Combining total variation and wavelet packet approaches for image deblurring*

François Malgouyres
Dept of Mathematics
UCLA, 6363 Math science building
Los Angeles, CA 90095-1555, USA
http://www.math.ucla.edu/∼malgouy

Abstract

We show two ways to combine wavelet packets and total variation based deblurring methods. For this purpose, we first recall that it is possible to approximate a convolution by mean of an operator diagonal in a wavelet packet basis. Then, we show two possibilities, which use this property, for combining wavelet packets and total variation approaches. We then show on experiments that, doing this we can expect to have the advantages of both approaches while avoiding their drawbacks.

1. Introduction

This paper is mainly concerned with image deblurring and with the application of a property of operator diagonal in wavelet-packet bases for this purpose. More precisely, it has been shown in [11] that the average over translations of an operator which is diagonal in a wavelet packet basis is a convolution. We will investigate two variational applications of this property to the issue of image deblurring.

The deblurring problem under our scope is to restore a convolved and noisy image u, given the data

$$u_0 = s * u + n\,,$$

where s is a low-pass filter and n is a noise. Expressing this in the Fourier Domain (we recall that the Fourier basis diagonalizes the convolution operator), we obtain

$$\widehat{u_0} = \widehat{s}\,\hat{u} + \hat{n}\,,$$

where we note with a hat the Fourier transform of a function. We clearly see here that, since $\widehat{s}$ can be very small or even be zero, this problem is ill-posed.

*This work has been financed by NSF grant DMS-9973341.
I would like to thank B. Rougé, S. Durand and J.M. Morel for their encouragements as well as for all the fruitful discussions we had on these subjects.
The results of the experiments of this paper are available at http://www.math.ucla.edu/∼malgouy

The first reason why people have used the framework of wavelet packet for image deblurring is that it permits to have both a sparse representation of an image (and therefore to separate the information and the noise) and a good frequencial localization. This has first been noticed by B. Rougé and has already been used under various ways (see [7, 9, 11, 15]). The method described in these articles are based on a shrinkage of the wavelet packet coefficients similar to the wavelet shrinkage approach, for the purpose of denoising, of Donoho and Johnstone (see [6]). These methods have recently been formalized in a way which permits to envisage to combine wavelet packet and variational methods. This paper investigates two possibilities for such a combination.

There is an abundant literature on image deblurring. The reader is referred to [1] for most of the linear methods and to [5, 8] for overviews on the subject. In few words, the first approach consists in enhancing images without regard to the convolution kernel [10]. The other methods are based on regularization approaches of the problem: using statistical properties (Wiener and Kalman filters) or regularity measurements of the images such as the entropy (see [5] and references there), the total variation (see [16]) or the characterization of Besov spaces by wavelets coefficients (see [3, 6]).

In Section 2, we make some recalls on wavelet packet bases and state the result saying that it is possible to approximate a convolution operator by averaging over translations an operator diagonal in a wavelet packet basis. Then, in Section 3, we propose a first application of this result which combines the total variation and the wavelet packet approach. This approach, compared to the usual Rudin-Osher-Fatemi method (see [16]), can be interpreted as the adaptation of the parameter λ (see (6) where we recall the form of the Rudin-Osher-Fatemi functional) according to a criterion on the wavelet packet decomposition of the image. In Section 4, we study a second approach where the role of wavelet packet is more important. Indeed, in this case for $\lambda = \infty$ the result of the method is very similar to the

result of the FCNR (see [9, 15]) and as λ decreases the influence of the total variation appears. The advantage here is that the total variation permits to remove the ringing which can appear with the FCNR. At last, in Section 5, we display some experiments which show to evidence the role of the parameters in the considered methods. Moreover, we compare the methods introduced in this paper to both a wavelet packet and Rudin-Osher-Fatemi methods and obtain satisfactory results. One of the method permits to simultaneously avoid ringing and preserve textures.

2. Approximation of the convolution in a wavelet packet basis

As we said in the introduction, it has already been discussed in [11] that it is possible to approximate a convolution by mean of the average over translations of an operator diagonal in a wavelet packet basis. Let us make some recalls on this property and introduce the notations we will use for the wavelet packet decomposition.

For simplicity, we only describe wavelet packet bases in the case of functions of $\mathbb{R}$, higher dimensional cases and wavelet packet bases on an interval can be deduced from this one by respectively taking tensor products and (for instance) periodizing functions out of the interval (see [12]). For more details the reader is referred to [4] or to Section 8 of [12].

In the following, we will denote by (h, g) a pair of conjugate mirror filters related with a multi-resolution analysis (for instance $g_n = (-1)^{1-n} h_{1-n}$) and by ϕ the associated scaling function. Letting $\psi_0^0 = \phi$, we can define recursively, for $j \in \mathbb{N}$ and $p \in \{0, ..., 2^j - 1\}$

$$\psi_{j+1}^{2p}(x) = \sum_{n=-\infty}^{\infty} h_n \psi_j^p(x - 2^j n)\,, \tag{1}$$

and

$$\psi_{j+1}^{2p+1}(x) = \sum_{n=-\infty}^{\infty} g_n \psi_j^p(x - 2^j n)\,. \tag{2}$$

Therefore, if we note $\psi_{j,n}^p(x) = \psi_j^p(x - 2^j n)$ and $\mathbf{W}_j^p$ the vectorial subspace of $L^2(\mathbb{R})$ generated by $\{\psi_{j,n}^p,\ n \in \mathbb{Z}\}$, we know that $\{\psi_{j,n}^p,\ n \in \mathbb{Z}\}$ is an orthonormal basis of $\mathbf{W}_j^p$. Moreover, we have

$$\mathbf{W}_{j+1}^{2p+1} \oplus \mathbf{W}_{j+1}^{2p} = \mathbf{W}_j^p.$$

We also know that for any admissible tree (see Section 8 of [12]) $(p_l, j_l)_{1 \le l \le L}$, $\{\psi_{j_l,n}^{p_l}\}_{n \in \mathbb{Z}, 1 \le l \le L}$, is an orthonormal basis of $\mathbf{W}_0^0$.

In the following, we will identify any $(u_n)_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})$ with $\tilde{u} = \left(\sum_{n \in \mathbb{Z}} u_n \psi_{0,n}^0\right) \in \mathbf{W}_0^0$. Therefore, noting $u_{j,n}^p = \langle \tilde{u}, \psi_{j,n}^p \rangle$ and $(u_j^p)_n = u_{j,n}^p$, we can deduce from (1) and (2) that for any admissible tree $(p_l, j_l)_{1 \le l \le L}$, there exists some kernels $H_{j_l}^{p_l}$ such that

$$u_{j_l,n}^{p_l} = H_{j_l}^{p_l} * u(2^{j_l} n)\,.$$

We can now state the following proposition (which has already been introduced in [11]).

Proposition 1 *Let $(\psi_{j_l,n}^{p_l})_{n \in \mathbb{Z}, 1 \le l \le L}$ be a wavelet packet basis. Let $\widetilde{D}$ be a linear continuous operator from $l^2(\mathbb{Z})$ into $l^2(\mathbb{Z})$, diagonal in the basis $(\psi_{j_l,n}^{p_l})_{n \in \mathbb{Z}, 1 \le l \le L}$. Assume moreover that the eigenvalues $(\lambda_{j_l,n}^{p_l})_{n \in \mathbb{Z}, 1 \le l \le L}$ (respectively associated to the eigenvector $\{\psi_{j_l,n}^{p_l}\}_{n \in \mathbb{Z}, 1 \le l \le L}$) do not depend on n. Then, the operator D defined, for any $u \in l^2(\mathbb{Z})$, by*

$$D(u) = 2^{-J} \sum_{k=0}^{2^J - 1} \tau_{-k} \circ \widetilde{D} \circ \tau_k(u)\,, \tag{3}$$

where $J = \max_{1 \le l \le L} j_l$ and τ_k represents the translation operator of $k \in \mathbb{Z}$, is a convolution continuous from $l^2(\mathbb{Z})$ into $l^2(\mathbb{Z})$. Moreover, the Fourier transform of the convolution kernel $\tilde{s}$ defining D is given, for $\xi \in [-\pi, \pi]$, by

$$\widehat{\tilde{s}}(\xi) = \sum_{l=1}^{L} \lambda_{j_l}^{p_l} \frac{|\widehat{H_{j_l}^{p_l}}(\xi)|^2}{2^{j_l}}\,, \tag{4}$$

where we note, for any $l \in \{1, ..., L\}$ and $n \in \mathbb{Z}$, $\lambda_{j_l}^{p_l} = \lambda_{j_l,n}^{p_l}$.

Once again, the proof of this result and some experiments showing that (4) can be used to properly approximate a convolution are given in [11].

Therefore, we can use a wavelet packet basis as an intermediate step for the Fourier basis. Of course, the advantage of this intermediate step is to have the possibility to decorrelate the noise and the information, which is of a great interest for the issue of image deblurring. In order to do so, we need to choose a tree (we will only use the cubic spline wavelet (see [12])) and the values of $\lambda_{j_l}^{p_l}$. For simplicity, we will always take the "best tree" introduced in [9]. Moreover, we will estimate the $\lambda_{j_l}^{p_l}$ by

$$\lambda_{j_l,n}^{p_l} = \langle s * \psi_{j_l,n}^{p_l}, \psi_{j_l,n}^{p_l} \rangle\,, \tag{5}$$

which permits to minimize $\|S - \widetilde{D}\|_2$, where $S(u) = s * u$ and $\tilde{D}$ is given in Proposition 1. Of course, all these choices can be improved by using (4) to design the approximation.

The first application of this result is, of course, to redefine the FCNR which has already been studied in [9, 15]. In the following, we will investigate the possibility of using this result in the framework of variation methods.

3. First combination of wavelet packet and total variation methods

We are now going to introduce a variational application of the approximation of the convolution to the problem of deconvolution. This consists in introducing a wavelet packet term in the method introduced by Rudin, Osher and Fatemi in [16]. In order to have a well defined variational problem, we boil down to the finite dimensional case where the signals are assumed to be of size $N \in \mathbb{N}$. Let us first make some recalls on this latter method.

Rudin, Osher and Fatemi introduced the total variation based deconvolution method, which consists in minimizing, for $N \in \mathbb{N}$ and a data $g \in \mathbb{R}^N$, the functional

$$TV(u) + \lambda \|s * u - g\|_2^2 \,, \tag{6}$$

among $u \in \mathbb{R}^N$, where λ can be interpreted as a Lagrange multiplier (see [2]) and the total variation is defined by

$$TV(u) = \sum_{m=0}^{N-1} |u_{m+1} - u_m| \,.$$

The main advantage of this method is that, since the total variation does not expect too much smoothness at edges, it permits to avoid ringing artifacts at their vicinity.

One of the possible improvement of this method is to adapt the value of the parameter λ to the region of the image we are considering (this has been investigated in [17]). Indeed, ideally, we would prefer to have a λ larger on region with and a λ smaller on smooth region where the image contains almost no information. However, the segmentation of the image is not a simple problem. Indeed, the segmentation needs to be adapted to the local spatial behavior of the image (for instance, if it is a textured or smooth region or an edge) and to local frequencial information (typically, a texture which has a "local frequency" belonging to a region where the Fourier transform of the kernel $\hat{s}$ is "large" should be preserved and those corresponding to a "local frequency" belonging to a region where the Fourier transform of the kernel $\hat{s}$ is "small" or zero should be erased since they correspond to a ringing artifact). This has led us to use the wavelet packet transform to "segment" our image.

More precisely, we do not really "segment" the image in a rigorous sense but make λ depend on the value of the wavelet packet coefficient of the data. Let us explain this in details.

First, we remark that in the Rudin-Osher-Fatemi model is hidden the fact that λ is adapted to the frequencial localization. This has first been remarked in [7] and is simply due to the fact that, for instance, when the convolution with s is invertible we can rewrite the functional

$$TV(u) + \sum_{k=0}^{N-1} \lambda |\hat{s}_k|^2 |\hat{u}_k - \hat{g\prime}_k|^2 \,,$$

where $\hat{g\prime}_k = \frac{\hat{g}_k}{\hat{s}_k}$ for $k \in \{0, \cdot, N-1\}$. Therefore, Rudin-Osher-Fatemi functional can be interpreted as a denoising (of a roughly deblurred version of g) with a value λ varying with the frequency localization. We can rewrite this functional under the form

$$TV(u) + \|(\sqrt{\lambda} s) * (u - g\prime)\|_2^2 \,. \tag{7}$$

We remark here that we can use Proposition 1 to modify and adapt the convolution in the right hand side term of (7) in order to include some information on the spatial behavior of the image. Therefore, noting $\boldsymbol{\lambda} = (\lambda_{j_l}^{p_l})_{1 \le l \le L}$ a set of eigenvalue permitting to approximate the convolution with s, and $J = \max_{1 \le l \le L} j_l$, we can modify this functional and minimize

$$TV(u) + \| \sum_{k=0}^{2^J-1} \tau_{-k} \circ S_{\boldsymbol{\lambda}, \tau_k g, \sigma, \sqrt{\lambda_1}, \sqrt{\lambda_2}} \circ \tau_k (u - g\prime) \|_2^2 \,, \tag{8}$$

where $S_{\boldsymbol{\lambda}, g, \sigma, \sqrt{\lambda_1}, \sqrt{\lambda_2}}$ is defined by its coordinates, $\langle S_{\boldsymbol{\lambda}, g, \sigma, \sqrt{\lambda_1}, \sqrt{\lambda_2}}(v), \psi_{j_l, n}^{p_l} \rangle$, which are equal to

$$\begin{cases} \sqrt{\lambda_2}\, \lambda_{j_l}^{p_l} \langle v, \psi_{j_l,n}^{p_l} \rangle & \text{, if } |\langle g, \psi_{j_l,n}^{p_l} \rangle| \ge \sigma, \\ \sqrt{\lambda_1}\, \lambda_{j_l}^{p_l} \langle v, \psi_{j_l,n}^{p_l} \rangle & \text{, otherwise.} \end{cases} \tag{9}$$

Heuristically, and with comparison to the Rudin-Osher-Fatemi functional (6), we simply take $\lambda = \lambda_1$ where we consider we do not have information and $\lambda = \lambda_2$ where we consider there are informations (be it a texture or an edge).

Note that we know there exists a minimum of the functional (8) since this latter is convex. As usual, we cannot guaranty the uniqueness of the result since the functional is not necessarily strictly convex. However, we could state, about this issue, results similar to the one given in [2, 7].

We present in Section 5 some experiments on this method. In our opinion, these experiments are very stimulating. However, we feel that there are a lot of possible improvements for this kind of method. One of the most obvious is probably the "segmentation" which is currently done with regard to the size of $\langle g, \psi_{j_l,n}^{p_l} \rangle$. This does not discriminate textures and edges while we may like to have different values of λ for these two kinds of structures. Another possible improvement (at least in cases where the approximation with s is not satisfactory enough) could be to do the exact convolution with s and to modify $S_{\boldsymbol{\lambda}, g, \sigma, \sqrt{\lambda_1}, \sqrt{\lambda_2}}$ in such a way that it does not take this convolution into account. Indeed we only do the convolution with an approximation of s in (9) (the $\lambda_{j_l}^{p_l}$) in order to gain in algorithmic complexity.

4. Second combination of wavelet packet and total variation methods

One of the thing we did not mention in the recall on the Rudin-Osher-Fatemi method of the preceding section is that

the main known drawback of this method is to create staircasing artifacts. This means that it tends to create large homogeneous zones and therefore to erase some textures. This has been studied by several authors among which we can cite [13, 14]. If we look in detail at the arguments given in [14], we see that one of the key properties which causes this staircasing is the fact that we cannot have a "reasonable local[1]" solution to the equation

$$\overline{s} * (s * u - g) = 0\,, \tag{10}$$

where $(\overline{s})_n = (s)_{-n}$. This is, in general, the case since g contains noise and s is regular (for instance a low-pass filter).

These considerations leads us to modify the functional in order to have a data fidelity term whose derivative (the left term in (10)) can be null. With that in mind, in (6), we change the convolution operator in $\|s * u - g\|_2^2$ by an "adaptative convolution", using Proposition 1. More precisely, given a wavelet packet basis of the interval $(\psi^{p_l}_{j_l,n})_{1\le l\le L, 0\le n<2^{-j_l}N}$, we compute some eigenvalues $\boldsymbol{\lambda} = (\lambda^{p_l}_{j_l})_{1\le l\le L}$ (for instance with (5)) in order to approximate the convolution with s by an operator D (defined in Proposition 1).

Given a data $g \in \mathbb{R}^N$, we can define an adaptative convolution by averaging, over translations of u, the operator $\widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta}$ which is defined by the coordinates of its image $\langle \widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta}(u), \psi^{p_l}_{j_l,n}\rangle$ and are equal to

$$\begin{cases} \lambda^{p_l}_{j_l}(\langle u, \psi^{p_l}_{j_l,n}\rangle - \sigma) + \delta\sigma & \text{, if } \langle g, \psi^{p_l}_{j_l,n}\rangle \ge \sigma, \\ \delta\langle u, \psi^{p_l}_{j_l,n}\rangle & \text{, if } \sigma > \langle g, \psi^{p_l}_{j_l,n}\rangle \ge -\sigma, \\ \lambda^{p_l}_{j_l}(\langle u, \psi^{p_l}_{j_l,n}\rangle + \sigma) - \delta\sigma & \text{, if } -\sigma > \langle g, \psi^{p_l}_{j_l,n}\rangle, \end{cases}$$

for $\sigma > 0$ and $\delta > 0$, if $\lambda^{p_l}_{j_l} \neq 0$, and 0, if $\lambda^{p_l}_{j_l} = 0$. We call

$$S_{g,\boldsymbol{\lambda},\sigma,\delta} = 2^{-J} \sum_{k=0}^{2^J-1} \tau_{-k} \circ \widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta} \circ \tau_k\,,$$

where $J = \max_l j_l$.

Note that, in order to define a convex data fidelity term $\|S_{g,\boldsymbol{\lambda},\sigma,\delta}(u) - g\|_2^2$, the criterion which determines whether we do the convolution or not deals with $\langle g, \psi^{p_l}_{j_l,n}\rangle$ and not $\langle u, \psi^{p_l}_{j_l,n}\rangle$, which would appear more natural. Indeed, in this other case $\|A_{\boldsymbol{\lambda},\sigma,\delta}(u) - g\|_2^2$ is not convex as soon as one of the $\lambda^{p_l}_{j_l}$ is lower than 1, which is, in general, the case. Therefore, we propose to minimize, among $u \in \mathbb{R}^N$, the functional

$$TV(u) + \lambda\|S_{g,\boldsymbol{\lambda},\sigma,\delta}(u) - g\|_2^2\,. \tag{11}$$

[1]W. Ring wrote his paper in the continuous framework of an open set $\Omega \subset \mathbb{R}$ (instead of $\{1, ..., N\}$). Therefore, he can assume that the equation $\overline{s} * (s * u - g) = 0$ does not have any solution on any open subset of Ω. The heuristic translation of this hypothesis in our discrete framework could be that a solution of (10) does not contain too much noise.

Note that $S_{g,\boldsymbol{\lambda},\sigma,\delta}$ is affine and that (11) is convex and admits a minimum. Moreover, we can make concerning the uniqueness of this minimum the same conclusion as for (8).

One of the advantages of this functional is that, this time, there exists a reasonably smooth solution u_∞ to

$$S'_{g,\boldsymbol{\lambda},\sigma,\delta}(S_{g,\boldsymbol{\lambda},\sigma,\delta}(u) - g) = 0\,, \tag{12}$$

where $S'_{g,\boldsymbol{\lambda},\sigma,\delta}$ is the derivative of $S_{g,\boldsymbol{\lambda},\sigma,\delta}$. This solution is close to the solution of the wavelet shrinkage method described in [15, 11][2].

Therefore, the role of the parameter λ, σ and δ is clear: σ and δ are used to control the noise and λ is used to control the ringing artifacts. It is also a point which is satisfactory. Indeed, in Rudin-Osher-Fatemi method, when letting λ as a parameter, we, in practice, fix it in order to have a reasonably low amount of noise in homogeneous regions (where the noise is the most visible). Though, we know that the main advantage in the use of the total variation is its ability to remove Gibbs effects (see [7]).

Moreover, with regard to the causes of the staircasing given in [14], the existence of u_∞ cancels the reason of the staircasing. We will see in the experiments that the images restored by means of (11) are indeed free of staircasing.

5. Numerical results

5.1. Description of the data and notations

The experiments are based on two realistic degradation models (the same as the ones presented in [7]) which are derived from satellite imaging. They correspond to two different satellites.

In both cases, the Fourier transform of the impulse response is supported over $[-\pi, \pi] \times [-\pi, \pi]$. Moreover, we will assume the noise Gaussian, even if the real noise is the sum of three noises having different structures. The assumed standard deviation of this Gaussian noise is, in both cases, realistic and gives rise to the same difficulty as the real noise.

- The convolution kernel of the first model is given by

$$\widehat{s_1}(\xi,\eta) = e^{-2\gamma_\xi|\xi| - 2\gamma_\eta|\eta|}\left(\frac{sin(2\xi)}{2\xi}\right)\left(\frac{sin(2\eta)}{2\eta}\right)\left(\frac{sin(\eta)}{\eta}\right), \tag{13}$$

  for $\xi, \eta \in [\pi, \pi]$, where $\gamma_\xi = 0.479$, $\gamma_\eta = 0.450$ and the standard deviation of the noise is $\sigma_1 = 2.4$ (see Figure 1).

[2]In fact, if, in (12), we take $\widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta}$, instead of $S_{g,\lambda,\sigma,\delta}$, u_∞ is the result of the FCNR for a particular set of parameter which depends on $\boldsymbol{\lambda}$, σ and δ.

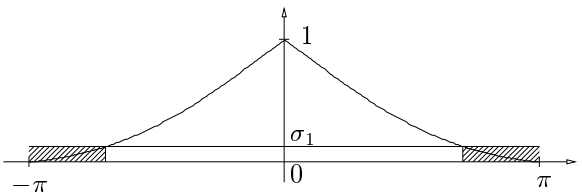

Figure 1. Profile of the Fourier transform of s_1 (see (13)). The hatching represents the frequencies which are, in practice, lost during the degradation.

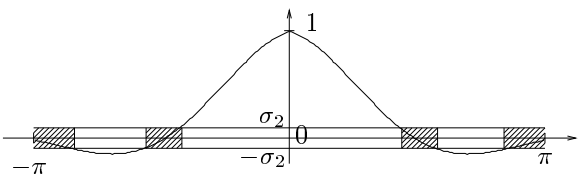

Figure 2. Profile of the Fourier transform of s_2 (see (14)). The hatching represents the frequencies which are, in practice, lost during the degradation.

- The convolution kernel of the second model is given by

$$\widehat{s_2}(\xi,\eta) = e^{-2\gamma_\xi|\xi|-2\gamma_\eta|\eta|} \left(\frac{sin(4\xi)}{4\xi}\right)\left(\frac{sin(4\eta)}{4\eta}\right), \quad (14)$$

  for $\xi,\eta \in [\pi,\pi]$, with the same values for γ_ξ and γ_η. The standard deviation of the noise is $\sigma_2 = 0.5$ (see Figure 2).

We have already shown to evidence in [7] that the main difference between these two convolution kernels is that in the first case the Fourier transform of the convolution kernel only vanishes when one of the Fourier coordinates is in the vicinity of $-\pi$ or π while in the second case we also miss some intermediate frequencies (see the hatched zones on Figure 1 and 2).

We also showed in this paper that variational methods are better suited to this second degradation model (due to there ability to retrieve lost frequencies) even if they tend to erase some textures.

We will compare the results of the methods in front of the reference image, which is the best sampled image, we can expect to recover, given the initial landscape and the sampling rate. (Once again, see its definition in [7].) We display some parts of this reference in the upper right corners of Figures 7, 8.

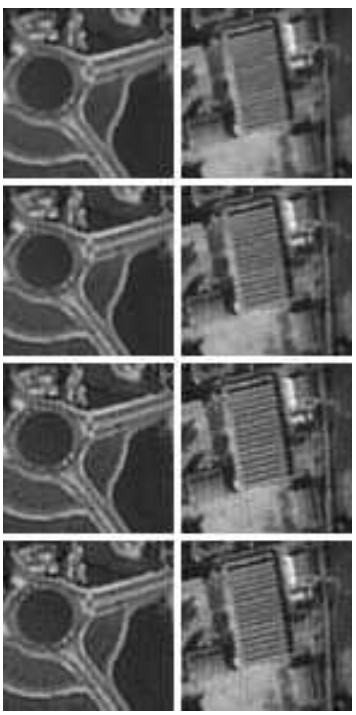

Figure 3. Restoration of an image degraded by (13) when minimizing (8) for $\sigma = 3$, $\lambda_1 = 0$ and (from up to down) $\lambda_2 = 5$, $\lambda_2 = 10$, $\lambda_2 = 60$ and $\lambda_2 = 1000$. (On left and right are displayed two extracted parts of the same image.)

In the following sections, we will show experiments using a wavelet packet basis. This basis is always defined by the mirror tree (see [9]), or its adaptation to s_2, with a spline wavelet (see [12], pp. 236).

5.2. Role of the parameters in the proposed methods

We display on Figure 3 and 4 the result of the method proposed in Section 3 when trying to restore the image obtained by the first model of degradation (see (13)). We have always let $\lambda_1 = 0$ to simplify the study of the role of the other parameters. Note that, in practice, it could be interesting to tune this parameter correctly. However, with $\lambda_1 = 0$, the best set of parameters, we have found, is $\lambda_2 = 10$ and $\sigma = 3$.

In order to illustrate the role of these parameters, we display on Figure 3, restoration for a value of $\sigma = 3$ for different values of λ_2. More precisely, we show:

- on left and right: two different parts of the same restored image;
- from up to down: the result when $\lambda_2 = 5$, $\lambda_2 = 10$, $\lambda_2 = 60$ and $\lambda_2 = 1000$.

Figure 4. Restoration of an image degraded by (13) when minimizing (8) for $\lambda_1 = 0$, $\lambda_2 = 10$ and for (from up to down) $\sigma = 0$, $\sigma = 3$, $\sigma = 7$ and $\sigma = 15$.

We see here that λ_2 has a role similar to the one λ is playing in Rudin-Osher-Fatemi method. The larger is λ_2 the more we can see noise. However, the difference is that here the noise does not completely blows up since wavelet packet coefficients of the blurred image whose modulus are smaller than σ have not been deconvolved (they no longer appear in (8), since $\lambda_1 = 0$). Indeed, for λ_2 large and $\lambda_1 = 0$, the minimization of (8) yields a kind of hard thresholding of the wavelet packet coefficients.

On Figure 4, we represent a part of the restored image for the same parameter $\lambda_2 = 10$, but for, from up to down, $\sigma = 0$, $\sigma = 3$, $\sigma = 7$ and $\sigma = 15$. We clearly see here the fact that σ permits to "segment" the image and to chose what we want to deconvolve. For a σ small, we obtain a result similar to the one obtained by the minimization of Rudin-Osher-Fatemi functional (except the fact that the convolution is approximated by an $\tilde{s}$). When σ increases, we see more the effect of the total variation in regions where the spatial/frequencial localization contains

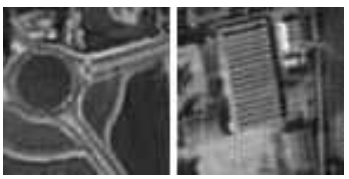

Figure 5. Restoration of an image degraded by (13) when minimizing (11) for $\sigma = 2$ and $\lambda = 1000$. (On left and right are displayed two extracted parts of the same image.)

only noise, while textures and edges are almost not affected by this change. Of course, when σ is too large, The result is similar to the one we would have obtained with Rudin-Osher-Fatemi method for $\lambda = \lambda_1 = 0$ which is in practice too small.

We can make for the method introduced in Section 4 the same experiments as for the one introduced in Section 3. Moreover, we would have almost the same comments and results when letting σ change. However, we illustrate the role of the parameter λ in (11) on Figures 6 and 5.

Indeed, on Figure 6, we display several restoration of the image blurred by the second degradation model (see (14)). All these images are computed for the same parameter $\sigma = 1$, but for different values of λ. More precisely, we display

- On left and right: two different parts of the same restored image.
- From up to down: the restoration for $\lambda = 2$, $\lambda = 8$, $\lambda = 30$ and $\lambda = 1000$.

We clearly see that for $\lambda = 1000$ the result has the same characteristics as the result of a restoration by a wavelet packet method such as the one described in [15, 9]. (These latter are basically some soft thresholding of the wavelet packet coefficient.) It retrieves the texture and contains ringing artifact. When λ decreases, we see that the texture is still preserved but the ringing artifact vanishes. Of course, for a value of λ too small (here $\lambda = 2$), both the texture and the ringing are removed, since the total variation term contributes too much to the functional. However, there is an interval of values in which we preserve the texture and remove the ringing.

We can also understand here a difference between the minimization of (8) and (11) by comparing the lowest row of images on Figure 3 and images of Figure 5. These latter images are extracted from a restoration of an image blurred with (13) by minimizing (11) with $\sigma = 2$ and $\lambda = 1000$. (This σ is the one of the best set of parameters (with $\lambda = 10$).) The result is more noisy for the minimization of (8). This correspond to the intuitive interpretation

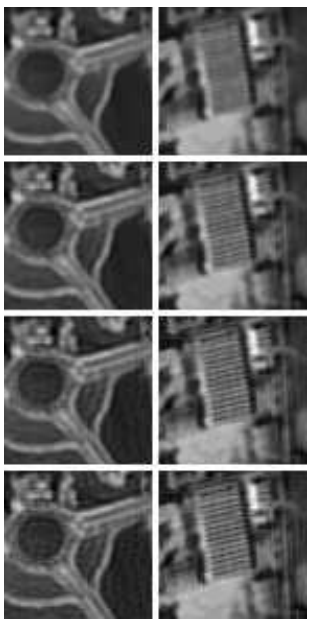

Figure 6. Restoration of an image degraded by (14) when minimizing (11) for $\sigma = 1$ and (from up to down) $\lambda = 2$, $\lambda = 8$, $\lambda = 30$ and $\lambda = 1000$. (On left and right are displayed two extracted parts of the same image.)

of (8) and (11). Indeed, for λ_2 large (in (8)), we compute an "inverse filter" where we have information, while for a λ large in the minimization of (11) we have a result close to a wavelet packet soft thresholding (which contains less noise).

5.3. Comparison of the proposed methods with Rudin-Osher-Fatemi and classical wavelet packet methods

We display on Figures 7 and 8 restorations of an image degraded with (14). Once again, on Figure 7 and 8 (Figure 8 has been sharpened) are two different extracted parts of images which are:

- Up-Left: the blurred image.
- Up-Right: the reference.
- Middle-Left: restoration by mean of a wavelet packet with $\sigma = 1$.
- Middle-Right: restoration by minimizing Rudin-Osher-Fatemi functional with $\lambda = 8$.
- Down-Left: restoration by minimizing (11) with $\lambda = 8$ and $\sigma = 1$.

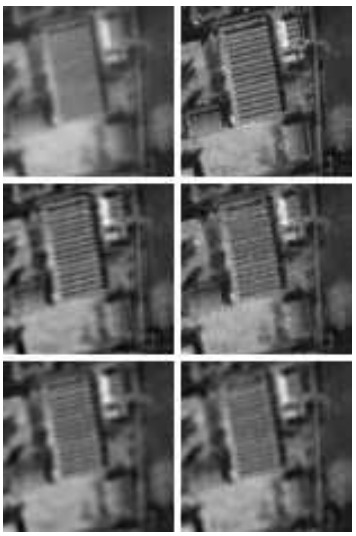

Figure 7. Experiments on the restoration of an image degraded by (14). Up-Left: the blurred image. Up-Right: the reference. Middle-Left: wavelet packet coefficient shrinkage method. Middle-Right: Rudin-Osher-Fatemi method. Down-Left: minimization of (11). Down-Right: minimization of (8).

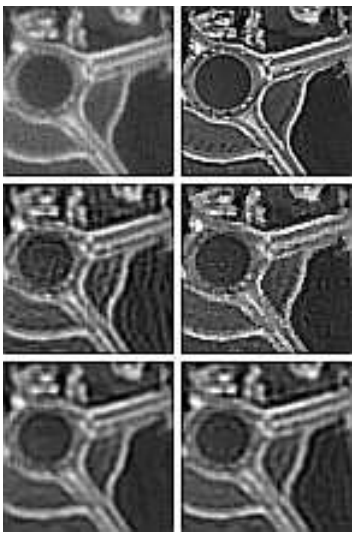

Figure 8. Experiments on the restoration of an image degraded by (14). (The images have been sharpened.) Up-Left: blurred image. Up-Right: the reference. Middle-Left: wavelet packet coefficient shrinkage method. Middle-Right: Rudin-Osher-Fatemi method. Down-Left: minimization of (11). Down-Right: minimization of (8).

- Down-Right: restoration by minimizing (8) with $\lambda_1 = 0$, $\lambda_2 = 40$ and $\sigma = 0.7$.

The only method which permits to restore the texture and avoid ringing is the one which consists in minimizing (11). The fact that the minimization of (8) yields ringing is probably due to the fact that the data fidelity term of (8) tends to create more ringing than the one of (11). Indeed for a value of λ_2 large, the minimum of (8) is to a hard thresholding of the wavelet packet coefficients what the minimum of (11) is to a soft thresholding in wavelet packet coefficients. It therefore contains more ringing artifacts. However, it could be interesting to investigate a little more this aspect since, theoretically, this ringing should not be present (even if it is harder to remove it than in the case of (11)).

For the same experiments with the image degraded by (13), the proposed methods yield better results than the classical ones (the images are available at http://www.math.ucla.edu/∼malgouy). Indeed, they both are free of ringing and properly restore the texture, simultaneously. The main difference between the results of these two methods is that (8) tends to yield sharper result but present some staircasing (even if it is small).

Note also that for both restoration with (8) and (11), we take a value for λ (or λ_2) larger than for the Rudin-Osher-Fatemi restoration and a value of σ smaller than for the wavelet packet restoration. We also take a value for σ larger in the case of (8) than in the case of (11), since in (8) the only denoising in the data fidelity term concerns the small coefficient which are not taken into account if they are smaller than σ.

References

[1] H. Andrews and B. Hunt. *Digital signal processing*. Technical Englewood Cliffs, NJ: Prentice-Hall, 1977.

[2] A. Chambolle and P. Lions. Image recovery via total variation minimisation and related problems. *Numerische Mathematik*, 76(2):167–188, 1997.

[3] A. Chambolle, R. D. Vore, N. Lee, and B. Lucier. Nonlinear wavelet image processing: Variational problems, compression and noise removal through wavelet shrinkage. Technical report, CEREMADE, 1998. short version in: IEEE Trans. Image Processing, Vol. 7, No. 3, pp. 319-335, 1998.

[4] R. Coifman, Y. Meyer, and M. Wickerhauser. Wavelet analysis and signal processing. In *Wavelets and their Applications*, pages 153–178. Jones and Barlett. B. Ruskai et al. eds, 1992.

[5] G. Demoment. Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Transactions on acoustics, speech and signal processing*, pages 2024–2036, 1989.

[6] D. Donoho and I. Johnstone. Minimax estimation via wavelet shrinkage. Technical report, Departement of Stat., Stanford University, 1992.

[7] S. Durand, F. Malgouyres, and B. Rougé. Image deblurring, spectrum interpolation and application to satellite imaging. *Control, Optimisation and Calculus of Variation*, 5(445-475), 2000. A preliminary version is available at http://www.math.ucla.edu/∼malgouy.

[8] T. Kailath. A view of three decades of linear filtering theory. *IEEE transaction on information theory*, IT20(2), March 1974.

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

[10] M. Lindenbaum, M. Fischer, and A. Bruckstein. On gabor's contribution to image enhancement. *PR.*, 27(1):1–8, 1994.

[11] F. Malgouyres. Convolution approximation by mean of an operator diagonal in a wavelet packet basis and application to image deblurring. CAM report at UCLA, available at http://www.math.ucla.edu/∼malgouy, November 2000.

[12] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, Boston, 1998.

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

[14] W. Ring. Structural properties of solutions of total variation regularization problems. Technical report, University of Graz, Austria, 1999. Available at http://www.kfunigraz.ac.at/imawww/ring/.

[15] B. Rougé. Fixed chosen noise restauration (fcnr). In *IEEE 95 Philadelphia*, 1995.

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

[17] S. Tramini. *Problemes inverses et EDP pour le decodage et la convolution d'images*. PhD thesis, Université de Nice, Sophia-Antipolis, November 1999.