INTRODUCTION TO THE RESTORATION OF ASTROPHYSICAL IMAGES BY MULTISCALE TRANSFORMS AND BAYESIAN METHODS

A. Bijaoui[1]

Abstract. The image restoration is today an important part of the astrophysical data analysis. The denoising and the deblurring can be efficiently performed using multiscale transforms. The multiresolution analysis constitutes the fundamental pillar for these transforms. The discrete wavelet transform is introduced from the theory of the approximation by translated functions. The continuous wavelet transform carries out a generalization of multiscale representations from translated and dilated wavelets. The *à trous* algorithm furnishes its discrete redundant transform. The image denoising is first considered without any hypothesis on the signal distribution, on the basis of the *a contrario* detection. Different softening functions are introduced. The introduction of a regularization constraint may improve the results. The application of Bayesian methods leads to an automated adaptation of the softening function to the signal distribution. The MAP principle leads to the *basis pursuit*, a sparse decomposition on redundant dictionaries. Nevertheless the posterior expectation minimizes, scale per scale, the quadratic error. The proposed deconvolution algorithm is based on a coupling of the wavelet denoising with an iterative inversion algorithm. The different methods are illustrated by numerical experiments on a simulated image similar to images of the deep sky. A white Gaussian stationary noise was added with three levels. In the conclusion different important connected problems are tackled.

1 Introduction

The astrophysical images observed by modern instruments are today currently enhanced by digital processing. In particular, many efforts are done for their denoising and their deblurring. These operations are often coupled in a global

---

[1] University of Nice Sophia Antipolis, UMR CNRS 6202, OCA, BP. 4229, 06304 Nice Cedex 04, France

image restoration. Generally the problem is written as:

$$\mathbf{Y} = \mathbf{AX} + \mathbf{N};\qquad\qquad(1.1)$$

where $\mathbf{Y}$, $\mathbf{X}$, $\mathbf{A}$ and $\mathbf{N}$ are respectively the observed image, the image to be restored, the blurring linear operator and the noise image.

The inversion is conditioned by two features: the matrix singularities and the noise level. In the case of a regular blurring matrix and a signal without noise, the inversion is unique, the problem being only to minimize the number of operations. In general, the noise level and the matrix singularities lead to inconsistency and instability. Statistical rules are needed to define the correct solution. A regularity constraint is also required to select the best one from a given criterion.

In this context, the image representation plays an important part. In the general case of a space invariant *Point Spread Function* (PSF), the matrix product corresponds to a convolution in the direct space and a filtering in the Fourier one. Thus, the matrix singularities are associated to the frequency holes. Thus, a method based on the Fourier transform can not fill these holes without a constraint. In the case of a representation different from a Fourier series, this filling becomes possible. This is the case for the CLEAN algorithm (Högbom 1974) based on the consideration that the information consists in sparse Dirac peaks.

The representation plays an important part for the denoising. Its quality depends on the efficiency to concentrate the information into the minimum number of coefficients; these coefficients being obtained by a suitable transform.

Multiscale transforms were early developed and applied for the image processing (Starck *et al.* 1998; Mallat 1998). The *Multiresolution Theory* developed in 80's is a beautiful framework to get multiscale representations (Mallat 1989). It leads to the *Discrete Wavelet Transform* (DWT). Closely related redundant transforms, connected to the *continuous wavelet transform* (CWT) (Morlet *et al.* 1982) carried out better results (Raphan & Simoncelli 2008). Nevertheless a correct CWT development needs the multiresolution theory.

The present paper constitutes an introduction to this large topic. In Section 2, the multiresolution theory is developed in the context of the approximation theory from translated scaling functions. In Section 3, the CWT is then described. It is shown that the use of scaling functions unifies DWT and CWT. The denoising is examined in Section 4 from different thresholding rules. In Section 5, a first Bayesian approach, derived from the Maximum a Posteriori (MAP), leads to the Basis Pursuit (BP). The Bayesian posterior mean is applied in Section 6. In Section 7, an application to a deconvolution problem is developed. Finally, in the conclusion, different uncovered problems are scanned.

2   The multiresolution theory and the DWT

2.1   The approximation by translated functions

The Shannon interpolation. The Shannon interpolation theorem was a milestone in the signal processing progress (Shannon 1948). A function $f(x) \in L^2(\mathbb{R})$ is

interpolated from regularly spaced samples thanks to the relation:

$$f_0(x) = \sum_{k=-\infty}^{+\infty} f(kh)\mathrm{sinc}\left(\frac{x}{h} - k\right);\qquad(2.1)$$

where $\mathrm{sinc}(x)$ is the sine cardinal function ($\frac{\sin \pi x}{\pi x}$) and h the sampling step. The interpolation is perfect ($f_0(x) \equiv f(x)$) if $h \leq h_0$, h_0 being the Nyquist-Shannon step:

$$h_0 = \frac{1}{2\nu_0};\qquad(2.2)$$

where ν_0 is the cut-off frequency of the function $f(x)$. In practice, this theorem is not directly applied due to the slow convergence of the sine cardinal function. It introduced the idea of interpolations based on translated functions that played a fundamental part for the building of the multiresolution theory.

The L^2 approximation by translated functions. We set now (Schoenberg 1946; Strang & Fix 1971):

$$f_0(x) = \sum_{k=-\infty}^{+\infty} a(k)\varphi(x - k).\qquad(2.3)$$

Compared to Equation (2.1), the sampling step is set to 1, the sine cardinal function is changed to the φ one, the values at the interpolation mesh $f(nh)$ are changed to the $a(k)$ coefficients. The goal is not to get $f_0(k) \equiv f(k)$. Here, we search the coefficients $a(k)$ such that the distance between the functions $f(x)$ and $f_0(x)$ is minimum in the $L^2(\mathbb{R})$ space, *i.e.*:

$$R = \int_{-\infty}^{+\infty} |f(x) - f_0(x)|^2 dx\qquad(2.4)$$

is minimum. Taking into account Equation (2.3) we get:

$$\int_{-\infty}^{+\infty} \varphi(x - k)[f(x) - \sum_{l=-\infty}^{+\infty} a(k)\varphi(x - l)]dx = 0.\qquad(2.5)$$

The following equation is derived:

$$c(k) = \sum_{l=-\infty}^{+\infty} a(l)A(k - l);\qquad(2.6)$$

with:

$$A(k - l) = \int_{-\infty}^{+\infty} \varphi(x - k)\varphi(x - l)dx;\qquad(2.7)$$

and

$$c(k) = \int_{-\infty}^{+\infty} f(x)\varphi(x - k)dx \equiv\, <f(x), \varphi(x - k)> .\qquad(2.8)$$

Equation (2.6) is a discrete convolution which can be solved by the application of the Fourier transform:

$$\hat{c}(\nu) = \hat{a}(\nu) \sum_{n=-\infty}^{+\infty} \hat{A}(\nu + n); \qquad (2.9)$$

with $\hat{A}(\nu) = |\hat{\varphi}(\nu)|^2$. The inversion is possible if:

$$\hat{S}(\nu) \equiv \sum_{n=-\infty}^{+\infty} |\hat{\varphi}(\nu + n)|^2 \neq 0. \qquad (2.10)$$

$\varphi(x)$ is called the scaling function. $c(k)$ is a weighted mean of $f(x)$ around k. $f_0(x)$ is the projection of $f(x)$ into a subspace V_0 of $L^2(R)$.

Duality and orthogonal scaling functions. If $\hat{S}(\nu) = 1$ $a(k) = c(k)$, the approximation is easily computed from the scalar products $c(k)$.

If, more generally, Relation 2.10 is satisfied for all frequencies, we can derive a new scaling function $\bar{\varphi}(x)$ from the Fourier transform of the initial one (Daubechies *et al.* 1986):

$$\hat{\bar{\varphi}}(\nu) = \frac{\hat{\varphi}(\nu)}{\sqrt{\hat{S}(\nu)}}. \qquad (2.11)$$

The set $\{\bar{\varphi}(x - k)\}$ is an orthonormal basis of the V_0 subspace. Here the $a(k)$ coefficients are identical to the $c(k)$ ones. The same scaling function is used for the analysis $(c(k))$ and the synthesis $(a(k))$. This is the case for the Shannon interpolation, the sine cardinal function being an orthogonal scaling function.

In the case of a non orthogonal scaling function, it is also convenient to introduce the dual scaling function:

$$\hat{\tilde{\varphi}}(\nu) = \frac{\hat{\varphi}(\nu)}{\hat{S}(\nu)}. \qquad (2.12)$$

In this framework, it results that:

$$f_0(x) = \sum_{k=-\infty}^{+\infty} c(k)\tilde{\varphi}(x - k). \qquad (2.13)$$

Here the $c(k)$ coefficients are also identical to the $a(k)$ ones. But it is not the same scaling function used for the analysis $(c(k))$ and the synthesis $(a(k))$.

Normalization of the scaling function. In Equation (2.3) the scaling function is considered without normalization. The orthonormal scaling functions associated to Equation (2.11) have, by construction, their square integral equal to 1. This is the general setting in the framework of the multiresolution theory. Nevertheless, it could be also convenient to choose the integral equal to 1. In this case the approximation coefficients are local means, weighted by the scaling function.

The Shannon scaling function. It is easy to show that the shifted sine cardinal functions with integers are orthogonal. The scaling function is here:

$$\varphi(x) = \frac{\sin \pi x}{\pi x}. \tag{2.14}$$

The corresponding V_0 subspace is the one of the functions having a frequency support in $[-0.5, 0, 5]$. The approximation for a function belonging to V_0 corresponds to the Shannon interpolation. It can be noted that this approximation is invariant by translation.

The Haar scaling function. The characteristic function, $H(x) = 1$ for $x \in [0, 1]$ and null outside this interval, is called the Haar scaling function. The functions shifted with integers are orthogonal. The corresponding V_0 subspace is the space of the staircase functions. Note that this approximation is only invariant by an integer shift.

2.2   The pyramid of resolution

Scale modification and the dilation equation. The scaling function is dilated by a factor a, the approximation coefficients become:

$$c(a, k) = < f(x), \frac{1}{a}\varphi\left(\frac{x}{a} - k\right) > . \tag{2.15}$$

The factor $\frac{1}{a}$ is introduced to keep constant the integral of the dilated scaling function. In the case of an orthonormal scaling function, the factor becomes $\frac{1}{\sqrt{a}}$ in order to keep the square integral equal to 1.

There is a linear relation between $c(k)$ and $c(a, k)$ if the scaling function satisfies to the dilation equation (Strang 1989):

$$\frac{1}{a}\varphi(\frac{x}{a}) = \sum_{n=-\infty}^{+\infty} h_a(n)\varphi(x - n). \tag{2.16}$$

It results that

$$c(a, k) = \sum_{n=-\infty}^{+\infty} h_a(n)c(ak + n). \tag{2.17}$$

In this framework, the function $f(x)$ has to be known only by its approximation coefficients $c(k)$. Note that the number of coefficients (for a finite set) is reduced by a factor a. Most often $a = 2$, this leads to the so-called dyadic analysis. The resulting approximation $f_a(x)$ belongs to a subset V_a which is embedded in V_0.

The resolution pyramid. The dilation of the scaling function may be iterated, leading to the approximations $f_0(x)$, $f_a(x)$, $f_{a^2}(x)$, ... These functions constitute the pyramid of resolution associated to this analysis. The functions belong to the subsets $V_0 \supset V_a \supset V_{a^2} \ldots$. For a finite initial number of approximation coefficients, their number is divided by a at each iteration step.

Examples of scaling functions. It is easy to show that the sine cardinal function obeys to the dilation equation, whatever the integer a. The generated subspace corresponds to the functions with a frequency bandwidth 1, $\frac{1}{a}$, $\frac{1}{a^2}$, ...

The Haar scaling function obeys also to the dilation equation:

$$H_a(x) = \frac{1}{a}[H(x) + H(x-1) + \ldots + H(x-a+1)]. \tag{2.18}$$

The B-spline functions (Hou & Andrews 1978) generalize the Haar one. Its centered version is defined by its Fourier transform:

$$\hat{B}_l(\nu) = \text{sinc}^{l+1}(\nu). \tag{2.19}$$

The Fourier transform of its dilated version is:

$$\hat{B}_{l,a} = \text{sinc}^{l+1}(a\nu). \tag{2.20}$$

Its quotient with $B_l(\nu)$ is:

$$\hat{h}_{l,a}(\nu) = \frac{\text{sinc}^{l+1}(a\nu)}{\text{sinc}^{l+1}(\nu)}. \tag{2.21}$$

It is easy to show that it is a 1-periodic function. In particular, for $a = 2$ we get:

$$\hat{h}_l = \cos^{l+1}(\nu); \tag{2.22}$$

which leads to:

$$h_l(n) = \frac{1}{2^{l+1}} C_{l+1}^{\frac{l+1}{2}-n}. \tag{2.23}$$

The cubic B-spline is often used (Starck *et al.* 1998). Its coefficients are:

$$h_3(n) = \frac{1}{16} C_4^{2-n}. \tag{2.24}$$

Case of an orthonormal scaling function. From Equation (2.12) we derive:

$$\sum_{n=-\infty}^{+\infty} |\hat{\varphi}(\nu+n)|^2 = 1. \tag{2.25}$$

That leads directly to:

$$\sum_{n=-\infty}^{+\infty} |\hat{\varphi}(2\nu+n)|^2 = 1. \tag{2.26}$$

The dilation equation in the Fourier space is written as:

$$\hat{\varphi}(2\nu) = \hat{h}(\nu)\hat{\varphi}(\nu). \tag{2.27}$$

We have:

$$\sum_{n=-\infty}^{+\infty} |\hat{\varphi}(2\nu + n)|^2 = \sum_{n=-\infty}^{+\infty} |\hat{\varphi}(2\nu + 2n)|^2 + \sum_{n=-\infty}^{+\infty} |\hat{\varphi}(2\nu + 2n + 1)|^2 = 1 \quad (2.28)$$

Relation 2.27 is applied:

$$\sum_{n=-\infty}^{+\infty} |\hat{h}(\nu + n)|^2 |\hat{\varphi}(\nu + n)|^2 + \left|\hat{h}\left(\nu + n + \frac{1}{2}\right)\right|^2 \left|\hat{\varphi}\left(\nu + n + \frac{1}{2}\right)\right|^2 = 1 \quad (2.29)$$

Taking Relation 2.25 and taking into account the periodicity of the function $\hat{h}(\nu)$ it results finally:

$$|\hat{h}(\nu)|^2 + \left|\hat{h}\left(\nu + \frac{1}{2}\right)\right|^2 = 1. \quad (2.30)$$

2.3   The 1D multiresolution analysis

The complementary subspace. As $V_1 \subset V_0$, we can write:

$$f_0(x) = f_1(x) + g_1(x); \quad (2.31)$$

where $f_0 \in V_0$ and $f_1 \in V_1$. g_1 is a function of the complementary subspace W_1 of V_1 in V_0, *i.e.* $V_0 = V_1 + W_1$.

The wavelet basis. $g_1(x)$ can be written as:

$$g_1(x) = \sum_{k=-\infty}^{\infty} w(1,k)\tilde{\psi}\left(\frac{x}{2} - k\right). \quad (2.32)$$

The detail coefficients $w(1,k)$ are obtained by projection on a translated set:

$$w(1,k) = < f, \frac{1}{2}\psi\left(\frac{x}{2} - k\right) > . \quad (2.33)$$

The $w(1,k)$ computation from the $c(k)$ ones, requires that:

$$\frac{1}{2}\psi\left(\frac{x}{2}\right) = \sum_{n} g(n)\varphi(x - n); \quad (2.34)$$

i.e. that $\frac{1}{2}\psi(\frac{x}{2})$ belongs to V_0. That leads to the relation:

$$w(1,k) = \sum_{n} g(n)c(2k + n). \quad (2.35)$$

Orthogonal wavelets. In this case we have:

$$\sum_{n=-\infty}^{+\infty} |\hat{\psi}(\nu + n)|^2 = 1. \quad (2.36)$$

Equation (2.34) is equivalent to:

$$\hat{\psi}(2\nu) = \hat{g}(\nu)\hat{\psi}(\nu). \tag{2.37}$$

Thus, it is derived that:

$$|\hat{g}(\nu)|^2 + |\hat{g}\left(\nu + \frac{1}{2}\right)|^2 = 1. \tag{2.38}$$

The subspace V_1 and W_1 being orthogonal we have:

$$\sum_n \hat{\varphi}(\nu + n)\hat{\psi}^*(\nu + n) = 0. \tag{2.39}$$

The following relation is derived:

$$\hat{h}(\nu)\hat{g}^*(\nu) + \hat{h}\left(\nu + \frac{1}{2}\right)\hat{g}^*\left(\nu + \frac{1}{2}\right) = 0. \tag{2.40}$$

Filters h and g satisfying Relations 2.30, 2.38 and 2.40 generate conjugate orthogonal scaling and wavelet functions. For a given filter h obeying to 2.30, we can associate the filter g given by the relation:

$$\hat{g}(\nu) = e^{-2i\pi\nu}\hat{h}^*\left(\nu + \frac{1}{2}\right) \tag{2.41}$$

h and g are called *Quadrature Mirror Filters* (QMF) (Esteban & Galland 1977).

Reconstruction. Any V_0 basis function $\varphi(x - k)$ can be written as a sum of V_1 and W_1 base functions:

$$\varphi(x - k) = 2[\sum_l h(k - 2l)\varphi_1(x - 2l) + g(k - 2l)\psi_1(x - 2l)]. \tag{2.42}$$

That leads to get the approximation coefficients by projection:

$$c(k) = 2[\sum_l h(k - 2l)c(1, l) + g(k - 2l)w(1, l)]. \tag{2.43}$$

The multiresolution analysis. From the approximation coefficients $c(1, k)$, it is possible to iterate by a new dilation of the scaling and of the wavelet functions. By iteration we obtained a set of details $w(i, k)$ such that the function $f(x)$ can be written as:

$$f(x) = \sum_{i,k=-\infty}^{\infty} w(i, k)\tilde{\psi}\left(\frac{x}{2^i} - k\right) \tag{2.44}$$

for any function of the $L^2(\mathbb{R})$ space (Mallat 1989).

The recurrence formulae and the filter bank. The previous developments are summarized as:

Approximation $c(i,k) = \sum_n h(n)c(i-1, 2k+n)$;

Wavelets $w(i,k) = \sum_n g(n)c(i-1, 2k+n)$;

Reconstruction $c(i,k) = 2[\sum_l \tilde{h}(k+2l)c(i+1,l) + \tilde{g}(k+2l)c(i+1,l)].]$

In the case of orthogonal scaling and wavelet functions, $\tilde{h}(n) = h(-n)$ and $\tilde{g}(n) = g(-n)$. The resulting algorithm flow-chart is drawn in Figure 1. This algorithm is known as the filter bank one (Vitterli 1986). The data vector inputs at the top left. It is convolved with the two filters H (low passband) and G (high passband). The resulting vectors are decimated, by removing every other point. The smoothed values are convolved again, and so on, up to get one value. The restoration consists to introduce a 0 between two approximation or two detail coefficients. We start from the bottom right, and progressively the signal is restored from the largest scale to the smallest one. The convolutions are done with the filter $\tilde{H}$ and $\tilde{G}$.
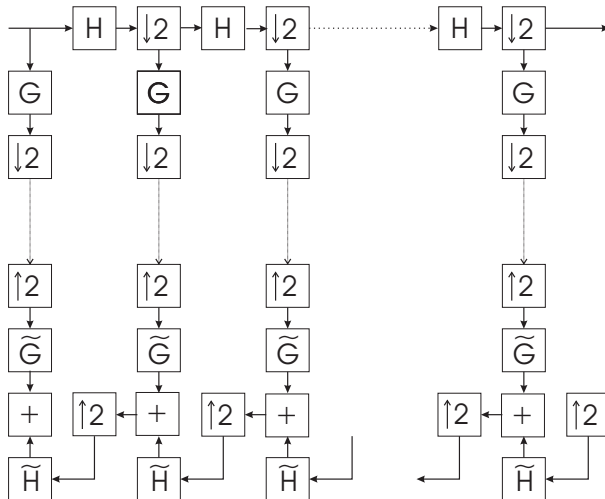

Fig. 1. Flow-chart of the filter bank algorithm.

The algorithm was developed from the multiresolution theory. But it is more general. The restoration is perfect only if the filters h, $\tilde{h}$, g and $\tilde{g}$ satisfied the following conditions called perfect reconstruction and antialiasing conditions:

$$\hat{h}(\nu)\hat{\tilde{h}}(\nu) + \hat{g}(\nu)\hat{\tilde{g}}(\nu) = 1; \tag{2.45}$$

$$\hat{h}(\nu)\hat{\tilde{g}}(\nu) + \hat{h}\left(\nu + \frac{1}{2}\right)\hat{\tilde{g}}\left(\nu + \frac{1}{2}\right) = 0. \tag{2.46}$$

The Daubechies wavelets. The Haar transform is associated to the filters:

$$h(0) = h(1) = \frac{1}{\sqrt{2}} \qquad h(n) = 0 \quad n \neq (0,1). \tag{2.47}$$

The corresponding high pass filter is

$$g(0) = \frac{1}{\sqrt{2}} \qquad g(1) = -\frac{1}{\sqrt{2}} \qquad h(n) = 0 \quad n \neq (0,1). \tag{2.48}$$

The algorithm is very fast. K operations are required for the transform and its inverse for a signal with K elements.

Daubechies (1988) generalized the Haar transform with compact filters. They are widely applied in modern signal processing. Later on in this paper, Daubechies' filters of length 8 are applied.

2.4 The 2D multiresolution

The 2D approximation by translated scaling functions. The concept of approximation by translated scaling functions in $L^2(\mathbb{R})$ is easily extended to 2 (and more) dimensions. If $f(x, y)$ is the function to be approximated and $\varphi(x, y)$ the scaling function, the approximation coefficients are:

$$c(0, k, l) = < f(x, y), \varphi(x - k, y - l) > . \tag{2.49}$$

The corresponding approximation is:

$$f_0(x, y) = \sum_{k,l} c(0, k, l) \tilde{\varphi}(x - k, y - l) \tag{2.50}$$

where $\tilde{\varphi}(x, y)$ is the dual scaling function. This function exists if:

$$\sum_{n,m} |\hat{\varphi}(u + n, v + m)|^2 \neq 0. \tag{2.51}$$

The approximation is a $f(x, y)$ projection on the V_0 subspace of $L^2(\mathbb{R}^2)$.

The 2D dilation equation. The approximation for a scaling function dilated by a factor a in each direction can be computed from the $c(0, k, l)$ coefficients if:

$$\frac{1}{a^2} \varphi(\frac{x}{a}, \frac{y}{a}) = \sum_{n,m} h(n, m) \varphi(x - n, y - m). \tag{2.52}$$

Most often the variables are separated:

$$\varphi(x, y) \equiv \varphi(x) \varphi(y) \tag{2.53}$$

where $\varphi(x)$ satisfies the 1D dilation equation.

The wavelets. Taking into account the variable separation the V_0 subspace is divided in four subspaces:

V_1 the subspace corresponding to the approximation at scale 2. It is computed with the filter $h(n)h(m)$. The scaling function is $\frac{1}{4}\varphi(\frac{x}{2})\varphi(\frac{y}{2})$.

$W_{1,h}$ associated to the horizontal details. The wavelet coefficients are computed with the filter $g(n)h(m)$. The wavelet function is $\frac{1}{4}\psi(\frac{x}{2})\varphi(\frac{y}{2})$.

$W_{1,v}$ associated to the vertical details. The wavelet coefficients are computed with the filter $h(n)g(m)$. The wavelet function is $\frac{1}{4}\varphi(\frac{x}{2})\psi(\frac{y}{2})$.

$W_{1,d}$ associated to the diagonal details. The wavelet coefficients are computed with the filter $g(n)g(m)$. The wavelet function is $\frac{1}{4}\psi(\frac{x}{2})\psi(\frac{y}{2})$.

In this framework, the 2D filter bank algorithm is easily deduced from the 1D one.

3   The continuous wavelet transform

3.1   Generalities

Definition and main properties. The Morlet-Grossmann definition of the continuous wavelet transform (Grossmann & Morlet 1984) for a 1D signal $f(x) \in L^2(\mathbb{R})$ is:

$$W(a,b) = N(a) \int_{-\infty}^{+\infty} f(x)\psi^* \left(\frac{x-b}{a} \right) dx; \qquad (3.1)$$

where z^* notes the complex conjugate of z, $\psi^*(x)$ is the analyzing wavelet, $a\ (>0)$ is the scale parameter and b is the position parameter. Grossmann & Morlet set $N(a) = \frac{1}{\sqrt{a}}$, but it is often convenient to set $N(a) = \frac{1}{a}$. The transformation is linear, covariant under translations and under dilations. The last property makes the wavelet transform very suitable for analyzing hierarchical structures. It is like a mathematical microscope with properties that do not depend on the magnification.

Inversion. Consider now a function $W(a,b)$ which is the wavelet transform of a given function $f(x)$. $f(x)$ can be restored by using the formula (Grossmann & Morlet 1984):

$$f(x) = \frac{1}{C_\psi} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{a}} W(a,b)\psi \left(\frac{x-b}{a} \right) \frac{da.db}{a^2}; \qquad (3.2)$$

where:

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\nu)|^2}{\nu} d\nu. \qquad (3.3)$$

The reconstruction is only available if C_ψ is defined (admissibility condition). This condition is generally true if $\hat{\psi}(0) = 0$, *i.e.* the mean of the wavelet function is 0.

3.2   The discrete wavelet transform

The transform sampling. The image sampling is generally made according to the Shannon theorem. The discrete wavelet transform (DWT) can be derived from

this theorem. If the wavelet function has no cut-off frequency, the transform cut-off frequency is the signal one, whatever the scale. So if K is the number of signal elements and I the number of scales the transform has KI elements.

In the case of a wavelet function having the cut-off frequency $\frac{1}{2}$, at each dyadic scale the cut-off frequency is divided by two. Thus, the sampling step can be multiplied by a factor 2. The total transform length becomes about $2K$.

Generally the scales are sampled according to a 2^i law. Nevertheless it is not guaranteed that all the information on the CWT is kept by this sampling.

Direct transformations. The DWT can be obtained directly by convolution, using a compact wavelet function. As the scale increases, the CPU time increases in proportion. So, in practice this method is not easy to use.

It is possible to work in the Fourier space, computing the transform scale by scale. The number of elements for a scale can be reduced for a wavelet having a cut-off frequency. Here, the CPU time is proportional to $K \log(K)$.

The transform from the filter bank. In fact, in the previous section, beyond the multiresolution theory, we examined a fast DWT algorithm based on the filter bank. The transform size is K and the computing time is proportional to K.

3.3   The wavelet approximation and the à trous algorithm

The sampled wavelet function at scale 1. Let us consider a real wavelet function which can be written as an approximation from translated scaling functions (Eq. (2.34)). Let us admit that we know the sampled approximation coefficients $c(0, k) = < f(x), \varphi(x - k) >$. The sampled continuous wavelet function at scale $a = 2$ can written as $(N(a) = 1/a)$:

$$w(1, k) = \sum_n g(n)c(0, k + n). \tag{3.4}$$

This expression is similar to 2.3, but the array is not decimated, the factor 2 being not present. Note that in the following, the scale of the wavelet transform will design the exponent i of the true scale $a = 2^i$.

The recurrence expressions. We want to compute $w(2, k)$ using a similar formula. The scaling function $\varphi(x)$ is chosen to satisfy the dilation Equation (2.16). In this framework, we have for the approximation coefficients:

$$c(i + 1, k) = \; < f(x), \frac{1}{2^{i+1}} \varphi \left(\frac{x}{2^{i+1}} - k \right) > \; = \; < f(x), \frac{1}{2^i} \sum_n h(n) \varphi \left(\frac{x}{2^i} - k \right) >; \tag{3.5}$$

which leads to:

$$c(i + 1, k) = \sum_n h(n)c(i, k + 2^i n). \tag{3.6}$$

Similarly we get:

$$w(i + 1, k) = \sum_n g(n)c(i, k + 2^i n). \tag{3.7}$$

The interpolation and the wavelet coefficients are computed with a linear operation similar to a convolution but we jump a set of $2^i - 1$ points. For that reason, this algorithm is called the *à trous* algorithm (algorithm with holes) (Holschneider *et al.* 1989). The flow-chart of this algorithm is drawn in Figure 2. A set of $K \log_2 K$ values of the wavelet transform is obtained, with a number of operations proportional to $K \log_2 K$.
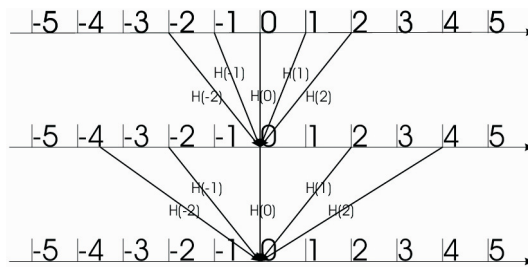

Fig. 2. Flow-chart of the à trous algorithm.

The inversion. Here, the transform is undecimated. Thus we get a transform size larger than the input signal one. The inverse system is over-determined. If the $\{w(i,k)\}$ set is a wavelet transform, it is easy to restore the approximation scale by scale using dual filters $\tilde{h}$ and $\tilde{g}$ which have to satisfy the relation:

$$\hat{h}(\nu)\hat{\tilde{h}}(\nu) + \hat{g}(\nu)\hat{\tilde{g}}(\nu) = 1. \tag{3.8}$$

The filters choice is large. A simple algorithm consists in making the difference between two approximations (Bijaoui *et al.* 1994):

$$w(i+1,k) = c(i,k) - c(i+1,k). \tag{3.9}$$

Here the inversion is obvious.

The inversion of a modified wavelet transform. If the $\{w(i,k)\}$ set is not the wavelet transform of a given signal, nevertheless a signal $\{c(0,k)\}$ will be obtained by inversion. But its wavelet transform $\{w_s(i,k)\}$ can be different from $\{w(i,k)\}$. This point is important for image restoration. In this framework most often a softening rule is applied on the wavelet coefficients. There is a duality between the wavelet transform and the signal for the orthogonal DWT. But this duality vanishes for the redundant undecimated wavelet transform (UDWT) associated to the *à trous* algorithm. Some cautions have to be taken for the inversion. A classical solving method consists to obtain the orthogonal projection of $\{w(i,k)\}$ in the space generated by the wavelet transforms of all the signals. That corresponds to get the set $\{c(0,k)\}$ such that its wavelet transform $\{w_s(i,k)\}$ has the minimum distance to the input set $\{w(i,k)\}$. Obviously the inversion algorithm is slowed.

The shift-invariant wavelet transform. Coifman & Donoho (1995) introduced a variant of the *à trous* algorithm based on the orthogonal wavelet transform. The

transform becomes shift invariant by removing the decimation of the coefficients (approximation and wavelet). The inversion is done by taking into account the coefficient interleaving generated by the lack of decimation. A same pixel is thus many times reconstructed, the mean is done in the proposed algorithm. This algorithm has the advantage to inverse from the filter bank.

The pyramidal transform. The undecimated wavelet transform may correspond to a too important data array for large images. At each step of the algorithm the approximation coefficients can be removed without decimating the wavelet array. We get a pyramidal set of values. The number of data is now $2N$ and the number of operations is proportional to N. The inversion is based on an orthogonal projection, obtained by an iterative algorithm.

3.4   The two-dimensional continuous wavelet transform

General definition. The wavelet dilation is not necessarily isotropic, *i.e.* identical whatever the direction. But it can be seen as a dilation in two orthogonal directions. The reference frame can be also rotated with a θ angle. That leads to the coordinate transform:

$$R(x, y, a_x, a_y, \theta) = \begin{pmatrix} \frac{1}{a_x} & 0 \\ 0 & \frac{1}{a_y} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{3.10}$$

The wavelet transform becomes:

$$w(b_x, b_y, a_x, a_y, \theta) = N(ab) < f(x, y), \psi^*(|R(x - b_x, y - b_y, a_x, a_y, \theta)|) >; \tag{3.11}$$

where $|R|$ designs the module of the vector R. The resulting transform is thus a 5D function. Its sampling rules are not evident, especially for the angular variable.

The isotropic 2D CWT. Most often the 2D CWT is applied in its simplified isotropic version:

$$w(b_x, b_y, a) = N(a^2) < f(x, y), \psi^*\left(\frac{x}{a}, \frac{y}{a}\right) > . \tag{3.12}$$

The two dimensional à trous algorithm. The *à trous* and the pyramidal algorithms can be easily transposed in two dimensions taking into account the use of a scaling function which satisfies to the 2D dilation equation. The algorithms are simplified in the case of a variable separation ($\varphi(x, y) \equiv \varphi(x)\varphi(y)$). The *à trous* computation of the approximation coefficients is obtained with the expression:

$$c(i + 1, k, l) = \sum_{n,m} h(n)h(m)c(i, k + 2^i n, l + 2^i m); \tag{3.13}$$

while the wavelet coefficients are computed with:

$$w(i + 1, k, l) = \sum_{n,m} g(n, m)c(i, k + 2^i n, l + 2^i m). \tag{3.14}$$

The wavelet function is not necessarily separated in the two directions. For example, the wavelet associated to the difference between two successive approximations is not separable.

Quasi isotropic scaling functions. For the coherence of the method it would be convenient to process the data with an isotropic scaling function. The Gaussian is the single function which is separable and isotropic. But this function does not satisfy the dilation equation.

The centered B-splines tends for an increasing index to the Gaussian. Thus, its use allows a quasi isotropic analysis with fast computations, for the *à trous* and for the pyramidal algorithms.

4 Image denoising from significant coefficients

4.1 The significant coefficients

The quality criterion. Let us consider a discrete noisy signal $\mathbf{Y} = \mathbf{X} + \mathbf{N}$. $\mathbf{X}$ is the true signal vector and $\mathbf{N}$ its associated noise. The signal denoising consists into the operation $O(\mathbf{Y}) \rightarrow \bar{\mathbf{X}}$ such that this vector is the closest to $\mathbf{X}$. The distance criterion depends on the noise statistics. The case of a stationary white Gaussian noise is only examined in the present paper. Some methods adapted to other noise statistics are indicated in the conclusion. For this statistical distribution, the Euclidian distance is the universal criterion. It is converted into the Signal to Noise Ratio (SNR) defined as:

$$SNR = 10 \log_{10} \frac{|\bar{\mathbf{X}} - \mathbf{X}|^2}{|\mathbf{X}|^2}. \tag{4.1}$$

The distribution of the transform coefficients. Let us admit that an orthonormal transform is applied on $\mathbf{Y}$. That corresponds to apply a rotation in the signal space. Thus, the noise is still Gaussian, stationary and white. This operation seems to be useless. But the transform can deeply modify the signal statistics. For example, let us consider a signal which is spatially quite uniformly distributed. Even if the pixel distribution law seems to do not depend on the position, its Fourier transform at the lowest frequencies correspond generally to the highest coefficient values; while the values at the highest frequencies can appear very faint compared to the noise deviation. The Wiener denoising (Wiener 1949) is a filtering based on the ratio between the signal and the noise at each frequency. This filter takes thus into account the information content.

This separation between the low and the high frequencies comes at the cost of a space delocalization. At the contrary of the Fourier transform, the DWT allows both a space and a frequency (scale) representation. Even if there is no global information detected at small scales, few coefficients could be significant. A wavelet denoising would restore this information while the Fourier filtering would remove it.

The best transform for the denoising is the one which optimizes the separation between the signal and the noise. That depends on the considered images; but many experiments showed that the DWT is well-adapted for the astronomical images.

Mean coefficient property. Let us consider a DWT coefficient, it can be written as:

$$w(i, k, l) = \sum_{n,m} g_i(n, m) c(k + n, l + m);$$ (4.2)

where $\{c(k, l)\}$ is the discrete signal and $g_i(n, m)$ is the discrete wavelet filter at scale i. The filter is a pass-band one, thus we have:

$$\sum_n g_i(n, m) = 0.$$ (4.3)

By consequence the mean DWT coefficient is also equal to 0 whatever the image background. The distribution of the wavelet coefficients is centered at each scale.

Now, if the signal is constant on the support of the filter (admitted to be compact), the wavelet coefficient is null. Due to the noise, the distribution of the observed coefficients would be a centered Gaussian with a deviation equal to the noise deviation σ for an orthonormal transform.

Significant coefficients. Let us consider a coefficient $w(i, k, l)$. If its value is positive, we consider the probability $p = Prob[W > w(i, k, l)]$, where W is the stochastic variable associated to the noise distribution of the wavelet coefficient. $p < \epsilon$ means that the probability of getting the value from a constant signal is fainter that the significance level ϵ. That leads to introduce a threshold $T(\epsilon)$ such that:

$$Prob(w(i, k) > T) < \epsilon \qquad \text{or} \qquad Prob(w(i, k) < -T) < \epsilon.$$ (4.4)

For a Gaussian distribution, the significance ϵ is translated into a factor of the noise deviation ($T = \kappa\sigma$). Note that, if the image has 1000×1000 pixels and $\epsilon = 0.001$, statistically 1000 positive coefficients (false alarms) appear significant for a noisy uniform image. 1000 negative coefficients also appear significant. The false alarm rate is identical to the significance threshold. Here this threshold is equal to 3.09σ, σ being the standard deviation of the Gaussian noise distribution.

4.2   Denoising from thresholdings

The material for the experiments. The restoration tests were done on a simulated image (Mel1) composed as a sum of $2D$ Gaussian functions. Their amplitudes are distributed according to a power law, in order to get an image like astronomical ones. A white Gaussian noise at the levels 0.007, 0.07 and 0.7 was added. In Figure 3 the simulated 256×256 images are displayed. Their SNRs are respectively 14.73, -5.27 and -25.27 dB.

Hard thresholding (HT). The basic method consists into the image reconstruction from only the significant coefficients, according to the threshold T (Starck & Bijaoui 1994).
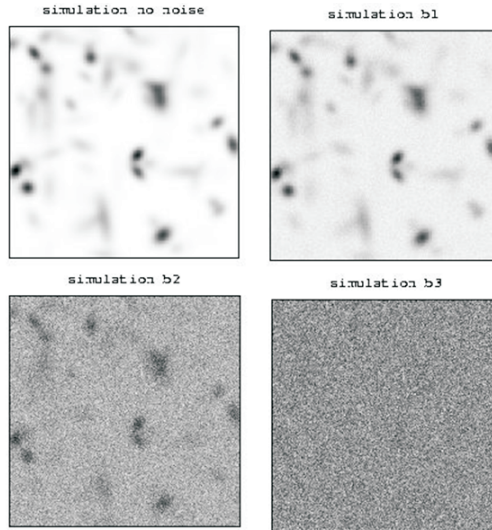
Fig. 3. The simulated images used for the tests by lexicographic order: reference and noisy images with increasing noise level b1, b2 and b3.

Soft thresholding (ST). The images restored with the previous method display some punctual defaults due to the discontinuities introduced by the hard thresholding. Donoho (1995) proposed to soften them with the following rules

$$w(i,k,l) \quad > \quad T \qquad \tilde{w}(i,k,l) = w(i,k,l) - T; \tag{4.5}$$

$$w(i,k,l) \quad < \quad -T \qquad \tilde{w}(i,k,l) = w(i,k,l) + T; \tag{4.6}$$

$$|w(i,k,l)| \quad < \quad T \qquad \tilde{w}(i,k,l) = 0. \tag{4.7}$$

Modified soft thresholding (MST). In previous rules the coefficients are modified even if they are largely significant. In a modified softening we proposed rules with two thresholds to keep them (Bijaoui *et al.* 1997).

$$|w(i,k,l)| \quad \geq \quad T_2 \qquad \tilde{w}(i,k,l) = w(i,k,l); \tag{4.8}$$

$$|w(i,k,l)| \quad \leq \quad T_1 \qquad \tilde{w}(i,k,l) = 0; \tag{4.9}$$

$$T_1 < |w(i,k,l)| \quad < \quad T_2 \qquad \tilde{w}(i,k,l) = w(i,k,l)\frac{|w(i,k,l)| - T_1}{T_2 - T_1}. \tag{4.10}$$

The thresholds. The denoising depends on the chosen κ parameter. In the presented experiments, we set different values, often the same for the whole scales. In the two thresholds case, we set $\kappa_1 = 3.5$ and $\kappa_2 = 4.5$. The corresponding false alarm rates are respectively $4.7 10^{-4}$ and $6.7.10^{-6}$, taking into account the two signs.

Donoho & Johnstone (1994) introduced a thresholding rule (DST) based on the minimum risk leading to a threshold depending on the number K of independent

wavelet coefficients (which is the case for an orthogonal DWT):

$$\kappa = \sqrt{2\log_2(K)}. \tag{4.11}$$

As K decreases with the scale by a factor 4 in 2D, the threshold decreases with it. A similar rule was also applied to the à trous algorithm with MST (SMST).

The experiments. In Figure 4 The Haar transform of the Mel1 and the b1 images are plotted. That shows the effect of the noise with the scales. At right of the figure, the denoised b1 image seems quite good. A faint block effect can be identified on this image. On Figure 5 the denoised images obtained for b2 (left) and b3 (right) are displayed. The best images were selected on the different thresholding methods. The block effect largely increases with the noise. This is due to the fact that the noise increasing more and more coefficients become insignificant. The images are thus reconstructed by less and less coefficients, displaying the staircases associated to the Haar scaling function.
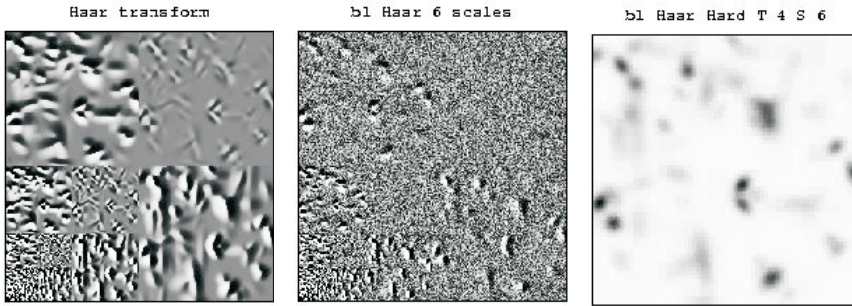

Fig. 4. The Haar transform of the Mel1 and b1 images. At right, the denoised image with the Haar transform and a hard thresholding at 3σ.

In Figure 6 the best denoised images obtained for b1 (left), b2 (middle) and b3 (right) with the Dauchechies 8 transform are displayed. There is no block effect, but a ringing appears around the bright objets. Due the reduction of the number of coefficients, these objects are characterized by peaks in the wavelet transform. Their reconstruction corresponds to the wavy wavelet pattern.

In Figure 7 the wavelet transform obtained with the *à trous* algorithm on b1 is displayed on 6 scales. The noise is clearly identified at the first scale. In Figure 8 the best denoised images obtained for b1 (left), b2 (middle) and b3 (right) with the *à trous* algorithm are displayed. There is no block effect, neither ringing. Some faint holes appeared around bright objects.

In Table 1 the SNRs obtained from different experiments are given. We can note that the denoisings obtained from the Haar transform are generally the worst ones. The effect of the thresholdings depends on the transform, the threshold and the initial SNR. The application of the redundant *à trous* algorithm seems to improve the denoising; but for a low SNR the Dauchechies 8 transform carries out the best result. Thus, the analysis shows that the choice of the best method
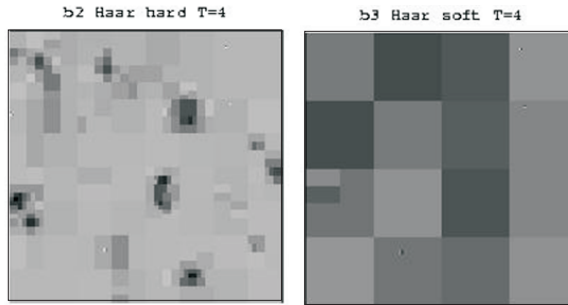
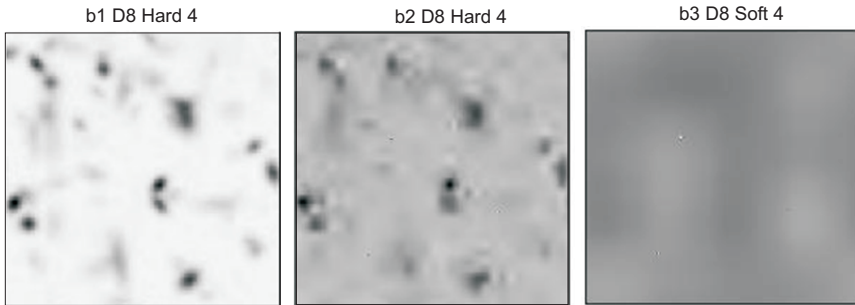**Fig. 5.** The best denoised images for b2 and b3 with the Haar transform.

Fig. 6. The best denoised images with the Daubechies 8 transform.

Table 1. SNR obtained on the three different images with different algorithms.

Method	b1	b2	b3
Haar HT $\kappa = 3$	19.49	6.06	−11.18
Haar HT $\kappa = 4$	18.77	7.82	−0.60
Haar ST $\kappa = 3$	17.25	7.06	0.37
Haar ST $\kappa = 4$	15.89	5.94	1.09
Daubechies 8 HT $\kappa = 3$	24.26	7.12	−10.78
Daubechies 8 HT $\kappa = 4$	24.46	9.33	−0.19
Daubechies 8 ST $\kappa = 3$	21.28	8.03	0.52
Daubechies 8 ST $\kappa = 4$	19.72	6.69	1.23
AT MST	26.46	12.25	−0.53
AT SMST	28.22	13.83	−2.96

depends on the input SNR. The adaptation needs to implement a thresholding algorithm taking into account the prior signal distribution, *i.e.* an algorithm based on a Bayesian statistics.
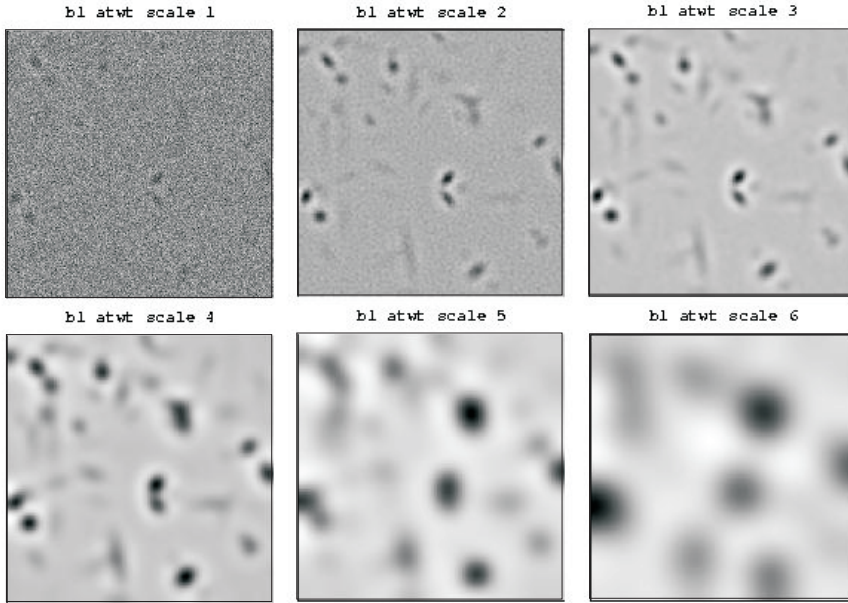
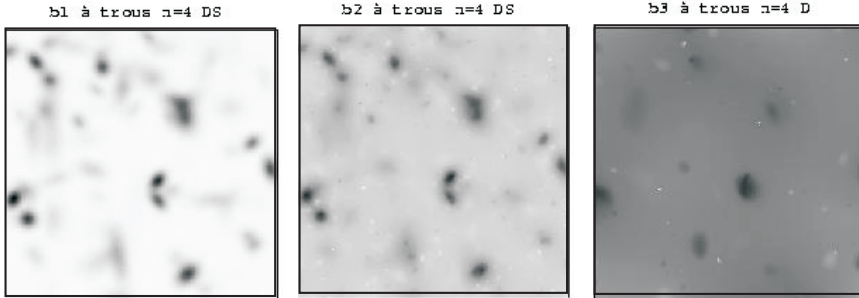**Fig. 7.** The *à trous* wavelet transform of the b1 image.

Fig. 8. The best denoised images with the *à trous* algorithm.

4.3   The regularization and the thresholding

Relation between close coefficients. In the previous methods the denoising was based on a local approach, the pixels were thresholded independently of each other. The environment played a role only in the computation of the wavelet coefficient. There are different ways to exploit the correlation between close wavelet coefficients. Here, a simple method based on the regularization is presented.

Values of non significant coefficients. In the HT case the wavelet coefficients are separated in two classes according to their values compared to the threshold. The raw reconstruction consists into the application of the inverse transform with

setting 0 for the non significant coefficients. Some artifacts, like block effects for the Haar transform, appeared in the restoration. In reality, the true value of a non significant coefficient is not null but faint.

Instead of inversing with null values we search to get a function $f(k,l)$ which minimizes a given objective function $C(f)$. The problem can be written as: *determine $f(k,l)$ such that $C(f)$ is minimum and $w_f(i,k,l) = w(i,k,l)$ for each significant coefficient.*

The application of the Tikhonov objective criterion. We set (Bobichon & Bijaoui 1997):

$$C(f) \equiv |D_k(f)|^2 + |D_l(f)|^2; \qquad (4.12)$$

where $D_k(f)$ and $D_l(f)$ are respectively the derivatives on the k and l directions. The $C(f)$ minimization is equivalent to:

$$L(f) = 0 \qquad (4.13)$$

where L is the image Laplacian. The Van-Cittert algorithm (see Sect. 7.2) leads to an iterated solution:

$$f^{(n+1)} = f^{(n)} + \alpha[0 - L(f^{(}n))] = f^{(n)} - \alpha L(f^{(n)}). \qquad (4.14)$$

where α is an adapted factor. For significant wavelet coefficients, we set:

$$w_{f^{(n+1)}}(i,k,l) = w(i,k,l). \qquad (4.15)$$

This operation allows the reduction of the block effects for the Haar transform. The restoration is also improved for the other DWT. Few iterations are generally needed.

Case of a softening function. Here, the wavelet coefficients are softened by a relation:

$$\tilde{w}(i,k,l) = \varpi w(i,k,l) \quad \text{with} \quad \varpi = S(w(i,k,l)). \qquad (4.16)$$

Here $S(w)$ is called the softening function. It takes values in the interval $[0,1]$. The application of the regularization can be done by considering ϖ as a weight. So after applying 4.14, we set (Jammal & Bijaoui 2004):

$$\tilde{w}_{f^{n+1}}(i,k,l) = \varpi w(i,k,l) + (1 - \varpi))w_{f^{n+1}}(i,k,l). \qquad (4.17)$$

For the highly significant coefficients, $\varpi \simeq 1$, no modification is done. While, for non significant ones $\varpi \simeq 0$, the algorithm furnishes the values given from the regularization.

The experiments. In Figure 9 the best denoised images obtained for b1 (left), b2 (middle) and b3 (right) with the *à trous* algorithm with regularization are displayed. There is no block effect, neither ringing. Some faint holes appeared around bright objects.

In Table 2 the results obtained with the application of the regularization, with the *à trous* algorithm are given. 10 iterations were applied. The regularization
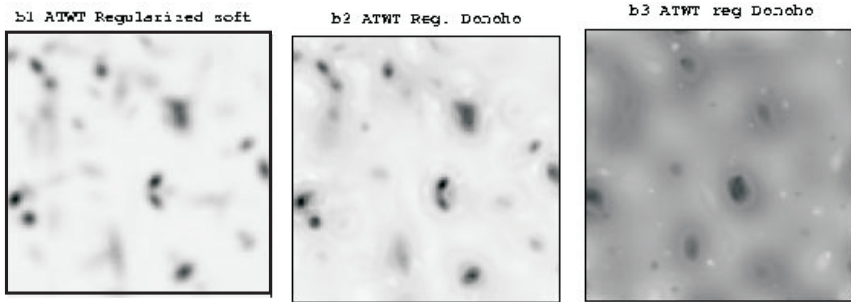
Fig. 9. The best denoised images with the *à trous* algorithm and Tikhonov regularization
of the image.

Table 2. The SNRs obtained after regularization with the Tikhonov constraint.

Method	b1	b2	b3
AT-HT	28.29	11.67	−5.48
AT-ST	27.27	10.85	−7.70
AT-DST	24.67	12.55	1.2

allowed a gain in the SNR for the hard thresholding at high SNR level. The gain
is less clear at low SNR.

The method was applied to a deconvolution in order to examine the possibility
to restore images compressed with the Haar transform (Bobichon & Bijaoui 1997;
Jammal & Bijaoui 2004; Dollet *et al.* 2004).

5   Image denoising from the maximum *a posteriori*

5.1   *The Bayesian estimations*

The posed problem. Let us consider a statistical variable x. It is observed with a
noise having a dispersion law $q(y|x)$. The denoising needs to answer the question
of what we can say about x knowing y. In the previous sections, only significant
coefficients were kept, possibly after a softening. The signal distribution was not
taken into account. It was noted that the method carrying out best results depends
on the signal-to-noise ratio. Thus, it is necessary to take into account the prior
distribution $p_x(x)$. That leads to apply the Bayes rule to get the conditional
posterior PDF:

$$p_{x|y} = \frac{p_x(x)q(y|x)}{p_y(y)}; \qquad \text{with} \qquad p_y(y) = \int_{-\infty}^{+\infty} p_x(x)q(y|x)dx. \qquad (5.1)$$

From the observation y, knowing the dispersion and the x prior PDF, we get
the x posterior PDF. As it is irrelevant to furnish this law for each observed value,

a simple estimate is derived. The maximum of the posterior PDF (MAP) is the most often furnished estimate.

The MAP and the regularization theory. The MAP estimate can be written as:

$$\hat{x} = \text{Arg} \quad \min_x[-\log(q(y|x)) - \log(p_x(x))]. \tag{5.2}$$

This expression is similar to the one introduced in the regularization theory. The first term corresponds to the data attachment $J_2(y, x)$; the second one to the objective function $J_1(x) = -\log(p_x(x))$. In the general case of the regularization theory, the objective function can be formalized independently of a prior distribution.

The case of a Gaussian white noise. From Equation (5.2) we get directly:

$$\hat{x} = y + \sigma^2 \frac{\partial \log(p_x(\hat{x}))}{\partial x}; \tag{5.3}$$

where σ is the standard deviation of the Gaussian noise distribution. Thus, $\hat{x}$ is obtained by solving an equation which may have many roots. That depends on the prior signal law. This law could be determined from the observed y PDF, by solving a deconvolution equation. This operation is delicate due to the histogram fluctuations.

The generalized Gaussian function and the L_q regularization. Experiments on natural images have been carried to model the out PDFs of their wavelet coefficients with extended tails. The prior PDFs were fitted with generalized Gaussian (Moulin & Liu 1998):

$$p_x(x) = ae^{-\frac{|x|^q}{b}}. \tag{5.4}$$

Coupled to Equation (5.2) that leads to the relation:

$$\hat{x} = \text{Arg} \quad \min_x \left[\frac{(y - x)^2}{2\sigma^2} + \lambda|x|^q \right]. \tag{5.5}$$

In this relation $\lambda = 1/b$ In the Gaussian case ($q = 2$), with a signal with a variance s, we get the Wiener filter (Wiener 1949):

$$\hat{x} = \frac{s^2}{s^2 + \sigma^2} y. \tag{5.6}$$

In the case of a Laplacian distribution ($q = 1$) of parameter b, the solution is the soft thresholding with the threshold $\frac{\sigma^2}{b}$. For a lower exponent, the filter tends to become a hard thresholding.

5.2   The basis pursuit

Principles. Chen *et al.* (1998) posed the restoration problem as:

$$\mathbf{Y} = \mathbf{A\Psi Z} + \mathbf{N} \tag{5.7}$$

where $\mathbf{\Psi}$ is a set (dictionary) of atoms and $\mathbf{Z}$ the related coefficients allowing the restoration of the image $\mathbf{X} = \mathbf{\Psi}\mathbf{Z}$. In this section, only the case $\mathbf{A} = \mathbf{I}$ is examined. An image atom is a given function, generally, but not necessarily, with a null mean. It can be obtained by translation and dilation of a generative function, like for the wavelet transform. A dictionary may be the union of primary dictionaries. A dictionary can be redundant, in which case the Gram matrix computed with these atoms is singular. In this case, a selection has to be done in order to restore the image with few atoms. There are many atom combinations leading to restore the same signal. The basis pursuit consists into the application of the MAP principle. That leads to search the atoms such that:

$$|\mathbf{Y} - \mathbf{\Psi}\mathbf{Z}|^2 + \lambda|\mathbf{Z}|_q \qquad (5.8)$$

is minimum. $q = 0$ corresponds to minimize the number of atoms (ℓ_0). Chen *et al.* (1998) proposed $q = 1$ which furnishes also a sparse representation. λ is the Lagrangian parameter. Its value results from the respect of the data attachment constraint.

The matching pursuit algorithm. In the ℓ_0 case, Equation (5.8) can be approximatively solved through a matching pursuit algorithm (Mallat & Zhang 1993). It is a greedy algorithm which progressively identifies the different atoms. Whatever the algorithm used to solved 5.8, generally it can not be proved that the minimum number of components is reached. Nevertheless, the matching pursuit algorithms, and specifically the orthogonal matching pursuit (OMP) (Pati *et al.* 1993) algorithms are very popular (Tropp 2004).

The ℓ_1 case. Let us consider the case of an orthogonal wavelet dictionary. Equation (5.8) leads to:

$$\mathbf{\Psi}^T Y - (\mathbf{\Psi}^T\mathbf{\Psi})^{-1}\mathbf{Z} - \lambda sign(\mathbf{Z}) = 0. \qquad (5.9)$$

$\mathbf{\Psi}_T\mathbf{Y}$ is the image wavelet transform $\mathbf{W}$. $\mathbf{\Psi}^T\mathbf{\Psi}$ is the identity matrix. Thus we write:

$$\mathbf{Z} = \mathbf{W} - \lambda sign(\mathbf{Z}). \qquad (5.10)$$

The algorithm corresponds to a soft thresholding; λ being chosen to satisfy the data attachment condition.

In the case of redundant dictionaries, the basis pursuit algorithm allows the minimization with a sparse representation. Different optimization algorithms were proposed. In particular the Block-Coordinate Relaxation (BCR) method (Bruce *et al.* 1998) leads to fast computations (Starck *et al.* 2004).

The Morphological Component Analysis (MCA). (Starck *et al.* 2004) This method was developed from the use of different transforms, such as the wavelet, the ridgelet (Candès 1998), the curvelet (Candès & Donoho 1999) and DCT (Ahmed *et al.* 1974) ones. BCR (Bruce *et al.* 1998) was used for obtaining the optimal representation. MCA was developed both with the ℓ_0 and the ℓ_1 norms.

The matching pursuit with the à trous algorithm. Instead of using dictionaries of orthonormal bases, it is possible to develop a sparse representation from the à

trous or the pyramidal algorithms (Bijaoui 2008). The corresponding atoms are progressively identified taking into account a threshold which decreases at each iteration. It is not guaranteed that the minimum number of atoms is reached with this greedy algorithm.

In Figure 10 the denoised images obtained for b1 (left), b2 (middle) and b3 (right) with this algorithm are displayed. The atoms are identified with the *à trous* wavelet transform using a matching pursuit algorithm A hard thresholding is performed with a threshold equal to 4. Thanks to the representation, the images appear very clean, without noise. Nevertheless the SNRs are not the best ones, 26.17 for b1, 12.07 for b2 and -1.09 for b3. The number of pyrels are respectively 702, 149 and 25. Compared to the number of pixels (65736) that corresponds to a high compression factor (94, 441 and 2629).
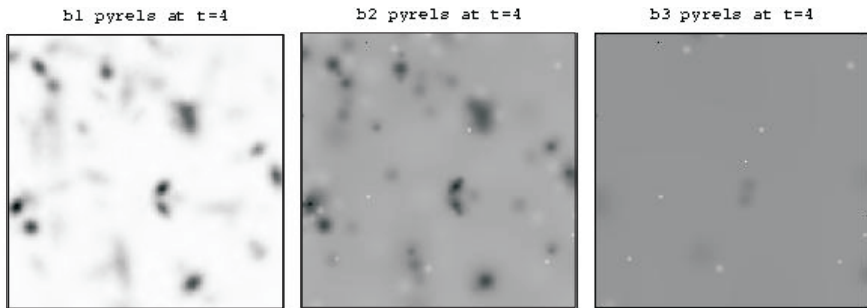

Fig. 10. The images resulting from the matching pursuit algorithm.

6 Image denoising from the posterior mean

6.1 The posterior mean

The minimum mean square estimator. MAP takes into account only the posterior distribution around its maximum which does not characterize the whole distribution. The expectation minimizes the mean square error (MMSE). Its value is:

$$\hat{x} = \int_{-\infty}^{+\infty} x \frac{p_x(x)q(y|x)dx}{\int_{-\infty}^{+\infty} p_x(x)q(y|x)dx}. \tag{6.1}$$

The MMSE evaluation. From Equation (6.1) we note that the evaluation of $\hat{x}$ needs to know the dispersion law $q(y|x)$ and the prior one $p_x(x)$. In this paper, it is admitted that the noise is white and Gaussian. As for the MAP estimation we have to evaluate the prior distribution.

Since Robbins' seminal work (Robbins 1956), it is known that the MMSE can be determined directly from the observed posterior distribution $p_y(y)$ for different dispersion PDFs. This is the case for the Gaussian one.

The Miyasawa relation. Taking into account that:

$$q(y|x) = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(y-x)^2}{2\sigma^2}};\tag{6.2}$$

we get:

$$\hat{x} = y + \frac{1}{\sqrt{2\pi}\sigma}\frac{\int_{-\infty}^{+\infty}(x-y)p_x(x)e^{-\frac{(y-x)^2}{2\sigma^2}}dx}{p_y(y)}.\tag{6.3}$$

As:

$$\frac{\partial p_y(y)}{\partial y} = \frac{1}{\sigma^2}\int_{-\infty}^{+\infty}(x-y)p_x(x)q(y|x)dx;\tag{6.4}$$

we get the Miyasawa relation (Miyasawa 1961):

$$\hat{x} = y + \sigma^2\frac{\partial \log p_y(y)}{\partial y}.\tag{6.5}$$

Note that the estimate depends only on the distribution of the observed variable y. We can also note the similarity between the Equations (5.3) and (6.5). But in the first case (MAP) the estimate is the solution of an equation which requires the knowledge of the prior PDF, while for the MMSE the estimate is directly furnished by a relation which takes into account the PDF of the observed variable.

6.2 Application of the Miyasawa relation

Application to the denoising with the DWT. Here, it is set that the variable y is the wavelet coefficient at a given scale. Thus, each wavelet plane is analyzed separately. It exists some correlation between the coefficients at the different scales, even for an orthogonal DWT. Thus, even if the MMSE is obtained at each scale, that does not guarantee that it is globally reached.

The estimation of the coefficients distribution. Relation 6.5 appears at the first glance very easy to exploit. However, the estimate depends on the derivative of the logarithm of the PDF, which is hard to correctly estimate. It is posed that the image is the realization of a stationary process. Thus, the coefficient histogram is the empirical $p(y)$ statistics. Its noise results from a Bernoulli distribution. That leads to very bad estimations for the distribution tails. Different approaches were proposed to improve the estimation (Bijaoui 2006, 2009):

- The Parzen method based on a sum of shifted windows (Parzen 1962). This method gives bad estimation for the tails. The window size has to be adapted to the event frequency.

- A denoising based on the wavelet transform (Bijaoui 2006). The Bernoulli noise increases the difficulty.

- A PDF model based on truncated distributions. Raphan & Simoncelli (2007) proposed exponentials. The use of Gaussians leads to easier computations (Bijaoui 2009).

- In the case of astronomical images, the PDF of the observed wavelet coefficients may be fitted by a Voigt function (García 2006), a convolution of a Gaussian with a Lorentzian function (Laplace PDF).

- The PDF can be approximated by a sum of Gaussians (Bijaoui 2002). The EM algorithm can be applied to determine the parameters (Bijaoui 2011). A simple mixture model furnishes a nice approximation to compute the softening filter (Bijaoui 2002).

A simple model for a MMSE estimation. First experiments showed on astronomical images that two Gaussian functions were extracted from the histogram of the wavelet transform at small scales: i/the Gaussian corresponding to the noise, ii/ a second one larger which corresponds to the sum of a signal with the noise. Thus, the PDF can be written as:

$$p(y) = (1 - a)G(y, N) + aG(y, S + N); \qquad (6.6)$$

where $G(y, V)$ is a centered Gaussian with a variance V.

The noise variance N is supposed to be known. Only a and S have to be determined at each scale. These parameters can be estimated from the variance M_2 and the 4th-order moment M_4:

$$
\begin{aligned}
M_2 &= (1 - a)N + a(S + N); & (6.7) \\
M_4 &= 3(1 - a)N^2 + 3a(S + N)^2. & (6.8)
\end{aligned}
$$

Thus:

$$
S = \frac{\frac{M_4}{3} - N^2}{M_2 - N}; \qquad (6.9)
$$

$$
a = \frac{(M_2 - N)^2}{\frac{M_4}{3} - N^2}. \qquad (6.10)
$$

If $M_2 - N < 0$ or if $M_4 < 3N^2$, S and a are set to 0. If $a > 1$, a is set to 1 and S is estimated only from the variance. We used these simple rules for estimating the model parameters at each wavelet scales.

Experimentations. In Figure 11 the denoised images obtained for b1 (left), b2 (middle) and b3 (right) with this algorithm (FONDW) are displayed. The SNRs are respectively 28.38, 13.15 and 1.84. The softening function is scale/scale determined by an algorithm free of tuning parameter. It is automatically adapted to the SNR.

At this experimentation level, the redundant à trous algorithm with MMSE carried out the best denoisings.
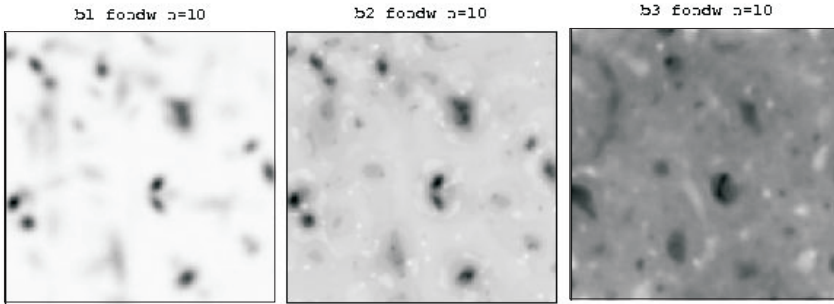
Fig. 11. The images resulting from the algorithm based on a scale/scale posterior expectation.

7   Application to astronomical image deconvolution

7.1   Image deconvolution using the wavelet transform

The direct inversion. The trivial image restoration would consist to denoise the image and to inverse the result (Starck & Bijaoui 1994). This method furnishes correct solutions for a regular blurring operator; but, it is not adapted to the case of a PSF with frequency holes.

The application of iterative inversions. Instead to inverse using the FFT, it is possible to apply an iterative scheme. Necessarily, the number of iterations is limited. This number plays a regularization role. For example, two classical iterative algorithms, the Van-Cittert and the Landweber ones are further presented. The information is not similarly restored for each frequency. If the modulation transfer function (MTF) is high the convergence is very fast; in contrast, the convergence is very low for the smallest MTF values. The number of iterations allows the obtention of a frequency filter depending of the MTF.

Many other iterative algorithms were proposed, such that the Richardson-Lucy one (Lucy 1974); which is very popular in the astronomical laboratories, due to its tendency to carry out images with point-like structures.

The wavelet-vaguelette decomposition. The multiresolution analysis carries out a linear representation with wavelet functions. The inversion is a linear operation so that the previous wavelet representation can be directly translated into a vaguelette decomposition; the vaguelettes being the wavelet functions deconvolved with the PSF (Donoho 1992). This is also convenient for a regular operator, but not adapted to a singular one.

The mirror wavelet. Among the different proposed methods, Kalifa *et al.* (2003) proposed first to deconvolve the image and then to denoise taking account a soft thresholding with mirror wavelet bases. This decomposition is a specific wave packet (Coifman & Wickerhauser 1992) for which the frequency band is decomposed like a mirror compared to the wavelet decomposition. This property allowed them to improve the signal localization both spatially and in frequency.

The multi resolution CLEAN. Wakker *et al.* (1988) proposed a restoration method for aperture synthesis images based on a two-stages CLEAN algorithm, with identification of Dirac peaks and extended Gaussians. Starck *et al.* (1994) developed this idea in the framework of the wavelet transform. The algorithm is similar to the matching pursuit one, with specific wavelet functions.

The application of the basis pursuit. In the previous method, the goal was to represent the image with few wavelet patterns. The basis pursuit algorithm allows the obtention of a solution, with an ℓ_1 minimization (Daubechies *et al.* 2004).

7.2 The Van Cittert and the Landweber iterative inversions

The basic relation. Let us consider the classical inverse problem without noise $\mathbf{Y} = \mathbf{AX}$. Van Cittert (1931) introduced a simple iterative inversion algorithm. The idea consists into writing $\mathbf{B} = \mathbf{I} - \mathbf{A}$; where $\mathbf{I}$ is the identity matrix. Thus, the solution is written as:

$$\mathbf{X} = [\mathbf{I} - \mathbf{B}]^{-1}\mathbf{Y}. \tag{7.1}$$

If all the $\mathbf{B}$ eigenvalues are in the open interval $]-1, +1[$, it can be derived that:

$$\mathbf{X} = [\mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \ldots]\mathbf{Y}. \tag{7.2}$$

The development is limited at the order n:

$$\mathbf{X}^{(n)} = [\mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \ldots + \mathbf{B}^{(n)}]\mathbf{Y}. \tag{7.3}$$

That leads to:

$$\mathbf{X}^{(n)} = \mathbf{Y} + \mathbf{B}[\mathbf{I} + \mathbf{B} + \mathbf{B}^2 + \ldots + \mathbf{B}^{(n-1)}]\mathbf{Y}; \tag{7.4}$$

Or:

$$\mathbf{X}^{(n)} = \mathbf{Y} + \mathbf{B}\mathbf{X}^{(n-1)} = \mathbf{Y} + [\mathbf{I} - \mathbf{A}]\mathbf{X}^{(n-1)}. \tag{7.5}$$

Thus, finally:

$$\mathbf{X}^{(n)} = \mathbf{X}^{(n-1)} + [\mathbf{Y} - \mathbf{AX}^{(n-1)}]. \tag{7.6}$$

The convergence factor. It is also possible to accelerate the convergence by the introduction of a convergence factor α_n:

$$\mathbf{X}^{(n)} = \mathbf{X}^{(n-1)} + \alpha_n[\mathbf{Y} - \mathbf{AX}^{(n-1)}]. \tag{7.7}$$

Its best value is reached for the minimum norm of $\mathbf{R}^{(n)} = \mathbf{Y} - \mathbf{AX}^{(n)}$:

$$\alpha_n = \frac{\mathbf{R}^{(n-1)}\mathbf{AR}^{(n-1)}}{|\mathbf{AR}^{(n-1)}|^2}. \tag{7.8}$$

The distance minimization. With the previous algorithm, the algorithm does not converge if $\mathbf{A}$ is singular, *i.e.* it exists null eigenvalues ($\lambda_p = 0$). In the direction of the corresponding eigenvectors the equation is written as $Y_p = \lambda_p X_p = 0$. If

$Y_p \neq 0$ the equation is not consistent. Thus, the problem is modified by writing $r = |\mathbf{Y} - \mathbf{AX}|^2$ is minimum. That leads to:

$$\mathbf{A}^T[\mathbf{Y} - \mathbf{AX}] = 0; \qquad (7.9)$$

which is written as:

$$\mathbf{A}^T\mathbf{Y} = \mathbf{A}^T\mathbf{AX}. \qquad (7.10)$$

For $\lambda_p = 0$, we get $0 = 0$. The system is now underdetermined but consistent. A unique solution can be obtained by regularization. A classical constraint consists into setting $X_p = 0$ for $\lambda_p = 0$ (minimum energy).

The Landweber algorithm. The Van-Cittert algorithm is applied to Equation (7.10). That leads to:

$$\mathbf{X}^{(n)} = \mathbf{X}^{(n-1)} + \alpha_n\mathbf{A}^T[\mathbf{Y} - \mathbf{AX}^{(n-1)}] = \mathbf{X}^{(n-1)} + \alpha_n\mathbf{A}^T\mathbf{R}^{(n-1)}. \qquad (7.11)$$

This relation is known as the Landweber algorithm, which can be directly developed from a gradient descent (Landweber 1951). α_n is computed as it is upper indicated for the Van Cittert The previous approach allows the comparison between the two algorithms for the convergence. In Relation 7.10 the data are smoothed by the joint matrix. For an eigendirection p, the left value is $\lambda_p Y_p$. If $\lambda_p = 0$, its value becomes also null in this direction. The eigenvalue of $\mathbf{A}^T\mathbf{A}$ is λ_p^2, always positive or null. If we consider a non null eigenvalue, the convergence is always assumed. For a null eigenvalue the algorithm keeps the initial value. If we set $\mathbf{X}^{(0)} = 0$, no new information is added. Thus, the algorithm furnishes the solution which minimizes the energy.

The convergence speed depends on the matrix conditioning, *i.e.* the ratio between the highest eigenvalue and its lowest (and different from 0) one of the matrix $\mathbf{A}^T\mathbf{A}$. Many methods were proposed to accelerate the convergence. The conjugate gradient is the most popular one (Hestenes & Stiefel 1952).

7.3   Deconvolution from the significant residual

The significant residuals. The Landweber algorithm consists into the addition of the residual $\mathbf{R}^{(n-1)}$ smoothed with the adjoint matrix $\mathbf{A}^T$ to the previous approximation. Even if the solution at step $(n-1)$ is not noisy we add a noisy residual. The smoothing by the joint matrix removes only a part of the noise. Thus it is needed to denoise the residual in order to avoid adding noise to the solution. It is considered that it is the same noise for the successive residuals than for the signal $\mathbf{Y}$. In (Murtagh *et al.* 1995) different iterative inversion algorithms were examined. The Van-Cittert and the Richardson-Lucy ones are also available for that purpose. I prefer to present the algorithm using the Landweber algorithm, more stable than the other ones. In the previous sections different denoising methods were examined. The best ones have to be applied. In (Murtagh *et al.* 1995) a hard thresholding was applied. Upper, it was clear that the Bayesian posterior mean leads to the best results for the different SNRs. It is thus chosen for the deconvolution algorithm.

The deconvolution algorihm. It is the following:

1. Set $\mathbf{X}^{(0)} = 0$ and $n = 1$.

2. $\mathbf{Z} = \mathbf{A}\mathbf{X}^{(n-1)}$.

3. $\mathbf{R} = \mathbf{Y} - \mathbf{Z}$.

4. $\mathbf{R}$ is denoised to $\tilde{\mathbf{R}}$.

5. $\mathbf{S} = \mathbf{A}^T \tilde{\mathbf{R}}$.

6. $\alpha_n = \mathbf{S}.\tilde{\mathbf{R}}/|\mathbf{S}|^2$.

7. $\mathbf{X}^{(n)} = \mathbf{X}^{(n-1)} + \alpha_n \mathbf{S}$ and $n = n + 1$.

8. According to the chosen convergence criterion (number of iterations, residual energy ...) the algorithm stops or comes back to step 2.

The positivity constraint. Generally the astrophysical sources are positive functions. A background is preliminary subtracted in order to get a positive image function. The positivity constraint can be easily satisfied by a thresholding to positive values at each iteration. The application of this constraint to a deconvolution with a PSF having frequency holes may lead to a significant gain in resolution.

The deconvolution experiments. The simulated image was smoothed with a Gaussian PSF having the size 1 (Msg1), 2 (Msg2) and 4 (Msg4). A white Gaussian noise was added with the deviations 0.007, 0.07 and 0.7. The applied deconvolution program was based on the FONDW denoising. The residuals are scale by scale examined. The wavelet coefficients are softened with a filter derived from the Miyasawa relation. Their histograms are fitted with the Gaussian mixture associated to the simple model. After deconvolution, the SNR for the deconvolved image compared to the initial one (Mel1) is computed. The SNR for the denoised image compared to the blurred image, free of noise, is also determined.

The results. In Figure 12 the denoised and restored images obtained with this algorithm (dgondw) are displayed for the Mel1 image blurred with a Gaussian PSF with $\sigma = 1$. The lines correspond respectively to the b1, b2 and b3 images. At left the smoothed noisy images, at middle the smoothed denoised ones and at right the Mel1 restoration. In Table 3 the resulting SNRs are given. It can be noted that the deconvolved images for b1 and b2 lead to a better SNR than the images obtained without blurring. That is probably due to the a regularization effect introduced by the Landweber iterations. In Table 3 the SNRs for the restored images with a blurring at $\sigma = 2$ and $\sigma = 4$ are also indicated. The positive constraint brings a significant gain, especially for a low SNR. For a large PSF, the number of iterations inside the Landweber algorithm was also increased to improve the results for a high SNR. On Figure 13 the denoised and restored images are displayed for $\sigma = 4$.

The denoising from deconvolution taking into account an hypothetic PSF. In the previous paragraph the denoising derived from a deconvolution with a known PSF.
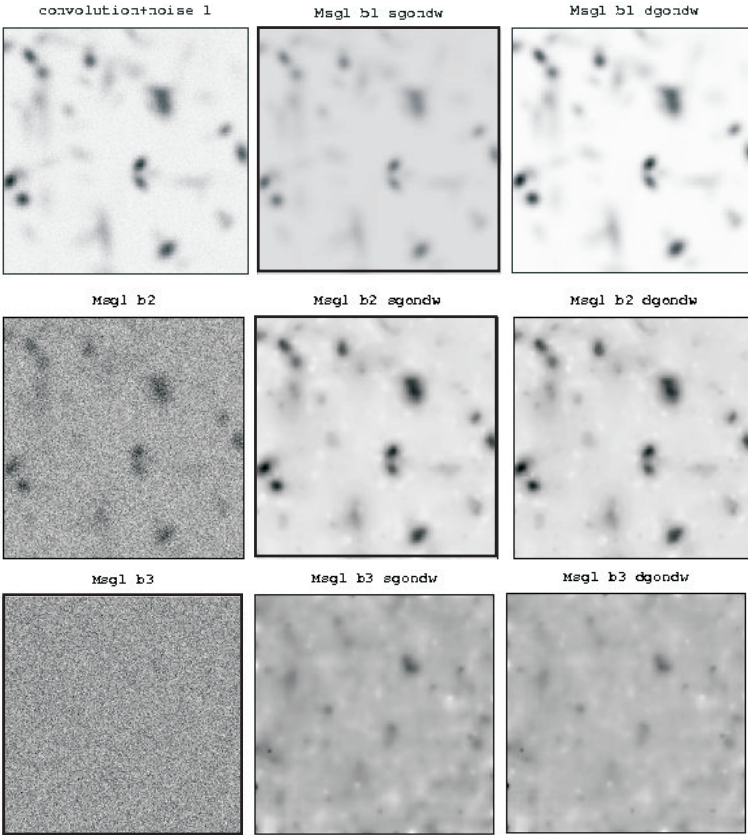
Fig. 12. The blurred noisy images at $\sigma = 1$ (*left*), its resulting denoised (*middle*) and restored images (*right*).

Table 3. SNR resulting from the deconvolution experiments. p: the positivity constraint is applied; x: 100 iteration steps were done instead to 10. (s) means smoothing and (d) deconvolution.

Image	(s)b1	(d)b1	(s)b2	(d)b2	(s)b3	(d)b3
Msg1	29.90	28.89	14.47	14.03	−1.12	−1.27
Msg2	30.02	27.25	13.87	12.78	−0.16	−0.42
Msg2 p	30.67	28.29	15.54		1.79	1.52
Msg4 p	30.56	14.07	15.32	11.48	0.82	0.08
Msg4 px	30.90	22.35	14.97	11.30	−0.35	−2.12

It can be noted that the SNRs were very fine. But the comparison was done to the blurred images without noise (Msg1, Msg2, Msg4) and not to the initial one (Mel1).
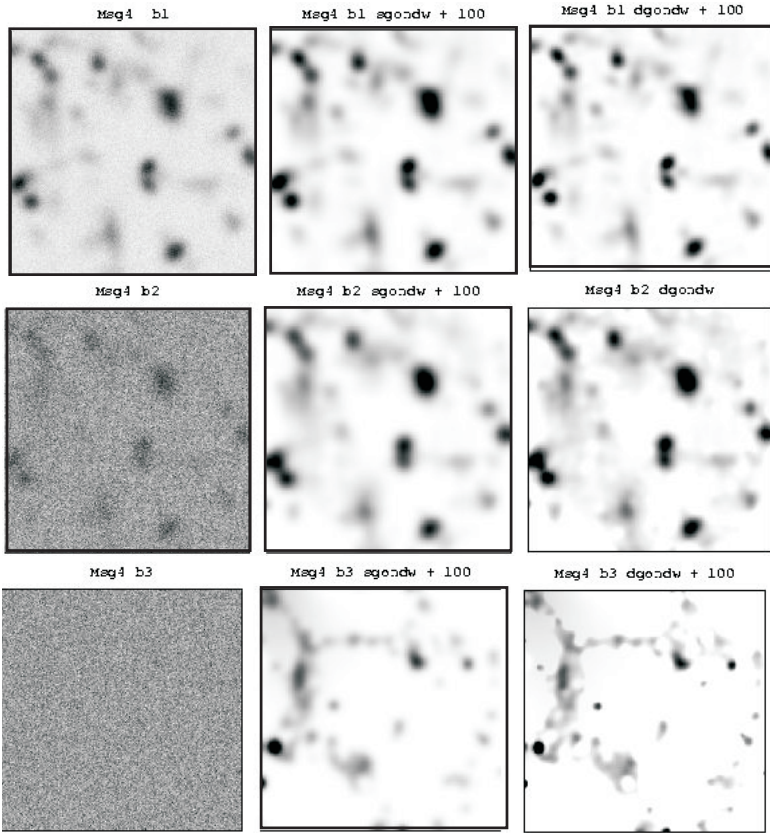
Fig. 13. The blurred noisy images at $\sigma = 4$ (*left*), its resulting denoised (*middle*) and restored images (*right*).

Table 4. The SNRs obtained after denoising taking into account an hypothetic PSF. The first number is the size of the hypothetic Gaussian PSF. x means that 100 iteration steps were done instead of 10.

PSF size	b1	b2	b3
1	29.99	15.55	1.65
2	30.11	15.24	1.72
4 x	30.11	15.80	2.69

It is posed that the observed image (for example b1) is the noisy blurred version of an image with a Gaussian PSF but of unknown width. Different widths are tested. The denoised image obtained after deconvolution is compared to Mel1. If the chosen PSF is close to a Dirac peak, no gain can be obtained compared to a direct denoising. For a too extended chosen PSF, information is lost on the details
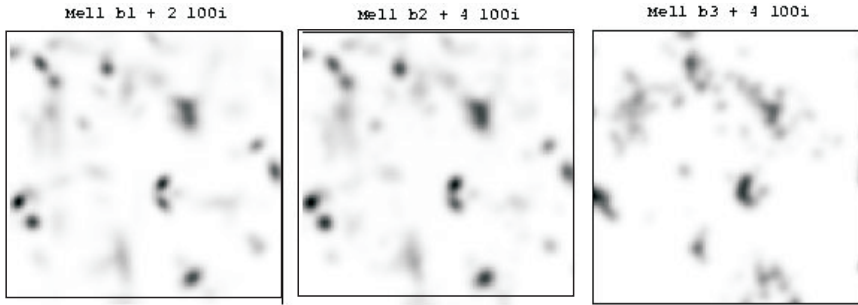
Fig. 14. The denoisings obtained from a deconvolution from an hypothetic PSF at $\sigma = 4$.

at small scales. In Table 4 the resulting SNRs are given for different PSF. In the whole cases the positivity constraint was applied. On Figure 14 the denoised Mel1 images are displayed for $\sigma = 4$. They correspond to the best denoised images.

It can be noted that the best SNRs were obtained with the present method.

8   Conclusion

The study limits. In the present paper, an introduction to the image restoration using multiscale methods was given. A large number of papers were published since two decades on this topic. The paper was centered on the use of a peculiar algorithm, the *à trous* one, for the restoration of astrophysical images corrupted by a white Gaussian noise. In the literature a large panel of transforms and different noises were examined.

The Poisson case. Many modern astronomical detectors furnish images for which the noise statistics is dominated by the photon noise. That led to introduce specific methods to restore these images with a Poisson noise (Murtagh *et al.* 1995). Different methods were proposed to quantify the significance of a wavelet coefficient. A simple way consists into the transformation of the initial pixel values such its variance becomes constant. The Anscombe transform (1948) was first proposed. The results are fine for a mean number of photons greater than about 10. Below this value, other strategies were proposed (Fadda *et al.* 1998; Bijaoui & Jammal 2001; Jammal & Bijaoui 2004; Zhang *et al.* 2008).

Other cases. Different other noise processes were also considered. A generalization of the Anscombe transform was proposed to process images with a mixing of a Gaussian and a Poisson noise (Murtagh *et al.* 1995). More generally the noise properties can be provided by a table. That leads to laborious computations to estimate for each wavelet coefficient its distribution.

Oriented wavelets. Orthogonal DWT carries out simple directional information. Discrete CWT allows the application of oriented wavelets using the FFT (Antoine *et al.* 1996). The steerable pyramids were introduced in order to overcome the limitations of orthogonal separable wavelet decompositions that do not represent

oblique orientations well (Simoncelli *et al.* 1992). This transform is shift invariant and also rotation invariant. Kingsbury (2002) generalized the use of complex DWT for the image processing. This transform allows an approximate shift invariance and some oriented information. More recently an oriented transform based on wavelets and the lifting scheme was introduced by Chapellier & Guillemot (2006) for image compression. This work brings a general framework for oriented DWTs.

Other representations. This paper was oriented to the application of the discrete transforms derived from the continuous one. Other multiscale representations were mentioned such that the ridgelets, the curvelets or the wave packets. Many other multiscale geometric representations were developed with a directional and frequency selectivity (Jacques *et al.* 2011).

A quality criterion for a given image. Beyond the proposed representations, the main problem resides in their capability to best restore the images. But, what is "a best restored image"? Here, the SNR was the alone considered criterion. The search of sparse representations leads to images which a clean appearance, fully denoised. But that does not mean that the results are better than the ones obtained with other algorithms leading to a faint noisy appearance. The posterior expectation leads to a better estimation than the MAP in term of quadratic error, even if residual fluctuations can be identified.

In fact, the main question concerns the use of the restored images. The image restoration is only a step in the image analysis. The astrophysicists are interested by the detection and the characterization of cosmic sources. They apply on the restored images programs for the source identification. Thus, more accurate quality criteria may derive from the detection rate, the false-alarm rate and the quality of the resulting measurements. This feature is fully outside this present introduction. Nevertheless the readers can get useful information on these questions inside the reference papers.

References

Ahmed, N., Natarajan, T., & Rao, K.R., IEEE Comp., 32, 90

Anscombe, F.J., 1948, Biometrika, 15, 246

Antoine, J.-P., Vandergheynst, P., & Murenzi, R., 1996, Int. J. Im. Sys. Tech., 152

Bijaoui, A., 2002, Signal Proc., 82, 709

Bijaoui, A., 2008, Astronomical Data Analysis V, ed. F., Murtagh & J.L. Starck, `http://www.ics.forth.gr/ada5/pdf_files/Bijaoui_talk.pdf`, 2008

Bijaoui, A., 2009, GRETSI Conference, ed. D. Ginhac (Dijon University)

Bijaoui, A., 2011, GRETSI Conference, ed. Y. Berthoumieu & M. Najim (Bordeaux University)

Bijaoui, A., Bobichon, Y., Fang, Y., & Rué, F., 1997, Traitement du Signal, 14, 179

Bijaoui, A., & Jammal, G., 2001, Signal Proc., 81, 1789

Bijaoui, A., Nowak, R., & Martin, A., 2006, Astronomical Data Analysis IV, ed. C. Surace, `http://www.oamp.fr/conf/ada4/pub/19/DSD/ADAIV_Bijaoui.pdf`

Bijaoui, A., Starck, J.L., & Murtagh, F., 1994, Traitement du Signal, 11, 229

Bobichon, Y., & Bijaoui, A., 1997, Exp. Astr., 7, 239

Bruce, A., Sardy, S., & Tseng, P., 1998, Proc. SPIE, 3391, 75

Candès, E., 1998, Ridgelets: theory and applications, Ph.D. Thesis, Statistics (Standford University)

Candès, E.J., & Donoho, D.L., 1999, in Curves and Surfaces, ed. Schumaker *et al.* (Vanderbilt University Press, Nashville, TN)

Chappellier V., & Guillemot, C., 2006, IEEE IP, 15, 2892

Chen, S.S., Donoho, D.L., & Saunders, M.A., 1988, SIAM J., 20, 33

Coifman, R.R., & Donoho, D.L., 1995, Technical report 475, Dpt. of Statistics (Stanford University)

Coifman, R.R., & Wickerhauser, M.V., 1992, IEEE IT, 38, 876

Daubechies, I., 1988, Com. Pure Appl. Math., 41, 909

Daubechies, I., Grossmann, A., & Meyer, Y., 1986, J. Math. Phys., 27, 1271

Daubechies, I., Defrise, M., & De Mol, C., 2004, Com. Pure Appl. Math., 57, 1413

Dollet, C., Bijaoui, A., & Mignard, F., 2004, A&A, 2004, 426

Donoho, D.L., 1992, Appl. Comput. Harmonic Anal., 2, 101

Donoho, D.L., 1995, IEEE IT, 41, 613

Donoho, D.L., & Johnstone, J.M., 1994, Biometrika, 81, 425

Fadda, D., Slezak, E., & Bijaoui, A., 1998, A&ASS, 127, 335

García, T.T., 2006, MNRAS, 369, 2025

Grossmann, A., & Morlet, J., 1984, SIAM J., 15, 723

Hestenes, M.R., & Stiefel, E., 1952, J. Res. National Bureau Stand., 49, 409

Högbom, J., 1974, A&ASS, 15, 417

Holschneider, M., Kronland-Martinet, R., Morlet, J., & Tchamichian, P., 1989, in Wavelets Combes, ed. J.M. *et al.* (Springer-Verlag), 286

Hou, H., & Andrews, H., 1978, IEEE ASSP, 26, 508

Jammal, G., & Bijaoui, A., 2004, Signal Proc., 84, 1049

Esteban, D., & Galland, C., 1977, in Proc. ICASSP, 191

Kalifa, J., Mallat, S., & Rougé, B., 2003, IEEE IT, 12, 446

Kingsbury, N.G., 2001, J. Appl. Comp. Harmonic Anal., 10, 234

Landweber, L., 1951, Am. J. Math., 73, 615

Jacques, L., Duval, L., Chaux, C., & Peyré, G., 2011, Signal Proc., 91, 2699

Lucy, L.B., 1974, AJ, 79, 745

Mallat, S., 1989, IEEE PAMI, 11, 674

Mallat, S., 1998, A wavelet tour of signal processing (Academic Press, San Diego)

Mallat, S., & Zhang, Z., 1993, IEEE SP, 41, 3397

Miyasawa, K., 1961, Bull. Inst. Internat. Statist., 38, 181

Morlet, J., Arens, G., Fourgeau, E., & Giard, D., 1982, Geophysics, 47, 203

Moulin, P., & Liu, J., 1998, IEEE, IT, 45, 909

Murtagh F., Starck J.L., & Bijaoui A., 1995, A&ASS, 112, 179

Parzen, E., 1962, Ann. Math. Stat., 33, 1065

Pati, Y., Rezaiifar, R., & Krishnaprasad, P., 1993, in Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers

Raphan, M., & Simoncelli, E.P., 2007, Comp. Sc. Tech. Report TR2007-900, Courant Institute of Math. Sci. (New York University)

Raphan, M., & Simoncelli, E.P., 2008, IEEE IP, 17, 1342

Robbins, H., 1956, Proc. Third Berkley Symp. on Math. Stat., 1, 157

Shannon, C.E., 1948, Bell Syst. Tech. J., 27, 379

Schoenberg, I.J., 1946, Quart. Appl. Math., 4, 45, 112

Simoncelli, E.P., Freeman, W.T., Adelson, E.H., & Heeger, D.J., 1992, IEEE IT, Special Issue on Wavelets, 38, 587

Starck, J.L., & Bijaoui A., 1994, Signal Proc., 35, 195

Starck, J.L., Bijaoui, A., Lopez, B., & Perrier, C., 1994, A&A, 283, 349

Starck, J.L., Murtagh, F., & Bijaoui, A., 1998, Image Processing and Data Analysis in the Physical Sciences. The Multiscale Approach (Cambridge University Press)

Starck, J.L., Elad, M., & Donoho, D.L., 2004, Adv. Imaging Electr. Phys., 132

Strang, G., 1989, SIAM Review, 31, 614

Strang, G., & Fix, G., 1971, in Constructive Aspects of Functional Analysis, Edizioni Cremonese, Rome, 796

Tropp, J.A., 2004, IEEE IT, 50, 2231

Van Cittert, P.H., 1931, Z. Phys., 69, 298

Vitterli, M., 1986, Signal Proc., 10, 219

Wakker, B.P., & Schwarz, U.J., 1988, A&A, 200, 312

Wiener, N., 1949, Extrapolation, Interpolation, and Smoothing of Stationary Time Series (Wiley, New York)

Zhang, B., Fadili, J.M., & Starck J.L., 2008, IEEE IT, 17, 1093