Stabilised wavelet transforms for non-equispaced data smoothing

Evelyne Vanraes*, Maarten Jansen†, Adhemar Bultheel*

July 2000

Abstract

This paper discusses wavelet thresholding in smoothing from non-equispaced, noisy data in one dimension. To deal with the irregularity of the grid we use so called second generation wavelets, based on the lifting scheme. The lifting scheme itself leads to a grid-adaptive wavelet transform. We explain that a good numerical condition is an absolute requisite for successful thresholding. If this condition is not satisfied the output signal can show an arbitrary bias. We examine the nature and origin of stability problems in second generation wavelet transforms. The investigation concentrates on lifting with interpolating prediction, but the conclusions are extendible. The stability problem is a cumulated effect of the three successive steps in a lifting scheme: split, predict and update. The paper proposes three ways to stabilise the second generation wavelet transform. The first is a change in update and reduces the influence of the previous steps. The second is a change in prediction and operates on the interval boundaries. The third is a change in splitting procedure and concentrates on the irregularity of the data points. Illustrations show that reconstruction from thresholded coefficients with this stabilised second generation wavelet transform leads to smooth and close fits.

Keywords: noise reduction, wavelet, lifting, adaptive, irregular meshes, stability, conditioning

1 Introduction

The classical wavelet based methods for data smoothing mostly assume the input to be a dyadic vector of equispaced, homoscedastic data. The wavelet basis functions, used in these methods, possess smoothness properties on regular, dyadic grids. When used for data on irregular point

*Department of Computer Science, K.U.Leuven, Belgium. The work of these authors is partially supported by the Fund for Scientific Research (FWO), project "MISS: Multiresolution on arbitrary grids via subdivision and Powell-Sabin splines", grant #G.0211.02 and the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with the authors. Corresponding author: evelyne.vanraes@cs.kuleuven.ac.be

†Department of Mathematics and Computer Science, T.U. Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

sets, *remapping* these basis functions to the actual grid, makes the irregularities show up in the output [1, 2].

Most existing wavelet based regression of non-equispaced data combines a traditional equispaced algorithm with a "translation" of the non-equispaced input into an equispaced problem. Possible techniques to do so are:

1. Interpolation in equidistant points [3, 4, 5]
2. Projection of the equispaced result onto the irregular grid [6, 7, 8, 9, 10, 11]. Some of these methods pay special attention to the approximation of the scaling basis and the projection coefficients therein.

This paper follows a different approach, based on so-called second generation wavelet transforms [12, 13]. Second generation wavelets extend the familiar concepts of multiresolution, sparsity, fast algorithms to data on irregular point sets. The key behind this extension is the lifting scheme [14]. Apart from a few publications [15, 16] that we know of, the use of second generation wavelets in statistical applications is quite new.

At the core of our noise reduction technique lies simple thresholding. The idea of thresholding is based on the concept of sparsity: the majority of wavelet coefficients is small, and can be replaced by zero. In the second generation setting, however, the transform may be unstable, i.e. 'far from orthogonal', no Riesz-basis is guaranteed. This turns out to be a challenging problem in applications of smoothing: the lack of orthogonality makes it hard to predict the effect of a threshold after reconstruction, and small coefficients may carry important information. Although the lifting scheme guarantees a *smooth* reconstruction, *closeness of fit* remains a problem, creating a considerable bias. Specifically for noise reduction, irregularity creates an additional complication: the noise in the wavelet domain becomes heteroscedastic (i.e. with fluctuating variance), even for homoscedastic input noise and even within each subband (resolution level). Correcting for this heteroscedasticity, though computationally feasible, may result in additional instability. Staying with the heteroscedastic noise on the other hand, leads to inferior fitting when using thresholds.

This paper proposes a novel scheme to deal with instability. Section 2 first puts the different components of our approach together. Section 3 explains why instability matters. Section 4 explores the mechanism behind the instability. This helps in finding and motivating the new scheme in Section 5: unlike classical lifting, the proposed scheme does not split the data into even and odd indexed samples. It rather takes the local grid structure into account. In this sense, the new scheme is more design adapted than classical lifting. Second, we reduce boundary problems by symmetric prediction rule near the signal end points. We conclude that section with some simulations on real data and conclude the paper with comments on the interesting results.

2 Our approach

2.1 Lifting

The input samples $s_{J,k}$ or f_k can be associated with a function f living in the space spanned by basis functions $\varphi_{J,k}$

$$f(x) = \sum_{k=1}^{2^J} s_{J,k}\varphi_{J,k}(x). \tag{1}$$

Here J stands for the highest resolution level used. The application of one step of the wavelet transform results in a low pass and a high pass signal. A wavelet decomposition algorithm has the structure of a repeated filter bank algorithm on the low pass output. Transforming from the highest resolution level J to the lowest resolution L gives the decomposition

$$f(x) = \sum_{k=1}^{2^L} s_{L,k}\varphi_{L,k}(x) + \sum_{j=L}^{J-1}\sum_{k=1}^{2^j} w_{j,k}\psi_{j,k}(x), \tag{2}$$

where $s_{L,k}$ are the scaling coefficients and $w_{j,k}$ are the wavelet coefficients. This is a multiresolution analysis (MRA).

The lifting scheme decomposes the filterbank operation in consecutive lifting steps [17]. The main difference with the classical construction is that is does not rely on the Fourier transform. All classical wavelet transforms can be implemented using the lifting scheme. The basic idea is simple. It starts by splitting the signal in points with an odd index and points with an even index. The lifting scheme then gradually builds the filterbank transform. The building blocks are lifting steps as shown in Figure 1.

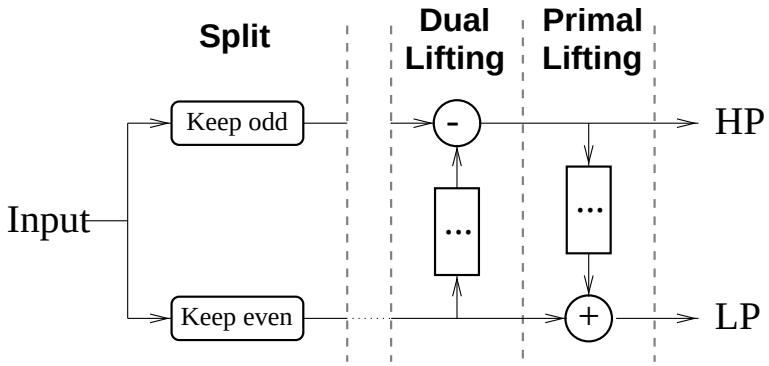


Figure 1: Decomposition of a filterbank into lifting steps. The first type of lifting is called *dual lifting* or *prediction*. The other type is *primal lifting* or *update*.

Dual lifting subtracts a filtered version of the even samples from the odd samples. The filter applied to the evens acts as a *prediction* of the odds. The prediction formula used in this paper is an interpolating polynomial, as proposed by Deslauriers and Dubuc [18, 19, 20]. Primal lifting, also known as *update* step, adds a filtered version of the dual lifting output to the so far untouched even samples.

2.2 Subdivision

Inverting a lifted transform is straightforward: run through the scheme backwards replacing plus with minus-signs, and merge what had been split [13]. Unlike the classical filterbank setup, the same filters appear in forward and inverse transform. Running through all scales of the inverse transform starting with all zeros except for one coefficient equal to one at a particular position, reveals the basis functions corresponding to the coefficient at that position. Indeed, inverse transform on a Kronecker sequence of wavelet coefficients $w_{j,k} = \delta_{ji}\delta_{kl}$ synthesises the function:

$$\sum_{k=1}^{2^L} 0 \cdot \varphi_{L,k} + \sum_{j=L}^{J-1} \sum_{k=1}^{2^j} \delta_{ji}\delta_{kl}\psi_{j,k} = \psi_{i,l}. \tag{3}$$

This procedure is known as subdivision. Applying this to the lifting decomposition of a wavelet transform, reveals the effect of the primal lifting step. Without this update step, the unique non-zero coefficient would flow unchanged and without any effect through the filter bank and arrive at the low pass branch of the (inverse) filter bank at the next (finer) scale. In other words, the wavelet function at scale j would simply coincide with the scaling function at the next, finer scale:

$$\psi_{j,k}^{[0]} = \varphi_{j+1,2k+1}. \tag{4}$$

Although in the forward transform the dual lifting step creates the detail or *wavelet coefficients*, it leaves the odd scaling basis functions untouched. The *background* (meaning, interpretation) of the detail coefficients *before* the update has taken place is still a scaling function. *After* the update step, this changes. Consider now the inverse transform including the update step. A two taps update filter with (possibly non-stationary) coefficients $A_{j,k}$, $B_{j,k}$ adds two non-zeros to the even branch, namely $-A_{j,k}$ and $-B_{j,k}$. The unique non-zero in the odd branch corresponds to the unlifted wavelet basis function, i.e. the odd, fine scaling function. This allows to write:

$$\psi_{j,k} = \psi_{j,k}^{[0]} - B_{j,k}\varphi_{j,k} - A_{j,k}\varphi_{j,k+1}. \tag{5}$$

The extension to longer update filters is obvious.

2.3 Thresholding

Thresholding, especially soft-thresholding [21] is a successful non-linear technique in wavelet based noise reduction of piecewise continuous signals. A central issue in this kind of smoothing procedures is how to find a suitable value for the smoothing parameter, in this case the threshold λ. This article opts for a minimum mean square error (MSE) approach. The expected MSE (also known as *risk*) combines two effects:

$$\text{Risk} \quad = \quad \text{bias}^2 \quad + \quad \text{variance}, \tag{6}$$

with:

$$\text{bias}^2(\lambda) \quad := \quad \frac{1}{N}\|\mathrm{E}\boldsymbol{w}_{\boldsymbol{\lambda}} - \boldsymbol{v}\|^2 \tag{7}$$

$$\text{variance}(\lambda) \quad := \quad \frac{1}{N}\mathrm{E}\|\boldsymbol{w}_{\boldsymbol{\lambda}} - \mathrm{E}\boldsymbol{w}_{\boldsymbol{\lambda}}\|^2. \tag{8}$$

In these equations, $\boldsymbol{w}$ stands for the vector of noisy wavelet coefficients and $\boldsymbol{w_\lambda}$ is the vector of thresholded wavelet coefficients. The vector $\boldsymbol{v}$ has the noise-free coefficients and N is the length of all these vectors. The variance stands for the noise: it decreases when the threshold grows. The bias on the other hand increases when the threshold grows. The minimum MSE threshold is the best trade-off between variance and bias in ℓ_2-norm sense.

In practical applications, the mean square error cannot be computed exactly, since the noise free data are unknown. Therefore, in our tests, we use the method of the Generalised Cross Validation (GCV) to estimate the minimum mean square error threshold [22].

2.4 Non-equidistant data

Lifting steps are by no means limited to equidistant data. Interpolating prediction, for instance, can trivially be extended to non-equispaced samples.

In the non-equispaced case, the lifting filters are no longer stationary. The standard deviation of the noise will be different for every wavelet coefficient even if the noise on the input had a constant standard deviation. Therefore we need a noise stationarity compensation: computing the noise covariance matrix S in the wavelet domain according to

$$S = \tilde{W} Q \tilde{W}^T, \tag{9}$$

where $\tilde{W}$ is the forward wavelet transform matrix, can be performed with linear complexity if the input correlation matrix Q is banded, for instance, if the input noise is uncorrelated. The matrix $\tilde{W}$ can be constructed by multiplying the filter matrices of each step. The entries follow directly from analysing the (time varying) impuls response of the lifted filter bank. The matrix Q is definied as $E\eta\eta^T$ where η is the vector with the noise on the input. Dividing each coefficient w_i by the corresponding diagonal element $\sqrt{S_{ii}}$ results in homoscedastic noise.

2.5 Example

Unlike Fourier-based procedures, wavelets perform particularly well in catching discontinuities in signals, but the traditional theory is directed to equispaced samples only. This paper tries to estimate piecewise continuous signals corrupted by noise, and sampled at non-equispaced locations. Figure 2(a) has an example of such a signal, the so called 'heavisine' test function

$$f(x) = 4\sin(4\pi x) - \text{sign}(x - 0.3) - \text{sign}(0.72 - x) \tag{10}$$

sampled on 2048 points. This test function is a piecewice continuous signal, a typical member of the class of signals for which wavelets are the optimal approach. The points were drawn from a uniform density on $[0, 1]$. To this signal, white and stationary noise was added with a standard deviation of $\sigma = 0.3$. Figure 2(b) shows the output from thresholding the 5 finest resolution level of a classical wavelet transform. The wavelet transform here was constructed by lifting with cubic prediction followed by 2-taps-update, but no grid structure was taken into account. Remapping the estimation to this grid makes the irregularity show up in the result. A much smoother result follows if the transform is based on the grid. Lifted wavelet transforms can be made grid-adaptive. The prediction operates on the real positions of each observation. Just as in classical threshold algorithms, coefficients are being thresholded and the output is obtained

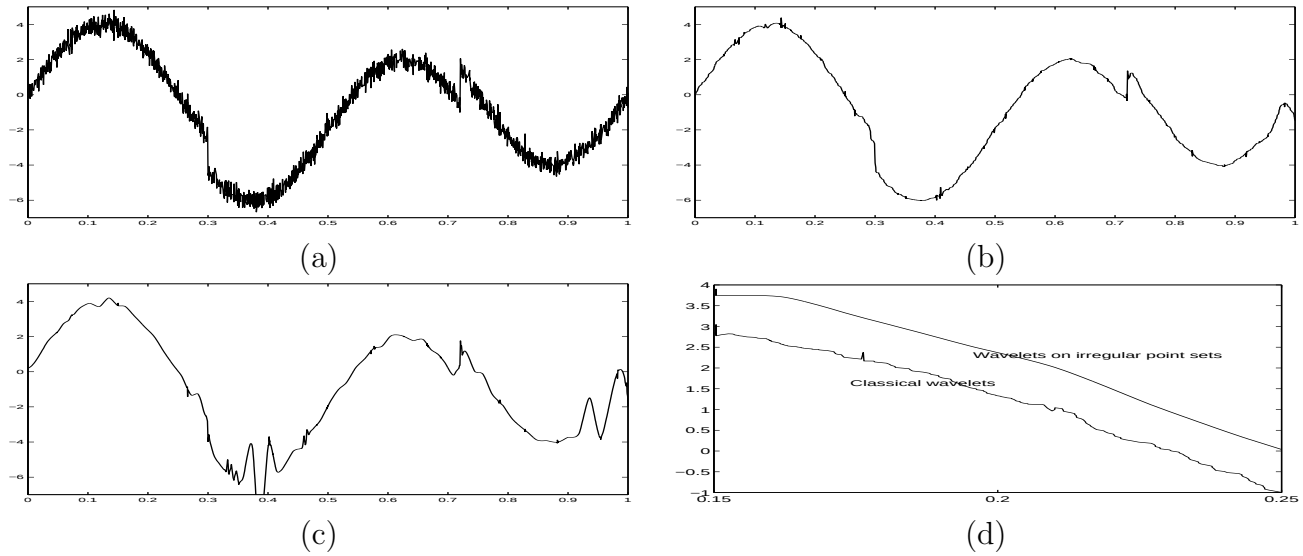


Figure 2: (a) A noisy version of the 'heavisine' test function. (b) Classical wavelets cannot take the grid structure into account. (c) Second generation wavelets, based on lifting, are smooth on the actual grid but lead to a tremendously biased reconstruction. (d) A detail of both reconstructions in a region without bias.

by inverse transform from these thresholded coefficients. From the detailed comparison of Figure 2(d) we see that the second generation approach is intrinsically superior. Figure 2(c) however shows that the reconstruction from these second generation wavelets contains strange artifacts, or — in statistical terminology — a strange bias.

3 The problem: instability

3.1 Numerical condition

A classical wavelet transform guarantees a norm semi-equivalence between the input and the wavelet coefficients: if $\boldsymbol{w}$ is the wavelet transform of $\boldsymbol{y}$, then the ℓ_2 norms of these vectors satisfy:

$$c \cdot \|\boldsymbol{w}\| \leq \|\boldsymbol{y}\| \leq C \cdot \|\boldsymbol{w}\|, \tag{11}$$

with $0 < c, C < \infty$ independent of the vector length. This relates to the concept of Riesz bases. Loosely speaking, a Riesz basis, also known as stable basis, is a basis in which the basis vectors or functions cannot be arbitrarily close to each other. This notion becomes important in vector spaces with infinite dimension, or, as in our case, when dealing with situations where the dimension is finite but arbitrarily large. The constants c and C are closely related to the condition number of the wavelet transform matrix.

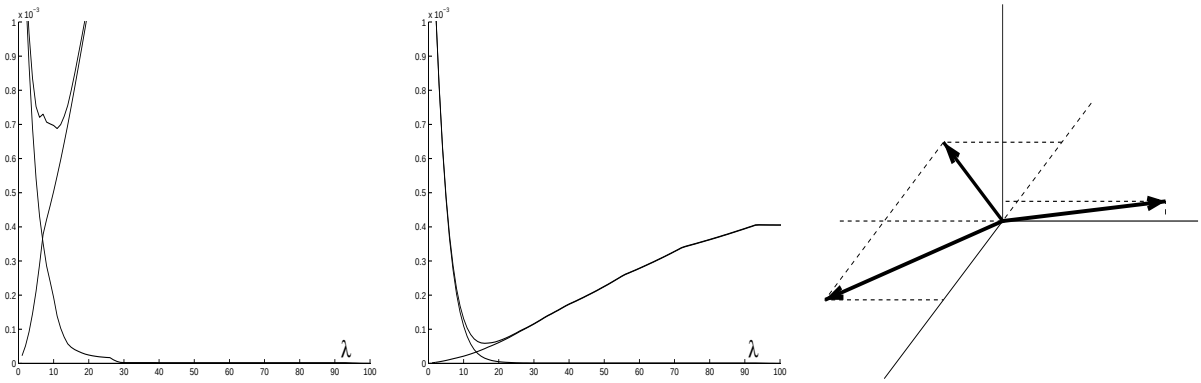
The extension to non-equispaced data through lifting gives no guarantee for the preservation of this comfortable Riesz basis background. As a matter of fact, Table 1 illustrates that condition numbers can be quite high. The condition number of the multiscale transform matrix $\tilde{W}$ is defined as $\kappa = \|\tilde{W}\|\|\tilde{W}\|^{-1}$.

$\tilde{n}$	equidistant points	random points
2	$1.98.10^2$	$3.25.10^1$
4	$6.11.10^1$	$1.48.10^4$
6	$1.78.10^2$	$6.02.10^3$
8	$1.51.10^3$	$1.22.10^5$

Table 1: Condition numbers of multiscale transforms on regular and irregular multilevel meshes for increasing number of dual vanishing moments ($\tilde{n}$)

3.2 Unpredictable effect

High condition numbers mean that a small modification of wavelet coefficient values may result in an unpredictable effect on the output. In the case of thresholding, this means that a small coefficient may carry substantial signal information. Since thresholding only works well on homoscedastic data (i.e. coefficients with constant noise variance), the wavelet coefficients have to be renormalised according to their variances. This makes the problem analysis even more complex.



(a) (b) (c)

Figure 3: (a) MSE plot in the signal domain. (b) MSE plot in the wavelet domain. The transform is unstable: the optimal threshold in the wavelet domain results in an unacceptable bias in the signal domain. (c) An arbitrarily unstable basis in $\mathbb{R}^3$.

Figure 3 illustrates this observation: it compares the MSE plot in the wavelet domain with the MSE plot in the signal domain. Whereas the MSE in the wavelet domain is smooth, small changes in threshold value may cause an important increase in error of the output in the signal domain. The threshold minimising this output error is also smaller than the minimum MSE threshold in the wavelet domain. This is because the bias increases faster in the signal domain. This small threshold is not really able to remove all the noise.

Not only do some individual coefficients have a wide impact, the interaction between the coefficients may be unpredictable, due to the fact that the transform is far from orthogonal. A simple example in $\mathbb{R}^3$ makes clear what happens. Suppose we have the basis vectors $\{(-1/2, \sqrt{3}/2, 0), (-1/2, -\sqrt{3}/2, 0), (1, 0, \varepsilon)\}$, as in Figure 3(c). If ε is small, this basis has an extremely bad condition. Suppose the noise is $(0, 0, \varepsilon)$ in the canonical basis, then its

coordinates in this oblique basis are $(1, 1, 1)$. If one or two of these coordinates are thresholded, "hidden components" become clear. This bad condition can only be detected with a global analysis: none of the basis vectors is close to another one. In the example of Figure 2, the noise added large coefficients to the wavelet representation. In the unthresholded set of coefficients, the effect of one coefficient is cancelled by a combination of other wavelet coefficients. We could state that the noise does not fit well into the oblique basis, thereby causing these large, mutually annihilating coefficients. Removing a part of these coefficients uncovers these hidden components.

4 The mechanism behind the instability

A more quantitative analysis of the instability problem follows from considering the lifting steps throughout. It turns out that the instability gradually builds up in the subsequent filter stages, culminating in the last, update step.

4.1 Large update coefficients

The expression for the lifted wavelet function in Section 2.2 shows that if the update coefficients are large, for instance, if:

$$B_{i,j} \gg \frac{\|\psi_{j,k}^{[0]}\|}{\|\varphi_{j,k}\|}, \tag{12}$$

the lifted wavelet $\psi_{j,k}$ nearly falls within the vector space spanned by its neighbouring scaling functions at the same scale. This creates a detail space which is far from orthogonal to the coarse scaling space. Large update coefficients result in a considerable overlap of scaling and wavelet function at a given scale. When the scaling functions are further decomposed into a wavelet basis at coarser scales, the immediate correlation between basis functions becomes hidden. The bad condition is then hard to detect in advance and hard to localise.

4.2 Prediction determines update coefficients

The classical implementation of the lifting scheme finds update filters such that they meet conditions of vanishing moments. With a two taps update filter, for instance, we can impose the primal wavelets to have two vanishing moments. The update coefficients are then:

$$A_{j,k} = \frac{M_{j+1,2k+1}}{M_{j,k}} \frac{\overline{x}_{j,k+1} - \overline{x}_{j+1,2k+1}}{\overline{x}_{j,k+1} - \overline{x}_{j,k}} \tag{13}$$

$$B_{j,k} = \frac{M_{j+1,2k+1}}{M_{j,k+1}} \frac{\overline{x}_{j+1,2k+1} - \overline{x}_{j,k}}{\overline{x}_{j,k+1} - \overline{x}_{j,k}}. \tag{14}$$

In these expressions, $M_{j,k}$ stands for the scaling function integrals:

$$M_{j,k} = \int_{-\infty}^{\infty} \varphi_{j,k}(x) dx, \tag{15}$$

and $\overline{x}_{j,k}$ stands for the first normalised moment:

$$\overline{x}_{j,k} = \frac{\int_{-\infty}^{\infty} x\varphi_{j,k}(x)dx}{\int_{-\infty}^{\infty} \varphi_{j,k}(x)dx}. \tag{16}$$

Through these expressions, the update coefficients depend on the primal scaling functions. These basis functions in their turn are determined by the prediction operator. Hence, the prediction has an influence on the stability of the transform.

An interesting example is a lifting scheme with linear interpolating prediction, followed by a 2-taps update. On an equidistant data set, this scheme reduces to the well known biorthogonal wavelet basis of Cohen, Daubechies, and Feauveau [23] with two dual and two primal vanishing moments (CDF 2,2). Subdivision shows that in this simple case the scaling functions are always positive. The L_1-normalised scaling functions all have integral one. On the other hand, the values of $\overline{x}_{j,k}$ can be interpreted as the mean value of the corresponding scaling function. Taking that into account, it is easy to check that $\overline{x}_{j,k} < \overline{x}_{j+1,2k+1} < \overline{x}_{j,k+1}$i. As a consequence, $0 < A_{j,k} < 1$ (and similarly for $B_{j,k}$), and the transform is stable on any grid.

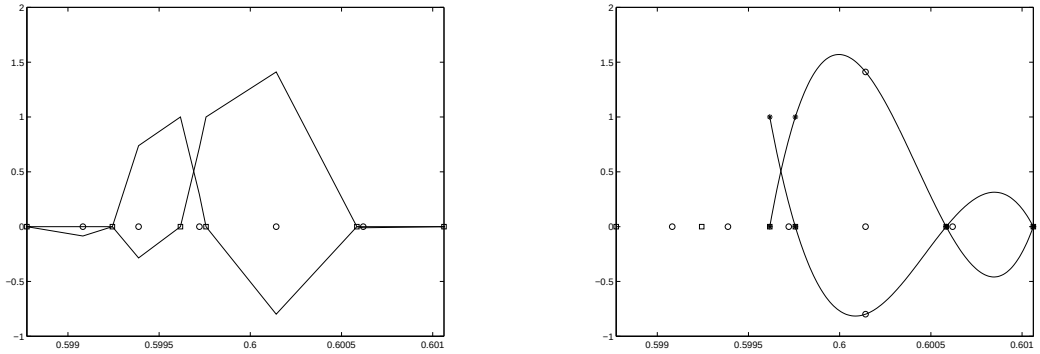
4.3 Splitting causes prediction to mix scales

If we use interpolating prediction with polynomials of higher order than linear, the scaling functions may show unwanted features. Mixing of scales creates scaling functions with heavy side blobs. Figure 4(a) has plots of two adjacent scaling functions after one subdivision step. 'Even' points at this scale are marked with a box, while the 'odd' points appear as circles. The first scaling function shows a heavy side blob, resulting in its integral being negative. The second scaling function does not have its maximum in its central point (i.e. the even point from which the subdivision started), which causes an unexpected move of its balancing point. Both phenomena persist in subsequent subdivision steps. Update steps using combinations with this kind of scaling functions are likely to use high update coefficients in orde to get wavelet functions with a given number of vanishing moments.

A closer look to the grid points reveals that the same 'odd' point creates both phenomena: the gap between this odd point and its immediate even neighbours is wider than the gaps between the even points on both sides. As a consequence, the prediction in this point mixes two scales. Figure 4(b) illustrates what happens to cubic interpolating polynomials on such a grid. If the gap between the point of prediction and the nearest interpolation points is larger than the gaps between the interpolation points at both sides, the interpolating polynomial stretches over two scales and may show high values at the position where the prediction takes place. Fine scale information is in some sense extrapolated to coarse scales, resulting in unexpected values.

5 Stabilising the lifting scheme

Because the stability problem is a combination of cumulating effects, there are several points in which the lifting procedure can be modified in order to enhance stability. A combination of modifications may further reduce the problem.



(a) (b)

Figure 4: (a) A scaling function with a negative integral and a scaling function of which the maximum does not coincide with the initial central point of the subdivision scheme. (b) Illustration for cubic interpolating polynomials.

5.1 Update step

There are several ways to control the destabilising effect of the update step. A method of local semi-orthogonalisation has been proposed [24] and applied to curve estimation problems [25].

A more direct method to reduce the magnitude of update coefficients, is by looking for a minimum norm update filter. In order to meet the same number of vanishing moments, the filter length has to be extended. For an update with two vanishing moments we can use a three taps update filter with coefficients $A_{j,k}, B_{j,k}, C_{j,k}$. This leads to the wavelet function:

$$\psi_{j,k} = \psi_{j,k}^{[0]} - C_{j,k}\varphi_{j,k-1} - B_{j,k}\varphi_{j,k} - A_{j,k}\varphi_{j,k+1}. \tag{17}$$

Two degrees of freedom are used to impose the vanishing moments. The general solution for the update coefficients then involves a parameter, u, that is left to choose:

$$A_{j,k} = A_{j,k}^{2taps} - u\frac{L}{N}, \qquad \text{with} \qquad L = M_{j,k-1}I_{j,k} - M_{j,k}I_{j,k-1}, \tag{18}$$

$$B_{j,k} = B_{j,k}^{2taps} - u\frac{K}{N}, \qquad K = M_{j,k+1}I_{j,k-1} - M_{j,k-1}I_{j,k+1}, \tag{19}$$

$$C_{j,k} = u. \qquad N = M_{j,k+1}I_{j,k} - M_{j,k}I_{j,k+1}. \tag{20}$$

$$\tag{21}$$

The minimum norm solution for this system follows from

$$u = \frac{N}{L^2 + K^2 + N^2}.\left(A_{j,k}^{2taps}L + B_{j,k}^{2taps}K\right). \tag{22}$$

A third possibility is to look for the closest pair of coarse scaling functions $\varphi_{j,l}$ and $\varphi_{j,r}$ that satisfy

$$\overline{x}_{j,l} < \overline{x}_{j+1,2k+1} < \overline{x}_{j,r}. \tag{23}$$

All three procedures expand the support width of the wavelet basis functions. Since the actual problem originate from splitting and prediction, a solution based on the update only has no guarantee of being effective in all cases.

5.2 Prediction and interval boundaries

The interpolation scheme for the prediction of 'odd' points needs to be adapted near the boundary of the interval. It is no longer possible to choose the interpolating points symmetrically around the point of prediction. The standard lifting procedure then chooses the interpolating points as close as possible to the prediction point, allowing for asymmetrical interpolation and even extrapolation as prediction, as illustrated in Figure 5(b). As a consequence, some prediction points use the same interpolating polynomial and the prediction in points close to the boundary is influenced by points relatively far.

We therefore propose to give up some vanishing moments in the neighbourhood of the boundaries, in order to preserve a symmetric prediction. This new prediction at the boundaries is illustrated in Figure 5(c). This approach is also used in [26, 27] in the framework of adaptive wavelet transforms.

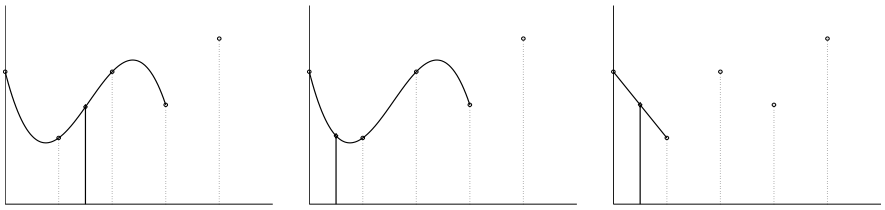


Figure 5: (a) Symmetrical interpolation points away from the boundary. (b) Close to the boundary the standard lifting procedure chooses the interpolation points as close as possible to the prediction point, but asymmetrically. (c) We preserve a symmetric prediction, giving up some vanishing moments.

Smooth functions, or smooth pieces of functions, are well approximated by polynomials and therefore have small coefficients if the wavelet basis reproduces polynomials exactly up to a certain degree. This degree is the number of dual vanishing moments: it measures the approximation capacity of the wavelet basis. Shortening the prediction stencil near the boundaries may therefore lead to a less sparse representation in the neighbourhood of these boundaries.

5.3 Splitting and scale mixing

As explained in Section 4.3 a good subdivision algorithm should not mix scales in one step. To this end, the splitting of samples can be reorganised. Changing the splitting procedure also influences the subsequent prediction and update steps.

5.3.1 At the splitting stage

Following the analysis of Figure 4, we wish to exclude odd points from the list of points whose function value is predicted, if this prediction would involve different scales. Figure 6 illustrates what we do. Point d is moved from the list of 'odd' points (points in which the value is predicted) towards the list of 'even' points (points used for prediction) since the distance C between this point and its even neighbour e is larger than the minimum distance between the

evens used for the prediction (in this case: $C > D + E$). Adding point d to the list of evens however introduces a new problem for the prediction of point b, since the distance between b and even neighbour c is larger than the distance between this even neighbour and new even d. We could rerun the resplitting procedure until we have reached a point where no scale mixture occurs. This would make some lattices unsplittable, or may lead to a slow progress in the wavelet coefficient computation.

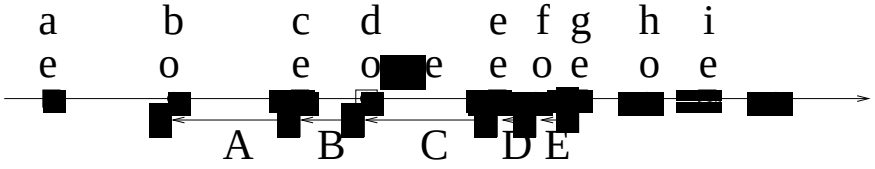


Figure 6: Re-arranging the split procedure: odd point d is added to the even list, because predicting the value in this point with a cubic polynomial would involve two scales. Indeed, the distance $C > D + E$.

5.3.2 Prediction stencil

We therefore take a different approach to deal with odd points that get into an unbalanced scale situation after new evens were added. Those new evens are not used for prediction if they are too close to an existing even point. The even point originally destinated to do this job is used instead. In the example of Figure 6, the value in d is only used for the prediction in f, not for b. The computation of the detail coefficient in b involves the values in c and in e, as originally planned.

One could ask why we do not apply the same procedure to predict the value in d. We could leave this point in the 'odd' state and predict it using the values in a, c, e and i. This would reduce instability indeed, but not completely. It could introduce new problems, since it would create a heavy and far side blob in the scaling function associated with i. If we apply this procedure on d, on the other hand, the scaling function associated with e stretches up to a, but that is no further than it was originally.

5.4 Examples and extensions

We now apply the proposed scheme to the 'heavisine' example from Figure 2. Using the symmetric prediction near the boundaries, as proposed in Section 5.2, in combination with the resplit procedure of this section, reduces the bias to acceptable level, as shown in Figure 7. This reconstruction involves approximately the same number of thresholded coefficients and none of them reveals hidden effects. The output is still a bit noisy, due to the imperfections of the brute threshold approach. More sophisticated coefficient selection remove a great deal of this remaining noise, typically by looking accross scales [28, 29].

Although the problem analysis had a four points prediction in mind, the method also solves the even worse instabilities at higher prediction orders. Figure 8 investigates the effect of our proposed modifications on a transform with eight dual vanishing moments, i.e. the predic-

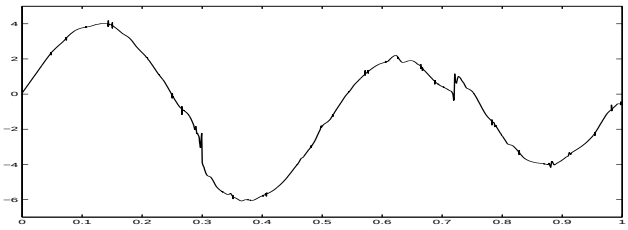


Figure 7: The proposed lifting scheme applied to the example of Figure 2.

tion was a seventh degree polynomial. The proposed stabilising methods bring the estimated function back to finite values, in a smooth and close fit.

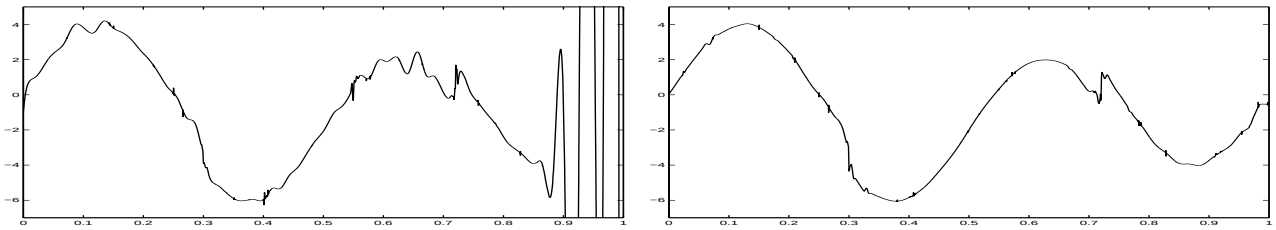
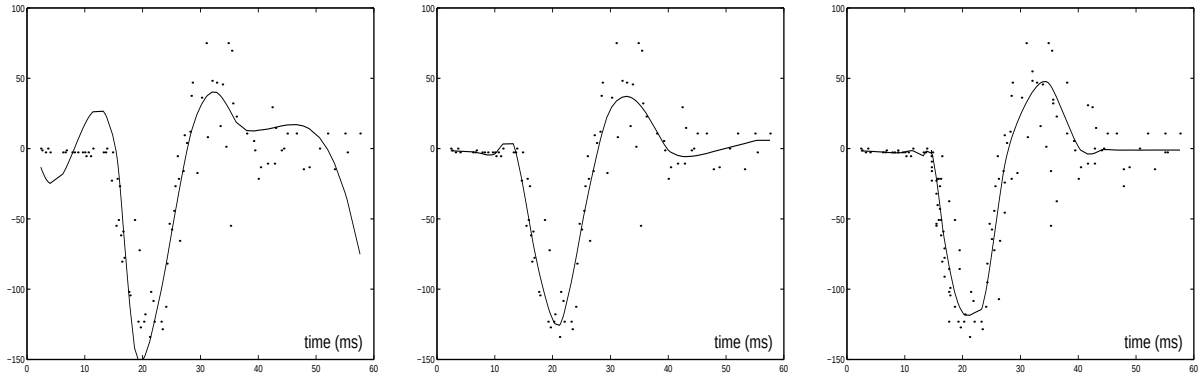


Figure 8: Left: the classical implementation of a second generation wavelet with 8 dual and 2 primal vanishing moments leads to an unacceptable bias. Right: the proposed lifting scheme has a stabilising effect, also for 8 dual vanishing moments.

A second example is a real data set [30]. It has 133 observations of motorcyclists' heads during a crash. The set represents the values of the head acceleration as a function of time (in milliseconds). This data set contains several simultaneous observations (i.e. different acceleration values at the same time). The classical second generation transform crashes when three or more abscis points coincide (the return values were ∞). Therefore, we first omit double observations and fit the data in 94 single observations. We could apply a threshold procedure. The observations are clearly heteroscedastic, so we first need to estimate the standard deviation in each point, for instance, using a *pilot estimator* as in [25]. Our findings were that the final outcome heavily depends on this pilot estimator and on the number of levels being thresholded. As a matter of fact, the non-linearity of thresholding added little effect to a simple, linear estimation: remove all coefficients at the first say M levels. A cross validation procedure finds a good value for M: we first apply one step of the (new or classical) splitting scheme without prediction or update. We then run the algorithm on the reduced set, using a forward wavelet transform of $M-1$ levels. We compute the values of the inverse transform with M levels in the left-out data points. The quality of this prediction (w.r.t. the observed values) measures the quality of the wavelet estimator with M levels of scale. Figure 9(a) shows that also on small data sets, the classical implementation suffers from instability. Our modified algorithm finds a smooth and quite unbiased curve in 9(b). Thanks to the improved splitting scheme, no problem arises from simultaneous data, as shown in the output in Figure 9(c).

In a third example we repeated an experiment a hundred times, i.e. we generated 100 random grids with uniformly distributed sample points and additive Gaussian noise. We compared to



(a) (b) (c)

Figure 9: Fitting the motorcyclists' acceleration data. (a) Using a classical second generation wavelet transform (cubic interpolating prediction, 2-taps update) leads to bias. (b) Applying the corresponding, modified transform, preventing mixing of scales and with symmetric, lower order prediction near the boundaries. (c) The original data set contained simultaneous observations, which pose no problem in the proposed scheme.

	no-grid	simple lifting	stabilized lifting
mean SNR	29.71 dB	10.38 dB	29.20 dB
stand. dev.	0.666 dB	9.95 dB	0.86 dB

Table 2: Average output SNR-values and standard deviations for three methods applied to 100 experiments. Input SNR was 20 dB.

output of three methods: the procedure without taking the grid into account (brute remapping), the non-stabilized lifting procedure, and our stabilized version. Table 2 compares the average output SNR-values, as well as the estimated standard deviation (i.e. sum of squared deviations from the average, devided by total number of experiments minus one). The input SNR for all these experiments was 20 dB. On the average, the straightforward lifting procedure sees a dramatic decrease in SNR, because of the occasionally severe bias. Moreover, there is large variation in output quality: some experiments generate a reasonable output, but in others, the bias may even destroy the signal completely, resulting in a negative SNR value. Both the no-grid approach and our stabilzed lifting procedure show a regular performance. The no-grid approach has a slightly higher signal-to-noise ratio, which seems to indicate that minor bias effects are still present after stabilsation. On the other hand, the SNR was computed point by point, so this measure has no direct influence from smoothness.

6 Conclusions

This paper has presented an original analysis of the instability problems of second generation wavelet transforms. This problem is the cumulated effect of several factors: the three successive steps in a lifting scheme — split, prediction and update — together are responsible for the

instability.

Based on this analysis, the paper proposed three novel adaptations for the lifting scheme:

1. the first modification is a minimisation of the update coefficients in order to reduce the effect of the previous steps,

2. the second modification is a relaxation of the prediction operation near the boundaries,

3. while the third modification starts from an alternative splitting scheme, followed by an according prediction and update, in order to deal with the irregularity.

Although the analysis concentrates on cubic polynomial prediction, the experiments illustrate that the adaptations are applicable for a wider range of prediction operators. The combination of our proposed modifications reduces the bias after reconstruction to the order of magnitude of bias on the wavelet coefficients: this compares to the classical, (bi)orthogonal situation.

These results leave us with a couple of mainly theoretical questions, none of which has a trivial answer, given the fact that subdivision on arbitrary grids in general is still a quite open field. More specifically, when applying lifting to non-equispaced data, at least four issues are important:

1. Convergence of the subdivision scheme is necessary for proper definition of the multiresolution basis functions. Whereas convergence is well understood in regular or semi-regular design, far less is known about convergence of subdivision on arbitrary point sets. Since Fourier techniques no longer apply, the proofs follow a totally different path [1]. The present results mainly concentrate on the cubic interpolation case and the extension to higher order scheme is not automatic. Moreover, the results are based on certain assumptions on the homogeneity and balancing of the grid. In our rearranged transform, it is unclear what the effect is of skipping prediction points in order to prevent scale mixing. The present results do however not depend on the splitting strategy, so one might expect that the new scheme does converge, possibly under less stringent conditions, because of the scale mix prevention. Proving this using the present techniques, would require a considerable effort.

2. What can be said about the smoothness of the limiting curve? If the rearranged scheme converges, what do the basis functions look like? Looking at the results of our experiments, we conjecture that the basis functions exist and are smooth, but they may have an interval of strict positive length inside their support where they are exactly zero.

3. What result can be found for the numerical condition? Is there any upper bound for the condition number? This could help in establishing a more precise criterion for splitting the input into interpolation points and prediction points. Also note that convergence results give information about the scaling basis, and the corresponding nested spaces. The step to a wavelet decomposition, i.e. update filters, may add new instabilities, as discussed in this paper.

4. What are the approximation properties of the basis? Dual vanishing moments play an important role in approximation. Since our adaptation includes a relaxation on the number of dual moments near the boundary, this could have (minor) impact on the local approximation capacity and the sparsity of the wavelet decomposition.

Acknowledgement

This work is partially supported by the Belgian Program on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with the authors. Evelyne Vanraes is funded as a Research Assistant of the Fund for Scientific Research Flanders (Belgium) (FWO). Maarten Jansen is postdoctoral research fellow of the Fund for Scientific Research Flanders (Belgium) (FWO). We wish to thank V. Delouille for kindly pointing us to some real data sets.

References

[1] I. Daubechies, I. Guskov, and W. Sweldens, "Regularity of irregular subdivision," *Constructive Approximation* **15**, pp. 381–426, 1999.

[2] I. Daubechies, I. Guskov, P. Schröder, and W. Sweldens, "Wavelets on irregular point sets," *Phil. Trans. R. Soc. Lond. A* **357**, pp. 2397–2413, 1999.

[3] P. Hall and B. A. Turlach, "Interpolation methods for nonlinear wavelet regression with irregularly spaced design," *Annals of Statistics* **25**(5), pp. 1912 – 1925, 1997.

[4] A. Kovac and B. W. Silverman, "Extending the scope of wavelet regression methods by coefficient-dependent thresholding," *J. Amer. Statist. Assoc.* **95**, pp. 172–183, 2000.

[5] B. Delyon and A. Juditsky, "On the computation of wavelet coefficients," *J. of Approx. Theory* **88**, pp. 47–79, 1997.

[6] T. Cai and L. Brown, "Wavelet shrinkage for nonequispaced samples," *Annals of Statistics* **26**(5), pp. 1783–1799, 1998.

[7] A. Antoniadis, G. Grégoire, and W. McKeague, "Wavelet methods for curve estimation," *J. Amer. Statist. Assoc.* **89**, pp. 1340–1353, 1994.

[8] S. Sardy, D. Percival, A. Bruce, H.-Y. Gao, and W. Stuetzle, "Wavelet de-noising for unequally spaced data," *Statistics and Computing* **9**, pp. 65–75, 1999.

[9] M. Pensky and B. Vidakovic, "On non-equally spaced wavelet regression," preprint, Duke University, Durham, NC, 1998.

[10] A. Antoniadis and D.-T. Pham, "Wavelet regression for random or irregular design," *Computational Statistics and data analysis* **28**(4), pp. 333–369, 1998.

[11] A. Antoniadis, G. Grégoire, and P. Vial, "Random design wavelet curve smoothing," *Statistics and Probability Letters* **35**, pp. 225–232, 1997.

[12] W. Sweldens, "The lifting scheme: A construction of second generation wavelets," *SIAM J. Math. Anal.* **29**(2), pp. 511–546, 1997.

[13] W. Sweldens and P. Schröder, "Building your own wavelets at home," in *Wavelets in Computer Graphics*, ACM SIGGRAPH Course Notes, pp. 15–87, ACM, 1996.

[14] W. Sweldens, "The lifting scheme: A custom-design construction of biorthogonal wavelets," *Appl. Comput. Harmon. Anal.* **3**(2), pp. 186–200, 1996.

[15] V. Delouille, J. Franke, and R. von Sachs, "Nonparametric stochastic regression with design-adapted wavelets," tech. rep., Université Catholique de Louvain, 2000.

[16] M. Jansen, *Wavelet thresholding and applications*. PhD thesis, Department of Computer Science, K.U.Leuven, Belgium, 2000.

[17] I. Daubechies and W. Sweldens, "Factoring wavelet transforms into lifting steps," *J. Fourier Anal. Appl.* **4**(3), pp. 245–267, 1998.

[18] S. Dubuc, "Interpolation through an iterative scheme," *J. Math. Anal. Appl.* **114**, pp. 185–204, 1986.

[19] G. Deslauriers and S. Dubuc, "Interpolation dyadique," in *Fractales, Dimensions Non-entières et Applications*, pp. 44–55, Masson, Paris, 1987.

[20] G. Deslauriers and S. Dubuc, "Symmetric iterative interpolation processes," *Constructive Approximation* **5**, pp. 49–68, 1989.

[21] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory* **41**, pp. 613–627, May 1995.

[22] M. Jansen, M. Malfait, and A. Bultheel, "Generalized cross validation for wavelet thresholding," *Signal Processing* **56**, pp. 33–44, Jan. 1997.

[23] A. Cohen, I. Daubechies, and J. Feauveau, "Bi-orthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.* **45**, pp. 485–560, 1992.

[24] J. Simoens and S. Vandewalle, "Average-interpolating wavelet bases on irregular meshes on the interval," TW Report 310, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, 2000.

[25] V. Delouille, J. Simoens, and R. von Sachs, "Smooth design-adapted wavelets for nonparametric stochastic regression," tech. rep., Université Catholique de Louvain, 2001.

[26] R. L. Claypoole, R. Baraniuk, and R. D. Nowak, "Adaptive wavelet transforms via lifting," in *Proceedings of the 1998 International Conference on Acoustics, Speech, and Signal Processing - ICASSP '98*, 1998.

[27] R. Claypoole, G. M. Davis, W. Sweldens, and R. Baraniuk, "Nonlinear wavelet transforms for image coding via lifting," *Submitted to IEEE Trans. on Image Proc.* , 1999.

[28] R. Baraniuk, "Optimal tree approximation with wavelets," in *Wavelet Applications in Signal and Image Processing VII*, M. A. Unser, A. Aldroubi, and L. A. F., eds., vol. 3813 of *SPIE Proceedings*, pp. 196–207, July 1999.

[29] Y. Xu, J. B. Weaver, D. M. Healy, and J. Lu, "Wavelet transform domain filters: a spatially selective noise filtration technique," *IEEE Transactions on Image Processing* **3**(6), pp. 747–758, 1994.

[30] B. Silverman, "Some aspects of the spline smoothing approach to non-parametric regression curve fitting," *Journal of the Royal Statistical Society, Series B* **47**, pp. 1–52, 1985.