Restoration of wavelet coefficients by minimizing a specially designed objective function

Durand Sylvain

LAMFA UMR 6140–Université de Picardie
33 rue Saint-Leu, 90039 Amien Cedex
CMLA UMR 8536–ENS de Cachan
61 av. President Wilson, 94235 Cachan Cedex
e-mail: sdurand@cmla.ens-cachan.fr

Mila Nikolova

CNRS URA820–ENST Dpt. TSI
46 rue Barrault, 75634 Paris Cedex 13
CMLA UMR 8536–ENS de Cachan
61 av. President Wilson, 94235 Cachan Cedex
e-mail: nikolova@tsi.enst.fr

Abstract

We consider the denoising of a function (an image or a signal) containing smooth regions and edges. Classical ways to solve this problem are variational methods and shrinkage of a representation of the data in a basis or in a frame. We propose a method which combines the advantages of both approaches. Following the wavelets shrinkage method of Donoho and Johnstone, we set to zero all frame coefficients which are smaller than a threshold. Then the frame representation involves both large coefficients corresponding to noise (outliers) and some coefficients, erroneously set to zero, leading to Gibbs oscillations in the estimate. We design a specialized (non-smooth) objective function allowing all these coefficients to be selectively restored, without modifying the other coefficients. We also propose an approximation of this method which is accurate enough and very fast. We present numerical experiments with signals and images corrupted with white Gaussian noise, which are decomposed into an orthogonal basis of Daubechies wavelets. The obtained results demonstrate the advantages of our approach over many alternative methods.

1 Introduction

We consider the restoration of an original (unknown) function $u_o(s)$ defined on a (possibly finite) domain Ω—an image or a signal containing smooth zones and edges—from noisy data $v = u_o + n$, where n represents a perturbation. In variational methods, the restored function is defined as the minimizer of an objective function which balances trade-off between closeness to data and a priori smoothness constraints, see e.g. [14, 8, 2]

$$\mathcal{F}_v(u) = \lambda \int_\Omega |u(s) - v(s)|^2 ds + \int_\Omega \phi(|\nabla u(s)|)\, ds, \quad (1)$$

where ∇ stands for gradient, $\phi : \mathrm{R}_+ \to \mathrm{R}_+$ is called a potential function and $\lambda > 0$ is a parameter.

Another family of methods is shrinkage estimators. Consider a frame of $L^2(\Omega)$, say $\{\widetilde{w}_i\}$, for i belonging to an index set I. The corresponding frame operator $\widetilde{W}$ is defined by $(\widetilde{W}v)[i] = \langle v, \widetilde{w}_i \rangle$, $\forall i \in I$. The frame coefficients of our noisy function v are

$$y = \widetilde{W}v. \quad (2)$$

Let W be a left inverse of $\widetilde{W}$ and let $\{w_i\}$ denote the associated dual frame. Therefore,

$$v = W\widetilde{W}v = \sum_{i \in I} y[i]\, w_i.$$

Given a symmetric function $\tau : \mathrm{R} \to \mathrm{R}$, such that $0 \leq \tau(t) \leq t$ if $t \geq 0$, the denoised function is defined as

$$v_\tau = \sum_{i \in I} \tau(y[i])\, w_i. \quad (3)$$

The hard-thresholding method, introduced by Donoho and Johnstone in [10] when $\widetilde{W}$ is an orthogonal wavelet transform, corresponds to

$$\tau(t) = \begin{cases} 0 & \text{if } |t| \leq T, \\ t & \text{otherwise,} \end{cases} \quad (4)$$

where $T > 0$ is a threshold. Soft-thresholding, introduced also in [10], corresponds to $\tau(t) = 0$ if $|t| \leq T$ and $\tau(t) = t - T\mathrm{sign}(t)$ otherwise. Both soft and hard thresholding are asymptotically optimal in the minimax sense if Ω is of finite cardinality, n is white Gaussian noise of standard deviation σ and

$$T = \sigma\sqrt{2\log_e \#\Omega}. \quad (5)$$

Refinements of these methods have been proposed in order to adapt thresholding to the scale of the coefficients [11]. On the other hand, many other shrinkage functions τ have been derived by considering the denoising of the frame coefficients as a maximum a posteriori estimation [20, 3, 1]. Then the restored coefficients minimize a cost-function similar to (1), namely

$$F(x) = \|y - x\|^2 + \sum_i \lambda_i \phi(|x_i|),$$

where ϕ corresponds to the priors on $\{x_i\}$ and $\{\lambda_i\}$ are weights related to the scale. Notice that $\phi' = \tau$ at every t where ϕ is differentiable. An useful class of priors is given by $\phi(t) = |t|^\alpha$ for $0 < \alpha \leq 2$. The relationship between shrinkage methods for wavelets and variational methods of the form (1) has also been investigated in [7].

A major problem with these methods is that shrinked coefficients entail oversmoothing of edges, coefficients set to zero yield Gibbs oscillations in the vicinity of edges, while coefficients which remain corrupted (called *outliers*) generate artifacts with the shape of the functions w_i of the frame. Even if ϕ accounts faithfully for the distribution of the coefficients x_i, the local features of the restored function, such as the presence of edges and smooth regions, are not properly addressed. Hence the idea to combine the useful information contained in the large coefficients $y[i]$ with pertinent prior smoothness constraints [12, 15, 6, 9].

We focus on post-processing noisy coefficients y_T obtained by hard thresholding,

$$y_T[i] = \tau(y[i]), \tag{6}$$

where τ is the function given in (4) for a loosely chosen T, smaller than (5) in order to have a richer information content. In such a case, the main problems to deal with are (a) to selectively shrink the outliers and (b) to restore the thresholded coefficients yielding Gibbs oscillations. Notice that, although these two phenomena are different by nature, they share the property to be local in the sense that they concern isolated zones of coefficients. In [17], a new family of objective functions is considered which allows outliers to be selectively removed, without modifying the other samples. Following this idea, we design a (non-smooth) objective function F_y specially adapted to deal with problems (a)-(b). We restrict our attention to *convex objective functions* in order to guarantee the uniqueness of the minimum. The effectiveness of our method is demonstrated by comparing our results with existing image de-noising schemes.

2 A specialized objective function

Our input data set is y_T, defined as in (6):

$$y_T[i] = \begin{cases} y[i] & \text{if} \quad i \in I_1, \\ 0 & \text{if} \quad i \in I_0, \end{cases} \tag{7}$$

where

$$I_0 = \{i \in I : \, |y[i]| < T\}$$

and $I_1 = I \setminus I_0$. The restored function, denoted $\hat{u}$, is defined as

$$\hat{u} = \sum_{i \in I} \hat{x}[i] \; w_i = Wx, \tag{8}$$

where $\hat{x}$ minimizes an objective function F_y of the form

$$\begin{aligned} F_y(x) &= \Psi(x) + \Phi(x), \\ \Psi(x) &= \sum_{i \in I_1} \psi_i \left(|x[i] - y[i]| \right) + \sum_{i \in I_0} \psi_i \left(|x[i]| \right), \\ \Phi(x) &= \int_\Omega \phi(|\nabla (Wx)(s)|) \, ds. \end{aligned} \tag{9}$$

Here Ψ is a data-fidelity term and Φ is a regularization term. Some general requirements are that $\{\psi_i\}$ and ϕ are $\mathcal{C}^1$, convex, increasing functions from $\mathbf{R}_+$ to $\mathbf{R}_+$. Below we discuss how the information in y_T must be considered. Recall that the original unknown function is denoted u_o.

- I_1 addresses two types of coefficients $y[i]$:
  - Large coefficients which bear the main features of the sought-after function. They verify $y[i] \approx (\widetilde{W} u_o)[i]$ and must be kept intact.
  - Outliers are characterized by $|y[i]| \gg |(\widetilde{W} u_o)[i]|$. They must be removed and replaced by values according to the priors conveyed by Φ.
- The coefficients $y[i]$ for $i \in I_0$ are usually high-frequency components which can be
  - Noise coefficients, if $(\widetilde{W} u_o)[i] \approx 0$. These coefficients must remain zero.
  - Coefficients which correspond to edges and which are erroneously set to zero. They generate Gibbs oscillations and need to be restored, based on the priors conveyed by Φ.

We will specify $\{\psi_i\}$ and ϕ in (9) in such a way that minimizers $\hat{x}$ of F_y achieve these goals. All requirements on $\hat{u}$ are formulated in terms of the minimizers $\hat{x}$ of F_y.

2.1 The regularization term Φ

This term brings the priors about the local features of the restored function. Its role is critical on the regions corresponding to wavelet coefficients which are either outliers or are erroneously set to zero. The images and signals we wish to restore are supposed to involve smooth regions and edges. To this end, we focus on *edge-preserving* convex potential functions ϕ which have been studied by many authors [18, 5, 14, 4, 8]. An essential distinction between these potential functions is the differentiability of $t \to \phi(|t|)$ at the origin. Since [16], it is known that if $t \to \phi(|t|)$ is non-smooth at zero, restored images and signals $W\hat{x}$ involve constant regions. Such a property does not correspond to real-world images and signals. In contrast, if Φ is smooth, they contain smoothly varying regions and possibly edges. We hence focus on potential functions ϕ of the latter kind, which means that $\phi'(0^+) = 0$. Examples of such functions

are [13, 5, 4, 8, 21]

$$\begin{aligned}\phi(t) &= \sqrt{\alpha+t^2}, \\ \phi(t) &= t^{(1+\alpha)},\ 0<\alpha\leq 1, \\ \phi(t) &= \log(\cosh(\alpha t)), \\ \phi(t) &= \begin{cases} t^2/2 & \text{if } |t|\leq\alpha, \\ \alpha|t|-\alpha^2/2 & \text{if } |t|>\alpha, \end{cases}\end{aligned} \tag{10}$$

where $\alpha > 0$ in order to have $\phi'(0^+) = 0$.

2.2 Conditions for a minimum

Given a vector x, its transposed will be denoted x^T. Consider F_y as given in (9) where ψ_1, ψ_0 and ϕ are $\mathcal{C}^1$, convex, increasing functions from R_+ to R_+ and $\phi'(0^+) = 0$. Then F_y reaches its minimum at $\hat{x}$ if and only if

$\forall i \in I_1$,

$$\hat{x}[i] = y[i] \Rightarrow |\partial_i\Phi(\hat{x})| \leq \psi_i'(0^+), \tag{11}$$

$$\hat{x}[i] \neq y[i] \Rightarrow \partial_i\Phi(\hat{x}) = -\psi_i'(|(\hat{x}-y)[i]|)\frac{(\hat{x}-y)[i]}{|(\hat{x}-y)[i]|}; \tag{12}$$

$\forall i \in I_0$,

$$\hat{x}[i] = 0 \Rightarrow |\partial_i\Phi(\hat{x})| \leq \psi_i'(0), \tag{13}$$

$$\hat{x}[i] \neq 0 \Rightarrow \partial_i\Phi(\hat{x}) = -\psi_i'(|\hat{x}[i]|)\frac{\hat{x}[i]}{|\hat{x}[i]|}, \tag{14}$$

where $\partial_i\Phi(x) = \partial\Phi/\partial x[i]\,(x)$ reads

$$\partial_i\Phi(x) = \int_\Omega \phi'(|\nabla W x|)\,(\nabla w_i)^T\,\frac{\nabla W x}{|\nabla W x|}\,ds.$$

Notice that Ψ is smooth only if $\psi_i'(0^+) = 0$, $\forall i$.

2.3 The data-fidelity term

The shape of Ψ will be determined in order to deal with two simple but revealing situations.

Wavelet-shaped artifacts. Let $e_k[k] = 1$ and $e_k[i] = 0$ if $i \neq k$, and $\mathbb{1}$ be a constant vector. Suppose that on a neighborhood of the index k, our input data y_T, obtained by (7), are of the form

$$y_T = \widetilde{W}\mathbb{1} + \delta e_k, \tag{15}$$

where $y[k] = \delta > 0$ is an outlier. The function denoised by hard-thresholding is

$$W y_T = W\widetilde{W}\mathbb{1} + \delta W e_k = \mathbb{1} + \delta w_k.$$

Clearly, it contains an artifact with the shape of w_k. Smoothing the outlier $y[k]$, without destroying the other coefficients, means that F_y is minimized by an $\hat{x}$ such that

$$\hat{x} = \widetilde{W}(\mathbb{1} + \hat{x}[k]\ w_k) \text{ with } \hat{x}[k] \approx 0, \tag{16}$$

even if δ is arbitrarily large. *Suppose henceforth that $\hat{x}$ is as given in (16).* Clearly, $k \in I_1$. By (8), our estimate is $\hat{u} = W\hat{x} = \mathbb{1} + \hat{x}[k]\ w_k \approx \mathbb{1}$.

1. Since $|\hat{x}[k]| \neq \delta$, then $\hat{x}[k]$ satisfies (12), hence

$$\partial_k\Phi(\hat{x}) = \psi_k'(|\hat{x}[k]\ - \delta|), \tag{17}$$

where we notice that $\hat{x}[k] - \delta < 0$ and

$$\partial_k\Phi(\hat{x}) = \int_\Omega \phi'(|\hat{x}[k]\ \nabla w_k|)\,|\nabla w_k|\ ds. \tag{18}$$

The outlier cannot be penalized unless $\psi_k'(t) \geq \psi_k'(0^+)$ for $t \in \mathrm{R}_+$. The right side of (17) is hence positive. The fact that $\phi'(t) = 0$ only if $t = 0$ in (18) shows that $\hat{x}[k] \neq 0$.

2. Since $\hat{x}[j] = 0$, for all $j \in I_1 \setminus \{k\}$, by (11) we have

$$\left|\int_\Omega \phi'(|\hat{x}[k]\ \nabla w_k|)\,(\nabla w_j)^T\,\frac{\nabla w_k}{|\nabla w_k|}\right| ds \leq \psi_k'(0^+).$$

Since $\hat{x}[k] \neq 0$, the left-side in the expression above is strictly positive. This inequality cannot be satisfied unless ψ_k is such that

$$\psi_k'(0^+) > 0. \tag{19}$$

Hence $t \to \psi_k(|t|)$ is non-smooth at 0.

3. Having $\hat{x}[k]$ in (17) *independent* of the value of δ requires that $\psi_k'(t)$ is constant for all $t > 0$. Since $\psi_k'(t) \geq \psi_k'(0^+)$ by the convexity of ψ_k, (19) leads to

$$\psi_k(t) = \lambda_k t,\ \ \forall t \in \mathrm{R}_+,\, k \in I_1, \lambda_k > 0. \tag{20}$$

4. It follows that (17) reads $|\partial_k\Phi(\hat{x})| = \lambda_k$. We have $\hat{x}[k] \approx 0$ provided that ϕ' has steep increase near to zero. E.g., when ϕ is of the form (10), we have $\hat{x}[k] = \sqrt{\alpha/C}$, where C is the unique solution of the equation

$$\int_\Omega \frac{|\nabla w_k|^2}{\sqrt{|\nabla w_k|^2 + C}}\,ds = \lambda_k.$$

Clearly, $\hat{x}[k]$ decreases to zero when $\alpha \searrow 0$.

5. Similarly, for all $j \in I_0$, $\hat{x}_j = 0$ means that

$$\left|\int_\Omega \phi'(|\hat{x}[k]\ \nabla w_k|)\,(\nabla w_j)^T\,\frac{\nabla w_k}{|\nabla w_k|}\right| ds \leq \psi_j'(0^+). \tag{21}$$

Using the same arguments as in item 2 above, we deduce that ψ_j must satisfy

$$\psi_j'(0^+) > 0.$$

Gibbs effect. Consider now a coefficient $y_T[j] = 0$ corresponding to a large $(\widetilde{W}u_o)[j]$. Since y_T contains no information on the true value of this coefficient, the best choice is that $\hat{x}[j]$ fits the prior, i.e. that it minimizes Φ. Consequently, we require that $\partial_j\Phi(\hat{x})$ is as close as possible to zero. On the

other hand, $\hat{x}[j]$ must satisfy (4), since we wish that $\hat{x}[j] \neq 0 = y_T[j]$. Combining these two requirements entails that the right-hand side of (4) must be as close as possible to zero. Noticing that $\psi_j'(t) \geq \psi_j'(0^+) > 0$ for all $t \in \mathrm{R}_+$ and using (15), we will choose $\psi_j'(t) = \psi_j'(0^+) = \lambda_j > 0$, for all $t \in \mathrm{R}_+$, i.e.

$$\psi_j(t) = \lambda_j t, \ \ \forall t \in \mathrm{R}_+, \, j \in I_0.$$

2.4 The objective function

Taking together all these elements, Ψ in (9) reads

$$\Psi(x) = \sum_{i \in I_1} \lambda_i \left| (x-y)[i] \right| + \sum_{i \in I_0} \lambda_i \left| x[i] \right|. \qquad (22)$$

Let us sumarize: given the coefficients y_T obtained by hard thresholding as given in (7), the restored function $\hat{u}$ reads $\hat{u} = W\hat{x}$ where $\hat{x}$ minimizes

$$\begin{aligned} F_y(x) &= \sum_{i \in I_1} \lambda_i \left| (x-y)[i] \right| + \sum_{i \in I_0} \lambda_i \left| x[i] \right| \qquad (23) \\ &+ \int_\Omega \phi(|\nabla W x|) \, ds. \end{aligned}$$

This function F_y is non-smooth for every x such that $x[i] = y_T[i]$ for some i. Using the results of [17], we can expect that minimizers $\hat{x}$ of F_y involve many indexes $i \in I_1$ for which $\hat{x}[i] = y[i]$ exactly, and that likewise, $\hat{x}[i] = 0$ for many $i \in I_0$. It can be deduced that the smoothing of aberrant coefficients, as well as the restoration of erroneously thresholded coefficients, is stable with respect to small perturbations of the input data.

We have observed that the minimizers $\hat{x}$ of F_y are very stable with respect to the choice of the parameters $\{\lambda_i\}$. This can be explained by the fact that since F_y is nonsmooth, minimizers $\hat{x}$ are located at "kinks" which are stable with respect to parameters and data. Some orientations for the choice of λ_i can be derived from the conditions for minimum (11)-(14). Let us come back to the data considered in (15) and focus on an index $k \in I_1$ relevant to an outlier. Notice that many edge-preserving functions ϕ satisfy $\|\phi'\|_\infty \leq 1$. Introducing this in (18) shows that

$$|\partial_k \Phi(x)| \leq \int_\Omega |\nabla w_k| \ ds, \ \ \forall x, \, \forall y.$$

In 4 in § 2.3 we saw that $|\partial_k \Phi(\hat{x})| = \lambda_k$. If λ_k is larger than the right-side of the expression given above, for any $y_T[k]$ we can write that $|\partial_k \Phi(\hat{x} + (y_T[k] - \hat{x}[k])e_k)| \leq \lambda_k$ which means that $\hat{x}[k] = y_T[k]$, i.e. that the outlier at k cannot be removed. In this way, we find that it is necessary that

$$\lambda_k < \int_\Omega |\nabla w_k| \ ds, \quad \forall k \in I_1. \qquad (24)$$

When k addresses frames of the same scale, $\int_\Omega |\nabla w_k| \ ds$ is constant. Thus (24) provides an upper bound for each scale. On the other hand, (21) is guaranteed if

$$\lambda_j \geq \left| \int_\Omega (\nabla w_j(s))^T \, \frac{\nabla w_k}{|\nabla w_k|} \, ds \right|, \quad \forall j \in I_0, \, \forall k \in I_1. \qquad (25)$$

If either j and k correspond to different scales, or if j and k are quite distant, the integral above is close to zero. So, for every $j \in I_0$, the inequality above is tested only for a restricted number of indexes $k \in I_1$.

Parameters when $\{w_k\}$ is a wavelet basis. We consider henceforth a wavelet basis generated by $2^d - 1$ mother wavelets w^m for $m \in \{1, \ldots, 2^d - 1\}$, defined on $\Omega \subset \mathrm{R}^d$. Let j and κ denote the scale and the space (or time) parameters, respectively. In such a case, I is an arrangement of all indexes (j, κ, m) and w_k in (24) is of the form

$$w_{j,\kappa}^m(s) = 2^{-\frac{dj}{2}} w^m(2^{-j}s - \kappa).$$

Using a change of variables, the upper bound in (24) is

$$\int_\Omega |\nabla(w_{j,\kappa}^m)| \, ds = 2^{\left(\frac{d}{2}-1\right)j} \int_\Omega |\nabla w^m| \, ds.$$

This suggests we take

$$\lambda_{j,\kappa}^m = 2^{\left(\frac{d}{2}-1\right)j} \lambda_1^m, \quad \forall (j, \kappa, m) \in I_1,$$

where $\lambda_1^m \leq \int_\Omega |\nabla w^m(s)| \, ds$.

Similarly, (25) leads to

$$\lambda_{j,\kappa}^m = 2^{\left(\frac{d}{2}-1\right)j} \lambda_0^m, \quad \forall (j, \kappa, m) \in I_0,$$

where for all $m' \in \{1, \ldots, 2^d - 1\}$ and for all $(j,k) \neq (0,0)$,

$$\lambda_0^m \geq \left| \int_\Omega (\nabla(w_{k,j}^{m'}))^T \, \frac{\nabla w^m}{|\nabla w^m|} \, ds \right|.$$

3 Experiments

3.1 Minimization scheme

Since the kernel of the gradient operator ∇ is the set of all constant functions, $u \mapsto \int_\Omega \phi(|\nabla u|) \, ds$ is constant in the direction of the mean value of u. It is strictly convex in every other direction provided that ϕ is strictly convex as well. $\widetilde{W}$ being a wavelet transform, the mean value of any estimate $\hat{u} = W\hat{x}$ is supported by a single wavelet coefficient. It is therefore determined by Ψ and equals the mean value of v_τ. We conclude that *the minimizer $\hat{x} = W\hat{u}$ of F_y is unique.*

We compute $\hat{x}$ using a subgradient descent scheme. Put $x_0 = y_T$ and, for all $k \in \mathrm{N}$, compute

$$x_{k+1} = x_k - t_k g_k,$$

where g_k is a subgradient of F_y at x_k and $t_k > 0$. Using classical results on minimization methods (see [19]), we can prove that, if Ω is finite, $\lim_{k \to \infty} t_k = 0$ and $\sum_{k=0}^{\infty} t_k = \infty$, then

$$\lim_{k \to \infty} x_k = \hat{x}.$$

3.2 Denoising of a signal

We consider the restoration of the 512-length original signal in Fig. 1 from the data shown there, contaminated with white Guassian noise with standard deviation $\sigma = 10$. The restoration in Fig. 2 is obtained using the sure-shrink method [11] and the toolbox WaveLab. The result displayed in Fig. 3 is the minimizer of a function $\mathcal{F}_v$ of the form (1) where ϕ is as given in (10), for $\alpha = 0.1$ and $\lambda = 0.01$. Smooth zones are rough, edges are slightly smoothed and spikes are eroded, while some diffused noise is still visible on the signal.

The restorations presented next are based on thresholded wavelet coefficients: $\widetilde{W}$ is an orthogonal basis of Daubechies wavelets with 8 vanishing moments and y_T is obtained according to (7). The optimal T, as given in (5), reads $T = 35$. The wavelet-thresholding estimate Wy_T is shown in Fig. 4. It involves important Gibbs artifacts, as well as wavelet-shaped oscillations due to aberrant coefficients. Using the same coefficients y_T, we calculated the minimizer $\hat{x}$ of F_y as given in (23) where ϕ is as given in (10), $\alpha = 0.05$, $\lambda_{j,\kappa} = 0.5 \times 2^{j/2}$ if $(j,\kappa) \in I_0$ and $\lambda_{j,\kappa} = 1.5 \times 2^{j/2}$ if $(j,\kappa) \in I_1$. The resultant restoration $\hat{u} = W\hat{x}$, shown in Fig. 5, involves sharp edges and well denoised smooth pieces.

Next we consider y_T, obtained by (7) for $T = 23$. These coefficients have a richer information content, but the relevant estimate Wy_T, seen in Fig. 6, manifests Gibbs artifacts and many wavelet-shaped artifacts. Below we compare restorations where F_y is of the form (9) for different choices of ψ_i. In spite of the considerations developed in § 2.3, it seems intuitive to take $\psi_{j,\kappa}(t) = \lambda_{j,\kappa} t^2$ in (9). Such a restoration is displayed in Fig. 7 where $\alpha = 0.05$ $\lambda_{j,\kappa} = 0.1$ if $(j,\kappa) \in I_0$, and $\lambda_{j,\kappa} = 0.2$ if $(j,\kappa) \in I_1$. The Gibbs oscillations are well removed but, because of the quadratic form of $\psi_{j,\kappa}$ for $(j,\kappa) \in I_1$, outliers overcontribute to F_y and biases the estimate. Another possibility which may seem reasonable is to cancel the term indexed by I_0, i.e. to consider $\psi_{j,\kappa}(t) = 0$ for $(j,\kappa) \in I_0$. The result can be seen in Fig. 8 where $\psi_{j,\kappa}(t) = 0.2\,t$ for all $(j,\kappa) \in I_1$ and $\alpha = 0.05$. Once again, the thresholded coefficients are well restored but we observe that leaving too much freedom to these coefficients prevents the method from removing the outliers efficiently. Fig. 9 illustrates the proposed method: F_y is of the form (23) with ϕ as given in (10), and the same parameters as in Fig. 5, namely $\alpha = 0.05$, $\lambda_{j,\kappa} = 0.5 \times 2^{j/2}$ if $(j,\kappa) \in I_0$ and $\lambda_{j,\kappa} = 1.5 \times 2^{j/2}$ if $(j,\kappa) \in I_1$. In this restoration, edges are neat and polynomial parts are well recovered. Fig. 10 illustrates how restored coefficients $\hat{x}$ are placed with respect to y_T and the coefficients of the original signal $\widetilde{W}u_o$. In particular, we observe how erroneously thresholded coefficients are restored and how outliers are smoothed.

3.3 Denoising of an image

In this experiment we consider the denoining of the 256×256 picture of Lena, Fig. 11 (a), from noisy data obtained by adding white Gaussian noise with standard de-

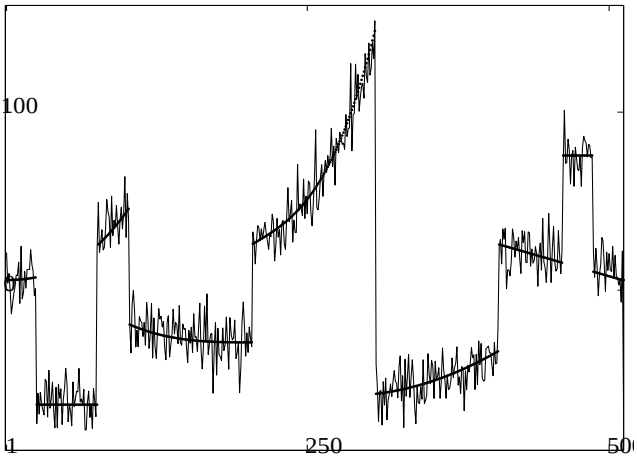

Figure 1: Original signal (dotted line) and noisy data (solid line).

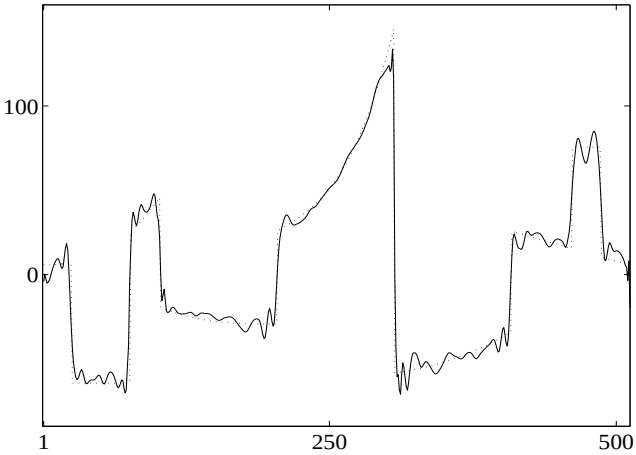

Figure 2: Denoising using the Donoho-Johnstone's *Sure-shrink* method.

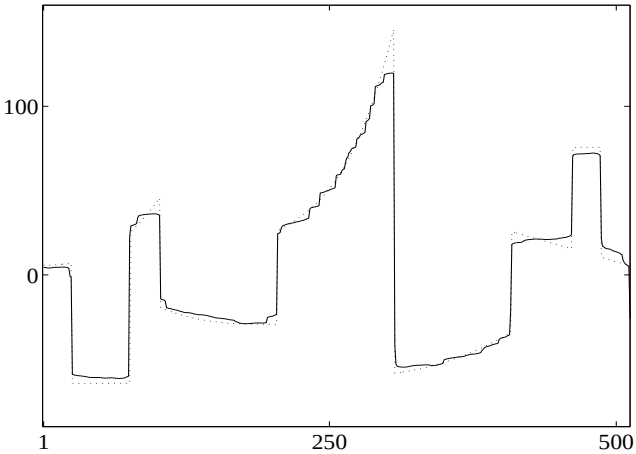

Figure 3: Denoising by minimizing $\mathcal{F}_v$ as given in (1) where $\phi(t) = \sqrt{0.05 + t^2}$ and $\lambda = 0.01$.

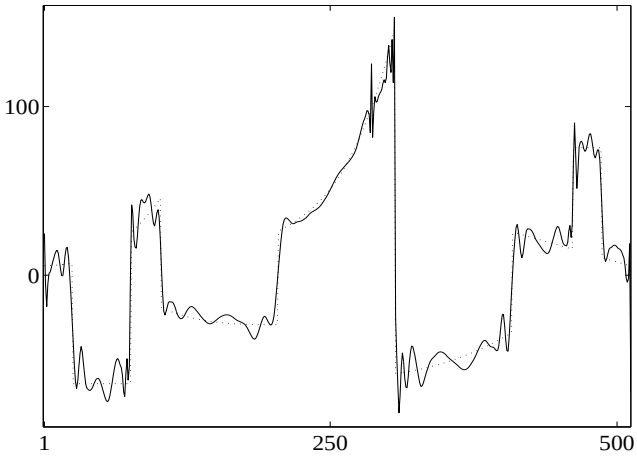

Figure 4: Denoising using wavelets thresholding with Donoho-Johnstone's optimal threshold $T = 35$.

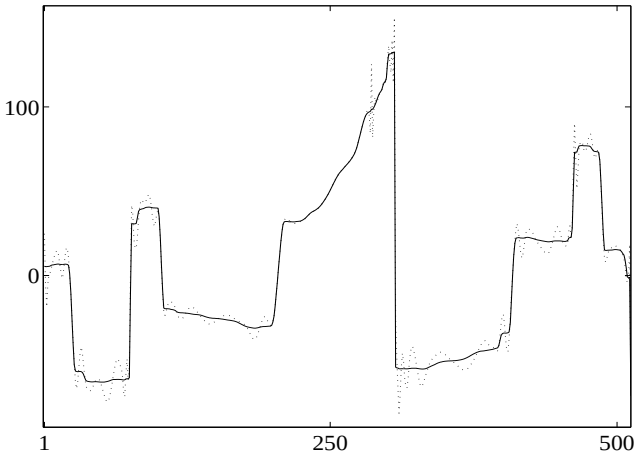

Figure 5: Denoising by restoration of the wavelet coefficients relevant to Fig. 4 using F_y in (23) with $\phi(t) = \sqrt{0.05 + t^2}$, $\lambda_{j,\kappa} = 0.5 \times 2^{j/2}$ if $(j,\kappa) \in I_0$, $\lambda_{j,\kappa} = 1.5$ if $(j,\kappa) \in I_1$.

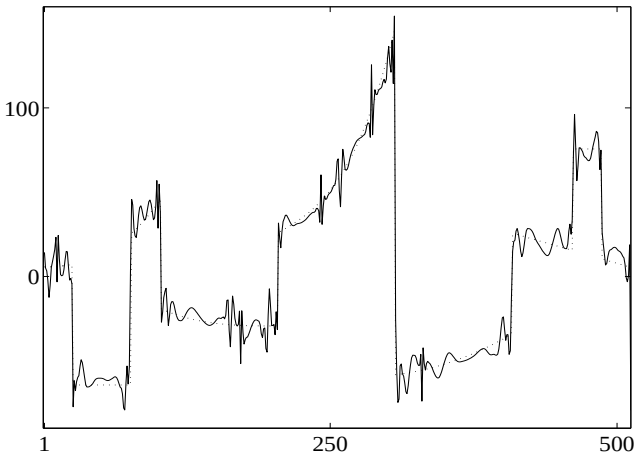

Figure 6: Denoising using wavelets thresholding with an under-optimal threshold $T = 23$.

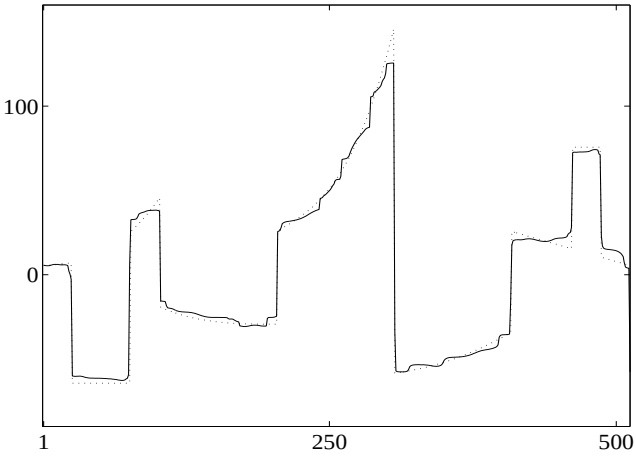

Figure 7: Restoration of the wavelet coefficients relevant to Fig. 6 by minimizing F_y in (9) with $\phi(t) = \sqrt{0.05 + t^2}$, $\psi_i(t) = 0.1t^2$ if $i \in I_0$ and $\psi_i(t) = 0.2t^2$ if $i \in I_1$.

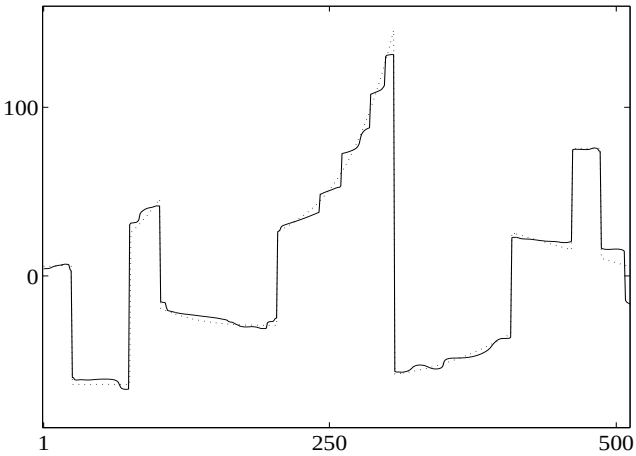

Figure 8: Restoration of Fig. 6 using F_y in (9) where $\phi(t) = \sqrt{0.05 + t^2}$, $\psi_i(t) = 0$ if $i \in I_0$ and $\psi_i(t) = 0.2t$ if $i \in I_1$.

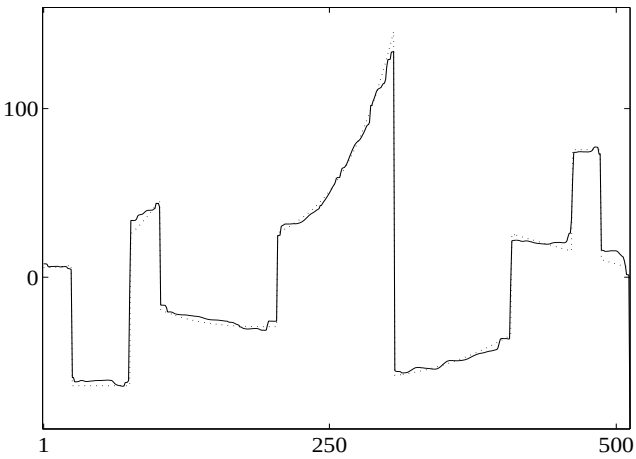

Figure 9: The proposed method: restoration of Fig. 6 using F_y in (23) with $\phi(t) = \sqrt{0.05 + t^2}$, $\lambda_{j,\kappa} = 0.5 \times 2^{j/2}$ if $(j,\kappa) \in I_0$ and $\lambda_{j,\kappa} = 1.5 \times 2^{j/2}$ if $(j,\kappa) \in I_1$.

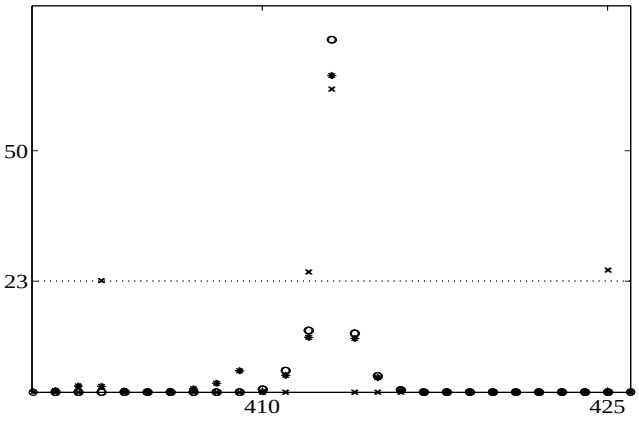

Figure 10: Magnitude of wavelet coefficients: $*$ signal restored by the proposed method (Fig. 9), $\circ$ original signal, $\times$ thresholded noisy signal (Fig. 6).

viation 20. The restoration in Fig. 12 (a) is obtained by thresholding the wavelet coefficients, see (7), with respect to Donoho-Johnstone's threshold, given in (5), which now reads $T = 100$. This image is very smooth, a lot of details are lost, and Gibbs oscillations are visible near the edges. In Fig. 12 (b) we show the result from total-variation restoration which corresponds to $\mathcal{F}_v$ of the form (1) with $\phi(t) = t$ and $\lambda = 0.03$. As expected, this restoration exhibits a staircasing effect since it is constant on many regions. The image in Fig. 13 (a) is obtained by thresholding the wavelet coefficients with respect to $T = 50$. This T is smaller than Donoho-Johnstone's threshold and the image presents many wavelet-shaped oscillations due to aberrant wavelet coefficients, as well as some Gibbs oscillations. It is used as input data for the specialized objective function F_y given in (23), where ϕ is as given in (10). The restoration in Fig. 13 (b) is obtained for $\lambda_i = 0.5$ if $i \in I_0$, $\lambda_i = 1.5$ if $i \in I_0$. This image has a quite natural appearance, and edges and texture are better preserved.

The numerical cost of variational methods become a real burden when images have a large size. In order to circumvent this problem, we have tested an approximation of the proposed method. Let y_T be the wavelet transform of the thresholded image. According to (11), the minimizer $\hat{x}$ of F_y satisfy

$$|\partial_i \Phi(\hat{x})| \leq \lambda_i, \quad \forall i \in I_1.$$

The idea of this approximation is to test for every $i \in I_1$ whether or not $|\partial_i \Phi(y_T)| > \lambda_i$. If $|\partial_i \Phi(y_T)| \leq \lambda_i$, we take simply $\hat{x}[i] = y_T[i]$. Otherwise, if $|\partial_i \Phi(y_T)| > \lambda_i$, we consider that $y_T[i]$ is an outlier. To restore such an outlier, we can take for the relevant $\hat{x}[i]$ either the median or the mean of the neighboring coefficients at the same scale. When outliers arise in homogeneous regions, we can just set $\hat{x}[i] = 0$. The Gibbs oscillations are not considered in this approximated method, so we have $\hat{x}[i] = y_T[i] = 0$ for all $i \in I_0$. The image obtained by this method for $T = 50$ and $\lambda_i = 5$ for all $i \in I_1$, is displayed on Fig. 14 (a). Let us emphasize that the

(a) Original image. (b) Noisy image.

Figure 11: Original and noisy images.

(a) Wavelets thresholding with the optimal thereshold $T = 100$. (b) Total-variation restoration: $\mathcal{F}_v$ as in (1) with $\phi(t) = t$.

Figure 12: Classical denoising methods.

image of the error $v_\tau - \hat{u}$, presented in Fig. 14 (b), exhibits the oscillations due to aberrant wavelet coefficients and that it does not present any structural information. This approximated method being computationally fast, it can be extended to translation invariant wavelets [9]. In Fig. 15 (a) we show the restoration obtained by the standard translation invariant wavelets thresholding, corresponding to $T = 50$ again. Although its quality is improved with respect to the image in Fig. 13 (a), it involves a lot of wavelet-shaped artifacts. This image is used as input data to our fast approximated method. The obtained restoration, shown in Fig. 15 (b), is of high quality, since edges and details are nicely recovered.

4 Conclusion

We proposed a method to denoise images and signals by restoring the thresholded frame coefficients of the noisy data. The restored coefficients minimize a specially designed objective function which allows the erroneously thresholded coefficients to be restored and the outliers to be removed, without substantially modifying the remaining coefficients. Our method is not sensitive to the probability distribution of the noise. We present numerical experiments with orthogonal bases of Daubechies wavelets. These experiments demonstrate the effectiveness of our method over alternative denoising methods.

(a) Wavelets thresholding with $T = 50$ (b) Restoration of (a) by the proposed method

Figure 13: The proposed method.

(a) Restoration of Fig. 13 (a) by the fast method (b) Outliers detected by the fast method

Figure 14: Fast approximation of the proposed method.

(a) Translation invariant wavelet thresholding ($T = 50$) (b) Fast method adapted to translation invariant wavelets

Figure 15: Translation invariant wavelets.

References

[1] A. Antoniadis and Jianqing Fan. Regularization of wavelet approximations. *Journal of Acoustical Society America*, 96(455):939–967, Sep. 2001.

[2] Gilles Aubert and Pierre Kornprobst. *Mathematical problems in images processing*. Springer-Verlag, Berlin, 2002.

[3] Murat Belge, Misha Kilmer, and Eric Miller. Wavelet domain image restoration with adaptive edge-preserving regularization. *IEEE Transactions on Image Processing*, 9(4):597–608, Apr. 2000.

[4] Michael Black and Anand Rangarajan. On the unification of line processes, outlier rejection, and robust statistics with applications to early vision. *International Journal of Computer Vision*, 19(1):57–91, 1996.

[5] Ch. Bouman and K. Sauer. A generalized Gaussian image model for edge-preserving MAP estimation. *IEEE Transactions on Image Processing*, 2(3):296–310, July 1993.

[6] E. J. Candès and F. Guo. New multiscale transforms, minimum total variation synthesis. Applications to edge-preserving image reconstruction. *Signal Processing*, 82, Mar. 2002.

[7] Antonin Chambolle, Ronald DeVore, Nam-yong Lee, and Bradley Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise remouval through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–335, Mar. 1998.

[8] Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, Feb. 1997.

[9] R. R. Coifman and D. Donoho. Translation-invariant de-noising. Technical Report Report 475, Stanford University, Dept. of Statistics, 1995.

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

[11] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of Acoustical Society America*, 90, 1995.

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

[13] Peter J. Green. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Transactions on Medical Imaging*, MI-9(1):84–93, Mar. 1990.

[14] S.Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, New York, 1 edition, 1995.

[15] F. Malgouyres. Minimizing the total variation under a general constraint for image restoration. *IEEE Transactions on Image Processing*, 11(12):1450–1456, Dec. 2002.

[16] Mila Nikolova. Local strong homogeneity of a regularized estimator. *SIAM Journal on Applied Mathematics*, 61(2):633–658, 2000.

[17] Mila Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM Journal on Numerical Analysis*, 40(3):965–994, 2001.

[18] L. Rudin, S. Osher, and C. Fatemi. Nonlinear total variation based noise removal algorithm. *Physica*, 60 D:259–268, 1992.

[19] N. Z. Shor. *Minimization Methods for Non-Differentiable Functions*, volume 3. Springer-Verlag, 1985.

[20] E. P. Simoncelli and E. H. Adelson. Noise removal via bayesian wavelet coding. In *Proceedings of the IEEE International Conference on Image Processing*, pages 379–382, Lausanne, Switzerland, Sep. 1996.

[21] C. R. Vogel and Mary. E. Oman. Fast, robust total variation-based reconstruction of noisy, blurred images. *IEEE Transactions on Image Processing*, 7(6):813–824, Mar. 1998.