UCLA
COMPUTATIONAL AND APPLIED MATHEMATICS

Convolution Approximation by Mean of an Operator Diagonal in a Wavelet Packet Basis and Application to Image Deblurring

Francois Malgouyres

November 2000

CAM Report 00-39

Department of Mathematics
University of California, Los Angeles
Los Angeles, CA. 90095-1555

http://www.math.ucla.edu/applied/cam/index.html

Convolution approximation by mean of an operator diagonal in a wavelet packet basis and application to image deblurring*

François Malgouyres†

November 7, 2000

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 then investigate on several applications of these properties to the issue of image deblurring. 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

$$u_0 = s_1 * u + n\,,$$

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

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

where we note with a hat the Fourier transform of a function. We clearly see here that, since $\widehat{s_1}$ can be very small or even be zero, this problem is ill-posed.

The main interest of the wavelet packet framework 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 in several articles (see [10, 12, 13, 21]). The methods, proposed 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 [9]). A part of this paper is somehow a continuation to the articles on "wavelet packet based deblurring" and, we hope, permits to understand and analyze them in a simple way.

There is an abundant literature on image deblurring. The reader is referred to [1] for most of the linear methods and to [8, 11] for overviews on the subject. In few words, the first approach consists in enhancing images without regard to the convolution kernel [14]. The other methods are based on

*This work has been partially financed by the CNES through the ENS Cachan and by UCLA.

†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

regularization approaches of the problem: using statistical properties (Wiener and Kalman filters) or regularity measurements of the images such as the entropy (see [8] and references there), the total variation (see [22]) or the characterization of Besov spaces by wavelets coefficients (see [4, 9]).

For simplicity, all the results, of this paper, are stated in the case of 1D signals. Note that they can be generalized to higher dimensions. The paper is organized as follows:

We will give in Section 2 the statement of the main result of the paper which is that we can approximate a convolution by averaging translations of an operator diagonal in a wavelet packet basis. 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

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

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 the kernel defining this latter. Note that this proposition is a new argument in favor of the cycle spinning introduced in [6]. 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 in [10].

In Section 3, we expose two models, for the image deblurring, based on the properties of Section 2. The first one is equivalent to the usual wavelet-packet shrinkage. The second is new and more audacious. The idea is to use the approximation of the convolution in a wavelet packet basis to approximate the convolution in the Rudin-Osher-Fatemi functional and to use the ability of the wavelet packet decompositions to sparsely represent information to modify this functional. If we present this modification under the point of view of Rudin-Osher-Fatemi 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 the convolution 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 Rudin-Osher-Fatemi model and the modification, we suggested, of this latter.

2 Approximation of the convolution in a wavelet packet basis

2.1 Wavelet packet bases

As we said previously, we approximate a convolution operator by the average over translations of an operator which is diagonal in a wavelet packet basis. Let us first 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. For more details the reader is referred to [7] or to Section 8 of [16].

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 $\psi^p_{j,n}(x) = \psi^p_j(x-2^j n)$ and $\mathbf{W}^p_j$ the vectorial subspace of $L^2(\mathbb{R})$ generated by $\{\psi^p_{j,n},\ n \in \mathbb{Z}\}$, we know that $\{\psi^p_{j,n},\ n \in \mathbb{Z}\}$ is an orthonormal basis of $\mathbf{W}^p_j$. Moreover, we have

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

We also know that for any admissible tree (see Section 8 of [16]) $(p_l, j_l)_{1\le l\le L}$, $\{\psi^{p_l}_{j_l,n}\}_{n\in\mathbb{Z},1\le l\le L}$, 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 $\tilde{u} = \left(\sum_{n\in\mathbb{Z}} u_n \psi^0_{0,n}\right) \in \mathbf{W}^0_0$. Therefore, noting $u^p_{j,n} = \langle \tilde{u}, \psi^p_{j,n}\rangle$ and $(u^p_j)_n = u^p_{j,n}$, we can deduce from (1) and (2) that

$$u^{2p}_{j+1,n} = \sum_{m\in\mathbb{Z}} h_m\ u^p_{j,2n+m} = \bar{h} * u^p_j(2n) \tag{3}$$

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

$$u^{2p+1}_{j+1,n} = \sum_{m\in\mathbb{Z}} g_m\ u^p_{j,2n+m} = \bar{g} * u^p_j(2n) \tag{4}$$

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 the kernel $H^{p_l}_{j_l}$ such that

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

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

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

In other words, noting, for any $u \in l^2(\mathbb{Z}^2)$,

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

we have

$$u^p_{j,n} = (h * (u^{2p}_{j+1})^\vee)_n + (g * (u^{2p+1}_{j+1})^\vee)_n\,. \tag{5}$$

2.2 The approximation of a convolution

The first idea which, in fact, comes from B. Rougé and is the departure point of the deblurring in wavelet packet bases, is that, due to their frequencial localization, it is possible to approximate the convolution in a wavelet packet basis. Therefore, for a suitable basis $\{\psi^{p_l}_{j_l,n}\}_{n\in\mathbb{Z},1\le l\le L}$ and suitable eigenvalues $(\lambda^{p_l}_{j_l})_{1\le l\le L}$ (which do not depend on n), we can define the linear operator $\widetilde{D}$ by

$$\langle \widetilde{D}(u), \psi^{p_l}_{j_l,n}\rangle = \lambda^{p_l}_{j_l} \langle u, \psi^{p_l}_{j_l,n}\rangle$$

for $u \in l^2(\mathbb{Z})$, $l \in \{1, ..., L\}$ and $n \in \mathbb{Z}$.

One of the very important property we loose, when approximating by such a $\widetilde{D}$ the convolution, is the translation invariance. The first simple way to solve this drawback is to use the Shannon wavelet (see [16], 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.

However, as soon as, for $l \in \{1, ..., L\}$, $\lambda^{p_l}_{j_l,n}$ do not depend on n, 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[1].

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

Proposition 1 *Let $(\psi^{p_l}_{j_l,n})_{n\in\mathbb{Z},1\leq l\leq L}$ 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 $(\psi^{p_l}_{j_l,n})_{n\in\mathbb{Z},1\leq l\leq L}$. Assume moreover that the eigenvalues $(\lambda^{p_l}_{j_l,n})_{n\in\mathbb{Z},1\leq l\leq L}$ (respectively associated to the eigenvector $\{\psi^{p_l}_{j_l,n}\}_{n\in\mathbb{Z},1\leq l\leq L}$) do not depend on n. Then, the operator D defined, for any $u\in l^2(\mathbb{Z})$, by*

$$D(u) = 2^{-J}\sum_{k=0}^{2^J-1} \tau_{-k}\circ\widetilde{D}\circ\tau_k(u)\,, \tag{6}$$

where $J=\max_{1\leq l\leq L} j_l$ 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 $\bar{h}$ defining D is given, for $\xi\in[-\pi,\pi]$, by

$$\widehat{\bar{s}}(\xi) = \sum_{l=1}^{L}\lambda^{p_l}_{j_l}\frac{|\widehat{H^{p_l}_{j_l}}(\xi)|^2}{2^{j_l}}\,, \tag{7}$$

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

Proof. For simplicity, we will only demonstrate the result in the case $J=1$. The proof of the general result is similar to this latter[2].

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

$$v^p_{j,n} = \langle\sum_{m\in\mathbb{Z}} v_m\psi^0_{0,m},\psi^p_{j,n}\rangle\,.$$

Let $(u_n)_{n\in\mathbb{Z}}\in l^2(\mathbb{Z})$, for any $n\in\mathbb{Z}$, it is clear, using (3), (4) and (5), that

$$(\widetilde{D}(u))^0_{0,n} = [h*(\lambda^0_1(\bar{h}*u)(2.))^\vee]_n + [g*(\lambda^1_1(\bar{g}*u)(2.))^\vee]_n \tag{8}$$

(note that we abuse of the notation u instead of u^0_0). 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 (8) in Fourier domain, we have, for $\xi\in[-\pi,\pi]$,

$$\widehat{\widetilde{D}(u)}(\xi) = \lambda^0_1\hat{h}(\xi)\,[(\bar{h}*u)\widehat{\sum_{k\in\mathbb{Z}}\delta_{n-2k}}](\xi) + \lambda^1_1\hat{g}(\xi)\,[(\bar{g}*u)\widehat{\sum_{k\in\mathbb{Z}}\delta_{n-2k}}](\xi)\,.$$

We can simplify this latter, by mean of the Poisson formula ([16], pp. 259), and we obtain

$$\begin{aligned}\widehat{\widetilde{D}(u)}(\xi) &= \lambda^0_1\hat{h}(\xi)\frac{\hat{\bar{h}}(\xi)\hat{u}(\xi)+\hat{\bar{h}}(\xi+\pi)\hat{u}(\xi+\pi)}{2} + \lambda^1_1\hat{g}(\xi)\frac{\hat{\bar{g}}(\xi)\hat{u}(\xi)+\hat{\bar{g}}(\xi+\pi)\hat{u}(\xi+\pi)}{2}\,,\\ &= \frac{1}{2}[\lambda^0_1|\hat{h}(\xi)|^2+\lambda^1_1|\hat{g}(\xi)|^2]\,\hat{u}(\xi) + \frac{1}{2}[\lambda^0_1\hat{h}(\xi)\hat{\bar{h}}(\xi+\pi)+\lambda^1_1\hat{g}(\xi)\hat{\bar{g}}(\xi+\pi)]\,\hat{u}(\xi+\pi)\,. \end{aligned}\tag{9}$$

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

$$\widehat{D(u)}(\xi) = [\lambda^0_1\frac{|\hat{h}(\xi)|^2}{2}+\lambda^1_1\frac{|\hat{g}(\xi)|^2}{2}]\,\hat{u}(\xi)\,,$$

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

$$\widetilde{D}(\psi^p_{J,n}) = \widetilde{\lambda^p_J}\,\psi^p_{J,n}\,,$$

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

which achieves the proof. □

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.

Figure 3 and 4 represent the Fourier transforms of two convolution kernels and the corresponding convolution kernel after the approximation by diagonal operators in different wavelet packet bases. We see here that, in the case of the Shannon wavelet (the dashed line), the initial kernel is approximated by a kernel which is constant on dyadic intervals of the Fourier domain (this is also visible in (7) and gives the intuitive meaning of $(\lambda_{j_l}^{p_l})_{1\leq l\leq L}$). Therefore, as long as the Fourier transform of the initial kernel does not vary too much inside these dyadic intervals, the approximation we are doing when approximating 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 $\{\psi_{j_l,n}^{p_l}\}_{n\in\mathbb{Z},l\in\{0,...,L\}}$ according to this criterion.

Remark that, in practice, in order to approximate the convolution operator with $s \in l^1(\mathbb{Z})$, by an operator D which is the average of a operator diagonal in a given wavelet packet basis $\{\psi_{j_l,n}^{p_l}\}_{n\in\mathbb{Z},1\leq l\leq L}$, we can compute the $(\lambda_{j_l}^{p_l})_{1\leq l\leq L}$ in several ways[3]. However, it is easy to check that

$$\lambda_{j_l,n}^{p_l} = \langle s * \psi_{j_l,n}^{p_l}, \psi_{j_l,n}^{p_l} \rangle , \tag{10}$$

permits to minimize $\|S - \widetilde{D}\|_2$ (where $S(u) = s * u$) and can therefore be considered as a good candidate. Remark that these $\lambda_{j_l,n}^{p_l}$ do not depend on n so that we can forget the index n and denote them by $\lambda_{j_l}^{p_l}$.

However, Proposition 1 can also be used to determine the approximated eigenvalues $(\lambda_{j_l}^{p_l})_{1\leq l\leq L}$. We can for instance choose these eigenvalues in such a way that they minimize the error between $\tilde{s}$ and s. Note that we can also use this formula 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, we could determine $(\lambda_{j_l}^{p_l})_{1\leq l\leq L}$ in such a way that $\tilde{s} * s_1$ is positive (in order to avoid ringing effects in the vicinity of edges).

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 separates information from the 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 [12, 13]). The second one was introduced in [10] 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 $(p_l, j_l)_{1\leq l\leq L}$ and $(p'_l, j'_l)_{1\leq l\leq L'}$ be two admissible trees, we say that*

$$(p_l, j_l)_{1\leq l\leq L} \geq (p'_l, j'_l)_{1\leq l\leq L'}$$

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

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

This relation simply means that the elements of $\{\psi_{j_l,n}^{p_l}\}_{n\in\mathbb{Z},1\leq l\leq L}$ correspond to a higher level of decomposition than the ones of $\{\psi_{j'_l,n}^{p'_l}\}_{n\in\mathbb{Z},1\leq l\leq L'}$. Note that if the admissible trees are indexed with regard to there position in the binary tree (for instance from left to right) then the $I_{l'}$ are of the form $\{t_{l'-1}, ..., t_{l'} - 1\}$ with $1 = t_0 < ... < t_{l'} < t_{l'+1} < ... < t_{L'} = L + 1$.

Using this definition, we can state,

[3] One can refer to [10] for examples of such computation.

Proposition 2 *Let $\{\psi^{p_l}_{j_l,n}\}_{n\in\mathbb{Z},1\leq l\leq L}$ be a wavelet packet basis and $\widetilde{D}$ be an operator linear, continuous and diagonal in the basis $\{\psi^{p_l}_{j_l,n}\}_{n\in\mathbb{Z},1\leq l\leq L}$, which goes from $l^2(\mathbb{Z})$ into itself. If we note $\lambda^{p_l}_{j_l,n}$ the eigenvalue of $\widetilde{D}$ associated with the eigenvector $\psi^{p_l}_{j_l,n}$, then for any admissible tree $(p'_l,j'_l)_{1\leq l\leq L'}$, such that $(p_l,j_l)_{1\leq l\leq L}\geq (p'_l,j'_l)_{1\leq l\leq L'}$, and any $(\mu^{p'_{l'}}_{j'_{l'}})_{1\leq l\leq L'}\in(\mathbb{R}\setminus\{0\})^{L'}$, 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 $l'\in\{1,...,L'\}$,

$$\widetilde{D}_1(\psi^{p'_{l'}}_{j'_{l'},n})=\mu^{p'_{l'}}_{j'_{l'}}\;\psi^{p'_{l'}}_{j'_{l'},n}\,,$$

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),

$$\widetilde{D}_2(\psi^{p_l}_{j_l,n})=\frac{\lambda^{p_l}_{j_l,n}}{\mu^{p'_{l'}}_{j'_{l'}}}\;\psi^{p_l}_{j_l,n}\,.$$

Proof. This is a simple consequence of the fact that $(p_l,j_l)_{1\leq l\leq L}\geq(p'_l,j'_l)_{1\leq l\leq L'}$. 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

$$\psi^{p_l}_{j_l,n}=\sum_{m\in\mathbb{Z}}\alpha_m\;\psi^{p'_{l'}}_{j'_{l'},m}\,.$$

Therefore,

$$\begin{aligned}\widetilde{D}_1\circ\widetilde{D}_2(\psi^{p_l}_{j_l,n}) &= \frac{\lambda^{p_l}_{j_l,n}}{\mu^{p'_{l'}}_{j'_{l'}}}\sum_{m\in\mathbb{Z}}\alpha_m\;\widetilde{D}_1(\psi^{p'_{l'}}_{j'_{l'},m})\\ &= \widetilde{D}(\psi^{p_l}_{j_l,n})\end{aligned}$$

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

This proposition proves that, the operator $\widetilde{D}$, diagonal in a wavelet packet basis, of which we average the translation, 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.

Remark also that the $\lambda^{p_l}_{j_l,n}$ generally do not depend on n (for instance in the case of (10)) and that Proposition 2 could be consequently simplified.

For instance, 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 that all the $\lambda^p_{j_0}$ are positive and do not depend on n. We can let $\mu^p_{j_0}=\lambda^p_{j_0}$ for $p\in\{0,...,2^{j_0}-1\}$ and then recursively define $\mu^p_{j-1}=min(|1-\mu^{2p+1}_j|,|1-\mu^{2p}_j|)$ 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 for levels $i\leq j$). Therefore, we can decompose $\widetilde{D}=\widetilde{D}_1\circ\cdots\circ\widetilde{D}_{j_0}$ where the $\widetilde{D}_j$ are some continuous and linear operator, diagonal in the basis $\{\psi^p_{j,n}\}_{n\in\mathbb{Z},0\leq p<2^j}$, with the eigenvalues $\frac{\mu^{2p+1}_j}{\mu^p_{j-1}}$ and $\frac{\mu^{2p}_j}{\mu^p_{j-1}}$ 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 FCNR

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 sparse representation of the image. For instance, in the case of the deconvolution, if we define (by any appropriate mean) a "pseudo-inverse[4]" r of the convolution kernel s_1.

Once we have chosen an appropriate wavelet packet basis $\{\psi_{j_l,n}^{p_l}\}_{n\in\mathbb{Z},1\leq l\leq L}$ and chosen a sequence $\boldsymbol{\lambda} = (\lambda_{j_l}^{p_l})_{1\leq l\leq L}$ of real numbers which permit to approximate the convolution with r efficiently (or so that the convolution it defines is itself an acceptable "pseudo-inverse"). We can, according to Proposition 1, define an approximated convolution, of $u \in l^2(\mathbb{Z})$, by averaging, over several translations of u, the operator $\widetilde{D}$, defined by

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

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

Remark that, according to Proposition 2, we can now decompose

$$\widetilde{D} = \widetilde{D}_1 \circ \widetilde{D}_2 \circ \widetilde{D}_3 \circ \cdot\,.$$

Remark now that, for any of these $\widetilde{D}_i$, we can define an "adaptative convolution" in a way similar to what follows. This permits to justify the multi-level wavelet packet shrinkage which has been named FCNR and described in [10, 21].

Let us now define what we mean by "adaptative convolution". For $\sigma > 0$ and $u \in l^2(\mathbb{Z})$, we can define an adaptative convolution, by averaging, over several translations of u, the operator

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

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

Heuristically, we convolve the information contained in u and leave unchanged what we consider being noise. 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, $\delta \in [0, 1]$, in order to lower the remaining noise. Therefore, we will prefer, to the above operator, an operator of the kind

$$\langle A_{\boldsymbol{\lambda},\sigma,\delta}(u), \psi_{j_l,n}^{p_l}\rangle = \begin{cases} 0 & \text{, if } \lambda_{j_l}^{p_l} = 0, \\ \lambda_{j_l}^{p_l}(\langle u, \psi_{j_l,n}^{p_l}\rangle - \sigma) + \delta\sigma & \text{, if } \langle u, \psi_{j_l,n}^{p_l}\rangle \geq \sigma \text{ and } \lambda_{j_l}^{p_l} \neq 0, \\ \delta\langle u, \psi_{j_l,n}^{p_l}\rangle & \text{, if } \sigma > \langle u, \psi_{j_l,n}^{p_l}\rangle \geq -\sigma \text{ and } \lambda_{j_l}^{p_l} \neq 0, \\ \lambda_{j_l}^{p_l}(\langle u, \psi_{j_l,n}^{p_l}\rangle + \sigma) - \delta\sigma & \text{, if } -\sigma > \langle u, \psi_{j_l,n}^{p_l}\rangle \text{ and } \lambda_{j_l}^{p_l} \neq 0, \end{cases} \tag{11}$$

for $u \in L^2(\mathbb{R})$, $\sigma > 0$, $\delta \in [0, 1]$, $n \in \mathbb{Z}$ and $l \in \{1, ..., L\}$.

Remark that the average over the translations of the usual wavelet thresholding methods falls under the scope of (11). Therefore, the framework of the "adaptative convolution" can probably give some tools to understand what these algorithms do.

This "adaptative convolution" is probably the most natural and immediate application of the results stated in the preceding section. In fact, this framework is only a new point of view on existing methods. However, we do believe, and will try to show it in the experiments, that this permits a better understanding of what these methods really do and to explain there drawbacks and properties.

[4]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.

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 in [22]. In order to have 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 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

$$TV(u) + \lambda \|s_1 * u - g\|_2 \,, \tag{12}$$

among $u \in \mathbb{R}^N$, where λ can be interpreted as a Lagrange multiplier (see [3]) 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 in the 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 [19, 20]. If we look in detail at the arguments given in [20], we see that one of the key properties which causes this staircasing is the fact that we can not have a "reasonable local[5]" solution to the equation

$$\overline{s_1} * (s_1 * u - g) = 0 \,, \tag{13}$$

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 leads us to modify the functional in order to have a data fidelity term whose derivative (the left term in (13)) can be null. With that in mind, we change the convolution operator in $\|s_1 * u - g\|_2$ by an "adaptative convolution" similar to the one defined in (11). More precisely, given a wavelet packet basis[6] $(\psi^{p_l}_{j_l,n})_{1\leq l\leq L, 0\leq n<2^{-j_l}N}$, we compute some eigenvalues $\boldsymbol{\lambda} = (\lambda^{p_l}_{j_l})_{1\leq l\leq L}$ (for instance with (10)) in order to approximate the convolution with s_1 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 $\widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta}$

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

for $\sigma > 0$ and $\delta > 0$. 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 $J = \max_l j_l$.

Note that, in order to define a convex data fidelity term $\|S_{g,\boldsymbol{\lambda},\sigma,\delta}(u) - g\|_2$, we have not taken the adaptative convolution defined by (11) ($\|A_{\boldsymbol{\lambda},\sigma,\delta}(u) - g\|_2$ is not convex as soon as one of the $\lambda^{p_l}_{j_l}$ is lower

[5] 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 $\overline{s_1} * (s_1 * u - g) = 0$ does not have any solution on any open subset of Ω. The heuristic translation of this hypothesis in our discrete framework could be that we do not have any "reasonable local" solution.

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

than 1, which is, in general, the case). Therefore, we propose to minimize, among $u \in \mathbb{R}^N$, a functional of the kind

$$TV(u) + \lambda \| S_{g,\boldsymbol{\lambda},\sigma,\delta}(u) - g \|_2 \,. \tag{14}$$

Note that $S_{g,\boldsymbol{\lambda},\sigma,\delta}$ is affine and so that the functional (14) 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 [3, 10].

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

$$S'_{g,\boldsymbol{\lambda},\sigma,\delta}(S_{g,\boldsymbol{\lambda},\sigma,\delta}(u) - g) = 0 \,, \tag{15}$$

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

Moreover, this solution is close to the solution of the wavelet shrinkage method described in the preceding section. In fact, if, in (15), we take $\widetilde{S}_{g,\boldsymbol{\lambda},\sigma,\delta}$, instead of $S_{g,\lambda,\sigma,\delta}$, and if $\delta \neq 0$ and $\lambda_{j_l}^{p_l} \neq 0$, for any $l \in \{1, ..., L\}$,

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

is a solution of (15).

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 [10]). That is one of the oddness which is solved by our new approach.

Moreover, with regard to the causes of the staircasing given in [20], the existence of u_∞ cancels one of the reason of the staircasing. We will see in the experiments that the image restored by means of (14) are indeed free of staircasing.

Remark: We have chosen here to present (14) under a variational point of view. We are conscious of the fact that (14) can appear redundant to readers who are usually interested in wavelet shrinkage. 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 close to (12)[7]. However, the drawback of these methods in the case of deblurring is that they cannot recover lost frequencies (see [10]) (we can however mention the attempt to oversample images by means of wavelet transforms made in [2]). 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 (14) 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 (14) 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 (14) decrease. We could probably have a better algorithm by using methods such as the ones introduced in [5, 17]

Compared to the usual Rudin-Osher-Fatemi algorithm, the computational cost increases due to the translations made 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.

[7] For instance, it is shown in [4] that, in the case of the denoising, the usual wavelet coefficient soft-thresholding is equivalent at minimizing

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

where $B_1^1(L^1)$ is a Besov space (see [18]) close to BV.

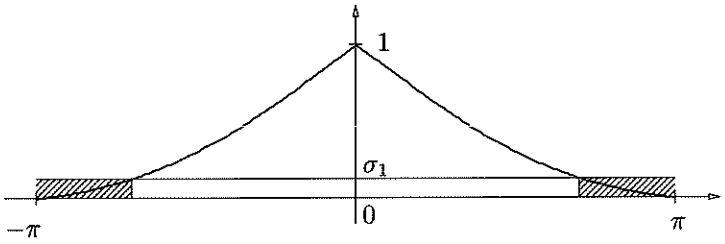

Figure 1: Profile of the Fourier transform of s_1 (see (16)). The hatching represents the frequencies which are, in practice, lost during the degradation.

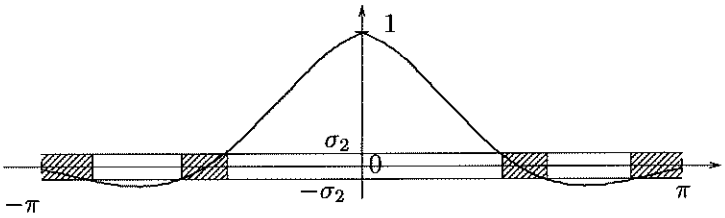

Figure 2: Profile of the Fourier transform of s_2 (see (17)). The hatching represents the frequencies which are, in practice, lost during the degradation.

4 Numerical results

This section is split into four parts. They are organized as follows. The first part describes the data and notations which permit to understand the experiments of other sections. In the second part, we display experiments which show that we can, in reasonable cases, approximate a convolution operator efficiently in a wavelet packet basis. In the third part, we illustrate the practical interest of Proposition 2. We display in the last section some experiments on the possibility to modify Rudin-Osher-Fatemi method according to the preceding section and compare its results with the method of Rudin-Osher-Fatemi and with two wavelet packet shrinkage.

4.1 Description of the data and notations

The experiments are based on two realistic degradation models (the same as the ones presented in [10]) which are derived from satellite imaging. They correspond to two different satellites.

In both cases, the Fourier transform of the impulse response is supported on $[-\pi,\pi]\times[-\pi,\pi]$. 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

$$\widehat{s_1}(\xi,\eta) = e^{-2\gamma_\xi|\xi|-2\gamma_\eta|\eta|}\left(\frac{sin(2\xi)}{2\xi}\right)\left(\frac{sin(2\eta)}{2\eta}\right)\left(\frac{sin(\eta)}{\eta}\right), \text{ for } \xi,\eta\in[-\pi,\pi], \tag{16}$$

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

- The convolution kernel of the second model is given by

$$\widehat{s_2}(\xi,\eta) = e^{-2\gamma_\xi|\xi|-2\gamma_\eta|\eta|}\left(\frac{sin(4\xi)}{4\xi}\right)\left(\frac{sin(4\eta)}{4\eta}\right), \text{ for } \xi,\eta\in[\pi,\pi], \tag{17}$$

  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 [10] 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

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.

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 (at least in the case of the total variation).

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

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

for $i = 1, 2$, where the s_i are defined by (16) or (17).

We will also use 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 [10], we display some parts of this reference on Figures 12.b, 13.b, 14.b).

In the following sections, we will show experiments using wavelet packet bases. Since these bases have an important impact on the results of the experiments, we have chosen to summarize their definitions in Table 1.

We describe these wavelet packet bases in terms of a tree and a wavelet. We will use two trees: the mirror like tree of a given depth, which is exactly the mirror tree described in [12] or its adaptation to s_2 (see (17)); the full tree of a given depth (or pseudo local cosine transform). Concerning wavelets, we will use the Shannon wavelet (see [16], pp. 245) and the cubic spline (see[16], pp. 236).

4.2 Approximation of the convolution

We will 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 (6) and the second one shows to evidence the practical importance of the average over translations in Proposition 1.

In order to highlight the difference between the different kind of approximations of the kernel, we will approximate the two "high pass" filters defined in the preceding section: r_1 or r_2.

As we said in Proposition 1, the average over several translations of an operator diagonal in a wavelet packet basis is a convolution. Therefore, we have computed the Fourier transform of its kernel. We display on Figure 3 the Fourier transform of r_1 (the hard line) and of several of its approximations. In order to create these signals, we have averaged the corresponding diagonal operator over translations of a Dirac delta function. The displayed signals are the profile of the Fourier transforms of the obtained kernels, on the line $\eta = 0$.

Here is the detailed and commented description of what is displayed on Figure 3:

- The hard line represents $\widehat{r_1}$.
- The dashed line represents the Fourier transform of the kernel when we approximate the convolution using the $(\lambda_{j_l}^{p_l})_{1 \leq l \leq L}$ defined by (10) in Basis 3 (defined in Table 1). We clearly see that in this case, we simply approximate $\hat{s}$ by a piecewise constant function (the pieces corresponding to dyadic intervals). Remark that this corresponds to the announced result (see (7)). Note also that in this

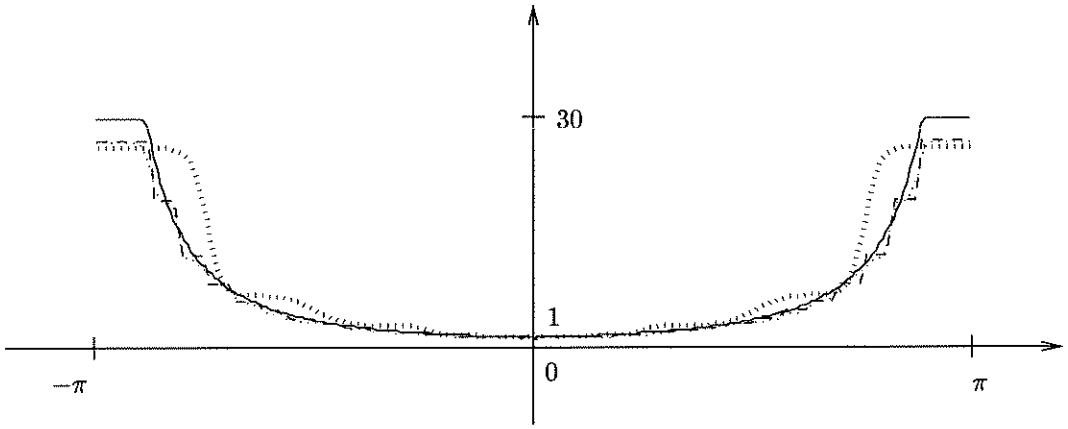

Figure 3: Profile of the Fourier transforms of the convolution kernels derived from different wavelet packets based approximations (see description on page 11). The initial convolution kernel is r_1 and is represented by the hard line.

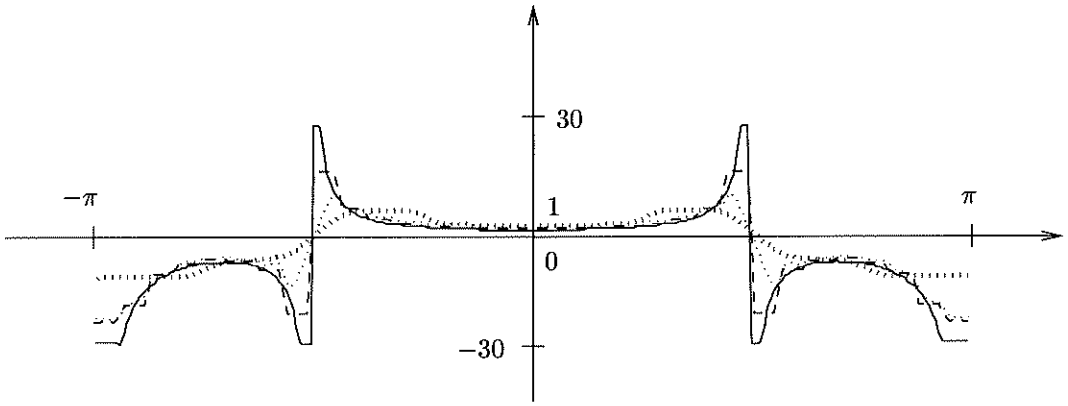

Figure 4: Profile of the Fourier transforms of the convolution kernels derived from different wavelet packets based approximations (see description on page 11). The initial convolution kernel is r_2 and is represented by the hard line.

case, since we use the Shannon wavelet, we do not need to do all the translation (see (9), in the proof of Proposition 1).

- The dotted line represents the Fourier transform of the kernel when we approximate the convolution using the same $(\lambda_{j_l}^{p_l})_{1\leq l\leq L}$ as previously but when using Basis 2. This kernel is very close to the previous one but is smoother (which is normal with regard to (7) and since the cubic spline is less localized in frequency domain than the Shannon wavelet). 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 the convolution using the $(\lambda_{j_l}^{p_l})_{1\leq l\leq L}$ defined in (10)[8] in the case of 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, this approximation is made in a basis whose elements have a better spatial localization than the previous basis, which is interesting for the purpose of denoising.

Note that we have not used (7) to compute the $(\lambda_l)_{1\leq l\leq L}$ and that we could clearly improve our approximations by doing so.

[8]Note that, for the calculus of $(\lambda_{j_l}^{p_l})_{1\leq l\leq L}$, we have approximated the spline by the Shannon wavelet.

Bases	convolution with r_1	convolution with r_2
Basis 1	7.9	22.8
Basis 2	3.1	6.3
Basis 3	3.9	5.3

Table 2: Mean square error between the exact and the approximated convolution.

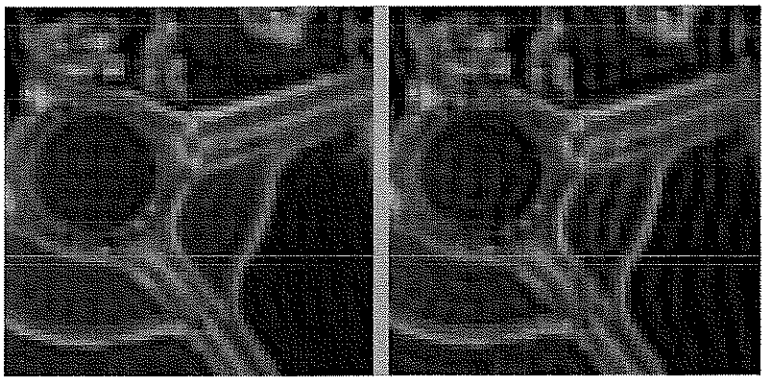

Figure 5: Illustration of the error made when we approximate the convolution kernel.
Left: the convolution with r_2.
Right: its approximation in Basis 1.

We display on Figure 4 exactly the same experiments 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 it mirror-like tree). Therefore, with this tree, we decompose more the wavelet packet whose frequencial localization is in the vicinity of $\frac{\pi}{2}$ and $-\frac{\pi}{2}$.

The approximation is worth 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, this time, 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 in a wavelet packet basis. Therefore, in order to show to evidence that the error we are doing when approximating the convolution kernel is not too important, we have computed the convolution of the reference image with s_1 (respectively s_2) and convolved it with its pseudo-inverse r_1 (respectively r_2) or by its approximation using different wavelet packet bases (described in Table 1). We summarize in Table 2 the error in terms of the means square error between the exact and the approximated convolution. Note that the main comments concerning these statistics are that they are reasonably small and that Basis 1 yields not as good statistics as the other ones (but on the other hand it has a better spatial localization).

Remark that, we can however see the difference when we approximate the convolution with r_2 in Basis 1 (the profile of the approximated kernel in this case is represented by the dotted strong line on Figure 4). In order to show this difference, we take as an initial image the reference image convolved with s_2 and we display, on the left part of Figure 5, its convolution with r_2; and, on the right hand side of Figure 5, the convolution approximated in Basis 1 (as described for Figure 4). Here, we see that the difference is a kind of ringing artifact which is due to the fact that the composition of the two convolutions (with s_2 and the approximation of r_2) is a convolution whose kernel can be negative.

The next experiments illustrate the need of doing the translations. First, note that in practice three translations (one pixel on the right, down and diagonal) are 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

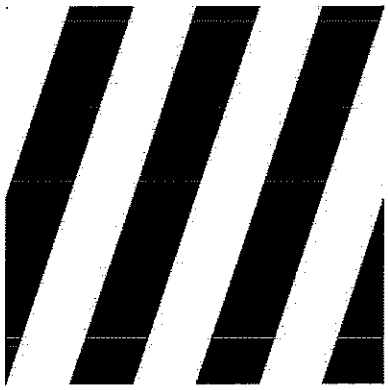

Figure 6: Cylindrical image.

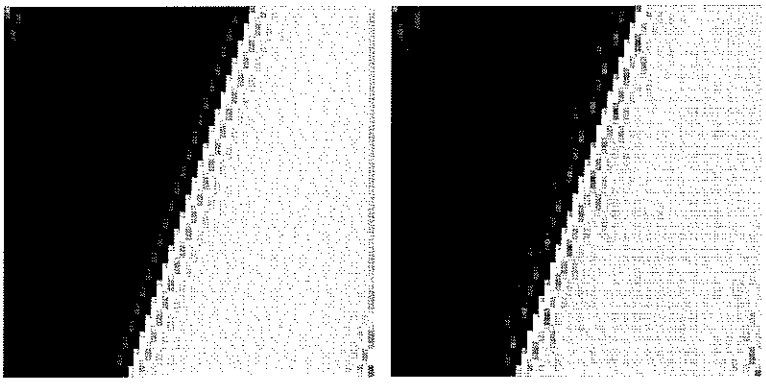

Figure 7: Illustration of the need of the averaging over translations. Extracted and sharpened part of:
Left: Approximation of the convolution with r_1 in Basis 2, average over several translations.
Right: Same calculus, without any translation.

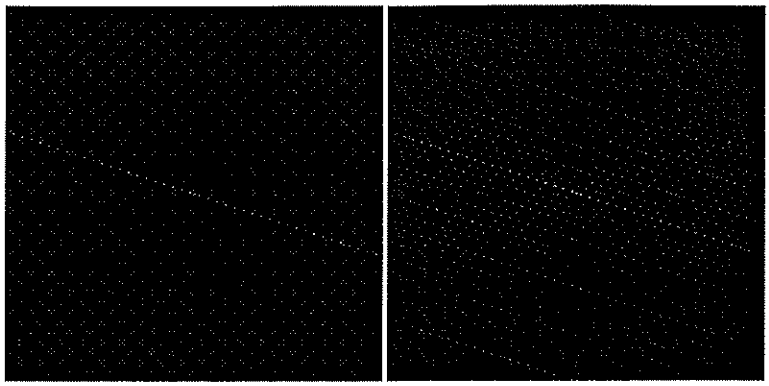

Figure 8: Modulus of the Fourier transform (raise at the power 0.01) of the images displayed on Figure 7.

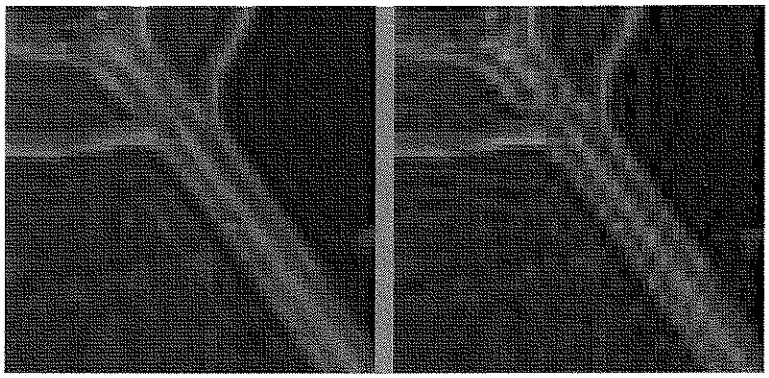

Figure 9: Illustration of the need of averaging over translations.
Left: Approximation of the convolution with r_2 in Basis 1, average over several translations.
Right: Same calculus, without any translation.

(9) it is clear that these artifacts are due to the aliasing occuring during the wavelet packet decomposition.

In order to illustrate this, we use a cylindrical function as introduced in [15] (the Fourier transform of a cylindrical function is supported by a line (just like on the left image of Figure 8)). Therefore, we convolved the cylindrical function displayed on Figure 6 with s_1. We then compute the image of this function by an approximation of r_1 in Basis 2 (note that the result would have been worth with Basis 1). We display on Figure 7 an extracted part of a sharpened version (the sharpening of "xv") of the result (on left) and the result when doing the same approximation without any translations (on right). The hatching which is not in the direction of the edge is clearly due to the aliasing inside the wavelet packet. We can see on the modulus of the Fourier transform of these images (see Figure 8) the effect of the aliasing inside the wavelet packet when we do not average different translations.

We also show on Figure 9 the same experiment when the initial image is the reference image, the first convolution kernel is s_2 and we approximate the convolution with r_2 in Basis 1. This time the aliasing is translated into the hatching on the road.

4.3 Need of spatial localization

As we said in the introduction of Proposition 2 it is often important to have a wavelet packet decomposition which decorrelates efficiently the noise from the information. We then presented two ways to achieve this goal. In order to illustrate these, we focus on the deconvolution, using the wavelet shrinkage method defined by (11), of the blurred image whose degradation model is described by (16).

First, we display on Figure 10 two restorations of this blurred image, when using a multi-level approach, with $\delta = 1$, on left in the case of Basis 3 (with $\sigma = 10$) and on right with Basis 2 (with $\sigma = 3.5$). We see

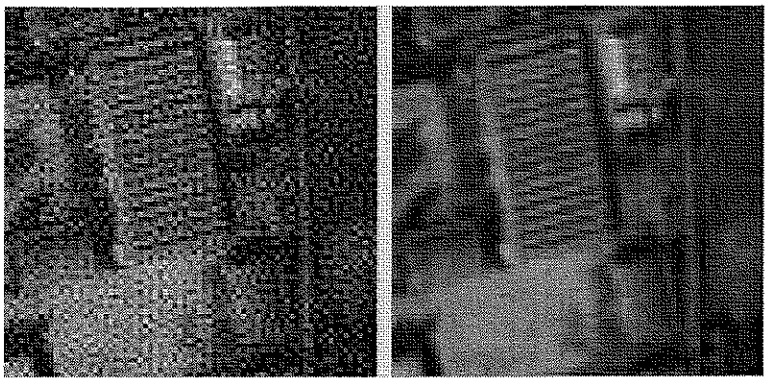

Figure 10: Illustration of the need of spatial localization of the elements of the wavelet packet basis. Deconvolution/Shrinkage of a noisy image in:
Left: Basis 3.
Right: Basis 2.

that the left image contains noise while the right one does almost not. This is due to the fact (which by the way is well known) that, just like the Fourier transform, the Shannon wavelet decomposition does not decorrelate the noise from the information because of its weak spatial localization. Therefore, the Shannon wavelet can not be used (at least for the purpose of denoising) to avoid the translations which are needed to properly approximate a convolution when using other wavelets.

In addition to this rough aspect of the need of spatial localization, we illustrate on Figure 11 the interest of Proposition 2. With that in mind, we have restored using (11) the image obtained with the degradation model described by (16). We display on Figure 11 two extracted parts of three images (remark that all the images are sharpened for the need of the display).

- Up: The two images are extracted from a restoration of the blurred image when shrinking in Basis 2, without the multi-level approach. Here, we take $\delta = 1$ and $\sigma = 4$. We see on both images some vertical structures[9].
- Middle: The image is computed with the same basis except that we take $\sigma = 3.5$ and use the multi-level shrinkage introduced in [10] and justified by Proposition 2. The vertical structures present in the preceding image have disappeared. Remark also that we obtain less noise with a smaller value of σ, since we remove some noise in different bases.
- Down: The two images are computed using the wavelet packet shrinkage in Basis 1, with $\delta = 1$ and $\sigma = 4$, and without any multi-level shrinkage. The structures present in the first images have also almost disappeared. You can also remark that the line in the middle of the road is better restored than on the previous image. This is probably due to the fact that, at least in our experiments , the approximated convolution kernel has, on almost all the frequency domain, a Fourier Transform larger in the case of Basis 1 than in other cases (see Figure 3).

4.4 Modifying the Rudin-Osher-Fatemi variational method

We present in this section some results on the modified Rudin-Osher-Fatemi variational method. In order to illustrate the ability of this method to both recover textures and avoid ringing, we have chosen to restore the image degraded by the second degradation model (see (17)).

We have restored this image by several methods. Let us describe the methods under consideration

- A wavelet shrinkage method applied in Basis 1, without multi-level approach, with $\delta = 1$ and $\sigma = 1$.

[9]Remark that the simulated noise is not really Gaussian and has an "inter-column" component, which yields these structures.

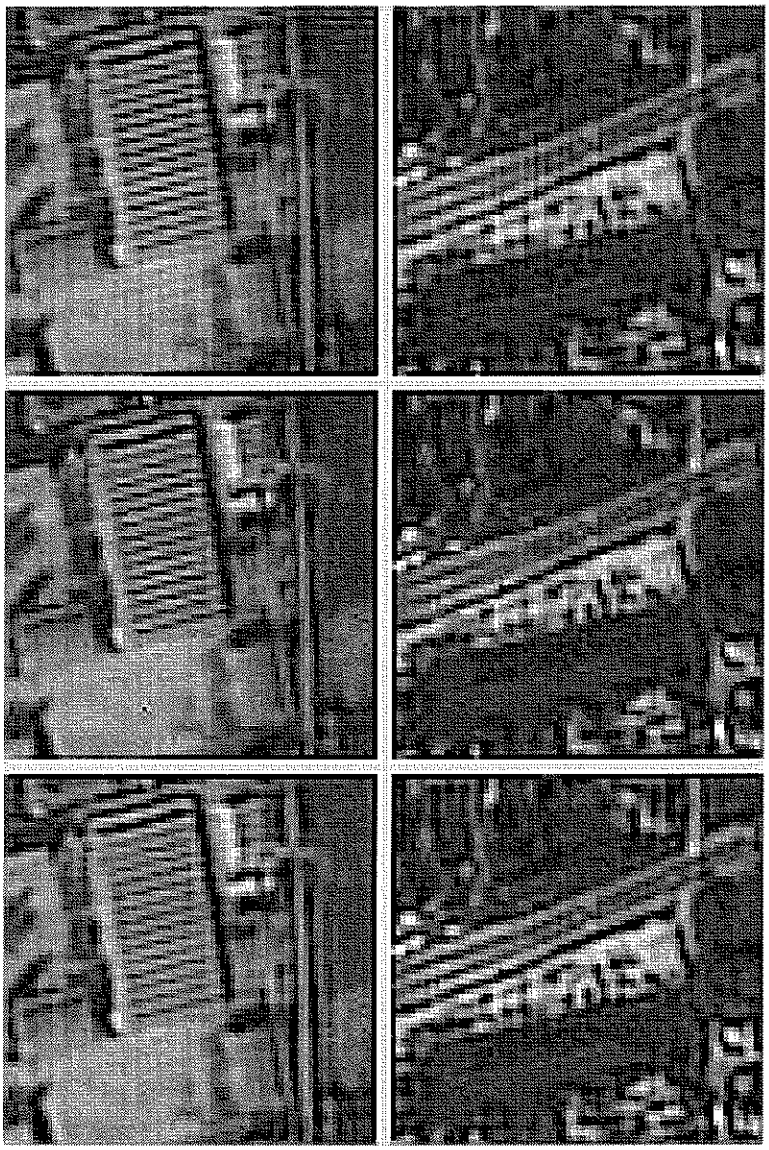

Figure 11: Illustration of the use of Proposition 2. Deconvolution/Shrinkage of a noisy image in:
Up: Basis 2, without multi-level shrinkage.
Middle: Basis 2, with multi-level shrinkage.
Down: Basis 1, without multi-level shrinkage.

- A wavelet shrinkage method applied in Basis 2, with the multi-level approach, with $\delta = 1$ and $\sigma = 1$.
- The usual Rudin-Osher-Fatemi method with $\lambda = 8$.
- The modified Rudin-Osher-Fatemi using Basis 1, with $\delta = 1$, $\sigma = 1$ and $\lambda = 8$.

We display on Figures 12, 13 and 14 extracted parts of these restorations as well as the reference and the blurred images. On Figure 12, we see that the modified Rudin-Osher-Fatemi and the wavelet packet shrinkage methods preserve textures better than the Rudin-Osher-Fatemi one. We also see that the modified Rudin-Osher-Fatemi method does not suffer from staircasing artifacts (like the usual Rudin-Osher-Fatemi method). On Figure 13 (which has been sharpened for the display), we see that, despite the modification, the methods based on the total variation still permits to remove the ringing effects which is present in the other wavelet packet methods. Once again, we can see the staircasing on Figure 13.e but not on 13.f. At last, we see on Figure 14 that, in wavelet packet methods, the approximation of the convolution kernel yields to a result which is still more blurry than the classical Rudin-Osher-Fatemi method. This is due to the bad approximation of the convolution (see Figure 4) and can maybe be improved by optimizing the basis and the $(\lambda_{j_l}^{p_l})_{1 \leq l \leq L}$.

Acknowledgments

I would like to thank B. Rougé for all the fruitful discussions we had on this subject and for all his encouragements.

References

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

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

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

[4] 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.

[5] 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.

[6] R.R. Coifman and D.L. Donoho. Translation-invariant de-noising. Technical Report 475, Standford University, May 1995.

[7] 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.

[8] 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.

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

[10] S. Durand, F. Malgouyres, and B. Rougé. Image de-blurring, spectrum interpolation and application to satellite imaging. *COCV*, 5, 2000. A preliminary version is available at http://www.math.ucla.edu/~malgouy.

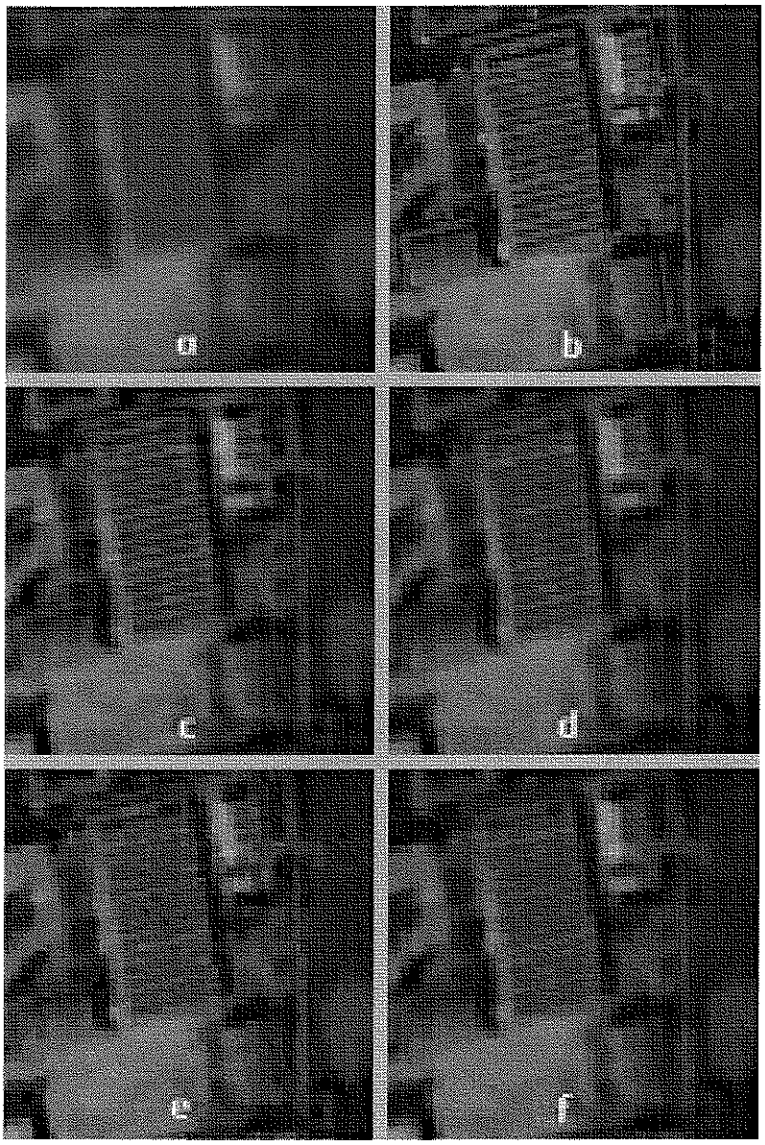

Figure 12: Restoration of the degradation model described in (17).
a: Blurred image.
b: Reference image.
c: Shrinkage in Basis 1, without multi-level approach.
d: Multi-level shrinkage in Basis 2.
e: Rudin-Osher-Fatemi method.
f: Modified Rudin-Osher-Fatemi method.

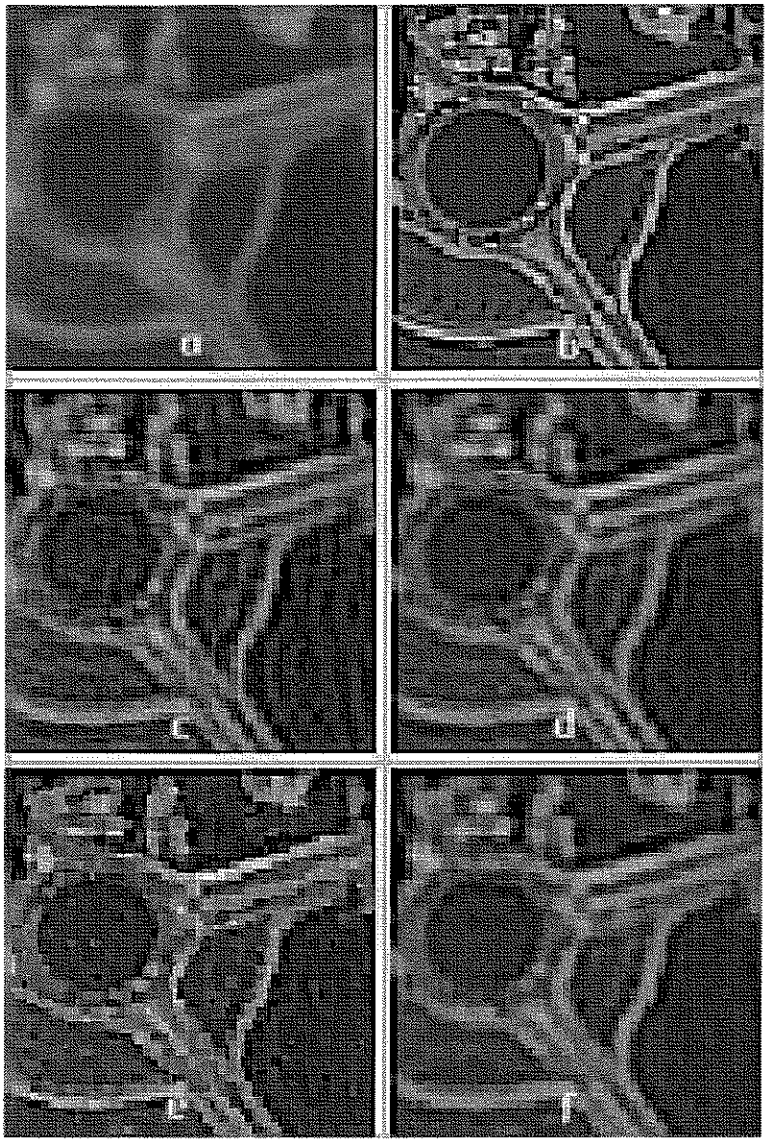

Figure 13: Restoration of the degradation model described in (17) (the images have been sharpened).
a: Blurred image.
b: Reference image.
c: Shrinkage in Basis 1, without multi-level approach.
d: Multi-level shrinkage in Basis 2.
e: Rudin-Osher-Fatemi method.
f: Modified Rudin-Osher-Fatemi method.

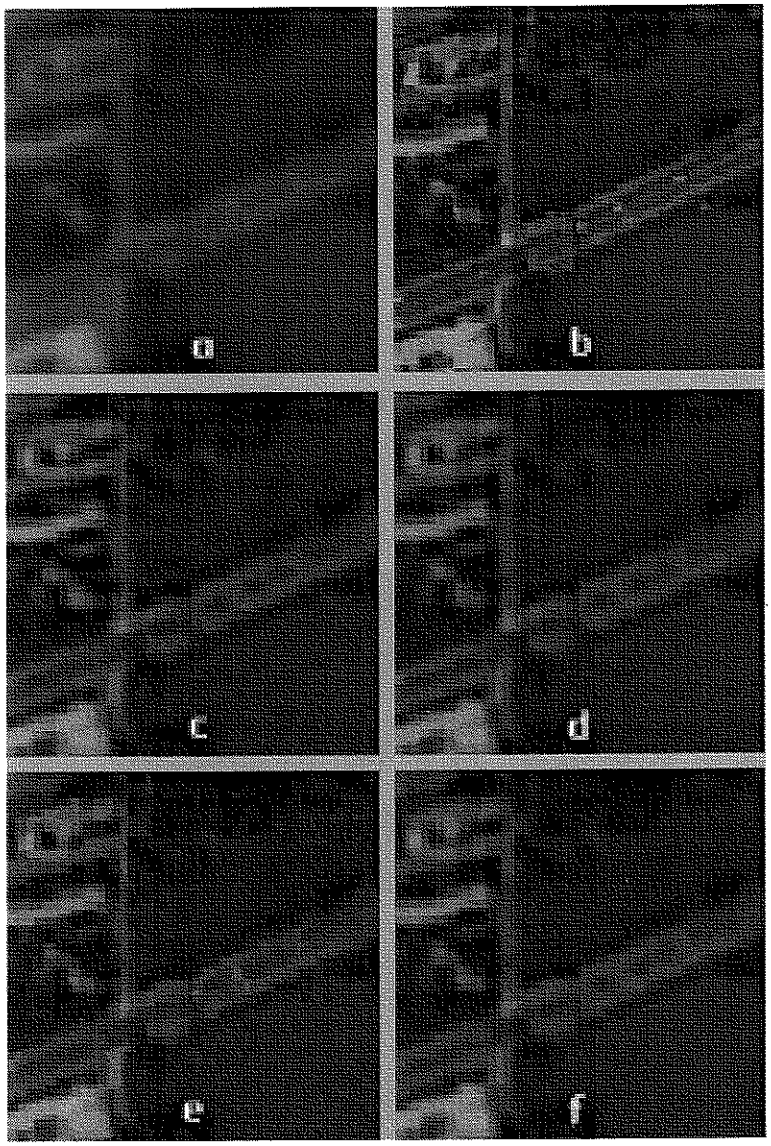

Figure 14: Restoration of the degradation model described in (17).
a: Blurred image.
b: Reference image.
c: Shrinkage in Basis 1, without multi-level approach.
d: Multi-level shrinkage in Basis 2.
e: Rudin-Osher-Fatemi method.
f: Modified Rudin-Osher-Fatemi method.

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

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

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

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

[15] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *SIAM, J. Num. Anal.*, 2000. To appear, a preliminary version is available at: http://www.cmla.ens-cachan.fr/~malgouy.

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

[17] 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.

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

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

[20] 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/.

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

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