Applied and Computational Harmonic Analysis **12**, 1–24 (2002)
doi:10.1006/acha.2001.0366, available online at http://www.idealibrary.com on IDEAL®

Smoothness Estimates for Soft-Threshold Denoising via Translation-Invariant Wavelet Transforms

Kathrin Berkner

Ricoh Innovations, Inc., 2882 Sand Hill Road, Suite 115, Menlo Park, California 94025
E-mail: berkner@rii.ricoh.com

and

Raymond O. Wells, Jr. [1]

International University Bremen, Campus Ring 1, 28759 Bremen, Germany

Communicated by Iain M. Johnstone

Received March 24, 1998; revised April 19, 2001

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.

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.

[1] This work was performed while the authors were with the Computational Mathematics Laboratory at Rice University, Houston, TX, partly supported by Texas ATP, DARPA/AFOSR F49620-97-1-0513, and Alexander von Humboldt-Stiftung.

1063-5203/02 $35.00

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 [27] 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 arose: 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 have 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 and related redundant discrete and continuous wavelet transforms and introduce the necessary definitions. A relation between discrete wavelet coefficients of sample values and theoretical wavelet coefficients is given by Donoho in [11]. In Section 3, 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 4 contains the stochastic 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 5 demonstrates the presented results in some numerical experiments. Finally, concluding remarks follow in Section 6.

2. ORTHOGONAL, TRANSLATION-INVARIANT, AND AUTOCORRELATION WAVELET TRANSFORMS

In the first part of this section, we briefly introduce the notations related to those wavelet systems that are used throughout the rest of this paper. We focus on the compactly supported Daubechies wavelet systems [9]. Most of the following results also apply to biorthogonal wavelet systems. However, different vanishing moments of analysis and synthesis wavelet systems or the different frame bounds can enter into some estimates, e.g., the smoothness parameter range estimation in Section 3 and the thresholding selection in Section 4. This is left to the reader.

The scaling function ϕ and wavelet function ψ of an orthonormal Daubechies wavelet system satisfy the scaling equations

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

with scaling filter $\mathbf{h} = \{h_k : k \in \mathbf{Z}\}$, wavelet filter $\mathbf{g} = \{g_k : k \in \mathbf{Z}\}$, and $g_k = (-1)^k h_{1-k}$. The scaling function ϕ is compactly supported in $[0, 2D-1]$ and of degree $D-1$. Using the common notations $\phi_{j,k}(x) = 2^{j/2}\phi(2^j x - k)$ and $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$, a function $f \in L^2(\mathbf{R})$ can be expanded into a *wavelet series* by

$$f = \sum_{j,k} \langle \psi_{j,k}, f\rangle \psi_{j,k} \qquad \text{or} \qquad f = P^{j_0} f + \sum_{j \ge j_0} Q^j f, \tag{1}$$

where $P^j f = \sum_k \langle \phi_{j,k}, f\rangle \phi_{j,k}$ and $Q^j f = \sum_k \langle \psi_{j,k}, f\rangle \psi_{j,k}$ denote the projections onto the subspaces spanned by the family of scaling functions $\{\phi_{j,k}\}$ and their orthogonal complements. For further reading we refer to [9]. Throughout the rest of this paper, for wavelet series, we choose the convention: the finer the scales (higher resolution), the larger the index j.

It follows from the definition of the projection that $P^j f$ is at least as smooth as the scaling function ϕ. 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., [28]).

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 $C^n(\mathbf{R})$, $1 \le n \le D$, and has compact support. Then there exists a constant $C > 0$ such that*

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

If f is in $C^{n+\alpha}(\mathbf{R})$ for $0 < \alpha < 1$ and $0 \le n \le D-1$, we obtain the approximation order $O(2^{-j(n+\alpha)})$.

One disadvantage of this classical orthogonal wavelet transform is that it is not translation invariant. To eliminate this disadvantage, we consider all shifted versions of the orthogonal transform and average over all those shifts. The result is a transform that is not orthogonal and, moreover, contains redundant information. In the discrete situation, this idea was introduced independently in [2, 20, 25, 30]. 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 complete Daubechies family and embed it into the *(continuous) autocorrelation transform*. Following [7], we denote the shift of size h by S_h. The redundant projections $P_R^j[f]$ and $Q_R^j[f]$ are defined by

$$P_R^j[f] = 2^j \int_0^{2^{-j}} S_{-h} P^j S_h[f]\, dh, \qquad Q_R^j[f] = 2^j \int_0^{2^{-j}} S_{-h} Q^j S_h[f]\, dh.$$

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

LEMMA 2.1. *Let ϕ, ψ be the Daubechies scaling and wavelet function of degree $D-1$. Then*

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

where

$$\Phi(s) = \int_{-\infty}^{+\infty} \phi(x)\phi(x+s)\,dx \qquad \textit{and} \qquad \Psi(s) = \int_{-\infty}^{+\infty} \psi(x)\psi(x+s)\,dx$$

are the autocorrelation functions of ϕ and ψ.

The proof is a straightforward calculation using the definitions of $P_R^j[f]$ and $Q_R^j[f]$ and is left to the reader. In the terminology of [19], the projections Q_R^j are called *single voices* of the continuous wavelet transform.

LEMMA 2.2. (1) *The projections $P_R^j[f]$ and $Q_R^j[f]$ are translation invariant, i.e., for $\delta \in \mathbf{R}$, we have*

$$P_R^j[f(\cdot+\delta)](y) = P_R^j[f](y+\delta), \qquad Q_R^j[f(\cdot+\delta)](y) = Q_R^j[f](y+\delta).$$

(2) *A function $f \in L^2(\mathbf{R})$ has the representation*

$$f = P_R^{j_0}[f] + \sum_{j \geq j_0} Q_R^j[f]. \tag{2}$$

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

These results follow straightforwardly from the definitions of P_R^j and Q_R^j.

Since the maximal decimated discrete orthogonal wavelet transform (DWT) is commonly called the discrete version of the continuous orthogonal wavelet transform in (1), the discrete transform introduced in [29] can be considered to be the discrete version of the autocorrelation transform in (2). In the context of wavelet theory, the functions Φ and Ψ were first studied by Beylkin in [1] and their applications were studied in [29]. They are also known as *Deslauriers–Dubuc fundamental functions* [10]. In [11], they are considered as special *interpolating scaling functions.*

2.1. 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.*

$$\Phi(x) = \sum_{k\in\mathbf{Z}} a_k \Phi(2x-k), \qquad \Psi(x) = \sum_{k\in\mathbf{Z}} (-1)^k a_k \Phi(2x-k),$$

where $a_k = \sum h_l h_{l+k}$ are the *autocorrelation coefficients* of $\mathbf{h}$. These coefficients further satisfy $\sum a_{2k} = \delta_{0,k}$ and $\sum a_k = 2$. They are also known as "a-trous" filters [20].

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

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

(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 $\hat{\Phi} = |\hat{\phi}|^2$. Using the relations between Sobolev and Hölder spaces, a similar statement can be made for the Hölder regularity of Φ for a wavelet system of larger degree than the Haar system.

(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 [34]. The correlation functions of a pair of biorthogonal scaling functions and their relations to the wavelet used in Donoho's interpolating wavelet transform (Section 3) are presented in [33] in the context of the lifting scheme.

Based on the number of vanishing moments $2D$ of the wavelet function Ψ, we obtain 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 state the generalization of that result for $D > 1$. The proof proceeds along the same lines as that of Theorem 2.1.

THEOREM 2.2. *Suppose that f is in $C^n(\mathbf{R})$, $1 \le n \le 2D$, and has compact support. Let P_R^j be the redundant projection based on the Daubechies scaling function of order D. Then there exists $C > 0$ such that*

$$\|f - P_R^j[f]\|_{L_2} \le C2^{-jn}.$$

If f is in $C^{n+\alpha}(\mathbf{R})$ for $0 < \alpha < 1$ and $0 \le n \le 2D - 1$, we obtain the approximation order $O(2^{-j(n+\alpha)})$. A comparison with the result for the projections P^j of the classical nonredundant WT in Theorem 2.1 shows that introduction of redundancy increases the approximation order for Hölder continuous functions of order β for $n < \beta < 2D$. If, e.g., $f \in C^{D+\alpha}(\mathbf{R})$ for $0 < \alpha < 1$, the nonredundant projection has the approximation order $O(2^{-jD})$ whereas the redundant projection has approximation order $O(2^{-j(D+\alpha)})$. For the Haar system, this observation was already mentioned in [7].

In the following, we choose a matrix representation for the discrete versions of the continuous transform. For the orthogonal case, the discrete transform, known as the maximal decimated orthogonal transform, of a vector $\mathbf{x} \in \mathbf{R^N}$ is represented by an orthogonal $N \times N$ matrix W (see, e.g., [5, 32]). A discrete version of the continuous translation-invariant WT (2) is the translation-invariant DWT (TIDWT) which relinquishes down-sampling at each scale ℓ.[2] The scaling coefficients at scale ℓ are given by the vector

$$\mathbf{c}^{(\ell)} = (c_1^{(\ell)}, \ldots, c_N^{(\ell)}), \qquad \text{with } \mathbf{c}_k^{(\ell)} = \sum_{m \in \mathbf{Z}} h_m c_{k+2^{\ell-1}m}^{(\ell-1)}, \tag{3}$$

and the wavelet coefficients are given by

$$\mathbf{d}^{(\ell)} = (d_1^{(\ell)}, \ldots, d_N^{(\ell)}), \qquad \text{with } \mathbf{d}_k^{(\ell)} = \sum_{m \in \mathbf{Z}} g_m c_{k+2^{\ell-1}m}^{(\ell-1)}, \tag{4}$$

where $\ell \leq \log_2 N$.

This TIDWT reflects 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 general discussion on the TIDWT and its applications to denoising, we refer to [4, 7, 18, 23]. In a matrix representation, the TIDWT can be defined as follows [18, 23].

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

$$M_L = (R_1^T, \ldots, R_L^T, S_L^T)^T. \tag{5}$$

An inverse transform is represented by the pseudoinverse matrix

$$M_L^\dagger = (M_L^T M_L)^{-1} M_L^T. \tag{6}$$

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

$$M_L^\dagger = \left(\frac{1}{2} R_1^T, \ldots, \frac{1}{2^L} R_L^T, \frac{1}{2^L} S_L^T \right). \tag{7}$$

There exists a relation to the *discrete autocorrelation transform* introduced, e.g., in [29]. This transform in our context is represented by the matrix

$$A_L := \left(\frac{1}{2} R_1 R_1^T, \ldots, \frac{1}{2^L} R_L R_L^T, \frac{1}{2^L} S_L S_L^T \right)^T. \tag{8}$$

We call $(t_{L,k}) = T_L := (1/2^L) S_L^T S_L$ the *discrete autocorrelation projection at scale L* and recall the following equality:

$$T_L(\mathbf{x}) = M_L^\dagger \circ (\mathbf{0}, \ldots, \mathbf{0}, S_L^T)^T(\mathbf{x}). \tag{9}$$

Setting complete scales of the TIDWT to zero can therefore be interpreted as switching of wavelet basis functions, since we obtain a representation of the approximized sequence as

[2] In contrast to the continuous wavelet transforms, in the literature on discrete wavelet transforms coarser scales are usually denoted by increasing indices. Therefore, we choose the index notation ℓ instead of j for scales of discrete wavelet transforms.

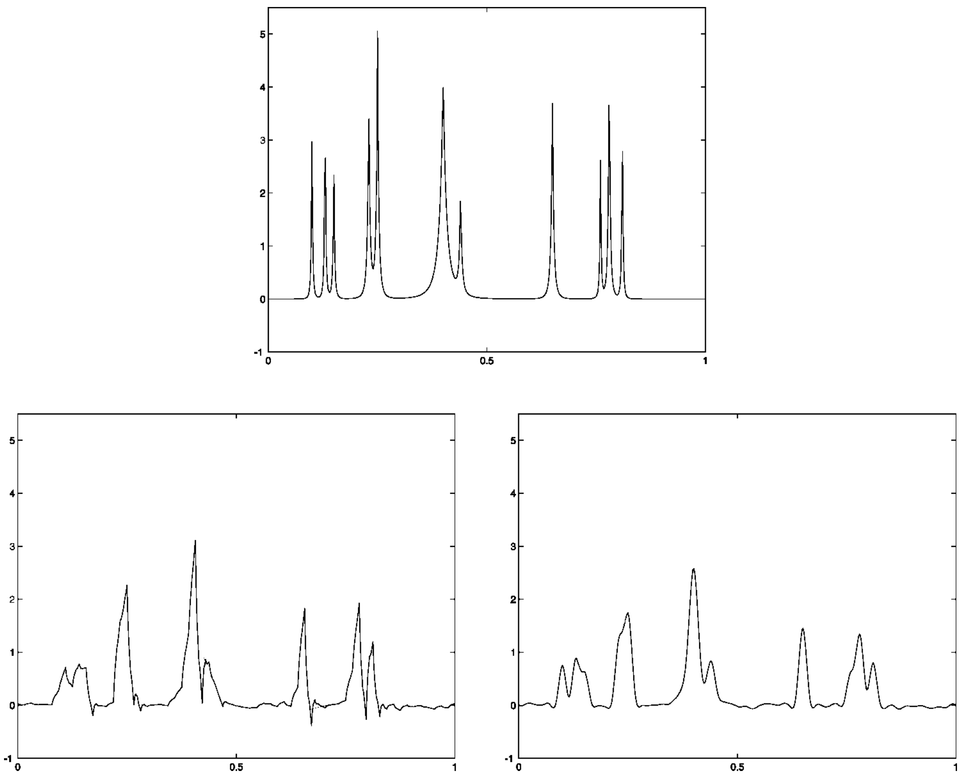

FIG. 1. The original signal *Bumps* from the WaveLab package for $N = 512$ (*top*), orthogonal discrete projection for $L = 3$ (*left*), discrete autocorrelation projection T_L for $L = 3$ (*right*) for the Daubechies-4 wavelet ($D = 2$).

the output of a filtering algorithm using not the Daubechies coefficients h_k, but the a-trous filter coefficients a_k. Figure 1 shows the discrete projections for the nonredundant and the redundant case using the Daubechies wavelet system of order $D = 2$.

Summarizing this section, we have two related transforms in the continuous setting: the orthogonal and the autocorrelation transform. In the discrete setting, we have three transforms: the orthogonal, the translation-invariant, and the autocorrelation transform. The last two are both redundant versions of the orthogonal DWT, but have different coefficients.

3. 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, 26, 35]. These function spaces contain Hölder (–Zygmund) and Sobolev spaces. Given the representation of f in scaling and wavelet functions from (1), we define the vector

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

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

$$\delta_{j,k} = \langle f, \psi_{j,k}\rangle \qquad \text{for } j \geq j_0. \tag{10}$$

It is shown in [16, 26] that the norm of the sequence

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

where $s = \sigma + 1/2 - 1/p$, is equivalent $\|f\|_{B^\sigma_{p,q}}$ $(\|\theta\|_{b^\sigma_{p,q}} \asymp \|f\|_{B^\sigma_{p,q}})$ for $1/p < \sigma < \min(\mathrm{Reg}(\phi),\ D-1)$, $p, q \in (\theta, \infty]$, where $\mathrm{Reg}(\phi)$ measures the regularity of ϕ as its number of continuous derivatives. Similarly, the norm

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

where $s = \sigma + 1/2$ and $\chi_{j,k}$ is the characteristic function of the interval $[k/2^j, (k+1)/2^j]$, is equivalent to $\|f\|_{F^\sigma_{p,q}}$ for $1/p < \sigma < \min(\mathrm{Reg}(\phi),\ D-1)$, $p, q \in (0, \infty]$. For further details on Besov and Triebel spaces and their relations to wavelet coefficients, we refer to [16, 26, 35]. An implication of those equivalences is the well-known fact that whenever we shrink some of the wavelet coefficients to zero, we decrease the norm and, therefore, "smooth" the corresponding function (see, e.g., [13]). In the following, we refer to this implication as *shrinkage* → *smoothing*. It is one important feature that makes wavelet transforms such a successful tool for image compression and denoising.

In applications, we usually compute the discrete wavelet transform of a finite number of function samples of a function $f \in C[0, 1]$. An important question is whether, just from this sequence of discrete wavelet transform coefficients that are not of the form as in (10), we can draw conclusions on the smoothness of the underlying function f and, as a consequence, can derive the same important implication *shrinkage* → *smoothing* as for the theoretical wavelet coefficients.

The goal is to derive a transform that allows a smoothness characterization as in [16, 26] and has given samples of a function as scaling coefficients at a fine scale and not as inner products as in (10). We could use the results from [16, 26] only if we knew that the given samples of a function f^* were indeed inner products of a function f with a scaling function $\phi_{j_1,k}$ for some $j_1 > j_0$. Then the coefficients computed via the pyramid algorithm for scales $j_0 \le j < j_1$ would be exactly the leading coefficients $\gamma_{j_0,\cdot}$ and $\delta_{j_0,\cdot}, \ldots, \delta_{j_1-1,\cdot}$ from (10).

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 transform of continuous functions, the *hybrid wavelet transform.* In the following, we summarize the main steps and theorems from [11] that lead to the construction of that transform in order to provide all the definitions and results necessary in Section 4. First Donoho uses the autocorrelation function Φ in a decomposition of a function and defines *interpolating wavelet transforms* [11, Thm. 2.2].

THEOREM 3.1. *Let j_0 be a nonnegative 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*

$$f = \sum_k \beta_{j_0,k}\tilde\phi_{j_0,k} + \sum_{j\ge j_0}\sum_k \alpha_{j,k}\tilde\psi_{j,k} \tag{11}$$

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

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

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* $\Phi_{j,k} = 2^j\Phi(2^jx-k)$ *as follows. At the interior of the interval,*

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

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

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

as defined in [11]. *The functions* $\psi_{j,k}$ *are constructed as follows. In the interior of the interval* $[0,1]$, *they are defined as*

$$\tilde{\psi}_{j,k} = \Phi_{j+1,\,2k+1}|_{[0,1]}, \qquad D-1 \leq k < 2^j - D + 1, \tag{12}$$

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

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

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* (11) *is called an* interpolating wavelet transform of f.

In this transform, the scaling coefficients $\beta_{j_0,.}$ are samples of the function f, not inner products. The scaling function of the decomposition is, except at the boundaries, the autocorrelation function Φ, and the wavelet functions in (12) are scaled versions of the function Φ. Note that these wavelet functions $\tilde{\psi}_{j,k}$ are different from the autocorrelation wavelet functions $\Psi_{j,k}$. The projection $\sum_k \beta_{j_0,k}\tilde{\phi}_{j_0,k}$ of the interpolating wavelet transform has an approximation property similar to the one in Theorem 2.2 (see [11]). For applications, it is important to notice that interpolating wavelet transforms are not orthogonal wavelet transforms. Similar to the classical orthogonal wavelet transform, there exists a characterization of smoothness in Besov/Triebel spaces [11, Thms. 2.7 and 2.8].

THEOREM 3.2. *Starting with the same assumptions as in Theorem* 3.1, *we define* θ *as*

$$\theta = ((\beta_{j_0,}),\ (\alpha_{j_0,.}),\ (\alpha_{j_0+1,.}),\ \ldots). \tag{13}$$

Then

$$\|\theta\|_{b^{\sigma}_{p,q}} = \|(\beta_{j_{0,\cdot}})\|_{l^p} + \left(\sum_{j\ge j_0}\left(2^{jsq}\left(\sum_{k=0}^{2^j-1}|\alpha_{j,k}|^p\right)^{1/p}\right)^q\right)^{1/q}, \tag{14}$$

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

$$\|\theta\|_{f^{\sigma}_{p,q}} = \|(\beta_{j_{0,\cdot}})\|_{l^p} + \left\|\left(\sum_{j\ge j_0}2^{jsq}\sum_{k=0}^{2^j-1}|\alpha^q_{j,k}|\chi_{j,k}\right)^{1/q}\right\|_{l^p}, \tag{15}$$

where $s=\sigma+1/2$ *and* $\chi_{j,k}$ *is 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* $1/p<\sigma<\min(\mathrm{Reg}(\Phi),\ 2D-1)$, $p,q\in(0,\infty]$.

If we computed the interpolating wavelet coefficients of f, we would be sure that stopping the decomposition at scale j_0 would yield scaling coefficients $\beta_{j_{0,\cdot}}$ that would be samples of f at sampling rate 2^{-j_0}. Then we could apply the argument *shrinkage* → *smoothing* to that decomposition. However, there is still no link to the actual DWT pyramid algorithm that uses the orthogonal wavelet filters **h** and **g**. That link is provided by the construction of a hybrid wavelet transform that is defined in the next theorem [11, Thm 4.2].

THEOREM 3.3. *We consider a function* $f\in C[0,1]$ *and its sample values* $f_{j_1,k}=n^{-j_1/2}f(k/2^{j_1})$ *for fixed integer* j_1 *and define the vector*

$$\tilde\theta = ((\tilde\beta_{j_{0,\cdot}}),\ (\tilde\alpha_{j_{0,\cdot}}),\ \ldots,\ (\tilde\alpha_{j_1-1,\cdot}),\ (\tilde\alpha_{j_{1,\cdot}}),\ \ldots)$$

of coefficients $\tilde\beta_{j_{0,\cdot}},\tilde\alpha_{j,\cdot}\in\mathbf{R}$, *where the coefficients* $(\tilde\beta_{j_{0,\cdot}}),(\tilde\alpha_{j_{0,\cdot}}),(\tilde\alpha_{j_0+1,\cdot}),\ldots,(\tilde\alpha_{j_1-1,\cdot})$ *are those computed via the discrete wavelet transform corresponding to* ϕ *and* ψ *of the samples* $f_{j_1,k}$. *Then, there are smooth functions* $\phi^*_{j_0,k}$, $\psi^*_{j,k}$, $0\le k<2^j$, $j\ge j_0$, *which have at least the same regularity and degree as* ϕ *and have compact support of width* $\le 2^{-j}$, *such that every function* $f\in C[0,1]$ *has an expansion*

$$f=\sum_k\tilde\beta_{j_0,k}\phi^*_{j_0,k}+\sum_{j\ge j_0}\sum_k\tilde\alpha_{j,k}\psi^*_{j,k} \tag{16}$$

with uniform convergence of partial sums.

For $j_0\le j<j_1$, *the functions* $\phi^*_{j_0,k}$, $\psi^*_{j,k}$ *are constructed as special linear combinations of* $\phi_{j,k}$, *under the constraints that* $\phi^*_{j_0,k}$ *and* $\psi^*_{j,k}$ *are orthonormal with respect to the sampling inner product* $\langle f,g\rangle_{j_1}=\sum_{k\in\mathbf{Z}}f(k/2^{j_1})g(k/2^{j_1})$. *For* $j\ge j_1$, *the functions* $\psi^*_{j,k}$ *are the interpolating wavelet functions* $\tilde\psi_{j,k}$ *from Theorem* 3.1 *and* $\tilde\alpha_{j,k}=\alpha_{j,k}$. *The representation in* (16) *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 $f(k/2^{j_1})$. The smoothness of the functions $\phi^*_{j_0,k}$ and $\psi^*_{j,k}$ for $j_0\le j<j_1$ is the same as that of ϕ, whereas for $j\ge j_1$ the smoothness of $\psi^*_{j,k}$ is the same as that of Φ. In general, for $\{\phi^*_{j_0,k},\psi^*_{j,k},\ j\ge j_0\}$, we can guarantee only the smoothness

of ϕ. Again we get a characterization in terms of Besov/Triebel norms [11, Thms. 4.5 and 4.6].

THEOREM 3.4. *Under the assumptions of Theorem* 3.3*, the equivalence*

$$\|f\|_{B^{\sigma}_{p,q}} \asymp \|\tilde{\theta}\|_{b^{\sigma}_{p,q}} \tag{17}$$

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

$$\|f\|_{F^{\sigma}_{p,q}} \asymp \|\tilde{\theta}\|_{f^{\sigma}_{p,q}} \tag{18}$$

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

As a consequence of this result, the implication *shrinkage* $\rightarrow$ *smoothing* is valid, but only for the norm parameter range $1/p < \sigma < \min(\mathrm{Reg}(\phi), \ D-1)$ and not for $1/p < \sigma < \min(\mathrm{Reg}(\Phi), \ 2D-1)$ as in Theorem 3.2. That means using hybrid wavelet transforms estimation of smoothness from wavelet coefficients is restricted to a smaller range of smoothness classes than using interpolating wavelet transforms. The properties of these two transforms will be used in the smoothness characterization of the denoised function in the following section, when we extend the results from [13] for orthogonal wavelet to translation-invariant wavelet transforms.

4. DENOISING BY SHRINKAGE OF REDUNDANT WAVELET COEFFICIENTS

4.1. The Stochastic Model

In this section, we analyze how the Donoho–Johnstone denoising model can be modified and transferred 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

$$d_i = f_i + \epsilon z_i, \tag{19}$$

where $\{z_i\}$ is an i.i.d. sequence of $N(0,1)$ random variables with $\epsilon > 0$ 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 obtain 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

$$W(\mathbf{d}) = W(\mathbf{f}) + \epsilon W(\mathbf{z}), \tag{20}$$

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, 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 distributed random variables, uncorrelatedness is equivalent to independence, the $\tilde{z}_i$ are i.i.d. $N(0, 1)$ random variables. At this point, we 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 reader interested in this topic to [13]. The following theorem from [24] gives the famous general threshold $\sqrt{2 \log N}$ for denoising via orthogonal wavelet transforms.

THEOREM 4.1. *Let $\{z_i\}$ be an i.i.d. sequence of $N(0, 1)$ random variables. Then*

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

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

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

removes a.s. all the noise in the wavelet coefficients as $N \to \infty$. This result, combined with the characterization of smoothness in Besov and Triebel spaces via coefficients of hybrid wavelet transforms from Theorems 3.3 and 3.4, leads to the following theorem [13, Thm. 1.1].

THEOREM 4.2. *Let $\{f_i^*\}_{i=1}^N$ be the sequence of estimated function values produced by soft-thresholding the wavelet coefficients $W(\mathbf{d})$ with the threshold $t_N = \epsilon\sqrt{2 \log N}$. 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,*

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

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 $1/p < \sigma < \min(\operatorname{Reg}(\phi), D)$.

The function $\tilde{f}$ is constructed in such a way that the shrunk coefficients $\eta_{t_N} \circ W(\mathbf{d})$ are the leading coefficients (i.e., $j_0 \le j < j_1$) of the hybrid wavelet transform of $\tilde{f}$ (see, [13, Sect. V.B.]). With Theorem 3.4, the *shrinkage* $\to$ *smoothing* implication is valid and yields the result in the above theorem. 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, Theorem 4.1 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 4.3. *Let $\{\xi_i\}_{i=1}^N$ be an i.d. normal sequence of random variables with $\operatorname{Var}(\xi_i) = 1$ and covariances $r_{ij} = \operatorname{Cov}(\xi_i, \xi_j)$ such that*

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

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

$$|P\{\max_{i=1,\dots,N} \xi_i \le u\} - F(u)^N| \le K \sum_{1\le i<j\le N} |r_{ij}| \exp\left(-\frac{u^2}{1+|r_{ij}|}\right). \tag{21}$$

If $z_1,\dots,z_N$ are i.i.d. $N(0,1)$ random variables, then $P(\max_{i=1,\dots,N} z_i \le u) = F(u)^N$. Therefore, we can interpret the result from Theorem 4.3 as follows: it tells us how much the probability $P\{\max_{i=1,\dots,N} \xi_i \le u\}$ 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-hand side of (21) tends to zero for $N \to \infty$ and such that $u_{(L+1)N} \ge \sqrt{2\log((L+1)N)}$, then we know that $P\{\max_{i=1,\dots,N} \xi_i \le u\} \to 1$ for $N \to \infty$. A possible choice for such a threshold is provided in the next lemma.

LEMMA 4.1. *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 $r_{ij} = \mathrm{Cov}(\xi_i,\xi_j)$. If $\delta = \max_{i\ne j}|r_{ij}| < 1$ and*

$$u = u_{(L+1)N} = \sqrt{2(1+\delta)\log((L+1)N)}, \tag{22}$$

then

$$\sum_{1\le i<j\le (L+1)N} |r_{ij}| \exp\left(-\frac{u^2_{(L+1)N}}{1+|r_{ij}|}\right) \to 0 \qquad \textit{for } N \to \infty. \tag{23}$$

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 $1 \le i < j \le (L+1)N$, we obtain

$$\begin{aligned}
\#\{r_{ij} \mid r_{ij} \ne 0\} &= \#\{\mathrm{Cov}(c_k^{(L)}, c_l^{(L)}) \mid \mathrm{Cov}(c_k^{(L)}, c_l^{(L)}) \ne 0\} \\
&\quad + \#\{\mathrm{Cov}(c_k^{(L)}, d_l^{(m)}) \mid \mathrm{Cov}(c_k^{(L)}, d_l^{(m)}) \ne 0,\ m=1,\dots,L\} \\
&\quad + \#\{\mathrm{Cov}(d_k^{(m)}, d_l^{(n)}) \mid \mathrm{Cov}(d_k^{(m)}, d_l^{(n)}) \ne 0, \\
&\qquad\quad m=1,\dots,L,\ n=1,\dots,m\} \\
&\le DN(2^L-1) + DN2^L\sum_{l=1}^{L}\frac{2^l-1}{2^l} + DN\sum_{m=1}^{L}2^m\sum_{l=1}^{m}\frac{2^l-1}{2^l} \\
&\le DN(2^L - 1 + 2^L L + L(2^{L+1}-2)) \\
&\le DN(2^{L+1} + 2^{L+1}L + 2^{L+1}L) \\
&\le DN2^{L+1}(1+2L).
\end{aligned} \tag{24}$$

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

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

$$\leq \sum_{1\leq i<j\leq (L+1)N} |r_{ij}| \exp(-2\log((L+1)N))$$
$$\leq \frac{1}{((L+1)N)^2} DN2^{L+1}(1+2L) \quad \text{with (23).}$$

For $L < \log_2 N$:

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

For $L = \log_2 N$:

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

for $N \to \infty$. ■

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

$$\sum_{1\leq i<j\leq (L+1)N} |r_{ij}| \exp\left(-\frac{u^2_{(L+1)N}}{1+|r_{ij}|}\right) \leq \sum_{1\leq i<j\leq (L+1)N} |r_{ij}| \exp\left(-\frac{u^2_{(L+1)N}}{1+\delta}\right)$$
$$\leq \left[\frac{1}{((L+1)N)^2}\right]^{1/(1+\delta)} \sum_{1\leq i<j\leq (L+1)N} |r_{ij}|$$
$$\leq \frac{D2^{L+1}(1+2L)}{(L+1)^{2/(1+\delta)} N^{(1-\delta)/(1+\delta)}}.$$

For fixed $L < \log_2 N$, the right-hand side would still go to 0, so the threshold $u_{(L+1)N} = \sqrt{2\log((L+1)N)}$ would work, but for $L = \log_2 N$ we cannot necessarily guarantee that the right-hand term vanishes. Namely,

$$\frac{D2N(1+2\log_2 N)}{(\log_2 N+1)^2 1+\delta N^{(1-\delta)/(1+\delta)}} = \frac{D2(1+2\log_2 N)}{(\log_2 N+1)^2 1+\delta} N^{2\delta/(1+\delta)}$$
$$\leq C\left((\log_2 N)^{(\delta-1)/(1+\delta)} N^{2\delta/(1+\delta)}\right) \to \infty$$
$$\text{for } N \to \infty.$$

In order to determine the threshold $u_{(L+1)N}$ from Lemma 4.1, we need the upper bound δ 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 4.2. *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* (3) *and* (4). *Then the following formulas hold,*

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

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

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

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

(2) The equality $\mathrm{Cov}(d_k^{(j)}, d_{k+r}^{(j)}) = \Psi(2^{-j}r)$ is proven analogously to the previous case.
(3) First we show $\mathrm{Cov}(c_k^{(j)}, d_{k+r}^{(j)}) = \langle \phi, \psi(\cdot - 2^{-j}r)\rangle$ by induction: For $j = 1$, we get

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

On the other hand,

$$\begin{aligned}\langle \phi, \psi(\cdot - 2^{-1}r)\rangle &= \int \phi(x)\psi\left(x - \frac{r}{2}\right) dx \\ &= 2\sum_{n,m} h_n g_m \int \phi(2x-n)\phi(2x-r-m)\, dx \\ &= \sum_{n,m} h_n g_m \int \phi(x-n)\phi(x-r-m)\, dx \\ &= \sum_m h_{m+r} g_m.\end{aligned}$$

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

$$\begin{aligned}\mathrm{Cov}(c_k^{(j)}, d_{k+r}^{(j)}) &= \sum_{n,m} h_n g_m \,\mathrm{Cov}(c_{k+2^{j-1}n}^{(j-1)}, c_{k+r+2^{j-1}m}^{(j-1)}) \\ &= \sum_{n,m} h_n g_m \langle \phi, \phi(\cdot - 2^{-j+1}(r + 2^{j-1}m - 2^{j-1}n))\rangle \\ &= \sum_{n,m} h_n g_m \langle \phi, \phi(\cdot - 2^{-j+1}r - m + n)\rangle.\end{aligned}$$

On the other hand,

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

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

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

As a consequence of this lemma, it follows that the upper bound of the covariances of the random variables $\xi_1, \ldots, \xi_{(L+1)N}$ depends on the maximal decomposition level L. Therefore, we set $\delta = \delta_L$ in Lemma 4.1. An upper bound on δ_L is guaranteed by the following lemma.

LEMMA 4.3. *Given the sequence* $\mathbf{z} = (z_1, \ldots, z_n)$ *of i.i.d.* $N(0,1)$ *random variables, for each* $r \neq 0$, $r \in \mathbf{Z}$, *there exists a constant* $C_r < 1$ *such that all the covariances from Lemma* 4.2 *have an absolute value smaller than* C_r.

Proof. For $r \neq 0$ and w.l.o.g. $r > 0$, 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,

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

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

At this point, we mention that there are very few 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 estimates on upper bounds of the correlation functions are known. For $L = \log_2 N$, the autocorrelations $\Phi(2^{-L}r) \to 1$ and $\Psi(2^{-L}r) \to 1$ for $N \to \infty$. Table 1 shows the cross-correlation bound for various levels and wavelet systems.

Remark. From the results of the two previous lemmas, it follows that the threshold in (22) 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.

TABLE 1
Cross-Correlation Bound for Various Levels L and Daubechies Wavelet Systems of Degree $D-1$ for $D = 1, 2, 3$

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

4.2. A Generalization of the Donoho Theorem to the TIDWT

In this section, we combine all above results to formulate and prove the analog result to Theorem 4.2 for the TIDWT.

If we computed the interpolating wavelet coefficients starting from samples of f, we would obtain a smoothness result similar to that in Theorem 4.2, but for an increased range of the parameter σ. However, shrinkage of coefficients of the TIDWT is different from shrinkage of interpolating wavelet transform coefficients and also different from shrinkage of the coefficients of the discrete autocorrelation transform. In other words, we are interested in the reconstruction of a function that is represented by shrunk wavelet coefficients $\eta_t \circ (R_1, \ldots, R_L, S_L)$ of the TIDWT and neither by shrunk coefficients $\eta_t \circ ((\alpha_{j_1+1,.}), \ldots, (\alpha_{j_{0,.}}), (\beta_{j_{0,.}}))$ of an interpolating wavelet transform nor by shrunk coefficients $\eta_t \circ (\frac{1}{2} R_1^T R_1, \ldots, \frac{1}{2^L} R_L^T R_L, \frac{1}{2^L} S_L^T S_L)$ of an autocorrelation transform.

The following theorem covers the smoothness properties of an estimated denoised function after thresholding the redundant wavelet coefficients.

THEOREM 4.4. *Let $\mathbf{g}^* = \{g_i^*\}_{i=1}^N$ be the sequence of estimated function values produced by soft-thresholding of the redundant wavelet coefficients $M_L(\mathbf{d})$ with the threshold $t_{(L+1)N} = \epsilon\sqrt{2(1+\delta_L)\log((L+1)N)}$ and suppose $N = 2^{j_1}$. 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,*

$$P\{\|g^*\|_{\mathcal{F}} \leq C_2 \|f\|_{\mathcal{F}}\ \forall \mathcal{F} \in \mathcal{S}\} \to 1 \qquad \text{for } N \to \infty, \tag{25}$$

where the constant $C_2 = C(\phi, p)$ depends only on ϕ and p, but is independent of f and 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 $1/p < \sigma < \min(\mathrm{Reg}(\phi),\ D-1)$.

If

$$R_\ell^T \circ \eta_{t_{LN}} \circ R_\ell(\mathbf{d}) = \mathbf{0} \qquad \text{for } \ell = 1, \ldots, L, \tag{26}$$

where R_ℓ is the matrix that yields the redundant wavelet coefficients at scale ℓ from Eq. (5), then

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

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 $1/p < \sigma < \min(\mathrm{Reg}(\Phi),\ 2D-1)$. The constant $C_L = C(\phi, p, L)$ depends on ϕ, p, and L, but is independent of f and N.

Proof. Combining Theorem 4.3 with Lemmas 4.1, 4.2, and 4.3, we know that

$$P\{\|M_L(\mathbf{d}) - M_L(\mathbf{f})\|_{l^\infty} \le \epsilon\sqrt{2(1+\delta_L)\log((L+1)N)}\} \to 1 \qquad \text{for } N \to \infty. \tag{28}$$

Because the TIDWT consists of all shifts of the orthogonal DWT, the result in (25) follows from the same arguments used hi the proof of Theorem 4.2.

Now we consider the special case that $R_\ell^T \circ \eta_{t_{LN}} \circ R_\ell(\mathbf{d}) = \mathbf{0}$ for $\ell \le L$. 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

$$\lambda^T = (\lambda_1, \ldots, \lambda_N)^T = \eta_{t_{(L+1)N}} \circ S_L(\mathbf{d}), \qquad \mu^T = (\mu_1, \ldots, \mu_N)^T = S_L(\mathbf{f}). \tag{29}$$

With (28), it follows that the absolute values of the components of λ are smaller than those of μ. There exists a matrix $S_L^\dagger = (S_L^T S_L)^{-1} S_L^T$ such that

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

Here $S_L^\dagger$ is the pseudoinverse of S_L. That means that

$$\mathbf{g}^* = \frac{1}{2^L} S_L^T S_L(\mathbf{g}) = T_L(\mathbf{g}) \tag{30}$$

is the discrete autocorrelation projection from (9). In the classical orthogonal situation, we would obtain only discrete orthogonal projections and the smoothness results from Theorem 4.2. Now we show that the “jump” from orthogonal to autocorrelation projections influences the smoothness of the function g^* constructed to interpolate the samples $\mathbf{g}^*$. We derive from (28) the following estimates for $\|\mathbf{g}\|_{l_P}$,

$$\begin{aligned} \|\mathbf{g}\|_{l^p} &\le C_1(L, p, \phi)\|\lambda\|_{l^p} \\ &\le C_1(L, p, \phi)\|\mu\|_{l^p} \\ &\le C_2(L, p, \phi)\|\mathbf{f}\|_{l^p} \end{aligned} \tag{31}$$

and from (30)

$$\|\mathbf{g}^*\|_{l^p} \le C_3(L, p, \phi)\|\mathbf{g}\|_{l^p} \le C_4(L, p, \phi)\|\mathbf{f}\|_{l^p}. \tag{32}$$

From Theorem 3.1, we know that there exist interpolating wavelet transforms of the data $\mathbf{g}$ and $\mathbf{g}^*$, namely,

$$g = \sum_k \beta_{j_1,k}\Phi_{j_1,k} \qquad \text{and} \qquad g^* = \sum_k \beta^*_{j_1,k}\Phi_{j_1,k}$$

with $\beta_{j_1,k} = g_k$ and $\beta^*_{j_1,k} = g^*_k$. The coefficients $\beta_{j_1,k}$ and $\beta^*_{j_1,k}$ are linked by $\beta^*_{j_1,k} = T_L((\beta_{j_{1,\cdot}}))$. From (9), it follows that

$$t_{j,k}(\beta_{j_{1,\cdot}}) = \sum_{m_1,\ldots,m_L} a_{m_1}\cdots a_{m_L}\beta_{j_1,\, k+m_1+\cdots+2^L m_L}.$$

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

$$
\begin{aligned}
g^*(x) &= 2^{j_1/2}\sum_k \sum_{m_1,\dots,m_L} a_{m_1}\cdots a_{m_L}\beta_{j_1,\, k+m_1+\cdots+2^L m_L}\Phi(2^{j_1}x-k)\\
&= 2^{j_1/2}\sum_t \sum_{m_1,\dots,m_L} a_{m_1}\cdots a_{m_L}\beta_{j_1,t}\Phi(2^{j_1}x+m_1+\cdots+2^L m_L-t)\\
&= 2^{j_1/2}\sum_t \beta_{j_1,t}\sum_{m_2,\dots,m_L} a_{m_2}\cdots a_{m_L}\sum_{m_1} a_{m_1}\Phi(2^{j_1}x+m_1+\cdots+2^L m_L-t)\\
&= 2^{j_1/2}\sum_t \beta_{j_0,t}\sum_{m_2,\dots,m_L} a_{m_2}\cdots a_{m_L}\Phi\left(-2^{j_1-1}x-m_2-\cdots-2^{L-1}m_L+\frac{t}{2}\right)\\
&\;\;\vdots\\
&= 2^{j_1/2}\sum_t \beta_{j_1,t}\Phi(-2^{j_1-L}x+2^{-L}t)\\
&= 2^{j_1/2}\sum_t \beta_{j_1,t}\Phi(2^{j_1-L}x-2^{-L}t). \qquad (33)
\end{aligned}
$$

Now we define the vectors of scaling and wavelet coefficients of the interpolating wavelet transforms of the functions $f,\ g,\ g^*$ as $\theta_f = ((\gamma_{j_1,\cdot}), (\delta_{j_1,\cdot}), (\delta_{j_1+1,\cdot}), \dots)$, $\theta_g = ((\beta_{j_1,\cdot}), \mathbf{0}, \mathbf{0}, \dots)$, and $\theta_g^* = ((\beta^*_{j_1,\cdot}), \mathbf{0}, \mathbf{0}, \dots)$. The equivalences $\|\theta_g\|_{b^\sigma_{p,q}} \asymp \|g\|_{B^\sigma_{p,q}}$ and $\|\theta_f\|_{b^\sigma_{p,q}} \asymp \|f\|_{B^\sigma_{p,q}}$ are shown in [11]. Consequently, the equivalence

$$\|\theta_g^*\|_{b^\sigma_{p,q}} \asymp \|g^*\|_{B^\sigma_{p,q}} \qquad (34)$$

also holds. From (32), we obtain

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

With the norm equivalence in (34), the desired result

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

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-thresholding of the orthogonal wavelet coefficients is a discrete orthogonal projection as in Fig. 1 and apply Theorem 4.2, the smoothness of the estimated function always depends on the function ϕ. Therefore, the range of the smoothness parameter σ is bounded by $\min(\mathrm{Reg}(\phi),\ D-1)$. The averaging process in the translation-invariant case makes it possible to extend the range of the smoothness parameter to $\min(\mathrm{Reg}(\Phi),\ 2D-1)$. This can be seen in the example when f is a polynomial of order P with $D < P < 2D$. In this case, the coefficients of the TIDWT are not necessarily zero since the wavelet system, based on ϕ and ψ, is of order D. However, the wavelet coefficients of the autocorrelation transform are zero since that system is of order $2D$. In this case, we are exactly in the situation (26) in Theorem 4.4. Using the TIDWT, we can guarantee a smoother interpolation through denoised data samples than using the DWT.

Returning to Theorem 4.3 and the Remark related to it, we know that the threshold $t_{(L+1)N} = \epsilon\sqrt{2(1+\delta_L)\log((L+1)N)}$ is sufficient, but not necessary for $L = \text{const}$ and $N \to \infty$, because the threshold $\epsilon\sqrt{2\log((L+1)N)}$ would already be sufficient in that case. But the smaller threshold is not necessarily sufficient in the case $L = \log_2 N$ and $N \to \infty$. For small N, however, as, e.g., $N = 512$ in the experiments, applying the threshold $\epsilon\sqrt{2(1+\delta_L)\log((L+1)N)}$, one is more likely to threshold wavelet coefficients completely in subintervals of $(0, 1)$ where f is smooth. That means in those intervals we make the transition to the interpolation by the smoother autocorrelation functions. In singular parts of f, as, e.g., in neighborhoods of the singularities in the *Bumps* example, the transition becomes more unlikely and an interpolation using the more irregular orthogonal wavelet and scaling functions is visible (Fig. 2).

5. NUMERICAL EXPERIMENTS

In this section, we show by 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 $\ell \leq L$. Therefore, the chosen threshold in the soft-threshold operator is $t_{\ell N}$ for the wavelet coefficients at scale l. The experiments were done using the `Rice Wavelet Toolbox for Matlab`.

Figure 2 shows the *Bumps* example created with Donoho's software package `WaveLab` for the sampling rate $N = 512$. The added noise has the standard deviation $\epsilon = 1/10$. The denoising results for different thresholds, namely, the two fixed thresholds $\epsilon\sqrt{2\log N}$, $\epsilon\sqrt{2\log(LN)}$ and the level-dependent threshold $\epsilon\sqrt{2(1+\delta_\ell)\log(\ell N)}$ $(\ell = 1, \ldots, L)$, show the different smoothness properties of the denoised signal. Using a DWT with the threshold $\epsilon\sqrt{2\log N}$ in (b), the smoothness of the reconstruction is limited by the smoothness of the functions ϕ and ψ. Using the TIDWT in (c) and (d), a much smoother reconstruction is obtained in smooth regions away from the singularities, while singularities are kept sharp. Using the threshold $\epsilon\sqrt{2\log(LN)}$, the smoothness suffers close to the singularities in regions where the original function had a regularity $> \text{Reg}(\Phi)$, while using the threshold $\epsilon\sqrt{2(1+\delta_\ell)\log(\ell N)}$ results in a smooth reconstruction everywhere except in the singularities (see closeups of (c) and (d)). This analysis was also done for the various standard test signals in the `WaveLab` package, and we have chosen the *Bumps* example to explain the results from the previous section since it contains singularities with regularities $<\text{Reg}(\phi)$ and smooth regions with regularities $>\text{Reg}(\Phi)$.

Figure 3 shows the elements of $M_L(\mathbf{f})$, $M_L(\mathbf{d})$, and the thresholded coefficients $\eta_{t_{\ell N}} \circ M_L(\mathbf{d})$. In the last case, we see that in the interval bounded by the dotted lines after thresholding, all wavelet coefficients are zero at all levels $\ell \leq 3$. Therefore, for a maximal level of decomposition $L = 3$, we result in the situation described in (26) of Theorem 4.4. That means there exists a function that interpolates the sequence given by $M_L^\dagger \circ \eta_{t_{\ell N}} \circ M_L(\mathbf{d})$ that is, in the dotted interval, a linear combination of the dilated and translated autocorrelation function Φ. Using a DWT, only the existence of an interpolating

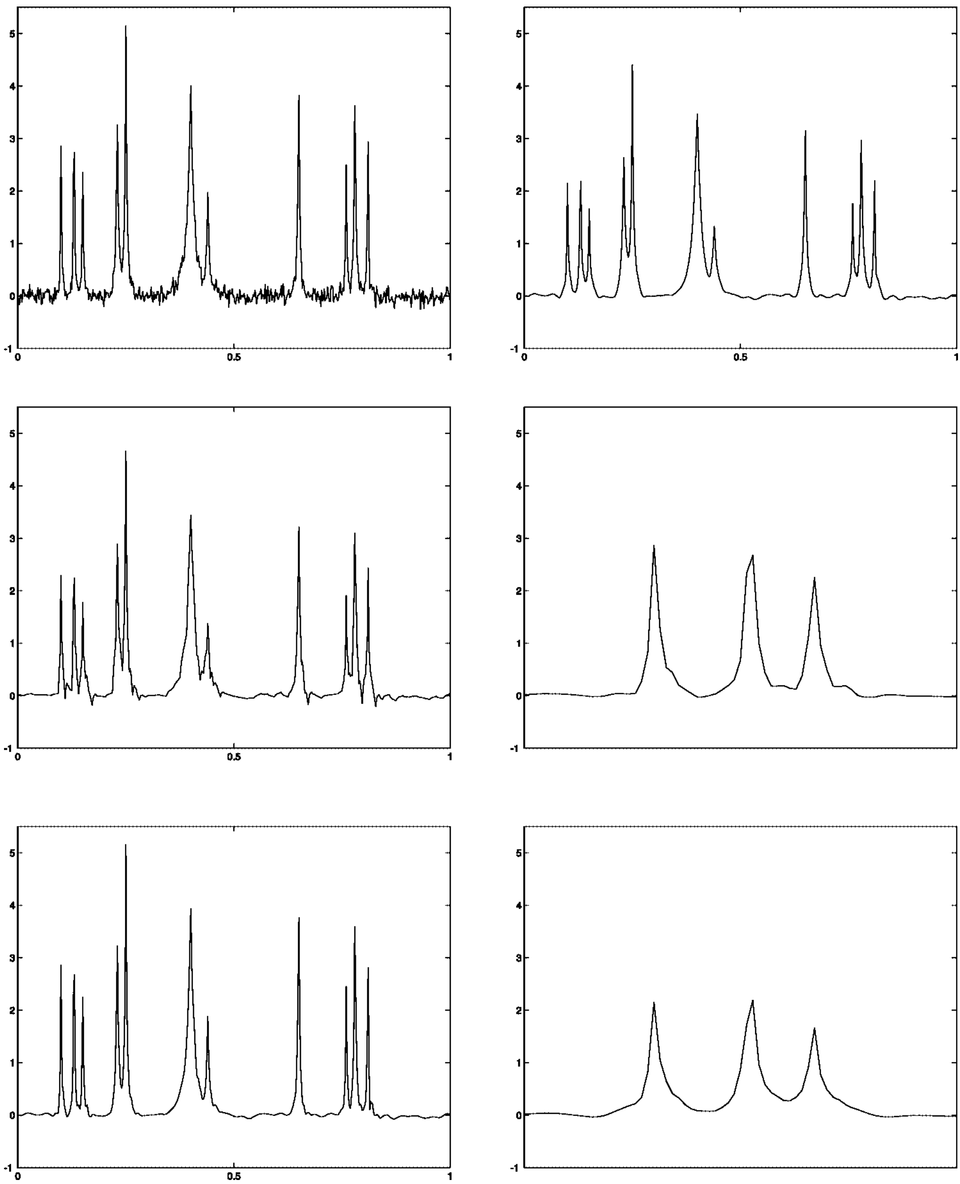

FIG. 2. (a) Noisy *Bumps* signal for $N = 512$ with $\epsilon = 1/10$ (*top left*), (b) orthogonal denoised with $t = \sqrt{2\log N}$ (*top right*), (c) translation-invariant denoised with $t = \epsilon\sqrt{2\log(LN)}$ (*middle left*); (d) translation-invariant denoised with $t = \epsilon\sqrt{2(1+\delta_\ell)\log(\ell N)}$ (*middle right*); closeups of the far left three bumps in (b) (*bottom left*) and (d) (*bottom right*). The chosen wavelet is Daubechies-4 ($D = 2$) and $L = 3$ is the maximal level of decomposition.

function being a linear combination of the dilated and translated version of ϕ can be guaranteed in the dotted interval.

6. CONCLUDING REMARKS

The reader will have noticed that we did not study the second denoising result in [13, 14], which is a statement on the optimality of the soft-thresholding technique in the field

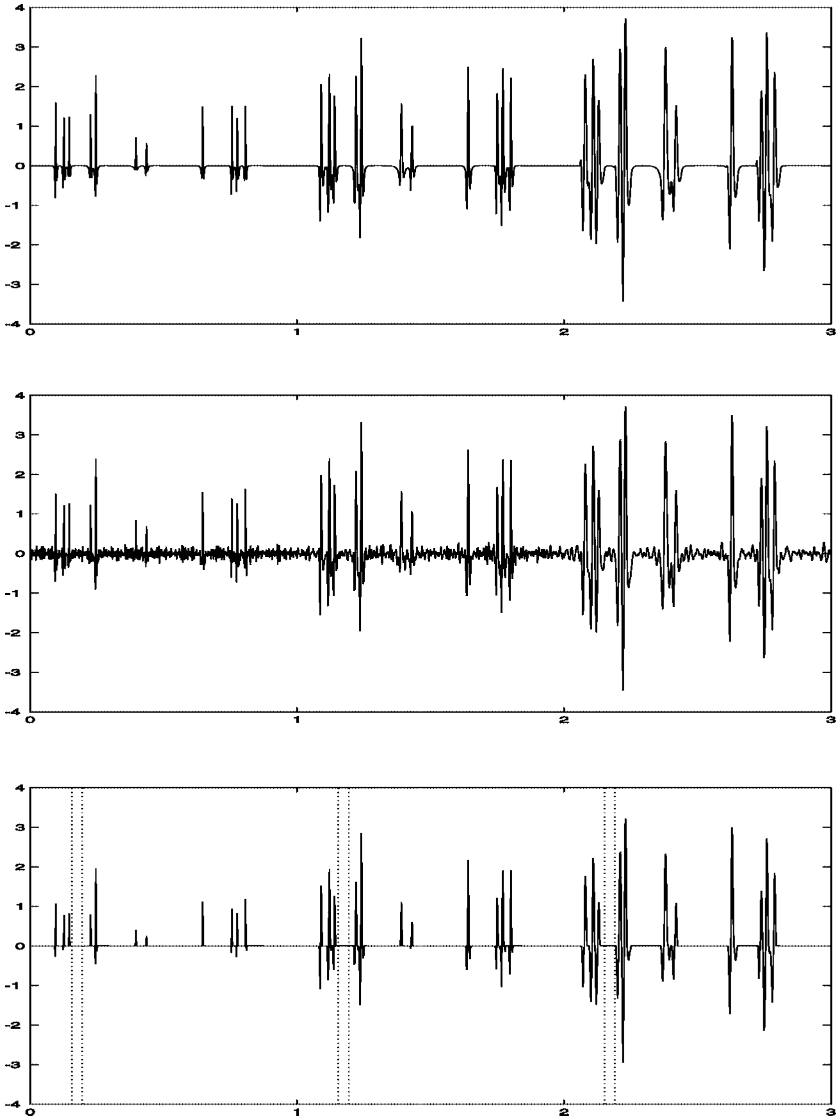

FIG. 3. Wavelet coefficients of the experiments from Fig. 2 for three levels of decomposition. Coefficients of the original signal (*top*), noisy signal (*middle*), thresholded with $t = \epsilon\sqrt{2(1+\delta_\ell)\log(\ell N)}$ (*bottom*). The dotted lines mark an interval where the wavelet coefficients vanish at all levels 1, 2, 3.

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 than does 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 cannot 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 at Rice University for inspiring ideas on various topics in wavelet theory. Furthermore, we thank the anonymous reviewers for their helpful comments.

REFERENCES

1. G. Beylkin, On the representation of operators in bases of compactly supported wavelets, *SIAM J. Numer. Anal.* **6** (1992), 1716–1740.
2. G. Beylkin, R. R. Coifman, and V. Rohlkin, Fast wavelet transforms and numerical algorithms I, *Comm. Pure Appl. Math.* **44** (1991), 141–183.
3. G. Beylkin and B. Torréssani, Implementation of operators via filter bank, autocorrelations shell and hardy wavelets, Tech. rep., Centre de Physique Théorique–CNRS–Lyminy, Marseille, 1994.
4. C. S. Burrus, R. A. Gopinath, and H. Guo, "Introduction to Wavelets and Wavelet Transforms," Prentice-Hall, Englewood Cliffs, NJ, 1998.
5. A. Cohen, I. Doubechies, B. Jawerth, and P. Vial, Multiresolution analysis, wavelets, and fast algorithms on an interval, *Compt. Rend. Acad. Sci. Paris (A)* **316** (1992), 417–421.
6. R. Coifman and M. Wickerhauser, Entropy-based algorithms for best basis selection, *IEEE Trans. Inform. Theory* **38** (1992), 713–719.
7. R. R. Coifman and D. L. Donoho, Translation invariant denoising, *in* "Wavelets and Statistics," Springer Lecture Notes, Vol. 103, Springer-Verlag, Berlin, 1995.
8. M. S. Crouse, R. D. Novak, and R. G. Baraniuk, Wavelet-based statistical signal processing using hidden Markov models, *IEEE Trans. Signal Process.* **46** (1998), 886–902.
9. I. Daubechies, "Ten Lectures on Wavelets," SIAM, Philadelphia, 1992.
10. G. Deslauriers and S. Dubuc, Symmetric iterative interpolation processes, *Constr. Approx.* **5** (1989), 49–68.
11. D. L. Donoho, Interpolating wavelet transforms, Tech. rep., Dept. of Statistics, Stanford University, 1992.
12. D. L. Donoho, Smooth wavelet decompositions with blocky coefficient kernels, *in* "Recent Advances in Wavelet Analysis" (L. L. Schumaker and G. Webb, Eds.), Academic Press, New York, 1994.
13. D. L. Donoho, De-noising by soft-thresholding, *IEEE Trans. Inform. Theory* **41** (1995), 613–627.
14. D. L. Donoho and I. M. Johnstone, Ideal spatial adaption by wavelet shrinkage, *Biometrika* **81** (1994), 425–455.
15. D. L. Donoho and T. P. Y. Yu, Robust nonlinear wavelet transform based on median-interpolation, *in* "Proceedings of the 31st. Asilomar Conference on Signals, Systems, and Computers," Pacific Grove, CA, 1997.
16. M. Frazier, B. Jawerth, and G. Weiss, "Littlewood–Paley Theory and the Study of Function Spaces," NSF-CBMS Regional Conf. Series in Mathematics, Vol. 79, Amer. Math. Soc., Providence, RI, 1991.
17. R. A. Gopinath, J. E. Odegard, and C. S. Burrus, On the correlation structure of multiplicity M scaling functions and wavelets, *in* "Proceedings of ISCAS-92," 1992.
18. H. Guo, Theory and applications of the shift-invariant, time-varying and undecimated wavelet transforms, M.S. thesis, George R. Brown School of Engineering, Rice University, 1995.
19. M. Holschneider, "Wavelets—An Analysis Tool," Oxford Univ. Press, New York, 1995.
20. M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, A real-time algorithm for signal analysis with the help of the wavelet transform, *in* "Wavelets: Time-Frequency Methods and Phase Space," pp. 286–297, Springer-Verlag, Berlin, 1989.
21. I. M. Johnstone and B. W. Silverman, Wavelet threshold estimators for data with correlated noise, *J. R. Statist. Soc. B* **59** (1997), 319–351.
22. M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, Jr., Nonlinear processing of a shift invariant dwt for noise reduction, *in* "Proccedings of SPIE," Orlando, FL, 1995.

23. M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, Jr., Noise reduction using an undecimated discrete wavelet transform, *IEEE Signal Process. Lett.* **3** (1996), 10–12.

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

25. S. Mallat, Zero-crossings of a wavelet transform, *IEEE Trans. Inform. Theory* **37** (1991), 1019–1033.

26. Y. Meyer, "Wavelets and Operators," Cambridge Univ. Press, Cambridge, UK, 1992.

27. R. D. Novak, Optimal signal estimation using cross-validation, *IEEE Signal Process. Lett.* **41** (1997), 23–25.

28. H. L. Resnikoff and R. O. Wells, Jr., "Wavelet Analysis: The Scalable Structure of Information," Springer-Verlag, New York, 1998.

29. N. Saito and G. Beylkin, Multiresolution representations using the auto-correlation functions of compactly supported wavelets, *IEEE Trans. Signal Process.* **41** (1993), 3584–3590.

30. M. J. Shensa, The discrete wavelet transform: Wedding the à trous and Mallat algorithms, *IEEE Trans. Signal Process.* **40** (1992), 2464–2482.

31. G. Strang, "Linear Algebra and Its Applications," Academic Press, New York, 1976.

32. G. Strang and T. Nguyen, "Wavelets and Filter Banks," Wellesley-Cambridge Press, Cambridge, MA, 1996.

33. W. Sweidens, The lifting scheme: A custom-design construction of biorthogonal wavelets, *Appl. Comput. Harmon. Anal.* **3** (1996), 186–200.

34. J. Tian and R. O. Wells, Jr., Dyadic rational biorthogonal Coifman wavelet systems, Tech. Rep. CML TR96-07, Computational Mathematics Laboratory, Rice University, 1996.

35. H. Triebel, "Theory of Function Spaces," Monographs in Math., Vol. 78, Birkhäuser Verlag, Basel, 1983.