

Deconvolution of astronomical images using the multiscale maximum entropy method

E. Pantin and J.-L. Starck

CEA/DSM/DAPNIA, 91191 Gif-sur-Yvette Cedex, France

Received August 18; accepted December 18, 1995

Abstract. — Following the ideas of Bontekoe et al. who noticed that the classical Maximum Entropy Method (MEM) had difficulties to efficiently restore high and low spatial frequency structure in an image at the same time, we use the wavelet transform, a mathematical tool to decompose a signal into different frequency bands. We introduce the concept of multi-scale entropy of an image, leading to a better restoration at all spatial frequencies. This deconvolution method is flux conservative and the use of a multiresolution support solves the problem of MEM to choose the α parameter, i.e. the relative weight between the goodness-of-fit and the entropy. We show that our algorithm is efficient too for filtering astronomical images. A range of practical examples illustrate this approach.

Key words: techniques: image processing

1. Introduction

In the field of image deconvolution, one of the most popular techniques is the Maximum Entropy Method (MEM) introduced initially by Burg (1967) and Ables (1974) for spectral analysis. Subsequently MEM was reformalized and improved a lot to provide an efficient technique of deconvolution by Skilling & Gull (1984), and Skilling & Bryan (1984). It is based on the idea that a given realization of a random variable (a non deterministic signal for instance) carries a certain amount of information measured by the entropy (Shannon 1948; Jaynes 1957). Thus, when trying to invert an ill-posed problem like the deconvolution of a signal, with the difficulties inherent in the presence of additive noise characteristic of an experimental signal, the entropy is used as a regularizing functional to constrain the solution, and give the simplest (in the sense of the amount of contained information) possible compatibility with the data. However the classical Maximum Entropy Deconvolution encounters technical problems such as finding an optimal value of α , i.e. the relative weight between the goodness-of-fit and the entropy. It has been observed also that a “low” value of α favours high frequency reconstructions, but gives a poorly regularized result, while a “high” α leads to a restored image with good regularization but in which the high frequency structures are under-reconstructed. Therefore, Bontekoe et al. (1994) have introduced the concept of Pyramid Maximum Entropy reconstruction which is a special application

of multi-channel maximum entropy image reconstruction techniques (Gull & Skilling 1991). They consider an image f as a weighted sum of a visible space pyramid of resolution $f = \sum_i f_i$, $i = 1, K$, which corresponds via a set of Intrinsic Correlation Functions (ICFs) to a hidden-space pyramid h_i , $i = 1, K$ on which the constraint of maximum entropy is applied. A major difficulty arises when summing the contributions corresponding to the different channels: the weights appear to be somewhat arbitrary or at least, difficult to determine theoretically. Another difficulty they encountered lies in the choice of the default constant (model) in each channel. Trying to advance further, we have reformulated this idea, using the appropriate mathematical tool to decompose a signal into channels of spectral bands, the wavelet transform. We show that the default value (model) at each wavelet scale is linked physically to the standard deviation of the noise present at this scale. Introducing the concept of multiscale entropy, we show that we minimize a functional depending on the desired solution regularized by minimizing the total amount of information contained at each resolution. We also use the concept of multiresolution support (Starck et al. 1995) which preserves the significant wavelet coefficients from a regularization, and leads to a fixed α for all types of images, removing the problem of its determination. Finally, we show that this method is very simple to use since there is no parameter to be determined by the user, and we give significant examples of deconvolution of blurred astronomical images showing the power of the method, especially to reconstruct weak structures and strong ones

simultaneously. We point out that one can derive a very efficient filtering method.

2. Image restoration using the MEM

We consider the data characterized by its intensity distribution $I(x, y)$, corresponding to the observation of an object $O(x, y)$ through an optical system. We assume that the process of degradation is linear with additive noise:

$$I(x, y) = (O * P)(x, y) + N(x, y) \quad (1)$$

where $P(x, y)$ is the point spread function (PSF) of the imaging system, and $N(x, y)$ is additive noise. We want to determine $O(x, y)$ knowing $I(x, y)$ and $P(x, y)$. For this paper, we shall consider Gaussian noise but in the case of Poisson, or Poisson plus Gaussian, we can reduce the problem to the Gaussian case (Murtagh et al. 1995) using the Anscombe transform (1948) and its generalization. We also consider that the number of “points” in the data is the same as the number of pixels in the original image O . This inverse problem is ill-posed because of the noise, making the previous system of equations lack a unique solution. The Maximum Entropy Method (MEM) which is a stochastic approach to the problem was initially suggested by Jaynes (1957). Among an infinite number of solutions to Eq. (1), it helps to choose the one that maximizes its entropy i.e. minimizes the amount of information contained in the solution. In other words, it is often said that this method gives the simplest solution compatible with the data, I .

Using Bayes’s theorem to evaluate the probability of the realization of the original image O , knowing the data I , we have

$$\text{Prob}(O|I) = \frac{\text{Prob}(I|O).\text{Prob}(O)}{\text{Prob}(I)} \quad (2)$$

$\text{Prob}(I|O)$ is the conditional probability of getting the data I given an original image O i.e. it represents the distribution of the noise. It is given, in the case of uncorrelated Gaussian noise with variance σ_I^2 by:

$$\text{Prob}(I|O) = \exp\left(-\sum_{\text{pixels}} \frac{(I - P * O)^2}{2\sigma_I^2}\right) \quad (3)$$

The Eq. (2) denominator is independent of O and is considered as a constant. $\text{Prob}(O)$ is the a priori distribution of the solution O . In the absence of any information on the solution O except its positivity, the maximum entropy principle suggests to take

$$\text{Prob}(O) = \exp(\alpha S(O)) \quad (4)$$

where $S(O)$ denotes the entropy of the image O .

Given the data, the maximum likelihood principle applies by maximizing $\text{Prob}(O|I)$, or equivalently by maximizing the product of the two previous equations. Taking the logarithm, we thus need to maximize

$$\ln(\text{Prob}(O|I)) = \alpha S(O) - \sum_{\text{pixels}} \frac{(I - P * O)^2}{2\sigma_I^2} \quad (5)$$

which is a linear combination of two terms: the entropy of the image, and a quantity corresponding to χ^2 in statistics measuring the discrepancy between the data and the predictions of the model.

The solution is found by minimizing

$$J(O) = \sum_{\text{pixels}} \frac{(I - P * O)^2}{2\sigma_I^2} - \alpha S(O) = \frac{\chi^2}{2} - \alpha S(O) \quad (6)$$

where α is a parameter that can be seen alternatively as a Lagrangian parameter or a value fixing the relative weight between the goodness-of-fit and the entropy S . Several entropy definitions have been proposed:

– Burg (1987):

$$S_b = -\sum_{\text{pixels}} \ln(O) \quad (7)$$

– Frieden (1975):

$$S_f = -\sum_{\text{pixels}} O \ln(O) \quad (8)$$

– Gull & Skilling (1991):

$$S_g = \sum_{\text{pixels}} O - m - O \ln(O/m) \quad (9)$$

The last definition of the entropy has the advantage of having a zero maximum when O equals m , but requires the concept of a model, m , which is in practice the value of the background. The determination of the α parameter is not an easy task and in fact it is a very serious problem that faces the ME method. In the historic MAX-ENT algorithm of Skilling and Gull, the choice of α is such that it must satisfy the *ad hoc* constraint $\chi^2 = N$ when the deconvolution is achieved, N being the number of degrees of freedom of the system i.e. the number of data points in the deconvolution problems. But this choice systematically leads to an under-fitting of the data (Titterton 1985) which is clearly apparent for imaging problems with little blurring. In reality, the χ^2 statistic is expected to vary in the range $N \pm \sqrt{2N}$ from one data realization to another. In the Quantified Maximum Entropy point of view (Skilling 1989), the optimum value of α is determined by including its probability $P(\alpha)$ in Bayes’ equation and then by maximizing the marginal probability of having α , knowing the data and the model m . In practice, a value of α which is too large gives a resulting image which is too regularized with a large loss of

resolution. A value which is too small leads to a poorly regularized solution showing unacceptable artifacts. Taking a flat model of the prior image softens the discontinuities which may appear paradoxical for astronomical images containing often some clusters of stars. Therefore the basic ME method appears to be not very appropriate for this kind of image which contains high and low spatial frequencies at the same time. Another point to be noted is a ringing effect of the MEM algorithm producing artifacts around bright sources. To solve these problems while still using the maximum entropy concept, some enhancements of the MEM method have been proposed. Noticing that neighbouring pixels of reconstructed images with MAXENT could have values differing a lot in expected flat regions (Charter 1990), Gull & Skilling introduced the concepts of hidden image h and Intrinsic Correlation Function C (ICF) (Gaussian or cubic spline-like) in the Preblur MAXENT algorithm. The ICF describes a minimum scale length of correlation in the desired image O which is achieved by assuming that

$$O = C * h \quad (10)$$

This corresponds to impose a minimum resolution in the solution O . Since the hidden space image h is not spatially correlated, this can be regularized by the entropy $S_g(h) = \sum h - m - h \ln(\frac{h}{m})$.

Since in astronomical images many scale lengths are present, the *Multi-channel Maximum Entropy Method* (Weir 1991, 1992) uses a set of ICFs having different scale lengths, each of them defining a channel. The visible-space image is now formed by weighted summing the visible-space image channels O_j :

$$O = \sum_{j=1}^K p_j O_j \quad (11)$$

where K is the number of channels. Like in Preblur MAXENT, each solution O_j is supposed to be the result of the convolution between a hidden image h_j with a low pass filter (ICF) C_j :

$$O_j = C_j * h_j \quad (12)$$

But such a method has several drawbacks:

1. The solution depends on the width of the ICFs (Bontekoe et al. 1994).
2. There is no rigorous way to fix the weights p_j for non unitary ICFs (Bontekoe et al. 1994).
3. The computation time increases linearly with the number of pixels.
4. The solution obtained depends on the choice of the models m_j ($j = 1..K$) which were chosen to be a constant c independent of the channel.

In 1993, Bontekoe et al. (1994) used a special application of this method they called Pyramid Maximum Entropy on IRAS data from the survey database. The pyramidal approach allows to have constant ICF width, and the computation time is reduced. It is demonstrated (Bontekoe et al. 1994) that all weights can be fixed ($p_j=1$ for each channel).

This method eliminates the three first drawbacks, and gives a better reconstruction of the sharp and smooth structures. But in addition to the last drawback, a new one is added: as the images O_j have different sizes (due to the pyramidal approach), the solution O is built by duplicating the pixels of the sub-images O_j of each channel. This procedure is known to produce artifacts due to the appearance of high frequencies incompatible with the real spectrum of the true image \hat{O} . However, this inconvenient can be easily overcome by duplicating the pixels before convolving with the ICF (Kester, private communication), or expanding the channels using linear interpolation. Thus the introduction of the ‘‘pyramid of resolution’’ has solved some problems and brought lots of improvements from the classic ME, but also raised other questions. In the following developments, we propose another way to use the information at different scales of resolution using the appropriate mathematical tool, the wavelet transform. We show that the problems encountered by Bontekoe et al. are overcome with this approach, especially the reconstruction of the object O which becomes natural. Furthermore, the wavelet transform gives a good framework for using a noise modeling. This modeling allows to preserve the significant wavelet coefficients (not due to the noise) from a regularization. The regularization becomes adaptive, depending on both the position in the image and the scale.

3. Formalism of multiscale MEM

3.1. Multiscale entropy

The concept of entropy following Shannon’s or Skilling and Gull’s definition is a global quantity calculated on the whole image O . It is not matched to quantify the distribution of the information at different scales of resolution. Therefore, we have proposed the concept of multi-scale entropy of a set of wavelet coefficients $\{w_j\}$ by

$$S_m(O) = \frac{1}{\sigma_I} \sum_{\text{scales } j} \sum_{\text{pixels}} \sigma_j (w_j(x, y) - m_j - |w_j(x, y)| \ln \frac{|w_j(x, y)|}{m_j}) \quad (13)$$

The multi-scale entropy is the addition of the entropy at each scale.

The coefficients w_j are wavelet coefficient, and we take the absolute value of w_j in that definition because the values of w_j can be positive or negative and a negative signal contains also some information in the wavelet transform.

The advantage of such a definition is the fact we can use previous works concerning the wavelet transform and image restoration (Starck & Murtagh 1994), (Murtagh et al. 1995). The noise behaviour has been studied in the wavelet transform (Starck & Bijaoui 1994) and we can estimate the standard deviation of the noise σ_j at the scale j from the standard deviation of the noise σ_I in the data (see in (Starck & Pantin 1995) how to have a robust estimation of σ_I).

These estimations can be naturally introduced in our models m_j

$$m_j = k_m \sigma_j \quad (14)$$

The model m_j at the scale j represents the value taken by a wavelet coefficient in the absence of any relevant signal and in practice, it must be a small value compared to any significant signal value. Following Gull & Skilling procedure, we take m_j as a fraction of the noise because the value of σ_j can be considered as a physical limit under which a signal cannot be distinguished from the noise ($k_m = \frac{1}{100}$).

The term σ_j which appears in Eq. (13) in front of the ‘‘classical entropy’’ of Skilling & Gull can be considered as a renormalization/weighting coefficient from one scale of resolution to another one.

Extensive literature exists on the wavelet transform and its applications (Chui 1992; Meyer 1992; Ruskai et al.; 1992; Starck 1993; Cohen et al. 1992). In this application, we will consider the discrete *à trous* algorithm (described in Holdschneider et al. 1989; Bijaoui et al. 1994) for its simplicity to use. An image $O(x, y)$ is decomposed into $w_j(x, y)j = 1, n_p$ scales (where n_p is the total number of wavelet scales) and a smooth image $c_{n_p}(x, y)$ and we can write:

$$O(x, y) = c_{n_p}(x, y) + \sum_{j=1}^{n_p} w_j(x, y) \quad (15)$$

Each scale $w_j(x, y)$ is obtained by calculating the difference between the last smoothed plane and the new one resulting from the application of a low-pass filter h on the last smoothed scale:

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

The signal difference w_j between two consecutive resolutions is:

$$w_j(k) = c_{j-1}(k) - c_j(k) \quad (17)$$

Thus, each wavelet scale w_j contains the structures of the image O having a given range of resolution/spatial frequencies. The above *à trous* algorithm has been discussed in terms of a single index, x , but is easily extendable to

two-dimensional space. The use of the B_3 spline leads to a convolution with a mask of 5×5 :

$$\begin{pmatrix} \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{1}{128} & \frac{3}{32} & \frac{6}{64} & \frac{3}{32} & \frac{1}{128} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \end{pmatrix}$$

In one dimension, this mask is: $(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$.

To facilitate computation, a simplification of this wavelet is to assume separability in the 2-dimensional case. In the case of the B_3 spline, this leads to a row by row convolution with $(\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16})$; followed by column by column convolution.

3.2. Multiscale entropy and multiresolution support

3.2.1. Multiresolution support definition

The multiresolution support was introduced in image restoration to hold the significant structures at each scale of resolution (i.e. in each wavelet scale). It consists in a cube of boolean images, each one corresponding to a wavelet scale. Such an image has the value TRUE if some information is present at a given location in the scale, and FALSE if not. The multiresolution support depends on the initial image I , the noise level, the chosen algorithm for the wavelet transform, and perhaps on some external constraints that can be imposed by the experimenter having some a priori knowledge of the solution.

The multiresolution support is generated using the following process:

- Calculate the wavelet transform of the image I using the chosen transform algorithm (the *à trous* algorithm in our case).
- Estimate the standard deviation of the noise in a wavelet scale j and deduce some significative thresholds (Starck et al. 1994).
- The multiresolution support M is then obtained by transforming the wavelet scales into binary planes using the rule:

$$M(j, x, y) = \begin{cases} 1 & \text{if } w_j(x, y) \geq k\sigma_j \\ 0 & \text{if } w_j(x, y) < k\sigma_j \end{cases} \quad (18)$$

The parameter k fixes the threshold level. When considering a Gaussian and centered (with mean equal to 0) noise, the probability of having a realization of the noise greater than 3σ is about 10^{-3} . This value is considered as a practical limit for a decision to distinguish between the noise and a real signal. Thus a value $k = 3$ is generally used.

3.2.2. Multiresolution support and multiscale entropy

If the definition 13 is used for the multi-scale entropy, the regularization acts on the whole image. We want to fully reconstruct significant structures, without imposing strong regularization, while eliminating efficiently the noise. Thus the introduction of the multiresolution support in another definition of the multi-scale entropy leads to a functional that answers these requirements:

$$S_{\text{ms}}(O) = \frac{1}{\sigma_I} \sum_{\text{scales } j} \sum_{\text{pixels}} A(j, x, y) \sigma_j (w_j(x, y) - m_j - |w_j(x, y)| \ln \frac{|w_j(x, y)|}{m_j}) \quad (19)$$

The A function of the scale j and the pixels (x, y) is $A(j, x, y) = 1 - M(j, x, y)$ i.e. the reciprocal of the multiresolution support M . In order to avoid some discontinuities in the A function created by such a coarse threshold of $3 \sigma_j$, one may possibly impose some smoothness by convolving it with a B-spline function with a FWHM varying with the scale j .

The degree of regularization will be determined at each scale j , and at each location (x, y) , by the value of the function $A(j, x, y)$: if $A(j, x, y)$ has a value near 1 then we have strong regularization; and it is weak when A is around 0.

The entropy S_s measures the amount of information only at scales and in areas where we have a low signal-to-noise ratio. We will show in the next section how these notions can be taken together to yield efficient methods for filtering and image deconvolution.

4. Deconvolution using multiscale entropy

4.1. Method

We assume that the blurring process of an image is linear. In our tests, the PSF was space invariant but the method can be extended to space-variant PSFs.

As in the ME method, we will minimize a functional of O , but considering an image as a pyramid of different scales of resolution in which we try to maximize its contribution to the multiscale entropy. The functional to minimize is

$$J(O) = \sum_{\text{pixels}} \frac{(I - P * O)^2}{2\sigma_I^2} - \alpha S_{\text{ms}}(O) \quad (20)$$

Then the final difficulty lies in finding an algorithm to minimize the functional $J(O)$. We have used the iterative “one step gradient” method due to its simplicity.

The solution is found by computing the gradient

$$\nabla(J(O)) = -P^* * \frac{(I - P * O)}{\sigma_I^2}$$

$$+ \alpha \frac{1}{\sigma_I} \sum_{\text{scale } j} [A(j) \sigma_j \text{sgn}(w_j^{(O)}) \ln(\frac{|w_j^{(O)}|}{m_j})] * \psi_j^* \quad (21)$$

where $\psi_j(x, y) = \frac{1}{2^j} \psi(\frac{x}{2^j}, \frac{y}{2^j})$, and ψ is the wavelet function corresponding to the *à trous* algorithm (see (Starck & Bijaoui 1994) for more details about this algorithm).

Then the “one step gradient” algorithm gives the iterative scheme:

$$O^{n+1} = O^n - \gamma \nabla(J(O^n)) \quad (22)$$

Note that the second part of the Eq. (21) has always a null mean value due to the fact that each scale is convolved by ψ_j^* (and ψ_j^* has a null mean value due to the admissibility condition of the wavelet function). Then the flux is not modified in the object O when applying the iterative scheme (Eq. (22)).

The positivity of the restored image can be ensured during the iterative process of functional minimization, by applying a simple positivity constraint (threshold) to the intermediate solution O^n . The iterative process is stopped when the standard deviation of the residuals doesn’t show significant change (relative variation $\leq 10^{-3}$, and we check that the χ^2 value is in the range $N \pm \sqrt{2N}$ (it was always the case in all the experiments, if not it would mean that σ_I value is wrong).

4.2. Choice of the α parameter

In the classic ME, the α parameter quantifies the relative weight between the goodness-of-fit, or chi-square, and the degree of smoothness introduced by the entropy. This parameter is generally constant over the whole image and therefore depends on the data (signal+noise). In our case, the degree of regularization level applied on the non significant structures at each wavelet scale j is controlled by the term $\frac{\alpha}{\sigma_I} A(j, x, y)$, and depends therefore on both the scale j and the location (x, y) . The regularization is performed only at scales and positions where no signal is detected ($A(j, x, y) = 1$). Then the α parameter has not the same importance than in the classical MEM: α has only to be high enough to regularize the solution at positions and scales where no signal has been detected. We found experimentally that

$$\alpha = 0.5 * \max(\text{PSF}) / \sigma_I \quad (23)$$

produces good results, and for any kind of images.

4.3. Experiments

4.3.1. Application 1

We have tested our algorithm with simulated data. The simulated image contains an extended object, and several smaller sources (Fig. 1, upper left). It has been convolved by a gaussian PSF ($\sigma = 2$), and Gaussian noise has been

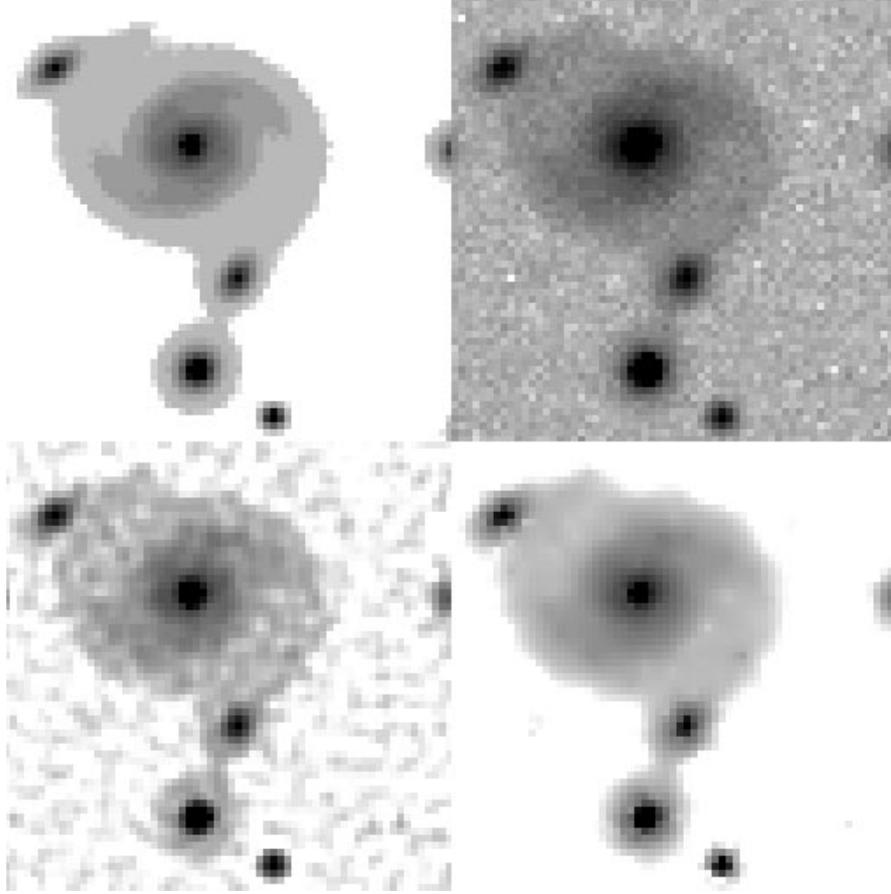


Fig. 1. Simulated object (upper left), blurred image (upper right), deconvolved image by MEM (lower left), and deconvolved image by Multiscale MEM (lower right)

added (Fig. 1, upper right). The ratio between the maximum in the convolved image and the standard deviation of the noise is 50. The results of the deconvolution by MEM and multiscale MEM are shown in Fig. 1 left and right. The multiscale MEM method leads to a better regularization, and the final resolution in the deconvolved image is better (Fig. 1).

4.3.2. Application 2

In order to give quantitative results on the possibilities of the multiscale MEM, we have tested it on a simulated image of the ISOCAM mid-infrared camera put on the infrared satellite ISO. The simulation consists in a faint extended object (galaxy, integrated flux=1 Jy) near a bright star (point-like: 10 Jy) (Fig. 2 upper left). It has been blurred using the $1.5''$ pixel field of view PSF, and noise was added (Fig. 2 upper right): the two objects are mixed and the galaxy is barely detectable. After deconvolution using the multiscale MEM (see Fig. 2 bottom right), the two objects are separated. The restored star has a flux of 10.04 Jy, and the SNR of the reconstructed image is 22.4

dB ($\text{SNR} = 10 \log_{10}(\frac{\sigma_{\text{Signal}}^2}{\sigma_{\text{Residuals}}^2})$). These results show that multiscale MEM is very efficient when we compare it to other methods (see Tab. 1), and prove the reliability of the photometry after deconvolution. It is clear in this example that photometry cannot always be done directly from the data, and a deconvolution is often necessary, especially when objects are mixed.

4.3.3. Application 3

We have tested our deconvolution method on astronomical 64 by 64 pixels images obtained with an mid-infrared camera: TIMMI placed on the 3.6 ESO telescope (Chile). The object studied is the β Pictoris dust disk. The image was obtained by integrating 5h on-source. The raw image has a peak signal to noise ratio of 80. It is strongly blurred by a combination of seeing, diffraction (0.7 arcsec on a 3 m class telescope) and additive Gaussian noise. The initial disk shape in the original image has been lost after the convolution with the PSF. Thus, we need to deconvolve them to get the best information on this object i.e. the exact profile and thickness of the disk and compare

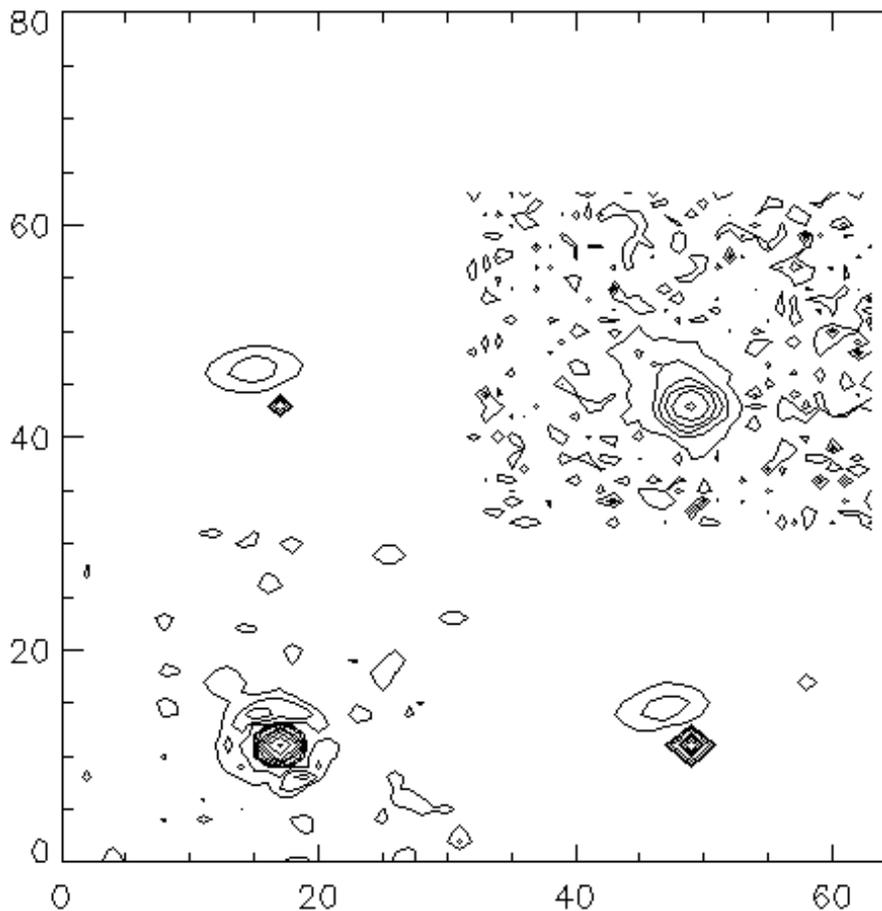


Fig. 2. Simulated ISOCAM image (upper left), blurred image using the 1.5 pixel field of view PSF (upper right), the deconvolved image by MEM (lower left), and deconvolved image by multiscale MEM (lower right)

it to models of thermal dust emission (Lagage & Pantin 1994). We used the multiscale ME method to deconvolve this image. The algorithm took about thirty iterations to converge. The deconvolved image (Fig. 3) shows that the disk is extended at $10 \mu\text{m}$ and asymmetrical (the right side is more extended than the left side). We have compared our method to the standard Richardson-Lucy (Richardson 1972), (Lucy 1974) algorithm which shows poor regularization (see Fig. 3, at upper right) and an inability to restore faint structures; and also to the classical MEM. The deconvolved image using the multiscale ME method proves to be more efficient for regularizing than the other standard methods, and leads to a good reconstruction of the faintest structures of the dust disk.

5. Conclusion

In the field of signal deconvolution, the maximum entropy method gave an attractive way of regularization and improved a lot the existing techniques. However, several difficulties were remaining: the most important is perhaps the inability to find an optimal regularizing parameter (α) to

reconstruct efficiently the high and low spatial frequencies at the same time while having a good regularization.

Compared to the classical MEM, our method has a fixed α parameter and there is no need to determine it: it is the same for every image. The iterative scheme always converged in our experiments. Furthermore, this new method is flux-conservative and thus reliable photometry can be done on the deconvolved image. In Bontekoe et al. 1994, it was noticed that the “models” in the multi-channel MEM deconvolution should be linked to a physical quantity. We have shown here that this is the case since it is a fraction of the standard deviation of the noise at a given scale of resolution. Bontekoe et al. have opened a new way of thinking in terms of multiresolution decomposition, but they did not use the appropriate mathematical tool which is the wavelets decomposition. Using such an approach, we have proven that many problems they encountered are naturally solved. The result is an efficient “easy to use” algorithm since the user has no parameter to supply.

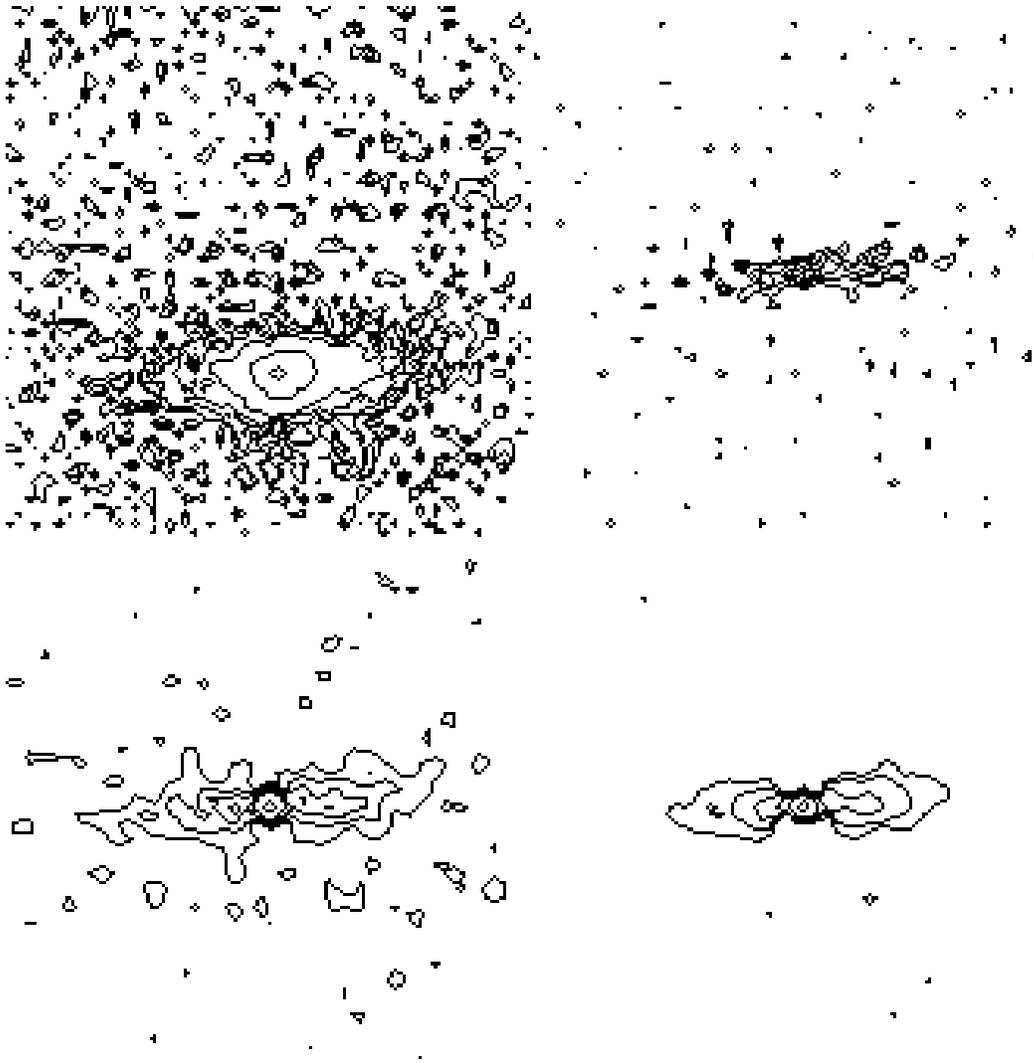


Fig. 3. Beta Pictoris: raw image (upper left) and deconvolved images using: Richardson-Lucy's method (upper right), classical MEM (lower left), and multiscale MEM (lower right)

Acknowledgements. We wish to thank F. Murtagh for useful discussions and comments. This paper was improved by the extensive comments of the referees, D.J.M. Kester and Tj.R. Bontekoe.

Appendix 1: Multiscale entropy applied to filtering

Filtering using the multiscale entropy leads to the minimization of:

$$J(\tilde{I}) = \sum_{\text{pixels}} \frac{(I - \tilde{I})^2}{2\sigma_I^2} - \alpha S_{\text{ms}}(\tilde{I}) \quad (24)$$

where I and \tilde{I} are the noisy and filtered image, α fixed experimentally to $0.5/\sigma_I$, σ_I the standard deviation of the noise in the data I .

Then, the gradient of the functional with respect to \tilde{I} must be calculated:

$$\begin{aligned} \nabla(J(\tilde{I})) &= -\frac{(I - \tilde{I})}{\sigma_I^2} \\ &+ \frac{\alpha}{\sigma_I} \sum_{\text{scales } j} [A(j)\sigma_j \text{sgn}(w_j^{(\tilde{I})}) \log(\frac{|w_j^{(\tilde{I})}|}{m_j})] * \psi_j^* \end{aligned} \quad (25)$$

where $\psi_j(x, y) = \frac{1}{2^j} \psi(\frac{x}{2^j}, \frac{y}{2^j})$, and ψ is the wavelet function corresponding to the *à trous* algorithm.

Then the “one step gradient” algorithm gives the iterative scheme:

$$\tilde{I}^{n+1} = \tilde{I}^n - \gamma \nabla(J(\tilde{I}^n)) \quad (26)$$

where γ is its step.

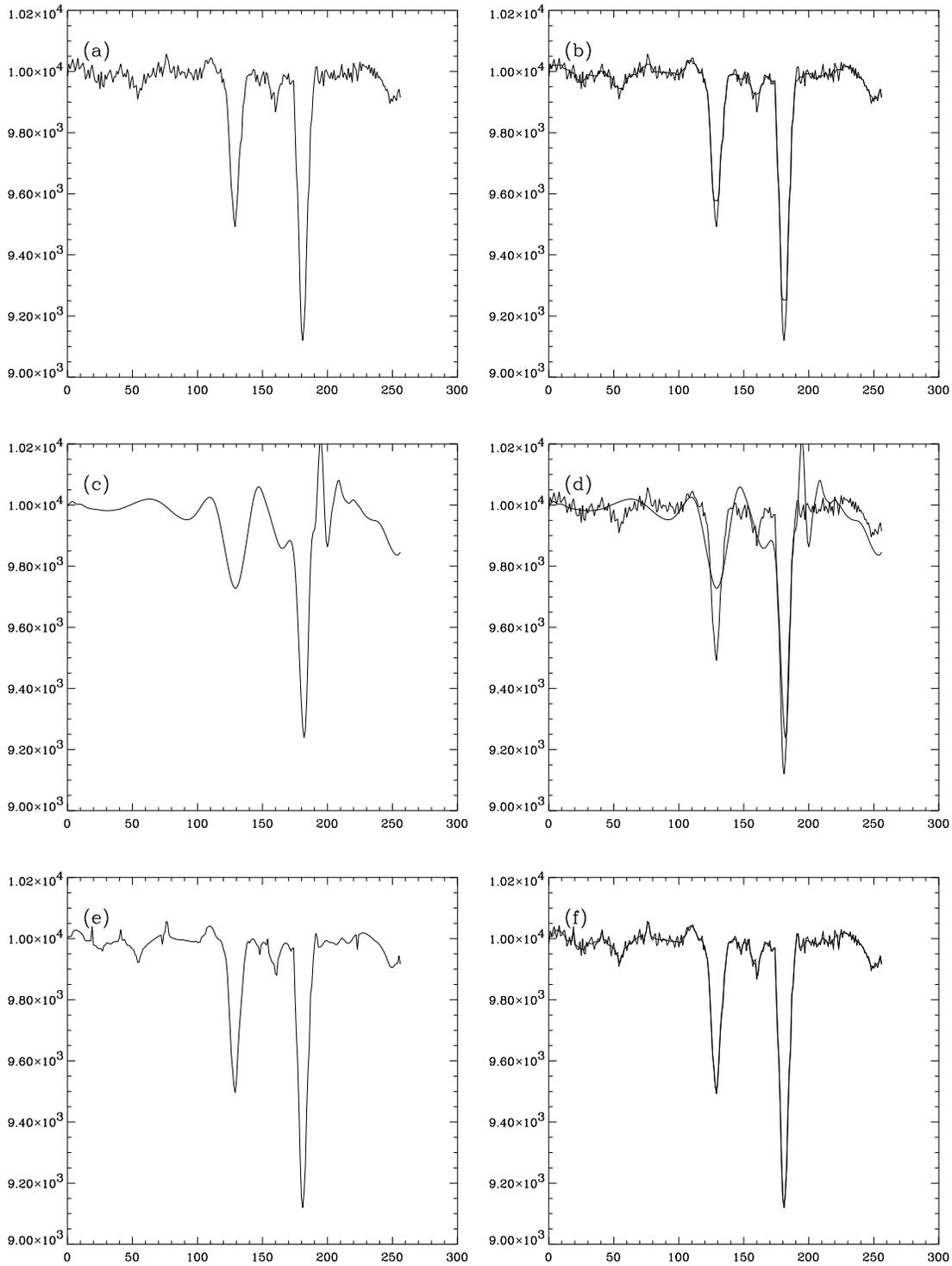


Fig. 4. **a)** original noisy spectrum; **b)** original spectrum and filtered spectrum by the median superimposed. **c)** filtered spectrum (using Daubechies coefficient 8, and Donoho and Johnstone “universal” thresholding); **d)** both **a)** and **c)** superimposed. **e)** filtered spectrum by multiscale MEM; **f)** both **a)** and **f)** superimposed

Table 1. Quantitative results extracted from three images deconvolved by Lucy’s method, MEM, and multiscale MEM. The first line gives the flux of the star. The second one, the integrated flux in the extended object, and the last line, the signal to noise ratio of the deconvolved images. Since the two objects are mixed in the degraded image, it is impossible to attribute a photometric measure to each one

	Original image	Data	Lucy	MEM	Multiscale MEM
star Flux (Jy)	10		4.27	14.1	10.04
ext. object Flux (Jy)	1		0.33	1.33	0.94
SNR (dB)	∞	1.03	4.51	4.45	22.4

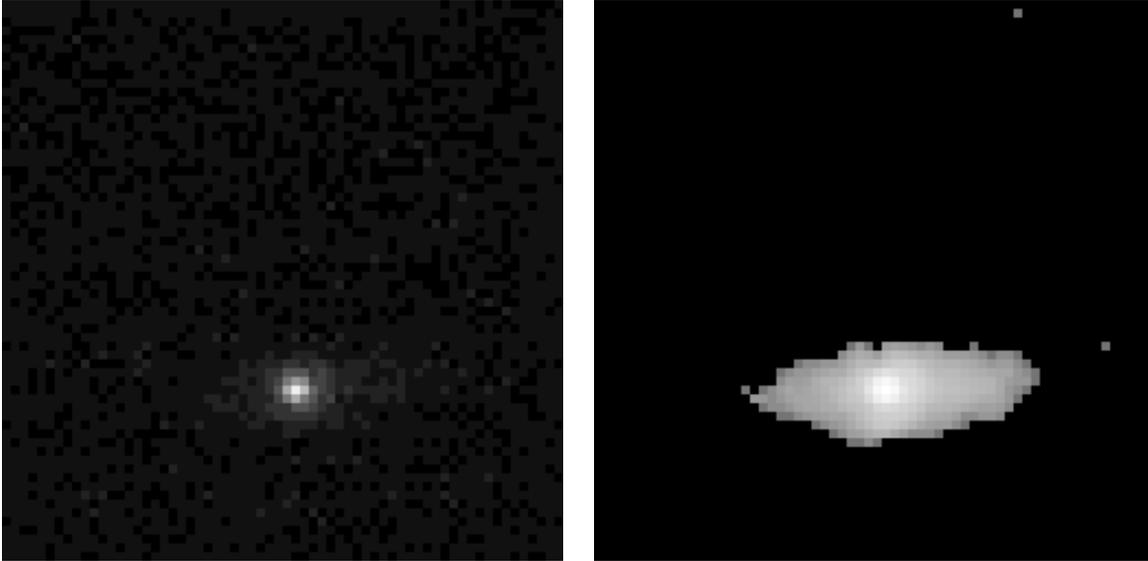


Fig. 5. β Pictoris dust disk: Raw image (left) and filtered image (right) using multiscale ME

5.1. Experiments

Figure 4a shows a noisy spectrum. Upper right (4b), we see the original and filtered (by the median) spectrum. The median filtering is relatively efficient but it does not restore well the lines, i.e. some interesting information have been suppressed during the filtering. Lower left (e), we see the same spectrum, but filtered by the multiscale maximum entropy method, and right (f), both (a) and (e) superimposed. The filtering is efficient, the lines are kept, and no artifacts appears. To illustrate the damage that can result from another wavelet transform, and another noise suppression policy, the middle (c) version shows the result of applying Daubechies’s coefficient 8 (Cohen et al. 1992), a compactly-supported orthonormal wavelet. This was followed by thresholding based on estimated variance of the coefficients (Donoho & Johnstone 1993), but not taking into account the image’s noise properties as we have done (see Nason 1993). One sees immediately that a problem- (or image-) driven choice of wavelet and filtering strategy is indispensable.

We have also tested our filtering method on a mid-infrared image of the β Pictoris dust disk described in

section “deconvolution experiments”, but obtained with only 1h of integration time (see Fig. 5). The peak signal to noise ratio is around 30. After filtering, the disk appears clearly. For detection of faint structures (the disk here), one can calculate that the application of such a filtering method on this image provides a gain of observing time of a factor around 60 (in the case of gaussian additive noise leading to a signal to noise ratio

varying like the square root of the integration time).

Figure 6 shows a profile of the object (crosses). The profiles of the filtered images of the dust disk using multiscale maximum entropy filtering (plain line) and Wiener filtering (dots) are superimposed. Contrary to the Wiener filtering, the multiscale maximum entropy algorithm doesn’t degrade the resolution, while filtering efficiently.

References

- Ables J.G., 1974, A&AS 15, 383
 Anscombe F.J., 1948, “The transformation of Poisson, binomial and negative-binomial data”, Biometrika 15, 246

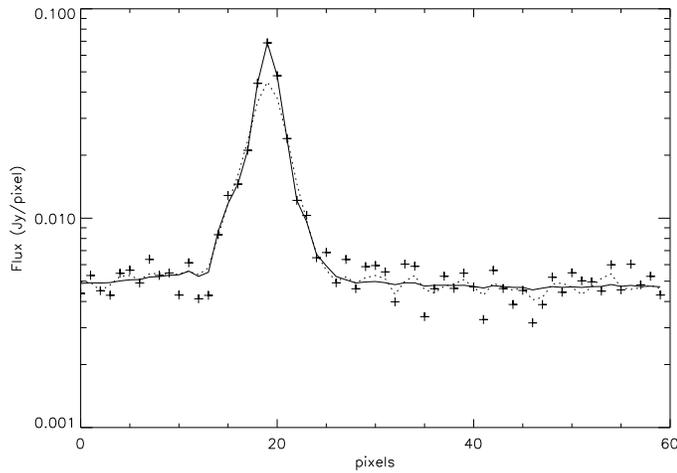


Fig. 6. Raw image profile of β Pictoris dust disk (crosses). The profiles of the filtered images using multiscale maximum entropy filtering (plain line) and Wiener filtering (dots) are superimposed

Bijaoui A., Starck J.L., Murtagh F., 1994, "Restauration des Images Multi-Echelles par l'Algorithme à Trous", *Traitement du Signal* 3, 11

Bontekoe Tj.R., Koper E., Kester D.J.M., 1994, "Pyramid Maximum Entropy Images of IRAS Survey Data", *A&A* 294, 1037-1053

Burg J.P., 1967, *Ann. Meet. Int. Soc. Explor. Geophys.* Reprinted in *Modern Spectral Analysis*, 1978. In: Childers D.G. (ed.). New York: IEEE press, pp. 34-41

Charter M.K., 1990, in: Fougere P.F. (ed.) *Maximum Entropy and Bayesian Methods*, Dartmouth 1989. Kluwer, Dordrecht, p. 325

Chui C.H., 1992, *Wavelet Analysis and its Application*. Academic Press, New York

Cohen A., Daubechies I., Feauveau J.C., 1992, "Biorthogonal bases of compactly supported wavelets", *Comm. Pure Appl. Math.* 45, 485

Donoho D.L., Johnstone I.M., 1993, "Ideal spatial adaptation by wavelet shrinkage", Stanford University, Technical Report 400, (available by anonymous ftp from play-fair.stanford.edu:/pub/donoho)

B.R. Frieden, 1975, "Image Enhancement and Restoration", *Picture Processing and Digital Filtering*. In: Huang T.S. (ed.). Springer-Verlag Berlin, pp. 177-249

Gull S.F., Skilling J., 1991, *MEMSYS5 User's Manual*

Holdschneider M., Kronland-Martinet R., Morlet J.,

Tchamitchian Ph., 1989, "A real-time algorithm for signal analysis with the help of the wavelet transform", in *Wavelets*. In: Combes J.M., Grossmann A. and Tchamitchian Ph. (eds.). Springer-Verlag, Berlin, pp. 286-297

Jaynes E.T., 1957, *Phys. Rev.* 106, 620

Lagage P.O., Pantin E., 1994, *Nat* 369, 628

Meyer Y., 1992, *Ondelettes: Algorithmes et Applications*, Collection Acquis Avancés de l'Informatique

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

Murtagh F., Starck J.L., Bijaoui A., 1995, "Image Restoration with Noise Suppression Using a Multiresolution Support", *A&AS* 112, 179

Nason G.P., 1993, "The discrete wavelet transform in S", Version 2.1, 1 (Software and documentation available from Statlib repository, <http://lib.stat.cmu.edu/>)

Richardson W.H., 1972, *J. Opt. Soc. Amer.* 62, 55

Ruskai M.B., Beylkin G., Coifman R., Daubechies I., Mallat S., Meyer Y., Raphael L., 1992, *Wavelets and their Applications*. Jones and Barlett

Shannon C.E., *Bell System Tech J.* 27, 379

Skilling J., Gull S.F., 1984, *SIAM Proc. Am. Math. Soc.* 14, 167

Skilling J., Bryan R.K., 1984, *MNRAS* 211, 111-24

Skilling J., 1989, in: Skilling, J. (ed.) *Maximum Entropy and Bayesian Methods*, Cambridge 1988. Kluwer, Dordrecht, p. 45

Starck J.L., 1993, in *MIDAS Manual*, Release 93NOV, Image Processing Group, European Southern Observatory, Garching bei München, 1993

Starck J.L., Bijaoui A., 1994, "Filtering and deconvolution by the wavelet transform", *Signal Proc.* 35, 195

Starck J.L., Murtagh F., 1994, "Image restoration with noise suppression using the wavelet transform", *A&A* 288, 343

Starck J.L., Murtagh F., Bijaoui A., 1995, "Multiresolution Support Applied to Image Filtering and Deconvolution", in *CVIP: Graph. Models Image Proc.* 57, 420

Starck J.L., Pantin E., 1995, "Multiscale Maximum Entropy Images Restoration", *First Converging Computing Methodologies in Astronomy Workshop on Vision Modeling and Information Coding*, Nice

Titterington D.M., 1985, *A&A* 144, 381

Weir N., 1991, "Application of Maximum Entropy Techniques to HST Data", 3th ESO/ST-ECF Data Analysis Workshop, Garching

Weir N., in: Worrall D.M., Biemesderfer C., Barnes J. (eds.) *Astronomical Data Analysis Software and System 1*, PASPC 25, 186