs of the signal in Figure 5 obtained using Daubechies 5 and 10.

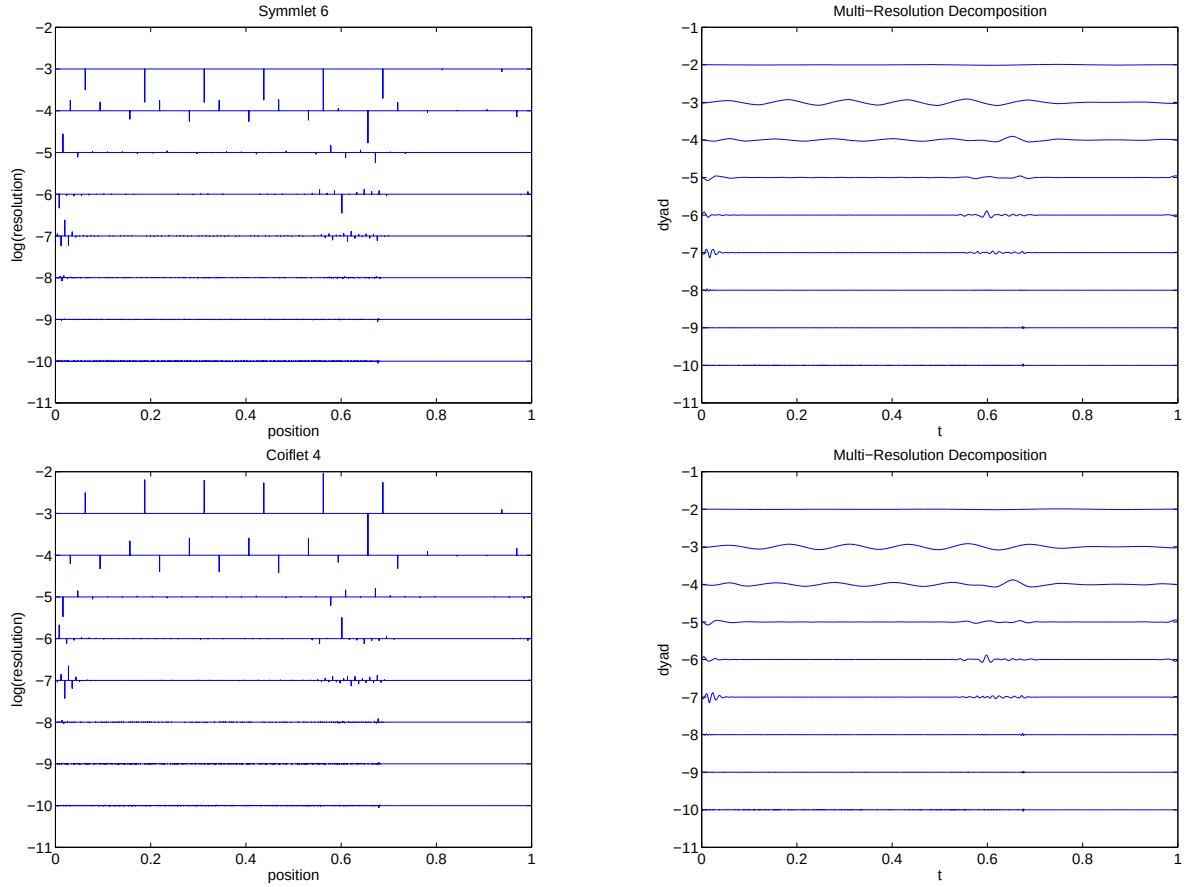

Figure 17: Comparison of the coefficients and reconstructed wavelet spaces in the DWTs of the signal in Figure 5 obtained using Symmlet 6 and Coiflet 4.

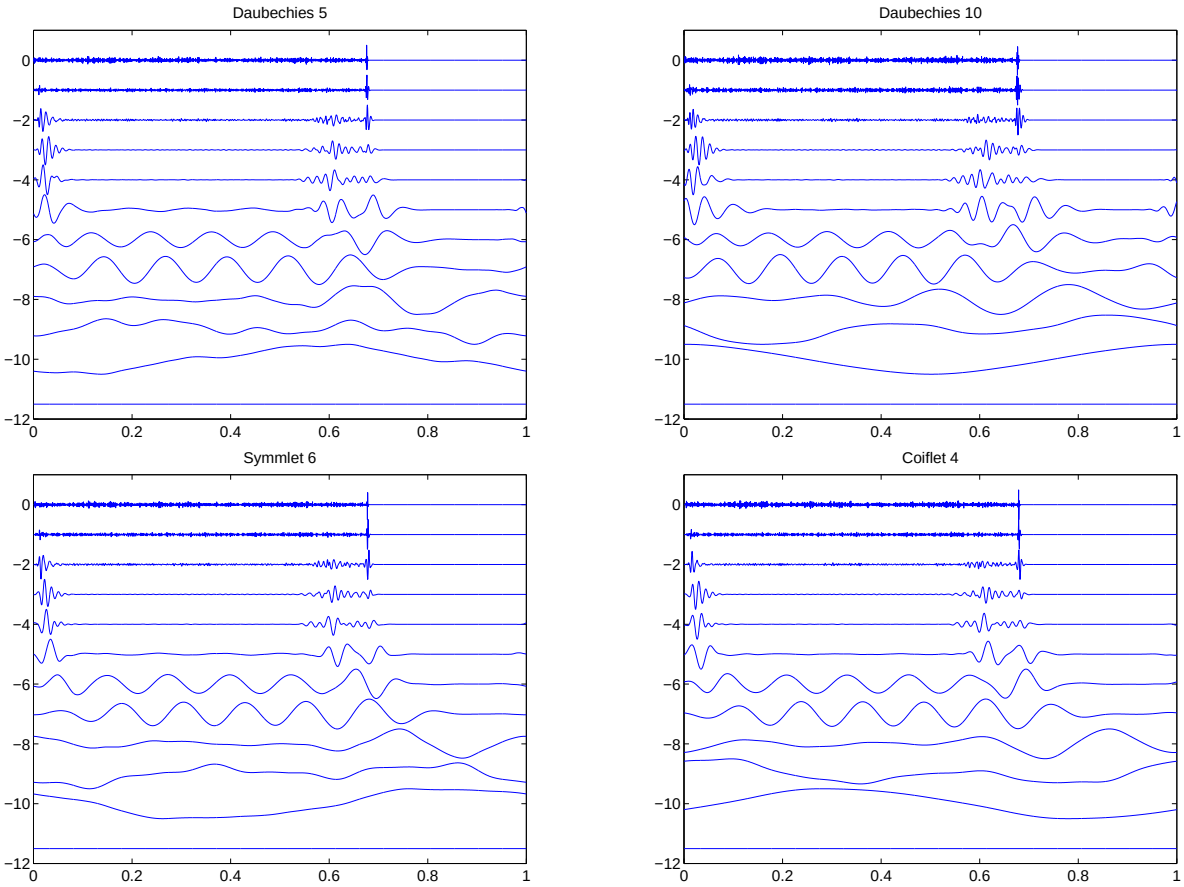

Figure 18: Comparison of the reconstructed wavelet spaces in the NDWTs of the signal in Figure 5 obtained using Daubechies 5 and 10, Symmlet 6 and Coiflet 4.

2.8 Extension to Higher Dimensions

Note: Again, it is not essential for the reader to understand all of the mathematical details contained within this section.

While the previous exposition deals mainly with one-dimensional signals, wavelets can also be used to analyse signals in higher dimensions. We consider only two dimensions, but extension to higher dimensions is straightforward. This section is based on [9, Chapter 10].

If

$$\{\psi_{j,k}(t)|j,k\in\mathbb{Z}\}=\{2^{-j/2}\psi(2^{-j}t-k)|j,k\in\mathbb{Z}\}$$

represents an orthonormal wavelet basis for $L^2(\mathbb{R})$ (the space of square-integrable complex-valued functions on $\mathbb{R}$), then an orthonormal basis for $L^2(\mathbb{R}^2)$ is obtained by taking the tensor product functions generated by two one-dimensional bases

$$\Psi_{j_1,k_1;j_2,k_2}(t_1,t_2)=\psi_{j_1,k_1}(t_1)\psi_{j_2,k_2}(t_2). \tag{14}$$

Note that in this basis, the two variables t_1 and t_2 are dilated separately.

Another way to form an orthonormal basis for $L^2(\mathbb{R}^2)$ involves the tensor product of two one-dimensional multiresolution analyses. Define the bivariate spaces $\mathbf{V}_j$ by $\mathbf{V}_j=V_j\otimes V_j$ so that

- $\mathbf{V}_{j+1}\subset\mathbf{V}_j$,
- $\cup_{j=-\infty}^{\infty}\mathbf{V}_j$ is dense in L^2 and $\cap_{j=-\infty}^{\infty}\mathbf{V}_j=0$.

These properties have the same meaning as described in the one-dimensional case in section 2.4.

Since $\{\phi(t-k)|k\in\mathbb{Z}\}$ are an orthonormal basis for V_0, the product functions

$$\Phi_{0;n_1,n_2}(x,y)=\phi(x-n_1)\phi(y-n_2),\qquad n_1,n_2\in\mathbb{Z}, \tag{15}$$

form an orthogonal basis for $\mathbf{V}_0$.

Similarly

$$\begin{aligned}\Phi_{j;n_1,n_2}(x,y) &= \phi_{j,n_1}(x)\phi_{j,n_2}(y) &(16)\\ &= 2^{-j}\Phi(2^{-j}x-n_1,2^{-j}y-n_2),\qquad n_1,n_2\in\mathbb{Z}, &(17)\\ & &(18)\end{aligned}$$

form an orthogonal basis for $\mathbf{V}_j$.

Similar to the one-dimensional case, we define $\mathbf{W}_j$ to be the orthogonal complement in $\mathbf{V}_{j-1}$ of $\mathbf{V}_j$.

$$\begin{aligned}\mathbf{V}_{j-1} &= V_{j-1}\otimes V_{j-1} &(19)\\ &= (V_j\oplus W_j)\otimes(V_j\oplus W_j) &(20)\\ &= (V_j\otimes V_j)\oplus(V_j\otimes W_j)\oplus(W_j\otimes V_j)\oplus(W_j\otimes W_j) &(21)\\ &= \mathbf{V}_j\oplus\mathbf{W}_j, &(22)\end{aligned}$$

i.e. $\mathbf{W}_j$ is made up of 3 parts, with orthogonal bases given by

$$\begin{aligned}\psi_{j,n_1}(x)\phi_{j,n_2}(y) &\quad\text{for}\quad W_j\otimes V_j, &(23)\\ \phi_{j,n_1}(x)\psi_{j,n_2}(y) &\quad\text{for}\quad V_j\otimes W_j, &(24)\\ \psi_{j,n_1}(x)\psi_{j,n_2}(y) &\quad\text{for}\quad W_j\otimes W_j, &(25)\end{aligned}$$

and we can define 3 wavelets

$$\Phi^h(x,y) = \phi(x)\psi(y), \tag{26}$$
$$\Phi^v(x,y) = \psi(x)\phi(y), \tag{27}$$
$$\Phi^d(x,y) = \phi(x)\phi(y), \tag{28}$$

and $\{\Phi^\lambda_{j;n_1,n_2} | n_1, n_2 \in \mathbb{Z}, \lambda = h, v \text{ or } d\}$ form an orthogonal basis for $\mathbf{W}_j$.

In a two-dimensional array, filtering can be done on the 'rows' and the 'columns' (corresponding to horizontal and vertical directions in images) as shown in Figure 19, and we obtain the two-dimensional analogue of Figure 2.

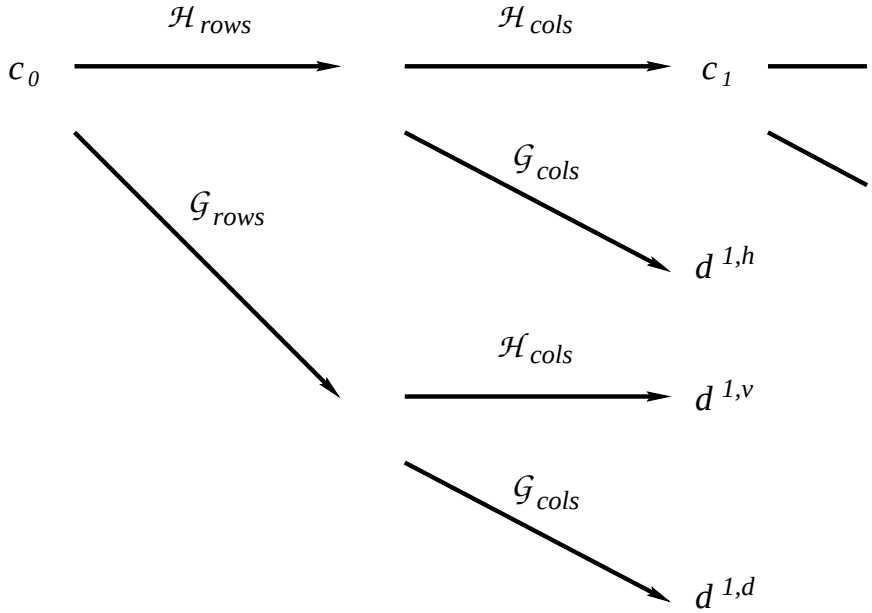

Figure 19: Schematic view of a two-dimensional wavelet decomposition achieved by repeated low- and high-pass filtering, on rows and columns.

The $d^{1,h}$ correspond to the wavelet coefficients $\langle F, \Phi^\lambda_{1;n_1,n_2}\rangle$, with

$$F = \sum_{n_1,n_2} c^0_{n_1,n_2} \Phi_{0;n_1,n_2}.$$

In an image analysis, horizontal edges show up in $d^{1,h}$, vertical edges in $d^{1,v}$ and diagonal edges in $d^{1,d}$. At each scale, we therefore obtain 3 images ($d^{j,h}$, $d^{j,v}$ and $d^{j,d}$) and an approximation image c^j, and if the original image consists of $N \times N$ elements, then the $d^{1,\lambda}$ are $N/2 \times N/2$ arrays. The decomposition can then be represented as shown in Figure 20.

Figure 21 shows the level-2 decomposition (bottom) of an example image (top) using the Daubechies 2 wavelet performed by the MATLAB Wavelet Toolbox. The image is a contour map of out-of-plane displacements of an injection-moulded plastic plate. Note how the boundary of the box is detected in the horizontal and veritcal coefficients in the decomposition.

2.9 Wavelets as a Computational Basis

One obvious application for wavelets is as a replacement for the standard bases used in numerical computations. Wavelets are attractive for numerical computations since they offer a natural

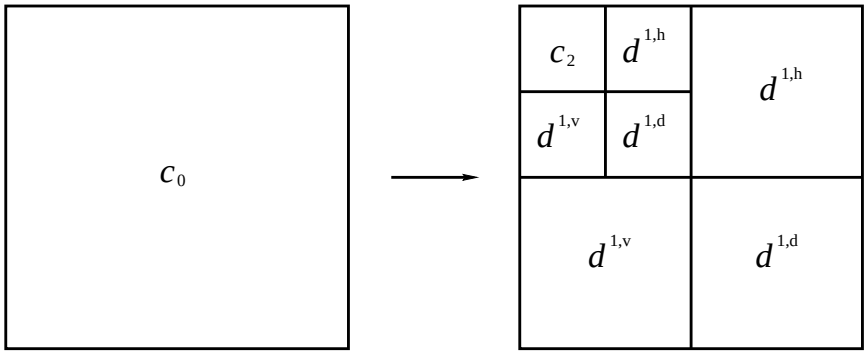

Figure 20: Schematic view of the visualisation of a two-dimensional wavelet transform, with decomposition into two layers.

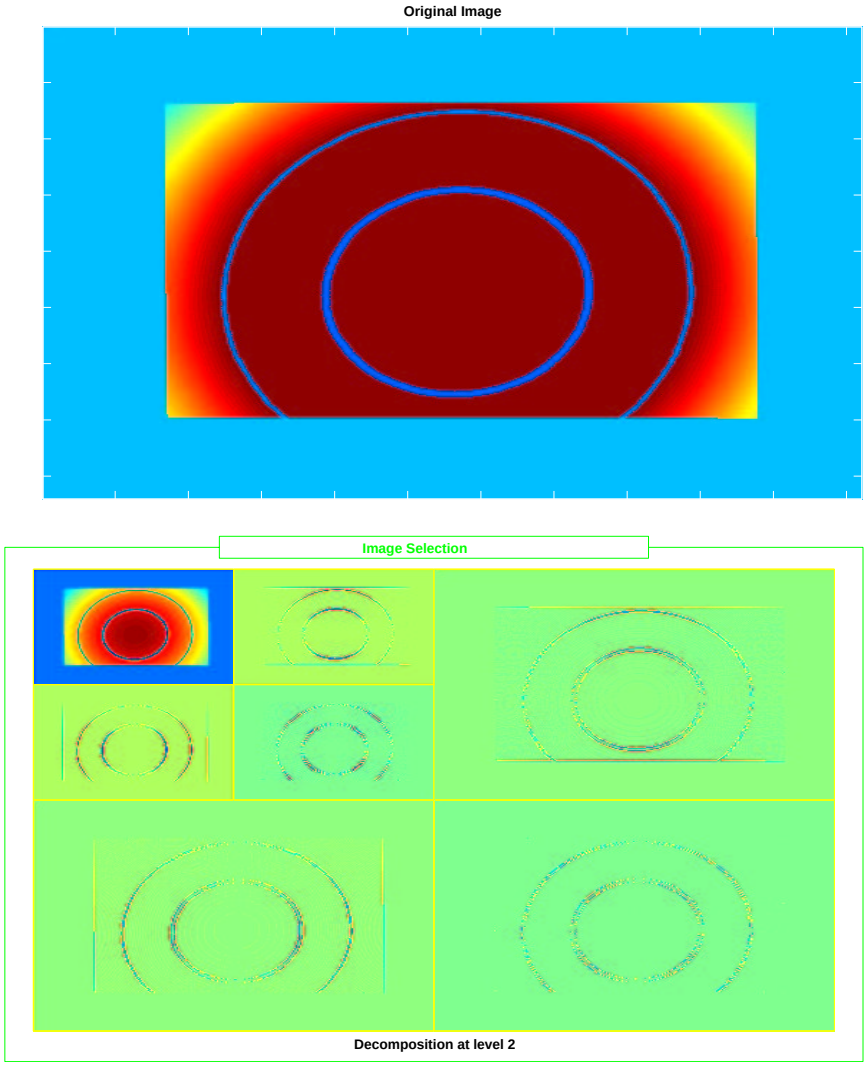

Figure 21: Original image (top) and wavelet decomposition (bottom) using Daubechies 2.

adaptability of the computational grid and hence would be expected to perform well for problems with shocks or steep boundary layers. A basic overview of wavelet methods for the solution of partial differential equations (PDEs) may be found in [36]; a more detailed mathematical view may be found in [7].

In general, two classes of wavelet methods are identified: *Galerkin* or *collocation* wavelet schemes. For a Galerkin method, the computational parameters may be thought of as the point values in a (wavelet) transformed space and so the wavelet coefficients are determined directly. In a wavelet collocation method, the computational parameters are point values of the variables in physical space and the wavelet coefficients are used to determine where to adapt the grid in the physical domain. As pointed out in [20], there are a number of variants for each of these methods; however in each case the key is that a multiresolution analysis is performed where the depth of the resolution is determined by the wavelet coefficients.

Wavelet Galerkin methods differ depending on the choice of wavelets. A spline wavelet Galerkin method corresponds to an adaptive finite element scheme (h refinement) or an adaptive finite difference scheme. For wavelet Galerkin methods, a key difficulty at present is dealing effectively with both non-periodic boundaries and non-linear terms.

Early work in this area concentrated on Burgers equation (for an early comparison of methods, see [34]), a typical example of a hyperbolic PDE that exhibits a steepening shock wave as time progresses. Other work has concentrated on time-independent problems (typified by elliptic PDEs such as Poisson's equation). There is a growing body of applications in the literature, including Stokes and Navier-Stokes equation (see [23]).

Wavelet methods can also be applied to solve integral equations. In many cases, finding the solution to a partial differential equation can be turned into a problem of finding a solution to an integral equation over the boundary of the domain. Using wavelets (like pre-conditioning) this linear system reduces to a sparse linear system, leading to fast algorithms (see [7]).

3 Statistical properties

3.1 De-noising and smoothing

Strictly speaking, de-noising and smoothing are two different operations. De-noising implies the reduction of noise in observations in order to highlight the underlying signal. Smoothing is an operation that produces an estimate of the trend that is less variable than the noisy signal itself. In spite of these considerations, from a statistical point of view both terms are used interchangeably, because the noise is usually modelled like a high frequency component and is removed by smoothing. In fact, statistical non-parametric regression techniques are often referred to as smoothers, while a signal processing engineer (say) will define the same kind of operations as de-noising.

Consider a set of noisy observations $\{y_i\}$. We can write

$$y_i = g(t_i) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{29}$$

where g is a modelling function and $\{\epsilon_i\}$ represents the noise, generally modelled as independent and identically distributed (often denoted by *iid*) random variables with zero mean and variance σ^2. Smoothing $\{y_i\}$ also implies de-noising, and non-parametric regression is a statistical term that includes both operations.

The reason why wavelets are an appealing methodology for smoothing (de-noising) has been explained in detail in the previous section (see, for example, Figure 1).

The non-parametric regression problem (i.e. smoothing) can be treated using wavelets by linear estimation or non-linear estimation [1].

The basic idea behind the *linear* estimation approach is that if the noise is represented by small scale variability, one way to eliminate the noise is to set to zero all the coefficients of the high (finest) levels in the wavelet representation. This idea is the same as considering the truncated wavelet estimator:

$$\hat{g}_M(t) = \hat{c}_0\phi(t) + \sum_{j=0}^{M}\sum_{k=0}^{2^j-1}\hat{\gamma}_{jk}\psi_{jk}(t), \tag{30}$$

where " ˆ " in this case denotes an estimate of the coefficients or function. Equation (30) can be compared with equation (10), where $J_0 = 0$ is usually assumed by most of the available software packages.

The linear or truncated estimator represented in equation (30) faces two main problems:

1. Optimal choice of M − Values of M that are either too small or too large result in oversmoothing or undersmoothing respectively. The problem is similar to the selection of the bandwidth in kernel smoothing.
2. Truncated estimators will give a poor representation of local singularities (if present) in the underlying function, because an accurate representation of these singularities requires the high-level terms that have been removed by the truncation.

Because of the problem faced by truncation, Donoho and Johnstone [12, 13, 14], proposed a *nonlinear* wavelet estimator based on the selective reconstruction of empirical wavelets coefficients, i.e. only coefficients that meet a given criterion are selected, irrespective of their level. When a signal is corrupted by white noise, the empirical coefficients of its wavelet representation will also be corrupted by white noise. Then the observed wavelet coefficients satisfy

$$\hat{\gamma}_{jk} = \gamma_{jk} + \epsilon_{jk}, \qquad j = 0, \ldots, J-1, \qquad k = 0, \ldots, 2^J - 1, \tag{31}$$

where $\{\epsilon_{jk}\}$ are normal *iid* random variables with zero mean and a given variance σ^2.

It seems reasonable to assume that only large $\hat{\gamma}_{jk}$ coefficients contain information about the underlying function g, while small $\hat{\gamma}_{jk}$ coefficients can be attributed to the noise which uniformly contaminates all the coefficients. Thus, smoothing will be performed by keeping the coefficients with absolute values greater than a given threshold while setting all the other coefficients to zero. The important problem of optimal selection of the threshold λ will be discussed in Section 3.3.

The non-linear parameter estimate can be written:

$$\hat{g}_\lambda(t) = \hat{c}_0\phi(t) + \sum_{j=0}^{J-1}\sum_{k=0}^{2^J-1}\hat{\gamma}^*_{jk}\psi_{jk}(t), \tag{32}$$

where $\{\hat{\gamma}^*_{jk}\}$ are coefficients with absolute value greater than the threshold. Because coefficients on all levels are kept (provided that their absolute value is larger than the threshold), local singularities in the underlying function are well approximated. Image de-noising can be carried out in the same way as for one-dimensional signals.

Wavelet Packet De-noising: This is an extension to the standard de-noising procedure. Since in a wavelet packet analysis, each detail space is decomposed, thresholding of the coefficients can be applied on decomposition. The advantage of this approach is that less of the signal is lost.

3.2 Compression in One and Two Dimensions

Wavelets offer excellent compression, the reason being the sparseness of the wavelet representation of functions, i.e. most of the wavelet coefficients will be nearly zero. Then, in order to have an

economical representation of the original signal, it will be enough to keep only a small number of coefficients together with their location, and assume that all the others are zero.

The aim of compression is to have a representation of the original function that is economical (i.e. requiring a smaller amount of data in the representation than in the original function) and yet still allows the recovery of the original function with acceptable detail. Here, the number of wavelet coefficients stored needs to be minimised while the detail of the function recovered from the coefficients needs to be maximised.

Although the problem has similarities with the smoothing problem reviewed in the previous subsection, the goals of both operations are different. Wavelets can deal with both problems because they provide a sparse representation of a signal. However, in the compression problem, the recovered function is desired to be an accurate (in some sense) approximation of the original function; the task is not to smooth the function and, in fact, the recovered function may be less smooth than the original. Although the problem of threshold selection is important (and will be treated in the next subsection), an equally important aspect is which family of wavelets to choose for the wavelet representation of the function. In general, wavelets with more vanishing moments are better for smooth signals, while wavelets with a small number of vanishing moments are better suited for the compression of signals with local singularities.

It can be shown that wavelets provide a much smaller residual error than the Fourier approximation for the same compression ratio. The compression ratio p (usually expressed as a percentage) is defined as the ratio

$$p = 100m/n, \tag{33}$$

where m is the number of coefficients kept in order to approximate a signal of n data points, while the residual error r (again usually expressed as a percentage) is defined as the ratio of the sum of squared errors and the sum of squares:

$$r = 100 \sum_{i=1}^{n} \left(\hat{g}(t_i) - g(t_i)\right)^2 / \sum_{i=1}^{n} \left(g(t_i)\right)^2 . \tag{34}$$

SURE thresholding is based on the concept of unbiased risk estimate for the threshold and scale dependent (hard) thresholding is applied to the wavelet coefficients. Hybrid thresholding combines SURE and soft thresholding and the two different types of threshold values.

Examples of the performances of different wavelet representations against the Fourier representation for different signals may be found in [28]. All of the previous ideas are valid for both signals (one-dimensional) and images (two-dimensional), and thus wavelets are very useful in practice for compression, with Netscape and the FBI fingerprint database being typical examples. By way of simple illustration, we consider the compression of the image in Figure 21, again using the Daubechies 2 wavelet. Applying a threshold based on the level-1 detail coefficients gives a compressed image (Figure 22) in which 88.97% of the coefficients are zero.

3.3 Thresholding

In the two previous subsections, it has been explained how the sparseness of the wavelet representation can be exploited for smoothing in the context of non-parametric regression (noise removal) and for compression. It was also stated that the basic idea is to keep only the coefficients of the wavelet representation larger than a threshold (wavelet shrinkage), but we have yet to describe how to determine this threshold. That is the subject of this subsection. In all the cases the wavelet shrinkage will produce an estimate $\hat{g}(t_i)$ of the original signal, and the task is to minimise the mean-square accuracy M defined by

$$M = \sum_{i=1}^{n} \left(\hat{g}(t_i) - g(t_i)\right)^2 , \tag{35}$$

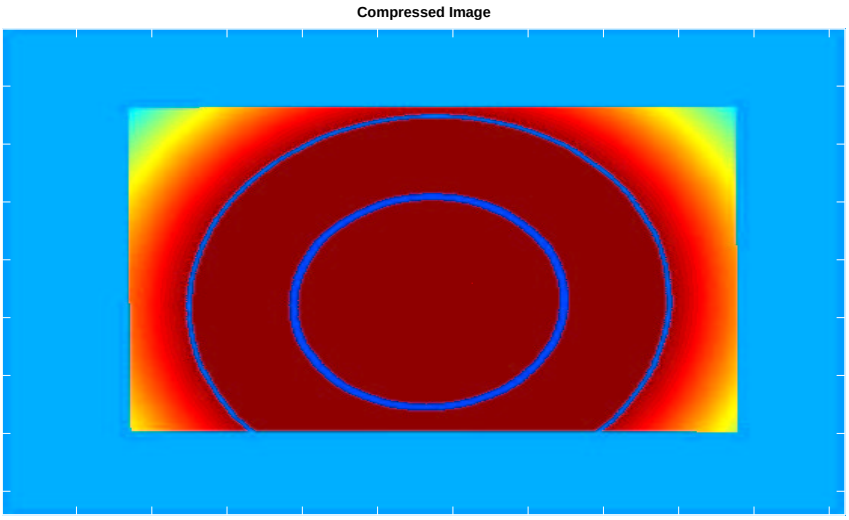

Figure 22: Example of an image compressed using wavelets (compare to the original in Figure 21. In this image 88.97 % of the coefficients have been set to zero.

which by the orthogonality of the wavelet transform can also be expressed as [28]:

$$M = \sum_j \sum_k \left(\hat{\gamma}_{jk} - \gamma_{jk}\right)^2 . \tag{36}$$

In the smoothing problem, what we observe is not $g(t_i)$, which is the unknown underlying signal, but the data corrupted by noise according to the model given by equation (29). In compression, $g(t_i)$ is the signal that we observe and that we want to compress to obtain a more economical representation.

The basic idea of thresholding [12, 13, 14] is to set to zero those wavelet coefficients which have absolute values below a threshold $\lambda > 0$ and keep the remaining coefficients. The values of the coefficients which are retained can either be kept unmodified (hard thresholding):

$$\hat{\gamma}^*_{jk} = \begin{cases} \hat{\gamma}_{jk} & \text{if } |\hat{\gamma}_{jk}| > \lambda \\ 0 & \text{otherwise} \end{cases} , \tag{37}$$

or be shrunk (soft thresholding) [1]:

$$\hat{\gamma}^*_{jk} = \text{sign}\left(\hat{\gamma}_{jk}\right) \max\left(0, |\hat{\gamma}_{jk}| - \lambda\right) . \tag{38}$$

Gao and Bruce [17] suggested a method that marries the best of both types of thresholding. There is a long list of available methods for choosing the threshold. Some of them are:

- **Universal threshold:** (Donoho and Johnstone [12]),
- **Risk shrink:** (Donoho and Johnstone [12]),
- **SURE shrink:** (Donoho and Johnstone [12]),
- **Cross-validation:** (Nason [28]),
- **Hypothesis tests:** (Abramowich and Benjamini [2]),
- **Bayesian methods:** (Chipman, Kolaczyk and McCulloch [5]).

A complete description of the methods can be found in the original references; here we are going to give a brief description of two of the methods most used in practice: the universal threshold and the cross-validation technique.

Donoho and Johnstone [12] originally proposed the universal threshold:

$$\lambda_{\text{uni}} = \sigma\sqrt{2\log n}, \tag{39}$$

where σ^2 is the variance of the noise in the model described by equation (29). This model is often used in practice as, despite its simplicity, it gives good results with both hard and soft thresholding. Donoho and Johnstone [12] showed that using the universal threshold, the wavelet estimator is asymptotically near-minimax in terms of L^2-risk. The standard deviation σ is usually estimated by a robust procedure that uses the absolute deviation from the median (**MAD** statistic), taking into account the coefficients of the high levels of the wavelet transform, since signal will be rare at these levels. The estimator is

$$\hat{\sigma} = \mathbf{MAD}/0.6745,$$

where

$$\mathbf{MAD} = \text{median}|\hat{\gamma}_{jk} - \tilde{\gamma}_{jk}|,$$

with

$$\tilde{\gamma}_{jk} = \text{median}(\hat{\gamma}_{jk}).$$

Although the universal threshold has important asymptotic properties, it depends on the data only through the estimate $\hat{\sigma}$, otherwise ignoring the data, and so cannot be tuned for specific problems. In general, for large samples, the universal threshold tends to bias the estimates by oversmoothing.

To improve the performance of the universal threshold, Donoho and Johnstone [12] suggested that one should always retain coefficients from some coarse scale levels even if they do not exceed the threshold. The level at which the thresholding begins is called the *primary resolution*. The primary resolution could be chosen by cross-validation methods [28].

The cross-validation method [28] splits the data into two halves: odds and evens. One of the sets (say odds) is then used in the estimation of the other set (evens) using a threshold λ, and the error is calculated. The role of the two sets is interchanged and an error is calculated that together with the previous error gives the total error for the threshold λ. The optimum threshold level is the one that minimises the total error.

3.4 Correlated data

In the model for noisy observations represented by equation (29), the noise was modelled as independent and identically distributed (*iid*) random variables. However, in many practical applications, the *iid* assumption is not realistic because the noise has a correlation structure. In the previous subsections, the thresholding techniques were designed to work with white noise; hence an adaptation is necessary to cater for the presence of correlated noise.

For white noise, the same threshold is applied to all the wavelet coefficients or at least to all the coefficients above a primary resolution level. However, when the noise is correlated, the variance of the wavelet coefficients will depend on the level of the wavelet representation, but will remain constant within each level. Thus it seems natural to apply a level-dependent approach, i.e. the threshold is different for each level. A feature that provides intuitive support to the level-dependent approach is that wavelets decorrelate data, i.e. the wavelet coefficients at a given level are in general uncorrelated (the autocorrelation dies away rapidly) and also there is little or no correlation between the wavelet coefficients at different levels. A complete practical and theoretical study of wavelet shrinkage with correlated data can be seen in [22]. For applying

level-dependent thresholding, the noise variance at each level j must be estimated from the data. One possibility is to use the robust estimator ([22]):

$$\hat{\sigma}_j^2 = \mathbf{MAD}\left\{\hat{\gamma}_{jk} | k = 1, \ldots, 2^j\right\} / 0.6745. \tag{40}$$

The threshold methods presented in subsection 3.3 can be adapted for the level-dependent thresholding, such as the universal threshold

$$\lambda_j = \sigma_j \sqrt{2 \log n}. \tag{41}$$

More sophisticated thresholding methods such as Stein unbiased risk estimation (SURE) can be found in [22].

3.5 Examples

Figure 23 shows: (a) an uncorrupted signal, (b) the signal of (a) corrupted with white noise and (c) the signal of (a) corrupted with correlated noise. The variance of the noise is the same in Figure 23(b) and Figure 23(c). However, it is clear that the effect of the correlation in the noise is going to make the underlying signal more difficult to extract from Figure 23(c).

The de-noised signals using the universal threshold can be seen in Figure 24(a) for the signal with white noise, and in Figure 24(b) for the signal with correlated noise. It is easy to see how with white noise, the de-noising operation using a single threshold is quite successful in recovering the original signal (i.e. the wave function in the first half of the signal and the constant value in the second half). However when the noise is correlated, part of the noise is interpreted like signal and the de-noised signal wiggles more than the underlying signal. Signals with correlated noise are best de-noised using multi-level thresholding as can be seen in Figure 24(c). The de-noised signal using the universal threshold applied to each resolution level gives a much better account of the underlying signal.

In Figure 25, we examine some data from a roundness measurement machine. Multi-level soft thresholding with the universal threshold for each level and a primary resolution level of 3 has been applied in order to de-noise the data.

4 Time series analysis

Classically in time series analysis, in order to make a problem statistically treatable, the hypothesis of stationarity is frequently assumed (see, for example, [4]). Loosely speaking, the stationary assumption implies that the statistical properties of the series remain unchanged over time. There are different degrees of stationarity, although second-order stationarity is the most commonly used in applications. Second-order stationarity (or wide sense stationarity) implies constant mean:

$$\mathbf{E}\{X_t\} = \mu, \tag{42}$$

and the autocovariance is a function of the lag between two variables and not of the absolute location of the variables:

$$\mathbf{E}\{(X_t - \mu)(X_{t+\ell} - \mu)\} = C(\ell), \tag{43}$$

where X_t represents a time series with expectation μ, $\mathbf{E}$ denotes the mathematical expectation operator and $C(\ell)$ is the covariance function for lag ℓ.

For the study of a stationary time series, a two-pronged approach has been used that allows the characterisation of the time series in the time (or space) and frequency domain. In the time domain, the covariance function is the main statistical function used while in the frequency domain

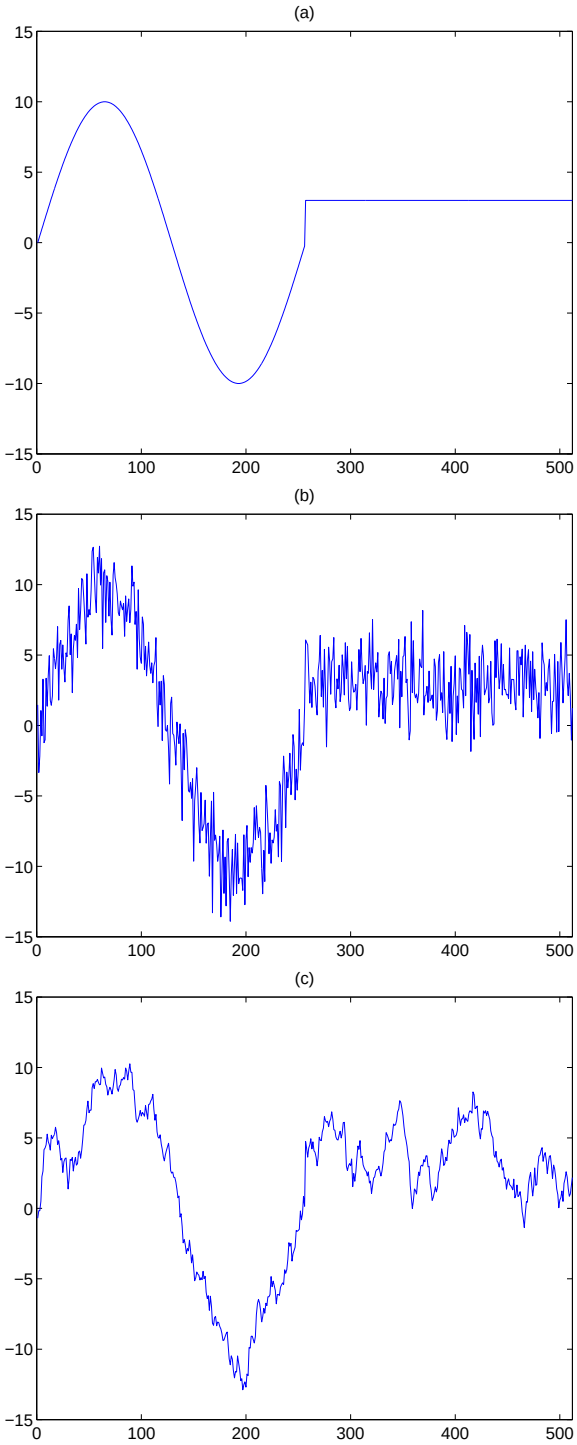

Figure 23: (a) A sine signal; (b) Same signal as (a) but with white noise added; (c) Same signal as (a) but with correlated noise added.

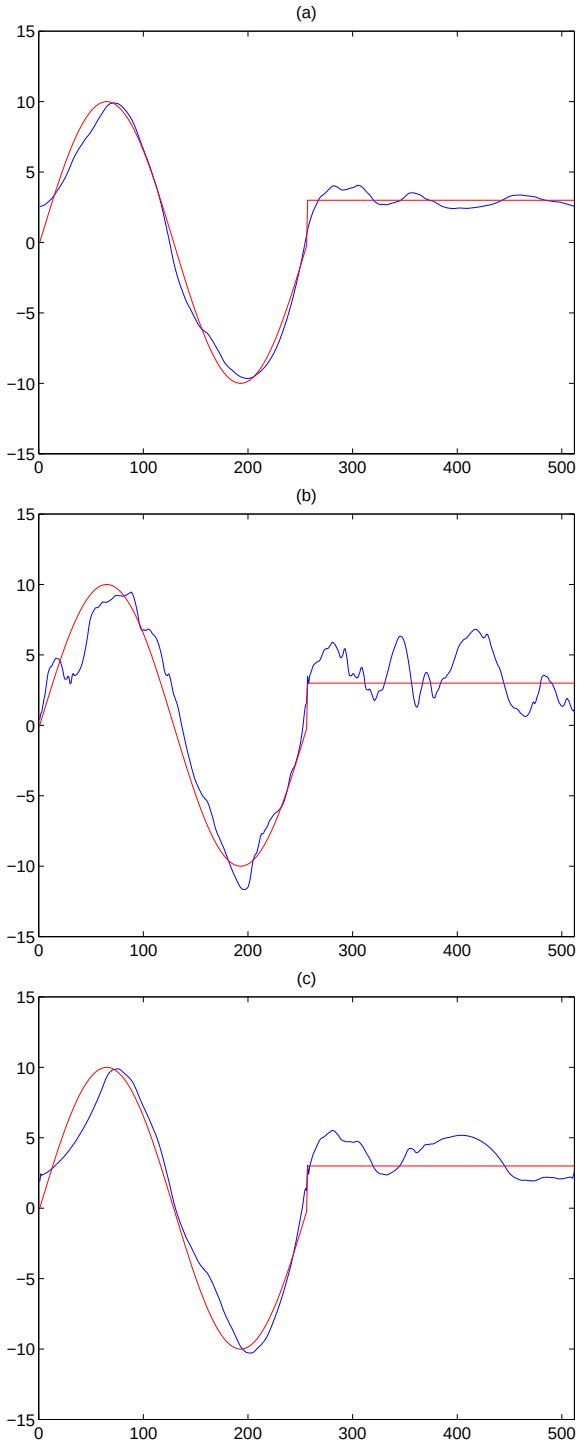

Figure 24: Examples of different de-noising: (a) Signal with white noise de-noised using the universal threshold; (b) Signal with correlated noise de-noised using the universal threshold; (c) Signal with correlated noise de-noised by applying the universal threshold to each level. The soft thresholding technique was applied in each case.

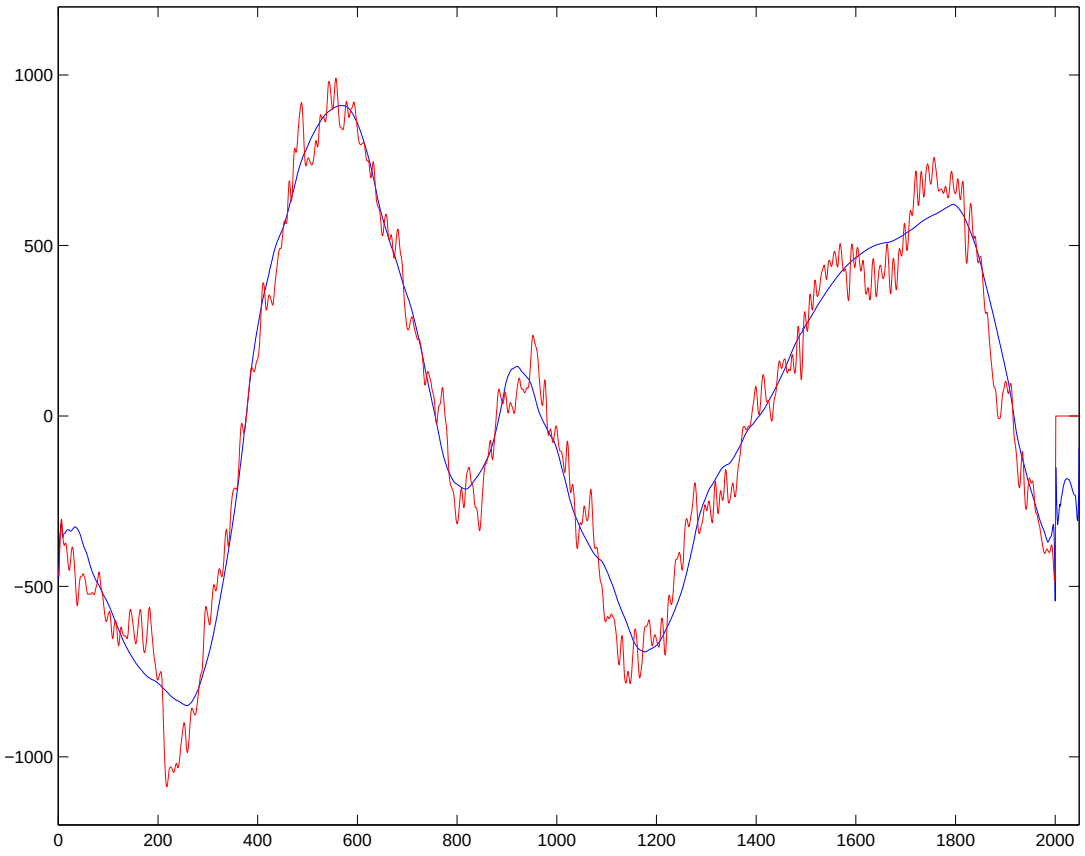

Figure 25: Example of multi-level thresholding on roundness data. Although the "true" signal is not known in this case, it is known that the data has correlated noise.

the power spectrum $f(\omega)$ has been used extensively. The power spectrum can be calculated as the Fourier transform of the covariance function:

$$f(\omega) = \sum_{\ell=-\infty}^{\infty} C(\ell) e^{-i\omega\ell}. \tag{44}$$

The power spectrum represents the distribution of the variance of the stationary time series for the different frequencies. Loosely speaking, it can be said that every stationary time series can be represented as a combination of sine and cosine functions for the different frequencies.

However, many of the time series that appear in practice are non-stationary. A non-stationary time series is typical of a phenomenon that evolves in time. One possibility for studying non-stationary time series is to use wavelets, the main reason being that wavelets reveal information in a signal at a particular time and scale. The wavelet representation of the time series:

$$X_t = \sum_j \sum_k A_{jk} \psi_{jk}(t), \tag{45}$$

where $\{\psi_{jk}\}$ are based on non-decimated wavelets, can be used to analyse if there is any change in the spectrum of the series with time. In this sense, wavelets could be used to test non-stationarity [28].

Other applications of wavelets in a time series analysis context are in the estimation of the power spectrum [16], modelling a time series in terms of another [29], and the study of $1/f$ noise [30].

4.1 Feature Detection

Techniques for signature detection in a signal depend on the form of feature. One of the origins of wavelets is in the analysis of seismic data and wavelets have now been applied in a large number of areas ranging from the detection of detonations in car engines to electrocardiogram data. Once again it is the localization in time and frequency that make wavelets attractive for detecting features. As stated in Section 2.7, the size of the wavelet coefficient γ_{jk} is related to the size of the derivative of the signal around position k. Thus large wavelet coefficients are related to features in the signal. To detect a discontinuity in the mth derivative, it is necessary to select a wavelet with at least m vanishing moments.

Trends in signals are often detected by removing the high frequencies. Consider once again the wavelet decompositions of the signal in Figure 5. In Figures 7 and 8, we have plotted the wavelet coefficients from the continuous and discrete wavelet transforms respectively and in Figure 12, the multiresolution analysis from the non-decimated wavelet transform (NDWT). In all three figures we clearly see large coefficients where the signal is padded to zero and also at "turn-on". Furthermore, from the CWT and NDWT we can identify the underlying undamped oscillation. In examining the figures we have relied on our vision to pick out the contours of large value coefficients. For a more systematic method for detecting features from the time-scale pictures, ridge detection algorithms have been developed; these extract the local maxima in the wavelet decomposition.

Wavelets have also been applied to the robust detection of transient signals [3]. In this work, scale-dependent tests on a statistic based on the energy and estimate of the noise in the wavelet decomposition appear to be the most successful.

In images and 2D signals, often the most important features for pattern recognition are edges. At edge points, the image intensity has sharp transitions, so the modulus of the gradient of the image intensity has a local maximum. The 1D algorithms can be extended to 2D [25], allowing techniques based on wavelet maxima to be used for signal reconstruction and de-noising.

5 Applications

5.1 Applications in Literature

There are papers on the use of wavelets in most areas of mathematics: they can be applied wherever a basis for a function space is required and, because of the practical nature of the fast wavelet transform, they have been applied to a large number of real world modelling and data analysis problems.

Typical wavelet applications include:

- Image processing:
  1. New JPEG standard,
  2. FBI wavelet fingerprint compression standard,
  3. Astronomical image processing,
  4. Brain function imaging,
  5. Texture processing,
  6. Medical imaging.
- General signal processing:
  1. Meteorology,
  2. Astronomy,
  3. Vibration analysis,
  4. Seismic analysis,
  5. Speech processing.
- Computer graphics
- Econometrics
- Fractals
- Communication theory
- Differential equations
- Feature detection
- De-noising
- Template matching [15]
- Materials work: There is already some work in the literature on the use of wavelets in materials problems. See, for example, the work of Addison:

  Wavelet Transforms for Low Strain Integrity Testing of Foundation Piles, Dr. P. S. Addison, Prof. A. Sibbald and J. N. Watson, Funding: EPSRC (Grant GR/M21881).

  Analysis of Concrete Cracking using Fractal and Wavelet Techniques, Dr. P. S. Addison and L. Dougan.
- Acoustic emission: Wavelets are currently being used as part of the AETSAD and AEMS strategic research projects into acoustic emission. Demonstrator projects have been taken from both NPL and industry.

5.2 Acoustics Application at NPL

In [18], a review on the application of wavelets in the acoustics area at NPL is presented. The review includes further examples of de-noising and feature detection in signals. We present details here on one potential use of wavelets to extend the amount of signal that can be used in a conventional analysis. To achieve this, we filter the signal using wavelets to extract the main frequency of interest. In Figure 26, top left, we have plotted a padded sample signal which has mixed frequencies from a switch on and first echo arriving. In Figure 26, top right, we have plotted the wavelet representation of this signal using the Symmlet 6 wavelet. In the rest of the figure we have plotted the effect of crudely setting whole levels of coefficients to zero and the corresponding reconstructed function. This approach may prove useful in extending the range of applicability of current techniques.

5.3 Potential Application Areas in Metrology

Although wavelets are well suited to a general signal, that does not mean they always offer the best representation. Where there is a detailed knowledge of the physics involved in a problem, this can often be applied to good effect. However, even in these cases there are potential roles for wavelets, for example, in de-noising of signals. Furthermore, not all metrology data is well fitted by traditional approaches. Often, the noise structure in a signal is complicated or there exist discontinuities or varying frequencies that are not well represented by traditional techniques. In these situations we expect wavelets will have a great deal to offer. The most effective use of wavelet techniques is to integrate them into existing methodologies.

We expect the following to be areas in metrology where wavelet techniques might prove effective:

- De-noising of metrological data
- Examining the correlation in data
- Feature extraction and detection
- Peak fitting
- Characterization of data into frequency levels
- Template matching

5.4 Areas for Future Investigation

One area critical for wavelet de-noising of signals to be adopted by the metrology community is the estimation of the uncertainty in a denoised sigal, so an investigation into possible approaches is definitely required.

Harmonic wavelets: at present there is no code available to NPL that implements a harmonic wavelet analysis such as described by Newland in [32, 31]. These are complex wavelets and hence contain phase information. In [32], it is claimed that the generalized harmonic wavelets are well adapted to detailed time-frequency analysis.

At present, when fitting data, it is useful to introduce additional constraints on the process that arise from the physics. One method for introducing extra properties on a wavelet basis has already been introduced and that is to use the Lifting Scheme [35]. Other work includes that of [33], in which *a priori* information on signal shape and noise correlation is taken into account. This is a possible area for future activity in CMSC since [33] claims improved signal de-noising and baseline drift minimization.

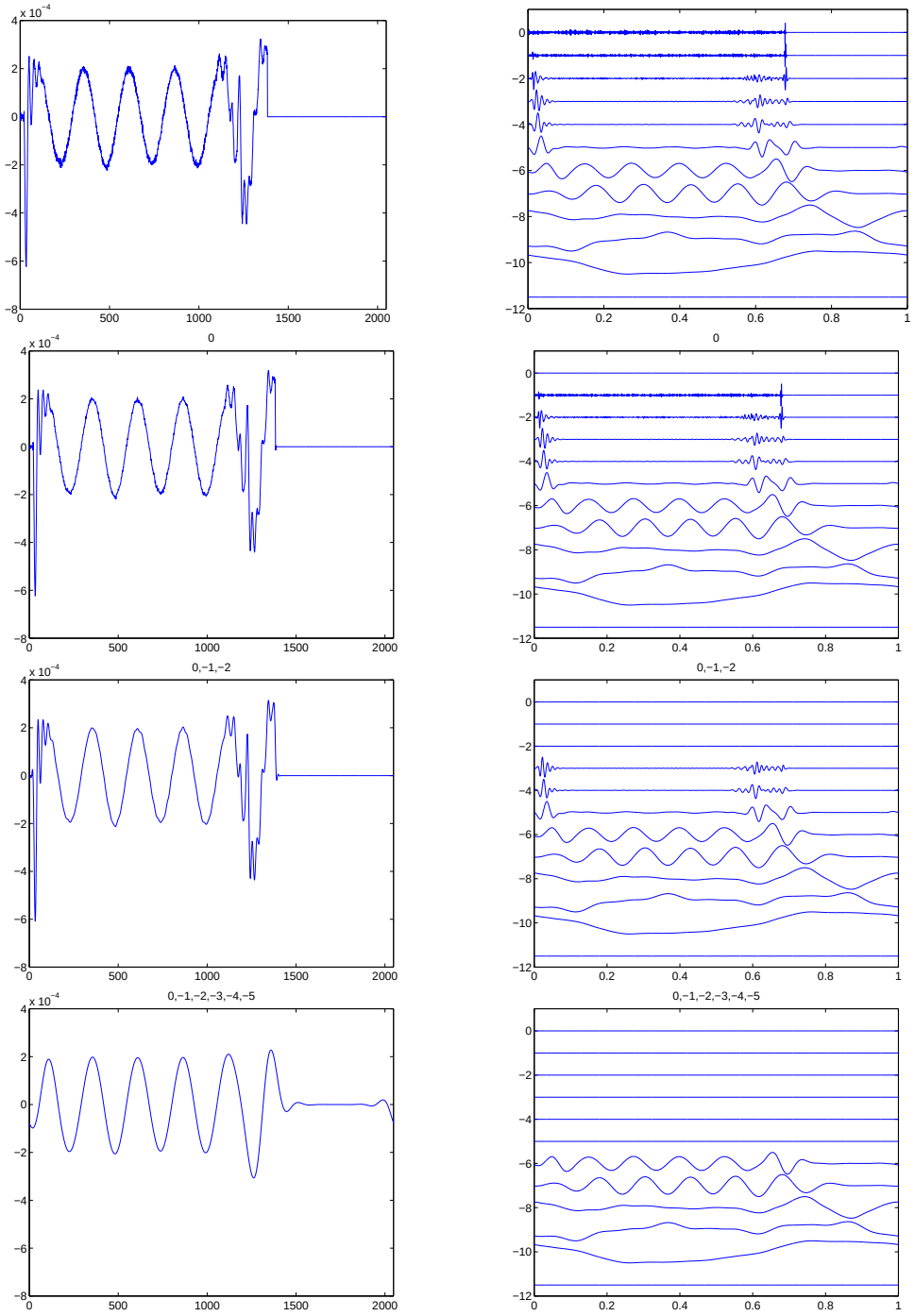

Figure 26: Taking out fine scales (high frequencies) to obtain an extended sine wave.

6 Review of available software

6.1 MATLAB Wavelet Toolbox

[26] (*MATLAB*) This is a fairly comprehensive toolbox that requires the MATLAB Signal Processing Toolbox, making it an expensive option. The graphics are good and using the graphical interface is easy, with good on-line help on wavelets. This toolbox is a new MATLAB facility and there are some glitches in the software. The package does not include the NDWT.

6.2 WaveLab

[11] (*MATLAB*) This is a versatile MATLAB suite with plenty of examples and access to all the source code (distributed under a gnu license). It has a full range of features but requires a greater knowledge of MATLAB to customize. The on-line help facility is not as functional as the **MATLAB Wavelet Toolbox** as it relies on the command line help in MATLAB. WaveLab does, however, have a lot of examples, sample scripts and programs, all with documentation. There is an overview document as well as a comprehensive reference manual of all the programs. The NDWT (or stationary wavelet transform) is available in WaveLab.

6.3 WaveThresh

[27] *S-PLUS* Written by Dr Guy Nason, this is an add-on package for the statistical package S-Plus, and is issued under a gnu licence. The package has the advantage of including the NDWT. However, unfamiliarity with S-Plus makes the package more difficult to use. The source for the routines is available and [28] includes some sample code.

6.4 Dataplore/Santis

[8] Santis was the gnu version of this software which had a limited wavelet capacity (only two wavelet families in the available version). Santis is a general purpose time series analysis package that includes standard Fourier-based techniques and some non-linear time series analysis. It has now been released as a commerical package (Dataplore) and this has not been investigated.

References

[1] F. Abramovich, T. C. Bailey, and T. Sapatinas, *Wavelet analysis and its statistical applications*, J. Roy. Statist. Soc. D, 49 (2000), pp. 1–29.

[2] F. Abramovich and Y. Benjamini, *Thresholding of wavelet coefficients as multiple hypothesis testing procedure*, in Wavelets and Statistics, A. Antoniadis and G. Oppenheim, eds., vol. 103 of Lecture Notes in Statistics, Springer, New York, 1995, pp. 5–14.

[3] R. A. Carmona, *Wavelet detection of transients in noisy time series*, tech. rep., Department of Mathematics, University of California, Irvine, 1991.

[4] C. Chatfield, *The Analysis of Time Series: An Introduction*, Chapman and Hall, fifth ed., 1996.

[5] H. A. Chipman, E. D. Kolaczyk, and R. E. McCulloch, *Adaptive bayesian wavelet shrinkage*, J. Am. Statist. Assoc., 92 (1997), pp. 1413–1421.

[6] R. R. Coifman and M. V. Wickerhauser, *Entropy-based algorithms for best basis selection*, IEEE Trans. on Inf. Theory, 38 (1992), pp. 713–718.

[7] W. Dahmen, *Wavelet and multiscale methods for operator equations*, Acta Numerica, 6 (1997), pp. 55–228.

[8] *Dataplore*. http://www.datan.de/dataplore/.

[9] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA., 1992.

[10] R. A. DeVore and B. J. Lucier, *Wavelets*, Acta Numerica, 1 (1991), pp. 1–56.

[11] D. Donoho, M. Duncan, X. Huo, and O. Levi, *Wavelab*, October 1999. http://www-stat.stanford.edu/∼wavelab/. Version 802.

[12] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaptation by wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.

[13] ——, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assn., 90 (1995), pp. 1200–1224.

[14] ——, *Minimax estimation via wavelet shrinkage*, Ann. Statist., 26 (1998), pp. 879–921.

[15] T. R. Downie, L. Shepstone, and B. W. Silverman, *A wavelet based approach to deformable templates*, in Image Fusion and Shape Variability Techniques, K. V. Mardia, C. A. Gill, and I. L. Dryden, eds., Leeds University Press, April 1996, pp. 163–169.

[16] H.-Y. Gao, *Choice of thresholds for wavelet shrinkage estimate of the spectrum*, J. Time Series Anal., 18 (1997), pp. 231–251.

[17] H.-Y. Gao and A. G. Bruce, *Waveshrink with firm shrinkage*, tech. rep., Research Report 39, Statistical Sciences Division, MathSoft, Inc., 1996.

[18] P. M. Harris, G. J. Lord, S. P. Robinson, and I. M. Smith, *The application of wavelet techniques to the calibration of underwater electroacoustic transducers in reverberant laboratory tanks*, tech. rep., NPL, to appear.

[19] I. Intrator, Q. Q. Huyn, Y. S. Wong, and B. H. K. Lee, *Wavelet feature extraction for discrimination tasks*, Proceedings of the 1997 Canadian Workshop on Information Theory., (1997).

[20] L. Jameson, *A wavelet-optimized, very high order adaptive grid and order numerical method*, SIAM J. Sci. Comput., 19 (1998), pp. 1980–2013.

[21] B. Jawerth and W. Sweldens, *An overview of wavelet based multiresolution analysis*, SIAM Rev., 36 (1994), pp. 377–412. http://cm.bell-labs.com/who/wim/papers/papers.html.

[22] I. M. Johnstone and B. W. Silverman, *Wavelet threshold estimators for data with correlated noise*, J. Roy. Statist. Soc. B, 59 (1997), pp. 319–351.

[23] F. Koster, M. Griebel, N.-R. Kevlahan, M. Farge, and K. Schneider, *Towards an adaptive wavelet-based 3d Navier–Stokes solver*, in Notes on Numerical Fluid Mechanics, E. H. Hirschel, ed., Vieweg, Braunschweig, 1998.

[24] *Mathsoft*. http://www.mathsoft.com/wavelets.html.

[25] S. Mallat and W. L. Hwang, *Singularity detection and processing with wavelets*, IEEE Transactions on Information Theory, 38 (1992), pp. 617–643.

[26] MathWorks, *Matlab wavelet toolbox*. http://www.mathworks.com/products/wavelet/.

[27] G. Nason, *Wavethresh*. http://www.stats.bris.ac.uk/pub/software/wavethresh.

[28] G. Nason, *Everything applied statisticians want to know about wavelets but are afraid to ask*. Lecture Course Notes, March 1999. Dept Mathematics, University of Bristol.

[29] G. P. Nason, T. Sapatinas, and A. Sawczenko, *Wavelet packet modelling of nonstationary time series*, tech. rep., University of Bristol, 1999.

[30] G. P. Nason and R. von Sachs, *Wavelets in time series analysis*, Phil. Trans. R. Soc. Lond. A, 357 (1999), pp. 2511–2526.

[31] D. E. Newland, *Harmonic wavelets in vibrations and acoustics*, in Wavelets: The Key to Intermittent Information, The Royal Society, 1999.

[32] ——, *Ridge and phase identification in the frequency analysis of transient signals by harmonic wavelets*, Journal of Vibration and Acoustics, 121 (1999), pp. 1–7.

[33] L. V. Novikov, *Adaptive wavelet analysis of signals*, tech. rep., Institute for Analytical Instrumentation, RAS, 1999.

[34] R. L. Schult and H. W. Wyld, *Using wavelets to solve the Burgers equation: A comparative study*, Phys. Rev. A, 46 (1992), pp. 7953–7958.

[35] W. Sweldens, *The lifting scheme: A construction of second generation wavelets*, SIAM J. Math. Anal., 29 (1998), pp. 511–546.

[36] O. V. Vasilyev, D. A. Yuen, and S. Paolucci, *The solution of pdes using wavelets*, Computers in Physics, 11 (1997), pp. 429–435. http://landau.mae.missouri.edu/∼vasilyev/Publications/.

[37] *Wavelet digest*. http://www.wavelet.org.

[38] M. V. Wickerhauser, *Acoustic signal compression with wavelet packets*, in Wavelets: A tutorial in Theory and Applications, C. K. Chui, ed., Academic Press, San Diego, CA., 1992, pp. 679–700.