Bayesian image reconstruction with space-variant noise suppression

J. Núñez[1] and J. Llacer[2]

[1] Departament d'Astronomia i Meteorologia, Universitat de Barcelona, Av. Diagonal 647, E-08028 Barcelona and Observatorio Fabra, Barcelona, Spain
   e-mail: jorge@fajnm1.am.ub.es
[2] Engineering Division, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720, U.S.A.
   e-mail: j_llacer@lbl.gov

Abstract. In this paper we present a Bayesian image reconstruction algorithm with entropy prior (FMAPE) that uses a space-variant hyperparameter. The spatial variation of the hyperparameter allows different degrees of resolution in areas of different statistical characteristics, thus avoiding the large residuals resulting from algorithms that use a constant hyperparameter. In the first implementation of the algorithm, we begin by segmenting a Maximum Likelihood Estimator (MLE) reconstruction. The segmentation method is based on using a wavelet decomposition and a self-organizing neural network. The result is a predetermined number of extended regions plus a small region for each star or bright object. To assign a different value of the hyperparameter to each extended region and star, we use either feasibility tests or cross-validation methods. Once the set of hyperparameters is obtained, we carried out the final Bayesian reconstruction, leading to a reconstruction with decreased bias and excellent visual characteristics. The method has been applied to data from the non-refurbished Hubble Space Telescope. The method can be also applied to ground-based images.

Key words: techniques: image processing — methods: data analysis

1. Introduction

Bayesian and maximum entropy solutions are increasingly being used for the reconstruction of images from noisy and incomplete data. The Bayesian strategy seeks the image of highest probability given the data. The Bayesian target function and the likelihood are related through Bayes' rule, which includes the probability distribution of the image, also known as image prior.

Send offprint requests to: J. Núñez

During the past several years a substantial amount of work has been carried out in image reconstruction in the areas of medical tomography and in optical astronomy. In the latter case, the discovery of spherical aberration in the Hubble Space Telescope in 1990 (White & Allen 1990; Hanisch & White 1993) generated a strong effort in the image reconstruction community. Since 1988 our group has been working on the development of statistically based algorithms for Image Reconstruction (Núñez 1993; Núñez & Llacer 1991, 1993a,b, 1994, 1995a,b; Llacer & Núñez 1990; Llacer et al. 1993). In particular, we have developed Bayesian algorithms with entropy prior (FMAPE), and methods based on feasibility and cross-validation in order to compute the balancing parameter between the entropy prior and the likelihood term.

In this paper we will describe the results of our effort in the reconstruction of optical astronomy data by a Bayesian method with a space variant hyperparameter which allows different degrees of resolution in different regions of the image. Regions representing bright objects, like stars, are allowed to have hyperparameters that lead to reconstructions near Maximum Likelihood, while extended objects are segmented into regions with hyperparameters adjusted in such way that lead to featureless uncorrelated normalized residuals with mean values near 1.0 in each region. A first implementation of the algorithm will be described with the reconstruction of data from the non-refurbished Hubble Telescope.

2. Notation and imaging model

The notation used in this paper is the following:

D	Number of detectors in the array
B	Number of elements (pixels) in the reconstruction
$p_j, j = 1, \cdots D$	projection (measured) data
$a_i, i = 1, \cdots B$	emission density in the image

	(parameters to be estimated)
f_{ji}	Point Spread Function (PSF) or probability that an emission in pixel i in the source be detected at detector j
$b_j, j = 1, \cdots D$	background in the data
$n_j, j = 1, \cdots D$	readout noise in the data
$C_j, j = 1, \cdots D$	detector gain corrections (flatfield)
$f'_{ji} = \frac{f_{ji}}{C_j}$	corrected PSF
$q_i = \sum_{j=1}^{D} f'_{ji}$	total detection probability for an emission from pixel i
$h_j = \sum_{i=1}^{B} f'_{ji} a_i + b_j$	forward projection or blurring operation
Q_i	prior distribution density or default image (if any)

let $\mathbf{p}, \mathbf{a}, \mathbf{f}, \mathbf{b}, \mathbf{n}, \mathbf{C}, \mathbf{f'}, \mathbf{q}, \mathbf{h}, \mathbf{Q}$ be the corresponding arrays.

We shall work with the following imaging model: an object emits light with an intensity given by a spatial distribution $\mathbf{a}$. The light is focused by the optical system over a detector array consisting of individual, discrete, independent detectors. Each detector has a different quantum efficiency characterized by a gain correction distribution $\mathbf{C}$. A certain background radiation $\mathbf{b}$, coming mainly from the sky but also from sources inside the detector system, is detected along with the spatial distribution $\mathbf{a}$. We assume that the detection process is Poisson distributed. Finally, the detector is read by an electronic process which adds a Gaussian readout noise $\mathbf{n}$ with zero mean and known standard deviation σ. The imaging equation corresponding to this model is:

$$\mathbf{f}' \, \mathbf{a} + \mathbf{b} + \mathbf{n} = \mathbf{p} \ . \tag{1}$$

Equation (1) in discrete form becomes:

$$\sum_{i=1}^{B} \frac{f_{ji}}{C_j} \, a_i + b_j + n_j = p_j \quad j = 1, \cdots, D \ .$$

Most imaging systems are described by Eq. (1), particularly those based on Charge Coupled Device (CCD) cameras, and Image Pulse Counting Systems (IPCS).

The background in Eq. (1) is an input in our algorithm. Some authors (Bontekoe et al. 1994; Narayan & Nityananda 1986) have raised questions about the introduction of the background in Eq. (1). Bontekoe et al. demonstrated that the solution depends on the background in standard Maximum Entropy Method (MEM) algorithm. However, in our approach, the background term includes not only light from the sky but also light from sources inside the camera. In the case of a CCD camera, the background term $\mathbf{b}$ can be considered as: $\mathbf{b} = \mathbf{b}_{\text{ext}} + \mathbf{b}_{\text{int}} + \mathbf{b}_{\text{dark}} + \mathbf{b}_{\text{bias}}$ (Snyder et al. 1993). The term $\mathbf{b}_{\text{ext}}$ is the external background radiation. The term $\mathbf{b}_{\text{int}}$ is the internal background radiation from luminiscent

radiation on the CCD chip itself. The term $\mathbf{b}_{\text{dark}}$ is the number of thermoelectrons that are generated by heat in the CCD and $\mathbf{b}_{\text{bias}}$ is the number of electrons that are due to bias or "fat zeros". Those terms are all Poisson distributed random variables and their sum can be represented by a single background term. We have not observed in our algorithm the background dependence effect reported by Bontekoe et al. (1994). In our opinion, if the background is large and acurately known, it is better to include it in the equation and in the reconstruction algorithm. Otherwise, it is always possible to set it to zero in the algorithm and reconstruct the background as part of the image.

3. The FMAPE algorithm with variable hyperparameter

3.1. The Bayesian approach

We use the Bayesian strategy to obtain an iterative algorithm for image reconstruction. The application of Bayes' theorem to the image reconstruction problem gives:

$$\mathbf{P}(\mathbf{a}|\mathbf{p}) = \frac{\mathbf{P}(\mathbf{p}|\mathbf{a})\mathbf{P}(\mathbf{a})}{\mathbf{P}(\mathbf{p})} \ . \tag{2}$$

The most probable image $\mathbf{a}$, given data $\mathbf{p}$, is obtained by maximizing $\mathbf{P}(\mathbf{a}|\mathbf{p})$ in (2) or the product $\mathbf{P}(\mathbf{p}|\mathbf{a})\mathbf{P}(\mathbf{a})$ since $\mathbf{P}(\mathbf{p})$ is constant.

3.2. The likelihood

The conditional probability $\mathbf{P}(\mathbf{p}|\mathbf{a})$ describes the noise in the data and its possible object dependence. It is fully specified in the problem by the likelihood function. As indicated above, we have two independent processes: image formation on the detector array and detector readout. Taking the two processes into account, the compound likelihood is (Núñez & Llacer 1993a):

$$\mathbf{L} = \mathbf{P}(\mathbf{p}|\mathbf{a}) = \prod_{j=1}^{D} \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \, e^{-\frac{(k-p_j)^2}{2\sigma^2}} e^{-h_j} \, \frac{(h_j)^k}{k!}$$

and its logarithm is:

$$\log \mathbf{L} = \sum_{j=1}^{D} \left[-\log(\sqrt{2\pi}\sigma) - h_j + \right.$$
$$\left. + \log \sum_{k=0}^{\infty} \left(e^{-\frac{(k-p_j)^2}{2\sigma^2}} \, \frac{(h_j)^k}{k!} \right) \right] \ . \tag{3}$$

This compound likelihood was first introduced for reconstructions of Poisson data in the presence of readout noise by Llacer & Núñez (1990). Snyder et al. (1993a) use the same likelihood form for CCD cameras. If the process were

pure Poisson (no readout noise), the logarithm of the likelihood would be the classical expression:

$$\log \mathbf{L} = \sum_{j=1}^{D} [-h_j + p_j \log h_j - \log(p_j!)] \ .$$ (4)

3.3. The prior

We use entropy to define the prior probability $\mathbf{P}(\mathbf{a})$ in a generalization of the concepts originally described by Frieden (1972).

Let N be the total energy (usually the number of counts or photons) in the object. Assume that there is an intensity increment Δa describing an intensity jump to which we can assign some appropriate physical meaning. Frieden (1972) describes Δa as the finest known intensity jumps that are possible in the object, but we prefer to relax that definition and consider it an adjustable hyperparameter of the Bayesian function.

Assume, in addition, that we have prior information regarding the statistical makeup of the object obtained, for example, from previous observations or from observations at other wavelengths from which one could infere realistic information about the makeup of the object at the wavelength at which the observation is made. If $Q_i \ i = 1, \cdots, B$ is the prior energy distribution, the prior probability of a photon going to pixel i is given by:

$$g_i = \frac{Q_i}{\sum_{l=1}^{B} Q_l} \ .$$

In that case, the a priori source probability of $\frac{N}{\Delta a}$ quanta distributing randomly over the B pixels is then characterized by the multinomial law (Lloyd 1980; Handbook of Applicable Mathematics, Vol II, p. 92) as:

$$P(\mathbf{a}) = \frac{(N/\Delta a)!}{\prod_{i=1}^{B}(a_i/\Delta a)!} \prod_{i=1}^{B}(g_i)^{a_i/\Delta a} \ .$$ (5)

The a priori probability $P(\mathbf{a})$ of (5) reflects the statistical distribution process of source units $\frac{a_i}{\Delta a}$ given the a priori probabilistic information map g_i.

If the prior information is not reliable, it is best to ignore it and give a uniform value to $Q_i = \frac{\sum_{l=1}^{B} Q_l}{B}$ that means $g_i = 1/B$. In that case, the multinomial distribution of (5) reduces to a normalized version of the Boltzmann distribution used by Frieden (1972).

Assume, also, that the intensity increment Δa is space variant in the object space. The introduction of the variability of Δa is the core of this paper. The parameter Δa controls the amount of prior information that is introduced into the solution. If the values of Q_i represent a uniform field, the parameter Δa will control the degree of smoothness of the solution. Its spatial variability will allow the local control of smoothness avoiding noise amplification in the different areas of the image.

Let $\Delta a_i \ i = 1, \cdots, B$ be the variable increment. Now, in place of $\frac{N}{\Delta a}$ units we distribute $\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}$ units. The a priori source probability of $\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}$ quanta distributing randomly over the B pixels is then:

$$P(\mathbf{a}) = \frac{(\sum_{i=1}^{B} \frac{a_i}{\Delta a_i})!}{\prod_{i=1}^{B}(\frac{a_i}{\Delta a_i})!} \prod_{i=1}^{B}(g_i)^{a_i/\Delta a_i} \ .$$ (6)

A natural choice for the prior energy distribution is:

$$g_i = \frac{\frac{Q_i}{\Delta a_i}}{\sum_{i=1}^{B} \frac{Q_i}{\Delta a_i}} \ .$$ (7)

Taking logarithms in (6)

$$\log \mathbf{P}(\mathbf{a}) = \log\left(\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}\right)! - \sum_{i=1}^{B} \log\left(\frac{a_i}{\Delta a_i}\right)! +$$
$$+ \sum_{i=1}^{B} \frac{a_i}{\Delta a_i} \log g_i \ .$$

To compute the factorials, we use Stirling's approximation

$$x! = (2\pi x)^{\frac{1}{2}} x^x e^{-x}; \quad \log(x!) \approx x \log x - x$$ (8)

in which we discarded the term $\frac{1}{2} \log 2\pi x$. This approximation is adequate for $x > 10$. Also, the use of the complete approximation would give a term in $\frac{1}{x}$ in the algorithms that would produce a numerical instability at the background pixels. In some cases (p.e. in regions of very low flux or background in photon counting devices) the approximation used may produce some error but we prefer to use this approximation to avoid the menctioned numerical instability. Using (8)

$$\log \mathbf{P}(\mathbf{a}) = \left(\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}\right) \log\left(\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}\right) - \sum_{i=1}^{B} \frac{a_i}{\Delta a_i} -$$
$$- \sum_{i=1}^{B} \left[(\frac{a_i}{\Delta a_i}) \log(\frac{a_i}{\Delta a_i}) - \frac{a_i}{\Delta a_i}\right] +$$
$$+ \sum_{i=1}^{B} \frac{a_i}{\Delta a_i} \log g_i$$

and taking into account (7) the logarithm of the probability $\mathbf{P}(\mathbf{a})$ is:

$$\log \mathbf{P}(\mathbf{a}) = \left(\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}\right) \log\left(\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}\right) -$$
$$- \sum_{i=1}^{B} \frac{a_i}{\Delta a_i} \log \frac{a_i Q'}{Q_i} \ ,$$ (9)

where

$$Q' = \sum_{i=1}^{B} \frac{Q_i}{\Delta a_i} \ .$$

If we do not consider the space variation of the intensity increment i.e. $\Delta a = $ const., and taking into account that

both $\sum a_i$ and $\sum Q_i$ are constant, the log of the prior probability is:

$$\log \mathbf{P}(\mathbf{a}) = -\sum_{i=1}^{B} \frac{a_i}{\Delta a} \log \frac{a_i}{Q_i} + \text{const. terms.} \qquad (10)$$

This form of the entropy that includes the parameter Δa and the spatial prior information $\mathbf{Q}$ is often called cross-entropy.

3.4. The general algorithm

As discussed in Sect. 3.1, the maximum *a posteriori* (MAP) probability will be obtained by maximizing $\mathbf{P}(\mathbf{a}|\mathbf{p}) = \mathbf{P}(\mathbf{a})\mathbf{P}(\mathbf{p}|\mathbf{a})$, or equivalently, the logarithm of that probability, $\log \mathbf{P}(\mathbf{a}|\mathbf{p}) = \log \mathbf{P}(\mathbf{a}) + \log \mathbf{L}$ with the constraint of energy conservation:

$$\sum_{i=1}^{B} q_i a_i = \sum_{j=1}^{D} p_j - \sum_{j=1}^{D} b_j \quad . \qquad (11)$$

Taking into account (3), (9) and (11), the Bayesian function to be maximized is:

$$BY = (\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}) \log(\sum_{i=1}^{B} \frac{a_i}{\Delta a_i}) - \sum_{i=1}^{B} \frac{a_i}{\Delta a_i} \log \frac{a_i Q'}{Q_i} +$$

$$+ \sum_{j=1}^{D} \left[-\log(\sqrt{2\pi}\sigma) - h_j + \log \sum_{k=0}^{\infty} \left(e^{-\frac{(k-p_j)^2}{2\sigma^2}} \frac{(h_j)^k}{k!} \right) \right] -$$

$$- \mu \left(\sum_{i=1}^{B} q_i a_i - \sum_{j=1}^{D} p_j + \sum_{j=1}^{D} b_j \right) \quad , \qquad (12)$$

where h_j is given by:

$$h_j = \sum_{i=1}^{B} f'_{ji} a_i + b_j$$

and μ is a Lagrange multiplier for the conserconservation of counts. Note that the relative weight of the two terms in the Bayesian function (12) is controlled by the set of hyperparameters Δa_i.

The nonlinear nature of the reconstruction problem described by (12) suggests an iterative algorithm. We have found that gradient methods, like steepest ascent or conjugate gradient algorithms, although well established, produce negative results for most of the pixels in the low light level part of the image. Also, the well-behaved Expectation Maximization (EM) algorithm -developed by Dempster et al. (1977) and extensively used for maximum likelihood reconstruction, after the work of Shepp & Vardi (1982)- is difficult to use in the Bayesian case with entropy because it requires the solution of transcendental equations in the M-step. Instead, we have found that the method of Successive Substitutions described by Hildebrand (1974) and Meinel (1986), and used by the authors in several previous papers, affords us greater flexibility than other methods and results in rapidly converging algorithms.

The Successive Substitutions method can be described as follows: given a series of equations on the unknowns $a_i, \cdots, a_B$ in the form

$$a_i = KF(\{a_m\}), \quad i = 1, \cdots, B \qquad (13)$$

where F is some function, $(\{a_m\})$ is the complete set of variables $a_i, \cdots, a_B$, and K is a normalization constant, (13) can be transformed into a recursive relation by

$$a_i^{(k+1)} = KF(\{a_m^{(k)}\}), \quad i = 1, \cdots, B \quad . \qquad (14)$$

Each of the new values of $a_i^{(k+1)}$ is calculated from all the known B values of $\{a_m^{(k)}\}$, and the complete set is updated at once. The constant K is obtained by invoking the energy conservation law:

$$\sum_{i=1}^{B} q_i a_i^{(k+1)} = \sum_{i=1}^{B} K q_i F(\{a_i^{(k)}\}) = \sum_{j=1}^{D} p_j - \sum_{j=1}^{D} b_j \quad . \quad (15)$$

To obtain the maximum of (12), we set $\partial BY/\partial a_i = 0$, taking into account that

$$\frac{\partial l}{\partial a_i} = \frac{\partial l}{\partial h_j} \frac{\partial h_j}{\partial a_i} = \frac{\partial l}{\partial h_j} f'_{ji} \quad ,$$

$$\frac{\partial BY}{\partial a_i} = -\frac{1}{\Delta a_i} \log \left(\frac{a_i}{Q_i} \frac{Q'}{\sum_{l=1}^{B} \frac{a_l}{\Delta a_l}} \right) +$$

$$+ \sum_{j=1}^{D} \left[-f'_{ji} + \frac{1}{\sum_{k=0}^{\infty} \left(e^{-\frac{(k-p_j)^2}{2\sigma^2}} \frac{(h_j)^k}{k!} \right)} \times \right.$$

$$\left. \times \sum_{k=0}^{\infty} \left(e^{-\frac{(k-p_j)^2}{2\sigma^2}} \frac{k(h_j)^{k-1}}{k!} f'_{ji} \right) \right] - \mu q_i = 0 \quad .$$

Multiplying and dividing by h_j in the last summation:

$$\frac{\partial BY}{\partial a_i} = -\frac{1}{\Delta a_i} \log \left(\frac{a_i}{Q_i} \frac{Q'}{\sum_{l=1}^{B} \frac{a_l}{\Delta a_l}} \right) -$$

$$- q_i + \sum_{j=1}^{D} \frac{f'_{ji} p'_j}{h_j} - \mu q_i = 0 \quad i = 1, \cdots, B \quad (16)$$

where

$$p'_j = \frac{\sum_{k=0}^{\infty} \left(k \, e^{-\frac{(k-p_j)^2}{2\sigma^2}} \frac{(h_j)^k}{k!} \right)}{\sum_{k=0}^{\infty} \left(e^{-\frac{(k-p_j)^2}{2\sigma^2}} \frac{(h_j)^k}{k!} \right)} \quad . \qquad (17)$$

We could now solve for a_i in the first term of (16) and obtain a relation of the type of (13), but an exponential function would then appear. That exponential function causes instability in the solution (see, e.g., Núñez & Llacer 1990). Instead we divide Eq. (16) by q_i to obtain

$$- \frac{1}{q_i \Delta a_i} \log \left(\frac{a_i}{Q_i} \frac{Q'}{\sum_{l=1}^{B} \frac{a_l}{\Delta a_l}} \right) +$$

$$+\frac{1}{q_i}\sum_{j=1}^{D}\frac{f'_{ji}p'_j}{\sum_{l=1}^{B}f'_{jl}a_l+b_j} \;=\mu+1 \quad i=1,\cdots,B \; . \qquad (18)$$

Multiplying and dividing by C_j in the first summation, then adding a constant C to both sides of the equation, raising both sides to the power n, and finally multiplying both sides by a_i results in

$$a_i\left[\mu+1+C\right]^n \;=\; a_i\left[\frac{1}{q_i}\sum_{j=1}^{D}\frac{f_{ji}p'_j}{\sum_{l=1}^{B}f_{jl}a_l+C_jb_j}\;-\right.$$
$$\left.-\frac{1}{q_i\Delta a_i}\log\left(\frac{a_i}{Q_i}\frac{Q'}{\sum_{l=1}^{B}\frac{a_l}{\Delta a_l}}\right)+C\right]^n$$
$$i=1,\cdots,B. \qquad (19)$$

The expression (19) is of the type indicated by (13). We can solve for the unknowns by the recursive relation (14), yielding the Bayesian maximum a posteriori algorithm with entropy prior (FMAPE) which is given by the iterative formula:

$$a_i^{(k+1)} \;=\; Ka_i^{(k)}\left[\frac{1}{q_i}\sum_{j=1}^{D}\frac{f_{ji}p'_j}{\sum_{l=1}^{B}f_{jl}a_l^{(k)}+C_jb_j}\;-\right.$$
$$\left.-\frac{1}{q_i\Delta a_i}\log\left(\frac{a_i^{(k)}Q'}{Q_i\sum_{l=1}^{B}\frac{a_l^{(k)}}{\Delta a_l}}\right)+C\right]^n$$
$$i=1,\cdots,B. \qquad (20)$$

In the case of $\Delta a = $ const., to compute $\log\mathbf{P}(\mathbf{a})$ in the Bayesian expresion (12) we should use (10) in place of (9). In that case, following the same steps as before, the algorithm (20) becomes:

$$a_i^{(k+1)} \;=\; Ka_i^{(k)}\left[\frac{1}{q_i}\sum_{j=1}^{D}\frac{f_{ji}p'_j}{\sum_{l=1}^{B}f_{jl}a_l^{(k)}+C_jb_j}\;-\right.$$
$$\left.-\frac{1}{q_i\Delta a}\left(\log\frac{a_i^{(k)}}{Q_i}+1\right)+C\right]^n \quad i=1,\cdots B. \; (21)$$

In (20) and (21) k is the index of the iteration and K is a constant to conserve the energy by the relation $\sum_{i=1}^{B}q_ia_i=\sum_{j=1}^{D}p_j$, computed at the end of each iteration. Since

$$K=\frac{1}{[\mu+1+C]^n} \quad ,$$

computing the constant K is equivalent to computing the Lagrange multiplier μ.

The iterative process defined by (20), (17) is the main result of this paper. We call this algorithm FMAPE. The algorithm has a number of desirable characteristics: It solves the cases of both pure Poisson data and Poisson data with Gaussian readout noise. The algorithm allows us to adjust the hyperparameters Δa_i in order to determine the degree of smoothing in different regions or, in connection with known default images, to adjust the proximity between the reconstruction and the default image $\mathbf{Q}$.

The algorithm maintains the positivity of the solution; it includes flatfield corrections; it removes background and can be accelerated to be faster than the Richardson-Lucy algorithm (Lucy 1974). The main loop (projection and backprojection) of the algorithm is similar in nature to the Expectation Maximization algorithm. The algorithm can be applied to a large number of imaging situations, including CCD and Pulse-counting cameras both in the presence and in the absence of background.

3.5. Positivity and Speed of convergence

The algorithm contains two arbitrary constants n and C. Both have been introduced in a way that the effect of changing the values of the constants n and C only affects the calculated values of K at the end of each iteration. The point of convergence of the algorithm (20), (17) is then solely dependent on the hyperparameters Δa_i. This can be seen as follows: Eq. (19) represents a set of B simultaneous equations in the unknowns a_i. Once the problem is solved by some appropriate method, Eq. (19) will hold, independently of the values of C and n. Equation (20) provides a method of solving the set of Eq. (19). If it converges to a solution of (19), that solution will therefore be independent of the values of C and n.

The constant C is introduced to insure the positivity of the solution. The presence of the negative entropy term

$$-\frac{1}{q_i\Delta a_i}\log\left(\frac{a_i^{(k)}Q'}{Q_i\sum_{l=1}^{B}\frac{a_l^{(k)}}{\Delta a_l}}\right)$$

on the right hand of (20) means that the positivity of the solution is not automatically guaranteed unless an appropriate constant is introduced. As indicated above, different choices of C will not affect the convergence point. Such a problem is often present in maximum entropy approaches, either by gradient or EM algorithms.

Since the first term in (20) is always positive, a value of C somewhat larger than

$$\max\left[-\frac{1}{q_i\Delta a_i}\log\left(\frac{a_i^{(k)}Q'}{Q_i\sum_{l=1}^{B}\frac{a_l^{(k)}}{\Delta a_l}}\right)\right] \quad i=1,\cdots,B$$

will suffice. In practice, we find that changing C can affect the speed of convergence. Increasing C by a factor of 2 over the approximate smallest value necessary to maintain non-negativity results in slowing down convergence by approximately the same factor. The use of $C = 1$ has usually resulted in adequate convergence speed and non-negativity of the pixels.

The constant n affects the rate of convergence. Since n affects the correction factor in (20), one can expect a range of values of n over which the iterative process is stable. We have observed that for $n = 1$, the convergence rate improves by a factor of approximately n with respect to the

rate for $n = 1$. At the beginning of the process, when the correcting factors are rather high, a large n could result in an instability of the algorithm. After several iterations, when the correcting factors are close to unity, large values of n can be used. We have obtained stability using up to $n \approx 3$ at the beginning of the iterative process and larger values later.

3.6. The modified data $p_j^{'}$

The complete form of the algorithm includes the computation of $p_j^{'}$ $j = 1, \cdots, D$ using (17) at each iteration. These variables were introduced in Llacer & Núñez (1990) for reconstructions in the presence of readout noise and further studied by Snyder et al. (1993a,b) and Núñez & Llacer (1993a). After every iteration, we must compute the modified data $p_j^{'}$ $j = 1, \cdots, D$ needed for the following iteration. This represents a pixel by pixel filtering operation in which the original data are substituted by a modified version.

In the absence of readout noise or when it is negligible, $\sigma \to 0$ in (17). Then the exponentials are dominant at $k = p_j$ and $p_j^{'} \to p_j$. We note that by its definition, $p_j^{'}$ $j = 1, \cdots, D$ is always positive, while the original data p_j $j = 1, \cdots, D$ can be negative. For example, if the background is small, as in the case of the Hubble Space Telescope, the mean background is just above zero, but due to the readout noise, we may find a large number of pixels with negative values. Negative values are not allowed in any algorithm with projection-backprojection operations, like in the classical Richardson-Lucy algorithm. A preprocessing of the data or other approximations are common. We do not need any approximation or preprocessing. In fact, $p_j^{'}$ $j = 1, \cdots, D$ is a representation of the data that is positive and close to the data, with a degree of closeness that depends on the projections h_j $j = 1, \cdots, D$ and σ. In this sense $p_j^{'}$ $j = 1, \cdots, D$ can be considered as a Bayesian filtered version of the data.

The practical computation of $p_j^{'}$ $j = 1, \cdots, D$ by Eq. (17) is not a trivial problem of numerical analysis. The $p_j^{'}s$ are numbers in the same range of the data, but to compute them, we need to compute summations from 0 to ∞. In addition, exponentials and factorials can easily give either machine overflow or underflow, even using quadruple precision. To solve the problem Snyder et al. (1993b) use Poisson and saddlepoint approximations. We preferred to use a recursive technique without any approximation with very good results. Also, it is easy to see that the summations from $k = 0$ to ∞ can be reduced to summations from $k = \max(0, p_j - n\sigma)$ to $k = p_j + n\sigma$ with $n \approx 3$ without loss of precision. By using this method, we can compute the $p_j^{'}$ values with high accuracy, without increasing the CPU time apreciably.

3.7. Maximum likelihood algorithms

We can obtain a maximum likelihood algorithm from the general FMAPE by taking the limit when $\Delta a_i \to \infty$ $i = 1, \cdots, B$. In that case the weight of the prior information (cross-entropy term) in the Bayesian function (12) is zero and the solution becomes a Maximum Likelihood one. The algorithm for the maximum likelihood (MLE) case in the presence of readout noise and background is:

$$a_i^{(k+1)} = K a_i^{(k)} \left[\frac{1}{q_i} \sum_{j=1}^{D} \frac{f_{ji} p_j^{'}}{\sum_{l=1}^{B} f_{jl} a_l^{(k)} + C_j b_j} \right]^n$$
$$i = 1, \cdots, B, \qquad (22)$$

where $p_j^{'}$ is given by Eq. (17).

In the case of no background and no readout noise, algorithm (22) becomes:

$$a_i^{(k+1)} = a_i^{(k)} \left[\frac{1}{q_i} \sum_{j=1}^{D} \frac{f_{ji} p_j}{\sum_{l=1}^{B} f_{jl} a_l^{(k)}} \right]^n \quad i = 1, \cdots, B. \quad (23)$$

For $n = 1$ and disregarding the gain (flatfield) corrections ($q_i = 1$), (23) is identical to the Richardson-Lucy algorithm.

4. Determination of hyperparameters

4.1. The set of hyperparameters

The hyperparameters Δa_i have a fundamental role in the Bayesian framework. They determine the relative weight between the prior information and the likelihood in the Bayesian function and therefore determine the chosen solution between the default image and the maximum likelihood solution. Ultimately, they define the degree of smoothness of the solution.

If we use a single value for all the hyperparameters ($\Delta a_i = $ const. $i = 1, \cdots, B$), we could obtain a reasonable global fit of the data, but the parts of the reconstructed image with low light levels would appear noisy (over-reconstructed) while bright objects (stars) would be too smooth (under-reconstructed). This effect is also present in the Richardson-Lucy algorithm when stopped according to some appropriate stopping rule and can easily be observed in the normalized residuals comparing the projection of the solution and the data. To compute the normalized residual over a set of S detectors we use the expression:

$$r = \frac{1}{S} \sum_{j=1}^{S} \frac{(h_j - p_j)^2}{h_j} \quad . \qquad (24)$$

The topic of the bias in the residuals has been addressed in the last few years by several authors (Lucy 1993; Núñez & Llacer 1993b; Starck et al. 1993; Starck & Pantin 1995;

White 1993), who developed different approaches in order to suppress noise amplification during the reconstruction process.

The FMAPE algorithm avoids noise amplification by the use of a set of space-variant hyperparameters. A variable Δa_i allows different degrees of smoothing in different regions, ranging from very smooth (or very similar to the prior $\mathbf{Q}$) regions ($\Delta a_i \rightarrow 0$) to maximum likelihood regions ($\Delta a_i \rightarrow \infty$).

In our implementation of the algorithm, we have computed the Δa_i for the different regions by using segmentation techniques in conjunction with: a) Regional cross-validation and b) Regional measurements of average normalized residuals. We present here a segmentation technique based on the wavelet decomposition and a method of classification (Kohonen self-organizing maps). However the problem of segmentation of an astronomical image may be as difficult as the reconstruction itself. The segmentation problem is, thus, open and improvements of the technique described here will be needed in the future.

Stars are essentially point sources but their large range of intensities does not allow us to consider all the stars as belonging to a single region. Thus, each star (or bright object) has been considered to be a different region. All other elements in the image should be separated into a reasonable number regions of similar characteristics for assignment of Δa_i values. What is meant by similar characteristics will be described in the next section.

As stated above, it is possible to use diverse techniques to segment the image. We have experimented with several methods, some of them from the field of medical imaging. However, at this stage, we prefer to use a general purpose method of classification. The self-organizing network (Kohonen 1990) is one of the best understood and it appears to be useful to clasify astronomical objects.

The hyperparameters Δa_i belong to the object space (as well as the emission densities a_i). Thus, the segmentation must be done in object space and not in data space. In most cases in astronomy the data (raw image) resemble the object and segmentation could be performed in data space, but this is incorrect (for example, in the problem of tomographic imaging the aspect of the data has nothing to do with the aspect of the object and segmenting the data makes no sense). Thus, some kind of previous deconvolution should be performed prior to the segmentation process. We use the Maximum Likelihood method (MLE or Richardson-Lucy) for this deconvolution because it is the most widely studied method of restoration in astronomy and it gives very good results. We stopped the algorithm using the cross-validation or feasibility tests for the whole image. Our starting point for the segmentation process is, thus, the result of a standard reconstruction (Richardson-Lucy method + statistically based stopping criteria).

4.2. Wavelet decomposition

We have assumed that regions with pixels having a similar mean value should form one region, though two regions with similar mean values but differing content of high frequencies should be different. The use of a wavelet decomposition at the larger scales provides the segmentation process with information about average values over extended regions and low frequencies, while the use of the smaller scales provides information about high frequency information. The joint use of all the scales helps us achieve the segmentation objective. As stated above, we have carried out the actual segmentation step by using a Kohonen self-organizing network. That method allows the automatic use of all the information contained in the wavelet decomposition of the image and results in excellent convergence. Then, in order to prepare the MLE image for the segmentation by a Kohonen self-organizing network, we represent it first by a block of planes containing its wavelet decomposition.

Following Starck & Murtagh (1994), we decompose the MLE image using the discrete wavelet transform algorithm known as "à trous" ("with holes"). An image $\mathbf{a}$ is successively transformed as follows:

$$f(\mathbf{a}) = \mathbf{a}_1, \ \ f(\mathbf{a}_1) = \mathbf{a}_2, \ \ f(\mathbf{a}_2) = \mathbf{a}_3, \ \ ....$$

We use a scaling function which has a $\mathbf{B}_3$ cubic spline profile. Letting

$$\mathbf{w}_l = \mathbf{a}_{l-1} - \mathbf{a}_l \ \ (l = 1, \cdots, n) \ ,$$

in which $\mathbf{a}_0 = \mathbf{a}$, we can write:

$$\mathbf{a} = \sum_{l=1}^{n} \mathbf{w}_l + \mathbf{a}_r \ . \tag{25}$$

In this representation, the images $\mathbf{a}_l \ \ (l = 0, \cdots, n)$ are versions of the original image at increasing scales (decreasing resolution levels), $\mathbf{w}_l \ \ (l = 1, \cdots, n)$ are the multiresolution wavelet planes and $\mathbf{a}_r$ a residual image. Specifically for our work we represent the MLE image using the first three wavelet planes plus a residual image:

$$\mathbf{a} = \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_3 + \mathbf{a}_r \ .$$

4.3. Separation of stars

As stated above, we first separate the stars and other prominent objects. There are several software packages to locate stars (INVENTORY, DAOPHOT, etc.) but we have found the separation using a wavelet plane to be practical and quantitatively and qualitatively easy to understand.

Which wavelet plane to use can depend on the size of the point spread function. In this paper we use the second wavelet plane $\mathbf{w}_2$ which has little noise but keeps enough signal to be a good representation of the stars and prominent objects. Also, in our experience, the plane $\mathbf{w}_2$

was always adequate for a wide range of PSF sizes coming from images from both the Hubble Space Telescope and the ground. The method consists simply of isolating the regions with a number of counts over certain threshold.

The threshold is set automatically. To do this, we use the IRAF routine "imstatistics" to compute the standard deviation σ of the data in the background and set the threshold to 4σ. The threshold parameter could be tuned by interaction with the user but this interaction would introduce a bias and a user-dependent solution.

4.4. Neural networks and segmentation

The MLE image is represented as a stack of 5 planes: the first three wavelet planes, the residual image and one plane containing only all the stars with an intensity value equal to the most intense pixel of any star. This choice insures that the Kohonen network will place all the stars in one single region, to be broken down into one region per star after the segmentation is carried out, and before the final reconstruction. In other words: we exclude the stars from this stage of the segmentation by forcing them into one single weight vector, while the Kohonen learning process proceeds to segment all the other extended areas. Each image pixel is thus represented by a "codebook" vector of dimension 5, (Kohonen 1990) of which we have as many pixels as in the MLE image. By using the three wavelet planes plus the residual image we prepare the data set for a segmentation that takes into consideration both the structural elements in the image and the average number of counts in different regions in some weighted manner.

The codebook vectors are presented to a self-organizing map with a square array of weight vectors which we have defined to be of dimensions 3×3, 4×4 or 5×5. The network uses square learning regions, with each plane of the image block normalized to a maximum range of 1.0. The "best matching" weight vector is selected by Euclidean distance. Before the training process starts, the order of the codebook vectors is randomized taking care that no vector is left out in any of the training cycles. The learning parameter starts at 0.1 and ends at 0.05 over 75 complete passes through the codebook vectors.

The outcome of the learning process is 9, 16 or 25 weight vectors that are representative of all the codebook vectors. Invariably, one of the vectors will represent all the stars (this is not the case if the extra plane with all the stars at the same high intensity is not included). A testing step in which all the codebook vectors are classified in terms of their proximity to the weight vectors follows the learning process. A map of regions is then generated, in which all MLE image pixels corresponding to a given weight vector appear with the same region number. The stars are made to have regions numbers above the 8, 15, or 24 of the smoother elements.

Thus, the final result of this process is a map of few extended regions (9, 16 or 25) plus one more region for each star or bright object.

4.5. Computation of the regional hyperparameters and final reconstruction

Once the segmentation process is completed, we can either compute the regional hyperparameters by cross-validation tests or adjust them dynamically by feasibility tests monitoring the residuals. In both cases, we carry out the final Bayesian reconstruction with the FMAPE algorithm given by Eqs. (20), (17) using the original data.

4.5.1. Feasibility tests

The method for computing the regional hyperparameters by feasibility tests is based on controlling the residuals of each region.

We start using an appropriate initial guess for Δa_i, indicated by experience. For astronomical images, an initial guess of $\Delta a_i = 100$ for all the pixels is generally a good choice. During the reconstruction, we allow each of the Δa_i values to change: At the end of each iteration, we use Eq. (24) to compute the mean normalized residuals of each region. Then, we adjust the corresponding value of Δa_i in such a manner that the residuals in each region approach 1.0, within a band of ± 0.05. This is achieved by the following simple algorithm: In the regions in which the residuals are higher than 1.05 we increase the Δa_i by a certain amount. Likewise, If the residuals are below 0.95, we decrease the Δa_i. These corrections are carried every few iterations. The reconstruction is finished when the image of the normalized residuals is featureless and the average residual values for all regions are within the prescribed band.

We have observed that the above methodology results in reconstructions in which the boundaries of each region become somewhat visible in the final results. In order to avoid this effect, we keep incrementing or decrementing a "master" map of Δa_i, but use a smoothed set of that current master map at each iteration. The smoothing is carried out with a Gaussian kernel of approximately one pixel of standard deviation.

4.5.2. Cross-validation tests

The cross-validation test is much more robust than the feasibility test. It is based in the concept of likelihood cross-validation introduced by Coakley (1991). The application of cross-validation to the reconstruction of astronomical images is described in Núñez & Llacer (1991, 1993a).

The method for obtaining the regional hyperparameters by cross-validation is the following: Consider the data

being split into two halves, A and B. This splitting can be obtained in two ways: a) as a result of two independent but successive exposures, and b) as a result of a thinning process if only one data set is available (see Núñez & Llacer 1993a).

Once we have obtained the two data sets, the method consists of the following steps: 1) Reconstruct data set A with an initially low uniform Δa_i for all the pixels, carrying out the reconstruction to near convergence. 2) For each segmented region compute the following two quantities: a) the log likelihood of data set A with respect to the reconstructed image A (direct likelihood), and b) the log likelihood of data set B with respect to the reconstructed image A (cross-likelihood). The computations of regional log likelihoods are carried out by Eq. (3) applied to the specific regions separately. If there is no readout noise, Eq. (4) can be used. 3) Increase the uniform Δa_i to a higher value and repeat steps 1) and 2) until the conditions described in the following paragraphs are met.

A plot of the values of the direct likelihood and cross-likelihood as a function of Δa_i should be made for each region. When the uniform values of Δa_i are low, the iterative reconstruction process carried out to convergence reconstructs mostly features that are common to data sets A and B (the signal), and both direct likelihood and cross-likelihood for all regions will increase with increasing values of the Δa_i. As the values of Δa_i increase further, the reconstruction of some regions of data set A will begin to yield results that correspond to the specific statistical noise of data set A, but do not correspond to the noise of data set B. The direct likelihood of those regions will always continue to increase with higher Δa_i, but the cross-likelihood will reach a maximum and start to decrease. Figure 7 shows the cross-validation test curve for region No. 3 of the first example presented in the next section.

For each region, then, the maximum of the cross-likelihood determines the proper value of Δa_i to be used in a final reconstruction. In some cases, mainly for some small regions corresponding to stars, the cross-validation curve does not show a maximum. In those cases a standard high value for Δa_i is used. Once the set of hyperparameters is obtained, we carry out the Bayesian reconstruction of data set A with the FMAPE algorithm given by Eqs. (20), (17).

After reconstructing data set A, the process could be repeated for data set B, obtaining the cross-likelihood curves with respect to data set A. In practice, once a set of Δa_i have been obtained from data set A, the same values can be used to reconstruct data set B. After obtaining the Bayesian reconstruction of the two data sets, both reconstructions are added to obtain the final result. During the final addition of the two half images, residual artifacts due to the reconstruction process that are not common to both reconstructions can also be easily eliminated.

The cross-validation method is robust for the following reasons: a) both data sets have the same signal; b) the
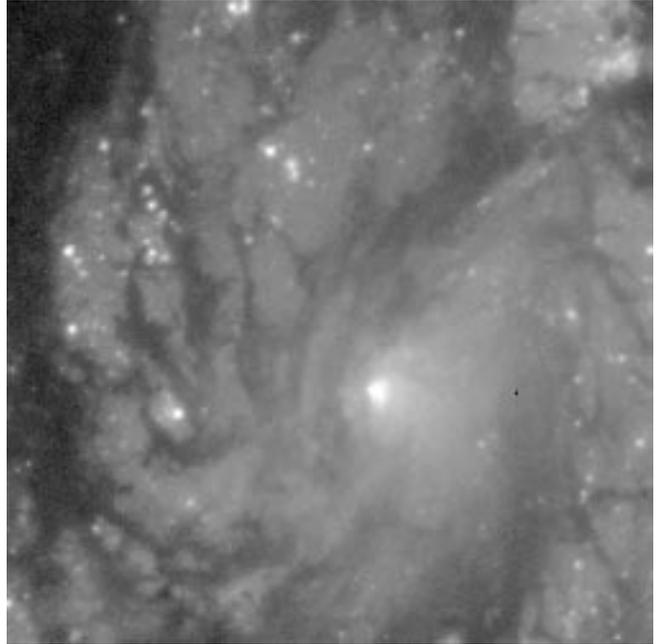

Fig. 1. True image of NGC 4321

statistical nature of the noise is identical in both cases; c) the noise is independent in the two data sets, and d) they have the same Point Spread Function.

5. Examples

We show here two reconstruction examples using the above methodology. For the first example we have worked with a 256×256 pixel image of the core of galaxy NGC 4321 (M100). The data were generated in the following way: an image of NGC 4321 obtained by the refurbished Hubble Space Telescope (HST) was used as source (true image), shown in Fig. 1. The raw data for the reconstruction were then obtained by convolving the true image with the Point Spread Function (PSF) of the aberrated HST, computed by the STScI TinyTim Program and adding Poisson noise. We generated two data sets: A and B using two different seeds to generate two independent Poisson noise data sets. Figure 2 show the raw image A. This example combines a diffused object and several bright sources. All the figures corresponding to this example are shown in a logarithmic scale, except where noted.

For the segmentation step, we first carried out a MLE reconstruction stopped after 20 iterations. The stopping point was given by the cross-validation test for the whole image. Then, we carried out the wavelet decomposition into three planes plus the residual image. Figure 3 shows wavelet plane $\mathbf{w}_2$ and Fig. 4 shows the residual image $\mathbf{a}_r$.

Wavelet plane $\mathbf{w}_2$ is used to identify the stars and other prominent features. Figure 5 shows the extracted
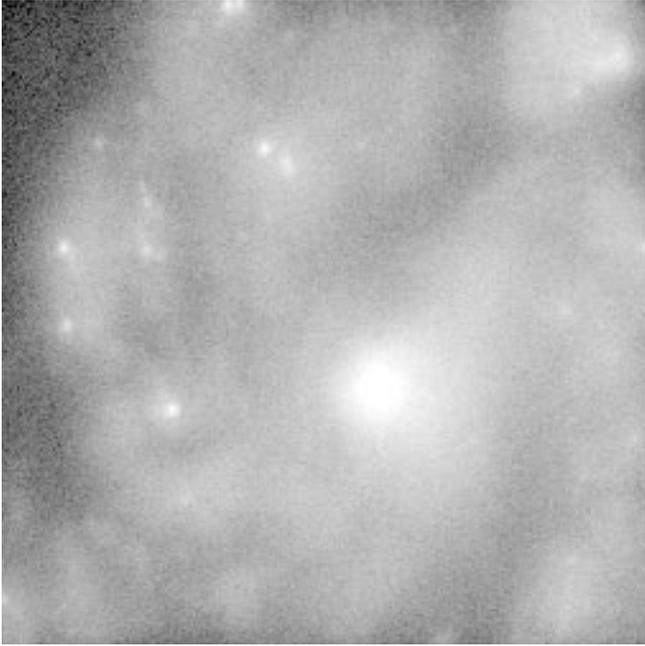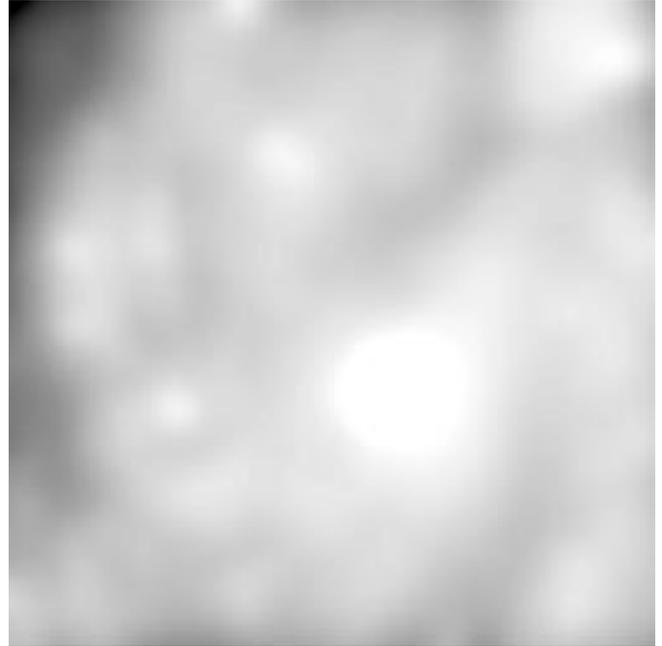
Fig. 2. Raw image of NGC 4321 to be reconstructed

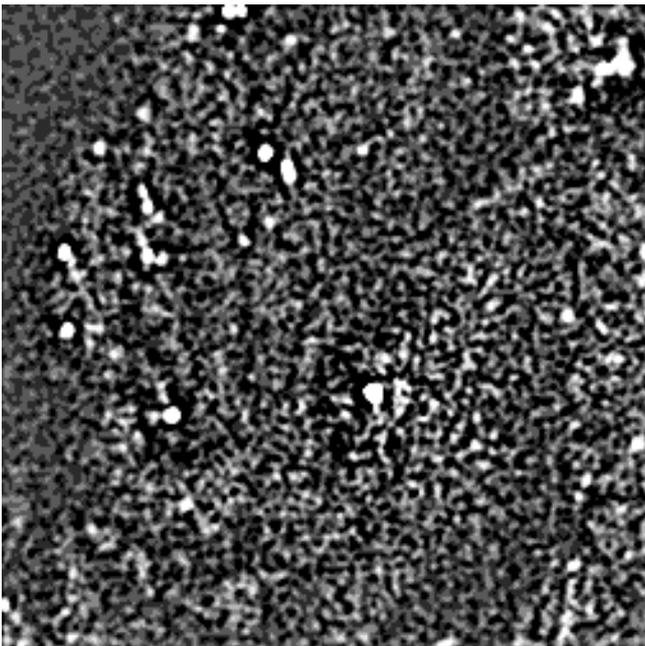**Fig. 4.** Residual image of the wavelet decomposition

Fig. 3. Second wavelet plane of the MLE reconstruction after 20 iterations

Fig. 5. Stars and other prominent objects extracted

objects obtained by setting the threshold to 22.8 counts. The threshold was computed using the automatic method described in Sect. 4.3. Finally, we carried out the segmentation using the self-organizing network described in Sect. 4.4. Figure 6 shows the final segmentation into 8 extended regions plus an additional region for each bright object totaling 50 regions.

The regional hyperparameters were computed by the cross-validation technique described in Sect. 4.5.2.

Figure 7 shows the cross-validation test curve for region No. 3 (corresponding to the spiral arms of the galaxy aproximately) from which an optimum value of $\Delta a_i = 65$ was obtained. In the 50 cross-validation tests carried out for this example, the following optimum values were obtained: for the eight extended regions $\Delta a_i = 20, 30, 70, 65, 60, 150, 70, 70$ respectively; for the other 42 regions (stars and bright objects) the Δa_i was set to the maximum of 300 in 29 cases while the other 13 objects received Δa_i values between 50 and 200.
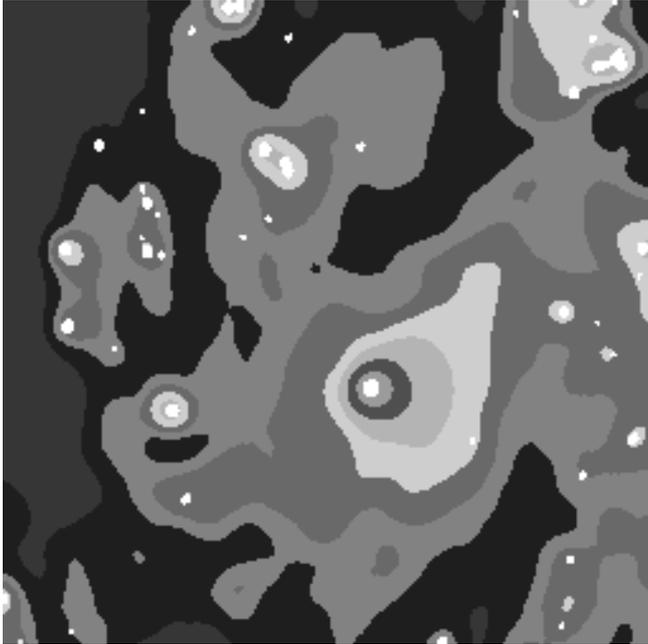
Fig. 6. Final segmentation

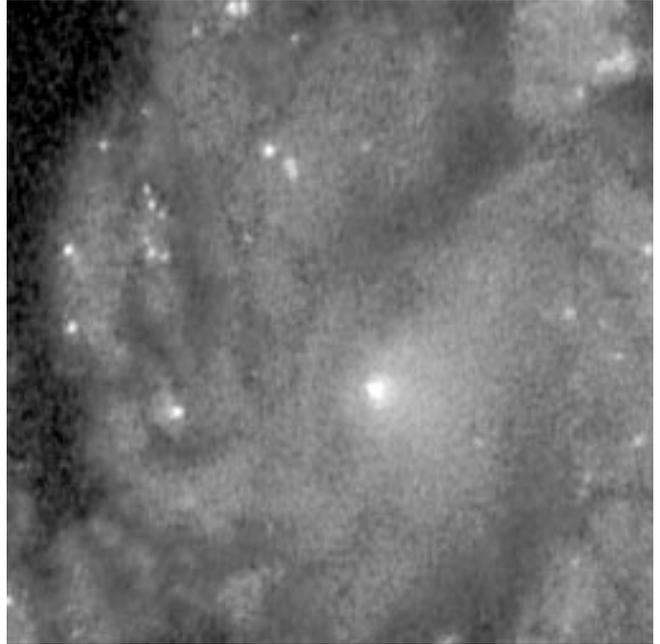**Fig. 8.** Final reconstruction of NGC 4321 using the FMAPE algorithm with variable hyperparameters defined by the segmentation

Fig. 7. Cross-validation test for region No. 3

Using the above set of hyperparameters, we carried out the final Bayesian reconstruction of data sets A and B with the FMAPE algorithm given by Eqs. (20), (17). After adding both reconstructions we obtained the final result, as shown in Fig. 8. A smooth background, a well reconstructed nebula, and sharp images of the stars are obtained. Noise amplification in the background and in the nebula was suppressed by the algorithm, while images of the bright objects were reconstructed to near Maximum Likelihood. Finally, Fig. 9 shows the image of the residuals in a linear scale, which exhibits no visible structure or correlation with the solution.

In order to compare the FMAPE method with variable hyperparameter with the method with a single hyperparameter given by algorithm (21) and with the Maximum Likelihood method given by algorithm (22), we carried out also the reconstructions using those methods. To make the results comparable, we reconstructed both data sets A and

B and added the results to obtain the final reconstructions. The single hyperparameter for Bayesian case (21), and the stopping point for the Maximum Likelihood method (22) were also determined by cross-validation for the entire image (Núñez & Llacer 1993a). We obtained a value of $\Delta a = 100$ for the Bayesian case and a stopping point of 30 iterations for the Maximum Likelihood method. Figure 10 and Fig. 11 show the results for the Bayesian with single hyperparameter and Maximum Likelihood cases respectively.

If we compare the results of the three methods, we can observe than the reconstruction with a single hyperparameter (Fig. 10 left) is noisier both in the nebula and in the background than in the variable hyperparameter case. Also, the images of the stars are under-reconstructed. The MLE image (Fig. 10 right) is clearly under-reconstructed in the bright objects. If we increase the number of iterations to correct that effect, the noise in the lower intensity regions increases excessively, making the result unacceptable. The Bayesian method with variable hyperparameter appears to be the better method of reconstruction.

The second example is the reconstruction of a 512×512 pixel real image of the planet Saturn obtained with the WF/PC camera of the Hubble Space Telescope, before the servicing mission. All images for this example are shown in a linear grey scale, except where noted. Figure 12 shows the raw data to be reconstructed. The Point Spread Function is shown in a logarithmic scale in Fig. 13.

Since the image of Saturn has no stars we segmented the image into 9 extended regions. For the segmentation step, we first carried out a MLE reconstruction stopped
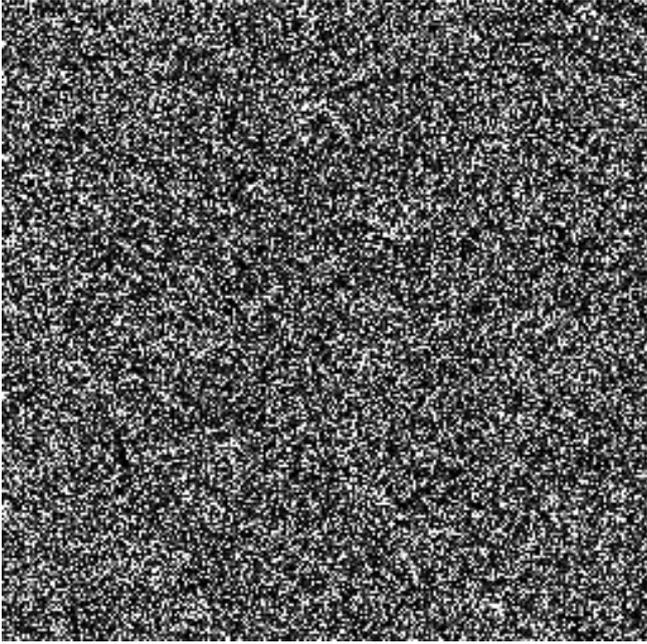
Fig. 9. Map of residuals of the reconstruction using the FMAPE algorithm with variable hyperparameters (linear scale)
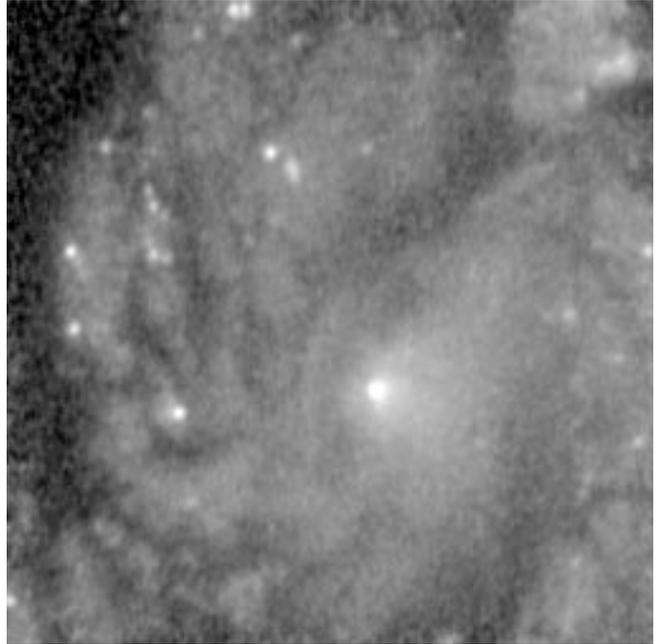

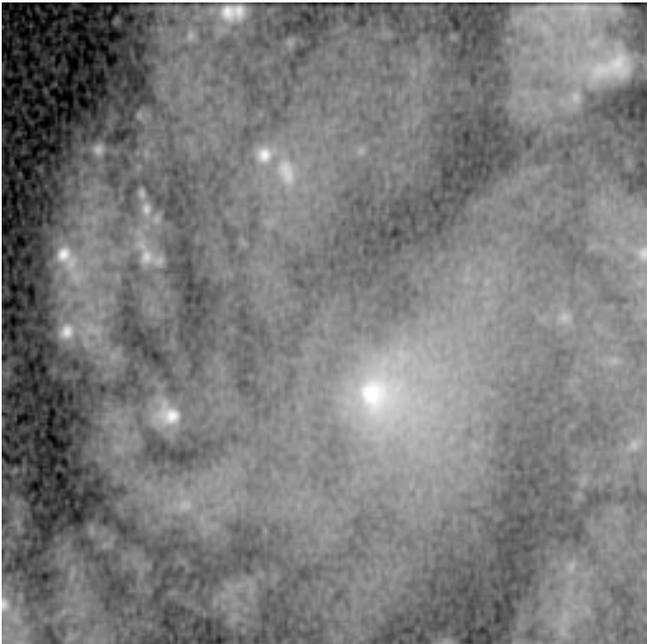**Fig. 11.** Reconstruction of NGC 4321 using the MLE algorithm

Fig. 10. Reconstruction of NGC 4321 using the FMAPE algorithm with simple hyperparameter

Fig. 12. Raw image for the planet Saturn

after 50 iterations. The stopping point was given by the feasibility test for the whole image. Then, as in the first example, we decomposed the image into three planes plus the residual image and carried out the segmentation by the self-organizing network described in Sect. 4.4. Figure 14 shows 9 regions resulting from the segmentation. In this example we do not have a second data set for cross-validation. In addition, the image was obtained from a

CCD camera with readout noise, rendering the process of splitting the image into two halves by thinning questionable (Núñez & Llacer 1993a). Thus, in this example we have used the feasibility test described in Sect. 4.5.1 to determine the set of hyperparameters.

Figure 15 shows the result of the 9-region reconstruction. We have obtained a smooth background, a well reconstructed image of the planet and sharp divisions in the rings. By the use of the variable resolution FMAPE algorithm, we have avoided noise amplification in all regions of the image.

Again, we compare the results of our new methodology with the constant Δa approach and with the MLE method. For that purpose, we used the Saturn data for reconstructions by the FMAPE algorithm with constant values of $\Delta a = 200$ and 500, and a reconstruction by the Maximum Likelihood method at 100 iterations. The results show characteristics similar to the ones described in the previous example: the solution with variable
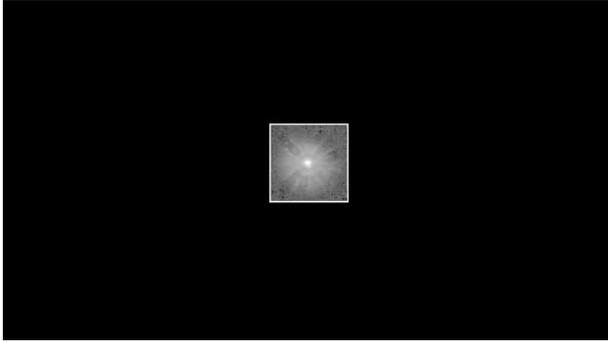
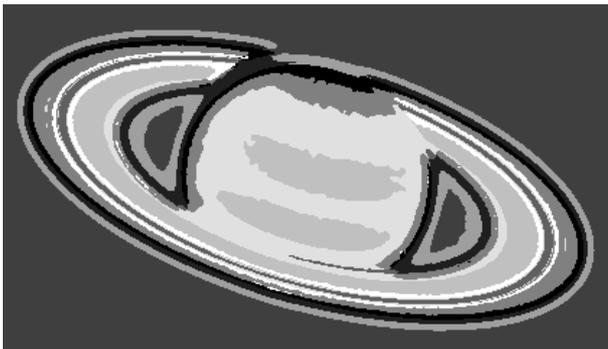**Fig. 13.** Observed PSF used for the reconstruction

Fig. 14. Segmentation of the image in 9 regions of Bayesian hyperparameters using the method described in the text
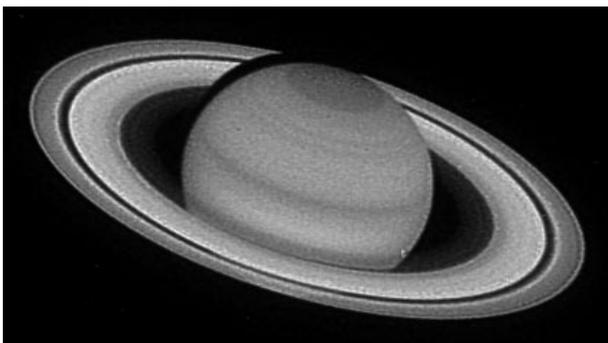

Fig. 15. Reconstruction of the image of Saturn using the FMAPE algorithm with the 9 channels defined by the segmentation

hyperparameter is better reconstructed and noise amplification is well controlled. In order to characterize the solutions, we computed the mean normalized residuals for each of the 9 regions of the segmentation. Figure 16 shows the plot of the mean residual as a function of region number. An ideal reconstruction would give a straight line at a mean value of 1.0. The reconstruction obtained by MLE has mean residuals that are increasing from below to above the 1.0 value. The reconstruction with a single uniform $\Delta a = 200$ has mean residuals that are too high. With $\Delta a = 500$ most of the values are too low but some are still too high. However, the reconstruction with the vari-
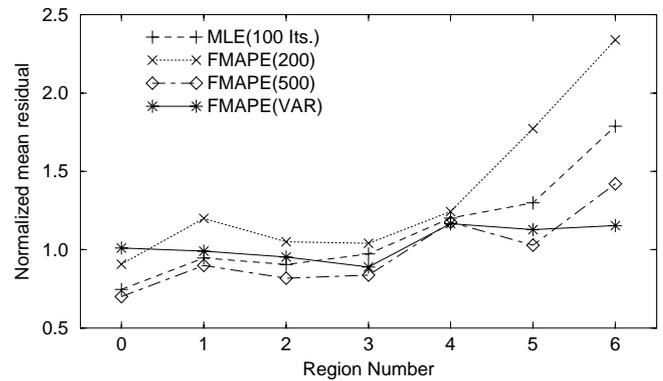

Fig. 16. Mean residuals for the 9 regions obtained with the FMAPE algorithm with variable hyperparameter in comparison with other methods. An ideal reconstruction would give a straight line at normalized mean residual 1.0

able Δa_i, although not perfect, is much improved, with most of the mean residuals close to the ideal value of 1.0.

6. Conclusions

In this paper we report the development of a Bayesian algorithm with entropy prior (FMAPE) that uses a space-variant balancing hyperparameter. In order to define the hyperparameters, we have used a segmentation technique based on the wavelet decomposition and artificial neural networks. We have found the Kohonen self-organizing maps quite suitable for clasification of astronomical objects, although it is important that the network implementation be carefully done with regards to plane normalization and randomization of the input vectors.

The combined use of the FMAPE algorithm and the segmentation allows us to obtain different degrees of smoothing in different regions of the image. This implies a variation in resolution in the reconstruction that can prevent amplification of noise while maintaining the photometric accuracy in the whole image (background, diffuse objects and stars).

In astronomical image reconstruction there are two basic requirements: 1 - the recovery of the object shape and 2 - the recovery of the correct brightness of the objects without amplifying the noise. The second requirement is largely related to the correctness of the residuals between the data and the projection of the solution. The method described here goes a long way towards satisfying those two needs.

We would like to conclude by indicating that we continue working on the better understanding of what the FMAPE requires from the segmentation and from the hyperparameter assignment steps with a view to provide a reliable method for reconstruction of images with a wide range of intensity values. Since there is no proof of uniqueness of the solutions obtained by the FMAPE with

space-variant hyperparameters, we also need to explore the effects of using different starting images, different initial values for the hyperparameters and updating schemes.

References

Bontekoe Tj.R., Koper E., Kester D.J.M., 1994, A&A 294, 1037

Coakley K.J., 1991, IEEE Trans. Nuclear Sci. NS-38, 1, 9

Dempster A.D., Laird N.M., Rubin D.B., 1977, J. Roy. Statist. Soc. B39, 1

Frieden B.R., 1972, J. Opt. Soc. Amer. 62, 511

Hanisch R.J., White R.L. (eds.), 1993, The restoration of HST Images and Spectra, STScI, Baltimore, U.S.A.

Hildebrand F.B., 1974, Introduction to Numerical Analysis 2d. New York: McGraw-Hill

Kohonen T., 1990, Proceedings of the IEEE 78, No. 9, 1464

Llacer J., Núñez J., 1990, in: White R.L., Allen R.J. (eds.). The restoration of HST Images and Spectra. STScI, Baltimore, p. 62

Llacer J., Veklerov E., Coakley K.J., Hoffman E.J., Núñez J., 1993, IEEE Trans. Medical Imaging 12, No. 2, 215

Lloyd E., 1980, Handbook of Applicable Mathematics, Vol. II. Chichester: John Wiley & Sons

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

Lucy L.B., 1993, in: Hanisch R.J., White R.L. (eds.). The restoration of HST Images and Spectra II. STScI, Baltimore, p. 79

Meinel E.S., 1986, J. Opt. Soc. Amer. A, 3, No. 6, 787

Narayan R., Nityananda R., 1986, ARA&A 24, 127

Núñez J., 1993, Newslett. STScI's Image Rest. Proj. 1, 24

Núñez J., Llacer J., 1990, IEEE Trans. Medical Imaging, 9, 159

Núñez J., Llacer J., 1991, PASPC 25, 210

Núñez J., Llacer J., 1993a, PASP 105, 1192

Núñez J., Llacer J., 1993b, in: Hanisch R.J., White R.L. (eds.) The restoration of HST Images and Spectra II. STScI, Baltimore, p. 123

Núñez J., Llacer J., 1994, in: Bernardo J.M., Berger J., Dawid A.P., Smith A.F.M. (eds.) Fifth International Meeting on Bayesian Statistics. Oxford University Press, p. 713

Núñez J., Llacer J., 1995a, in: Pitas I. (ed.) Proceedings of the 1995 IEEE Workshop on Nonlinear Signal and Image Processing. Neos Marmaras-Halkidiki, Greece, p. 106

Núñez J., Llacer J., 1995b, Vistas Astron. 40, No. 4, 547

Shepp L.A., Vardi Y., 1982, IEEE Trans. Medical Imaging, MI-1, 113

Snyder D.L., Hammoud A.M., White R.L., 1993a, J. Opt. Soc. Amer. A 10, 1014

Snyder D.L., Helstrom C.W., Lanterman A.D., Faisal M., White R.L., 1993b, in: Hanisch R.J., White R.L. (eds.). The restoration of HST Images and Spectra II. STScI, Baltimore, p. 139

Starck J.L., Murtagh F., 1994, A&A 288, 342

Starck J.L. Murtagh F., Bijaoui A., 1993, in: Hanisch R.J., White R.L. (eds.). The restoration of HST Images and Spectra II. STScI, Baltimore, p. 111

Starck J.L., Pantin E., 1995, Vistas Astron. 40, No. 4, 563

ter Haar Romeny B.M., Florack L.M.J., Salden A.H., Viergever M.A., 1993, Lect. Notes Comp. Sci. 687, 77

White R., 1993, in: Hanisch R.J., White R.L. (eds.). The restoration of HST Images and Spectra II. STScI, Baltimore, p. 104

White R.L., Allen R.J. (eds.), 1990, The restoration of HST Images and Spectra. STScI, Baltimore