Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational and Applied Mathematics 210 (2007) 222–231
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Edge-preserving wavelet thresholding for image denoising☆

D. Lazzaro*, L.B. Montefusco

Departement of Mathematics, University of Bologna, Italy

Received 19 July 2005; received in revised form 8 June 2006

Abstract

In this paper we consider a general setting for wavelet based image denoising methods. In fact, in both deterministic regularization methods and stochastic maximum a posteriori estimations, the denoised image $\hat{f}$ is obtained by minimizing a functional, which is the sum of a data fidelity term and a regularization term that enforces a roughness penalty on the solution. The latter is usually defined as a sum of potentials, which are functions of a derivative of the image. By considering particular families of dyadic wavelets, we propose the use of new potential functions, which allows us to preserve and restore important image features, such as edges and smooth regions, during the wavelet denoising process. Numerical results are presented, showing the optimal performance of the denoising algorithm obtained.

MSC: 65D; 65Y20; 65F

Keywords: Image denoising; Dyadic wavelets; Thresholding estimators; Potential functions

1. Introduction

We consider a classical problem of image processing: find an estimate $\hat{f}$ of an unknown function f, called the original image, from its noisy measurement $\bar{f}$, where $\bar{f} = f + e$ and e denotes an additive noise function. We want to improve the image quality by removing noise without sacrificing important image features, such as edges and homogeneous regions.

A number of approaches have been proposed to this aim, including stochastic and variational methods, non-linear diffusion filtering and wavelet techniques. The relations between some of the different approaches have also been investigated. In [3] it is proven that the variational formulation in the Besov space $B_1^1(L_1^1(I))$ leads to the classical Donoho and Johnstone wavelet shrinkage method [7], and in [1] the authors show that hard and soft thresholding wavelet estimators correspond to the lower and upper envelopes of a class of penalized least-squares estimators. The relations between anisotropic diffusion and robust statistics are analyzed in [2], where it is shown that the anisotropic diffusion equation is closely related to the error norm and influence function in the robust estimation framework. In [9] the authors consider the correspondence between wavelet shrinkage and non-linear diffusion and derive new wavelet shrinkage functions from existing diffusivity functions, while in [14,15] the relations between soft wavelet shrinkage and total variation denoising are analyzed and conditions for the equivalence of these methods are studied.

☆ This research was supported by MIUR, Cofin 2004 and R.F.O. projects.
* Corresponding author.
E-mail addresses: damiana.lazzaro@unibo.it (D. Lazzaro), laura.montefusco@unibo.it (L.B. Montefusco).

0377-0427/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2006.10.069

Motivated by the statistical interpretation of anisotropic diffusion, as well as the correspondence between statistical approaches and thresholding wavelet estimators, the purpose of this work is to develop new wavelet denoising methods that exploit the interplay between non-linear diffusion filtering and variational methods in the wavelet domain.

To this end, we consider a variational approach in a deterministic wavelet setting, consisting in the minimization of a functional that is the sum of two terms. The first one ensures that the estimated $\hat{f}$ is a faithful approximation of the original image, and the second represents an a priori constraint which enforces a roughness penalty on the estimate. In order to obtain edge-preserving denoising, the latter is defined as a sum of potentials, which are functions of a derivative of the image. Differently from [1], we make a particular choice of the wavelet transform, which is particularly well suited to maintaining important image features. In fact, we use the two-dimensional (2D) discrete dyadic transform introduced in [8], in which the two components of the wavelet transform of a function $f(x, y)$, at each scale, are proportional to the two components of the gradient of f smoothed at that scale.

This allows us to use for the penalty term the well-known potential functions of the non-linear diffusion filtering methods, hence obtaining new non-linear wavelet estimators that inherit the edge recovery properties of the chosen potential.

In spite of the non-linearity of the method, the minimization of the chosen functional can be realized quite efficiently. In fact, in the wavelet domain the original minimization problem uncouples into a family of 2D independent optimization problems. For the particular choice of the penalty function, these problems are non-quadratic but they can easily be solved by using a simple iterative approach. Making use of the results of [4], it is shown that the proposed iterative procedure turns out to be a simplified version of the half-quadratic regularization method known in the literature, from which it inherits all the convergence properties. The resulting denoising algorithm, consisting in a two-step iterative shrinkage procedure for each couple of dyadic wavelet coefficients, has been applied for the denoising of several test images corrupted by Gaussian noise. The results obtained are very encouraging as they are, both in terms of an objective estimate and from the point of view of the visual quality , among the best results in the existing literature.

The paper is organized as follows: in Sections 2 and 3 we briefly review the main ideas of non-linear diffusion filtering, and we define the discrete dyadic wavelet transform. The variational approach is presented in Section 4, while the characterization of the minimizer and the numerical algorithm are given in Section 5. Section 6 presents a representative set of results obtained for some classical test images.

2. Non-linear diffusion filtering

Diffusion algorithms remove noise from an image $\bar{f}$ by modifying the image via a partial differential equation. In the non-linear case, the method obtains a family $u(x, y, t)$ of filtered versions of $\bar{f}$ as the solution of a non-linear diffusion equation

$$\partial_t = \operatorname{div}[g(|\nabla u|)\nabla u], \quad u(x, y, 0) = \bar{f}(x, y),$$

where $|\nabla u|$ is the gradient magnitude and $g(|\nabla u|)$ is a diffusivity function, whose task is to reduce smoothing as the gradient becomes ever larger, as occurs near the object edges [16]. This allows for the preservation of their contrast and location. Typically g is a non-negative, non-increasing function of the gradient magnitude and its properties characterize the behavior of the corresponding non-linear filter.

Well known diffusivity functions are, for example,

$$g_{\mathrm{Ch}}(|\nabla u|) = \frac{\mu}{\sqrt{\mu^2 + |\nabla u|^2}}, \quad g_{\mathrm{PM}_1}(|\nabla u|) = \frac{\mu^2}{\mu^2 + |\nabla u|^2},$$

$$g_{\mathrm{PM}_2}(|\nabla u|) = \mu^2(1 - \mathrm{e}^{-c|\nabla u|^2/\mu^2})$$

due to Charbonnier [4] and Perona–Malik [12]. All these functions are continuously differentiable, bounded from above by 1, positive, monotonously decreasing and approaching zero for $|x| \to \infty$, but they produce different filtering results.

This is due to the different behavior of the corresponding potential functions [16]. In fact, for the Charbonnier diffusivity, the corresponding potential $\psi(x)$ is convex, while the Perona–Malik diffusivities correspond to potential functions $\psi(x)$ that are only convex for $|x| \leqslant \mu$. (See Fig. 1 for the 1D case.)

This means that the Perona–Malik model may sharpen edges, if their gradient is larger than the *contrast parameter* μ, while the Charbonnier model at most preserves edges by reducing diffusion in correspondence with large gradients.

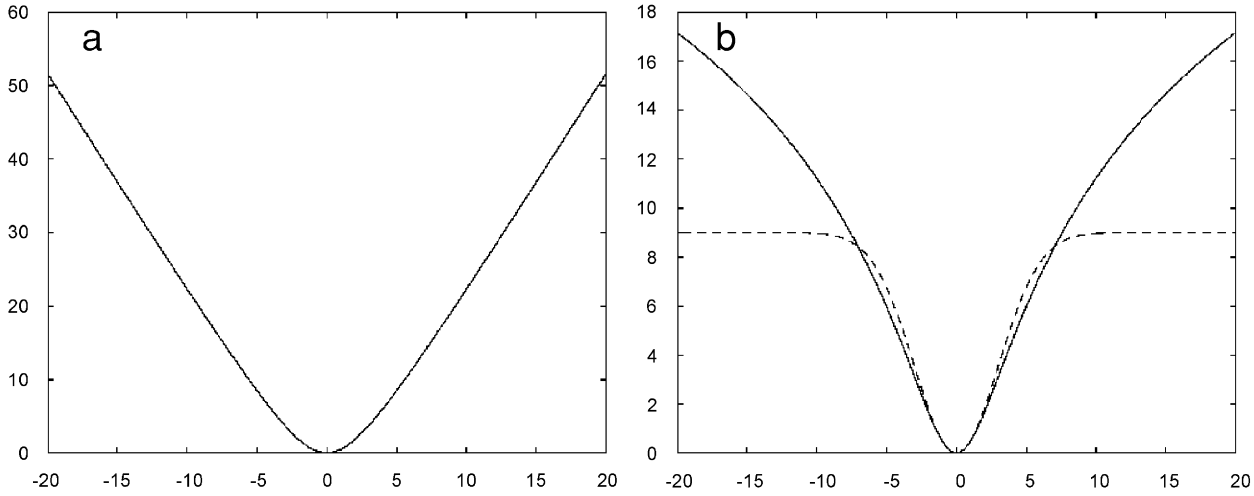

Fig. 1. (a) Charbonnier convex potential function $\mu = 3$; (b) Perona–Malik non-convex potential functions: PM_1, $\mu = 3$ solid, PM_2, $c = 0.5$, $\mu = 3$ (dashed).

A drawback of the non-linear diffusion model is that the discretization of the differential operator yields a large non-linear system that must be solved at each step, with a consequent high computational cost.

3. Dyadic wavelets and multiscale analysis

Edges in an image are located at its sharp variation points, namely, in correspondence with gradient magnitude maxima. They generally mark the contours of important image structures. In order to detect the contours of small structures, as well as the boundaries of larger regions, it is necessary to perform a *multiscale edge detection*. Most multiscale edge detectors smooth the signal at various scales and detect sharp variation points from their first or second order derivative. A simple and elegant instrument to perform multiscale edge detection is the discrete dyadic wavelet transform introduced in [8]. It is a redundant, translation invariant transform that, at each scale, yields two images whose elements are proportional to the coordinates of the gradient vector of the original image smoothed at that scale.

More precisely, given a two-dimensional smoothing function $\theta(x, y)$ (a Gaussian function, for example), two *mother wavelets* are defined as

$$\psi^1(x, y) = \frac{\partial\theta(x, y)}{\partial x}, \quad \psi^2(x, y) = \frac{\partial\theta(x, y)}{\partial y}. \tag{1}$$

Setting $\theta_s(x, y) = \frac{1}{s^2}\theta(x/s, y/s)$, $\psi^1_s(x, y) = \frac{1}{s^2}\psi^1(x/s, y/s)$, and $\psi^2_s(x, y) = \frac{1}{s^2}\psi^2(x/s, y/s)$, and imposing that the scale s varies only along the dyadic sequence $(2^j)_{j\in Z}$, we obtain that the *dyadic wavelet transform* of $f(x, y) \in L^2(\mathbb{R}^2)$ is defined by

$$\mathbf{W}f = \begin{pmatrix} f * \psi^1_{2^j}(x, y) \\ f * \psi^2_{2^j}(x, y) \end{pmatrix}_{j\in Z} = 2^j \begin{pmatrix} \dfrac{\partial}{\partial x}(f * \theta_{2^j})(x, y) \\ \dfrac{\partial}{\partial y}(f * \theta_{2^j})(x, y) \end{pmatrix} = 2^j \vec{\nabla}(f * \theta_{2^j})(x, y).$$

Its discrete version, the *discrete dyadic wavelet transform*, is easily derived by noting that an $N \times N$ image is measured at a finite resolution that represents a physical lower limit for the possible scales. By normalizing to 1 the finest scale and denoting with J the coarsest scale, $(J \leqslant \log(N))$, the finite discrete dyadic wavelet transform of an image f is given by:

$$\mathbf{W}^d f = \{[S_{2^J} f], ([W^1_{2^j} f])_{1\leqslant j\leqslant J}, ([W^2_{2^j} f])_{1\leqslant j\leqslant J}\},$$

Fig. 2. Dyadic wavelet decomposition: gradient magnitude for $j = 1, 2, 3$.

where $[S_{2^J} f]$ represents the image smoothed at scale 2^J and, at each scale j, the values of $[W^1_{2^j} f]$ and $[W^2_{2^j} f]$ are proportional to the coordinates of the gradient vector of $(f * \theta_{2^j})$. In [8] a fast algorithm is given for evaluating the discrete dyadic wavelet transform of an $N \times N$ image with a numerical complexity of $N^2 \log N$ (see [8] for details) (Fig. 2).

4. The variational approach to denoising

Wavelets are a well-established powerful tool in the area of signal and image processing, especially due to their sparseness and decorrelation properties.

In the denoising context, the decorrelation property of the discrete wavelet transform suggests processing the coefficients independently of each other, while the sparseness property leads to the use of thresholding and shrinkage methods aimed at removing, or attenuating, those coefficients that are small compared with the noise level.

The classical thresholding methods proposed by Donoho and Johnstone in [7] are the well-known hard and soft thresholding rules

$$S^{\text{hard}}_\tau(x) = x \cdot I(|x| > \tau), \quad S^{\text{soft}}_\tau(x) = \text{sign}(x) \cdot (|x|) - \tau)_+,$$

where τ is the thresholding parameter.

Even if developed in a statistical context, these rules can also be deduced from a variational approach. In [3,1,6], the authors consider the denoising problem, both from a deterministic and a statistical point of view, obtaining a similar minimization problem. More precisely, let $\mathscr{W}$ be a given discrete wavelet transform and $\bar{c} = \mathscr{W}\bar{f}$, $\hat{d} = \mathscr{W}\hat{f}$ be the wavelet transform of the data $\bar{f}$ and of the estimate $\hat{f}$, respectively. The variational approach to the denoising problem, formulated in the wavelet domain, consists in finding $\hat{d}$ as the minimizer of a penalized least-squares functional $F_p(d)$

$$\hat{d} = \arg\min F_p(d) = \arg\min(\|\bar{c} - d\|_2^2 + \lambda p(d)), \tag{2}$$

where $p(\cdot)$ is a given penalty function, which enforces a roughness penalty on the estimate, and the positive parameter λ balances the effect of the data fidelity and the penalization terms.

When the penalty $p(d)$ is chosen to be additive, i.e., $p(d) = \sum_{j,k} p(|d_{j,k}|)$, the minimization problem becomes separable. Minimizing $F_p(d)$ is equivalent to minimizing, for each coordinate j, k, the 1D functional

$$F_p(d_{j,k}) = |\bar{c}_{j,k} - d_{j,k}|^2 + \lambda p(|d_{j,k}|). \tag{3}$$

The choice of the penalty function strongly influences the behavior of the minimizer. The l_1 penalty function, i.e., $p(|d_{j,k}|) = |d_{j,k}|$, leads to the soft thresholding rule, while the l_2 penalty $p(|d_{j,k}|) = |d_{j,k}|^2$ usually produces an oversmooth solution. Other possible choices, better suited to denoising images with sharp features, are analyzed in [1] in a statistical wavelet setting.

The following general result states that the minimizer of (3) is either zero or a shrinked version of the noisy coefficient.

Theorem 4.1. *Let λ be a positive parameter and $\psi : \mathbb{R} \to \mathbb{R}$ satisfy*:

(I) $\psi(s) = \psi(-s)$, ψ *is continuously differentiable in* $(0, +\infty)$ *and* $\psi'(s) \geqslant 0\ \forall s > 0$.
(II) *The function* $s + \lambda\psi'(s)$ *is strictly unimodal on* $(0, +\infty)$. *Then the solution of the minimization problem*

$$\hat{s} = \arg\ \min(|t - s|^2 + \lambda\psi(|s|)) \tag{4}$$

exists and satisfies

$$\hat{s} = (|t| - 2^{-1}\lambda\psi'(|\hat{s}|))\,\mathrm{sign}(t) \quad \text{if} \quad |t| > \varepsilon_0; \quad \hat{s} = 0 \quad \text{if} \quad |t| \leqslant \varepsilon_0,$$

where $\varepsilon_0 = \min_{s \geqslant 0}(s + 2^{-1}\lambda\psi'(|s|))$. *Moreover, if* ψ *is convex in* $[0, +\infty]$, *the minimizer is unique*.

On the contrary, if the potential function ψ is non-convex, the minimizer is not unique, but, as shown in [10,11], it may have better edge detection properties.

5. Edge preserving wavelet thresholding estimators

In this section we use the previous results to develop a new family of wavelet denoising methods, that possess good edge preserving capabilities.

Since important edges are characterized by high gradient magnitude, an efficient edge preserving denoising method must reduce shrinkage at points where the magnitude of the gradients exceeds certain thresholds, while shrinking coefficients corresponding to small values of the gradient, that are probably due to noise. On the other hand, the dyadic wavelet transform of an image allows us to simply evaluate, at each scale, the magnitude of the gradient of the original image smoothed at that scale, since, at each pixel, it is proportional to

$$M_{jk} = \sqrt{|W^1_{2^j} f(k)|^2 + |W^2_{2^j} f(k)|^2} \qquad \text{with } k = 1, \ldots, N^2.$$

Motivated by this observation, our proposal is to consider the variational approach (2) in the dyadic wavelet domain, and use as penalty term the following expression:

$$p(d) = \sum_{j,k} \psi_j(M_{jk})$$

with ψ_j chosen to be one of the most used potential functions of the non-linear diffusion filtering methods, as, for example, those shown in Fig. 1, i.e.,

$$\psi_{j,\mathrm{Ch}}(M_{jk}) = \mu_j \left(\sqrt{\mu_j^2 + M_{jk}^2} - \mu_j\right), \tag{5}$$

$$\psi_{j,\mathrm{PM}_1}(M_{jk}) = \frac{\mu_j^2}{2} \log\left(1 + \left(\frac{M_{jk}}{\mu_j}\right)^2\right), \tag{6}$$

$$\psi_{j,\mathrm{PM}_2}(M_{jk}) = \mu_j^2 \left(1 - \mathrm{e}^{-c(M_{jk}/\mu_j)^2}\right), \quad c > 0, \tag{7}$$

where the parameter μ_j plays the role of a scale-dependent contrast parameter, that, for each scale, represents the threshold between *edge* and *not-an-edge*.

With this choice the minimization problem decouples into a family of two-dimensional minimizations, that is

$$\begin{cases} \hat{d}^1_{jk} = \arg\ \min F^1(d^1_{jk}, d^2_{jk}) = \arg\ \min(|\bar{c}^1_{jk} - d^1_{jk}|^2 + \lambda_j\psi_j(M_{jk})), \\ \hat{d}^2_{jk} = \arg\ \min F^2(d^1_{jk}, d^2_{jk}) = \arg\ \min(|\bar{c}^2_{jk} - d^2_{jk}|^2 + \lambda_j\psi_j(M_{jk})), \end{cases} \tag{8}$$

where

$$\hat{d}^r_{jk} = W^r_{2^j}\hat{f}(k), \quad \bar{c}^r_{jk} = W^r_{2^j}\bar{f}(k), \quad r = 1, 2 \quad \text{and} \quad M_{jk} = \sqrt{|d^1_{jk}|^2 + |d^2_{jk}|^2}.$$

In (8) also the parameter λ_j is considered scale-dependent in order to let both the form and the weight of the penalty term change according to the scale.

Since all the above potential functions satisfy the assumptions of Theorem 4.1, and have $\varepsilon_0 = 0$, the denoising algorithm consists in finding, for each pair of dyadic wavelet coefficients and at each scale, the solution of the following couple of non-linear Eulero–Lagrange equations

$$\begin{cases} |\hat{d}^1_{jk}| + \frac{\lambda_j}{2}\psi'_j(\hat{M}_{jk}) = |\bar{c}^1_{jk}|, \\ |\hat{d}^2_{jk}| + \frac{\lambda_j}{2}\psi'_j(\hat{M}_{jk}) = |\bar{c}^2_{jk}|. \end{cases} \tag{9}$$

By omitting, for the sake of simplicity, the subscript j, k and using the auxiliary variables

$$\hat{y}^1 = \frac{1}{2}\frac{\psi'(\hat{M})}{|d^1|}, \quad \hat{y}^2 = \frac{1}{2}\frac{\psi'(\hat{M})}{|d^2|}$$

the above non-linear system can be written as

$$\begin{aligned} |\hat{d}^1| &= |\bar{c}^1|/(1 + \lambda\hat{y}^1), \\ |\hat{d}^2| &= |\bar{c}^2|/(1 + \lambda\hat{y}^2). \end{aligned} \tag{10}$$

In order to approximate its solution, we propose an iterative approach that, at each scale j and for each position k, generates two sequences $(d^1)_n$ and $(d^2)_n$ according to the following scheme:

Given two positive initial values d^1_0 and d^2_0,
  for $n = 0, 1, \ldots$
    $M_n = \sqrt{(d^1_n)^2 + (d^2_n)^2}$
    for r = 1,2

$$y^r_{n+1} = \psi'(M_n)/(2d^r_n) \tag{11}$$

$$d^r_{n+1} = |\bar{c}^r|/(1 + \lambda y^r_{n+1}) \tag{12}$$

    untila convergence criterion is satisfied for $n = \bar{n}_q$.

Then, according to Theorem 4.1,

$$\hat{d}^r_{jk} = (\hat{d}^r_{jk})_{\bar{n}} \cdot \operatorname{sign}(\bar{c}^r_{jk}), \quad r = 1, 2.$$

The following proposition gives a rigorous justification of the proposed iterative strategy, by inserting it into the context of half quadratic minimization.

Proposition 5.1. *Let the potential function $\psi(s)$ satisfy*:

(1) $\psi(s) \geqslant 0\ \forall s$ *with* $\psi(0) = 0$, $\psi(s) = \psi(-s)$, ψ *is continuously differentiable in* $(0, +\infty)$ *and* $\psi'(s) \geqslant 0\ \forall s \geqslant 0$;
(2) $\frac{\psi'(s)}{2s}$ *is continuous and strictly decreasing in* $[0, +\infty)$, $\lim_{s\to 0^+}\frac{\psi'(s)}{2s} = M$, $0 < M < \infty$, *and* $\lim_{s\to\infty}\frac{\psi'(s)}{2s} = 0$;
(3) $\psi'''(s) = 0$ *and* $\psi^{\mathrm{iv}}(0)$ *exists*.

Then

(i) *there exists a strictly convex and decreasing function θ defined on* $(0, M)$, *such that*

$$\psi(s) = \inf_{0<y\leqslant M}(y\,s^2 + \theta(y)); \tag{13}$$

(ii) *there exist two functionals*

$$F_*^r(d^1, d^2, y^r) = |\bar{c}^r - d^r|^2 + \frac{\lambda}{2}\left[y^r (d^r)^2 + \theta(y^r)\right], \quad r = 1, 2,$$

strictly convex in y^r for d^r fixed, and quadratic in d^r for y^r fixed, for which

$$\begin{cases} y_{n+1}^r = \arg\ \min_{y^r}\ F_*^r(d_n^1, d_n^2, y^r), \\ d_{n+1}^r = \arg\ \min_{d^r}\ F_*^r(d^1, d^2, y_{n+1}^r), \end{cases} \quad r = 1, 2$$

and the minimum is reached for the values of y_{n+1}^r and d_{n+1}^r as given in (11) *and* (12), *respectively.*

Proof. The proof is straightforward if we make use of the results stated in [4]. In fact, in [4, Theorem 1], the existence of the function θ satisfying (13) is proved, under similar assumptions, in a more general context, and the expression of the unique minimum of F_*^r, for fixed values of d^1 and d^2, is shown to be

$$y^r = \frac{\psi'(M)}{2d^r}.$$

By fixing the value of y, the functional F_*^r becomes quadratic in d^r and its minimization yields immediately the value of d^r.

Concerning the convergence of the sequences generated by the iterative algorithm, we refer again to [4], where it is also shown that, if the potential $\psi(s)$ is convex, the convergence is to the unique minimum, while, for non-convex potential functions, the algorithm probably computes a local minimum.

It is worthwhile noting that the auxiliary variables y^r play an important role in our denoising method. In fact, they determine the amount of shrinkage to be applied to each dyadic wavelet coefficient, i.e., to the image gradient components. For large gradient magnitude the original coefficient remains almost unchanged, while, where the gradient magnitude is small, a strong reduction is applied.

The interesting fact, that explains the analogy between the present approach and non-linear diffusion filtering, is that the variables $y^r = \frac{\psi'(M)}{2d^r}$ coincide with the values of the diffusivity functions of the non-linear diffusion methods. In both cases they act by preserving large gradients, while reducing the small ones. □

6. Numerical results

In this section, we present one representative set of results illustrating the performance characteristics of the new wavelet thresholding estimators we have presented so far. For our experimentation we have used the three 512×512 grayscale images Lena, Boat, and Edges, shown in Fig. 3, corrupted by adding white Gaussian noise of standard deviation σ ranging from 15 to 30. The performance of the estimators was measured by using the classical peak signal

Fig. 3. Original test images.

Table 1
PSNR values for test image "Edges" with different noise levels

Egdes	Noisy	SureShrink	PM_1	PM_2	Ch
$\sigma = 15$	24.65	30.13	36.73	**36.90**	34.72
$\sigma = 20$	22.14	27.70	33.84	**34.00**	32.31
$\sigma = 25$	20.17	25.57	31.67	**31.75**	30.54
$\sigma = 30$	18.62	24.21	29.99	**30.07**	29.12

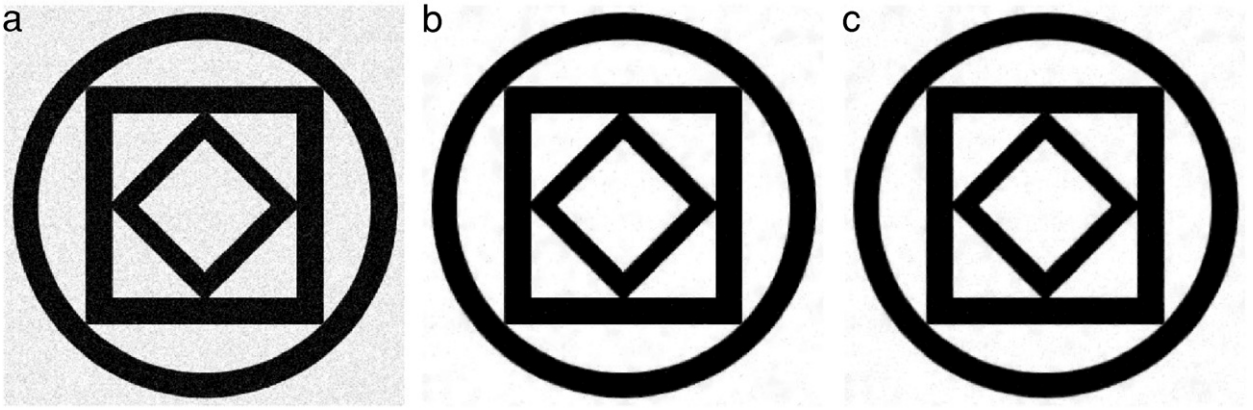

Fig. 4. Denoising results obtained with: (b) the non-convex potential function PM_2; (c) the convex potential function Ch: (a) Noisy image $\sigma = 30$; (b) PSNR = 30.07; (c) PSNR = 29.12.

to noise ratio, defined as

$$\mathrm{PSNR} = 20 \log_{10} \frac{255}{\mathrm{RMSE}} \quad \text{with } \mathrm{RMSE} = \sqrt{\frac{\sum_i \sum_j (f_{ij} - \hat{f}_{ij})^2)}{N^2}}.$$

The results of the new method are obtained by choosing as mother wavelets of the diadic wavelet decomposition two functions ψ^1 and ψ^2 defined as the partial derivatives of a 2D cubic spline function. This choice is proposed in [8], where the filters for the fast algorithm are also given. We have considered three decomposition steps and used both the non-convex Perona–Malik potential functions (6) and (7), denoted by PM_1 and PM_2, respectively, as well as the convex Charbonnier potential (5), denoted by Ch. In all cases the scale dependent contrast parameter μ_j, $j = 1, \ldots, 3$ was automatically estimated, by means of a classical tool from the "robust statistics" [2], according to

$$\mu_j = 1.4826 \cdot \left(\mathrm{MAD}\left(\left[\left(\bar{c}^1_{jk} \right)^2 \right]_{k=1}^{N^2} \right) + \mathrm{MAD}\left(\left[\left(\bar{c}^2_{jk} \right)^2 \right]_{k=1}^{N^2} \right) \right),$$

where MAD denotes the median absolute deviation. On the contrary, the values of the parameter λ_j, balancing, for each decomposition level, the tradeoff between the data fidelity and the penalization terms, were chosen manually, according to the image characteristics. Typically its value decreases as j increases and it depends on the noise level.

The results presented are obtained making approximatively five iterations of the proposed algorithm, since after few steps the corresponding PSNR value remains stationary.

In the first set of experiments we have used the geometric image "Edges" in order to highlight the edge preservation properties of the new wavelet thresholding estimators. The values of the PSNR of the denoised images, corresponding to different denoising levels, are given in Table 1, where we also present, for comparison purposes, the results obtained using the SureShrink method with the undecimated orthonormal Daubechies wavelets (4 vanishing moments and 4 decomposition steps).

Table 2
PSNR values for test images Lena and Boat perturbed with different noise levels and denoised with: VisuShrink (VS), SureShrink (SS), HMT system, Bivariate shrinkage function (BS), and our new thresholding estimators

Test images	Methods in the literature					Proposed methods		
	Noisy	VS	SS	HMT	BS	PM_1	PM_2	Ch
Lena								
$\sigma = 15$	24.65	27.48	31.59	31.76	**32.06**	31.66	31.64	30.85
$\sigma = 20$	22.14	26.46	30.22	30.39	**30.73**	30.56	30.54	30.17
$\sigma = 25$	20.17	25.67	29.14	29.24	29.81	30.01	**30.04**	29.90
$\sigma = 30$	18.62	25.14	28.38	28.35	28.94	29.51	**29.51**	29.44
Boat								
$\sigma = 15$	24.65	25.34	29.48	30.31	**30.25**	29.61	30.04	28.46
$\sigma = 20$	22.15	24.43	28.14	28.84	28.93	28.80	**28.96**	28.11
$\sigma = 25$	20.15	23.76	27.15	27.68	27.91	28.11	**28.17**	27.75
$\sigma = 30$	18.62	23.33	26.52	26.83	27.11	27.50	**27.55**	27.40

Fig. 5. Denoising results obtained with: (b), (e) the non-convex potential function PM_2; (c), (f) the convex potential function Ch: (a) Noisy image $\sigma = 30$; (b) PSNR = 29.51; (c) PSNR = 29.44; (d) Noisy image $\sigma = 30$; (e) PSNR = 27.55; (f) PSNR = 27.40.

As expected, the thresholding estimators corresponding to the non-convex potential functions PM_1 and PM_2 yield better results than the convex one, and all the proposed methods outperform the traditional techniques, both in terms of PSNR values and visual quality (see Fig. 4).

We have, then, compared the results obtained for the Lena and Boat images, with those presented in [13] obtained with the VisuShrink, the SureShrink, the hidden Markov tree (HMT) model [5], and a new bivariate shrinkage function proposed in [13]. The results are given in Table 2.

We note that the results obtained for the higher noise levels are particularly impressive and outperform those obtained with the others methods in literature. In fact, also for this severe perturbation, the proposed estimators (especially using the non-convex potentials) succeed in eliminating the noise component, while preserving the important image features, such as edges and homogeneous regions (Fig. 5).

References

[1] A. Antoniadis, J. Fan, Regularization of wavelet approximations, J. Amer. Stat. Assoc. 961 (455) (2001) 939–967.
[2] M.J. Black, G. Sapiro, D.H. Marimont, D. Heeger, Robust anisotropic diffusion, IEEE Trans. Image Process. 7 (3) (1998) 421–432.
[3] A. Chambolle, R.A. DeVore, N. Lee, B.L. Lucier, Nonlinear wavelet image processing: variational problems compression and noise removal through wavelet shrinkage, IEEE Trans. Image Process. 7 (3) (1998) 319–335.
[4] P. Charbonnier, L. Blanc-Feraud, G. Aubert, M. Barlaud, Deterministic edge-preserving regularization in computed imaging, IEEE Trans. Image Process. 6 (2) (1997) 298–311.
[5] H. Choi, J.K. Romberg, R.G. Baraniuk, N.G. Kingsbury, Hidden Markov tree modeling of complex wavelet transform, in: Proc. ICASSP, vol. 1, Instanbul, Turkey, June 2000, pp. 133–136.
[6] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure Appl. Math. LVII (2004) 1413–1457.
[7] D. Donoho, I. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (1994) 425–455.
[8] S. Mallat, S. Zhong, Characterization of signals form multiscale edges, IEEE Trans. Pattern Anal. Mach. Intelligence 14 (7) (1992) 710–732.
[9] P. Mrázek, J. Weickert, G. Steidl, Correspondences between wavelet shrinkage and nonlinear diffusion, in: L.D. Griffin, M. Lillholm (Eds.), Scale-Space 2003, Lecture Notes in Computer Science 2695, Springer, Berlin, Heidelberg, 2003, pp. 101–116.
[10] M. Nikolova, Regularization functions and estimators, in: Proceedings of the IEEE International Conference on Image Processing, Lausanne, Switzerland, September 1996, pp. 457–460.
[11] M. Nikolova, Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares, Report CMLS no. 2004–17.
[12] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Mach. Intelligence 12 (7) (1990) 629–639.
[13] L. Sendur, I.W. Selesnick, A bivariate shrinkage functions for wavelet based denoising, IEEE Trans. Signal Process. 50 (2002) 2744–2756.
[14] G. Steidl, J. Weickert, Relations between soft wavelet shrinkage and total variation denoising, in: L. Van Gool (Ed.), Pattern Recognition, Lecture Notes in Computer Science, vol. 2449, Springer, Berlin, 2002, pp. 198–205.
[15] G. Steidl, J. Weickert, T. Brox, P. Mrázek, M. Welk, On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs, Technical Report, Series SPP-1114, Department of Mathematics, University of Bremen, Germany, 2003.
[16] J. Weickert, A review of nonlinear diffusion filtering, in: B. ter Haar Romeny, L. Florack, J. Koenderink, M. Viergever (Eds.), Scale-Space Theory in Computer Vision, Lecture Notes in Computer Science, vol. 1252, Springer, Berlin, 1997, pp. 3–28.