# The Starlet Transform in Astronomical Data Processing: Application to Source Detection and Image Deconvolution

J.-L. Starck [*], F. Murtagh [†], M. Bertero [‡]

August 5, 2011

# Contents

[*]J.-L. Starck is with the CEA, Laboratoire AIM, CEA/DSM-CNRS-Université Paris Diderot, CEA, IRFU, Service d'Astrophysique, Centre de Saclay, F-91191 Gif-Sur-Yvette cedex, France

[†]F. Murtagh is with Science Foundation Ireland, Wilton Park House, Wilton Place, Dublin 2, Ireland and the Department of Computer Science, Royal Holloway University of London, Egham TW20 0EX, UK

[‡]M. Bertero is with DISI, Università di Genova, Via Dodecaneso 35, 16146 Genova, Italy

**Abstract**

We begin with traditional source detection algorithms in astronomy. We then introduce the sparsity data model. The starlet wavelet transform serves as our main focus in this article. Sparse modeling, and noise modeling, are described. Applications to object detection and characterization, and to image filtering and deconvolution, are discussed. The multiscale vision model is a further development of this work, which can allow for image reconstruction when the point spread function is not known, or not known well. Bayesian and other algorithms are described for image restoration. A range of examples is used to illustrate the algorithms.

# 1   Introduction

Data analysis is becoming more and more important in astronomy. This can be explained by detector evolution, which concerns all wavelengths. In the 1980s, CCD (charge coupled device) images had a typical size of $512 \times 512$ pixels, while astronomers now have CCD mosaics with $16000 \times 16000$ pixels or even more. At the same time, methods of analysis have become much more complex, and the human and financial efforts to create and process the data can sometimes be of the same order as for the construction of the instrument itself. As an example, for the ISOCAM camera of the Infrared Space Observatory (ISO), the command software of the instrument, and the online and offline data processing, required altogether 70 person years of development, while 200 person years were necessary for the construction of the camera. The data analysis effort for the PLANCK project is even larger. Furthermore, the quantity of outputs requires the use of databases, and in parallel sophisticated tools are needed to extract ancillary astrophysical information, generally now through the web. From the current knowledge, new questions emerge, and it is necessary to proceed to new observations of a given object or a part of the sky. The acquired data need to be calibrated prior to useful information for the scientific project being extracted.

Data analysis acts during the calibration, the scientific information extraction process, and the database manipulation. The calibration phase consists of correcting various instrumental effects, such as the dark current (i.e. in the absence of any light, the camera does not return zero values, and the measured image is called the dark image, and needs to be subtracted from any observation), or the flat field correction (i.e. for uniform light, the detector does not return the same value for each pixel, and a normalization needs to be performed by dividing the observed image by the "flat" image). Hence, it is very important to know well the parameters of the detector (flat field image, dark image, etc.), because any error on the these parameters will propagate to the measurements. Other effects can also be corrected during this phase, such as the removal of the cosmic ray impacts or the field distortion (the pixel surface for each pixel does not correspond to the same surface on the sky). Depending on the knowledge of the instrument, each of these tasks may be more or less difficult.

Once the data are calibrated, the analysis phase can start. Following the scientific objectives, several kinds of information can be extracted from the data, such as for example the detection of stars and galaxies, the measurement of their position, intensity, and various morphological parameters. The results can be compared to existing catalogs, obtained from previous observations. It is obviously impossible to cite all operations we may want to carry

through on an astronomical image, and we have just mentioned the most common. In order to extract the information, it is necessary to take into account noise and point spread function. Noise is the random fluctuation which is added to the CCD data, and comes partially from the detector, and partially from the data. In addition to the errors induced by the noise on the different measurements, noise also limits the detection of objects, and can be responsible for false detections. The point spread function is manifested in how the image of a star (for example) is generally spread out on several pixels, caused by the atmosphere's effect on the light path. The main effect is a loss of resolution, because two sufficiently close objects cannot be separated. Once information has been extracted, such details can be compared to our existing knowledge. This comparison allows us to validate or reject our understanding of the universe.

In this chapter, we will discuss in detail how to detect objects in astronomical images and how to take into account the point spread function though the deconvolution processing.

## Source Detection

As explained above, source (i.e., object) extraction from images is a fundamental step for astronomers. For example, to build catalogs, stars and galaxies must be identified and their position and photometry must be estimated with good accuracy. Catologs comprise a key result of astronomical research. Various methods have been proposed to support the construction of catalogs. One of the now most widely used software packages is SExtractor [5], that is capable of handling very large images. A standard source detection approach, such as in SExtractor, consists of the following steps:

1. Background estimation.

2. Convolution with a mask.

3. Detection.

4. Deblending/merging.

5. Photometry.

6. Classification.

These different steps are described in the next section. Astronomical images contain typically a large set of point-like sources (the stars), some quasi point-like objects (faint galaxies, double stars), and some complex and diffuse structures (galaxies, nebulae, planetary stars, clusters, etc.). These objects are often hierarchically organized: a star in a small nebula, itself embedded in a galaxy arm, itself included in a galaxy, and so on.

The standard approach, which is presented in detail in section 2, presents some limits, when we are looking for faint extended objects embedded in noise. Fig. 1 shows a typical example where a faint extended object is under the detection limit. In order to detect such objects, more complex data modeling needs to be defined. Section 3 presents another approach to model and represent astronomical data, by using a sparse model in a wavelet dictionary. A specific wavelet transform, called the *starlet transform* or the isotropic undecimated wavelet transform, is presented. Based on this new modeling, several approaches are proposed in sections 4 and 5.
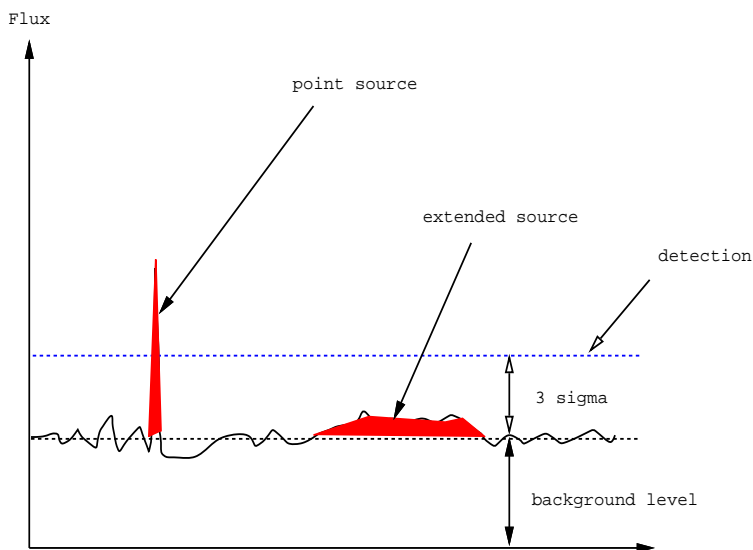
Figure 1: Example of astronomical data: a point source and an extended source are shown, with noise and background. The extended object, which can be detected by eye, is undetected by a standard detection approach.

## 2 Standard Approaches to Source Detection

We describe here the most popular way to create a catalog of galaxies from astronomical images.

**The Traditional Data Model**

The observed data $Y$ can be decomposed into two parts, the signal $X$, and the noise $N$:

$$Y[k,l] = X[k,l] + N[k,l] \tag{1}$$

The imaging system can also be considered. If it is linear, the relation between the data and the image in the same coordinate frame is a convolution:

$$Y[k,l] \quad = \quad (HX)[k,l] + N[k,l] \tag{2}$$

where $H$ is the matrix related to the Point Spread Function (PSF) of the imaging system.

In most cases, objects of interest are superimposed on a relatively flat signal $B$, called *background signal*. The model becomes:

$$Y[k,l] = (HX)[k,l] + B[k,l] + N[k,l] \tag{3}$$

**PSF Estimation**

The PSF $H$ can be estimated from the data, or from an optical model of the imaging telescope. In astronomical images, the data may contain stars, or one can point towards a reference star

in order to reconstruct a PSF. The drawback is the "degradation" of this PSF because of unavoidable noise or spurious instrument signatures in the data. So, when reconstructing a PSF from experimental data, one has to reduce very carefully the images used (background removal for instance). Another problem arises when the PSF is highly variable with time, as is the case for adaptive optics (AO) images. This means usually that the PSF estimated when observing a reference star, after or before the observation of the scientific target, has small differences from the perfectly correct PSF.

Another approach consists of constructing a synthetic PSF. Various studies [8, 29, 16, 30] have suggested a radially symmetric approximation to the PSF:

$$P(r) \propto (1 + \frac{r^2}{R^2})^{-\beta} \tag{4}$$

The parameters $\beta$ and $R$ are obtained by fitting the model with stars contained in the data.

In the case of AO systems this model can be used for the tail of the PSF (the so-called *seeing* contribution), while in the central region the system provides an approximation of the diffraction-limited PSF. The quality of the approximation is measured by the Strehl ratio (SR), which is defined as the ratio of the observed peak intensity in the image of a point source to the theoretical peak intensity of a perfect imaging system.

## Background Estimation

The background must be accurately estimated, otherwise it will introduce bias in flux estimation. In [6, 21], the image is partitioned into blocks, and the local sky level in each block is estimated from its histogram. The pixel intensity histogram $p(Y)$ is modeled using three parameters, the true sky level $B$, the RMS (root mean square) noise $\sigma$, and a parameter describing the asymmetry in $p(Y)$ due to the presence of objects, and is defined by [6]:

$$p(Y) = \frac{1}{a} \exp(\sigma^2/2a^2) \exp\left[-(Y - B)/a\right] \operatorname{erfc}\left(\frac{\sigma}{a} - \frac{(Y - B)}{\sigma}\right) \tag{5}$$

Median filtering can be applied to the 2D array of background measurements in order to correct for spurious background values. Finally the background map is obtained by a bi-linear or a cubic interpolation of the 2D array. The blocksize is a crucial parameter. If it is too small, the background estimation map will be affected by the presence of objects, and if too large it will not take into account real background variations.

In [11, 5], the local sky level is calculated differently. A 3-sigma clipping around the median is performed in each block. If the standard deviation is changed by less than 20% in the clipping iterations, the block is uncrowded, and the background level is considered to be equal to the mean of the clipped histogram. Otherwise, it is calculated by $c_1 \times \operatorname{median} - c_2 \times \operatorname{mean}$, where $c_1 = 3, c_2 = 2$ in [11], and $c_1 = 2.5, c_2 = 1.5$ in [5]. This approach has been preferred to histogram fitting for two reasons: it is more efficient from the computation point of view, and more robust with small sample size.

## Convolution

In order to optimize the detection, the image must be convolved with a filter. The shape of this filter optimizes the detection of objects with the same shape. Therefore, for star detection,

the optimal filter is the PSF. For extended objects, a larger filter size is recommended. In order to have optimal detection for any object size, the detection must be repeated several times with different filter sizes, leading to a kind of multiscale approach.

## Detection

Once the image is convolved, all pixels $Y[k,l]$ at location $(k,l)$ with a value larger than $T[k,l]$ are considered as significant, i.e. belonging to an object. $T[k,l]$ is generally chosen as $B[k,l] + K\sigma$, where $B[k,l]$ is the background estimate at the same position, $\sigma$ is the noise standard deviation, and $K$ is a given constant (typically chosen between 3 and 5). The thresholded image is then segmented, i.e. a label is assigned to each group of connected pixels. The next step is to separate the blended objects which are connected and have the same label.

An alternative to the thresholding/segmentation procedure is to find peaks. This is only well-suited to star detection and not to extended objects. In this case, the next step is to merge the pixels belonging to the same object.

## Deblending/Merging

This is the most delicate step. Extended objects must be considered as single objects, while multiple objects must be well separated. In SExtractor, each group of connected pixels is analyzed at different intensity levels, starting from the highest down to the lowest level. The pixel group can be seen as a surface, with mountains and valleys. At the beginning (highest level), only the highest peak is visible. When the level decreases several other peaks may become visible, defining therefore several structures. At a given level, two structures may become connected, and the decision whether they form only one (i.e. merging) or several objects (i.e. deblending) must be taken. This is done by comparing the integrated intensities inside the peaks. If the ratio between them is too low, then the two structures must be merged.

## Photometry and Classification

### Photometry.

Several methods can be used to derive the photometry of a detected object [6, 22]. Adaptive aperture photometry uses the first image moment to determine the elliptical aperture from which the object flux is integrated. Kron [22] proposed an aperture size of twice the radius of the first image moment radius $r_1$, which leads to recovery of most of the flux ($> 90$ %). In [5], the value of $2.5r_1$ is discussed, leading to loss of less than 6% of the total flux. Assuming that the intensity profiles of the faint objects are Gaussian, flux estimates can be refined [26, 5]. When the image contains only stars, specific methods can be developed which take the PSF into account [14, 33].

### Star-galaxy separation.

In the case of star–galaxy classification, following the scanning of digitized images, Kurtz [23] lists the following parameters which have been used:

1. mean surface brightness;

2. maximum intensity, area;

3. maximum intensity, intensity gradient;

4. normalized density gradient;

5. areal profile;

6. radial profile;

7. maximum intensity, $2^{nd}$ and $4^{th}$ order moments, ellipticity;

8. the fit of galaxy and star models;

9. contrast versus smoothness ratio;

10. the fit of a Gaussian model;

11. moment invariants;

12. standard deviation of brightness;

13. $2^{nd}$ order moment;

14. inverse effective squared radius;

15. maximum intensity, intensity weighted radius;

16. $2^{nd}$ and $3^{rd}$ order moments, number of local maxima, maximum intensity.

References for all of these may be found in the cited work. Clearly there is room for differing views on parameters to be chosen for what is essentially the same problem. It is of course the case also that aspects such as the following will help to orientate us towards a particular set of parameters in a particular case: the quality of the data; the computational ease of measuring certain parameters; the relevance and importance of the parameters measured relative to the data analysis output (e.g. the classification, or the planar graphics); and, similarly, the importance of the parameters relative to theoretical models under investigation.

**Galaxy morphology classification.**

The inherent difficulty of characterizing spiral galaxies especially when not face-on has meant that most work focuses on ellipticity in the galaxies under study. This points to an inherent bias in the potential multivariate statistical procedures. In the following, it will not be attempted to address problems of galaxy photometry per se [13, 35], but rather to draw some conclusions on what types of parameters or features have been used in practice.

From the point of view of multivariate statistical algorithms, a reasonably homogeneous set of parameters is required. Given this fact, and the available literature on quantitative galaxy morphological classification, two approaches to parameter selection appear to be strongly represented:

1. The luminosity profile along the major axis of the object is determined at discrete intervals. This may be done by the fitting of elliptical contours, followed by the integrating of light in elliptical annuli [24]. A similar approach was used in the ESO-Uppsala survey. Noisiness and faintness require attention to robustness in measurement: the radial

profile may be determined taking into account the assumption of a face–on optically–thin axisymmetric galaxy, and may be further adjusted to yield values for circles of given radius [53]. Alternatively, isophotal contours may determine the discrete radial values for which the profile is determined [51].

2. Specific morphology-related parameters may be derived instead of the profile. The integrated magnitude within the limiting surface brightness of 25 or 26 mag. arcsec$^{-2}$ in the visual is popular [50, 24]. The logarithmic diameter ($D_{26}$) is also supported by Okamura [34]. It may be interesting to fit to galaxies under consideration model bulges and disks using, respectively, $r^{\frac{1}{4}}$ or exponential laws [51], in order to define further parameters. Some catering for the asymmetry of spirals may be carried out by decomposing the object into octants; furthermore the taking of a Fourier transform of the intensity may indicate aspects of the spiral structure [50].

The following remarks can be made relating to image data and reduced data.

- The range of parameters to be used should be linked to the subsequent use to which they might be put, such as to underlying physical aspects.

- Parameters can be derived from a carefully-constructed luminosity profile, rather than it being possible to derive a profile from any given set of parameters.

- The presence of both partially reduced data such as luminosity profiles, and more fully reduced features such as integrated flux in a range of octants, is of course not a hindrance to analysis. However it is more useful if the analysis is carried out on both types of data separately.

Parameter data can be analyzed by clustering algorithms, by principal components analysis or by methods for discriminant analysis. Profile data can be sampled at suitable intervals and thus analyzed also by the foregoing procedures. It may be more convenient in practice to create dissimilarities between profiles, and analyze these dissimilarities: this can be done using clustering algorithms with dissimilarity input.

## 3   Mathematical Modeling

Different models may be considered to represent the data. One of the most effective is certainly the sparsity model, especially when a specific wavelet dictionary is chosen to represent the data. We introduce here the sparsity concept, as well as the wavelet transform decomposition which is the most used in astronomy.

### 3.1   Sparsity Data Model

A signal $X$, $X = [x_1, \cdots, x_N]^T$, is sparse if most of its entries are equal to zero. For instance, a $k$-sparse signal is a signal where only $k$ samples have a non-zero value. A less strict definition is to consider a signal as weakly sparse or compressible when only a few of its entries have a large magnitude, while most of them are close to zero.

If a signal is not sparse, it may be *sparsified* using a given data representation. For instance, if $X$ is a sine, it is clearly not sparse but its Fourier transform is extremely sparse

(i.e. 1-sparse). Hence we say that a signal $X$ is sparse in the Fourier domain if its Fourier coefficients $\hat{X}[u]$, $\hat{X}[u] = \frac{1}{N}\sum_{k=-\infty}^{+\infty} X[k]e^{2i\pi\frac{uk}{N}}$, are sparse. More generally, we can model a vector signal $X \in \mathbb{R}^N$ as the linear combination of $T$ elementary waveforms, also called *signal atoms*: $X = \mathbf{\Phi}\alpha = \sum_{i=1}^{T}\alpha[i]\phi_i$ ,, where $\alpha[i] = \langle X, \phi_i \rangle$ are called the decomposition coefficients of $X$ in the dictionary $\mathbf{\Phi} = [\phi_1, \cdots, \phi_T]$ (the $N \times T$ matrix whose columns are the atoms normalized to a unit $\ell_2$-norm, i.e. $\forall i \in [1,T], \|\phi_i\|_{\ell^2} = 1$).

Therefore to get a sparse representation of our data we need first to define the dictionary $\mathbf{\Phi}$ and then to compute the coefficients $\alpha$. $x$ is sparse in $\Phi$ if the sorted coefficients in decreasing magnitude have fast decay; i.e. most coefficients $\alpha$ vanish except for a few.

The best dictionary is the one which leads to the sparsest representation. Hence we could imagine having a huge overcomplete dictionary (i.e. $T \gg N$), but we would be faced with prohibitive computation time cost for calculating the $\alpha$ coefficients. Therefore there is a trade-off between the complexity of our analysis step (i.e. the size of the dictionary) and the computation time. Some specific dictionaries have the advantage of having fast operators and are very good candidates for analyzing the data.

The Isotropic Undecimated Wavelet Transform (IUWT), also called *starlet wavelet transform*, is well known in the astronomical domain because it is well adapted to astronomical data where objects are more or less isotropic in most cases [43, 46]. For more astronomical images, the starlet dictionary is very well adapted.

## 3.2  The Starlet Transform

The starlet wavelet transform [42] decomposes an $n \times n$ image $c_0$ into a coefficient set $W = \{w_1, \ldots, w_J, c_J\}$, as a superposition of the form

$$c_0[k,l] = c_J[k,l] + \sum_{j=1}^{J} w_j[k,l],$$

where $c_J$ is a coarse or smooth version of the original image $c_0$ and $w_j$ represents the details of $c_0$ at scale $2^{-j}$ (see Starck et al. [47, 45] for more information). Thus, the algorithm outputs $J+1$ sub-band arrays of size $N \times N$. (The present indexing is such that $j = 1$ corresponds to the finest scale or high frequencies).

The decomposition is achieved using the filter bank $(h_{2\mathrm{D}}, g_{2\mathrm{D}} = \delta - h_{2\mathrm{D}}, \tilde{h}_{2\mathrm{D}} = \delta, \tilde{g}_{2\mathrm{D}} = \delta)$ where $h_{2\mathrm{D}}$ is the tensor product of two 1D filters $h_{1\mathrm{D}}$ and $\delta$ is the Dirac function. The passage from one resolution to the next one is obtained using the "à trous" ("with holes") algorithm [47]

$$
\begin{aligned}
c_{j+1}[k,l] &= \sum_{m}\sum_{n} h_{1D}[m]h_{1D}[n]c_j[k+2^j m, l+2^j n], \\
w_{j+1}[k,l] &= c_j[k,l] - c_{j+1}[k,l] ,
\end{aligned}
\tag{6}
$$

If we choose a $B_3$-spline for the scaling function:

$$\phi(x) = B_3(x) =$$

$$\frac{1}{12}(\mid x-2 \mid^3 -4 \mid x-1 \mid^3 +6 \mid x \mid^3 -4 \mid x+1 \mid^3 + \mid x+2 \mid^3) \tag{7}$$
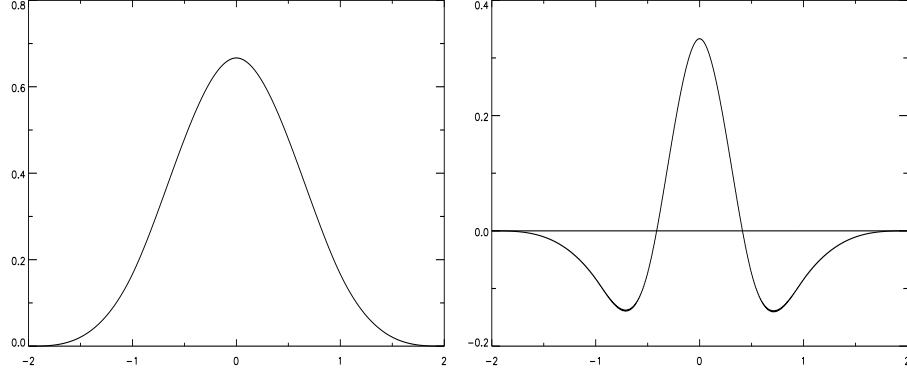
Figure 2: Left, the cubic spline function $\phi$; right, the wavelet $\psi$.

the coefficients of the convolution mask in one dimension are $h_{1D} = \left\{\frac{1}{16}, \frac{1}{4}, \frac{3}{8}, \frac{1}{4}, \frac{1}{16}\right\}$, and in two dimensions:

$$
h_{2D} = \begin{pmatrix} 1/16 & 1/4 & 3/8 & 1/4 & 1/16 \end{pmatrix} \begin{pmatrix} \frac{1}{16} \\ \frac{1}{4} \\ \frac{3}{8} \\ \frac{1}{4} \\ \frac{1}{16} \end{pmatrix} = \begin{pmatrix} \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{3}{128} & \frac{3}{32} & \frac{9}{64} & \frac{3}{32} & \frac{3}{128} \\ \frac{1}{64} & \frac{1}{16} & \frac{3}{32} & \frac{1}{16} & \frac{1}{64} \\ \frac{1}{256} & \frac{1}{64} & \frac{3}{128} & \frac{1}{64} & \frac{1}{256} \end{pmatrix}
$$

Fig. 2 shows the scaling function and the wavelet function when a cubic spline function is chosen as the scaling function $\phi$.

The most general way to handle the boundaries is to consider that $c[k + N] = c[N - k]$ ("mirror"). But other methods can be used such as periodicity ($c[k+N] = c[N]$), or continuity ($c[k + N] = c[k]$).

The starlet transform algorithm is:

1. We initialize $j$ to 0 and we start with the data $c_j[k, l]$.

2. We carry out a discrete convolution of the data $c_j[k, l]$ using the filter ($h_{2D}$), using the separability in the two-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. The distance between the central pixel and the adjacent ones is $2^j$.

3. After this smoothing, we obtain the discrete wavelet transform from the difference $c_j[k, l] - c_{j+1}[k, l]$.

4. If $j$ is less than the number $J$ of resolutions we want to compute, we increment $j$ and then go to step 2.

5. The set $\alpha = \{w_1, ..., w_J, c_J\}$ represents the wavelet transform of the data.

This starlet transform is very well adapted to the detection of isotropic features, and this explains its success for astronomical image processing, where the data contain mostly isotropic or quasi-isotropic objects, such as stars, galaxies or galaxy clusters.

Fig. 4 shows the starlet transform of the galaxy NGC 2997 displayed in Fig. 3 . Five wavelet scales are shown and the final smoothed plane (lower right). The original image is given exactly by the sum of these six images.
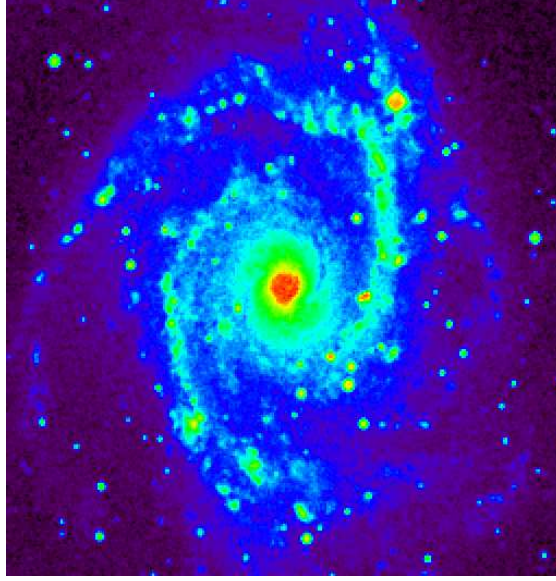
Figure 3: Galaxy NGC 2997.

## 3.3 The Starlet Reconstruction

The reconstruction is straightforward. A simple co-addition of all wavelet scales reproduces the original map: $c_0[k,l] = c_J[k,l] + \sum_{j=1}^{J} w_j[k,l]$. But because the transform is non-subsampled, there are many ways to reconstruct the original image from its wavelet transform [42]. For a given wavelet filter bank $(h,g)$, asssociated with a scaling function $\phi$ and a wavelet function $\psi$, any synthesis filter bank $(\tilde{h},\tilde{g})$, which satisfies the following reconstruction condition

$$\hat{h}^*(\nu)\hat{\tilde{h}}(\nu) + \hat{g}^*(\nu)\hat{\tilde{g}}(\nu) = 1 , \tag{8}$$

leads to exact reconstruction. For instance, for isotropic $h$, if we choose $\tilde{h} = h$ (the synthesis scaling function $\tilde{\phi} = \phi$) we obtain a filter $\tilde{g}$ defined by [42]:

$$\tilde{g} = \delta + h .$$

If $h$ is a positive filter, then $g$ is also positive. For instance, if $h_{1D} = [1,4,6,4,1]/16$, then $\tilde{g}_{1D} = [1,4,22,4,1]/16$. That is, $\tilde{g}_{1D}$ *is positive*. This means that $\tilde{g}$ is no longer related to a wavelet function. The 1D detail synthesis function related to $\tilde{g}_{1D}$ is defined by:

$$\frac{1}{2}\tilde{\psi}_{1D}\left(\frac{t}{2}\right) = \phi_{1D}(t) + \frac{1}{2}\phi_{1D}\left(\frac{t}{2}\right) . \tag{9}$$

Note that by choosing $\tilde{\phi}_{1D} = \phi_{1D}$, any synthesis function $\tilde{\psi}_{1D}$ which satisfies

$$\hat{\tilde{\psi}}_{1D}(2\nu)\hat{\psi}_{1D}(2\nu) = \hat{\phi}_{1D}^2(\nu) - \hat{\phi}_{1D}^2(2\nu) \tag{10}$$

leads to an exact reconstruction [27] and $\hat{\tilde{\psi}}_{1D}(0)$ can take any value. The synthesis function $\tilde{\psi}_{1D}$ does not need to verify the admissibility condition (i.e. to have a zero mean).
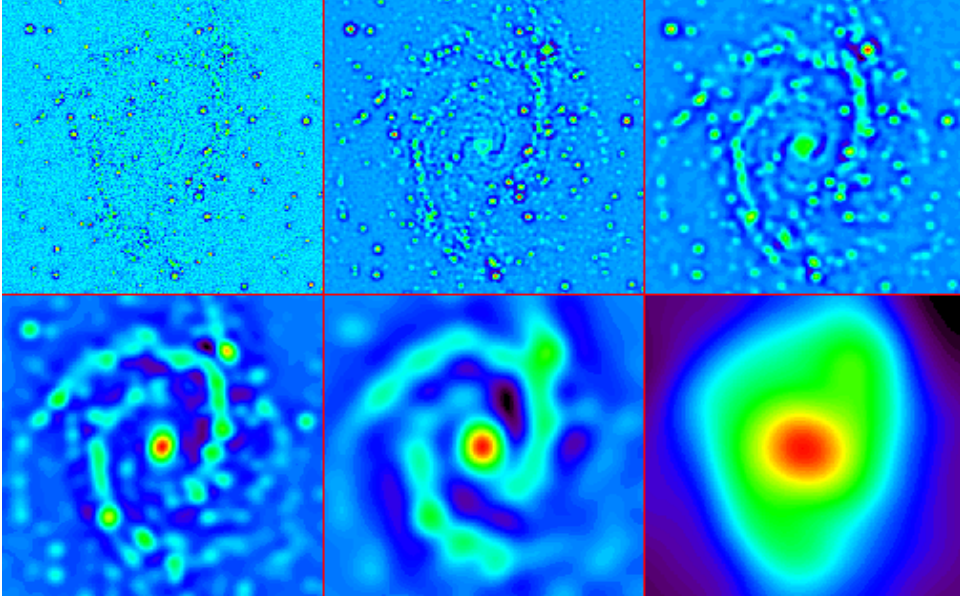
Figure 4: Wavelet transform of NGC 2997 by the IUWT. The co-addition of these six images reproduces exactly the original image.
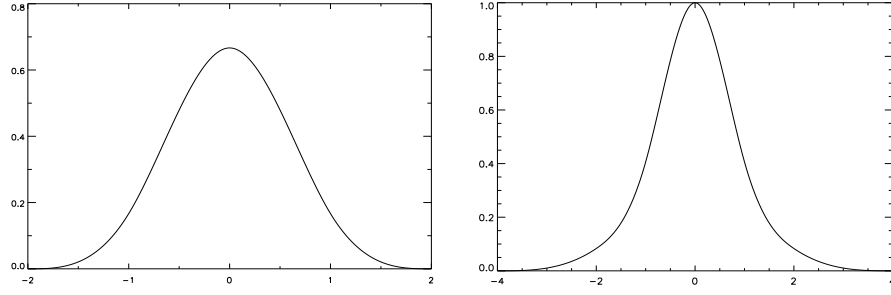


Figure 5: Left, $\tilde{\phi}_{1D}$ the 1D synthesis scaling function and right, $\tilde{\psi}_{1D}$ the 1D detail synthesis function.

Fig. 5 shows the two functions $\tilde{\phi}_{1D}$ $(= \phi_{1D})$ and $\tilde{\psi}_{1D}$ used in the reconstruction in 1D, corresponding to the synthesis filters $\tilde{h}_{1D} = h_{1D}$ and $\tilde{g}_{1D} = \delta + h_{1D}$. More details can be found in [42].

## 3.4   Starlet Transform: Second Generation

A particular case is obtained when $\hat{\tilde{\phi}}_{1D} = \hat{\phi}_{1D}$ and $\hat{\psi}_{1D}(2\nu) = \frac{\hat{\phi}^2_{1D}(\nu) - \hat{\phi}^2_{1D}(2\nu)}{\hat{\phi}_{1D}(\nu)}$, which leads to a filter $g_{1D}$ equal to $\delta - h_{1D} \star h_{1D}$. In this case, the synthesis function $\tilde{\psi}_{1D}$ is defined by $\frac{1}{2}\tilde{\psi}_{1D}(\frac{t}{2}) = \phi_{1D}(t)$ and the filter $\tilde{g}_{1D} = \delta$ is the solution to (8).

We end up with a synthesis scheme where only the smooth part is convolved during the reconstruction.

Deriving $h$ from a spline scaling function, for instance $B_1$ ($h_1 = [1, 2, 1]/4$) or $B_3$ ($h_3 = [1, 4, 6, 4, 1]/16$) (note that $h_3 = h_1 \star h_1$), since $h_{1D}$ is even-symmetric (i.e. $H(z) = H(z^{-1})$),

12

the $z$-transform of $g_{1D}$ is then:

$$G(z) = 1 - H^2(z) = 1 - z^4 \left( \frac{1 + z^{-1}}{2} \right)^8$$

$$= \frac{1}{256} \left( -z^4 - 8z^3 - 28z^2 - 56z + 186 - 56z^{-1} - 28z^{-2} - 8z^{-3} - z^{-4} \right) , \qquad (11)$$

which is the $z$-transform of the filter

$$g_{1D} = [-1, -8, -28, -56, 186, -56, -28, -8, -1]/256.$$

We get the following filter bank:

$$h_{1D} = h_3 = \tilde{h} = [1, 4, 6, 4, 1]/16$$
$$g_{1D} = \delta - h \star h = [-1, -8, -28, -56, 186, -56, -28, -8, -1]/256 \qquad (12)$$
$$\tilde{g}_{1D} = \delta .$$

The second generation starlet transform algorithm is:

1. We initialize $j$ to 0 and we start with the data $c_j[k]$.

2. We carry out a discrete convolution of the data $c_j[k]$ using the filter $h_{1D}$. The distance between the central pixel and the adjacent ones is $2^j$. We obtain $c_{j+1}[k]$.

3. We do exactly the same convolution on $c_{j+1}[k]$, and we obtain $c'_{j+1}[k]$.

4. After this two-steps smoothing, we obtain the discrete starlet wavelet transform from the difference $w_{j+1}[k] = c_j[k] - c'_{j+1}[k]$.

5. If $j$ is less than the number $J$ of resolutions we want to compute, we increment $j$ and then go to step 2.

6. The set $\alpha = \{w_1, ..., w_J, c_J\}$ represents the starlet wavelet transform of the data.

As in the standard starlet transform, extension to 2D is trivial. We just replace the convolution with $h_{1D}$ by a convolution with the filter $h_{2D}$, which is performed efficiently by using the separability.

With this specific filter bank, there is a no convolution with the filter $\tilde{g}_{1D}$ during the reconstruction. Only the low-pass synthesis filter $\tilde{h}_{1D}$ is used.

The reconstruction formula is

$$c_j[l] = (h_{1D}^{(j)} \star c_{j+1})[l] + w_{j+1}[l] , \qquad (13)$$

and denoting $L^j = h^{(0)} \star \cdots \star h^{(j-1)}$ and $L^0 = \delta$, we have

$$c_0[l] = \left( L^J \star c_J \right)[l] + \sum_{j=1}^{J} \left( L^{j-1} \star w_j \right)[l] . \qquad (14)$$

Each wavelet scale is convolved with a low-pass filter.

The second generation starlet reconstruction algorithm is:

13

1. The set $\alpha = \{w_1, ..., w_J, c_J\}$ represents the input starlet wavelet transform of the data.

2. We initialize $j$ to $J - 1$ and we start with the coefficients $c_j[k]$.

3. We carry out a discrete convolution of the data $c_{j+1}[k]$ using the filter $(h_{1D})$. The distance between the central pixel and the adjacent ones is $2^j$. We obtain $c'_{j+1}[k]$.

4. Compute $c_j[k] = c'_{j+1}[k] + w_{j+1}[k]$.

5. If $j$ is larger than 0, $j = j - 1$ and then go to step 3.

6. $c_0$ contains the reconstructed data.

As for the transformation, the 2D extension consists just in replacing the convolution by $h_{1D}$ with a convolution by $h_{2D}$.
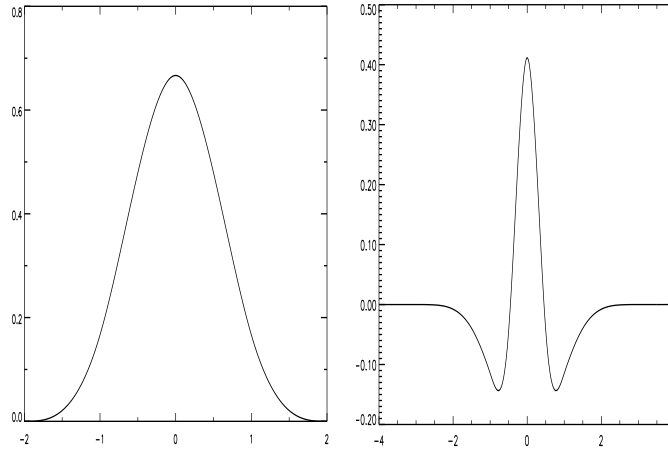


Figure 6: Left, the $\phi_{1D}$ analysis scaling function and right, the $\psi_{1D}$ analysis wavelet function. The synthesis functions $\tilde{\phi}_{1D}$ and $\tilde{\psi}_{1D}$ are the same as those in Fig. 5.

Fig. 6 shows the analysis scaling and wavelet functions. The synthesis functions $\tilde{\phi}_{1D}$ and $\tilde{\psi}_{1D}$ are the same as those in Fig. 5. As both are positive, we have a decomposition of an image $X$ on positive scaling functions $\tilde{\phi}_{1D}$ and $\tilde{\psi}_{1D}$, but the coefficients $\alpha$ are obtained with the starlet wavelet transform, and have a zero mean (except for $c_J$), as a regular wavelet transform.

In 2D, similarly, the 2nd generation starlet transform leads to the representation of an image $X[k, l]$:

$$X[k, l] = \sum_{m,n} \phi_{j,k,l}^{(1)}(m, n)c_J[m, n] + \sum_{j=1}^{J}\sum_{m,n} \phi_{j,k,l}^{(2)}(m, n)w_j[m, n] \ , \tag{15}$$

where $\phi_{j,k,l}^{(1)}(m, n) = 2^{-2j}\tilde{\phi}_{1D}(2^{-j}(k-m))\tilde{\phi}_{1D}(2^{-j}(l-n))$, and $\phi_{j,k,l}^{(2)}(m, n) = 2^{-2j}\tilde{\psi}_{1D}(2^{-j}(k-m))\tilde{\psi}_{1D}(2^{-j}(l-n))$.

$\phi^{(1)}$ and $\phi^{(2)}$ are positive, and $w_j$ are zero mean 2D wavelet coefficients.

The advantage of the second generation starlet transform will be seen in subsection 3.6 below.

14

## 3.5 Sparse Modeling of Astronomical Images

Using the sparse modeling, we now consider that the observed signal $X$ can be considered as a linear combination of a few atoms of the wavelet dictionary $\mathbf{\Phi} = [\phi_1, \cdots, \phi_T]$. The model of Eq. 3 is then replaced by the following:

$$Y = H\mathbf{\Phi}\alpha + N + B \tag{16}$$

and $X = \mathbf{\Phi}\alpha$, and $\alpha = \{w_1, \cdots, w_J, c_J\}$. Furthermore, most of the coefficients $\alpha$ will be equal to zero. Positions and scales of active coefficients are unknown, but they can be estimated directly from the data $Y$. We define the multiresolution support $M$ of an image $Y$ by:

$$M_j[k,l] = \left\{ \begin{array}{ll} 1 & \text{if } w_j[k,l] \text{ is significant} \\ 0 & \text{if } w_j[k,l] \text{ is not significant} \end{array} \right. \tag{17}$$

where $w_j[k,l]$ is the wavelet coefficient of $Y$ at scale $j$ and at position $(k,l)$. Hence, $M$ describes the set of active atoms in $Y$. If $H$ is compact and not too extended, then $M$ describes also well the active set of $X$. This is true because the background $B$ is generally very smooth, and therefore a wavelet coefficient $w_j[k,l]$ of $Y$, which does not belong to the coarsest scale is only dependent on $X$ and $N$ (the term $< \phi_i, B >$ being equal to zero).

### Selection of significant coefficients through noise modeling

We need now to determine when a wavelet coefficient is significant. Wavelet coefficients of $Y$ are corrupted by noise, which follows in many cases a Gaussian distribution, a Poisson distribution, or a combination of both. It is important to detect the wavelet coefficients which are "significant", i.e. the wavelet coefficients which have an absolute value too large to be due to noise.

For Gaussian noise, it is easy to derive an estimation of the noise standard deviation $\sigma_j$ at scale $j$ from the noise standard deviation, which can be evaluated with good accuracy in an automated way [44]. To detect the significant wavelet coefficients, it suffices to compare the wavelet coefficients $w_j[k,l]$ to a threshold level $t_j$. $t_j$ is generally taken equal to $K\sigma_j$, and $K$, as noted in section 2, is chosen between 3 and 5. The value of 3 corresponds to a probability of false detection of 0.27%. If $w_j[k,l]$ is small, then it is not significant and could be due to noise. If $w_j[k,l]$ is large, it is significant:

$$\begin{array}{ll} \text{if } |w_j[k,l]| \geq t_j & \text{then } w_j[k,l] \text{ is significant} \\ \text{if } |w_j[k,l]| < t_j & \text{then } w_j[k,l] \text{ is not significant} \end{array} \tag{18}$$

When the noise is not Gaussian, other strategies may be used:

- **Poisson noise:** if the noise in the data $Y$ is Poisson, the transformation [1] $\mathcal{A}(Y) = 2\sqrt{Y + \frac{3}{8}}$ acts as if the data arose from a Gaussian white noise model, with $\sigma = 1$, under the assumption that the mean value of $Y$ is sufficiently large. However, this transform has some limits and it has been shown that it cannot be applied for data with less than 20 counts (due to photons) per pixel. So for X-ray or gamma ray data, other solutions have to be chosen, which manage the case of a reduced number of events or photons under assumptions of Poisson statistics.

- **Gaussian + Poisson noise:** the generalization of variance stabilization [31] is:

$$\mathcal{G}((Y[k,l]) = \frac{2}{\alpha}\sqrt{\alpha Y[k,l] + \frac{3}{8}\alpha^2 + \sigma^2 - \alpha g}$$

where $\alpha$ is the gain of the detector, and $g$ and $\sigma$ are the mean and the standard deviation of the read-out noise.

- **Poisson noise with few events using the MS-VST:** For images with very few photons, one solution consists in using the Multi-Scale Variance Stabilization Transform (MS-VST) [54]. The MS-VST combines both the Anscombe transform and the starlet transform in order to produce *stabilized* wavelet coefficients, i.e. coefficients corrupted by a Gaussian noise with a standard deviation equal to 1. In this framework, wavelet coefficients are now calculated by:

$$\begin{array}{c} \text{Starlet} \\ + \\ \text{MS-VST} \end{array} \left\{ \begin{array}{rcl} c_j &=& \sum_m \sum_n h_{1D}[m]h_{1D}[n] \\ && c_{j-1}[k + 2^{j-1}m, l + 2^{j-1}n] \\ w_j &=& \mathcal{A}_{j-1}(c_{j-1}) - \mathcal{A}_j(c_j) \end{array} \right. \tag{19}$$

where $\mathcal{A}_j$ is the VST operator at scale $j$ defined by:

$$\mathcal{A}_j(c_j) = b^{(j)}\sqrt{|c_j + e^{(j)}|} \tag{20}$$

where the variance stabilization constants $b^{(j)}$ and $e^{(j)}$ only depend on the filter $h_{1D}$ and the scale level $j$. They can all be pre-computed once for any given $h_{1D}$ [54]. The multiresolution support is computed from the MS-VST coefficients, considering a Gaussian noise with a standard deviation equal to 1. This stabilization procedure is also invertible as we have:

$$c_0 = \mathcal{A}_0^{-1}\left[\mathcal{A}_J(a_J) + \sum_{j=1}^{J} w_j\right] \tag{21}$$

For other kinds of noise (correlated noise, non-stationary noise, etc.), other solutions have been proposed to derive the multiresolution support [46].

## 3.6 Sparse Positive Decomposition

Many astronomical images can be modeled as a sum of positive features, like stars and galaxies, which are more or less isotropic. The previous representation, based on the starlet transform, is well adapted to the representation of isotropic objects, but does not introduce any prior relative to the positivity of the features contained in our image. A positive and sparse modeling of astronomical images is similar to Eq. 16:

$$Y = H\mathbf{\Phi}\alpha + N + B \tag{22}$$

or

$$Y = \mathbf{\Phi}\alpha + N + B \tag{23}$$

16

if we do not take into account the point spread function. All coefficients in $\alpha$ are now positive, and all atoms in the dictionary $\mathbf{\Phi}$ are positive functions. Such a decomposition normally requires computationally intensive algorithms such as Matching Pursuit [28]. The second generation starlet transform offers us a new way to perform such a decomposition. Indeed, we have seen in section 3.4 that, using a specific filter bank, we can decompose an image $Y$ on a positive dictionary $\mathbf{\Phi}$ (see Fig. 5) and obtain a set of coefficients $\alpha^{(Y)}$, where $\alpha^{(Y)} = \mathbf{W}Y = \{w_1, ..., w_J, c_J\}$, $\mathbf{W}$ being the starlet wavelet transform operator. $\alpha$ coefficients are positive and negative, and are obtained using the standard starlet wavelet transform algorithm. Hence, by thresholding all negative (respectively, positive) coefficients, the reconstruction is always positive (respectively, negative), since $\mathbf{\Phi}$ contains only positive atoms.

Hence, we would like to have a sparse set of positive coefficients $\tilde{\alpha}$ which verify $\mathbf{\Phi}\tilde{\alpha} = Y$. But in order to take into account the background and the noise, we need to define the constraint in the wavelet space (i.e. $\mathbf{W}\mathbf{\Phi}\tilde{\alpha} = \mathbf{W}Y = \alpha^{(Y)}$), and this constraint must be applied only to the subset of coefficients in $\alpha^{(Y)}$ which are larger than the detection level. Therefore, to get a sparse positive decomposition on $\mathbf{\Phi}$, we need to minimize:

$$\tilde{\alpha} = \min_{\alpha} \| \alpha \|_1 \quad \text{s.t. } M\mathbf{W}\mathbf{\Phi}\alpha = M\alpha^{(Y)} \quad , \tag{24}$$

where $M$ is the multiresolution support defined in the previous section (i.e. $M_j[k,l] = 1$ is a significant coefficient is detected at scale $j$ and at position $(k,l)$, and zero otherwise). To remove the background, we have to set $M_{J+1}[k,l] = 0$ for all $(k,l)$.

It was shown that such optimization problems can be efficiently solved through an iterative soft thresholding (IST) algorithm [19, 41, 10]. The following algorithm, based on the IST, allows to take into account the noise modeling through the multiresolution support and force the coefficients to be all positive.

1. Take the 2nd generation starlet wavelet transform of the data $Y$, we obtain $\alpha^{(Y)}$.

2. From a given noise model, determine the multiresolution support $M$.

3. Set the number of iterations $N_{iter}$, the first threshold, $\lambda^{(0)} = MAX(\alpha^{(Y)})$, and the solution $\tilde{\alpha}^{(0)} = 0$.

4. For $0 = 1, N_{iter}$ do

   - Reconstruct the image $\tilde{Y}^{(i)}$ from $\tilde{\alpha}^{(i)}$: $\tilde{Y}^{(i)} = \mathbf{\Phi}\tilde{\alpha}^{(i)}$.

   - Take the 2nd generation starlet wavelet transform of the data $\tilde{Y}^{(i)}$, we obtain $\alpha^{\tilde{Y}^{(i)}} = \mathbf{W}\mathbf{\Phi}\tilde{\alpha}^{(i)}$.

   - Compute the significant residual $r^{(i)}$:

$$r^{(i)} = M\left(\alpha^{(Y)} - \alpha^{\tilde{Y}^{(i)}}\right) = M\left(\alpha^{(Y)} - \mathbf{W}\mathbf{\Phi}\tilde{\alpha}^{(i)}\right) \tag{25}$$

   - Calculate the value $\lambda^{(i)} = \lambda^{(0)}(1 - i/N_{iter})$

   - Update the solution, by adding the residual, applying a soft thresholding on positive coefficients using the threshold level $\lambda^{(i)}$, and setting all negative coefficients

17

to zero.

$$\begin{aligned}
\tilde{\alpha}^{(i+1)} &= \left( \tilde{\alpha}^{(i)} + r^{(i)} - \lambda^{(i)} \right)_+ \\
&= \left( \tilde{\alpha}^{(i)} + M \left( \alpha^{(Y)} - \mathbf{W}\mathbf{\Phi}\tilde{\alpha}^{(i)} \right) - \lambda^{(i)} \right)_+
\end{aligned} \tag{26}$$

- $i = i + 1$.

5. The set $\tilde{\alpha} = \tilde{\alpha}^{(N_{iter})}$ represents the sparse positive decomposition of the data.

The threshold parameter $\lambda^{(i)}$ decreases with the iteration number and it plays a role similar to the cooling parameter of the simulated annealing techniques, i.e. it allows the solution to escape from local minima.

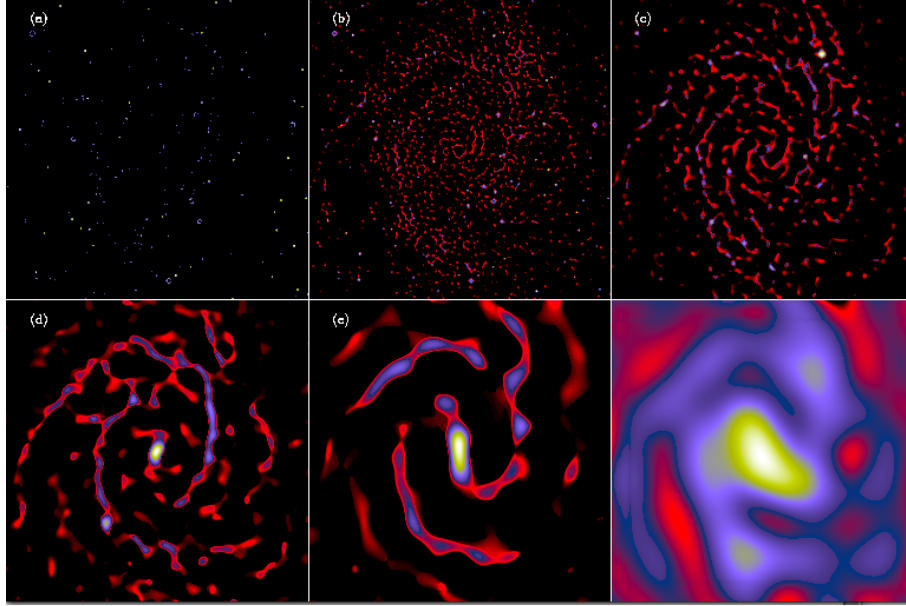**Example 1: Sparse Positive Decomposition of NGC2997**



Figure 7: Positive starlet decomposition of the galaxy NGC2997 with six scales.

Fig. 7 shows the position starlet decomposition, using 100 iterations, and can be compared to Fig. 4.

**Example 2: Sparse Positive Starlet Decomposition of a simulated image**

The next example compares the standard starlet transform to the positive starlet decomposition (PSD) on a simulated image.

Fig. 8 shows respectively from top to bottom and left to right, a) the original simulated image, b) the noisy data, c) the reconstruction from the PSD coefficients, and d) the residual between the noisy data and the PSD reconstructed image (i.e. image b − image c). Hence, the PSD reconstructed image gives a very good approximation of the original image. No structures can be be seen in the residual, and all sources are well detected.
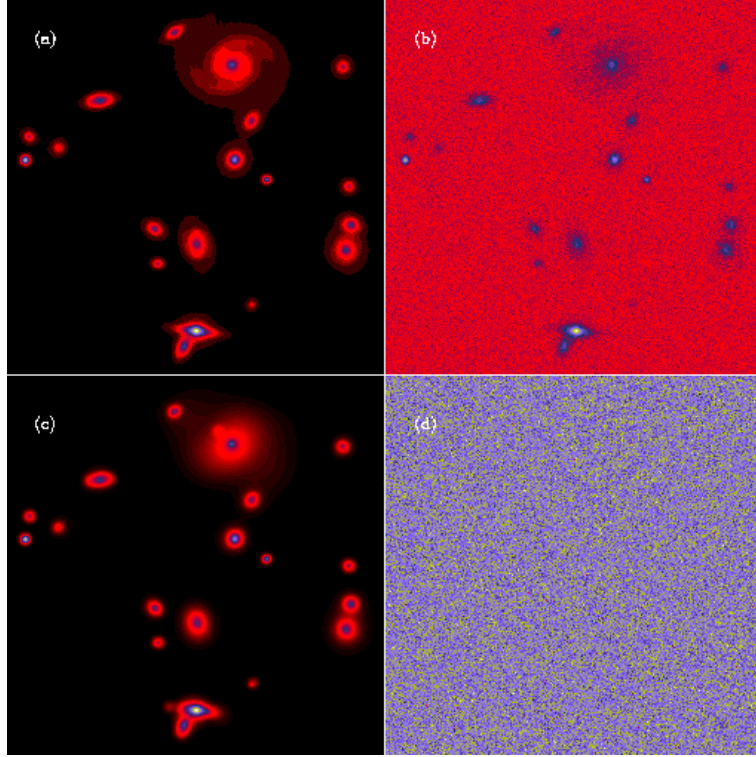
Figure 8: Top left and right, original simulated image and the same image contaminated by a Gaussian noise. Bottom left and right, reconstructed image for the positive starlet coefficients of the noisy image using 50 iterations, and residual (i.e. noisy image - reconstructed image).

The first PSD scale does not contain any non-zero coefficient. Fig. 9 top shows the four first scales of the wavelet transform, and Fig. 9 bottom the four first scales of the PSD.

## 4    Source Detection using a Sparsity Model

As described is the previous section, the wavelet coefficients of $Y$ which do not belong to the coarsest scale $c_J$ are not dependent on the background. This is a serious disadvantage, since the background estimation can be sometimes very problematic.

Two approaches have been proposed to detect sources, assuming the signal is sparse in the wavelet domain. The first consists in first removing the noise and the background, and then applying the standard approach described in section 2. It has been used for many years for X-ray source detection [48, 36]. The second approach, called Multiscale Vision Model [7], attempts to define directly an astronomical object in the wavelet space.

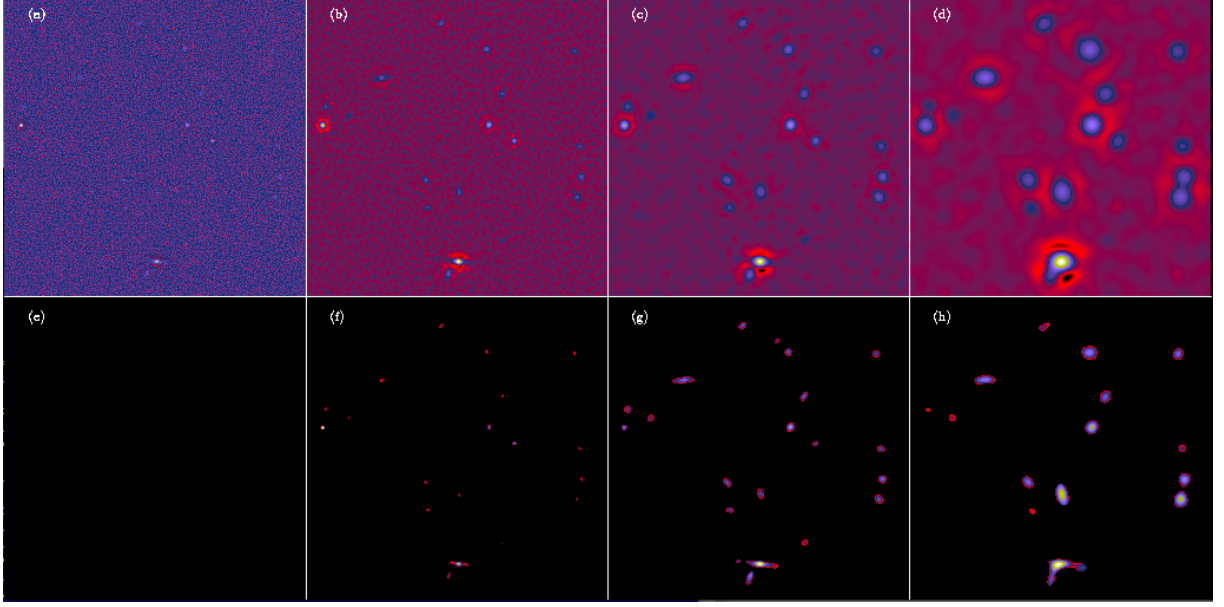### 4.1    Detection through Wavelet Denoising

19

Figure 9: Top, starlet transform, and bottom, positive starlet decomposition of a simulated astronomical image.

The most commonly used filtering method is hard thresholding, which consists of setting to 0 all wavelet coefficients of $Y$ which have an absolute value lower than a threshold $t_j$:

$$\tilde{w}_j[k,l] = \begin{cases} w_j[k,l] & \text{if } \mid w_j[k,l] \mid > t_j \\ 0 & \text{otherwise} \end{cases} \tag{27}$$

More generally, for a given sparse representation (i.e. wavelet) with its associated fast transform $\mathbf{W}$ and fast reconstruction $\mathbf{R}$, we can derive a hard threshold denoising solution $X$ from the data $Y$, by first estimating the multiresolution support $M$ using a given noise model, and then calculating:

$$X = \mathbf{R}M\mathbf{W}Y. \tag{28}$$

We transform the data, multiply the coefficients by the support and reconstruct the solution.

The solution can however be improved by considering the following optimization problem, $\min_X \parallel M(\mathbf{W}Y - \mathbf{W}X) \parallel_2^2$, where $M$ is the multiresolution support of $Y$. A solution can be obtained using the Landweber iterative scheme [40, 47]:

$$X^{n+1} = X^n + \mathbf{R}M\left[\mathbf{W}Y - \mathbf{W}X^n\right] \tag{29}$$

If the solution is known to be positive, the positivity constraint can be introduced using the following equation:

$$X^{n+1} = P_+ \left(X^n + \mathbf{R}M\left[\mathbf{W}Y - \mathbf{W}X^n\right]\right) \tag{30}$$

where $P_+$ is the projection on the cone of non-negative images.

This algorithm allows us to constrain the residual to have a zero value within the multiresolution support [47]. For astronomical image filtering, iterating improves significantly

the results, especially for the photometry (i.e. the integrated number of photons in a given object).

Removing the background in the solution is straightforward. The algorithm does not need to be modified. We only need to set to zero the coefficients related to the coarsest scale in the multiresolution support: $\forall k \ \ M_J[k,l] = 0$.

## 4.2 The Multiscale Vision Model

### 4.2.1 Introduction

The wavelet transform of an image $Y$ by the starlet transform produces at each scale $j$ a set $\{w_j\}$. This has the same number of pixels as the image. The original image $I$ can be expressed as the sum of all the wavelet scales and the smoothed array $c_J$ by the expression

$$Y[k,l] = c_J[k,l] + \sum_{j=1}^{J} w_j[k,l]. \tag{31}$$

Hence, we have a *multiscale pixel representation*, i.e. each pixel of the input image is associated with a set of pixels of the multiscale transform. A further step is to consider a *multiscale object representation*, which would associate with an object contained in the data a volume in the multiscale transform. Such a representation obviously depends on the kind of image we need to analyze, and we present here a model which has been developed for astronomical data. It may however be used for other kinds of data, to the extent that such data are similar to astronomical data. We assume that an image $Y$ can be decomposed into a set of components:

$$Y[k,l] = \sum_{i=1}^{N_o} X_i[k,l] + B[k,l] + N[k,l] \tag{32}$$

where $N_o$ is the number of components, $X_i$ are the components contained in the data (stars galaxies, etc.), $B$ is the background image, and $N$ is the noise.

To perform such a decomposition, we have to detect, to extract, to measure and to recognize the significant structures. This is done by first computing the multiresolution support of the image (i.e. the set of significant active coefficients), and by applying a segmentation scale by scale. The wavelet space of a 2D direct space is a 3D volume. An object, associated to a component, has to be defined in this space. A general idea for object definition lies in the connectivity property. An object occupies a physical region, and in this region we can join any pixel to other pixels based on significant adjacency. Connectivity in direct space has to be transported into wavelet transform space. In order to define the objects we have to identify the wavelet transform space pixels we can attribute to the objects. We describe in this section the different steps of this method.

### 4.2.2 Multiscale Vision Model Definition

The multiscale vision model, MVM [7], described an object as a hierarchical set of structures. It uses the following definitions:

- *Significant wavelet coefficient*: a wavelet coefficient is significant when its absolute value is above a given detection limit. The detection limit depends on the noise model (Gaussian noise, Poisson noise, and so on). See section 3.5 for more details.

- *Structure*: a structure $S_{j,k}$ is a set of significant connected wavelet coefficients at the same scale $j$:

$$S_{j,k} = \{w_j[k_1, l_1], w_j[k_2, l_2], \cdots, w_j[k_p, l_p]\} \tag{33}$$

  where $p$ is the number of significant coefficients included in the structure $S_{j,k}$, and $w_{j,x_i,y_i}$ is a wavelet coefficient at scale $i$ and at position $(x_i, y_i)$.

- *Object*: an object is a set of structures:

$$O_l = \{S_{j_1,k_1}, \cdots, S_{j_n,k_n}\} \tag{34}$$

  We define also the operator $\mathcal{L}$ which indicates to which object a given structure belongs: $\mathcal{L}(S_{j,k}) = l$ is $S_{j,k} \in O_l$, and $\mathcal{L}(S_{j,k}) = 0$ otherwise.

- *Object scale*: the scale of an object is given by the scale of the maximum of its wavelet coefficients.

- *Interscale relation*: the criterion allowing us to connect two structures into a single object is called the "interscale relation".

- *Sub-object*: a sub-object is a part of an object. It appears when an object has a local wavelet maximum. Hence, an object can be composed of several sub-objects. Each sub-object can also be analyzed.

### 4.2.3  From Wavelet Coefficients to Object Identification

**Multiresolution support segmentation.**

Once the multiresolution support has been calculated, we have at each scale a boolean image (i.e. pixel intensity equals 1 when a significant coefficient has been detected, and 0 otherwise). The segmentation consists of labeling the boolean scales. Each group of connected pixels having a "1" value gets a label value between 1 and $L_{\max}$, $L_{\max}$ being the number of groups. This process is repeated at each scale of the multiresolution support. We define a "structure" $S_{j,i}$ as the group of connected significant pixels which has the label $i$ at a given scale $j$.

**Interscale connectivity graph.**

An object is described as a hierarchical set of structures. The rule which allows us to connect two structures into a single object is called "interscale relation". Fig. 10 shows how several structures at different scales are linked together, and form objects. We have now to define the interscale relation. Let us consider two structures at two successive scales, $S_{j,k}$ and $S_{j+1,l}$. Each structure is located in one of the individual images of the decomposition and corresponds to a region in this image where the signal is significant. Denoting $(x_m, y_m)$ the pixel position of the maximum wavelet coefficient value of $S_{j,k}$, $S_{j,k}$ is said to be connected
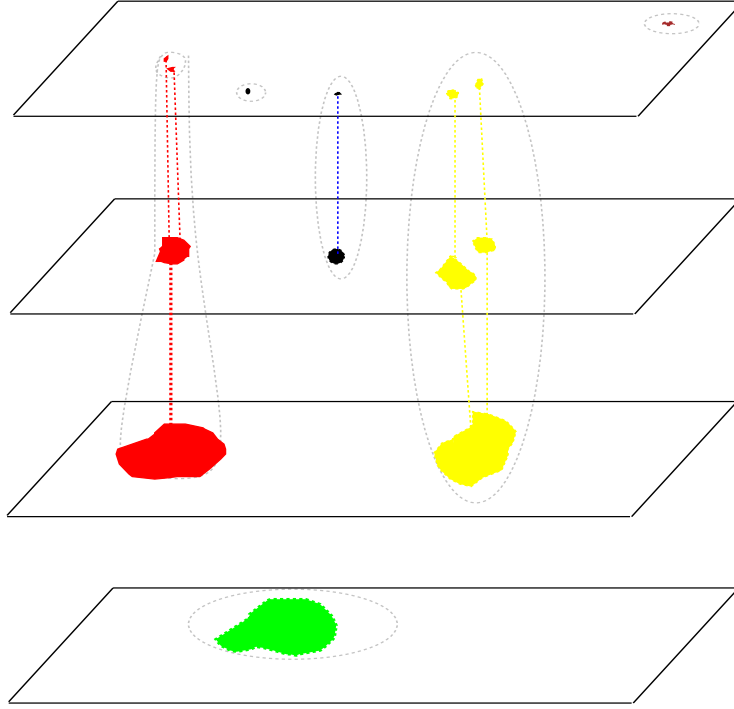
Figure 10: Example of connectivity in wavelet space: contiguous significant wavelet coefficients form a structure, and following an interscale relation, a set of structures forms an object. Two structures $S_j, S_{j+1}$ at two successive scales belong to the same object if the position pixel of the maximum wavelet coefficient value of $S_j$ is included in $S_{j+1}$.

to $S_{j+1,l}$ if $S_{j+1,l}$ contains the pixel position $(x_m, y_m)$ (i.e. the pixel position of the maximum wavelet coefficient of the structure $S_{j,k}$ must also be contained in the structure $S_{j+1,l}$). Several structures appearing in successive wavelet coefficient images can be connected in such a way, which we call an object in the interscale connectivity graph. Hence, we identify $n_o$ objects in the wavelet space, each object $O_i$ being defined by a set of structures, and we can assign to each structure a label $i$, with $i \in [1, n_o]$: $\mathcal{L}(S_{j,k}) = i$ if the structure $S_{j,k}$ belongs to the $i$th object.

**Filtering.**

Statistically, some significant structures can be due to the noise. They contain very few pixels and are generally isolated, i.e. connected to no field at upper and lower scales. So, to avoid false detection, the isolated fields can be removed from the initial interscale connection graph. Structures at the border of the images may also have been detected because of the border problem, and can be removed.

**Merging/deblending.**

23

As in the standard approach, true objects which are too close may generate a set of connected structures, initially associated with the same object, and a decision must be taken whether to consider such a case as one or two objects. Several cases may be distinguished:

- Two (or more) close objects, approximately of the same size, generate a set of structures. At a given scale $j$, two separate structures $S_{j,1}$ and $S_{j,2}$ are detected while at the scale $j+1$, only one structure is detected $S_{j+1,1}$, which is connected to the $S_{j,1}$ and $S_{j,2}$.

- Two (or more) close objects of different sizes generate a set of structures, from scale $j$ to scale $k$ ($k > j$).

In the wavelet space, the merging/deblending decision will be based on the local maxima values of the different structures belonging to this object. A new object (i.e. deblending) is derived from the structure $S_{j,k}$ if there exists at least one other structure at the same scale belonging to the same object (i.e. there exists one structure $S_{j+1,a}$ and at least one structure $S_{j,b}$ such that $\mathcal{L}(S_{j+1,a}) = \mathcal{L}(S_{j,b}) = \mathcal{L}(S_{j,k})$), and if the following relationship is verified: $w_j^m > w_{j-1}^m$ and $w_j^m > w_{j+1}^m$, where:

- $w_j^m$ is the maximum wavelet coefficient of the structure $S_{j,k}$: $w_j^m = \text{Max}(S_{j,k})$.

    - $w_{j-1}^m = 0$ if $S_{j,k}$ is not connected to any structure at scale $j-1$.
    - $w_{j-1}^m$ is the maximum wavelet coefficient of the structure $S_{j-1,l}$, where $S_{j-1,l}$ is such that $\mathcal{L}(S_{j-1,l}) = \mathcal{L}(S_{j,k})$ and the position of its highest wavelet coefficient is the closest to the position of the maximum of $S_{j,k}$.

- $w_{j+1}^m = \text{Max}\{w_{j+1,x_1,y_1}, \cdots, w_{j+1,x_n,y_n}\}$, where all wavelet coefficients $w_{j+1,x,y}$ are at a position which belongs also to $S_{j,k}$ (i.e. $w_{j,x,y} \in S_{j,k}$).

When these conditions are verified, $S_{j,k}$ and all structures at smaller scales which are directly or indirectly connected to $S_{j,k}$ will define a new object.

**Object identification.**

We can now summarize this method allowing us to identify all the objects in a given image $Y$:

1. We compute the wavelet transform with the starlet algorithm, which leads to a set $\alpha = \mathbf{W}Y = \{w_1, \ldots, w_J, c_J\}$. Each scale $w_j$ has the same size as the input image.

2. We determine the noise standard deviation in $w_1$.

3. We deduce the thresholds at each scale from the noise modeling.

4. We threshold scale-by-scale and we do an image labeling.

5. We determine the interscale relations.

6. We identify all the wavelet coefficient maxima of the wavelet transform space.

7. We extract all the connected trees resulting from each wavelet transform space maximum.

### 4.3  Source Reconstruction

**Partial reconstruction as an inverse problem.**

A set of structures $\mathcal{S}_i$ ($\mathcal{S}_i = \{S_{j,k}, \cdots, S_{j',k'}\}$) defines an object $O_i$ which can be reconstructed separately from other objects, in order to provide the components $X_i$. The coaddition of all reconstructed objects is a filtered version of the input data. We will denote $\alpha_i$ the set of wavelet coefficients belonging to the object $O_i$. Therefore, $\alpha_i$ is a subset of the wavelet transform of $X_i$, $\tilde{\alpha}_i = \mathbf{W} X_i$. Indeed, the last scale of $\tilde{\alpha}_i$ is unknown, as well as many wavelet coefficients which have not been detected. Then the reconstruction problem consists of searching for an image $X_i$ such that its wavelet transform reproduces the coefficients $\alpha_i$ (i.e. they are the same as those of $\mathcal{S}_i$, the detected structures). If $\mathbf{W}$ describes the wavelet transform operator, and $P_w$ the projection operator in the subspace of the detected coefficients (i.e. having set to zero all coefficients at scales and positions where nothing was detected), the solution is found by minimization of :

$$\min_{X_i} \parallel \alpha_i - P_w \left( \mathbf{W} X_i \right) \parallel^2 \tag{35}$$

The size of the restored image $X_i$ is arbitrary and it can be easily set greater than the number of known coefficients. It is certain that there exists at least one image $X_i$ which gives exactly $\alpha_i$, i.e. the original one. But generally we have an infinity of solutions, and we have to choose among them the one which is considered as correct. An image is always a positive function, which leads us to constrain the solution, but this is not sufficient to get a unique solution. More details on the reconstruction algorithm can be found in [7, 46].

### 4.4  Examples

**Band extraction.**

We simulated a spectrum which contains an emission band at 3.50 $\mu$m and non-stationary noise superimposed on a smooth continuum. The band is a Gaussian of width FWHM = 0.01 $\mu$m (FWHM = full width at half-maximum), and normalized such that its maximum value equals ten times the local noise standard deviation.

Fig. 11 (top) contains the simulated spectrum. The wavelet analysis results in the detection of an emission band at 3.50 $\mu$m above $3\sigma$. Fig. 11 (middle) shows the reconstruction of the detected band in the simulated spectrum. The real feature is over-plotted as a dashed line. Fig. 11 (bottom) contains the original simulation with the reconstructed band subtracted. It can be seen that there are no strong residuals near the location of the band, which indicates that the band is well-reconstructed. The center position of the band, its FWHM, and its maximum, can then be estimated via a Gaussian fit. More details about the use of MVM for spectral analysis can be found in [49].

**Star extraction in NGC2997.**

We applied MVM to the galaxy NGC2997 (Fig. 12, top left). Two images were created by coadding objects detected from scales 1 and 2, and from scales 3 to 6. They are displayed respectively in Fig. 12, top right, and bottom left. Fig. 12, bottom right, shows the difference between the input data and the image which contained the objects from scales 1 and 2. As we can see, all small objects have been removed, and the galaxy can be better analyzed.
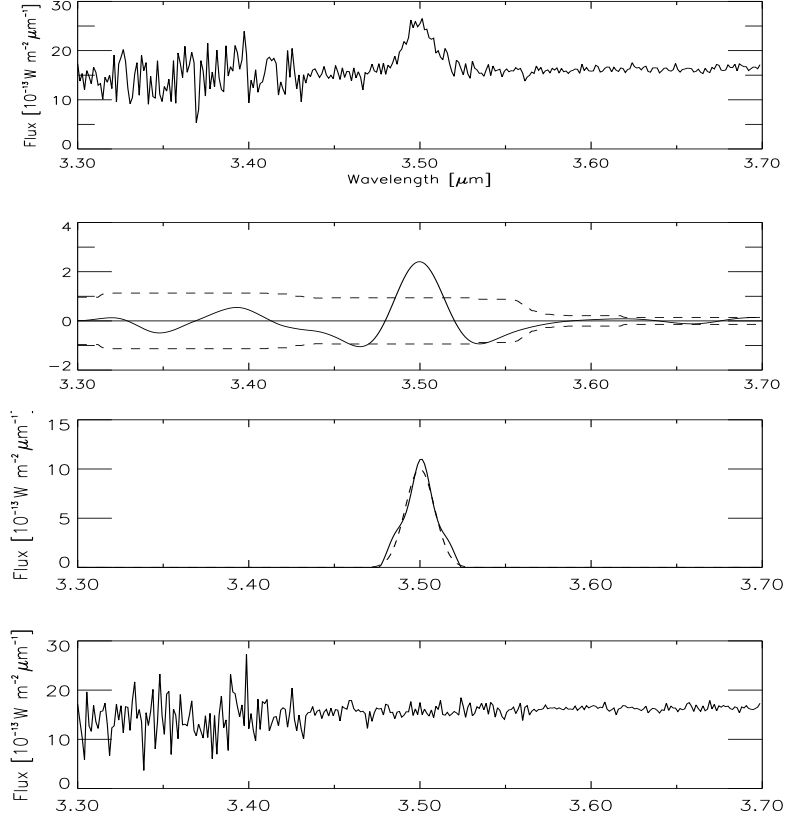
Figure 11: Top: simulated spectrum. Middle: reconstructed simulated band (full line) and original band (dashed line). Bottom: simulated spectrum minus the reconstructed band.

**Galaxy nucleus extraction.**

Fig. 13 shows the extracted nucleus of NGC2997 using the MVM method, and the difference between the galaxy image and the nucleus image.

# 5 Deconvolution

Up to now, the PSF $H$ has not been considered in the source detection. This means that all morphological parameters (size, ellipticity, etc.) derived from the detected objects need to be corrected from the PSF. Very close objects may also be seen as a single object because $H$ acts as a blurring operator on the data. A solution may consist in deconvolving first the data, and carrying out the source detection afterwards.

The problem of image deconvolution is ill-posed [3] and, as a consequence, the matrix $H$ modeling the imaging system is ill-conditioned. If $Y$ is the observed image and $X$ the unknown object, the equation $HX = Y$ has not a unique and stable solution. Therefore one must look for approximate solutions of this equation that are also physically meaningful. One approach is Tikhonov regularization theory [18]; however, a more general approach is provided by the so-called Bayes paradigm [20], even if it is applicable only to discrete problems. In
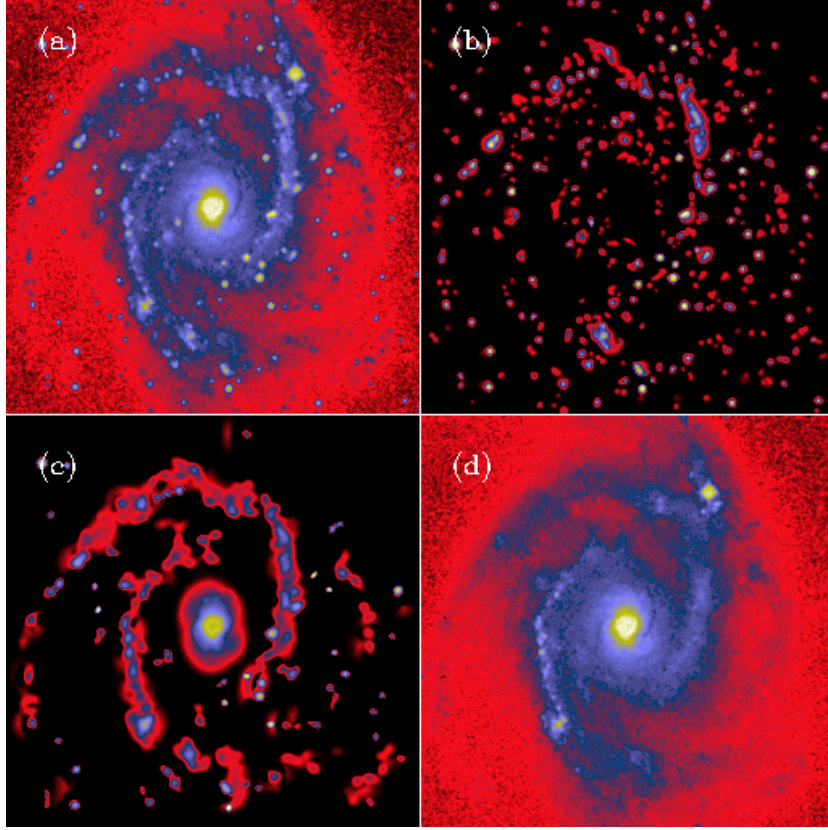
Figure 12: (a) Galaxy NGC2997, (b) objects detected from scales 1 and 2, (c) objects detected from scales 3 to 6, and (d) difference between (a) and (b).

this framework one can both take into account statistical properties of the data (Tikhonov regularization is obtained by assuming additive Gaussian noise) and also introduce *a priori* information on the unknown object.

## 5.1  Statistical Approach to Deconvolution

We assume that the detected image $Y$ is the realization of a multi-valued random variable $I$ corresponding to the (unknown) value $X$ of another multi-valued random variable, the object $O$. Moreover we assume that the *conditional probability distribution* $p_I(Y|X)$ is known. Since the unknown object appears as a set of unknown parameters, the problem of image deconvolution can be considered as a classical problem of parameter estimation. The standard approach is the *maximum likelihood* (ML) method. In our specific application, for a given detected image $Y$, this consists of introducing the *likelihood function* defined by

$$L_Y(X) = p_I(Y; X) \ .  \tag{36}$$

Then the ML estimate of the unknown object is any maximizer $X^*$ of the likelihood function

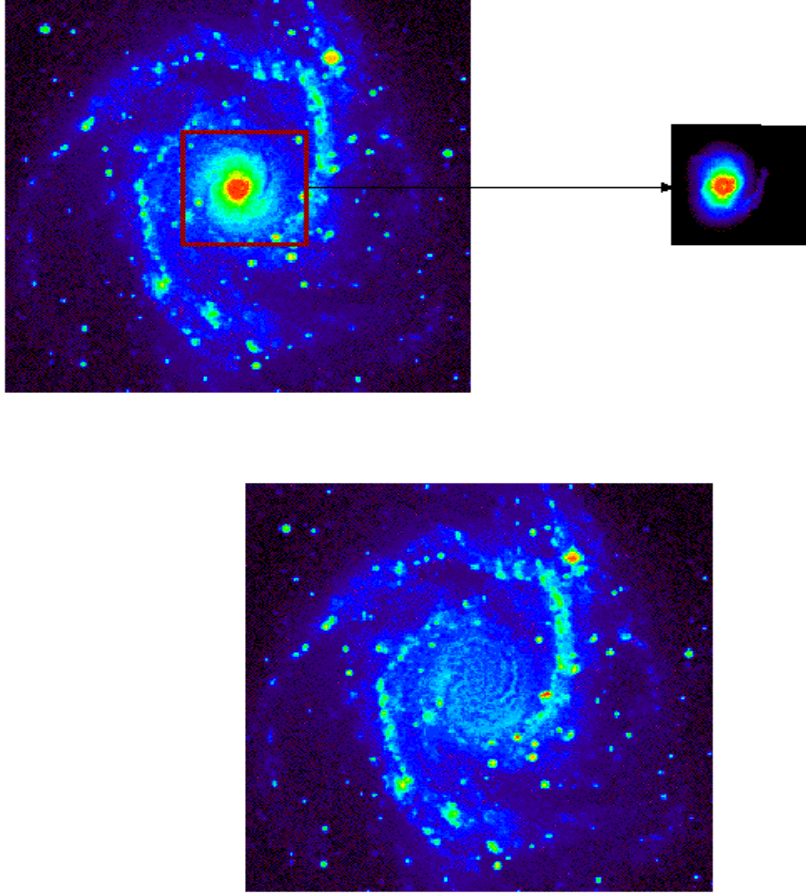$$X^* = \arg\max_{X \in \mathbb{R}^n} \ L_Y(X) \ ,  \tag{37}$$

Figure 13: Upper left, galaxy NGC2997; upper right, extracted nucleus; bottom, difference between the two previous images.

if it exists.

In our applications the likelihood function is the product of a very large number of terms (the data components are assumed to be statistically independent), so that it is convenient to take the logarithm of this function; moreover, if we consider the negative logarithm (the so-called neglog), the maximization problem is transformed into a minimization one. Let us consider the function

$$J_0(X;Y) = -A \ln L_Y(X) + B \quad , \tag{38}$$

where $A, B$ are suitable constants. They are introduced in order to obtain a function which has a simpler expression and is also nonnegative since, in our applications, the neglog of the likelihood is bounded from below. Then, it is easy to verify that the problem of Eq. (37) is equivalent to the following one:

$$X^* = \arg\min_{X \in \mathbb{R}^n} \; J_0(X;Y) \quad . \tag{39}$$

We consider now the model of Eq. (2) with three different examples of noise.

**Example 1** *In the case of additive white Gaussian noise, by a suitable choice of the constants*

$A, B$, we obtain (we assume here that the background $B$ is not subtracted even if it must be estimated)

$$J_0(X;Y) = ||HX + B - Y||^2 \quad , \tag{40}$$

and therefore the ML approach coincides with the well-known least-squares (LS) approach. It is also well-known that the function of Eq. (40) is convex, and strictly convex if and only if the equation $HX = 0$ has only the solution $X = 0$. Moreover it has always absolute minimizers, i.e. the LS-problem has always a solution; but the problem is ill-conditioned because it is equivalent to the solution of the Euler equation:

$$H^T H\ X = H^T(Y - B) \quad . \tag{41}$$

We remark that the ill-posedness of the LS-problem is the starting point of Tikhonov regularization theory (see, for instance, [52, 18]), and therefore this theory is based on the tacit assumption that the noise affecting the data is additive and Gaussian.

We remark that, in the case of object reconstruction, since objects are non-negative, we should consider the minimization of the function of Eq. (40) on the non-negative orthant. With such a constraint the problem is not treatable in the standard framework of regularization theory.

**Example 2** In the case of Poisson noise, if we introduce the so-called generalized Kullback-Leibler (KL) divergence of a vector $Z$ from a vector $Y$, defined by

$$D_{KL}(Y, Z) = \sum_{i=1}^{m} \left\{ y_i\ \ln \frac{Y_i}{Z_i} + Z_i - Y_i \right\} \quad , \tag{42}$$

then, with a suitable choice of the constants $A, B$, the function $J_0(X;Y)$ is given by

$$\begin{aligned} J_0(X;Y) &= D_{KL}(Y; HX + B) = \\ &= \sum_{i=1}^{m} \left\{ Y_i\ \ln \frac{y_i}{(HX + B)_i} + (HX + B)_i - y_i \right\} \quad . \end{aligned} \tag{43}$$

It is quite natural to take the non-negative orthant as the domain of this function. Moreover, it is well-known that it is convex (strictly convex if the equation $HX = 0$ has only the solution $X = 0$), non-negative, coercive and locally bounded. Therefore it has absolute minimizers. However, these minimizers are strongly affected by noise and the specific effect of the noise in this problem is known as checkerboard effect [32], since many components of the minimizers are zero.

**Example 3** In the case of Gauss+Poisson noise, the function $J_0(X;Y)$ is given by a much more complex form. This function is also convex (strictly convex if the equation $Hx = 0$ has the unique solution $x = 0$), non-negative, locally bounded and coercive [2]. Therefore it also has absolute minimizer on the non-negative orthant.

The previous examples demonstrate that, in the case of image reconstruction, ML problems are ill-posed or ill-conditioned. That means that one is not interested in computing

the minimum points $X^*$ of the functions corresponding to the different noise models because they do not provide sensible estimates $\bar{X}$ of the unknown object.

The previous remark is not surprising in the framework of inverse problem theory. Indeed it is generally accepted that, if the formulation of the problem does not use some additional information on the object, then the resulting problem is ill-posed. This is what happens in the maximum likelihood approach because we only use information about the noise with, possibly, the addition of the constraint of non-negativity.

The additional information may consist, for instance, of prescribed bounds on the solution and/or its derivatives up to a certain order (in general not greater than two). These prescribed bounds can be introduced in the problem as additional constraints in the variational formulation provided by ML. However, in a quite natural probabilistic approach, called the *Bayesian approach*, the additional information is given in the form of statistical properties of the object [20].

In other words, one assumes that the unknown object $X$ is a realization of a vector-valued random variable $O$, and that the probability distribution of $O$, the so-called *prior* denoted by $p_O(X)$, is also known or can be deduced from known properties of the object. The most frequently used priors are Markov random fields or, equivalently, Gibbs random fields, i.e. they have the following form:

$$p_O(X) = \frac{1}{Z} e^{-\mu \Omega(X)} \quad , \tag{44}$$

where $Z$ is a normalization constant, $\mu$ is a positive parameter (a hyperparameter in statistical language, a regularization parameter in the language of regularization theory), while $\Omega(X)$ is a function, possibly convex.

The previous assumptions imply that the joint probability density of the random variables $O, I$ is given by

$$p_{OI}(X, Y) = p_I(Y|X) p_O(X) \quad . \tag{45}$$

If we introduce the marginal probability density of the image $I$

$$p_I(Y) = \int p_{OI}(X, Y) \, dX \quad , \tag{46}$$

from *Bayes' formula* we obtain the conditional probability density of $O$ for a given value $Y$ of $I$:

$$p_O(X|Y) = \frac{p_{OI}(X, Y)}{p_I(Y)} = \frac{p_I(Y|X) p_O(X)}{p_I(Y)} \quad . \tag{47}$$

If in this equation we insert the detected value $Y$ of the image, we obtain the *a posteriori* probability density of $X$:

$$P_Y(X) = p_O(X|Y) = L_Y(X) \frac{p_O(X)}{p_I(Y)} \quad . \tag{48}$$

Then, a *maximum a posteriori* (MAP) estimate of the unknown object is defined as any object $X^*$ that maximizes the a posteriori probability density:

$$X^* = \arg\max_{X \in \mathbb{R}^n} P_Y(X) \quad . \tag{49}$$

As in the case of the likelihood it is convenient to consider the neglog of $P_Y(X)$. If we assume a Gibbs prior as that given in Eq. (44) and we take into account the definition of Eq. (38), we can introduce the following function

$$
\begin{aligned}
J(X;Y) &= -A \ln P_Y(X) + B - A \ln Z - \\
&- A \ln p_I(Y) = J_0(X;Y) + \mu J_R(X) \ ,
\end{aligned}
\tag{50}
$$

where $J_R(X) = A\Omega(X)$. Therefore the MAP estimates are also given by

$$
X^* = \arg \min_{X \in \mathbb{R}^n} \ J(X;Y)
\tag{51}
$$

and again one must look for the minimizers satisfying the non-negativity constraint.

## 5.2 The Richardson-Lucy Algorithm

One of the most frequently used methods for image deconvolution in astronomy is an iterative algorithm known as the *Richardson-Lucy* (RL) algorithm [37, 25]. In emission tomography it is also denoted as *expectation maximization* (EM) because, as shown in [38], it can be obtained by applying to the ML problem with Poisson noise a general EM method introduced in [15] for obtaining ML estimates.

In [38] it is shown that, if the iteration converges, then the limit is just a ML estimate in the case of Poisson data. Subsequently the convergence of the algorithm was proved by several authors in the case $B = 0$. An account can be found in [32].

The iteration is as follows: it is initialized with a positive image $X^0$ (a constant array, in general); then, given $X^n$, $X^{n+1}$ is computed by

$$
X^{n+1} = X^n H^T \frac{Y}{HX^n + B} \ .
\tag{52}
$$

This algorithm has some nice features. First, the result of each iteration is automatically a positive array; second, in the case $B = 0$, the result of each iteration has the same flux of the detected image $Y$, and this property is interesting from the photometric point of view.

The limit of the RL iteration is, in general, very noisy (see the remark at the end of Example 2), but a reasonable solution can be obtained by a suitable stopping of the algorithm before convergence. This can be seen as a kind of regularization [3]. An example of RL-reconstruction is shown in Fig. 10 (lower left panel).

Several iterative methods, modeled on RL, have been introduced for computing MAP estimates corresponding to different kinds of priors. A recent account can be found in [4].

## 5.3 Deconvolution with a Sparsity Prior

Another approach is to use the sparsity to model the data. A sparse model can be interpreted from a Bayesian standpoint, by assuming the coefficients $\alpha$ of the solution in the dictionary $\Phi$ follow a leptokurtic PDF with heavy tails such as the generalized Gaussian distribution form:

$$
\text{pdf}_\alpha(\alpha_1, \ldots, \alpha_K) \propto \prod_{k=1}^{K} \exp\left(-\lambda \|\alpha_i\|_p^p\right) \quad 0 \le p < 2 \ .
\tag{53}
$$

Between all possible solutions, we want the one which has the sparsest representation in the dictionary $\Phi$. Putting together the log-likelihood function in the case of Gaussian noise $\sigma$ and the priors on $\alpha$, the MAP estimator leads to the following optimization problem:

$$\min_{\alpha_1,\cdots,\alpha_K} \frac{1}{2\sigma} \|Y - \Phi\alpha\|^2 + \lambda \sum_{k=1}^{K} \|\alpha_k\|_p^p \ , 0 \leq p < 2 \ . \tag{54}$$

The sparsity can be measured through the $\|\alpha\|_0$ norm (i.e. $p = 0$). This counts in fact the number of non-zero elements in the sequence. It was also proposed to convexify the constraint by substituting the convex $\|\alpha\|_1$ norm for the $\|\alpha\|_0$ norm [9]. Depending on the $H$ operator, there are several ways to obtain the solution of this equation.

A first iterative thresholding deconvolution method was proposed in [40] which consists of the following iterative scheme:

$$X^{(n+1)} = P_+ \left( X^{(n)} + H^T \left( \mathbf{WDen}_{M^{(n)}} \left( Y - HX^{(n)} \right) \right) \right) \tag{55}$$

where $P_+$ is the projection on the cone of non-negative images. and **WDen** is an operator which performs a wavelet thresholding, i.e. applies the wavelet transform of the residual $R^{(n)}$ (i.e. $R^{(n)} = Y - HX^{(n)}$), thresholds some wavelet coefficients, and applies the inverse wavelet transform. Only coefficients that belong to the multiresolution support $M^{(n)}$ [40] are kept, while the others are set to zero. At each iteration, the multiresolution support $M^{(n)}$ is updated by selecting new coefficients in the wavelet transform of the residual which have an absolute value larger than a given threshold. The threshold is automatically derived assuming a given noise distribution such as Gaussian or Poisson noise.

More recently, it was shown [19, 12, 10] that a solution of Eq. 54 for $p = 1$ can be obtained through a thresholded Landweber iteration:

$$X^{(n+1)} = P_+ \left( \mathbf{WDen}_\lambda \left( X^{(n)} + H^T \left( Y - HX^{(n)} \right) \right) \right) \ , \tag{56}$$

with $\|H\| = 1$. In the framework of monotone operator splitting theory, it was shown that for frame dictionaries, a slight modification of this algorithm converges to the solution [10]. Extension to constrained non-linear deconvolution is proposed in [17].

**Constraints in the object or image domains**

Let us define the *object domain* $\mathcal{O}$ as the space in which the solution belongs, and the *image domain* $\mathcal{I}$ as the space in which the observed data belongs (i.e. if $X \in \mathcal{O}$ then $HX \in \mathcal{I}$). The constraint in (55) was applied in the image domain, while in (56) we have considered constraints on the solution. Hence, two different wavelet based strategies can be chosen in order to regularize the deconvolution problem. The constraint in the image domain through the multiresolution support leads to a very robust way to control the noise. Indeed, whatever the nature of the noise, we can always derive robust detection levels in the wavelet space and determine scales and positions of the important coefficients. A drawback of the image constraints is that there is no guarantee that the solution is free of artifacts such as ringing around point sources. A second drawback is that image constraints can be used only if the point spread function is relatively compact, i.e. does not smear the information over the whole image.

The property of introducing robust noise modeling is lost when applying the constraint in the object domain. For example, in the case of Poisson noise, there is no way (except using time consuming Monte Carlo techniques) to estimate the level of the noise in the solution and to adjust properly the thresholds. The second problem with this approach is that, in fact, we try to solve two problems simultaneously (noise amplification and artifact control in the solution) with one parameter (i.e. $\lambda$). The choice of this parameter is crucial, while such a parameter is implicit when using the multiresolution support.

Ideally, constraints should be added in both the object and image domains in order to better control the noise by using the multiresolution support and avoid artifact such ringing.
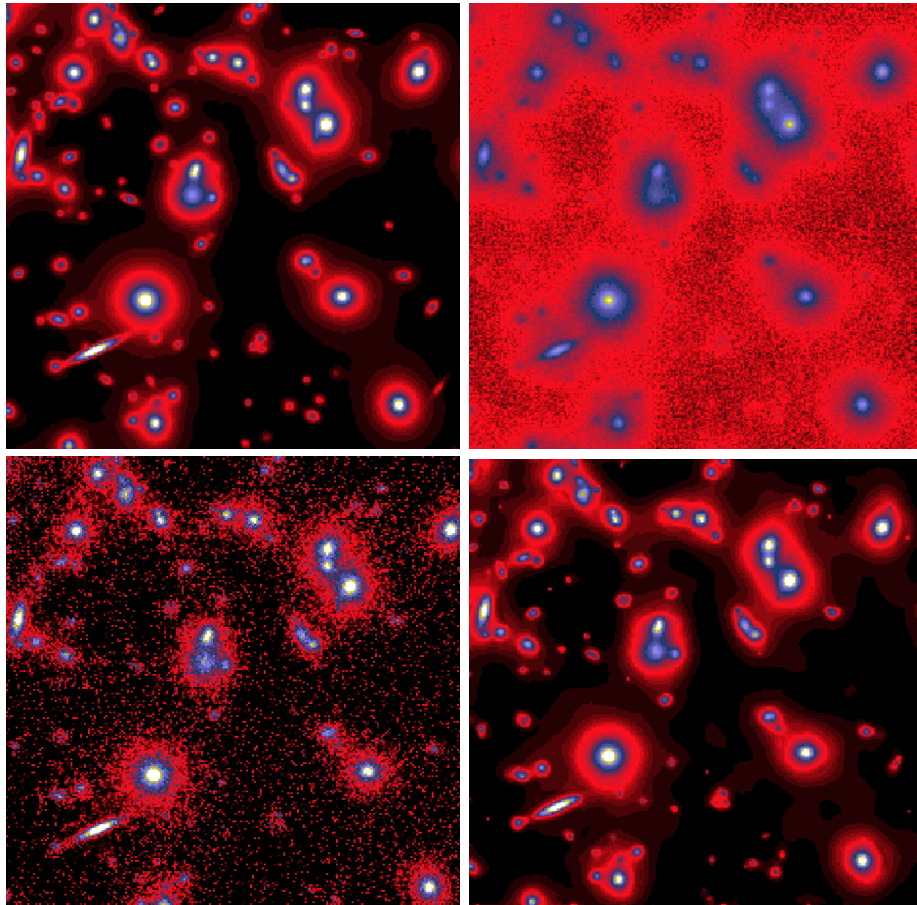
**Example**



Figure 14: Simulated Hubble Space Telescope Wide Field Camera image of a distant cluster of galaxies. Upper left: original, unaberrated and noise-free. Upper right: input, aberrated, noise added. Lower left: restoration, Richardson-Lucy. Lower right, restoration starlet-deconvolution.

A simulated Hubble Space Telescope Wide Field Camera image of a distant cluster of galaxies is shown in Fig. 14, upper left. The simulated data are shown in Fig. 14, upper right. The Richardson-Lucy and the wavelet solutions are shown respectively in Fig. 14,

lower left and right. The Richardson-Lucy method amplifies the noise, which implies that the faintest objects disappear in the deconvolved image, while the wavelet starlet solution is stable for any kind of PSF, and any kind of noise modeling can be considered.

## 5.4   Detection and Deconvolution

The PSF is not needed with MVM. This is an advantage when the PSF is unknown, or difficult to estimate, which happens relatively often when it is space-variant. However, when the PSF is well-determined, it becomes a drawback because known information is not used for the object reconstruction. This can lead to systematic errors in the photometry, which depends on the PSF and on the source signal-to-noise ratio. In order to preempt such a bias, a kind of calibration must be performed using simulations [39]. This section shows how the PSF can be used in the MVM, leading to a deconvolution.

### Object Reconstruction using the PSF

A reconstructed and deconvolved object $X_i$ can be obtained by searching for a signal $X_i$ such that the wavelet coefficients of $HX_i$ are the same as those of the detected structures $\alpha_i$. If $\mathbf{W}$ describes the wavelet transform operator, and $P_w$ the projection operator in the subspace of the detected coefficients, the solution is found by minimization of

$$\min_{X_i} \| \, \alpha_i - P_w \left( \mathbf{W} H X_i \right) \, \|^2 \tag{57}$$

where $\alpha_i$ represents the detected wavelet coefficients of the object $O_i$, and $H$ is the PSF. In this approach, each object is deconvolved separately. The flux related to the extent of the PSF will be taken into account. For point sources, the solution will be close to that obtained by PSF fitting. This problem is also different from global deconvolution in the sense that it is well-constrained. Except for the positivity of the solution which is always true and must be used, no other constraint is needed. This is due to the fact that the reconstruction is performed from a small set of wavelet coefficients (those above a detection limit). The number of objects is the same as those obtained by the MVM, but the photometry and the morphology are different. The astrometry may also be affected.

### The Algorithm

Any minimizing method can be used to obtain the solution $X_i$. Since there is no problem of convergence, noise amplification, or ringing effect, the Van Cittert method was proposed on the grounds of its simplicity [46]. It leads to the following iterative scheme:

$$X_i^{(n+1)} = X_i^{(n)} + \mathbf{R} \left( \alpha_i - P_w \left( \mathbf{W} H X_i^{(n)} \right) \right) \tag{58}$$

where $\mathbf{R}$ is the inverse wavelet transform, and the algorithm is:

1. Set $n$ to 0.

2. Find the initial estimation $X_i^{(n)}$ by applying an inverse wavelet transform to the set $\alpha_i$ corresponding to the detected wavelet coefficients in the data.

3. Convolve $X_i^{(n)}$ with the PSF $H$: $Y_i^{(n)} = HX_i^{(n)}$.

4. Determine the wavelet transform $\alpha^{(Y_i^{(n)})}$ of $Y_i^{(n)}$.

5. Threshold all wavelet coefficients in $\alpha^{(Y_i^{(n)})}$ at position and scales where nothing has been detected (i.e. $P_w$ operator). We get $\alpha_t^{(Y_i^{(n)})}$.

6. Determine the residual $\alpha_r = \alpha_i - \alpha_t^{(Y_i^{(n)})}$.

7. Reconstruct the residual image $R^{(n)}$ by applying an inverse wavelet transform.

8. Add the residual to the solution: $X_i^{(n+1)} = X_i^{(n)} + R^{(n)}$.

9. Threshold negative values in $X_i^{(n+1)}$.

10. If $\sigma(R^{(n)})/\sigma(X_i^{(0)}) < \epsilon$ then $n = n+1$ and go to step 3.

11. $X_i^{(n+1)}$ contains the deconvolved reconstructed object.

In practice, convergence is very fast (less than 20 iterations). The reconstructed image (not deconvolved) can also be obtained just by reconvolving the solution with the PSF.

## Space-Variant PSF

Deconvolution methods generally do not take into account the case of a space-variant PSF. The standard approach when the PSF varies is to decompose the image into blocks, and to consider the PSF constant inside a given block. Blocks which are too small lead to a problem of computation time (the FFT cannot be used), while blocks which are too large introduce errors due to the use of an incorrect PSF. Blocking artifacts may also appear. Combining source detection and deconvolution opens up an elegant way for deconvolution with a space-variant PSF. Indeed, a straightforward method is derived by just replacing the constant PSF at step 3 of the algorithm with the PSF at the center of the object. This means that it is not the image which is deconvolved, but its constituent objects.

## Undersampled Point Spread Function

If the PSF is undersampled, it can be used in the same way, but results may not be optimal due to the fact that the sampled PSF varies depending on the position of the source. If an oversampled PSF is available, resulting from theoretical calculation or from a set of observations, it should be used to improve the solution. In this case, each reconstructed object will be oversampled. Eq. (57) must be replaced by

$$\min_{X_i} \| \alpha_i - P_w \left( \mathbf{W} \mathcal{D}_l H X_i \right) \|^2 \tag{59}$$

where $\mathcal{D}_l$ is the averaging-decimation operator, consisting of averaging the data in the window of size $l \times l$, and keeping only one average pixel for each $l \times l$ block.
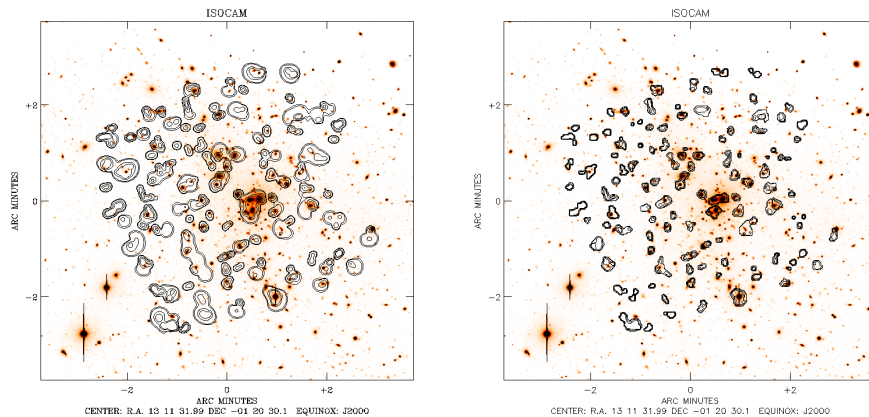
Figure 15: Abell 1689: left, ISOCAM source detection (isophotes) overplotted on an optical image (NTT, band V). The ISOCAM image is a raster observation at 7 $\mu$m. Right, ISOCAM source detection using the PSF (isophotes) overplotted on the optical image. Compared to the left panel, it is clearly easier to identify the detected infrared sources in the optical image.

**Example: Application to Abell 1689 ISOCAM Data**

Fig. 15 (left) shows the detections (isophotes) obtained using the MVM method without deconvolution on ISOCAM data. The data were collected using the 6 arcsecond lens at $6.75\mu$m. This was a raster observation with 10s integration time, 16 raster positions, and 25 frames per raster position. The noise is non-stationary, and the detection of the significant wavelet coefficients was carried out using the root mean square error map $R_\sigma(x,y)$ by the method described in [39]. The isophotes are overplotted on an optical image (NTT, band V) in order to identify the infrared source. Fig. 15 (right) shows the same treatment but using the MVM method with deconvolution. The objects are the same, but the photometry is improved, and it is clearly easier to identify the optical counterpart of the infrared sources.

# 6 Recommended Readings

- Mario Bertero and Patrizia Boccacci, *Introduction to Inverse Problems in Imaging*, Institute of Physics, 1998.

- Jean-Luc Starck and Fionn Murtagh, *Astronomical Data Analysis*, second edition, Springer, 2006.

- Stephane Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, third edition, 2008.

- Jean-Luc Starck, Fionn Murtagh and Jalal Fadili *Sparse Image & Signal Processing*, Cambridge University Press, in press.

# References

[1] F.J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 15:246–254, 1948.

[2] F. Benvenuto, A. La Camera, C. Theys, A. Ferrari, H. Lantéri, and M. Bertero. The study of an iterative method for the reconstruction of images corrupted by poisson and gaussian noise. *Inverse Problems*, 24(035016), 2008.

[3] M. Bertero and P. Boccacci. *Introduction to Inverse Problems in Imaging*. Institute of Physics, 1998.

[4] M. Bertero, P. Boccacci, G. Desiderá, and G. Vicidomini. Image deblurring with poisson data: from cells to galaxies. *Inverse Problems*, 25(123006), 2009.

[5] E. Bertin and S. Arnouts. SExtractor: Software for source extraction. *Astronomy and Astrophysics, Supplement Series*, 117:393–404, June 1996.

[6] A. Bijaoui. Sky background estimation and application. *Astronomy and Astrophysics*, 84:81–84, April 1980.

[7] A. Bijaoui and F. Rué. A multiscale vision model adapted to astronomical images. *Signal Processing*, 46:229–243, 1995.

[8] R. Buonanno, G. Buscema, C.E. Corsi, I. Ferraro, and G. Iannicola. Automated photographic photometry of stars in globular clusters. *Astronomy and Astrophysics*, 126:278–282, October 1983.

[9] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1999.

[10] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4):1168–1200, 2005.

[11] G.S. Da Costa. Basic photometry techniques. In S.B. Howel, editor, *ASP Conference Series 23, Astronomical CCD Observing and Reduction Techniques*, volume 23, page 90. Astronical Society of the Pacific, 1992.

[12] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57:1413–1541, 2004.

[13] E. Davoust and W.D. Pence. Detailed bibliography on the surface photometry of galaxies. *Astronomy and Astrophysics, Supplement Series*, 49:631–661, 1982.

[14] B. Debray, A. Llebaria, R. Dubout-Crillon, and M. Petit. CAPELLA: Software for stellar photometry in dense fields with an irregular background. *Astronomy and Astrophysics*, 281:613–635, January 1994.

[15] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc. B*, 39(1):1–38, 1977.

[16] S. Djorgovski. Modelling of seeing effects in extragalactic astronomy and cosmology. *Journal of Astrophysics and Astronomy*, 4:271–288, December 1983.

[17] F.-X. Dupé, M.J. Fadili, and J-L. Starck. A proximal iteration for deconvolving Poisson noisy images using sparse representations. *IEEE Transactions on Image Processing*, 18(2):310–321, 2009.

[18] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and Its Applications*. Kuwer Academic Press, 1996.

[19] M.A. Figueiredo and R. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003.

[20] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

[21] M. J. Irwin. Automatic analysis of crowded fields. *Monthly Notices of the Royal Astronomical Society*, 214:575–604, June 1985.

[22] R. G. Kron. Photometry of a complete sample of faint galaxies. *Astrophysical Journal Supplement Series*, 43:305–325, June 1980.

[23] M.J. Kurtz. Classification methods: an introductory survey. In *Statistical Methods in Astronomy*, pages 47–58. European Space Agency Special Publication 201, 1983.

[24] O. Lefèvre, A. Bijaoui, G. Mathez, J.P. Picat, and G. Lelièvre. Electronographic BV photometry of three distant clusters of galaxies. *Astronomy and Astrophysics*, 154:92–99, 1986.

[25] L.B. Lucy. An iteration technique for the rectification of observed distributions. *Astronomical Journal*, 79:745–754, 1974.

[26] S. J. Maddox, G. Efstathiou, and W. J. Sutherland. The APM Galaxy Survey – Part Two – Photometric corrections. *Monthly Notices of the Royal Astronomical Society*, 246:433, October 1990.

[27] S. Mallat. *A Wavelet Tour of Signal Processing, The Sparse Way*. Academic Press, 3rd edition, 2008.

[28] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[29] A.F.J. Moffat. A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry. *Astronomy and Astrophysics*, 3:455, December 1969.

[30] R. Molina, B.D. Ripley, A. Molina, F. Moreno, and J.L. Ortiz. Bayesian deconvolution with prior knowledge of object location – applications to ground-based planetary images. *Astrophysical Journal*, 104:1662–1668, October 1992.

[31] F. Murtagh, J.-L. Starck, and A. Bijaoui. Image restoration with noise suppression using a multiresolution support. *Astronomy and Astrophysics, Supplement Series*, 112:179–189, 1995.

[32] F. Natterer and F. Wûbbeling. *Mathematical Methods in Image Reconstruction*. SIAM, 2001.

[33] T. Naylor. An optimal extraction algorithm for imaging photometry. *Monthly Notices of the Royal Astronomical Society*, 296:339–346, May 1998.

[34] S. Okamura. Global structure of Virgo cluster galaxies. In *ESO Workshop On The Virgo Cluster of Galaxies*, pages 201–215, 1985.

[35] W.D. Pence and E. Davoust. Supplement to the detailed bibliography on the surface photometry of galaxies. *Astronomy and Astrophysics, Supplement Series*, 60:517–526, 1985.

[36] M. Pierre, I. Valtchanov, B. Altieri, S. Andreon, M. Bolzonella, M. Bremer, L. Disseau, S. Dos Santos, P. Gandhi, C. Jean, F. Pacaud, A. Read, A. Refregier, J. Willis, C. Adami, D. Alloin, M. Birkinshaw, L. Chiappetti, A. Cohen, A. Detal, P. Duc, E. Gosset, J. Hjorth, L. Jones, O. LeFevre, C. Lonsdale, D. Maccagni, A. Mazure, B. McBreen, H. McCracken, Y. Mellier, T. Ponman, H. Quintana, H. Rottgering, A. Smette, J. Surdej, J. Starck, L. Vigroux, and S. White. The XMM-LSS survey. Survey design and first results. *Journal of Cosmology and Astro-Particle Physics*, 9:11–+, September 2004.

[37] W.H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62:55–59, 1972.

[38] L.A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, MI-2:113–122, 1982.

[39] J.-L. Starck, H. Aussel, D. Elbaz, D. Fadda, and C. Cesarsky. Faint source detection in ISOCAM images. *Astronomy and Astrophysics, Supplement Series*, 138:365–379, 1999.

[40] J.-L. Starck, A. Bijaoui, and F. Murtagh. Multiresolution support applied to image filtering and deconvolution. *CVGIP: Graphical Models and Image Processing*, 57:420–431, 1995.

[41] J.-L. Starck, M. Elad, and D.L. Donoho. Redundant multiscale transforms and their application for morphological component analysis. *Advances in Imaging and Electron Physics*, 132, 2004.

[42] J.-L Starck, J. Fadili, and F. Murtagh. The undecimated wavelet decomposition and its reconstruction. *IEEE Transactions on Image Processing*, 16:297–309, 2007.

[43] J.-L. Starck and F. Murtagh. Image restoration with noise suppression using the wavelet transform. *Astronomy and Astrophysics*, 288:343–348, 1994.

[44] J.-L. Starck and F. Murtagh. Automatic noise estimation from the multiresolution support. *Publications of the Astronomical Society of the Pacific*, 110:193–199, 1998.

[45] J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer-Verlag, 2002.

[46] J.-L. Starck and F. Murtagh. *Astronomical Image and Data Analysis*. Springer, 2006. 2nd edn.

[47] J.-L. Starck, F. Murtagh, and A. Bijaoui. *Image Processing and Data Analysis: The Multiscale Approach*. Cambridge University Press, 1998.

[48] J.-L. Starck and M. Pierre. Structure detection in low intensity X-ray images. *Astronomy and Astrophysics, Supplement Series*, 128, 1998.

[49] J.-L. Starck, R. Siebenmorgen, and R. Gredel. Spectral analysis by the wavelet transform. *Astrophysical Journal*, 482:1011–1020, 1997.

[50] B. Takase, K. Kodaira, and S. Okamura. *An Atlas of Selected Galaxies*. University of Tokyo Press, 1984.

[51] M. Thonnat. INRIA Rapport de Recherche, Centre Sophia Antipolis, No. 387, 1985. Automatic morphological description of galaxies and classification by an expert system.

[52] A.N. Tikhonov, A.V. Goncharski, V.V. Stepanov, and I.V. Kochikov. Ill-posed image processing problems. *Soviet Physics – Doklady*, 32:456–458, 1987.

[53] M. Watanabe, K. Kodaira, and S. Okamura. Digital surface photometry of galaxies toward a quantitative classification. I. 20 galaxies in the Virgo cluster. *Astronomy and Astrophysics, Supplement Series*, 50:1–22, 1982.

[54] B. Zhang, M.J. Fadili, and J.-L. Starck. Wavelets, ridgelets and curvelets for Poisson noise removal. *IEEE Transactions on Image Processing*, 17(7):1093–1108, 2008.