Multiple wavelet threshold estimation by generalized cross validation for data with correlated noise

Maarten Jansen, Adhemar Bultheel
K.U.Leuven, Department of Computer Science

January 1997, Revision: May 1997

Abstract

De-noising algorithms based on wavelet thresholding replace small wavelet coefficients by zero and keep or shrink the coefficients with absolute value above the threshold. The optimal threshold minimizes the error of the result as compared to the unknown, exact data. To estimate this optimal threshold, we use Generalized Cross Validation. This procedure does not require an estimation for the noise energy. Originally, this method assumes uncorrelated noise. In this paper we describe how we can extend it to images with correlated noise.

keywords:
Noise reduction, wavelets, thresholding, cross validation, correlated noise

Contents

1 Introduction

Thanks to the combination of a nice theoretical foundation and the promising applications, wavelets have become a popular tool in many research domains. In fact, wavelet theory combines many existing concepts into a global framework. This new theoretical basis reveals new insights and throws a new light on several domains of applications.

One of these applications is image enhancement. In this manuscript we concentrate on the problem of noise reduction. Among other methods to suppress noise with wavelets, we distinguish two important classes:

1. The first type of algorithms, such as described in [1] uses a library of regular waveforms. These methods assume that the signal without noise fits well in this library, whereas the noisy contribution cannot be well represented. These methods try to select the regular waveform from the library that is as close as possible to the input.

2. The other class of methods has the following scheme: first the algorithm performs a wavelet transform. Then, in contrast to methods of the first class, it *manipulates* the wavelet coefficients. Finally an inverse transform yields — hopefully — a de-noised signal or image. This paper describes a method that belongs to the latter case.

The *manipulation* of the wavelet coefficients is mostly based on a *classification*. This classification is often *binary*: the coefficients are divided into two groups. The first group contains important, regular coefficients, while the other group consists of coefficients that were catalogued as "too noisy". These two groups are then processed in a different way. Noisy coefficients are often replaced by zero.

To classify wavelet coefficients, the procedure needs a *criterion* to distinguish noisy from regular coefficients. We mention some of the possible criteria:

1. The most straightforward procedure uses the *absolute values* of the coefficients as a measure of regularity: the most important coefficients are also the most regular ones. This method assumes that a regular signal or image can be represented by a small number of large coefficients. Donoho e.a. [2] showed that this method has statistical optimality properties.

2. Another class of methods computes the *correlation* between coefficients at successive scales [3]. These methods are based on the assumption that regular signal or image features show correlated coefficients at different scales, whereas irregularities due to noise do not.

3. A third class of methods is based on the characterization of the Lipschitz or Hölder regularity of a function by its (continuous) wavelet

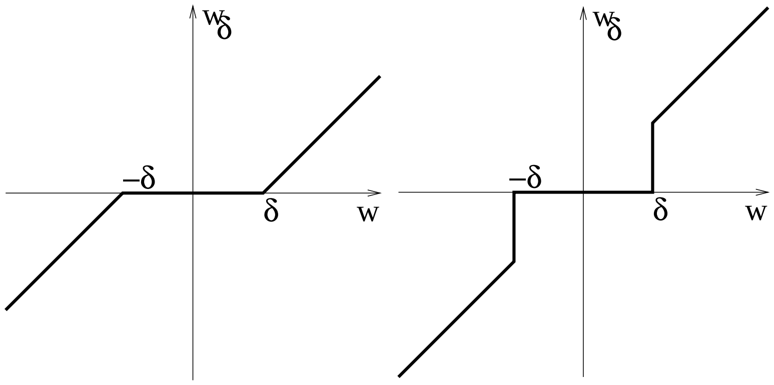

Figure 1: Noise reduction by wavelet shrinking. On the left: Soft-thresholding: a wavelet coefficient w with an absolute value below the threshold δ is replaced by 0. Coefficients with higher absolute values are shrunk. On the right: Hard-thresholding: Coefficients with an absolute value above the threshold are kept.

> transform [4, 5]. These methods look at the *evolution* of the coefficients across the different scales to distinguish regular from noisy contributions. Loosely spoken, a regular image or signal singularity has a long-term range and therefore the corresponding wavelet coefficients at coarse scales are large. Noise on the contrary, is local and therefore its singularities have larger coefficients at finer scales.

In principle, all these methods rely on a binary decision: a coefficient is affected by noise *or* sufficiently clean. To construct a more continuous approach, one can try to compute the *probability* for a coefficient to be sufficiently clean *according to* the criterion. To this end, we need an a priori probability model for regular wavelet coefficients. This Bayesian approach allows us to use the input data as an observation and so to compute for each coefficient the a posteriori probability to be sufficiently clean. We can use the a priori model to incorporate spatial coherence conditions [6, 7].

In this text, we restrict ourselves to simple threshold procedures, which replace the coefficients with small absolute value by zero. The "large" coefficients are kept in the hard-thresholding case and shrunk in the soft-thresholding case. Figure 1 shows the difference between these two. 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.

A natural question arising from this procedure is how to choose the threshold. Weyrich and Warhola [8] proposed to use a generalized cross validation (GCV) [9] algorithm. In [10] we showed that, under certain conditions, this procedure is asymptotically optimal, i.e. for a large number of data points it yields a threshold for which the result minimizes the expected mean square error as compared with the unknown, noise-free data. Moreover this GCV procedure does not need any value or estimation of the noise energy. Other cross validation procedures are discussed in [11].

In this paper we extend the use of generalized cross validation to the case of correlated noise. The classic shrinking procedure assumes white noise. Johnstone and Silverman [12] showed that a resolution-level dependent choice of the threshold allows to remove correlated noise. We thus investigate the possibility to choose these level-dependent thresholds by generalized cross validation.

This paper is organized as follows: we first repeat some basics about wavelets as far as we need it for further discussion. Then we explain the idea of generalized cross validation and we discuss the assumptions that are necessary for this method to be successful. In section 4, we discuss the properties of wavelet transforms in relation to data with correlated noise. We propose a modification of the generalized cross validation function for correlated noise, based on these properties. In Section 5 we introduce the redundant wavelet transform as a method to improve the results. In Section 6 we discuss some of the results and we end with a brief conclusion.

2 The discrete, dyadic, non-redundant wavelet transform

In this section we repeat some basic wavelet material, as far as we need it in our further discussion. For simplicity, we mainly restrict ourselves to a one-dimensional terminology. Extension to more dimensions is possible, for instance by a tensor product approach. For more information or a more complete overview, we refer to the extensive literature. We mention [13, 14, 15].

An *orthogonal, discrete, non-redundant* wavelet transform maps an array of numbers onto a new array of equal length. It tries to recombine the data in such a way that the result is a more compact representation of the information. To this end, the algorithm uses the local correlation between the input numbers. For instance, neighbour pixels in a digital image normally have approximately the same grey values.

In a first step the algorithm convolves the input with a lowpass filter $\boldsymbol{h}$, which results in the *scaling* coefficients. These coefficients are a smoothed version of the input data. The difference between this coarser scale representation and the original data is captured by a convolution of the input with a highpass filter $\boldsymbol{g}$. 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, we can omit half of these data by sub-sampling, as indicated in Figure 2. Of course, some conditions on $\boldsymbol{g}$ and $\boldsymbol{h}$ are necessary to make a

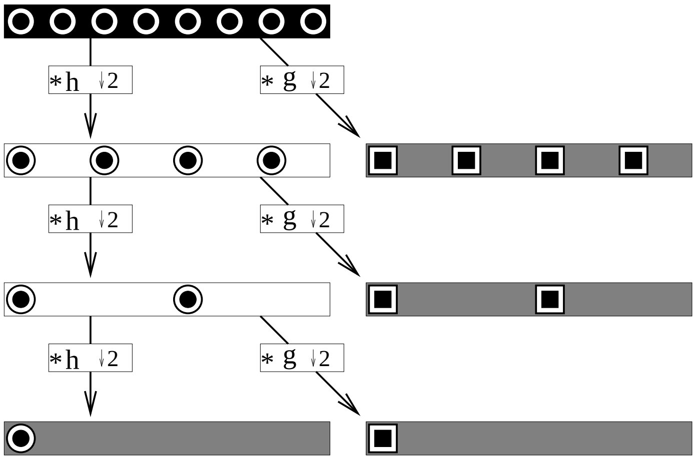

Figure 2: Successive steps of a fast decimated wavelet transform. After convolving the input with two filters $\boldsymbol{h}$ and $\boldsymbol{g}$, the algorithm drops half of the result (*down-sampling*). Under certain conditions, perfect reconstruction of the original signal is possible from this decimated transform

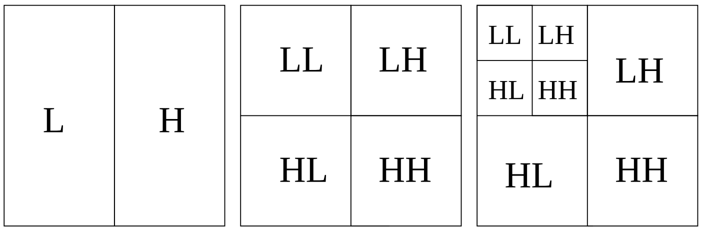

Figure 3: A two dimensional wavelet transform. First we apply one step of the one dimensional transform to all rows (left). Then, we repeat the same for all columns (middle). In the next step, we proceed with the coefficients that result from a convolution with $\boldsymbol{h}$ in both directions (right).

perfect reconstruction of the input possible from the resulting data. We do not go into detail on this problem.

In the second and following steps the algorithm repeats the same procedure on the reduced set of smoothed data. In two dimensions, we first apply one step on the row vectors and then on the column vectors. Figure 3 shows how this results in coefficients of four classes of coefficients. Coefficients that result from a convolution with $\boldsymbol{g}$ in both directions (HH) represent diagonal features of the image, whereas a convolution with $\boldsymbol{h}$ in one direction and with $\boldsymbol{g}$ in the other, reflects vertical and horizontal information (HL and LH). In the next step we proceed with the lowpass (LL) coefficients. Instead of proceeding with the LL-coefficients of the previous step only, we could also further transform all rows and all columns in each step. This leads to

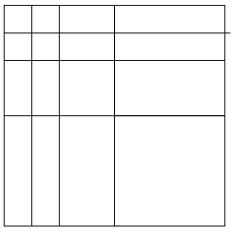

Figure 4: Grafical representation of wavelet coefficients after three steps of the rectangular wavelet transform: in each step all rows and all columns are completely further transformed.

the *rectangular* two dimensional wavelet transform, illustrated in Figure 4. This alternative not only requires more computation, it is also less usefull in applications: in the *square* wavelet transform, the HL and LH components contain more specific information on horizontal or vertical structures.

We now return to one dimension and suppose that the input vector $\boldsymbol{y}$ has length $N = 2^J$ for some integer J. If we denote by $\boldsymbol{w}$ the wavelet transform of $\boldsymbol{y}$, the linearity of this operation permits to write:

$$\boldsymbol{w} = W\,\boldsymbol{y}. \tag{1}$$

W is the wavelet transform matrix. Such a matrix-vector product has a quadratic complexity. However, the filter algorithm as described above only requires a linear amount of work, and is therefore called a *fast wavelet transform* (FWT). The inverse transform has also linear complexity and uses the same filters (h) and (g).

With each discrete wavelet transform, one can associate a pair of functions $\phi(x)$ and $\psi(x)$. The first is a smooth function with local support such that

$$y(x) := \sum_{k=0}^{2^J-1} y_k \phi(2^J x - k) \tag{2}$$

can be seen as a continuous representation of discrete data. $\phi(x)$ is called the *scaling* or *father* function. If we now denote by $w_{j,i}$ the ith wavelet coefficient in step j (j going from $J-1$ down to L), then, for an orthogonal wavelet transform, there exists one function $\psi(x)$ such that:

$$w_{j,i} = \langle y(x), \psi_{j,i}(x)\rangle = \int_{\mathbb{R}} y(x)\,\psi_{j,i}(x)\,\mathrm{d}x, \tag{3}$$

where

$$\psi_{j,i}(x) = 2^{j/2}\psi(2^j x - i).$$

The *mother* or *wavelet* function $\psi(x)$ has a typical waveform with a compact support. The function $y(x)$ can thus be written as:

$$y(x) := \sum_{j=L}^{J-1} \sum_{i=0}^{2^j-1} w_{j,i}\psi_{j,i}(x) + \sum_{k=0}^{2^L-1} s_{L,k}\phi(2^L x - k), \tag{4}$$

where $s_{L,k}$ represents the remaining scaling coefficients after the last step L. Each step of the discrete transform thus yields coefficients for different scales of the mother function. In this way, we get a multiresolution analysis of the original signal. If we continue as far as possible, i.e. until $L = 0$, we can write the function $y(x)$ as a DC-component plus combinations of dilations and translations of one function $\psi(x)$.

The decomposition in (4) remains unchanged if the wavelet transform is no longer orthogonal. However, in this case, we need a *dual* wavelet function $\tilde{\psi}(x)$ to find the coefficients like in (3):

$$w_{j,i} = \langle y(x), \tilde{\psi}_{j,i}(x) \rangle = \int_{\mathbb{R}} y(x)\, \tilde{\psi}_{j,i}(x)\, \mathrm{d}x. \tag{5}$$

The fast wavelet decomposition now uses *dual* filters $\tilde{\boldsymbol{h}}$ and $\tilde{\boldsymbol{g}}$, while $\boldsymbol{h}$ and $\boldsymbol{g}$ still appear in the reconstruction formula. This is the *biorthogonal* wavelet transform.

3 Shrinking and Generalized Cross Validation for white noise

We start from the following, additive model for data with noise.

$$\boldsymbol{y} = \boldsymbol{f} + \boldsymbol{\varepsilon}.$$

The vector $\boldsymbol{f}$ represents unknown, deterministic and structured data and $\boldsymbol{y}$ is the input for our algorithm. The noise $\boldsymbol{\varepsilon}$ is *stationary*. In principle, this means that all components of the vector should have the same distribution. In practice, we want the mean $\mathrm{E}\varepsilon_i$ and the variance $\sigma^2 = \mathrm{E}\varepsilon_i^2$ to be constants, i.e. independent of i. This is *second order stationarity*. In the first instance, we also assume that the noise is *white* or *uncorrelated*: $\mathrm{E}\varepsilon_i\varepsilon_j = \delta_{ij}\sigma^2$.

The algorithm starts with a wavelet transform. By the linearity of this operation, the model in terms of wavelet coefficients remains unchanged:

$$\boldsymbol{w} = \boldsymbol{v} + \boldsymbol{\omega}.$$

The vector $\boldsymbol{v} = W\,\boldsymbol{f}$ contains the wavelet coefficients of the original data, $\boldsymbol{\omega} = W\,\boldsymbol{\varepsilon}$ are the noise coefficients and $\boldsymbol{w} = W\,\boldsymbol{y}$. Only for orthogonal W the wavelet decomposition of the noise $\boldsymbol{\omega}$ is stationary and white. In Section 4 we discuss the biorthogonal case.

After applying a threshold δ, we get the modified wavelet coefficients $\boldsymbol{w_\delta}$, for which the inverse transform yields the restored data $\boldsymbol{y_\delta}$. As threshold value, we choose the minimizer of the following Generalized Cross Validation function:

$$GCV(\delta) = \frac{\frac{1}{N}\|\boldsymbol{y} - \boldsymbol{y_\delta}\|^2}{[\frac{N_0}{N}]^2}, \tag{6}$$

where N is the total number of wavelet coefficients and N_0 the number of these coefficients that were replaced by zero. This function only depends on input and output data. A priori knowledge about the amount of noise energy is not necessary. If we use orthogonal wavelet transforms, then we can compute this formula in the "wavelet-domain", and hence minimization can be done completely in this domain. In that case, the amount of work, due to this minimization is comparable to or even less than the number of computations, necessary for the wavelet transform.

In [10], we proved that this threshold choice is asymptotically optimal, i.e. for a large number of wavelet coefficients, the minimizer of $GCV(\delta)$ also minimizes the mean square error function (or risk function) $R(\delta)$, where

$$R(\delta) = \frac{1}{N}\|\boldsymbol{y_\delta} - \boldsymbol{f}\|^2. \tag{7}$$

More precisely, we have:

Theorem 1 *If $\delta^* = arg \min R(\delta)$ and $\tilde{\delta} = arg \min GCV(\delta)$, then for $N \to \infty$:*

$$\frac{\mathrm{E}R(\tilde{\delta})}{\mathrm{E}R(\delta^*)} \downarrow 1, \tag{8}$$

and in the neighbourhood of δ^*:

$$\mathrm{E}GCV(\delta) \approx \mathrm{E}R(\delta) + \sigma^2. \tag{9}$$

Figure 5 illustrates this principle.
To give the proof in [10], we had to make several assumptions:

1. The original data $\boldsymbol{f}$ are *smooth* in the sense that they can be represented compactly by taking a wavelet transform. In fact, this assumption justifies the use of wavelets, since the localizing properties of these basis functions guarantee such a compact representation for most inputs. Without this assumption, the wavelet transform would not be necessary.

2. We need an orthogonal wavelet transform.

3. As mentioned before, the noise should be second order stationary.

4. The noise should be white.

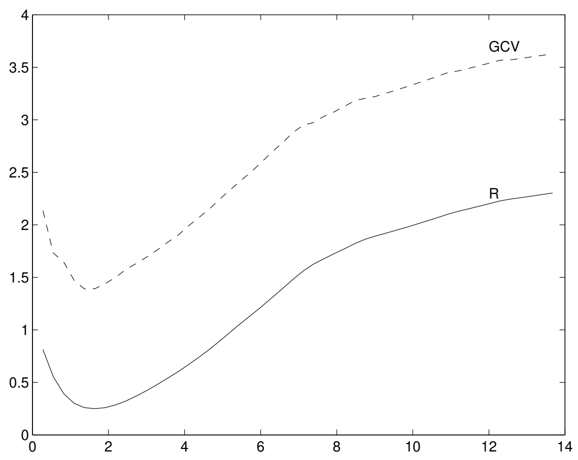

Figure 5: GCV and mean square error of the result in function of the threshold δ .

5. The noise should be Gaussian with zero mean. Experiments showed that, in practice, the GCV-method performs well for other zero mean stationary distributions of the noise.

6. We use soft-thresholding. The method fails in the hard-thresholding case, because this operation is not continuous, as illustrated in Figure 1.

4 Correlated noise

Experiments show that this Generalized Cross Validation procedure fails in cases with correlated noise. In this section, we study the properties of wavelet transforms in relation to correlated noise, we explain why a GCV-procedure needs uncorrelated noise and propose a modification of the method for the case of correlated noise.

We start with the following observation:

If $R = \mathrm{E}\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T$ is the correlation matrix of vector $\boldsymbol{\varepsilon}$ of random numbers, and $\boldsymbol{\omega} = W\,\boldsymbol{\varepsilon}$ is a linear transformation of this vector, then it is easy to proof that the correlation matrix S of this vector equals:

$$S = WRW^T. \tag{10}$$

If W is a wavelet transform, this says that the correlation matrix of the one-dimensional wavelet transform of a vector equals the *rectangular* wavelet transform of the correlation matrix of the input vector.

If W is orthogonal and $R = \sigma^2 I$, then we have that $S = \sigma^2 I$. This means that:

Observation 1 *The wavelet transform of* stationary *AND* white *noise is* stationary *AND* white.

However, if the noise is not stationary or not white, then the wavelet transform could be neither white nor stationary.

To prove that GCV yields the optimal threshold if the number of wavelet coefficients tends to infinity, we do not need uncorrelated wavelet coefficients at any moment, but we do need stationary noise in the wavelet domain [10]. This is because the GCV-estimator is built on the fact that an estimator for $R(\delta)$ is given by [16]:

$$\mathrm{SURE}(\delta) := T(\delta) + \sigma^2 \cdot \frac{2\mathrm{Tr}(D') - N}{N}. \tag{11}$$

In this equation we used $T(\delta)$ to mean:

$$T(\delta) = \frac{1}{N}\|\boldsymbol{w}_{\boldsymbol{\delta}} - \boldsymbol{w}\|^2, \tag{12}$$

and D' stands for a diagonal matrix, where $D'_{ii} = 0$ if the corresponding wavelet coefficient is replaced by zero and $D'_{ii} = 1$ otherwise. Thus $\mathrm{Tr}(D') = N_1$, where $N_1 = N - N_0$ is the number of wavelet coefficients that are not replaced by zero.

For proper use of this SURE estimator, we need stationarity. Anyway, it is obvious that a wavelet threshold method fails when the noise on the coefficients is not stationary. Indeed the optimal choice of the threshold depends on the present noise energy. The more noise, the higher the threshold should be. Some threshold selection methods, like the "universal threshold" of Donoho and Johnstone [17]:

$$\delta = \sqrt{2\log(N)}\,\sigma, \tag{13}$$

use this dependency explicitly. Equation (13) chooses a threshold *in proportion to σ*. If the amount of noise is different for all coefficients, it is difficult to remove it decently by only one threshold. As a matter of fact, this is also the reason why we do need orthogonal wavelet transforms: a non-orthogonal transform yields non-stationary noise.

We now suppose that the original noise is stationary and more precisely that the correlation between two points only depends on the distance between them. This means that the correlation matrix R is a (symmetric) Toeplitz matrix. If this is true, the multiresolution structure of a wavelet transform allows to prove that:

Lemma 1 *If $\omega_{j,i}$ represents a wavelet coefficient at place i and resolution level j (scale 2^{-j}) of a random vector $\boldsymbol{\varepsilon}$, then*

$$\mathrm{E}\omega_{j,i}^2 =: \sigma_j^2 \tag{14}$$

only depends on the resolution level j.

Proof: We only consider the one dimensional case here. Extension to more dimensions is straightforward.

Since the correlation matrix is symmetric Toeplitz, we have that $R_{i,j} = r_{|i-j|}$.

The wavelet coefficients at the finest resolution level are then:

$$\begin{aligned} \mathrm{E}\omega_{J-1,k}\omega_{J-1,l} &= \sum_i \sum_j g_{i-2k} g_{j-2l} \mathrm{E}\varepsilon_i \varepsilon_j \\ &= \sum_i \sum_j g_{i-2k} g_{j-2l} r_{|i-j|}. \end{aligned}$$

Substitutions $m = i - 2k$ and $n = j - 2l$ then yield that:

$$\mathrm{E}\omega_{J-1,k}\omega_{J-1,l} = \sum_m \sum_n g_m g_n r_{|2(k-l)+m-n|}.$$

From this formula, it follows immediately that for all integer r:

$$\mathrm{E}\omega_{J-1,k+r}\omega_{J-1,l+r} = \mathrm{E}\omega_{J-1,k}\omega_{J-1,l}.$$

In particular, we have that:

$$\mathrm{E}\omega^2_{J-1,k+r} = \mathrm{E}\omega^2_{J-1,k} =: \sigma^2_{J-1}.$$

A similar argument holds for the scaling coefficients at resolution level $J-1$. We can thus repeat the same procedure for the wavelet coefficients at coarser levels, thereby completing the proof. $\square$

We have proven that the wavelet transform of stationary correlated noise is stationary within each resolution level. Since stationarity is a condition for a successful GCV-estimation of the optimal threshold, this result suggests choosing a different threshold for each resolution level. In our proof, we did not use orthogonality. This means that a biorthogonal wavelet transform of white or colored, stationary noise is stationary within each resolution level.

The mean square error now becomes a function of a vector of thresholds $\boldsymbol{\delta}$. If $\boldsymbol{w_j}$ denotes the vector of wavelet coefficients at resolution level j, then we can write:

$$R(\boldsymbol{\delta}) = \sum_j \frac{N_j}{N} R_j(\delta_j), \tag{15}$$

where N_j represents the number of wavelet coefficients on level j and

$$R_j(\delta_j) = \frac{1}{N_j} \|\boldsymbol{w_{j,\delta}} - \boldsymbol{v_j}\|^2. \tag{16}$$

Equation (16) suggests an optimization in wavelet domain. Only for orthogonal wavelets, this is equivalent to minimization of $R(\boldsymbol{\delta})$ as defined in (7). Since all terms in equation (15) are positive, minimisation of $R(\boldsymbol{\delta})$ is equivalent to successive one dimensional minimisations of $R_j(\delta_j)$ for all j.

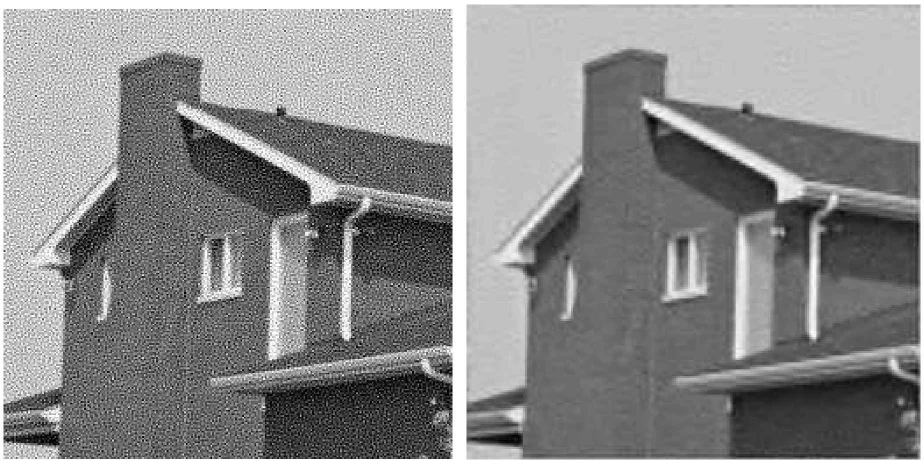

Figure 6: Left: an image with artificial, correlated noise. The noise is the result of a convolution of white noise with FIR highpass filter. Right: the result after level-dependent wavelet thresholding. We use Haar-wavelets.

A similar argument as in [10] leads to an estimation

$$\mathrm{SURE}_j(\delta_j) := T_j(\delta_j) + \sigma_j^2 \cdot \frac{2\mathrm{Tr}(D') - N_j}{N_j}, \tag{17}$$

with:

$$T_j(\delta_j) = \frac{1}{N_j} \|\boldsymbol{w_{j,\delta}} - \boldsymbol{w_j}\|^2.$$

Based on this estimator, we can construct:

$$GCV_j(\delta_j) = \frac{\frac{1}{N_j}\|\boldsymbol{w_j} - \boldsymbol{w_{j,\delta}}\|^2}{[\frac{N_{j0}}{N_j}]^2}, \tag{18}$$

as a function of which the minimum is an asymptotically optimal estimator for the minimum risk threshold.

In two dimensions, we minimize a GCV-function not only at each scale but also for each of the three components (horizontal, vertical, diagonal) at each resolution level.

We now illustrate the procedure with a testcase. To a clean image we added artificial colored noise. This noise was the result of a convolution of white noise with a FIR-highpass-filter. The signal-to-noise ratio is 4.56 dB. Figure 6 shows that the algorithm achieves a signal-to-noise ratio of 14.77 dB. Figure 7 compares the GCV-function with the mean square error for the vertical component at the one but finest resolution level.

The idea of a level-dependent threshold for data with correlated noise also appears in a paper by Johnstone and Silverman [12]. These authors

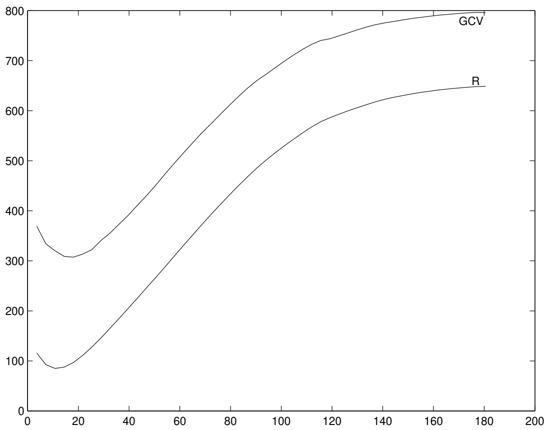

Figure 7: Mean square error and Generalized Cross Validation for vertical component coefficients at the one but finest resolution level.

start from the "universal threshold" by Donoho and Johnstone [17]:

$$\delta = \sqrt{2\log(N)}\,\sigma. \tag{19}$$

This formula essentially chooses the threshold in proportion to the amount of noise. If, according to (14), the amount of noise is level-dependent, it is straightforward to make the threshold selection level-dependent too.

Also for data with white noise, a level-dependent threshold may turn out to be better, since it is more adaptive.

5 Redundant Wavelet Transforms

Although a level dependent threshold selection works fine for correlated noise, problems may occur from the fact that the GCV-estimation is only asymptotically optimal. Indeed, the number of available wavelet coefficients decreases if the scale get coarser, because of the sub-sampling step in the wavelet transform algorithm. In this section, we discuss a redundant alternative for the classical wavelet transform that deals with this problem. This alternative also has other advantages, which we discuss below. Several authors have introduced the same modification [18, 19, 7].

Such a redundant transform results from omitting the sub-sampling step in our procedure (See Figure 8). Of course this transform should be *consistent* with the decimated transform in the sense that all the decimated coefficients re-appear in our new transform. To compute, for instance, the wavelet coefficients on the one but finest resolution level, we cannot, like in the decimated case, just convolve the scaling coefficients of the previous step with the high frequency filter $\boldsymbol{g}$. If we want to get the original coefficients

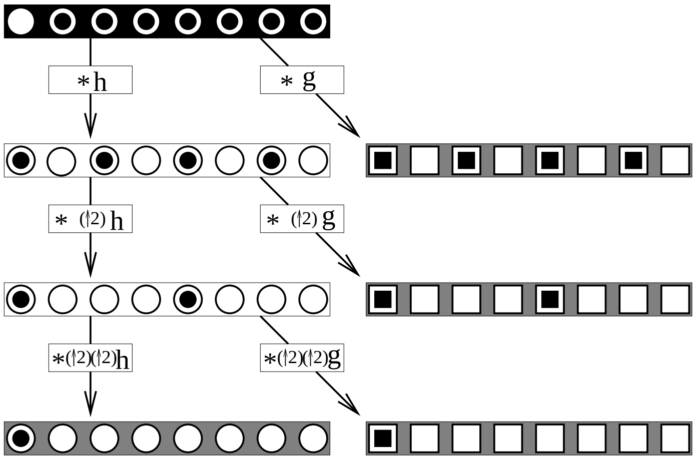

Figure 8: The redundant wavelet transform. The points with a black center represent coefficients that also appear in the decimated transform. To be consistent with this decimated transform, we should make sure that we only combine intermediate results from the original transform in our computation of coefficients "with a black center". To this end, we insert at each level, new zero elements between the filter coefficients of the previous step. This up-sampling operation is represented by $(\uparrow 2)$.

among our redundant set, we have to skip the extra coefficients of the previous step before the actual convolution. Of course these extra coefficients serve in their turn to complete the redundant set of wavelet coefficients at the given resolution level. A similar procedure is necessary for the computation of the scaling coefficients at the next level. At each level, the number of coefficients to skip, increases as a power of two minus one. As a matter of fact, instead of sub-sampling the coefficients, this alternative introduces up-sampling of the filters $\boldsymbol{h}$ and $\boldsymbol{g}$. Indeed, the wavelet and scaling coefficients at a certain resolution level can be seen as the result of a convolution with filters that are obtained by inserting zeros between the filter coefficients of the previous step. This adaptation preserves the multiresolution character of the wavelet transform. More precisely, in one dimension the coefficients of decimated wavelet transform of a discrete signal $\boldsymbol{y}$ are equal to:

$$w_{j,i} = \langle y(x), 2^{j/2}\psi(2^j x - i)\rangle = 2^{j/2}\int_{\mathbb{R}} y(x)\,\psi(2^j x - i)\,\mathrm{d}x, \qquad (20)$$

where $\psi(x)$ is the mother-wavelet-function and $y(x)$ is the function associated with the input vector $\boldsymbol{y}$ as described in (2). The redundant set is then:

$$w_{j,i} = \left\langle y(x), 2^{j/2}\psi(2^j x - 2^{j-J} i)\right\rangle = 2^{j/2}\int_{\mathbb{R}} y(x)\,\psi(2^j x - 2^{j-J} i)\,\mathrm{d}x, \quad (21)$$

Figure 9: Result of level dependent wavelet thresholding on the redundant wavelet transform of the image with noise in Figure 6. Signal-to-noise ratio is now 17.52 dB.

The functions $\psi(2^j x - 2^{j-J} i)$ do not constitute a basis, since they are not linearly independent. This set of functions is an example of a more general theory of over-complete representations or *frames*. For more details, we refer to [14, Chapter 3].

It is easy to prove that a redundant wavelet transform of stationary noise is still stationary within each scale.

We now have the same number of coefficients at all levels. This number is equal to the size of the original input data. This procedure guarantees a successful application of an asymptotic estimator.

Secondly, we know that in each step, we could omit one half of the (wavelet and scaling) coefficients before reconstruction of the scaling coefficients at the previous level. This means that these coefficients can be reconstructed in two different ways. If we manipulate the wavelet coefficients, for instance to remove noise, then the result will probably not be an exact redundant wavelet transform of one function. As a consequence the two possible reconstruction schemes at each level generate two different scaling coefficients at the previous level. Experiments show that taking a linear combination of these two possibilities causes an extra smoothing. Figure 9 illustrates this effect for an inverse transform that takes at every scale the mean value of the two reconstructions.

Thirdly, this redundant transform is immediately extensible for cases where the number of data is not a power of two.

Moreover, unlike the decimated transform, this redundant transform is translation invariant.

Figure 10: Aerial photograph with noise (512×512 pixels).

6 Results and discussion

We now illustrate the method for two "realistic" images: the first is an aerial photograph of 512×512 pixels. As can be expected, the algorithm does not distinguish real noise from the apparently noisy texture in the foliage of the trees in the wood. The second image is an MRI-image of a knee. Although wavelet thresholding is a very simple noise removal strategy, the result is quite fair. More sofisticated methods do not use the absolute value of the coefficients to distinguish between noise and regular image or signal. If there exists for these more complicated criteria a possibility to incorporate a GCV-selection procedure, we still do not know whether a level-dependent application makes sense in this case.

7 Conclusion

This paper has presented an automatic selection of level-dependent thresholds for wavelet coefficients, to remove stationary, correlated noise. We use the multiresolution character of a wavelet transform to justify the choice of one threshold per multiresolution level. Most existing algorithms use an estimation of the noise energy at all scales and all components (horizontal, vertical, diagonal) to choose these thresholds. Since we need one threshold for each scale and each component the benefits of an automatic selection of

Figure 11: Result of level-dependent wavelet thresholding for the aerial photograph.

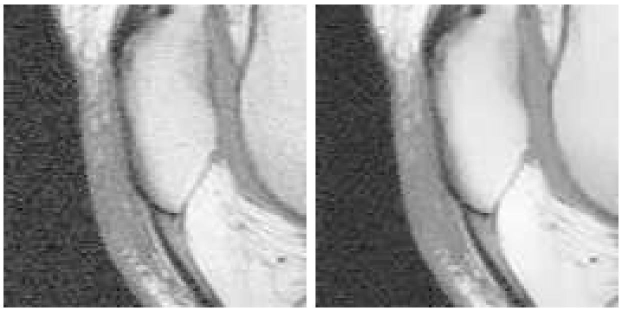

Figure 12: Left: MRI image with noise. Right: result after thresholding

these thresholds are still more important than for the choice of one threshold, like in [10]. However, our selection is based on an asymptotic method, which fails at coarse scales, since only a few wavelet coefficients correspond to these levels. Therefore, we introduced the redundant wavelet transform. This alternative provides, at the expense of more computation time and higher memory requirements, sufficiently coefficients at all scales.

Acknowledgement

This research was financed by a grant from the Flemish Institute for the Promotion of Scientific and Technological Research in the Industry (IWT).

References

[1] M. V. Wickerhauser, *Adapted Wavelet Analysis from Theory to Software*, A. K. Peters, Wellesley, MA, 1994.

[2] D. L. Donoho and I. M. Johnstone, "Wavelet shrinkage: Asymptopia?," *Journal of the Royal Statistical Society, Series B*, vol. 57, no. 2, pp. 301–369, 1995.

[3] 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*, vol. 3, no. 6, pp. 747–758, 1994.

[4] S. Mallat and W. L. Hwang, "Singularity detection and processing with wavelets," *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 617–643, 1992.

[5] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 710–732, 1992.

[6] M. Malfait, "Using wavelets to suppress noise in biomedical images," in *Wavelets in Medicine and Biology*, A. Aldroubi and M. Unser, Eds. April 1996, pp. 191–208, CRC Press, Boca Raton, FL.

[7] M. Malfait and D. Roose, "Wavelet based image denoising using a markov random field a priori model," *IEEE Transactions on Image Processing*, vol. 6, no. 4, pp. 549–565, 1997.

[8] N. Weyrich and G. T. Warhola, "De-noising using wavelets and cross validation," in *Approximation Theory, Wavelets and Applications*, S.P. Singh, Ed., 1995, vol. 454 of *NATO ASI Series C*, pp. 523–532.

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

[10] M. Jansen, M. Malfait, and A. Bultheel, "Generalized cross validation for wavelet thresholding," *Signal Processing*, vol. 56, no. 1, pp. 33–44, January 1997.

[11] G. P. Nason, "Wavelet shrinkage using cross validation," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 463–479, 1996.

[12] I. M. Johnstone and B. W. Silverman, "Wavelet threshold estimators for data with correlated noise," *Journal of the Royal Statistical Society, Series B*, vol. To Appear, 1997.

[13] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.

[14] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Series in Appl. Math., Vol. 61. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1992.

[15] G Strang and T. Nguyen, *Wavelets and filter banks*, Wellesley-Cambridge Press, Box 812060, Wellesley MA 02181, fax 617-253-4358, 1996.

[16] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, 1995.

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

[18] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Wavelets and Statistics*, A. Antoniadis and G. Oppenheim, Eds., 1995, Lecture Notes in Statistics, pp. 281–299.

[19] J.-C. Pesquet, H. Krim, and H. Carfantan, "Time invariant orthonormal wavelet representations," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 1964–1970, 1996.

Correspondence address:
Maarten Jansen
K.U.Leuven
Department of Computer Science

Celestijnenlaan 200A
B-3001 Heverlee
BELGIUM
phone ++32 16 32 7080; Fax ++32 16 32 7996
e-mail `maarten.jansen@cs.kuleuven.ac.be`
`http://www.cs.kuleuven.ac.be/~maarten/`