Smoothness estimates for soft-threshold denoising via translation invariant 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

February 25, 1998

Abstract

In this paper we study a generalization of the Donoho-Johnstone denoising model for the case of the translation invariant wavelet transform. Instead of soft-thresholding coefficients of the classical orthogonal discrete wavelet transform, we study soft-thresholding of the coefficients of the translation invariant discrete wavelet transform. This latter transform is not an orthogonal transformation. As a first step we construct a level-dependent threshold to remove all the noise in the wavelet domain. Subsequently, we use the theory of interpolating wavelet transforms to characterize the smoothness of an estimated denoised function. Based on the fact that the inverse of the translation invariant discrete transform includes averaging over all shifts, we use smoother autocorrelation functions in the representation of the estimated denoised function in place of Daubechies scaling functions.

*Technical Report CML TR 98-01, Computational Mathematics Laboratory, Rice University.

This work was supported in part by Texas ATP, DARPA/AFOSR F49620-97-1-0513, and Alexander von Humboldt-Stiftung.

1 Introduction

During the last decade, wavelet analysis has become a powerful tool in the analysis of functions and applications in signal processing. Via the wavelet transform, characteristics of a function are separated into wavelet coefficients at several scales. The manipulation and analysis of wavelet coefficients yields the desired compression, edge localization or denoising effects.

The theoretical basis for denoising was introduced by Donoho and Johnstone [13, 14] under the assumption of underlying Gaussian White Noise. In several papers, Donoho, together with his co-authors, developed a complete theory on denoising via orthogonal wavelet transforms, including different thresholding techniques and especially results concerning the smoothness of the estimated function as well as optimality properties of the presented estimators. For a list of references see [12] and the discussion in [4]. Inspired by this work several wavelet-based techniques were developed for different models, e.g. for Poisson [28] or Cauchy distributed noise [15] or using statistical methods such as hidden Markov models [8] .

Based on the original work done by Donoho and Johnstone, the principle of translation invariant denoising was introduced in [7, 22, 23]. It was shown in many experiments that thresholding of the coefficients of the translation invariant wavelet transform yields better results in terms of mean-squared-error and signal-to-noise-ratio. From the experimental work in [7, 22, 23] several questions have arisen: Why do we have to choose in the invariant case a higher threshold than the famous $\sqrt{2 \log N}$ in the orthogonal situation? Why do the results get "smoother"? For the orthogonal case the problem of how to find a proper threshold and the problem of characterizing the smoothness of the estimated function has been solved [13, 14]. But a similar theory for the translation invariant wavelet transform was still missing. In this paper, we want to close that gap and present a denoising model similar to that in [13, 14], but adapted to the translation invariant wavelet transform.

This paper is organized as follows. In Section 2 we give a brief review of the theory of orthonormal wavelet systems and introduce the required definitions. Characteristic properties of continuous translation invariant wavelet transforms and their relations to autocorrelation functions are presented in Section 3. The discrete translation invariant wavelet transform is introduced in Section 4. We summarize the matrix representation of this redundant wavelet transform and state relations to a-trous filters and the undecimated a-trous algorithm. A relation between discrete wavelet coefficients of sample values and theoretical wavelet coefficients is given by Donoho in [11]. In Section 5, we summarize the main theorems from [11] regarding interpolating and hybrid wavelet transforms and their significance in the characterization of smoothness in Besov and Triebel spaces. Section 6 contains the stochastical model for the translation invariant denoising. In contrast to the orthogonal situation, we have to consider correlations between the redundant wavelet coefficients. From these correlations we derive a level dependent threshold which also depends on the chosen wavelet. Combining all the results from the previous sections, we formulate and prove a generalization of the Donoho smoothness estimate for denoising via soft-thresholding for the translation invariant wavelet transform. Section 7 demonstrates the presented results in some numerical experiments. Finally, concluding remarks follow in Section 8.

2 Definition and properties of orthogonal wavelet systems

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 $\mathbf{g} = \{g_k : k \in \mathbb{Z}\}$. In this paper, we focus on the compactly supported Daubechies wavelet systems ([9]). In that case the scaling function ϕ is compactly supported in $[0, 2D-1]$. A collection of formal sums $\sum \beta_{j,k} 2^{j/2}\phi(2^j x - k)$ contains all polynomials of degree less than D. Under these conditions we say the wavelet system or ϕ is of degree $D-1$. The coefficients h_k and g_k are linked by $g_k = (-1)^k h_{1-k}$. Since we will have to compute the fast discrete wavelet transform in other parts of this paper we choose the normalization

$$\sum h_k = \sqrt{2}.$$

We define $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$. Then the orthogonality conditions on ϕ and ψ are

$$\begin{aligned} \langle \phi, \phi(\cdot - k)\rangle &= \delta_{0,k} && (1)\\ \langle \phi_{0,l}, \psi_{j,k}\rangle &= 0 \\ \langle \psi_{m,l}, \psi_{j,k}\rangle &= \delta_{m,j}\delta_{l,k} \quad \text{for all } j,k,l,m \in \mathbb{Z}. \end{aligned}$$

From the first condition the well-known relation

$$\sum h_k h_{k+2l} = \delta_{0,l}$$

follows. The scaling function ϕ induces a multiresolution analysis $\{V_m\}_{m\in\mathbb{Z}}$ where $V_m := \text{span}_{L^2}\{\phi_{m,k}\}_{k\in Z}$. 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 into 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 $P^j f = \sum_k \langle \phi_{j,k}, f\rangle \phi_{j,k}$ onto the subspace V_j and projections $Q^m f = \sum_k \langle \psi_{m,k}, f\rangle \psi_{m,k}$ onto the subspaces $W_m, m \geq j$, by

$$f = P^j f + \sum_{m\geq j} Q^m f. \tag{2}$$

From this expansion it follows that the first D moments of the wavelet function ψ vanish. When wavelet 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) ([5, 25]). If the DWT is applied to a sequence of N sampled data, the computational complexity of this algorithm is $O(N)$.

It follows from the definition of the projection that $P^j f$ is at least as smooth as the scaling function ϕ. Since the regularity of ϕ increases with D (see e.g. [9]) the corresponding projections get also smoother. With $Reg(\phi) \in \mathbb{N}_0$ we denote the highest order of continuous differentiability of ϕ, i.e. $\phi \in C^{Reg(\phi)}, \phi \notin C^{Reg(\phi)+1}$. The approximation property of $P^j f$ to the function f is bounded by the number of vanishing moments of the wavelet function and is given by the following well known theorem (see e.g. [29]).

Theorem 2.1 *Let P^j be the projection based on the Daubechies scaling function of degree $D-1$. Further suppose that f is in $C^n, 1 \leq n \leq D$ and has compact support. Then*

$$\| f - P^j f \|_{L_2} \leq C 2^{-jn}.$$

3 The continuous translation invariant wavelet transform

One disadvantage of the classical orthogonal wavelet transform from the previous section is that it is not translation invariant. To eliminate this fact we consider all shifted versions of the orthogonal transform and average over all those shifts. The result is a transform that is not orthogonal anymore and, moreover, contains redundant information. For the discrete situation (see Section 4) this idea was introduced independently in [2, 20, 26, 31]. Some aspects of the continuous case were studied in [7]. In the following we show how the redundancy affects characteristics of projections such as vanishing moments, smoothness and approximation order. In [7], some of the following results were stated for the Haar system ($D = 1$) in the context of denoising. We generalize that approach to the whole Daubechies family and embed it in the *autocorrelation transform*. Following [7], we denote the shift of size h by S_h. Then the redundant projections $P_R^j[f]$ and $Q_R^j[f]$ are defined by

$$\begin{aligned} P_R^j[f] &= 2^j \int_0^{2^{-j}} S_{-h} P^j S_h[f] dh \quad \text{and} \\ Q_R^j[f] &= 2^j \int_0^{2^{-j}} S_{-h} Q^j S_h[f] dh. \end{aligned}$$

These projections can be expressed in terms of the autocorrelation functions of the scaling function ϕ and the wavelet function ψ.

Lemma 3.1 *Let ϕ, ψ be the Daubechies scaling and wavelet function of order N. Then*

$$\begin{aligned} P_R^j[f](y) &= 2^j \int_{-\infty}^{+\infty} f(x)\Phi(2^j(x-y))dx, \\ Q_R^j[f](y) &= 2^j \int_{-\infty}^{+\infty} f(x)\Psi(2^j(x-y))dx, \end{aligned}$$

where

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

are the autocorrelation functions of ϕ and ψ.

Proof:

$$2^{-j} P_R^j[f](y) = \int_0^{2^{-j}} S_{-h} P^j S_h[f](y) dh$$

$$
\begin{aligned}
&= \int_0^{2^{-j}} \sum_{k\in\mathbb{Z}} \int_{-\infty}^{\infty} f(x+h)\phi_{j,k}(x)dx\phi_{j,k}(y-h)dh \\
&= \int_0^{1} \sum_{k\in\mathbb{Z}} \int_{-\infty}^{\infty} f(x+2^{-j}h)\phi_{j,k}(x)dx\phi_{j,k}(y-2^{-j}h)2^{-j}dh \\
&= \int_0^{1} \sum_{k\in\mathbb{Z}} \int_{-\infty}^{\infty} f(x)\phi(2^jx-k-h)dx\phi(2^jy-k-h)dh \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x)\phi(2^jx-s)\phi(2^jy-s)ds \\
&= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{\infty} \phi(2^jx+t-2^jy)\phi(t)dt \\
&= \int_{-\infty}^{\infty} f(x) \int_{-\infty}^{\infty} \phi(2^j(x-y)+t)\phi(t)dt \\
&= \int_{-\infty}^{\infty} f(x)\Phi(2^j(x-y))dx
\end{aligned}
$$

The result for $Q_R^j[f]$ follows in the same way. $\square$

In the terminology of [19], the projections P_R^j, Q_R^j are *single voices* of the continuous wavelet transform.

Lemma 3.2 *1. The projections $P_R^j[f]$ and $Q_R^j[f]$ are translation invariant, i.e. for $\delta \in \mathbb{R}$ we have*

$$
\begin{aligned}
P_R^j[f(\cdot+\delta)](y) &= P_R^j[f](y+\delta) \quad \textit{and} \\
Q_R^j[f(\cdot+\delta)](y) &= Q_R^j[f](y+\delta).
\end{aligned}
$$

2. A function $f \in L^2(\mathbb{R})$ has the following representation

$$
f = P_R^j[f] + \sum_{m\geq j} Q_R^m[f]. \tag{3}
$$

We call this representation the continuous translation invariant wavelet transform *or the* autocorrelation transform of f with respect to *ϕ.*

Proof:

1.

$$\begin{aligned}
P_R^j[f(\cdot+\delta)](y) &= 2^j \int_{-\infty}^{\infty} f(x+\delta)\Phi(2^j(x-y))dx \\
&= 2^j \int_{-\infty}^{\infty} f(x)\Phi(2^j(x-\delta-y))dx \\
&= P_R^j[f](y+\delta)
\end{aligned}$$

2. From the translation invariance and Eq.(2) it follows:

$$\begin{aligned}
\int_0^{2^{-j}} f(y)dh &= 2^{-j}f(y) = \int_0^{2^{-j}} P^j[f(\cdot+h)](y-h)dh + \int_0^{2^{-j}} \sum_{m\geq j} Q^m[f(\cdot+h)](y-h)dh \\
&= 2^{-j}P_R^j[f](y) + \sum_{m\geq j} \sum_{l=0}^{2^{-j+m}-1} \int_{l2^{-m}}^{(l+1)2^{-m}} Q^m[f(\cdot+h)](y-h)dh \\
&= 2^{-j}P_R^j[f](y) + \sum_{m\geq j} \sum_{l=0}^{2^{-j+m}-1} \int_0^{2^{-m}} Q^m[f(\cdot+h+l2^{-m})](y-h-l2^{-m})dh \\
&= 2^{-j}P_R^j[f](y) + \sum_{m\geq j} 2^{-j+m} Q^m[f(\cdot+h)](y-h)dh \\
&= 2^{-j}P_R^j[f](y) + 2^{-j} \sum_{m\geq j} Q_R^m[f](y)
\end{aligned}$$

□

In the context of wavelet theory the functions Φ and Ψ were first studied by Beylkin in [1] and its applications in [30], but they are also known as *Deslauriers-Dubuc Fundamental functions* [10]. In [11], they are considered as special *interpolating scaling functions.*

Properties of the functions Φ and Ψ: ([1, 3])

1. *Symmetry.* It follows from the definition that Φ and Ψ are symmetric functions.

2. *Scaling equation.*

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

where

$$a_k = \sum h_l h_{l+k}$$

are the *autocorrelation coefficients* of $\mathbf{h}$. These coefficients further satisfy

$$\sum a_{2k} = \delta_{0,k} \quad \text{and} \quad \sum a_k = 2.$$

They are also known as "a-trous" filters ([20]).

3. *Vanishing moments.* Φ and Ψ have the Coiflet property of order $2D$, i.e.

$$\begin{aligned} \int \Phi(x)dx &= 1 \\ \int x^p \Phi(x)dx &= 0 \quad \text{for} \quad p = 1, \ldots 2D-1 \\ \int x^p \Psi(x)dx &= 0 \quad \text{for} \quad p = 0, \ldots 2D-1. \end{aligned}$$

4. *Smoothness.* Roughly speaking Φ is twice as smooth as ϕ in the Sobolev-norm, because this norm is measured by the decay of the Fourier transform of ϕ and $\hat{\Phi} = |\hat{\phi}|^2$.

5. *Interpolation property.* Because of the property $\sum a_{2k} = \delta_{0,k}$, Φ and Ψ are interpolating functions. That means we can start with the autocorrelation coefficients as special interpolation points of Φ and compute more interpolation points using the scaling equation. Note that the functions ϕ and ψ are not interpolating functions.

6. We want to mention that Φ and Ψ also occur as synthesis scaling and wavelet functions of biorthogonal Coiflet wavelet systems [35]. Another relation between a pair of biorthogonal scaling functions and its correlation function is presented in [34] in the context of the lifting-scheme.

Based on the number of vanishing moments $2D$ of the wavelet function Ψ we get the following result for the approximation order of $P_R^j[f]$. The result was stated in [7] for the case $D = 1$ where ϕ is the Haar function. Here we generalize that result for $D > 1$.

Theorem 3.1 *Suppose that f is in $C^{n+\alpha}, n \in \mathbb{N}_0,\ 0 < \alpha \leq 1$, and has compact support. Let P_R^j be the redundant projection based on the Daubechies scaling function of order D. Then*

$$\| f - P_R^j[f] \|_{L_2} \leq C 2^{-j\gamma}$$

where $\gamma = \min(\alpha, 2D)$.

Proof: With Eq.(3) we get:

$$
\begin{aligned}
\| f - P_R^j[f] \|_{L_2}^2 &= \| \sum_{m \geq j} Q_R^j[f] \|_{L_2}^2 \\
&= \int \left[\sum_{m \geq j} 2^m \int f(x) \Psi(2^m(x-y)) dx \right]^2 dy \\
&= \int \left[\sum_{m \geq j} 2^m \int 2^{-m} f(2^{-m}t + y) \Psi(t) dt \right]^2 dy
\end{aligned}
$$

Because of the number of vanishing moments of Ψ we can substitute $\int f(2^{-m}t+y)\Psi(t)dt$ by $\int [f(2^{-m}t+y) - T_M[y](t)]\Psi(t)dt$, where $T_M[y]$ is the Taylor polynomial of order M of f in y with $M = \min(n, 2D-1)$. Therefore

$$
\begin{aligned}
\| f - P_R^j[f] \|_{L_2}^2 &= \int \left[\sum_{m \geq j} \int (f(2^{-m}t+y) - T_M[y](t)) \Psi(t) dt \right]^2 dy \\
&\leq \int \left[\sum_{m \geq j} \int |f(2^{-m}t+y) - T_M[y](t)| |\Psi(t)| dt \right]^2 dy \\
&\leq \int \left[\sum_{m \geq j} \int C_1 |2^{-m}t|^{\gamma} |\Psi(t)| dt \right]^2 dy \quad \text{for} \quad \gamma = \min(\alpha, 2D) \\
&\leq \int \left[\sum_{m \geq j} C_2 2^{-\gamma m} \right]^2 dy \\
&\leq C_3 \left[\sum_{m \geq j} 2^{-\gamma m} \right]^2 \\
&= C_3 \left[\sum_{m=0}^{\infty} 2^{-\gamma m} - \sum_{m=0}^{j-1} 2^{-\gamma m} \right]^2 \\
&= C_3 \left[\frac{1}{1 - 2^{-\gamma}} - \frac{2^{-j\gamma} - 1}{2^{-\gamma} - 1} \right]^2 \\
&= C_3 \left[\frac{1 + 2^{-j\gamma} - 1}{1 - 2^{-\gamma}} \right]^2 \\
&= C_4 \, 2^{-2j\gamma}
\end{aligned}
$$

□

A comparison to the result for the projections P^j of the classical non-redundant WT in Theorem 2.1 shows that introduction of redundancy decreases the approximation order for hölder continuous functions of non-integer order α and also for functions $f \in C^k$ for $k > D$.

4 The translation invariant discrete wavelet transform

The discrete wavelet transform of a vector $\mathbf{x} \in \mathbb{R}^{\mathbf{N}}$, i.e. the result of the pyramid transformation derived from an orthonormal wavelet basis having compact support, can be represented by an orthogonal $N \times N$ matrix W. The inverse transform is simply given by W^T. The pyramid algorithm includes down-sampling at each scale (see e.g. [5, 33]).

The discrete analogue to the continuous translation invariant WT (Eq.(3)) is the translation invariant DWT (TIDWT) which relinquishes down-sampling at each scale, but computes all shifts and, therefore, keeps all wavelet coefficients at each scale and all scaling coefficients at the coarsest scale. The scaling coefficients at scale j are given by the vector

$$c^{(j)} = (c_1^{(j)}, \ldots, c_N^{(j)}) \quad \text{with} \quad c_k^{(j)} = \sum_{l \in \mathbb{Z}} h_l c_{k+2^{j-1}l}^{(j-1)} \tag{4}$$

and the wavelet coefficients by

$$d^{(j)} = (d_1^{(j)}, \ldots, d_N^{(j)}) \quad \text{with} \quad d_k^{(j)} = \sum_{l \in \mathbb{Z}} g_l c_{k+2^{j-1}l}^{(j-1)} \tag{5}$$

where $j \leq \log_2 N$.

This is one possible way to include redundant information into the wavelet transforms. Other approaches are, e.g., wavelet packets or best basis representations [6]. The TIDWT was studied independently by many authors. For a discussion we refer to [4]. In the context of denoising the TIDWT was first studied in [7, 18, 23]. We follow [18, 23] and define the TIDWT and its inverse in a matrix representation setting.

The TIDWT is given by a $(L+1)N \times N$ matrix M_L, where L is the maximal level of decomposition. If we denote by R_j the $N \times N$ matrix which yields the wavelet coefficients at scale j, i.e. $R_j(\mathbf{x}) = \mathbf{d}^{(j)}$, and by S_j the matrix which produces the scaling coefficients at scale j, i.e. $S_j(\mathbf{x}) = \mathbf{c}^{(j)}$, then

$$M_L = (R_1^T, \ldots, R_L^T, S_L^T)^T. \tag{6}$$

It has been shown that the TIDWT has the computational complexity of order $O(N \log N)$.

Since M_L is not an orthogonal matrix anymore as in the decimated case, the inverse transform is not simply given by the transposed matrix. From matrix theory (see e.g. [32, 33]) it follows that there exist generalized (*Moore-Penrose*) inverses. In order to keep the degrees of freedom as small as possible, from all generalized inverses of M_L we choose the *pseudo inverse*

$$M_L^\dagger = (M_L^T M_L)^{-1} M_L^T. \tag{7}$$

The matrix $M_L^\dagger$ has the form ([18])

$$M_L^\dagger = (\frac{1}{2}R_1^T, \ldots, \frac{1}{2^L}R_L^T, \frac{1}{2^L}S_L^T). \tag{8}$$

Now we want to analyze if there exists, similar to the continuous situation, a relation to a "discrete" autocorrelation transform. In Section 3, we showed that the autocorrelation functions Φ and Ψ satisfy a scaling equation with coefficients a_k and $(-1)^k a_k$. These coefficients are also known as *a-trous filters* ([20, 31]). We denote the transform which produces the scaling coefficients computed by the undecimated a-trous-algorithm at scale j by T_j and the transform for the wavelet coefficients by Q_j. If we denote the elements of $\mathbf{x}$ by x_i and the elements of $T_j(\mathbf{x})$ by $t_{j,k}(\mathbf{x})$, then

$$t_{j,k}(\mathbf{x}) = \sum_{m_1,\ldots,m_j} a_{m_1} \cdots a_{m_j} x_{k+m_1+\cdots+2^j m_j}. \tag{9}$$

We observe the following relations between the matrices T_j, S_j and Q_j, R_j:

$$2^{-j} S_j^T S_j = T_j \quad \text{and} \quad 2^{-j} R_j^T R_j = Q_j. \tag{10}$$

Therefore, we call $T_j(\mathbf{x})$ the *discrete translation invariant projection at scale* j and remember the following equality:

$$T_j(\mathbf{x}) = M_j^\dagger \circ (\mathbf{0}, \ldots, \mathbf{0}, S_j^T)^T(\mathbf{x}). \tag{11}$$

Setting whole scales of the TIDWT to zero can therefore be interpreted as a kind of lifting step, because we get a representation of the approximized sequence as the output of a filtering algorithm using not the Daubechies coefficients h_k, but the the a-trous filter coefficients a_k. Figure 1 shows the discrete projections for the non-redundant and the redundant case.

5 Donoho's interpolating wavelet transform theory

One of the properties of wavelet coefficients that makes wavelet transforms such a successful tool for describing smoothness of functions is their characterization of Besov and Triebel spaces ([16, 27, 36]). These function spaces contain Hölder (-Zygmund) and Sobolev spaces. Given the representation of f in scaling and wavelet functions from Eq.(2) we define the vector

$$\theta = ((\gamma_{j_0,\cdot}), (\delta_{j_0,\cdot}), (\delta_{j_0+1,\cdot}), \ldots)$$

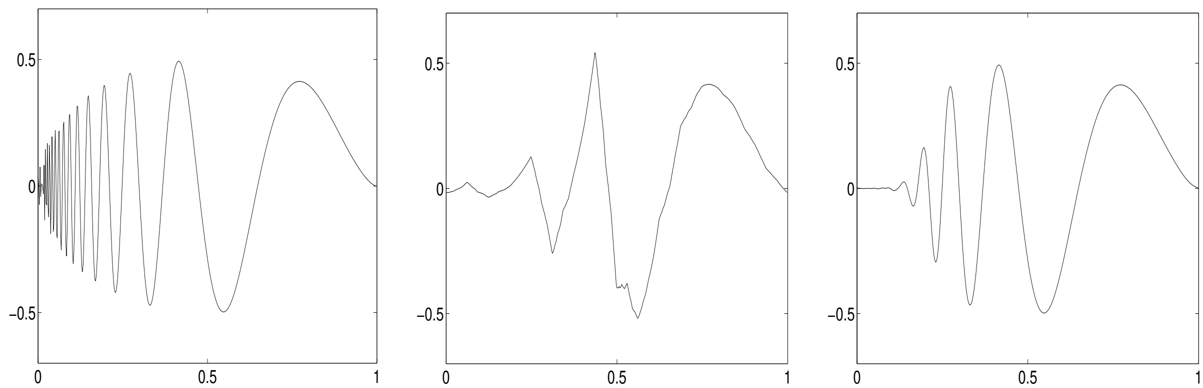

Figure 1: The original signal *Doppler* from the WaveLab package for N=512 (*left*), orthogonal discrete projection for $j = 5$ (*middle*), discrete translation invariant projection T_j for $j = 5$ (*right*).

of scaling coefficients $\gamma_{j_0,k} = \langle f, \phi_{j_0,k} \rangle$ and wavelet coefficients

$$\delta_{j,k} = \langle f, \psi_{j,k} \rangle \quad \text{for} \quad j \geq j_0. \tag{12}$$

It is shown in [16, 27] that the sequence norm

$$\| \theta \|_{b^\sigma_{p,q}} := \| (\gamma_{j_0,.}) \|_{l^p} + (\sum_{j \geq j_0} (2^{jsq} (\sum_{k=0}^{2^j - 1} |\delta_{j,k}|^p)^{1/p})^q)^{1/q},$$

where $s = \sigma + 1/2 - 1/p$, is equivalent to $\| f \|_{B^\sigma_{p,q}}$ ($\| \theta \|_{b^\sigma_{p,q}} \asymp \| f \|_{B^\sigma_{p,q}}$) for $1/p < \sigma < \min(Reg(\phi), D-1), p, q \in (0, \infty]$. Similarly, the norm

$$\| \theta \|_{f^\sigma_{p,q}} := \| (\gamma_{j_0,.}) \|_{l^p} + \| (\sum_{j \geq j_0} 2^{jsq} \sum_{k=0}^{2^j - 1} |\delta_{j,k}|^q \chi_{j,k})^{1/q} \|_{l^p},$$

where $s = \sigma + 1/2$ and $\chi_{j,k}$ the indicator function of the interval $[k/2^j, (k+1)/2^j]$, is equivalent to $\| f \|_{F^\sigma_{p,q}}$ for $1/p < \sigma < \min(Reg(\phi), D-1), p, q \in (0, \infty]$. For further details on Besov and Triebel spaces and their relations to wavelet coefficients, we refer to [16, 27, 36].

In applications we usually compute the discrete wavelet transform of a finite number of function samples of a function $f \in C[0, 1]$. The question is whether from this sequence of discrete wavelet transform coefficients, that are not of the form in Eq.(12), we can draw conclusions on the smoothness of the underlying function. In [29], results similar to that in Theorem 2.1 are derived where $P^j f$ is replaced by $S^j f = \sum_k f(2^{-j}k) 2^{j/2} \phi(2^j x - k)$. But these results do not capture the more detailed characterization of smoothness in Besov/Triebel norms. Donoho showed in [11] that the wavelet coefficients derived from the DWT of samples of f are in fact the leading theoretical coefficients for a nicely behaved transform of continuous functions. In the following we summarize the main theorems from [11]. First Donoho uses the autocorrelation function Φ to define *interpolating wavelet transforms.*

Theorem 5.1 *Let j_0 be a non-negative integer satisfying $2^{j_0} > 4D$. Then there exists a collection of functions $\tilde{\phi}_{j_0,k}$ and $\tilde{\psi}_{j,k}$ such that every $f \in C[0,1]$ has a representation*

$$f = \sum_k \beta_{j_0,k} \tilde{\phi}_{j_0,k} + \sum_{j \geq j_0} \sum_k \alpha_{j,k} \tilde{\psi}_{j,k} \tag{13}$$

with uniform convergence of partial sums $j \leq j_1$ as $j_1 \to \infty$. The scaling coefficients $\beta_{j_0,k}$ are given by

$$\beta_{j_0,k} = 2^{-j_0/2} f(2^{-j_0} k) \quad \text{for } 0 \leq k \leq 2^{j_0}.$$

and, with $P_I^j[f] = \sum_{k=0}^{2^j-1} \beta_{j,k} \tilde{\phi}_{j,k}$

$$\alpha_{j,k} = 2^{-j/2}(f(2^{-j}(k+1)/2) - P_I^j[f](2^{-j}(k+1)/2)).$$

The functions $\tilde{\phi}_{j,k}$ are derived from the function $\Phi_{j,k} = 2^j \Phi(2^j x - k)$ as follows. At the interior of the interval,

$$\tilde{\phi}_{j,k} = \Phi_{j,k}|_{[0,1]}, \quad 2D - 1 < k < 2^j - 2D - 2$$

and at the "edges" of the interval, they are dilations of certain special boundary-adjusted functions

$$\begin{aligned} \tilde{\phi}_{j,k} &= 2^{j/2} \phi_k^\sharp (2^{-j} x - k), \quad 0 \leq k \leq 2D - 1 \\ \tilde{\phi}_{j,2^j-k-1} &= 2^{j/2} \phi_k^\flat (2^{-j} x - 2^j - k - 1), \quad 0 \leq k \leq 2D - 1. \end{aligned}$$

The functions $\psi_{j,k}$ are constructed as follows. In the interior of the interval $[0,1]$ they are defined as

$$\tilde{\psi}_{j,k} = \Phi_{j+1,2k+1}|_{[0,1]}, \quad D - 1 \leq k < 2^j - D + 1 \tag{14}$$

and at the "edges" of the interval, they are dilations of certain special boundary-adjusted functions

$$\begin{aligned} \tilde{\psi}_{j,k} &= 2^{j/2} \psi_k^\sharp (2^{-j} x - k), \quad 0 \leq k \leq D - 1 \\ \tilde{\phi}_{j,2^j-k-1} &= 2^{j/2} \psi_k^\flat (2^{-j} x - 2^j - k - 1), \quad 0 \leq k \leq D - 1. \end{aligned}$$

The functions $\tilde{\phi}_{j_0,k}$ and $\tilde{\psi}_{j,k}$ have the same degree and the same regularity as Φ. The representation in Eq.(13) is called an interpolating wavelet transform of f.

Note that the wavelet functions $\tilde{\psi}_{j,k}$ (Eq.(14)) are different from the autocorrelation wavelet functions $\Psi_{j,k}$. Similar to the classical orthogonal wavelet transform there exists a characterization of smoothness in Besov/Triebel spaces.

Theorem 5.2 *Starting with the same assumptions as in Theorem 5.1, we define θ as*

$$\theta = ((\beta_{j_0,.}), (\alpha_{j_0,.}), (\alpha_{j_0+1,.}), \ldots). \tag{15}$$

Then

$$\| \theta \|_{b^{\sigma}_{p,q}} = \| (\beta_{j_0,.}) \|_{l^p} + (\sum_{j \geq j_0} (2^{jsq} (\sum_{k=0}^{2^j-1} |\alpha_{j,k}|^p)^{1/p})^q)^{1/q}, \tag{16}$$

where $s = \sigma + 1/2 - 1/p$, gives an equivalent norm for the Besov spaces $B^{\sigma}_{p,q}[0,1]$ where $1/p < \sigma < \min(Reg(\Phi), 2D-1), p, q \in (0, \infty]$. For Triebel-Lizorkin spaces we get a similar result:

$$\| \theta \|_{f^{\sigma}_{p,q}} = \| (\beta_{j_0,.}) \|_{l^p} + \| (\sum_{j \geq j_0} 2^{jsq} \sum_{k=0}^{2^j-1} |\alpha^q_{j,k}| \chi_{j,k})^{1/q} \|_{l^p}, \tag{17}$$

where $s = \sigma + 1/2$ and $\chi_{j,k}$ the indicator function of the interval $[k/2^j, (k+1)/2^j]$, gives an equivalent norm for the Triebel spaces $F^{\sigma}_{p,q}[0,1]$ where $1/p < \sigma < \min(Reg(\Phi), 2D-1), p, q \in (0, \infty]$.

These interpolating wavelet transforms are parts of the *hybrid wavelet transforms* that are defined in the next theorem.

Theorem 5.3 *We consider a function $f \in C[0,1]$ and its samples values $f_i = n^{-1/2} f(i/n)$. We denote the coefficients that are computed via the discrete wavelet transform of the samples f_i by $((\tilde{\beta}_{j_0,.}), (\tilde{\alpha}_{j_0,.}), (\tilde{\alpha}_{j_0+1,.}), \ldots, (\tilde{\alpha}_{j_1-1,.}))$. Then there are smooth functions $\phi^*_{j_0,k}, \psi^*_{j,k}$, $0 \leq k < 2^j, j \leq j_0$ which have the same order of differentiability and degree as ϕ and have compact support width $\leq 2^j$, such that every function $f \in C[0,1]$ has an expansion*

$$f = \sum_k \tilde{\beta}_{j_0,k} \phi^*_{j_0,k} + \sum_{j \geq j_0} \sum_k \tilde{\alpha}_{j,k} \psi^*_{j,k} \tag{18}$$

with uniform convergence of partial sums.

For $j_1 < j \leq j_0$ the functions $\phi^*_{j_0,k}, \psi^*_{j,k}$ are constructed as special linear combinations of $\phi_{j,k}$, but with the constrains that $\phi^*_{j_0,k}$ and $\psi^*_{j,k}$ are orthonormal with respect to the sampling inner product $\langle f, g \rangle_n = 1/n \sum_{k \in \mathbb{Z}} f(2^{-n}k) g(2^{-n}k)$. For $j \geq j_1$ the functions $\psi^*_{j,k}$ are the interpolating wavelet functions $\tilde{\psi}_{j,k}$ from Theorem 5.1 and $\tilde{\alpha}_{j,k} = \alpha_{j,k}$. The representation in Eq.(18) is called the hybrid wavelet transform *of f.*

The leading coefficients of the hybrid wavelet transform of f are the DWT coefficients computed from the samples $f(2^{-j_1}k)$. The smoothness of the hybrid scaling functions is determined by the smoothness of ϕ. That is easily seen in the Fourier domain, since $\hat{\phi}^*_j(\omega) = \hat{\phi}(\omega)\ \overline{\hat{\phi}2^{-j}\omega)}$. Again we get a characterization in Besov/Triebel norms:

Theorem 5.4 *Under the assumptions of Theorem 5.3 the equivalence*

$$\| f \|_{B^\sigma_{p,q}} \asymp \| \tilde{\theta} \|_{b^\sigma_{p,q}} \tag{19}$$

holds for $\tilde{\theta} = (\tilde{\beta}_{j_0,.}, \tilde{\alpha}_{j_0,.}, \tilde{\alpha}_{j_0+1,.} \ldots)$ *with* $1/p < \sigma < \min(Reg(\phi), D-1)$ *and a constant of equivalence independent of* $j_0 - j_1 > 0$*. A similar result is valid for Triebel spaces:*

$$\| f \|_{F^\sigma_{p,q}} \asymp \| \tilde{\theta} \|_{f^\sigma_{p,q}} \tag{20}$$

with $1/p < \sigma < \min(Reg(\phi), D-1)$ *and a constant of equivalence independent of* $j_0 - j_1 > 0$*.*

In [11], Donoho pointed out that the extension of a finite sequence of wavelet coefficients of a function f, produced by the pyramid transform algorithm, by infinite many zeros shrinks the norm $\| \,.\, \|_{b^\sigma_{p,q}}$ ($\| \,.\, \|_{f^\sigma_{p,q}}$) and makes the function, which corresponds to the infinite sequence, smoother than f in the sense of the considered norm.

The projection $\sum_k \tilde{\beta}_{j_0,k}\phi^*_{j_0,k}$ of the interpolating wavelet transform has an approximation property similar to the one in Theorem 2.1, but the parameter n is bounded by $2D$ instead of D. This property is responsible for the different ranges of the parameter σ for interpolating and hybrid wavelet transforms. In the first case, σ is bounded by the regularity properties of the autocorrelation function Φ, in the hybrid wavelet transform by the less regular function ϕ. This will be the basis for the smoothness characterization of the denoised function in the following section.

6 Denoising by shrinkage of redundant wavelet coefficients

6.1 The stochastic model

In this section, we want to analyze how the Donoho-Johnstone denoising model can be modified and transfered to the TIDWT. As in [13, 14], we consider samples of a function $f \in C[0,1]$ that are corrupted by Gaussian White noise. In particular, we consider a sequence of given data points $\{d_i\}_{i=1}^N$ with

$$d_i = f_i + \sigma z_i \tag{21}$$

where $\{z_i\}$ is an i.i.d. sequence of N(0,1) random variables with $\sigma > 0$ and $f_i = f(i/N)$. We are interested in a denoising of the samples $\mathbf{d} = \{d_i\}$ via the TIDWT. That means we compute the

TIDWT of $\mathbf{d}$, manipulate the wavelet coefficients and get back a sequence of function sample values f_i^* that are in some norm "close" to the samples f_i.

Following [13, 14], the first step is to characterize the magnitude of noisy wavelet coefficients and find a threshold that removes the noise almost surely as N goes to infinity. In the decimated case, the wavelet transform of the data is given by

$$W(\mathbf{d}) = W(\mathbf{f}) + \sigma W(\mathbf{z}) \tag{22}$$

where $\mathbf{f}$ denotes the vector of the function sample values $f_i = f(i/N)$. Because of the orthogonality properties of the scaling and wavelet functions (Eq.(1)) it follows that $W(\mathbf{z})$ yields a sequence $\{\tilde{z}_i\}$ of identically distributed N(0,1) variables that are uncorrelated. Since for normal random variables, uncorrelatedness is equivalent to independence, the $\tilde{z}_i$ are i.i.d. N(0,1) random variables. At this point we want to mention that in order to emphasize the difference between the nonredundant and redundant denoising models we neglect the slight modifications that are made in [13] and that are due to the treatment of boundary effects. We refer the interested reader concerning this topic to [13]. The following theorem in [24] gives the famous general threshold $\sqrt{2\log N}$ for denoising via orthogonal wavelet transforms.

Theorem 6.1 *Let $\{z_i\}$ be an i.i.d. sequence of N(0,1) random variables. Then*

$$P\{\max_{i=1,\dots n} z_i \le \sqrt{2\log n}\} \quad \to \quad 1 \quad \textit{for} \quad n \to \infty.$$

With this result it is shown that applying the soft-thresholding operator

$$\eta_t(y) = sgn(y)(|y| - t)_+ \quad \text{for } t = t_N = \sqrt{2\log N}$$

removes a.s. all the noise in the wavelet coefficients as $N \to \infty$. This results combined with the characterization of smoothness in Besov and Triebel spaces via wavelet coefficients from Theorem 5.4, leads to the following theorem in [13].

Theorem 6.2 *Let $\{f_i^*\}$ be the sequence of estimated function values produced by soft-thresholding the wavelet coefficients $W(\mathbf{d})$ with the threshold $t_N = \sigma\sqrt{2\log N}$. Then there exists a smooth interpolation $\tilde{f}$ of the sequence $\{f_i^*\}$ on $[0,1]$ which is, with probability tending to 1, at least as smooth as f. In particular,*

$$P\{\| \tilde{f} \|_{\mathcal{F}} \le C_1 \| f \|_{\mathcal{F}} \ \forall \mathcal{F} \in \mathcal{S}\} \quad \to \quad 1 \quad \textit{for} \quad N \to \infty.$$

The class $\mathcal{S}$ is defined as the scale of all spaces $B^\sigma_{p,q}$, $F^\sigma_{p,q}$ which embed continuously in $C[0,1]$, so that $1/p < \sigma < \min(Reg(\phi), D)$.

The function $\tilde{f}$ is constructed in the way that the thresholded coefficients $\eta_{t_N} \circ W(\mathbf{d})$ are the leading coefficients of the hybrid wavelet transform of $\tilde{f}$. In the remaining part of this section we follow the presented steps from [13] to transfer the model to the TIDWT.

The first difference compared to the decimated situation is that the TIDWT of $\mathbf{z}$ produces not N, but $(L+1)N$ new random variables $\{\xi_i\}_{i=1}^{(L+1)N}$. These random variables are identically distributed $N(0,1)$, but not uncorrelated anymore. Therefore, the result from Theorem 6.1 does not apply. This situation is different from the model considered in [21] where the authors studied the orthogonal DWT of correlated noise. The following theorem from [24] covers the case of correlated random variables.

Theorem 6.3 *Let $\{\xi_i\}_{i=1}^n$ be an i.d. normal sequence of random variables with $Var(\xi_i) = 1$ and covariances $r_{ij} = Cov(\xi_i, \xi_j)$ such that*

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

Furthermore let F be the distribution function of the normal distribution. Then there exists a constant K, such that

$$|P\{\max_{i=1,\dots n} \xi_i \leq u\} - F(u)^n| \leq K \sum_{1 \leq i < j \leq n} |r_{ij}| \exp\left(-\frac{u^2}{1 + |r_{ij}|}\right). \tag{23}$$

If $z_1, \dots, z_n$ are i.i.d. N(0,1) random variables then $P(\max_{i=1,\dots,n} z_i \leq u) = F(u)^n$. Therefore, we can interpret the result from Theorem 6.3 in that way that it tells us how much the probability $P\{\max_{i=1,\dots n} \xi_i \leq u\}$ in the correlated case differs from the one in the independent case. If we find a threshold $u = u_{(L+1)N}$ such that the right term in Eq.(23) goes to zero for $N \to \infty$ and such that $u_{LN} \geq \sqrt{2\log((L+1)N)}$, then we know that $P\{\max_{i=1,\dots n} \xi_i \leq u\} \to 1$ for $N \to \infty$. A possible choice for such a threshold is given in the next lemma.

Lemma 6.1 *Let $\{\xi_i\}_{i=1}^{(L+1)N} = M_L(\mathbf{z})$ be the discrete redundant wavelet transform of the sequence $\mathbf{z} = \{z_i\}_{i=1}^N$ of i.i.d. N(0,1) random variables. We set $r_{ij} = Cov(\xi_i, \xi_j)$. If $\delta_L = \max_{i \neq j} |r_{ij}| < 1$ and*

$$u = u_{(L+1)N} = \sqrt{2(1+\delta_L)\log((L+1)N)} \tag{24}$$

then

$$\sum_{1 \leq i < j \leq (L+1)N} |r_{ij}| \exp\left(-\frac{u^2_{(L+1)N}}{1 + |r_{ij}|}\right) \quad \to \quad 0 \quad \textit{for} \quad N \to \infty. \tag{25}$$

Proof: From the orthogonality conditions on the scaling vectors $\mathbf{h}$ and $\mathbf{g}$ it follows that several of the correlations $r_{i,j}$ are zero. Particularly, for $1 \leq i < j \leq (L+1)N$ we obtain

$$\#\{r_{ij} \,|\, r_{ij} \neq 0\} \quad = \quad \#\{Cov(c_k^{(L)}, c_l^{(L)}) \,|\, Cov(c_k^{(L)}, c_l^{(L)}) \neq 0\}$$

$$
\begin{aligned}
&+ \quad \#\{Cov(c_k^{(L)}, d_l^{(m)}) \mid Cov(c_k^{(L)}, d_l^{(m)}) \neq 0, m = 1, \ldots, L\} \\
&+ \quad \#\{Cov(d_k^{(m)}, d_l^{(n)}) \mid Cov(d_k^{(m)}, d_l^{(n)}) \neq 0, m = 1, \ldots, L, n = 1, \ldots, m\} \\
&\leq \quad DN(2^L - 1) + DN2^L \sum_{l=1}^{L} \frac{2^l - 1}{2^l} + DN \sum_{m=1}^{L} 2^m \sum_{l=1}^{m} \frac{2^l - 1}{2^l} \\
&\leq \quad DN(2^L - 1 + 2^L L + L(2^{L+1} - 2)) \\
&\leq \quad DN(2^{L+1} + 2^{L+1} L + 2^{L+1} L) \\
&\leq \quad DN2^{L+1}(1 + 2L)
\end{aligned} \tag{26}
$$

With $u_{(L+1)N} = \sqrt{2(1+\delta_L)\log((L+1)N)}$ we get

$$
\begin{aligned}
\sum_{1 \leq i < j \leq (L+1)N} |r_{ij}| \exp\left(-\frac{u^2_{(L+1)N}}{1 + |r_{ij}|}\right) &= \sum_{1 \leq i < j \leq (L+1)N} |r_{ij}| \exp\left(-2\log((L+1)N)\frac{1+\delta_L}{1+|r_{ij}|}\right) \\
&\leq \sum_{1 \leq i < j \leq (L+1)N} |r_{ij}| \exp\left(-2\log((L+1)N)\right) \\
&\leq \frac{1}{((L+1)N)^2} DN2^{L+1}(1+2L) \quad \text{with (26)}
\end{aligned}
$$

For $L < \log_2 N$:

$$
\sum_{1 \leq i < j \leq (L+1)N} |r_{ij}| \exp\left(-\frac{u^2_{(L+1)N}}{1 + |r_{ij}|}\right) \leq \frac{D2^{L+1}(1+2L)}{(L+1)^2 N} \quad \to \quad 0 \quad (N \to \infty)
$$

For $L = \log_2 N$:

$$
\sum_{1 \leq i < j \leq (L+1)N} |r_{ij}| \exp\left(-\frac{u^2_{LN}}{1 + |r_{ij}|}\right) \leq \frac{2DN^2(1 + 2\log_2 N)}{N^2(\log_2 N + 1)^2} = \frac{2D(1 + 2\log_2 N)}{(\log_2 N)^2} \quad \to \quad 0
$$

for $N \to \infty$. $\square$

Remark: If we choose $u_{(L+1)N} = \sqrt{2\log((L+1)N)}$ we get

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

Using the estimation in Eq.(26) we can not a-priori guarantee that $((L+1)N)^{-1} \sum_{1 \leq i < j \leq (L+1)N} |r_{ij}|$ tends to zero for $N \to \infty$. Therefore, the threshold $u_{(L+1)N} = \sqrt{2\log((L+1)N)}$ does not work in general.

In order to determine the threshold $u_{(L+1)N}$ from Lemma 6.1, we need the upper bound δ_L of the covariances r_{ij}. First we show that the covariances occur as the values of the correlation functions between $\phi_{j,k}$ and $\psi_{j,k}$ at dyadic rationals.

Lemma 6.2 *Let* $\mathbf{z} = (z_1, \ldots, z_N)$ *be an i.i.d. sequence of* $N(0,1)$ *random variables. Furthermore, let* $c_k^{(j)}, d_k^{(j)}$ *be the scaling and wavelet coefficients computed with the filtering scheme described in Eq.(4) and (5). Then the following formulas hold:*

$$\begin{aligned}
Cov(c_k^{(j)}, c_{k+r}^{(j)}) &= \langle \phi, \phi(\cdot - 2^{-j}r)\rangle = \Phi(2^{-j}r) \\
Cov(d_k^{(j)}, d_{k+r}^{(j)}) &= \langle \psi, \psi(\cdot - 2^{-j}r)\rangle = \Psi(2^{-j}r) \\
Cov(c_k^{(l)}, d_{k+r}^{(j)}) &= \langle \phi_{j-l}, \psi(\cdot - 2^{-j}r)\rangle \\
Cov(d_k^{(l)}, d_{k+r}^{(j)}) &= \langle \psi_{j-l}, \psi(\cdot - 2^{-j}r)\rangle.
\end{aligned}$$

Proof:

1. We show $Cov(c_k^{(j)}, c_{k+r}^{(j)}) = \Phi(2^{-j}r)$ by induction: For $j=1$ we get

$$\begin{aligned}
Cov(c_k^{(1)}, c_{k+r}^{(1)}) &= \sum_{l,m} h_l h_m Cov(z_{k+l}, z_{k+r+m}) \\
&= \sum_m h_{m+r} h_m \\
&= a_r = \Phi(\frac{r}{2})
\end{aligned}$$

$$\begin{aligned}
Cov(c_k^{(j)}, c_{k+r}^{(j)}) &= \sum_{l,m} h_l h_m Cov(c_{k+2^{j-1}l}^{(j-1)}, c_{k+r+2^{j-1}m}^{(j-1)}) \\
&= \sum_{l,m} h_l h_m \Phi(2^{-j+1}r + m - l) \\
&= \sum_t \Phi(2^{-j+1}r + t) \sum_m h_{m+t} h_m \\
&= \sum_t \Phi(2^{-j+1}r + t) a_t \\
&= \Phi(2^{-j}r)
\end{aligned}$$

2. The equality $Cov(d_k^{(j)}, d_{k+r}^{(j)}) = \Psi(2^{-j}r)$ is proven analogously to the previous case.

3. First we show $Cov(c_k^{(j)}, d_{k+r}^{(j)}) = \langle \phi, \psi(\cdot - 2^{-j}r)\rangle$ by induction: For $j=1$ we get

$$\begin{aligned}
Cov(c_k^{(1)}, d_{k+r}^{(1)}) &= \sum_{l,m} h_l g_m Cov(z_{k+l}, z_{k+r+m}) \\
&= \sum_m h_{m+r} g_m
\end{aligned}$$

On the other hand

$$
\begin{aligned}
\langle \phi, \psi(\cdot - 2^{-1}r)\rangle &= \int \phi(x)\psi(x-\frac{r}{2})dx \\
&= 2\sum_{l,m} h_l g_m \int \phi(2x-l)\phi(2x-r-m)dx \\
&= \sum_{l,m} h_l g_m \int \phi(x-l)\phi(x-r-m)dx \\
&= \sum_{m} h_{m+r} g_m
\end{aligned}
$$

Now we consider the step from $j-1$ to j:

$$
\begin{aligned}
Cov(c_k^{(j)}, d_{k+r}^{(j)}) &= \sum_{l,m} h_l g_m Cov(c^{(j-1)}_{k+2^{j-1}l}, c^{(j-1)}_{k+r+2^{j-1}m}) \\
&= \sum_{l,m} h_l g_m \langle \phi, \phi(\cdot - 2^{-j+1}(r+2^{j-1}m-2^{j-1}l))\rangle \\
&= \sum_{l,m} h_l g_m \langle \phi, \phi(\cdot - 2^{-j+1}r - m + l)\rangle
\end{aligned}
$$

On the other hand

$$
\langle \phi, \psi(\cdot - 2^{-j}r)\rangle = \sum_{l,m} h_l g_m \int \phi(x-l)\phi(x-2^{-j+1}r-m)dx
$$

The result for $Cov(c_k^{(l)}, d_{k+r}^{(j)}), l \neq j$, we get in a similar way by induction and using the previous results.

4. The equality $Cov(d_k^{(l)}, d_{k+r}^{(j)}) = \langle \psi_{j-l}, \psi(\cdot - 2^{-j}r)\rangle$ is proven also by induction.

□

The following lemma guarantees the existence of the parameter δ_L from Lemma 6.1.

Lemma 6.3 *For each $r \neq 0$, $r \in \mathbb{Z}$, there exist a constant $C_r < 1$ such that all the covariances from Lemma 6.2 have an absolute value smaller than C_r.*

Proof: For $r \neq 0$ the function $\phi(x)\cdot\phi(x-2^{-j}r)$ has smaller support than the individual functions ϕ and $\phi(\cdot - 2^{-j}r)$. In particular

$$
\begin{aligned}
\int \phi(x)\phi(x-2^{-j}r)dx &= \int \phi(x)\chi_{[0,2D-1]}\phi(x-2^{-j}r)\chi_{[2^{-j}r,2D-1+2^{-j}r]}dx \\
&= \int \phi(x)\phi(x-2^{-j}r)\chi_{[\max(0,2^{-j}r),\min(2D-1,2D-1+2^{-j}r)]}dx.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
|\int \phi(x)\phi(x-2^{-j}r)dx|^2 &\leq \int \phi^2(x)\chi_{[\max(0,2^{-j}r),\min(2D-1,2D-1+2^{-j}r)]}dx \\
&\quad \cdot \int \phi^2(x-2^{-j}r)\chi_{[\max(0,2^{-j}r),\min(2D-1,2D-1+2^{-j}r)]}dx \\
&= \int \phi^2(x)\chi_{[\max(0,2^{-j}r),\min(2D-1,2D-1+2^{-j}r)]}dx \cdot \\
&\quad \cdot \int \phi^2(x-2^{-j}r)\chi_{[\max(-2^{-j}r,0),\min(2D-1-2^{-j}r,2D-1)]}dx \\
&= C_r < 1.
\end{aligned}
$$

The proofs for the other correlation functions follow in a similar way and are left to the reader. □

At this point we want to mention that there are not many results on the structure of the correlation functions between scaling and wavelet functions. In [17], the number of zeros of those functions is studied. But to our knowledge, no general sharp estimations on upper bounds of the correlation functions are known.

Remark: From the results of the two previous lemmata it follows that the threshold in Eq.(24) is not only level dependent, but also wavelet dependent. The threshold for the orthogonal situation is, neglecting the boundary corrections, neither level nor wavelet dependent.

6.2 A generalization of the Donoho theorem to the TIDWT

In this section we combine all the results from the previous sections to formulate and prove the analogue result to Theorem 6.2 for the TIDWT. The theorem covers the smoothness properties of an estimated denoised function after thresholding the redundant wavelet coefficients.

Theorem 6.4 *Let* $\mathbf{g}^* = \{g_i^*\}$ *be the sequence of estimated function values produced by soft-thresholding the redundant wavelet coefficients* $M_L(\mathbf{d})$ *with the threshold* $t_{(L+1)N} = \sigma\sqrt{2(1+\delta_L)\log((L+1)N)}$ *and suppose* $N = 2^{j_0}$. *Then there exists a smooth interpolation* g^* *of the sequence* $\{g_i^*\}$ *on* $[0,1]$ *which is, with probability tending to* 1, *at least as smooth as* f. *In particular,*

$$P\{\| g^* \|_{\mathcal{F}} \leq C_2 \| f \|_{\mathcal{F}} \; \forall \mathcal{F} \in \mathcal{S}\} \quad \to \quad 1 \quad \text{for} \quad N \to \infty \tag{27}$$

where the constant $C_2 = C(\phi, p)$ *depends only on* ϕ *and* p, *but not on* f *or* N. *The class* $\mathcal{S}$ *is defined as the scale of all spaces* $B^{\sigma}_{p,q}$, $F^{\sigma}_{p,q}$ *which embed continuously in* $C[0,1]$, *so that* $1/p < \sigma < \min(Reg(\phi), D-1)$.

If

$$\eta_{t_{LN}} \circ R_j(\mathbf{d}) = \mathbf{0} \quad \textit{for} \quad j = 1, \dots L, \tag{28}$$

where R_j is the matrix that yields the redundant wavelet coefficients as scale j from Eq.(6), then

$$P\{\| g^* \|_{\mathcal{F}} \leq C_L \| f \|_{\mathcal{F}} \, \forall \mathcal{F} \in \mathcal{S}\} \quad \rightarrow \quad 1 \quad \textit{for} \quad N \rightarrow \infty. \tag{29}$$

where $\mathcal{S}$ is the scale of all spaces $B^\sigma_{p,q}$, $F^\sigma_{p,q}$ which embed continuously in $C[0,1]$, so that $1/p < \sigma < \min(Reg(\Phi), 2D-1)$. The constant $C_L = C(\phi, p, L)$ depends on ϕ, p and L, but not on f or N.

Proof: Combining Theorem 6.3 with Lemma 6.1, 6.2 and 6.2 we know that

$$P\{\| M_L(\mathbf{d}) - M_L(\mathbf{f}) \|_{l^\infty} \leq \sigma \sqrt{2(1+\delta_L)\log((L+1)N)}\} \quad \rightarrow \quad 1 \quad \text{for} \quad N \rightarrow \infty. \tag{30}$$

Because the TIDWT consist of all shifts of the orthogonal DWT,the result in Eq.(27) follows from the same arguments used in the proof of Theorem 6.2.
Now we consider the special case that $\eta_{t_{(L+1)N}} \circ R_j(\mathbf{d}) = \mathbf{0}$ for $j \leq L$. Then

$$\mathbf{g}^* = M_L^\dagger \circ \eta_{t_{(L+1)N}} M_L(\mathbf{d}) = \frac{1}{2^L} S_L^T \circ \eta_{t_{(L+1)N}} \circ S_L(\mathbf{d}).$$

Furthermore, we define

$$\begin{aligned} \lambda^T &= (\lambda_1, \dots \lambda_N)^T = \eta_{t_{(L+1)N}} \circ S_L(\mathbf{d}) \quad \text{and} \\ \mu^T &= (\mu_1, \dots \mu_N)^T = S_L(\mathbf{f}). \end{aligned} \tag{31}$$

With Eq.(30) it follows that λ is element-wise smaller than μ. There exists a matrix $S_L^\dagger = (S_L^T S_L)^{-1} S_L^T$ such that

$$S_L(\mathbf{g}) = \lambda \quad \text{with} \quad \mathbf{g} = S^\dagger(\lambda).$$

$S_L^\dagger$ is the pseudo inverse of S_L. That means that

$$\mathbf{g}^* = \frac{1}{2^L} S_L^T S_L(\mathbf{g}) = T_L(\mathbf{g}) \tag{32}$$

is the discrete translation invariant projection from Eq.(11). In the classical orthogonal situation we could obtain only discrete orthogonal projections. Now we show that this “jump” to the discrete translation invariant projections influences the smoothness of the interpolating function g^*. We derive the following estimations for $\| \mathbf{g} \|_{l^p}$:

$$\begin{aligned} \| \mathbf{g} \|_{l^p} &\leq C_1(L, p, \phi)) \| \lambda \|_{l^p} \\ &\leq C_1(L, p, \phi) \| \mu \|_{l^p} \\ &\leq C_2(L, p, \phi) \| \mathbf{f} \|_{l^p} \quad \text{with Eq.(31)} \end{aligned} \tag{33}$$

and

$$\| \mathbf{g}^* \|_{l^p} \leq C_3(L,p,\phi) \| \mathbf{g} \|_{l^p} \leq C_4(L,p,\phi) \| \mathbf{f} \|_{l^p} \quad \text{with Eq.(32).} \tag{34}$$

With Theorem 5.1 we know that there exist interpolating wavelet transforms of the data $\mathbf{g}, \mathbf{g}^*$, namely

$$g = \sum_k \beta_{j_0,k} \Phi_{j_0,k} \quad \text{and } g^* = \sum_k \beta^*_{j_0,k} \Phi_{j_0,k}.$$

with $\beta_{j_0,k} = g_k$ and $\beta^*_{j_0,k} = g^*_k$. The coefficients $\beta_{j_0,k}$ and $\beta^*_{j_0,k}$ are linked by $\beta^*_{j_0,k} = T_L((\beta_{j_0,.}))$ From Eq.(9) it follows that

$$t_{j,k}(\beta_{j_0,.}) = \sum_{m_1,\ldots,m_L} a_{m_1} \cdots a_{m_L} \beta_{j_0,k+m_1+\cdots+2^L m_L}.$$

Therefore, in the interior of the interval $[0,1]$ we get

$$\begin{aligned}
g^*(x) &= 2^{j_0/2} \sum_k \sum_{m_1,\ldots,m_L} a_{m_1} \cdots a_{m_L} \beta_{j_0,k+m_1+\cdots+2^L m_L} \Phi(2^{j_0}x - k) \\
&= 2^{j_0/2} \sum_t \sum_{m_1,\ldots,m_L} a_{m_1} \cdots a_{m_L} \beta_{j_0,t} \Phi(2^{j_0}x + m_1 + \cdots + 2^L m_L - t) \\
&= 2^{j_0/2} \sum_t \beta_{j_0,t} \sum_{m_2,\ldots,m_L} a_{m_2} \cdots a_{m_L} \sum_{m_1} a_{m_1} \Phi(2^{j_0}x + m_1 + \cdots + 2^L m_L - t) \\
&= 2^{j_0/2} \sum_t \beta_{j_0,t} \sum_{m_2,\ldots,m_L} a_{m_2} \cdots a_{m_L} \Phi(-2^{j_0-1}x - m_2 - \cdots - 2^{L-1} m_L + \frac{t}{2}) \\
&= \ldots \\
&= 2^{j_0/2} \sum_t \beta_{j_0,t} \Phi(-2^{j_0-L}x + 2^{-L}t) \\
&= 2^{j_0/2} \sum_t \beta_{j_0,t} \Phi(2^{j_0-L}x - 2^{-L}t)
\end{aligned} \tag{35}$$

Now we define the vectors of scaling and wavelet coefficients of the interpolating wavelet transforms of the functions f, g, g^* as $\theta_f = ((\gamma_{j_0,.}), (\delta_{j_0,.}), (\delta_{j_0+1,.}), \ldots)$, $\theta_g = ((\beta_{j_0,.}), \mathbf{0}, \mathbf{0}, \ldots)$ and $\theta^*_g = ((\beta^*_{j_0,.}), \mathbf{0}, \mathbf{0}, \ldots)$. With the same arguments as the equivalence $\| \theta_g \|_{b^\sigma_{p,q}} \asymp \| g \|_{B^\sigma_{p,q}}$ and $\| \theta_f \|_{b^\sigma_{p,q}} \asymp \| f \|_{B^\sigma_{p,q}}$ is shown in [11], the equivalence

$$\| \theta^*_g \|_{b^\sigma_{p,q}} \asymp \| g^* \|_{B^\sigma_{p,q}} \tag{36}$$

follows. From (34) we obtain

$$\| \theta^*_g \|_{b^\sigma_{p,q}} \leq C_5 \| \theta_g \|_{b^\sigma_{p,q}} \leq C_6 \| \theta_f \|_{b^\sigma_{p,q}} .$$

With the norm equivalence in Eq.(36) the desired result

$$\| g^* \|_{B^\sigma_{p,q}} \leq C_7 \| f \|_{B^\sigma_{p,q}}$$

follows where the constant C_7 depends on ϕ, p and L. In a similar way the result for Triebel spaces is proven. □

We want to give a heuristic interpretation of the above theorem. Aside from the thresholds t_N and $t_{(L+1)N}$, the difference between the orthogonal and the translation invariant denoising techniques is contained in the "jump" possibility in the latter case. If we assume that the result of soft-tresholding of the orthogonal wavelet coefficients is a discrete orthogonal projection as in Figure 1 and apply Theorem 6.2, the smoothness of the estimated function always depends on the function ϕ. Therefore, the range of the smoothness parameter σ is bounded by $\min(Reg(\phi), D-1)$. The averaging process in the translation invariant case makes it possible to extend the range of the smoothness parameter to $\max(Reg(\Phi), 2D-1)$. Going back to Theorem 6.3 we know that we threshold the wavelet coefficients "a bit more than necessary" when applying the threshold $t_{(L+1)N} = \sqrt{2(1+\delta_L)\log((L+1)N)}$ because the threshold $\sqrt{2\log((L+1)N)}$ would already be sufficient for the case of $(L+1)N$ independent random variables. Usually, the wavelet coefficients of a function f are small in smooth and large in singular parts of f. Therefore, it is likely that thresholding the wavelet coefficients with $t_{(L+1)N}$ results in the situation from Eq.(28) for wavelet coefficients coming from subintervals of $[0,1]$, where f is smooth. That means we make the transition to the interpolation by the smoother autocorrelation functions. In singular parts of f the transition becomes more unlikely.

7 Numerical experiments

In this section we show in some numerical examples the results of translation invariant denoising via soft-thresholding with the level-dependent threshold $t_{(L+1)N}$. In the applications we do not threshold the scaling coefficients, but only the wavelet coefficients. The parameter δ_l is computed for each level $l \leq L$. Therefore the chosen threshold in the soft-threshold operator is t_{lN} for the wavelet coefficients at scale l. The experiments were done using the `Rice Wavelet Toolbox for Matlab`*.

Figure 2 shows the *Doppler* example created with Donoho's software package `WaveLab` for the sampling rate $N = 512$. The added noise has the standard deviation 1/10. The denoising results for different thresholds, namely the two fixed thresholds $\sqrt{2\log N}$, $\sqrt{2\log(LN)}$ and the level dependent threshold $\sqrt{2(1+\delta_l)\log(lN)}(l = 1,\ldots,L)$, show that the new level dependent threshold is necessary to remove all the noise in the translation invariant denoising model. The lower threshold still keep some of the noise in the thresholded wavelet coefficients. The analysis

*Available from http://www-dsp.rice.edu/software/

was also done for the various standard test signals in the WaveLab package, and we have chosen the *Doppler* example to explain the results from the previous section.

Figure 3 shows the elements of $M_L(\mathbf{f}), M_L(\mathbf{d})$ and the thresholded coefficients $\eta_{t_{lN}} \circ M_L(\mathbf{d})$. In the last case we see that, almost in the middle of the interval [0,1], there exist a left and right bound b_l, b_r such that the thresholded wavelet coefficients $\eta_{t_{lN}} \circ d_k^{(l)}$ are equal to zero for all k with $b_l \leq k \leq b_r$ and all $l \leq 5$. If we choose the maximal level of decomposition $L = 5$ we result in the situation described in Eq.(28) of Theorem 6.4. That means there exists a function that interpolates the sequence given by $M_L^\dagger \circ \eta_{t_{lN}} \circ M_L(\mathbf{d})$ and is in the interval $[b_l/N, b_r/N]$ a linear combination of the dilated and translated autocorrelation function Φ. Figure 4 shows the denoising result for $L = 5$ via soft-thresholding with the TIDWT and the orthogonal DWT. The increasing smoothness in the middle part of the interval compared to the corresponding orthogonal situation is visible.

8 Concluding remarks

The reader will have noticed that we did not study the the second denoising result in [13, 14] which is a statement on the optimality of the soft-thresholding technique in the field of minimax/risk estimators, in the translation invariant setting. In [18, 23], an analysis of the risk estimation for the TIDWT is done not for soft-thresholding, but for hard-thresholding. The soft-thresholding case for the TIDWT will be part of our future work. It is demonstrated in many examples that hard-thresholding of the coefficients of the TIDWT yields better denoising results in terms of mean-square-error and signal-to-noise-ratio as soft-thresholding ([7, 23]). We have not studied hard-thresholding in this paper, because we wanted to focus on the characterization of the smoothness of the estimated function. This can not be done with the technique of wavelet shrinkage presented in this paper when replacing soft- by hard-thresholding.

Acknowledgments. We thank Jan Odegard for many discussions on the translation invariant discrete wavelet transform and the members of the Computational Mathematics Laboratory for inspiring ideas on various topics in wavelet theory.

References

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

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

[3] 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.

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

[5] Cohen, A., Daubechies, I., Jawerth, B., Vial, P. Multiresolution analysis, wavelets, and fast algorithms on an interval. *Compt. Rend. Acad. Sci. Paris (A)*, 316:417–421, 1992.

[6] Coifman, R., Wickerhauser, M. Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory*, 38:713–718, 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 Transactions on Signal Processing*, 1998. To appear.

[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. *Appl. Computat. Harmonic Anal.*, 1994.

[12] Donoho, D.L. Smooth wavelet decompositions with blocky coefficient kernels. In Schumaker, L.L., Webb, G., editor, *Recent Advances in Wavelet Analysis.* Academic Press, Inc., 1994.

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

[14] Donoho, D.L., Johnstone, I.M. Ideal spacial adaption by wavelet shrinkage. Technical report, Department of Statistics, Stanford University, 1992.

[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] Frazier, M., Jawerth, B., Weiss, G. *Littlewood-Paley Theory and the Study of Function Spaces*, volume 79 of *NSF-CBMS Regional Conf. Series in Mathematics.* American. Math. Society, 1991.

[17] Gopinath, R.A., Odegard, J.E., Burrus, C.S. On the Correlation Structure of Multiplicity M Scaling Functions and Wavelets. *Proc. ISCAS-92*, 1992.

[18] 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.

[19] Holschneider, M. *Wavelets - An Analysis Tool.* Oxford University Press, 1995.

[20] 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.

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

[22] Lang, M., Guo, H., Odegard, J.E., Burrus, C.S., Wells Jr, R.O. Nonlinear processing of a shift invariant dwt for noise reduction. In *Proceedings of SPIE, Olando 1995*, 1995.

[23] 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.

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

[25] Mallat, S. Multiresolution approximation and wavelets. *Trans. Amer. Math. Soc.*, 315:69–88, 1989.

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

[27] Meyer, Y. *Wavelets and Operators.* Cambridge University Press, 1992.

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

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

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

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

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

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

[34] Sweldens, W. The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets. *Applied and Computational Harmonic Analysis*, 3:186–200, 1996.

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

[36] Triebel, H. *Theory of function spaces*, volume 78 of *Monographs in Math.* Birkhäuser Verlag, Basel, 1983.

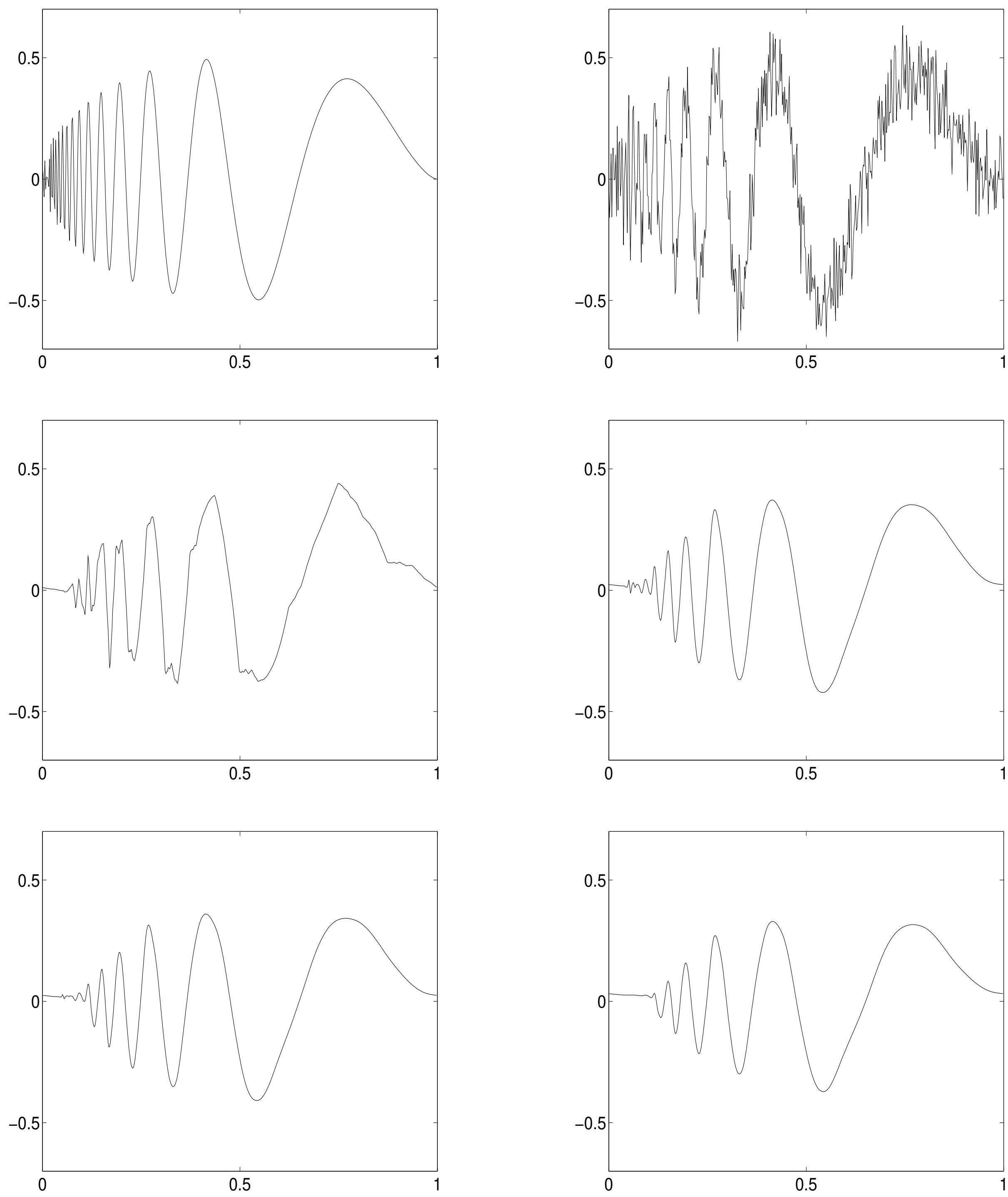

Figure 2: The original signal *Doppler* from the WaveLab package for N=512 (*top,left*), noisy signal (*top,right*), orthogonal denoised with $t = \sqrt{2 \log N}$ (*middle,left*), translation invariant denoised $t = \sqrt{2 \log N}$ (*middle,right*), translation invariant denoised $t = \sqrt{2 \log(LN)}$ (*bottom,left*), translation invariant denoised $t = \sqrt{2(1+\delta_l) \log(lN)}$ (*bottom,right*). The chosen wavelet is Daubechies-3 (D=3) and $L = 9$ is the maximal level of decomposition.

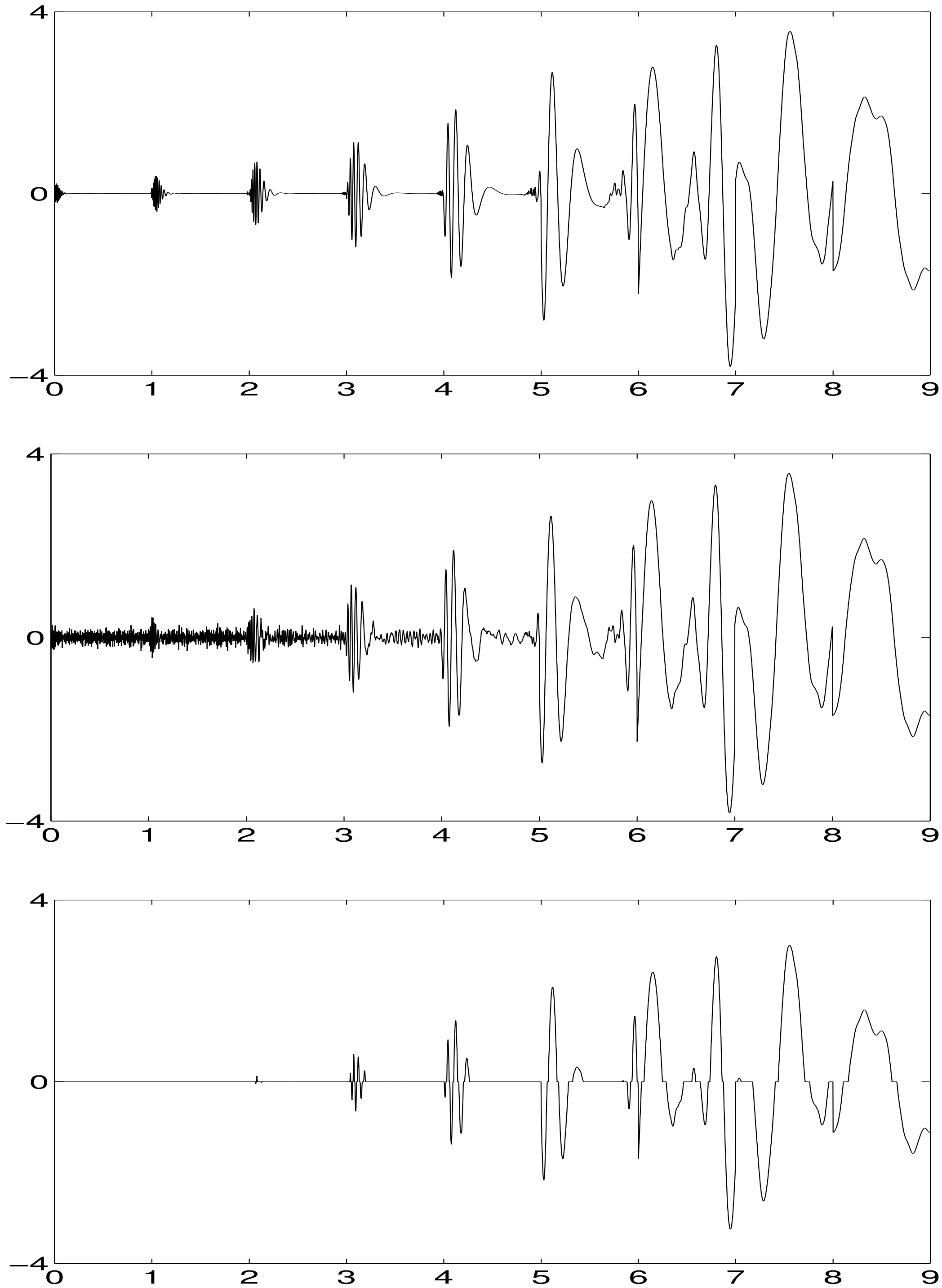

Figure 3: Wavelet coefficients of the experiments from Figure 2 for 9 levels of decomposition. Coefficients of the original signal (*top*), noisy signal (*middle,*), thresholded with $t = \sqrt{2(1+\delta_l)\log(lN)}$ (*bottom*).

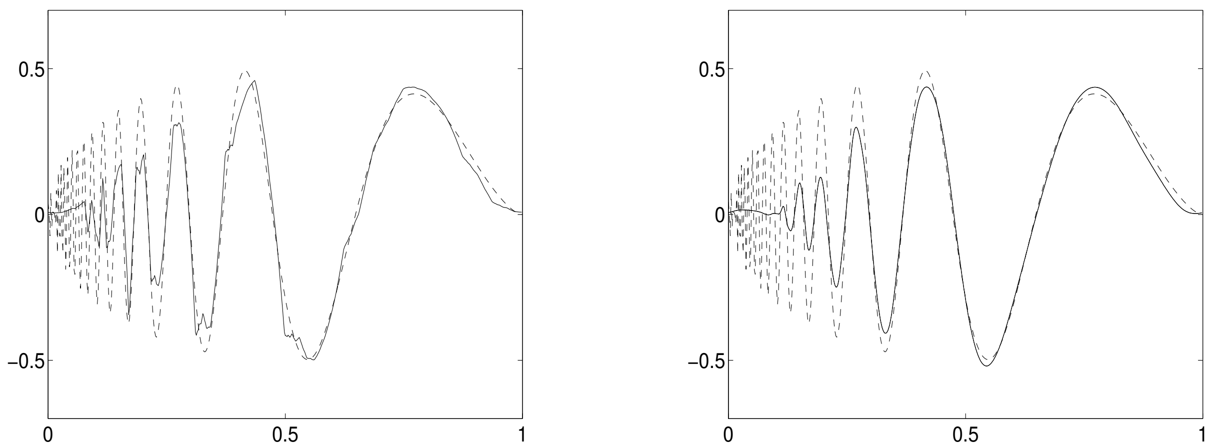

Figure 4: Original signal (dotted) and signal after denoising applying a wavelet transforms with 5 levels of decomposition (solid): DWT with $t = \sqrt{2 \log N}$ (*left*), TIDWT with $t = \sqrt{2(1+\delta_l)\log(lN)}$, $(l = 1, \ldots, 5)$ (*right*).