

## On Iterations and Scales of Nonlinear Filters

Pavel Mrázek<sup>1</sup>, Joachim Weickert<sup>1</sup>, Gabriele Steidl<sup>2</sup> and Martin Welk<sup>1</sup>

<sup>1</sup>Faculty of Mathematics and Computer Science, Building 27.1  
Saarland University, 66123 Saarbruecken, Germany  
<http://www.mia.uni-saarland.de>  
{mrazek—weickert—welk}@mia.uni-saarland.de

<sup>2</sup>Faculty of Mathematics and Computer Science, D7, 27  
University of Mannheim, 68131 Mannheim, Germany  
steidl@math.uni-mannheim.de

**Abstract** *Methods modeled by partial differential equations (PDEs), and wavelet-based shrinkage procedures belong to the most successful approaches to nonlinear signal processing. The two groups of methods are quite different on the first glance: PDE based methods find the solution iteratively, while wavelet shrinkage performs a single step on multiple scales of the signal.*

*On the examples of total-variation (TV) diffusion on one hand, and translation-invariant soft Haar wavelet shrinkage on the other, we study the role of iterations and multiple scales for nonlinear filtering. Iterations and multi-scale processing go in the same direction in the sense that they allow simple, local operations to lead to global effects. We demonstrate that it may be advantageous to combine both, and create a powerful and efficient iterative multi-scale nonlinear procedure.*

### 1 Introduction

For the task of signal simplification or restoration, we typically want to remove insignificant, small-scale variations while preserving important features such as discontinuities or edges. A good discontinuity-preserving solution cannot be obtained by any linear method. From the previously proposed nonlinear methods, we concentrate in this paper on total variation (TV) diffusion, and translation-invariant soft Haar wavelet shrinkage.

TV diffusion filter (also called total variation flow) [1, 2] belongs to the class of nonlinear diffusion methods which start from the noisy input, and the possible solution develops with a process described by a nonlinear partial differential equation. In the discrete setting, the nonlinear equation is approximated by local operations on the data samples, and these operations have to be iterated many times in order to approach the desired continuous behaviour: signal smoothing with preservation of important structures, such as discontinuities.

Wavelet transform expresses the signal in terms of wavelet coefficients, describing the signal variation at different scales; this transform comes down to a simple change of basis. If

the basis is chosen properly, a signal will be described by only a few significant wavelet coefficients, while moderate white Gaussian noise pollutes all the wavelet coefficients to a small extent. Signal denoising by wavelet shrinkage [6, 5] starts from this assumption, and creates a smoothed version of the processed signal by removing smaller wavelet coefficients on all scales.

Both the TV flow (and other nonlinear diffusion methods) and wavelet shrinkage represent very successful nonlinear filters serving the same purpose. However, not much research has been devoted to their comparison. Steidl *et al.* proved in a recent paper [11] the equivalence of a discrete algorithm for TV flow and translation-invariant soft Haar wavelet thresholding in 1D, on the conditions that the wavelet transform was limited to a single decomposition scale, and that the result was obtained using a single iteration. However, both classical wavelet filters and TV flow algorithms differ from this setting: the wavelet transform is computed on multiple scales, and the TV flow scheme is iterated.

In this paper we concentrate on this significant difference (single scale and many iterations of TV flow, versus multiple scales and single step of wavelet shrinkage), and try to clarify the practical meaning of iterations and multi-scale processing for the above mentioned nonlinear filters. We will see that the result of these two approaches is not identical; however, iterations and multiple scales go in the same direction in the sense of allowing local filtering operations to lead to global effects. Combining both approaches seems advantageous: we introduce a hybrid, multi-scale iterated nonlinear filter, and demonstrate experimentally that it may retain the high degree of nonlinearity and adaptability of the iterated method while using much fewer iterations than necessary at a single scale. This leads to a higher computational efficiency.

A related topic has been studied by Chambolle and Lucier [3] who showed that continuous wavelet shrinkage creates a scale-space. Our work differs from their in concentrating on the connection between iterated wavelet shrinkage and TV flow in the discrete setting, and comparing the practical properties of the two filters.

The paper is organized as follows. Sections 2–3 give a slightly more detailed introduction to TV flow and wavelet shrinkage. The role of iterations and scales for the nonlinear filters are discussed in Section 4. The practical impact of multiple scales and iterations on the filtering result is signal-dependent; however, the experiment in Section 5 suggests that a multi-scale iterated wavelet filter might be used as a highly efficient approximation to the TV flow filter. The denoising performance of single-scale iterated, multi-scale iterated and multi-scale noniterated filters is compared in Section 6. Section 7 concludes the paper.

## 2 TV flow

The TV flow belongs to the class of nonlinear diffusion methods which can be described by the following partial differential equations:

$$\frac{\partial u}{\partial t} = \operatorname{div}(g(|\nabla u|) \nabla u) \quad \text{on } \Omega \times (0, \infty) \quad (1)$$

$$u(0) = f \quad \text{on } \Omega \quad (2)$$

$$\frac{\partial u}{\partial n} = 0 \quad \text{on } \partial\Omega \times (0, \infty) \quad (3)$$

where  $f$  is the original (noisy) signal,  $u$  the smoothed signal which evolves with time  $t$ ,  $\Omega$  is the signal (or image) domain, and  $n$  stands for the normal to the domain boundary  $\partial\Omega$ . In words, the solution  $u$  starts from the noisy signal  $f$  at time  $t = 0$  (2), and evolves locally in time with the flux to the neighbouring positions (1); employing homogeneous Neumann boundary conditions (3), no flow passes through the domain boundary. The diffusivity function  $g$  in (1) controls the amount of smoothing: typically, it is a positive decreasing function of the signal gradient magnitude, and the resulting procedure tends to preserve important discontinuities while removing small-scale fluctuations caused by noise. For TV flow [1, 2],  $g = \frac{1}{|\nabla u|}$  is chosen.

When modeled numerically, the TV diffusion equations (1)–(3) have to be discretized both in space and time. The solution  $u$  is then computed iteratively, leading to a series of solutions  $u^k \approx u(k\tau)$  where  $\tau$  is the iteration time step. The result  $u^{k+1}$  is created from  $u^k$  by operations with mostly local effect: for the simplest, explicit scheme, a pixel influences only its neighbours.

## 3 Wavelet shrinkage

The discrete wavelet transform represents a one-dimensional signal  $f(x)$  in terms of shifted versions of a dilated low-pass scaling function  $\phi(x)$ , and shifted and dilated versions of a bandpass wavelet function  $\psi(x)$ . For orthogonal wavelets, this gives

$$f(x) = \sum_{i \in Z} \langle f, \phi_i^M \rangle \phi_i^M + \sum_{j=-\infty}^M \sum_{i \in Z} \langle f, \psi_i^j \rangle \psi_i^j \quad (4)$$

where the lower index  $i$  stands for spatial position, upper index  $j$  represents the level of scale (up to a chosen maximum  $M$ ), a scaled wavelet is obtained by  $\psi_i^j(x) = 2^{-j/2} \psi(2^{-j}x - i)$ , and  $\langle \cdot, \cdot \rangle$  denotes the inner product in

$L_2(R)$ . If the measurements  $f$  are corrupted by white Gaussian noise, the noise will effect all the wavelet coefficients by a small amount, while the original signal is generally determined by a few significant wavelet coefficients. Wavelet shrinkage therefore attempts to eliminate noise from the signal using the following 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  $c_i^M = \langle f, \phi_i^M \rangle$ .
2. *Shrinkage*: apply a shrinkage function  $S_\theta$  to the wavelet coefficients  $d_i^j$ .
3. *Synthesis*: reconstruct a denoised version  $u$  of  $f$  from the shrunken wavelet coefficients:

$$u(x) = \sum_{i \in Z} \langle f, \phi_i^M \rangle \phi_i^M + \sum_{j=-\infty}^M \sum_{i \in Z} S_\theta(\langle f, \psi_i^j \rangle) \psi_i^j. \quad (5)$$

Wavelet shrinkage was developed mainly by Donoho *et al.* [6, 5]. In this paper, we limit our attention to Haar wavelets, well suited for piecewise constant signals with discontinuities. The Haar wavelet and scaling functions are given respectively by

$$\psi(x) = \mathbf{1}_{[0, \frac{1}{2})} - \mathbf{1}_{[\frac{1}{2}, 1)}, \quad (6)$$

$$\phi(x) = \mathbf{1}_{[0, 1)} \quad (7)$$

where  $\mathbf{1}_{[a, b)}$  denotes the characteristic function, equal to 1 on  $[a, b)$  and zero everywhere else. From the existing shrinkage functions  $S_\theta$ , we concentrate on soft shrinkage [5]:

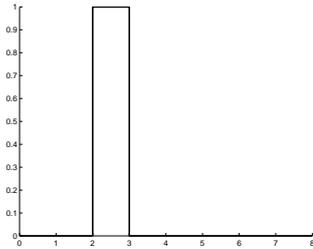
$$S_\theta(x) = \begin{cases} x - \theta \operatorname{sgn}(x) & \text{if } |x| > \theta \\ 0 & \text{if } |x| \leq \theta. \end{cases} \quad (8)$$

The shrinkage parameter  $\theta$  is chosen with respect to the amount of noise in the input signal. The classical approach [9] is to use the same threshold  $\theta$  for the shrinkage function at each spatial level of the coefficients. However, using a uniform threshold in this way leads to oscillations (Gibbs phenomenon) near discontinuities of the reconstructed signal. It was shown in [11] that for soft wavelet shrinkage, these oscillations are significantly reduced if we use a different threshold  $\theta_m$  at each level  $m$  according to the rule

$$\theta_m = \frac{\theta}{\sqrt{2}^{m-1}}. \quad (9)$$

We employ this scaled thresholding in all multiscale shrinkage experiments below.

Let us stress once more two notions that will appear in the text. We will call *scale* or *level* the spatial scale of a wavelet coefficient; scale was denoted by upper index  $j$  in (4), (5). The wavelet representation of a discrete signal may consist of coefficients from  $j = 1$  up to a chosen maximum level  $M$ . The maximum level may lie in the range  $M \in \{1, \dots, \log_2(N)\}$  where  $N$  is the input data size. (On the assumption that  $N$  is a dyadic length, all wavelet coefficients at scales higher than  $\log_2(N)$  are equal to zero.)



**Figure 1:** A simple input to demonstrate the role of iterations and spatial levels.

Then, we will use the term *iteration* for the cycle consisting of forward wavelet transform, shrinkage of the wavelet coefficients using a threshold  $\theta$ , and inverse wavelet transform. One such iteration starts from a signal  $u(t)$  and creates a signal  $u(t + \theta)$ . Successively repeating these iteration cycles, we obtain an *iterated wavelet shrinkage filtering procedure*. Note that iterating the wavelet shrinkage is meaningful only if we are using a shift-invariant formulation of the transform, where some form of averaging of shifted versions of the filtering result takes place (see cycle spinning by Coifman and Donoho [4], or an efficient implementation using the ‘algorithme à trous’ [8, 7]). If a classical (decimated) discrete transform is employed,  $n$  iterations of soft shrinkage with parameter  $\theta$  described above collapse to a single iteration with a shrinkage parameter  $n\theta$ . With shift-invariant transform, the results of the iterated and non-iterated filtering procedures will not be identical.

We have mentioned that a single-scale iterated soft Haar wavelet shrinkage is equivalent to a discrete algorithm for TV flow. From now on, we use only the wavelet terminology with scales and iterations to describe the methods discussed.

## 4 Role of iterations and scales

The role of iterations and of the number of spatial scales employed can be best explained by an example. Figure 1 shows simple input data, Fig. 2 presents the influence of iterations and number of spatial levels on the shrinkage result.

With the single-level decomposition (left column of Fig. 2), one iteration of the shrinkage has a local effect only. We have to repeat the iteration many times (or, in other words, to divide the total threshold  $\theta$  into many small steps) to (i) spread the information globally over the function domain (assemble local effects into global action), and (ii) ensure that the resulting process corresponds to the continuous model of a TV flow.

Using multiple levels of wavelet decomposition, already a single iteration results in global effects (Fig. 2 top centre and right). Iterating the multi-scale procedure further flattens homogeneous regions, as desired. Note on the other hand that the multiscale wavelet filter does not correspond to the explicit discretization of TV diffusion precisely; it was shown in [10] that the multiscale approach corresponds to the TV flow on a Laplacian pyramid of the signal. Values at more distant positions influence each other in the multiscale wavelet setting, whereas in the explicit scheme for TV flow only the neighbours communicate. As an example, the re-

gions to the left and to the right from the peak have different sizes, and with TV flow would not end at the same height (Fig. 2 right).

In both single- and multi-level cases, too small a number of iterations has the effect that a diffusion-like smoothing process dominates over the TV flow, and discontinuities are blurred. How many iterations are sufficient depends on the particular signal, its dimension, and on the number of decomposition levels.

Unfortunately, the exact behaviour of single- and multi-scale iterated transforms is signal-dependent. In the next two sections we study the practical filter properties on generic examples of noisy data with discontinuities.

## 5 Experiment on iterations vs. levels

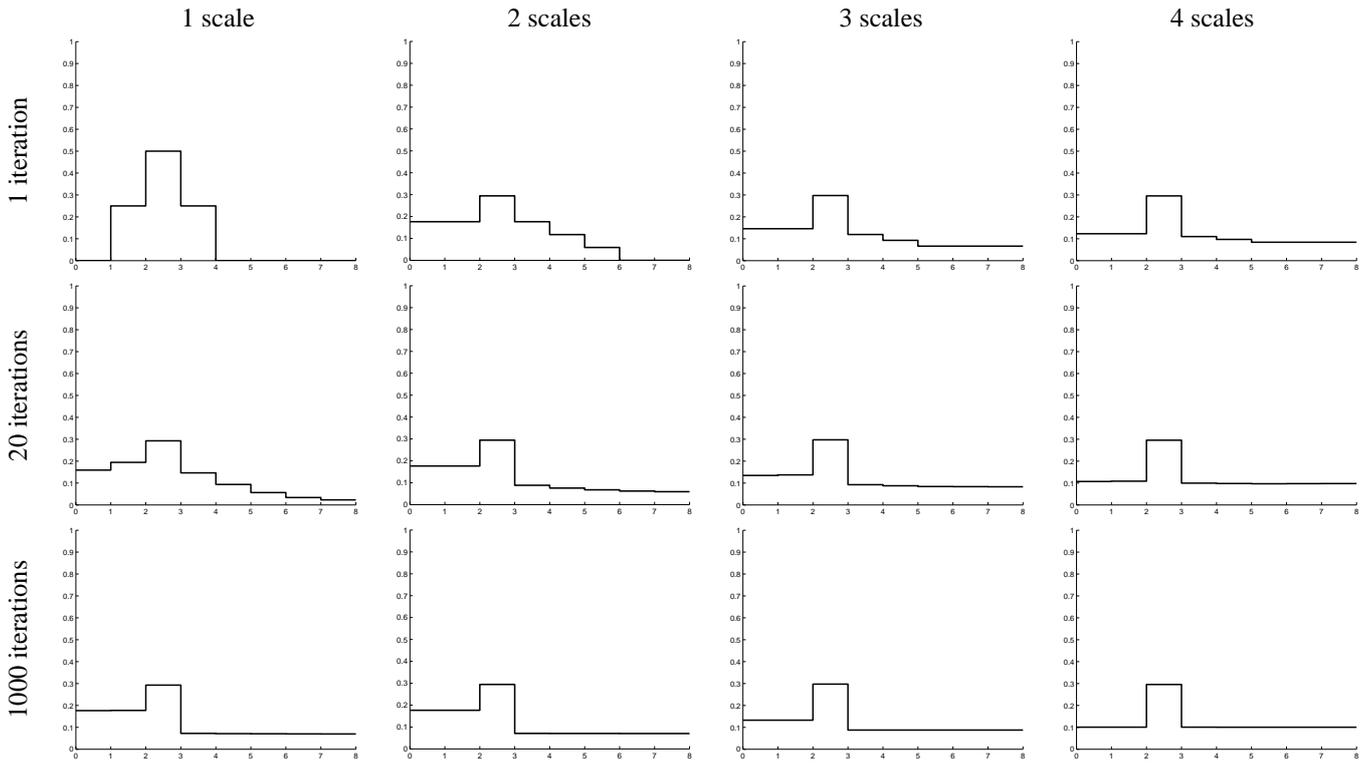
We have seen in so far that iterated wavelet-shrinkage transforms on a single and multiple levels may provide similar results, although at different values of the (accumulated) threshold parameter.<sup>1</sup> We have also experienced that fewer iterations of the multi-level transform were sufficient to yield an effect for which more iterations of a single-level transform were needed. In this section we analyse this point in more detail, and study the influence of the number of iterations and the number of levels on filtering results.

We start the filtering process from the noisy data in Fig. 4, run the iterated filters (with the chosen parameters  $M$  and  $\theta$ ), and stop when the filtered signal is closest to a precomputed ‘ground truth’ obtained using single-level iterated filter with a small time step ( $M = 1$ ,  $\theta = 0.001$ , 53800 iterations; this filter represents the TV flow). Numerical evaluation of the results obtained with varied number of levels and varied iteration time step can be seen in Tables 1 and 2.

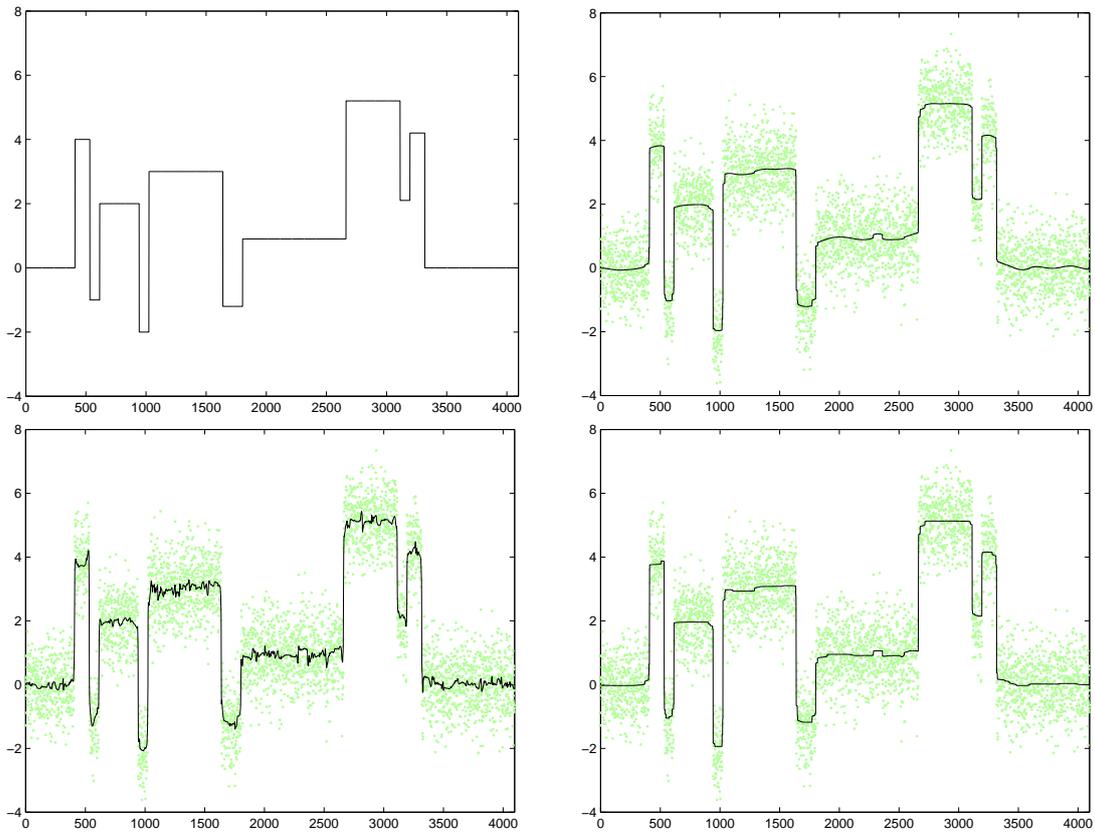
Table 2 gives the values of the accumulated threshold  $T$  (sum of single-iteration thresholds  $\theta$  over all iterations) at which the filtered signal was closest to the ground truth. Note how the scaling of the accumulated threshold depends on the maximum level  $M$ . Using the scaled thresholds from formula (9), we should ideally (i.e. according to theoretical spatial scaling between levels) approach the single-scale result of accumulated threshold  $T$  using a multi-scale iterative shrinkage with a threshold  $T/M$ , on two assumptions: that we use enough iterations to approximate the continuous process, and that the filtered signal contains equal amount of energy at all levels up to  $M$ . This scaling rule is approximately valid in the first two rows of Table 2 up to  $M = 4$ .

We can see in Table 1 that with enough iterations, filters with any number of levels lead to only small relative error (measured with  $l_1$  norm from the single-level result with  $\theta = 0.001$ ). Interesting effect can be observed if the number of iteration decreases: then, the filters with small number of spatial levels deteriorate faster. As one iteration of the filter has the complexity of order  $O(M \cdot N)$ , we would propose to balance the number of iterations needed to obtain a desired precision and the number of levels. A good solution

<sup>1</sup>There is not a simple correspondence between the threshold, or ‘time’, of the iterated transforms using different number of levels. Generally, using a threshold  $\theta$  with a multi-level filter changes the signal more than for the same threshold and a single-level transform, but the exact scaling is signal-dependent.



**Figure 2:** Demonstration of the role of iterations and scales; the chosen shrinkage threshold was divided into (top to bottom) 1, 20, 1000 iterations. Left column: single scale,  $T = 1$ . Second column: two scales,  $T = 0.66$ . Third column: three scales,  $T = 0.60$ . Right column: four scales (maximum),  $T = 0.58$ . The bottom left figure corresponds to TV flow, top right to classical wavelet shrinkage.



**Figure 3:** Piecewise constant signal, length 4096, noise of SNR=8 filtered by several procedures. Top left: original without noise. Top right: result of iterated single-level filter (equivalent to TV flow), 19564 iterations with  $\theta = 0.001$ , SNR=25.2. Bottom left: filtered in single step on 12 levels (classical wavelet shrinkage),  $\theta = 3.0$ , SNR = 20.9. Bottom right: iterated multi-scale procedure, 12 levels, 3011 iterations with  $\theta = 0.001$ , SNR = 24.8.

$\theta$	Number of levels				
	1	2	4	8	13
0.01	0.13	0.15	0.17	0.28	0.30
0.1	0.53	0.43	0.31	0.35	0.36
1	1.08	0.84	0.53	0.47	0.49
2	1.35	1.02	0.59	0.50	0.52
5	1.75	1.37	0.73	0.72	0.75
10	2.06	1.66	0.94	1.06	1.10

**Table 1:** Relative error in terms of  $l_1$  norm (per cent), measured from the ‘ground truth’ TV flow result obtained on a single level using 53800 iterations with  $\theta = 0.001$ .

$\theta$	Number of levels				
	1	2	4	8	13
0.01	53.7	27.2	14.4	11.3	11.1
0.1	58.2	28.0	14.5	11.2	11.2
1	129.0	49.0	18.0	13.0	13.0
2	148.0	62.0	20.0	14.0	14.0
5	160.0	80.0	25.0	20.0	20.0
10	240.0	90.0	40.0	30.0	30.0

**Table 2:** Accumulated threshold (or ‘time’) needed to get nearest to the ‘ground truth’ using filters with varied number of spatial levels and with varied iterated threshold  $\theta$ .

to reduce computational complexity may be represented by a filter of medium number of spatial levels (e.g.  $\approx \frac{\log_2(N)}{2}$ ); the dependence between relative precision and computation cost is plotted in Fig. 5.

## 6 Signal-denoising experiments

In the next experiment, we run several types of wavelet-based filters on a noisy version of the *blocks* data in Fig. 3 top left. The following types of soft-shrinkage wavelet-based filters were employed:

**Single-level iterated.** The wavelet decomposition is performed at the finest spatial scale only; the cycle (decomposition + coefficient shrinkage + inverse transform) is iterated, taking the filtered result from the previous step as input for the next iteration. This iterative procedure corresponds to a numerical scheme for TV flow.

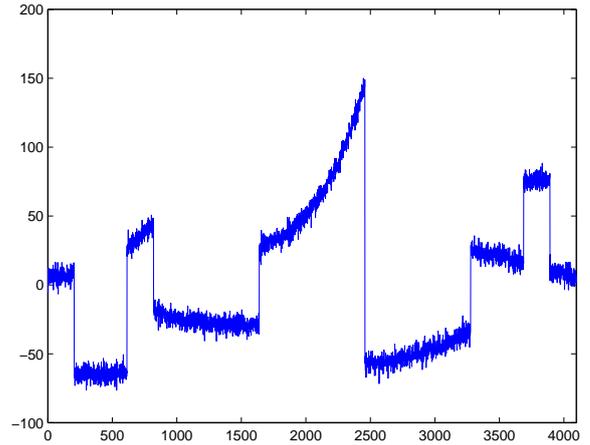
**Multiple levels, single iteration.** This is the classical procedure for shift-invariant soft wavelet shrinkage: coefficients are obtained on multiple scales (here  $\log_2(N)$  where  $N$  is the number of data samples), soft thresholding is applied to all of them (with scaled thresholds  $\theta_m$  at each spatial level  $m$ , as given by Eq. (9)), and the result is transformed back to the original space.

### Multiple levels iterated.

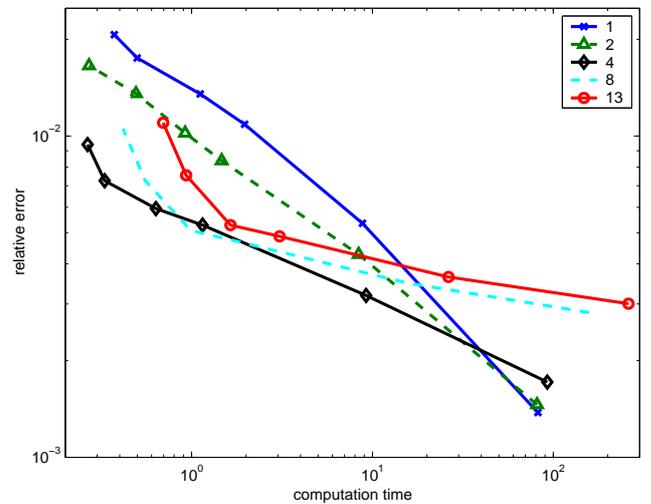
An iterated version of the multi-level procedure.

In all cases of iterated transforms, the shrinkage parameter of a single iteration was set to a small value ( $\theta = 0.001$ ) so that the discretization effects can be neglected.

The filtered signals are presented in Fig. 3. The best restoration result in terms of signal-to-noise ratio (defined



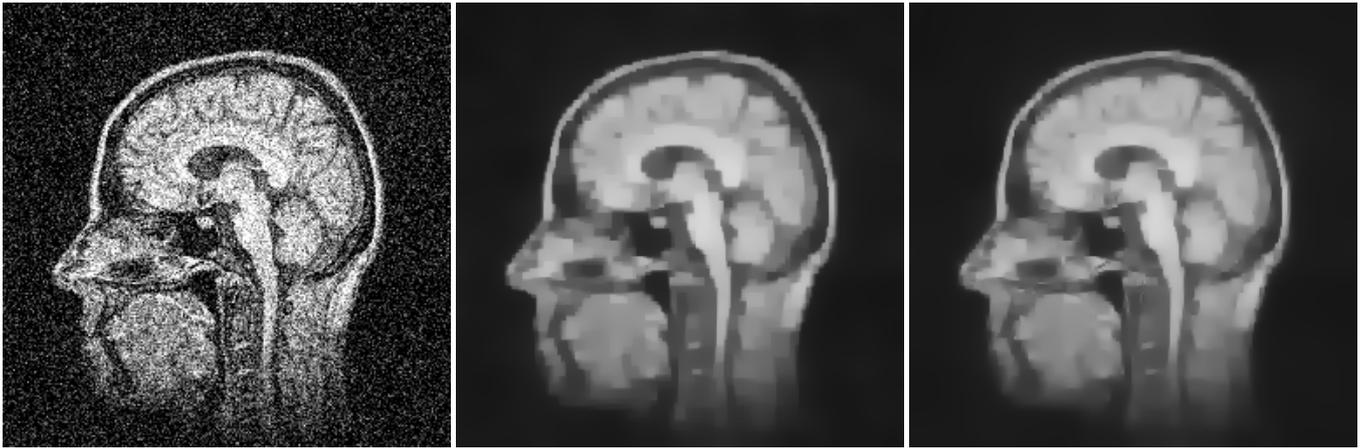
**Figure 4:** Piecewise polynomial signal with additive Gaussian white noise (SNR = 21.9).



**Figure 5:** Computation time vs. relative error of the procedure for different number of spatial levels (see legend in the graph) and for varied time-step  $\theta$  (increasing from right to left). At the price of a factor 4 in the relative error, the computations can be speeded up by a factor 100: instead of 5000 iterations of a single-level filter (with  $\theta = 0.01$ ), 18 iterations of a 4-level filter (using  $\theta = 1$ ). (Computation time in seconds, matlab code, AMD Athlon 1800+)

by  $\text{SNR} = 20 \log_{10} \frac{|a - \bar{a}|_2}{|n|_2}$ , with  $a$  standing for the ideal signal with mean  $\bar{a}$ , and  $n$  representing noise) is obtained using the single-level iterated transform (Fig. 3 top right, SNR=25.2), and a similar result is reached using iterated transform on multiple levels with scaled threshold (Fig. 3 bottom right, SNR=24.8). Although the two methods are not exactly equivalent, they also reveal a high level of visual similarity, and provide a good piecewise-constant approximation to the desired signal. The multi-level transform with scaled threshold and a single iteration (Fig. 3 bottom left, SNR=20.9) performs worse.

Another noise-filtering example, this time on a 2D medical image, is presented in Fig. 6. Also there, the single-scale and multi-scale iterated procedures reveal a high level of visual similarity, although at different values of the accumulated shrinkage parameter.



**Figure 6:** Left: noisy input image. Middle: filtered with single-level iterated soft shrinkage, 200 iterations with  $\theta = 1$ . Right: iterated wavelet shrinkage using 4 scales and 15 iterations with  $\theta = 10$ .

## 7 Conclusion

In this paper we studied practical properties of a wavelet-based multi-scale iterated nonlinear filter: iterated shift-invariant soft Haar wavelet shrinkage with scaled thresholds between levels. Depending on the chosen parameters, the filter represents a discrete scheme for total variation flow (single scale, many iterations) or classical wavelet shrinkage (single iteration on multiple scales). In this framework, we examined the role of multi-scale processing and iterations for nonlinear filters. Iterations and multi-scale processing have a similar effect in the sense that they allow simple, local operations to lead to global effects. We have shown that combining both is advantageous: a multi-scale iterated filter retains the high degree of nonlinearity and adaptability of the iterated method while using much fewer iterations than necessary at a single scale. Although the two methods are not identical in theory and the actual performance is signal-dependent, we have seen that an iterated soft Haar wavelet shrinkage with a moderate number of scales can be used as a computationally efficient approximation to TV flow.

TV flow is only one example of nonlinear diffusion methods modeled by partial differential equation, and the method presented here is not necessarily the best of wavelet-based filters. In the future we plan to explore more connections between the two worlds and contribute to a fruitful exchange of ideas between them.

## Acknowledgement

This research was partly funded by the project *Relations between nonlinear filters in digital image processing* within the DFG–Schwerpunktprogramm 1114: *Mathematical methods for time series analysis and digital image processing*. This is gratefully acknowledged.

## References

- [1] F. Andreu, C. Ballester, V. Caselles, and J. M. Mazón. Minimizing total variation flow. *Differential and Integral Equations*, 14(3):321–360, Mar. 2001.
- [2] F. Andreu, V. Caselles, J. I. Diaz, and J. M. Mazón. Qualitative properties of the total variation flow. *Journal of Functional Analysis*, 188(2):516–547, 2002.
- [3] A. Chambolle and B. J. Lucier. Interpreting translation-invariant wavelet shrinkage as a new image smoothing scale space. *IEEE Transactions on Image Processing*, 10:993–1000, 2001.
- [4] R. R. Coifman and D. L. Donoho. Translation-invariant de-noising. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, volume 103 of *Lecture Notes in Statistics*, pages 125–150. Springer-Verlag, New York, 1995.
- [5] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. on Information Theory*, 41(3):613–627, 1995.
- [6] D. L. Donoho and I. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.
- [7] P. Dutilleul. An implementation of the "algorithme à trous" to compute the wavelet transform. In J. Combes, A. Grossman, and P. Tchamitchian, editors, *Wavelets: Time-Frequency Methods and Phase Space*, pages 298–304. Springer-Verlag, 1987.
- [8] 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. Combes, A. Grossman, and P. Tchamitchian, editors, *Wavelets: Time-Frequency Methods and Phase Space*, pages 286–297. Springer-Verlag, 1987.
- [9] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
- [10] 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.
- [11] 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, Series SPP-1114, Department of Mathematics, University of Bremen, Germany, 2003.