Smoothing Non-Equidistantly Sampled Data using Wavelets and Cross Validation

Maarten Jansen and Adhemar Bultheel

Department of Computer Science
K.U.Leuven
Celestijnenlaan 200A, B3001 Heverlee - Belgium
`phone: ++32 16 32 7080 fax: ++32 16 32 7996`
`e-mail: maarten.jansen@cs.kuleuven.ac.be`

Abstract

Coefficient thresholding is a popular method in wavelet based noise reduction. A wavelet decomposition is typically a sparse representation of noise-free signals: the essential information is captured by a limited number of large, important coefficients, while the main part of coefficients is close to zero. Replacing these small coefficients by zero is a straightforward way to reduce noise variance without affecting the noise-free signal too much. Recently, algorithms have been developed for wavelet decompositions of non-equidistant samples, using so called second generation wavelets. This paper investigates how to apply these algorithms to reduce noise in signals on a non-equidistant grid.

1 INTRODUCTION

In the last several years, wavelet thresholding has shown remarkable results in digital signal de-noising. A wavelet threshold procedure [2] starts with a discrete wavelet transform of the input. In a second step, coefficients beneath a certain threshold are replaced by zero. Inverse transform yields the result. The main issue in this procedure is the selection of the threshold. This parameter should be chosen so that the eventual result is as close as possible to the unknown noise-free signal. This paper uses the method of generalised cross validation [8, 9, 4], which does not need an estimate of the amount of noise.

Classical wavelet theory and algorithms are based on Fourier analysis, as a tool for filter design. A wavelet decomposition writes a function as a linear combination of dilations and translations of one function $\psi(x)$. This procedure algorithm requires equidistant samples. Recently, a new algorithm for wavelet transforms has been developed, called the *lifting scheme* [6]. Not only this algorithm is faster, but it also allows an extension of wavelet theory to the case of data on an irregular grid [7].

This paper investigates how to adapt a wavelet threshold procedure for coefficients on a non-equidistant grid.

2 LIFTING AND NON-EQUIDISTANT DATA

A classical wavelet decomposition algorithm has the structure of a repeated filter bank algorithm. In the first step, the input is convolved with a high pass filter $\tilde{\boldsymbol{g}}$ and a low pass filter $\tilde{\boldsymbol{h}}$. The result of the latter convolution is a smoothed version of the input. The high frequency part is captured by the first convolution, and normally contains a lot of very small numbers. Since these convolutions both give a result with a size equal to that of the input, this procedure doubles the total number of data. Therefore, the *Fast Wavelet Transform* (FWT) omits half of these data by *subsampling*. The resulting high frequency coefficients are wavelet coefficients at the *finest level* or *finest scale*. The low frequency output are scaling coefficients. In the next step the algorithm repeats the same procedure on this smoothed version of the input.

The basic building block of a wavelet transform is a filter bank with filters $\tilde{\boldsymbol{g}}$ and $\tilde{\boldsymbol{h}}$. This is represented in Figure 1. This figure also shows the reconstruction, corresponding to one step of the filter bank decomposition. This reconstruction starts by *upsampling* the rows of coefficients: a zero is put in between two elements. The next step is a convolution with filters $\boldsymbol{g}$ and $\boldsymbol{h}$. The results of these convolutions are added. Of course, some conditions on $\tilde{\boldsymbol{g}}$, $\tilde{\boldsymbol{h}}$, $\boldsymbol{g}$, and $\boldsymbol{h}$ are necessary to make a perfect reconstruction of the input possible from the resulting data. We do not go into detail on this problem.

The lifting scheme decomposes a filter bank operation in a number of consecutive lifting steps [1]. This series starts by splitting the input vector into points with odd and points with even index (Figure 2).

There exists two kinds of lifting operations. The first, called dual lifting, subtracts a filtered version of the "even" input to the "odd" input. The second, called primal lifting, adds a filtered version of the dual lifting output to the so far untouched "even" input. One way to interpret the dual lifting step, is the following: we assume that the "even" input and the "odd"

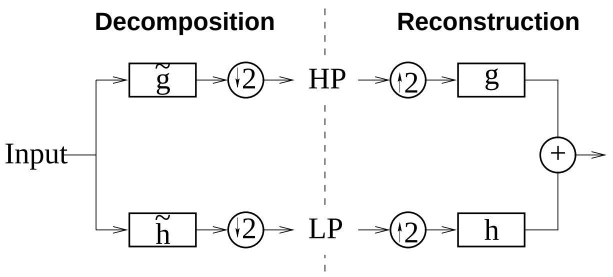

Figure 1: One step of a wavelet decomposition and its reconstruction. This is a filter bank: The input is filtered and downsampled to get a low pass signal LP and a high pass signal HP. Reconstruction starts with upsampling by introducing zeroes between every pair of points in LP and HP.

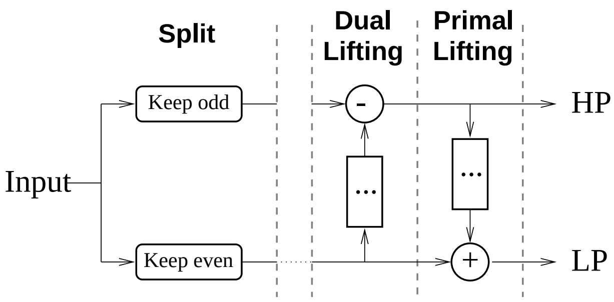

Figure 2: Decomposition of a filter bank into lifting steps. The first type of lifting is called *dual lifting* or a *prediction step*. The other type is *primal lifting* or *update*.

input are highly correlated. This is certainly true in the first lifting step, where the "even" and the "odd" input are directly originating from one input signal. We now try to *predict* the odd samples by a prediction operator (a filter) on the even ones. By subtracting this prediction from the odd samples, we reduce correlation. These differences are high-pass coefficients. An example of a prediction formula is an interpolating polynomial. The second lifting step is an update step. It can be interpreted as a way to preserve the average (and higher moments) in the low-pass coefficients.

The lifting philosophy is by no means limited to equidistant samples.

3 WAVELET THRESHOLDING

3.1 Threshold procedures

A wavelet transform has strong decorrelating properties, i.e. it uses the correlation between neighbouring samples to obtain a sparse representation of a noise-free signal: the essential information is captured by a limited number of large, important coefficients, while the main part of coefficients is close to zero. Replacing these small coefficients by zero is a straightforward way to reduce noise variance without affecting the noise-free signal too much.

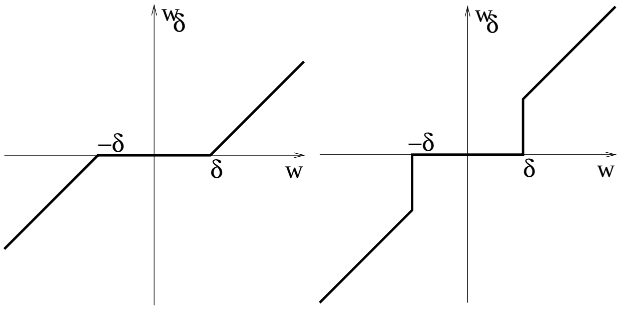

Figure 3: Noise reduction by wavelet shrinking. On the left: Soft-thresholding: a wavelet coefficient w with an absolute value below the threshold δ is replaced by 0. Coefficients with higher absolute values are shrunk. Right: Hard-thresholding: Coefficients with an absolute value above the threshold are kept.

Wavelet coefficients with a magnitude above a given threshold can be kept or shrunk: the first option is called *hard-thresholding*, illustrated in Figure 3, left. The most simple example of a wavelet *shrinkage* procedure is *soft thresholding*: wavelet coefficients with absolute value above a threshold are reduced towards zero with an amount equal to the threshold. While at first sight hard-thresholding may seem a more natural approach, soft-thresholding is a more continuous operation, and it is mathematically more tractable. Moreover, hard-thresholding often leads to a result with more spikes and artifacts.

3.2 Threshold selection

Threshold assessment is a central question in each wavelet threshold procedure. The ultimate goal is to produce a result as close as possible to the noise-free signal. Since this is unkown in practical applications, we try to find an approximation which is smooth and fits the input as well as possible. Note that a small threshold produces a noisy result close to the input. A big threshold makes a smooth result: it reduces variance, but on the other hand, it introduces bias. The optimal threshold makes the best compromise: it minimises the error of the output as compared to the unknown, noise-free signal.

This paper uses a method based on *Generalised Cross Validation* (GCV) [8, 9, 4]. In the framework of a threshold algorithm, GCV is a function of the threshold value δ:

$$GCV(\delta) = \frac{\frac{1}{N}\|\boldsymbol{w} - \boldsymbol{w_\delta}\|^2}{[\frac{N_0}{N}]^2}. \tag{1}$$

$\boldsymbol{w}$ is the vector of wavelet coefficients at the input, whereas $\boldsymbol{w_\delta}$ is the vector after applying threshold δ. N is the vector length and N_0 counts the number of zeroes in $\boldsymbol{w_\delta}$. The evaluation of this function only uses input data. It is important to note that no value or estimation of the noise variance σ^2 is needed. The minimiser of this function is an asymptotically optimal threshold in ℓ_2-sense, i.e. for N sufficiently large, the minimiser of $GCV(\delta)$ also minimises the mean square error of the result, as compared to the unknown, noise-free wavelet coefficients. For orthogonal wavelet transforms, this is equivalent to a mean square error minimisation of the result after inverse transform.

Estimation of the optimal threshold by minimising $GCV(\delta)$ can be done in a linear amount of time, and so, the threshold procedure has the same complexity as the fast wavelet transform.

4 THRESHOLDING SECOND GENERATION WAVELET COEFFICIENTS

While applying threshold procedures to second generation wavelet coefficients, we have to deal with two problems. The first of these two is related to the *noise*. Since the wavelet *analysis* takes into account the lattice on which the samples are defined, the threshold *estimation* step turns out to be more complicated.

The second problem has to do with the inverse transform or *synthesis*. This inverse transform may cause a considerable *bias* in the result, even if bias is relatively small in terms of wavelet coefficients. Part of this problem has to do with the threshold *procedure*: shrinking a large coefficient may have a wide range effect.

4.1 noise stationarity compensation

Obviously, each thresholding and threshold selection procedure assumes the noisy coefficients to be at least second order stationary. This means that all coefficients have the same noise variance: if the amount of noise is different for each coefficient, it is impossible to remove it decently by only one threshold. Whereas all threshold estimators implicitly or explicitly use the noise standard deviation, this becomes most clear in Donoho's and Johnstone's "universal threshold" [3]:

$$\delta = \sqrt{2\log(N)}\,\sigma. \tag{2}$$

It estimates the optimal threshold by δ, which is proportional to the amount of noise σ. N is the number of coefficients. For wavelets on a regular grid, it is easy to prove that the wavelet transform of second-order stationary noise is second-order stationary within each scale. This justifies a level-dependent threshold assessment [5]. In this case, an automatic threshold selection, i.e. a procedure that does not use a noise variance estimation explicitely, becomes even more advantageous.

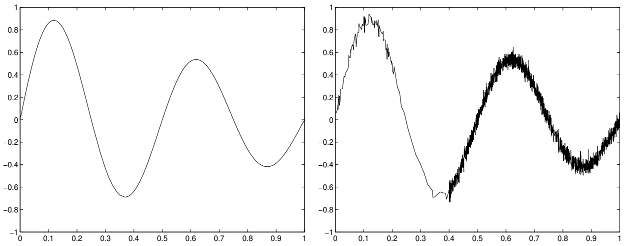

Figure 4: A test case: a damped sign ($f(x) = \mathrm{e}^{-x}\sin 4\pi x$) on an irregular grid and a noisy version of this signal.

Since stationarity is no longer preserved in the non-equidistant case, the procedure has to compute the variance fluctuation. A wavelet transform, classical or second generation, is a linear operation:

$$\boldsymbol{w} = W\boldsymbol{y}.$$

In this equation $\boldsymbol{y}$ is the input vector. If C is the correlation matrix of the input, then the correlation matrix in the wavelet domain equals:

$$D = WCW^T.$$

If C is a banded matrix, D can be computed in a linear amount of time. In most practical cases, the exact values of C are unknown, but the strucure of C may be known, up to constant. For instance, the case of white, stationary noise corresponds to $C = cI$, whit I the unit matrix.

We now apply the classical threshold procedure on the vector with elements $w_i/\sqrt{D_{ii}}$.

Figure 4 shows a damped sine ($f(x) = \mathrm{e}^{-x}\sin 4\pi x$) on an irregular grid and a noisy version of this signal. This irregular grid was constructed as follows: we choose approximately 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. We assume that the noise is additive, white and stationary. Figure 5 compares the results of a classical wavelet transform with the result of a threshold procedure on scaled coefficients from the corresponding second generation procedure. Second generation wavelets allow a smoother reconstruction, but our solution shows an awful bias.

4.2 The impact of thresholding

In a classical wavelet transform, each coefficient has about the same impact: in the orthogonal case, the inverse transform even preserves signal "energy" (the sum of squares). The big coefficients are also the important ones. In the second generation setting, a small coefficient may have a wide impact, especially when this coefficient is related to a region with only a few samples. Secondly, dividing wavelet coefficients

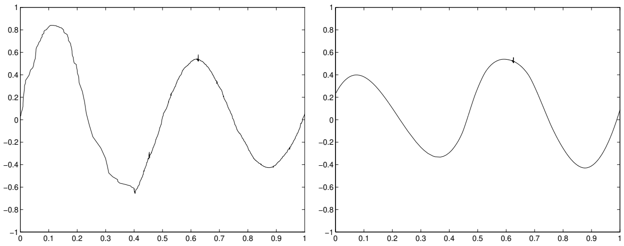

Figure 5: Left: result of classical wavelet transform. Right: result of a threshold procedure on scaled coefficients from a second generation procedure. The classical wavelet transform was obtained by applying the second generation procedure on a regular grid. Because this does not correspond to the real grid structure, the result is stil "noisy". On the other hand, the second generation procedure shows a terrible bias.

by their standard deviation may cause important coefficients to become relatively small. Thirdly, soft-thresholding a coefficient may cause a considerable bias in the result, even if this bias is small in terms of wavelet coefficients.

It seems almost impossible to modify the scaled coefficients according to their impact, while preserving stationarity. Therefore, we take an heuristic approach: we use the GCV-procedure to estimate a threshold for the equidistant transform, i.e. the transform of the input data without taking into account that the samples are non-equidsitant. For each coefficient w_i we compute the shrinking ratio $w_{\delta i}/w_i$. Next, we multiply every coefficient in the second-generation transform by this ratio. Inverse second-generation transform yields the result. This method combines the advantage of thresholding in optimal conditions with the smooth reconstruction of second-generation wavelets. Moreover, this approach does not require the computation of the variance for each coefficient. Thirdly, an extension to colored noise and level-dependent thresholding is quite obvious.

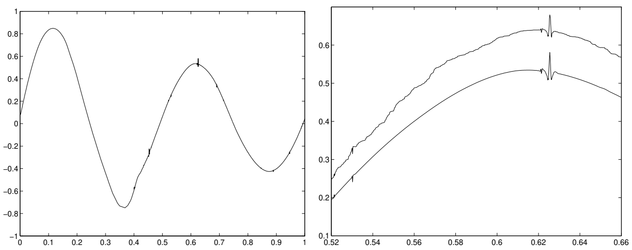

Figure 6: Left: result of heuristic approach, combining soft-thresholding for coefficients of an classical, equidsitant transform with the smooth second generation reconstruction. This method reduces bias. Right: detail of reconstruction with classical procedure (up) and 2nd generation wavelets (below).

5 CONCLUSIONS

We present an algorithm of linear complexity, that removes noise in non-equidistantly spaced data. A wavelet transform based on lifting takes into account the irregular structure of the input data and thereby introduces less artifacts than a classical transform on the same data.

Acknowledgements

This paper presents research results of 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 its authors. The first author is financed by a grant from the Flemish Institute for the Promotion of Scientific and Technological Research in the Industry (IWT).

References

[1] I. Daubechies and W. Sweldens. Factoring wavelet transforms into lifting steps. Technical report, Bell Laboratories, Lucent Technologies, 1996.

[2] D. L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[3] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

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

[5] I. M. Johnstone and B. W. Silverman. Wavelet threshold estimators for data with correlated noise. *Journal of the Royal Statistical Society, Series B*, 59:319–351, 1997.

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

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

[8] G. Wahba. *Spline Models for Observational Data*, chapter 4, pages 45–65. CBMS-NSF Regional Conf. Series in Appl. Math. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.

[9] N. Weyrich and G. T. Warhola. De-noising using wavelets and cross validation. In S.P. Singh, editor, *Approximation Theory, Wavelets and Applications*, volume 454 of *NATO ASI Series C*, pages 523–532, 1995.