Monograph:

Coca, D. and Billings, S.A. (1995) Estimating Derivatives from Noisy Data: A Wavelet Multiresolution Decomposition Approach. Research Report. ACSE Research Report 606 . Department of Automatic Control and Systems Engineering

Estimating Derivatives from Noisy Data: A Wavelet Multiresolution Decomposition Approach

D. Coca S.A. Billings

Department of Automatic Control and Systems Engineering,
University of Sheffield
Sheffield, S1 3JD,
UK

Research Report No:606
November 1995

Abstract

Wavelet decompositions provide an excellent tool for localised approximation of functions with any degree of regularity at different scales and with a desired accuracy. This representation allows operations like differentiation and integration to be replaced by simple algebraic operations performed over the wavelet coefficients. In this context the task of differentiating discrete noisy data can be performed more efficiently. The present study analyses from both a theoretical and practical point of view the problem of smoothing and differentiating experimental noisy signals within the framework of multiresolution decomposition. The algorithm proposed is tested on a numerical simulated example.

1 Introduction

Numerical differentiation is required in many fields of science and engineering such as mechanical engineering for the computation of velocity and acceleration from position data (Harrison and McMahon, 1993), in thermodynamics where in many situations a realistic thermodynamic model based on physical principles is not available (Nunhez et al., 1993), in processing signals of biomedical interest or in human motion analysis (Winter and Sidwall, 1974), (Pezzack et al., 1977).

Because in many situations direct measurements of the signal derivatives is very difficult, these values must usually be obtained using signal observations only. In most cases the observed data is corrupted by noise which is amplified by numerical differentiation making it very difficult to obtain good derivative estimates.

To obtain accurate high order derivatives several methods have been employed in the past. One approach makes use of frequency domain techniques to prefilter the data with a low-pass filter prior to a first order finite difference approximation which is performed to estimate the derivatives (Winter and Sidwall, 1974). This method however cannot remove the noise superimposed over the signal band and may, in some cases, alter the interesting features of the original signal such as local high frequency oscillations.

The signal derivatives can also be determined by interpolating the signal. Due to the fact that experimental data usually contains regions of sharply different behaviour the use of a simple polynomial representation is not satisfactory for the entire set of data. The use of the orthogonal Chebysev polynomials for example, was reported by Pezzack et al (1977) to produce an oversmoothed estimate of the second derivative. Splines perform better in this case and were used to smooth discrete noisy data in conjunction with a method to control the degree of smoothing. The optimum degree of smoothing is obtained by minimising a certain functional which depends on a parameter often referred to as the regularisation parameter. Such a method is described in Craven and Wahba (1979) and Hutchinson and Hoog(1985) where the critical part of the procedure namely choosing the parameter, is done with the help of a Generalised Cross-Validation technique.

Good results were reported by Fioretti and Jetto (1989) using a state space representation which allows the use of a fixed-lag Kalman smoother to obtain an optimal state estimate. The use of the smoother leads to good estimates provided that an accurate signal model is available and that a high signal-to-noise ratio is assumed.

The present work presents an improved method for the estimation of derivatives from experimental data, corrupted by additive white noise, using a multiresolution wavelet decomposition approach. In this context the paper evaluates first the effects of noise on the wavelet coefficients. Based on this, a new procedure to smooth the data is proposed, in which the first stage is performed by simply setting to zero the coefficients of the wavelet functions at higher scales. This is equivalent to a low-pass filtering operation with the advantage that because of the time frequency localisation of the basis functions, this can be done locally in time. At the second stage the noise superimposed over the signal band can be reduced using the mutual information between succesive data points. The new algorithm is performed over the wavelet coefficients taking advantage of the fast pyramidal decomposition and reconstruction algorithms of a multiresolution decomposition. The whole approach is derived with the use of a multiresolution decomposition constructed using B-spline basis functions which are well suited for smoothing purposes. In this case the derivatives can be estimated in the same efficient way by simply differencing the coefficients of the multiresolution approximation. Simulation results are included to demonstrate the performance of the algorithms.

2 Wavelets

Multiresolution decompositions have spurred a lot of interest in recent years in mathematics, in signal processing and numerical analysis.

A multiresolution decomposition of L^2 is a nested sequence of closed subspaces $...V_{-1} \subset V_0 \subset V_1 \subset V_2...$ of L^2 having an empty intersection and a dense union and satisfying the translation and scaling properties

$$\begin{gathered} f(x) \in V_0 \Leftrightarrow f(2x) \in V_1 \\ f(x) \in V_0 \Leftrightarrow f(x-k) \in V_0 \end{gathered} \tag{1}$$

for all $j, k \in \mathbb{Z}$

Any function f from L^2 can be approximated at resolution j by its orthogonal projection, denoted $P_j f$, on V_j.

The importance of the scaling property arises from the fact that a multiresolution approximation can be described by means of a single function ϕ and its translates and dilates. In this way at resolution j the projection P_j of a function f can be represented as a series which converges in L^2 norm,

$$P_j f = \sum_k c_{j,k} \phi_{j,k} \tag{2}$$

where $\phi_{j,k} = 2^{j/2} \phi(2^j x - k)$

When the multiresolution decomposition was first defined by Mallat and Meyer (see (Mallat, 1989) and (Meyer, 1993)) the scaling function $\phi(x)$ was orthonormal to all translates $\phi(x-k)$ thus leading to anorthonormal multiresolution decomposition. However the orthonormality condition imposes tight constrains on the scaling function and moreover this condition is not essential in many applications.

The wavelet subspaces W_j 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}

$$V_{j+1} = V_j \bigoplus W_j \tag{3}$$

where $\bigoplus$ denotes the orthogonal sum of subspaces. Each wavelet subspace W_j is generated by a single function $\psi(x)$ and its dilates and translates in the same way the scaling function $\phi(x)$ generates subspaces V_j.

In this way the projection $Q_j f$ of a function f on the wavelet subspace W_j has a series representation in terms of the dilates and translates of the wavelet function $\psi(x)$ as follows

$$Q_j f = \sum_k d_{j,k} \psi_{j,k} \tag{4}$$

Following the nested structure of subspaces V_j and the fact that W_j is the orthogonal complement of the subspace V_j with respect to V_{j+1} we can write

$$P_{j+1} f = P_j f + Q_j f \tag{5}$$

This gives the following series representation of a function in terms of scaling and wavelet functions

$$f(x) = \sum_k c_{j,k}\phi_{j,k}(x) + \sum_k \sum_{i \geq j} d_{i,k}\psi_{i,k}(x) \tag{6}$$

Both the scaling and wavelet function coefficients are given by

$$\begin{aligned} c_{j,k} &= \int_{-\infty}^{\infty} f(x)\tilde{\phi}_{j,k}(x)dx \\ d_{j,k} &= \int_{-\infty}^{\infty} f(x)\tilde{\psi}_{j,k}(x)dx \end{aligned} \tag{7}$$

with $\tilde{\phi}(x)$ and $\tilde{\psi}(x)$ known as the duals of the scaling and wavelet functions. There are just a few particular cases when the dual functions are known explicitly. In the orthonormal case for example the scaling and wavelet functions are identical with their duals.

Since $V_1 = V_0 \bigoplus W_0$, both the scaling and wavelet function can be expressed in terms of the scaling function at resolution $j = 1$

$$\begin{aligned} \phi(x) &= \sum_k p_k \phi(2x - k) \\ \psi(x) &= \sum_k q_k \phi(2x - k) \end{aligned} \tag{8}$$

These are the two scale relations of the scaling and wavelet function and $\{p_k\}_{k\in\mathbb{Z}}$, $\{q_k\}_{k\in\mathbb{Z}}$ are known as the two scale reconstruction sequences. Reciprocally any scaling function from V_1 can alternatively be written using the scaling and wavelet functions from V_0 and W_0 respectively as

$$\phi(2x - k) = \sum_l \{a_{k-2l}\phi(x - l) + b_{k-2l}\psi(x - l)\}, \quad k \in \mathbb{Z} \tag{9}$$

This is referred to as the decomposition relation with $\{a_k\}_{k\in\mathbb{Z}}$ and $\{b_k\}_{k\in\mathbb{Z}}$ the decomposition sequences. Based on eq (8) and (9) fast decomposition and reconstruction algorithms have been derived.

Consider the approximation of a function $f(x) \in L^2(\mathbb{R})$ by the orthogonal projection $P_j f$ on a finer approximation subspace V_j. This can be decomposed into a coarser approximation onto V_{j-1} together with the difference or detailed information between the succesive approximation levels $j-1$ and j which is represented by the orthogonal projection onto the wavelet subspace W_{j-1}. This can be written as

$$\sum_k c_{j,k}\phi_{j,k}(x) = \sum_k c_{j-1,k}\phi_{j-1,k}(x) + \sum_k d_{j-1,k}\psi_{j-1,k}(x) \tag{10}$$

The scaling and wavelet function coefficients $c_{j-1,k}$ and $d_{j-1,k}$ at resolution $j-1$ can be computed in an efficient way using the fast decomposition algorithm

$$\begin{cases} c_{j-1,k} = \sum_l a_{l-2k} c_{j,l} \\ \\ d_{j-1,k} = \sum_l b_{l-2k} c_{j,l} \end{cases} \tag{11}$$

derived from (9). These equations describe a moving average process involving the scaling coefficients $c_{j,k}$ at resolution j and the decomposition sequences $\{a_k\}_{k\in\mathbb{Z}}$ and $\{b_k\}_{k\in\mathbb{Z}}$. The resulting data sequence has to be subsequently downsampled or decimated by two by taking every second value in the sequence in order to obtain the desired coefficients.

This decomposition can be continued as long as desired using this time the coefficients $c_{j-1,k}$ to perform the moving average procedure.

The reconstruction algorithm is a consequence of the two scale relations (8) and is used to calculate the scaling function coefficients at the coarser level.

The computation involves the following moving average scheme

$$c_{j,l} = \sum_k [p_{l-2k} c_{j-1,k} + q_{l-2k} d_{j-1,k}] \tag{12}$$

this time with the weighting sequences $\{p_k\}_{k\in\mathbb{Z}}$, $\{q_k\}_{k\in\mathbb{Z}}$. In this case upsampling (inserting a zero between every two consecutive terms in the input sequence $\{c_{j-1,k}\}_{j,k\in\mathbb{Z}}$ and $\{d_{j-1,k}\}_{j,k\in\mathbb{Z}}$ is required before the moving average scheme is performed.

3 Smoothing

The problem of measurement is in most of the cases subject to errors. These errors are partly of a systematic and partly of an accidental nature. While the systematic errors are due to poor calibration of the instruments employed for taking the measurments, accidental errors arise from the fact that the parameters which characterise a certain physical process are subject to uncontrollable fluctuations which can normally be described by statistical laws. These accidental errors which occur during measurments are usually referred to as the noise on the measurements.

A model of a sampled signal affected by measurement errors is given by the equation

$$y_\epsilon(t_i) = y(t_i) + \epsilon(t_i) \tag{13}$$

where $y(t)$ is the function representing the signal of interest and $y_\epsilon(t_i)$ is the set of noisy observations obtained by uniformly sampling the analog signal. The random errors are represented by $\epsilon(t_i)$ and satisfy

$$\begin{gathered} E[\epsilon(t_i)] = 0 \\ E[\epsilon(t_i)\epsilon(t_j)] = C\delta_{i,j} \end{gathered} \tag{14}$$

where E denotes mathematical expectation and $\delta_{i,j}$ is the Kronecker symbol.

Considering the discrete sequence of equidistant noisy observations it is desired to reconstruct y and the associated first two derivatives. This can be done by first separating the stochastic part of the signal (the noise) from the deterministing signal. The application of such a procedure is designed to reduce the effects of observational errors over a data set and is usually referred to as a smoothing operation.

The fact that noise is a highly non-analytical phenomenon and that it has an unpredictable nature paradoxically gives a very good basis of separation from the deterministic signal. A separation method is proposed which makes use of a multiresolution decomposition of the signal. The smoothing procedure involves two stages; at the first stage a part of the noise is rejected using the assumption that the signal is bandlimited while the second stage exploits the mutual information between succesive samples of the signal.

3.1 The effect of noise on the wavelet coefficients

The discrete noisy signal can be approximated by a wavelet series at resolution j leading to the following expression

$$y_\epsilon(t_i) = \sum_k c_{j-p,k}\phi_{j-p,k}(t_i) + \sum_k \sum_{l=j-p}^{j} d_{l,k}\psi_{l,k}(t_i) \tag{15}$$

Although the noise signal does not belong to the space of square integrable functions having infinite energy, the right hand side series will converge in L^2 norm (Cambanis and Masry, 1994) since stationary random processes are square integrable over every finite interval.

For the case of an infinite length signal, assuming that the probability distribution function of the noise signal has compact support, the noise signal will have bounded amplitude variation and it is possible to calculate the noisy wavelet cofficients since the integral

$$\int_{-\infty}^{\infty} \epsilon(t)\tilde{\psi}_{j,k}(t)dt \tag{16}$$

is convergent. Of course the coresponding wavelet series will not converge in L^2 norm but in L^∞.

Because of the presence of noise the coefficients of the wavelet series representation of the noisy signal will also have a stochastic character. The noisy wavelet coefficients can be calculated theoretically in this case using the integral formula (7)

$$\hat{d}_{j,k} = \int_{-\infty}^{\infty} y_\epsilon(t)\tilde{\psi}_{j,k}(t)dt = \int_{-\infty}^{\infty} y(t)\tilde{\psi}_{j,k}(t)dt + \int_{-\infty}^{\infty} \epsilon(t)\tilde{\psi}_{j,k}(t)dt \tag{17}$$

Considering $\epsilon(t)$ to be white noise with $\mathbf{E}[\epsilon(t)] = 0$, the expected value of the wavelet coefficients $\hat{d}_{j,k}$ can be computed

$$\mathbf{E}\left[\hat{d}_{j,k}\right] = \mathbf{E}\left[\int_{-\infty}^{\infty} y(t)\tilde{\psi}_{j,k}(t)dt\right] + \mathbf{E}\left[\int_{-\infty}^{\infty} \epsilon(t)\tilde{\psi}_{j,k}(t)dt\right] = \tag{18}$$

$$= \quad d_{j,k} + \int_{-\infty}^{\infty} \mathbf{E}\left[\epsilon(t)\right] \tilde{\psi}_{j,k}(t)dt = d_{j,k}$$

and the result demonstrates that the stochastic coefficients are unbiased.

The error variance of the wavelet coefficiets can be calculated as

$$\begin{aligned} \sigma^2_{\hat{d}_{j,k}} &= \mathbf{E}\left[(\hat{d}_{j,k} - d_{j,k})^2\right] = \mathbf{E}\left[\left(\int_{-\infty}^{\infty} \epsilon(t)\tilde{\psi}_{j,k}(t)dt\right)^2\right] = \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathbf{E}\left[\epsilon(u)\epsilon(v)\right] \tilde{\psi}_{j,k}(u)\tilde{\psi}_{j,k}(v)du dv \end{aligned} \tag{19}$$

But $\mathbf{E}\left[\epsilon(u)\epsilon(v)\right] = \gamma_{\epsilon\epsilon}(u-v) = \gamma_{\epsilon\epsilon}(\tau)$ where $\gamma_{\epsilon\epsilon}(\tau)$ is the autocorrelation function of the noise signal. Since $\epsilon(t)$ is white

$$\gamma_{\epsilon\epsilon}(\tau) = C\delta(\tau) \tag{20}$$

where $\delta(\tau)$ is the Dirac function. The expression of $\gamma_{\epsilon\epsilon}(\tau)$ can be substituted into (19) leading to the following relation

$$\begin{aligned} \sigma^2_{\hat{d}_{j,k}} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \gamma_{\epsilon\epsilon}(\tau)\tilde{\psi}_{j,k}(\xi)\tilde{\psi}_{j,k}(\xi+\tau)d\xi d\tau \\ &= \int_{-\infty}^{\infty} C\delta(\tau)d\tau \int_{-\infty}^{\infty} \tilde{\psi}_{j,k}(\xi)\tilde{\psi}_{j,k}(\xi+\tau)d\xi \\ &= C\int_{-\infty}^{\infty} \delta(\tau)\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau)d\tau \end{aligned} \tag{21}$$

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 Dirac function, that is

$$\int_{-\infty}^{\infty} \delta(t)f(t)dt = f(0) \tag{22}$$

the integral in (21) becomes

$$\sigma^2_{\hat{d}_{j,k}} = C\int_{-\infty}^{\infty} \delta(\tau)\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau)d\tau = C\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(0) \tag{23}$$

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 the following proposition can be proved:

Proposition : *The variance of the wavelet coefficients of a noisy signal is scale invariant*

This is somehow to be expected since the Fourier spectrum of a noise signal has esentially a constant amplitude, that is the Fourier components will remain of essentially constant order of magnitude,with irregular fluctuations around this constant.

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

$$\tilde{\psi}_{j,k}(t) = 2^{j/2}\tilde{\psi}\left(2^j x - k\right) \tag{24}$$

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

$$\gamma_{\tilde{\psi}_{j,k}\tilde{\psi}_{j,k}}(\tau) = \int_{-\infty}^{\infty} \tilde{\psi}_{j,k}(\xi)\tilde{\psi}_{j,k}(\xi+\tau)d\xi =$$

$$= \int_{-\infty}^{\infty} \tilde{\psi}(\xi')\tilde{\psi}(\xi' + 2^j\tau)d\xi' = \gamma_{\tilde{\psi}\tilde{\psi}}(2^j\tau) \tag{25}$$

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

In practice, when finite length signals are involved, the infinite integrals (17) are truncated and the effect of these truncations will be more and more obvious the lower the scale, since the support of the wavelet functions widens in this case. Generally the effect will be a slight decrease with the scale of the variance of the noise coefficients. However for all practical purposes the variance can be considered as an invariant for a given range of scales.

3.2 High frequency noise reduction

The behaviour of the wavelet coefficients discussed above yields a very effective method for the separation of the true signal $y(t)$ from the observed signal which as in equation (13) is the true function plus noise.

If the noisy signal is approximated in terms of a wavelet series it can be observed that above a certain scale j the wavelet coefficients do not have any tendency to become smaller but remain within a certain band of fairly constant amplitude. These coefficients will account for the high frequency components which are purely caused by noise. This is because the deterministic signal is by definition a bandlimited signal so the high frequency components of such a signal above the cut-off frequency make a negligible contribution to the spectrum of the signal. This in turn means that the wavelet coefficients of the noise-free signal coresponding to that frequency band should be very small, nearly zero.

By simply setting to zero and thus omitting the corresponding wavelet functions from the series representation, the signal will be sensibly smoothed. Because the genuine signal components are not zero above the frequency we have chosen to define the noise band, it is expected that in this way some errors will be introduced in estimating the deterministing signal. However these errors will generally be very small considering the fact that at higher scales the contibution of the signal is almost zero.

Here it is important to note that the above procedure has a global character since it involves all the coefficients representing the signal at a given scale. This in turn means that some nice properties of the wavelet functions are not yet fully exploited. The most important is the time frequency localisation offered by a multiresolution decomposition. This is a result of the time-frequency localisation of the wavelet function which acts as a window function both in the time and frequency domain and provides a local Fourier analysis which takes place at every single scale.

This feature is especially useful when dealing with deterministic signals with a time-varying frequency behaviour, with local oscillations or discontinuities. In such cases wavelets, due to the compact support (or at least spatial localisation) have the ability to locally adapt to the features of the signal, the amplitude of the wavelet coefficients in this case reflecting the contribution of each function to the approximation. A local high frequency burst of the signal can in this way be identified and isolated from noise. In such cases instead of setting all the coefficients at the corresponding scale to zero, we can choose to preserve the high magnitude coefficients which represent a feature of the deterministic signal. However, if the multiresolution decomposition is not orthonormal, it is more correct to say that the corresponding wavelet functions are preserved since the respective coefficients have to be re-estimated under the additional constraint that all other coefficients at the same scale are zero.

After completing the above procedure the noise will not be eliminated completely since the remaining wavelet coefficients at a lower scale are also affected by the noise. Next a method to purify these coefficients is proposed which will further improve the accuracy of estimating the noise-free signal.

3.3 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 at 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.

Suppose that noisy data $y_\epsilon(t_1), ..., y_\epsilon(t_kn)$ are given from sampling an analog 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

$$y_\epsilon(t_i) = y(t_i) + \epsilon(t_i), \qquad i = 1, kn \tag{26}$$

The noise content of the signal can be described by the signal to noise ratio defined as

$$SNR = 20log\frac{\sigma_y^2}{\sigma_\epsilon^2} \tag{27}$$

or alternatively we can use the noise to signal ratio measure defined as

$$NSR_{\%} = \frac{\sigma_\epsilon}{\sigma_y} \times 100 \tag{28}$$

where σ_y and σ_ϵ are the standard deviation of the signal and noise respectively.

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

$$\left\{y_\epsilon\left(t_{k(i-1)+p}\right)\right\}_{i=1,n} = \{y_\epsilon^p(t_i)\}_{i=1,n} \tag{29}$$

with $p = 1, ..., k$. This is done in fact by downsampling by k the original data sequence with k successive starting points $y_\epsilon(t_1), ..., y_\epsilon(t_k)$. As a result $y_\epsilon^1(t_i), ..., y_\epsilon^k(t_i), \quad i = 1, ..., n$ 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 sample 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

$$\frac{1}{k}\sum_{p=1}^{k} y^p(t_i) \cong y^{(p+1)/2}(t_i) \tag{30}$$

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

Each of the k signals can be represented independently as a wavelet series (15).

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

$$\bar{\hat{d}}_{j,k} = \frac{1}{k}\sum_{p=1}^{k} \hat{d}^{p}_{j,k} = \frac{1}{k}\sum_{p=1}^{k} d^{p}_{j,k} + \frac{1}{k}\sum_{p=1}^{k} d\epsilon^{p}_{j,k} \tag{31}$$

where $\left\{d^{p}_{j,k}\right\}$ are the noise-free wavelet coefficients and $\left\{d\epsilon^{p}_{j,k}\right\}$ 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

$$\frac{1}{k}\sum_{p=1}^{k} d^{p}_{j,k} = d^{\frac{k+1}{2}}_{j,k} \tag{32}$$

for k odd. As a result the expected value of the mean is given by

$$\mathbf{E}\left[\bar{\hat{d}}_{j,k}\right] = \mathbf{E}\left[d^{\frac{k+1}{2}}_{j,k}\right] + \frac{1}{k}\sum_{p=1}^{k} \mathbf{E}\left[d\epsilon^{p}_{j,k}\right] = d^{\frac{k+1}{2}}_{j,k} \tag{33}$$

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

$$\mathbf{E}\left[\left(\bar{\hat{d}}_{j,k} - d^{\frac{k+1}{2}}\right)^{2}\right] = \frac{1}{k^{2}}\sum_{p=1}^{k} \mathbf{E}\left[\left(d\epsilon^{p}_{j,k}\right)^{2}\right] = \frac{\sigma^{2}_{d\epsilon_{j,k}}}{k} \tag{34}$$

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 variance deviation of the stochastic part of the wavelet coefficients is reduced k times. The overall effect is that the variance of the noise affecting the signal is also reduced by the same amount.

4 A B-Spline Multiresolution Approximation

4.1 Construction and properties of B-spline Multiresolution Approximations

For many applications of 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 $\phi(x)$ 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

$$\beta^m(x) = \int_{-\infty}^{\infty} \beta^{m-1}(x-t)\beta^1(t)dt \tag{35}$$

where $\beta^1(x)$ is the characteristic function of the interval (called also the indicator function) $\chi(x)$.

$$\beta^1(t) = \chi(t) = \begin{cases} 1 & \text{if } x \in (0,1) \\ 0 & \text{otherwise} \end{cases} \tag{36}$$

Relation (35) thus becomes

$$\beta^m(x) = \int_0^1 \beta^{m-1}(x-t)dt \tag{37}$$

The B-spline basis functions of higher order can be defined starting with the first order basis function (of degree zero) given explicitly in (36) by the following recursive algorithm

$$\beta^m(x) = \frac{x}{m-1}\beta^{m-1}(x) + \frac{m-x}{m-1}\beta^{m-1}(x-1) \tag{38}$$

Using this formula it is easy to see that the scaled and translated version of a B-spline function can be calculated using a similar recurrence relation

$$\beta^m_{j,k}(x) = \frac{2^j x - k}{m-1}\beta^{m-1}_{j,k}(x) + \frac{m+k-2^j x}{m-1}\beta^{m-1}_{j,k}(x-1) \tag{39}$$

where as usual $\beta^m_{j,k} = 2^{j/2}\beta^m(2^j x - k)$. The B-spline function β^m (of order m) consists in fact of m nontrivial polynomial pieces of degree $m-1$ so that

$$\beta^m \mid_{[k-1,k)} = P_{m-1,k}, \qquad k = 1, ..., m \tag{40}$$

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 $\phi(x) = \beta^m(x)$ as the subspace of all functions $f \in C^{m-2} \bigcap L^2(\mathbb{R})$such that the restriction of f to any interval $[k-1, k]$ 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.

A nonorthogonal multiresolution approximation is generated by an associated scaling function $\phi(x)$ which together with its translates forms a basis (frame) for the subspace V_0. In our case, for m a positive integer, the scaling function $\phi(x) = \beta^m(x)$ satisfies this condition and 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 wavelet subspace at resolution j. A B-spline wavelet basis can then be defined for any W_j in terms of translates and dilates of a mother wavelet function. This mother wavelet function can be constructed (Chui, 1992) as a linear combination of B-spline scaling functions. It can be shown that up to multiplication by a constant the compactly supported wavelets 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 $2m - 1$ and all wavelets are symmetric for m even and anti symmetric for m odd.

Because of the total positivity property of the B-spline scaling functions, the oscillations in a data sequence are diminished when convolved with a B-spline interpolation kernel. 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.

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 function widens with the scale while 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. It can be shown (Chui, 1992)

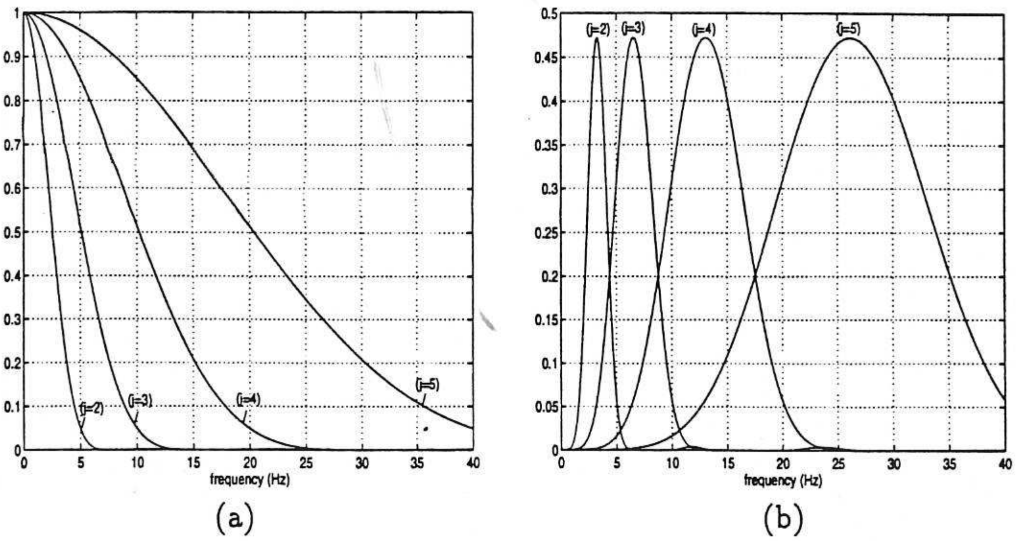

Figure 1: The Fourier transform of cubic B-spline scaling and wavelet function at different scales j function increases with the scale.

that both the B-spline scaling and wavelet functions satisfy this requirement.

4.2 Computation of the coefficients of a B-spline multiresolution decomposition

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

$$y(t_i) \cong \sum_k c_{j,k}\phi_{j,k}(t_i) \tag{41}$$

where $\phi_{j,k} = 2^{j/2}\phi(2^j t - k)$ and $\phi(t) = \beta^m(t)$ the m-th order cardinal B-spline function defined in (38). The right hand side series in equation (41) is in fact the projection of the signal onto the approximation subspace V_j.

To find the coefficients involved in the series (41), one way is to calculate the coefficient sequence corresponding to the scaling basis functions by convoluting the data sequence (moving average procedure) with an appropriate spline interpolation operator constructed as in (Chui, 1992). At this stage the resolution subspace V_j should be chosen so that the expansion (41) gives a sufficiently good approximation of the signal. For example for $h = 2^{-j}$ the approximation error is bounded by a constant multiple of h^m as $j \to \infty$. The coefficients of the series (41) can be also computed by using a least squares algorithm since this is a linear-in-the-parameters expression. This gives the orthogonal projection $P_j y$ of $y(t_i)$ 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, a standard practice, for example in image processing, is to extend the signal on both sides by using the mirror image of the signal. In our case it will be sufficient to consider the approximation of our signal on the interval $[0, 1]$ using the observations of $y(t)$ over a larger interval $[-N, 1 + N]$. Once the coefficients have been determined the signal will be approximated, with an accuracy depending on the chosen resolution subspace V_j, by the series (41). At this stage presuming that the signal is corrupted by noise, the coefficients of the series representation are also noisy.

A multiresolution pyramidal decomposition applied p times leads to the following equivalent expression of (41)

$$y(t_i) \cong \sum_k c_{j-p,k}\phi_{j-p,k}(t_i) + \sum_k \sum_{l=j-p}^{j} d_{l,k}\psi_{l,k}(t_i) \tag{42}$$

To perform such a decomposition the weight sequences $\{a_k\}$ and $\{b_k\}$ must be determined. In the case of B-spline multiresolution approximation the weighting sequences are infinite length sequences (Chui, 1992) which means that the moving average 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

$$G(z) = \frac{1}{2}\sum_{n=-\infty}^{\infty} g_n z^n = z^{-1}\left(\frac{1+z}{2}\right)^m \frac{E_{2m-1}(z)}{E_{2m-1(z^2)}}$$

$$H(z) = \frac{1}{2}\sum_{n=-\infty}^{\infty} h_n z^n = -z^{-1}\left(\frac{1-z}{2}\right)^m \frac{(2m-1)!}{E_{2m-1}(z^2)} \tag{43}$$

where the decomposition sequences are given by

$$a_n = \frac{1}{2} g_{-n}$$

$$b_n = \frac{1}{2} h_{-n} \tag{44}$$

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 coefficients $c_{j,k}$ in equation (41) at the finer resolution j.
- Following equation (9) 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 $c_{j-1,k}$. In this way at the p-th iteration the coefficients $c_{j-p+1,k}$ are used to compute $c_{j-p,k}$ and $d_{j-p,k}$.

After performing the smoothing operation over the wavelet coefficients, in order to calculate the signal derivatives, the signal should be expressed just in terms of the scaling functions at resolution j, as in (41), for ease of computation. The modified coefficients of the series (41) can be calculated using the pyramidal reconstruction algorithm, which this time involves the reconstruction sequences $\{p_k\}$ and $\{q_k\}$. The finite reconstruction sequences are given in this case by the following expressions

$$\begin{aligned} p_n &= 2^{-m+1}\binom{m}{n}, \qquad n = 0, m \\ q_n &= (-1)^n 2^{-m+1} \sum_{l=0}^{m} \binom{m}{l} \beta^{2m}(n+1-l), \qquad n = 0, 3m-2 \end{aligned} \tag{45}$$

which can be easily computed for any order m (Chui, 1992).

Considering that the decomposition procedure has been performed p times starting from the resolution level j the reconstruction algorithm can be summarised as follows

- Upsample with two both coefficient sequences $\{c_{j-p,k}\}$ and $\{d_{j-p,k}\}$ by inserting zeros between every two succesive values.
- Perform a moving average procedure as described by equation (8), involving the reconstruction sequences $\{p_k\}$ and $\{q_k\}$ to obtain the coefficients $\{c_{j-p+1,k}\}$.
- Repeat the first two steps using this time the newly calculated coefficients $\{c_{j-p+1,k}\}$ and the wavelet coefficients $\{d_{j-p+1}\}$ at scale $j-p+1$. This means that at the i-th iteration the coefficients used will be $\{c_{j-p+i,k}\}$ and $\{d_{j-p+i}\}$. The procedure should be repeated p times until the coefficients $\{c_{j,k}\}$ at resolution j are calculated.

4.3 Estimation of signal derivatives

The derivatives of the signal can be determined, after smoothing, by simply differentiating the reconstructed B-spline series. Following de Boor (de Boor, 1978) The derivatives of a B-spline series at scale $j = 0$ can be written as

$$D\left[\sum_k c_k \beta_k^m\right] = \sum_k (c_k - c_{k-1})\beta_k^{m-1} \tag{46}$$

This shows that the first derivative of a B-spline series can be found simply by differencing its B-spline coefficients, to obtain the coefficients of a B-spline series of one order lower which represents this derivative. Since formula (46) applies for infinite length signals, the derivatives of a finite B-spline series are obtained by making the series bi-infinite through adding B-spline functions with zero coefficients.

The formula to compute the derivatives of a B-spline series at any resolution j is as follows

$$D\left[\sum_k c_{j,k}\beta_{j,k}^m\right] = 2^j \sum_k (c_{j,k} - c_{j,k-1})\beta_{j,k}^{m-1} \tag{47}$$

and can be derived easily from (46). By repeating the application of (47), the following formula of the i-th derivative of a spline series can be written:

$$D^i\left[\sum_k c_{j,k}\beta_{j,k}^m\right] = \sum_k c_{j,k}^{i+1}\beta_{j,k}^{m-1} \tag{48}$$

where

$$c_{j,k}^{i+1} = \begin{cases} c_{j,k} & \text{for } i = 0 \\ \\ 2^{ij}\left(c_{j,k}^i - c_{j,k-1}^i\right) & \text{for } i > 0 \end{cases} \tag{49}$$

5 Simulation results

This section is devoted to a numerical simulation of the smoothing procedure described in the previous sections. In order to be able to evaluate in a quantitative manner the performance of the algorithm, the following expression

$$y(t) = \sin(\omega_0 t) + \sin(2\omega_0 t - \pi/6) + \cos(5\omega_0 t + \pi/4) \tag{50}$$

with $\omega_0 = 2\pi/T$ and $T = 1s$ (signal period) was used to generate a sequence of 10000 points with a sampling period $dt = 1/10000s$. A white noise sequence was generated and added to the original data to simulate the noisy observed signal. The noise variance was chosen in order that the signal-to-noise ratio of the resulting signal was $SNR = 40$dB that is equivalent to $NSR_{\%} = 10\%$. For this level of noise the resulting signal can be considered to be heavily contaminated by noise. In order to measure the effectiveness of the smoothing scheme the following quantity has been used

$$NSR^{(i)}_{\%} = \frac{1}{\sigma_{y^{(i)}}}\sqrt{\frac{1}{n}\sum_{k=1}^{n}\left((\overline{y}^{(i)}(t_k) - y^{(i)}(t_k)\right)^2} \tag{51}$$

which is the noise to signal ratio of $\overline{y}^{(i)}$, the i-th derivative of the estimated, smoothed signal, $i = 0, 1, 2$ where $\sigma_{y^{(i)}}$ is the standard deviation of the i-th derivative of the simulated signal.

To furher illustrate the efficiency of the smoother another measure, the noise reduction rate (NRR) described by the following relation

$$NRR = \sqrt{\frac{\sum_k (y_\epsilon(t_k) - y(t_k))^2}{\sum_k (\overline{y}(t_k) - y(t_k))^2}} \tag{52}$$

was used. All the above quantities can be used of course only in a simulation context since in practice the true signal would be unknown.

For the example chosen the whole smoothing procedure can be summarised as follows:

- The (oversampled) signal of interest is separated in $k = 9$ signals following the procedure described in (29)
- Each signal is approximated by a six order B-spline series at resolution $j = 10$
- For each signal a wavelet decomposition algorithm is performed $p = 8$ times
- 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 noise-free signal
- The signal is reconstructed using the mean values of the wavelet coefficients corresponding to the 9 signals computed as in (31).
- Compute the derivatives of the reconstructed B-spline series using relation (49)

One of the $k = 9$ signals resulting by downsampling the original signal is represented in Fig.(2) while the associated wavelet decomposition is illustrated in Fig.(3). The initial signal is the sum of the signals resulting after decomposition.

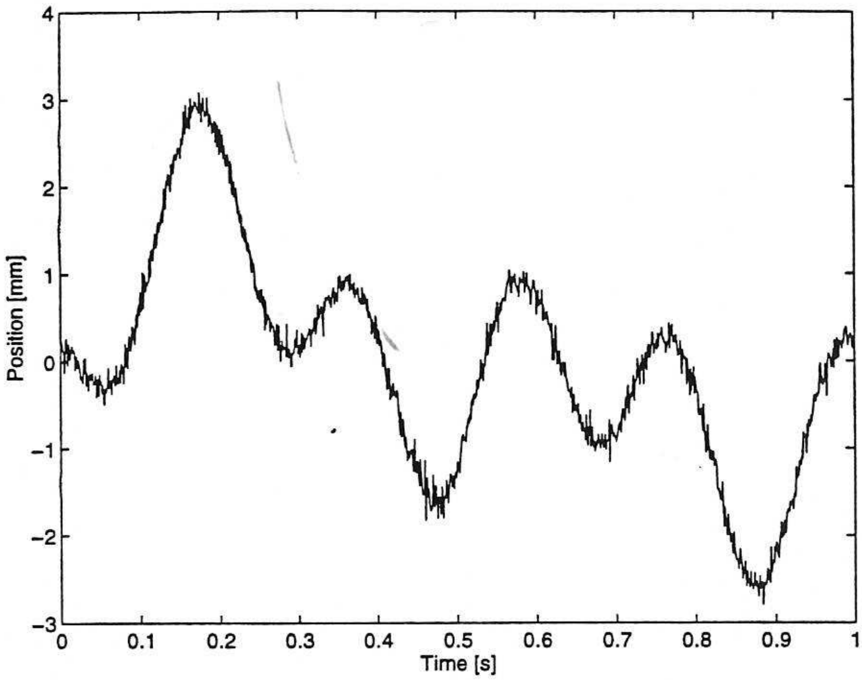

Figure 2: The 5-th noisy signal resulting from downsampling

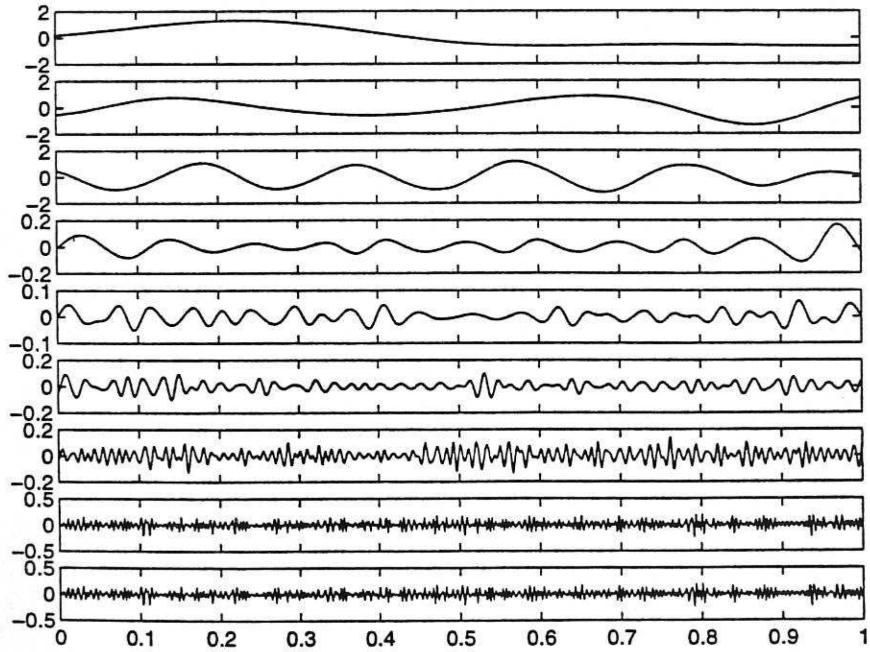

Figure 3: Wavelet decomposition

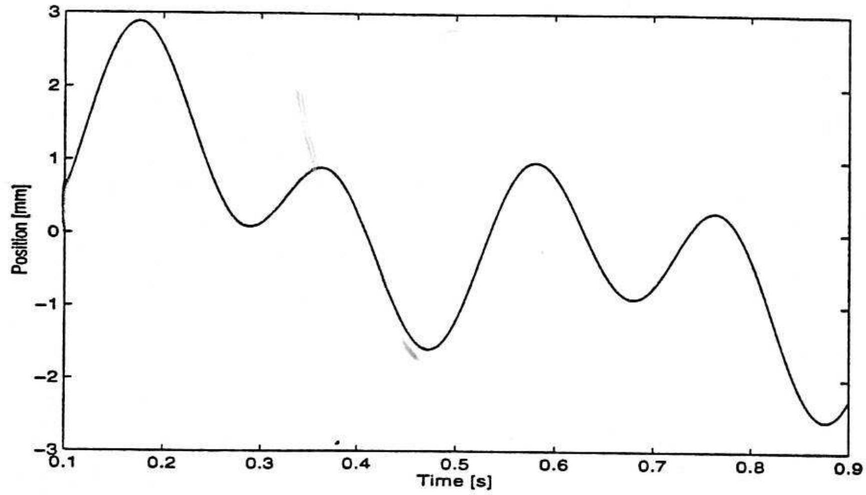

Figure 4: Smoothed position estimates (solid) and simulated (dashed)

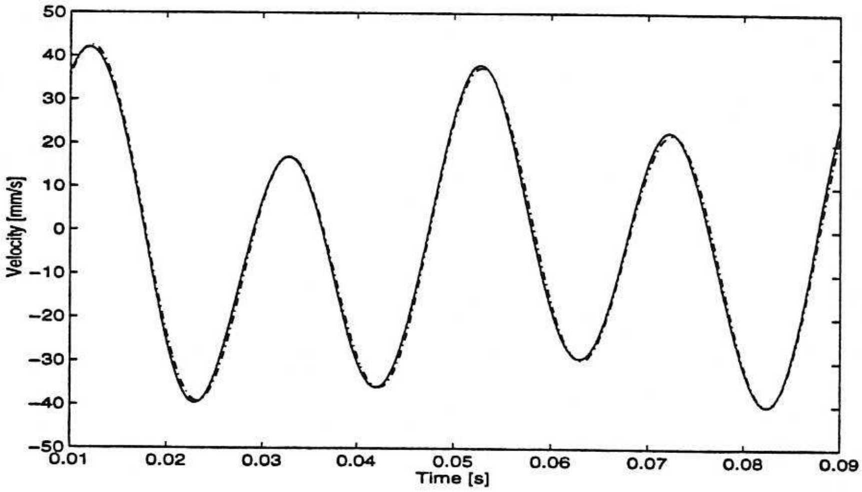

Figure 5: Smoothed velocity estimates (solid) and simulated (dashed)

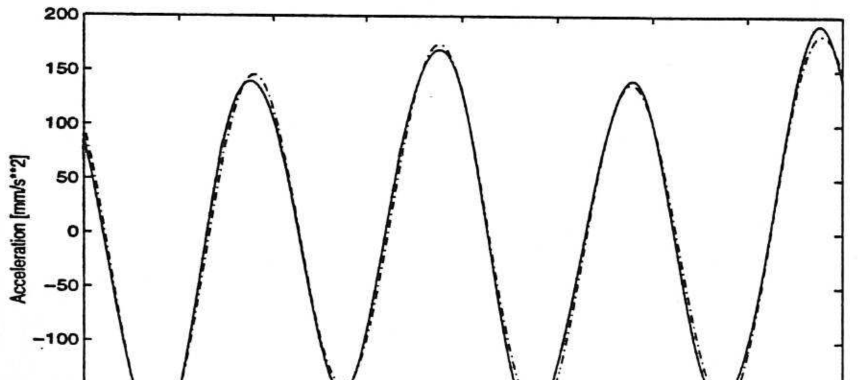

Figure 6: Smoothed acceleration estimates (solid) and simulated (dashed)

Figures (4) (5) and (6) represent the signal and the first and second derivatives together with the estimated (smoothed) signal and the corresponding derivatives. The plots clearly demonstrate excellent accuracy of the estimates.

Numerical values of the performance indexes defined are

$$\begin{aligned} NSR_{\%}^{(0)} &= 0.6045\% \quad \text{(position)} \\ NSR_{\%}^{(1)} &= 2.4400\% \quad \text{(velocity)} \\ NSR_{\%}^{(2)} &= 7.2070\% \quad \text{(acceleration)} \end{aligned} \tag{53}$$

The resulting noise reduction rate is $NRR = 16.5405$

5.1 Conclusions

In this paper, the concept of multiresolution approximation has been shown both theoretically and by example to be an effective method to solve the problem of smoothing and numerical differentiation of discrete noisy data. Using a new fast and efficient wavelet based algorithm, very good estimates of the derivatives are obtained in an example involving a heavily noise corrupted signal (10%).

5.2 Acknowledgements

SAB gratefully acknowledges that part of this work was supported by ESPRC. DC would express his thanks to the University of Sheffield for the award of a scholarship which made this study possible.

References

Akansu, A. (1992). *Multiresolution Signal Decomposition, Transforms, Subbands, and Wavelets.* Academic Press, London.

Cambanis, S. and Masry, E. (1994). Wavelet approximations of deterministic and random signals: convergence properties and rates. *IEEE Transactions On Information Theory*, 40(4):1013–1029.

Chui, C. (1992). *An introduction to wavelets.* Academic Press, New York.

Craven, P. and Wahba, G. (1989). Smoothing noisy data with spline functions. *Numerishe Mathematik*, 31:377–403.

de Boor, C. (1978). *A practical guide to spline.* Springer Verlag.

Fioretti, S.and Jetto, L. (1989). Accuate derivative estimation from noisy data:a state space approach. *Int. J. Systems Sci.*, 20(1):33–53.

Harrison, A. and McMahon, C. (1993). Estimation of acceleration from data with quantisation errors using central finite-difference methods. *Proc. Instn. Mech. Engrs; Part 1: Journal of Systems and Control Engineering*, 207(12):77–86.

Hutchinson, M.F.and Hoog, F. (1985). Smoothing noisy data with spline functions. *Numerishe Mathematik*, 47(1):99–106.

Mallat, S. G. (1989). A theory of multiresolution signal decomposition; the wavelet representation. *IEEE Pattern Anal. and Machine Intelligence*, 11:674–693.

Meyer, Y. (1993). *Wavelets and operators.* Cambridge studies in advanced mathematics. Cambridge Univ. Press.

Nunhez, J., Mori, M., and D'Avila, S. (1993). Fitting thermodynamic data using the modified spline technique. *Computers Chem. Engineering*, 17(11):1091–1099.

Pezzack, J., Norman, R., and Winter, D. (1977). An assessment of derivative determining techniques used for motion analysis. *Journal of Biomechanics*, 10.

Ronald, A., DeVore, Jawerth, B., and Popov, V. (1992). Compression of wavelet decompositions. *American Journal of Mathematics*, (114):737–785.

Unser, M. and Aldroubi, A. (1992). Polynomial splines and wavelets. a signal processing perspective. In Chui, C. K., editor, *Wavelets. A tutorial in theory and applications*, pages 91–122. Academic Press, New York.

Winter, D. and Sidwall, H. (1974). Measurement and reduction noise in kinematics of locomotion. *Journal of Biomechanics*, 7:157–159.