A framework for image deblurring using wavelet packet bases*

François Malgouyres†

July 19, 2001

Abstract

We show in this paper that the average over translations of an operator diagonal in a wavelet packet basis is a convolution. We also show that an operator diagonal in a wavelet packet basis can be decomposed into several operators of the same kind, each of them being better conditioned. We investigate the possibility of using such a convolution to approximate a given convolution (in practice an image blur). Then we use these approximations to deblur images. First, we show that this framework permits to redefine existing deblurring methods. Then, we show that it permits to define a new variational method which combines the wavelet packet and the total variation approaches. We argue and show on experiments that this permits to avoid the drawbacks of both approaches which are respectively the ringing and the staircasing.

1 Introduction

This paper is mainly concerned with image deblurring and with the use of wavelet-packet bases for this purpose. More precisely, we will show that the average over translations of an operator which is diagonal in a wavelet packet basis is a convolution. We will investigate several 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

$$u_0 = s_1 * u + n\,, \tag{1}$$

where s_1 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

$$\widehat{u_0} = \hat{s_1}\hat{u} + \hat{n}\,, \tag{2}$$

where we note with a hat the Fourier transform of a function. We clearly see here that, since at some points $\hat{s}_1$ can can be very small or even take the value zero, this problem is ill-posed. Moreover, due to (2), we can regularize u_0 by independently modifying each of its Fourier coefficient in such a way that after the deconvolution the noise does not blow-up. This is one of the reason why linear filters (such as the Wiener filter, see [1]) have been so popular. On the other hand, for the purpose of denoising, wavelet style bases are now very popular because of there ability to yield a "sparse representation" of the information contained in an image. In order to adapt wavelet style methods to image deblurring, we need to find a way to express (1) on the coordinates of the image in a wavelet style basis in a way similar to (2). This is why, due to their good frequencial localization, wavelet packet bases appear to be a natural framework for image deblurring. Remark however that when the convolution is a high pass filter (such as for the inversion of the Radon transform) we would use a wavelet basis. In this case our approach is an alternative to the wavelet/vaguelet decomposition (see [11]).

As far as we know, the possibility to use wavelet packet bases for image deblurring has first been noticed by B. Rougé. It has then been studied in several articles (see [13, 15, 16, 23]). The methods proposed in

*This work has been partially financed by the CNES through the ENS Cachan, by NSF DMS-9973341 and by ONR grant N00014-96-1-0277.

†UCLA Dept of Mathematics, 6363 Math science building, Box 951555, Los Angeles, CA 90095-1555, USA; and CMLA, ENS Cachan, 61 avenue du président Wilson, 94235 Cachan Cedex, France;
http://www.math.ucla.edu/∼malgouy
malgouy@math.ucla.edu

these articles are based on a shrinkage of the wavelet packet coefficients similar to the wavelet shrinkage approach of Donoho and Johnstone (see [12]) for the purpose of denoising. A part of this paper gives a new interpretation of the "wavelet packet based deblurring" and permits a better understanding of some of the parameters of these methods.

There is an abundant literature on image deblurring. The reader is referred to [1] for most of the linear methods and to [10, 14] for overviews on the subject. Among these and despite numerous links binding them (see [6]), we will distinguish two kinds :

- The one based on the decomposition in an appropriate basis. The most famous in this category are probably the Wiener filter (see [1]), the wavelet shrinkage (see [12]) and its adaptation to deblurring (see [15, 23]).
- The variational ones : where we can mention some based on the entropy (see [10] and references there) and more recently the total variation [24].

The paper is organized as follows. (For simplicity, all our results are stated in the case of 1D signals but they can be generalized to higher dimensions.)

We will give in Section 2 the statement of the main result of the paper which is that the averaging over translations of an operator diagonal in a wavelet packet basis is a convolution. More precisely, if we note $\widetilde{D}$ an operator which is diagonal in a wavelet packet basis of depth J, we define the operator D by

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

where $u \in l^2(\mathbb{Z})$ and τ_k represents the translation operator of $k \in \mathbb{Z}$. We show that this operator is a convolution and give the explicit form of its kernel. Note that this proposition is also a new argument in favor of the cycle spinning introduced in [8]. We also show that an operator which is diagonal in a wavelet packet basis can be written as the composition of several operators which are diagonal in other wavelet packet bases. We show that this property permits to justify the multi-level thresholding proposed by B. Rougé in [13].

In Section 3, we expose two models for image deblurring which are based on the results of Section 2. The first one is equivalent to the usual wavelet-packet shrinkage. In the second one, we approximate (1) using an operator of the form (3) to improve the conditioning of the data fidelity term in the method of Rudin-Osher-Fatemi (ROF). If we present this modification under the point of view of ROF functional, this permits to avoid staircasing while if we present it under the wavelet packet shrinkage point of view, this permits to avoid ringing artifacts.

We display in Section 4 several experiments which show to evidence that the approximation of a convolution by another convolution defined by mean of (3) is often a good approximation and that its analysis permits a better understanding of the existing wavelet packet shrinkage algorithms. We also show the importance of the average over translations and the advantage of the multi-level thresholding. We finish with some comparison between two wavelet packet shrinkage, the ROF model and our modification of this latter. We also describe the role of the parameters in the modified ROF method.

2 Approximation of the convolution in a wavelet packet basis

2.1 Wavelet packet bases

Let us now define the notations that we will use in order to describe wavelet packet bases. Once again, for simplicity, we only describe wavelet packet bases in the case of function of $\mathbb{R}$, higher dimensional cases can be deduced from this one by taking tensor products[1]. For more details the reader is referred to [9] or to Section 8 of [18].

[1] Remark that we could imagine smarter ways to generalize our result to higher dimensions. One can for instance take ideas from ridgelets (see [2]) or complex wavelet packet transform (see [17]).

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 $p \in \{0, ..., 2^j - 1\}$

$$\psi_{j+1}^{2p}(x) = \sum_{n=-\infty}^{\infty} h_n \psi_j^p(x - 2^j n)\,, \tag{4}$$

and

$$\psi_{j+1}^{2p+1}(x) = \sum_{n=-\infty}^{\infty} g_n \psi_j^p(x - 2^j n)\,. \tag{5}$$

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 [18]) $(p_l, j_l)_{1 \le l \le L}$, $\{\psi_{j_l,n}^{p_l}\}_{n \in \mathbb{Z}, 1 \le l \le L}$, is an orthonormal basis of $\mathbf{W}_0^0$. Since in the following we will mostly use wavelet packet bases associated with a particular tree and in order to simplify notations, we will denote the leaves of a tree by $t_l = (p_l, j_l)$ and index the elements of a wavelet packet basis defined by a tree $(t_l)_{1 \le l \le L}$ using the notation

$$\psi_{t_l,n} = \psi_{j_l,n}^{p_l}\,,$$

for $l \in \{1, ..., L\}$ and $n \in \mathbb{Z}$.

In the following, we will identify any sequence $u = (u_n)_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})$ with $\tilde{u} = \left(\sum_{n \in \mathbb{Z}} u_n \psi_{0,n}^0\right) \in \mathbf{W}_0^0$. Therefore, noting $u_{j,n}^p = \langle \tilde{u}, \psi_{j,n}^p \rangle$, we can deduce from (4) and (5) that

$$u_{j+1,n}^{2p} = \sum_{m \in \mathbb{Z}} h_m\ u_{j,2n+m}^p = \bar{h} * u_j^p(2n) \tag{6}$$

where, for any $n \in \mathbb{Z}$, $\bar{h}_n = h_{-n}$, and

$$u_{j+1,n}^{2p+1} = \sum_{m \in \mathbb{Z}} g_m\ u_{j,2n+m}^p = \bar{g} * u_j^p(2n) \tag{7}$$

where, for any $n \in \mathbb{Z}$, $\bar{g}_n = g_{-n}$.

Therefore, for any admissible tree $(p_l, j_l)_{1 \le l \le L}$, we can recursively define a kernel $H_{j_l}^{p_l}$ such that

$$u_{j_l,n}^{p_l} = H_{j_l}^{p_l} * u_0^0(2^{j_l} n)\,.$$

Similarly, we can rebuild $u_{j,n}^p$ from $u_{j+1,n}^{2p}$ and $u_{j+1,n}^{2p+1}$ using

$$u_{j,n}^p = \sum_{m \in \mathbb{Z}} h_{n-2m}\ u_{j+1,m}^{2p} + \sum_{m \in \mathbb{Z}} g_{n-2m}\ u_{j+1,m}^{2p+1}\,.$$

In other words, noting

$$\check{u}_n = \begin{cases} u_{\frac{n}{2}} & \text{, if } n \text{ is even,} \\ 0 & \text{, if } n \text{ is odd,} \end{cases}$$

for any $u \in l^2(\mathbb{Z})$, we have

$$u_{j,n}^p = (h * (u_{j+1}^{2p})^\vee)_n + (g * (u_{j+1}^{2p+1})^\vee)_n\,. \tag{8}$$

2.2 The approximation of a convolution

The first idea is that, due to their frequencial localization, it is possible to approximate the convolution in a wavelet packet basis. Therefore, for a suitable basis $\{\psi_{t_l,n}\}_{n\in\mathbb{Z},1\leq l\leq L}$ and suitable eigenvalues $(\lambda_{t_l})_{1\leq l\leq L}$, we can define the linear operator $\widetilde{D}$ by

$$\langle \widetilde{D}(u), \psi_{t_l,n}\rangle = \lambda_{t_l}\langle u, \psi_{t_l,n}\rangle$$

for $u \in l^2(\mathbb{Z}), l \in \{1, ..., L\}$ and $n \in \mathbb{Z}$. Note that the eigenvalues λ_{t_l} do not depend on n since we consider a uniform blur. However, it could be interesting to study the possibility of using the wavelet packet framework for deblurring cases where the blurring kernel slowly varies with the location in the image.

One of the very important property we loose, when approximating the convolution by such a $\widetilde{D}$ is the translation invariance. In practice this yields to strong and unacceptable artifacts on textures and in the vicinity of edges (see Section 4).

The first simple way to solve this drawback is to use the Shannon wavelet (see [18], pp. 245). In this case, we have $\hat{h} = \sqrt{2}\,1_{[-\frac{\pi}{2},\frac{\pi}{2}]}$ and $\hat{g} = \sqrt{2}\,1_{[\frac{\pi}{2},\frac{3\pi}{2}]}$ and therefore the wavelet packet analysis itself is translation invariant. The problem with the Shannon wavelet is that it has a slow decay at infinity and therefore, in a noisy case, poorly decorrelates information and noise.

Another simple way to turn around this drawback is to average $\widetilde{D}$ over some translations of the image. The following proposition proves that the so defined operator is a convolution and gives the form of its convolution kernel[2].

Proposition 1 *Let $(t_l)_{1\leq l\leq L} = (p_l, j_l)_{1\leq l\leq L}$ be an admissible tree and let $(\psi_{t_l,n})_{n\in\mathbb{Z},1\leq l\leq L}$ be a wavelet packet basis. Let $\widetilde{D}$ be a linear operator continuous from $l^2(\mathbb{Z})$ into $l^2(\mathbb{Z})$, diagonal in the basis $(\psi_{t_l,n})_{n\in\mathbb{Z},1\leq l\leq L}$. Assume moreover that for $n \in \mathbb{Z}$ and $l \in \{1, \cdots, L\}$ the eigenvalue associated with the eigenvector $\psi_{t_l,n}$ does not depend on n (we note it $(\lambda_{t_l})_{n\in\mathbb{Z},1\leq l\leq L}$). Then, the operator D defined, for any $u \in l^2(\mathbb{Z})$, by*

$$D(u) = 2^{-J}\sum_{k=0}^{2^J-1} \tau_{-k}\circ\widetilde{D}\circ\tau_k(u)\,, \tag{9}$$

where $J = \max_{1\leq l\leq L} j_l$ and τ_k represents the translation operator of $k \in \mathbb{Z}$, is a continuous convolution 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

$$\widehat{\tilde{s}}(\xi) = \sum_{l=1}^{L} \lambda_{t_l}\frac{|\widehat{H_{t_l}}(\xi)|^2}{2^{j_l}}\,. \tag{10}$$

The proof of this proposition is given in appendix.

Proposition 1 ensure that 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.

The issue which has now to be considered is to find a way to design $\tilde{D}$ in order to achieve a good approximation of the convolution with a kernel s.

The formula (10) can of course be used to solve this problem. We can for instance imagine an optimization process which would minimize the error between $\tilde{s}$ and s. Note that we can also use (10) in such a way that the approximated convolution avoids specific artifacts. For instance, when approximating a kernel which inverts the convolution with a kernel s_1 and in order to avoid Gibbs phenomena, we could determine $\tilde{D}$ in such a way that $\tilde{s} * s_1$ is positive[3].

Note that until now people were designing $\tilde{D}$ in an empirical way. The wavelet was chosen in such a way that it has a good frequencial localization (in practice a cubic spline, see [18], pp. 236). The tree was either a fully decomposed tree of a sufficiently large depth (see [13]) or the mirror tree (see [15]).

[2]Remark that the average over translations of any linear operator is a linear and translation invariant operator. Therefore it is a convolution. The main interests of Proposition 1 is due to the nature of wavelet packet bases (sparse representation of the image and frequencial localization). Moreover, we only have to average over 2^J translations, with $J = \max_{1\leq l\leq L} j_l$.

[3]Like this the convolution of a Heavyside function with s_1 and then s does not overshoot.

There have been several attempts to determine for a given basis $(\psi_{t_l,n})_{1\leq l\leq L, n\in\mathbb{Z}}$ some good eigenvalues[4] $(\lambda_{t_l})_{1\leq l\leq L}$. It is for instance easy to check that

$$\lambda_{t_l} = \langle s * \psi_{t_l,n}, \psi_{t_l,n}\rangle\,, \tag{11}$$

permits to minimize $\|S-\widetilde{D}\|_2$ (where $S(u) = s*u$ is the operator we want to approximate) and can therefore be considered as a good candidate. Note that these λ_{t_l} do indeed not depend on n since the convolution with s is translation invariant.

Figure 4 and 5 represent the Fourier transforms of two convolution kernels and the corresponding convolution kernel after the approximation by some diagonal operators in different wavelet packet bases. We see here that, in the case of the Shannon wavelet (the dotted line), the initial kernel is approximated by a kernel which is constant on dyadic intervals of the Fourier domain (this is also visible on (10) and gives the intuitive meaning of $(\lambda_{t_l})_{1\leq l\leq L}$). Therefore, as long as the Fourier transform of the initial kernel does not vary too much inside these dyadic intervals, the approximation of the convolution in a wavelet packet basis will yield good results. It seems therefore a good idea to choose the tree which defines the basis $\{\psi_{t_l,n}\}_{n\in\mathbb{Z}, l\in\{0,\ldots,L\}}$ according to this criterion.

Let us now investigate the issue of the spatial localization of the wavelet packet basis (versus its frequencial one). Indeed, in the case of the deconvolution (s is the pseudo-inverse of a low pass filter s_1), we usually want the elements of the wavelet packet basis to have: a good frequencial localization, in order to define a good approximation of the deconvolution; and a good spatial localization, in order to properly separate information and noise. As far as we know there have been two attempts to cope with these incompatible properties. The first one consists in finding the "best basis", that is the basis which separates the most the information from the deconvolved noise (see [15, 16]). The second one was introduced in [13] and consists in shrinking the image at different scales.

The following proposition, despite its simplicity, permits to justify and generalize this second approach. Let us first define a partial order among admissible trees.

Definition 1 *Let $(t_l)_{1\leq l\leq L} = (p_l, j_l)_{1\leq l\leq L}$ and $(t'_l)_{1\leq l\leq L'} = (p'_l, j'_l)_{1\leq l\leq L'}$ be two admissible trees, we say that*

$$(t_l)_{1\leq l\leq L} \geq (t'_l)_{1\leq l\leq L'}$$

if and only if there exists a partition of $\{1,\ldots,L\}$ into L' subsets $(I_{l'})_{1\leq l'\leq L'}$ such that for any $l'\in\{1,\ldots,L'\}$,

$$\mathbf{W}_{j'_{l'}}^{p'_{l'}} = \oplus_{l\in I_{l'}} \mathbf{W}_{j_l}^{p_l}.$$

This relation simply means that the elements of $\{\psi_{t_l,n}\}_{n\in\mathbb{Z}, 1\leq l\leq L}$ correspond to a higher level of decomposition than the ones of $\{\psi_{t'_l,n}\}_{n\in\mathbb{Z}, 1\leq l\leq L'}$. Note that if the admissible trees are indexed with regard to their position in the binary tree (for instance from the left to the right) then the $I_{l'}$ are of the form $\{t_{l'-1},\ldots,t_{l'}-1\}$ with $1 = t_0 < \ldots < t_{l'} < t_{l'+1} < \ldots < t_{L'} = L+1$.

Using this definition, we can state,

Proposition 2 *Let $\{\psi_{t_l,n}\}_{n\in\mathbb{Z}, 1\leq l\leq L}$ be a wavelet packet basis and $\widetilde{D}$ be an operator linear, continuous and diagonal in the basis $\{\psi_{t_l,n}\}_{n\in\mathbb{Z}, 1\leq l\leq L}$, which goes from $l^2(\mathbb{Z})$ into itself. If we note $\lambda_{t_l,n}$ the eigenvalue of $\widetilde{D}$ associated with the eigenvector $\psi_{t_l,n}$, then for any admissible tree $(p'_l, j'_l)_{1\leq l\leq L'}$, such that $(t_l)_{1\leq l\leq L} \geq (t'_{l'})_{1\leq l'\leq L'}$, and any $(\mu_{t'_{l'}})_{1\leq l'\leq L'} \in (\mathbb{R}\setminus\{0\})^{L'}$, we have*

$$\widetilde{D} = \widetilde{D}_1 \circ \widetilde{D}_2\,,$$

where $\widetilde{D}_1$ and $\widetilde{D}_2$ are linear and continuous from $l^2(\mathbb{Z})$ into itself and are such that: for any $n\in\mathbb{Z}$ and any $l'\in\{1,\ldots,L'\}$,

$$\widetilde{D}_1(\psi_{t'_{l'},n}) = \mu_{t'_{l'}}\ \psi_{t'_{l'},n}\,,$$

[4] One can refer to [13] for examples.

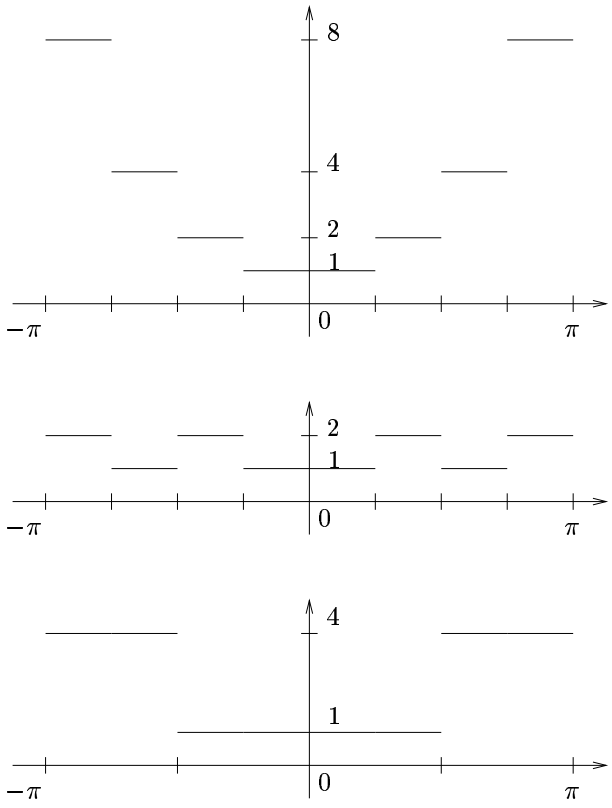

Figure 1: Heuristically, Proposition 2 permits to deconvolve a few as possible when the basis has a good localization in Fourier domain. Example with convolution : The convolution with the kernel on top is equal to the composition of the convolution with the two other kernels.

and for any $n \in \mathbb{Z}$, *any* $l' \in \{1, ..., L'\}$ *and any* $l \in I_{l'}$ *(we take here the notations of Definition 1),*

$$\widetilde{D}_2(\psi_{t_l,n}) = \frac{\lambda_{t_l,n}}{\mu_{j'_{l'}}^{p'_{l'}}} \, \psi_{t_l,n} \, .$$

The proof of this result is given in appendix.

Remark that in Proposition 1 the $\lambda_{t_l,n}$ do not depend on n and that Proposition 2 could be consequently simplified.

This proposition proves that, the operator $\widetilde{D}$, diagonal in a wavelet packet basis can be written as a composition of similar operators $\widetilde{D} = \widetilde{D}_1 \circ \widetilde{D}_2 \circ \widetilde{D}_3 \circ \cdots$. In practice, it can be used to obtain some $\widetilde{D}_i$ which are better conditioned than $\widetilde{D}$. Moreover, in a noisy case, we can apply the $\widetilde{D}_i$ and smooth the image successively. The advantage of this approach is that the operator $\widetilde{D}_i$, for small indexes i, separate the noise from the information very efficiently since they correspond to low decomposition levels and therefore have a good spatial localization.

We present on Figure 1 a simple case of such a decomposition (with Shannon wavelet). In practice, we choose to "deconvolve" as few as possible when the frequencial localization of the wavelet packet basis is good (in which case the wavelet packet basis is similar to a Fourier basis and poorly decorrelates noise and information). This can in practice be achieved by decomposing $\tilde{D}$ according to the following process.

Let us consider the case of the approximation of a deconvolution using the averaging over translations of an operator $\widetilde{D}$ diagonal in a basis $\{\psi^p_{j_0,n}\}_{n\in\mathbb{Z},0\leq p<2^{j_0}}$. For simplicity, we assume here that all the $\lambda_{(p,j_0)}$ are positive. We can let $\mu^p_{j_0} = \lambda_{(p,j_0)}$ for $p \in \{0, ..., 2^{j_0} - 1\}$ and then recursively define $\mu^p_{j-1} = min(|1 - \mu_j^{2p+1}|, |1 - \mu_j^{2p}|)$ for $p \in \{0, ..., 2^{j-1} - 1\}$ and for $j = j_0, j_0 - 1, \cdots, 1$ and let $\mu^0_0 = 1$ (the μ_j's have to be understood as the remaining convolution at the level j). Therefore, we can decompose

$\widetilde{D} = \widetilde{D}_1 \circ \cdots \circ \widetilde{D}_{j_0}$ where the $\widetilde{D}_j$ are diagonal in the basis $\{\psi^p_{j,n}\}_{n\in\mathbb{Z}, 0\leq p<2^j}$, with the eigenvalues $\frac{\mu_j^{2p+1}}{\mu_{j-1}^p}$ and $\frac{\mu_j^{2p}}{\mu_{j-1}^p}$ respectively associated with the eigenvectors $\psi^{2p+1}_{j,n}$ and $\psi^{2p}_{j,n}$, for $p \in \{0, ..., 2^{j-1}-1\}$. This decomposition permits to deconvolve as few as possible at coarse scales, where the spatial localization is weak.

3 Application to the issue of image deconvolution

3.1 The Fixed Chosen Noise Restoration

As we said in the introduction, the approximation of the convolution, by mean of the wavelet packet decomposition of the image, permits to use the ability of these decompositions to yield a sparse representation of the image. For instance, in the case of the deconvolution, if we define a "pseudo-inverse" (by any appropriate mean[5]) r of the convolution kernel s_1.

Once we have chosen an appropriate wavelet packet basis $\{\psi_{t_l,n}\}_{n\in\mathbb{Z}, 1\leq l\leq L}$ and chosen a sequence $\boldsymbol{\lambda} = (\lambda_{t_l})_{1\leq l\leq L}$ of real numbers which permit to approximate the convolution with r efficiently (or so that it defines a convolution which is an acceptable "pseudo-inverse"). According to Proposition 1 and Proposition 2, the convolution $r * u$ is approximated by

$$D(u) = 2^{-J} \sum_{k=0}^{2^J-1} \tau_{-k} \circ \widetilde{D}_1 \circ \cdots \circ \widetilde{D}_J \circ \tau_k \,,$$

where

$$\langle \widetilde{D}_1 \circ \cdots \circ \widetilde{D}_J(u), \psi_{t_l,n}\rangle = \lambda_{t_l} \langle u, \psi_{t_l,n}\rangle \,,$$

for $l \in \{1, \cdots, L\}$ and $n \in \mathbb{Z}$. In practice, we defined here the $\widetilde{D}_i$ according to the procedure which was described at the end of the preceding section.

Remark now that when u contains noise we want to regularize $D(u)$. In order to do so, we improve the conditioning of all the operators[6] $\tilde{D}_i$. To avoid notations, we only consider in the next formula the improvement of $\tilde{D}$. So we replace it by

$$\langle \widetilde{A}_{\boldsymbol{\lambda},\sigma}(u), \psi_{t_l,n}\rangle = \begin{cases} \lambda_{t_l}\langle u, \psi_{t_l,n}\rangle & \text{, if } |\langle u, \psi_{t_l,n}\rangle| \geq \sigma \text{ or } \lambda_{t_l} = 0 \\ \langle u, \psi_{t_l,n}\rangle & \text{, if} |\langle u, \psi_{t_l,n}\rangle| < \sigma \text{ and } \lambda_{t_l} \neq 0 \end{cases}$$

for $\sigma > 0$, $n \in \mathbb{Z}$ and $l \in \{1, ..., L\}$.

Heuristically, we first segment information and noise according to the size of $|\langle u, \psi_{t_l,n}\rangle|$ and then convolve (partially since $\widetilde{D}_i$ is involved instead of $\tilde{D}$) the information and leave the noise unchanged. Due to this heuristic we will call "adaptative convolution" a regularized version of $D(u)$ of the kind described above. Note that according to the usual framework of wavelet denoising, we can for instance take $\sigma = \sigma_b\sqrt{2\log N}$ where σ_b is the standard deviation of the noise. However, in practice we will leave it as a parameter.

Of course, it is in general preferable to have a continuous operator, instead of $\widetilde{A}_{\boldsymbol{\lambda},\sigma}$. We can moreover introduce a parameter, $\delta \in [0, 1]$ by which we multiply the small coefficients (which are mostly noise). Therefore, noting the soft thresholding function

$$f_\sigma(t) = \begin{cases} t + \sigma & \text{, if } t \leq -\sigma \\ 0 & \text{, if } -\sigma \leq t \leq \sigma \\ t - \sigma & \text{, if } \sigma \leq t \,, \end{cases}$$

[5] By pseudo-inverse, we mean any kernel r such that $r * s_1$ is close to the identity (restricted to $l^2(\mathbb{Z}) \setminus Ker(s_1)$) which would by the way satisfy suitable properties (for instance $r * s_1 \geq 0$ or/and spatial localization), depending on the user's expectations.

[6] This is similar to the multi-level wavelet packet shrinkage which has been described in [13].

we will prefer to $\widetilde{A}_{\boldsymbol{\lambda},\sigma}$, an operator of the kind

$$\langle A_{\boldsymbol{\lambda},\sigma,\delta}(u), \psi_{t_l,n}\rangle = \begin{cases} 0 & \text{, if } \lambda_{t_l} = 0, \\ (\lambda_{t_l} - \delta) f_\sigma(\langle u, \psi_{t_l,n}\rangle) + \delta\langle u, \psi_{t_l,n}\rangle, & \text{if } \lambda_{t_l} \neq 0, \end{cases} \tag{12}$$

for $u \in L^2(\mathbb{R})$, $\sigma > 0$, $\delta \in [0,1]$, $n \in \mathbb{Z}$ and $l \in \{1, ..., L\}$ (for simplicity of notations, we have only considered the case $\lambda_{t_l} \geq 0$). In practice δ does not play an important role but allowing $\delta \neq 0$ (in the usual thresholding $\delta = 0$) permits to define an invertible operator. This will be useful in the next section.

We remark that the average over the translations of the usual wavelet thresholding methods (called cycle spinning, see [8]) falls under the scope of "adaptative convolutions". This framework therefore provides a new tool to understand these algorithms.

This "adaptative convolution" is probably the most natural and immediate application of the results stated in the preceding section. In this case, our work is only a new framework for these methods. However, we will show in the experiments that this permits to better understand them.

3.2 A modification of Rudin-Osher-Fatemi functional

We are now going to introduce another 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 (ROF) in [24]. In order to have some well defined variational problems, we are forced to 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 ROF method.

ROF method consists in minimizing, for $N \in \mathbb{N}$ and a data $g \in \mathbb{R}^N$, the functional

$$TV(u) + \mu \|s_1 * u - g\|_2^2 , \tag{13}$$

among $u \in \mathbb{R}^N$, where μ can be interpreted as a Lagrange multiplier (see [5]) and the total variation is defined by

$$TV(u) = \sum_{n=0}^{N-1} |u_{n+1} - u_n| .$$

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.

On the other hand, its main drawback is that it tends to create staircasing artifacts and therefore to remove some textures. This has been studied by several authors among which we can cite [21, 22].

If we look in detail at the arguments given in [22], we see that one of the key properties which causes this staircasing is the fact that we cannot have a "reasonable local solution[7]" to the equation

$$\overline{s_1} * (s_1 * u - g) = 0 , \tag{14}$$

where $(\overline{s_1})_n = (s_1)_{-n}$. This is, in general, the case since g contains noise and s_1 is regular (for instance a low-pass filter).

These considerations lead us to modify the functional in order to have a data fidelity term whose derivative (the left term in (14)) can be null. With that in mind, we improve the conditioning of the

[7] W. Ring wrote his paper in the continuous framework of an open set $\Omega \subset \mathbb{R}$ (instead of $\{1, ..., N\}$). In this framework, the key argument he gave to explain the staircasing is that if there exists a solution which is differentiable and monotone on an open subset of Ω, then we must have on this subset,

$$2\mu\, \overline{s_1} * (s_1 * u - g) = -\left(\frac{u'}{|u'|}\right)' = 0 .$$

Moreover, this is in practice impossible since $s_1 * u$ is smooth and g contains noise. Therefore, he concludes that the absolutely continuous part with regard to the Lebesgue measure of the derivative of the result of ROF method is always zero.

Note that the fact that the staircasing is related to the existence of noise has also been noticed in [4].

The heuristic translation of the hypothesis that $\overline{s_1} * (s_1 * u - g) = 0$ is impossible on any open set in our discrete framework could be that we do not have any "reasonable local solution".

convolution operator in $\|s_1 * u - g\|_2^2$ using an "adaptative convolution" similar to the one defined by (12). (The difference with (12) is that in order to have a convex functional, we segment the information and the noise with regard to g while applying the convolution to u.) More precisely, given a wavelet packet basis[8] $(\psi_{t_l,n})_{1\leq l\leq L, 0\leq n<2^{-j_l}N}$ and some eigenvalues $\boldsymbol{\lambda} = (\lambda_{t_l})_{1\leq l\leq L}$ such that D (see Proposition 1) approximates the convolution with s_1.

Given a data $g \in \mathbb{R}^N$, we can define an adaptative convolution by averaging, over translations of u, the operator $\widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta}$

$$\langle \widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta}(u), \psi_{t_l,n}\rangle = \begin{cases} 0 & \text{, if } \lambda_{t_l} = 0, \\ \lambda_{t_l}(\langle u, \psi_{t_l,n}\rangle - \sigma) + \delta\sigma & \text{, if } \langle g, \psi_{t_l,n}\rangle \geq \sigma \text{ and } \lambda_{t_l} \neq 0, \\ \delta\langle u, \psi_{t_l,n}\rangle & \text{, if } \sigma > \langle g, \psi_{t_l,n}\rangle \geq -\sigma \text{ and } \lambda_{t_l} \neq 0, \\ \lambda_{t_l}(\langle u, \psi_{t_l,n}\rangle + \sigma) - \delta\sigma & \text{, if } -\sigma > \langle g, \psi_{t_l,n}\rangle \text{ and } \lambda_{t_l} \neq 0, \end{cases}$$

for $\sigma > 0$ and $\delta > 0$ (once again, for simplicity of notations, we have only considered the case $\lambda_{t_l} \geq 0$). We call it

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

where $J = \max_l j_l$.

Therefore, we propose to minimize, among $u \in \mathbb{R}^N$, a functional of the kind

$$TV(u) + \mu \|S_{g,\boldsymbol{\lambda},\sigma,\delta}(u) - g\|_2^2 \,. \tag{15}$$

Note that $S_{g,\boldsymbol{\lambda},\sigma,\delta}$ is affine and that therefore (15) is convex and admits a minimum. 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 [5, 13].

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

$$S'_{g,\boldsymbol{\lambda},\sigma,\delta}(S_{g,\boldsymbol{\lambda},\sigma,\delta}(u) - g) = 0 \,, \tag{16}$$

where $S'_{g,\boldsymbol{\lambda},\sigma,\delta}$ is the derivative of $S_{g,\boldsymbol{\lambda},\sigma,\delta}$.

Moreover, this solution is close to the solution of the wavelet shrinkage method described in the preceding section. In fact, if in (16) we take $\widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta}$ instead of $S_{g,\lambda,\sigma,\delta}$, and if $\delta \neq 0$ and $\lambda_{t_l} \neq 0$, for any $l \in \{1, ..., L\}$,

$$u_\infty = A_{(\frac{1}{\lambda_{t_l}})_{1\leq l\leq L}, \frac{\sigma}{\delta}, \frac{1}{\delta}}(g)$$

is a solution to (16). We will see in the experiments in Section 4.4 that the image restored by means of (15) are indeed free of staircasing.

Therefore, the role of the parameter μ, σ and δ is clear: σ and δ are used to control the noise and μ is used to control the ringing artifacts. The role of the parameters is highlighted in Section 4. Note that the fact that μ controls the ringing artifacts is a point which is satisfactory. Indeed, in practice, in ROF method we fix μ 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 [13]). That is one of the oddness which is solved by our new approach.

A drawback of this model is of course that it involves two parameters (δ is easy to tune) which makes it harder to tune. However, in practice, we tune the two parameters separately (first σ then μ) which makes. Moreover, this drawback is not very important when the degradation is known and fixed (for instance in the case of satellite images). Indeed, in this case, we only have to tune the parameters once.

Remark: We have chosen here to present (15) under a variational point of view. We are conscious of the fact that (15) can appear redundant to readers who are usually interested in wavelet shrinkage.

[8] This time we have to take a wavelet packet basis of the interval.

Indeed, in the case of denoising, one can consider the characterization of Besov spaces by semi-norms on wavelet coefficients to show that wavelet shrinkage algorithms are equivalent to minimization problems similar to (13)[9]. However, the drawback of these methods in the case of deblurring is that they cannot recover lost frequencies (see [13]) (we can however mention the attempt to oversample images by means of wavelet transforms made in [3]). Therefore, it seems interesting to reintroduce the total variation term for spatial/frequencial location where the regularity needed by the Besov semi-norm is too important. We will see in the section devoted to the experiments (Section 4.4) that (15) permits to avoid ringing artifacts where wavelet packet shrinkage method does not.

In the experiments presented in Section 4.4 we have computed a solution to (15) by mean of a gradient algorithm with an optimal step. This means that at each iteration we compute the gradient of the functional and then compute the optimal move in that direction in order to make (15) decrease. We could probably have a better algorithm by adapting methods such as the ones introduced in [7, 19]

Compared to the usual ROF algorithm, the computational cost increases due to the translations in the operator $S_{g,\boldsymbol{\lambda},\sigma,\delta}$. Fortunately, in practice, we only need to average over four translations of u to obtain a sufficiently nice approximation of the convolution.

4 Numerical results

This section is split into four parts. They are organized as follow. The first part describes the data and notations which permit to understand the experiments of the other sections. In the second part, we display experiments which show that we can approximate a convolution operator efficiently in a wavelet packet basis. In the third part, we show the importance of Proposition 2. In the last section, we display some results on the role of the parameters in the modification of ROF exposed in Section 3.2. We also compare the result provided by this method with two wavelet packet algorithms and ROF method.

4.1 Description of the data and notations

The experiments of Section 4.2 and 4.4 deal with two degradation models. For simplicity we neglect the aliasing in the creation of the images and assume that the Fourier transform of the convolution kernel is supported on $[-\pi,\pi]\times[-\pi,\pi]$.

- The first convolution kernel is a characteristic function a square of size 2 pixels. Therefore, its Fourier transform is

$$\widehat{s_1}(\xi,\eta)=\left(\frac{sin(\xi)}{\xi}\right)\left(\frac{sin(\eta)}{\eta}\right), \tag{17}$$

  for ξ and $\eta\in[-\pi,\pi]$. For the experiments of Section 4.3 we also add a Gaussian noise of standard deviation 5.

- The Fourier transform of the second convolution kernel is given by

$$\widehat{s_2}(\xi,\eta)=\left(\frac{sin(2\xi)}{2\xi}\right)\left(\frac{sin(2\eta)}{2\eta}\right), \tag{18}$$

  for ξ and $\eta\in[-\frac{\pi}{2},\frac{\pi}{2}]$ and 0 otherwise. For the experiments on restoration of this degradation model, we also add a Gaussian noise of standard deviation 2. Note that this degradation model is particularly not adapted to wavelet packet methods since it cancels a wide band of frequencies (see Figure 2). It fact, we know that, because of their ability to reconstruct some lost frequencies (see [13]), variational methods are better suited to this kind of degradation model.

[9] For instance, it is shown in [6] that, in the case of the denoising, the usual wavelet coefficient soft-thresholding is equivalent to the minimization of

$$\|u\|_{B_1^1(L^1)}+\mu\|u-g\|_2^2,$$

where $B_1^1(L^1)$ is a Besov space (see [20]) close to BV.

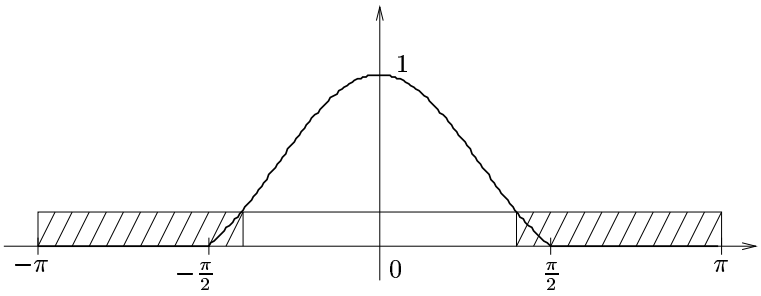

Figure 2: Profile of the Fourier transform of s_2 (see (18)). The hatching represents the frequencies which are, in practice, lost during the degradation.

Figure 3: The reference image.

We also define some simple "pseudo-inverse" operator to the convolutions presented above by truncating the inverse of the Fourier transform at the value 30. More precisely, we take

$$\widehat{r_i}(\xi,\eta) = \begin{cases} \frac{1}{\widehat{s_i}(\xi,\eta)} & \text{, if } |\widehat{s_i}(\xi,\eta)| > \frac{1}{30}, \\ 30 & \text{, if } 0 < \widehat{s_i}(\xi,\eta) \leq \frac{1}{30}, \\ -30 & \text{, if } -\frac{1}{30} \leq \widehat{s_i}(\xi,\eta) < 0, \\ 0 & \text{, if } \widehat{s_i}(\xi,\eta) = 0, \end{cases} \tag{19}$$

for $i = 1, 2$, where s_i is defined by either (17) or (18).

All the experiments using images are done using a part of the image called "Barbara" which is displayed on Figure 3. We chose this particular part because it both contains some textures and a contrasted edge.

Finally, Table 1 summarizes the definition of the wavelet packet bases we will use in the following. We describe these wavelet packet bases in terms of a tree and a wavelet. We use two trees: the mirror-like tree of a given depth, which is exactly the mirror tree described in [15] or its adaptation to s_2 (see (18)); the full tree of a given depth (or pseudo local cosine transform). Concerning wavelets, we use the Shannon wavelet (see [18], pp. 245) and the cubic spline (see[18], pp. 236).

4.2 Approximation of the convolution

We display in this section two kinds of experiments whose aim is to illustrate Proposition 1. The first one shows that we can approximate a convolution efficiently when using (9) and the second shows to evidence the practical importance of the averaging over translations in Proposition 1.

Name	Tree	Wavelet
Basis 1	mirror-like tree of depth 4	cubic spline
Basis 2	full tree of depth 4	cubic spline
Basis 3	full tree of depth 4	Shannon

Table 1: Definition of the wavelet packet bases.

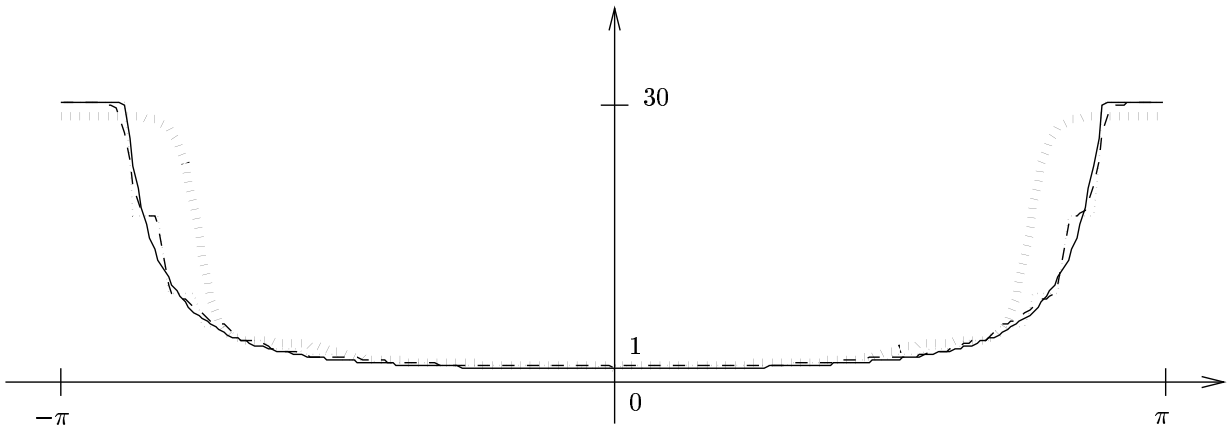

Figure 4: Profile of the Fourier transforms of the convolution kernels derived from different wavelet packets based approximations (see description on page 12). The actual convolution kernel r_1 is represented by the hard line.

In order to highlight the difference between the different kinds of approximation of the kernel, we will approximate the two "high pass" filters r_1 and r_2 (see (19)).

As we said in Proposition 1, the average over several translations of an operator diagonal in a wavelet packet basis is a convolution. We compute and display on Figure 4 the Fourier transform of r_1 (the hard line) and of several of its approximations. In order to create these signals, we averaged the corresponding diagonal operator over translations of a Dirac delta function. The displayed signals are the profile (on the line $\eta = 0$) of the Fourier transforms of the obtained kernels.

The approximations are done in different bases and for each basis we compute the eigenvalues[10] $(\lambda_{t_l})_{1 \leq l \leq L}$ with (11).

Here is a commented description of what is displayed on Figure 4:

- The hard line represents $\widehat{r_1}$.
- The dotted line represents the Fourier transform of the kernel when we approximate the convolution in Basis 3 (see Table 1). In this case, we approximate $\widehat{r_1}$ by a piecewise constant function (the pieces corresponding to dyadic intervals). Note that this corresponds to the announced result (see (10)).
- The dashed line represents the Fourier transform of the kernel when we approximate $\widehat{r_1}$ using Basis 2. This kernel is very close to the previous one but is smoother (which is normal with regard to (10)). Note that both this approximation and the previous one are very close to the initial convolution.
- The dotted and strong line represents the Fourier transform of the kernel when we approximate $\widehat{r_1}$ in Basis 1. This approximation is, of course, less efficient since this time we do not decompose all the frequencial dyadic intervals as much as possible. However, it decorrelates the noise and the information more efficiently than the two previous bases.

We display on Figure 5 exactly the same experiments as in the case of the convolution with r_2. The only difference is that this time we replace the mirror tree by a tree adapted to the special case r_2 (we call

[10]Note that we have not used (10) to compute the $(\lambda_l)_{1 \leq l \leq L}$ and that we could clearly improve our approximations by doing so.

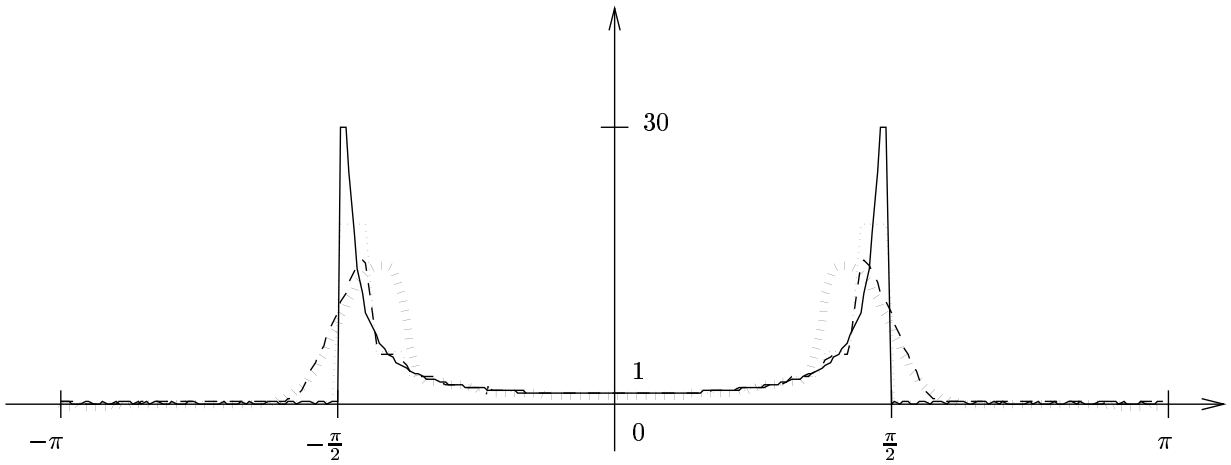

Figure 5: Profile of the Fourier transforms of the convolution kernels derived from different wavelet packets based approximations (see description on page 12). The actual convolution kernel is r_2 and is represented by the hard line.

it mirror-like tree). This tree is fully decomposed for wavelet packet whose frequencial localization is in the vicinity of $\frac{\pi}{2}$ and $-\frac{\pi}{2}$.

The approximation is less accurate in this case than in the previous one because of the large variation of $\widehat{r_2}$. This is especially true for the one made in Basis 1 which poorly approximates the real kernel. Moreover, we partly loose the advantage of the mirror-like tree approach since we must have $\min_l j_l = 3$ in order to decompose more the intermediate frequencies.

It is visually almost impossible to see the difference between a convolved image and its approximation using a wavelet packet basis. However, when it is visible, since it is a modification of a convolution kernel it yields to blurring and/or ringing artifacts.

The next experiment illustrates the importance of the translations. First note that, in our experiments, three translations (one pixel on the right, down and diagonal) have always been sufficient to obtain a reasonably good results. However, if we do not average over any translations the result contains aliasing-like artifacts. With regard to (21) it is clear that these artifacts are due to the aliasing occuring during the wavelet packet decomposition.

Taking notations of Proposition 1, we display on Figure 6 an extracted part of the result of D (on left) and $\tilde{D}$ (on right). The hatching along the edge and the change in the orientation of the texture are two typical aspects of the aliasing.

4.3 Need of spatial localization

We illustrate now the interest of Proposition 2. With that in mind, we have restored using (12) for the same parameters $\sigma = 10$ and $\delta = 0.01$ the image obtained with the degradation model described by (17). We display on Figure 7 an extracted parts of three images (remark that all the images are sharpened for the need of the display).

- Up : Restoration in Basis 2 without the multi-level approach. There is still a lot of noise.
- Down-Left : Restoration in Basis 1 without the multi-level approach. There is less noise than in the previous image but some wavelet packet coefficients are still noisy and we can in practice see the shape of the corresponding basis elements.
- Down-Right : Restoration in Basis 2 with the multi-level approach. This time, the information and the noise have sufficiently been decorrelated and the result does not contain noise.

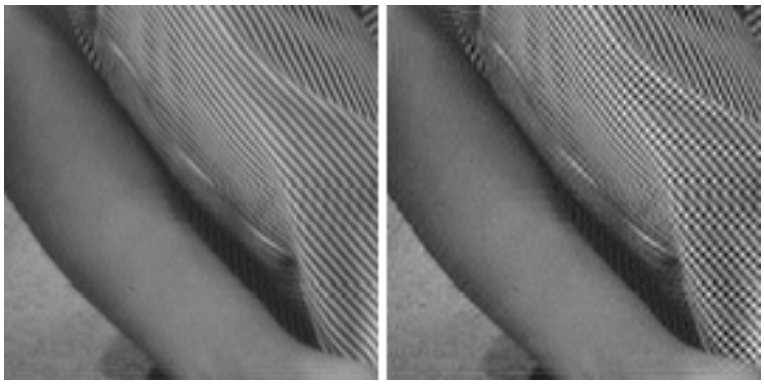

Figure 6: Illustration of the need of the averaging over translations. Extracted and sharpened part of: Left: Approximation of the convolution with r_1 in Basis 1 with the averaging over several translations. Right: Same calculus, without any translation.

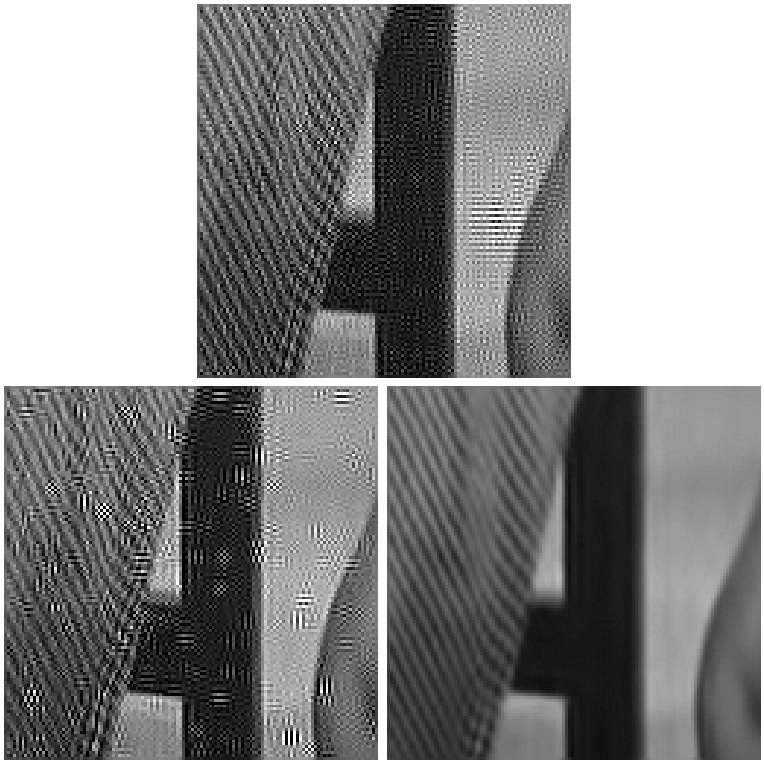

Figure 7: Illustration of the importance of Proposition 2. Deconvolution/Shrinkage of the image degraded with (17) for the same set of parameters in : Up : Basis 2, without multi-level shrinkage. Down-Left : Basis 1, without multi-level shrinkage. Down-Right : Basis 2, with multi-level shrinkage.

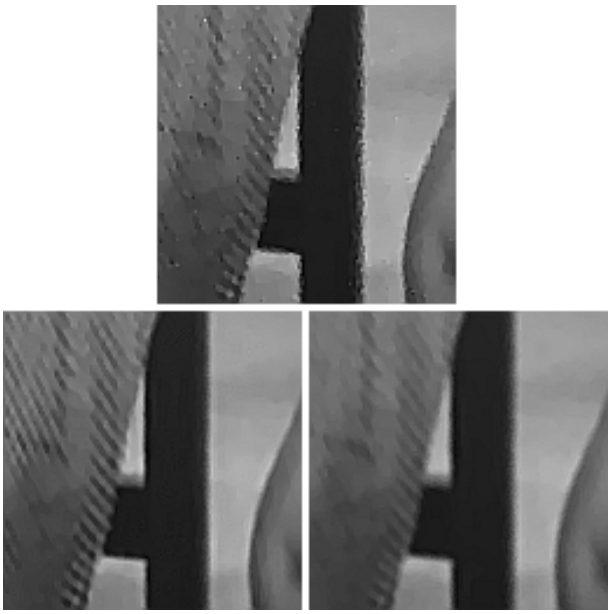

Figure 8: Role of the parameter σ in the modified ROF method (the images are sharpened). Up : $\sigma = 0.001$; Down-Left : $\sigma = 5$; Down-Right : $\sigma = 30$.

4.4 Modifying the Rudin-Osher-Fatemi variational method

We first display two experiments which illustrate the role of the parameter σ and μ of the method. For this purpose, we have computed some restoration of the image degraded with (18) with several sets of parameter.

On Figure 8, we display some sharpened (sharpening of "xv") restorations with $\delta = 1$, $\mu = 0.1$ and : Up : $\sigma = 0.001$; Down-Left : $\sigma = 5$; Down-Right : $\sigma = 30$. We clearly see that for σ too small the staircasing reappears while for σ too large some information is no longer constrained by the data fidelity term and is therefore removed by the total variation.

On Figure 9, we display some sharpened (sharpening of "xv") restorations with $\delta = 1$, $\sigma = 5$ and : Up : $\mu = 0.0001$; Down-Left : $\mu = 0.1$; Down-Right : $\mu = 100$. This time, we see that for μ too large we obtain a result very similar to the result of the wavelet packet shrinkage (which is normal, see comments on page 9). When μ decreases, we remove the remaining noise (we chose $\delta = 1$) and more importantly the ringing and some texture and for μ too small the texture is completely removed. However, for some intermediate μ the texture is well preserved and the ringing has disappeared.

Finally, we present now a comparison of the modified ROF method with two wavelet packet methods and ROF method.

First, note that all the images displayed on Figure 10 have been sharpened (sharpening of "xv"). Let us describe these images in detail.

- Up-Left : A wavelet shrinkage method applied in Basis 1, without multi-level approach, with $\delta = 1$ and $\sigma = 10$.
- Up-Right : A wavelet shrinkage method applied in Basis 2, with the multi-level approach, with $\delta = 1$ and $\sigma = 5$.
- Down-Left : The usual ROF method with $\mu = 0.1$.
- Down-Right The modified ROF using Basis 1, with $\delta = 1$, $\sigma = 5$ and $\mu = 0.1$.

It is clear that the two wavelet packet shrinkage permit to better preserve the texture while they suffer from Gibbs phenomena in the vicinity of the contrasted edge. On the other hand, the two restorations which use the total variation do not present this Gibbs phenomena. However, on the result of ROF

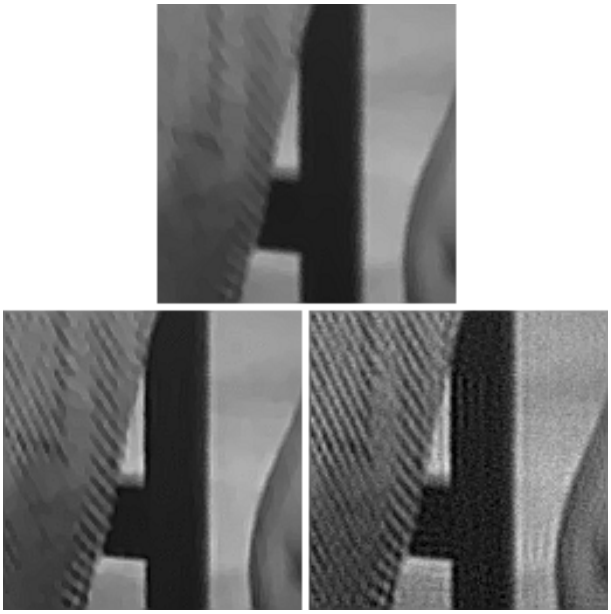

Figure 9: Role of the parameter μ in the modified ROF method (the images are sharpened). Up : $\mu = 0.0001$; Down-Left : $\mu = 0.1$; Down-Right : $\mu = 100$.

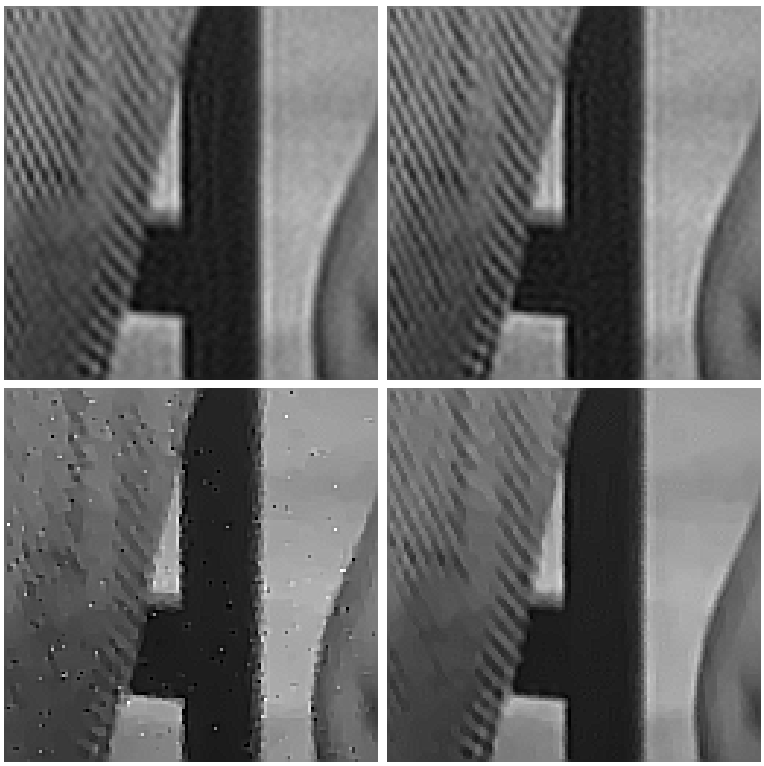

Figure 10: Comparison of restoration methods (the images are sharpened). Up : two wavelet packet shrinkage; Down-Left : ROF method; Down-Right : modification of ROF method.

method we can see some isolated points and flat areas which are due to the staircasing. The texture has also almost completely been removed. On the modification of ROF method, due to the reconditioning of the convolution operator, we no longer see the staircasing and this yields a better preservation of the textures.

Acknowledgments

I would like to thank B. Rougé for all the fruitful discussions we had on this subject and for all his encouragements. I would also like to thank J.M. Morel and S. Durand for the respective parts they have in his work.

Appendix A

Proof of Proposition 1

Proof. For simplicity, we will only prove the result in the case $J = 1$. The proof of the general result is similar to this latter[11].

Similarly to Section 2.1, we will denote, for any $v \in l^2(\mathbb{Z})$, $j \in \mathbb{N}$ and $p \in \{0, ..., 2^j - 1\}$,

$$v_{j,n}^p = \langle \sum_{m \in \mathbb{Z}} v_m \psi_{0,m}^0, \psi_{j,n}^p \rangle \,.$$

We note $t_0 = (0,1)$ and $t_1 = (1,1)$ the two leaves of the tree we are considering.

Let $(u_n)_{n \in \mathbb{Z}} \in l^2(\mathbb{Z})$, for any $n \in \mathbb{Z}$, using (6) and (7) we know that

$$(\widetilde{D}(u))_{1,n}^0 = \lambda_{t_0} \bar{h} * u_0^0(2n) \,,$$

and

$$(\widetilde{D}(u))_{1,n}^1 = \lambda_{t_1} \bar{g} * u_0^0(2n) \,,$$

for $n \in \mathbb{Z}$ (from now one we will abuse of the notation u instead of u_0^0). Therefore, using (8), we have

$$(\widetilde{D}(u))_{0,n}^0 = [h * (\lambda_{t_0} \, (\bar{h} * u)(2.))^\vee]_n + [g * (\lambda_{t_1} \, (\bar{g} * u)(2.))^\vee]_n \,, \tag{20}$$

for $n \in \mathbb{Z}$.

Of course, we have

$$((\bar{h} * u)(2.))^\vee = (\bar{h} * u) \sum_{k \in \mathbb{Z}} \delta_{n-2k}$$

where δ denotes the Dirac delta function (a similar statement holds for $((\bar{g} * u)(2.))^\vee$).

Therefore, expressing (20) in Fourier domain, we have for $\xi \in [-\pi, \pi]$

$$\widehat{\widetilde{D}(u)}(\xi) = \lambda_{t_0} \, \hat{h}(\xi) \, [(\bar{h} * u) \widehat{\sum_{k \in \mathbb{Z}} \delta_{n-2k}}](\xi) + \lambda_{t_1} \, \hat{g}(\xi) \, [(\bar{g} * u) \widehat{\sum_{k \in \mathbb{Z}} \delta_{n-2k}}](\xi) \,.$$

We can simplify this latter, by mean of the Poisson formula ([18], pp. 259), and we obtain

$$\widehat{\widetilde{D}(u)}(\xi) = \lambda_{t_0} \, \hat{h}(\xi) \, \frac{\hat{\bar{h}}(\xi)\hat{u}(\xi) + \hat{\bar{h}}(\xi + \pi)\hat{u}(\xi + \pi)}{2} + \lambda_{t_1} \, \hat{g}(\xi) \, \frac{\hat{\bar{g}}(\xi)\hat{u}(\xi) + \hat{\bar{g}}(\xi + \pi)\hat{u}(\xi + \pi)}{2}$$

[11] Indeed, if $J > 1$, we can simply define $(\widetilde{\lambda_{(p,J)}})_{0 \leq p < 2^J}$ such that, for $p \in \{0, ..., 2^J - 1\}$,

$$\widetilde{D}(\psi_{J,n}^p) = \widetilde{\lambda_{(p,J)}} \, \psi_{J,n}^p \,,$$

in order to paraphrase the proof in the case $J = 1$. However, this yields more complicated notations.

Which can be written

$$\widehat{\widetilde{D}(u)}(\xi) = \frac{1}{2}[\lambda_{t_0}|\hat{h}(\xi)|^2 + \lambda_{t_1}|\hat{g}(\xi)|^2]\,\hat{u}(\xi) + \frac{1}{2}[\lambda_{t_0}\hat{h}(\xi)\hat{\tilde{h}}(\xi+\pi) + \lambda_{t_1}\hat{g}(\xi)\hat{\tilde{g}}(\xi+\pi)]\,\hat{u}(\xi+\pi)\,. \quad (21)$$

Therefore, since $\widehat{(\tau_k u)}(\xi) = e^{ik\xi}\hat{u}(\xi)$, we have, for $\xi \in [-\pi, \pi]$,

$$\widehat{D(u)}(\xi) = [\lambda_{t_0}\frac{|\hat{h}(\xi)|^2}{2} + \lambda_{t_1}\frac{|\hat{g}(\xi)|^2}{2}]\,\hat{u}(\xi)\,,$$

which achieves the proof. □

Appendix B

Proof of Proposition 2

Proof. This is a simple consequence of the fact that $(t_l)_{1\le l\le L} \ge (t'_l)_{1\le l\le L'}$. Indeed, let $n \in \mathbb{Z}$ and $l \in \{1, ..., L\}$, there exists $l' \in \{1, ..., L'\}$ and $(\alpha_m)_{m\in\mathbb{Z}} \in l^2(\mathbb{Z})$, such that

$$\psi_{t_l,n} = \sum_{m\in\mathbb{Z}} \alpha_m\ \psi_{t'_{l'},m}\,.$$

Therefore,

$$\begin{aligned}\widetilde{D}_1 \circ \widetilde{D}_2(\psi_{t_l,n}) &= \frac{\lambda_{(t_l,n)}}{\mu^{p'_{l'}}_{j'_{l'}}} \sum_{m\in\mathbb{Z}} \alpha_m\ \widetilde{D}_1(\psi_{t'_{l'},m}) \\ &= \widetilde{D}(\psi_{t_l,n})\end{aligned}$$

So, $\widetilde{D}_1 \circ \widetilde{D}_2$ and $\widetilde{D}$ are continuous and linear over $l^2(\mathbb{Z})$. They coincide on a basis of $l^2(\mathbb{Z})$. So they are equal. □

References

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

[2] E.J. Candes. *Ridgelets: Theory and Applications.* PhD thesis, Department of Statistics, Stanford University, 1998.

[3] W.K Carey, D. B. Chuang, and S. S. Hemami. Regularity-preserving image interpolation. In *IEEE international conference on image processing*, 1997.

[4] A. Chambolle and P.L. Lions. Restauration de données par minimisation de la variation total et variantes d'ordre supérieur. In *Procedings of GRETSI*, September 1995.

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

[6] 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. Technical report, CEREMADE, 1998. short version in: IEEE Trans. Image Processing, Vol. 7, No. 3, pp. 319-335, 1998.

[7] 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*, 36(2):354–367, 1999.

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

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

[10] 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.

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

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

[13] S. Durand, F. Malgouyres, and B. Rougé. Image de-blurring, 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.

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

[15] 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.

[16] J. Kalifa, S. Mallat, and B. Rougé. Image deconvolution in mirror wavelet bases. In *IEEE, ICIP*, 1998.

[17] N. Kingsbury. Image processing with complex wavelets. *Phil. Trans. Roy. Soc. London A*, 1999.

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

[19] 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*, 22(2):387–405, 2000.

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

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

[22] 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/.

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

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