The Stationary Wavelet Transform and some Statistical Applications

G.P. Nason and B.W. Silverman[1]

Department of Mathematics, University of Bristol, Bristol BS8 1TW, UK

Abstract

Wavelets are of wide potential use in statistical contexts. The basics of the discrete wavelet transform are reviewed using a filter notation that is useful subsequently in the paper. A 'stationary wavelet transform', where the coefficient sequences are not decimated at each stage, is described. Two different approaches to the construction of an inverse of the stationary wavelet transform are set out. The application of the stationary wavelet transform as an exploratory statistical method is discussed, together with its potential use in nonparametric regression. A method of local spectral density estimation is developed. This involves extensions to the wavelet context of standard time series ideas such as the periodogram and spectrum. The technique is illustrated by its application to data sets from astronomy and veterinary anatomy.

1 Introduction

In this paper we discuss some aspects of wavelets with a particular view to their statistical application. In particular we shall be concerned with an extension of the standard discrete wavelet transform, which we term the *stationary wavelet transform*. The stationary wavelet transform itself is discussed by Pesquet, Krim and Carfantan [PKC] from more of an 'engineering' point of view, and we do not make any claims of originality for it. We do, however, feel that the potential value of the technique for statistical purposes will be enhanced by a 'tutorial' introduction both to the standard wavelet transform and to the stationary wavelet transform, and we provide this in Sections 2 and 3 below, in a way that is particularly angled towards the uses of the techniques that we will discuss subsequently. The basic idea of the stationary wavelet transform is to 'fill in the gaps' caused by the decimation step in the standard wavelet transform. This leads to an over-determined, or redundant, representation of the original data, but one which has considerable statistical potential, which is explored in the remainder of the paper. We hope that our development, based on a particular 'filter' notation, will be of value in understanding the various transforms.

Two main examples, one from astronomy and one from veterinary science, are presented in Section 4, and the value of the stationary wavelet transform for exploring these data sets is discussed. The potential uses of the stationary wavelet transform in regression and curve estimation are then considered in Section 5. In order to do this, we need to set out ways of inverting the stationary wavelet transform, again building on work of Pesquet *et al.* [PKC]. The stationary wavelet transform has a valuable role in the exploration and spectral analysis of non-stationary time series. In Section 6, we give a heuristic discussion of the way in which wavelets can be used to extend to non-stationary data the notions of periodogram and spectrum that arise from the Fourier analysis of stationary series. This discussion is illustrated by the application of the methods proposed to the two practical examples introduced in Section 4.

2 The standard wavelet transform

2.1 The discrete wavelet transform

Suppose we are given a sequence $c_0, c_1, \ldots, c_{N-1}$ where $N = 2^J$ for some integer J. For the moment we shall concentrate entirely on the algorithmic aspects of the wavelet expansion, without making any explicit or tacit assumptions about the sequence $\{c_i\}$.

The standard discrete wavelet transform is based on filters $\mathcal{H}$ and $\mathcal{G}$ and on a 'binary decimation' operator $\mathcal{D}_0$. The filter $\mathcal{H}$ is a low pass filter, defined by a sequence conventionally denoted $\{h_n\}$. Typically, only a small number of the $\{h_n\}$ are non-zero. The action of the low pass filter on a doubly infinite sequence $\{\ldots, x_{-1}, x_0, x_1, x_2, \ldots\}$ is defined by

$$(\mathcal{H}x)_k = \sum_n h_{n-k} x_n. \tag{1}$$

The definitions for sequences of finite length depend on a choice of treatment at the boundaries. For the present paper, periodic boundary conditions will be used, but others are possible; see, for example, Nason and Silverman [NS1].

The filter is assumed to satisfy the internal orthogonality relation

$$\sum_n h_n h_{n+2j} = 0 \tag{2}$$

for all integers $j \neq 0$, and to have sum of squares $\sum h_n^2 = 1$.

The high-pass filter $\mathcal{G}$ is defined by the sequence

$$g_n = (-1)^n h_{1-n} \tag{3}$$

for all n. It is easy to show that the filter $\mathcal{G}$ satisfies the same internal orthogonality relations as $\mathcal{H}$ and in addition that the filters obey the mutual orthogonality relation

$$\sum_n h_n g_{n+2j} = 0 \tag{4}$$

for *all* integers j.

Filters constructed in the way we have described are called *quadrature mirror filters.* For further details of suitable constructions of particular quadrature mirror filters, see Vaidyanathan [Va1] or Daubechies [Da1].

The binary decimation operator $\mathcal{D}_0$ simply chooses every even member of a sequence, so that

$$(\mathcal{D}_0 x)_j = x_{2j} \tag{5}$$

for all integers j. It follows from the internal and mutual orthogonality properties of the quadrature mirror filters that the mapping of a sequence x to the pair of sequences $(\mathcal{D}_0\mathcal{G}x, \mathcal{D}_0\mathcal{H}x)$ is an orthogonal transformation. If x is a finite sequence of length 2^m with periodic boundary conditions applied, then each of $\mathcal{D}_0\mathcal{G}x$ and $\mathcal{D}_0\mathcal{H}x$ will be sequences of length 2^{m-1}.

The discrete wavelet transform is derived from a multiresolution analysis, performed as follows. For clarity of exposition, we shall impose periodic boundary conditions throughout. Define the *smooth at level J*, written c^J to be the original data

$$c^J_n = c_n \text{ for } n = 0, 1, \ldots, 2^J - 1.$$

Now, for $j = J-1, J-2, \ldots, 0$, recursively define the *smooth c^j at level j* and the *detail d^j at level j* by

$$c^j = \mathcal{D}_0\mathcal{H}c^{j+1} \text{ and } d^j = \mathcal{D}_0\mathcal{G}c^{j+1}. \tag{6}$$

Note that c^j and d^j will both be sequences of length 2^j. It can be seen from (6) that the smooth at each level is fed down to the next level to give rise to the smooth and detail at that level.

Because the mapping $(\mathcal{D}_0\mathcal{H}, \mathcal{D}_0\mathcal{G})$ is an orthogonal transform it is very easily inverted to find c^{j+1} in terms of c^j and d^j; write the transform as a matrix and take the transpose. If the original filter sequence $\{h_n\}$ has H non-zero members, then the matrix will have (at most) $2^{j+1}H$ non-zero elements. We shall write $\mathcal{R}_0$ for the inverse transform so that

$$c^{j+1} = \mathcal{R}_0(c^j, d^j) \text{ for each } j\ . \tag{7}$$

The usual discrete wavelet transform is obtained by continuing this process to obtain the detail at each level together with the smooth at the zero level, so that the original sequence is orthogonally transformed to the sequence of sequences $d^{J-1}, d^{J-2}, \ldots, d^0, c^0$, of total length 2^J. It is immediate from (7) that the process can be reversed by reconstructing c^1 from d^0 and c^0, then reconstructing c^2 from c^1 and d^1, and so on. Provided the filter sequence $\{h_n\}$ has a finite number of non-zero elements, the step at level j in both the transform and its inverse will require $O(2^j)$ operations; summing over j it follows that the overall number of arithmetic operations for both the transform and its inverse is $O(2^J)$. For further details on the discrete wavelet transform see, for example, Mallat [Ma1] and Smith and Barnwell [SB1].

2.2 Some simple modifications

Since the discrete wavelet transform is an orthogonal transform it corresponds to a particular choice of basis for the space $\Re^N$ in which the original sequence lies. There are several possible modifications of the DWT each corresponding to a somewhat different choice of basis.

We first of all note that it is not necessary to continue the decomposition as far as level 0. An equally valid orthogonal transformation of the data is given by stopping the process at any level R, so that we map the original sequence to the sequence of sequences $d^{J-1}, d^{J-2}, \ldots, d^R, c^R$. We shall call this the discrete wavelet transform *curtailed at level R*.

Secondly it is immediate that the decomposition could equally well be carried out by selecting every odd member of each sequence rather than every even member. To be precise, define $\mathcal{D}_1$ to be the operator defined for sequences x by

$$(\mathcal{D}_1 x)_j = x_{2j+1} \tag{8}$$

for all integers j. The mapping $(\mathcal{D}_1\mathcal{H}, \mathcal{D}_1\mathcal{G})$ is still an orthogonal transformation, and the multiresolution analysis can be carried out by successively applying this operator in (6) instead of $(\mathcal{D}_0\mathcal{H}, \mathcal{D}_0\mathcal{G})$. The results will not be the same, but the overall transformation will still be an orthogonal transformation. The reconstruction can be obtained by successive application of the corresponding inverse operator, denoted $\mathcal{R}_1$.

Indeed it is not even necessary for the same choice of 'odd' or 'even' to be used throughout. Suppose that $\epsilon_{J-1}, \epsilon_{J-2} \ldots, \epsilon_0$ is a sequence of 0's and 1's. We can then use the operator $\mathcal{D}_{\epsilon_j}$ at level j, and perform the reconstruction by using the corresponding sequence of operators $\mathcal{R}_{\epsilon_j}$. For each choice of the sequence ϵ this will give a different orthogonal transformation of the original sequence. We shall refer to this transformation as *the ϵ-decimated discrete wavelet transform.*

Some elucidation of this procedure can be obtained by considering shift operators. Let $\mathcal{S}$ be the shift operator defined by

$$(\mathcal{S}x)_j = x_{j+1}.$$

If x is a finite sequence, define the shift periodically at the boundary. It then immediate from the definitions that $\mathcal{D}_1 = \mathcal{D}_0\mathcal{S}$ and hence that $\mathcal{R}_1 = \mathcal{S}^{-1}\mathcal{R}_0$. It is also easy to see that $\mathcal{S}\mathcal{D}_0 = \mathcal{D}_0\mathcal{S}^2$ and that the operator $\mathcal{S}$ commutes with $\mathcal{H}$ and $\mathcal{G}$.

Now define S to be the integer whose binary representation is $\epsilon_0\epsilon_1 \ldots \epsilon_{J-1}$. It can easily be shown that the coefficient sequences c^j and d^j yielded by the ϵ-decimated discrete wavelet transform are all shifted versions of those yielded by the ordinary DWT applied to the *shifted* sequence $\mathcal{S}^S x$.

To see this, consider any fixed j and let s_1 and s_2 be the integers with binary representations $\epsilon_0\epsilon_1 \ldots \epsilon_{j-1}$ and $\epsilon_j\epsilon_{j+1} \ldots \epsilon_{J-1}$. In the standard DWT the sequence $d^j = \mathcal{D}_0\mathcal{G}(\mathcal{D}_0\mathcal{H})^{J-j-1}c^J$. In the ϵ-decimated case, we have

$$\begin{aligned} d^j &= \mathcal{D}_{\epsilon_j}\mathcal{G}\mathcal{D}_{\epsilon_{j+1}}\mathcal{H}\mathcal{D}_{\epsilon_{j+2}}\mathcal{H}\ldots\mathcal{D}_{\epsilon_{J-1}}\mathcal{H}c^J \\ &= \mathcal{D}_0\mathcal{S}^{\epsilon_j}\mathcal{G}\mathcal{D}_0\mathcal{S}^{\epsilon_{j+1}}\mathcal{H}\mathcal{D}_0\mathcal{S}^{\epsilon_{j+2}}\mathcal{H}\ldots\mathcal{D}_0\mathcal{S}^{\epsilon_{J-1}}\mathcal{H}c^J \qquad (9) \\ &= \mathcal{D}_0\mathcal{G}(\mathcal{D}_0\mathcal{H})^{J-j-1}\mathcal{S}^{s_2}c^J. \qquad (10) \end{aligned}$$

To obtain (10) from (9) we have repeatedly commuted the shift operators with $\mathcal{G}$ and $\mathcal{H}$, and used the property that $\mathcal{S}\mathcal{D}_0 = \mathcal{D}_0\mathcal{S}^2$. Now apply the operator $\mathcal{S}^{s_1}$. We have

$$\mathcal{S}^{s_1}d^j = \mathcal{S}^{s_1}\mathcal{D}_0\mathcal{G}(\mathcal{D}_0\mathcal{H})^{J-j-1}\mathcal{S}^{s_2}c^J = \mathcal{D}_0\mathcal{G}(\mathcal{D}_0\mathcal{H})^{J-j-1}\mathcal{S}^S c^J,$$

since $S = 2^{J-j}s_1 + s_2$. Thus d^j shifted by an amount s_1 is the jth detail sequence of the standard DWT applied to the original data shifted by an amount S. The corresponding result for c^j is derived exactly similarly, replacing $\mathcal{D}_0\mathcal{G}(\mathcal{D}_0\mathcal{H})^{J-j-1}$ by $(\mathcal{D}_0\mathcal{H})^{J-j}$ throughout.

It follows that the basis vectors of the ϵ-decimated DWT can be obtained from those of the standard DWT by applying the shift operator $\mathcal{S}^S$, and the choice of ϵ thus corresponds to a choice of 'origin' with respect to which the basis functions are defined.

2.3 The continuous wavelet transform

Some elucidation both of the standard DWT and of the modifications described in Section 2.2 can be obtained by relating the discrete wavelet transform of a sequence to the continuous wavelet decomposition of a corresponding function.

Wavelets are based on functions called *scaling functions* ϕ which have two key properties. Firstly, $\phi(t)$ and all its integer translates $\phi(t+j)$ form an orthonormal set in L^2, so that $\int \phi(t)^2 dt = 1$ and $\int \phi(t)\phi(t+j)dt = 0$ for integers $j \neq 0$. Secondly, ϕ can be expressed as a linear combination of half-integer translates of itself at double the scale:

$$\phi(t) = \sqrt{2}\sum_j h_j \phi(2t-j). \tag{11}$$

Defining the sequence $\{g_j\}$ as the 'mirror' of the sequence $\{h_j\}$ by the relation (3), the *mother wavelet* ψ is then defined by

$$\psi(t) = \sqrt{2}\sum_j g_j \phi(2t-j). \tag{12}$$

The key properties of the scaling function imply that the sequence $\{h_j\}$ satisfies the properties required of the corresponding sequence in the discrete wavelet transform.

Bases of various function spaces can now be constructed from appropriate dilations and translations of ϕ and ψ. For any integer j, define

$$\phi_j(t) = 2^{j/2}\phi(2^j t) \text{ and } \psi_j(t) = 2^{j/2}\psi(2^j t).$$

Now, associate with a sequence c^J the function

$$f(t) = \sum_k c_k^J \phi_J(t - 2^{-J}k). \tag{13}$$

For any level $R < J$, perform a discrete wavelet transform curtailed at level R. It follows from the properties of scaling functions and the definition of the DWT that the resulting coefficients give the expansion for f in orthonormal functions

$$f(t) = \sum_k c_k^R \phi_R(t - 2^{-R}k) \; + \; \sum_{j=R}^{J-1} d_k^j \psi_j(t - 2^{-j}k).$$

Because of the orthonormality properties of the wavelets, it is also true that

$$d_k^j = \int \psi_j(t - 2^{-j}k) f(t) dt$$

so that the detail coefficient d_k^j gives information about f at 'scale' 2^{-j} near position $t = 2^{-j}k$. In terms of the original sequence c^J this corresponds to scale 2^{J-j} near position $2^{J-j}k$. The coefficients c^R of the smooth at level R give an approximation to the original function in terms of orthogonal translates of ϕ_R.

Now consider the ϵ-decimated discrete wavelet transform. Defining S as before, let $t_0 = 2^{-J}S$. Then the coefficient sequences obtained will give an expansion of f in terms of $\phi_R(t - t_0 - 2^{-R}l)$ and $\psi_j(t - t_0 - 2^{-j}l)$ for integers l and for $j = R, R+1, \ldots, J-1$. In terms of the original sequence, this means that, at scale 2^{J-j}, the positions at which the detail coefficients will be localised are of the form $2^{J-j}l + S$, a grid of integers of gauge 2^{J-j} shifted to have origin S.

3 The stationary wavelet transform

In this section we explain how the basic DWT algorithm can be modified to give a *stationary wavelet transform* that no longer depends on the choice of origin.

3.1 Definition

The basic idea is extremely simple. We simply apply appropriate high and low pass filters to the data at each level to produce two sequences at the next level. We do not decimate, and the two new sequences each have the same length as the original sequence. Instead, we modify the filters at each level, by padding them out with zeroes, in a way that we now define.

Let $\mathcal{Z}$ be the operator that alternates a given sequence with zeroes, so that, for all integers j, $(\mathcal{Z}x)_{2j} = x_j$ and $(\mathcal{Z}x)_{2j+1} = 0$. Define filters $\mathcal{H}^{[r]}$ and $\mathcal{G}^{[r]}$ to have weights $\mathcal{Z}^r h$ and $\mathcal{Z}^r g$ respectively. Thus the filter $\mathcal{H}^{[r]}$ has weights $h^{[r]}_{2^r j} = h_j$ and $h^{[r]}_k = 0$ if k is not a multiple of 2^r. The filter $\mathcal{H}^{[r]}$ is obtained by inserting a zero between every adjacent pair of elements of the filter $\mathcal{H}^{[r-1]}$, and similarly for $\mathcal{G}^{[r]}$. It is immediate that $\mathcal{H}^{[r]}$ and $\mathcal{G}^{[r]}$ commute with $\mathcal{S}$ and that

$$\mathcal{D}_0^r \mathcal{H}^{[r]} = \mathcal{H}\mathcal{D}_0^r \text{ and } \mathcal{D}_0^r \mathcal{G}^{[r]} = \mathcal{G}\mathcal{D}_0^r. \tag{14}$$

To define the stationary wavelet transform, we start by setting a^J to be the original sequence c^J in our previous notation. For $j = J, J-1, \dots, 1$ we then recursively define

$$a^{j-1} = \mathcal{H}^{[J-j]} a^j \text{ and } b^{j-1} = \mathcal{G}^{[J-j]} a^j. \tag{15}$$

If the vector a^J is of length 2^J, then all the vectors a^j and b^j will be of the same length, rather than becoming shorter as j decreases as in the standard DWT. Thus to find $b^{J-1}, b^{J-2}, \dots, b^0, a^0$ will take $O(J2^J)$ operations rather than the $O(2^J)$ of the standard DWT.

3.2 Relation with the decimated DWT

It is easy to see that the stationary wavelet transform contains the coefficients of the ϵ-decimated DWT for *every* choice of ϵ. For any given ϵ and corresponding 'origin' S, the detail at level j are a shifted version of $\mathcal{D}_0^{J-j}\mathcal{S}^S b^j$ and the data at level j the same shifted version of $\mathcal{D}_0^{J-j}\mathcal{S}^S a^j$.

To be precise, define s_1 and s_2 as in Section 2.2 above. Let $d^j(\epsilon)$ be the j detail sequence obtained from the ϵ-decimated discrete wavelet transform. We then have, for each j,

$$\begin{aligned}
\mathcal{S}^{-s_1}\mathcal{D}_0^{J-j}\mathcal{S}^S b^j &= \mathcal{D}_0^{J-j}\mathcal{S}^{s_2} b^j = \mathcal{D}_0^{J-j}\mathcal{S}^{s_2}\mathcal{G}^{[J-j-1]}\mathcal{H}^{[J-j-2]}\dots\mathcal{H}^{[0]}c^J \\
&= \mathcal{D}_0^{J-j}\mathcal{G}^{[J-j-1]}\mathcal{H}^{[J-j-2]}\dots\mathcal{H}^{[0]}\mathcal{S}^{s_2}c^J \\
&= \mathcal{D}_0\mathcal{G}\mathcal{D}_0^{J-j-1}\mathcal{H}^{[J-j-1]}\dots\mathcal{H}\mathcal{S}^{s_2}c^J \\
&= \mathcal{D}_0\mathcal{G}\mathcal{D}_0\mathcal{H}\mathcal{D}_0^{J-j-2}\dots\mathcal{H}\mathcal{S}^{s_2}c^J = \dots \\
&= \mathcal{D}_0\mathcal{G}(\mathcal{D}_0\mathcal{H})^{J-j-1}\mathcal{S}^{s_2}c^J = d^j,
\end{aligned}$$

from (10) above, with the corresponding calculation linking a^j and c^j.

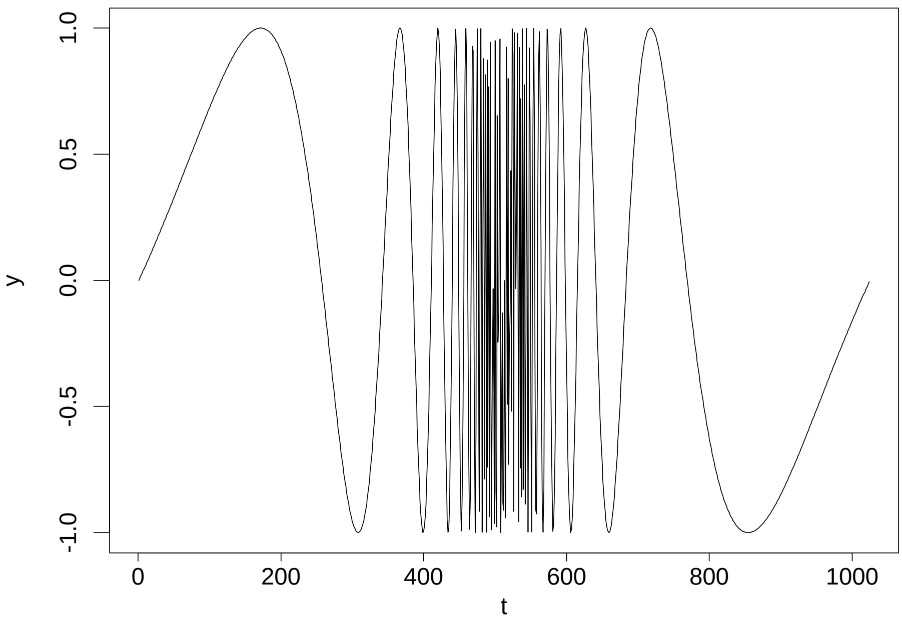

Figure 1: A simulated chirp signal.

The relation with the ϵ-decimated DWT allows a clear interpretation of the stationary wavelet transform coefficients. Associate the data sequence c^J with the function f as in (13) above. For any j and k, by considering a decimated DWT with $S = k$ and hence $t_0 = 2^{-J}k$, we can see at once that

$$b_k^j = \int \psi_j(t - 2^{-J}k)f(t)dt \qquad (16)$$

so that, in terms of the original sequence, b_k^j gives information at scale 2^{J-j} localised at position k. There is no longer any restriction of the localisation position to a grid of integers. The stationary wavelet transform 'fills in the gaps' between the coefficients in any particular ϵ-decimated DWT.

4 Using the wavelet transform as an exploratory method

In this section we consider two examples which illustrate the use of the stationary wavelet transform for the exploration of observed signals. The examples will be considered again in the context of local spectral estimation later in the paper.

The first example is a 'chirp' signal, which is of interest in many scientific fields, for example, in gravity wave detection in astronomy and signal recognition problems in ecology. Our particular concern arose out of discussions with David Nicholson of the Department of Astronomy, University College, Cardiff. The signal is shown in Figure 1. The basic feature is that the frequency increases up to the singularity in the middle of the graph, and then decreases again. The standard wavelet transform of this signal (Figure 2) does not make this property very clear. On the other hand, in the stationary wavelet transform (Figure 3) the increasing and decreasing frequncy can be seen. As t

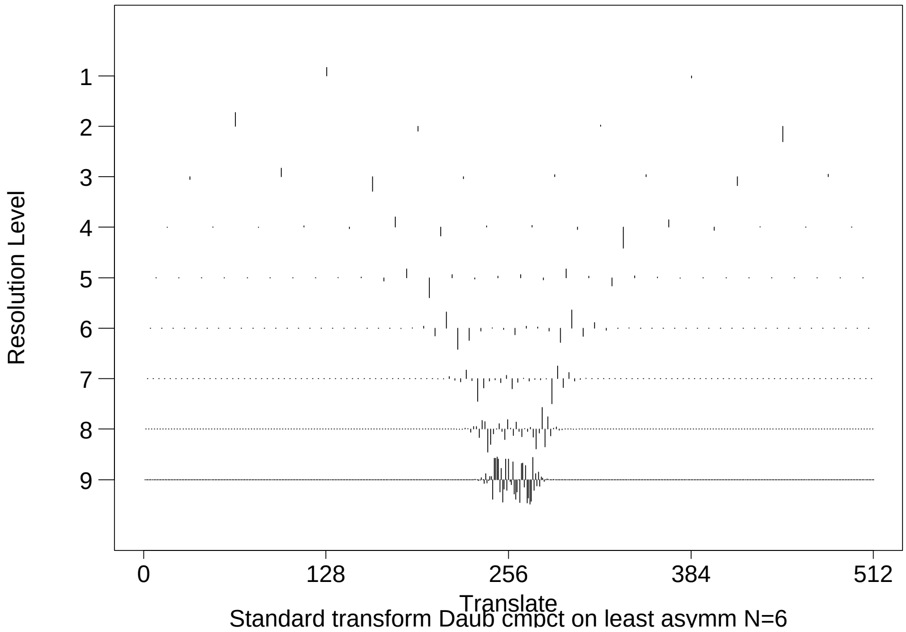

Figure 2: Discrete wavelet transform of the chirp signal.

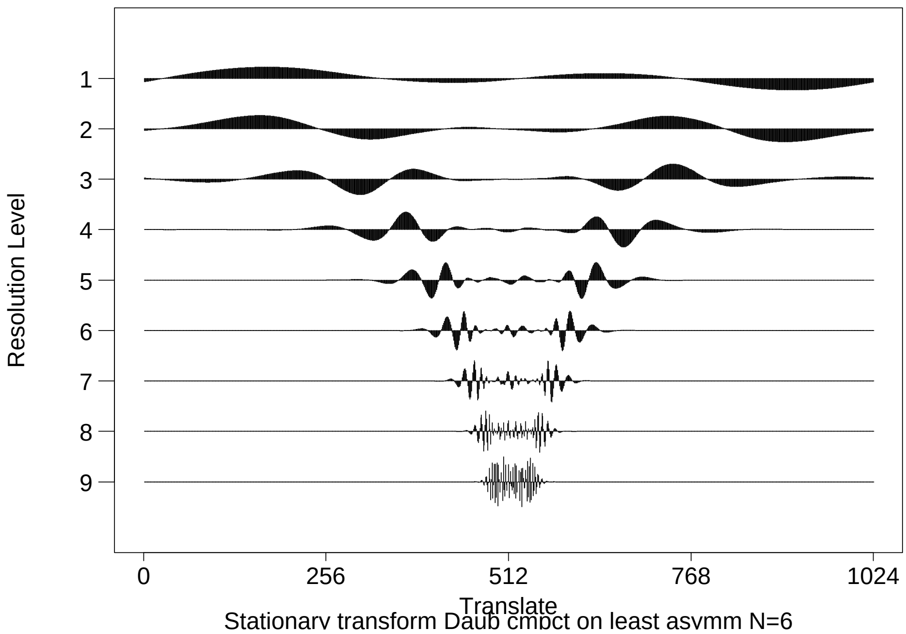

Figure 3: Stationary wavelet transform of the chirp signal.

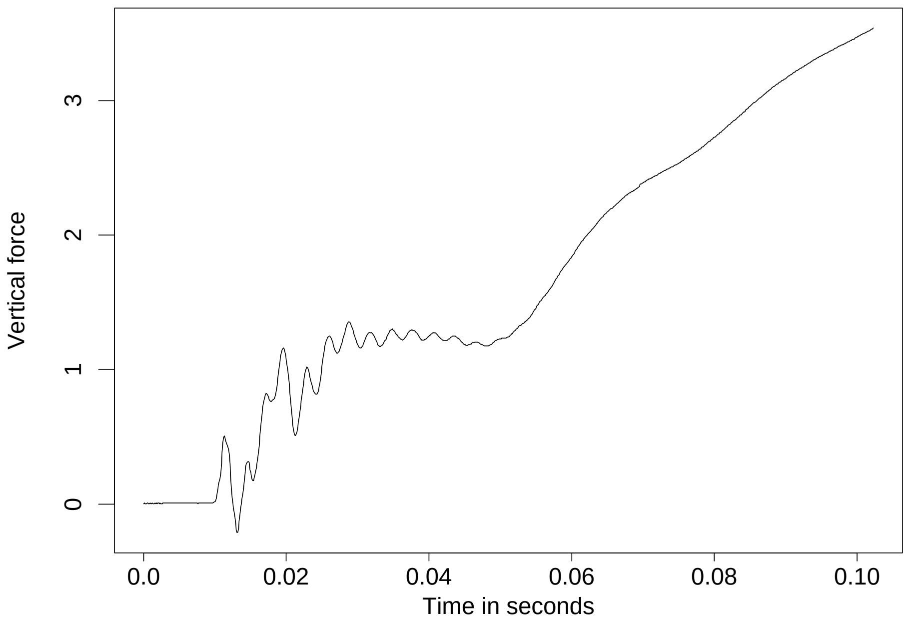

Figure 4: Vertical force pattern for part of a stride of a trotting horse.

increases, the amplitude of the oscillation within any particular frequency level increases and then dies off, and the region of high amplitude becomes closer to the singularity as the frequency band increases. In Section 6 below, we will pursue this example somewhat further in the context of providing a 'local spectral density estimate' for a nonstationary time series.

Another example comes from the study of equine gait. This data was collected at the Bristol Equine Sports Medicine Centre by Dr Alan Wilson and colleagues of the Department of Anatomy, Bristol University. The particular data we discuss concerns the study of the initial impact spike when the hoof of a trotting horse hits the ground. A typical plot of the vertical force is given in Figure 4. It can be seen that, after the initial impact, the force pattern falls into two phases. At first there is a phase in which the force increases and then reaches a plateau; this is followed by a second phase in which the full weight of the horse is put down on the leg. It is of considerable interest to characterise the way in which the initial impact energy, represented by the high frequency oscillations, is absorbed by the horse's leg. An important aspect for future scientific investigation is the contrast between these patterns for horses landing on different surfaces.

The main interest is in understanding the frequency behaviour of the signal in the period immediately after the initial impact. In order to do this, an overall trend curve is fitted to the data (starting at the time of the initial impact) and the residual curve about this trend determined, as shown in Figure 5. The discrete and stationary wavelet transforms of this curve are shown in Figure 6. At the frequency bands at which there is any substantial detail at all, the discrete wavelet transform has a sampling rate that is essentially too low to give any clear picture of the data. It is better to sacrifice the

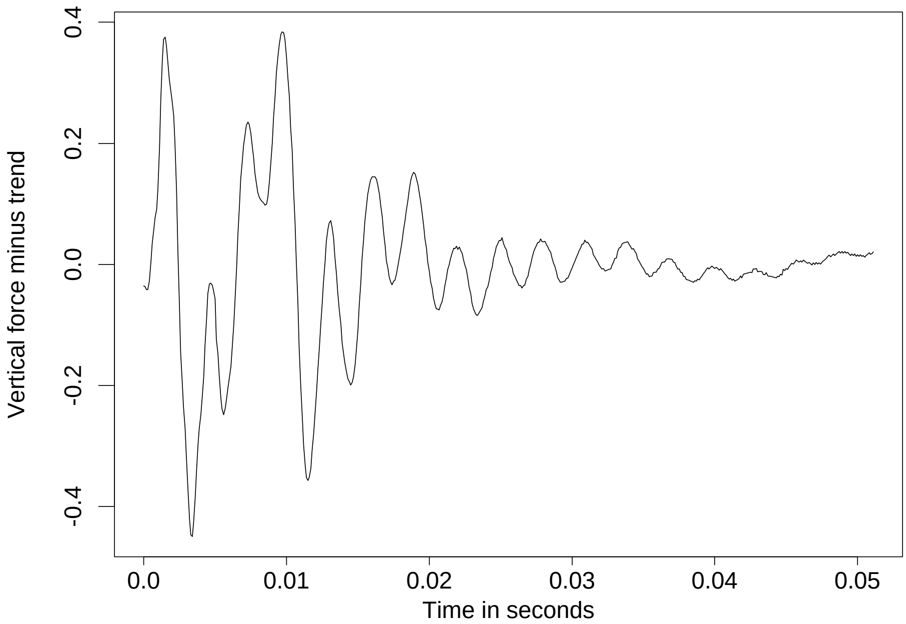

Figure 5: Vertical force after initial impact with overall.

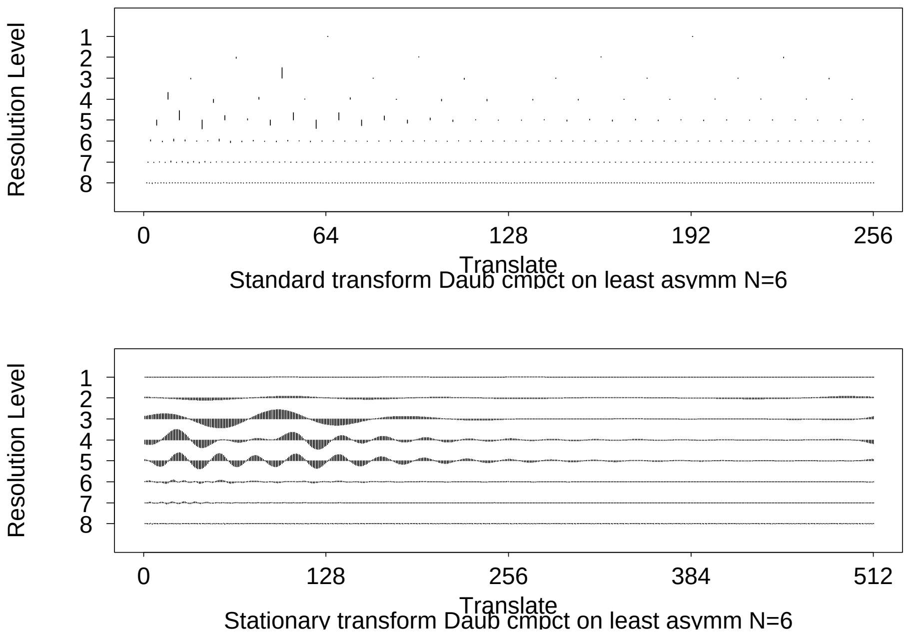

Figure 6: Discrete and stationary wavelet transforms of the equine vertical force curve shown in Figure 5

orthogonality of the successive filters in order to attain the higher sampling rate of the stationary wavelet transform. Again, we shall consider this example again below.

The main applied interest of this paper will be in the use of the stationary wavelet transform for exploratory purposes and for local spectral estimation. However, some comments about the possible use of the method for regression will also be made.

5 Inverses, overdetermined representations and nonparametric regression

5.1 Inverting the stationary wavelet transform

Consider, now the reconstruction of a sequence from its stationary wavelet transform. As we have seen in Section 3.2 the stationary wavelet transform is an overdetermined representation of the original sequence and contains its coefficients relative to many different bases. Therefore the inverse operator will be far from unique, and it will be convenient to consider different approaches to constructing an inverse. Of course if the various inverses are applied to an exact SWT they will give the same result, but later in the paper we shall be considering their application to *estimates* of the SWT of a function of interest. The estimates of the SWT may not themselves be SWTs of anything, and so the choice of exactly which inverse to use will itself be part of the estimation process.

The first approach is to choose some sequence ϵ of 0's and 1's, to select from the SWT coefficients those corresponding to the ϵ-decimated DWT, and then to invert that transform to perform the reconstruction. We call this the *ϵ-basis inverse* of the SWT.

The second approach, the *average basis inverse*, corresponds to finding the ϵ-basis inverse for every ϵ and to averaging the result. For a series of length 2^J, there are 2^J possible values of ϵ and any particular ϵ-basis inverse would take $O(2^J)$ operations to compute. However, the average basis inverse can be calculated far more quickly than the $O(2^{2J})$ operations that a naive approach would suggest, as we now explain.

For any particular r, the operator $(\mathcal{H}^{[r]}, \mathcal{G}^{[r]})$ that maps a^{J-r} to (a^{J-r-1}, b^{J-r-1}) is not an orthogonal transform. However, it follows at once from the basic properties of the filters $\mathcal{H}$ and $\mathcal{G}$ that the decimated transforms $(\mathcal{D}_0\mathcal{H}^{[r]}, \mathcal{D}_0\mathcal{G}^{[r]})$ and $(\mathcal{D}_1\mathcal{H}^{[r]}, \mathcal{D}_1\mathcal{G}^{[r]})$ are each orthogonal. The first of these yields the even members of the sequences a^{J-r-1} and b^{J-r-1} and the second the odd. Denote the inverses of the two transformations by $\mathcal{R}_0^{[r]}$ and $\mathcal{R}_1^{[r]}$ respectively. Because of the relation $\mathcal{D}_1 = \mathcal{D}_0\mathcal{S}$, and the fact that $\mathcal{S}$ commutes with $\mathcal{H}^{[r]}$ and $\mathcal{G}^{[r]}$, it follows that $\mathcal{R}_1^{[r]} = \mathcal{S}^{-1}\mathcal{R}_0^{[r]}$. The operators $\mathcal{R}_0^{[r]}$ and $\mathcal{R}_1^{[r]}$ each require $O(2^J)$ operations for sequences of length 2^J. It also follows easily from the relation (14) that

$$\mathcal{D}_0^r\mathcal{R}_0^{[r]} = \mathcal{R}_0\mathcal{D}_0^r. \tag{17}$$

Define the operator

$$\bar{\mathcal{R}}^{[r]} = \frac{1}{2}(\mathcal{R}_0^{[r]} + \mathcal{R}_1^{[r]}),$$

so that $\bar{\mathcal{R}}^{[r]}$ requires $O(2^J)$ operations, and reconstructs a^{J-r} in turn from the odd and even members of the sequences a^{J-r-1} and b^{J-r-1}) and averages the results. The original data can be obtained by recursively evaluating

$$a^j = \bar{\mathcal{R}}^{[J-j]}(a^{j-1}, b^{j-1}) \text{ for } j = 1, 2, \ldots, J. \tag{18}$$

This recursion requires $O(J2^J)$ operations, and it is equivalent to the average basis inverse of the stationary wavelet transform. A proof, using the properties of the various operators involved, is left as an exercise for the reader.

5.2 Potential application to nonparametric regression

Hitherto the main statistical use of wavelets has been in nonparametric regression. Suppose that, for some $N = 2^J$, we make observations $Y_1, \ldots, Y_N$ assumed to satisfy

$$Y_i = f(i/N) + \epsilon_i$$

for some function f, where the ϵ_i are independent $N(0, \sigma^2)$ random variables. The classic wavelet paradigm for estimating f is to find the discrete wavelet transform of $\{Y_i\}$, apply a soft or hard thresholding rule, and then to invert to yield an estimate of f. See Donoho and Johnstone [DJ2, DJ1], Donoho, Johnstone, Kerkyacharian and Picard [DJKP]. Nason [Na2, Na3], Abramovich and Benjamini [AB1], Fan, Hall, Martin and Patil [FHMP], Johnstone and Silverman [JS], Neumann and Spokoiny [NS2], Ogden [Og1], Vidakovic [Vi1], Wang [Wa1] and Weyrich and Warhola [WW1].

In this paper we only mention regression in passing, but note that the two inverse methods set out in the above section yield two contrasting approaches to the nonparametric regression problem.

Once the stationary wavelet transform of the data has been worked out, an approach related to the 'best basis' approach for wavelet packets can be used. The choice of whether to choose the 'odd' or 'even' coefficients at each level to pass down to the next level of the transform can be considered as a binary tree. Each possible choice of ϵ can be considered as a different path to the bottom of the tree from its root. A suitable entropy function can be used to choose between possible bases. If this entropy is additive, then a dynamic programming idea can be used to find the path that corresponds to the basis with the lowest entropy; see Pesquet *et al.* [PKC] for further details. The basic idea is exactly the same as that used by Coifman and Wickerhauser [CW1] for wavelet packets. An additional possibility is that of curtailing the decomposition at any particular level R, as discussed in Section 2.2 above. This involves adding an additional possibility at each node of the tree, that of curtailing the decomposition at this point; this third choice is then a terminal node. The additional burden in the dynamic programming algorithm is trivial.

The other approach is to use the average basis inverse of the stationary wavelet transform. Take the SWT of the data, and apply a hard or soft thresholding rule to the individual coefficients. Each element of the resulting array will be an estimate of the corresponding element of the SWT of the function f. In order to obtain an estimate of f, apply the average basis inverse of the SWT. This procedure is, of course, equivalent to performing a standard wavelet regression on each ϵ-decimated wavelet transform, and then averaging over ϵ. Some preliminary work (joint with Donoho and Coifman) suggests that the averaging approach retains the adaptivity of the standard wavelet approach but improves its performance on smooth parts of the function of interest.

6 Local spectral density estimation

Suppose $X_1, X_2, \ldots, X_N$ is an observed time series with $N = 2^J$. In this section we shall discuss a possible approach to the local spectral analysis of the series. Another approach, also using wavelets, is discussed by Sachs and Schneider [vSS]. The use of wavelets in the estimation of the spectral density of a stationary series is considered by Gao [Ga1, Ga2].

6.1 The wavelet periodogram and the wavelet spectrum

For each fixed j, consider the sequence b^j_k as defined in (16). As was pointed out in Section 3.2 above, this sequence provides information about the original data at scale 2^{J-j}. A way of seeing this more clearly is to recall that the mother wavelet ψ is band-limited in the frequency domain, and hence, from equation (16), we can see that b^j_k is a band-limited filtering of the original sequence; as j varies, the frequency response function of the filter is dilated by a factor of 2^j, and so, roughly speaking, the frequencies allowed through by the band-limited filter are in an interval whose endpoints are proportional to 2^j.

In the standard analysis of stationary time series, the periodogram is defined to be the squared modulus of the discrete Fourier transform of the observed series. By analogy, we define the array $\{(b^j_k)^2\}$ to be the *wavelet periodogram* of the series.

If the series is defined by a stochastic mechanism with $E(X_i) = 0$ and $\Gamma_{ij} = \mathrm{cov}(X_i, X_j)$, then the *stationary wavelet spectrum* $\{\eta_{jk}\}$ is defined by

$$\eta_{jk} = E(b^j_k)^2.$$

Since the coefficients b^j_k depend linearly on the observed series, it follows immediately that, if the series is Gaussian, we have

$$(b^j_k)^2 \sim \eta_{jk}\chi^2_1$$

and hence, just as in classical time series analysis, the wavelet periodogram will not be a consistent estimator of the wavelet spectrum.

Each element of the wavelet spectrum is of course a quadratic form dependent on the variance matrix Γ. Some brief remarks about this dependence are made in Section 6.3 below, and a more detailed discussion will be provided in subsequent work.

Some heuristic insight into the behaviour of the wavelet spectrum and periodogram may be obtained by considering the case where the series consists of a pure sinusoidal signal at a particular frequency. For simplicity of exposition, associate with the discrete series a continuous function $Y(t)$ by the construction (13). Neglect the high frequency effects introduced into Y by this construction and assume that for some constants a and ω,

$$Y(t) = A\sin(2^J\omega t),$$

where ω is the frequency of variation of the original sequence. Assume that the mother wavelet ψ has Fourier transform $\tilde{\psi}(s)$, so that the function ψ_j has Fourier transform $2^{-j/2}\tilde{\psi}(2^{-j}s)$. Assume that, for $s > 0$, $\tilde{\psi}(s)$ is negligible outside a range (s_L, s_H). For the Littlewood–Paley wavelet, defined by

$$\tilde{\psi}(s) = \begin{cases} (2\pi)^{-1/2} & \text{for } \pi \leq |s| \leq 2\pi \\ 0 & \text{otherwise} \end{cases}$$

we will have $s_H = 2s_L$ and for other wavelets s_H/s_L will be somewhat larger but not usually dramatically so.

The stationary wavelet transform b^j at level j is simply the convolution of the functions ψ_j and Y, sampled at frequency 2^J. By standard Fourier series arguments, therefore, b^j will be a sine wave of frequency ω and amplitude A_j proportional to $2^{-j/2}A|\tilde{\psi}(2^{J-j}\omega)| = Aw_j(\omega)$, say. It can now easy to see that A_j will be negligible if $2^{J-j}\omega$ is outside the range (s_L, s_H), in other words if $J-j$ is outside the range $(\log_2(\omega)-\log_2(s_H), \log_2(\omega)-\log_2(s_L))$. If $s_H = 2s_L$ then only a single A_j will be non-zero; more generally the signal will be spread over a small number of levels of the stationary wavelet transform.

Squaring the stationary wavelet transform will yield, at level j, a sequence of the form $\frac{1}{2}Aw_j(\omega)\{1 + \sin(2\omega k + \tau_j)$ where τ_j is a possible phase shift. Thus the stationary wavelet periodogram at level j will consist of the combination of a 'signal' $Aw_j(\omega)$ with the addition of a 'noise component' oscillating at frequency 2ω.

Now let us consider a more general signal with slowly varying amplitude $A(t)$. Since the various wavelet coefficients depend only locally on the series, the amplitudes of the individual levels of the stationary wavelet transform can be expected, roughly speaking, to vary in proportion to $A(t)$.

If we now consider the case where the signal consists of a mixture of a number of pure frequency components each with varying amplitude, we can expect that the total 'signal' in the wavelet periodogram at level j will be $\frac{1}{2}\sum_\omega w_j(\omega)A_\omega(t)$ with a 'noise component' that will oscillate at frequencies double those of the frequencies for which $w_j(\omega)$ is non-zero, ie the range $(2^{j-J}s_L, 2^{j-J}s_H)$. Thus in smoothing the wavelet periodogram to obtain an estimate of the wavelet spectrum, we aim at level j to remove frequencies above $2^{j+1-J}s_L$, by a suitable filtering approach. Rather unusually for smoothing problems, we have a reasonably objective criterion for deciding which part of the 'data' (in this case the wavelet periodogram) is 'signal' and which is 'noise'. In this paper we have performed the smoothing by a non-adaptive wavelet approach. Each level j of the wavelet periodogram is considered separately and its standard discrete wavelet transform found. All the coefficients at levels $j+1$ and above are set to zero and the inverse found. The effect is, approximately, to apply a low-pass filter removing frequencies above $2^{j+1-J}s_L$, as required. Refinements of this rather crude procedure are the subject of current investigation.

The heuristic remarks made in this section will of course be the subject of further theoretical investigation. In the present paper we shall illustrate and investigate them by reference to some practical examples.

6.2 Examples

The first example is the chirp data discussed in Section 4. The wavelet periodogram for this data is given in Figure 7. It can immediately be seen that the oscillations at each level occur in the manner predicted by our discussion above. As the frequency of the signal increases, the power in the wavelet periodogram moves to the higher frequency levels, and similarly in the second half of the plot the power moves back down again. The smoothing method described above can be applied to yield the local spectral estimate given in Figure 8. The spurious high frequency effects within each level have now been removed, and the overall picture of a signal with increasing and then decreasing frequency is both clear and quantifiable.

Turning to the equine gait example, the local spectral estimate obtained by our approach is given in Figure 9. At frequency band 3, there is a single episode of oscillation.

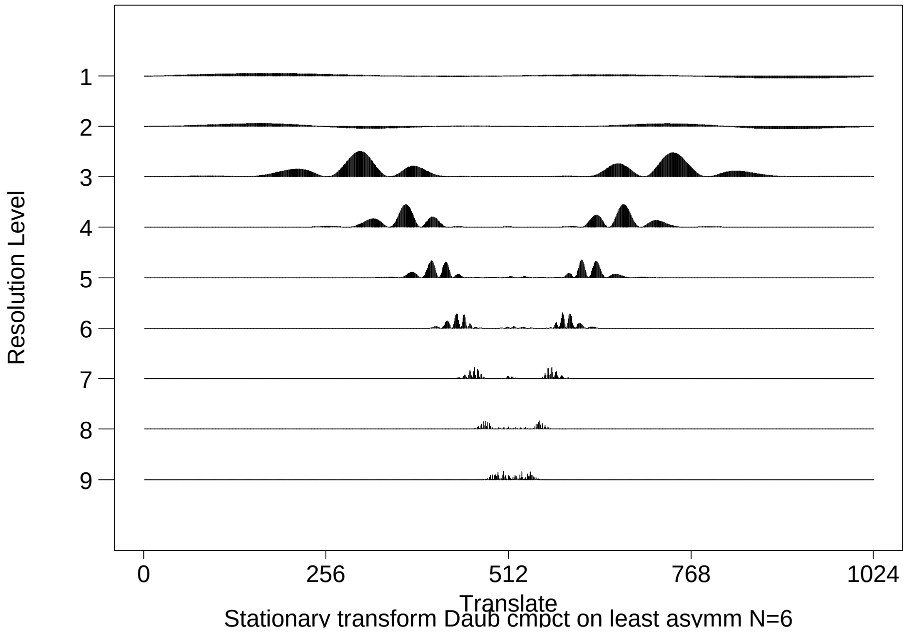

Figure 7: Wavelet periodogram for the simulated chirp.

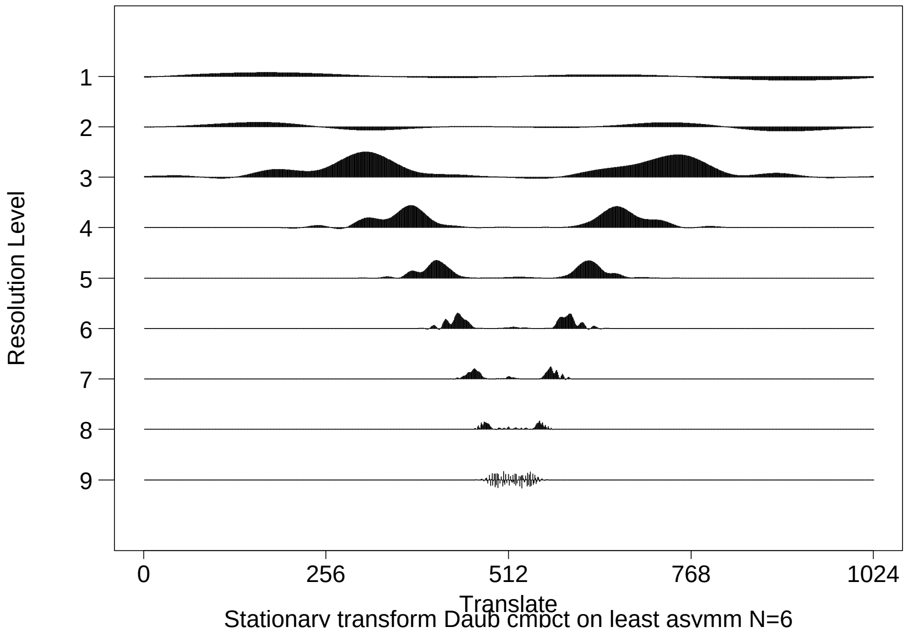

Figure 8: Local spectral estimate for the simulated chirp.

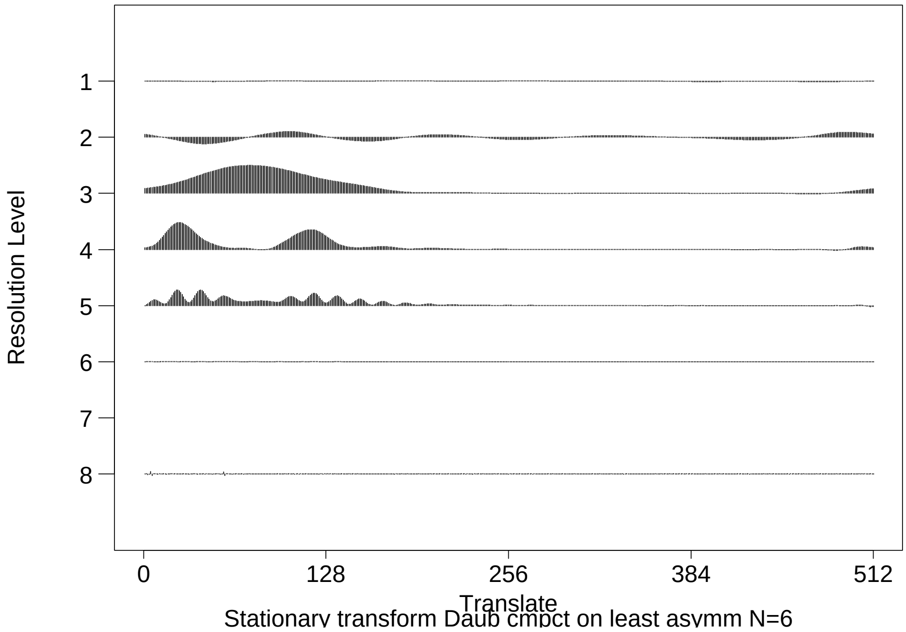

Figure 9: Local spectral estimate for equine force plate data.

Comparison with the original data as shown in Figure 4 demonstrates that this episode lasts about as long as the first phase of the stride. By contrast, frequency band 4 contains two bursts of oscillation, one at the initial impact and one subsequently at the end of the first phase. Our discussion of the appropriate amount of smoothing to apply to the wavelet periodogram indicates that this double bump is not spurious. This interesting feature of the data has important implications for the biomechanics of the horse's leg, and further data collection and analysis is in progress to investigate it further.

6.3 Some theoretical remarks

In this section we make some brief theoretical remarks, relating our approach to some concepts in the analysis of nonstationary time series.

Consider, for the moment, a continuous time random process $X(t)$ with covariance function $\Gamma(s,t)$. It is then natural to consider $\gamma_t(\tau) = \Gamma(t-\tau/2, t+\tau/2)$ as being the covariance of the process at lag τ at time t. In this sense $\gamma_t(\tau)$ is the *local covariance* of the series. If the stochastic structure of the process is in fact stationary, then $\gamma_t(\tau)$ does not depend on τ and is simply the covariance of the process at lag τ. If the process is observed at discrete time points, then $\gamma_t(\tau)$ is of course only defined for certain values of t and τ, but the general principle remains.

Having defined a local covariance, we can now define the local spectrum in an obvious way. The spectrum of a stationary process is defined to be the Fourier transform of the autocovariance function; so it is natural to define the local spectrum to be the Fourier transform of the local covariance:

$$\psi_t(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \gamma_t(\tau) e^{-i\omega\tau} d\tau.$$

The function $\psi_t(\omega)$ is called the *Wigner distribution* of the process; see Section 11.1 of Priestley [Pr1]. The wavelet spectrum can then be defined in terms of the Wigner distribution. The element η_{jk} is a linear functional of the form

$$\eta_{jk} = \int \int W_j(t - 2^{-J}k, \omega)\psi_t(\omega)dtd\omega$$

for some suitable weight function W_j that can be expressed in terms of the particular wavelet being used. We are at present investigating the properties of the local spectral estimate by considering the weight functions W_j in more detail.

7 Conclusions and acknowledgements

The use of wavelets for statistical purposes is still in its infancy, and it will be some years before their genuine practical advantages and disadvantages are understood properly. In particular the statistical ideas presented in this paper are clearly in need of further development but the results so far are extremely promising.

We would like to thank the organizers of the "XV^{e} Recontres Franco-Belges de Statisticiens" for financial support and for organizing a most enjoyable and intellectually stimulating meeting. The support of EPSRC in carrying out this research is gratefully acknowledged.

References

[AB1] Abramovich, F., Benjamini, Y.: Adaptive thresholding of wavelet coefficients. (submitted for publication), (1994).

[BH1] Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Statist. Soc. B. **57** (1995), 289–300.

[Bu1] Burman, P.: A comparative study of ordinary cross-validation, ν-fold cross-validation and the repeated learning-testing methods. Biometrika. **76** (1989), 503–514.

[CDJV] Cohen, A., Daubechies, I., Jawerth, B., Vial, P.: Multiresolution analysis, wavelets, and fast algorithms on an interval. Compt. Rend. Acad. Sci. Paris A. **316** (1993), 417–421.

[CW1] Coifman, R.R., Wickerhauser, M.V.: Entropy-based algorithms for best basis selection. IEEE Transactions on Information Theory. **38** (1992), 713–718.

[Da1] Daubechies, I.: Orthonormal bases of compactly supported wavelets. Comms Pure Appl. Math.. **41** (1988), 909–996.

[Da2] Daubechies, I.: *Ten lectures on Wavelets*. SIAM, (1992).

[DJ1] Donoho, D.L., Johnstone, I.M.: Adapting to unknown smoothness via wavelet shrinkage. J. Am. Statist. Ass., (to appear).

[DJ2] Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika, **81** (1994), 425–455.

[DJKP] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., and Picard, D.: Wavelet shrinkage: asymptopia? (with discussion). J. R. Statist. Soc. B. **57** (1995), to appear.

[Do1] Donoho, D.L.: Nonlinear solution of linear-inverse problems by wavelet-vaguelette decomposition. Technical Report 403, Department of Statistics, Stanford University, Stanford, (1992).

[FHMP] Fan, J., Hall, P., Martin, M., Patil, P.: Adaption to high spatial inhomogeneity based on wavelets and on local linear smoothing. Technical Report CMA-SR18-93, Centre for Mathematics and Its Applications, Australian National University, Canberra, (1993).

[Ga1] Gao, H.-Y.: Wavelet estimation of spectral densities in time series analysis. PhD thesis, University of California, Berkeley. (1993).

[Ga2] Gao, H.-Y.: Choice of thresholds for wavelet estimation of the log spectrum. Technical Report number 438, Department of Statistics, Stanford University. (1993).

[JS] Johnstone, I.M., Silverman, B.W.: Wavelet threshold estimators for data with correlated noise. (submitted for publication).

[KT] Kwong, M.K., Tang, P.T.P.: W-matrices, nonorthogonal multiresolution analysis, and finite signals of arbitrary length. Technical Report MCS-P449-0794, Argonne National Laboratory, (1994).

[Ma1] Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattn Anal. Mach. Intell. **11** (1989), 674–693.

[Me1] Meyer, Y.: *Wavelets and Operators.* Cambridge University Press, Cambridge, (1992).

[Na1] Nason, G.P.: The `WaveThresh` package; wavelet transform and thresholding software for S. Available from the StatLib archive, (1993).

[Na2] Nason, G.P.: Wavelet function estimation using cross-validation. (submitted for publication), (1994).

[Na3] Nason, G.P.: Wavelet regression by cross-validation. Technical Report 447, Department of Statistics, Stanford University, Stanford, (1994).

[NS1] Nason, G.P., Silverman, B.W.: The discrete wavelet transform in S. Journal of Computational and Graphical Statistics, **3** (1994), 163–191.

[NS2] Neumann, M.H., Spokoiny, V.G.: On the efficiency of wavelet estimators under arbitrary error distributions. The IMS Bulletin, **23** (1994) 218.

[Og1] Ogden, R.T.: Wavelet thresholding in nonparametric regression with change point applications. PhD thesis, Texas A&M University, (1994).

[PKC] Pesquet, J.C., Krim, H., Carfantan, H.: Time invariant orthonormal wavelet representations. (1994). (submitted for publication)

[PTVF] Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: *Numerical Recipes in C, the Art of Scientific Computing.* Cambridge University Press, Cambridge, (1992).

[Pr1] Priestley, M.B.: *Spectral Analysis and Time Series.* Academic Press, London, (1981).

[vSS] von Sachs, R., Schneider, K.: Wavelet smoothing of evolutionary spectra by non-linear thresholding. Technical report, University of Kaiserslautern.

[SB1] Smith, M.J.T., Barnwell, T.P.: Exact reconstruction techniques for tree-structured subband coders. IEEE Transactions on Acoustics, Speech and Signal Processing. **34** (1986), 434–441.

[St1] Stein, C.: Estimation of the mean of a multivariate normal distribution. Ann. Statist., **9** (1981), 1135–1151.

[St1] Stone, M.: Cross-validatory choice and assessment of statistical predictions (with discussion). J. R. Statist. Soc. B, **36** (1974), 111–147.

[TM1] Taswell, C., McGill, K.C.: Wavelet transform algorithms for finite duration discrete-time signals. Technical Report Numerical Analysis Project Manuscript NA-91-07, Department of Computer Science, Stanford University, (1991).

[Va1] Vaidyanathan, P.P.: Multirate digital filters, filter banks, polyphase networks, and applications: a tutorial. Proceedings of the IEEE, **78** (1990), 56–93.

[Vi1] Vidakovic, B. Nonlinear wavelet shrinkage with Bayes rules and Bayes factors. (submitted for publication), (1994).

[Wa1] Wang, Y.: Function estimation via wavelets for data with long-range dependence. Technical Report, Univeristy of Missouri, Columbia, (1994).

[WW1] Weyrich, N., Warhola, G.T.: De-noising using wavelets and cross-validation. Technical Report AFIT/EN/TR/94-01, Department of Mathematics and Statistics, Air Force Institute of Technology, AFIT/ENC, 2950 P ST, Wright-Patterson Air Force Base, Ohio, 45433-7765, (1994).