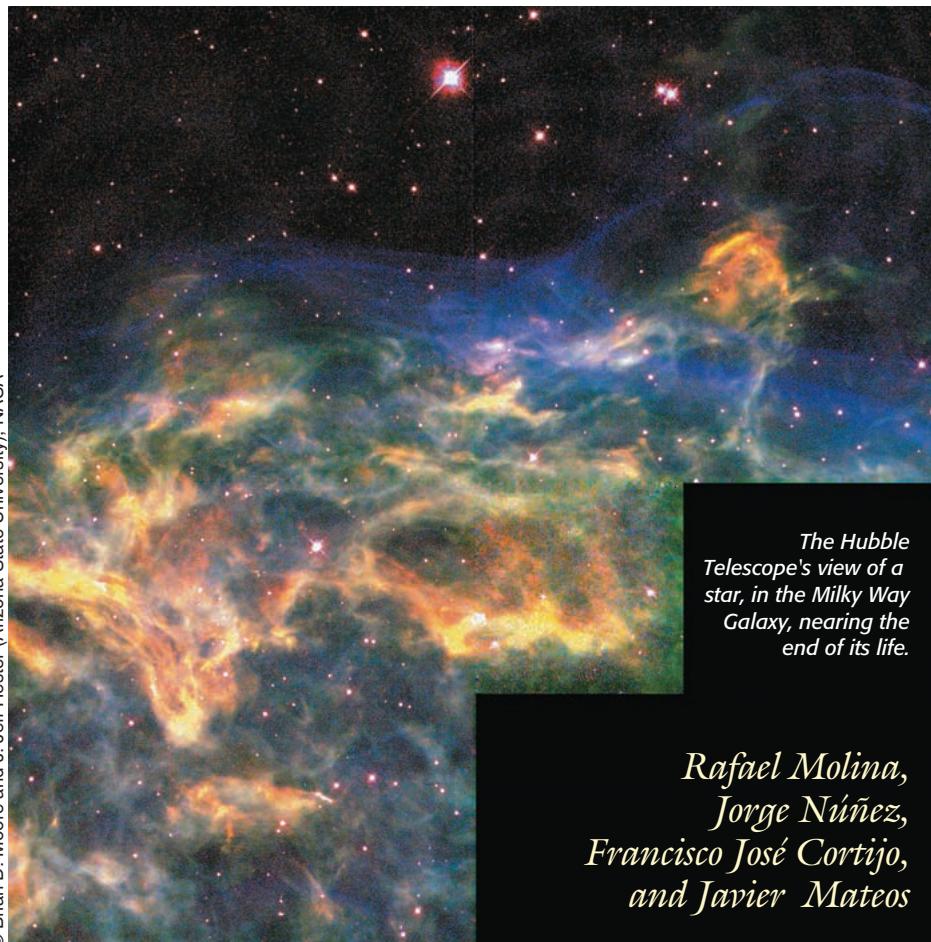


Image Restoration in Astronomy

A Bayesian Perspective



*Rafael Molina,
Jorge Núñez,
Francisco José Cortijo,
and Javier Mateos*

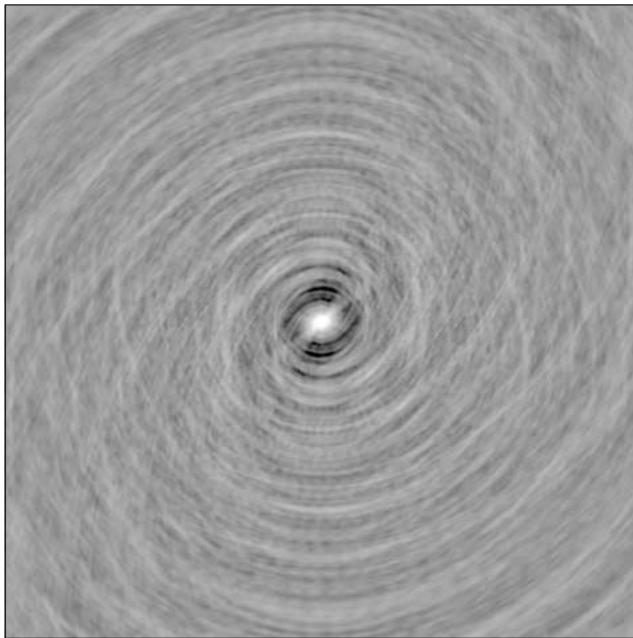
The field of digital image restoration has a quite long history that began in the 1950s with the space program. The first images of the Earth, Moon (mainly of the opposite side), and planet Mars were, at that time, of unimaginable resolution. The images were obtained under big technical difficulties, however, such as vibrations, bad pointing, motion due to spinning, etc. These difficulties resulted, in most cases, in medium to large degradations that could be scientifically and economically devastating. The need to retrieve as much information as possible from such degraded images was the aim of the early efforts to adapt the one-dimen-

sional signal processing algorithms to images, creating a new field that is today known as digital image restoration and reconstruction.

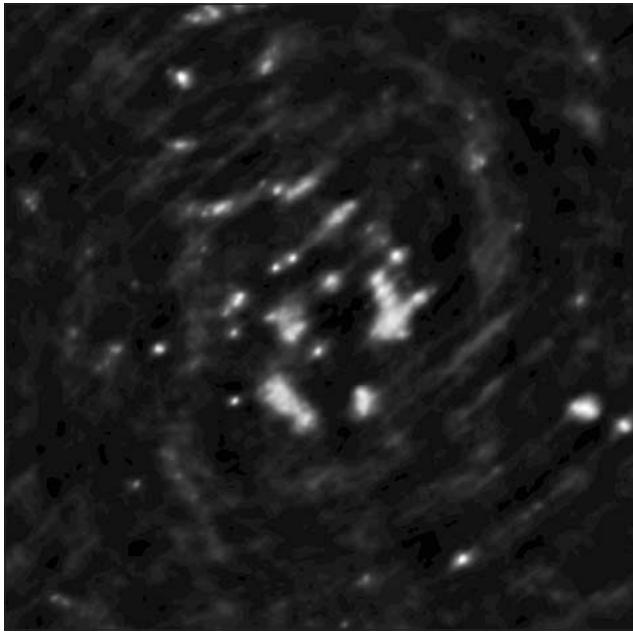
The application of early image restoration techniques to these images was very successful. If we compare the raw data obtained by the spacecraft with the “final” products (obtained also by adding several images) one could think that they were obtained using different equipment and in different epochs.

Since their introduction, the techniques of image reconstruction and restoration have been a “must” in all scientific disciplines involving projections or interferometric

data as medical tomography, seismology, magnetic resonance imaging, and even some astronomical applications such as mapping in radio astronomy. If we look at the “dirty map” of a radio-interferometric observation (see Fig. 1) obtained with the very large array (VLA), it is almost impossible to distinguish anything but the image of the beam. After the restoration process, however, the images of tiny extended sources (Fig. 2) can be easily recognized.



▲ 1. Dirty map of the radio-interferometric observation of the star LSI +61~303 obtained with the VLA. The image is in logarithmic scale.



▲ 2. Restoration of the image shown in Fig. 1 by the FMAPE method. The image of the star itself was removed to show the extended sources. The image is in logarithmic scale.

For a long time, image restoration was considered a luxury in other fields such as optical astronomy. Nonetheless image restoration was applied to images coming from space such as the case of the images taken in 1986 of comet Halley by the spacecraft Vega and Giotto. Again, the raw images of the comet obtained, for example, by the Vega mission have almost nothing to do with the elegant “peanut shape” images of the comet Halley nucleus.

In 1990 something happened which changed the situation of image restoration in the field of optical astronomy. After the launch of the \$2000 million Hubble Space Telescope (HST), an “impossible” mistake was discovered in the main mirror. The mirror had a severe problem of spherical aberration because it was polished with the help of a faulty device and checked with the same faulty device. Thus, the checking was perfectly coherent with the polishing but the curvature of the mirror was wrong (see Fig. 3). Since a single minute of observing telescope time cost about \$100,000, any effort to improve the images was cheap. Since then, a substantial amount of work has been done in image restoration directed towards optical astronomy (see [70], [16], [46], [1], and [19]). As result of such efforts, it was possible to correct the aberrant HST images.

Today, restoration is routinely carried out on many astronomical observations. As another example, the final product of the HST Deep Space Survey (HST-DSP), recently delivered, is an elegant and colorful series of very deep images of a selected area. The raw data (needed for some astronomical measurements) contain a great number of cosmic-ray hits, readout noise, low signal to noise ratio, etc. Again the final product requires restoration, dithering, and co-addition.

The restoration techniques used have improved enormously since the beginnings of the digital image restoration era. The goal of this article is to present some of the methods most widely used to restore images in optical astronomy.

To better understand the image restoration problem, it is important to recognize it as a particular case of the more general field of inverse problems which arise in many scientific fields, with astronomy being a prime example. In this field, one of the fundamental characteristics of the inverse problems is that the theoretical understanding of the formation of astronomical entities is limited, which precludes the direct approach of predicting the experimental data using a theoretical model developed from first principles.

The restoration problem can be cast in the form of a Fredholm integral equation of the first kind:

$$\mathbf{g}(x) = \int \mathbf{H}(x, y) \mathbf{f}(y) dy \quad (1)$$

in which $\mathbf{f}(y)$ is the function of interest (image to be recovered), $\mathbf{g}(x)$ is the function accessible to measurement (observed image), and $\mathbf{H}(x, y)$ is the kernel of the integral equation (point spread function). Equation (1) is the imaging equation of the restoration problem.

Since the data are obtained by measurements and therefore subject to accidental errors, the direct inversion of the ill-posed problem represented by (1) magnifies the accidental errors giving unacceptable results. This fact forces us to recognize that the image restoration problem should be treated by the methods of statistical estimation theory or regularization principles. The statistical estimation methods most used in astronomy are the maximum likelihood method (MLE), the maximum entropy (ME), and the Bayesian methods. MLE seeks the maximization of the likelihood function, resulting in an image for which the measurements have the highest probability. Given the ill-posed nature of the image reconstruction problem, however, the maximum likelihood solution is (as the direct inversion of the imaging equation) very sensitive to noise. To avoid this problem, it is necessary to stop the iterations of the algorithm before reaching the point of maximum likelihood.

Another approach is to use Bayesian probability theory to define a different target function to be maximized. Bayesian solutions are increasingly being used for the restoration of images from noisy and incomplete data. The process 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.

Most MLE, ME, and regularization-based algorithms can be seen as special cases of the Bayes theorem: in the MLE case by adopting a flat distribution as prior, in the case of ME by adopting an entropic form as prior distribution, and in the case of regularization, the constraints defining the prior model. Thus, we will use the Bayesian strategy as a framework to present the different approaches to manage the image restoration problem. To examine the correct likelihood function is also necessary, since in optical astronomy different types of cameras are being used.

When preparing an article on image restoration in astronomy, it is obvious that some topics have to be dropped to keep the work at reasonable length. We have decided to concentrate on image and noise models and on the algorithms to find the restoration. Topics like parameter estimation and stopping rules are also commented on but much more briefly. Because of the length of this article, we have decided to omit a section on multichannel image restoration in astronomy; the interested reader can consult [19] for a recent, although short, overview. We include a sidebar on "Software and Online Information in Astronomy."

We will start by describing the Bayesian paradigm and then proceed to study the noise and blur models used by the astronomical community. Then the prior models used to restore astronomical images are examined. In the next section we describe the algorithms used to find the restoration for the most common combinations of degradation and image models. Then we briefly comment on important issues such as acceleration of algorithms, stop-

ping rules, and parameter estimation. We comment on the huge amount of information available to, and made available by, the astronomical community.

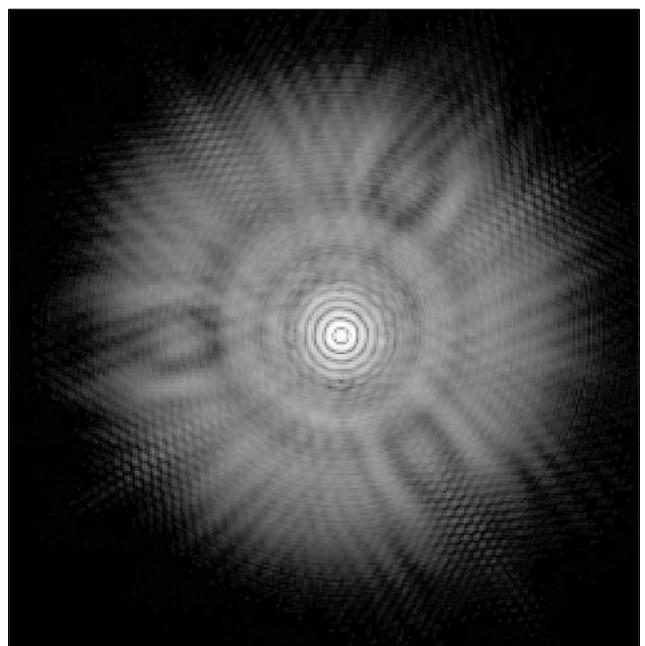
Bayesian Methods

The philosophy within statistics known as Bayesian inference has a very long history, but has only relatively recently been used explicitly in image processing in astronomy. It is distinguished from the perhaps more familiar classical statistical ideas by using prior information about the images being studied.

Image restoration in astronomy involves considering the image being made up of point sources (stars) and smoothly varying luminosity. Bayesian methods are distinguished from others most particularly by the inclusion of specific spatial information about the objects present in the image.

We will distinguish between f , the "true" image which would be observed under ideal conditions (i.e., no noise and no distortions produced by blurring and instrumental effects), and g , the observed image. The aim is then to reconstruct f from g . Bayesian methods start with a prior distribution, a probability distribution over images f . It is here where we incorporate information on the expected structure within an image. It is also necessary to specify $P(g|f)$, the probability distribution of observed images g if f were the true image. The Bayesian paradigm dictates that inference about the true f should be based on $P(f|g)$ given by

$$P(f|g) = P(g|f)P(f) / P(g) \propto P(g|f)P(f). \quad (2)$$



▲ 3. Blurring of the HST Faint Object Camera working at f/96 modeled by Tiny Tim. The blurring is shown in logarithmic scale and corresponds to the aberrated optics (before the servicing mission which added correction lenses).

To show just one restoration it is common (but not obligatory) to choose the mode of $P(\mathbf{f}|\mathbf{g})$, that is, to display the image $\hat{\mathbf{f}}$ which satisfies

$$\hat{\mathbf{f}} \text{ maximizes } P(\mathbf{g}|\mathbf{f})P(\mathbf{f}). \quad (3)$$

This is known as the maximum *a posteriori* (MAP) estimate of \mathbf{f} .

Equivalently, we can choose $\hat{\mathbf{f}}$ as

$$\hat{\mathbf{f}} \text{ minimizes } -\log P(\mathbf{g}|\mathbf{f}) - \log P(\mathbf{f}). \quad (4)$$

The first term of (4) will be familiar as the data log-likelihood. The second term can be thought of as a “roughness penalty,” as images \mathbf{f} which do not correspond to our prior conceptions will have been assigned a small $P(\mathbf{f})$ and hence a large penalty.

In statistical physics it is common to define probabilities by the “energy,” U , of a system, so that

$$P(\mathbf{f}) \propto \exp[-\alpha U(\mathbf{f})], \quad (5)$$

where α is $1/kT$, with T being temperature and k Boltzmann’s constant. If we adopt this notation, then

$$\hat{\mathbf{f}} \text{ minimizes } -\log \text{likelihood} + \alpha U(\mathbf{f}). \quad (6)$$

We can recognize this as a Lagrangian form, in other words as a regularization method, so its solution is equivalent to solving

$$\max \text{ likelihood} \text{ subject to } \text{energy} \leq \text{constraint} \quad (7)$$

and to

$$\min \text{ energy} \text{ subject to } \text{likelihood} \geq \text{constraint}. \quad (8)$$

As we have already mentioned, many other deconvolution principles fit into one of other of these forms, in particular maximum entropy (see [45] for a review) and the classical Richardson-Lucy restoration method [57], [26], where a flat improper distribution over all images is used as prior model.

Obviously, not all the methods presented here have been developed within the Bayesian paradigm. Their authors have used different approaches to tackle the restoration problem. In any case, we will try to describe them within the Bayesian paradigm and also provide information about the approach used. Using the Bayesian paradigm as framework for this article will allow us to describe in a systematic way all the ingredients of an image restoration problem in astronomy, that is, blur and noise models, image models and optimal restoration, in some statistical sense.

Having described the Bayesian process to estimate the real underlying image we proceed to study in depth the different observation models, the image models, and the

restoration algorithms that have appeared in the literature to restore astronomical images.

Observation Model

The modeling of the observation model can be divided in two parts: blurring function, in other words, point spread function (PSF), and noise modeling. Let us start with the blurring function.

Blurring Function

Ideal image formation, without noise, can be modeled as the convolution of the true brightness distribution of the object with the PSF. The PSF is the image of an ideal point source such as a star.

The PSF fitted on each particular image must be based on two main goals: to estimate its functional form and to estimate the flux from each star in the image. For ground-based telescopes, a detailed review of the atmospheric processes which lead to the distortion has been presented in [73] (see also [74]). No exact expression describing the shape of the PSF is known. However, studies presented in [5], [10], [33], and [41]-[43] have suggested a radially symmetric approximation for the PSF h of the form

$$h(r) \propto \left(1 + \frac{r^2}{R^2}\right)^{-\beta}. \quad (9)$$

This can be checked, and β and R chosen, by extracting what are clearly stars from a displayed image and fitting this function by weighted nonlinear least squares. The full width half maximum (FWHM) for this function has the value $\text{FWHM} = 2R\sqrt{2^{1/\beta} - 1}$. In Fig. 4 we plot the light profile of a star together with the fit to the PSF in (9) as an example of the good fit of this analytical function.

Other authors prefer to model the blurring process due to atmospheric turbulence by a Gaussian function of the form

$$h(r) \propto \exp\left[-\frac{r^2}{2\sigma^2}\right] \quad (10)$$

where σ^2 determines the severity of the blur (see [25]).

The Telescope Image Modeling (TIM) software was developed at the Space Telescope Science Institute (STScI) (<http://www.stsci.edu>) to model the output of an imaging telescope viewing a general object with a detector. Although the software is rather general, it was specifically developed with the science instruments of the HST viewing a star field in mind. It is used, among other things, to test data reduction algorithms and evaluate photometric reduction accuracy as well as to generate theoretical PSFs for deconvolution purposes. A similar use is made of the package Tiny TIM also developed at STScI. Fig. 3 depicts a PSF generated by the Tiny TIM software.

Since PSF modeling is not the main topic of this article, we will proceed to model the noise. The interested reader can examine, for instance, the contributions presented in [16] on PSF modeling.

We note that if $p = m \times n$ is the size of the image, the blurring process can be described by $\mathbf{H}\mathbf{f}$, where \mathbf{f} is a $p \times 1$ vector and \mathbf{H} is the $p \times p$ point spread matrix defining the systematic blur (for simplicity we are assuming that the sizes of the observed and restored images are the same). Since each detector can have a different quantum efficiency characterized by a gain correction distribution C , we will have

$$\mathbf{H} = (\text{diag}(C))^{-1} \mathbf{B}, \quad (11)$$

where $\text{diag}(C)$ is a diagonal matrix with entries c_j and \mathbf{B} is the blurring matrix modeled by (9) and (10) or a software package.

So, \mathbf{H} models the so-called flatfield corrected PSF, and we will assume that the gain correction distribution C is known. Together with $\mathbf{H}\mathbf{f}$ a certain background radiation \mathbf{b} , coming mainly from the sky but also from sources internal to the detector is normally detected. The blurring together with the background radiation can be described as $\mathbf{H}\mathbf{f} + \mathbf{b}$ where \mathbf{b} is a known column vector of size $p \times 1$. We will assume that the background radiation is known and will remove it from the observation process (see [47]).

Noise Models

Let us now proceed to examine the noise models used in the astronomical community. For the observed vector \mathbf{g} we could use the normal distribution $\mathbf{g} \sim N(\mathbf{H}\mathbf{f}, \sigma_n^2 \mathbf{I})$ and so we would have

$$P(\mathbf{g}|\mathbf{f}) \propto \beta^{p/2} \exp\left[-\frac{\beta}{2} \|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2\right], \quad (12)$$

with $\beta = 1/\sigma_n^2$.

From here we can extend the degradation model to having the form

$$P(\mathbf{g}|\mathbf{f}) \propto \exp\left[-\|\mathbf{g} - \mathbf{H}\mathbf{f}\|_{A(\mathbf{f})}^2\right], \quad (13)$$

where $A(\mathbf{f})$ is a weighting matrix whose form is discussed in [22] and [23]. Two points are worth mentioning; first, this sort of model also appears, as we will see later, when using signal-dependent noise and, second, the normalizing constant of the conditional distribution depends on the signal, which is not always used when using the regularization approach to image restoration.

From the degradation model in (12), the maximum residual likelihood (MRL) criterion [50] has been proposed to restore astronomical images being, in fact, an alternative degradation model. Let $n_i, i=1,\dots,p$, be p independent Gaussian variables with zero mean and vari-

ance σ_n^2 , Piña and Puettter [50] considered the autocorrelation function defined by

$$A_n(k) = \sum_{j=1}^p n_{j+k} n_j = \sum_{j=1}^p C_j^n(k). \quad (14)$$

They showed that the distribution of $A_n(k), k \neq 0$, approaches a Gaussian distribution due to the central limit theorem and that the statistic

$$\chi^2_{A_n} = \sum_{k=1}^m \frac{[A_n(k) - E(A_n(k))]^2}{\sigma_{A_n}^2(k)} \quad (15)$$

follows an χ^2 distribution with m degrees of freedom. The values of $E(A_n(k))$ and $\sigma_{A_n}^2(k)$ can be found in [50], together with some comments on how m should be selected taking into account the blurring function.

Then, for an image restoration problem we can define the residual image given by

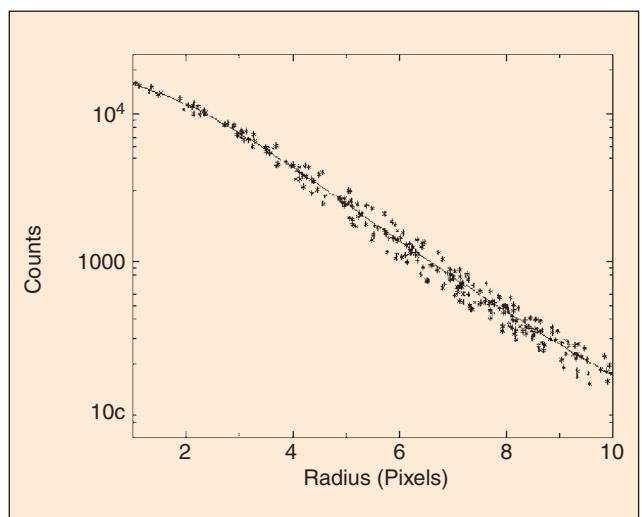
$$\mathbf{r} = \mathbf{g} - \mathbf{H}\mathbf{f} \quad (16)$$

and form the statistic

$$E_r = \chi^2_{A_r} = \sum_{k=1}^p \frac{[A_r(k) - E(A_r(k))]^2}{\sigma_{A_r}^2(k)}. \quad (17)$$

If the restored image is an accurate representation of the true image, r_i will show the behavior of independent Gaussian distribution $n_i, i=1,\dots,p$, with zero mean and standard deviation $\beta^{-1/2}$. We can then use $P_{\chi^2}(E_r)$ as $P(\mathbf{g}|\mathbf{f})$. We note that for this $P(\mathbf{g}|\mathbf{f})$, if $P(\mathbf{f}) \propto \text{const}$, the MAP and MRL images are the same.

We now move on to consider signal-dependent noise. To take into account the charge-coupled device detectors used in astronomy, each component of the observed vector \mathbf{g} is assumed to follow a Poisson distribution, and so



▲ 4. Fitted radial profile PSF function (solid line) for the observed values (asterisks)

$$P(\mathbf{g}|\mathbf{f}) = \prod_{i=1}^p e^{-(\mathbf{H}\mathbf{f})_i} \frac{(\mathbf{H}\mathbf{f})_i^{\mathcal{J}_i}}{\mathcal{J}_i!}, \quad (18)$$

which is widely used in the astronomical community.

This model can be approximated, at least for high brightness values, by the Gaussian distribution $\mathcal{N}((\mathbf{H}\mathbf{f})_i, (\mathbf{H}\mathbf{f})_i)$. An alternative model is to assume $\mathbf{g}_i = \lambda \mathbf{z}_i$ where $\mathbf{z}_i \sim \mathcal{P}(\mathbf{H}\mathbf{f})_i$. This is the case in many images provided by the astronomers. For this model we can use the Gaussian approximation $\mathbf{g}_i|\mathbf{f} \sim \mathcal{N}((\mathbf{H}(\lambda\mathbf{f}))_i, \lambda(\mathbf{H}(\lambda\mathbf{f}))_i)$, and substituting $\lambda\mathbf{f}$ by \mathbf{f} , we have

$$\mathbf{g}_i|\mathbf{f} \sim \mathcal{N}((\mathbf{H}\mathbf{f})_i, \lambda(\mathbf{H}\mathbf{f})_i). \quad (19)$$

Finally, the noise model

$$\mathbf{g}_i|\mathbf{f} \sim \mathcal{P}((\mathbf{H}\mathbf{f})_i) + \mathcal{N}(0, \sigma^2)$$

takes into account both the detection process, Poisson model, and the reading process by an electronic device which adds a Gaussian read-out noise with mean zero and variance σ^2 (see [47]). The conditional model, in other words the data likelihood, has then the form

$$P(\mathbf{g}|\mathbf{f}) = \prod_{i=1}^p \sum_{k=0}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{[-(\mathbf{g}_{i-k})^2/2\sigma^2]} e^{-(\mathbf{H}\mathbf{f})_i} \frac{(\mathbf{H}\mathbf{f})_i^k}{k!}. \quad (20)$$

Note that for each component, i , this distribution can be approximated by a $\mathcal{N}((\mathbf{H}\mathbf{f})_i, a + b(\mathbf{H}\mathbf{f})_i)$ for appropriate constants a and b . In this case,

$$-2 \log p(\mathbf{g}|\mathbf{f}) = \text{const} + \sum_i E_i^2, \quad (21)$$

with $E_i = [(\mathbf{g}_i - (\mathbf{H}\mathbf{f})_i)/\sigma((\mathbf{H}\mathbf{f})_i)]$, $\sigma((\mathbf{H}\mathbf{f})_i)$ being the standard deviation, $\sqrt{a + b(\mathbf{H}\mathbf{f})_i}$ in the noise model just described (see [43] and [41] for methods on how to estimate a and b).

This Gaussian approximation allows the easy incorporation of robust statistics concepts to deal with detector errors like bad lines or hot pixels [41], but that should be used very carefully in regions with point sources which can be mistakenly taken as hot pixels. The idea is to downweight observations which are far away from their means. Such values are given too much weight in (21). The squared term in E_i represents the number of standard deviations that \mathbf{g}_i is away from its mean. In robust statistics $\sum_i E_i^2$ is replaced by $p(E_i)$ for a function p which penalizes extreme values less severely. A typical function p is Huber's "proposal 2" function defined by

$$p(x) = \begin{cases} x^2, & \text{for } |x| \leq c \\ 2c|x| - c^2, & \text{for } |x| > c. \end{cases} \quad (22)$$

This is quadratic in the center, but penalizes large deviations linearly rather than quadratically. Equivalently, observations E_i are downweighted if $|E_i|$ exceeds c . In practice c is chosen at about two, which downweights

only those observations more than two standard deviations away from their means. Note that Huber's distribution has been used as image model in image restoration (see [66]) but it has not been used much to model the noise process.

For Poisson noise, White [72] introduces a modification of the noise model to make it flatter in the vicinity of a good data fit. Instead of using $\log P(\mathbf{g}|\mathbf{f})$ from (19) he uses

$$\log P'(\mathbf{g}|\mathbf{f}) = \sum_i z(s_i(\mathbf{g}|\mathbf{f})), \quad (23)$$

where

$$s_i(\mathbf{g}|\mathbf{f}) = -\frac{2}{T^2} \left[\mathcal{J}_i \log \frac{(\mathbf{H}\mathbf{f})_i}{\mathcal{J}_i} - (\mathbf{H}\mathbf{f})_i + \mathcal{J}_i \right] \quad (24)$$

with

$$z(x) = \begin{cases} \frac{N-1}{N+1} (1-x^{N+1}) + x^N, & x < 1 \\ x, & x \geq 1 \end{cases} \quad (25)$$

with $z(x)$ being the damping function. This function is chosen to be a simple function, that is x for $x > 1$, is approximately constant for $x \approx 0$ and has continuous first and second derivatives at $x=1$. The constant N determines how suddenly the function z becomes flat for $x < 1$, for $N=0$, $z(x)=x$ and there is no flattening in the vicinity of a good data fit. The larger the value of N , the flatter the function is. The constants and multiplicative factors in $s_i(\mathbf{g}|\mathbf{f}), i=1, \dots, p$, are chosen to determine at what level the damping turns on (see [72]).

Finally, to complete the degradation models used in astronomy, we describe the use of multiresolution techniques to model the conditional distribution. A wavelet transform for discrete data that is used in the astronomical community is the à trous (with holes) algorithm (see [18] and [59]). The wavelet transform of an image by the à trous algorithm produces a set of coefficients $\{w^l\}$ at each scale l which has the same number of pixels as the image. The original image \mathbf{c} can be expressed as the sum of all the wavelet planes and the smoothed array \mathbf{c}^k , so,

$$\mathbf{c} = \mathbf{c}^k + \sum_{l=1}^k \mathbf{w}^l. \quad (26)$$

We can then consider the residuals defined in (16) and perform the à trous transform algorithm on them (see [2] and [64]) to obtain

$$\mathbf{r} = \mathbf{c}^k + \sum_{l=1}^k \mathbf{w}^l. \quad (27)$$

Any iterative scheme that uses the residuals as defined in (16) can instead use the modified residuals defined at each pixel position j by

$$\bar{r}_j = c_j^k + \sum_{l=1}^k T(w_j^l) w_j^l, \quad (28)$$

where T is a function which is defined by

$$T(w_j^l) = \begin{cases} 1, & \text{if } w_j^l \text{ is significant} \\ 0, & \text{if } w_j^l \text{ is not significant} \end{cases} \quad (29)$$

where methods to define whether w_j^l is significant or not can be found in [64].

Image Models

In this section we examine the image models used to restore astronomical images (for a short recent review of some multichannel models see [19]).

It is important to note that the introduction of prior models means making assumptions about the real underlying astronomical image and that testing the validity of those image models means, in fact, the introduction of distributions on the prior models. We will not discuss that problem here. Furthermore, before using a prior model to restore an astronomical image, it is important to know what each prior means. For example, as we will see now, some priors introduce smoothness constraints which make sense when restoring smoothly varying objects but that should not be applied when restoring point sources like stars. From our point of view there is not a so-called “best prior model” and to restore an observed image, several methods should be tried, examining then their restorations, and residuals, on the regions of interest.

Maximum Entropy and Related Methods

Frieden [11] was the first author to use the Shannon form of entropy in image restoration. He assumed that in each intensity value $f_j, j=1, \dots, p$, there is an allowed uncertainty $\pm\alpha^{-1}$. Following [11] it was then sufficient to restore each n_j defined by $\alpha f_j = n_j$ (we note that α^{-1} can be understood as a variance and that we will see later how this uncertainty has been made signal dependent).

Let $N = \sum n_j$, and let us assume that the occurrence of one object unit α^{-1} does not affect the possible location of any other object units. Thus, the object units are statistically independent within each cell and from cell to cell. If $\mathbf{n} = (n_1, \dots, n_p)$ then, the number of ways, $S(\mathbf{n})$, that a typical object can occur is

$$S(\mathbf{n}) = \frac{N!}{n_1! n_2! \cdots n_p!}$$

and so

$$P(\mathbf{n}) = S(\mathbf{n}) \left[\frac{1}{p} \right]^N. \quad (30)$$

Then, using the approximation, $n! \approx n^n e^{-n}$ we find that

$$\log P(\mathbf{f}) = -\alpha \sum_{i=1}^p f_i \log f_i + \text{const},$$

where we have used the fact that N is a fixed number known approximately by conservation of energy principles that is, $N = \alpha E_0$ where $E_0 = \sum_i f_i$ with E_0 known. This produces

$$P(\mathbf{f}) \propto \exp \left[-\alpha \sum_{i=1}^p f_i \log f_i \right]. \quad (31)$$

We note that this prior probability can also be written as

$$P(\mathbf{f}) \propto \exp \left[-\alpha E_0 \sum_{i=1}^p p_i \log p_i \right], \quad (32)$$

with $p_j = f_j / E_0$ (see [15] for the classical “team of monkeys” justification of the maximum entropy principle and [62] for a general description of maximum entropy algorithms).

In 1983, Frieden [12] (see also [13]) refined his initial formulation of the maximum entropy approach and allowed the introduction of a probability distribution over object units going to pixels. He denoted such probability by $\mathbf{r} = (r_1, \dots, r_p)$. Then (30) becomes

$$P(\mathbf{n}) = \frac{N!}{n_1! n_2! \cdots n_p!} r_1^{n_1} \cdots r_p^{n_p}, \quad (33)$$

and the prior model defined in (31) becomes

$$P(\mathbf{f}) \propto \exp \left[-\alpha \sum_{i=1}^p f_i \log(f_i / r_i) \right]. \quad (34)$$

To work at the same scale for \mathbf{f} and \mathbf{r} , this equation can be rewritten as

$$P(\mathbf{f}) \propto \exp \left[-\alpha \sum_{i=1}^p f_i \log(f_i / Q_i) \right], \quad (35)$$

where $\mathbf{Q} = E_0 \mathbf{r}$. Note that \mathbf{Q} represents an *a priori* image (not distribution) for the real underlying image. Equation (34) can also be written as

$$P(\mathbf{f}) \propto \exp \left[-\alpha E_0 \sum_{i=1}^p p_i \log(p_i / r_i) \right]. \quad (36)$$

As Frieden [13] pointed out, the need to know the distribution \mathbf{r} is both the drawback and strength of the procedure (see [45] for a full account on the use in astronomy of the maximum entropy methods described so far).

In 1993, Núñez and Llacer [47] used Frieden’s formulation and noted that in (36) when $\alpha \rightarrow 0$ the maximum likelihood restoration provided by the degradation model used was obtained and that when $\alpha \rightarrow \infty$ the solution ob-

tained was the trivial one provided by the prior model, that is, $f_i = Q_i$. They applied this prior model in conjunction with the noise model defined in (21) and proposed the so-called fast maximum *a posteriori* with entropy prior (FMAPE) algorithm to find the restoration.

To obtain an space-variant α , Nuñez and Llacer [48], [49] proceeded as follows. They redefined n_i as $\alpha_i f_i$ and obtained, from similar derivations to the ones used above, that

$$P(\mathbf{f}) \propto \exp \left[-\sum_{i=1}^p \alpha_i f_i \log(f_i / Q_i) \right]. \quad (37)$$

The authors then applied this prior model in conjunction with the noise model defined in (21) to obtain the so-called FMAPEVAR algorithm to find the restoration (see [48]). They later renamed the algorithm as the multiple output channel FMAPE algorithm [49].

In 1988 Skilling [61] considered the traditional “team of monkeys” throwing balls at p cells at random with Poisson expectation μ_i , $i=1,\dots,p$, having in this case

$$P(\mathbf{n}|\boldsymbol{\mu}) = \prod_i \mu_i^{n_i} e^{-\mu_i} / n_i!, \quad (38)$$

where $\alpha f_i = n_i$ and $\alpha m_i = \mu_i$. Then using again the factorial approximation, we have

$$P(\mathbf{f}|\boldsymbol{\mu}) \propto \exp \left[-\alpha \sum_i (f_i - m_i - f_i \log(f_i / m_i)) \right]. \quad (39)$$

Note that this prior is similar, in some sense, to the ones studied above with the main difference being the introduction of the term $f_i - m_i$.

The image model just described was first applied to image restoration by Gull [14] in a companion paper to [61]. There, the author introduced some ideas about how to modify the underlying image model $\boldsymbol{\mu}$ and also commented on the possibility of generating $\mathbf{m} = (m_1, \dots, m_p)$ from a blurred image of hidden variables.

Gull [14] and Charter [7] incorporated the concept of preference for spatial correlation in maximum entropy restoration with the use of the intrinsic correlation function (ICF). In this model the restored image is defined as the convolution of some hidden image \mathbf{i} with the ICF, \mathbf{G} ,

$$\mathbf{f} = \mathbf{G}\mathbf{i}, \quad (40)$$

where \mathbf{G} is some blurring function, such as a Gaussian. Note that the idea is to introduce correlation between neighboring pixels. Then the prior model becomes

$$P(\mathbf{i}|\boldsymbol{\mu}) \propto \exp \left[-\alpha \sum_k (i_k - m_k - i_k \log(i_k / m_k)) \right], \quad (41)$$

and the blurring process is modified from $\mathbf{H}\mathbf{f}$ to $\mathbf{H}\mathbf{G}\mathbf{i}$.

The ICF was later extended by Weir [69] to a multi-channel decomposition, where \mathbf{f} was written as

$$\mathbf{f} = \sum_k w_k \mathbf{f}^k, \quad (42)$$

where k was the number of channels and w_k the weight of image channel k . Then each channel was considered to be obtained as the convolution of the corresponding hidden space channel having the form

$$\mathbf{f}^k = \mathbf{G}_k \mathbf{i}^k, \quad (43)$$

and the image prior reformulated as

$$P(\mathbf{i}|\boldsymbol{\mu}) \propto \exp \left[-\alpha \sum_k \sum_n (i_n^k - m_n^k - i_n^k \log(i_n^k / m_n^k)) \right]. \quad (44)$$

The method has also been extended to a pyramidal decomposition of the images (see [4] for details).

Starck et al. (see [64] and references therein and [63]) have proposed the use of the so-called multiscale entropy to restore and filter astronomical images. Considering the à trous decomposition of the original image in a similar manner as was done for the residuals in (26), we can write

$$\mathbf{f} = \mathbf{c}^k + \sum_{l=1}^k \mathbf{w}^l \quad (45)$$

and so define the prior model

$$\log P(\mathbf{f}) = \frac{1}{\sigma_{\text{scale}}^2} \sum_{\text{scale } l} \sum_j \sigma_l \left(w_j^l - m^l - |w_j^l| \log \frac{|w_j^l|}{m^l} \right) + \text{const} \quad (46)$$

where the wavelet coefficients are taken in absolute value since they can be positive and negative (see [65]), σ_{scale}^2 is an overall variance and the σ_l are scale-dependent deviations. Following [64], it is possible to estimate the scale deviations and also the model at each scale m^l .

This multiscale entropy can also be combined with a multiresolution support by multiplying each term in the sum in (46) by a term of the form A_j^l (initially zero or one but later convolved with a B-spline function) to introduce regularization on the corresponding wavelet coefficient [64].

To finish the entropic priors, let us now describe the pixon approach. In this approach, the original image is replaced by a pseudoimage smoothed locally by a function with position-dependent scale, that is,

$$f_i = \sum_x K \left[\frac{x-i}{\delta(i)} \right] f_x^{\text{pseudo}}, \quad (47)$$

where $\delta(i)$ is the locally variable scale and $K[(x-i)/(\delta(i))]$ is a smoothing function normalized to volume one, for instance,

$$K\left[\frac{x-i}{\delta(i)}\right] \propto \begin{cases} 1 - \frac{\|x-i\|^2}{\delta^2(i)}, & \|x-i\| \leq \delta(i) \\ 0, & \text{elsewhere} \end{cases} \quad (48)$$

where $K[(x-i)/(\delta(i))]$ is called the pixon shape function (see [51]) where pixons are basically variable cells. Note that the classical pixels are pixons for particular values of the scaling function but the concept of pixon is more general and, in some sense, can be considered as spreading, in some cases, over several pixels.

It is then possible to define an entropic prior on the units of signal on the pixons (see for instance [51], [55], [56], and [32]) and in combination with a degradation model proceed to find the MAP. Obviously, there are two important questions: how to calculate the scale parameters and the pseudo image, $\mathbf{f}^{\text{pseudo}}$. That will be discussed in the algorithm section.

Noninformative Priors and Extensions

One of the most widely used image restoration methods in astronomy is the so-called Richardson-Lucy (R-L) algorithm. This algorithm [57], [26] was initially derived from Bayes theorem and shown to increase the likelihood, as defined in [26], assigned to the observed image. The same algorithm was later obtained by Shepp and Vardi [60] and developed further by Vardi et al. [68] as an EM algorithm ([9], see also [30]) for medical image reconstruction with Poisson noise.

From the Bayesian point of view, the R-L algorithm can be obtained assuming that $P(\mathbf{f}) \propto \text{const}$, a so-called noninformative prior, and that the degradation model is the one given in (19). In this case we have

$$\log P(\mathbf{f}|\mathbf{g}) = \text{const} - \sum_{i=1}^p (\mathbf{H}\mathbf{f})_i + \sum_{i=1}^p g_i \log(\mathbf{H}\mathbf{f})_i, \quad (49)$$

and so

$$\hat{\mathbf{f}}_{RL} = \arg \max_{\mathbf{f}} \left[- \sum_{i=1}^p (\mathbf{H}\mathbf{f})_i + \sum_{i=1}^p g_i \log(\mathbf{H}\mathbf{f})_i \right]. \quad (50)$$

Note that the image model used is improper, it does not integrate to one, and that the term noninformative is rather misleading, since assuming that all the images have the same probability is quite informative.

The R-L algorithm was later adapted to incorporate designated sources having specified shape and known positions [27]. The designated sources are point sources, and the specified shape is the delta function. The idea is to use this *a priori* information, at least for the designated point sources, and so try to eliminate the ringing associated with recovered delta functions. The method described below is the so-called PLUCY algorithm. To incorporate this *a priori* knowledge the real underlying image is divided into two parts, that is,

$$\mathbf{f} = \mathbf{f}^s + \mathbf{f}^*,$$

where \mathbf{f}^s represents the smooth contribution and \mathbf{f}^* the designated point sources. This representation is achieved by defining \mathbf{f}^* to be zero everywhere except at the locations of the N known point sources. Then a smoothness constraint is introduced on \mathbf{f}^s by defining an entropic prior of the form (see for instance [67])

$$P(\mathbf{f}) = P(\mathbf{f}^s, \mathbf{f}^*) \propto \exp \left[-\alpha \sum_i \frac{f_i^s}{\sigma} \log \frac{f_i^s}{\psi_i} \right],$$

where $\sigma = \sum f_i^s$ and $\psi_i = \sum_m R_{im} f_i^s$. The resolution kernel R_{im} is a Gaussian function whose width is the required resolution limit.

There are two extensions of the PLUCY method. In the first one, called the CPLUCY algorithm, the point sources are not δ -functions but simply X, Y positions at arbitrary subpixel locations, see [19] and CPLUCY man pages at <ftp://ecf.hq.eso.org/rhook/cplucy>. GIRA is the second extension of PLUCY: this algorithm handles the deficiencies of PLUCY and CPLUCY when using the entropic constraints on the background (the smooth signal). In GIRA, the background is represented not as a set of independent pixels but as the sum of a set of Gaussians, one for each pixel. The width of these Gaussians is a free parameter which constrains the maximum fine structure in the background (see [52]). Note that in the astronomical literature some authors refer to PLUCY, CPLUCY and GIRA, together with the method to be described just below, as two-channel restoration algorithms. We prefer to use this terminology with restoration algorithms that combine different images.

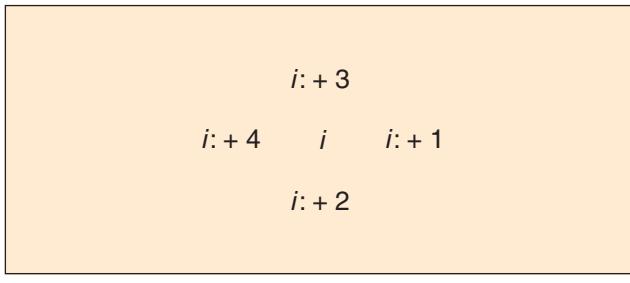
There is an alternative image model which is, in some sense, similar to the ones described in this subsection but which has been obtained from a different starting point. To solve the problem that appears when the observed data obey the sampling theorem but the deconvolved data do not, Magain et al. [29] propose to deconvolve not with the total PSF, \mathbf{H} , but with a narrower function, \mathbf{S} , chosen so that the deconvolved image has its own PSF, \mathbf{R} , compatible with the adopted sampling. The blurring matrices are related by

$$\mathbf{H} = \mathbf{SR}. \quad (51)$$

The shape and width of the PSF corresponding to \mathbf{R} are chosen by the user. The only constraint is that (51) admits a solution \mathbf{S} . The process is very intuitive: think about the decomposition of a normal distribution with variance σ^2 as the sum of two Gaussians with variances σ_1^2 and σ_2^2 such that $\sigma^2 = \sigma_1^2 + \sigma_2^2$, one corresponding to the scale of the original image σ_1^2 and represented by \mathbf{S} and the other to the scale of the deconvolution process, corresponding to σ_2^2 and represented by \mathbf{R} .

The original image is then decomposed using

$$\mathbf{f} = \mathbf{s} + \sum_{k=1}^M \alpha_k \mathbf{Rc}_k, \quad (52)$$



▲ 5. Notation used for the four pixels around a given one i .

where M is the number of point sources, for which α_k and the vectors \mathbf{c}_k are free parameters corresponding to their intensities and positions and \mathbf{s} is the extended component of the solution, that is, generally a rather smooth background. Note that the \mathbf{c}_k are basically column vectors which are zero everywhere except in one component where it takes the value one (a subpixel approximation is also possible).

We could now write $\mathbf{s} = \mathbf{R}\mathbf{s}'$ and use this knowledge to try to find \mathbf{s}' , but \mathbf{s}' may violate the sampling theorem, so the authors prefer to use this equation to impose smoothness on \mathbf{s} on the scale length of \mathbf{R} , and so in [29] the following was used

$$P(\mathbf{g}|\mathbf{f}) \propto \exp\left[-\frac{1}{2}\left\|\mathbf{g} - \mathbf{S}\left(\mathbf{s} + \sum_{k=1}^M \alpha_k \mathbf{R}\mathbf{c}_k\right)\right\|_W^2\right], \quad (53)$$

where W is a weighting diagonal matrix with diagonal entries $1/\sigma_i^2$. The image model is defined by

$$P(\mathbf{f}) \propto \exp\left[-\frac{\lambda}{2}\|\mathbf{s} - \mathbf{R}\mathbf{s}\|^2\right]. \quad (54)$$

Regularization and Spatial Statistics Models

We now move on to examine the image models that have appeared from the point of view of regularization or from the explicit use of spatial statistical models to restore astronomical images.

To do so, we denote by $i+1, i+2, i+3, i+4$ the four pixels around pixel i (see Fig. 5) and consider the following quadratic form

$$\|\mathbf{C}\mathbf{f}\|^2 = \sum_i \left(f_i - \frac{1}{4}(f_{i+1} + f_{i+2} + f_{i+3} + f_{i+4}) \right)^2. \quad (55)$$

We can then restore an image by minimizing $\alpha\|\mathbf{C}\mathbf{f}\|^2 + \beta\|\mathbf{g} - \mathbf{H}\mathbf{f}\|^2$ (see [21]). This is equivalent to using $P(\mathbf{g}|\mathbf{f})$ as defined in (12) and

$$P(\mathbf{f}) \propto \exp\left[-\frac{\alpha}{2}\|\mathbf{C}\mathbf{f}\|^2\right]. \quad (56)$$

From the point of view of regularization, $\|\mathbf{C}\mathbf{f}\|^2$ is used as a way to set constraints on the Laplacian of the restora-

tion or to limit the high frequencies in the restoration. Note that this model is the so-called simultaneous autoregressive model (SAR) within the statistical community (see [58]).

From the point of view of regularization a model that has been used to restore astronomical images is the one given by

$$P(\mathbf{f}) \propto \exp\left[-\alpha_w(\mathbf{f})\|\mathbf{C}\mathbf{f}\|_{B(\mathbf{f})}^2\right], \quad (57)$$

where $B(\mathbf{f})$ has been included to make the restoration algorithm spatially adaptive and where the regularization parameter $\alpha(\mathbf{f})$ is a function of the original image (in practice it becomes a function of an estimate of the original image).

This prior model in combination with the noise model defined in (13) has been used by Katsaggelos et al. ([22], see also [23] and [1]) to find the restoration by minimizing the functional

$$M(\alpha(\mathbf{f}), \mathbf{f}) = \left\|\mathbf{g} - \mathbf{H}\mathbf{f}\right\|_{A(\mathbf{f})}^2 + \alpha_w(\mathbf{f})\|\mathbf{C}\mathbf{f}\|_{B(\mathbf{f})}^2, \quad (58)$$

where to satisfy some optimality criteria for the functional defined in (58), it is shown in [20] that, among others, a possible choice for the regularization parameter is

$$\alpha_w(\mathbf{f}) = \frac{\left\|\mathbf{g} - \mathbf{H}\mathbf{f}\right\|_{A(\mathbf{f})}^2}{(1/\gamma_1) - \|\mathbf{C}\mathbf{f}\|_{B(\mathbf{f})}^2}, \quad (59)$$

where $\gamma_1 = 1/(2\|\mathbf{g}\|^2)$.

We note that from the point of view of regularization the normalizing constants of the prior and degradation models are not taken into account when finding the MAP. This can be justified by the weak dependence of the restoration on them (only the logarithm of those normalizing constants should be present in the objective function to be minimized to find the MAP).

From the point of view of regularization we can also use the prior model defined by

$$P(\mathbf{f}) \propto \exp\left[-\alpha \sum_i [(f_i - f_{i+1})^2 + (f_i - f_{i+2})^2]\right], \quad (60)$$

which means setting constraints on first derivatives. We will reencounter this model again from the point of view of spatial statistics.

This prior model has been extended to

$$P(\mathbf{f}) \propto \exp\left[-\alpha \sum_i [\varphi(f_i - f_{i+1})^2 + \varphi(f_i - f_{i+2})^2]\right], \quad (61)$$

where $\varphi(x)$ is an edge preserving function satisfying some constraints (see [3] and [6]). The MAP is then found by an algorithm, ARTUR, proposed by the authors [3], [6].

From the point of view of spatial statistics, the simplest prior model to restore astronomical images consists of working at linear scale. Consider an image with no stars but regions of smoothly varying luminosity. We then expect $\mathbf{f}_i \geq 0$ and \mathbf{f} to be spatially smooth. In this case smoothness is modeled by spatial autoregressions [58].

The conditional autoregression, CAR, model is defined by

$$p(\mathbf{f}) \propto \exp \left\{ -\frac{\alpha}{2} \mathbf{f}^t (\mathbf{I} - \phi \mathbf{N}) \mathbf{f} \right\},$$

where α^{-1} is the unknown prior variance, matrix \mathbf{N} is such that $\mathbf{N}_{ij} = 1$ if cells i and j are spatial neighbors (pixels at distance one), zero otherwise and scalar ϕ is just less than 0.25. The term $\mathbf{f}^t (\mathbf{I} - \phi \mathbf{N}) \mathbf{f}$ represents in matrix notation the sum of squares of the values \mathbf{f}_i minus ϕ times the sum of $\mathbf{f}_i \mathbf{f}_j$ for neighboring pixels i and j .

We note that

$$\begin{aligned} \mathbf{f}^t (\mathbf{I} - \phi \mathbf{N}) \mathbf{f} = \\ \sum_i \left[\phi(f_i - f_{i+1})^2 + \phi(f_i - f_{i+2})^2 + (1 - 4\phi)f_i^2 \right] \end{aligned} \quad (62)$$

and so a CAR model in the limit case, $\phi=0.25$, is equivalent to setting constraints on first derivatives. We note that when $\phi=0.25$, $(\mathbf{I} - \phi \mathbf{N})$ is equal to \mathbf{C} as defined in (59).

The parameters in a CAR model can be interpreted by the following expressions describing the conditional distribution

$$E(f_i | f_j, j \neq i) = \phi \sum_{j \text{ nbr } i} \mathbf{f}_j \quad \text{and} \quad \text{var}(f_i | f_j, j \neq i) = \alpha^{-1}.$$

It is also possible to modify the smoothness scale when working with a CAR image model. The luminosity distribution of pure disk in a galaxy can be modeled by “the exponential law” written as

$$I(r) = I(0) \exp(-b_0 r). \quad (63)$$

Furthermore, the luminosity distribution of elliptical galaxies has been investigated by many astronomers, and several analytical functions have been proposed to model this distribution. Among others, the most commonly used is the “ $r^{1/4}$ law” proposed by de Vaucouleurs

$$\log(I(r) / I_E) = -3.33 \left[(r / r_E)^{1/4} - 1 \right], \quad (64)$$

where r is the distance from the center of the galaxy and I_E and r_E are parameters which differ from galaxy to galaxy.

These results suggest that the luminosity of galaxies is most naturally considered on log scale (see [43]). The smoothness of the luminosity is then modeled on log scale except for very small values of luminosity (see [43] and [41]). This means using a CAR prior for $\mathbf{y} = \ln(\mathbf{f} + \gamma \mathbf{1})$, where γ is a small constant, that is,

$$-2 \ln P(\mathbf{y}) = \text{const} + \alpha_y \mathbf{y}^t (\mathbf{I} - \phi \mathbf{N}) \mathbf{y}. \quad (65)$$

It is very important to note that we are proposing an image model for \mathbf{y} and then obtaining an estimate of the source image as $\hat{\mathbf{f}} = \exp[\hat{\mathbf{y}}] - \gamma \mathbf{1}$, where $\hat{\mathbf{y}}$ is the MAP restoration for the log-scale formulation.

The CAR model either on linear or log scale can be extended to restore astronomical images when the location of objects is known in advance [44].

Compound Gauss Markov Random Fields image models (CGMRF, see [21]) have also been applied to the restoration of astronomical images. The image model has the form

$$\begin{aligned} -\log P(\mathbf{f}, \mathbf{l}) = \text{const} + \frac{\alpha}{2} \sum_i \left[\phi(f_i - f_{i+1})^2 (1 - l_{[i,i+1]}) \right. \\ \left. + \beta l_{[i,i+1]} + \phi(f_i - f_{i+2})^2 \right. \\ \times (1 - l_{[i,i+2]}) + \beta l_{[i,i+2]} \\ \left. + (1 - 4\phi)f_i^2 \right], \end{aligned} \quad (66)$$

where the function $l_{[i,j]}$ takes the value zero if pixels i and j are not separated by an active line and one otherwise and ϕ is just less than 1/4. The introduction of the line element $[i,j]$ is then penalized by the term $\beta l_{[i,j]}$ since otherwise the expression in (67) would obtain its minimum value by setting all line elements equal to one. The intuitive interpretation of this line process is simple; it acts as an activator or inhibitor of the relation between two neighbor pixels depending on whether or not the pixels are separated by an edge. In the process of restoring the image, this method requires the estimation of the line process too. Since the energy of this model for \mathbf{f} and \mathbf{l} is no longer convex finding the MAP becomes more complicated (see [36] and [38]).

To finish this section on image priors we note that the following prior has been used to restore images of planets

$$P(\mathbf{f}) \propto \exp \left\{ \beta \sum_{<i,j>} \frac{1}{1 + [(f_i - f_j) / \delta]^2} \right\}, \quad (67)$$

where $<i,j>$ denotes the set of pixel pairs (i,j) at unit distance apart (see [35]).

Algorithms

As discussed previously, the MAP restoration will be obtained by

$$\text{minimizing} -\log P(\mathbf{g}|\mathbf{f}) - \log P(\mathbf{f}), \quad (68)$$

sometimes with the constraint of energy conservation that can be expressed as

$$\sum_{i=1}^p g_i = \sum_{i=1}^p (\mathbf{H}\mathbf{f})_i = \sum_{j=1}^p q_j f_j \quad (69)$$

where

$$q_j = \sum_{i=1}^p b_{ij} \quad (70)$$

which usually satisfies $q_j = 1, j = 1, \dots, p$.

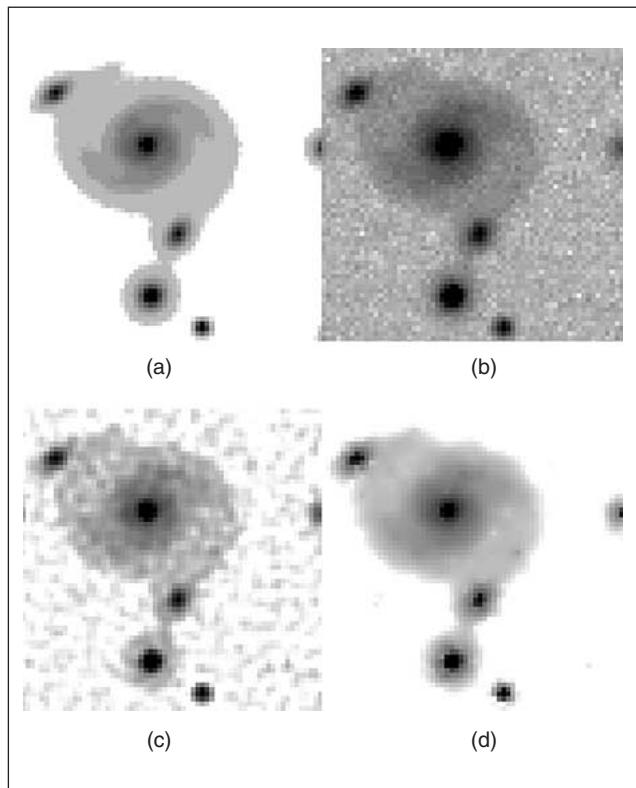
The exact form of the Bayesian function to be minimized will depend on the specific models used for both the prior distribution and the likelihood, as described in the previous sections. Here we will comment on the general aspects of the minimization methods and present some of the most used algorithms.

Classical Gradient Techniques

To obtain the MAP the simplest way is to solve

$$\frac{\partial}{\partial f} \log P(f|g) = 0, \quad (71)$$

which aims at finding the stationary points of the posterior distribution. This so-called direct minimization has the problem that in most cases, mainly when using entropic forms as prior, it leads to compute exponentials that produce large instabilities when used in iterative formulae. Finding the solution of (71) it is possible in some cases, for instance when the blurring is space invariant, the noise model is the classical Gaussian independent one, see (12), and the prior model is the SAR model described in (56). In this case, the MAP restoration can be found by solving



▲ 6. Simulated object (a), blurred image (b), deconvolved image by MEM (c), and deconvolved image by MEM multiresolution (d).

$$(\lambda C^t C + \mathbf{H}^t \mathbf{H}) \mathbf{f} = \mathbf{H}^t \mathbf{g}, \quad (72)$$

with $\lambda = \alpha / \beta$. The solution can be found using Fourier transform. This method, however, does not necessarily satisfy the nonnegativity constraint of the restoration.

The simplest method to incorporate gradient based techniques into the problem of finding the solution of (69) is to use the steepest descent algorithm [53]. The method seeks to minimize the Bayesian function along the direction of the steepest descent, defined by a vector proportional to the gradient.

The main advantage of the steepest descent method over direct methods is that usually it is not needed to compute exponentials avoiding the instabilities of the direct minimization. A serious disadvantage, however, is that the method does not guarantee positivity. In almost every iteration in which the gradient step is sufficiently large to enable significant development of the intensities in reasonable time, there are pixels with negative values. Since intensity negative values have no physical meaning and at the next iteration it would not be possible to compute the logarithms appearing in the gradient (especially using entropic priors), after each iteration any negative intensity should be reset to a small positive value. By doing this, we are deviating the iteration point from the search direction given by the gradient vector. The consequence is that neither the decrease of the Bayesian function nor the convergence are guaranteed. If many pixels become negative, the effect is to stop the algorithm progressing towards a minimum.

Gradient descent techniques have been applied to, among others, the minimization of the functional $M(\alpha(f), f)$ defined in (58) (see [22] and [23] for details), to the CAR image model in log scale [see (65)] together with multiplicative noise model [see (21)] described in [43] and [41] and to the multiscale entropy prior described in [64]. Figure 6 shows the restoration of a simulated image by the method described in [64].

NGC 450/UGC 807 is a case of two galaxies having a small angular separation but with very discrepant redshifts. The original and deconvolved frames using the log CAR prior are presented in Fig. 7. As can be seen from the figure, many small scale features are strongly enhanced mainly at the disk of NGC 450 and the arms of its companion are now clearly delimited.

The conjugate gradient method has also been applied to image restoration in astronomy. It starts like the steepest descent (indeed, the first iteration is exactly the same) but in place of minimizing along the direction of the gradient, the conjugate gradient minimizes in the vectorial subspace generated by the steepest descent vector corresponding to the iteration and the steepest descent vectors of all (or several) the previous iterations (see [53] for details). The conjugate gradient represents a clear improvement over the steepest descent algorithm at no extra computational cost. Nevertheless, the problem of nega-

tive values for the intensities is not solved and in most cases the solution is plagued by them, although several authors have presented variants of the conjugate gradient algorithm (see, for instance, [62]) to prevent the appearance of negative values.

Expectation Maximization and the R-L-EM Algorithm

The expectation-maximization (EM) algorithm was introduced by Dempster et al. [9] and first applied to image reconstruction by Shepp and Vardi [60] in the field of medical imaging (positron emission tomography). The algorithm was identical to the algorithms obtained independently by Richardson [57] and Lucy [26], the R-L algorithm.

The R-L algorithm can be seen as a particular case in the Bayesian framework by considering the case of pure Poisson noise distribution [see (18)] and taking as constant the prior information [see (49)].

The solution of (68), in this case by the EM algorithm, gives the following iterative scheme

$$f_i^{(k+1)} = f_i^{(k)} \left[\frac{1}{q_i} \sum_{j=1}^p \frac{H_{ji} \mathcal{G}_j}{\sum_{l=1}^p H_{jl} f_l^{(k)}} \right] \quad i=1, \dots, p, \quad (73)$$

where k denotes iteration. This equation can be rewritten as

$$f_i^{(k+1)} = f_i^{(k)} \left[\frac{1}{q_i} \left(\mathbf{H}^T \left(\frac{\mathbf{g}}{\mathbf{H}\mathbf{f}^k} \right) \right)_i \right], \quad i=1, \dots, p. \quad (74)$$

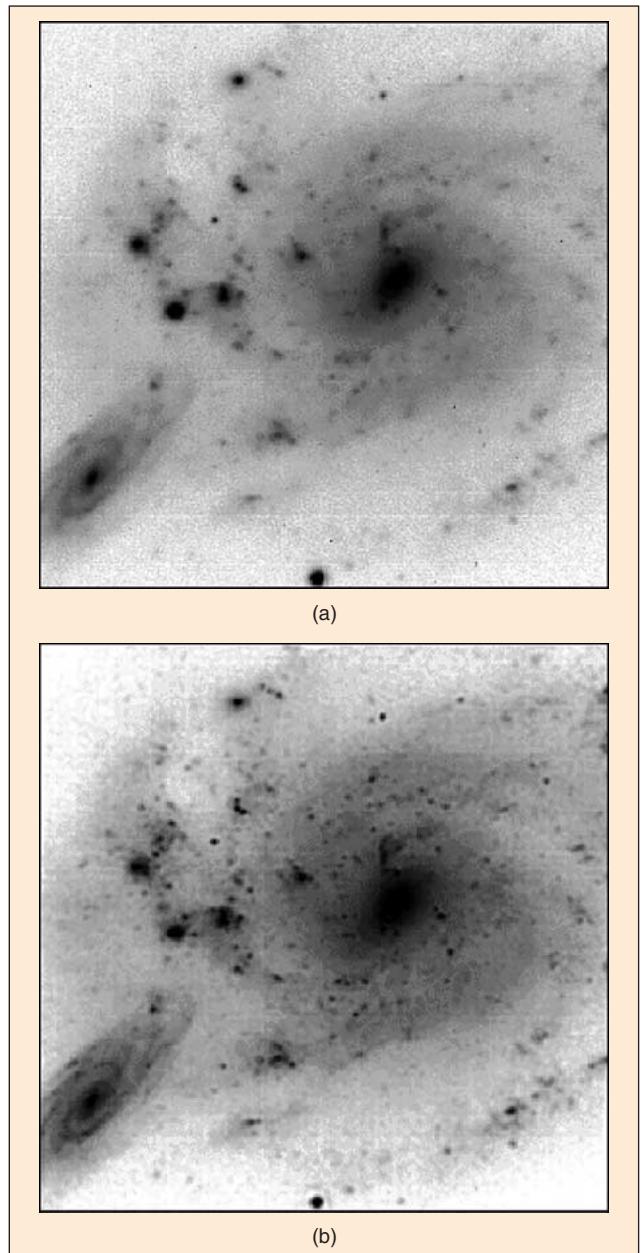
Equation (73) is the R-L-EM algorithm which is extensively used for maximum likelihood in image reconstruction and restoration. Since the observational data (\mathcal{G}_j) are positive in the Poisson case, the form of the algorithm guarantees that the solution is always positive (or zero) at each pixel. Also the energy is preserved at global and local scale. These are two of the most important properties of this algorithm.

The R-L algorithm has the general form of an iterative formula in which the value of a pixel at each iteration is equal to the previous one multiplied by a modifying factor. The modifying factor between brackets in (73) consists of a projection and a backprojection. In the case of space-invariant PSF, the projection and backprojection operations can be performed using the fast Fourier transform (FFT). From the point of view of the computation requirements, in the case of space-invariant PSF and assuming that the FFT of the PSF is calculated at the beginning of the process and stored in memory, each iteration needs two FFTs in the projection step and two more in the backprojection step. All the other computations are vector operations that have little impact on CPU time. Thus the R-L algorithm is basically an algorithm with four FFTs per iteration.

Figure 8 shows the raw data of an image of Saturn obtained with the WF/PC camera of the HST, before the refurbishing mission in 1994. Figure 9 shows its reconstruction using the R-L algorithm with 50 iterations. The restoration shows a well-reconstructed image of the planet and sharp separations in the rings.

The R-L algorithm has been extended to the methods PLUCY, CLUCY, and GIRA described earlier, to the noise model described in (23) and also to incorporate smoothness constraints in the form of a CAR prior model (see [39] for details).

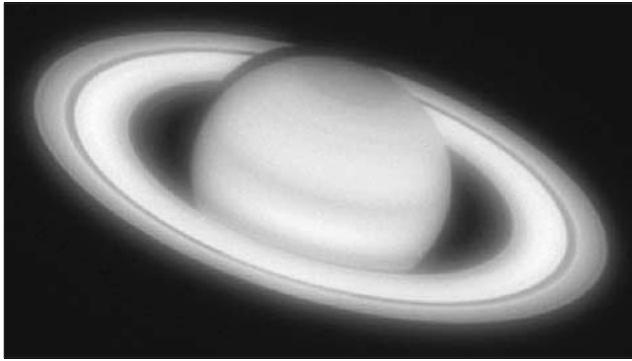
Note that in (74), \mathbf{g} can be substituted by $\mathbf{H}\mathbf{f}^k + \mathbf{r}^k$ where \mathbf{r}^k are the residuals at iteration k . A wavelet decomposition can be applied to those residuals and the wavelet



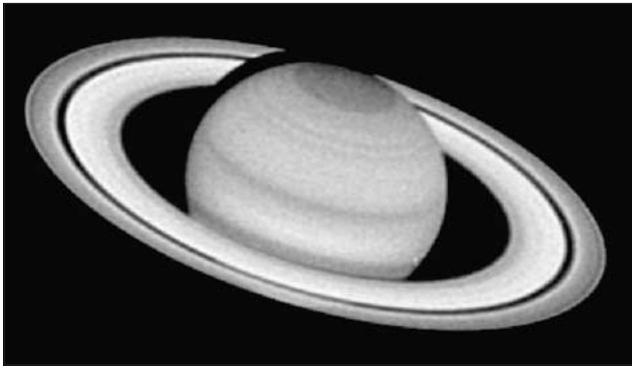
▲ 7. (a) Original and (b) deconvolved images of NGC 450/UGC 807 by the method proposed in [43].

coefficients can be modified as described in (28). Figure 10 shows an example of the use of such an algorithm.

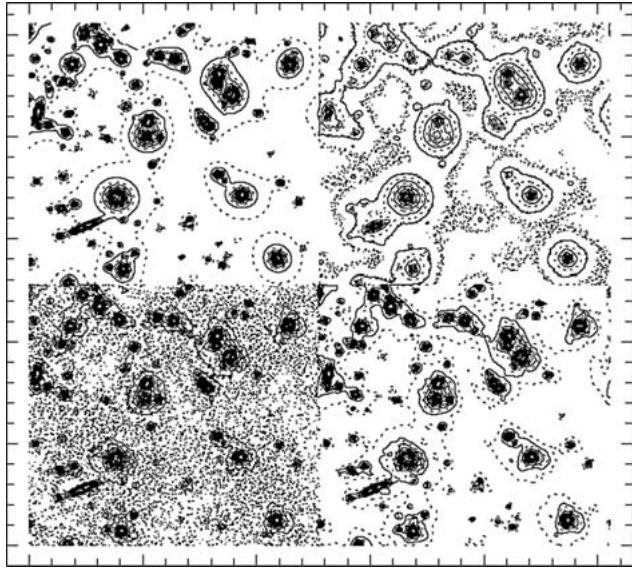
The algorithm given by (73) was extended in [47] to the case Poisson + Gaussian readout noise [see (20)].



▲ 8. Raw image of planet Saturn obtained with the WF/PC camera of the HST.



▲ 9. Reconstruction of the image of Saturn using the R-L algorithm.



▲ 10. Simulated HST wide field camera image of a distant cluster of galaxies. Four quadrants. Upper left: original, unaberrated, and noise-free. Upper right: input, aberrated, noise added. Lower left: restoration, Richardson-Lucy method with noise suppression, 28 iterations. Lower right: restoration, Richardson-Lucy method without noise suppression, 40 iterations. Intensities logarithmically transformed.

The obtained algorithm has the same form as the R-L algorithm but changing the data \mathcal{J}_j to

$$\mathcal{J}'_j = \frac{\sum_{k=0}^{\infty} \left(k e^{-\frac{(k-\mathcal{J}_j)^2}{2\sigma^2}} \frac{((\mathbf{Hf})_j^k)}{k!} \right)}{\sum_{k=0}^{\infty} \left(e^{-\frac{(k-\mathcal{J}_j)^2}{2\sigma^2}} \frac{((\mathbf{Hf})_j^k)}{k!} \right)}. \quad (75)$$

In the absence of readout noise or when it is negligible [$\sigma \rightarrow 0$ in (75)], the exponentials are dominant at $k = \mathcal{J}_j$ and $\mathcal{J}'_j \rightarrow \mathcal{J}_j$. We note that by its definition, \mathcal{J}'_j , $j = 1, \dots, p$, is always positive, while the original data \mathcal{J}_j , $j = 1, \dots, p$, can be negative. For example, if the background of the image is small, 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 R-L algorithm. A preprocessing of the data or other approximations are common. Using (75), however, approximation or preprocessing is not needed. In fact, \mathcal{J}'_j , $j = 1, \dots, p$, is a representation of the data that is positive and close to the data, with a degree of closeness that depends on the projections $(\mathbf{Hf})_j$, $j = 1, \dots, p$, and the standard deviation of the Gaussian readout noise σ . In this sense \mathcal{J}'_j , $j = 1, \dots, p$, can be considered as a Bayesian filtered version of the data.

Successive Substitutions

Although the EM algorithm is well behaved, it is difficult to use in the Bayesian case (mainly with entropy priors) because it requires the solution of transcendental equations in the M-step. Instead, it is possible to use the method of successive substitutions ([17], [31]), which has been used in image restoration in astronomy (see, for instance, [47] and [49]). This method allows greater flexibility than other methods and results, sometimes, in rapidly converging algorithms.

The successive substitutions method can be described as follows: given a series of equations with the unknowns f_i , $i = 1, \dots, p$, in the form

$$f_i = K F_i(\mathbf{f}), \quad i = 1, \dots, p, \quad (76)$$

where F_i is some function and K is a normalization constant, (76) can be transformed into a recursive relation by

$$f_i^{(k+1)} = K F_i(\mathbf{f}^{(k)}), \quad i = 1, \dots, p. \quad (77)$$

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

$$\sum_{i=1}^p q_i f_i^{(k+1)} = \sum_{i=1}^p K q_i F_i(\mathbf{f}^{(k)}) = \sum_{j=1}^p g_j. \quad (78)$$

This optimization technique has been applied to the image model defined in (35) and the noise model defined in (20). The method, as we already know, is the so-called FMAPE algorithm ([47]). This optimization method has also been applied to the image model defined in (37) producing the FMAPEVAR algorithm or the multiple output channel FMAPE algorithm (see [49]).

As we have already seen in the introduction, Fig. 1 shows in logarithmic scale the “dirty map” of a radio-interferometric observation obtained with the VLA of the star LSI + 61 303. Note that the image of the beam dominates practically all the scene. Figure 2 shows, also in logarithmic scale, the restoration of Fig. 1 by the FMAPE method described in the text. The image of the star itself has been removed to better show the tiny extended sources surrounding the star.

Figure 11 shows the result of the reconstruction of Saturn in Fig. 8 using a nine-channel multichannel FMAPE algorithm, that is, nine different hyperparameters α_i in (37). The restoration shows a well-reconstructed image of the planet and sharp divisions in the rings. By the use of the variable resolution FMAPE algorithm, noise amplification was avoided in all regions of the image.

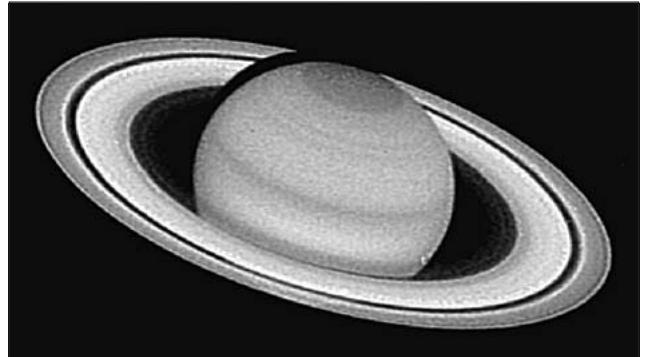
Other Methods

Deterministic techniques to find the MAP for edge preserving priors [see (61)] have been proposed in [3] and [6], producing the ARTUR algorithm. Stochastic techniques like simulated annealing (SA) have also been used to restore astronomical images when the noise model is the one given in (12) and the prior model is the CGMRF given by (66).

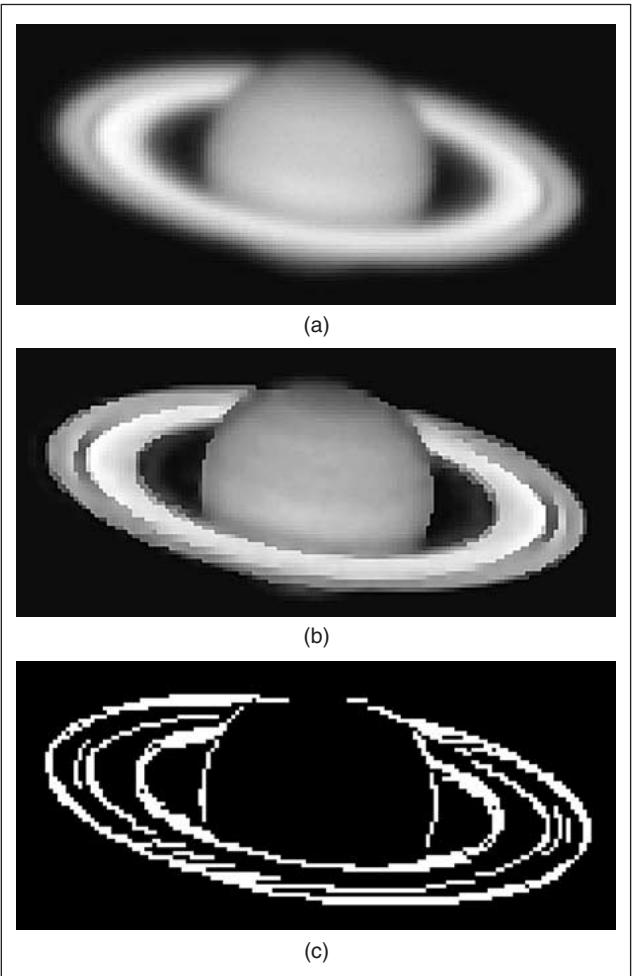
Figure 12(a) shows a 96×250 image of Saturn taken at the Cassegrain f/8 focus of the 1.52m telescope at Calar Alto Observatory (Spain) in July 1991. Results are presented in a image taken through a narrow-band interference filter centered at the wavelength 9500 \AA . Figure 12(b) shows the restoration obtained when using the modification of SA proposed in [38]. The obtained line process for this restoration is shown in Fig. 12(c). In all the images the improvement in spatial resolution is evident. In particular, ring light contribution has been successfully removed from equatorial regions close to the actual location of the rings and amongst the rings of Saturn, the Cassini division is enhanced in contrast, and the Encke division appears on the ansae of the rings in all deconvolved images.

We describe, briefly, here how the pixon based approach [see (47)] is used to restore astronomical images. The pixon prior is normally used in combination with a Gaussian noise model [see (12)] or the noise model described in (15). The method starts by finding a restoration (maximum entropy, for instance) to be used as a starting point for $\mathbf{f}^{\text{pseudo}}$ in (47). Then, the scales of the

pixons, $\delta(i)$, initially set equal to one in (47), are modified to increase the posterior probability. This is done taking into account that by increasing the length of the scales we increase the pixon prior probability but that will usually decrease the log likelihood of the data. Once the new lengths of the scales maximizing the posterior have been found, a new pseudo image in (47), and so a new restoration, is calculated trying always to maximize the posterior probability (see [54]).



▲ 11. Reconstruction of the image of Saturn using the FMAPE algorithm with nine channels.



▲ 12. (a) Observed 96×250 image of Saturn from Calar Alto (Spain), (b) restoration by the modified SA algorithm proposed in [38], and (c) its line process.

Software and Online Information in Astronomy

Professional Free Software

Comprehensive professional implementations of both R-L and ME methods for image restoration can be obtained from IRAF. IRAF (Image Reduction and Analysis Facility) is a general purpose software system for the reduction and analysis of astronomical data. IRAF is written and supported by the IRAF programming group at the National Optical Astronomy Observatories (NOAO) in Tucson, AZ. NOAO is operated by the Association of Universities for Research in Astronomy (AURA), Inc. under cooperative agreement with the National Science Foundation. IRAF is freely available on all common platforms via anonymous ftp from the server at <http://iraf.noao.edu> (see also the mirror at <http://ecf.hq.eso.org/iraf/web>). The software for image restoration can be found in the STSDAS package `stsdas/analysis/restore`. There, the reader will find implementations for the R-L method (Lucy), an implementation of the ME method (MEM), the primitive Wiener filtering Fourier linear method (Wiener), and the CLEAN method of PSF fitting and removal (clean).

Extensions of the R-L algorithm PLUCY, CPLUCY and GIRA can also be obtained. PLUCY is available under IRAF `stsdas/contrib` package. Using the code available at [ftp://ecf.hq.eso.org/rhook/cplucy](http://ecf.hq.eso.org/rhook/cplucy) CPLUCY can be incorporated into IRAF. The IRAF implementation of GIRA is under construction.

Several restoration algorithms including the R-L one in a multiresolution context can be obtained from the ESO-MIDAS package. ESO-MIDAS is the acronym for the European Southern Observatory Munich Image Data Analysis System which is developed by the European Southern Observatory (see <http://www.eso.org/projects/esomidas>) The ESO-MIDAS system provides general tools for image processing and data reduction with emphasis on astronomical applications including imaging and special reduction packages for ESO instrumentation at La Silla and the Very Large Telescope at Paranal. In addition it contains applications packages for stellar and surface photometry, image sharpening and decomposition, statistics and other packages. ESO-MIDAS is available under the GNU General Public License (GPL). In particular, the Richardson-Lucy implementation is under the LUCY/WAVE package.

In radio astronomy the most widely used image processing system are the AIPS/AIPS++ packages. The AIPS, Astronomical Image Processing System, is a software package for calibration, data analysis, image display, plotting, and a variety of ancillary tasks on astronomical data. It comes from the National Radio Astronomy Observatory. Although it is primarily for radio astronomy, AIPS has also routines for general image processing. AIPS++ (C++ based) is a software to process (again primarily radio) astronomical data. It is developed by an international consortium of observatories. AIPS/AIPS++ packages are available at <http://www.cv.nrao.edu/aips> and <http://aips2.nrao.edu>, respectively.

Commercial Software

In recent years the number of commercial companies offering software for image deconvolution and restoration for astronomy has increased dramatically. Some companies offer software directly specialized while others offer software that includes image deconvolution as a part of a general package. Here we present only some of them.

Special software for ME image restoration (MEMSYS) has been commercially available for a long time. The MEMSYS package is produced by Maximum Entropy Data Consultants Ltd. (MEDC) (see <http://www.maxent.co.uk>). MEDC was created in 1981 by two Cambridge University academics: John Skilling, of the Department of Applied Mathematics and Theoretical Physics, and Steve Gull of the Department of Physics. The company specializes in custom and semicustom software for Bayesian Data Analysis, including the maximum entropy method and massive inference. Applications include image processing, using either direct or Fourier data (deconvolution, reconstruction from sparse data, motion deblurring, etc.). MEMSYS is also available through the Starlink system (<http://star-www.rl.ac.uk>).

Multi Resolutions Ltd. (<http://www.multiresolution.com>) offers the commercial wavelet packages MR/1 and MR/2 which include many deconvolution methods such as Lucy, MEM, the regularized Lucy method, and the multiscale entropy method for demanding and high-performance scientific, engineering, medical, financial, media and biological applications.

Pixon LLC (<http://www.pixon.com/>) offers software to restore images using the pixon approach. The company offers software for several geometric-based data sets such as images. Applications include image reconstruction in astronomy. The software also includes application to other imaging systems such as medical, geophysical, surveillance, radar, and sonar.

Most of the foregoing packages run on PC Windows platform and here are a few more: CCDSOFT of Software Bisque (<http://www.bisque.com>) which includes the R-L algorithm; MAXIMDL of Cyanogen Productions Inc. (<http://www.cyanogen.com>) which includes a version of the ME method and ASTROART of MSB software (<http://www.msb-astroart.com>) which includes versions of R-L and ME methods.

Online Information

A few web areas which can be used as a starting point in the search for data, bibliographies, and software are as follows.

AstroWeb: Astronomy/Astrophysics Resources on the Internet can be accessed from <http://www.cv.nrao.edu/fits/www/astronomy.html> and a number of mirror sites. AstroWeb currently contains about 3000 distinct resources.

The *American Astronomical Society*, <http://www.aas.org>, can be used as starting point for major meetings and other resources. The International Astronomical Union, the association of professional astronomy, is at <http://www.iau.org> and includes an online directory of members worldwide.

The most useful tool for searching and retrieving bibliographies is *The NASA Astrophysics Data System*. The Astrophysics Data System (ADS) is a NASA-funded project with an Abstract Service, which includes four sets of abstracts: 1) astronomy and astrophysics, 2) instrumentation, 3) physics and geophysics, and 4) Los Alamos preprint server. Each data set can be searched by author, object name (astronomy only), title, or abstract text words. In addition, it is possible to retrieve scanned images of over 40,000 articles appearing in most of the major astronomical journals. This service is available at <http://adsabs.harvard.edu>. A preprint repository is at <http://xxx.lanl.gov>.

Optimization of the Results

The optimization of the algorithms presented above requires us to address several questions related to the speed of convergence of the iterative algorithms toward the maximum, the relative weight of the prior and the conditional probability and the stopping point of the maximization/minimization.

Acceleration of the Algorithms

Almost all the iterative image reconstruction algorithms have the problem of slow convergence rate. This is especially important in the algorithms based on the EM algorithm and on the successive substitutions.

Techniques for accelerating the classical R-L method have been discovered independently by several authors (see, for example, [71], [24], and [28]). The basic idea is to substitute (73) by

$$f_i^{(k+1)} = f_i^{(k)} + A \left[\frac{f_i^{(k)}}{q_i} \sum_{j=1}^p H_{ji} \left(\frac{g_j}{\sum_{l=1}^B H_{jl} f_l^{(k)}} - 1 \right) \right], \\ i=1, \dots, B \quad (79)$$

where the value of A is chosen to maximize the likelihood given by (18). If $A=1$, then (79) reduces to the standard R-L algorithm. Interesting ideas on acceleration have also been applied to the FMAPE algorithm and described in [48] and [49].

Adjustment of the Hyperparameter and Stopping Rules

An image is defined as strongly feasible if the data can be considered as a Poisson (or Gaussian) sample with means given by the projection of the image. An image is defined as weakly feasible of order n if the first n moments of the distribution are consistent with the Poisson/Gaussian hypothesis. Tests based on the feasibility criteria have been implemented. The most usual is the χ^2 test which for pure Poisson data has the form

$$\chi^2 = \sum_{j=1}^p \frac{(g_j - (\mathbf{Hf})_j)^2}{(\mathbf{Hf})_j} \approx p.$$

These tests give as a result the desired Bayesian hyperparameter and/or a stopping point for the algorithms. To work well, however, two conditions should be met: a) the hypothesis about the data (Poisson/Gaussian) should be satisfied by the true statistics of the data, and b) the PSF should be perfectly known.

The feasibility tests work well with computer generated data in which it is possible to control the statistics and the PSF is exactly known. The feasibility tests often fail with real data, however, because the nature of the statistics and the PSF are not sufficiently well known. Disturbances caused by readout noise, corrections for

geometric distortion, flatfield corrections, detector failures, reseaux marks, etc. contribute to that. Also, the image of a field star, often used as PSF, has its own noise, background, different color, etc. A robust alternative to feasibility is cross validation.

The likelihood cross validation [8] is a new approach that answers the questions posed above in a more robust way. Núñez and Llacer have applied the cross-validation technique successfully in both medical emission tomography and astronomy (see [47] and [49] for details).

The minimization of the functional $M(\alpha(\mathbf{f}), \mathbf{f})$ defined in (58) together with the estimation of the weighting matrices there and the corresponding hyperparameter have been addressed by Katsaggelos et al. ([22], see also [23] and [1]).

Finally, we would like to mention that it is also possible to treat the unknown parameters of the image and degradation models as realizations of given distributions and then apply the so-called hierarchical Bayesian approach to estimate both the image and the unknown parameters (see [34] for the application to astronomical images and [37] for the general framework).

Acknowledgments

This work has been partially supported by the Comisión Nacional de Ciencia y Tecnología under contracts TIC97-0989 and PB97-0903. The authors thank the anonymous reviewers for their comments and Fionn Murtagh for proofreading the article.

Rafael Molina was born in Granada in 1957. He received the degree in mathematics (statistics) and the Ph.D. degree in mathematics from the University of Granada in 1979 and 1983, respectively. He became Professor of Computer Science and Artificial Intelligence at the University of Granada in 2000 and is currently the Dean of the Computer Engineering Faculty. His areas of research interest are image restoration (applications to astronomy and medicine), parameter estimation, image and video compression, and blind deconvolution. He is a member of SPIE, Royal Statistical Society, and the Spanish Association of Pattern Recognition and Image Analysis.

Jorge Núñez received the B.A. and the Ph.D. degrees in physics from the University of Barcelona, Spain in 1975 and 1981, respectively. Since 1984 he has been a Professor at the Department of Astronomy of the University of Barcelona. In 1998 he was selected member of the Academy of Sciences and Arts of Barcelona. Presently, he is conducting research on digital image processing including image reconstruction and restoration using Bayesian techniques and data fusion with applications in astronomy, medicine, and remote sensing.

Francisco José Cortijo was born in Valencia, Spain, in 1967. He received the M.S. and the Ph.D. degrees in

computer science from the University of Granada in 1990 and 1995, respectively. He is currently a Professor at the Department of Computer Science and Artificial Intelligence at the University of Granada. His research work is with the Digital Image Analysis and its Applications group at the University of Granada. His main fields of interest are image classification and restoration, learning and pattern recognition. He is a member of the Spanish Association of Pattern Recognition and Image Analysis and the International Association for Pattern Recognition.

Javier Mateos was born in Granada, Spain, in 1968. He received the M.S. and Ph.D. degrees in computer science at the University of Granada in 1991 and 1998, respectively. Since September 1992, he has been Assistant Professor at the Department of Computer Science and Artificial Intelligence of the University of Granada. His research interests include image restoration and image and video recovery and compression. He is a member of the Spanish Association of Pattern Recognition and Image Analysis and the International Association for Pattern Recognition.

References

- [1] M.R. Banham and A.K. Katsaggelos, "Digital image restoration," *IEEE Signal Processing Mag.*, vol. 14, pp. 24-41, 1997.
- [2] A. Bijaoui, J.L. Starck, and F. Murtagh, "Restauration des images multi-échelles par l'algorithme à trous," *Traitem. du Signal*, vol. 11, pp. 229-243, 1994.
- [3] L. Blanc-Féraud and M. Barlaud, "Edge preserving restoration of astrophysical images," *Vistas Astron.*, vol. 40, pp. 531-538, 1996.
- [4] T.J.R. Bontekoe, E. Koper, and D.J.M. Kester, "Pyramid maximum entropy images of IRAS survey data," *Astron. Astrophys.*, vol. 284, pp. 1037-1053, 1994.
- [5] R. Bouanno, G. Buscema, C.E. Corsi, I. Ferraro, and G. Iannicola, "Automated photographic photometry of stars in globular clusters," *Astron. Astrophys.*, vol. 126, pp. 278-282, 1983.
- [6] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *IEEE Trans. Image Processing*, vol. 6, pp. 298-311, 1997.
- [7] M.K. Charter, "Drug absorption in man, and its measurement by Max-Ent," in *Maximum Entropy and Bayesian Methods, Proc. of the 9th International Workshop*, P.F. Fougeres, Ed. Norwell, MA: Kluwer, 1990, pp. 325-339.
- [8] K.J. Coakley, "A cross-validation procedure for stopping the EM algorithm and deconvolution of neutron depth profiling spectra," *IEEE Trans. Nuclear Sci.*, vol. 38, pp. 9-15, 1991.
- [9] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *J. Royal Stat. Soc. B*, vol. 39, pp. 1-38, 1977.
- [10] S. Djorgovski, "Modelling of seeing effects in extragalactic astronomy and cosmology," *J. Astrophys. Astron.*, vol. 4, pp. 271-288, 1983.
- [11] B.R. Frieden, "Restoring with maximum likelihood and maximum entropy," *J. Opt. Soc. Amer.*, vol. 62, pp. 511-518, 1972.
- [12] B.R. Frieden, "Unified theory for estimating frequency-of-occurrence laws and optical objects," *J. Opt. Soc. Amer.*, vol. 73, pp. 927-938, 1983.
- [13] B.R. Frieden, "Stochastic approaches to inversion problems in optics," *J. Applied Stat.*, vol. 16, pp. 243-266, 1989.
- [14] S.F. Gull, "Development in maximum entropy data analysis," in *Maximum Entropy and Bayesian Methods*, J. Skilling, Ed. Dordrecht: Kluwer, 1988, pp. 53-71, 1988.
- [15] S.F. Gull and G.J. Daniell, "Image reconstruction from incomplete and noisy data," *Nature*, vol. 272, pp. 686-690, 1978.
- [16] R.J. Hanisch and R.L. White, "The restoration of HST images and spectra-II," in *Proc. Workshop held at the Space Telescope Science Institute*, Baltimore, MD, Nov. 1993.
- [17] F.B. Hildebrand, *Introduction to Numerical Analysis*, 2nd ed. New York: McGraw Hill, 1974.
- [18] M. Holschneider, R. Kronland-Martinet, J. Morlet, and Ph. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in *Wavelets: Time-Frequency Methods and Phase Space*, J.M. Combes, A. Grossman, and P. Tchamitchian, Eds. Berlin, Germany: Springer-Verlag, 1989, pp. 286-297.
- [19] R. Hook, "An overview of some image restoration and combination methods," *ST-ECF Newsletter*, vol. 26, pp. 3-5, 1999.
- [20] M.G. Kang and A.K. Katsaggelos, "General choice of the regularization functional in regularized image restoration," *IEEE Trans. Image Processing*, vol. 4, pp. 594-602, 1995.
- [21] A.K. Katsaggelos, Ed., *Digital Image Restoration*. New York: Springer-Verlag, 1991.
- [22] A.K. Katsaggelos, M.G. Kang, and M.R. Banham, "Iterative evaluation of the regularization parameter applied to HST images" in *Proc. The Restoration of HST Images and Spectra* R.J. Hanisch and R.L. White, Eds. Space Telescope Science Institute, 1994, pp. 3-13.
- [23] A.K. Katssagelos and M.G. Kang, "A spatially adaptive iterative algorithm for the restoration of astronomical images," *Int. J. Imaging Syst. Technol.*, vol. 6, pp. 305-313, 1995.
- [24] L. Kaufman, "Implementing and accelerating the EM algorithm for positron emission tomography," *IEEE Trans. Med. Imag.*, vol. MI-6, pp. 37-51, 1987.
- [25] R.L. Lagendijk and J. Biemond, *Iterative Identification and Restoration of Images*. Norwell, MA: Kluwer, 1991.
- [26] L.B. Lucy, "An iterative technique for the rectification of observed distribution," *Astron. J.*, vol. 79, pp. 745-754, 1974.
- [27] L.B. Lucy, "Image restoration of high photometric quality," in *Proc. The Restoration of HST Images and Spectra*, R.J. Hanisch and R.L. White, Eds. Space Telescope Science Institute, 1994, pp. 79-85.
- [28] L.B. Lucy, "Optimum strategies for inverse problems in statistical astronomy," *Astron. Astrophys.*, vol. 289, pp. 983-994, 1994.
- [29] P. Magian, F. Courbin, and S. Sohy, "Deconvolution with correct sampling," *Astrophys. J.*, vol. 494, pp. 472-477, 1998.
- [30] G.J. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*. New York: Wiley, 1997.
- [31] E.S. Meinel, "Origins of linear and nonlinear recursive restoration algorithms," *J. Opt. Soc. Amer. A*, vol. 3, pp. 787-799, 1986.
- [32] T.R. Metcalf and H.S. Hudson, "Pixon-based multiresolution image reconstruction of YOHKO's hard X-ray telescope," *Astrophys. J.*, vol. 466, pp. 585-594, 1996.
- [33] A.F.J. Moffat, "A theoretical investigation of focal stellar images in the photographic emulsion and application to photographic photometry," *Astron. Astrophys.*, vol. 3, p. 455, 1969.
- [34] R. Molina, "On the hierarchical Bayesian approach to image restoration. Applications to astronomical images," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, pp. 1122-1128, 1994.
- [35] R. Molina and F.J. Cortijo, "On the Bayesian deconvolution of planets," in *Proc. Int. Conf. Pattern Recognition, ICPR'92*, 1992, vol. 3, pp. 147-150.
- [36] R. Molina, A.K. Katsaggelos, J. Mateos, and J. Abad, "Compound Gauss Markov random fields for astronomical image restoration," *Vistas Astron.*, vol. 40, pp. 539-546, 1996.

- [37] R. Molina, A.K. Katsaggelos, and J. Mateos, "Bayesian and regularization methods for hyperparameter estimation in image restoration," *IEEE Trans. Image Processing*, vol. 8, pp. 231-247, 1999.
- [38] R. Molina, A.K. Katsaggelos, J. Mateos, A. Hermoso, and A. Segall, "Restoration of severely blurred high range images using stochastic and deterministic relaxation algorithms in compound Gauss Markov random fields," *Pattern Recognit.*, vol. 33, pp. 555-571, 2000.
- [39] R. Molina, J. Mateos, and J. Abad, "Prior models and the Richardson-Lucy restoration method," in *Proc. The Restoration of HST Images and Spectra*, R.J. Hanisch and R.L. White, Eds. Space Telescope Science Institute, 1993, pp. 118-122.
- [40] R. Molina, J. Mateos, J. Abad, N. Pérez de la Blanca, A. Molina, and F. Moreno, "Bayesian image restoration in astronomy. Applications to images of the recent collision of comet Shoemaker-Levy 9 with Jupiter," *Int. J. Imag. Syst. Technol.*, vol. 6, pp. 370-375, 1995.
- [41] R. Molina, A. del Olmo, J. Perea, and B.D. Ripley, "Bayesian deconvolution in optical astronomy," *Astron. J.*, vol. 103, pp. 666-675, 1992.
- [42] R. Molina, N. Pérez de la Blanca, and B.D. Ripley, "Statistical restoration of astronomical images," *Data Analysis in Astronomy III*, V. di Gesu, L. Scarsi, P. Crane, J.H. Friedman, S. Levialdi and M.C. Maccarone, Eds. New York: 1988, pp. 75-82.
- [43] R. Molina, and B.D. Ripley, "Using spatial models as priors in astronomical images analysis," *J. Appl. Statist.*, vol. 16, pp. 193-206, 1989.
- [44] 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," *Astron. J.*, vol. 104, pp. 1662-1668, 1992.
- [45] R. Narayan and R. Nityananda, "Maximum entropy image restoration in astronomy," *Astron. Astrophys.*, vol. 24, pp. 127-170, 1986.
- [46] Special Issue on Image Reconstruction and Restoration in Astronomy, *Int. J. Imag. Syst. Technol.*, vol 6, no. 4, pp. 295-420, 1995.
- [47] J. Núñez and J. Llacer, "A general Bayesian image reconstruction algorithm with entropy prior. Preliminary application to HST data," *Pub. Astron. Soc. Pacific*, vol. 105, pp. 1191-1208, 1993
- [48] J. Núñez, and J. Llacer, "HST image restoration with variable resolution", in *Proc. The Restoration of HST Images and Spectra*, R.J. Hanisch, and R.L. White, Eds., Space Telescope Science Institute, 1994, pp. 123-130, 1994.
- [49] J. Núñez and J. Llacer, "Bayesian image reconstruction with space-variant noise suppression," *Astron. Astrophys. Supplement Series*, vol. 131, pp. 167-180, 1998.
- [50] R.K. Piña and R.C. Puetter, "Incorporation of the spatial information in Bayesian image reconstruction. The maximum residual likelihood criterion," *Pub. Astron. Soc. Pacific*, vol. 104, pp. 1096-1103, 1992.
- [51] R.K. Piña and R.C. Puetter, "Bayesian image reconstruction: The pixon and optimal image modeling," *Pub. Astron. Soc. Pacific*, vol. 105, pp. 630-637, 1993.
- [52] N. Pirzkal, R.N. Hook, and L.B. Lucy, "Two channel photometric restoration algorithm—GIRA," *ST-ECF Web Newsletter*, no. 26 [Online]. Available: <http://ecf.hq.eso.org/newsletter/webnews1/gira>
- [53] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C*, 2nd. ed. Cambridge, U.K.: Cambridge Univ. Press, 1996.
- [54] R.C. Puetter, "Pixon-based multiresolution image reconstruction and the quantification of picture information content," in *Int. J. Imaging Syst. Technol.*, vol. 6, pp. 314-331, 1995.
- [55] R.C. Puetter, "Information, language, and pixon-based image reconstruction," in *Nonlinear Signal and Image Analysis*, J. R. Buchler and H. Kandrup, Eds. New York: New York Academy of Sciences, 1997, vol. 808, pp. 160-183.
- [56] R.C. Puetter, "Information, language, and pixon-based image reconstruction" *Proc. SPIE*, vol. 2827, pp. 12-31, 1996.
- [57] W.H. Richardson, "Bayesian-based iterative method of image restoration," *J. Opt. Soc. Amer.*, vol. 62, pp. 55-59, 1972.
- [58] B.D. Ripley, *Spatial Statistics*. New York: Wiley, 1981, pp. 88-90.
- [59] M.J. Shensa, "Discrete wavelet transforms: Wedding the à trous and Mallat algorithms," *IEEE Trans. Signal Processing*, vol. 40, pp. 2464-2482, 1992.
- [60] L.A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography," *IEEE Trans. Medical Imag.*, vol. 1, pp. 113-122, 1982.
- [61] J. Skilling, "Classic maximum entropy," in *Maximum Entropy and Bayesian Methods (8th Workshop)*, J. Skilling, ed. Norwell, MA: Kluwer, 1989, pp. 45-52.
- [62] J. Skilling and R.K. Bryan, "Maximum entropy image reconstruction: General algorithm," *Monthly Notices Royal Astrotron. Soc.*, vol. 211, pp. 111-124, 1984.
- [63] J-L. Starck and F. Murtagh, "Multiscale entropy filtering," *Signal Processing*, vol. 76, pp. 147-165, 1999.
- [64] J-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis. The Multiscale Approach*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
- [65] J-L. Starck and E. Pantin, "Multiscale maximum entropy image restoration," *Vistas Astron.*, vol. 40, pp. 563-571, 1996.
- [66] R. Stevenson and E. Delp, "Fitting curves with discontinuities," in *Proc. 1st Int. Workshop Robust Computer Vision*, Oct. 1990, pp. 127-136.
- [67] D.M. Titterington, "General structure of regularization procedures in image reconstruction," *Astron. Astrophys.*, vol. 144, pp. 381-387, 1985.
- [68] Y. Vardi, L.A. Shepp, and L. Kaufman, "A statistical model for positron emission tomography (with discussion)," *J. Amer. Statist. Assoc.*, vol. 80, pp. 8-37, 1985.
- [69] N. Weir, "A multi-channel method of maximum entropy image restoration," in *Astronomical Data Analysis Software and Systems I, A.S.P. Conference Series*, D.M. Worrall, C. Biemesderfer, and J. Barnes, Eds., 1992, vol. 25, p. 186.
- [70] R.L. White and R.J. Allen, Eds., *The Restoration of HST Images and Spectra, Proc. Workshop held at the Space Telescope Science Institute*, Baltimore, MD, Aug. 1990.
- [71] R.L. White, "Improvements to the Richardson-Lucy method," *Newsletter STScI's Image Restoration Project*, no. 1, pp. 11-23, 1993.
- [72] R.L. White, "Image restoration using the damped Richardson-Lucy method," in *Proc. The Restoration of HST Images and Spectra*, R.J. Hanisch and R.L. White, Eds., 1994, pp. 104-110.
- [73] N.J. Woolf, "High resolution imaging from the ground," *Ann. Rev. Astron. Astrophys.*, vol. 20, pp. 367-398, 1982.
- [74] A.T. Young, "Seeing: its cause and cure," *Astrophys. J.*, vol. 189, pp. 587-604, 1974.