Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 96

Diffusion-Inspired Shrinkage Functions and Stability Results for Wavelet Denoising

Pavel Mrázek, Joachim Weickert and Gabriele Steidl

Saarbrücken 2003

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 96
submitted: October 27, 2003

Diffusion-Inspired Shrinkage Functions and Stability Results for Wavelet Denoising

Pavel Mrázek
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Building 27
Saarland University, 66041 Saarbrücken, Germany
mrazek@mia.uni-saarland.de

Joachim Weickert
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Building 27
Saarland University, 66041 Saarbrücken, Germany
weickert@mia.uni-saarland.de

Gabriele Steidl
Faculty of Mathematics and Computer Science, D7, 27
University of Mannheim, 68131 Mannheim, Germany
steidl@math.uni-mannheim.de

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

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

Abstract

We study the connections between discrete 1-D schemes for nonlinear diffusion and shift-invariant Haar wavelet shrinkage. We show that one step of a (stabilised) explicit discretisation 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. In this paper we derive new wavelet shrinkage functions from existing diffusivity functions, and identify some previously used shrinkage functions as corresponding to well known diffusivities. We demonstrate experimentally that some of the diffusion-inspired shrinkage functions are among the best for translation-invariant multiscale wavelet denoising. Moreover, by transferring stability notions from diffusion filtering to wavelet shrinkage, we derive conditions on the shrinkage function that ensure that shift invariant single-level Haar wavelet shrinkage is maximum-minimum stable, monotonicity preserving, and variation diminishing.

AMS 2000 Subject Classification: 68T45, 68U10, 35K55, 65T60, 65N04
Key words: image denoising, wavelet shrinkage, diffusion filtering, finite differences, stability

Contents

1 Introduction

We consider a classical task of signal denoising: create an estimate u of an original signal z from its noisy measurement f, where

$$f = z + n,$$

and n denotes an additive noise function. Various methods have been proposed to remove the noise from z without sacrificing important structures such as edges, including rank-order filtering, mathematical morphology, adaptive smoothing, stochastic methods, wavelet techniques, partial differential equations (PDEs) and variational methods. Although these classes of methods serve the same purpose, relatively few publications examine their similarities and differences, in order to transfer results from one of these classes to the others, or to design hybrid methods that combine the advantages of different classes. The present paper is a contribution in this direction, where we concentrate on two of these methods, namely nonlinear diffusion techniques and wavelet shrinkage.

Nonlinear diffusion creates a family of restored signals $u(t)$ by starting from the noisy signal f, and evolving it locally according to a process described by a nonlinear partial differential equation. This process is controlled by a diffusivity function g of the signal gradient. Typically, $g(s)$ is a nonnegative, nonincreasing function of the gradient magnitude, approaching zero as $s \to \infty$. This setting leads to the effect that smoothing of u proceeds faster in homogeneous regions (where the gradient is small, possibly caused by noise), and discontinuities (large gradient, hopefully corresponding to important features of the underlying signal) tend to be preserved. Depending on the choice of the diffusivity function g, a single nonlinear diffusion equation may cover a variety of nonlinear filters, including the original nonlinear diffusion of Perona and Malik [36] and its regularised variants [10, 44], total variation (TV) diffusion [2], balanced forward-backward (BFB) diffusion [28] and a number of others. When applied to discrete data $\mathbf{f} = (f_i)_{i=0}^{N-1}$, the nonlinear diffusion filter creates a series of smoothed signals $\mathbf{u}^k := \mathbf{u}(k\tau)$ iteratively, starting from the noisy signal, $\mathbf{u}^0 = \mathbf{f}$.

Wavelet transforms express the signal in terms of wavelet coefficients, describing the signal variation at different scales. If the wavelet basis is chosen properly, a signal will be generally described by only a few significant wavelet coefficients, while moderate white Gaussian noise pollutes all the wavelet coefficients by a small amount. Signal denoising by wavelet shrinkage [15, 16] starts from this assumption, and creates a smoothed version of the processed signal by the following three-step procedure:

1. *Analysis:* transform the noisy data f to the wavelet coefficients d_i^j, representing the signal at various scales j and positions i.

2. *Shrinkage:* apply a shrinkage function S_θ to the wavelet coefficients d_i^j, thus reducing the relative importance of small coefficients.

3. *Synthesis:* reconstruct a denoised version u of f from the shrunken wavelet coefficients.

The shrinkage parameter θ is chosen with respect to the amount of noise in the input signal. In general, the denoised solution u is obtained from f using a single step of this multiscale procedure, i.e. the method is applied noniteratively. The specific choice of the wavelets and the shrinkage functions allows a large variability of wavelet shrinkage methods.

In the present paper, we show equivalence between a single iteration of a 1-D explicit scheme for nonlinear diffusion on one side, and translation-invariant wavelet shrinkage with a single level of Haar wavelet decomposition on the other. This equivalence is obtained by constructing an appropriate shrinkage function S_θ to an existing diffusivity g, and vice versa. Such a relation does not only allow us to prove some diffusion-inspired stability properties for wavelet shrinkage, it also enable us to generalise a variation-diminishing result known from explicit linear diffusion schemes [22] to the nonlinear setting.
Having asserted the equivalence between wavelet shrinkage and nonlinear diffusion for this special situation, it remains to be seen whether this connection brings any advantages in more general settings. We demonstrate numerically that the shrinkage functions derived from diffusivities are able to provide some of the best results when used for classical (i.e. multi-level, one step) translation-invariant wavelet shrinkage.

This paper is organised as follows. Section 2 sketches nonlinear diffusion filtering and develops its explicit discretisation in 1-D, while Section 3 provides a brief introduction to translation-invariant Haar wavelet shrinkage. The connections between these two types of methods are exploited in Section 4: We establish conditions on diffusivities and shrinkage functions under which the two methods (restricted to one step / one scale) are equivalent. Section 5 describes how this relation can be used for transferring stability results between both paradigms. In Section 6 some novel, diffusion-inspired shrinkage functions are tested experimentally, and compared to previously used ones. The paper is concluded with a summary in Section 7.

Related work. Analysing the relations between regularisation methods and *continuous* wavelet shrinkage of functions, Chambolle *et al.* [7] showed that

one may interpret wavelet shrinkage of functions as regularisation processes in suitable Besov spaces. In the case of Haar wavelets, Cohen *et al.* [11] showed that this approximates total variation regularisation. Later on, Chambolle and Lucier [8] considered iterated translation-invariant wavelet shrinkage and interpreted it as a nonlinear scale-space that differs from other scale-spaces by the fact that it is not given in terms of PDEs.

Regarding the relations between wavelet shrinkage denoising of *discrete* signals and nonlinear diffusion, not much research has been done so far. A recent paper by Coifman and Sowa [14] proposes TV diminishing flows that act along the direction of Haar wavelets. Bao and Krim [3] addressed the problem of texture loss in diffusion scale-spaces by incorporating ideas from wavelet analysis. Recent work in which the authors are involved [5, 40, 41] investigates conditions under which equivalence between wavelet shrinkage of discrete signals, space-discrete TV diffusion or regularisation, and SIDEs (stabilised inverse diffusion equations) holds true.

Some recently proposed hybrid methods are based on combining wavelet shrinkage and TV regularisation methods [1, 37]. Durand and Froment [17] proposed to address the problem of pseudo-Gibbs artifacts in wavelet denoising by replacing the thresholded wavelet coefficients by coefficients that minimise the total variation. Their method is also close in spirit to approaches by Chan and Zhou [9] who postprocessed images obtained from wavelet shrinkage by a TV-like regularisation technique. Coifman and Sowa [13] used functional minimisation with wavelet constraints for postprocessing signals that have been degraded by wavelet thresholding or quantisation. Candes and Guo [6] also presented related work, in which they combined ridgelets and curvelets with TV minimisation strategies. Recently, Malgouyres [30, 31] proposed a hybrid method that uses both wavelet packets and TV approaches. His experiments showed that it may restore textured regions without introducing visible ringing artifacts.

This discussion shows that the previous papers typically focus on TV-based denoising techniques on the PDE side. Moreover, most of them present a continuous analysis rather than a discrete one. Our paper differs from previous work in this field by the fact that we do not restrict ourselves to a single diffusivity or shrinkage function, but introduce and analyse a general connection between a discrete diffusion scheme and Haar wavelet shrinkage. To this end, we investigate a large number of diffusivities and shrinkage functions.

A shorter, preliminary version of the present paper has been published in the proceedings of the Scale-Space 2003 Conference [34]. The current version is extended by a new section on stability issues and it analyses two new diffusivity functions.

2 Nonlinear Diffusion

2.1 Basic Concept

The basic idea behind nonlinear diffusion filtering [36] is to obtain a family $u(x,t)$ of filtered versions of the signal $f(x)$ as the solution of a suitable diffusion process

$$u_t = (g(|u_x|)\, u_x)_x \tag{1}$$

with f as initial condition:

$$u(x,0) = f(x).$$

Here subscripts denote partial derivatives, and the diffusion time t is a simplification parameter: larger values correspond to stronger filtering.
The diffusivity $g(|u_x|)$ is a nonnegative function that controls the amount of diffusion. In most cases, it is a decreasing function in $|u_x|$. This ensures that strong edges are less blurred by the diffusion filter than noise and low-contrast details. Depending on the choice of the diffusivity function, equation (1) covers a variety of filters. Here are some of the previously employed diffusivity functions:

A. Linear diffusivity [25]: $g(|x|) = 1$,

B. Charbonnier diffusivity [10]: $g(|x|) = \dfrac{1}{\sqrt{1 + \frac{|x|^2}{\lambda^2}}}$,

C. Perona–Malik diffusivity [36]: $g(|x|) = \dfrac{1}{1 + \frac{|x|^2}{\lambda^2}}$,

D. Weickert diffusivity [44]: $g(|x|) = \begin{cases} 1 & |x| = 0, \\ 1 - \exp\left(\frac{-3.31488}{(|x|/\lambda)^8}\right) & |x| > 0, \end{cases}$

E. Tukey diffusivity [4]: $g(|x|) = \begin{cases} \left(1 - (x/\lambda)^2\right)^2 & |x| \le \lambda, \\ 0 & |x| > \lambda, \end{cases}$

F. TV diffusivity [2]: $g(|x|) = \dfrac{1}{|x|}$,

G. BFB diffusivity [28]: $g(|x|) = \dfrac{1}{|x|^2}$,

H. FAB diffusivity [21, 39]: $g(|x|) = 2\exp(-|x|^2/\lambda_1^2) - \exp(-|x|^2/\lambda_2^2)$, $\lambda_1 < \lambda_2$.

Note that the diffusivities A–E are bounded from above by 1, while the diffusivities F and G are unbounded. In order to avoid theoretical and numerical

difficulties, it is common to replace the latter ones by regularisations that make them bounded: e.g. one may use $g(|x|) = 1/\sqrt{\epsilon^2 + |x|^2}$ instead of the TV diffusivity. The forward-and-backward (FAB) diffusivity H differs from the other diffusivities by the fact that it may even attain negative values. First diffusivities of such a type have been proposed by Gilboa et al. [20]. Well-posedness results are available for the diffusivities A, B and F, since they result from convex potentials. For the diffusivities C, D and G, which can be related to nonconvex potentials, some well-posedness questions are open in the continuous setting [29, 27], while already a space-discretisation creates well-posed processes [45]. In case of the Tukey diffusivity E, well-posedness results are more difficult to establish, since it may degenerate to 0. The FAB diffusivity H goes one step further by allowing even negative values. However, at extrema the FAB diffusivity is in the forward diffusion region which is responsible for a certain degree of stability.

2.2 Explicit Discretisation Scheme

When applied to discrete signals, the partial differential equation (1) has to be discretised. In this paper, we focus on explicit finite difference schemes. Substituting the spatial partial derivatives in (1) by finite differences (with the assumption of unit distance between neighboring pixels), and employing explicit discretisation in time, an explicit 1-D scheme for nonlinear diffusion can be written in the form

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

where τ is the time step size and the upper index k denotes the approximate solution at time $k\tau$. Separating the unknown u_i^{k+1} on one side, we obtain

$$u_i^{k+1} = u_i^k - \tau\, g(|u_i^k - u_{i+1}^k|)\,(u_i^k - u_{i+1}^k) + \tau\, g(|u_{i-1}^k - u_i^k|)\,(u_{i-1}^k - u_i^k). \quad (2)$$

The initial condition reads $u_i^0 = f_i$ for all i.

3 Wavelet Shrinkage

3.1 Basic Concept

The discrete wavelet transform represents a one-dimensional signal f in terms of shifted versions of a dilated lowpass scaling function φ, and shifted and dilated versions of a bandpass wavelet function ψ. In case of orthonormal

wavelets, this gives

$$f = \sum_{i\in\mathbb{Z}} \langle f, \varphi_i^n \rangle \, \varphi_i^n + \sum_{j=-\infty}^{n} \sum_{i\in\mathbb{Z}} \langle f, \psi_i^j \rangle \, \psi_i^j, \tag{3}$$

where $\psi_i^j(s) := 2^{-j/2}\psi(2^{-j}s - i)$ and where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L_2(\mathbb{R})$. If the measurement f is corrupted by moderate 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 a few significant wavelet coefficients [32]. Therefore, wavelet shrinkage attempts to eliminate noise from the wavelet coefficients by the following three-step procedure:

1. *Analysis*: transform the noisy data f to the wavelet coefficients $d_i^j = \langle f, \psi_i^j \rangle$ and scaling function coefficients $c_i^n = \langle f, \varphi_i^n \rangle$ according to (3).

2. *Shrinkage*: apply a shrinkage function S_θ with a threshold parameter θ to the wavelet coefficients, i.e., $S_\theta(d_i^j) = S_\theta(\langle f, \psi_i^j \rangle)$.

3. *Synthesis*: reconstruct the denoised version u of f from the shrunken wavelet coefficients:

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

In this paper we restrict our attention to *Haar wavelets*, well suited for piecewise constant signals with discontinuities. The Haar wavelet and scaling functions are given respectively by

$$\psi(x) = \mathbf{1}_{[0,\frac{1}{2})} - \mathbf{1}_{[\frac{1}{2},1)}, \tag{4}$$

$$\phi(x) = \mathbf{1}_{[0,1)} \tag{5}$$

where $\mathbf{1}_{[a,b)}$ denotes the characteristic function, equal to 1 on $[a, b)$ and zero everywhere else. 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 conversely:

$$c_i^j = \frac{c_{2i}^{j-1} + c_{2i+1}^{j-1}}{\sqrt{2}}, \qquad d_i^j = \frac{c_{2i}^{j-1} - c_{2i+1}^{j-1}}{\sqrt{2}}, \tag{6}$$

and

$$c_{2i}^{j-1} = \frac{c_i^j + d_i^j}{\sqrt{2}}, \qquad c_{2i+1}^{j-1} = \frac{c_i^j - d_i^j}{\sqrt{2}}. \tag{7}$$

This results in a fast algorithm for the analysis step and synthesis step. Various shrinkage functions leading to qualitatively different denoised functions u have been considered in literature, e.g.,

A. Linear shrinkage: $S(x) = \lambda x \qquad (\lambda \in [0,1])$,

B. Soft shrinkage [15]: $S_\theta(x) = \begin{cases} 0 & |x| \le \theta, \\ x - \theta\ \mathrm{sgn}(x) & |x| > \theta, \end{cases}$

C. Garrote shrinkage [18]: $S_\theta(x) = \begin{cases} 0 & |x| \le \theta, \\ x - \frac{\theta^2}{x} & |x| > \theta, \end{cases}$

D. Firm shrinkage [19]: $S_{\theta_1,\theta_2}(x) = \begin{cases} 0 & |x| \le \theta_1, \\ \mathrm{sgn}(x)\frac{\theta_2(|x|-\theta_1)}{\theta_2-\theta_1} & \theta_1 < |x| \le \theta_2, \\ x & \theta_2 < |x|, \end{cases}$

E. Hard shrinkage [32]: $S_\theta(x) = \begin{cases} 0 & |x| \le \theta, \\ x & |x| > \theta. \end{cases}$

3.2 Discrete Translation-Invariant Scheme

In practice one deals with discrete signals $\mathbf{f} = (f_i)_{i=0}^{N-1}$, where, for simplicity, N is a power of 2. Then Haar wavelet shrinkage starts by setting $c_i^0 = f_i$ and proceeds by analysis (6), shrinkage, and synthesis (7). Let us just consider a *single* wavelet decomposition level, i.e., we set $n = 1$. Then, using the convention that $c_i = c_i^1$ and $d_i = d_i^1$, we can drop the superscripts $j = 0$ and $j = 1$. By (6) and (7), Haar wavelet shrinkage on one level produces the signal $\mathbf{u}^+ = (u_i^+)_{i=0}^{N-1}$ with coefficients

$$u_{2i}^+ = \frac{c_i + S_\theta(d_i)}{\sqrt{2}} = \frac{f_{2i} + f_{2i+1}}{2} + \frac{1}{\sqrt{2}} S_\theta\left(\frac{f_{2i} - f_{2i+1}}{\sqrt{2}}\right), \tag{8}$$

$$u_{2i+1}^+ = \frac{c_i - S_\theta(d_i)}{\sqrt{2}} = \frac{f_{2i} + f_{2i+1}}{2} - \frac{1}{\sqrt{2}} S_\theta\left(\frac{f_{2i} - f_{2i+1}}{\sqrt{2}}\right). \tag{9}$$

Note that the single Haar wavelet shrinkage step (8)–(9) decouples the input signal into successive pixel pairs: the pixel at position $2i-1$ has no direct connection to its neighbour at position $2i$, and the procedure is not invariant to translation of the input signal. To overcome this problem, Coifman and Donoho [12] introduced the so-called *cycle spinning*: The input signal is shifted, denoised using wavelet shrinkage, shifted back, and the results of all such shifts are averaged. This procedure is equivalent to thresholding of nondecimated wavelet coefficients which can be implemented efficiently using the *algorithme à trous* [23]. For our single decomposition level, we need

only one additional shift to acquire translation invariance. The shifted Haar wavelet shrinkage yields the signal $\mathbf{u}^- = (u_i^-)_{i=0}^{N-1}$ with coefficients

$$u_{2i-1}^- = \frac{f_{2i-1} + f_{2i}}{2} + \frac{1}{\sqrt{2}}\, S_\theta\!\left(\frac{f_{2i-1} - f_{2i}}{\sqrt{2}}\right),$$
$$u_{2i}^- = \frac{f_{2i-1} + f_{2i}}{2} - \frac{1}{\sqrt{2}}\, S_\theta\!\left(\frac{f_{2i-1} - f_{2i}}{\sqrt{2}}\right).$$

Averaging the shifted results, one cycle of shift-invariant Haar wavelet shrinkage can be summarised into

$$\begin{aligned} u_i &= \frac{u_i^- + u_i^+}{2} \\ &= \frac{f_{i-1} + 2f_i + f_{i+1}}{4} + \frac{1}{2\sqrt{2}}\, S_\theta\!\left(\frac{f_i - f_{i+1}}{\sqrt{2}}\right) - \frac{1}{2\sqrt{2}}\, S_\theta\!\left(\frac{f_{i-1} - f_i}{\sqrt{2}}\right). \end{aligned} \tag{10}$$

4 Correspondence of Diffusivities and Shrinkage Functions

4.1 Basic Considerations

In order to derive the relation between the explicit diffusion scheme and translation-invariant Haar wavelet shrinkage, we rewrite the first iteration step in (2) using the initial condition $u_i^0 = f_i$ and the simplified notation $u_i^1 = u_i$ as

$$\begin{aligned} u_i &= \frac{f_{i-1} + 2f_i + f_{i+1}}{4} + \frac{f_i - f_{i+1}}{4} - \frac{f_{i-1} - f_i}{4} \\ &\quad - \tau\, g(|f_i - f_{i+1}|)\,(f_i - f_{i+1}) + \tau\, g(|f_{i-1} - f_i|)\,(f_{i-1} - f_i) \\ &= \frac{f_{i-1} + 2f_i + f_{i+1}}{4} \\ &\quad + (f_i - f_{i+1})\left(\frac{1}{4} - \tau\, g(|f_i - f_{i+1}|)\right) \\ &\quad - (f_{i-1} - f_i)\left(\frac{1}{4} - \tau\, g(|f_{i-1} - f_i|)\right). \end{aligned} \tag{11}$$

This coincides with (10) if and only if

$$\boxed{\frac{1}{2\sqrt{2}}\, S_\theta\!\left(\frac{x}{\sqrt{2}}\right) = x\,\left(\frac{1}{4} - \tau\, g(|x|)\right)} \tag{12}$$

Equation (12) is of central importance in our paper: It relates the shrinkage function S_θ of wavelet denoising to the diffusivity g of nonlinear diffusion. Provided that relation (12) holds true, a single step of wavelet shrinkage is equivalent to a single step of explicitly discretised nonlinear diffusion. The following two formulas are derived from (12) and can be used to obtain a shrinkage function S_θ from a diffusivity g, or vice versa.

$$S_\theta(x) = x\left(1 - 4\tau g(|\sqrt{2}x|)\right), \tag{13}$$

$$g(|x|) = \frac{1}{4\tau} - \frac{\sqrt{2}}{4\tau x}\, S_\theta\left(\frac{x}{\sqrt{2}}\right). \tag{14}$$

These equations will be essential for the next two subsections.

4.2 From Diffusivities to Shrinkage Functions

Let us first investigate equation (13) in detail. The examples from Section 3.1 show that typical shrinkage functions from the literature are odd functions (i.e. $S_\theta(-x) = -S_\theta(x)$) that satisfy

$$0 \le S_\theta(x) \le x \quad \text{for} \quad x > 0. \tag{15}$$

In Section 5 we will see that these conditions are responsible for ensuring certain stability properties of the shrinkage process. We can now specify the time step size τ in (13) such that these two conditions are always satisfied for bounded nonnegative diffusivities. In Section 2.1 we have seen that the diffusivities A–E are nonnegative and bounded from above by 1. In order to ensure that the corresponding shrinkage functions satisfy (15), the time step size has to fulfil $\tau \le 0.25$. Let us now investigate the shrinkage functions that correspond to the diffusivities A–E.

We observe that the linear diffusivity corresponds to the linear shrinkage function

$$S(x) = (1 - 4\tau)x.$$

Nonlinear shrinkage functions such as soft, garrote, firm and hard shrinkage satisfy $S'(0) = 0$, since their goal is to set small wavelet coefficients to zero. In order to derive shrinkage functions that correspond to the bounded nonlinear diffusivities B–E and satisfy $S'(0) = 0$ as well, we fix $\tau := 0.25$. Then we obtain the following novel shrinkage functions:

- The Charbonnier diffusivity corresponds to the shrinkage function

$$S_\lambda(x) = x\left(1 - \sqrt{\frac{\lambda^2}{\lambda^2 + 2x^2}}\right).$$

- The Perona–Malik diffusivity leads to

$$S_\lambda(x) = \frac{2x^3}{2x^2 + \lambda^2}.$$

- The Weickert diffusivity gives

$$S_\lambda(x) = \begin{cases} 0 & x = 0, \\ x \exp\left(-\frac{0.20718\,\lambda^8}{x^8}\right) & x \neq 0. \end{cases}$$

- The Tukey diffusivity leads to the shrinkage function

$$S_\lambda(x) = \begin{cases} \frac{4x^3}{\lambda^2} - \frac{4x^5}{\lambda^4} & |x| \leq \lambda/\sqrt{2}, \\ x & |x| > \lambda/\sqrt{2}. \end{cases}$$

Since the FAB diffusivity may attain negative values, it is not surprising that its corresponding shrinkage function

$$S_{\lambda_1,\lambda_2}(x) = x\left(1 - 2\exp(-2|x|^2/\lambda_1^2) + \exp(-2|x|^2/\lambda_2^2)\right)$$

with $\lambda_1 < \lambda_2$ may violate the condition (15).
Figures 1 and 2 illustrate these bounded diffusivities and their shrinkage functions.

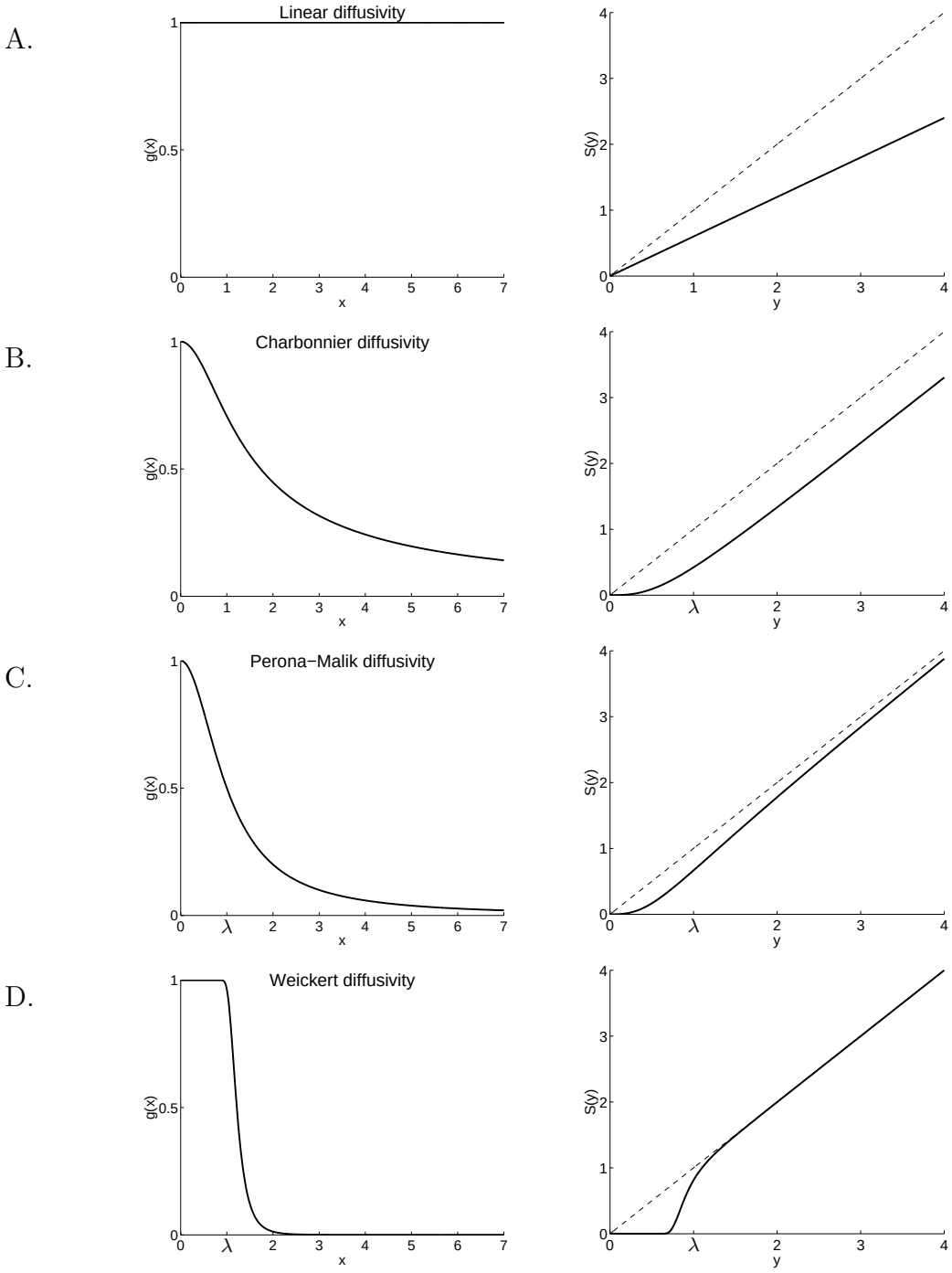

Figure 1: Diffusivity functions (left), corresponding shrinkage functions (right).
A. Linear diffusion. B. Charbonnier diffusivity. C. Perona-Malik diffusivity. D. Weickert diffusivity. The functions are plotted for $\tau = 0.1$ (linear diffusion), and $\tau = 0.25$, $\lambda = 1$ (all others).

E.

F.

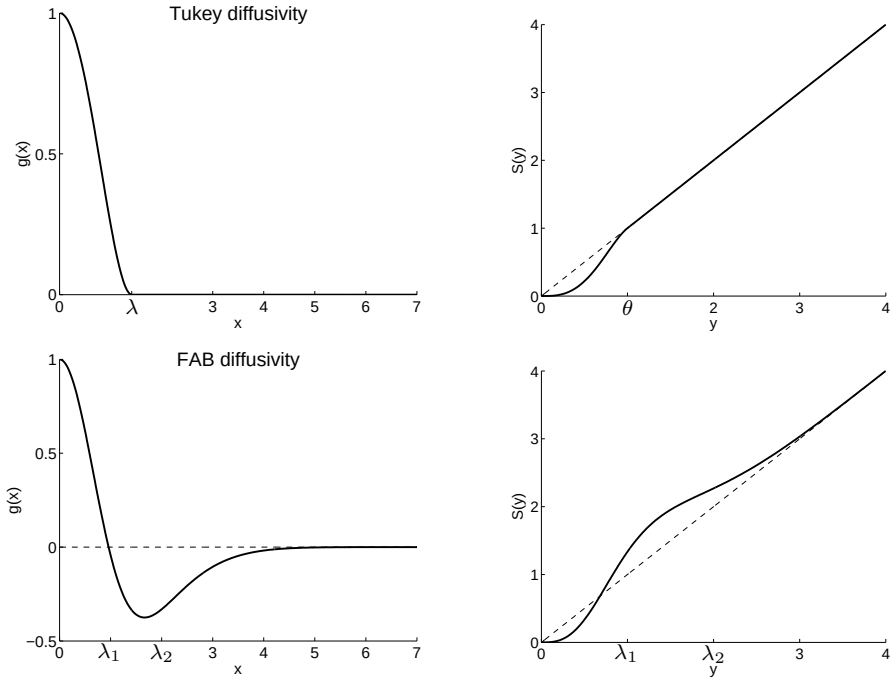

Figure 2: Diffusivity functions (left), corresponding shrinkage functions (right).
E. Tukey diffusivity F. Forward-and-backward (FAB) diffusivity. The functions are plotted for $\lambda/\sqrt{2} = \theta = 1$ (Tukey) and $\lambda_1 = 1$, $\lambda_2 = 2$ (FAB).

4.3 From Shrinkage Functions to Diffusivities

Having derived shrinkage functions from nonlinear diffusivities, let us now derive diffusivities from frequently used shrinkage functions. To this end, all we have to do is to plug in the specific shrinkage function into (14).
In the case of soft shrinkage, this gives the diffusivity

$$g(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq \theta\sqrt{2}, \\ \frac{\sqrt{2}\theta}{4\tau|x|} & |x| > \theta\sqrt{2}. \end{cases}$$

If we select the time step size τ such that $\theta = 2\sqrt{2}\tau$, we obtain a stabilised TV diffusivity (see also [40, 41] for an alternative derivation):

$$g(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq 4\tau, \\ \frac{1}{|x|} & |x| > 4\tau. \end{cases}$$

In the same way one can show that garrote shrinkage leads to a stabilised BFB diffusivity for $\theta = \sqrt{2\tau}$:

$$g(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq 2\sqrt{\tau}, \\ \frac{1}{|x|^2} & |x| > 2\sqrt{\tau}. \end{cases}$$

Firm shrinkage yields a diffusivity that degenerates to 0 for sufficiently large gradients:

$$g(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq \sqrt{2}\theta_1, \\ \frac{\theta_1}{4\tau(\theta_2-\theta_1)}\left(\frac{\sqrt{2}\theta_2}{|x|} - 1\right) & \sqrt{2}\theta_1 < |x| \leq \sqrt{2}\theta_2, \\ 0 & |x| > \sqrt{2}\theta_2. \end{cases}$$

This degeneracy resembles the behaviour of the Tukey diffusivity from Section 2.1. Another diffusivity that degenerates to 0 can be derived from hard shrinkage:

$$g(|x|) = \begin{cases} \frac{1}{4\tau} & |x| \leq \sqrt{2}\theta, \\ 0 & |x| > \sqrt{2}\theta. \end{cases}$$

All diffusivities in this subsection are depicted in Figure 3.

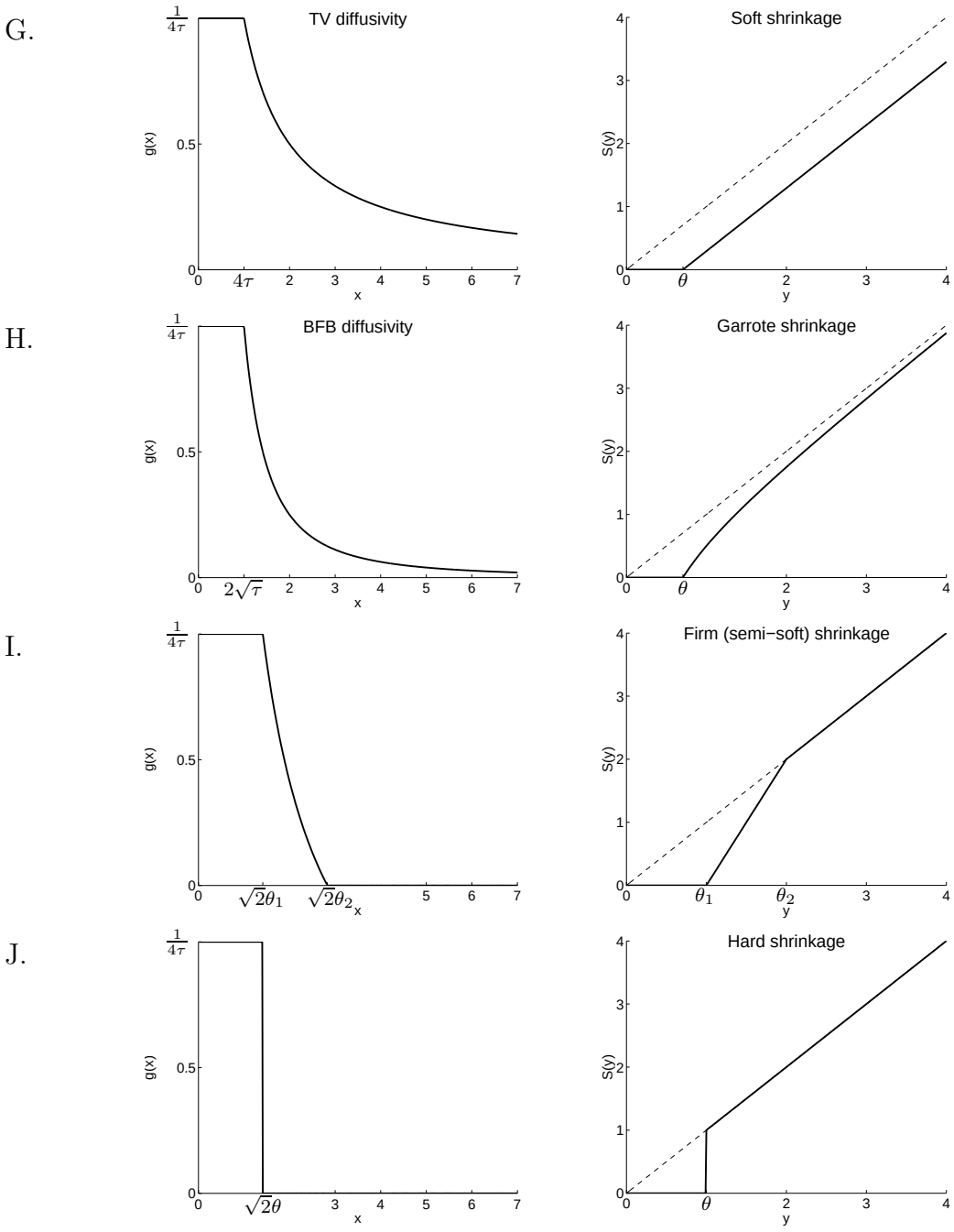

Figure 3: Diffusivity functions (left), corresponding shrinkage functions (right).
G. TV flow and soft shrinkage. H. Balanced forward-backward (BFB) diffusivity and garrote shrinkage. I. Firm shrinkage. J. Hard shrinkage.
The functions are plotted for $\tau = 0.25$ (which corresponds to $\theta = \tau 2\sqrt{2}$ for soft shrinkage, and to $\theta = \sqrt{2\tau}$ for garrote; the other use $\theta = \theta_1 = 1$, $\theta_2 = 2$).

5 Stability Analysis of Wavelet Shrinkage

In this section we exploit the connections between wavelet shrinkage and diffusion filtering in order to establish three stability properties for wavelet shrinkage: maximum–minimum stability, monotonicity preservation, and sign stability.

5.1 Maximum–Minimum Stability

Maximum–minimum stability of a discrete diffusion process states that the filtered signal at time step $k+1$ stays within the range of the data at step k. For a single step output $\mathbf{u}$ computed from the input signal $\mathbf{f}$ this maximum–minimum principle gives

$$\min_j f_j \leq u_i \leq \max_j f_j \qquad (i = 0, ..., N-1). \tag{16}$$

Maximum–minimum principles are essential in scale-space theory since they guarantee *causality* of the continuous evolution [24]: Level lines can be traced back in scale. The following proposition states conditions under which wavelet shrinkage is maximum–minimum stable.

Proposition 1 (Maximum–Minimum Stability).
The shift invariant single level Haar wavelet shrinkage (10) with an odd shrinkage function S_θ satisfies the discrete maximum–minimum principle (16) if

$$-x \leq S_\theta(x) \leq x \qquad \text{for all } x \geq 0. \tag{17}$$

If this condition does not hold, it is possible to construct counterexamples that violate the discrete maximum–minimum principle.

The proof for this result can be found in the appendix.
We observe that odd shrinkage functions satisfying (15) are maximum–minimum stable. By exploiting the relation (13), let us now re-interpret the stability condition (17) in terms of stability for nonlinear diffusion filtering. From

$$x(1 - 4\tau g(\sqrt{2}x)) = S_\theta(x) \leq x \qquad \forall\, x > 0$$

one immediately obtains

$$\tau g(x) \;\geq\; 0 \qquad \forall\, x > 0. \tag{18}$$

In the same way, the condition

$$-x \leq S_\theta(x) = x(1 - 4\tau g(\sqrt{2}x)) \qquad \forall\, x > 0$$

can be simplified to

$$\tau g(x) \;\leq\; \frac{1}{2} \qquad \forall\, x > 0. \tag{19}$$

Since the explicit scheme may be rewritten as

$$\begin{aligned} u_i \;&=\; \tau g(|f_{i+1}-f_i|)\, f_{i+1} \;+\; \tau g(|f_i-f_{i-1}|)\, f_{i-1} \\ &\quad + \big(1 - \tau g(|f_{i+1}-f_i|) - \tau g(|f_i-f_{i-1}|)\big)\, f_i \end{aligned} \tag{20}$$

it becomes clear that (18) and (19) guarantee that the weights in front of f_{i+1}, f_i and f_{i-1} are nonnegative. Since these weights sum up to 1, we have a convex combination that ensures stability in terms of a discrete maximum–minimum principle:

$$\min(f_{i-1}, f_i, f_{i+1}) \leq u_i \leq \max(f_{i-1}, f_i, f_{i+1}).$$

This can be regarded as an alternative to the proof in the appendix which is performed within the wavelet setting.

5.2 Preservation of Monotonicity

Maximum–minimum stability guarantees that the filtered signal does not leave the range of the original signal, but it does not give any statements on the behaviour within these bounds. One stability notion that takes into account such a behaviour is monotonicity preservation.
Assume that the input signal $\mathbf{f} = (f_0, f_1, ..., f_{N-1})^\top$ is monotonically increasing (or decreasing), then a filter is called *monotonicity preserving*, if the processed signal $\mathbf{u} = (u_0, u_1, ..., u_{N-1})^\top$ is monotonically increasing (or decreasing) as well.

From this definition it follows that monotonicity preserving filters cannot create oscillations for monotonic signals. The subsequent proposition shows that monotonicity preservation of wavelet shrinkage can be established under the same conditions as maximum–minimum stability.

Proposition 2 (Preservation of Monotonicity).
The shift invariant single level Haar wavelet shrinkage (10) with an odd shrinkage function S_θ is monotonicity preserving if

$$-x \leq S_\theta(x) \leq x \qquad \textit{for all } x \geq 0. \tag{21}$$

The proof of this proposition is a direct consequence of the connection between Haar wavelet shrinkage and explicit diffusion filtering: In the previous

subsection we have seen that single level Haar wavelet shrinkage satisfying condition (21) can be expressed as the explicit diffusion filter (20) with

$$0 \leq \tau g(x) \leq \frac{1}{2} \qquad \forall\, x > 0.$$

For such a diffusion scheme, monotonicity preservation has already been established in [45].

5.3 Sign Stability

The preceding discussion shows that the condition (17), which ensures maximum–minimum stability as well as positivity preservation, is less restrictive than the typical design property (15) for shrinkage functions. This gives rise to the conjecture that (15) may imply additional stability properties. Therefore, let us consider the notion of sign stability next.
A filter transforming an initial signal $\mathbf{f} = (f_0, f_1, ..., f_{N-1})^\top$ into a processed signal $\mathbf{u} = (u_0, u_1, ..., u_{N-1})^\top$ is called *sign stable* or *variation diminishing* in the sense of Schoenberg [38], if the number of sign changes in the components of $\mathbf{u}$ does not exceed the number of sign changes in $\mathbf{f}$. In this definition, zeros are not taken into account.
It should be noted that sign stability is a stronger stability requirement than it may seem at first glance: Since wavelet shrinkage is invariant under shifts of the average grey value, sign stability also implies that the number of level crossings does not increase for any level. In particular, by choosing the smallest and largest value of the initial signal as reference levels, sign stability always ensures maximum–minimum stability. Moreover, it also guarantees that a monotone input signal does not create an oscillatory filter output.
Sign stability results for wavelet shrinkage are established in the following proposition, which is proved in the appendix.

Proposition 3 (Sign Stability).
The shift invariant Haar wavelet shrinkage (10) with an odd shrinkage function S_θ is sign stable if

$$0 \leq S_\theta(x) \leq x \qquad \textit{for all } x \geq 0. \tag{22}$$

By means of our equivalence results, sign stability directly carries over to the explicit discretisation on nonlinear diffusion filters. This extends Glashoff's and Kreth's results [22] for finite difference schemes for *linear* diffusion to the *nonlinear* setting. It also gives rise to the conjecture that such a variation

diminishing property does not only hold in the discrete framework, but carries over to the continuous PDE formulation as well. To the best of our knowledge, continuous sign stability results are only available in the *linear* case so far [42].

It is instructive to analyse the sign stability condition (22) in detail. Using

$$S_\theta(x) = x(1 - 4\tau g(\sqrt{2}x)),$$

the condition $0 \leq S_\theta(x)$ for all $x > 0$ can be simplified to

$$\tau g(x) \;\leq\; \frac{1}{4} \qquad \forall\, x > 0.$$

This result explains why soft, garrote, firm and hard shrinkage correspond to diffusivities that have been cut off at $\frac{1}{4\tau}$: It is a direct consequence from the fact that wavelet shrinkage is sign stable.

6 Denoising Experiment

To test the applicability of the newly derived shrinkage functions from Section 4.2, we perform experiments with signal-denoising using the shift-invariant multiscale Haar wavelet transform from Section 3. The input signal *blocks*, one of the standard signals in wavelet denoising, mimics a scan line through a 2-D image depicting an object with several edges [16]. The signal is shown in Fig. 4. The same figure then shows examples of the results of multiscale Haar wavelet denoising when combined with several shrinkage functions introduced in previous sections.

Table 1 and Fig. 5 present additional experimental results obtained with the *blocks* data. Here we performed a series of experiments with several levels of additive zero-mean Gaussian noise in the input signal. The noise varies between SNR=1 and SNR=32, where the *signal-to-noise ratio (SNR)* is defined by

$$\mathrm{SNR} = 20 \log_{10} \frac{|z - \overline{z}|_2}{|n|_2},$$

with z standing for the ideal signal with mean $\bar{z}$, and n representing noise. The noise is generated five times for each input SNR level and the resulting output SNR was averaged. Then we used multiscale wavelet denoising with various shrinkage functions, and searched for the optimal parameters that maximise the signal-to-noise ratio in the filtered signal.

Table 1 summarises the average optimal SNR after filtering obtained with different shrinkage functions; Fig. 5 presents the same information graphically, together with the standard deviation of the results. We observe that for

SNR_{in} / Shrinkage method	1	2	4	8	16	32
Linear	3.6	4.2	5.5	8.7	16.1	32.0
Charbonnier	8.7	9.6	11.2	14.8	22.4	38.4
Soft (TV)	10.1	10.9	12.6	16.2	24.0	39.9
Perona-Malik	9.9	10.8	12.8	16.8	25.8	44.5
FAB	10.5	11.3	12.8	15.7	23.6	45.1
Tukey	11.6	12.6	14.7	18.9	28.0	45.8
Garrote (BFB)	11.9	12.9	15.0	19.3	28.5	46.0
Firm	12.8	13.8	15.9	20.2	29.0	46.1
Weickert	12.9	13.9	16.0	20.2	29.1	46.1
Hard	12.9	13.9	15.9	20.2	29.1	46.1

Table 1: Numerical results (measured by mean signal-to-noise ratio in the filtered signal) of wavelet denoising for the *blocks* data of length 1024. Each column represents a given level of noise in the input image; each row contains the results for one shrinkage function.

all noise levels, the best signal-to-noise ratio is obtained by those shrinkage functions which put small wavelet coefficients to zero and keep larger coefficients almost unaffected. The functions with these properties include hard shrinkage, firm shrinkage and – to some extent – the garrote shrinkage on the wavelet side. Of the diffusion origin, the experimentally best shrinkage functions correspond to the Weickert diffusivity, followed by the stabilised BFB diffusivity (which is equivalent to garrote shrinkage), the Tukey, the FAB and the Perona-Malik diffusivity. Interestingly, these are diffusivities with nonmonotone flux functions that allow even contrast enhancement.
The second group of shrinkage functions decreases even large wavelet coefficients by a constant (or almost constant) value; the functions with this behaviour include soft shrinkage, TV flow corresponding to it, and Charbonnier diffusivity. It seems that this strategy is less successful numerically. These diffusivities lead to monotonically increasing flux functions and well-posed diffusion filters.
As a group of its own, the denoising performance of linear diffusion (or its shrinkage function) is far worse than that of the nonlinear methods.

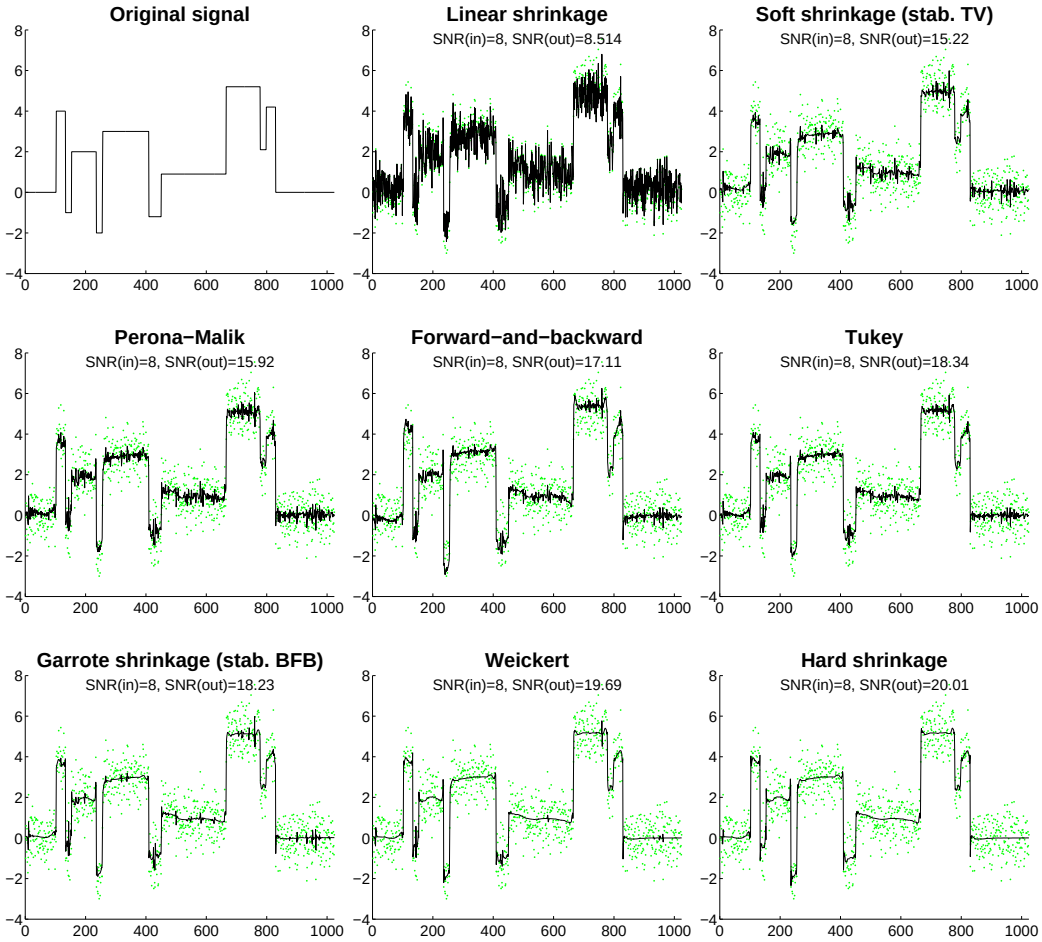

Figure 4: Example of multiscale translation-invariant Haar wavelet denoising. Normal noise of SNR=8 was added to the ideal signal, and different shrinkage functions have been applied. The noisy signal is represented by dots, reconstructed signal by solid line.

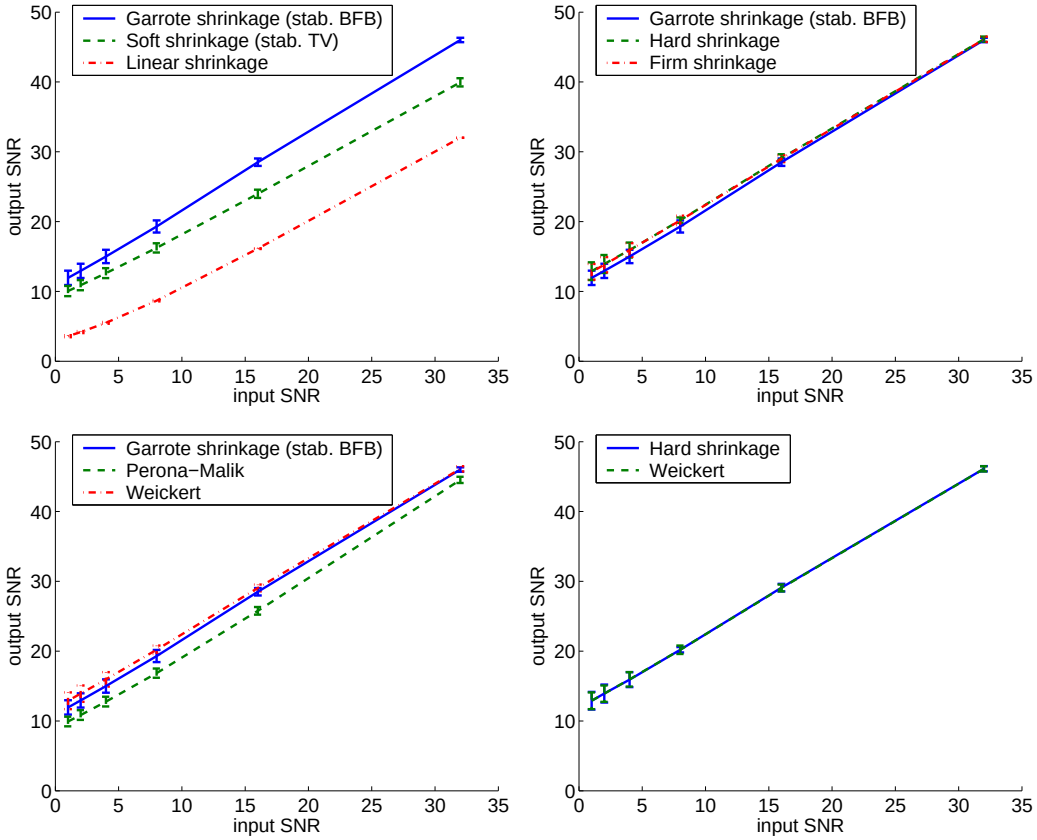

Figure 5: Comparing the optimal denoising performance of shift-invariant multiscale wavelet shrinkage with various shrinkage functions. SNR of the filtered signal is plotted against SNR of the input; the higher the graph, the better the result. The input signal was *blocks*, length 1024.
Top left: garrote shrinkage (BFB diffusivity), soft shrinkage (TV flow) and linear diffusion. **Top right:** garrote (BFB), hard and firm shrinkages. **Bottom left:** garrote (BFB), Perona-Malik and Weickert functions. **Bottom right:** best from either world, hard shrinkage and Weickert diffusivity give comparable results.

7 Conclusions

We have analysed correspondences between explicit one-dimensional schemes for nonlinear diffusion and discrete translation-invariant Haar wavelet shrinkage. We have shown that if we restrict the methods to one discrete step and a single spatial level, the two approaches can be made equivalent, if suitable diffusivities or shrinkage functions are chosen.

This connection between nonlinear diffusion and wavelet shrinkage opens the gate for a fruitful exchange of ideas between the two worlds. In this paper, we derived new wavelet shrinkage functions from frequently used nonlinear diffusivities. Vice versa, we showed that soft and garrote shrinkage may be regarded as stabilised TV or BFB diffusion, respectively. We experienced that some novel shrinkage functions inspired from rapidly decreasing diffusivities are competitive with the best previously known shrinkage methods when applied to signal denoising with multiscale wavelet procedures.

The connection between Haar wavelet shrinkage and explicit schemes for nonlinear diffusion filtering has also enabled us to establish three stability results for single scale wavelet shrinkage: maximum-minimum stability, monotonicity preservation, and sign stability. The proofs of these stability properties illustrate that we have a nice transfer of ideas in both directions: While maximum–minimum stability can be proved without difficulties in both the wavelet and the diffusion framework, monotonicity preservation of wavelet shrinkage has been derived from findings for diffusion filtering, while sign stability of nonlinear diffusion filtering followed from sign stability of wavelet shrinkage.

The results in this paper can be extended in several directions. One can study iterated multi-scale wavelet shrinkage as a hybrid method combining the efficiency of multi-scale wavelet shrinkage with the quality of iterated diffusion filtering [35]. This hybrid method may be also explained as nonlinear diffusion applied to the Laplacian pyramid of the signal [40, 41]. While the present paper focuses on the 1-D case, first 2-D results are reported in [33], where explicit schemes for nonlinear diffusion filtering are used to construct coupled shrinkage rules with improved rotation invariance. In our ongoing work we consider other wavelet bases and analyse the multiscale setting in more detail.

Acknowledgements

Our joint research is supported by the project *Relations between Nonlinear Filters in Digital Image Processing* within the DFG Priority Programme

1114: *Mathematical Methods for Time Series Analysis and Digital Image Processing*. This is gratefully acknowledged. J.W. also thanks Nir Sochen (Tel Aviv University) for drawing his attention on the FAB diffusivity.

References

[1] R. Acar and C. R. Vogel. Analysis of bounded variation penalty methods for ill–posed problems. *Inverse Problems*, 10:1217–1229, 1994.

[2] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, March 2001.

[3] Y. Bao and H. Krim. Towards bridging scale-space and multiscale frame analyses. In A. A. Petrosian and F. G. Meyer, editors, *Wavelets in Signal and Image Analysis*, volume 19 of *Computational Imaging and Vision*, chapter 6. Kluwer, Dordrecht, 2001.

[4] M. J. Black, G. Sapiro, D. H. Marimont, and D. Heeger. Robust anisotropic diffusion. *IEEE Transactions on Image Processing*, 7(3):421–432, March 1998.

[5] T. Brox, M. Welk, G. Steidl, and J. Weickert. Equivalence results for TV diffusion and TV regularisation. In L. D. Griffin and M. Lillholm, editors, *Scale-Space Methods in Computer Vision*, volume 2695 of *Lecture Notes in Computer Science*, pages 86–100, Berlin, 2003. Springer.

[6] E. J. Candés and F. Guo. New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction. *Signal Processing*, 82(11):1519–1543, 2002.

[7] 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):319–335, March 1998.

[8] A. Chambolle and B. L. Lucier. Interpreting translationally-invariant wavelet shrinkage as a new image smoothing scale space. *IEEE Transactions on Image Processing*, 10(7):993–1000, 2001.

[9] T. F. Chan and H. M. Zhou. Total variation improved wavelet thresholding in image compression. In *Proc. Seventh International Conference on Image Processing*, Vancouver, Canada, September 2000.

[10] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Proc. 1994 IEEE International Conference on Image Processing*, volume 2, pages 168–172, Austin, TX, November 1994. IEEE Computer Society Press.

[11] A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear approximation and the space $\mathrm{BV}(\mathrm{R}^2)$. *American Journal of Mathematics*, 121:587–628, 1999.

[12] R. R. Coifman and D. Donoho. Translation invariant denoising. In A. Antoine and G. Oppenheim, editors, *Wavelets in Statistics*, pages 125–150. Springer, New York, 1995.

[13] R. R. Coifman and A. Sowa. Combining the calculus of variations and wavelets for image enhancement. *Applied and Computational Harmonic Analysis*, 9(1):1–18, July 2000.

[14] R. R. Coifman and A. Sowa. New methods of controlled total variation reduction for digital functions. *SIAM Journal on Numerical Analysis*, 39(2):480–498, 2001.

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

[16] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrica*, 81(3):425–455, 1994.

[17] S. Durand and J. Froment. Reconstruction of wavelet coefficients using total-variation minimization. *SIAM Journal on Scientific Computing*, 24(5):1754–1767, 2003.

[18] H.-Y. Gao. Wavelet shrinkage denoising using the non-negative garrote. *Journal of Computational and Graphical Statistics*, 7(4):469–488, 1998.

[19] H.-Y. Gao and A. G. Bruce. WaveShrink with firm shrinkage. *Statistica Sinica*, 7:855–874, 1997.

[20] G. Gilboa, N. A. Sochen, and Y. Y. Zeevi. Forward-and-backward diffusion processes for adaptive image enhancement and denoising. *IEEE Transactions on Image Processing*, 11(7):689–703, 2002.

[21] G. Gilboa, N. A. Sochen, and Y. Y. Zeevi. Regularized shock filters and complex diffusion. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen,

editors, *Computer Vision – ECCV 2002*, volume 2350 of *Lecture Notes in Computer Science*, pages 399–413. Springer, Berlin, 2002.

[22] K. Glashoff and H. Kreth. Vorzeichenstabile Differenzenverfahren für parabolische Anfangswertaufgaben. *Numerische Mathematik*, 35:343–354, 1980.

[23] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian. A real-time algorithm for signal analysis with the help of the wavelet transform. In J. M. Combes, A. Grossman, and P. Tchamitchian, editors, *Wavelets: Time-Frequency Methods and Phase Space*, pages 286–297. Springer, Berlin, 1989.

[24] R. A. Hummel. Representations based on zero-crossings in scale space. In *Proc. 1986 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 204–209, Miami Beach, FL, June 1986. IEEE Computer Society Press.

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

[26] S. Karlin. *Total Positivity*. Stanford University Press, 1968.

[27] B. Kawohl and N. Kutev. Maximum and comparison principle for one-dimensional anisotropic diffusion. *Mathematische Annalen*, 311:107–123, 1998.

[28] S. L. Keeling and R. Stollberger. Nonlinear anisotropic diffusion filters for wide range edge sharpening. *Inverse Problems*, 18:175–190, January 2002.

[29] S. Kichenassamy. The Perona–Malik paradox. *SIAM Journal on Applied Mathematics*, 57:1343–1372, 1997.

[30] F. Malgouyres. Combining total variation and wavelet packet approaches for image deblurring. In *Proc. First IEEE Workshop on Variational and Level Set Methods in Computer Vision*, pages 57–64, Vancouver, Canada, July 2001. IEEE Computer Society Press.

[31] F. Malgouyres. Mathematical analysis of a model which combines total variation and wavelet for image restoration. *Inverse Problems*, 2(1):1–10, 2002.

[32] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, second edition, 1999.

[33] P. Mrázek and J. Weickert. Rotationally invariant wavelet shrinkage. In B. Michaelis and G. Krell, editors, *Pattern Recognition*, volume 2781 of *Lecture Notes in Computer Science*, pages 156–163, Berlin, 2003. Springer.

[34] P. Mrázek, J. Weickert, and G. Steidl. Correspondences between wavelet shrinkage and nonlinear diffusion. In L. D. Griffin and M. Lillholm, editors, *Scale-Space Methods in Computer Vision*, volume 2695 of *Lecture Notes in Computer Science*, pages 101–116, Berlin, 2003. Springer.

[35] P. Mrázek, J. Weickert, G. Steidl, and M. Welk. On iterations and scales of nonlinear filters. In O. Drbohlav, editor, *Proc. Eighth Computer Vision Winter Workshop*, pages 61–66, Valtice, Czech Republic, February 2003. Czech Pattern Recognition Society.

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

[37] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[38] I. J. Schoenberg. Über variationsvermindernde lineare Transformationen. *Mathematische Zeitschrift*, 32:321–328, 1930.

[39] B. Smolka. Combined forward and backward anisotropic diffusion filtering of color images. In L. Van Gool, editor, *Pattern Recognition*, volume 2449 of *Lecture Notes in Computer Science*, pages 314–320. Springer, Berlin, 2002.

[40] G. Steidl and J. Weickert. Relations between soft wavelet shrinkage and total variation denoising. In L. Van Gool, editor, *Pattern Recognition*, volume 2449 of *Lecture Notes in Computer Science*, pages 198–205. Springer, Berlin, 2002.

[41] 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. Technical Report 94, Department of Mathematics, Saarland University, Saarbrücken, Germany, August 2003. Submitted to *SIAM Journal on Numerical Analysis*.

[42] C. Sturm. Sur une classe d'equations á differences partielles. *Journal de Mathématiques Pures et Appliquées*, 1:373–444, 1836.

[43] R. S. Varga. *Matrix Iterative Analysis*. Prentice Hall, Englewood Cliffs, 1962.

[44] J. Weickert. *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

[45] J. Weickert and B. Benhamouda. A semidiscrete nonlinear scale-space theory and its relation to the Perona–Malik paradox. In F. Solina, W. G. Kropatsch, R. Klette, and R. Bajcsy, editors, *Advances in Computer Vision*, pages 1–10. Springer, Wien, 1997.

Appendix

A.1 Proof of Proposition 1 (Maximum–Minimum Stability)

1. Let the shrinkage function S_θ satisfy (17). To prove the maximum-minimum principle, we show for all $i = 0, \ldots, N-1$ that (17) implies

$$\min\{f_{i-1}, f_i, f_{i+1}\} \le u_i \le \max\{f_{i-1}, f_i, f_{i+1}\}. \tag{23}$$

If $f_{i-1} \le f_i \le f_{i+1}$, then we obtain by (17) and since S_θ is odd that

$$\begin{aligned} \frac{f_i - f_{i+1}}{\sqrt{2}} &\le S_\theta\left(\frac{f_i - f_{i+1}}{\sqrt{2}}\right) \le \frac{f_{i+1} - f_i}{\sqrt{2}}, \\ \frac{f_{i-1} - f_i}{\sqrt{2}} &\le S_\theta\left(\frac{f_{i-1} - f_i}{\sqrt{2}}\right) \le \frac{f_i - f_{i-1}}{\sqrt{2}} \end{aligned}$$

and consequently by (10) that

$$\min\{f_{i-1}, f_i\} \le \frac{f_{i-1} + f_i}{2} \le u_i \le \frac{f_i + f_{i+1}}{2} \le \max\{f_i, f_{i+1}\}.$$

Similarly it follows in case $f_{i-1} \ge f_i \ge f_{i+1}$ that

$$\min\{f_i, f_{i+1}\} \le \frac{f_i + f_{i+1}}{2} \le u_i \le \frac{f_{i-1} + f_i}{2} \le \max\{f_{i-1}, f_i\},$$

in case $f_{i-1} \le f_i$ and $f_i \ge f_{i+1}$ that

$$\min\{f_{i-1}, f_{i+1}\} \le \frac{f_{i-1} + f_{i+1}}{2} \le u_i \le f_i,$$

and in case $f_{i-1} \geq f_i$ and $f_i \leq f_{i+1}$ that

$$f_i \leq u_i \leq \frac{f_{i-1} + f_{i+1}}{2} \leq \max\{f_{i-1}, f_{i+1}\}.$$

This completes the proof of (23).

2. Conversely, let the shift invariant Haar wavelet shrinkage (10) satisfy a discrete maximum-minimum principle. Assume that there exists $\tilde{x} \geq 0$ where (17) is violated in such a way that $S_\theta(\tilde{x}) > \tilde{x}$.

   In order to construct a counterexample, we consider an input signal f which is zero except for $f_i = \sqrt{2}\tilde{x}$ such that $\max\limits_{i=0,\ldots,N-1} f_i = f_i$ and $\min\limits_{i=0,\ldots,N-1} f_i = 0$. Then we have by (10) that

$$u_i = \frac{f_i}{2} + \frac{1}{2\sqrt{2}} S_\theta(\tilde{x}) - \frac{1}{2\sqrt{2}} S_\theta(-\tilde{x}) > f_i$$

   which contradicts the maximum principle.
   Now assume that there exists $\tilde{x} \geq 0$ such that (17) is violated by $S_\theta(\tilde{x}) < -\tilde{x}$. Then we show for the same input sequence that $u_i < 0$ which contradicts the minimum principle. This completes the proof. □

A.2 Proof of Proposition 3 (Sign Stability)

By (10) the values u_i are the averages of u_i^+ and u_i^-. We start by considering u_i^+. By (8), (9) and (22), we obtain for $f_{2i} \geq f_{2i+1}$ that

$$\begin{aligned} \frac{f_{2i} + f_{2i+1}}{2} &\leq u_{2i}^+ \leq f_{2i}, \\ f_{2i+1} &\leq u_{2i+1}^+ \leq \frac{f_{2i} + f_{2i+1}}{2}. \end{aligned}$$

Similarly we can handle the case $f_{2i} \leq f_{2i+1}$. Hence we obtain u_{2i}^+, u_{2i+1}^+ by moving the pixels f_{2i}, f_{2i+1} the same distance into the direction of the mean value $(f_{2i} + f_{2i+1})/2$ without passing this value. In other words, for arbitrary fixed f there exist a_{2i}, b_{2i} with $1/2 \leq a_{2i} \leq 1$ and $b_{2i} = 1 - a_{2i}$ such that

$$\begin{pmatrix} u_{2i}^+ \\ u_{2i+1}^+ \end{pmatrix} = \begin{pmatrix} a_{2i} & b_{2i} \\ b_{2i} & a_{2i} \end{pmatrix} \begin{pmatrix} f_{2i} \\ f_{2i+1} \end{pmatrix} \qquad (i = 0, ..., N/2 - 1).$$

Similarly we can handle u_i^-, where we group f_{2i-1} and f_{2i} together:

$$\begin{pmatrix} u_{2i-1}^- \\ u_{2i}^- \end{pmatrix} = \begin{pmatrix} a_{2i-1} & b_{2i-1} \\ b_{2i-1} & a_{2i-1} \end{pmatrix} \begin{pmatrix} f_{2i-1} \\ f_{2i} \end{pmatrix} \qquad (i = 1, ..., N/2 - 1),$$

with $1/2 \leq a_{2i-1} \leq 1$ and $b_{2i-1} = 1 - a_{2i-1}$.
Since $N = 2^n$ is even, we have two unaffected boundary values in this case:

$$\begin{aligned} u_0^- &= f_0, \\ u_{N-1}^- &= f_{N-1}. \end{aligned}$$

Now $u_i = (u_i^+ + u_i^-)/2$ implies that

$$u = Qf$$

where $Q = Q(f)$ is the symmetric tridiagonal matrix

$$\frac{1}{2} \begin{pmatrix} c_0 & b_0 & & & & & \\ b_0 & c_1 & b_1 & & & & \\ & \ddots & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & b_{N-3} & c_{N-2} & b_{N-2} \\ & & & & b_{N-2} & c_{N-1} \end{pmatrix},$$

with

$$\begin{aligned} c_i &:= a_{i-1} + a_i \qquad (i = 0, ..., N-1), \\ a_{-1} &:= 1, \\ a_{N-1} &:= 1. \end{aligned}$$

Obviously we have $c_i \geq 1$ and $0 \leq b_i \leq 1/2$ for $i = 0,...,N-1$. Thus, from Gershgorin's theorem [43] it immediately follows that Q is positive semidefinite. Interestingly, this also implies that Q is totally positive, i.e. all its minors are nonnegative: Theorem 3.2 in [26] states that a symmetric tridiagonal matrix is totally positive, if it is positive semidefinite and if all subdiagonal elements b_i are nonnegative. It is shown in a classical paper by Schoenberg [38] that if Q is totally positive, then the corresponding linear transform is sign stable. Thus, $u = Qf$ cannot have more sign changes than f and we are done. □