RECONSTRUCTION OF WAVELET COEFFICIENTS USING TOTAL VARIATION MINIMIZATION

SYLVAIN DURAND* AND JACQUES FROMENT†

Abstract. We propose a model to reconstruct wavelet coefficients using a total variation minimization algorithm. The approach is motivated by wavelet signal denoising methods, where thresholding small wavelet coefficients leads pseudo-Gibbs artifacts. By replacing these thresholded coefficients by values minimizing the total variation, our method performs a nearly artifact free signal denoising. In this paper, we detail the algorithm based on a subgradient descent combining a projection on a linear space. The convergence of the algorithm is established and numerical experiments are reported.

Key words. wavelet, total variation, denoising, subgradient method

AMS subject classifications. (MSC 2000) 26A45, 65K10, 65T60, 94A12

1. Introduction. Let us consider the problem of denoising nearly piecewise smooth functions presenting sharp discontinuities, following the additive noise model

$$\tilde{u} = u + \epsilon, \tag{1.1}$$

where $\tilde{u}$ represents the observed data, u the noiseless function to estimate and ϵ the noise or, more generally, an unknown error. This problem occurs in number of applications, especially in the signal and image processing community, where one tries to recover the original smoothness of a signal while the main discontinuities are preserved. Among the various solutions that have been proposed, we will focus our attention on two promising approaches which recently appear : wavelet thresholding and total variation minimization.

Wavelet denoising consists in decomposing the noisy data into an orthogonal wavelet basis, in suppressing the wavelet coefficients smaller than a given amplitude using a so-called soft or hard thresholding, and in transforming the data back into the original domain : let $\left\{\{\psi_{j,n}\}_{(j,n)\in\Psi}, \{\phi_{J,n}\}_{n\in\Phi}\right\}$ be an orthogonal basis of wavelets and scaling functions on the interval $I=[a,b]$ as described by Cohen, Daubechies and Vial [7], so that we can write any function $\tilde{u} \in L^2(I)$ as the sum of the series

$$\tilde{u} = \sum_{(j,n)\in\Psi} \langle \tilde{u}, \psi_{j,n}\rangle \psi_{j,n} + \sum_{n\in\Phi} \langle \tilde{u}, \phi_{J,n}\rangle \phi_{J,n} \tag{1.2}$$

where, $\forall u, v \in L^2(I)$,

$$\langle u, v\rangle = \int_I u(x)\, v(x)\, dx. \tag{1.3}$$

In $\psi_{j,n}$, j is the index of scale and n the translation factor, taking values in the countable set Ψ. The term $\phi_{J,n}$ denotes the scaling function on scale 2^J and translated

* CMLA, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94235 Cachan cedex, France (`sdurand@cmla.ens-cachan.fr`) & LAMFA, Université de Picardie-Jules Verne, 33 rue Saint Leu, 80039 Amiens cedex 1, France.

† PRISME, UFR Math. et Info. Université Paris 5 R. Descartes, 45 rue des Saints-Pères, 75270 Paris cedex 06, France (`froment@math-info.univ-paris5.fr`) & CMLA, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94235 Cachan cedex, France.

by n, which is indexed to a finite set Φ. The family $\{\phi_{J,n}\}_{n\in\Phi}$ is an orthonormal basis of a space V_J that belongs to a multiresolution approximation of $L^2(I)$.

The hard thresholding operator τ is defined by

$$\tau(x) = \begin{cases} x & \text{if } |x| \geq \lambda, \\ 0 & \text{if } |x| < \lambda, \end{cases} \tag{1.4}$$

while in the case of soft thresholding, the operator τ is

$$\tau(x) = \begin{cases} x - \operatorname{sgn}(x)\lambda & \text{if } |x| \geq \lambda, \\ 0 & \text{if } |x| < \lambda. \end{cases} \tag{1.5}$$

The denoised signal using wavelet thresholding is simply

$$u_0 = \sum_{(j,n)\in\Psi} \tau(\langle \tilde{u}, \psi_{j,n}\rangle)\psi_{j,n} + \sum_{n\in\Phi} \langle \tilde{u}, \phi_{J,n}\rangle \phi_{J,n}. \tag{1.6}$$

We will denote M the map that records the indexes of retained coefficients :

$$M = \{(j,n) \in \Psi \ : \ |\langle \tilde{u}, \psi_{j,n}\rangle| \geq \lambda\}. \tag{1.7}$$

Because of its simplicity, the algorithm sketches in (1.6) has been widely used by engineers since the beginning of wavelet in signal processing. It has been formalized by Donoho and Johnstone in [13], where they proved that the performance associated to non-linear thresholding estimator in orthogonal bases is close to an ideal coefficient selection and attenuation. In addition, among classical orthogonal bases, wavelet series outperforms Fourier or cosine series in the representation of piecewise-smooth functions (see e.g.[6, 8]) : the efficiency of the estimator depends on the rate of decay of the sorted decomposition coefficients and, thanks to the wavelets time-localization, the decay of wavelet coefficients in the neighborhood of discontinuities is faster that the decay of Fourier coefficients. However, wavelet thresholding method is still a regularization process and the estimator presents oscillations in the vicinity of function's discontinuities. Such oscillations are very close to the Gibbs phenomena exhibited by Fourier thresholding, although they are more local and of smaller amplitude. For this reason, they are called pseudo-Gibbs phenomena. These oscillations do not affect too much the L^2 error between the original noiseless signal and the estimated one, but they do affect the visual quality of the result : it is often impossible to perform a complete denoising while keeping the threshold small enough to avoid the pseudo-Gibbs phenomena. If ϵ is a Gaussian white noise of standard deviation σ, the threshold should be set to

$$\lambda = \sigma\sqrt{2\log N}, \tag{1.8}$$

N being the number of samples of the digital signal, so that the estimator is the best in the min-max sense as N tends to infinity [13]. The use of the soft thresholding operator (1.5) instead of the more intuitive hard one (1.4) allows to partially reduce the pseudo-Gibbs phenomena [12]: thanks to the continuity of the soft thresholding operator, the structure of the wavelet coefficients is better preserved. However, the soft thresholding operator introduces another type of artifact : since all wavelet coefficients are lowered, local averages are not preserved, leading peaks to be eroded.

Total variation denoising follows a completely different approach. It has been introduced for the first time by Rudin, Osher and Fatemi in [21], in the context of image denoising. They proposed to minimize the total variation (TV) of a function $u \in L^2(\Omega)$

$$\mathrm{TV}(u) := \int_\Omega |\nabla u| \, dx, \tag{1.9}$$

where Ω is a bounded and convex region of $\mathbb{R}^d$, subject to the fidelity constraint

$$||u - \tilde{u}||_{L^2(\Omega)} = \sigma, \tag{1.10}$$

σ being an estimated error level. The noise is reduced while discontinuities are preserved, in contrast with other regularization techniques, generally using a L^2 norm, where discontinuities are smoothed. Although the TV functional seems to be particularly relevant in regularizing piecewise smooth functions, it generates artifact as well, known as the staircase effect (see [4, 10] for numerical evidence, and [20] for theoretical analysis in a general framework) : TV-based algorithms tend to restore piecewise constant functions. For example and though they have the same total variation, a staircase will be preferred to a ramp (see Figure 5.5). This is mainly due to the lack of regularity of the TV.

In [23] (see also [1, 3, 11]), Vogel and Oman propose to replace the TV by the regularized functional

$$J_\beta(u) := \int_\Omega \sqrt{|\nabla u|^2 + \beta^2} \, dx \tag{1.11}$$

where β is a small positive parameter. Being differentiable, the TV-regularized problem may be stated using various optimization techniques and the staircase effect may be removed. However, the experiments we have performed tend to show the difficulty of finding a suitable value for β : if β is significantly greater than 0, the staircase effect is effectively removed but the noise is diffused, and for β larger again, discontinuities are smoothed. If not, the regularization is similar to the one performed with the TV, but in any case the computation (due to the introduction of the square root in (1.11)) is much slower. Another staircase reduction strategy is to introduce higher order derivatives in the functional, so that inopportune jumps are penalized : see e.g. [2, 4].

The denoising algorithm we are presenting in this article combines the wavelet and the total variation approaches. In our knowledge, a few articles only use these complementary tools in the context of signal or image processing. In [15], Malgouyres, Rougé and one of us point out the complementarity of these tools and, in [17], Malgouyres proposed a deblurring algorithm based on both tools, in the context of satellite imaging. Another close approach is exposed by Chan and Zhou in [5] ; let us precise that our algorithm was developed independently. As we do, Chan and Zhou noticed that a total variation minimization may be applied to remove the pseudo-Gibbs phenomenon generated by wavelet thresholding. They proposed to solve the following convex and unconstrained problem

$$\min_{(c_{j,n})} \ \alpha \int_\Omega |\nabla u_{(c_{j,n})}| \, dx + \frac{1}{2} ||u - \tilde{u}||^2_{L^2(\Omega)} \tag{1.12}$$

where $u_{(c_{j,n})}$ is the function reconstructed using the wavelet coefficients $(c_{j,n})$, these coefficients satisfying $c_{j,n} = 0$ if $(j,n) \notin M$. The first term of this functional reduces the oscillations of the estimated function by modifying the values of the retained wavelet coefficients, so that the total variation is diminished. The second term is a classical L^2 fitting term. The regularization parameter α is used to balance the respective influence of these two terms.

The model we are presenting is somewhat opposite to the one of Chan and Zhou. We propose to reconstruct a function with minimal total variation such that for indexes belonging to M, its wavelet coefficients are the same than the wavelet coefficients of the observed function $\tilde{u}$. That is, retained wavelet coefficients are not modified while canceled coefficients are no more set to 0, but to values that minimize the total variation. This main idea results simply in the following remark : apart from the pseudo-Gibbs phenomenon, wavelet thresholding works well in denoising functions. Therefore, we can assume that the unknown original noiseless signal u has the same wavelet coefficients than the thresholded one in the location given by the map M. Wavelet denoising algorithm makes the choice of setting the coefficients outside M to 0, leading oscillations in the vicinity of discontinuities. Because of the strong dependency between wavelet coefficients in the original noiseless signal, this is far from an optimal choice. By proposing to set the coefficients outside M to the values that minimize the total variation of the reconstructed function, occurrence of oscillations is discouraged : a structure of coefficients compatible with sharp discontinuities is recovered. Notice that, in contrary to the model of Chan and Zhou, we define a constrained problem but which is free of regularization parameter. Since the important wavelet coefficients are kept unchanged, there is no need for a fitting term. The only parameter is the threshold λ, and it can be fixed according to (1.8).

We have exposed this denoising model in [14], in a mathematically simplified version. The aim of the current article is to present and justify the algorithm used to solve the model in the 1D case, which is based on a subgradient descent combining a projection on a linear space. Section 2 recalls the model in the continuous case, while section 3 gives its counterpart in the discrete case, which defines the effective algorithm. Section 4 is devoted to the convergence study of the algorithm. It is proved that, provided the stepsize is not decreasing too fast to 0, the sequence of computed vectors converges to a solution of the discrete model. Section 5 presents numerical results on two signals and comparison with wavelet and total variation classical denoising algorithms is performed.

2. The continuous model. For simplicity, we present the model in the 1D case. But a similar development can be performed on $\mathbb{R}^d$, for example to denoise images. The total variation of any unidimensional function u in I is defined by

$$\mathrm{TV}(u) = \sup_{(x_l)} \sum_{l=1}^{L} |u(x_l) - u(x_{l-1})| \tag{2.1}$$

where the supremum is on all sequences (x_l) such that $a \leq x_1 < x_2 < \ldots < x_L \leq b$. Let X be the space of $L^2(I)$ functions such that their TV norm is finite, which is an Hilbert space for the scalar product $\langle .,. \rangle$ defined on $L^2(I)$. We denote by U the constraint space

$$U = \{u \in X : \forall (j,n) \in M, \langle u, \psi_{j,n} \rangle = \langle \tilde{u}, \psi_{j,n} \rangle;\ \forall n \in \Phi, \langle u, \phi_{J,n} \rangle = \langle \tilde{u}, \phi_{J,n} \rangle\}, \tag{2.2}$$

where M is defined in (1.7). The set U is an affine space with direction given by the linear space

$$V = \{v \in X : \forall (j,n) \in M, \langle v, \psi_{j,n} \rangle = 0;\ \forall n \in \Phi, \langle v, \phi_{J,n} \rangle = 0\}. \tag{2.3}$$

Consider u_0 introduced in (1.6). Since $u_0 \in U$, we have

$$U = \{u_0\} + V. \tag{2.4}$$

We propose to solve the variational problem

PROBLEM 2.1. *Find* $u^* \in U$ *such that* $TV(u^*) = \min_{u \in U} TV(u)$.

A small technical difficulty comes from the fact that the TV-functional is not differentiable. In order to avoid the use of a subgradient instead of a gradient, we could make the choice of a regularization as in (1.11), but as we wrote it in the introduction, we didn't find that this would increase the performance of the algorithm. Since TV is a convex function, we can define a subgradient of TV at u as any function $g_{\mathrm{TV}}(u)$ satisfying

$$\forall v \in X,\ \mathrm{TV}(v) \geq \mathrm{TV}(u) + \langle g_{\mathrm{TV}}(u), v - u \rangle, \tag{2.5}$$

and we get the classical result

THEOREM 2.2. *Any solution* u^* *of Problem 2.1 is given by*

$$\forall t > 0,\ u^* = P\left(u^* - t\, g_{TV}(u^*)\right), \tag{2.6}$$

for P *the affine projector onto* U *that minimizes the distance.*

Proof. As TV is convex, u^* is a solution of Problem 2.1 iff

$$\begin{aligned}
& u^* \in U \text{ and } \forall v \in U,\ \langle g_{\mathrm{TV}}(u^*), v - u^* \rangle \geq 0 \\
\Longleftrightarrow\ & u^* \in U \text{ and } \forall v \in U,\ \forall t > 0,\ \langle u^* - (u^* - t g_{\mathrm{TV}}(u^*)), v - u^* \rangle \geq 0 \\
\Longleftrightarrow\ & \forall t > 0,\ u^* = P(u^* - t\, g_{\mathrm{TV}}(u^*)).
\end{aligned}$$

□

3. The discrete model. We now assume that functions are signals with N samples, that is $X = \mathbb{R}^N$. For such a discrete signal we use the vectorial notation $u = (u^1, u^2, \ldots, u^N)$, we denote by $(.,.)$ the standard scalar product on $\mathbb{R}^N$

$$\forall u, v \in \mathbb{R}^N,\ (u, v) = \sum_{n=1}^{N} u^n v^n, \tag{3.1}$$

and $||.||$ is the associated Euclidean norm. The discrete total variation of u is given by

$$\mathrm{TV}(u) = \sum_{n=1}^{N-1} |u^{n+1} - u^n|. \tag{3.2}$$

Let u^c be the continuous, or analog function in the approximation space V_0 (scale 1), associated to the discrete signal u by $u^c = \sum_n u^n \phi_{0,n}$ and $u^n = \langle u^c, \phi_{0,n} \rangle$. The sets U and V are defined as in (2.2) and (2.3), using a FWT (Fast Wavelet Transform [18]) to compute from u up to the coarse scale 2^J (for a given $J < 0$), the sequence of wavelet coefficients $(\langle u^c, \psi_{j,n} \rangle)_{j,n \in \Psi}$ (with $J \leq j < 0$), plus the remaining approximation $(\langle u^c, \phi_{J,n} \rangle)_{n \in \Phi}$:

$$\begin{aligned}
U = \{u \in X : \forall (j,n) \in M,\ & \langle u^c, \psi_{j,n} \rangle = \langle \tilde{u}^c, \psi_{j,n} \rangle \text{ and} \\
& \forall n \in \Phi,\ \langle u^c, \phi_{J,n} \rangle = \langle \tilde{u}^c, \phi_{J,n} \rangle \} \\
V = \{v \in X : \forall (j,n) \in M,\ & \langle v^c, \psi_{j,n} \rangle = 0 \text{ and} \\
& \forall n \in \Phi,\ \langle v^c, \phi_{J,n} \rangle = 0 \}.
\end{aligned} \tag{3.3}$$

Assuming this new notation, we define the discrete problem as in Problem 2.1 and thus Theorem 2.2 follows unchanged. From now, Problem 2.1 and Theorem 2.2 will refer to the discrete formulation. Theorem 2.2 leads us to define, as an approximation method of the solution of Problem 2.1, the following subgradient descent scheme with a projection on the constraint:

PROBLEM 3.1.

$$u_{k+1} = P\left(u_k - t_k g_{TV}(u_k)\right), \tag{3.4}$$

where $u_0 \in U$ is the denoised signal by wavelet thresholding obtained from the discrete formulation of (1.6), and where $t_k > 0$ is the step chosen in order to obtain the convergence.

THEOREM 3.2. *Problem 3.1 can be solved using the following algorithm*

$$u_{k+1} = u_k - t_k P_V(g_k) \tag{3.5}$$

where P_V is the orthogonal projection onto V and where $g_k \in X$ satisfies

$$\begin{aligned} \forall n = 2, \ldots N-1, \quad g_k^n &= sgn(u_k^n - u_k^{n-1}) - sgn(u_k^{n+1} - u_k^n), \\ g_k^1 &= -sgn(u_k^2 - u_k^1), \\ g_k^n &= sgn(u_k^n - u_k^{n-1}). \end{aligned} \tag{3.6}$$

Proof. Equation (3.5) comes from the equality

$$\forall u \in X, \ P(u) = u_0 + P_V(u - u_0), \tag{3.7}$$

and from the fact that g_k, as defined in (3.6), is a subgradient of TV at u_k. Indeed, remark that

$$|y| \geq |x| + \operatorname{sgn}(x)(y - x) \ \ (\forall x, y \in \mathbb{R}), \tag{3.8}$$

Let $y = v^{n+1} - v^n$, $x = u_k^{n+1} - u_k^n$, and sum over n. One obtains

$$TV(v) \geq TV(u_k) + \langle g_k, v - u_k \rangle. \tag{3.9}$$

□

The subgradient g_k is projected onto V using the FWT, followed by the cancellation of coefficients belonging to M and by an inverse FWT. If we assume a fixed number of iterations in (3.5), the complexity of the algorithm is of the same order than the complexity of the FWT, that is $O(N)$.

4. Convergence of the algorithm. In this section, we establish the following result

THEOREM 4.1. *Let the sequence (t_k) satisfies*

$$t_k > 0, \ \lim_{k \to +\infty} t_k = 0, \ \sum_{k=0}^{+\infty} t_k = +\infty. \tag{4.1}$$

Then, the algorithm given by Theorem 3.2 converges in the sense that

$$\lim_{k \to +\infty} \min_{u^* \in U^*} ||u_k - u^*|| = 0 \textit{ and } \lim_{k \to +\infty} TV(u_k) = \min_{u \in U} TV(u), \tag{4.2}$$

where

$$U^* = \{u^* \in U \; : \; TV(u^*) = \min_{u \in U} TV(u)\}. \tag{4.3}$$

This theorem is an adaptation of a classical result for the subgradient method, that we recall below (see [22] for a demonstration).

THEOREM 4.2. *Let X be a vectorial space of finite dimension and J a convex function defined on X which has a bounded set of minimum point X^*. Assume that the sequence of positive numbers (t_k) satisfies the conditions (4.1) and suppose that a sequence (x_k) of X is generated according to the formula*

$$x_{k+1} = x_k - t_k g_J(x_k), \tag{4.4}$$

where $g_J(x_k)$ is a subgradient of J at x_k and for $x_0 \in X$ an arbitrary starting point. Then either
(a) the sequence $(g_J(x_k))$ is bounded and the algorithm converges in the sense that

$$\lim_{k \to +\infty} \min_{x^* \in X^*} ||x_k - x^*|| = 0 \text{ and } \lim_{k \to +\infty} J(x_k) = \min_{x \in X} J(x), \tag{4.5}$$

or
(b) the sequence $(g_J(x_k))$ is unbounded and there is no convergence.

The key point is to transform the constrained minimization Problem 2.1 to an unconstrained one, so that Theorem 4.2 can be applied. Since U is an affine space, this is simply done using the obvious lemma

LEMMA 4.3. *u^* is a solution of Problem 2.1 iff $u^* = u_0 + v^*$ where $v^* \in V$ satisfies*

$$J(v^*) = \min_{v \in V} J(v) \quad \text{for} \quad J(v) := TV(v + u_0). \tag{4.6}$$

The corresponding algorithm is given by

LEMMA 4.4. *Let (v_k) be the sequence defined on V by $v_k = u_k - u_0$, the sequence (u_k) being defined as in (3.5). Then,*

$$v_{k+1} = v_k - t_k g_J(v_k), \; v_0 = 0 \in V, \tag{4.7}$$

where $g_J(v_k)$ is a subgradient of J at v_k.

Proof. We have $v_{k+1} = v_k - t_k P_V(g_k)$ and $g_J(v_k) := P_V(g_k)$ is a subgradient of J at v_k. Indeed, from

$$\forall u \in U, \; \mathrm{TV}(u) - \mathrm{TV}(u_k) \geq (g_k, u - u_k) \tag{4.8}$$

one gets, with $v = u - u_0$,

$$\begin{aligned} \forall v \in V, \; J(v) - J(v_k) &\geq (g_k, v - v_k) \\ &\geq (g_k, P_V(v - v_k)) \quad \text{(since } v - v_k \in V) \\ &\geq (P_V(g_k), v - v_k) \quad \text{(since } P_V \text{ is symmetric).} \end{aligned}$$

□

Now, to prove Theorem 4.1, it suffices to show that the sequence (v_k) satisfies the conditions of Theorem 4.2, where one sets $X = V$. This is done by lemmas 4.5 and 4.6 :

LEMMA 4.5. *The set* $V^* := \{v^* \in V \,:\, J(v^*) \leq J(v) \;\forall v \in V\}$ *is bounded and* J *is a convex functional.*

Proof. The functional TV defines a (convex) semi-norm on $\mathbb{R}^N$ and hence on V. Let us prove that for any $v \in V$, we have $TV(v) = 0 \Rightarrow v = 0$. Since linear combination of wavelets $(\psi_{j,n})$ have a zero average and since $\langle v^c, \phi_{J,n}\rangle = 0$, one gets

$$\int_I v^c = \int_I \sum_{(j,n)\in\Psi} \langle v^c, \psi_{j,n}\rangle \psi_{j,n} + \int_I \sum_{n\in\Phi} \langle v^c, \phi_{J,n}\rangle \phi_{J,n} = 0. \tag{4.9}$$

If $TV(v) = 0$, then $v^n = \langle v^c, \phi_{0,n}\rangle = \gamma$ ($\forall n = 1\ldots N$), γ being constant. Since one can write

$$v^c = \sum_{j\geq 0,n} \langle v^c, \psi_{j,n}\rangle \psi_{j,n} + \sum_n \langle v^c, \phi_{0,n}\rangle \phi_{0,n}, \tag{4.10}$$

we get

$$0 = \int_I v^c = \gamma \sum_n \int_I \phi_{0,n} \implies \gamma = 0. \tag{4.11}$$

Therefore, TV is a norm on V which is equivalent to the Euclidean norm. In particular,

$$\exists a > 0 \,:\, \forall v \in V,\; a\,||v|| \leq TV(v). \tag{4.12}$$

Hence,

$$\forall v \in V,\; J(v) = TV(u_0 + v) \geq TV(v) - TV(u_0) \geq a\,||v|| - TV(u_0),$$

and

$$\forall v^* \in V^*, ||v^*|| \leq \left(\min_{v\in V} J(v) + TV(u_0)\right)/a.$$

□

LEMMA 4.6. *The sequence* $(g_J(v_k))$ *is bounded.*

Proof. Since P_V is a continuous operator, it suffices to prove that the sequence (g_k) is bounded. This is obviously true from (3.6). □

5. Numerical results. The algorithm described in Section 3 has been implemented using the tools given by the free and open-source MegaWave2 software [16]. A forthcoming update of this software will contain all the modules used to obtain the following experiments, so that everyone will be able to reproduce the results and to perform additional experiments.

We report two experiments. The first one consists in denoising a synthetic signal containing two discontinuities of second order (a ramp), followed by a sharp discontinuity (a step), followed by a peak. A Gaussian white noise has been added following the model $\tilde{u} = u + \epsilon$. Figure 5.1 displays the original noiseless signal u, Figure 5.2 the noisy signal $\tilde{u}$, Figure 5.3 the wavelet-denoised signal u_0 and Figure 5.4 the restored signal u_k for $k = 10000$. The estimator u_k is far better than u_0, either in terms of SNR (Signal to Noise Ratio) or visually. Fair results are still obtained with much lower k (as low as $k \simeq 10$).

Let us now compare our result with three standard methods. On Figure 5.5, the signal $\tilde{u}$ has been denoised by Rudin-Osher-Fatemi's variational method, exhibiting the staircase phenomena. Then, in the latter method, the Total Variation has been replaced by the regularized-TV J_β given in (1.11), with $\beta = 10^{-4}$. The obtained result is displayed on Figure 5.6. The data fidelity weight has been chosen large enough in order not to smooth the jump, but, as a consequence, the noise is diffused. Finally, Figure 5.7 shows the signal $\tilde{u}$ denoised by Coifman and Donoho's translation invariant wavelet thresholding algorithm [9]. This procedure, called *SpinCycle*, consists in applying the thresholding process to translated versions of the original signal and averaging them. The pseudo-Gibbs phenomena have been reduced compared to the standard wavelet thresholding (Figure 5.4), but they are still visible.

The second experiment is obtained from a natural noisy signal $\tilde{u}$, which follows our assumption of a piecewise smooth noiseless signal u. The signal $\tilde{u}$ shows in Figure 5.8 corresponds to a line of a digital image, which is a snapshot of an office. Figure 5.9 displays the signal u_0 and Figure 5.10 the signal u_k. Once again, the visual aspect of u_k is far better than u_0.

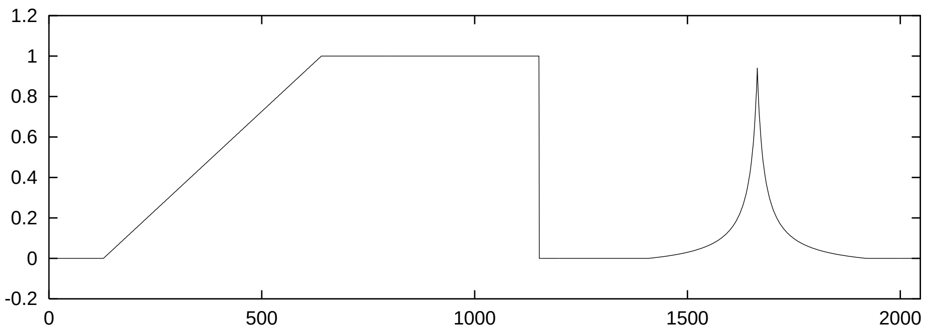

FIG. 5.1. *Original function u.*

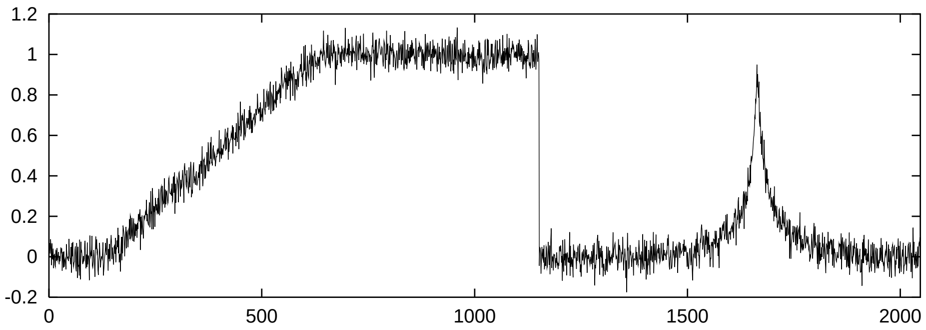

FIG. 5.2. *Noisy function $\tilde{u}$, obtained by adding to u a Gaussian white noise ϵ of $\sigma = 0.05$. SNR=18.7 db.*

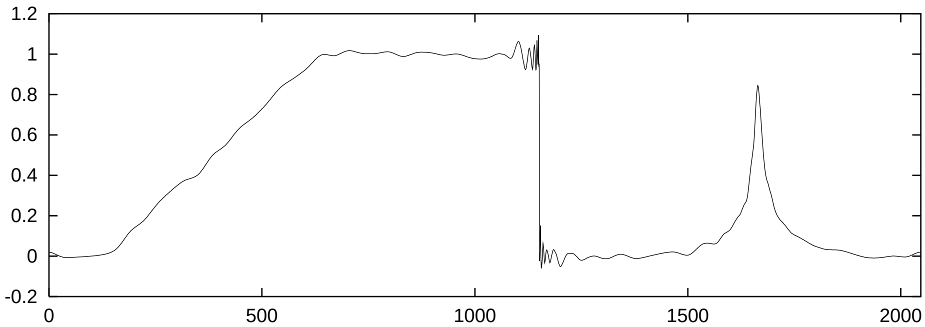

FIG. 5.3. *Denoised function* u_0 *, obtained by wavelet hard thresholding. SNR*=29.0 *db. NTV (Normalized Total Variation)*=0.0031*.*

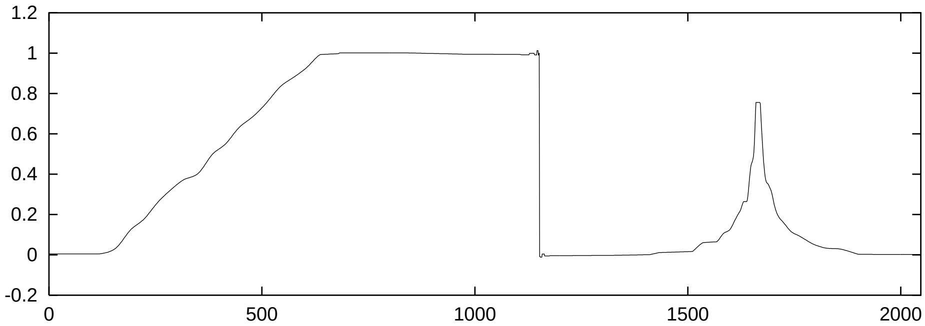

FIG. 5.4. *Denoised function* u_k*, obtained by our method. SNR*=31.1 *db. NTV*=0.0018*.*

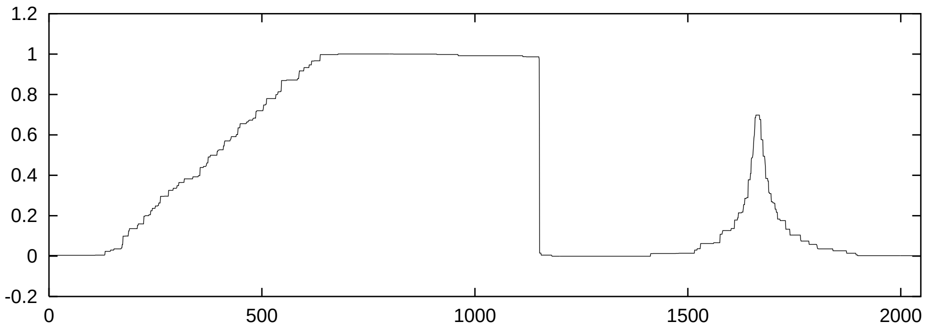

FIG. 5.5. *Rudin-Osher-Fatemi's algorithm : denoised function* $argminTV(v)$ *subject to* $||v - \tilde{u}||_{l^2} = \sigma$*. SNR*=29.3 *db.*

6. Concluding remarks. We have presented a method to reconstruct wavelet coefficients using a total variation minimization algorithm. This approach performs a nearly artifact free signal denoising : the pseudo-Gibbs phenomena vanish almost totally while the sharpness of the signal is preserved, without staircase effect.

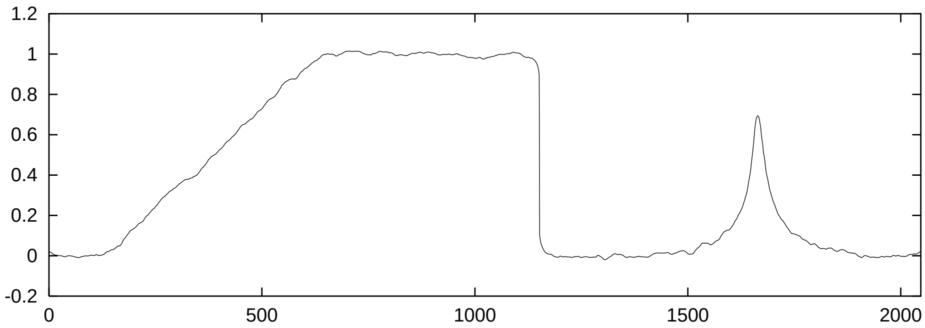

FIG. 5.6. *Regularized Total Variation algorithm : denoised function* $argmin J_\beta(v)$ *subject to* $||v - \tilde{u}||_{l^2} = \sigma$. *SNR*=28.7 *db.*

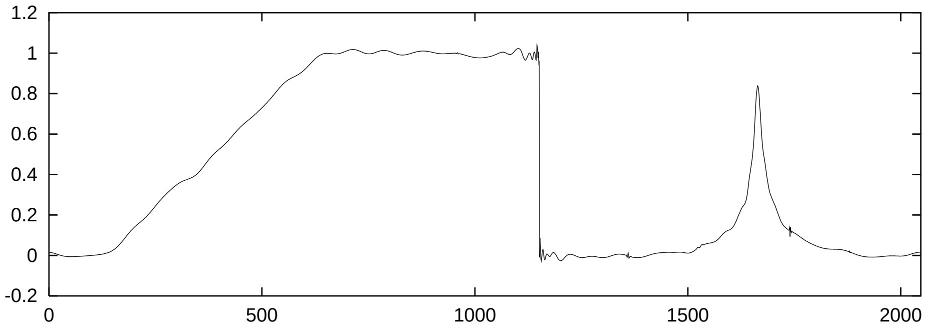

FIG. 5.7. *SpinCycle : denoised function, obtained by translation invariant wavelet thresholding. SNR*=31.4 *db.*

However, better results may be obtained by improving our algorithm. Observe that the peak on the right side of Figure 5.4 has been slightly eroded compared to the original one. This drawback is shared by all TV regularization approaches; in our case it may be lowered by keeping small coefficients vanished when they are not in the vicinity of jumps. In this way, a standard wavelet denoising would be applied on regular parts. Another subject of research would be to introduce a weighted TV functional, so that the regularization would be relaxed in transient parts and reinforced in smooth ones, following the idea recently proposed by L. Moisan in [19] in the context of spectral extrapolation.

Let us emphasize that, although our algorithm was justified and illustrated in the case of unidimensional signals, it can be easily extended to signals of higher dimensions, and in particular to images for which the piecewise-smooth assumption is highly relevant. A slight modification of the constraint may also be performed in order to achieve restoration of signals and images that have been compressed within an orthogonal basis, although in such case care should be taken in order to keep the convergence of the algorithm.

Acknowledgments. The authors are grateful to Lionel Moisan, Mila Nikolova and Bernard Rougé for useful discussions.

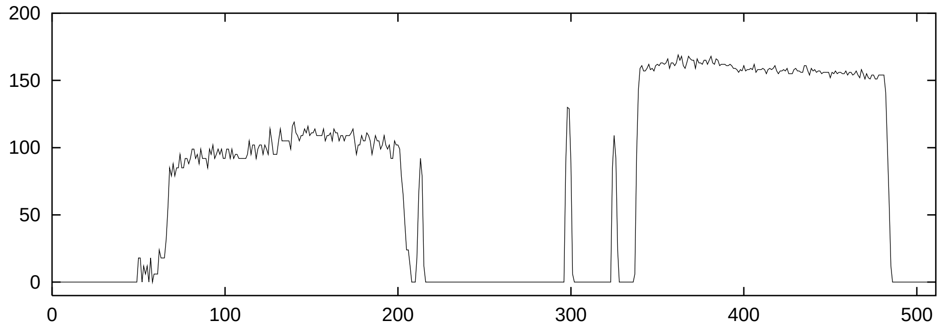

Fig. 5.8. *Real signal u extracted from a line of a digital image (view of an office).*

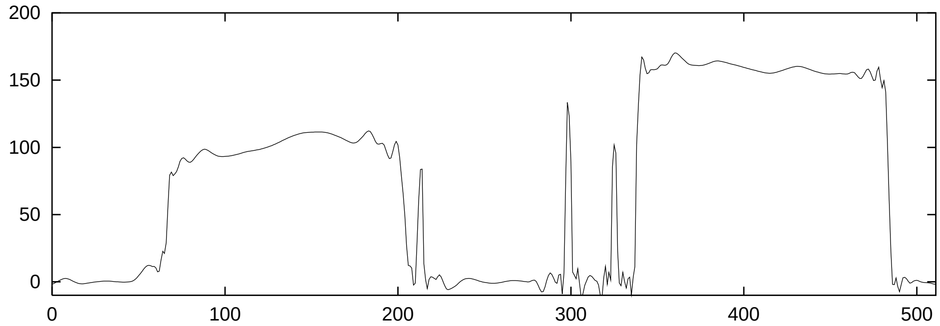

Fig. 5.9. *Denoised line of the image, obtained by wavelet hard thresholding. NTV*=3.244.

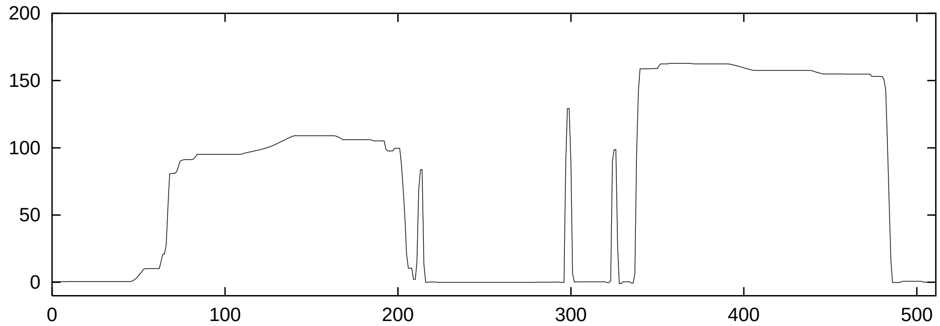

Fig. 5.10. *Denoised line of the image, obtained by our method. NTV*=2.365.

REFERENCES

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

[2] A. Chambolle and P-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76:167–188, 1997.

[3] T.F. Chan, H.M. Zhou, and R.H. Chan. A continuation method for total variation denoising problems. UCLA CAM Report 95-18, University of California, Losa Angeles, CA, 1995.

[4] Tony Chan, Antonio Marquina, and Pep Mulet. High-order total variation-based image restoration. *SIAM J. Sci. Comput.*, 22(2):503–516, 2000.

[5] Tony F. Chan and H.M. Zhou. Total variation improved wavelet thresholding in image compression. In *Proc. of ICIP'2000*, 2000.

[6] A. Cohen and J.P. D'Alès. Nonlinear approximation of random functions. *SIAM J. Appl. Math.*, 57(2):518–540, 1997.

[7] A. Cohen, I. Daubechies, and P. Vial. Wavelets and fast wavelet transforms on an interval. *Applied and Comput. Harmonic Ana.*, 1:54–81, 1993.

[8] A. Cohen, J. Froment, and J. Istas. Analyse multirsolution des signaux alatoires. *Note aux C.R. Acad. Sci. Paris*, t.312(I):567–570, 1991. Analyse mathématique.

[9] R.R. Coifman and D.L. Donoho. Translation-invariant de-noising. In A. Antoniadis and G. Oppenheim, editors, *Wavelets and Statistics*, pages 125–150. Springer-Verlag, 1995.

[10] D. Dobson and F. Santosa. Recovery of blocky images from noisy and blurred data. *SIAM J. Appl. Math.*, 56:1181–1199, 1996.

[11] David C. Dobson and Curtis R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM J. Numer. Anal.*, 34(5):1779–1791, 1997.

[12] D. Donoho. De-noising by soft-thresholding. *IEEE Trans. on Info. Theory*, 41:613–627, 1995.

[13] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81:425–455, 1994.

[14] S. Durand and J. Froment. Artifact free signal denoising with wavelets. In *Proc. of ICASSP'01*, volume 6, 2001.

[15] S. Durand, F. Malgouyres, and B. Rougé. Image deblurring, spectrum interpolation and application to satellite imaging. *ESAIM:COCV Control, Opt. and Cal. of Var.*, 5:445–475, 2000.

[16] J. Froment and L. Moisan (ed.). Megawave2 v.2.01. A free and open-source Unix image processing software for reproducible research, available at `http://www.cmla.ens-cachan.fr`, 2001.

[17] F. Malgouyres. Combining total variation and wavelet packet approaches for image deblurring. In *Proc. of VLSM 2001*, 2001.

[18] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1997.

[19] Lionel Moisan. Extrapolation de spectre et variation totale pondérée. In *Proc. of Gretsi'01, Toulouse, France*, 2001.

[20] Mila Nikolova. Local strong homogeneity of a regularized estimator. *SIAM J. Appl. Math.*, 61(2):633–658, 2000.

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

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

[23] C.R. Vogel and M.E. Oman. Iterative method for total variation denoising. *SIAM J. Sci. Comput.*, 17(1):227–238, 1996.