Minimizing the total variation under a general convex constraint for image restoration

F. Malgouyres

Abstract— **In this paper, we present a general framework for image restoration; despite its simplicity, certain variational and certain wavelet approaches can be formulated within this framework. This permits the construction of a natural model, with only one parameter, which has the advantages of both approaches. We give a mathematical analysis of this model, describe our algorithm and illustrate this by some experiments.**

Keywords— **image, signal, restoration, wavelet, wavelet packet, total variation, deblurring, deconvolution, Maximum A Posteriori.**

I. Introduction

This paper is concerned with image and signal restoration; more precisely, we consider two classical methods (the wavelet thresholding and the Rudin-Osher-Fatemi (ROF) method) in a unified framework. In this framework, it appears that wavelet methods focus mainly upon the data fidelity term, while variational methods are more concerned with the regularity criterion. This leads us to propose a new method which combines the advantages of both.

In the following, we will understand "restoration" as methods whose aim is to recover an image (similarly a signal) $u \in L^2(\mathbb{T})$, from data

$$v = H(u) + b\,,$$

where $\mathbb{T}$ is the torus (the periodization of $[0,1[^2$), H is a continuous linear operator which goes from $L^2(\mathbb{T})$ into itself and b is a Gaussian noise of standard deviation σ.

There is a large number of papers which consider this problem. Among these papers there are two opposing "families": the wavelet and the variational approaches. Among the variational approaches, those based on the minimization of the total variation, as introduced in [1], are often considered as being the most efficient (see [2], [3], [4], [5], [6], [7]). On the other hand, wavelet soft thresholding methods were introduced by Donoho and Johnstone and have been studied and extended in several papers (see [8], [9], [10], [11], [12], [13]).

The paper is organized as follows:

- In section II, we describe the unified framework and propose the model.
- In section III, we state the mathematical results which guarantee that the proposed model has a computable solution.
- In section IV, we describe the algorithm used to compute a solution.
- In section V, we illustrate the algorithm by some experiments.

LAGA, Université Paris 13, 99, avenue Jean-Baptiste Clément, 93430 Villetaneuse, FRANCE; Email: malgouy@math.univ-paris13.fr; http://zeus.math.univ-paris13.fr/~malgouy/ .

I started this work while I participated to the "Geometrically Based Motion" program of the "Institute for Pure and Applied Mathematics", it was continued while I was supported by the grant ONR-N00014-97-1-0027, and was finished while I was at the Université Paris 13.

II. A unified framework

We describe here a framework into which many restoration methods fit. Amongst these are Maximum A Posteriori methods (such as the ROF method) and wavelet approaches (such as the wavelet soft-thresholding). The model is very simple and describes a method in terms of the minimization of an energy $E(w)$ under a convex constraint. The constraint forces the residual $(H(w) - v)$ to belong to a set $\mathcal{N}_{\mathcal{D},\tau}$ defined by

$$\mathcal{N}_{\mathcal{D},\tau} = \{w \in L^2(\mathbb{T}), \forall \Psi \in \mathcal{D}, |\langle w, \Psi\rangle| \leq \tau\},$$

for a dictionary $\mathcal{D} \subset L^2(\mathbb{T})$.

The proposed constraint is simply a convex constraint (defined by its envelope); several data fidelity terms used in image restoration can be expressed under this form (for instance, some of those proposed in [14]). The result depends strongly upon both the definition of the energy and the content of the dictionary $\mathcal{D}$. Let us explain this through two examples: the wavelet soft-thresholding and the ROF model.

The wavelet soft-thresholding method:

For simplicity, we consider only the case of denoising in a wavelet basis; one can refer to [12], [11] for the deblurring case. The usual wavelet soft-thresholding $W(v)$ (see [9]) can be rewritten in the above framework by letting $\mathcal{D} = \{\Psi_l\}$, an orthonormal wavelet basis, and by minimizing any energy of the kind[1]

$$E(w) = \sum_l f_l\left(|\langle w, \Psi_l\rangle|\right) \tag{1}$$

for increasing functions f_l.

This is indeed clear since the latter problem is equivalent to independently solving, for all l (the wavelet basis is orthonormal),

$$\begin{cases} \text{Minimize } f_l\left(|\langle w, \Psi_l\rangle|\right), \\ \text{under the constraint } \langle v, \Psi_l\rangle - \tau \leq \langle w, \Psi_l\rangle \leq \langle v, \Psi_l\rangle + \tau; \end{cases}$$

the solutions of which are the soft-thresholded wavelet coefficients.

[1] This has already been observed in [15], in the case of an l^2 constraint.

We see from (1) that this wavelet method does not pay too much attention to the energy but focuses on the dictionary; although, among such energies, for an appropriate wavelet basis, are some Besov norms (see [16]). We also remark that, following the same reasoning, a soft-thresholding of the coordinates of an image in any orthonormal basis (for instance: for deconvolution, a wavelet packet basis) can always be expressed as the minimization of an energy of the form (1).

A major drawback of wavelet soft-thresholding methods is that they are local in the wavelet domain ($\langle W(v), \Psi_l\rangle$ only depends on $\langle v, \Psi_l\rangle$). This drawback is of importance in the case of image deblurring when some information is lost during the degradation. In this case, we could expect to restore $\langle W(v), \Psi_l\rangle$ according to some information contained in other coordinates of v. (That is one of the advantages of the next method.) Another drawback of the method is that the constraint, $(W(v) - v) \in \mathcal{N}_{\mathcal{D},\tau}$, only constrains the movement along some orthogonal directions, while we could expect to constrain it in other directions (typically, we would like to use a real dictionary instead of a single basis).

Finally, a good choice of $\mathcal{D}$ yields a small threshold; for instance, in the case of finite dimensional images of size $N \times N$, $\tau = \sigma\sqrt{4 \ln N}$ is considered (see [9]) an optimal choice for the threshold.

The ROF model:

The ROF restoration method uses the total variation as a regularity criterion. The total variation is usually defined by duality (see [17]), for certain functions of $L^1(\mathbb{T})$. However, the total variation of any continuously differentiable functions of the torus w is simply

$$\int_{\mathbb{T}} |\nabla w|.$$

For simplicity, we will abuse of this notation throughout the paper. Moreover, we will write

$$BV(\mathbb{T}) = \{w \in L^1(\mathbb{T}), \int_{\mathbb{T}} |\nabla w| < \infty\}.$$

The ROF restoration method takes several equivalent forms. One of those (see [2]) is the constrained minimization problem:

$$\text{Minimize, } \int_{\mathbb{T}} |\nabla w|,$$

among functions w satisfying

$$\int_{\mathbb{T}} |H(w) - v|^2 \leq \sigma^2 .$$

This method can be expressed in the presented framework since

$$\begin{aligned} \{w \in L^2(\mathbb{T}), \int_{\mathbb{T}} |H(w) - v|^2 \leq \sigma^2\} \\ = \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\sigma}\} \end{aligned}$$

with

$$\mathcal{D} = \{\Psi \in L^2(\mathbb{T}), \|\Psi\|_2 = 1\} .$$

The main advantage of this method is that it reconstructs some lost information (see [5], [6]).

However, when presented this way, it is clear that the constraint could be improved. Indeed, it restricts changes along directions which are not autocorrelated (such as the direction of the noise $\frac{b}{\|b\|_2}$!). Therefore, the parameter τ, for ROF method, has to be much larger than in the previous case. For instance, in the case of the restoration of a discrete image of size $N \times N$, if we want $b \in \mathcal{N}_{\mathcal{D},\tau}$ to be likely, we must take $\tau = \sigma N$. This corresponds to a bound of the expectation

$$\begin{aligned} E\left(\sup_{\|\Psi\|_2=1} |\langle b, \Psi\rangle|\right) &= E\left(|\langle b, \frac{b}{\|b\|_2}\rangle|\right) \\ &= E\left(\|b\|_2\right) \\ &\leq \sqrt{E\left(\|b\|_2^2\right)} = \sigma N. \end{aligned}$$

Once again, this has to be compared to

$$E\left(\sup_{l\in\{1,\ldots,N^2\}} |\langle b, \Psi_l\rangle|\right) \leq \sigma\sqrt{4 \ln N}$$

(see [18]) when using a single basis (for instance, in the case of soft-thresholding).

An hybrid model:

This leads us to a model where we minimize the total variation under a constraint defined by a dictionary smaller than that used in the ROF method. Moreover, since, for such a model, we do not use any reconstruction formula (such as in the case of the wavelet thresholding), we can use a dictionary containing more than one basis. This leads us to

$$\begin{array}{rc} \text{Minimize,} & \int_{\mathbb{T}} |\nabla w| \\ \text{under the constraint} & (H(w) - v) \in \mathcal{N}_{\mathcal{D},\tau} \end{array} \tag{2}$$

for a dictionary $\mathcal{D}$ and a parameter $\tau > 0$. Such a model has recently and independently been evoked in [19] for the purpose of de-quantization and [13] for the purpose of de-noising. The design of the dictionary is of course now the key problem. It has to be understood as the set of all the "structures" which we do not want to erase. Therefore, it seems important to have very different types of elements such as wavelet, ridgelet, textures,... However, in the case of the deconvolution (or the inversion of an operator H), we want to control the noise in some particular directions, along which the noise is enhanced during the inversion of H. For instance for the deconvolution problem it is advisable to have some wavelet packet or Fourier bases in the dictionary.

It might be possible to apply the proposed modification of the data fidelity term, to other fields of image processing (such as segmentation, inpainting, ...). This should work as soon as the method is variational and contains a convex data fidelity term.

III. Mathematical analysis of the model

In this section, we show that the proposed model has a solution and that we can compute an approximation to this solution. Sketches of the proofs of the theorems are given in the appendix and the complete proofs are available in [20].

Theorem 1: Let $v \in L^2(\mathbb{T})$ and H be a continuous linear operator from $L^2(\mathbb{T})$ into itself. Let $\mathcal{D} \subset L^2(\mathbb{T})$ and $\tau > 0$. Assume that $BV(\mathbb{T}) \cap \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\} \neq \emptyset$ and that there exists $C > 0$ such that, for any $w \in \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$, $|\int_{\mathbb{T}} w| \leq C$. Then (2) admits a solution $u_\infty \in BV(\mathbb{T}) \cap \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$.

The hypothesis on the mean of the elements of $\{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$ is obviously satisfied when dealing with discrete functions and $\mathcal{D}$ contains a basis. Moreover, when H is a convolution with a kernel h whose Fourier transform satisfies $\hat{h}(0,0) \neq 0$, putting $1_{|\mathbb{T}}$ in $\mathcal{D}$ is sufficient.

We can unfortunately not guarantee the uniqueness of a solution to this model; indeed, neither $\{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$ nor the total variation are strictly convex. However, we can obtain some results similar to those given in [2] and [4].

In order to find a solution to (2), we approximate (2) by a model, whose solution yields an approximate solution to (2); this is done by penalization. We will see, however, that in some cases we have difficulties in actually computing solutions to the approximating model. This is explained in the next section and will lead us to some restrictions on $\mathcal{D}$.

Theorem 2: Let $v \in L^2(\mathbb{T})$ and H be a continuous linear operator from $L^2(\mathbb{T})$ into itself. Let $\mathcal{D} \subset L^2(\mathbb{T})$ be a countable set and $\tau > 0$. Assume that $BV(\mathbb{T}) \cap \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau'}\} \neq \emptyset$, for a $\tau' < \tau$, and that there exists $C > 0$ such that, for any $w \in \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$, $|\int_{\mathbb{T}} w| \leq C$. Then, for any $\epsilon > 0$,

$$E_\epsilon(w) = \int_{\mathbb{T}} |\nabla w| + \frac{1}{\epsilon} \sum_{\Psi \in \mathcal{D}} \Big(\sup\left(|\langle H(w) - v, \Psi \rangle| - \tau, 0\right) \Big)^2 \quad (3)$$

has a solution $w_\epsilon \in BV(\mathbb{T}) \cap L^2(\mathbb{T})$. Moreover, we can extract a sequence $(w_{\epsilon_n})_{n \in \mathbb{N}}$ (with $\lim_{n \to \infty} \epsilon_n = 0$) that converges in $L^1(\mathbb{T})$ and converges weakly in $L^2(\mathbb{T})$ to a function $w_0 \in BV(\mathbb{T}) \cap \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$. Moreover, for any such sub-sequence, its limit is a solution to (2).

IV. Numerical algorithm

In order to minimize (3), we apply a steepest descent algorithm. Let us describe this algorithm; let us denote

$$\begin{aligned} E(w) &= \sum_{i,j=0}^{N-1} |\nabla w_{i,j}| \\ &\quad + \frac{1}{\epsilon} \sum_{\Psi \in \mathcal{D}} \Big(\sup\left(|\langle H(w) - v, \Psi \rangle| - \tau, 0\right) \Big)^2 \\ &= TV(w) + \frac{1}{\epsilon} \sum_{\Psi \in \mathcal{D}} J_\Psi(w) \end{aligned}$$

a discrete version of (3), where

$$J_\Psi(w) = \sup\left(|\langle H(w) - v, \Psi \rangle| - \tau, 0\right)^2.$$

We also denote by $(\varphi_l)_{l=1,\ldots,N^2}$ a basis of $\mathbb{R}^{N^2}$, consisting of Dirac delta functions. Given $w^0 = v$ the initial image, we need to compute, at each iteration and for all $l \in \{1, ..., N^2\}$,

$$\frac{\partial E(w^n)}{\partial \varphi_l} = \frac{\partial TV(w^n)}{\partial \varphi_l} + \frac{1}{\epsilon} \sum_{\Psi \in \mathcal{D}} \frac{\partial J_\Psi(w^n)}{\partial \varphi_l}$$

We then let

$$\nabla E(w^n) = \sum_{l=1}^{N^2} \frac{\partial E(w^n)}{\partial \varphi_l} \varphi_l$$

We then compute (with a dichotomy algorithm) the optimal step

$$s^n = argmin_{t \in \mathbb{R}}\; E\left(w^n - t \nabla E(w^n)\right).$$

Finally, we set

$$w^{n+1} = w^n - s^n \nabla E(w^n).$$

We remark that it is important here to compute the optimal step s^n, since, when ϵ is very small (which should be the case), we would otherwise have to choose a very small step to ensure convergence.

The computation of $\frac{\partial TV(w^n)}{\partial \varphi_l}$ has been studied widely; for instance, such computations are considered in [3] and [7]. Moreover, in practice, we approximate the total variation by $\sum_{i,j=0}^{N-1} \sqrt{\varepsilon^2 + |\nabla w_{i,j}|^2}$, for a small $\varepsilon \in \mathbb{R}$. This common approximation, which was introduced and discussed in [21], permits the minimization of a differentiable energy. It is clear that this energy satisfies sufficient hypotheses for the convergence of the steepest descent algorithm (see, for instance, page 29 of [22]).

We remark also that, even if we use only one orthonormal basis $(\varphi_l)_{l=1,\ldots,N^2}$ to express the gradient, we can of course switch from one basis to another basis. For instance, when $\mathcal{D}$ contains an orthonormal basis $(\varphi'_k)_{k=1,\ldots,N^2}$ (which is always the case in our experiments), in order to compute $\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi_l}$, we use

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi_l} = \sum_{j=1}^{N^2} \frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_j} \langle \varphi'_j, \varphi_l \rangle.$$

This means that we can compute $\left(\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi_l}\right)_{l=1,\ldots,N^2}$ from $\left(\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_j}\right)_{j=1,\ldots,N^2}$ with a fast transform.

Consider $\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_j}$; a simple calculation yields

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_j} = 0,$$

if $|\langle H(w^n)-v,\varphi'_k\rangle|<\tau$, and otherwise

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_j} = 2sign\left(\langle H(w^n)-v,\varphi'_k\rangle\right)\langle H(\varphi'_j),\varphi'_k\rangle \left(|\langle H(w^n)-v,\varphi'_k\rangle|-\tau\right) \quad (4)$$

where $sign(t)=1$, if $t\geq 0$, and -1 otherwise.

In (4), all the terms except $\langle H(\varphi'_j),\varphi'_k\rangle$ are, in practice, easy to compute and store. Indeed, $H(w^n)$ is known (for instance a convolution) and the scalar products $\langle H(w^n)-v,\varphi'_k\rangle$ are computed with fast transforms depending on the chosen dictionary. (In our experiments, we take wavelet packet bases and transforms.)

We observe also that, when H is the identity (the denoising case), we have

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_k} = 0,$$

if $|\langle w^n-v,\varphi'_k\rangle|<\tau$,

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_k} = 2sign\left(\langle w^n-v,\varphi'_k\rangle\right)\left(|\langle w^n-v,\varphi'_k\rangle|-\tau\right),$$

if $|\langle w^n-v,\varphi'_k\rangle|\geq\tau$, and

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_j} = 0$$

for $j\neq k$. Therefore, all the $\sum_{k=1}^{N^2}\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi_l}$, for $l\in\{1,...,N^2\}$, are easily computed as the inverse transform of the coefficients $\left(\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_k}\right)_{k=1,\dots,N^2}$.

For a general H, it becomes difficult to compute and store all the $\langle H(\varphi'_j),\varphi'_k\rangle$. In some particular cases, this could probably be achieved with the help of a wavelet-vaguelet kind of decomposition (see [8]). However, in general, too much computation and memory are required. This leads to a constraint on the elements of the dictionary.

We seek a dictionary made of orthonormal bases $\mathcal{B}$ whose elements "almost" diagonalize H. That is, there exists $\lambda_{H,\Psi}\in\mathbb{R}$, such that

$$H(\Psi)\sim\lambda_{H,\Psi}\Psi,$$

for every $\Psi\in\mathcal{B}$. For instance, for convolutions this can be achieved by a dictionary made of Fourier and/or wavelet packet bases (see [10], [11]). Using such a dictionary, we know $\langle H(\varphi'_j),\varphi'_k\rangle$ approximately. So, we have approximately

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_k} = 0,$$

if $|\langle H(w^n)-v,\varphi'_k\rangle|<\tau$,

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_k} = 2\lambda_{H,\varphi'_k}sign\left(\langle H(w^n)-v,\varphi'_k\rangle\right)\left(|\langle H(w^n)-v,\varphi'_k\rangle|-\tau\right)$$

if $|\langle H(w^n)-v,\varphi'_k\rangle|\geq\tau$, and

$$\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_j} = 0,$$

for all $j\neq k$. Once again, all the $\sum_{k=1}^{N^2}\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi_l}$, for $l\in\{1,...,N^2\}$, are then computed easily as the inverse transform of the coefficients $\left(\frac{\partial J_{\varphi'_k}(w^n)}{\partial \varphi'_k}\right)_{k=1,\dots,N^2}$.

We will see in the experiments of Section V-B that the approximation $H(\varphi'_k)=\lambda_{H,\varphi'_k}\varphi'_k$ seems reasonable.

Moreover, in practice, we choose a small value for ϵ, we fix it at a small value (say 0.001); with this value, after convergence, we obtain actually $|\langle H(w_0)-v,\Psi\rangle|\leq\tau$. The algorithm speed could probably be improved, by starting from a larger ϵ and letting it decrease during the process.

V. Experimental results

All the wavelet packet bases used in this section are based on the cubic spline wavelet; their trees are always of maximum depth 3. For image deblurring, we use the mirror tree which is described in [10]. To be fair, it should be remarked that, all the methods to be compared have only one parameter. This parameter has always been chosen empirically, in order to have good results. Of course, we could add new parameters to improve the results further.

A. Experiments on denoising

All the experiments of this section are made with the data displayed in Figure 1. In this figure, the image on the right is obtained by adding a Gaussian noise, of standard deviation 20, to the image on the left. This corresponds to a signal to noise ratio (SNR) of 8.12.

We display the results of several denoising methods in Figure 2. Here is the description of the methods:

- Top: ROF method[2], with $\lambda=0.01$.
- Middle-Left: The wavelet soft-thresholding, with a parameter $\tau=70$.
- Middle-Right: The restoration, using (2), with dictionary consisting of a single wavelet basis and a parameter $\tau=70$.
- Bottom-Left: The noise selection approach, with a wavelet packet dictionary consisting of fully decomposed wavelet packet bases of depth 1, 2 and 3 and a parameter $\tau=70$. (The idea is just to "compose" restorations in all the wavelet packet bases (see [23]). We would like to say that, since we wrote [23], we have discovered that identical ideas had been published in [24] and [25]).
- Bottom-Right: The restoration, using (2), with dictionary consisting of 4 translations of a wavelet packet dictionary (consisting of fully decomposed wavelet packets bases of depth 1, 2 and 3), for a parameter $\tau=120$.

[2]Here, we consider the ROF method as the minimization of

$$\int_{\mathbb{T}}|\nabla w|+\lambda\int_{\mathbb{T}}|H(w)-v|^2.$$

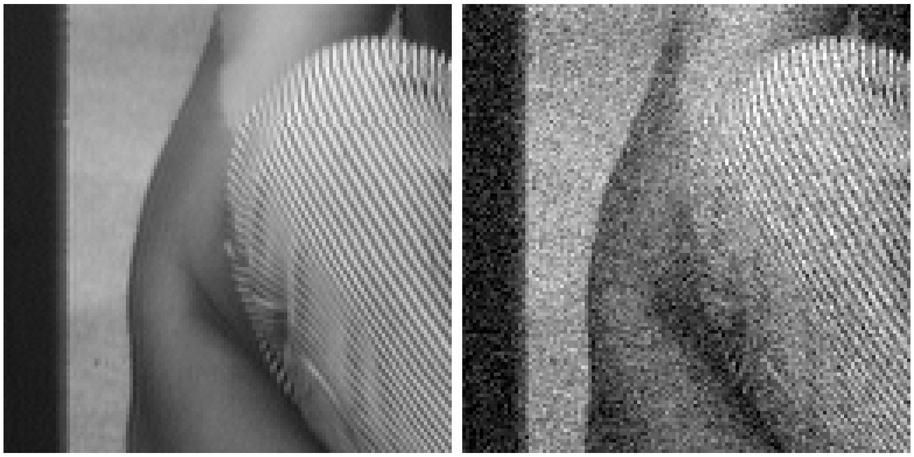

Fig. 1. Left: Original image; Right: Noisy image.

First we remark that, either with a wavelet basis or a wavelet packet dictionary, the introduction of the total variation compared to the thresholding (or noise selection) yields sharper edges. We also remark that they do not present Gibbs phenomena in the vicinity of these edges. This is due to the fact that (2) allows the reconstruction of some small coefficients, which are canceled by the thresholding.

Moreover, if we compare the images of the middle with those at the bottom of Figure 2, we see clearly that we gain in putting a larger dictionary. Indeed, the texture of the pans can be represented by few large wavelet packet coefficients (well localized in the Fourier domain) or a lot of small correlated wavelet coefficients. This explains why they are preserved with the larger dictionary. However, we do not claim that the wavelet packet dictionary is particularly appropriate and it is even likely that, in general, a dictionary made of very different kind of elements (wavelets, ridgelets, textures, ...) would be more efficient.

Finally, the texture is better preserved in the image on the bottom right than in the one denoised with ROF method. This is due to the fact that we restrict the evolution of the result in the direction of the texture much more in the case of (2) than with the usual ROF model. This shows that the dictionary should not be too large.

B. *Experiments on deblurring*

We present here some experiments on image deconvolution. Therefore, we want to recover u, given

$$v = h * u + b.$$

Since we do not want to consider the aliasing in the creation of the image and we want to demonstrate the ability of the method to avoid Gibbs phenomena, we take, for h, a function whose Fourier transform cancels all the frequencies outside $[-\frac{\pi}{2}, \frac{\pi}{2}] \times [-\frac{\pi}{2}, \frac{\pi}{2}]$. More precisely, the Fourier transform of h is given by

$$\widehat{h}(\xi, \eta) = \left(\frac{sin(2\xi)}{2\xi}\right)\left(\frac{sin(2\eta)}{2\eta}\right), \qquad (5)$$

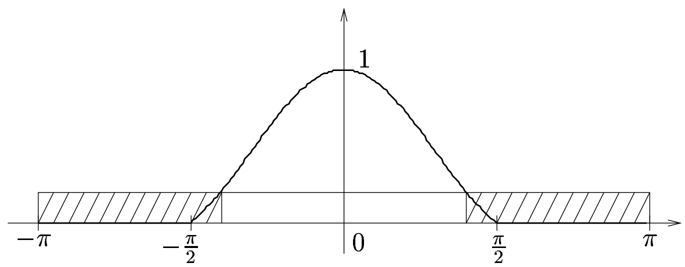

Fig. 3. Profile of the Fourier transform of h (see (5)). The hatching represents the frequencies which are, in practice, lost during the degradation.

for ξ and $\eta \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ and 0 otherwise (see Figure 3). We add a Gaussian noise of standard deviation 2. This corresponds to a SNR (between the convolved and the noisy convolved images) of 27.4. Remark that this degradation model is particularly ill-adapted to wavelet packet methods since it cancels a wide band of frequencies (see Figure 3). In fact, we know that, because of their ability to reconstruct some lost frequencies (see [4]), variational methods are better suited to this kind of degradation model.

We display in Figure 4:

- Top: the reference image.
- Middle-Left: the degraded image.
- Middle-Right: The noise selection approach, with a subdictionary of the wavelet packet dictionary (see [23]) and a parameter $\tau = 12$. The idea is just to compose the restoration in all the wavelet packet bases which are more decomposed than the mirror basis (see [10]) and whose maximum depth is 3. We also average the results of the algorithm over 4 translations.
- Bottom-Left: The restoration with ROF method, with a parameter $\lambda = 0.1$.
- Bottom-Right: The restoration, with (2), for a parameter $\tau = 12$ and dictionary comprising all the wavelet packet bases which are more decomposed than the mirror basis and whose maximum depth is 3 (as well as 4 of its translations).

On this latter image, we do not see artifacts similar to aliasing, Gibbs phenomena or blurring. Since they are typi-

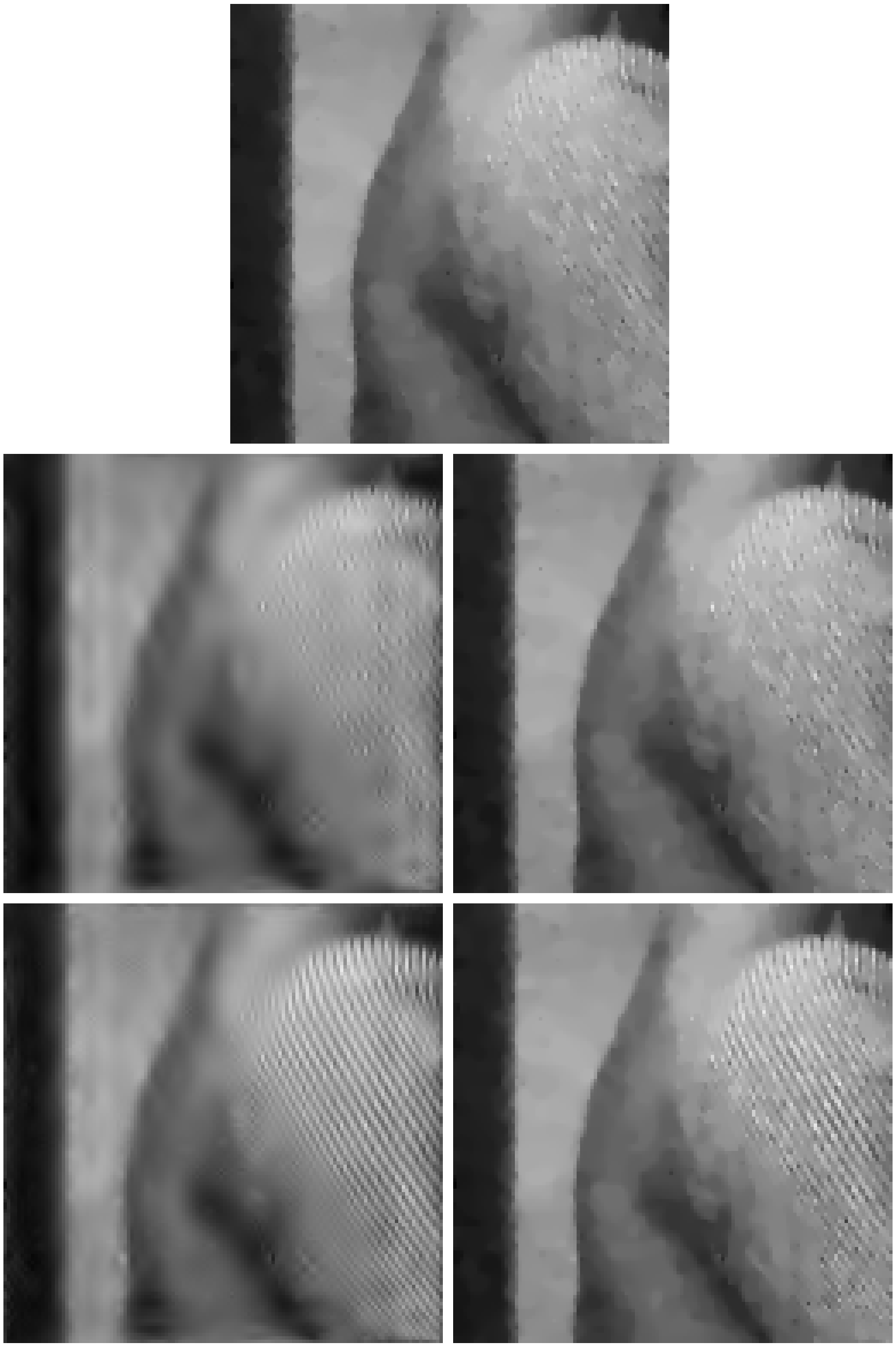

Fig. 2. Restoration with: Top: Rudin-Osher-Fatemi method; Middle-Left: wavelet soft-thresholding; Middle-Right: solution to (2), with a single wavelet basis; Bottom-Left: noise selection with a wavelet packet dictionary; Bottom-Right: solution to (2), with a wavelet packet dictionary.

cal artifacts of the approximation $H(\Phi) = \lambda_{H,\Psi}\Psi$, we think this approximation (see [11]) is acceptable in our case.

Moreover, it is clear that compared to ROF method, the new model preserves the texture better, since this latter is more constrained by the data fidelity term. Compared to the wavelet packet method, we have a similar constraint but, since our model permits the extrapolation of the lost frequencies, we do not have any Gibbs phenomena. This is of course also the case with ROF method.

Appendix

Once again, all the details of the proofs are given in [20].

Sketch of the proof of theorem 1: The proof breaks down as follows:

- Build a minimizing sequence $(u_n)_{n\in\mathbb{N}}$.
- Show that $(u_n)_{n\in\mathbb{N}}$ is bounded in $L^2(\mathbb{T})$ and $BV(\mathbb{T})$.
- Extract a sub-sequence of $(u_n)_{n\in\mathbb{N}}$ which converges in $L^1(\mathbb{T})$ and converges weakly in $L^2(\mathbb{T})$ to a limit u_∞.
- Show that u_∞ is solution to (2).

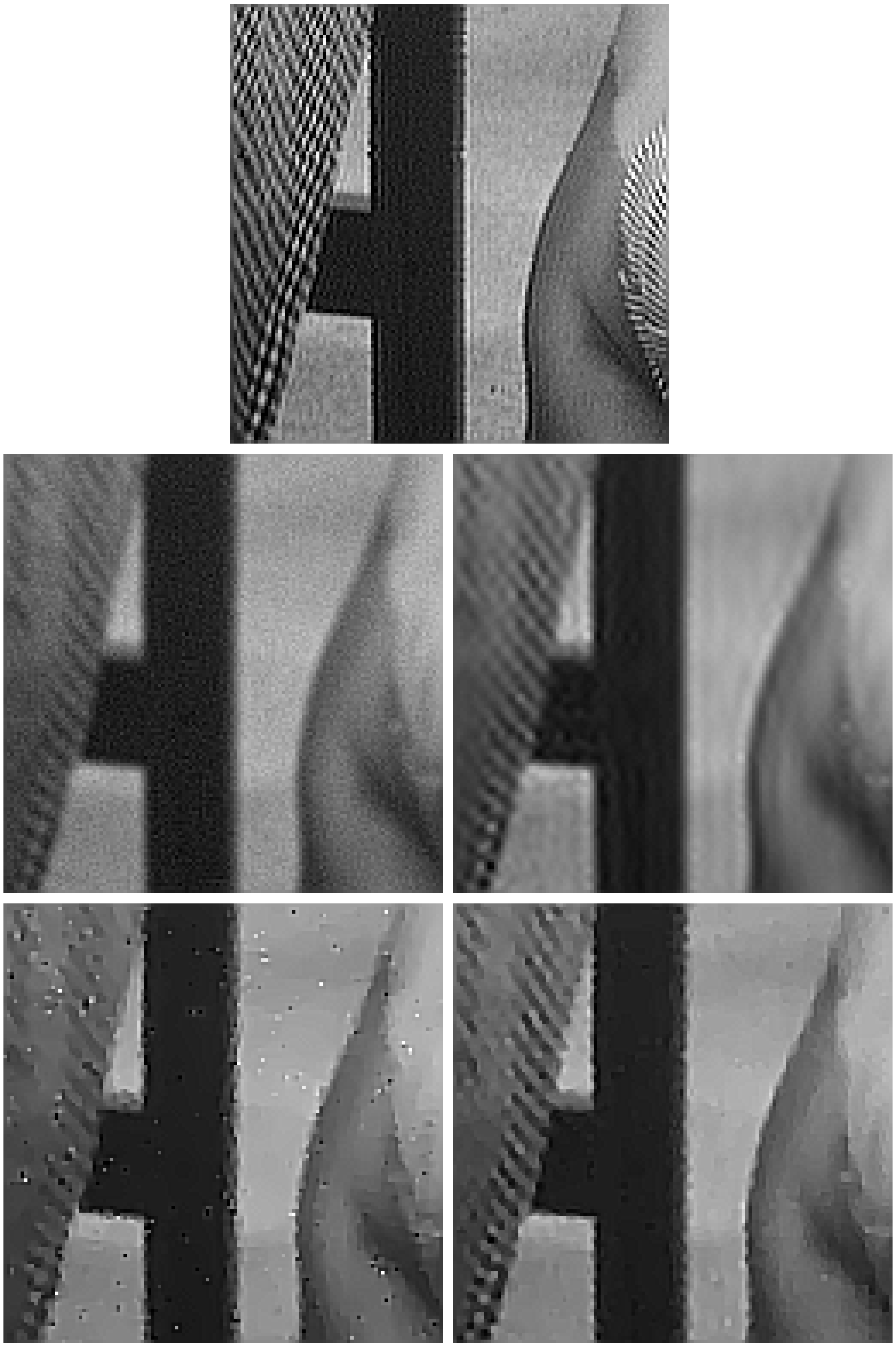

Fig. 4. All the images have been **sharpened** for the display. Top: Original image; Middle-Left: Degraded image; Middle-Right: Noise selection in a sub-dictionary of the wavelet packet dictionary; Bottom-Left: Rudin-Osher-Fatemi method; Bottom-Right: solution to (2), with a sub-dictionary of the wavelet packet dictionary.

Sketch of the proof of theorem 2: The proof breaks down as follows:

- Show that, for any $\epsilon > 0$, there exists w_ϵ minimizing (3). This proof follows the same sketch as the proof of Theorem 1. It is, however, a little more technical.
- Prove that $(w_\epsilon)_{\epsilon>0}$ is bounded in $L^2(\mathbb{T})$ and $BV(\mathbb{T})$.
- Conclude that there exists $w_0 \in L^2(\mathbb{T})$ and a sub-sequence $(w_{\epsilon_n})_{n\in\mathbb{N}}$ of $(w_\epsilon)_{\epsilon\in\mathbb{R}}$ (with $\lim_{n\to\infty} \epsilon_n = 0$), such that the sub-sequence converges in $L^1(\mathbb{T})$ and converges weakly in $L^2(\mathbb{T})$ to w_0.
- Show that $w_0 \in \{w \in L^2(\mathbb{T}), H(w) - v \in \mathcal{N}_{\mathcal{D},\tau}\}$.
- Prove that w_0 is a solution to (2).

References

[1] L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D*, vol. 60, pp. 259–268, 1992.

[2] A. Chambolle and P.L. Lions, "Image recovery via total variation minimisation and related problems," *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.

[3] T. F. Chan and P. Mulet, "On the convergence of the lagged diffusivity fixed method in total variation image restoration," *SIAM Journal of Numerical Analysis*, vol. 36, no. 2, pp. 354–367, 1999.

[4] S. Durand, F. Malgouyres, and B. Rougé, "Image deblurring, spectrum interpolation and application to satellite imaging," *Control, Optimisation and Calculus of Variation*, vol. 5, no. 445-475, 2000, A preliminary version is available at http://www.zeus.math.univ-paris13.fr/~malgouy.

[5] F. Malgouyres, *Increase in the Resolution of Digital Images: Variational Theory and Applications*, Ph.D. thesis, Ecole Normale Supérieure de Cachan, Cachan, France, 2000, Available at http://www.zeus.math.univ-paris13.fr/~malgouy.

[6] F. Malgouyres and F. Guichard, "Edge direction preserving image zooming: a mathematical and numerical analysis," *SIAM, J. Num. Anal.*, vol. 39, no. 1, pp. 1–37, 2001, A preliminary version is available at: http://www.zeus.math.univ-paris13.fr/~malgouy.

[7] A. Marquina and S. Osher, "Explicit algorithms for a new time dependent model based on level set motion for nonlinear deblurring and noise removal," *SIAM, Journal of Sientific Computing*, vol. 22, no. 2, pp. 387–405, 2000.

[8] D. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Applied and Computational Harmonic Analysis*, vol. 2, pp. 101–126, 1995.

[9] D. Donoho and I.M. Johnstone, "Minimax estimation via wavelet shrinkage," Tech. Rep., Departement of Stat., Stanford University, 1992.

[10] J. Kalifa, *Restauration minimax et déconvolution dans une base d'ondelettes miroirs*, Ph.D. thesis, Ecole Polytechnique, 1999, Available at http://www.cmap.polytechnique.fr/~kalifa.

[11] F. Malgouyres, "A framework for image deblurring using wavelet packet bases," *Applied and Computational Harmonic Analysis*, vol. 12, no. 3, pp. 309–331, 2002.

[12] J. Kalifa, S. Mallat, and B. Rougé, "Image deconvolution in mirror wavelet bases," in *Int. Conf. on Image Processing (ICIP'98)*, Chicago, Oct. 1998, pp. 565–569, IEEE.

[13] J.L. Starck, D. Donoho, and E. Candes, "Very high quality image restoration," in *Wavelet and applications in signal and image processing IX*, San Diego, August 2001, SPIE, vol. 4478, pp. 9–19.

[14] P.L. Combettes and H.J. Trussell, "The use of noise properties in set theoretic estimation," *IEEE Transactions on Signal Processing*, vol. 39, no. 7, pp. 1630–1641, July 1991.

[15] A. Chambolle, R.A. De Vore, N. Lee, and B.J. Lucier, "Nonlinear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage," *IEEE Transactions on Image Processing*, vol. 7, pp. 319–355, 1998, Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis.

[16] Y. Meyer, *Ondelettes et opérateurs*, Hermann, 1990.

[17] L.C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Studies in Advanced Mathematics. CRC Press, Boca Raton, 1992.

[18] S.M. Berman, "Sojournes and extremes os stochastic processes," *Wadsworth, Reading, MA*, 1989.

[19] R. R. Coifman and A. Sowa, "Combining the calculus of variations and wavelets for image enhancement," *Applied and computational harmonic analysis*, vol. 9, pp. 1–18, 2000.

[20] F. Malgouyres, "Mathematical analysis of a model which combines total variation and wavelet for image restoration," *Journal of information processes*, vol. 2, no. 1, pp. 1–10, 2002, available at http://www.zeus.math.univ-paris13.fr/~malgouy.

[21] R. Acar and C. Vogel, "Analysis of bounded variation methods for ill-posed problems," *Inverse Problems*, vol. 10, pp. 1217–1229, 1994.

[22] E. Polak, *Computational methods in optimization*, vol. 7 of *Mathematics in science and engineering*, Academic Press, New York, 1971.

[23] F. Malgouyres, "A noise selection approach of image restoration," in *SPIE, International Conference on Wavelets IX*, Unser Laine and Aldroubi, Eds., San Diego, August 2001, vol. 4478, pp. 34–41.

[24] P. Ishwar and P. Moulin, "Multiple-domain image modeling and restoration," in *Proc. IEEE Intl. Conf. Image Processing*, Kobe, Japan, Oct. 1999, vol. 1, pp. 362–366.

[25] L. Woog, *Adapted Waveform Algorithms for Denoising*, Ph.D. thesis, Yale University, April 1996, Available at http://www.cs.yale.edu/homes/woog.html.

François Malgouyres had his PhD in mathematics co-advised by J.M. Morel and B.Rougé, at the Ecole Normale Supérieur de Cachan (France), in 2000. He went to the department of mathematics of the University of California, Los Angeles (USA), during the academic year 2000-2001, as a CAM assistant professor (within the teams of S. Osher and T. Chan). He has been working as a "maître de conférence" at the LAGA, L2TI and LIPN, at the University Paris 13, since 2001.