

## RECONSTRUCTION OF WAVELET COEFFICIENTS USING TOTAL VARIATION MINIMIZATION\*

SYLVAIN DURAND<sup>†</sup> AND JACQUES FROMENT<sup>‡</sup>

**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 to 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.** 26A45, 65K10, 65T60, 94A12

**PII.** S1064827501397792

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

$$(1.1) \quad \tilde{u} = u + \epsilon,$$

where  $\tilde{u}$  represents the observed data,  $u$  is the noiseless function to estimate, and  $\epsilon$  is the noise or, more generally, an unknown error. This problem occurs in a 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 appeared: wavelet thresholding and total variation minimization.

Wavelet denoising consists of decomposing the noisy data into an orthogonal wavelet basis, suppressing the wavelet coefficients smaller than a given amplitude using a so-called soft or hard thresholding, and transforming the data back into the original domain: Let  $\{\{\psi_{j,n}\}_{(j,n) \in \Psi}, \{\phi_{J,n}\}_{n \in \Phi}\}$  be an orthogonal basis of wavelets and scaling functions on the interval  $I = [a, b]$  as described by Cohen, Daubechies, and Vial [8], so that we can write any function  $\tilde{u} \in L^2(I)$  as the sum of the series

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

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

$$(1.3) \quad \langle u, v \rangle = \int_I u(x) v(x) dx.$$

---

\*Received by the editors November 12, 2001; accepted for publication (in revised form) October 18, 2002; published electronically April 15, 2003.

<http://www.siam.org/journals/sisc/24-5/39779.html>

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

<sup>‡</sup>LMAM, Université de Bretagne Sud, Centre Y. Coppens, campus de Tohannic, BP 573, 56017 Vannes, France (Jacques.Froment@univ-ubs.fr) and CMLA, Ecole Normale Supérieure de Cachan, 61 avenue du Président Wilson, 94235 Cachan cedex, France.

In  $\psi_{j,n}$ ,  $j$  is the index of scale and  $n$  the translation factor, taking values in the countable set  $\Psi$ . The term  $\phi_{J,n}$  denotes the scaling function on scale  $2^J$  translated by  $n$ , which is indexed to a finite set  $\Phi$ . 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  $\tau$  is defined by

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

while in the case of soft thresholding, the operator  $\tau$  is

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

The denoised signal using wavelet thresholding is simply

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

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

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

Because of its simplicity, the algorithm sketched 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 [15], where they proved that the performance associated with the nonlinear thresholding estimator in orthogonal bases is close to an ideal coefficient selection and attenuation. In addition, among classical orthogonal bases, wavelet series outperform Fourier or cosine series in the representation of piecewise smooth functions (see, e.g., [7, 9]): 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 than the decay of Fourier coefficients. However, the wavelet thresholding method is still a regularization process and the estimator presents oscillations in the vicinity of the 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  $\epsilon$  is a Gaussian white noise of standard deviation  $\sigma$ , the threshold should be set to

$$(1.8) \quad \lambda = \sigma \sqrt{2 \log N},$$

$N$  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 [15]. The use of the soft thresholding operator (1.5) instead of the more intuitive hard one (1.4) allows us to partially reduce the pseudo-Gibbs phenomena [14]: 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 to peaks being eroded.

Total variation denoising follows a completely different approach. It has been introduced for the first time by Rudin, Osher, and Fatemi in [23] in the context of image denoising. They proposed minimizing the total variation (TV) of a function  $u \in L^2(\Omega)$ ,

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

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

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

$e$  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 an artifact as well, known as the staircase effect (see [5, 12] for numerical evidence and [22] for theoretical analysis in a general framework): TV-based algorithms tend to restore piecewise constant functions. For example, and though they have the same TV, a staircase will be preferred to a ramp (see Figure 5). This is mainly due to the lack of regularity of the TV.

In [25] (see also [1, 4, 13]), Vogel and Oman propose replacing the TV by the regularized functional

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

where  $\beta$  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  $\beta$ : if  $\beta$  is significantly greater than 0, the staircase effect is effectively removed but the noise is diffused, and for  $\beta$  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., [3, 5].

The denoising algorithm we are presenting in this article combines the wavelet and the TV approaches. To our knowledge, only a few articles use these complementary tools in the context of signal or image processing. In [17], Durand, Malgouyres, and Rougé point out the complementarity of these tools and in [19] Malgouyres proposed a deblurring algorithm based on both tools, in the context of satellite imaging. Another close approach is proposed by Chan and Zhou in [6]; let us state that our algorithm was developed independently. As we do, Chan and Zhou noticed that a TV minimization may be applied to remove the pseudo-Gibbs phenomenon generated by wavelet thresholding. They proposed to solve the following convex and unconstrained problem:

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

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 TV is diminished. The second term is a classical  $L^2$  fitting term. The regularization parameter  $\alpha$  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. Whereas our article is concerned with signal denoising, [6] deals with image compression and therefore their algorithm does not touch zero wavelet coefficients. We propose to reconstruct a function with minimal TV such that for indexes belonging to  $M$ , its wavelet coefficients are the same as the wavelet coefficients of the observed function  $\tilde{u}$ . That is, retained wavelet coefficients are not modified while canceled coefficients are set to 0 no more, but rather to values that minimize the TV. 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 as the thresholded one in the location given by the map  $M$ . The 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 TV of the reconstructed function, occurrence of oscillations is discouraged: a structure of coefficients compatible with sharp discontinuities is recovered. Notice that, in contrast to the model of Chan and Zhou, we define a constrained problem which is free of a regularization parameter. Since the important wavelet coefficients are kept unchanged, there is no need for a fitting term. The only parameter is the threshold  $\lambda$ , and it can be fixed according to (1.8). However, in the following we will assume  $\lambda$  set in order to obtain, after the wavelet thresholding, a free of noise signal. The aim of our TV minimization is therefore not to denoise the signal, but to remove the pseudo-Gibbs phenomena due to wavelet thresholding.

While our manuscript was under review, we discovered the work of Coifman and Sowa [11], which proposed, based on similar ideas, a more general method to combine the calculus of variations and wavelets for image enhancement. This method appears to be a generalization of the one introduced by Bobichon and Bijaoui in [2]. The article [11] differs from ours in the following way: Coifman and Sowa seem to be more concerned with regularizing the restored signal rather than with removing the pseudo-Gibbs phenomena, as they prefer for their algorithms the use of a Sobolev norm instead of the TV. In addition, they do not address the problem of the convergence to the minimizer.

We have proposed our denoising model in [16], 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 one dimensional (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 TV 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 TV of any unidimensional function  $u$  in  $I$  is defined by

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

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

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

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

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

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

$$(2.4) \quad U = \{u_0\} + V.$$

We propose to solve the variational problem.

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

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 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  $g_{\text{TV}}(u)$  satisfying

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

and we get the classical result.

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

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

for  $P$  the affine projector onto  $U$  that minimizes the distance:

$$(2.7) \quad \forall v \in X, P(v) = \arg \min_{u \in U} \|u - v\|.$$

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

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

**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  $u = (u^1, u^2, \dots, u^N)$ , we denote by  $(\cdot, \cdot)$  the standard scalar product on  $\mathbb{R}^N$ ,

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

and  $\|\cdot\|$  is the associated Euclidean norm. The discrete TV of  $u$  is given by

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

Let  $u^c$  be the continuous, or analog, function in the approximation space  $V_0$  (scale 1) associated with 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 [20]) 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  $(\langle u^c, \phi_{J,n} \rangle)_{n \in \Phi}$ :

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

Assuming this new notation, we define the discrete problem as in Problem 2.1 and thus Theorem 2.2 follows unchanged. From now on, 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.

$$(3.4) \quad u_{k+1} = P(u_k - t_k g_{\text{TV}}(u_k)),$$

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 algorithm*

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

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

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

*Proof.* First observe that  $g_k$ , as defined in (3.6), is a subgradient of TV at  $u_k$ . Indeed, we have

$$(3.7) \quad |y| \geq |x| + \text{sgn}(x)(y - x) \quad (\forall x, y \in \mathbb{R}).$$

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

$$(3.8) \quad \text{TV}(v) \geq \text{TV}(u_k) + \langle g_k, v - u_k \rangle.$$

Now, (3.5) comes from the equality

$$(3.9) \quad \forall u \in X, P(u) = u_0 + P_V(u - u_0),$$

with  $u = u_k - t_k g_k$ :  $u_{k+1} = P(u) = u_0 + P_V(u_k - u_0) - t_k P_V(g_k)$  and  $P_V(u_k - u_0) = u_k - u_0$  since  $u_k - u_0 \in V$ .  $\square$

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:

$$(3.10) \quad P_V(g_k) = \sum_{(j,n) \notin M} \langle g_k, \psi_{j,n} \rangle \psi_{j,n}.$$

If we denote by  $K$  the number of iterations in (3.5), the complexity of the algorithm is  $O(KN)$  since the complexity of the FWT is  $O(N)$ . However,  $K$  may be considered as fixed, independent of the signal size  $N$ . Indeed, as  $t_k$  is not chosen to be the optimal stepsize, it does not depend on the signal and therefore the computation of  $u_{k+1}$  from  $u_k$  remains local. (We are using compactly supported wavelets with a limited number of scales.)

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

**THEOREM 4.1.** *Let the sequence  $(t_k)$  satisfy*

$$(4.1) \quad t_k > 0, \quad \lim_{k \rightarrow +\infty} t_k = 0, \quad \sum_{k=0}^{+\infty} t_k = +\infty.$$

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

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

where

$$(4.3) \quad U^* = \left\{ u^* \in U : \text{TV}(u^*) = \min_{u \in U} \text{TV}(u) \right\}.$$

This theorem is an adaptation of a classical result for the subgradient method, which we recall below (see [24] for a discussion).

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

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

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*

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

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 into 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

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

The corresponding algorithm is given by the following lemma.

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,

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

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

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

$$(4.8) \quad \forall u \in U, \text{TV}(u) - \text{TV}(u_k) \geq (g_k, u - u_k)$$

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

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

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

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

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

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

we get

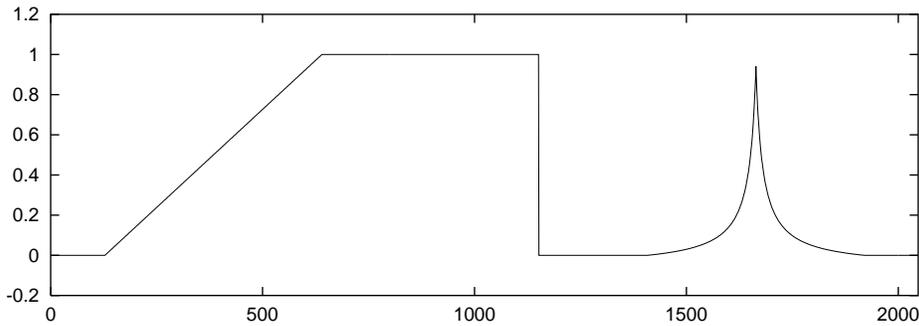
$$(4.11) \quad 0 = \int_I v^c = \gamma \sum_n \int_I \phi_{0,n} \implies \gamma = 0.$$

Therefore,  $\text{TV}$  is a norm on  $V$  which is equivalent to the Euclidean norm. In particular,

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

Hence,

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

FIG. 1. Original function  $u$ .

and

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

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).  $\square$

**5. Numerical results.** The algorithm described in section 3 has been implemented using the tools given by the free and open-source MegaWave2 software [18]. 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. Both of them have been performed using Daubechies' filter with 16 coefficients (8 vanishing moments) and 3 levels of scales. The stepsize  $(t_k)$  has been chosen linearly decreasing with  $1/k$  in order to satisfy (4.1).

The first one consists of 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 1 displays the original noiseless signal  $u$ , Figure 2 the noisy signal  $\tilde{u}$ , Figure 3 the wavelet-denoised signal  $u_0$ , and Figure 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 results with three standard methods. On Figure 5, the signal  $\tilde{u}$  has been denoised by Rudin, Osher, and Fatemi's variational method, exhibiting the staircase phenomena. Then, in the latter method, the TV has been replaced by the regularized-TV  $J_\beta$  given in (1.11), with  $\beta = 10^{-4}$ . The obtained result is displayed in Figure 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 7 shows the signal  $\tilde{u}$  denoised by Coifman and Donoho's translation invariant wavelet thresholding algorithm [10]. This procedure, called *SpinCycle*, consists of 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 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  $\tilde{u}$  shown in Figure 8

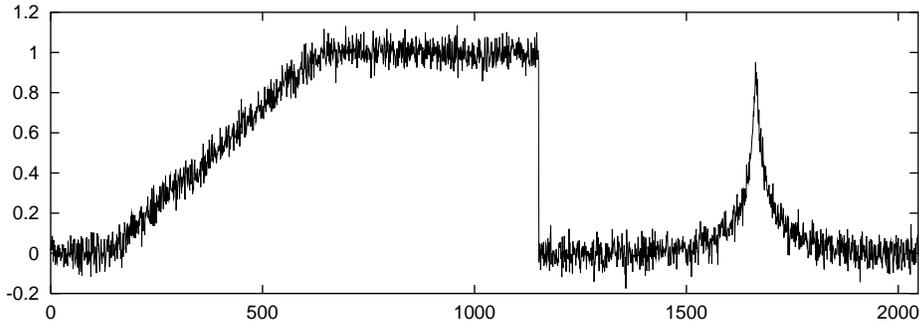


FIG. 2. Noisy function  $\tilde{u}$ , obtained by adding to  $u$  a Gaussian white noise  $\epsilon$  of  $\sigma = 0.05$ .  $SNR = 18.7$  db.

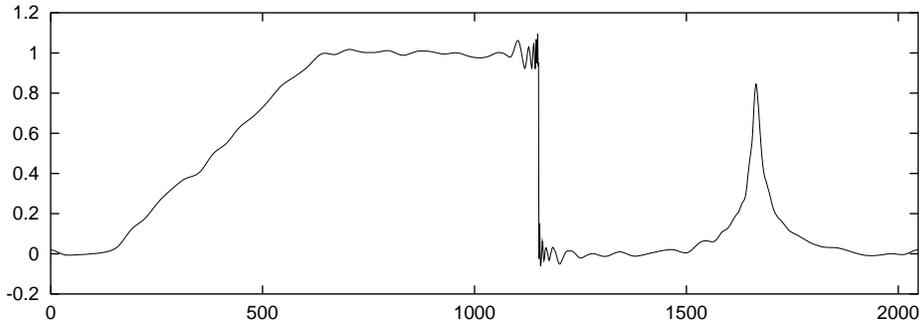


FIG. 3. Denoised function  $u_0$ , obtained by wavelet hard thresholding.  $SNR = 29.0$  db.  $NTV$  (normalized total variation)=0.0031.

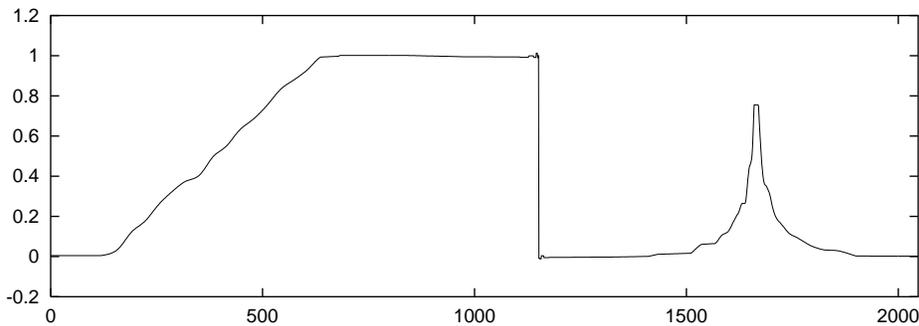


FIG. 4. Denoised function  $u_k$ , obtained by our method.  $SNR = 31.1$  db.  $NTV = 0.0018$ .

corresponds to a line of a digital image, which is a snapshot of an office. Figure 9 displays the signal  $u_0$  and Figure 10 the signal  $u_k$ . Once again, the visual aspect of  $u_k$  is far better than  $u_0$ .

**6. Concluding remarks.** We have presented a method to reconstruct wavelet coefficients using a TV minimization algorithm. This approach performs a nearly

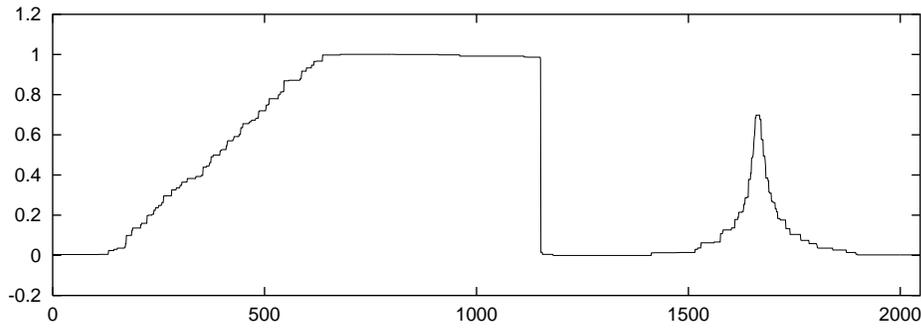


FIG. 5. *Rudin, Osher, and Fatemi's algorithm: denoised function  $\text{argminTV}(v)$  subject to  $\|v - \tilde{u}\|_2 = e$ .  $\text{SNR} = 29.3$  db.*

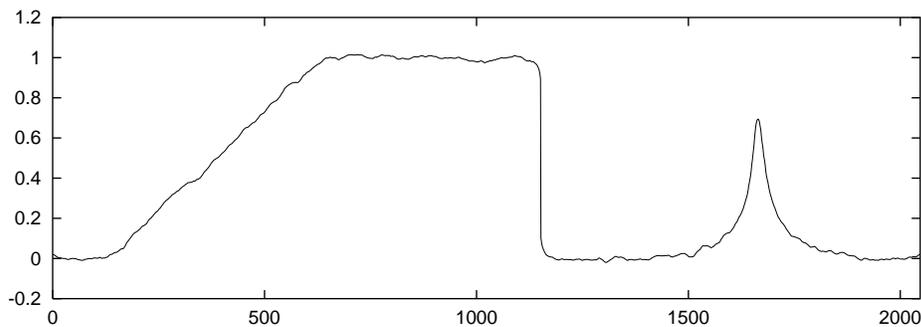


FIG. 6. *Regularized TV algorithm: denoised function  $\text{argmin}J_\beta(v)$  subject to  $\|v - \tilde{u}\|_2 = e$ .  $\text{SNR} = 28.7$  db.*

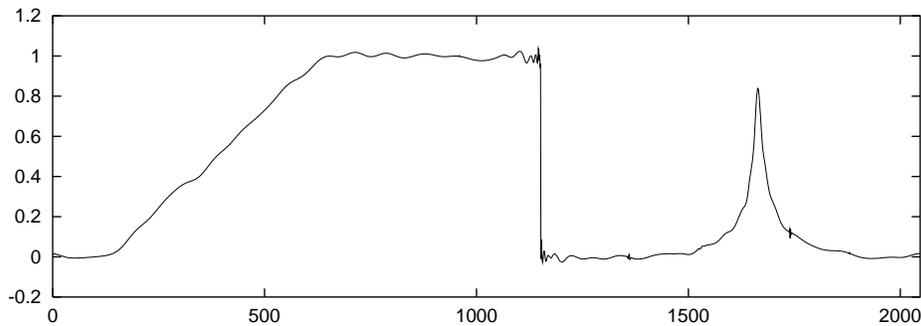


FIG. 7. *SpinCycle: denoised function, obtained by translation invariant wavelet thresholding.  $\text{SNR} = 31.4$  db.*

artifact-free signal denoising: the pseudo-Gibbs phenomena vanish almost totally while the sharpness of the signal is preserved, without the staircase effect.

Observe that the solution to our problem is not always unique. Indeed, two nondecreasing (or nonincreasing) functions  $u$  and  $v$  such that  $u(a) = v(a)$  and  $u(b) =$

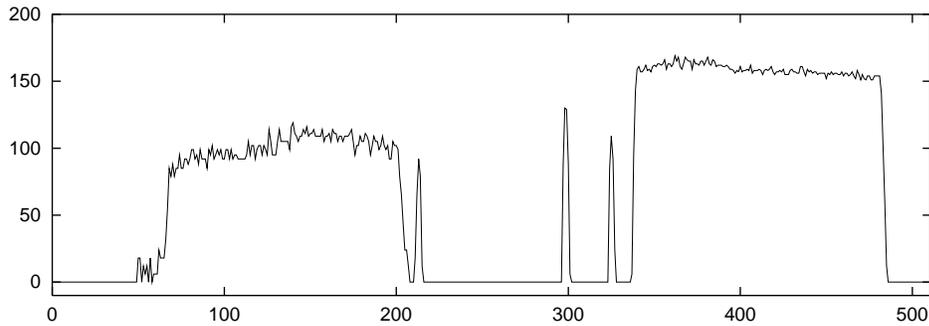


FIG. 8. Real signal  $u$  extracted from a line of a digital image (view of an office).

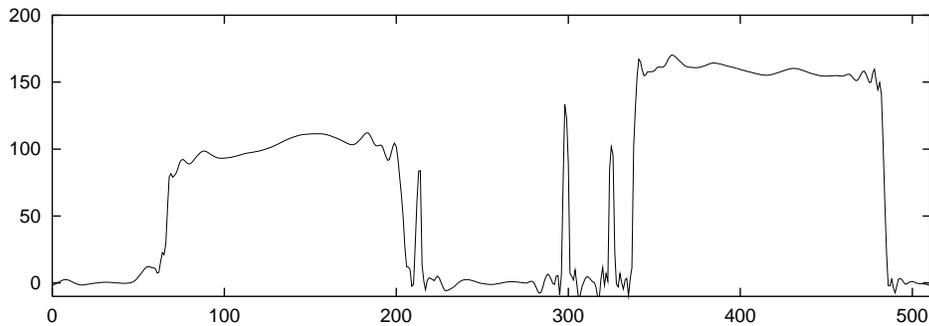


FIG. 9. Denoised line of the image, obtained by wavelet hard thresholding.  $NTV = 3.244$ .

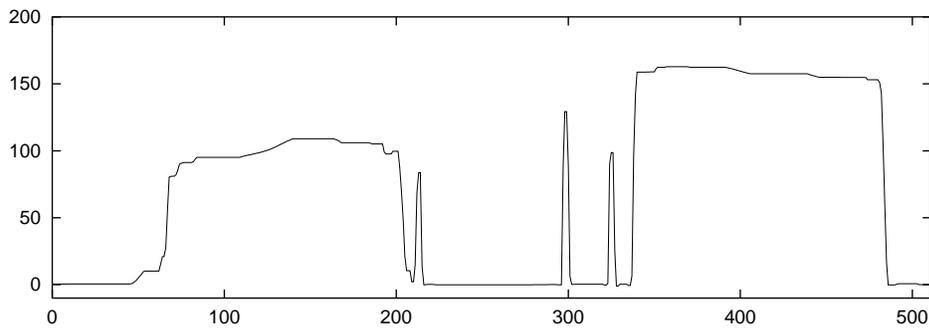


FIG. 10. Denoised line of the image, obtained by our method.  $NTV = 2.365$ .

$v(b)$  have the same TV. In such a case (see for instance the ramp in Figure 3), one may modify a small wavelet coefficient without changing the TV and the constraint. Thanks to the nonuniqueness of the solution, our algorithm avoids staircasing in regular regions. These regions are well restored by wavelet thresholding, and under the nondecreasing (or nonincreasing) assumption, they have already minimal TV. This shows the significance of the initial function  $u_0$ . Taking for  $u_0$  the noisy signal instead of the thresholded one will lead to staircase effects.

However, better results may be obtained by improving our algorithm. Notice that there is no guarantee that the reconstructed signal will be close to the original one. Whereas some structures may be removed, no new structure can appear thanks to the TV minimization. Observe that the peak on the right side of Figure 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. Indeed, in the location where TV is minimized, the resulting correlation of wavelet coefficients leads to an opposite peak, so that the original one is eroded (see Figure 3 where small wavelet coefficients are kept vanished while the peak is not eroded). By keeping small wavelet coefficients vanished, 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 Moisan in [21] 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 to keep the convergence of the algorithm.

**Acknowledgments.** The authors are grateful to Lionel Moisan, Mila Nikolova, and Bernard Rougé for useful discussions. Thanks also to François Malgouyres for pointing out the work of Coifman and Sowa.

#### REFERENCES

- [1] R. ACAR AND C.R. VOGEL, *Analysis of total variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.
- [2] Y. BOBICHON AND A. BIJAOU, *Regularized multiresolution methods for astronomical image enhancement*, Exper. Astron., 7 (1997), pp. 239–255.
- [3] A. CHAMBOLLE AND P.-L. LIONS, *Image recovery via total variation minimization and related problems*, Numer. Math., 76 (1997), pp. 167–188.
- [4] 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, Los Angeles, CA, 1995.
- [5] T. CHAN, A. MARQUINA, AND P. MULET, *High-order total variation-based image restoration*, SIAM J. Sci. Comput., 22 (2000), pp. 503–516.
- [6] T.F. CHAN AND H.M. ZHOU, *Total variation improved wavelet thresholding in image compression*, in Proceedings of the 2000 International Conference on Image Processing, Vol. 2, IEEE Press, Piscataway, NJ, 2000, pp. 391–394.
- [7] A. COHEN AND J.-P. D’ALES, *Nonlinear approximation of random functions*, SIAM J. Appl. Math., 57 (1997), pp. 518–540.
- [8] A. COHEN, I. DAUBECHIES, AND P. VIAL, *Wavelets and fast wavelet transforms on an interval*, Appl. Comput. Harmon. Anal., 1 (1993), pp. 54–81.
- [9] A. COHEN, J. FROMENT, AND J. ISTAS, *Analyse multirésolution des signaux aléatoires*, C. R. Acad. Sci. Paris Sér. I Math, 312 (1991), pp. 567–570.
- [10] R.R. COIFMAN AND D.L. DONOHO, *Translation-invariant de-noising*, in Wavelets and Statistics, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, 1995, pp. 125–150.
- [11] R.R. COIFMAN AND A. SOWA, *Combining the calculus of variations and wavelets for image enhancement*, Appl. Comput. Harmon. Anal., 9 (2000), pp. 1–18.
- [12] D.C. DOBSON AND F. SANTOSA, *Recovery of blocky images from noisy and blurred data*, SIAM J. Appl. Math., 56 (1996), pp. 1181–1198.
- [13] D.C. DOBSON AND C.R. VOGEL, *Convergence of an iterative method for total variation denoising*, SIAM J. Numer. Anal., 34 (1997), pp. 1779–1791.

- [14] D. DONOHO, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–627.
- [15] D. DONOHO AND I. JOHNSTONE, *Ideal spatial adaptation via wavelet shrinkage*, Biometrika, 81 (1994), pp. 425–455.
- [16] S. DURAND AND J. FROMENT, *Artifact free signal denoising with wavelets*, in Proceedings of the 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 6, IEEE Press, Piscataway, NJ, 2001, pp. 3685–3688.
- [17] S. DURAND, F. MALGOUYRES, AND B. ROUGÉ, *Image deblurring, spectrum interpolation and application to satellite imaging*, ESAIM. Control Optim. Cal. Var., 5 (2000), pp. 445–475.
- [18] J. FROMENT AND L. MOISAN, EDS., *MegaWave2 v.2.10*, <http://www.cmla.ens-cachan.fr/Cmla/Megawave/index.html> (17 December 2002).
- [19] F. MALGOUYRES, *Combining total variation and wavelet packet approaches for image deblurring*, in Proceedings of the IEEE Workshop on Variational and Level Set Methods, IEEE Press, Piscataway, NJ, 2001, pp. 57–64.
- [20] S. MALLAT, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1997.
- [21] L. MOISAN, *Extrapolation de spectre et variation totale pondérée*, in Proceedings of the 12th Symposium GRETSI'01 on Signal and Image Processing, Toulouse, France, 2001.
- [22] M. NIKOLOVA, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658.
- [23] L.I. RUDIN, S. OSHER, AND E. FATEMI, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60 (1992), pp. 259–268.
- [24] N.Z. SHOR, *Minimization Methods for Non-Differentiable Functions*, Springer-Verlag, Berlin, 1985.
- [25] C.R. VOGEL AND M.E. OMAN, *Iterative method for total variation denoising*, SIAM J. Sci. Comput., 17 (1996), pp. 227–238.