Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 214

Integrodifferential Equations for Multiscale Wavelet Shrinkage: The Discrete Case

Stephan Didas, Gabriele Steidl and Joachim Weickert

Saarbrücken 2008

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 214
submitted: July 22, 2008

Integrodifferential Equations for Multiscale Wavelet Shrinkage: The Discrete Case

Stephan Didas
Mathematical Image Analysis Group
Department of Mathematics and Computer Science
Saarland University
D-66041 Saarbrücken
Germany
didas@mia.uni-saarland.de

Gabriele Steidl
University of Mannheim
Department of Mathematics and Computer Science
D-68131 Mannheim
Germany
steidl@math.uni-mannheim.de

Joachim Weickert
Mathematical Image Analysis Group
Department of Mathematics and Computer Science
Saarland University
D-66041 Saarbrücken
Germany
weickert@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

Integrodifferential Equations for Multiscale Wavelet Shrinkage: The Discrete Case

S. Didas* G. Steidl† J. Weickert*

July 22, 2008

Abstract

We investigate the relations between wavelet shrinkage and integrodifferential equations for image simplification and denoising in the discrete case. Previous investigations in the continuous one-dimensional setting are transferred to the discrete multidimentional case. The key observation is that a wavelet transform can be understood as derivative operator in connection with convolution with a smoothing kernel. In this paper, we extend these ideas to the practically relevant discrete formulation with both orthogonal and biorthogonal wavelets. In the discrete setting, the behaviour of the smoothing kernels for different scales is more complicated than in the continuous setting and of special interest for the understanding of the filters. With the help of tensor product wavelets and special shrinkage rules, the approach is extended to more than one spatial dimension. The results of wavelet shrinkage and related integrodifferential equations are compared in terms of quality by numerical experiments.

Keywords: Image Denoising, Wavelet Shrinkage, Integrodifferential Equations
AMS subject classification: 68U10,45K05,65T60

1 Introduction

Since the beginning of the 1990s, wavelet shrinkage and nonlinear diffusion filtering are two established classes of methods for signal and image simplification and denoising [35, 11, 26, 37].
The idea behind wavelet shrinkage is to denoise an image by performing very simple pointwise operations in a suitable multiresolution representation of the data

*Mathematical Image Analysis Group, Department of Mathematics and Computer Science, Saarland University, D-66041 Saarbrücken, Germany, {didas,weickert}@mia.uni-saarland.de
†Faculty of Mathematics and Computer Science, University of Mannheim, D-68131 Mannheim, Germany, steidl@math.uni-mannheim.de

[35]. This representation is obtained by using the wavelet transform. Depending on the application, the use of different types of wavelets might be suitable. Via the concept of multiresolution analysis [18, 19, 21] the shrinkage technique is closely related to earlier signal processing methods like filter banks and subband coding [5, 6, 22, 28, 33, 32].

Nonlinear diffusion filtering simplifies and denoises an image by solving a partial differential equation which is typically done without changing the spatial representation of the image. In this setting, first or higher order derivatives of the image are used to formalise the desired smoothness and to detect and eliminate the noise [26, 37, 17, 10].

The close relationship between both methods is emphasised, for example, by the fact that wavelet shrinkage can also be understood as energy minimisation [3, 4, 2]. This fact already relates it to the context of scale-spaces [15, 39, 26, 1] and PDE-based methods. In the discrete setting, translationally invariant wavelet shrinkage on the finest scale is even equivalent to total variation regularisation and diffusion [29].

The connections between multiscale wavelet shrinkage and corresponding integrodifferential equations in the continuous one-dimensional setting have been the topic of an earlier publication by the authors [9]. The goal of this paper is to transfer the ideas and results from the continuous to the practically relevant discrete setting. Since the dilation operation on the wavelets can only be approximated on a discrete grid, the formulation is slightly more technical here. Moreover, we will not restrict ourselves to orthogonal wavelets, but also have a look at biorthogonal ones allowing for more general integrodifferential equations. Preliminary results concerning this transfer have been presented at a conference [8]. In addition, we will transfer the one-dimensional case to two dimensions using tensor product wavelets and special shrinkage rules to increase rotational invariance. We also discuss in detail the behaviour of the appearing smoothing kernels at different scales. Numerical experiments will be shown to compare the resulting methods in terms of denoising quality.

This paper is organised as follows: Section 2 introduces some notations used throughout the paper. Sections 3 and 4 describe classical wavelet shrinkage and nonlinear diffusion filtering in a discrete setting. The factorisation of a discrete wavelet into a convolution kernel and a derivative approximation is derived in Section 5. In Section 6, this idea is used to derive relations between discrete wavelet shrinkage and integrodifferential equations. Section 7 shows how these ideas can be generalised two higher dimensions. Numerical experiments in Section 8 display the behaviour of the presented filters in practice. The paper is concluded with a summary in Section 9.

2 Preliminaries and Notations

Let us start with the notations used
throughout this paper. Let $f \in \ell^2(\mathbb{Z}) := \{(f_n)_{n\in\mathbb{Z}} \mid \sum_{n=-\infty}^{\infty} f_n^2 < \infty\}$ be a real signal of infinite length. Then

$$\hat{f}(\omega) \;:=\; \sum_{n=-\infty}^{\infty} f_n \exp(-in\xi) \quad \text{and} \quad F(z) \;:=\; \sum_{n=-\infty}^{\infty} f_n z^{-n} \tag{1}$$

denote the *Fourier- and the z-transform* of f, respectively. The importance of the z-transform in this context results from the fact that it allows for an easy formulation of convolutions as multiplications of formal Laurent series. More precisely, the k-th component of the *convolution* $a * f$ given by

$$(a * f)_k := \sum_{j\in\mathbb{Z}} a_j f_{k-j}$$

corresponds to the coefficient of z^{-k} in $A(z)F(z)$.
In practice, we will work with signals of finite length N and assume N-periodic extensions of the signals. Then the k-th component of the *cyclic convolution* $a*f$ of the vectors $a, f \in \mathbb{R}^N$ given by

$$(a * f)_k := \sum_{j=0}^{N-1} a_j f_{(k-j) \bmod N}$$

corresponds to the coefficient of z^{-k} in $A(z)F(z) \bmod z^N - 1$. On the other hand, the cyclic convolution of $a, f \in \mathbb{R}^N$ can be expressed as multiplication of f with the *circulant matrix* corresponding to a [14]:

$$A \;:=\; \begin{pmatrix} a_0 & a_1 & a_2 & \dots & a_{N-1} \\ a_{N-1} & a_0 & a_1 & \dots & a_{N-2} \\ a_{N-2} & a_{N-1} & a_0 & \dots & a_{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & a_3 & \dots & a_0 \end{pmatrix} \;\in \mathbb{R}^{N,N} \quad .$$

Each circulant matrix can be written as

$$A \;:=\; \sum_{j=0}^{N-1} a_j C^j, \qquad \text{where} \quad C \;:=\; \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 0 \end{pmatrix}$$

denotes the so-called *basic circulant permutation matrix.* Multiplication with C performs a periodic left-shift of a vector.

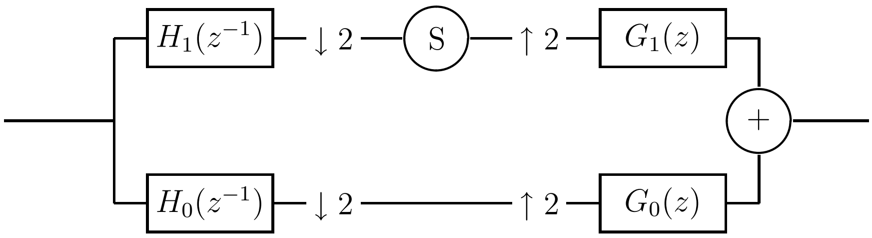

Figure 1: Filter bank for wavelet shrinkage on the finest scale.

In the following we will often use some vector $a \in \mathbb{R}^N$ in connection with its corresponding N-dimensional circulant matrix $A = \sum_{j=0}^{N-1} a_j C^j$ and its z-transform $A(z) = \sum_{j=0}^{N-1} a_j z^{-j}$. The circulant $N \times N$-matrices can be diagonalised by the same matrix, namely the N-th Fourier matrix. Hence, the multiplication of circulant matrices is commutative.

3 Discrete Wavelet Shrinkage

In this section, we review the three steps of wavelet shrinkage in the discrete setting [35]: Figure 1 shows the corresponding filter bank for wavelet shrinkage on the finest scale, where the z-transform notation of the filters is used.

1. **Analysis:** In the analysis step, the initial signal is transferred to a wavelet coefficient representation. This decomposition is done with the help of the analysis filters h_0 and h_1 which can be obtained as scaling coefficients of the corresponding scaling function. The filter h_0 plays the role of a low-pass filter, and h_1 plays the role of the corresponding high-pass filter. In addition, both channels are sampled down by leaving out all components with an odd index. This is indicated in the filter bank with the symbol $\downarrow 2$.

2. **Shrinkage:** The wavelet coefficients of the signal are shrunken towards zero in this step while the low-frequency components are kept. This is modelled as applying a nonlinear *shrinkage function* $S : \mathbb{R} \to \mathbb{R}$ to each of the wavelet coefficients.

3. **Synthesis:** In this step, the resulting signal is synthesised from the wavelet coefficients. First an upsampling is used by introducing zeros between each pair of neighbouring signal components. This is written as $\uparrow 2$ here. For the synthesis, the filter pair g_0 and g_1 is used.

We note that the analysis filters h_0 and h_1 are mirrored in our notation. To ensure perfect reconstruction of the signal, the analysis and the synthesis filters

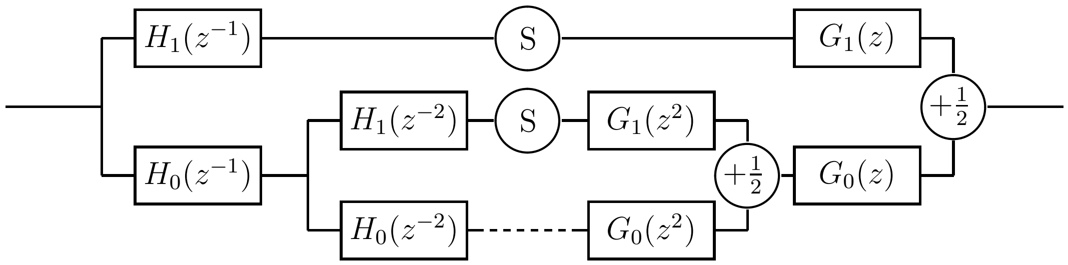

Figure 2: Filter bank for translational invariant wavelet shrinkage with multiple scales using the algorithme à trous.

have to satisfy the following properties, [34, 30, 20]:

$$G_0(z)H_0(z^{-1}) + G_1(z)H_1(z^{-1}) = 2, \tag{2}$$
$$G_0(z)H_0(-z^{-1}) + G_1(z)H_1(-z^{-1}) = 0 \ . \tag{3}$$

For filters of finite length, one can further show (see [34, p. 120] or [20, Theorem 7.9], for example) that there are numbers $\alpha \neq 0$ and $k \in \mathbb{Z}$ such that

$$G_0(z) = \frac{2}{\alpha} z^{2k+1} H_1(-z^{-1}) \quad \text{and} \quad G_1(z) = -\frac{2}{\alpha} z^{2k+1} H_0(-z^{-1}) \ . \tag{4}$$

For simplicity, we assume without loss of generality that $\alpha = 2$ and $k = 0$. This gives us the simple relations between analysis and synthesis filters:

$$G_0(z) = zH_1(-z^{-1}), \quad G_1(z) = -zH_0(-z^{-1}) \ . \tag{5}$$

It immediately follows that

$$H_0(z) = zG_1(-z^{-1}) \ . \tag{6}$$

These equations hold for the general biorthogonal case with filters of finite length. In order to have orthonormal filters, we have the additional requirement that

$$G_i(z) = H_i(z) \quad \text{for } i \in \{0, 1\} \tag{7}$$

which allows us to determine all four filters with one prototype.
To make wavelet methods compatible to PDE approaches we need a translation invariant wavelet shrinkage process. This can be obtained by skipping the down- and up-sampling procedure as shown in Figure 2. For the synthesis, the result has to be multiplied with 1/2 at each scale. This is also known as *algorithme à trous*, cf. Holschneider et al. [13, 20]. We see that the analysis and synthesis filters are widened by inserting zeros into the filters.

4 Discrete Higher Order Nonlinear Diffusion

Next, let us have a look at the discretisations of nonlinear diffusion which we will need in this section.
Here we use a discretisation of the nonlinear higher order diffusion equation

$$\partial_t u \;=\; (-1)^{p+1}\, \partial_x^p \Big(g\left((\partial_x^p u)^2\right) \partial_x^p u \Big) \tag{8}$$

with initial condition $u(\cdot, 0) = f$ as it has been described in [10], for example.
We restrict our attention to N-periodic signals on the interval $[0, N-1]$. To discretise this equation, we consider the sampled version $u \in \mathbb{R}^N$ of u at an equidistant grid $\{jh \;:\; j = 0, \ldots, N-1\}$ with spatial step size $h = 1$.
To approximate the spatial derivatives in (8), we use a forward difference as approximation of the first derivative. It can be expressed in matrix-vector form as $\partial_x \sim Du$, where

$$D \;:=\; \begin{pmatrix} -1 & 1 & 0 & \ldots & 0 \\ 0 & -1 & 1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & -1 & 1 \\ 1 & 0 & \ldots & 0 & -1 \end{pmatrix} \;=\; C - I \tag{9}$$

and in terms of the z-transform as $D(z)u(z)\text{mod}(z^N - 1)$ with $D(z) = z^{-1} - 1$. Then the transposed matrix D^T corresponds to the z-transform $D^T(z) = z - 1$ and yields an approximation of the negated first derivative with a backward difference. Further D^p and $(D^T)^p$ serve as approximations of p-th derivatives with appropriate sign. For time discretisation we use a simple Euler forward scheme. Then the discrete iterative scheme can be written as

$$\begin{aligned} u^0 &= f \\ u^{k+1} &= u^k - \tau (D^T)^p \Phi_{D^p}(u^k) D^p u^k, \qquad k \in \mathbb{N}\ . \end{aligned} \tag{10}$$

The diagonal matrix $\Phi_{D^p}(u^k) := \text{diag}\Big(g(|(D^p u^k)_j|)\Big)_{j=0,\ldots,N-1}$ stands for the multiplication with the nonlinear diffusivity function. In our computations we use the Perona-Malik function [26] defined as

$$g(s^2) \;=\; \left(1 + \frac{s^2}{\lambda^2}\right)^{-1} \ . \tag{11}$$

See for example [25] for a list of other possible diffusivity functions.

5 Discrete Wavelets and Convolution Kernels

In this section, we formulate the key idea of factorising discrete wavelets into derivative approximations of smoothing kernels. We make the assumption that the wavelet has p vanishing moments to relate the wavelet transform to an approximation of the p-th derivative. In the discrete setting, this condition reads as: A signal $f \in \ell^2(\mathbb{Z})$ is said to have $p \in \mathbb{N}$ *vanishing moments* if

$$\sum_{n=-\infty}^{\infty} n^j f_n = 0 \qquad \text{for } j \in \{0, \ldots, p-1\} \quad \text{and} \quad \sum_{n=-\infty}^{\infty} n^p f_n \neq 0 \ . \tag{12}$$

Let us now factorise the z-transform of a wavelet with p vanishing moments such that we obtain a derivative approximation filter and a convolution or smoothing kernel. Since the number of vanishing moments is directly connected with regularity properties, such factorisations are often used in the design of wavelets (see [7, 30, 20, 16], for example). It should also be noticed that the number of vanishing moments of the filter coefficients is the same as the number of (continuous) vanishing moments of the continuous wavelet function; see [20, Theorem 7.4].

Proposition 5.1 (Wavelet Filter Factorisation)
Let $f \in \ell^2(\mathbb{Z})$ be a filter of finite length and p vanishing moments. Then its z-transform can be decomposed as

$$F(z) \ = \ (z-1)^p \, K(z), \quad K(1) \neq 0 \ ,$$

where K is the z-transform of the corresponding filter k which will be understood as smoothing kernel.

Although it is standard in wavelet analysis, we attach the simple proof in order to make the paper more self-contained:
Proof: Since f has finite length, the Fourier transform $\hat{f} \in \mathscr{C}^\infty$ is infinitely many times differentiable. The j-th derivative of $\hat{f}$ at the point 0 is then

$$\hat{f}^{(j)}(0) \ = \ (-i)^j \sum_{n=-\infty}^{\infty} n^j f_n \tag{13}$$

which is the j-th moment of f times the nonzero constant $(-i)^j$. Our assumption about f then reads as $\hat{f}^{(j)} = 0$ for $j \in \{0, \ldots, p-1\}$. This means the Fourier transform of f is a trigonometric polynomial which has a zero of order p in 0. Thus it can be factorised as

$$\hat{f}(\xi) \ = \ (\exp(i\xi) - 1)^p \, K(\exp(i\xi)) \tag{14}$$

with a suitable (Laurent-) polynomial K. Replacing $\exp(i\xi)$ by z directly yields the desired factorisation $F(z) = (z-1)^p K(z)$ of the z-transform. $\square$

With the help of this proposition, we can understand the convolution with a wavelet as derivative approximation of a presmoothed signal. We remember that

$z-1$ is the z-transform of the finite difference matrix D^T approximating the negated first derivative. Thus $(z-1)^p$ can be used as approximation of $(-1)^p$ times the p-th derivative. This reasoning of understanding the wavelet as derivative of a smoothing kernel is in accordance with the approach in the previous section and the continuous considerations in [9]. For details on such factorisations, see [20, Section 7.2] for orthogonal wavelets and [20, Section 7.4.2] for the biorthogonal case, for example. Let p and q be the number of vanishing moments of our analysis and synthesis highpass filters H_1 and G_1. Then Proposition 5.1 allows us to write the filters as

$$H_1(z) \;=\; (z-1)^p K^H(z) \quad \text{and} \quad G_1(z) \;=\; (z-1)^q K^G(z) \tag{15}$$

where K^H and K^G are the z-transforms of two smoothing kernels k^H and k^G of the synthesis and analysis wavelet. For orthogonal wavelets, we simply have $K^H(z) = K^G(z)$ and $p = q$. With the two relations (5) and (6) between low- and highpass we see that for the lowpass filters H_0 and G_0, the following relations hold:

$$H_0(z) \;=\; (-1)^q \, z(z^{-1}+1)^q \, K^G(-z^{-1}), \tag{16}$$

$$G_0(z) \;=\; (-1)^p \, z(z^{-1}+1)^p \, K^H(-z^{-1}). \tag{17}$$

To make these formulae a bit more intuitive, let us now give some examples of kernels K^H and K^G for commonly used orthogonal wavelets on the finest scale:

Example 5.2 (Discrete Wavelets and Convolution Kernels)
(a) Haar Wavelet: For the discrete Haar wavelet, we have $H_1(z) = \frac{1}{\sqrt{2}}(z-1)$. The kernel on the finest scale is in this case just a scalar factor $K^H(z) = \frac{1}{\sqrt{2}}$.
(b) Daubechies Wavelets: The Daubechies wavelet [7] with $p = 2$ is represented by the filter

$$H_1(z) \;=\; \frac{1}{4\sqrt{2}}\left(\sqrt{3}-1+(3-\sqrt{3})z-(3+\sqrt{3})z^2+(1+\sqrt{3})z^3\right) \tag{18}$$

which can be factorised as $H_1(z) = (z-1)^2 K^H(z)$ leading to

$$K^H(z) \;=\; \frac{1}{4\sqrt{2}}\left(\sqrt{3}-1+(\sqrt{3}+1)z\right) \;. \tag{19}$$

Let us briefly say a few words about the differences between our idea and previous approaches to relations between shrinkage on the finest scale and nonlinear diffusion. In contrast to the idea in this paper, Weickert et al. [38] have directly considered the wavelet filter H_1 as stencil for a derivative approximation. With a Taylor expansion, one can directly prove that any filter with p vanishing moments yields an approximation of the p-th derivative up to a constant factor. This works well as long as only the finest scale is considered, but it does not help to explain

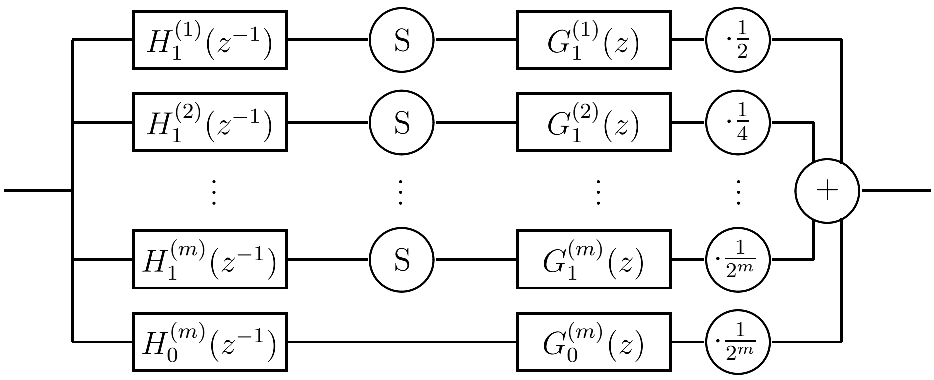

Figure 3: Filter bank for translation invariant wavelet shrinkage, written with multiple channels.

what happens on coarser scales. Here, we try to model coarser scales by separating the derivative approximation from the smoothing kernel which yields a coarse scale approximation of our signal. In the continuous setting considered in [9], the smoothing kernel is a function for which the scaling operation is invertible without loss of information. In contrast to this, the discrete wavelets on coarser scales treated in this paper can change their appearance due to discretisation effects.

Following [34, Section 3.3], we introduce *wavelets on coarser scales*: starting from the filters G_0 and G_1 on the finest scale, we define the wavelet filters $G_0^{(\sigma)}$ and $G_1^{(\sigma)}$ on coarser scales $\sigma \in \mathbb{N}$ as

$$G_0^{(\sigma)}(z) \;=\; \prod_{r=0}^{\sigma-1} G_0(z^{2^r}) \quad \text{and} \quad G_1^{(\sigma)}(z) \;=\; G_1(z^{2^{\sigma-1}})G_0^{(\sigma-1)}(z) \;, \qquad (20)$$

and use the same formulae for $H_0^{(\sigma)}$ and $H_1^{(\sigma)}$.

The exponents 2^r come from the fact that the algorithme à trous inserts the corresponding number of zeros between two samples of the filter at scale r. In addition, we have to multiply the z-transforms of all filters lying on the path from the input to the middle of the filter bank for H_i in Figure 2, or from the middle to the output for G_i, $i = 0, 1$.

Having these formulae at hand we can rewrite the filter bank in Figure 2 with $m+1$ different paths as shown in Figure 3. Now we are interested in the changes of the shape of the convolution kernels corresponding to the wavelets when the scale increases. Our starting point are the relations (20), and we firstly consider the scaling coefficients using the factorisation (17):

$$\begin{aligned} G_0^{(\sigma)}(z) &= \prod_{r=0}^{\sigma-1} G_0(z^{2^r}) = \prod_{r=0}^{\sigma-1} \left((-1)^p z^{2^r} (z^{-2^r}+1)^p \, K^H(-z^{-2^r})\right) \\ &= (-1)^{\sigma p}\, z^{2^\sigma-1} \left(\sum_{r=0}^{2^\sigma-1} z^{-r}\right)^p \prod_{r=0}^{\sigma-1} K^H(-z^{-2^r}) \ . \end{aligned} \tag{21}$$

We see that the scaling filter on larger scales can be decomposed into four parts: The sign given by $(-1)^{\sigma p}$ and the pure shift $z^{(2^\sigma)-1}$ do not change the shape of the convolution kernel. This shape is determined by the rightmost two factors: The second one is a product of the kernels k^H with alternating signs and with inserted zeros. This is actually the wavelet-dependent part. The first factor is independent of the wavelet: It is the p times convolution of a box filter of width 2^σ with itself. This can be understood as a discrete B-spline kernel of order p.
Let us see how this decomposition looks for the wavelet coefficients:

$$\begin{aligned} G_1^{(\sigma)}(z) &= G_1(z^{2^{\sigma-1}}) G_0^{(\sigma-1)}(z) \\ &= (z^{2^{\sigma-1}}-1)^q \, K^G(z^{2^{\sigma-1}})\, G_0^{(\sigma-1)}(z) \\ &= (z-1)^q \left(\sum_{r=0}^{2^{\sigma-1}-1} z^r\right)^q K^G(z^{2^{\sigma-1}})\, G_0^{(\sigma-1)}(z) \\ &= (z-1)^q\,(-1)^{(\sigma-1)p} \left(z^{-2^{\sigma-1}+1}\right)^{p-1} \left(\sum_{r=0}^{2^{\sigma-1}-1} z^r\right)^{p+q} \cdot \\ &\qquad \cdot K^G(z^{2^{\sigma-1}}) \prod_{r=0}^{\sigma-2} K^H(-z^{-2^r}) \ . \end{aligned}$$

Let us also analyse the ingredients of this product: The first factor $(z-1)^q$ tells us that the wavelet can be understood as approximation of the q-th derivative (with sign $(-1)^q$). It is the z-transform of the finite difference matrix $(D^T)^q$ defined above. Again, the sign and the shift do not change the shape of the convolution kernel. As for the scaling function, we also find a spline kernel of order $p+q$ and a wavelet-dependent part.
Let us now give some examples of commonly used wavelets to see how the related convolution kernels look like:

Example 5.3 (Haar Wavelet on Coarser Scales)
We have already seen that for a Haar wavelet we have $p = q = 1$ and the kernels $K^G(z) = K^H(z) = \frac{1}{\sqrt{2}}$ are just constants. Thus the wavelet on scale σ can be seen as

$$G_1^{(\sigma)}(z) = (-1)^{(\sigma-1)}\,(z-1)\frac{1}{2^{\frac{\sigma}{2}}} \left(\sum_{r=0}^{2^{\sigma-1}-1} z^r\right)^2 \ . \tag{22}$$

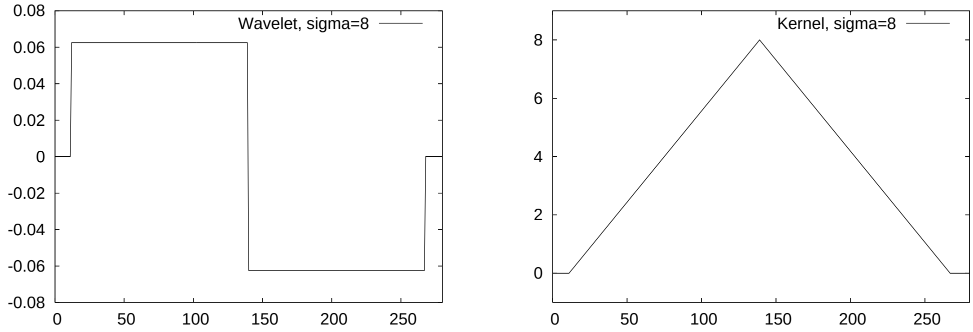

Figure 4: Convolution kernel corresponding to the Haar wavelet. *Left column:* Haar wavelet on scale 8. *Right column:* Corresponding smoothing kernel: a hat function.

This means that in complete analogy to the continuous case, the discrete Haar wavelet is the derivative approximation of a hat function. This hat is created by multiplying a box filter with itself. An example for the scale $\sigma = 8$ is shown in Figure 4.

Example 5.4 (Daubechies Wavelets on Coarser Scales)
For some representatives of the family of Daubechies wavelets [7], we display the corresponding kernels obtained by numerical calculations in Figure 5. One can see that the smoothing kernels have a shape similar to a Gaussian kernel with a perturbation at the right side where they even change the sign. Daubechies has proven that the Haar wavelets are the only symmetric or antisymmetric orthonormal wavelets with compact support [7], and so it is clear that the corresponding kernels of Daubechies wavelets of higher order can not be symmetric.

The following two examples consider the convolution kernels corresponding to biorthogonal filter pairs. These filters can be symmetric or antisymmetric with compact support. Hence, the convolution kernels can be symmetric.

Example 5.5 (Compactly Supported Spline Wavelets)
Figure 6 presents the compactly supported spline wavelet filters h_1 and g_1 with 3 and 7 vanishing moments. Details on these filters can be found in [20, p. 271], for example. We see that the corresponding kernel to h_1 has negative parts while the kernel derived from g_1 is positive and resembles a Gaussian kernel.

Example 5.6 (Perfect Reconstruction Filters of Most Similar Length)
These biorthogonal filters are displayed in Figure 7 and details can be found in [20, p. 273], for example. The filter corresponding to g_1 has some small negative parts.

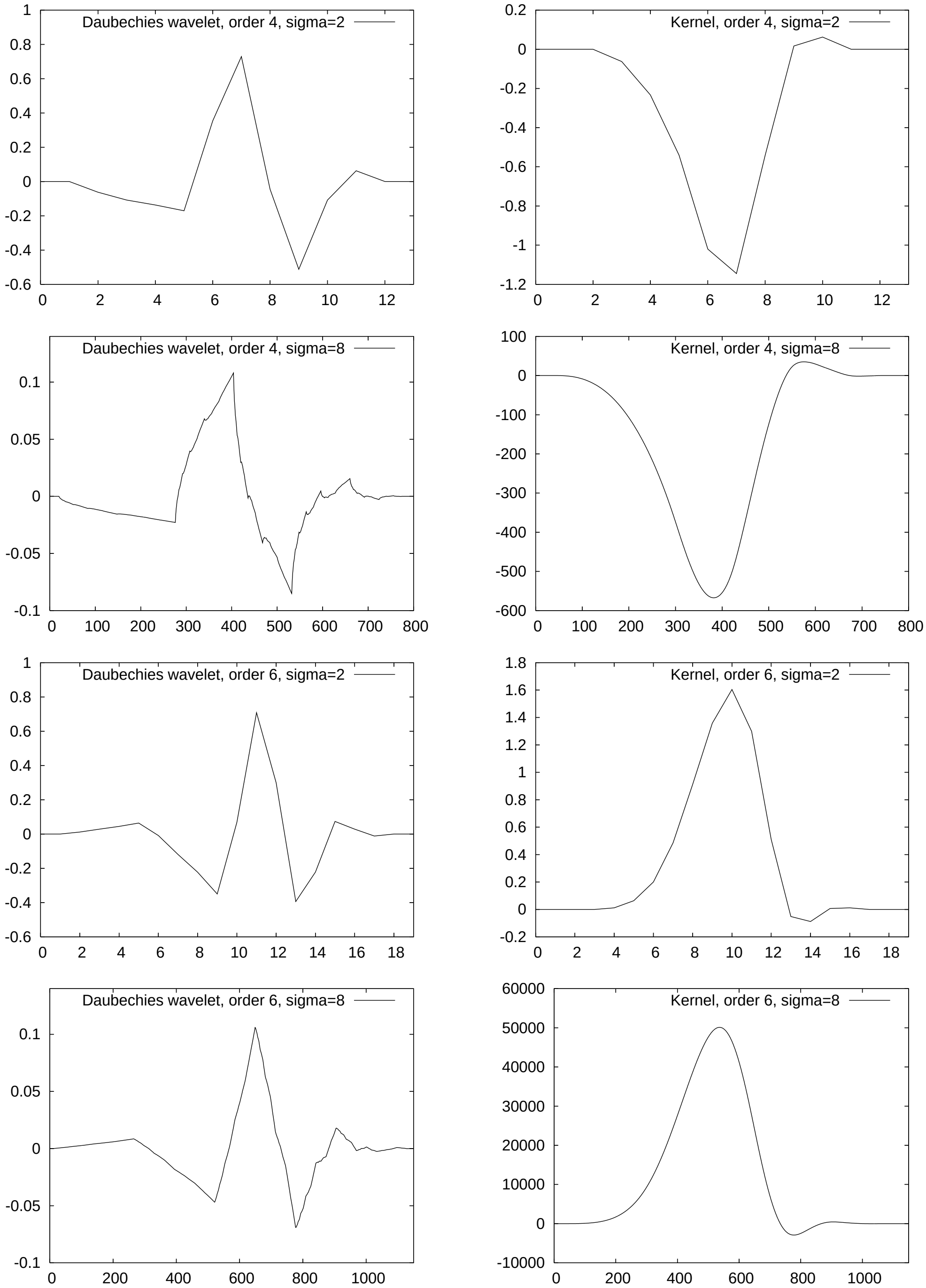

Figure 5: Convolution kernels corresponding to Daubechies wavelets on larger scales. *Left column:* Daubechies wavelets of orders 4 and 6 on scales 2 and 8. *Right column:* Corresponding smoothing kernels. The scaling comes from the fact that wavelets are normalised with respect to the ℓ^2-norm.

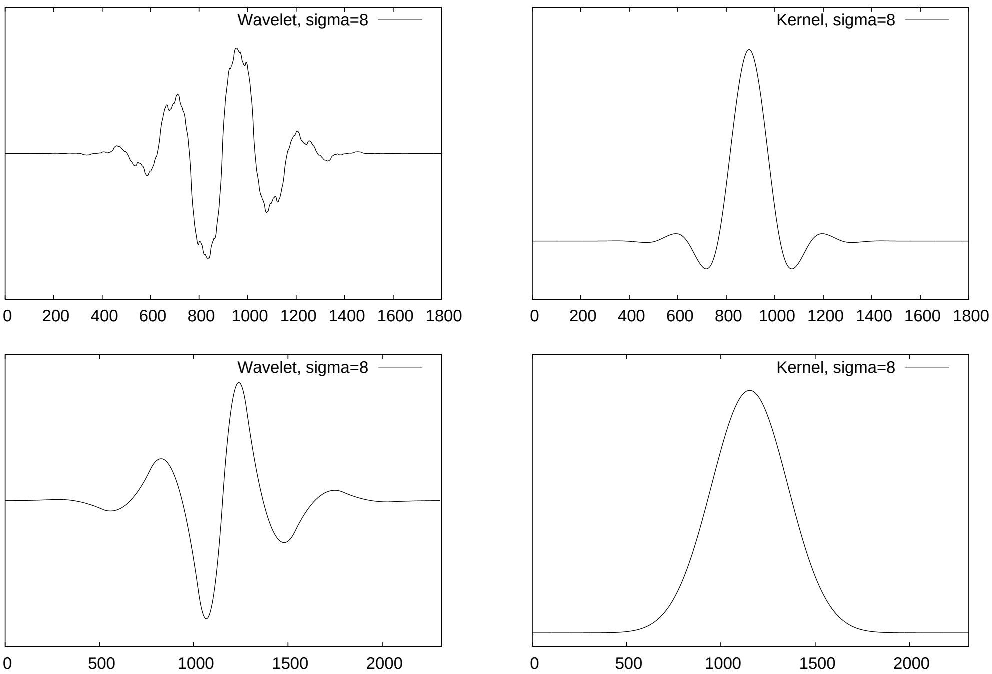

Figure 6: Convolution kernels corresponding to compactly supported spline wavelets on scale 8. *Top left:* Filter h_1 with 3 vanishing moments. *Bottom left:* Filter g_1 with 7 vanishing moments. *Right column:* Corresponding smoothing kernels.

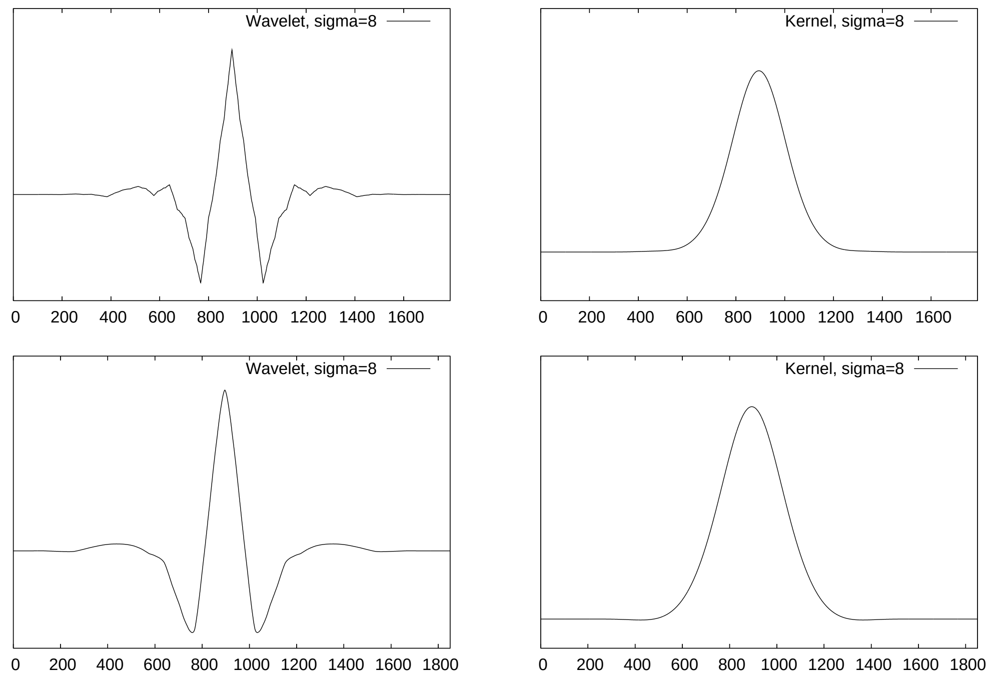

Figure 7: Convolution kernels corresponding to perfect reconstruction filters of most similar length on scale 8. *Left column:* Analysis and synthesis filter. *Right column:* Corresponding smoothing kernels.

Regardless the shape of the convolution kernels, it will be important for our considerations in the next section that we can write the analysis and the synthesis wavelet as

$$G_1^{(\sigma)}(z) \;=\; (1-z)^q \, K^{G,(\sigma)}(z) \quad \text{and} \quad H_1^{(\sigma)}(z) \;=\; (1-z)^p \, K^{H,(\sigma)}(z) \;. \tag{23}$$

We use the notions $K^{G,(\sigma)}$ and $K^{H,(\sigma)}$ to denote the corresponding convolution kernels on scale σ. With the finite difference matrices introduced in (9), we can rewrite (23) in matrix notation as

$$G_1^{(\sigma)} \;=\; \left(D^q\right)^T K^{G,(\sigma)} \quad \text{and} \quad H_1^{(\sigma)} \;=\; \left(D^p\right)^T K^{H,(\sigma)} \;. \tag{24}$$

We will use these equations in the next section to rewrite iterated wavelet shrinkage as discretisation of an integrodifferential equation.

6 Relations Between Both Methods

In this section, let $f, u \in \mathbb{R}^N$ be vectors and $H_i^{(\sigma)}, G_i^{(\sigma)}$, $i = 0, 1$ denote the $N \times N$ circulant matrices corresponding to the filters $H_i^{(\sigma)}(z), G_i^{(\sigma)}(z)$ modulo $z^N - 1$. Then we can rewrite wavelet shrinkage according to Figure 3 as

$$u \;=\; \sum_{\sigma=1}^{m} \frac{1}{2^\sigma} \, G_1^{(\sigma)} S\left(\left(H_1^{(\sigma)}\right)^T f\right) \;+\; \frac{1}{2^m} \, G_0^{(m)} \left(H_0^{(m)}\right)^T f \;. \tag{25}$$

The analysis matrices are transposed to reflect the fact that we have used $H_i(z^{-1})$ for $i = 0, 1$ for the analysis part of our filter banks in Figures 1, 2, and 3. The function S is meant to act componentwise on the vector entries.

Without shrinking the coefficients, the filter bank will allow for a perfect reconstruction, which means that

$$f \;=\; \sum_{\sigma=1}^{m} \frac{1}{2^\sigma} \, G_1^{(\sigma)} \left(H_1^{(\sigma)}\right)^T f \;+\; \frac{1}{2^m} \, G_0^{(m)} \left(H_0^{(m)}\right)^T f \tag{26}$$

for all $f \in \mathbb{R}^N$. Similar to [25, 9] we use

$$S(x) \;=\; (1 - \tau g(|x|))x \tag{27}$$

to rewrite our shrinkage function with the help of a function g which will play the role of a diffusivity later on. This leads to pairs of shrinkage functions and diffusivities which are studied in detail in [25]. Plugging (27) into (25) we obtain

$$\begin{aligned} u \;=\; & \sum_{\sigma=1}^{m} \frac{1}{2^\sigma} \, G_1^{(\sigma)} \left(H_1^{(\sigma)}\right)^T f \;+\; \frac{1}{2^m} \, G_0^{(m)} \left(H_0^{(m)}\right)^T f \\ & - \tau \sum_{\sigma=1}^{m} \frac{1}{2^\sigma} \, G_1^{(\sigma)} \, \Phi\left(\left(H_1^{(\sigma)}\right)^T f\right) \left(H_1^{(\sigma)}\right)^T f \;, \end{aligned} \tag{28}$$

where Φ is a diagonal matrix such that $\Phi(v) := \mathrm{diag}\Big(g(|v_j|^2)\Big)_{j\in J}$. By property (26) the first part is just the reconstruction of the initial signal f, and we obtain

$$u \;=\; f - \tau \sum_{\sigma=1}^{m} \frac{1}{2^\sigma}\, G_1^{(\sigma)}\, \Phi\left(\left(H_1^{(\sigma)}\right)^T f\right)\, \left(H_1^{(\sigma)}\right)^T f \tag{29}$$

for one multilevel shrinkage step. Iterating these multilevel shrinkage steps leads to the scheme

$$\begin{aligned} u^0 \;&=\; f \\ u^{k+1} \;&=\; u^k - \tau \sum_{\sigma=1}^{m} \frac{1}{2^\sigma}\, G_1^{(\sigma)}\, \Phi\left(\left(H_1^{(\sigma)}\right)^T u^k\right)\, \left(H_1^{(\sigma)}\right)^T u^k \;, \quad k\in\mathbb{N} \end{aligned} \tag{30}$$

which has a similar structure as the discretisation of the nonlinear diffusion equation (10). Using (24), the iteration rule can be written as

$$\begin{aligned} u^{k+1} \;&=\; u^k \\ &\quad -\tau \sum_{\sigma=1}^{m} \frac{1}{2^\sigma}\, (D^q)^T\, K^{G,(\sigma)}\, \Phi\left(D^p \left(K^{H,(\sigma)}\right)^T u^k\right)\, D^p \left(K^{H,(\sigma)}\right)^T u^k \;. \end{aligned} \tag{31}$$

A continuous equivalent, the integrodifferential equation

$$u^{k+1} \;=\; u^k - \tau(-1)^{p+1} \int_0^\infty \sigma^{2p}\, \partial_x^p\, \theta_\sigma * \left(g\left(\sigma^p\, \partial_x^p\, \tilde\theta_\sigma\right)\, \left(\partial_x^p\, \tilde\theta_\sigma * u\right)\right)\, \frac{d\sigma}{\sigma^2}$$

with a smoothing kernel θ_σ and its mirrored version $\tilde\theta_\sigma$ has been derived in [9]. It becomes evident that (31) can be considered as discrete version of this integrodifferential equation. As in the continuous case, we see also in our discrete setting two differences between discrete wavelet shrinkage (31) and nonlinear diffusion filtering (10), namely all derivatives are presmoothed and we sum over all scales σ. In contrast to the continuous considerations, we have worked with two different kernels to allow for biorthogonal wavelets. This can lead to partial differential equations with different orders of the inner and the outer derivative.

In the PDE-based image processing context, similar ideas, but without presmoothing, have been used in the filters of Tumblin and Turk [31] and Wei [36]. They proposed to use evolution equations of the form

$$u_t \;=\; -\,\mathrm{div}\Big(g(m)\nabla\Delta u\Big)$$

where m is the squared gradient norm or the squared Frobenius norm of the Hessian matrix of u. In this respect these approaches even go one step further: They do not only allow the derivative orders in front of the nonlinear function and behind to be different, but the argument can also be a third order one, while m depends on first or second order derivatives. By the construction (31) this is not included in our framework since the argument of the diffusivity is always the same as its multiplier.

Remark 6.1 (Orthogonal Wavelets)
In the case of orthogonal wavelets, (31) simplifies to

$$\begin{aligned} u^{k+1} \;&=\; u^k \\ &-\tau \sum_{\sigma=1}^{m} \frac{1}{2^\sigma} \left(D^p\right)^T K^{H,(\sigma)}\, \Phi\left(D^p \left(K^{H,(\sigma)}\right)^T u^k\right) D^p \left(K^{H,(\sigma)}\right)^T u^k. \end{aligned} \tag{32}$$

Besides the smoothing kernels and the sum over all scales, this is identical to an explicit discretisation of a higher order nonlinear diffusion equation. Since the outer matrices are the adjoints of the inner ones, this approach can be understood as arising from an energy function of the form

$$E(u) \;=\; \sum_{i \in J} (u_i - f_i)^2 \;+\; \alpha \sum_{\sigma=1}^{m} \frac{1}{2^\sigma} \sum_{i \in J} \Psi\Big(\left(D^p K^{H,(\sigma)} u\right)_i^2 \Big) \tag{33}$$

with $\Psi'(s^2) = g(s^2)$. Continuous analoga to this equation can be found in [9, 4], for example. For biorthogonal wavelets such a formulation does not exist.

7 Generalisation to Higher Dimensions

So far, the ideas in this paper have been considered in one spatial dimension only. Let us turn to the two-dimensional case. For one single scale of Haar wavelet shrinkage, relations to nonlinear diffusion equations have been discussed by Mrázek and Weickert [24]. Here we follow the strategy sketched in [24], but apply it not only to one scale of Haar wavelet shrinkage, but to multiple ones with general biorthogonal filters.
It is common to use tensor product wavelets for the processing of two-dimensional images; see [20, Subsections 7.7.2 and 7.7.3] or [12, Section 7.5], for example. With the one-dimensional analysis scaling coefficients h_0 and wavelet coefficients h_1, the tensor product analysis filters h_s, h_h, h_v, and h_d in 2-D read as

$$\begin{aligned} h_s(i,j) \;&:=\; h_0(i)h_0(j), \qquad & h_h(i,j) \;&:=\; h_1(i)h_0(j), \\ h_v(i,j) \;&:=\; h_0(i)h_1(j), \qquad & h_d(i,j) \;&:=\; h_1(i)h_1(j). \end{aligned}$$

Here, the subscript s stands for scaling function, h for the horizontal, v for the vertical, and d for the diagonal wavelet. The same definition applies for the synthesis coefficients with g instead of h. It is a classical result that these filters on multiple scales yield a biorthogonal family in 2-D. In analogy to (26), the perfect reconstruction property for m scales in 2-D can be formulated as

$$f \;=\; \sum_{\sigma=1}^{m} \frac{1}{4^\sigma} \Big(\sum_{\delta \in \{h,v,d\}} G_\delta^{(\sigma)} \left(H_\delta^{(\sigma)}\right)^T f \Big) \;+\; \frac{1}{4^m} G_s^{(m)} \left(H_s^{(m)}\right)^T f. \tag{34}$$

Then one step of shrinkage reads as

$$u \;=\; \sum_{\sigma=1}^{m} \frac{1}{4^{\sigma}} \Big(\sum_{\delta\in\{h,v,d\}} G_{\delta}^{(\sigma)} S^{\delta}\Big(\Big(H_{\delta}^{(\sigma)}\Big)^{T} f \Big)\Big) \;+\; \frac{1}{4^{m}} G_{s}^{(m)} \left(H_{s}^{(m)}\right)^{T} f \tag{35}$$

with the shrinkage functions S^h, S^v, and S^d applied to the corresponding wavelet coefficients.
To give a motivation for using different shrinkage functions S^{δ} in the three directions, we have a look at the approximation properties of the wavelet coefficients in 2-D. Convolution of an image with the filters given above can also be understood as derivative approximation with presmoothing where the derivative order and the smoothing kernel depend on h_1 and h_0. For example, let p be the number of vanishing moments of h_1. Convolution of a discrete image u with h_h and h_v approximates presmoothed p-th derivatives of u in x- and y-direction. The filter h_d yields the approximation of the derivative $\partial_x^p \partial_y^p u$ with additional smoothing. That means this derivative in diagonal direction has twice the order than the other ones. This fact suggests to follow the shrinkage rule described in [23] to improve rotational invariance. Inspired by nonlinear diffusion filtering, it is suggested in [23] to couple the horizontal and vertical coefficients in the argument of the shrinkage function and not to shrink the diagonal ones at all. Let w_h, w_v, and w_d stand for the wavelet coefficients in horizontal, vertical and diagonal direction at a given scale and position. The corresponding shrinkage functions applied to the horizontal, vertical and diagonal coefficients can be written as:

$$\begin{aligned} S^h(w_h, w_v) &:= w_h(1 - \tau g(w_h^2 + w_v^2)), && (36)\\ S^v(w_h, w_v) &:= w_d(1 - \tau g(w_h^2 + w_v^2)), && (37)\\ S^d(w_d) &:= w_d \;. \end{aligned}$$

In contrast to [23] we avoid the additional factor 4 in front of the function g here. This factor can be explained as compensation of the factor $\frac{1}{4}$ appearing in (34) and (35) together with the fact that only the finest scale is considered in [23]. We avoid the factor here since we work on multiple scales and prefer to use the same shrinkage function on all scales.
With these shrinkage functions and the perfect reconstruction property (34), wavelet shrinkage (35) can be transformed into

$$u \;=\; f - \tau \sum_{\sigma=1}^{m} \frac{1}{4^{\sigma}} \left(G_h^{(\sigma)}\, \Phi_h^{(\sigma)} \left(H_h^{(\sigma)}\right)^{T} f + G_v^{(\sigma)}\, \Phi_v^{(\sigma)} \left(H_v^{(\sigma)}\right)^{T} f \right) \;. \tag{38}$$

Here, $\Phi_h^{(\sigma)}$ and $\Phi_v^{(\sigma)}$ represent the pointwise multiplication of the wavelet coefficients in horizontal and vertical direction on scale σ with diffusivity g in (36) and (37). Note that this diffusivity depends on the squared sum of the horizontal and vertical wavelet coefficients at the corresponding position and scale. Understood as discretisation of an integrodifferential equation, one would use it iteratively yielding

$$u^{k+1} \;=\; u^k - \tau \sum_{\sigma=1}^{m} \frac{1}{4^\sigma} \left(G_h^{(\sigma)}\, \Phi_h^{(\sigma)} \left(H_h^{(\sigma)}\right)^T u^k + G_v^{(\sigma)}\, \Phi_v^{(\sigma)} \left(H_v^{(\sigma)}\right)^T u^k \right) \;. \qquad (39)$$

This is a 2-D analogue of (31).

Example 7.1 (Orthogonal Wavelets in 2-D)
Let us consider the case of orthogonal wavelets, i. e., $G_h = H_h$ and $G_v = H_v$, with p vanishing moments. If we neglect the presmoothing introduced by the wavelets, the shrinkage process is obviously connected to a continuous equation of the form

$$\begin{aligned} \partial_t u \;=\; (-1)^{p+1} \Big(& \partial_x^p \left(g(|\partial_x^p u|^2 + |\partial_y^p u|^2)\, \partial_x^p u \right) \\ & + \partial_y^p \left(g(|\partial_x^p u|^2 + |\partial_y^p u|^2)\, \partial_y^p u \right) \Big) \end{aligned} \qquad (40)$$

which only considers the derivatives with respect to the coordinate axes. For $p = 1$, this is the classical Perona-Malik equation. For higher derivative orders $p > 1$, it only involves the derivatives of order p in coordinate directions and no mixed derivatives.

8 Numerical Experiments

In this section we want to investigate experimentally the differences between nonlinear diffusion filtering and our discrete version of the integrodifferential equations related to wavelet shrinkage described in this paper. In 1-D, we perform detailed qualitative comparisons for the denoising of a signal with additive Gaussian noise. Experiments for image simplification in 2-D show that the same effects appear for higher spatial dimensions. All implementations have been written in `C`.
Let us first describe our experiments in 1-D: Figure 8 shows our test signal `piecepoly` taken from the Wavelab library[1] and its noisy version with additive Gaussian noise of standard deviation 20.
In our first experiment, we compare the quality of presmoothed iterative denoising methods at a single scale σ given by the equation

$$u^{k+1} \;=\; u^k - \tau\, (D^p)^T\, K^{H,(\sigma)}\, \Phi^\sigma \left(D^p \left(K^{H,(\sigma)}\right)^T u^k \right) D^p \left(K^{H,(\sigma)}\right)^T u^k. \qquad (41)$$

In our experiments, we have used the order $p = 1$ and the hat function as kernel in the matrices K^H. As we have seen in Section 5, this corresponds to Haar

[1] Wavelab is available under the address http://www-stat.stanford.edu/~wavelab/.

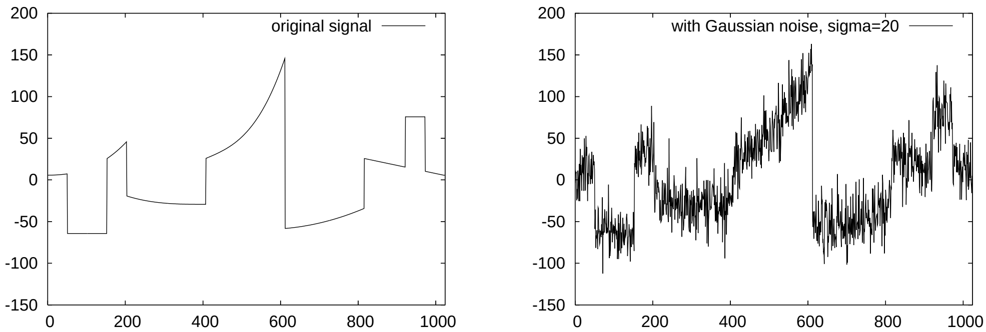

Figure 8: Test signals. *Left:* Piecewise polynomials signal with 1024 pixels. *Right:* With additive Gaussian noise, standard deviation 20.

Table 1: Error norms for denoising results with presmoothed diffusion and one single scale.

Scale σ	ℓ^1-error per pixel			ℓ^2-error per pixel		
	error	λ	iterations	error	λ	iterations
1	**2.740**	**1.02**	**4593**	**0.141**	**1.67**	**1265**
2	5.087	0.10	247000	0.227	0.10	233000
3	6.515	0.10	351000	0.285	0.10	263000

wavelets. The kernel length is $l = 2\sigma$. Moreover, we have applied the Perona-Malik diffusivity in the diagonal matrix Φ. Notice that $\sigma = 1$ corresponds to the classical diffusion filtering. We have used one single scale for presmoothing, and thus in contrast to (32), there is no sum and no weight factor on the right-hand side. The parameters have been optimised in order to obtain minimal errors in both the ℓ^1- and ℓ^2-norms. The optimal parameters and the corresponding mimimal error measures can be found in Table 1. We see that the minimal errors are obtained for classical nonlinear diffusion filters without presmoothing. To visualise the differences some of the corresponding signals are displayed in Figure 9. It is clearly visible that using single-scale presmoothing kernels for all derivatives leads to artefacts. The process is not able to remove the noise on the small scales which leads to oscillations. Only the general shape of the signal is restored for larger scales. This is in accordance with the results reported by Scherzer and Weickert [27].

In our second experiment, we do not only filter with one larger scale, but involve all dyadic scales $\sigma = 2^l$ for $l = 0, \ldots, k$ and use (32) for filtering. The corresponding optimal error measures are shown in Table 2. We have used a time step size $\tau = 1/2$. We see that involving larger scales does not influence the minimal error as severly as in the first experiment. For the ℓ^1-error, it is even possible to obtain better values by using $k = 2$. We notice that using only the finest scale requires half the number of iterations than in the first experiment: This is caused

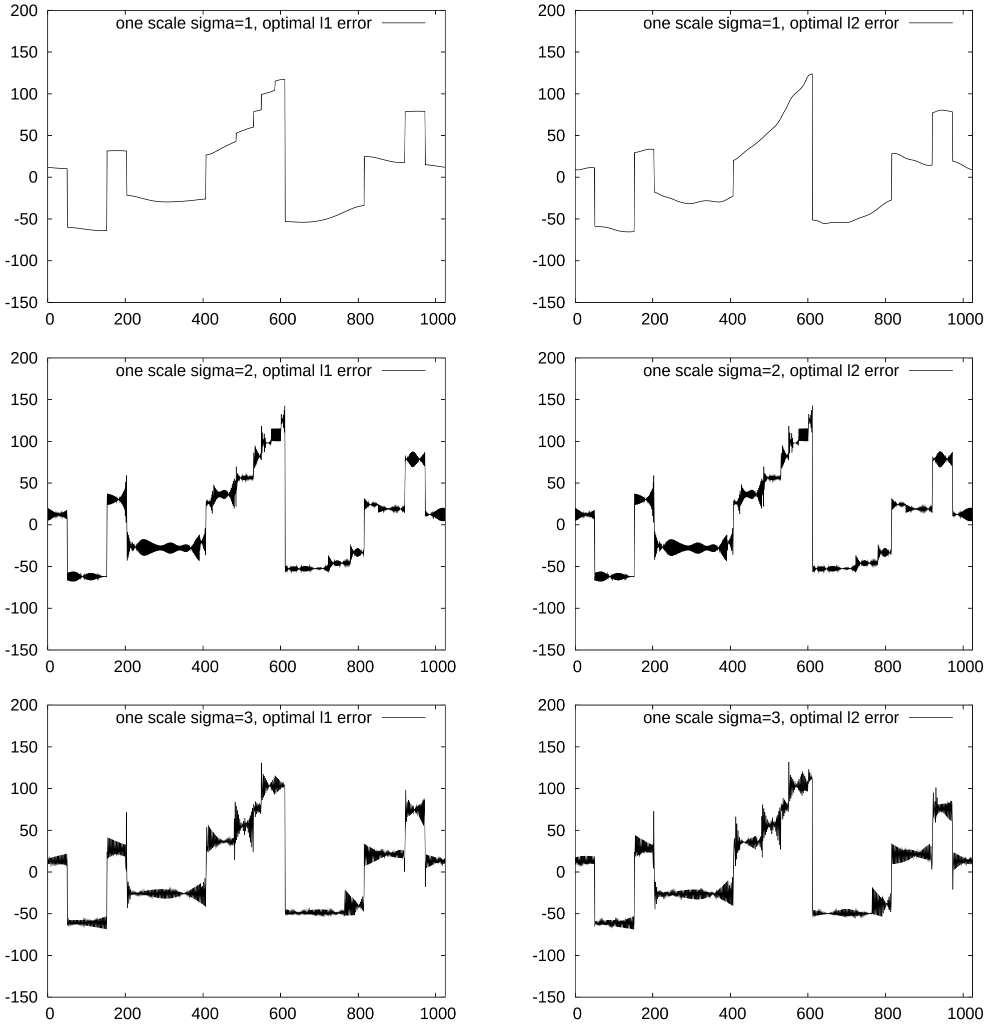

Figure 9: Denoising results with presmoothed diffusion and one single scale σ. *Left column:* Results with optimal ℓ^1-error. *Right column:* Results with optimal ℓ^2-error. *Top row:* $\sigma = 1$. *Middle row:* $\sigma = 2$. *Bottom row:* $\sigma = 3$.

Table 2: Error norms for denoising results using presmoothed diffusion on dyadic scales.

Largest scale $\sigma = 2^k$	ℓ^1-error per pixel error	λ	iterations	ℓ^2-error per pixel error	λ	iterations
$k = 0$	2.740	1.02	9197	**0.140**	**1.67**	**2604**
$k = 1$	2.824	1.47	1904	0.142	2.11	677
$k = 2$	**2.717**	**2.39**	**495**	145.03	3.57	200
$k = 3$	2.791	4.02	153	0.143	4.95	95
$k = 4$	3.000	6.36	53	0.146	5.84	61
$k = 5$	3.184	8.95	27	0.150	6.47	48

by the additional factor $\frac{1}{2}$ in (32) on the finest scale which was not present in the last experiment. The necessary number of iterations reduces by two orders of magnitude by involving larger scales. This can be understood as approximative numerical method for speeding up the process. The corresponding signals are shown in Figure 10. We see that for larger scales, some smaller artefacts appear. Nevertheless, it seems that the presence of smaller scales at the right-hand side can help to suppress most of them.
In our 2-D experiment, we also display results for smoothing on one larger scale and on all dyadic scales. For one larger scale, we use the filter

$$u^{k+1} \;=\; u^k - \tau \left(K_h^{(\sigma)} \, \Phi_h^{(\sigma)} \left(K_h^{(\sigma)} \right)^T u^k + K_v^{(\sigma)} \, \Phi_v^{(\sigma)} \left(K_v^{(\sigma)} \right)^T u^k \right) \; . \tag{42}$$

This corresponds to (39) where the factor and the sum on the right-hand side are left out. We use $p = 1$ and hat functions in the directions of the derivative and box filters in the other direction which implements tensor product Haar wavelets. Figure 11 shows the resulting images if we fix all parameters and only vary the scale. We see that using larger scales only introduces artefacts in the image which can be compared to those appearing also in the 1-D case.
For involving all scales we directly use (39). Some results for involving all dyadic scales up to a certain order are displayed in Figure 12. Here we see that more and more small details are removed by using the larger scales while the artefacts are suppressed.

9 Summary

In this paper, we have investigated the relation between discrete multiscale wavelet shrinkage on the one hand and discretised nonlinear diffusion filters of arbitrary order and their variational counterparts on the other hand. To this end we exploited the fact that the wavelet transform using wavelets with a finite number of vanishing moments represents smoothed derivative operators. The resulting

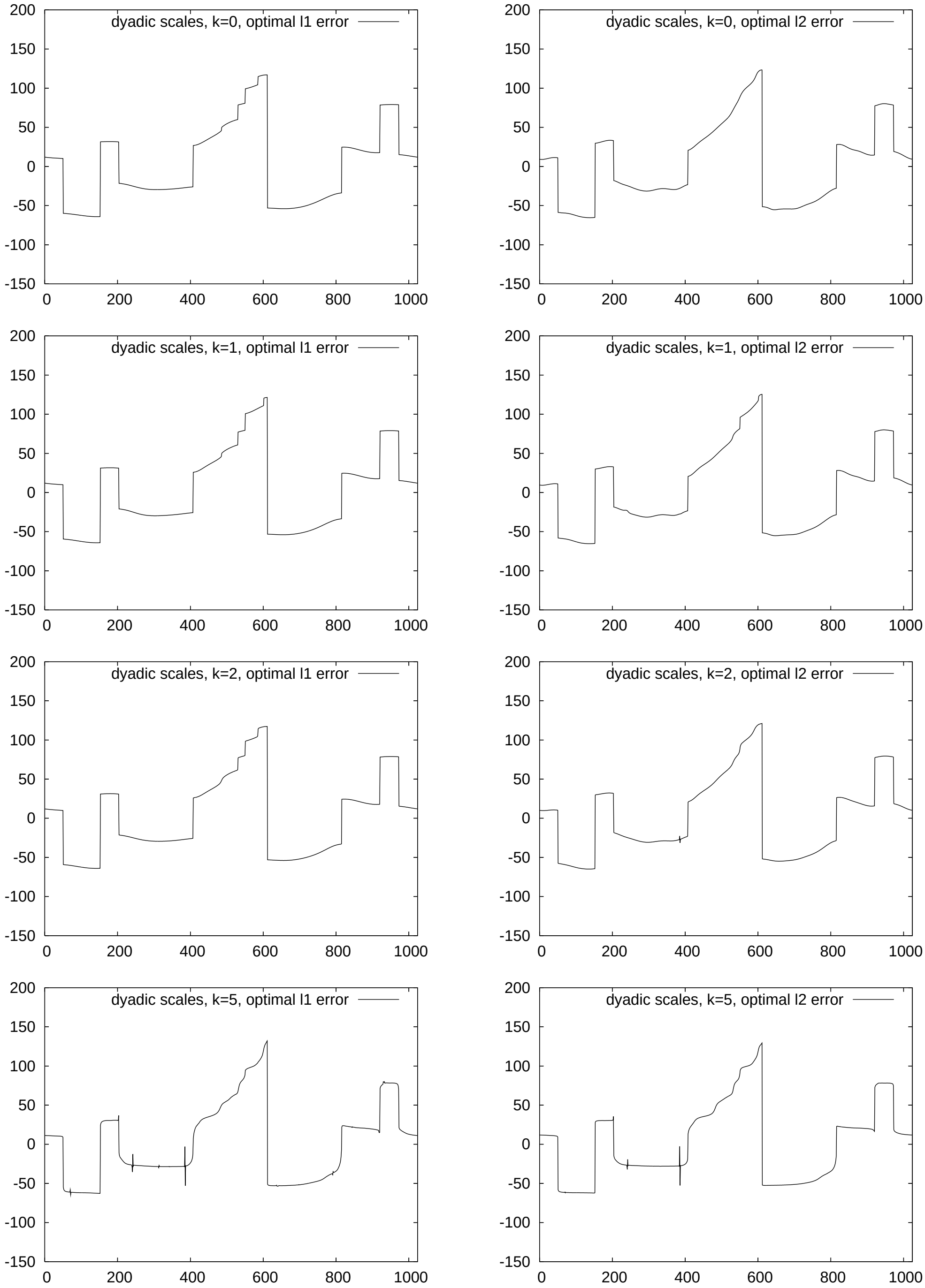

Figure 10: Denoising results with presmoothed diffusion and dyadic scale up to $\sigma = 2^k$. *Left column:* Results with optimal ℓ^1-error. *Right column:* Results with optimal ℓ^2-error. *Top row:* $k = 0$. *Second row:* $k = 1$. *Third row:* $k = 2$. *Bottom row:* $k = 5$.

Figure 11: Image simplification results with presmoothed diffusion and one single scale σ, Perona-Malik diffusivity $g(s^2) = 1/(1+s^2/\lambda^2)$ with $\lambda = 10$, and stopping time $t = 5$. *Top left:* Original image, 512×512 pixels. *Top right:* $\sigma = 1$. *Bottom left:* $\sigma = 2$. *Bottom right:* $\sigma = 3$.

Figure 12: Image simplification results with presmoothed diffusion, $g(s^2) = (1 + s^2/\lambda^2)^{-1}$ for $\lambda = 10$, stopping time $t = 20$, and dyadic scales up to $\sigma = 2^k$. *Top left:* Original image, 512×512 pixels. *Top right:* $k = 0$. *Bottom left:* $k = 1$. *Bottom right:* $k = 5$.

discrete integrodifferential equations differ from their nonlinear diffusion counterparts by the additional presmoothing of derivatives and integration over a larger number of scales. The shape of the corresponding convolution kernels changes for coarser scales in the discrete setting due to sampling. We have extended the considerations from orthogonal to biorthogonal wavelets: Here, the corresponding discrete versions of integrodifferential equations are no longer related to diffusion equations, but to more general PDE models like the methods by Tumblin and Turk [31] or Wei [36]. Using tensor product wavelets and special shrinkage rules to improve the rotation invariance, the relations have been carried over to the 2-D setting. Numerical experiments have shown that presmoothed nonlinear diffusion on one single larger scale gives worse results than classical nonlinear diffusion. However, involving all dyadic scales up to a certain order, as it is done in wavelet shrinkage, almost keeps the good quality and significantly reduces the number of required iterations. In this sense, discrete multiscale wavelet shrinkage can be understood as numerical method for discrete integrodifferential equations.

Acknowledgements

We gratefully acknowledge partly funding by the *Deutsche Forschungsgemeinschaft (DFG)*, project WE 2602/2-3.

References

[1] Y. Bao and H. Krim, *Towards bridging scale-space and multiscale frame analyses*, in Wavelets in Signal and Image Analysis, A. A. Petrosian and F. G. Meyer, eds., vol. 19 of Computational Imaging and Vision, Kluwer, Dordrecht, 2001, ch. 6, pp. 169–192.

[2] K. Bredies, D. A. Lorenz, and P. Maass, *Mathematical concepts of multiscale smoothing*, Applied and Computational Harmonic Analysis, 19 (2005), pp. 141–161.

[3] A. Chambolle, R. DeVore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Transactions on Image Processing, 7 (1998), pp. 319–335.

[4] A. Chambolle and B. L. Lucier, *Interpreting translation-invariant wavelet shrinkage as a new image smoothing scale space*, IEEE Transactions on Image Processing, 10 (2001), pp. 993–1000.

[5] R. E. Crochiere and L. R. Rabiner, *Digital coding of speech in subbands*, Bell Systems Technical Journal, 55 (1976), pp. 1069–1085.

[6] A. Croisier, D. Esteban, and C. Galand, *Perfect channel splitting by use of interpolation / decimation / tree decomposition techniques*, in International Conference on Information Sciences and Systems, Patras, Greece, Aug. 1976, pp. 443–446.

[7] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, 1992.

[8] S. Didas, G. Steidl, and J. Weickert, *Discrete multiscale wavelet shrinkage and integrodifferential equations*, in Optical and Digital Image Processing – Photonics Europe, vol. 7000 of Proceedings of SPIE, 2008, in print.

[9] S. Didas and J. Weickert, *Integrodifferential equations for continuous multiscale wavelet shrinkage*, Inverse Problems and Imaging, 1 (2007), pp. 47–62.

[10] S. Didas, J. Weickert, and B. Burgeth, *Stability and local feature enhancement of higher order nonlinear diffusion filtering*, in Pattern Recognition, W. Kropatsch, R. Sablatnig, and A. Hanbury, eds., vol. 3663 of Lecture Notes in Computer Science, Springer, 2005, pp. 451–458.

[11] D. L. Donoho and I. M. Johnstone, *Ideal spatial adaption by wavelet shrinkage*, Biometrica, 81 (1994), pp. 425–455.

[12] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Prentice Hall, second ed., 2002.

[13] 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, J. M. Combes, A. Grossmann, and P. Tchamitchian, eds., Springer, Berlin, 1989, pp. 286–297.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[15] T. Iijima, *Basic theory on normalization of pattern (in case of typical one-dimensional pattern)*, Bulletin of the Electrotechnical Laboratory, 26 (1962), pp. 368–388. (In Japanese).

[16] A. K. Louis, P. Maass, and A. Rieder, *Wavelets*, B. G. Teubner Stuttgart, second ed., 1998.

[17] M. Lysaker, A. Lundervold, and X.-C. Tai, *Noise removal using fourth-order partial differential equation with applications to medical magnetic resonance images in space and time*, IEEE Transactions on Image Processing, 12 (2003), pp. 1579–1590.

[18] S. Mallat, *Multiresolution approximations and wavelet orthonormal bases of* $\mathscr{L}^2(\mathbb{R})$, Transactions of the Americal Mathematical Society, 315 (1989), pp. 69–87.

[19] ——, *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11 (1989), pp. 674–693.

[20] ——, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, second ed., 1999.

[21] Y. Meyer, *Wavelets and Operators*, vol. 37 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, 1992.

[22] F. Mintzer, *Filters for distrotion-free two-band multirate filter banks*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 33 (1985), pp. 626–630.

[23] P. Mrázek and J. Weickert, *Rotationally invariant wavelet shrinkage*, in Pattern Recognition, B. Michaelis and G. Krell, eds., vol. 2781 of Lecture Notes in Computer Science, Springer, Berlin, 2003, pp. 156–163.

[24] ——, *From two-dimensional nonlinear diffusion to coupled Haar wavelet shrinkage*, Journal of Visual Communication and Image Representation, 18 (2007), pp. 162–175.

[25] P. Mrázek, J. Weickert, and G. Steidl, *Diffusion-inspired shrinkage functions and stability results for wavelet denoising*, International Journal of Computer Vision, 64 (2005), pp. 171–186.

[26] P. Perona and J. Malik, *Scale space and edge detection using anisotropic diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12 (1990), pp. 629–639.

[27] O. Scherzer and J. Weickert, *Relations between regularization and diffusion filtering*, Journal of Mathematical Imaging and Vision, 12 (2000), pp. 43–63.

[28] M. Smith and T. Barnwell III, *Exact reconstruction techniques for tree-structures subband coders*, IEEE Transactions on Acoustics, Speech, and Signal Processing, 34 (1986), pp. 434–441.

[29] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs*, SIAM Journal on Numerical Analysis, 42 (2004), pp. 686–713.

[30] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, 1997.

[31] J. Tumblin and G. Turk, *LCIS: A boundary hierarchy for detail-preserving contrast reduction*, in SIGGRAPH '99: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, ACM Press/Addison-Wesley Publishing Co., 1999, pp. 83–90.

[32] M. Vetterli, *Multidimensional subband coding: some theory and algorithms*, Signal Processing, 6 (1984), pp. 97–112.

[33] ——, *Filter banks allowing perfect reconstruction*, Signal Processing, 10 (1986), pp. 219–244.

[34] M. Vetterli and J. Kovačevic, *Wavelets and Subband Coding*, Prentice-Hall, Upper Saddle River, 1995.

[35] J. B. Weaver, Y. Xu, D. M. Healy, and L. D. Cromwell, *Filtering noise from images with wavelet transforms*, Magnetic Resonance in Medicine, 21 (1991), pp. 288–295.

[36] G. W. Wei, *Generalized Perona-Malik equation for image restoration*, IEEE Signal Processing Letters, 6 (1999), pp. 165–167.

[37] J. Weickert, *Anisotropic Diffusion in Image Processing*, B. G. Teubner, Stuttgart, 1998.

[38] J. Weickert, G. Steidl, P. Mrázek, M. Welk, and T. Brox, *Diffusion filters and wavelets: What can they learn from each other?*, in Handbook of Mathematical Models in Computer Vision, N. Paragios, Y. Chen, and O. Faugeras, eds., Springer, New York, 2006, pp. 3–16.

[39] A. P. Witkin, *Scale-space filtering*, in Proc. Eighth International Joint Conference on Artificial Intelligence, vol. 2, Karlsruhe, Germany, August 1983, pp. 945–951.