A Correlation-Dependent Model for Denoising via Nonorthogonal Wavelet Transforms

Kathrin Berkner and Raymond O. Wells, Jr.
Computational Mathematics Laboratory
Department of Mathematics, Rice University
Houston, TX 77005-1892, USA
berkner@math.rice.edu, wells@rice.edu

June 1998

Abstract

In many applications it is desirable to study nonorthogonal wavelet transforms. A translation-invariant wavelet transform is a nonorthogonal variant of the classical wavelet transform which plays an important role in denoising algorithms. However, it has been observed in many experiments that the thresholding scheme for the orthogonal DWT should be slightly modified for use in the translation-invariant setting. These observations motivate us to study denoising schemes for nonorthogonal wavelet transforms. In this paper we derive a thresholding scheme for denoising that incorporates correlations between nonorthogonal wavelet coefficients and specify these thresholds for translation-invariant and biorthogonal wavelet systems. The new scheme includes a scale- and wavelet-dependent threshold which is larger than the one normally used in combination with the orthogonal discrete wavelet transform.

Keywords

Nonorthogonal wavelet transforms, Gaussian noise, denoising.

EDICS Numbers: SP-2.4.4 Wavelets and filter banks, SP-2.7.2 Nonlinear analysis

Technical Report CML TR 98-07, Computational Mathematics Laboratory, Rice University.
This work was supported in part by DARPA/AFOSR F49620-97-1-0513, and Alexander von Humboldt-Stiftung.

I. Introduction

In many applications in signal and image processing the observed data are influenced by noise and the question arises of how to estimate the underlying "clear" signal from the noisy observations. An important step in solving this problem was achieved by the work of Donoho and Johnstone ([12], [13]) using wavelet transforms of the noisy data. They studied the situation of detecting a signal in additive Gaussian white noise. In particular, the basic idea involved special thresholding schemes for the noisy wavelet coefficients using properties of orthogonal wavelet transforms and Gaussian random variables. Using the inverse wavelet transform, an estimate for the original signal was obtained and bounds on the L^2-error and smoothness were presented ([12], [13]). The advantage of this denoising method is that it does not depend on the underlying function or on the specific chosen wavelet or the level of decomposition in the wavelet tree, but only on the sample size of the given data.

It is shown in [7], [19] that the use of a translation-invariant wavelet transform improves the denoising mentioned above using the classical orthogonal wavelet transform. Experiments in [19] have also indicated that a different threshold than the one for the Donoho-Johnstone model achieved better results. This modified threshold was found experimentally and was adapted to a special wavelet system and a special data set. These observations have lead to the question we want to study in this paper: Can the Donoho-Johnstone thresholding model be simply transfered to nonorthogonal wavelet transforms, do we have to modify the model, and can we explain the phenomena observed in [19]? The investigation of this problem is also motivated by the recent results on multiwavelet-denoising ([29]) which involves the study of covariance matrices in order to calculate averaged variances.

In this paper, we present a new thresholding scheme for noisy wavelet coefficients of nonorthogonal wavelet transforms. This scheme is based on the results proposed in [1] where the authors derived a smoothness characterization for an estimate of a function after scale dependent soft-threshold denoising with translation-invariant wavelet transforms. The new framework provides a general approach for denoising via discrete wavelet transforms that includes various nonorthogonal as well as classical orthogonal discrete wavelet transforms. Our approach utilizes a new scale- and wavelet-dependent thresholding scheme for nonorthogonal wavelet coefficients. This modification of the uniform thresholding scheme in [12], [13] is due to the correlations between wavelet matrix components.

This paper is organized as follows. In Section 2 we give a brief review on the theory of orthogonal, biorthogonal and translation-invariant wavelet transforms and introduce the necessary definitions. In Section 3 we introduce the Donoho-Johnstone denoising model for general wavelet transforms and derive conditions on thresholds such that these thresholds are sufficient to remove the noise in wavelet coefficients obtained from nonorthogonal wavelet transforms. Specific thresholds for biorthogonal and translation-invariant wavelet transforms are derived in Section 4. The correlation structure of these transforms is investigated in Section

5. Section 6 discusses numerical experiments using the new model including hard- and soft-thresholding techniques. The paper is concluded in Section 7.

II. Decimated and undecimated wavelet transforms

A. The orthogonal wavelet transform

This paper uses the following notation for an orthogonal wavelet system. An orthogonal wavelet system consists of a scaling function ϕ and a wavelet function ψ, such that ϕ satisfies the scaling equation

$$\phi(x) = \sqrt{2}\sum_{k\in Z} h_k \phi(2x-k)$$

where ϕ and ψ are related by

$$\psi(x) = \sqrt{2}\sum_{k\in Z} g_k \phi(2x-k)$$

with scaling filter $\mathbf{h} = \{h_k : k \in \mathbb{Z}\}$ and wavelet filter $\mathbf{g} = \{g_k : k \in \mathbb{Z}\}$. Throughout this paper, we assume ϕ to be compactly supported. This implies that the filter $\mathbf{h}$ and $\mathbf{g}$ are finite.

Defining $\phi_{j,k}(x)$ and $\psi_{j,k}(x)$ as $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x-k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x-k)$ the orthogonality conditions on ϕ and ψ are written as

$$\begin{aligned} \langle\phi,\phi(\cdot-k)\rangle &= \delta_{0,k} \\ \langle\phi_{0,l},\psi_{j,k}\rangle &= 0 \\ \langle\psi_{m,l},\psi_{j,k}\rangle &= \delta_{m,j}\delta_{l,k} \quad \text{for all } j,k,l,m\in\mathbb{Z}. \end{aligned} \tag{2.1}$$

The scaling function ϕ induces a multiresolution analysis $\{V_m\}_{m\in\mathbb{Z}}$ where $V_m := \mathrm{span}_{L^2}\{\phi_{m,k}\}_{k\in Z}$. The orthogonal complement of V_m is denoted by W_m and satisfies the condition

$$V_{m+1} = V_m \oplus W_m \quad \text{and} \quad V_m \perp W_m.$$

The set of all translations and dilations $\{\psi_{j,k}, j,k\in\mathbb{Z}\}$ of the wavelet function ψ, forms an orthonormal basis of $L^2(\mathbb{R})$. As a consequence, a function $f\in L^2(\mathbb{R})$ can be expanded in a *wavelet series* by

$$f = \sum_{j,k}\langle\psi_{j,k},f\rangle\psi_{j,k}.$$

Using the scaling function ϕ, the function f can be split into a projection $\sum_k\langle\phi_{j,k},f\rangle\phi_{j,k}$ onto the subspace V_j and projections $\sum_k\langle\psi_{m,k},f\rangle\psi_{m,k}$ onto the subspaces $W_m, m\geq j$, by

$$f = \sum_k\langle\phi_{j,k},f\rangle\phi_{j,k} + \sum_{m\geq j}\sum_k\langle\psi_{m,k},f\rangle\psi_{m,k}. \tag{2.2}$$

For further details on orthogonal wavelet systems we refer, e.g., to [9], [24].

When scaling coefficients at one scale j_0 are given, then the wavelet coefficients at coarser scales can be computed via a pyramid transform algorithm, the *discrete wavelet transform* (DWT). For an overview on the DWT and related subjects we refer, e.g., to [5], [24], [28]. If the DWT is applied to a sequence of N sampled data points, the computational complexity of this algorithm is $O(N)$.

The DWT of a data vector $\mathbf{x} \in \mathbb{R}^{\mathbb{N}}$ is represented by an orthogonal $N \times N$ matrix W. The vector of wavelet coefficients $\mu = (\mu_1, \ldots, \mu_N) \in \mathbb{R}^{\mathbb{N}}$ of the vector $\mathbf{x} \in \mathbb{R}^{\mathbb{N}}$ using the DWT is then simply given by

$$\mu = W\mathbf{x}. \tag{2.3}$$

The matrix W usually includes the corrections due to boundary treatments such as periodization or reflection (see, e.g., [5], [24]). In order to focus on the main features of the wavelet matrix we neglect those corrections in the remainder of this paper. Since W is an orthogonal matrix there exists an inverse transform $W^{-1} := W^T$ and the original data are obtained by

$$\mathbf{x} = W^{-1}\mu. \tag{2.4}$$

The DWT can be implemented as a filter bank using the filter $\mathbf{h}$ as a low-pass and the filter $\mathbf{g}$ as a high-pass filter. The coefficients h_k and g_k are linked by $g_k = (-1)^k h_{1-k}$. The orthogonality conditions on ϕ and ψ in (2.1) yield the well-known relations

$$\sum h_k h_{k+2l} = \delta_{0,l} \quad \text{and} \quad \sum h_k g_{k+2l} = 0 \quad \text{for all} \quad l \in \mathbb{Z}. \tag{2.5}$$

For the DWT we call the number of iterations of the filter bank the *scale* or the *level of resolution.*

B. The biorthogonal wavelet transform

A transform based on biorthogonal wavelet systems is constructed similar to the orthogonal wavelet transform, but using different scaling and wavelet functions for decomposition and reconstruction. The duals of the given scaling and wavelet functions are typically denoted by $\tilde{\phi}$ and $\tilde{\psi}$, resp.

For biorthogonal wavelet systems the orthogonality conditions in (2.1) are then modified to

$$\begin{aligned} \langle \tilde{\phi}, \phi(\cdot - k) \rangle &= \delta_{0,k} \\ \langle \tilde{\phi}_{0,l}, \psi_{j,k} \rangle &= 0 \\ \langle \tilde{\psi}_{m,l}, \psi_{j,k} \rangle &= \delta_{m,j}\delta_{l,k} \quad \text{for all } j,k,l,m \in \mathbb{Z} \end{aligned} \tag{2.6}$$

and in terms of the discrete filter this implies

$$\sum \tilde{h}_k h_{k+2l} = \delta_{0,l} \quad \text{and} \quad \sum \tilde{h}_k g_{k+2l} = 0 \quad \text{for all} \quad l \in \mathbb{Z}.$$

The DWT of a vector $\mathbf{x} \in \mathbb{R}^{\mathbb{N}}$ using a biorthogonal wavelet system is represented by an $N \times N$ matrix W, and the inverse transform by a matrix $\widetilde{W}$ with

$$\widetilde{W}W = I \quad \text{but} \quad \widetilde{W} \neq W^T.$$

There exist several families of biorthogonal wavelet systems such as spline-based and almost-orthogonal wavelet systems ([6]) or biorthogonal Coiflets ([30], [31]). Biorthogonal wavelets are the common bases used in compression. In denoising biorthogonal wavelet systems are used less frequently. For further details on biorthogonal wavelet systems we refer, e.g., to [6], [24].

C. The translation-invariant discrete wavelet transform

One disadvantage of the classical decimated DWT is its lack of shift invariance. As it is shown, e.g. in [19], the wavelet decomposition of a signal can differ enormously when simply shifting the signal by one data point. To eliminate this disadvantage the translation-invariant discrete wavelet transform (TIDWT) was introduced independently by various authors in [3], [17], [21], [26]. This transform is also known as the undecimated wavelet transform. The transform does not include down-sampling at each scale, but computes all shifts of the classical orthogonal DWT. It is shown, e.g., in [26] that the computation of the TIDWT can be done by an algorithm with order of complexity $O(N \log N)$. More specifically, the algorithm for calculating the TIDWT computes not all, but only special shifts, depending on the level of resolution. On the first level $j = 1$ only the shifts of size $h = 0, 1$ have to be considered, on the second level $j = 2$ all the shifts of size $h = 0, 1, 2, 3$ and so on (see e.g. [26]). A translation-invariant wavelet transform can also be obtained by considering shifts of biorthogonal wavelet transforms, but in this paper we focus on the shifts of the orthogonal DWT and restrict the wavelet systems to the Daubechies family. The results obtained in the remainder of this paper can be easily transfered to the biorthogonal translation-invariant situation.

Given a maximal level of decomposition L, the TIDWT is represented by an $(L+1)N \times N$ matrix M_L defined by

$$M_L = (R_1^T, \dots, R_L^T, S_L^T)^T \tag{2.7}$$

where R_j is the $N \times N$ matrix which yields the wavelet coefficients at scale j, and S_L is the matrix which produces the scaling coefficients at the coarsest scale L. Since the matrix M_L is not a square matrix no inverse is defined. Consequently, the matrix M_L is not an orthogonal matrix as given above for the orthogonal DWT. But there exists the *pseudo-inverse* or *Moore-Penrose inverse* (see e.g. [27]) of M_L

$$M_L^\dagger = (M_L^T M_L)^{-1} M_L^T$$

which satisfies the following equality

$$M_L^\dagger M_L = I.$$

Using the fact that M_L consist of all shifts of the matrix W representing the DWT, the explicit form of the matrix $M_L^\dagger$ reduces to ([16])

$$M_L^\dagger = (\frac{1}{2} R_1^T, \dots, \frac{1}{2^L} R_L^T, \frac{1}{2^L} S_L^T).$$

That means that the inverse TIDWT averages over all shifts of the orthogonal DWT.

The matrices R_j and S_L have a close relation to the so-called *à-trous filters*. These filters were developed in [17] to compute wavelet coefficients on a denser grid than the DWT does. The class of à-trous filters $\mathbf{a} = \{a_k : k \in Z\}$ is characterized by the property $a_{2k} = \delta_{0k}$. In [26], Shensa showed the relation between the à-trous and the Daubechies scaling and wavelet filters. In particular, the à-trous filters can be defined as the autocorrelation filters of the Daubechies filters, i.e.

$$a_k = \sum_l h_l h_{l+k} \quad \text{and} \quad b_k = \sum_l g_l g_{l+k}. \tag{2.8}$$

For the Haar basis ($D = 1$) these filters are given by

$$\mathbf{a} = \begin{pmatrix} 0.5 & 1 & 0.5 \end{pmatrix} \quad \text{and} \quad \mathbf{b} = \begin{pmatrix} -0.5 & 1 & -0.5 \end{pmatrix}. \tag{2.9}$$

For the Daubechies-4 wavelet system ($D = 2$), we obtain

$$\mathbf{a} = \begin{pmatrix} -1/16 & 0 & 9/16 & 1 & 9/16 & 0 & -1/16 \end{pmatrix} \quad \text{and} \tag{2.10}$$

$$\mathbf{b} = \begin{pmatrix} 1/16 & 0 & -9/16 & 1 & -9/16 & 0 & 1/16 \end{pmatrix}. \tag{2.11}$$

Some properties of these filters and relations to autocorrelation functions of the Daubechies scaling and wavelet functions will be studied in the following subsection.

We will express the computation done by *undecimated à-trous algorithm* at scale j, i.e. running a filter bank with the filter coefficients a_k up to level j, but not down-sampling, in a matrix representation. We denote the matrix for the scaling coefficients by T_j, and the corresponding wavelet part by Q_j. Then these matrices satisfy the following equalities

$$T_j = 2^{-j} S_j^T S_j \quad \text{and} \quad Q_j = 2^{-j} R_j^T R_j.$$

These equations imply that after setting all wavelet coefficients of the TIDWT to zero, applying the inverse transform $M_L^\dagger$ yields the same result as simply multiplying the input data $\mathbf{x}$ with the matrix T_L, i.e.

$$T_L\mathbf{x} = M_L^\dagger \circ (\mathbf{0}, \dots, \mathbf{0}, S_L^T)^T \mathbf{x}.$$

Therefore, we refer to $T_L\mathbf{x}$ as the *translation-invariant projection of* $\mathbf{x}$. These projections play an important role in the smoothness characterization for translation-invariant denoised functions. For further details on this topic we refer the interested reader to [1].

The examples of the à-trous filters corresponding to the Haar and Daubechies-4 wavelet system in Eq.(2.9) and (2.10) seem to show some relation between the scaling and the wavelet filters. In fact, the á-trous filters have many important properties for signal analysis and were studied under various purposes in the literature ([2], [4], [11], [21]). Since the filters $\mathbf{a}$ and $\mathbf{b}$ are autocorrelation filters of the Daubechies filters $\mathbf{h}$ and $\mathbf{g}$ we can use the equation $g_k = (-1)^k h_{1-k}$ and obtain the following relation between the filter coefficients a_k and b_k ([2])

$$b_k = (-1)^k a_k.$$

The Daubechies filters occur in the scaling equations for the scaling and wavelet function. Analogously, the coefficients a_k and b_k occur as coefficients in scaling equations for the autocorrelation functions of ϕ and ψ. In particular, we denote the autocorrelation functions by

$$\begin{aligned} \Phi(s) &= \int_{-\infty}^{+\infty} \phi(x)\phi(x+s)dx \quad \text{and} \\ \Psi(s) &= \int_{-\infty}^{+\infty} \psi(x)\psi(x+s)dx. \end{aligned} \tag{2.12}$$

and obtain the following scaling equations ([2])

$$\begin{aligned} \Phi(x) &= \sum_{k\in\mathbb{Z}} a_k \Phi(2x-k) \\ \Psi(x) &= \sum_{k\in\mathbb{Z}} (-1)^k a_k \Phi(2x-k). \end{aligned}$$

One important property of these autocorrelation functions is that they are interpolating functions. This means that starting with an impulse signal and iterating a filter bank with the á-trous filters yields exact values of Φ,Ψ, resp. at dyadic rationals ([10]). Applying this iterative scheme to orthogonal Daubechies scaling functions we obtain only approximations of function values of ϕ and ψ. The autocorrelation functions were used, e.g., in [25] for reconstruction of a given signal from zeros of a multiresolution analysis formed by autocorrelation functions. The filter coefficients a_k were used in [2] to represent differential operators. For a summary of properties of these functions we refer to [1].

III. A correlation-dependent model for denoising via nonorthogonal wavelet transforms

Wavelet coefficients of a function are, in general, large in regions where the function is irregular and small in smooth regions. If a function is corrupted by additive noise this noise dominates the wavelet coefficients at small scales. That means most of those coefficients contain the noisy part of the signal and only a few large coefficients are related to strong singularities in the underlying function. Thresholding the noisy wavelet coefficients using the right thresholding scheme removes most of the noise and preserves the large coefficients. A smooth approximation of the function f is obtained via the inverse wavelet transform.

The most common model for noise is additive Gaussian white noise. This leads to the following stochastic model. We observe data points $\mathbf{d} = (d_1, \ldots, d_N)$ given by

$$d_i = f_i + \sigma z_i \tag{3.1}$$

where $f_i = f(i/N)$ are sample values of a continuous function supported on the unit interval, $\sigma > 0$, and $z_1, \ldots, z_N$ are i.i.d. N(0,1) random variables. That means we observe noisy samples of an unknown function with noise level σ. The goal is to recover the samples f_i from the given data.

A theoretical solution to this denoising problem via thresholding of wavelet coefficients has been developed by Donoho and Johnstone in several papers ([11], [12], [13]) and has lead to state-of-the-art denoising techniques. Inspired by these results further issues such as modeling of non-Gaussian noise or applying wavelet-based Hidden Markov Models to noisy data has been studied in [15], [22], [8]. All these denoising schemes use wavelet coefficients of the orthogonal DWT. In this section, we derive a general thresholding scheme for the coefficients of the biorthogonal and translation-invariant DWT and obtain a generalizations of the Donoho-Johnstone denoising model.

A. The Donoho-Johnstone model for the orthogonal DWT

The main idea developed in [12], [13] is to consider the noise in the wavelet domain rather than in the original data set $\mathbf{d}$ in order to get estimations on the L_2-error and on the smoothness of the denoised function. The latter estimation is possible because of the property that orthogonal wavelet systems form unconditional bases for a wide range of smoothness spaces. This means that a sequence norm on wavelet coefficients is equivalent to a norm on the function itself. Therefore, a shrinkage of wavelet coefficients yields a decreasing norm and a related increasing smoothness of the corresponding function. For further details on the norm equivalence and its effect on denoising we refer to [11], [12], [14]. In the following we will focus on how to estimate an upper bound on the noise in the wavelet domain.

In the wavelet transform domain the data $\mathbf{d} = (d_1, \ldots, d_N)$ transform to

$$W\mathbf{d} = W\mathbf{f} + \sigma\, W\mathbf{z}. \tag{3.2}$$

where $\mathbf{f}$ denotes the vector of samples $f_1, \ldots, f_N$ and $\mathbf{z}$ the vector of random variables $z_1, \ldots, z_N$. Since W is an orthogonal matrix the components of the vector $\tilde{\mathbf{z}} = W\mathbf{z}$ are identically distributed Gaussian random variables with zero mean and variance one. Furthermore, the N variables $\tilde{z}_i$ are uncorrelated. Since for normal variables, uncorrelatedness is equivalent to independence (e.g. [23]) the $\tilde{z}_i$ form a sequence of i.i.d N(0,1) random variables.

Our goal is to remove all the noise in the wavelet domain, i.e. we would like to find an operator that sets the noise term $W\mathbf{z}$ to zero. To find this operator Donoho and Johnstone used the following result from [20] which describes how the maximal magnitude of i.i.d N(0,1) random variables behaves as N goes to infinity. Namely,

$$P\{\max_{i=1,\ldots,N} |z_i| \leq \sqrt{2\log N}\} \to 1 \quad \text{for} \quad N \to \infty. \tag{3.3}$$

Therefore, applying the soft-threshold nonlinearity

$$\eta_t(y) = \mathrm{sgn}(y)(|y| - t)_+ \tag{3.4}$$

with $t = t_N = \sigma\sqrt{2\log N}$ to the wavelet coefficients $W\mathbf{d}$ removes the noise in the wavelet domain almost surely as N goes to infinity. As the final step we apply the inverse transform W^{-1} to the denoised wavelet coefficients and obtain a vector $\mathbf{f}^*$ defined by

$$\mathbf{f}^* := W^{-1}\eta_{t_N} W\mathbf{d}. \tag{3.5}$$

The elements of this denoised vector $\mathbf{f}^*$ can be interpreted as samples of a continuous function f^* which is at least as smooth as the original function f (see [12]).

The threshold $t_N := \sigma\sqrt{2\log N}$ is a general threshold which depends only on the number of random variables N and the strength of the noise, but neither on the underlying function f nor the chosen level of decomposition or the wavelet filter. Therefore, we call the threshold $t_N = \sigma\sqrt{2\log N}$ a *uniform threshold.*

B. A general thresholding scheme for nonorthogonal wavelet transforms

In order to study denoising properties of biorthogonal and undecimated wavelet transforms we will derive a general thresholding scheme for nonorthogonal wavelet transforms. In general, a nonorthogonal DWT is given by an $P_N \times N$ matrix W with $P_N \geq N$. We assume that there exists an inverse transform given by an $N \times P_N$ matrix $\widetilde{W}$ such that

$$\widetilde{W}W = I \tag{3.6}$$

The existence of this inverse matrix is guaranteed for the biorthogonal and the translation invariant DWT (in general, the Moore-Penrose inverse provides such a pseudo-inverse). The wavelet transform of the data $\mathbf{d}$ is given by

$$W\mathbf{d} = W\mathbf{f} + \sigma W\mathbf{z} = \theta + \sigma\xi \tag{3.7}$$

where $\theta := W\mathbf{f}$ and $\xi := W\mathbf{z}$. The vector ξ consist of P_N random variables $\xi_1, \ldots \xi_{P_N}$. These random variables have a normal distribution with expected value $E(\xi_i) = 0$ and a covariance matrix

$$\Sigma := WW^T. \tag{3.8}$$

The random variables ξ_i are not necessarily uncorrelated, i.e. $\Sigma \neq I$ in general, and the variances of ξ_i are in general not the same. These variances can depend on the level of resolution in the wavelet decomposition. If Σ is a diagonal matrix with elements σ_i, that are not all equal, then the random variables ξ_i are independent with variance σ_i and the Donoho thresholding model can be modified by rescaling the wavelet coefficients at each level of resolution. After this normalization we can use result from (3.3) to remove the noise part in the wavelet coefficients. After thresholding the coefficients have to be scaled to the original variance σ_i. This can be simply done using the level-dependent threshold $\sigma_i\sqrt{2\log N}$.

In the case that the random variables ξ_i are not independent the result from (3.3) is not valid anymore. We need a result similar to (3.3), but for correlated random variables. For simplicity, we assume that the ξ_i are identically distributed $\mathrm{N}(0,1)$. In [20], the maximum of the magnitude of those correlated random variables ξ_i is characterized in the following way by comparing $\max|\xi_i|$ with the maximum of the magnitude of the same number of i.i.d. N(0,1) random variables z_i (see also [1])

$$|P\{\max_{i=1,\ldots P_N}|\xi_i|\leq t\}-P\{\max_{i=1,\ldots P_N}|z_i|\leq t\}|\leq K\sum_{1\leq i<j\leq P_N}|r_{ij}|\exp\left(-\frac{t^2}{1+|r_{ij}|}\right), \tag{3.9}$$

where K is a constant, r_{ij} denotes the correlation $Cor(\xi_i,\xi_j)$ and we have to assume that

$$\delta:=\max_{i\neq j}|r_{ij}|<1. \tag{3.10}$$

The above inequality (3.9) describes the difference between the correlated and the uncorrelated case of N(0,1) random variables are expressed by the *error term*

$$R(t)=\sum_{1\leq i<j\leq P_N}|r_{ij}|\exp\left(-\frac{t^2}{1+|r_{ij}|}\right). \tag{3.11}$$

Our goal is now to find for a given wavelet system a threshold $t^*_{P_N}$ such that the following two conditions are satisfied

$$t^*_{P_N}\geq t_{P_N} = \sqrt{2\log P_N} \quad \text{and} \tag{3.12}$$

$$\lim_{N\to\infty}R(t^*_{P_N}) = 0 \quad . \tag{3.13}$$

Given a threshold $t^*_{P_N}$ that satisfies the first condition we know from (3.3) that

$$P\{\max_{i=1,\ldots,P_N}|z_i|\leq t^*_{P_N}\}\to 1 \quad \text{for} \quad N\to\infty. \tag{3.14}$$

If the second condition (3.13) is also satisfied we obtain the desired asymptotic behavior for the correlated random variables ξ_i

$$P\{\max_{i=1,\ldots,P_N}|\xi_i|\leq t^*_{P_N}\}\to 1 \quad \text{for} \quad N\to\infty. \tag{3.15}$$

The error term R depends on the correlations r_{ij} and the number of variables P_N. This indicates that we should use different thresholds $t^*_{P_N}$ for different wavelet systems and transforms.

Our approach is different from the level-dependent threshold derived in [18]. Johnstone and Silverman studied minimax properties of estimators for data with colored noise, but used an orthogonal DWT. In this context they could neglect the correlations in the original data. In this paper, we deal with correlations introduced through the transform and address the question of how to choose the threshold in order to remove the noise instead of measuring risk estimates.

IV. Special thresholds of for selected nonorthogonal wavelet transforms

In this section we explicitly derive thresholds t^* depending on a given wavelet system such that the conditions (3.12) and (3.13) are satisfied. We will see that, due to the fact that biorthogonal and translation-invariant wavelet transform form are not orthogonal transforms, we have to choose a larger threshold than for orthogonal wavelet transforms.

A. Biorthogonal wavelet systems

If we apply a biorthogonal DWT to our data then the matrix W is of size $N \times N$, i.e. $P_N = N$ in the previous section. The biorthogonality implies that there exists an $N \times N$ matrix $\widetilde{W}$ with $\widetilde{W}W = I$, but $\widetilde{W} \neq W^T$. Therefore, the correlation matrix Σ from (3.8) does not reduce to the identity matrix. As a consequence, the error term $R(t)$ from Eq.(3.11) does not vanish. In order to apply the result from (3.9) we assume that the random variables ξ_i all have the same variance 1. We now analyze the size of the error term $R(t^*)$ when we set t^* equal to the uniform threshold $t_N = \sqrt{2\log N}$, i.e. $t^* = t_N^* = \sqrt{2\log N}$. We obtain the following inequality

$$\begin{aligned} \sum_{1\le i<j\le N} |r_{ij}| \exp\left(-\frac{(t_N^*)^2}{1+|r_{ij}|}\right) &\le \sum_{1\le i<j\le N} |r_{ij}| \exp(-\log N) \\ &\le \frac{1}{N} \sum_{1\le i<j\le N} |r_{ij}|. \end{aligned}$$

To find a general upper bound for $\sum_{1\le i<j\le N} |r_{ij}|$ we count those correlations between wavelet and scaling coefficients that are nonzero for every wavelet system due to the finite length of the filters.

$$\#\{(k,r)|Cov(c_k^{(L)}, c_{k+r}^{(L)} \neq 0\} \le 2^{-L}N(D-1) \tag{4.1}$$

$$\#\{(k,r)|Cov(c_k^{(L)}, d_{k+r}^{(l)} \neq 0,\, r \neq 0\} \le 2^{-L}ND \cdot 2^{-l}ND \tag{4.2}$$

$$\#\{(k,r)|Cov(d_k^{(l)}, d_{k+r}^{(l)} \neq 0\} \le 2^{-l}N(D-1) \tag{4.3}$$

$$\#\{(k,r)|Cov(d_k^{(l)}, d_{k+r}^{(m)} \neq 0,\, m<l, r \neq 0\} \le 2^{-l}ND \cdot 2^{-m} \tag{4.4}$$

where D denotes half of the support-width of ϕ. Combining these results we obtain

$$\begin{aligned} \sum_{1\le i<j\le N} |r_{ij}| &\le \delta[2^{-L}N(D-1) + 2^{-L}ND\sum_{l=1}^{L} \cdot 2^{-l}ND + \sum_{l=1}^{L} 2^{-l}N(D-1) + \sum_{l=1}^{L}\sum_{m=1}^{l-1} 2^{-l}ND \cdot 2^{m}] \\ &= \delta N[(D-1)(2^{-L} + (1-2^{-L}) + D2^{-L}(1-2^{-L}) + D\sum_{l=1}^{L} 2^{-l}(2^{l}-2)] \\ &= \delta N[(D-1)(2^{-L}(1-2^{-L})) + D2^{-L}(1-2^{-L}) + D\sum_{l=1}^{L}(1-2^{-l+1})] \\ &= \delta N[(D-1)(2^{-L} + (1-2^{-L})) + D2^{-L}(1-2^{-L}) + D(L-2(1-2^{-L}))] \end{aligned}$$

Using this estimate it follows that $R(t_N^*) \not\to 0$ for $N \to \infty$. However, if we incorporate the correlation bound δ into the threshold t^*, namely $t_N^* = \sqrt{2(1+\delta)\log N}$ we obtain the desired asymptotic behavior of R

$$\begin{aligned}
R(t_N^*) &\leq K \sum_{1\leq i<j\leq N} |r_{ij}| \exp(-2\log N) \\
&\leq K\frac{1}{N^2} \sum_{1\leq i<j\leq N} |r_{ij}| \\
&\leq K\delta\frac{1}{N}[2^{-L}(D-1)+(D-1)(1-2^{-L})+D2^{-L}(1-2^{-L})+D(L-2(1-2^{-L}))] \\
&\to 0 \quad \text{for} \quad N\to\infty.
\end{aligned}$$

The threshold t_N^* depends on N and the correlation bound δ. Furthermore, the bound δ depends on the modified covariance matrix $\Sigma - I$ and, therefore, on the chosen wavelet system. Moreover, the correlation structure changes in general from one level to the next level. Thus, the bound δ depends on L and on ϕ and we write $\delta = \delta_{L,\phi}$. Using this notation the threshold

$$t_{N,\phi,L}^* = \sqrt{2(1+\delta_{L,\phi})\log N} \tag{4.5}$$

is sufficient to remove all noise in the normalized wavelet coefficients. At the beginning of this section we made the assumption that $Var(\xi_i) = 1$ for all $i = 1, \ldots, N$. This is in general not satisfied for biorthogonal wavelet transforms since the variances can differ from level to level. All random variables that correspond to wavelet coefficients of the first scale have variance σ_1^2 and are denoted by $\xi_{1,1}, \ldots, \xi_{N_1,1}$, those corresponding to wavelet coefficients at the second scale have variance σ_2^2 and are denoted by $\xi_{1,2}, \ldots, \xi_{N_2,2}$, and so on. The random variables that correspond to the scaling coefficients at the coarsest scale have variance σ_{L+1}^2. Incorporating this information into the threshold in (4.5) we obtain the following *scale-wavelet dependent* threshold for biorthogonal wavelet systems

$$t_{N,\phi,L}^*(l) = \sigma_l\sqrt{2(1+\delta_{L,\phi})\log N} \quad \text{for wavelet coefficients at scale } l, l=1,\ldots,L, \text{ and} \tag{4.6}$$

$$t_{N,\phi,L}^*(L+1) = \sigma_{L+1}\sqrt{2(1+\delta_{L,\phi})\log N} \quad \text{for scaling coefficients at scale } L. \tag{4.7}$$

Here $\delta_{L,\phi}$ denotes the bound of the cross-correlations between the normalized random variables.

B. Translation-invariant wavelet transforms

The translation-invariant DWT consists of all shifts of an orthogonal DWT. For that reason the transform matrix $W = M_L$ is of size $(L+1)N \times N$, i.e. $P_N = (L+1)N$, where L is the maximal level of decomposition. The matrix $\widetilde{W} = M_L^\dagger$ is the pseudo-inverse $(M_L^T M_L)^{-1} M_L^T$ o fM_L. As a consequence, the correlation matrix $\Sigma = WW^T$ does not reduce to the identity matrix. However, many entries of the correlation matrix are zero due to the relations of the TIDWT to the corresponding orthogonal DWT. Following the strategy from the previous sections the following two conditions have to be satisfied for a threshold t^* that is sufficient to remove the noise in the redundant wavelet coefficients

$$t^*_{(L+1)N} \geq t_{(L+1)N} = \sqrt{2\log((L+1)N)} \quad \text{and} \tag{4.8}$$

$$\lim_{N\to\infty} R(t^*_{(L+1)N}) = 0. \tag{4.9}$$

Similar to the case of biorthogonal wavelets we have to count the correlations that are not necessarily zero. If we incorporate the knowledge of zero-entries in the correlation matrix into our computations we obtain the following bound for non-zero correlations

$$\#\{(i,j)\,|\,r_{ij} \neq 0, i \neq j\} \leq DN2^{L+1}(1+2^L) \tag{4.10}$$

where D denotes half of the support-width, i.e. D is the order of the Daubechies wavelet system ([9]). For the exact computations we refer to [1]. Again we see that this bound is not enough to derive the desired result $R(t^*_{(L+1)N}) \to 0$. However, incorporating the correlation bound $\delta = \delta_{L,\phi}$ into the threshold as

$$t^*_{(L+1)N,\phi} = \sqrt{2(1+\delta_{L,\phi})\log((L+1)N)} \tag{4.11}$$

the two conditions (4.8) and (4.9) are satisfied. We want to mention that we do not have to incorporate a variance factor σ_l into the threshold as we had to do in the biorthogonal case. Since the TIDWT consists of shifts of an orthogonal DWT the variances at each level are equal to 1. The increasing amount of redundancy introduced from level to level requires the conditional factor $(L+1)$ in the log-term in the threshold. As for biorthogonal wavelets the derived threshold is a *scale-wavelet dependent* threshold.

We want to mention that the orthogonal uniform-threshold model is embedded in the new scale-wavelet dependent model. For the orthogonal situation the correlation bound $\delta_{L,\phi}$ is zero and we have to consider only N noisy wavelet coefficients, i.e. $P_N = N$. Then the thresholds

$$\sqrt{2(1+\delta_{L,\phi})\log((L+1)N)} \quad \text{(translation-invariant DWT)} \qquad \text{and}$$

$$\sqrt{2(1+\delta_{L,\phi})\log N} \quad \text{(normalized biorthogonal DWT)}$$

reduce to the uniform-threshold

$$\sqrt{2\log N} \quad \text{(orthogonal DWT).}$$

For this reason we interpret the new thresholding model as a generalization of the classical Donoho-Johnstone model.

V. The correlation structure

In the previous section we showed that the key to a scale-wavelet dependent threshold for nonorthogonal wavelet transforms is the correlation matrix. The lack of orthogonality in the biorthogonal and translation-invariant wavelet DWT produces non zero cross-correlations and requires a larger threshold than that for the orthogonal DWT. In this section we will study the correlation matrices for translation-invariant and biorthogonal DWT.

A. Translation-invariant DWT

We have shown that the correlations between noisy wavelet coefficients contribute to the threshold by entering in the factor $(1+\delta_{L,\phi})$. In the following we study this factor for the TIDWT in more detail.

First we look at the correlations between scaling coefficients $c_k^{(j)}$ at a fixed scale j, i.e. between the elements of the vector $S_j\,\mathbf{z}$ in (2.7). For $j=1$ we have to compute

$$Cov(c_k^{(1)}, c_{k+r}^{(1)}) = \sum_{l,m} h_l h_m Cov(z_{k+l}, z_{k+r+m})$$

From the independence of the input random variables z_i we know that only the terms where $l = r+m$ contributes to the sum. Therefore, we obtain

$$Cov(c_k^{(1)}, c_{k+r}^{(1)}) = \sum_m h_{m+r} h_m .$$

Comparing this expression with (2.8) we notice that the covariance $Cov(c_k^{(1)}, c_{k+r}^{(1)})$ is exactly the à-trous filter coefficient a_r. Because the corresponding scaling function Φ from Eq.(2.12) is an interpolating function, a_r is the function value of $\Phi(x)$ at the point $x = r/2$. In [1] we generalized this approach and showed that the covariances between scaling coefficients at scale j satisfy the following equality

$$Cov(c_k^{(j)}, c_{k+r}^{(j)}) = \Phi(2^{-j} r). \tag{5.1}$$

In an analogue way we obtain the equation

$$Cov(d_k^{(j)}, d_{k+r}^{(j)}) = \Psi(2^{-j} r). \tag{5.2}$$

for the correlations between wavelet coefficients $d_k^{(j)}$ at a fixed scale j. It remains to study the structure of cross-correlations $Cov(c_k^{(l)}, d_{k+r}^{(j)})$ and $Cov(d_k^{(l)}, d_{k+r}^{(j)})$. We showed in [1] that these correlations occur as values of the corresponding cross-correlation functions at dyadic rationals. In particular

$$\begin{aligned} Cov(c_k^{(l)}, d_{k+r}^{(j)}) &= \langle \phi_{j-l}, \psi(.-2^{-j}r)\rangle \quad \text{and} \\ Cov(d_k^{(l)}, d_{k+r}^{(j)}) &= \langle \psi_{j-l}, \psi(.-2^{-j}r)\rangle . \end{aligned} \tag{5.3}$$

In order to estimate the threshold $t^*_{L,N,\phi}$ we have to compute the maximum of all these cross-correlation terms for all $r \in \mathbb{Z}\backslash\{\nvdash\}$. First we start to analyze the maximum of the correlations between the scaling coefficients. From the structure of the autocorrelation function Φ it becomes clear that for large j the maximum is $\Phi(2^{-j})$, and it tends to one for $j \to \infty$. Using the relation ([4])

$$2\Phi(2x) = \Phi(x) + \Psi(x)$$

for large j it follows that the value $|\max_{r\in\mathbb{Z}\setminus\{\nvdash\}}\Psi(2^{-j}r)|$ is smaller than $\Phi(2^{-j})$. Therefore, the value $\Phi(2^{-j})$ should dominate the correlation terms of scaling and wavelet coefficients at a scale j. But we have not been able to find a similar result for the cross-correlation terms in (5.3). Many experiments have indicated that the term $\Phi(2^{-j})$ dominates also the cross-correlation terms, but we could not found a proof for these experimental results. However, we showed in [1] that there exist an upper bound of these cross-correlations which is less than 1. Therefore, the assumption in (3.10) is satisfied. At this point it becomes clear that the correlation bound $\delta_{L,\phi}$ does not depend on N, but only on the maximal level of resolution and the chosen wavelet system. Figure 1 shows the correlation bound $\delta_{L,\phi}$ for various levels L and the Daubechies scaling functions ϕ of order $1, 2$ and 3.

B. Biorthogonal DWT

For biorthogonal wavelet systems the characterization of the estimate of the correlations via continuous correlation functions as given above for the TIDWT in the previous section is not valid. This is due to the fact that the correlation functions of biorthogonal wavelet and scaling functions are not interpolating functions because the analysis filters of biorthogonal wavelet systems do not satisfy the orthogonality condition (2.5).

The correlation matrices can have very different structures. In Figure 2 we show the values for σ_l and the associated correlation bounds $\Delta_\phi(l)$ for normalized cross-correlations $Cov(\xi_{i,m}/\sigma_m, \xi_{j,n}/\sigma_n), m, n \leq N$ and scales i, j between wavelet coefficients for two biorthogonal wavelet systems. One is the spline based wavelet system with decomposition scaling filter $(-1/8, 1/4, 3/4, 1/4, -1/8)$ and reconstruction filter $(1/4, 1/2, 1/4)$, and the other one is the Daubechies7-9 wavelet system (see [6]). The change in the variances from level to level in the spline example is clearly visible. In order to keep the correlation bound δ in the threshold $t^*_{N,\phi,L}$ (4.6) as small as possible for the purpose of denoising we should use a biorthogonal system that has a correlation matrix close to the identity matrix. This strategy is also followed in [29], but using a different criterion based on averaging over the different variances and neglecting the cross-correlations.

VI. Uniform smoothness via nonorthogonal denoising applying the scale-wavelet dependent threshold

In this section we compare the performance in denoising using the uniform and the new scale-wavelet dependent threshold. The experimental results were computed using the `Matlab` toolboxes `Wavelab` (Stanford) and the `Rice Wavelet Toolbox`.

A. Translation-invariant DWT

Lang et al. mentioned in [19] that best results of translation-invariant denoising with a uniform threshold were achieved when using a slightly larger threshold than the one which worked best for denoising via an orthogonal DWT. The results of the previous sections give insight into this phenomenon. The correlations between the noisy wavelet coefficients require a larger threshold than the one which is sufficient for uncorrelated noise. This new threshold $t^*_{L,N,\phi}$ increases slowly with increasing L. For example, the threshold $t^*_{1,N,\phi}$ is sufficient to remove all the noise in the first scale of wavelet and scaling coefficients. For a larger level of resolution $l > 1$ we have to choose the corresponding threshold $t^*_{l,N,\phi}$. It is well-known that in order to preserve the mean we usually do not threshold the low-pass component, i.e. the scaling coefficients. If we denote the maximum of the magnitude of the cross-correlations between all wavelet coefficients $R_j\mathbf{z}$ up to level l $(j \leq l)$ by $\Delta_\phi(l)$ then the threshold

$$t^*_{N,\phi}[l] = \sqrt{2(1+\Delta_\phi(l))\log(lN)}$$

is sufficient to remove the noise in the high-pass components. The values of $\Delta_\phi(l)$ corresponding to different level and wavelet systems are given in Figure 3.

The threshold in [19] was computed from experiments done for a special wavelet and a special maximal level of resolution for specific data sets. In this paper we have derived a general framework for computing thresholding schemes for coefficients of the TIDWT associated to a Daubechies wavelet system.

In classical wavelet denoising there exist two thresholding techniques, *hard-* and *soft-thresholding* ([13]). The hard-thresholding operator is a "keep-or-kill" operator

$$\eta^H_T(y) = \begin{cases} y & , \quad y \geq t \\ 0 & , \quad y < t \end{cases} , \tag{6.1}$$

i.e. it keeps all coefficients larger than a given threshold t and sets all coefficients smaller than t to zero. The alternative soft-thresholding operator

$$\eta^S_t(y) = \mathrm{sgn}(y)(|y| - t)_+ \tag{6.2}$$

shrinks all coefficients larger than the threshold by the same amount (the threshold), while the smaller ones are mapped to zero.

For classical orthogonal wavelet denoising estimates of the original sample vector $\mathbf{f}$ are obtained by hard- and soft-thresholding from

$$\mathbf{f}^* = W^{-1}\eta^H_{t_N} W\mathbf{d} \quad \text{or} \quad \mathbf{f}^* = W^{-1}\eta^S_{t_N} W\mathbf{d} \tag{6.3}$$

where $t_N = \sqrt{2\log N}$ is the uniform threshold. When using soft-thresholding, smoothness properties of the estimate $\mathbf{f}^*$ are guaranteed. In particular, Donoho showed in [12] that there exists a smooth function $f^* \in C[0,1]$ which interpolates the samples $\mathbf{f}^*$ and is at least as smooth as f in a wide range of smoothness classes. This smoothness characterization can not be obtained using hard-thresholding. In applications soft-thresholding generally tends to oversmooth signal details, whereas hard-thresholding exhibits spurious oscillations. The experiments in [7] and [19] show that this disadvantage of hard-thresholding vanishes if the operator $\eta^H_{t_N}$ is applied to coefficients of the TIDWT. In general, soft-thresholding of the coefficients of the TIDWT amplifies the oversmoothing effect obtained using the DWT, and hard-thresholding yields better results in terms of the mean-square error. Therefore, hard-thresholding is usually preferred in combination with the TIDWT.

We showed in [1] that the increasing smoothness using soft-thresholding of the TIDWT compared to soft-thresholding of the DWT coefficients is due to the change of the representation of the estimated denoised function. In particular, we switch from Daubechies scaling and wavelet functions ϕ and ψ to their autocorrelation functions Φ and Ψ, respectively.

The Figures 6,7,8 and 9 demonstrate denoising results using the TIDWT and different thresholds. We analyze the four test signal from the matlab package `Wavelab`. The results using the uniform threshold t_N are obtained by computing

$$\mathbf{f}^* = M_L^\dagger(\eta^H_{t_N} R_1^T, \dots, \eta^H_{t_N} R_L^T, S_L^T)^T \mathbf{d} \tag{6.4}$$

whereas in the case of the scale-wavelet dependent threshold we compute

$$\mathbf{f}^* = M_L^\dagger(\eta^H_{t^*_{N,\phi}[1]} R_1^T, \dots, \eta^H_{t^*_{N,\phi}[L]} R_L^T, S_L^T)^T \mathbf{d}. \tag{6.5}$$

The results using the uniform thresholding model still show outliers at various sampling rates. The adaptive thresholding model yields an overall smooth estimation of the signal without any outliers. Figure 4 shows the mean-square error computed from the results of both methods using the uniform (*uni*) and the correlation dependent (*c-d*) threshold. Following a mean-square error criteria the uniform model slightly outperforms the correlation dependent model, but the mean-square error does not capture the clearly visible advantage of our new method.

Results applying the soft-thresholding operators $\eta^S_{t_N}$ and $\eta^S_{t^*_{N,\phi}[l]}$ are shown in Figure 10 for the *Doppler* test signal. This example shows the most significant difference of the four test signals between the uniform and the correlation-dependent model. The effect of oversmoothing the details in high-frequency parts becomes visible. Furthermore, the shrinkage of the wavelet coefficients lowers the amplitude of the signal. However, ranking the soft-thresholding results using uniform and scale-wavelet dependent thresholds based on the visible smoothness especially in the high-frequency part we prefer the correlation-dependent model. The uniform model still keeps some oscillations which are rather due to the noise than to the original signal. The estimates of the L^2-error are shown in Figure 4. As for hard-thresholding, the correlation-dependent model yields larger L^2-error than the uniform model.

Whatever type of thresholding, hard- or soft-thresholding, we use, we have to compute the same scale-wavelet dependent $t^*_{N,\phi}[l]$ threshold for a given wavelet system, sample size of the signal, and given level of resolution. Once computed it can be used for all kinds of test signals.

At this point we want to mention that it was observed by Coifman and Donoho in [7] that results using the *Sure Shrink* denoising method for the TIDWT were worse than the same method applied to the orthogonal DWT. There was still a high portion of noise in the denoise signal. This observation matches the content of this paper since we showed that a different thresholding scheme is necessary when using the TIDWT instead of the orthogonal DWT. Our intuition is that adapting our thresholding scheme to the Sure-Shrink method would eliminate the phenomenon observed in [7]. But this is far beyond the scope of this paper and is part of our future work.

B. Biorthogonal DWT

We study the same examples as in the previous subsection, but using biorthogonal wavelet systems. In analogy to the TIDWT case we do not threshold the scaling coefficients at the coarsest scale, and we define the level-dependent correlation $\Delta_\phi(l)$ as the upper bound on the correlations between wavelet coefficients up to level l. The Figures 6, 7,8 and 9 shows hard-thresholding denoising results obtained by applying the scale-wavelet dependent threshold $t^*_{N,\phi}[l] = \sigma_l\sqrt{2(1+\Delta_\phi(l))\log N}$ and the uniform threshold $\sqrt{2\log N}$ to the Daub7-9 and the Spline2-2 wavelet systems. The Daubechies7-9 wavelet system is close to an orthogonal one. This is also demonstrated in the parameters in Figure 2. The "almost-orthogonality" allows us to use a threshold that is closer to the uniform threshold because the correlation bounds $\Delta_\phi(l)$ are very small. The correlation structure of the Spline2-2 wavelet system is different. The variances and the maxima of the normalized cross-correlations increase and require a larger scale-wavelet dependent threshold. The difference in the denoising performance measured in terms of the mean-squared-error is demonstrated in Figure 4. As in the translation-invariant case, the slightly larger threshold produces a larger mean-squared-error, but removes the clearly visible outliers as observed in the experimental results.

VII. Conclusions

We have introduced a correlation-dependent model for thresholding noisy coefficients of nonorthogonal wavelet transforms. This model is based on the correlation structure of the transform and includes a scale-wavelet dependent threshold that reduces to the famous uniform threshold $\sqrt{2\log N}$ in the case of an orthogonal wavelet transform. The threshold is explicitly calculated for biorthogonal and translation invariant wavelet systems. We compared denoising results for hard- and soft-thresholding techniques using the new thresholding scheme.

Using hard-thresholding the new correlation model clearly outperforms the uniform model since outliers, that occur in the latter case, vanish. When applying soft-thresholding a more global smoothness of the estimate is achieved and small oscillations, that disturb the global smoothness after soft-thresholding with the uniform threshold, are removed. The better uniform smoothness in both cases does not become transparent in the estimation of the L^2-error.

Ongoing work is investigating the correlation matrices of the biorthogonal DWT in greater detail in order to find more general estimates on the correlation bound δ.

References

[1] Berkner, K., Wells Jr, R.O. Smoothness estimates for soft-threshold denoising via translation invariant wavelet transforms. Technical Report CML TR98-01, Computational Mathematics Laboratory, Rice University, 1998.

[2] Beylkin, G. On the Representation of Operators in Bases of Compactly Supported Wavelets. *SIAM J. Numer. Anal.*, 6:1716–1740, 1992.

[3] Beylkin, G., Coifman, R. R., Rohlkin, V. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.

[4] Beylkin, G., Torréssani, B. Implementation of Operators via Filter Bank, Autocorrelations Shell and Hardy Wavelets. Technical report, Centre de Physique Théorique - CNRS- Lyminy, Marseille, 1994.

[5] Burrus, C.S., Gopinath, R.A., Guo, H. *Introduction to Wavelets and Wavelet Transforms.* Prentice Hall, Englewood Cliffs, HJ, 1998.

[6] Cohen, A., Daubechies, I., Feauveau, J.-C. Biorthogonal Bases for Compactly Supported Wavelets. *Comm. Pure Appl. Math.*, 45:485–560, 1992.

[7] Coifman, R.R., Donoho, D.L. Translation invariant denoising. In A. Antoniades, editor, *Wavelets and Statistics, Springer Lecture Notes.* Springer-Verlag, 1995.

[8] Crouse, M.S., Novak, R.D., Baraniuk, R.G. Wavelet-based statistical signal processing using hidden markov models. *IEEE Trans. Signal Processing*, 46:886–902, 1998.

[9] Daubechies, I. *Ten Lectures on Wavelets.* SIAM Publ., 1992.

[10] Deslauriers, G., Dubuc, S. Symmetric iterative interpolation processes. *Constructive Approximation*, 5:49–68, 1989.

[11] Donoho, D.L. Interpolating wavelet transforms. Technical report, Department of Statistics, Stanford University, 1992.

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

[13] Donoho, D.L., Johnstone, I.M. Ideal spacial adaption by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[14] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D. Wavelet shrinkage: Asymptotia. *J.R. Statist. Soc. B*, 57, 1995.

[15] Donoho, D.L., Yu, T.P.Y. Robust Nonlinear Wavelet Transform based on Median-Interpolation. In *Proc. of the 31st Asilomar Conference on Signals, Systems, and Computers*, 1997.

[16] Guo, H. Theory and Applications of the Shift-Invariant, Time-Varying and Undecimated Wavelet Transforms. Master's thesis, George R. Brown School of Engineering, Rice University, 1995.

[17] Holschneider, M., Kronland-Martinet, R., Morlet, J., Tchamitchian, P. A real-time algorithm for signal analysis with the help of the wavelet transform. In *Wavelets: Time-Frequency Methods and Phase Space*, pages 286–297. Springer Verlag, Berlin, 1989.

[18] Johnstone, I.M., Silverman, B.W. Wavelet threshold estimators for data with correlated noise. Technical report, Department of Statistics, Stanford University, 1992.

[19] Lang, M., Guo, H., Odegard, J.E., Burrus, C.S., Wells Jr, R.O. Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Processing Letters*, 3:10–12, 1996.

[20] Leadbetter, M.R., Lindgren, G., Rootzen, H. *Extremes and Related Properties of Random Sequences and Processes.* Springer Verlag, New York, 1983.

[21] Mallat, S. Zero-crossings of a wavelet transform. *IEEE Trans. Inform. Theor.*, 37(4):1019,–1033, 1991.

[22] Novak, R.D. Optimal singal estimation using cross-validation. *IEEE Signal Processing Letters*, 4(1):23–25, 1997.

[23] Papoulis, A. *Probability, Random Variables, and Stochastic Processes.* McGraw-Hill, New York, 1984.

[24] Resnikoff, H.L., Wells Jr, R.O. *Wavelet Analysis: The Scalable Structure of Information.* Springer-Verlag, New York, 1998. To appear.

[25] Saito, N., Beylkin, G. Multiresolution representations using the auto-correlation functions of compactly supported wavelets. *IEEE Transactions on Signal Processing*, 41:3584–3590, 1993.

[26] Shensa, M.J. The discrete wavelet transform: wedding the à trous and Mallat algorithms. *IEEE Trans. Sig. Proc.*, 40(10):2464–2482, 1992.

[27] Strang, G. *Linear Algebra and Its Applications.* Academic Press, New York, 1976.

[28] Strang, G., Nguyen, T. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, 1996.

[29] Strela, V., Walden, A.T. Orthogonal and Biorthogonal Multiwavelets for Signal Denoising and Image Compression. Technical report, Department of Mathematics, Darmouth College, 1997.

[30] Tian, J., Wells Jr, R.O. Dyadic Rational Biorthogonal Coifman Wavelet Systems. Technical Report CML TR96-07, Computational Mathematics Laboratory, Rice University, 1996.

[31] Wei, D., Tian, J., Wells Jr., R.O., Burrus, C.S. A New Class of Biorthogonal Wavelet Systems for Image Transform Coding. *IEEE Trans. Image Proc.*, 7:1000–1013, 1998.

L	1	2	3	4	5	6	7	8	9	10
D=1	0.5	0.75	0.825	0.9375	0.9688	0.9844	0.9922	0.9961	0.998	0.9999
D=2	0.5625	0.5625	0.7383	0.9141	0.9731	0.9921	0.9977	0.9993	0.9998	1
D=3	0.5859	0.8708	0.9637	0.9902	0.9973	0.9992	0.9996	0.9998	1	1

Fig. 1. TIDWT: Cross-correlation bound $\delta_{L,\phi}$ for various levels and Daubechies wavelet systems of order $D = 1, 2, 3$.

Wavelet system	l	1	2	3	4	5	6	7	8
Daub7-9	σ_l	0.9830	1.1186	1.0443	1.0037	0.9901	0.9861	0.9841	0.9848
	$\Delta[l]$	0.0577	0.1382	0.1330	0.1261	0.1233	0.1225	0.1223	0.1222
Spline2-2	σ_l	0.75	1.3359	1.7593	2.0132	2.1575	2.2374	2.2811	2.3050
	$\Delta[l]$	0.1667	0.2656	0.2770	0.2903	0.2967	0.2999	0.3017	0.3025

Fig. 2. Standard deviation σ_l and normalized cross-correlation bound $\Delta_\phi(l)$ for various levels of decomposition of the Daubechies7-9 and the Spline2-2 wavelet system.

l	1	2	3	4	5	6	7	8	9	10
D=1	0.5	0.5	0.625	0.8125	0.9063	0.9531	0.9766	0.9883	0.9941	0.9971
D=2	0.5625	0.5625	0.7383	0.9141	0.9731	0.9921	0.9977	0.9993	0.9998	1
D=3	0.5859	0.5859	0.7779	0.9374	0.9833	0.9957	0.9989	0.9997	0.9999	1

Fig. 3. TIDWT: Correlation bound $\Delta_\phi(l)$ for various levels and Daubechies wavelet systems of order $D = 1, 2, 3$.

	signal	D4 uni	D4 c-d	D7-9 uni	D7-9 c-d	spline2-2 uni	spline2-2 c-d
hard-thresh.	Doppler ($\sigma = 0.1$)	0.0015	0.0022	0.0024	0.0024	0.0050	0.0065
	HeaviSine ($\sigma = 0.2$)	0.0027	0.0046	0.0090	0.0090	0.0332	0.0319
	Blocks ($\sigma = 0.1$)	0.0022	0.0024	0.0077	0.0082	0.0089	0.0094
	Bumps ($\sigma = 0.1$)	0.0028	0.0033	0.0061	0.0070	0.0070	0.0073
soft-thresh.	Doppler ($\sigma = 0.1$)	0.0064	0.0128	0.0083	0.0091	0.0149	0.0178
	HeaviSine ($\sigma = 0.2$)	0.0069	0.0425	0.0302	0.0330	0.0705	0.0827
	Blocks ($\sigma = 0.1$)	0.0175	0.0342	0.0318	0.0347	0.0322	0.0390
	Bumps ($\sigma = 0.1$)	0.0173	0.0336	0.0299	0.0327	0.0324	0.0398

Fig. 4. Mean-square error for the four test signals using different wavelet systems and thresholding techniques.

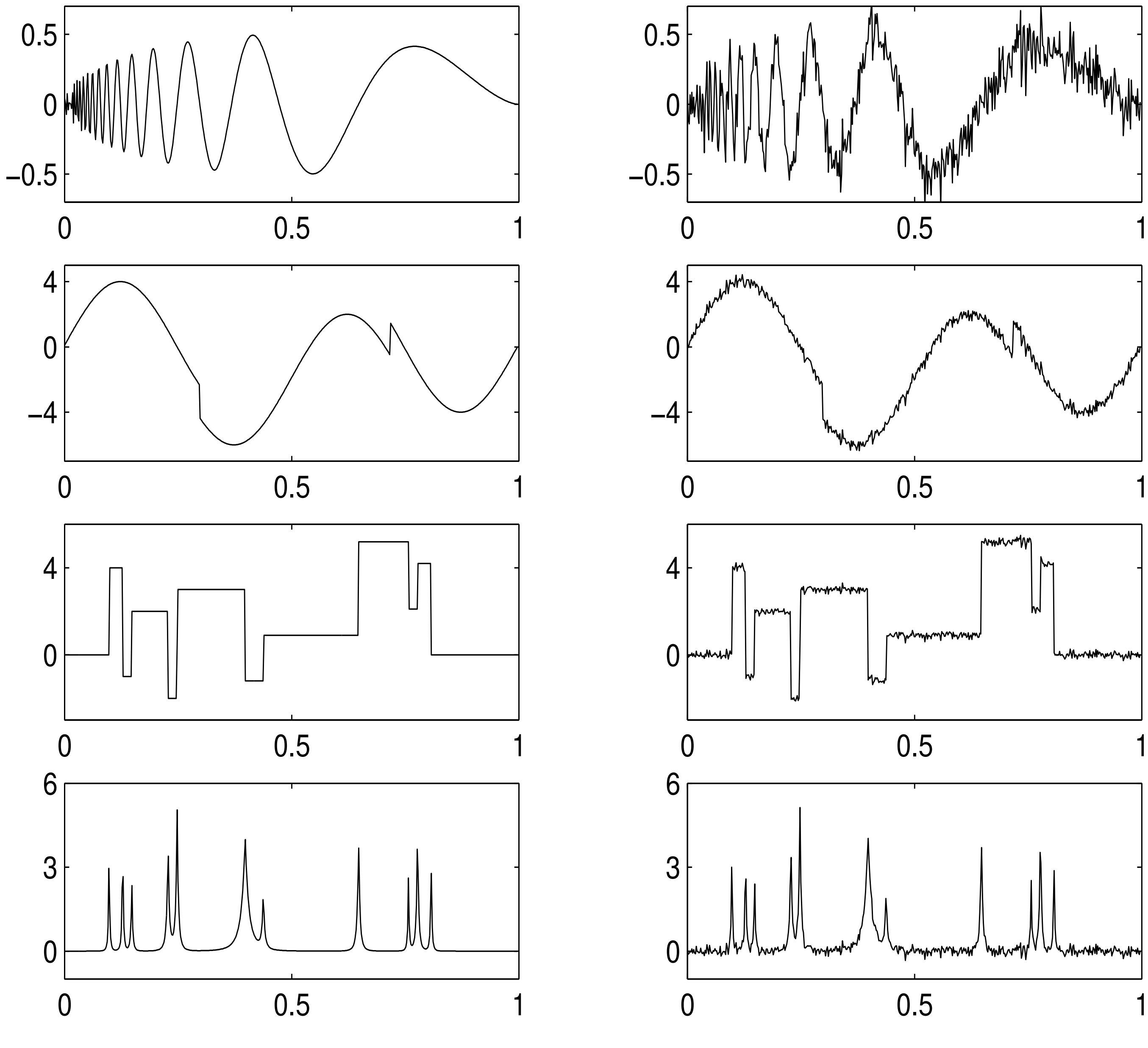

Fig. 5: The original test signals *Doppler, HeaviSine, Blocks* and *Bumps* (*left*) and noisy versions (*right*) with standard deviations $\sigma = 0.1$ (*Doppler, Blocks, Bumps*) and $\sigma = 0.2$ (*HeaviSine*).

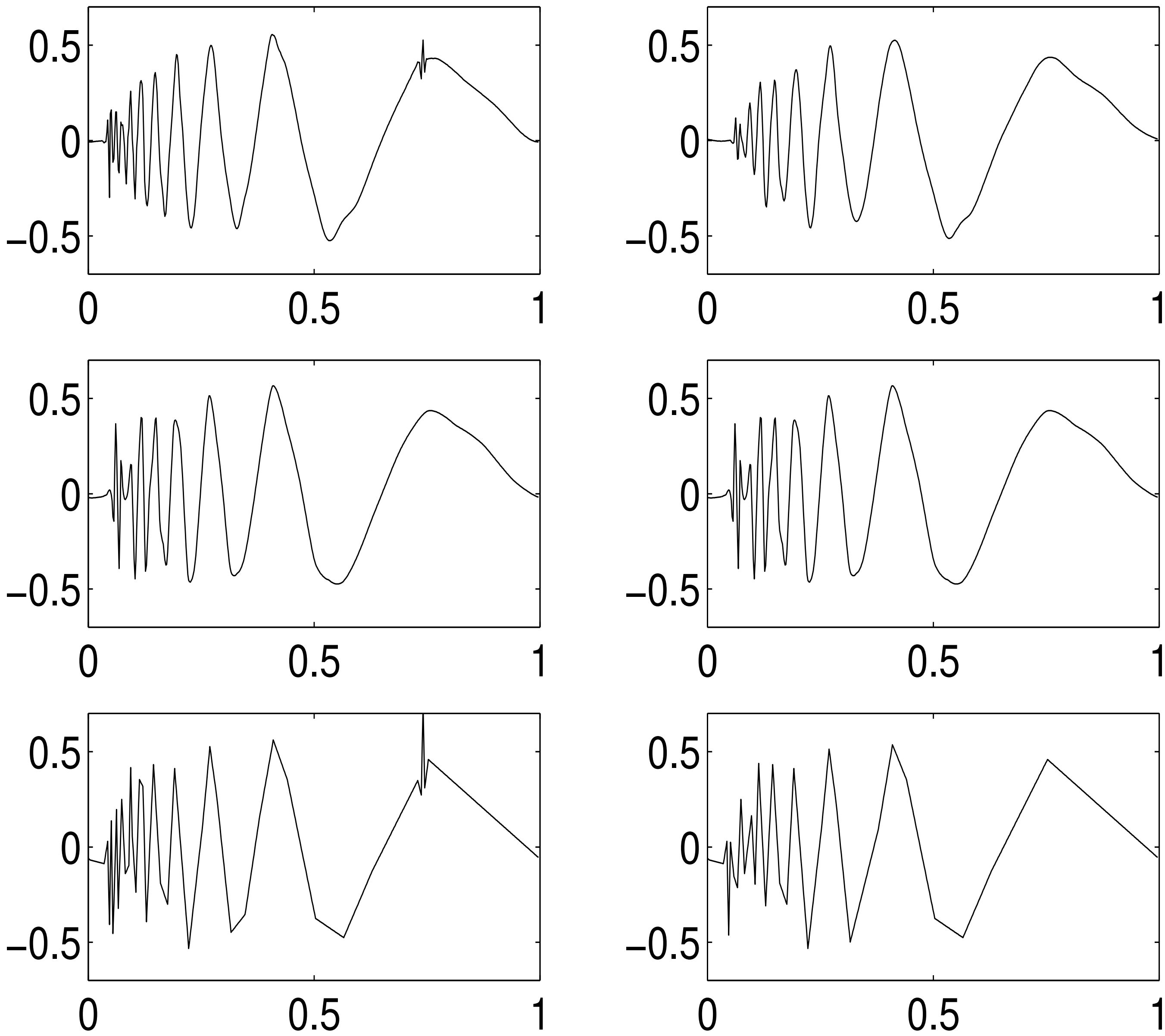

Fig. 6: Hard-threshold denoising of the *Doppler* test signal in Gaussian white noise with $\sigma = 0.1$ using different wavelet systems (redundant with D4 wavelet (*top*), biorthogonal D7-9 (*middle*), biorthogonal spline2-2 (*bottom*)). The leftmost plots use the uniform threshold $\sigma_l\sqrt{2\log N}$. The rightmost plots use the correlation-dependent threshold $\sqrt{2(1+\Delta_\phi(l))\log(lN)}$, $l = 1, \ldots, L$.

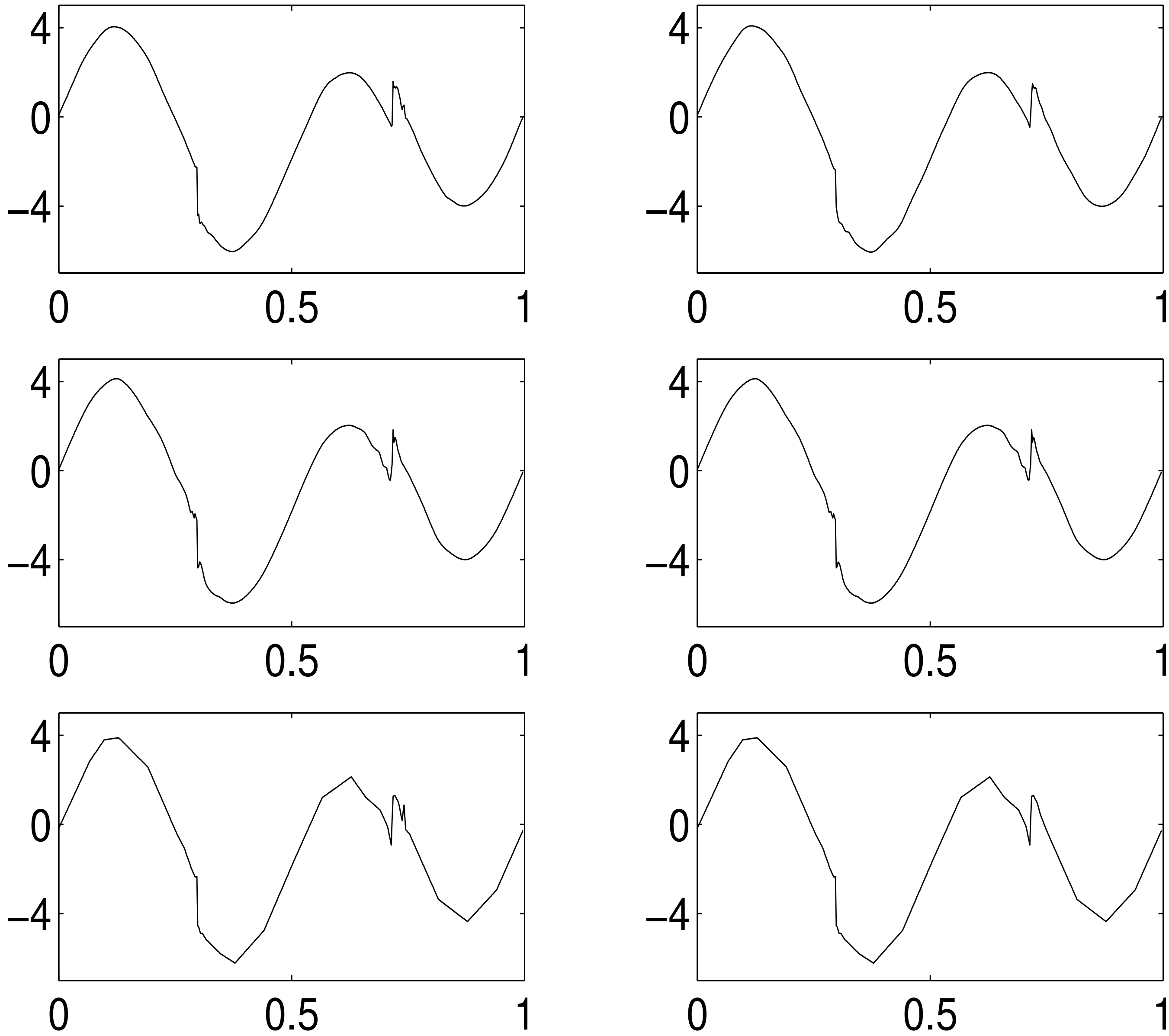

Fig. 7: Hard-threshold denoising of the *HeaviSine* test signal in Gaussian white noise with $\sigma = 0.2$ using different wavelet systems (redundant with D4 wavelet (*top*), biorthogonal D7-9 (*middle*), biorthogonal spline2-2 (*bottom*)). The leftmost plots use the uniform threshold $\sigma_l\sqrt{2\log N}$. The rightmost plots use the correlation-dependent threshold $\sqrt{2(1+\Delta_\phi(l))\log(lN)}$, $l = 1, \ldots, L$.

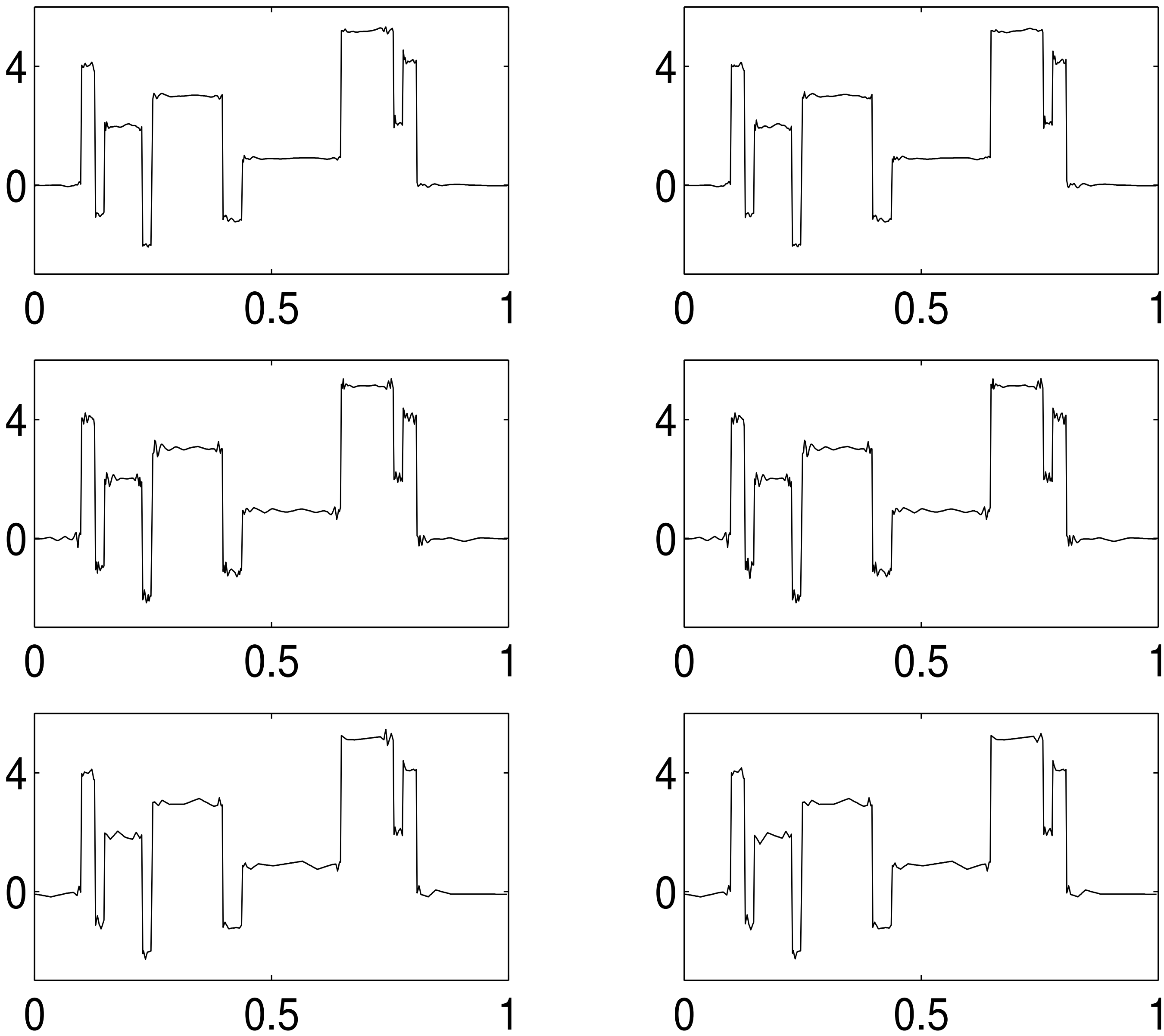

Fig. 8: Hard-threshold denoising of the *Blocks* test signal in Gaussian white noise with $\sigma = 0.1$ using different wavelet systems (redundant with D4 wavelet (*top*), biorthogonal D7-9 (*middle*), biorthogonal spline2-2 (*bottom*)). The leftmost plots use the uniform threshold $\sigma_l\sqrt{2\log N}$. The rightmost plots use the correlation-dependent threshold $\sqrt{2(1+\Delta_\phi(l))\log(lN)}$, $l = 1, \ldots, L$.

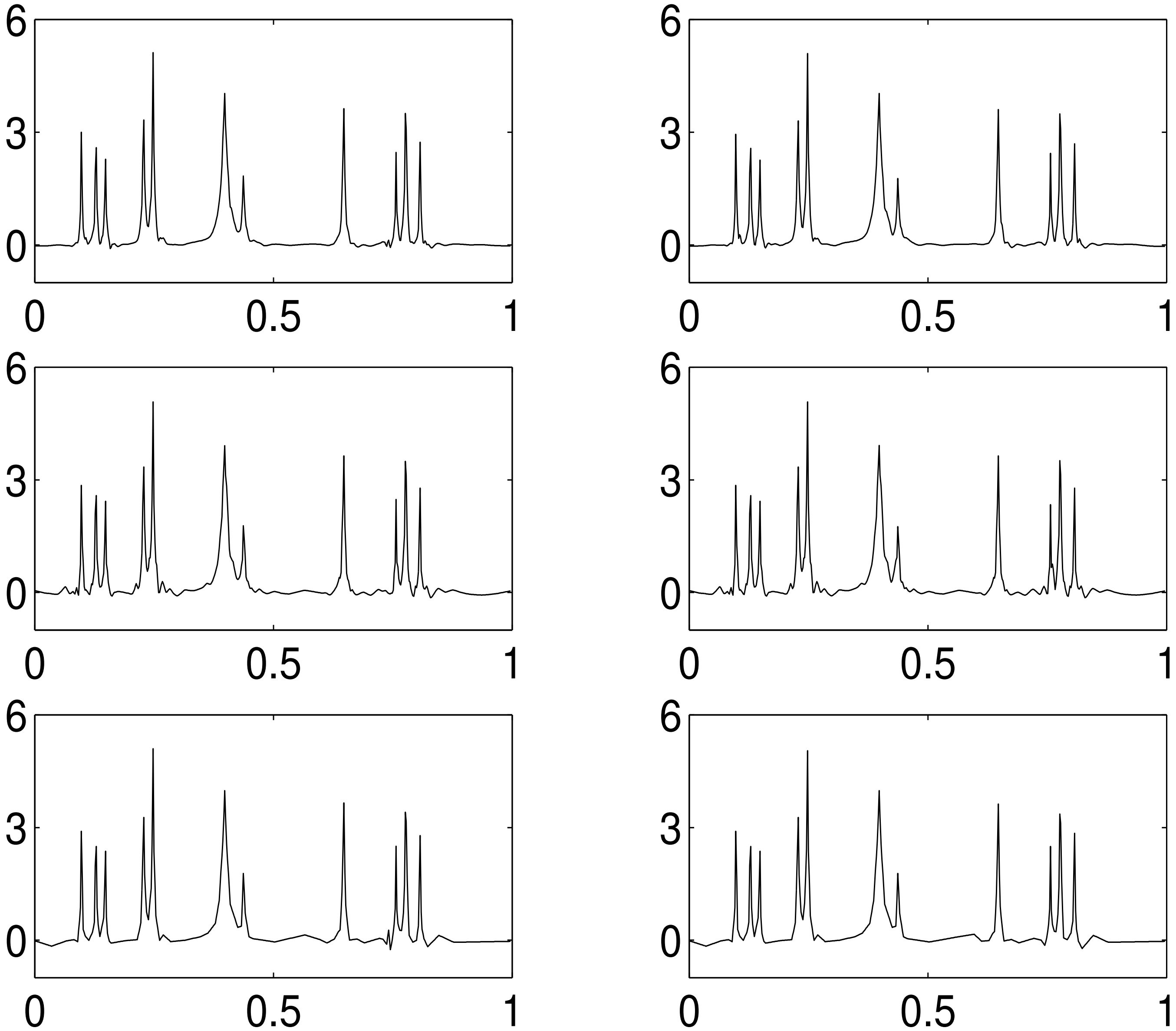

Fig. 9: Hard-threshold denoising of the *Bumps* test signal in Gaussian white noise with $\sigma = 0.1$ using different wavelet systems (redundant with D4 wavelet (*top*), biorthogonal D7-9 (*middle*), biorthogonal spline2-2 (*bottom*)). The leftmost plots use the uniform threshold $\sigma_l\sqrt{2\log N}$. The rightmost plots use the correlation-dependent threshold $\sqrt{2(1+\Delta_\phi(l))\log(lN)}$, $l = 1, \dots, L$.

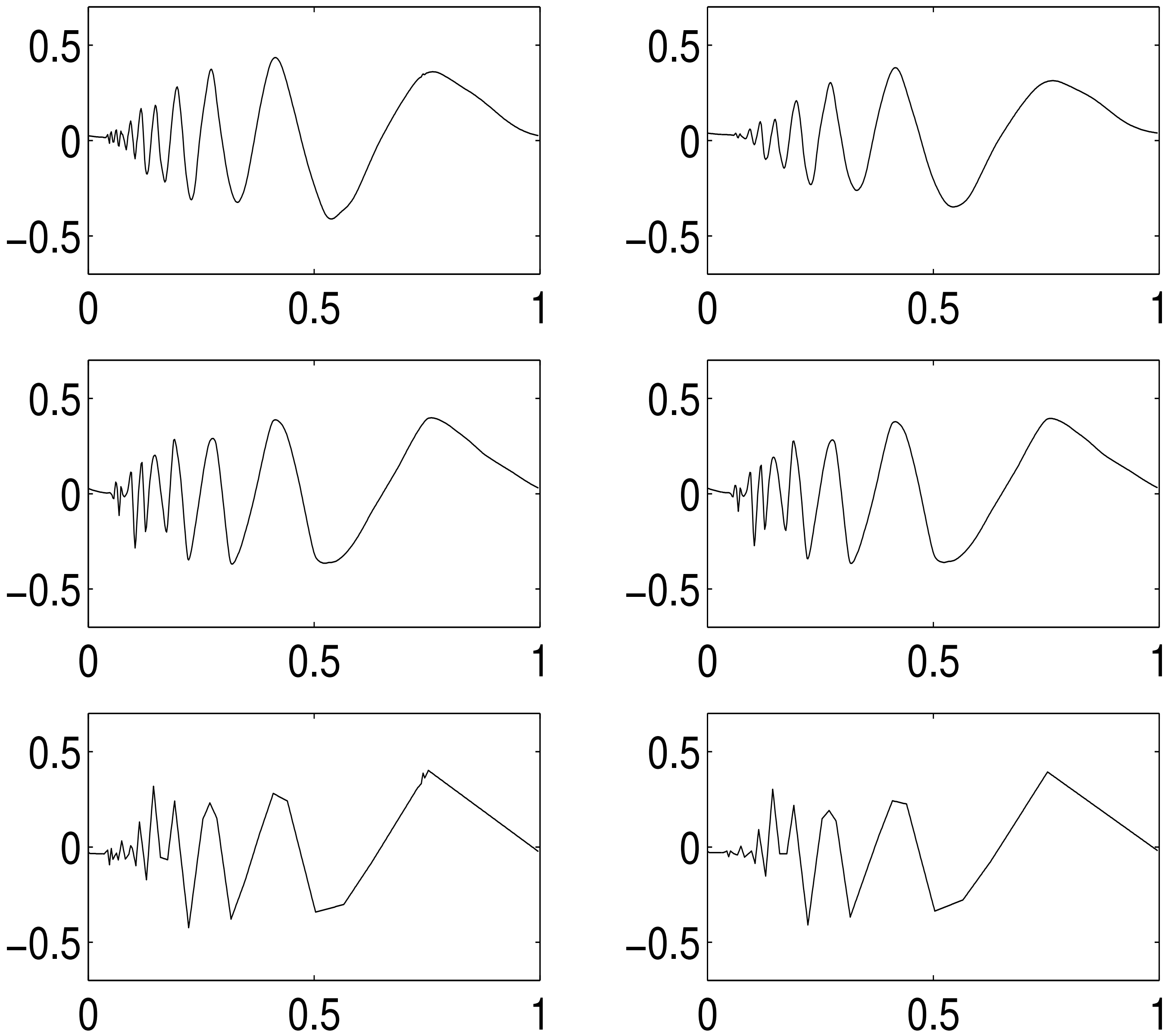

Fig. 10: Soft-threshold denoising of the *Doppler* test signal in Gaussian white noise with $\sigma = 0.1$ using different wavelet systems (redundant with D4 wavelet (*top*), biorthogonal D7-9 (*middle*), biorthogonal spline2-2 (*bottom*)). The leftmost plots use the uniform threshold $\sigma_l\sqrt{2\log N}$. The rightmost plots use the correlation-dependent threshold $\sqrt{2(1+\Delta_\phi(l))\log(lN)}$, $l = 1, \ldots, L$.