Smoothing Chaotic Signals for System Identification : A Multiresolution Wavelet Decompositon Approach

D. Coca, S. A. Billings

Depatment of Automatic Control & Systems Engineering

University of Sheffield

Mappin Street

Sheffield, S1 3JD

United Kingdom

June 1995

Research Report No. 587

Abstract

A new wavelet based smoothing algorithm is introduced to reduce the noise affecting chaotic signals prior to system identification. The algorithm involves a multiresolution decomposition of the signal using B-spline wavelets which makes use of the mutual information between neighbouring points in time. Rigurous results concerning the effects of noise on the wavelet coefficients and the efficiency of the smoothing procedure are derived. Noise corrupted signals generated from Chua's circuit and the Lorenz equation are used to test the new method. The correlation dimension and the largest Lyapunov exponent of the models, estimated using the smoothed signals, are compared with the true values to demonstrate that the new algorithm preserves the dynamical properties of the underlying system.

1. Introduction

Many real life systems can, under certain conditions, exhibit a very irregular behaviour known as chaos. Chaotic behaviour has for example been detected in systems ranging from electronic circuits, electric power stations and lasers to chemical reactors, economic growth and heart activity. Initially chaotic motion was regarded as a special type of noise but later work showed that it is in fact a deterministic process. To distinguish such apparent randomness from a true stochastic process, to understand and explain this behaviour, it was necessary to develop new mathematical techniques within the framework of nonlinear systems theory.

In many practical situations the only information available from a dynamical process takes the form of a time series and this has to be used to extract both qualitative and quantitative features of the underlying system. Because of the erratic structure it is not easy to quantify and characterise chaotic motion. The most typical measures of chaos are the correlation dimension and the largest Lyapunov exponent and these can often be estimated directly from the recorded time series. The correlation dimension, characterises the fractal nature of the geometric structure generated by a chaotic system in phase space. The largest Lyapunov exponent, a generalisation of the concept of Lyapunov exponents of a system around an equilibrium point, reflects the sensitivity of the chaotic motion with respect to initial conditions. Given a sufficiently long record of clean data, both of these dynamical invariants can be estimated with reasonable accuracy.

Unfortunately in many practical situations only a limited amount of data, often corrupted by noise is available and this makes the estimation of the above dynamical invariants very difficult. In this situation, the only way to gain some insight into the process under investigation is to try to estimate a mathematical model based on a short length of noisy observations. The presence of noise is in this case the main factor that can limit the possibility of estimating a dynamically valid model.

Traditional Wiener or Kalman filtering techniques can be used in an attempt to separate the signal from the noise. But this tends to be limited by the fact that one characteristic of chaotic signals is aperiodic behaviour and a nearly flat broad spectrum that resembles a pure stochastic process. Simple low pass filtering for example, can severely damage the dynamical properties of the original signal with direct implications on the dynamical validity of estimated models.

Methods that are superior to low pass filtering make use of singular value decomposition (SVD) (Vautard1992) or are prediction-based algorithms that attempt to correct the noise perturbed trajectory at one point using both future and past information to predict the true value (Schreiber and Grassberger 1991, Aguirre et al 1994).

More recently wavelet decompositions have provided a useful tool for recovering nonlinear non-stationary signals blurred by noise with high accuracy based on the fact that a multiresolution wavelet decomposition provides a very useful time-frequency analysis which can be exploited to reduce the stochastic part of a signal (Donoho1992).

In the present paper a new wavelet smoothing algorithm is introduced and applied to the noise reduction of chaotic signals. The main objective is to develop an approach which reduces the noise on the signals without altering the dynamical invariants or characteristics of the underlying system. This is achieved using a multiresolution decomposition of the signal using B-spline wavelets followed by a smoothing operation using the wavelet coefficients. It is shown that enhanced performance can be obtained by smoothing rather than filtering. Both the correlation dimension and the largest Lyapunov exponent are computed for the original system and the model estimated using the smoothed noisy signal to illustrate that the dynamic invariants of the system remain unchanged by the wavelet smoothing routine. Chua's circuit and Lorenz's equations are used as examples to demonstrate the performance of the algorithm.

The paper is organised as follows. Section 2 briefly introduces dynamical system representations and discusses the effects of noise. Some specific problems encountered with noisy chaotic signals, which make wavelet decomposition a feasible tool for smoothing such signals, are also discussed. In Section 3 the concept of wavelet multiresolution decomposition is introduced. Section 4 describes the implementation of such a decomposition based on B-spline wavelet functions that are used to generate a nonorthogonal multiresolution approximation. The special properties of B-spline wavelets which make them attractive in this application, optimal time-frequency localisation and a linear phase characteristic, are also presented. In the same section, rigorous results concerning the effects of noise on the wavelet coefficients of a chaotic signal are derived and a technique to reduce the stochastic part of the noisy wavelet coefficients and subsequently smooth the underlying signal is introduced. The new method is illustrated in two numerical examples which are contained in Section 5. Finally Section 6 involves comments and conclusions while acknowledgements are contained in Section 7.

3. Noise and Chaotic Systems

2.1 Dynamical systems representation

The notion of a dynamical system covers any system evolving in time which can be described mathematically either as a set of ordinary differential equations or by discrete mappings or difference equations.

Differential equations are used to model systems which evolve continuously in time. Such systems occur frequently in practice and are represented in general by the following equations

$$\begin{cases} \dot{x}(t) = f(x(t), u(t)) \\ y(t) = h(x(t), u(t)) \end{cases} \tag{2.1.1}$$

with $f: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $h: \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

The solution or flow of the dynamical system described by equation (2.1.1) is defined locally for each $\xi \in \mathbf{R}^n \times \mathbf{R}^m$ as the function

$$\varphi(t):(-\varepsilon,\varepsilon) \rightarrow \mathbb{R}^n \tag{2.1.2}$$

with $\varepsilon > 0$ such that

$$\frac{d\varphi(t)}{dt} = f(t,\varphi(t)), \text{ and } \varphi(0) = \xi \tag{2.1.3}$$

The function $\varphi(t)$ is unique with respect to the initial condition ξ and can be denoted as $\varphi(t,\xi)$ or as $\varphi(t,x_0,u_0)$ where $x_0 = x(0)$, $u_0 = u(0)$.

From equation (2.1.3) is easy to see that since $f(t,x(t),u(t))$ is C^r then

$$\varphi(t) \in C^r \tag{2.1.4}$$

Although dynamical systems are continuous in most practical situations, such systems are usually observed at discrete instants in time. This means that the flow of a continuous dynamical system has to be sampled in order to analyse, model or control it. Mathematically the discretisation of a continuous flow is described by Euler's integral equation

$$\varphi(t) = \xi + \int_0^t f(s,\varphi(s))ds \tag{2.1.5}$$

where ξ is the initial condition. Equation (2.1.4) defines a function

$$\Phi(\varphi) = \xi + \int_0^t f(s,\varphi(s))ds \tag{2.1.6}$$

so that $\varphi(t)$ is the function which solves the fixed point equation

$$\varphi = \Phi(\varphi) \tag{2.1.7}$$

The solution map satisfies the conditions

$$\begin{aligned} &\varphi(0,\xi) = \xi \\ &\varphi(t,\varphi(s,\xi)) = \varphi(t+s,\xi), \quad t,s \in \mathbb{R} \end{aligned} \tag{2.1.8}$$

as long as it exists since $\varphi(t)$ may not be defined in general for all $t>0$

Following (2.1.7) we can define a discrete dynamical system as

$$\begin{cases} x(k+1) = f(x(k),u(k)) \\ y(k) = h(x(k),u(k)) \end{cases} \tag{2.1.9}$$

where once again $x \in \mathbb{R}^n$ is the state vector, $u \in \mathbb{R}^m$ is the input or control of the system and y the output of the system.

A dynamical system can under certain conditions, (for a particular choice of parameters) behave chaotically The trajectory of such a system gives birth to a geometric structure in the state space which is neither a fixed point or a limit cycle nor a quasi periodic torus. Such a structure, known as a chaotic attractor, which is not even a manifold, is related to a Cantor set and it possesses a fractional dimension. The motion on such attractor can be associated with random behaviour, although the equations of motion are fully deterministic. An intrinsic property of a chaotic system is the sensitive dependence of initial conditions which is illustrated by the fact that nearby points on the attractor will separate exponentially in time, until they become totally uncorrelated.

When the signal is chaotic some assumptions can be made concerning the smoothness of the chaotic signal. Such assumptions are based on the fact that chaos can only appear if the dynamical process is described by, at least, a third-order differential equation (autonomous case). This is a *sine qua non* condition for the existence of chaos in a system governed by ordinary differential equations. Following this and (2.1.4) we can state the following proposition

Proposition 1: *A chaotic signal y(t) is at least three times continuously differentiable with respect to time, almost everywhere.*

Almost everywhere means that $y(t)$ is not C^r, $r \geq 3$ for a countable set of points. This reflects the fact that for a given dynamical system described by a set of ordinary differential equations a global solution may not exist. An example in this sense is given by the Chua's circuit described by the equations

$$\begin{cases} \dot{x} = \alpha \cdot (y - l(x)) \\ \dot{y} = x - y + z \\ \dot{z} = -\beta \cdot y \end{cases} \tag{2.1.10}$$

where

$$l(x) = \begin{cases} m_1 x + (m_0 - m_1) & x \geq 1 \\ m_0 x & |x| \leq 1 \\ m_1 x - (m_0 - m_1) & x \leq -1 \end{cases} \tag{2.1.11}$$

From (2.1.10) and (2.1.11) it is clear that $\frac{d^2x}{dt^2}$ does not exist for $x(t)=\{-1,1\}$ since $\frac{\partial f}{\partial x}$, with $f = \alpha(y - h(x))$, is not defined at these points. Taking the observation function to be $h(x,y,z) = x$ it is clear that in this case the output of the system $y(t)=x(t)$ is C^3 everywhere except at the points for which $x(t)=-1$ or $x(t)=1$.

In practice our ability to observe and analyse a dynamical system is limited by the presence of unknown perturbations. Such perturbations are generally referred to as noise and were formally described by the concept of innovations introduced by Kolmogorov in 1941. In this sense an innovation process at each step generates new information which cannot be predicted from previous measurements. The innovation part of an observation is orthogonal to the previous measurement so an innovation process is comparable to a Gram-Schmidt orthogonalisation for vectors.

For a dynamical system, the noise can be classified as two types namely dynamical and observational.

2.2 Dynamical noise vs. observational noise

There is a fundamental difference between observational and dynamical noise. Observational noise is related to the accuracy the output of a dynamical system can be measured to. The presence of observational noise can be introduced in the general representation of the dynamical system by augmenting the observation function $h(x(t), y(t))$ in (2.1.1). In the general case this can be expressed as

$$y(t) = h(x(t), u(t), \varepsilon_h(t)) \tag{2.2.1}$$

where $\varepsilon_h(t)$ is white noise.

However, in most practical cases the observational noise enters additively, that is

$$y(t) = h(x(t), u(t)) + \varepsilon_h(t) \tag{2.2.2}$$

Dynamical noise (known also as plant noise) is the noise that perturbs the states of the system and the effect of such perturbation can be included in the state space equations given in (2.1.1) which become

$$\dot{x}(t) = f(x(t), u(t), \varepsilon_f(t)) \tag{2.2.3}$$

with $\varepsilon_f(t)$ again white noise. Here observational and dynamical noise can be considered to be generated by two independent random processes.

As an example observational noise can be considered as generated in the measurement process by round-off errors or by the imprecision of the instruments, dynamical noise can occur because of a random variation of a parameter in the dynamical system. Dynamical noise generally creates greater problems than observational noise. However, in most of practical cases observational noise accounts for most of the perturbations affecting the system. In both cases the noise can be viewed as an unknown (unobserved) input which perturbs the system.

The present study investigates the possibility of removing the additive part of the observational noise present in a measured chaotic signal, prior to the identification stage. The smoothing method introduced in the paper deals in this sense with an unfavourable situation, when the observed signal is generated by a chaotic system.

Generally the task of filtering or smoothing a chaotic signal poses some specific problems due to the typical aperiodic behaviour and broadband spectrum of such signals. The broadband spectrum of a chaotic signal reflects the fact that the chaotic motion can be represented by a continuum of frequencies. This property means traditional filtering techniques often fail when applied to a chaotic signal. It is well known that a purely random or noisy process also has a broadband spectrum so that the characteristic frequencies of a chaotic signal are practically inseparable from that of a random process.

The new approach introduced in this paper makes use of the multiresolution wavelet decomposition of the signal, which is represented as a wavelet series, to remove the noise. The smoothing procedure is performed indirectly by using just the coefficients of the wavelet series representation. First it is important to find how the noise is reflected by the wavelet coefficients of the wavelet series representation of the signal. New results concerning this are derived in section *4.2* using only the assumption that the noise is white and additive. Next a way to reduce the noise using mutual information between neighbouring points in the data sequence is introduced in *4.3* and it is shown both analytically and numerically that this does not affect the dynamical properties of the underlying signal.

The advantage of using the wavelet approach is twofold. First of all the multiresolution analysis based on wavelets provides a time-frequency representation of the signal of interest which is particularly suitable for non-stationary signals. While simple Fourier analysis provides global information about the frequency content of the signal, wavelet analysis can relate a particular frequency content to the moment in time when it occurred. It is possible therefore to break down the apparently continuos spectrum of a chaotic signal so that the frequency content of the signal at different instances in time depends on the particular unstable periodic orbit the trajectory follows at that time instant.

A period one or period two orbit for example is characterised by very sharp peaks in the frequency domain. Such peaks which cannot be distinguished clearly in the spectrum of the chaotic signal, can be separated using the time-frequency window provided by wavelet functions. This makes the task of separating the noise from the true signal easier. Moreover such a separation can be made not just on the grounds of the characteristic spectrum. An important property of the wavelet representation is the fact that the magnitude of the wavelet coefficients reflects in an unequivocal way the properties of the signal. This feature is not available in the classical Fourier representation where it is impossible to predict the properties of a function like size and regularity solely from the values of the Fourier coefficients.

An important aspect, which has to be considered before attempting to filter any signal from a dynamical system, concerns the smoothness of the signal. In general the smoothness of the data determines the smoothness needed for the wavelet basis. Preserving the smoothness of the original signal is important especially in system identification where the derivatives of the signal contain essential information about the

dynamical motion. In this sense by performing the smoothing procedure over the wavelet coefficients, simultaneously, depending on the smoothness class of the wavelet basis, the derivatives of the signal are also smoothed.

The main constraint on the smoothing procedure is that it should not affect the dynamical characteristics of the signal so that the qualitative and quantitative properties of the underlying system remain unchanged. Some quantitative characteristics of chaotic signals are traditionally obtained directly from the data set and this includes estimates of the correlation dimension and the largest positive Lyapunov exponent. Usually to obtain good estimates of these invariants a very long data record is required.

Another approach which has proven to be more reliable especially when a long time series is not available is to estimate a model and to use this to compute the dynamical measures. There are many characteristics of the system which can be determined in this way providing the model accurately captures all the dynamical characteristics of the underlying system. For example the full Lyapunov spectrum can be evaluated using the estimated model and then information such as the Lyapunov dimension can be readily computed. Qualitative aspects of the dynamical motion, concerning the equilibrium points and the structural stability of the system, presented in the form of a bifurcation diagram can also be calculated.

3. Wavelet decomposition

3.1. Multiresolution approximation

The main ingredient of a multiresolution analysis is the sequence of nested subspaces V_j satisfying the following conditions:

(1.) $\ldots \subset V_{-1} \subset V_0 \subset V_1 \subset \ldots;$

(2.) $\bigcup_{j=-\infty}^{\infty} V_j = L^2(\mathbf{R})$

(3.) $\bigcap_{j \to -\infty} V_j = \{\emptyset\}$

(4.) $f(x) \in V_j \Leftrightarrow f(2^{-1}x) \in V_{j-1}$

Conditions (1) and (2) state that successive projections $P_j f$ of a function f onto V_j approximate f with increasing accuracy. Actually property (1) states that for $i<j$, $P_j f$ can be viewed as a finer approximation of f than $P_i f$.

It follows from property (2) that

$$f = \lim_{j \to \infty} P_j f \qquad (3.1.1)$$

By decreasing the resolution and taking coarser and coarser approximations, the resulting projection $P_j f$ will contain less and less information about the function f and will eventually converge to zero.

$$\lim_{j\to-\infty} P_j f = \{\emptyset\} \tag{3.1.2}$$

Condition (4) is the so called scaling property and introduces a link between the successive approximation subspaces V_j . Rewrite the condition as

$$f(x)\in V_j \Leftrightarrow f(2^{-j}x)\in V_0 \tag{3.1.3}$$

to see that all the subspaces V_j are scaled versions of a central subspace V_0.

The importance of the scaling property arises from the fact that a multiresolution analysis can be induced by means of a single function $\varphi(x)$ and its translates generating a basis $\{\varphi(x-k)\}_{k\in Z}$ which span the subspace V_0. Due to the scaling property the same function or rather a scaled version of it can generate the successive resolution subspaces V_j in a similar way.

When the multiresolution approximation was first defined by S.Mallat and Y. Meyer the scaling function $\varphi(x)$ was orthonormal to all translates $\varphi(x-k)$ thus leading to an orthonormal multiresolution approximation. However the orthonormality condition imposes tight constrains on the scaling function and moreover this condition is not essential in many applications. The orthonormal multiresolution approximation can be viewed as a particular case included in the more general framework of nonorthogonal multiresolution approximations generated in terms of a redundant but still useful basis known as a frame (see Chui 1992 for a definition).

Orthonormal or not, the scaling functions spanning the resolution subspace V_j can be expressed as $\{2^{j/2}\cdot\varphi(2^j x-k)\}_{k\in Z}$ where $2^{j/2}$ is a normalisation parameter required to ensure that the following relation, known as the partition of unity property, holds

$$\int_{-\infty}^{\infty}\varphi_{j,k}(x)dx = 1 \tag{3.1.4}$$

Wavelet subspaces W_i can be introduced in the context of multiresolution analysis as the orthogonal complement of V_j with respect to the next resolution subspace V_{j+1}. This can be written as

$$V_j \oplus W_j = V_{j+1} \tag{3.1.5}$$

where $\oplus$ denotes the orthogonal sum of subspaces. We would like to have a function $\psi(x)$ to provide a basis for W_0 in the same way the scaling function $\varphi(x)$ provides a basis for V_0 . Such a function will be a linear combination of the functions $\{2^{1/2} \cdot \varphi(2 \cdot x - k)\}_{k \in Z}$ which form a frame basis for V_1 since $\psi(x) \in W_0 \subset V_1$. This is also true for the scaling function $\varphi(x)$ which can be expressed in a similar way in terms of the basis functions spanning V_1.

The formulas which describe this

$$\begin{aligned} \varphi(x) &= \sum_k p_k \varphi(2x-k); \\ \psi(x) &= \sum_k q_k \varphi(2x-k); \end{aligned} \tag{3.1.6}$$

are called the two scale relations of the scaling and wavelet functions and $\{p_k\}$ $\{q_k\}$ are known as the two scale reconstruction sequences. Moreover any basis function $\varphi(2x-k)$ from V_1 can alternatively be written using the scaling and wavelet basis from V_0 and W_0 respectively as

$$\varphi(2x-k) = \sum_{l=-\infty}^{\infty} \{a_{k-2l}\varphi(x-l) + b_{k-2l}\psi(x-l)\}, \quad k \in \mathbb{Z} \tag{3.1.7}$$

This is referred to as the decomposition relation with $\{a_k\}$ and $\{b_k\}$ the decomposition sequences.

Once the scaling function is known, equation (3.1.6) gives a constructive method to determine the wavelet function such that each subspace W_i is the closure in $L^2(\mathbb{R})$ determined by the linear span of the collection of functions

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \tag{3.1.8}$$

It is interesting to note, however, that the scaling and wavelet functions are not unique and that different choices are available.

Following the inclusion condition $\ldots V_{-1} \subset V_0 \subset V_1 \subset V_2 \ldots$ and the fact that W_j is the orthogonal complement for the subspace V_j with respect to V_{j+1}

$$V_j \oplus W_j = V_{j+1} \tag{3.1.9}$$

we can write

$$\bigoplus_{i=\{0,j\}} W_i \oplus V_0 = V_{j+1} \tag{3.1.10}$$

or equivalently,

$$V_0 \oplus W_0 \oplus \ldots \oplus W_j = V_{j+1} \tag{3.1.11}$$

where in contrast to the subspaces V_j which are nested

$$W_j \cap W_{j+1} = \{\varnothing\} \tag{3.1.12}$$

The different resolution subspaces W_j are mutually orthogonal and decompose the Hilbert space $L^2(\mathbb{R})$ in a similar way, for example, the orthonormal basis $e^{j\omega k}$ decomposes $L^2((0,2\pi))$. Unlike the Fourier series the major characteristic of the decomposition is that the direct sum of the subspaces W_i, up to a given index, say j, can be replaced by V_{j+1} that is

$$V_{j+1} = \bigoplus_{i=\{-\infty,j\}} W_i \tag{3.1.13}$$

This relation gives an alternative way of representing the approximation of a function at resolution V_j (or the projection of f onto V_j) by adding up the projections of f on each of the subspaces W_i with $i = \{-\infty, j\}$. From a practical point of view we can start with a fixed resolution V_j and later improve the approximation by adding the detail or the projections onto W_i. Such details could be seen as variations or oscillations of arbitrarily small energy, of the function analysed, which in the frequency domain are represented by frequency bands.

In the wavelet representation the coefficients play the same role as the Fourier coefficients in Fourier analysis, the difference being that here they give information about the local properties of the function analysed.

The expansion of a function in terms of scaling and wavelet functions (wavelet series) is given by the following equation

$$f(x) = \sum_k c_{j,k}\varphi_{j,k}(x) + \sum_{i \geq j,k} d_{i,k}\psi_{i,k}(x) \tag{3.1.14}$$

where

$$P_j f = \sum_k c_{j,k} \varphi_{j,k}(x) \tag{3.1.15}$$

is the projection of f onto V_j and

$$Q_j f = \sum_{k \in \mathbb{Z}} d_{j,k} \psi_{j,k}(x) \tag{3.1.16}$$

is the projection of f onto W_j, with

$$\begin{aligned} \varphi_{j,k}(x) &= 2^{j/2} \varphi(2^j x - k), \\ \psi_{i,k}(x) &= 2^{i/2} \psi(2^i x - k) \end{aligned} \tag{3.1.17}$$

Finally we can write

$$f(x) = P_j f + \sum_{i \geq j} Q_i f = P_j f + Q_j f + Q_{j+1} f + \ldots \tag{3.1.18}$$

This decomposition enables us to choose how much of the details contained in $Q_j f$, $Q_{j+1} f$,.... we want to incorporate in our approximation.

2.2. Multiresolution Pyramid Decomposition

Multiresolution analysis leads to a fast scheme for the computation of the wavelet coefficients of a given function.

Starting with a finer projection of a function f onto V_j this can be decomposed into a coarser approximation onto V_{j-1} together with the difference or detailed information between the successive levels V_j and V_{j-1} that is the orthogonal projection onto the wavelet subspace W_{j-1}. This decomposition can be continued as long as desired so that the initial projection $P_j f$ will be written as a projection $P_{j-i} f$ onto the coarser subspace V_{j-n} plus a sum of projections onto the complementary subspaces $\{W_{j-i}\}_{i=1,n}$

$$P_j f = P_{j-n} f + Q_{j-n} f + \ldots + Q_{j-1} f, \quad j, n \in \mathbb{Z} \tag{3.2.1}$$

At the finer resolution the function f will be approximated as

$$f(x) \cong P_j f = \sum_k c_{j,k} \varphi_{j,k}(x) \tag{3.2.2}$$

Using the decomposition relation we can alternatively express (3.2.2) in terms of scaling and wavelet basis functions at resolution j-1 as follows

$$f(x) \cong P_j f = \sum_k \left\{ c_{j-1,k} \varphi_{j-1,k}(x) + d_{j-1,k} \psi_{j-1,k}(x) \right\} \tag{3.2.3}$$

The scaling and wavelet function coefficients at resolution j-1 can be computed in an efficient way using the fast decomposition algorithm

$$\begin{cases} c_{j-1,k} = \sum_l a_{l-2k} c_{j,l} \\ d_{j-1,k} = \sum_l b_{l-2k} c_{j,l} \end{cases} \tag{3.2.4}$$

derived from (3.1.7). Relations (3.2.4) describe a moving average process, involving the scaling coefficients at resolution j and the decomposition sequences $\{a_l\}_{l \in \mathbf{Z}}$ and $\{b_l\}_{l \in \mathbf{Z}}$. Downsampling or decimating the data sequence which results after the moving average procedure has been performed, by taking every second value, is required to obtain the desired coefficients.

The reconstruction algorithm is a consequence of the two scale relations (3.1.6) and is used to calculate the scaling function coefficients at the finer resolution level from the scaling and wavelet function coefficients at the coarser level. The computation involves the following moving average scheme,

$$c_{j,l} = \sum_k \left[p_{l-2k} c_{j-1,k} + q_{l-2k} d_{j-1,k} \right] \tag{3.2.5}$$

this time with the weighting sequences $\{p_k\}$ and $\{q_k\}$. In this case upsampling is required before the MA scheme is used and this is done by inserting a zero between every two consecutive terms of the input sequences $\{c_{j-1,k}\}$ and $\{d_{j-1,k}\}$.

The scaling and wavelet functions can be considered as filter functions. While the scaling function behaves like a low-pass filter the wavelet function has the characteristic of a band-pass filter. This is why the projection of a function onto a subspace V_j can be viewed as a smoothing operation. The importance of the decomposition and reconstruction algorithm is that by separating the details of the approximated signal and storing them in different subspaces W_i of V_j using the decomposition algorithm, a better analysis of the signal can be carried out.

An important feature of the wavelet coefficients is that the energy of the signal is reflected locally by their magnitude. In this way wavelet decompositions will compress the energy of the signal into fewer numbers of large coefficients. This property is extremely valuable in data compression where simply by discarding the small coefficients (small relative to the other coefficients) the original signal can be represented more economically.

4. Nonorthogonal B-spline Multiresolution Approximation

4.1. Construction and properties of B-spline Multiresolution Approximation

For many applications of the multiresolution analysis, orthonormality is not essential. Wavelets need not be orthonormal. Relaxing the orthonormality condition leads to nonorthogonal multiresolution approximations and provides a more flexible framework for function approximation.

A typical example of scaling functions $\varphi(x)$ are the m-th order cardinal B-spline functions $\beta_m(x)$ with $m \in \mathbb{Z}$ which are defined recursively by the integral convolution

$$\beta_m(x) = \int_{-\infty}^{\infty} \beta_{m-1}(x-t)\beta_1(t)dt \tag{4.1.1}$$

where $\beta_1(x)$ is the characteristic function of the interval (called also the indicator function) $\chi(t)$. Relation (4.1.1) thus becomes

$$\beta_m(x) = \int_0^1 \beta_{m-1}(x-t)dt \tag{4.1.2}$$

since

$$\beta_1(t) = \chi(t) = \begin{cases} 1 & x \in (0,1) \\ 0 & \text{otherwise} \end{cases} \tag{4.1.3}$$

Based on equations (4.1.1) and (4.1.2) the B-spline basis functions can be defined starting with the first order basis function (of degree zero) given explicitly in (4.1.3) and using the following recursive algorithm to construct basis functions of higher order

$$\beta_m(x) = \frac{x}{m-1}\beta_{m-1}(x) + \frac{m-x}{m-1}\beta_{m-1}(x-1). \tag{4.1.4}$$

The B-spline function β_m (of order m) consists in fact of m nontrivial polynomial pieces of degree m-1 so that

$$\beta_m|_{[k-1,k)} = P_{m-1,k}, \quad k = 1,\ldots,m \tag{4.1.5}$$

where $P_{m-1,k}$ is a polynomial of degree m-1. Denoting π_n as the collection of all polynomials of degree at most n and with C^n the space of continuous functions having up to n continuous derivatives, allows us to define the subspace V_0 generated by $\varphi(x) = \beta_m(x)$ as the subspace of all functions $f \in C^{m-2} \cap L^2(\mathbf{R})$ such that the restriction of f to any interval $[k\text{-}1,k]$ is a polynomial with degree at most m-1. Each polynomial piece can be computed analytically which gives an alternative way to compute the value of the B-spline functions at any point.

To generate a nonorthogonal multiresolution approximation, the associated scaling function $\varphi(x)$ has to generate a frame over V_0. In the case above, for m a positive integer, the scaling function $\varphi(x) = \beta_m(x)$ generates a frame for the subspace V_0 and moreover it can be proven that a whole multiresolution approximation can be defined based on B-spline basis functions (see Chui 1992 for details).

Let W_j denote the sequence of orthogonal complementary subspaces with respect to V_j,. A B-spline wavelet basis can then be defined for any W_j , which satisfies the scaling property. It can be shown that up to multiplication by a constant the compactly supported wavelets $\psi(x)$ with minimum support that correspond to the m-th order cardinal B-spline are unique. The support of the m-th order B-spline wavelet is an interval of length $2m$-1. Moreover all wavelets are symmetric for even m and anti symmetric for odd m.

Because of the total positivity property of B-spline functions (analysed in Chui(1992) , the B-spline series

$$\sum_k c_k \cdot \beta_m(x-k) \tag{4.1.4}$$

is one of the most suitable tools for smoothing purposes. In contrast the B-spline wavelet functions have a strong oscillatory character so, normally, a B-spline wavelet series will detect variations in the data. The smoothing effect of the B-spline scaling functions arises because oscillations of a data sequence are diminished when convolved with a discrete B-spline interpolation kernel.

In fact scaling functions and wavelets can be considered as filter functions. The difference between them is that while the scaling function acts as a low pass filter, the corresponding wavelet function behaves like a bandpass filter. The frequency characteristics of B-spline scaling functions is presented in Fig. 1(a) for the case of a cubic B-spline (m=4). Fig.1(a) shows how the low frequency band of the scaling basis

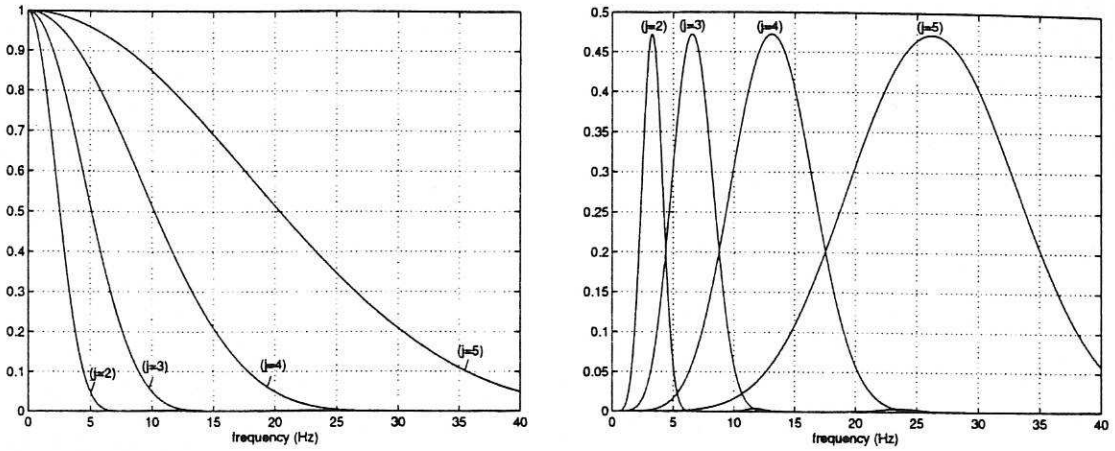


Fig.1. The Fourier transform of cubic B-spline (a) scaling and (b) wavelet function at different scales j

function increases with the scale. Fig.1(b) illustrates the multiband structure generated by the corresponding wavelets, which in the frequency domain cover higher and wider frequency ranges when the scale is increased.

If a signal is represented using the multiresolution approximation approach, such as a wavelet series, the reconstructed signal is nothing other than the result of a linear filtering process. From this perspective we should be aware of possible distortions in the reconstructed signal. A usual requirement in such cases, to avoid distortions, is that the filter should have linear or at least generalised linear phase.

By definition $f \in L^2(\mathbb{R})$ is said to have linear phase if its Fourier Transform $\hat{f}$ satisfies

$$\hat{f}(\omega) = \pm\left|\hat{f}(\omega)\right| \cdot e^{-i a\omega} \tag{4.1.5}$$

where $a \in \mathbb{R}$ and "+ "or "-" sign is independent of ω. Moreover $f \in L^2(\mathbb{R})$ is said to have generalised linear phase if

$$\hat{f}(\omega) = F(\omega) \cdot e^{-i(a\omega+b)} \qquad a,b \in \mathbb{R} \tag{4.1.6}$$

where $F(\omega)$ is a real valued function. In both cases the constant a represents the phase of $\hat{f}$.

The Fourier transform of the B-spline scaling function β_m of order m can be written as

$$\hat{\beta}_m = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^m \cdot e^{-i \cdot m \cdot \omega/2} \tag{4.1.7}$$

which clearly shows that β_m has linear phase which in this case equals *m/2*. On the other hand the B-spline wavelet functions, can all be shown (Chui 1992) to have generalised linear phases, due to the fact that all wavelets are either symmetric (for odd *m*) or antisymmetric (for even *m*).

Another essential characteristic of the B-spline wavelet functions is the excellent time-frequency (or space-frequency) localisation. This is why they are most suitable for time-frequency analysis of non-stationary signals. B-splines provide a flexible time-frequency window which narrows (in time) at high centre-frequencies and widens at low centre-frequencies.

By definition a function $f \in L^2(\mathbb{R})$ qualifies as a window function if $x \cdot f(x) \in L^2(\mathbb{R})$. Such a window function can be characterised by the centre x_{center} defined as

$$x_{center} = \frac{1}{\|f\|_2^2} \cdot \int_{-\infty}^{\infty} x \cdot |f(x)|^2 \, dx \tag{4.1.8}$$

and the width of the window $2 \cdot \Delta_f$ where

$$\Delta_f = \frac{1}{\|f\|_2^2} \cdot \left\{ \int_{-\infty}^{\infty} (x - x_{center}) \cdot |f(x)|^2 \, dx \right\}^{1/2} \tag{4.1.9}$$

Such a window will localise the signal in time with a time window

$$\left[x_{center} - \Delta_f, x_{center} + \Delta_f \right] \tag{4.1.10}$$

Therefore for wavelet functions at scale *j*, $\psi_{j,k} = 2^{j/2} \cdot \psi(2^{-j}x - k)$ the time window localisation is given by

$$\left[\frac{k}{2^j} + \frac{x_{center\,\psi}}{2^j} - \frac{1}{2^j} \cdot \Delta_\psi, \frac{k}{2^j} + \frac{x_{center\,\psi}}{2^j} + \frac{1}{2^j} \cdot \Delta_\psi \right] \tag{4.1.11}$$

where $x_{center\,\psi}$ and Δ_c are the centre and the radius of the window function ψ respectively.

The wavelet function $\psi_{j,k}$ defines a window function in both the time and frequency domain. In the frequency domain the window of $\hat{\psi}_{j,k}$ (the Fourier transform of $\psi_{j,k}$) can be defined by the centre-frequency $\omega_{center\,\hat{\psi}}$, the window width $2 \cdot \Delta_{\hat{\psi}}$ of the mother wavelet and by the scaling factor

$$\left[2^j \cdot (\omega_{center\hat{\psi}} - \Delta_{\hat{\psi}}), 2^j \cdot (\omega_{center\hat{\psi}} + \Delta_{\hat{\psi}})\right] \tag{4.1.12}$$

This is why an important property of the frequency window is that the ratio between the centre frequency and the width of the frequency band is an independent quantity with respect to the scale, and is given by

$$Q = \frac{\omega_{center} / a}{2 \cdot \Delta_{\hat{\psi}} / a} = \frac{\omega_{center}}{2 \cdot \Delta_{\hat{\psi}}} \tag{4.1.13}$$

The ratio Q, known as the quality factor remains constant for all resolutions and in this way the multiresolution analysis is said to provide a constant Q frequency analysis.

However due to the *Uncertainty Principle* we cannot have simultaneously perfect localisation in both the time and frequency domains. This can be expressed in terms of the smallest possible area achieved by a window function in the time frequency plane where the lower bound is given by

$$2 \cdot \Delta_{\hat{f}} \cdot \Delta_f \geq 1 \tag{4.1.14}$$

The equality is attained only when the window function f is Gaussian.

From this point of view the B-spline wavelet functions are in many respects close to the optimal time-frequency localisation attained by Gaussian functions. Since Gaussian functions are not suitable for use as a basic wavelet, the B-spline wavelets of larger order m are the best candidates. For $m=6$ for example

$$\Delta_{\psi_m} \cdot \Delta_{\hat{\psi}_m} = 0.500367 \cong \frac{1}{2} \tag{4.1.15}$$

which is very close to the optimum *0.5*.

4.2. Computation of B-spline wavelet coefficients of a signal

In previous sections the attractive properties that make B-spline wavelet decomposition the optimum candidate for smoothing purposes and time-frequency analysis of non-stationary signals was presented. The practical implementation of such a decomposition must now be considered.

Consider a discrete time series $f(t_i)$.

The first step consists in expanding the signal in terms of the scaling basis functions corresponding to the resolution subspace V_j. This involves the computation of the coefficients of the series

$$f(t_i) \equiv \sum_k c_{j,k} \cdot \varphi_{j,k}(t_i) \tag{4.2.1}$$

where $\varphi_{j,k}(t) = 2^{j/2}\varphi(2^j t - k)$ and $\varphi(t) = \beta_m(t)$ the m-th order cardinal B-spline function defined in (4.1.1).

Equation (4.2.1) can be seen as a projection of the signal $f(t)$ onto the approximation subspace V_j. This is in fact a B-spline interpolation problem and there are two practical approaches to find the corresponding coefficients involved in the equation.

One way is to calculate the coefficient sequence corresponding to the basis functions spanning the subspace V_j by convoluting the data sequence (moving average procedure) with an appropriate spline interpolation operator constructed as in Chui(1992). It is important to choose the resolution subspace V_j so that the expansion (4.2.1) will give a sufficiently good approximation of the signal $f(t_i)$. For example for $h=2^{-j}$ the approximation error is bounded by a constant multiple of h^m as $j \to \infty$.

Another possibility is to compute the coefficients of the series (4.2.1) using a least squares algorithm since it is a linear-in-the-parameters expression. This gives the orthogonal projection of $f(t_i)$ onto the approximation subspace V_j.

In both situations it is important to take into account that the sequence we want to approximate has in practice a finite extent, so in order to avoid distortions, due to the lack of continuity at the boundaries which occur in many applications such as image processing a standard practice is to extend the signal on both sides by using the mirror image of the signal. In our case it is sufficient that we use a longer data sequence than needed which will contain the signal of interest centred within the sequence so in this way we avoid discontinuities at the ends of the signal of interest.

Once the coefficients have been determined the signal will be represented, with an accuracy depending on the resolution subspace V_j chosen, by the series (4.2.1). At this stage presuming that the signal is corrupted by noise, the approximation represents the noisy signal.

A multiresolution pyramidal decomposition leads to the following equivalent expression of (4.2.1)

$$f(t_i) \equiv \sum_k c_{j-p,k} \cdot \varphi_{j-p,k}(t_i) + \sum_k \sum_{l=j-p}^{j} d_{l,k} \cdot \psi_{l,k}(t_i) \tag{4.2.2}$$

To perform such a decomposition the weight sequences $\{a_k\}$ and $\{b_k\}$ must be determined, which for the case of the B-spline multiresolution approximation are

infinite length sequences (Chui 1992). The moving average procedure used to calculate the wavelet and scaling coefficients at the coarser resolution level is an IIR (Infinite Impulse Response) filter. Usually IIR filters can be implemented as ARMA (Autoregressive Moving Average) filters providing that the Z-transform of the weight sequence is a rational function. Otherwise the infinite weight sequence has to be truncated to give an FIR filter. Truncation coupled with round-off errors will usually induce errors which can be estimated (Chui 1992) and made arbitrarily small.

It has been proved (Chui 1992) that the decomposition sequences for B-spline wavelets are ARMA since the Z-transform of the weight sequences can be described by the rational functions

$$\begin{aligned} G(z) &= \frac{1}{2}\sum_{n=-\infty}^{\infty} g_n \cdot z^n = z^{-1}\cdot\left(\frac{1+z}{2}\right)^m \cdot \frac{E_{2m-1}(z)}{E_{2m-1}(z^2)} \\ H(z) &= \frac{1}{2}\sum_{n=-\infty}^{\infty} h_n \cdot z^n = -z^{-1}\cdot\left(\frac{1-z}{2}\right)^m \cdot \frac{(2m-1)!}{E_{2m-1}(z^2)} \end{aligned} \tag{4.2.3}$$

where the decomposition sequences are given by

$$\begin{cases} a_n = \frac{1}{2}\cdot g_{-n} \\ b_n = \frac{1}{2}\cdot h_{-n} \end{cases} \tag{4.2.4}$$

and E_{2m-1} is the Euler-Frobenius polynomial of order $2m$-1 (of degree $2m$-2).

Formulas to calculate the truncated decomposition sequences are provided in Chui(1992) along with error bound estimates for the cases m=2,3,4 (where m is the B-spline order).

Having computed the truncated decomposition sequences $\{a_k\}$ and $\{b_k\}$ the wavelet decomposition of a signal can be performed step by step as follows:

- Compute the scaling function coefficients $c_{j,k}$ in equation (4.2.1) at the finer resolution j.
- Following equation (3.2.4) perform a moving average algorithm over the coefficients $c_{j,k}$, using the decomposition sequences $\{a_k\}$ and $\{b_k\}$ as weights, and downsample or decimate by two (take every other point) the results to obtain the scaling and wavelet basis function coefficients $c_{j-1,k}$ and $d_{j-1,k}$ at resolution j-1.
- Repeat the second step with $c_{j-1,k}$ replacing $c_{j,k}$. In this way at the p-th iteration the coefficients $c_{j-p+1,k}$ are used to compute

$c_{j-p,k}$ and $d_{j-p,k}$. The decomposition can be continued as long as desired.

It should be observed at this stage that theoretically, when the signal is considered continuous, the coefficients of the wavelet series are obtained by integral convolutions of the signal

$$c_{j,k} = \int_{-\infty}^{\infty} f(t)\tilde{\varphi}_{j,k}(t)dt$$
$$d_{j,k} = \int_{-\infty}^{\infty} f(t)\tilde{\psi}_{j,k}(t)dt \qquad (4.2.5)$$

where $\tilde{\varphi}_{j,k}(x)$ and $\tilde{\psi}_{j,k}(x)$ are the dual functions of the scaling and wavelet functions $\varphi_{j,k}(x)$ and $\psi_{j,k}(x)$.

4.3 The effects of noise on the wavelet coefficients

Because of the presence of noise, the coefficients of the wavelet series representation of the noisy signal will also have a stochastic character. To see how noise is reflected by the wavelet coefficients consider the continuous noisy signal $\hat{f}(t)$ given by

$$\hat{f}(t) = f(t) + \varepsilon(t) \qquad (4.3.1)$$

where $f(t) \in L^2(\mathbb{R})$ and $\varepsilon(t)$ is the noise signal.

The noisy wavelet coefficients can be calculated theoretically in this case using the integral formula (4.2.5)

$$\hat{d}_{j,k} = \int_{-\infty}^{\infty} \hat{f}(t)\tilde{\psi}_{j,k}(t)dt = \int_{-\infty}^{\infty} f(t)\tilde{\psi}_{j,k}(t)dt + \int_{-\infty}^{\infty} \varepsilon(t)\tilde{\psi}_{j,k}(t)dt \qquad (4.3.2)$$

where $\tilde{\psi}_{j,k}$ is the dual of the wavelet function $\psi_{j,k}$.

Because $\tilde{\psi}_{j,k}$ is a square integrable function the integral

$$\int_{-\infty}^{\infty} \varepsilon(t)\tilde{\psi}_{j,k}dt \qquad (4.3.3)$$

is also convergent.

Considering $\varepsilon(t)$ to be white noise with $\mathbf{E}[\varepsilon(t)] = 0$, the expected value for $\hat{d}_{j,k}$ is given by

$$\mathbf{E}\left[\hat{d}_{j,k}\right] = \mathbf{E}\left[\int_{-\infty}^{\infty} f(t)\tilde{\psi}_{j,k}(t)dt\right] + \mathbf{E}\left[\int_{-\infty}^{\infty} \varepsilon(t)\tilde{\psi}_{j,k}(t)dt\right] =$$
$$= d_{j,k} + \int_{-\infty}^{\infty} \mathbf{E}[\varepsilon(t)]\tilde{\psi}_{j,k}(t)dt = d_{j,k} \tag{4.3.4}$$

which means that the wavelet coefficients are unbiased.

The error variance of the noisy wavelet coefficients can be calculated as

$$\sigma_{\hat{d}_{j,k}}{}^2 = \mathbf{E}\left[(\hat{d}_{j,k} - d_{j,k})^2\right] = \mathbf{E}\left[\left(\int_{-\infty}^{\infty} \varepsilon(t)\tilde{\psi}_{j,k}(t)dt\right)^2\right] =$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{E}[\varepsilon(u)\varepsilon(v)]\tilde{\psi}_{j,k}(u)\tilde{\psi}_{j,k}(v)dudv \tag{4.3.5}$$

But $\mathbf{E}[\varepsilon(u)\varepsilon(v)] = \gamma_{\varepsilon\varepsilon}(u-v) = \gamma_{\varepsilon\varepsilon}(\tau)$ where $\gamma_{\varepsilon\varepsilon}(\tau)$ is the autocorrelation function of the noise signal.

Since $\varepsilon(t)$ is white

$$\gamma_{\varepsilon\varepsilon}(\tau) = C \cdot \delta(\tau) \tag{4.3.6}$$

where $\delta(\tau)$ is the Dirac function. The expression of $\gamma_{\varepsilon\varepsilon}(\tau)$ in (4.3.6) can be substituted into (4.3.5) leading to the following relation

$$\sigma_{\hat{d}_{j,k}}{}^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \gamma_{\varepsilon\varepsilon}(\tau)\tilde{\psi}_{j,k}(\zeta)\tilde{\psi}(\zeta+\tau)d\tau d\xi = \int_{-\infty}^{\infty} C\delta(\tau)d\tau \int_{-\infty}^{\infty} \tilde{\psi}_{j,k}(\zeta)\tilde{\psi}(\zeta+\tau)d\xi =$$
$$= C\int_{-\infty}^{\infty} \delta(\tau)\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau)d\tau \tag{4.3.7}$$

with $\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau)$ the autocorrelation function of $\tilde{\psi}_{j,k}$.

Using the sifting property of the delta sequence given by

$$\int_{-\infty}^{\infty} \delta(t)h(t)dt = h(0) \tag{4.3.8}$$

the integral in (4.3.7) can be evaluated as

$$\sigma^2{}_{\hat{d}_{j,k}} = C\int_{-\infty}^{\infty} \delta(\tau)\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau)d\tau = C\cdot\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(0) \tag{4.3.9}$$

where $\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(0)$ is the maximum of the autocorrelation function $\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau)$.

Next we can state the following proposition

Proposition2: *The variance of the wavelet coefficients of a white noise signal* $\varepsilon(t)$ *is scale invariant*

To prove this it is sufficient to show that the maximum of the autocorrelation function of the dual wavelet function is scale invariant. Using the fact that $\tilde{\psi}_{j,k}$ is generated by translating and dilating the dual of the mother wavelet function, that is

$$\tilde{\psi}_{j,k}(t) = 2^{j/2}\tilde{\psi}(2^j t - k) \tag{4.3.10}$$

by a simple change of variable the autocorrelation function of $\tilde{\psi}_{j,k}$ can be expressed as

$$\begin{aligned}\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau) &= \int_{-\infty}^{\infty}\tilde{\psi}_{j,k}(\xi)\tilde{\psi}_{j,k}(\xi+\tau)d\xi = \int_{-\infty}^{\infty}\tilde{\psi}(\xi')\tilde{\psi}(\xi'+\tau')d\xi' = \\ &= \gamma_{\tilde{\psi}\tilde{\psi}}(\tau') = \gamma_{\tilde{\psi}\tilde{\psi}}(2^j\tau)\end{aligned} \tag{4.3.11}$$

that is $\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau) = \gamma_{\tilde{\psi}\tilde{\psi}}(2^j\tau)$. Consequently we can write $\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(0) = \gamma_{\tilde{\psi}\tilde{\psi}}(0)$ so the proposition is proven.

Because in practice the length of the signals are finite in time the infinite integrals in (4.3.5) are truncated which in turn means that the variance of the wavelet coefficients of a finite noise signal will in general decrease slightly with the scale.

4.4 Using mutual information to denoise wavelet coefficients

The results of the previous section showed how additive noise is reflected by the wavelet coefficients of a signal represented as a wavelet series. As a result, to reject the noise affecting a signal it is sufficient to remove or at least attenuate the stochastic part present in the wavelet coefficients. Such a procedure is equivalent to a smoothing operation performed over the noisy signal.

The advantage of reducing the noise by using the coefficients of the wavelet decomposition of the signal is that since the signal is represented in terms of smooth basis functions, by reducing the noise contribution to the wavelet coefficients the derivatives of the signal are smoothed in the same time. Another advantage comes from the fact that a multiresolution wavelet decomposition is equivalent in the frequency domain, to a multiband representation of the signal. This is because each of the orthogonal signals which result after decomposition have a frequency spectrum lying in different frequency bands. Thus by minimising the effects of noise on the wavelet coefficients the noise will be rejected at each frequency band.

The algorithm presented here makes use of the mutual information present in the signal to minimise the noise contribution to the wavelet coefficients.

The noisy signal is in practice a result of a data acquisition procedure. At this stage the continuous signal produced by a physical process has to be sampled with a sampling frequency f_s which satisfies the Whittaker-Shannon condition $f_s > 2f_h$ where f_h is the highest frequency present in the continuous signal considered to be band limited. In most cases the performance of data acquisition equipment can handle fairly high sampling rates which allows the user to oversample the continuous signal.

During the estimation stage the oversampled data is usually decimated or downsampled before a model is fitted in order to avoid ill-conditioned regression matrices so only a part of the measured data is actually used for identification.

Suppose that noisy data $\hat{y}(1),\ldots,\hat{y}(kn)$ are given from sampling a continuous dynamical signal with a sampling time δt so that the data is oversampled $f_s > 2kf_h$ where k is the oversampling factor.

It is assumed that the signal is corrupted by additive white noise

$$\hat{y}(i) = y(i) + \varepsilon(i) \tag{4.4.1}$$

The signal to noise ratio is defined as

$$SNR = 20\log_{10}\frac{\sigma^2_y}{\sigma^2_\varepsilon} \quad (\text{dB}) \tag{4.4.2}$$

At the first stage the initial data set is separated in k different subsets

$$\{\hat{y}(k(i-1)+p)\}_{i=1,n} = \{\hat{y}_p(i)\}_{i=1,n} \tag{4.4.3}$$

with p=1,...,k. This is done in fact by downsampling by k the original data sequence with k successive starting points $\hat{y}(1),\ldots,\hat{y}(k)$. As a result $\hat{y}_1(i),\ldots,\hat{y}_k(i)$, i=1,...,n represent k successive samples of the original data record stored in k different data sets. Because the sampling time $\delta t = 1/f_s$ is sufficiently small, k successive samples can be considered to have a linear variation since the Taylor series expansion of the signal around the central value within the k samples interval can be truncated to the first derivative. As a result the mean value of k successive samples of the noise-free signal is the central value of the interval that is the $(k+1)/2$-*th* sample for k odd. This can be expressed as

$$\frac{1}{k}\sum_{p=0}^{k-1} y_p(i) \cong y_{(p+1)/2}(i) \tag{4.4.4}$$

for any $i = 1,\ldots,n$.

Each of the k signals can be represented independently as a wavelet series following the procedure described in *4.2*.

Consider the noisy wavelet coefficients of the *p-th* signal at scale j to be denoted as $\{\hat{d}^p{}_{j,k}\}$ and consider the properties of the mean

$$\bar{\hat{d}}_{j,k} = \frac{1}{k}\sum_{p=1}^{k}\hat{d}^p{}_{j,k} = \frac{1}{k}\sum_{p=1}^{k} d^p{}_{j,k} + \frac{1}{k}\sum_{p=1}^{k} d\varepsilon^p{}_{j,k} \tag{4.4.5}$$

where $\{d^p{}_{j,k}\}$ are the noise-free wavelet coefficients and $\{d\varepsilon^p{}_{j,k}\}$ represents the stochastic part of the wavelet coefficients. Since the coefficients are the result of a linear transformation performed over the signal it follows that

$$\frac{1}{k}\sum_{p=1}^{k} d^p{}_{j,k} = d^{\frac{k+1}{2}}{}_{j,k} \tag{4.4.6}$$

for k odd.

As a result the expected value of the mean is given by

$$\mathbf{E}\left[\bar{\bar{d}}_{j,k}\right]=\mathbf{E}\left[d^{\frac{k+1}{2}}{}_{j,k}\right]+\frac{1}{k}\sum_{p=1}^{k}\mathbf{E}\left[d\varepsilon^{p}{}_{j,k}\right]=d^{\frac{k+1}{2}}{}_{j,k} \tag{4.4.7}$$

and this provides an unbiased estimate of the wavelet coefficients corresponding to the signal having the index $p=(k+1)/2$.

The variance of the same variable is given by

$$\mathbf{E}\left[\left(\bar{\bar{d}}_{j,k}-d^{\frac{k+1}{2}}{}_{j,k}\right)^{2}\right]=d^{\frac{k+1}{2}}{}_{j,k}+\frac{1}{k^{2}}\sum_{p=1}^{k}\mathbf{E}\left[(d\varepsilon^{p}{}_{j,k})^{2}\right]=d^{\frac{k+1}{2}}{}_{j,k}+\frac{\sigma^{2}{}_{d\varepsilon_{j,k}}}{k} \tag{4.4.8}$$

since the noise components affecting each of the k signals can be considered uncorrelated.

The results show that the effects of taking the mean, are that the standard deviation of the stochastic part of the wavelet coefficients is reduced $\sqrt{k}$ times. The overall effect is that the standard deviation of the noise affecting the signal is also reduced by the same amount The improvement of the signal to noise ratio can therefore expressed as

$$SNR_{new}=20\log_{10}k+SNR_{old} \tag{4.4.9}$$

Further improvements can be obtained by considering the fact that the noisy signal is usually a band limited signal so at high frequencies the energy of the signal is very small. In turn this leads to wavelet coefficients at higher scales which reflect just the noise. By discarding these coefficients the stochastic part of the signal is further reduced. This is equivalent to high frequency filtering with the advantage that the underlying signal is not affected.

5. Wavelet Smoothing of Noisy Chaotic Attractors

This section is devoted to a practical application of the results derived in the first part to reduce the observational noise affecting chaotic signals. To be able to compare the results the signals used were obtained from simulation and not from a practical experiment. The whole procedure can be summarised as follows

- The observed noisy signal is oversampled and then separated into k signals following the procedure described in (4.4.3).
- For each signal a wavelet decomposition is performed as in *4.2*
- Perform a high frequency smoothing by setting to zero the wavelet coefficients at higher scales which reflect high frequency components above the characteristic frequency band of the clean signal.

- The clean signal is reconstructed using the mean values of the wavelet coefficients corresponding to the k signals computed as in (4.4.5).

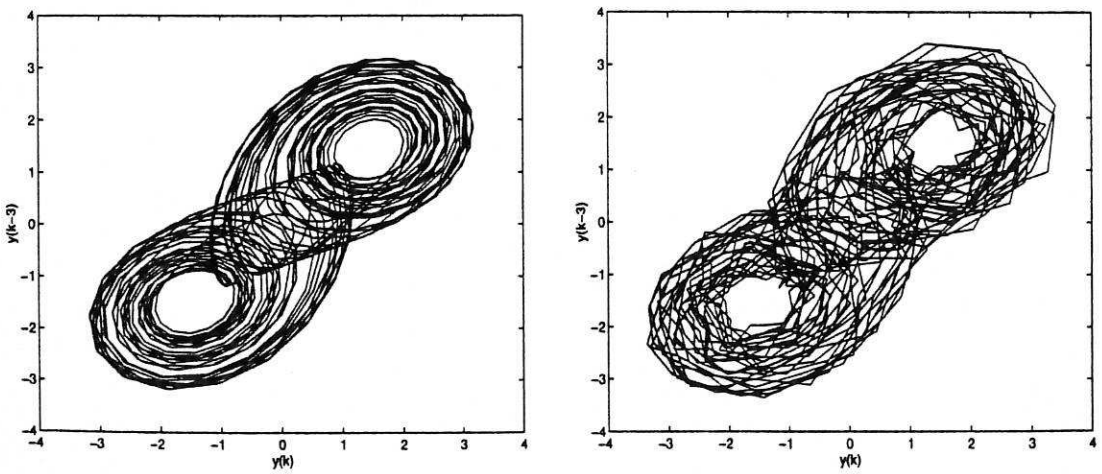
Using the smoothed signal a model is estimated and compared in terms of the characteristic invariants namely the correlation dimension and the largest Lyapunov exponent with the original system which produced the data.

5.1. Chua's circuit

The first example uses the well known chaotic system known as Chua's circuit. The system was intensively studied in the past (see for example Chua and Matsumoto(1986), Chua(1993) Matsumoto(1993)) due to the rich dynamical behaviour exhibited by this simple electronic circuit which can be represented by a system of three differential equations (2.1),(2.2)

To generate a data set the system was simulated using a fourth-order Runge Kutta algorithm with an integration step of Δt=1/400. As a result 90,000 measurements of the z component were obtained. Gaussian noise was added to the discrete data set leading to a signal to noise ratio SNR=42.7dB. Finally by downsampling by a factor of 60, as described in equation (4.4.3), only 1500 points sampled at $T_s = \Delta t \cdot 60 = 0.15s$ were retained for further processing. Instead of using all k=60 signals which were produced, only the first 9 signals were utilised.

The presence of noise can easily be distinguished by plotting the three dimensional attractor in the phase space using the classical embedding technique. Fig.2(b) clearly shows that the embedded trajectory is affected by noise.



(a) (b)

Fig. 2: Chua's double Scroll Attractor: (a) noise free and (b) superimposed with Gaussian noise

Initially each signal was represented as a series of the form of (4.2.1) using a least squares algorithm to calculate the scaling function coefficients.

The level of the initial approximation subspace V_j was chosen to be j=10 which leads to an approximation of the signal characterised by a RMS error level of $4 \cdot 10^{-15}$.

Next a pyramidal decomposition algorithm was applied using the scaling function coefficients at a resolution j=10. The decomposition was continued down to the resolution space j=2 leading to the following representation of the signals.

$$y(t_i) = \sum_k c_{2,k} \varphi_{2,k}(x) + \sum_{i=2}^{9} \sum_k d_{i,k} \psi_{i,k}(t_i) \qquad (5.1.1)$$

As a result, the original signals can be expressed as the sum of the ten orthogonal signals illustrated in Fig. 3.

The wavelet coefficients of the smoothed signal were calculated using (4.4.4) and the recovered signal was then used to reconstruct the chaotic attractor plotted in Fig.4(a). This compares well with the original in Fig.3(a) and shows that the noise has been removed from the signal.

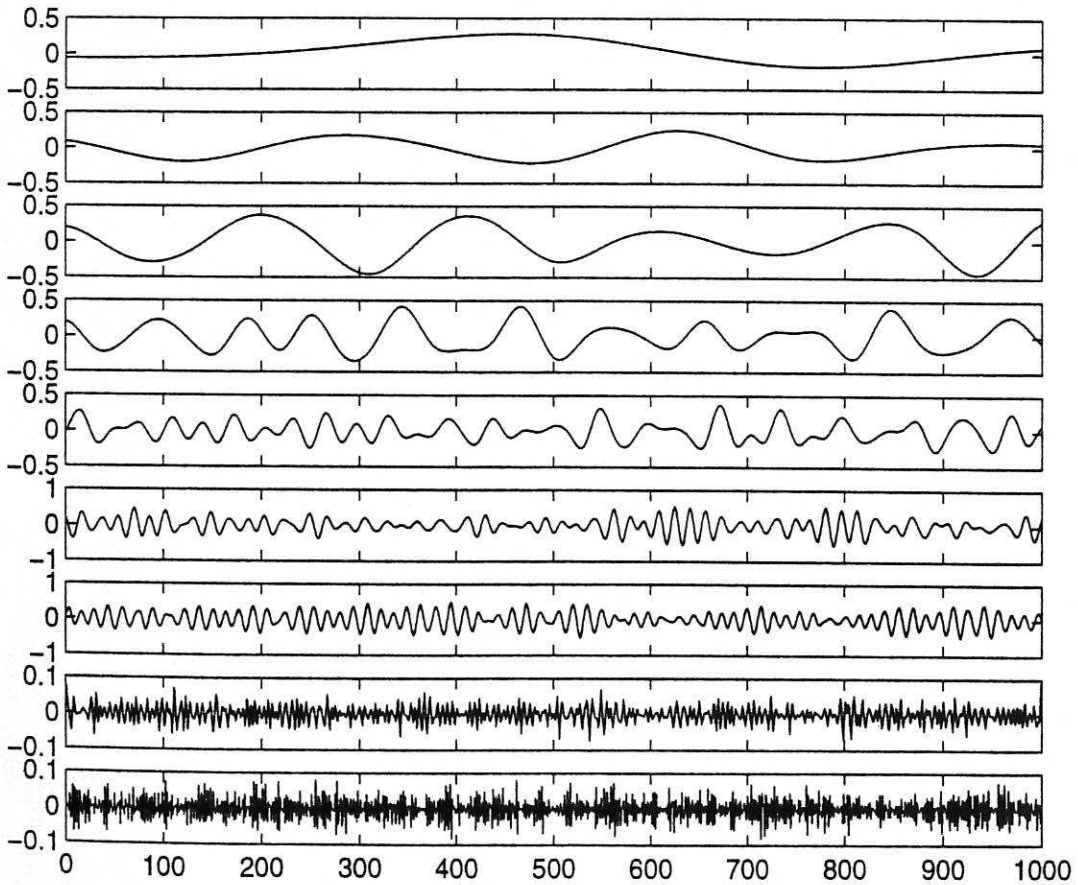


Fig. 3: Wavelet decomposition of the noisy chaotic signal starting with resolution level j=2.The projection onto the scaling subspace at resolution j=2 (top) can be added with the successive projections onto the wavelet subspaces j=2,..,9 (from top to bottom) to recover the original signal.

One thousand data samples of the smoothed signal were then used to estimate a nonlinear input/output model of the form

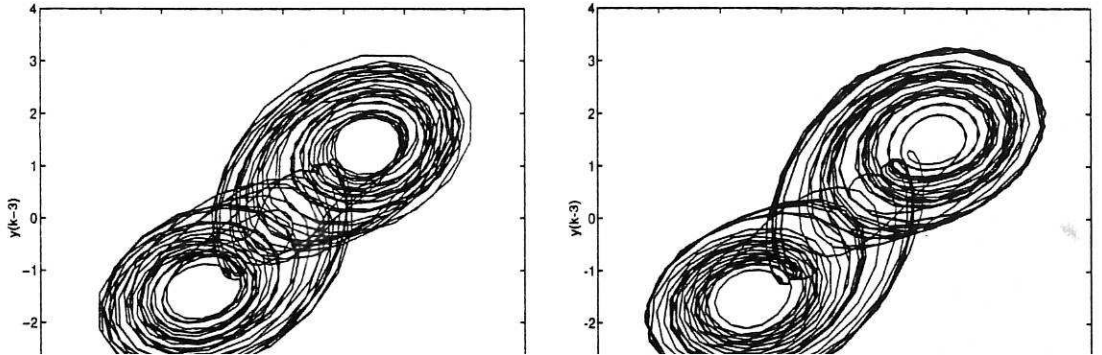
$$y(t+1) = F(y(t), y(t-1), y(t-2)) \tag{5.1.2}$$

Where $F: \mathbb{R}^3 \rightarrow \mathbb{R}$ was represented as a three dimensional multiresolution approximation of the form

$$F(y(t), y(t-1), y(t-2)) = \sum_{j,k} d_{j,k} \Phi_{j,k}(y(t), y(t-1), y(t-2)) \tag{5.1.3}$$

where the basis functions $\Phi_{j,k}$ are generated by taking the tensor product of the one dimensional B-spline scaling and wavelet functions. Since this expansion is linear in the parameters the value of the parameters were estimated using a simple least squares algorithm.

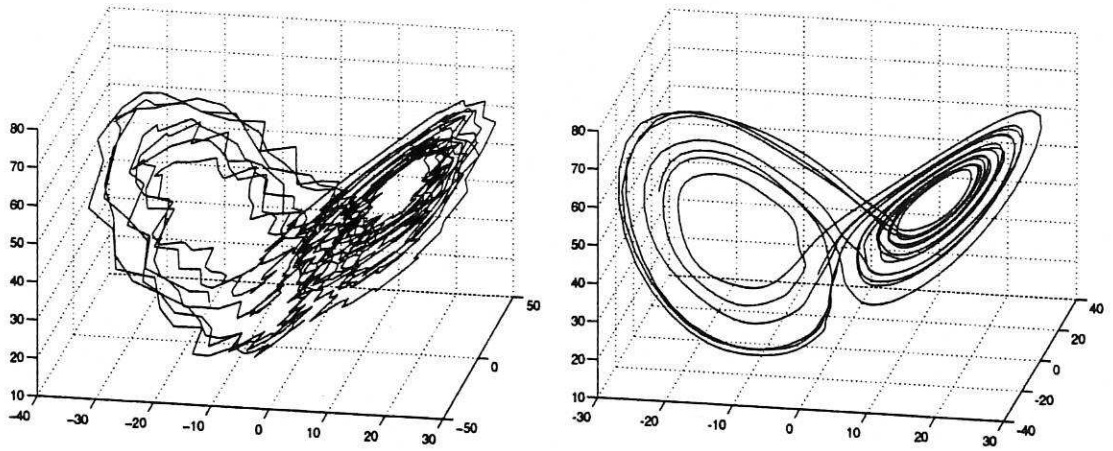
The estimated model was then used to generate a time series and the attractor reconstructed based upon this is illustrated in Fig.4(b).



$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = \rho x - y - xz \\ \dot{z} = xy - \beta z \end{cases} \tag{5.2.1}$$

Depending on the value of the control parameter, we can encounter stable and unstable fixed points, chaotic motion, bistability and hysteresis, coexistence of stable limit cycles and chaotic regions.

Again we chose one of the chaotic regimes to simulate and collect the data needed for estimation. A typical choice for the parameters is $\sigma = 16$, $\beta = 4$, $\rho = 45.92$ for which the system trajectory settles to a chaotic 'butterfly' shaped attractor. For these values of parameters, 15000 data points, for each coordinate x, y, and z, were simulated from the system using the Runge-Kutta algorithm with an integration step $\Delta t = 0.001$. At this stage white noise was added to each of the three time series so that the resulting signal to noise ratio was SNR=40dB. The three dimensional noisy chaotic attractor is plotted in Fig.5. Using equation (4.4.3) with k=10, each of the three time series, corresponding to x, y and z coordinates, was separated into k=10 data sets, each consisting of 1500 points having a sampling period $T=\Delta t \cdot 10=0.01$s.



(a) (b)

Fig. 5: Lorenz Attractor:(a) superimposed with Gaussian noise and (b) after smoothing

The smoothing algorithm was performed separately for each phase of the system as in the previous example using again only the first k=9 signals which were produced by downsampling. In this case the order of the basis functions used to represent the signals was m=4.

Using the pyramidal decomposition algorithm each signal was expressed alternatively as a sum of successive orthogonal approximations consisting of projections of the signals onto the wavelet subspaces W_i, i=2,...,10, together with the coarse approximation onto the subspace V_2. This is illustrated in Fig. 6, 7 and 8 which contain the decomposition of one out of nine signals for each phase x, y, and z produced by downsampling.

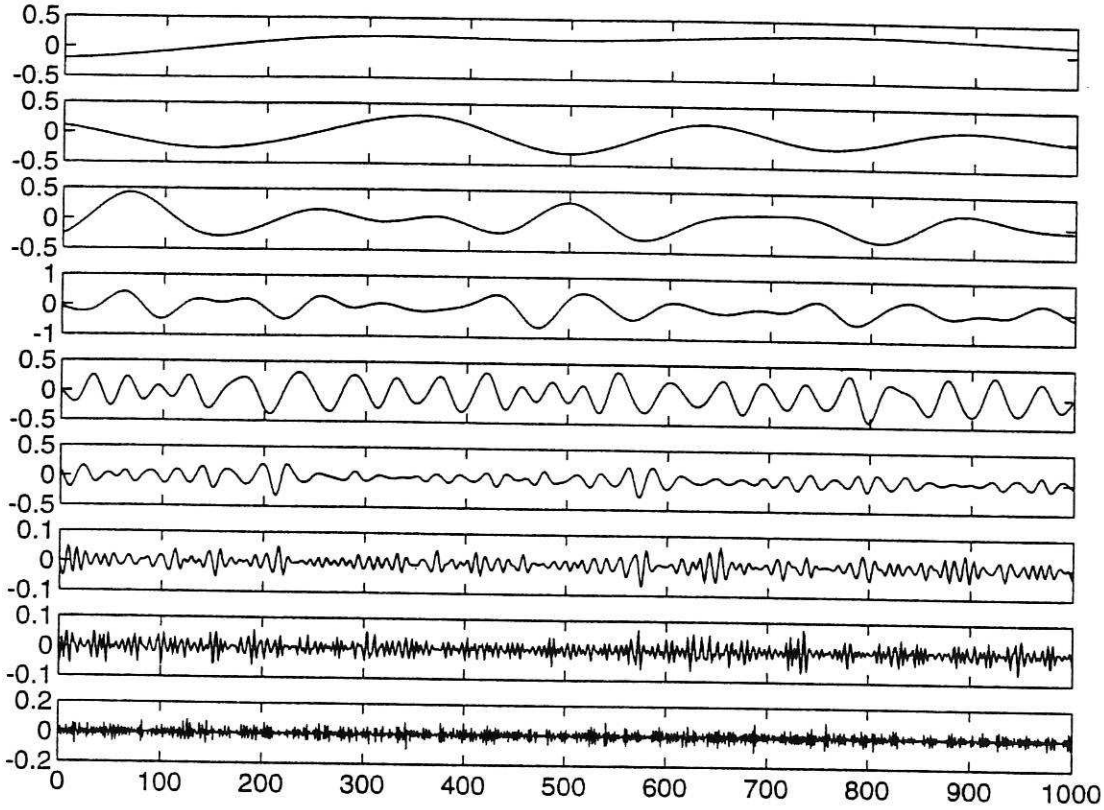


Fig. 6: Wavelet decomposition of the x component of the Lorenz system starting at resolution j=2 (top)

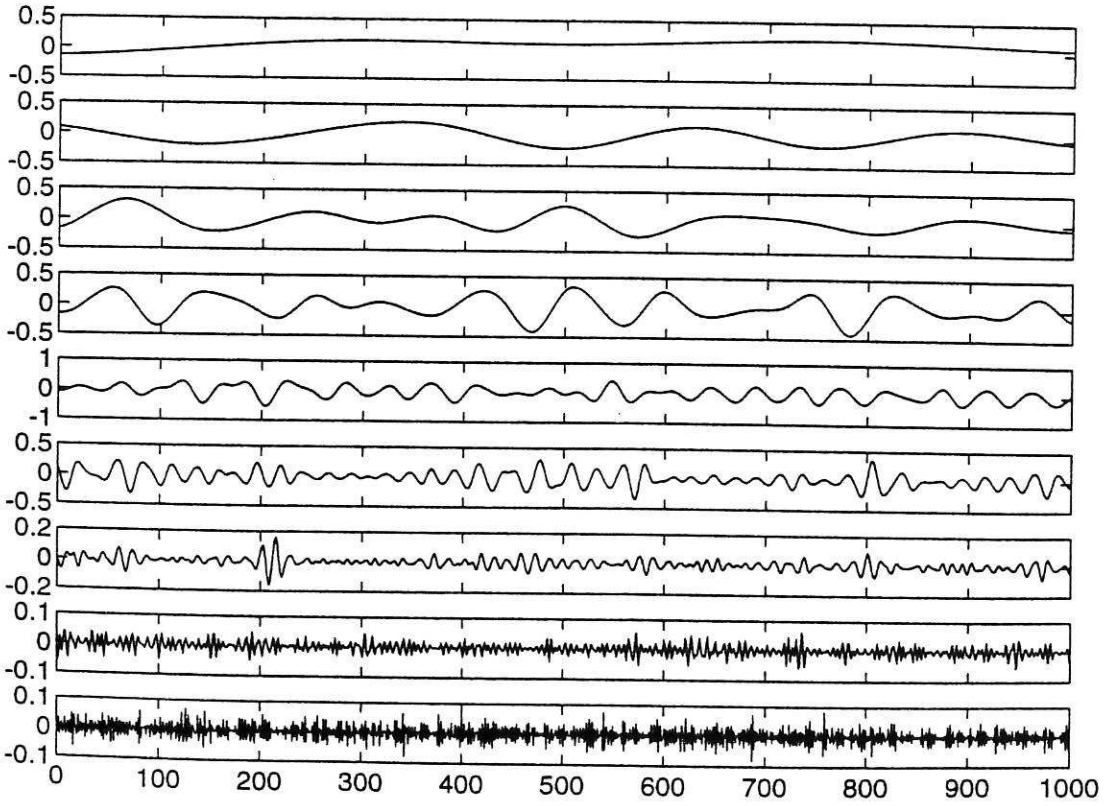


Fig. 7: Wavelet decomposition of the y component of the Lorenz system starting at resolution j=2 (top)

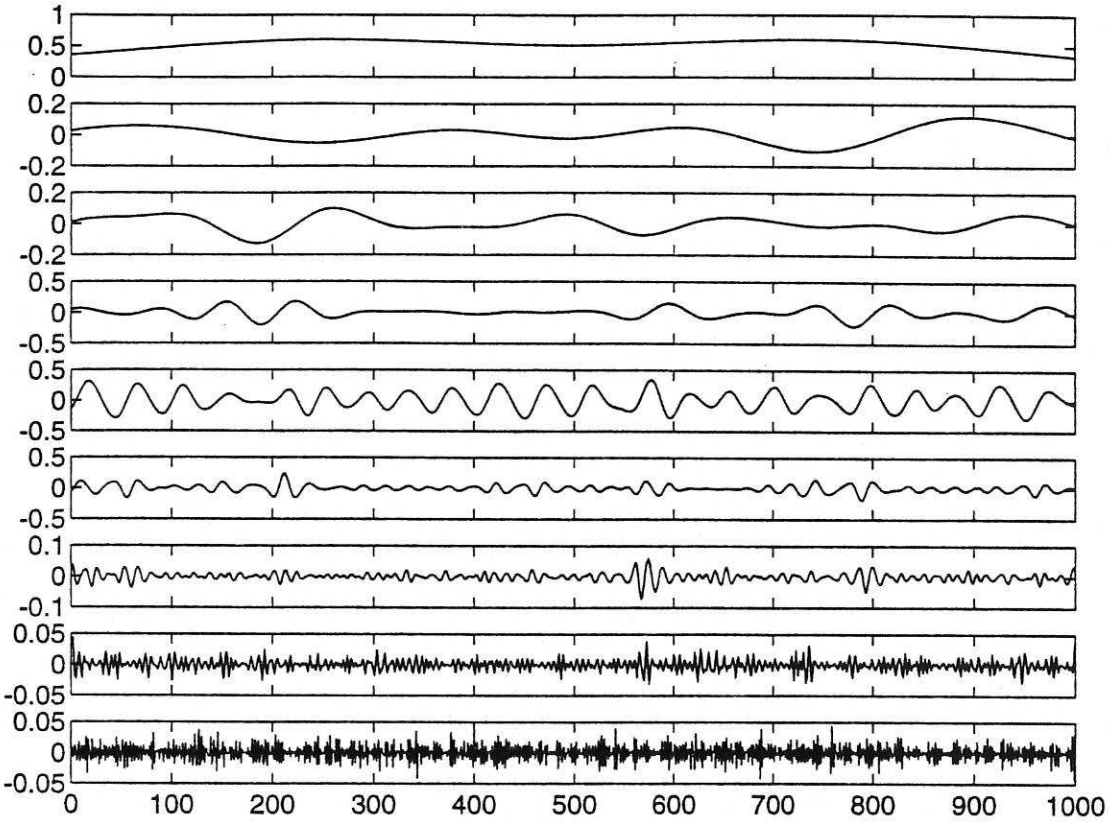


Fig. 8: Wavelet decomposition of the z component of the Lorenz system starting at resolution j=2 (top)

The new wavelet coefficients calculated as in (4.4.5) were used to reconstruct each phase of the chaotic system.

The smoothed versions of the signals show a major improvement in terms of noise reduction when plotted to generate the chaotic attractor illustrated in Fig.5(b).

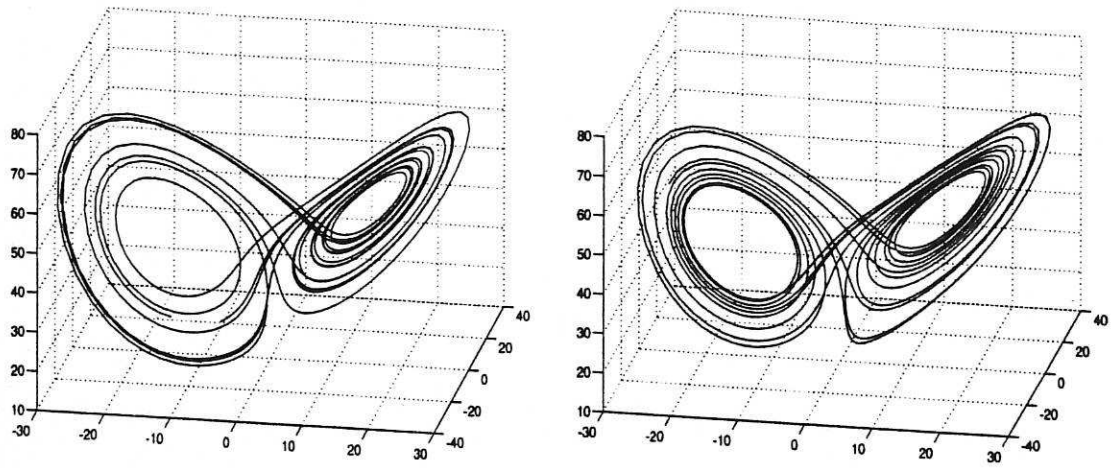
Using the smoothed signals a multi-input multi-output (MIMO) representation of the system was estimated using a wavelet approximation approach to give a model of the form

$$\begin{cases} x(t+1) = F^1(x(t), y(t), z(t)) = \sum_{j,k} d^1{}_{j,k} \Phi_{j,k}(x(t), y(t), z(t)) \\ y(t+1) = F^2(x(t), y(t), z(t)) = \sum_{j,k} d^2{}_{j,k} \Phi_{j,k}(x(t), y(t), z(t)) \\ z(t+1) = F^3(x(t), y(t), z(t)) = \sum_{j,k} d^3{}_{j,k} \Phi_{j,k}(x(t), y(t), z(t)) \end{cases} \tag{5.2.2}$$

with $F^k: \mathbb{R}^3 \to \mathbb{R}$, k=1,..,3.

Simulation of the estimated model gives three time series for the three components x, y and z which can then be plotted in state space. The resulting chaotic attractor compares very well to the original Fig.9(a) and this confirms that the

smoothing procedure has not affected the properties of the underlying chaotic signal Fig.9(b).



(a) (b)

Fig.9: Lorenz attractor:(a)original and (b) reconstructed using the model predicted output

Furthermore there is a good agreement between the largest Lyapunov exponent of the identified model, estimated to be $\lambda_+ = 1.51$ and that of the original system calculated to be $\lambda_+ = 1.5$. The correlation dimension of the model was also computed and found to be $D_{corr} \cong 2.028$ while the original system has $D_{corr} \cong 2.02$.

6. Conclusions

This paper has investigated the use of multiresolution wavelet decompositions, based on B-spline functions, for smoothing chaotic signals prior to identification. It is well known that the presence of noise poses serious problems not only in determining the characteristics of experimental chaotic signals but also when the signal is used to estimate a mathematical model of the process. The algorithm derived in this paper provides a new approach for rejecting observational noise on chaotic signals. The effectiveness of the new wavelet smoothing procedure has been tested in practice using two well known chaotic examples. The simulation results show that the models estimated using the smoothed data have the same dynamical characteristics as the original noise free systems demonstrating that the dynamical information contained in the initially noisy signals has been preserved.

7. Acknowledgements

DC gratefully acknowledges financial support from a scholarship awarded by Sheffield University. SAB gratefully acknowledge that part of this work was supported by ESPRC.

References

ADOMAITIS, R.A.,1990, FARBER, R.M., HUDSON, J.L., KEVREKIDIS, KUBE, M., LAPEDES, A.S., Application of neural nets to system identification and bifurcation analysis of real world experimental data, *Neural Networks: Biological computers or electronic brains*, pp. 87-97, Springer Verlag, Paris

AGUIRRE, L.A., BILLINGS, S.A.,1993, *Validating identified nonlinear models with chaotic dynamics.* Submitted for publication.

AGUIRRE, L.A., BILLINGS, S.A.,1995, Identification of models for chaotic systems from noisy data: Implications for performance and nonlinear filtering, *PhysicaD*, 85, pp.239-259

AGUIRRE, L.A., BILLINGS, S.A.,1994, *Retriving dynamical invariants from chaotic data using NARMAX models.* Submitted for publication

AGUIRRE, L.A., Mendes, E.M., BILLINGS, S.A.,1994, *Smoothing data with local insatabilities for the identification of chaotic systems.* Submitted for publication.

ANOSOV, D.V., ARNOLD, V.I., (editors), *Dynamical Systems (I), Encyclopaedia of Mathematical Sciences*, Vol 1, Springer Verlag:Berlin Heideberg

CASDAGLI, M.,1989, Nonlinear prediction of chaotic time series. *Physica D*, 35,pp.335-356

CHUA, L.O., Matsumoto, T., 1986, The double scroll family,, *IEICE Trans. Circuits Syst.* 33(11), pp.1072-1118

CHUA, L.O., 1993, Global unfolding of Chua's circuit, *IEICE Trans.on Fundamentals Electronics, Communications and Computer Sciences,* 76-A(5), pp.704-734

CHUI, C. K., LI, X., MHASKAR, 1994, Neural networks for localized approximation. *Mathematics of computing,* vol 63, No.208, pp 607-623

CHUI, C.K., 1992, *An introduction to wavelets*, Academic Press ,New York

COHEN, A., 1992, Biorthogonal wavelets, *Wavelets. A tutorial in Theory and applications*, Chui CK (editor), pp 123-152

CRUCHFIELD, J.P. and MCNAMARA, B.S., 1987, Equations of motions from a data series, *Complex Systems*, 1,417

DEAUBECHIES, I., 1993, Wavelet transforms and orthonormal wavelet bases. *Proceedings of Symposia in Applied Mathematics*, vol 47

DONOHO, D.L., 1992, *De-noising by soft-thresholding.* Technical Report, Department of Statistics, Stanford University

ELSNER,J.B., 1992, Predicting time series using Neural networks as a method of distinguishing chaos from noise. *J. Physics A:Math.Gen.*, 25, pp.843-850,1992

FRIEDMAN, J.H., 1979, A Tree-structured aproach to Non-parametric multiple regression. *Smoothing Techniques for Curve Estimation*, Gasser T., Rosenblatt, M. (editors), Springer-Verlag, New York

GOUPILLAUD, P., GROSSMAN, A., MORLET, J., 1984/85, Cycle octave and related transforms in seismic signal analysis, *Geoexploration*, 23, pp 85-102

GRASSBERGER, P., SCHREIBER, J., and SCHAFFRATH, C., 1991, Nonlinear time sequence analysis. *Int. J. Bifurcation and Chaos*, 1(3), pp.521-547

JUDITSKY, A., ZHANG, Q., DELYON, B., GORENNEC, P-Y., and BENEVISTE, A., 1994, *Wavelets in identification. Wavelets, Splines, Neurons, Fuzzies: How good for identification?*, Technical report No 849, IRISA

LEMARIE-RIEUSSET, P.G., 1993 Projection Operators in Multiresolution Analysis in Proceeedings of Symposia in Applied Mathematics, vol 47

MALLAT, S. G., 1989, A theory of multiresolution signal decomposition; the wavelet representation. *IEEE Pattern Anal. and Machine Intelligence*, vol 11, pp 674-693

MALLAT, S. G., 1989, Multiresolution approximations and wavelet orthogonal basis of $L^2(R)$. *Trans. Amer Math. Soc*, 315, pp 69-87

MATSUMOTO, T., KOMURO, M., KOKUBU, M., TOKUNAGA, R., 1993, *Bifurcations. Sights Sounds and Mathematics.* Springher Verlag:Tokyo

MAYNE, D.O., BROCKETT, B.W.,1973, *Geometric Methods in Systems Theory.* Boston, Mass:Reidel.

MEYER, Y., 1993, *Wavelets and operators*, Cambridge studies in advanced mathematics

NASON, G.P., 1994, *Wavelet regression by cross-validation*, Technical Report, Department of Mathematics, University of Bristol

PARKER, T.S., CHUA, L.O.,1989, *Practical numerical algorithms for chaotic systems.* Springer Verlag:Berlin

RONALD, A., DEVORE, JAWERTH, B., POPOV, V, 1992, Compression of wavelet decompositions. *American Journal of Mathematics*, 114, pp.737-785

SCHREIBER, J., GRASSBERGER, P., 1991, A simple noise reduction method for real data. *Phys. Lett.*, 160 A(5), pp.411-418

STRANG, G., 1989, Wavelets and dilation equation. A brief introduction, in *SIAM Review*, vol 31, No.4,pp 614-627

TAKENS, F., 1980, Detecting strange attractors in turbulence. *Dynamical systems and turbulence* , D.A. Rand and L.S. Young (editors), Lecture Notes in Mathematics, vol 898, Springer Verlag:Berlin, pp.366-381.

UNSER, M., ALDROUBI, A., 1992, Polynomial splines and wavelets.A signal processing perspective. *Wavelets. A tutorial in theory and applications*, Chui C.K. (editor), pp 91-122

VAUTARD, R., YIOU, P., GHIL, M., 1992, Singular spectrum analysis:A toolkit for short, noisy chaotic signals, *Physica D*, **58**, pp.95-126

WEINER, N.,1958, *Nonlinear Problems in Random Theory.* New York:Technology Press

WIGGINS, S ,1990, *Introduction to Applied Nonlinear Dynamical Systems and Chaos*, Springer Verlag:New York

WIGGINS, S. ,1980,*Global Bifurcation and chaos*, Springer Verlag:New York