

## ON THE EQUIVALENCE OF SOFT WAVELET SHRINKAGE, TOTAL VARIATION DIFFUSION, TOTAL VARIATION REGULARIZATION, AND SIDES\*

GABRIELE STEIDL<sup>†</sup>, JOACHIM WEICKERT<sup>‡</sup>, THOMAS BROX<sup>‡</sup>, PAVEL MRÁZEK<sup>‡</sup>,  
AND MARTIN WELK<sup>‡</sup>

**Abstract.** Soft wavelet shrinkage, total variation (TV) diffusion, TV regularization, and a dynamical system called SIDEs are four useful techniques for discontinuity preserving denoising of signals and images. In this paper we investigate under which circumstances these methods are equivalent in the one-dimensional case. First, we prove that Haar wavelet shrinkage on a single scale is equivalent to a single step of space-discrete TV diffusion or regularization of two-pixel pairs. In the translationally invariant case we show that applying cycle spinning to Haar wavelet shrinkage on a single scale can be regarded as an absolutely stable explicit discretization of TV diffusion. We prove that space-discrete TV diffusion and TV regularization are identical and that they are also equivalent to the SIDEs system when a specific force function is chosen. Afterwards, we show that wavelet shrinkage on multiple scales can be regarded as a single step diffusion filtering or regularization of the Laplacian pyramid of the signal. We analyze possibilities to avoid Gibbs-like artifacts for multiscale Haar wavelet shrinkage by scaling the thresholds. Finally, we present experiments where hybrid methods are designed that combine the advantages of wavelets and PDE/variational approaches. These methods are based on iterated shift-invariant wavelet shrinkage at multiple scales with scaled thresholds.

**Key words.** soft wavelet shrinkage, total variation diffusion, total variation regularization, stabilized inverse diffusion equation

**AMS subject classifications.** 94A08, 65M06, 65T60, 49-99, 94A12

**DOI.** 10.1137/S0036142903422429

**1. Introduction.** Image denoising is a field where one is typically interested in removing noise without sacrificing important structures such as edges. This goal cannot be achieved with linear filters. Consequently, a large variety of nonlinear strategies has been proposed including, among others, wavelet techniques [22, 23, 31], PDEs [2, 37, 41, 47], and variational methods [5, 6, 9, 36].

Although these method classes serve the same purpose, relatively few publications exist where their similarities and differences are juxtaposed and their mutual relations are analyzed. However, such an analysis is highly desirable, since it can help to transfer results from one of these classes to the others. Moreover, a deeper understanding of the differences between these classes might be helpful for designing novel hybrid methods that combine the advantages of the different classes.

The goal of the present paper is to address this problem by analyzing relations between four important representatives of discontinuity-preserving denoising methods:

- wavelet soft thresholding [22],
- space-discrete total variation (TV) diffusion [3, 4],

---

\*Received by the editors February 4, 2003; accepted for publication (in revised form) September 16, 2003; published electronically May 17, 2004. This research was partially funded by DFG projects We 2602/2-1 and We 2602/1-1/So 363/9-1.

<http://www.siam.org/journals/sinum/42-2/42242.html>

<sup>†</sup>Faculty of Mathematics and Computer Science, University of Mannheim, 68131 Mannheim, Germany (steidl@math.uni-mannheim.de, <http://kiwi.math.uni-mannheim.de>).

<sup>‡</sup>Mathematical Image Analysis Group, Faculty of Mathematics and Computer Science, Saarland University, Building 27, 66041 Saarbrücken, Germany (weickert@mia.uni-saarland.de, brox@mia.uni-saarland.de, mrazek@mia.uni-saarland.de, welk@mia.uni-saarland.de; <http://www.mia.uni-saarland.de>).

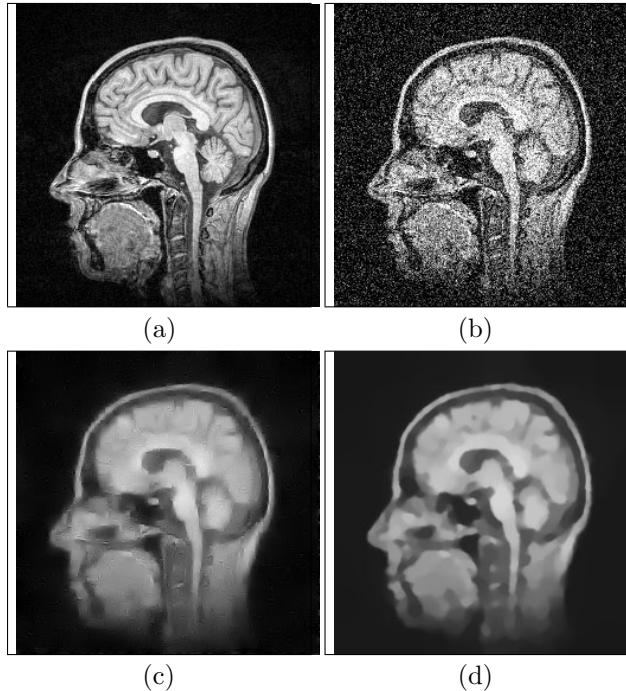


FIG. 1. (a) *Top left: original magnetic resonance image.* (b) *Top right: magnetic resonance image degraded with Gaussian noise with standard deviation 50.* (c) *Bottom left: wavelet denoising of (b) using translation-invariant soft shrinkage with Haar wavelets.* (d) *Bottom right: TV diffusion of (b).*

- discrete TV regularization [40, 1],
- SIDes, a dynamical system that has been inspired from space-discrete stabilized inverse diffusion equations [38].

Figure 1 gives an illustration of the denoising properties of soft wavelet shrinkage and TV regularization methods. The original image is available from [31]. We observe that the results do not differ very much. Indeed, we shall prove in our paper that all four aforementioned methods are very closely related.

In order to keep things as simple as possible we base our analysis on the one-dimensional (1-D) case. Our basic strategy is to start with the simplest cases for which we can establish equivalence. Afterwards, we extend these results to more general situations. The higher-dimensional case is beyond the scope of the present paper, since it cannot be treated as a straightforward generalization of the 1-D ideas. For some preliminary results in two dimensions, we refer the reader to [34], where diffusion-inspired wavelet shrinkage with improved rotation invariance is introduced.

Our paper is organized as follows. In section 2 we give a very brief description of the general ideas behind wavelet shrinkage, nonlinear diffusion filtering, variational image denoising, and SIDes. In section 3 we specify these paradigms to the simplest cases where equivalence can be shown. In this section we restrict ourselves to two-pixel signals, soft Haar wavelet shrinkage, TV diffusivity, and its corresponding regularizer. Under these circumstances we prove equivalence between wavelet shrinkage, TV diffusion, and TV regularization. These results are extended in section 4 to the translationally invariant case with  $N$ -pixel signals. In the wavelet setting, we use

a Haar wavelet-based technique on a single scale with cycle spinning. We show that it can be regarded as a single iteration of a stabilized explicit scheme for TV diffusion, and we prove that this TV diffusion is equivalent to both TV regularization and SIDes with an appropriate force function. In section 5 we extend our wavelet results from a single scale to multiple scales. We show that multiple scale Haar wavelet soft shrinkage can be regarded as TV diffusion, TV regularization, or SIDes applied to a Laplacian pyramid decomposition of the signal. Moreover, we propose and analyze a strategy for avoiding Gibbs-like artifacts by scaling the shrinkage thresholds. In section 6 we present experiments where we compare iterated single-scale filtering with noniterated and iterated multiscale filtering. The paper is concluded with a summary in section 7.

*Related work.* Recently, a number of interesting connections between wavelet shrinkage of functions, regularization methods, and PDEs has been established. A book by Meyer [33] presents a unified view on wavelets and nonlinear evolutions, and Shen and Strang [43] have included wavelets into the solution of the linear heat equation. Chambolle et al. [13] showed that one may interpret wavelet shrinkage of functions as regularization processes in suitable Besov spaces. In particular, Haar thresholding was considered in [18]. Furthermore, Cohen et al. [17] showed that the space of functions of bounded variation can be “almost” characterized by wavelet expansions. Chambolle and Lucier [15] considered iterated translationally invariant wavelet shrinkage and interpreted it as a nonlinear scale-space, which differs from other scale-spaces by the fact that it is not given in terms of PDEs.

There has also been a rapidly increasing interest in designing hybrid methods using both wavelet shrinkage and TV denoising methods. Durand and Froment [24] proposed to address the problem of pseudo-Gibbs artifacts in wavelet denoising by replacing the thresholded wavelet coefficients by coefficients that minimize the total variation. Their method is also close in spirit to approaches by Chan and Zhou [16], who postprocessed images obtained from wavelet shrinkage by a TV-like regularization technique. Coifman and Sowa [20] used functional minimization with wavelet constraints for postprocessing signals that have been degraded by wavelet thresholding or quantization. Candés and Guo [12] also presented related work, in which they combined ridgelets and curvelets with TV minimization strategies. Recently, Malgouyres [30] proposed a hybrid method that uses both wavelet packets and TV approaches. His experiments showed that it may restore textured regions without introducing ringing artifacts.

Regarding the relations between wavelet shrinkage denoising of discrete signals and TV reduction, not much research has been done so far. One notable exception is a recent paper by Coifman and Sowa [21], where they propose TV diminishing flows that act along the direction of Haar wavelets. Bao and Krim [7] addressed the problem of texture loss in diffusion scale-spaces by incorporating ideas from wavelet analysis. An experimental evaluation of the denoising capabilities of three-dimensional wavelet shrinkage and nonlinear diffusion filters is presented in a paper by Frangakis, Stoschek, and Hegerl [27].

This discussion shows that our paper differs from preceding work by the fact that we investigate conditions under which we can prove equivalence between wavelet shrinkage of discrete signals, space-discrete TV diffusion or regularization, and SIDes. Some preliminary results in this paper have been presented at conferences [44, 10].

**2. The basic methods.** The goal of this section is to give a brief introduction to the methods that are considered in this paper: soft Haar wavelet shrinkage, TV diffusion, TV denoising, and SIDes.

**2.1. Wavelet shrinkage.** During recent years wavelet methods have proved their use in various signal processing tasks. One of them is discontinuity-preserving denoising. The discrete wavelet transform represents a 1-D signal  $f(x)$  in terms of shifted versions of a dilated lowpass scaling function  $\varphi(x)$ , as well as shifted and dilated versions of a bandpass wavelet function  $\psi(x)$ . In the case of orthogonal wavelets, this gives

$$f(x) = \sum_{i \in \mathbb{Z}} \langle f, \varphi_i^{j_e} \rangle \varphi_i^{j_e}(x) + \sum_{j=-\infty}^{j_e} \sum_{i \in \mathbb{Z}} \langle f, \psi_i^j \rangle \psi_i^j(x),$$

where  $\psi_i^j(x) := 2^{-j/2} \psi(2^{-j}x - i)$  and where  $\langle \cdot, \cdot \rangle$  denotes the inner product in  $L_2(\mathbb{R})$ . The wavelet representation employs scaling components only at one level  $j_e$ , and wavelet components at levels  $j \leq j_e$  add higher resolution details to the signal.

If the measurements  $f$  are corrupted by white Gaussian noise, then this noise is contained to a small amount in all wavelet coefficients  $\langle f, \psi_i^j \rangle$ , while the original signal is in general determined by few significant wavelet coefficients. Therefore wavelet shrinkage attempts to eliminate noise from the wavelet coefficients by the following three-step procedure:

- *Analysis.* Transform the noisy data  $f$  to the wavelet coefficients  $d_i^j = \langle f, \psi_i^j \rangle$  and scaling function coefficients  $c_i^{j_e} = \langle f, \varphi_i^{j_e} \rangle$ .
- *Shrinkage.* Apply a shrinkage function  $S_\tau$  with a threshold parameter  $\tau$  related to the variance of the Gaussian noise to the wavelet coefficients, i.e.,  $S_\tau(d_i^j) = S_\tau(\langle f, \psi_i^j \rangle)$ .
- *Synthesis.* Reconstruct the denoised version  $u$  of  $f$  from the shrunken wavelet coefficients

$$u(x) := \sum_{i \in \mathbb{Z}} \langle f, \varphi_i^{j_e} \rangle \varphi_i^{j_e}(x) + \sum_{j=-\infty}^{j_e} \sum_{i \in \mathbb{Z}} S_\tau(\langle f, \psi_i^j \rangle) \psi_i^j(x).$$

In the literature a number of different shrinkage functions have been considered. In this paper we focus on one of the most popular strategies, namely Donoho's soft shrinkage [22]. It uses the soft thresholding with threshold parameter  $\tau > 0$ :

$$(2.1) \quad S_\tau(x) = \begin{cases} x - \tau \operatorname{sgn}(x) & \text{if } |x| > \tau, \\ 0 & \text{if } |x| \leq \tau, \end{cases}$$

which shrinks all coefficients towards zero. Other shrinkage functions will be considered in a forthcoming paper.

Furthermore, in this paper we restrict our attention to Haar wavelets. They are well suited for recovering piecewise constant signals with discontinuities. The Haar wavelet  $\psi(x)$  and the corresponding scaling function  $\varphi(x)$  are given by  $\psi(x) := 1_{[0, \frac{1}{2})} - 1_{[\frac{1}{2}, 1)}$  and  $\varphi(x) := 1_{[0, 1)}$ , where  $1_{[a, b)}$  is the characteristic function of  $[a, b)$ :

$$1_{[a, b)}(x) := \begin{cases} 1 & \text{if } x \in [a, b), \\ 0 & \text{else.} \end{cases}$$

Using the so-called two-scale relation of the wavelet and its scaling function, the coefficients  $c_i^j$  and  $d_i^j$  at higher level  $j$  can be computed from the coefficients  $c_i^{j-1}$  at lower level  $j-1$  and vice versa. This results in fast algorithms for the analysis step

and the synthesis step. For the Haar wavelets, we obtain

$$(2.2) \quad c_i^j = \frac{c_{2i}^{j-1} + c_{2i+1}^{j-1}}{\sqrt{2}}, \quad d_i^j = \frac{c_{2i}^{j-1} - c_{2i+1}^{j-1}}{\sqrt{2}},$$

$$(2.3) \quad c_{2i}^{j-1} = \frac{c_i^j + d_i^j}{\sqrt{2}}, \quad c_{2i+1}^{j-1} = \frac{c_i^j - d_i^j}{\sqrt{2}}.$$

**2.2. Diffusion filtering.** Let us now consider a function  $f(x)$  on some interval  $[a, b]$ . The basic idea behind nonlinear diffusion filtering is to obtain a family  $u(x, t)$  of filtered versions of the signal  $f(x)$  as the solution of a suitable diffusion process with  $f(x)$  as initial condition and homogeneous Neumann boundary conditions [37]:

$$(2.4) \quad \begin{aligned} u_t &= (g(u_x^2) u_x)_x && \text{on } (a, b) \times (0, \infty), \\ u(x, 0) &= f(x) && \text{for all } x \in [a, b], \\ u_x(a, t) &= u_x(b, t) = 0 && \text{for all } t \in (0, \infty), \end{aligned}$$

where subscripts denote partial derivatives, and the diffusion time  $t$  is a simplification parameter: larger values correspond to stronger filtering.

The diffusivity  $g(u_x^2)$  is a nonnegative function that steers the amount of diffusion. Usually, it is decreasing in  $u_x^2$ . This ensures that strong edges are less blurred by the diffusion filter than noise and low-contrast details. In the present paper, we focus on the *TV diffusivity*

$$(2.5) \quad g(u_x^2) := \frac{1}{|u_x|}.$$

The resulting *TV diffusion filter* (also called *TV flow*) has a number of interesting properties. It requires no additional parameters (besides  $t$ ), it is well posed [3, 8, 25], it preserves the shape of some objects [8], and it leads to constant signals in finite time [4].

**2.3. Regularization methods.** Regularization methods constitute an alternative to diffusion filters when one is interested in a discontinuity-preserving denoising method for a continuous signal  $f(x)$  with  $x \in [a, b]$ . Here the basic idea is to look for the minimizer  $u$  of the energy functional

$$(2.6) \quad E(u; \alpha, f) := \int_a^b \left( (u - f)^2 + \alpha \Psi(u_x^2) \right) dx.$$

The first term of this functional encourages similarity between the original signal  $f(x)$  and its filtered version  $u(x)$ , while the second term penalizes deviations from smoothness. The increasing function  $\Psi$  is called the *penalizer (regularizer)*, and the nonnegative *regularization parameter*  $\alpha$  serves as smoothness weight: larger values correspond to a more pronounced filtering.

As is explained in detail in [42], there are strong relations between regularization methods and diffusion filters: A minimizer of (2.6) satisfies necessarily the Euler–Lagrange equation

$$\frac{u - f}{\alpha} = (\Psi'(u_x^2) u_x)_x,$$

with homogeneous Neumann boundary conditions. This equation may be regarded as a fully implicit time discretization of the diffusion equation (2.4) with diffusivity

$g(u_x^2) = \Psi'(u_x^2)$ , initial value  $f(x)$ , and stopping time  $\alpha$ . Thus, one would expect that the minimizer of (2.6) *approximates* the diffusion filter (2.4) but is not identical to it.

In the present paper, we are interested in one of the most popular nonlinear regularization methods, namely *TV regularization* [40, 1]. It uses the penalizer  $\Psi(u_x^2) := 2|u_x|$ , which corresponds to the TV diffusivity (2.5). This regularization is well known for its good denoising capabilities and its tendency to create blocky, segmentation-like results. Well-posedness results can be found in [14].

**2.4. SIDEs.** A SIDE is a dynamical system that has been inspired from a stabilized limiting case of a space-discrete nonlinear diffusion filter [38]. The name SIDE is an acronym for *stabilized inverse diffusion equation*.

Let us consider a discrete signal  $f = (f_i)_{i=0}^{N-1}$ . Then its SIDE evolution produces a sequence of filtered images  $u(t) = (u_i(t))_{i=0}^{N-1}$ , with  $u(0) = f$ . Increasing the time  $t$  leads to a consecutive merging of regions. The evolution between two merging events is governed by a dynamical system with a discontinuous right-hand side.

Assume that at some time  $t_j$  a pixel with index  $i$  belongs to a constant region of size  $m_{i,t_j}$ ; i.e., there exist  $l \geq 1$  and  $r \geq 0$  with  $m_{i,t_j} = l + r$ ,

$$u_{i-l+1} = \dots = u_i = u_{i+1} = \dots = u_{i+r},$$

$$u_{i-l} \neq u_{i-l+1} \text{ if } i-l \geq 0, \quad u_{i+r} \neq u_{i+r+1} \text{ if } i+r \leq N-2.$$

Then the SIDEs algorithm proceeds as follows:

(i) *Initialization.* Start at time  $t_0 = 0$  with the trivial segmentation, where each pixel  $i$  is regarded as a region of size  $m_{i,0} = 1$ :

$$u_i(0) = f_i.$$

(ii) *Evolution.* Given a segmentation at time  $t_j$ , the signal evolves according to

$$(2.7) \quad \dot{u}_i = \begin{cases} \frac{1}{m_{i,t_j}} F(u_{i+r+1} - u_{i+r}) & \text{if } i-l = -1, \\ \frac{-1}{m_{i,t_j}} F(u_{i-l+1} - u_{i-l}) & \text{if } i+r = N-1, \\ \frac{1}{m_{i,t_j}} (F(u_{i+r+1} - u_{i+r}) - F(u_{i-l+1} - u_{i-l})) & \text{else,} \end{cases}$$

where  $\dot{u}_i$  denotes the derivative of  $u_i$  with respect to  $t$  and  $F$  is a so-called *force function* that satisfies a number of formal requirements [38]. The first case in (2.7) describes the evolution of the region at the left signal boundary, the second case applies for the right boundary region, and the third case specifies the evolution of all inner regions.

In [38], only the third case has been specified. We have supplemented the other two cases here in order to be able to treat the boundary regions in a proper way as well. The evolution is stopped when two neighboring regions attain equal grey values. This determines the new merging time  $t_{j+1}$ .

(iii) *Merging.* Merge the neighboring regions with equal grey values.

(iv) *Loop control.* Stop if all regions are merged to one; else go back to step (ii).

We see that the stabilization in SIDEs is achieved by an additional definition that results in merging neighboring regions when they approach each other. This step is crucial for the performance of SIDEs, as it can be used for reducing the state variables of the dynamical system. The analytical solutions in the following sections will provide further theoretical justification for this region-merging step.

In [38] several theoretical results for SIDEs are proved, including a maximum principle, well-posedness properties, and a finite extinction time.

The dynamic system suggests that for the specific case  $m_i = 1$  one may regard a 1-D SIDE as a space discretization of the PDE

$$u_t = (F(u_x))_x$$

with homogeneous Neumann boundary conditions. This is a diffusion equation with flux function  $F$ . Since we are specifically interested in the TV case, we do not consider the specific choice in [38] but restrict ourselves to the *TV force function*

$$F(v) := \begin{cases} 1 & \text{if } v > 0, \\ -1 & \text{if } v < 0. \end{cases}$$

Then it is evident that if  $m_i = 1$  for all  $i$ , TV diffusion is approximated.

**3. Two-pixel signals.** In this section, we analyze relations between soft wavelet shrinkage, TV diffusion, TV regularization, and SIDEs for the simplest signals, namely discrete signals with only two pixels. We will see that the restriction to two pixels allows us to find analytical solutions for these degenerated nonlinear processes.

**3.1. Soft Haar wavelet shrinkage of two-pixel signals.** Let us now consider a discrete two-pixel signal  $f = (f_0, f_1)$  and study its change under soft Haar wavelet shrinkage. The analysis step produces the coefficients  $c = (f_0 + f_1)/\sqrt{2}$  and  $d = (f_0 - f_1)/\sqrt{2}$  of the scaling function and the wavelet. For simplicity, we have dropped the sub- and superscripts for  $c$  and  $d$ . This step is followed by the shrinkage operation  $S_\tau(d)$  with the soft shrinkage function (2.1). Then the synthesis step  $u_0 = (c + S_\tau(d))/\sqrt{2}$ ,  $u_1 = (c - S_\tau(d))/\sqrt{2}$  gives the final result:

$$(3.1) \quad u_0 = \begin{cases} f_0 + \frac{\tau}{\sqrt{2}} \operatorname{sgn}(f_1 - f_0) & \text{if } \tau < |f_1 - f_0|/\sqrt{2}, \\ (f_0 + f_1)/2 & \text{else,} \end{cases}$$

$$(3.2) \quad u_1 = \begin{cases} f_1 - \frac{\tau}{\sqrt{2}} \operatorname{sgn}(f_1 - f_0) & \text{if } \tau < |f_1 - f_0|/\sqrt{2}, \\ (f_0 + f_1)/2 & \text{else.} \end{cases}$$

This shows that by increasing the shrinkage threshold  $\tau$  the grey values of both pixels approach each other. For  $\tau = |f_1 - f_0|/\sqrt{2}$  they merge, and for larger  $\tau$  they remain merged.

**3.2. TV diffusion of two-pixel signals.** Next, we are interested in the space-discrete diffusion of two-pixel signals  $(f_0, f_1)$ . The homogeneous Neumann boundary conditions are discretized by setting flows over the signal boundary to zero. In this case a space-discrete version of the TV diffusion equation

$$u_t = \left( \frac{u_x}{|u_x|} \right)_x$$

can be written as

$$(3.3) \quad \dot{u}_0 = \frac{u_1 - u_0}{|u_1 - u_0|}, \quad \dot{u}_1 = -\frac{u_1 - u_0}{|u_1 - u_0|},$$

with initial conditions  $u_0(0) = f_0$  and  $u_1(0) = f_1$ . Here the dot denotes again temporal differentiation, and the pixel size is set to 1. Setting  $w(t) := u_1(t) - u_0(t)$  and

$\eta := f_1 - f_0$ , and subtracting  $\dot{u}_0$  from  $\dot{u}_1$  in (3.3), we obtain the following initial value problem:

$$(3.4) \quad \dot{w} = -2 \frac{w}{|w|}, \quad w(0) = \eta.$$

The right-hand side of this differential equation is discontinuous for  $w = 0$  and thus requires a generalization of the concept of solution. We say that  $w$  is a solution of (3.4) if it is an absolutely continuous function which fulfills

$$(3.5) \quad \dot{w} = -2 \operatorname{sgn}(w), \quad w(0) = \eta$$

almost everywhere, where

$$(I) \quad \operatorname{sgn}(w) := 1 \text{ if } w > 0,$$

$$\operatorname{sgn}(w) := -1 \text{ if } w < 0$$

and may take any value in  $[-1, 1]$  if  $w = 0$ .

Note that this definition is in agreement with the frequently used concept of differential inclusions for differential equations with discontinuous right-hand sides [26], where absolutely continuous solutions of

$$-\frac{1}{2} \dot{w} \in \begin{cases} \{1\} & \text{if } w > 0, \\ \{-1\} & \text{if } w < 0, \\ [-1, 1] & \text{if } w = 0 \end{cases}$$

were considered. The solution of (3.5) can be obtained as follows: If  $\eta \neq 0$ , then we have by straightforward computation for  $t < |\eta|/2$  that  $w(t) = \eta - 2t \operatorname{sgn}(\eta)$  and in particular, by continuity of  $w$ , that  $w(|\eta|/2) = 0$ . Assume that  $w(t) \neq 0$  for some  $t > |\eta|/2$ . Let without loss of generality  $w(t) > 0$ . The opposite assumption  $w(t) < 0$  can be handled in the same way. Then  $w(t) = -2t + C$ , where we get by continuity of  $w$ , if  $t$  approaches  $|\eta|/2$ , that  $C = |\eta|$  and, consequently,  $w(t) = 2(|\eta|/2 - t) < 0$ . This contradicts our assumption. Thus  $w(t) = 0$  for  $t \geq |\eta|/2$ . In summary, we obtain the solution

$$(3.6) \quad w(t) = \begin{cases} \eta - 2t \operatorname{sgn}(\eta) & \text{if } t < |\eta|/2, \\ 0 & \text{if } t \geq |\eta|/2. \end{cases}$$

This equation shows that the grey value difference  $w(t) = u_1(t) - u_0(t)$  tends linearly to 0. Both pixels merge at time  $t = |f_1 - f_0|/2$  and remain merged afterwards. Thus, already the simple two-pixel model indicates a finite extinction time for TV diffusion. Since  $\dot{u}_0 + \dot{u}_1 = 0$  and  $u_0(0) + u_1(0) = f_0 + f_1$ , we see further that the average grey value is preserved:

$$(3.7) \quad u_0(t) + u_1(t) = f_0 + f_1 \quad \forall t \geq 0.$$

Using (3.6) and (3.7), we obtain the analytical solution

$$(3.8) \quad u_0(t) = \begin{cases} f_0 + t \operatorname{sgn}(f_1 - f_0) & \text{if } t < |f_1 - f_0|/2, \\ (f_0 + f_1)/2 & \text{else,} \end{cases}$$

$$(3.9) \quad u_1(t) = \begin{cases} f_1 - t \operatorname{sgn}(f_1 - f_0) & \text{if } t < |f_1 - f_0|/2, \\ (f_0 + f_1)/2 & \text{else.} \end{cases}$$

Interestingly, this result is identical to the results (3.1)–(3.2) for soft Haar wavelet shrinkage if one identifies the diffusion time  $t$  with the threshold parameter  $\tau = \sqrt{2}t$ .



**3.3. TV regularization of two-pixel signals.** Let us now turn our attention to the regularization framework. Again we are interested only in the two-pixel model  $(f_0, f_1)$ . We consider a space-discrete variant of (2.6) with a TV penalizer:

$$(3.10) \quad E(u_0, u_1; \alpha, f) = (f_0 - u_0)^2 + (f_1 - u_1)^2 + 2\alpha |u_1 - u_0|.$$

Straightforward computation results in the following minimizer of (3.10):

$$u_0 = \begin{cases} f_0 + \alpha \operatorname{sgn}(f_1 - f_0) & \text{if } \alpha < |f_1 - f_0|/2, \\ (f_0 + f_1)/2 & \text{else,} \end{cases}$$

$$u_1 = \begin{cases} f_1 - \alpha \operatorname{sgn}(f_1 - f_0) & \text{if } \alpha < |f_1 - f_0|/2, \\ (f_0 + f_1)/2 & \text{else.} \end{cases}$$

This result coincides with the outcome of a single Haar wavelet shrinkage step with shrinkage parameter  $\tau = \sqrt{2}\alpha$ . Moreover, it is identical to TV diffusion if one replaces the diffusion time  $t$  by the regularization parameter  $\alpha$ . Thus, all three methods are equivalent by setting  $\tau = \sqrt{2}t = \sqrt{2}\alpha$ . It is remarkable that TV diffusion and TV regularization give identical evolutions in the two-pixel case. From the considerations in section 2.3 one would expect only that the processes approximate each other. In section 4.3 we will investigate if this equivalence also holds in the general space-discrete case with  $N$  pixels.

**3.4. SIDEs for two-pixel signals.** If we consider the SIDE evolution of a two-pixel signal  $(f_0, f_1)$ , we obtain for the case of a TV force function the dynamical system

$$\dot{u}_0 = \frac{u_1 - u_0}{|u_1 - u_0|}, \quad \dot{u}_1 = -\frac{u_1 - u_0}{|u_1 - u_0|},$$

with initial conditions  $u_0(0) = f_0$  and  $u_1(0) = f_1$ .

This is the same evolution as in the TV diffusion case. Hence, its solution is given by (3.8)–(3.9), and there is a finite merging time  $t = |f_1 - f_0|/2$ .

**4. N-pixel signals.** So far we have focused on the two-pixel case. Let us now investigate which of the equivalences carry over to the general 1-D case with  $N$  pixels. To this end we will consider shift invariant wavelet shrinkage on a single scale, show that it performs a numerical approximation to TV diffusion, prove the equivalence of space-discrete TV diffusion and discrete TV regularization by deriving analytical solutions of both processes, and show that this solution coincides with SIDEs with TV force functions.

**4.1. Shift invariant wavelet shrinkage on a single scale.** Let us first reconsider the soft Haar wavelet shrinkage on a single scale with  $N$  pixels, where  $N$  is even. Figure 2 illustrates this computation as the two-channel filter bank. As usual we apply the  $z$ -transform notation  $f(z) = \sum_{i=0}^{N-1} f_i z^{-i}$ . Then  $\boxed{H_i(z)}$  ( $i = 0, 1$ ) denotes the convolution of  $f$  with the lowpass filter ( $i = 0$ ) and the highpass filter ( $i = 1$ ), i.e.,  $f(z)H_i(z)$ ,  $\boxed{2 \downarrow}$  and  $\boxed{2 \uparrow}$  downsampling and upsampling by 2, respectively, and the circle soft thresholding by  $S_\tau$ . Finally,  $\bullet$  signifies addition.

The use of Haar wavelets creates a natural decomposition of the signal into two-pixel pairs of type  $(f_{2j}, f_{2j+1})$  ( $j = 0, \dots, N/2-1$ ). This two-pixel clustering, however, also causes a lack of translation invariance which may be responsible for visual artifacts. One method to improve the quality of the denoised signal considerably is to

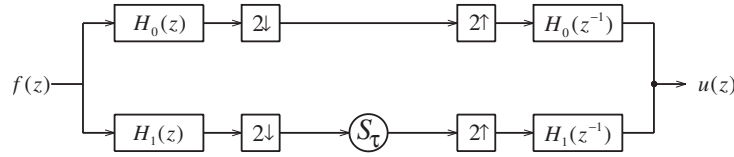


FIG. 2. Two-channel filter bank with  $H_0(z) = \frac{1+z}{\sqrt{2}}$  and  $H_1(z) = \frac{1-z}{\sqrt{2}}$ .

“average out” the translation dependence. This method was termed *cycle spinning* by Coifman and Donoho [19]. For a single wavelet decomposition step, the basic idea of cycle spinning on a single scale reads as follows:

- (a) perform wavelet shrinkage (3.1), (3.2) on successive pairs of the original signal;
- (b) shift the signal one pixel to the right;
  - perform wavelet shrinkage on successive pairs of the shifted signal;
  - shift the resulting signal one pixel back to the left;
- (c) average both results.

The shifting process requires the incorporation of boundary conditions for  $f$ . Again we mirror the signal  $f$  at its ends. Steps (a)–(c) are equivalent to denoising the signal using a nonsubsampled filter bank. More sophisticated material on oversampled filter banks, corresponding wavelet frames, and undecimated wavelet transforms can be found in [31].

**4.2. Equivalence to a numerical scheme for TV diffusion.** We have seen that, in order to improve the performance of wavelet shrinkage and to make wavelet-based denoising translationally invariant, cycle spinning can be used. Since there is an equivalence between Haar wavelet shrinkage and TV diffusion in the two-pixel case, it would be natural to ask if there is a TV diffusion scheme equivalent to translationally invariant soft Haar wavelet shrinkage on a single level. This leads us to an interesting novel scheme for TV diffusion.

*Derivation of the scheme.* We have been able to derive an analytical solution for TV diffusion in the two-pixel case. We can use this two-pixel solution to create a numerical scheme for  $N$  pixels. In order to derive such a scheme for some time step size  $\Delta t$ , we proceed in three steps that are inspired by the cycle spinning technique:

- (a) perform TV diffusion with step size  $2\Delta t$  on all pixel pairs  $(u_{2j}, u_{2j+1})$ ;
- (b) perform TV diffusion with step size  $2\Delta t$  on all pixel pairs  $(u_{2j-1}, u_{2j})$ ;
- (c) average both results.

Obviously, one step of this iterative scheme is equivalent to a translationally invariant Haar wavelet shrinkage with threshold  $\tau = 2\sqrt{2}\Delta t$  on a single level. So let us investigate this scheme in more detail.

At iteration level  $k$ , we assume that our signal is given by  $(u_i^k)_{i=0}^{N-1}$ . We denote the resulting signal of step (a) by  $(v_i^{k+1})_{i=0}^{N-1}$  and the spatial grid size by  $h$ . From our analysis of the two-pixel situation, it follows that  $v_i$  in some even pixel  $i = 2j$  is given by

$$(4.1) \quad v_i^{k+1} = \frac{u_i^k + u_{i+1}^k}{2} - \begin{cases} \max\left(\frac{u_{i+1}^k - u_i^k}{2} - \frac{2\Delta t}{h}, 0\right) & \text{if } u_{i+1}^k \geq u_i^k, \\ \min\left(\frac{u_{i+1}^k - u_i^k}{2} + \frac{2\Delta t}{h}, 0\right) & \text{if } u_{i+1}^k < u_i^k. \end{cases}$$

To simplify the notation, we assume only in this subsection instead of the third agreement in (I) that  $\text{sgn}(0) := 0$ . It is not difficult to see that (4.1) can be rewritten

as

$$(4.2) \quad v_i^{k+1} = u_i^k + \frac{2\Delta t}{h} \operatorname{sgn}(u_{i+1}^k - u_i^k) \min\left(1, \frac{h}{4\Delta t} |u_{i+1}^k - u_i^k|\right).$$

Step (b) leads to a resulting signal  $(w_i^{k+1})_{i=0}^{N-1}$ . For  $i = 2j$  it is given by

$$(4.3) \quad w_i^{k+1} = u_i^k - \frac{2\Delta t}{h} \operatorname{sgn}(u_i^k - u_{i-1}^k) \min\left(1, \frac{h}{4\Delta t} |u_i^k - u_{i-1}^k|\right).$$

Thus, the averaging step (c) gives the final scheme for TV diffusion:

$$(4.4) \quad \begin{aligned} u_i^{k+1} &= u_i^k + \frac{\Delta t}{h} \operatorname{sgn}(u_{i+1}^k - u_i^k) \min\left(1, \frac{h}{4\Delta t} |u_{i+1}^k - u_i^k|\right) \\ &\quad - \frac{\Delta t}{h} \operatorname{sgn}(u_i^k - u_{i-1}^k) \min\left(1, \frac{h}{4\Delta t} |u_i^k - u_{i-1}^k|\right). \end{aligned}$$

The same scheme can also be derived if  $i$  is odd, since the construction (a)–(c) in this subsection ensures that the result is translationally invariant. Hence it holds for every inner pixel  $i \in \{1, \dots, N-2\}$ . It is even valid for the boundary pixels  $i = 0$  and  $i = N-1$  if we realize the homogeneous Neumann boundary conditions by introducing dummy values  $u_{-1}^k := u_0^k$  and  $u_N^k := u_{N-1}^k$ .

*Stability.* Let us now investigate the stability properties of the explicit finite difference scheme (4.2). Since (4.2) satisfies

$$\min(u_i^k, u_{i+1}^k) \leq v_i^{k+1} \leq \max(u_i^k, u_{i+1}^k)$$

and (4.3) fulfills the estimate

$$\min(u_{i-1}^k, u_i^k) \leq w_i^{k+1} \leq \max(u_{i-1}^k, u_i^k),$$

we can conclude that

$$\min(u_{i-1}^k, u_i^k, u_{i+1}^k) \leq u_i^{k+1} \leq \max(u_{i-1}^k, u_i^k, u_{i+1}^k).$$

With the initial condition  $u_j^0 = f_j$  for  $j = 0, \dots, N-1$ , it follows that the two-pixel scheme (4.2) satisfies the discrete maximum–minimum principle

$$\min_j f_j \leq u_i^{k+1} \leq \max_j f_j$$

for all pixels  $i \in \{0, \dots, N-1\}$ , all iteration levels  $k = 0, 1, 2, \dots$ , and all time step sizes  $\Delta t > 0$ . In particular, this shows that the scheme is absolutely stable in the maximum norm.

We may regard (4.4) as a stabilization of the naive explicit scheme

$$(4.5) \quad u_i^{k+1} = u_i^k + \frac{\Delta t}{h} \operatorname{sgn}(u_{i+1}^k - u_i^k) - \frac{\Delta t}{h} \operatorname{sgn}(u_i^k - u_{i-1}^k),$$

which becomes unstable for arbitrary small time steps if neighboring values become arbitrarily close.

*Consistency.* The absolute stability in scheme (4.4) is at the expense that its consistency is no longer unconditional. This effect is typical for absolutely stable explicit schemes; see, for example, the DuFort–Frankel scheme for linear diffusion. In

our case, (4.4) is an  $O(\Delta t + h^2)$  approximation to the continuous TV diffusion for  $\Delta t \leq \frac{h}{4} \min(|u_{i+1}^k - u_i^k|, |u_i^k - u_{i-1}^k|)$ , since it coincides with scheme (4.5) then. For larger time step sizes, the scheme performs averaging within the neighborhood of each pixel. By using small time step sizes, these averaging effects appear only in regions that are already almost flat such that the difference from real TV diffusion becomes invisible. This two-pixel scheme may be regarded as an alternative to classical finite difference schemes that are based on the regularized TV flow

$$(4.6) \quad u_t = \left( \frac{u_x}{\sqrt{\epsilon^2 + u_x^2}} \right)_x.$$

The  $\epsilon$ -regularization is necessary for making the diffusivity bounded. It has an effect similar to the deviation from consistency in the two-pixel scheme (4.4): For small  $|u_x|$ , a PDE is approximated that differs from TV diffusion and has better stability properties. Indeed, in section 6 we shall see that both schemes give very similar results.

*Related schemes.* The idea to split up a diffusion process into pairwise interactions has also proved to be fruitful in other fields. In the context of fluid dynamic problems, related schemes have been formulated by Richardson, Ferrell, and Long [39]. These authors, however, use multiplicative splittings; i.e., they first compute the diffusion of the pairs of type  $(u_{2j}, u_{2j+1})$ , which is then used as the initial state for the subsequent diffusion of the shifted pairs. In a general nonlinear setting, such a scheme is not translationally invariant. Our approach computes the diffusion of the pairs and the shifted pairs in parallel and averages afterwards. This additive splitting guarantees translation invariance. The splitting into two-pixel interactions distinguishes scheme (4.4) from other additive operator splittings [29, 48]. They use directional splittings along the coordinate axis.

**4.3. Equivalence of space-discrete TV diffusion and discrete TV regularization.** The equivalence of TV diffusion and TV regularization in the two-pixel case gives rise to the question of whether this equivalence also holds in the  $N$ -pixel situation. In order to prove this, we now show that both processes have the same analytical solutions.

**4.3.1. Space-discrete TV diffusion.** We consider the following dynamical system designed to describe space-discrete TV flow on a 1-D signal with  $N$  pixels:

$$(4.7) \quad \begin{aligned} \dot{u}_0 &= \operatorname{sgn}(u_1 - u_0), \\ \dot{u}_i &= \operatorname{sgn}(u_{i+1} - u_i) - \operatorname{sgn}(u_i - u_{i-1}) \quad (i = 1, \dots, N-2), \\ \dot{u}_{N-1} &= -\operatorname{sgn}(u_{N-1} - u_{N-2}), \\ u(0) &= f. \end{aligned}$$

In the following, we further set  $u_{-1} := u_0$  and  $u_N := u_{N-1}$ . Since the right-hand side of this system is discontinuous, we need again a more detailed specification of when a system of functions is said to satisfy these differential equations. A vector-valued function  $u$  is said to fulfill the system (4.7) over the time interval  $[0, T]$  if the following holds true:

- (II)  $u$  is an absolutely continuous vector-valued function which satisfies (4.7) almost everywhere, where  $\operatorname{sgn}$  is defined by (I) in subsection 3.2.
- (III) If  $\dot{u}_i(t)$  and  $\dot{u}_{i+1}(t)$  exist for the same  $t$ , and  $u_{i+1}(t) = u_i(t)$  holds, then the expression  $\operatorname{sgn}(u_{i+1}(t) - u_i(t))$  occurring in both the right-hand sides for  $\dot{u}_i(t)$  and  $\dot{u}_{i+1}(t)$  must take the same value in both equations.

With this notation we can establish the following results.

PROPOSITION 4.1 (properties of space-discrete TV diffusion). *The system (4.7) has a unique solution  $u(t)$  in the sense of (II) and (III). This solution has the following properties:*

(i) Finite extinction time. *There exists a finite time  $T \geq 0$  such that for all  $t \geq T$  the signal becomes constant:*

$$(4.8) \quad u_i(t) = \frac{1}{N} \sum_{k=0}^{N-1} f_k \quad \text{for all } i = 0, \dots, N - 1.$$

(ii) Finite number of merging events. *There exists a finite sequence  $0 = t_0 < t_1 < \dots < t_{n-1} < t_n = T$  such that the interval  $[0, T)$  splits into subintervals  $[t_j, t_{j+1})$  with the property that for all  $i = 0, \dots, N - 2$  either  $u_i(t) = u_{i+1}(t)$  or  $u_i(t) \neq u_{i+1}(t)$  throughout  $[t_j, t_{j+1})$ . The absolute difference between neighboring pixels does not become larger for increasing  $t \in [t_j, t_{j+1})$ .*

(iii) Analytical solution. *In each of the subintervals  $[t_j, t_{j+1})$  constant regions of  $u(t)$  evolve linearly:*

*For a fixed index  $i$  let us consider a constant region given by*

$$(4.9) \quad \begin{aligned} u_{i-l+1} &= \dots = u_i = u_{i+1} = \dots = u_{i+r} & (l \geq 1, r \geq 0), \\ u_{i-l} &\neq u_{i-l+1} \text{ if } i-l \geq 0, & u_{i+r} \neq u_{i+r+1} \text{ if } i+r \leq N-1 \end{aligned}$$

*for all  $t \in [t_j, t_{j+1})$ . We call (4.9) a region of size  $m_{i,t_j} = l + r$ . For  $t \in [t_j, t_{j+1})$  let  $\Delta t = t - t_j$ . Then  $u_i(t)$  is given by*

$$u_i(t) = u_i(t_j) + \mu_{i,t_j} \frac{2\Delta t}{m_{i,t_j}},$$

*where  $\mu_{i,t_j}$  reflects the relation between the region containing  $u_i$  and its neighboring regions. It is given as follows:*

*For inner regions (i.e.,  $i - l \geq 0$  and  $i + r \leq N - 1$ ) we have*

$$(4.10) \quad \mu_{i,t_j} = \begin{cases} 0 & \text{if } (u_{i-l}, u_i, u_{i+r+1}) \text{ is strictly monotonic,} \\ 1 & \text{if } u_i \text{ is minimal in } (u_{i-l}, u_i, u_{i+r+1}), \\ -1 & \text{if } u_i \text{ is maximal in } (u_{i-l}, u_i, u_{i+r+1}), \end{cases}$$

*and in the boundary case ( $i - l + 1 = 0$  or  $i + r = N - 1$ ) the evolution is half as fast:*

$$(4.11) \quad \mu_{i,t_j} = \begin{cases} 0 & \text{if } m = N, \\ \frac{1}{2} & \text{if } u_i \text{ is minimal in } (u_{i-l}, u_i, u_{i+r+1}), \\ -\frac{1}{2} & \text{if } u_i \text{ is maximal in } (u_{i-l}, u_i, u_{i+r+1}). \end{cases}$$

*Proof.* Let  $u$  be a solution of (4.7). We show that  $u$  is uniquely determined and satisfies the rules (i)–(iii). Our proof proceeds in four steps.

*Step 1.* If  $\dot{u}(t)$  exists at a fixed time  $t$  and  $u_i(t)$  lies at this time in some region

$$\begin{aligned} u_{i-l+1}(t) &= \dots = u_i(t) = \dots = u_{i+r}(t) & (l \geq 1, r \geq 0), \\ u_{i-l}(t) &\neq u_{i-l+1}(t) \text{ if } i-l \geq 0, & u_{i+r}(t) \neq u_{i+r+1}(t) \text{ if } i+r \leq N-1 \end{aligned}$$

of size  $m_{i,t}$ , then it follows by (4.7) and (III) in the nonboundary case  $i - l \geq 0$  and  $i + r \leq N - 1$  that

$$u_i(t) = \frac{1}{m_{i,t}} \sum_{k=-l+1}^r u_{i+k}(t),$$

and therefore

$$\begin{aligned} \dot{u}_i(t) &= \frac{1}{m_{i,t}} \sum_{k=-l+1}^r \dot{u}_{i+k}(t) = \frac{1}{m_{i,t}} (\operatorname{sgn}(u_{i+r+1}(t) - u_i(t)) - \operatorname{sgn}(u_i(t) - u_{i-l}(t))) \\ (4.12) \quad &= \mu_{i,t} \frac{2}{m_{i,t}}, \end{aligned}$$

where  $\mu_{i,t}$  describes the relation between the region containing  $u_i$  and its neighbors at time  $t$  as in (4.10). In the boundary case  $i - l + 1 = 0$  or  $i + r = N - 1$  we follow the same lines and obtain (4.12) with  $\mu_{i,t}$  defined by (4.11).

*Step 2.* Let  $\dot{u}(t)$  exist in some small interval  $(\tau_0, \tau_1)$ , and assume that  $u_i(t) \neq u_{i+1}(t)$  for some  $i \in \{0, \dots, N - 2\}$  and all  $t \in (\tau_0, \tau_1)$ . By continuity of  $u$  we may assume that  $u_i(t) < u_{i+1}(t)$  throughout  $(\tau_0, \tau_1)$ . The opposite case  $u_i(t) > u_{i+1}(t)$  can be handled in the same way. Then we obtain by (4.12) and the definition of  $\mu_{i,t}$  for all  $t \in (\tau_0, \tau_1)$  that

$$(4.13) \quad \dot{u}_i(t) \geq 0 \quad \text{if} \quad i - l \geq 0,$$

$$(4.14) \quad \dot{u}_i(t) > 0 \quad \text{if} \quad i - l + 1 = 0,$$

$$(4.15) \quad \dot{u}_{i+1}(t) \leq 0 \quad \text{if} \quad i + r \leq N - 2,$$

$$(4.16) \quad \dot{u}_{i+1}(t) < 0 \quad \text{if} \quad i + r = N - 1.$$

Set  $w(t) := u_{i+1}(t) - u_i(t)$ . Then the mean value theorem yields

$$w(\tau_1) - w(\tau_0) = (\tau_1 - \tau_0) \dot{w}(t^*)$$

for some  $t^* \in (\tau_0, \tau_1)$ , and we get by (4.13)–(4.16) that

$$w(\tau_1) - w(\tau_0) \leq 0$$

with strict inequality in the boundary case. Consequently, the difference between pixels cannot become larger in the considered interval. In particular, by continuity of  $u$ , pixels cannot be split. Once merged they stay merged.

*Step 3.* Now we start at time  $t_0 = 0$ . Let  $t_1$  be the largest time such that  $\dot{u}(t)$  exists and no merging of regions appears in  $(0, t_1)$ . Then, for all  $i \in \{0, \dots, N - 1\}$ , a function  $u_i$  is in the same region with the same relations to its neighboring regions throughout  $[0, t_1)$ . Thus, we conclude by (4.12) that

$$\dot{u}_i(t) = \mu_{i,0} \frac{2}{m_{i,0}} \quad (t \in (0, t_1))$$

and, consequently,

$$u_i(t) = \mu_{i,0} \frac{2t}{m_{i,0}} + C_{i,0} = f_i + \mu_{i,0} \frac{2t}{m_{i,0}} \quad (t \in [0, t_1]),$$

where the last equality follows by continuity of  $u_i$  if  $t$  approaches 0.

*Step 4.* We are now in the position to analyze the entire chain of merging events successively.

Next, we consider the largest interval  $(t_1, t_2)$  without merging events in the same way, where we take the initial setting  $u(t_1)$  into account instead of  $f$ . Then we obtain

$$u_i(t) = \mu_{i,t_1} \frac{2t}{m_{i,t_1}} + C_{i,t_1},$$

where, by continuity of  $u_i$ ,  $u_i(t_1) = \mu_{i,t_1} \frac{2t_1}{m_{i,t_1}} + C_{i,t_1}$  and, consequently,

$$u_i(t) = u_i(t_1) + \mu_{i,t_1} \frac{2(t - t_1)}{m_{i,t_1}}.$$

Now we can continue in the same way by considering  $[t_2, t_3)$  and so on. Since we have only a finite number  $N$  of pixels and some of these pixels merge at the points  $t_j$ , the process stops after a finite number of  $n$  steps with output (4.8). Conversely, it is easy to check that a function  $u$  with (i)–(iii) is a solution of the system (4.7). This completes the proof of the proposition.  $\square$

**4.3.2. Discrete TV regularization.** Next, we will prove that discrete TV regularization satisfies the same rules as space-discrete TV diffusion. For given initial data  $f = (f_0, \dots, f_{N-1})$  discrete TV regularization consists of constructing the minimizer  $u(\alpha) = \min_u E(u; \alpha, f)$  of the functional

$$(4.17) \quad E(u; \alpha, f) = \sum_{i=0}^{N-1} ((u_i - f_i)^2 + 2\alpha|u_{i+1} - u_i|),$$

where we suppose again Neumann boundary conditions  $u_{-1} = u_0$  and  $u_N = u_{N-1}$ .

For a fixed regularization parameter  $\alpha \geq 0$ , the minimizer of (4.17) is uniquely determined since  $E(u; \alpha, f)$  is strictly convex in  $u_0, \dots, u_{N-1}$ . Further,  $E(u, \alpha; f)$  is a continuous function in  $u_0, \dots, u_{N-1}, \alpha$ . Consequently,  $u(\alpha)$  is a (componentwise) continuous function in  $\alpha$ .

The following proposition implies, together with Proposition 4.1, the equivalence of space-discrete TV diffusion and discrete TV regularization.

PROPOSITION 4.2 (properties of discrete TV regularization). *The minimizing function  $u(\alpha)$  of (4.17) is uniquely determined by the following rules:*

(i) Finite extinction parameter. *There exists a finite  $A \geq 0$  such that for all  $\alpha \geq A$  the signal becomes constant:*

$$u_i(\alpha) = \frac{1}{N} \sum_{k=0}^{N-1} f_k \quad \text{for all } i = 0, \dots, N - 1.$$

(ii) Finite number of merging events. *There exists a finite sequence  $0 = a_0 < a_1 < \dots < a_{n-1} < a_n = A$  such that the interval  $[0, A)$  splits into subintervals  $[a_j, a_{j+1})$  with the property that for all  $i = 0, \dots, N - 2$  either  $u_i(\alpha) = u_{i+1}(\alpha)$  or  $u_i(\alpha) \neq u_{i+1}(\alpha)$  throughout  $[a_j, a_{j+1})$ . The absolute difference between neighboring pixels does not become larger for increasing  $\alpha \in [a_j, a_{j+1})$ .*

(iii) Analytical solution. *In each of the subintervals  $[a_j, a_{j+1})$  constant regions of  $u(\alpha)$  evolve linearly:*

*For a fixed index  $i$  let us consider a constant region given by*

$$(4.18) \quad u_{i-l+1} = \dots = u_i = u_{i+1} = \dots = u_{i+r} \quad (l \geq 1, r \geq 0),$$

$$(4.19) \quad u_{i-l} \neq u_{i-l+1} \text{ if } i - l \geq 0, \quad u_{i+r} \neq u_{i+r+1} \text{ if } i + r \leq N - 2$$

*for all  $\alpha \in [a_j, a_{j+1})$ . We call (4.18) a region of size  $m_{i,a_j} = l + r$ . For  $\alpha \in [a_j, a_{j+1})$  let  $\Delta\alpha = \alpha - a_j$ . Then  $u_i(\alpha)$  is given by*

$$u_i(\alpha) = u_i(a_j) + \mu_{i,a_j} \frac{2\Delta\alpha}{m_{i,a_j}},$$

where  $\mu_{i,a_j}$  reflects the relation between the region containing  $u_i$  and its neighboring regions. It is given as follows:

For inner regions (i.e.,  $i - l \geq 0$  and  $i + r \leq N - 2$ ) we have

$$(4.20) \quad \mu_{i,a_j} = \begin{cases} 0 & \text{if } (u_{i-l}, u_i, u_{i+r+1}) \text{ is strictly monotonic,} \\ 1 & \text{if } u_i \text{ is minimal in } (u_{i-l}, u_i, u_{i+r+1}), \\ -1 & \text{if } u_i \text{ is maximal in } (u_{i-l}, u_i, u_{i+r+1}), \end{cases}$$

and in the boundary case ( $i - l + 1 = 0$  or  $i + r = N - 1$ ) the evolution is half as fast:

$$(4.21) \quad \mu_{i,a_j} = \begin{cases} 0 & \text{if } m = N, \\ \frac{1}{2} & \text{if } u_i \text{ is minimal in } (u_{i-l}, u_i, u_{i+r+1}), \\ -\frac{1}{2} & \text{if } u_i \text{ is maximal in } (u_{i-l}, u_i, u_{i+r+1}). \end{cases}$$

*Proof.* A proof that is in complete analogy with our proof for the TV diffusion case is presented in [10, 45].  $\square$

Similar results have also been established in a different way by Strong [46] for the case of continuous TV regularization methods with step functions as initializations. It should be noted that TV regularization by using the taut-string algorithm was also considered by Mammen and van de Geer [32]; see also [28].

**4.4. Equivalence to SIDEs with TV force functions.** In section 2.4 we have seen that 1-D SIDEs with region size 1 and TV force function are identical to space-discrete TV diffusion. Moreover, in section 4.3 we have derived analytical solutions of space-discrete TV diffusion and discrete TV regularization that show the same merging behavior as SIDEs with TV force functions. Consequently, 1-D SIDEs can be interpreted as an exact solution of space-discrete TV diffusion or regularization in general.

This also confirms that the merging steps in the SIDE evolution are much more than a heuristic stabilization that speeds up the evolution: They are a natural consequence of the degenerated diffusivities that are unbounded in 0. Last but not least, our considerations can be regarded as a theoretical justification of region merging in terms of variational and PDE-based techniques.

**5. Multiple scales.** So far we have considered only soft wavelet shrinkage on a single scale. In almost all practical applications, however, wavelet shrinkage is performed on multiple scales. In this section, we interpret *multiscale* soft shrinkage with Haar wavelets as the application of nonlinear TV-based diffusion to two-pixel groups of *hierarchical* signals. First, we consider the standard situation without shift invariance; then we discuss the shift-invariant case. Finally, we address a frequent problem that occurs with wavelet shrinkage on multiple scales: the presence of Gibbs-like artifacts. We analyze ways to circumvent this phenomenon by using scale-dependent thresholds.

Throughout this section we deal with signals of length  $N = 2^n$  ( $n \in \mathbb{N}$ ).

**5.1. Standard case without shift invariance.** Haar wavelet shrinkage on two scales is described by the filter bank in Figure 3. To obtain more than two scales we further split up the upper branch of the inner filter bank and so on until we arrive at scale  $n = \log_2 N$ , where the successive downsampling by 2 results in a one-pixel signal.

Next, we briefly recall the concept of *Gaussian* and *Laplacian pyramids* [11] with respect to the Haar filters. The Gaussian pyramid we are interested in is the sequence



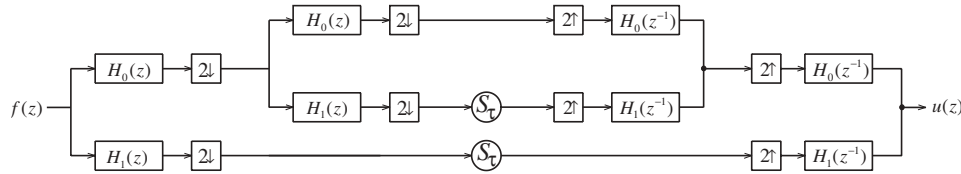


FIG. 3. Two scales of Haar wavelet shrinkage with  $H_0(z) = \frac{1+z}{\sqrt{2}}$  and  $H_1(z) = \frac{1-z}{\sqrt{2}}$ .

of  $H_0$ -smoothed and downsampled versions of an initial signal  $f$  given by

$$f = f^{(0)} \longrightarrow f^{(1)} = Rf \longrightarrow \dots \longrightarrow f^{(n)} = R^n f,$$

where  $R$  denotes the operator for  $H_0$ -smoothing and subsequent downsampling by 2, i.e.,

$$f_i^{(j+1)} = (Rf^{(j)})_i = (f_{2i}^{(j)} + f_{2i+1}^{(j)})/\sqrt{2} \quad (j = 0, \dots, n-1; i = 0, \dots, N/2^{j+1} - 1).$$

Let  $Pf^{(j)}$  denote the prolonged version of  $f^{(j)}$  given by

$$(5.1) \quad (Pf^{(j)})_{2i} = (Pf^{(j)})_{2i+1} = f_i^{(j)}/\sqrt{2} \quad (j = 1, \dots, n; i = 0, \dots, N/2^j - 1).$$

Then the corresponding Laplacian pyramid is the sequence

$$f - Pf^{(1)} \longrightarrow f^{(1)} - Pf^{(2)} \longrightarrow \dots \longrightarrow f^{(n-1)} - Pf^{(n)} \longrightarrow f^{(n)}.$$

By

$$f^{(j)} = Pf^{(j+1)} + (f^{(j)} - Pf^{(j+1)}) \quad (j = n-1, \dots, 0)$$

we can reconstruct  $f$  from its Laplacian pyramid.

Let  $\text{diff}_t$  denote the operator of nonlinear diffusion with TV diffusivity and stopping time  $t$  applied to the successive two-pixel parts of a signal. By subsection 3.2 we know that  $\text{diff}_t$  performs like a single wavelet shrinkage step with soft threshold parameter  $\tau = \sqrt{2}t$ . In other words, the result of the filter bank in Figure 2 is  $u = \text{diff}_t(f)$ . Further, we see that the upper branch of this filter bank produces  $Pf^{(1)}$  so that the lower branch must produce  $\text{diff}_t(f) - Pf^{(1)}$ . By (5.1) and (3.1) it is easy to check that the nonlinear operator  $\text{diff}_t$  fulfills  $\text{diff}_t(f) - Pf^{(1)} = \text{diff}_t(f - Pf^{(1)})$ . Thus, one wavelet shrinkage step is given by  $u = Pf^{(1)} + \text{diff}_t(f - Pf^{(1)})$ . Now the multiscale Haar wavelet shrinkage up to scale  $n$  can be described by successive application of  $\text{diff}_t$  to the Laplacian pyramid:

$$(5.2) \quad u^{(n)} = f^{(n)},$$

$$(5.3) \quad u^{(j)} = Pu^{(j+1)} + \text{diff}_t(f^{(j)} - Pf^{(j+1)}) \quad (j = n-1, \dots, 0).$$

The result of the multiscale wavelet shrinkage is  $u = u^{(0)}$ .

**5.2. Shift-invariant case.** Now we consider translation-invariant multiscale wavelet shrinkage. In the multiscale setting we apply cycle spinning over the range of all  $N$  shifts of  $f$ . The filter bank which corresponds to two scales of translation-invariant Haar wavelet shrinkage is shown in Figure 4. Note that the inner filter bank uses  $z^2$  instead of  $z$  in  $H_i$  ( $i = 0, 1$ ). In general we have to replace  $z$  by  $z^{2^{j-1}}$  at

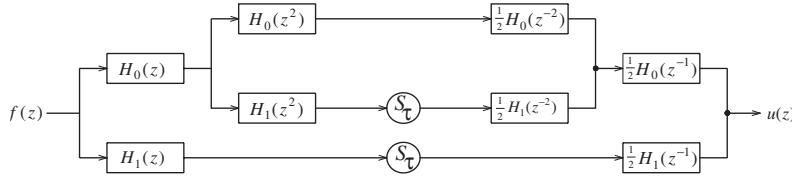


FIG. 4. Two scales of shift-invariant Haar wavelet shrinkage with  $H_0(z) = \frac{1+z}{\sqrt{2}}$  and  $H_1(z) = \frac{1-z}{\sqrt{2}}$ .

scale  $j$ . While ordinary wavelet shrinkage requires  $O(N)$  arithmetic operations, its translation-invariant version needs  $O(N \log_2 N)$  arithmetic operations.

In subsection 4.2 we have deduced a numerical scheme for TV diffusion. Each iteration is given by (4.4). This coincides with a single translation-invariant Haar wavelet shrinkage step with threshold  $\tau = 2\sqrt{2}t$ . Using our operator  $\text{diff}$  and the operator  $S$ , which shifts a signal one pixel to the right, the result  $u$  of the single-scale translation-invariant filter bank is given by

$$u = \frac{1}{2} (\text{diff}_{2t}(f) + S^{-1} \text{diff}_{2t}(Sf)).$$

Now the multiscale translation-invariant Haar wavelet shrinkage can be interpreted as application of  $\text{diff}$  to a multiple Laplacian pyramid. We define a multiple Gaussian pyramid by

$$f^{(0,0)} \rightarrow (f^{(1,0)}, f^{(1,1)}) \rightarrow (f^{(2,0)}, f^{(2,1)}, f^{(2,2)}, f^{(2,3)}) \rightarrow \dots \rightarrow (f^{(n,0)}, \dots, f^{(n,2^n-1)}),$$

where  $f = f^{(0,0)}$ . Here  $f^{(j,k)}$  is obtained by successive application of the operators  $R$  and  $RS$  on  $f$  as follows: Let 0 denote the application of  $R$  and 1 the application of  $RS$ ; then these operators are applied to  $f$  in the order of the binary representation  $(k_{j-1}, \dots, k_0)_2$  of  $k$ , where we start from the left. For example, we get  $f^{(2,1)} = f^{(2,(0,1)_2)} = RS Rf$  and  $f^{(2,2)} = f^{(2,(1,0)_2)} = RRSf$ . Then the multiple Laplacian pyramid is given by

$$\begin{aligned} & (f^{(0,0)} - Pf^{(1,0)}, Sf^{(0,0)} - Pf^{(1,1)}) \\ & \rightarrow (f^{(1,0)} - Pf^{(2,0)}, Sf^{(1,0)} - Pf^{(2,1)}, f^{(1,1)} - Pf^{(2,2)}, Sf^{(1,1)} - Pf^{(2,3)}) \rightarrow \dots \rightarrow \\ & (f^{(n,0)}, \dots, f^{(n,2^n-1)}), \end{aligned}$$

and the translation-invariant version of (5.2)–(5.3) can be obtained from this multiple Laplacian pyramid by

$$\begin{aligned} u^{(n,k)} &= f^{(n,k)} \quad (k = 0, \dots, 2^n - 1), \\ u^{(j,k)} &= \frac{1}{2} \left( Pu^{(j+1,2k)} + \text{diff}_{2t}(f^{(j,k)} - Pf^{(j+1,2k)}) \right. \\ & \quad \left. + S^{-1}(Pu^{(j+1,2k+1)} + \text{diff}_{2t}(Sf^{(j,k)} - Pf^{(j+1,2k+1)})) \right) \end{aligned}$$

for  $j = n - 1, \dots, 0; k = 0, \dots, 2^j - 1$ . The result is  $u = u^{(0,0)}$ .

**5.3. Scale-dependent thresholds.** Cycle spinning techniques can be used to make wavelet shrinkage not only translationally invariant, but they can also reduce artifacts. However, it is still possible that oscillatory (Gibbs-like) artifacts appear if

multiple scales are used. We want to demonstrate that the use of the scale-dependent thresholds

$$(5.4) \quad \tau_j = \tau / \sqrt{2^{j-1}} \quad (j = 1, \dots, n)$$

suppresses oscillations in the shrinkage process.

In this subsection, we consider signals  $f = (f_0, \dots, f_{N-1})$  with periodic boundary conditions. Note that mirror boundary conditions can easily be transferred into periodic ones by doubling the signal. The decimated Haar wavelet shrinkage with full  $n$ -scale decomposition and thresholds (5.4) consists of three operations. It starts with the linear transform (2.2) of  $f$  yielding the wavelet coefficients  $(c^n, d^n, d^{n-1}, \dots, d^1)$ , where  $d^j := (d_0^j, \dots, d_{N/2^j-1}^j)$ . The wavelet coefficients then undergo the soft wavelet thresholding  $s_i^j := S_{\tau_j}(d_i^j)$  ( $j = 1, \dots, n; i = 0, \dots, N/2^j - 1$ ) followed by the inverse linear transform (2.3) of  $(c^n, s^n, s^{n-1}, \dots, s^1)$  which gives the denoised signal  $u(\tau)$ . In particular we have  $u(0) = f$ . Note that by the semigroup property  $S_{\tau+\bar{\tau}}(x) = S_{\bar{\tau}}(S_{\tau}(x))$  of our shrinkage function (2.1) the signal  $u(K\tau)$  obtained by one  $n$ -scale wavelet shrinkage cycle with threshold  $K\tau$  coincides with the signal which results from  $K$  times repeating one  $n$ -scale wavelet shrinkage cycle with smaller threshold  $\tau$ . Of course, this is no longer true for the translation-invariant wavelet transform. In our examples in the next section we will consider iterated translation-invariant Haar wavelet shrinkage with small thresholds  $\tau$ .

Since oscillatory (Gibbs-like) artifacts are characterized by the emergence of new local extrema, we study the behavior of local extrema of the signal under the shrinkage process. We call  $u_i$  an *extremal pixel* if either  $u_{i-1} < u_i, u_i > u_{i+1}$  or  $u_{i-1} > u_i, u_i < u_{i+1}$ .

First, we consider the dynamics of “infinitesimal translation-invariant soft Haar wavelet shrinkage,” i.e., the speed at which pixels of the signal evolve with respect to the threshold  $\tau \in [0, T]$  in the limit case  $T \rightarrow 0$ .

PROPOSITION 5.1 (suppression of Gibbs-like artifacts by scaled thresholds). *Under infinitesimal translation-invariant soft Haar wavelet shrinkage, an extremal pixel  $f_i$  evolves as follows:*

- (i) *The value of the extremal pixel decreases, i.e.,  $\dot{u}_i < 0$ , if it is a maximum and increases, i.e.,  $\dot{u}_i > 0$ , if it is a minimum. Here the dot denotes differentiation with respect to  $\tau$ .*
- (ii) *The absolute value of the difference of the extremal pixel to each of its two neighbors decreases; i.e.,  $\dot{u}_i - \dot{u}_{i\pm 1} < 0$  for a maximum and  $\dot{u}_i - \dot{u}_{i\pm 1} > 0$  for a minimum.*

Statement (i) holds also for the decimated Haar wavelet shrinkage, while statement (ii) cannot be established in that setting.

*Proof.* For the decimated Haar wavelet shrinkage with full  $n$ -scale decomposition and thresholds (5.4) it is easy to check that the resulting signal  $\tilde{u}_i$  is given by

$$\tilde{u}_i = \mu + \sum_{j=1}^n 2^{-j/2} \varepsilon_j(i) s_{\lfloor i/2^j \rfloor}^j,$$

where  $\lfloor x \rfloor$  denotes the largest integer  $\leq x$ . Moreover,  $\mu := \frac{1}{N} \sum_{i=0}^{N-1} f_i$  is the average value, and

$$\varepsilon_j(i) := \begin{cases} 1 & \text{if } \lfloor i/2^{j-1} \rfloor \text{ is even,} \\ -1 & \text{if } \lfloor i/2^{j-1} \rfloor \text{ is odd.} \end{cases}$$

For the translation-invariant Haar wavelet shrinkage, the sum on the right-hand side of this equation is replaced by the average of  $N$  sums of the same kind containing the back-shifted shrunken wavelet coefficients of  $N$  forward-shifted initial signals, i.e.,

$$(5.5) \quad u_i = \mu + \frac{1}{N} \sum_{\nu=0}^{N-1} \sum_{j=1}^n 2^{-j/2} \varepsilon_j(i + \nu) s_{\lfloor (i+\nu)/2^j \rfloor, \nu}^j,$$

where  $s_{i,\nu}^j$  denotes the  $i$ th coefficient of the  $j$ th level of the  $\nu$ -shifted initial signal, and the coefficients are treated  $N/2^j$ -periodic with respect to  $i$ . Of course, some coefficients coincide for different  $\nu$ ; more precisely,

$$s_{\lfloor (i+\nu)/2^j \rfloor, \nu}^j = s_{\lfloor (i+\nu+r2^j)/2^j \rfloor, \nu+r2^j}^j \quad (\nu = 0, \dots, 2^j - 1; \quad r = 0, \dots, N/2^j - 1).$$

This equation allows us to rewrite (5.5) as

$$\begin{aligned} u_i &= \mu + \frac{1}{2\sqrt{2}} \left( \varepsilon_1(i) s_{\lfloor i/2 \rfloor, 0}^1 + \varepsilon_1(i+1) s_{\lfloor (i+1)/2 \rfloor, 1}^1 \right) \\ &\quad + \frac{1}{N} \sum_{j=2}^n 2^{-j/2} \sum_{\nu=0}^{2^j-1} \sum_{r=0}^{N/2^j-1} \varepsilon_j(i + \nu + r2^j) s_{\lfloor (i+\nu+r2^j)/2^j \rfloor, \nu+r2^j}^j \\ &= \mu + \frac{s_{i,+} - s_{i,-}}{2\sqrt{2}} + \sum_{j=2}^n 2^{-3j/2} \sum_{\nu=0}^{2^j-1} \varepsilon_j(i + \nu) s_{\lfloor (i+\nu)/2^j \rfloor, \nu}^j, \end{aligned}$$

where  $s_{i,+} := S_\tau(d_{i,+}) = S_\tau((f_i - f_{i+1})/\sqrt{2})$  and  $s_{i,-} := S_\tau(d_{i,-}) = S_\tau((f_{i-1} - f_i)/\sqrt{2})$ . Now the evolution of  $u_i$  under infinitesimal soft wavelet shrinkage is described by

$$(5.6) \quad \dot{u}_i = \frac{\dot{s}_{i,+} - \dot{s}_{i,-}}{2\sqrt{2}} + \sum_{j=2}^n 2^{-3j/2} \sum_{\nu=0}^{2^j-1} \varepsilon_j(i + \nu) \dot{s}_{\lfloor (i+\nu)/2^j \rfloor, \nu}^j,$$

where

$$\dot{s}^j = \frac{dS_{\tau_j}(d^j)}{d\tau_j} \cdot \frac{d\tau_j}{d\tau} = \frac{-\operatorname{sgn}(d^j)}{\sqrt{2}^{j-1}}.$$

Inserting this into (5.6), we obtain

$$(5.7) \quad \dot{u}_i = \frac{-\operatorname{sgn}(d_{i,+}) + \operatorname{sgn}(d_{i,-})}{2\sqrt{2}} - A_i,$$

where

$$A_i := \sqrt{2} \sum_{j=2}^n 4^{-j} \sum_{\nu=0}^{2^j-1} \varepsilon_j(i + \nu) \operatorname{sgn} \left( d_{\lfloor (i+\nu)/2^j \rfloor, \nu}^j \right).$$

By the triangle inequality we can estimate

$$(5.8) \quad |A_i| \leq \sqrt{2} \sum_{j=2}^n 2^{-j} < \frac{1}{\sqrt{2}}.$$

If  $f_i$  is an extremal pixel, then we have that  $\text{sgn}(d_{i,+}) = -\text{sgn}(d_{i,-}) = 1$  for a maximum and  $-1$  for a minimum. This implies by (5.7) and (5.8) that

$$(5.9) \quad \text{sgn}(\dot{u}_i) = -\text{sgn}(d_{i,+}),$$

proving statement (i) of the proposition.

By subtracting from (5.7) its counterpart for pixel  $u_{i+1}$ , we obtain by  $d_{i,+} = d_{i+1,-}$  that

$$(5.10) \quad \dot{u}_i - \dot{u}_{i+1} = \frac{\text{sgn}(d_{i,-}) - 2\text{sgn}(d_{i,+}) + \text{sgn}(d_{i+1,+})}{2\sqrt{2}} - (A_i - A_{i+1}).$$

In

$$\begin{aligned} & A_i - A_{i+1} \\ &= \sqrt{2} \sum_{j=2}^n 4^{-j} \sum_{\nu=0}^{2^j-1} \left( \varepsilon_j(i + \nu) \text{sgn} \left( d_{\lfloor (i+\nu)/2^j \rfloor, \nu}^j \right) - \varepsilon_j(i + 1 + \nu) \text{sgn} \left( d_{\lfloor (i+1+\nu)/2^j \rfloor, \nu}^j \right) \right) \end{aligned}$$

the values in the inner brackets cancel except for the two indices  $\nu = \nu_k^j \in \{0, \dots, 2^j - 1\}$  ( $k = 0, 1$ ) with  $\nu_k^j + 1 + i \equiv 0 \pmod{2^{j-1}}$ . For these indices the signs of  $\varepsilon_j(i + \nu_k)$  and  $\varepsilon_j(i + 1 + \nu_k)$  are opposite. Consequently, for each  $j$ , the inner sum contains only four summands, and we can estimate

$$(5.11) \quad |A_i - A_{i+1}| \leq \sqrt{2} \sum_{j=2}^n 4^{-j} \cdot 4 < \frac{\sqrt{2}}{3}.$$

By inserting this into (5.10), it becomes clear that for an extremal pixel  $f_i$  we get

$$(5.12) \quad \text{sgn}(\dot{u}_i - \dot{u}_{i+1}) = -\text{sgn}(d_{i,+}).$$

We have therefore proven that the difference of an extremal pixel to its right neighbor decreases under infinitesimal soft wavelet shrinkage. Analogous considerations apply to the left neighbor, which completes the proof of (ii).  $\square$

It follows particularly from Proposition 5.1 that under iterated infinitesimal soft wavelet shrinkage no oscillatory (Gibbs-like) artifacts can appear. Any artifact of this type would include at least one local extremum evolving from a flat region which would, for continuity, have to grow over a finite time interval in contradiction to Proposition 5.1.

It should be noted that a single step of infinitesimal shrinkage does not effectively change the signal any more since  $T \rightarrow 0$ . To investigate true changes of the signal by the shrinkage procedure, one has to consider iterated shrinkage. Summing up  $\tau$  over all iteration steps, a “total evolution time”  $t$  is obtained; for fixed  $t$ , the number of iteration steps tends to infinity as  $\tau$  goes to zero. Infinitesimal translation-invariant soft Haar wavelet shrinkage thus becomes a dynamic process parametrized by  $t$ , and Proposition 5.1 describes its behavior at a single point of time.

Of course, this analysis can be extended to a time interval. Then one has to take care of the discontinuity of  $\text{sgn}$  at 0. Similarly, as in the proof of Proposition 4.1, this can be done by splitting the time axis into intervals in which no sign changes of wavelet coefficients occur. However, since once-merged pixels can split again in the process considered here,  $\text{sgn}(0)$  will in most cases occur only in discrete time points.

Now we turn to consider finite-size shrinkage steps  $\tau$ . The ideas used in the proof of Proposition 5.1 can also be applied to analyze soft wavelet shrinkage with finite threshold  $\tau$  by simply replacing the derivatives  $\dot{u}_i$ ,  $\dot{s}_{i,\nu}^j$  by differences  $\Delta u_i := u_i(\tau) - u_i(0)$  and  $\Delta s_{i,\nu}^j := s_{i,\nu}^j - d_{i,\nu}^j$ , respectively. Then we obtain instead of (5.7) that

$$\Delta u_i = \frac{(s_{i,+} - d_{i,+}) - (s_{i,-} - d_{i,-})}{2\sqrt{2}} + A_i,$$

where

$$A_i = \sum_{j=2}^n 2^{-3j/2} \sum_{\nu=0}^{2^j-1} \varepsilon_j(i+\nu) \Delta s_{[(i+\nu)/2^j],\nu}^j.$$

By (5.4) and (2.1) we obtain instead of (5.8) the estimate

$$|A_i| \leq \tau\sqrt{2} \sum_{j=2}^n 2^{-j} < \frac{\tau}{\sqrt{2}}.$$

However, the implication from inequality (5.8) to (5.9) can be transferred only if  $|d_{i,+}| \geq \tau$  and  $|d_{i,-}| \geq \tau$ . Similarly, we conclude instead of (5.10) that

$$(5.13) \quad \Delta u_i - \Delta u_{i+1} = \frac{-\Delta s_{i,-} + 2\Delta s_{i,+} - \Delta s_{i+1,+}}{2\sqrt{2}} + (A_i - A_{i+1})$$

and estimate the latter difference by

$$(5.14) \quad |A_i - A_{i+1}| \leq \tau\sqrt{2} \sum_{j=2}^n 4^{-j} < \frac{\tau\sqrt{2}}{3}.$$

However, the conclusion from (5.11) to (5.12) can be transferred only if  $|\Delta s_{i,-} - 2\Delta s_{i,+} + \Delta s_{i+1,+}| \geq 4\tau/3$ . The latter holds true if (but not only if)  $|d_{i,+}| \geq \tau$  and  $|d_{i,-}| \geq \tau$ , i.e., if  $f_i - f_{i\pm 1} \geq \sqrt{2}\tau$ . In this case we obtain by (5.13), (5.14), and their counterparts for the left neighbors of  $f_i$  that

$$-\frac{\tau\sqrt{2}}{3} \leq u_i(\tau) - u_{i\pm 1}(\tau) \leq f_i - f_{i\pm 1} - \frac{\tau\sqrt{2}}{6}$$

if  $f_i$  is a maximum. Analogous inequalities hold true if  $f_i$  is a minimum. We can therefore state the following corollary.

**COROLLARY 5.2** (behavior of extrema under Haar wavelet shrinkage). *Under translation-invariant soft Haar wavelet shrinkage with thresholds (5.4) an extremal pixel  $f_i$ , which differs at least by  $\sqrt{2}\tau$  from each of its neighbors, evolves as follows:*

- (i) *The value of the extremal pixel decreases, i.e.,  $\Delta u_i < 0$ , if it is a maximum and increases, i.e.,  $\Delta u_i > 0$ , if it is a minimum.*
- (ii) *The absolute value of the difference of the extremal pixel to each of its two neighbors decreases; in particular, one has  $\Delta u_i - \Delta u_{i\pm 1} < 0$  for a maximum and  $\Delta u_i - \Delta u_{i\pm 1} > 0$  for a minimum.*

It can be shown by examples that each of the statements (i) and (ii) of the corollary can be violated if the extremal pixel  $f_i$  differs from its neighbors by not more than  $\sqrt{2}\tau$ . In summary, it follows that Gibbs-like artifacts can in principle still occur under finite-size steps of soft Haar wavelet shrinkage but are restricted in amplitude.

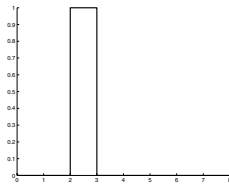
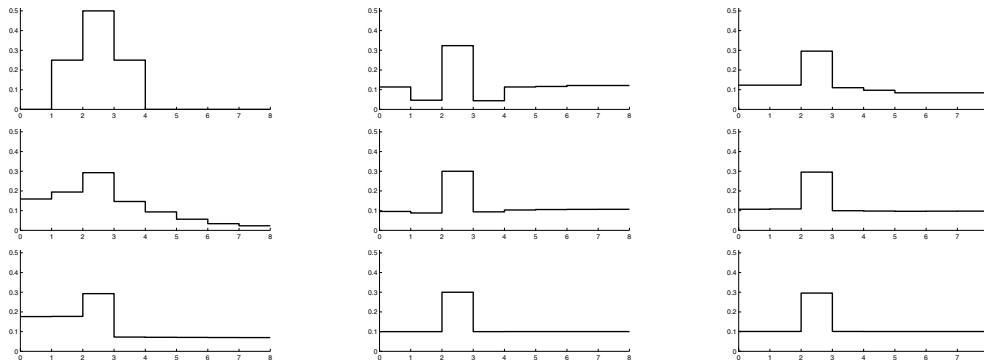
FIG. 5. Test signal with  $N = 8$  pixels.

FIG. 6.  $K = 1, 20, 1000$  iterations (top to bottom) of translation-invariant soft Haar wavelet shrinkage with thresholds  $\tau/K$  applied to the signal in Figure 5. Left column: single-scale wavelet shrinkage with  $\tau = 1$ . Center column: multiscale wavelet shrinkage ( $m = 4$ ) with uniform threshold  $\tau = 0.48$  on all scales. Right column: multiscale wavelet shrinkage ( $m = 4$ ) with  $\tau = 0.585$  and scale-adapted thresholds according to (5.4).

**6. Experiments.** In this section we illustrate the interplay of iterations and multiscale soft Haar wavelet shrinkage by two examples. As in the previous section we consider initial signals  $f = (f_0, \dots, f_{N-1})$ , where  $N = 2^n$  is a power of 2. Furthermore, we restrict our attention to reflecting (Neumann) boundary conditions. Then we can perform multiscale wavelet shrinkage up to some assigned scale  $m \leq n$ .

We start with a simple example which demonstrates the influence of the interplay between iterations and multiscale wavelet shrinkage on Gibbs-like artifacts and its relation to TV diffusion. We consider the initial signal in Figure 5 and apply iterated translation-invariant single-scale and multiscale soft Haar wavelet shrinkage with various threshold parameters. The resulting signals are presented in Figure 6.

Consider the left column of Figure 6. In subsection 4.2 we have shown that translation-invariant soft Haar wavelet shrinkage corresponds to a stable numerical scheme for TV diffusion which represents real TV diffusion if the shrinkage parameter  $\tau$  is small enough. The first row demonstrates the local effect of the single-scale wavelet shrinkage with threshold  $\tau = 1$ . The  $K$ -times iterated processes with thresholds  $\tau = 1/K$  in the second and third rows spread the information globally over the signal. For  $K = 1000$ , the scheme is a very good approximation to TV diffusion.

The middle and the right columns of Figure 6 deal with multiscale wavelet shrinkage which does not fully correspond to TV diffusion. Already a single iteration results in global effects here. Iterating the multiscale wavelet shrinkage flattens homogeneous regions, as desired also in TV diffusion. In the middle column, we can observe Gibbs-like phenomena. In the right column, they are avoided by scaling the thresholds.

In our second example we are concerned with the initial signal in Figure 7 ob-

tained using the *WaveLab* package [31]. Figure 8 presents the denoised signal, where the parameter of each method (threshold value or number of iterations) was chosen to optimize the signal-to-noise ratio on the output. We have applied the following techniques:

- A. 1 level, regularized TV scheme (4.6) with  $\epsilon = \frac{0.04}{2\sqrt{2}}$ , iterated with  $\tau = \frac{0.01}{2\sqrt{2}}$ ,  $K = 53707$  iterations.
- B. 1 level, two-pixel scheme (4.4), iterated with  $\tau = 0.01$ ,  $K = 53707$  iterations.
- C. 13 levels, 1 iteration, uniform threshold  $\tau = 37.4$ .
- D. 13 levels, iterated,  $\tau = 0.01$ ,  $K = 3244$  iterations.
- E. 13 levels, 1 iteration, scaled thresholds,  $\tau = 92.6$ .
- F. 13 levels, iterated,  $\tau = 0.01$ ,  $K = 7800$  iterations.

The best restoration results in terms of the signal-to-noise ratio are obtained using the regularized TV diffusion scheme (A, SNR=24.6dB), iterated single-scale wavelet shrinkage (B, SNR=24.5dB), or the iterated  $n$ -scale wavelet shrinkage with adapted thresholds (F, SNR=24.3dB). Although these methods are not exactly equivalent, they reveal a high level of visual similarity and provide a good piecewise constant approximation to the original signal. The single step multiscale wavelet shrinkage with scale-adapted threshold (E, SNR=21.9dB) performs slightly worse. The single step and iterated multiscale wavelet shrinkage techniques with a uniform threshold on all scales (C, SNR=18.3dB and D, SNR=21.3dB, respectively) are less satisfactory, also visually.

These experiments show that TV denoising outperforms many soft wavelet shrinkage strategies. On the other hand, this is at the expense of a relatively high numerical effort. In order to make wavelets competitive, the shrinkage should be shift invariant, iterative, and use multiple scales with scaled thresholds. In those cases where it is possible to reduce the number of iterations without severe quality degradations, one obtains a hybrid method that combines the speed of multiscale wavelet techniques with the quality of variational or PDE-based denoising methods. For more experiments on multiscale ideas versus iterations we refer the reader to [35].

**7. Summary.** The goal of the present paper was to investigate under which conditions one can prove equivalence between four discontinuity preserving denoising techniques in the 1-D case: soft wavelet thresholding, TV diffusion, TV regularization, and SIDEs. Starting from a simple two-pixel case we were able to derive analytical solutions. These two-pixel solutions have been used for the following purposes:

- They establish equivalence between soft Haar wavelet shrinkage with threshold parameter  $\tau$  and TV diffusion of two-pixel signal pairs with diffusion time  $t = \tau/\sqrt{2}$ .
- They prove also equivalence to TV regularization of two-pixel pairs with regularization parameter  $\alpha = \tau/\sqrt{2}$ .
- They conjecture equivalence of space-discrete TV diffusion and discrete TV regularization for general  $N$ -pixel signals. This conjecture has been proven in subsection 4.3.
- They prove that space-discrete TV diffusion and discrete TV regularization are also equivalent to a SIDE evolution with a TV-based force function. This gives a sound theoretical justification for the heuristically introduced evolution rules for SIDEs.
- They design a novel numerical scheme for TV diffusion of  $N$ -pixel signals. It is based on an additive operator splitting into two-pixel interactions where analytical solutions exist for arbitrary large time step sizes. Thus, the numerical scheme is



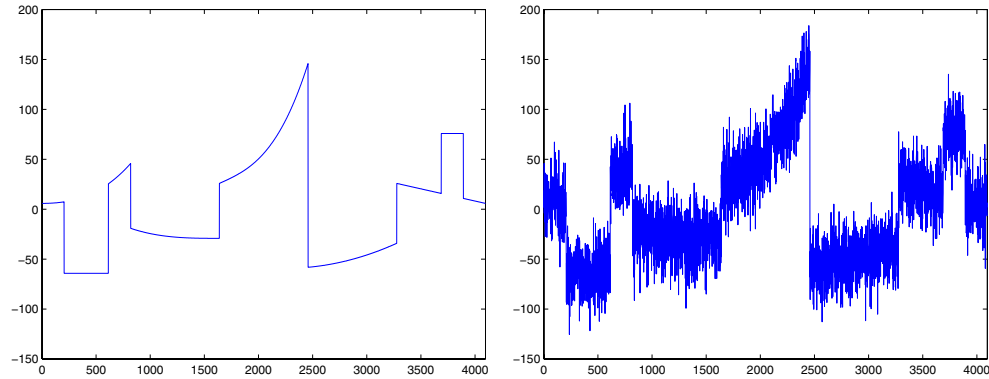


FIG. 7. Piecewise polynomial signal. Left: original. Right: with additive Gaussian white noise ( $\text{SNR} = 8 \text{ dB}$ ) as input for the filtering procedures.

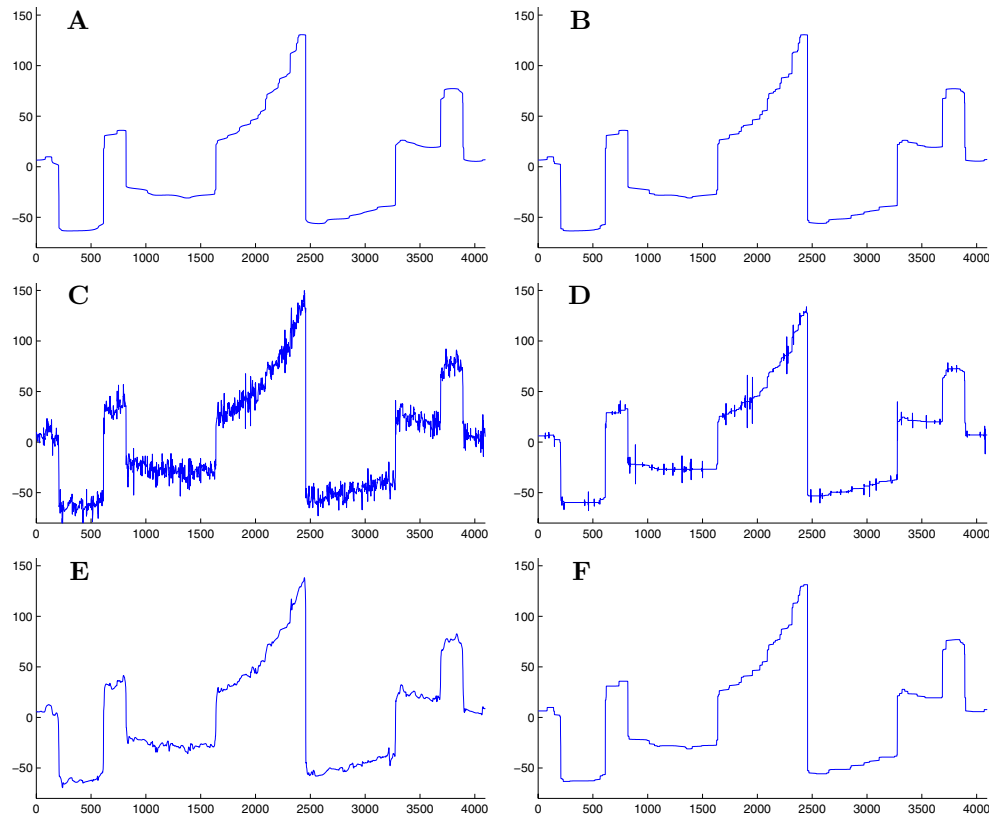


FIG. 8. Optimal filtering results of several variants of procedures based on TV or wavelet filtering when run on the noisy data of Figure 7.

- A. Iterated classical scheme for the regularized TV flow (4.6).
- B. Iterated single-level shrinkage (equivalent to the scheme (4.4) for TV flow).
- C. Multiple levels with a single threshold, single step (i.e., noniterated).
- D. Iterated multiple level with a single threshold at each of the levels.
- E. Multiple levels with thresholds scaled according to (5.4), single step.
- F. Iterated multiple level with scaled thresholds.

See text for the explanation and numerical evaluation of the results.

explicit and absolutely stable.

We showed that wavelet shrinkage on multiple scales can also be regarded as two-pixel TV diffusion or regularization on the Laplacian pyramid of the signal. On the wavelet side, our experiments show that one can improve the denoising performance by rescaling the thresholds for each wavelet level and by iterating the translation-invariant wavelet shrinkage. On the PDE side, it is possible to achieve a speed-up without significant quality deterioration by using iterated multiple scales instead of iterated single-scale denoising. Thus, the resulting hybrid methods combine the advantages of wavelet and PDE-based denoising.

In our future work we intend to consider more advanced wavelet methods (other shrinkage functions, different wavelets) and to analyze the multidimensional case. In two dimensions, first results on diffusion-inspired wavelet shrinkage with improved rotation invariance are presented in [34]. We will also consider extensions of the numerical two-pixel schemes for TV diffusion.

**Acknowledgment.** Joachim Weickert thanks Stephen Keeling (Graz, Austria) for interesting discussions on two-pixel signals.

#### REFERENCES

- [1] R. ACAR AND C. R. VOGEL, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.
- [2] L. ALVAREZ, P.-L. LIONS, AND J.-M. MOREL, *Image selective smoothing and edge detection by nonlinear diffusion. II*, SIAM J. Numer. Anal., 29 (1992), pp. 845–866.
- [3] F. ANDREU, C. BALLESTER, V. CASELLES, AND J. M. MAZÓN, *Minimizing total variation flow*, Differential Integral Equations, 14 (2001), pp. 321–360.
- [4] F. ANDREU, V. CASELLES, J. I. DIAZ, AND J. M. MAZÓN, *Qualitative properties of the total variation flow*, J. Funct. Anal., 188 (2002), pp. 516–547.
- [5] G. AUBERT AND P. KORNPBST, *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*, Appl. Math. Sci. 147, Springer, New York, 2002.
- [6] G. AUBERT AND L. VESE, *A variational method in image recovery*, SIAM J. Numer. Anal., 34 (1997), pp. 1948–1979.
- [7] Y. BAO AND H. KRIM, *Towards bridging scale-space and multiscale frame analyses*, in Wavelets in Signal and Image Analysis, Comput. Imaging Vision 19, A. A. Petrosian and F. G. Meyer, eds., Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001.
- [8] G. BELLETTINI, V. CASELLES, AND M. NOVAGA, *The total variation flow in  $R^N$* , J. Differential Equations, 184 (2002), pp. 475–525.
- [9] A. BLAKE AND A. ZISSERMAN, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.
- [10] T. BROX, M. WELK, G. STEIDL, AND J. WEICKERT, *Equivalence results for TV diffusion and TV regularisation*, in Scale-Space Methods in Computer Vision, Lecture Notes in Comput. Sci. 2695, L. D. Griffin and M. Lillholm, eds., Springer, Berlin, 2003, pp. 86–100.
- [11] P. J. BURT AND E. H. ADELSON, *The Laplacian pyramid as a compact image code*, IEEE Trans. Comm., 31 (1983), pp. 532–540.
- [12] E. J. CANDÉS AND F. GUO, *New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction*, Signal Process., 82 (2002), pp. 1519–1543.
- [13] A. CHAMBOLLE, R. A. DEVORE, N. LEE, AND B. L. LUCIER, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image Process., 7 (1998), pp. 319–335.
- [14] A. CHAMBOLLE AND P.-L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.
- [15] A. CHAMBOLLE AND B. L. LUCIER, *Interpreting translationally-invariant wavelet shrinkage as a new image smoothing scale space*, IEEE Trans. Image Process., 10 (2001), pp. 993–1000.
- [16] T. F. CHAN AND H. M. ZHOU, *Total variation improved wavelet thresholding in image compression*, in Proceedings of the Seventh International Conference on Image Processing, Vancouver, BC, Canada, 2000.
- [17] A. COHEN, W. DAHMEN, I. DAUBECHIES, AND R. DEVORE, *Harmonic analysis in the space BV*,

- Rev. Mat. Iberoamericana, 19 (2003), pp. 235–262.
- [18] A. COHEN, R. DEVORE, P. PETRUSHEV, AND H. XU, *Nonlinear approximation and the space  $BV(R^2)$* , Amer. J. Math., 121 (1999), pp. 587–628.
- [19] R. R. COIFMAN AND D. DONOHO, *Translation invariant denoising*, in *Wavelets in Statistics*, A. Antoine and G. Oppenheim, eds., Springer, New York, 1995, pp. 125–150.
- [20] R. R. COIFMAN AND A. SOWA, *Combining the calculus of variations and wavelets for image enhancement*, Appl. Comput. Harmon. Anal., 9 (2000), pp. 1–18.
- [21] R. R. COIFMAN AND A. SOWA, *New methods of controlled total variation reduction for digital functions*, SIAM J. Numer. Anal., 39 (2001), pp. 480–498.
- [22] D. L. DONOHO, *De-noising by soft thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.
- [23] D. L. DONOHO AND I. M. JOHNSTONE, *Minimax estimation via wavelet shrinkage*, Ann. Statist., 26 (1998), pp. 879–921.
- [24] S. DURAND AND J. FROMENT, *Reconstruction of wavelet coefficients using total variation minimization*, SIAM J. Sci. Comput., 24 (2003), pp. 1754–1767.
- [25] X. FENG AND A. PROHL, *Analysis of total variation flow and its finite element approximations*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 533–556.
- [26] A. F. FILIPPOV, *Differential Equations with Discontinuous Righthand Sides*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1988.
- [27] A. S. FRANGAKIS, A. STOSCHEK, AND R. HEGERL, *Wavelet transform filtering and nonlinear anisotropic diffusion assessed for signal reconstruction performance on multidimensional biomedical data*, IEEE Transactions on Biomedical Engineering, 48 (2001), pp. 213–222.
- [28] W. HINTERBERGER, M. HINTERMÜLLER, K. KUNISCH, M. VON OEHSSEN, AND O. SCHERZER, *Tube methods for BV regularization*, J. Math. Imaging Vision, 19 (2003), pp. 219–235.
- [29] T. LU, P. NEITTAANMÄKI, AND X.-C. TAI, *A parallel splitting up method and its application to Navier–Stokes equations*, Appl. Math. Lett., 4 (1991), pp. 25–29.
- [30] F. MALGOUYRES, *Mathematical analysis of a model which combines total variation and wavelet for image restoration*, Inverse Problems, 2 (2002), pp. 1–10.
- [31] S. MALLAT, *A Wavelet Tour of Signal Processing*, 2nd ed., Academic Press, San Diego, 1999.
- [32] E. MAMMEN AND S. VAN DE GEER, *Locally adaptive regression splines*, Ann. Statist., 25 (1997), pp. 387–413.
- [33] Y. MEYER, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, Univ. Lecture Ser. 22, AMS, Providence, RI, 2001.
- [34] P. MRÁZEK AND J. WEICKERT, *Rotationally invariant wavelet shrinkage*, in *Pattern Recognition*, Lecture Notes in Comput. Sci. 2781, B. Michaelis and G. Krell, eds., Springer, Berlin, 2003, pp. 156–163.
- [35] P. MRÁZEK, J. WEICKERT, G. STEIDL, AND M. WELK, *On iterations and scales of nonlinear filters*, in *Proceedings of the Eighth Computer Vision Winter Workshop*, O. Drbohlav, ed., Valtice, Czech Republic, 2003, Czech Pattern Recognition Society, Peršlák, Czech Republic, 2003, pp. 61–66.
- [36] N. NORDSTRÖM, *Biased anisotropic diffusion – A unified regularization and diffusion approach to edge detection*, Image and Vision Computing, 8 (1990), pp. 318–327.
- [37] 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.
- [38] I. POLLAK, A. S. WILSKY, AND H. KRIM, *Image segmentation and edge enhancement with stabilized inverse diffusion equations*, IEEE Trans. Image Process., 9 (2000), pp. 256–266.
- [39] J. L. RICHARDSON, R. C. FERRELL, AND L. N. LONG, *Unconditionally stable explicit algorithms for nonlinear fluid dynamics problems*, J. Comput. Phys., 104 (1993), pp. 69–74.
- [40] L. I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
- [41] G. SAPIRO, *Geometric Partial Differential Equations and Image Analysis*, Cambridge University Press, Cambridge, UK, 2001.
- [42] O. SCHERZER AND J. WEICKERT, *Relations between regularization and diffusion filtering*, J. Math. Imaging Vision, 12 (2000), pp. 43–63.
- [43] J. SHEN AND G. STRANG, *On wavelet fundamental solutions to the heat equation—Heatlets*, J. Differential Equations, 161 (2000), pp. 403–421.
- [44] G. STEIDL AND J. WEICKERT, *Relations between soft wavelet shrinkage and total variation denoising*, in *Pattern Recognition*, Lecture Notes in Comput. Sci. 2449, L. Van Gool, ed., Springer, Berlin, 2002, pp. 198–205.
- [45] 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*, Tech. Report 26, Series SPP-1114, Department of Mathematics, University of Bremen,

- Bremen, Germany, 2003.
- [46] D. M. STRONG, *Adaptive Total Variation Minimizing Image Restoration*, Ph.D. thesis, Department of Mathematics, University of California, Los Angeles, CA, 1997.
  - [47] J. WEICKERT, *Anisotropic Diffusion in Image Processing*, Teubner, Stuttgart, 1998.
  - [48] J. WEICKERT, B. M. TER HAAR ROMENY, AND M. A. VIERGEVER, *Efficient and reliable schemes for nonlinear diffusion filtering*, IEEE Trans. Image Process., 7 (1998), pp. 398–410.