Relations between Soft Wavelet Shrinkage and Total Variation Denoising

Gabriele Steidl[1] and Joachim Weickert[2]

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

[2] Faculty of Mathematics and Computer Science,
Building 27.1, Saarland University, 66041 Saarbrücken, Germany,
weickert@mia.uni-saarland.de

Abstract. Soft wavelet shrinkage and total variation (TV) denoising are two frequently used techniques for denoising signals and images, while preserving their discontinuities. In this paper we show that – under specific circumstances – both methods are equivalent. First we prove that 1-D Haar wavelet shrinkage on a single scale is equivalent to a single step of TV diffusion or regularisation of two-pixel pairs. Afterwards we show that wavelet shrinkage on multiple scales can be regarded as a single step diffusion filtering or regularisation of the Laplacian pyramid of the signal.

1 Introduction

Image denoising is a field where one is frequently interested in removing noise without sacrificing important structures such as edges. Since this is not possible with linear techniques many nonlinear strategies have been proposed in the last two decades. Two of these classes are wavelet methods [4,6,9] and techniques based on partial differential equations (PDEs) [10,11,14].

Although both classes serve the same purpose, not many results are available where their similarities and differences are juxtaposed and their mutual relations are analysed. However, such an analysis is highly desirable, since it will help to transfer results from one of these classes to the others. Moreover, a deeper understanding of the differences between these classes might be helpful for designing novel hybrid methods that combine the advantages of the different classes.

The goal of the present paper is to address this problem by analysing relations between two of the most popular wavelet and PDE based methods: soft wavelet shrinkage [6] and total variation (TV) denoising [11] in its formulation as a diffusion flow or a regularisation process. Figure 1 gives an illustration of the denoising properties of wavelet and TV methods. We observe that the results do not differ very much. Indeed, we shall prove in our paper that both methods are very closely related. In order to keep things as simple as possible we base our analysis on the 1-D case and consider only Haar wavelets. Generalisations and extensions will be considered in forthcoming publications.

L. Van Gool (Ed.): DAGM 2002, LNCS 2449, pp. 198–205, 2002.

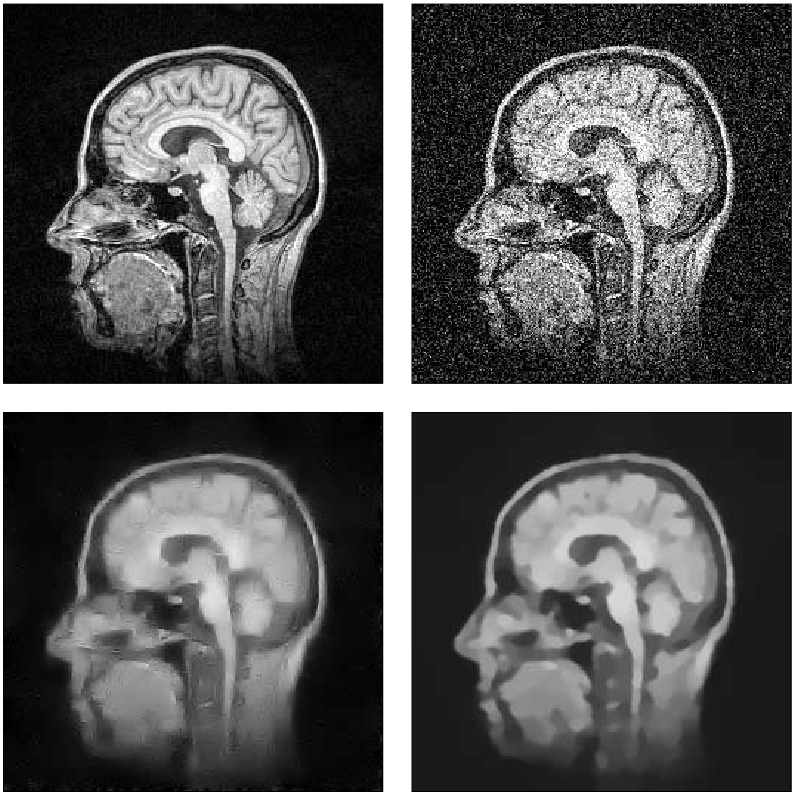

Fig. 1. (a) TOP LEFT: Original MR image. (b) TOP RIGHT: MR image degraded with Gaussian noise with standard deviation 50. (c) BOTTOM LEFT: Wavelet denoising of (b) using translation invariant soft shrinkage with Haar wavelets. (d) BOTTOM RIGHT: Total variation diffusion of (b).

Our paper is organised as follows. In Section 2 we prove a fundamental relation between soft wavelet shrinkage, nonlinear diffusion with TV diffusivity, and TV regularisation. While Section 2 is concerned with a single wavelet shrinkage step, Section 3 deals with the multiscale approach. Here we show the equivalence between multiscale wavelet shrinkage and TV diffusion / regularisation on a Laplacian pyramid. The paper is concluded with a summary in Section 4.

Related Work. Although we are not aware of any method in the literature that investigates the relations between discrete wavelet shrinkage and TV denoising, there are some interesting related techniques that should be mentioned in this context. Chambolle et al. [4] showed that one may interpret continuous wavelet shrinkage as regularisation processes in suitable Besov spaces. Durand and Froment [7] 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 an approach by Chan and Zhou [5] who postprocessed images obtained from wavelet shrinkage by a TV-like regularisation technique. Recently, Malgouyres [8] proposed a hybrid method that uses both wavelet packets and TV approaches. His experiments showed that it may restore textured regions without introducing ringing artifacts.

2 Soft Thresholding, TV Diffusion and TV Regularization

2.1 Soft Thresholding

We start by recalling a single Haar wavelet shrinkage step. Let $f = (f_i)_{i=0}^{N-1}$ be our initial signal, where $N = 2^n$. Then the *analysis step* produces the coefficients

$$c_i = \frac{f_{2i} + f_{2i+1}}{\sqrt{2}}, \quad d_i = \frac{f_{2i} - f_{2i+1}}{\sqrt{2}} \qquad (i = 0, ..., N/2 - 1)$$

of the scaling functions and the wavelets on the next coarser grid. This step is followed by the *shrinkage operation* $S_\tau(d_i)$, where S_τ denotes in general a nonlinear function which depends on a threshold parameter τ. In this paper we are interested in the *soft thresholding*

$$S_\tau(\eta) = \begin{cases} \eta - \tau \operatorname{sgn} \eta & \text{if} \quad |\eta| \geq \tau, \\ 0 & \text{if} \quad |\eta| < \tau. \end{cases} \tag{1}$$

Other shrinkage functions will be considered in a forthcoming paper. After the *synthesis step*

$$\begin{aligned} u_{2i} &= \frac{c_i + S_\tau(d_i)}{\sqrt{2}} &= \frac{f_{2i} + f_{2i+1}}{2} - \frac{1}{\sqrt{2}} S_\tau\left(\frac{f_{2i+1} - f_{2i}}{\sqrt{2}}\right), \\ u_{2i+1} &= \frac{c_i - S_\tau(d_i)}{\sqrt{2}} &= \frac{f_{2i} + f_{2i+1}}{2} + \frac{1}{\sqrt{2}} S_\tau\left(\frac{f_{2i+1} - f_{2i}}{\sqrt{2}}\right) \end{aligned} \tag{2}$$

we obtain a new signal u with smaller wavelet coefficients at the first decomposition level. The basic idea behind this procedure is that small wavelet coefficients mainly correspond to the noise contained in f while larger ones really signify basic features, e.g., edges, so that u can be considered as denoised version of f with preserved edges.

The Haar wavelet transform in (2) introduces a splitting of the signal f into successive two–pixel parts $(f_{2i}\ f_{2i+1})$ $(i = 0, ..., N/2)$. In the following we want to interpret a single wavelet shrinkage step as nonlinear diffusion of these successive two–pixel signals.

2.2 TV Diffusion

The basic idea behind nonlinear diffusion filtering is to obtain a family $u(x,t)$ of filtered versions of a signal $f(x)$ as the solution of a suitable diffusion process with $f(x)$ as initial condition [10]:

$$\begin{aligned} u_t &= (g(u_x)u_x)_x, \\ u(x,0) &= f(x) \end{aligned} \tag{3}$$

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

Motivated by the wavelet splitting in the Haar basis, we are interested in space-discrete diffusion of two–pixel signals (f_0, f_1). We do not allow any flow over the signal boundary, i.e. we deal with the extended sequence f_0, f_0, f_1, f_1 where the boundary values have been mirrored. For this simple setting a space-discrete version of the diffusion equation (3) in both pixels can be written as

$$\dot{u}_0 = g(u_1 - u_0)(u_1 - u_0), \quad \dot{u}_1 = -g(u_1 - u_0)(u_1 - u_0), \tag{4}$$

where $u_0(0) = f_0$, $u_1(0) = f_1$, and the pixel size is assumed to be 1. Setting $w(t) := u_1(t) - u_0(t)$ and $\eta := f_1 - f_0$, we obtain the initial value problem

$$\begin{aligned} \dot{w} &= -2\, g(w) w, \\ w(0) &= \eta \end{aligned}$$

by subtracting both equations in (4).

We are interested in the TV diffusivity $g(w) = 1/|w|$ since – unlike most other commonly used diffusivities – it does not require to specify additional contrast parameters. Moreover, it has a number of favourable qualitative properties [1,2]. By straightforward computation, the corresponding initial value problem

$$\begin{aligned} \dot{w} &= -2 \operatorname{sgn} w, \\ w(0) &= \eta. \end{aligned}$$

has the solution

$$w(t) = \begin{cases} \eta \; - \; 2\,t \operatorname{sgn} \eta & \text{if } t \leq |\eta|/2, \\ 0 & \text{if } t > |\eta|/2. \end{cases}$$

Since $\dot{u}_0 + \dot{u}_1 = 0$ and $u_0(0) + u_1(0) = f_0 + f_1$, we see further that the average grey value is preserved:

$$u_0(t) + u_1(t) = f_0 + f_1.$$

By the definition of w it follows that

$$\begin{aligned} u_i(t) &= \frac{f_0 + f_1}{2} \; - \; (-1)^i \, \frac{w(t)}{2} \qquad\qquad (i = 0, 1) \\ &= \frac{f_0 + f_1}{2} \; - \; (-1)^i \begin{cases} |\eta|/2 \; - \; t \operatorname{sgn} \eta & \text{if } t \leq |\eta|/2, \\ 0 & \text{if } t > |\eta|/2. \end{cases} \end{aligned}$$

By (2) and (1) this coincides with the Haar wavelet shrinkage with soft thresholding, where the threshold parameter τ is related to the diffusion time t by $\tau = \sqrt{2}\, t$.

2.3 TV Regularization

Nonlinear diffusion filtering of signals is related to variational methods for signal restoration [12]. Here the basic idea is to use the minimiser u of

$$E_f(u) := ||f - u||_{L_2}^2 \; + \; \alpha \int \varphi(u_x)\, dx \tag{5}$$

as denoised version of the initial signal $f(x)$. Via the Euler–Lagrange equation it follows that this minimiser coincides with the solution of

$$\frac{u-f}{\alpha} = (g(u_x)u_x)_x,$$

where $g(s) = \frac{\varphi'(s)}{2s}$. This can be considered as a time discretisation of the diffusion filter (3), where the regularization parameter α approximates the stopping time of the diffusion process [12].

Again we are only interested in the two–pixel model (f_0, f_1). We consider a space-discrete variant of (5), namely

$$E_f(u_1, u_2) = (f_0 - u_0)^2 + (f_1 - u_1)^2 + \alpha\,\varphi(u_1 - u_0), \tag{6}$$

for the TV potential function $\varphi(s) = 2|s|$ corresponding to the TV diffusivity $g(s) = 1/|s|$. Straightforward computation results in the following minimiser of (6)

$$\begin{aligned} u_i &= f_i \; + \; (-1)^i\,\alpha \qquad\qquad\qquad\qquad\qquad\qquad (i = 0,1) \\ &= \frac{f_0 + f_1}{2} \; - \; (-1)^i \begin{cases} |\eta|/2 - \alpha \operatorname{sgn} \eta & \text{if } \alpha \le |\eta|/2, \\ 0 & \text{if } \alpha > |\eta|/2. \end{cases} \end{aligned}$$

By (2) and (1) this coincides with a single Haar wavelet shrinkage step on (f_0, f_1) with soft threshold S_τ, where the threshold parameter τ is related to the regularisation parameter by $\tau = \sqrt{2}\,\alpha$.

In summary, the nonlinear diffusion with TV diffusivity and the variational method (6) with TV regularisation applied to the successive two–pixel parts (f_{2i}, f_{2i+1}) of f coincide with a single step of Haar wavelet shrinkage with soft thresholding. The threshold parameter τ is related to the diffusion time t and to the regularisation parameter α by

$$\tau = \sqrt{2}\,t = \sqrt{2}\,\alpha.$$

It is remarkable that TV diffusion and TV regularisation give identical evolutions in the two-pixel case, if one identifies the diffusion time t with the regularisation parameter α. From the considerations in [12] one would only expect that the processes approximate each other.

3 Multiscale Approach

So far we have only considered soft wavelet shrinkage on a single scale. In this section, we interpret *multiscale* soft shrinkage with Haar wavelets as application of nonlinear TV based diffusion to two–pixel groups of *hierarchical* signals.

Let us start with wavelet shrinkage again. Two steps of Haar wavelet shrinkage are described by the filter bank in Figure 2. As usual we apply the z-transform notation $f(z) = \sum_{i=0}^{N-1} f_i z^{-i}$. Then $\boxed{H_i(z)}$ $(i = 0,1)$ denotes the convolution of f with the low pass filter $(i = 0)$ and the high pass filter $(i = 1)$,

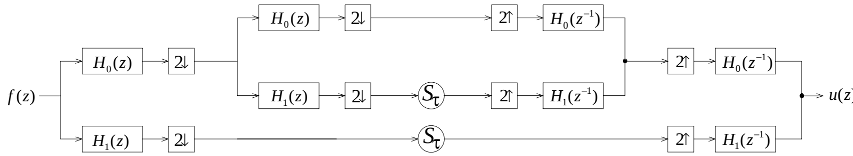

Fig. 2. Two steps of Haar wavelet shrinkage with $H_0(z) = \frac{1+z}{\sqrt{2}}$ and $H_1(z) = \frac{1-z}{\sqrt{2}}$.

i.e. $f(z)H_i(z)$, $\boxed{2\downarrow}$ and $\boxed{2\uparrow}$ downsampling and upsampling by 2, respectively, and the circle soft thresholding by S_τ. Finally $\bullet$ signifies addition; see also [13]. To obtain more scales we further split the upper branch of the inner filter bank cycle and so on.

We briefly recall the concept of *Gaussian* and *Laplacian pyramids* [3]. The Gaussian pyramid we are interested in is the sequence of H_0–smoothed and subsampled versions of an initial signal f given by

$$f = f^{(0)} \longrightarrow f^{(1)} \longrightarrow \ldots \longrightarrow f^{(n)},$$

where

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

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

$$Pf_{2i}^{(j)} = Pf_{2i+1}^{(j)} = f_i^{(j)}/\sqrt{2} \quad (j = 1, \ldots, n;\ i = 0, \ldots, N/2^j - 1). \tag{7}$$

Then the corresponding Laplacian pyramid is the sequence

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

By

$$f^{(j)} = Pf^{(j+1)} + \left(f^{(j)} - Pf^{(j+1)}\right) \qquad (j = n-1, \ldots, 0)$$

we can reconstruct f from its Laplacian pyramid.

Let diff_t denote the operator of nonlinear diffusion with TV diffusivity and stopping time t, applied to the successive two–pixel parts of a signal. By Subsection 2.2 we know that diff_t performs like a single wavelet shrinkage step with soft threshold parameter $\tau = \sqrt{2}t$. Further, we see that the upper branch and the lower branch of the filter bank in Figure 2 are given by $Pf^{(1)}$ and $\mathrm{diff}_t(f) - Pf^{(1)} = \mathrm{diff}_t(f - Pf^{(1)})$, respectively, where the later equation follows (although diff_t is a nonlinear operator) by (7) and (2). Thus, one wavelet shrinkage step is given by

$$u = Pf^{(1)} + \mathrm{diff}_t(f - Pf^{(1)}).$$

Now the multiscale Haar wavelet shrinkage up to scale n can be described by successive application of diff_t to the Laplacian pyramid:

$$\begin{aligned} u^{(n)} &= f^{(n)} \\ u^{(j)} &= Pu^{(j+1)} + \mathrm{diff}_t(f^{(j)} - Pf^{(j+1)}) \quad (j = n-1, \ldots, 0). \end{aligned} \tag{8}$$

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

4 Summary

In this paper we have seen that wavelet soft shrinkage on a single scale with Haar wavelets and threshold parameter τ is equivalent to TV-based nonlinear diffusion of two-pixel signal pairs with diffusion time $t = \tau/\sqrt{2}$. Moreover, it is also equivalent to TV regularisation of two-pixel pairs with regularisation parameter $\alpha = \tau/\sqrt{2}$. This might give rise to the conjecture that TV diffusion and regularisation yield identical results in general. Finally we showed that wavelet shrinkage on multiple scales is nothing but applying two-pixel TV diffusion or regularisation on the Laplacian pyramid of the signal.

These results are not only theoretically interesting, they may also have a number of practically relevant consequences. Firstly, they may help to make TV-based methods more popular for tasks such as image compression where wavelets constitute the state-of-the-art. Wavelet ideas may also help to make such PDE methods computationally more efficient. On the other hand, it is worth noting that PDE-based methods have no problems with translation and rotation invariance. Understanding their relation to wavelet methods might help to solve such well-known problems in the wavelet setting in a better way.

In our future work we intend to consider more advanced wavelet methods, to analyse the multidimensional case in detail, and to investigate possibilities to design hybrid methods that share the advantages of PDE-based techniques and wavelets.

Acknowledgements

Our joint research is partly funded by the project *Relations Between Nonlinear Filters in Digital Image Processing* within the the *DFG-Schwerpunktprogramm 1114: Mathematical Methods for Time Series Analysis and Digital Image Processing.* This is gratefully acknowledged.

References

1. F. Andreu, V. Caselles, J. I. Diaz, and J. M. Mazón. Qualitative properties of the total variation flow. *Journal of Functional Analysis*, 2002. To appear.
2. G. Bellettini, V. Caselles, and M. Novaga. The total variation flow in R^N. *Journal of Differential Equations*, 2002. To appear.
3. P. J. Burt and E. H. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31:532–540, 1983.
4. 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, Mar. 1998.
5. 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, Sept. 2000.
6. D. Donoho. De-noising by soft thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.

7. S. Durand and J. Froment. Reconstruction of wavelet coefficients using total-variation minimization. Technical Report 2001–18, Centre de Mathématiques et de Leurs Applications, ENS de Cachan, France, 2001.
8. 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.
9. S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, San Diego, 1998.
10. 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.
11. L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
12. O. Scherzer and J. Weickert. Relations between regularization and diffusion filtering. *Journal of Mathematical Imaging and Vision*, 12(1):43–63, Feb. 2000.
13. G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley–Cambridge Press, Wellesley, 1996.
14. J. Weickert. *Anisotropic Diffusion in Image Processing.* Teubner, Stuttgart, 1998.