Stabilized wavelet transforms for non-equispaced data smoothing

Evelyne Vanraes[1*], *Maarten Jansen*[2] *and Adhemar Bultheel*[3]

[1,3] Dep. of Comp. Science, K.U.Leuven, Belgium

[2] Dep. of Elect. & Comp. Eng., Rice U., U.S.A. and
Dep. of Comp. Science, K.U.Leuven, Belgium

Technical Report TW319, February 2001

Abstract

This paper discusses wavelet thresholding in smoothing from non-equispaced, noisy data using so called second generation wavelets. We explain that a good numerical condition is an absolute requisite for successful thresholding. We examine the nature and origin of stability problems in second generation wavelet transforms, i.e. when applying a lifting scheme to non-equispaced data. The investigation concentrates on lifting with interpolating prediction, but the conclusions are extendible. The paper proposes two ways to stabilize the second generation wavelet transform. The first operates on the interval boundaries, while the second concentrates on the irregularity of the data points. Illustrations show that reconstruction from thresholded coefficients with this stabilized second generation wavelet transform leads to smooth and close fits.

1 Introduction

The classical wavelet based methods for data smoothing mostly assume the input to be a dyadic vector of equispaced, homoscedastic data, i.e., they model the input as:

$$y_i = f(i \cdot \Delta_t) + \eta_i, \qquad i = 1, \dots, N,$$

where $N = 2^J$ for some integer J.

The wavelet basis functions, used in these methods, possess smoothness properties on regular, dyadic grids. When used for data on irregular point sets, *remapping* these basis functions to the actual grid, makes the irregularities show up in the output [1, 2].

*Contact information: Evelyne Vanraes, Dep. of Comp. Science, Celestijnenlaan 200A, B-3001 Leuven, Belgium, phone: ++32 16 32 7080, fax: ++32 16 32 7996, evelyne.vanraes@cs.kuleuven.ac.be

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 'far from orthogonal' (i.e. 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 feasable, may result in additional instability.

2 Our approach

2.1 Lifting

A classical wavelet decomposition algorithm has the structure of a repeated filter bank algorithm. The repeated block (Figure 1) is a filterbank consisting of two filters. The input, a discrete signal S, is convolved with a low pass filter $\tilde{\boldsymbol{h}}$ and a high pass filter $\tilde{\boldsymbol{g}}$, and downsampled. The filter $\tilde{\boldsymbol{h}}$ results in a smoothed version of the input. The filter $\tilde{\boldsymbol{g}}$ gives the detail information. Downsampling the filter outputs leads to low pass coefficients LP and high pass or *wavelet coefficients* HP. The low pass coefficients can be considered as a smoothed version of the input. These *scaling coefficients* are further processed in the next step, using the same filter bank.

One step of the inverse transform starts by upsampling, i.e. putting zeroes in between every two samples. This is followed by a convolution in the two filter bank branches with filters $\boldsymbol{g}$ and $\boldsymbol{h}$ respectively and a final summing of these convolutions. The *perfect reconstruction* requires the reconstruction from unmodified wavelet coefficients to be equal to the input signal S. Therefore conditions rest on the filters $\tilde{\boldsymbol{g}}$, $\tilde{\boldsymbol{h}}$, $\boldsymbol{g}$ and $\boldsymbol{h}$.

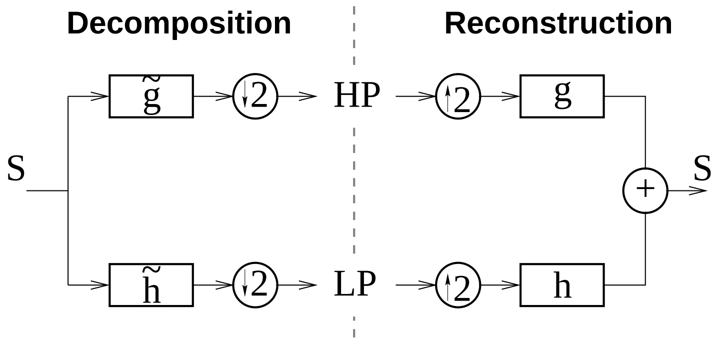


Figure 1: One step of a wavelet decomposition and reconstruction. This is a filterbank: the input is filtered and down-sampled to get a low pass signal *LP* and a high pass signal *HP*. Reconstruction starts with up-sampling by introducing zeroes between every pair of points in *LP* and *HP*.

The original signal S lives in the space V_J with basis functions $\varphi_{J,k}$, where J stands for the highest resolution level used. The low pass signal and the high pass signal resulting from one step of the wavelet transform are each in another of two complementary subspaces of V_J. The low pass subspace V_{J-1} has the scaling functions $\varphi_{J-1,k}$ as basis functions. The high pass subspace W_{J-1} has the wavelet functions $\psi_{J-1,k}$ as basis functions. A signal $S(x)$ can be decomposed as

$$\begin{aligned} S(x) &= \sum_{k=1}^{2^J} s_{J,k}\varphi_{J,k}(x) \\ &= \sum_{k=1}^{2^{J-1}} s_{J-1,k}\varphi_{J-1,k}(x) + \sum_{k=1}^{2^{J-1}} w_{J-1,k}\psi_{J-1,k}(x) \end{aligned}$$

where $s_{j,k}$ are the scaling coefficients and $w_{j,k}$ are the wavelet coefficients. Repeating the same filterbank procedure on the low pass signal at a certain resolution j decomposes the space V_j again in V_{j-1} and W_{j-1}. This is a multiresolution analysis (MRA). Transforming from the highest resolution level J to the lowest resolution L gives the decomposition

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

For an orthogonal transform the subspaces V_j and W_j are orthogonal to each other and the basis functions within each subspace are also orthogonal. If $\varphi_{j,k}$ and $\psi_{j,k}$ are orthonormal:

$$\begin{aligned} V_j &\perp W_j, \\ \langle \varphi_{j,k}, \varphi_{j,k'} \rangle &= \delta_{kk'}, \\ \langle \psi_{j,k}, \psi_{j',k'} \rangle &= \delta_{jj'}\delta_{kk'}. \end{aligned}$$

In general a MRA has no orthogonality, but biorthogonality. In that case we also have a dual scaling function $\tilde{\varphi}$ and a dual wavelet function $\tilde{\psi}$ that fit in a dual MRA with spaces $\tilde{V}_j$ and $\tilde{W}_j$. The biorthogonality conditions become:

$$\begin{aligned}
\tilde{V}_j &\perp W_j, \\
V_j &\perp \tilde{W}_j, \\
\langle \tilde{\varphi}_{j,k}, \varphi_{j,k'} \rangle &= \delta_{kk'}, \\
\langle \tilde{\psi}_{j,k}, \psi_{j',k'} \rangle &= \delta_{jj'}\delta_{kk'}.
\end{aligned}$$

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 very simple. It starts with a trivial wavelet, the Lazy wavelet, that is just splitting the signal in points with an odd index and points with an even index. This corresponds to subsampling in the classical filterbank. The lifting scheme then gradually builds a new wavelet with improved properties. The building blocks are lifting steps (Figure 2).

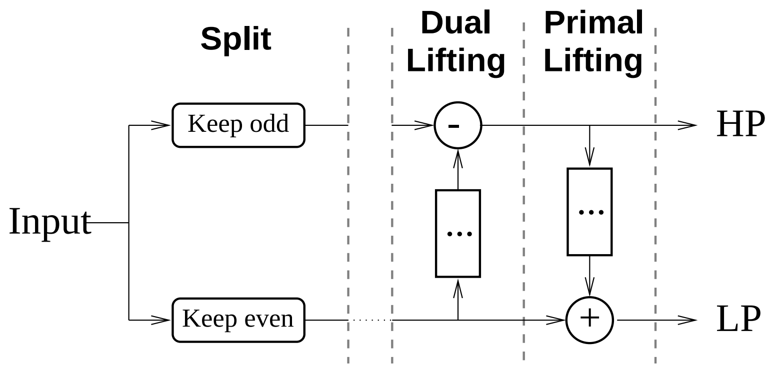


Figure 2: Decomposition of a filterbank into lifting steps. The first type of lifting is called *dual lifting* or *prediction*. The other type is *primal lifting* or *update*

Dual lifting substracts a filtered version of the even samples from the odd samples. Primal lifting adds a filtered version of the dual lifting output to the so far untouched even samples. Dual and primal lifting are often called prediction and update due to an interpretation one can give to it. Before prediction the even and odd samples are highly correlated. We then try to predict the odd samples by a prediction filter on the even samples. Substracting this prediction from the odd samples reduces the correlation. The differences are the detail coefficients, the high pass information. The prediction formula used in this paper is an interpolating polynomial. The polynomial interpolates the even samples and the prediction is the evaluation in the odd point.

The update step can be interpreted as a way to preserve the average and higher moments in the low pass coefficients.

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. Unlike the classical filterbank setup, the same filters appear in forward and inverse transform. The inverse diagram is shown in Figure 3.

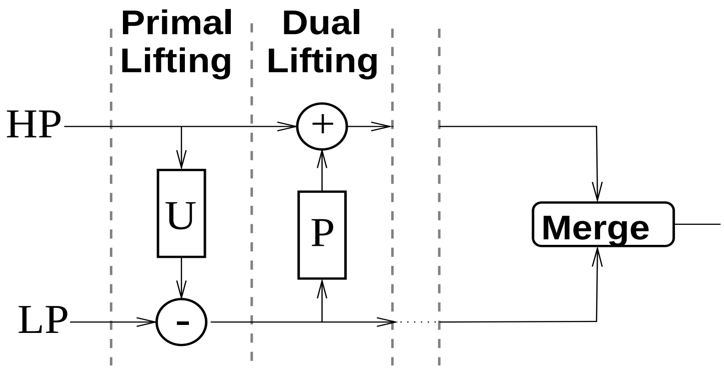


Figure 3: The inverse lifting scheme: run through the lifting scheme backwards, replacing plus with minus-signs and merge what had been split. Running an inverse wavelet transform from a sequence of zeros except for one coefficients equal to one leads to the basis function at that position.

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

$$w_{j,k} = \delta_{ji}\delta_{kl}$$

synthesizes the function:

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

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:

$$\psi^{[0]}_{j,k} = \varphi_{j+1,2k+1}.$$

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:

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

The extension to longer update filters is obvious.

2.3 Thresholding

A wavelet transform has strong decorrelating properties. It uses the correlation between neighbouring samples to obtain a sparse representation of the noise free signal. The main part of the coefficients is close to zero and the essential information is captured by a limited number of large, important coefficients. Replacing the small coefficients with an absolute value below a certain treshold with zero reduces the noise without affecting the noise free signal too much. Coefficients with an absolute value above the threshold are shrunk with the threshold. This approach is called soft thresholding. 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:

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

with:

$$\begin{aligned}
\text{bias}^2(\lambda) \quad &:= \quad \frac{1}{N}\|\mathrm{E}\boldsymbol{w}_{\boldsymbol{\lambda}} - \boldsymbol{v}\|^2 \\
\text{variance}(\lambda) \quad &:= \quad \frac{1}{N}\mathrm{E}\|\boldsymbol{w}_{\boldsymbol{\lambda}} - \mathrm{E}\boldsymbol{w}_{\boldsymbol{\lambda}}\|^2.
\end{aligned}$$

In these equations, $\boldsymbol{w}$ stands for the vector of noisy wavelet coefficients and $\boldsymbol{w}_{\boldsymbol{\lambda}}$ 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 [18].

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. This is shown in figure 4 for the linear case. Also the update step is not limited to equidistant data.

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

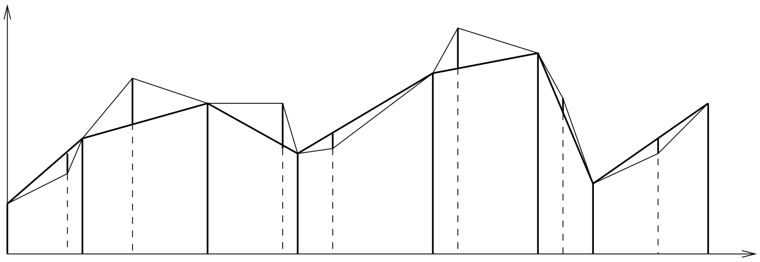


Figure 4: Linear prediction operator an an irregular grid.

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

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

can be performed with linear complexity if the input correlation matrix Q is banded [16], for instance, if the input noise is uncorrelated. Dividing each coefficient w_i by the corresponding diagonal element $\sqrt{S_{ii}}$ results in homoscedastic noise.

2.5 Example

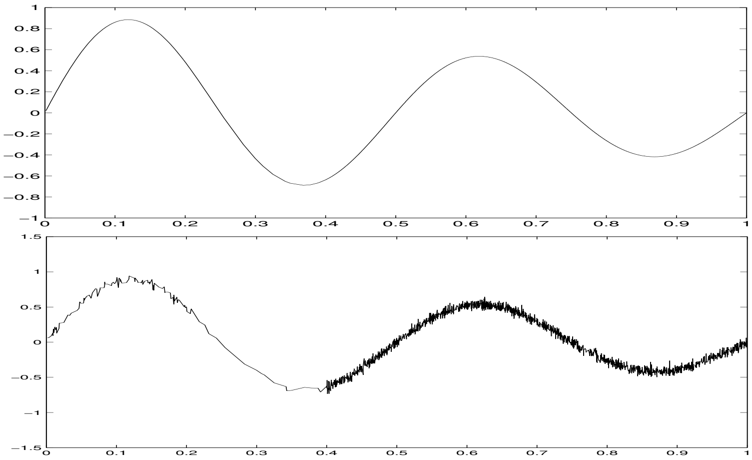


Figure 5: A damped sine ($f(x) = e^{-x} \sin 4\pi x$) and a noisy version of this signal.

We now apply this to an example. Figure 5 shows a damped sine ($f(x) = e^{-x} \sin 4\pi x$) on an irregular grid and a noisy version of this signal. For this grid we choose approximatly 100 samples at random between 0 and 0.2., about 10 samples between 0.2 and 0.4 and about 1940 samples between 0.4 and 1.

Figure 6 compares the result of a classical wavelet transform with the result of a second generation wavelet transform. Second generation wavelets allow a smoother reconstruction, but in this example the result shows a bias.

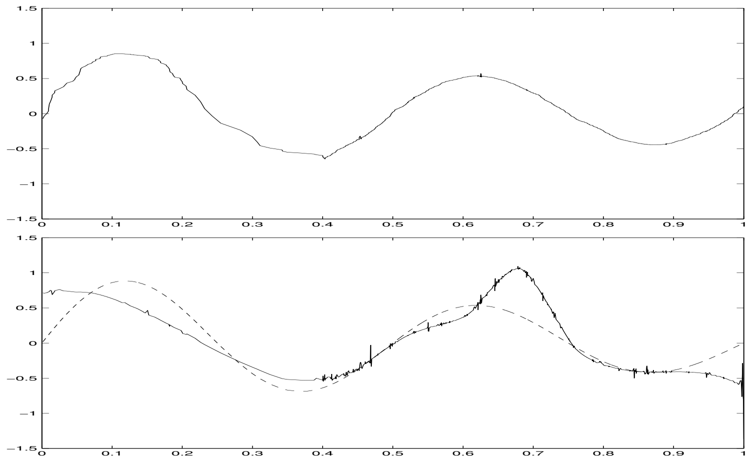


Figure 6: Top: result of a threshold procedure with a classical wavelet transform. The real grid structure is neglected, the result is still noisy. Bottom: result of a threshold procedure with a second generation wavelet transform. The result is smoother but shows a bias.

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:

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

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 3.1 illustrates that condition numbers can be quite high. The condition number of the multiscale transform matrix $\tilde{W}$ is defined as

$$\kappa = \|\tilde{W}\|\|\tilde{W}\|^{-1}.$$

$\tilde{n}$	regular $\gamma = 1$	random $\gamma = 10032$	irregular $\gamma = 13930$
2	$1.98.10^2$	$3.25.10^1$	$5.70.10^1$
4	$6.11.10^1$	$1.48.10^4$	$2.34.10^4$
6	$1.78.10^2$	$6.02.10^3$	$7.32.10^7$
8	$1.51.10^3$	$1.22.10^5$	$1.05.10^{11}$

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

As illustrated in [19] the condition gets worse in function of the homogeneity constant. This constant captures ratios of neighbouring intervals without paying any attention to the refinement procedure. It is defined as follows for a multilevel mesh [1] :

$$\gamma = \sup_{j,k} \frac{\max(I_{j,k+1}, I_{j,k-1})}{I_{j,k}}$$

with j the level of the grid point and $I_{j,k}$ the distance between $x_{j,k}$ and $x_{j,k+1}$:

$$I_{j,k} = x_{j,k+1} - x_{j,k}.$$

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 renormalized according to their variances. This makes the problem analysis even more complex.

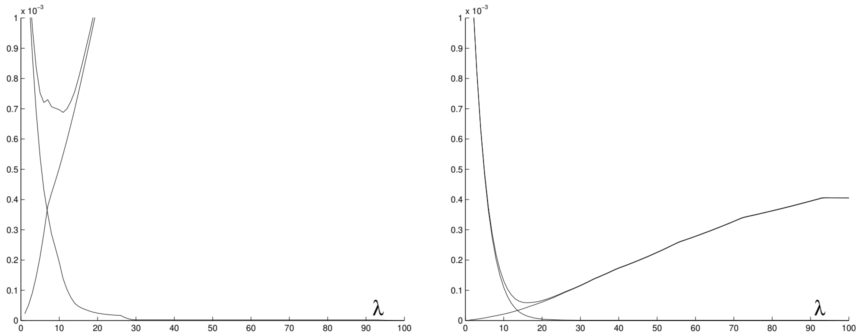


Figure 7: Left: MSE plot in the signal domain. Right: MSE plot in the wavelet domain. The transform is unstable: the optimal threshold in the wavelet domain results in an unacceptable bias in the signal domain.

Figure 7 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

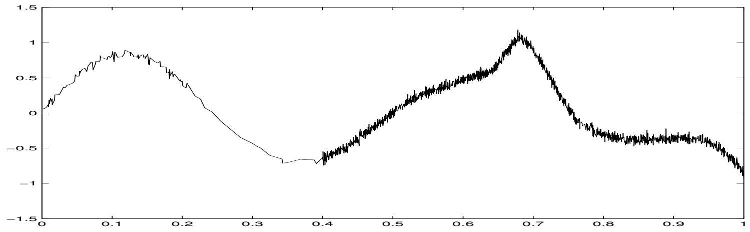


Figure 8: Reconstruction after removing one coefficient from the noisy transform. The effect is enormous, but the coefficient was rather big.

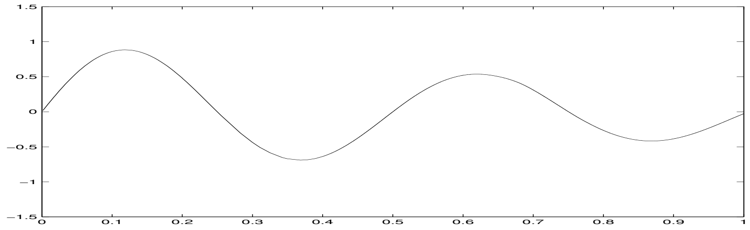


Figure 9: Reconstruction after removing the same coefficient as in Figure 8 from the noise-free transform. The effect is quasi nihil.

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 minimizing 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: the sharp, deep MSE plot makes it hard to find any good threshold value.

3.3 Hidden components

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. Figure 8 shows an experiment where one particular second-generation wavelet coefficient of the noisy signal was replaced by zero. Inverse transform reveals a tremendous effect. The coefficient had a rather large magnitude, and apparently also a wide impact, but comparison of the results in Figure 8 and Figure 6 indicates that the same coefficient was classified as not important by the threshold algorithm and therefore discarded. This is because not only its magnitude was large, but so was its variance. If we replace the same coefficient in the noise-free set by zero, we get the reconstruction in Figure 9. The difference with the original function is hardly visible. The threshold algorithm was right to remove it. 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 10. 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 exam-

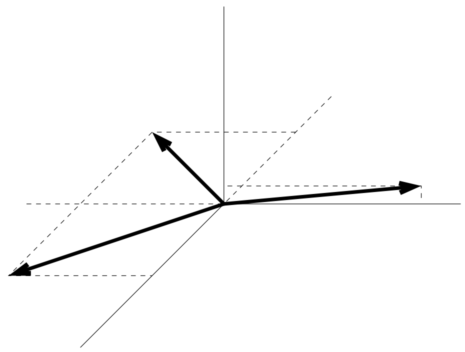


Figure 10: An arbitrarily unstable basis in $\mathbb{R}^3$.

ple of Figure 8, the noise added large coefficients to the wavelet representation. In the unthresholded set of coefficients, the effect of one coefficient is canceled by a combination of other wavelet coefficients. As for Figure 10, 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

In order to remedy the stability problem, we need to have a closer look at its precise origin. It turns out that the high condition numbers are the outcome of an interaction between several factors. We may consider at least two forms of interaction: on one hand, there is the interaction between the irregularity of the grid and the careless treatment of the boundaries of the interval. The second type of interaction is between the three successive steps in a lifting scheme: split, predict and update.

4.1 Irregularity versus boundary

If we perform the smoothing procedure, neglecting the underlying grid, the algorithm seeks for a smooth output on an equidistant grid. Remapping to the true data structure then introduces all the irregularity of the grid into the data, thereby creating a non-smooth curve. This curve however shows far less bias, so the irregularity is definitely a factor contributing to the instability.

On the other hand, most of the problems occur near the interval end points, or at least, seem to origin from that area.

4.2 Split, predict or update

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

4.2.1 Large update coefficients

If the update coefficients are large, for instance, if:

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

Equation (1) shows that 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. 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 localize.

4.2.2 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:

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

$$B_{j,k} = \frac{M_{j+1,2k+1}}{M_{j,k+1}} \frac{\overline{x}_{j+1,2k+1} - \overline{x}_{j,k}}{\overline{x}_{j,k+1} - \overline{x}_{j,k}}. \tag{3}$$

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

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

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

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

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.

If we use linear interpolating prediction, the scaling functions are positive. This becomes clear from running the subdivision scheme with linear prediction. The scaling functions follow from the prediction step only. No update is involved in their computation. In that case, the value of $\overline{x}_{j,k}$ can be interpreted as the mean value of the corresponding scaling function. From Figure 11, it is clear that

$$\overline{x}_{j,k} < \overline{x}_{j+1,2k+1} < \overline{x}_{j,k+1}$$
$$M_{j+1,2k+1} < M_{j,k}.$$

As a consequence, $0 < A_{j,k} < 1$ (and similarly for $B_{j,k}$), and the transform is stable on any grid.

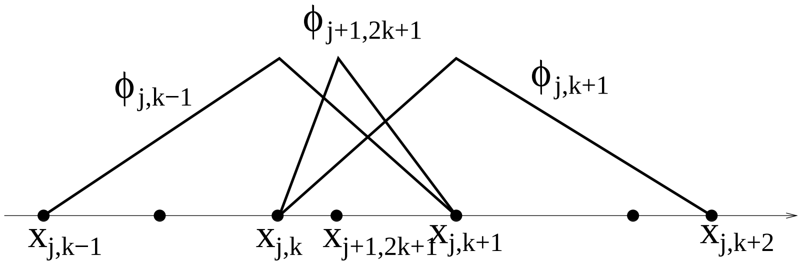


Figure 11: The CDF 2,2 is always stable.

4.2.3 Splitting and scale mixing

If we use interpolating prediction with polynomials of higher order than just linear, the scaling functions may show unwanted features. Figure 12(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.

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 12(b) illustrates what happens to cubic interpolating polynomials on such a grid: fine scale information is in some sense extrapolated to coarse scales, resulting in unexpected values.

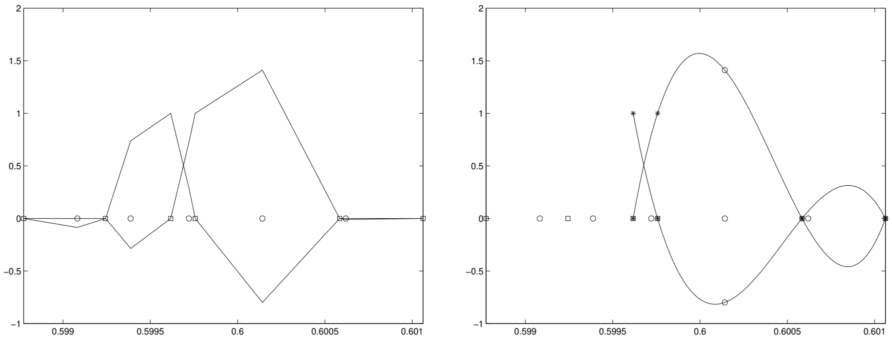
A good subdivision algorithm should not mix scales in one step. To this end, the splitting of samples could be reorganized.

5 Stabilizing 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.

We try to modify the transform so that it becomes more orthogonal. It is not necessary to orthogonalize the whole transform. Often just a few basis functions are responsible for the bad condition. In examples where we can isolate these basis functions, orthogonalising this set improves the condition number considerably.

As explained in Sections 3.3 and 4.2.1 these unstable combinations of wavelet functions mostly form a set across the scales at a fixed location. This set originates from a large update coefficient, resulting in a considerable overlap of scaling and wavelet function at a given scale, followed by further decomposition of the scaling function. Hence, interscale orthogonality is a more important issue than orthogonality within a single level. Orthogonalizing subspaces at



(a) (b)

Figure 12: (a) Mixing of scales creates scaling functions with heavy side blobs. These scaling functions may have negative integrals or may have a maximum which does not coincide with the initial central point of the subdivision scheme. Update steps using combinations with this kind of scaling functions are likely to use high update coefficients in order to get wavelet functions with a given number of vanishing moments. (b) 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.

subsequent levels without necessarily orthogonalizing the wavelets within each level is called *semi-orthogonalization*.

5.1 Update

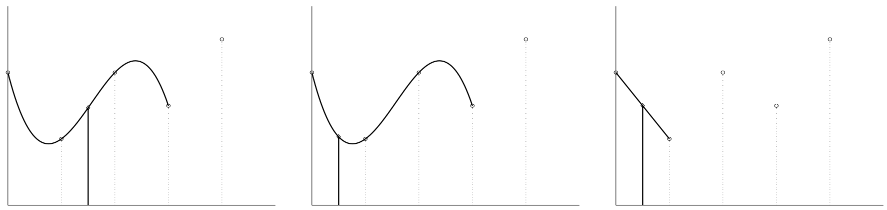
As discussed before, the classical implementation of the lifting scheme spends all degrees of freedom in the update step on vanishing moments. This objective gives no guarantee whatsoever for stability. Relaxing on this objective and leaving some degrees of freedom for other objectives can reduce the number of large update coefficients.

A procedure of *local* semi-orthogonalization [20] concentrates on the locations of the instabilities. This procedure yields no vanishing moments: all degrees of freedom are used for the orthogonalisation. The update step is defined by its *stencil*, which determines the detail coefficients involved in updating the coarse scaling coefficients. Local semi-orthogonalization while keeping the number of primal vanishing moments is achieved by incorporating more coefficients and thus more degrees of freedom into the stencil. The price to pay, however, is that longer stencils create basis functions with wider support.

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 13. 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 13. Figure 14 shows the result on the example of Section 2.5. This approach is also used in [21, 22] in the framework of adaptive wavelet transforms.



(a) (b) (c)

Figure 13: (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 assymetrically. (c) We preserve a symmetric prediction, giving up some vanishing moments.

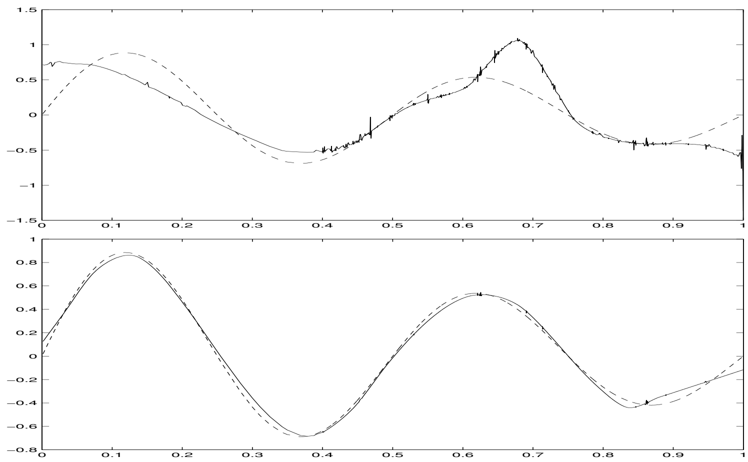


Figure 14: Top: result of a threshold procedure with standard prediction. Bottom: result of a threshold procedure with lower order prediction near the boundaries.

The number of dual vanishing moments measures the approximation capacity of the wavelet basis: 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 (the number of dual vanishing moments). Shortening the prediction stencil near the boundaries may therefore lead to a less sparse representation in the neighbourhood of these boundaries.

5.3 Reorganizing the transform

5.3.1 Predicting one point at the same time

In order to avoid scale mixing effects, we consider a scheme in which only one point is split off from the remaining set in each step. The function value in this point is then predicted using its closest neighbours.

The central question in this procedure is which point to select. The goal is to find a point for which the prediction does not involve different scales. For example we can choose the point which minimizes the distance between the left most and right most interpolation point.

A similar approach is applied in a two-dimensional wavelet transform for scattered data smoothing [23]. The prediction in this transform uses a least square plane in the points immediately adjacent to the selected point. Converted into the one dimensional setting, this procedure would reduce to the CDF 2,2 (linear prediction - linear update), which is stable on any grid.

The extension of this approach to a higher order prediction turns out to be 'greedy': since it only considers local geometry, it tends to create additional inhomogeneity. Figure 15 illustrates this effect on a equidistant set of sample positions.

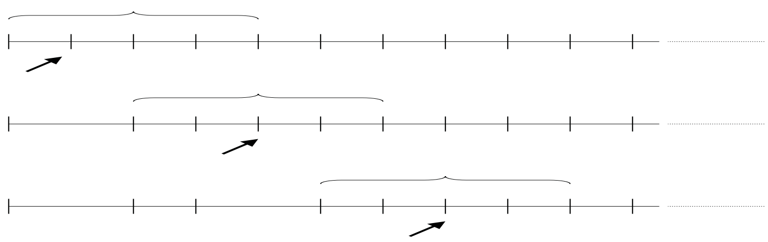


Figure 15: This selection procedure only considers local geometry. It tends to create additional inhomogeneity.

5.3.2 Preventing the mixing of scales

At the splitting stage

Following the analysis of Figure 12, we wish to exclude odd points from the list of points whose function value is predicted, if this prediction would involve different scales. Figure 16 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.

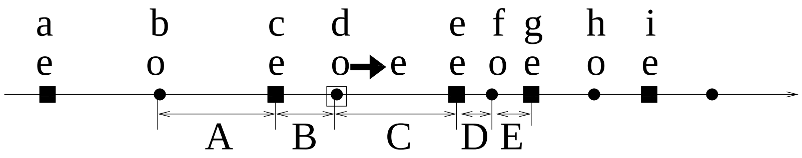


Figure 16: Re-arranging the split procedure: odd point d is added to the even list, because predicting the value in this point with a cubic polynomial would involve two scales. Indeed, the distance $C > D + E$.

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 16, 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.

Update stencil

The modifications in split and predictions affects the scaling functions. Further theoretical analysis is needed to find what exact conditions are necessary to guarantee 'manageable' scaling functions (in whatever sense). An ultimate modification in the update stencil deals with remaining artifacts such as those near boundaries.

If the values of $\overline{x}_{j,k}$ in the equations (2) and (3) do not satisfy

$$\overline{x}_{j,k} < \overline{x}_{j+1,2k+1} < \overline{x}_{j,k+1},$$

the update coefficients may be (arbitrarily) large, with respect to the scaling function integrals. The modifications in the previous steps reduce the number of places where this occurs, and they also allow to find two scaling functions close to the point to be predicted, which do satisfy this condition.

Examples and extensions

Figure 17 on top shows noisy samples of the 'Heavisine' function

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

on the same grid as in Figure 5. The plot in the middle is the reconstruction from thresholded coefficients on six resolution levels, using cubic prediction and

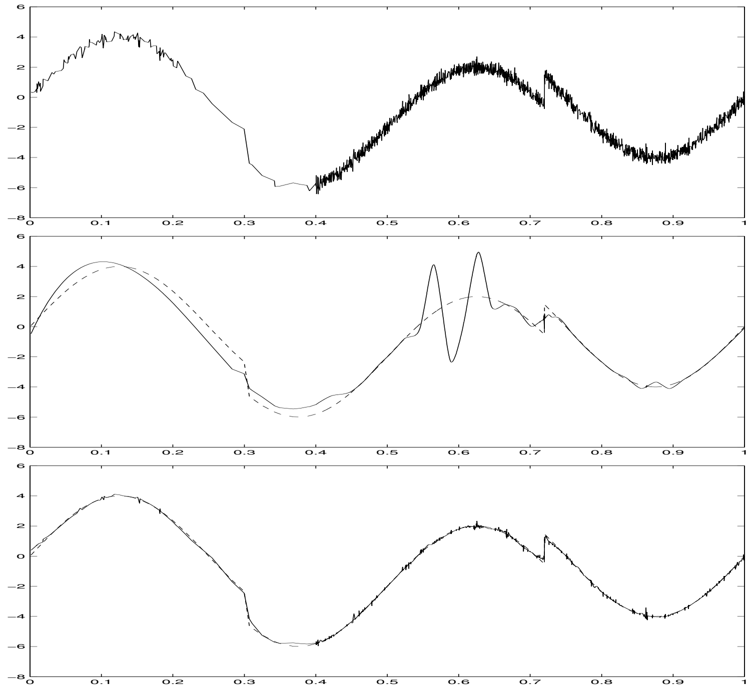


Figure 17: Top: noisy, highly non-equispaced samples of piecewise smooth signal. Middle: biased reconstruction from thresholded coefficients, using multiscale cubic polynomial prediction. 2016 out of 2048 coefficients were subject to thresholding. Bottom: stabilized transform using the two proposed methods. 2013 out of 2048 coefficients were subject to thresholding.

linear update. 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 in the bottom of Figure 17. 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 a brute threshold approach. More sophisticated coefficient selection could remove a great deal of this remaining noise.

Although the problem analysis had a four points prediction in mind, the method also solves the even worse instabilities at higher prediction orders. Figure 18 on top has the useless output from the non-stabilized transform with 8 dual vanishing moments. The proposed stabilizing methods bring the estimated function back to finite values, in a smooth and close fit.

6 Conclusions and directions of further research

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. It can also be seen as a combination of boundary effects and effects due to the non-equidistant

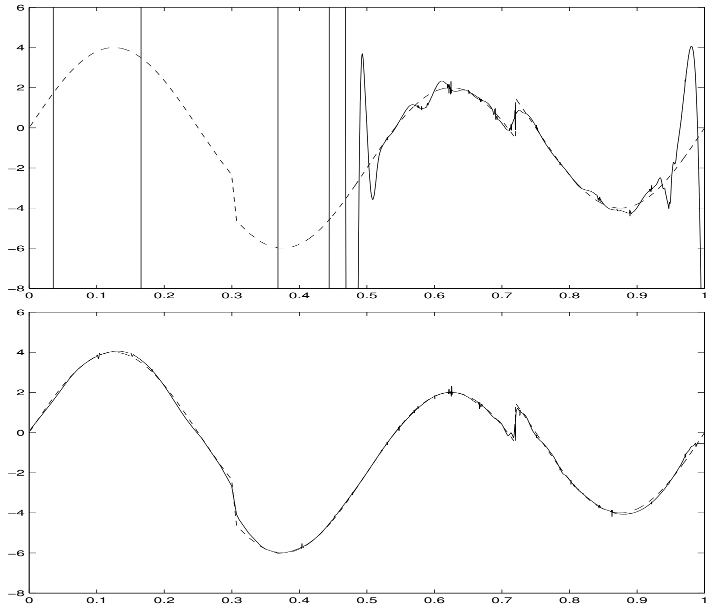


Figure 18: Output for a wavelet transform with 8 dual vanishing moments. Top: non-stabilized. Bottom: stabilized.

sampling.

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

1. the first modification is a relaxation of the prediction operation near the boundaries,

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

In addition, we also discussed a earlier stabilization in the update step [20], in the framework of our analysis.

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 two 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: when applying lifting to non-equispaced data, at least four issues are important:

1. Does the proposed scheme converge? In our rearranged transform, it is unclear what the effect is of skipping prediction points in order to prevent scale mixing.

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.

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 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 of Scientific Research Flanders (Belgium) (FWO) and postdoctoral researcher, Department of Electrical and Computer Engineering, Rice University, Houston, TX, and Department of Computer Science, K.U.Leuven, B-3000 Leuven, Belgium.

References

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

[2] I. Daubechies, I. Guskov, P. Schröder, and W. Sweldens. Wavelets on irregular point sets. *Phil. Trans. R. Soc. Lond. A*, 357(1760):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):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.*, 95:172–183, 2000.

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

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

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

[8] S. Sardy, D.B. Percival, A.G. Bruce, H-Y. Gao, and W. Stuetzle. Wavelet de-noising for unequally spaced data. *Statistics and Computing*, 9:65–75, 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):333–369, 1998.

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

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

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

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

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

[16] M. Jansen and A. Bultheel. Smoothing irregularly sampled signals using wavelets and cross validation. TW Report 289, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, April 1999.

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

[18] M. Jansen, M. Malfait, and A. Bultheel. Generalized cross validation for wavelet thresholding. *Signal Processing*, 56(1):33–44, January 1997.

[19] 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.

[20] J. Simoens and S. Vandewalle. On the stability of wavelet bases in the lifting scheme. TW Report 306, Department of Computer Science, Katholieke Universiteit Leuven, Belgium, 2000.

[21] R. L. Claypoole, R.G. 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.

[22] R. L. Claypoole, W. Davis, G. Sweldens, and R.G. Baraniuk. Nonlinear wavelet transforms for image coding. In *Proc. 31st Asilomar Conference*, Nov. 1997.

[23] M. H. Jansen, G. P. Nason, and B. W. Silverman. Second generation wavelets and empirical bayes estimates for scattered data smoothing. Technical report, Bristol University, 2001.