

A Multiscale Wavelet–Inspired Scheme for Nonlinear Diffusion

GERLIND PLONKA¹ AND GABRIELE STEIDL²

¹ Department of Mathematics, University of Duisburg-Essen, Campus Duisburg, 47048
Duisburg, Germany

plonka@math.uni-duisburg.de

² Faculty of Mathematics and Computer Science, A5, University of Mannheim, 68131
Mannheim, Germany

steidl@math.uni-mannheim.de

Abstract

Nonlinear diffusion filtering and wavelet shrinkage are two successfully applied methods for discontinuity preserving denoising of signals and images. Recently, relations between both methods have been established taking into account wavelet shrinkage at one scale. In this paper, we propose a new explicit scheme for nonlinear diffusion which directly incorporates ideas from multiscale Haar wavelet shrinkage. We prove that our scheme permits larger time steps while preserving convergence to the mean signal value. Numerical examples demonstrate the behavior of our scheme for two and three scales.

Mathematics Subject Classification 2000. 65T60, 65M06, 65M12, 94A12.

Key words. Nonlinear diffusion, Haar wavelet filter bank, signal denoising, wavelet shrinkage, nonlinear diffusion filtering.

1 Introduction

Nonlinear diffusion filtering has become a powerful tool in signal and image restoration, e.g., for denoising purposes. The choice of nonlinear diffusion filters leads to impressive results by removing insignificant, small-scale variations while preserving important features such as discontinuities, see [4, 10, 14, 16] and the references therein. On the other hand, wavelet shrinkage methods were also successfully applied for removing noise without sacrificing important structures such as edges, see [1, 5, 7, 11] and the references therein.

Recently, the connections between explicit discrete one-dimensional schemes for nonlinear diffusion and shift-invariant Haar wavelet shrinkage were established in [15, 12]. The authors showed that one step of a (stabilized) explicit discretization of nonlinear diffusion can be expressed in terms of wavelet shrinkage on a single spatial level. This equivalence allows a fruitful exchange of ideas between the two fields. For example, new wavelet shrinkage functions can be derived from existing diffusivity functions and vice versa leading to explicit schemes with well understood stability properties. These considerations were generalized to two spatial dimensions in [13, 18]. In particular, diffusion inspired wavelet shrinkage schemes with improved rotation invariance were proposed.

However, while these approaches rely on *one scale* of wavelet shrinkage, wavelet filter banks are usually applied with multiple scales. Of course, multiscale shrinkage with Haar wavelets can be interpreted as application of the one scale nonlinear explicit diffusion scheme to hierarchical signals [15]. In this paper, we propose a wavelet inspired explicit diffusion scheme which directly incorporates *multiple scales*. The new discretization admits larger time steps than the original explicit scheme and is therefore more efficient. It has been shown already in [2] that translation-invariant wavelet shrinkage can be interpreted as a smoothing scale-space. We prove that the new explicit scheme satisfies a lot of discrete scale-space criteria as the average level invariance, a certain extremum principle and convergence to spatial average. Note that the ideas for proving scale-space properties from [16] can not be applied here. Compared to the model of discrete diffusion filtering presented in [16], our iteration matrices do not satisfy all conditions given there. In particular, the nonnegativity of all elements of the iteration matrix may be violated.

In this paper, we restrict our attention to the one-dimensional spatial setting. Moreover, we do not take semi-implicit discretization schemes into account which do not suffer from time step restrictions, but require to solve a linear system of equations, see, e.g., [17, 16].

This paper is organized as follows. Section 2 reviews the relation between explicit schemes for nonlinear diffusion and shift-invariant one-scale Haar wavelet shrinkage in a matrix-vector notation. We extend this scheme to multiple scales in Section 3. This results in a new explicit diffusion scheme which allows larger time steps while preserving convergence. Section 4 deals with convergence and stability properties of our new scheme. Finally, we present numerical results in Section 5.

2 Explicit discretization of the filter process

We consider the diffusion process

$$u_t = (g(|u_x|)u_x)_x \quad \text{on} \quad (0, N) \times (0, \infty) \quad (2.1)$$

with a given noisy signal f as initial state

$$u(x, 0) = f(x) \quad x \in [0, N],$$

and periodic boundary conditions

$$u(0, t) = u(N, t), \quad u_x(0, t) = u_x(N, t).$$

Here subscripts t and x denote partial derivatives. The time t is a scale parameter. Increasing t corresponds to stronger filtering. The diffusivity function $g(|u_x|)$ is a non-negative function that determines the amount of diffusion. It is decreasing in $|u_x|$ in order to ensure that strong edges are hardly blurred by the diffusion filter while small variations (noise) are smoothed much stronger.

In this paper, we restrict our attention to bounded diffusivities g . Without loss of generality let

$$0 < g(x) \leq 1, \quad x \in \mathbb{R}. \quad (2.2)$$

Frequently applied bounded diffusivities are

$$\begin{aligned} \text{Perona - Malik - diffusivity } g(x) &:= \frac{1}{1 + x^2/\lambda^2} \quad (\text{see [14]}), \\ \text{Charbonnier - diffusivity } g(x) &:= \frac{1}{\sqrt{1 + x^2/\lambda^2}} \quad (\text{see [3]}). \end{aligned}$$

For further diffusivity functions which can be applied in this context, see, e.g., [12].

In this paper, we restrict our attention to discrete signals $\mathbf{f} = (f_0, \dots, f_{N-1})^T$ representing, e.g., the sampled values of a function f at the locations $x_i = i$, $i = 0, \dots, N-1$. Moreover, the periodic boundary conditions are taken into account by considering spatial indices modulo N . Replacing the spatial derivative in (2.1) by simple forward differences, we obtain

$$u'_i(t) = g(u_{i+1}(t) - u_i(t)) (u_{i+1}(t) - u_i(t)) - g(u_i(t) - u_{i-1}(t)) (u_i(t) - u_{i-1}(t)),$$

where $u_i(t)$ approximates $u(x_i, t)$. For the time discretization we apply the Euler-forward scheme with time step size τ . Using $t_k := k\tau$, $k \in \mathbb{N}_0$ and $u_i^{(k)}$ as approximation of $u(x_i, t_k)$, we obtain

$$\frac{u_i^{(k+1)} - u_i^{(k)}}{\tau} = g(u_{i+1}^{(k)} - u_i^{(k)}) (u_{i+1}^{(k)} - u_i^{(k)}) - g(u_i^{(k)} - u_{i-1}^{(k)}) (u_i^{(k)} - u_{i-1}^{(k)}).$$

Hence, the approximate solutions $u_i^{(k+1)}$, $i = 0, \dots, N-1$ at time step $(k+1)\tau$ can be computed from $u_i^{(k)}$, $i = 0, \dots, N-1$ by

$$u_i^{(k+1)} = u_i^{(k)} + \tau \left(g(u_{i+1}^{(k)} - u_i^{(k)}) (u_{i+1}^{(k)} - u_i^{(k)}) - g(u_i^{(k)} - u_{i-1}^{(k)}) (u_i^{(k)} - u_{i-1}^{(k)}) \right)$$

with initial conditions $u_i^{(0)} := f_i$, $i = 0, \dots, N-1$.

Set $\mathbf{u}^{(k)} := (u_0^{(k)}, \dots, u_{N-1}^{(k)})^T$ and let \mathbf{I} be the identity matrix of order N and \mathbf{V} the circulant shift matrix

$$\mathbf{V} := \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & 0 \\ & & & \ddots & \\ 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & & 0 \end{pmatrix}$$

with $\mathbf{V}\mathbf{x} = (x_1, x_2, \dots, x_{N-1}, x_0)^T$ for $\mathbf{x} := (x_i)_{i=0}^{N-1}$. Then we can write the explicit scheme in matrix-vector notation as

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \tau(\mathbf{I} - \mathbf{V}^T) \mathbf{D}^{(k)} (\mathbf{I} - \mathbf{V}) \mathbf{u}^{(k)}, \quad (2.3)$$

where the diagonal *diffusivity matrix* is given by

$$\mathbf{D}^{(k)} := \mathbf{D} \left((\mathbf{I} - \mathbf{V}) \mathbf{u}^{(k)} \right)$$

and

$$\mathbf{D}(\mathbf{y}) := \text{diag}(g(|y_0|), \dots, g(|y_{N-1}|)).$$

In summary, one iteration step of the explicit scheme is of the form $\mathbf{u}^{(k+1)} = \mathbf{A}^{(k)} \mathbf{u}^{(k)}$ with the iteration matrix

$$\mathbf{A}^{(k)} = \mathbf{A}^{(k)}(\mathbf{u}^{(k)}) := \mathbf{I} - \tau(\mathbf{I} - \mathbf{V}^T) \mathbf{D}^{(k)} (\mathbf{I} - \mathbf{V}).$$

Let us briefly explain the connection with translation-invariant Haar wavelet shrinkage investigated in [12]. We consider the Haar wavelet filter bank with a single scale shown in Figure 1. Applying the z -transform to the discrete signal $\mathbf{u}^{(k)} = (u_i^{(k)})_{i=0}^{N-1}$ we obtain $u^{(k)}(z) = \sum_{i=0}^{N-1} u_i^{(k)} z^{-i}$. Further, let $H(z) := (1+z)/\sqrt{2}$ and $G(z) := (1-z)/\sqrt{2}$ be the low-pass and the high-pass filter of the filter bank, $2 \downarrow$ and $2 \uparrow$ denote downsampling and upsampling by 2 and S the shrinkage function.

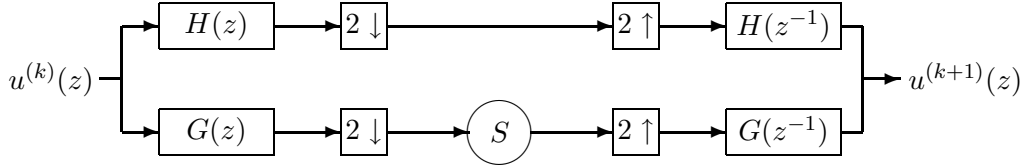


Figure 1: Two-channel filter bank with $H(z) = \frac{1+z}{\sqrt{2}}$ and $G(z) = \frac{1-z}{\sqrt{2}}$.

We obtain

$$u^{(k)}(z)H(z) = \sum_{i=0}^{N-1} \frac{u_i^{(k)} + u_{i+1}^{(k)}}{\sqrt{2}} z^{-i}, \quad u^{(k)}(z)G(z) = \sum_{i=0}^{N-1} \frac{u_i^{(k)} - u_{i+1}^{(k)}}{\sqrt{2}} z^{-i},$$

such that downsampling, application of S , and upsampling lead to

$$u^{(k+1)}(z) = H(z^{-1}) \sum_{i=0}^{N/2-1} \frac{u_{2i}^{(k)} + u_{2i+1}^{(k)}}{\sqrt{2}} z^{-2i} + G(z^{-1}) \sum_{i=0}^{N/2-1} S\left(\frac{u_{2i}^{(k)} - u_{2i+1}^{(k)}}{\sqrt{2}}\right) z^{-2i}.$$

Identifying z with the shift matrix \mathbf{V} , the convolution of $\mathbf{u}^{(k)}$ with the filter H yields $H(\mathbf{V}) \mathbf{u}^{(k)}$, where $H(\mathbf{V}) := \frac{1}{\sqrt{2}}(\mathbf{I} + \mathbf{V})$. Similarly, the convolution with the filter G can be written as multiplication with $G(\mathbf{V}) := \frac{1}{\sqrt{2}}(\mathbf{I} - \mathbf{V})$. Downsampling is represented by multiplication with the matrix

$$\mathbf{P} := \begin{pmatrix} 1 & 0 & 0 & & \\ 0 & 0 & 1 & 0 & 0 \\ & & & & \\ 0 & & & 0 & 0 & 1 & 0 \end{pmatrix} \in \mathbb{R}^{N/2 \times N}$$

and upsampling by multiplication with \mathbf{P}^T . In summary, we obtain the matrix-vector representation of the filter bank

$$\mathbf{u}^{(k+1)} = H(\mathbf{V}^T) \mathbf{P}^T \mathbf{P} H(\mathbf{V}) \mathbf{u}^{(k)} + G(\mathbf{V}^T) \mathbf{P}^T \mathbf{S} \left(\mathbf{P} G(\mathbf{V}) \mathbf{u}^{(k)} \right),$$

where $\mathbf{S}(\mathbf{y}) := (S(y_0), \dots, S(y_{N/2-1}))$.

A translation-invariant filter bank can be simply interpreted as a filter bank without downsampling and upsampling, where the result is "normalized" by the factor 1/2, i.e.,

$$\mathbf{u}^{(k+1)} = \frac{1}{2} H(\mathbf{V}^T) H(\mathbf{V}) \mathbf{u}^{(k)} + \frac{1}{2} G(\mathbf{V}^T) \mathbf{S}^{(k)}, \quad (2.4)$$

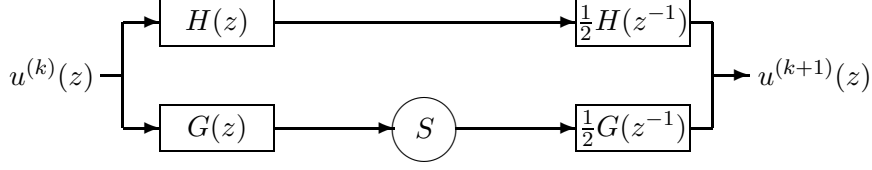


Figure 2: Nonsubsampled two-channel filter bank with $H(z) = \frac{1+z}{\sqrt{2}}$ and $G(z) = \frac{1-z}{\sqrt{2}}$.

where $\mathbf{S}^{(k)} := \mathbf{S}(G(\mathbf{V})\mathbf{u}^{(k)})$ and the shrinkage operator is again applied componentwise. The translation-invariant filter bank is shown in Figure 2.

Since

$$\frac{1}{2} (H(\mathbf{V}^T) H(\mathbf{V}) + G(\mathbf{V}^T) G(\mathbf{V})) = \mathbf{I} \quad (2.5)$$

which indeed holds true for every filter bank with *perfect reconstruction property*, we can rewrite (2.4) as

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \frac{1}{2} G(\mathbf{V}^T) (G(\mathbf{V})\mathbf{u}^{(k)} - \mathbf{S}^{(k)}).$$

Comparison with (2.3) leads with $\mathbf{x}^{(k)} := G(\mathbf{V})\mathbf{u}^{(k)} = \frac{1}{\sqrt{2}}(\mathbf{I} - \mathbf{V})\mathbf{u}^{(k)}$ to

$$\mathbf{x}^{(k)} - \mathbf{S}(\mathbf{x}^{(k)}) = 4\tau \mathbf{D} \left(\sqrt{2}\mathbf{x}^{(k)} \right) \mathbf{x}^{(k)}.$$

Consequently, we have that

$$S(x) = x (1 - 4\tau g(\sqrt{2}x)).$$

This connection between diffusivity and shrinkage function was established in [12].

The diffusivity functions considered before correspond to the following shrinkage functions:

$$\begin{aligned} \text{Perona-Malik diffusivity} &\Rightarrow S(x) = x \left(1 - \frac{4\tau\lambda^2}{\lambda^2 + 2x^2} \right), \\ \text{Charbonnier diffusivity} &\Rightarrow S(x) = x \left(1 - \frac{4\tau\lambda}{\sqrt{\lambda^2 + 2x^2}} \right). \end{aligned}$$

The diffusivities and corresponding shrinkage functions are depicted in Figure 3, cf. [12].

3 The new explicit scheme

In general, wavelet shrinkage is applied to multiple scales. We want to use the idea of multilevel wavelet shrinkage to develop a new explicit diffusion scheme. Let us consider two levels of translation-invariant wavelet shrinkage as described in Figure 4. Since we can apply different shrinkage functions at different levels, we introduce S_i , ($i = 1, 2$).

Translation-invariant Haar wavelet shrinkage at two scales produces for a given input vector $\mathbf{u}^{(k)}$ the output vector

$$\mathbf{u}^{(k+1)} = \frac{1}{4} H(\mathbf{V}^T) H((\mathbf{V}^2)^T) H(\mathbf{V}^2) H(\mathbf{V}) \mathbf{u}^{(k)} + \frac{1}{4} H(\mathbf{V}^T) G((\mathbf{V}^2)^T) \mathbf{S}_2^{(k)} + \frac{1}{2} G(\mathbf{V}^T) \mathbf{S}_1^{(k)},$$

where

$$\mathbf{S}_2^{(k)} := \mathbf{S}_2(G(\mathbf{V}^2) H(\mathbf{V}) \mathbf{u}^{(k)}) = \left(S_2 \left(\frac{1}{2}(u_i^{(k)} + u_{i+1}^{(k)} - u_{i+2}^{(k)} - u_{i+3}^{(k)}) \right) \right)_{i=0}^{N-1},$$

and $\mathbf{S}_1^{(k)} := \mathbf{S}_1(G(\mathbf{V})\mathbf{u}^{(k)})$ as before. Using the perfect reconstruction property (2.5) this can be rewritten as

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \frac{1}{2} G(\mathbf{V}^T) \left(G(\mathbf{V})\mathbf{u}^{(k)} - \mathbf{S}_1^{(k)} \right) - \frac{1}{4} H(\mathbf{V}^T) G((\mathbf{V}^2)^T) \left(H(\mathbf{V}) G(\mathbf{V}^2) \mathbf{u}^{(k)} - \mathbf{S}_2^{(k)} \right). \quad (3.1)$$

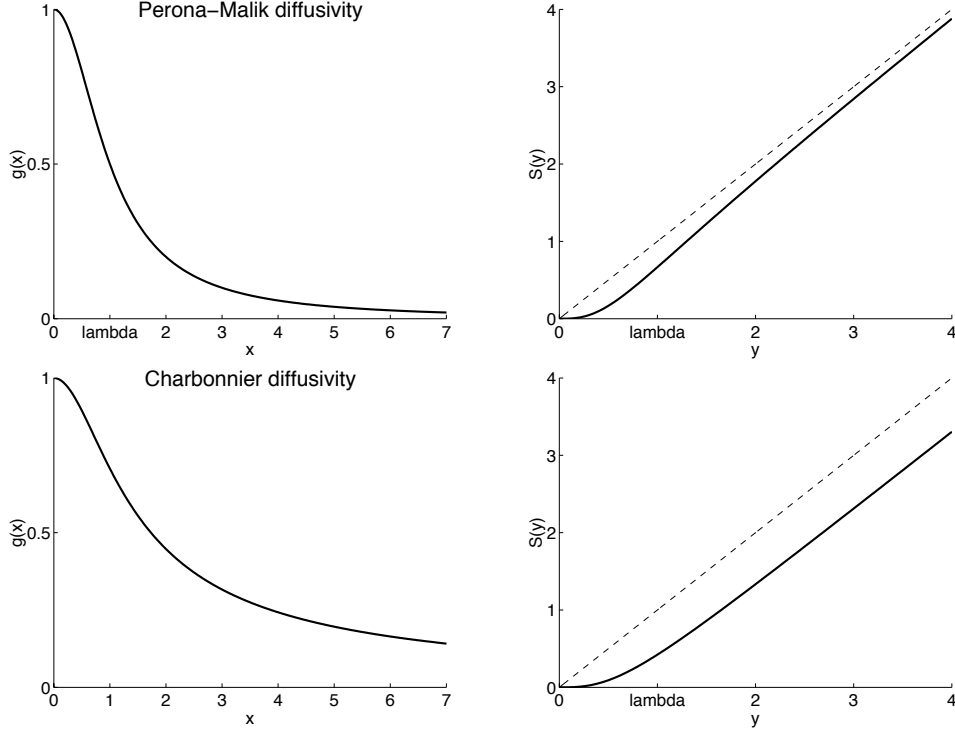


Figure 3: **Left:** diffusivity functions, **Right:** corresponding shrinkage functions for Perona-Malik (top) and Charbonnier (bottom). The functions are plotted for $\tau = 0.25$, $\lambda = 1$.

In order to see how (3.1) can be interpreted as diffusion scheme we verify by Taylor expansion that both $(\mathbf{I} - \mathbf{V})\mathbf{u}$ and $\frac{1}{4}(\mathbf{I} + \mathbf{V})(\mathbf{I} - \mathbf{V}^2)\mathbf{u}$ are consistent discretizations of u_x at i , $i = 0, \dots, N-1$ of order 1. Now we consider the shrinkage parts as suitable splitting of the discretization of u_x into these parts. More precisely, we replace the shrinkage parts by

$$\begin{aligned} \mathbf{x}^{(k)} - \mathbf{S}_1(\mathbf{x}^{(k)}) &= \alpha_1 4 \tau \mathbf{D} \left(\sqrt{2} \mathbf{x}^{(k)} \right) \mathbf{x}^{(k)} = \alpha_1 4 \tau \mathbf{D}_1^{(k)} \mathbf{x}^{(k)}, \\ \mathbf{y}^{(k)} - \mathbf{S}_2(\mathbf{y}^{(k)}) &= \alpha_2 \tau \mathbf{D} \left(\mathbf{y}^{(k)} / 2 \right) \mathbf{y}^{(k)} = \alpha_2 \tau \mathbf{D}_2^{(k)} \mathbf{y}^{(k)}, \end{aligned}$$

where $\mathbf{x}^{(k)} := G(\mathbf{V})\mathbf{u}^{(k)}$, $\mathbf{y}^{(k)} := H(\mathbf{V})G(\mathbf{V}^2)\mathbf{u}^{(k)}$ and $\alpha_1 + \alpha_2 = 1$. Then (3.1) becomes

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \tau \left(\alpha_1 (\mathbf{I} - \mathbf{V}^T) \mathbf{D}_1^{(k)} (\mathbf{I} - \mathbf{V}) + \frac{\alpha_2}{16} (\mathbf{I} + \mathbf{V}^T) (\mathbf{I} - (\mathbf{V}^2)^T) \mathbf{D}_2^{(k)} (\mathbf{I} - \mathbf{V}^2) (\mathbf{I} + \mathbf{V}) \right) \mathbf{u}^{(k)}.$$

This can be rewritten as

$$\mathbf{u}^{(k+1)} = \mathbf{u}^{(k)} - \tau (\mathbf{I} - \mathbf{V}^T) \left[\alpha_1 \mathbf{D}_1^{(k)} + \frac{\alpha_2}{16} (\mathbf{I} + \mathbf{V}^T)^2 \mathbf{D}_2^{(k)} (\mathbf{I} + \mathbf{V})^2 \right] (\mathbf{I} - \mathbf{V}) \mathbf{u}^{(k)}. \quad (3.2)$$

These considerations can simply be generalized to more scales. Let $\sum_{j=1}^n \alpha_j = 1$. For $j \geq 2$ we introduce the following explicit discretization of the diffusion process (2.1) with initial condition $\mathbf{u}^{(0)} = \mathbf{f}$ and periodic boundary conditions,

$$\begin{aligned} \mathbf{u}^{(k+1)} &= \mathbf{u}^{(k)} - \tau \left[\alpha_1 (\mathbf{I} - \mathbf{V}^T) \mathbf{D}_1^{(k)} (\mathbf{I} - \mathbf{V}) \right. \\ &\quad \left. + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \left(\prod_{l=0}^{j-2} (\mathbf{I} + (\mathbf{V}^{2^l})^T) \right) (\mathbf{I} - (\mathbf{V}^{2^{j-1}})^T) \mathbf{D}_j^{(k)} (\mathbf{I} - \mathbf{V}^{2^{j-1}}) \left(\prod_{l=0}^{j-2} (\mathbf{I} + \mathbf{V}^{2^l}) \right) \right] \mathbf{u}^{(k)} \\ &= \mathbf{u}^{(k)} - \tau (\mathbf{I} - \mathbf{V}^T) \left[\alpha_1 \mathbf{D}_1^{(k)} \right. \end{aligned}$$

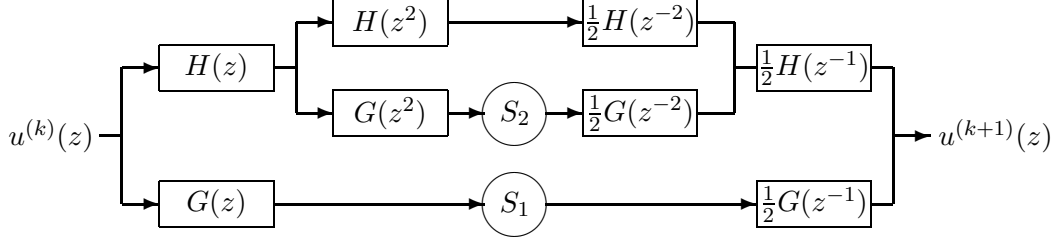


Figure 4: Two scales of translation-invariant Haar wavelet shrinkage with $H(z) = \frac{1+z}{\sqrt{2}}$ and $G(z) = \frac{1-z}{\sqrt{2}}$.

$$+ \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \left(\prod_{l=0}^{j-2} (\mathbf{I} + (\mathbf{V}^{2^l})^T)^2 \right) \mathbf{D}_j^{(k)} \left(\prod_{l=0}^{j-2} (\mathbf{I} + \mathbf{V}^{2^l})^2 \right) (\mathbf{I} - \mathbf{V}) \mathbf{u}^{(k)}. \quad (3.3)$$

with diagonal diffusivity matrices

$$\begin{aligned} \mathbf{D}_j^{(k)} &:= \mathbf{D} \left(\frac{1}{4^{j-1}} (\mathbf{I} - \mathbf{V}^{2^{j-1}}) \left(\prod_{l=0}^{j-2} (\mathbf{I} + \mathbf{V}^{2^l}) \right) \mathbf{u}^{(k)} \right) \\ &= \text{diag} \left(g \left(\frac{1}{4^{j-1}} \sum_{l=0}^{2^{j-1}-1} (u_{i+l}^{(k)} - u_{i+l+2^{j-1}}^{(k)}) \right) \right)_{i=0}^{N-1}, \quad j = 1, \dots, n. \end{aligned} \quad (3.4)$$

For $n = 1$, this is just (2.3). For $n = 2$, equation (3.3) coincides with (3.2).

The consistency of the discretization can be seen as follows. Considering the i -th component ($i = 0, \dots, N-1$) of

$$\frac{1}{4^{j-1}h} (\mathbf{I} - \mathbf{V}^{2^{j-1}}) \prod_{l=0}^{j-2} (\mathbf{I} + \mathbf{V}^{2^l}) \mathbf{u}^{(k)}$$

we obtain by Taylor expansion assuming a sufficiently smooth function $u^{(k)}$ that

$$\begin{aligned} &\frac{1}{4^{j-1}h} \sum_{l=0}^{2^{j-1}-1} u^{(k)}((i+l)h) - u^{(k)}((i+l+2^{j-1})h) \\ &= \frac{1}{4^{j-1}h} \sum_{l=0}^{2^{j-1}-1} (u^{(k)}(ih) + u_x^{(k)}(ih)lh + \mathcal{O}(h^2)) - (u^{(k)}(ih) + u_x^{(k)}(ih)(l+2^{j-1})h + \mathcal{O}(h^2)) \\ &= -u_x^{(k)}(ih) + \mathcal{O}(h). \end{aligned}$$

Analogously, for a sufficiently smooth periodic function w , the vector

$$\frac{1}{4^{j-1}h} (\mathbf{I} - (\mathbf{V}^{2^{j-1}})^T) \prod_{l=0}^{j-2} (\mathbf{I} - \mathbf{V}^{2^l}) \mathbf{w}$$

with $\mathbf{w} := (w(ih))_{i=0}^{N-1}$ approximates $\mathbf{w}_x := (w_x(ih))_{i=0}^{N-1}$ by backward differences.

The new discretization scheme (3.3) can be interpreted as follows. While the first term $\tau \alpha_1 (\mathbf{I} - \mathbf{V}^T) \mathbf{D}_1^{(k)} (\mathbf{I} - \mathbf{V}) \mathbf{u}^{(k)}$ is the weighted discretization (2.3), the further terms act more globally and smooth the signal stronger. In the following, we will see that we can use larger time steps τ for appropriately chosen parameters α_j ($j = 1, \dots, n$). In other words, our wavelet inspired scheme needs much less iterations than the original discretization scheme (2.3) in order to achieve an equivalent denoising result.

4 Properties of the new explicit scheme

It is well-known that the solution of the continuous diffusion process (2.1) converges to the spatial average of the initial data f for $t \rightarrow \infty$, see, e.g., [16]. Therefore, the explicit discretization should also have this property. In this section, we derive conditions for the parameters $\alpha_1, \dots, \alpha_n$ and the time step τ such that the iteration process (3.3) converges for $k \rightarrow \infty$ to the average of the initial discrete signal $\mathbf{f} = (f_0, \dots, f_{N-1})^T$, i.e., to $\mu \mathbf{1}$ with $\mathbf{1} := (1, 1, \dots, 1)^T$ and

$$\mu := \frac{1}{N} \sum_{i=0}^{N-1} f_i.$$

Lemma 4.1 (Conservation of the average value) *The average value is preserved by the diffusion process (3.3), that is*

$$\frac{1}{N} \sum_{i=0}^{N-1} u_i^{(k)} = \mu \quad \forall k \in \mathbb{N}_0.$$

Proof. Since $\mathbf{1}^T(\mathbf{I} - \mathbf{V}^T) = \mathbf{0}^T$ we see by (3.3) that $\mathbf{1}^T \mathbf{u}^{(k+1)} = \mathbf{1}^T \mathbf{u}^{(k)}$. *q.e.d.*

4.1 Linear diffusion

Let us consider the convergence for the linear diffusivity $g(x) \equiv 1$ first. In this case, we have $\mathbf{D}_j^{(k)} = \mathbf{I}$, $j = 1, \dots, n$ and (3.3) is of the form

$$\mathbf{u}^{(k+1)} = \mathbf{A}_n \mathbf{u}^{(k)}$$

with

$$\mathbf{A}_n := \mathbf{I} - \tau(\mathbf{I} - \mathbf{V}^T) \left[\alpha_1 \mathbf{I} + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \left[\prod_{l=0}^{j-2} (\mathbf{I} + (\mathbf{V}^{2^l})^T)^2 (\mathbf{I} + \mathbf{V}^{2^l})^2 \right] \right] (\mathbf{I} - \mathbf{V}).$$

Observe that \mathbf{A}_n is a symmetric circulant matrix, i.e., it is diagonalizable by the N -th Fourier matrix, see [6], and all its eigenvalues are real. Since $(\mathbf{I} - \mathbf{V})\mathbf{1} = \mathbf{0}$, we have that $\mathbf{A}_n \mathbf{1} = \mathbf{1}$. Thus, 1 is an eigenvalue of \mathbf{A}_n with corresponding eigenvector $\mathbf{1}$. In order to show that $\mathbf{u}^{(k)} = \mathbf{A}_n^k \mathbf{u}^{(0)}$ with $\mathbf{u}^{(0)} = \mathbf{f}$ converges to $\mu \mathbf{1}$ for $k \rightarrow \infty$ it remains to ensure that all further eigenvalues of \mathbf{A}_n lie inside $(-1, 1)$.

Theorem 4.2 *The iteration scheme (3.3) with linear diffusivity $g(|x|) \equiv 1$ converges for every $\mathbf{u}^{(0)} = \mathbf{f} \in \mathbb{R}^N$ to the spatial average,*

$$\lim_{k \rightarrow \infty} \mathbf{A}_n^k \mathbf{u}^{(0)} = \mu \mathbf{1}, \quad (4.1)$$

if and only if the parameters $\alpha_j \in [0, 1]$ with $\sum_{j=1}^n \alpha_j = 1$ and the time step $\tau > 0$ fulfill

$$\tau \sum_{j=1}^n \frac{\alpha_j}{16^{j-1}} \frac{\left(\sin \frac{\pi 2^{j-1} m}{N} \right)^4}{\left(\sin \frac{\pi m}{N} \right)^2} < \frac{1}{2}, \quad m = 1, \dots, N-1. \quad (4.2)$$

Proof. The eigenvalues of \mathbf{A}_n are given by $p_n(e^{\frac{2\pi i m}{N}})$, $m = 0, \dots, N-1$, where

$$\begin{aligned} p_n(z) &:= 1 - \tau(1 - z^{-1}) \left[\alpha_1 + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \prod_{l=0}^{j-2} (1 + z^{-2^l})^2 (1 + z^{2^l})^2 \right] (1 - z) \\ &= 1 - \tau(2 - z^{-1} - z) \left[\alpha_1 + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \prod_{l=0}^{j-2} (2 + z^{-2^l} + z^{2^l})^2 \right]. \end{aligned}$$

Hence, the assertion (4.1) is satisfied if and only if $|p_n(e^{\frac{2\pi im}{N}})| < 1$ for all $m = 1, \dots, N-1$. Now $p_n(e^{\frac{2\pi im}{N}})$ can be rewritten as

$$\begin{aligned} p_n(e^{\frac{2\pi im}{N}}) &= 1 - \tau \left(2 - 2 \cos \frac{2\pi m}{N} \right) \left[\alpha_1 + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \prod_{l=0}^{j-2} \left(2 + 2 \cos \frac{2\pi 2^l m}{N} \right)^2 \right] \\ &= 1 - 4\tau \left(\sin \frac{\pi m}{N} \right)^2 \left[\alpha_1 + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \prod_{l=0}^{j-2} \left(2 \cos \frac{\pi 2^l m}{N} \right)^4 \right] \end{aligned} \quad (4.3)$$

and further, using that $2 \cos \alpha = \sin 2\alpha / \sin \alpha$, as

$$p_n(e^{\frac{2\pi im}{N}}) = 1 - \frac{4\tau}{\left(\sin \frac{\pi m}{N} \right)^2} \sum_{j=1}^n \frac{\alpha_j}{16^{j-1}} \left(\sin \frac{\pi 2^{j-1} m}{N} \right)^4.$$

Obviously, we have that $p_n(e^{\frac{2\pi im}{N}}) < 1$. The restriction $p_n(e^{\frac{2\pi im}{N}}) > -1$ yields the assertion. *q.e.d.*

4.2 Nonlinear Diffusion

Now we consider bounded nonlinear diffusivity functions $g(|x|)$ satisfying (2.2). Recall that the explicit discretization scheme (3.3) has the form

$$\mathbf{u}^{(k+1)} = \mathbf{A}_n^{(k)} \mathbf{u}^{(k)},$$

where

$$\mathbf{A}_n^{(k)} := \mathbf{I} - \tau(\mathbf{I} - \mathbf{V}^T) \left[\alpha_1 \mathbf{D}_1^{(k)} + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \left(\prod_{l=0}^{j-2} (\mathbf{I} + (\mathbf{V}^{2^l})^T)^2 \right) \mathbf{D}_j^{(k)} \left(\prod_{l=0}^{j-2} (\mathbf{I} + \mathbf{V}^{2^l})^2 \right) \right] (\mathbf{I} - \mathbf{V})$$

strongly depends on the input vector $\mathbf{u}^{(k)}$.

We will apply the following Theorem of Weyl (see e.g. [9], page 181).

Theorem 4.3 (Weyl) *Let $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{N \times N}$ be Hermitian matrices and let the eigenvalues of \mathbf{A}, \mathbf{B} and $\mathbf{A} + \mathbf{B}$ be denoted in increasing order by $\lambda_i(\mathbf{A}), \lambda_i(\mathbf{B}), \lambda_i(\mathbf{A} + \mathbf{B}), i = 1, \dots, N$. Then we have for $i = 1, \dots, N$ that*

$$\lambda_i(\mathbf{A}) + \lambda_1(\mathbf{B}) \leq \lambda_i(\mathbf{A} + \mathbf{B}) \leq \lambda_i(\mathbf{A}) + \lambda_N(\mathbf{B}).$$

Now we can prove the convergence of our scheme for nonlinear diffusivities. By (2.2), the diagonal elements of the matrices $\mathbf{D}_j^{(k)}$ are positive and bounded from above by 1. Therefore it appears that the parameters have to satisfy nearly the same conditions as in the linear setting.

Theorem 4.4 *The iteration scheme (3.3) with diffusivity function g satisfying (2.2) converges for $\mathbf{u}^{(0)} = \mathbf{f} \in \mathbb{R}^N$ to the spatial average,*

$$\lim_{k \rightarrow \infty} \mathbf{u}^{(k)} = \mu \mathbf{1},$$

if the diagonal elements of $\mathbf{D}_j^{(k)}$ are uniformly bounded from below by $c > 0$ for all $k \in \mathbb{N}_0$ and the parameters $\alpha_j \in [0, 1]$ with $\sum_{j=1}^n \alpha_j = 1$ and the time step $\tau > 0$ fulfill condition (4.2).

Proof. 1. Since $(\mathbf{I}-\mathbf{V})\mathbf{1} = \mathbf{0}$, the vector $\mathbf{1}$ is an eigenvector of $\mathbf{A}_n^{(k)}$ with eigenvalue $\lambda_N(\mathbf{A}_n^{(k)}) = 1$. Let $\lambda_1(\mathbf{A}_n^{(k)}) \leq \dots \leq \lambda_{N-1}(\mathbf{A}_n^{(k)})$ denote the other eigenvalues of $\mathbf{A}_n^{(k)}$. We show that there exists some $\varepsilon > 0$ such that $\lambda_j(\mathbf{A}_n^{(k)}) \in [-1 + \varepsilon, 1 - \varepsilon]$ for all $j = 1, \dots, N-1$. By (2.2) and assumption on $\mathbf{D}_j^{(k)}$, it follows that

$$d_{\min}^{(k)} := \min \{1/\|\mathbf{D}_j^{(k)}\|_\infty : j = 1, \dots, n\} \geq c > 0, \quad (4.4)$$

$$d_{\max}^{(k)} := \max \{\|\mathbf{D}_j^{(k)}\|_\infty : j = 1, \dots, n\} \leq 1, \quad (4.5)$$

and $d_{\min}^{(k)} \leq d_{\max}^{(k)}$. We consider the two matrices

$$\begin{aligned} \mathbf{A}_{\min}^{(k)} &:= \mathbf{I} - \tau d_{\min}^{(k)} (\mathbf{I} - \mathbf{V}^T) \left[\alpha_1 \mathbf{I} + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \prod_{l=0}^{j-2} (\mathbf{I} + (\mathbf{V}^{2^l})^T)^2 (\mathbf{I} + \mathbf{V}^{2^l})^2 \right] (\mathbf{I} - \mathbf{V}) \\ \mathbf{A}_{\max}^{(k)} &:= \mathbf{I} - \tau d_{\max}^{(k)} (\mathbf{I} - \mathbf{V}^T) \left[\alpha_1 \mathbf{I} + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \prod_{l=0}^{j-2} (\mathbf{I} + (\mathbf{V}^{2^l})^T)^2 (\mathbf{I} + \mathbf{V}^{2^l})^2 \right] (\mathbf{I} - \mathbf{V}). \end{aligned}$$

Obviously, these are the iteration matrices for the n level linear diffusion schemes with time steps $\tau d_{\min}^{(k)}$ and $\tau d_{\max}^{(k)}$, respectively. Thus, we obtain by (4.2) and (4.4), (4.5) that for some $\varepsilon > 0$ (being independent of k) the relations $\lambda_i(\mathbf{A}_{\max}^{(k)}) \geq -1 + \varepsilon$ and $\lambda_i(\mathbf{A}_{\min}^{(k)}) \leq 1 - \varepsilon$ hold true for all $i = 1, \dots, N-1$. Further, the matrix

$$\begin{aligned} \mathbf{A}_n^{(k)} - \mathbf{A}_{\max}^{(k)} &= \tau (\mathbf{I} - \mathbf{V}^T) \left[\alpha_1 (d_{\max}^{(k)} \mathbf{I} - \mathbf{D}_1^{(k)}) \right. \\ &\quad \left. + \sum_{j=2}^n \frac{\alpha_j}{16^{j-1}} \left(\prod_{l=0}^{j-2} (\mathbf{I} + (\mathbf{V}^{2^l})^T)^2 \right) (d_{\max}^{(k)} \mathbf{I} - \mathbf{D}_j^{(k)}) \left(\prod_{l=0}^{j-2} (\mathbf{I} + \mathbf{V}^{2^l})^2 \right) \right] (\mathbf{I} - \mathbf{V}) \end{aligned}$$

is symmetric and positive semidefinite. Analogously, $\mathbf{A}_n^{(k)} - \mathbf{A}_{\min}^{(k)}$ is symmetric and negative semidefinite. Hence, using the Theorem of Weyl, we obtain for all $i = 1, \dots, N-1$ that

$$-1 + \varepsilon \leq \lambda_1(\mathbf{A}_n^{(k)} - \mathbf{A}_{\max}^{(k)}) + \lambda_i(\mathbf{A}_{\max}^{(k)}) \leq \lambda_i(\mathbf{A}_n^{(k)}).$$

Similarly, we have for $i = 1, \dots, N-1$ that

$$\lambda_i(\mathbf{A}_n^{(k)}) \leq \lambda_N(\mathbf{A}_n^{(k)} - \mathbf{A}_{\min}^{(k)}) + \lambda_i(\mathbf{A}_{\min}^{(k)}) \leq 1 - \varepsilon.$$

2. Now we prove that our scheme converges to $\mu \mathbf{1}$. Since $\mathbf{1}^T \mathbf{u}^{(k)} = N\mu$, the unique decomposition of $\mathbf{u}^{(k)}$ with respect to the orthogonal sum $\text{Ker}(\mathbf{I} - \mathbf{V}) \oplus \text{Im}(\mathbf{I} - \mathbf{V}) = \mathbb{R}^N$ is given by

$$\mathbf{u}^{(k)} = \mu \mathbf{1} + \mathbf{r}^{(k)},$$

where $\mathbf{1}^T \mathbf{r}^{(k)} = 0$. Since $\text{Im}(\mathbf{I} - \mathbf{V}) = \text{span}\{\mathbf{b}_1, \dots, \mathbf{b}_{N-1}\}$, with orthonormal eigenvectors \mathbf{b}_j of $\mathbf{A}_n^{(k)}$ corresponding to the eigenvalues $\lambda_j(\mathbf{A}_n^{(k)})$, we can represent $\mathbf{r}^{(k)}$ as

$$\mathbf{r}^{(k)} = c_1 \mathbf{b}_1^{(k)} + \dots + c_{N-1} \mathbf{b}_{N-1}^{(k)}.$$

Then it follows by part 1 of the proof and by

$$\mathbf{u}^{(k+1)} = \mu \mathbf{1} + \mathbf{r}^{(k+1)} = \mathbf{A}_n^{(k)} (\mu \mathbf{1} + \mathbf{u}^{(k)}) = \mu \mathbf{1} + \mathbf{A}_n^{(k)} \mathbf{r}^{(k)}$$

that

$$\|\mathbf{r}^{(k+1)}\|_2^2 = \|\mathbf{A}_n^{(k)} \mathbf{r}^{(k)}\|_2^2 = \sum_{j=1}^{N-1} |\lambda_j(\mathbf{A}_n^{(k)})|^2 c_j^2 \leq (1 - \varepsilon)^2 \sum_{j=1}^{N-1} c_j^2 = (1 - \varepsilon)^2 \|\mathbf{r}^{(k)}\|_2^2.$$

Thus, $\lim_{k \rightarrow \infty} \mathbf{r}^{(k)} = \mathbf{0}$ which implies the assertion. *q.e.d.*

Remark. Going through the proof of Theorem 4.4 without the assumption that the elements of $\mathbf{D}_j^{(k)}$ are uniformly bounded from below by a positive number for all $k \in \mathbb{N}_0$, we can ensure nevertheless that $-1 + \varepsilon \leq \lambda_i(\mathbf{A}_n^{(k)}) \leq 1$ for all $k \in \mathbb{N}_0$. Hence $\|\mathbf{r}^{(k+1)}\|_2 \leq \|\mathbf{r}^{(k)}\|_2$ as before, i.e.,

$$\|\mathbf{u}^{(k)}\|_2^2 = \|\mu \mathbf{1} + \mathbf{r}^{(k)}\|_2^2 = \|\mu \mathbf{1}\|_2^2 + \|\mathbf{r}^{(k)}\|_2^2 \leq \|\mu \mathbf{1}\|_2^2 + \|\mathbf{r}^{(0)}\|_2^2 = \|\mathbf{f}\|_2^2,$$

and in particular, $|u_i^{(k)}| \leq \|\mathbf{f}\|_2$ for $i = 0, \dots, N-1, k \in \mathbb{N}_0$. Considering the diagonal diffusivity matrices in (3.4) it follows that the above assumption on boundedness of $\|(\mathbf{D}_j^{(k)})^{-1}\|$ for all $k \in \mathbb{N}_0$ is satisfied for Perona-Malik diffusivity and Charbonnier diffusivity.

For the special cases $n = 1, 2, 3$, we obtain the following corollary.

Corollary 4.5 *The iteration scheme (3.3) with diffusivity function g satisfying (2.2) converges for $\mathbf{u}^{(0)} = \mathbf{f} \in \mathbb{R}^N$ to the spatial average, if the diagonal elements of $\mathbf{D}_j^{(k)}$, $j = 1, 2, 3$ are uniformly bounded from below by $c > 0$ for all $k \in \mathbb{N}_0$ and*

i) case $n = 1$: $0 < \tau < \frac{1}{2}$

ii) case $n = 2$:

if $\alpha_1 \geq \frac{1}{4}$, then $\tau \alpha_1 < \frac{1}{2}$,

if $\alpha_1 < \frac{1}{4}$, then $\tau \left(1 + 8\alpha_1 + (1 - 4\alpha_1) \sqrt{(1 - 4\alpha_1)/(1 - \alpha_1)} \right) < \frac{27}{4}$.

iii) case $n = 3$:

$$\tau \left(\alpha_1 u_m + \alpha_2 u_m (1 - u_m)^2 + \alpha_3 u_m (1 - u_m)^2 (1 - 2u_m)^4 \right) < \frac{1}{2},$$

where $u_m := \left(\sin \frac{\pi m}{N} \right)^2$, $m = 1, \dots, N-1$.

Proof. For $n \leq 3$, condition 4.2 reads

$$\tau \left(\alpha_1 \left(\sin \frac{\pi m}{N} \right)^2 + \frac{\alpha_2 \left(\sin \frac{2\pi m}{N} \right)^4}{16 \left(\sin \frac{\pi m}{N} \right)^2} + \frac{\alpha_3 \left(\sin \frac{4\pi m}{N} \right)^4}{16^2 \left(\sin \frac{\pi m}{N} \right)^2} \right) < \frac{1}{2}, \quad m = 1, \dots, N-1.$$

By $\sin 2\alpha = 2 \sin \alpha \cos \alpha$ this can be rewritten as

$$\tau \left(\alpha_1 u_m + \alpha_2 u_m (1 - u_m)^2 + \alpha_3 u_m (1 - u_m)^2 (1 - 2u_m)^4 \right) < \frac{1}{2}.$$

For $n = 3$, this yields the assertion iii). For $n = 1$, i.e., $\alpha_1 = 1$ and $\alpha_2 = \alpha_3 = 0$, assertion i) follows immediately since $u_m \in (0, 1]$. For $n = 2$, i.e., $\alpha_1 + \alpha_2 = 1$ and $\alpha_3 = 0$, the function $\varphi(u_m) := \alpha_1 u_m + \alpha_2 u_m (1 - u_m)^2$ is monotonically increasing in $(0, 1]$ if $\alpha_1 \geq \frac{1}{4}$. Consequently, in this case, the condition $\tau \varphi(1) = \tau \alpha_1 < \frac{1}{2}$ must be fulfilled. For $\alpha_1 < \frac{1}{4}$, the maximum of the function φ in $(0, 1]$ is

$$\frac{2}{27} \left(1 + 8\alpha_1 + (1 - 4\alpha_1) \sqrt{(1 - 4\alpha_1)/(1 - \alpha_1)} \right).$$

This completes the proof. *q.e.d.*

For $n = 1$, the above convergence result for the explicit scheme is well-known, see e.g. [15].

For $n = 2$ and given parameters α_1, α_2 we obtain from Corollary 4.5 the bounds for the time step τ given in Table 1 (left). To get large time steps τ we have to choose α_1 small, e.g., for $\alpha_1 = 0.2$, we can use five times larger time steps than in the original diffusion scheme.

$n = 2$			$n = 3$			
α_1	α_2	τ	α_1	α_2	α_3	τ
1.00	0.00	< 0.50	0.20	0.4	0.40	< 2.5000
0.50	0.50	< 1.00	0.10	0.4	0.50	< 5.0000
0.40	0.60	< 1.25	0.10	0.3	0.60	< 5.0000
0.25	0.75	< 2.00	0.07	0.3	0.63	< 6.8927
0.20	0.80	< 2.50	0.05	0.2	0.75	< 10.000

Table 1: Bounds for the time step τ for $n = 2$ and given parameters α_1, α_2 (left) and for $n = 3$ and given parameters $\alpha_1, \alpha_2, \alpha_3$ (right).

For $n = 3$ and given parameters $\alpha_1, \alpha_2, \alpha_3$, simple computations result in the bounds for τ in Table 1 (right). For example, for $\alpha_1 = 0.05$ and $\alpha_2 = 0.2$, we can use 20 times larger time steps than in the original diffusion scheme.

Compared to the model for discrete diffusion filtering in [16], we can not ensure for $n \geq 2$ that all entries of $\mathbf{A}_n^{(k)}$ are non-negative and all diagonal elements of $\mathbf{A}_n^{(k)}$ are positive. Therefore, a rigorous extremum principle of the form $a \leq u_i^{(k)} \leq b$ for all $i = 0, \dots, N-1$ and all $k \in \mathbb{N}_0$ with $a := \min \{f_i : i = 0, \dots, N-1\}$, $b := \max \{f_i : i = 0, \dots, N-1\}$ holds only true for one scale $n = 1$.

However, the following estimates for $u_i^{(k)}$, $i = 0, \dots, N-1$ follow from Theorem 4.4.

Corollary 4.6 *Let $\mathbf{u}^{(0)} = \mathbf{f} \in \mathbb{R}^N$ and $(\mathbf{u}^{(k)})_{k \in \mathbb{N}_0}$ be the sequence of filtered vectors with $\mathbf{u}^{(k+1)} = \mathbf{A}_n^{(k)} \mathbf{u}^{(k)}$. Let the assumptions of Theorem 4.4 be satisfied. Then we have for all $k \in \mathbb{N}_0$ that*

$$\|\mathbf{u}^{(k+1)}\|_2 < \|\mathbf{u}^{(k)}\|_2, \quad (\mathbf{u}^{(k)} \neq \mu \mathbf{1}), \quad \text{and} \quad \|\mathbf{u}^{(k)}\|_2 \geq \mu \sqrt{N}.$$

Proof. By the proof of Theorem 4.4, we see for $\mathbf{u}^{(k)} \neq \mu \mathbf{1}$ that

$$\|\mathbf{u}^{(k)}\|_2^2 = \mu^2 N + \|\mathbf{r}^{(k)}\|_2^2 < \mu^2 N + \|\mathbf{r}^{(k+1)}\|_2^2 = \|\mathbf{u}^{(k+1)}\|_2^2.$$

Now the assertion follows by $\|\mathbf{u}^{(k)}\|_2 \geq \|\mu \mathbf{1}\|_2 = \mu \sqrt{N}$. *q.e.d.*

Furthermore, we can show that the central moments of $\mathbf{u}^{(k)}$ are also decreasing.

Corollary 4.7 *Let $\mathbf{u}^{(0)} = \mathbf{f} \in \mathbb{R}^N$ and $(\mathbf{u}^{(k)})_{k \in \mathbb{N}_0}$ be the sequence of filtered vectors with $\mathbf{u}^{(k+1)} = \mathbf{A}_n^{(k)} \mathbf{u}^{(k)}$. Let the assumptions of Theorem 4.4 be satisfied. Then the central moments*

$$\mathcal{M}_{2m}(\mathbf{u}^{(k)}) := \frac{1}{N} \sum_{i=0}^{N-1} (u_i^{(k)} - \mu)^{2m} = \frac{1}{N} \|\mathbf{u}^{(k)} - \mu \mathbf{1}\|_2^{2m}$$

are decreasing with k , i.e.,

$$\mathcal{M}_{2m}(\mathbf{u}^{(k+1)}) \leq \mathcal{M}_{2m}(\mathbf{u}^{(k)}) \quad \text{for all } k \in \mathbb{N}_0, m \in \mathbb{N}.$$

Proof. By Theorem 4.4 we have $\|\mathbf{A}_n^{(k)}\|_2 = 1$ for all $k \in \mathbb{N}_0$ and consequently

$$\mathcal{M}_{2m}(\mathbf{u}^{(k+1)}) = \frac{1}{N} \|\mathbf{u}^{(k+1)} - \mu \mathbf{1}\|_2^{2m} = \frac{1}{N} \|\mathbf{A}_n^{(k)} \mathbf{u}^{(k)} - \mu \mathbf{1}\|_2^{2m} \leq \frac{1}{N} \|\mathbf{u}^{(k)} - \mu \mathbf{1}\|_2^{2m} = \mathcal{M}_{2m}(\mathbf{u}^{(k)}). \quad \text{q.e.d.}$$

5 Numerical Examples

In this section we want to show how our algorithms perform in contrast to the usual diffusion scheme. We start with the linear case where our algorithm with appropriately chosen parameters leads to nearly the same results as the ordinary diffusion scheme but requires less arithmetic operations.

	$k = 1$		$k = 2$		$k = 3$	
m	α_1	J_α	α_1	J_α	α_1	J_α
2	0.55	0.1000	0.59	0.0385	0.63	0.0181
3	0.36	0.1036	0.39	0.0289	0.41	0.0126
4	0.26	0.0779	0.27	0.0064	0.27	0.0012
5	0.21	0.0678	0.19	0.0226	0.17	0.0124
6	0.15	0.1447	0.14	0.0430	0.13	0.0257
7	0.12	0.2544	0.12	0.0607	0.10	0.0356
8	0.09	0.3607	0.09	0.1126	0.09	0.0442

Table 2: Optimal parameters minimizing $J_\alpha = \|\mathbf{A}_1^{mk}(\tau) - \mathbf{A}_2^k(m\tau)\|_2$ for $\tau = 0.25$.

5.1 Linear Diffusion

We want to compare the ordinary linear diffusion with mk iterations of step size τ

$$\mathbf{u}_1^{(mk)} = \mathbf{A}_1^{mk}(\tau)\mathbf{u}^{(0)}$$

with our new n -scale scheme with only k iterations of larger step size $m\tau$

$$\mathbf{u}_n^{(k)} = \mathbf{A}_n^k(m\tau)\mathbf{u}^{(0)}.$$

Here we use $\mathbf{A}_n(\tau)$ instead of \mathbf{A}_n to emphasize the dependence of the iteration matrix on the step size τ . Since $\|\mathbf{u}_1^{(mk)} - \mathbf{u}_n^{(k)}\|_2 \leq \|\mathbf{A}_1^{mk}(\tau) - \mathbf{A}_n^k(m\tau)\|_2 \|\mathbf{u}^{(0)}\|_2$ we consider the spectral matrix norm

$$J_\alpha = J_\alpha(\tau, k, m, n) := \|\mathbf{A}_1^{mk}(\tau) - \mathbf{A}_n^k(m\tau)\|_2.$$

By Theorem 4.2, we know that this norm converges for appropriately chosen α_j and $k \rightarrow \infty$ to zero. Now we want to fit for small k and m the parameters α_j so that J_α becomes small. To this end, we have computed for $n = 2, 3$, $\tau = 0.25$ and fixed m, k the optimal parameters α_j (up to 2 digits) for which J_α becomes minimal. The results are contained in Tables 2 and 3. Here we have taken $N = 64$. For larger N the values J_α differ slightly, while the optimal parameters α_j remain the same. Remember that for linear diffusion all $\mathbf{A}_n^{(k)}(\tau)$ are circulant matrices. By (4.3) with $\alpha_1 + \alpha_2 = 1$ and $x_j = \cos \frac{2\pi j}{N}$ we have for the two-scale case $n = 2$ that

$$\begin{aligned} J_\alpha &= \max_{j=1, \dots, N} \left\{ \left| (1 - 2\tau(1 - x_j))^{km} - \left(1 - 2m\tau(1 - x_j)\left(\alpha_1 + \frac{\alpha_2}{4}(1 + x_j^2)\right)\right)^k \right| \right\} \\ &= \max_{j=1, \dots, N} \left\{ \left| (1 - 2\tau(1 - x_j))^{km} \right. \right. \\ &\quad \left. \left. - \left(1 - \alpha_2 m\tau + \left(\frac{5}{2}\alpha_2 - 2\right)m\tau(1 - x_j) + \frac{\alpha_2}{2}m\tau x_j^2(1 + x_j)\right)^k \right| \right\}. \end{aligned}$$

Minimization of this functional over $x_j \in [-1, 1]$ leads to the same optimal parameters α_1 as given in Table 2.

In the two-scale case $n = 2$ the parameters $m = 4$ and $\alpha_1 = 0.27$ lead to the best fit. For this choice our time step is 4 times larger.

At the left-hand side of Figure 5 we have plotted some columns of $\mathbf{A}_1^{mk}(\tau)$ (dashed) and of $\mathbf{A}_2^k(m\tau)$ (dotted) with $m = 4$, $\tau = 0.25$ and $\alpha_1 = 0.27$ in order to compare these diffusion kernels; see column 8 for $k = 2$, column 16 for $k = 1$ and column 24 for $k = 3$ (left to right). We have chosen the small matrix size $N = 32$ to visualize the differences.

m	$k = 1$			$k = 2$		
	α_1	α_2	J_α	α_1	α_2	J_α
5	0.21	0.76	0.0500	0.19	0.75	0.0017
6	0.17	0.68	0.0510	0.16	0.71	0.0027
7	0.15	0.58	0.0500	0.13	0.68	0.0029
8	0.13	0.52	0.0478	0.12	0.63	0.0030
9	0.11	0.49	0.0578	0.10	0.59	0.0029
10	0.10	0.44	0.0563	0.09	0.54	0.0031
15	0.07	0.28	0.0541	0.07	0.31	0.0012
20	0.05	0.21	0.0528	0.06	0.13	0.0081

Table 3: Optimal parameters minimizing $J_\alpha = \|\mathbf{A}_1^{mk}(\tau) - \mathbf{A}_3^k(m\tau)\|_2$ for $\tau = 0.25$.

At the right-hand side of Figure 5 we have plotted some columns of $\mathbf{A}_1^{mk}(\tau)$ (dashed) and of $\mathbf{A}_3^k(m\tau)$ (dotted) with $\tau = 0.25$, matrix size $N = 64$ and the following parameters; column 16 for $(k, m, \alpha_1, \alpha_2) = (2, 10, 0.09, 0.54)$, column 32 for $(k, m, \alpha_1, \alpha_2) = (1, 10, 0.1, 0.44)$, and column 48 for $(k, m, \alpha_1, \alpha_2) = (1, 20, 0.05, 0.21)$ (left to right).

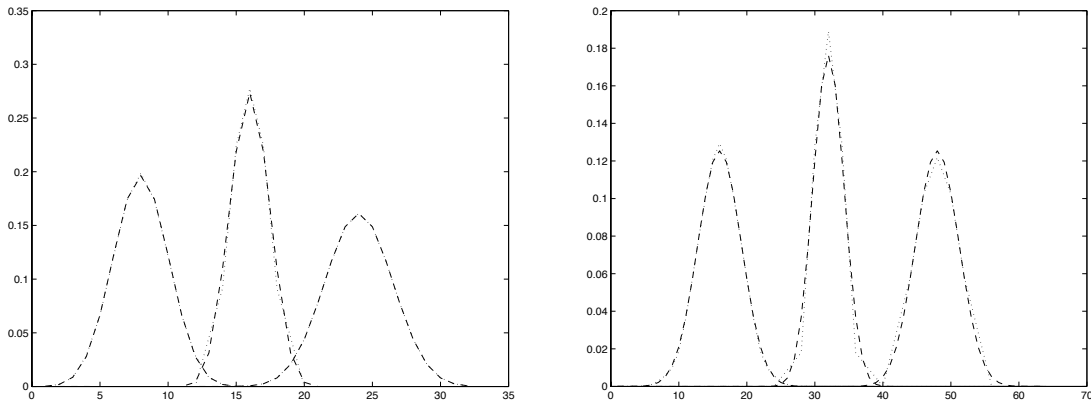


Figure 5: **Left:** Columns 8,16,24 (left to right) of $\mathbf{A}_1^{4k}(\tau)$ (dashed) and of $\mathbf{A}_2^k(4\tau)$ (dotted) for $k = 2, 1, 3$ (left to right) with $\alpha_1 = 0.27$. **Right:** Columns 16,32,48 (left to right) of $\mathbf{A}_1^{mk}(\tau)$ (dashed) and of $\mathbf{A}_3^k(m\tau)$ (dotted) for $k = 2, 1, 1$ and $m = 10, 10, 20$ (left to right) and α_1, α_2 from Table 3. We have always used $\tau = 0.25$.

5.2 Nonlinear Diffusion

In the nonlinear case, the 'best' choice of the parameters α_j depends on the input signal and the chosen diffusivity function. In the following, we present some denoising results to sketch the behavior of our multiscale approach. As input signals we choose a ramp signal of length $N = 512$ with central jump and the signal 'blocks' of length $N = 1024$, one of the standard signals in wavelet denoising which mimics a scan line through a 2-D image depicting an object with several edges [7]. We added zero-mean Gaussian noise with SNR of approximately 10 and 8, respectively. Here the *signal-to-noise ratio* (SNR) is defined by

$$\text{SNR} = 20 \log_{10} \frac{\|\mathbf{z} - \bar{\mathbf{z}}\|_2}{\|\mathbf{n}\|_2},$$

with \mathbf{z} standing for the ideal signal with mean $\bar{\mathbf{z}}$, and \mathbf{n} representing noise. The input signals are depicted in Figure 6.

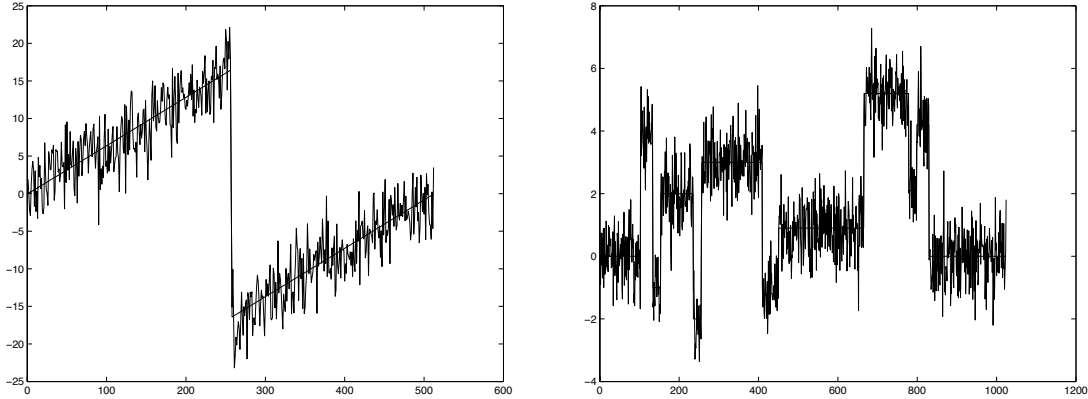


Figure 6: **Left:** noisy ramp (SNR = 10.2126). **Right:** noisy blocks (SNR = 7.9277).

We applied the diffusion schemes with Perona-Malik diffusivity and Charbonnier diffusivity until we reached the maximal SNR. As filter parameters we applied $\lambda = 0.25$ for the ramp and $\lambda = 0.05$ for the blocks to achieve a good SNR. The ordinary diffusion algorithm (2.3) was applied with time step size $\tau = 0.25$. As in the linear case this ensures the 'sign stability' of the algorithm, see [12]. We applied the two-scale algorithm (3.2) with five times larger step size $\tau = 1.25$ and $\alpha_1 = 0.27$ and the three-scale algorithm ((3.3) with $n = 3$) with 20 times larger step size $\tau = 5$ and $\alpha_1 = 0.02, \alpha_2 = 0.2$. The denoising results are contained in Table 4. The corresponding signals are shown in the Figures 7 and 8. We observe the following: In the two-scale case, the results are rather similar to the original diffusion results. A slight superiority of the two-scale case is considered for the ramp and Perona-Malik diffusivity, an observation which cannot be generalized because of the strong dependence of the results from the input signal. However, we have used a five times larger time step which always leads to fewer iterations. In the three-scale case we have applied a 20 times larger time step. Consequently, the number of iterations further decreases. However, this method tends to incorporate specially localized artifacts.

Algorithm	Ramp		Blocks	
	iterations	SNR	iterations	SNR
Perona-Malik-1	2102	22.2072	5588	28.2526
Perona-Malik-2	526	27.7631	553	25.4474
Perona-Malik-3	121	27.9471	136	21.4541
Charbonnier-1	412	23.5583	216	20.0121
Charbonnier-2	85	23.5451	45	21.8166
Charbonnier-3	54	25.8718	32	19.8807

Table 4: Number of iterations and the achieved SNR. Perona-Malik/Charbonnier-1: ordinary diffusion algorithm with $\tau = 0.25$; Perona-Malik/Charbonnier-2: two-scale algorithm with $\tau = 1.25$ and $\alpha_1 = 0.27$; Perona-Malik/Charbonnier-3: three-scale algorithm with $\tau = 5$ and $\alpha_1 = 0.05, \alpha_2 = 0.2$. We have used $\lambda = 0.25$ for the ramp signal and $\lambda = 0.05$ for the blocks.

Finally, let us give some remarks on the efficiency of the new algorithms.

Our filter in \mathbf{A}_2 has length 7, i.e., \mathbf{A}_2 has 7 nonzero entries per row, while the filter in \mathbf{A}_1 has only length 3. If the time step for the two-scale algorithm is chosen five times larger, this leads to 7 versus 15 multiplications per pixel to obtain $\mathbf{u}_2^{(1)}$ and $\mathbf{u}_1^{(5)}$, respectively. However, in the linear case we may also apply \mathbf{A}_1^m with filter length $2m + 1$ (here 11) at once. For larger $m \geq 6$ the fit is still very good and we need fewer arithmetic operations.

For the three-scale algorithm, note that our filter in \mathbf{A}_3 has length 15 while 20-fold application of \mathbf{A}_1 requires 60 multiplications.

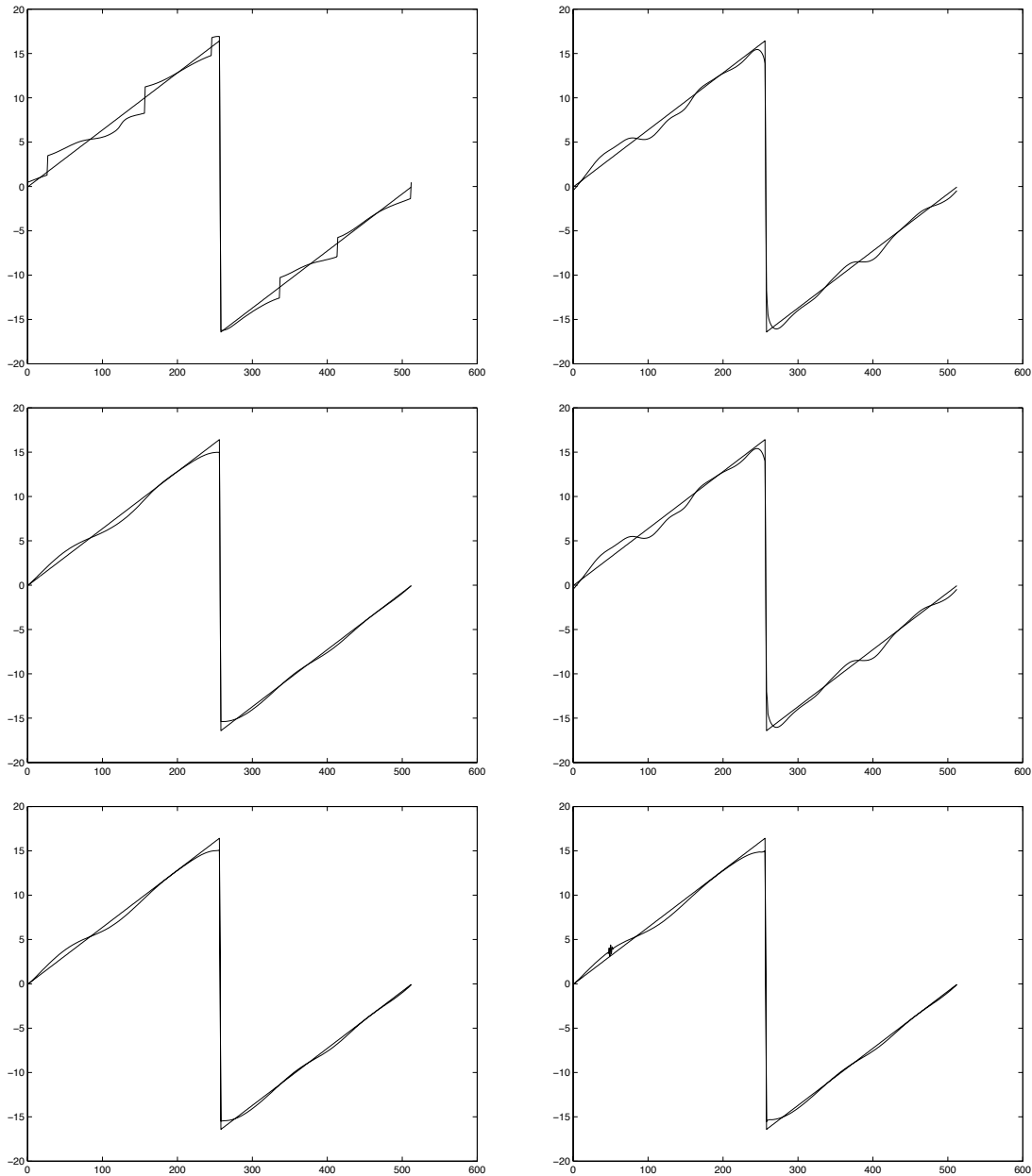


Figure 7: Denoising results for the ramp signal. Perona-Malik/Charbonnier-1: ordinary diffusion algorithm with $\tau = 0.25$; Perona-Malik/Charbonnier-2: two-scale algorithm with $\tau = 1.25$ and $\alpha_1 = 0.27$; Perona-Malik/Charbonnier-3: three-scale algorithm with $\tau = 5$ and $\alpha_1 = 0.05, \alpha_2 = 0.2$. We have used $\lambda = 0.25$.

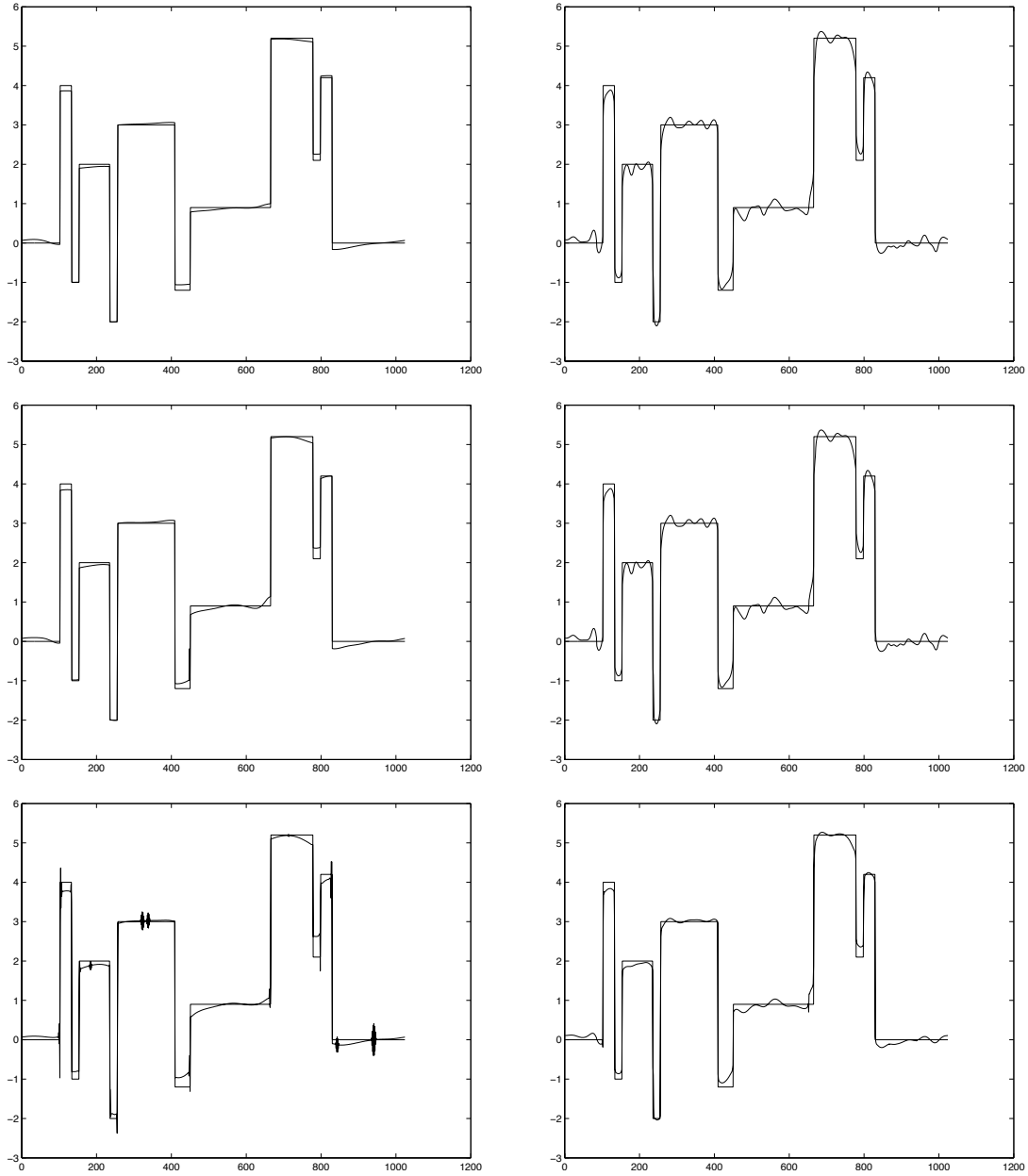


Figure 8: Denoising results for the blocks signal. Perona-Malik/Charbonnier-1: ordinary diffusion algorithm with $\tau = 0.25$; Perona-Malik/Charbonnier-2: two-scale algorithm with $\tau = 1.25$ and $\alpha_1 = 0.27$; Perona-Malik/Charbonnier-3: three-scale algorithm with $\tau = 5$ and $\alpha_1 = 0.05, \alpha_2 = 0.2$. We have used $\lambda = 0.05$.

References

- [1] 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 Transactions on Image Processing*, **7**(3) (1998), 319–335.
- [2] A. Chambolle and B.L. Lucier, Interpreting translation-invariant wavelet shrinkage as a new image smoothing scale space, *IEEE Transactions on Image Processing*, **10**(2) (2001), 993–1000.
- [3] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, Deterministic edge-preserving regularization in computed imaging, *IEEE Transactions on Image Processing*, **6**(2) (1997) 298–311.
- [4] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, Image selective smoothing and edge detection by nonlinear diffusion, *SIAM J. Numer. Anal.*, **29** (1992), 182–193.
- [5] R.R. Coifman and D. Donoho, Translation-invariant denoising, In: A. Antoine and G. Oppenheim, editors, *Wavelets in Statistics*, Springer, New York, 1995, 125–150.
- [6] P. Davis, *Circulant Matrices*, J. Wiley and Sons, New York 1979.
- [7] D. Donoho, De-noising by soft thresholding, *IEEE Transactions on Information Theory*, **41** (1995), 613–627.
- [8] D. Donoho and I.M. Johnstone, Minimax estimation via wavelet shrinkage, *Annals of Statistics*, **26** (1998), 879–921.
- [9] Horn and Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [10] S.L. Keeling and R. Stollberger, Nonlinear anisotropic diffusion filters for wide range edge sharpening, *Inverse Problems* **18** (2002), 175–190.
- [11] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 2nd edition, 1999.
- [12] P. Mrázek, J. Weickert, and G. Steidl, Diffusion-inspired shrinkage functions and stability results for wavelet denoising, *International Journal of Computer Vision*, 2005, to appear.
- [13] P. Mrázek and J. Weickert: Rotationally invariant wavelet shrinkage. In B. Michaelis, G. Krell (Eds.): *Pattern Recognition*. Lecture Notes in Computer Science, Vol. 2781, Springer, Berlin, 2003, 156–163.
- [14] P. Perona and J. Malik, Scale space and edge detection using anisotropic diffusion, Proc. IEEE Computer Society Workshop on Computer Vision, IEEE Computer Society Press, 1987, 16–22.
- [15] G. Steidl, J. Weickert, T. Brox, P. Mrázek, M. Welk, On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides, *SIAM J. Numer. Anal.*, **42/2** (2004), 686–713.
- [16] J. Weickert, *Anisotropic Diffusion in Image Processing*, Teubner, Stuttgart, 1998.
- [17] J. Weickert, B.M. ter Haar Romeny, and M.A. Viergever, Efficient and reliable schemes for nonlinear diffusion filtering, *IEEE Transactions on Image Processing* **7**(3) (1998), 398–410.
- [18] M. Welk, G. Steidl and J. Weickert: A four-pixel scheme for singular differential equations. In R. Kimmel, N. Sochen, J. Weickert (Eds.), *Scale-Space and PDE Methods in Computer Vision*. Lecture Notes in Computer Science, Springer, Berlin, 610–621 (2005).