

Filtering and deconvolution by the wavelet transform

Jean-Luc Starck

Cisi-Ingenierie, 1360 route des Dolines, 06560 Valbonne, France and Observatoire de la Côte d'Azur B.P. 229, F-06304 Nice Cedex 4, France

Albert Bijaoui

Observatoire de la Côte d'Azur, B.P. 229, F-06304 Nice Cedex 4, France

Received 9 March 1992

Revised 4 September 1992, 5 January 1993 and 16 April 1993

Abstract. The wavelet transform gives information in both spatial and frequency domains and is a very useful tool for describing the hierarchical structures. A new approach for filtering based on the wavelet transform is presented in this paper, and several algorithms are proposed. A criterion of quality, which takes into account the resolution, is used to compare these algorithms. We show that deconvolution can be done using filtered wavelet coefficients. By computing the wavelet from the point spread function, we find a new transform algorithm and a reconstruction method related to it.

Zusammenfassung. Die Wavelettransformation liefert Informationen sowohl über den Raum- als auch über den Frequenzbereich und ist ein sehr nützliches Werkzeug zur Beschreibung hierarchischer Strukturen. Ein neuer Ansatz zur Filterung auf der Grundlage der Wavelettransformation wird in diesem Beitrag vorgestellt, und es werden verschiedene Rechenverfahren vorgeschlagen. Ein Gütemaß, das die Auflösung berücksichtigt, wird zum Vergleich dieser Algorithmen verwendet. Wir zeigen, daß eine Entfaltung mittels gefilterter Waveletkoeffizienten durchgeführt werden kann. Indem wir die Wavelets aus der Impulsantwort berechnen, finden wir einen neuen Transformationsalgorithmus und eine damit zusammenhängende Rekonstruktionsmethode.

Résumé. La transformée en ondelettes donne de l'information à la fois spatialement et fréquentiellement. Elle est donc un outil extrêmement intéressant pour décrire les structures hiérarchiques. Une nouvelle approche du filtrage à partir des ondelettes est présentée dans cet article, et différents algorithmes sont proposés. Un critère de qualité, qui tient compte de la résolution, est utilisé pour comparer ces algorithmes. On montre qu'une déconvolution peut aussi être envisagée à partir des coefficients d'ondelettes filtrés. En calculant l'ondelette à partir de la réponse impulsionnelle, un nouvel algorithme de transformée en ondelettes est proposé, ainsi qu'une reconstruction au sens des moindres carrés.

Keywords. Multiresolution analysis; wavelet; image processing; image restoration; filtering; deconvolution.

1. Introduction

The sky contains many kinds of non-stellar components: planets and other objects belonging to the solar system, planetary nebulae, interstellar clouds, star clusters, galaxies, clusters of galaxies, etc. Each of these components shows irregular pat-

terns which are generally associated with a hierarchical structure. For example, molecular clouds, where clouds are thought to originate, have very complex structures often described as fractals. Many galaxies have sets of non-stellar components which seem to be graded in an ordered fashion. These hierarchical structures are connected with physical processes, in particular non-linear ones such as those which give rise to turbulent flows.

Correspondence to: Mr. J.-L. Starck, Observatoire de la Côte d'Azur, B.P. 229, F-06304 Nice Cedex 4, France.

Images of astronomical objects, even when obtained by large telescopes, are generally noisy and blurred. Many smoothing and restoration techniques are currently used to enhance structural features of images. The algorithms used are generally based on stationary statistical processes, but astronomical images cannot be considered an achievement of such a process. Consequently, one cannot get an optimal restoration using the classical Wiener filtering, which is based on stationary Gaussian processes [21]. Some adaptive filterings [15] and many non-linear restoration methods [5] furnish better results.

The continuous wavelet transform [9] allows one to decompose a signal in a position-scale space separating the components of different sizes. An analysis of the significant features in this space leads to adaptive filterings. In this paper we present different methods which take into account the hierarchy, i.e. the link between successive scales. First, we make use of the wavelet transform in order to introduce the various algorithms. After summarizing how the transform can be used, we give as an example a Wiener-type filtering. Adaptive filterings are discussed and a comparison is done using a multiresolution quality criterion. Finally, we derive deconvolution methods adapted to hierarchical structures.

2. The wavelet transform

Extensive literature exists on the wavelet transform and its applications [3, 6, 17–19, 23]. We summarize here the main features which we make use of.

2.1. The continuous wavelet transform

The Morlet–Grossmann definition of the continuous wavelet transform [9] for a 1-D signal $f(x) \in L^2(\mathbb{R})$ is

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x) \psi^* \left(\frac{x-b}{a} \right) dx, \quad (1)$$

where z^* denotes the complex conjugate of z , $\psi^*(x)$

is the analyzing wavelet, $a (> 0)$ is the scale parameter and b is the position parameter. The transform is characterized by the following three properties:

1. it is a linear transformation;
2. it is covariant under translations:

$$f(x) \rightarrow f(x-u), \quad W(a, b) \rightarrow W(a, b-u); \quad (2)$$

3. it is covariant under dilations:

$$f(x) \rightarrow f(sx), \quad W(a, b) \rightarrow s^{-1/2} W(sa, sb). \quad (3)$$

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.

In Fourier space, we have

$$\hat{W}(a, v) = \sqrt{a} \hat{f}(v) \hat{\psi}^*(av). \quad (4)$$

When the scale a varies, the filter $\hat{\psi}^*(av)$ is only reduced or dilated while keeping the same pattern.

Consider now a function $W(a, b)$ which is the wavelet transform of a given function $f(x)$. It has been shown [9, 11] that $f(x)$ can be restored using the formula

$$f(x) = \frac{1}{C_x} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{a}} W(a, b) \chi \left(\frac{x-b}{a} \right) \frac{da db}{a^2}, \quad (5)$$

where

$$\begin{aligned} C_x &= \int_0^{+\infty} \frac{\hat{\psi}^*(v) \hat{\chi}(v)}{v} dv \\ &= \int_{-\infty}^0 \frac{\hat{\psi}^*(v) \hat{\chi}(v)}{v} dv. \end{aligned} \quad (6)$$

Generally $\chi(x) = \psi(x)$, but other choices can enhance certain features for some applications.

The reconstruction is only available if C_x is defined (admissibility condition). In the case of $\chi(x) = \psi(x)$, this condition implies $\hat{\psi}(0) = 0$, i.e. the mean of the wavelet function is 0.

2.2. The discrete wavelet transform

For processing classical images, the sampling is made in accordance with Shannon's [24]

well-known theorem. The discrete wavelet transform (DWT) can be derived from this theorem if we process a signal which has a cut-off frequency. For such images the frequency band is always limited by the size of the camera aperture.

A digital analysis is provided by the discretisation of formula (1), with some simple considerations on the modification of the wavelet pattern by dilation. Usually, the wavelet function $\psi^*(x)$ has no cut-off frequency and it is necessary to suppress the values outside the frequency band in order to avoid aliasing effects. We can work in Fourier space, computing the transform scale by scale. The number of elements for a scale can be reduced, if the frequency bandwidth is also reduced. This is possible only for a wavelet which also has a cut-off frequency. The decomposition proposed by Littlewood and Paley [13] provides a very good illustration of the reduction of elements scale by scale. This decomposition is based on an iterative dichotomy of the frequency band. The associated wavelet is well localized in Fourier space where it allows a reasonable analysis to be made though not in the original space. The search for a discrete transform which is well localized in both spaces leads to the multiresolution analysis.

The multiresolution analysis [14] results from the embedded subsets generated by the approximations of a given function $f(x)$ at different scales. $f(x)$ is projected at each step j on the subset V_j ($j \leq 0$). This projection is defined by the scalar product $c_j(k)$ of $f(x)$ with the scaling function $\phi(x)$ which is dilated and translated:

$$c_j(k) = \langle f(x), 2^{-j}\phi(2^{-j}x - k) \rangle. \tag{7}$$

If we start from the set $\{c_0(k)\}$, we compute all the sets $\{c_j(k)\}$, with $j < 0$, without directly computing any other scalar product

$$c_j(k) = \sum_n h(n - 2k)c_{j+1}(n). \tag{8}$$

At each step, the number of scalar products is divided by 2. Information is lost, and the signal is smoothed step by step. The remaining information

can be restored using the complementary subspace W_j of V_j in V_{j+1} . This subspace can be generated by a wavelet function $\psi(x)$ with translation and dilation. We have

$$\frac{1}{2}\psi\left(\frac{x}{2}\right) = \sum_{k=-\infty}^{+\infty} g(k)\phi(x - k). \tag{9}$$

We compute the scalar products $\langle f(x), 2^j\psi(2^jx - k) \rangle$ with

$$w_j(k) = \sum_n g(n - 2k)c_{j+1}(n). \tag{10}$$

With this analysis, we have built the first part of a filter bank [26]. In order to restore the original data, Mallat [14] uses the properties of orthogonal wavelets, but the theory has been generalized to a large class of filters [4]. The restoration is performed with

$$c_{j+1}(k) = 2 \sum_l [c_j(l)\tilde{h}(k - 2l) + w_j(l)\tilde{g}(k - 2l)], \tag{11}$$

where \tilde{h} and \tilde{g} are the conjugate filters of h and g . The set of filters must satisfy the following relations:

- *Dealizing condition:*

$$\hat{h}(v + \frac{1}{2})\hat{\tilde{h}}(v) + \hat{g}(v + \frac{1}{2})\hat{\tilde{g}}(v) = 0, \tag{12}$$

- *Exact restoration:*

$$\hat{h}(v)\hat{\tilde{h}}(v) + \hat{g}(v)\hat{\tilde{g}}(v) = 1. \tag{13}$$

Many sets of filters were proposed especially for coding and it was shown [6] that the choice of these filters must be guided by the regularity of the scaling and the wavelet functions. The complexity is proportional to N . The algorithm provides a pyramid of N elements.

There exists a set of extensions of the wavelet transform to 2-D signals [14, 20], etc.

- Dilate the wavelet function independently in two dimensions.
- Use an isotropic wavelet function and only radial dilation.

- Transform using an anisotropic wavelet with a set of rotations [20].
- Process separate variables, but with only one scale for each coordinate.

In the case of separate variables, it is easy to extend the theory of the multiresolution analysis [14]. The scaling function $\phi(x, y)$ can be written as

$$\phi(x, y) = \phi(x)\phi(y). \quad (14)$$

Details are now given for three different wavelet functions. The filter bank associated with separated variables is similar to the 1-D case. The algorithm is performed line by line then column by column leading to four arrays decimated in x and y . The available 1-D filters are used. If we wish to use isotropic filters, we must exclude separated variables. It is possible to give relations for a non-separable filter bank [7], but it is not evident how to get an isotropic filter bank. The use of a quincunx decimation allows us to work with only one wavelet [7].

The multiresolution analysis furnishes a remarkable framework to code a signal, and more generally an image, with a pyramidal set of values. But this analysis is not invariant under translations. That has no importance for a signal coding which does not modify the data, but the situation is not the same in a strategy where we want to clean an image of spurious features. In our paper, the problem of the redundancy is not critical, but we wish to keep an isotropic vision. Therefore, we choose the *à trous* algorithm [2, 11, 25] instead of the filter bank.

2.3. The *à trous* algorithm

The discrete approach of the wavelet transform can be done with the special version of the so-called *à trous* algorithm (with holes) [11, 25]. One assumes that the sampled data $\{c_0(k)\}$ are the scalar products at pixels k of the function $f(x)$ with a scaling function $\phi(x)$ which corresponds to a low-pass filter.

The first filtering is then performed by a twice magnified scale leading to the $\{c_1(k)\}$ set. The

signal difference $\{c_0(k)\} - \{c_1(k)\}$ contains the information between these two scales and is the discrete set associated with the wavelet transform corresponding to $\psi(x)$. The associated wavelet is therefore $\psi(x)$:

$$\frac{1}{2}\psi\left(\frac{x}{2}\right) = \phi(x) - \frac{1}{2}\phi\left(\frac{x}{2}\right). \quad (15)$$

The distance between samples increasing by a factor 2 from the scale $(i - 1)$ ($i > 0$) to the next one, $c_i(k)$ is given by

$$c_i(k) = \sum_l h(l)c_{i-1}(k + 2^{i-1}l) \quad (16)$$

and the discrete wavelet transform $w_i(k)$ by

$$w_i(k) = c_{i-1}(k) - c_i(k). \quad (17)$$

The coefficients $\{h(k)\}$ derive from the scaling function $\phi(x)$:

$$\frac{1}{2}\phi\left(\frac{x}{2}\right) = \sum_l h(l)\phi(x - l). \quad (18)$$

The B-spline of degree 3 scaling function was used in our calculations. The algorithm allowing

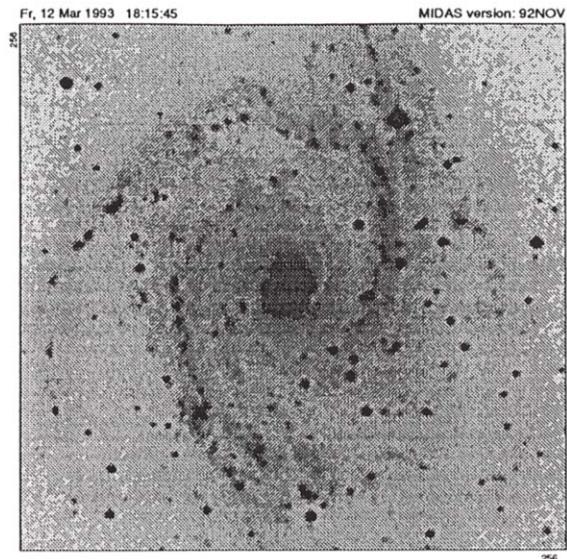


Fig. 1. Galaxy NGC 2997.

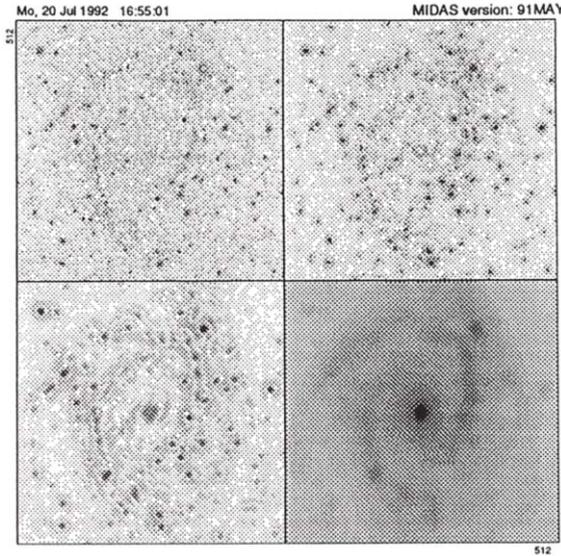


Fig. 2. Wavelet transform of the Galaxy.

one to rebuild the data frame is evident: the last smoothed array is added to all the differences.

The above *à trous algorithm* is easily extensible to the 2-D space. This leads to convolution with a mask of 5×5 pixels for the wavelet connected to B_3 -spline. At each scale i , we obtain a set $\{w_i(k, l)\}$ (we will call it wavelet plane in the next), which has the same number of pixels as the image.

Figure 2 shows the wavelet transform of a galaxy (Fig. 1). We compute only four planes, three wavelet planes and the image at lower resolution.

3. The noise reduction from the wavelet transform

3.1. The convolution from the continuous wavelet transform

We will examine here the computation of a convolution by using the continuous wavelet transform in order to get a framework for linear smoothings. Let us consider the convolution product of two functions:

$$h(x) = \int_{-\infty}^{+\infty} f(u)g(x - u) dx. \quad (19)$$

We introduce two real wavelet functions $\psi(x)$ and $\chi(x)$ such that

$$C = \int_0^{+\infty} \frac{\hat{\psi}^*(v)\hat{\chi}(v)}{v} dv \quad (20)$$

is defined. $W_g(a, b)$ designs the wavelet transform of g with the wavelet function $\psi(x)$:

$$W_g(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} g(x)\psi^*\left(\frac{x-b}{a}\right) dx. \quad (21)$$

We restore $g(x)$ with the wavelet function $\chi(x)$:

$$g(x) = \frac{1}{C} \int_0^{+\infty} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{a}} W_g(a, b)\chi\left(\frac{x-b}{a}\right) \frac{da db}{a^2}. \quad (22)$$

The convolution product can be written as

$$h(x) = \frac{1}{C} \int_0^{+\infty} \frac{da}{a^{5/2}} \int_{-\infty}^{+\infty} W_g(a, b) db \times \int_{-\infty}^{+\infty} f(u)\chi\left(\frac{x-u-b}{a}\right) du. \quad (23)$$

Let us denote $\tilde{\chi}(x) = \chi(-x)$. The wavelet transform $\tilde{W}_f(a, b)$ of $f(x)$ with the wavelet $\tilde{\chi}(x)$ is

$$\tilde{W}_f(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x)\tilde{\chi}\left(\frac{x-b}{a}\right) dx. \quad (24)$$

This leads to

$$h(x) = \frac{1}{C} \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} \tilde{W}_f(a, x-b) W_g(a, b) db. \quad (25)$$

Then we get the final result

$$h(x) = \frac{1}{C} \int_0^{+\infty} \tilde{W}_f(a, x) \otimes W_g(a, x) \frac{da}{a^2}. \quad (26)$$

In order to compute a convolution with the continuous wavelet transform:

we compute the wavelet transform $\tilde{W}_f(a, b)$ of the function $f(x)$ with the wavelet function $\tilde{\chi}(x)$;

- we compute the wavelet transform $W_g(a, b)$ of the function $g(x)$ with the wavelet function $\psi(x)$;
- we sum the convolution product of the wavelet transforms, scale by scale.

The wavelet transform permits us to perform any linear filtering. Its efficiency depends on the number of terms in the wavelet transform associated with $g(x)$ for a given signal $f(x)$. If we have a filter where the number of significant coefficients is small for each scale, the complexity of the algorithm is proportional to N . For a classical convolution, the complexity is also proportional to N , but the number of operations is also proportional to the length of the convolution mask. The main advantage of the present technique lies in the possibility of having a filter with long-scale terms without computing the convolution on a large window. If we achieve the convolution with the FFT algorithm, the complexity is in $N \log_2 N$, the computing time is longer than the one obtained with the wavelet transform if we concentrate the energy on very few coefficients.

3.2. The Wiener-like filtering in the wavelet space

Let us consider a measured wavelet coefficient w_i at the scale i . We assume that its value, at a given scale and a given position, results from a noisy process, with a Gaussian law of a mathematical expectation W_i , and a standard deviation B_i :

$$P(w_i/W_i) = \frac{1}{\sqrt{2\pi B_i}} e^{-(w_i - W_i)^2/2B_i^2}. \quad (27)$$

Now, we assume that the set of expected coefficients W_i for a given scale also follows a Gaussian law, with a null mean and a standard deviation S_i :

$$P(W_i) = \frac{1}{\sqrt{2\pi S_i}} e^{-w_i^2/2S_i^2}. \quad (28)$$

The null mean value results from the wavelet property

$$\int_{-\infty}^{+\infty} \psi^*(x) dx = 0. \quad (29)$$

Signal Processing

We want to get an estimation of W_i knowing w_i . Bayes' theorem gives

$$P(W_i/w_i) = \frac{P(W_i)P(w_i/W_i)}{P(w_i)}. \quad (30)$$

We get

$$P(W_i/w_i) = \frac{1}{\sqrt{2\pi\beta_i}} e^{-(W_i - \alpha_i w_i)^2/2\beta_i^2}, \quad (31)$$

where

$$\alpha_i = \frac{S_i^2}{S_i^2 + B_i^2}, \quad (32)$$

the probability $P(W_i/w_i)$ is a Gaussian law with a mean

$$m = \alpha_i w_i \quad (33)$$

and a variance

$$\beta_i^2 = \frac{S_i^2 B_i^2}{S_i^2 + B_i^2}. \quad (34)$$

The mathematical expectation of W_i is $\alpha_i w_i$.

With a simple multiplication of the coefficients by the constant α_i , we get a linear filter. The algorithm is

1. Compute the wavelet transform of the data. We get w_i .
2. Estimate the standard deviation of the noise B_0 of the first plane from the histogram of w_0 . As we process oversampled images, the values of the wavelet image corresponding to the first scale (w_0) are due mainly to the noise. The histogram shows a Gaussian peak around 0. We compute the standard deviation of this Gaussian function, with a 3σ clipping, rejecting pixels where the signal could be significant.
3. Set i to 0.
4. Estimate the standard deviation of the noise B_i from B_0 . This is done from the study of the variation of the noise between two scales, with an hypothesis of a white Gaussian noise.
5. $S_i^2 = s_i^2 - B_i^2$ where s_i^2 is the variance of w_i .
6. $\alpha_i = S_i^2/(S_i^2 + B_i^2)$.
7. $W_i = \alpha_i w_i$.
8. $i = i + 1$ and go to 4.
9. Reconstruct the picture from W_i .

We simulate an image by convolving the galaxy NGC 2997 by a point spread function (PSF). A noise is added in order to get a signal-to-noise ratio of 4 (see Fig. 3 at the right). In Fig. 4 we compare the image obtained after the multiresolution Wiener filtering to the original one (galaxy convolved by the PSF, but without any noise).

3.3. The hierarchical Wiener filtering

In the above process, we do not use the information between the wavelet coefficients at different scales. We modify the previous algorithm by introducing a prediction w_h of the wavelet coefficient from the upper scale. This prediction could be de-

termined from the regression [1] between the two scales but better results are obtained when we only set w_h to W_{i+1} . Between the expectation coefficient W_i and the prediction, a dispersion exists where we assume that it is a Gaussian law:

$$P(W_i/w_h) = \frac{1}{\sqrt{2\pi T_i}} e^{-(W_i-w_h)^2/2T_i^2} \tag{35}$$

The relation which gives the coefficient W_i knowing w_i and w_h is

$$P(W_i/w_i \text{ and } w_h) = \frac{1}{\sqrt{2\pi\beta_i}} e^{-(W_i-\alpha_i w_i)^2/2\beta_i^2} \times \frac{1}{\sqrt{2\pi T_i}} e^{-(W_i-w_h)^2/2T_i^2}, \tag{36}$$

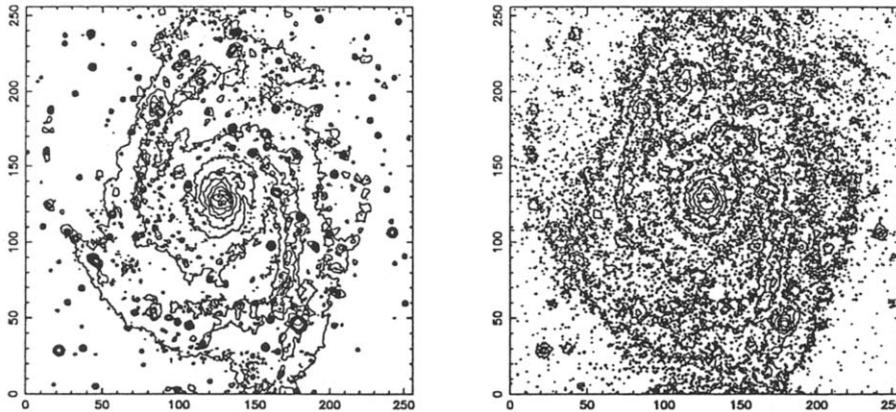


Fig. 3. Galaxy NGC 2997 (left), and noisy image (right).

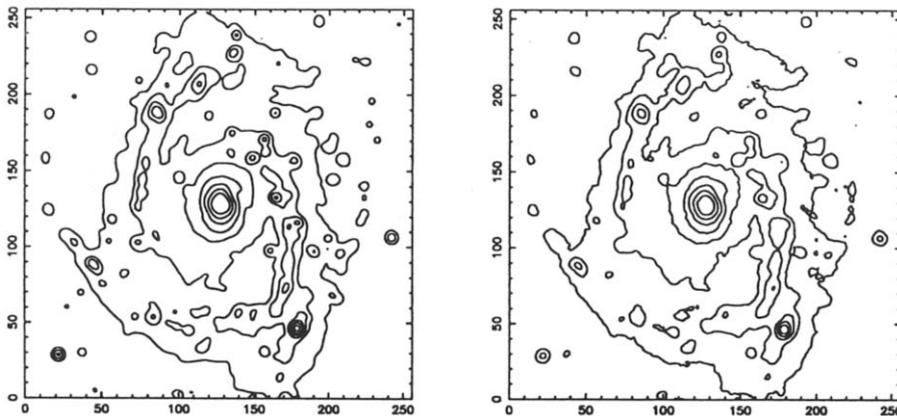


Fig. 4. Original image (left), and image obtained after multiresolution Wiener filtering (right).

with

$$\beta_i^2 = \frac{S_i^2 B_i^2}{S^2 + B_i^2} \tag{37}$$

and

$$\alpha_i = \frac{S_i^2}{S_i^2 + B_i^2}. \tag{38}$$

It is a Gaussian law with a mathematical expectation:

$$W_i = \frac{T_i^2}{B_i^2 + T_i^2 + Q_i^2} w_i + \frac{B_i^2}{B_i^2 + T_i^2 + Q_i^2} w_h, \tag{39}$$

with

$$Q_i^2 = \frac{T_i^2 B_i^2}{S_i^2}. \tag{40}$$

W_i is the barycentre of the three values $w_i, w_h, 0$ with the weights T_i^2, B_i^2, Q_i^2 . The particular cases are

- if the noise is important ($S_i \ll B_i$) and even if the correlation between the two scales is good (T_i is low), we get $W_i \rightarrow 0$,
- if $B_i \ll S_i \ll T$ then $W_i \rightarrow w_i$,
- if $B_i \ll T_i \ll S$ then $W_i \rightarrow w_i$,
- if $T_i \ll B_i \ll S$ then $W_i \rightarrow w_h$.

At each scale, by changing all the wavelet coefficients w_i of the plane by the estimate value W_i , we

get a hierarchical Wiener filter. The algorithm is

- (1) Compute the wavelet transform of the data. We get w_i .
- (2) Estimate the standard deviation of the noise B_0 of the first plane from the histogram of w_0 .
- (3) Set i to the index associated with the last plane: $i = n$.
- (4) Estimate the standard deviation of the noise B_i from B_0 .
- (5) $S_i^2 = s_i^2 - B_i^2$ where s_i^2 is the variance of w_i .
- (6) Set w_h to W_{i+1} and compute the standard deviation T_i of $w_i - w_h$.
- (7) $W_i = T_i^2 / (B_i^2 + T_i^2 + Q_i^2) w_i + B_i^2 / (B_i^2 + T_i^2 + Q_i^2) w_h$.
- (8) $i = i - 1$. If $i > 0$ go to 4.
- (9) Reconstruct the picture.

We compare in Fig. 5 the image obtained after the hierarchical Wiener filtering to the original one.

3.4. The adaptive filtering from the wavelet transform

In the preceding algorithm we have assumed the properties of the signal and the noise to be stationary. The wavelet transform was first used to obtain an algorithm which is faster than classical Wiener filtering. Then we took account of the correlation between two different scales. In this way we got a filtering with stationary properties. In fact, these hypotheses were too simple, because in general the

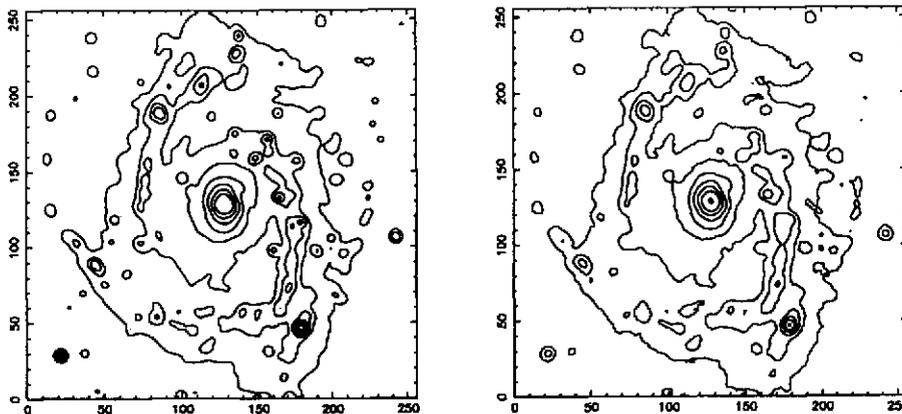


Fig. 5. Original image (left), and image obtained after hierarchical Wiener filtering (right).

signal may not arise from a Gaussian stochastic process. Knowing the noise distribution, we can determine the statistically significant level at each scale of the measured wavelet coefficients. If $w_i(x)$ is very weak, it is not significant and could be due to noise. Then the hypothesis that the value $W_i(x)$ is null is not forbidden. In the opposite case where $w_i(x)$ is significant, we keep its value. If the noise is Gaussian, we write

$$W_i = \begin{cases} 0 & \text{if } |w_i| < kB_i, \\ w_i & \text{if } |w_i| \geq kB_i. \end{cases} \quad (41)$$

Generally, we choose $k = 3$.

With a filter bank we have a biunivocity between the image and its transform, so that the thresholded transform leads to only one restored image. Some experiments show us that uncontrolled artifacts appear for high-level thresholding ($k = 3$). The decimation done at each step on the wavelet transform takes into account the knowledge of the coefficients at further resolutions. The thresholding sets to zero the intrinsic small terms which play their part in the reconstruction. With the lattice filter the situation is very different. No decimation is done and the thresholding keeps all significant coefficients. Where the coefficients are set to zero, we do not put zero, but we say that these values are unknown. The redundancy is used to restore them. Before the thresholding we have a redundant transform, which can be decimated; after the thresholding we get a set of coefficients from which we wish to restore an image.

If one applies the reconstruction algorithm, then it is not guaranteed that the wavelet transform of the restored image will give the same values for the coefficients. This is not important in the case where they are not significant, but otherwise the same values must be found. If $W_i^{(s)}$ are the coefficients obtained by the thresholding, then we require $W_i(x)$ such that

$$PW_i(x) = W_i^{(s)}(x), \quad (43)$$

where P is the non-linear operator which performs the inverse transform, the wavelet transform, and the thresholding. An alternative is to use the following iterative solution which is similar to Van Cittert's algorithm:

$$W_i^{(n)}(x) = W_i^{(s)}(x) + W_i^{(n-1)}(x) - PW_i^{(n-1)}(x) \quad (44)$$

for the significant coefficients ($W_i^{(s)}(x) \neq 0$) and

$$W_i^{(n)}(x) = W_i^{(n-1)}(x) \quad (45)$$

for the non-significant coefficients ($W_i^{(s)}(x) = 0$).

The algorithm is the following one:

- (1) Compute the wavelet transform of the data. We get w_i .
- (2) Estimate the standard deviation of the noise B_0 of the first plane from the histogram of w_0 .
- (3) Estimate the standard deviation of the noise B_i from B_0 at each scale.
- (4) Estimate the significant level at each scale, and threshold.
- (5) Initialize: $W_i^{(0)}(x) = W_i^{(s)}(x)$.
- (6) Reconstruct the picture by using the iterative method.

The thresholding may introduce negative values in the resulting image. A positivity constraint can be introduced in the iterative process by thresholding the restored image. The algorithm converges after five or six iterations. In Fig. 6, we can see the resulting image after adaptive filtering.

We remark that point-like artifacts remain which are the main defect of this adaptive filtering. The number of independent wavelet coefficients is proportional to $(N/a)^2$ (a is the scale). The thresholding does not cause all insignificant coefficients to vanish since a proportion depending on k remains. For $k = 3$ the proportion of remaining coefficients for a Gaussian process is around 0.25%. The number of remaining insignificant values is important for scale 1 and it is reduced by a factor 4 from one scale to another. This leads to point-like artifacts. We can increase the value of k , but this leads to a great smoothing of the large-scale structures. Furthermore, we may increase k only for scale 1.

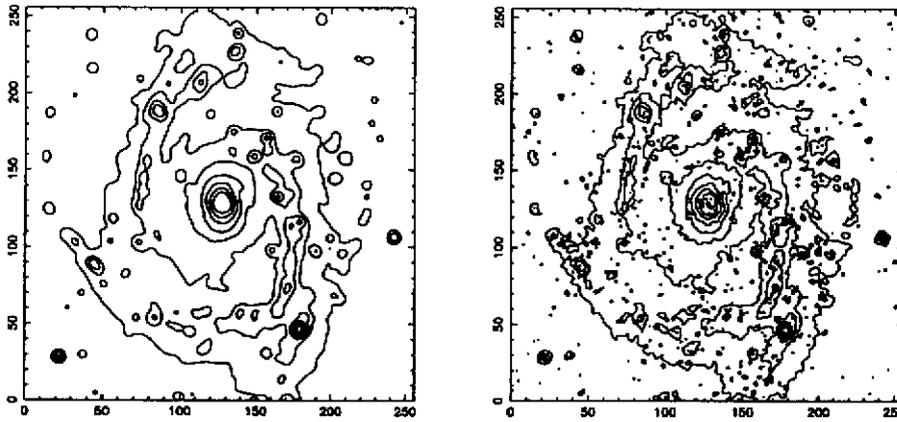


Fig. 6. Original image (left), and image obtained after adaptive filtering (right).

3.5. The hierarchical adaptive filtering

In the previous algorithm we do not use the hierarchy of structures. We have explored many approaches for introducing a non-linear hierarchical law in the adaptive filtering and we found that the best way was to link the threshold to the wavelet coefficient of the previous plane w_h . We get

$$W_i(x) = \begin{cases} w_i(x) & \text{if } |w_i(x)| \geq L, \\ 0 & \text{if } |w_i(x)| < L \end{cases}$$

and L is a threshold estimated by

$$\text{if } |w_i(x)| \geq kB_i \quad \text{then } L = kB_i,$$

$$\text{if } |w_i(x)| < kB_i \quad \text{then } L = kB_i t\left(\frac{w_h}{S_h}\right),$$

where S_h is the standard deviation of w_h . The function $t(a)$ must return a value between 0 and 1. A possible function for t is

$$\begin{aligned} - t(a) &= 0 & \text{if } a \geq k, \\ - t(a) &= 1 - \frac{1}{k}a & \text{if } a < k. \end{aligned}$$

Figure 7 shows the image resulting from hierarchical adaptive filtering.

3.6. Comparison from a multiresolution quality criterion

It is important to evaluate the quality of the restoration. Very few quantitative parameters can be extracted for that. The correlation between the

original image $I(i, j)$ and the restored one $\tilde{I}(i, j)$ gives a classical criterion. The correlation coefficient is

$$C_{or} = \frac{\sum_{i=1}^N \sum_{j=1}^N I(i, j) \tilde{I}(i, j)}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N I^2(i, j) \sum_{i=1}^N \sum_{j=1}^N \tilde{I}^2(i, j)}}. \quad (46)$$

The correlation is 1 if the images are identical, and less if some differences exist. Another way to compare two pictures is to determine the mean-square error

$$E_{ms}^2 = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (I(i, j) - \tilde{I}(i, j))^2. \quad (47)$$

E_{ms}^2 can be normalized by

$$E_{nms}^2 = \frac{\sum_{i=1}^N \sum_{j=1}^N (I(i, j) - \tilde{I}(i, j))^2}{\sum_{i=1}^N \sum_{j=1}^N I^2(i, j)}. \quad (48)$$

The signal-to-noise ratio (SNR) corresponding to the above error is

$$\text{SNR}_{dB} = 10 \log_{10} \frac{1}{E_{nms}^2} \text{ dB}. \quad (49)$$

Table 1 indicates the correlation rate, the quadratic error between the original image and the filtered ones, and the SNR.

These criteria are not sufficient, they give no information on the resulting resolution. A complete criterion must take into account the resolution. We

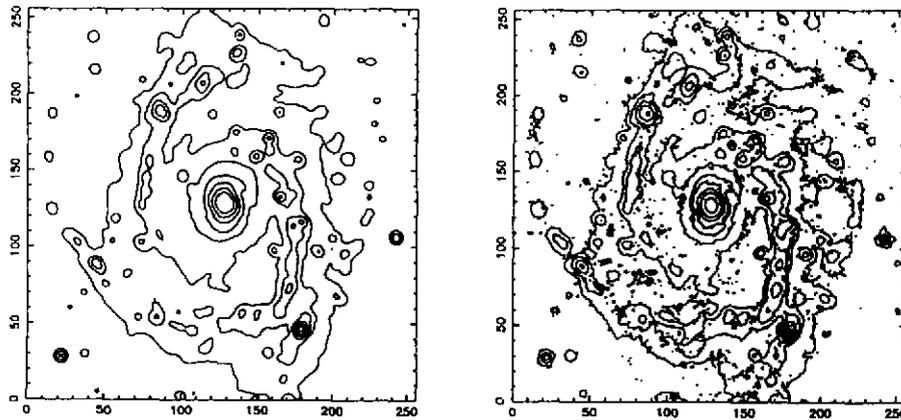


Fig. 7. Original image (left), and image obtained after hierarchical adaptive filtering (right).

Table 1
Comparison between the original image and the filtered ones

Images	Noisy image	Multiresolution Wiener	Thresholding	Hierarchical Thresholding	Hierarchical Wiener	Median filter
C_{or}	0.981612	0.998624	0.994335	0.991629	0.998855	0.996716
SNR(dB)	14.21	25.59	19.41	17.55	26.39	21.83

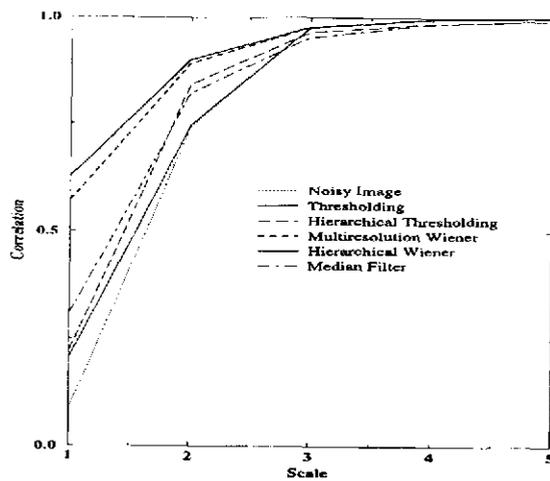


Fig. 8. Correlation between the wavelet coefficients of original image and the filtered ones.

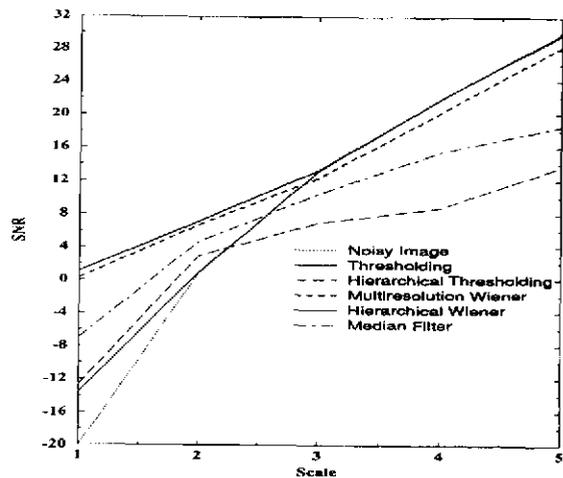


Fig. 9. Reconstruction error.

can compute for each dyadic scale the correlation coefficient and the quadratic error between the wavelet transforms of the original and the restored images, so that we can compare for each resolution,

the quality of the restoration. Figures 8 and 9 present the same information, but scale by scale.

From Table 1, we conclude that the hierarchical Wiener filtering gives on our image the best results.

We see from Figs. 8 and 9 that it is the best at all scales. The hierarchical Wiener filtering has the advantage of producing smooth images, without any artefact. The hierarchical thresholding filtering is, in our simulation, the worst method. It is, however, sometimes useful when the image contains edges which are quite preserved in the filtering.

4. The deconvolution from the wavelet transform

4.1. The direct deconvolution

Let us consider an image characterized by its intensity distribution $I(x, y)$. It corresponds to the observation of an object distribution $O(X, Y)$ through an optical system. We reduced the coordinates of the object to the one of the image. If the imaging system is linear and shift-invariant, the relation between the object and the image in the same coordinates frame is a convolution:

$$I(x, y) = \int_{-\infty}^{+\infty} O(X, Y) P(x - X, y - Y) dX dY.$$

$P(x, y)$ is the point spread function (PSF) of the imaging system. We want to determine $O(X, Y)$ knowing $I(x, y)$ and $P(x, y)$. This inverse problem led to a large number of papers from which it follows that the main difficulties lie on the existence of (i) a cut-off frequency of the PSF and (ii) an intensity noise (see e.g. [5]).

We remark that the equation can be written as

$$I(x, y) = \langle O(X, Y), \phi(X - x, Y - y) \rangle, \quad (50)$$

where $\phi(x, y) = P(-x, -y)$. The image is sampled and we get an array $I(n, m)$ such that

$$I(n, m) = \langle O(X, Y), \phi(X - n, Y - m) \rangle. \quad (51)$$

If we want to restore $O(X, Y)$ using the dual basis associated to $\phi(x, y)$, we get

$$\hat{\phi}(u, v) = \frac{\hat{\phi}^*(u, v)}{D(u, v)}, \quad (52)$$

where

$$D(u, v) = \sum_{k, l} |\hat{\phi}(u + k, v + l)|^2. \quad (53)$$

Taking into account the existence of a cut-off frequency smaller than $\frac{1}{2}$, for a well-sampled image

$$D(u, v) = |\hat{\phi}(u, v)|^2,$$

so that

$$\hat{\phi}(u, v) = \frac{1}{\hat{\phi}(u, v)}. \quad (54)$$

We found the trivial inverse solution. This needs $\hat{P}(u, v) \neq 0$. We remark that the concept of the dual basis is only available for undersampled data. For oversampled ones, there exists a dependency between the data which leads to perfect solution. A classical solution to this problem lies in the generalized inverse [22]. Landweber's iterative scheme [12] is a typical solution. It corresponds to

$$\begin{aligned} O^{(n)}(X, Y) &= O^{(n-1)}(X, Y) + \omega \tilde{P}(X, Y) \\ &\otimes [I(X, Y) - O^{(n-1)}(X, Y) \\ &\otimes P(X, Y)]. \end{aligned} \quad (55)$$

where ω is a converging parameter generally taken as 1 and $\tilde{P}(x, y) = P(-x, -y)$. We start with $O^{(0)}(X, Y) = I(X, Y)$.

In Fourier space we get

$$\begin{aligned} \hat{O}^{(n)}(u, v) &= \hat{O}^{(n-1)}(u, v) + \omega \hat{P}^*(u, v) \\ &\times [\hat{I}(u, v) - \hat{O}^{(n-1)}(u, v) \hat{P}(u, v)]. \end{aligned} \quad (56)$$

Above the cut-off frequency $\hat{P}(u, v) = 0$, so $\hat{O}^{(n)}(u, v) = \hat{O}^{(n-1)}(u, v)$. If $\hat{I}(u, v) = 0$, we always get $\hat{O}^{(n)}(u, v) = 0$: the inversion does not add information above the frequency cut-off. For small $\hat{P}(u, v)$ values the convergence is slow.

The filterings seen above give smoothed images where the noise is removed for the frequencies above the cut-off one. We may use Landweber's method with a positivity constraint in order to deconvolve the image. The resulting algorithm is

- (1) noise reduction with a hierarchical filtering;
- (2) iterative deconvolution with Landweber's method.

We plot in Fig. 10 the resulting image after 50 iterations. We remark that we get an enhancement of the resolution and that we do not have any noise in the restored image.

4.2. Wavelets generated by the point spread function

We remark that the PSF is similar to a scaling function, and we may introduce a wavelet function generated by this function. If the sampling step is 1, the cut-off frequency ν_c is less than $\frac{1}{2}$. We admit that $\hat{P}(u, v)$ is never null below this frequency. We set $\phi(x, y) = P(-x, -y)$. Let us define $\hat{h}(u, v)$ such that

$$\hat{h}(u, v) = \frac{\hat{\phi}(2u, 2v)}{\hat{\phi}(u, v)} \tag{57}$$

for $|u|, |v| \leq \nu_c/2$, and $\hat{h}(u, v) = 0$ for any other frequencies where $|u|, |v| \leq \frac{1}{2}$. For other frequencies we set

$$\hat{h}(u + n, v + m) = \hat{h}(u, v). \tag{58}$$

With this definition, we can always write

$$\hat{\phi}(2u, 2v) = \hat{h}(u, v)\hat{\phi}(u, v). \tag{59}$$

The PSF is a scaling function associated to fast multiresolution algorithms. In the direct space we write

$$\frac{1}{4} \phi\left(\frac{x}{2}, \frac{y}{2}\right) = \sum_{n,m} h(n, m)\phi(x - n, y - m). \tag{60}$$

This allows us to define wavelet transforms generated by the PSF with the relation

$$\hat{\psi}(2u, 2v) = \hat{g}(u, v)\hat{\phi}(u, v), \tag{61}$$

where $\hat{g}(u + n, v + m) = \hat{g}(u, v)$.

In the square of side $]-\frac{1}{2}, +\frac{1}{2}[$ in the two dimensions, $\hat{h}(u, v)$ is necessarily limited in the interval $]-\frac{1}{4}, +\frac{1}{4}[$ in the two dimensions, so that its Fourier transform $h(n, m)$ has an infinite length. We must process the transform in the Fourier space. The algorithm is the following one:

- (1) Compute the corresponding image FFT. We name T_0 the resulting complex array.
- (2) Set i to 0. We iterate.
- (3) Take the inner product of T_i with $\hat{g}(2^i u, 2^i v)$. We get the complex array W_{i+1} . The inverse FFT gives the wavelet coefficients at the scale 2^i .
- (4) Take the inner product of T_i with $\hat{h}(2^i u, 2^i v)$. We get the array T_{i+1} . Its inverse FFT gives the image at the scale 2^{i+1} . The frequency band is reduced by a factor 2.
- (5) We increment i and we go back to Step 3.

We decimate only the interpolation array scale by scale. We get a pyramidal set of data for the wavelet transform. The reconstruction can be done using all the wavelet transform coefficients, or scale by scale. In this last case, we use the relations

$$T_{i+1}(u, v) = \hat{h}(2^i u, 2^i v)T_i(u, v),$$

$$W_{i+1}(u, v) = \hat{g}(2^i u, 2^i v)T_i(u, v).$$

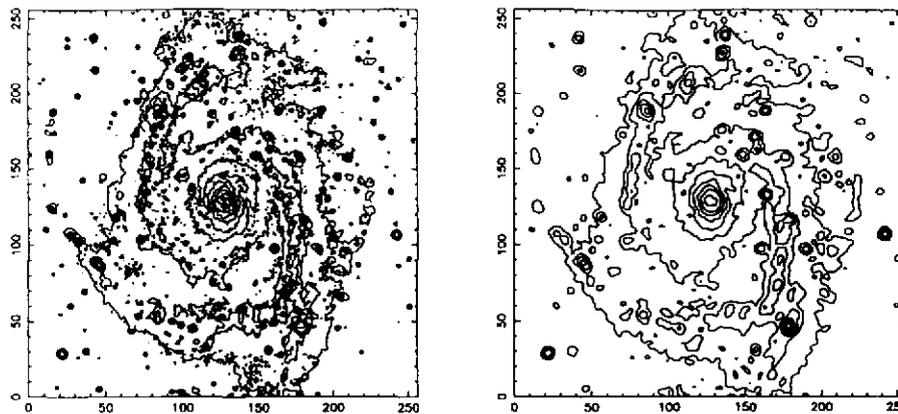


Fig. 10. Original object (left), and image deconvolved (right).

Knowing the arrays $T_{i+1}(u, v)$ and $W_{i+1}(u, v)$, we can determine $T_i(u, v)$. We weight each equation with the functions $p_h(2^i u, 2^i v)$ and $p_g(2^i u, 2^i v)$. Using the least-mean-square estimator we get

$$T_i(u, v) = \frac{p_h(u, v) \hat{h}^*(2^i u, 2^i v)}{D(2^i u, 2^i v)} T_{i+1}(u, v) + \frac{p_g(u, v) \hat{g}^*(2^i u, 2^i v)}{D(2^i u, 2^i v)} W_{i+1}(u, v),$$

where

$$D(u, v) = p_h(u, v) |\hat{h}(u, v)|^2 + p_g(u, v) |\hat{g}(u, v)|^2.$$

This leads to the conjugate filters

$$\hat{h}(u, v) = \frac{p_h(u, v) \hat{h}^*(u, v)}{D(u, v)},$$

$$\hat{g}(u, v) = \frac{p_g(u, v) \hat{g}^*(u, v)}{D(u, v)}.$$

We verify that the exact restoration condition for a filter bank is true:

$$\hat{h}(u, v) \hat{h}(u, v) + \hat{g}(u, v) \hat{g}(u, v) = 1.$$

But the dealiasing condition is not verified. We can decimate the low-resolution image but we cannot decimate the wavelet image. We get a pyramidal algorithm but not a classical filter bank. Generally, we choose

$$p_h(u, v) = p_g(u, v) = 1.$$

In a first case, we considered the filter $g(u, v)$ resulting from the difference between the approximations at two following scales:

$$\hat{g}(u, v) = 1 - \hat{h}(u, v).$$

In Fig. 11 we plot the chosen scaling function derived from a B-spline of degree 3 in frequency space and its resulting wavelet function. Their conjugate functions are plotted in Fig. 12. The

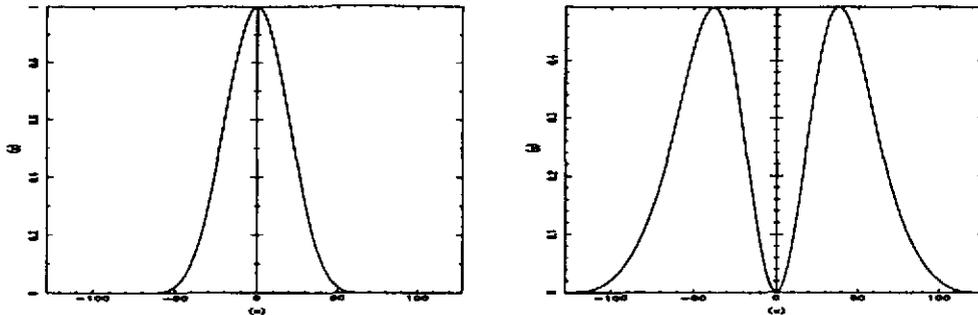


Fig. 11. The scaling function $\hat{\phi}$ (left), and the wavelet $\hat{\psi}$ (right).

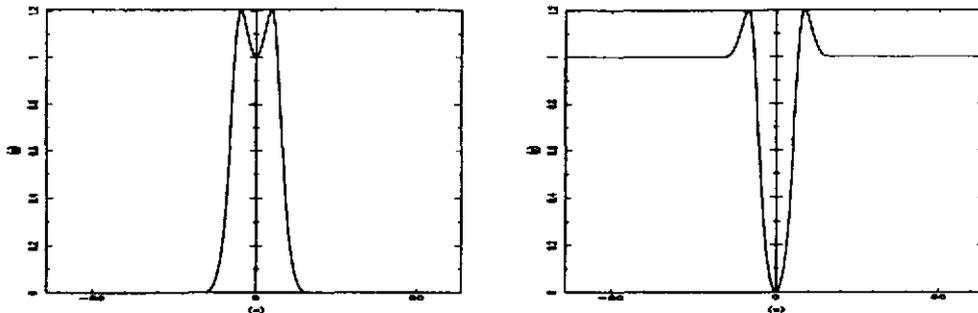


Fig. 12. The filter \hat{h} (left), and the filter \hat{g} (right).

reconstruction is done using

$$T_i(u, v) = T_{i+1}(u, v) + W_{i+1}(u, v),$$

with the following weighting functions:

$$p_h(u, v) = \frac{1}{\hat{h}^*(u, v)}, \quad p_g(u, v) = \frac{1}{\hat{g}^*(u, v)}.$$

Then we considered the filter

$$\hat{g}(u, v) = \sqrt{1 - |\hat{h}(u, v)|^2},$$

which leads to

$$\tilde{h}(u, v) = \hat{h}^*(u, v), \quad \tilde{g}(u, v) = \hat{g}^*(u, v).$$

We plot the resulting wavelet function in Fig. 13.

4.3. Deconvolution with the wavelet transform generated by the PSF

We do the wavelet transform with a wavelet function generated by the PSF. We get a pyramidal set of data

$$w_i(n, m) = \frac{1}{2^{2i}} \left\langle O(x, y), \psi \left(\frac{x}{2^i} - n, \frac{y}{2^i} - m \right) \right\rangle.$$

For each scale we get

$$c_i(n, m) = \frac{1}{2^{2i}} \left\langle O(x, y), \phi \left(\frac{x}{2^i} - n, \frac{y}{2^i} - m \right) \right\rangle.$$

This allows to restore the data at the lower scale 2^{i-1} .

We can apply on this pyramid the hierarchical smoothing algorithms in order to reduce the noise.

We get at the lowest scale a set of interpolation coefficients:

$$f(n, m) = \langle O(x, y), \phi(x - n, y - m) \rangle.$$

We come back to the inversion of this equation. We saw that it was impossible to give a solution such that

$$O(x, y) = \sum_{n,m} f(n, m) \phi(x - n, y - m).$$

We may restore using Landweber's iterative scheme.

We see in Fig. 14 the original object (left), and the deconvolved image (right) after 50 iterations with the Landweber algorithm. The wavelet coefficients used for the deconvolution were those obtained after the filtering by the hierarchical Wiener filter. Compared to Fig. 10, the results are not significantly better. So we can see that the use of the PSF as the scaling function does not yield the best restoration.

We considered below a regular PSF, without zero-value before the cut-off frequency. If it is not the case, we are not able to compute the filter $\hat{h}(u, v)$, because the term $\phi(2u, 2v)/\phi(u, v)$ is undefined for some frequencies. For instance, for the aperture synthesis used in radio astronomical mapping, there always exists 'frequency holes' (i.e. we do not have a full coverage of the u, v plane), and it is impossible to restore the image by a simple inversion. Nevertheless, the wavelet transform should be used, but in a different way than has been developed in this paper [27]. The deconvolution

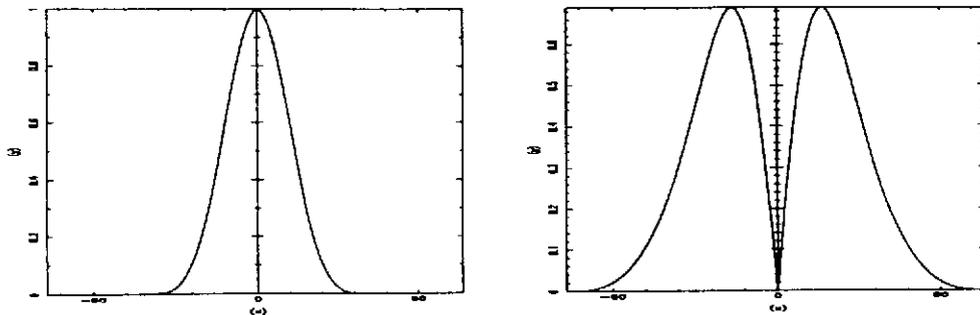


Fig. 13. The scaling function $\hat{\phi}$ is shown on the left, and the wavelet $\hat{\psi}$ on the right.

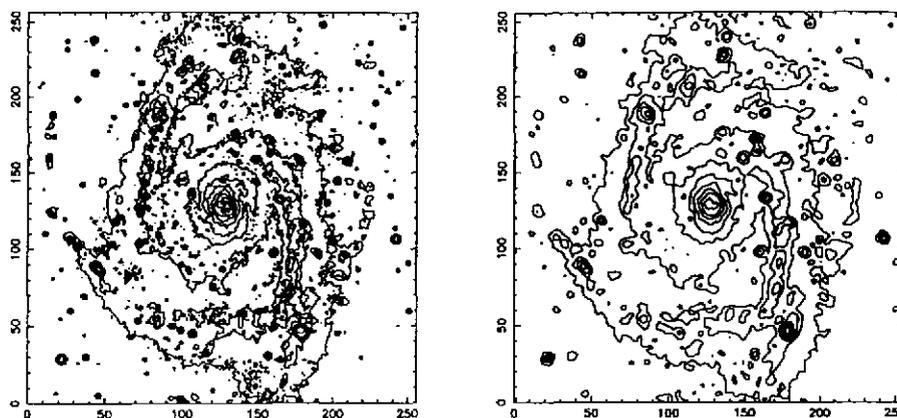


Fig. 14. Original object (left), and image obtained after deconvolution (right).

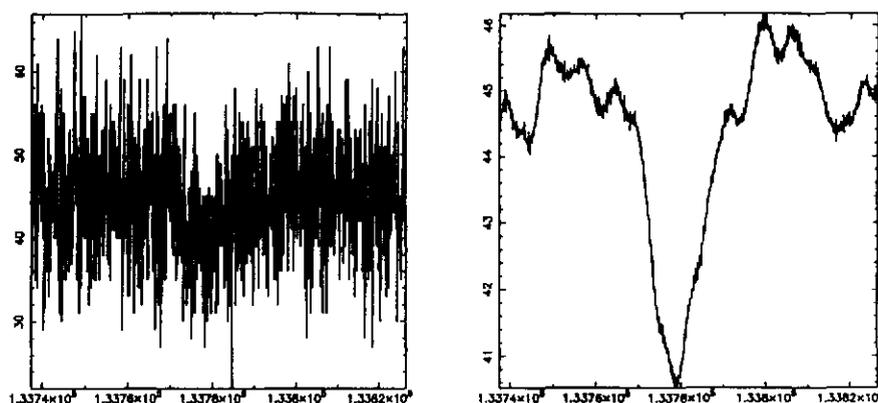


Fig. 15. PPS data of Voyager 2 on the Encke Division in Saturn's rings. On the left the real data with a resolution of 90 m, and on the right the filtered signal. These plots give counts versus the radial distance from the center of Saturn.

has been done here with Landweber's algorithm with a positivity constraint. Other regularization constraints such as maximum entropy [8] or Thikonov's [28] could also be introduced, which can be achieved at the full restoration, or scale by scale.

5. Conclusion

In this paper we show some applications of the wavelet transform for filtering and deconvolution. We use discrete transforms derived from the lattice

Signal Processing

or pyramidal algorithms. The redundancy which arises is used to take account of the interscale constraint relations.

We show in Fig. 15 a current application of the hierarchical Wiener filtering applied to a one-dimensional signal. No parameter was introduced to get the smoothed signal and the noise has been well reduced.

When images present hierarchical structures, as is generally the case in astronomy, tools which take into account this hierarchy are of significant interest and allow one to optimize the restoration.

Acknowledgments

We wish to thank J. Tully and J.M. Petit for useful discussions and comments.

References

- [1] M. Antonini, M. Barlaud and P. Mathieu, "Image coding using vector quantization in the wavelet transform domain", *Proc. Internat. Conf. Acoust. Speech Signal Process.*, Albuquerque, USA, 1990.
- [2] A. Bijaoui, "Algorithmes de la transformation en ondelettes, Application à l'imagerie astronomique", *Proc. Ondelettes et Paquets d'Ondes*, Inria, Rocquencourt, 1991.
- [3] C.H. Chui, "An introduction to wavelets", in: *Wavelet Analysis and its Application*, Academic Press, Harcourt Brace Jovanovich, 1992.
- [4] A. Cohen, I. Daubechies and J.C. Feauveau, "Biorthogonal bases of compactly supported wavelets", *Commun. Pure Appl. Math.*, Vol. 45, 1992, pp. 485–560.
- [5] T.J. Cornwell, "Image restoration", *Proc. NATO Advanced Study Institute on Diffraction-Limited Imaging with Very Large Telescopes*, Cargèse, 1988, pp. 273–292.
- [6] I. Daubechies, "Orthogonal bases of compactly supported wavelets", *Commun. Pure Appl. Math.*, Vol. 41, 1988, pp. 909–996.
- [7] J.C. Feauveau, *Analyse Multiresolution par Ondelettes non Orthogonales et Bancs de Filtres Numériques*, Thèse de doctorat de l'université, Paris Sud, 1990.
- [8] B.R. Frieden, "Image enhancement and restoration", in: *Topics in Applied Physics*, Springer, Berlin, Vol. 6, 1975, pp. 177–249.
- [9] A. Grossmann and J. Morlet, "Decomposition of Hardy functions into square integrable wavelets of constant shape", *SIAM J. Math. Anal.*, Vol. 15, 1984, pp. 723–736.
- [10] M. Holschneider, R. Kronland-Martinet, J. Morlet and Ph. Tchamitchian, "The à trous Algorithm", *CPT-88/P.2215*, Berlin, 1988, pp. 1–22.
- [11] M. Holschneider and P. Tchamitchian, *Les ondelettes en 1989*, PG Lemarié, Springer, Berlin, 1990, p. 102.
- [12] L. Landweber, "An iteration formula for Fredholm integral equations of the first kind", *Amer. J. Math.*, Vol. 73, 1951, pp. 615–624.
- [13] J. Littlewood and R. Paley, *J. Lond. Math. Soc.*, Vol. 6, 1931, p. 230.
- [14] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation", *Proc. IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 11, No. 7, 1989.
- [15] G.A. Mastin, "Adaptative filters for digital Image" *Comput. Vision Graph. Image Process.*, Vol. 31, 1985, pp. 103–121.
- [16] Y. Meyer, in: J.M. Combes et al., eds. *Wavelets*, Springer, Berlin, 1989, p. 21.
- [17] Y. Meyer, *Ondelettes*, ed. Hermann, Paris, 1990.
- [18] Y. Meyer, "Méthodes temps-fréquence et méthode temps-échelle en traitement du signal et de l'image", *Proc. Ondelettes et Paquets d'Ondes*, Inria, Rocquencourt, 1991.
- [19] Y. Meyer, *Ondelettes: Algorithmes et Applications*, A. Colin, Paris, 1992.
- [20] R. Murenzi, in: J.M. Combed et al., eds., *Wavelets*, Springer, Berlin, 1988.
- [21] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill Series in System Science, 1965.
- [22] W.K. Pratt, *Digital Image Processing*, Wiley/Interscience, New York, 1978, p. 378.
- [23] M.B. Ruskai, G. Beylkin, R. Coifman, I. Daubechies, S. Mallat, Y. Meyer and L. Raphael, *Wavelets and their Applications*, Jones and Barlett Publishers, Boston, 1992.
- [24] C.E. Shannon, *Bell System Tech.*, Vol. 27, 1948, p. 379.
- [25] M.J. Shensa, "Discrete wavelet transforms: Wedding the à trous and Mallat Algorithms", *IEEE Trans. Signal Process.*, Vol. 40, No. 10, 1992, pp. 2464–2482.
- [26] M.J.T. Smith and T.P. Barnwell, *III IEEE Acoust. Speech Signal Process.*, 1988, Vol. 34, p. 434.
- [27] J.L. Starck and A. Bijaoui, "Wavelets and multiresolution clean", *Proc. ESO Conf. High Res. Imag. by Interferometry II*, Garchingh, 1991.
- [28] A.N. Tikhonov and V.Y. Arsenin, *Solution of Ill-Posed Problems*, Winston, Washington, DC, 1977.
- [29] P.H. Van Cittert, *Z. Physik*, Vol. 69, 1931, p. 298.